Modeling retail browsing sessions and wearables data

Rolando Medellin-Gasque, Henrik Nordmark, Anthony Mullen, Luca Citi, Aris Perperoglou and Berthold Lausen

Abstract The advent of wearable non-invasive sensors for the consumer market has made it cost-effective to conduct studies that integrate physiological measures such as heart rate into data analysis research. In this paper we investigate the predictive value of heart rate measurements from a commercial wrist wearable device in the context of e-commerce. We look into a dataset comprised of browser-logs and wearables data from 28 individuals in a field experiment over a period of ten days. We are particularly interested in finding predictors for starting a retail session, such as the heart rate at the beginning of a web browsing session. We describe preprocessing tasks applied to the dataset and logistic regression and survival analysis models to retrieve the probability of starting a retail browsing session. Preliminary results show that heart rate has a significant predictive value on starting a retail session if we consider increased and decreased heart rate individual values and the time of day.

Rolando Medellin-Gasque, Luca Citi, Aris Perperoglou and Berthold Lausen
University of Essex, Wivenhoe Park, Colchester CO4 3SQ
✉ (rmedel,lciti,aperpe,blausen)@essex.ac.uk

Henrik Nordmark
Profusion, Octavia House, 50 Banner Street, London EC1Y8ST
✉ henrikn@profusion.com

Anthony Mullen
Gartner, Tamesis, The Glanty, Egham, Surrey TW20 9AW
✉ ant@flyingmountain.co.uk

ARCHIVES OF DATA SCIENCE (ONLINE FIRST)
KIT SCIENTIFIC PUBLISHING ISSN 2363-9881
Vol. -, No. -, -

1 Introduction

The Internet of Things (IoT) is defined as a network of physical devices embedded with electronics, sensors, and network connectivity that enables the collection and exchange of data. One of the most widespread type of devices in the IoT spectrum are activity trackers or “wearable” devices that are typically worn next to the body.

Some of these trackers are able to collect not just movement data through accelerometers and pedometers but also heart rate, location and skin temperature (Ghaoui, 2005). The use case presented in this paper refers to a wrist activity tracker with heart rate measurement capabilities. This device was worn by participants throughout their day capturing heart rate when performing a number of activities including the browsing of retail pages. Since its inception, e-commerce has revolutionized the way consumers shop online by providing a secure fast and convenient way to shop for products from home. Retail websites can improve the shopping experience of the consumers, and therefore maximize revenues, by analyzing the data generated by online behavior. Online activity can also reveal interests or preferences that can be analyzed to produce tailored advertising. On e-commerce websites product browsing and other behaviors can be used to create tailored recommendations (e.g., Pazzani and Billsus (2007)) in real-time. This paper presents preliminary research that focuses on merging and analyzing a dataset comprised of wearables data and browser-logs from a field experiment where 28 individuals wore wearable devices and were “tracked” continuously for a period of ten days. The objective of this paper is threefold 1) describe a dataset comprised of aggregated data from wearable devices and online browsing activities from a field experiment 2) describe the techniques and algorithms used to process and align browser-logs and heart rate data and 3) provide a methodology to model retail-browsing and physical data provided by wearable devices.

The are two parallel motivations for this research. On the one hand, this field experiment provides an opportunity to deal with large datasets coming from IoT devices and learn how to process this data effectively, which is important given that data generated from IoT devices is likely to grow dramatically in the next few years. The other parallel motivation is to attempt to answer a simple business question: “Is heart rate a predictive variable for the likelihood to shop?”

The accuracy and data availability limitations of wearable devices is still under scrutiny. Related to the accuracy, as these are not medical grade devices the accuracy is expected to be low and although some of these issues can be alleviated using statistical methods such as smoothing of the data, it nevertheless introduces noise. But these devices are becoming more accurate and widespread with companies like Fitbit and Jawbone investing 150 million USD in the past year in new research and development. With regards to the data availability limitation, these devices are often protected or restricted in some way by the manufacturers making it hard to obtain data directly from the device. It will be interesting to see whether the data availability issue will continue to stand. The main market for these devices is currently with early adopters, who tend to be rather forgiving. However, the predictions for wearable devices sales by the end of 2016 is expected to grow 18.6 percent compared to 2015 figures (Inc., 2016). Consequently, some of the benefits of this research lie in the fact that we start exploring problems related to the acquisition, pre-processing and normalization of the data obtained from these consumer grade devices with the aim of applying this in a context that is slightly different than its mainstream use as a health and fitness device .

Predicting e-commerce behavior poses several modeling challenges, for example in (Sismeiro and Bucklin, 2004) the authors discuss an approach to model purchase behavior using data from what users do online, but in reality the act of buying online is influenced by several factors, online stores reach a diverse user population and models must account for the corresponding user heterogeneity. While the likelihood of people buying online has increased and it is easier nowadays to know what users do (i.e., what they click, download, etc.), for how long and to what they are exposed, online purchases are sparse events in the context of all web behavior and this can lead to a lack of predictive and explanatory power from models (Sismeiro and Bucklin, 2004).

The weblogs dataset is comprised of user identifiable browsing logs obtained by exporting web browser history logs from the web browser Chrome [1] for 28 participants. Browsing activity was tracked continuously both at home and at work. This enabled us to identify web sessions for a user even across desktop devices. Unfortunately, we were unable to track web behavior done on smartphones or tablets.

While the weblogs dataset we obtained is not necessarily as rich as what can be tracked using web analytics technology such as Webtrends, Omniture

[1] www.google.com/chrome/

or Google Analytics, for our purpose it is sufficient since it contains user identifiable webpage access time-stamps of URLs accessed over a period of time. Furthermore, browser logs give us access to all the webpages users accessed disregarding the type of webpage, which gives us a better picture of an individual's entire web journey. While browser logs are usually private and getting access to large datasets of such individual level web behavior is difficult because of privacy issues, we want to define a methodology taking advantage of the diversity of sources of the dataset in order to understand when individuals decide to engage in a retail session. Most data analysis research on online behavior focuses on extracting patterns using interaction metrics such as: time spent on a webpage, crossing from one device to another and tracking events like scrolling, clicking or hovering with a mouse. More detailed models sometimes incorporate a semantic layer by including webpage content. Interest in specific types of content is gauged by observing the length of time spent looking at various product types, the items searched for using the search field and items added to the virtual shopping basket whether they are purchased or not and of course the items that are actually purchased if any. However incorporating "external" (i.e. not generated from online behavior) datasets to the online model behavior presents several challenges in terms of acquisition and relevance.

Our external dataset is one obtained from a wrist activity tracker wearable device. As consumer grade devices provide automated measurements of heart rate variability, researchers have been provided with a simple tool for both research and clinical studies. However, streamlining and automating the process of obtaining and cleaning the raw data from an IoT device is an important and frequently non-trivial task since these devices don't always have good documentation for their API's or worse the documentation is sometimes non-existent. Wearable devices are becoming more common and widespread creating new demands and context for data-driven systems (Billinghurst and Starner, 1999). Wearable technology provides a continuous stream of behavioral data, typically by minute or second, which opens up new avenues for data analysis. At the moment, wearable devices are capable of tracking steps, measure heart rate and use GPS to track location. These measurements can be used by "affective computing" (Picard, 1997) algorithms for unobtrusive monitoring of the user's physiological and cognitive state. In particular, continuous monitoring of heart rate dynamics can provide a non-invasive observation window on the autonomic nervous system activity, which in turn provides information on the emotional state of the individual (Kim et al, 2004). Previous studies have shown that the analysis of an individual's heart rate can be used to track performance and alertness during sleep deprivation (Citi et al, 2010), for the instantaneous estimation of the depth of anesthesia (Citi et al, 2012; Valenza et al, 2014a), to monitor signatures of sleep fragmentation (Citi et al, 2011), and for the automatic recognition of different mood states, both in normal and pathological conditions (Valenza et al, 2015). Particularly relevant to the analysis of browsing behavior and online shopping are the studies investigating the use of heart rate for emotion recognition (Calvo et al, 2010; Valenza et al, 2013, 2014b). Despite these promising preliminary results, the reliable measurement and the correct interpretation of these physiological correlates is more challenging than commonly appreciated. In particular, there is a risk of drawing excessive or unfounded conclusions from the data (Malik, 1996).

Data science consultancy Profusion[2] ran a project in 2015 to explore the limits and applications of data collected and analyzed from wearable devices. For ten days Profusion tracked behaviors and attitudes of 28 employees who volunteered to wear a Fitbit™Charge HR [3] device. The location of the volunteers was tracked using the smart-phone app Moves. Finally, the volunteers were also asked to self-report their emotional state and other contextual information at regular intervals. In addition to providing some interesting datasets to analyze, this field experiment provided an opportunity to better understand human behavior in relation to an emerging piece of technology that is likely to become much more widespread in the near future. It allowed the employees of Profusion to gain first hand experience with the use of wearable devices in the workplace including thinking about the ethical implications of conducting such a study and experiencing the psychological impact of being tracked continuously for ten days. For a further discussion on the Profusion wearables experiment of 2015 more information and on the use of wearable devices in the workplace see (OConnor, 2015).

The remaining of the paper is as follows: Section 2 presents descriptive statistics of the dataset and describes some pre-processing tasks. Section 3 presents linear regression and survival analysis models related to the time to engage in retail sessions using heart rate as a feature. Section 4 presents a discussion of the results and their implication in the context of retail. Section 5 concludes the paper.

---

[2] www.profusion.com

[3] www.fitbit.com/uk/chargehr

2 Dataset Definition

The dataset was obtained by tracking behavior and attitudes of 28 employees who volunteered to wear an activity tracker device, self-report their emotional state, use location tracking software on their smart-phone and record their online browsing behavior. Profusion followed a strict data confidentiality process. Participants signed a consent form with details informing them of the data that would be collected from them, the frequency at which it would be collected and who would have access to this data. From an on-boarding questionnaire, we obtained demographic details of such as age (M=34.54 SD=9.04), sex (14 males, 14 females), address, height, weight, marital/family status, distance to work, living arrangements, etc. Also, we installed an app in their smart-phone that prompted them to report every 2 hours starting from 7am their emotions, attitudes, stress and food consumption (including alcohol and cigarettes). From the Fitbit ™wearable device we obtained: heart-rate measurements, number of steps, sleep duration, and calories burned. From their browser history, we obtained raw logs of every webpage accessed at work and at home from desktop devices. Finally, the names of the employees was replaced with a user id number to provide some degree of anonymization even though this is imperfect with such a small group of individuals since the identity of an employee could in many instances be inferred if for example they are the only person in a particular age bracket.

The *wearables dataset* comprises number of steps, heart rate measurements in beats per minute (bpm), and calorie consumption every 15 minutes, which was the frequency at which we could obtain a measurement for all variables being tracked by the device. The total number of heart rate records in the dataset was 26,379 for 19 users who agreed to provide their browsing history (average age is M=32.74, gender 9 women and 10 men) . The raw data was obtained by accessing the personal cloud storage account provided for every Fitbit user. This was done using Fitbit's API for developers. Since the frequency of the measurements was reported every 15 minutes we used linear interpolation to obtain minute by minute data that could aligned with the time stamps of the weblog entries. Measurements below 46 bpm were considered to be very low heart rate measurements that might be inaccurate due to the limitations of the device were rounded up to 46 bpm. Figure 1 presents the individual heart rate variability in beats per min for each participant at the beginning of a retail session.

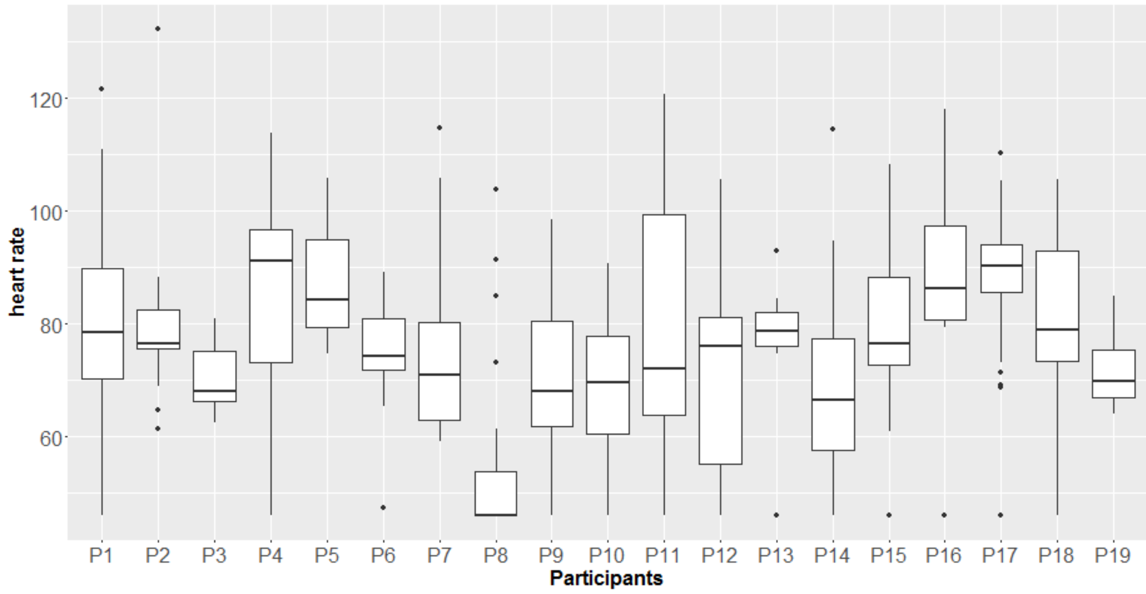

Fig. 1 Heart rate individual variability for each participant *P* at the beginning of a retail session. The bottom and the top of the box are the 25th and 75th percentile and the band near the middle of the box is the 50th percentile. The whiskers extend to the most extreme data point which is no more than 1 times the interquartile range from the box.

The weblog dataset consists of 27,624 URL entries from 19 users over a ten day period. We considered only retail URLs from Amazon UK [4], Asos [5] and Argos [6] (following a manual analysis of the most important retail domains), resulting in 2044 "retail" URLs in the dataset. A raw browser log entry contains: an anonymous user identification, the full URL accessed, the webpage title, the date and time of access, a visit count for the number of visits to the same URL, a typed count for the number of times that the user typed the URL address directly to visit the page, the referrer URL for visits that originated on some other page, an alphanumerical identification for the visit, the Chrome profile used and the URL character length. We added to these raw variables, the domain name of the URL, a main category and a subcategory for the URL obtained from an external URL domain tagger, a binary variable indicating that the URL was from a retail website and the time spent on a webpage (estimated by calculating the difference between the start time of two consecutive webpages). The next challenge is defining what is meant by a "web browsing

[4] www.amazon.co.uk
[5] www.asos.co.uk
[6] www.argos.co.uk

session”. Time stamps of when a webpage is accessed are punctual events. How does one group punctual events into a session? A simplistic approach would be to group URLs that occurred during the same hour of the day or some other arbitrary time frame, but this approach presents the problem of possibly fragmenting webpages that intuitively belong to the same browsing session if what we intuitively think of a web browsing session happens to span two or more of these time frames. Conversely, it could group together webpages that were accessed at the beginning of the time frame and at the end of the time frame but for which there is a big gap in between the two time stamps in which no webpage was accessed and would thus indicate that they actually belong to distinct web browsing sessions. This is the reason why we choose to define web browsing sessions in terms of idle periods of time between the access time of consecutive web pages. A browsing session for our purpose is defined as a sequence of time ordered URLs whose access time is separated by less than α minutes. We chose α to be 60 minutes. We experimented with several values for α deciding on 60 minutes as this value returned a fair number of sessions per user but also did not fragment sessions into very small web browsing stints. Therefore, any two webpage visits that occur within 60 minutes or less of one another belong to the same web browsing session. Once the time between two consecutive webpages is more than 60 minutes this constitutes a rupture that demarcates the end of one browsing session and the beginning of the next browsing session, see Figure 2. Once the sessions were calculated in the way we just described, we labeled sessions as “retail sessions” if they had at least one webpage from the set of targeted retail URL domains. We left out other known retail domains and focused only on these three since they represent 96% percent of the retail domains in the dataset. We obtained 533 browsing sessions (M=164 URLs per day per user) from which 47 were retail sessions. These 47 retail sessions were distributed among only 8 users. And one user had a total of 18 retail sessions. Thus, the retail sessions are not distributed very evenly among users. Once the sessions were created we added the following features to the weblog dataset: a session identification, the time to retail, i.e. the time it takes for the user to go from non-retail pages to the first retail page in the session and the total session time. We merged both datasets using the date, the time and the user in order to calculate what the estimated heart rate was at the beginning of each session. Retail sessions are associated with a higher heart rate when the session starts but only if we account for the time of day (i.e., mornings, afternoons, evenings). Figure 3 shows an increased heart rate in the afternoon for retail sessions relative to the non-retail sessions.

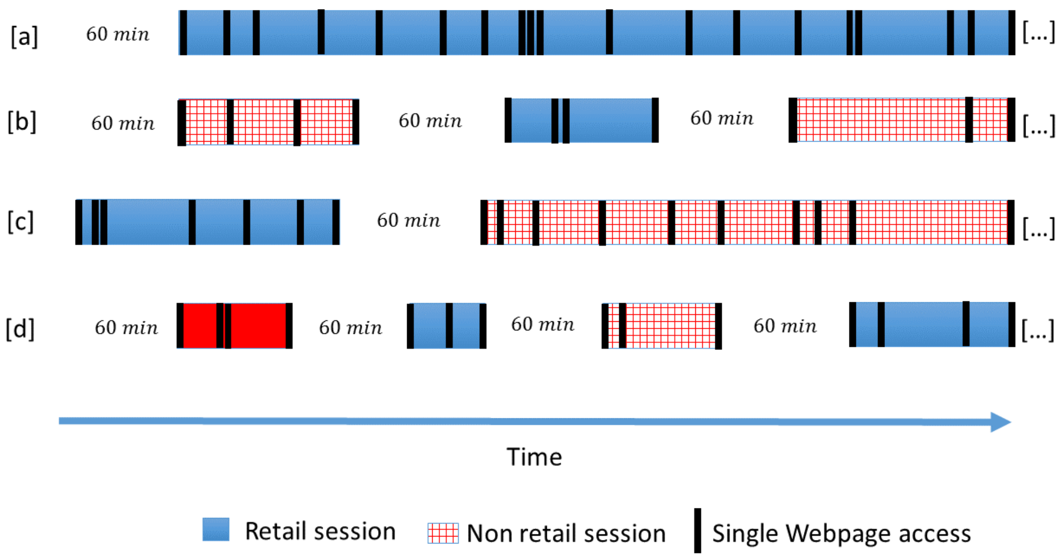

Fig. 2 Rectangles represent sessions and black lines represent the access time of webpages. Access times sessions are less than 60 minutes. Sessions can be comprised of retail and non retail URLs but if the session has at least one URL with a retail domain is considered as a retail session. Therefore any given period of time can have one (e.g., [a], [b], [c]) or several retail sessions (e.g., [d]).

3 Models

In this chapter we discuss some specific challenges in modeling and predicting the probability of entering a retail session. The purpose of our models is to predict individual level online retail engagement behavior. We present a logistic regression model to estimate the probabilities of entering a retail session using features such as the time of the day and the heart rate. We also present a survival analysis model to account for right censored events (i.e. sessions where users do not enter retail URLs) and investigate if the same features tested in the logistic regression model have a significant value if we account for the time to retail.

3.1 Logistic Regression Models

We present a logistic regression model to check whether some of the variables in the wearables dataset are predictive of engaging in a retail session. We fit a mixed effects logistic regression model using a discretized heart rate variable with three different levels: increased, decreased and baseline. Additionally, we

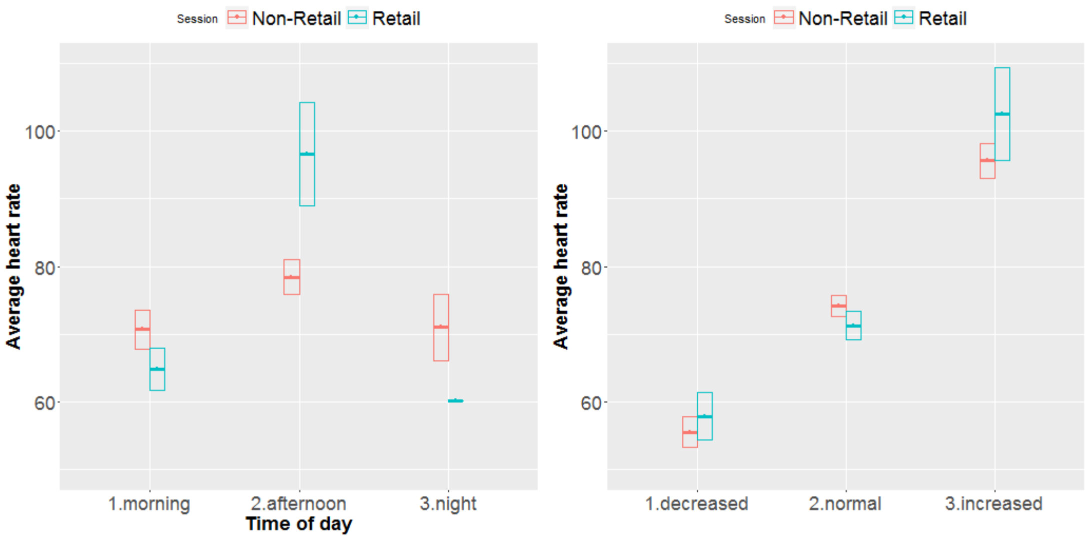

Fig. 3 Retail sessions are associated with a higher heart rates in the afternoons. The plot on the left shows heart rate distributions across the day for retail and non-retail sessions. The plot on the right presents heart rate distributions of sessions with increased and decreased heart rate measurements. The increased heart rate is defined by more than +/- 10 bpm away from individual average heart rate.

also have time of day as a discretized categorical variable with three levels: morning, afternoon and night. And finally, a variable to capture the effect of individual users. The increased/decreased heart rate is defined by more than +/- 10 bpm away from each individual's average heart rate. This leads to 15 events with an increased heart rate, 13 events with a decreased heart rate and 19 events with a baseline heart rate. A simple ANOVA analysis, see Table 1, show that both the time of day and the heart rate are statistically significant. The logistic regression model, see Table 2 shows that an increased heart rate at the beginning of web browsing session indicates a higher likelihood of entering a retail session.

Table 1 ANOVA between variables.

Variable	mean sq	F-Value	Pr(>F)
$time\_of\_day\_afternoon$	0.41012	4.0512	0.04482
$time\_of\_day\_morning$	0.96017	9.4846	0.00221
$increasedheart\_rate$	0.44168	4.3629	0.03738
$decreasedheart\_rate$	0.11435	1.1295	0.28853
Residuals		0.10123	

Table 2 Logistic regression with random effects using the increased and decreased heart rate

Random effects Groups	**Name**	**Variance**	**Std.Dev.**	
user	(Intercept)	7.415	2.723	
Number Observations	395	groups:user,19		
Fixed Effects Variable	**Estimate**	**Std.Error**	**z**	**Pr(>|z|)**
(Intercept)	-3.3913	1.1215	-3.024	0.00250 **
increased_heartrate	1.2827	0.6373	2.013	0.04414 *
normal_heart-rate	-0.3318	0.5249	-0.632	0.52729
time_of_day afternoon	-1.7410	0.5436	-3.203	0.00136 **
time_of_day night	-2.1028	1.1191	-1.879	0.06024 *

Significance codes: *** 0.001, ** 0.01, * 0.05, . 0.1

3.2 Survival to retail analysis

The survival function in this context tells us the probability that the time to enter a retail session is later than some specified time t see Figure 4. A lifetime is defined from the time of starting a browsing session until the time of entering a retail webpage or censoring, whatever comes first. Of course, these lifetimes are correlated since several sessions are created by the same user. As such, we have an analysis of repeated events for a specific group of people that may be prone to experience the event again and again at a different stage in their browsing history. To account for this correlation structure we will be using a random effects model (shared frailty Cox model Hougaard (1995)).

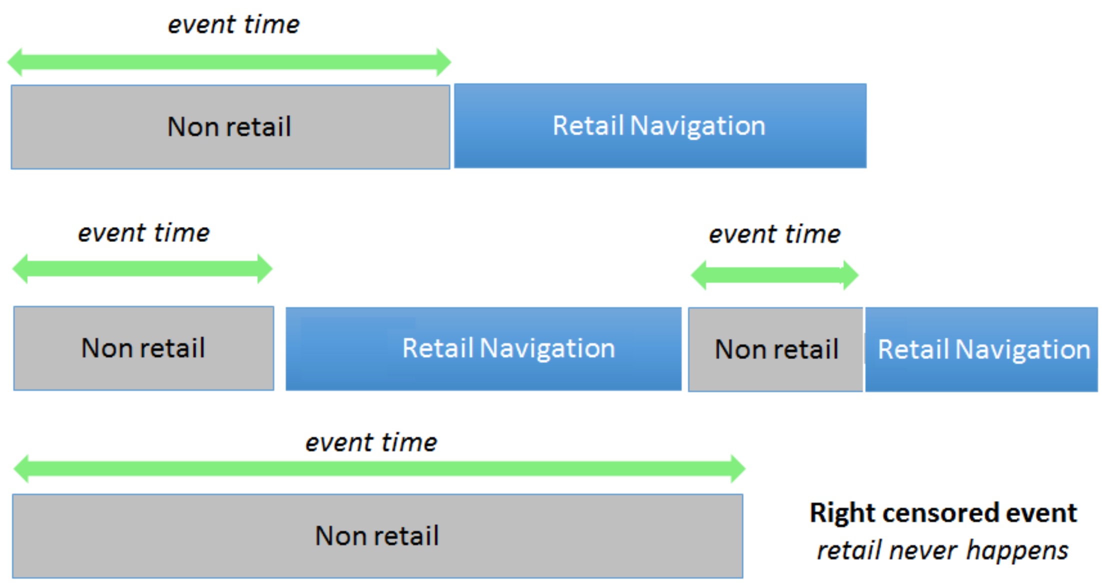

Fig. 4 Event time for web-log sessions

In our context, sessions “die” when a retail page is visited. Defining “death” in terms of visiting a certain type of webpage is not entirely natural for the usual interpretation of survival analysis models since death is usually considered to be a one time event per user. In our case the user can return to browsing non-retail webpages and is thus “resuscitated”. Finally, a censored session in our context is one in which a retail page is not visited within time frame of the web browsing session. For the purpose of building the model, we did not take into account sessions starting directly at retail domains (7 such sessions in total) since there is no transition signal from non-retail to retail domains. Using the hospital patient metaphor, this would be equivalent to starting out with patients who are already dead at the beginning of the study. We created two survival models, the first one presents sessions related to increased/decreased heart rates and the second one presents the sessions' time of day of the sessions as features. In the Kaplan-Meier plots the censoring is taking place early on (sessions tend to be shorter than 100 minutes). The median session lasts 42 minutes (min = 0, max = 527) and the median survival time is approximately 350 minutes.

In the first model, as in the logistic regression model, an increased or decreased heart rate increases the probability to engage with a retail webpage (p=0.00065 and p=0.0028 respectively), see Table 3 and Figure 5. Increased and decreased heart rates are both deviations from a normal situation and increase the probability to enter retail webpages. We have checked if this can be explained by the non-valid heart rates (equal to 46 bpm) which seem to decrease the probability, but this does not explain the effect since we have significant values again if we replace the time to retail with the number of pages accessed, see Table 5. It is also interesting to see that both, increased heart rate and decreased heart rate are significant relative to the normal heart rate. This suggest that having an increased or decreased heart rate increases the probability of entering a retail session.

Table 3 Survival analysis frailty model using increased and decreased heart rate.

	Using time to retail			Using number of pages		
Variable	**coef**	**se**	**p**	**coef**	**se**	**p**
increased.heartrate	1.841	0.530	**0.0052**	2.575	0.557	**3.7e-06**
decreased.heartrate	1.220	0.427	**0.00430**	1.760	0.455	**0.00011**
timeofdayafternoon	-1.270	0.453	**0.00504**	-0.844	0.491	**0.08564**
timeofdaynight	-1.686	1.073	0.11593	-0.954	1.088	0.38032
frailty(user)			5.8e-10			5.8e-07
Variance of random effect	5.15			4.54		

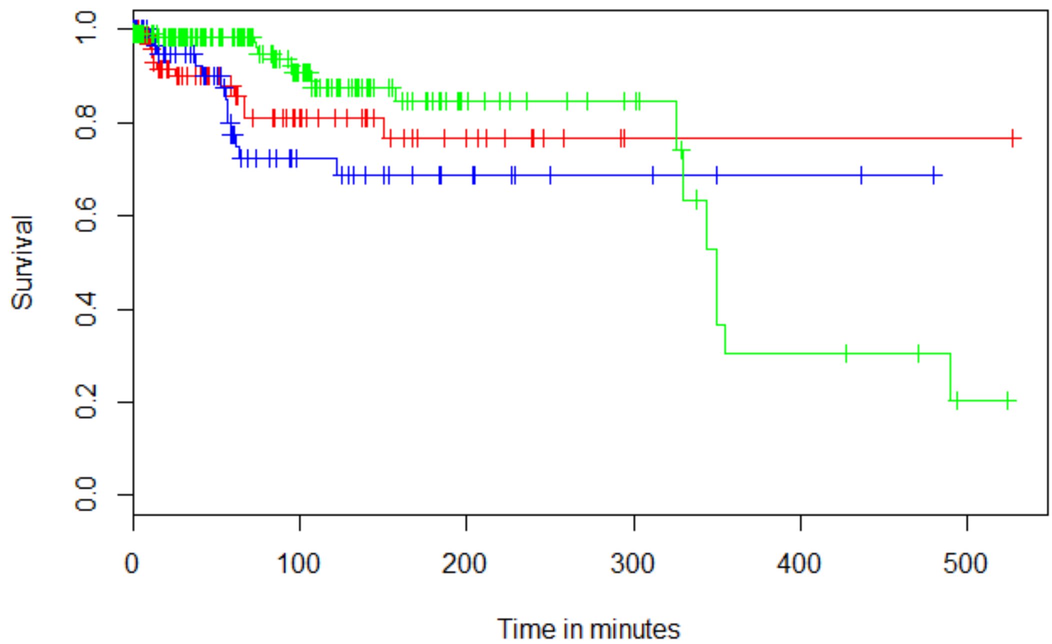

Fig. 5 Kaplan-Meier estimates of the survival functions for sessions with increased (red curve), normal (green curve) and decreased (blue curve) heart rates.

In the time of day analysis (see Figure 6 by fitting a frailty model, the heart-rate is significant and increases the hazard of going into a retail session. Also, it is more likely to enter a retail session in the morning than in the afternoon.

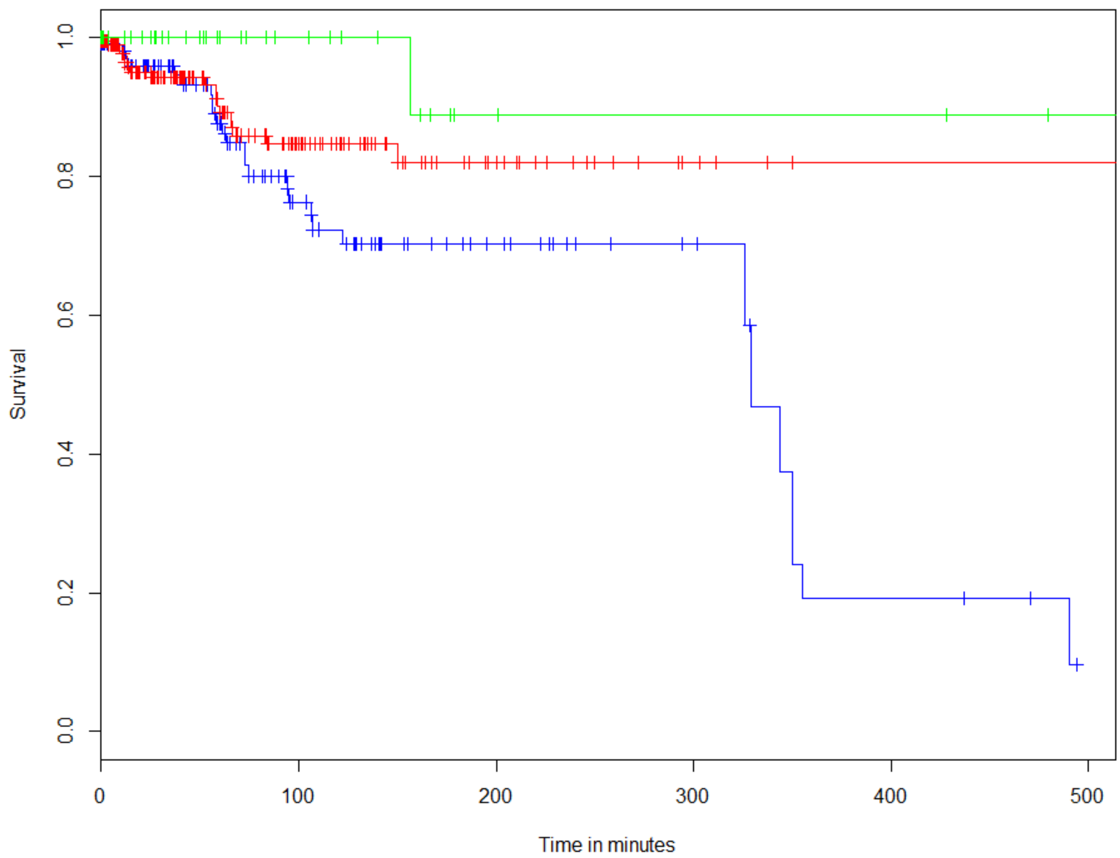

Fig. 6 Kaplan Meier curves for session lifespan survival accounting for the session time of day. The blue curve indicates morning sessions, the red curve indicates afternoon sessions and the green curve indicates night sessions.

4 Discussion

We used browser logs data and aligned wearable heart rate data to predict the start of retail browsing. Heart rate measures were used as a predictor on both the linear and the survival models. Preliminary results show that heart rate has a significant predictive value on starting a retail session which is stronger for sessions with increased heart rate and accounting for the time of day. Particularly, increased and decreased heart rate increase the probability of engaging in a retail session in both models suggesting a non-linear effect of heart rate on the likelihood of entering a retail session.

Wearables provide a new type of rich behavioral data that industries outside of fitness applications e.g. retail are still considering how to use. Fitness applications have emerged as the front-runner in the adoption of wearable technology. Retail organizations are already beginning to implement wearable technology to improve productivity across business processes. Wearable technology will to become more widespread over the next several years, and retailers should prepare to exploit data coming from these devices by offering products and services that re-use these data feeds. While wearable technology encompasses several types of devices (e.g., fitness monitors, activity trackers and smart clothing) retailers can benefit most from smart watches, as they present new opportunities for customer engagement providing contextual information about products or services.

In this context retailers and IoT data analysts need to provide evidence of the capabilities that wearable technology adds to the larger retail ecosystem, offline and online, and how systems and devices can complement each other with their own data and open datasets by running trials, POCs, innovation labs or partnerships. While most of the research on online retail sales focus on modeling customer journeys the research presented in this paper discusses models that predict the probability to engage with an online retail website given heart rate readings from a wearable device. This preliminary research aims to influence online retail marketing and advertising by providing statistical models that inform targeted interventions. While this is a preliminary study and the datasets used are usually private or not widely available, we foresee wearables data will be exploited in several areas, retail sales included.

We are fully aware of the limitations of this study but expanding sample size and granularity of measures for this model will provide changing insights and methods for the retail industry. One important limitation in the weblogs dataset is that the dataset did not have the number of opened browser-tabs at a particular time nor the time each webpage was closed, which prevents us from building a complete survival analysis model. We worked around this limitation by defining sessions that grouped URL entries using an arbitrary threshold, limiting the impact of not having the time when a webpage was closed and using the access time between webpages and fitted the data into a survival analysis problem. One limitation regarding the wearables dataset was the frequency with which heart rate was sampled, a limitation imposed by the device API at the time of the controlled experiment. A more frequent sampling would have helped to detect anomalies in the dataset and apply a smoothing strategy to the heart rate dataset. Another limitation was that the time spent in a single webpage can only be inferred by the differences between the access times for two consecutive webpages. We cannot be sure that a user effectively spends that time actively looking at a single webpage given that we do not track for concurrent open tabs within the browser or any activity within the webpage such as mousing or scrolling.

5 Conclusions

The motivation of this paper was to understand what triggers a retail session using predictors such as the heart rate at the beginning of the session. We employed logistic regression and survival analysis to model people engaging in retail sessions. The fact that we had a wearables dataset aligned to these sessions presents a novel approach to analyzing wearables data and allowed us to present preprocessing and data formatting tasks for both datasets. Preliminary results show that the increased heart rate is a strong predictor to engage in a retail session in both models. Wearable data is set to provide a much richer contextualized view of human behaviour. As consumers begin to use more complex devices that provide heart rate variability and skin galvanic response techniques will become more accurate to predict the emotional state of users. The study described was in the position of having access to a wealth of data about consumers therefore this rich depth of shopping behaviours complimented viewed through a physiological lens will not come easily for many brands. Even though these systems may not be used in real-time for some years it is clear that the approach is a preliminary step where closed private results can be applied to the many.

The implications of wearables in retail are more relevant for in-store shopping. According to a report (PowerReviews, 2000) 82% of people surveyed wanted wearable devices to enhance their in-store experience, and nearly half wanted to leverage wearables to save time when shopping. Among the desired features in the report are wearables receiving event reminders while shopping, alerts to long lines in stores, touch-less or one-click payments and interactive maps. Wearable devices can carry physiological metrics and as well information about the preferences and needs of the person for companies to offer products or services. Furthermore wearable data could be used to predict the likelihood of other behaviours closer to personal physiological conditions of the users. This preliminary research allow us to start thinking about models to start widen out the scope and focus on environmental and social dynamic models emerging from these datasets.

References

Billinghurst M, Starner T (1999) Wearable devices: new ways to manage information. Computer 32(1):57–64

Calvo R, D'Mello S, et al (2010) Affect detection: An interdisciplinary review of models, methods, and their applications. Affective Computing, IEEE Transactions on 1(1):18–37

Citi L, Klerman EB, Brown EN, Barbieri R (2010) Point process heart rate variability assessment during sleep deprivation. In: Proceedings of Computing in Cardiology, Belfast, URL `http://web.cinc.org/2010/preprints/315.pdf`

Citi L, Bianchi MT, Klerman EB, Barbieri R (2011) Instantaneous monitoring of sleep fragmentation by point process heart rate variability and respiratory dynamics. In: Proceedings of the 33rd IEEE Engineering in Medicine and Biology Society Conference, EMBC, Boston, DOI 10.1109/IEMBS.2011.6091906

Citi L, Valenza G, Purdon P, Brown E, Barbieri R (2012) Monitoring heartbeat nonlinear dynamics during general anesthesia by using the instantaneous dominant Lyapunov exponent. In: Proceedings of the 34th IEEE Engineering in Medicine and Biology Society Conference, EMBC, San Diego, pp 3124–3127, DOI 10.1109/EMBC.2012.6346626

Ghaoui C (2005) Encyclopedia of human computer interaction. IGI Global

Hougaard P (1995) Frailty models for survival data. Lifetime data analysis 1(3):255–273

Inc G (2016) Gartner Says Worldwide Wearable Devices Sales to Grow 18.4 Percent in 2016. `https://www.gartner.com/newsroom/id/3198018`, [Online; accessed 17-June-2016]

Kim KH, Bang S, Kim S (2004) Emotion recognition system using short-term monitoring of physiological signals. Medical and biological engineering and computing 42(3):419–427

Malik M (1996) Heart rate variability. Annals of Noninvasive Electrocardiology 1(2):151–181, DOI 10.1111/j.1542-474X.1996.tb00275.x, URL http://dx.doi.org/10.1111/j.1542-474X.1996.tb00275.x

OConnor S (2015) Wearables at work: the new frontier of employee surveillance. http://www.ft.com/cms/s/2/d7eee768-0b65-11e5-994d-00144feabdc0.html, [Online; accessed 17-June-2016]

Pazzani MJ, Billsus D (2007) Content-based recommendation systems. In: The adaptive web, Springer, pp 325–341

Picard RW (1997) Affective computing, vol 252. MIT press Cambridge

PowerReviews (2000) Mobile, wearable tech and hyper-relevance: Transforming consumer behavior and retailer opportunities. Tech. rep., URL http://pages.powerreviews.com/rs/powerreviewsinc/images/USWearableTech.pdf

Sismeiro C, Bucklin R (2004) Modeling purchase behavior at an e-commerce web site: A task-completion approach. Journal of marketing research 41(3):306–323

Valenza G, Citi L, Lanata A, Scilingo EP, Barbieri R (2013) A nonlinear heartbeat dynamics model approach for personalized emotion recognition. In: Proceedings of the 35th IEEE Engineering in Medicine and Biology Society Conference, EMBC, IEEE, pp 2579–2582, DOI 10.1109/EMBC.2013.6610067

Valenza G, Akeju O, Pavone KJ, Citi L, Hartnack KE, Sampson A, Purdon PL, Brown EN, Barbieri R (2014a) Instantaneous monitoring of heart beat dynamics during anesthesia and sedation. Journal of Computational Surgery 3(1):1–18, DOI 10.1186/s40244-014-0013-2

Valenza G, Citi L, Lanatá A, Scilingo EP, Barbieri R (2014b) Revealing real-time emotional responses: a personalized assessment based on heartbeat dynamics. Scientific Reports 4, DOI 10.1038/srep04998

Valenza G, Citi L, Gentili C, Lanata A, Scilingo EP, Barbieri R (2015) Characterization of depressive states in bipolar patients using wearable textile technology and instantaneous heart rate variability assessment. IEEE Journal of Biomedical and Health Informatics 19(1):263–274, DOI 10.1109/JBHI.2014.2307584