Periodicity Detection and its Application in Lifelog Data

Feiyan Hu, B.Sc. (hons)

A dissertation submitted in fulfilment of the requirements for the award of

Doctor of Philosophy (Ph.D.)

to the

Dublin City University
School of Computing

Supervisor:

Prof. Alan F. Smeaton

Dr. Eamonn Newman

September 16, 2016

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed:

ID No: 12211604

Date:

Contents

Acknowledgements

Firstly I'd like to thank my supervisor, Prof. Alan Smeaton for all his guidance and support during my time producing this work, offering valuable advice at numerous crossroads. Thanks also to the European Community's Seventh Framework Programme under grant agreement 288199 (Dem@Care) and to Science Foundation Ireland under grant number SFI/12/RC/2289 for funding to support my research. Also, thanks to the ASU-DCU Catalyst Fund for supporting the collaboration I had with Matt Buman at Arizona State University.

Thanks to all the current and past members of Insight who offered advice and support. I would particularly like to thank Eamonn Newman, whose advice, feedback and collaboration were invaluable to me in my work. Many thanks to Haolin Wei, Tengqi Ye, Owen Corrigan, Philip Scanlon, David Azcona, Eva Mohedano, Suzanne Little, Kevin McGuinness, David Monaghan Rami Albatal and Na Li who made PhD work an enjoyable experience.

I would also like give special thanks to Meggan King and Ihar Afanasenka for offering help and respite, and being understanding when I was occupied.

My parents deserve all the thanks I can give them for the years of providing the good home and environment that led me to where I am today. Thanks also to my cousin Junwei Wang who has broaden my horizon of science in different areas.

Abbreviations

AR Autoregressive. 74–76, 82, 101, 102, 123

ARMA Autoregressive Moving-average. xii, 90, 92, 94, 95, 101, 102, 106, 107, 109, 123

ASU Arizona State University. ii–iv, xiii, 33, 35, 40, 41, 49, 74, 80–82, 123–126, 145, 149, 154, 172, 185, 186

CC Correlation Coefficient. xi, xiv, xvi, 8, 9, 79, 80, 84–88, 94, 135, 136, 141, 143–145, 184

CCC Concordance Correlation Coefficient. xi, 8, 79, 80, 83–88, 94

CNN Convolutional Neural Networks. xi, 3, 23, 58–60, 67, 68, 183

CUDA Compute Unified Device Architecture. 64

DFA Detrended Fluctuation Analysis. 25

DFT Discrete Fourier transform. 99, 100, 134

DI-HMM Duration and Interval Hidden Markov Model. 26

DT Decision tree. 63, 68

DTFT Discrete-time Fourier transform. 136, 142

FFT Fast Fourier transform. 24, 99

GDP Gross Domestic Product. 29

GNB Gaussian Naïve Bayes. 68

HMMs Hidden Markov Models. 26

LS Lomb-Scargle. 103, 120

LSTM Long Short Term Memory. xi, xii, 9, 57, 60, 61, 63, 64, 67, 68, 78, 82, 83, 86–89, 92–96, 183

LTI Linear time-invariant. 102

MA Moving-average. 74–76, 101, 102, 123

MLP Multilayer Perceptron. xi, 66, 68

MODWT Maximum Overlap Discrete Wavelet Transform. 25

MPEG Moving Picture Experts Group. 20, 23

NN Neural Networks. 63

RF Random forest. 63, 68

RMSE Root Mean Square Error. xvi, 8, 79, 94, 95, 136, 141, 144

RNN Recurrent Neural Network. 57, 59, 60

RQ Research Question. iv, v, 5, 6, 8–10, 32, 33, 36, 38, 148, 180, 181

SVM Support Vector Machine. xi, 9, 18, 55, 56, 62–64, 68, 77, 82–92, 94, 95, 183

VAR Vector Autoregressive. xii, 76, 90, 92–94

List of Figures

List of Tables

Abstract

Periodicity Detection and its Application in Lifelog Data
Feiyan Hu

Wearable sensors are catching our attention not only in industry but also in the market. We can now acquire sensor data from different types of health tracking devices like smart watches, smart bands, lifelog cameras and most smart phones are capable of tracking and logging information using built-in sensors. As data is generated and collected from various sources constantly, researchers have focused on interpreting and understanding the semantics of this longitudinal multi-modal data. One challenge is the fusion of multi-modal data and achieving good performance on tasks such activity recognition, event detection and event segmentation. The classical approach to process the data generated by wearable sensors has three main parts: 1) Event segmentation 2) Event recognition 3) Event retrieval. Many papers have been published in each of the three fields.

This thesis has focused on the longitudinal aspect of the data from wearable sensors, instead of concentrating on the data over a short period of time. The following aspects are several key research questions in the thesis. Does longitudinal sensor data have unique features than can distinguish the subject generating the data from other subjects ? In other words, from the longitudinal perspective, does the data from different subjects share more common structure/similarity/identical patterns so that it is difficult to identify a subject using the data. If this is the case, what are those common patterns ? If we are able to eliminate those similarities among all the data, does the data show more specific features that we can use to model the data series and predict the future values ? If there are repeating patterns in longitudinal data, we can use different methods to compute the periodicity of the recurring patterns and furthermore to identify and extract those patterns. Following that we could be able to compare local data over a short time period with more global patterns in order to show the regularity of the local data. Some case studies are included in the thesis to show the value of longitudinal lifelog data related to a correlation of health conditions and training performance.

Chapter 1

Introduction

> Not until we are lost do we begin to understand ourselves.
>
> *Henry David Thoreau*

The three basic philosophical questions: Who am I? Why am I here? and Where am I going? have been haunting us since the very beginning of our civilization. Scientists were, are and will be trying to understand ourselves from different perspectives such as "things that can be materialized" or "thoughts that cannot be". We have developed mathematics, philosophy, biology, physics, psychology and other enormous research areas to help us better understand ourselves. The appearance of computer science seemed act the same way, except computers do everything faster. There were times that human calculators were used iteratively to solve functions that cannot be solved analytically which is so hard to think about it now with computers doing the same tasks. Using computers to build and validate models or simulate proposed models faster than ever before has accelerated developments in nearly every discipline and brought different disciplines together to create some things that were not here before.

Devices or sensors with computational power can be made smaller, faster, more energy efficient and more affordable. This is the wave we are experiencing now.

The trend of small yet powerful devices enables us to wear sensors comfortably and continuously while data can be captured with adequate accuracy. The use of multi-model sensory technology has brought challenges and opportunities in personal sensing such as management and analysis of personal big data. Researchers are now addressing issues in indexing and retrieval of personal sensing data. Some of the fundamental problems in this have rarely been touched before, which is what this thesis is trying to address.

1.1 Personal Sensing Models

How can we build a model to simulate human beings in the sense that it will react in the similar way that a human being would do ? This is indeed a very interesting and complicated questions in so many ways. At least using the technology we have so far it is an unachievable task. Maybe we are able to train computer algorithms to recognize faces, cars and many other objects or learn how to generate text according to learned models. The performance for this may be even better than some human beings, but we have some profound beauties such as creativity, innovation and there are so many parameters from our uncertain environment and from ourselves that might influence us, that a single trained model for a complex task cannot solve for now.

In order to confront less complicated questions, personal sensing is treating human beings as a black box. By using different sensors that take measurements that can be logged, we can analyse the data we capture and process it to form some sort of knowledge about the data, and this knowledge will be eventually fed back to the person who generated the data in many different possible ways.

The rising research interest into personal sensing has brought challenges in this research area. Those challenges inclue but not limited to data privacy, sourcing and stroring personal sensing data, organizing sensing data. For data privacy, research

by Klasnja *et al.* (2009) shows privacy concerns about such personal sensing. In the study 24 participants who took part in a three month study that used personal sensing to detect their physical activities are interviewed. Those participants show concerns on what is recorded. They also concern about the context in which participants worked and lived and thus would be sensed. For sourcing and storing personal sensing data, the challenges come from continious data capture would consume large amount of energy thus requiring carrying battery that could support such function (Ra *et al.*, 2012). On the other hand those wearable devices need to be made as small as possible. While data science now are heavily data driven, capture high quality data is vitally important for a good research.

In personal sensing, we would like to capture clean data with good enough resolution and low noise. If those data with decent quality need to be stored on the device, large storage is need (Gemmell *et al.*, 2004). Personal sensing generates continuous streams of data on a per-person basis, however despite the potential for real-time interactions, most of the applications for lifelogging we have seen to date do not yet operate in a real-time mode. So while lifelogging does not yet have huge volume, this volume of data is constantly increasing as more and more people lifelog. For a single individual, the data volumes can be large when considered as a Personal Information Management challenge, but in terms of big-data analysis, the data volumes for a single individual are small. Considering a lifelog of many people, thousands, perhaps millions, all centrally stored by a service provider, then the data analytics over such huge archives becomes a real big-data challenge in terms of volume of data.

There are a number of challenges that immediately arise when organising personal sensing data. A key challenge for processing of lifelog data is in extracting meaningful information from the content to make it useful. Although deep learning such as CNN has imporved performance of classification of images greatly, due to lack of annotation and incompletness of data, it is still hard to retrieve useful

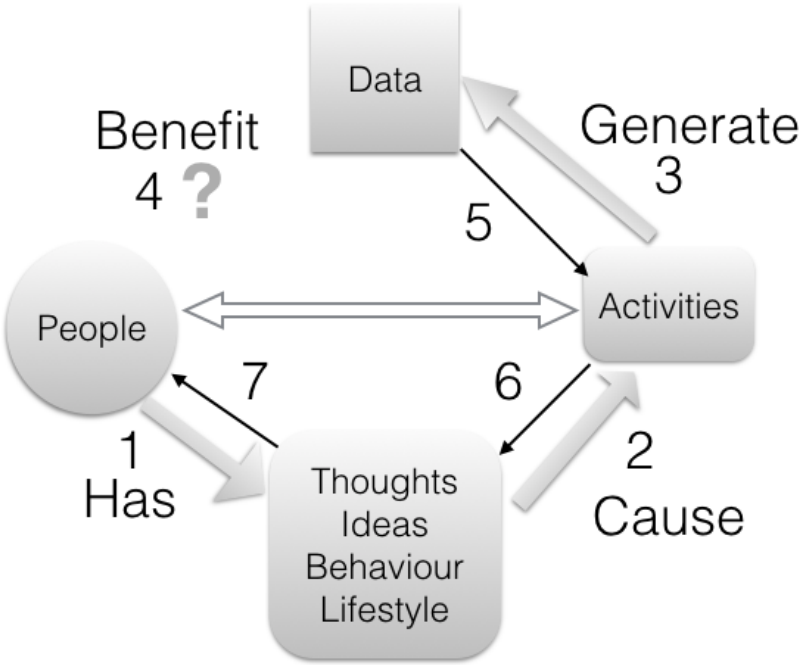


Figure 1.1: Personal Sensing Model

information.

Figure 1.1 shows the general data processing model for personal sensing. If we have a look at the outer circle of data flow from 'People' to 'Data' via steps 1, 2, 3. Generally speaking, a subject has to have some ideas in mind and impulses make those ideas into activities and then those activities would be captured by different wearable or ambient sensors. Let us then have a look at the inner circle from 'Data' to 'People' via steps 5, 6, 7. Following the same principle, the data being collected can be processed into activities and then activities can then be processed into higher semantic level events with more abstract features such as lifestyle and then can we identify who generated the data. This chain is rather complicated, because there are so many uncertainties might affect the final results of chain. If we start from people and end at data, going anti-clockwise, we have to investigate human lifestyle and activities. Those two can actually be treated as separate researches that we are not familiar with. In order avoid going deep into those studies, we are going a different direction. We are trying to bridge the gap between people and data directly to see

what benefit data can bring to people. We will explore this by omitting manually investigating each intermediate nodes between data and subject who generated the data. Instead we use computer hardware and algorithms to model those intermediate modes to bridge the gap directly.

1.2 Hypothesis and Research Questions

The overall hypothesis in this thesis is that

> "Human lifestyles have in-built periodicities of various frequencies including daily, weekly, monthly, seasonal, and annual and that using lifelogging sensors to collect data, even though it may have gaps in coverage, we can use time series analysis to detect these periodicities and use them to good effect."

To explore this hypothesis, 5 research questions (RQs) have been formulated and we present these in sequence, and the sets of experiments to be carried out associated with these RQs.

1.2.1 RQ1: Relationship between people and data

Does each individual generate a unique distribution of their data ? Does the data generated by a subject have unique features that can represent them ? This research question is trying to find out if there is a bijective corresponding between 'People' and 'Data' by omitting intermediate factors that might affect the result. In order to examine this research question, we will:

1. Show that the data generated by each subject is different.

2. Show that the data can be projected back to the subject.

In order to study RQ1, we will be using low level data, which is generated by more than 20 subjects. The point of RQ1 is to prove different subjects generate different data and the uniqueness of the data can somehow be detected using computer algorithms and mapping it back to the subjects who generate the data, in other words to identify the subjects. If the bijective correspondence were to prove to exist to some extent, we will be able to confirm that sensors are able to capture some unique features of the subject and therefore any knowledge mined from those data does connect with the subject. This fundamental question is seldom raised, which can sometimes make the validity of personal sensing questionable.

1.2.2 RQ2: Predict lifelog data

Is human sensing data predictable? This research question is trying to build models for personal sensing data. We will see how different models such as different machine learning models, time series or deep learning models do affect the predicability. Assume there is a one-to-one correspondence between people who generate the data, this research question is trying to address the issues that the capability of uniqueness of the personal sensing data to be modelled. If we are able to model the data, what would be the prediction to be? Experiments will be conducted on both low-level data and high-level data.

1.2.3 RQ3: Periodicity in lifelog data

Are there any repeating structures in lifelog data? How can we detect periodicity in personal sensing data? In this research question we are interested in detecting periodicity in personal sensing data in order to reveal hidden patterns that cannot been seen directly. Several methods to detect and examine periodicity will be tested such as parametric and non-parametric methods. Both low-level and high level datasets will be used since they are all longitudinal data.

1.2.4 RQ4: Intensity of Periodicity

How can we measure intensity of periodicity ? In this research question we will compute the strength of periodicity for various datasets. Note that periodicity is assumed to be reported positive in the previous research question. Since the periodicity detected can only represent an overall global repetitive structure, there are many variations locally compared with the global pattern. In the places we think the bigger the variation is, the smaller the strength of local periodicity is. In this research question we will define several metrics to compute the strength of periodicity in lifelog data.

1.2.5 RQ5: Applications in real life

Do the detected patterns of periodicity make sense for users in the real world ? In this research question, we will be using two case studies to demonstrate how the patterns of periodicity and/or intensity of those patterns could correlate with subjects' interpretations of their own real life. In the low-level data case, the correlation between intensity of circadian periodicity and bio-markers will be investigated. In the case of high-level data, we will conduct a focus group interview with the subject and present the results of periodicity analysis by quoting the response of the subject. Participants will review their lifelog data in order to stir up past memories for example in the form of an interview and detected patterns will be fed back to participants as triggers for deciding if these periodicities are right or are irrelevant. This research question is closely related to the qualitative evaluation of periodicity detection. How can we give feedback to users after discovering periodicities so that we can conduct an assessment of the performance of detection according to users' reaction and/or their biomarkers ?

1.3 Methodologies

The thesis sets out to study personal sensing data and it's latent patterns that might be used to good effect. We used both quantitative and qualitative analysis. For research questions 1 to 4 we study the data and investigate some algorithms using quantitative methods to evaluate. In research question 5, we implement qualitative evaluation for one case.

Evaluation is a challenging task, especially when we are trying to evaluate the performance of periodicity and intensity of periodicity. The relevance of the periodicity is quite subjective since every individual has his/her understanding of periodicity. That is why both qualitative and quantitative evaluation will be used. In order to use quantitative assessment we need to annotate and record accurate meta data. In RQ1, this is been done by evaluation using precision, recall and F1 measures to evaluate the result of mapping data to subject by using the identity of each subject as an annotation. The similarity matrix is used to examine the projection from subject to data.If we are able to map the data back to people who generate those data, it is feasible to conclude that any analysis result coming from the data are more or less related with the subject. In RQ2, models using different algorithms are evaluated by RMSE, CC and CCC. If the predicted lifelog data is close to ground truth, then we are able to to use trained model to predict lifelog data. But on the other hand, it also proves that there are some features machine can pick up and use it to predict data. In RQ3, the periodograms are firstly evaluated by using synthetic data and then one chosen method is applied to several datasets in order to show the feasibility of periodicity detection. If we are able to use empirically selected sophisticated methods used in other research fields to identify periodicity, at the same time those periodicity are quite significant, we are able to say that there is a occurring structure in lifelog data on which information is modulated on. Further more we can then identify the periodicity. In RQ4, the intensity metrics will be evaluated on

simulated data generated by a model that we propose. The performance of different metric definitions and window sizes of local intensity will be evaluated. Then we will choose one metric to generate intensity of periodicity for high-level data and low-level data collected in real life. If we are able to use measurements to capture periodicity intensity theoretically, we can then apply those measurement on real life data, to measure intensity of periodicity of real life. In RQ5, the results of real life data we generated in RQ3 and RQ4, will be used to find out the connection with real life. High-level data will be analysed using qualitative analysis. The subject will be interviewed and asked to answer some questions which are relevant to the patterns and intensity coming from previous processes. Low-level data will be analyzed by computing the CC between the strength of circadian periodicity and biomarkers. If the response from real life is positive, we can then verify the pipe line of the theory is working. In other words, the framework of lifelog system is valid.

1.4 Organization of the thesis

This thesis will begin by outlining the currently existing work that relates to our research in Chapter 2. In Chapter 3 we look at the data we will be working on by examining the data distribution and entropy of the data and a similarity matrix and confusion matrix will be computed to explore RQ1. Several machine learning and deep learning algorithms will be used for the classification problem. In Chapter 3, we will model both high-level and low-level lifelog data. Several machine learning methods such as SVM, LSTM or time series models such as autoregression types of model, will be used. This regression problem will provide some reference that we need to tackle RQ2. Chapter 4 then looks at different methods to detect periodicity. Parametric methods and non-parametric methods will be applied on synthetic data. One method will be selected to show the periodicity in real life data across different datasets (RQ3). Next, Chapter 5 will look at the intensity of detected periodicities.

Here we compute the strength of significant periodicity/periodicities. Several metrics will be defined to measure the intensity/strength of periodicity. This chapter is related to RQ4. Chapter 6 will bridge between the discovered periodicity and/or intensity of periodicity, and real life interpretation of that data. We use two cases here to illustrate the practical impacts and validity of processing personal sensing data. This is trying to answer RQ5. Finally we will conclude with a summary and look at how in answering our research questions, we have formed a contribution to the field.

Chapter 2

Background

Lifelogging is the ambient, digital capture of any of several possible data sources which log the ordinary day to day activities of a person performing their typical activities of daily living. There are several possible purposes for such recording of the daily activities which vary from subsequent searching for events to automatically determining patterns of a person's life. Lifelogging is already widely used in medical and therapeutic applications such as a support for reminiscence therapy, a tool used in memory reinforcement and helping to promote healthy lifestyle, as described by, for example, Kerr *et al.* (2013). In lifelog applications which form part of reminiscence therapy, participants are shown items from earlier periods in their life, either very recently which is where lifelogs can be used, or from their distant past, and seeing these items which can include photos, helps to evoke memories and to re-live that past. This, in turn, provides cognitive stimulus as well as having social benefits like encouraging conversation and storytelling.

In Doherty (2009) created a framework which enables effective memory retrieval as a form reminiscence therapy for people with memory impairments such as acquired brain injury or some form of dementia. The framework can be simplified as shown in Figure 2.1, and can support what is in effect an event-based model of lifelogging. In this framework the lifelog consists of a series of images captured from

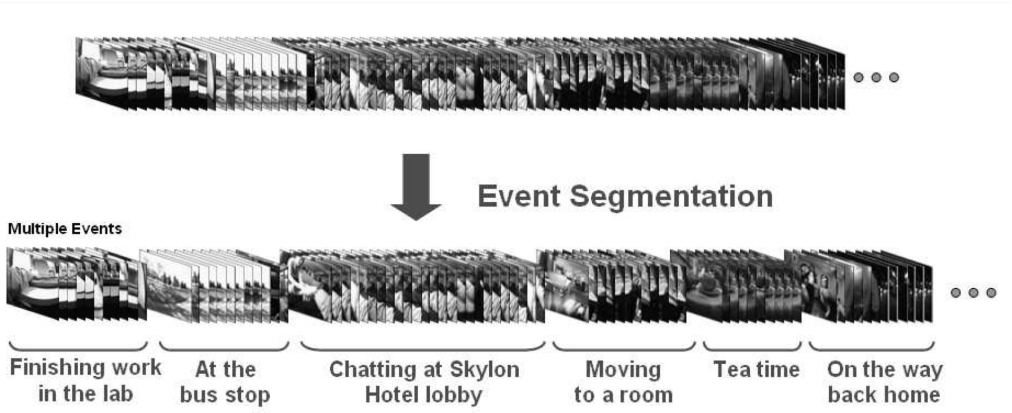


Figure 2.1: Lifelogging Events Detection.

a wearable, front-facing camera and taken automatically on a regular basis, say 2 images per minute. This stream of images is then automatically segmented into discrete events such as *"finishing work in the lab", "waiting at the bus stop"*, and *"chatting with friend at the Skylon hotel lobby"*. From each one of these discrete, non-overlapping events, a single key frame is selected to illustrate the event and the event as a whole can be subject to a process to determine its novelty or uniqueness in terms of the rest of the person's life, as well as to a process of automatic analysis by tagging it with semantic concepts, which are detected automatically.

The SenseCam[1] is, or rather it was, a popular wearable device used to take pictures automatically with a certain minimal frequency in order to create personal, visual lifelogs or archives. The challenges of effective structuring of a lifelog, followed by useful searching and browsing of this image collection for the purpose of locating important or significant events in a person's life has been addressed by Lee *et al.* (2008). A media process was proposed based on the following simple three-step process with a specific focus on visual lifelogging:

1. Capture and upload images;

2. Post-processing of images;

[1] `http://research.microsoft.com/en-us/um/cambridge/projects/sensecam/introduction.htm`

3. Accessing images.

This series of processes maps to almost all uses of lifelogging and the first of these steps, capture and upload, is common whereas the others will vary depending on the lifelogging application. We will now look at the capture and upload process in more detail.

2.1 Lifelog Data Capture

The automatic capture of data for lifelogging has been investigated by many researchers for some time. The first such work is probably that by Mann (1997), who is recognised as a pioneer who tried to capture what he saw through video cameras mounted on his head. Over time, these have evolved from head-mounted cameras to more discreet recorders built into his eyeglasses, but all serve the same purpose, the continuous capture of his life from his viewpoint as a stream of images. Sellen *et al.* (2007) from Microsoft Research in Cambridge has used a more modest device, the SenseCam, to capture everyday life and in their work the authors report evidence that images captured from a SenseCam can improve the memory recall of wearers when they are shown their own SenseCam images at some time later. At MIT, Eagle and Pentland (2006) carried out an experiment using Bluetooth-enabled mobile telephones to capture Bluetooth signals from surrounding devices such as the mobile phones of other people and they used these signals to measure information context in order to identify the deep social patterns that occur in a user's typical activities. In effect the mobile phone was recording who, or rather whose phones, a person would typically be in the same vicinity as the main subject, so the phone's Bluetooth signal was acting as a form of proxy for social interaction.

Vemuri and Bender (2004) presented a memory re-finding use for lifelogging which is called "iRemember". In their research, the authors recorded audio clips as the main information used to navigate a subject's memory and it is their use of

audio which makes this one of the more unusual cases in the history of lifelogging as audio is not normally recorded. Patterson *et al.* (2004) employed a similar kind of technology to provide real-time transport information to individuals with mild cognitive disorders and to improve efficiency of their travel and their safety as well.

Mobile phones and smartpones have become very popular nowadays and these form a useful computing resource because of their in-built sensors like GPS and accelerometers and because they offer continuous network connectivity they can form a ubiquitous infrastructure for capturing our digital life. The DietSense project by Reddy *et al.* (2007) at UCLA makes use of a mobile phone with a camera embedded to capture pictures automatically. The images collected as the log of a wearer's mealtimes are used to capture and then subsequently analyze the subject's diet intake in order to give the subject feedback on their food intake and to improve their diet choices. The WayMarkr project by Bukhin and DelGaudio (2006) also makes use of a mobile phone affixed to a strap to take pictures automatically. Rawassizadeh *et al.* (2013) proposed Ubiqlog using mobile phone as a device for life logging. The project provides extension to other devices, such as ebook readers, T.V.s, etc. Further work by Eagle and Pentland (2006) studied social dynamics by using mobile Bluetooth as the measure in lifelogging.

Data capture technology is a topic under active development. For instance, Google glass[2] is a wearable computer with a head-mounted display that displays information for the wearer in a smartphone-like, hands-free format. The user can interact with the glass via natural language voice commands. For instance, the wearer can ask the glass to take a picture or record a video. Another lifelogging tool that has come to the market recently is a wearable camera called Narrative[3] . Narrative is a small wearable camera about one inch square that clips to a wearer's clothing like a shirt lapel, and takes a photo every 30 seconds and automatically

[2] `https://www.google.com/glass/start/`
[3] `http://getnarrative.com`

uploads these to an online service. The Narrative also has other in-built sensors. When capturing data, the data set can be annotated, which is helpful especially for the use of supervised methods of image analysis as the annotations can provide a ground truth of description of content. Each observation in an annotated data set is drawn from a set of pre-defined annotated concepts.

When annotating a (visual) lifelog it is possible to do the annotation while capturing the data by asking the participant to perform a set of activities, and then to label the captured data based on the performed activity. This method has been used in many studies, where a researcher supervises and observes the participant during the set of activities corresponding to the data collecting process but it is only useful in lab-type settings or other controlled environments where the aim is not to do ambient lifelogging of a subject in ordinary settings.

Another approach to annotating a lifelog data set is to use manual annotation protocols. In this approach a vocabulary for annotation would be created and then the subject could subsequently look over their data, perform the data annotation process using an annotation tool, and annotate the metadata based on the controlled vocabulary. For instance, a study which uses such a manual annotation protocol was performed by Kerr *et al.* (2013) to assess sedentary behaviour and relaying on worn accelerometer data and SenseCam images. The data was collected by 40 users, and images were coded later for sitting and standing posture and 12 activity types. The coded image data was then compared to the accelerometer data. The authors reported that manual coding of the images was time-consuming and coding errors can easily occur.

Although they are successful in solving some design considerations, algorithms for automatically detecting contexts and situations from lifelog data tend to lack flexibility. This means that the systems developed cannot generally adapt to the semantics of different contexts. Whatever context information is captured is not fully used to receive more flexible approaches for context classification and recognition

for the task of labeling the semantic meaning of the users' events in their lifelog.

The size and scope of the research in this thesis shows that there is a very active research community in lifelogging, exploring a range of techniques and using a variety of lifelogging devices. Yet lifelogging research needs to address more than just the data capture technology, it needs to also investigate and create new techniques for the analysis of lifelogs and to provide search, browsing, and navigation through the lifelogs. Thus indexing and retrieval are just as important as the lifelog capture devices.

In order to manage accumulated lifelogs we need good information management tools, and much related work has been done in multimedia retrieval where low-level feature-based multimedia queries using image features such as colour, texture, edges and other attributes have been studied extensively. Examples of this work include McGuinness *et al.* (2014). However, despite the progress in this area since reported by Smeulders *et al.* (2000), there is still no foolproof means to reflect the coincidence between features automatically extracted from visual data and the interpretation that they have for a user in a given situation.

2.2 Time-Independent Lifelogs

When we say *time-independent lifelogs*, we refer to event based lifelogs. Users use SenseCam or other personal wearable sensors to generate a huge amount of data. Annotators are then asked to segment the collection of personal sensing data into different episodes. The annotations are normally 1) the start and end of an event 2) Semantic concepts (objects) contained in each lifelog unit, such as object(s) in each image 3) Semantic concepts corresponding to each segment of the data. The concepts are normally high level, namely activity of each segment of the data. We can then implement various machine learning methods to learn how to segment the data and/or how to map the data into higher semantic space.

2.2.1 Event-based Lifelogging

As researchers started to develop the technology necessary to build a lifelog and then to make it searchable and browsable by automatically constructing an index, one of the first issues faced was to decide on the unit of retrieval. This could be a single image, or something more meaningful. The current approach for lifelog access is to create an index which uses each of the daily activities of the participant as its basic unit for retrieval or browsing. Thus, in this task, research is concentrates on two challenges:

- **Event Segmentation:** this is the identification of each episode of behaviour, i.e., a continuous period of time where the participant is engaged in a single activity. By identifying the points in time where the primary activity changes (i.e., the boundary between events) we can delineate the periods of time where a single event started and stopped.

- **Event Identification:** having determined the time periods where a single activity is underway, these periods need to be classified, or labelled, with the identity of the activity taking place (e.g., walking, sleeping, meal-time, on a bus, etc.).

Ó'Conaire *et al.* (2007) describe a visual diary of lifelog images constructed by clustering images based on low-level image features such as a colour spatiogram and block-based correlation between images. Their experiments also incorporated additional sensor data from accelerometers which can be worn or in the case of SenseCam, are built into the camera which capture images. These features allow the clustering of images into events, allowing the user to search, browse or review their day by event, rather than as a stream of images.

The study by Doherty and Smeaton (2008a) showed a method for detecting event boundaries, based on an adaptation of TextTiling algorithm by Hearst (1997), in which images (or blocks of images) are compared to their neighbours to determine

their dissimilarity. The system identifies boundaries where the dissimilarity has exceeded a threshold. The paper investigates the optimal size for the block of images, the optimal distance metric to measure dissimilarity and the optimal threshold for successfully detecting event boundaries. The authors described a method that can automatically segment an image collection. Features used to compare images selected were MPEG-7 descriptors namely color layout, color structure, scalable color and edge histogram. Similarity scores among current and adjacent images were calculated using those features. The authors used a technique called *peak scoring* to enlarge the dissimilarity and some automatic thresholding methods were applied to determine the boundaries between events. In the final steps of the process, boundaries that are too close to each other are removed. Later, researchers would apply machine learning techniques such as SVM to learn a classifier to identify the boundaries between events in a stream of lifelog images. Also, data from other sources could be used in the segmentation process including data from accelerometers, GPS co-ordinates or other metadata.

Doherty *et al.* (2007) examined event-boundary detection using multimodal data. In that research, the data consists of images, accelerometer data, light sensor values, temperature and recorded audio. They also experiment with various kinds of data fusion of these different data sets. From their results they were able to identify three main types of activity boundaries:

1. a change of activities within the same location,
2. a change of location, and
3. engagement in social interaction.

They show that sensor accuracy is related to the type of activity boundary. For example, recorded audio is significantly better at identifying changes in social interaction, but not activity changes in the same location. Chennuru *et al.* (2010)

developed a lifelogging system using mobile phones to record and index lifelogs using activity language. By converting sensory data such as accelerometer and GPS readings into activity language, they apply statistical natural language processing techniques to index, recognize, segment, cluster, retrieve, and infer high-level semantic meanings of the collected lifelogs. Their lifelog system supports easy retrieval of log segments representing past similar activities and automatic lifelog segmentation for efficient browsing and activity summarization. Aghazadeh *et al.* (2011) developed a novelty detection algorithm based on identifying deviations from the wearer's normal behaviour. Daily activities are logged in the usual fashion. Sequences of similar images can be matched over days and weeks (e.g., the wearer's daily commute to work). Novelty in day-to-day activities can be detected when a sequence of images from the lifelog cannot be matched to a previously recorded sequence. Assuming that the adjoining sequences do match previously recorded sequences, the novel sequence can be seen as a temporary deviation from the participant's normal daily behaviour, and so the event can be emphasized as significant for that participant. Yamauchi and Akiba (2016) proposed method to detect repeated event in newest NTCIR lifelog data.

In much of the research into processing lifelogs, participants have used SenseCam or other personal wearable visual sensors to generate a huge amount of images. To test the accuracy of event segmentation, it is best if users themselves rather than a researcher or a third party, are asked to annotate part of their personal lifelog images, due to ethical issues like privacy. Typically the annotations generated correspond to markers for the start and end of an event and, optionally, a set of semantic concepts corresponding to each lifelog image, or event.

Following the process by which images are segmented into events, a single image is generally selected to represent the whole event, in order to facilitate event-based queries from users. Doherty *et al.* (2008a) investigated several selection methods including, selecting the middle image, selecting the image that is most representative

of all the images in the event, selecting the image that is most representative but also most different to other events. Also, the quality of the image has been considered as an important criterion in selecting the key frame image (Ratsamee *et al.*, 2015) and different image quality measures have been compared and included.

Novelty detection gives us an indication of which events are different and possibly more important in a lifelog as a result of their uniqueness. Doherty and Smeaton (2008b) uses both frequency of face detected and novelty score calculated using just the key frame image from each event. The window size and the sampling rules were defined by the authors to compute similarity between the current event and adjacent events. Face detection was used to determine the factor of human involvement in conversation.

A set of classifiers for pre-defined concepts have been trained on lifelog data and each classifier assigns images a score as to the confidence of that concept being present in the image. Byrne *et al.* (2008) was among the first reported work to determine whether a lifelog image belongs to a set of concept occurrences or not. One of the most important statistics as a result of concept detection is the author-calculated average number of concepts detected for each event and compared among users. The authors of (Byrne *et al.*, 2008) concluded that the average numbers of concepts detected for each event gives an outline of the differing lifestyles of users.

Doherty *et al.* (2008b) described a way that a user can retrieve events by using queries which are based around the 4 search axes of when, where, who, and what. Moreover, similar events could be retrieved automatically by computing and ranking similarity between lifelog events using features like Moving Picture Experts Group (MPEG)-7 colour and texture, combined with data from worn sensors.

2.2.2 Event Identification

Bridging the gaps between different levels of semantic representation of lifelogs is a challenge for researchers in content-based information retrieval in all application domains, not just lifelogging. High-level features refer to features that are semantically meaningful for the end user. While low-level features are never readable by the end user, high-level features can express the semantics of media in a more acceptable way as concepts, such as, for example, "indoor", "outdoor", "vegetation", "computer screen", etc. These features can provide a meaningful link between low-level features of images like colour and texture, and user interpretations. The extraction of such high-level features as these demands filling the gap between low-level feature occurrences and high-level features, which is called the semantic gap in multimedia retrieval.

Semantic concepts are usually automatically detected based on a mathematical model by mapping low-level features to high-level features. The popular pipeline approach as described by Snoek *et al.* (2007) is to apply discriminative machine learning algorithms such as Support Vector Machines (SVMs) to decide the most likely concepts given the extracted features. Hori and Aizawa (2003) develops a context-based video retrieval system for life-log applications. This wearable system is capable of continuously capturing data not only from a wearable camera and a microphone, but also from various kinds of sensors such as a brain-wave analyzer, a GPS receiver, an acceleration sensor, and a gyro sensor to extract the user's contexts. By mapping data to higher semantic the system will be able to browsing or retrieve videos using conversational query. Compared to a discriminative model which is more task-oriented, generative statistical models such as Markov Models try to analyze the joint probability of variables as described by Li and Wang (2003), which are also proposed in concept annotations. Both generative and discriminative approaches have their pros and cons. A generative model is a full probabilistic

model of all variables whereas a discriminative model has limited modeling capability. This is because a discriminative model provides a model only for the target variable(s), conditional on the observed variables hence cannot generally explain the more complex dynamics underlying the generation of data for a given class. However, discriminative models are often easier to learn and perform faster than generative models. Besides, it has been shown that discriminative classifiers often get better classification performance than generative classifiers with large training volume (usually including positive and negative samples).

A limitation for building classifiers is for them to reveal the higher-level semantics of images when they have multiple concepts with high correlation. The concepts involved in lifelogging cover numerous aspects of our daily lives and the choice of concepts is very broad.

Although individuals may have different contexts and personal characteristics, the common understanding of concepts that is already socially constructed and allows people to communicate according to Lakoff (1990) and Huurnink *et al.* (2008), also makes it possible for users to choose suitable concepts relevant to their own daily activities. In everyday concept detection and validation as described by Byrne *et al.* (2010), concepts are suggested by several SenseCam users after they have reviewed several days' worth of their own lifelogged events. The set of concepts used are those that can be detected with an accuracy rate above a particular threshold.

To find a set of candidate concepts related to each activity in a set of everyday activities, Wang (2012) and Wang and Smeaton (2012b) carried out user experiments on concept selection where candidate concepts related to each of the activities above were pooled based on user investigation. Byrne *et al.* (2008) used low-level features of SenseCam images to define high-level semantic concepts such as eating, road, sky, office, etc. 27 semantic concepts have been defined and used as a source for improving the segmentation task. Everyday concept detection consists of supervised learning, visual feature extraction and then finally feature and classifier fusion. The validation

in the work by Byrne *et al.* (2008) was done by 9 participants who manually judged the accuracy of the detection on a subset of 95,507 lifelog images. The results showed an average precision of 57% for positive matches and 93% for negative matches within such a collection. The authors see these results as encouraging, and they suggested that automatic concept detection methods translate well to the domain of visual lifelogs.

Krizhevsky *et al.* (2012) proposed Convolutional Neural Network to train convolutional kernels on large scale dataset. The CNN and some variant of the network such as VGG (Simonyan and Zisserman, 2014) is one of the most widely implemented network of deep neural network in computer vision. One of state-of-art neural network approach in computer vision is proposed by He *et al.* (2015). Before mentioned network all provide open access to pre-trained models in several model zoos.

Doherty (2009) relied on SenseCam sensor readings and low-level features of images to create clusters of distinct activities throughout a day. The MPEG-7 visual descriptors of color layout, scalable color, edge histogram, and color structure information for each image is extracted to give an indication of what image features can represent this image. SenseCam sensor readings (including accelerometer, ambient temperature, light level, and passive infrared detector) are then associated with each image based on time. The values of sensor readings and images features are normalized to ensure that they are all on the same scale for comparison. The adjacent image/sensor values are then compared against each other to determine how dissimilar they are. When the dissimilarity value is higher than a threshold value, a boundary for a new activity is considered.

Bolaños *et al.* (2015) used a CNN for egocentric videos as an unsupervised feature extractor for discovering objects. The author reported that the F-1 score of the classifier to score objectiveness in egocentric video sequences, are better than state-of-art techniques.

Accelerometer data has been used in many studies such as by Yang *et al.* (2008) and by Atallah *et al.* (2011) to provide recognition of everyday activities such as walking, running, sitting and lying. Many researchers utilized machine-learning techniques to segment each day into activities based on accelerometer data. Bao and Intille (2004) performed a study to detect activities of daily living using accelerometer data. Participants were asked to perform 20 different activities and five accelerometers were placed on the upper arm, lower arm, hip, thigh and ankle. Data from the accelerometers was labelled during the data collection process based on the performed activity. Features derived from both the time and frequency domain were extracted from the raw accelerometer data and used to train a number of classifiers including the C4.5 decision tree, decision tables, naive Bayes and nearest neighbor classifiers. The authors succeeded in classifying 20 different activities with an accuracy of 86% using the decision tree classifier. Long *et al.* (2009) used a single accelerometer to detect occurrences of five daily activities including walking, running, cycling, driving, and playing sports by applying a Bayesian classifier. The data was collected from wearable accelerometers without intervention and annotated thereafter by just examining the values of the three axial accelerators manually. Features of the data were extracted and used as input to the automatic classifier and the accuracy obtained from this was around 80%. Preece *et al.* (2009) did a similar study to recognize activities including walking, going up and down stairs, running, hopping on left or right leg and jumping, all based on wearable accelerometers. The highest activity recognition accuracy for a single sensor (97%) was achieved using an ankle-mounted accelerometer. Preece *et al.* (2009) also relied on features extracted from the raw accelerometer data namely the Fast Fourier transform (FFT) component feature set.

2.3 Time-Dependent Lifelogs

Time dependent lifelog research is more likely to consider lifelog data as time series data and therefore it is likely to apply analysis techniques originated from time series modeling. Those time series analysis techniques include, but are not limited to, detrended fluctuation analysis, autoregression, moving average, autoregression moving average, and some methods from signal processing.

Li *et al.* (2013b) used time series analysis methods to study chronologically-presented lifelogging images. Although the authors did not describe detailed experiments or show convincing data, their conclusion is still very heuristic: Detrended Fluctuation Analysis (DFA) shows lifelogging data is a time series with a cyclic fluctuation and it is not a random walk. This statement inspires us to explore time series analysis, given an assumption that lifelogging data is not totally random and this is the basis for our proposed work. Hu *et al.* (2009) apply DFA analysis to acceleromery data in order to test disturbances of scale-invariant activity fluctuation in elderly humans and elderly people with Alzheimer Disease based on the hypothesis that the master circadian pacemaker or suprachiasmatic nucleus is responsible for the scale-invariant activity fluctuation in humans.

Li *et al.* (2013c) applied cross-correlation matrix and Maximum Overlap Discrete Wavelet Transform (MODWT) to analyse SenseCam lifelog data streams. Eigenvalue decomposition was used at various wavelet scales to determine significant changes at different granularity in order to find out boundaries between events in lifelog data. Li *et al.* (2014) extend the previous work and improve the ability to remove noises by applying random matrix to cross-correlation matrices.

Wang and Smeaton (2012a) built an ontology of semantic concepts using automatically detected concepts in lifelog image collections. The author included semantic reasoning in order to deduce a set of concepts to represent lifelog content for applications like searching, browsing or summarization Wang and Smeaton (2013) used

Hidden Markov Models (HMMs) to capture the temporal information of detected concepts in lifelog data. The HMMs are thereafter used to detect complex/high-level activities.

Narimatsu and Kasai (2015) proposed a new dubbed Duration and Interval Hidden Markov Model (DI-HMM) trying to model sequential data by defining the event of continuous duration as 'state duration', and defining the discontinuous interval with no observation as 'state interval'.

Wang *et al.* (2016) explored and compared methods both using and without using temporal information. The authors applied concept detectors firstly and then used the outputs of concept detectors as inputs into a complex/high-level activities classifier. The result showed some interesting results, such as that with the improvement of basic concept detectors, the performance of time-series based concepts could be improved but if appropriate approaches are chosen, relatively poorer performing concept detectors can still achieve good performances. In other words, an appropriate way to choose, extract or model temporal information can compensate for the loss of information through concept detection.

Besides using images as lifelog data, other types of sensor data have also been investigated by researchers such as accelerometry, pressure and electrophysiological data. Cheng *et al.* (2016) extracting subtle features from a textile-based pressure sensor array. The author was expecting that the results of an activity recognition classifier could be improved. One feature utilized was extracted from the time domain to improve the performance of the classifier but the importance and properties of each feature are not discussed in the paper.

Krafty *et al.* (2015) used electrophysiological time series by polysomnography to characterize sleep and elucidate the pathways through which sleep affects, and can be treated to improve, health and functioning. The authors analysed data to quantify associations between power spectra of time series from different sleep periods with cross-sectional clinical and behavioral variables.

2.4 Periodicity

Periodicity is a natural phenomenon observed in biology, astronomy, financial and sociology such as sunspots which have a 12-years period, the recurring of financial crises whose periodicity can vary, signals emitted by pulsars, tides, circadian rhythms, seasons and the depths of a river, all which have a regular periodicity. In this section, some examples and applications of periodicity will be reviewed to show how periodicity is used across variety research areas.

2.4.1 Periodicity in Human Behaviour

Intuitively, we think our routine daily life is composed of another form of recurring events. Without the help of lifelogging devices, analyzing the periodicity of human life is a challenging proposition as it is not possible to ask a participant to keep accurate self-reports of their activities to the right level of detail over a period of months or years.

A study conducted by Czeisler *et al.* (1999) at Harvard University estimated the natural human rhythm to be closer to 24 hours and 11 minutes which is much closer to the solar day but still not perfectly in sync. Studies by Figueiro *et al.* (2006) have also shown that exposure to sunlight has a direct effect on human health namely breast cancer in the paper because of the way it influences the circadian rhythms.

Brocklebank *et al.* (2015), Bassett *et al.* (2015), Healy *et al.* (2011) and Knutson *et al.* (2011) have shown that human behaviors that are measurable using an accelerometer — sleep, sedentary behavior, and more active behaviors — are consistently associated with cardiometabolic risk biomarkers and health-related quality of life. Accelerometers can also capture the patterns in which sleep, sedentary, and active behaviors are accumulated. For example, Wolff-Hughes *et al.* (2015) showed that physical activity accumulated in bouts of $\geq$ 10 minutes have stronger relationships with health outcomes than total physical activity. Finally, in work on analysis

of sleep, Cappuccio *et al.* (2010) used accelerometers to quantify measures of sleep quality (e.g., sleep efficiency, wake after sleep onset) which typically provide greater predictive value of health outcomes than sleep duration alone.

Despite the capabilities when using accelerometers to measure behaviors across the 24 hour spectrum, less is known about metrics that encapsulate the full 24 hours that could be derived from accelerometer data. These metrics may identify unique patterns of behavior that could further explain relationships with health outcomes. One such known metric that is ascertained from accelerometry is the rest-activity cycle, that can represent the human circadian system. Gu *et al.* (2014) state that disruptions in the circadian system consistently show profound and detrimental impacts on health. Mormont *et al.* (2000) and Innominato *et al.* (2009) conducted studies using accelerometry and have shown relationships with health-related quality of life and better survival following metastatic colorectal cancer chemotherapy treatment.

Recently, with the growth of more wearable accelerometers that accommodate larger storage capacities, waterproofing, and more unobtrusive wear locations, long-term monitoring of behavior (i.e. much greater than one week) throughout the 24 hour spectrum have become more feasible. Indeed, consumer-based accelerometers (e.g., Fitbit, Jawbone) are already achieving long-term population-level data collection of these health behaviors. With the collection of long-term data, it may now be possible to characterize weekly, seasonal, and even annual patterns of behavior that encapsulate the full 24 hour spectrum that extend beyond traditional methods (e.g. accelerometry thresholds, sleep/wake rhythms). Hu *et al.* (2014) conducted experiments on several lifelog data collections and found that periodicities (i.e., repeating patterns) are observed in many human behaviors and may be derived from various forms of lifelog data. However, such methods have not been applied to long-term monitoring of accelerometry data as yet.

2.4.2 Periodicity in Astronomy, Finance and Biology

Pulsars are known to emit pulse signals throughout the universe. Since the first discovery of pulsars by Bell Burnell (1979), the study of pulsars is an important topic in Astronomy. The periodicity of a pulsar's signal can be used to determine the age of a pulsar and it thought to be as accurate as an atomic clock. 1993 the Nobel prize was awarded to Russell Alan Hulse and Joseph Hooton Taylor, Jr. for their discovery of a new type of pulsar, a discovery that has opened up new possibilities for the study of gravitation. Current research by Zhu *et al.* (2014) applied image pattern recognition algorithms training a classifier using statistical features of the image and using the trained classifier to find new pulsars.

An economic cycle is the upward and downward movement of Gross Domestic Product (GDP) of a country around its long-term growth trend. In Capitalism theory we always see words like "booms" or "expansions" which correspond to rapid economic growth and "recessions" or "contractions" which correspond to economic decline or stagnation. The booms and recessions shifting over time are the representations of the economic cycle. Those fluctuations in economic activities are proven to be unpredictable.

Chronobiological studies include, but are not limited to comparative anatomy, physiology, genetics, molecular biology and behavior of organisms within biological rhythms mechanics. The circadian rhythm is important and investigated in chronobiology. The process is related to an oscillation around 24-hour rhythms which are driven by a circadian clock, and they have been widely observed in plants, animals, tides and so on. Circadian rhythmicity is present in the sleeping and feeding patterns of animals, including human beings. There are also clear patterns of core body temperature, brain wave activity, hormone production, cell regeneration, and other biological activities which use this same periodicity. We are able to find a clock-related genome in mice, clock-mutant mice are hyperphagic and obese, and

have altered glucose metabolism Arendt (1997).

In the next section we will present the various datasets constituting lifelogs of different people, that will be used in the remainder of the work in this thesis.

2.5 Datasets Used in this Thesis

In order to fully explore the hypothesis and research questions set out earlier in this thesis (see section 1.2), a set of lifelog datasets with different characteristics are needed, characteristics which differ in terms of length of recording, frequency of recording, nature of data captured, nature of the subject(s), and variety across subjects. To this end, we have gathered five different datasets which we outline briefly here, and which we will return to later in the thesis. The several datasets we use are chosen so as to support an investigation into our different research questions. In the literature we see many works being done to analysis personal lifelog data on per-person base. But researches rarely mentioned the validity of data for each subject, namely the question why is the analysis of the data can be applied to the subject. In other words, the validity of personal data framework hasn't been studied completely which is one of the focus of the thesis. We have also seen several deep learning methods are used in computer vision community. Several researches apply deep learning on lifelog images, but few of them use this approach on low level longitudinal acceleration data. We bring state-of-art deep learning methods to longitudinal acceleration data. In the literature, numerous researcher trying to build connections between people and data through events, but relate people and data directly using raw data is not common. We explore raw data to identify periodicity and intensity of periodicity to bridge the gap between people and data.

2.5.1 Athletic Data

This dataset represents a 10-year record of physical exercise and training activities including running, cycling and swimming, from an international triathlete (now retired from competition). The log contains a daily entry for distance covered for 1 or more of the three sports as well as daily text comments which can indicate mood, training effort, relative performance, weather, etc. and these can be analysed for sentiment. This sports dataset capture 100% of activity log in the 10 years, i.e. there are no missing entries and the log is complete.

Obvious periodicities which could be detected from this data include seasons, performance at targeted sports events, perturbations caused by occasional injury and overall decline over the decade from aging.

We manually annotated the daily reports by the athlete for both mood and for performance level to create a third dataset (the other two being distance covered in each of 3 sports). Four annotators were asked to annotate the text for mood by following the following strategy: if a comment provides an indication of mood (e.g. " feeling great", "not well" or "ok"), the annotators assign a rating between 1 and 5, where 1 indicates the worst feelings and 5 indicates the best feelings. If there is no indication of mood in the text, the annotators assign a rating of 0. For annotation of performance level, the four annotators were given the following instruction: when a comment provides an indication of performance level (e.g. "personal best", "strong finish", or "stopped early"), they give a rating between 1 and 5 where 1 indicates poorest performance and 5 indicates best performance. If there is no indication of performance in the text, the annotators assign a rating of 0. The purpose to assign 0 to missing values is to identify between missing values and non-missing values. The assigned zeros are not used in the processing of data.

Comments made by the athlete during the year 2007 were randomised and presented to 4 annotators. Because the marks for mood and performance given by

annotators are highly subjective and have biases, inter-annotator agreement namely Cohen's Kappa co-efficient as described by Cohen (1968), was calculated across the annotators.

This dataset is used for RQ2 for the prediction of high-level lifelog data. Because it's logged for more than 10 year. It is used in RQ3, RQ4 and RQ5 as well, because we would like to see if there is an periodicity in sports data and annotation of sports data or not. If an periodicity is present in this data set we would like to compute the intensity of the detected periodicity. The result of intensity of periodicity will be then summarized to questions which is asked to the subject who generate those data in an interview.

2.5.2 Sleep Data

The dataset represents 2.5 years of continuous nightly sleep monitoring for an individual (male, middle age, daytime 5-day week worker) with a $\geq$80 % capture rate. Data was collected using the wrist-worn Lark sleep sensor[4] and contains the following information:

1. Time to sleep - represents the time taken between going to bed and falling asleep;

2. Time to rise - represents the time taken between waking and getting out of bed;

3. Time asleep - represents the duration of sleep;

4. Quality - a numeric indicator of sleep quality computed as a function of how well the night's sleep mapped to the circadian sleep (90-minute) rhythm and how many cycles of that rhythm were completed;

[4] `http://www.lark.com`

5. Times woken up - represents the number of instances of a wake-up during sleep, where "wake up" represents even a turning over in the bed;

An obvious periodicity we would expect to detect in this data is based on the weekly cycle where the subject tends to sleep longer at weekends than during work-days because he has a regular work schedule of Monday to Friday.

This dataset is used only in RQ3 to show that even in unevenly sampled lifelog data we are able to detect periodicity. We didn't apply it in other RQs is because it has only 2.5 year with about 20% of data missing. And sleep conditions are hard to remember especially there is no comment or annotation which will help the subject to remember anything during an interview.

2.5.3 ASU Accelerometer Data

The ASU runs a program to monitor sleep quality for US Veterans. Participants in the program used a smartphone-based, multi-component behavioral intervention which targeted changes in sleep, sedentary behavior, and more active behaviors. In other words, the smartphone's on board sensors were used to determine message interventions to help subjects improve their sleep and reduce the amount of their sedentary behaviour.

The target population in this study was US Veterans currently receiving clinical care at a regional Veterans Health Administration (VHA) hospital in the Southwestern United States, aged 35-65 years, measured as either overweight or obese (BMI $\geq$ 25 kg/m2), with a fasting glucose of $\geq$ 100 mg/dL. Eligibility criteria for the study also included reporting of (a) insufficient physical activity (defined as endorsing activity ranking categories $\leq$ 4 on the Stanford Brief Activity Survey by Taylor-Piliae *et al.* (2006), which closely aligns with US national physical activity guidelines), excessive sitting (defined as $\geq$ 8 hours of sitting from the International Physical Activity Questionnaire [IPAQ] by Hagströmer *et al.* (2006)), and short

sleep duration (< 7 hours/night) or mild/moderate sleep complaint (modified version of the Insomnia Severity Index [ISI] by Morin *et al.* (2011)). All participants completed a telephone screening to determine eligibility. Institutional review boards governing the local VHA hospital and the Arizona State University in which some of the researchers on the study were affiliated, approved all study procedures. All participants provided written informed consent.

Participants were initially screened by telephone and this was followed by an in-person visit to confirm eligibility and complete informed consent procedures. At this visit participants were given a wrist-worn accelerometer for three consecutive weeks. This period constituted the "un-in" period of the behavioral intervention and baseline data collection. Participants were instructed to wear the monitor continuously during both sleep and wake. As part of the run-in period, participants were asked to self-monitor their sleep, sedentary, and active behaviors using a customized smartphone application designed for this purpose. After two weeks, participants were mailed a second accelerometer and asked to return the first accelerometer in a pre-paid envelope. At three weeks participants returned for a second in-person visit where the second accelerometer was returned and all other study measures including questionnaires, blood draws, and clinical measurements were completed. Participants received 25 USD for completing study measures at this visit.

Following this visit, participants were randomized to receive active elements of the behavioral intervention. A full description of the intervention can be found in the work by Buman *et al.*, but briefly, participants were randomized into a full-factorial 2x2x2 screening experiment where smartphone-based interventions targeting sleep, sedentary behavior, and physical activity were delivered for 8 weeks. All participants maintained self-monitoring of their behaviors using the custom application during the intervention phase. Participants also attended two additional visits during the eight weeks to complete study-related assessments and to return/exchange accelerometers to maintain continuous wear and data collection. To take advantage

of the continuous and longitudinal nature of the data, the full accelerometer data for the run-in and intervention periods were leveraged for this analysis and the effect of the intervention was statistically controlled for in all analyses.

During the data-gathering, movements during sleep and wake by participants were monitored objectively and continuously throughout the study period using the GENEactiv accelerometer[5]. The GENEActiv is an open source, wave-form wrist-worn accelerometer that is fully waterproof, allowing the monitor to be worn continuously, 24 hours a day without the need to be removed during water activities or be shifted from hip to wrist for daytime and nighttime measurement. Since the GENEactiv provides continuous forms of data recordings for periods of at least 1 month it can be considered a valid form of lifelogging. Raw data captures acceleration of X, Y and Z 3 axis along side temperature and lumen. All raw data is encoded with HEX format. Data captured on board the device were initially sampled at 40Hz and summarized to 60 second epochs using a gravity-subtracted sum of vector magnitudes and provided through the Activinsights software package. Periods of non-wear were screened for and removed based upon variability in the monitor temperature outputs (i.e., low variability indicates lack of normal fluctuation in temperatures indicated human wearing), and visual inspection. Additional removal occurred for overlapping wear periods that occurred when the monitors were in transit by post.

Before data processing, there are several practical difficulties in the data set that need to be addressed. First of all, during the data capture process, some data entries have become corrupted or some data attributes have not been logged due to various reasons. Beside this, in the ASU data there are some large data gaps due to the fact that participants were encourage to wear the sensor as long as possible, but they were still free to decide when and where to wear the device. There is also an overlap problem in the ASU data. The device cannot log data constantly for 3 months, which is the upper limit of the logging period, so it has to be changed at some point,

[5] `http://www.geneactiv.org`

perhaps more than once. This is done by either the participants visiting the lab or a new sensor was posted to them. The overlap problem is because the new sensor will have started to log data while the old sensor was still attached to the participant and the exact time when the participants changed the sensor was not recorded. If the participants change the sensor on site in the lab, the overlapping problem still exists but is minor, but if the new sensor was posted to a participant, the overlap could be large and must be fixed. Under current experimental settings, we manually decide the time points where the changing of sensors happened. This dataset poses some interesting possibilities for periodicity which may improve or disimprove before and after the sleep intervention and it will be interesting to examine this to see if the intervention has an impact on the participants' periodicity.

This dataset is used in all the RQs. It is used in RQ1, because 25 subjects participated in the experiment. For RQ2, each subject is logged for months' data with high sampling rate. We use it for RQ3 because we are hypothesis that there are periodicities in such longitudinal raw acceleration data. For RQ4 and RQ5, we are interested to see if the intensity of a certain periodicity would be correlated with some health related biomarkers.

2.6 Conclusion

In this chapter, related lifelog research was reviewed, including state-of-art data capturing methods, detection of boundaries in lifelogs, event detection and applying time series analysis method to lifelog data. Periodicity, the phenomenon of regularly occurring patters in other fields was also introduced and reviewed. Despite much research carried out in lifelogging and post-processing of lifelog images, research concentrating on analysis of lifelogging data which uses the chronological aspect and investigates longitudinal aspects and its causality and impact on lifestyle, is not widespread. Consider a sequence of collected data where we could potentially

discover some patterns such as trends, or even full periodicity, seasonality, partial periodicity, periodicity association rules, frequent patterns and cross correlation. Detecting such patterns would be very helpful to reveal aspects of a person's lifestyle. One of the most interesting patterns of a person's lifestyle is their daily routine, it conveys a near-truth of your lifestyle by your recurring activities or events, in other words, the periodicity of your life. This is one of the things we set out to investigate in this thesis.

Chapter 3

Uniqueness of Users and their Lifelog Data

This chapter is related to the first of the research questions (RQ1) introduced at the start of this thesis, which concentrates on the question: *Does longitudinal log data have some unique features than can distinguish the subject generating the data from other subjects ?* In other words, from a longitudinal perspective, does the lifelog data captured from different subjects share more common structural or similarity or even identical patterns so that it is difficult to identify subjects from the data ? For instance, if we assume two subjects are generating two time series of lifelog data and if the two data series are very distinct from each other, we could say that the two subjects share less common properties and there is a lesser chance for us to discover similarity in lifelog patterns between them. In another extreme example, if two subjects generate exactly the same data series, we could say that the patterns found in one subject can be applied to another relatively easily, because they are identical. Before looking at methods for finding common patterns and periodicity within lifelog data, as introduced in Chapter 5, it is very important to look into lifelog data itself first. The purpose of doing this is to discover some unseen and perhaps unknown properties of lifelog data that we are dealing with. Each of the

sections in this chapter will try to test or examine one of the following statements:

1. Lifelog data generated by one person is different from lifelog data generated by other people.

2. Lifelog data generated by one person changes with time.

3. Computer algorithms are able to identify different persons given lifelog data series.

Section 3.1 is related to the first statement. The overall distribution of data generated by different subjects will be computed and metrics that can measure similarity between the distribution of the subjects will show the uniqueness of each subject in terms of data distribution of their lifelog data.

Section 3.2 is related to the second statement. The distribution of data generated by each subject will be studied. Each data series of lifelog data will be divided into window size of a day. The distribution within each window will be computed and the distribution changing with time will be calculated, in order to see how lifelog data does change with time.

Section 3.3 is related the third statement. Several machine learning methods will be used to classify different subjects. We expect to see that the result of this classification will be better than a random guess.

It is legitimate to envisage a more or less bijective correspondence between each subject and the data generated by the subject. If the one-to-one mapping exists on a degree, it implies that each person has enough uniqueness in their accelerometer data to distinguish themselves from others, that this then makes the case for a need for visualisation and analysis on a per-subject, individual basis. On the other hand, the bijection between subjects and data could also means that patterns mined on each data series do represent the subject, because of the uniqueness carried in the time series. This could potentially be the bases and foundation of analyzing longitudinal-populational lifelog data.

3.1 Data distribution

In this section, data distribution refers to a probability density distribution of lifelog data and we will compute the data distribution for each unique subject for whom we have lifelog data. The reason we are doing this is because we are interested to see if there are any differences among subjects in terms of their data distributions. Several methods to measure differences between distributions will be used.

3.1.1 Experiment Settings

We will use the ASU dataset, introduced earlier in Section 2.5.3, to examine the distribution of lifelog data across different subjects. The ASU data is a dataset which contains data on more than 25 subjects. Each subject was asked to wear a wristband like wearable sensor freely for several weeks without interruption. The dataset includes data gathered by an accelerometer, luminance, temperature, voltage of the device itself and all these are timestamped. In a data distribution analysis experiment, only the 3-dimensional accelerometer data, and its timestamp, will be used. In the first experiment the time variable is not considered, which means the data generated from the beginning until the end of the data collection period will be used as a whole.

1D histogram is computed using 999 bins with range from -8.5 to 8.5. Three histograms are computed from each axis of the accelerometer. For 2D histogram, we use 999×999 lattice as bins with range from -8.5 to 8.5 for each dimension. Three histograms are computed for X-Y axis, X-Z axis and Y-Z axis of the accelerometer. As for 3D histogram, due to limited memory, we apply $199 \times 199 \times 199$ cubes as bins with range from -8.5 to 8.5 for each dimension. There is one histogram computed for the 3D case.

3.1.2 Results of Data Distribution

We firstly visualize and examine the ASU data. Figure 3.1 is a 2D histogram of the distribution of all the X and Y axis of accelerometer data collected. The colour shows the probability of a value falling into an interval indicated by the X and Y axis. The value of the probability is indicated using colour coding. The colour mapping is indicated on the right side of the figure. The lighter the colour, the higher the probability value is. We can see there is a vague white circle in the middle because data has a smaller chance to fall on the most of the intervals. Thus we enlarged the centre part where the interval ranges from around -2.5G to +2.5G, so that we can see a clearer picture of the distribution in the centre. Figure 3.2 shows the result of the enlarged centre of the histogram. The same idea was applied to a histogram of X and Z axis in Figure3.3 and Figure 3.4, and also a histogram of Y and Z axis in Figure 3.5 and Figure 3.6.

At first glance, the distribution of data seems to look like a 2D Gaussian distribution, where there is a peak in the centre and probability values getting lower when the distance between points and centre becoming bigger. Although the histograms of X-Y axis, X-Z axis and Y-Z axis are not exactly the same, they are more or less following similar pattern, which is to have a smaller chance to have values around the edges of the histogram.

We can observe something very interesting after we enlarge the histogram in all three settings namely that there is a circle that can be seen among all three settings. The reason we can see the circle is because the colour of the circle is lighter than the colour of the background. From a 3D perspective, we can imagine the higher value or lighter colour as hills and darker background as valleys, so most of the data that the subjects generated are bounded between -g and g. And we also can see some hills in the area bounded by the hills with radius of one gravity.

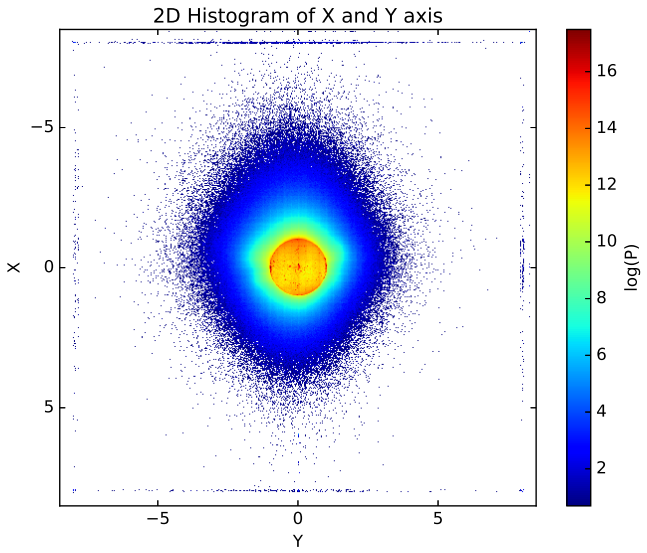


Figure 3.1: Histogram for X and Y axis of raw acceleration from all subjects. X axis is the range of Y-axis acceleration from -8.5G to 8.5G. Y axis is the range of X-axis acceleration from -8.5G to 8.5G. Colour scale represents the logarithm of probability.

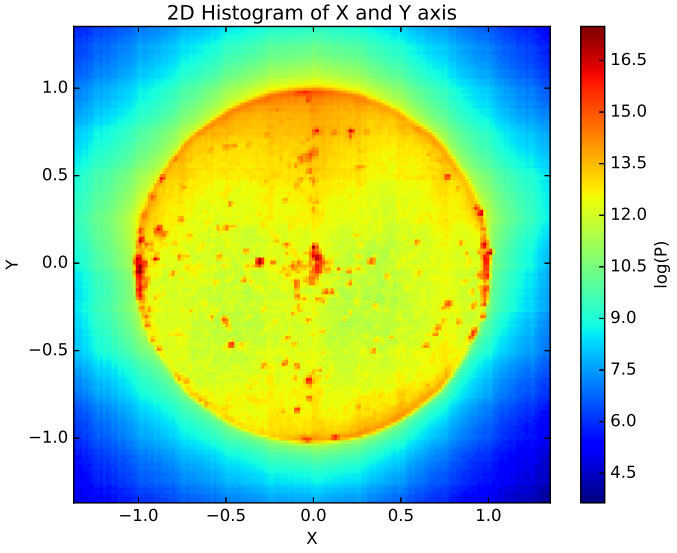


Figure 3.2: Enlarged histogram for X and Y axis of raw acceleration from all subjects. X axis is the range of Y-axis acceleration from -1.4G to 1.4G. Y axis is the range of X-axis acceleration from -1.4G to 1.4G. Colour scale represents the logarithm of probability.

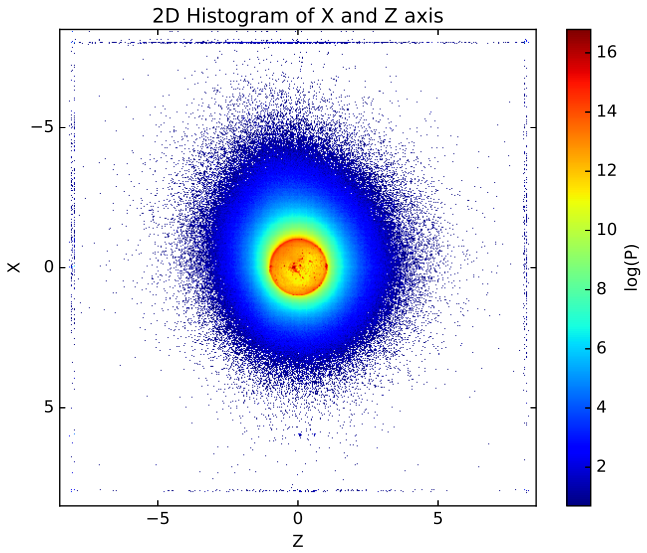


Figure 3.3: Histogram for X and Z axis of raw acceleration from all subjects. X axis is the range of Z-axis acceleration from -8.5G to 8.5G. Y axis is the range of X-axis acceleration from -8.5G to 8.5G. Colour scale represents the logarithm of probability.

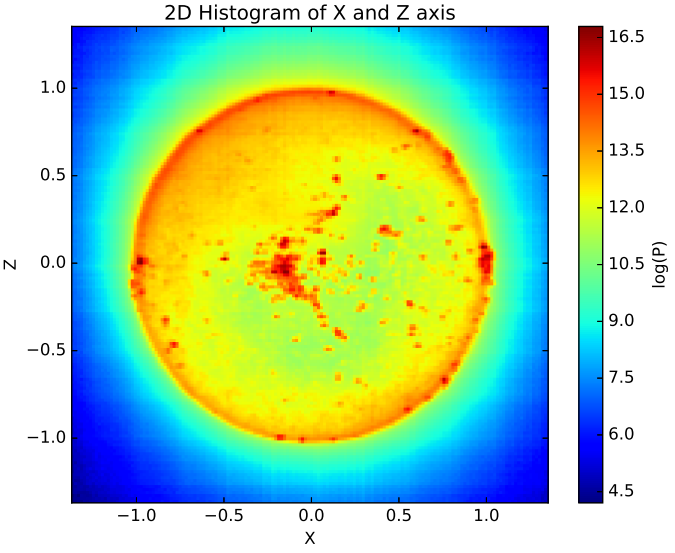


Figure 3.4: Enlarged histogram for X and Z axis of raw acceleration from all subjects. X axis is the range of Z-axis acceleration from -1.4G to 1.4G. Y axis is the range of X-axis acceleration from -1.4G to 1.4G. Colour scale represents the logarithm of probability.

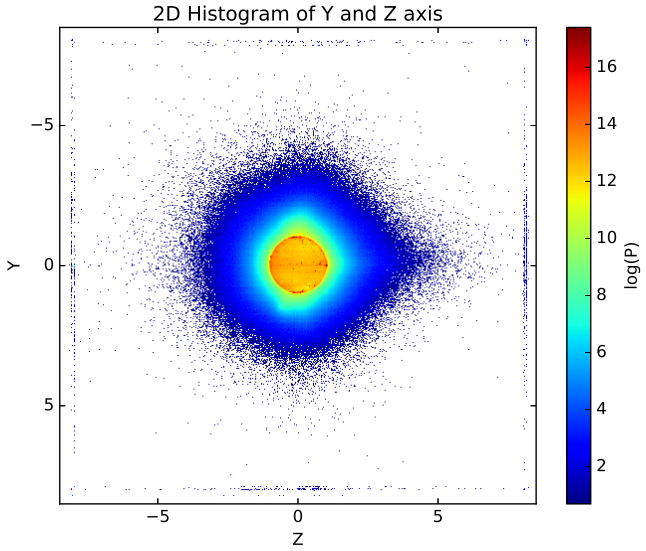


Figure 3.5: Histogram for Y and Z axis of raw acceleration from all subjects. X axis is the range of Z-axis acceleration from -8.5G to 8.5G. Y axis is the range of Y-axis acceleration from -8.5G to 8.5G. Colour scale represents the logarithm of probability.

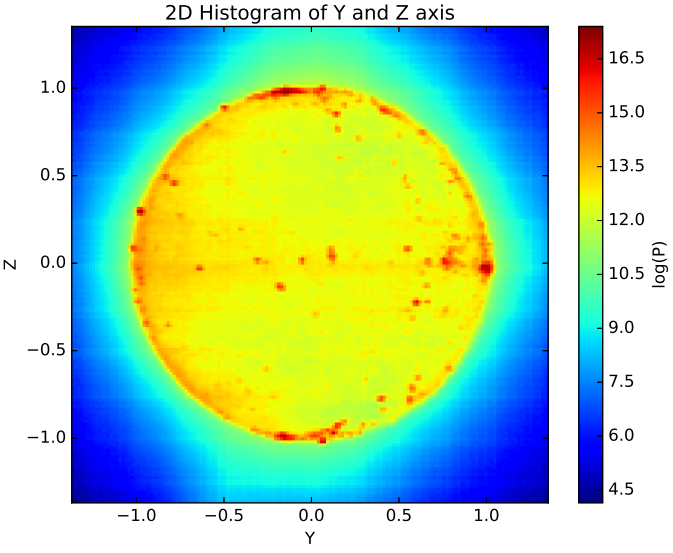


Figure 3.6: Enlarged histogram for Y and Z axis of raw acceleration from all subjects. X axis is the range of Z-axis acceleration from -1.4G to 1.4G. Y axis is the range of Y-axis acceleration from -1.4G to 1.4G. Colour scale represents the logarithm of probability.

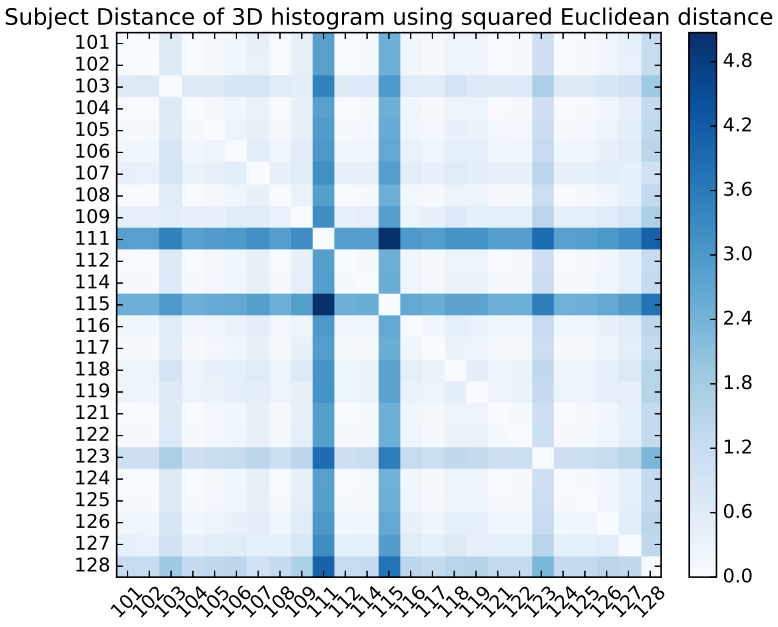


Figure 3.7: Distances between 3D histograms of subjects using squared Euclidean distance. X and Y axis are subject number. Colour scale represent the normalized distance between subjects.

3.1.3 Similarity Metrics

A 3D histogram will be used to calculate the distance in terms of lifelog data, between each subject. In order to see how different distance matrices work on our 3D histogram, we will use and compare several distance metrics. We can hypothesis now that different metrics could reveal or emphasize different aspects of difference. Distance metrics are always used to measure dissimilarity, for instance the squared Euclidean distance is a norm 2 distance which is widely used to measure similarity, where $s(x, y)$ measures the distance between x and y:

$$s(x, y) = ||x - y||_2^2$$

Cosine similarity measures the similarity between two vectors by calculating the cosine of the angle between two vectors. The cosine of $0°$ is 1 which means the two

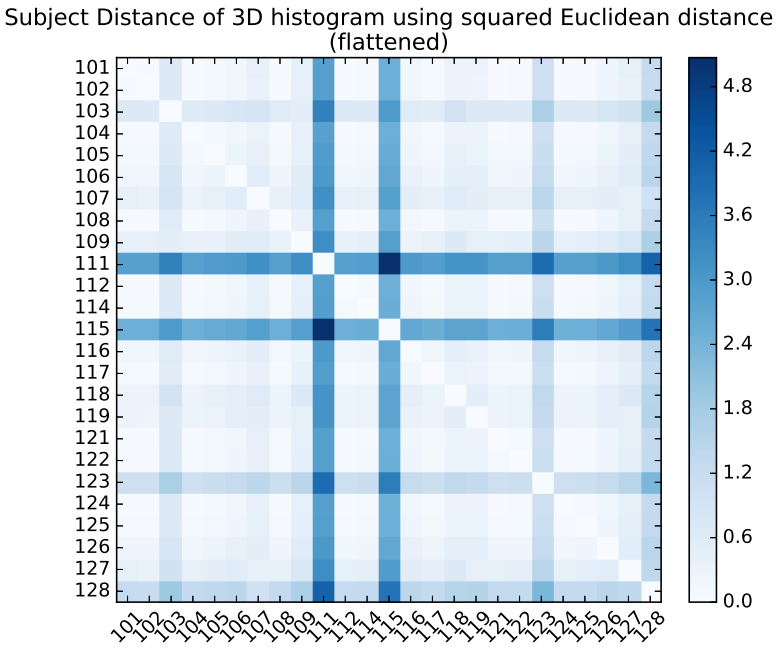


Figure 3.8: Distances between 3D histograms of subjects using squared Euclidean distance (flattened). X and Y axis are subject number. Colour scale represent the normalized distance between subjects.

vectors point in the same direction while the cosine of 180° is 1 which means the two vectors point in the opposite direction to each other. The cosine distance of between vectors A and B can be written as:

$$\cos(\theta) = 1 - \frac{A \cdot B}{||A|| * ||B||}$$

In probability theory and in information theory, the Kullback-Leibler divergence or KL divergence (Kullback, 1987) is a measure of the difference between two probability distributions P and Q. The Q distribution is normally referred to as the theoretical distribution while P is considered as the actual distribution. The KL divergence is given by the formula

$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

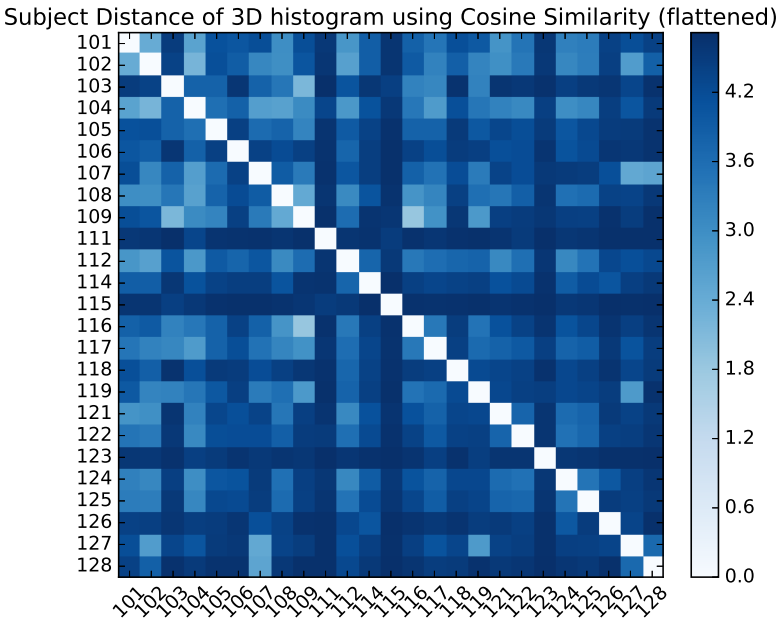


Figure 3.9: Distances between 3D histograms of subjects using cosine similarity (flattened). X and Y axis are subject number. Colour scale represent the normalized distance between subjects.

It should be noted that the measure between distribution P and Q is not symmetric, which means that the KL divergence from P to Q is not equal to the KL divergence from Q to P.

We applied the above-mentioned metrics to measure distances between each subject pair. Note that distance is the negative value of similarity. We used colour coding in the visualization of the similarity metrics, where the lighter the colour, the more similar two subjects are. The result of the original metrics are adjusted by normalization and divided by standard deviation using the following equation:

$$A_{ij} = \frac{A_{ij}^{original}}{\sigma * \sum_{i,j} A_{ij}}$$

where A_{ij} is the distance matrix and σ is the standard deviation of the flattened distance matrix A_{ij}. This step is to make different measurements are within a comparable range of values. Figure 3.7 and Figure 3.8 show that according to the

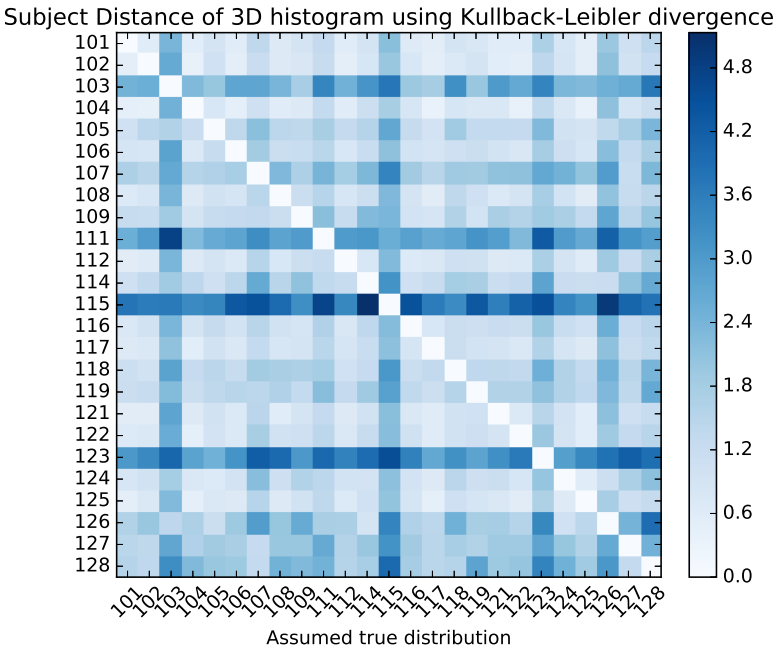


Figure 3.10: Distances between 3D histograms of subjects using Kullback-Leibler divergence. X and Y axis are subject number. Colour scale represent the normalized distance between subjects.

squared Euclidean distance, most of the subjects are similar to each other, except for subjects 103, 111, 115, 123 and 128. The metric might be good at locating subjects that are very different. Figure 3.9 uses cosine similarity to measure the distance between subjects and it turns out that according to this metric, most subjects are quite different from each other. Figure 3.10 uses Kullback-Leibler divergence, a metric which is known to be asymmetric, which could cause some problems in some circumstances. The metrics shows subjects 103, 111, 115 and 123 are very different from the rest of subjects. The result is similar to squared Euclidean distance while KL divergence also reveals the difference of other subjects from each other in general. To summarize, of the metrics for measuring distribution similarity, squared Euclidean distance is least capable of showing the difference, while cosine similarity has the strongest ability, and KL divergence somewhere in-between.

3.2 Data distribution through time

In the previous section, our experiment showed how the ASU accelerometer data distributed differently across participants and how different metrics showed the discrepancy different subjects. In this section, we try to show how lifelog data changes over time. A similar experimental setting will be used as in the previous section. Time series accelerometer lifelog data from each of our subjects will be separated into each day. We take advantage of the knowledge we acquired about the data distribution from the previous section, so that distribution in this section is calculated by the 3D histogram where each axis has 199 bins which range from -2.5G to +2.5G. Because we observed that most of the acceleration data falls within a circle with 1G radius with the centre located at zero gravity, in 3D, it should be a sphere with a 1G radius with the centre located at zero gravity. To make the boundary loose, we use -2.5G to +2.5G instead of -G to +G as the range of the histogram. As for the choice of number of bins, since 3D histogram takes much more memory than 2D and 1D, so that we apply small 199 bins on 3D histogram. On the other hand, a reduced interval and a reduced number of bins can reduced the size of histogram significantly while speeding up the calculation of similarity and reducing memory usage. Then the 3D distribution of each day from each subject would be calculated in order to show the change of data generated each day, for every subject.

Figure 3.11 and Figure 3.12 show two examples of the change in lifelog data for individual subjects' data over time. The red curve is the change of entropy. Entropy represents how much information is being carried in the data. When the entropy line goes down, it means accelerometer data carries less information. In other words, there is an abnormality at the point where entropy is lower than most of the days but we can't tell the exact reason for the low entropy day. This might because the subject switched off the device, the subject didn't wear the device, or it is a switching overlap day and so forth. For example in Figure 3.11 there are

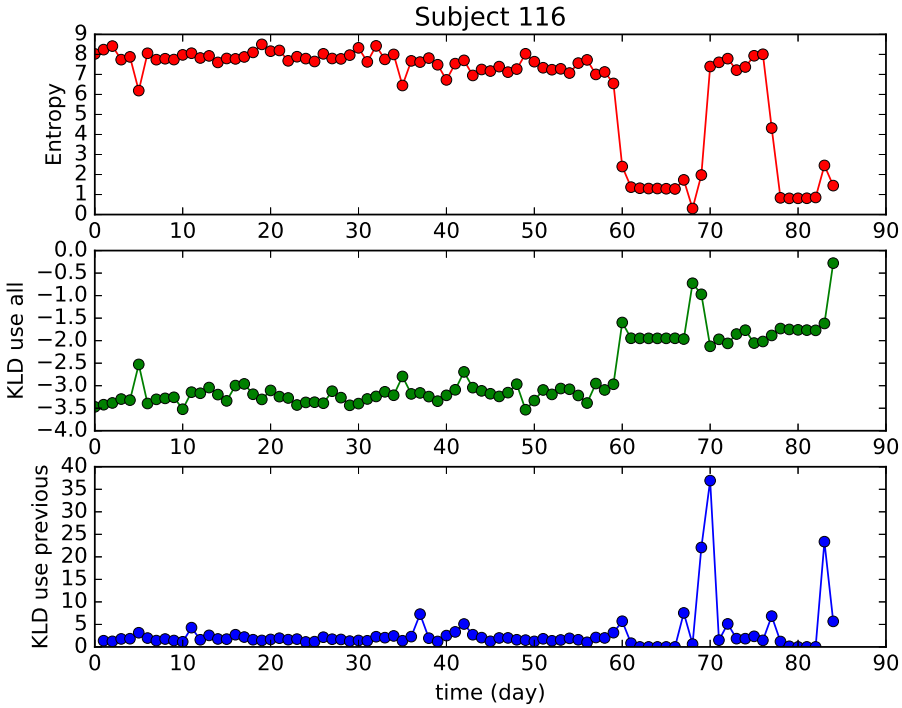


Figure 3.11: Change of distribution (Subject 116). X axis for all three panels are time in days. Y axis for first panel is entropy. Y axis for first panel is KLD using all data as true distribution. Y axis for first panel is entropy using previous day data as true distribution.

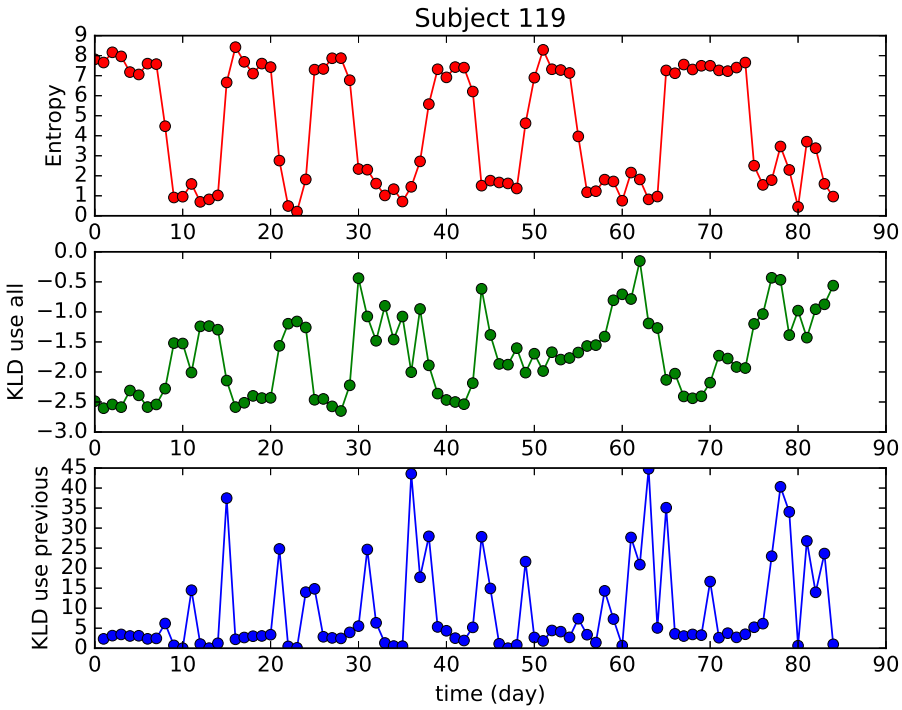


Figure 3.12: Change of distribution (Subject 119). Y axis for first panel is entropy. Y axis for first panel is KLD using all data as true distribution. Y axis for first panel is entropy using previous day data as true distribution.

low entropy periods between 60-70 days and at around 80 days. In Figure 3.12 the red line or entropy line is oscillating with a period of around 12 days. This might indicate that the subject repeats some action periodically. The blue curve in the figures indicates the KL divergence between the current day and the previous day and the higher the value, the bigger the difference. In both Figure 3.11 and Figure 3.12, we observe the big change in distribution happens at the borderline of high entropy and low entropy days, while there are some small changes corresponding to small entropy changes during high entropy days. The green curves show the KL divergence between each day and the total distribution, namely the sum of distribution of each day, for each subject. This indicates the abnormality of each day but shows more trending features instead of just local changes. For example, for subject 116, we can see the abnormality is increased from a point at around 60 days.

3.2.1 Change of Distribution / Change Detection

The objective of change point detection is to discover abrupt property changes lying behind or embedded within time series data. This is the process of identifying differences in the state of an object or phenomenon by observing it at different times. Change detection can be categorized into real-time detection (Garnett *et al.*, 2009) and retrospective detection (Yamanishi and Takeuchi, 2002; Moskvina and Zhigljavsky, 2001). Real-time change-point detection targets applications that require immediate responses such as, for example, robot control. On the other hand, although retrospective change-point detection requires longer reaction periods, it tends to give more robust and accurate detection. Retrospective change point detection accommodates various applications that allow certain delays, for example, climate change detection (Reeves *et al.*, 2007), genetic time-series analysis (Wang *et al.*, 2011), signal segmentation (Basseville *et al.*, 1993), and intrusion detection

in computer networks (Yamanishi *et al.*, 2004). Change detection is also very important in geo-informatics using remotely-sensed data obtained from Earth-orbiting satellites (Verbesselt *et al.*, 2010), such as land use change analysis, monitoring of shifting cultivation, assessment of deforestation, study of changes in vegetation phenology, seasonal changes in pasture production, damage assessment, crop stress detection, disaster monitoring snow-melt measurements, day/night analysis of thermal characteristics and other environmental changes.

Researchers have developed methods to either directly estimate probability density or the ratio of probability densities. Probability density estimation can be tackled using parametric methods or non-parametric methods. The first method requires an assumption of pre-defined density functions like probability density models, autoregressive models, and state-space models. Within statistical framework, several parametric methods are proposed such as generalized likelihood ratio (Basseville *et al.*, 1993) and cumulative sum (Gustafsson, 1996; Gustafsson and Gustafsson, 2000). Those approaches focusing on logarithm of the likelihood ratio between two consecutive intervals in time-series data to detect change points. In data mining community has applied those parametric method to real life data and application such as novelty detection Guralnik and Srivastava (1999) and maximum-likelihood estimation Markou and Singh (2003) and autoregressive models trained with online learning Takeuchi and Yamanishi (2006). Some researchers are also focusing on methods using analysis of subspaces in which time-series sequences are constrained by Moskvina and Zhigljavsky (2003); Kawahara *et al.* (2007). This approach has a strong connection with a system identification method called subspace identification, which has been thoroughly studied in the area of control theory. Some non-parametric approaches to change detection problems have been discussed by Brodsky and Darkhovsky (2013) to compute the likelihood ratio. However, due to the properties of non-parametric density estimation Härdle *et al.* (2012), it is hard to put this naive approach into practice. Instead of estimate distributions

separately, Kawahara and Sugiyama (2009); Liu *et al.* (2013) used relative Pearson divergence as a divergence measurement estimated by a method to estimate density ratio directly, which avoids expensive non-parametric density estimation. In other words, the ratio of distributions are calculated directly instead of computing the distribution separately.

The point of change detection in lifelog data can be explained in Figure 3.11 and Figure 3.12 where we can see a clear entropy change. The detection could potentially reveal points where changes in lifelog data reflecting a reach change in the subject's lifestyle or behaviour, happen. On the other hand, it provides us with a prospect that the activity or behavioural change associated with the lifelog data change could also possibly be detected, and this is something that we shall return to later in the thesis.

3.3 Automatic Classification

Machine learning techniques are usually used to solve classification problems. A weak classifier can be as simple as using a threshold value where above, or below, this value signals a classification. However, most machine learning methods are more complicated than just computing a threshold value. To formalize the machine learning problem, we denote the following notations: assume there exists a function $g : X \to Y$ that can map input instance $x \in X$ to an output label $y \in Y$. If Y is discrete then this is defined as a classification problem otherwise if Y is continuous, then this is defined as a regression problem. Machine learning techniques try to learn a mapping function $h : X \to Y$ as close as the actual function g. One common way to learn an approximate function g is achieved by defining a loss function that will assign a value that indicates the difference between ground truth labels and labels being computed by an approximate function g. Then the goal of machine learning algorithms are to try to minimize the expected loss function. For example

1-0 loss is a common loss function. This loss function simply assigns a loss of 1 to a label that does not match annotation while assigning a loss of 0 to a label that matches. From a probability point of view the problem becomes how to estimate the probability of each possible output label given a certain input instance. The formalization of the problem becomes:

$$p(l|x,\theta) = f(x,\theta)$$

where l is the label and f is parametrized by θ. Discriminative methods compute the probability directly from x and θ such as the general linear method or even a Support Vector Machine (SVM), while generative approaches compute the probability using Bayesian rules:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}$$

So the generative model will try to predict the probability of every possible label given input and parameters, as follows:

$$P(y|x,\theta) = \frac{p(x|y,\theta)p(y|\theta)}{\sum_{y \in Labels} p(x|y)p(y|\theta)}$$

The purpose of machine learning is to learn the best θ:

$$\theta^* = \arg\max_{\theta} p(\theta|X,Y)$$

where:

$$p(\theta|X,Y) = [\prod_i p(y_i|x_i,\theta)]p(\theta)$$

3.3.1 Machine Learning Methods

The following sub-sections will briefly introduce how different methods would tackle the machine learning problem.

3.3.1.1 Support Vector Machine (SVM)

The Support Vector Machine (SVM) was and is still a popular machine learning algorithm originated from Vapnik-Chervonenkis' theory (Vapnik and Chervonenkis, 2015). The theory tries to explain what is the condition that makes data learnable and the SVM implementation is one of the products of that theory. An implementation of a SVM tries to find a hyperplane $w \cdot x - b = 0$ given a training set $x \in X$ and $y \in Y$. One can find the intermediate equations and techniques to assist with implementation such as Lagrange multipliers in the paper by Cortes and Vapnik (1995). The original optimization problem becomes:

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$s.t., \alpha_i \geq 0, i = 1, ..., n$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0$$

where α_i are the Lagrange multipliers. If we solve the α_i, we can then get the hyperplane by $w = \sum_i \alpha_i y_i x_i$ and $b = w \cdot x_i - y_i$. Sequential minimal optimization (SMO) as described by Platt *et al.* (1998) is normally used to solve the optimization problem by fixing two Lagrange multipliers at one time. The SVM with soft margins and kernel trick is then used to solve the situation that the training data are not separable. The whole problem can then be reformulated as:

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j k < x_i^T x_j >$$

$$s.t., 0 \leq \alpha_i \leq C, i = 1, ..., n$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0$$

where C is the parameter that would allow some training data to be mis-classified. k is the kernel that needs to be determined.

3.3.1.2 Neural Networks (Recurrent Neural Network (RNN)/LSTM/Auto-Encoder)

The technique of using neural networks, especially deep neural networks, for classification is very popular among the research community. Researches have found that the human brain has a deep structure and deep neural networks are designed to imitate this deep structure of human brain. This makes deep neural networks good for solving the same kinds of problems the brain solves, like vision and recognition. In order to achieve the deep structure, many layers of neural networks need to be stacked together but a stacked deep structure had brought up some new issues and problems. These problems arise from the backpropagation algorithm which we use to train neural networks. Backpropagation first calculates the error or loss between output of forward propagation and ground truth, and then computes the gradient of weights, layer by layer, applying a chain rule so that the error propagates from output to input. The weights between layers are then updated by minus the portion of the gradient. After the weights are updated iteratively using gradient descent, the loss function will be close to local minimum. Some methods can also give a momentum while we perform gradient descent in order to jump out of the local minimum.

During backpropagation, the error would be disappear from output to input. Traditional activation functions such as the hyperbolic tangent function have gradients in the range $(-1, 1)$ or $(0, 1)$, and backpropagation computes gradients by the chain rule. This has the effect of multiplying n of these small numbers to compute gradients of the layers close to input layer in an n-layer network, meaning that the gradient (error signal) decreases exponentially with n and the front layers train very slowly. Many new network structures have been brought up to tackle those problems

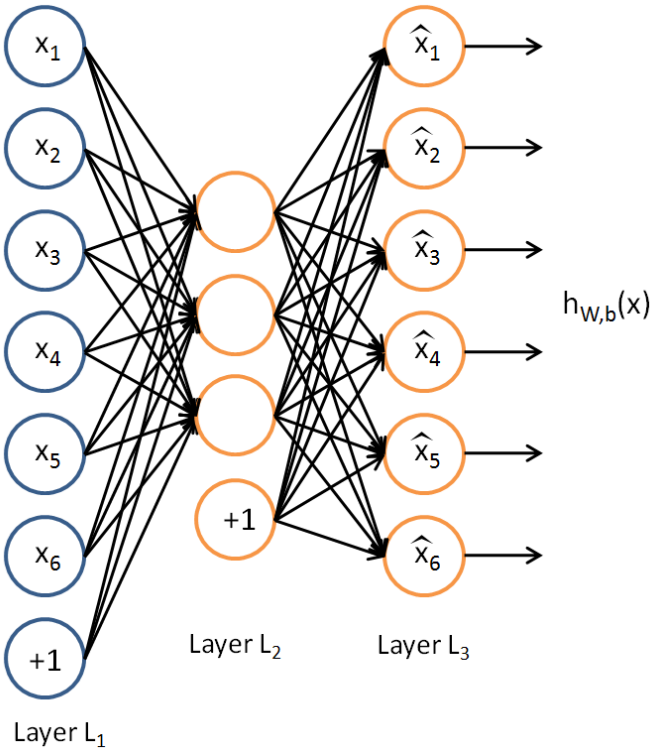


Figure 3.13: One auto-encoder

like autoencoder or restricted Boltzman Machines.

Auto-encoder (Bengio, 2009) is a technique that works as an unsupervised learning method. Figure 3.13 shows one basic configuration of auto-encoder. For a basic autoencoder, there are three layers, one input, one hidden and one output as in the figure L_1, L_2 and L_3. The output $L3$ is the same or almost same as input $L1$, so the network is trying to learn how to encode itself by give input and use the input as output. Many autoencoders like this can be stacked together to generate a deep structure. If many autoencoders stacked together, each layer is trained as an autoencoder by minimizing the error in reconstructing its input. Once the first k layers are trained, we can train the $k + 1^{th}$ layer because we can now compute the code or latent representation from the layer below. The fact that the output is the same as input and the way that hidden layer stacked and trained makes autoencoder a good way to do unsupervised learning to automatically extract features.

Convolutional Neural Networks (CNNs) (Krizhevsky *et al.*, 2012) are another

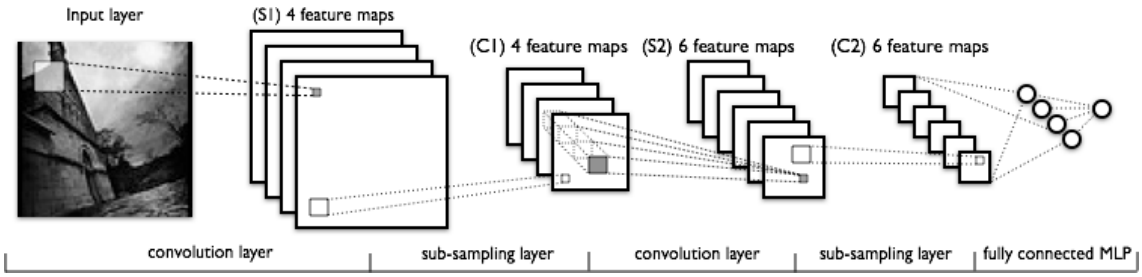


Figure 3.14: Convolutional Neural Network

popular neural network with deep structure which is growing in terms of widespread used in the computer vision community. A convolutional layer and a max pooling layer are two features of CNNs. Unlike a fully connected layer, only a small region of the input data are fully connected with each node in the convolutional layer. Using this small region or local receptive field as a convolutional kernel enables a CNN to capture local features. Partially connected layers also greatly reduce the number of weights to be trained and this brings benefit to performance.

Another concept used in CNNs is that of max-pooling, which is a form of non-linear down-sampling. Max-pooling partitions the input data into a set of non-overlapping sub-sets and, for each such sub-region or sub-set, outputs the maximum value. Figure 3.14 shows a CNN network with 2D convolutional layers and max pooling layers. In the example the input is a grey-scale 2D matrix image. S1 is a convolutional layer with 4 convolutional kernels, 4 kernels correspond to 4 feature maps, and the size of each feature map is related to strides defined to move convolutional kernel. C1 is the maxpooling layer to reduce the size of each feature map. Each feature map is divided into non-overlapping blocks, the maximum value of each block forms the maxpooling layer. S2 and C2 are the convolutional layer and the maxpooling layer, C2 is then fully connected with next hidden layer in order to create a softmax classifier.

A Recurrent Neural Network (RNN) (Cruse, 2006) is normally used to train sequence data in applications like machine translation, image restoration or language modelling (Mikolov *et al.*, 2010). RNN is a type of neural network in which con-

nections between units form a directed cycle. This creates an internal state of the network which allows it to exhibit dynamic temporal behavior. In an RNN, the input and output of the previous step will be fed into a cell along with the current input in order to generate the current output. Such a mechanism enables the network to encode the temporal correlation between current data and previous data.

A feedforward neural network like autoencoder or CNN can only input data with a fixed length, but RNNs can use their internal memory to process arbitrary sequences of input. The fact that an RNN does not require its every input sample to have same length makes it popular in speech recognition and modeling language. For instance, in order to build a machine translate model, the first input could be a sentence with 10 words and the next sentence could be 20 words, which could not be processed by CNN or autoencoder, but it's applicable for RNN. Vanilla RNN is not good at encoding long range correlations also known as long-term memory which is why Long Short Term Memory (LSTM) and its variations are widely used to improve long range encoding by using different types of gates.

LSTM is a RNN architecture published by Hochreiter and Schmidhuber (1997). Figure 3.15 shows the connected three cells and the internal structure of a cell of LSTM. The key to LSTMs is the cell state, the horizontal line running through the top of the diagram. The cell state works like a conveyor belt for information. It runs straight down the entire chain, with little linear interactions. The linear interactions come from structures called gates. Gates are a way to optionally let information through. They are composed of a sigmoid neural net layer and a pointwise multiplication operation. For example the output of the sigmoid layer σ is between zero and one, describing how much of each component should be let through. A value of 0 means letting nothing through, in other words forgetting, while a value of 1 means letting everything through which is equal to remembering the value. The LSTM showen in the figure has three of these gates, to protect and control the cell state.

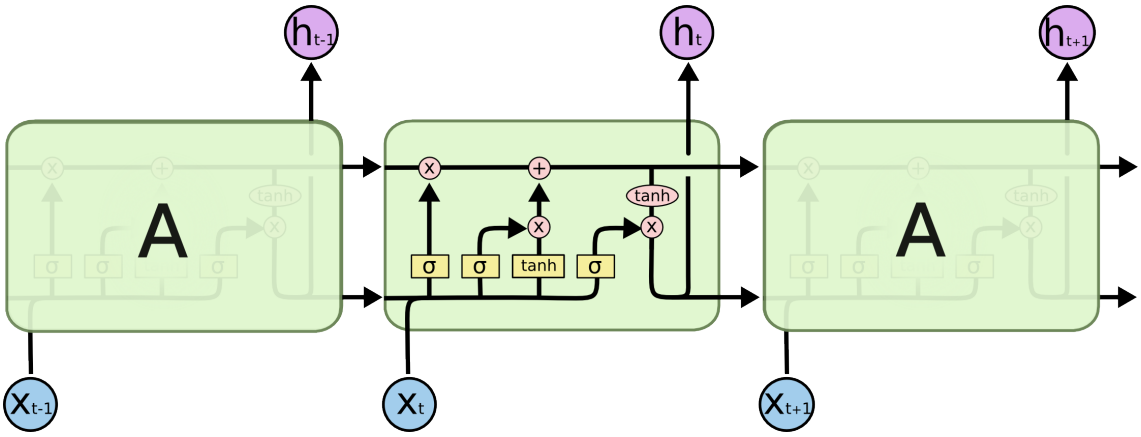


Figure 3.15: LSTM

3.3.2 Classification Based on Features vs. Based on Raw data

Automatic classification using data from a wearable accelerometer has many applications. Mathie *et al.* (2004) used the data from a triaxial accelerometer to classifier daily movements using a pre-defined binary decision tree. Their experiment was done in controlled lab environment while Choi *et al.* (2011) also conducted experiments in a lab to validate classification algorithms for wear and non-wear time based on statistical analysis.

Despite the success of deep neural networks on some tasks, not many papers have yet explored the application of deep neural networks which work good in computer vision, on accelerometer data. We will implement a deep neural network approach on the classification task on accelerometer data and the result will be presented later.

As we have seen earlier, accelerometer data is actually quite noisy and may have many gaps and missing days or even days which might not contain useful information. Such days normally have low information entropy, so we manually examined entropy curves for each subject and determined an entropy threshold for each subject. This corresponds to a personalised threshold of activity level, per subject. We then only use days which have an entropy value bigger than the

Subject ID	101	102	103	104	105	106	108	109	111	112
Entropy threshold (nat)	7	6	3	6	5	6	7	5	6	7

Table 3.1: Entropy threshold for 10 subjects

threshold, to yield our training set and testing set. Table 3.1 shows the entropy threshold for only 10 subjects due to the space. The entropy threshold range from 3 to 7 with mean 5.932 and standard deviation 1.12.

Entropy represents how much information is being carried in the data. When the entropy line goes down, it means the accelerometer data carries less or minor information, which does not contribute much to our knowledge of the subject. After we have chosen only days with high entropy values, we get more than 1,000 days of lifelog data as our input. Our next experiment will try to identify each subject using their own longitudinal accelerometer data which will be used as one data entry. The label of the data entry is the subject who generated the data entry. In total we have 25 subjects. Note that we only use data of a day exceeding a certain entropy threshold value as stated in Table 3.1. Since our data is sampled at 40Hz, each day of data will have 3,456,000 data points. Since some of the machine learning algorithms need to load all their training data in to memory, such as for example an SVM, in order to calculate distance from one point to the rest of the points, we need to reduce the dimensionality of the data. The memory needed to load all the training data corresponds to more than 30 GB and since the computer we use to do the experiments is only equipped with 8GB memory this would cause a lot of page swopping and thrashing, which will increase computation time. In order to reduce the dimensionality, for every 100 non-overlapping points we calculate a mean and a standard deviation for each X,Y,Z axis as two features to represent those 100 points. The reason to choose downsampling rate between 1/2 to 1/3 Hz is because we suspect downsampling less than this wouldn't be enough to capture enough information. While down to 1 Hz or more would still be time consuming for computers to calculate classifiers. Also, we need to split the data into training as

well as testing sets. In the end, this yields a training set of size 862 × 207,360 and a testing set of size 292 × 207,360. this is calculated as 862 days for training and 292 days for testing and each day has extracted features of 207,360 dimensions. Please note that raw data can be only used in neural networks, because Neural Networks (NN)s do not require all their training data to be loaded to memory at once. Also the data we have is imbalanced across subjects since not every subject recorded the same length of lifelog data. Some subjects have collected data for an extremely short period, like a couple of days while most others are of the order of several weeks. We ran several machine learning algorithms to see if those subjects can be identified from their accelerometer data as described in the next subsection.

3.3.3 Experiments

We applied SVM (C=1) with a linear kernel, Gaussian Naïve Bayes (glsgnb), Decision Trees (Decision tree (DT)) and Random Forest Random forest (RF), to the task of classifying users from accelerometer data. Each of the classifiers used default settings. We also constructed a perceptron with two hidden layers where each layer had 100 and 100 hidden nodes. The drop-out rates between each layer are 0.2, 0.5 and 0.5 respectively.

We also constructed a 1D Convolutional Neural Network (glscnn) using the following settings. First, the convolutional layer has 500 filters with size 200 and stride 80, without dropout. This is followed by a max pooling layer with each region of size 10, and then another 1D convolutional layer with 150 nodes with size 100 and stride 10. This is followed by a max pooling layer with size of 5. The last pooling layer is fully connected with next hidden layer with 50 nodes and a dropout rate 0.5. The current neural network settings are taken from the best performance in three different settings.

Long-Short Term Memory (LSTM) is a popular type of recurrent neural network

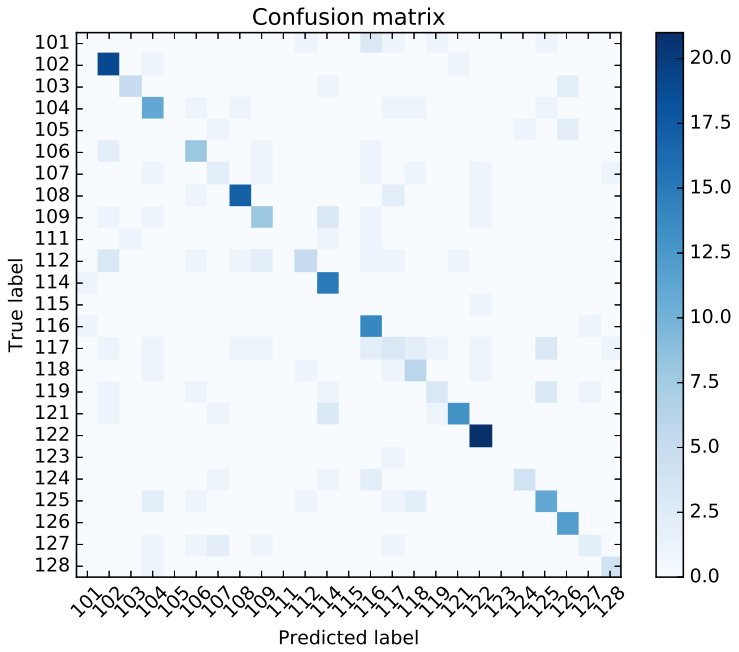


Figure 3.16: Visualization of confusion Matrix for SVM with linear kernel.

and works well when modelling sequences like sentences, or DNA sequences. In order to make LSTM work better for our application we need to reorganize our training samples. If we consider each data point from each day as a sentence of our life, then our sentence of life has more than 200,000 words, which is not possible to train on our lab computer due to hardware restrictions. So we reorganize our data to 120 "chunks" and each of the chunks has a vector with 1,278 dimensions. The size of the training set then become 862 $\times$ 120 $\times$ 1,278, which is much more manageable. All the neural networks we implemented are trained using a hardware accelerator, an NVIDIA GTX 660 Ti with 2GB memory which is compatible with Compute Unified Device Architecture (CUDA). CUDA[6] is a library which accelerates computation by parallelization using a graphic card. Each network is trained with 100 epochs.

Figures 3.16 to 3.22 show the confusion matrices for each machine learning algorithm we implemented, for each subject. On Figure 3.16, 3.19 and all neural networks we can clearly see a strong line on the diagonal, which means the machine

[6]https://developer.nvidia.com/about-cuda

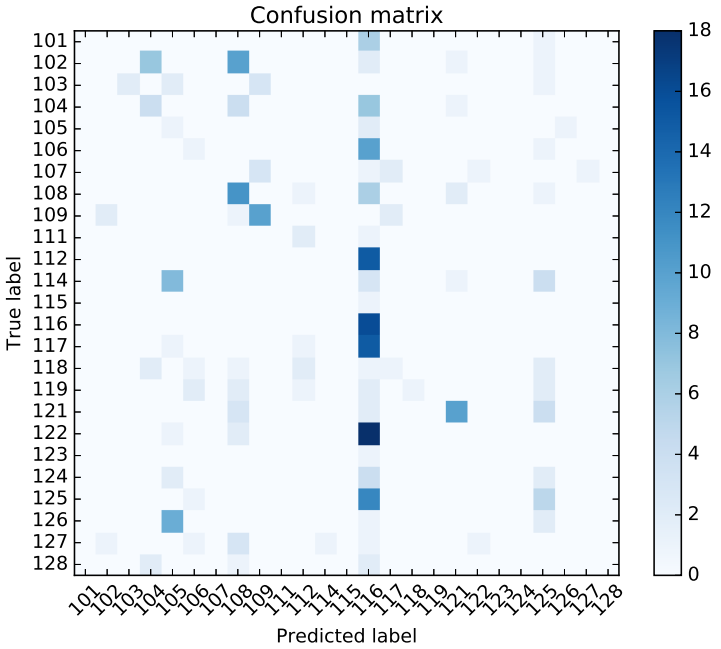


Figure 3.17: Visualization of confusion Matrix for Gaussian Naïve Bayes

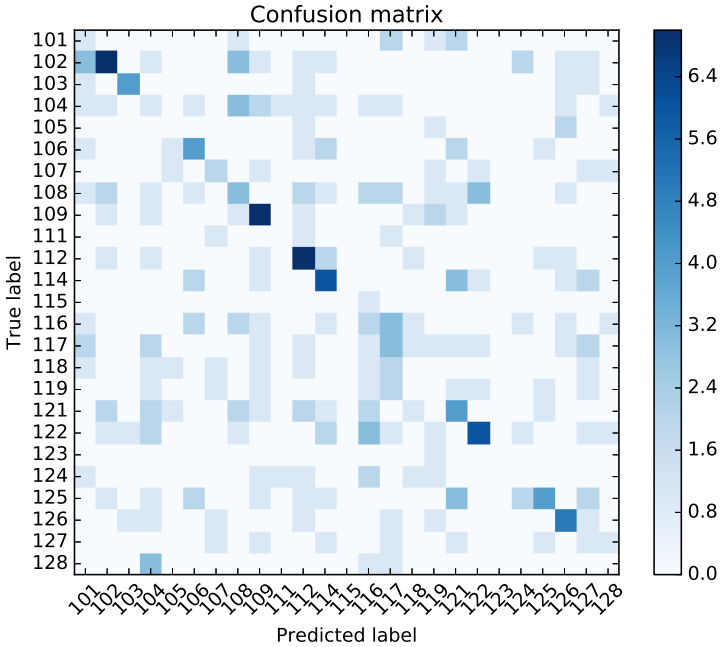


Figure 3.18: Visualization of confusion Matrix for Decision Tree

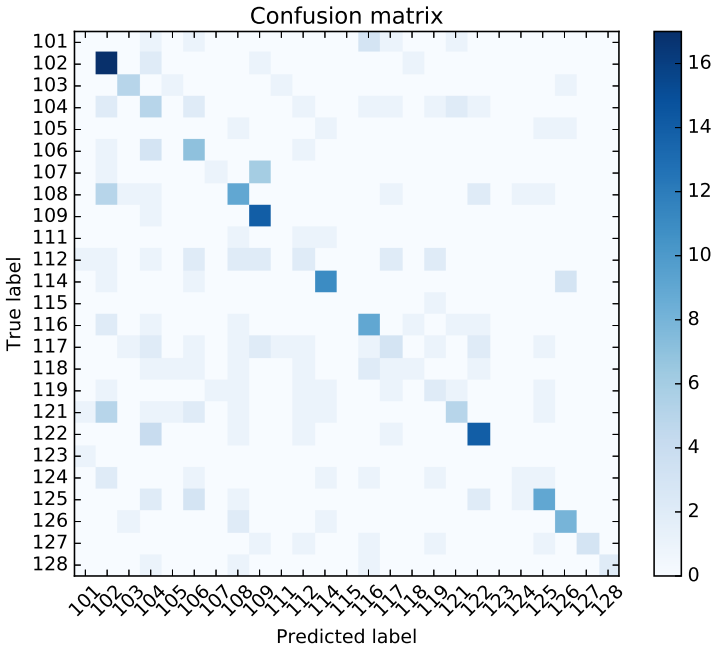


Figure 3.19: Visualization of confusion Matrix for Random Forest

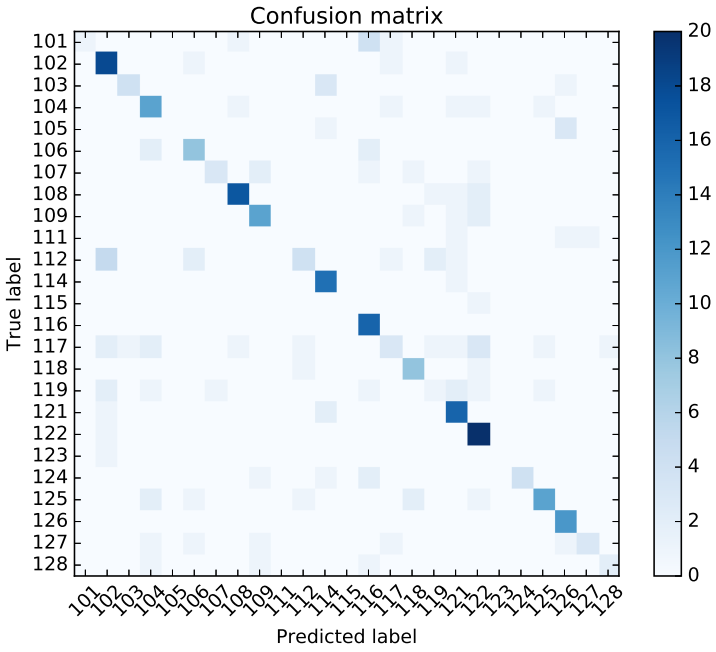


Figure 3.20: Visualization of confusion Matrix for MLP

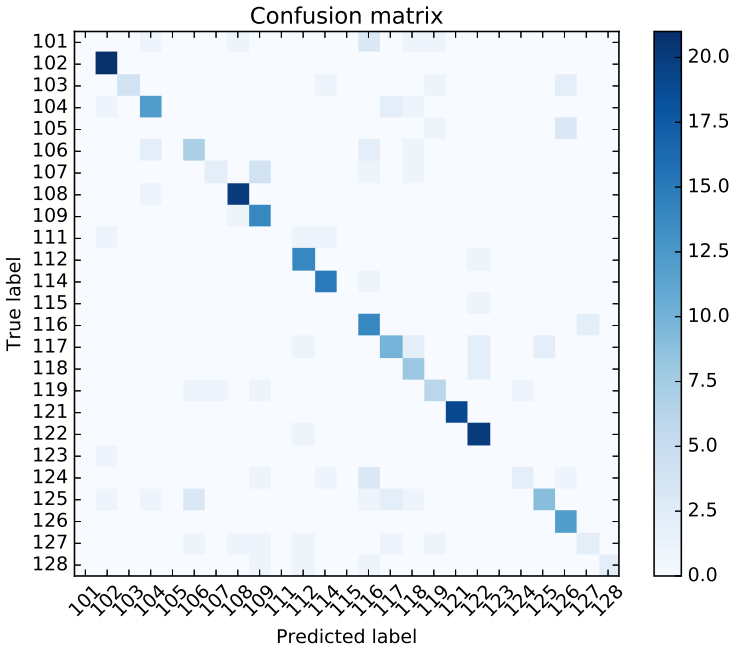


Figure 3.21: Visualization of confusion Matrix for CNN

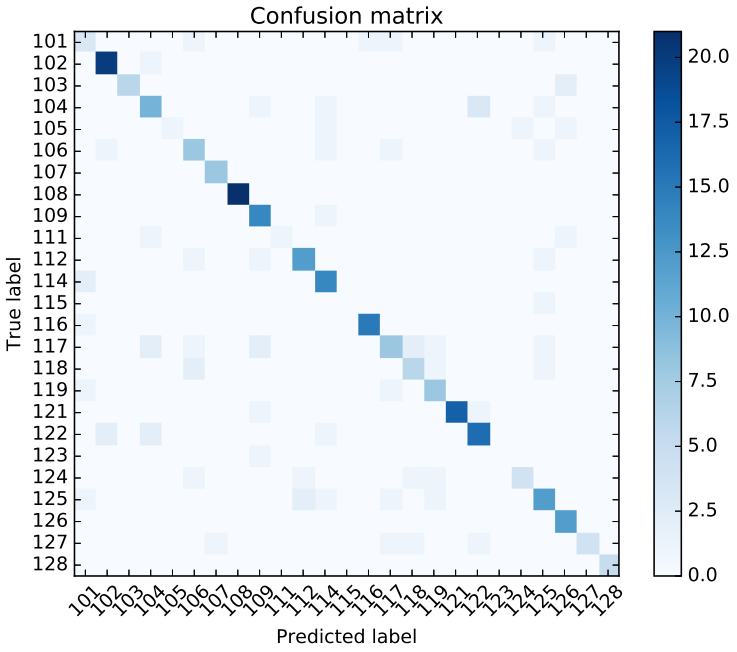


Figure 3.22: Visualization of confusion Matrix for LSTM

learning algorithms do recognize most of our subjects when given testing data. On Figure 3.17, Gaussian Naïve Bayes (Gaussian Naïve Bayes (GNB)) can be seen to recognize most of the subjects as subjects 116. Decision tree (in Figure 3.18) works much worse than Random forest (Figure 3.19). We can see the subjects who are mis-classified against other different subjects, but the confusion matrix from random forest is much clearer than that from decision tree, indicating there are fewer pairwise confusions.

In terms of the confusion matrices for the three neural network approaches, LSTM seems to work best. This is actually not surprising because LSTM is designed to model sequences, which is the nature of our data.

	Precision	Recall	F1-score
SVM	0.59	0.63	0.59
GNB	0.16	0.21	0.15
DT	0.24	0.23	0.23
RF	0.44	0.44	0.41
MLP	0.63	0.64	0.6
CNN	0.7	0.73	0.69
LSTM	0.77	0.77	0.76

Table 3.2: Evaluation of performance of different classifiers

Table 3.2 shows the average performance of each classifier which we implemented. From the Table we can see that LSTM works best among all the classifiers and that CNN, MLP and SVM with a linear kernel also work well. Since we have collected only a limited number of samples, the performances of different classifiers might vary within an interval. In the experiment, the capability of CNN and LSTM to model data has over-performed the rest of the machine learning algorithms. In CNN, filters moving at the convolutional layer work the same way as we examine the data from one small part to another, so that local features are kept and remain connected. A fully connected layer could aggregate those locally connected features together to extract global features. As for LSTM, or the other variation of recurrent neural network approaches, these would also work very well on time series data. If we consider

the application of recurrent neural networks to applications like machine translation, of which input and output are sequences, time series data like accelerometer data that we are dealing with shares some similar properties with sentences in language processing problems, and this might offer a reason for the performances we observe.

3.4 Conclusions and RQ1

In this chapter, we have examined wrist-worn accelerometer data collected from 25 subjects over extended, but varying periods of time on a 24 hours/day basis. Data distribution of 2D histograms from all subjects shows that accelerometer data are mostly bounded in a circle which has a radius of one gravity (1G). This means that most of our subjects' activities are of such magnitude that they are within positive and negative one gravity. If we use a 3D histogram to visualise the aggregated data, then most our subjects' activities are bounded by a sphere with a radius which is also one gravity.

We calculated the difference between different subjects and their accelerometer data (i.e. their activities) using several metrics. Different metrics do show the difference between each subject, but the metrics themselves have some specialization. For example squared Euclidean distance tends to show the most different person, cosine similarity shows that everyone is very different, while KL divergence is somewhere in-between. For this reason, we chose KL divergence for the subsequent calculations, because the result of KL divergence is quite balanced in that it does not exaggerate the difference nor ignore any small discrepancies.

As part of our analysis of this data we split the accelerometer data into days, and computed entropy and KL distance between the current day and previous day, and the KL distance between each day and the summation of distribution of each subject. In this experiment we saw some days with low entropy which means low information carried in those days. We do see changes of entropy across each day or

changes of distribution between one and the previous day. Through those changes we hypothesize that some days can be dismissed due to their low entropy value. If we are going to use those days to identify a subject from among others, then it might interfere with the training and test process, so we choose only days in which entropy is higher than a certain threshold.

We then generated the training set and the testing set. Input data is the data from a day, and the output or label is the enumerated subject. For every 100 points in the input data, the mean and standard deviation is computed and used as features. Several machine learning algorithms were used to train classifiers to identify subjects. Our results showed that classifiers based on neural networks have higher performance than the others and for all cases, the result is better than random guess. We can conclude that each subject will generate data with some features that makes it identifiable and can be detected and captured by computer algorithms.

To summarize, we can simply say that everyone, and the data they generate for their own lifelogs, is different. What's more a more or less bijective/ one-to-one correspondence between subjects and their data has been proven to exist in this thesis chapter. The uniqueness of subjects are represented in the data they generated, and the ability of capturing this uniqueness can be improved by improving algorithms to find better mapping functions between subjects and data. The uniqueness of each subject and their data thus provides a case for requiring individualized analysis.

Chapter 4

Predicability

In previous chapters we used machine learning to discover a one-to-one match between individual subjects and longitudinal population-based activity data and we were able to identify individuals from within the (small) population. Based on the fact that this shows the uniqueness of a subject is represented in the data we collected, an interesting question is now raised: is it possible to use the uniqueness from within the longitudinal data in order to model lifelog data sets. If computer algorithms are able to create generative models based on previous data by using the uniqueness in the data, it is reasonable to hypothesise that there are patterns existing in the data, and the patterns can be perceived by algorithms. If such models exist, it could facilitate applications such as subject differentiation, abnormality detection or recommendation. In this chapter we further explore lifelog data and its possible application by using different methods to potentially model longitudinal data.

If we are able to build a model for longitudinal lifelog data for each individual subject, we would have the following benefits: 1) predicting future trends 2) filling missing gaps for current data. The first benefits would be applicable in the scenario where long-term monitoring is needed, such as when caring for people with dementia. Collected longitudinal lifelog data can be modeled and thereafter used to predict

future trends of development by measuring for instance activity level, sleep quality or heart rate. That is to say we could build a model that we then can use to give predictions of future response based on current interventions. The second benefit would facilitate us with the data cleaning process, since lifelog data are usually quite messy and have many gaps during the collection periods, especially when there is a long period during which no data is collected. On the other hand, the longer the whole collection duration is, the more gaps that we will likely have. We can use a built model using the data we have collected to fill in whatever data gaps with parameters like the length (duration) of gaps or our confidence of the prediction result and so on. Although it cannot be used to reliably fill every gap in the data, it is a considerable approach to improving the integrity of the collected data. Even if it cannot solve the data cleaning problem entirely, it still provides a perspective to tackle the challenge.

In order to show how regular and predictable lifelog data can be and how we can model lifelog data especially longitudinal data, in this chapter, we will firstly state our regression problem in section 4.1, and the approach we will be using in our approach to the regression problem. In sections 4.2, 4.3 and 4.4 several popular models are introduced. These models have been used in some other research fields such as computer vision, language processing and finance to model data or to solve regressing problems. In the experimental part of the chapter which is section 4.6, two data sets will be used to test our hypothesis, that is: are we able to build a generative model to predict lifelog data based on data we already have. In the experiments, we build a general predictor for every subject, i.e. not each individual one, that we are going to build, although building a predictor for each subject might improve the performance.

4.1 Regression

To predict values in lifelog data, we need to explore those approaches which can handle continuous data as output because that is the nature of lifelogging, it needs to be continuous. In order to unify the entire thesis, we denote using the same notations as in previous chapters. Assume there exists a function $g : X \rightarrow Y$ that can map input instance $x \in X$ to an output label $y \in Y$. If Y is discrete, it is defined as a classification problem, otherwise Y is continuous, and this is a regression problem instead. More specific in our regression problem, there are two approaches to describe our effort to model lifelog data.

1. Mapping function g is trying to model a correlation between absolute time and data. That is to say input data is one dimensional time and output is values of acceleration of axis X and/or Y and/or Z in our case. There are some methods such as local weighted regression to tackle this type of requirement but local weighted regression would be utterly impractical as our lifelog data is longitudinal and is likely to be extended over time.

2. Mapping function g is trying the model a relationship between relative time and data. In this case, input data is observed as previous data and output is the current data for example. In other words, we are using data from x_{t-n} to x_{t-1} (where $n < 1$) to predict x_t. The application of such temporal information could capture temporal and dynamic information such as seasonality, or periodicity. It is because periodicity or seasonality appears in the lifelog data, we find that the repeating cycle of a season or a period is stable and with minor changes. The occurrence of patterns could re-enforce the model and thus we need to learn parameters to fit the periodic signal.

Since the capability of the second approach has those mentioned properties, that seems to be appropriate for our data. Thus this is the type of model that we will

use to capture the relative correlation between current and previous data. Using this approach, the length of data that will be used to used as input of the model in our ASU data for example, introduced earlier in Section 2.5.3, is one day, so we are trying to build a model based on the one day data. We will then predict future data based on the model we learned and then evaluate the model using ground truth.

4.2 Time Series Models

Time series analysis has been widely used to model sets of observations generated sequentially in time. Many sets of data appear as time series including weekly lottery sales (Farrell *et al.*, 1999), hourly observed temperature of a meteorological station (Easterling and Peterson, 1995) or GDP per capita of a country (Perron, 1997). Lifelogging researchers concentrate on multi-platform data collected from multiple wearable sensors, and on event detection with little investigation into data from a longitudinal point of view. Li *et al.* (2013c) applied time series analysis techniques such as cross correlation matrix and Wavelet transforms to investigate how Eigenvector and Eigenvalues calculated from a correlation matrix in different scales can be used to detect statistically significant events Li *et al.* (2013a,b). Li *et al.* (2014) introduced random matrix theory to further analyze cross correlations across whole data streams. Yet in lifelogging, the data logged is not always complete and might have gaps of some hours, a day or several days. It is known that filling missing data gaps would greatly improve the performance of any form of data processing. Thus it is vitally important to clean data by predicting missing data using time series or machine learning techniques. For stationary processes with uni-variate data there exist Autoregressive Models (Autoregressive (AR)) and Moving Average (Moving-average (MA)) models and even a mixture of the two, which we introduce later in this section, and all these Autoregression Models model time series by computing parameters which map previous values to current ones based on some assumptions.

In other words, having previous data, can we estimate the parameters of the model and therefore predict missing values.

4.2.1 Autoregressive Models (AR)

AR models could be helpful for some types of time series data. Compared to regression, this stochastic model tries to express current values using a finite, linear aggregate of previous values of the process and a random noise a_t. The AR model uses the following equation:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \cdots + \phi_p \tilde{z}_{t-p} + a_t$$

where $\tilde{z}_t = z_t - \mu$ which is the series of deviations from μ. $t, t_1, t_2, \cdots$ are equally spaced time intervals. We define B as a lag operator $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ and so the model could be written in the form:

$$\phi(B)\tilde{z}_t = a_t$$

4.2.2 Moving Average Models (MA)

Another kind of model, MA, represents observed time sequences by using a linear combination of current and previous random 'shock(s)'. Thus,

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_p a_{t-p}$$

is called an MA process of order q. We define a moving average operator of order q by $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ and so a moving average model my be written as:

$$\tilde{z}_t = \theta(B) a_t$$

4.2.3 Mixed Autoregressive-Moving Average Models (ARMA)

In order to better fit time series using AR models, we can combine AR and MA thus:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \cdots + \phi_p \tilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_p a_{t-p}$$

or

$$\phi(B)\tilde{z}_t = \theta(B)a_t$$

For this, we need to estimate parameters μ,$\phi_1, \cdots, \phi_p$, and $\theta_1, \cdots, \theta_q$ which is the variance of white noise from data. It has been proven in economics that adequate models like autoregressive, moving average or mixed can be used to successfully model stationary time series.

4.2.4 Vector Autoregressive (VAR)

Multivariate time series analysis is the study of statistical models and methods of analysis that describe relationships among several time series. To fully utilize the lifelogged data and correlation between each variable or sensor stream recored, VAR (Hamilton, 1994) need to be introduced. In previous uni-variate AR models, each random variable $\tilde{z}_t$ only corresponds to one value, but for multi-modal datasets the random variable would contains several elements, so that $\tilde{z}_t$ becomes a random vector in which each element is a random variable. The VAR model for a stationary multi-variate time series, with a zero mean vector, represented by $x_t = (x_{1t}, x_{2t}, \cdots, x_{nt})$ is of the form

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}$$

where

- x_t is $n \times 1$ column vectors, and a_t represents multivariate white noise. ϕ_p and θ_q are $n \times n$ parameter matrix.

- $E[a_t] = 0$, $E[a_t a'_{t-k}] = 0$ when $k \neq 0$, $E[a_t a'_{t-k}] = \sum_a$ when $k = 0$ where $\sum_a$ is the covariance matrix of a_t.

4.3 SVM regression

A support vector machine (SVM), so-called "traditional" machine learning approach, can be also used for regression purposes. The difference between an SVM used for classification and used for regression is originated from the use of different loss functions. The classification task uses a so-called hinge loss function where for an intended output $\hat{y} = \pm 1$ and a classifier score, the hinge loss of the prediction $\hat{y}$ is defined as:

$$\ell(\hat{y}, y) = max(0, 1 - \hat{y} \cdot y)$$

Note that $\hat{y}$ is the predicted label and y is the ground truth label. If the predicted label $\hat{y}$ and the gold standard y have the same sign (meaning y predicts the right class) and $|y| \geq 1$, the hinge loss $\ell(y) = 0$, but when they have opposite signs, $\ell(y)$ increases linearly with y.

On the other hand, a different type of loss function is used in a SVM when it is used for regression such as the ϵ-insensitive loss function described by Müller *et al.* (1997). The details of comparison between these loss functions can be seen in the work by Suykens and Vandewalle (1999). The equation of ϵ-insensitive loss function is:

$$\ell(\hat{y}, y) = \begin{cases} 0 & \text{if } |y - \hat{y}| \leq \epsilon \\ |y - \hat{y}| - \epsilon & \text{otherwise} \end{cases}$$

4.4 Neural networks (RNN/LSTM)

The idea behind a Recurrent Neural Network was introduced in the previous chapter in section 3.3.1.2. To explain it again briefly, this type of network would share all the weight of the network during the propagation through time. The key element of this type of network is a cell which contains different types of gate to control flow of information. In the previous chapter we worked on applying a neural network to a classification task, but in this chapter we will be using LSTM to solve a regression problem. The way to achieve this is to modify the loss function which is used to compute the loss between ground truth and output of forward propagation and the loss can then be back propagated through time using chain rules. In a classification task, the loss function we used is categorical cross entropy. The cross entropy between two probability distributions measures the average number of bits needed to identify an event from a set of possibilities, if a coding scheme is used based on a given probability distribution $\hat{Y}$, rather than the true distribution Y. The equation to calculate categorical cross entropy of two distributions is:

$$H(Y, \hat{Y}) = -\sum_{x} Y(x) \log \hat{Y}(x)$$

In a regression task, we just need to change this loss function to continuous error instead of categorical type of error, such as root square error or square error. Here we choose square error:

$$L(Y, \hat{Y}) = (Y - \hat{Y})^2$$

4.5 Evaluations

Evaluation plays a very important part in classification an/or regression problems. Because different evaluation metrics can tell us the discrepancy among different approaches we will now briefly introduce several popular evaluation metrics that we

will use in our experiments.

4.5.1 RMSE

RMSE is defined literally as the square root of the mean square error. The mean square error calculates the absolute difference between predicted values and ground truth. It measures the average of the squares of the errors or deviations, that is, the difference between the estimator and what is estimated. The equation to calculate the RMSE is as follows:

$$RMSE = \sqrt{MSE}$$

$$MSE = \frac{1}{n}\sum_{i}(\hat{y}_i - y_i)^2$$

CCC measures the absolute difference between two random variables. If CCC is 0, in our case the predicted time series would be the exact same match as the ground truth.

4.5.2 Correlation Coefficient

Correlation Coefficient (CC) is also known as Pearson product-moment correlation coefficient and is a measure of the linear correlation between two variables X and Y. The values of correlation coefficient range from -1 to $+1$, where $+1$ is total positive correlation, 0 is no correlation, and -1 is total negative correlation. The equation to calculate CC is:

$$\rho_{\hat{Y},Y} = \frac{cov(\hat{Y}, Y)}{\sigma_{\hat{Y}}\sigma_Y} = \frac{E[(\hat{Y} - \mu_{\hat{Y}})(Y - \mu_Y)]}{\sigma_{\hat{Y}}\sigma_Y}$$

CC is widely used to measure the degree of linear dependence between predicted values and ground truth. If CC is $+1$, which implies fully linear dependence, it does not yet mean the predicted values are exactly the same as the ground truth. For instance, in an extreme example if every predicted value is 10 times bigger than the

actual value, the CC is 1. Or, put another way, if every predicted value is 5 times larger than true labels, the CC would still be 1. Informally speaking, CC is really about measuring the difference of 'shape' of data.

4.5.3 Concordance Correlation Coefficient

Due to limitations of CCC and CC, Concordance Correlation Coefficient (CCC) Lawrence and Lin (1989) was proposed to take into account both absolute and shape differences. As an evaluation measure, Concordance Correlation Coefficient (CCC) combines the Pearson's correlation coefficient (CC) with the square difference between the mean of the two compared time series:

$$\rho_c = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2 + (\mu_x - \mu_y)^2}$$

where ρ is the Pearson correlation coefficient between two time series in our case it's predicted and ground truth time series, σ_x^2 and σ_y^2 are the variance of each time series, and μ_x and μ_y are the mean value of each. Therefore, predictions that are well correlated with the gold standard but shifted in value are penalised in proportion to the deviation.

4.6 Experiments

A series of experiments were conducted on ASU raw accelerometer data and on athletic data, both described earlier. The reason to use two data sets is to see the predicability on both raw data (ASU) and on higher level data (athletic), since the athletic data consists of detected activities using an accelerometer. We expect to see algorithms that are able to learn data models using previous data and to predict new or missing data.

4.6.1 Using ASU Data

The same data preprocessing is applied to ASU data as was described in Chapter 3 section 3.1. The cleaned data is used to generate training and testing data. We applied fixed training and testing data (which will be rotated into an n-fold cross-validation in the future study). The splitting mechanism is as following:

1. For each subject, data is split into non-overlapping time series with an interval of 1.5 days.

2. The 1.5-day data chunks are randomly shuffled. Each chunk is called a basic data unit.

3. For each subject, 80% of all basic data units are selected as the training set and 20% of all basic data units are selected as the testing set. If the number of testing sets for each subject is less than one basic data unit, then a basic data unit will be chosen randomly.

Randomization of training and testing data is a standard approach in machine learning to evaluate classifiers. We choose 1.5 days is because we are trying to explore the maximum capability of predictive model we think. In this case, we are using one day of data to predicting half day data, so that is 1.5 day in total. We firstly investigate the performance of different regression methods by giving one day data to predict one and then giving one day data to predict many points after.

4.6.1.1 Predicting one point

After the process, we have 576 basic data units in training sets and 157 in testing sets, with each basic data unit of length 1.5 days. The reason to shuffle the order of basic data units at the beginning is to eliminate the situation that subjects keep to a pattern frequently within a period of time. In this case, if the training basic data units are not randomized, there is a chance that the model can only model a

certain pattern, which could constrain the generalization of the model. Due to the limitations of each regression method, each basic data unit has to be reformed for different algorithms.

AR models are not used in our experiments on ASU data since the ASU data has gaps and for the best performance of AR models, we need to use data series without gaps. However we will apply AR-type models in data series with higher semantics because this data is both clean and has no gaps.

For SVM regression and for linear regression, a window size with one day within each basic data unit will be used as input and the next point will be used as a label with a continuous value. We used 40,000 points as stride to move windows and generate new training data within the basic data unit. Since SVM or linear regression needs all data to be fed into the regressor at once, and due to limitations of our computer's memory, raw input data needs to be further reduced. We choose to divide the 1-day input data into non-overlapping 2,000 pieces with each piece having 1,728 data points, and for each piece the mean value is calculated. So at the end of this process, training data input is formed with size 25,344 $\times$ 1,728, and labelled with size 25,344. Please note that our task is regression, so the label here is continuous.

For LSTM, we use two approaches: 1) a similar strategy is used to generate training data. Each axis will be trained as a LSTM 2) An LSTM will be trained using data from three axis. The performance of SVM and linear regression are guaranteed by the underlying theories but the applying real life data with noises would always bring challenges in any kind of machine learning algorithms. As for neural networks, they are normally data driven models. That is to say give enough good quality data, less complicated linear models with deep enough layers can achieve quite good result. Event without considering the noises, the number of samples that are correctly collected often caused issues, such as under-fitting and over-fitting. Also the parameter of those models can be further optimized by grid search to improve

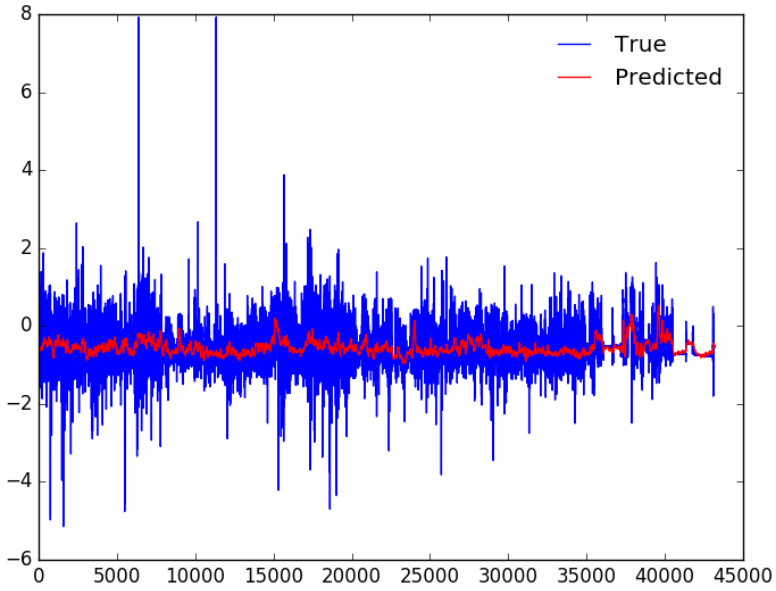


Figure 4.1: One of best results to predict X-axis using SVM regression (RBF kernel) evaluated by CCC. X axis is time. Y axis is acceleration value.

the final performance, but it's a very computational intensive task which cannot be done on our machine with reasonable time.

In the training process, for SVM regression and for linear regression, default settings are used. For LSTM we use a single axis, and the input data uses a mini batch with size 10. One day data is used as input with length 1 and dimension 1,728. For LSTM we use data from all three axises, input data uses a mini batch with size 10 and each data series consisting of a day of each axis is combined and used as input with length 3 and dimension 1,728. The output of the network is three for LSTM using data from all three axis. the LSTM model consists of 2 layers with 50 nodes in each layer. The models saved are the model with least loss achieved on training data.

Figures 4.1 and 4.2 show one of best and one of the worst predicted results compared with ground truth measured by CCC. Figures 4.3 and 4.4 show one of best and one of the worst predicted results compared with ground truth measured

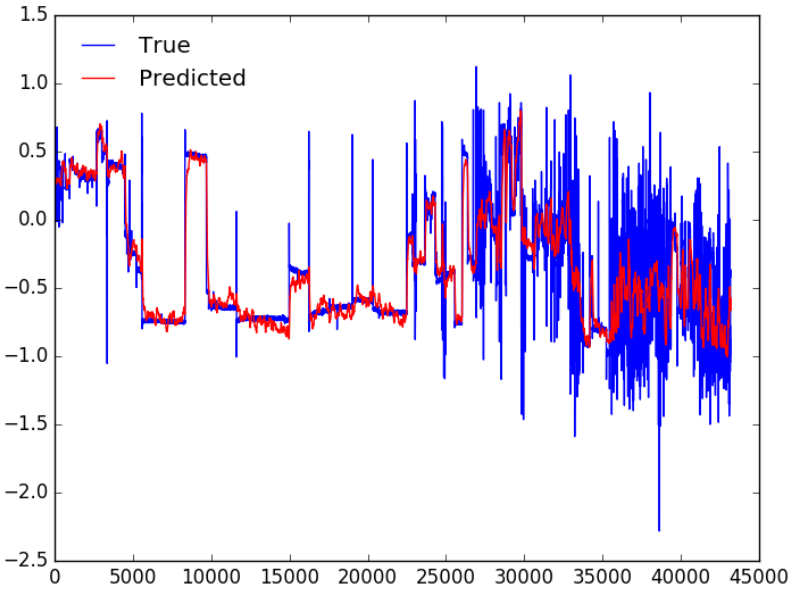


Figure 4.2: One of worst results to predict X-axis using SVM regression (RBF kernel) evaluated by CCC. X axis is time. Y axis is acceleration value.

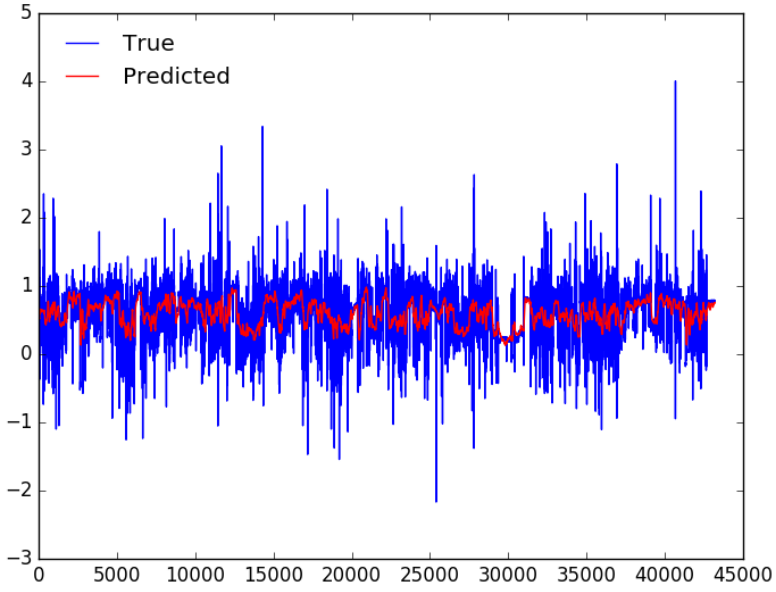


Figure 4.3: One of best results to predict X-axis using SVM regression (RBF kernel) evaluated by CC. X axis is time. Y axis is acceleration value.

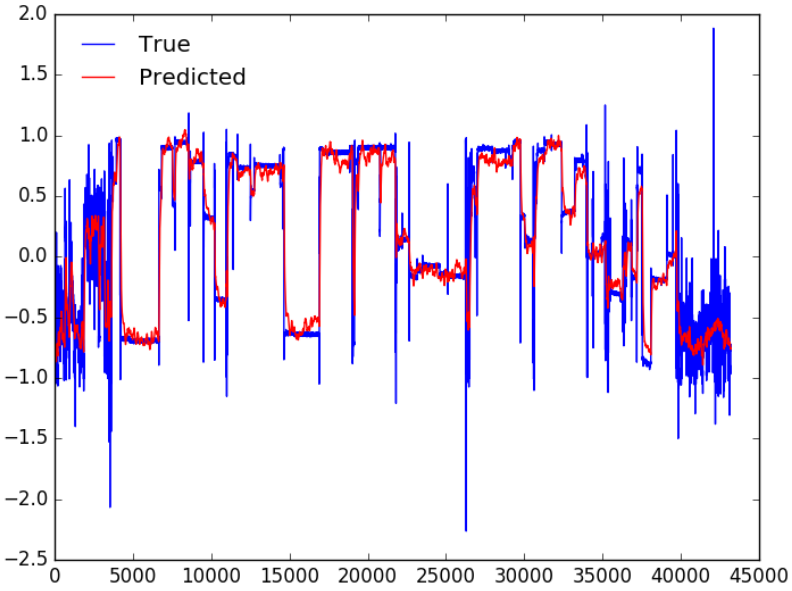


Figure 4.4: One of worst result to predict X-axis using SVM regression (RBF kernel) evaluated by CC. X axis is time. Y axis is acceleration value.

by coefficient correlation. In those figures red lines are predicted values and blue lines are ground truth with 40 times sub-sampling. Since the predictors only give prediction points after one day data, we compute the prediction for the testing set by a 40-point step within a basic data unit in order to generate a time series to be evaluated. In the figures, it can be seen that the predictor can well predict the trend of the time series even in the two worst performing examples evaluated by CCC and CC. We can also see some sudden changes in the time series which would be hard to predict. Evaluation results of different methods are shown in Tables 4.1, 4.2, and 4.3.

Tables 4.1, 4.2, and 4.3 show the performance of different regressors in modelling the data. CCC indicates the error of predicted data series vs. ground truth. The lower the CCC value, the better the result is predicted. CC and CCC show the correlation between predicted data and ground truth, in a range from -1 to $+1$. The higher the value, the more correlation between predicted and true data. Cross

	CCC	CC	CCC
SVM RBF	0.2546	0.8087	0.7832
SVM Linear	0.2836	0.7912	0.7742
Linear Regression	0.2452	0.8192	0.8080
LSTM	0.3696	0.6560	0.6471
LSTM (3 axises)	0.3338	0.7033	0.6955

Table 4.1: Evaluation of performance of different classifiers for X-axis

	CCC	CC	CCC
SVM RBF	0.2786	0.7981	0.7724
SVM Linear	0.2941	0.7921	0.7798
Linear Regression	0.2738	0.8011	0.7872
LSTM	0.4316	0.5949	0.5867
LSTM (3 axises)	0.3922	0.6604	0.6537

Table 4.2: Evaluation of performance of different classifiers for Y-axis

referencing the three tables, some approaches show similar performance on different axis. Linear regression performs best among all the methods, while SVM regression on the second tier and LSTM does not seem to work well. We hypothesise that his phenomenon comes from the task in the experiment. In this experiment a whole day of data with 3,456,000 data points is used to predict the next one point. There is a big chance that the value of a predicted point is close to those points chronologically near it. Linear regression performs best among those methods indicating that there exists a linear correlation between predicted value and the data from the previous day.

LSTM does not work well for this task but what should be noticed is that the LSTM model trained using 3 axis combined works better than separately trained models on each axis. More information about training data thus seems to improve the LSTM performance. The reason that LSTM does not work as well as linear models might be because

1. LSTM is good at modeling sequence-to-sequence relationships. The model we

	CCC	CC	CCC
SVM RBF	0.2746	0.8022	0.7753
SVM Linear	0.2983	0.7802	0.7628
Linear Regression	0.2683	0.8094	0.7967
LSTM	0.4169	0.6170	0.6091
LSTM (3 axises)	0.3753	0.6712	0.6658

Table 4.3: Evaluation of performance of different classifiers for Z-axis

are training is a sequence-to-point;

2. As we mentioned before, chronologically near data points might be more correlated than distant ones. LSTM models long term correlation well but does not model short term correlations as well, especially when there are sudden changes in behaviour within a short period of time;

3. Since the number of samples we have is not large, there are chances that the best model on training data might be overfitting.

There are various of ways to improve the performance of classifier we use. Firstly, we could change the way we generate more training and testing samples such as reduce the step size we choose to generate more data. Secondly cross validation would be considered to train those regressor with less bias. Last but not least we could use grid search to optimize those machine learning algorithms. Those parameters can be different kernel functions in SVM or different combination of layer in neural networks.

4.6.1.2 Predicting many points

Another experiment was conducted to see how well each model is at predicting many points in the data stream instead of just predicting one point. We use trained SVM regression and linear regression models as in the previous experiment but we use a different testing method as we are predicting many points after giving one day of

data and once a new prediction of the next point is generated, the first point in the training data will be removed and we will predict the next point using a sliding window of training data. The newly-formed training data with the predicted next point will be then used to predict the next one, and so on.

For LSTM we will train a new LSTM model whose output is a sequence instead of a value, so that the LSTM model can model the sequence-to-sequence correlation to take full advantages of LSTM's of long term memory. The implementation of multiple output LSTM uses raw accelerator data as input. Firstly, one day's data is used as input and for every non-overlapping 100 points in the input, the mean value is calculated and then the newly-formed data will be resized to 1,728 $\times$ 20 and used as input to LSTM. The output is the raw accelerator data consisting of the next 34,560 points. We evaluate the ability of the model to predict 1,000 points. Due to the computation limitations, we are able to compute only 1,000 points in a reasonable time.

	CCC	CC	CCC
SVM RBF	0.2186	-0.0035	0.0041
SVM Linear	0.2598	-0.0057	-0.0081
Linear Regression	0.2249	-0.0002	-0.0006
LSTM	0.4828	0.0491	0.0107

Table 4.4: Evaluation of performance of different classifiers for Y-axis to predict 1,000 points at once

Table 4.4 shows a less satisfying result compared with predicting just one point. LSTM seems better than the others, but all the results are not promising. Although absolute error looks not that bad, correlation scores indicate that the predicted curve shows almost no correlation with the ground truth. The reasons for this might be

1. For one point prediction escalation, we sampled several windows with one day within a basic data unit, and here in a 1,000-point prediction test, we are only able to use one 1-day window due to limited computational power;

2. LSTM could be better trained with more layers and nodes on a graphics card with bigger memory;

3. The loss function used to train the neural network is using error instead of correlation. In other words the network is trying to lower the square error instead of correlating. It is interesting to see the result using correlation as loss function;

4. The experiment shows the difficulty of predicting multiple points in lifelog data even with state-of-art approaches. Detecting or anticipating sudden changes in lifelog data is still a challenge to model, especially when we are trying to predict many points given only one day's data in advance.

4.6.2 Using Athletic Data

In the previous section we showed how we conducted the regression task on raw accelerometer data but we are able to model the raw accelerometer data with reasonable loss only in the situation when we are predicting one data point at a time. Another interesting perspective is to see if a similar result can be observed wen working at a higher semantic level. In this scenario, raw accelerometer data is converted into different activities such as running, cycling and swimming. We used our Athletic dataset to explain and conduct this next experiment. This dataset was collected from a triathlete who use an wearable device to record daily training for over 10 years. The data recorded consists of distance of running, cycling and swimming every day. This data will be used in the next experiments to see if a model can be built to predict the distances covered for different activities.

SVM regression with a linear kernel and radial basis function (RBF) kernel were used and linear regression was used to model the time series data. For these models, the whole dataset is split into 10 chunks and the first 9 chunks are used for training. The last chunk is used to evaluate trained model. For the training set, we apply

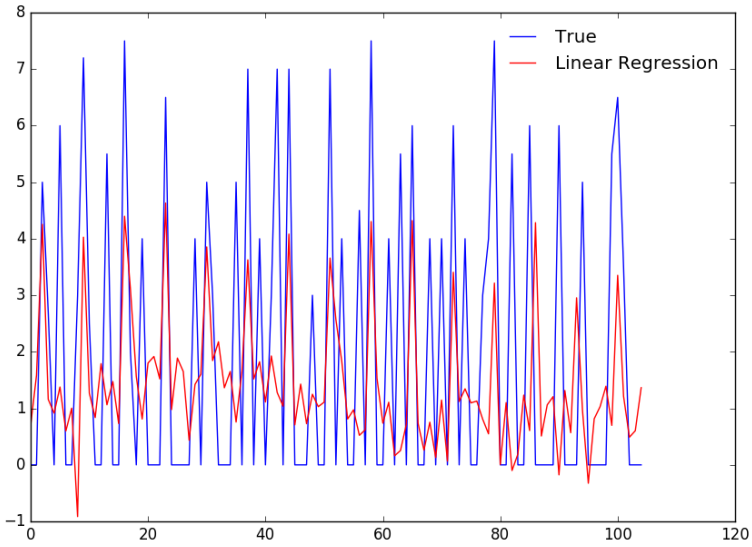


Figure 4.5: Prediction with Linear regression for running data. X axis is time. Y axis is running distance.

a window with size of 360 days, and the next day as the target or output of the models. The window moves forward with a stride of size 1. In the last chunk, first 360 points of data will be used as input for the model, and the next 105 data points will be the prediction to be evaluated. In the experiment, predicted values will be added into the previous input with size 360 as the new input data. Figures 4.5, 4.6, 4.7 show the results of predicted values and the ground truth.

In the figures, we can observe that both Linear regression and SVM regression with linear kernels show better results than SVM regression with an RBF kernel. We assume this is because the data along a chronological order has more linear correlation. The SVM with a linear kernel seems to learn fluctuations better than linear regression while linear regression tends to assume the data are more flat. Both the models have the ability to capture the linear correlations showing the existence of periodicity with the training data.

Vector Autoregression and ARMA models are also used to build predictive models. The data set is once again split into 10 chunks and first 9 used to train VAR

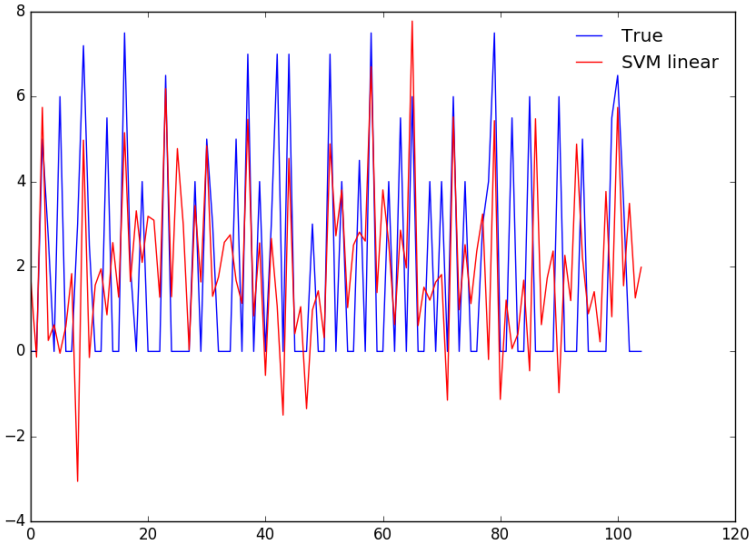


Figure 4.6: Prediction with SVM Linear for running data. X axis is time. Y axis is running distance.

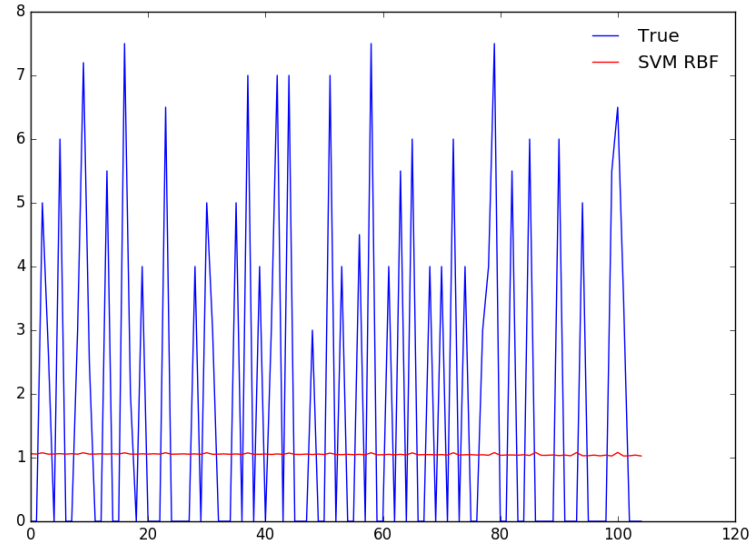


Figure 4.7: Prediction with SVM RBF for running data. X axis is time. Y axis is running distance.

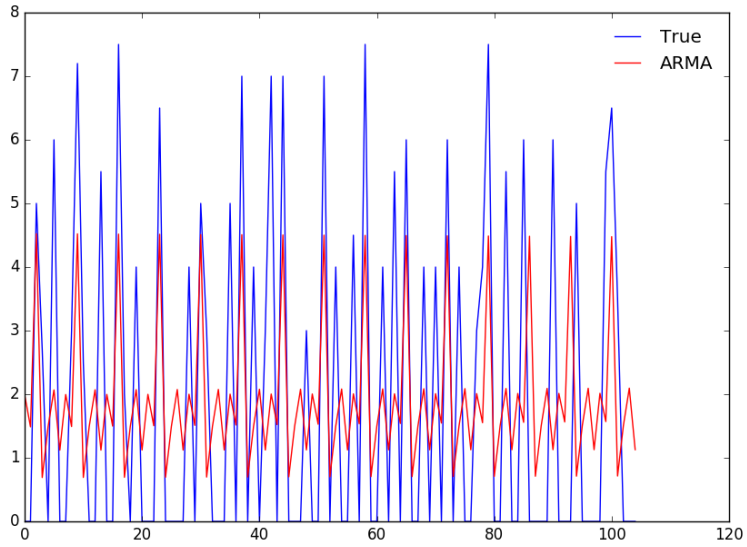


Figure 4.8: Prediction with ARMA for running data. X axis is time. Y axis is running distance.

and ARMA models and the same pieces of time series data as in SVM regression and linear regression will be predicted using VAR and ARMA models.

As we can see the results of VAR and ARMA in Figure 4.8 and 4.9 respectively. The ARMA model tends to generate a more stable model which has an obvious repetitive structure and very little fluctuation, while VAR has more dynamics which fluctuate along the time axis. LSTM is also used to train a sequence-to-sequence model. Figure 4.10 shows the result of LSTM.

Table 4.5 lists the evaluation of different approaches using data for different activities. In the table we can see the approaches have different capability to model and predict activities with higher semantics than just raw accelerometer data. Also, the difficulty for methods to predict each activity varies. Generally speaking, running is easiest to model amongst the three, with cycling in the middle, and swimming is hardest to model and predict. We guess running is more frequent than cycling or swimming, so that more data about running is gathered hence a better model can be generated. The ARMA model has the best performance for running and

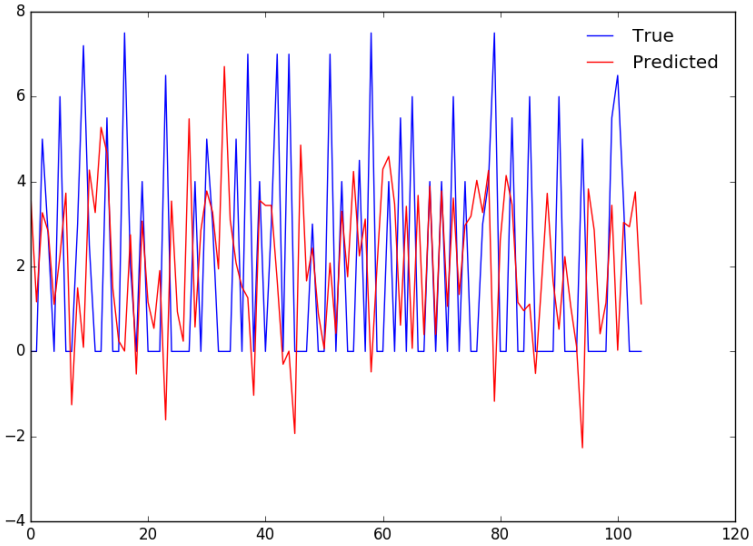


Figure 4.9: Prediction with VAR for running data. X axis is time. Y axis is running distance.

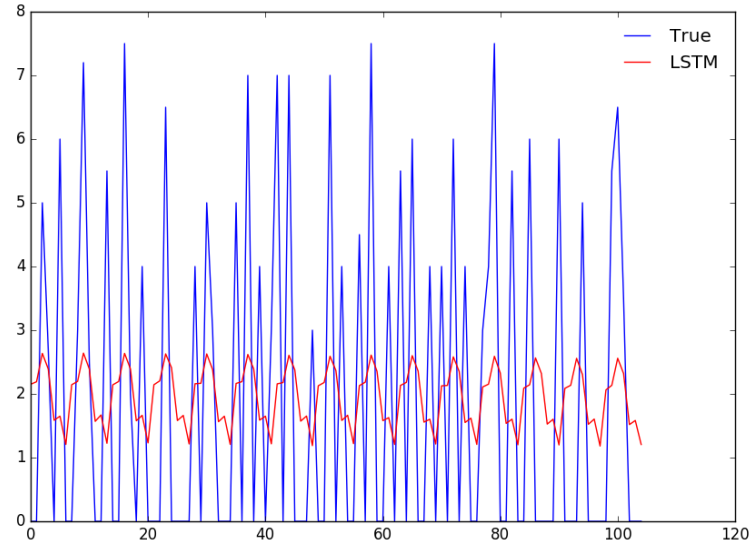


Figure 4.10: Prediction with LSTM for running data. X axis is time. Y axis is running distance.

swimming data, while linear regression works quite well for cycling data. LSTM does not seem to work well on this task. The reasons might be 1) There are many short-term sudden changes in the data which LSTM is not quite good at expressing those sudden short term changes. Meanwhile long-term changes are quite stable for the subject and LSTM tends to model the data in a stable manner. 2) The task has a limited number of training and test samples. We can argue that the ability of LSTM is constrained by the number of samples presented in the network. Normally the corpus for training LSTM in natural language processing tasks are enormous. The details about qualitative analysis of this data set will be described in the case study presented later in Chapter 7.

		RMSE	CC	CCC
Run	SVM Linear	2.5470	0.4089	0.3901
	SVM RBF	2.7859	0.4090	0.0037
	Linear Regression	2.3872	0.4674	0.3350
	VAR	2.8683	0.2294	0.2178
	ARMA	**2.1921**	**0.5684**	**0.4154**
	LSTM	2.4897	0.3819	0.1298
Cycle	SVM Linear	18.9751	0.4191	0.3825
	SVM RBF	19.3863	0.0001	0.0001
	Linear Regression	**15.8767**	**0.4785**	**0.4315**
	VAR	18.3266	0.3567	0.3376
	ARMA	16.8373	0.3852	0.2107
	LSTM	17.9980	0.1082	0.0313
Swim	SVM Linear	35.9941	0.0170	0.0159
	SVM RBF	31.6449	0.0001	0.0001
	Linear Regression	28.0336	0.1796	0.1303
	VAR	32.0294	0.0894	0.0771
	ARMA	**27.8164**	**0.1956**	**0.1347**
	LSTM	28.1985	0.0578	0.0002

Table 4.5: Evaluation of performance of different classifiers for different activities

Despite different approaches showing different predicabilities, one interesting common fact is that several methods such as linear regression, ARMA and even LSTM tend to model the training data as a periodic signal or more precisely as a quasi-periodic signal and those methods tend to have better performance than other

approaches. The predicted periodic properties can be seen in Figure 4.5, 4.8 and 4.10, which raise an interesting question ... if machine tend to 'think' the signal is periodic, is the signal actually periodic ? We will investigate this in the next chapter.

4.7 Conclusion and RQ2

We have shown more or less bijective correspondence between subject and data in the previous chapter. Based on the findings in the previous chapter, we conducted a further analysis which is to see if the uniqueness presented in longitudinal data could potentially be modeled from raw accelerometer data. Different models which could tackle regression problems are used to model or capture the uniqueness of the data and the models are further used to predict future values for both one point and for many points scenarios. RMSE, Coefficient Correlation and Concordance Correlation Confident are used to evaluate how well the data are predicted. In the experiment to predict one point given one-day data, basic linear regression seems to have the best performance, which could be the closeness of chronological order between the point to be predicted and previous points. The result show that all predicted results using different algorithms show quite high correlation with the ground truth. We then conducted experiments to predict more than one point, by adding new predicted values into training data for SVM regression with a linear kernel, SVM regression with an RBF kernel and linear regression. For LSTM in this situation, we implemented a multiple output network. In the example of predicting many points, overall performance of all approaches are not satisfactory by giving low error but almost no correlation, which indicates this is still a tough task even with state-of-the-art approaches.

Experiments were also conducted on data with higher semantic meaning instead of just raw accelerometer data. In this experiment, ARMA models and Linear

Regression seem to work well for running and cycle data. For swimming data, all the methods perform worse than running and cycling data. LSTM does not work best in this task, the reason we hypothesise is because it is less expressive of sudden short term changes and there is a limited amount of training data. For running and cycling data, based on the results of different methods, the data is still quite predictable when choosing suitable models. For swimming data, the result is even worse than for running and cycling data, if appropriate approaches are chosen, the predicted result still shows correlation with the ground truth to some extent. Another interesting finding is that several good performance models tend to model the data with periodic features, which make us raise the question: is the nature of the longitudinal signal actually periodic ?

To summarize, longitudinal lifelog data can be modeled and predicted with suitable models and under several constraints. In the experiment the number of points to be predicted yields very different results. In other words, the uniqueness of longitudinal lifelog data can captured by computer algorithms under the condition of how 'greedy' the model is to be and some of the best performing methods model lifelog data with higher semantics as a quasi-periodic signal.

Chapter 5

Spectrum Estimation

In the previous chapter, we have shown several time series and/or machine learning models that can be used to model longitudinal lifelog data, especially to predict missing data, i.e. the next several points over a short period of time. Interestingly, we found the best-performing models tend to model longitudinal time series as quasi-periodic signals which gives us encouragement to investigate this further. In this chapter, we investigate several spectrum estimators in order to examine if there is/are periodicity/periodicities in longitudinal lifelog data.

Detecting patterns of periodicity would give huge insights and could help to reveal aspects of a person's lifestyle. However, periodicity detection usually relies on data which is both complete and has no missing values, and is accurate with no probabilities associated with values of the data. With lifelogging data, this isn't always the case as people can simply decide not to switch on their logging devices or there can be calibration errors with lifelog sensors. In this chapter we address how to detect repeating patterns of lifestyle from lifelogs when 1) the data is fully logged and 2) the underlying data has missing or incomplete data, or even data which is erroneous. Once such patterns and periodicities have been detected it is beyond the scope of this chapter as to how to use them or present them back to users. The details of presenting the periodicity results back to the participant and/or

illustrating the result of periodicity detection will be introduced in a case study in chapter 7.

To illustrate our work on detecting periodicity from such noisy data we work with real lifelog datasets[7] which have in-built gaps and noise. Our work demonstrates that even with very noisy data which is also far from being continuous, we can detect repeating patterns and periodicities.

Our aim is to detect and report longitudinal patterns in lifelogs which we can regard as a form of time series, and these patterns can be referred to as periodicities. In terms of lifelogging, the periodogram can be used to detect the natural cycles that occur in lifestyle, behaviour, and activities. Periodicity can be observed in many natural phenomena, such as circadian rhythms associated with our sleep, annual seasons and so on. Intuitively, we think of our routine daily lives as composed of various forms of recurring events with obvious periodicities around daily, weekly, monthly, seasonal and annual cycles. In any kind of spectral analysis of a lifelog we expect to see periodicity around these frequencies. However, without the help of lifelogging devices and the resulting lifelog, analyzing the periodicity of human life is not a practical proposition.

5.1 Non-Parametric Methods

Signal processing theory tells us that in order to detect low-level periodicities in any time-series, we calculate its power spectral density (PSD or power spectrum) (Vlachos *et al.*, 2005). The PSD essentially tells us how strong is the expected signal power at each possible frequency of the signal. Because frequency is the inverse of period, we wish to identify frequencies that carry most of the energy and then from that to detect the most dominant periods. Two estimators of the PSD

[7]As scientists our philosophy is always to make our research data openly available to others in the interests of transparently and reproducibility but because this is *personal* data from a personal lifelog we cannot publish this easily.

could be used to detect and present periodicities; the *periodogram* and the *circular autocorrelation* or *full cross correlation*. The power spectral density can be computed using the Discrete Fourier transform (DFT) (Discrete Fourier Transform) or FFT (Fast Fourier Transform). PSD is also called a periodogram and we can detect and visualise periodicity using a periodogram. The periodogram was first proposed by Schuster (1898) and is visualised as a 2D plot with spectral frequencies on the x-axis and the strength of the pattern at each frequency measured on the y-axis.

5.1.1 Autocorrelation

In statistics, correlation is basically measuring how similar two sequences are. The quantitative measurement of similarity of signal 1 and signal 2 can be defined as:

$$r_{12} = \frac{1}{N} \sum_{n=1}^{N-1} x_1[n] x_2[n]$$

Cross correlation between time shifted sequences, can be defined as:

$$r_{12}(k) = \frac{1}{N} \sum_{n=1}^{N-1} x_1[n] x_2[n+k]$$

All possible k-shifted time series could generate another sequence of numbers only changing with k, which is called full cross-correlation. The correlation between a signal and a time-shifted version of itself is called an auto-correlation. A lag operator is used to generate the time-shifted signal and '0 lag' equals to mean-square signal power. Auto-correlation can be defined as:

$$r_{11}(k) = \frac{1}{N} \sum_{n=1}^{N-1} x_1[n] x_1[n+k]$$

5.1.2 Periodogram

The normalised Discrete Fourier Transform (DFT) of a sequence $x(n)$, $n = 0, 1, \ldots, N-1$ is a sequence of complex numbers $X(f)$:

$$X(f_{k/N}) = \frac{1}{\sqrt{n}} \sum_{n=0}^{N-1} x(n) e^{-\frac{j2\pi kn}{N}}$$

where the subscript k/N denotes the frequency that each coefficient captures. Suppose that X is the DFT of a sequence $x(n)$. The periodogram P is provided by the squared length of each Fourier coefficient:

$$P(f_{k/N}) = \|X(f_{k/N})\|^2 \quad k = 0, 1, \ldots, \lceil \frac{N-1}{2} \rceil$$

Notice here that k ranges from 0 to $\frac{N-1}{2}$. In order to find the k dominant periods, we need to pick the k largest values of the periodogram. This works well for short to medium length periods but for long periods or low frequencies, performance is worse because each value in the periodogram indicates the power at frequency interval $[\frac{N}{k}, \frac{N}{k-1}]$ which is too wide to capture large periodicity. Thus the accuracy of periodicity detection at low frequency will be lower than at higher frequency. For lifelogging, this means there is difficulty in detecting patterns measured in cycle times of years.

Another difficulty when using periodograms is *spectrum leakage*, which causes frequencies that are not integer multiples of the DFT bin width to disperse over the entire spectrum which could result in false alarms being detected in the periodogram. However, the periodogram is still a good way to guarantee the accuracy of detected periods with short to medium frequency.

5.2 Parametric Methods

Parametric methods are an aggregation of different methods that fit data into an assumed model first, and then a periodogram will be computed based on the model parameterised by parameters that make the data fit best. One reason to choose parametric methods is that non-parametric spectral analysis methods are limited to that fact that they do not incorporate information that may be available about the process generating the data series. Yet this is also a disadvantages of the parametric methods. If data cannot fit into the chosen model, there is a big chance that the spectrum based on the parameter of the model is not accurate. We will show this in synthetic data examples later. We will now introduce Autoregressive models since these kinds of models show high performance when modeling high-level lifelog data. Also, a Least-Squares Spectral Analysis or Lomb-Scargle method is also introduced since these can handle data with gaps or can work with unevenly sampled data.

5.2.1 Autoregressive Models

With a parametric approach, the first step is to select an appropriate model for the process. This selection may based on *a priori* knowledge about the way the process was generated or based on experimental results indicating that the particular model suits well. Commonly used models are autoregressive (AR), moving average (MA), and autoregressive moving average (ARMA). After the model is selected, the parameters that make the data best suited to the model will be estimated as the next step. The final step is to estimate the power spectrum by using the estimated parameters. Although it is possible to significantly improve the resolution of the spectral estimation with a parametric method, it is important to realise that unless the model that is used is appropriate for the analysed process, inaccurate or misleading estimation may result.

If there is a time-domain signal Y_N that consists of N data samples $Y_0, ..., Y_{N-1}$,

we assume it is possible to model the signal precisely with a model with a number of parameters. Therefore we can efficiently use the parameter of the model to transmit or store the data which can be represented by those parameters, from which we can reconstruct the original signal. Assuming Y_N is generated by an Linear time-invariant (LTI) system, which is obtained by filtering white noise $e(t)$ of power σ^2 through the rational stable and causal filter with the transfer function $H(\omega) = B(\omega)/A(\omega)$, where

$$A(\omega) = 1 + a_1 e^{-j\omega} + ... + a_n e^{-jn\omega}$$

$$B(\omega) = 1 + b_1 e^{-j\omega} + ... + b_n e^{-jn\omega}$$

In the time domain, the above filtering can be represented as:

$$y(t) + \sum_{i=1}^{n} a_i y(t-i) = \sum_{i=0}^{m} b_j e(t-j)$$

If both m and n are non-zero, then the signal is said to be auto-regressive moving average (ARMA) and is denoted by $ARMA(n, m)$. If $m = 0$, then the signal is an auto-regressive (AR) signal and is denoted by $AR(n)$. If $n = 0$, the signal is a moving average (MA) signal and is denoted by $MA(m)$. The spectrum of an $ARMA(n, m)$ process may be estimated using estimates of the model parameters:

$$\hat{P}_{ARMA}(e^{j\omega}) = \frac{|\sum_{k=0}^{q} \hat{b}_k e^{-jk\omega}|^2}{1 + |\sum_{k=1}^{p} \hat{a}_k e^{-jk\omega}|^2}$$

The AR model parameters can be estimated from modified Yule-Walker equations either directly or by using a least squares. Once the AR parameter $\hat{a}_k$ has been estimated, an MA modeling technique, such as Durblin's method may be used to estimate the MA parameters $\hat{b}_k$

5.2.2 Least-Squares Spectral Analysis

In the context of our work on periodicity detection from lifelogs, one of the challenges we are faced with is missing or erroneous data from the lifelog. For such a scenario, the Lomb-Scargle periodogram Scargle (1982) can be used to detect periodicity in signals with missing, or unequally spaced data. Past work Zechmeister and Kürster (2009) has shown that the Lomb-Scargle periodogram that handles missing data values can be successfully applied to generate periodograms from non-continuous lifelog data. Least-squares spectral analysis or Lomb-Scargle (LS) periodogram is a very different method developed by Lomb and Scargle based on work by Barning and Vanicek to handle continuous data with missing parts Press *et al.* (2007); Scargle (1982). To formalise the problem, suppose we have a data sequence with N data points: $X_n = X_{t_n}, n = 0, 1, ..., N-1$. The mean and variance of the data sequence needs to be calculated first. The Lomb-Scargle periodogram is defined formally as:

$$P_X(\omega) = \frac{1}{2}\left\{ \frac{[\sum_{n=1}^{N} y(t_n)\cos(\omega(t_n-\tau))]^2}{\sum_{n=1}^{N}\cos^2(\omega(t_n-\tau))} + \frac{[\sum_{n=1}^{N} y(t_n)\sin(\omega(t_n-\tau))]^2}{\sum_{n=1}^{N}\sin^2(\omega(t_n-\tau))} \right\}$$

where τ is defined as:

$$\tan(2\omega\tau) = \frac{\sum_{n=1}^{N}\sin(2\omega t_n)}{\sum_{n=1}^{N}\cos(2\omega t_n)}$$

5.3 Experiments on Synthetic Data

We will first show periodicity detection using synthetic data as a form of worked example. The purpose of using synthetic data is to see 1) how different methods could potentially be used as methods to detect periodicity and 2) how noise in the

data series could affect the detection of periodicity. These two points are critical in the analysis of real life data. The choice of methods such as parametric or non-parametric could give misleading information when illustrating a spectrum. For example, in parametric methods, the accuracy of the spectrum is based on an assumption that the model chosen can be used to model data series as well. So if the model is not good at modeling the data series, parametric methods can give results that cannot represent the spectrum very well. Also, as real life data is almost always noisy, it is important to observe how noise in the data series could influence the resulting spectrum.

In the first experiment, the data series is generated using cosine waves mixed with a level of noise. These synthetic waves have sampling frequency of 1 which corresponds to one sample per day, and the wave frequency is smaller than 1/2, in order to fulfill the Nyquist sampling theorem, which is that the sampling frequency should be equal or more than twice the maximum signal frequency. In resulting spectrum or periodogram, the X-axis is frequency and Y-axis is energy carried by the corresponding frequency. For all the periodograms in this experiment, the frequency is converted to period using $T = \frac{1}{f}$, the unit of X-axis is day and the 0-day period corresponds to the DC component of the periodogram.

Figure 5.1 is generated by the equation:

$$x[t] = \cos(2\pi t * f_0) + A * w[t]$$

where f_0 is the frequency of the signal and $w[t]$ is the white noise with variance 1. In the example data in Figure 5.1, f_0 is 1/14, in other words the period is 14 days and the strength of noise A is set to 4.

Figure 5.2 is the periodogram of the synthetic data with a 14-day period, and we can observe there is a peak at the 14-day period, followed by several others, which are probably harmonics. Although the strength of noise is four times the signal

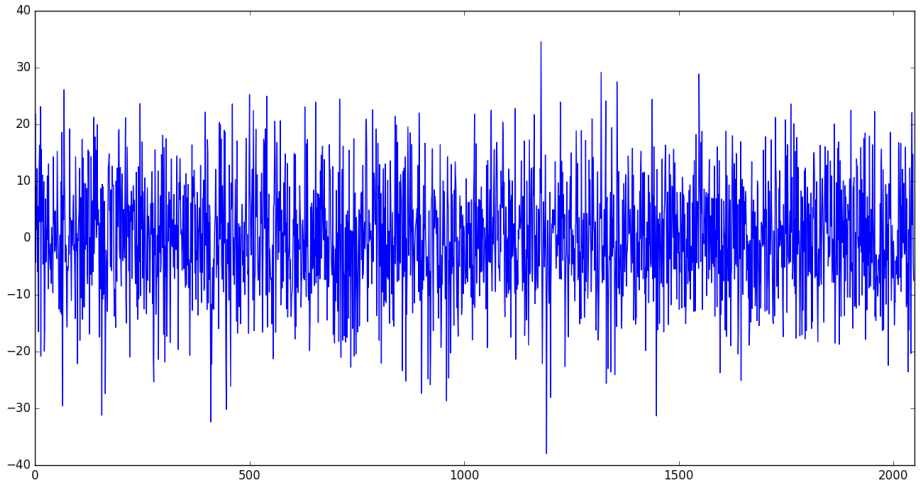


Figure 5.1: Synthetic data visualization.

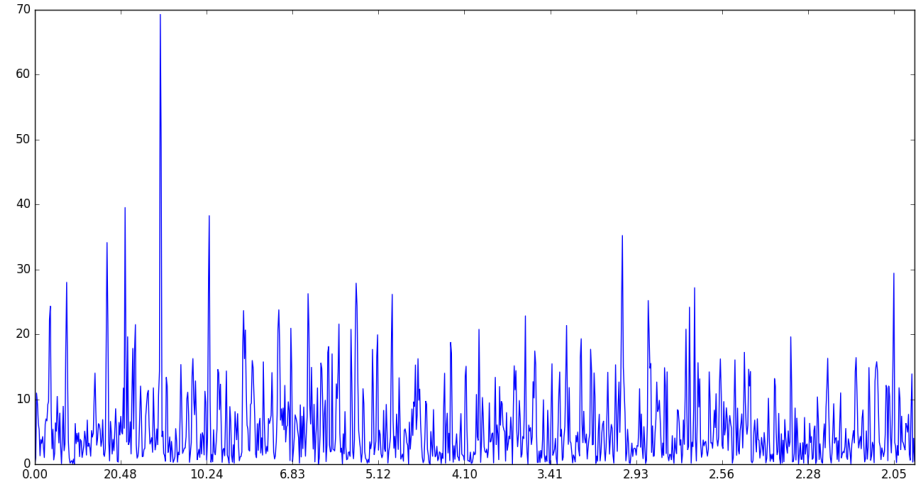


Figure 5.2: Spectrum using periodogram. X axis is frequency. Y axis is energy.

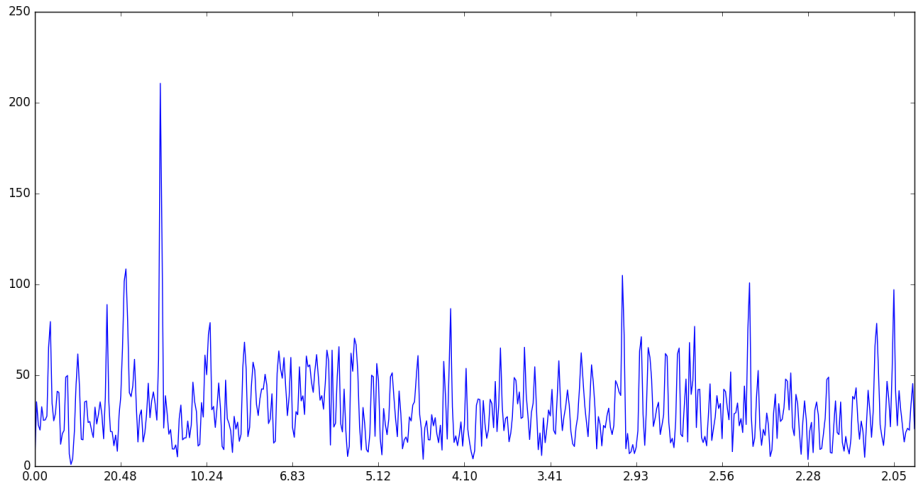


Figure 5.3: Spectrum using Welch smoothed periodogram. X axis is frequency. Y axis is energy.

amplitude, we can still identify the 14-day period. Despite the rich information contained in a simple periodogram, we can also use the Welch periodogram to smooth the simple periodogram. What is needed here is to set up a sliding window over the whole time series and the periodogram within each window is then summed and the average of all the periodograms is computed in order to achieve a smoothed periodogram. Figure 5.3 shows the smoothed periodogram.

Parametric analysis methods using ARMA are shown in Figure 5.4 and Figure 5.5. Since parametric methods need to model data using the Autoregressive Moving Average method to model the data, parameters need to be specified in order to fit into the model. Two tuples of parameters are selected to show how the choice of parameter could impact the data modeling and therefore affect the estimated spectrum.

We can observe in both parametric estimators that the spectrums generated are quite 'clean', with little noise. In Figure 5.4 there is an obvious and significant periodicity at the 14-day period. The significance is more than non-parametric methods achieved. Figure 5.5 uses the same ARMA model but different parameters

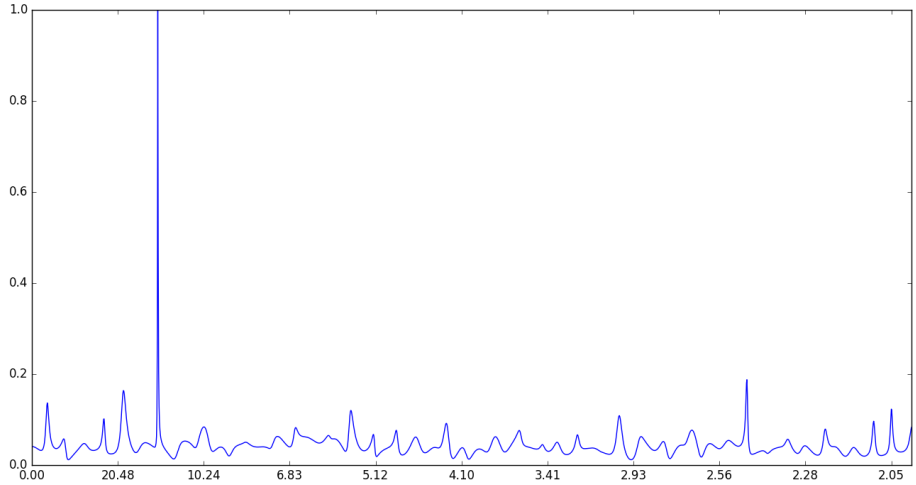


Figure 5.4: Spectrum using ARMA - part 1. X axis is frequency. Y axis is energy.

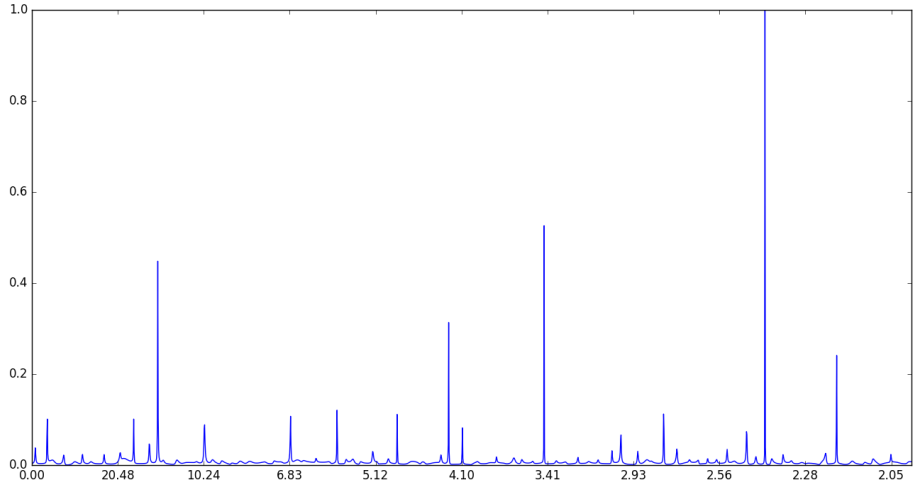


Figure 5.5: Spectrum using ARMA - part 2. X axis is frequency. Y axis is energy.

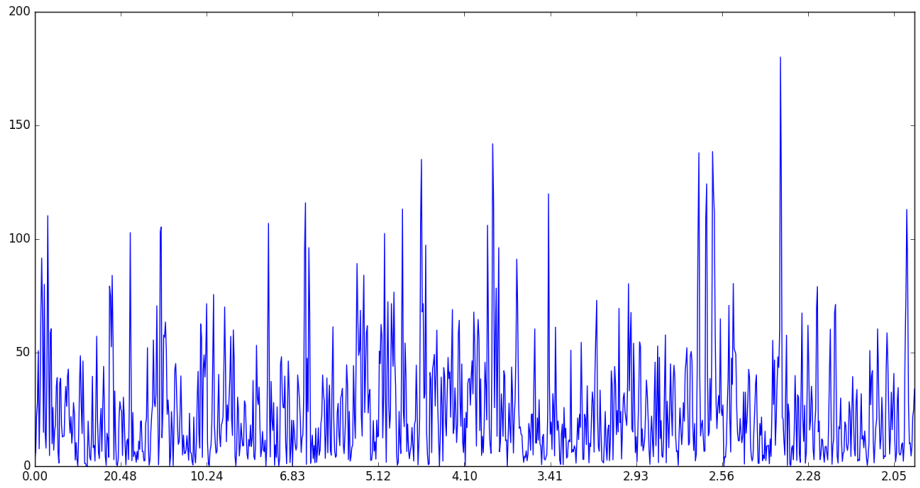


Figure 5.6: Spectrum using periodogram with higher noise strength. X axis is frequency. Y axis is energy.

to fit the data, and while the result is still quite clean there are several false alarms, which do not actually carry in the signal we generated. That tells us that parametric methods are very sensitive to the quality of the model. If the model does not fit very well, like in Figure 5.4, the final estimation of the spectrum could be very misleading.

If we increase the strength of noise from 4 to 8, both parametric and non-parametric methods struggle to identify the right periodicities. The results are shown in Figure 5.6 and Figure 5.7. Based on the results of synthetic data, we will not be using parametric methods in our analysis of real life data because we are not confident enough to make the assumption that the data can fit into certain types of model due to the natural character of real life data such as it being noisy and less structured. In addition, we know that the spectrum of parametric methods is very sensitive to the selection of the models.

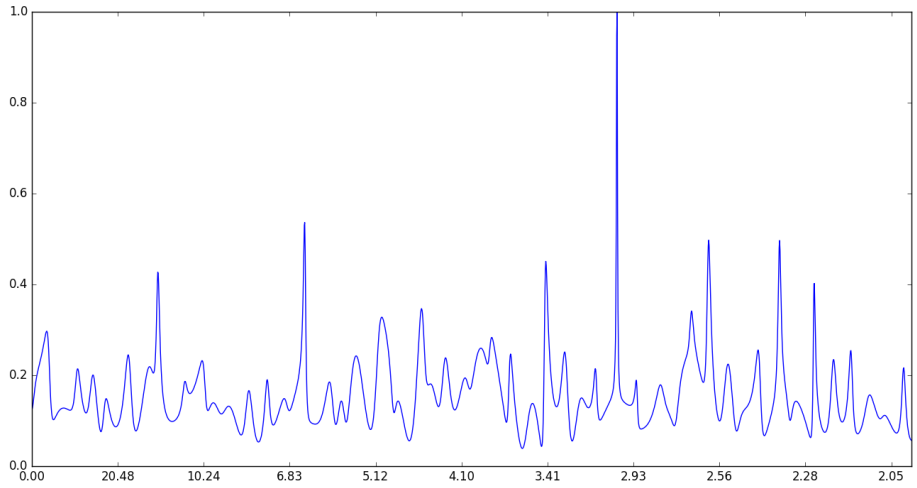


Figure 5.7: Spectrum using ARMA with higher noise strength. X axis is frequency. Y axis is energy.

5.4 Experiments on Real Lifelog Data

We have demonstrated the capability of different methods to detect periodicity on synthetic data. Now we will move to data collected in real life in order to see if there is any latent periodicity. Since this is real life data, the detected periodicity or periodicities can be related to a real life entity or scenario or participant. It will be interesting and practical to illustrate those detected periodicities with real examples.

5.4.1 Periodicities in Athletic Data

We now describe the athletic data briefly. The data consists of 1) running distance 2) cycling distance 3) swimming distance 4) comments made by an athlete to indicate his performance or mood. This data is recorded on a daily basis. In Figure 5.8, the raw distances for running, cycling, swimming and for aggregated activity effort is shown. The latter of these plots accounts for days where the athlete would exercise or compete in more than one activity or discipline and the aggregated activity is computed according to the metabolic equivalent (MET) where the unit of MET is 1

kcal/kg*h. To calculate this the average speed for each of the three sports activities of the athlete is used. In Jette *et al.* (1990), the MET for each sport activity at the average speeds indicated by the athlete are shown in Table 5.1.

Table 5.1: MET TABLE

Activity	Speed (kph)	MET
Running	13	12.9
Cycling	25	8.4
Swimming	3	8.9

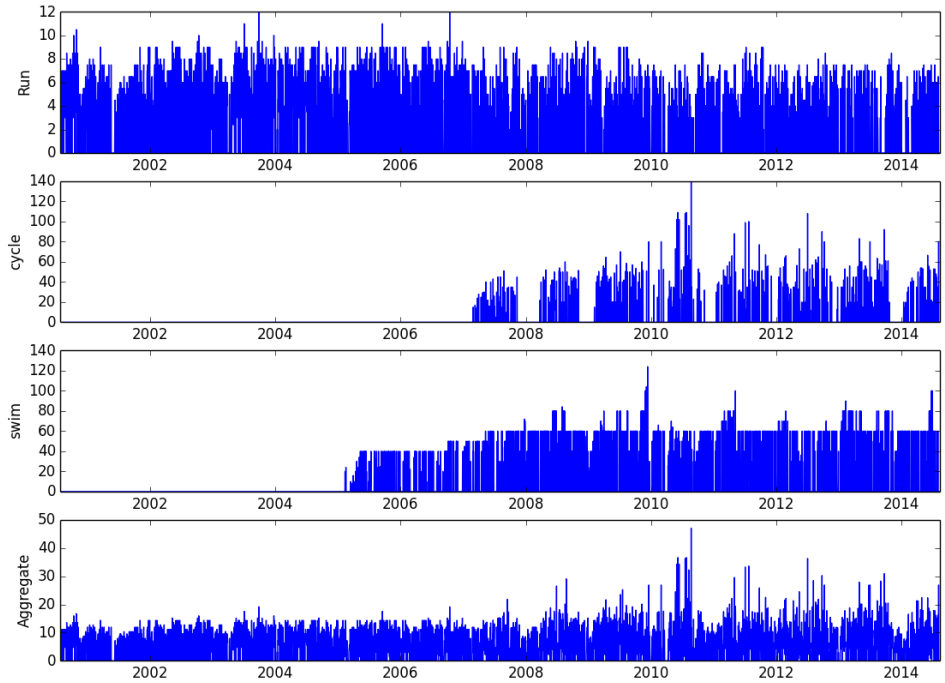


Figure 5.8: Visualization of raw data in the athletic activity dataset. X axis is time. Y axis is performance of different activities.

In the running, cycling, swimming and aggregated data graphs, visualised in Figure 5.8 the X-axis represents time, while Y-axis is the distance for the corresponding activity. Note that this data has been gathered over a period of over 7 years so Figure 5.8 presents a very compacted picture. From the visualization, no

obvious periodicity can be observed in running, swimming or aggregated data but there seems to be an annual periodicity in the cycling data.

For each sporting activity and for the aggregated data, we applied window sizes of 7, 14, 30, 120 and 365 days to calculate moving averages. Figures 5.9, 5.10, 5.11 and 5.12 show the results of this. Running, cycling and swimming start from 2000, 2007 and 2005 respectively. Moving average calculates the mean value of a fixed size window and then moves the window one day forward to get the new value. Moving average works like a low-pass filter; the bigger the window size, the lower the frequency can pass. Because of this, it is easier to find long-term trends using a larger window size because short term shocks in the data (competitions, vacation, short-term injuries) will be smoothed. From the moving average results, we can see that running distance decreased over time, while the cycling and swimming distances increased. The total amount of energy expenditure according to MET fluctuates, and no obvious trends can be seen in the aggregated data. We can infer from this data that after the athlete started to train for swimming in 2005 and for cycling in 2007, he adapted himself to this by reducing the amount of training for running.

One major difference between the sleep dataset that we used earlier and the sports dataset that we are using here is that the sports dataset has 100% capture rate of activity over 10 years, while the sleep dataset captures just over 80% of the nights in a 2.5 year period. The raw figures on sporting activity distances are augmented by the athlete annotating most days with text comments which summarise the day and occasionally report on performance or mood. These reports are infrequent (25–30%), and so provide sparse data which we can also examine for periodic patterns.

We manually annotated the text comments for mood and for performance in order to create a third dataset. Four annotators were asked to annotate the text for mood by following the following strategy:

- if a text comment provides an indication of mood (*"feeling great"* or *"not*

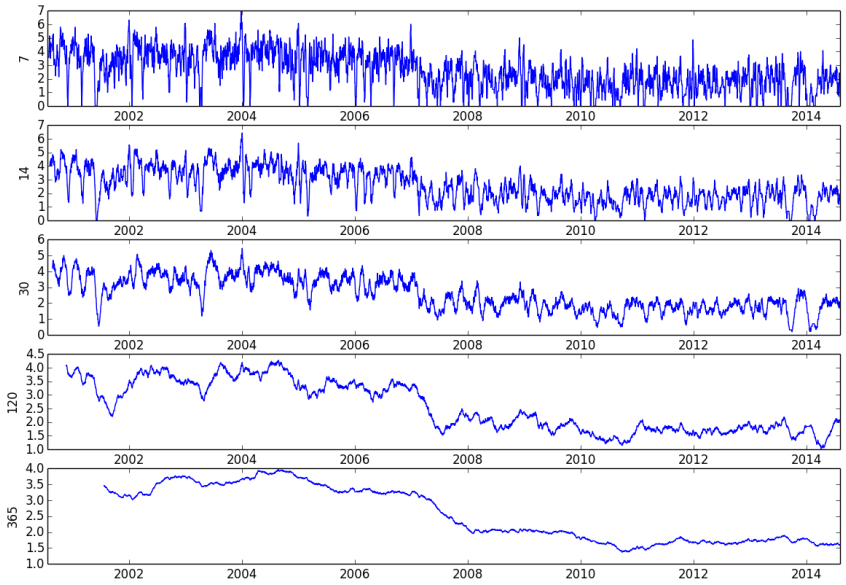


Figure 5.9: Moving average values for run. X axis is time. Y axis is filtered result for running. From top to bottom the window sizes used to filter are 7, 14, 30, 120, 365.

well", *"ok"*), give a rating between 1 and 5, where 1 indicates the worst mood feelings and 5 indicates the best mood feelings. If there is no indication of mood in the text, give a rating of 0.

For annotation of performance the four annotators were given the following instruction:

- when a text comment provides an indication of performance (*"personal best"*, *"strong finish"*, *"stopped early"*), give a rating between 1 and 5 where 1 indicates poorest performance and 5 indicates best performance. If there is no indication of performance in the text, give a rating of 0.

Comments made by the athlete during the year 2007 were randomised and presented to 4 annotators. Because the marks for mood and performance given by annotators are highly subjective and have biases, inter-annotator agreement, namely Cohen's Kappa co-efficient Cohen (1968) was calculated across the annotators and

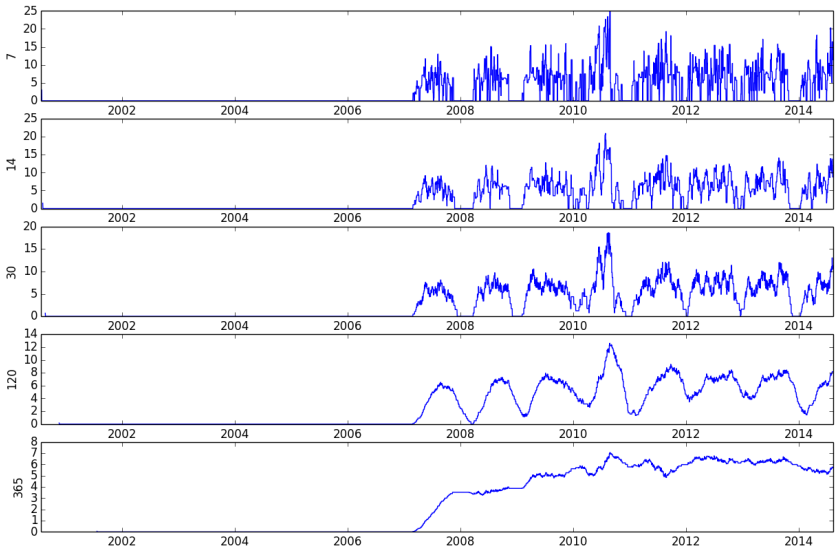


Figure 5.10: Moving average values for cycle. X axis is time. Y axis is filtered result for cycling. From top to bottom the window sizes used to filter are 7, 14, 30, 120, 365.

is presented in Tables 5.2 (mood) and 5.3 (performance).

Table 5.2: Inter annotation agreement for Mood

Annotator	A	B	C	D
A	1.00	0.47	0.60	0.47
B	0.47	1.00	0.41	0.48
C	0.60	0.41	1.00	0.36
D	0.47	0.48	0.36	1.00

Cohen's Kappa coefficient ranges from 0 to 1, where a value of 1 indicates complete agreement between a pair of annotators, and 0 denotes complete disagreement. For mood, we can see that annotator A highly agrees with annotator C, while C and D are agree least with each other, though all values are greater than 0.3. For annotation of performance, it is obvious that annotator A has low agreement with all three other 3 annotators. Based on this assessment of inter-annotator agreement, we apply the following fusion strategy:

- For *Mood*, for each annotated comment, discount the value which is the great-

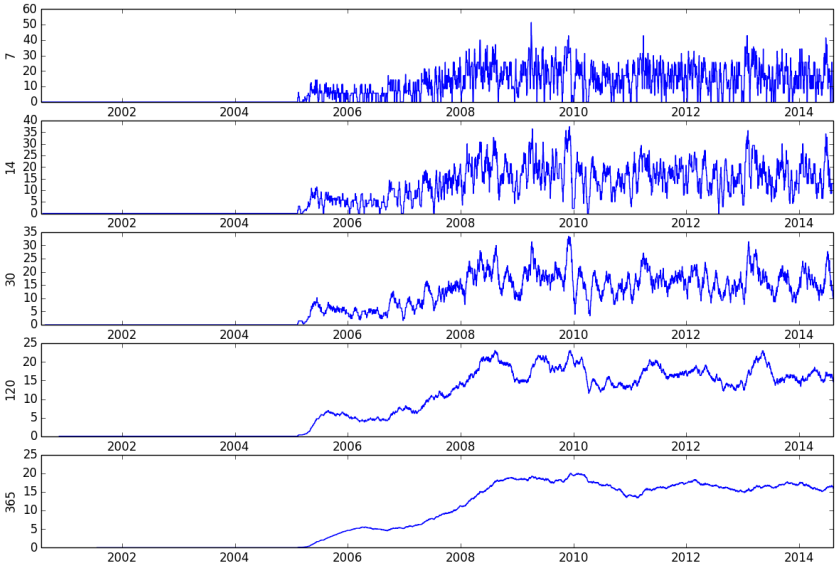


Figure 5.11: Moving average values for swim. X axis is time. Y axis is filtered result for swimming. From top to bottom the window sizes used to filter are 7, 14, 30, 120, 365.

Table 5.3: Inter annotation agreement for Performance

Annotator	A	B	C	D
A	1.00	0.15	0.12	0.17
B	0.15	1.00	0.35	0.37
C	0.12	0.35	1.00	0.36
D	0.17	0.37	0.36	1.00

est outlier and average the remainder;

- For *Performance*, discount annotator A completely and then for the other three (B,C,D) annotations on each comment, discount the one who is the greatest outlier, then average the remainder.

The fused mood and performance data from the 4 annotators are sparse and have large amounts of missing data and gaps as shown in Figure 5.13 where a black horizontal line represents a mood or performance value while whitespace indicates there is either no comment made by the subject for that day's activity or the mood and/or performance indicators are absent. The gap sizes for fused mood and per-

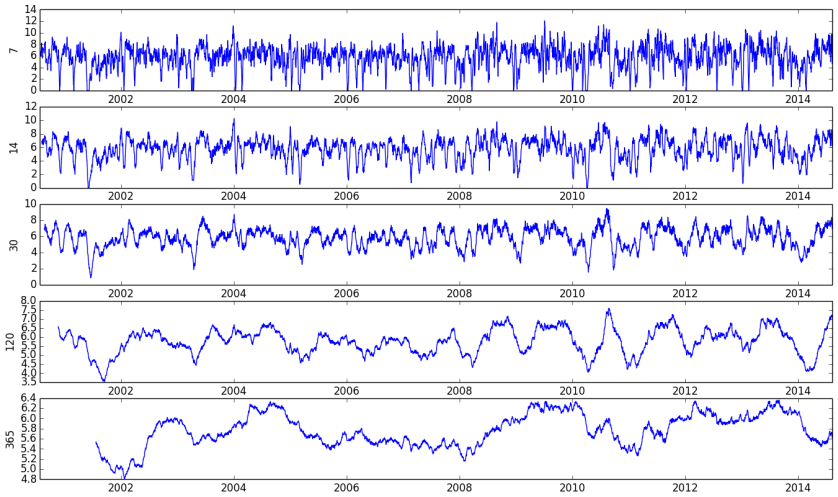
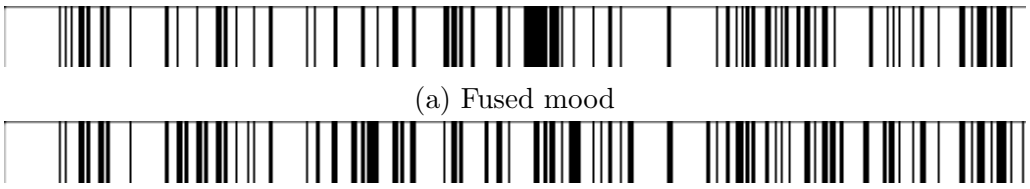


Figure 5.12: Moving average values for aggregated. X axis is time. Y axis is filtered result for aggregated data. From top to bottom the window sizes used to filter are 7, 14, 30, 120, 365.



(a) Fused mood

(b) Fused performance

Figure 5.13: Distribution of fused data from annotators

formance vary between 1 and 19 days, while mood has a mean gap of size 4.15 days compared to 3.15 for performance. This unevenly sampled data makes it a real challenge to detect periodicity from this data and an ideal target for the Lomb-Scargle periodogram.

Since the sampling rate of our sports activity dataset is 1 day, the minimum periodic pattern of this dataset we can detect is 2 days. In Figure 5.14, periodograms for the sports dataset which does not have missing data and is consistently and regularly sampled for the three sport activities and for the aggregated data MET levels, shows interesting results. We can observe three significant energy levels carried by three different frequencies consistently across all 4 subplots. These three frequencies

are around 0.14, 0.28, 0.43, which corresponding to periods of 7 days, 3.5 days and 2.3 days. Moreover, if we look at the plots more thoroughly, there exists a frequency at circa 0.0027 located near the left end of the cycling and aggregated data subplots. This frequency corresponds to the annual period (ca. 365 days) that we observed in the visualization of the cycling data.

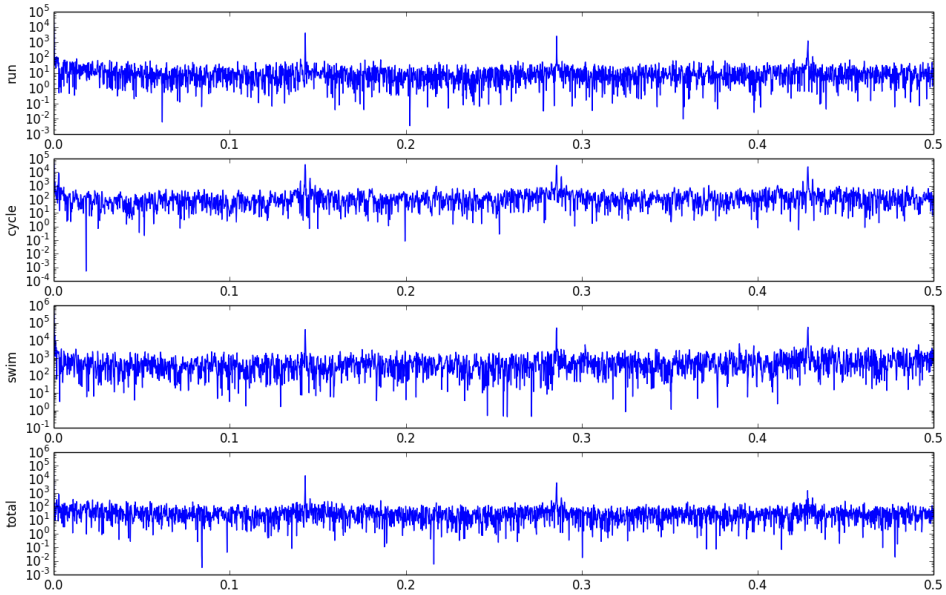


Figure 5.14: Sports dataset periodograms. X axis is frequency (1/day). Y axis is energy.

In order to investigate periodicity in irregularly sampled data, a second tool we use is autocorrelation. Autocorrelation of 10 years data is plotted in Figure 5.15.

An autocorrelation computes the correlation between the signal and a time-shifted version of the same signal. The X-axis of the autocorrelation plot is time lagged and the Y-axis is a measure of the correlation of the original signal and the lagged signal. If the original signal is periodic then the autocorrelation of the signal should also be periodic and the periods will be located at the peaks of the autocorrelation plot. From Figure 5.15, there are no periodicities observed in the running, swimming or MET score aggregated data, but an annual periodicity can be

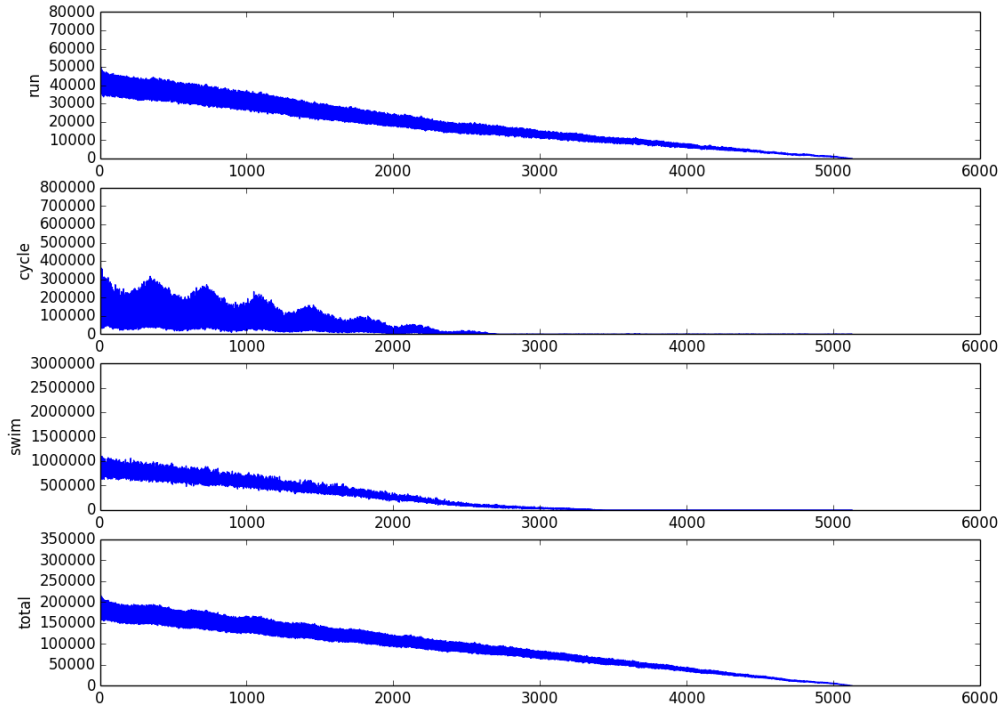


Figure 5.15: Sports dataset autocorrelations. X axis is time lag (day). Y axis is correlation between original signal and signal with time lag.

found in the autocorrelation of the cycling data. Curious as to where the periodicities over 7, 3.5 and 2.3 days which were found in periodograms from running, swimming and the aggregated data, we took one year of data from 2007 to see if we could detect periodicity in periodograms for just that year. An autocorrelation plot for data from the year 2007 is shown in Figure 5.16.

The autocorrelation plot of sports data from 2007 shows that there is a very regular weekly periodicity in running, cycling and in the total energy expenditure of activities, but a less regular weekly periodicity for swimming. We can also find smaller peaks between the two obviously large peaks from running and cycling data, which may correspond to the 3.5- and 2.3-day periodicities also detected in the periodogram. However there are no obvious smaller peaks found in the autocorrelation

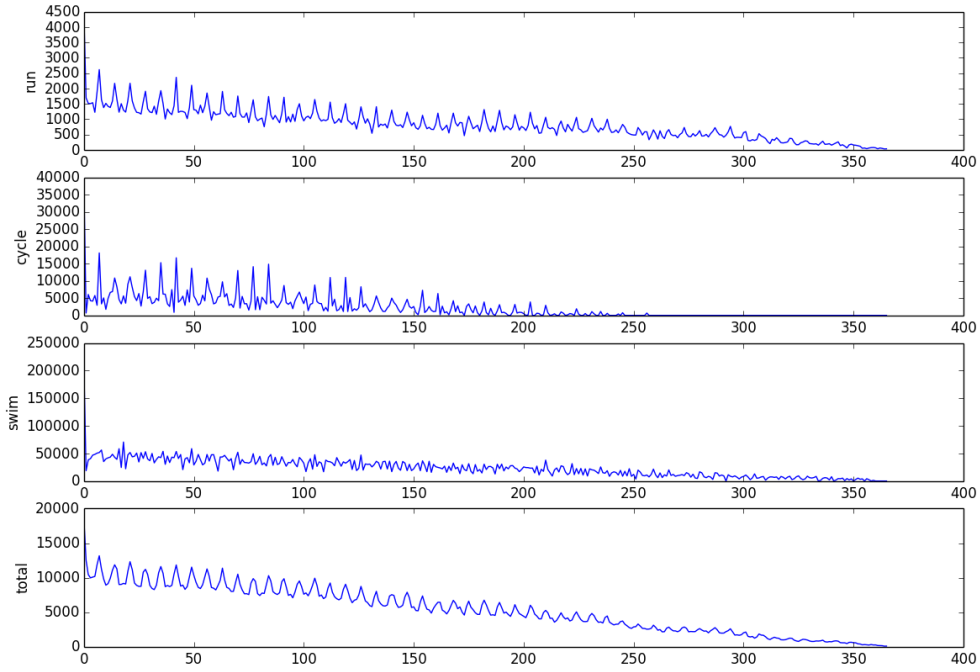
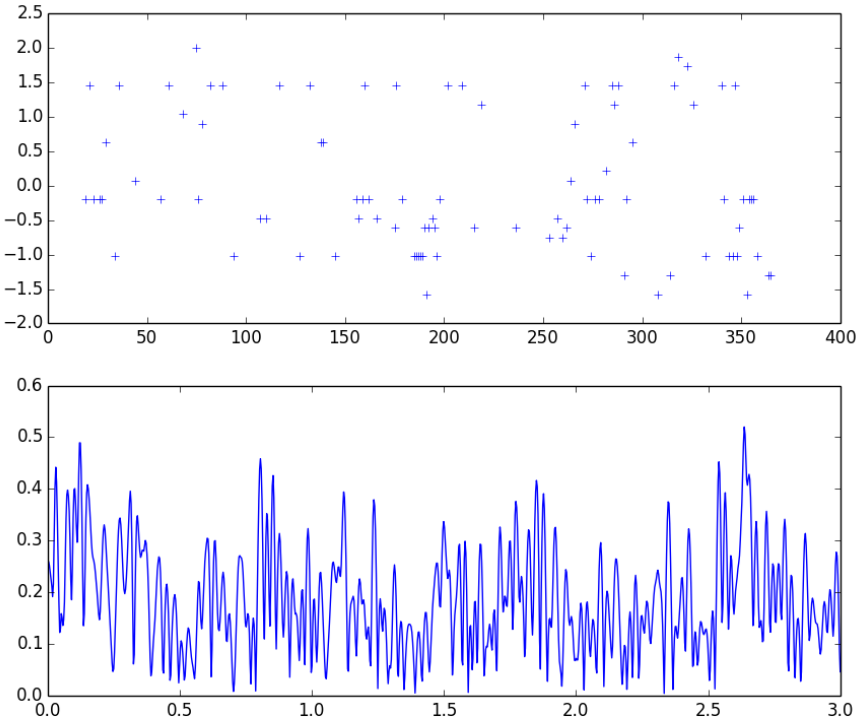


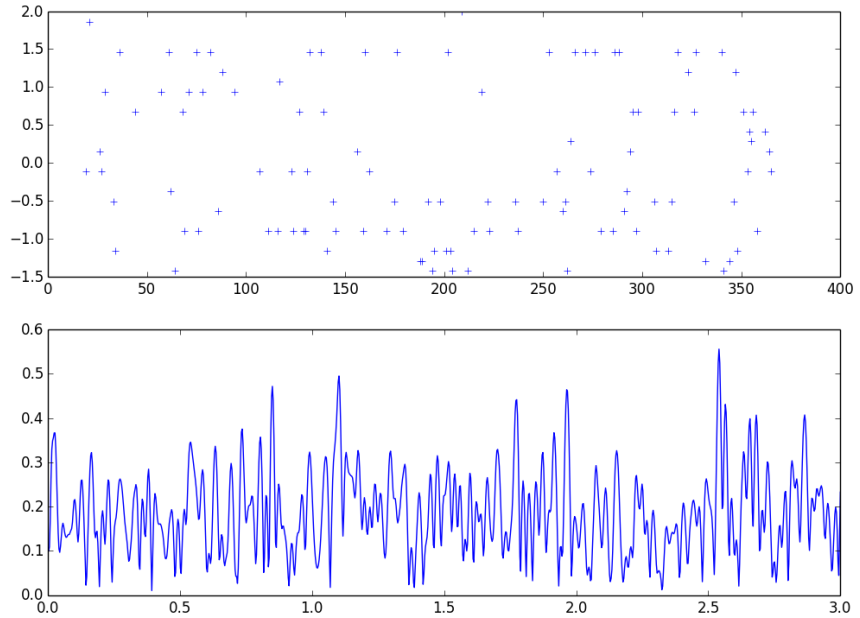
Figure 5.16: Autocorrelation plot of sports data from the year 2007. X axis is time lag (day). Y axis is correlation between original signal and signal with time lag.

of aggregated data. A possible explanation may be that these detected periodicities indicate the lifestyle of the subject such as regular scheduled training sessions for running, cycling and swimming. Another explanation might be that there exists an inherent timetable that the subject follows in order to balance participation in the three different activities. For instance the timetable could be every 2 or 3 days run, cycle or swim once. Determining this requires going back to the subject to confirm this, though this falls into the category of exploiting rather than determining periodicities which as mentioned earlier, is beyond the scope of this chapter.

In order to detect periodicity in mood and performance data, which are unevenly sampled and have gaps in the data, the Lomb-Scargle periodogram is applied to the mood and to the performance data. For the Lomb-Scargle periodogram, the period is $T = \frac{2\pi}{f}$. In Figure 5.17, we can see that there are no statistically significant



(a) Mood. X axis of first panel is time and Y axis is score for mood. X axis of second panel is frequency and Y axis is energy.



(b) Performance. X axis of first panel is time and Y axis is score for performance. X axis of second panel is frequency and Y axis is energy.

Figure 5.17: Mood and performance data

energy levels carried by any of the frequencies in the LS periodogram for mood or for performance. In other words, no periodicity is detected in either mood or performance data which has been fused from the manual annotators of the subject's text comments. Trying to rationalise this by going back to the subject might reveal that the training schedule is oriented to have peak performance during the months of competition, typically the Summer months, so there could be an annual cycle for performance which could be tied to a mood performance cycle. There may also be peaks in performance, and in mood, around regular seasonal targets such as Winter, Spring, Summer and Autumn events. The fact that such periodicities did not appear does not mean that they do not exist, it just means that they were not detected, most probably because of the sparsity of our mood and performance data with large gaps and irregular sampling. Not even the Lomb-Scargle periodogram was able to overcome this disadvantage.

5.4.2 Periodicities in Sleep Data

The sleep dataset that we used represents 2.5 years of continuous nightly sleep monitoring for an individual and contains the following information:

1. Time to sleep - represents the time taken between going to bed and falling asleep;

2. Time to rise - represents the time taken between waking and getting out of bed;

3. Time asleep - represents the duration of sleep;

4. Quality - a numeric indicator of sleep quality computed as a function of how well the night's sleep mapped to the circadian sleep (90-minute) rhythm and how many cycles of that rhythm were completed;

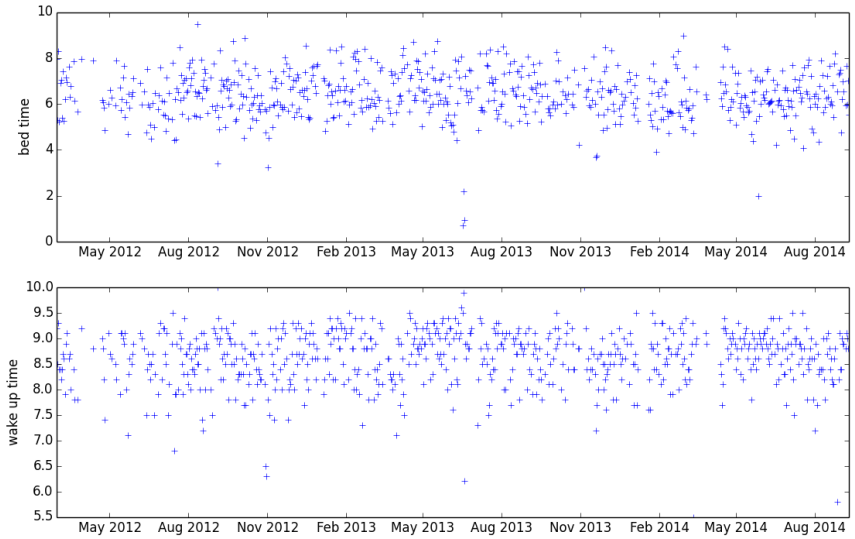


Figure 5.18: Some of the raw sleep data. X axis is time. Y axis of the first panel is bed time. Y axis of the second panel is wake up time.

5. Times woken up - represents the number of instances of a wake-up during sleep, where ’“wake up” represents even a turning over in the bed.

The distribution of some of these metrics (3 and 4) is shown in Figure 5.18 and the frequency of data capture is shown in Figure 5.19 where a thin horizontal black line represents an instance of captured data, and where days are contiguous these form the wide black bands in the figure. An obvious periodicity we would hope to detect is based on the weekly cycle where the subject tends to sleep longer at weekends than during workdays because the subject has a regular work schedule of Monday to Friday.

Figure 5.19: Frequency of capture of sleep data

Each of the metrics from sleep logging (duration, quality, number of wakes, time in bed, etc.) has been analysed for periodicity but rather than present all of them, we limit ourselves to just two. For time asleep, a weekly periodicity is clearly detected

as can be seen in Figure 5.20. This can be explained by the weekday/weekend cycle which is the basis for the subject's lifestyle of working during weekdays and having to get up early to commute to work and then leisure activities with later rising at the weekend. There is also a relatively weaker periodicity at around the 120 day frequency, about every 4 months but without going back to the subject to investigate, this remains unexplained for the moment.

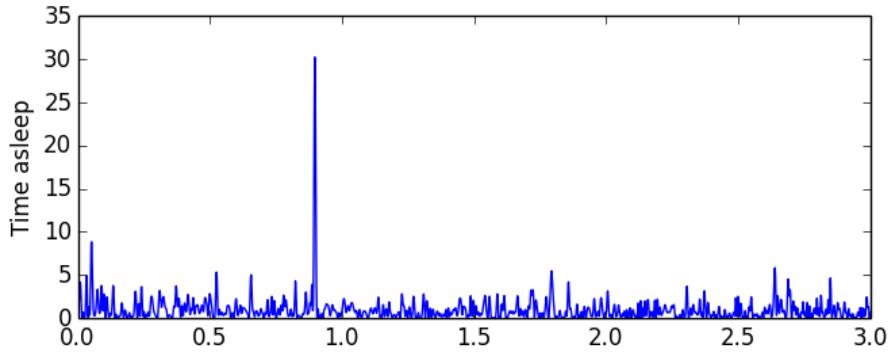


Figure 5.20: Time asleep periodogram. X axis is frequency. Y axis is energy.

For sleep quality as shown in Figure 5.21 there is no weekly periodicity which tells us that even though the subject sleeps more at weekends, he does not actually sleep with better quality. We also observe a periodicity around 128 days (ca. 4 months) for sleep quality but at the time of writing, without conferring with the subject, this is something we cannot explain.

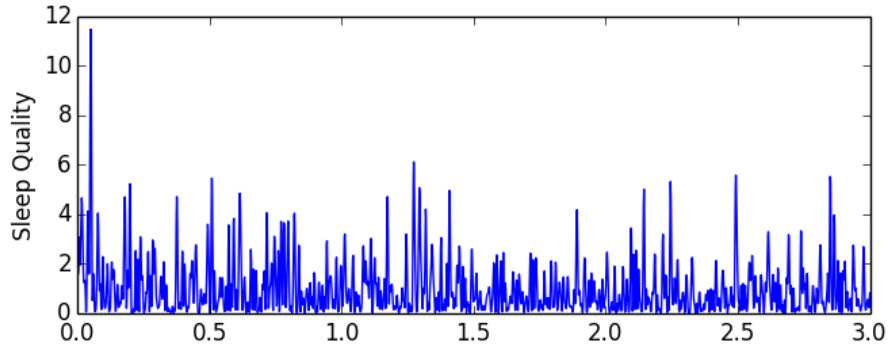


Figure 5.21: Sleep quality periodogram. X axis is frequency. Y axis is energy.

While the other sleep parameters such as *time spent in bed*, *time going to bed* have also yielded interesting results, the point we wanted to make is already made, namely that we can detect credible periodicities from real lifelogs even though there is missing data and irregular sampling.

5.4.3 Periodicities in Sleep Data from ASU

For this experiment, 25 adults took part and gathered data from a wrist-worn accelerometer, continuously, 24 hours per day, for between 28 and 72 consecutive days. Each of our participants took part in this study conducted by Arizona State University (ASU) because they have some sleep issues and during the lifelogging period there were some sleep improvement interventions introduced. We are interested to see if the effects of the interventions on behavior can be detected in the periodicity analysis. The computation of strength of periodicity will be introduced in the next chapter.

The non-parametric methods are used to compute periodogram for subjects that have complete data. If there are gaps in the data we can either set the missing data as zeros or use the Lomb-Scargle periodogram. The reason we are not using parametric methods is because 1) Time series models such as AR/MA/ARMA would possibly not be good at modelling this type of longitudinal data with high sampling frequency and 2) a large lag is needed to compute this type of model which could lead to enormous calculation time.

Since the collected data contains measures of acceleration of 3 axis, the raw data from the accelerometer was summarised to 1 second epochs. We firstly computed gravity-subtracted sum of vector magnitude for each 3 axis acceleration. And then we sum every 1 second of the computed gravity-subtracted sum of vector magnitudes. A sample of this for subject 102 from the experiment is shown in Figure 5.22. This plot of the overall activity levels illustrates several isolated periods of high activity,

probably exercise of some form, throughout the 12- week period of data logging but there is no evidence of changes in behavior. It provides visualization of the sum of vector magnitudes (1 second epochs) along the y-axis and time along the x-axis over the course of the monitoring period. Sleep and wake periods are evident visually from these data.

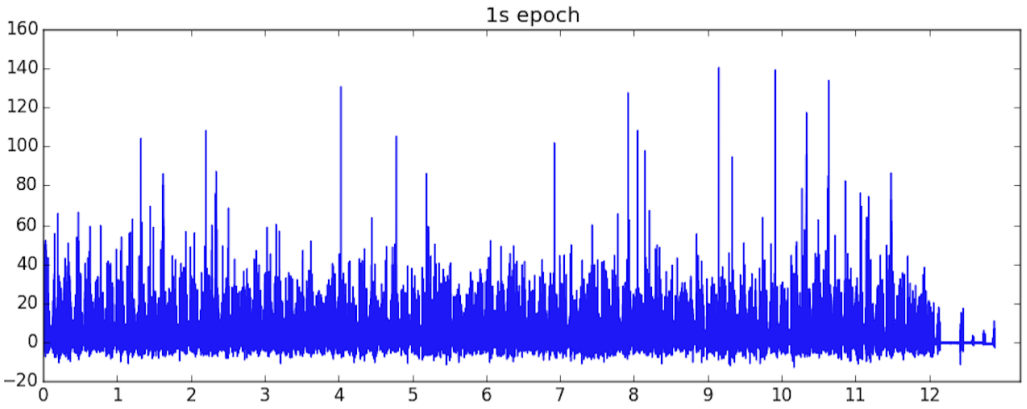


Figure 5.22: ASU sleep data - 1 second epoch. X axis is time. Y axis is summed 1s epoch.

For each of the 25 participants we generated a periodogram like Figure 5.23, and a sample periodogram for the same participant, number 102, is shown in Figure 5.23. The figure displays a periodogram calculated from 1 minute epochs. The x-axis is frequency and y-axis is energy of the frequency, namely, how strong the corresponding frequency is. This shows a reasonably strong energy level around the 1-day point and a smaller peak at around the 12-hour point. Compared to some of our other participants who have gathered similar data, the regularity of this individual's daily cycle is not particularly strong for the whole of the 12-week period, suggesting that s/he may work shifts or just have a very disorganised and irregular lifestyle. In this figure, we observe strong circadian periodicity followed by a 12 h periodicity which is the harmonic of the circadian. Interestingly, there is a peak standing at the weekly period. This could be caused by a high level of noise, but no within-day patterns were observed.

Figure 5.24 shows 60s epoch of subject 108 and the result of Lomb-Scargle pe-

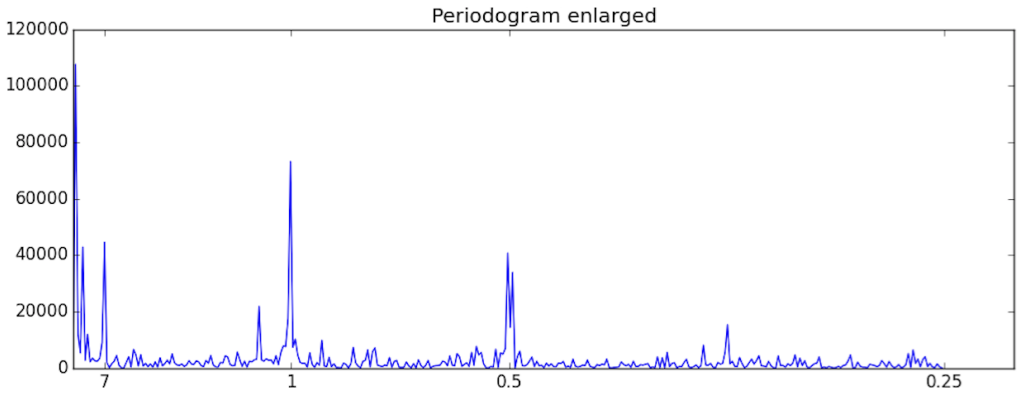


Figure 5.23: Periodogram from ASU Participant. X axis is frequency (1/day). Y axis is energy.

riodogram conducted on this subject which is with missing gaps. In the 60s epoch sub-figure, we can see the gap is at around the 8th week and in the Lomb-Scargle periodogram, we need to specify the starting and stoping frequency that we will examine. In this case we choose the frequency between 3 days and 0.3 day, and the number of frequency in between is set to 1,000. In this sub-figure below, we can see that a strong response resides at the 1-day period, followed by a half-day period, but there is no peak at around the 0.25 day mark. We can observe a similar spectrum in the periodogram and in the L-S periodogram.

5.5 Conclusions and RQ3

In the chapter, inspired by the results of modeling longitudinal lifelog data and works by other researchers, we applied periodicity detection on several longitudinal datasets, which include

1. synthetic data to show the character of different periodicity detection methods;
2. distances covered for athletic training and competition for an international triathlete, over a 10-year period;
3. sleep quality, duration and timing data from a subject over a 2.5 year period;

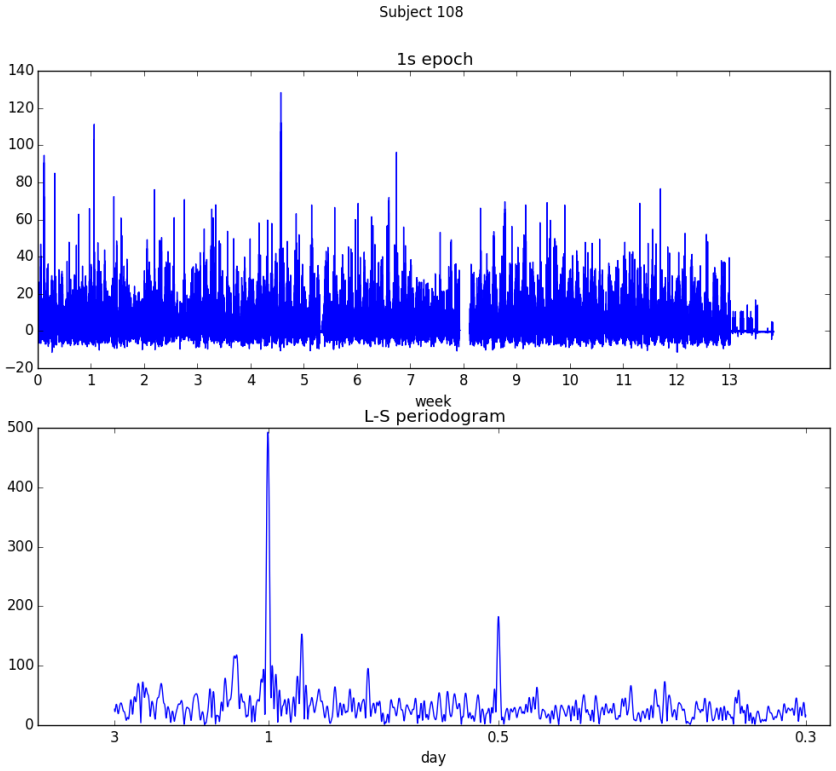


Figure 5.24: ASU Lomb-Scargle Periodogram. First panel is 1s epoch. X axis of first panel is time (week). Y axis is 1s-epoch. Second panel is L-S periodogram. X axis of second panel is frequency (1/day). Y axis of second panel is energy.

4. raw accelerometer data for 25 participants from an Arizona State University sleep study dataset.

Experiments on synthetic data has shown the advantages and disadvantages of several important methods to tackle periodicity detection. For the reason that parametric methods have higher requirements which is that the data has to fit into an assumed model, and our typical lifelog data does not necessary fit into such models, we choose to use parametric methods for cases where we have complete data and we choose to use the Lomb-Scargle method for cases where we have incomplete data.

The athletic training dataset (number 2 above) was augmented with a pool-based annotation of the triathlete's daily text commentary on his training and perfor-

mance, from which we were able to get annotations for mood, and for performance. This gave us a collection of datasets which are rich in the variability of their regularity of logging, from consistent and regular daily entries to much more sporadic data with missing data and irregular sampling.

Applying moving average, we discovered that after starting cycling and swimming at a point several years ago, the subject decreased the amount of running while the distances for swimming and cycling kept increasing. The use of periodograms revealed that there are rhythms of repeating patterns at 7, 3.5 and 2.3 days for the running, cycling and swimming data, as well as for when the individual activity data is aggregated based on MET scores. An annual periodicity was also detected in the cycling data. Using an autocorrelation plot for data from the year 2007, an obvious weekly periodicity was detected in running, cycling and aggregated MET data but the weekly pattern for swimming is weak suggesting less rigour and regularity associated with training in that sport. An autocorrelation plot of running and cycling shows an unexpected periodicity at a cycle of less than a week (2 or 3 days). This infra-week periodicity may be caused by training schedules for different sports in order to achieve a balanced exercise portfolio. There are no significant periodicities detected in the Lomb-Scargle periodogram for mood or for performance when fused from the annotations of a set of four annotators.

In the sleep dataset (number 3 above), several metrics related to sleep were examined. There are not a common or unified pattern detected across different metrics, that is to say each metric used to detect periodicity either has periodicity with different periods or does not have periodicity. For example in time asleep, a weekly periodicity is clearly detected and for sleep quality there is a periodicity at around 4 month detected.

In the Arizona State University raw accelerometer dataset (number 4 above), the raw data was firstly preprocessed into 60s epochs, and then a periodogram was computed for both complete data and data with gaps. 1-day periodicity is revealed in

most of the participants, and in some participants there is even weekly periodicity detected. The fact that a periodogram can be calculated directly from low-level accelerometer data without converting it into a higher semantic level facilitates a method for characterizing behavioral patterns that encapsulate behaviors across the 24h spectrum, longitudinally. Our results indicate a pattern that resembles natural circadian rhythms; however, this framework is flexible enough to identify additional periodicities that may emerge and characterises the strength of these patterns.

Evaluating the detected periodicities using some form of qualitative evaluation will be detailed in Chapter 7. Case study can be often seen in qualitative studies. Researchers usually study case-by-case independently and then draw findings separately. An interview is a common way to collect sufficient data to for qualitative analysis. After a case-by-case study, a cross case analysis could be conducted to discover common phenomena. Evaluation is always a challenging part of such research and the relevance of any detected periodicities is quite subjective since every individual has his/her own understanding of their own periodicity.

We have demonstrated in this chapter that automatic detection of periodicities from lifelog data can be achieved, even when there is substantial missing data. We have shown that methods based on periodograms and/or autocorrelation can be used to detect periodicity on complete datasets, while Lomb-Scargle periodograms can be used to detect periodicity on datasets with missing data. Experiments conducted on real life datasets with different levels of sparsity show that we are able to detect periodicity in these datasets. Since we are able to detect periodicity in real life data, the next chapter will address how to compute the strength of the periodicity in order to reveal potential behavioural changes in longitudinal data.

Chapter 6

Intensity of Periodicity

In the previous chapter of this thesis, we have shown that there is periodicity in longitudinal lifelog data, even when we use a basic periodogram to detect this. In this chapter, we are focusing on how to define and calculate the degree to which a frequency is periodic, in a lifelog. For this reason we introduce the idea of *intensity of periodicity* where the ideal output of intensity of periodicity would be a number or value that indicates the regularity of a certain periodic frequency within a given timespan or duration. If we use a window, sliding through a whole dataset or lifelog with incremental steps, the output would be a series of numbers or values indicating the change of regularity of a certain period, within the whole dataset. Changes in intensity of periodicity could potentially reveal changes in periodic data which could potentially correspond to changes of behaviour. In the literature, we haven't seen many works in lifelog to detection periodicity, not even mentioning the intensity of periodicity. Since we discovered the periodicity in lifelog data, to measure the strength/intensity to enable more application would be reasonable next step.

The intensity of periodicity potentially provides a different and practical perspective for researchers to review lifelog data. After identifying the significant periodicity frequency (e.g., weekly, daily) we would like to compute how strong or weak that period is and/or how the strength of the period changes over time. Tempo-spectral

analysis is popular in signal processing to understand how the frequency of a signal changes with time. Suppose we are able to detect the most significant periods in lifelog data, it is interesting to see how the strength of the periods changes, so that we can identify when is high/low regularity. The intuitive idea is to calculate the energy carried by those significant periods using a moving window. Here are the high-level steps we follow to calculate period intensity change with time:

1. Choose a suitable length of window, within which a periodogram could be calculated. If the window size is too long, the time resolution of the signal would be poor while good frequency resolution could be achieved and vice versa. One way to trade off between time and frequency is to overlap windows, but what needs to be addressed is that overlapping of windows brings issues of time lag into the intensity of periodicity. This may cause delays or advances of the resulting intensity graph, comparing it to actual data.

2. From the periodogram, extract the frequency that is exactly and/or close to detected significant periods of the lifelog, and calculate the corresponding energy. Depending on the size of the window, the most significant period within a window may differ from the most significant period detected, using all data. Also, the most important periods detected using all data could be affected by spectral leakage due to the fact that real frequencies may not be integer times the frequencies of cosine/sine basis, in FFT. Using sophisticated methods such as adding various window functions could decrease the spectral leakage problem.

3. Move windows and repeat the second step, until there are no more data point available. The y-axis in the resulting figure stands for energy, namely the regularity of the selected frequency. Note that the high total amount of energy within a window might lead to high intensity values.

6.1 Definition of Intensity

In order to define what intensity is, we will try to understand what factors may related to modeling lifelog data, and those factors may result in changes in lifelog data. We shall then focus on using those factors to generate simulated data with known intensity. With the simulated data can we then compare the performance of different metrics to measure the intensity of periodicity quantitatively, against ground truth.

6.1.1 Activity trends in data

The first factor that may affect the measurement of intensity of lifelog data is the trend of activity in the lifelog data. That is to say, the quantity of activity either increases, declines or stays the same within a period of time locally and/or longitudinally. For instance, for a triathlete this situation may happen due to the need for intensive training during a short period of time when preparing for competition. Or, people suffering from aging symptoms might decrease their overall activity levels due to constrained mobility. The previously mentioned longitudinal trend or global trend is the trend over a long period of time compared with the local trend which is within a short period of time. The trend can be linear or non-linear. The reason that activity might affect the intensity of periodicity is that the activity level itself might change the results of spectrum analysis and we suspect this may affect the low frequency energy. As we mentioned earlier, both global and local trends could be linear or non-linear but in our experiments, we only consider linear trends, which might or might not fit the real world experience, but is clearer to illustrate the problem we are facing, namely to compute intensity.

6.1.2 Local Activity Patterns

Local activity patterns refer to activity patterns repeated within a short period of time. For example a pattern can be exercise on Friday and Sunday for 6 months. So the patterns can be:

1. Activities repeated at a certain day in a week over a certain period of time. For instance routine activity on every Monday for a year. In this case, within the one year the intensity of weekly periodicity should remain the same, although the local trend or global trend might change.

2. Activities repeated at any day in a week within a certain period of time. For instance activities that are done once every week for 6 months. In this case we still would like to think the activity is routinely executed, and although on an irregular day but at the same time being done regularly on a weekly basis.

The first situation can be simulated by sampling data using a Gaussian distribution and the second one can be achieved using uniform distribution. In theory the first one would have less randomness than the second one. Because uniform distribution has bigger entropy than Gaussian distribution, thus more information is carried. In the matter of factors even if the activity is always repeated at some day, there are chances that the strength of each activity can vary at each repeated day. Because of this, even more uncertainty can be brought into the data.

6.1.3 Missing Activity Routine

It is practically impossible to maintain any activity routine without missing some planed activity in real life situations due to a variety of reasons. The chances that a routine activity would be missed could potentially cause a change of intensity of periodicity. The probability of missing a routine activity thus bring another form of uncertainty into our model. The example of probably of missing an activity could

be as much as a 20% chance that a weekly activity would be missed within a certain month. In other words, the intensity of the weekly periodicity would have been 0.8 of the original intensity if the original routine was kept. We assume the probability of the chance that missing a routine activity follows a uniform distribution, which will be examined in the synthetic data experiment.

6.1.4 Intensity of Periodicity

Having discussed the factors might change the intensity of periodicity, we have to define what is intensity of periodicity and how can it be computed. Starting from the point where we have already determined which periodicity is significant in our lifelog data, we shall then decide the intensity of the periodicity. That is to say, to have a detected periodicity is a condition for computing intensity. In fact it is quite intuitive to understand, since we have to know what is the main periodicity in the lifelog data, then can we decide how strong the periodicity is. However here is the key issue, how do we know how "strong" a periodicity is, namely the intensity of the periodicity? Firstly, let us think about the requirement of intensity of periodicity. It should be able to show:

1. The difference of patterns in terms of quantity. For example two patterns may be the same shape, but with different sizes, and the intensity should show the difference.

2. The integrity of patterns. For instance, two patterns could be the same shape, but only if missing points were filled in. The intensity should be able to show the completeness of a pattern.

Based on the requirement of strength of a periodicity, we can find out that correlation between the original signal and local patterns could achieve both. In the next Section we will design some metrics to compute such a correlation in detail.

6.2 Metrics of Strength of Periodicity

Since periodicity is observed and significant in almost all lifelog data generated by human subjects, and we suppose we have identified those significant periodicities in the periodogram, we would like to use the lifelog to compute the strength of a certain periodicity for each participant at different points in time or as a whole. If we are to compute the intensity change through time, we can use a window and compare the local data with discovered global repeating patterns. The similarity between local data and global patterns is the intensity. So, the intensity could be a correlation or error between local data and the global pattern with a significant periodicity. We use the following denotation to explain how we calculate the strength of periodicity. F denotes the DFT of signal $x(n), n = 0, 1, ..., N-1$, and F' denotes the inverse transformation. S stands for the strength of periodicity. Suppose the periodicity we are concerned about is circadian periodicity. The specified periodicity can be changed according to different periodicities detected in real life. The metric we developed was calculated using six different approaches, described as follows:

$$\begin{aligned} \mathcal{A}_1(k) &= \frac{1}{k}\sum_{n=1}^{N-1} x[n]x[n+k] \\ \mathcal{A}_2(k) &= \sum_{n=1}^{N-1} x[n]x[n+k] \end{aligned} \tag{6.1}$$

Method 1:

$$\begin{aligned} S &= P(f) \\ P(f) &= \frac{1}{N}\mathcal{F}(x_n)^2 \end{aligned} \tag{6.2}$$

where: $f = \frac{1}{day}$

Method 2:

$$\begin{aligned} S &= P(f) \\ P(f) &= \frac{1}{N}\mathcal{F}(\mathcal{A}_1(x_n))^2 \end{aligned} \tag{6.3}$$

$$\text{where: } f = \tfrac{1}{day}$$

Method 3:

$$\begin{aligned} S &= P(f) \\ P(f) &= \frac{1}{N}\mathcal{F}(\mathcal{A}_2(x_n))^2 \end{aligned} \tag{6.4}$$

$$\text{where: } f = \tfrac{1}{day}$$

Method 4:

$$\begin{aligned} S &= max(P(f)) \\ P(f) &= \frac{1}{N}\mathcal{F}(x_n)^2 \end{aligned} \tag{6.5}$$

Method 5:

$$\begin{aligned} S &= \frac{1}{2}\sqrt{\sum_n (x_n - x'_n)^2} \\ x'_n &= \mathcal{F}'(P(f)), iff \neq \frac{1}{day}, P(f) = 0 \\ P(f) &= \mathcal{F}(x_n) \end{aligned} \tag{6.6}$$

Method 6:

$$\begin{aligned} S &= CC(x_n, x'_n) \\ x'_n &= \mathcal{F}'(P(f)), iff \neq \frac{1}{day}, P(f) = 0 \\ P(f) &= \mathcal{F}(x_n) \end{aligned} \tag{6.7}$$

where CC is correlation coefficient.

Method 1 uses the power carried by 1/day frequency as the strength of the circadian periodicity, namely, the correlation between signal and sinusoid with daily periodicity. Methods 2 and 3 use $\mathcal{A}_1$ and $\mathcal{A}_2$ to calculate autocorrelation, respectively. Using the result of autocorrelation as input to compute the periodogram, we thereafter use power of daily periodicity as strength of the circadian periodicity. It should be noted that $\mathcal{A}_1$ is normalized autocorrelation. Method 4 uses the maximum power in the periodogram to represent strength of periodicity, though in this case it is not assured that daily periodicity will carry maximum power all of the time. Method 5 calculates a sinusoid with daily periodicity that is most correlated to the data and then computes root mean square error (RMSE) between the signal and the most fitting sinusoid with daily period. Finally, Method 5 finds out a sinusoid with daily periodicity that is most correlated with the data, and then computes CC between the signal and the most fitting sinusoid with daily period.

If we consider the informal formulation of spectrum estimation as estimating how the total power is distributed over the frequency, the definition of intensity of periodicity can be thought of as the power corresponding to a certain periodicity or several periodicities. Method 1 comes directly from the definition of power spectral density, which uses Discrete-time Fourier transform (DTFT) to calculate how power is distributed over frequency directly and here in Method 1 we only take the power carried by 24-hour periodicity. Methods 2 and 3 derive from another definition of power spectrum which shows that spectrum can be achieved as the DTFT of the autocorrelation. Method 3 is normally used to calculate autocorrelation in signal processing. The reason we also use Method 2 is because when we lag signal to calculate autocorrelation, the bigger the lag is, the fewer the number of points that are involved in the calculation. Method 2 tries to eliminate this effect by using averaged value. Both Methods 2 and 3 use power of circadian periodicity as the intensity. Method 4 uses power of frequency with maximum power as intensity. The reason we use this method is that we are trying to see how the frequency with

maximum power would be correlated with biomarkers. In particular, it is interesting to see a comparison of the result between Methods 1 and 4. Method 5 uses a different way to calculate how the signal is different from the 24-hour periodicity. The rest of the methods use correlation as a metric to quantify the difference while Method 5 uses summed error as the metric to quantify the deviation.

6.3 Experiments

An experiment is firstly conducted on synthetic data before moving to real data. The data is modeled by the three factors described in Section 6.1 and intensity of periodicity is computed on both low-level and high-level accelerometer data.

6.3.1 Experiment on synthetic high-level data

In this section we conduct experiment on synthetic data. In order to generate synthetic data, we used the model described before and randomness is introduced by adding noise at different synthetic stages. Assume we are going to simulate N data points where N is a number which is 7 times an integral and we use this because in the last chapter, we discovered a strong weekly periodicity in high-level accelerometer data. The prior is used here and the 7-day pattern is a basic pattern which could compose the whole signal with different scaling and time shifting. The value of N is assumed to consist of local patterns with different smaller sizes. For easy understanding we assume the local pattern is also the integral times 7. The problem can be formalized as: assuming there is a signal $x_n, n = 0, 1, ..., N-1$. The equation to split the signal would be:

$$X_m = \sum_n x_n S_n \delta(m-n)$$

where S_t is the function used to split data, δ is a Dirac delta function, and m is the number of segmentations of the original signal. X_m is one segment of the signal. $|X_m|$ is the notation for the length of X_m, namely how many points in each segment. The size of each segment is sampled from uniform distribution. S_t is a function similar to a window function, but its function is to split the data into local patterns. If we are going to split the data into X_m segments, S_n is defined as:

$$S_n = \begin{cases} 1 & \text{if } n \text{ is in a segmentation} \\ 0 & \text{Otherwise} \end{cases} \tag{6.8}$$

Each segment X_m is composed of a global trend, local trend and local patterns with missing probabilities. The following settings are used to generate each segment of synthetic data:

1. The size of a segment is firstly chosen from 2 to 10 uniform distributions. The chosen number will then be multiplied by 7 in order to make a week as a unit of the local pattern. 50 segments are generated to simulate high-level lifelog data.

2. At the stage of generating local and global trends, a linear model $X_m = wn+b$ is chosen as the trend. Parameters w and b from the local linear model are chosen randomly from a uniform distribution range from -0.005 to 0.005 plus Gaussian noise with 0 mean and 0.3 std to the output of the linear model. Parameters w and b of the global linear model are chosen randomly from a uniform distribution range from -0.0005 to 0.0005 plus Gaussian noise with 0 mean and 0.3 std to the output of the linear model.

3. The local patterns chosen, is a pattern with a weekly repeating period. For example, activities can be done every Monday for 4 weeks. In order to intro-

duce randomness into the model, the day when the activity is repeated can be randomly chosen from a uniform distribution, which means that each day has equal probability of doing the activity and the day chosen can be based on a Gaussian distribution, which means activities are more likely done on a certain day. In reality, the prior of an activity being done on each day could be useful, but here we do not want to make any strong assumption. The quantity of the level of activity is selected from a uniform distribution with range from 2 to 7, plus normal distribution with 0 mean and 0.5 std. The probability of missing some activities within each segment is randomly chosen from a uniform distribution range from 0 to 1. For example if the probability of missing an activity within a segment could be 0.2. so we randomly set 20% of activities done in that segment, to 0.

4. The intensity of each week will then computed using the quantity of each activity level multiplied by the probability of missing the activity.

The reason to select the above mentioned settings is based on the assumption that a participant have each training plan/segmentation last from 2 to 10 weeks. The choice of local and global linear model parameters is to introduce trend but minimize effect of trend, especially global trend, to actually training performance. Within each training plan/segmentation we assume the subject is doing one exercise every week and the performance of each exercise range from 2KM to 7KM. The day of which exercise is done is randomly chosen within each training plan/segmentation and has random chance of missing weekly exercise. To compare the fact intensity we used to simulate lifelog data and the computed intensity with different proposed methods, firstly let us have a look at the periodogram from the example simulated data. Figure 6.1 shows the randomly generated global and local trend data with Gaussian noise with zero mean and 0.3 standard deviation.

Figure 6.2 shows the result signal with global and local trends, and local patterns

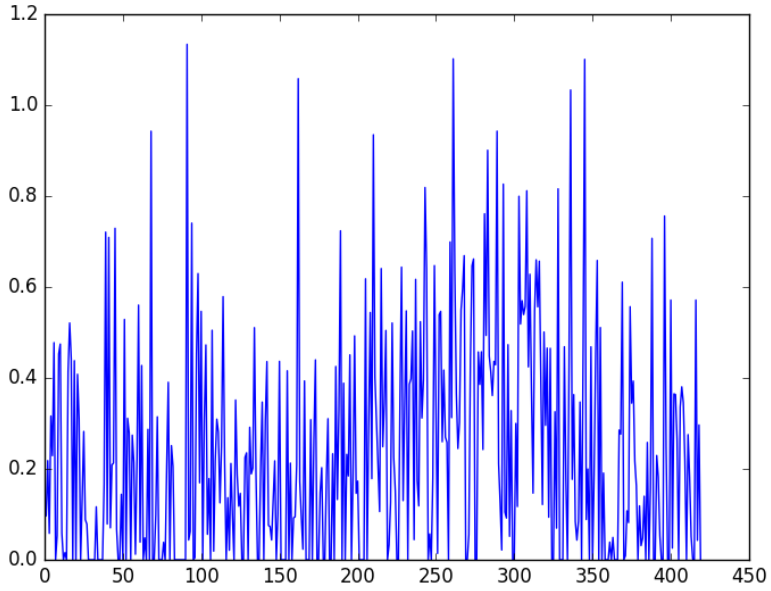


Figure 6.1: Global and Local Trend with gaussian noise

with local patterns. The local patterns are activities that repeat weekly and the repeating day of doing the activity is chosen randomly from a uniform distribution, and the chances of missing activities within a local pattern is randomly chosen from a uniform distribution as well. Figure 6.3 shows the periodogram of the final simulated data. Interestingly, we can see that the periodogram is similar to the real data generated by the triathlete in Fig. 5.14. The similarity refers to high energy carried at low frequency and two harmonics following a weekly periodicity. The similarity of simulated data and real life data could possibly provide an explanation for the correlation between the time domain signal and spectrum. The power of spectrum in the lower frequency, namely on the left side of the periodogram, indicates the trend of the activity. We can also observe on both periodograms, harmonics on 7-day, 2/7-day and 3/7-day frequencies. It is legitimate to hypothesise that this is caused by regular weekly activity conducted on different days of a week, because this is how we generated the simulated data.

Figure 6.4 indicates the result using the 6 different methods to compute that we

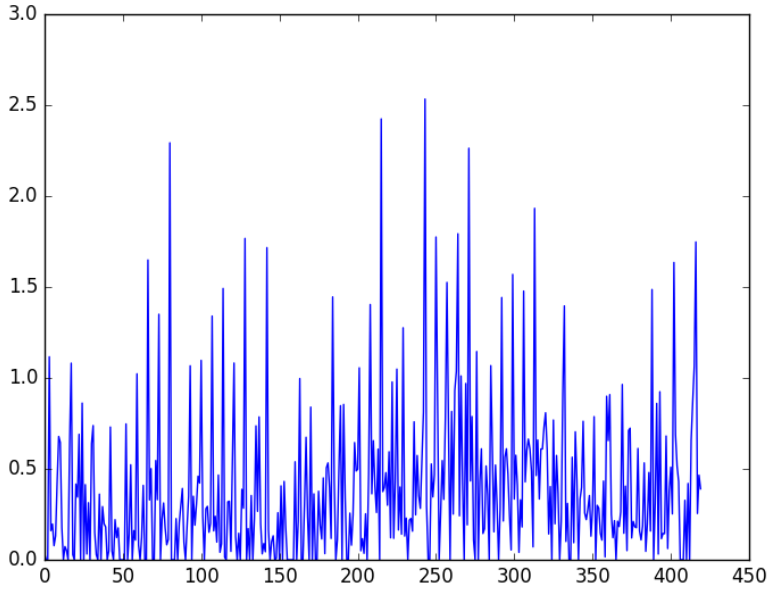


Figure 6.2: Signal with Trends and local patterns

introduced in Section 6.2 and the result is compared with ground truth intensity used to generate the data. We can observe in the result that some methods tend to enlarge the peak while some methods are quite close to the ground truth. Since this is just one run of random generated data, there are chances that this run might not reveal the true effect of different methods. In order to evaluate those metrics with more confidence, we will run the randomly generated data for 1,000 iterations and the result will be shown in a table. The performance of those proposed metric methods are evaluated with correlation coefficient (CC) and root-mean-square error (RMSE) as measures.

Table 6.1 shows the result of the performance of the different methods, measured by CC and RMSE between computed intensity using the previously mentioned methods and the intensity used to generate the lifelog data. In the table the window is fixed to 35 days with zero overlapping. We compare the result using two different distributions of days, namely normal and uniform. The choice of normal distribution is to simulate the situation that most of the occurrence of an activity is on one

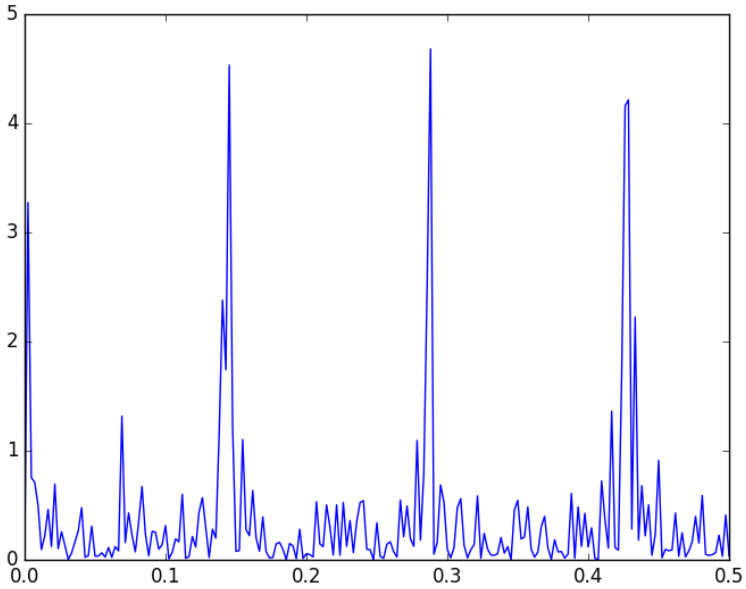


Figure 6.3: Periodogram of simulated data. X axis is frequency. Y axis is energy.

day and the distribution of the other days follows a normal distribution. The choice of uniform distribution corresponds to the scenario that activities could be done on any day of a week with equal probability. Not surprisingly, normal distribution has better performance than uniform because uniform distributions introduce yet more uncertainties. Based on the periodogram we have using real life data, we tend to think that normal distributions are more common than uniform distributions but because a 7-day periodicity normally carries the highest energy than its harmonics, this is the case using normal distribution.

The definition of intensity of periodicity can be thought of as the power corresponding to a certain periodicity or several periodicities. In this case, method 1 comes directly from the definition of power spectral density, which uses DTFT to calculate how power is distributed over frequency directly. Both Methods 2 and 3 use power of circadian periodicity as the intensity. They derive from another definition of power spectrum which shows that spectrum can be achieved as the DTFT of the autocorrelation. The difference between method 2 and 3 is that they use

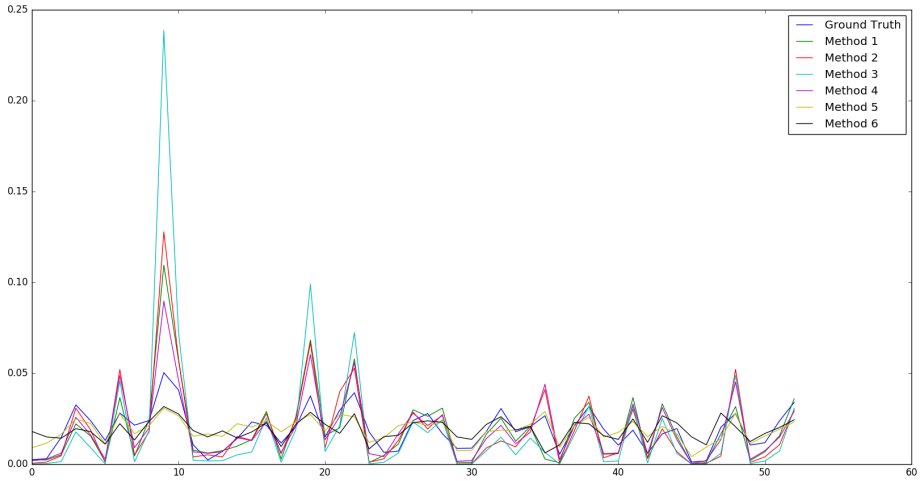


Figure 6.4: Intensity computed with different methods. X axis is time. Y axis is periodicity intensity.

different ways to compute autocorrelation. Method 4 uses power of frequency with maximum power as intensity. In particular, it is interesting to see a comparison of the result between Methods 1 and 4. Method 5 uses a different way to calculate how the signal is different from the 24-hour periodicity. The rest of the methods use correlation as a metric to quantify the difference while Method 5 uses summed error as the metric to quantify the deviation. Method 6 is similar to method 5 but using correlation as the deviation from the pattern signal.

Figure 6.5 and Figure 6.6 show the correlation coefficient between computed intensity levels using different metrics and the intensity used to generate the lifelog data. The X-axis is the window size with a unit of a week. Please note that the starting of window size is 1 week which corresponds to 0 on the X-axis. We can see that the CC changes with the window size. The CC is increasing at the beginning with increasing of the window size but at a certain point, the CC is not improving as we continue to increase the window size. For some metrics, the performance is even decreasing. Figure 6.5 used normal distribution to generate the day on which activities would be conducted. In this setting, the performance of CC tends to

Table 6.1: Performance of 6 methods as measured by CC and RMSE using a 35-day window

		CC		RMSE	
Distribution	Method	Mean	Std.	Mean	Std.
Normal	Method 1	0.791	0.053	0.0015	0.0002
	Method 2	0.757	0.057	0.0017	0.0002
	Method 3	0.725	0.056	0.0027	0.0005
	Method 4	0.827	0.044	0.0011	0.0002
	Method 5	0.740	0.059	0.0009	0.0001
	Method 6	0.677	0.083	0.0009	0.0002
Uniform	Method 1	0.747	0.063	0.0017	0.0003
	Method 2	0.763	0.059	0.0017	0.0003
	Method 3	0.700	0.062	0.0029	0.0006
	Method 4	0.832	0.042	0.0011	0.0002
	Method 5	0.741	0.059	0.0008	0.0001
	Method 6	0.576	0.098	0.0011	0.0002

converge at window size of 4 or 5 weeks. Figure 6.6 uses uniform distribution to generate the day on which activities would be done. A similar trend can be observed in both figures. We can observe that the performance of method 6 is always the worse among all the methods. We suspect that it is the summed absolute values like RMSE that matters when computing the deviation of real data from patterns. That leads poor result of CC here.

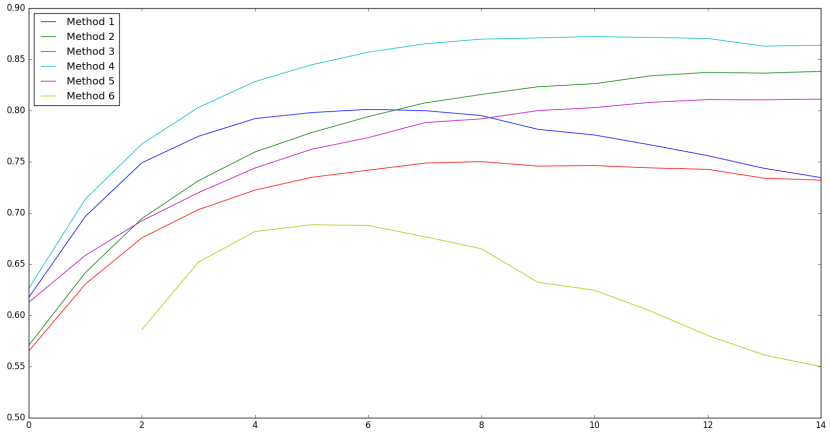


Figure 6.5: CC trend with window size changing using normal distribution. X axis is window size. Y axis is CC between estimated intensity and ground truth.

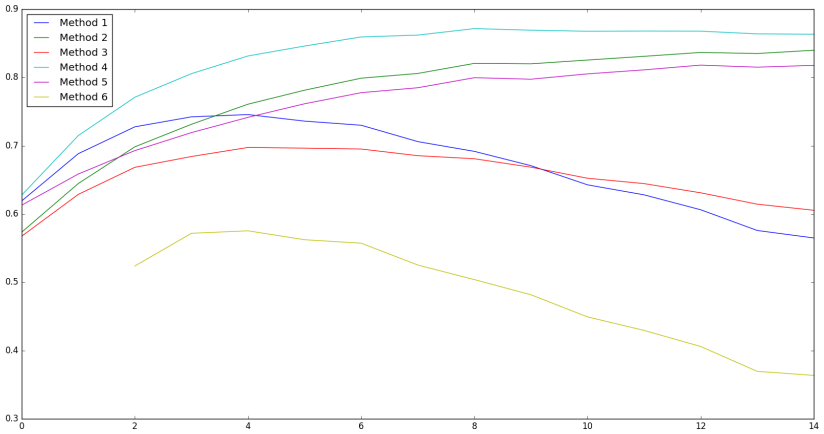


Figure 6.6: CC trend with window size changing using uniform distribution. X axis is window size. Y axis is CC between estimated intensity and ground truth.

6.3.2 Experiments on the ASU data

In this section we use the ASU data which we introduced in chapter 2. To explain the results we obtained more concretely, we use one participant, number 102, shown in Figure 6.7 which outlines the methodological steps for identifying periodicities and visualizing periodicity strength that we followed. From the top to bottom of that Figure we see 4 sub-figures, and we name these Panel A, B, C and D correspondingly. Panel A provides a visualization of the sum of vector magnitudes (at 1 min epochs) along the Y-axis and time along the X-axis over the course of the monitoring period. Sleep and wake periods are evident visually from these data, which runs for 12 weeks of data collection. The raw data from the accelerometer was summarized to 60 second epochs using the gravity-subtracted sum of vector magnitudes. This plot of the overall activity levels illustrates several isolated periods of high activity, probably exercise of some form, throughout the 12-week period of data logging but there is no evidence of changes in behavior.

Panel B displays a periodogram calculated from 1 minute epochs. Here, the X-axis is frequency and Y-axis is energy of the frequency, namely, how strong the

corresponding frequency is. In this panel, we observe strong circadian periodicity followed by a 12 hour periodicity. This shows a reasonably strong energy level around the 1-day point and a smaller peak at around the 12-hour point which is the harmonic of the circadian. No within-day or weekly patterns were observed. Compared to some of our other participants who have gathered similar data, the regularity of this individual's daily cycle is not particularly strong for the whole of the 12-week period, suggesting that s/he may work shifts or just have a very disorganized and irregular lifestyle.

Panel C plots time (X-axis) by the strongest periodicity observed over the 3-day time lagged window. The Y-axis of panel C is the frequency that carries maximum power within a window. In this example, the 24 hour periodicity held consistently for the majority of 3-day windows with small breaks at the beginning of the monitoring period.

Panel D describes the strength of the periodicity using Method 1 (Y-axis) over time (X-axis). The strength/intensity of the 24 hour circadian periodicity changes throughout the lifelogged observation period, showing, for example, a weaker period of regular circadian cycle from day 0 to day 14 and again from day 32 to day 44.

The raw acceleration data is firstly summarized using 1s epoch in order to capture information from all 3 axis acceleration over 1s time. After the raw data is summarized, we then computed periodogram to detect periodicity of 1s epoch data. Although different way to summarize the data may have impact on the result of periodicity detection, but the refinement of periodicity detection is not included in the research questions. After the periodicity is detected, we further examined which frequency carrying maximum energy to provide some extra insight for the subject. In the example we showed in the Fig. 6.7 we can see the subject is not very regular in terms of circadian periodicity. By applying method 1 we can then plot intensity of circadian periodicity. We also computed the intensity of the accelerometer data using Methods 1 to 5 for every subject and the detailed result will be introduced in

next chapter.

6.3.3 Experiments on Athletic data

The metric using Method 1 with a 28-day window and zero overlapping is used to show the example results on the Athletic data, which we introduced in chapter 2. Figures 6.8, 6.9, 6.10 and 6.11 are the results of such settings. The result of the intensity is normalized with the sum of the intensity data series. The reason for selecting a 28-day as window is because this window size showed some good quality in the simulated experiment described earlier in Section 6.3.1. The result of processing this data set will then be summarized into questions and those questions will be then used in an interview with the athlete who generated those data. The detailed results will be described in next chapter.

In Figure 6.8 we are able to see several running patterns in the intensity graph. For example there are peaks at the beginning of 2004 and 2006 and at the end of 2008 and beginning of 2009. Generally speaking, the intensity is increasing until 2007 where it maxes and then starts to decrease. In Figure 6.9 we show the intensity of weekly periodicity for cycling. We are able to see the subject started to cycle from 2007 onwards and the intensity level increases until 2014 when there is a sudden decrease. In the middle of 2010 there is very high intensity. Figure 6.10 shows fewer changes compared with running and cycling. But still we can see peaks at around the beginning of 2008, the end of 2009 and the end of 2012 or beginning of 2013. In aggregated data shown in Figure 6.11, peaks seems repeated with a period of 2 years.

Since in the previous experiment we have shown that there is a strong weekly periodicity in the athletic data, we focus on computing the intensity of weekly periodicity of this dataset. The weekly periodicity intensity graph for running, swimming, cycling and aggregated data show a oscillating effect consisting of numerous peaks

and valleys. Those changing places are especially interesting for us to further investigating to confirm the validity of intensity graph, which will be used in qualitative analysis.

6.4 Conclusions and RQ4

In this chapter, we further studied the intensity of periodicity in lifelog data. We proposed a model to simulate lifelog data and we used the generated data in our first experiments. We proposed several methods to calculate the intensity of periodicity. Selected methods were executed on real life high-level and low-level accelerometer data.

In the model we proposed to simulate high-level lifelog data, we considered factors that we think might be close to the real data we collected. The factors are global trends, local trends, local patterns, and the probability of missing activities not being logged or recorded. Each of the factors is determined by further parameters with introduced randomness deliberately. We use known intensity which is also randomly generated, to simulate high level lifelog data. Because of this, we are able to compare the performance of different metrics with the intensity used to generate data. Not only can we see the result of different metrics, but also we can observe how the changing of window size can affect the performance of those metrics. We firstly compared the periodogram of simulated data and real high-level lifelog data, and discovered the similarity between them. Local and/or global trending tend to affect the low frequency, which is indicated on the graph as the high significant peaks on the left side of the periodogram. Also, the distribution of the day on which activity is conducted can affect the harmonics in the periodogram. If the distribution of the day of the activity is uniform, that is to say that the activity is likely to happen on any day of a week with equal probability, the harmonic tends to have the same level of significance of power. On the other hand, if the distribution is normal, that is to

say activity is more likely to happen on a certain day, the 7-day period tends to carry more energy than 3.5-day or 2.3-day period harmonics. In terms of a periodogram, we have achieved a similar periodogram as for the real data, and also the simulated data provided some explanation of harmonics which we observed in real lifelog data not once but repeatedly.

Different methods performed on simulated data are able to achieve similar results, except method 6 is always the worst method compared with the rest. In general, methods 1 - 5 show quite correlated results compared with the intensity used to generate the lifelog data for both uniform distribution of days and normal distribution of days. In the experiment to find out the correlation between window size and performance, we explored different window sizes. We discovered that the performances are increasing rapidly at the beginning, but after window size exceeded ca. 4 weeks, the improvement in the performance tends to be stable, and for some methods even to decrease.

For experiments on real life data, we conducted experiments on ASU data and on Athletic data. The experiment is conducted based on a pipeline. The first step is to compute 1 second epochs from the X, Y and Z axis, and a periodogram is computed to reveal periodicity. Then, we define a window with fixed window size and step size to slide through all the data. Within each window, we extract the period that carries maximum energy. At the same time, we also compute the intensity of the circadian periodicity within each window. As we can see, the figures are able to reveal the change of intensity. But a question remains, does the intensity of periodicity have realistic use? Or, more precisely, what are the underlying facts about participants that the intensity of periodicity is trying to tell? We will investigate this in the next Chapter.

The experiments conducted on Athletic data have also shown some of the intensity of each sport event that the triathlete is changing. We need to use some of the significant changes found in the intensity graph, as feedback to the subject. There

will be an interview asking the athlete about those abnormalities in the intensity graph. The results of this interview will also be revealed in the next chapter.

In this chapter, we proposed several methods to measure the intensity of periodicity, and the experiments on synthetic data help us to understand where the harmonics in the periodogram come from and how the peak on lower frequencies in the periodogram, are generated. The result of different methods are compared using the synthetic data and the result is generally good. Experiments conducted on real lifelog data have suggested that further exploration with the subject involved in the experiment, is needed.

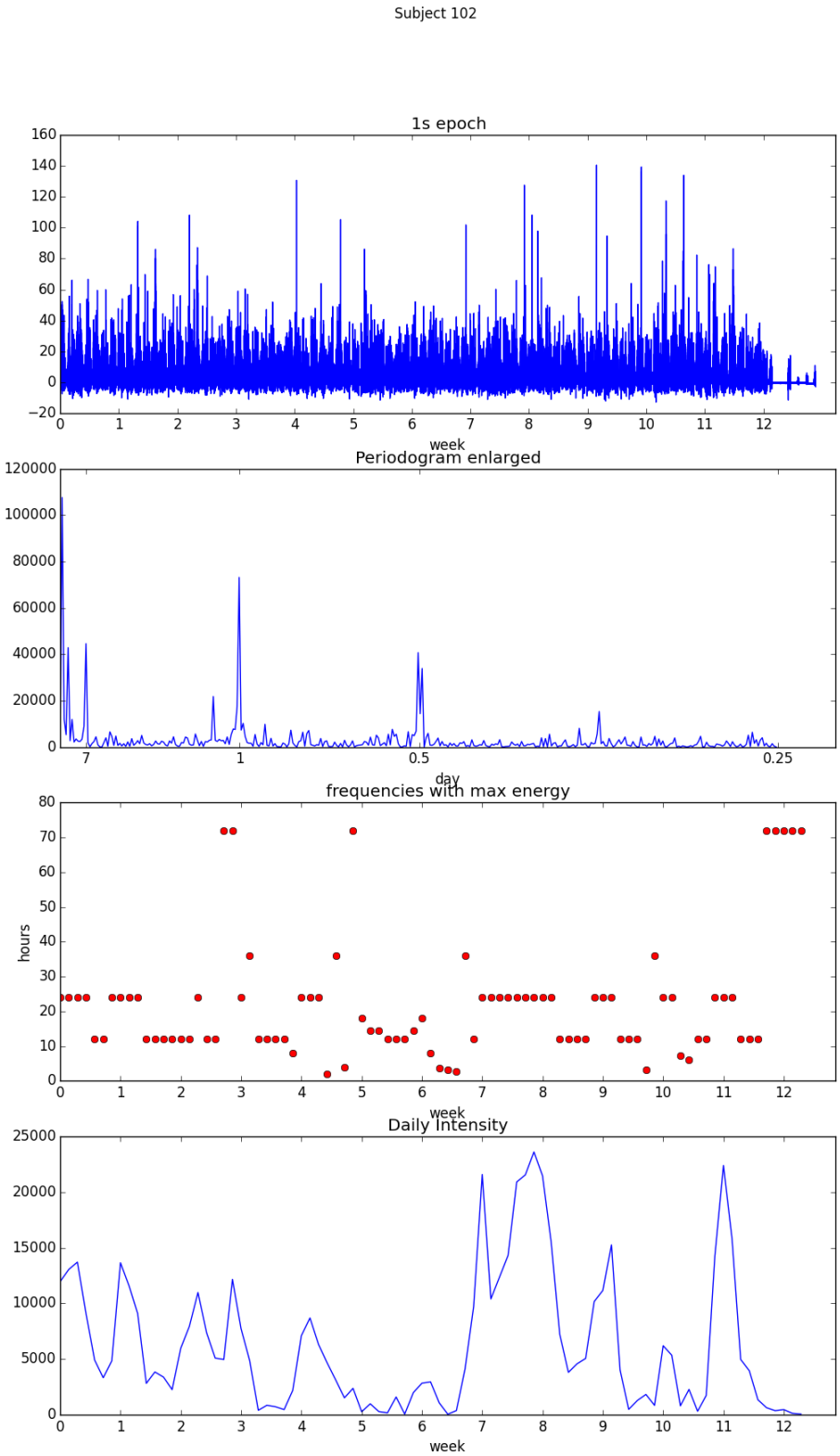


Figure 6.7: Result for Subject 102. Panel A provides a visualization of the sum of vector magnitudes. Panel B displays a periodogram calculated from 1 minute epochs. Panel C plots time (X-axis) by the strongest periodicity observed over the 3-day time lagged window. Panel D describes the strength of the periodicity using Method 1 (Y-axis) over time (X-axis)

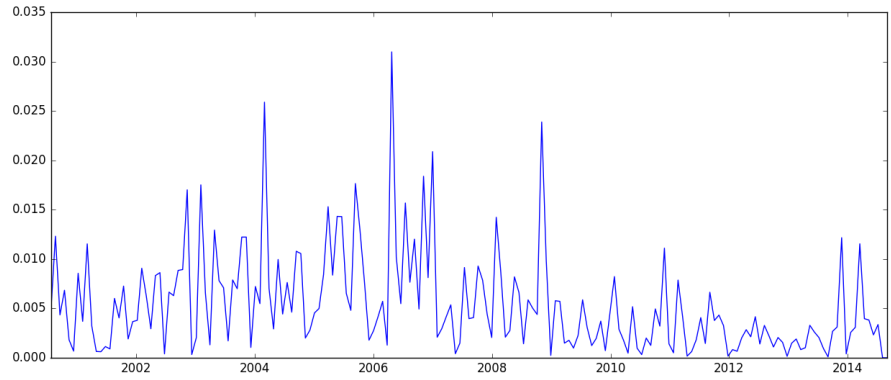


Figure 6.8: Intensity Graph for Running Data. X axis is time. Y axis is periodicity intensity.

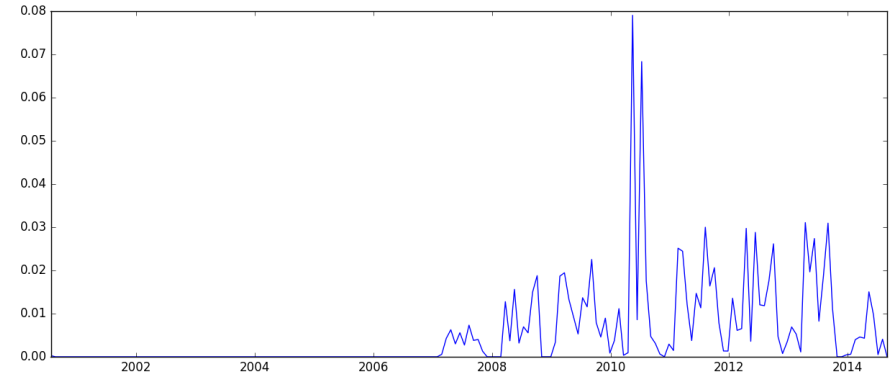


Figure 6.9: Intensity Graph for Cycling Data. X axis is time. Y axis is periodicity intensity.

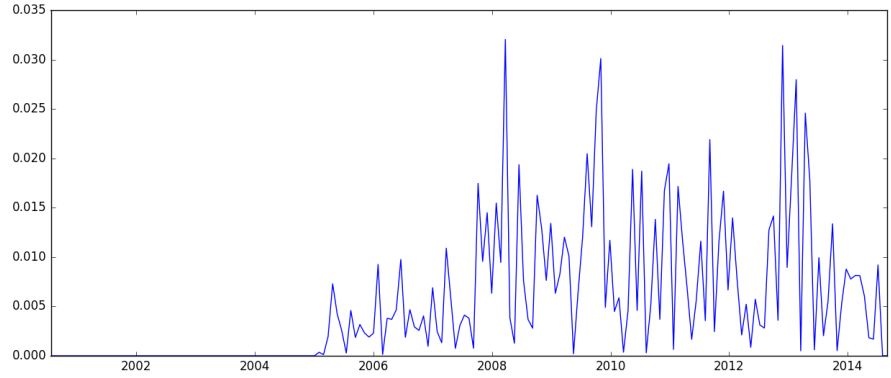


Figure 6.10: Intensity Graph for Swimming Data. X axis is time. Y axis is periodicity intensity.

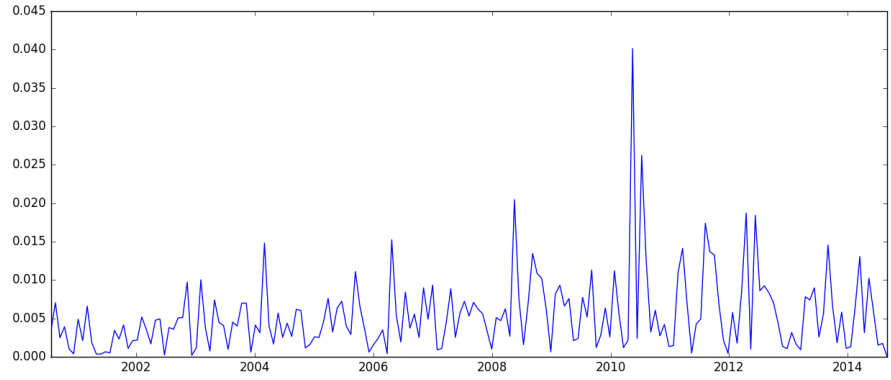


Figure 6.11: Intensity Graph for Aggregated Data. X axis is time. Y axis is periodicity intensity.

Chapter 7

Case Studies

In carrying out some case studies on our data analysis, we are trying to illustrate the usefulness of the computed patterns identified from user data in the previous chapters. The first case study is an interview with the triathlete from whom the high-level lifelog data was collected. The questions used in the interview which form part of the case study are generated by using the detected periodicity and intensity of periodicity with different window sizes and various step sizes. The second case study we focus on is the low-level ASU accelerometer data. The question we are trying to address in this case study is how does the intensity or strength of periodicity correlate with the subjects who were involved in the experiment. The methods we developed to compute periodicity intensity have been shown in the previous chapter to have high correlation coefficient using simulated lifelog data and in this chapter we will see the correlation between results of those methods and several biomarkers collected from the subjects.

7.1 Case Study: Athletic Data

In order to determine what questions to be asked in an interview with the subject who generated the data, periodograms and graphs of intensity of periodicity need to

be generated first. After the response to the questions is answered by the athlete, the results of the interview were studied in order to find a correspondence between computed graphs and the actual training intensity or regularity. We would also like to see if the mined result would potentially reveal some facts that the athlete would have ignored or not known about.

7.1.1 Graphs of Intensity

The data from which intensity graphs are computed are originated from the athletic data which consists of running, cycling, swimming and aggregated data. These 4 different sub-datasets are all time series data which are sampled daily. The corresponding units for running, cycling, swimming and aggregation are miles, kilometres, laps and MET. Power Spectral Density is calculated to determine the frequencies that carry significant energy. In other words, we conducted PSD calculations on the whole data length (more than 10 years) in order to detect periodicities that are important. After determining the significant frequencies, we can observe a common phenomenon. Energy carried by sine/cosine functions with weekly periods are significant in all 4 subsets. Therefore, we generate intensity graphs based on weekly periodicity. The intensity graphs show the strength of weekly periodicity, changing with time. Different window sizes and overlapping sizes are chosen to calculate the intensity graph, namely window sizes of 14, 28, 70, 28 corresponding with overlapping sizes of 7, 21, 63, 0. The unit for both window and overlapping sizes, are one day.

Figures 7.1, 7.2, 7.3 and 7.4 show the results of running, cycling, swimming and aggregation respectively, generated by using different window and overlapping sizes. The abnormalities, namely peaks and valleys, are the special interests that we may raise questions from.

In order to investigate the subjects' interpretations of intensity graphs, we de-

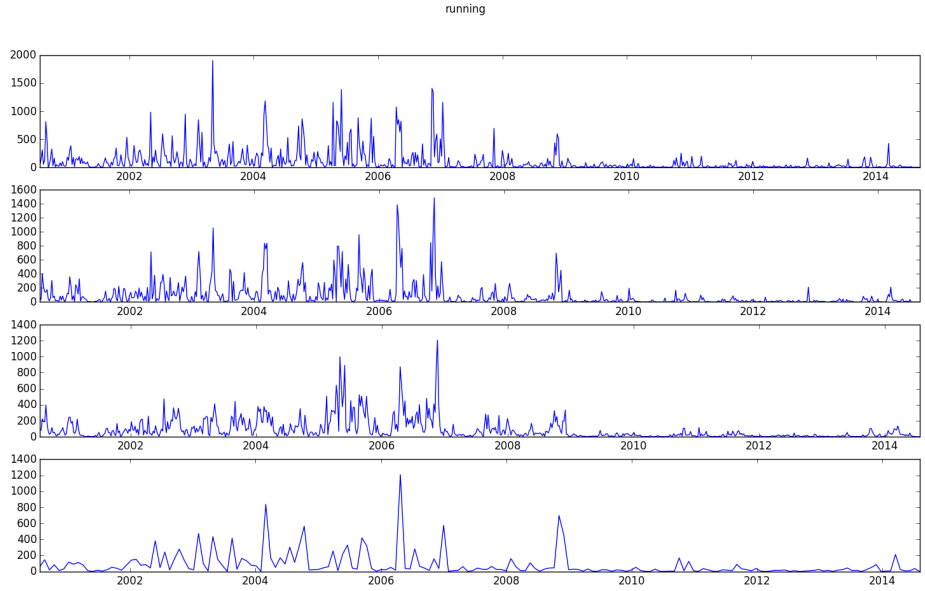


Figure 7.1: Graph of intensity of periodicity for running. Panel A is intensity graph computed with window size 14 overlapping size 7. Panel B is intensity graph computed with window size 28 overlapping size 21. Panel C is intensity graph computed with window size 70 overlapping size 63. Panel D is intensity graph computed with window size 28 overlapping size 0. X axis is time and Y axis is periodicity intensity.

signed and conducted an interview with the subject who generated the data. The interview would help us to understand how well the algorithms could be used to generate intensity graphs, and at the same time we cannot interpret the intensity graph without building a connection with the personal experience of the subject. In other words, we could confirm some of the results coming from the computation of the PSD/periodogram and interpret the intensity using personal response as input.

As we can see from the figures, there are many peaks and valleys that we are interested in. In order to achieve the objective, we pay attention to these peaks and valleys in the intensity graphs. The reasons for this are:

1. Peaks correspond to high regularity that the subjects might be aware of and/or keep on record. For example if you look at Fig. 7.1, there is a peak between

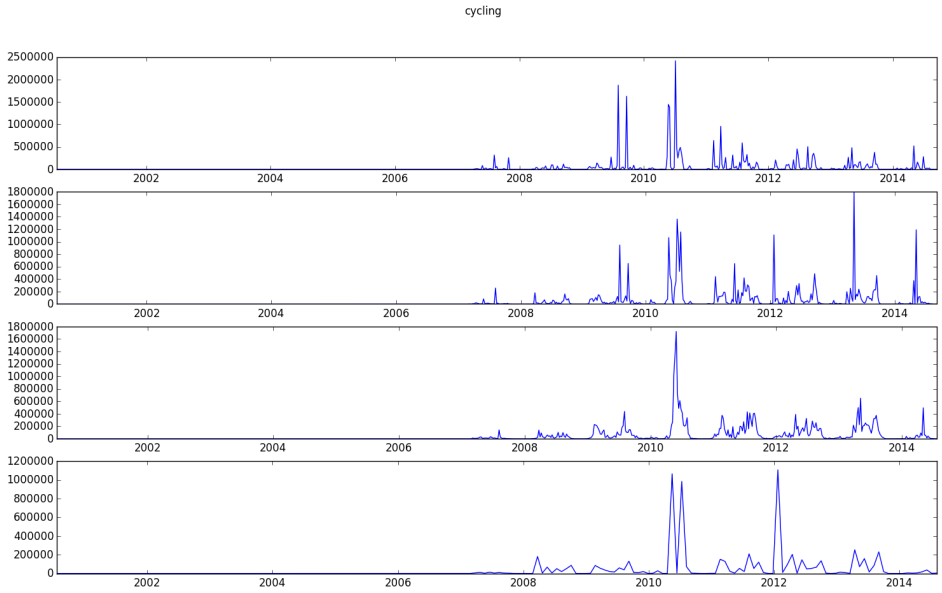


Figure 7.2: Graph of intensity of periodicity for cycling. Panel A is intensity graph computed with window size 14 overlapping size 7. Panel B is intensity graph computed with window size 28 overlapping size 21. Panel C is intensity graph computed with window size 70 overlapping size 63. Panel D is intensity graph computed with window size 28 overlapping size 0. X axis is time and Y axis is periodicity intensity.

2008 and 2010, and the intensity before and after are very low, that make this peak interesting, and we will ask question about the subject what happened at the peak.

2. Valleys correspond to breaks/injury/tiredness/travel/holidays in the training which the subject might remember and/or keep on record. For example in Fig. 7.1, there is a valley at end of 2005 and beginning of 2006, we guess something might happened to stop the athlete from training.

Based on the idea that peaks and valleys are important for the triathlon athlete and the periods to which those peaks and valleys correspond could make sense for the subject from his/her own perspective, we designed our interview around the peaks and valleys in the intensity graph. Due to huge amount of work to identify every

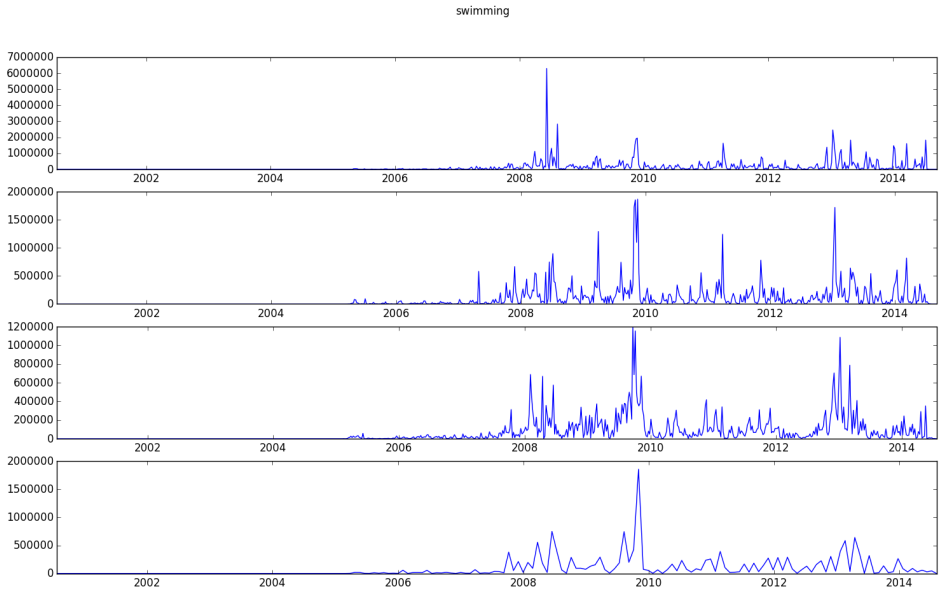


Figure 7.3: Graph of intensity of periodicity for swimming. Panel A is intensity graph computed with window size 14 overlapping size 7. Panel B is intensity graph computed with window size 28 overlapping size 21. Panel C is intensity graph computed with window size 70 overlapping size 63. Panel D is intensity graph computed with window size 28 overlapping size 0. X axis is time and Y axis is periodicity intensity.

single peaks and valleys from intensity graph, we pay attention to those peaks and valleys appear on larger window size first, because we hypothesis that the athlete might have better memory about those changes in training last longer.

7.1.2 Questions

In order to raise questions against the discovered patterns, we will describe the reasons why we have these questions and what questions will be asked in the interview. All the questions will be categorized into three types: 1) short period 2) long period 3) general questions. The first type is related to intensity changes within a short period of time such as a sudden increase or decrease of intensity. In the section we refer this as a 'point' change. The second type of question asks about the longer

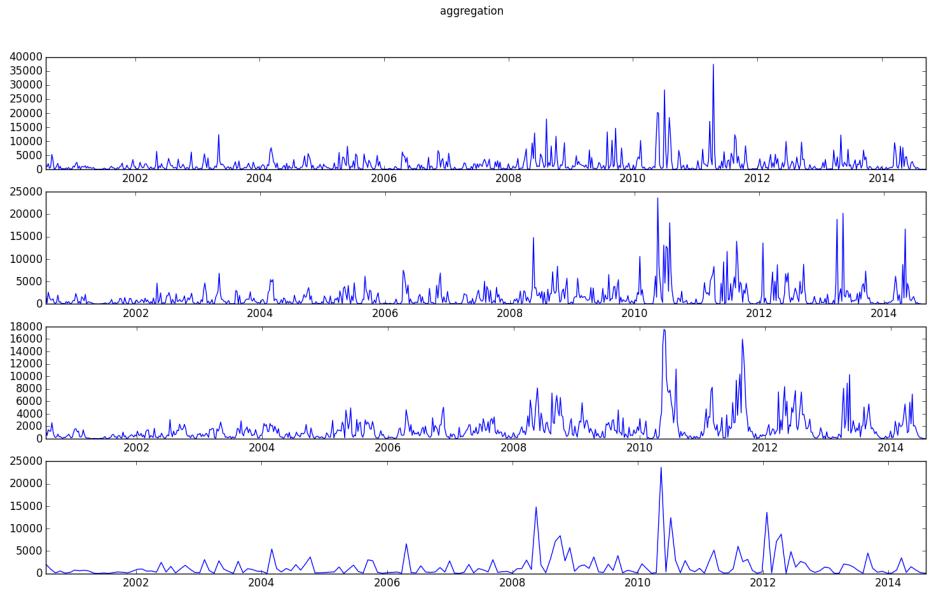


Figure 7.4: Graph of intensity of periodicity for aggregate data. Panel A is intensity graph computed with window size 14 overlapping size 7. Panel B is intensity graph computed with window size 28 overlapping size 21. Panel C is intensity graph computed with window size 70 overlapping size 63. Panel D is intensity graph computed with window size 28 overlapping size 0. X axis is time and Y axis is periodicity intensity.

term changes such as interesting patterns within a three-month period. We will ask the subject to explain the changes s/he might think of. We refer to these types of questions as 'interval' change. The last question type refers to some general questions regarding general patterns that the subject has, such as training schedules. Although we are only able to put resized Fig. 7.1 to Fig. 7.4 due to limited of size of pages, in fact while generating those questions we can view enlarged intensity graph. That means the scale of X-axis which represent date, can be seen with greater granularity. So that we can locate the corresponding months to peaks and valleys on intensity graph.

The following questions concern running data:

1. General questions

(a) Ask the subject to confirm detected periodicity.

(b) Ask the subject to segment his/her running training.

2. Interval change

(a) Why was running not regular from beginning of running to April 2002 ?

(b) Ask the subject to explain the high regularity between May 2002 to October 2003 ?

(c) The intensity repeated about every 3 months between May 2002 to October 2003. Ask the subject to explain those oscillating peaks during this period of time ?

(d) What happened between December 2004 to March 2006 related to low intensity ?

3. Point change

(a) Why was running getting regular from the start of 2002 ?

(b) What happened in May 2002 ?

(c) What happened in February 2003 ?

(d) What happened in May 2003 ?

(e) What happened in March 2004 ? It's very regular.

(f) Why is it declining in March 2004 ?

(g) What happened in October 2004?

(h) What happened in April and May 2005?

(i) What happened in April 2006

(j) What happened three months later in July 2006 ?

(k) What happened in January 2007 ? It's a peak.

(l) Did you stop running after March 2007 ?

(m) What happened on November 2008 ?

(n) Did you stop running after November 2008 ?

(o) Was the running after November 2008 regular ?

(p) Regularity is increasing from the end of 2010. What's the reason for that ?

The following questions concern cycling data:

1. General questions

   (a) When did you start cycling ?

   (b) Do you have schedule for cycling ?

2. Interval change

   (a) Did you stop cycling between May and June 2010 ?

   (b) Did you follow same pattern in 2009 and 2011 ?

3. Point change

   (a) It was regular in August 2009. What happened ?

   (b) What happened in September 2009 ?

   (c) What happened at beginning of 2012 ?

   (d) What happened in May 2013 ?

   (e) What happened in May 2014 ?

The following questions concern swimming data:

1. General questions

   (a) When did you start swimming ?

(b) Were you aware of the fact that you run less while swimming more ?

(c) Ask the subject to segment swimming in the past 10 years.

2. Interval change

(a) Were you regular at the beginning ?

(b) From year 2008 to year 2010 would you call it regular training of swimming ?

(c) In general, would you say the regularity of training is increasing ?

3. Points change

(a) What happened in January 2008 ?

(b) What happened in the end of 2009 ?

(c) What happened in February or March 2008 ?

(d) What happened at the end of 2012 or beginning of 2013 ?

The following questions concern aggregated data:

1. General questions

(a) Is the total regularity of training increasing over time ?

(b) Do you think your volume of training is increasing ?

2. Points change

(a) What happened in May 2008 ?

(b) What happened in May 2010 ?

(c) What happened in June 2010 ? It seem to be decreasing fast.

(d) What happened in April 2011 ?

(e) What happened in April 2013 ?

7.1.3 Running Responses

General questions:

1. There are weekly periodicity and several harmonics detected in the periodogram. The subject responded positively and said "In general over the decades, I would run reasonably regularly", which corresponds to detected periodicity, And s/he further explained that "I have a fixed time of the day, that's how I manage to get the regularity. So it's the fixed time of the day during the weekdays or at the weekends". Also the subject denied a deliberate training plan. The athlete said: "No. I never had weekly or monthly plan or a long-term plan. I just do run on a day-to-day basis. I may have targets or may go to run at race or to do events, and then I increase the training in the month beforehand., but I don't plan out a schedule for this, or this is what I am going to do this week. I have targets in my mind of how and what would I like to do or should do. And I try to reach these targets".

2. The subject was asked to segment his/her running. The athlete tended to split into 4 different phases, s/he said "You would have to group them into 4 which is prior to taking up swimming, prior to taking up bike, and then most of recent years I am doing all three. Those are the phases of my life. When we build the pool where I work was when I started to swimming and then I changed to that. And then when I got the bike when I sort of doing triathlons, then I changed again. And then around about that period when I started to doing things that blended different sports you know they changed again. So those tipping points in my life are explained by the dairy entries".

Interval change:

1. On the intensity graph it seems the intensity was kept at a low level until April 2002. But the subject couldn't provide detailed information from beginning of running to April 2002 to explain the intensity graph.

2. From May 2002 to October 2003, the intensity is oscillating with a period around 3 months. The subject explained that it might because of the travel he/she did. The subject confirmed that he/she would try to run during trips. She/He said "But it's more difficult, because it depends on where you are". And the subject said "I would try. I hate using a gymnasium. I wouldn't use them and that's absolutely". The subject thinks the valleys of the oscillation in intensity graph related to long haul trips every 3 months during that period. She/He said "I'd say every 3 months I am doing a long trip. Every 3 month I am between one long trip and next long trip".

3. From December 2004 to March 2006, we can see an oscillation on the intensity graph. The subject said he/she "was in Australia for a part of that month" (December 2004) and back in January 2005. The subject confirmed that "there weren't that much travels for that calendar year 2005. And when there were travels, there were travels where it's easy to do runs". And the subject said the regularity ended in March 2006 because "there was a couple of holidays".

Point change:

1. From the start of 2002, the intensity gets higher than at the beginning on the intensity graph, the subject said "There were no major events or races or anything to explain any deviation in that". He/she did confirm that "I started to do some weekly gym sessions and that probably made my overall training for everything more regular".

2. There is high intensity at around May 2002. The subject reported that there is "Nothing special. But the mileage was quite a lot at that period May 2002".

3. There is a small peak at around February 2003. According the to subject "I was doing not lot of miles but I was doing shorter faster 600 meter 800 meter races and interval training at that period. So the training was quite intense"..

4. Further peaks reside at around May 2003. The subject reported it's "big mileage then. Yeah 40+ something a week. But between those months, between February and May I was off for an entire month without anything training because of travel". Interestingly, the months that the athlete are off correspond to valleys on intensity graphs during this period of time.

5. There is a high peak at around March 2004. The subject responded that "The training was intense because I was doing a lot of miles and I was also in the gym".

6. After March 2004, the intensity graph declines suddenly and the subject reported it's because of travel. S/he confirmed "it was a decline overall, because there is no regularity, because of the travel schedule".

7. At round October 2004 the intensity went back to ahigh level. The subject confirmed that the regularity went back and he/she went to gym "twice a week at that point".

8. At some time around April and May 2005 the intensity starts to go up. The subject mentioned that the swimming pool opened in February. But he/she didn't mention regularity change in April for running. She/he mentioned for swimming "That was April and May and then I guess what happened by May. I started to getting used to it, because during those two months there were entries there".

9. There is a high intensity peak around April 2006. The subject responded that "April was a period of not too much running activity at all. And then I started getting back in May". The subject confirmed that "I started getting back in May".

10. There is a small peak at around July 2006. The subject said "There is a bit travel. And running was good I think".

11. After March 2007 the intensity graph shows low level for a long time, the subject explained " That's when I got a bike". The subject said " And the thing about the cycling and running is that the running reduced in terms of its volume or the miles completed and was compensated by the increase in the energy spent on cycling".

12. We can see a peak at around January 2007. To explain this the subject said "There was no travel. And there is a really large amount of running. Because over the Christmas break from about the Christmas Eve through to about 5th or 6th of January, I ran with a group from home".

13. A peak can be observed at around November 2008. The athlete said "It's because in the winter time I put away the bike". And "I literally every winter would put away the bike for 2, 3, 4 months".

14. After November 2008, the running intensity graph shows a very low regularity. The subject said " I cycle more but every year I run approximately the same". The subject said "It's not regular in the daily sense. It's regular in the amount in total sense".

15. At around End of 2010 regularity is increasing. The subject confirmed and said "It's due to Christmas intensive running".

7.1.4 Cycling Responses

General questions:

1. The subject confirmed that start of cycling "around about March 2007. I started swimming in February 2005.

2. The subject said s/he has a schedule for cycling which is "Weekly slot, weekends, Sundays".

Interval change:

1. Although from April 2010 the intensity is kept at a high level, 3 out of 4 intensity graphs show that there is a deep valley at around May or June 2010. The subject confirmed the high-level of intensity and said "Having left all for January, February and done a little bit in the gym in March it's now starting to get out regularly". But the valley can't be explained by the subject. We specially asked if the subject was aware of a sudden decline, the subject said "No. I was very regular. Even though there were trips. I remember there were trips but we rent bikes there. And I was cycling always".

2. On the 3rd sub-figure of the intensity graph, the pattern of 2009 looks similar to 2011. The athlete answered "Not deliberately. But probably accidentally doing. So yes. Because I go for the same triathlon. I go for the same targets. And following the same kind of patterns. Or maybe even some travel plans. That would give an illusion of, yeah I suppose the expression of the regularity. Appearance of regularity".

Point change:

1. Intensity is very high in August 2009 and the subject confirmed that "It was a very regular period. I was doing 50 something kilometres every Sunday. That was the Summer like. That was the long evenings. And prior to that it's building up for those kinds of distances and that's just the way cycling is".

2. In September 2009 intensity started to decrease and the subject reported it's because of the weather and travel. The athlete said "Well it started to get cool after then. So it isn't nice but you know in the case of when you look at September. I went away at the end of August and that messed up the . . . yeah in August I went away in France, would have been to Brussels, would have been to France again. Yeah, so it was the travel things in September 2009 which disrupted the regularity".

3. At the beginning of 2012 the intensity went up again and the subject couldn't associate it with a specific reason at the beginning but s/he suggested that "the weather must be good, must be a mild Winter and because of that it was appealing to start cycling in January. And you can see I cycled every week through the January into February into March and April". And the subject said the distances weren't good, because of some trips s/he made. The subject afterwards said the reason for the peaks at the beginning of 2012 is "There was triathlon on the 25th of March 2012 in Melbourne and that would be earlier than I would normally do but I decided because I was going to Melbourne I do that triathlon in March that's why I started cycling in early January deliberately for that".

4. May 2013 the intensity was quite high. The athlete confirmed that It is "very regular just once a week cycle of more than 50 kilometres. Week on week on weeks on Sundays. Building up for the first triathlon which would have been later in the year which was later in the Summer. I was actually aiming for a cycling event in Italy that I took part in at the end of June so May would have been regular. Longer and longer distance Sunday spins".

5. Intensity in May 2014 was quite obviously high as well. The athlete confirmed another event s/he took part in. "Yeah it was aiming for triathlon at the end of May. So it was lots of regular cycles. Short not very long but twice a week at the end of the week. Thats why I would get the regularity".

7.1.5 Swim Responses

General questions:

1. The subject confirmed that the start of swimming "would have started in February 2005. Not doing a lot".

2. The subject confirmed that s/he is aware of that s/he runs less while swimming more.

3. The subject was asked to segment the swimming training. The athlete responded "I would say the first period there from 2006 through to 2008, it's just getting used to it. Learning how to swim and from 2008 through to 2010 I wouldn't say it's changed that much even since then 2008 there have been waves of enthusiasm where regularity is gone up or down on a monthly basis but I'd say swimming is very consistent. From about 2008 through to now, I am not changing anything. I might start doing sea swimming on the weekends but it is very recent. It's not captured that data there. I think my swimming has been regularly consistent".

Interval change:

1. The intensity graph shows low intensity at the beginning of swimming. But the subject said s/he was quite regular at the beginning. "Not regular as in every day of the week, but large number of days with a regular slot". And s/he added that "I swim a couple of times a week. But not on a given day".

2. From year 2008 to year 2010 the intensity is quite high, the athlete said because "I got used to it. Because it takes that long to be able to get used to swimming. It became part of, yeah, you are able to do it without falling asleep for the rest of the day. And it took that year to do that. And when I got the that stage I was comfortable doing that".

3. The intensity graph does not show a significant trend of intensity in the data. The subject said "I would say I am still getting 2 or 3 days of swimming a week but I think that remains constant throughout the calendar year unlike the cycling. And I think that varies from one day to the next. So on this

week that might be Monday, Wednesday, Friday and next week it might be Tuesday, Thursday because it depends upon things that happen".

Point change:

1. January 2008 is high intensity. The subject confirmed that "Yeah. There is a lot of lengths on then practically every day. Two days a week. Two times every second day. I just do a lot of swimming. Oh I think that was the target. I think I set myself target of doing 1,000 lengths in a month. That's what that peak is. 1,000 lengths in a calendar month 25 kilometers a month".

2. There is high intensity in February or March 2008. The subject responded positively "Yeah. Again it was lot of regularity because there is no real travel. And its just 3 times a week (repeat) yeah very regular between about 1st January through to about 13th of March just a lot of regularity".

3. At the end of 2009 the intensity was very high. And the athlete confirmed that "Wow look at that. Oh yeah. Look at those days. I think because I was injured, running, yeah that's it was. So look at that month it's got the huge number of swims. No running right and it's because I was injured from running. Yeah. It pulled me out of running for 4 weeks and I was swimming instead".

4. At the end of 2012 or beginning of 2013, the intensity gets very high again. The athlete said "End of 2012 there wasn't a lot because I was away. Because I don't do much swimming in the Christmas period, because the place is closed. And I had a bad back I had to get needle acupuncture in December 2012. And then from about 9th of January onwards I was doing lots of regular swimming for about a month. Yeah from 9th of January through to about 8th February I was swimming every second day. No particular reason. Just very regular".

7.1.6 Aggregated Data Responses

General:

1. The athlete cannot assure that the intensity is increasing. But s/he said "I don't know if the regularity is increasing. I think regularity is my annual cycle". The graph suggests an annual pattern with various intensity.

2. The subject does not think the volume of training is increasing. She/He said "Probably isn't. Over the years you do less but maybe by diversifying the sports it's allowing me to lose as much".

Point change:

1. In May 2008 the intensity of aggregated data is high. The subject confirmed with "Even thought there is trip there. A trip to Iceland, I was swimming there most days there, yeah everyday no interruptions to regularity because travel facilitated activities". The subject mentioned only running, cycling and swimming seemed dismissed.

2. In May 2010 the intensity is high. The subject said "There is an event. There is a triathlon event. Oh two events in May 2010".

3. There is a valley around June 2010. The subject did endorse that with "Yeah recovery. And it was a week holiday. In the month of June 2010." But the athlete report that s/he did "a lot of sports instead in August 2010", which could correspond to a peak in the intensity graph.

4. There is a high peak in April 2011. The athlete said "a lot of sports instead in August 2010". She/He report high swim mileages "Yeah the distance in swim were hard. It was 80 lengths. Yeah it was moderate".

5. There is another peak around April 2013. The athlete reports another event. "I was in Australia. Did another triathlon in April 2013".

7.2 Conclusion

By referring back to the subject in an interview, we are able to see for most of the patterns, no matter whether intensity peaks, intensity valleys or general trends of intensity over a period of time, these are responded to positively by the athlete. That is to say that the knowledge we learned from the periodogram and intensity graph has fit greatly with the athlete's self-reported training career. Even though we didn't know any training experience as prior knowledge. The questions that the subject didn't respond to are mostly due to the fact that the athlete cannot remember, and some questions are rejected by the subject but endorsed by the subject later after a time delay. As a qualitative study of periodicity and intensity of periodicity, we can conclude from the data and response from the athlete that the periodicity and intensity of periodicity are somewhat correlated with the training experience of a real person.

7.3 ASU Data

The intensity or strength of circadian periodicity is computed in the ASU accelerometer data using Method 1-5 as described in the previous chapter. The intensity of circadian periodicity is computed for each subject applied on full length of collected data without using sliding windows. If there are gaps in the data, the missing data gap is set to zero. The reason we use all data for each subject is because there are several biomarkers associated with each subject and we would like to see how the overall intensity of circadian periodicity could be related to those biomarkers.

7.3.1 Participants

Table 7.1 provides demographic information regarding the final sample of participants. In total, 25 participants were enrolled for this analysis; however, five were

excluded due to not presenting for the study measure completion following the initial three weeks of accelerometer wear. The final sample (N = 20) were middle-aged, primarily men and Caucasian, inactive, and with moderate levels of insomnia symptoms. Continuous accelerometer wear time varied from 13.9 (minimum) to 102.0 (maximum) days (mean wear: days). Non-wear time was minimal across the 24-h period in the sample (0.03 ± 0.07 percent of days). Of the 25 participants, 15 had complete data (58.8 ± 26.4 days). The remaining nine participants had 73.6 ± 23.1 days of data collection and $9.2\% \pm 9.0\%$ days of missing data. Overall missing data across the full data collection period was $3.5\% \pm 7.1\%$.

Table 7.1: Participant demographics (N = 20).

Age, M±SD	49.7±9.1
Men, N(%)	17 (85.0)
Race/ethnicity, N(%)	
Caucasian	14(70.0)
African-American	3 (15.0)
Hispanic	2 (10.0)
Asian American	1 (5.0)
Leisure-time physical activity (MET-min/week), M±SD	878.6±1680.9
Insomnia symptoms (ISI), M±SD	14.8±6.4

ISI = Insomnia Severity Index (range: 0 - 28).

7.3.2 Periodicity Strength Metrics

Table 7.2 presents descriptive statistics and intercorrelations among the five methods for calculating periodicity. Methods 1 - 4 displayed very high correlations among methods. In particular, Method 1 was strongly correlated with Method 4 and Method 2 was strongly correlated with Method 3. Method 5 was not strongly correlated ($r's < 0.40$) with any of the other methods. Normalized versions of these metrics were calculated and similar patterns of results was observed (not pictured).

Table 7.2: Means, standard deviations, and Pearson correlations among five periodicity strength metrics (N = 20)

	Method 1	Method 2	Method 3	Method 4	Method 5
Mean	0.2	0.10	0.10	0.23	0.45
SD	0.24	0.22	0.21	0.24	0.15
Method 1					
Method 2	0.930				
Method 3	0.93	0.99			
Method 4	0.98	0.90	0.92		
Method 5	0.29	0.14	0.19	0.38	

7.3.3 Associations with Cardiometabolic and Quality of Life Outcomes

Tables 7.3 and 7.4 present descriptive data for cardiometabolic and quality of life outcomes and partial correlation coefficients for each of the periodicity strength metrics and cardiometabolic and quality of life outcomes. The profile of participants' descriptive data suggests that the sample was at moderate to high risk for cardiometabolic risk diseases. The strongest and most consistent correlations were observed between the periodicity strength metrics and LDL-cholesterol and triglycerides outcomes. Consistent-yet only moderate in strength-relationships were observed for hs-CRP and health-related quality of life. HDL-cholesterol, plasma glucose, and insulin were not consistently associated with the periodicity strength metrics. As expected (due to high intercorrelations), Methods 1 - 4 displayed very similar patterns of results. In contrast to Methods 1 - 4, Method 5 displayed a moderately strong relationship with systolic BP and HDL-cholesterol and no relationship with hs-CRP or triglycerides.

7.3.4 Discussion

In the previous chapter we have tried to develop a framework for identifying periodicities (i.e., repeating patterns) from longitudinal wrist-worn accelerometer data and to establish whether these periodicities were independently associated with key

Table 7.3: Partial correlation coefficients, between cardiometabolic biomarkers and health-related quality of life indices, and periodicity strength metrics (N=20) for Method 1 - 3.

		Periodicity strength metrics		
	M±SD	Method 1	Method 2	Method 3
Waist circumference, in	66.82 ± 35.10	0.28	0.27	0.25
Systolic BP, mm Hg	138.6 ± 17.13	‡	‡	‡
Diastolic BP, mm Hg	89 ± 16.32	‡	‡	‡
Total cholesterol, mg/dL	177.4 ± 50.51	$0.52^{\dagger}$	0.68^{**}	0.57^{*}
HDL cholesterol, mg/dL	33.9 ± 11.76	‡	‡	‡
LDL cholesterol, mg/dL	109.7 ± 37.64	$0.45^{\dagger}$	0.57^{*}	$0.46^{\dagger}$
hs-CRP, mg/dL	7.76 ± 5.60	$0.47^{\dagger}$	0.38	0.30
Triglycerides, mg/dL	168.7 ± 74.06	0.77^{**}	0.86^{***}	0.81^{***}
Plasma glucose, mg/dL	117.2 ± 50.69	‡	‡	‡
Insulin, pmol/L	44.58 ± 73.01	‡	‡	‡
Health-related quality of life	47.25 13.03	0.37	0.54^{*}	0.55^{*}

$^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$; $\dagger P < 0.10$; $^{\ddagger}r < 0.25$ and $P > 0.0$.
All models are adjusted for age, gender, race/ethnicity, leisure-time physical activity, insomnia symptoms, and intervention assignment.

Table 7.4: Partial correlation coefficients, between cardiometabolic biomarkers and health-related quality of life indices, and periodicity strength metrics (N=20) for Method 4 - 5.

		Periodicity strength metrics	
	M±SD	Method 4	Method 5
Waist circumference, in	66.82 ± 35.10	0.30	‡
Systolic BP, mm Hg	138.6 ± 17.13	‡	0.57^{*}
Diastolic BP, mm Hg	89 ± 16.32	‡	‡
Total cholesterol, mg/dL	177.4 ± 50.51	$0.46^{\dagger}$	$0.47^{\dagger}$
HDL cholesterol, mg/dL	33.9 ± 11.76	‡	$0.51^{\dagger}$
LDL cholesterol, mg/dL	109.7 ± 37.64	0.40	0.42
hs-CRP, mg/dL	7.76 ± 5.60	$0.53^{\dagger}$	‡
Triglycerides, mg/dL	168.7 ± 74.06	0.75^{**}	‡
Plasma glucose, mg/dL	117.2 ± 50.69	‡	‡
Insulin, pmol/L	44.58 ± 73.01	‡	‡
Health-related quality of life	47.25 13.03	0.37	$0.52^{\dagger}$

$^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$; $\dagger P < 0.10$; $^{\ddagger}r < 0.25$ and $P > 0.0$.
All models are adjusted for age, gender, race/ethnicity, leisure-time physical activity, insomnia symptoms, and intervention assignment.

markers of cardiometabolic health and health-related quality of life. The resultant periodograms demonstrated a consistent 24-h pattern representing a typical rest-activity cycle; however, the strength of this 24-h rest-activity pattern varied within and between individuals. Using varying methods of quantifying periodicity strength, we found preliminary evidence that the strength of the rest-activity cycle was associated with key cardiometabolic risk biomarkers and health-related quality of life, independent of self-rated physical activity and insomnia symptoms.

Despite different methodologies in characterizing the rest-activity cycle and health outcomes, the result from this section is consistent with other studies. Mormont *et al.* (2000) examined the rest-activity cycle in metastatic colorectal cancer patients using an autocorrelation coefficient at 24-h and a dichotomy index that compared activity in bed and out of bed. These metrics were positively correlated with improved quality of life, response to treatment, and survival. In a follow-up to this study, Innominato *et al.* (2009) further clarified the importance of the rest-activity cycle, as measured via accelerometry, by demonstrating the stronger correlations observed between the rest-activity cycle metrics compared to mean counts of physical activity for health-related quality of life and survival outcomes in metastatic colorectal cancer patients. Our approach extends these findings in some important ways. First, these studies sampled behavior over 3-4 consecutive days. Therefore, our investigation substantially lengthens the monitoring period and therefore provides a clearer picture of habitual 24-h rest-activity cycles. Second, we have explored these relationships and found associations with a broader set of health outcomes in a group at elevated cardiometabolic risk. Finally, our framework for the development of periodograms and metrics to characterize periodicity strength represents a more sophisticated and nuanced approach that may provide a more precise determination of the 24-h rest-activity cycle.

One of the most interesting findings from the current investigation was the differences and similarities in correlation of the various periodicity strength metrics and

health outcomes. Methods 1 - 4 yielded very similar results due to high intercorrelations among these related methods for quantifying periodicity strength. Methods 1 - 4 computed correlation as the sum of products $(s * r)$ over all points, s_i , in the pattern against corresponding points r_i, in the original signal. This tells us how close the shape is between the original signal and the detected pattern. These metrics were consistently and strongly correlated with cardiovascular physiology outcomes such as LDL-cholesterol, triglycerides, and inflammation (hs-CRP). In contrast, Method 5 produced a different profile of correlation with health outcomes. Method 5 computed correlation as the root-mean-square error of the difference $(s - r)$ between corresponding points in the original data and the pattern. This tells us the sum of absolute differences between the pattern and the original signal. This metric was associated with HDL-cholesterol and systolic blood pressure, while Methods 1 - 4 were not. While it is not directly clear why these unique correlates were identified for the various periodicity strength metrics, it does suggest that nuances in the rest-activity cycle may uniquely contribute to cardiometabolic disease risk.

7.3.5 Strengths and Limitations

An important strength of our approach was the long-term, longitudinal nature of the collection of accelerometry data. Typically, accelerometer data are collected for seven or fewer consecutive days. These analyses demonstrate a novel methodology for harnessing longitudinal accelerometry data with demonstrated additional explanatory power for health outcomes beyond what has typically been reported in reports of accelerometry data and health outcomes. An additional strength was the minimal amount of missing accelerometer data. While the methods employed here were relatively robust to missing data, the trivial missing data demonstrates the feasibility of collecting long-term monitoring data. A final strength of our intensity based approach was the reliance on a completely open-source, raw data collection

methodology with no proprietary algorithms. An important limitation of this preliminary experiment was the relatively small sample size and limited duration of the monitoring period. While there was substantial within- and between-person variability in periodicity strength observed, a larger sample may have yielded stronger and more definitive patterns in the rest-activity cycle. Relatedly, the sample was exclusively those with elevated cardiometabolic risk and the results may not generalize to a healthy population. Furthermore, while the length of the monitoring was indeed longer than typically what is reported, longer monitoring periods may have yielded more interesting month, seasonal, or annual patterns of data as have been observed in other forms of lifelog data (Hu *et al.*, 2014). Additionally, while the periodicity strength metrics were calculated based on longitudinal data, the cardiometabolic and health-related quality of life metrics were measured concurrently, and therefore the relationships reported represent cross-sectional associations. Finally, these data were collected in the context of a behavioral intervention. While the effect of this intervention was statistically adjusted for, residual confounding may exist.

7.3.6 Conclusion

The use of periodograms and periodicity strength represents a novel methodology for understanding long-term monitoring of 24-h accelerometry data. The intensity based analytical framework can be used with minimally processed accelerometer data and, in this sample, demonstrated moderate to strong independent associations with key cardiometabolic and health-related quality of life outcomes. This framework and preliminary work may be useful as long-term monitoring of accelerometer data across the 24-h period becomes more commonplace in epidemiological and intervention research.

Chapter 8

Conclusion

The chapter summarizes the thesis and highlights the contribution to knowledge that is made. Some future work to be done is also included as well.

8.1 Hypothesis and Research Questions

Please visualize this in your mind. Imagine every one of us as a particle. Every particle has different size and rotates differently. The thesis has proven the uniqueness of the rotation of each particle and, also, that the rotation can some how be modelled and predicted given enough data. Interestingly, each particle seems to rotate at near the same periodicity such as one day per circle. Yet, every particle rotates slightly differently and even for the same particle, the rotation of each day is different. It is almost like the information from each rotation is modulated on circadian periodicity. Even for circadian periodicity, each particle may rotate differently because the particles have different sizes and may rotate suddenly faster, or slower. Let's acknowledge that there is a difference between standard circadian rotation and actual rotation intensity. If we map each particle into an acceleration domain, this is data we have collected. This can be also projected onto a different biomarker domain. We found those two domains are highly correlated by intensity.

Last, but not least, the intensity has correspondence with how the particle actually rotates.

8.1.1 RQ1

For our first research question, we found that everyone, and the data they generate for their own lifelogs, is different. We also proved the existence of a more or less bijective/one-to-one correspondence between subjects and their data. We found out that the uniqueness of subjects are represented in the data they generate, and the ability of capturing this uniqueness can be improved by improved algorithms to find better projection functions between subjects and data. Because of the uniqueness of each subject and their data, we endorse the idea to provide individualized analysis.

8.1.2 RQ2

We found that longitudinal lifelog data can be modelled and predicted with suitable models and under several constraints. We have found that in the experiments in section 4.6 we conducted, the number of points to be predicted yields very different results. That is to say the uniqueness of longitudinal lifelog data can be captured by computer algorithms under the condition of how 'greedy' the model is to be. We also observed that some of the best performing methods model high-level lifelog data as a quasi-periodic signal.

8.1.3 RQ3

We have found it is possible to achieve detection of periodicities from lifelog data, even when there is substantial missing data. To be more precise, we have found out that methods based on periodograms and/or autocorrelation can be used to detect periodicity on complete datasets, while Lomb-Scargle periodograms can be used to detect periodicity on datasets with missing data. We have observed periodicity

across several lifelog datasets collected in real life by conducting experiments on those datasets. We showed that even in the cases with different levels of sparsity, periodicity can still be detected.

8.1.4 RQ4

We found that our proposed model can simulate real life high-level data quite well. We explained the harmonics we observe in periodograms and the peaks on lower frequency in periodograms. We have showed that the performance of different metrics and how to choose window size by using the synthetic data, and the result is generally good. Experiments conducted on real lifelog data have suggested that further exploration with subjects involved in the experiment is needed.

8.1.5 RQ5

In the first case study we have found that the knowledge we learned from the periodogram and intensity graph has fit quite well with the athlete's self-reported training career, even though we didn't know any training experience as prior knowledge. The result is drawn from an interview with the subject who generated the data. We have demonstrated in this case study that the athlete responded positively to most of the questions, except for those questions that the athlete wasn't be able to answer due to loss of memory. We found some questions were rejected by the subject initially but responded positively later, after a short time delay.

In the second case study, we revealed the use of periodograms and periodicity strength represents a novel methodology for understanding long-term monitoring of 24-h accelerometry data. The intensity based analytical framework can be used with minimally processed accelerometer data. We have demonstrated moderate to strong independent associations with key cardiometabolic and health-related quality of life outcomes. We therefore can conclude that the intensity computed from longitudinal

lifelog data may be useful for long-term monitoring of accelerometer data across the 24-h becomes more commonplace in epidemiological and intervention research.

8.2 Contribution

In the literature we did not see many of the works to validate the feasibility of lifelog data existing. The validity of such a lifelog framework is so important that the result of analysis could be questioned sometimes. In the thesis we are trying to provide convincible evidence by using different datasets. The aim to use cross dataset validation is to show that the fundamental ideas of lifelog is theoretical examinable. In order to achieve this we have applied several state-of-art deep learning methods on raw accelerometer data which is rarely discussed in the literature. The fact that we directly operate, manipulate and process raw low-level and high-level acceleration data is a new direction in lifelog data processing. It is the first time in lifelog research that we discover and identify periodicity in lifelog data. Beside this we further proposed several measurement to capture periodicity intensity which is one of the frontier research in lifelog research.

The main contribution of the thesis is to provide evidence to support lifelog data framework. This include prove bijective correlation between data and people, predicability of lifelog data and find periodicity and periodicity intensity of lifelog data. Beside the research questions presented in the thesis which showed validity of data flows from people and to people, we take advantage of this framework and demonstrate how to correlate data with people using periodicity and intensity of periodicity. There are several contributions that this thesis makes to knowledge and they are enumerated as follows:

1. We have discovered that raw accelerometer data as a distribution of 2D histograms from all subjects shows that accelerometer data are mostly bounded in a circle which has a radius of one gravity (1G).

2. We have demonstrated that data generated by each subject is different as measured by various metrics and the local variance of the data generated by each subject fluctuates through time. Local variance can be measured by metrics such as KL divergence or entropy.

3. We have used different machine learning techniques such as SVM and deep learning methods such as CNN and LSTM to classify data by using the identity of each subject as the target class. The performance of different methods have shown quite good results in terms of precision, recall and F1-score. The application of deep learning methods has been rarely applied to longitudinal lifelog data in the literature.

4. We have used machine learning, deep learning methods to model lifelog data in order to build a generative model for low-level data in section 4.6. We also built models for high-level data using various machine learning, deep learning and time series models. These models have shown satisfactory results when evaluated by root-mean-square error, Correlation Coefficient and Concordance Correlation Coefficient to predict data points with short time lag. We have discovered that the learned parameters tend to model the data as a periodic signal.

5. We have shown that different algorithms can tackle the periodicity detection problem even for irregularly or unevenly sampled data. We have demonstrated that periodicity is a common, natural feature of lifelog data by using several lifelog datasets, both regularity sampled and unevenly sampled.

6. We have proposed a lifelog data model composing global trends, local trends, local patterns and missing possibility of local patterns. Periodograms have shown the proposed models have similar features to real life high level lifelog data. We provided a possible explanation of harmonics occurring in real life

data periodograms.

7. We have proposed several metrics to measure intensity of periodicity. Evaluated by the proposed model, we have shown the performance of different intensity metrics measured by root-mean-square error and CC. We have also studied the optimal relationship between window size and performance of different metrics. We have used selected intensity metrics to generate intensity graphs for both low-level and high-level data.

8. We have demonstrated two case studies to show the correlation between computed intensity and real life meaning. A low-level data case study has shown that there is a correlation between intensity of circadian periodicity and several biomarkers. For high-level data, an interview was conducted with the subject and the results have shown the subject has respond positively to most of the abnormalities in the periodicity intensity graph, and, some insightful explanations have been made by the subject.

8.3 Limitations

In this thesis, most of research is evaluated quantitatively while few points are addressed by qualitative method. There are some limitations that is inevitable which we would like to discuss it in this section. From dataset point of view, we used three datasets, 2 of them are contributed by one subject, and 1 of them are collected from 25 subjects. For the dataset contributed from one subject, the biggest limitation is generalization of the model. Since we did an interview with the subject using the intensity graph computed from 1 method, if we need to conduct more interviews with the subject with different measurements described in the thesis, those interviews are not considered independent interviews, which means the result of those interviews are more or less correlated. For the dataset contributed from 25 subjects,

although the variety of the data is ensured, but since the data collection process is less controlled so that there are many overlapping and noisy segments that needed to be manually removed. To make sure the data we have has enough information to be captured by machine learning algorithms we have to remove those days with low entropy to balance between quantity and quality. As for the evaluation of ASU dataset, the results show some measurement are quite correlated with different biomarkers. But the limitation is that as the number of subject we collect, can we still see the correlation of those biomarkers. In this case we require the experiment is conducted on larger scale. To collect more data from more people for longer period of time would not only provide more evidence of the existence of bijective correlation and predicability of lifelog data, but also enable large scale unsupervised learning.

As for the methods we choose, if we had enough space and computational power, we could generate more training and testing samples so that each machine learning models can converge more to the expected result. And please note that performance of different machine learning methods are bounded by the data we have. In some extreme cases, even simplest model can achieve good result with well captured and annotated data. Limitation also comes with the training process of different algorithms. Due to large parameter space of each algorithms, we are not able to explore the whole space using methods like grid search, which means some of the methods might be biased. But in the thesis we are focused on if the model works rather than which model works best makes this limitation less arguable.

For the methods used to detect periodicity, the limitation is to evaluate those methods quantitatively. This issue arises from the nature of lifelog data. In signal processing, we can assume for instance high frequency part of a signal is noise. But in lifelog data, we are not able to make this statement, because data generated by human is rather a complex system. It's is hard to defined what is noise.

We also proposed some periodicity intensity measurements. Those measurement are evaluated by synthetic data, again several aspect may affect it, such as adjust-

ing the calculation of intensity from ground truth, change of noise distribution or even change range of parameters to generate trends. But the evolution of the measurement can be designed to tackle those problems. Again due to the lifelog data property it is still a hard task to evaluate those measurement. That's why we decide to correlate with real life scenario such as biomarkers captured.

8.4 Future Directions

Logical next steps for this work are threefold. First, replication of these methods on larger and more diverse samples is warranted. In the thesis most experiments is conducted on ASU low-level data generated by more than 20 subjects and an athletic high-level data generated by a single subject. This may include the use of existing cohorts where raw data collection protocols of 24-h accelerometry are in place (even protocols that only include seven days of wear) with health-related outcomes measured in a cross-sectional or longitudinal fashion (e.g., US National Nutrition and Health Examination Survey, UK Biobank).

Second, an expansion of our theory to longitudinal multimodal data is needed. We are interested to see periodicity of visual logs combined with other sensors. Also, how would the periodicity of each modality correlate with each other. It is also interesting to see how can we expand our proposed model to low-level data and to compare the proposed model with machine learned models could also be a next step. It is also very important to find a probability model that can be used to generate periodic signals to simulate lifelog data and then we can use the probability model and Monte Carlo Markov chain to generate better simulation data.

Finally, it is not known whether this metric may be more sensitive to some biomarkers or activities such as changes in sleep, sedentary behavior, physical activity, some combination of these behaviors, or some alternative feature not currently being considered. Further clarification is needed for why these metrics were associ-

ated with certain biomarkers and not others, as well as why there was such variability in the strength of these associations with various biomarkers. Causality cannot be established between biomarkers and intensity metrics, and therefore it is imperative that these relationships be followed longitudinally where periodicity strength is experimentally manipulated in a manner to invoke changes in cardiometabolic risk biomarkers.

Bibliography

O. Aghazadeh, J. Sullivan, and S. Carlsson. Novelty detection from an ego-centric perspective. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 3297–3304. IEEE, 2011.

J. Arendt. Biological rhythms: the science of chronobiology. *Journal of the Royal College of Physicians of london*, 32(1):27–35, 1997.

L. Atallah, B. Lo, R. King, and G.-Z. Yang. Sensor positioning for activity recognition using wearable accelerometers. *Biomedical Circuits and Systems, IEEE Transactions on*, 5(4):320–329, 2011.

L. Bao and S. S. Intille. Activity recognition from user-annotated acceleration data. In *Pervasive Computing*, pages 1–17. Springer, 2004.

D. R. Bassett, R. P. Troiano, J. J. McClain, and D. L. Wolff. Accelerometer-based physical activity: total volume per day and standardized measures. *Med Sci Sports Exerc*, 47(4):833–8, 2015.

M. Basseville, I. V. Nikiforov, *et al. Detection of abrupt changes: theory and application*, volume 104. Prentice Hall Englewood Cliffs, 1993.

S. J. Bell Burnell. Little Green Men, White Dwarfs or Pulsars? *Cosmic Search*, 1: 16, 1979.

Y. Bengio. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127, 2009.

M. Bolaños, M. Garolera, and P. Radeva. Object Discovery Using CNN Features in Egocentric Videos. In *Pattern Recognition and Image Analysis*, pages 67–74. Springer, 2015.

L. A. Brocklebank, C. L. Falconer, A. S. Page, R. Perry, and A. R. Cooper. Accelerometer-measured sedentary time and cardiometabolic biomarkers: A systematic review. *Preventive Medicine*, 76:92–102, 2015.

E. Brodsky and B. S. Darkhovsky. *Nonparametric methods in change point problems*, volume 243. Springer Science & Business Media, 2013.

M. Bukhin and M. DelGaudio. WayMarkr: acquiring perspective through continuous documentation. In *Proceedings of the 5th International Conference on Mobile and Ubiquitous Multimedia*, page 9. ACM, 2006.

M. P. Buman, D. R. Epstein, M. Gutierrez, C. Herb, K. Hollingshead, J. L. Huberty, E. B. Hekler, S. Vega-López, P. Ohri-Vachaspati, A. C. Hekler, *et al.* BeWell24: Development and Process Evaluation of a Smartphone App to Improve Sleep, Sedentary, and Active Behaviors in US Veterans with Increased Metabolic Risk. *Translational Behavioral Medicine*, pages 1–11.

D. Byrne, A. R. Doherty, C. G. Snoek, G. G. Jones, and A. F. Smeaton. Validating the detection of everyday concepts in visual lifelogs. In *Semantic Multimedia*, pages 15–30. Springer, 2008.

D. Byrne, A. R. Doherty, C. G. Snoek, G. J. Jones, and A. F. Smeaton. Everyday concept detection in visual lifelogs: validation, relationships and trends. *Multimedia Tools and Applications*, 49(1):119–144, 2010.

F. P. Cappuccio, L. D'Elia, P. Strazzullo, and M. A. Miller. Quantity and quality of sleep and incidence of type 2 diabetes a systematic review and meta-analysis. *Diabetes Care*, 33(2):414–420, 2010.

J. Cheng, M. Sundholm, B. Zhou, M. Hirsch, and P. Lukowicz. Smart-surface: Large scale textile pressure sensors arrays for activity recognition. *Pervasive and Mobile Computing*, 2016.

S. Chennuru, P.-W. Chen, J. Zhu, and J. Y. Zhang. Mobile lifelogger–recording, indexing, and understanding a mobile user?s life. In *International Conference on Mobile Computing, Applications, and Services*, pages 263–281. Springer, 2010.

L. Choi, Z. Liu, C. E. Matthews, and M. S. Buchowski. Validation of accelerometer wear and nonwear time classification algorithm. *Medicine and science in sports and exercise*, 43(2):357, 2011.

J. Cohen. Weighted Kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4):213, 1968.

C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

H. Cruse. Neural networks as cybernetic systems. *Brain, minds, and media. See http://www. brains-minds-media. org/archive/289*, 2006.

C. A. Czeisler, J. F. Duffy, T. L. Shanahan, E. N. Brown, J. F. Mitchell, D. W. Rimmer, J. M. Ronda, E. J. Silva, J. S. Allan, J. S. Emens, *et al.* Stability, precision, and near-24-hour period of the human circadian pacemaker. *Science*, 284(5423):2177–2181, 1999.

A. R. Doherty. *Providing effective memory retrieval cues through automatic structuring and augmentation of a lifelog of images*. PhD thesis, Dublin City University, 2009.

A. R. Doherty and A. F. Smeaton. Automatically segmenting lifelog data into events. In *Image Analysis for Multimedia Interactive Services, 2008. WIAMIS'08. Ninth International Workshop on*, pages 20–23. IEEE, 2008a.

A. R. Doherty and A. F. Smeaton. Combining face detection and novelty to identify important events in a visual lifelog. In *Computer and Information Technology Workshops, 2008. CIT Workshops 2008. IEEE 8th International Conference on*, pages 348–353. IEEE, 2008b.

A. R. Doherty, A. F. Smeaton, K. Lee, and D. P. Ellis. Multimodal segmentation of lifelog data. In *RIAO 2007: Large Scale Semantic Access to Content (Text, Image, Video, and Sound)*, pages 21–38, 2007.

A. R. Doherty, D. Byrne, A. F. Smeaton, G. J. Jones, and M. Hughes. Investigating keyframe selection methods in the novel domain of passively captured visual lifelogs. In *Proceedings of the 2008 International Conference on Content-Based Image and Video Retrieval*, pages 259–268. ACM, 2008a.

A. R. Doherty, C. Ó'Conaire, M. Blighe, A. F. Smeaton, and N. E. O'Connor. Combining image descriptors to effectively retrieve events from visual lifelogs. In *Proceedings of the 1st ACM international conference on Multimedia information retrieval*, pages 10–17. ACM, 2008b.

N. Eagle and A. Pentland. Reality mining: Sensing complex social systems. *Personal and Ubiquitous Computing*, 10(4):255–268, 2006.

D. R. Easterling and T. C. Peterson. A new method for detecting undocumented discontinuities in climatological time series. *International journal of climatology*, 15(4):369–377, 1995.

L. Farrell, E. Morgenroth, and I. Walker. A time series analysis of uk lottery sales: Long and short run price elasticities. *oxford Bulletin of Economics and Statistics*, 61(4):513–526, 1999.

M. G. Figueiro, M. S. Rea, and J. D. Bullough. Does architectural lighting contribute to breast cancer ? *Journal of Carcinogenesis*, 5(1):20, 2006.

R. Garnett, M. A. Osborne, and S. J. Roberts. Sequential bayesian prediction in the presence of changepoints. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 345–352. ACM, 2009.

J. Gemmell, L. Williams, K. Wood, R. Lueder, and G. Bell. Passive capture and ensuing issues for a personal lifetime store. In *Proceedings of the the 1st ACM workshop on Continuous archival and retrieval of personal experiences*, pages 48–55. ACM, 2004.

F. Gu, J. Han, S. Hankinson, E. Schernhammer, N. H. S. Group, *et al.* Rotating night shift work and cancer mortality in the nurses' health study. *Cancer Research*, 74(19 Supplement):2178–2178, 2014.

V. Guralnik and J. Srivastava. Event detection from time series data. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 33–42. ACM, 1999.

F. Gustafsson. The marginalized likelihood ratio test for detecting abrupt changes. *IEEE Transactions on automatic control*, 41(1):66–78, 1996.

F. Gustafsson and F. Gustafsson. *Adaptive filtering and change detection*, volume 1. Wiley New York, 2000.

M. Hagströmer, P. Oja, and M. Sjöström. The International Physical Activity Questionnaire (IPAQ): a study of concurrent and construct validity. *Public Health Nutrition*, 9(06):755–762, 2006.

J. D. Hamilton. *Time series analysis*, volume 2. Princeton university press Princeton, 1994.

W. K. Härdle, M. Müller, S. Sperlich, and A. Werwatz. *Nonparametric and semiparametric models*. Springer Science & Business Media, 2012.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015.

G. N. Healy, C. E. Matthews, D. W. Dunstan, E. A. Winkler, and N. Owen. Sedentary time and cardio-metabolic biomarkers in US adults: NHANES 2003–06. *European Heart Journal*, page ehq451, 2011.

M. A. Hearst. Texttiling: Segmenting text into multi-paragraph subtopic passages. *Computational linguistics*, 23(1):33–64, 1997.

S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

T. Hori and K. Aizawa. Context-based video retrieval system for the life-log applications. In *Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*, pages 31–38. ACM, 2003.

F. Hu, A. F. Smeaton, and E. Newman. Periodicity detection in lifelog data with missing and irregularly sampled data. In *Bioinformatics and Biomedicine (BIBM), 2014 IEEE International Conference on*, pages 16–23. IEEE, 2014.

K. Hu, E. J. Van Someren, S. A. Shea, and F. A. Scheer. Reduction of scale invariance of activity fluctuations with aging and alzheimer's disease: Involvement of the circadian pacemaker. *Proceedings of the National Academy of Sciences*, 106 (8):2490–2494, 2009.

B. Huurnink, K. Hofmann, and M. De Rijke. Assessing concept selection for video retrieval. In *Proceedings of the 1st ACM international conference on Multimedia information retrieval*, pages 459–466. ACM, 2008.

P. F. Innominato, C. Focan, T. Gorlia, T. Moreau, C. Garufi, J. Waterhouse, S. Giacchetti, B. Coudert, S. Iacobelli, D. Genet, *et al.* Circadian rhythm in rest and activity: a biological correlate of quality of life and a predictor of survival in

patients with metastatic colorectal cancer. *Cancer Research*, 69(11):4700–4707, 2009.

M. Jette, K. Sidney, and G. Blümchen. Metabolic equivalents (mets) in exercise testing, exercise prescription, and evaluation of functional capacity. *Clinical cardiology*, 13(8):555–565, 1990.

Y. Kawahara and M. Sugiyama. Change-point detection in time-series data by direct density-ratio estimation. In *SDM*, volume 9, pages 389–400. SIAM, 2009.

Y. Kawahara, T. Yairi, and K. Machida. Change-point detection in time-series data based on subspace identification. In *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, pages 559–564. IEEE, 2007.

J. Kerr, S. J. Marshall, S. Godbole, J. Chen, A. Legge, A. R. Doherty, P. Kelly, M. Oliver, H. M. Badland, and C. Foster. Using the SenseCam to improve classifications of sedentary behavior in free-living settings. *American Journal of Preventive Medicine*, 44(3):290–296, 2013.

P. Klasnja, S. Consolvo, T. Choudhury, R. Beckwith, and J. Hightower. Exploring privacy concerns about personal sensing. In *International Conference on Pervasive Computing*, pages 176–183. Springer, 2009.

K. L. Knutson, E. Van Cauter, P. Zee, K. Liu, and D. S. Lauderdale. Cross-sectional associations between measures of sleep and markers of glucose metabolism among subjects with and without diabetes the coronary artery risk development in young adults (CARDIA) sleep study. *Diabetes Care*, 34(5):1171–1176, 2011.

R. T. Krafty, O. Rosen, D. S. Stoffer, D. J. Buysse, and M. H. Hall. Conditional spectral analysis of replicated multiple time series with application to nocturnal physiology. *arXiv preprint arXiv:1502.03153*, 2015.

A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

S. Kullback. Letter to the editor: the kullback-leibler distance. 1987.

G. Lakoff. *Women, fire, and dangerous things: What categories reveal about the mind*. Cambridge Univ Press, 1990.

I. Lawrence and K. Lin. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, pages 255–268, 1989.

H. Lee, A. F. Smeaton, N. E. O'Connor, G. Jones, M. Blighe, D. Byrne, A. Doherty, and C. Gurrin. Constructing a SenseCam visual diary as a media process. *Multimedia Systems*, 14(6):341–349, 2008.

J. Li and J. Z. Wang. Automatic linguistic indexing of pictures by a statistical modeling approach. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(9):1075–1088, 2003.

N. Li, M. Crane, and H. J. Ruskin. Visual experience for recognising human activities. In *Evaluating AAL Systems Through Competitive Benchmarking*, pages 173–185. Springer, 2013a.

N. Li, M. Crane, and H. J. Ruskin. Automatically Detecting "Significant Events" on SenseCam. *International Journal of Wavelets, Multiresolution and Information Processing*, 11(06), 2013b.

N. Li, M. Crane, H. J. Ruskin, and C. Gurrin. Multi-scaled cross-correlation dynamics on SenseCam lifelogged images. In *Advances in Multimedia Modeling*, pages 490–501. Springer, 2013c.

N. Li, M. Crane, H. J. Ruskin, and C. Gurrin. Random matrix ensembles of time correlation matrices to analyze visual lifelogs. In *MultiMedia Modeling*, pages 400–411. Springer, 2014.

S. Liu, M. Yamada, N. Collier, and M. Sugiyama. Change-point detection in time-series data by relative density-ratio estimation. *Neural Networks*, 43:72–83, 2013.

X. Long, B. Yin, and R. M. Aarts. Single-accelerometer-based daily physical activity classification. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 6107–6110. IEEE, 2009.

S. Mann. Wearable computing: A first step toward personal imaging. *Computer*, 30(2):25–32, 1997.

M. Markou and S. Singh. Novelty detection: a review?part 1: statistical approaches. *Signal processing*, 83(12):2481–2497, 2003.

M. Mathie, B. G. Celler, N. H. Lovell, and A. Coster. Classification of basic daily movements using a triaxial accelerometer. *Medical and Biological Engineering and Computing*, 42(5):679–687, 2004.

K. McGuinness, E. Mohedano, Z. Zhang, F. Hu, R. Albatal, C. Gurrin, N. E. O'Connor, A. F. Smeaton, A. Salvador, X. Giró-i Nieto, *et al.* Insight centre for data analytics (dcu) at trecvid 2014: instance search and semantic indexing tasks. 2014.

T. Mikolov, M. Karafiát, L. Burget, J. Cernockỳ, and S. Khudanpur. Recurrent neural network based language model. In *INTERSPEECH*, volume 2, page 3, 2010.

C. M. Morin, G. Belleville, L. Bélanger, and H. Ivers. The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response. *Sleep*, 34(5):601–608, 2011.

M.-C. Mormont, J. Waterhouse, P. Bleuzen, S. Giacchetti, A. Jami, A. Bogdan, J. Lellouch, J.-L. Misset, Y. Touitou, and F. Lévi. Marked 24-h rest/activity rhythms are associated with better quality of life, better response, and longer survival in patients with metastatic colorectal cancer and good performance status. *Clinical Cancer Research*, 6(8):3038–3045, 2000.

V. Moskvina and A. Zhigljavsky. Application of the singular spectrum analysis for change-point detection in time series. *Journal of Time Series Analysis, submitted*, 2001.

V. Moskvina and A. Zhigljavsky. An algorithm based on singular spectrum analysis for change-point detection. *Communications in Statistics-Simulation and Computation*, 32(2):319–352, 2003.

K.-R. Müller, A. J. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, and V. Vapnik. Predicting time series with support vector machines. In *International Conference on Artificial Neural Networks*, pages 999–1004. Springer, 1997.

H. Narimatsu and H. Kasai. Duration and interval hidden markov model for sequential data analysis. In *Neural Networks (IJCNN), 2015 International Joint Conference on*, pages 1–8. IEEE, 2015.

C. Ó'Conaire, N. E. O'Connor, A. F. Smeaton, and G. J. Jones. Organising a daily visual diary using multifeature clustering. In *Electronic Imaging 2007*, pages 65060C–65060C. International Society for Optics and Photonics, 2007.

D. J. Patterson, L. Liao, K. Gajos, M. Collier, N. Livic, K. Olson, S. Wang, D. Fox, and H. Kautz. Opportunity knocks: A system to provide cognitive assistance with transportation services. In *UbiComp 2004: Ubiquitous Computing*, pages 433–450. Springer, 2004.

P. Perron. Further evidence on breaking trend functions in macroeconomic variables. *Journal of econometrics*, 80(2):355–385, 1997.

J. Platt *et al.* Sequential minimal optimization: A fast algorithm for training support vector machines. 1998.

S. J. Preece, J. Y. Goulermas, L. P. Kenney, D. Howard, K. Meijer, and R. Crompton. Activity identification using body-mounted sensors ? A review of classification techniques. *Physiological Measurement*, 30(4):R1, 2009.

W. H. Press, S. A. Teukolsky, T. Veterling, and B. Flannery. Numerical recipes 3rd edition. *The art of scientific computing*, 2007.

M.-R. Ra, B. Priyantha, A. Kansal, and J. Liu. Improving energy efficiency of personal sensing applications with heterogeneous multi-processors. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 1–10. ACM, 2012.

P. Ratsamee, Y. Mae, A. Jinda-apiraksa, M. Horade, K. Kamiyama, M. Kojima, and T. Arai. Keyframe selection framework based on visual and excitement features for lifelog image sequences. *International Journal of Social Robotics*, 7(5):859–874, 2015.

R. Rawassizadeh, M. Tomitsch, K. Wac, and A. M. Tjoa. Ubiqlog: a generic mobile phone-based life-log framework. *Personal and ubiquitous computing*, 17(4):621–637, 2013.

S. Reddy, A. Parker, J. Hyman, J. Burke, D. Estrin, and M. Hansen. Image browsing, processing, and clustering for participatory sensing: lessons from a dietsense prototype. In *Proceedings of the 4th workshop on Embedded Networked Sensors*, pages 13–17. ACM, 2007.

J. Reeves, J. Chen, X. L. Wang, R. Lund, and Q. Q. Lu. A review and comparison of

changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology*, 46(6):900–915, 2007.

J. D. Scargle. Studies in astronomical time series analysis. ii-statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, 263:835–853, 1982.

A. Schuster. On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Terrestrial Magnetism*, 3 (1):13–41, 1898.

A. J. Sellen, A. Fogg, M. Aitken, S. Hodges, C. Rother, and K. Wood. Do life-logging technologies support memory for the past?: an experimental study using SenseCam. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 81–90. ACM, 2007.

K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(12):1349–1380, 2000.

C. G. Snoek, B. Huurnink, L. Hollink, M. De Rijke, G. Schreiber, and M. Worring. Adding semantics to detectors for video retrieval. *Multimedia, IEEE Transactions on*, 9(5):975–986, 2007.

J. A. Suykens and J. Vandewalle. Least squares support vector machine classifiers. *Neural processing letters*, 9(3):293–300, 1999.

J.-i. Takeuchi and K. Yamanishi. A unifying framework for detecting outliers and change points from time series. *IEEE transactions on Knowledge and Data Engineering*, 18(4):482–492, 2006.

R. E. Taylor-Piliae, L. C. Norton, W. L. Haskell, M. H. Mahbouda, J. M. Fair, C. Iribarren, M. A. Hlatky, A. S. Go, and S. P. Fortmann. Validation of a new brief physical activity survey among men and women aged 60–69 years. *American Journal of Epidemiology*, 164(6):598–606, 2006.

V. N. Vapnik and A. Y. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. In *Measures of Complexity*, pages 11–30. Springer, 2015.

S. Vemuri and W. Bender. Next-generation personal memory aids. *BT Technology Journal*, 22(4):125–138, 2004.

J. Verbesselt, R. Hyndman, A. Zeileis, and D. Culvenor. Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. *Remote Sensing of Environment*, 114(12):2970–2980, 2010.

M. Vlachos, S. Y. Philip, and V. Castelli. On periodicity detection and structural periodic similarity. In *SDM*, volume 5, pages 449–460. SIAM, 2005.

P. Wang. *Semantic interpretation of events in lifelogging*. PhD thesis, Dublin City University, 2012.

P. Wang and A. F. Smeaton. Semantics-based selection of everyday concepts in visual lifelogging. *International Journal of Multimedia Information Retrieval*, 1 (2):87–101, 2012a.

P. Wang and A. F. Smeaton. Semantics-based selection of everyday concepts in visual lifelogging. *International Journal of Multimedia Information Retrieval*, 1 (2):87–101, 2012b.

P. Wang and A. F. Smeaton. Using visual lifelogs to automatically characterize everyday activities. *Information Sciences*, 230:147–161, 2013.

P. Wang, L. Sun, S. Yang, and A. F. Smeaton. What are the Limits to Time Series Based Recognition of Semantic Concepts? In *MultiMedia Modeling*, pages 277–289. Springer, 2016.

Y. Wang, C. Wu, Z. Ji, B. Wang, and Y. Liang. Non-parametric change-point method for differential gene expression detection. *PloS one*, 6(5):e20060, 2011.

D. L. Wolff-Hughes, E. C. Fitzhugh, D. R. Bassett, and J. R. Churilla. Total Activity Counts and Bouted Minutes of Moderate-to-Vigorous Physical Activity: Relationships With Cardiometabolic Biomarkers Using 2003–2006 NHANES. *Journal of Physical Activity & Health*, 12(5), 2015.

K. Yamanishi and J.-i. Takeuchi. A unifying framework for detecting outliers and change points from non-stationary time series data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 676–681. ACM, 2002.

K. Yamanishi, J.-I. Takeuchi, G. Williams, and P. Milne. On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms. *Data Mining and Knowledge Discovery*, 8(3):275–300, 2004.

K. Yamauchi and T. Akiba. Repeated event discovery from image sequences by using segmental dynamic time warping: experiment at the ntcir-12 lifelog task. *Proceedings of NTCIR-12, Tokyo, Japan*, 2016.

J.-Y. Yang, J.-S. Wang, and Y.-P. Chen. Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers. *Pattern Recognition Letters*, 29(16):2213–2220, 2008.

M. Zechmeister and M. Kürster. The generalised lomb-scargle periodogram-a new formalism for the floating-mean and keplerian periodograms. *Astronomy & Astrophysics*, 496(2):577–584, 2009.

W. Zhu, A. Berndsen, E. Madsen, M. Tan, I. Stairs, A. Brazier, P. Lazarus, R. Lynch, P. Scholz, K. Stovall, *et al.* Searching for Pulsars Using Image Pattern Recognition. *The Astrophysical Journal*, 781(2):117, 2014.

List of Publications

- Hu, F., Smeaton, A. F., Newman, E., & Buman, M. P.. Using periodicity intensity to detect long term behaviour change. *In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers (pp. 1069-1074). ACM*

- Hu, F., Smeaton, A. F., & Newman, E. (2014, November). Periodicity detection in lifelog data with missing and irregularly sampled data. In Bioinformatics and Biomedicine (BIBM), *2014 IEEE International Conference on (pp. 16-23). IEEE.*

- Hu, F., Smeaton, A. F., & Sun, Y. C. (2012). An image retrieval system based on explicit and implicit feedback on a tablet computer. *Irish Human Computer Interaction 2012, Galway, Ireland, 2012*

- Buman, M. P., Hu, F., Newman, E., Smeaton, A. F., & Epstein, D. R. (2015). Behavioral periodicity detection from 24h waveform wrist accelerometry. *In: International Conference on Ambulatory Monitoring of Physical Activity and Movement (ICAMPAM), 10-12 June 2015.*

- Albatal, R., McGuinness, K., Hu, F., & Smeaton, A. F. (2014, August). Formulating queries for collecting training examples in visual concept classification. *In Workshop On Vision And Language 2014.*

- McGuinness, K., Mohedano, E., Zhang, Z., Hu, F., Abatal, R., Gurrin, C., ... & Ventura, C. (2014). *Insight Centre for Data Analytics (DCU) at TRECVid 2014: instance search and semantic indexing tasks.*

- King, M., Hu, F., McHugh, J., Murphy, E., Newman, E., Irving, K., & Smeaton, A. F. (2013). Visibility of wearable sensors as measured using eye tracking glasses. *In Evolving Ambient Intelligence (pp. 23-32). Springer International Publishing.*