HILDA - A Health Interaction Log Data Analysis Workflow to Aid Understanding of Usage Patterns and Behaviours

Maurice D Mulvenna[1], Raymond R Bond[1], Alexander Grigorash[1], Siobhan O'Neill[2], Assumpta Ryan[3]

Abstract. Health and wellbeing products and services for individuals are becoming increasingly popular as people realise the benefits provided by lifelogging or quantified-self platforms in such areas as exercise, diet management and mood. However, in addition to the data that users record using these platforms, all user interactions and events can be elusively logged to represent usage. Such user interaction or event logs provide rich and large datasets that can fuel applied artificial intelligence. As products and services based on these digital interaction technologies are taken up across public healthcare provision, should healthcare policy and practice take more cognisance of the opportunities and risks in gathering interaction data? Is 'healthcare' ignorant that there is knowledge in such data? Are there differences between event logging in healthcare and other areas such as commerce, media and industry? In order to realise benefits in analysing such data, methods that help ensure consistency, accuracy, data protection, as well as reproducibility of knowledge derived from log data need to be examined. This paper presents methods to explore usage log data and a process workflow followed by a presentation of two real world case studies. The workflow has been coined Health Interaction Log Data Analysis (HILDA) and focuses on data prospecting and machine learning stages to show the opportunities realisable in analysing interactional or event data automatically recorded by digital healthcare services.

1 INTRODUCTION

Products and services based on digital interaction technologies typically include mobile device apps as well as browser-based apps to a lesser extent, and can include telephony-based services, text-based chatbots and voice activated chatbots. Many of these digital products and services are simultaneously available across many channels in order to maximise availability for users. The focus of this paper is to examine digital interaction technology-based products and services developed for use in the health and wellbeing domain, explore how these technologies can automatically log user interactions and events for subsequent analysis and business intelligence, and propose a methodology for data mining of digital products and service event logs in a healthcare context.

[1] School of Computing, Ulster Univ., BT37 0QB, UK. Email: {md.mulvenna, rb.bond}@ulster.ac.uk
[2] School of Psychology, Ulster Univ., BT52 1SA, UK. Email: sm.oneill@ulster.ac.uk
[3] School of Nursing, Ulster Univ., BT48 7JL, UK. Email: aa.ryan@ulster.ac.uk

Digital interaction technologies offer useful methods for real time data capture of the interactions of users with the products and services. We have become accustomed to using tools such as Google Analytics in order to generate aggregated usage reporting for our websites, but this kind of log analysis is available for all the digital interaction technologies we design. Indeed, we have the ability to design what data are recorded, how and where it may be stored, and crucially, how it can be analysed to reveal individual or collective usage patterns. This paper reports on two case studies in digital interaction technologies service analysis, where the usage data are examined using a structured data mining pipeline that is proposed and detailed in the methods.

The main focus of previous research on the analysis of usage logs for digital interaction technologies used in health and wellbeing has been to aid in usability analysis [1] or to reveal usage patterns in using technology [2]. Research has also been carried out to explore how rehabilitation devices can have data or event logging incorporated, but this has been more to support the goal of device monitoring [3]. More recent research has examined engagement data in web-based intervention platforms but has primarily focused on visualisation of the log data [4].

Our goal in incorporating usage or event logging or metadata logging is to provide relevant data sources as a basis for identification and exploration of individual and collective behavioural patterns of products and services developed using digital interaction technologies.

This work is potentially important since many national health departments including the National Health Institute (NHS) are looking to use digital technologies such as health apps for self-management of diseases and thus logging user interactions would allow for greater insight into user needs and may provide ideas for improving these digital interventions, for example through enhanced personalisation. The NHS would benefit since the data can be automatically and hence cost-effectively collected. Such data may facilitate new ways for epidemiological analyses and provide data to inform health policies. If the NHS promote health apps and log analysis is insightful, then perhaps there is a need for a standard to maximise the utility of recorded event logs for analysis in healthcare contexts.

The structure of the paper is as follows. In the following section, methods encompassing usage log data types and process workflows is described. In the subsequent results section, the two case studies in the paper are introduced, before data prospecting and machine learning stages are described and results presented. The final sections of the paper provide for discussion and conclusions. We aim to examine the value and limitations of user log analysis in the healthcare domain.

2 METHODS

A user event log must at least be comprised of three columns or variables as part of a tabular structure, comprising a unique identifier for the user (can be anonymous); the event that was recorded; and the date and time (preferable precision to seconds) of that event. This paper proposes a methodology for data mining such user event logs in a healthcare context (see Figure below). This study then takes a case study-based approach to examine the feasibility of using this model for analysing log data. Case study 1 involves an analysis of user event logs from a healthcare app that is used by people living with dementia and their carers for reminiscing about their past to encourage conversations, social connectedness, mutuality with their spouse and reflection as well as improving quality of life. This app was used in a 12 week trial during which all user events and features used were recorded. Log data includes the event/feature used (browsing photos, watching a video etc.), the date and time as well as a unique identifier for each user. Case study 2 involves analysis of ~3.5 million call logs made to a mental health and wellbeing helpline (Samaritans Ireland). Basic log data include time, date and duration of the call as well as a unique identifier for each caller. The results in this paper are very much autoethnographic of using our user log analysis workflow detailed in this paper. However, we do provide actual results of data analytic workflows in the case studies to provide an example of the insights revealed by the data mining process.

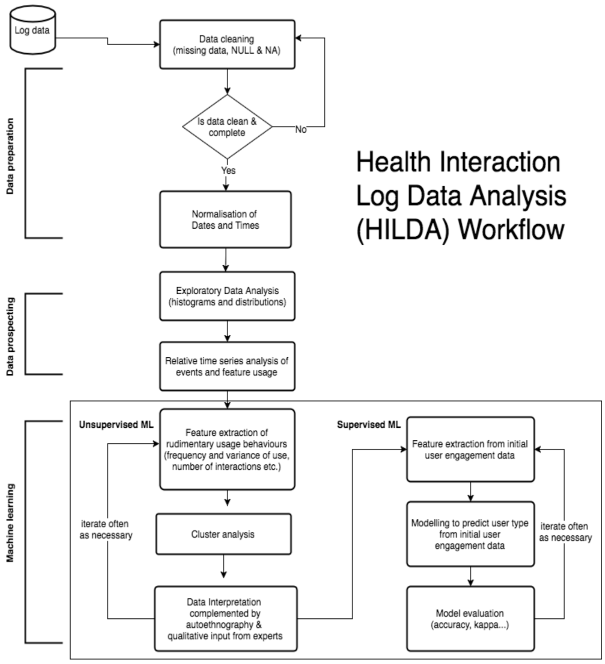

Figure 1. Workflow for Health Interaction Log Data Analysis (HILDA), involving three high level phases of data preparation, data prospecting and machine learning

2.1 Proposed workflow model for log data analytics

Standardising workflows are crucial in order to ensure consistency and that best practices are adopted in a domain. A number of standards for carrying out a data science, data mining or a machine learning project have been proposed. For example, the Cross-industry standard process for data mining (CRISP-DM) is a data mining process model encompassing the following stages: Business understanding, Data understanding, Data preparation, Modeling, Evaluation, Deployment [5]. CRISP-DM has been available in various guides since 1996. An updated variant developed by IBM, called Analytics Solutions Unified Method for Data Mining/Predictive Analytics (ASUM-DM) [6] expands on CRISP-DM.

However, to date, there has not been a standardised workflow that embraces data mining in log analysis. Figure 1 below depicts Health Interaction Log Data Analysis (HILDA) workflow, which has been applied to two health and wellbeing log datasets: a mental health helpline call log dataset; and a user log dataset from a healthcare app to facilitate reminiscence for people living with dementia and their carers.

The first phase of the HILDA workflow model involves data cleaning. This comprises of data quality checks and imputations for missing data (NULL and NA fields). This is followed by a normalisation of the date stamps. That is to say that each date of each event is converted to a relative date for each user. For example, the first logged event for each user will be transformed to 'day 1', allowing like-for-like temporal analyses for all user logs. Moreover, seasonality is also often determined as a separate feature. The time stamps are converted to a simple integer from 0 to 23 denoting hour of the day. Oftentimes, classification of morning, afternoon and evening are used.

Following data preparation, data prospecting is carried out. This comprises of analysing each variable for inherent characteristics by viewing histograms and probability density functions as well as identifying correlations via bivariate analysis. This is followed by time series analysis of a representative sample of users, or all if possible. This involves data visualisations of user events over their entire journey of using the service.

Following data prospecting is often a machine learning phase which can involve both unsupervised and supervised machine learning. Unsupervised machine learning comprises of extracting simple features that characterise basic usage behaviours such as frequency of use and number of interactions per week. These features are used in a clustering algorithm to determine the type of users the exist in using the service. This helps to understand the client base by behaviour. Having determined the type of service users that exist, one can proceed to predict service user type using features extracted from initial user behaviours, allowing for example the early detection of abstainers and adopters [7]. These predictive models can be developed using established supervised machine learning techniques such as decision trees and support vector machines which are very popular and arguably the most efficient methods that exist today.

The HILDA workflow model covers data preparation, data prospecting and machine learning in some detail. Deployment and usage is not explored in this version. Arguably, a focus on deployment was lacking in CRISP-DM as context for usage varies to such an extent that it is not helpful to be prescriptive about model deployment and usage.

3. RESULTS

The following sections outline the results and reflections of analysing user event log data in the health domain for each case study, describing the data prospecting and machine learning phases of the HILDA workflow in each.

3.1 Case study 1: Reminiscence health app user log data

This case study reports on the analysis of log data from a tablet application, specifically designed and developed to facilitate reminiscence for people with early to moderate stage dementia. Reminiscence is the sharing of memories relating to personal life experiences. It is the act of remembering and reflecting on real past events. The act of reminiscing can serve many functions that create bonds between people and in doing so, supports them to reflect on important life events and to attribute meaning to their lives [8]. The development of the app was a component part of a larger feasibility study to investigate the effects of individual specific reminiscence activity using a range of outcome measures, to explore users' views on the app; and to incorporate an economic analysis, examining the cost of implementing the app intervention in comparison with quality of life outcomes. The feasibility study incorporated a paired sample of 28 dyads (person living with dementia and their carer), and applied several scales at start, mid- and end-point of a 12-week use of the app in the homes of people living with dementia and their carers, with one-to-one interviews with participants carried out at the end of the 12 weeks.

Data prospecting

The app was designed to incorporate a logging facility for key events by users across 45 specific activities, covering five different types of events. The five different canonical events include: entry (Logging in), admin (Adding a photo, deleting an audio, etc.), reminiscing (Viewing a video, viewing a photo, etc.), in the moment (ITM) questions and exit (Logging out). Thus, the behaviour of users can be analysed within and across each usage session, over the 12-week trial. The ITM questions comprise items from the primary outcome measure for the study, the Mutuality Scale developed by Archbold et al. [9].

The data show that the app was primarily used for reminiscing as expected. A total of 71% of interactions from people living with dementia were within the reminiscing sections of the system whereas only 47% of interactions from carers were within the reminiscing sections ($p<0.001$). It is reassuring that people living with dementia mainly used the system for reminiscing. Only carers could carry out 'Admin' events such as adding a photo, as mandated by their access rights set at login. It can perhaps be seen as a positive sign that carers generally added to the music, pictures and videos that were uploaded to the app prior to the intervention beginning, rather than simply browsing those already there. There were twice as many interactions with photographs in comparison to music and five times as many interactions with photographs in comparison to video by people with dementia using the app. Reminiscing, with its history in photograph-based memory books, has been more about the image than music, sound or video, and this effect may be what is being seen in this data [10]. What is also interesting in this data is the popularity of music to people living with dementia. Again, this is known from the literature [11] and anecdotally from carers of people living with dementia but it is useful to see this behaviour replicated in this trial data. The most popular times that the dyads of people living with dementia and carers prefer to use the app peak around 11am, 3pm and 8pm. These times correspond to post-breakfast, post-lunch and post-evening meal times. We also calculated the number of unique days in which users interacted with the system, and there is a significant statistical correlation between the number of days the carer interacted with the system and the number of days the dyad's corresponding person living with dementia interacted ($r=5.77$, $p<0.001$).

Machine learning

In this study, we used K-means clustering algorithm given it is the most widely used and established clustering algorithm in the unsupervised machine learning literature. Using the elbow method, we discerned that 4 is a reasonably small number of clusters that would provide reasonable resolution in terms of explained variability. Clustering was based on the following features five features: number of interactions by person living with dementia, number of interactions by carer of person living with dementia, number of daily interactions by person living with dementia, the mean usage interval by a user and the standard deviation of usage interval by a user.

Four clusters were revealed by the k-means algorithm. The first cluster, 'the hooked adopter' constituted one dyad, who fully adopted the system. They had 7.2 times more interactions than their carer. Whilst the person with dementia obsessively used the app, the carer showed a normal amount of usage, hence the person with dementia was independently dedicated. The hooked adopter and dyad uses the app for over half the days in a month (55% of days) and with little variability uses the app every two days. The second cluster, labelled the 'typical user' encompassing the plurality of users, where 12 dyads or 43% fall into this cluster, hence making them the most typical user. These people living with dementia user only have 1.7 times more interactions with the app than their carer. This indicates that these users have some dependence on the carer for app usage. This dyad uses the app 15% of days in a month. This dyad are unpredictable when they will use the app but on average interacts with it every 6.61 days (approximately once per week). The third cluster, labelled 'disengaged irregular user' encompassed 7 dyads or 25% of users. These users had 25% fewer interactions with the app than the carer. Whilst the people with dementia had fewer interactions than their carers, the carers had fewer interactions than other carers in all other clusters. These dyads use the app 9% of the days in a month. However, typically they can go for 20 days without using the app making them the least consistent users of the app. The final cluster labelled the 'well supported dependent user' encompassed 8 dyads or 29% of users, the second largest group of users. These users have 36% fewer interactions with the app than their carers. The carers are very enthusiastic and have more interactions than other carers in all other clusters but they seem to struggle to get people with dementia users to the same engagement level. Similar to the typical users in cluster 2, these dyads interact with app 16% of the days in a month and on average use the app every 6.97 days. This unsupervised learning provided clusters that were clear and transparent to the health science researcher involved in the project. The next stage in this work is to seek to identify correlations between the post-trial interviews with the dyads and the clusters enumerated above.

3.2 Case study 2: Mental health helpline call log metadata

Helplines are key elements of mental wellbeing and suicide prevention efforts, however little is known about how these services are used. This study involves analysis of digital telephony data sourced from a mental health and wellbeing charity in Ireland. The charity operates a national helpline to provide emotional support to anyone in distress or at risk of suicide. Whilst support is also offered via SMS, email and face to face, 95% of the contacts remain via telephone. Data was provided for all calls made to the charity in the Republic of Ireland for almost a 4-year period (April 2013 to December 2016). A total of 3.449 million calls was analysed, amounting to 725 calls per 1,000 population.

Data prospecting

Novel data analytics and machine learning approaches were used to identify populations of callers based on caller behaviour variations. Key findings include the identification of five clusters of callers based on caller persistence, that persist regardless of which year or group of years is considered. The volume of calls exhibits strong intra-day and intra-week repetitive patterns, while intra-month repetitions are conspicuously absent. The influx of new helpline callers is remarkably stable, at a rate of about 1,200 per month. This rate remains constant over time, showing virtually no fluctuations across months or years. The observed probability distribution of call durations cannot be adequately explained using simple modelling techniques. The complexity of this distribution indicates that a mixture of distributions generated by several sub-populations of callers is being observed. The dataset comprises a number of fields, however only the following fields were used in this study: the date-time stamp of the call arrival precise to the last second; the Boolean engaged field meaning that the call was dropped with a busy tone; the answered flag meaning that the call was passed to a volunteer from the charity; the duration of the call in seconds; and the unique caller ID. The caller IDs allowed us to enumerate the callers uniquely while providing no personally identifying or sensitive details. The IDs were associated with most, but not all, call arrivals. About 20% of the calls had the caller ID missing. At a first glance, the duration of calls appears to follow an exponential decay distribution, where the volume of calls decreases at a rate proportional to call duration. However, the distribution follows a complicated decay pattern with a rate that associates with the call duration in a non-obvious manner.

Machine learning

In this study, cluster analysis involves grouping a set of objects (e.g,. callers based on their attributes) in such a way that objects in the same group (called a cluster) are more similar to each other in comparison to other groups (clusters). Clustering was initially performed for 2013-2015 timespan, and then on the 2016 data to check for stability of our findings. Only the calls from callers with a unique identifier were used (44,613 callers collected in 2013–2015, increasing to 61,287 unique callers by the end of 2016). Callers were clustered using 3 caller attributes: number of calls; mean call duration; and standard deviation of call duration. We selected these parameters due to their explanatory power: the number of calls a person makes indicates their frequency of help seeking behaviour; the mean call duration indicates call length; and the standard deviation of call durations indicates a person's variability and consistency in conversation length. We also used K-means clustering algorithm in this case study.

Five clusters were revealed by the k-means algorithm. The first cluster named 'Elite prolific callers' had the largest average number of calls per caller in the cluster and the smallest cluster size. A handful (fewer than 50 callers over the 4-year time span) of extremely prolific callers, responsible for 20% of the total call volume. They call thousands of times, each call on average lasts about 4 minutes, with a small minority of calls lasting 10 minutes. The second cluster named 'Typical callers' had the largest cluster size. The majority of callers who call 5-6 times and almost always have a short 3- to 4-minute conversation each time. This cluster accumulates 40 to 50 percent of all callers depending on the time slice under consideration. The third cluster named 'Standard prolific callers' had the second largest average number of calls per caller, middling average call duration and the largest unexplained variability encompassed by the cluster. About 12 to 15 percent of callers are prolific, each calling hundreds of times and having call durations that are moderate in length (from a few minutes to half an hour long). The fourth cluster named 'Unpredictable erratic callers' had the largest average standard deviation of the call duration. About 3 to 5 percent of callers whose call duration varies considerably, with some calls lasting 3 minutes and some up to 1 hour. The final fifth cluster named 'One-off chatty callers' had the smallest average number of calls per caller accompanied by the largest average call duration. About 13 percent of callers that only call 1-2 times, have a long 30 minutes to 1 hour conversation, and do not return for any sustained support (the operational opposite to prolific callers).

Data mining of the association of caller IDs with the volume and duration of calls revealed several caller clusters, each describing a distinctive behaviour type. The most striking of those clusters, Elite Prolific callers, encompasses a small number of caller IDs responsible for a substantial share of the total call volume that the charity receives. Early identification of callers of this type and routing their calls to specialized advisers provides insight in the modelling of healthcare service usage, offering insights for evidence-based practice and operational decision-making.

4. DISCUSSION

The work presented in this paper introduces a new data analytics process workflow, called Health Interaction Log Data Analysis (HILDA), designed to accommodate processing of user, system event and interaction data logged on products and services that use digital interaction technologies. The HILDA workflow is specifically designed for use in log data pertaining to health and wellbeing products and services. We are seeing a rapidly growing use by individuals of these health and wellbeing products and services as people realise benefits provided by lifelogging or quantified self platforms in such areas as exercise, diet management and mood. HILDA is designed to enable the capture of best practices in user, event and interaction log preparation, data prospecting and machine learning phases across a broad range of health and wellbeing usage scenarios. The workflow has been developed in response to the practical needs of our research to make use of more standardised approaches in data analytics research; but more broadly in recognition of the opportunities and threats involved as public health sector providers seek to make use of these kinds of technologies. Two case studies have been presented to illustrate the utility of the HILDA workflow in helping the transitioning from the data to actionable knowledge.

Workflows such as HILDA can help ensure reproducibility of findings as knowledge is derived from log data and also support consistency and accuracy. Broader topics such as data protection, compliance with ethical guidelines, etc., need to be accommodated within these kinds of workflows, especially as the EU General Data Protection Regulation (GDPR) (https://www.eugdpr.org/) comes into force in 2018. Other ethical concerns need to be discussed, for example the ethics of logging geolocation with each event or interaction with a smartphone health app. Moreover, can supposedly anonymous log analysis be used to identify individuals? Also, the ethics of consent or lack thereof for logging events need to be considered. Websites often log events and make use of logging tools like Google Analytics without obtaining consent from the user. However modern web browsers now seek consent for using cookies so the future for consenting on elusive data collection on the web has perhaps progressed.

The HILDA workflow describes three stages or phases, including data preparation, data prospecting and machine learning. Once the machine learning stage is complete, the results can lend themselves to being labelled and described in clear terms that are easily understood beyond the data science community. This is the case in our two case study examples. However, this is not always the case. Often the results of the machine learning stage if, for example generated using deep neural network methods offer little in terms of rationale for results, which can hamper take up, either because of the inherent complexity of the findings or the need for any deployment to be fair, accountable and transparent perhaps due to legal reasons. There are many other uses of log data not mentioned in this paper, for example, building a recommender engine to understanding the feature preferences for similar users, perhaps taking a collaborative filtering approach. Moreover perhaps a Markov model can be used for state predictions using a probability matrix for switching between states. In the second case study, we did use Fourier analysis to determine dominant frequencies of calls and to filter-out non-dominant frequencies to elicit a more noise free model of call volume over time. Hence, there are many techniques that be applied to log analysis that are not necessarily represented in this paper.

The HILDA workflow could arguably benefit from a fourth phase or stage that helps signpost how the knowledge and or model generated can be taken up by the sponsor or community. Variously, this could be called deployment, uptake or operationalisation.

A major limitation of event log analysis is that it is very focused on quantitative analysis and requires significant interpretation. As such, there is no way of knowing why users take a certain journey or prefer a certain feature or even why they have certain patterns and frequency of use. Put differently, log analysis does not provide any qualitative feedback. However, this can be partly addressed using Ecological Momentary Assessment (EMA aka experience sampling) to add qualitative to the quantitative. EMA involves prompts to users often as pop-up questions during the service experience to gather user opinion or state. As a result, we would recommend considering EMA to augment user log analysis. However, EMA could be a distractor to the service itself so it needs to be carefully considered.

5. CONCLUSIONS

In summary, this paper presented a relatively unsophisticated model for carrying out user log analysis and applied this process in two healthcare services that automatically logged events. We found that that both case studies provided business intelligence and insights that are useful for understanding service users and for improving the service itself. We would support the integration of event logging in digital health given such data can be easily collected and analysed.

In addition to the work presented on the two case studies in this paper, the research team are exploring the analysis processes involved in carrying out data analytics in several other projects. In one of these, the data pertains to the use of an app by people with learning disabilities as they rate their transport or consumer experience in their daily life. This data therefore has additional locational information, and requires careful ethical considerations as the data relates to vulnerable users. Future development of the method therefore needs to address ethical aspects as a core part of the process, perhaps to be considered horizontally across all technical phases of the process. In another project, users access a maternal mental health app that helps them to compile and curate an event diary relating to mood in addition to completion of health and wellbeing scales. The data in the project can, when analysed, offer insights back to the community of users and so the process need to take account of deployment issues.

ACKNOWLEDGEMENTS

The Reminiscence 'Inspired' app was co-created and designed with help and input from the Alzheimer's Society, Reminiscence Network Northern Ireland and people living with dementia and their carers. The authors gratefully acknowledge their support and the funding support provided by HSC R&D Grant COM/5016/14 for the project. Financial support for the metadata study was provided in part by Samaritans Ireland together with Ireland's National Office for Suicide Prevention. Our thanks to the research teams who worked on the projects with us, including Ciaran Moore, Colette Ramsey, Cherie Armour for the Samaritans project, and Aideen Gibson, Claire McCauley, Liz Laird, Brendan Bunting and Finola Ferry for the Inspired project.

REFERENCES

[1] K. Miller, P.J. Woollam, G. Powell, D. Hitchings, D., and J. Stallard (2007) A Rehabilitation Device Data Logging System, Disability and Rehabilitation: Assistive Technology 2(1):9-14.

[2] Vagner Figuerêdo de Santana, M. Cecília C. Baranauskas (2010) Summarizing observational client-side data to reveal web usage patterns. In Proceedings of the 2010 ACM Symposium on Applied Computing (SAC '10). ACM, New York, NY, USA, 1219-1223.

[3] Daniel Woo, and Joji Mori. 2004. Accessibility: A Tool for Usability Evaluation. In book. Computer Human Interaction: 6th Asia Pacific Conference, APCHI 2004, Rotorua, New Zealand, June 29-July 2,

2004. Proceedings, edited by Masood Masoodian, Steve Jones, and Bill Rogers, 531-39. Berlin, Heidelberg: Springer Berlin Heidelberg.
[4] Morrison C, Doherty G, Analyzing Engagement in a Web-Based Intervention Platform Through Visualizing Log-Data J Med Internet Res 2014;16(11):e252, DOI: 10.2196/jmir.3575.
[5] Shearer C., The CRISP-DM model: the new blueprint for data mining, J Data Warehousing (2000); 5:13—22.
[6] Jason Haffar (2015) Have you seen ASUM-DM?, SPSS Predictive Analytics, IBM, Located at: https://developer.ibm.com/predictiveanalytics/2015/10/16/have-you-seen-asum-dm/, Last accessed 24 January 2018.
[7] Thickett, J., 2006 Connecting Older People: Consumer Engagement with Digital Services, London, Ofcom.
[8] Butler, R.N. (1963). The life review: an interpretation of reminiscence in the aged. Psychiatry, 26: 65-76.
[9] P.G. Archbold B.J. Stewart M.R. Greenlick T. Harvath (1990) Mutuality and preparedness as predictors of caregiver role strain. Research in Nursing & Health. 13: 375-384.
[10] T. Wright (2009) Drawn from Memory: Reminiscing, Narrative and the Visual Image, Proceedings of the First International Workshop on Reminiscence Systems (RSW-2009), Cambridge, UK, 5 September, 2009 pp:37-42.
[11] A. Sixsmith, G. Gibson (2007) Music and the wellbeing of people with dementia, Ageing and Society, vol. 27, pp. 127-145.