MOTIVATION-BASED INTEREST RECOGNITION USING DIGITAL PHENOTYPING

A THESIS SUBMITTED TO THE UNIVERSITY OF MANCHESTER
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE FACULTY OF SCIENCE AND ENGINEERING

2022

By
Ahmed Ibrahim

Department of Computer Science

Contents

Word Count: 66991

List of Tables

List of Figures

Abstract

This work aims to recognise people's interests from their daily behaviours. Recognising personal interests contributes to the production of personalised behavioural interventions, which would have more traction and motivational value when compared to general interventions. Typically, self-reporting methods are used to understand people's interests. These methods rely on people's perception; despite that, in most cases, interests are demonstrated in an individual's daily activity. Moreover, self-reporting tools are discrete and hence do not capture interest dynamics, which require continuous observation; attempts to overcome this weakness through longitudinal analysis can be highly intrusive and prone to memory and recall biases.

Digital devices such as smartphones and wearables can overcome such limitations and hence have the potential to capture interests from daily behaviour in a continuous, longitudinal and unobtrusive (passive) manner. However, the daily routine is not only formed from actions motivated by personal interests. Instead, many of our daily actions are motivated by other reasons such as obligations and external rewards. Therefore, understanding the motives behind our daily activities is essential to distinguish behaviours driven by personal interests from those motivated by other factors.

In this work, we create a framework for recognising personal interests using smartphones. We create an approach that first derives behavioural features of individuals' daily routines from their smartphones' data (digital phenotyping). Then, we employ knowledge of human motivation to (1) infer interests without recourse of asking (unobtrusive) and (2) adapt to newly developed interests. We have conducted real-world experiments to inform and assess our method. The conducted studies were designed to longitudinally and continuously observe behaviours while people undertake their daily life. Our results showed the advantage of basing the recognition of personal interests on motivational knowledge. Compared to baseline methods, our approach significantly improved the recognition of interests by an average of 62% with $p < 0.05$.

The in-depth understanding of interests can be of value for personalisation in domains such as precision medicine and behavioural nudges. Future work can build upon our effort and measure the enhancement it may add to these domains. Moreover, the techniques applied in this work can be further investigated to infer similar cognitive phenomena.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright

Acknowledgements

I thank the Almighty Allah for his countless blessings. It would not have been possible for me to complete this work had it not been for the grace and will of Almighty Allah.

I want to thank my supervisor, Professor Simon Harper, for choosing me to do this research and helping me in it in the best way possible. My sincere gratitude and appreciation to you, Simon, for providing me with the chance to work with you and benefit from your valuable experiences. Thanks for all the advice, which was not limited to the academic aspect but also included the social and family aspects. Your understanding of these aspects and your advice in this regard helped me adapt and reconcile my duties and responsibilities. I could not have had and asked for a better supervisor. I extend my sincere gratitude and thanks to Dr Sarah Clinch, who provided valuable guidance during this PhD trip, whether related to the purely technical aspect or other aspects such as academic writing. I have never felt that you dealt with this research as a co-supervisor. You were always there in the meetings, discussing the ideas, providing precious inputs and taking the time to help whenever I needed that.

I thank my friends from and outside the Interaction Analysis and Modelling Laboratory (IAM) for making this PhD journey special. I also thank Dr Markel Vigo for sparing time for me whenever I needed assistance and Dr Bijan Parsia for the valuable inputs that he provided to this work.

My special thanks, love and gratitude to my mother. Thank you for all the support, prayers and love, without which this work would not have been possible. To my greatest uncles, Mohammad and Ali, thank you for all the support and guidance that you gave during the easy and challenging times of my life. To my brothers and sisters, I love you and thank you for all the support and love. To my wife and children, I would not have done and completed this work without you. Your support, love and presence with me during this journey made it unforgettable.

Chapter 1

Introduction

This work explores the use of smartphone-based digital phenotyping in recognising the personal interests of an individual. Digital phenotyping is the process of quantifying features of the person's daily behaviours from digital devices' data collected continuously, in the wild, and without the need to ask individuals (Onnela and Rauch, 2016; Vega-Hernandez, 2019). In this work, we quantify features of mobility, phone usage and buying behaviours in order to recognise interests and understand their dynamics. The work is the first step within a broader vision that seeks to benefit from individual interests in personalising behavioural nudging. Personalised nudges are hypothesised to work better in changing behaviour when compared to the generalised ones, as the former would have more traction and motivational value (Schoning et al., 2019; Mills, 2020). Using interests for such personalisation can encompass more than one field, whether educational, therapeutic, or professional.

The term interest is commonly used in our everyday life which can make its meaning sometimes too vague (Silvia, 2007; Ahmed and Srivastava, 2019). However, from cognitive and psychological aspects, interest expresses a mental and affective state (Silvia, 2007). It can reflect spontaneous emotions triggered by people's interaction with their surroundings (situational interests) or can describe preferences and intrinsic motivations that drive a range of personal behaviours (individual interests) (Renninger and Hidi, 2016). The latter (which is the target of this thesis) reflects deep-seated tendencies and convictions that people manifest in some of their daily behaviours. Hence, individual interests are more enduring and intrinsic compared to situational interests (Silvia, 2007; Renninger and Hidi, 2011). Differentiating behaviours that embody those individual interests from other behaviours requires a deeper understanding of the human motivations behind each action. Through motivations' understanding, behaviours that are motivated by individual interests can be recognised and separated from those driven by external factors such as rewards and obligations.

Assessing and recognising interests has been the concern of multiple self-reporting tools across different psychological sub-disciplines and applications (e.g. Amabile et al., 1994; Tyler-Wood et al., 2010; Ryan, 2018). These tools are limited by the fact that the provided answers may not necessarily reflect individuals' real interests. Instead, these answers may express perceptions that are not reflected in daily behaviours, or replies that people think are more socially acceptable (Northrup, 1997; Paulhus and Vazire, 2007). The evolution and changes of interests (i.e. interest dynamics) is another aspect that is not addressed by self-reporting inventories. Although in most cases, intrinsic interests are demonstrated in an individual's daily activity; self-reporting tools are not designed to extract interest from daily behaviours nor to observe interest dynamics over time. Doing so would be highly intrusive and may significantly impact the ecological validity of studies (Vega-Hernandez, 2019). For example, in the study of (Memedi et al., 2015), participants were asked to answer seven questions four times a day which caused high dropout (up to 42%) and median compliance rate (93%).

The use of self-reporting tools has been extended beyond behavioural and psychological studies to include computer science research. In this case, technology is used as a tool to ask individuals about their interests. The reported interests can be employed differently according to the underlying objectives (e.g. personalising user interfaces (Adu et al., 2018) or behavioural recommendations (Meixner et al., 2020)). However, this use of self-reporting methods is just a digitisation of the traditional medium (usually papers) without addressing the drawbacks inherent in the essence of this method.

Another intersection between computer science and interest recognition is in the use of recommender systems. In these systems, interests are studied through recurrent time episodes that represent periods of interaction between the user and the platform within which the recommender systems operate (Khan et al., 2017; Quadrana et al., 2018). Next, recommendations based on users' interests are delivered in several forms such as targeted ads, personalised movies or preferred news (Ricci et al., 2011; Lops et al., 2011; Zhang et al., 2019). However, for recognising individual interests, recommender systems typically have three features that differentiate them from the problem at hand: (i) data capture relates to a highly constrained set of behaviours taking place on a specified platform (e.g. all interactions with the website `Amazon.com`); (ii) these data are sporadic and hence do not reflect the continuity of daily behaviour (iii) predictions are made based on prior behaviour itself, rather than the underlying motivations that led to those behaviours.

The advancement and popularity of digital devices (such as smartphones and wearables) have advanced the use of digital phenotyping, which in turn has the potential to mitigate self-reporting and recommender systems shortcomings. Although various studies have used digital phenotyping (e.g. Vega-Hernandez et al., 2017; Barnett et al., 2018), to our knowledge, none of them utilises digital phenotyping to recognise human interest, which is the subject of this research. We aim to recognise personal interests through smartphone-based digital phenotyping. Specifically, we study three behaviours that can reveal individual interests: mobility, phone usage, and buying behaviours.

For each one of the three behaviours, we rely on the data source that is highly indicative of it. We use GPS data as the source for mobility behaviour, apps' interactions for phone usage behaviour, and notifications to analyse buying behaviour. Existing literature on mobility (Zandbergen, 2009; Wang et al., 2012) and phone usage behaviours (Falaki et al., 2010; Harari et al., 2017) support our decision for respectively selecting GPS data and app interactions as the data sources. The decision to select notifications for buying behaviour is stemmed from three reasons. First, online buying activities are usually proceeded by a receipt that is delivered as an email, a text message or a notification from the related app. In all these cases, a notification is typically generated to alert the user about the delivery of the receipt. Second, buying activities that occur in stores can lead to the generation of a notification. In this case, users may provide the retailers with their emails to receive receipts which represents another chance of capturing buying behaviour. Third, using notifications instead of relying on a specific shopping app enables us to build our method around the behaviour rather than a single platform. Designing the method around a single app would make it unresponsive to users' adding new apps or removing existing ones.

We set our goal based on the presumption that there is a mechanism by which smartphones can be used to understand everyday behaviours and to extract interests from these behaviours. More specifically, we hypothesise that if we computationally employ knowledge of human motivation and use smartphones to capture behavioural data, we will be able to differentiate behaviours that are motivated by internal interests from the ones driven by external factors such as obligations and rewards. Behaviours of the former are embodiments of individual interests and hence can lead to their recognition. Our presumption is supported by (i) what experiments have shown in terms of the possibility and effectiveness of extracting behavioural knowledge through digital phenotyping (Vega-Hernandez et al., 2017; Barnett et al., 2018), (ii) the fact that the majority of people carry smartphones almost continuously (Torous et al., 2014; Poushter et al., 2016), and (iii) personal interests are manifested in an individual's daily activity (Silvia, 2007; Renninger and Hidi, 2016).

1.1 Research questions

The research questions of this PhD thesis are:

RQ1: What data do we need to understand interests, and how can we obtain them through digital phenotyping? Digital devices, such as smartphones, can be used to collect data of daily routines and behavioural features effectively. The collected data can lead to the identification of personal interests without asking people directly (unlike self-reporting). However, behaviours of the daily routine are encoded and not directly observed in the data that are passively collected from digital devices. Detecting those behaviours and their features can be done through digital phenotyping, and that needs to be according to each source's properties. For instance, extracting mobility behaviour based on location data is different from detecting buying behaviour using receipts received as smartphone's notifications. Therefore, we need to research the possible ways of decoding behavioural events from the collected raw data for each of the three behaviours regarded by this thesis. The extracted behavioural events can then be used as the basis for understanding individual interests. The use of digital phenotyping to derive behavioural knowledge can contribute to providing an understanding of interests through an unobtrusive observation.

RQ2: How can we recognise individual interests using digital phenotyping? Personal interests are ingrained desires within people. Identifying these interests requires observing people's behaviours while they engage in their daily events. However, since not all events reflect individual interests, understanding the motives behind the daily events of each behaviour is essential to recognise the ones driven by internal interests. Knowledge of human motivation can help in understanding the properties of behaviours motivated by interests. Therefore, researching possible ways to build computational models from the motivation knowledge can help in bringing this knowledge from its behavioural and psychological context to the fields of technology and informatics. Doing so can contribute to a better understanding of interests using unobtrusive means.

RQ3: Does combining events of mobility, phone usage and buying behaviours improve the accuracy of interest detection? Each one of the three behaviours considered by this thesis is formed of multiple events. Mobility behaviour, for instance, may reflect events such as going to a movies theatre or visiting a hospital, whereas phone usage behaviour may exhibit a web browsing or gaming activity. Recognising personal interests using a single behaviour might not be sufficient since people may exhibit their interests through various channels (e.g. watching a sport and buying a sport-related product). Therefore, after we determine the behavioural data necessary to understand interests (RQ1), and investigate how to detect individual interests using digital phenotyping (RQ2), we explore how we can combine and weigh multiple behaviours into a framework that shows to what level of accuracy does mobility, phone usage and buying behaviours can imply interest. This combinatory approach needs to be compared to the existing methods that either rely on a single or multiple behavioural streams to understand interests.

RQ4: How can we personalise the contribution of each behaviour and conceive changes in people's interests? Individual differences influence the nature and realisation of people's interests. Some people are more willing to explore and try new interests, whereas others find peace in enjoying interests that they are familiar with. Also, the degree to which each behaviour contributes to fulfilling personal interests differs according to each person. An individual may prefer to not use their phone for pursuing a specific interest (e.g. watching movies). At the same time, another person may not mind the means as long as they satisfy an underlying interest. In addition to the personal preferences of a specific source, individual differences can be caused by constraints on data collection. Examples of these constraints include the people's desire to limit the data of a specific channel (privacy), the type of information that can be gathered (technological), the compliance of the method with ethical requirements (ethical). Therefore, we need an approach that can capture the interest dynamics and consider variation in collection quality. Such an approach can contribute toward better adaptability and personalisation of interests detection.

1.2 Contributions

The main contributions of this thesis are:

1. **Knowledge that supports the understanding of data needed for interest detection through digital phenotyping.**

   Through two papers, we contribute to the understanding of interest-related behaviours using digital phenotyping. The two papers (detailed in Chapter 3 and 4) present the use of digital phenotyping within the context of mobility and buying behaviours. For the phone usage behaviour, we have conducted an extensive review of the methods used to derive events of phone usage in Chapter 2. The paper in Chapter 3 details a framework that aims to derive mobility behaviours from the location data. The presented framework is based on a systematic literature review and aims to bridge the theory and practical implementation gap. The second paper details and compares (based on real-world data) the various methods of recognising buying behaviour from smartphone notifications (Chapter 4). The knowledge detailed on the three behaviours helps determine what and how much data is needed to recognise interests (RQ1).

2. **A framework for recognising interests that is better than existing methods.**

   The two papers that we detail in Chapters 5 and 6 contribute to the production of a framework that improves the recognition of personal interests. In Chapter 5, we present a paper that introduces our Motivation-based Interest Recognition (MIR) approach and realises it with mobility behaviour. Chapter 6 details the paper that combines the three behaviours and apply the MIR to them as a combinatory method (cMIR). In both papers, we infer interests without recourse to ask individuals (unobtrusive). As a result, it has become possible to rely more on individuals' behaviour derived from digital devices (smartphone in this case) in identifying interests (objective) rather than relying solely on self-reporting. The development of such a framework answers the second and third research questions of this thesis. It shows how interests can be detected from passively collected data (RQ2) and how combining multiple behaviours can improve the recognition of personal interests (RQ3).

3. **An adaptable and personalised approach to recognise interests from digital phenotyping.**

This thesis also contributes an approach that adapts to changes in people's interests and personalises the process of interests detection. The papers that we present in Chapter 6 shows the adaptivity of our approach according to each individual's data. By tailoring our combinatory MIR method according to an individual's data, we provide a flexible method capable of producing a personalised model from an overfitted set of motivation properties forming the general approach. The model, as it is adapted for one person, is not meant to be transferred to another. Also, the tool in Chapter 7 depicts how interests are changing over time according to each individual's data. The two papers contribute to the production of an adapted and personalised understanding of interests and their dynamics through digital phenotyping (RQ4).

1.3 Thesis overview

With the permission of the supervisory team from the Faculty of Science and Engineering, this thesis is presented in a journal/alternative format. This means that the main chapters (Chapters 3 to 7 inclusive) of this work are papers that are published or currently under review. Chapter 8 contains a sixth paper as a case study and part of the overall discussion of the presented work. Our selection of the journal format is driven by the ability to read and comprehend each paper individually. However, the chapters containing these papers connectedly contribute to the overall picture that forms the subject of this work.

In this work, we have focused on detecting people's interests using three behaviours: mobility, phone usage and buying. The selection of these behaviours stemmed from the fact that people are willing to move to different places (a physical cost), spend time (a mental cost) and pay money (a financial cost) in their pursuit of interests (Renninger and Hidi, 2016; Ryan and Deci, 2017). Respectively, mobility, phone usage and buying behaviours are proxies to these and hence were chosen as the interest detection base.

Our work has three main phases: 1) The feasibility phase, where we conducted a secondary analysis on a smartphone dataset to explore initial parameters necessary to identify interests. 2) A formative phase, where we set up our study's protocol and data collection tools. In this phase, we have conducted a three-month pilot study to produce initial results and inform our method. 3) A summative phase, where we performed a longitudinal study that lasted for six months. A summative evaluation has been conducted to assess the approach that we have developed based on the insights from the previous two phases.

1.3.1 Feasibility phase

Our feasibility stage relied on secondary analysis based on a dataset previously captured from seven adults using the AWARE mobile sensing framework[1] (Ferreira et al., 2015). We used this phase to inform our study design decisions. The dataset, collected over one year, contains 2.8 million measures of location and 22.8 thousand app usage datapoints that are passively sensed from two Android and five iOS devices (Vega-Hernandez, 2019). However, the dataset does not contain interest-specific ground truth nor notification texts to help with exploring buying behaviour. Also, the app usage data points were collected from two participants (out of the seven), which limited the exploration of the phone usage behaviour. Nonetheless, the locations measures were a good starting point as they were collected from all seven participants and covered most of the study period.

As a result, we focused on mobility behaviour and how its features can be derived from smartphones to understand specific behavioural or cognitive phenomena, including interest. We culminated that with the paper titled *"From GPS to Semantic Data: How and Why"* which proposes a structural framework for understanding mobility behaviours from the smartphone's GPS sensor (Chapter 3). The paper is based on a systematic review of the relevant literature as well as our initial exploration. It discusses the process of semantically enriching raw location data to extract features of mobility behaviour (RQ1).

The location data of the secondary dataset also helped the initial investigation of interest-related aspects, such as: the required duration to spot personal interests, the basis for selecting the top N items from a set of ranked interests, and the relations between potential variables that determine interests. The initial findings guided our planning and execution of the formative phase described next.

1.3.2 Formative phase

The main goals of this phase were to explore the technological challenges, examine our modelling and design decisions, suggest appropriate measures for tunable parameters and leverage our gained knowledge to plan the main study. While the feasibility phase relied on secondary data analysis, we have conducted a pilot study and collected

[1] https://awareframework.com

a novel dataset from new participants in our exploration phase. We mirror the method used by (Vega-Hernandez, 2019), using AWARE (Ferreira et al., 2015) to collect location readings, app usage data, and notification texts.

We benefited from the pilot study in investigating the methods of extracting phone usage and buying behaviours, respectively, from apps' data and notification texts. As the former is well studied (as we shall see in Chapter 2), extracting buying behaviour from the notification text was yet to be investigated; therefore, we used the collected data to explore it. In our endeavour, we have submitted a paper titled *"Extracting Behavioural Features From Smartphone Notifications"* (Chapter 4). The paper provides end-to-end processing of notifications to understand behavioural aspects. We apply knowledge-based and machine learning techniques to those notifications to assess the recognition of buying behaviour from smartphone notifications (RQ1).

Besides the phone usage and buying behaviours, and as we started investigating the location data from the feasibility phase, the pilot study helped initialise the assessment of our modelling with the mobility behaviour. We published the results in a paper titled *"Recognising Intrinsic Motivation using Smartphone Trajectories"* (Chapter 5). The paper details our Motivation-based Interest Recognition (MIR) approach that infers personal interests from continuously- and passively-sensed smartphones location data (RQ2).

1.3.3 Summative phase

Guided by insights from both phases, we conducted our main study to perform a summative evaluation of our combinatory MIR (cMIR) that goes beyond a single behaviour (i.e. mobility) to integrate the three. Our evaluation included longitudinal mobile data collected – using AWARE – from eight participants going about their normal daily activities. The details, methods and results of our main study have been elaborated in a paper titled *"Recognising Personal Interests: A Combinatory Approach based on Smartphone-derived Behaviours and Intrinsic Motivation"* (Chapter 6). The paper shows how the multiplicity of behavioural streams improves the recognition of personal interests (RQ3). It also shows how the cMIR can be personalised and adapted based on each participant's data in addition to the integration of the three behaviours (RQ4).

To implement our approach, we develop a web-based tool that extracts behavioural events from the raw data and applies the proposed model accordingly. A paper about the tool titled *"Smartphone Data Analytics: A Behaviour and Motivation Centric Implementation"* is included in Chapter 7. The tool visualises the results of applying each one of the measurements used to model human interest and depicts the dynamics of the interest over time (RQ4). Visualising the individual measures and their integration facilitates the comparison of each measure's impact on the overall modelling of interest.

We conclude the thesis by solidifying the knowledge that has been produced (Chapter 8). We explain the advantages and limitations of our work and the suggested improvements for future research opportunities.

Chapter 2

Background and Related Work

Detection of interest forms the goal of this thesis, whereas motivation is the means by which behaviours motivated by personal interests can be detected and distinguished from other behaviours. Therefore, before embarking on discussing the related work, we present the theoretical underpinnings on which this thesis is based. Specifically, we cover interest and motivation. This primer helps explain the theoretical pillars of the research and the grounds on which we chose these pillars.

After the primer, we introduce a detailed review of current studies that use smartphone's digital phenotyping. The provided review serves the goal of introducing the work related to understanding daily behaviours from smartphones data without recourse to ask people. In this regard, we present an extensive survey of the literature related to the three behaviours considered by this thesis: mobility, buying and phone usage behaviours. The review details existing methods of forming behavioural events from smartphone's raw data. These events (a.k.a. behavioural items or units) are the basis for recognising interests from human behaviour.

We conclude this chapter by detailing existing methods of recognising interests from smartphones' data (using the extracted behavioural events). Specifically, we present a comprehensive review of the existing studies that base their detection of interests on smartphone-derived data. We analyse the detection mechanisms used by the studies and compare their works to the overall goal of this thesis. Accordingly, we identify the existing shortcomings and the research gap that this work aims to fill.

2.1 Theoretical underpinnings and primer

In this subsection, we provide the background necessary to understand human interest and motivation. Although the papers included throughout the thesis share some of the information presented in this subsection, the knowledge detailed here sets the stage for a better and broader understanding of the theoretical background. It shows the relationship between human interest and motivation from a theoretical perspective, which influenced our work's modelling and implementation details.

2.1.1 Human interest

The term *interest* may refer to several meanings in our everyday life, which can sometimes be too vague (Savickas, 1999; Benedict, 2001; Silvia, 2007). Saying that "we live in an *interesting* time due to COVID-19" indicates a meaning that is different from "we are not *interested* in eating at this place". The former refers to the particularity of the COVID-19 period, whereas the latter expresses a motivation state (Silvia, 2007). Interest also has other meanings – such as the financial one – that are outside psychology and this thesis's scope.

Psychologists' study of interest has a long and shifting history (Silvia, 2007). In the early nineties, research investigated the role of interest in the educational, motivational and cognitive fields (Arnold, 1905; Dewey, 1913). The advent of behaviourism[1] in the twenties of the nineteenth century contributes to hindering psychological studies on interest until a group of psychologists revive those studies between the period of 1960s and 1970s (Silvia, 2007). Since then, there have been many studies aimed at understanding interest (e.g. Tomkins, 1962; Schumacher, 1963; Berlyne, 1966). These studies slowly began to revolve around two main areas: (1) interest as part of the emotion and (2) interest as a determinant part of the individual's personality, goals and dispositions (Silvia, 2007). Eventually, researchers started to refer to the first as situational interests and the second as individual interests (Hidi, 1990; Krapp et al., 1992; Schraw et al., 1995; Silvia, 2007). In the next subsections, we detail each one and discuss the methods used to measure (i.e. operationalise) them.

[1]Behaviourism is a behavioural psychology school that focuses only on observable stimulus-response behaviors.

2.1.1.1 A primer on situational interest

Situational interest is defined as a temporary liking or mindful engagement in an activity (either mental or physical) that arises due to environmental factors (Schraw et al., 2001). Triggers of situational interests exist outside the self in the environment, which is the key feature of this type (Krapp et al., 1992). Responses to these triggers are mediated through abstract qualities that exist in individuals, which determine the arousal of interest feelings. Interest emotion that arises is assumed to be temporal and linked to specific contexts whose demise may lead to losing interest in the accompanying activity (Krapp et al., 1992; Schraw et al., 2001).

Since situational interest is activated by external environmental stimuli (i.e. not intrinsic to the self), it is measured by observing the reactions to these stimuli (Silvia, 2007). These reactions include facial expressions such as eyelid widening and head stillness (Reeve, 1993; Reeve and Nix, 1997; Li et al., 2017b). They also include vocal expressions such as loudness of speech, the pitch of the voice and speech rate (Frick, 1985; Scherer, 1986). As situational interests start to last longer, they may develop into enduring ones (Renninger and Hidi, 2011). If so, then traditional methods such as questionnaires or interviews become more appropriate.

Situational interest has a motivational value that plays a pivotal role in various domains (Silvia, 2007). Examples of that are: visiting the zoo to cultivate the student interest in biology class (Dohn et al., 2009); performing a group choral reading to motivate people with reading difficulties (Paige, 2011); and competing against time to boost individuals' motivation to complete tedious tasks such as copying letter (Sansone et al., 1992). In addition to its motivational value, situational interests can be the seeds that contribute to producing long-lasting individual interests which are discussed next.

2.1.1.2 A primer on individual interest

Individual interests are stable predispositions toward specific topics or objects that reflect a close association between positive feelings and those topics or objects (Ann Renninger, 2000; Schiefele, 2009). Activating those interests can be due to internal or external factors (Ann Renninger, 2000; Palmer et al., 2017). Unlike how external cues work on situational interests, however, external factors of individual interests activate inclinations that already exist within the self rather than create or help develop new ones (Schraw et al., 2001; Hidi and Renninger, 2006). On the other hand, internal cues are intrinsic stimuli latent in the self and motivate the pursuit of the positive feeling that comes from experiencing those interests (Ann Renninger, 2000; Schraw et al., 2001).

The possession of this type of interest often comes from the development of situational interests into enduring ones (Linnenbrink-Garcia et al., 2013). Studies on interest development discuss the process of internalising situational interests (i.e., making them intrinsic to the self) (Silvia, 2007). For instance, Hidi and Renninger (2006) proposes a four-phase model of interest development. Situational interests are first triggered through external stimuli. The transition to the second phase, "maintained situational interest", is observed through the willingness of a person to spend more time on the targeted object or topic. Developing curiosity and asking questions are indicators of emerging individual interests (the third phase). Lastly, interest is culminated as "individual" when a reliable and lasting association between the positive feeling and the object is established (Hidi and Renninger, 2006).

Typically, self-reporting inventories (e.g. Amabile et al., 1994; Tyler-Wood et al., 2010; Ryan, 2018) are used to assess individual interests (Ainley et al., 2002; Renninger and Hidi, 2016). Inventories of this form require individuals to respond to specific prompts, which can be about assessing positive feelings toward a single interest or selecting the most interesting object or subject from multiple alternatives (Schiefele, 2009; Ashton, 2013). Since individual interests are enduring dispositions, the detected interests are widely used in personalisation. This includes, for instance, tailoring behavioural interventions based on an individual's interest (Mohr et al., 2013; Rabbi et al., 2015) and improving the learning outcome through content personalisation (Walkington and Bernacki, 2014; Harackiewicz et al., 2016; Reber et al., 2018).

2.1.2 Human motivation

The motivation literature is replete with a large body of studies that seek, in its general direction, to create theories and frameworks capable of explaining the reasons behind our actions (McClelland, 1987; Weiner, 1992; Ryan and Deci, 2000). Some of these theories attribute the causes of our actions to purely physiological factors (e.g. Yerkes and Dodson, 1908; Hull, 1943), while others place these causes within a more comprehensive framework. This comprehensive framework integrates physiological factors with other psychological, cognitive and social constructs to explain the reasons that drive our behaviour (e.g. Maslow, 1943; Ryan and Deci, 2017). The emergence and ability of the latter trend to comprehend both cognitive and biological factors led to a departure from the reliance on physiological theories towards more cognitive and thorough explanations (Weiner, 2004; Strombach et al., 2016; Ryan and Deci, 2017).

In line with that, we have focused our work on studying and investigating these comprehensive theories and frameworks, which we refer to simply as psychological interpretations in this thesis.

Psychological interpretations, typically, relate motives to human behaviour in one of two approaches. *Static approaches* use a fairly rigid classification to match behaviour to underlying physiological or psychological needs. Examples include Murray's (1938) need theory and Maslow's (1943) hierarchy. These methods focus on the essence of actions, regardless of how individuals actualise them. Accordingly, behaviours are classified based on the relation of their nature to a set of human needs; such as food, housing, and family. For example, Maslow views actions related to eating, such as cooking, as physiological-driven, while classifying profession among the psychological needs that fuel the need to feel safe.

Contrary to static methods, *dynamic approaches* quantify motivation based on what the individual's subjective experiences can unveil about the reasons behind the performed behaviour. Factors such as contexts and rewards may impact the participant's attitude toward an activity (McClelland, 1987). Examples of dynamic approaches are Fogg's (2012) motivational waves and Self-Determination Theory (Ryan and Deci, 2017).

Benefiting of a static method in understanding the nature of actions and reinforcing that with a dynamic one to include subjective experiences would improve the analysis of motives. In this work, we draw on two dominant psychological explanations: Maslow's Hierarchy of Needs (Maslow, 1943) and Self-Determination Theory (Ryan and Deci, 2017). In particular, this work focuses on determining an individual's intrinsic motivation (Ryan and Deci, 2000) – activities that inherently bring satisfaction to an individual (commonly referred to as interests; Renninger and Hidi, 2016).

2.1.2.1 A primer on Maslow's Hierarchy of Needs

Maslow (1943) proposed a theory that explains motivation in terms of five levels of needs: *physiological needs*, *safety needs*, *belongingness needs*, *esteem needs* and *self-actualisation needs*. He postulated that higher needs could not be considered unless the lower ones are satisfied (Maslow, 1943; McClelland, 1987).

To explain how needs evolve; the example of primitive people is typically used (McClelland, 1987). At first, people in such an environment are concerned about their basic survival needs – such as food and sex – which corresponds to the lower level of the hierarchy (i.e. *physiological needs*). Once they are met, people start work on securing those needs which is the second phase of their development (i.e. *safety needs*). Upon a successful building of the secured environment, they reinforce their existence through love and social relationships (i.e. *belongingness needs*). As a result, the need for achievement and *self-esteem* arises which leads to the top of the pyramid and pushes them toward autonomy and *self-actualisation* (McClelland, 1987).

Some researchers criticized Maslow's view for several reasons (Neher, 1991). Some of them thought it applies only to the creative individuals; while others found Maslow's prerequisites for the high-level needs to be contrary to some of the evidence in the real world (McClelland, 1987; Neher, 1991). For instance, lack of security in some communities – due to war or similar reasons – does not prevent their inhabitants from developing social ties and pursue the fulfilment of belongingness needs. Nevertheless, Maslow's hierarchy of needs is considered as one of the most critical theories in the field which is widely adopted (McClelland, 1987).

2.1.2.2 A primer on Self-Determination Theory (SDT)

SDT identifies *competence*, *autonomy* and *relatedness* as the three basic psychological needs that differentiate and represent motivation states (Ryan and Deci, 2017). *Competence* satisfies the need to be able to perform; *autonomy* relates to the extent to which a person controls a behaviour; *relatedness* is concerned with feelings of connection with others and is an essential driver for social behaviour (Ryan and Deci, 2000, 2017).

SDT (Ryan and Deci, 2017) is one of a number of contemporary theories that build on the distinction between intrinsic and extrinsic motivation. However, in contrast to others, SDT treats these concepts not as a dichotomy, but instead as a continuum that ranges from amotivation, through a set of extrinsic motivation states, to a fully internalised intrinsic motivation (Figure 2.1).

As suggested by SDT, intrinsic motivation may evolve from internalising extrinsically motivated actions. In contrast to the four-phase model of interest development (Section 2.1.1.2), SDT's internalisation includes five stages that do not require sequential progression through them (Ryan and Deci, 2017). "External regulation" is a subtype of extrinsic motivation that lacks autonomy, whereas "intrinsic regulation" represents autonomous behaviours that produce positive feelings of interest. Between the two, SDT states "introjection", "identification", and "integration" as three phases. "Introjection" refers to regulating behaviours due to obligations and a desire to avoid the negative feelings that may result from abandoning them. Introjected regulation of an action may lead to discovering positive aspects of it, leading to an eagerness

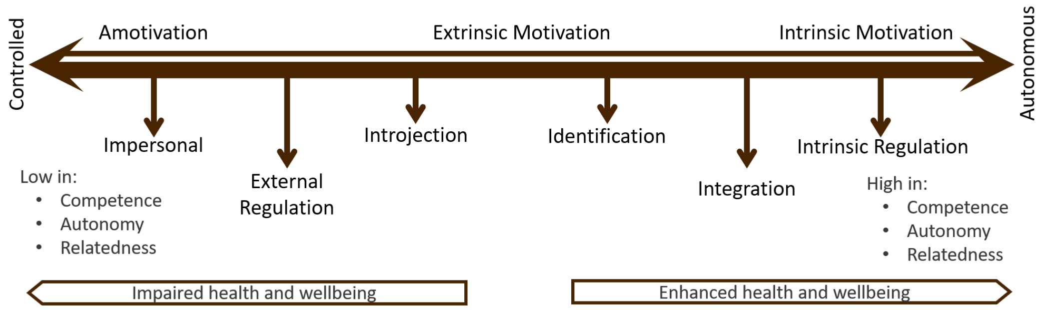


Figure 2.1: The motivation continuum proposed in SDT that shows the phases of internalising extrinsically motivated actions: External regulation, introjection, identification, integration and intrinsic regulation.

to maintain that action (identified regulation). As positive feelings grow, behaviours become more integrated and self-motivated (integration). Ultimately, the person does behaviour merely out of interest and to get the internal satisfaction it produces (Ryan and Deci, 2017). As an example of this internalisation process, consider a boy whose parents are controlling and sacredly value religion. The boy may practise religion and attend related events compliantly to avoid his parents' potential upsets (e.g. Brambilla et al., 2015). As the boy practises and learns more about the religion, he may start to accept it and internalise its practices. As a result, he would become more enthusiastic about it, assimilate the religious values and transfer them to his daily routine (Ryan and Deci, 2017).

2.1.3 Individual interest and intrinsic motivation

The broader shift in psychology, caused by the decline of behaviourism, contributes to the rising of studies on interest and motivation (Silvia, 2007). Although connecting these concepts becomes more difficult due to their acquisitions of many meanings over the years, the relation between individual interest defined as "enduring predisposition to reengage particular contents over time" (Hidi and Renninger, 2006) and intrinsic motivation seems intuitive and solid (Hidi, 2000; Bye et al., 2007; Silvia, 2007). This relation is evident in Ryan and Deci (2000)'s definition of intrinsically motivated behaviours as "those that are freely engaged out of interest without the necessity of separable consequences" (Ryan and Deci, 2000). This articulation of intrinsic motivation as *self-rewarding activities that are driven by individual interests* is also supported by the existing literature (Silvia, 2007).

Between interest and human motivation, this thesis focuses on *individual interest* and *intrinsic motivation*. Individual interest is what we aim to detect, and intrinsic motivation characterises actions that embody those individual interests. We adopt Ryan and Deci (2000)'s definition of intrinsic motivation and Hidi and Renninger (2006)'s of individual interests (quoted above) as they are widely dominants in their corresponding areas (Silvia, 2007; Alharthi et al., 2017).

2.2 Related work: Deriving behaviours from smartphones

Behaviours are the analysis units that are used to detect interests. Therefore, before studying the existing work on recognising interest from smartphones, we review studies that detect behavioural events from location, apps and notifications – the sources of the three behaviours targeted by this thesis.

Smartphones have transformed personal data collection. The majority of the population near-continuously carry a smartphone device featuring multiple specialised sensors such as: accelerometer, gyroscope, ambient light sensor, proximity sensing (e.g. Bluetooth, NFC); these are in addition to the microphone and camera that are considered critical to the devices' functionality (Lane et al., 2010). These sensors allow for passive data collection (i.e. without intervention from a user) that can be considered highly indicative of the user's environment and behavior. This data has a multitude of applications (Khan et al., 2013), including extensive use for health and well-being (Cornet and Holden, 2018). It is in this context that Jain et al. (2015) coined the term *digital phenotyping* to refer to the process of using an individual's interaction with digital technologies to derive indicative behavioural markers.

This thesis uses digital phenotyping to reveal behavioural aspects of mobility, phone usage and buying activities. Therefore, in the next sections, we review studies related to these three behaviours.

2.2.1 Mobility behaviour

Deriving mobility behaviour from smartphone location data is a multitask enrichment process (Baglioni et al., 2008). Semantic enrichment of Geographical Positioning System (GPS) data is the process of transforming streams of geographic coordinates into human behaviours (Krueger et al., 2015). The transformation process involves segmentation and annotation processes whose implementations are domain-specific (Montoliu and Gatica-Perez, 2010). However, before we identify these processes, it is essential to formally clarify some of the concepts that are typically used by the related literature. ***A trajectory*** is a sequence of geographic coordinates (i.e. longitude and latitude) over a period of time (Krueger et al., 2015). A trajectory could be divided into subtrajectories, and the ones that meet a predefined set of features are called ***episodes*** (Parent et al., 2013). For instance, a daily trajectory could be segmented into various episodes based on time spent at specific locations. ***A stay-point*** is a specific type of episodes that is characterised by predefined time and distance thresholds. So, if 10 minutes is the time limit and 100 meters is the distance threshold, then a stay-point is a place where the person spent at least 10 minutes without travelling more than 100 meters (Solomon et al., 2018).

Segmentation is the process of dividing trajectories into episodes based on the requirements of a target application (Parent et al., 2013). Stay-points are one type of segmentation output. Move-episode is another possible output which could represent a transition between two different stay-points or a movement that starts from and end at the same stay-point such as walking a dog (Parent et al., 2013; Krueger et al., 2015). Our scope focuses on understanding intrinsically motivated behaviour mainly from stay-points.

Annotation is the process of consulting external data sources to associate episodes with semantics (e.g. places details) (Parent et al., 2013). Most applications employ an external knowledge source to add context-specific data to raw coordinates (Nogueira et al., 2018). Foursquare and Google places are examples of external sources which provide reverse geocoding services to facilitate the annotation of stay-points (Krueger et al., 2015).

As stated in Chapter 1 (see Section 1.3), we started our work by considering the mobility behaviour as a case study to understand how behaviours are derived from the smartphone location data. Therefore, a deeper analysis of the related studies that covers the existing work on both segmentation and annotation is included as part of the next chapter (Chapter 3). The analysis is based on a systematic review that also covers the existing work on deriving cognitive and behavioural phenomena from the GPS data.

2.2.2 Phone usage behaviour

Phone usage behaviour refers to the active interactions between users and their smartphones' screens. Interactions include gestures, such as long clicks and scrolling the screen up and down, which always occur within the context of an app (Mehrotra et al., 2017b; Shah et al., 2020). To understand how behavioural events and features of phone usage are utilised, we have extensively reviewed the literature on phone usage behaviour within the context of digital phenotyping and passive sensing. Accordingly, we found 39 papers that analyse phone usage behaviour based on passive sensing. These studies can be classified as either health-related or behavioural. Health-related studies use phone usage behaviour to understand features associated with symptoms of specific health conditions such as schizophrenia (Wang et al., 2016), anxiety (Moshe et al., 2021), and depression (Chikersal et al., 2021), whereas behavioural research targets non-pathological conditions such as alertness (Abdullah et al., 2016), personality (Gao et al., 2020), and student behaviours in class (Shah et al., 2020).

Of the 39 papers, 19 are health-related. More than 47% of the 19 (9 papers) target different aspects of depression, such as its states and dynamics (Mehrotra et al., 2016a; Wang et al., 2018a). The remaining 10 studies address symptoms of schizophrenia (e.g. Wang et al., 2016), stress and mental health (e.g. Sano et al., 2018), social isolation (e.g. Doryab et al., 2019), social anxiety (e.g. Moshe et al., 2021), and impulsive behaviour (e.g. Wen et al., 2021). On the other hand, behavioural studies (20 papers) centred around two themes. The first focuses on understanding the personal aspects of individuals' behaviours. Examples of that include: alertness (Abdullah et al., 2016), understanding affect states (Cai et al., 2018), and eating episodes (Meegahapola et al., 2020a) of individuals. The other theme's topic is focused around the features of a specific group or community—for instance, detecting discrimination in college students (Sefidgar et al., 2019) and quantifying smartphone usage in adolescents (Domoff et al., 2021).

The segmentation of the captured interactions into events varies with the objective of the study. Part of the reviewed studies segment an event (i.e. determine the unit of analysis) based on the unlocking and locking of the phone's screen (e.g. Moukaddam et al., 2019; Domoff et al., 2021; Di Matteo et al., 2021) . In this case, an event may represent interactions with multiple apps as long as switching between apps happen before locking/unlocking the screen. Some studies divide interactions based on the app in which the interaction takes place such that an event is formed every time the app is changed (e.g. Tseng et al., 2016; Sano et al., 2018; Sükei et al., 2021). A third category divides the interactions series into morning, noon, and evening events based on the time of those interactions (e.g. Murnane et al., 2016; Mirjafari et al., 2019; Moshe et al., 2021). The number of clicks and screen swipes in the morning, for instance, can then be compared to those that take place in the evening (Murnane et al., 2016). In this thesis, we rely on the app as the basis for segmenting users' interactions with their phones into events. An event of phone usage behaviour is formed of all continuous interactions that are taken place on the same app in a single session (Shah et al., 2020). Changing from one app to another, locking the screen, and turning off the device would separate sessions and produce multiple events. Collectively, all events of using various smartphone apps form the phone usage behaviour (Tseng et al., 2016; Shah et al., 2020).

The studies that segment events based on the used app differ in how they add semantic meanings to these events. In some of these studies (e.g. Tseng et al., 2016; Sultana et al., 2020; Meegahapola et al., 2020a), statistics of the used apps is what matters regardless of their names or types. An example of that is knowing the number of apps (despite their names or types) used during eating episodes (Meegahapola et al., 2020a). The contrary to that is when the study's objective is related to the type of either the used app or the target behaviour (Mehrotra et al., 2017a; Rhim et al., 2020; Sarsenbayeva et al., 2020). For example, in Mehrotra et al. (2017a)'s study of mobile phone interactions, events of phone usage behaviour are labelled based on the type of the app rather than its name. The analysis is then conducted among multiple app groups.

Based on the reviewed studies, adding semantic types is done in three ways: (i) based on previous knowledge of the related apps (ii) consulting an external source (iii) or a combination of the two. In the first, researchers rely on prior knowledge to identify a set of apps (e.g. Skype, Snapchat, Chrome) as related to specific types (e.g. social, communication), and accordingly label the apps (e.g. Doryab et al., 2019; Sarsenbayeva et al., 2020; Wang et al., 2020). The second use an external source such as GooglePlay and AppStore to label the apps based on the retrieved category (e.g. Rhim et al., 2020; Gao et al., 2020). When combining the two, researchers first consult the external source and then combining the types into more consolidated categories (e.g. Mehrotra et al., 2017a). In this work, we rely on GooglePlay to categorise events based on their type so that we can link them to their possible motives (as we will see in the next chapters).

In analysing their topics, the reviewed studies relied on features of the events that form smartphone usage behaviour. These features heavily depend on the basis used in segmenting the interaction series and creating these events. When locking/unlocking is the selected base, then examples of the derived features are the frequency and duration of screen time (Mirjafari et al., 2019; Obuchi et al., 2020). Events that are formed based on the running app may produce a different set of features, such as the frequency and duration of using a specific app or category (e.g. Cai et al., 2018; Sarsenbayeva et al., 2020). However, properties of the interactions themselves might be used as the analysis features. For instance, the frequency of checking (i.e. touching) the screen is used as an indicator of impulsive behaviour (Wen et al., 2021). Also, the number of locks/unlocks per minute is studied as a possible indicator of loneliness and social isolation (Doryab et al., 2019). In this thesis, we extract features of the events that are based on app usage, such as the persistence of using a specific app.

Lastly, the studies included in the review employed the collected ground truths as either a formative or summative tool. Researches that utilise ground truths for formative purposes study the relationship between the responses collected from self-reporting inventories and the behavioural features of the smartphone usage events. For instance, participants were asked to complete the Patient Health Questionnaire (PHQ-9), a well-known tool for measuring depression (Kroenke et al., 2001; Kroenke and Spitzer, 2002), at the beginning of a study in order to correlate depression states with phone usage behaviour (Mehrotra et al., 2016a; Wang et al., 2018a). Another example is using participants' responses to the Generalized Anxiety Disorder (GAD-7) (Spitzer et al., 2006) or the Social Interaction Anxiety Scale (SIAS) (Mattick and Clarke, 1998) questionnaires to understand symptoms related to anxiety from smartphone usage data (Jacobson et al., 2020; Di Matteo et al., 2021). On the other hand, a summative use of the ground truth is manifested in studies that model and predict specific symptoms or behavioural features (similar to what we do in this thesis). These studies use the collected responses to assess the validity of the proposed model in predicting health-related symptoms or behavioural features. Examples include using questionnaires to assess the prediction accuracy of alertness (Abdullah et al., 2016), a schizophrenia relapse (Barnett et al., 2018), and depression states (Chikersal et al., 2021).

Table 2.1 summarises the reviewed studies. The table shows the topic of the included paper, the type of those studies, the features that are used for predicting aspects related to the investigated topic and how each study uses ground truth.

Table 2.1: Summary of the reviewed study related to phone usage behaviour.

Paper	Topic	Duration	Sample	Duration (app)	Frequency (app)	Diversity (apps)	Installed apps count	Duration (category)	Frequency (category)	Duration (screen)	frequency (screen)	Frequency (actions)	Time (actions)
Abdullah et al. (2016)	Alertness	40 days	20								✓		
Mehrotra et al. (2016a)	Depression	30 days	25	✓	✓		✓					✓	
Murnane et al. (2016)	Alertness	40 days	20	✓	✓								
Tseng et al. (2016)	Mental health	4 months	22	✓	✓								
Wang et al. (2016)	Schizophrenia	2–8.5 months	21							✓	✓		
Mehrotra et al. (2017a)	Usage/interaction Patterns	2 Weeks	26					✓					✓
Wang et al. (2017a)	Schizophrenia	2–12 months	36							✓	✓		
Wang et al. (2017b)	User recognition	2–10 weeks	20									✓	
Cai et al. (2018)	State affect recognition	2 weeks	220		✓								
Sano et al. (2018)	Stress and mental health	1 month	201	✓									✓
Wang et al. (2018a)	Depression	18 weeks	83							✓	✓		
Wang et al. (2018b)	Personality	14 days	646							✓	✓		
Das Swain et al. (2019)	Job performance	6 months	757							✓	✓		

Continued on next page

Table 2.1 – Continued from previous page

Paper	Topic	Duration	Sample	Duration (app)	Frequency (app)	Diversity (apps)	Installed apps count	Duration (category)	Frequency (category)	Duration (screen)	frequency (screen)	Frequency (actions)	Time (actions)
DaSilva et al. (2019)	Stress	2 semesters	94							✓	✓		
Doryab et al. (2019)	Loneliness and social isolation	1 semester	160							✓	✓	✓	✓
Mirjafari et al. (2019)	Job performance	2–8.5 months	554							✓	✓		
Moukaddam et al. (2019)	Depression and anxiety	8 weeks	22							✓	✓		
Rooksby et al. (2019)	Acceptability for mental health	7 days	15	✓									
Sefidgar et al. (2019)	Discrimination	2 semesters	209								✓	✓	✓
Xu et al. (2019)	Depression	219	455							✓	✓		
Gao et al. (2020)	Personality	1 week	183			✓	✓	✓	✓				
Levine et al. (2020)	Anxiety	1 month	10							✓			
Meegahapola et al. (2020a)	Usage during eating episodes	NR	206	✓									
Meegahapola et al. (2020b)	Food diary	NR	160								✓		
Obuchi et al. (2020)	Brain functional connectivity	79 days	105							✓	✓		
Rhim et al. (2020)	Subjective well-being	4 months	75	✓				✓		✓			
Sarsenbayeva et al. (2020)	Emotion	2 weeks	30	✓	✓			✓	✓				

Continued on next page

Table 2.1 – Continued from previous page

Paper	Topic	Duration	Sample	Duration (app)	Frequency (app)	Diversity (apps)	Installed apps count	Duration (category)	Frequency (category)	Duration (screen)	frequency (screen)	Frequency (actions)	Time (actions)
Shah et al. (2020)	Student behavioral patterns	45 days	47	✓	✓								
Wang et al. (2020)	Schizophrenia	12 months	75					✓			✓		
Chikersal et al. (2021)	Depression	16 weeks	138							✓	✓		✓
Domoff et al. (2021)	Phone usage of adolescents	7 days	46	✓	✓					✓			
Mack et al. (2021)	COVID-19	20 weeks	217							✓			
Melcher et al. (2021)	COVID-19	28 days	100							✓			
Moshe et al. (2021)	Depression and anxiety	2 weeks	60							✓	✓		
Nickels et al. (2021)	Depression	12 weeks	415							✓			
Sükei et al. (2021)	Emotional states	30 days	943	✓									
Wang et al. (2021)	Depression	30 days	120						✓				
Wen et al. (2021)	Impulsive behaviour	21 days	26									✓	
Xu et al. (2021)	Depression	16 weeks	395							✓	✓		

2.2.3 Buying behaviour

Buying behaviour can be deduced from smartphones through (1) the analysis of user's interactions with a shopping channel or (2) shopping-related messages received on smartphones. A shopping channel is what the person uses to complete a purchase such as a website or an app (Zhao et al., 2019). Data that result from interacting with a shopping channel can reveal information about the viewed and purchased products and hence help understand buying behaviour (Lemon and Verhoef, 2016). The other way of deducing buying behaviour is shopping-related messages. Smartphones' messages that are related to buying behaviour may include receipts or promotions (Grbovic et al., 2015). These messages can be delivered as SMS text, emails or messages within a shopping app (Kooti et al., 2016; Drossos et al., 2013). Regardless of the delivery means; a smartphone user typically would receive a notification each time a receipt is delivered (Li et al., 2018b). Therefore, in addition to studies around possible delivery means of these messages, we review studies on smartphone's notifications and their usages.

2.2.3.1 Mobile shopping

Mobile shopping is a recent form of transformations that occur in commerce and are caused by advances in technology and digital devices (Tang, 2019; Tyrvainen and Karjaluoto, 2019). Buying is only a motive among others that drive the use of shopping apps (Huang and Zhou, 2018; Tang, 2019). Examples of other motives may include prices comparison, products sharing, and reviews probing (Huang and Zhou, 2018; Chopdar et al., 2018). However, studies suggest that buying through shopping apps may still encounter some obstacles that fuel consumers' reluctance to embrace this method (Tyrvainen and Karjaluoto, 2019; Sohn and Groß, 2020). Usability issues and privacy concerns are common examples of these obstacles in the literature of smartphone shopping (Chopdar et al., 2018; Tyrvainen and Karjaluoto, 2019). Nonetheless, studies show that people's adoption of mobile buying is growing as retailers invest in addressing these obstacles to benefit from the ubiquity of smartphones among people (Tang, 2019).

Studies on mobile shopping widely address the topic of shopping intentions (Tyrvainen and Karjaluoto, 2019). They focus on identifying the motives behind using shopping apps, including the motive for buying (Sohn and Groß, 2020). Instead of understanding what does a user buy, participants of these studies are asked about the reasons behind (1) using the shopping apps or (2) their in-app purchase transactions (Tang, 2019; Sohn and Groß, 2020). Studies show positive correlations between the buying intention and the person's desire to install and use a shopping app (Kim et al., 2017; Tyrvainen and Karjaluoto, 2019). Also, predicting the persistence of the buying intention has been shown as an essential factor that may help retailers better guess future buying activities (Prashar et al., 2018). With respect to understanding the reasons of in-app purchase transactions; studies show that factors such as brand name and ease of use have strong correlations with buying through smartphones apps (Tang, 2019). For instance, brand names such as Amazon and eBay make users less reluctant to buy through their apps compared to local retailers with no popularity (Smith and Chen, 2018). Besides the lack of fame, perceived risk and privacy concerns are other bars that may hinder the execution of in-app buying activities (Sohn and Groß, 2020; Tang, 2019). However, studies suggest that users may still buy items over smartphones by directly using the web channels rather than the apps of retailers (Groß, 2020).

The relative preference for a buying channel within a smartphone (e.g. Amazon app) over another (e.g. Amazon website) is a subject of studies related to mobile shopping (Groß, 2020; Lemon and Verhoef, 2016). These studies look at what activities are preferred for each smartphone channel. For instance, buying activities have been seen more through websites compared to apps, whereas users rely more on apps for searching and browsing products (Tyrvainen and Karjaluoto, 2019; Lemon and Verhoef, 2016). Studies relate these two trends to hedonic and utilitarian motives (Tang, 2019). Hedonic motives refer to the positive emotions that drive people's desire to use the phone. When users are not motivated by an actual desire to buy, they may use their smartphone apps to browse and check items if doing so would cause the arousal of positive feelings. In contrast, websites are used more like a utility that a person utilises to actualise a buying intention (Lemon and Verhoef, 2016). Reasons for not preferring apps as a buying tool intersect with the ones behind the reluctance of conducting in-app buying, such as ease of use and privacy (Tyrvainen and Karjaluoto, 2019; Lemon and Verhoef, 2016). It is noteworthy that studies typically set off from theoretical and cognitive bases in their analysis of utilitarian and hedonic motives. Examples include relying on self-determination theory to understand intention for hotel booking (Ozturk et al., 2016) and the theory of planned behaviour to study the acceptance of shopping apps (Yang, 2013).

The comparison with website shopping has been extended to include devices such as desktops and laptops. In this case, shopping through mobile apps is compared to using website shopping from these devices (Tyrvainen and Karjaluoto, 2019). One example is about studying the impact of interacting through screen touching on selecting the shopping channel (Brasel and Gips, 2014). Studies suggest a role for the smartphones' feature of screen touching in preferring smartphones over desktops and laptops (Brasel and Gips, 2014; Tyrvainen and Karjaluoto, 2019). Another example is related to the mobility feature of smartphones. Ease of access to smartphones, caused by their near-continuous carrying, could support their use for shopping. This mobility feature could prompt the use of shopping apps for non-purchasing reasons as users may exploit smartphones' near-continuity to view and search products instantly. Accordingly, attributing smartphone apps' non-purchasing usages to hedonic motives (that we stated earlier) may arise from the feature of smartphone's mobility. This feature allows prompt responses to these motives compared to other less mobile devices such as laptops.

According to our review of the mobile shopping literature, an extensive portion of the studies relied on surveys as the basis for data collection. This note of reliance on surveys is supported by two systematic reviews (Tang, 2019; Tyrvainen and Karjaluoto, 2019) and a literature review (Lemon and Verhoef, 2016) published between November 2016 and August 2019. According to the systematic review published in 2019 by Tyrvainen and Karjaluoto; 41 out of 56 studies (73%) of mobile shopping studies used surveys as the primary source of data. The remaining portion comes from data collected during an experimental study or using a secondary database. We found a smaller portion of studies that relies on data from interviews and secondary databases (e.g. Andrews et al., 2012; Cao, 2014).

Two papers, however, that we are aware of have relied on data collected from smartphone passive sensing (Bang et al., 2013; Kim et al., 2017). The first one (Bang et al., 2013) used the collected data to understand the products' features that influence the selection of mobile over an online channel, whereas the other (Kim et al., 2017) studied the relation between buying behaviour and apps' possession. Although Kim et al. (2017) aimed to understand buying behaviour, they did so by correlating the users' self-reported inputs to the collected data (i.e. formative approach). This method is in contrast with our goal of being unobtrusive and relying on data from smartphones. Understanding what a user buys from the data and consequently analysing buying behaviour, are yet to be researched.

2.2.3.2 Smartphones' messages

The second aspect that might lead to understanding smartphone buying behaviours is the analysis of smartphones' messages that are related to shopping. When a purchase is made, the user usually receives a confirmation message (i.e. a receipt) that contain the transaction details (Kooti et al., 2016). Retailers may also send out promotional messages containing purchasing recommendations related to what has been previously purchased (Rossi et al., 1996; Zhang and Wedel, 2009). Both receipts and promotional messages can help in understanding the buying behaviour of a specific user. The former provides a direct way to infer the purchased items from the receipt content, while the latter can indirectly lead to the perception of previously purchased items (Grbovic et al., 2015).

The advent of emails and online shopping has contributed to the retailers' adoption of digital receipts in a greater capacity. In turn, digital receipts and retailers' ability to electronically communicate with users (e.g. through email) facilitate the personalisation and delivery of promotions (Grbovic et al., 2015).

Emails are the most common way of delivering digital receipts and promotions (Kooti et al., 2016). Although most of the users' emails are machine-generated and sent from a business to human (Sun et al., 2018; Whittaker et al., 2019; Gupta et al., 2019); understanding buying behaviour from emails is limited to companies that host email services and have access to the private contents of these messages (Botta, 2016). For instance, Yahoo users' emails have been used to identify templates of digital receipts and extract purchase data (Grbovic et al., 2015). Next, the extracted data was used to predict the behavioural patterns and demographic data of the users (Kooti et al., 2016; Grbovic et al., 2015). The templates and contents of promotions have also been studied from emails of Gmail users. These studies predict the category of the promoted contents (Sun et al., 2018), classify the promotions templates (Whittaker et al., 2019; Kocayusufoglu et al., 2019) and use promotions to create recommendations (Potti et al., 2018). However, previous studies conducted the analysis without relying on smartphones as a data channel. We could not locate a study that analyse emails collected from personal smartphones to understand what an individual buys.

In contrast, Short Message Services (SMS) is a smartphone-specific data channel that enables the exchange of short messages. The ability to communicate through SMS without the need for Internet service has contributed to the use of these messages in contextual applications; such as offering context-aware promotions via text messaging (Tyrvainen and Karjaluoto, 2019; Grbovic et al., 2015); which is another topic that is relevant to our research. In this case, the context controls where and when to deliver the SMS containing the promotion to be released. For example, if a user is near a specific store, a text message promoting that store's products is sent (Duzgun and Yamamoto, 2017; Shareef et al., 2015; Tyrvainen and Karjaluoto, 2019).

As for understanding buying behaviour by analysing the content of receipts and promotions sent as text messages, we could not find research papers on this topic. We also have not been able to locate any study aiming to classify receipts and promotions from all SMS messages received by a user. However, it should be noted that there are studies that have focused on analysing the content of text messages in other contexts, such as sentiment analysis for teaching evaluation (Leong et al., 2012) and recognising friends and family (Min et al., 2013).

So far, we have reviewed emails and SMS as possible smartphone data sources that can lead to understanding buying behaviour. However, regardless of how receipts and promotions land (i.e. as emails or SMS); typically, a notification is issued whenever a message is received on the user's phone (Pielot et al., 2014). Therefore, notifications can provide a more holistic source of short messages that might help understand buying behaviour.

The content of smartphones' notifications varies according to the apps issuing these notifications. News apps, for instance, may provide breaking news about the stock market, while videos recommendations may form the content for notifications of entertainment apps (Pielot et al., 2014). Interactions of users with these notifications, on the other hand, are affected by several factors. Examples of these factors include the contexts in which notifications arrived (Visuri et al., 2019); the importance of the topic to the recipient (Sahami Shirazi et al., 2014); and the content clarity (whether or not notifications are touched for viewing incomplete or unclear ones) (Mehrotra et al., 2016b). These factors impact the users interactions with the received notifications.

A large body of studies on notifications analysis seeks to understand interruptibility from users' contexts and interactions with notifications (Mehrotra et al., 2016b). Analysis outcomes can be used to tailor notifications delivery based on each user's situation (Mehrotra et al., 2015; Sahami Shirazi et al., 2014). When a user is at a movie theatre or business meeting, notifications can be deferred in order not to interrupt or disturb the user (Mehrotra et al., 2017a). Changing the smartphone's tone to a less disturbing mode (such as the silent one), rather than delaying notifications, is an alternative way of tailoring the notification delivery according to the user's situation (Visuri et al., 2019).

Besides the spatial context, the user's situation may be based on other contexts. For instance, temporal context can be understood by studying a relationship between users' past interactions with notifications and the times of these interactions (Mehrotra et al., 2016b). Understanding this relationship helps prevent notifications delivery at times when a response is not expected by the user, such as bedtime (Fischer et al., 2010). These studies regardless of the contexts and interactions used to analyse notifications, consider interruptibility as either a binary state or a multifaceted case. The former classifies situations as either interruptible or uninterruptible (Poppinga et al., 2014). In contrast, the latter includes instances where a user might accept or even prefer to be partially interrupted if the notifications are related to a specific topic of interest (Turner et al., 2017).

Some notifications analysis studies go beyond merely understanding interruptibility to study factors controlling the user's response to notifications (receptivity) (Mehrotra et al., 2016b; Westermann et al., 2016; Schulze and Groh, 2014). Users' responses are interactions such as viewing, touching and dismissing notifications. A third category of studies proposes applications and frameworks that can help in collecting notification in-the-wild (Weber et al., 2019). However, previous studies are not oriented toward understanding aspects of an individual's daily behaviour from their content. The closest work that aimed to do so is the one introduced by Li et al. (2018b). In that work, notifications are classified into templates, and then knowledge entities are recognised as parameters of these templates. However, to conduct the classification task as suggested by Li et al., 2018b, the existence of a large corpus of smartphone notifications generated by a large number of apps and people is required. Unlike Li et al. (2018b)'s work, we rely only on the notifications generated from the user's device to understand behaviour aspects while preserving privacy. Moreover, this is the first effort, to our knowledge, that aims to understand what individuals buy from the notifications received on their devices.

As we notice from the previous discussion, studies on smartphone buying behaviour do not explore the content of the receipts or the features of the bought items.

2.3 Related work: Recognising interest using smartphones

In this part, we review studies that use the smartphone to detect human interest. We do not include research that infers interests from other sensor-based devices nor recommender systems that rely on datasets collected from devices other than smartphones. Also, various methods of employing the detected interests to personalise recommendations, such as targeted ads, are not reviewed in this section, although those recommendations are typically subsumed by recommender systems (Ricci et al., 2011; Lops et al., 2011; Zhang et al., 2019). Those types of research are outside the scope of this work and we refer the reader to following and recent reviews (Khan et al., 2017; Ding et al., 2018; Zhang et al., 2019; Babiker et al., 2021) that address them. The first three reviews (Khan et al., 2017; Ding et al., 2018; Zhang et al., 2019) address different aspects related to recommender systems. Specifically, the work of Khan et al. (2017), presented a systematic review of the literature on cross domain recommenders. The second paper (Ding et al., 2018) conducted a comprehensive review on recommending items based on data from Location Based Social Networks (LSBN). As recommender systems moved toward more adoptations of deep learning methods, Zhang et al. (2019) reviewed the existing recommenders that use these methods. Unlike the previous reviews on recommender systems, the last review (Babiker et al., 2021) focused on studies that detect human interests using sensors attached to individuals such EEG and ECG. The review looks at how psychological and physiological aspects of interests has been addressed by studies authored between 2009 and 2019 (Babiker et al., 2021). Nonetheless, we have researched all the reviews for studies that detect interests from smartphone data. Accordingly, we found four studies that meet that goal.

Besides the four studies included from these reviews, we have conducted an extensive review of the literature. As discussed in the first part of this chapter, interest is a common word that can have many meanings in our everyday language, which sometimes makes its use ambiguous (Silvia, 2007; Ahmed and Srivastava, 2019). Therefore, we limit our search to studies that aim to detect situational or individual interests (see Section 2.1.1). Also, our review included studies that use smartphone's data as the primary and only source for detecting interests. As we detailed earlier in Section 2.2, smartphones, as data sources, have unique characteristics due to their ubiquitous functionalities and near-continuously carrying property which make them different from other discrete sources (e.g. tweets, and check-ins data) (Lane et al., 2010; Cornet and

Holden, 2018). So, deriving behavioural events and their features from smartphones (i.e. digital phenotyping) will be determined specifically by the characteristics of those devices.

Based on the above, we searched for studies that detect: “human interest”, “individual interest”, “personal interest”, “user interest” or “situational interest” from “smartphone”, “mobile phone”, or “cell phone” from ACM, IEEE, Springer and Scopus databses. We combined the studies from our extensive review of the literature with other studies that we identified as relevant from the aforementioned reviews. Collectively, we have identified 17 studies as relevant to our work. We only included studies that focus primarily on using smartphone’s data to explore aspects related to human interests (e.g. detecting them or analysing their dynamics). We excluded studies that are not interest-specific or use interests data from sources other than smartphones, such as user profiles in social media.

The included studies are published between 2013 and 2021. The majority of those studies (11 papers) relied only on passively sensed smartphone data to detect aspects related to human interests. Two of the remaining six detected interests by solely relying on smartphone-based questionnaires, and the remaining four combined both questionnaire and passive smartphone data in their analysis of interest.

Also, six of the 17 studies focus only on detecting interests and/or aspects related to them (e.g. Zhao et al., 2013; Akkerman et al., 2020). The six studies attempt to elicit individuals’ interests using experimentation and through people’s interactions with technology. These interactions, in one example, are used along with textual data (e.g. tweets) to detect interests (Tu et al., 2020). Another form of inferring interests can be achieved through the analysis of the visited sites that are captured passively using sensors such as GPS (Shi et al., 2019).

The remaining 11 studies used smartphone-detected interests for personalisation purposes. Researchers, in this second type, go beyond the mere detection of interests to using and employing them in various applications. An example of this is customising mobile advertising based on smartphone-detected interests (Lee and So, 2014). Another example is the use of intereset in news personalisation (De Pessemier et al., 2016). Both examples detect interests first from the data and then use them to customise outputs.

Although the included studies suggest methods of detecting interests, an in-depth analysis of these methods is required to assess their validity. Given the subject matter of this thesis, we will discuss, next, these studies from four facets: unobtrusiveness, personalisation, indicator selection, and interest dynamics. These facets characterise both digital phenotyping (Onnela and Rauch, 2016; Vega-Hernandez, 2019) and human interests (Renninger and Hidi, 2011; Zhao et al., 2013) and hence can lead to a detection method that we argue is suitable for real-world implementations.

2.3.1 Unobtrusiveness

A measure is said to be unobtrusive if it does not involve a direct extraction of knowledge from the user (Webb et al., 2000). Based on this definition, 11 of the 17 studies adopted an unobtrusive measure of human interest. Four of the 11 studies relied on the app usage frequency as the interest determinant (Lee and So, 2014; De Pessemier et al., 2016; Huang et al., 2019; Tu et al., 2020). One study determined interests based on the app usage duration (Jia et al., 2015). Two studies combined both frequency and duration of app usage as determinants of interest (Zhao et al., 2013; Tu et al., 2021). Of the remaining four studies, one relied on the frequency of visited places as its interest indicator (Xia et al., 2014). Another study combined the frequency of visited places with app usage and call frequencies (Shi et al., 2019). Khusumanegara et al. (2015) determined interests based on the entries of browsing history, and the last paper (Gulla. et al., 2014) used the frequency and duration of micro-interactions collected from their news app (e.g. clicked categories, article view time and starred articles) to determine interests. Only three of the 11 papers included ground truths for evaluating their interest detection methods. The other eight seem to assume that the detected interests are correct without the need to confirm the results with their users (Table 2.2).

On the other hand, two studies (Akkerman et al., 2020; Draijer et al., 2020) relied solely on self-reporting to detect interest (i.e. an obtrusive measure). Specifically, the authors of the two papers used smartphone's Experience Sampling Method (ESM) to identify interests. ESM is a self-reporting method that benefits from smartphone's mobility and availability to deliver questionnaires under specific contexts. The used questionnaire defines time, value, agency, frequency, intensity, and mystery as interest determinants. Both studies targeted students in their interest analysis.

Combining unobtrusive and obtrusive measures occur in four studies that combined interaction data and self-reporting methods. The authors of those studies relied on self-reported interests as formative tools to detect interests from the interaction data. Two papers relied on focus group observations and informal interviews and correlated that to the app usage frequency to determine interests (Rosales and Fernandez-Ardevol, 2016a,b). One of these two studies used the findings that resulted from combining the two measures to emphasise the importance of including apps that are of personal interest for older adults (Rosales and Fernandez-Ardevol, 2016a). The other study detected the interesting apps for older adults and indicated the importance of mixing self-reporting and interaction data to confirm the validity of the detected interests (Rosales and Fernandez-Ardevol, 2016b). The other two studies that integrated unobtrusive and obtrusive measures used smartphone questionnaires for interest reporting (Frey et al., 2017; Aoude et al., 2018). The reported interests of one study (Frey et al., 2017) along with the installed apps of the users are then integrated to build a collaborative model that predicts interests of new users based on their installed apps. The other correlates mobility and connectivity patterns with the interests reported through questionnaire (Aoude et al., 2018) (Table 2.2).

2.3.2 Base for indicator selection

Our review shows that frequency and duration are the common indicators used for detecting interests from smartphones' data. Typically, frequency and duration are calculated based on an extracted feature such as the usage of apps (Huang et al., 2019) or visiting a specific places (Xia et al., 2014). This occurred in 13 out of the 17 studies that are included in this review. However, the base on which it has been assumed that frequency or duration are sufficient determinants of interests is undefined in these 13 studies.

As we discussed in Chapter 1, many of our daily activities that we do frequently and spend long periods doing them can be caused by reasons other than personal interests (Ryan and Deci, 2017). To address this issue, one of the included studies (Tu et al., 2020) ranked the app usage frequency for all users and then dropped out the top ten apps that are mostly used assuming they may not show personalised preferences. Although this approach may improve the detection of interests as suggested by the paper, it risks losing the crucial ones. A different study (Rosales and Fernandez-Ardevol, 2016a) confirmed the need to further investigate the causality between the usage patterns and personal interests as the conducted interviews proposed that the basic metrics may not be enough to deduce personal interests. Nonetheless, the sufficiency of the metrics implicitly assumed by the other included studies has not been supported by a base or profound knowledge from the existing literature on interests.

Also, because the discussed studies did not built their analysis on a profound knowledge of interests, they typically do not distinguish between individual and situational interests. As discussed earlier in this chapter (Section 2.1.1), the distinction is important as the former describes an emotion whereas personal interests are enduring dispositions or traits integrated within the self (Silvia, 2007; Ryan and Deci, 2017). Failure to distinguish between the two types directly impact the analysis. For instance, it is not clear whether the interests that are predicted from browsing history (Khusumanegara et al., 2015) represent situational or individual interests. Also, studies that used the detected interests to personalise advertisement (Lee and So, 2014; Huang et al., 2019) do not distinguish between the two types which may hinder the studies' ability to adapt to the short and long term interests.

A profound analysis of interests occurred only in the two studies that adopted ESM to report interests (Akkerman et al., 2020; Draijer et al., 2020). Despite the downside of self-reporting that we discussed in Chapter 1, the two papers proposed a deeper analysis based on six variables which include frequency and time. Although some of these variables can be computationally modelled in a relatively straightforward manner (e.g. frequency and time), others such as mystery and agency can be problematic. This is due to the cognitive nature of these constructs and the lack of a reliable method for quantifying them using smartphones. Nonetheless, studies that target a proper detection of interests should not ignore these constructs or others to improve the results.

2.3.3 Personalisation

The sensitivity of people's data and the existence of individual differences (Silvia, 2007) call for a personalised method of detecting interests. The methods should take into account the importance of preserving people's privacy as well as the ability to detect their interests without the need to share their data with others (i.e. collaborative analysis).

However, collaborative detection of interests can help in finding general patterns (traits) of interests . In this context, some studies relied on data collected from Internet Service Providers (ISPs). Although these studies do not collect ground truths, they can form a base of a general understanding of interest. For example, in one study (Tu et al., 2020), the smartphone data of 32,000 people collected over a six-day period have been used to understand interests. The data was provided by a telecommunications company in the country of the study. Another study provided an analysis of the interest dynamics through a one-month data of more than 19,000 people (Zhao et al., 2013). These studies relied on a collaborative analysis that combines the collected data to build a general model for predicting interests.

This collaborative analysis is not limited to data collected by ISP companies. For example, in an academic study, data collected from 904 participants were used to build a model that predicts interest through collaborative analysis of software installed on users' computers (Frey et al., 2017). Despite the benefit of understanding interests through a collaborative analysis of this data, this understanding does not necessarily reflect personal differences between individuals. In fact, a large body of the interest's literature shows that general patterns, such as personality traits, are not guaranteed to detect individual differences specific to each person (Silvia, 2007). Moreover, the collaborative analysis relies on data sharing, which contradicts the importance of preserving the privacy of the person's data.

In contrast, there are a number of studies that have taken into account data privacy by conducting experiments that aim to understand interests without the need for data sharing. For example, to provide personalised recommendations for new mobile applications, an individual-based analysis has been conducted on data from 40 people that were collected for eight months (Tu et al., 2021). Another study attempted to elicit interest by analysing the content of each person's interactions with the news app that was developed specifically for the experiment purposes (Gulla. et al., 2014). However, these studies share the issue of not providing a profound analysis of interest that covers aspect such as interest dynamics (Gulla. et al., 2014) or properly validates the accuracy of the detected interests through participants' ground truths (Tu et al., 2021).

2.3.4 Interest dynamics

Interest dynamics is another aspect that influences the detection of human interest. The detection of such dynamics is expected to improve through a longer observation (i.e. longitudinal studies). Since individual interests are relatively stable dispositions (Renninger and Hidi, 2011), a study needs to be *long enough* to be able to capture potential dynamics caused by acquiring new interests or abandoning existing ones (Zhao et al., 2013; Tu et al., 2021).

Only four of the included studies have considered the dynamics of the interest in their analysis. In Zhao et al. (2013)'s work, the dynamics are analysed based on one month of data collection. A longer period appeared in Tu et al. (2021)'s work in which the study lasted for eight months providing a better chance of discovering the dynamics. Both studies relied on smartphone's interaction data that were collected continuously during the study period which improve the chances of capturing existing dynamics.

The third study (Gulla. et al., 2014) designed a specific news app to detect interest and its dynamic from users' interactions. Specifically, the paper included additional functionalities to capture the micro-interactions data such as clicks and sharing a specific article on social media. Lastly and unlike the previous three, the fourth study depended on self-reporting in capturing interests dynamics (Akkerman et al., 2020). The study delivered a questionnaires to the participants' smartphones at two time-points that are 3 months apart. However, the lack of continuous observation of how things might have evolved or vanished during the three months between the time points weakens the reliability of this approach for dynamic detection.

Although individual interests are not expected to change rapidly (i.e. daily or weekly), they are expected to be temporally or permanently replaced in the long run. Interest dynamics can capture such changes with the appropriate observation time. Existing studies suggest an observation period that is longer than two months (Sarker et al., 2019). However, only one study has met that criteria (Tu et al., 2021). The importance of longer observation can also support the detection of developing situational interests into enduring and personal ones.

In Table 2.2 we summarise the reviewed studies and classify them based on the four facets discussed in this part of the Chapter. We also indicate the papers' adoption of a continuous collection of data and the existence of ground truth.

The works reviewed and discussed in this last section share the goal of this thesis which is detecting interests from smartphone's data. However, the deeper analysis of unobtrusiveness, indicator selection, personalisation and dynamics shows that the closest to ours are probably (Zhao et al., 2013 and Akkerman et al., 2020). Although Zhao et al. (2013)'s work does not adopt a specific psychological nor cognitive framework as the base for their work, they provided, to the best of our knowledge, the only study that digs into human interest as a behavioural phenomenon. However, the paper adopted a collaborative approach rather than a personalised method. Also, as shown earlier, one month is a relatively short period to discover changes in individual interests. Lastly, the paper did not support the selection of their indicators on a profound knowledge nor validate the detection with collected ground truths.

Unlike the Zhao et al. (2013)'s study, the work of Akkerman et al. (2020) was built on a profound knowledge to define interest determinants. The work also used properly collected ground truths for the analysis. However, smartphones have been used to digitise self-reporting which does not solve the downsides of self-reporting described in Chapter 1.

This work fills the gap by binding together the theoretical foundation of human interests with the need to detect them unobtrusively, and continuously. Doing so facilitates the discovery of interests and their dynamics. This is important not only for solving self-reporting issues but also for serving applications that target indirect behavioural changes (i.e. nudges).

Table 2.2: Summary of the reviewed study related to recognising interest from smartphone. *NR*: Not Reported.

Paper	Study focus	Duration	Sample	Unobtrusiveness	Continuousness	Dynamics	Ground truth	Personalisatin	Profound base
Zhao et al. (2013)	Detect personal interests and its dynamic from smartphone's data	1 month	19067	✓	✓	✓			
Gulla. et al. (2014)	Use interests detected from users' interactions with a news app to personalise news recommendations	NR	NR	✓	✓	✓		✓	
Lee and So (2014)	Detect personal interests and use them to personalise mobile advertising	3 days	15	✓	✓		✓	✓	
Xia et al. (2014)	Detect personal interests from the visited places data derived from GPS sensor	2 weeks	NR	✓	✓		✓	✓	
Jia et al. (2015)	Detect personal interests to personalise video recommendation	NR	NR	✓	✓		✓	✓	
Khusumanegara et al. (2015)	Predict interests from smartphone's browsing history	1 month	30	✓	✓			✓	

Continued on next page

Table 2.2 – Continued from previous page

Paper	Study focus	Duration	Sample	Unobtrusiveness	Continuousness	Dynamics	Ground truth	Personalisatin	Profound base
De Pessemier et al. (2016)	Compare interests detected from smartphone to static users' profiles for news recommendation	35 days	110	✓	✓				
Rosales and Fernandez-Ardevol (2016b)	Recognise personal interests to provide insights on improving the smartphone usability among older people	1 month	216		✓				
Rosales and Fernandez-Ardevol (2016a)	Detect interests of older adults by comparing results of tracking mobile app usage with focus group discussions	1 month of tracking and 3 sessions	25		✓		✓	✓	
Frey et al. (2017)	The interests are self-reported through an app and then the list of installed apps are used to predict them	NR	904				✓	✓	
Aoude et al. (2018)	Correlate mobility and connectivity patterns with content interests that are self-reported	11 weeks	38				✓	✓	

Continued on next page

Table 2.2 – Continued from previous page

Paper	Study focus	Duration	Sample	Unobtrusiveness	Continuousness	Dynamics	Ground truth	Personalisatin	Profound base
Huang et al. (2019)	Detect interest from smartphone's data and use that to select how to distribute ads across multiple regions	20 days	431,928	✓	✓				
Shi et al. (2019)	Use the interests detected from smartphone usage data to provide a personalised location recommendations	NR	32	✓	✓			✓	
Akkerman et al. (2020)	Measure interests of students and their development over time through smartphone's Experience Sampling Mehtod (ESM)	2 time points that are 3 months apart	204			✓	✓	✓	✓
Draijer et al. (2020)	Use smartphone's ESM to detect the multidimensional structure of students' interests	2 weeks	94				✓	✓	✓
Tu et al. (2020)	Detect interests from app usage data and users' tweets to make personalised app usage estimation	6 days	32,000	✓	✓				

Continued on next page

Table 2.2 – Continued from previous page

Paper	Study focus	Duration	Sample	Unobtrusiveness	Continuousness	Dynamics	Ground truth	Personalisatin	Profound base
Tu et al. (2021)	Study the changes in user interests and app functionalities and suggest recommendation based on the dynamics of both	8 months	40	✓	✓	✓		✓	

2.3.5 Conclusion

Our review indicated the need to examine interest more closely. Also, it showed the extensive work on extracting behavioural events from phone usage data. This observation shifted our focus toward researching the other two behaviours, especially as the occurrence of both mobility and buying behaviours is not directly related to the existence of a smartphone device, unlike the phone usage (in the former, the device is a means of capturing behaviour rather than an execution tool of it as in the latter).

Therefore, a deeper investigation that goes beyond a mere review of these two behaviours is essential to recognise interests, which we do in the next two chapters. First, regarding mobility behaviour, a systematic review of the literature on how the raw location data is enriched to understand features of cognitive or behavioural phenomena has been conducted (the next chapter). We synthesis the results into a framework that streamline the process of enriching the GPS data.

With regard to buying behaviour, and in view of the absence of any studies aimed at extracting it from the smartphone notifications, we do that in Chapter 4, taking advantage of the data that we collect in analysing the possible ways to do so. In this context, our study included methods based on prior knowledge of the installed apps and compared them with those based on machine learning techniques.

Once the methods of extracting behavioural events are set up properly, applying the indicators extracted from the interest and motivation literature are applied on those events to detect interests (Chapter 5 and Chapter 6).

Chapter 3

Smartphone-Derived Mobility Behaviour

In the previous chapter, we provided a general background that covered (1) theoretical underpinnings of interest and human motivation, (2) the extraction of events from smartphones' data and (3) the recognition of interests using smartphones. The remaining chapters of this thesis expand the last two aspects as they are crucial to the process of interests recognition. Events of the daily routine are the units that need to be analysed to recognise actions motivated by personal interests. Those events are encoded into the smartphone's raw data. Therefore, the first step is to extract those events such that they can be classified as motivated or not motivated by personal interest. In this chapter, we focus on events of mobility behaviour and how they can be extracted. We conduct a systematic literature review and synthesise the existing processes into a structural framework. The knowledge presented in this chapter helps understand existing methods of extracting mobility events and the requirements that need to be considered. Also, we show how errors can propagate throughout the framework and the impact of that on the overall extraction task.

The main content of this chapter is a paper authored by: *Ahmed Ibrahim, Heng Zhang, Sarah Clinch and Simon Harper*. The title of the paper is: ***From GPS to Semantic Data: How and Why—a Framework for Enriching Smartphone Trajectories***. The paper is published in Computing, Aug 2021. ISSN: 1436-5057. DOI: 10.1007/s00607-021-00993-z. URL: https://doi.org/10.1007/s00607-021-00993-z. For this thesis, we edited some formatting styles, such as the sizes of some tables for consistency and readability reasons.

Author contribution

Ahmed Ibrahim designed the study, carried out the data collection, analysed and synthesised the results and wrote the paper. Heng Zhang reviewed the papers separately to select the related ones. Sarah Clinch and Simon Harper provided continuous feedback throughout all the stages of the study, offered advice and discussion and contributed vital edits to the paper's writing.

Abstract

Deriving human behaviour from smartphone location data is a multitask enrichment process that can be of value in behavioural studies. Optimising the algorithmic details of the enrichment tasks has shaped the current advances in the literature. However, the lack of a processing framework built around those advances complicates the planning for implementing the enrichment. This work fulfils the need for a holistic and integrative view that comprehends smartphone-specific requirements and challenges to help researchers plan the implementation. We propose a structural framework from a systematic literature review conducted to pinpoint the main challenges and requirements of research on enriching location data. We classify findings based on the enrichment task and integrate them accordingly into workflows that facilitate the task's implementation. These workflows help researchers better streamline their implementations of the enrichment process and analyse errors within and across tasks. Moreover, researchers can integrate the presented findings with the proposed opportunities to better predict the impact of their research.

3.1 Introduction

Semantic enrichment of location data is the process of transforming raw data collected from mobility tracking devices into behaviours (Baglioni et al., 2008). These behaviours may express human activities, or they may be descriptions of non-human actions (such as animal behaviours or ship traffic and air navigation) (Fileto et al., 2015). The former can be derived from sources such as Geographical Positioning System (GPS) (Nogueira et al., 2018) and Call Detail Records (CDR) (Dashdorj et al., 2013) and has a multitude of applications (Khan et al., 2013), including extensive use for health and well-being (Cornet and Holden, 2018).

This paper focuses on semantic enrichment of GPS data collected using smartphones since the majority of the population near-continuously carries a smartphone featuring a GPS sensor (Lane et al., 2010). The enrichment process involves several sub-processes whose implementations are domain-specific (Montoliu and Gatica-Perez, 2010). For instance, segmentation is a sub-process that aims to divide GPS streams (a.k.a. trajectories) into episodes that serve specific application purposes. Some applications may split episodes based on their duration, while others may specify them based on the distance to previously determined points of interest. Consequently, different domains use different requirements to produce application-specific meanings of trajectories (Parent et al., 2013).

Smartphones trajectories reflect a naturalistic representation of human mobility and introduce unique semantic enrichment challenges. Smartphone-based GPS tracking is particularly problematic since individuals' mobility do not necessarily represent constrained roads and can have more variable trajectories (Yan et al., 2013). Additionally, data collection is negatively impacted by factors that are unique to smartphones. For instance, people can explicitly turn off sensors to prioritise the battery consumption (Rawassizadeh et al., 2016; Do and Gatica-Perez, 2014). However, it is not always necessary that the collected data is an actual representation of mobility behaviour. This is because people are expected to leave or forget their phones in different places such as home or car (Dey et al., 2011). Lastly, implicit factors, such as power management and software modules, degrade the resilience of the enrichment process and requires a more profound analysis of how each reason could hinder the semantic understanding of the raw data (Nogueira et al., 2017).

To enrich smartphones trajectories, we need to consider the above challenges in conjunction with requirements scattered across the literature of semantic enrichment.

In this paper, we approach this goal through a structural framework based on a meta-analysis of our systematic review of the literature. We go beyond the mere introducing and surveying of the general knowledge related to the semantic enrichment operation to synthesising the findings into a structural model. Our analysis of the literature is human-centric that provides the following contributions:

- We introduce a structural framework for enriching smartphone location data based on a systematic review of the literature. This framework presents a holistic and integrative view to help researchers plan the semantic enrichment requirements and address the smartphone-specific challenges. We synthesise findings scattered across the existing studies into workflows corresponding to the enrichment tasks. These workflows streamline the implementation of the enrichment process and facilitate the tracing of errors throughout the entire process.

- We provide a systematic literature review of enriching smartphone location data. To the best of our knowledge, this is the first review that targets smartphone trajectories and organises the findings according to the semantic enrichment task. The reported results introduce the researchers with a comprehensive analysis of the state-of-the-art of each task and help them identify the characteristics and limitations of the existing methods.

- We provide a planning strategy derived from the conducted review and the created model. We identify the Strengths, Weaknesses, Opportunities, and Threats of the existing studies according to the SWOT analysis framework. As a well-known planning framework, a SWOT analysis based on the review findings can help researchers better envisage the potentials of future contributions.

3.2 Background and related work

Traditionally, Geographical Information Systems (GIS) provide tools that analyse and understand spatial data (Burrough et al., 2015). These applications map longitude and latitude to place labels; and provide several functions that facilitate the users' interactions with maps, such as location query and map edit. GIS systems use different methods to capture and store the large amount of locations' meta-data they need to support their functionalities. Recently, due to the proliferation of mobile devices (e.g. smartphones), people are becoming primary data collectors for GIS data as they check in their visited locations (Burrough et al., 2015).

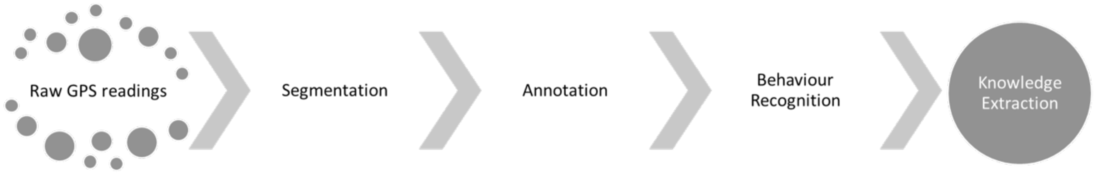


Figure 3.1: Semantic enrichment of smartphone trajectories.

Mobile devices, however, foster a new paradigm of spatial analysis centred around individuals' behaviour (Wilmer et al., 2017). In this paradigm, the spatial analysis of raw data is tightly coupled to high-level behaviour conducted by humans (Santani et al., 2018). If a person moves from one place to the other, the captured raw data is enriched to answer human-centred questions such as how long does the person stay, does the stay duration significant enough to be considered, what defines significance and how to decode that from data. These types of analysis go beyond the mere labelling of GPS data to build a semantic enrichment process that is human-centric.

This new paradigm is commonly discussed using the concept of trajectories and episodes. ***A trajectory*** is a continuous temporal stream of geographical coordinates collected from GPS sensors (such as smartphone-embedded GPS). The temporal boundaries of a trajectory are application-specific. Some applications are interested in daily behaviour, and accordingly, each trajectory record the mobility behaviour of one day. Other implementations may consider weekly or monthly behaviour and consequently define a trajectory. ***Episode*** is another commonly used concept which determines a segment of the trajectory (i.e. sub-trajectory) that represents a specific event. For instance, a daily trajectory may consist of home, work and walking episodes. ***A stay-point*** is a particular type of episode used to divide trajectories based on time and distance threshold. For instance, if the distance between adjacent points in a trajectory is less than 10 metres and the duration between the start and end of the adjacent points – that meet the distance constraint – is greater than 5 minutes, then the underlying segment is considered a stay-point. However, as we shall explore in this paper, decisions about thresholds values are application-dependent and impacted by the selected algorithm and the collection media.

Besides the basic concepts, the process of enriching raw data involves one or more of the following tasks to facilitate knowledge extraction: segmentation, annotation, and behaviour recognition (Figure 3.1). Segmentation and annotation sub-processes are driven by the target behaviour and thus facilitate the mining of behavioural knowledge. For instance, if the target behaviour is walking, then the segmentation step divides location data into walking and non-walking episodes. Next, contextual data sources are consulted to associate episodes with places details (i.e. annotation). Most applications employ an external knowledge source to add context-specific data to raw coordinates (Nogueira et al., 2018). We refer to these additional sources as a context data source (CDS). Foursquare – a geographical information repository – is a CDS example that maps a pair of longitude and latitude values to a place's details such as name and category. Consequently, knowledge – such as the person's preferences for walking (e.g. park, lake) – are extracted from the annotated trajectories.

Segmentation, annotation and behaviour recognition are not the only way of classifying studies related to semantic enrichment. Other studies related to semantic trajectories are classified into modelling, computation, and applications (Albanna et al., 2015; Chakri et al., 2015). Modelling class groups studies that focus on how GPS data is modelled and used in the database. Studies that focus on the segmentation and annotation of trajectories are assigned to the computation class. Lastly, studies of predicting or visualising behaviours that are derived from GPS data are classified under applications.

To this end, we recognise several studies that contribute to the goal of better understanding the challenges of enriching location data. Some of which partially address the enrichment processes (Prelipcean et al., 2017), while others consider trajectories in a broader domain that include human and non-human trajectories (Parent et al., 2013). Nevertheless, this is the first effort, to the best of knowledge, that systematically target smartphone-based trajectories.

In the next section, we introduce the general framework proposed by this paper. We articulate the main layers and components of the model. Then, since our framework is motivated by a systematic literature review, we explain the methods and analyse the results (Section 3.4 and Section 3.5) before we dive into the details of each component in our model (Section 3.6).

3.3 Design

We propose a layered and structural design to detail the semantic enrichment processes. Our work is built on a systematic literature review of enriching smartphones' location collected in-the-wild. We expand processes in Figure 3.1 to lay out the internal structure of each process as well as the interactions across processes. We map each one of those processes to a layer in the proposed framework and derive the details from the

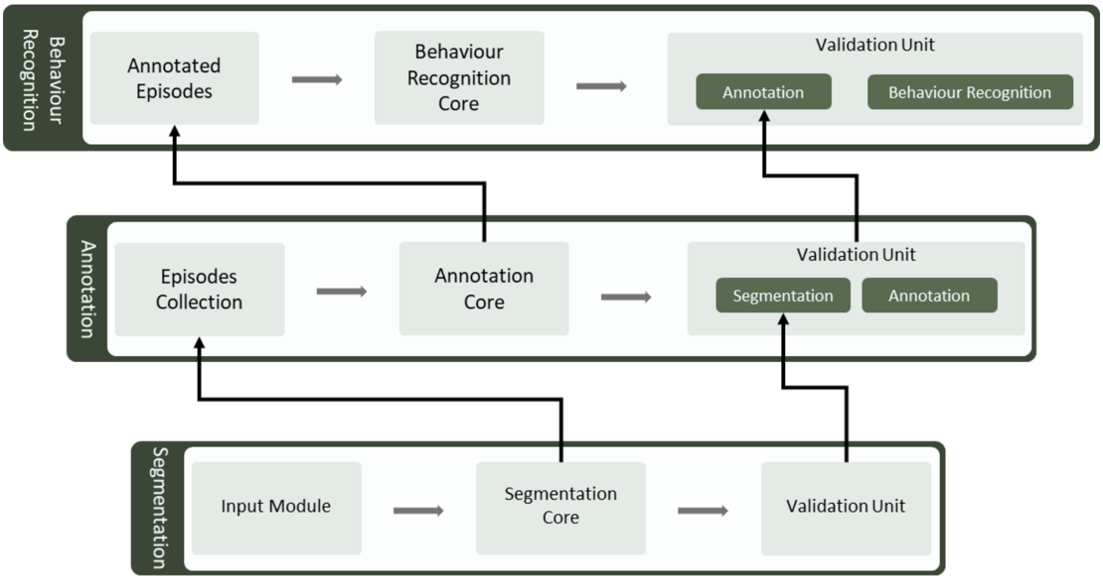


Figure 3.2: Structural framework for enriching GPS trajectories.

conducted review.

As the first task of the enrichment process, segmentation is the base layer of our model. Within this layer, we have three main components (Figure 3.2). An input module that interfaces with the collection device and stores the movement logs according to the collection requirements. Off-device-based enrichment may have constraints for collecting and offloading GPS data that differ from the online-based enrichment (Yan et al., 2011). The collected raw data are passed to the segmentation core, which manages the activities responsible for dividing the spatiotemporal stream. These activities include tasks such as data cleansing, compression and episode identification. Once the core unit produces the application-specific episodes, the validation step assesses the correctness of the extracted episodes by comparing them against the available ground truth. When no ground truth data is available, episodes extracted from other sources such as CDR or accelerometers can be compared against the ones extracted from the GPS sensor. Accordingly, the number of matches can determine the correctness of the extracted episodes. In the absence of data from these sensors, it may not be possible to validate the exact time of the extracted behavioural events (i.e. stay-points). However, it is still possible to know whether a person has visited a particular place, although we are not sure about the correct times of this visit.

An episodes collection is generated from the segmentation layer and used as input to the annotation core component. Episodes in the collection may be represented by one or more GPS points. A stay-point is referenced by a longitude and latitude pair that represents the mean value of the multiple GPS readings within a boundary of d meters. In contrast, move-points contain multiple GPS references that form the route taken by an individual to travel from a stay-point to another. The core unit defines the annotation rules to filter out episodes that do not require annotation. For instance, if the application is interested in stay-points only, then move-points will be ignored during the annotation process. Accordingly, episodes are annotated either externally or internally using the appropriate CDS. Decisions about selecting the best candidates and the reliability of the semantic labels are made within the annotation core units.

The validation step assesses the accuracy of the annotation (according to the CDS selected by the core unit) and evaluates the impact of segmentation errors on the overall results. Measuring the accuracy can be done differently according to the experiment design and goals. For instance, in our previous work (Ibrahim et al., 2021a), participants are asked to confirm the correctness of the detected and annotated stay-points. Accordingly, the number of corrected places are used to estimate the accuracy of the external CDS (Foursquare in this case). Also, participants can see the start and end time of the recognised events (i.e. stay-points) and report potential segmentation errors concerning the start and end of those events. This integrative evaluation enables a more comprehensive analysis of the results and enhances the ability to separate segmentation errors from the ones caused by the annotation process.

The annotated episodes are used as inputs to the core unit of behaviour recognition layer [1]. The implementation of this unit is tightly coupled with the application goal. Recognition of social anxiety (Huang et al., 2016) differs from the identification of user routine, and therefore they yield different implementation of the core component. The behaviour recognition layer also has a validation unit to measure the accuracy of recognising behaviour. Similar to the annotation layer, errors are either produced by the process of behaviour recognition or propagated from lower layers.

In this paper, we propose workflows for each one of the core units described above. These workflows are built on the insights extracted from the systematic review of the literature. Next, we explain how this review is conducted before we dive into the details of the workflows later.

[1]We used the term *behaviour recognition* to specifically refer to the human behaviour targeted by this process and to avoid potential confusion that could result from the use of more general terms such as pattern recognition.

3.4 Method

We conduct a broad review of the literature and adopt the PRISMA statement for reporting the systematic review of enriching GPS data collected via smartphone devices. To comply with the objective of understanding smartphone-specific requirements and challenges, we include studies if:

- They use smartphones devices as the source of raw GPS data.
- They analyse multi-day continuous real-world data. Short studies do not reflect a continuous and longitudinal data collection that can help understand daily behaviours of individuals.
- They collect data continuously and unobtrusively (i.e. in a passive manner). Studies that require smartphones to be in a specific posture or attached to the participants' bodies are excluded.
- The movement data are collected using GPS sensors only. Studies of location data gathered by other means – such as location dairy delivered through smartphones or check-ins tweets – are excluded.
- They analyse smartphone trajectories and are not restricted by specific conditions such as vehicle-only trajectories.
- They are full papers written in English published before March 2020.

Guided by the above inclusion criteria, two researchers have reviewed the papers separately and selected the related papers. A second cycle of the review was conducted to resolve disagreements about the selected papers.

3.5 Results

We report the results of a cross-domain search using Google Scholar and two-domain specific searches in ACM and ScienceDirect. The search query and retrieved results are detailed in Table 3.1.

We selected 21 papers that meet the inclusion criteria that we specified later. The details of the process through which these 21 papers were selected are illustrated in Figure 3.3.

Table 3.1: Search query and returned results.

Search query	*Google Scholar*	*ACM*	*ScienceDirect*
("semantic enrichment" OR "semantic annotation" OR "trajectory segmentation") AND (trajectories OR trajectory) AND GPS AND (smartphone OR "mobile phone" OR "cell phone")	639	49	47

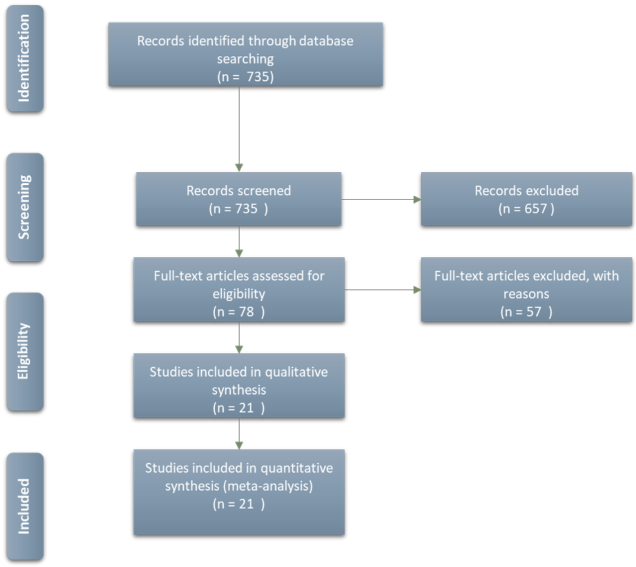


Figure 3.3: Summary of the literature systematic review.

We classify the selected papers according to the semantic enrichment task. If a paper, for instance, focuses mainly on dividing movement records into episodes, then it is categorised as segmentation only. It is possible to have a paper that covers more than one process. In that case, the paper category would be based on the process order in the chain (e.g. segmentation and annotation classified as annotation). Figure 3.4 shows the distribution of selected studies across processes. Papers about annotation contribute the most to the enrichment process; whereas behaviour recognition and segmentation-specific papers are studied equally. However, 84% of the included papers refer to the segmentation process within the context of the papers' main contributions.

Among the selected works, the most recent publication of a segmentation-only paper was in 2018 (Wang and McArthur, 2018). Between 2016 and 2018, 83% of the behaviour recognition papers were published. The first paper about annotation

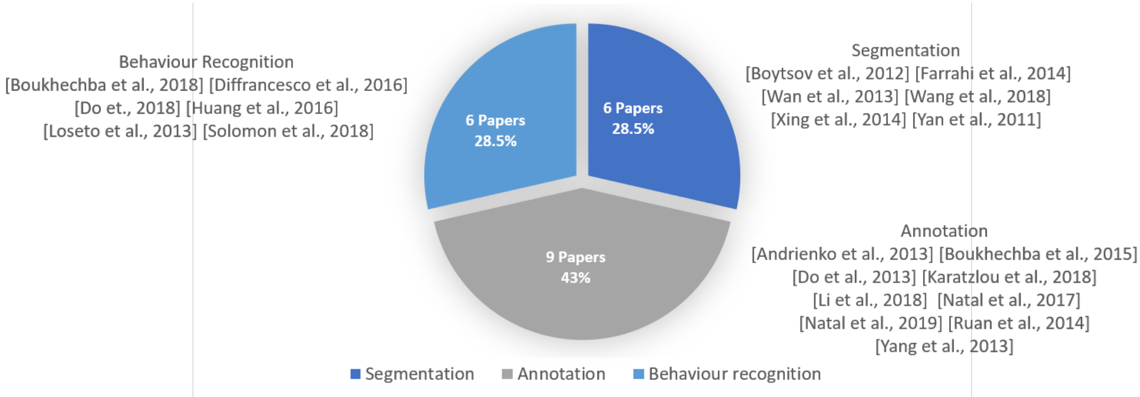


Figure 3.4: Papers distribution across the three sub-processes of the semantic enrichment.

published in 2013, and since then, every year except 2016 has at least one annotation-related article (Figure 3.5).

3.5.1 Duration and Sample Size

Papers vary in their studies duration with a minimum of 5 days and a maximum of 18 months. The mean and median of studies duration are 200 and 75 days, respectively. These statistics differ significantly according to the dataset property. Analysis based on public data set such as Lausanne campaign (Kiukkonen et al., 2010) and reality mining (Eagle and Sandy Pentland, 2006) has a mean and a median of 405 days; while these statistics change significantly to become 58 and 30 days for the mean and median respectively when those in charge of the experiment collect the data. However, process-based analysis of duration does not reveal any significant differences compared to the overall duration results.

Similar to the duration statistics, the analysis of the sampling size of the overall process is consistent across subprocesses. It ranges from 1 to 228 participants with a median of 9 participants and a mean of 37. Two studies have not specified the sampling size, and no study rationalised the determination of the selected size through statistical analysis such as power analysis.

3.5.2 Validation

In 71% of the papers, results are evaluated based on ground truths collected directly from the participants. In the absence of participants' inputs, researchers substitute the

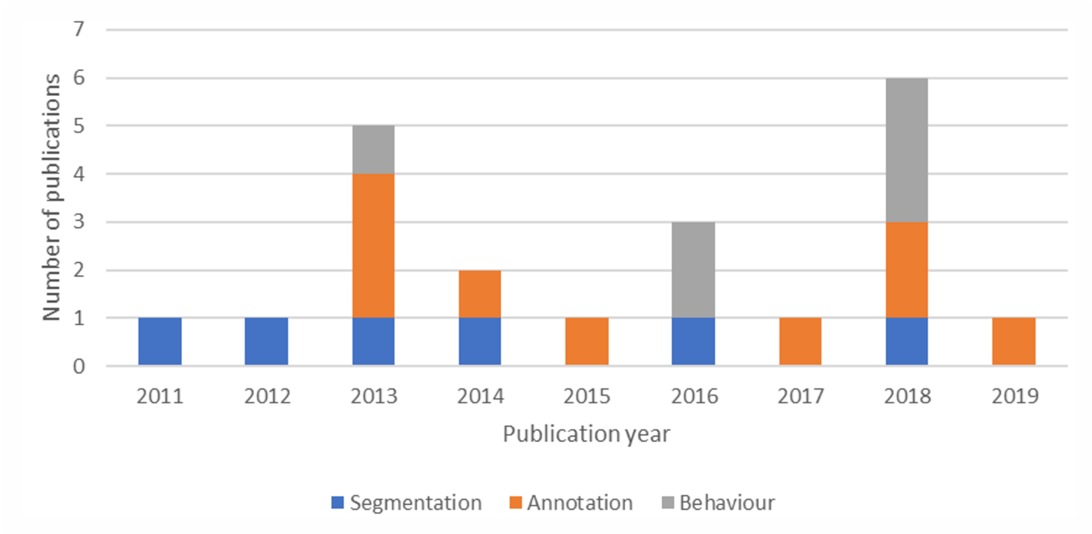


Figure 3.5: Publications per year.

ground with synthetic data (e.g. ask external assessors to predict the trajectory details and compare their generated results with human prediction). Two out of six papers on segmentation (Wang and McArthur, 2018; Wan and Lin, 2013) collect ground truth, while most of the annotation papers (78%) built a ground truth to evaluate their inferences. All papers about behaviour recognition analyse their results based on ground truths collected about the examined behaviour; however, they do not gather data about other sub-process to investigate the possibility of error propagation and how that may impact the accuracy of the behaviour recognition process.

The reported results can be divided into three categories (Table 3.2). The first one is descriptive results that explain and clarify the outputs based on the collected ground truth, mainly in terms of precision/recall or general statistics. The main theme of this category is the absence of results comparison in which the outputs are not compared with papers of a similar process or any other baselines. On the contrary, the second type of papers depends on a comparison that distinguishes its proposed method from a comparable process in the literature. Between the two categories, the third one is based, where ground truths and extracted features are modelled as a supervised learning task. The contributing factor under this approach is measured as to how feature engineering based on semantic enrichment techniques improves the classification task. Consequently, the results of various machine learning algorithms are compared based on baseline features and enrichment-based features but not against papers of similar interests.

Table 3.2: Papers distribution based on the category of the reported results.

Category	***Percent***	***Papers***
Descriptive results	57% (12 Papers)	(Yan et al., 2013), (Do and Gatica-Perez, 2014), (Yan et al., 2011), (Huang et al., 2016), (Wan and Lin, 2013), (Andrienko et al., 2013), (Boukhechba et al., 2015), (Boukhechba et al., 2018), (Boytsov et al., 2012), (Difrancesco et al., 2016), (Farrahi and Gatica-Perez, 2014), (Loseto et al., 2013)
Comparative results	19% (4 Papers)	(Wang and McArthur, 2018), (Karatzoglou et al., 2018), (Li et al., 2018a), (Xing et al., 2014)
Machine learning-based	24% (5 Papers)	(Natal et al., 2017), (Natal et al., 2019), (Ruan et al., 2014), (Santani et al., 2018), (Solomon et al., 2018)

3.5.3 Study Data

Lastly, 71% of studies conduct real-time experiments to collect location data. The remaining studies use public dataset collected longitudinally under natural settings. Also, 33% of the papers that reported the use of a public dataset did not provide details about their utilisation of data (e.g. if they use the entire dataset for evaluation or how they split evaluation/test folds when training models).

80% of the reported studies conduct an off-device analysis of the collected data to one or more of the semantic enrichment sub-processes. The annotation holds the most significant portion of the off-device analysis, with 90% of the papers consult external APIs to annotate episodes. Google, Foursquare and OSM APIs are the primary annotation providers reported by these articles. The on-device analysis starts to emerge recently (the first study was published in 2017) to improve data privacy and mainly tackle the segmentation process.

3.5.4 Summary of selected papers

In this part of the results, we summarise the selected papers in Table 3.3 to set the stage for the in-depth discussion reported in the next section.

Table 3.3: Summary of the reviewed papers (ordered first by the sequence of the workflow operations and then by the year). *GT*: Ground Truth; *SG*: Segmentation; *AN*: Annotation; *BR*: Behaviour Recognition; *NR*: Not Reported.

Author	**Process**	**Purpose**	**Size**	**Duration**	**Dataset**	**GT**
Yan et al. (2011)	SG	Perform real-time cleaning, compression and segmentation of trajectories	1	NR	Public (Kiukkonen et al., 2010)	No
Boytsov et al. (2012)	SG	Examine the configuration values of the segmentation clustering algorithm	NR	NR	Public (Kiukkonen et al., 2010)	No
Wan and Lin (2013)	SG	Propose an approach to segment trajectories based on the performed activities	1	4 months	Private	Yes
Xing et al. (2014)	SG	Apply topic modelling to segment trajectories	10	11 months	Public (Kiukkonen et al., 2010)	No
Farrahi and Gatica-Perez (2014)	SG	Apply topic modelling to segment trajectories	25	1 year	Public (Kiukkonen et al., 2010)	No

Continued on next page

Table 3.3 – Continued from previous page

Author	Process	Purpose	Size	Duration	Dataset	GT
Wang and McArthur (2018)	SG	Propose a density-based segmentation of trajectories	3	27 to 68 days	Private	Yes
Andrienko et al. (2013)	AN	Extract places characteristics from movement records	1	NR	Private	No
Do and Gatica-Perez (2014)	AN	Label trajectory without relying on geo-location information	114	18 months	Public (Kiukkonen et al., 2010)	Yes
Yan et al. (2013)	AN	An application independent platform for segmentation and annotation	185; 6 with GT.	18 months	Public (Kiukkonen et al., 2010)	No
Ruan et al. (2014)	AN	Predict labels of the visited places	8	6 months	Private	Yes
Boukhechba et al. (2015)	AN	Annotate the visited places and predict the performed activity	1	NR	Private	Yes

Continued on next page

Table 3.3 – Continued from previous page

Author	Process	Purpose	Size	Duration	Dataset	GT
Natal et al. (2017)	AN	Annotate segments based on the performed activities	10	10 days	Private	Yes
Karatzoglou et al. (2018)	AN	Combine data and knowledge driven approaches to annotate episodes	6	5 weeks	Private	Yes
Li et al. (2018a)	AN	Annotate the visited places and predict the performed activity	1	13 days	Private	Yes
Natal et al. (2019)	AN	Annotate segments based on the performed activities	22	20 days	Private	Yes
Loseto et al. (2013)	BR	Recognise the user habits from location data	1	14 months	Private	Yes
Difrancesco et al. (2016)	BR	Predict social functioning in schizophrenic patients	5	5 days	Private	Yes

Continued on next page

Table 3.3 – Continued from previous page

Author	Process	Purpose	Size	Duration	Dataset	GT
Huang et al. (2016)	BR	Predict social anxiety based on movement records	18	10 days	Private	Yes
Boukhechba et al. (2018)	BR	Predict social anxiety based on movement records	228	2 weeks	Private	Yes
Santani et al. (2018)	BR	Recognise drinking behaviour from movement records	241	10 weeks	Private	Yes
Solomon et al. (2018)	BR	Predict demographics from movement records	45	6 months	Public (Mirsky et al., 2016)	Yes

3.6 Discussion

Motivated by the reported results, we extracted insights from each sub-process and accordingly present the details of the core units in the proposed framework. By doing so, we facilitate the planning of the enrichment process as well as the tracking of potential errors. In the subsequent sections, we describe the core units of the framework (Section 3.3) as task-based workflows. These workflows integrate the extracted insights per sub-process into a consistent set of steps to facilitate a proper semantic enrichment of smartphone trajectories. Later in this section, we provide a SWOT analysis of the extracted findings to help researchers identify and plan future directions.

3.6.1 Segmentation

In most cases, segmentation is the first process toward enriching GPS trajectories. It divides movement records into episodes that reflect behavioural units in the real world. Behavioural units are cognitive-driven segments that compose behavioural sequences. Trajectory segments represent behavioural units within the context of GPS data. Based on our analysis of the selected papers, we identify three main perspectives to segmentation, namely, segmentation base, segmentation algorithm and collection strategy.

3.6.1.1 Segmentation base

The segmentation base is the reference point that guides the segmentation process. It could be a behavioural reference, such as walking, or a statistical-based point inferred from calculations on movement records. Trajectories represent continuous behavioural units in real life captured through GPS devices (Nogueira et al., 2017). Behavioural-based referencing implement top-down approaches to trajectory segmentation that divide GPS sequences based on the goal of a particular behaviour. If the motive is to find the places where a user prefers to stay, then stillness and movement are potential segmentation references that divide movement records to stay and move points. The choice of stillness and movement (i.e. behavioural references) and the variables (aka. hyperparameters) that identify those references (e.g. time and distance threshold) are decided according to heuristics and prior behavioural knowledge (Montoliu and Gatica-Perez, 2010).

On the other hand, bottom-up approaches adopt a statistical mechanism to merge atomic segments and form a larger one consisting of statistically homogeneous state. An atomic segment is a small unit of the captured trajectory that is used as the building

Table 3.4: Segmentation bases and behavioural references adopted by the selected papers.

Paper(s)	*Method*	*Reference*	*Episode*	*Variables*
Yan et al. (2011)	Bottom-up	Features correlation	• Jogging • Walking • Standing	• Sliding window size • Atomic segment size
Boukhechba et al. (2015)	Bottom-up	Cluster similarities	• Stay-point • Move-point • Transition-point	• Sliding window size
Xing et al. (2014)	Bottom-up	Topic modelling	• Stay-point • Move-point	• Word distribution • Topic distribution • Transition tendency
• Yan et al. (2013) • Do and Gatica-Perez (2014) • Huang et al. (2016) • Wang and McArthur (2018) • Wan and Lin (2013) • Andrienko et al. (2013) • Boukhechba et al. (2018) • Boytsov et al. (2012) • Difrancesco et al. (2016) • Li et al. (2018a) • Natal et al. (2017) • Natal et al. (2019)	Top-down	Stillness	• Stay-point • Move-point	• Distance • Duration
Ruan et al. (2014)	Top-down	Stillness	• Stay-point • Move-point	• Distance • GPS points count

block of an episode. For each atomic segment, a feature vector is calculated, and a sliding window is used to compare segments by their underlying features such as duration or covered distances. If, for instance, a mobile device collects movement logs every two minutes, an atomic segment of six minutes would contain three data points. If we define the duration of the sliding window to be 30 minutes, then each sliding window would cover five atomic segments. Atomic segments within the same sliding window are sequentially compared based on a similarity measure of their feature vectors. Based on the similarity score, consecutive segments are merged if they are identified as 'similar'.

Table 3.4 summarises the papers in our review based on the segmentation reference. As general observations, stillness is the most common reference among studies that adopt top-down segmentation approaches. 98% of papers following this approach employ the covered distance and duration as the episode determinants, with one paper rely on the number of GPS readings inside the cluster instead of the duration to define the episode boundaries.

In contrast, bottom-up approaches focus more on the movement patterns and classifying episodes based on their movement status. Bottom-up approaches are built under the hypothesis of sampling rate regularity. They address sampling irregularities through data imputation; a process that aims to fill the frequency gaps in data collection. However, this leads to different issues related to the reliability of the imputation process and how errors may propagate through the entire process.

3.6.1.2 Segmentation algorithms

Statistical and behavioural referencing just set the guidelines for the subsequent processes. Each segmentation reference has several implementation options, and the selection among them is dependent upon other factors such as the application domain and the collection media. In this section, we discuss the various implementations from an algorithmic perspective.

We classify the segmentation algorithms into two classes, density-based and sequence-based. Density-based algorithms (e.g. DBSCAN and K-means) employ clustering techniques to group similar locations entries. As these are parametric algorithms that rely on hyperparameters to accomplish their tasks and compute items similarities, the type of segmentation reference determine the values of those hyperparameters. Topic modelling is another type of clustering that stems from the literature of natural language processing (Xing et al., 2014). In this approach, point similarities correspond to latent topics and episodes are formed (i.e. clustered) based on their closeness to each topic.

Sequential algorithms preserve the temporal order of trajectories' points during the process of generating segments. These algorithms study the relations between consecutive entries of movement records and rely on behavioural rules applied on spatiotemporal features embedded into trajectories. An example is a rule to define stillness behaviour applied to the distance between temporally adjacent points. If the geodetic distance[2] of two points is close to 0, then a stillness behaviour is detected; otherwise, the person is moving, and a stay-point is defined accordingly.

All bottom-up approaches (Table 3.4) are sequential in nature as they adopt a sliding window to process the movement data sequentially. On the other hand, top-down approaches utilise both algorithmic types to divide trajectories.

Both algorithmic classes, however, are mostly built on the assumption that movement records are sampled at regular time intervals. Although this assumption may go well with some controlled implementations, it does not reflect a real-life smartphone-based collection of location data as we shall explain next.

3.6.1.3 Collection strategy

The third perspective, collection strategy, emphasises the crucial role of the collecting mechanism on the semantic enrichment operation. Different devices export different challenges to the process of collecting and processing trajectories. Smartphones, as the epicentre of this paper, introduce power optimisation techniques to increase the battery life, which in turn influence the sampling rates of GPS sensors. This shows why segmentation methods should consider how movement tracks are captured and sampled to improve their performance.

Generally, we discuss two types of sampling strategies for collecting GPS data. The first one is a time-based strategy that assumes a fixed and guaranteed sampling rate of collecting location's data. Algorithms written under these assumptions do not have to deal with irregularities of sampling intervals as the device is configured to enforce the sampling constraints. An event-based strategy is a different approach in which recording GPS data is only triggered if predefined conditions are met. For instance, an app may be set to collect the GPS data only if the participant is connected to a WiFi network. Event-based strategy imposes an additional data preparation task to deal with sampling irregularities and potential data loss. However, unobtrusive observing

[2]Geodetic distance is the shortest path between two GPS readings.

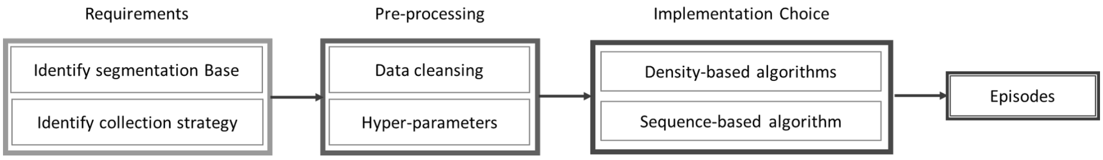


Figure 3.6: Workflow of the segmentation process.

may involve both strategies since built-in power optimisation techniques as well as interaction preferences impact the streaming of data.

Each one of these perspectives addresses one of the challenges specific to the enrichment process. Segmentation-base deals with the contextual ambiguity. By identifying the reference of the segmentation, we limit the scope of the possible outcomes and orient the operation based on the specified reference. Within the context of smartphones, segmentation algorithms should be designed to deal with the application-specific sampling rate challenges. If the time intervals regularity of recording GPS data is guaranteed, state-of-the-art density-based algorithms may fit well. However, if such regularity is not guaranteed, as in many naturalistic smartphone-based settings, then density-based algorithms are more likely to fail (Nogueira et al., 2017; Montoliu and Gatica-Perez, 2010). This last point shows how the algorithmic and data collection perspectives interplay to enrich raw GPS data.

Figure 3.6 shows a workflow that we propose to illustrate the smartphone-based segmentation process. Before dividing trajectories, it is essential first to identify which segmentation reference – behavioural or statistical – is more appropriate to the enrichment objective. At the same time, the requirements of the collection process should be identified based on the device and application capabilities. Once the collection criteria and segmentation reference are decided, a pre-processing step is initiated to smooth and clean the collected data. This stage is essential as the collection strategy influences the expected noise, and therefore the applied pre-processing techniques may differ. The segmentation process and collection constraints drive the choice of the implementation algorithm. A density-based algorithm could be the right choice when the sampling rate is guaranteed, while sequential-based is more flexible when dealing with unexpected sampling rates. After applying the segmentation step, application requirements may require additional postprocessing of the resultant episodes (e.g. merging consecutive stay-point episodes if they are separated by a move-point that is less than 2 minutes long).

3.6.2 Annotation

Annotation is the process of assigning descriptive labels to behavioural units extracted from trajectories. The goal of this process is to bridge the semantic gap between raw location data and human cognition by naming the extracted episodes. This descriptive annotation refers to more than the segmentation-based driven annotation. For instance, stillness as a behavioural reference for the segmentation process implies the existence of two basic labels: move episode and stay episode. These two labels are embodied into the segmentation base and therefore do not provide additional knowledge. Semantic annotation of such trajectories would go beyond these built-in labels to include more descriptive data like the type of place (e.g. restaurant, café) or the purpose of visit (e.g. socialising, studying). In this article, we classify research as annotation-related if they target filling the semantic with information different than the one presented by the segmentation phase.

Within the context of smartphone-based trajectories, we found two main basics for the annotation process, namely, activity-based and land-based. The former aims to understand the activity performed within the episode's boundaries and annotate the episode accordingly. If a person is having a meeting at a café place, then the corresponding activity is labelled as 'meeting'. In contrast, the land-based method would have labelled the same episode as 'café' as its emphasis is on the land use of the property on where the episode takes place.

It is noteworthy that the primary affordance of the property may sometimes describe both the activity and the land-use, such as in the case of a dance club. Although the two approaches may seem to overlap in this case, their outputs differ according to the target user. If the episode is extracted from a trajectory of a worker in that dancing club, the activity-based annotation yields 'working' episode. Alternatively, customers' episodes are annotated as "dancing" since they are expected to do so. This example shows why one approach cannot be substituted for the other.

Annotation source is another annotative aspect that considers the contextual data source (CDS) necessary to enrich movement trajectories. Traditionally, CDSs are classified as either external or internal sources. When the annotation data is retrieved from a remote conduit, that exists outside the phone such as Google or Foursquare spatial APIs, the source is considered external. Inputs to external sources are either single or multiple coordinates per episode based on the output of the segmentation process. Segmentation algorithms do not necessarily produce a single representation for episodes and consequently shift the burden of this task to the annotation phase. In

Table 3.5: Annotation methods and Contextual Data Sources (CDS) as reported by the selected papers.

Paper(s)	***Method***	***Source***	***CDS***
Andrienko et al. (2013)	Land use	Internal	Temporal features.
Yan et al. (2013) Boukhechba et al. (2015) Boukhechba et al. (2018) Difrancesco et al. (2016) Loseto et al. (2013)	Land use	External	Open Street Map API.
Huang et al. (2016) Ruan et al. (2014)	Land use	External	Foursquare API.
Loseto et al. (2013) Karatzoglou et al. (2018)	Land use	External	Google places API.
Wang and McArthur (2018)	Land use	External	Barefoot.
Loseto et al. (2013)	Land use	External	LinkedGeoData
Natal et al. (2017) Natal et al. (2019)	Activity	External	Google places API.
Do and Gatica-Perez (2014) Li et al. (2018a)	Activity	Internal	Self-reported labels.

that case, the label for each point within the episode is first retrieved from the external provider. Then a postprocessing task is initiated to select the representative label based on application-specific criteria.

Internal sources employ contextual data collected explicitly or implicitly alongside the GPS data. To annotate episodes based on explicitly collected data, users are required to annotate the extracted segments, and then a classification task is conducted to train a model that utilises additional features (e.g. temporal features) to predict the annotation. However, this approach requires users to update the extracted episodes regularly. Alternatively, sensor data collected passively along with location data, are used as a source for annotation. For instance, Wi-Fi labels may contain useful information such as the name or category of the place, which provides a valuable source for

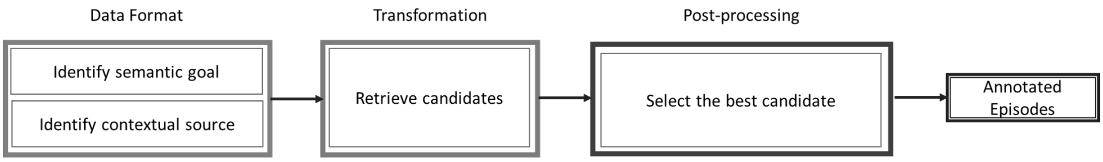


Figure 3.7: Workflow of the annotation process.

annotation. These contextual data are used by the annotation algorithm to predict the labels of the extracted episodes. Table 3.5 summarise annotation papers based on the discussed views.

Although the above perspectives suggest multiple methods to the challenge of filling the semantic gaps in the location data, only one paper Karatzoglou et al. (2018) provides a mechanism to facilitate the evaluation of the external Geo-location provider. Nevertheless, none of the selected papers rationalised the selection of specific CDS nor provides a comparison or inter-reliability test of the accuracy of various Geolocation APIs, despite the significance of this matter.

Based on the above, Figure 3.7 shows our proposed workflow that integrates the elements and perspectives of the annotation process. First, it is essential to identify the goal of the annotation task as it determines the details of the subsequent processes. Point of interests (POI) systems that aim to provide suggestions based on users' preferences (e.g. preferred cuisine) may adopt a land-based approach to extract visited places and generate recommendations accordingly. On the other hand, behavioural informatics systems may focus more on annotating episode based on the underlying activity to serve their objectives. Once the goal is identified, labels are generated either internally or externally. Although external sources typically provide APIs to facilitate their functions, raw data may require additional pre-processing and manipulation to utilise these functions. The produced results are post-processed to select the best annotation candidate. This step may include synthesising data from several sensors (e.g. Wi-Fi and Bluetooth) to select the most probable description of episodes under consideration, or it may vote on the best candidates from labels provided by external annotators based on inter-reliability tests.

3.6.3 Behaviour Recognition

Existing studies have addressed the behaviour and knowledge extraction from GPS trajectories. The application domains addressed by these studies shape their differences. Requirements for extracting knowledge from health-related applications differ from the ones within the context of marketing, for instance. Moreover, some of those researches reside outside the context of enriching raw location data. For example, instead of going through the process of transforming raw GPS data to semantically improved trajectories, an application may utilise check-ins data as input to the behavioural mining task. This approach does not address the challenges caused by the potential limitation of enriching raw data and how that may affect the knowledge extraction process. Therefore, to account for the influence of potential challenges inherited from other sub-processes (e.g. segmentation), in this article, we address the mining of behavioural knowledge that arise as a result of the semantic enrichment. Other location-based knowledge extraction studies lie outside the scope of this analysis.

Accordingly, we find that the analysis granularity is the central aspect that distinguishes smartphone-based behaviour identification methods. Episode-based behavioural analysis mine features related to the trajectory components and how these components – and their latent features – correlate with each other to form a behaviour. Trajectory components are the different episodes' types that compose a trajectory. If a trajectory is segmented based on the stillness attribute of the embodied event, then stay-points and move-points are the components of that trajectory. Accordingly, episode-based knowledge extraction may study episodes of similar types, such as counting the frequency of similar episodes to get the number of visits to a specific place. The place in this instance represents a stay-point extracted from the collected trajectories. Alternatively, the knowledge extraction may target the inter-relations across different episodes type. In this case, multiple episodes' types (e.g. stay-point and move-point) are investigated to determine behavioural phenomena such as preferred transportation mode (i.e. move-point features) for each visited place (i.e. stay-point).

Although episode-based approaches may study the temporal relation between sub-components of the trajectory, these methods do not preserve the full sequentiality of the entire trajectory. To clarify this idea, consider the example of extracting the preferred transportation mode for each place. Episode-based approaches would study the relationship between the episode representing the visited place and its surroundings to understand how a user moves to and leave the target place. With multiple stay and

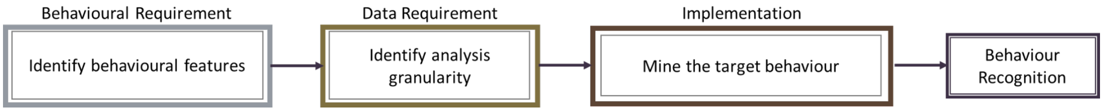


Figure 3.8: Workflow of the behaviour recognition process.

move points (i.e. places and transportation modes) reside on a single trajectory, a similar approach is conducted to extract knowledge. However, from an episode-based perspective, only the temporal aspect between the adjacent components is required by the analysis, as other sequential features (e.g. temporal sequence of two places) does not contribute to the learning process.

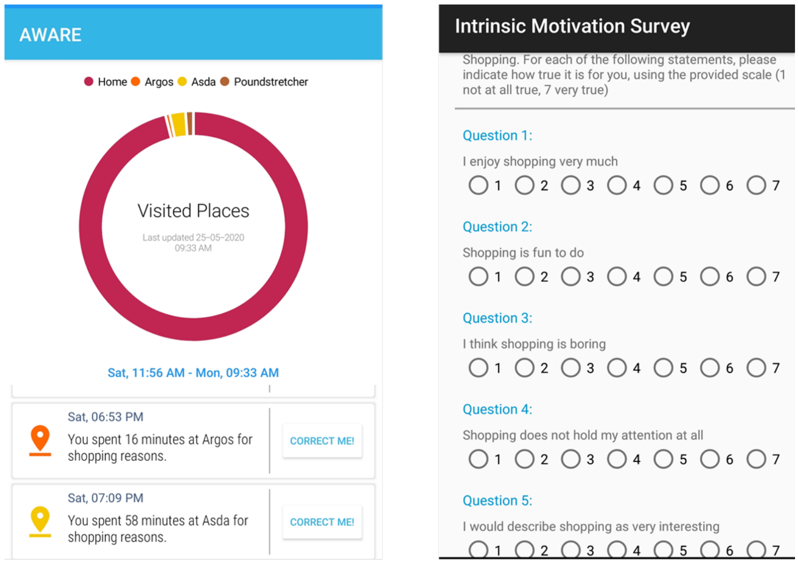
In contrast to episode-based analysis, the trajectory-based approaches extract knowledge encoded in an entire trajectory rather than its building components. Accordingly, the sequentiality of episodes is preserved to facilitate the mining of behavioural patterns. An example of this method would be the extraction of daily habits from multiple daily trajectories. In this scenario, the behavioural habits may be extracted based on aggregating similar trajectory and performing sequential pattern analysis.

Moreover, trajectory-based mining may target movement records co-located across multiple devices. One example would be a trajectory modelling to discover chasing behaviour from two smartphones. In this case, two trajectories are examined to decide whether a person is being followed by another person. This is also an example of inter-personal analysis that involves more than one person in the mining process.

Figure 3.8 concludes the proposed framework by depicting the workflow of the last semantic enrichment process. The first step is to identify the features of the target behaviour since this will impact the granularity choice, as explained above. Recognising episode-based behaviour has different requirements than trajectory-based. Once the granularity level is decided, the mining strategies vary according to the selected methodology. Rule-based and machine-learning approaches are possible mechanisms to achieve this goal.

3.6.4 Validation and Error handling

The correctness of the outputs for each process in our framework is essential to the semantic enrichment validity. Therefore, studies related to semantic enrichment should be designed in a way that facilitates the understanding of how potential errors propagate across the framework. In this subsection, we discuss the design of a real-world



a) Places b) IMI

Figure 3.9: Interfaces for the two developed plugins: (a) the Places plugin allows participants to examine and correct their annotated locations; (b) the IMI plugin presents participants with a set of validated questions that can be used to evaluate the correctness of the recognised behaviour.

study that we have conducted to extract personal interests from GPS data [3]. As part of the experimentation process, seven participants were asked to assess the correctness of the semantic enrichment processes. The collection period lasted for three months, and 200,000 GPS data points were collected.

To locate the errors of each layer's processes, we provide a plugin within the study app to examine and correct the enriched GPS data (Figure 3.9a). Each time a visit to a new place is detected (i.e. a stay-point), the participant receives a notification inviting them to confirm or correct the detected place. To validate the segmentation correctness; the start and end time of the visits are provided. Also, the names of the nearby places are shown if a participant decides to correct the label that is assigned to a detected stay-point. To support the analysis of errors related to behaviour recognition (in this case, behaviours of personal interests), we add a further plugin within the study app (Figure 3.9b). This plugin presents an adaptation of an Interest/Enjoyment subscale that is widely used to assess interest associated with a given activity (Monteiro et al., 2015; Ryan, 2018) [4].

[3] The details of this study and how interests are recognised are published in (Ibrahim et al., 2021a).
[4] The original scale is called Intrinsic Motivation Inventory (IMI) and developed by (Ryan, 2018).

This design allows us to separate errors caused by a process such as annotation from errors caused by an algorithm intended to recognise behaviour from raw GPS data. For example, in the same work, the extracted places are analysed to extract behaviours motivated by personal interests. Without separating errors, the performance of the ranking algorithm could be impacted by the segmentation and/or the annotation errors. This is because the algorithm can classify wrongly identified stay-points as a potential interest. When we rely on the corrected data, the algorithm's performance can better reflect its ability to recognise behaviours motivated by interests. This is a result of avoiding errors that propagate from segmentation and annotation layers.

3.6.5 SWOT Analysis

To better benefit from this review findings in helping future research on semantic enrichment of GPS trajectories, we summarise and organise limitations and opportunities found in the selected papers into a SWOT analysis framework. SWOT framework is a decision-making technique used to identify Strengths, Weakness, Opportunities and Threats related to a specific application (Dyson, 2004). Researchers can use this tool strategically to analyse and plan their research through (i) embracing strengths and potential opportunities, (ii) addressing weaknesses and (iii) mitigating potential threats (Thomas et al., 2014; Sondaal et al., 2016). We provide a planning strategy for semantic enrichment of smartphone-based location. Our implementation of SWOT derived from the conducted review and the created model. The first researcher drafted the analysis, and through an iterative feedback process with the third and fourth authors, the final analysis was reached. The presented analysis can help researchers better envisage the potentials of future contributions.

3.7 Conclusion

We propose a structural framework and planning strategy to streamline the semantic enrichment process of smartphone location data. Our work helps in understanding the challenges and limitations of the existing methods and how they interrelate within the entire process. Moreover, the layered approach and workflows facilitate the understanding of error propagation through the enrichment operation. Next, we plan to instantiate this framework with real-world smartphone data to examine the effectiveness of the proposed methodology in facilitating the analysis of mobile-specific challenges.

Future reviews can be conducted on smartphone-based digital phenotypes such as device usage and notifications. These reviews could study the extraction of behavioural units from other smartphones' sensors and organise the involved process in a human-oriented manner. Collectively, this work and the suggested reviews on streamlining the processes of extracting human behaviour from digital phenotypes can improve the human-centric research based around smartphone's longitudinal data.

SWOT analysis: Strengths and Weakness

Strengths

- The existence of reliable segmentation algorithms for detecting stay and move points. These algorithms show a good performance across different implementation settings with different threshold values.
- External APIs, such as Google Places and Foursquare, provide an easy to use interface to enrich the extracted episodes with places details.
- The existence of public dataset contributes to the research in general. This contribution is evident in activity-based annotation studies since such studies do not require the exact coordinates to analyse the underlying activities.

Weaknesses

- The impact of collection strategy is mostly overlooked in several studies, and if mentioned, the impact is reduced to battery-related issues although other factors such as explicitly turning the sensor off and on are possible reasons for data loss.
- None of the included studies that employ external sources measure the reliability of the provided annotations nor the potential role of such reliability on the subsequent process (i.e. behaviour recognition).
- Data-driven approach to annotation relies on participants periodic inputs to build a training model. Such an approach is intrusive and impacted by confirming and cognitive biases.
- Anonymised public datasets do not contribute to annotation based on external sources since it requires the exact coordinates to get the place's semantic.

SWOT analysis: Opportunities and Threats

Opportunities

- Collect and publish smartphones' datasets collected unobtrusively and longitudinally to facilitate the study of smartphone-based challenges of semantic enrichment.
- Study the privacy issues when consulting external APIs for episodes annotation
- Measure the impact of collection strategies on the performance of the segmentation algorithms.
- Study the impact of the recent reliance and heavy usage of smartphones on the data collection reliability.
- Estimate and improve the reliability of the external annotator.
- Improve the behavioural inferences based on places' categories.
- Develop a dynamic approach to facilitate the discovery of error propagation and distinguish enrichment-based errors from behavioural-based ones.

Threats

- Conducting an experiment without supporting the sampling size (especially small sizes) with statistical analysis may negatively impact the generalisability of results.
- Designing an experiment under the assumption of a guaranteed sampling rate can fail under naturalistic settings.
- Disregarding the reliability of the external annotator may impact the produced results
- Direct mapping of land use categories to behaviour can reduce the accuracy of behaviour recognition since it does not consider the issue of multipurpose places (e.g. Cafe can be mapped to studying, relaxing or socialising)
- Population-based assumptions could ignore personalised routines (e.g. a person may have Tuesday and Wednesday as her weekend as opposed to Saturday and Sunday)

Chapter 4

Smartphone-Derived Buying Behaviour

For each one of the three behaviours targeted by this thesis, we need to extract features of behavioural events that are derivable from the corresponding smartphone's sources. In Chapter 2, we have extensively reviewed the methods of extracting events of phone usage behaviour from smartphones. Unlike phone usage events that fully depend on the existence of smartphones, events of mobility and buying behaviours do not require smartphones to occur. Therefore, in the previous chapter, we presented a framework for extracting events of mobility behaviour encoded into GPS data. The proposed framework is based on a systematic review of the literature as location-based phenotyping is a relatively well-known topic. The extraction of buying events (the third behaviour targeted by this work) from notification texts is explored in this chapter. Features of these events (e.g. product names, channels) need to be extracted in order to proceed with our approach of detecting interest. However, notification-based digital phenotyping is relatively new, with few studies compared to the location-based. Therefore, we wrote a paper showing how features of buying behaviour can be extracted from the notification text. We benefited from the data that we collected from the pilot and the main study. The primary purposes of the paper are to (1) compare different techniques of filtering out notifications that are irrelevant to buying behaviour and (2) assess various methods of extracting features of buying events through notification-based digital phenotyping.

We have included the data from the pilot and main studies. Including data from the main study supports the achievement of the above purposes and, at the same time, does not influence and contradict the summative evaluation detailed in the following chapters. As stated in Chapter 1, we used the data of the pilot study to inform our main experiment's decisions. This includes the methods of extracting features related to buying behaviour that we selected. This paper evaluates methods for extracting buying-related information without associating that with the person issuing the notifications.

The main content of this chapter is a paper authored by: *Ahmed Ibrahim, Sarah Clinch and Simon Harper*. The title of the paper is: ***Extracting Behavioural Features from Smartphone Notifications***. The paper is currently under review. For this thesis, we edited some formatting styles, such as the sizes of some tables for consistency and readability reasons.

Author contribution

Ahmed Ibrahim designed the study, carried out the data collection, analysed and synthesised the results and wrote the paper. Sarah Clinch and Simon Harper provided continuous feedback throughout all the stages of the study, offered advice and discussion and contributed vital edits to the paper's writing.

Abstract

A significant proportion of smartphone notifications are indicative of human behaviour (e.g., delivery updates for purchased items, physical activity summaries, and notification of updates to subscribed content). However, present attempts to understand human behaviour from smartphone traces typically focus on sensors such as location, accelerometer and proximity, overlooking the potential for notifications as a valuable data source. In this paper, we propose a general framework that provides end-to-end processing of notifications to understand behavioural aspects. We realise the framework with an implementation that tackles the specific use case of establishing prior buying behaviour from associated notifications. To evaluate the framework and implementation, we conduct a longitudinal user study in which we collect more than $250,000$ notifications, from twelve users, over an average of three months. We apply knowledge-based and machine learning techniques to those notifications to assess the tasks of the proposed framework. The results show a substantial difference in the performance between the methods used to extract behavioural features from the collected notifications.

4.1 Introduction

Smartphone notifications are short messages that can convey a variety of information about the phone holder (Sahami Shirazi et al., 2014; Pielot et al., 2014). Notifications, unlike other forms of short messages, are issued by an underlying app, not the user. Their issuing can be caused by a random event (e.g. arrival of general public health emails) or be a direct and immediate consequence of a user actively engaging in a behaviour (e.g. purchasing confirmation email). This work leverages the latter (we refer to it as active notifications) in an effort to recognise the features of users' actions that cause the issuing of notifications. The recognised knowledge can be of value for applications that require personalisation, such as recommender systems.

Existing studies on smartphone notifications focus on analysing: their topics based on the generating apps (what) (Fischer et al., 2010; Mehrotra et al., 2015); users' interactions with them (how) (Sahami Shirazi et al., 2014; Visuri et al., 2019) and the contexts in which they are received (where) (Mehrotra et al., 2017a; Turner et al., 2017). These analyses serve the goals of finding interesting notifications (Visuri et al., 2019), predicting appropriate delivery times (Mehrotra et al., 2015; Turner et al., 2017) and understanding factors controlling the user's response to notifications (Mehrotra et al., 2016b). However, those studies do not consider events that cause the issuing of notifications (why). For instance, buying behaviour may cause the issuing of notifications about receipts emails. Also, notifying a user of new chat messages is an example of notifications caused by communication behaviour. Using notifications to extract knowledge of behaviours that cause their issuing can boost studies centred around the smartphone holder's behaviour (rather than those solely centred around the person's interactions with the notifications themselves).

In this work, we aim to recognise behavioural features for actions conducted by the smartphone's holder, a process known as digital phenotyping (Jain et al., 2015). More specifically, we focus on notification-based digital phenotyping within the context of buying behaviour. We propose a framework to determine and analyse the content of notifications that result from buying activities. For instance, when buying a product, a notification about the receipt could be generated by (1) an email app, (2) an SMS client or/and (3) the related shopping app if installed. The issuing of notifications is not dependent on whether a behaviour is conducted using the smartphone itself or the shopping app. A user may buy a product from a desktop and still receive a notification on the phone if the related app is connected to the same account. Once relevant notifications are determined and extracted, their contents are analysed to understand behavioural features such as: what a user buys and how often similar products are purchased.

To evaluate our method, we conducted an in-the-wild study and passively collected more than $250,000$ notifications using an app installed on the participant's smartphone. Twelve participants joined our longitudinal study that lasted for an average of three months. We assess the feasibility of knowledge-based and machine learning techniques in classifying notifications as either relevant or irrelevant to buying behaviour. Text mining techniques are used to extract the features from the notification text. More specifically, we design a Natural Language Processing (NLP) pipeline to recognise the product names from the texts. Semantic data can be added to the extracted features through an external source. The additional information about the product names can be used to better understand the bought items and group the ones representing the same interest together.

Our approach showed that relying on machine learning algorithms enables us to filter out irrelevant notifications more efficiently when compared to a knowledge-based approach. Moreover, the application of NLP techniques facilitates the extraction of product names from the notification content. We compare a method that relies on the global frequencies of the words in the notification text against the use of named entity and a clustering-based approach suggested by (Li et al., 2018b). The results show that the word frequency approach substantially outperforms the other alternatives.

To summarise, our contributions are threefold:

- **A framework for notification-based digital phenotyping** that introduces an approach to extract behavioural features from the text of smartphone notifications. The proposed framework has the potential to process notifications captured unobtrusively over an extended time period using smartphone-based passive sensing.

- **An application to understand buying behaviour from notification-based digital phenotyping** that provides an implementation case of using the proposed framework to understand features of buying behaviours.

- **An evaluation based on a real-world data set** that provides results based on a naturalistic setting reflecting and addressing potential issues in the notification texts. Also, the naturalistic setting provides a way to understand the proportion of notifications a specific behaviour represents from the total number of notifications received.

4.2 Related work

The subject of this work intersects with smartphone studies in two aspects: notification analysis and digital phenotyping.

A large body of studies on notifications analysis seeks to understand interruptibility from users' contexts and interactions with notifications (Mehrotra et al., 2016b). Analysis outcomes are used to tailor notifications delivery based on each user's situation (Mehrotra et al., 2015; Sahami Shirazi et al., 2014). When a user is at a movie theatre or business meeting, notifications can be deferred in order not to interrupt or disturb the user (Mehrotra et al., 2017a). Changing the smartphone's tone to a less disturbing mode (such as the silent one), rather than delaying notifications, is an alternative way of tailoring the notification delivery according to the user's situation (Visuri et al., 2019). Besides the spatial context, the user's situation may be based on other contexts. For instance, temporal context can be understood by studying a relationship between users' past interactions with notifications and the times of these interactions (Mehrotra et al., 2016b). Understanding this relationship helps prevent notifications delivery at times when a response is not expected by the user, such as bedtime (Fischer et al., 2010). These studies regardless of the contexts and interactions used to analyse notifications, consider interruptibility as either a binary state or a multifaceted case. The former classifies situations as either interruptible or uninterruptible (Poppinga et al., 2014). In contrast, the latter includes instances where a user might accept or even prefer to be partially interrupted if the notifications are related to a specific topic of interest (Turner et al., 2017).

Some notifications analysis studies go beyond merely understanding interruptibility to study factors controlling the user's response to notifications (receptivity) (Mehrotra et al., 2016b; Westermann et al., 2016; Schulze and Groh, 2014). Users' responses are interactions such as viewing, touching and dismissing notifications. A third category of studies proposes applications and frameworks that can help in collecting notifications in-the-wild (Weber et al., 2019). Applications of this last category can support studies related to the other aspect in which our work intersects, digital phenotyping.

Smartphones allow for passive data collection (i.e. without intervention from a user) that can be considered highly indicative of the user's environment and behaviour. It is in this context that Jain et al. (2015) coined the term *digital phenotyping* to refer to the process of using an individual's interaction with digital technologies to derive indicative behavioural markers. Digital phenotyping can be realised through longitudinal studies and digital tools such as smartphones and smartwatches. For the scope of this work, we focus on smartphones' related studies.

Potential features of buying behaviour from smartphone interactions can be seen through shopping apps. Mobile shopping is a recent form of transformations that occur in commerce and are caused by advances in technology and digital devices (Tang, 2019; Tyrvainen and Karjaluoto, 2019). Buying is only a motive among others that drive the use of shopping apps (Huang and Zhou, 2018; Tang, 2019). Examples of other motives may include prices comparison, products sharing, and reviews probing (Huang and Zhou, 2018; Chopdar et al., 2018). However, studies around the use of mobile shopping apps rely primarily on surveys and interviews rather than digital phenotyping. In contrast, studies use digital phenotyping with other sources, such as location data, to predict shopping activities. For instance, if location data point to a clothing store, the person's activity is predicted as shopping related (Ibrahim et al., 2021a).

The closest work to ours is the one introduced by Li et al., (2018b). In that work, notifications are classified into templates, and then knowledge entities are recognised as parameters of these templates. However, to conduct the classification task as suggested by Li et al., (2018b), the existence of a large corpus of smartphone notifications generated by a large number of apps is required. Unlike the proposed approach, we present a general framework that provides end-to-end processing and instantiates it with a use case. Our method can be tailored based on each individual's data. Therefore, we rely only on the notifications generated from the user's device to understand behaviour aspects. This is the first approach, to our knowledge, that aims to understand what individuals buy from the notifications received on their devices.

4.3 Our approach

We adopt a novel approach that processes notifications from a behavioural aspect. Unlike the user interaction aspect used by the existing literature, *our approach seeks to extract behavioural features of actions that can cause the production of notifications*. We start from the behaviour and accordingly signify the importance of notifications. The relation between the issued notification and the target behaviour is the essential part. For instance, suppose a user dismisses an active notification (that is issued as a result of actively engaging in a target behaviour such as buying a product). In that case, that notification is still more critical and relevant to our analysis than a clicked one that is not related to the same target behaviour. This behaviour-centric analysis is precious

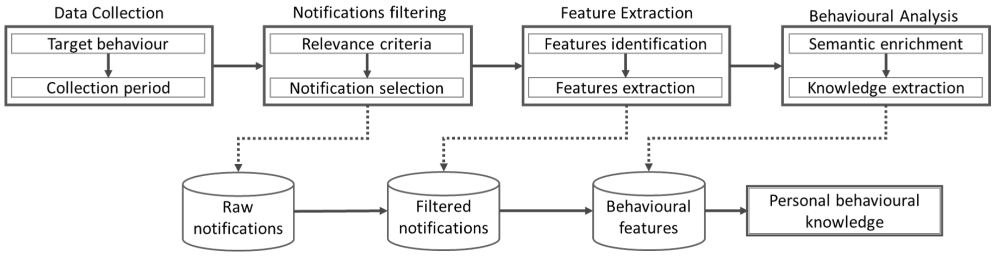


Figure 4.1: The proposed framework for extracting behavioural knowledge from raw notifications texts. Dashed arrows point to the data sources on which the framework steps are applied.

for understanding personal preferences, whether related to buying or other behaviour such as reading and communication. Understanding these behavioural preferences are the basis for building personal recommendations from smartphone notifications.

Since no work to our knowledge provides a way for processing notifications this way, we propose a general framework for doing so. Then, we use the proposed framework to extract features of buying behaviours and discuss the methods and implementation details.

4.3.1 The proposed framework

The proposed framework (Figure 4.1) is based on four main tasks: data collection; notifications filtering; features extraction; and behavioural analysis. The first task is to know the period required to collect data for the behaviour that is targeted by the analysis. In the second task, we define the criteria for filtering out notifications such that only those related to the targeted behaviour are kept. The texts of the kept notifications are processed next, in the third task, to extract the target behaviour's features (e.g. the book name if reading is the target behaviour). In the fourth task, semantic information about the extracted features (e.g. the book category) is obtained to be used for further behavioural analysis. For instance, the category of purchased books can be used to analyse which books a person prefers. Each one of the four tasks has two steps. Decisions about each step's implementation are heavily influenced by the target behaviour (step 1) and by decisions made in steps precedent to the one under consideration.

4.3.1.1 Data collection

Identify target behaviour As centred around behaviour, our framework starts by identifying behaviour that is targeted by the analysis. When selecting a behaviour, it is essential to consider the possibility of extracting indicators of that behaviour from notifications. This is because indicators of some behaviours might rarely be present in smartphones' notifications, which may hinder the ability to discover them. Also, the determination of target behaviour does not have to be limited by installing related apps on the smartphone. As we shall see in the next section, receipts of buying behaviour can be seen in emails rather than apps related to buying activities such as Amazon or eBay.

Determine the collection period The notifications amount issued by smartphones' apps varies based on the behaviours that cause their issuing. Notifications caused by social communications behaviours (e.g. using WhatsApp or Instagram) may form a larger portion if compared to shopping-caused ones (Li et al., 2018b). Also, the nature of how people practise behaviour can play a role in this variation. For instance, a daily reader may receive more notifications compared to another who reads on a monthly or weekly basis. Therefore, knowledge about the target behaviour plays a pivotal role in determining the period length of notifications data needed to obtain behavioural insights.

It is essential, however, to emphasise the importance of not limiting the collection of notifications at this point of the analysis on any constraints (e.g. collecting notifications from reading apps only to analyse reading behaviour) . As we shall see next, some filtering techniques may require the entire dataset to filter out irrelevant notifications.

4.3.1.2 Notifications filtering

Identify the base of selecting relevant notifications This task is a preparation step before filtering out irrelevant notifications. A decision is made on how notifications related to a target behaviour are selected. Knowledge-based and machine learning are two potential approaches to doing so. An example of the former would be relying on notifications from shopping apps in understanding buying behaviour. Machine learning approaches start from the data and aim to find patterns in notifications related to the target behaviour. NLP techniques are typically used to prepare the texts for processing, and then a classification or clustering task is conducted to spot patterns.

Filter out irrelevant notifications In this step, notifications that are irrelevant to the target behaviour are filtered out. Using a knowledge-based method, a simple choice would be selecting notifications based on the relevance of apps issuing them to target behaviour. For instance, notifications of apps categorised as reading-related are selected to analyse reading behaviour.

A more advance method of filtering out irrelevant notifications using a knowledge-based approach would be through the notification content. In this approach, notifications' contents are searched based on predefined keywords known to be related to the target behaviour. For instance, the content can be searched for keywords such as 'book' and 'author' to select notifications related to reading behaviour. A more dynamic approach may search the notification content using the names of apps categorised as related to reading. For instance, Google PlayStore categorises 'Kindle' app as 'Books and References'. Hence, if a user has the Kindle app installed, notifications containing 'Kindle' might be considered related to reading.

Alternatively, machine learning can be used to filter out irrelevant notifications. Using supervised learning, notifications are labelled and a model can be trained and used to classify new instances as either relevant or irrelevant. If not labelled, clustering provides a second option of applying machine learning that can group notifications. A decision can then be made on which notification groups are relevant to the investigated behaviour.

4.3.1.3 Feature extraction

Identify behavioural features In this step, the behavioural features that need to be extracted are identified. For reading behaviour, the features would be the title of the book or the book author, whereas the product name and its cost could be the features of the buying behaviour. Knowledge of the notification contents can help determine the possibility of extracting a behavioural feature. For example, although a notification from a shopping app may contain the cost of a purchased book; extracting the cost from a notification issued by a reading app might be unlikely. Although this step relies only on identifying the target behaviour, we place it later in our framework due to its direct relation to the final step.

Extract features from notifications To extract the behavioural features from the raw texts of notifications; a Natural Language Processing (NLP) pipeline needs to be implemented. With respect to this work; we propose an NLP pipeline that has tokenisation, cleaning and extraction tasks. In tokenisation, the raw text is converted to individual words. The cleaning task processes these words to signify the important ones. This processing might involve tasks such as lemmatisation and removing stop-words. The last step in our proposed pipeline, extraction, aims to process the cleaned text to extract the target behaviour's features. Extracting features may require, in some cases, a numeric representation of words. In that case, an optional step of vectorisation can be added. In the vectorisation step, transforming words into a numeric form can be done through techniques such as word2vec (Mikolov et al., 2013), words of bags (Zhang et al., 2010) and TF-IDF (Joachims, 1996).

4.3.1.4 Behavioural analysis

Feature enrichment What extracted from notifications in the previous step are the features of the target behaviour. If, for instance, a notification is related to buying behaviour, a potential feature would be what a person buys (i.e. the product name). However, the number of extractable features is limited by what is contained in the notification text. Therefore, to better understand the extracted features, additional semantic information about these features is needed. In this step, the extracted features are enriched with semantic information. Since this information is not present in the notification text, an external source is used to retrieve the needed information. For example, the product category can be retrieved from an external API to add a semantic value to the bought product's name. Hence, a name feature of a bought product such as "A Promised land by Obama" can be understood as being a book.

Knowledge extraction In this final step, behavioural analysis is conducted to obtain knowledge about the person. This analysis is based on the extracted features and the semantic information about them. For instance, if a person reads many books related to sports, this may indicate an interest in the sport. The breadth and depth of this analysis depend on the semantic enrichment level. As more information about the extracted feature is added, a more in-depth analysis can be achieved. We can illustrate this using the example of reading sport-related books. Suppose a second level of the book category is retrieved (e.g. the sport type) as part of the enrichment process. In that case, we can use notifications to study the type of books, within a specific category, that interests a person (e.g. tennis books within the category of sports).

4.3.2 Buying behaviour

Our proposed framework can be used to extract behavioural features from the notification text. In this section, we identify buying as the target behaviour and detail the following steps according to the proposed framework. Buying behaviour includes any act that involves a payment to get a product or a service. So, the act of paying for a taxi ride or for getting a gaming device are examples of buying behaviour. Following the identification of the target behaviour, the second step aims to determine the appropriate data collection period. Studies related to buying via smartphones show a dependency between the frequency of buying and the type of the purchased product (Beatriz, 2021; Mohsin, 2021). For example, individuals use their smartphones to buy their basic needs of food and drink on a weekly basis. In comparison, lower purchase rates are noticed (bi-monthly or monthly) when it comes to other commodities such as clothes and electronic devices. Based on that, a study of purchasing behaviour should not be less than a month if it aims to observe different types of purchased products. However, this period may be shortened or extended according to the analysis goal. Studies that only target the basic needs of food and drink may decide that two or three weeks are enough, whereas others may decide to go for longer periods. Also, the basis on which the collection period is determined may differ. Although we choose to rely on previous studies, others may decide to conduct a dedicated study to determine the collection period based on a specific set of requirements.

To filter out irrelevant notifications, we manually classified notifications into three classes[1]: actual_buying, marketing_only and non_buying. Notifications labelled as actual_buying are the ones that are generated as a direct result of a buying activity (i.e. active notifications). Examples of those include receipts and payment appreciations. The class of marketing_only includes buying recommendations that contain at least one product or brand name. Since recommendations are typically personalised based on previous buying activities (Behera et al., 2020), we leverage those notifications as they might be of interest to the buying behaviour of a person. Since both actual_buying and marketing_only represent notifications of buying behaviour, we will use the term buying_related when we want to refer to both of these classes. The third category, non_buying, contains notifications that do not include marketing offers nor result directly from a buying behaviour. Based on these categories, a knowledge-based

[1] The first author conducted the classification of these notifications. Since buying notifications are typically generated by apps related to that, the need for additional annotators is minimised. Nonetheless, future work may improve the classification with such a procedure.

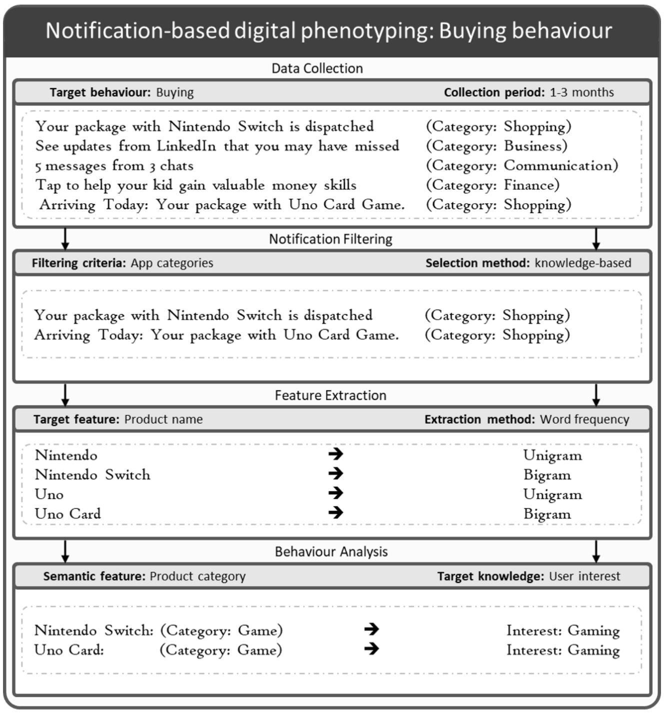


Figure 4.2: Using passively collected notifications to extract behavioural features of buying behaviour.

approach (that benefits from the notification' meta-data) or machine learning (mainly classification) can be used to filter out irrelevant notifications.

Once the set of relevant notifications is determined, we process the texts to extract behavioural features. We aim at recognising what a person buys from her/his notifications. Therefore, a natural language processing task has been designed to extract the product name from the notification text. We propose an approach that relies on the global word frequency to extract what a person buys. Accordingly, product names, such as those illustrated in Figure 4.2, are expected to be extracted.

In the proposed approach, we benefit from the short and direct nature characterising the notifications' text. Within the context of buying behaviour, notifications may form order receipts, status updates or delivery information. The sent text is expected to have a combination of common words (e.g. "your order" or "delivery updates") and the product names. For instance, a message such as "your package with *brand_name* blanket will be delivered tomorrow" is issued by Amazon when you buy a blanket. The brand_name is expected to be less common than the other words in similar notifications. Therefore, we take advantage of this observation and rely on the word commonality to extract the product name. Specifically, we compute the word frequency in the English language for each word on the notification text and select the n words that have the least values.

As the extracted features (e.g. product name) may need further information to make sense of them, their semantic can be improved through an external knowledge provider. SerpApi[2], for instance, can provide additional information about the product names and be used for a better understanding of the extracted features. Doing so would make it easier to understand that Nintendo Switch, for instance, is a gaming device. Also, such understanding can help in grouping similar products together under one category (e.g. Nintendo Switch and Uno cards are gaming products).

4.4 Evaluation

We use the proposed framework to understand features of buying behaviour from 250,532 smartphones' notifications collected from 12 participants in-the-wild. Our evaluation targets the notification filtering and the feature extraction steps of our proposed framework. Specifically, we aim to assess the following:

1. The performance of knowledge-based and machine learning methods in selecting notifications relevant to buying behaviour.

2. The performance of multiple NLP techniques in extracting the features of buying behaviour.

4.4.1 Method

Our experimentation includes longitudinal mobile data that was passively sensed using the AWARE mobile sensing framework (Ferreira et al., 2015). We recruited 12 participants and collected notifications from them passively as they go about their normal daily activities. All data captured within the app was collected from Android devices and sent to a secure server at The University of Manchester. Procedures for our study was reviewed and approved by the Department of Computer Science Ethics Committee at The University of Manchester (Reference: 2019-7817-12726).

Participants were recruited using poster advertisements displayed in public areas of The University of Manchester and surrounding buildings, and on social media. Due to COVID-19 restrictions enforced in United Kingdom in early 2020, the recruitment

[2] https://serpapi.com/

Table 4.1: The total number of notifications per participant (Total), collection duration in days (Duration), average number of daily notifications per participant (Daily). ($\bar{x}$) is the calculated sample mean.

	Pre-processing			**Post-processing**		
	Total	Duration	Daily	Total	Duration	Daily
P1	11,804	96	122.96	2,760	96	28.75
P2	16,794	65	258.37	3,809	65	58.60
P3	7,750	58	133.62	314	52	6.04
P4	4,762	53	89.85	24	10	2.40
P5	4,889	55	90.85	178	29	4.36
P6	53,180	144	369.31	27,894	144	193.71
P7	30,709	117	262.47	13,044	117	111.49
P8	97,262	134	725.84	40,419	134	301.63
P9	9,137	55	166.13	3,230	55	58.73
P10	9,646	68	141.85	7,232	68	106.35
P11	4,486	50	89.72	1,957	50	39.14
x̄	**22765.36**	**81.36**	**222.81**	**9169.18**	**74.55**	**82.84**

process had been impacted, and we had to only rely on online advertisement. A total of 12 participants ultimately agreed to participate, all but one of whom were students at The University of Manchester (seven undergraduates, four postgraduates). Participants were supplied with an information sheet prior to participation and had the opportunity to ask further questions prior to consent. Participants were rewarded for their participation with three months of Netflix subscription or equivalent Amazon voucher.

All participants provided written consent before being guided to install the AWARE app on their personal mobile devices. Participants were asked to keep the installed application running and to carry their phones as they normally do. They were advised that the application would automatically send their data to our backend server, but only when connected via WiFi, and that we anticipated no noticeable negative effects on battery life.

Participants enrolled on the study over a staggered period based on when they chose to respond to recruitment advertisements. The first participant (P1) began data collection on January 22, 2020, and the final participant (P12) began on March 3, 2021. One participant (P12) was excluded because of technical difficulties related to the app and phone compatibility. The collection period for the remaining 11 participants ranges from 50 days (P11) to 144 days (P7), with an average period length of approximately three months (81.36 days).

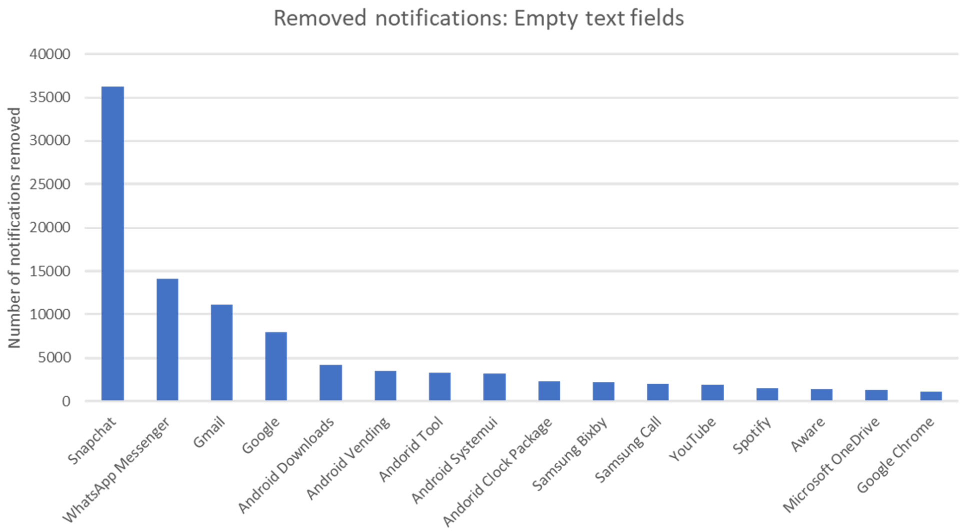


Figure 4.3: The per app distribution of notifications removed due to empty text fields. Only apps having more than 1000 notifications with empty texts are shown.

4.4.2 Dataset

After excluding P12's data, the entries in our dataset has become 250,419 notification entries. The mean and median number of notifications per participant are 22765.36 and 9646.00, respectively. The largest number of notifications collected from a single participant was from P8 (97,262 notifications), whereas P11 has the least contribution of (4,486 notifications). Table 4.1 details the notifications per participant.

Each entry of the dataset contains the notification time, title, app, category and text. To collect these data, we have added a plugin within the AWARE app. Our plugin uses Android capabilities and Google PlayStore to populate notifications' data. To retrieve the category of the app issuing the notification, our plugins queries the PlayStore website using the package name that is provided by Android system[3]. If the app's category is not found, we label the category as "Unknown". Typically, "Unknown" categories result from installing apps that are not listed on Google Play. Lastly, we anonymise numbers and emails contained in the notification text to preserve the participant's privacy. Consequently, a notification that says "You received 2 emails from john@example.com", is stored as "You received * emails from ****". This step is essential and needed to comply with the university's ethics requirements.

In this work, we rely on the text field of each notification entry to extract and

[3]Package names are different from the app names. For example, "eBay" is an app name while "com.ebay.mobile" is the package name as retrieved by Android.

Table 4.2: Examples of notifications classified as actual_buying, marketing_only and non_buying.

Class	Examples
actual_buying	• Your package will be delivered tomorrow. • Your parcel from Whitworth Pharmacy is due today.
marketing_only	• Just landed the size exclusive adidas Originals BC Trainer. • Deal on Symphonized NRG Wood from your wish list.
non_buying	• Check out all our offers before they end. • Tomorrow Leap Day How will you use it.

analyse data related to buying behaviour. To properly conduct the analysis, we first perform a pre-processing task that aims to remove notifications with empty text or those that only contain symbols and special characters. The latter mainly results from the non-English text that is not correctly encoded. Accordingly, a total of $149,558$ notifications were excluded. Of those, $113,324$ entries have empty text fields (Figure 4.3 shows the distribution of the excluded empty text notifications per app) and $36,234$ contain only symbols and special characters . As a result, $100,861$ notifications were kept. Table 4.1 shows the notification details per participant before and after the pre-processing step.

The remaining notifications were manually classified into the three classes: actual_buying, marketing_only and non_buying (see section 4.3.2 for details). As a result, we found out that 23 apps generated 387 actual_buying notifications, and 462 marketing_only notifications were generated by 31 apps. This makes the total number of buying_related notifications 849 generated by 41 apps. The remaining $100,012$ entries, classified as non_buying, are generated by 256 apps. Table 4.2 shows examples per class from the collected dataset and Table 4.3 details those classes per participant.

Lastly, the majority of notifications (34.11%) labelled as actual_buying are generated from apps categorised as “Communication”. “Shopping” apps generated (26.36%) of actual_buying notifications. The remaining 39.53% is distributed across “Finance”, “Food and Drink”, “Travel and Local”, “Maps and Navigation” and “Business” categories with percentages of 16.54%, 16.02%, 4.91%, 1.03% and 1.03% respectively.

Figure 4.4 shows the ratio of actual_buying and marketing_only notifications for each app that generates them to the number of remaining notifications produced by these apps.

Table 4.3: Number of notifications and apps issuing them per participant for each one of the classes (actual_buying, marketing_only, buying_related and non_buying).

	actual_buying		**marketing_only**		**buying_related**		**non_buying**	
	Entries	Apps	Entries	Apps	Entries	Apps	Entries	Apps
P1	23	3	231	9	254	10	2,506	23
P2	10	3	45	5	55	7	3,754	37
P3	27	5	10	3	37	7	277	24
P4	0	0	2	1	2	1	22	6
P5	2	1	5	2	7	2	171	15
P6	15	2	0	0	15	2	27,879	78
P7	106	11	117	11	223	15	12,821	104
P8	52	3	17	4	69	5	40,350	38
P9	52	5	26	4	78	5	3,152	38
P10	1	1	0	0	1	1	7,231	25
P11	99	4	9	2	108	4	1,849	53
$\bar{x}$	**35.18**	**3.46**	**42.00**	**3.72**	**71.18**	**5.36**	**9092.00**	**40.09**

4.4.3 Results

We detail our experimentation results based on the two aims specified earlier in this section. These aims mainly target the tasks of notification filtering and feature extraction from our proposed framework.

4.4.3.1 Notification filtering

Labelling the notifications allows us to assess the ability to distinguish buying_related entries (i.e. actual_buying and marketing_only) from other non_buying data. In so doing, we use a baseline knowledge-based approach and compares it to machine learning algorithms. The knowledge-based approach benefits from the meta-data captured with the collected notifications. Specifically, we rely on the categories of notifications (retrieved from GooglePlay) and special keywords that are typically associated with buying notifications. We have checked the list of apps categories on GooglePlay[4] and classified them as either related to buying or not. Our classification is based on the categories description as provided by Google and examples of apps found under each category. Accordingly, notifications from apps categorised as "Shopping" and "Food and Drink" are classified and considered as buying_related.

[4] https://support.google.com/googleplay/android-developer/answer/9859673?hl=en

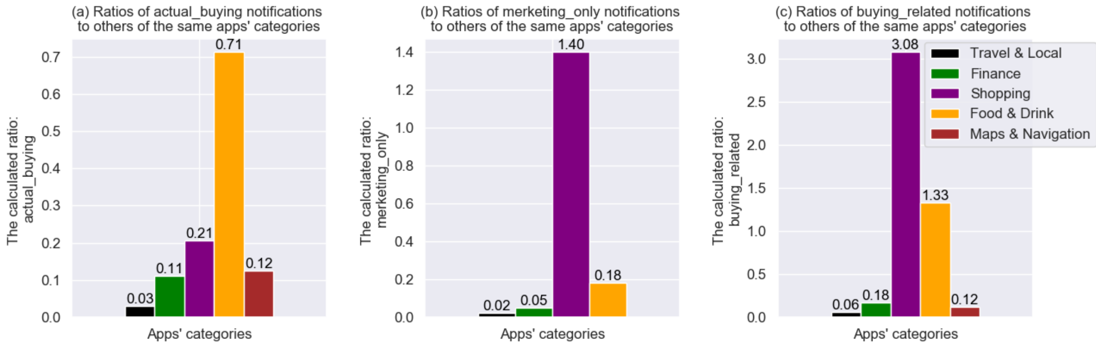


Figure 4.4: Relation between the number of buying behaviour notifications to the number of non_buying notifications issued by the same apps' category. In (b) and (c) the number of marketing_only of "Shopping" apps is higher than the remaining number of notifications issued by the same categories.

To identify actual_buying entries, notifications are searched based on keywords typically associated with buying activities. A notification is classified as actual_buying if the content contains one of the keywords specified in Table 4.4. The selection of these keywords was based on analysing emails from Enron email dataset, a public dataset that contains emails from more than 100 users (Klimt and Yang, 2004). Specifically, we randomly selected 100 emails sent by retailers such as "Amazon" and searched for receipts, order confirmations and shipping notices. Based on that, the keywords in Table 4.4 were identified.

The second approach for distinguishing notifications of buying behaviour is based on machine learning algorithms. However and as noticed in the dataset section, labelling the collected entries produced an imbalanced dataset[5]. Learning from an imbalanced dataset is typically done through undersampling, oversampling or re-weighting techniques. In undersampling, a subset of the majority class that has a number of entries close to the ones of the minority class is used to train and test the classifier. Oversampling increases the number of entries in the minority class to closely match the one of the majority. Lastly, the minority class can be assigned a higher weight (through the re-weighting technique) to mitigate the impact of the majority class.

In this work, we rely on undersampling as it is recommended when the majority class is less critical for modelling (Aggarwal, 2015). Also, unlike oversampling, undersampling has the advantage of including all the more valuable entries of the minority class (Aggarwal, 2015; Estabrooks et al., 2004). As we focus on buying behaviour,

[5]In an imbalanced dataset, the number of entries of one class is much larger than the entries of another class.

the aim is to distinguish buying_related notifications from non_buying. In so doing and guided by (Bayerstadler et al., 2016; Weihs and Buschfeld, 2021), we have randomly generated 50 subsets from the collected dataset[6]. Each set contains an equal number of entries per class. Specifically, to distinguish actual_buying from non_buying, we have generated 50 subsets that contain 387 entries per class. Also, another 50 subsets for learning marketing_only from non_buying have been generated. The total number of entries per the latter subset is 924 with 462 in each class (i.e. marketing_only and non_buying). Lastly, 50 subsets that include all buying_related notifications (849 entries) with a similar number of non_buying entries have also been created through undersampling. In total, we have created 150 subsets, 50 subsets per class. These subsets were randomly undersampled based on the entire dataset without constraints on the app's category. Hence, we refer to that as *uncategorised undersampling*.

In contrast, and to conduct an in-depth analysis, we have generated subsets based on the categories of apps producing notifications of buying behaviours (we refer to this randomisation as *categorised undersampling*). In each subset, entries are sampled only from apps issuing notifications of buying behaviour. The number of entries in each subset varies based on the number of actual_buying and marketing_only notifications that are generated by an underlying app category. For instance, shopping apps generated 102 actual_buying notifications and hence the number of entries in the corresponding subset is 204 (102 actual_buying and 204 non_buying). We have excluded subsets that contain less than 100 notifications of buying behaviours to avoid small groups. As a result, we have created additional 120 subsets from four categories ("Communication", "Shopping", "Finance" and "Food and Drink"). For each category, we have generated 30 subsets (10 per each one of the three classes: actual_buying, marketing_only and buying_related versus non_buying). The number of entries in each sample ranges from 124 to 354. Lastly, each one of the generated subsets is trained and tested independently to avoid overfitting. Therefore, the results of training and testing a specific subset do not impact the results of another subset.

The selection of the machine learning algorithms was based on the ones that H2O[7] supports. H2O is an open-source and auto-ml platform that facilitates the building of machine learning models (LeDell and Poirier, 2020). Namely, we have used: logistic regression from the Generalised Linear Models (GLM)[8], Gradient Boosting Machine (GBM), Distributed Random Forest (DRF), Extremely Randomised Trees (XRT) and

[6]We use python imblearn package for generating the subsets.
[7]`https://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html`
[8]We use GLM to refer to logistic regression to be consistent with H2O terminology.

Table 4.4: The keywords used for searching the notifications content.

Keywords	Examples
"your order"	"We shipped this portion of your order separately"
"your purchase"	"We are writing to confirm your purchase of the following... "
"your payment"	"We have received Your payment and will be shipping out ... "
"your package"	"You can come down to ... and pick up Your package ..."

Deep Neural Net (DNN). To run the data with these algorithms, we have applied an NLP pipeline (tokenisation, stop-word removal and stemming). Then, train and test vectors have been produced using the TfidfVectorizer of python sklearn. These vectors were used to train and test the various models supported by H2O[9].

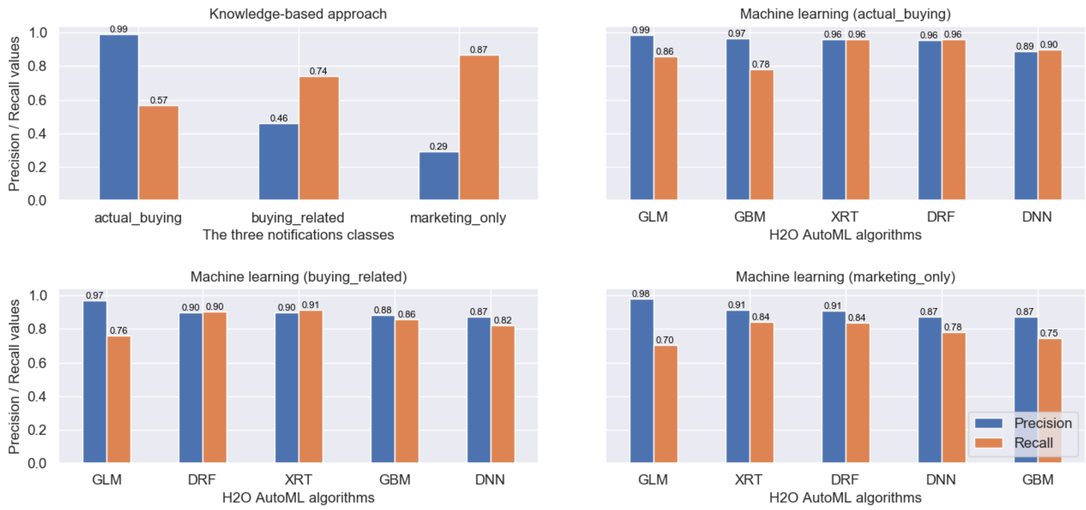


Figure 4.5: The performance of the knowledge-based approach and the machine learning algorithms (ordered by the precision values). Precision and recall values of the machine learning algorithms are averaged based on the 50 uncategorised and 40 categorised subsets in each one of notification class (actual_buying, marketing_only and buying_related).

Based on the above, we have evaluated the ability to filter out notifications irrelevant to buying from the collected dataset. We have calculated the precision and recall values for the knowledge-based approach and compared it to the five machine learning algorithms (see Figure 4.5). The machine learning scores in Figure 4.5 are calculated based on the mean precision and recall scores of the subsets under each category. For

[9]We use the default parameters for each algorithm as specified by H2O (80% used for training and 20% for testing).

instance, actual_buying class contains 50 subsets that are undersampled based on the entire dataset (i.e. *uncategorised undersampling*) and 40 that are undersampled based on the categorised one. Therefore, the precision and recall values represent the mean values of the 90 subsets. Similarly, the precision and recall values have been averaged for the marketing_only and buying_related classes (Precision: min std = 0.029, max std = 0.100, mean std = 0.043, min CI = 0.004, max CI = 0.007, mean CI = 0.005; Recall: min std = 0.050, max std = 0.143, mean std = 0.098, min CI = 0.008, max CI = 0.013, mean CI = 0.010)[10]. For each of those classes, the total number of subsets in which precision and recall values are averaged is 90. Accordingly, we found out that on the actual_buying, both GLM and the knowledge-based approach (i.e. relying on common keywords associated with buying behaviour) have the best precision performance (0.99). The remaining H2O algorithms also perform well on the precision scores for all classes of buying behaviour (scores range between 0.97 and 0.89). The precision scores for marketing_only and buying_related under the knowledge-based approach (0.29 and 0.46 respectively) are the worst among all methods. However, the recall of marketing_only notifications under the same approach (0.87) is better than all their corresponding machine learning algorithms (XRT and DRF have the highest score, 0.84).

To show the results per the subsets' groups, Figure 4.6 compares the mean scores of the uncategorised undersampling to the categorised one. We compute the mean scores of predicting buying_related notifications (i.e. both actual_buying and marketing_only from non_buying). Moreover, the mean scores for separately predicting actual_buying and marketing_only from the remaining notifications are also depicted in the figure. We show the scores of accuracy and Area Under the Precision-Recall Curve (AUCPR) as they are commonly used metrics for classification (LeDell and Poirier, 2020) (Algorithms are ranked first by the accuracy and then by the AUCPR). Of the five algorithms, logistic regression (GLM) consistently outperforms the others. The highest average accuracy score of the 150 subsets is 0.99 for the GLM, whereas GBM has the least average accuracy score (0.90). However, GLM is ranked behind in only one of the groups (categorised sampling: actual_buying vs non_buying) in which GLM ties with XRT and DRF in the accuracy score but falls behind in the AUCPR score. Nonetheless, the mean performance of all the 120 subsets under the categorised undersampling group puts back the GLM on the lead with an average accuracy score of 0.94 (compared to 0.93 for both XRT and DRF). Lastly, of both the uncategorised and

[10] CI is calculated at confidence level of 95%.

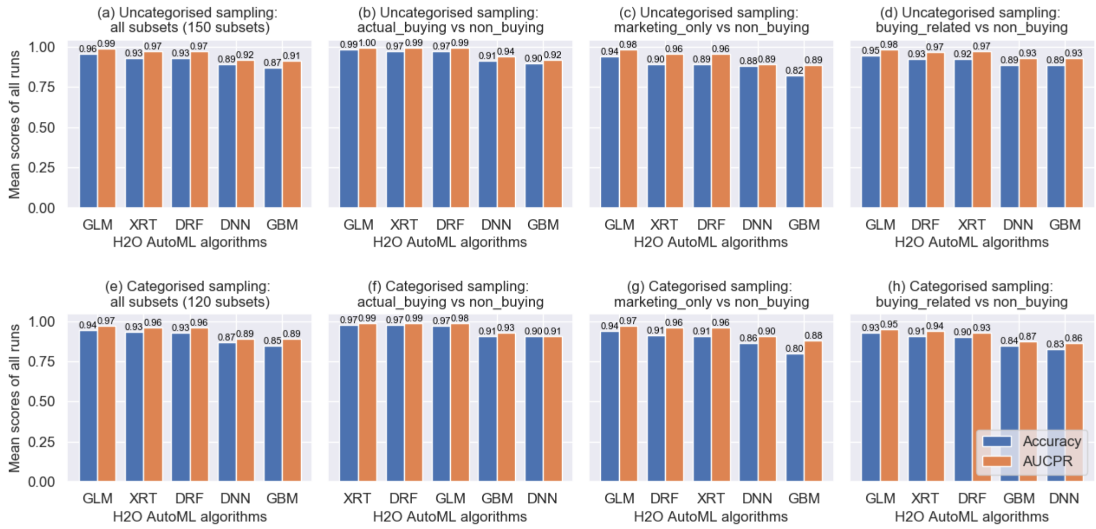


Figure 4.6: Filtering out irrelevant notifications. The top four sub-figures show the mean scores based on the *uncategorised undersampling* whereas the bottom four show the results of the *categorised undersampling*.

categorised undersampling, the least accuracy score results from using GBM to predict marketing_only notifications from the non_buying entries.

4.4.3.2 Feature extraction

The feature extraction step has two sub-tasks: features identification and the actual extraction (section 4.3.1.3). With respect to buying behaviour, we identify: "what a person buys" and "the time of buying" as the features that need to be extracted from the notifications. The event time can be directly obtained from the timestamp field of the notification entry. The product or service name, however, represents what a person buys and exists within the notification text. In this section, we present the evaluation of three methods aiming to extract the product names from the notification text.

First, we had to manually extract the names of the bought products or services from all the buying_related notifications (i.e. actual_buying and marketing_only) to serve as ground truth. Here, we will use "product name(s)" to refer to both the product and/or the service name. The total number of product names in the actual_buying notifications is 283. The majority of actual_buying entries (93.54%) contain only one product name; 4.94% contains two product names, while 1.52% contains three product names. With respect to the marketing_only, the texts under this class contain 555 product names (as discussed in section 4.3.2, each marketing_only entry is expected to have at least one product or service name). 84.20% of those marketing_only entries contain one product name only, 11.47% have two product names and 4.33% have notifications recommending three product names.

To extract the names of the bought items, we use and compare the word frequency approach (that relies on the global frequency of the word in the English language) with two baselines: template-based and named entity-based. In the word frequency approach, we first extract the named entities from the text (using python NLTK) and then rely on the Python *wordfreq* library[11] to retrieve the frequency of a word in the English language. The three words that have the least frequent values are considered as the product name. We limit the number of words to three because the product or the service names are typically expected to be short (Robertson, 1989).

This approach is compared to a basic named entity implementation. We use python NLTK to implement the NLP pipeline described earlier (see section 4.3.2) and to extract the named entities. Similar to what we do in the previous approach, we limit the number of words forming the product name to three. These three words are randomly chosen from the list of words retrieved as named entities (unlike the previous one that picks the globally least frequent words). We compute the accuracy based on the average of 10 runs that randomly selects the three words.

In addition to the previous two approaches, we also use a template-based approach suggested by Li et al., (2018b) to extract what a person buys. This approach is the only one in the literature that we found targeting the text of smartphones' notifications. The template-based approach clusters notifications into templates according to the text similarity. As per Li et al., (2018b), each cluster is expected to have notifications with a similar structure, and hence the variable part of the template structure is considered as the feature's name. To implement this approach, we use Mean-Shift clustering of the python scikit-learn package. The bandwidth parameter of the Mean-Shift algorithm controls the number of notifications per cluster. A larger value may produce a cluster that contains all notifications generated by a specific app, whereas a smaller one may generate clusters with one notification each. Therefore, we have instantiated the implementation with various bandwidth values (1.20, 1.00, 0.90, and 0.80), and we stopped when the accuracy values of both actual_buying and marketing_only notifications get below 0.50. Figure 4.7 shows how the number of actual_buying notifications varies per cluster as the bandwidth value changes. For instance, when the bandwidth

[11] https://pypi.org/project/wordfreq/

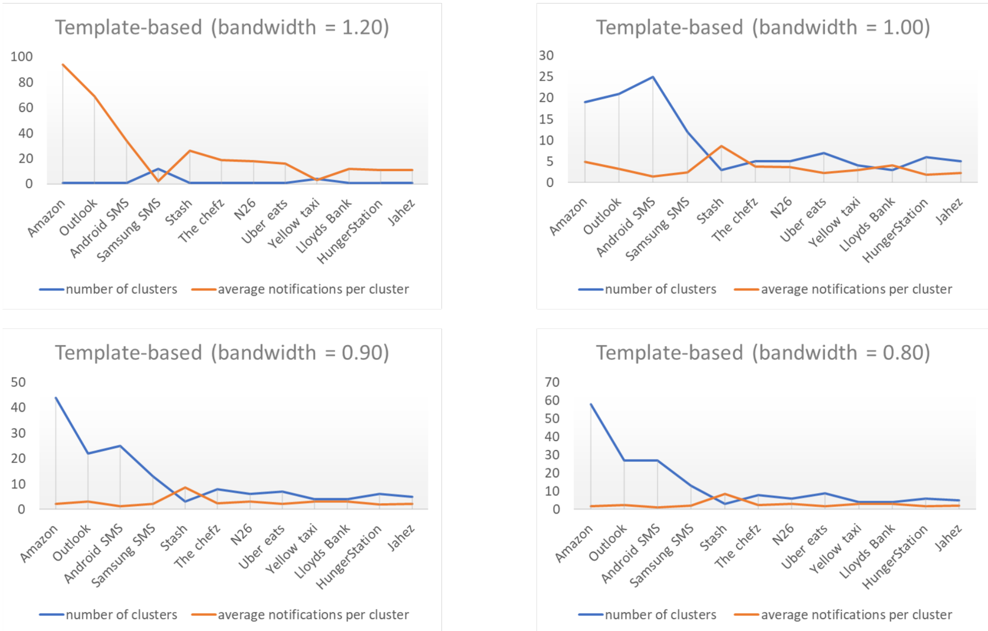


Figure 4.7: The impact of changing bandwidths on both the number of cluster per app and the average number of notifications per cluster. The average number is computed by dividing the total notifications per cluster on the number of clusters per app.

is set to 1.20, all 94 notifications of Amazon are assigned to the same cluster. In contrast, setting the bandwidth to 0.80 produces 58 Amazon's clusters with a mean of 1.62 notification in each one.

Based on these three approaches, we calculate the accuracy score as the number of correct results divided by the total number of notifications in each class. Since not all the notifications of actual_buying class contain a product name (see the first example in Table 4.2), each one of the three approaches may return one of the following:

- An empty text that correctly describes the absence of a product name in the notification entry,
- An empty text that fails to capture a product name exists in the notification entry,
- A three-words feature name that correctly matches a product name,
- Or, a three-word text that does not match the correct product name.

The accuracy scores show the ability of each approach to correctly (i) extract at least one of the product names found in the notifications texts and (ii) discover notifications not containing product names.

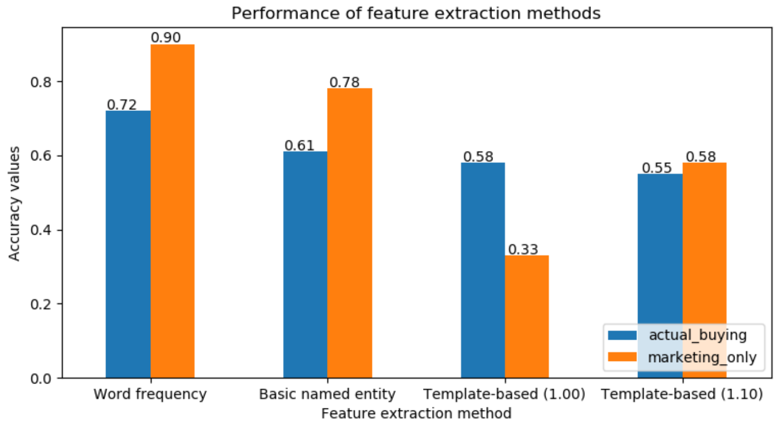


Figure 4.8: Accuracy values for the basic named-entities-, word frequency- and template-based approaches. The manually extracted features are used as the ground truth to calculate the values.

Accordingly, we have evaluated the three approaches and reported the results in Figure 4.8. The word frequency approach substantially outperforms the other two approaches in both classes of buying_related notifications (0.72 for actual_buying and 0.90 for marketing_only). The average accuracy values of randomly selecting three named entities as the product name for the actual_buying and marketing_only classes are 0.61 and 0.78, respectively. The accuracy score of the actual_buying class under the basic named entity approach is slightly better than its corresponding values under the template-based approach (0.58 if bandwidth = 1.00 and 0.55 at a bandwidth of 1.10). The accuracy values for both the basic named entities and the word frequency approach are better than the template-based approach. However, the performance of extracting the product names from marketing_only notifications using the template-based class is the worst among the three approaches (0.33 and 0.58 at the bandwidths of 1.00 and 1.10 respectively).

4.5 Discussion

The results of selecting notifications relevant to buying (Figure 4.5) show that machine learning algorithms perform better than the knowledge-based approach. The precision scores of the knowledge-based approach are substantially worse than their corresponding values of the machine learning algorithms except for the actual_buying class. When

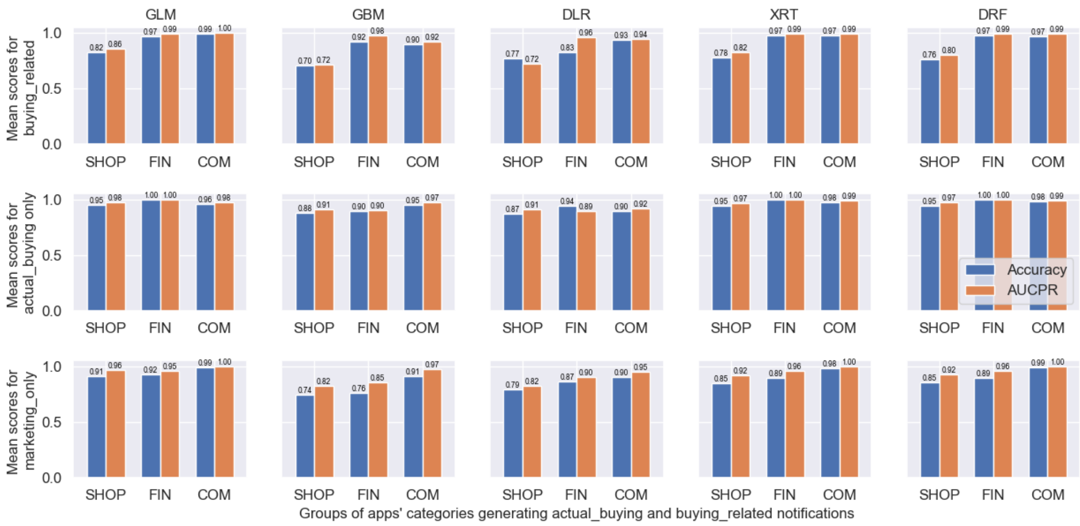


Figure 4.9: The accuracy and Area Under the Precision-Recall Curve (AUPRC) for each apps' category producing notifications of buying behaviour. The top three categories included in this figure are: communication (COM), finance (FIN) and shopping (SHOP).

all notifications from a specific category are identified as either buying_related or marketing_only, we include many messages that do not include the brand name. Examples of those messages are: "Your package will be delivered tomorrow" and "Check out all our offers before they end.". Also, the improvement in the buying_related scores is mainly driven by the high accuracy achieved under the actual_buying class (buying_related includes both actual_buying and marketing_only, see section 4.3.2 for details). However, improving the recall score of actual_buying class under the knowledge-based approach is visible by adding more keywords that are expected to appear in the notifications of buying behaviour.

Figure 4.6 shows that the average accuracy scores of all categorised undersampling subsets are below or equal to the ones corresponding to them in the uncategorised undersampling subsets. Therefore, we have further investigated the categorised undersampling. In Figure 4.9, we looked at how algorithms perform per each one of the top three categories that issue buying related notifications. Namely, "Finance", "Communication", and "Shopping"). We found out that predicting actual_buying entries of the "Communication" apps using GLM is consistent with what we have reported previously (see section 4.4.3.1). However, the accuracy and AUCPR scores of predicting marketing_only entries – generated by shopping apps using GLM – have dropped to 0.87 and 0.82 respectively. Similarly, both scores for marketing_only notifications of

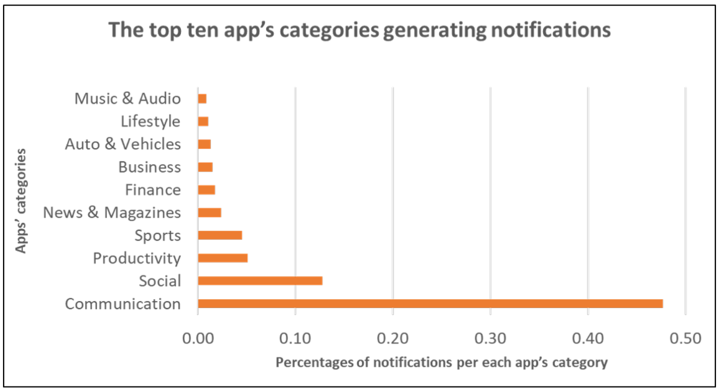


Figure 4.10: The percentages of the top ten apps' categories based on the total number of notifications in the dataset.

shopping apps are negatively impacted across all the other algorithms, and the smallest recorded values were by the GBM algorithm (0.72 for accuracy and 0.70 for AUCPR). This suggests that predicting notifications of buying behaviour from the same apps issuing them is not as clear to spot as recognising them from other apps' notifications.

Moreover, marketing_only entries are expected to have text shared with non_buying entries generated by the same category (since both's goal is to attract the user toward shopping). So, phrases such as "shop now" or "check out your ... offer" can be found in notifications that contain brand or service names (i.e. marketing_only). Also, these phrases may exist in notifications generated by shopping apps but do not contain a specific brand (e.g. "Check out all our offers before they end"). In contrast to that, actual_buying notifications are limited by the messages that are needed to be conveyed (i.e. order confirmation, shipping notice or delivery notes). Hence, they are expected to have more specific keywords (e.g. "your item" or "you paid") and consistent structure. In fact, the importance of these keywords and others can be measured and derived when applying the classification task on the actual_buying entries. Subsequently, additional keywords can be used to improve the performance of the knowledge-based approach by including them.

With respect to the machine learning-based approach, we can improve the algorithms' performance per apps' categories by collecting more data from each category. Additional techniques that deal with imbalanced datasets can also be used and compared against the undersampling used in this paper. Nonetheless, it is noteworthy that although extracting notifications of buying behaviour produces an imbalanced dataset, other behaviours have the potential of also producing a similar type of dataset. For instance, Figure 4.10 shows the percentages of the notifications of the top ten categories to the entire entries dataset. If we want to identify the behaviours that correspond to each one of those categories, it is expected that we will end up with an imbalanced dataset for most of them. Therefore, selecting the suitable technique to mitigate the problem of this type of dataset should be based on the characteristics of each case. Undersampling can be the right choice when the minority class represents the target behaviour. However, when the majority class is the target, oversampling or other techniques can be better choices.

For the feature extracting, we show that the approach based on the word frequency outperforms the other alternatives. However, the word frequency approach may be negatively impacted by the existence of famous brand names such as “Apple”, which may negatively impact the accuracy score. According to Python’s wordfreq library, the Zipf frequency for the word *package* is 4.45, whereas *Apple’s* Zipf value is 4.69. So, a notification that says something like “Your package with Apple device ...” may not retrieve “Apple” as the product name. Also, notifications that include named entities other than the product or the service name can lead to an increased number of false positives. Examples of those would be identifying the word “shipment” included in a notification text as a product name.

With respect to the template-based approach, one of the reasons behind the low accuracy value could be the actual_buying notifications that are clustered together but differ slightly. For instance, the following two notifications were classified under the same cluster:

- “You paid ** GB for Apple iPhone Unlocked Smartphone device”
- “Your Apple iPhone Unlocked Smartphone device will be delivered tomorrow”

. Although these notifications are clustered together, the variable parts can not lead to the correct identification of the product or the service name. This is because the produced template would be “`$feature1$` Apple iPhone Unlocked Smartphone device `$feature2$`” and therefore, neither feature would carry what has been bought. However, a dataset with a larger number of buying_related notifications may mitigate the impact of this issue. Also, as the notification text gets longer (e.g. order receipts emails), extracting what a person buys can become harder, especially under the template-based approach. This is mainly due to the increased number of both the variables expected in each template and the unique templates (each sender has its own template). Mitigating the latter can be done by adding the sender as a criterion when generating templates derived from email notifications.

However, we notice that extracting product names from marketing_only entries has a better performance for the three approaches (except when the bandwidth = 1.00) when compared to the actual_buying class. One reason behind that could be the fact that notifications under marketing_only category are expected to have a product or a service name. Therefore, the potential of incorrectly extracting a product name from an entry that does not actually include one is reduced (every notification has a brand name). In contrast, we may predict a three-word text as the product name from an actual_buying entry that does not include one. In this case, we have an additional way that can lead to the production of incorrect predictions.

To examine this latter point, we have further analysed the feature extraction results and found out that only 19.23% of actual_buying entries with no products were detected correctly by the word frequency approach (i.e. the approach correctly return empty text). This number is the same for the basic named entity approach since it produces the named entities similarly but differs in the selection of three-word text. The template-based approach has the best performance of detecting entries that do not include a product name (34.62%). The reason behind that could be the absence of variable parts in those notifications if clustered together. For instance, a cluster containing 10 notifications that say "your item will be delivered tomorrow" will not contain a variable part; hence, the returned three-word text will be empty.

Lastly, when recognising what a person buys (the feature extraction step), we noticed that multiple notifications could represent a single buying transaction. For instance, when a person buys a product X from Amazon, the following three notifications are typically received. (1) "Your package with product X has been dispatched", (2) "Your package with product X will be delivered tomorrow", and (3) "Your package with product X is out for delivery". Although from a feature extraction perspective, the product X is what the person buys regardless of the number of notifications. However, behavioural analysis requires a more careful look at these various notifications. Ignoring that may mislead the analysis. For instance, if we rely on the number of times a product is bought to derive personal interest, not addressing the same transaction's notifications may lead to incorrect conclusions.

4.6 Limitation

In this work, we had to get rid of more than 50% of the collected notifications as they contain empty texts or invalid characters. As shown in Figure 4.3, more than $50,000$ notifications were removed from three apps only (Snapchat, WhatsApp and Gmail). This limitation can be addressed by looking at each app separately and investigate how the notification text is stored. Also, the impact can be mitigated through the inclusion of the notification's title as an alternative that may lead to the context of the text (the analysis of the title is beyond the scope of this work).

The feature extraction results (section 4.4.3.2) are based on finding at least one product name. We limit our feature extraction on finding at least one item because most notifications contain only one product or service name. More in-depth analysis can target the extraction of more than a product name from the notification text.

Although we have conducted the analysis of notification filtering and feature extraction based on the entire dataset, the proposed framework is expected to be implemented on an individual basis. Our decision to use the entire dataset for the analysis stems from two reasons. First is the relatively small number of entries that we will get if the dataset is analysed on a per-participant basis. Second, the fact that text of buying_related notifications is not user-specific. Regardless of the user, the app will generate a similar text if a different user buys the same item. This observation is expected to be valid as long as the same app issues those notifications (Li et al., 2018b). What is user-specific, however, is the analysis that can be derived from the extracted feature. Examples include the products that a specific person likes and how often a person buys his/her favourite items.

Behavioural analysis that is based on the extracted features is not presented in this work due to the lack of ground truth data. For instance, P8 regularly buys his movie tickets in the late evening (in one instance around 10:00 PM and in three others around midnight). However, confirming and evaluating similar observations require additional ground truth about the investigated behavioural aspects. A more in-depth analysis of a specific behavioural aspect (e.g. buying habits and favourite brands) can be the subject of future studies. However, the framework proposed in this paper can be applied for notification-based digital phenotyping. The analysis can be related to buying behaviour or other behaviour captured by smartphone's notifications.

Lastly, we acknowledge that the selection of three words as the product name may cause the inclusion of unnecessary and unrelated words as part of the product name. This is especially true when the product name is formed of only one or two words.

Also, different ordering of the selected words may produce different results when additional semantic is added to the product name. Therefore, future work may look at the different methods of selecting the words compromising the product names and compare the impact on understanding buying behaviour.

4.7 Conclusion

In this study, we present a novel approach to recognising behavioural features from the smartphone's notifications. Specifically, we propose a framework that details the general steps needed to filter out notifications based on specific behavioural requirements and extract features accordingly. This could be particularly valuable for understanding personal preferences, whether related to buying or other behaviour such as reading and communication. Understanding these behavioural preferences are the basis for building personal recommendations.

Using this framework, we provide a use case for recognising behavioural features from buying-related notifications. By experimenting with knowledge-based and machine learning methods, we show how the latter provides better results in filtering out notifications irrelevant to buying behaviour. The data also shows that notifications of a specific behaviour are expected to produce an imbalanced dataset. The class representing the target behaviour can either be the minority or the majority class. Buying behaviour is an example of the former, whereas the class representing communication behaviour is expected to be the majority. For the feature extraction, the approach that relies on the least word frequency substantially outperforms other alternatives. These extracted features can be enriched using external data sources. More experimentation on those data sources' accuracy and their in-depth behavioural analysis values are planned for future work.

As notifications text can reveal high sensitive data such as financial data, this framework's applicability should be considered within such constraints. Also, limitations such as the inability to collect text from specific apps can be mitigated through the use of additional data (e.g. the title or the category of the issuing app). The informative value of these alternatives compared to the text is yet to be investigated.

Chapter 5

Interest Recognition from Smartphone-Derived Mobility Behaviour

The previous chapters indicated how events of the three behaviours targeted by this thesis are extracted. The next step is to understand and detect personal interests using the extracted events. We do so in this chapter and propose our Motivation-based Interest Recognition (MIR) approach to detect personal interests using events of mobility behaviour. We built our MIR method on the knowledge of human motivation (reviewed in Chapter 2) and the events extracted based on what has been described in Chapter 3. We analysed a secondary dataset for an initial exploration of interest-related features such as the time it takes to detect personal interests, the criteria for picking the top N items from a list of ordered interests, and the relationships between possible factors that influence interests. We benefited from the initial findings in the planning and execution of a three-month study that provided a summative evaluation of our method. Details about the MIR methods, the used datasets and the results are reported in this chapter.

The main content of this chapter is a paper authored by: *Ahmed Ibrahim, Sarah Clinch and Simon Harper*. The title of the paper is: ***Recognising Intrinsic Motivation using Smartphone Trajectories***. The paper is published in International Journal of Human-Computer Studies, Sep 2021, Volume: 153; and is made available under the CC-BY-NC-ND 4.0, license :https://creativecommons.org/licenses/by-nc-nd/4.0/. ISSN: 1071-5819. DOI: doi.org/10.1016/j.ijhcs.2021.102650. URL: https://www.sciencedirect.com/science/article/pii/S1071581921000689.

For this thesis, we edited some formatting styles, such as the sizes of some tables for consistency and readability reasons.

Author contribution

Ahmed Ibrahim designed the study, carried out the data collection, analysed and synthesised the results and wrote the paper. Sarah Clinch and Simon Harper provided continuous feedback throughout all the stages of the study, offered advice and discussion and contributed vital edits to the paper's writing.

Copyright

Abstract

Human behaviours that are motivated by and indicative of personal interests can be utilised to personalise behavioural recommendations used to promote health and well-being. Behavioural and psychological studies show that (1) personal interests are demonstrated differently in individuals' daily activities; and (2) drawbacks of self-reporting methods, such as forgetfulness and providing socially accepted answers rather than actual ones, may negatively impact the reliability and validity of the recognition process. To address these two challenges, we propose an adaptive approach that infers personal interests from continuously- and passively-sensed smartphones location data. We evaluate our approach based on two longitudinal datasets gathered by human participants going about their normal daily activities. Our results indicate that our approach successfully identifies interests consistent with those reported by participants, matching or outperforming alternative approaches. We also see high inter-personal variation, suggesting a future role for personalisation in our approach.

5.1 Introduction

Motivation is important for almost every aspect of human behaviour – every human action is shaped by, and indicative of some aspects of motivation. These motivations may be driven by external rewards or obligations (*extrinsic motivation*) or by personal interests and curiosity (*intrinsic motivation*) (Ryan and Deci, 2017). Identifying motivation in this latter case, in particular recognising an individual's personal interests, has value not only in understanding the individual, but also in shaping their future behaviour.

Assessing interests has been the concern of multiple inventories across different psychological sub-disciplines and applications (e.g. Ryan, 2018; Amabile et al., 1994; Tyler-Wood et al., 2010). However, inventories of this form require individuals to respond to specific prompts (e.g. "I enjoyed doing this activity very much") despite the fact that in most cases, intrinsic interests are demonstrated in an individual's daily activity.

A close neighbour, to this desire to derive intrinsic interests from individuals' behaviour, can be seen in many of our everyday interactions with technology. *Recommender systems* gather data about an individual's behaviours as part of a process intended to predict what an individual might like (or dislike) in the future (Zhang et al., 2019; Raza and Ding, 2019). Such systems are heavily used in a variety of technology platforms including e-commerce (Lu et al., 2014; Schafer et al., 2001; Zhou et al., 2018) and digital media consumption (Beam, 2014; Gomez-Uribe and Hunt, 2016). However, recommender systems typically have two features that differentiate them from the problem at hand: (i) data capture relates to a highly constrained set of behaviours taking place on a specified platform (e.g. all interactions with the website `Amazon.com`); and (ii) prior behaviours are used to make future predictions rather than to understand the underlying motivations that led to those behaviour. Therefore, while most recommender systems focus on figuring out the behaviours and making predictions based on those behaviours; we want to go one step further and go from the behaviours to the underlying motivation. Then given that motivation, future behavioural recommendations can be personalised.

In so doing, we propose to leverage individuals' smartphones for continuous, unobtrusive data collection that reflects behaviours undertaken in daily life. In particular, this chapter focuses on location behaviours as described by readings of smartphone positioning (GPS) captured at frequent and regular intervals. Behavioural events, such as visiting a cafe place or going to a movie theatre, are then extracted from these raw GPS data. Next, we identify specific measures that can be used to operationalise properties of two psychological models of human motivation: Self-Determination Theory (Ryan and Deci, 2017) and Maslow's Hierarchy of Needs (Maslow, 1943). We apply those measures on the behavioural events extracted from GPS traces to understand the

Table 5.1: Acronyms frequently used in the paper published in International Journal of Human-Computer Studies.

Acronym	***Description***
IMB	Intrinsically Motivated Behaviour
IMI	Intrinsic Motivation Inventory questionnaire
MIR	Motivation-based Interest Recognition[1]
SDT	Self-Determination Theory

underlying motivations and rank the prior behaviours accordingly. For example, if the extracted event is shopping, the rating produced from applying the measures of motivation properties would indicate how much a person is motivated by shopping. Higher ratings imply more internalised actions and therefore a better chance of the person being intrinsically motivated.

This distinction of internally motivated behaviours from externally motivated ones such as obligations is essential for applications of personalised behavioural change. Such applications aim to promote health and well-being by targeting actions that are motivated by and indicative of personal interests (i.e. intrinsically motivated). This is in contrast to other applications (e.g. e-commerce) where both personal interest and obligations would be considered as long as they serve the underlying goal (e.g. increasing revenues).

Our Motivation-based Interest Recognition (MIR[2]) model, summarised in Table 5.2, is derived based on a formative dataset (~2.8 million datapoints collected over one year from seven participants) and then evaluated through a summative evaluation using data from a further seven participants (~0.2 million datapoints collected over three months). Although two different populations have been used for formative and summative evaluations, our results suggest that the MIR approach can be used to detect interest for both groups of participants – despite the potential differences in the type of interests and how they are realised by each group.

To summarise, our contributions are threefold:

- **An approach for computationally modelling motivational properties**. We identify a set of behavioural measurements that reflect aspects of previously-articulated models of human motivation. These measurements have the potential

[1] Our approach.
[2] For a list of acronyms used throughout this paper, see Table 5.1.

Table 5.2: Motivation properties used in our approach, together with their corresponding behavioural measures. Integration of these measures to a single MIR value is discussed in Section 5.4.3.

Property	Description	Measurements	Theory	Details
Needs	Categorisation of internal needs	Needs level	Maslow	Section 5.4.1
Competence	Perceived ability to perform	Intensity	SDT	Section 5.4.2.1
Autonomy	Voluntarily performance of actions	Sustainability	SDT	Section 5.4.2.2
Novelty	Exploring of new behaviour	Recency	SDT	Section 5.4.2.3

to be captured unobtrusively over an extended time period using smartphone-based passive sensing.

- **An algorithm for Motivation-based Interest Recognition (MIR)** that aggregates the identified behavioural measures into a ranked set of Intrinsically Motivated Behaviours (IMBs) (i.e. where the top-ranked item is the one for which the user has the strongest interest).

- **An evaluation based on two distinct real world datasets** that validates ranked IMBs against participant ground truth and two alternative ranking approaches.

Our results provide a strong indication that our approach produces ranked interests that align closely with those elicited from participants through self-reports. Over the two evaluations we achieve higher precision and at-least-comparable recall when compared to alternative approaches.

5.2 Background and related work

5.2.1 Human Motivation

Theories of human motivation attempt to describe why humans do what they do (Ryan and Deci, 2000; McClelland, 1987; Weiner, 1992). Biological approaches focus on physiological state and processes (Cofer and Appley, 1964; Petri and Govern, 2013), and examples include Yerkes-Dodson (Yerkes and Dodson, 1908), drive reduction (Hull, 1943, 1952) and operant-conditioning (Skinner, 1953; Cooper et al., 1987). For the purposes of understanding and leveraging individual differences, however, these may be considered overly reductive (Strombach et al., 2016; Eccles and Wigfield, 2002).

Psychological mappings from human behaviour to motivation typically take one of two approaches. *Static approaches* use a fairly rigid classification to match behaviour to underlying physiological or psychological needs. Examples include Maslow's (1943) hierarchy and Murray's (1938) need theory. By contrast, *Dynamic approaches* quantify motivation based on the subjective impression of a participant toward a performed behaviour; factors such as contexts and rewards may impact the participant's attitude toward an activity (Fogg, 2012; Ryan and Deci, 2017). Examples include Self-Determination Theory (Ryan and Deci, 2017) and Fogg's (2012) motivational waves.

For the purposes of this work, we draw on two dominant psychological explanations: Maslow's Hierarchy of Needs (Maslow, 1943) and Self-Determination Theory (Ryan and Deci, 2017). In particular, this work focuses on determining an individual's intrinsic motivation (Ryan and Deci, 2000) – activities that inherently bring satisfaction to an individual (commonly referred to as interests; Renninger and Hidi, 2016). Key concepts extracted from these theories, and used in this work, can be found in Table 5.2.

5.2.1.1 Maslow's Hierarchy

Maslow (1943) discusses motivation in terms of five needs, the lowest of which must be fulfilled before the next comes into focus. These five needs (presented lowest to highest) are as follows: physiological, safety, belongingness, self-esteem and self-actualisation need. *Physiological* needs relate to survival at an individual and species level, such as food, drink, sleep and sex. *Safety* needs include stability, security and protection from fear. *Belongingness* needs are driven by the desire for interpersonal relationships, and feelings such as love, friendship and acceptance. *Self-esteem* needs drive our desire for respect, dignity and independence. Finally, *self-actualisation* needs drive our ambition and desire for personal growth (Maslow, 1943; McLeod, 2007).

Critics of Maslow (e.g. Neher, 1991) suggest that experiencing these needs in the proposed order is contrary to evidence in the real world. For instance, lack of security in some communities - due to war, civil unrest or similar - does not prevent their inhabitants from developing social ties and pursue the fulfilment of belongingness needs. Despite this, Maslow's hierarchy continues to be highly influential (including, for example, in recent attempts to understand technology: Houghton et al. 2020; Kang and Jung 2014). In light of the identified limitation, in this chapter, we focus on mapping behaviours to needs as nominal categories rather than concerning ourselves with an ordinal progression between the levels.

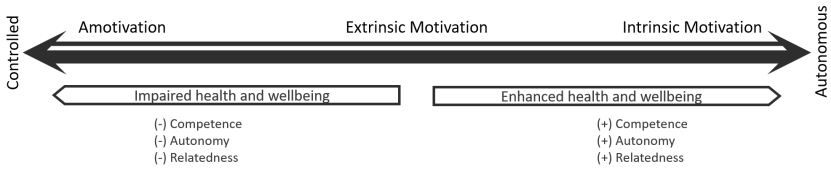


Figure 5.1: The motivation continuum proposed in SDT runs from amotivation through to intrinsic motivation.

5.2.1.2 Self-Determination Theory

SDT (Ryan and Deci, 2017) is one of a number of contemporary theories that build on the distinction between intrinsic and extrinsic motivation. However, in contrast to others, SDT treats these concepts not as a dichotomy, but instead as a continuum that ranges from amotivation, through a set of extrinsic motivation states, to a fully internalised intrinsic motivation (Figure 5.1).

SDT identifies competence, autonomy and relatedness as the three basic psychological needs that differentiate and represent motivation states: the need for competence (also called self-efficacy), the need for autonomy, and the need for social relatedness (Ryan and Deci, 2017):

1. *The need for competence* refers to one's belief in their ability to perform (Bandura, 1971). Self-perceived success, satisfaction or efficiency when engaging in tasks helps to satisfy the need for competence (Ryan and Deci, 2017; White, 1959, 1963).

2. *The need for autonomy* relates to the extent to which a person controls a behaviour (Ryan and Deci, 2017), and self-regulates goals and the process of attaining them (Schunk et al., 2008).

3. *The need for relatedness* is concerned with feelings of connection with others and is an essential driver for social behaviour (Ryan and Deci, 2017).

In addition to these three basic needs, proponents of SDT have noted the importance of novelty in motivating individuals to pursue and possibly change personal interests (Ryan and Deci, 2000, 2017; González-Cutre et al., 2016; Silvia, 2007). This has in turn led some to propose *the need for novelty* as a futher innate psychological need (González-Cutre et al., 2016).

As a popular and "living" theory (Vansteenkiste et al., 2010), SDT has been applied in a wide variety of domains, including many related to technology. For example, SDT can be used to guide the interface design of mobile apps (e.g. Zuckerman and Gal-Oz 2014; Rooksby et al. 2015), encourage the use of health apps (e.g. Saksono et al. 2020), or propose behavioural intervention (e.g. Gustafson et al. 2014). Unlike these applications, we employ SDT to classify behaviours, that are passively sensed, as either extrinsically or intrinsically motivated.

5.2.2 Passive Sensing, Digital Phenotyping and Mobility Traces

Smartphones have transformed personal data collection. The majority of the population near-continuously carries a device featuring multiple specialised sensors such as: accelerometer, gyroscope, ambient light sensor, proximity sensing (e.g. Bluetooth, NFC); these are in addition to the microphone and camera that are considered critical to the devices' functionality (Lane et al., 2010). These sensors allow for passive data collection (i.e. without intervention from a user) that can be considered highly indicative of the user's environment and behavior. This data has a multitude of applications (Khan et al., 2013), including extensive use for health and well-being (Cornet and Holden, 2018). It is in this context that Jain et al. (2015) coined the term *digital phenotyping* to refer to the process of using an individual's interaction with digital technologies to derive indicative markers for human health and wellbeing (Figure 5.2).

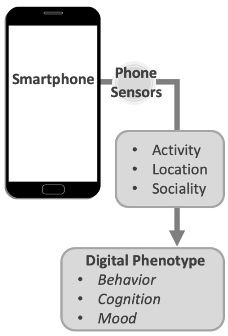


Figure 5.2: The use of passive smartphone sensors for digital phenotyping. Adapted from Insel (2017).

In this paper, we focus specifically on the use of passively-sensed location, one of the most popular approaches applied in health and well-being to date (Cornet and Holden, 2018). Most commonly, this will take the form of a set of time-location pairs captured using the GPS sensor on a smartphone. From this spatiotemporal data, a series of features are typically extracted. For example, "distance travelled" and "time spent at home" are mobility features that have been used as indicators of fatigue and social anxiety (Vega-Hernandez, 2019; Andrienko et al., 2013), while "fraction of the day spent stationary" and "maximum distance from home" have been used as indicators for relapse behaviours in schizophrenia (Barnett et al., 2018).

A core part of the feature extraction process is the *segmentation* of a trace into a set of *episodes* and *trajectories*:

- An **episode** is the abstraction over a set of data points that represent a stationary or motion period based on some specific criteria. The term *stay-point* can be used to refer to a stationary episode.
- A **trajectory** is a sequence of episodes that represent an individual's movement through geographic space over a period of time.

Once extracted, episodes can be semantically *annotated* using external data sources (Nogueira et al., 2018). For example, services such as Foursquare[3] can allow a collection of proximate GPS readings (a stay-point) to recognised as a public park, residential area, or even a specific shop or restaurant.

5.2.3 Recommender Systems

Recommender systems are software applications that aim to make predictions about the items or behaviours that might be of interest to a specific individual (Zhang et al., 2019; Raza and Ding, 2019). Recommender systems draw on a variety of computer science techniques including data mining, user modelling and machine learning. However, the fundamental concept centres on the use of existing indicators of a target user's interest (e.g. ratings, purchases, frequency of interaction), together with knowledge about all of the items that could be recommended (e.g. object classifications, features, other users' ratings or interactions) to derive a rating for each item in a set of possible recommendations. Based on this rating, the top n items can be presented to the target user (or user group) (Adomavicius and Tuzhilin, 2005). There are two principal approaches to recommender systems. *Content-based* systems suggest items based on the user's profile, while *collaborative* systems considers information of similar users to predict the recommended items (Adomavicius and Tuzhilin, 2005; Villegas et al., 2018).

Although they are most prominently used in e-commerce (Zhou et al., 2018; Schafer

[3] https://foursquare.com

et al., 2001; Lu et al., 2014) and digital media consumption (Beam, 2014; Gomez-Uribe and Hunt, 2016), recommender systems can be used to encourage broader behavioural change. For example, Rabbi et al.'s (2015) smartphone application, *MyBehavior*, used a combination of passive sensing and manual logging to record physical activity and food intake. Statistical machine learning was then used to identify and recommend high calorie loss behaviours similar to the user's existing behaviors, resulting in a statistically significant increase in physical activity and corresponding decrease in calorie intake compared to a control condition.

Context-aware recommender systems employ contextual information to improve the predictability of recommendations (Adomavicius and Tuzhilin, 2015). For behavioural recommender systems, such as that described above, this contextual information is especially critical to disambiguate users' activities. Villegas and Müller (2010) classified contexts into five categories: individual, location, time, activity, and social/relational. Contexts, such as time of the day or the activity performed at specific locations, are used as indicators that facilitate the extraction of behaviours or locations of interest; the first step toward building high quality recommendations (Raza and Ding, 2019). Of greatest relevance to this paper, are those recommender systems that leverage location and time.

Existing recommender systems that make use of location typically consider the frequency of visitation (Musto et al., 2018; Yu and Chen, 2015; Li et al., 2015a), duration of visit, (Boytsov et al., 2012) or combination of both (Do and Gatica-Perez, 2014; Li et al., 2008). Annotations of the form described in Section 5.2.2 can be used to help generalise from one specific location to similar places (Do and Gatica-Perez, 2014; Karatzoglou et al., 2018). When combined with time, a recommender system can also deliver recommendations based on the both frequency and duration of visits at a specified time of day (Yuan et al., 2013; Li et al., 2017a; Natal et al., 2019), or based on recency of visit (Li et al., 2015b; Logesh and Subramaniyaswamy, 2017).

Despite their complexity, the majority of recommender systems (including context-based and behavioural systems) fail to consider the underlying motivation that led to the user behaviours used as indicators. Whilst many of these indicators may reflect personal interests, others will be the result of obligations (e.g. buying a gift for others, visiting a workplace). In this chapter, we set out to identify Intrinsically Motivated Behaviours (IMBs) – behaviours that reflect intrinsic motivation (Ryan and Deci, 2000). In so doing, we aim to enable future recommendations that align with the individuals' interests and values and can facilitate sustained behaviour change (Michie et al., 2011).

5.3 Inferring Motivation from Smartphone Sensor Data

In this chapter, we set out to measure aspects of human motivation, continuously and unobtrusively, by using the location data that is captured by smartphones as individuals go about their daily activities. We focus specifically on trying to identify behaviours that are internally motivated. To the best of our knowledge, this is the first approach to capturing personal interests that integrates psychoanalytical and data-driven techniques to rank behaviour based on motivation properties.

We identify three specific subprocesses necessary to enable our approach: *indicator identification*, *item identification*, and *interest determination*. At their most abstract, these subprocesses align with those of conventional recommender systems, but each requires substantial rethinking to address the unique challenges that come from both (a) operating over human mobility behaviour, and (b) attempting to extract intrinsic motivation from a set of intrinsically- and extrinsically-motivated behaviours.

Indicator identification A set of common indicators (e.g. item features, user interaction, and user ratings) have widespread applicability in conventional recommender systems (Zhang et al., 2019; Raza and Ding, 2019) but are of limited relevance to the motivation needs described in Section 5.2.1). Further, unlike other behavioural indicators, concepts such as competence, autonomy, and novelty are not immediately captured in smartphone-observable measurements. Therefore, for each property, we need to identify indicators that are behaviourally observable and, at the same time, strongly indicative of that underlying property.

Item identification Unlike traditional computer systems that deal with 'interest' (i.e. recommender systems), our approach must operate over a diverse set of behaviours rather than predefined 'items' (e.g. products or online movies/TV). These behavioural items are encoded into streams of low-level data (for location this is typically a GPS trace). Therefore, an initial step is the identification of individual behaviour 'items' from this low-level data. Within the context of this paper, stay-points are the 'items' that we extract from the raw GPS data and on which our approach would operate.

Interest determination Traditional recommender systems seek only to identify patterns of behaviour, and then to use this information to rank potential future behaviours. By contrast, we seek to differentiate within those patterns of behaviour, to *specifically identify behaviours that reflect underlying personal interest* (i.e. intrinsic motivation). Therefore, we apply indicators, that are amenable to computations, on the items extracted from digital phenotyping. An indicator is amenable to computation if there exists a behavioural measure that strongly correlates with the underlying cognitive indicator (i.e. the motivation property). As a result, we encode smartphone GPS data as cognitive traits within our behavioural model and then aggregate these traits to rank and personalise behaviour.

5.3.1 Indicator Identification

As noted in the previous section, we first concern ourselves with the identification of appropriate indicators and items, before utilising these to identify the likely underlying interests.

To identify indicators, we conducted a review of the literature to identify behavioural measures for human motivation. Specifically, we sought out any measure that was considered to be strongly indicative (i.e. the majority of existing literature supports an indicative relationship) of any of the underlying cognitive needs identified in Section 5.2.1.2, namely *competence*, *autonomy* and *novelty*. We then consider each of these in terms of their applicability to location data, and determine the final set of indicators accordingly. Note that we deliberately exclude *relatedness* due its inherently social nature – although smartphone sensor data may be indicative of social behaviours, it is unlikely that location data alone will provide a meaningful measure of this construct.

Table 5.2 lists the identified behavioural measures together with the motivation concepts to which they are associated. High levels of perceived competence are associated with completing an action more often (Ryan and Deci, 2017; Rabbi et al., 2015), and the time required to complete the action (Nicholls, 1984; Fishbach and Hofmann, 2015). Thus, we combine frequency and duration into a measure of intensity (Wolf and Hopko, 2008). Voluntary performance of action (i.e. autonomy) is associated with action sustainability over time (Pelletier et al., 2001; Seguin et al., 1999), whereas the propensity to seek out novelty is manifested in the exploration of new behaviours and gaining new interests (i.e. recency) (Ryan and Deci, 2013, 2017).

5.3.2 Item Identification

Next, we use event segmentation to decompose location traces into distinct visitation events (i.e. item identification). Specifically, we apply sequential processing of GPS data to extract stay-points based on predefined time t and distance d thresholds (appropriate values for t and d may be implementation specific and are explored in Section

5.6.1). The resulting stay-points are places where an individual has lingered for a period greater than t minutes within a boundary of d meters. Note that for the purposes of this work, we do not consider the intermediate periods (i.e. periods of motion) [4].

Once extracted, stay-points are semantically labelled using a location annotation service. Google Places, Foursquare, and other platforms each provide reverse geocoding services that can be used to annotate stay-points (e.g. converting a location reading of `51.5194° N, 0.1270° W` to `The British Museum`) and then appropriately categorising that as a `history museum`. In this work we use Foursquare as our location annotation service due to its richer categorisation (950 categories) than other services (e.g. Google provides 96 categories).

5.3.3 Interest Determination

For interest determination, we build on our prior contrast of Maslow (1943) and SDT (Ryan and Deci, 2017) as static and dynamic approaches respectively (Section 5.2.1), and develop a solution that engages both mechanisms in parallel. A *static modelling* step (Section 5.4.1) builds on Maslow's heirachy, whilst our *dynamic modelling* (Section 5.4.2) is based on SDT. Computed values from both dynamic and static measures are combined to form the motivation score (referred to as the MIR score).

5.4 Modelling

Interest determination is achieved by generating a measurement, the MIR score, for each participant behaviour over a given time period. This measure can be used to classify actions as intrinsically or extrinsically motivated, based on the SDT continuum depicted in Figure 5.1. Specifically, higher ratings imply more internalised actions and therefore a better chance of the person being intrinsically motivated. In this section, we provide a detailed description of how the MIR score is assembled through use of two modelling steps; both steps are grounded in the literature, targeting patterns of behaviour that are considered positive indicators of IMB. Measures emerging from the modelling steps are combined to form the overall MIR score (Section 5.4.3). This integration process may be generalised (i.e. all elements are weighted identically for each

[4]These move-points could represent either a necessary transition between two stay-points, or one of a limited set of (typically fitness or sporting) activities that are intrinsically motivated (e.g. walking, running). In the case of the former, there are more appropriate sensors that could detect these specific activities (e.g. accelerometer: Wannenburg and Malekian, 2017).

individual), or adapted to reflect individual differences (i.e. weighting is determined based on differences in an individual's need satisfaction: Deci and Ryan, 2000).

5.4.1 Static Modelling

Our static modelling is a participant-independent mapping between location episodes and motivation based on Maslow's hierarchy. We assign an ordinal value to each of Maslow's five need levels, based on the degree to which the level reflects intrinsic motivation. Behaviours that serve needs at the top of the hierarchy are considered to display more intrinsic and autonomous behaviour (McClelland, 1987) and are scored more highly than those at the base. Specifically, we assign a value of one for physiological needs and two for safety needs. The remaining three levels relate to belongingness, esteem and self-actualisation needs; since all three represent intrinsic and self-determined actions (Barbuto Jr and Scholl, 1998; McClelland, 1987), we score each equally with respect to intrinsic motivation. We assign behaviours associated with these levels a value of four (i.e. a mean derived from the values three, four and five if one simply incremented the score as one progresses up the hierarchy).

To identify behaviours that correspond to these levels, we use a popular location annotation provider (Foursquare) to identify different semantic classes of location (e.g. art gallery, casino, mosque). Then, guided by motivation-based taxonomies of behaviour (Tinsley and Eldredge, 1995; Barbuto Jr and Scholl, 1998; Talevich et al., 2017), we match semantic classes to a corresponding generalised category (e.g. art, games/gambling, spiritual) and locate the category within Maslow's hierarchy. Each location stay-point captured in a participant's dataset can then be mapped from specific location to semantic class (as determined by Foursquare)[5], and then from semantic class to category. The resulting category determines hierarchy level and associated intrinsic motivation score. The output of this process is reflected in Table 5.3 which lists the categories identified, together with examples of semantic classes within those categories, and the intrinsic motivation score as derived from Maslow's hierarchy.

The described classification should accurately distinguish between intrinsically- and extrinsically-motivated activities in the majority of cases (Tinsley and Eldredge, 1995; Barbuto Jr and Scholl, 1998; Talevich et al., 2017). However, visits to locations that would score highly for intrinsic motivation may occur for reasons other than personal interest (e.g. to accompany a friend). This limitation is addressed through the

[5] We select the first place in the list of categories as produced and ranked by the Foursquare API.

Table 5.3: Categories used in static modelling. Each category is allocated to a level within Maslow's hierarchy and scored appropriately. Values of one and two denote physiological and safety needs respectively; all other levels suggest some intrinsic motivation and are scored with the value four.

Motivation Category	**Exemplar Locations**	**Maslow Score**
Art	art gallery, public arts	Intrinsic - 4
Culture and History	museum, historic site	Intrinsic - 4
Dance	dance studio, salsa club	Intrinsic - 4
Dining out	fine dining, family style dining	Physiological - 1
Education	School, university	Safety - 2
Entertainment	aquarium, circus	Intrinsic - 4
Games and gambling	casino, gaming cafe	Intrinsic - 4
Health and Fitness	gym, weight loss center	Safety - 2
Movies	movie theater, indie theater	Intrinsic - 4
Music	concert hall, jazz club	Intrinsic - 4
Outdoors and Recreation	national park, mountain	Intrinsic - 4
Profession	bank, day care	Safety - 2
Reading	bookstore, library	Intrinsic - 4
Residence	home, hotel	Physiological - 1
Shopping	shopping mall, auto dealership.	Safety - 2
Socializing and Drinking	pub, lounge	Intrinsic - 4
Spiritual	church, mosque	Safety - 2
Sport	hockey arena, stadium	Intrinsic - 4
Travel and Transport	airport, train station	Physiological - 1

addition of dynamic modelling.

5.4.2 Dynamic Modelling

Dynamic modelling is used to determine the degree to which an exhibited behaviour aligns with concepts associated with intrinsic motivation (competence, autonomy and novelty) and to capture naturally-occurring variation in motivation (Fogg, 2012). Unlike static modelling, these measurements are instantiated from the participants' data and hence actual values vary according to the behaviour exhibited by each individual. Each participant's data is subdivided into week-long analysis windows (Monday-Sunday), reflecting this naturally-emerging determinant of human behaviour (i.e. most people exhibit consistent patterns of behaviour on weekdays vs. weekends: Cho et al., 2011; Sarker et al., 2019).

Within each analysis window, we then identify instances of the indicators summarised in Table 5.2. To overcome limitations associated with any one motivation property, and/or it's associated indicator, our dynamic modelling integrates all of the identified properties and indicators. Existing literature shows each to be positively correlated with others, and with intrinsic motivation, at an individual and aggregated level (Ryan and Deci, 2017). However, we do not consider this indicator set to be exhaustive; future work may demonstrate the utility of alternative or additional measures. For example, an experiment may use the diversity and flexibility in times of a behaviour as a measure of autonomy (Ryan and Deci, 2017) along with sustainability. Similarly, we envisage future integration of a measure for relatedness (omitted from this work due to its focus on location behaviour), through indicators derived from proximity or other sensors.

5.4.2.1 Competence

Feelings of competence arise from self-perceived achievement rather than the activity itself (Oudeyer and Kaplan, 2009) and are predicted by both frequency and duration of engagement in the activity (Fishbach and Hofmann, 2015; Nicholls, 1984; Wolf and Hopko, 2008; Rabbi et al., 2015; Ryan and Deci, 2017). To integrate duration and frequency into a measure of *intensity*, we count, for each stay-point, the total number of visits and multiply that by the average duration per day.

Formally, if f_w is the weekly frequency of a behaviour x, and m_w is the weekly average duration for the same behaviour, then the intensity of x up to the current last week of an ongoing study t is computed as:

$$Intensity_x = \sum_{w=1}^{t} f_w * m_w \tag{5.1}$$

5.4.2.2 Autonomy

Autonomous actions are characterised as being volitional and self-initiated and are demonstrated through repeated and continuing engagement in the activity Ryan and Deci (2017). We express this as the *sustainability* of behaviours. To generate a measure of *sustainability*, we assign a Boolean value that indicates whether a specific behaviour occurs within the analysis window. We then count the number of subsequent windows (i.e. weeks) in which a behaviour is observed and divide that by the total number of subsequent analysis windows. The closer is the result to 1, the more sustained the behaviour.

Formally, if n_{xw} denotes the existence of a behaviour x in week w, and d_{xw} represents the current number of weeks in a study, then the sustainability score of a behaviour x is computed as:

$$Sustainabilty_x = \sum_{w=1}^{t} n_{xw}/d_{xw} \tag{5.2}$$

where t represents the current last week of an ongoing study. Next, we see the impact of novelty on shaping the current period.

5.4.2.3 Novelty

Individuals' interests and behaviours change periodically, previous behaviours and interests are abandoned and new ones adopted (Ryan and Deci, 2017; Sarker et al., 2019). To ensure that IMBs reflect current interests, we therefore consider novelty through a measure of the *recency* of observed behaviours. Thus, we segment the entire study duration into periods (these periods are distinct from the previously mentioned week-long analysis windows and may vary in duration either statically or dynamically[6]; in this work we adopt a static approach based on findings reported in Zhao et al., 2013 and Srinivasan et al., 2014). We then weight each behaviour according to its occurrence

[6]See Srinivasan et al. (2014) and Sarker et al. (2019) for examples of static and dynamic recency thresholds respectively.

within each period, with the most recent period accumulating the highest value; values associated with prior periods gradually decline as their distance from the present period increases. This gradual retrospective degradation ensures currency whilst also accounting for the fact that individuals typically revisit previous behaviours rather than entirely abandoning them (Zhao et al., 2013). This is in direct contrast with prior work that has considered only the most recent period (e.g. Sarker et al., 2019)

5.4.3 Integration

We bring together our static and dynamic models to compute an overall score for any given behaviour. The resulting *MIR Score* indicates the degree to which the specified behaviour is intrinsically motivated (for a given individual). The MIR Score takes into account the needs level derived from Maslow's hierarchy (Section 5.4.1) and the three SDT indicators (Section 5.4.2).

To the best of our knowledge, prior computational models of these properties have not been realised. Thus, we suggest a straightforward approach that uses linear summation as the aggregation function and weights each feature equally (i.e. coefficients = 1 for all properties). The resulting formalisation operates over a set of periods p, such that i ranges from 1 to p where 1 designates the most recent period, and p represents the outmoded interval.

$$\text{MIR score} = \sum_{i=1}^{p} \frac{1}{i} \left(needs + intensity + sustainability \right) \tag{5.3}$$

where *needs* is determined as specified in Table 5.3, and $intensity$ and $sustainability$ are determined in accordance with equations (5.1) and (5.2) respectively, before being adjusted to account for *recency*. Thus, both $intensity$ and $sustainability$ are calculated for each period before being weighted and summed as described in Section 5.4.2.3.

Note that Equation 5.3 is just one potential combination of the submeasures to form the MIR Score, and assumes a single generalised approach is applicable to all. SDT literature suggests that differences in individual's need satisfaction may shape their pursuit of a given need (Deci and Ryan, 2000), such that weighting sub-measures identically for different individuals may not be appropriate. Thus, we suggest a role for *personalisation* in the integration phase (as shown in Figure 5.3).

Finally, the resulting MIR score for each behaviour can be used to assemble a

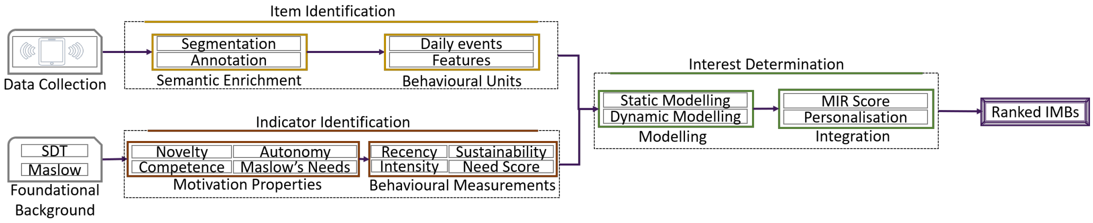


Figure 5.3: The overall process of extracting and ranking IMBs from phenotyped GPS raw data.

ranked list of IMBs as follows:

$$Behaviour \times \text{MIR Score} \rightarrow Ratings. \tag{5.4}$$

where the ratings for each user is predicted as a function of the motivation properties and the performed actions. For example, if the behaviour is related to football, then the rating score would indicate the degree to which a football-related behaviour is intrinsically motivated (for a given individual).

5.5 Implementation: The MIR algorithm

Figure 5.3 and Algorithm 1 provide an end-to-end view of our approach, whereby raw GPS data is ultimately transformed into a ranked list of IMBs. Passively-sensed location data is semantically enriched to produce a set of stay-points that correspond to life events such as dining out, and shopping. Extracted stay-points are used as input to the aforementioned models of motivation properties (determined based on the indicator identification) to rank the behavioural events based on the underlying motivation properties. As a result, we produce a ranked list of motivated behaviour that represents the behaviour's place on the proposed motivation continuum.

Data input takes the form of raw GPS data collected continuously and longitudinally (`locationData`), which is then subdivided into periods (`updateRecencyPeriod`). The most recent data form period $P = 1$, which is preceded by $P = 2...P = N$. Data is then collected into stay-points and semantically annotated (`enrichData`), resulting in a collection of behavioural items (`behaviourList`). Together, these steps form the item identification process (Section 5.3.2).

Our indicators are embedded in a pair of motivation modelling steps (`applyMotivationModels`), which are applied to behavioural items (`computeMotivationScore`)

Algorithm 1: Algorithmic implementation of the proposed approach that uses a combination of static and dynamic models of motivation to derive a set of ranked interest behaviours (IMBs).

```
input: GPS locationData of size n
output: Ranked motivation list of size m

Algorithm GenerateRankedMotivations (locationData)
    periods = updateRecencyPeriods (threshold)
    behaviourList = enrichData (locationData)
    totalRatings = < period:periodRating >
    foreach period p in periods do
        periodBehaviour = getPeriodBehaviour (behaviourList, p)
        periodBehaviourScore = computeMotivationScore
          (periodBehaviour, periodLength (p))
        totalRatings = updateTotalRating (p, periodBehaviourScore)
    end
    IMBs = aggregateTotalRatings (totalRatings)
    return IMBs
Func computeMotivationScore (periodBehaviour, periodLength)
    periodBehaviourScore = < behaviour:score >
    foreach behaviour b in periodBehaviour do
        foreach week w in periodLength do
            applyMotivationModels (b)
        end
        score = aggregateWeeklyScores (b)
        periodBehaviourScore = updateBehaviourScore (b, score)
    end
    return periodBehaviourScore
```

in a given period (`getPeriodBehaviour`). Specifically, (`computeMotivationScore`) computes the motivation properties, *intensity* and *sustainability*, by iterating over each behaviour in the period and applying both the static and dynamic models (`applyMotivationModels`). Weekly *intensity* and *sustainability* scores for each behaviour observed in the period are stored as a key-value (behaviour-score) pair (`updateBehaviourScore`). For each period, the associated (`periodBehaviourScore`) pairs are themselves stored as key-value (period-score) pair (`updateTotalRating`).

Finally, (`aggregateTotalRatings`) applies the recency value on the motivation properties of each period to get the MIR score. Accordingly, each behaviour is valued based on its period such that recent, sustained, intense and intrinsically needed behaviours are at the top of our ranking.

5.6 Experimentation

Given a lack of existing mobile sensing approaches that use motivation to inform interest recognition, we first conduct a formative evaluation that allows us to examine our modelling and design decisions, contrasting the output with that from popular interest measurement approaches; this evaluation also helps to suggest appropriate measures for tuneable parameters. Guided by insights from this first study, we then conduct a further (summative) evaluation. Our evaluation includes longitudinal mobile data from fourteen participants going about their normal daily activities (seven participants in each phase). Procedures for both studies were reviewed and approved by the Department of Computer Science Ethics Committee at The University of Manchester (Reference: 2019-7817-12726).

5.6.1 Formative evaluation

For our formative evaluation, we conduct secondary analysis on a dataset previously captured from seven adults using the AWARE mobile sensing framework (Ferreira et al., 2015); this dataset was collected from seven older adults with Parkinson's disease (age $53-72$, median age 65, mean 65.71), over a period of one year. The participants had mild motor symptoms and none to slight involuntary movements (a.k.a. Dyskinesias) (Vega-Hernandez, 2019). The dataset, collected over one year, contains over 1000 million passively-sensed datapoints from two Android and five iOS devices

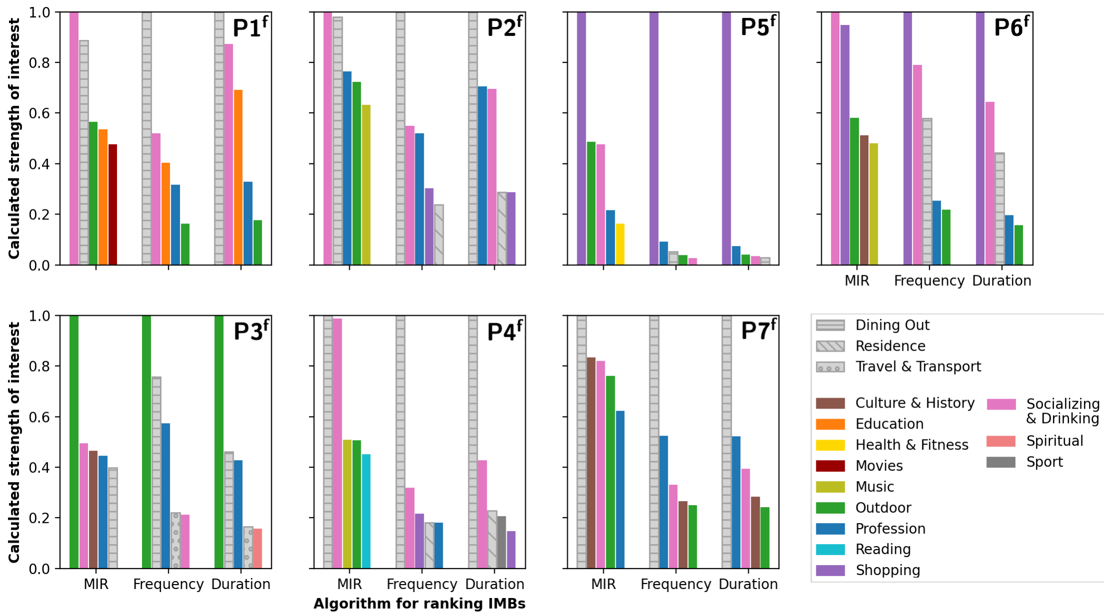


Figure 5.4: IMBs for each participant as determined by our MIR algorithm and by frequency-based and duration-based approaches for comparison. Participants in the bottom row ($P3^f$, $P4^f$, $P7^f$) are those for whom we also have ground truth.

(Vega-Hernandez, 2019). Of these datapoints, 2.8 million are measures of location (tuples comprised of latitude, longitude and timestamp). GPS sampling took place at an interval of one minute, although factors like battery outage and signal loss reduce the number of collected samples.

In addition to the location data, we use extracts from interviews with three of the seven participants[7], in which they describe their strongest interests and the frequency with which they engage in activities relating to those interests (Table 5.4). The "Reported interests" column represents the ground truth that are collected from the participants. The annotated interests (represented by the "Annotation" column in Table 5.4) are the results of mapping the ground truth data to our proposed taxonomy. This data was unfortunately not available for all seven participants, but should provide an indicative groundtruth nonetheless.

We use the algorithm proposed in Li et al. (2008) to extract stay-points: data points are processed sequentially, with stay-points determined in accordance with predefined time and distance thresholds. Guided by Boytsov et al. (2012), we set our time threshold at 15 minutes. We fix our distance threshold at 100 meters (Solomon et al., 2018),

[7]The researcher of the original study was able to collect interests from three participants only as they were the only ones who remained by the time of this work.

Table 5.4: Annotated interests for three participants in the formative dataset. Note that activities that would largely take place at home are not annotated, as these would not be detected by our approach (location data for home is discarded).

Participant	Reported interests	Annotation
P3[f]	Reading	Reading
	Gardening	-
	Walking	Outdoor and Recreation
	Music	Music
	Television	-
P4[f]	Gardening	-
	Spending time with family and friends	Socialising and drinking
	Singing	Music
	Dancing	Dance
	Walking	Outdoor and Recreation
	Reading	Reading
P7[f]	Bird watching	Outdoor and Recreation
	Playing bridge	Outdoor and Recreation
	Helping out local charity	Profession
	Learning Microsoft Access	-
	Walking	Outdoor and Recreation
	Reading novels	Reading
	Dancing	Dance
	Visiting National Trust properties	Culture and History
	Drinking craft beers	Socialising and drinking
	Learning bass guitar	Music
	Listen to music	Music

meaning that GPS readings within a 100-meter circumference are considered to be the same stay-point.

Staypoints and interview responses were then annotated using a common category set based on a semantic grouping of Foursquare annotations. We then follow the approach described in Sections 5.3 and 5.5 to determine a ranked set of IMBs. Guided by studies and findings on mobility behaviour (Srinivasan et al., 2014; Zhao et al., 2013; Song et al., 2010), we use three months as the recency threshold that segments the study duration into periods and apply equation (5.3) accordingly.

Finally, since IMBs conducted inside the home are impossible to identify using location alone, we seek to exclude the participant's likely residence from our analysis. Early exploration of the data using semantic annotation for this purpose (i.e. looking for places in a residences category) was often inconclusive. Thus, we instead identify the single location in which the participant spent the most time in any given week and exclude it as the participant's likely place of residence. Excluding locations on a weekly basis should account for temporary accommodation such as vacation or business trips.

5.6.1.1 Results

Figure 5.4 shows an ordered (left-to-right) set of the top five ranked IMBs for each of our seven participants (the three bottom-most participants are those for which we also have interview ground truth data). We also plot comparison results from two alternative algorithms: one frequency-based (used in e.g. Musto et al., 2018 and Liu et al., 2016) and one duration-based (used in e.g. Lim et al., 2015 and Gaonkar et al., 2018). MIR selects the same top-rated interest as both other algorithms for just over half of the participants ($P3^{f}$, $P4^{f}$, $P5^{f}$, $P7^{f}$). In other cases, the top two IMBs are transposed compared to one ($P2^{f}$) or both ($P1^{f}$, $P6^{f}$) other algorithms. However agreement on the top three IMBs for MIR and at least one of the two comparison algorithms (ignoring ordering) occurs only in one case (the two comparison algorithms themselves agree on the top three in five of the seven cases). Similarly, agreement on the top five IMBs for MIR and at least one of the two comparison algorithms (ignoring ordering) occurs only in one (different) case whereas the two comparison algorithms themselves agree on the top five in four of the seven cases.

Our original intention was to consider only the top three IMBs, but having calculated strength of interest using each method it was evident that clear differences in the strengths of consecutively ranked IMBs naturally emerged at different points for each participant. We therefore additionally establish a cut-off point based on the largest difference between consecutive behaviours (Figure 5.5), and in subsequent analysis compare the validity of this dynamic N (shown in Table 5.5) with fixed values. In many case the dynamic N value is equal to one, suggesting that the participant's behaviour reflects a single interest much more strongly than any others (this is especially true when considering frequency or duration alone). Whilst in some applications, identifying the strongest intrinsic motivator would be sufficient, we suggest that in many cases a broader understanding of IMB would be beneficial. Thus, for each case where the dynamic N would be 1, we also identify the next largest dynamic N using the approach previously described (Table 5.5). The resultant value, $N^{\mathrm{Dyn}'}$, is used in subsequent analyses.

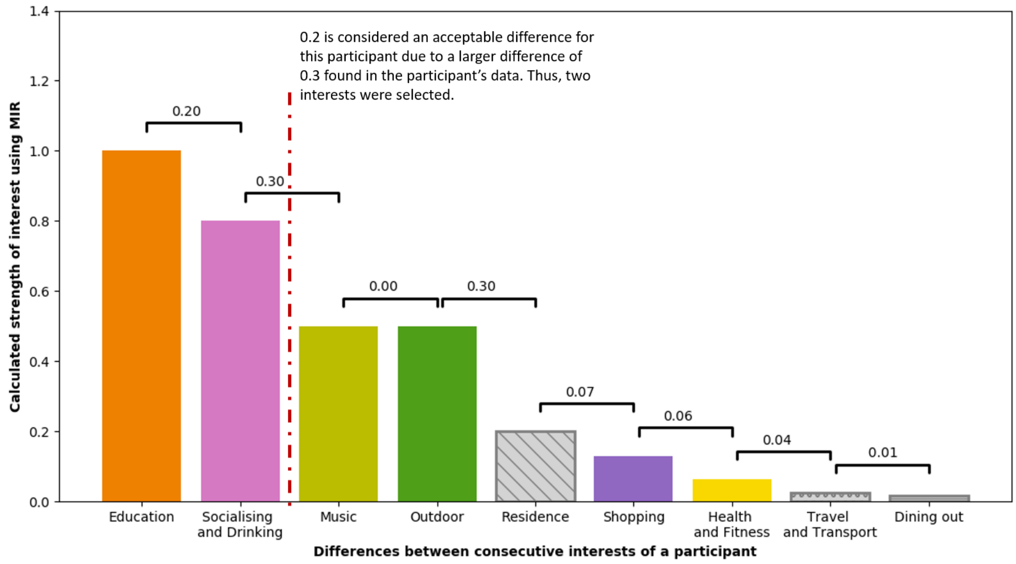


Figure 5.5: An example for the determination of the top interests based on the largest difference.

Examining the top-ranked interests themselves (using $N^{\mathrm{Dyn}'}$), we find that over half (min: 50%, max: 75% mean: 59%) match a Maslow needs level of four (intrinsic). However, a significant minority correspond to Maslow's physiological (mean: 24%) and safety needs (mean: 17%). Whilst one could interpret this as indicating that a location-based approach may struggle to filter out extrinsically-motivated behaviours, we note that in reality many of these kinds of activity can be intrinsically motivated (e.g. choosing to engage in shopping or fitness activities because they are enjoyable or align with personal values rather than out of necessity).

From the raw data reported in Figure 5.4 and Table 5.4, we calculate precision and recall for each of the three algorithms. For instance, the top three interests retrieved by the MIR approach when N = 3 for $\mathrm{P3^f}$ are: "Outdoor and Recreation"; "Socialising and Drinking"; and "Culture and History". Since only "Outdoor and Recreation" is retrieved and $\mathrm{P3^f}$ reported "Reading"; "Outdoor and Recreation"; and "Music" as interests (see Table 5.4), then both precision and recall are equal to $1/3$. Similarly, we compute the values for other participants and the results are summarised in Table 5.6. MIR outperforms both alternative algorithms, with a mean precision of $0.56 - 0.75$ and mean recall of $0.40 - 0.69$. The dynamic N performs best on recall, but this is not true for either frequency or duration. By contrast, use of a dynamic N produces the best precision values for both frequency and duration, but not for MIR. MIR's precision is

Table 5.5: Number of top IMBs determined dynamically (N^{Dyn}) for each participant ($\mathrm{P1^f}$-$\mathrm{P7^f}$) based on largest difference between interest score for consecutively ranked behaviours. Where N^{Dyn} is 1, the column N^{Alt} indicates the top N IMBs based on the next largest difference between ranked behaviours. Shaded values are those used as $N^{\mathrm{Dyn}\prime}$ in subsequent analyses.

	MIR		**Frequency**		**Duration**	
	$\mathrm{N^{Dyn}}$	$\mathrm{N^{Alt}}$	$\mathrm{N^{Dyn}}$	$\mathrm{N^{Alt}}$	$\mathrm{N^{Dyn}}$	$\mathrm{N^{Alt}}$
P1$^{\mathbf{f}}$	2	–	1	4	3	–
P2$^{\mathbf{f}}$	2	–	1	3	3	–
P3$^{\mathbf{f}}$	1	13	3	–	1	3
P4$^{\mathbf{f}}$	2	–	1	2	1	2
P5$^{\mathbf{f}}$	1	3	1	2	1	2
P6$^{\mathbf{f}}$	2	–	3	–	1	3
P7$^{\mathbf{f}}$	8	–	1	2	1	2

heavily impacted by the very large alternate N used for $\mathrm{P3^f}$. Excluding $\mathrm{P3^f}$, the mean precision increases to 0.62, higher than both fixed N values.

To better understand the impact of study duration on our results, we consider both temporal stability of our identified IMBs. We consider a behaviour "stable" if its weekly MIR scores fluctuation stays within ± 0.05 for three consecutive weeks. This stability represents the time needed for our algorithm to stabilise (i.e. calibration time). Table 5.7 reports the mean number of weeks between an IMB's first appearance in the participant's data and the point at which that IMB begins to stablilise (based on $N^{\mathrm{Dyn}\prime}$), demonstrating that it takes between two and fourteen weeks for measurement of an IMB to stabilise. Overall, it takes less time for IMBs to reach stability with MIR than comparative algorithms (mean 7.26 compared to 8.01 and 8.19 for frequency- and duration-based algorithms respectively). However, this varies across the sample (e.g. for $\mathrm{P4^f}$ and $\mathrm{P7^f}$ IMBs stabilise slower with MIR than the other two algorithms).

Similarly, to help inform the sample size for a summative evaluation, we calculate the Pearson correlation values amongst the individual behavioural measurements that form our dynamic (intensity, sustainability) and static models: (needs) of motivation. A Fisher transformation was applied to average the correlation coefficients, and the results are given in Table 5.8. The two dynamic measures have a strong positive correlation with each other (0.88) but weak negative relationships with the static measure (-0.27 and -0.43 for intensity and sustainability respectively). We see stronger positive correlation between MIR and dynamic (SDT-derived) motivation measures (0.75

Table 5.6: Precision (P) and Recall (R) values for the MIR, Frequency-, and Duration-based algorithms. Values calculated based on each participant ground truths, together with the sample mean ($\bar{x}$).

	MIR						Frequency						Duration					
	N=3		N=5		$N^{Dyn'}$		N=3		N=5		$N^{Dyn'}$		N=3		N=5		$N^{Dyn'}$	
	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R
P3^f	0.33	0.33	0.20	0.33	0.23	1.00	0.33	0.33	0.20	0.33	0.33	0.33	0.33	0.33	0.20	0.33	0.33	0.33
P4^f	0.67	0.60	0.80	0.80	0.50	0.20	0.33	0.20	0.20	0.20	0.50	0.20	0.33	0.20	0.20	0.20	0.50	0.20
P7^f	0.67	0.29	0.80	0.57	0.75	0.86	0.67	0.29	0.80	0.57	0.50	0.14	0.67	0.29	0.80	0.57	0.50	0.14
$\bar{x}$	**0.56**	**0.40**	**0.60**	**0.57**	**0.49**	**0.69**	**0.44**	**0.27**	**0.40**	**0.37**	**0.44**	**0.23**	**0.44**	**0.27**	**0.40**	**0.37**	**0.44**	**0.23**

Table 5.7: Mean number of weeks taken for before IMBs enter a period of stability for each participant and algorithm (using $N^{\mathrm{Dyn}'}$ in all cases), together with overall mean ($\bar{x}$) and median ($\tilde{x}$).

Method	**P1^f**	**P2^f**	**P3^f**	**P4^f**	**P5^f**	**P6^f**	**P7^f**	$\bar{x}$	$\tilde{x}$
MIR	7.50	6.50	4.92	10.50	2.00	9.00	10.38	**7.26**	7.50
Frequency	10.25	8.33	8.00	4.00	7.00	14.00	4.50	**8.01**	8.00
Duration	9.33	12.00	9.67	3.50	8.00	10.33	4.50	**8.19**	9.33

and 0.66 for intensity and sustainability respectively), than with our static (Maslow-derived) measure (0.39). Interestingly, we see variation in the degree to which these correlation trends hold for individual participants – for example, for P4^f it is only intensity that correlates strongly with MIR, and for P5^f it is sustainability.

5.6.1.2 Initial insights

As reported in Section 5.6.1.1, both MIR and the two comparison algorithms agree on the top-ranked IMB in the majority (57%) of cases, but they agree on the top three

Table 5.8: Correlations amongst the measurements that form our dynamic (intensity, sustainability) and static models: (needs) of motivation (top); and between each measurement and the final MIR score (bottom).

	P1^f	**P2^f**	**P3^f**	**P4^f**	**P5^f**	**P6^f**	**P7^f**	$\bar{x}$
Intensity/Sustainability	0.91	0.75	0.71	0.64	1.00	0.85	0.80	**0.88**
Intensity/Needs	−0.21	−0.60	−0.03	−0.44	−0.04	−0.19	−0.29	**−0.27**
Needs/Sustainability	−0.20	−0.77	−0.56	−0.75	−0.02	−0.28	−0.12	**−0.43**
Intensity/MIR	0.71	0.74	0.87	0.67	0.83	0.75	0.59	**0.75**
Sustainability/MIR	0.73	0.59	0.52	0.38	0.84	0.69	0.71	**0.66**
Needs/MIR	0.51	−0.05	0.39	0.23	0.52	0.46	0.56	**0.39**

and top five IMBs in only one case. Given that MIR outperforms both alternative algorithms on precision and recall, we suggest that MIR is better able to identify postliminary IMBs than either alternative algorithm.

Our dynamic N values vary considerably (min 1, max 8; Table 5.5), and differ by algorithm. For example, for P1^{f} $N^{\text{Dyn}\prime}$ is 2, 1, and 3 for MIR, frequency- and duration-based methods respectively. The propensity for N^{Dyn} to be one (true for 12 of our 21 calculations) is reflective of the fact that for many of our participants there was a stand out primary IMB (as indicated both by N^{Dyn} and the much higher algorithmic agreement on this IMB). However, as previously noted, identifying a single IMB is unlikely to be sufficient for many applications.

Considering $N^{\text{Dyn}\prime}$, we still see differences in values by algorithm in almost all cases. However, the values themselves are now almost all between two and four, with just two outliers (P3^{f} MIR: 13, and P7^{f} MIR: 8). Of these two, we note that P7^{f} does report considerably more interests than other participants (11 reported interests mapping to 7 distinct annotation categories); this is not true for P3^{f}, whose 13 interests emerge only as a result of considering N^{Alt}. The overall consistency in $N^{\text{Dyn}\prime}$ suggests that whilst there are naturally occurring cut offs in the ranking of IMBs, in practice a fixed N of four would be equivalent in the majority of cases. However, since both cases with a higher $N^{\text{Dyn}\prime}$ occur amongst our participants with ground truth, we are able to see that this increased N does yield better recall in both cases. Precision is reduced, but only mimimally for P7^{f} (compared to N=5, we add two reported IMBs and one non-reported, potentially erroneous IMB). For P3^{f} recall reaches 100%, but at a significant cost for precision (100% recall is reached with an N of 9 using this algorithm). Nonetheless, there may still be application-specific needs that prompt further minimum and maximum bounds on a dynamic N.

Based on the stability measurements, we suggest that studies of IMB should use a minimum duration of around eight weeks (the largest mean reported in Table 5.7). Studies of longer duration are likely to be more informative (e.g. our highest time to stability is fourteen weeks) and will capture a richer picture of the natural changes in the pursuit of IMBs over time (including, e.g. seasonal change).

Combining an understanding of the correlations between behavioural measurements, with the knowledge that SDT proposes a linear aggregation between motivation properties (Ryan and Deci, 2017), we suggest that a linear regression can provide an indicative sample size for future experimentation. Thus, we use the linear regression of GPower[8] with parameters determined based on the initial correlations reported in Table 5.8. On that basis, we suggest that seven participants would be needed to evaluate MIR at 80% power and 95% confidence level. This aligns with the suggestion from Barnett et al. (2020) that ten participants are needed to get 84% power in a longitudinal study lasting for 90 days.

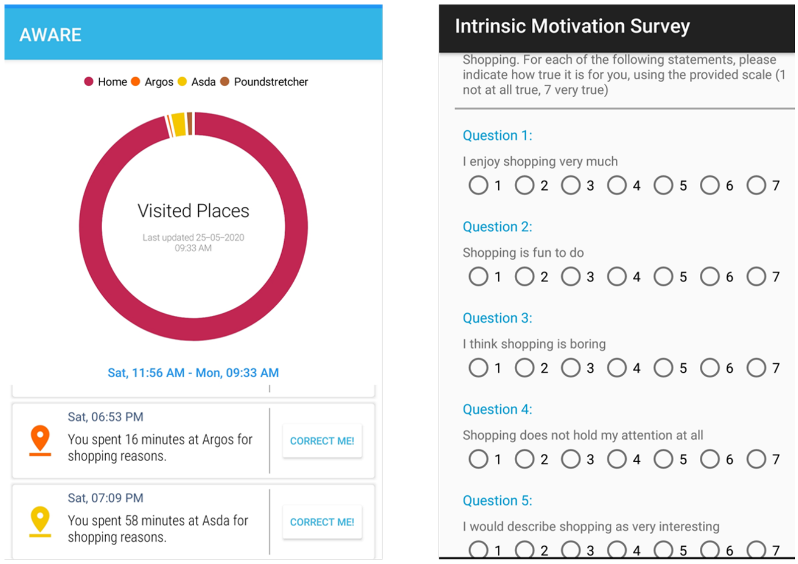
Finally, we note a general trend for strong positive correlations observed between MIR and both intensity and sustainability, suggesting that MIR is successfully reflecting these measures of intrinsic motivation. We see a weak positive correlation between MIR and needs. However, the correlation values meet the expectations stated earlier (Section 5.4.2) and show a strong correlation between the motivation properties quantified through those measurements and intrinsic motivation. Extremes and variations from this general trend for individual participants suggest that some participants (e.g. $P1^{f}$, $P5^{f}$-7^{f}) exhibit more behaviours that are considered to be more intrinsic by Maslow's hierarchy whilst others may be intrinsically motivated to engage in behaviours that could be interpreted as extrinsic when considering Maslow's need level alone ($P2^{f}$-4^{f}). This suggests that personalisation with regard to the weightings of different motivation properties may be valuable, and warrants further exploration in the summative evaluation.

5.6.2 Summative evaluation

Our formative evaluation suggests that MIR has potential to identify IMBs not captured by baseline approaches, and thus there is a need for further evaluation.

While the formative evaluation relied on secondary data analysis, in our summative evaluation we collect a novel dataset from new participants. We mirror the method used by Vega-Hernandez (2019), using AWARE (Ferreira et al., 2015) to collect location readings at one minute intervals. To maximise the probability of correctly identifying participant locations, and thus the derived motivators, we provide participants with a plugin to examine and correct their annotated locations (Figure 5.6a). Each time a visit to new place is detected, the participant receives a notification inviting them to confirm or correct the detected place. To support collection of ground truth, we add a further plugin within the AWARE app (Figure 5.6b). This plugin presents an adaptation of the Interest/Enjoyment subscale of the widely-used Intrinsic Motivation Inventory (IMI). IMI is a multidimensional scale developed by Ryan (2018) and is

[8] https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower.html



a) Places b) IMI

Figure 5.6: Interfaces for the two developed AWARE plugins: (a) the Places plugin allows participants to examine and correct their annotated locations; (b) the IMI plugin presents participants with a set of validated questions that can be used as ground truth.

widely used to assess intrinsic motivation associated with a given activity (Monteiro et al., 2015; Ryan, 2018). All data captured within the app was sent to a secure server at The University of Manchester.

5.6.2.1 Study Duration and Participants

Based on the study duration recommendations reported in Section 5.6.1.2, we select a study duration of three months. This exceeds the minimum period of eight weeks, and is approximately equal to the largest participant mean in Table 5.7. Also, the participants of the formative study may have reduced mobility as they had reasonably mild symptoms. Therefore, if they do have reduced mobility then the final parameters that we use for our summative study will be conservative. This is because the reduced amount of mobility means it is going to take longer to reach stability and thus we are going to overestimate the duration needed rather than underestimate it.

Similarly, we use the method described in Section 5.6.1.2 to determine an appropriate sample size. We recruit seven participants to achieve 80% power and 95% confidence level. We further note the previously noted arguments in support of smaller samples given an appropriately long study duration, shifting the focus towards data corpus that reflects real-world behaviour and spans over a long period of time. Upon conclusion, our study data consists of ~3.8 million, which is sufficient for meaningful analyses.

Participants were recruited using poster advertisements displayed in public areas of The University of Manchester and surrounding buildings, and on social media. A total of seven participants ultimately agreed to participate, all but one of whom were students at The University of Manchester (three undergraduates, three postgraduates). Participants were supplied with an information sheet prior to participation and had the opportunity to ask further questions prior to consent. Participants were rewarded for their participation with a total of £30 gift cards (approximate value 39 USD, 33 EUR) distributed across the study period.

Participants enrolled on the study over a staggered period based on when they chose to respond to recruitment advertisements, but each completed a full twelve weeks of data collection. The first participant ($P1^s$) began data collection on January 22, 2020, all but one had begun by February 10, 2020, and the final participant ($P7^s$) began on March 4, 2020. Unfortunately, due to COVID-19 restrictions enforced in United Kingdom in early 2020, all of our participants' mobility patterns significantly changed for at least some of their data collection period. From 23rd March, 2020, non-essential stores and services (e.g. gyms, cafés, bars, theatres) were closed, as were schools and other places of education. This mostly strongly impacts data from $P7^s$, but even $P1^s$ was only able to supply us with location data from 8.5 weeks of unrestricted mobility.

5.6.2.2 Method

All participants provided written consent before being guided to install the AWARE app on their personal mobile device. Participants were asked to keep the installed application running and to carry their phones as they normally do; they were advised that the application would automatically send their data to our backend server, but only when connected via WiFi, and that we anticipated no noticeable negative effects on battery life.

During the initial session, participants also engaged in a short audio-recorded interview in which they were asked a set of general questions designed to act as a supplementary source of ground truth. For example: what are the places they visit because they want to rather than they have to? What motivates them to practice their behaviours? How often do they practice them?

As noted in Section 5.6.2.1, participants were asked to run the AWARE app for three months, during which all data was synced to our backend server for offline analysis. This includes both the raw locations, post-correction annotations, and participants' IMI responses. Using the corrected locations, we exclude participants' place of residence using the method described in Section 5.6.1.

Seven IMI prompts (the complete Interest/Enjoyment subscale) were delivered monthly; each takes the form of a statement with an associated seven-point Likert-like scale (1=not very true, 7=very true). Statements were adapted from the original IMI (Ryan, 2018) in order to incorporate a specific IMB into the phrasing, e.g. from "I enjoyed doing this activity very much" to "I enjoy shopping very much" (Figure 5.6b)[9]. All seven statements reference the same IMB, which is selected randomly from the pool of top ranked N^{Dyn} IMBs for the previous month[10]. The resultant data is used as a ground truth for the target IMB.

At the end of the study, participants again participated in a short audio-recorded interview. Participants were asked about all specific IMBs recognised in their the dataset, and were additionally asked to identify personal interests that they felt had been omitted.

5.6.2.3 Results

In addition to participant interview and IMI ground truth, we again compare the output of our MIR algorithm with the baselines reported in Section 5.6.1. We begin by mirroring our formative evaluation, considering the final set of ranked IMBs derived from the full three months of collected data. MIR selects the same top-rated IMB as at least one other algorithms in all cases, agreeing with both other algorithms for just under half the participants ($P2^s$, $P4^s$, $P6^s$). Agreement on the top three IMBs for MIR and at least one of the two comparison algorithms (ignoring ordering) occurs for just over half the participants ($P1^s$, $P5^s$, $P6^s$, $P7^s$) as does agreement amongst the two comparison algorithms themselves ($P1^s$, $P2^s$, $P4^s$, $P5^s$). However, agreement on the top five IMBs for MIR and at least one comparison algorithm (ignoring ordering) occurs only in one case ($P6^s$), whereas the two comparison algorithms themselves agree on the top five in just over half of the seven cases ($P1^s$, $P2^s$, $P4^s$, $P6^s$). Overall, we see higher agreement

[9] The full question set is presented in Appendix C.5.

[10] As a slight variation on our formative analysis, top-ranked IMBs are selected based on participant corrected annotations. Over the course of our study, 6 participants correct a total of 124 locations (16.51% of all locations recorded). Although we discuss the impact of incorrect annotation on the produced results in Section 5.7, it is not the goal of this work to enhance the accuracy of external annotators.

between MIR and other algorithms in the summative study.

Selecting IMBs dynamically results in a N^{Dyn} of between 1 and 4 for MIR (mean: 2.00, median: 2.00, std: 1.07, iqr: $2.50 - 1.00 = 1.50$) with slightly lower N^{Dyn} values for the two alternate algorithms (frequency mean: 1.14, median: 1.00, std: 0.35, iqr: $1.00 - 1.00 = 0.00$; duration mean: 1.57, median: 1.00, std: 0.73, iqr: $2.00 - 1.00 = 1.00$). Note that in this case, we maintain the original N^{Dyn} to preserve only the strongest interests (although we again see a substantial number of cases in which $N^{\text{Dyn}} = 1$: MIR 3; frequency 6; duration 4). By so doing, we aim to constrain the maximum number of interests. The mean size of the difference used as a cut off for N^{Dyn} is 0.34 (MIR), 0.54 (frequency), 0.50 (duration). Overall intersection of dynamic ranked list of participant IMBs (N^{Dyn}) is 72.62% for MIR/frequency, 84.52% for MIR/duration, and 83.33% for frequency/duration.

Compared to the formative data set, we see a lower proportion of IMBs corresponding to a Maslow needs level of 4 (intrinsic): 44% (mean), and a much higher proportion of safety (mean 40%) needs. However, this considerably more variable across the sample when compared to the formative study.

Our summative study provides a much richer ground truth for the final set of IMBs, with an IMI score generated for each ranked IMB (N=1... N^{Dyn}). These scores can range from 1 (no interest/enjoyment) to 7 (very high interest/enjoyment) and are an aggregation of participants responses to the seven scale items listed in Appendix C.5. The resulting values are detailed in Table 5.9. If the IMI score falls below four (denoted with red shading in Table 5.9), then the behaviour is considered to be a false positive. Two such values occur, the final ranked interest for P1^{s}, and the only ranked interest for P7^{s}. In this latter case, collected data was extremely sparse (27 missing days; and mean daily GPS readings is 91.63 compared to 357.81 for other participants) due to mobility restrictions incurred as a result of COVID-19 (as noted in Section 5.6.2.1, only nineteen days, 23%, of this participant's collection period occurred outside of a COVID-19 lockdown). Participants also reported interests that they felt were not listed: P1^{s} added watching movies, P2^{s} added sports, and P4^{s} added music (others did not add to the recognised interests). All three participants noted that they had not practised these interests during the study period.

We determine the final ground truth where true interests include those items with an IMI score greater than or equal to four plus any identified as missing in the closing interviews. We also expand our top-ranked interest set to include any missed IMBs (this adds one interest to P1^s, P2^s and P4^s). Based on this, we calculate the precision and recall values reported in Table 5.10. MIR substantially outperforms both other algorithms on recall, however a Friedman test shows no significant difference ($Q(2) = 5.64$, $p = 0.06$). For precision, MIR is surpassed by the frequency-based approach, and a Friedman test again shows no significant difference between the algorithms ($Q(2) = 5.64$, $p = 0.06$), or precision ($Q(2) = 2.00$, $p = 0.37$). MIR's recall is consistent with results from the formative study (formative: 0.69, summative: 0.70), but both alternative algorithms perform more strongly this time around (formative: 0.23 in both cases, summative: 0.49 and 0.57 for frequency and duration respectively). All approaches perform considerably better on precision than in our formative study (values $0.81 - 0.86$ compared to $0.44 - 0.49$), although this is likely to be impacted both by our decision to use N^{Dyn} rather than $N^{\mathrm{Dyn}'}$.

Table 5.9: Participant IMI responses (from the final interview) for each of the interests identified by the MIR algorithm. False positives (i.e. IMI scores below 4) are shaded in red.

	Reported interests	MIR	IMI
P1^s	Education	1.00	5.29
	Dining out	0.87	6.00
	Socialising and Drinking	0.83	5.14
	Shopping	0.83	2.43
P2^s	Education	1.00	5.00
P3^s	Education	1.00	7.00
	Dining out	0.94	6.17
	Socialising and Drinking	0.90	4.83
P4^s	Education	1.00	5.83
	Socialising and Drinking	0.80	5.67
P5^s	Education	1.00	6.43
	Dining out	0.87	5.83
P6^s	Outdoor and Recreation	1.00	6.23
P7^s	Socialising and Drinking	1.00	2.43

Compared to the formative data set, we find that IMBs are quicker to enter a period of stability. Excluding P7^s (because the participant did not consider the recognised behaviours as interests), the mean weeks elapsed prior to stability is 5.92 for MIR, 3.92 for frequency, and 6.11 for duration (medians 6.13, 3.50 and 6.75 for MIR, frequency

Table 5.10: Precision (P) and Recall (R) values for the MIR, Frequency-based, and Duration-based algorithms based on the set of ranked interests (N^{Dyn}) generated by each *at the end of the summative data collection period*. Values are calculated based on IMI participant ground truths collected in the closing interview. Note that P7^{s} started two weeks before COVID-19 lockdown which making it impossible to determine IMB. Best and worst performers are highlighted with green and red shading respectively.

	MIR		**Frequency**		**Duration**	
	P	R	P	R	P	R
P1^{s}	0.75	0.75	1.00	0.25	0.67	0.50
P2^{s}	1.00	0.50	1.00	0.50	1.00	0.50
P3^{s}	1.00	1.00	1.00	0.33	1.00	0.67
P4^{s}	1.00	0.67	1.00	0.33	1.00	0.33
P5^{s}	1.00	1.00	1.00	1.00	1.00	1.00
P6^{s}	1.00	1.00	1.00	1.00	1.00	1.00
P7^{s}	0.00	0.00	0.00	0.00	0.00	0.00
$\bar{x}$	**0.82**	**0.70**	**0.86**	**0.49**	**0.81**	**0.57**

and duration respectively). However, in four cases, the recognised interests do not stabilize at all over the study period (i.e. fluctuation in strength of interest exceeds ±0.05 in all consecutive 3-week periods). Removing these behaviours from our calculations (P3^{s} dining out MIR and duration, P3^{s} MIR socialising and drinking, and P5^{s} dining out duration) reduces the means further (4.85 weeks MIR, 5.28 weeks; median 5.50 in both cases)[11].

We again calculate the correlation amongst the individual behavioural measurements that form our models with the overall MIR score. A Fisher transformation was applied to average the correlation coefficients, and the results are given in Table 5.11. Mirroring our formative results, we see strong positive correlation between MIR and dynamic (SDT-derived) motivation measures (0.67 and 0.73 for intensity and sustainability respectively), and moderate positive correlation with our static (Maslow-derived) measure (0.54). This difference, however, is smaller than the one seen in our formative data. Correlations from both formative and summative analysis suggest that a sample size of seven participants can achieve 80% power and 95% confidence level.

In addition to IMIs collected at the end of the study, participants each completed an IMI for one top-ranked interest each month. Availability of some participants was impacted by COVID-19, such that five monthly IMIs (23.81%) were not returned; P7^{s}

[11] In the previously reported means/medians, these interests were each assumed to stablilise immediately following the study period, i.e. we assign the final week as the stability value.

Table 5.11: Correlation between overall MIR scores and the measurements that form our dynamic (intensity, sustainability) and static models: (needs).

MIR Score/Feature	**P1^s**	**P2^s**	**P3^s**	**P4^s**	**P5^s**	**P6^s**	**P7^s**	**$\bar{x}$**
MIR/Intensity	0.63	0.54	0.62	0.63	0.79	0.89	0.42	**0.67**
MIR/Sustainability	0.62	0.45	0.47	0.23	0.80	0.93	0.95	**0.73**
MIR/Needs	0.32	0.53	0.48	0.59	0.20	0.73	0.76	**0.54**

returned zero IMIs, and two other participants also failed to return an IMI response in month two. Monthly IMI scores for participants improve as the quantity of collected data increases (month one: 5.31, month two: 5.78, month three: 5.93). However, a Kendall's tau-b correlation coefficient between the score improvement and study length indicates that this increase is not statistically significant ($\tau b = 0.22$, $p = 0.14$).

5.6.2.4 Discussion

Interest selection is consistent with the formative study – all three algorithms typically agree on the top-ranked IMB, but vary in their selection of lower ranked interests. We see more agreement in this study than the formative data analysis. MIR performs strongest of the three for recall, and comparably on precision. However, differences between the algorithms were not statistically significant ($p = 0.06$ for recall). Given considerable disruption to participants' mobility caused by COVID-19 (impacting 42% of the summative dataset), and a clear trend for strong precision/recall in both the formative and summative studies, we suggest that on balance MIR does outperform other algorithms, particularly in identifying postliminary IMBs. Future work (when mobility is less constrained) would be helpful to provide further evidence.

MIR selects between one and four top-ranked interests for our participants (mean and median both 2). We again see a high incidence of cases in which a dynamic cut off generates a single top-ranked IMB, although this is lower for MIR than other algorithms. Whilst this may not be ideal for some applications, our closing interviews suggest that some participants may genuinely consider themselves to have only a very limited set of intrinsic motivations. For example, when interviewed, P2^s reported that they did not consider themselves to have any interest outside of their studies. However, we do see a noticeable difference in the size of the difference used as a cut off for participants where a single IMB is selected (MIR mean 0.27) compared to those where multiple IMBs are selected (MIR mean 0.55). Our summative rankings show no outliers in terms of list length (max is 4), but we acknowledge the likelihood for

COVID-19 mobility restrictions to have artificially limited the number of locations and thus, potentially, identified IMBs. Overall, the formative and summative data suggests that for most individuals, a list of between one and four ranked items would encompass the dominant IMBs. Thus, for applications requiring a static N value, we propose that three to four is likely to be most appropriate.

Our results show more locations annotated as being related to safety and physiological needs according to Maslow's hierarchy when compared to the formative data (55% compared to 41%); broken down, we see a greater proportion of safety (40%) needs than physiological (15%), another reversal of the pattern seen in the formative dataset. Restrictions on mobility likely reduced visits to locations associated with higher level needs (since these overlap near completely with those considered non-essential by many societies). However, we also note that significant differences in age between the two participant sets may well have influenced this result – prior literature documents that fact that older participants (such as those in our formative dataset) are more likely to pursue intrinsic interests than younger (as in our summative data) (Sheldon and Kasser, 2001; Gomez et al., 2012). Further, during interviews our participants reported pursuit of these supposedly extrinsically-motivate behaviours for both intrinsic and extrinsic reasons – thus, the combination of needs with our other measures is important to appropriately produce an accurate MIR score. Unsurprisingly, both of our SDT measures positively correlate with the MIR score (Table 5.11).

Both the formative and summative data show considerable individual variation – for example P5^{s} and P1^{s} demonstrate a very high proportion of safety and physiological needs (combined these amount for 100% and 75% of top-ranked interests for P1^{s} and P5^{s} respectively), whilst P6^{s} and P7^{s} demonstrate no safety or physiological needs. This pattern is also evident in our correlations, where P6^{s} and P7^{s}'s propensity to exhibit IMBs that are meant primarily to satisfy Maslow's intrinsic needs manifests as a stronger correlation between MIR and the needs measurement (Table 5.11). We suggest that this variation does indicate that personalised weighting of individual elements could further improve the performance of MIR. Comparison with other single-measure approaches lends credence to our approach. For instance, the IMBs of P4^{s} relate to education, socialising and drinking, and music (Table 5.9). Considering frequency alone successfully identifies education as the strongest interests, places socialising and drinking fourth, but ranks music-related activities only ninth in the behaviour list. Similarly, duration alone ranks education first, socialising and drinking fourth, and musical activities eighth. The persistence of behaviours (an autonomy indicator under

SDT) associated with music, and with socialising and drinking, combined with their higher needs level (compared to, e.g., travel and transport which ranks third using both single-measure approaches), positively impact their rankings under our approach. The IMI score and participant's interview confirm our findings.

In addition to evaluating the final set of IMBs (i.e. those identified by the end of the study period), our summative study provides interim measures using the adapted IMI subscale to evaluate one (randomly-selected) top-ranked interest. Improvements in IMI scores for participants over time suggest that MIR does select more accurate IMBs later in the study. However, the limited scale of the data points involved suggests that further work is needed to confirm this trend.

Finally, we note that the one major area of difference in results from our two studies relates to the period of time before the calculated strength of an IMB stabilises. Specifically, we see that whilst on average behaviours stabilise more quickly, some behaviours fail to stabilise within the study period. The longer duration of the formative study maximised the potential for identifying behaviours that took a relatively large number of weeks to stabilise (e.g. MIR for $P7^f$ required 10.38 weeks to stabilise and frequency and duration for $P6^f$ required 14.00 and 10.33 weeks respectively). Further, we suggest that both patterns are potentially the result of COVID-19 restrictions – reduced mobility will mean that some behaviours could be pursued more consistently, at the expense of other behaviours that could not be pursued.

However, we acknowledge that such variation could be due to the demographic differences between the two groups. Conditions such as age, health condition and work may impact the type of interests and how they are realised by each group. Nevertheless, our results suggest that the MIR approach can be used to detect interest regardless of how the two groups may differ in the type and realisation of those interests. This is in line with the findings from the literature that do not relate the indicators of motivation properties – used in this work – to a specific type of interests or group of people (Ryan and Deci, 2017). Instead, these indicators are generic and describe behaviours that are motivated by personal interests. Comparison that shows how various aspects; such as the interests type and how they are realised; may differ across demographic groups can be further explored by future work.

5.7 Limitations and Future Work

In this chapter we proposed a generalisable set of static and dynamic motivation properties that are grounded in established psychology literature (needs, competence, autonomy and novelty). We then establish a set of behavioural measures for those properties, using smartphone location traces as the foundation for a proof-of-concept realisation. Results from two evaluative studies indicate that our approach successfully generates a ranked list of IMBs, and that the generated list more closely aligns with participant self-reports than two comparative single-measure approaches.

Although our motivation properties and behavioural measures should be applicable beyond location, our formative study showed that the semantic annotation upon which our location measures are based has considerable impact on the results. Thus, in our summative evaluation, we provided a means by which participants could correct the annotation of their extracted stay-points (i.e. item identification). All but one participant in the summative study applied corrections to at least one annotation (Table 5.12), with a total of 124 corrections made overall (impacting 16% of all locations). Enhancing the accuracy of smartphone-based location detection, and the associated semantic annotation, is beyond the scope of this work. However, we do note a specific challenge with regards to the use of location semantics as an indicator for IMBs. Our present approach assumes each location serves a single purpose and that purpose is consistent over time, but in practice some users will visit the same location multiple times, each with very different intentions. For example, one user may visit a coffee shop with the primary purpose of eating lunch (a physiological need), to engage in work or quiet study (a safety need), or to engage in a hobby (e.g. reading, knitting) alone or with others (intrinsic). Thus, the one-to-one mapping reported in this chapter may limit the recognition of the performed interest at these places. As mentioned in Section 5.4.1, we rely on the category that is retrieved by the Foursquare to enrich our modelling; approaches to overcome this challenge are left for future work, but will likely involve the integration of additional contextual data. Also, details of the performed interests can sometimes be difficult to recognise by relying only on location data and Foursquare annotations. For instance, being at a place related to football, by itself, may express a potential interest in football regardless of how that interest is being actualised (either through watching or playing). Future work can benefit from additional sensors and contextual data to improve the granularity of the detected interests.

61.29% of annotation corrections relate to the participants' home locations (i.e. incorrect identification of, or failure to correctly identify), a reflection of our decision

Table 5.12: Number of Stay-Points (SP) for each participant based on foursquare results and the participant corrections. Foursquare results are categorised as either residence SP (SP^{Res}) or non-residence SP ($SP^{!Res}$). Likewise the correction are either from non-residence SP to residence SP ($SP^{!Res \rightarrow Res}$) or from non-residence SP to another non-residence SP ($SP^{!Res \rightarrow !Res}$).

	Foursquare results		**Participant corrections**	
	SP^{Res}	$SP^{!Res}$	$SP^{!Res \rightarrow Res}$	$SP^{!Res \rightarrow !Res}$
P1^{s}	31	186	23	15
P2^{s}	27	73	6	2
P3^{s}	8	93	6	15
P4^{s}	11	91	10	1
P5^{s}	6	119	16	13
P6^{s}	6	53	0	0
P7^{s}	3	44	15	2

to allow correction prior to establishing the home location using the method described in Section 5.6.1; when detecting home, the percent of correctly annotated places has raised from 83.49% to 93.61%. Our rolling weekly window was intended to manage temporary changes in residence, but to reduce the correction burden on users we could easily modify our study to do some residential annotation on the device (e.g. showing the location considered to be prior week's residence, or the location at which the user had spent most time thus far in the present week). We also observe a high proportion of "education" IMBs appearing in our summative dataset, raising the question of whether frequently occurring places of work/study should be excluded. Our interviews suggest that participants were intrinsically (as well as extrinsically) motivated to visit places of education, and we also acknowledge that many individuals will pursue a career that relates to their hobbies and interests. Thus, we argue that omitting work and study locations completely may be problematic. Further studies at scale should help to validate if MIR is successfully differentiating between participants whose workplace/education visits are intrinsically motivated compared to those whose visits are made purely out of necessity.

We saw significant problems caused by mobility restrictions as a result of COVID-19, suggesting that there is still value in further longitudinal study (as restrictions ease) to provide confirmatory evidence for our approach. Such studies could also attempt to replicate some of the more novel findings from our summative dataset, for example whether trends for IMI scores improve as the study progresses. Additional IMI prompts could also be used to confirm the accuracy of low-ranked IMBs (i.e. asking participants about a low ranked IMB results in a low IMI). We also note that fixed study parameters (although based on prior literature) may warrant further exploration, for example using alternate or dynamically determined values for time/distance when extracting staypoints, or as a recency threshold.

Mobility restrictions also demonstrate the vulnerability of relying on a single sensor stream. However, the overall motivation properties and behavioural measures were intended to be applicable beyond location, meaning that other sensor data could replace or supplement that collected in this proof-of-concept. Monitoring engagement with phone apps or websites could be used to determine the same measures of needs (e.g. recipe app corresponds to a physiological need to eat while visits to a sports website suggests fulfilling an intrinsic need), intensity (e.g. frequent high-duration use of the same app or multiple semantically similar apps), sustainability (e.g. sustained visits to a website over time) and recency. Further, some sensor streams may open up opportunities for alternative or additional behavioural measures that correspond to the identified motivation properties (e.g. diversity and flexibility in behaviour patterns as alternative measures of autonomy: Ryan and Deci, 2017), and provide opportunities to integrate measures for relatedness (a key need in SDT that was omitted from this work due to its focus on location behaviour). Thus, we propose a key area for future work centers on (a) establishing new sensor streams for which the identified behavioural measures can be used to derive motivation properties, (b) considering novel behavioural measures that correspond to both those same motivation properties or to relatedness, and (c) determining how best to combine multiple sensor streams (and any novel measures) into a more robust MIR.

Lastly, this work is limited by what is referred to by researchers (Aeffner et al., 2017; Vega-Hernandez, 2019) as the “Gold Standard Paradox.” The contradiction arises from the fact that our study adopts digital phenotyping, which relies on sensors’ data that are objectively produced. The goal is to use knowledge derived from those data to replace the subjective self-reporting methods and avoid their issues (e.g. memory and recall biases). However, these self-reporting tools (IMI in our case) are the best instruments for assessing our method. As a result, our techniques must be assessed using the scales that they are attempting to replace.

Our studies both demonstrate considerable individual variation, particularly with regard to execution of activities that correspond to the lower levels of Maslow’s hierarchy. This was predicted by prior literature that suggests that while all individuals share the same fundamental needs, their pursuit of them may vary (e.g. based on the degree to which they have previously had that need fulfilled (Deci and Ryan, 2000)). Thus, our overall approach (Figure 5.3) incorporated a personalisation step that is not realised in our proof of concept. Future work should consider the degree to which coefficients within MIR can be personalised to deliver a more tailored IMB calculation.

Lastly, there are two interesting ways of generating the top *N* results and we explore both of them. In the first study, we explore the difference between a dynamic *N* and static *N* in a formative and detailed manner. In the second study, we focus on a dynamic *N* with one cut-off point to preserve only the strongest interests and avoid asking participants about too many and possibly irrelevant interests. However, we do not suggest at this point that either the dynamic or static approach is better than the other or a specific approach works best for a particular population. Instead, these deeper investigations are still worthy of further exploration.

5.8 Conclusion

In this study, we present a novel approach to technology-based identification of behaviours that reflect underlying personal interest (IMBs). Specifically, we build on current psychology theory to identify a set of core set of static and dynamic motivation properties (needs, competence, autonomy and novelty). From these properties, we use the literature to establish a set of behavioural measures that can be derived from passively-sensed smartphone location traces. We combine these measures in a single MIR score whose value (0-1) represents the strength of intrinsic motivation associated with a behaviour.

Through a combination of formative and summative evalution, we show how our approach can facilitate personalised understanding of IMB compared to frequency- and duration-based approaches. Our results indicate that our approach successfully identifies IMBs that are consistent with those reported by participants, matching or outperforming alternatives. Results also suggest that most IMBs can be detected over fairly short time periods (within 2 months) and adapts quickly to variations in mobility patterns as users' motivation changes.

Our proposed approach allows for unobtrusive detection of individual user's IMBs based on a standardised weighting of each motivation properties. To further improve personalisation, we suggest that variation in the weighting of different motivation properties could allow the model to better reflect difference in individuals' need satisfaction (i.e. the degree to which they pursue a particular need).

Personalised IMB identification of the kind enabled by our work has the potential to facilitate new applications that capitalise on individuals' intrinsic motivation. This could be particularly valuable for behaviour change, where existing evidence indicates leveraging intrinsic motivation leads to more effective and sustained change (Ryan and Deci, 2017). For example, we envisage fitness applications that prompt users to take a slightly longer journey home by plotting a route that is consistent with personal interests (e.g. passing by a local soccer club for a user for whom soccer is considered an IMB).

Relying on a single data indicator in behavioural analysis is inherently risky, and mobility restrictions imposed in response to COVID-19 have provided a clear demonstration of this vulnerability. Our discussion of future work highlights a set of promising smartphone data sources that could be used as measures of our fundamental motivation properties. Adoption of these data sources would see them used as alternative or complementary indicators of the identified motivation properties and associated behavioural measures. Thus, the overall approach taken in this paper is easily applicable irrespective of the underlying sensor set.

Chapter 6

Interest Recognition from Multiple Smartphone-Derived Behaviours

In the previous chapter, we presented our MIR method and showed how it could be used with a single behaviour (i.e. mobility behaviour). In this chapter, we extend our MIR approach to incorporate mobility, buying and phone usage behaviours. We show how our combinatory MIR (cMIR) is standardised to be implemented across multiple behaviours. We have benefited from the knowledge detailed in chapters 2, 3 and 4 as well as the findings that we stated in the previous chapter to design a six-month study and implement our method. The integration of multiple behaviours has been accomplished in a personalised manner, and the conducted study was used to evaluate our cMIR method.

The main content of this chapter is a paper authored by: *Ahmed Ibrahim, Sarah Clinch and Simon Harper*. The title of the paper is: ***Recognising Personal Interests: A Combinatory Approach based on Smartphone-derived Behaviours and Intrinsic Motivation***. The paper is currently under review. For this thesis, we edited some formatting styles, such as the sizes of some tables for consistency and readability reasons.

Author contribution

Ahmed Ibrahim designed the study, carried out the data collection, analysed and synthesised the results and wrote the paper. Sarah Clinch and Simon Harper provided continuous feedback throughout all the stages of the study, offered advice and discussion and contributed vital edits to the paper's writing.

Abstract

In this paper, we aim to recognise individual interests from daily behaviours. Knowledge of individual interests can be utilised to personalise behavioural recommendations and promote health and well-being. Typically, self-reporting methods are used to understand people's interests; despite that, in most cases, interests are demonstrated in an individual's daily activity. Moreover, interests are not the only motives behind daily behaviours. Instead, many of our daily actions are motivated by other reasons such as obligations and external rewards. Therefore, in this work, we present a motivation-based approach to recognise daily behaviours that are motivated by individual interests. The daily behaviours are derived from passively and continuously sensed smartphones data. Our approach combines behaviours from multiple smartphone data streams to overcome the limitations of relying on a single one that is acknowledged by the related literature. To evaluate our approach, we conduct a six-month real-world study. The results indicate that our approach performs significantly better than traditional interest recognition methods and presents a better understanding of internally interesting behaviours.

6.1 Introduction

Individual interests (i.e. the things that motivate and excite us) directly shape our everyday habits and behaviour, including behaviours with significant health or well-being implications. Thus, establishing an individual's interests can support the design of effective personalised behavioural interventions and nudges (Mills, 2020; Schoning et al., 2019). However, recognising these interests is not without challenge – psychological sub-disciplines tend towards the use of paper-based inventories/questionnaires (e.g. the Work Preference Inventory (Amabile et al., 1994), the Intrinsic Motivation Inventory (Ryan, 2018); and the STEM semantics survey (Tyler-Wood et al., 2010)), but these self-reports are limited by individual's biases (Vega-Hernandez, 2019) and sometimes by people being unaware of their own interests (Renninger and Hidi, 2016). Moreover, self-reporting tools are discrete and hence do not capture behavioural dynamics which require continuous observation (Vega-Hernandez, 2019); attempts to overcome this weakness through longitudinal analysis can be highly intrusive and prone to memory and recall biases (Matthews et al., 2020; Hassan, 2006). However, one of the most straightforward and representative indicators of individual interest is behaviour itself. Thus, in this paper, we seek to recognise individuals' interests from passively sensed smartphone's data that characterises their everyday activities.

Passive sensing (also referred to as digital phenotyping when used in health and well-being settings (Onnela and Rauch, 2016; Vega-Hernandez, 2019)) uses digital devices (often a smartphone) to continuously and longitudinally capture data describing an individual's behaviour and context. Crucially, this data capture occurs with little-to-no intervention from the user. Digital phenotyping has been applied in a wide variety of applications including monitoring of patients with Parkinson's disease (Vega-Hernandez, 2019), Schizophrenia (Difrancesco et al., 2016) and depression (Wahle et al., 2016; Burns et al., 2011).

Traces emerging from digital phenotyping describe the 'what' of individual behaviours, but in this work we also seek to establish 'why' – i.e., what are the interests (*intrinsic motivations*) implied by the patterns of behaviour observed. Many of our daily actions are motivated by obligations and external rewards (*extrinsic motivation*) (Ryan and Deci, 2017). Therefore, identifying the motives behind observed daily activities is essential to distinguish behaviours driven by personal interests from those motivated by other factors.

In this paper, we build on a previous work from Ibrahim et al. (2021a) that utilises motivational knowledge to recognise interests from behaviours identified from a single sensor trace (location). Unlike the existing method (which the authors refer to as Motivation-based Interest Recognition approach – MIR), we integrate multiple data streams to capture a broader set of behaviours. This multiplicity of behaviours overcomes the limitations of relying on a single one that is acknowledged by the previous work of Ibrahim et al. (e.g. the inability to capture phone usage interests from mobility behaviour) . In addition to the location traces used by Ibrahim et al., we additionally fold in data describing interactions with installed smartphone apps, and data extracted from notifications that relate to buying behaviours (e.g., "Your order of ... has been shipped"). The addition of these behaviours substantially impacted the realisation of the MIR subprocesses as well as the modelling details. We name the resulting measure *combinatory MIR* (cMIR).

In cMIR, events of the three behaviours, such as going to a movie theatre or buying a gaming device, are first extracted from the raw smartphone data. Next, we identify specific measures that can be used to operationalise properties of two psychological models of human motivation: Self-Determination Theory (Ryan and Deci, 2017) and Maslow's Hierarchy of Needs (Maslow, 1943). We apply those measures on the extracted behavioural events to understand the underlying motivations and rank the prior behaviours accordingly. For example, if the extracted event is shopping, the rating produced from applying the measures of motivation properties would indicate how much a person is motivated by shopping. Higher ratings imply more internalised actions and, therefore, a better chance of the person being intrinsically motivated.

Our cMIR model is evaluated using data from eight participants (~0.8 million datapoints collected over six months). Our results significantly indicates that our approach produces ranked interests that align closely with those elicited from participants through self-reports. We achieve higher precision and recall when compared to alternative approaches including the base MIR that relies on a single behavioural trace. We also show how our model adapts to each individual and discuss the significance of that on personalisation.

To summarise, our contributions are threefold:

- **An algorithm for a combinatory Motivation-based Interest Recognition (cMIR)** that aggregates multiple behaviours and the identified behavioural measures into a ranked set of Intrinsically Motivated Behaviours (IMBs) (i.e. where the top-ranked item is the one for which the user has the strongest interest).
- **The adaptivity of the cMIR to each individual's data**. By tailoring the cMIR method according to an individual's data, we provided a flexible method capable of reacting to the interest dynamics and producing a personalised model from multiple behaviours and from an overfitted set of motivation properties.
- **An evaluation based on a real world dataset** that validates ranked IMBs against participant ground truth and three alternative ranking approaches.

6.2 Related work

The subject of this work is interdisciplinary that from one side digs into the cognitive and psychological studies of human motivation and from the other side, investigates how interests are realised in technological platforms such as recommender systems. Therefore we discuss the related work with respect to these disciplines.

6.2.1 Human Motivation

Theories of human motivation attempt to describe why humans do what they do (Ryan and Deci, 2000; McClelland, 1987; Weiner, 1992). Biological approaches focus on physiological state and processes (Cofer and Appley, 1964; Petri and Govern, 2013), and examples include Yerkes-Dodson (Yerkes and Dodson, 1908), drive reduction (Hull, 1943, 1952) and operant-conditioning (Skinner, 1953; Cooper et al., 1987). For the purposes of understanding and leveraging individual differences, however, these may be considered overly reductive (Strombach et al., 2016; Eccles and Wigfield, 2002).

Psychological mappings from human behaviour to motivation typically take one of two approaches. *Static approaches* use a fairly rigid classification to match behaviour to underlying physiological or psychological needs. Examples include Maslow's hierarchy (Maslow, 1943) and Murray's need theory (Murray, 1938). By contrast, *Dynamic approaches* quantify motivation based on the subjective impression of a participant toward a performed behaviour; factors such as contexts and rewards may impact the participant's attitude toward an activity (Fogg, 2012; Ryan and Deci, 2017). Examples include Self-Determination Theory (Ryan and Deci, 2017) and Fogg's motivational waves (Fogg, 2012).

For the purposes of this work, we draw on two dominant psychological explanations: Maslow's Hierarchy of Needs (Maslow, 1943) and Self-Determination Theory (Ryan and Deci, 2017). In particular, this work focuses on determining an individual's intrinsic motivation (Ryan and Deci, 2000) – activities that inherently bring satisfaction to an individual (commonly referred to as interests, (Renninger and Hidi, 2016)).

6.2.1.1 Maslow's Hierarchy of Needs

Maslow (Maslow, 1943) discusses motivation in terms of five needs, the lowest of which must be fulfilled before the next comes into focus. These five needs (presented lowest to highest) are as follows: physiological, safety, belongingness, self-esteem and self-actualisation need. *Physiological* needs relate to survival at an individual and species level, such as food, drink, sleep and sex. *Safety* needs include stability, security and protection from fear. *Belongingness* needs are driven by the desire for interpersonal relationships, and feelings such as love, friendship and acceptance. *Self-esteem* needs drive our desire for respect, dignity and independence. Finally, *self-actualisation* needs drive our ambition and desire for personal growth (Maslow, 1943; McLeod, 2007).

Critics of Maslow (e.g. Neher, 1991) suggest that experiencing these needs in the proposed order is contrary to evidence in the real world. For instance, lack of security in some communities – due to war, civil unrest or similar – does not prevent their inhabitants from developing social ties and pursue the fulfilment of belongingness needs. Despite this, Maslow's hierarchy continues to be highly influential (including, for example, in recent attempts to understand technology: (Houghton et al., 2020); (Kang and Jung, 2014)). In light of the identified limitation, in this chapter, we focus on mapping behaviours to needs as nominal categories rather than concerning ourselves with an ordinal progression between the levels.

6.2.1.2 Self-Determination Theory

SDT (Ryan and Deci, 2017) is one of a number of contemporary theories that build on the distinction between intrinsic and extrinsic motivation. However, in contrast to others, SDT treats these concepts not as a dichotomy, but instead as a continuum that ranges from amotivation, through a set of extrinsic motivation states, to a fully internalised intrinsic motivation.

SDT identifies competence, autonomy and relatedness as the three basic psychological needs that differentiate and represent motivation states: the need for competence (also called self-efficacy), the need for autonomy, and the need for social relatedness (Ryan and Deci, 2017):

1. *The need for competence* refers to one's belief in their ability to perform (Bandura, 1971). Self-perceived success, satisfaction or efficiency when engaging in tasks helps to satisfy the need for competence (Ryan and Deci, 2017; White, 1959, 1963).

2. *The need for autonomy* relates to the extent to which a person controls a behaviour (Ryan and Deci, 2017), and self-regulates goals and the process of attaining them (Schunk et al., 2008).

3. *The need for relatedness* is concerned with feelings of connection with others and is an essential driver for social behaviour (Ryan and Deci, 2017).

In addition to these three basic needs, proponents of SDT have noted the importance of novelty in motivating individuals to pursue and possibly change personal interests (Ryan and Deci, 2000, 2017; González-Cutre et al., 2016; Silvia, 2007). This has in turn led some to propose *the need for novelty* as a further innate psychological need (González-Cutre et al., 2016).

As a popular and "living" theory (Vansteenkiste et al., 2010), SDT has been applied in a wide variety of domains, including many related to technology. For example, SDT can be used to guide the interface design of mobile apps (e.g. Zuckerman and Gal-Oz, 2014; Rooksby et al., 2015), encourage the use of health apps (e.g. Saksono et al., 2020), or propose behavioural intervention (e.g. Gustafson et al., 2014). Unlike these applications, we employ SDT to classify behaviours, that are passively sensed, as either extrinsically or intrinsically motivated.

6.2.2 Recommender systems

Recommender systems are software applications that aim to make predictions about the items or behaviours that might be of interest to a specific individual (Zhang et al., 2019; Raza and Ding, 2019). The fundamental concept centres on the use of existing indicators of a target user's interest (e.g. ratings, purchases, frequency of interaction), together with knowledge about all of the items that could be recommended (e.g. object classifications, features, other users' ratings or interactions) to derive a rating for each item in a set of possible recommendations. Based on this rating, the top n items can be presented to the target user (or user group) (Adomavicius and Tuzhilin, 2005).

Although they are most prominently used in e-commerce (Zhou et al., 2018; Schafer et al., 2001; Lu et al., 2014) and digital media consumption (Beam, 2014; Gomez-Uribe and Hunt, 2016), recommender systems can be used to encourage broader behavioural change. For example, Rabbi et al.'s smartphone application `MyBehavior` (Rabbi et al., 2015) used a combination of passive sensing and manual logging to record physical activity and food intake. Statistical machine learning was then used to identify and recommend high calorie loss behaviours similar to the user's existing behaviors, resulting in a statistically significant increase in physical activity and corresponding decrease in calorie intake compared to a control condition.

Existing recommender systems that make use of smartphone's data typically consider the frequency of behaviour (Musto et al., 2018; Yu and Chen, 2015; Li et al., 2015a), duration (Boytsov et al., 2012) or combination of both (Do and Gatica-Perez, 2014; Li et al., 2008). When combined with time, a recommender system can also deliver recommendations based on the both frequency and duration of actions at a specified time of day (Yuan et al., 2013; Li et al., 2017a), or based on recency of behaviour (Li et al., 2015b; Logesh and Subramaniyaswamy, 2017).

Despite their complexity, the majority of recommender systems fail to consider the underlying motivation that led to the user behaviours used as indicators. Whilst many of these indicators may reflect personal interests, others will be the result of obligations (e.g. buying a gift for others, visiting a workplace). In this chapter, we set out to identify Intrinsically Motivated Behaviours (IMBs) – behaviours that reflect intrinsic motivation (Ryan and Deci, 2000). In so doing, we aim to enable future recommendations that align with the individuals' interests and values and can facilitate sustained behaviour change (Michie et al., 2011).

6.2.3 Digital phenotyping

Digital phenotyping is defined as the naturalistic moment-by-moment quantification of an individual's behaviour (Onnela and Rauch, 2016). It is achieved using digital devices such as smartphones to sense human behaviour either passively (e.g. GPS) or actively (e.g. questionnaires) (Torous et al., 2016). Researchers use several terms to describe the same goal, such as "personal sensing" and "context sensing" (Mohr et al., 2017; Burns et al., 2011). mHealth studies that are designed for behavioural interventions but do not rely on digital phenotyping (e.g. Chen et al., 2018) are outside the scope of this work.

Traditionally, digital phenotyping is implemented by unobtrusively and naturalistically collecting and syncing data to a backend server. AWARE (Ferreira et al., 2015) and EARS (Lind et al., 2018) are open source tools designed for passively collecting digital data. Developers can extend AWARE to extract behavioural features through plugins (e.g. activity recognition and app usage) (Ferreira et al., 2015). Beiwe (Torous et al., 2016) is another tool that provides, in addition to basic sensing functionalities, codebase data analysis pipeline for conducting the behavioural analysis.

The data collected from sensing tools form the basis for digital phenotyping. For instance, app usage has been used as the basis for digital phenotyping to detect the mood of bipolar patients (Alvarez-Lozano et al., 2014). Symptoms of Parkinson's disease have been investigated from mobility-based phenotyping to improve the patient's quality of life (Vega-Hernandez et al., 2017). Loneliness indicators, as well as physical activities, have been investigated within the context of older adults (Seifert et al., 2017; Sanchez et al., 2015). Behavioural features of depression symptoms have been extracted and studied to detect depression (Wahle et al., 2016; Burns et al., 2011), whereas location features have been employed to detect out of home activities in schizophrenic Patients (Difrancesco et al., 2016).

Besides the above implementations, digital phenotyping has intersected with human motivation in several studies. For instance, the role of presenting walking time on incentivising a participant to walk has been studied through digital phenotyping

(Zuckerman and Gal-Oz, 2014). Other studies use self-determination theory and social rewards to encourage the use of a health app (Rooksby et al., 2015; Gustafson et al., 2014). However, rather than using motivational knowledge to understand interests, these studies focus on the role of elements, such as rewards and visualisation, in incentivising a behavioural change.

Lastly, the use of visualisation to support the implementation of digital phenotyping has been utilised. Health Mashups (Bentley et al., 2013) depicts the connection between sleep, weight, pain. In Visual Cuts (Epstein et al., 2014) data is summarised to help users identify meaningful findings. Passively sensed location and activity data had been visualised for reflection analysis (Tang and Kay, 2017).

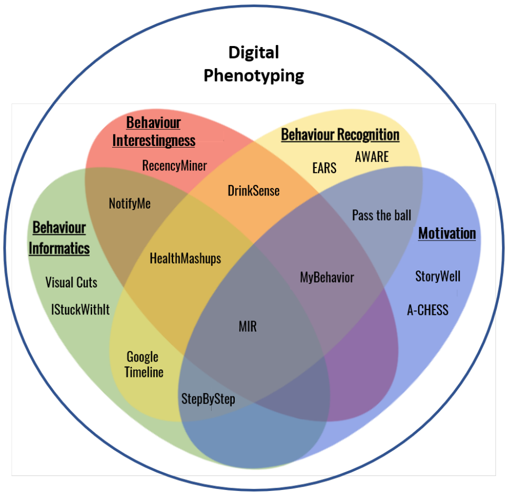


Figure 6.1: Existing tools that derive behavioural aspects from digital phenotyping.

To the best of our knowledge, this is the first work that uses digital phenotyping to recognise personal interests from multiple behaviours. Doing so can help mitigate the drawbacks of relying on a single behaviour that is acknowledged by the previous work of Ibrahim et al. (2021a). Also, our combinatory approach shares the goal of Ibrahim et al.'s work which is to avoid self-reporting drawbacks while focusing on an individual's behaviour.

6.3 Our approach

In this chapter, we set out to measure aspects of human motivation, continuously and unobtrusively, by using the smartphone data that is captured as individuals go about their daily activities. We focus specifically on trying to identify behaviours that are internally motivated. In particular, we build upon prior work from Ibrahim et al. (2021a), integrating psychoanalytical and data-driven techniques to rank behaviour based on motivation properties.

The processes required to recognise interests from multiple behaviours overlap with those used to recognise interest from smartphone trajectories proposed by Ibrahim et al., but with substantial rethinking to address the unique challenges that come from the differences between events of multiple behaviours. These processes are *indicator identification*, *item identification*, and *interest determination*. In this work, we propose a different realisation for the last two processes while using the same set of motivation properties and behavioural measures specified by the first process (i.e. indicator identification). Our realisation aims to improve the generalisability of the MIR method across multiple behaviours. In the next subsections, we summarise the three processes and detail the proposed improvements on item identification and interest determination processes. In each one of the two processes, we highlight the differences between this work and the one proposed by Ibrahim et al..

6.3.1 Indicator identification

We rely on the same set of motivation properties that are used by Ibrahim et al. (2021a), namely *Maslow's needs*, *competence*, *autonomy* and *novelty*. These properties are based on Maslow's hierarchy of needs and Self-determination theory. To operationalise those properties, needs level, intensity, sustainability and recency are identified as behavioural measures. High levels of perceived competence are associated with completing an action more often (Ryan and Deci, 2017; Rabbi et al., 2015), and the time spent on the action (Nicholls, 1984; Fishbach and Hofmann, 2015). Thus, frequency and duration are combined into a measure of intensity (Wolf and Hopko, 2008). Voluntary performance of action (i.e. autonomy) is associated with action sustainability over time (Pelletier et al., 2001; Seguin et al., 1999), whereas the propensity to seek out novelty is manifested in the exploration of new behaviours and gaining new interests (i.e. recency) (Ryan and Deci, 2013, 2017).

However and as stated by Ibrahim et al., this indicator set is not considered to be exhaustive; future work may demonstrate the utility of alternative or additional measures. For example, an experiment may use the diversity and flexibility in times of a behaviour as a measure of autonomy (Ryan and Deci, 2017) along with sustainability. Similarly, we envisage future integration of a measure for relatedness, through indicators derived from proximity or other sensors.

6.3.2 Item identification

The MIR approach proposed by Ibrahim et al. operates over events of a single behaviour (i.e. mobility). Compared to Ibrahim et al., our approach abstracts over an even more diverse set of behavioural events than that of prior work (in addition to location, we extend prior approaches to include smartphone application interactions and indicators of buying behaviour). These behaviours are encoded within streams of low-level data (GPS traces for location, screen touches for application interactions, and notifications for buying behaviour). Each low-level data stream contains its own unique 'noise' in addition to the target signal, and thus an initial processing step is required to identify individual events from the source data.

For the mobility behaviour, a sequential processing of GPS data is performed to extract stationary periods (known as stay-points) based on predefined time t and distance d thresholds (appropriate values for t and d may be implementation specific and are explored in Section 6.4). The resulting stay-points are places where an individual has lingered for a period greater than t minutes within a boundary of d meters. Note that for the purposes of this work, we do not consider the intermediate periods (i.e. periods of motion) [1].

Once extracted, stay-points are semantically labelled using a location annotation service. Google Places, Foursquare, and other platforms each provides reverse geocoding services that can be used to annotate stay-points (e.g. converting a location reading of `51.5194° N, 0.1270° W` to `The British Museum`) and then appropriately categorising that as a `history museum`. In this work, we use Foursquare as our location annotation service due to its richer categorisation (950 categories) compared to other services (e.g. Google provides 96 categories).

In a similar way, items of phone usage behaviour are identified through sequential processing of smartphones' interaction data to extract events of interaction based on the details of the used app (smartphones allow for interaction with one app at a time). Contextual information about battery and screen status is used for better identification of the segment boundaries. An event is formed when the phone is out of battery or the

[1] These move-points could represent either a necessary transition between two stay-points, or one of a limited set of (typically fitness or sporting) activities that are intrinsically motivated (e.g. walking, running). In the case of the former, there are more appropriate sensors that could detect these specific activities (e.g. accelerometer: Wannenburg and Malekian, 2017).

user turns off the screen. Otherwise, since smartphones allow interactions with one app at a time, changes between apps are used to identify the events. The resulting events represent the apps that an individual has interacted with for a period of t minutes. Once extracted, events are semantically labelled using an app annotation source. GooglePlay or AppStore can be used to annotate phone usage events based on the apps' categories (e.g. mapping a package name of `com.score365` to `sport`). In this work, we use GooglePlay as our app annotation service since we rely only on android devices to collect the data.

Extracting buying events from the notification text differs from the previous two. First, notifications are filtered to preserve the ones that are related to buying behaviour[2]. Specifically, a notification is related to buying if (i) it is generated by an app that has a category typically associated with a buying activity (e.g. `Shopping` and `Food_AND_Drink`) or (ii) the text contains a buying-related keyword (e.g. `'your order'` and `'your payment'`)[3]. Once the buying-related notifications are extracted, their texts are semantically labelled using an external annotation service. Cloud Natural Language API, Amazon and other platforms can be used to classify the text based on predefined categories. For instance, Cloud Natural Language API classifies the following notification as `Music`: "Bax Music — Confirmation of order Thank you for your order Dear ...". However, in many cases, the Google Cloud Natural Language API may find the notification text too short and hence fail to assign a category. Therefore, we first search Google Cloud Natural Language API for a category. If the text is too short and the API fails to get a category, we use Amazon Comprehend API to get the category. We start the search with Google as it is not e-commerce specific and hence is expected to cover a broader range of potential bought items' names (compared to Amazon, which is an e-commerce specific platform).

Besides how the events of each behaviour are extracted, the applicability of the behavioural measures is impacted by the type of behavioural events. For instance, intensity (the measure of competence) combine duration and frequency. Although events of phone usage and mobility behaviours derived from the smartphone are expected to possess the duration attribute, the derived buying events (using notifications) can not lead to the duration a specific person spends on a buying activity. Therefore, an estimation of the time a person takes to make a purchase is essential to standardise the measures' implementation across multiple behaviours. In this work, we rely on the

[2]Comparison between the different approaches and keywords used to filtering notifications are detailed in another paper that is under review.

[3]The list of the used keywords are explored in Section 6.4.

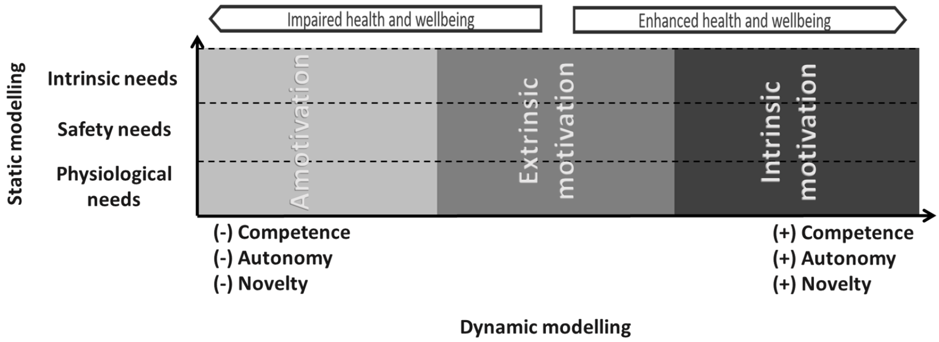


Figure 6.2: The motivation continuum proposed in SDT runs from amotivation through to intrinsic motivation.

existing knowledge and estimate that time (estimation details may be implementation-specific and are explored in Section 6.4).

6.3.3 Interest Determination

We build on Ibrahim et al. (2021a) prior contrast of Maslow (Maslow, 1943) and SDT (Ryan and Deci, 2017) as static and dynamic approaches respectively (Section 6.2.1), and develop a solution that engages both mechanisms in parallel. A *static modelling* step (Section 6.3.3.1) builds on Maslow's heirachy, whilst our *dynamic modelling* (Section 6.3.3.2) is based on SDT. Computed values from both dynamic and static measures are combined to form the motivation score based on behaviours derived from multiple traces (referred to as the cMIR score). The cMIR score is used to classify events of those behaviours as intrinsically or extrinsically motivated based on the interplay between the static and dynamic scores (Figure 6.2). Specifically, higher ratings imply more internalised actions and therefore a better chance of the person being intrinsically motivated. However, both static and dynamic modelling are heavily impacted by the multiplicity of behaviours. We detail each of these modelling steps and highlight the difference in the following subsections.

6.3.3.1 Static Modelling

The static aspect signifies the nature of the underlying needs and introduces a fixed mapping of behavioural events to one of Maslow's five levels of needs: physiological, safety, belongingness, esteem and self-actualisation need. The static property refers to the independence between the participants' data and the event-motivation mapping.

We adopt the scoring system proposed by Ibrahim et al. (2021a) that is built on the literature of human motivation. Specifically, we assign a value of one for physiological needs and two for safety needs. The remaining three levels relate to belongingness, esteem and self-actualisation needs; since all three represent intrinsic and self-determined actions (Barbuto Jr and Scholl, 1998; McClelland, 1987), we score each equally with respect to intrinsic motivation. We assign behavioural events associated with these levels a value of four (i.e. a mean derived from the values three, four and five if one simply incremented the score as one progresses up the hierarchy).

However, the multiplicity of behaviours introduced by this work complicates the mapping process. To identify behavioural events that correspond to Maslow's levels, we rely on the semantic classes of the extracted behavioural events. We rely on popular category providers to identify different semantic classes of events (e.g. art gallery, casino, mosque). More specifically, we rely on categories from Foursquare for categorising events of mobility behaviour, from GooglePlay for events of phone usage, and from both Google and Amazon for buying-related events. Each provider has its own taxonomy that suits the nature of services it provides. For instance, social networking category relates to events of phone usage behaviour and does not match a category under Foursquare taxonomy. On the other hand, Foursquare uses outdoor and recreation (which is not used by Google) to classify events of mobility behaviour. Another aspect that complicates the integration process relates to the providers' interpretation of categories. Using the same example of Google and Foursquare, libraries is a class that is used by Google to describe software libraries and technical demos. In contrast, libraries are reading places as categorised by Foursquare.

Therefore, we extend the method proposed by Ibrahim et al. (2021a) to account for the issues arise from the multiplicity of behaviours and taxonomies. Guided by motivation-based taxonomies of behaviour (Tinsley and Eldredge, 1995; Barbuto Jr and Scholl, 1998; Talevich et al., 2017), we first merge taxonomies of multiple service providers to synthesise and unify the classes based on the motivation-based interpretations. Then, we match and integrate semantic classes to a corresponding generalised category (e.g. art, games/gambling, spiritual) and locate the category within Maslow's hierarchy. Each behavioural event captured in a participant's dataset can then be mapped to semantic class (as determined by the corresponding categories, Foursquare for locations, GooglePlay for apps, and Google and Amazon for notifications), and then from semantic class to category. The resulting category determines hierarchy level and associated intrinsic motivation score.

The described classification should accurately distinguish between intrinsically- and extrinsically-motivated activities in the majority of cases (Tinsley and Eldredge, 1995; Barbuto Jr and Scholl, 1998; Talevich et al., 2017). However, conducting events that would score highly for intrinsic motivation may occur for reasons other than personal interest (e.g. visit a place to accompany a friend). This limitation is addressed through the addition of dynamic modelling.

6.3.3.2 Dynamic Modelling

The naturally-occurring variation in motivation (Fogg, 2012) are studied based on competence, autonomy and novelty; psychological indicators of intrinsic motivation. Unlike static modelling, these measurements are instantiated from the participants' data and hence actual values vary according to the behaviour exhibited by each individual. Similar to Ibrahim et al., participants' data is subdivided into week-long analysis windows (Monday-Sunday) as people mostly shape their behaviour around weekdays (Cho et al., 2011; Sarker et al., 2019). However, the introduction of multiple behaviours forces significant changes in the dynamic modelling as we shall see next.

Competence Competence is predicted from both frequency and duration (Fishbach and Hofmann, 2015; Nicholls, 1984; Wolf and Hopko, 2008; Rabbi et al., 2015; Ryan and Deci, 2017). Both metrics are integrated into a measure of *intensity*. We propose an integration method that is based on the extracted behavioural events and considers the multiplicity of behaviours. Specifically, we count, for each event of each behaviour, the total number of occurrences per week and multiply that by the weekly average duration.

Formally, if F_{wb_x} is the weekly frequency of a behavioural event x of a behaviour b, and M_{wb_x} is the weekly average duration, then the intensity of x up to the current last week of an ongoing study t is computed as:

$$Intensity_x = \sum_{b=1}^{h} \sum_{w=1}^{t} F_{wb_x} * M_{wb_x} \tag{6.1}$$

where h is the number of behaviour (three in our case).

Autonomy People display greater sustainability toward the internalised behaviours (Ryan and Deci, 2017). We express this as the *sustainability* of behaviours. Similar to how we adjusted intensity, we model sustainability based on the multiple behaviours studied by this paper. To generate a measure of *sustainability*, we assign a Boolean value that indicates whether a specific behavioural event occurs within the analysis window. We then count the number of subsequent windows (i.e. weeks) in which a behavioural event is observed and divide that by the total number of subsequent analysis windows. The closer is the result to 1, the more sustained the behaviour.

Formally, if N_{wb_x} denotes the existence of a behavioural event x of a behaviour b in week w, and D_w represents the current number of weeks in a study, then the sustainability score is computed as:

$$Sustainability_x = \sum_{b=1}^{h} \sum_{w=1}^{t} N_{wb_x} / D_w \tag{6.2}$$

where t represents the current last week of an ongoing study.

Novelty To account for interest dynamics, novelty, through the measure of *recency* is used. Similar to Ibrahim et al. (2021a)'s work, we segment the entire study duration into periods based on a predefined threshold (i.e. a static approach[4]). In this work, however, we weight each event of the multiple behaviours according to their occurrence within each period, with the most recent period accumulating the highest value; values associated with prior periods gradually decline as their distance from the present period increases. This gradual retrospective degradation is crucial because although people shift their attention to recent behaviours, they are expected to revisit previous behaviours rather than entirely abandoning them (Zhao et al., 2013).

6.3.3.3 Integration

We integrate the three behavioural traces used by this work as well as our static and dynamic models to compute an overall score for any given behaviour. The resulting *cMIR Score* indicates the degree to which the specified behaviour is intrinsically motivated. To integrate the behavioural traces, we weigh them differently for each participant based on their data coverage to account for missing days of data collection. More specifically, if the collected locations, apps' interactions and notifications data of a participant cover respectively 90%, 70% and 40% of the study period, the score for each trace is weighted based on those percentages and is combined to form the cMIR

[4]See Srinivasan et al. (2014) and Sarker et al. (2019) for examples of static and dynamic recency thresholds respectively.

score. Formally, the combinatory MIR (cMIR) score is calculated as:

$$\text{cMIR score} = \sum_{i=1}^{p} \frac{1}{i} \sum_{b=1}^{h} (needs + g_b((\underbrace{F_{wb_x} * M_{wb_x}}_{\text{intensity}}) + (\underbrace{N_{wb_x}/D_w}_{\text{sustainability}}))) \tag{6.3}$$

where *needs* is determined as specified in Section 6.3.3.1, and *intensity* and *sustainability* are determined separately for each behaviour before being adjusted to account for *recency* and personalisation (i.e. data coverage). The weight of a behavioural trace, g_b, is determined based on the data coverage as described, and h is the number of traces used to recognise interests (three in our case). Thus, both *intensity* and *sustainability* are personalised and calculated for each period before being weighted and summed as described in Section 6.3.3.2. The resulting formalisation operates over a set of periods p, such that i ranges from 1 to p where 1 designates the most recent period, and p represents the outmoded interval. So, if a person visits sport-related places, uses sports apps, and buys sport-related goods in a sustained, intense, and renewed manner, then the person is likely to be interested in the sport.

Finally, the resulting cMIR score can be used to assemble a ranked list of IMBs as follows:

$$Behaviour \times \text{cMIR Score} \rightarrow Ratings. \tag{6.4}$$

where the ratings for each user is predicted as a function of the motivation properties and the performed actions. For example, if the behaviour is related to sport, then the rating score would indicate the degree to which a sport-related behaviour is intrinsically motivated (for a given individual)

Figure 6.3 is adjusted from (Ibrahim et al., 2021a) to provide an end-to-end view of our approach, whereby raw smartphone's data is ultimately transformed into a ranked list of IMBs. Passively-sensed data is semantically enriched to produce a set of behavioural events that correspond to life events such as dining out, and shopping. Extracted events are used as input to the aforementioned models of motivation properties (determined based on the indicator identification) to rank the behavioural events based on the underlying motivation properties. As a result, we produce a ranked list of motivated behaviour that represents the behaviour's place on the proposed motivation continuum.

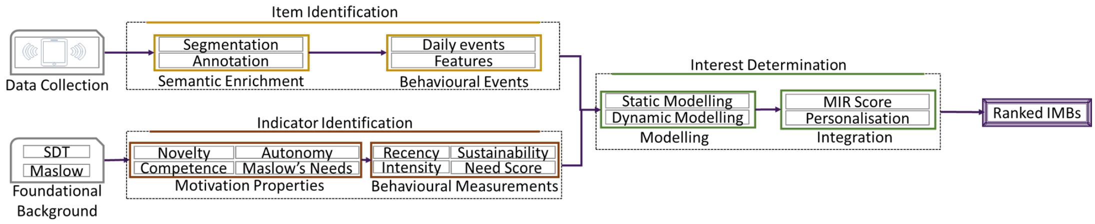


Figure 6.3: The overall process of extracting and ranking IMBs from passively sensed smartphone's data. Adapted from Ibrahim et al. (2021a).

6.4 Evaluation

Our evaluation includes longitudinal mobile data from eight participants going about their normal daily activities. We use the AWARE mobile sensing framework (Ferreira et al., 2015) to collect the data. Location readings are collected at one minute intervals, notifications are recorded once they are received and app usage data are captured each time a user interacts with the phone. Procedures for the study were reviewed and approved by the Department of Computer Science Ethics Committee at The University of Manchester (Reference: 2019-7817-12726).

6.4.1 Study Duration and Participants

We select a study duration of six months to capture a richer picture of the natural changes in the pursuit of IMBs over time (including, e.g. seasonal change). The long time allows us to examine the impact of the recency factor on IMB recognition as multiple periods will be formed. This is in direct contrast with the three-month study of Ibrahim et al. (2021a) which consists of one period only. In this work, we segment the study duration into two three-month periods (guided by previous studies (Srinivasan et al., 2014; Zhao et al., 2013; Song et al., 2010; Ibrahim et al., 2021a)) and apply equation (6.3) accordingly.

For the sample size, we use the methods described in (Barnett et al., 2020; Ibrahim et al., 2021a) to determine an appropriate sample size. The used methods (Barnett et al., 2020; Ibrahim et al., 2021a) support the arguments of smaller samples given an appropriately long study duration and the positive correlations between behavioural measurements, shifting the focus towards data corpus that reflects real-world behaviour and spans over a long period of time. Accordingly, we need at least seven participants to achieve 80% power and 95% confidence level. Upon conclusion, our study data consists of ~0.8 million, which is sufficient for meaningful analyses.

Participants were recruited using poster advertisements displayed in public areas of The University of Manchester and surrounding buildings, and on social media. Due to COVID-19 restrictions enforced in United Kingdom in early 2020, the recruitment process has been impacted, and we had to only rely on online advertisement. A total of nine participants ultimately agreed to participate, all but one of whom were students (four undergraduates, four postgraduates). Participants were rewarded for their participation with six months of Netflix subscription or equivalent Amazon voucher.

Participants enrolled on the study over a staggered period based on when they chose to respond to recruitment advertisements. The first participant (P1) began data collection on January 1, 2021, and the final participant (P9) began on April 1, 2021. One participant (P9) was excluded after one month because of technical difficulties related to the app and phone compatibility. Two participants (P6 and P8) had withdrawn from the study after three months. Accordingly, the collection period ranges from 90 days for P6 and P8 to 180 days for the remaining participants.

6.4.2 Method

Participants were supplied with an information sheet prior to participation and had the opportunity to ask further questions prior to consent. All participants provided written consent before being guided to install the AWARE app on their personal mobile devices. Participants were asked to keep the installed application running and to carry their phones as they normally do. They were advised that the application would automatically send their data to our backend server, but only when connected via WiFi, and that we anticipate no noticeable negative effects on battery life.

During the initial session, participants also engaged in a short audio-recorded interview in which they were asked a set of general questions designed to act as a supplementary source of ground truth. For example: what are the places they visit because they want to rather than they have to? What motivates them to practice their behaviours? How often do they practice them?

As noted in Section 6.4.1, participants were asked to run the AWARE app for six months, during which all data was synced to our backend server for offline analysis. This includes the raw locations, notifications, and app usage. All data captured within the app was sent to a secure server at The University of Manchester.

To collect ground truth, we send a monthly questionnaire that presents an adaptation of the Interest/Enjoyment subscale of the widely-used Intrinsic Motivation Inventory (IMI). IMI is a multidimensional scale developed by (Ryan, 2018) and is widely used to assess intrinsic motivation associated with a given activity (Monteiro et al., 2015; Ryan, 2018). Seven IMI prompts (the complete Interest/Enjoyment subscale) were delivered monthly; each takes the form of a statement with an associated seven-point Likert-like scale (1=not very true, 7=very true). Statements were adapted from the original IMI (Ryan, 2018) in order to incorporate a specific IMB into the phrasing, e.g. from "I enjoyed doing this activity very much" to "I enjoy shopping very much"[5]. All seven statements reference the same IMB. The resultant data is used as a ground truth for the target IMB.

To select the top IMBs, we mirror the method used by Ibrahim et al. (2021a), in which the data shows that clear differences in the strengths of consecutively ranked IMBs naturally emerged at different points for each participant (Ibrahim et al., 2021a). Therefore, we establish a cut-off point based on the largest difference between the top five consecutive behaviours. In subsequent analysis, we compare the validity of this dynamic N with fixed values. In some cases, the N value is equal to one, suggesting that the participant's behaviour reflects a single interest much more strongly than any others. Whilst in some applications, identifying the strongest intrinsic motivator would be sufficient, we suggest that in many cases, a broader understanding of IMB would be beneficial. Thus, for each case where the N would be 1, we also identify the next largest dynamic N using the approach previously described. We do so as long as the cMIR value is above 0.60[6] to preserve only the strongest interests. The resultant value, N^{Dyn}, is used in subsequent analyses.

At the end of the study, participants again participated in a short audio-recorded interview. Participants were asked about all specific IMBs recognised in the dataset and were additionally asked to identify personal interests that they felt had been omitted.

6.4.3 Dataset

After excluding the participant (P9), the number of entries in our dataset has become $782,875$. The mean and median number of entries per participant are 97859.38 and

[5]The full question set is presented in Appendix C.5.

[6]When normalising the IMI score between 0 and 1, the threshold of 0.60 corresponds to 4, the threshold used to determine an IMB from an IMI response.

Table 6.1: The total number of readings per source per participant: (Locations), (Apps), and (Notifications), collection duration in days (Duration). ($\bar{x}$) is the calculated sample mean.

	Location	Apps	Notifications	Duration
P1	55,810	66,5513	53,861	180
P2	392	19,092	120,156	180
P3	8,092	117,415	45,982	180
P4	19,175	1,269	0.00	180
P5	115,176	4,923	0.00	180
P6	15,937	55,335	6,760	90
P7	16,006	19,524	14,470	175
P8	22,226	6,031	6,183	90
x̄	**32758.50**	**39740.75**	**35334.13**	**156.88**

99065.50, respectively. The largest number of entries collected from a single participant was from P1 (176,222 records), whereas P4 has the least contribution of (20,444 entries). Table 4.1 details the entries per source per participant.

6.4.3.1 Location data

The location records are tuples comprised of latitude, longitude and timestamp. We use the algorithm proposed in (Li et al., 2008) to extract stay-points: data points are processed sequentially, with stay-points determined in accordance with predefined time and distance thresholds. Guided by (Boytsov et al., 2012), we set our time threshold at 15 minutes. We fix our distance threshold at 100 meters (Solomon et al., 2018), meaning that GPS readings within a 100-meter circumference are considered to be the same stay-point.

Since IMBs conducted inside the home are impossible to identify using GPS location alone, we seek to exclude the participant's likely residence from our analysis. Using semantic annotation for this purpose (i.e. looking for places in a residences category) was often inconclusive. Thus, we instead identify the single location in which the participant spent the most time in any given week and exclude it as the participant's likely place of residence. Excluding locations on a weekly basis should account for temporary accommodation such as vacation or business trips.

6.4.3.2 Phone usage data

Phone usage data are recorded as entries containing timestamp and package name. Since smartphones allow interactions with one app at a time, the package name is used as the base for identifying interaction events. Specifically, consecutive interactions with the same app forms a phone usage event.

The extracted events are then annotated based on the categories retrieved by GooglePlay. If the app's category is not found, we label the category as "Unknown" and exclude them from the analysis. Typically, Unknown categories result from installing apps that are not listed on GooglePlay.

6.4.3.3 Notification data

Each entry of the dataset contains the notification time, title, app, and text. Similar to the phone usage behaviour, we use GooglePlay to retrieve the category of the app issuing the notification. Also, we anonymise numbers and emails contained in the notification text to preserve the participant's privacy. Consequently, a notification that says "You received 2 emails from john@example.com", is stored as "You received * emails from ****". This step is essential and needed to comply with the university's ethics requirements.

We rely on the text field of each notification entry to extract and analyse data related to buying behaviour. We first remove notifications with empty text or those that only contain symbols and special characters. The latter mainly results from the non-English text that is not correctly encoded. The remaining notifications are then classified as either related or unrelated to buying. A notification is buying-related if (i) it is issued by an app categorised as *"Shopping"* or *"Food_AND_Drink"*, or (ii) it contains one of the following buying-related keywords: *"your package"*, *"your order"*, *"your payment"*, and *"your purchase"*. The Selection of these keywords was based on analysing emails from Enron email dataset, a public dataset that contains emails from more than 100 users (Klimt and Yang, 2004). Specifically, we randomly selected 100 emails sent by retailers such as "Amazon" and searched for receipts, order confirmations and shipping notices. Based on that, the keywords were identified. Lastly, Google Cloud Natural Language API and Amazon are used to annotate the notification text as described in section 6.3.2. Lastly, guided by (Anesbury et al., 2016; Hui et al., 2013), we estimated the duration per buying event at 15 minutes. This estimation is mainly based on online buying as it is mostly used under COVID-19 restrictions.

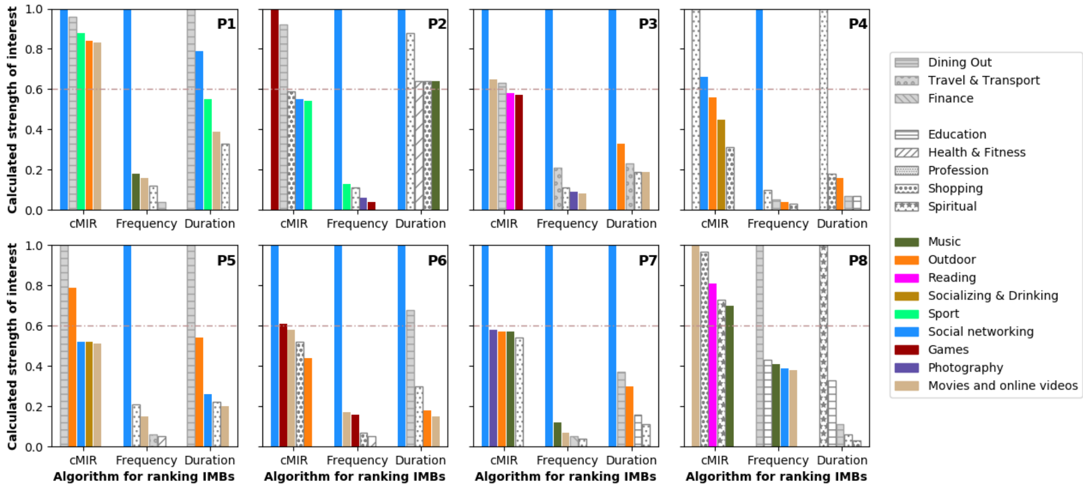


Figure 6.4: IMBs for each participant as determined by our cMIR algorithm and by frequency-based and duration-based approaches for comparison. The dashed line shows the 0.60 threshold that we use with N^{Dyn} selection.

6.4.4 Results

Figure 6.4 shows an ordered (left-to-right) set of the top five ranked IMBs for each of our eight participants[7]. We also plot comparison results from two alternative algorithms: one frequency-based (used in e.g. (Musto et al., 2018), and (Liu et al., 2016)) and one duration-based (used in e.g. (Lim et al., 2015) and (Gaonkar et al., 2018))[8]. cMIR selects the same top-rated interest as both other algorithms for three of the participants (P3, P6, P7). Agreement on the top three IMBs for cMIR and at least one of the two comparison algorithms (ignoring ordering) occurs in three cases (two of them with the duration-based and one with the frequency-based method). However, the two comparison algorithms themselves have no agreement on the selected top three. Similarly, there is no agreement on the top five IMBs between neither cMIR and any of the two comparison algorithms (ignoring ordering) nor the two comparison algorithms themselves.

Selecting IMBs dynamically results in a N^{Dyn} between 1 and 3 for cMIR (mean: 2.00, median: 2.00, std: 0.53, iqr: $2.00 - 2.00 = 0.00$) with lower N^{Dyn} values for the two alternate algorithms (frequency mean: 1.00, median: 1.00, std: 0.00, iqr:

[7] We adopted the styling of figures and tables used by Ibrahim et al. (2021a) to ease the readability and comparison between the two studies.

[8] Similar to cMIR, scores for the two alternative approaches are calculated based on the combination of behavioural traces unless explicitly specified (e.g. "frequency (mobility)").

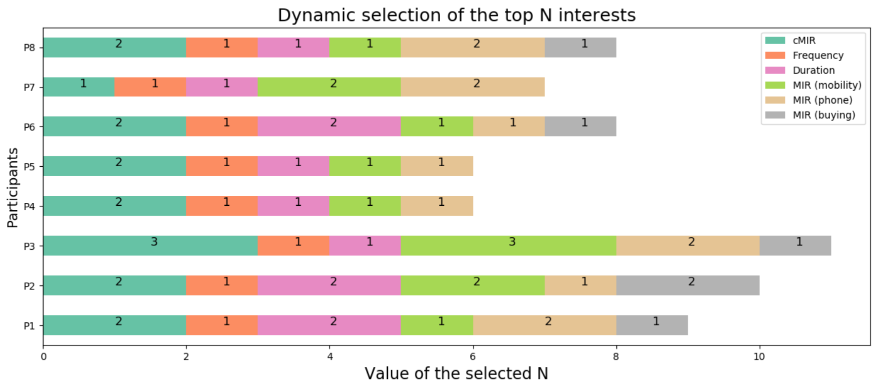


Figure 6.5: The selected N^{Dyn} for each participant based on each interest recognition method.

$1.00-1.00=0.00$; duration mean: 1.38, median: 1.00, std: 0.52, iqr: $2.00-1.00=1.00$). The mean size of the difference used as a cut off for N^{Dyn} is 0.18 (cMIR), 0.81 (frequency), 0.51 (duration). Overall intersection of dynamic ranked list of participant IMBs (N^{Dyn}) is 12.50% for cMIR/frequency, 37.50% for cMIR/duration, and 62.50% for frequency/duration. The dynamic N selection for the base MIR (i.e. using a single behavioural trace) results in a N^{Dyn} between 1 and 3 when using MIR with mobility behaviour, and between 1 and 2 for both MIR (phone) and MIR (buying). Figure 6.5 details the N^{Dyn} per method for each participant.

Examining the top-ranked interests themselves (using N^{Dyn}), we find that around 60% (min:50%, max:100% mean: 62.50%) match a Maslow needs level of four (intrinsic). However, a significant minority correspond to Maslow's physiological (mean: 25.00%) and safety needs (mean: 12.50%). Whilst one could interpret this as indicating that our approach may struggle to filter out extrinsically-motivated behaviours, we note that in reality many of these kinds of activity can be intrinsically motivated (e.g. choosing to engage in shopping or fitness activities because they are enjoyable or align with personal values rather than out of necessity).

Our study provides a ground truth for the final set of IMBs, with an IMI score generated for each ranked IMB (N=1... N^{Dyn}). These scores can range from 1 (no interest/enjoyment) to 7 (very high interest/enjoyment) and are an aggregation of participants responses to the seven scale items listed in Appendix C.5. We detail the resulting values for the top-ranked interests (using N^{Dyn}) in Table 6.2. If the IMI score equals or falls below four, then the behaviour is considered to be a false positive. Overall during the study period, we have collected 64 IMI responses from all the participants. Of those 64 responses, 51 were identified as interests according to the reported scores.

Table 6.2: Participant IMI responses for each of the interests identified by the cMIR algorithm.

	Reported interests	cMIR	IMI
P1^s	Social networking	1.00	4.43
	Dining out	0.96	4.71
P2^s	Games	1.00	5.71
	Dining out	0.92	6.00
P3^s	Social networking	1.00	5.43
	Movies and online videos	0.65	6.86
	Dining out	0.63	6.43
P4^s	Profession	1.00	4.14
	Social networking	0.66	3.14
P5^s	Dining out	1.00	6.00
	Outdoor and recreation	0.79	6.29
P6^s	Social networking	1.00	6.00
	Games	0.61	5.86
P7^s	Social networking	1.00	5.29
P8	Movies and online videos	1.00	6.00
	Shopping	0.97	5.14

The mean and median number of interests per participant is 6.38 and 6.00 respectively (std: 2.56, iqr: $8.25 - 5.00 = 3.25$).

We determine the final ground truth where true interests include those items with an IMI score greater than four. Based on this, we calculate the precision and recall values reported in Table 6.3. cMIR outperforms both duration and frequency algorithms on precision and recall. Also, cMIR outperforms the base MIR with a single behavioural trace (mobility, phone usage, and buying). For precision, however, cMIR shows no significant difference between the used methods ($Q = 3.14$, $p = 0.68$). In contrast, the cMIR recall is significantly better as a Friedman test shows a significant difference between the algorithms ($Q = 17.94$, $p < 0.01$). Recall values range from 0.11 for the MIR based on buying behaviour (single trace) and 0.34 for the cMIR. The former is impacted by the lack of buying-related activities in the collected data for P4, P5, and P7. While no notification data were collected from P4 and P5, P7 confirmed in the final interview that he does not trust online shopping and hence has never used it.

Table 6.3: Precision (P) and Recall (R) values for the cMIR, Frequency-based, and Duration-based algorithms, as well as the MIR methods based on a single behavioural trace (mobility, phone usage and buying behaviours). Values are based on the dynamic selection of the top-ranked interests (N^{Dyn}) generated by each at the end of the data collection period and are calculated based on IMI participant ground truths.

Participant	**cMIR**		**Frequency**		**Duration**		**MIR - mobility**		**MIR - phone**		**MIR - buying**	
	P	R	P	R	P	R	P	R	P	R	P	R
P1^{s}	1.00	0.22	1.00	0.11	1.00	0.22	1.00	0.11	1.00	0.22	1.00	0.11
P2^{s}	1.00	0.33	1.00	0.17	0.50	0.17	0.50	0.17	1.00	0.17	1.00	0.17
P3^{s}	1.00	0.50	1.00	0.17	1.00	0.17	0.67	0.33	1.00	0.33	1.00	0.17
P4^{s}	0.50	0.50	0.00	0.00	1.00	0.50	1.00	0.50	0.00	0.00	0.00	0.00
P5^{s}	1.00	0.20	1.00	0.20	1.00	0.10	1.00	0.10	1.00	0.10	0.00	0.00
P6^{s}	1.00	0.40	1.00	0.20	0.50	0.20	0.00	0.00	1.00	0.20	1.00	0.20
P7^{s}	1.00	0.13	1.00	0.13	1.00	0.13	1.00	0.25	0.50	0.13	0.00	0.00
P8^{s}	1.00	0.40	0.00	0.00	1.00	0.20	1.00	0.20	0.50	0.20	1.00	0.20
x̄	**0.94**	**0.34**	**0.75**	**0.15**	**0.88**	**0.21**	**0.77**	**0.21**	**0.75**	**0.17**	**0.63**	**0.11**

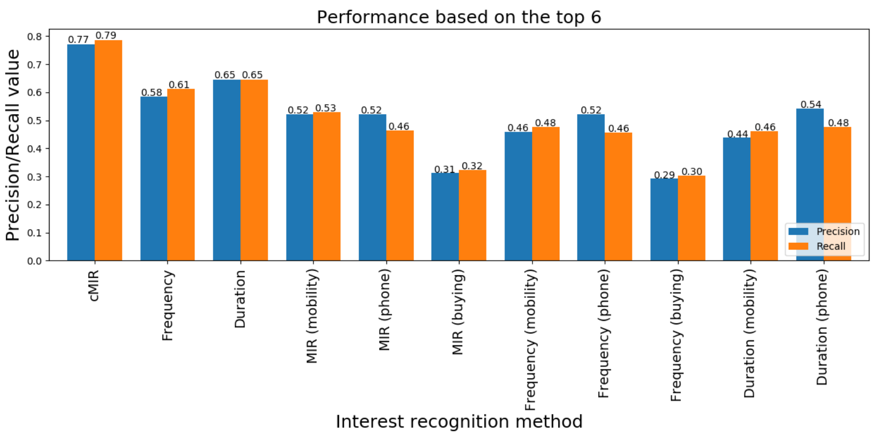


Figure 6.6: Precision and recall for the top six IMBs as determined by each one of the interest recognition methods. cMIR, frequency and duration are based on multiple behavioural traces, whereas the remaining are based on a single behavioural trace. Duration based on buying behaviour (i.e. duration (buying)) is not shown as the duration is not captured as discussed in Section 6.3.3.2.

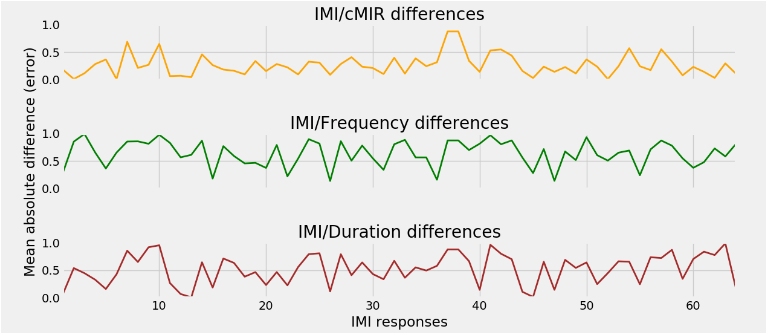


Figure 6.7: Differences between each of the methods that score behaviour based on multiple traces (i.e. cMIR, frequency and duration) and the reported IMI scores (64 responses).

As noted from the range of recall values, the dynamic selection of the top N produced a conservative set of interests (mean and median of 2.00 compared to 6.38 and 6.00 true interests per participant). Therefore and based on the mean and median of the true interests, we examined the impact of selecting the top six on the precision and recall scores. Figure 6.6 shows the results of doing so on all methods based on single and multiple behavioural traces. cMIR, frequency and duration are based on multiple behavioural traces, whereas the remaining are single behavioural traces. The results show that cMIR outperforms all alternatives. However, unlike the dynamic N selection, both precision and recall of the cMIR method are significantly better than all other alternatives ($Q = 18.82$ and $p < 0.01$ for both precision and recall). This shows a mean improvement of 62% (precision: 59% and recall: 66%) when comparing the cMIR to all methods and a mean improvement of 26% (precision: 26% and recall: 27%) when comparing the cMIR to the combinatory-based frequency and duration methods.

Based on the above, we examined the overall difference between the scores produced by the interest recognition methods and the 64 reported IMI scores. Specifically, we compare scores of cMIR, frequency and duration methods with the IMI scores in Figure 6.7. The values of the 64 IMI responses were first normalised to be in the same scale (between 0 and 1) and then the differences were calculated. The figure shows that cMIR is significantly a better *predictor* of the IMI responses when compared to alternatives ($p < 0.01$). The mean absolute difference (error) between the cMIR and IMI scores is 0.26 compared to 0.64 and 0.52 for frequency and duration respectively.

6.4.5 Discussion

Interest recognition algorithms used in this work typically agree on the top-ranked IMB but vary in their selection of lower-ranked interests. When selecting the top N dynamically, the precision scores are high for all methods (cMIR outperforms the others with no statistical significance). On the other side, cMIR significantly performs strongest of the remaining for recall ($p < 0.01$), although their overall scores are low. One reason behind both the high precision and low recall values is the small number of interests produced by selecting the top N dynamically. The mean number of interests retrieved by the dynamic approach using our cMIR (2.00) is substantially lower than the mean of the reported interests per participant (6.38). This caused a conservative selection of the top N and hence higher precision and lower recall scores. As we increase the number of the top interests, the significant difference between the methods starts to emerge. Selecting the top six (to match the median and mean of the true reported interests) improved the recall of the methods but showed that cMIR significantly outperforms the others. Similarly, cMIR outperforms on the precision side significantly, although the scores of all approaches have decreased. Moreover, the values differences between the reported interests (based on the IMI scale) and the interest recognition methods (Figure 6.7) confirms that increasing the number of the top N would improve the performance of cMIR compared to others if more IMBs are selected. These results indicate that cMIR outperforms other algorithms, particularly in identifying postliminary IMBs.

The combinatory approach has also improved the recognition of IMBs when compared to the reliance on a single behavioural trace. This is true for the cMIR method as well as the frequency- and duration-based when compared to alternatives based on a single trace. A deeper analysis of the data showed that the precision and recall values for MIR (mobility), MIR (phone) and MIR (buying) do not significantly differ from their corresponding frequency- and duration-based methods. MIR (the measure for a single behaviour) performs, on average, similar to or slightly better than their single trace correspondings on precision and recall except for phone usage. As shown in Figure 6.6, recognising interests from phone usage based on duration (i.e. duration (phone)) produces precision and recall values that are slightly better (precision: 0.54, recall: 0.48 compared to 0.52 and 0.46 for MIR (phone)). This could be the result of smartphones activities being mostly intrinsically motivated in nature. People are not expected to spend a long time on smartphones doing extrinsically motivated behaviours. Instead, they are expected to use a desktop or laptop if they have online activities that are extrinsically motivated and require a long time (Steeds et al., 2021).

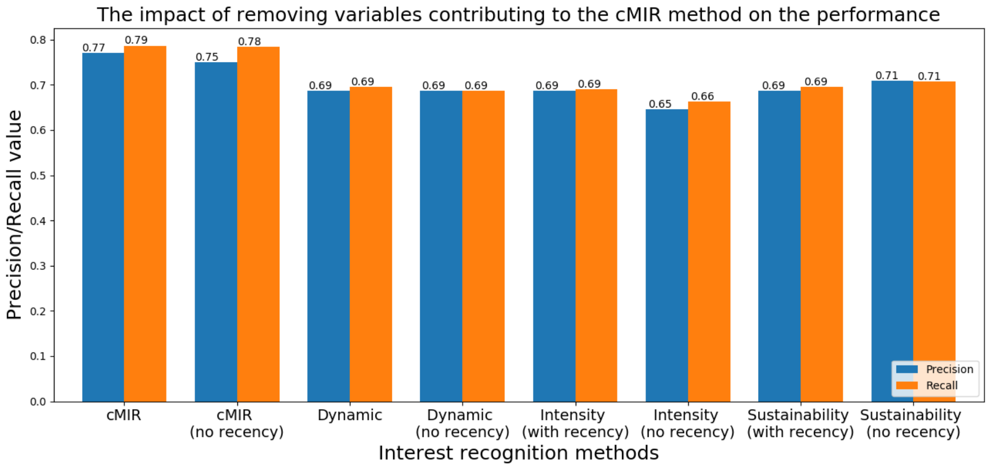


Figure 6.8: The precision and recall values for various variations of interest recognition methods that are based on our dynamic and static modelling.

Nonetheless, when integrating the three behaviours, we note the significance of our method (cMIR) as the precision and recall outperform their corresponding frequency and duration methods as well as the ones relying on a single behavioural trace. This shows the importance of integrating multiple behaviours in capturing interests that are exhibited differently.

With respect to the ranked list itself, our results show that 44% of the identified IMBs are related to safety and physiological needs according to Maslow's hierarchy. Although every participant has at least one in the top five IMBs, our closing interviews suggest that IMBs of lower Maslow's needs may indicate actual interests. During interviews, our participants reported pursuit of these supposedly extrinsically-motivate behaviours for both intrinsic and extrinsic reasons – thus, the combination of needs with our other measures is essential to appropriately produce accurate cMIR scores. Comparison with other approaches lends credence to our approach. For instance, the top five IMBs of P1 contains outdoor and recreation activities. Frequency and duration failed to identify outdoor as one of the top five or at least to be above the threshold of 0.6 (ranked tenth by frequency and sixth by duration). The persistence of behaviours (an autonomy indicator under SDT) associated with outdoor, combined with their higher needs level, positively impact their rankings under our approach. The IMI score and participant's interview confirmed our findings.

To investigate the interplay between the static and dynamic parts and the impact on the performance of our approach, we depict in Figure 6.8 the results of only relying on the dynamic modelling for identifying interests. The figure shows that the scores of both precision and recall have been negatively impacted. The removal of the static modelling has significantly decreased the scores by a mean of 13% (precision:12%, recall:14% and $p < 0.05$ for both). In the same figure, we also depict the impact of removing each one of the variables that form the cMIR equation. The figure shows that removing any of the variables would degrade the performance of the cMIR algorithm significantly. The performance degradation is statistically significant ($p < 0.05$) when we rely on intensity, intensity with recency or sustainability with recency for recognising interest (see Figure 6.8). When relying on sustainability alone, only the recall is significantly impacted, whereas removing the recency does not significantly impact the results, although it slightly degrades the performance. Nonetheless, these results show the importance of considering all the components of our method when detecting interests.

In our closing interview, we also discussed with the participants the impact of COVID-19 on their behaviours generally and on practising their interests specifically. Based on the participant's country of residence, the impact on the behaviour (mainly mobility) differ. On average, restrictions on mobility behaviour were eased toward the third month of participation. As these restrictions were lifted, additional interests started to emerge, such as outdoor for P1 and movies for P3. The presence of the recency factor helps signify the emergent behaviours and avoid shadowing them by currently abandoned ones. Capturing buying behaviour has also been negatively impacted by the COVID-19 pandemic. As stay-at-home policy likely reduced visits to locations associated with higher-level needs, travel restrictions of COVID-19 have negatively impacted the buying behaviour of some participants who returned to their home countries. According to them, their use of online shopping has significantly decreased as it is not a preferred and reliable method in their countries. However, one additional benefit of the behaviour multiplicity is its tolerance to similar issues. If buying behaviour is impacted, mobility and phone usage can mitigate that and help to identify interests.

Also, with respect to buying behaviour, different methods of determining the duration of buying events may improve the modelling of those events. In this work, we set the duration at 15 minutes. However, this duration may increase or decrease based on various factors. For instance, repeated buying from the same source may decrease this duration as the person becomes more familiar with and better at completing the transaction. Also, buying expensive products may require more time as the person may think carefully and take additional time comparing different options before completing a transaction. Therefore, future work may propose a more flexible method of determining the suitable duration of buying events.

Although our motivation properties and behavioural measures should be applicable beyond the three behaviours, our study showed that the semantic annotation upon which our measures are based has a considerable impact on the results. Specifically, item identification (Section 6.3.2) is heavily impacted and limited by the accuracy of the external annotator. In previous work, (Ibrahim et al., 2021a), the accuracy of annotating stay-point is discussed and showed to be above 90% using Foursquare. We have investigated the accuracy of extracting buying features from the notification text in a separate work that is currently under review. However, the accuracy of annotating the bought items extracted from the text is yet to be investigated and beyond the scope of this work. Additional improvements are expected to have a similar impact on other methods. The phone usage behaviour is less impacted as it depends entirely on the device (mobility and buying behaviours themselves are smartphone-independent, although they can be captured).

Our study demonstrates considerable individual variation, particularly with regard to the execution of activities that correspond to the lower levels of Maslow's hierarchy. This was predicted by prior literature that suggests that while all individuals share the same fundamental needs, their pursuit of them may vary (e.g. based on the degree to which they have previously had that need fulfilled) (Deci and Ryan, 2000). Thus, the personalisation step in our overall approach (Figure 6.3) can be extended in future to consider the degree to which coefficients within cMIR can be personalised in order to deliver a more tailored IMB calculation. This is in addition to how it is used in this work to weigh each behaviour according to the individual's data.

Lastly, there are two interesting ways of generating the top N results, and we explore both of them. In this study, we report the results using dynamic N and static N. However, we do not suggest at this point that either the dynamic or static approach is better than the other or a specific approach works best for a particular population. Instead, these deeper investigations are still worthy of further exploration.

6.5 Conclusion

In this study, we propose a combinatory approach that uses multiple smartphone-derived behaviours and intrinsic motivation to recognise personal interests. We extend the work of Ibrahim et al. (2021a) and integrate events of mobility, phone usage, and buying behaviours to identify Intrinsically Motivated Behaviours (IMBs), which reflect personal interests. Specifically, a core set of static and dynamic motivation properties (needs, competence, autonomy and novelty) are identified as indicators of IMBs and modelled based on behavioural measures derived from the literature. Then, we extract events of mobility, phone usage, and buying behaviours from raw smartphone data that are passively sensed. The extracted events are used as the basis for the analysis. We combine the behavioural events and the modelled indicators in a combinatory MIR (cMIR) score whose value (0-1) represents the strength of intrinsic motivation associated with multiple behaviours.

Through a real-world study, we show how our approach can facilitate personalised understanding of IMB compared to frequency- and duration-based approaches as well as the MIR method proposed by Ibrahim et al. (2021a). Our results indicate that our approach successfully identifies IMBs that are consistent with those reported by participants, outperforming alternatives. Results also suggest that most IMBs can be detected and adapted to variations in behavioural patterns as users' motivation changes.

Relying on a single data indicator in behavioural analysis is inherently risky (as acknowledged by Ibrahim et al. (2021a), and mobility restrictions imposed in response to COVID-19 have provided a clear demonstration of this vulnerability. Our approach addresses that through the integration of multiple behaviours and personalising the weight of each behaviour based on the individual's data. However, future work can propose and compare different ways of integrating those behaviours. Moreover and to further improve personalisation, we suggest that variation in the weighting of different motivation properties could allow the model to better reflect difference in individuals' need satisfaction (i.e. the degree to which they pursue a particular need).

Personalised IMB identification of the kind enabled by our work has the potential to facilitate new applications that capitalise on individuals' intrinsic motivation. This could be particularly valuable for behaviour change, where existing evidence indicates leveraging intrinsic motivation leads to more effective and sustained change (Ryan and Deci, 2017). For example, we envisage fitness applications that prompt users to take a slightly longer journey home by plotting a route that is consistent with personal interests (e.g. passing by a local soccer club for a user for whom soccer is considered an IMB).

Chapter 7

Behaviour and Interest Recognition Tool

In this chapter, we introduce the web-based tool that we use to implement our MIR approach on mobility and phone usage behaviours. The tool focuses on the two sources that require sequential processing of the data. Details about notifications (which rely on the analysis of textual data are explained in Chapter 4). The tool extracts behavioural events from raw smartphone data (detailed in chapters 2, 3, and 4) and implements our MIR method accordingly (chapters 5 and 6). Through exemplar case studies, the tool has been shown as an effective tool that can be used to (i) recognise behaviour from raw data and (ii) support the modelling of motivation properties necessary to understand behaviours driven by personal interests.

The main content of this chapter is a paper authored by: *Ahmed Ibrahim, Sarah Clinch and Simon Harper*. The title of the paper is: ***Smartphone Data Analytics: A Behaviour and Motivation Centric Implementation***. The paper is currently under review. For this thesis, we edited some formatting styles, such as the sizes of some tables for consistency and readability reasons. Also, footnotes 1, 2, and 3 on pages 207, 211 and 214 respectively refer to the paper that we included in Chapter 5. As this paper is under review, we anonymise that in the submitted version to comply with the reviewing requirements.

Author contribution

Ahmed Ibrahim designed and developed the tool, and wrote the paper. Sarah Clinch and Simon Harper provided continuous feedback throughout all the stages of the study, offered advice and discussion and contributed vital edits to the paper's writing.

Abstract

In this paper, we introduce a web-based analytical tool that enables users to understand behaviour based on an underlying interest when captured from smartphone data (digital phenotyping). Traditionally, users have found it challenging to differentiate between actions which are interesting to the user and those which are obligations (e.g. watching movies vs paying bills) to make a decision based on motivating interest. One reason for this is the lack of a software tool that enables users to perform the analytics. The presented tool adopts a motivation-based approach to recognise behaviours of interest that can be of value for behavioural interventions. Accordingly, users can predict whether the behaviour that an individual exhibits is externally or internally motivated. Moreover, users can generate behavioural rules personalised based on internally motivated actions (i.e. motivating interest). To show the versatility of the tool, we report three exemplar case studies that support the importance of the behaviour-centric processing of smartphone data. Moreover, they support the practical value of the tool in motivational analytics as well as in enabling personalised rules generation.

7.1 Introduction

This paper presents a Motivation-based Interest Recognition (MIR) analytic tool that uses smartphone data. The tool is based on a solid foundational theory of human motivation to help users (mainly researchers who are interested in digital phenotyping) recognise interest using smartphone-derived behaviours. Unlike obligations that are externally motivated, interests are internally motivated and performed to attain the satisfaction inherent in the underlying activities (Ryan and Deci, 2000). The MIR tool models and visualises behaviours from raw smartphone data based on motivation properties to help researchers distinguish actions that are motivated by personal interests (a.k.a Intrinsically Motivated Behaviours or IMB) from those motivated by external factors such as obligations. The distinction is significant for domains that need a reliable and deep understanding of personal interest, such as personalised behavioural intervention and nudges.

The reliability and depth of understanding interest rely heavily on the quality of the collected data (McCusker and Gunaydin, 2015). The advent and popularity of smartphones drive the development of logging tools that facilitate the acquisition of moment-by-moment personal and ecologically valid data (Miller, 2012). Those tools, however, focus either on personal and passive sensing (e.g. Miluzzo et al., 2007; Ferreira et al., 2015) or on extracting behavioural features relevant to specific application domains such as obesity (Rabbi et al., 2015). There is a lack of a tool that provides a holistic and modular view of how behaviour is first derived and then used to understand an underlying phenomenon. The way the MIR tool is designed fills that gap in addition to its facilitation of an in-depth analysis of behaviours motivated by personal interests.

The MIR tool facilitates the behavioural analysis through three processes: behaviour identification, indicator identification, and interest determination. Behaviour identification is the process of converting raw data into events. Each event corresponds to a real-life activity such as dining out and web browsing. Researchers can upload smartphone's raw data (e.g. location and smartphones' interactions data), and accordingly, behaviour-centric processing that transforms the uploaded data into events is performed. The extracted events are analysed based on properties characterising actions motivated by personal interests. These properties are extracted from the literature on human motivation following an extensive review of it (indicator identification). Researchers can recognise behaviours motivated by personal interests through the application of motivation properties on the extracted behavioural events (interest determination). The analytical approach introduced by the MIR tool depicts the results of applying those properties to each behavioural event while aiming to minimise the cognitive load. For each event, researchers can use the tool to examine the motivation score over a period of time, and visualise and correlate the impact of each property on the overall calculated score. Moreover, the tool generates personalised rules that associate actions motivated by personal interests with contextual data – through association rule mining (Agrawal et al., 1993).

The modular design based on the three processes helped expand the tool's usability. Researchers from various backgrounds can use one or more modules according to their needs. They can perform the behavioural analysis without having to do the motivational one. Hence, the tool can be used by a broader range of researchers rather than focusing on those interested in motivation. In that context, we present three exemplars in which one or more of the three processes supported by this tool have been used for purposes other than motivational analysis. We introduce these exemplars as case studies that are distributed across the paper, keeping the focus and flow of the paper's presentation centred around the motivational use of the tool. Later in the discussion section, we discuss them within the overall context of the tool's discussion.

The main contributions of the MIR tool are:

- The tool supports a motivation-based analysis of smartphone data. It helps researchers distinguish behaviours motivated by personal interests from those motivated by external factors.
- The tool provides a modular and behaviour-centric approach to help researchers apply each one of the three processes separately or collectively based on their needs.
- The tool helps researchers identify cues of behaviours motivated by personal interests through contextual data. It uses association rules to generate behavioural relations between contexts and IMBs.

7.2 Related work

Digital phenotyping is defined as the naturalistic moment-by-moment quantification of an individual's behaviour (Onnela and Rauch, 2016). It is achieved using digital devices such as smartphones to sense human behaviour either passively (e.g. through GPS) or actively using experience sampling methods (Torous et al., 2016). Researchers use several terms to describe the same goal, such as "personal sensing" and "context sensing" (Mohr et al., 2017; Burns et al., 2011). We compare our work with tools that facilitate digital phenotyping. Such tools collect smartphone data passively and continuously for behavioural inferences and studies. mHealth studies that are designed for behavioural interventions but do not rely on digital phenotyping (e.g. Chen et al., 2018) are outside the scope of this work.

Traditionally, digital phenotyping tools are designed mainly to unobtrusively and naturalistically collect and sync data to a backend server. AWARE (Ferreira et al., 2015) and EARS (Lind et al., 2018) are open source tools that run on Android and iOS devices. Developers can extend AWARE to extract behavioural features through plugins such as activity recognition and app usage plugins (Ferreira et al., 2015). Beiwe (Torous et al., 2016) is another tool that provides, in addition to basic phenotyping functionalities, codebase data analysis pipeline for conducting the behavioural analysis. Unlike Beiwe, the MIR tool provides easy to use visual functionalities to help researchers interact with the tool without coding effort.

The data collected from the sensing tools form the basis for digital phenotyping. For instance, app usage has been used as the base for detecting the mood of bipolar patients (Alvarez-Lozano et al., 2014). Symptoms of Parkinson's disease have been investigated from mobility-based phenotyping to improve the patient's quality of life (Vega-Hernandez et al., 2017). Loneliness indicators, as well as physical activities, have been investigated within the context of older adults (Seifert et al., 2017; Sanchez et al., 2015). Behavioural features of depression symptoms have been extracted and studied to detect depression (Wahle et al., 2016; Burns et al., 2011), whereas location features have been employed to detect out of home activities in schizophrenic Patients (Difrancesco et al., 2016). The use of visualisation to support the implementation of digital phenotyping has also been utilised. Health Mashups (Bentley et al., 2013) depicts the connection between sleep, weight, pain. In Visual Cuts (Epstein et al., 2014) data is summarised to help researchers identify meaningful findings. Passively sensed location and activity data had been visualised for reflection analysis (Tang and Kay, 2017).

Besides the above implementations, digital phenotyping has intersected with human motivation in several studies. For instance, MyBehavior (Rabbi et al., 2015) uses the frequency of the visits as an indicator of motivated behaviour. Notifyme (Mehrotra et al., 2015) aims to analyse notification's interest based on the acceptance rate and allows users to check the hourly acceptance rates. RecencyMiner (Sarker et al., 2019) adopts association rules and contextual features to model the behavioural patterns of smartphone usage. Self-Determination Theory (SDT), a well-known theory of motivation, is used to guide the interface design of mobile apps (e.g. Zuckerman and Gal-Oz, 2014; Rooksby et al., 2015), encourage the use of health apps (e.g. Saksono et al., 2020), or propose behavioural intervention (e.g. Gustafson et al., 2014).

Unlike these applications, we employ motivational knowledge to classify behaviours that are passively sensed, as either extrinsically or intrinsically motivated. To the best of our knowledge, the MIR tool is the first tool designed to recognise behaviours motivated by personal interests from digital phenotyping. Accordingly, the MIR tool proposes an approach to a tool that is domain-agnostic and user-specific. It is domain agnostic because it could be plugged into different situations (as shown by the case studies); and user-specific since it performs individual-oriented analytics. Moreover, although machine learning tools support association rules as part of their basic functionalities (e.g. Frank et al., 2009), the MIR tool provides a behaviour-oriented implementation of association rules.

Case study 1: Parkinson disease

Parkinson's is a progressive medical condition that worsens over time and impacts the patient's overall movement (Dauer and Przedborski, 2003). Patients with Parkinson's Disease (PD) exhibits physical and psychological symptoms such as stiffness, fatigue and depression (Cummings, 1992; Friedman et al., 2007). Recently, researchers utilise personal devices – such as smartphones and smartwatches – to sense and identify the behavioural indicators of PD symptoms (Arora et al., 2015). These devices have the potential to facilitate the understanding of the symptoms and their fluctuations through longitudinal studies (Vega-Hernandez et al., 2017).

In addition to symptoms monitoring, Espay et al. (2016) identify enhancing treatment and improving diagnosis and rehabilitation interventions as possible technological contributions to Parkinson's clinical problems.

Accordingly, Vega-Hernandez (2019) conducted a longitudinal study to unobtrusively and passively monitor the relationship between location and activity-based behavioural metrics derived from smartphone data and self-reported pain, gait and fatigue in Parkinson's. Results show that for some patients, clinically informed predictions are moderately correlated to daily increases and decreases of the self-reported severity of the aforementioned symptoms.

This study suggests that adding contextual features, such as the time of the day, could help patients and health professionals uncover individual correlations between human behaviour and Parkinson's symptoms, which in turn, is a step forward towards personalised medical counselling. The MIR tool can help researchers interested in analysing and understanding symptoms of Parkinson through its modelling and visualising of these features.

7.3 Theoretical underpinnings

The MIR tool is designed and built from a psychological background to help researchers recognise interest based on a deep understanding of the motivation behind behaviours. This section introduces the background knowledge necessary to grasp the psychological concepts related to the MIR tool functionalities.

Human motivation studies attempt to answer the question of why humans do what they do (McClelland, 1987; Weiner, 1992; Ryan and Deci, 2000). Intrinsic motivation is the type of motivation that is initiated out of interest to satisfy internal needs (Ryan and Deci, 2000). People who are intrinsically motivated toward a task are more likely to be interested in it (Ryan, 1982; Renninger and Hidi, 2016). Therefore intrinsic motivation plays a significant role in promoting health and well-being. In contrast, extrinsically motivated behaviours are not initiated mainly out of interest and, therefore, may contradict intrinsic motives and could significantly undermine them (Ryan and Deci, 2000). Nevertheless, most of the activities people do are extrinsically motivated (Ryan and Deci, 2017), which complicates the process of observing and recognising intrinsically motivated behaviour (i.e. interesting behaviour).

Measurement methods operationalise motivation, either statically or dynamically based on the underlying theoretical basis. Static operationalisations, such as Maslow's hierarchy (Maslow, 1943) and Murray's system of needs (Morgan and Murray, 1935), associate behaviour to a fixed taxonomy based on the psychological need satisfaction level, whereas dynamic measurements account for the features impacted by the motivation variability over time (Ryan and Deci, 2017). To avoid purely theoretical, and sometimes vague, details provided by existing studies (Oudeyer and Kaplan, 2009; Spruijt-Metz et al., 2015; Gneezy et al., 2011); we base our tool on two dominant and well-found theories in the field of human motivation (McClelland, 1987; Alharthi et al., 2017): Self-Determination Theory (SDT) and Maslow's hierarchy of needs.

We consider the psychological constructs of autonomy, competence, and intrinsic needs as motivation properties that facilitate the static and dynamic operationalisation of intrinsic motivation. Accordingly, the MIR tool is built and designed to depict intrinsically and extrinsically motivated behaviours through the operationalisation of those psychological constructs. SDT defines *autonomy* as the extent to which a person controls a behaviour (Ryan and Deci, 2017) and *self-regulates* goals and the process of attaining them (Schunk et al., 2008). *Competence* (also called self-efficacy) is the one's belief in his ability to perform (Bandura, 1971); the more frequent the action is, the more self-efficacious an individual is. To account for motive prioritisation and understand the nature of needs pursued by performed actions, Maslow (Maslow, 1943) introduces five levels of needs: physiological needs, safety needs, belongingness needs, esteem needs and self-actualisation needs. Maslow's needs are commonly depicted as a layered pyramid with a base representing physiological needs and self-actualisation at the top of the pyramid.

Table 7.1: Motivation properties and corresponding behavioural measurements.

Property	***Measurement***	***Base Theory***
Competence	Intensity of action	SDT
Autonomy	Sustainability of action.	SDT
Needs	Hierarchy-based scoring.	Maslow

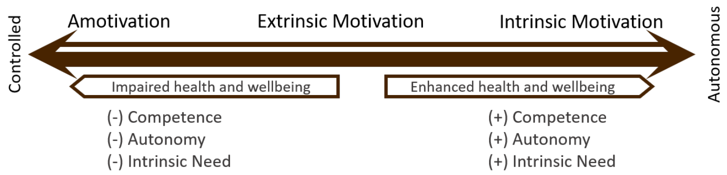


Figure 7.1: The motivation continuum proposed by SDT progresses from a motivation to intrinsic motivation. Performed actions are placed on the continuum based on the scores of motivation properties.

Each one of the motivation properties is mapped to one or more behavioural measurements to quantify its value. For example, SDT considers sustainability as a behavioural indicator of autonomous action, such that people display greater persistence toward the internalised behaviours (Ryan and Deci, 2017). Therefore, we use this behavioural proxy to operationalise autonomy. Table 7.1 shows the motivation properties used in our approach together with the measurements used as proxies to quantify them.

Our overall motivation score is calculated according to the interaction between motivation properties. We use the continuum proposed by SDT to represent various motivation states (Figure 7.1). The placement of action on the proposed continuum corresponds to the motivation state of the behaviour; higher ratings imply more internalised actions and, therefore, a better chance of the person being intrinsically motivated. The algorithmic details of our approach and how it calculates the motivation score have been validated through a real-world study and published separately[1]. This work details the tooling aspect of our work.

[1]We will cite the related paper upon approval to preserve anonymity.

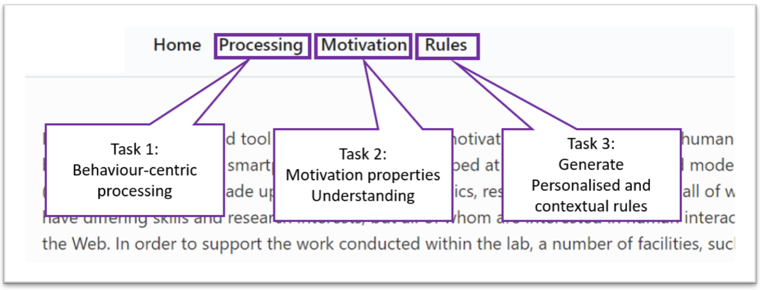


Figure 7.2: The three main tasks that are performed by the MIR tool.

7.4 Overview of the MIR tool

Researchers can use the MIR tool to initiate one of the below three tasks (Figure 7.2) that implement the three processes discussed earlier (i.e. behaviour identification, indicator identification and interest determination):

- Conduct a behaviour-centric processing of smartphone data (section 7.4.1).
- Visualise and investigate the results of implementing the models of motivation properties on passively sensed smartphone' data (section 7.4.2).
- Study the contextual features of IMBs and generate personalised rules based on those contextual features (section 7.4.3).

In designing the MIR tool, we derive the following Design Guidelines (DG) from (Smith and Mosier, 1986) and (Sonego et al., 2018). We aim to ensure developers and researchers can efficiently and quickly interact with the tool and extract behavioural knowledge.

DG1: Simplicity

Researchers who are familiar with timestamped logs and web-based interactions should not need any additional technical background to use the tool. Interpretation of the produced results and interactions with them should be presented in a behavioural-driven manner rather than presenting them technically. This is important to facilitate the insights extraction with a minimal cognitive load.

DG2: Consistency

The attributes of the raw data are expected to differ according to the collection source. For example, mobility data have different attributes than phone interaction data. However, despite the differences expected across input files, behavioural events should have a shared set of attributes. From a behavioural perspective, web browsing and dining out are examples of real-life events, and thus they share the same behavioural attributes (e.g. time of the day and frequency). Therefore, researchers should be able to investigate and understand behavioural events accordingly.

DG3: Flexibility

Although the MIR tool is designed and built based on SDT's and Maslow's interpretations of motivation properties, the tool should support the adaptation of various interpretations portrayed in the literature of human motivation. Suppose the approach preferred by a user proposes a different scoring baseline for the intrinsic needs or aggregates motivation properties differently. In that case, researchers should be able to tailor the tool to match their baselines requirements easily.

DG4: Modularity

Due to the cognitive complexity of interactions among motivation properties (Hekler et al., 2016; Patel et al., 2017), researchers should be able to integrate motivation properties as separate modules. Hence, researchers can examine all possible interactions between features contributing to the motivated behaviour. Moreover, additional properties can be developed as separate modules without impacting the overall interaction with the tool.

Figure 7.3 shows the overall architecture of the MIR tool in which the design guidelines were considered. First, a target behaviour and data sources needed to derive it are determined. Accordingly, based on the determined behaviour and its data source(s), an interface is created to upload the collected data. As a behaviour-centric tool, each interface should be simple and represents one behaviour (DG1). For instance, for the mobility behaviour, GPS is identified as the data source, and an interface to upload GPS data and complete requirements necessary to derive the target behaviour is created.

Passive sensing produces streams of timestamped data points. Segmenting those streams into behavioural events is done according to the target behaviour. Behavioural events formed from mobility streams, for example, may represent dining out or walking a dog. Alternatively, app usage or web browsing are events that can be derived from phone interactions data points. Nonetheless, the MIR tool produces, for each stream of data points, an output file in which the input data points are assigned episode ID and name (DG2). Data points with the same episode ID represent the same events and are named according to the underlying real-world event.

Motivation properties are modelled according to the behavioural measurements discussed earlier. For instance, the intensity of action is used as a measurement of

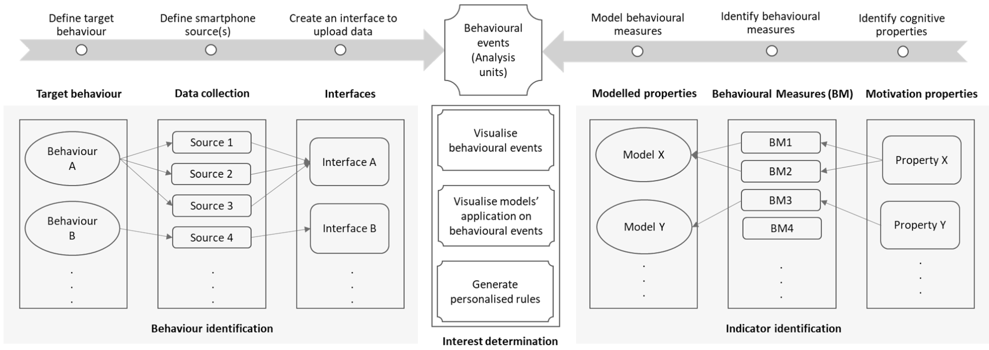


Figure 7.3: Architecture diagram of the MIR tool.

competence. Digital intensity is used accordingly as a proxy measure of competence. The results of applying each measure are depicted separately. Similarly, developers can customise existing measures or add additional properties and apply them to the behavioural events (DG3, DG4).

Motivations behind performing behavioural events are analysed through the visualisation of the operationalised properties. The MIR tool supports week by week developmental analysis of intrinsically motivated behaviour. Personalised rules are generated to associate contexts with behavioural events. Researchers can visually investigate the operationalisation of each property and th e impact on the overall score and its stability (DG4). Next, we dive into the details of the tool's functionalities and how researchers can utilise them for motivation-based interest recognition.

Figure 7.4 shows a screen sequence that depicts the processes of the MIR tool. When researchers select the processing tab on the Home page (see Figure 7.2), they can upload the data that they collect using the "Mobility Data" or "Phone Usage" interface (Figure 7.4a). The data are processed and segmented sequentially to form behavioural events. Researchers can save the results in either textual (i.e. CSV file) or visual formats. Only the textual files can be used for future processing. Next, researchers can either initiate the motivation analysis, visualise contextual features, or generate behavioural rules using one of the buttons on the screen (Figure 7.4b).

Another sequence is depicted in Figure 7.5 which shows how researchers can benefit from the files that they processed and saved. When a researcher selects the motivation tab, the interface in Figure 7.5a can be used to upload the CSV files of either the mobility or phone usage behaviour. The motivational analysis can then be applied to the uploaded files. The CSV files can also be uploaded to investigate the contextual features that are associated with motivated behaviours. If researchers want to generate

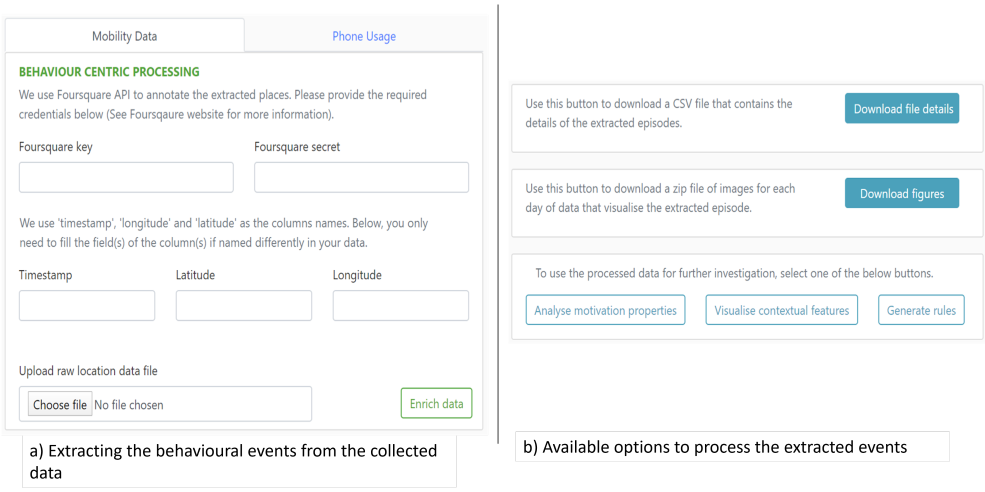


Figure 7.4: Screenshots that show two interfaces: (a) shows the interface used to upload the data, and (b) shows the interface that appears to the user after the uploaded data is segmented and annotated as a result of clicking “Enrich data” button in (a).

the rules directly from these CSV files, they can go to the rule tab of Figure 7.2, and upload the CSV file to do so using the interface in Figure 7.5b.

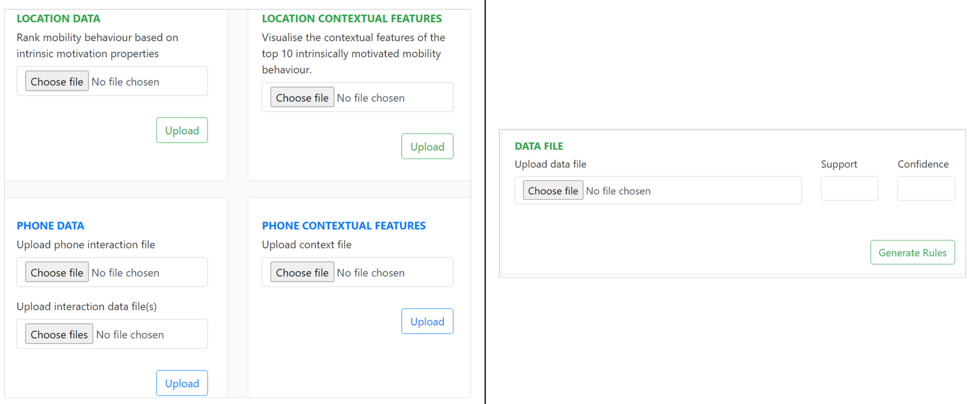
In the following subsections, we will detail each one of the main three tasks supported by the tool.

7.4.1 Behaviour-centric processing

The MIR tool semantically enriched the raw smartphone’s data to form behavioural events. The semantic enrichment typically involves: (1) a segmentation task that groups together raw data representing a behavioural event; and (2) a semantic annotation task to assign basic features (e.g. name, type) to the formed events. Currently, the MIR tool can semantically enrich location and interaction data.

Location data uploaded by researchers must contain timestamped longitude and latitude readings. The data are segmented according to a well-known stay-points extraction algorithm (Li et al., 2008). Data points are processed sequentially, and stay-points are defined based on predefined time and distance thresholds. Fifteen minutes and 100 meters are used as the default time and distance thresholds. However, researchers can change that according to their needs. The extracted stay-points are annotated based on Foursquare API. Researchers need to have the appropriate credentials to use the API service (Figure 7.4a) .

With respect to phone interactions, researchers need only to upload a timestamped



a) Uploading previously enriched data to apply measures of motivation properties.

b) Uploading previously enriched data to generate association rules.

Figure 7.5: Screenshots that show two interfaces: (a) shows the interface used to apply measures of motivation properties on a CSV file that is enriched previously using the MIR tool, and (b) shows the interface used to generate association rules based on a previously enriched and saved CSV file.

file with the phone interaction data, and the tool generates a CSV file that contains the details of the event. We build our tool to work seamlessly based on the data that are passively sensed using the AWARE platform. However, other formats should work as long as they contain similar entries. Researchers have the option to upload contextual information about battery and screen status for better identification of the segment boundaries. An event is formed when the phone is out of battery, or the user turns off the screen. Otherwise, since smartphones allow interactions with one app at a time, app names are used as the segmentation base. We use GooglePlay to retrieve the category of the identified event.

The content of the generated files contains more columns to detail the contextual features and data related to the semantic enrichment process. Researchers can examine the output details of the enrichment process by exploring the columns of the generated file. The MIR tool assigns a unique identifier to the data records that belong to the same behavioural event. Additionally, temporal features are included in the generated file to help researchers investigate associations between behavioural events and the extracted features (later, we discuss the contextual features in more detail).

Currently, researchers can visualise the behavioural events generated from mobility data only. For each day of mobility records, researchers will be provided with a figure to help them quickly envisage daily behavioural mobility patterns (figure 7.6). The shaded areas are exemplars of behavioural events extracted by the tool, which indicate

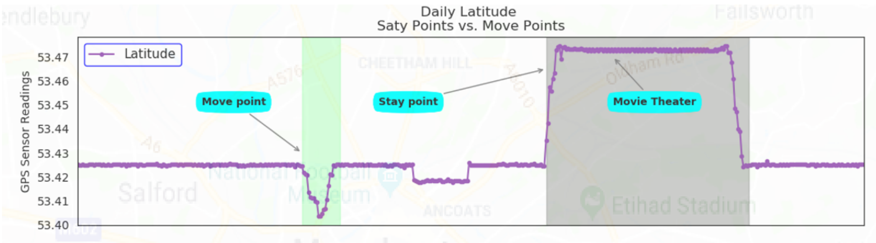


Figure 7.6: Daily behavioural mobility patterns. Shaded areas are examples of stay and move points.

various activities. Researchers can use this to conceive when the user stays at a specific location (a.k.a stay-point) and for how long to help them quickly compare the duration of multiple stay-points. Additionally, researchers can have a quick glimpse of the moving activity directly from the figure by comparing the slopes of the signal. A steeper slope indicates fast movement (e.g. driving), while a flatter one may express a slower transition produced by a walking or running activity. Lastly, developers may choose to apply a different stay-point detection algorithm and use the generated figures to investigate the segmentation accuracy.

7.4.2 Motivation properties

Researchers can apply the static and dynamic motivation properties to the identified behavioural events. They can initiate that after processing the raw data through one of the buttons shown in Figure 7.4b, or based on previously saved and enriched data as in Figure 7.5a. The tool visualises the connection between the motivation properties and the extracted events based on week-wise analysis windows. We select weekly basis as people mostly shape their behaviour around weekdays (Cho et al., 2011; Sarker et al., 2019). Researchers can perform week-wise comparisons to derive insights embedded in the relationship between behaviour and motivation properties when investigating the motivation properties. However, we plan to provide researchers with the ability to experiment with different bases that suit their needs (e.g. monthly or quarterly).

To help researchers reflect on static modelling, the MIR tool quantifies the needs based on Maslow's hierarchy. Behavioural events classified as physiological needs are assigned the value of one; safety needs events are assigned two, and the remaining events are assigned the value of four[2]. However, researchers may disagree with the scoring system and reference different motivation studies, or they may go beyond that

[2]The rationale and evaluation of Maslow's quantification are detailed in a different submission.

and conduct extensive studies about the best scoring that represent the mapping between behaviours and Maslow's layer. For these reasons, researchers may decide to modify the need values to align with the base they use for their scoring. However, unlike competence and autonomy, need realisation is not presented on a weekly basis as it is assumed to be fixed and participant-independent.

Competence is operationalised via the behaviour's intensity, which is a function of frequency and duration. Researchers can check the intensity – and its frequency and duration components – of a specific event at a particular point of time, as well as study how intensity evolves over time. Since performing some behaviour may impact the practising of others, researchers can analyse the correlation between various behaviour – represented by behavioural events – and draw conclusions about the relationship between the absence and presence of actions. Moreover, as higher intensity implies a higher competence level (Fishbach and Hofmann, 2015; Nicholls, 1984; Wolf and Hopko, 2008; Rabbi et al., 2015; Ryan and Deci, 2017), behavioural events are ranked accordingly to help researchers compare the score of intensity with the MIR score.

The MIR tool also enables researchers to examine the autonomy construct through its behavioural measure, sustainability. Intrinsically Motivated Behaviours (IMBs) are sustained over time. Individuals may differently exhibit sustainability. A person may practise an IMB on a weekly basis (e.g. moviegoing), while another individual may attend a book club every month. The visual depiction of sustainability over time enables researchers to understand how individuals actualise this property.

It is essential to distinguish the sustainability from the frequency component of intensity when inferring motivational insights. Sustainability is represented by a Boolean value that indicates whether a specific behaviour occurs within the analysis window. The cumulative aggregation of sustainability values of multiple analysis windows is different from the cumulative score of frequency component for the same set of windows. This distinction helps researchers assess how these two dynamic aspects interplay to extract in-depth motivational insights.

The motivational properties are normalised and linearly aggregated to rank the enriched behavioural events. Higher rankings do not always correspond to higher scores for all properties. Therefore, researchers can reflect on the MIR score and investigate how each motivation property affects the overall ranking. Also, the MIR tool enables researchers to correlate the dynamics of IMBs to the individual motivation properties applied to each event. Moreover, instead of collectively investigating all motivation properties and behavioural events, the tool enables researchers to selectively filter behavioural events to visually assess the impact of a motivational property on a subset of the extracted events.

Case study 2: Social isolation

Social isolation refers to a state when people are lack contact with other individuals or society (Nicholson Jr., 2009). Studying social isolation helps in understanding its impact on the individual's health and quality of life (Hawton et al., 2011). Smartphones are the hub of communication nowadays, so in a previous project, we deployed a monitoring application to longitudinally collect personal data related to social behaviour.

We used the MIR tool to generate higher-level features from the massive collected raw data. The tool enriched raw GPS data such that we were able to identify where did the participants go and how socially they were. The tool also gave information about the phone usage duration based on apps used by a participant. We used that to analyse the usage time of social media and communication apps (e.g. Facebook), which are good indicators of social isolation (Cho, 2015).

Using the generated graphs, the tool helped with the initial data analysis by providing an immediate impression of the changes. So, we can know the time points in where location and application pattern changes happened. These time points could be used as signals of social isolation that we can make further investigations into them.

7.4.3 Rules and Contextual properties

To analyse the relation between contextual properties and IMBs, the tool presents statistical and rule-based functionalities that target contextual features (Figure 7.5b). Researchers can enrich their understanding of interests (represented by IMBs) and use contextual information for better personalisation. For instance, the MIR calculations may suggest that movie-watching is an intrinsically motivated behaviour for a specific person. If the analysis of contextual features shows that the person dominantly watches movies on Saturday evenings, then that should be considered for personalised recommendations.

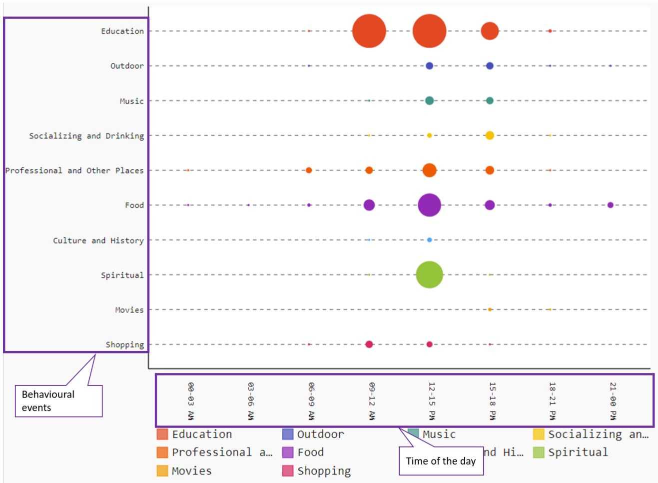


Figure 7.7: An example of contextual attribute: Time of the day.

Each contextual feature is represented by a graph that correlates a contextual feature with the top IMBs. For instance, researchers can explore the correlation between the times of the day and IMBs (Figure 7.7).

The contextual features are used to generate personalised rules based on association rules, a well-known technique from the literature of machine learning (Agrawal et al., 1993). One of the prominent challenges when applying association rules is how to avoid the proliferation of uninterested rules. As we aim at recognising IMBs, the rule engine of our tool operates only on behaviours motivated by personal interests to generate a minimal set of personalised rules directly related to the researchers' investigation goals (i.e. detecting personal interests). Besides, researchers can apply their own parameters for support and confidence (Srikant et al., 1997).

Lastly, to serve researchers from various backgrounds, the tool integrates two options for presenting the rules to researchers. Researchers with little or no knowledge of association rules can benefit from an expressive language that does not adopt technical details, such as confidence and support. In contrast, computer scientists may prefer a presentation that suits their technical background and understanding of the association rules. Therefore, we produce a WEKA-similar representation (Frank et al., 2009) of the generated rules since WEKA is one of the most popular tools used by machine learning practitioners (Hall et al., 2009).

7.5 Discussion

We use the described tool to implement our Motivation-based Interest Recognition method on location and phone usage data[3]. The tool was used as a trigger in a real-world study that is detailed in a separate paper. Eight participants were recruited, and their data were collected for six months. The collected data was used to derive behavioural features and subsequently understand interests from them. During the final interviews, we used the figures generated by the tool to show participants how their behaviours were analysed. Most of the participants requested a copy of their analysis which indicates that the visualisation of their interests was positively received. In fact, some of the participants discussed their perspectives on the visualisation during the interviews. For instance, a participant stated that the capturing of interests dynamics and the longitudinal tracking of his interests is very important for him to organise and set his overall priorities:

> *This information (about interests) is really useful to me; to have my interests tracked without me being writing out dairies. Because right now, I am at a part of my life where I want to structure the things that I do and set my priorities right. So, it is really useful information for me. If I would be able to have an app that tracks this for me, I would certainly use it.*

Another participant emphasised the importance of visualising the dynamics and changes of the behaviours motivated by personal interests compared to just showing historical trajectories of all activities that he does:

> *I would absolutely love an app about this. It is cool to see the ups and downs of my interests. I think this is better than just seeing what I do. My interests are what matters to me, and if I can track them, I can better take care of myself.*

Besides the interest-oriented use of the tool, the three exemplars distributed throughout the paper indicate the versatility of the MIR tool. They provide a short study for three cases of using the tool. The behaviour-centric processing facilitates the recognition of many behavioural features. Some of those features are indicative of the underlying symptoms of a specific health condition. Fatigue, for instance, is a disabling symptom associated with several health conditions such as Parkinson Disease (Case study 1). Since specific behavioural features, such as distance travelled and spatial

[3] We will cite the related paper upon approval to preserve anonymity

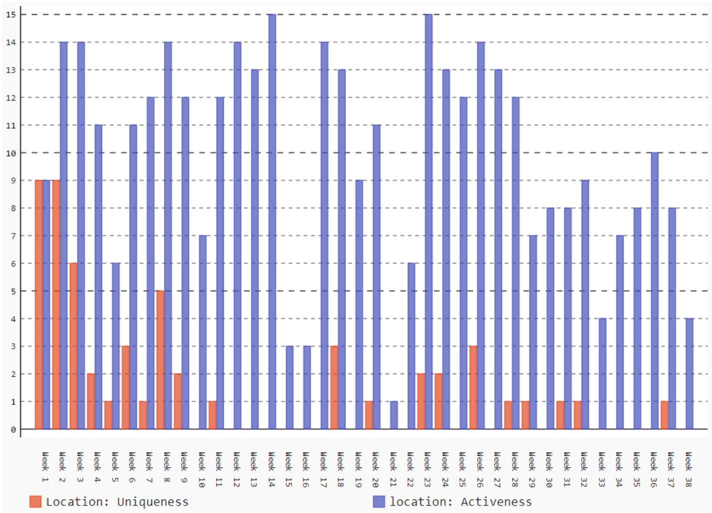


Figure 7.8: Activeness chart indicates a weekly-base uniqueness and visited locations.

coverage, are potential indicators of fatigue (Vega-Hernandez, 2019), researchers can quickly and easily investigate that through our tool. The daily figures produced at the end of the behaviour-centric processing (see Figure 7.6) is a conduit that researchers can use for similar analysis.

A different set of features can be derived from the participant's phone usage. The tool's visualisation of longitudinal apps usage enables researchers to study behavioural features related to the case they study. For instance, combing the usage pattern of social apps and communication logs may help different researchers characterise symptoms related to social isolation (Case study 2). The latter shows an example of how multiple behavioural features depicted by MIR may collectively facilitate the assessing of an underlying symptom.

Researchers may also conceive two different symptoms using the same indicator of a behavioural feature. For instance, a user interested in Parkinson disease may investigate the activeness chart (Figure 7.8) to envisage the existence of a symptom (e.g. fatigue). The same activeness chart may also help other researchers studying the developmental pattern of social isolation or the people's compliance with COVID-19 policies. In fact, a paper that used the tool to analyse the impact of COVID-19's stay at home policy on people's mobility, as well as their phone usage patterns, is published (Ibrahim et al., 2021b).

The interaction between behavioural traces and motivation properties may help researchers better analyse the underlying symptoms. This interaction is salient in application domains related to behavioural recommendations. Personalised nudges, for

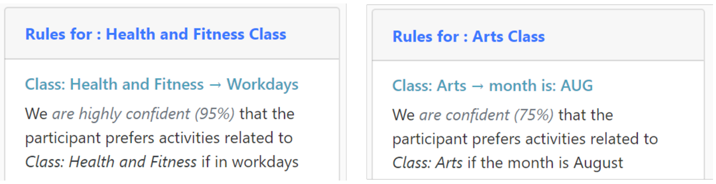


Figure 7.9: An example that two rules that are generated using the MIR tool.

instance, are indirect recommendations aim to influence the behaviour of individuals (Case study 3). If a user targets the fatigue symptom, behavioural features can be used to observe the symptom. Moreover, since our tool facilitates the recognition of IMBs, researchers can use those intrinsically motivated events as a tool to generate personalised nudges when a negative indicator is exhibited (e.g. declination in activeness level).

Such as how contextual features improve the recommendation process, they can improve the manipulation of nudging options. People may exhibit their motivated actions differently based on temporal features such as the time of the day. In that case, planning personalised nudges while considering the appropriate time to nudge would augment the possibility of a positive response from participants. Researchers can achieve that by associating the contextual features with the extracted IMBs. Moreover, they can investigate the produced association rules to use contextual information when nudging (Figure 7.9 shows an example of the generated rules).

The presented tool has some limitations. First, the tool is not designed to compare different implementations' approaches of the processed or operationalised properties. For instance, the behavioural processing of mobility data uses Foursquare as the annotation source. This makes the accuracy of the annotation strictly limited by the data returned from Foursquare. Although we plan to support additional external annotators such as Google Places API, our focus is not oriented toward the comparison between different implementations (i.e. Foursquare vs Google Places). Instead, researchers need to make a decision on their preferred implementation's approach and accordingly investigate the extracted properties.

A second limitation is related to the number of behaviours that are supported by the tool. Currently, we support the analysis of mobility and phone usage behaviours. Respectively, GPS and screen interactions are used to derive and analyse the two behaviours. Although additional sources can be used to improve the recognition of both behaviours, they are not supported by the tool. Interested developers, however, can benefit from the tool's modularity to accommodate additional sensors.

Lastly, each behaviour is analysed separately to preserve the modularity of the tool. However and as integrating multiple behaviours in the analysis process can be beneficial, we plan to incorporate the ability to integrate multiple behaviours as part of the tool's design while preserving its modularity. Within the same context, additional motivational properties can also be integrated using the appropriate behavioural measure(s). For instance, we plan to add recency as part of the tool to operationalise the motivation property of novelty. This property is essential for distinguishing behaviours motivated by personal interests since people are expected to abandon existing interests and acquire new ones (Zhao et al., 2013). Novelty is also a part of the algorithmic aspect in which this work is built.

Case study 3: Personalised nudge

Nudge theory (Thaler and Sunstein, 2008) proposes less coercive suggestions to change behaviour (Kosters and Van der Heijden, 2015). The idea is to arrange options in a way that advantages some alternatives over the others (i.e. paternalism) while preserving the people's ability to choose the unpreferred options or avoid the promoted ones (i.e.libertarianism) (Thaler and Sunstein, 2008). Choice architects are responsible for presenting options. These options are traditionally based on course-grained insights derived from domains such as psychology and economics (Thaler and Sunstein, 2008).

Digital nudging embraces adaptation of user-interface elements to deliver nudges (Weinmann et al., 2016). As a form of digital nudging (Schoning et al., 2019), personalised nudges utilise digital devices to 'architect' choices based on personalised insights. Accordingly, options are derived from an individual's behavioural and motivational features rather than a one-size-fits-all approach.

Recently, Caraban et al. (2019) published a systematic review classifying nudges based on those categories. None of the reported nudges is classified as personalised nudges (i.e. choice manipulation). Authors embrace the potential positive effect expected when personalising nudges based on a better behavioural and situational understanding of individuals.

Personalised nudging can be applied by using people's interests to design nudges. For example, if we want to reinforce a specific behaviour (such as walking) for a person who is interested in shopping, we can nudge by suggesting walking paths that pass through the shopping places that this person prefers.

Similarly, walking paths that pass through green and open spaces can be suggested to people who are attracted and intrinsically motivated by these views. The MIR tool can help us design these nudges through its ability to recognise interests and understand each person's commuting patterns and times.

7.6 Conclusion

The MIR tool enables researchers to extract insights based on a psychological understanding of people's motivation. Using the tool, researchers can conduct behaviour-centric processing of smartphone data to generate and visualise behavioural events, the counterparts of real-life events such as dining out and web browsing. Researchers can use the tool to facilitate the extraction of insights from behavioural events while preserving the ability to investigate the psychological components that compose those insights. Moreover, researchers can examine the association between contextual features and motivated behaviour through visualisation and behavioural rules.

Unlike existing work, the MIR tool enriches raw smartphone data with behavioural and motivational semantics to help researchers better investigate the collected data. This semantic enrichment provides analytical benefits to domains where behavioural and motivational insights are crucial (e.g. personalised nudges).

Since most digital phenotyping tools produce mobility data through GPS coordinates (i.e. longitude and latitude), we expect the MIR tool to work with location data seamlessly. However, for smartphone interaction, we work on improving the interoperability of the tool to accommodate different formats of apps' logs. Moreover, to better serve a holistic understanding of motivated behaviour, we plan to process additional behaviours such as buying. We also plan to improve the visualisation of association rules to help researchers investigate several configurations and compare them quickly and efficiently. Lastly, although developers can extend the tool to include additional features, we plan to improve the modularity of the tool to make it easier to integrate multiple behaviours in the analysis.

7.7 Acknowledgement

We thank the individuals who helped with writing the case studies and providing feedback on the presented tool (we will add their names upon approval).

Chapter 8

Conclusion and Future Work

In this thesis, we identified interests by recognising them from exhibited behaviours rather than relying on people's self-reporting. Smartphones data formed the basis for a digital phenotyping process that aimed to derive behavioural events and features. Indicators of intrinsic motivation were applied to the derived events to recognise behaviours motivated by personal interests. Those behaviours are the manifestations of personal interests and can support the design of effective personalised behavioural interventions and nudges while avoiding the downside of self-reporting methods (Mills, 2020; Schoning et al., 2019).

We have relied on events of mobility, phone usage and buying behaviours derived from smartphone data as the basis for the analysis. Using a variety of approaches, including systematic reviewing of the literature, the analysis of secondary data, and applying text mining and machine learning techniques on the notifications' text, we explored various methods of recognising the behavioural events. The attributes of the collected data and privacy preservation guided the selection of the methods that we used to extract behavioural events.

A Motivation-based Interest Recognition (MIR) approach has been proposed based on current psychology theory. Both MIR (used for a single behaviour) and cMIR (a combinatory MIR used for integrating the three behaviours) identify a set of static and dynamic motivation properties (needs, competence, autonomy and novelty). We have developed a set of behavioural measures that can be applied to the extracted behavioural events based on these properties. However, these measurements are combined and modelled differently according to the singularity (i.e. MIR) or multiplicity (i.e. cMIR) of behaviour. The formed score (ranged between 0 and 1) shows the level of intrinsic motivation connected with each behavioural event.

To inform and assess our method, we have conducted real-world studies. The studies were meant to track people's behaviour over time as they went about their regular lives. Our findings demonstrated the benefit of recognising personal interests based on motivational knowledge. When compared to baseline approaches, our approach boosted interest recognition by an average of 62% with $p < 0.05$.

Our work presented by this thesis could be particularly valuable for personalising behaviour change, where existing evidence indicates that leveraging personal interests leads to more effective and sustained change (Ryan and Deci, 2017). Future work can build upon our effort and measure the enhancement it may add to personalised intervention and nudging.

While each discussion was included through each chapter, it is helpful to present next some overarching thoughts that synthesise the presented work. We used three different datasets (a total of 22 participants) to evaluate both the extraction of behavioural events as well as the recognition of personal interests. Our results showed that the reliability of interest recognition is directly affected by the validity of extracting behavioural events from smartphone data. The steps for extracting these events were similar across the three behaviours studied by this thesis which mainly included segmentation and annotation tasks. However, the application of these tasks varied according to the target behaviour. Events of mobility and phone usage behaviours were segmented based on sequential processing of the collected GPS and phone interaction data, respectively. In contrast, we discussed in Chapter 4 how text mining and apps' categories were used to identify buying events from smartphone notifications. Despite these differences between the three behaviours, they all used external annotators to add semantic to the segmented data.

The annotation of behavioural events (segments) depended on the type of behaviour represented by these events. In the mobility behaviour, this was done by first identifying the user's home and then retrieving the other places using an external provider (Foursquare). This process helped increase the accuracy of detecting behavioural events, which positively affected the validity of the interest analysis (as we explained in the paper included in Chapter 5). Similarly, the annotation of buying events included more than one step. First, a query request was executed using Google, and then Amazon was used to find out the type of products that Google did not recognise. This two-step process contributed to recognising a larger number of products and overcoming the inability of Google to recognise short texts. Finally, and in contrast to the previous two behaviours, annotating phone usage events relied on a single step in which Google Play was consulted only. This is because the study was limited to smartphones that use the Android system.

The selection of segmentation and annotation details was driven by the goal of this thesis. We do not suggest that the selected details are better for a broader and general context of digital phenotyping. As an example, we used Foursquare to get a detailed taxonomy of places, which can better help understand interests from mobility events. Our selection of Foursquare does not imply a general advantage of it over other annotators such as Google Places. Also, in Chapter 4, we compared and discussed different methods of selecting notifications related to buying behaviour and extracting features from them. We then (in Chapter 6) selected the methods that provide acceptable accuracy and preserve the privacy of the participants' data in order to adhere to the goals of this thesis.

The results of recognising personal interests using our MIR method were based on the selected segmentation and annotation details. Both MIR and cMIR outperformed the alternative methods used typically to recognise interests. When determining the top N interests using a single behaviour, a small N produced a significantly better recall for the cMIR compared to MIR and other alternatives (Chapter 6). On the other hand, the precision has improved, but the improvement was not statistically significant using a small N. As we increased the selected number of interests (i.e. set a larger value for N), the precision and recall significantly outperformed other alternatives. This trend showed the advantage of our method in detecting interests, especially the postliminary ones.

In selecting the top N interests, we examined multiple methods. When determining the top interests based on a fixed N, it was evident that clear differences in the strengths of consecutively ranked Intrinsically Motivated Behaviours (IMBs) naturally emerged at different points for each participant. Establishing a cut-off point based on the largest difference between consecutive behaviours (i.e. dynamic N) solved that issue. However, and as noted in Chapter 5, dynamic selection of N in many cases was equal to one, suggesting that the participant's behaviour reflects a single interest much more strongly than any others (this is especially true when considering frequency or duration alone). Thus, for each case where the dynamic N would be 1, we identified the next largest difference between the remaining behaviours to get a broader understanding of IMB. The results of the final study (Chapter 6) confirmed the need to go beyond a single interest. The data showed the broader selection of the top N interests using our approach as a significant predictor of the participants' interests compared to alternatives.

When calculating the scores on which the top N were chosen, we analysed the score's stability—the time required for the calculated score to stabilise. In Chapter 5, the results of the secondary data analysis showed that the average stability time based on the mobility behaviour was three months. Based on that, the two real-world studies were designed and run. However, we acknowledged that variation in stability scores might differ according to demographic differences between different study groups or the analysed behaviour. Conditions such as age, health condition and work may impact the type of interests and how they are realised by each group. Also, scores calculated based on mobility behaviours may take longer to stabilise when compared to scores based on phone usage behaviour. Nevertheless, our results suggested that the MIR approach can be used to detect interest regardless of how groups may differ in the type and realisation of those interests. This is in line with the findings from the literature that do not relate the indicators of motivation properties – used in this thesis – to a specific type of interests or group of people (Ryan and Deci, 2017). Instead, these indicators are generic and describe behaviours that are motivated by personal interests. A comparison that shows how stability may differ across demographic groups can be further explored by future work.

Our study of stability guided the determination of the period length when we applied the recency factor. To adapt to the expected changes in interests, we examined novelty (through the measure of recency) in the final study. We adopted a static threshold of three months and divided the entire six months duration into two 3-months periods. Accordingly, we were able to detect newly developed interests and observe their stability. In Chapter 6, we discussed examples that showed the importance of doing so as part of the interest analysis. Also, we discussed the significance of degrading the value of older behaviours rather than entirely abandoning them when analysing behaviours. The final interviews and the related literature supported our decisions and findings that are related to recency.

Lastly, we noted a general trend for strong positive correlations observed between MIR and both intensity and sustainability, suggesting that MIR successfully reflects these intrinsic motivation measures as developed by the IMI authors (see Appendix C). We also noted a positive but a less strong correlation between MIR and needs. Extremes and variations from this general trend for individual participants suggest that some participants exhibit more behaviours that are considered to be more intrinsic by Maslow's hierarchy, whilst others may be intrinsically motivated to engage in behaviours that could be interpreted as extrinsic when considering Maslow's need level alone. Further, during interviews, our participants reported pursuing these supposedly extrinsically-motivate behaviours for both intrinsic and extrinsic reasons – thus, the combination of needs with our other measures is essential to produce an accurate MIR score appropriately. We suggest that this variation does indicate that personalised weighting of individual elements could improve the performance of our method, a step that warrants further exploration in the future.

8.1 Main findings

- **The reliability of detecting interests relies heavily on the reliability of deriving behavioural events.** Unlike traditional computer systems that deal with 'interest' using predefined items, our approach had to decode behavioural events encoded into streams of low-level smartphone data. The reliability of decoding and extracting these events influences the interest recognition process directly. Therefore, we systematically reviewed the literature on mobility behaviour to better detect events from GPS data. Also, using the data collected through our real-world studies, we compared different approaches to detecting events of buying behaviour. We studied the advantages and disadvantages of each method while considering the goal of this thesis. Accordingly, the methods that suit the purpose of this work were selected. For events of phone usage behaviour, we benefited from our review of the related literature in identifying them. As discussed, those events are smartphone-dependent. Hence, the reliability of their detection is affected to a lesser degree compared to mobility and buying, which are smartphone-independent. It is noteworthy to mention within this context that the more improvement that is gained in detecting events from smartphone data, the better the recognition of interest becomes. This was evident when we detected home and asked participants to correct the invalid annotation during the pilot study. As shown in the related paper (Chapter 5), the results of so doing helped improve the interest recognition because the analysis units (i.e. behavioural units) were better identified.

- **The integration between the type of behavioural events (static modelling) and how they are performed (dynamic modelling) is essential for interest**

recognition. Our approach of recognising behaviours motivated by personal interests included the two aspects that are typically covered by studies on human motivation. By relying on well-founded theories, we were able to computationally model properties of the two aspects. The positive correlation between the modelled properties that we got after applying them to the behavioural events matched the expectations supported by the theoretical underpinnings. The pilot study showed how the digital measures of static and dynamic aspects improved the recognition of interests. Also, it showed the ability of both aspects to capture various motivational cases. Examples included cases in which one person's work was motivated by extrinsic motives while another person considered the work to be a manifestation of personal interest. We showed in that pilot study how alternative methods failed to address similar cases, which negatively impacted their precision and recall scores. The improvement produced by our approach was consistent across different methods of selecting the top N interests, whether dynamic or fixed. In the main study, we benefited from the longer period and conducted a similar analysis that used single behaviour. The results of statistical significance using each behaviour were similar to what we got in the pilot study.

- **The multiplicity of behaviours improved the interest recognition.** To overcome the limitations of relying on a single behaviour (e.g. the inability to capture phone usage interests from mobility behaviour or the mobility restrictions imposed by COVID-19), we integrated multiple data streams to capture a broader set of behaviours. Our results in Chapter 6 showed that the combinatory approach had improved the recognition of IMBs when compared to the reliance on a single behavioural trace. The significance of this improvement relied on the number of interests retrieved. When the average N (using the dynamic selection) was two, only the recall of the combinatory (cMIR) was statistically significant. As we covered additional interests (by increasing N to six in order to match the average number of interests reported by participants), both precision and recall of our cMIR were significantly better (by 62%) than other alternatives.

- **Adaptability and personalisation are essential for interest recognition.** We discussed (in Chapter 6) the importance of personalising the integration of the three behaviours per participant based on the coverage days of each data source. Accordingly, we were able to provide a flexible method that is capable of reacting to the interest dynamics and producing a personalised model from multiple behaviours and from an overfitted set of motivation properties. The tool that we provided in Chapter 7 depicts the rise and fall of specific interest (i.e. its dynamics) and visualises the results of applying our approach per participant. Also, we provided quotes from participants that embraced the importance of adapting the analysis according to their interests' dynamics.

8.2 Limitations

Our studies showed that the semantic annotation upon which our measures are based had a significant impact on the results. In each of the three behaviours, recognition of interests is constrained by the accuracy of the external annotators. For example, the proximity of places in a small space can affect the correctness of annotating the place. This error, in turn, will lead to the misidentification of the mobility event and an application of the measures to incorrect information. Similarly, the annotation problem can happen with the buying and phone usage events. Enhancing the accuracy of the associated semantic annotators is beyond the scope of this work. However, we do note a specific challenge with regards to the use of external annotators.

Also, due to the sensitivity of personal data, we derived behavioural events through methods that considered the privacy of the data. These methods are not necessarily the best at deriving behavioural events, but they do ensure that privacy is preserved while maintaining a reliable derivation accuracy. As shown in Chapter 4, the machine learning algorithms did a better job in filtering out notifications that are irrelevant to buying behaviour. However, using them requires participants to share their data in order to train models or the installation of a pre-trained model on the participant devices. As both solutions are not visible, the best alternative was to rely on shopping apps and keywords related to buying to identify related events from the smartphone notifications.

Our static modelling assumes that each event serves a single purpose and that purpose is consistent over time, but in practice, some users will do the same event multiple times, each with very different intentions. For example, one user may visit a coffee shop with the primary purpose of eating lunch (a physiological need), engaging in work or quiet study (a safety need), or engaging in a hobby (e.g. reading, knitting) alone or with others (intrinsic). Thus, the one-to-one mapping reported throughout this thesis may limit the recognition of the performed interest at these places; approaches to overcome this challenge are left for future work, but will likely involve the integration of additional contextual data. Also, details of the performed interests can sometimes be difficult to recognise by relying only on a single source. For instance, being at a place related to football, by itself, may express potential interest in football regardless of how that interest is being actualised (either through watching or playing). Future work can benefit from additional sensors and contextual data to improve the granularity of the detected interests.

Although the measures used in our approach has been largely studied and supported by the literature on human motivation, we do not present them as the absolute and only measures of motivation properties. Instead, this work provided a novel and the first approach that benefited from human motivation literature and digital phenotyping for a better understanding of personal interests. Well-founded theories guided the modelling decisions, and hence the decisions are limited by those theories interpretations. Future work may propose and adopt different interpretations and compare how they may differ or agree with the one presented by this thesis.

This work is also limited by what is referred to by researchers (Aeffner et al., 2017; Vega-Hernandez, 2019) as the "Gold Standard Paradox." The contradiction arises from the fact that our study adopts digital phenotyping, which relies on sensors' data that are objectively produced. The goal is to use knowledge derived from those data to replace the subjective self-reporting methods and avoid their issues (e.g. memory and recall biases). However, these self-reporting tools (IMI in our case) are the best instruments for assessing our method. As a result, our techniques must be assessed using the scales that they are attempting to replace.

Lastly, our work mainly depends on capturing and analysing the behaviours of individuals. These behaviours (particularly mobility behaviour) are undoubtedly affected by the repercussions of the COVID-19 pandemic. Therefore, the validity of the findings of this work must take that into consideration. However, we affirm that any potential impact of this pandemic is expected to be primarily related to the number of interests that can be recognised, not to the methods in which those interests can be identified. These methods are what we analysed and detailed throughout this thesis.

8.3 Future work

In previous chapters, we have highlighted several recommendations for future work. Below, we synthesis those future opportunities under the bigger picture:

Annotation: The annotation step is critical for understanding interest. Our analysis of static and dynamic aspects are based on the semantic labels assigned to the behavioural events. In this work, we carefully selected the external annotators that can serve our goal. Although suggesting methods of improving the annotation based on external annotators is outside the scope of this thesis, it represents an important opportunity for future work that would have a significant impact on studies of behavioural analysis. Investigating methods of improving annotation could target finding the best ways of utilising and synthesising the information available by the external annotators. It also may include how information of other sensors such as WiFi and Bluetooth labels can contribute to the semantic enrichment of the behavioural events.

Interest properties: Although the primary goal of this research is to recognise interests from smartphone data, there are still many aspects related to interest properties that are worth further investigation. For instance, in Chapter 5 and Chapter 6, we detailed the use of the recency factor and its role in our method. Future work can go one step further and study the static and dynamic selection of recency factors and the impact of that on detecting interests. Also, the methods of selecting the top N can be further investigated by future work. As shown by this work, the investigated methods of determining the top N have their advantages and disadvantages. A deeper investigation of this topic may produce a better way of selecting the top N that overcome the limitations of the existing ones. Another aspect is the stability of interests which can be further explored and studied. Specifically, the factors that contribute to the stability of each behaviour and how to model them in a general method of deducing stability could be explored. Lastly, the experimentation of additional motivational properties and different scoring systems of Maslow's needs form an exciting area for future study. As stated throughout the thesis, we do not consider the selected motivation properties (and their measures) to be exhaustive; future work may demonstrate the utility of alternative or additional measures. For example, an experiment may use the diversity and flexibility in times of a behaviour as a measure of autonomy (Ryan and Deci, 2017) along with sustainability. Similarly, we envisage future integration of a measure for relatedness through indicators derived from proximity or other sensors.

Personalisation: Our studies showed considerable individual variation, particularly with regard to the execution of activities that correspond to the lower levels of Maslow's hierarchy. This was predicted by prior literature that suggests that while all individuals share the same fundamental needs, their pursuit of them may vary (e.g. based on the degree to which they have previously had that need fulfilled) (Deci and

Ryan, 2000). Thus, the personalisation step in our overall approach can be extended in future to consider the degree to which coefficients within cMIR can be personalised in order to deliver a more tailored IMB calculation. This is in addition to how it was used in this work to weigh each behaviour according to the individual's data. The personalisation of cMIR coefficients does not have to be limited to the motivation measures selected in this work. As discussed in Chapter 7, the tool provided a way of personalising the context based on the detected IMBs. A deeper look at how contextual data can be modelled and integrated with the cMIR method is worth further investigation in the future. Similarly, other behavioural metrics can be included as neutral measures whose significance (i.e. coefficients) in recognising interest is determined based on each person's data.

Nudging: As stated in the first chapter of this thesis, the presented work is the first step within a broader vision that seeks to benefit from individual interests in personalising behavioural nudging. The importance of detecting interests without directly asking individuals is aligned with the way behavioural nudging should be designed. Nudges are expected to be indirect, and hence the way of recognising interests should also be indirect. In Chapter D, we presented our paper which shows how behavioural knowledge of individuals can be used to personalise NHS nudges. Future research can go beyond behavioural knowledge and focus on personal interests when personalising nudges. The privacy concerns that may hinder the measurements and implementation of personalised nudges should be considered. Researchers may propose and evaluate various options similar to the ones that we discussed in the paper.

Bibliography

Abdullah, S., Matthews, M., Murnane, E.L., Gay, G., Choudhury, T., 2014. Towards circadian computing: "early to bed and early to rise" makes some of us unhealthy and sleep deprived, in: Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing, pp. 673–684.

Abdullah, S., Murnane, E.L., Matthews, M., Kay, M., Kientz, J.A., Gay, G., Choudhury, T., 2016. Cognitive rhythms: Unobtrusive and continuous sensing of alertness using a mobile phone, in: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Association for Computing Machinery, New York, NY, USA. p. 178–189. URL: https://doi.org/10.1145/2971648.2971712, doi:10.1145/2971648.2971712.

Adomavicius, G., Tuzhilin, A., 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering 17, 734–749. doi:10.1109/TKDE.2005.99.

Adomavicius, G., Tuzhilin, A., 2015. Context-aware recommender systems. Recommender Systems Handbook, Second Edition , 191–226doi:10.1007/978-1-4899-7637-6_6.

Adu, M.D., Malabu, U.H., Malau-Aduli, A.E.O., Malau-Aduli, B.S., 2018. Users' preferences and design recommendations to promote engagements with mobile apps for diabetes self-management: Multi-national perspectives. PLOS ONE 13, 1–21. URL: https://doi.org/10.1371/journal.pone.0208942, doi:10.1371/journal.pone.0208942.

Aeffner, F., Wilson, K., Martin, N.T., Black, J.C., Hendriks, C.L.L., Bolon, B., Rudmann, D.G., Gianani, R., Koegler, S.R., Krueger, J., Young, G.D., 2017. The Gold Standard Paradox in Digital Image Analysis: Manual Versus Automated Scoring as Ground Truth. Archives of Pathology and Laboratory Medicine 141, 1267–1275. URL: https://doi.org/10.5858/arpa.2016-0386-RA, doi:10.5858/arpa.2016-0386-RA.

Aggarwal, C.C., 2015. Data Mining: The Textbook. Springer Publishing Company, Incorporated.

Agrawal, R., Imieliundefinedski, T., Swami, A., 1993. Mining association rules between sets of items in large databases, in: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Association for Computing Machinery, New York, NY, USA. p. 207–216. URL: https://doi.org/10.1145/170035.170072, doi:10.1145/170035.170072.

Ahmed, T., Srivastava, A., 2019. Combining humans and machines for the future: A novel procedure to predict human interest. Future Generation Computer Systems 96, 713 – 730. URL: http://www.sciencedirect.com/science/article/pii/S0167739X18301353, doi:https://doi.org/10.1016/j.future.2018.01.043.

Ainley, M., Hidi, S., Berndorff, D., 2002. Interest, learning, and the psychological processes that mediate their relationship. Journal of Educational Psychology 94, 545–561. URL: https://doi.org/10.1037/0022-0663.94.3.545, doi:10.1037/0022-0663.94.3.545.

Akkerman, D.M., Vulperhorst, J.P., Akkerman, S.F., 2020. A developmental extension to the multidimensional structure of interests. Journal of Educational Psychology 112, 183.

Albanna, B.H., Moawad, I.F., Moussa, S.M., Sakr, M.A., 2015. Semantic Trajectories: A Survey from Modeling to Application. Springer International Publishing, Cham. pp. 59–76. URL: https://doi.org/10.1007/978-3-319-16667-4_4, doi:10.1007/978-3-319-16667-4_4.

Alharthi, R., Guthier, B., Guertin, C., El Saddik, A., 2017. A Dataset for Psychological Human Needs Detection from Social Networks. IEEE Access 5, 9109–9117. doi:10.1109/ACCESS.2017.2706084.

Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., Yang, D.Y., 2020. Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. NBER Working Paper .

Alvarez-Lozano, J., Osmani, V., Mayora, O., Frost, M., Bardram, J., Faurholt-Jepsen, M., Kessing, L.V., 2014. Tell me your apps and I will tell you your mood, in: Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments - PETRA '14, pp. 1–7. URL: http://dl.acm.org/citation.cfm?doid=2674396.2674408, doi:10.1145/2674396.2674408.

Amabile, T.M., Hill, K.G., Hennessey, B.A., Tighe, E.M., 1994. The work preference inventory: assessing intrinsic and extrinsic motivational orientations. Journal of personality and social psychology 66, 950.

Andersen, M., 2020. Early evidence on social distancing in response to covid-19 in the united states. Available at SSRN 3569368 .

Andrews, L., Drennan, J., Russell-Bennett, R., 2012. Linking perceived value of mobile marketing with the experiential consumption of mobile phones. European Journal of Marketing .

Andrienko, G., Andrienko, N., Fuchs, G., Raimond, A.M.O., Symanzik, J., Ziemlicki, C., 2013. Extracting semantics of individual places from movement data by analyzing temporal patterns of visits, in: 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, p. 9–16. URL: https://openaccess.city.ac.uk/id/eprint/2907/.

Anesbury, Z., Nenycz-Thiel, M., Dawes, J., Kennedy, R., 2016. How do shoppers behave online? an observational study of online grocery shopping. Journal of Consumer Behaviour 15, 261–270. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/cb.1566, doi:https://doi.org/10.1002/cb.1566, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/cb.1566.

Ann Renninger, K., 2000. Chapter 13 - individual interest and its implications for understanding intrinsic motivation, in: Sansone, C., Harackiewicz, J.M. (Eds.), Intrinsic and Extrinsic Motivation. Academic Press, San Diego. Educational Psychology, pp. 373–404. URL: https://www.sciencedirect.com/science/article/pii/B9780126190700500350, doi:https://doi.org/10.1016/B978-012619070-0/50035-0.

Aoude, L., Dawy, Z., Sharafeddine, S., Frenn, K., Jahed, K., 2018. Social-aware device-to-device offloading based on experimental mobility and content similarity models. Wireless Communications and Mobile Computing 2018, 5606829. URL: https://doi.org/10.1155/2018/5606829, doi:10.1155/2018/5606829.

Arnold, F., 1905. Interest and attention. Psychological Bulletin 2, 361.

Arora, S., Venkataraman, V., Zhan, A., Donohue, S., Biglan, K.M., Dorsey, E.R., Little, M.A., 2015. Detecting and monitoring the symptoms of parkinson's disease using smartphones: a pilot study. Parkinsonism & related disorders 21, 650–653.

Ashton, M.C., 2013. Chapter 11 - vocational interests, in: Ashton, M.C. (Ed.), Individual Differences and Personality (Second Edition). second edition ed.. Academic Press, San Diego, pp. 285–302. URL: https://www.sciencedirect.com/science/article/pii/B9780124160095000116, doi:https://doi.org/10.1016/B978-0-12-416009-5.00011-6.

Babiker, A., Baashar, Y., Alkahtani, A.A., Faye, I., Alkawsi, G., 2021. Towards detection of interest using physiological sensors. Applied Sciences 11. URL: https://www.mdpi.com/2076-3417/11/3/1318, doi:10.3390/app11031318.

Baglioni, M., Macedo, J., Renso, C., Wachowicz, M., 2008. An ontology-based approach for the semantic modelling and reasoning on trajectories, in: International conference on conceptual modeling, Springer. pp. 344–353.

Bandura, A., 1971. Social learning theory. volume 1. Prentice-hall Englewood Cliffs, NJ. doi:10.1111/j.1460-2466.1978.tb01621.x.

Bang, Y., Lee, D.J., Han, K., Hwang, M., Ahn, J.H., 2013. Channel capabilities, product characteristics, and the impacts of mobile channel introduction. Journal of Management Information Systems 30, 101–126. URL: https://doi.org/10.2753/MIS0742-1222300204, doi:10.2753/MIS0742-1222300204, arXiv:https://doi.org/10.2753/MIS0742-1222300204.

Barbuto Jr, J.E., Scholl, R.W., 1998. Motivation sources inventory: Development and validation of new scales to measure an integrative taxonomy of motivation. Psychological reports 82, 1011–1022.

Barnett, I., Torous, J., Reeder, H.T., Baker, J., Onnela, J.P., 2020. Determining sample size and length of follow-up for smartphone-based digital phenotyping studies. Journal of the American Medical Informatics Association doi:10.1093/jamia/ocaa201.

Barnett, I., Torous, J., Staples, P., Sandoval, L., Keshavan, M., Onnela, J.P., 2018. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. Neuropsychopharmacology 43, 1660–1666.

Bayerstadler, A., van Dijk, L., Winter, F., 2016. Bayesian multinomial latent variable modeling for fraud and abuse detection in health insurance. Insurance: Mathematics and Economics 71, 244–252. URL: https://www.sciencedirect.com/science/article/pii/S0167668715302845, doi:https://doi.org/10.1016/j.insmatheco.2016.09.013.

Beam, M.A., 2014. Automating the news: How personalized news recommender system design choices impact news reception. Communication Research 41, 1019–1041.

Beatriz, E., 2021. 16 fascinating online shopping statistics 2021. URL: https://www.bigcommerce.com/blog/online-shopping-statistics/#get-to-know-the-customers-who-shop-online.

Behera, R.K., Gunasekaran, A., Gupta, S., Kamboj, S., Bala, P.K., 2020. Personalized digital marketing recommender engine. Journal of Retailing and Consumer Services 53, 101799. URL: https://www.sciencedirect.com/science/article/pii/S0969698918307987, doi:https://doi.org/10.1016/j.jretconser.2019.03.026.

Benedict, B.M., 2001. Curiosity: A cultural history of early modern inquiry. University of Chicago Press.

Bentley, F., Tollmar, K., Stephenson, P., Levy, L., Jones, B., Robertson, S., Price, E., Catrambone, R., Wilson, J., 2013. Health mashups: Presenting statistical patterns between wellbeing data and context in natural language to promote behavior change. ACM Trans. Comput.-Hum. Interact. 20. URL: https://doi.org/10.1145/2503823, doi:10.1145/2503823.

Berlyne, D.E., 1966. Curiosity and exploration. Science 153, 25–33.

Botta, F., 2016. Quantifying human behaviour using complex social datasets. Ph.D. thesis. University of Warwick.

Boukhechba, M., Bouzouane, A., Bouchard, B., Gouin-Vallerand, C., Giroux, S., 2015. Online recognition of people's activities from raw gps data: Semantic trajectory data analysis, in: Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/2769493.2769498, doi:10.1145/2769493.2769498.

Boukhechba, M., Chow, P., Fua, K., Teachman, B., Barnes, L., 2018. Predicting social anxiety from global positioning system traces of college students: Feasibility study, Journal of Medical Internet Research. p. e10101.

Boukhechba, M., Huang, Y., Chow, P., Fua, K., Teachman, B.A., Barnes, L.E., 2017. Monitoring social anxiety from mobility and communication patterns, in: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, pp. 749–753.

Boytsov, A., Zaslavsky, A., Abdallah, Z., 2012. Where have you been? using location clustering and context awareness to understand places of interest, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 51,62.

Brambilla, M., Assor, A., Manzi, C., Regalia, C., 2015. Autonomous versus controlled religiosity: Family and group antecedents. The International Journal for the Psychology of Religion 25, 193–210. URL: https://doi.org/10.1080/10508619.2014.888902, doi:10.1080/10508619.2014.888902, arXiv:https://doi.org/10.1080/10508619.2014.888902.

Brasel, S.A., Gips, J., 2014. Tablets, touchscreens, and touchpads: How varying touch interfaces trigger psychological ownership and endowment. Journal of Consumer Psychology 24, 226 – 233. URL: http://www.sciencedirect.com/science/article/pii/S1057740813000934, doi:https://doi.org/10.1016/j.jcps.2013.10.003. sensory perception, embodiment, and grounded cognition: Implications for consumer behavior.

Briscese, G., Lacetera, N., Macis, M., Tonin, M., 2020. Compliance with covid-19 social-distancing measures in italy: the role of expectations and duration. Technical Report. National Bureau of Economic Research.

Burd, H., Coleman, C., 2020. Using behavioural insights to create a covid-19 text service for the nhs. URL: http://shorturl.at/hxJM1.

Burns, N.M., Begale, M., Duffecy, J., Gergle, D., Karr, J.C., Giangrande, E., Mohr, C.D., 2011. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 13, e55. URL: http://www.jmir.org/2011/3/e55/, doi:10.2196/jmir.1838.

Burrough, P.A., McDonnell, R., McDonnell, R.A., Lloyd, C.D., 2015. Principles of geographical information systems. Oxford university press.

Bye, D., Pushkar, D., Conway, M., 2007. Motivation, interest, and positive affect in traditional and nontraditional undergraduate students. Adult education quarterly 57, 141–158.

Cai, L., Boukhechba, M., Wu, C., Chow, P.I., Teachman, B.A., Barnes, L.E., Gerber, M.S., 2018. State affect recognition using smartphone sensing data, in: Proceedings of the 2018 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies, Association for Computing Machinery, New York, NY, USA. p. 120–125. URL: https://doi.org/10.1145/3278576.3284386, doi:10.1145/3278576.3284386.

Cao, L., 2014. Business model transformation in moving to a cross-channel retail strategy: A case study. International Journal of Electronic Commerce 18, 69–96. URL: https://doi.org/10.2753/JEC1086-4415180403, doi:10.2753/JEC1086-4415180403, arXiv:https://doi.org/10.2753/JEC1086-4415180403.

Caraban, A., Karapanos, E., Gonçalves, D., Campos, P., 2019. 23 ways to nudge: A review of technology-mediated nudging in human-computer interaction, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–15.

Chakri, S., Raghay, S., El Hadaj, S., 2015. Modeling, mining, and analyzing semantic trajectories: the process to extract meaningful behaviors of moving objects. International Journal of Computer Applications 124.

Chen, R.J., Orand, M., Choi, S.Y.L., Choi, L., 2018. An empirical exploration of mindfulness design using solo travel domain, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 1–12. URL: https://doi.org/10.1145/3173574.3173671, doi:10.1145/3173574.3173671.

Chikersal, P., Doryab, A., Tumminia, M., Villalba, D.K., Dutcher, J.M., Liu, X., Cohen, S., Creswell, K.G., Mankoff, J., Creswell, J.D., Goel, M., Dey, A.K., 2021. Detecting depression and predicting its onset using longitudinal symptoms captured by passive sensing: A machine learning approach with robust feature selection. ACM Trans. Comput.-Hum. Interact. 28. URL: https://doi.org/10.1145/3422821, doi:10.1145/3422821.

Cho, E., Myers, S.A., Leskovec, J., 2011. Friendship and mobility: user movement in location-based social networks, in: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1082–1090.

Cho, J., 2015. Roles of smartphone app use in improving social capital and reducing social isolation. Cyberpsychology, Behavior, and Social Networking 18, 350–355. URL: https://doi.org/10.1089/cyber.2014.0657, doi:10.1089/cyber.2014.0657, arXiv:https://doi.org/10.1089/cyber.2014.0657. pMID: 26075923.

Chopdar, P.K., Korfiatis, N., Sivakumar, V., Lytras, M.D., 2018. Mobile shopping apps adoption and perceived risks: A cross-country perspective utilizing the unified theory of acceptance and use of technology. Computers in Human Behavior 86, 109 – 128. URL: http://www.sciencedirect.com/science/article/pii/S0747563218301833, doi:https://doi.org/10.1016/j.chb.2018.04.017.

Cofer, C.N., Appley, M.H., 1964. Motivation: Theory and research. John Wiley, Oxford, England.

Cooper, J.O., Heron, T.E., Heward, W.L., 1987. Applied behavior analysis. Columbus : Merrill Pub. Co., Columbus.

Cornet, V.P., Holden, R.J., 2018. Systematic review of smartphone-based passive sensing for health and wellbeing. Journal of biomedical informatics 77, 120–132.

Culnane, C., Leins, K., Rubinstein, B.I., 2020. Tracking, tracing, trust: contemplating mitigating the impact of covid-19 through technological interventions. The Medical journal of Australia 213, 94—94.e1. URL: https://europepmc.org/articles/PMC7361786, doi:10.5694/mja2.50683.

Cummings, J.L., 1992. Depression and parkinson's disease: a review. The American journal of psychiatry .

Das Swain, V., Saha, K., Rajvanshy, H., Sirigiri, A., Gregg, J.M., Lin, S., Martinez, G.J., Mattingly, S.M., Mirjafari, S., Mulukutla, R., Nepal, S., Nies, K., Reddy, M.D., Robles-Granda, P., Campbell, A.T., Chawla, N.V., D'Mello, S., Dey, A.K., Jiang, K., Liu, Q., Mark, G., Moskal, E., Striegel, A., Tay, L., Abowd, G.D., De Choudhury, M., 2019. A multisensor person-centered approach to understand the role of daily activities in job performance with organizational personas. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 3. URL: https://doi.org/10.1145/3369828, doi:10.1145/3369828.

Dashdorj, Z., Serafini, L., Antonelli, F., Larcher, R., 2013. Semantic enrichment of mobile phone data records, in: Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/2541831.2541857, doi:10.1145/2541831.2541857.

DaSilva, A.W., Huckins, J.F., Wang, R., Wang, W., Wagner, D.D., Campbell, A.T., 2019. Correlates of stress in the college environment uncovered by the application of penalized generalized estimating equations to mobile sensing data. JMIR mHealth and uHealth 7, e12084. doi:10.2196/12084.

Dauer, W., Przedborski, S., 2003. Parkinson's disease: mechanisms and models. Neuron 39, 889–909.

De Pessemier, T., Courtois, C., Vanhecke, K., Van Damme, K., Martens, L., De Marez, L., 2016. A user-centric evaluation of context-aware recommendations for a mobile news service. Multimedia Tools and Applications 75, 3323–3351. URL: https://doi.org/10.1007/s11042-014-2437-9, doi:10.1007/s11042-014-2437-9.

Deci, E.L., Ryan, R.M., 2000. The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. Psychological inquiry 11, 227–268.

Dewey, J., 1913. Interest and effort in education. Houghton Mifflin.

Dey, A.K., Wac, K., Ferreira, D., Tassini, K., Hong, J.H., Ramos, J., 2011. Getting closer: An empirical investigation of the proximity of user to their smart phones, in: Proceedings of the 13th International Conference on Ubiquitous Computing, Association for Computing Machinery, New York, NY, USA. p. 163–172. URL: https://doi.org/10.1145/2030112.2030135, doi:10.1145/2030112.2030135.

Di Matteo, D., Fotinos, K., Lokuge, S., Mason, G., Sternat, T., Katzman, M.A., Rose, J., et al., 2021. Automated screening for social anxiety, generalized anxiety, and depression from objective smartphone-collected data: Cross-sectional study. Journal of Medical Internet Research 23, e28918. doi:10.2196/28918.

Difrancesco, S., Fraccaro, P., v. d. Veer, S.N., Alshoumr, B., Ainsworth, J., Bellazzi, R., Peek, N., 2016. Out-of-home activity recognition from gps data in schizophrenic patients, in: 2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS), pp. 324–328. doi:10.1109/CBMS.2016.54.

Ding, Z., Li, X., Jiang, C., Zhou, M., 2018. Objectives and state-of-the-art of location-based social network recommender systems. ACM Comput. Surv. 51. URL: https://doi.org/10.1145/3154526, doi:10.1145/3154526.

Do, T.M.T., Gatica-Perez, D., 2014. The places of our lives: Visiting patterns and automatic labeling from longitudinal smartphone data. IEEE Transactions on Mobile Computing 13, 638,648.

Dohn, N.B., Madsen, P.T., Malte, H., 2009. The situational interest of undergraduate students in zoophysiology. Advances in Physiology Education 33, 196–201. URL: https://doi.org/10.1152/advan.00038.2009, doi:10.1152/advan.00038.2009, arXiv:https://doi.org/10.1152/advan.00038.2009. pMID: 19745045.

Domoff, S.E., Banga, C.A., Borgen, A.L., Foley, R.P., Robinson, C., Avery, K., Gentile, D.A., 2021. Use of passive sensing to quantify adolescent mobile device usage: Feasibility, acceptability, and preliminary validation of the emoodie application. Human Behavior and Emerging Technologies 3, 63–74. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/hbe2.247, doi:https://doi.org/10.1002/hbe2.247, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbe2.247.

Doryab, A., Villalba, D.K., Chikersal, P., Dutcher, J.M., Tumminia, M., Liu, X., Cohen, S., Creswell, K., Mankoff, J., Creswell, J.D., et al., 2019. Identifying behavioral phenotypes of loneliness and social isolation with passive sensing: statistical analysis, data mining and machine learning of smartphone and fitbit data. JMIR mHealth and uHealth 7, e13209. doi:10.2196/13209.

Draijer, J., Bakker, A., Slot, E., Akkerman, S., 2020. The multidimensional structure of interest. Frontline Learning Research 8, 18–36.

Drossos, D.A., Giaglis, G.M., Vlachos, P.A., Zamani, E.D., Lekakos, G., 2013. Consumer responses to sms advertising: Antecedents and consequences. International Journal of Electronic Commerce 18, 105–136. URL: https://doi.org/10.2753/JEC1086-4415180104, doi:10.2753/JEC1086-4415180104, arXiv:https://doi.org/10.2753/JEC1086-4415180104.

Duzgun, F., Yamamoto, G., 2017. Sms promotion effects on consumer behaviour: A turkish case. Int J Econ Manag Sci 6, 2.

Dyson, R.G., 2004. Strategic development and swot analysis at the university of warwick. European journal of operational research 152, 631–640.

Eagle, N., Sandy Pentland, A., 2006. Reality mining: Sensing complex social systems. Personal Ubiquitous Comput. 10, 255–268. URL: https://doi.org/10.1007/s00779-005-0046-3, doi:10.1007/s00779-005-0046-3.

Eccles, J.S., Wigfield, A., 2002. Motivational Beliefs, Values, and Goals. Annual Review of Psychology 53, 109–132. doi:10.1146/annurev.psych.53.100901.135153.

Epstein, D., Cordeiro, F., Bales, E., Fogarty, J., Munson, S., 2014. Taming data complexity in lifelogs: Exploring visual cuts of personal informatics data, in: Proceedings of the 2014 Conference on Designing Interactive Systems, Association for Computing Machinery, New York, NY, USA. p. 667–676. URL: https://doi.org/10.1145/2598510.2598558, doi:10.1145/2598510.2598558.

Espay, A.J., Bonato, P., Nahab, F.B., Maetzler, W., Dean, J.M., Klucken, J., Eskofier, B.M., Merola, A., Horak, F., Lang, A.E., et al., 2016. Technology in parkinson's disease: challenges and opportunities. Movement Disorders 31, 1272–1282.

Estabrooks, A., Jo, T., Japkowicz, N., 2004. A multiple resampling method for learning from imbalanced data sets. Computational Intelligence 20, 18–36. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0824-7935.2004.t01-1-00228.x, doi:https://doi.org/10.1111/j.0824-7935.2004.t01-1-00228.x.

Falaki, H., Mahajan, R., Kandula, S., Lymberopoulos, D., Govindan, R., Estrin, D., 2010. Diversity in smartphone usage, in: Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, Association for Computing Machinery, New York, NY, USA. p. 179–194. URL: https://doi.org/10.1145/1814433.1814453, doi:10.1145/1814433.1814453.

Farhan, A.A., Yue, C., Morillo, R., Ware, S., Lu, J., Bi, J., Kamath, J., Russell, A., Bamis, A., Wang, B., 2016. Behavior vs. introspection: refining prediction of clinical depression via smartphone sensing data, in: 2016 IEEE Wireless Health (WH), IEEE. pp. 1–8.

Farrahi, K., Gatica-Perez, D., 2014. A probabilistic approach to mining mobile phone data sequences. Personal and ubiquitous computing 18, 223–238.

Ferreira, D., Kostakos, V., Dey, A.K., 2015. Aware: mobile context instrumentation framework. Frontiers in ICT 2, 6.

Fileto, R., May, C., Renso, C., Pelekis, N., Klein, D., Theodoridis, Y., 2015. The baquara2 knowledge-based framework for semantic enrichment and analysis of movement data. Data and Knowledge Engineering 98, 104–122. URL: https://www.sciencedirect.com/science/article/pii/S0169023X15000555, doi:https://doi.org/10.1016/j.datak.2015.07.010. research on conceptual modeling.

Fischer, J.E., Yee, N., Bellotti, V., Good, N., Benford, S., Greenhalgh, C., 2010. Effects of content and time of delivery on receptivity to mobile interruptions, in: Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services, Association for Computing Machinery, New York, NY, USA. p. 103–112. URL: https://doi.org/10.1145/1851600.1851620, doi:10.1145/1851600.1851620.

Fishbach, A., Hofmann, W., 2015. Nudging self-control: A smartphone intervention of temptation anticipation and goal resolution improves everyday goal progress. Motivation Science 1, 137–150. doi:10.1037/mot0000022.

Fogg, B., 2012. The health solution. Recorded Keynote, published on the online video-sharing platform YouTube: https://www.youtube.com/watch?v=fqUSjHjIEFg [Last accessed: September 2020].

Frank, E., Hall, M., Holmes, G., Kirkby, R., Pfahringer, B., Witten, I.H., Trigg, L., 2009. Weka-a machine learning workbench for data mining, in: Data mining and knowledge discovery handbook. Springer, pp. 1269–1277.

Frey, R.M., Xu, R., Ammendola, C., Moling, O., Giglio, G., Ilic, A., 2017. Mobile recommendations based on interest prediction from consumer's installed apps–insights from a large-scale field study. Information Systems 71, 152–163. URL: https://www.sciencedirect.com/science/article/pii/S0306437917304611, doi:https://doi.org/10.1016/j.is.2017.08.006.

Frick, R.W., 1985. Communicating emotion: The role of prosodic features. Psychological bulletin 97, 412.

Friedman, J.H., Brown, R.G., Comella, C., Garber, C.E., Krupp, L.B., Lou, J.S., Marsh, L., Nail, L., Shulman, L., Taylor, C.B., 2007. Fatigue in parkinson's disease: a review. Movement disorders: official journal of the Movement Disorder Society 22, 297–308.

Gao, S., Li, W., Song, L.J., Zhang, X., Lin, M., Lu, S., 2020. Personalitysensing: A multi-view multi-task learning approach for personality detection based on smartphone usage, in: Proceedings of the 28th ACM International Conference on Multimedia, Association for Computing Machinery, New York, NY, USA. p. 2862–2870. URL: https://doi.org/10.1145/3394171.3413591, doi:10.1145/3394171.3413591.

Gaonkar, R., Tavakol, M., Brefeld, U., 2018. Mdp-based itinerary recommendation using geo-tagged social media, in: International Symposium on Intelligent Data Analysis, Springer. pp. 111–123.

Gneezy, U., Meier, S., Rey-Biel, P., 2011. When and Why Incentives (Don't) Work to Modify Behavior. Journal of Economic Perspectives 25, 191–210. URL: http://pubs.aeaweb.org/doi/10.1257/jep.25.4.191, doi:10.1257/jep.25.4.191, arXiv:00368075.

Gomez, V., Allemand, M., Grob, A., 2012. Neuroticism, extraversion, goals, and subjective well-being: Exploring the relations in young, middle-aged, and older adults. Journal of Research in Personality 46, 317 – 325. URL: http://www.sciencedirect.com/science/article/pii/S0092656612000505, doi:https://doi.org/10.1016/j.jrp.2012.03.001.

Gomez-Uribe, C.A., Hunt, N., 2016. The netflix recommender system: Algorithms, business value, and innovation. ACM Trans. Manage. Inf. Syst. 6. doi:10.1145/2843948.

González-Cutre, D., Sicilia, Á., Sierra, A.C., Ferriz, R., Hagger, M.S., 2016. Understanding the need for novelty from the perspective of self-determination theory. Personality and individual differences 102, 159–169.

Grbovic, M., Radosavljevic, V., Djuric, N., Bhamidipati, N., Savla, J., Bhagwan, V., Sharp, D., 2015. E-commerce in your inbox: Product recommendations at scale, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, NY, USA. p. 1809–1818. URL: https://doi.org/10.1145/2783258.2788627, doi:10.1145/2783258.2788627.

Groß, M., 2020. Differences between mobile and non-mobile buyers: comparing attitudinal, motive-related, and media behaviour. International Journal of Electronic Marketing and Retailing 11, 50–80. URL: https://www.inderscienceonline.com/doi/abs/10.1504/IJEMR.2020.106433, doi:10.1504/IJEMR.2020.106433.

Gulla., J.A., Fidjestøl., A.D., Su., X., Castejon., H., 2014. Implicit user profiling in news recommender systems, in: Proceedings of the 10th International Conference on Web Information Systems and Technologies - Volume 2: WEBIST,, INSTICC. SciTePress. pp. 185–192. doi:10.5220/0004860801850192.

Gupta, R., Kondapally, R., Guha, S., 2019. Large-scale information extraction from emails with data constraints, in: Madria, S., Fournier-Viger, P., Chaudhary, S., Reddy, P.K. (Eds.), Big Data Analytics, Springer International Publishing, Cham. pp. 124–139.

Gustafson, D.H., McTavish, F.M., Chih, M.Y., Atwood, A.K., Johnson, R.A., Boyle, M.G., Levy, M.S., Driscoll, H., Chisholm, S.M., Dillenburg, L., et al., 2014. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA psychiatry 71, 566–572.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The weka data mining software: An update. SIGKDD Explor. Newsl. 11, 10–18. URL: https://doi.org/10.1145/1656274.1656278, doi:10.1145/1656274.1656278.

Harackiewicz, J.M., Smith, J.L., Priniski, S.J., 2016. Interest matters: The importance of promoting interest in education. Policy Insights from the Behavioral and Brain Sciences 3, 220–227. URL: https://doi.org/10.1177/2372732216655542, doi:10.1177/2372732216655542, arXiv:https://doi.org/10.1177/2372732216655542. pMID: 29520371.

Harari, G.M., Müller, S.R., Aung, M.S., Rentfrow, P.J., 2017. Smartphone sensing methods for studying behavior in everyday life. Current Opinion in Behavioral Sciences 18, 83–90. URL: https://www.sciencedirect.com/science/article/pii/S2352154617300505, doi:https://doi.org/10.1016/j.cobeha.2017.07.018. big data in the behavioural sciences.

Hassan, E., 2006. Recall bias can be a threat to retrospective and prospective research designs. The Internet Journal of Epidemiology 3, 339–412.

Hawton, A., Green, C., Dickens, A.P., Richards, S.H., Taylor, R.S., Edwards, R., Greaves, C.J., Campbell, J.L., 2011. The impact of social isolation on the health status and health-related quality of life of older people. Quality of Life Research 20, 57–67. URL: https://doi.org/10.1007/s11136-010-9717-2, doi:10.1007/s11136-010-9717-2.

Hekler, E.B., Michie, S., Pavel, M., Rivera, D.E., Collins, L.M., Jimison, H.B., Garnett, C., Parral, S., Spruijt-Metz, D., 2016. Advancing Models and Theories for Digital Behavior Change Interventions. American Journal of Preventive Medicine 51, 825–832. doi:10.1016/j.amepre.2016.06.013.

Hidi, S., 1990. Interest and its contribution as a mental resource for learning. Review of Educational research 60, 549–571.

Hidi, S., 2000. Chapter 11 - an interest researcher's perspective: The effects of extrinsic and intrinsic factors on motivation, in: Sansone, C., Harackiewicz, J.M. (Eds.), Intrinsic and Extrinsic Motivation. Academic Press, San Diego. Educational Psychology, pp. 309–339. URL: https://www.sciencedirect.com/science/article/pii/B9780126190700500337, doi:https://doi.org/10.1016/B978-012619070-0/50033-7.

Hidi, S., Renninger, K.A., 2006. The Four-Phase Model of Interest Development. Educational Psychologist 41, 111–127. URL: http://www.tandfonline.com/doi/abs/10.1207/s15326985ep4102{_}4, doi:10.1207/s15326985ep4102_4, arXiv:0710.4428v1.

Houghton, D., Pressey, A., Istanbulluoglu, D., 2020. Who needs social networking? an empirical enquiry into the capability of facebook to meet human needs and satisfaction with life. Computers in Human Behavior 104, 106153.

Huang, J., Zhou, L., 2018. Timing of web personalization in mobile shopping: A perspective from uses and gratifications theory. Computers in Human Behavior 88, 103 – 113. URL: http://www.sciencedirect.com/science/article/pii/S0747563218303121, doi:https://doi.org/10.1016/j.chb.2018.06.035.

Huang, M., Fang, Z., Xiong, S., Zhang, T., 2019. Interest-driven outdoor advertising display location selection using mobile phone data. IEEE Access 7, 30878–30889. doi:10.1109/ACCESS.2019.2903277.

Huang, Y., Xiong, H., Leach, K., Zhang, Y., Chow, P., Fua, K., Teachman, B.A., Barnes, L.E., 2016. Assessing social anxiety using gps trajectories and point-of-interest data, in: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Association for Computing Machinery, New York, NY, USA. p. 898–903. URL: https://doi.org/10.1145/2971648.2971761, doi:10.1145/2971648.2971761.

Hui, S.K., Huang, Y., Suher, J., Inman, J.J., 2013. Deconstructing the "first moment of truth": Understanding unplanned consideration and purchase conversion using in-store video tracking. Journal of Marketing Research 50, 445–462. URL: https://doi.org/10.1509/jmr.12.0065, doi:10.1509/jmr.12.0065, arXiv:https://doi.org/10.1509/jmr.12.0065.

Hull, C.L., 1943. Principles of behavior: an introduction to behavior theory. Appleton-Century, Oxford, England.

Hull, C.L., 1952. A behavior system; an introduction to behavior theory concerning the individual organism. Yale University Press, New Haven, CT, US.

Ibrahim, A., Clinch, S., Harper, S., 2021a. Recognising intrinsic motivation using smartphone trajectories. International Journal of Human-Computer Studies 153, 102650. URL: https://www.sciencedirect.com/science/article/pii/S1071581921000689, doi:https://doi.org/10.1016/j.ijhcs.2021.102650.

Ibrahim, A., Zhang, H., Clinch, S., Poliakoff, E., Parsia, B., Harper, S., 2021b. Digital phenotypes for understanding individuals' compliance with covid-19 policies and personalized nudges: Longitudinal observational study. JMIR Form Res 5, e23461. URL: https://formative.jmir.org/2021/5/e23461, doi:10.2196/23461.

Insel, T.R., 2017. Digital Phenotyping: Technology for a New Science of Behavior. JAMA 318, 1215–1216. doi:10.1001/jama.2017.11295.

Jacobson, N.C., Summers, B., Wilhelm, S., 2020. Digital biomarkers of social anxiety severity: digital phenotyping using passive smartphone sensors. Journal of medical Internet research 22, e16875. doi:10.2196/16875.

Jain, S.H., Powers, B.W., Hawkins, J.B., Brownstein, J.S., 2015. The digital phenotype. Nature biotechnology 33, 462–463.

Jia, X., Wang, A., Li, X., Xun, G., Xu, W., Zhang, A., 2015. Multi-modal learning for video recommendation based on mobile application usage, in: 2015 IEEE International Conference on Big Data (Big Data), pp. 837–842. doi:10.1109/BigData.2015.7363830.

Joachims, T., 1996. A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization. Technical Report. Carnegie-mellon univ pittsburgh pa dept of computer science.

Kang, S., Jung, J., 2014. Mobile communication for human needs: A comparison of smartphone use between the us and korea. Computers in Human Behavior 35, 376–387.

Karatzoglou, A., Koehler, D., Beigl, M., 2018. Purpose-of-visit-driven semantic similarity analysis on semantic trajectories for enhancing the future location prediction, in: 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), pp. 100–106. doi:10.1109/PERCOMW.2018.8480230.

Khan, M.M., Ibrahim, R., Ghani, I., 2017. Cross domain recommender systems: A systematic literature review. ACM Comput. Surv. 50. URL: https://doi.org/10.1145/3073565, doi:10.1145/3073565.

Khan, W.Z., Xiang, Y., Aalsalem, M.Y., Arshad, Q., 2013. Mobile phone sensing systems: A survey. IEEE Communications Surveys & Tutorials 15, 402–427.

Khusumanegara, P., Mafrur, R., Choi, D., 2015. Profiler for smartphone users interests using modified hierarchical agglomerative clustering algorithm based on browsing history, in: Khalil, I., Neuhold, E., Tjoa, A.M., Xu, L.D., You, I. (Eds.), Information and Communication Technology, Springer International Publishing, Cham. pp. 89–96.

Kim, M., Kim, J., Choi, J., Trivedi, M., 2017. Mobile shopping through applications: Understanding application possession and mobile purchase. Journal of Interactive Marketing 39, 55 – 68. URL: http://www.sciencedirect.com/science/article/pii/S1094996817300154, doi:https://doi.org/10.1016/j.intmar.2017.02.001.

Kiukkonen, N., Blom, J., Dousse, O., Gatica-Perez, D., Laurila, J., 2010. Towards rich mobile phone datasets: Lausanne data collection campaign. Proc. ICPS, Berlin 68.

Klimt, B., Yang, Y., 2004. The enron corpus: A new dataset for email classification research, in: Proceedings of the 15th European Conference on Machine Learning, Springer-Verlag, Berlin, Heidelberg. p. 217–226. URL: https://doi.org/10.1007/978-3-540-30115-8_22, doi:10.1007/978-3-540-30115-8_22.

Kocayusufoglu, F., Sheng, Y., Vo, N., Wendt, J., Zhao, Q., Tata, S., Najork, M., 2019. Riser: Learning better representations for richly structured emails, in: The World Wide Web Conference, Association for Computing Machinery, New York, NY, USA. p. 886–895. URL: https://doi.org/10.1145/3308558.3313720, doi:10.1145/3308558.3313720.

Kooti, F., Lerman, K., Aiello, L.M., Grbovic, M., Djuric, N., Radosavljevic, V., 2016. Portrait of an online shopper: Understanding and predicting consumer behavior, in:

Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, Association for Computing Machinery, New York, NY, USA. p. 205–214. URL: https://doi.org/10.1145/2835776.2835831, doi:10.1145/2835776.2835831.

Kosters, M., Van der Heijden, J., 2015. From mechanism to virtue: Evaluating Nudge theory. Evaluation 21, 276–291. URL: https://doi.org/10.1177/1356389015590218, doi:10.1177/1356389015590218.

Krapp, A., Hidi, S., Renninger, K.A., 1992. Interest, learning, and development.. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US. The role of interest in learning and development., pp. 3–25.

Kroenke, K., Spitzer, R.L., 2002. The phq-9: A new depression diagnostic and severity measure. Psychiatric Annals 32, 509–515. URL: https://journals.healio.com/doi/abs/10.3928/0048-5713-20020901-06, doi:10.3928/0048-5713-20020901-06.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The phq-9. Journal of General Internal Medicine 16, 606–613. URL: https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1525-1497.2001.016009606.x, doi:https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

Krueger, R., Thom, D., Ertl, T., 2015. Semantic Enrichment of Movement Behavior with Foursquare - A Visual Analytics Approach. IEEE Transactions on Visualization and Computer Graphics 21, 903–915. doi:10.1109/TVCG.2014.2371856.

Lane, N.D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., Campbell, A.T., 2010. A survey of mobile phone sensing. IEEE Communications magazine 48, 140–150.

LeDell, E., Poirier, S., 2020. H2O AutoML: Scalable automatic machine learning. 7th ICML Workshop on Automated Machine Learning (AutoML) URL: https://www.automl.org/wp-content/uploads/2020/07/AutoML_2020_paper_61.pdf.

Lee, H.C., So, H.J., 2014. Exploring interest-based personalization in mobile advertising, in: Proceedings of the 18th International Academic MindTrek Conference: Media Business, Management, Content and Services, Association for Computing Machinery, New York, NY, USA. p. 142–144. URL: https://doi.org/10.1145/2676467.2676510, doi:10.1145/2676467.2676510.

Lemon, K.N., Verhoef, P.C., 2016. Understanding customer experience throughout the customer journey. Journal of Marketing 80, 69–96. URL: https://doi.org/10.1509/jm.15.0420, doi:10.1509/jm.15.0420, arXiv:https://doi.org/10.1509/jm.15.0420.

Leong, C.K., Lee, Y.H., Mak, W.K., 2012. Mining sentiments in sms texts for teaching evaluation. Expert Systems with Applications 39, 2584 – 2589. URL: http://www.sciencedirect.com/science/article/pii/S0957417411012449, doi:https://doi.org/10.1016/j.eswa.2011.08.113.

Levine, L.M., Gwak, M., Kärkkäinen, K., Fazeli, S., Zadeh, B., Peris, T., Young, A.S., Sarrafzadeh, M., 2020. Anxiety detection leveraging mobile passive sensing, in: Alam, M.M., Hämäläinen, M., Mucchi, L., Niazi, I.K., Le Moullec, Y. (Eds.), Body Area Networks. Smart IoT and Big Data for Intelligent Health, Springer International Publishing, Cham. pp. 212–225.

Li, D., Gu, Y., Kamijo, S., 2018a. Smartphone based lifelog with meaningful place detection, in: 2018 IEEE International Conference on Consumer Electronics (ICCE), pp. 1–5. doi:10.1109/ICCE.2018.8326116.

Li, Q., Zheng, Y., Xie, X., Chen, Y., Liu, W., Ma, W.Y., 2008. Mining user similarity based on location history, in: Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems, pp. 1–10.

Li, X., Cong, G., Li, X.L., Pham, T.A.N., Krishnaswamy, S., 2015a. Rank-geofm: A ranking based geographical factorization method for point of interest recommendation, in: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 433–442.

Li, X., Jiang, M., Hong, H., Liao, L., 2017a. A time-aware personalized point-of-interest recommendation via high-order tensor factorization. ACM Transactions on Information Systems (TOIS) 35, 1–23.

Li, X., Xu, G., Chen, E., Zong, Y., 2015b. Learning recency based comparative choice towards point-of-interest recommendation. Expert Systems with Applications 42, 4274–4283.

Li, Y., Xu, P., Lagun, D., Navalpakkam, V., 2017b. Towards measuring and inferring user interest from gaze, in: Proceedings of the 26th International Conference on

World Wide Web Companion, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE. p. 525–533. URL: https://doi.org/10.1145/3041021.3054182, doi:10.1145/3041021.3054182.

Li, Y., Yang, Z., Guo, Y., Chen, X., Agarwal, Y., Hong, J.I., 2018b. Automated extraction of personal knowledge from smartphone push notifications, in: 2018 IEEE International Conference on Big Data (Big Data), pp. 733–742. doi:10.1109/BigData.2018.8621866.

LiKamWa, R., Liu, Y., Lane, N.D., Zhong, L., 2013. Moodscope: Building a mood sensor from smartphone usage patterns, in: Proceeding of the 11th annual international conference on Mobile systems, applications, and services, pp. 389–402.

Lim, K.H., Chan, J., Leckie, C., Karunasekera, S., 2015. Personalized tour recommendation based on user interests and points of interest visit durations, in: Twenty-Fourth International Joint Conference on Artificial Intelligence.

Lind, M.N., Byrne, M.L., Wicks, G., Smidt, A.M., Allen, N.B., 2018. The effortless assessment of risk states (ears) tool: An interpersonal approach to mobile sensing. JMIR Ment Health 5, e10334. URL: http://mental.jmir.org/2018/3/e10334/, doi:10.2196/10334.

Linnenbrink-Garcia, L., Patall, E.A., Messersmith, E.E., 2013. Antecedents and consequences of situational interest. British Journal of Educational Psychology 83, 591–614. URL: https://bpspsychub.onlinelibrary.wiley.com/doi/abs/10.1111/j.2044-8279.2012.02080.x, doi:https://doi.org/10.1111/j.2044-8279.2012.02080.x.

Liu, S., Jiang, Y., Striegel, A., 2013. Face-to-face proximity estimationusing bluetooth on smartphones. IEEE Transactions on Mobile Computing 13, 811–823.

Liu, S., Striegel, A., 2011. Accurate extraction of face-to-face proximity using smartphones and bluetooth, in: 2011 Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN), IEEE. pp. 1–5.

Liu, Y., Liu, C., Liu, B., Qu, M., Xiong, H., 2016. Unified point-of-interest recommendation with temporal interval assessment, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 1015–1024.

Logesh, R., Subramaniyaswamy, V., 2017. Learning recency and inferring associations in location based social network for emotion induced point-of-interest recommendation. Journal of Information Science & Engineering 33.

Lops, P., de Gemmis, M., Semeraro, G., 2011. Content-based Recommender Systems: State of the Art and Trends. Springer US, Boston, MA. pp. 73–105. URL: https://doi.org/10.1007/978-0-387-85820-3_3, doi:10.1007/978-0-387-85820-3_3.

Loseto, G., Ruta, M., Scioscia, F., Sciascio, E.D., Mongiello, M., 2013. Mining the user profile from a smartphone: a multimodal agent framework, in: WOA@AI*IA.

Lu, W., Chen, S., Li, K., Lakshmanan, L.V., 2014. Show me the money: dynamic recommendations for revenue maximization. Proceedings of the VLDB Endowment 7, 1785–1796.

Mack, D.L., DaSilva, A.W., Rogers, C., Hedlund, E., Murphy, E.I., Vojdanovski, V., Plomp, J., Wang, W., Nepal, S.K., Holtzheimer, P.E., et al., 2021. Mental health and behavior of college students during the covid-19 pandemic: Longitudinal mobile smartphone and ecological momentary assessment study, part ii. Journal of Medical Internet Research 23, e28892. doi:10.2196/28892.

Maslow, A.H., 1943. A theory of human motivation. Psychological Review 50, 370–396. doi:10.1037/h0054346.

Matthews, O., Davies, A., Vigo, M., Harper, S., 2020. Unobtrusive arousal detection on the web using pupillary response. International Journal of Human-Computer Studies 136, 102361. URL: https://www.sciencedirect.com/science/article/pii/S1071581919301181, doi:https://doi.org/10.1016/j.ijhcs.2019.09.003.

Mattick, R.P., Clarke, J., 1998. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety11editor's note: This article was written before the development of some contemporary measures of social phobia, such as the social phobia and anxiety inventory (turner et al., 1989). we have invited this article for publication because of the growing interest in the scales described therein. s.t. Behaviour Research and Therapy 36, 455–470. URL: https://www.sciencedirect.com/science/article/pii/S0005796797100316, doi:https://doi.org/10.1016/S0005-7967(97)10031-6.

McAuley, E., Duncan, T., Tammen, V.V., 1989. Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: A confirmatory factor analysis. Research Quarterly for Exercise and Sport 60, 48–58. URL: https://doi.org/10.1080/02701367.1989.10607413, doi:10.1080/02701367.1989.10607413, arXiv:https://doi.org/10.1080/02701367.1989.10607413. pMID: 2489825.

McClelland, D.C., 1987. Human motivation. Cambridge University Press, New York, NY, US.

McCusker, K., Gunaydin, S., 2015. Research using qualitative, quantitative or mixed methods and choice based on the research. Perfusion 30, 537–542.

McLeod, S., 2007. Maslow's hierarchy of needs. Simply psychology 1, 1–8.

Meegahapola, L., Ruiz-Correa, S., Gatica-Perez, D., 2020a. Alone or with others? understanding eating episodes of college students with mobile sensing, in: 19th International Conference on Mobile and Ubiquitous Multimedia, Association for Computing Machinery, New York, NY, USA. p. 162–166. URL: https://doi.org/10.1145/3428361.3428463, doi:10.1145/3428361.3428463.

Meegahapola, L., Ruiz-Correa, S., Gatica-Perez, D., 2020b. Protecting mobile food diaries from getting too personal, in: 19th International Conference on Mobile and Ubiquitous Multimedia, Association for Computing Machinery, New York, NY, USA. p. 212–222. URL: https://doi.org/10.1145/3428361.3428468, doi:10.1145/3428361.3428468.

Mehrotra, A., Hendley, R., Musolesi, M., 2016a. Towards multi-modal anticipatory monitoring of depressive states through the analysis of human-smartphone interaction, in: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, Association for Computing Machinery, New York, NY, USA. p. 1132–1138. URL: https://doi.org/10.1145/2968219.2968299, doi:10.1145/2968219.2968299.

Mehrotra, A., Müller, S.R., Harari, G.M., Gosling, S.D., Mascolo, C., Musolesi, M., Rentfrow, P.J., 2017a. Understanding the role of places and activities on mobile phone interaction and usage patterns. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 1. URL: https://doi.org/10.1145/3131901, doi:10.1145/3131901.

Mehrotra, A., Musolesi, M., Hendley, R., Pejovic, V., 2015. Designing content-driven intelligent notification mechanisms for mobile applications, in: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Association for Computing Machinery, New York, NY, USA. p. 813–824. URL: https://doi.org/10.1145/2750858.2807544, doi:10.1145/2750858.2807544.

Mehrotra, A., Pejovic, V., Vermeulen, J., Hendley, R., Musolesi, M., 2016b. My phone and me: Understanding people's receptivity to mobile notifications, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 1021–1032. URL: https://doi.org/10.1145/2858036.2858566, doi:10.1145/2858036.2858566.

Mehrotra, A., Tsapeli, F., Hendley, R., Musolesi, M., 2017b. Mytraces: Investigating correlation and causation between users' emotional states and mobile phone interaction. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 1. URL: https://doi.org/10.1145/3130948, doi:10.1145/3130948.

Meixner, C., Baumann, H., Wollesen, B., 2020. Personality traits, gamification and features to develop an app to reduce physical inactivity. Information 11. URL: https://www.mdpi.com/2078-2489/11/7/367, doi:10.3390/info11070367.

Melcher, J., Lavoie, J., Hays, R., D'Mello, R., Rauseo-Ricupero, N., Camacho, E., Rodriguez-Villa, E., Wisniewski, H., Lagan, S., Vaidyam, A., Torous, J., 2021. Digital phenotyping of student mental health during covid-19: an observational study of 100 college students. Journal of American College Health 0, 1–13. URL: https://doi.org/10.1080/07448481.2021.1905650, doi:10.1080/07448481.2021.1905650, arXiv:https://doi.org/10.1080/07448481.2021.1905650. pMID: 33769927.

Memedi, M., Nyholm, D., Johansson, A., Pålhagen, S., Willows, T., Widner, H., Linder, J., Westin, J., 2015. Validity and responsiveness of at-home touch screen assessments in advanced parkinson's disease. IEEE Journal of Biomedical and Health Informatics 19, 1829–1834.

Michie, S., Van Stralen, M.M., West, R., 2011. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implementation science 6, 42.

Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013. Efficient estimation of word representations in vector space. arXiv:1301.3781.

Miller, G., 2012. The Smartphone Psychology Manifesto. Perspectives on Psychological Science 7, 221–237. URL: http://journals.sagepub.com/doi/10.1177/1745691612441215, doi:10.1177/1745691612441215.

Mills, S., 2020. Personalized nudging. Behavioural Public Policy , 1–10doi:10.1017/bpp.2020.7.

Miluzzo, E., Lane, N.D., Eisenman, S.B., Campbell, A.T., 2007. Cenceme – injecting sensing presence into social networking applications, in: Kortuem, G., Finney, J., Lea, R., Sundramoorthy, V. (Eds.), Smart Sensing and Context, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 1–28.

Min, J.K., Wiese, J., Hong, J.I., Zimmerman, J., 2013. Mining smartphone data to classify life-facets of social relationships, in: Proceedings of the 2013 Conference on Computer Supported Cooperative Work, Association for Computing Machinery, New York, NY, USA. p. 285–294. URL: https://doi.org/10.1145/2441776.2441810, doi:10.1145/2441776.2441810.

Mirjafari, S., Masaba, K., Grover, T., Wang, W., Audia, P., Campbell, A.T., Chawla, N.V., Swain, V.D., Choudhury, M.D., Dey, A.K., D'Mello, S.K., Gao, G., Gregg, J.M., Jagannath, K., Jiang, K., Lin, S., Liu, Q., Mark, G., Martinez, G.J., Mattingly, S.M., Moskal, E., Mulukutla, R., Nepal, S., Nies, K., Reddy, M.D., Robles-Granda, P., Saha, K., Sirigiri, A., Striegel, A., 2019. Differentiating higher and lower job performers in the workplace using mobile sensing. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 3. URL: https://doi.org/10.1145/3328908, doi:10.1145/3328908.

Mirsky, Y., Shabtai, A., Rokach, L., Shapira, B., Elovici, Y., 2016. Sherlock vs moriarty: A smartphone dataset for cybersecurity research, in: Proceedings of the 2016 ACM workshop on Artificial intelligence and security, pp. 1–12.

Mohr, D.C., Burns, M.N., Schueller, S.M., Clarke, G., Klinkman, M., 2013. Behavioral intervention technologies: Evidence review and recommendations for future research in mental health. General Hospital Psychiatry 35, 332–338. URL: https://www.sciencedirect.com/science/article/pii/S0163834313000698, doi:https://doi.org/10.1016/j.genhosppsych.2013.03.008.

Mohr, D.C., Zhang, M., Schueller, S.M., 2017. Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. Annual Review of Clinical Psychology 13, 23–47. URL: https://doi.org/10.1146/annurev-clinpsy-032816-044949, doi:10.1146/annurev-clinpsy-032816-044949, arXiv:https://doi.org/10.1146/annurev-clinpsy-032816-044949. pMID: 28375728.

Mohsin, M., 2021. 10 online shopping statistics you need to know in 2021 [infographic]. URL: https://www.oberlo.co.uk/blog/online-shopping-statistics.

Monteiro, V., Mata, L., Peixoto, F., 2015. Intrinsic Motivation Inventory: Psychometric Properties in the Context of First Language and Mathematics Learning. Psicologia: Reflexão e Crítica 28, 434–443. doi:10.1590/1678-7153.201528302.

Montoliu, R., Gatica-Perez, D., 2010. Discovering human places of interest from multimodal mobile phone data, in: Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia, pp. 1–10.

Morgan, C., Murray, H., 1935. A method for investigating fantasies: The Thematic Apperception Test. Archives of Neurology and Psychiatry 34, 289. URL: http://dx.doi.org/10.1001/archneurpsyc.1935.02250200049005.

Moshe, I., Terhorst, Y., Opoku Asare, K., Sander, L.B., Ferreira, D., Baumeister, H., Mohr, D.C., Pulkki-Råback, L., 2021. Predicting symptoms of depression and anxiety using smartphone and wearable data. Frontiers in psychiatry 12, 625247–625247. doi:10.3389/fpsyt.2021.625247. 33584388[pmid].

Moukaddam, N., TRUONG, A., CAO, J., SHAH, A., SABHARWAL, A., 2019. Findings from a trial of the smartphone and online usage-based evaluation for depression (solvd) application: What do apps really tell us about patients with depression? concordance between app-generated data and standard psychiatric questionnaires for depression and anxiety. Journal of Psychiatric Practice® 25. URL: https://journals.lww.com/practicalpsychiatry/Fulltext/2019/09000/Findings_From_a_Trial_of_the_Smartphone_and_OnLine.5.aspx.

Murnane, E.L., Abdullah, S., Matthews, M., Kay, M., Kientz, J.A., Choudhury, T., Gay, G., Cosley, D., 2016. Mobile manifestations of alertness: Connecting biological rhythms with patterns of smartphone app use, in: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, Association for Computing Machinery, New York, NY, USA. p. 465–477. URL: https://doi.org/10.1145/2935334.2935383, doi:10.1145/2935334.2935383.

Murray, H.A., 1938. Explorations in personality: A clinical and experimental study of fifty men of college age. .

Musto, C., Semeraro, G., Lovascio, C., de Gemmis, M., Lops, P., 2018. A framework for holistic user modeling merging heterogeneous digital footprints, in: Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization, Association for Computing Machinery, New York, NY, USA. p. 97–101. doi:10.1145/3213586.3226218.

Natal, I., de Avellar Campos Cordeiro, R., Garcia, A.C.B., 2017. Activity recognition model based on gps data, points of interest and user profile, in: Kryszkiewicz, M., Appice, A., Ślezak, D., Rybinski, H., Skowron, A., Raś, Z.W. (Eds.), Foundations of Intelligent Systems, Springer International Publishing, Cham. pp. 358–367.

Natal, I., Correia, L., Garcia, A., Fernandes, L., 2019. Efficient out-of-home activity recognition by complementing gps data with semantic information. First Monday 24. URL: http://search.proquest.com/docview/2311913858/.

Neher, A., 1991. Maslow's theory of motivation. Journal of Humanistic Psychology 31, 89–112.

Nicholls, J.G., 1984. Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. Psychological review 91, 328.

Nicholson Jr., N.R., 2009. Social isolation in older adults: an evolutionary concept analysis. Journal of Advanced Nursing 65, 1342–1352. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2648.2008.04959.x, doi:https://doi.org/10.1111/j.1365-2648.2008.04959.x.

Nickels, S., Edwards, M.D., Poole, S.F., Winter, D., Gronsbell, J., Rozenkrants, B., Miller, D.P., Fleck, M., McLean, A., Peterson, B., et al., 2021. Toward a mobile platform for real-world digital measurement of depression: User-centered design, data quality, and behavioral and clinical modeling. JMIR Mental Health 8, e27589. doi:10.2196/27589.

Nogueira, T.P., Braga, R.B., de Oliveira, C.T., Martin, H., 2018. FrameSTEP: A framework for annotating semantic trajectories based on episodes. Expert Systems with Applications 92, 533–545. doi:10.1016/j.eswa.2017.10.004.

Nogueira, T.P., Martin, H., Andrade, R.M., 2017. A statistical method for detecting move, stop, and noise episodes in trajectories., in: GEOINFO, pp. 210–221.

Northrup, D.A., 1997. The problem of the self-report in survey research. Institute for Social Research, York University.

Obuchi, M., Huckins, J.F., Wang, W., daSilva, A., Rogers, C., Murphy, E., Hedlund, E., Holtzheimer, P., Mirjafari, S., Campbell, A., 2020. Predicting brain functional connectivity using mobile sensing. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 4. URL: https://doi.org/10.1145/3381001, doi:10.1145/3381001.

Onnela, J.P., Rauch, S.L., 2016. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 41, 1691–1696.

Oudeyer, P.Y., Kaplan, F., 2009. What is intrinsic motivation? A typology of computational approaches. Frontiers in Neurorobotics 3. doi:10.3389/neuro.12.006.2007, arXiv:arXiv:1410.5401v2.

Ozturk, A.B., Nusair, K., Okumus, F., Hua, N., 2016. The role of utilitarian and hedonic values on users' continued usage intention in a mobile hotel booking environment. International Journal of Hospitality Management 57, 106 – 115. URL: http://www.sciencedirect.com/science/article/pii/S0278431916300809, doi:https://doi.org/10.1016/j.ijhm.2016.06.007.

Paige, D.D., 2011. Engaging struggling adolescent readers through situational interest: A model proposing the relationships among extrinsic motivation, oral reading proficiency, comprehension, and academic achievement. Reading Psychology 32, 395–425. URL: https://doi.org/10.1080/02702711.2010.495633, doi:10.1080/02702711.2010.495633, arXiv:https://doi.org/10.1080/02702711.2010.495633.

Palmer, D., Dixon, J., Archer, J., 2017. Using situational interest to enhance individual interest and science-related behaviours. Research in Science Education 47, 731–753. URL: https://doi.org/10.1007/s11165-016-9526-x, doi:10.1007/s11165-016-9526-x.

Parent, C., Pelekis, N., Theodoridis, Y., Yan, Z., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., Damiani, M.L., Gkoulalas-Divanis, A., Macedo, J., 2013. Semantic trajectories modeling and analysis. ACM Computing Surveys 45, 1–32. doi:10.1145/2501654.2501656.

Patel, V.L., Arocha, J.F., Ancker, J.S., 2017. Cognitive Informatics in Health and Biomedicine. URL: http://link.springer.com/10.1007/978-3-319-51732-2, doi:10.1007/978-3-319-51732-2.

Paulhus, D.L., Vazire, S., 2007. The self-report method. Handbook of research methods in personality psychology 1, 224–239.

Pelletier, L.G., Fortier, M.S., Vallerand, R.J., Briere, N.M., 2001. Associations among perceived autonomy support, forms of self-regulation, and persistence: A prospective study. Motivation and emotion 25, 279–306.

Petri, H.L., Govern, J.M., 2013. Motivation : theory, research and application. 6th ed. ed., Belmont, CA : Wadsworth, Cengage Learning, Belmont, CA.

Pielot, M., Church, K., de Oliveira, R., 2014. An in-situ study of mobile phone notifications, in: Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices and Services, Association for Computing Machinery, New York, NY, USA. p. 233–242. URL: https://doi.org/10.1145/2628363.2628364, doi:10.1145/2628363.2628364.

Poppinga, B., Heuten, W., Boll, S., 2014. Sensor-based identification of opportune moments for triggering notifications. IEEE Pervasive Computing 13, 22–29. doi:10.1109/MPRV.2014.15.

Potti, N., Wendt, J.B., Zhao, Q., Tata, S., Najork, M., 2018. Hidden in plain sight: Classifying emails using embedded image contents, in: Proceedings of the 2018 World Wide Web Conference, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE. p. 1865–1874. URL: https://doi.org/10.1145/3178876.3186167, doi:10.1145/3178876.3186167.

Poushter, J., et al., 2016. Smartphone ownership and internet usage continues to climb in emerging economies. Pew research center 22, 1–44.

Prashar, S., Gupta, P., Parsad, C., Vijay, T.S., 2018. Predicting shoppers' continuous buying intention using mobile apps. International Journal of Strategic Decision Sciences (IJSDS) 9, 69–83.

Prelipcean, A.C., Gidófalvi, G., Susilo, Y.O., 2017. Transportation mode detection–an in-depth review of applicability and reliability. Transport reviews 37, 442–464.

Quadrana, M., Cremonesi, P., Jannach, D., 2018. Sequence-aware recommender systems. ACM Comput. Surv. 51. URL: https://doi.org/10.1145/3190616, doi:10.1145/3190616.

Rabbi, M., Aung, M.H., Zhang, M., Choudhury, T., 2015. MyBehavior. Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15 , 707–718doi:10.1145/2750858.2805840.

Rawassizadeh, R., Momeni, E., Dobbins, C., Gharibshah, J., Pazzani, M., 2016. Scalable daily human behavioral pattern mining from multivariate temporal data. IEEE Transactions on Knowledge and Data Engineering 28, 3098–3112. doi:10.1109/TKDE.2016.2592527.

Raza, S., Ding, C., 2019. Progress in context-aware recommender systems—an overview. Computer Science Review 31, 84–97.

Reber, R., Canning, E.A., Harackiewicz, J.M., 2018. Personalized education to increase interest. Current Directions in Psychological Science 27, 449–454. URL: https://doi.org/10.1177/0963721418793140, doi:10.1177/0963721418793140, arXiv:https://doi.org/10.1177/0963721418793140.

Reeve, J., 1993. The face of interest. Motivation and Emotion 17, 353–375. URL: https://doi.org/10.1007/BF00992325, doi:10.1007/BF00992325.

Reeve, J., Nix, G., 1997. Expressing intrinsic motivation through acts of exploration and facial displays of interest. Motivation and Emotion 21, 237–250. URL: https://doi.org/10.1023/A:1024470213500, doi:10.1023/A:1024470213500.

Renninger, K.A., Hidi, S., 2011. Revisiting the conceptualization, measurement, and generation of interest. Educational Psychologist 46, 168–184. doi:10.1080/00461520.2011.587723.

Renninger, K.A., Hidi, S.E., 2016. The power of interest for motivation and engagement. London Taylor and Francis. doi:10.4324/9781315771045.

Rhim, S., Lee, U., Han, K., 2020. Tracking and modeling subjective well-being using smartphone-based digital phenotype, in: Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, Association for Computing Machinery, New York, NY, USA. p. 211–220. URL: https://doi.org/10.1145/3340631.3394855, doi:10.1145/3340631.3394855.

Ricci, F., Rokach, L., Shapira, B., 2011. Introduction to Recommender Systems Handbook. Springer US, Boston, MA. pp. 1–35. URL: https://doi.org/10.1007/978-0-387-85820-3_1, doi:10.1007/978-0-387-85820-3_1.

Robertson, K., 1989. Strategically desirable brand name characteristics. Journal of Consumer Marketing 6, 61–71. URL: https://doi.org/10.1108/EUM0000000002563, doi:10.1108/EUM0000000002563.

Rooksby, J., Morrison, A., Murray-Rust, D., 2019. Student Perspectives on Digital Phenotyping: The Acceptability of Using Smartphone Data to Assess Mental Health. Association for Computing Machinery, New York, NY, USA. p. 1–14. URL: https://doi.org/10.1145/3290605.3300655.

Rooksby, J., Rost, M., Morrison, A., Chalmers, M., 2015. Pass the ball: Enforced turn-taking in activity tracking, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 2417–2426. doi:10.1145/2702123.2702577.

Rosales, A., Fernandez-Ardevol, M., 2016a. Beyond whatsapp: Older people and smartphones. Romanian Journal of Communication and Public Relations 18, 27–47.

Rosales, A., Fernandez-Ardevol, M., 2016b. Smartphones, apps and older people's interests: From a generational perspective, in: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, Association for Computing Machinery, New York, NY, USA. p. 491–503. URL: https://doi.org/10.1145/2935334.2935363, doi:10.1145/2935334.2935363.

Rossi, P.E., McCulloch, R.E., Allenby, G.M., 1996. The value of purchase history data in target marketing. Marketing Science 15, 321–340. URL: https://doi.org/10.1287/mksc.15.4.321.

Ruan, X.W., Lee, S.C., Peng, W.C., 2014. Exploring location-related data on smart phones for activity inference, in: 2014 IEEE 15th International Conference on Mobile Data Management, IEEE. pp. 73,78.

Ryan, R., 2018. Intrinsic motivation inventory. URL: {http://selfdeterminationtheory.org/intrinsic-motivation-inventory}[Lastaccessed:March2018].

Ryan, R.M., 1982. Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. Journal of Personality and Social Psychology 43, 450–461. doi:10.1037//0022-3514.43.3.450.

Ryan, R.M., Deci, E.L., 2000. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. Contemporary Educational Psychology 25, 54–67. doi:10.1006/ceps.1999.1020, arXiv:arXiv:1011.1669v3.

Ryan, R.M., Deci, E.L., 2013. Toward a social psychology of assimilation: Self-determination theory in cognitive. Self-regulation and autonomy: Social and developmental dimensions of human conduct 40, 191.

Ryan, R.M., Deci, E.L., 2017. Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness. Guilford Publications Inc.

Sahami Shirazi, A., Henze, N., Dingler, T., Pielot, M., Weber, D., Schmidt, A., 2014. Large-scale assessment of mobile notifications, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 3055–3064. doi:10.1145/2556288.2557189.

Saksono, H., Castaneda-Sceppa, C., Hoffman, J., Morris, V., Seif El-Nasr, M., Parker, A.G., 2020. Storywell: Designing for family fitness app motivation by using social rewards and reflection, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–13.

Sanchez, W., Martinez, A., Campos, W., Estrada, H., Pelechano, V., 2015. Inferring loneliness levels in older adults from smartphones. Journal of Ambient Intelligence and Smart Environments 7, 85–98. doi:10.3233/AIS-140297.

Sano, A., Taylor, S., McHill, A.W., Phillips, A.J., Barger, L.K., Klerman, E., Picard, R., et al., 2018. Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: observational study. Journal of medical Internet research 20, e9410. doi:10.2196/jmir.9410.

Sansone, C., Weir, C., Harpster, L., Morgan, C., 1992. Once a boring task always a boring task? interest as a self-regulatory mechanism. Journal of personality and social psychology 63, 379.

Santani, D., Do, T., Labhart, F., Landolt, S., Kuntsche, E., Gatica-Perez, D., 2018. Drinksense: Characterizing youth drinking behavior using smartphones. IEEE Transactions on Mobile Computing 17, 2279–2292. doi:10.1109/TMC.2018.2797901.

Sarker, I.H., Colman, A., Han, J., 2019. Recencyminer: mining recency-based personalized behavior from contextual smartphone data. Journal of Big Data 6, 49.

Sarsenbayeva, Z., Marini, G., van Berkel, N., Luo, C., Jiang, W., Yang, K., Wadley, G., Dingler, T., Kostakos, V., Goncalves, J., 2020. Does Smartphone Use Drive Our Emotions or Vice Versa? A Causal Analysis. Association for Computing Machinery, New York, NY, USA. p. 1–15. URL: https://doi.org/10.1145/3313831.3376163.

Savickas, M.L., 1999. The psychology of interests. .

Schafer, J.B., Konstan, J.A., Riedl, J., 2001. E-commerce recommendation applications. Data mining and knowledge discovery 5, 115–153.

Scherer, K.R., 1986. Vocal affect expression: a review and a model for future research. Psychological bulletin 99, 143.

Schiefele, U., 2009. Situational and individual interest.. Routledge/Taylor & Francis Group, New York, NY, US. Educational psychology handbook series., pp. 197–222.

Schoning, C., Matt, C., Hess, T., 2019. Personalised nudging for more data disclosure? on the adaption of data usage policies format to cognitive styles. Hawaii International Conference on System Sciences 2019 doi:10.24251/HICSS.2019.532.

Schraw, G., Bruning, R., Svoboda, C., 1995. Sources of situational interest. Journal of reading behavior 27, 1–17.

Schraw, G., Flowerday, T., Lehman, S., 2001. Increasing situational interest in the classroom. Educational Psychology Review 13, 211–224. URL: https://doi.org/10.1023/A:1016619705184, doi:10.1023/A:1016619705184.

Schulze, F., Groh, G., 2014. Studying how character of conversation affects personal receptivity to mobile notifications, in: CHI '14 Extended Abstracts on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 1729–1734. URL: https://doi.org/10.1145/2559206.2581320, doi:10.1145/2559206.2581320.

Schumacher, C.F., 1963. Interest and personlity factors as related to choice of medical career. Academic Medicine 38, 932–942.

Schunk, D.H., Pintrich, P.R., Meece, J.L., 2008. Motivation in Education: Theory, Research, and Applications. 3rd ed. ed., Pearson/Merrill Prentice Hall, Upper Saddle River, N.J.

Sefidgar, Y.S., Seo, W., Kuehn, K.S., Althoff, T., Browning, A., Riskin, E., Nurius, P.S., Dey, A.K., Mankoff, J., 2019. Passively-sensed behavioral correlates of discrimination events in college students. Proc. ACM Hum.-Comput. Interact. 3. URL: https://doi.org/10.1145/3359216, doi:10.1145/3359216.

Seguin, C., Pelletier, L.G., Hunsley, J., 1999. Predicting environmental behaviors: The influence of self-determined motivation and information about perceived environmental health risks 1. Journal of Applied Social Psychology 29, 1582–1604.

Seifert, A., Schlomann, A., Rietz, C., Schelling, H.R., 2017. The use of mobile devices for physical activity tracking in older adults' everyday life. DIGITAL HEALTH 3, 2055207617740088. URL: https://doi.org/10.1177/2055207617740088, doi:10.1177/2055207617740088, arXiv:https://doi.org/10.1177/2055207617740088.

Shah, D., Upasini, A., Sasidhar, K., 2020. Findings from an experimental study of student behavioral patterns using smartphone sensors, in: 2020 International Conference on COMmunication Systems NETworkS (COMSNETS), pp. 768–772. doi:10.1109/COMSNETS48256.2020.9027496.

Shareef, M.A., Dwivedi, Y.K., Rana, N.P., 2015. Consumer behaviour in the context of sms-based marketing. The Marketing Review 15, 135–160.

Sheldon, K.M., Kasser, T., 2001. Getting older, getting better? personal strivings and psychological maturity across the life span. Developmental psychology 37, 491–501.

Shi, H., Chen, L., Xu, Z., Lyu, D., 2019. Personalized location recommendation using mobile phone usage information. Applied Intelligence 49, 3694–3707. URL: https://doi.org/10.1007/s10489-019-01477-6, doi:10.1007/s10489-019-01477-6.

Silvia, P.J., 2007. Exploring the Psychology of Interest. Oxford ; New York : Oxford University Press, Oxford ; New York. doi:10.1093/acprof:oso/9780195158557.001.0001, arXiv:arXiv:1011.1669v3.

Skinner, B. F. (Burrhus Frederic), .., 1953. Science and human behavior. Macmillan, New York.

Smith, D.N., Chen, X., 2018. Brand experience, flow and brand app loyalty: Examining consumer decision making within branded mobile apps. Marketing Management Journal 28.

Smith, S.L., Mosier, J.N., 1986. Guidelines for designing user interface software. Citeseer.

Sohn, S., Groß, M., 2020. Understanding the inhibitors to consumer mobile purchasing intentions. Journal of Retailing and Consumer Services 55, 102129. URL: http://www.sciencedirect.com/science/article/pii/S0969698920300825, doi:https://doi.org/10.1016/j.jretconser.2020.102129.

Solomon, A., Bar, A., Yanai, C., Shapira, B., Rokach, L., 2018. Predict demographic information using word2vec on spatial trajectories, in: Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization, ACM, New York, NY, USA. pp. 331–339. doi:10.1145/3209219.3209224.

Sondaal, S.F.V., Browne, J.L., Amoakoh-Coleman, M., Borgstein, A., Miltenburg, A.S., Verwijs, M., Klipstein-Grobusch, K., 2016. Assessing the effect of mhealth interventions in improving maternal and neonatal care in low-and middle-income countries: a systematic review. PloS one 11.

Sonego, M., Echeveste, M.E.S., Galvan Debarba, H., 2018. The role of modularity in sustainable design: A systematic review. Journal of Cleaner Production 176, 196–209. doi:https://doi.org/10.1016/j.jclepro.2017.12.106.

Song, C., Koren, T., Wang, P., Barabási, A.L., 2010. Modelling the scaling properties of human mobility. Nature Physics 6, 818–823.

Spitzer, R.L., Kroenke, K., Williams, J.B.W., Löwe, B., 2006. A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. Archives of Internal Medicine 166, 1092–1097. URL: https://doi.org/10.1001/archinte.166.10.1092, doi:10.1001/archinte.166.10.1092.

Spruijt-Metz, D., Hekler, E., Saranummi, N., Intille, S., Korhonen, I., Nilsen, W., Rivera, D.E., Spring, B., Michie, S., Asch, D.A., Sanna, A., Salcedo, V.T., Kukakfa, R., Pavel, M., 2015. Building new computational models to support health behavior change and maintenance: new opportunities in behavioral research. Translational Behavioral Medicine 5, 335–346. URL: http://dx.doi.org/10.1007/s13142-015-0324-1, doi:10.1007/s13142-015-0324-1.

Srikant, R., Vu, Q., Agrawal, R., 1997. Mining association rules with item constraints., in: Kdd, pp. 67–73.

Srinivasan, V., Moghaddam, S., Mukherji, A., Rachuri, K.K., Xu, C., Tapia, E.M., 2014. Mobileminer: Mining your frequent patterns on your phone, in: Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing, pp. 389–400.

Steeds, M., Clinch, S., Jay, C., 2021. Device uses and device stereotypes. Computers in Human Behavior Reports 4, 100100. URL: https://www.sciencedirect.com/science/article/pii/S2451958821000488, doi:https://doi.org/10.1016/j.chbr.2021.100100.

Strombach, T., Strang, S., Park, S.Q., Kenning, P., 2016. Common and distinctive approaches to motivation in different disciplines, in: Progress in Brain Research. volume 229, pp. 3–23. doi:10.1016/bs.pbr.2016.06.007.

Sükei, E., Norbury, A., Perez-Rodriguez, M.M., Olmos, P.M., Artés, A., 2021. Predicting emotional states using behavioral markers derived from passively sensed data: Data-driven machine learning approach. JMIR mHealth and uHealth 9, e24465. doi:10.2196/24465.

Sultana, M., Al-Jefri, M., Lee, J., 2020. Using machine learning and smartphone and smartwatch data to detect emotional states and transitions: exploratory study. JMIR mHealth and uHealth 8, e17818. doi:10.2196/17818.

Sun, Y., Garcia-Pueyo, L., Wendt, J.B., Najork, M., Broder, A., 2018. Learning effective embeddings for machine generated emails with applications to email category prediction, in: 2018 IEEE International Conference on Big Data (Big Data), pp. 1846–1855. doi:10.1109/BigData.2018.8622048.

Talevich, J.R., Read, S.J., Walsh, D.A., Iyer, R., Chopra, G., 2017. Toward a comprehensive taxonomy of human motives. PLOS ONE 12, 1–32. doi:10.1371/journal.pone.0172279.

Tang, A.K., 2019. A systematic literature review and analysis on mobile apps in m-commerce: Implications for future research. Electronic Commerce Research and Applications 37, 100885. URL: http://www.sciencedirect.com/science/article/pii/S1567422319300626, doi:https://doi.org/10.1016/j.elerap.2019.100885.

Tang, L.M., Kay, J., 2017. Harnessing long term physical activity data—how long-term trackers use data and how an adherence-based interface supports new insights. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 1. URL: https://doi.org/10.1145/3090091, doi:10.1145/3090091.

Thaler, R.H., Sunstein, C.R., 2008. Nudge : improving decisions about health, wealth, and happiness. Yale University Press, New Haven.

Thomas, S., Chie, Q.T., Abraham, M., Jalarajan Raj, S., Beh, L.S., 2014. A qualitative review of literature on peer review of teaching in higher education: An application of the swot framework. Review of educational Research 84, 112–159.

Tinsley, H.E., Eldredge, B.D., 1995. Psychological benefits of leisure participation: A taxonomy of leisure activities based on their need-gratifying properties. Journal of Counseling Psychology 42, 123.

Tomkins, S., 1962. Affect imagery consciousness: Volume I: The positive affects. Springer publishing company.

Torous, J., Friedman, R., Keshavan, M., 2014. Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions. JMIR mHealth uHealth 2, e2. doi:10.2196/mhealth.2994.

Torous, J., Kiang, M.V., Lorme, J., Onnela, J.P., 2016. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR mental health 3, e16.

Tseng, V.W.S., Merrill, M., Wittleder, F., Abdullah, S., Aung, M.H., Choudhury, T., 2016. Assessing mental health issues on college campuses: Preliminary findings from a pilot study, in: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, Association for Computing Machinery, New York, NY, USA. p. 1200–1208. URL: https://doi.org/10.1145/2968219.2968308, doi:10.1145/2968219.2968308.

Tu, Z., Duan, B., Wang, Z., Xu, X., 2021. Bidirectional sensing of user preferences and application changes for dynamic mobile app recommendations. Neural Computing and Applications 33, 9791–9803. URL: https://doi.org/10.1007/s00521-021-05744-x, doi:10.1007/s00521-021-05744-x.

Tu, Z., Li, Y., Hui, P., Su, L., Jin, D., 2020. Personalized mobile app recommendation by learning user's interest from social media. IEEE Transactions on Mobile Computing 19, 2670–2683. doi:10.1109/TMC.2019.2929388.

Turner, L.D., Allen, S.M., Whitaker, R.M., 2017. Reachable but not receptive: Enhancing smartphone interruptibility prediction by modelling the extent of user engagement with notifications. Pervasive and Mobile Computing 40, 480 – 494. URL: http://www.sciencedirect.com/science/article/pii/S1574119217300640, doi:https://doi.org/10.1016/j.pmcj.2017.01.011.

Tyler-Wood, T., Knezek, G., Christensen, R., 2010. Instruments for assessing interest in stem content and careers. Journal of Technology and Teacher Education 18, 345–368.

Tyrvainen, O., Karjaluoto, H., 2019. A systematic literature review and analysis of mobile retailing adoption. Journal of Internet Commerce 18, 221–247. URL: https://doi.org/10.1080/15332861.2019.1595364, doi:10.1080/15332861.2019.1595364, arXiv:https://doi.org/10.1080/15332861.2019.1595364.

Usher, K., Bhullar, N., Jackson, D., 2020. Life in the pandemic: Social isolation and mental health. Journal of Clinical Nursing .

Van Bavel, J.J., Baicker, K., Boggio, P.S., Capraro, V., Cichocka, A., Cikara, M., Crockett, M.J., Crum, A.J., Douglas, K.M., Druckman, J.N., et al., 2020. Using social and behavioural science to support covid-19 pandemic response. Nature Human Behaviour , 1–12.

Vansteenkiste, M., Niemiec, C.P., Soenens, B., 2010. The development of the five mini-theories of self-determination theory: An historical overview, emerging trends, and future directions, in: The decade ahead: Theoretical perspectives on motivation and achievement. Emerald Group Publishing Limited, pp. 105–165.

Vega-Hernandez, J., 2019. Unobtrusive and Personalised Monitoring of Parkinson's Disease Using Smartphones. Ph.D. thesis. University of Manchester.

Vega-Hernandez, J., Jay, C., Vigo, M., Harper, S., 2017. Unobtrusive monitoring of parkinson's disease based on digital biomarkers of human behaviour, in: Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 351–352.

Villegas, N.M., Müller, H.A., 2010. Managing dynamic context to optimize smart interactions and services, in: Chignell, M., Cordy, J., Ng, J., Yesha, Y. (Eds.), The Smart Internet: Current Research and Future Applications. Springer Berlin Heidelberg, Berlin, Heidelberg. chapter 18, pp. 289–318. doi:10.1007/978-3-642-16599-3_18.

Villegas, N.M., Sánchez, C., Díaz-Cely, J., Tamura, G., 2018. Characterizing context-aware recommender systems: A systematic literature review. Knowledge-Based Systems 140, 173–200. doi:10.1016/j.knosys.2017.11.003.

Visuri, A., van Berkel, N., Okoshi, T., Goncalves, J., Kostakos, V., 2019. Understanding smartphone notifications' user interactions and content importance. International Journal of Human-Computer Studies 128, 72 – 85. URL: http://www.sciencedirect.com/science/article/pii/S1071581919300205, doi:https://doi.org/10.1016/j.ijhcs.2019.03.001.

Wahle, F., Kowatsch, T., Fleisch, E., Rufer, M., Weidt, S., 2016. Mobile sensing and support for people with depression: A pilot trial in the wild, in: JMIR mHealth and uHealth.

Walkington, C., Bernacki, M.L., 2014. Motivating Students by "Personalizing" Learning around Individual Interests: A Consideration of Theory, Design, and Implementation Issues. Emerald Group Publishing Limited. volume 18 of *Advances in Motivation and Achievement*. pp. 139–176. URL: https://doi.org/10.1108/S0749-742320140000018004, doi:10.1108/S0749-742320140000018004.

Wan, N., Lin, G., 2013. Life-space characterization from cellular telephone collected gps data. Computers, Environment and Urban Systems 39, 63 – 70. URL: http://www.sciencedirect.com/science/article/pii/S0198971513000045, doi:https://doi.org/10.1016/j.compenvurbsys.2013.01.003.

Wang, R., Aung, M.S.H., Abdullah, S., Brian, R., Campbell, A.T., Choudhury, T., Hauser, M., Kane, J., Merrill, M., Scherer, E.A., Tseng, V.W.S., Ben-Zeev, D., 2016. Crosscheck: Toward passive sensing and detection of mental health changes in people with schizophrenia, in: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Association for Computing Machinery, New York, NY, USA. p. 886–897. URL: https://doi.org/10.1145/2971648.2971740, doi:10.1145/2971648.2971740.

Wang, R., Wang, W., Aung, M.S.H., Ben-Zeev, D., Brian, R., Campbell, A.T., Choudhury, T., Hauser, M., Kane, J., Scherer, E.A., Walsh, M., 2017a. Predicting symptom trajectories of schizophrenia using mobile sensing. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 1. URL: https://doi.org/10.1145/3130976, doi:10.1145/3130976.

Wang, R., Wang, W., daSilva, A., Huckins, J.F., Kelley, W.M., Heatherton, T.F., Campbell, A.T., 2018a. Tracking depression dynamics in college students using mobile phone and wearable sensing. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2. URL: https://doi.org/10.1145/3191775, doi:10.1145/3191775.

Wang, W., Harari, G.M., Wang, R., Müller, S.R., Mirjafari, S., Masaba, K., Campbell, A.T., 2018b. Sensing behavioral change over time: Using within-person variability features from mobile sensing to predict personality traits. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2. URL: https://doi.org/10.1145/3264951, doi:10.1145/3264951.

Wang, W., Mirjafari, S., Harari, G., Ben-Zeev, D., Brian, R., Choudhury, T., Hauser, M., Kane, J., Masaba, K., Nepal, S., Sano, A., Scherer, E., Tseng, V., Wang, R., Wen, H., Wu, J., Campbell, A., 2020. Social Sensing: Assessing Social Functioning of Patients Living with Schizophrenia Using Mobile Phone Sensing. Association for Computing Machinery, New York, NY, USA. p. 1–15. URL: https://doi.org/10.1145/3313831.3376855.

Wang, X., Wong, A.K.s., Kong, Y., 2012. Mobility tracking using gps, wi-fi and cell id, in: The International Conference on Information Network 2012, pp. 171–176. doi:10.1109/ICOIN.2012.6164371.

Wang, X., Yu, T., Zeng, M., Tague, P., 2017b. Xrec: Behavior-based user recognition across mobile devices. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 1. URL: https://doi.org/10.1145/3130975, doi:10.1145/3130975.

Wang, Y., McArthur, D., 2018. Enhancing data privacy with semantic trajectories: A raster-based framework for gps stop/move management. Transactions in GIS 22, 975–990. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/tgis.12334, doi:10.1111/tgis.12334, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/tgis.12334.

Wang, Y., Ren, X., Liu, X., Zhu, T., 2021. Examining the correlation between depression and social behavior on smartphones through usage metadata: Empirical study. JMIR mHealth and uHealth 9, e19046. doi:10.2196/19046.

Wannenburg, J., Malekian, R., 2017. Physical activity recognition from smartphone accelerometer data for user context awareness sensing. IEEE Transactions on Systems, Man, and Cybernetics: Systems 47, 3142–3149. doi:10.1109/TSMC.2016.2562509.

Webb, E., Campbell, D., Schwartz, R., Sechrest, L., 2000. Unobtrusive measures, revised edition au. doi:10.4135/9781452243443.

Weber, D., Voit, A., Kollotzek, G., Henze, N., 2019. Annotif: A system for annotating mobile notifcations in user studies, in: Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/3365610.3365611, doi:10.1145/3365610.3365611.

Weihs, C., Buschfeld, S., 2021. Repeated undersampling in prindt (reprindt): Variation in undersampling and prediction, and ranking of predictors in ensembles. arXiv:2108.05129.

Weiner, B., 1992. Human motivation : metaphors, theories, and research. London : Sage, London.

Weiner, B., 2004. Intrapersonal and interpersonal theories of motivation from an attributional perspective. Educational Psychology Review 12, 1–14. URL: http://www.springerlink.com/index/M33X5M787715W503.pdf, doi:10.1023/A:1009017532121, arXiv:0005074v1.

Weinmann, M., Schneider, C., vom Brocke, J., 2016. Digital nudging. Business & Information Systems Engineering 58, 433–436.

Wen, H., Sobolev, M., Vitale, R., Kizer, J., Pollak, J., Muench, F., Estrin, D., 2021. mpulse mobile sensing model for passive detection of impulsive behavior: Exploratory prediction study. JMIR Mental Health 8, e25019. doi:10.2196/25019.

Westermann, T., Wechsung, I., Möller, S., 2016. Smartphone notifications in context: A case study on receptivity by the example of an advertising service, in: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 2355–2361. URL: https://doi.org/10.1145/2851581.2892383, doi:10.1145/2851581.2892383.

White, R.W., 1959. Motivation reconsidered: The concept of competence. Psychological Review 66, 297–333. doi:10.1037/h0040934.

White, R.W., 1963. Ego and reality in psychoanalytic theory. Psychological issues .

Whittaker, M., Edmonds, N., Tata, S., Wendt, J.B., Najork, M., 2019. Online template induction for machine-generated emails, in: PVLDB, pp. 1235–1248.

Wikipedia, 2020. Coronavirus disease 2019 — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=Coronavirus%20disease%202019&oldid=956033455. [Online; accessed 11-May-2020].

Wilmer, H.H., Sherman, L.E., Chein, J.M., 2017. Smartphones and cognition: A review of research exploring the links between mobile technology habits and cognitive functioning. Frontiers in Psychology 8, 605. URL: https://www.frontiersin.org/article/10.3389/fpsyg.2017.00605, doi:10.3389/fpsyg.2017.00605.

Wolf, N., Hopko, D., 2008. Predicting the frequency, duration, and intensity of running behavior: The inhibitory effects of cardiac anxiety. Journal of Behavior Analysis in Health, Sports, Fitness and Medicine 1. doi:10.1037/h0100367.

Xia, N., Miskovic, S., Baldi, M., Kuzmanovic, A., Nucci, A., 2014. Geoecho: Inferring user interests from geotag reports in network traffic, in: 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), pp. 1–8. doi:10.1109/WI-IAT.2014.73.

Xing, X., Li, M., Hu, W., Huang, W., Song, G., Xie, K., 2014. A spatial-temporal topic segmentation model for human mobile behavior, in: Li, F., Li, G., Hwang, S.w., Yao, B., Zhang, Z. (Eds.), Web-Age Information Management, Springer International Publishing, Cham. pp. 255–267.

Xu, P., Dredze, M., Broniatowski, D.A., 2020. The twitter social mobility index: Measuring social distancing practices from geolocated tweets. arXiv preprint arXiv:2004.02397 .

Xu, X., Chikersal, P., Doryab, A., Villalba, D.K., Dutcher, J.M., Tumminia, M.J., Althoff, T., Cohen, S., Creswell, K.G., Creswell, J.D., Mankoff, J., Dey, A.K., 2019. Leveraging routine behavior and contextually-filtered features for depression detection among college students. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 3. URL: https://doi.org/10.1145/3351274, doi:10.1145/3351274.

Xu, X., Chikersal, P., Dutcher, J.M., Sefidgar, Y.S., Seo, W., Tumminia, M.J., Villalba, D.K., Cohen, S., Creswell, K.G., Creswell, J.D., Doryab, A., Nurius, P.S., Riskin, E., Dey, A.K., Mankoff, J., 2021. Leveraging collaborative-filtering for personalized behavior modeling: A case study of depression detection among college students. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 5. URL: https://doi.org/10.1145/3448107, doi:10.1145/3448107.

Yan, Z., Chakraborty, D., Parent, C., Spaccapietra, S., Aberer, K., 2013. Semantic trajectories: Mobility data computation and annotation. ACM Trans. Intell. Syst. Technol. 4. URL: https://doi.org/10.1145/2483669.2483682, doi:10.1145/2483669.2483682.

Yan, Z., Giatrakos, N., Katsikaros, V., Pelekis, N., Theodoridis, Y., 2011. Setrastream: Semantic-aware trajectory construction over streaming movement data, in: Pfoser, D., Tao, Y., Mouratidis, K., Nascimento, M.A., Mokbel, M., Shekhar, S., Huang, Y.

(Eds.), Advances in Spatial and Temporal Databases, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 367–385.

Yang, H.C., 2013. Bon appétit for apps: Young american consumers' acceptance of mobile applications. Journal of Computer Information Systems 53, 85–96. URL: https://doi.org/10.1080/08874417.2013.11645635, doi:10.1080/08874417.2013.11645635, arXiv:https://doi.org/10.1080/08874417.2013.11645635.

Yerkes, R.M., Dodson, J.D., 1908. The relation of strength of stimulus to rapidity of habit-formation. Journal of Comparative Neurology and Psychology 18, 459–482. doi:10.1002/cne.920180503.

Yu, Y., Chen, X., 2015. A survey of point-of-interest recommendation in location-based social networks, in: Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence.

Yuan, Q., Cong, G., Ma, Z., Sun, A., Thalmann, N.M., 2013. Time-aware point-of-interest recommendation, in: Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval, pp. 363–372.

Zandbergen, P.A., 2009. Accuracy of iphone locations: A comparison of assisted gps, wifi and cellular positioning. Transactions in GIS 13, 5–25. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9671.2009.01152.x, doi:https://doi.org/10.1111/j.1467-9671.2009.01152.x.

Zhang, J., Wedel, M., 2009. The effectiveness of customized promotions in online and offline stores. Journal of Marketing Research 46, 190–206. URL: https://doi.org/10.1509/jmkr.46.2.190, doi:10.1509/jmkr.46.2.190, arXiv:https://doi.org/10.1509/jmkr.46.2.190.

Zhang, S., Yao, L., Sun, A., Tay, Y., 2019. Deep learning based recommender system: A survey and new perspectives. ACM Comput. Surv. 52. doi:10.1145/3285029.

Zhang, Y., Jin, R., Zhou, Z.H., 2010. Understanding bag-of-words model: a statistical framework. International Journal of Machine Learning and Cybernetics 1, 43–52.

Zhao, Y., Song, P., Feng, F., 2019. What are the revenue implications of mobile channel visits? evidence from the online travel agency industry. Electronic Commerce Research and Applications 36, 100865. URL: http://www.scienced irect.com/science/article/pii/S1567422319300420, doi:https://doi.org/10.1016/j.elerap.2019.100865.

Zhao, Z.D., Yang, Z., Zhang, Z., Zhou, T., Huang, Z.G., Lai, Y.C., 2013. Emergence of scaling in human-interest dynamics. Scientific reports 3, 3472.

Zhou, M., Ding, Z., Tang, J., Yin, D., 2018. Micro behaviors: A new perspective in e-commerce recommender systems, in: Proceedings of the eleventh ACM international conference on web search and data mining, pp. 727–735.

Zuckerman, O., Gal-Oz, A., 2014. Deconstructing gamification: Evaluating the effectiveness of continuous measurement, virtual rewards, and social comparison for promoting physical activity. Personal Ubiquitous Comput. 18, 1705–1719. doi:10.1007/s00779-014-0783-2.

Appendix A

Supplement for Chapter 5

A.1 Participant information sheet

This page is left blank.

MANCHESTER 1824
The University of Manchester

Participant Information Sheet (PIS)

This PIS should be read in conjunction with The University privacy notice:
http://documents.manchester.ac.uk/display.aspx?DocID=37095

You are being invited to take part in a research study as part of a student project that aims to recognise human motivation from smartphone's data. Before you decide whether to take part, it is important for you to understand why the research is being conducted and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please ask if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for taking the time to read this.

About the research

- **Who will conduct the research?**

This research is conducted by Ahmed Ibrahim, a PhD student at the University of Manchester – department of computer science. The research is supervised by professor Simon Harper and Dr Sarah Clinch - University of Manchester. Address details are below.

Name	Email	Address
Ahmed Ibrahim	**Ahmed.ibrahim-2@manchester.ac.uk**	IAM lab, Kilburn Building. M13 9PL. Manchester, UK
Simon Harper	**Simon.harper@manchester.ac.uk**	2.60 Kilburn building, M13 9PL Manchester, UK
Sarah Clinch	**Sarah.clinch@manchester.ac.uk**	Kilburn building, M13 9PL Manchester, UK

- **What is the purpose of the research?**

The purpose of this research is to recognise human motivation by monitoring the behaviour of the participants through their own smartphones

- **Will the outcomes of the research be published?**

Results will be published in conferences, journals and student thesis.

- **Who has reviewed the research project?**

This research project has been reviewed by the Ethics Committee of the Department of Computer Science.

- **What happens after the research project is finished?**

Once the project is finished, the app will still be working on your phone. However, we will stop synchronising the data to our server and it is up to you to keep the app or delete it. If you decide to delete it, we will be happy to help you with that.

What would my involvement be?

➢ What would I be asked to do if I took part?

You will attend the University of Manchester for an introductory session of 30 minutes. During this session we will ask you questions about your interests, help you install the app into your phone and explain how it works.

You will be asked to enable the collection of the data specified in the below table. The collected data will be stored locally on the phone and synched with a secure server hosted internally by the University of Manchester.

Source	Data
GPS	Location coordinates (longitude and latitude)
Weather	Temperature, humidity, pressure, wind speed, cloudiness, amount of rain and snow, times of sunrise and sunset.
Applications	App usage data
Notifications	All notifications generated by any app installed on the phone such as notifications from news or emails apps. All numbers and emails contained in the notifications are de-identified by replacing them with asterisks. For example, if you received "from 555-555-5555, your package has been delivered"; it will be stored as "from *, your package has been delivered".
Screen	Screen interactions, visited websites.
Wifi	Access points.
Bluetooth	Nearby devices.
Battery	Charging status, charging start time, charging end time, discharging start time, discharging end time.
Calls	Calls types (outgoing, incoming and missed calls) and times. Numbers are stored in an encrypted format.
Keyboard	Time and the typed letters. (numbers and emails are replaced with asterisks).
Consent form	Name and signature
Questionnaire	We ask about interests that we infer from the data
Activity recognition	Tilting, running, on vehicle, walking, on bicycle, on foot

You will be asked to keep the installed app running and carry your phone as you would normally do. We will send you weekly messages about your motivations and you will be asked to respond by replying to those messages which is expected to take between 3 and 5 minutes.

At the end of the experiment, to evaluate our work, you will be asked to join an interview to ask you about the interests that we recognised from your data and understand what interests did we miss and why did we miss them.

➢ Will I be compensated for taking part?

As a token of appreciation, participants will be provided with a £5 Amazon gift card at the beginning of each month for three months

As a token of appreciation, participants will be provided with a £5 Amazon gift card at the beginning of each month for the first three months. A £15 Netflix card will be given if you completed three months and joined the final interview at the end of the third month.

➢ What happens if I do not want to take part or if I change my mind?

It is up to you to decide whether or not to take part. If you do decide to take part, you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time without giving a reason and without detriment to yourself. You will need to let us know whether or not to keep your data. Your data will be removed from the dataset and will not be used in any future publications. This does not affect your data protection rights.

Data Protection and Confidentiality

- **What information will you collect about me?**

In order to participate in this research project we will need to collect your contact details and information that could indirectly identify you, such as location data. Please see the table above for more details.

- **Under what legal basis are you collecting this information?**

We are collecting and storing your information in accordance with data protection law which protects your rights. These state that we must have a legal basis (specific reason) for collecting your data. For this study, the specific reason is that it is a process necessary for research purposes only.

- **What are my rights in relation to the information you will collect about me?**

You have a number of rights under data protection law regarding your personal information. For example, you can request a copy of the information that we hold about you.

If you would like to know more about your different rights or the way we use your personal information to ensure we follow the law, please consult our Privacy Notice for Research.

- **Will my participation in the study be confidential and my personal identifiable information be protected?**

In accordance with data protection law, The University of Manchester is the Data Controller for this project. This means that we are responsible for making sure your personal information is kept secure, confidential and used only in the way you have been told it will be used. All researchers are trained with this in mind, and your data will be looked after in the following way:

- To ensure confidentiality of data in digital format:
    - De-identify the data via an assigned participant ID only known to the research team (also referred to as pseudonymised or coded data).
    - Encrypt the submission of data between the participant's phone and the university's server to ensure individuals cannot be readily identified.
    - Only the study team at The University of Manchester will have access to your data
    - De-identify numbers and emails contained in the keyboard and notification texts by replacing them with asterisks.
- To ensure confidentiality of data in non-digital format:
    - We will use locked cabinets inside Kilburn building to store your consent form and personal information.

    - Data in non-digital formats (e.g. your consent form) will be digitized within one month of joining the experiment and securely destroyed. The digitised data will be stored with the collected data in secure servers hosted by the University of Manchester.
- Data could be shared for academic research purposes only and will be anonymised before sharing with other institutions. Any identifiable raw data will not be shared and will be either encrypted or replaced by other unidentifiable data. For example, GPS coordinates will be replaced by the general category of the place which they represent (e.g. restaurant). The below table indicates how each collected data will be prepared for sharing.
- If during the study, we become aware of evidence about any current or future illegal activities, we have a legal obligation to report this and will, therefore, need to inform the relevant authorities.

Please also note that individuals from The University of Manchester or regulatory authorities may need to look at the data collected for this study to make sure the project is being carried out as

Sensor	Shared data
GPS	Sequences of places categories: e.g. home → restaurant → home
Weather	Status: cold, hot, nice
Applications	App usage data
Notifications	Will not be shared.
Screen	Screen status: On/Off
Wifi	Access points labels and ids will be shared in an encrypted form.
Bluetooth	Devices information will be encrypted before sharing.
Battery	Charging and discharging times will be shared
Calls	Calls duration, time, type (received/incoming/missed). Numbers will be shared in an encrypted format.
Keyboard	Only keystrokes times will be shared.
Consent form	Will not be shared.
Questionnaire	Responses associated with anonymised IDs will be shared
Activity Recognition	Activity types and times will be shared

planned. This may involve looking at identifiable data. All individuals involved in auditing and monitoring the study will have a strict duty of confidentiality to you as a research participant.

What if I have a complaint?

- **Contact details for complaints**

If you have a complaint that you wish to direct to members of the research team, please contact:

Name	Email	Address	Telephone
Simon Harper	Simon.harper@manchester.ac.uk	2.60 Kilburn building, M13 9PL Manchester, UK	0161 275 0599

Sarah Clinch	**Sarah.clinch@manchester.ac.uk**	2.24 Kilburn building, M13 9PL Manchester, UK	0161 275 7190

If you wish to make a formal complaint to someone independent of the research team or if you are not satisfied with the response you have gained from the researchers in the first instance then please contact

The Research Governance and Integrity Officer, Research Office, Christie Building, The University of Manchester, Oxford Road, Manchester, M13 9PL, by emailing: research.complaints@manchester.ac.uk or by telephoning 0161 275 2674.

If you wish to contact us about your data protection rights, please email dataprotection@manchester.ac.uk or write to The Information Governance Office, Christie Building, The University of Manchester, Oxford Road, M13 9PL at the University and we will guide you through the process of exercising your rights.

You also have a right to complain to the Information Commissioner's Office about complaints relating to your personal identifiable information Tel 0303 123 1113

Link for Information Commissioner's Office: https://ico.org.uk/concerns

Contact Details

If you have comments or questions, please contact:

Name	Email	Address
Ahmed Ibrahim	**ahmed.ibrahim-2@manchester.ac.uk**	IAM lab, Kilburn Building. M13 9PL. Manchester, UK

A.2 Consent form

This page is left blank.

MANCHESTER 1824
The University of Manchester

Consent Form

If you are happy to participate please complete and sign the consent form below

	Activities	Initials
1	I confirm that I have read the attached information sheet (Version 1.6, 16/12/2019) for the above study and have had the opportunity to consider the information and ask questions and had these answered satisfactorily.	
2	I understand that my participation in the study is voluntary and that I am free to withdraw at any time – up until two months after the end date of the experiment - without giving a reason and without detriment to myself. I agree to take part on this basis.	
3	I understand that when I withdraw from the experiment, my data will be removed from the dataset and will not be used in any future publications.	
4	I agree to conduct an interview about my personal interests at the beginning of the experiment to help researchers validate their results.	
5	I agree to join an interview about my interests at the end of the experiment to help researchers understand what interests did they miss and why did they miss them.	
6	I agree to receive notifications on my phone that ask questions about my interests.	
7	I understand that the installed app will collect: locations, apps usage, notifications, calls duration and times, screen interactions, keyboard interactions, performed activities, wifi and Bluetooth connections, and weather data, in order to understand my daily behaviour.	
8	I agree that any data collected may be published in anonymous form in academic books, conference, reports, journals or thesis.	
9	I understand that data collected during the study may be looked at by individuals from The University of Manchester or regulatory authorities. I give permission for these individuals to have access to my data.	
10	I agree that the researchers may retain my contact details in order to provide me with a summary of the findings for this study.	
11	I understand that there may be instances where during the course of the study information is revealed which means that the researchers will be obliged to break confidentiality and this has been explained in more detail in the information sheet.	
12	I agree to take part in this study.	

Data Protection

The personal information we collect and use to conduct this research will be processed in accordance with data protection law as explained in the Participant Information Sheet and the Privacy Notice for Research Participants.

___________________________	___________________________	_______________
Name of Participant	Signature	Date
___________________________	___________________________	_______________
Name of the person taking consent	Signature	Date

1 copy for the participant, 1 copy for the research team (original)

A.3 Interview guide

This page is left blank.

Interview Type: Semi Structure

Location: IAM lab

Interviewer: Ahmed Ibrahim

Expected duration: 10 – 15 minutes

General information:

- These interviews will be recorded using a laptop with an encrypted hard drive and will be deleted at the end date of the experiment.
- Each interview will be identified by a unique ID that represents the participant, location, start and end time of the interview.
- Transcription will be conducted by the Ahmed Ibrahim and will be stored on an encrypted hard drive.
- The interview will be used as a supplementary source of ground truth.

Questions of the first interview (when joining the study):

- Tell us please about your daily routine.?
- What are the places that you prefer to go to?
- How often do you go to those places?
- What motivates you to go to those places?
- In general, what are your interests and how often do you practice them?

Questions of the second interview (at the end of the experiment):

- How close are we in understanding your interests and what we missed?
- We will provide participants with the common places they visit which we discovered from their data and ask them to rank those places based on their interests.

Appendix B

Supplement for Chapter 6

B.1 Participant information sheet

This page is left blank.

Participant Information Sheet (PIS)

This PIS should be read in conjunction with The University privacy notice:
http://documents.manchester.ac.uk/display.aspx?DocID=37095

You are being invited to take part in a research study as part of a student project that aims to recognise human motivation and social engagement from smartphone data. Before you decide whether to take part, it is important for you to understand why the research is being conducted and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please ask if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for taking the time to read this.

About the research

- **Who will conduct the research?**

This research is conducted by Ahmed Ibrahim, a PhD student at the University of Manchester – department of computer science. The research is supervised by Professor Simon Harper and Dr. Sarah Clinch - University of Manchester. Address details are below.

Name	Email	Address
Ahmed Ibrahim	**Ahmed.ibrahim-2@manchester.ac.uk**	IAM lab, Kilburn Building. M13 9PL. Manchester, UK
Simon Harper	**Simon.harper@manchester.ac.uk**	2.60 Kilburn building, M13 9PL Manchester, UK
Sarah Clinch	**Sarah.clinch@manchester.ac.uk**	Kilburn building, M13 9PL Manchester, UK

- **What is the purpose of the research?**

The purpose of this research is to recognise human motivation and social engagement by monitoring the behaviour of the participants through their own passively sensed personal mobile devices (i.e. smartphones, smartwatches).

- **Will the outcomes of the research be published?**

Results will be published in conferences, journals and student thesis.

- **Who has reviewed the research project?**

This research project has been reviewed by the Ethics Committee of the Department of Computer Science.

- **Who is funding the research project?**

This research is funded by the Ministry of Education in Saudi Arabia, and by The CHERISH-DE Centre at Swansea University (UK EPSRC grant number EP/M022722/1).

- **What happens after the research project is finished?**

Once the project is finished, the app will still be working on your phone. However, we will stop synchronising the data to our server and it is up to you to keep the app or delete it. If you decide to delete it, we will be happy to help you with that. If you do not delete the app we will still not receive any further data from you.

Data that is synced to our server during your participation will be kept in anonymised form for a minimum of five years after study completion. This data may be used in future research or shared with researchers from other institutions upon request.

What would my involvement be?

- **What would I be asked to do if I took part?**

You will attend an introductory session of 30 minutes. The introductory session can be online due to the COVID-19 precautions. During this session we will ask you questions about your interests, help you install the app onto your phone and explain how the app works. We'll also ask you some short questions about your normal socialising and drinking behaviour. We'll be happy to answer any of your questions during this session, but if you have a subsequent query then feel free to contact the research team using the details on page 1 of this document.

You will be asked to enable the collection of the data specified in the below table. The collected data will be stored locally on the phone and synced with a secure server hosted internally at the University of Manchester.

Source	Data
GPS	Location coordinates (longitude and latitude).
Weather	Temperature, humidity, pressure, wind speed, cloudiness, amount of rain and snow, times of sunrise and sunset.
Applications	App usage data.
Notifications	All notifications generated by any app installed on the phone such as notifications from news or emails apps. All numbers and emails contained in the notifications are de-identified by replacing them with asterisks. For example, if you received "from 555-555-5555, your package has been delivered"; it will be stored as "from *, your package has been delivered".
Screen	Screen interactions, visited websites.
Wifi	Access points.
Bluetooth	Nearby devices.
Battery	Charging status, charging start time, charging end time, discharging start time, discharging end time.
Calls	Calls types (outgoing, incoming and missed calls) and times. Numbers are stored in an encrypted format.
Messages	Message types (outgoing, incoming) and times. Numbers are stored in an encrypted format. Message content is not recorded.
Keyboard	Time and the typed letters. (numbers and emails are replaced with asterisks).
Consent form	Name and signature.
Questionnaire	We ask about interests that we infer from the data.
Activity recognition	Tilting, running, on vehicle, walking, on bicycle, on foot.

Opening and closing interviews	Audio recordings (deleted following transcription). Your gender and responses to questions about your interests, how you socialise and how often you drink alcohol.

You will be asked to keep the installed app running for six months and carry your phone as you would normally do. We will send you monthly questionnaire (via email) about your motivations and you will be asked to respond by filling the questionnaire and send it back. Each questionnaire is expected to take between 3 and 5 minutes.

At the end of the study, to evaluate our work, you will be asked to join an interview to ask you about the interests that we recognised from your data and understand what interests did we miss and why did we miss them.

You also have the option to participate in two further ways:

1. You can opt-in to four mini-interviews (one every six weeks for around 10-15 mins each) in which we ask you about your recent socialising and drinking activities. These interviews will take place online.
2. You can choose to borrow a Fitbit smartwatch from the research team to wear during the study. At the end of the study period you will return this device to us and can choose whether or not to share with us any location traces captured by this device.

Both of the above are <u>optional</u>. You can participate in the study without doing either of these.

- **Will I be compensated for taking part?**

As a token of appreciation, participants will be provided with an £18 Netflix gift card at the beginning of the experiment that covers three months basic subscription. After the third month, participants will be given the second £18 Netflix card that covers three additional months of Netflix subscription.

Participants who participate in the optional mini-interviews will receive their choice of an additional £18 Netflix or Amazon gift card at the end of the study period.

- **What happens if I do not want to take part or if I change my mind?**

It is up to you to decide whether or not to take part. If you do decide to take part, you will be given this information sheet to keep and be asked to sign a consent form.

If you decide to take part, you are still free to withdraw at any time without giving a reason and without detriment to yourself. You will need to let us know whether or not to keep your data. If requested, your data will be removed from the dataset and will not be used in any future publications. This does not affect your data protection rights.

If you choose to participate in the optional mini-interviews, then you must also consent to those interviews being audio recorded. However, we aim to ensure that you are comfortable with the recording process at all times and you are free to stop the interview and/or recording at any time.

If you decide to withdraw before the end of the third month, you will not be eligible for further compensation, and unfortunately, we will not be able to give you the second gift card. Likewise, if you do not participate through to the end of the study, we will be unable to provide you with a gift card for mini-interview participation.

Data Protection and Confidentiality

- **What information will you collect about me?**

In order to participate in this research project we will need to collect your contact details and information that could indirectly identify you, such as location data. Please see the table above for more details.

If you choose to participate in the optional mini-interviews, then we will audio-record these using the researcher computer that is used to make the call. The hard drive on this device will be encrypted.

If you choose to borrow a FitBit device then you also have the option to contribute the location traces captured using this device.

- **Under what legal basis are you collecting this information?**

We are collecting and storing your information in accordance with data protection law which protects your rights. These state that we must have a legal basis (specific reason) for collecting your data. For this study, the specific reason is that it is a process necessary for research purposes only.

- **What are my rights in relation to the information you will collect about me?**

You have a number of rights under data protection law regarding your personal information. For example, you can request a copy of the information that we hold about you.

If you would like to know more about your different rights or the way we use your personal information to ensure we follow the law, please consult our Privacy Notice for Research.

- **Will my participation in the study be confidential and my personal identifiable information be protected?**

In accordance with data protection law, The University of Manchester is the Data Controller for this project. This means that we are responsible for making sure your personal information is kept secure, confidential and used only in the way you have been told it will be used. All researchers are trained with this in mind, and your data will be looked after in the following way:

- To ensure confidentiality of data in digital format:
    - De-identify the data via an assigned participant ID only known to the research team (also referred to as pseudonymised or coded data).
    - Encrypt the submission of data between the participant's phone and the university's server to ensure individuals cannot be readily identified.
    - Remove any identifying or sensitive data revealed in interviews during the process of audio transcription.
    - Audio recordings will be transferred securely to a University-authorised third-party transcription service. Audio recordings will be deleted following transcription.

    - De-identify numbers and emails contained in the keyboard and notification texts by replacing them with asterisks.
    - Only the study team at The University of Manchester will have access to your data
- To ensure confidentiality of data in non-digital format:
    - We will use locked cabinets inside Kilburn building to store your consent form and personal information.
    - Data in non-digital formats (e.g. your consent form) will be digitized within one month of joining the experiment and securely destroyed. The digitised data will be stored with the collected data in secure servers hosted by the University of Manchester.
- Data could be shared for academic research purposes only and will be anonymised before sharing with other institutions. Any identifiable raw data will not be shared and will be either encrypted or replaced by other unidentifiable data. For example, GPS coordinates will be replaced by the general category of the place which they represent (e.g. restaurant). The below table indicates how each collected data will be prepared for sharing.

Sensor	Shared data
GPS	Sequences of places categories: e.g. home → restaurant → home
Weather	Status: cold, hot, nice
Applications	App usage data
Notifications	Will not be shared.
Screen	Screen status: On/Off
Wifi	Access points labels and ids will be shared in an encrypted form.
Bluetooth	Devices information will be encrypted before sharing.
Battery	Charging and discharging times will be shared
Calls	Calls duration, time, type (received/incoming/missed). Numbers will be shared in an encrypted format.
Messages	Message types (outgoing, incoming) and times. Numbers will be shared in an encrypted format. Message content is not recorded.
Keyboard	Only keystrokes times will be shared.
Consent form	Will not be shared.
Questionnaire	Responses associated with anonymised IDs will be shared
Activity Recognition	Activity types and times will be shared
Interview Responses (optional)	Anonymised transcripts will be shared. Any personal or sensitive information (including, for example, names of people or places) will be redacted.
Fitbit Location Data (optional)	As GPS.

- If during the study, we become aware of evidence about any current or future illegal activities, we have a legal obligation to report this and will, therefore, need to inform the relevant authorities.

Please also note that individuals from The University of Manchester or regulatory authorities may need to look at the data collected for this study to make sure the project is being carried out as planned. This may involve looking at identifiable data. All individuals involved in auditing and monitoring the study will have a strict duty of confidentiality to you as a research participant.

What if I have a complaint?

- **Contact details for complaints**

If you have a complaint that you wish to direct to members of the research team, please contact:

Name	Email	Address	Telephone
Simon Harper	**Simon.harper@manchester.ac.uk**	2.60 Kilburn building, M13 9PL Manchester, UK	0161 275 0599
Sarah Clinch	**Sarah.clinch@manchester.ac.uk**	2.24 Kilburn building, M13 9PL Manchester, UK	0161 275 7190

Note that the above telephone numbers are the research team's work telephones may not be redirected during periods where the government advice is to work from home. Emails will continue to be monitored during these periods.

If you wish to make a formal complaint to someone independent of the research team or if you are not satisfied with the response you have gained from the researchers in the first instance then please contact

The Research Governance and Integrity Officer, Research Office, Christie Building, The University of Manchester, Oxford Road, Manchester, M13 9PL, by emailing: research.complaints@manchester.ac.uk or by telephoning 0161 275 2674.

If you wish to contact us about your data protection rights, please email dataprotection@manchester.ac.uk or write to The Information Governance Office, Christie Building, The University of Manchester, Oxford Road, M13 9PL at the University and we will guide you through the process of exercising your rights.

You also have a right to complain to the Information Commissioner's Office about complaints relating to your personal identifiable information Tel 0303 123 1113

Link for Information Commissioner's Office: https://ico.org.uk/concerns

Contact Details

If you have comments or questions, please contact:

Name	Email	Address
Ahmed Ibrahim	**ahmed.ibrahim-2@manchester.ac.uk**	IAM lab, Kilburn Building. M13 9PL. Manchester, UK

B.2 Consent form

This page is left blank.

MANCHESTER
1824
The University of Manchester

Consent Form

Please read the participant information sheet (Version 1.8, 14/11/2020) before signing this document.

If you are happy to participate in study then please complete and sign the consent form below

General		Initials
1	I confirm that I have read the participant information sheet (Version 1.8, 14/11/2020) for the above study. I have had the opportunity to consider the information and ask questions. I am happy that any questions have answered satisfactorily.	
2	I understand that my participation in the study is voluntary and that I am free to withdraw at any time – up until two months after the end date of the study - without giving a reason and without detriment to myself. I agree to take part on this basis.	
3	I understand that when I withdraw from the experiment, I can opt to remove my data from the dataset and prevent its use in any future publications.	

Data Collection		Initials
4	I agree to participate in two interviews: one at the beginning of the experiment about my personal interests, how I socialise and how much alcohol I consume; and one at the end of the experiment that reflects on the data captured and inferences made by the researchers.	
5	I agree to install and run the application for research purposes. I understand that the application will collect the following data: *location coordinates (longitude and latitude), app usage, notifications, call duration and times, message send/receive times, screen interactions, keyboard interactions, performed activities, WiFi and Bluetooth activity,* and *weather data*. This data will be to understand my daily behaviour.	
6	I agree to receive notifications on my phone that ask questions about my interests.	
7	**[OPTIONAL]** I agree to participate in four mini-interviews (one every six weeks) conducted using video conferencing software. The interviews will discuss my recent socialising and drinking behaviour. I understand that these will be audio recorded and sent to a third party for transcription.	*(optional)*

Data Use		Initials
8	I agree that any data collected may be published in anonymous form in academic books, conference, reports, journals or thesis.	
9	I understand that data collected during the study may be looked at by individuals from The University of Manchester or regulatory authorities. I give permission for these individuals to have access to my data.	
10	I understand that there may be instances where during the course of the study information is revealed which means that the researchers will be obliged to break confidentiality and this has been explained in more detail in the information sheet.	
11	**[OPTIONAL]** I agree that the researchers may retain my contact details in order to provide me with a summary of the findings for this study.	*(optional)*

Overall Statement of Consent

		Initials
12	I agree to take part in the study (six months duration).	

Data Protection

The personal information we collect and use to conduct this research will be processed in accordance with data protection law as explained in the Participant Information Sheet and the Privacy Notice for Research Participants.

______________________	______________________	____________
Name of Participant	Signature	Date
______________________	______________________	____________
Name of person taking consent	Signature	Date

You can choose to return an electronic copy of this form. We suggest doing this either as a non-editable document (e.g. PDF) with an electronic representation of your signature, or as a photograph of a signed paper copy. If you choose to email the electronic form then we will download the document and store it securely before deleting the email to disassociate it from your contact details.

Once completed, you should receive a copy of this form with both signatures on it. Please do keep this for your records.

B.3 Interview guide

This page is left blank.

Interview Type: Semi Structure
Location: IAM lab
Interviewer: Ahmed Ibrahim
Expected duration: 10 – 15 minutes
General information:

- These interviews will be recorded using a laptop with an encrypted hard drive and will be deleted at the end date of the experiment.
- Each interview will be identified by a unique ID that represents the participant, location, start and end time of the interview.
- Transcription will be conducted by the Ahmed Ibrahim and will be stored on an encrypted hard drive.
- The interview will be used as a supplementary source of ground truth.

Questions of the first interview (when joining the study):

Open questions

- How would you describe your gender?
- Tell us please about your daily routine.?
- What are the places that you prefer to go to?
- How often do you go to those places?
- What motivates you to go to those places?
- In general, what are your interests and how often do you practice them?

Questions of the second interview (at the end of the experiment):

- How close are we in understanding your interests and what we missed?
- We will provide participants with the common places they visit which we discovered from their data and ask them to rank those places based on their interests.
- To what extent do you think that the data we captured accurately describes your socialising activities? Can you tell us what we missed?
- To what extent do you think that the data we captured accurately describes your drinking (alcohol) activities? Can you tell us what we missed?

Appendix C

Supplement for the Evaluation Scale: Intrinsic Motivation Inventory

C.1 Intrinsic Motivation Inventory

Intrinsic Motivation Inventory (IMI) is a post-experimental and multidimensional scale used widely to assess intrinsic motivation toward activities (Ryan, 2018). Although the authors of the IMI scale suggest that slight modifications on the wording of the scale's items would not affect its reliability nor validity, we have conducted an experiment to ensure that our minor changes have no impact. Specifically, we have conducted a confirmatory factor analysis replicating to the one of McAuley et al. which is referenced by the IMI's authors and used to validate the published scale. 77 students and parents joined the experiment, which we conducted at the open days of the University of Manchester. During the open days, the university provides the chance for students and their parents to learn more about the study fields and environment[1]. We have advertised our IMI experiment as part of the activities that students and parents can do during the open days. The experiment was run during the open days of 2018 and 2019. Of the 77 participants, 22 joined in 2018 and the remaining 55 participated in the two open days of 2019.

[1] For more information about the open days, please visit: https://www.manchester.ac.uk/study/undergraduate/open-days-visits/open-days/

C.1.1 Method

The original version of IMI requires some tasks to be completed before conducting the survey. To account for the different dimensions covered by IMI, we chose three tasks that differ cognitively, and participants were free to select which one to perform. The proposed activities are Sudoku, Origami, and a 'HORSE' like basketball game whereby two players are asked to compete by replicating a shot. Missed shots result in gradually forming the word 'HORSE' and the first player to form the word loses the game.

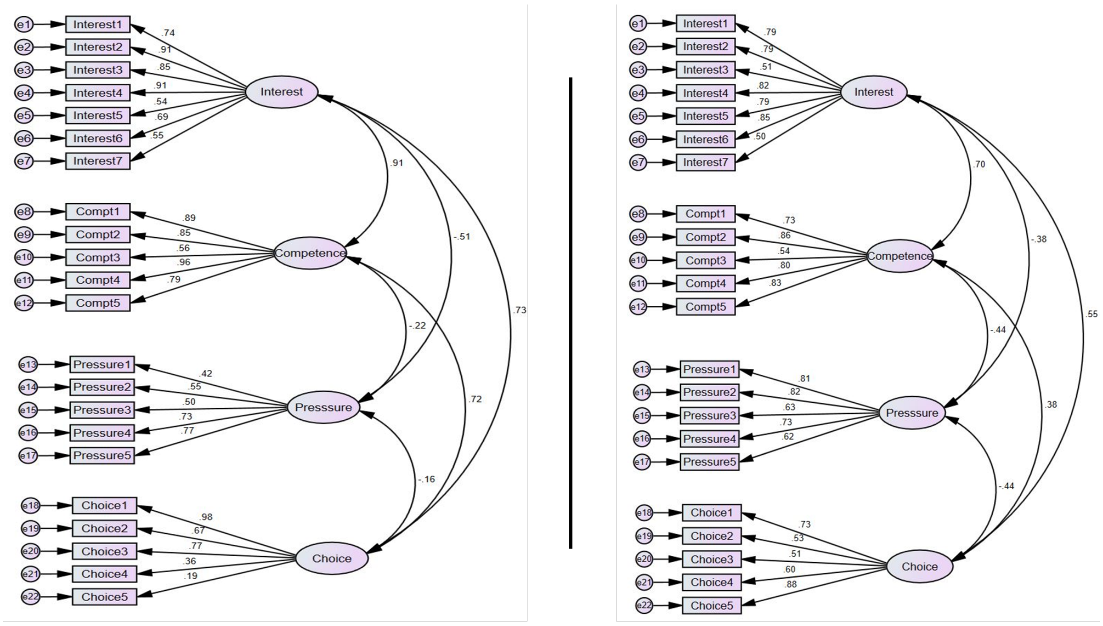
We based our study on the 22-items version of the IMI (see Section C.4). The 22-items version is widely used and shown to be reliable across multiple domains Ryan (2018). It covers four dimensions: interest/enjoyment, perceived competence, perceived choice and pressure/tension, and its scoring system is based on a seven-point Likert scale. we refer to the original post-experimental version as IMI_o, and we used it to evaluate the performed tasks. Unlike IMI_o, the modified version (IMI_m) explicitly integrates the tasks within the scale while reflecting the modifications required by our work. Specifically, we modify the phraseology from the past to the present tense to express the longevity aspect of interest.

Ethics application that contains a description of the study was completed to obtain the approval from the school of computer science at the University of Manchester (see the participant information sheet under Section C.2) . The study description - which includes steps of the experiment as well as information related to preserving the anonymity and confidentiality of the data - was shared with the participants. To comply with the ethics requirements, all participants had to provide formal consent before joining the study (Section C.3).

As noted earlier, we conducted the experiment in the open day at the University of Manchester, participants or their parents were asked to sign the consent form before joining the experiment. They had the chance to select which one of the three activities to perform. Upon the task completion, the participants completed the modified version of the IMI scale that is related to the completed task.

C.1.2 Results

We have conducted a Confirmatory Factor Analysis (CFA) to replicate the process performed by the related works that aimed at adapting the IMI scale. Rather than providing a conclusive analysis, our goal was to ensure that the trend of the calculated



a) Factor analysis after the first year open days (22 participants)

b) Factor analysis after the second year open days (77 participants)

Figure C.1: The factor analysis of the participants' responses to the IMI.

parameters does not contradict with ones of the original IMI. Specifically the loadings of the items and cross loadings between the four dimensions. Therefore, we have conducted the CFA after the open days of 2018 (with 22 participants) and then for the entire 77 participants after the open days of 2019.

Figures C.1a and C.1b show that the loadings values of most dimension's items are stabalising above the threshold identified by the original IMI scale (i.e. 0.6). Although this is true for all dimensions, we have paid a special attention to the interest subscale as it is the one that is used by the work of this thesis. The third and the seventh items of the interest subscale fall below the 0.6 threshold. However, we have noticed that all the loadings that fall below 0.6 belong to reversed items (i.e. the scores of these items are reversed with respect to the measured variable). This may indicate an impact of the reverse scoring on the participants' responses.

With respect to the cross loadings, Figures Figures C.1a and C.1b show the impact of increasing the number of participants in separating the contributing factors (i.e. dimensions). The threshold identified by the original IMI scale is 0.4 for cross loading. After the first open day, the factor separation was not evident as the calculated loadings were larger than the specified threshold. The inclusion of more participants changed that and started to indicate the independence between each factor. Moreover, the positive and negative relation between the factors seems to move to the expected direction.

Competence, choice and interest are positively related whereas pressure negatively related with the previous three factors.

C.2 Participant information sheet for the IMI experiment

This page is left blank.

An adapted scale for the intrinsic motivation inventory

Participant Information Sheet (PIS)

This PIS should be read in conjunction with The University privacy notice: http://documents.manchester.ac.uk/display.aspx?DocID=37095

You are being invited to take part in a research study as part of a student project that aims to recognise human motivation from smartphone's data. Before you decide whether to take part, it is important for you to understand why the research is being conducted and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please ask if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for taking the time to read this.

Who will conduct the research?

Ahmed Ibrahim, a PhD student at the University of Manchester (school of computer science).

What is the purpose of the research?

Assess the participants' opinions with respect to suggested motivations.

Why have I been chosen?

All participants were chosen randomly.

What would I be asked to do if I took part?

Perform one of the following tasks (Sudoku, origami, or basketball) and complete a questionnaire (called intrinsic motivation inventory or shortly IMI) that assesses your motivation toward the performed activity.

What will happen to my personal information?

In order to undertake the research project we will need to collect the following personal information/data about you:

- Age.
- Gender.
- Education.

Only the research team will have access to this information. We are collecting and storing this personal information in accordance with the General Data Protection Regulation (GDPR) and Data Protection Act 2018 which legislate to protect your personal information. The legal basis upon which we are using your personal information is "public interest task" and "for research purposes" if sensitive information is collected. For more information about the way we process your personal information and comply with data protection law please see our Privacy Notice for Research Participants: (http://documents.manchester.ac.uk/display.aspx?DocID=37095).

The University of Manchester, as Data Controller for this project takes responsibility for the protection of the personal information that this study is collecting about you. In order to comply with the legal obligations to protect your personal data the University has safeguards in place such as policies and procedures. All researchers are appropriately trained and your data will be looked after in the following way:

The study team at the University of Manchester will have access to your personal identifiable information, that is data which could identify you, but they will anonymise it as soon as practical. However your consent form, contact details, and your responses to the intrinsic motivation inventory items will be transferred to and retained in a secured server at the University of Manchester for 5 years.

You have a number of rights under data protection law regarding your personal information. For example you can request a copy of the information we hold about you. This is known as a Subject Access Request. If you would like to know more about your different rights, please consult our privacy notice for research: (http://documents.manchester.ac.uk/display.aspx?DocID=37095),and if you wish to contact us about your data protection rights, please email dataprotection@manchester.ac.uk or write to The Information Governance Office, Christie Building, University of Manchester, Oxford Road, M13 9PL. at the University and we will guide you through the process of exercising your rights.

You also have a right to complain to the Information Commissioner's Office, Tel 0303 123 1113

Will my participation in the study be confidential?

Your participation in the study will be kept confidential to the study team and those with access to your personal information as listed above.

- Individuals from the University, the site where the research is taking place and regulatory authorities may need to review the study information for auditing and monitoring purposes or in the event of an incident.
- In the event that there are concerns about the participant's safety or the safety of others you may need to contact their GP/care team/family member.

To ensure confidentiality, we will

- De-identify the data and linking to the individual via an assigned participant ID only known to research team (also referred to pseudonymised or coded data).
- Ensure the reporting of the data is done in such a way that individuals cannot be readily identified.

What happens if I do not want to take part or if I change my mind?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time without giving a reason and without detriment to yourself.

However, it will not be possible to remove your data from the project once it has been anonymised and forms part of the dataset as we will not be able to identify your specific data. This does not affect your data protection rights.

Will my data be used for future research?

When you agree to take part in a research study, the information about your health and care may be provided to researchers running other research studies in this organisation. The future research should not be incompatible with this research project and will concern Human Computer Interaction. These organisations may be universities, NHS organisations or companies involved in health and care research in this country or abroad. Your information will only be used by organisations and researchers to conduct research in accordance with the UK Policy Framework for Health and Social Care Research.

This information will not identify you and will not be combined with other information in a way that could identify you. The information will only be used for the purpose of health and care research, and cannot be used to contact you regarding any other matter or to affect your care. It will not be used to make decisions about future services available to you.

Will I be paid for participating in the research?

No.

What is the duration of the research?

It is expected to be 10 minutes (5 minutes for the task and 5 minutes for the questionnaire).

Where will the research be conducted?

At the University of Manchester – Kilburn building.

Will the outcomes of the research be published?

Statistical information and comparisons between the versions will be published.

Who has reviewed the research project?

The project has been reviewed by the University of Manchester Research Ethics Committee/ the school of computer science ethics committee.

What if I want to make a complaint?

Please contact Dr. Simon Harper: simon.harper@manchester.ac.uk

Minor complaints

If you have a minor complaint then you need to contact Ahmed Ibrahim in the first instance:

Email: ahmed.ibrahim-7@postgrad.manchester.ac.uk

Formal Complaints

If you wish to make a formal complaint or if you are not satisfied with the response you have gained from the researchers in the first instance then please contact

The Research Governance and Integrity Manager, Research Office, Christie Building, University of Manchester, Oxford Road, Manchester, M13 9PL, by emailing: research.complaints@manchester.ac.uk or by telephoning 0161 275 2674.

What Do I Do Now?

If you have any queries about the study or if you are interested in taking part then please contact

Ahmed Ibrahim (**Email:ahmed.ibrahim-7@postgrad.manchester.ac.uk**)

This Project Has Been Approved by the University of Manchester's Research Ethics Committee
[2018-3917-6214**]**

C.3 Consent form for the IMI experiment

This page is left blank.

MANCHESTER 1824
The University of Manchester

An adapted scale for the intrinsic motivation inventory

Consent Form

If you are happy to participate please complete and sign the consent form below

	Activities	Initials
1	I confirm that I have read the attached information sheet (Version 1.0, 04/06/2018) for the above study and have had the opportunity to consider the information and ask questions and had these answered satisfactorily.	
2	I understand that my participation in the study is voluntary and that I am free to withdraw at any time without giving a reason and without detriment to myself. I understand that it will not be possible to remove my data from the project once it has been anonymised and forms part of the data set. I agree to take part on this basis	
3	I agree that any data collected may be published in anonymous form in academic books, reports, study, conference or journals	
4	I agree that the researchers/researchers at other institutions may contact me in future about other research projects.	
5	I agree that the researchers may retain my contact details in order to provide me with a summary of the findings for this study.	
6	I agree to take part in this study	

Data Protection

The personal information we collect and use to conduct this research will be processed in accordance with data protection law as explained in the Participant Information Sheet and the [Privacy Notice for Research Participants](http://documents.manchester.ac.uk/display.aspx?DocID=37095) (http://documents.manchester.ac.uk/display.aspx?DocID=37095).

________________________ ________________________ _____________
Name of Participant Signature Date

________________________ ________________________ _____________
Name of the person taking consent Signature Date

1 copy for the research team (original)
1 copy for the participant

C.4 The original IMI scale

Following is the original IMI scale directly obtained as a pdf file from https://selfdeterminationtheory.org/intrinsic-motivation-inventory/. Although there is some typographic errors in the downloaded PDF, we have included the file as is since it represents the formal version of the scale.

Intrinsic Motivation Inventory (IMI)

Scale Description

The Intrinsic Motivation Inventory (IMI) is a multidimensional measurement device intended to assess participantsÕ subjective experience related to a target activity in laboratory experiments. It has been used in several experiments related to intrinsic motivation and self-regulation (e.g., Ryan, 1982; Ryan, Mims & Koestner, 1983; Plant & Ryan, 1985; Ryan, Connell, & Plant, 1990; Ryan, Koestner & Deci, 1991; Deci, Eghrari, Patrick, & Leone, 1994). The instrument assesses participantsÕ interest/enjoyment, perceived competence, effort, value/usefulness, felt pressure and tension, and perceived choice while performing a given activity, thus yielding six subscale scores. Recently, a seventh subscale has been added to tap the experiences of relatedness, although the validity of this subscale has yet to be established. The **interest/enjoyment subscale is considered the self-report measure of intrinsic motivation**; thus, although the overall questionnaire is called the Intrinsic Motivation Inventory, it is only the one subscale that assesses intrinsic motivation, per se. As a result, the interest/enjoyment subscale often has more items on it that do the other subscales. The perceived choice and perceived competence concepts are theorized to be positive predictors of both self-report and behavioral measures of intrinsic motivation, and pressure/tension is theorized to be a negative predictor of intrinsic motivation. Effort is a separate variable that is relevant to some motivation questions, so is used it its relevant. The value/usefulness subscale is used in internalization studies (e.g., Deci et al, 1994), the idea being that people internalize and become self-regulating with respect to activities that they experience as useful or valuable for themselves. Finally, the relatedness subscale is used in studies having to do with interpersonal interactions, friendship formation, and so on.

The IMI consists of varied numbers of items from these subscales, all of which have been shown to be factor analytically coherent and stable across a variety of tasks, conditions, and settings. The general criteria for inclusion of items on subscales have been a factor loading of at least 0.6 on the appropriate subscale, and no cross loadings above 0.4. Typically, loadings substantially exceed these criteria. Nonetheless, we recommend that investigators perform their own factor analyses on new data sets. Past research suggests that order effects of item presentation appear to be negligible, and the inclusion or exclusion of specific subscales appears to have no impact on the others. Thus, it is rare that all items have been used in a particular experiment. Instead, experimenters have chosen the subscales that are relevant to the issues they are exploring.

The IMI items have often been modified slightly to fit specific activities. Thus, for example, an item such as ÒI tried very hard to do well at this activityÓ can be changed to ÒI tried very hard to do well on these puzzlesÓ or Ò...in learning this materialÓ without effecting its reliability or validity. As one can readily tell, there is nothing subtle about these items; they are quite face-valid. However, in part, because of their straightforward nature, caution is needed in interpretation. We have found, for example, that correlations between self-reports of effort or interest and behavioral indices of these dimensions are quite modest--often around 0.4. Like other self-report measures, there is always the need to appropriately interpret how and why participants report as they do. Ego-involvements, self-presentation styles, reactance, and other psychological dynamics must be considered. For example, in a study by Ryan, Koestner, and Deci (1991), we found that when participants were ego involved, the engaged in pressured persistence during a free choice period and this behavior did not correlate with the self-reports of interest/enjoyment. In fact, we concluded that to be confident in oneÕs assessment of intrinsic motivation, one needs to find that the free-choice behavior and the self-reports of interest/enjoyment are significantly correlated.

Another issue is that of redundancy. Items within the subscales overlap considerably, although randomizing their presentation makes this less salient to most participants. Nonetheless, shorter versions have been used and been found to be quite reliable. The incremental R for every item above 4 for any given factor is quite small. Still, it is very important to recognize that multiple item subscales consistently outperform single items for obvious reasons, and they have better external validity.

On The Scale page, there are five sections. First, the full 45 items that make up the 7 subscales are shown, along with information on constructing your own IMI and scoring it. Then, there are four specific versions of the IMI that have been used in past studies. This should give you a sense of the different ways it has been used. These have different numbers of items and different numbers of subscales, and they concern different activities. First, there is a standard, 22-item version that has been used in several studies, with four subscales: interest/enjoyment, perceived competence, perceived choice, and pressure/tension. Second, there is a short 9-item version concerned with the activity of reading some text material; it has three subscales: interest/enjoyment, perceived competence, and pressure/tension. Then, there is the 25-item version that was used in the internalization study, including the three subscales of value/usefulness, interest/enjoyment, and perceived choice. Finally, there is a 29-item version of the interpersonal relatedness questionnaire that has five subscales: relatedness, interest/enjoyment, perceived choice, pressure/tension, and effort.

Finally, McAuley, Duncan, and Tammen (1987) did a study to examine the validity of the IMI and found strong support for its validity.

References

Deci, E. L., Eghrari, H., Patrick, B. C., & Leone, D. (1994). Facilitating internalization: The self-determination theory perspective. *Journal of Personality*, *62*, 119-142.

McAuley, E., Duncan, T., & Tammen, V. V. (1987). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport, 60,* 48-58.

Plant, R. W., & Ryan, R. M. (1985). Intrinsic motivation and the effects of self-consciousness, self-awareness, and ego-involvement: An investigation of internally-controlling styles. *Journal of Personality*, *53*, 435-449.

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, *43*, 450-461.

Ryan, R. M., Connell, J. P., & Plant, R. W. (1990). Emotions in non-directed text learning. *Learning and Individual Differences*, *2*, 1-17.

Ryan, R. M., Koestner, R., & Deci, E. L. (1991). Varied forms of persistence: When free-choice behavior is not intrinsically motivated. *Motivation and Emotion*, *15*, 185-205.

Ryan, R. M., Mims, V., & Koestner, R. (1983). Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *Journal of Personality and Social Psychology*, *45*, 736-750.

The Scales

THE POST-EXPERIMENTAL INTRINSIC MOTIVATION INVENTORY
(Below are listed all 45 items that can be used depending on which are needed.)

For each of the following statements, please indicate how true it is for you, using the following scale:

1 2 3 4 5 6 7

not at all true	somewhat true	very true

Interest/Enjoyment

I enjoyed doing this activity very much
This activity was fun to do.
I thought this was a boring activity. (R)
This activity did not hold my attention at all. (R)
I would describe this activity as very interesting.
I thought this activity was quite enjoyable.
While I was doing this activity, I was thinking about how much I enjoyed it.

Perceived Competence
I think I am pretty good at this activity.
I think I did pretty well at this activity, compared to other students.
After working at this activity for awhile, I felt pretty competent.
I am satisfied with my performance at this task.
I was pretty skilled at this activity.
This was an activity that I couldnÕt do very well. (R)

Effort/Importance
I put a lot of effort into this.
I didnÕt try very hard to do well at this activity. (R)

I tried very hard on this activity.
It was important to me to do well at this task.
I didnÕt put much energy into this. (R)

Pressure/Tension
I did not feel nervous at all while doing this. (R)
I felt very tense while doing this activity.
I was very relaxed in doing these. (R)
I was anxious while working on this task.
I felt pressured while doing these.

Perceived Choice
I believe I had some choice about doing this activity.
I felt like it was not my own choice to do this task. (R)
I didnÕt really have a choice about doing this task. (R)
I felt like I had to do this. (R)
I did this activity because I had no choice. (R)
I did this activity because I wanted to.
I did this activity because I had to. (R)

Value/Usefulness
I believe this activity could be of some value to me.
I think that doing this activity is useful for ______________________
I think this is important to do because it can ______________________
I would be willing to do this again because it has some value to me.
I think doing this activity could help me to ______________________
I believe doing this activity could be beneficial to me.
I think this is an important activity.

Relatedness
I felt really distant to this person. (R)
I really doubt that this person and I would ever be friends. (R)
I felt like I could really trust this person.
IÕd like a chance to interact with this person more often.
IÕd really prefer not to interact with this person in the future. (R)
I donÕt feel like I could really trust this person. (R)
It is likely that this person and I could become friends if we interacted a lot.
I feel close to this person.

Constructing the IMI for your study. First, decide which of the variables (factors) you want to use, based on what theoretical questions you are addressing. Then, use the items from those factors, randomly ordered. If you use the value/usefulness items, you will need to complete the three items as appropriate. In other words, if you were studying whether the person believes an activity is useful for improving concentration, or becoming a better basketball player, or whatever, then fill in the blanks with that information. If you do not want to refer to a particular outcome, then just truncate the items with its being useful, helpful, or important.

Scoring information for the IMI. To score this instrument, you must first reverse score the items for which an (R) is shown after them. To do that, subtract the item response from 8, and use the resulting number as the item score. Then, calculate subscale scores by averaging across all of the items on that subscale. The subscale scores are then used in the analyses of relevant questions.

* * * * * * * * * * * *

The following is a 22 item version of the scale that has been used in some lab studies on intrinsic motivation. It has four subscales: interest/enjoyment, perceived choice, perceived competence, and pressure/tension. The interest/enjoyment subscale is considered the self-report measure of intrinsic motivation; perceived choice and perceived competence are theorized to be positive predictors of both self-report and behavioral measures of intrinsic motivation. Pressure tension is theorized to be a negative predictor of intrinsic motivation. Scoring information is presented after the questionnaire itself.

TASK EVALUATION QUESTIONNAIRE

For each of the following statements, please indicate how true it is for you, using the following scale:

1	2	3	4	5	6	7
not at all true			somewhat true			very true

1. While I was working on the task I was thinking about how much I enjoyed it.
2. I did not feel at all nervous about doing the task.
3. I felt that it was my choice to do the task.
4. I think I am pretty good at this task.
5. I found the task very interesting.
6. I felt tense while doing the task.
7. I think I did pretty well at this activity, compared to other students.

8. Doing the task was fun.

9. I felt relaxed while doing the task.

10. I enjoyed doing the task very much.

11. I didnÕt really have a choice about doing the task.

12. I am satisfied with my performance at this task.

13. I was anxious while doing the task.

14. I thought the task was very boring.

15. I felt like I was doing what I wanted to do while I was working on the task.

16. I felt pretty skilled at this task.

17. I thought the task was very interesting.

18. I felt pressured while doing the task.

19. I felt like I had to do the task.

20. I would describe the task as very enjoyable.

21. I did the task because I had no choice.

22. After working at this task for awhile, I felt pretty competent.

Scoring information. Begin by reverse scoring items # 2, 9, 11, 14, 19, 21. In other words, subtract the item response from 8, and use the result as the item score for that item. This way, a higher score will indicate more of the concept described in the subscale name. Thus, a higher score on pressure/tension means the person felt more pressured and tense; a higher score on perceived competence means the person felt more competent; and so on. Then calculate subscale scores by averaging the items scores for the items on each subscale. They are as follows. The (R) after an item number is just a reminder that the item score is the reverse of the participantÕs response on that item.

Interest/enjoyment: 1, 5, 8, 10, 14(R), 17, 20
Perceived competence: 4, 7, 12, 16, 22
Perceived choice: 3, 11(R), 15, 19(R), 21(R)
Pressure/tension: 2(R), 6, 9(R), 13, 18

The subscale scores can then be used as dependent variables, predictors, or mediators, depending on the research questions being addressed.

* * * * * * * * * * * *

TEXT MATERIAL QUESTIONNAIRE I

For each of the following statements, please indicate how true it is for your, using the following scale as a guide:

1	2	3	4	5	6	7
not at all true			somewhat true			very true

1. While I was reading this material, I was thinking about how much I enjoyed it.
2. I did not feel at all nervous while reading.
3. This material did not hold my attention at all.
4. I think I understood this material pretty well.
5. I would describe this material as very interesting.
6. I think I understood this material very well, compared to other students.
7. I enjoyed reading this material very much.
8. I felt very tense while reading this material.
9. This material was fun to read.

Scoring information. Begin by reverse scoring items # 2 and 3. In other words, subtract the item response from 8, and use the result as the item score for that item. This way, a higher score will indicate more of the concept described in the subscale name. Then calculate subscale scores by averaging the items scores for the items on each subscale. They are shown below. The (R) after an item number is just a reminder that the item score is the reverse of the participantÕs response on that item.

Interest/enjoyment: 1, 3(R), 5, 7, 9
Perceived competence: 4, 6,
Pressure/tension: 2(R), 8

* * * * * * * * * * * *

The next version of the questionnaire was used for a study of internalization with an uninteresting computer task (Deci et al., 1994).

ACTIVITY PERCEPTION QUESTIONNAIRE

The following items concern your experience with the task. Please answer all items. For each item, please indicate how true the statement is for you, using the following scale as a guide:

1	2	3	4	5	6	7
not at all true			somewhat true			very true

1. I believe that doing this activity could be of some value for me.

2. I believe I had some choice about doing this activity.

3. While I was doing this activity, I was thinking about how much I enjoyed it.

4. I believe that doing this activity is useful for improved concentration.

5. This activity was fun to do.

6. I think this activity is important for my improvement.

7. I enjoyed doing this activity very much.

8. I really did not have a choice about doing this activity.

9. I did this activity because I wanted to.

10. I think this is an important activity.

11. I felt like I was enjoying the activity while I was doing it.

12. I thought this was a very boring activity.

13. It is possible that this activity could improve my studying habits.

14. I felt like I had no choice but to do this activity.

15. I thought this was a very interesting activity.

16. I am willing to do this activity again because I think it is somewhat useful.

17. I would describe this activity as very enjoyable.

18. I felt like I had to do this activity.

19. I believe doing this activity could be somewhat beneficial for me.

20. I did this activity because I had to.

21. I believe doing this activity could help me do better in school.

22. While doing this activity I felt like I had a choice.

23. I would describe this activity as very fun.

24. I felt like it was not my own choice to do this activity.

25. I would be willing to do this activity again because it has some value for me.

Scoring information. Begin by reverse scoring items # 8, 12, 14, 18, 20, and 24 by subtracting the item response from 8 and using the result as the item score for that item. Then calculate subscale scores by averaging the items scores for the items on each subscale. They are shown below. The (R) after an item number is just a reminder that the item score is the reverse of the participantÕs response on that item.

Interest/enjoyment:	3, 5, 7, 11, 12(R), 15, 17, 23
Value/usefulness:	1, 4, 6, 10, 13, 16, 19, 21, 25
Perceived choice:	2, 8(R), 9, 14(R), 18(R), 20(R), 22, 24(R)

* * * * * * * * * * * *

SUBJECT IMPRESSIONS QUESTIONNAIRE

The following sentences describe thoughts and feelings you may have had regarding the other person who participated in the experiment with you. For each of the following statement please indicate how true it is for you, using the following scale as a guide:

1	2	3	4	5	6	7
not at all true			somewhat true			very true

1. While I was interacting with this person, I was thinking about how much I enjoyed it.

2. I felt really distant to this person.

3. I did not feel at all nervous about interacting with this person.

4. I felt like I had choice about interacting with this person.

5. I would describe interacting with this person as very enjoyable.

6. I really doubt that this person and I would ever become friends.

7. I found this person very interesting.

8. I enjoyed interacting with this person very much.

9. I felt tense while interacting with this person.

10. I really feel like I could trust this person.

11. Interacting with this person was fun.

12. I felt relaxed while interacting with this person.

13. IÕd like a chance to interact more with this person.

14. I didnÕt really have a choice about interacting with this person.

15. I tried hard to have a good interaction with this person.

16. IÕd really prefer not to interact with this person in the future.

17. I was anxious while interacting with this person.

18. I thought this person was very boring.

19. I felt like I was doing what I wanted to do while I was interacting with this person.

20. I tried very hard while interacting with this person.

21. I donÕt feel like I could really trust this person.

22. I thought interacting with this person was very interesting.

23. I felt pressured while interacting with this person.

24. I think itÕs likely that this person and I could become friends.

25. I felt like I had to interact with this person.

26. I feel really close to this person.

27. I didnÕt put much energy into interacting with this person.

28. I interacted with this person because I had no choice.

29. I put some effort into interacting with this person.

Scoring information. Begin by reverse scoring items # 2, 3, 6, 12, 14, 16, 18, 21, 25, 27, and 28 by subtracting the item response from 8 and using the result as the item score for that item. Then calculate subscale scores by averaging the items scores for the items on each subscale. They are shown below. The (R) after an item number is just a reminder that the item score is the reverse of the participantÕs response on that item.

Relatedness:	2(R), 6(R), 10, 13, 16(R), 21(R), 24, 26
Interest/enjoyment:	1, 5, 7, 8, 11, 18(R), 22
Perceived choice:	4, 14(R), 19, 25(R), 28(R)
Pressure/tension:	3(R), 9, 12(R), 17, 23,
Effort:	15, 20, 27(R), 29

C.5 Adapted IMI Interest/Enjoyment scale

This page is left blank.

For each of the following statements,[2] please indicate how true it is for you, using the following scale:

1	2	3	4	5	6	7
not at all			somewhat true			true

- I enjoy [shopping] very much
- [Shopping] is fun to do.
- I think [shopping] is boring.[3]
- [Shopping] does not hold my attention at all.
- I would describe [shopping] as very interesting.
- I think [shopping] is quite enjoyable.
- While [shopping], I am thinking about how much I enjoy it.

[2] In each statement, the placeholder [shopping] is illustrative and would be replaced by any identified IMB.

[3] Scores for this item, and the one immediately following, should be reversed.

Appendix D

Case Study: Covid-19 and the MIR

In this appendix, we present a case related to COVID-19. Specifically, we include a paper that details how digital phenotyping and deriving behavioural and personal knowledge can benefit the personalisation of nudges related to COVID-19. The paper relies on our work to observe, extract and understand behavioural and personal aspects from smartphone's data. Although this paper is not part of the research questions or contributions that we detailed in the first chapter, it explains how personal knowledge similar to the one discussed in this thesis can be applied in different domains. This shows that our work is not overfitted to one particular domain.

The main content of this appendix is a paper authored by: *Ahmed Ibrahim, Heng Zhang, Sarah Clinch, Ellen Poliakoff, Bijan Parsia, Simon Harper*. The title of the paper is: ***Digital Phenotypes for Understanding Individuals' Compliance With COVID-19 Policies and Personalized Nudges: Longitudinal Observational Study***. The paper is published in JMIR Formative Research (https://formative.jmir.org), May 2021. Volume: 5. ISSN: 2561-326X. DOI: 10.2196/23461. URL: https://formative.jmir.org/2021/5/e23461. For this thesis, we edited some formatting styles, such as the sizes of some tables for consistency and readability reasons.

Author contribution

Ahmed Ibrahim and Heng Zhang contributed equally to this paper. They designed the included studies, analysed and synthesised the results and wrote the paper. Sarah Clinch, Ellen Poliakoff, Bijan Parsia, and Simon Harper provided continuous feedback throughout all the stages of the study, offered advice and discussion and contributed vital edits to the paper's writing.

Abstract

Background Governments promote behavioral policies such as social distancing and phased reopening to control the spread of COVID-19. Digital phenotyping helps promote the compliance with these policies through the personalized behavioral knowledge it produces.

Objectives This study investigated the value of smartphone-derived digital phenotypes in (1) analyzing individuals' compliance with COVID-19 policies through behavioral responses and (2) suggesting ways to personalize communication through those policies.

Methods We conducted longitudinal experiments that started before the outbreak of COVID-19 and continued during the pandemic. A total of 16 participants were recruited before the pandemic, and a smartphone sensing app was installed for each of them. We then assessed individual compliance with COVID-19 policies and their impact on habitual behaviors.

Results Our results show a significant change in people's mobility ($P < .001$) as a result of COVID-19 regulations, from an average of 10 visited places every week to approximately 2 places a week. We also discussed our results within the context of nudges used by the National Health Service in the United Kingdom to promote COVID-19 regulations.

Conclusions Our findings show that digital phenotyping has substantial value in understanding people's behavior during a pandemic. Behavioral features extracted from digital phenotypes can facilitate the personalization of and compliance with behavioral policies. A rule-based messaging system can be implemented to deliver nudges on the basis of digital phenotyping.

D.1 Introduction

COVID-19 is a highly contagious disease with confirmed cases in more than 188 countries as between December 2019 and June 2020, resulting in a global pandemic (Wikipedia, 2020). To control the spread of COVID-19, governments have enforced behavioral policies, such as stay-at-home and social distancing measures, which limit the usual patterns of human interaction (Andersen, 2020; Briscese et al., 2020). The potential risk of problems with social isolation (Usher et al., 2020) complicates the implementation of these policies, which places an additional responsibility on governments to maintain mental health throughout the pandemic.

Currently, governments rely on communication campaigns to persuade people to adhere to COVID-19 behavioral policies and reduce disease spread. Health agencies, such as the National Health service (NHS) in the United Kingdom, design communication in a way that encourages the application of the promoted behaviors while avoiding problems related to social isolation. This approach to communications design employs behavioral insights derived from scientific studies to deliver behavioral guidance (Van Bavel et al., 2020). The communications resulting from this process are called "nudges" (Thaler and Sunstein, 2008).

Despite the critical role of these campaigns in elevating community awareness, they are not designed to reflect differently when people exhibit different behavioral responses to the promoted procedures. Digital devices including smartphones can be used to recognize behavioral differences. Accordingly, communications can be personalized and contextualized on the basis of the individual's behavior. Smartphones facilitate the capturing of behavioral features through the continuous and unobtrusive collection of sensor and interaction data; this process is known as "digital phenotyping."

In this study, we show how an individual's behavioral reactions to COVID-19 policies can be observed through digital phenotyping. Subsequently, we suggest a personalized way of delivering nudges designed around the individual's reactions to the enforced regulations. We report 2 longitudinal studies that started before the outbreak of the pandemic to collect digital phenotypes. Our studies allow us to observe the impact on the overall behavior before and during the outbreak. Additionally, we observed the impact of COVID-19 on habitual behaviors and the uptake of new apps.

Our primary research contribution is the introduction of an approach that employs behavioral differences derived from digital phenotyping in the design of personalized nudges. Although we did not conduct an experiment to measure the real-time effects of personalized nudges, the proposed nudges conform to the general guidelines in behavioral science and are expected to improve individual compliance to them. Moreover, the development of mental health issues as a result of lockdown policies can be observed through digital phenotyping and better addressed through personalized nudges.

D.2 Related work

With the popularity and evolution of personal electronic devices, people are producing an increasing number of digital footprints such as those generated through web-based communication and mobile device usage. These footprints can be linked and analyzed with clinical data to create an individualized, nuanced view of human disease, which is called a "digital phenotype" (Jain et al., 2015). In 2015, a digital phenotype was defined by Jukka-Pekka Onnela as the "moment-by-moment quantification of the individual-level human phenotype in-situ using data from smartphones and other personal digital devices" (Torous et al., 2016). Digital phenotyping has become one of the most innovative approaches to enhance health and wellness via human-computer interactions through digital technology.

Nowadays, smartphones have become the one of the ideal tools for digital phenotyping. Smartphones are the hub of personal communication, and almost everyone has a smartphone. Although smartphones are not specially designed for behavioral research, they can collect a large amount of related data directly and instantly with ecological validity. Social interaction on smartphones, including calls, messages, emails, and social media usage, can be captured without difficulty. Thus, social sensing could be less intrusive on smartphones than on any other device. Embedded multiple and power sensors also empower smartphones as an efficient tool to record the surrounding social context. For example, raw data from sensors such as microphones, the global positioning system (GPS), and accelerometers can be gathered and interpreted as conversation engagement, mobility patterns, and the number of encounters to infer social interaction occurring outside of smartphones. Thus, smartphones could be one of the most applicable ways of passive societal digital phenotyping.

Digital phenotyping on smartphones has been utilized in various fields, especially psychological and health-related studies. Abdullah et al. (2014) collected phone usage patterns to detect and predict discrepancies in sleep rhythms. Furthermore, LiKamWa et al. (2013) analyzed call, message, or email contacts and location clusters from smartphones to infer users' daily mood. Farhan et al. (2016) combined the locations and activities from participants' smartphones to predict depression. Boukhechba et al. (2017) explored the association of social anxiety with GPS and communication patterns. To confirm the findings and observations of passively collected smartphone data, all these studies asked for participants' input through various means including interviews, focus groups, and questionnaires. All these studies claimed to have relatively high accuracy. Albeit with different aims, our study similarly implemented these smartphone monitoring technologies. We collected data before and during the COVID-19 lockdown, which provided us an opportunity to observe individual behavioral changes. We also conducted interviews with our study participants to verify our findings.

D.3 Methods

D.3.1 Methods overview

We used behavioral indicators for the COVID-19 policies as proxies that would help us observe the adoption of the desired change by people. Our approach relies on transforming raw smartphone data collected longitudinally (i.e. digital phenotypes) into behavioral features. Distance travelled and time spent at home by a person are examples of features derived from raw location data (i.e. timestamped longitude and latitude attributes). The detection of behavioral indicators is achieved at the level of behavioral features rather than the raw data. This is because behavioral indicators are manifested at a higher level of human understanding expressible by those features. In the following section, we detail the behavioral features and their roles in recognizing the behavioral indicators of the proposed policies.

For this disease, transmitted through close contact, reducing the possibility of an uninfected person having physical contact with an infected person may be the only effective way to suppress the transmission of the disease. Since the onset of the COVID-19 pandemic, governments worldwide enforced a series of behavioral policies based on this concept to control the spread of this highly infectious disease. For example, the government of the United Kingdom instructed individuals to stay home as much as possible, to limit contact with those from other households, and to maintain distance from others when stepping out of home (2 meters apart where possible)[1]. Other

[1] https://www.gov.uk/coronavirus

measures include school closures, working from home, cancellation of mass gatherings, and travel restrictions. These policies are referred to as “social distancing” or “physical distancing” policies.

D.3.2 Stay-at-Home measures

Deriving behavioral indicators of social distancing from smartphone data was our primary consideration. There are some existing studies on the mobility responses to COVID-19; for instance, a previous study (Xu et al., 2020) analyzed public geolocated Twitter data to measure the travel behaviors of users. Allcott et al. (2020) combined surveys and GPS foot traffic patterns to observe partisan differences in social distancing. They reported a substantial reduction in the mobility of people in the United States, albeit with partisan gaps in beliefs and behavior. Similarly, we can expect that our participants should spend almost all their time at home and to limit the time and number of places when stepping out, which is usually only for essential shopping owing to the implementation of social distancing measures. These behavioral changes can be acquired from raw GPS data. Since participants’ smartphones record latitude and longitude attributes continuously, their distance from home can always be calculated. Thus, we can determine the time and frequency of their trips outside of home.

Furthermore, social distancing measures can bring about adverse effects, especially on mental health. Some of these reported effects include stress, anxiety and depression, and panic[2]. To maintain mental well-being and while at home, people may find alternative methods of communication to replace their regular face-to-face interactions. Phone calls, messages, video chatting, and social media are possible substitutions people may choose; accordingly, a potential increase in the use of these communication methods is expected. With the various data sources, we could draw a comprehensive and personalized picture of how people react to the impact of COVID-19 restrictions.

D.3.3 Social distancing measures

Social distancing implies that people should meet fewer people than they would during normal times. Bluetooth signals are an effective reference for face-to-face interaction

[2] https://www.cdc.gov/coronavirus/2019-ncov/daily-life-coping/managing-stress-anxiety.html

recorded on smartphones. Nowadays, almost everyone carries a smartphone, and almost every smartphone is equipped with Bluetooth technology, which scans surrounding signals and reports its identity continuously in a short range. Thus, every newly captured Bluetooth entry could potentially represent a new person in close proximity (Liu et al., 2013). This technology has been widely used in the field to estimate face-to-face proximity (Liu and Striegel, 2011). Although it is not fully accurate because of the physical position of the smartphone and surrounding environments, it can still provide a trend that people have less face-to-face interactions. Hence, owing to the social distancing policy, a reduction in the number of unique Bluetooth signals is expected. Theoretically, this would indicate whether our participants adhere to the rules of staying at home and avoiding others visiting their household.

Moreover, social distancing has also affected people when they go for essential shopping. Many grocery stores have a limited number of people in their branches and have introduced directional floor markings to help shoppers maintain a 2-meter distance from one another[3]. This policy could reduce the capacity of crowded grocery stores, and fewer people are expected to be in close proximity to our participants compared to the time before social distancing measures were implemented. Thus, from Bluetooth signals, we could expect a reduction in the number of unique devices from a single scan.

D.3.4 Experiments

We report results from 2 longitudinal studies conducted to gather smartphones' digital phenotypes. Both studies were underway prior to, and continued through, large-scale transmission of COVID-19 and associated social distancing behaviors.

D.3.5 Participants

The studies were reviewed and approved by the Department of Computer Science Ethics Committee at the University. A total of 16 participants were recruited (4 males and 4 females per experiment) through the university database and websites. The 2 experiments recruited individuals from different populations in the United Kingdom: (1) students and (2) patients with a diagnosis of Parkinson disease (aged 63-75 years).

The 2 studies used smartphones to capture data on the participants' activities. Both experiments rely on the same sensing platform.

[3]https://www.tesco.com/help/covid-19/

Table D.1: Sources of the collected digital phenotypes with data description.

Source	***Data***
GPS	Location coordinates (longitude and latitude).
Weather	Temperature, humidity, pressure, wind speed, cloudiness, amount of rain and snow, times of sunrise and sunset.
Applications	App usage data.
Notifications	All notifications generated by any app installed on the phone .
Screen	Screen interactions, visited websites.
Wifi	Access points.
Bluetooth	Nearby devices.
Battery	Charging status, charging start time, charging end time, discharging start time, discharging end time.
Calls	Calls types (outgoing, incoming and missed calls) and times. Numbers are stored in an encrypted format.
Keyboard	Time and the typed letters. (numbers and emails are replaced with asterisks).
Consent form	Name and signature.
Questionnaire	We ask about interests that we infer from the data.
Activity recognition	Tilting, running, on vehicle, walking, on bicycle, on foot.

D.3.6 Instrument

In this study, we used smartphones as independent sensing tools to retrieve participants' behavioral data. The AWARE sensing platform (Ferreira et al., 2015) and developed plug-ins were deployed on participants' smartphone as a monitor app. Under the approval of the ethics committee, different kinds of data, including calls, messages, social media app usage, smartphone usage, notifications, locations, Bluetooth signals, and Wi-Fi signals were collected passively. The content of sensitive communications, such as calls, messages, and conversations, was not recorded. All these data were processed to maintain the anonymity and confidentiality of all participants. All data sources are summarized in Table D.1.

Participants were asked to attend an introductory interview to obtain information on our study and to clarify any of their doubts. On obtaining formal approval from the participants, the AWARE app was installed on their smartphones. Participants were asked to keep the installed app running and use their phones as they normally do. An offline analysis was conducted on data synced with the backend AWARE server.

D.4 Results

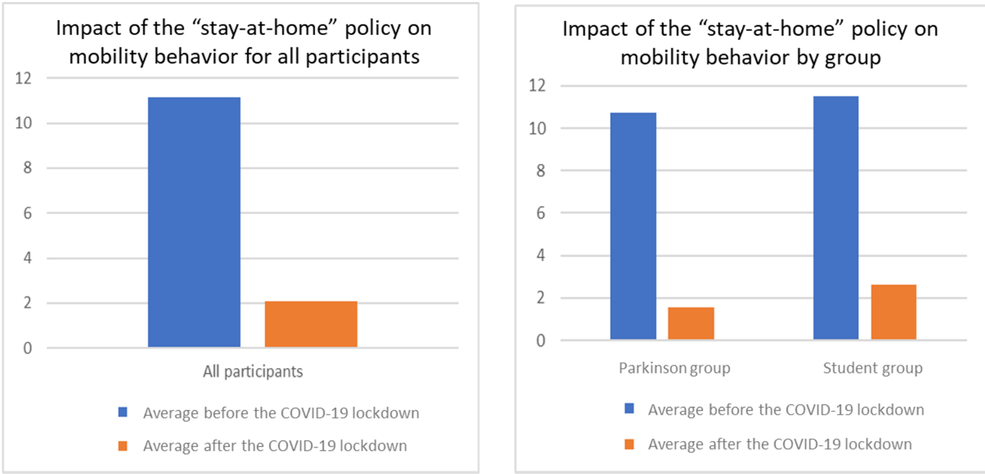
D.4.1 Results overview

This section discusses the results obtained from the responses to the stay-at-home and social distancing policies. To show how digital phenotyping can help understand behavioral responses to these policies, we selected a prototypical participant who exemplified the general behavioral responses exhibited by all participants, in each subsection, except for Figure D.1, which represents all participants. Our experiments started at different times; therefore, the lockdown timelines for each participant may differ. The behavioral responses to COVID-19 were captured despite the differences in the lockdown week. It was intended per our experimental design to have participants adhere to these policies at different times because participants were individually assessed, and no extrapolation among other participants was intended.

D.4.2 Stay-at-Home measures

Mobility patterns for participants in both experiments significantly decreased as a result of the compliance with the stay-at-home policy ($P < .001$) (Figure D.1A). Before the lockdown, the average number of places visited was slightly lesser among patients with Parkinson disease than among the students (Figure D.1B). However, a patient with Parkinson disease and a student may exhibit similar responses to the stay-at-home policy. Thus, individuals of the same group can exhibit a pattern that is different from the average behavior of their corresponding groups. Thus, individual analysis of digital phenotypes would help better understand people's compliance with the suggested policies.

Participants exhibited similar behavioral responses to COVID-19 regulations. We selected a participant who exemplifies the behavioral responses to present the results. We divided the participant's behavior window by week (Monday to Sunday), such that a whole cycle of a weekly social routine could be acquired. The stop point detection algorithms were applied for raw GPS data, such that the place of residence of the participant could be extracted. We used the algorithm proposed by Li et al. (2008) to extract stop points. The algorithm processes data points sequentially, and stop points are defined on the basis of predefined time and distance thresholds. Furthermore, we considered the location where participants spend most of their time of the day as their



(A) Average number of visited places before and after the lockdown for all participants.

(B) Average number of visited places before and after the lockdown for each group.

Figure D.1: Impact of the stay-at-home policy on mobility behaviour.

home. We used Foursquare[4] to determine the names of places, which allows for a better understanding of location semantics. By summing up the calculated results of the algorithm, the length of time participants spend at home and time spent by participants outside of home per week were obtained.

Another indicator is Bluetooth signals. As mentioned before, a scanned unique Bluetooth device could represent a person in close proximity. With everyone staying at home, fewer new identified Bluetooth entries were expected to be recorded. The time spent outside of home was usually below 30 minutes, but identified Bluetooth entries were all above 1000. To easily observe the similar trend of time spent outside of home and the number of new identified Bluetooth entries, we normalized the actual data so they can be plotted on the same graph. As illustrated in Figure D.2, a clear boundary was observed, in that the participant went outside of home fewer times and presented decreased unique Bluetooth entries. Although fluctuations continue, the edge appeared around week 9; that is, March 15-22. This was the week before a lockdown was officially declared in the United Kingdom. Thus, it was observed that this participant perceived the stay-at-home policy and obeyed it objectively.

Figure D.3 shows the impact of the "stay-at-home" policy on participant mobility. The figure represents the mobility behavior of participants who reside in the United Kingdom. Starting from week 12, the number of visited locations drastically decreased from an average of 7 locations to 2 locations. The 2 locations are the participant's home

[4] https://foursquare.com

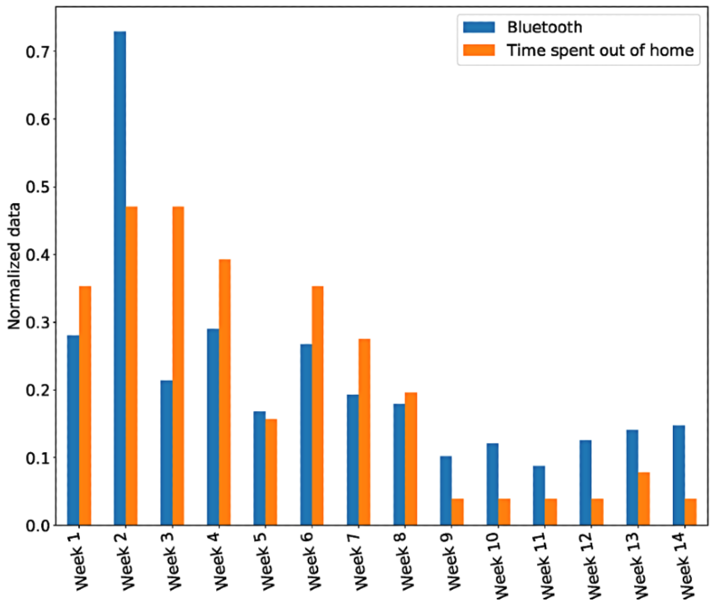


Figure D.2: Normalized unique Bluetooth signals and time spent out of home of a participant before and after the lock-down.

and a grocery store. To motivate this participant to comply with the stay-at-home policy, options for the delivery of grocery items or shopping times can be communicated.

D.4.3 Social distancing measures

As described before, in accordance with the social distancing policy, people have to stay further away from each other than they would during normal times. Because of the capability of the Bluetooth technology, fewer scanned entries would be expected at a time. In this example, we also separated the data into natural weeks and combined all Bluetooth records within that week. Then, we divided this number by the total times for the scans to calculate the average 1-time Bluetooth discovery. As shown in Figure D.4, the average 1-time Bluetooth entries decreased around week 9, which is the first week of the official lockdown in the United Kingdom. This potentially indicates that the participant maintained social distance with others and met fewer people during the lockdown.

The results of our experiment show that the participants complied with COVID-19 policies. Participants managed to stay at home and adapt to the requested changes. However, to stay connected, the participant data show corresponding changes in app usage. The usage of social media apps, phone calls, and video conferences increased for most participants compared to the period before the lockdown. Figure D.5 shows

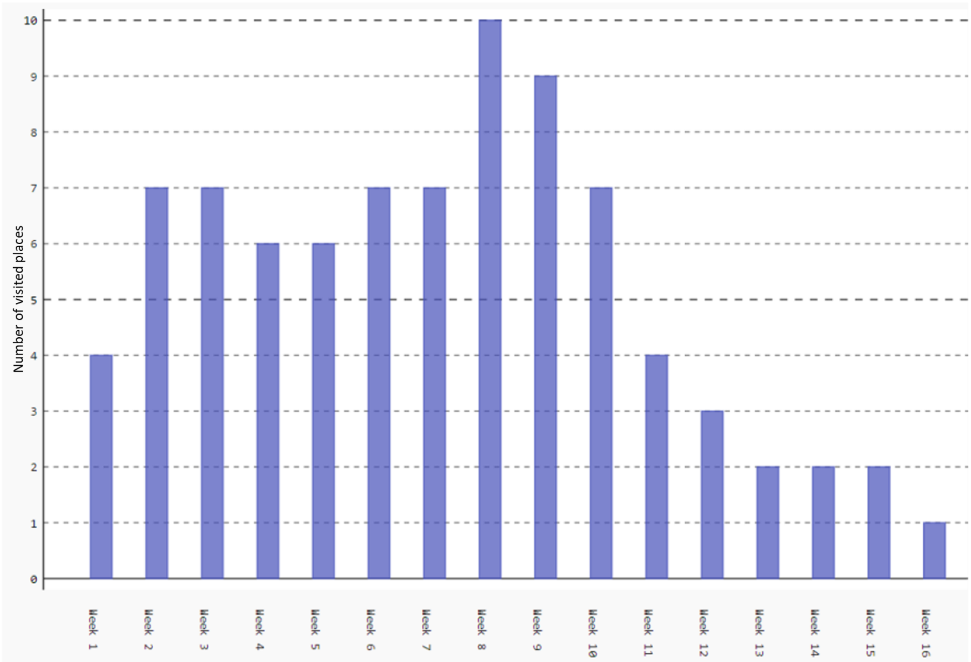


Figure D.3: Location visited by a participant before and after the lockdown.

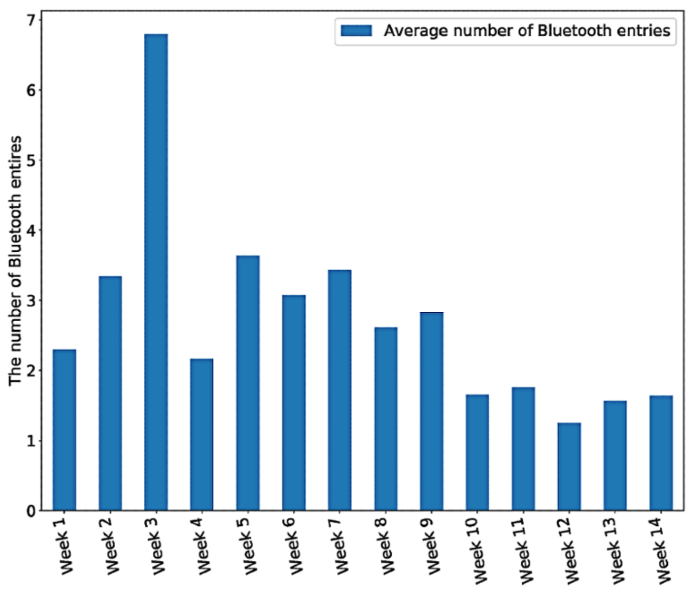


Figure D.4: Average 1-time Bluetooth entries before and after the lockdown.

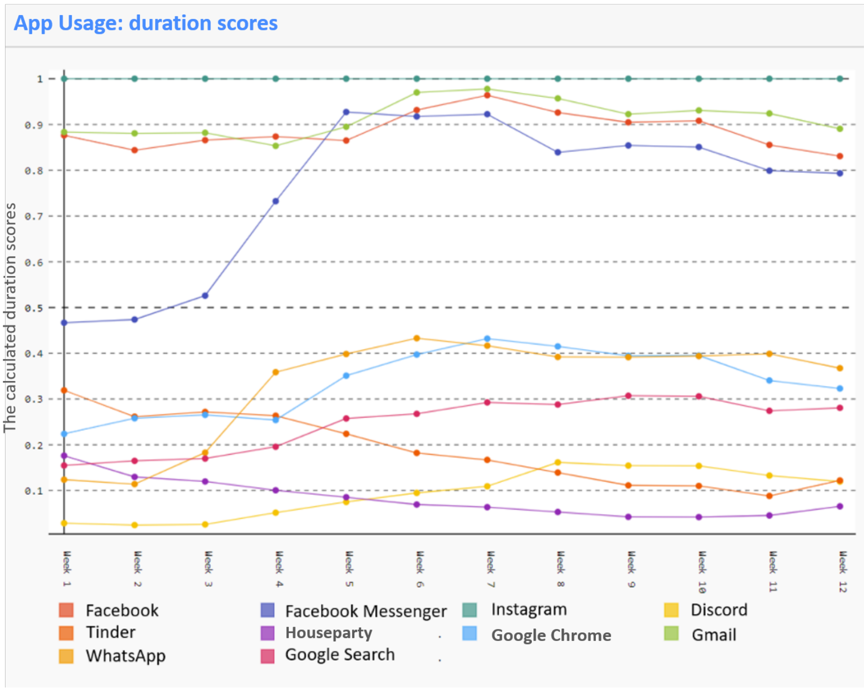


Figure D.5: Normalized duration scores for a participant before and during the pandemic. The participant was enrolled in the first week of March and the lockdown started after the third week of data collection.

the app usage of a participant before and during the pandemic. Instagram was used the longest at 19.50 hours of usage, whereas the time spent on the Houseparty app was 9.27 hours. Values were normalized to easily observe the trend and be consistent with observations from other sources. The lockdown started during week 3. Consequently, the usage of apps, such as Facebook Messenger, WhatsApp, and Discord, has increased.

In contrast, 2 participants presented a decline in phone usage during the lockdown. When interviewed, the participants indicated that they started to use their personal computers and smart televisions more to accomplish the same tasks they previously did with smartphones.

D.5 Discussion

D.5.1 Principal Findings

The reported results show that actionable information can be derived from digital phenotyping. The information derived from understanding participants' compliance, as well as the behavioral impact, can be used in personalized behavioral interventions. Behavioral nudges are used as an effective approach to promote behavioral changes. The NHS in the United Kingdom employs behavioral principles, such as reducing the cognitive load, to communicate nudges. We use actual text messages delivered by the NHS during the pandemic to demonstrate the potential benefit of personalization based on digital phenotyping. We show how a personalized understanding can be leveraged for more traction nudges and just-in-time intervention. The Behavioural Insights (BI) team[5] and the NHS have collaborated to nudge approximately 2 million people through text messages. The recipients of these nudges include people at the highest risk of developing critical complications should they contract the disease. The BI team employ the following behavioral principles to produce the content of a nudge (i.e. the delivered text message).

1. Selection of the appropriate communication channel: since smartphone apps introduce multiple communication channels (e.g. SMS, WhatsApp, and Messenger), personal preferences vary. The NHS and BI team have selected SMS as their preferred method on the basis of a study that shows that 85% of 600 participants do not mind receiving text messages on their personal devices from the NHS (Burd and Coleman, 2020).

2. Signifying the key points: owing to the limitation of text messages, the NHS and BI team have to summarize extended guidelines into short messages. Accordingly, they designed messages such that the key ideas are prioritized.

3. Minimization of confusion and the cognitive load: the key ideas should be delivered in a language that is understandable by laypeople. Additionally, the messages should be clear to avoid confusion and misunderstanding that may quickly spread and negatively impact people.

4. Drawing on scientific behavioral findings: insights derived from behavioral and psychological studies are used to design nudges. For instance, it has been suggested that providing the rationale can help manage people's mental health when quarantined. Accordingly, the NHS and BI team comply with that when designing nudges.

[5] https://www.bi.team/blogs/using-behavioural-insights-to-create-a-covid-19-text-service-for-the-nhs/

These behavioral principles are population-based, which has been reflected on the content of the nudge. M1, M2, and M3 (Table D.2) are examples of 3 nudges that are delivered in accordance with these principles. We hypothesize that digital phenotyping can better improve the content and delivery of these nudges through personalization. For instance, the predicate of M1 can be tailored in accordance with the participant's status as follows. We can predict whether or not a person lives alone from the digital phenotypes. Accordingly, 2 versions of the message can be prepared to deliver a personalized nudge. Versions can be tailored on the basis of the predicted status, age, or other demographics predictable through digital phenotyping.

Digital phenotypes can also improve M2. For instance, an individual used to go to the cinema on Saturdays. Instead of delivering a general nudge about adhering to the typical routine, we can nudge the participant to watch a movie every Saturday during the pandemic. Thus, the typical routine can be embraced, and the delivery of the nudge can be contextualized (i.e. just-in-time intervention). Adhering to typical routines can improve the mental health of individuals and reduce the negative impact of COVID-19 policies.

The information derived from digital phenotyping can also be used to prevent over-messaging. M1 encourages participants to chat with others to stay connected. If the derived data show that a participant regularly chats with others, there is no need to send M1. We speculate that crafting messages on the basis of both data and behavioral principles as well as introducing fewer messages is expected to provide better results. However, actual field testing is required to scientifically measure the real effect of doing so.

Although our approach demonstrates a potential way of producing personalized nudges, it can be reflected in existing behavioral change frameworks such as the behavioral change wheel (Michie et al., 2011). For instance, the framework of the behavioral change wheel identifies 3 main stages to the behavioral change: (1) understanding the behavior to be changed, (2) deciding on the intervention function, and (3) selecting the mode of delivery. We profile and understand the individuals' behaviors through digital phenotyping. Incentivization and persuasion are intervention functions that shape nudging (Liu and Striegel, 2011). Communication as a delivery mode is then used to deliver text messages that nudge people to exhibit the desired behavior.

We are aware of the privacy concerns that may hinder the measurements and implementation of personalized nudges. However, apps can be designed in a way that allows people to partially share information in accordance with their needs. For instance, an

Table D.2: The NHS text messages used for nudging and our proposed personalisation.

Code	*Goal*	*NHS text*	*Personalisation suggestions*
M1	Nudge to establish social responsibility and keep connected	"If you live alone, text a friend or a family member to let them know you are following advice to stay at home until it is safer to mix with others. Plan to chat to someone over the phone at least once a day."	If a participant chats regularly or lives with others, do not send it and prevent over-messaging.
M2	Nudge to keep normal routine and ease anxiety.	"Try to stick as closely as you can to your typical daily routine."	If a participant is a moviegoer, "Watch a movie and try to stick as closely as you can to your typical daily routine."
M3	Nudge to preserve mental health	"Are there things you enjoy doing at home that you usually don't have time for?"	If a participant reports home activities, do not send it and prevent over-messaging.

individual may choose to share the location data only if diagnosed with COVID-19, to trace and limit the spread of the disease to others. Another individual may choose to share his/her data to receive personalized nudges that help him/her adhere to the daily routine (M2). Nevertheless, in these cases and others, personal behaviors are privately phenotyped, and it is up to the person whether or not to share the collected data. Alternatively, messages can be packaged with the app and delivered to participants on the basis of the outcome of a decision tree.

Stay-at-home, social distancing, and other policies are primarily behavioral measures aimed at changing individuals' behaviors to ensure that the risk of contracting the disease is reduced. From this standpoint, behavioral change frameworks (e.g. nudging and the behavioral change wheel) can be relied upon to support the implementation of these behavioral policies. The use of digital phenotyping in activating these frameworks provides an opportunity to personalize the delivery of these policies on the basis of each individual's data. Individuals, institutions, and governments can benefit from such personalization in containing the spread of the virus. Governments may choose to develop apps that have behavioral policies implemented as built-in messages. The delivery of these messages is designed to adapt in accordance with the exhibited behaviors. Individuals who stayed at home (according to digital phenotyping) will not receive messages encouraging them to do so. This decision and others related to message delivery are made locally, on the individual's phone, without compromising his/her privacy. However, individuals who test positive can help governments reduce the potential impacts on others by voluntarily sharing their latest mobility behaviors.

Besides generating personalized nudges, digital phenotyping shows its capability to observe people's behavior on an individual level. In the contest of the COVID-19 pandemic, digital phenotyping has great potential for various implementations. Some of the COVID-19 tracking apps such as TraceTogether in Singapore and COVIDSafe in Australia have used Bluetooth technology embedded in smartphones as their primary contact tracing tool (Culnane et al., 2020). People are encouraged to install these apps so they can know if they have been in close contact with individuals who have tested positive for COVID-19. Institutions such as universities can implement digital phenotyping as innovative methods to study the traditional physiological or societal questions, since no face-to-face settlement is needed. Care facilities could also have digital phenotyping apps installed on their clients' smartphones, such that their issues can be noted without face-to-face reporting. Moreover, the large amount of personal and longitudinal digital phenotyping data could provide policymakers with a deeper understanding of the impact of COVID-19 on a sample of the population. This would shed light on how people actually react to these policies, rather than only determining the infection rate.

D.6 Conclusions

This study shows how digital phenotyping can be of value in understanding people's behavior during a pandemic. Behavioral features extracted from digital phenotypes represent the cornerstone that facilitates the personalization of and compliance with behavioral policies. We presented examples of using Bluetooth, GPS, and app usage data to analyze behavioral responses to COVID-19 policies. Additional sources can be further investigated, such as accelerometers and their role in understanding if people pause more to maintain safe distance.

To encourage the large-scale adaptation of digital phenotyping, governments can emphasize the potential benefits of public health and of maintaining mental health. To preserve privacy, an individual's data are stored locally, and he/she can make the ultimate decision on what to share and to whom the access is granted.

A rule-based messaging implementation can be used to deliver nudges on the basis of the analysis of digital phenotyping. In future studies, we intend to examine the impact of these suggested messages on a sample of the population to measure the impact of preventing overmessaging. Conducting a real-world experiment would also enable us to assess whether having more tailored messages would yield the expected benefits.

D.7 Abbreviation

BI: Behavioural Insights GPS: global positioning system NHS: National Health Service