VILNIAUS UNIVERSITETAS

MATEMATIKOS IR INFORMATIKOS FAKULTETAS

Magistro baigiamasis darbas

Širdies ritmo duomenų analizė funkciniais metodais

Heart Rate Data Analysis Using Functional Methods

Mantas Zabarauskas

VILNIUS 2020

MATEMATIKOS IR INFORMATIKOS FAKULTETAS
STATISTINĖS ANALIZĖS KATEDRA

Darbo vadovas Prof. habil. dr Alfredas Račkauskas ________________

Darbas apgintas 2020-01-13

Registravimo NR. ________________

Contents

1 Abstract

Heart Rate Data Analysis Using Functional Methods

Abstract

Heart rate is one of the most informative data point that we can collect non-invasably. Heart rate reflects anxiety, stress, sleep quality, physical exercise, illness, ingestion of drugs and many other events that we might experience on a day-to-day basis. In this master thesis a use for heart rate data collected by smart watches is suggested - an online algorithm that can detect an abnormal heart rate behavior and raise an alert to the wearable device user. Creation of an effective online detection algorithm could be implemented into wearable devices and be used as a preventative measure to combat the rise of the heart diseases. The thesis exploits functional data analysis to smooth the point-wise measurements into functions for each 24 hour period. Using additional data of active calories burned, the heart rate data set is grouped into specific activity level clusters using functional expectation maximization algorithm. The results of clustering are cleaned of outliers using functional outlier detection by depth measures. Then the deepest functions are found for each group and L^2-distances are calculated between the deepest function and all the rest in the group. The 99.9% percentile value of each group is set as the threshold for alerting. The testing data set is then taken one day at a time, assigned to a particular group using the methodology described, then the L^2-distance between the day considered and its group's deepest function is calculated. If the calculated value exceeds 99.9% percentile value the day is registered as an outlier if not - it is added to the group and the statistics are recalculated before the end of iteration. The suggested algorithm reached the accuracy of 91.67% with the testing data set.

Key words: functional data analysis, heart rate, online change detection, functional depth.

Širdies ritmo duomenų analizė funkciniais metodais

Santrauka

Širdies ritmas yra vienas iš informatyviausių duomenų, kuriuos mes galime surinkti ne invaziniu būdu. Širdies ritmas atspindi nerimą, stresą, miego kokybę, fizinį krūvį, ligas, vaistų poveikį ir dar daug kitų reiškinių, kuriuos mes patiriame kiekvieną dieną. Šiame magistro baigiamajame darbe yra siūlomas naujas panaudojimo būdas širdies ritmo duomenim surinktiem išmaniųjų laikrodžių - realiu laiku (angl. *online*) veikiantis algoritmas, kuris identifikuoja anomalų širdies ritmą ir apie tai įspėja išmaniojo įrenginio naudotoją. Efektyvus realaus laiko algoritmas galėtų būti implementuotas į išmaniuosius įrenginius ir tarnauti kaip prevencinė priemonė kovoje su širdies ligomis. Baigiamajame darbe naudojama funkcinė duomenų analizė suglodinti pataškiams širdies ritmo duomenims 24 valandų intervaluose. Naudojantis papildomais aktyvių sudegintų kalorijų duomenimis, širdies ritmo duomenys yra sugrupuojami į grupes su specifiniu aktyvumo lygiu, pasitelkiant funckinį tikėtinumo maksimizavimo algoritmą. Iš gautų grupių yra pašalinamos visos išskirtys naudojantis funkciniais gylio matavimais, tada giliausia funkcija yra randama kiekvienai grupei. L^2-atstumai yra suskaičiuojami tarp grupės giliausios funkcijos ir visų likusių. 99.9% procentilio reikšmė kiekvienoje grupėje yra nustatoma kaip riba siūlomame algoritme. Testavimo duomenų rinkinys yra imamas po vieną dieną iš eilės, toji diena yra priskiriama prie vienos iš suformuotų grupių pagal aktyvių sudegintų kalorijų duomenis, L^2-atstumas suskaičiuojamas tarp naujos dienos ir grupės giliausios funkcijos. Jei apskaičiuota reikšmė yra didesnė nei 99.9% procentilio reiškmė - diena yra išskirtis, jei ne - diena yra pridedama prie grupės ir statistikos perskaičiuojamos. Baigiamajame darbe pasiūlytas algoritmas pasiekia 91.67% tikslumą testavimo duomenų rinkinyje.

Raktiniai žodžiai: funkcinė duomenų analizė, širdies ritmas, pokyčio nustatymas realiu laiku, funkcinis gylis.

2 Introduction

Heart rate is one of the most informative data point that we can collect non-invasably. Heart rate reflects anxiety, stress, sleep quality, physical exercise, illness, ingestion of drugs and many more other event that we might experience on a day-to-day basis. The beforementioned reasons might be the main contributers, which determined the fact that more and more smart wearable devices (usually smart watches) available on the market now come with the heart rate monitoring capabilities. The rise in the availability of the functionality means that now it is not only the athletes who are voluntarily collecting their heart rate data, but everyone has this choice as well. In addition to the availability of functionality, the wearable smart device changed how we collect heart rate data as well. Now instead of collecting the data only for the period of interest, say for a duration of a physical exercise, we collect the data continuously with only the minor pauses for the reacharging of the said device. With this collection of the heart rate data comes an issue - the use of it. With the exceptions of high and low heart rate detection, providing a point-in-time value of the measurement and detection of atrial fibrilations, there has not been any significant uses of the data so far. In this master thesis another use for the data is suggested - an online algorithm that can detect an abnormal heart rate behavior and raise an alert to the wearable device user. As of the defense of the thesis, no academic papers have been found, which would tackle the same problem or even the same data as is being tackled with this thesis. It is extremely important to note that research in this area is very important as creation of an effective online detection algorithm of abnormal heart rate behavior could be implemented into wearable devices and be used as a preventative measure to combat the rise of the heart diseases. In addition to this, additional input from the medical professionals would allow to adjust the algorithm into the version that could track the progression of already diagnosed illnesses.

The master thesis exploits functional data analysis (hereinafter FDA) to smooth the point-wise measurements into functions for each 24 hour period, which minimizes the issue of measurement error. In addition to this, the abstraction allows FDA to make use of additional information from the functions, contrary to classical multivariate statistical techniques. The suggested approach constitutes of the clustering of the additional data set containing smoothed active calories burned using functional expectation maximization algorithm to get four interpretable groups of dates. The results of clustering are then used to split the heart rate data set into the same groups, which are cleaned of outliers using functional outlier detection by depth measures. Then the deepest functions are found for each group using random projections and Tukey's method and L^2-distances are calculated between the deepest function and all the rest in the group. The 99.9% percentile value of each group is set as the threshold for alerting. The testing data set is then taken one day at a time, assigned to a particular group using the same methodology as described, then the L^2-distance between the day considered and its group's deepest function is calculated. If the calculated value exceeds 99.9% percentile value the day is registered as an outlier if not - it is added to the group and the statistics are recalculated before the end of iteration. The suggested algorithm reached the accuracy of 91.67% with the testing data set if an offline method for an outlier detection is considered as a ground truth.

The paper is structured as follows: section 3 covers the research applying functional data analysis to the data collected from the heart muscle. Section 4 explains the data used for the suggested algorithm. Section 5 covers the step by step methodology. In section 6 the results of the analysis are presented and the errors are discussed. Section 7 provides a summary of the analysis, the benefits of the suggested approach, their importance and future development of the research.

3 Literature review

Academic literature surrounding the application of functional analysis for heart rate data is extremely sparse. The vast majority of the publications of FDA application for medical data are to do with either the analysis of electrocardiogram or functional magnetic resonance imaging data. This section will cover general papers applying functional data analysis approaches to data collected about the heart muscle.

One work that is marginally similar and is applying FDA methods to heart rate data is Matabuena et al. (2019). In the paper, the authors use functional data analysis to create a new method to predict maximum heart rate (MHR). According to the authors, MHR is widely used in the prescription and monitoring of exercise intensity, and also as a criterion for the termination of sub-maximal aerobic fitness tests in clinical populations. Matabuena et al. (2019) argue that the traditional way of predicting the MHR from the age-based formula, usually 220 - age, leads to very high error rates, which could mean that the training prescribed to an individual is inaccurate. In the paper, the authors use data gathered every 5 seconds during a 6-minute interval of low intensity sub-maximal exercise test. Using the data gathered, the authors fit both linear regression and functional regression see Table 1.

Table 1: Matabuena et al. (2019) regression models.

Regression type	Model (MHR=)	R^2	RMSE
Linear regression	$209.9 - 0.77 \times A$	0.238	7.044
Functional regression	$s(HR_{[0,6]} + s(dHR_{[0,6]}) + s(HR^{max}_{[0,6]}) + s(A)$	0.737	4.143

For the functional regression $HR_{[0,6]}$ and $dHR_{[0,6]}$ are the heart rate (hereinafter HR) and the derivative of HR for the first six minutes of the treadmill test, respectively. The function $s(\cdots)$ denotes the additive effect over the variable. Variable A in both cases is age. As can be seen from the above table the difference between the models is vast when comparing $R^2$and the lowest RMSE values.

According to the authors, this paper is the first one to apply FDA approaches to model and predict the maximum heart rate value of an athlete and their functional data model reduced the predictive error by more than 50% compared to current models within the literature.

The second paper applying FDA approach to heart rate data is Ratcliffe et al. (2002).

In their work, the authors use functional regression to model longitudinal data where the number of measurements on the functional covariate is much greater than the number of subject in the study. The aim of the study was to determine if the foetal heart rate responses to the stimulus are predictive of birth outcomes, and of the infant's development at 18 months of age. They have applied the technique to periodically stimulated foetal heart rates. According to the authors, the heart rate tracings were used as a predictor in child's psychomotor development. The data was gathered in late pregnancy of 73 women, the measurements were taken using Corometrics cardiotocograph. For each subject, the data was recorded over 19-minute period, 14 days or less before birth. The above-mentioned stimulus was introduced via a medical device placed on the mother's abdomen and given every minute for a total of 19 stimuli. According to Ratcliffe et al. (2002), this stimulus initially results in an increase in the foetal heart rate, with the response decreasing with repetition of the stimulus.

The authors chose to model the Psychomotor Development Index (hereinafter PDI), which helps to quantify the new-born's development. The higher the PDI score – the better the development. For the model itself, the authors apply the modified version of Ramsay and Silverman (Functional Data Analysis, 1997) method for functional linear regression:

$$y_i = z_i^T\alpha + \sum_{s=1}^{q}\gamma_s \int x_{i,s}(t)\beta(t)dt + \varepsilon_i, \qquad i = 1, \ldots, n$$

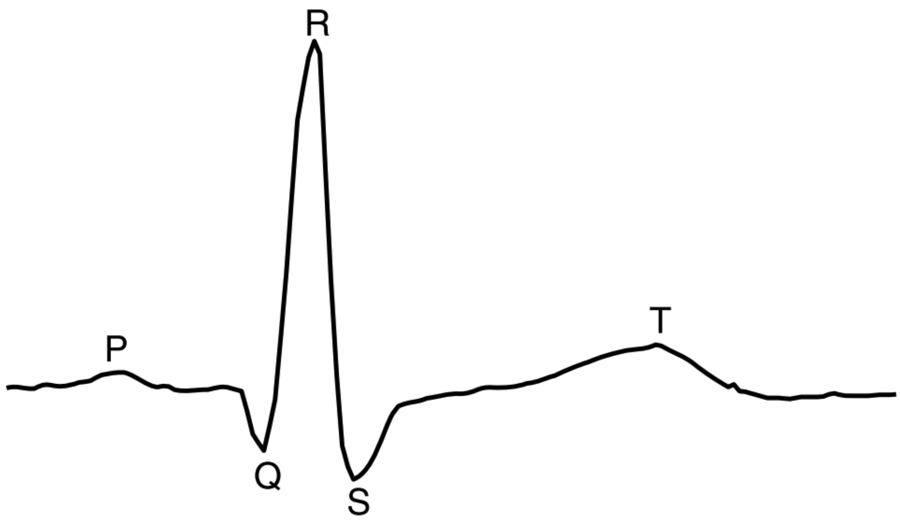
where y_i is the response from the ith subject, $\alpha$is the vector of parameters for the covariates, z_i is the vector of the covariates for the ith subject, γ_s is the parameter for the sth stimulus, $x_{i,s}(t)$ is the functional covariate for the ith subject, measured at time t within the sth stimulus, $\beta(t)$ is the functional parameter for the times and $\varepsilon_i \sim NID(0, \sigma_\varepsilon^2)$ is the error associated with subject i.

For comparison, the authors have fitted a standard linear regression model without the functional covariate. The mentioned model has achieved R^2equal to 7.10 percent. For the functional regression, the same parameter reached 68.77 percent, which is a significant improvement. The overall conclusion that the authors arrive to is that changes in foetal heart rate after stimulation, as well as infant's sex, are important factors in prediction of child's development at the age of 18 months.

Another work applying FDA approached to heart data is Zhou et al. (2009). In the paper, the authors are using functional data analysis to extract a common shape from

ECG[1]measurement (see Figure 1) for each subject and then models the deviation of each curve in sequence from that reference curve as a four-dimensional vector. According to the authors, the representation can be used to distinguish differences between beats or to model shape changes in a subject's T-wave over time. This has important implications as the T-wave of an ECG represents the ventricular repolarization that is critical in restoration of the hear muscle to a pre-heart-beat state prior to the next beat. Alterations to this part of the ECG measurements could mean various heart conditions.

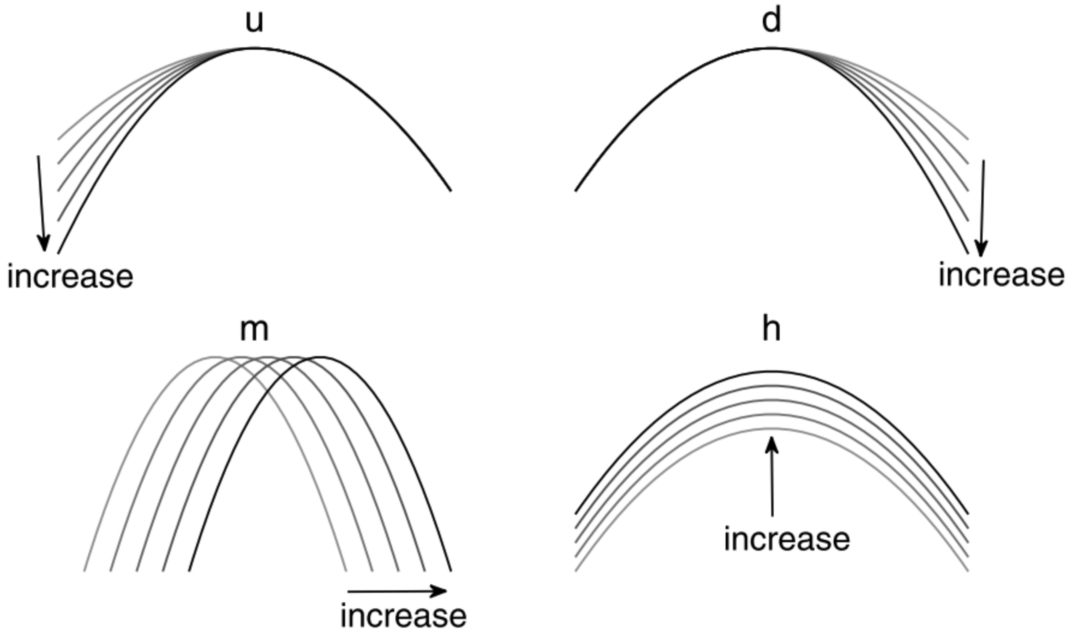
Figure 1: Example of different waves within an ECG recording. Zhou et al. (2009)



Zhou et al. (2009) motivated by Izem et al. (2005) create a model that decomposes the data into the above mentioned reference curve and four modes of variation that are of interest to the authors: uphill slope, downhill slope, horizontal location and vertical shift (see Figure 2).

[1]Electrocardiogram - a recording – a graph of voltage versus time – of the electrical activity of the heart using electrodes placed on the skin (Leonard (2010))

Figure 2: Four modes of variation of T-waves: uphill slope change (u), downhill slope change (d), horizontal location (m) and vertical shift (h). Zhou et al. (2009)



According to the authors, the reference curve represents the common shape of the T-wave; an element in the data matrix X is modelled as:

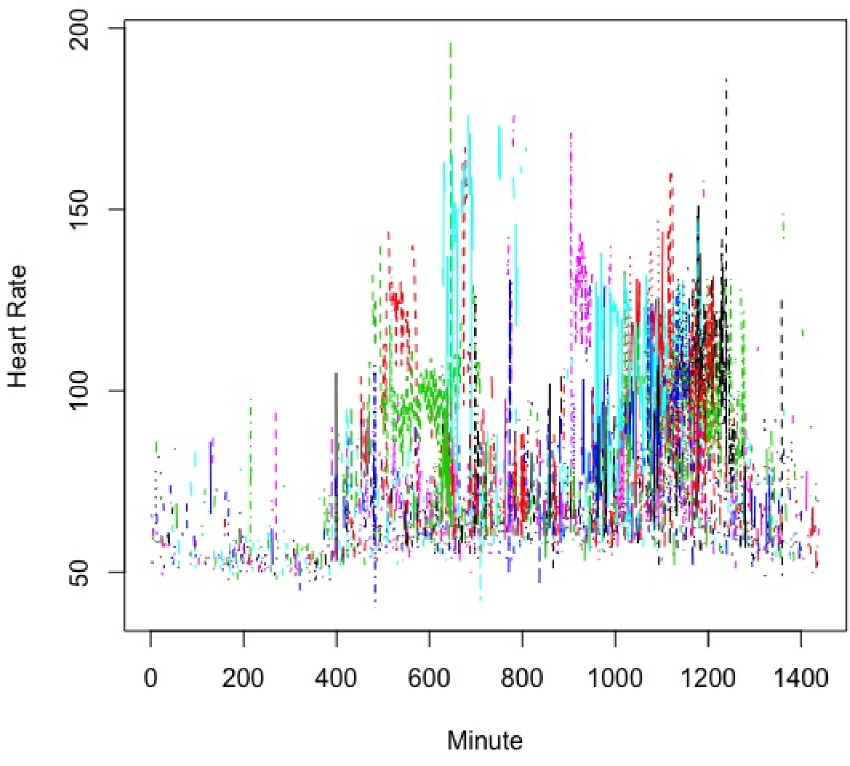
$$x_{it_j} = x_i(t_j) = \begin{cases} \sqrt{u_i d_i} K(u_i(t_j - m_i) + h_i + \varepsilon_i(t_j), & t_j \leq m_i \\ \sqrt{u_i d_i} K(d_i(t_j - m_i) + h_i + \varepsilon_i(t_j), & t_j > m_i \end{cases}$$

where $i = 1, \dots, I$, $j = 1, \dots, J$, K is the reference curve, the four parameters u_i, d_i, m_i, h_i are the four modes of variation as described above. As the reference curve the authors decide to take the Fréchet mean of the curves in the sample. According to Zhou et al. (2009), the Fréchet mean is a generalized mean on the nonlinear manifold, therefore, it represents the common shape of the curves. According to the authors, the model accounts for the entire T-wave morphology, making the estimation robust to the marking of T-wave boundaries. Applications such as classification of beats were illustrated by the authors showing that decomposition allows for a better detection of arrhythmia in subjects. Different modes of variation of T-waves allows to see correlation between the specific modes and length of QT (time from start of Q-wave until the end of T-wave – indicator of heart arrhythmia), which, according to the authors, is difficult to measure using other techniques. According to Zhou et al. (2009), the representation can be used to distinguish differences between beats or to model shape changes in a subject's T-wave over time. This model provides physically interpretable parameters characterizing T-wave shape, and is robust to the determination of the endpoint of the T-wave. Thus, this dimension reduction methodology offers the strong potential for definition of more robust and more informative biomarkers of cardiac abnormalities than the QT (or QT corrected) interval in current use.

4 Data

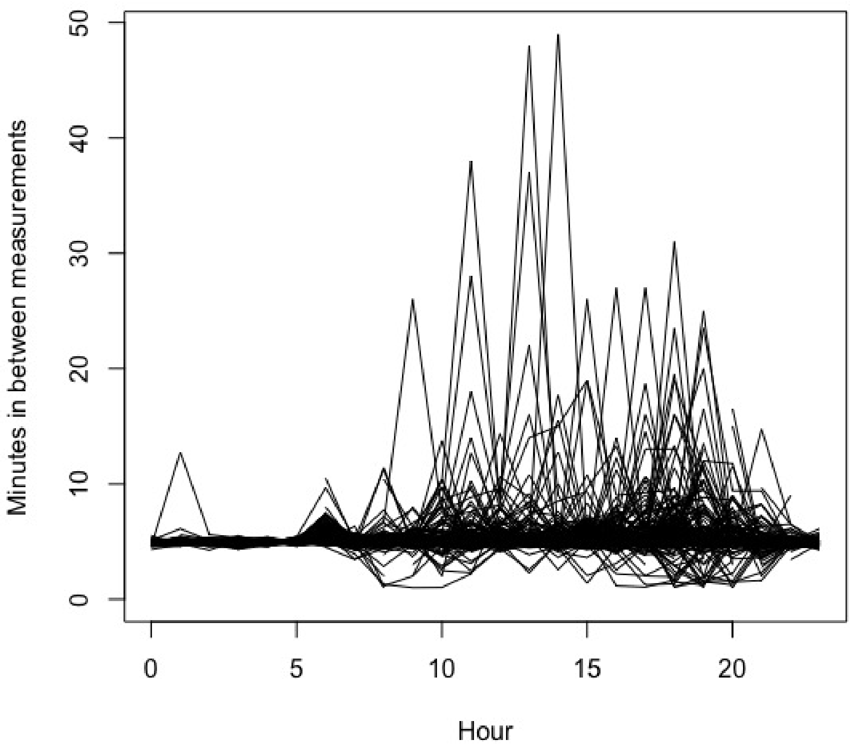
The training data set used in this paper is collected from 2018-04-08 until 2018-10-11 - this constitutes 187 days for the training data set. The testing data set was collected from 2019-06-01 until 2018-08-13 - 74 days. The data being considered is the heart rate and active calories burned data that was collected using Apple Watch Series 3 and Series 4 devices during the beforementioned periods. The heart rate data was taken 259 times per day on average, that is every 5 minutes and 33 seconds on average (see Figure 3).

Figure 3: Raw heart rate measurement data.



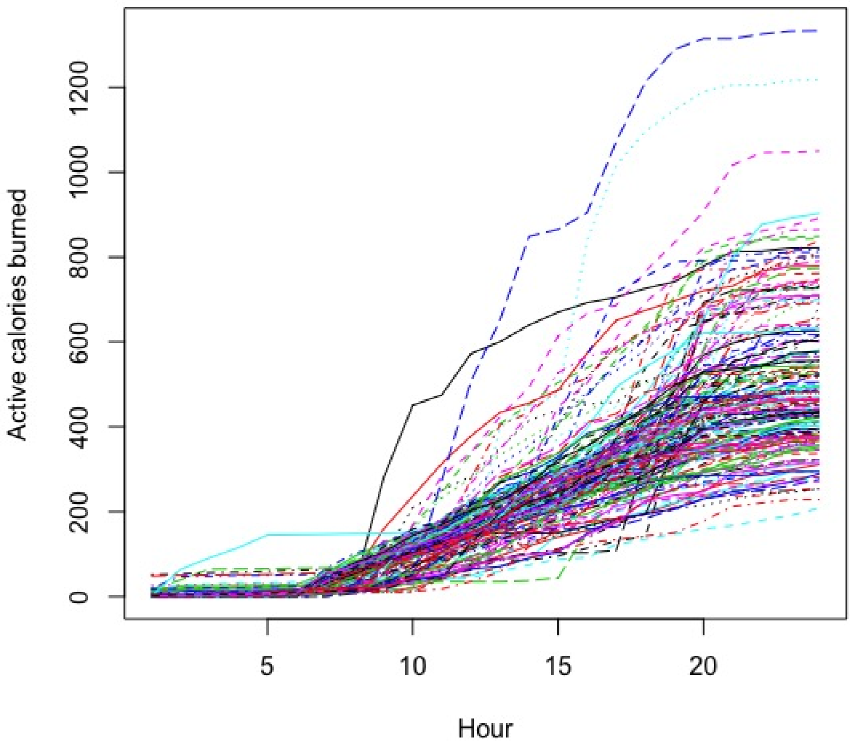
It is important to mention that these are averages as the data is being collected at semi-random intervals. From observations, the measurements are taken at a more stable time intervals during the night hours and at less stable intervals during the day hours (see Figure 4). The interval would become even shorter if the algorithm determines higher activity levels such as workouts.

Figure 4: Average time in minutes in between measurements of the heart rate.



The data for the active calories burned is reported for each hour and the value is equal to all active calories burned up until that time point, meaning that the data is monotonically increasing during time (see Figure 5). According to the manufacturer (see Apple Support (2019)), the active calories burned data is calculated by intaking the data about user, such as height, weight, gender and age. Every full minute of movement that equals or exceeds the intensity of a brisk walk counts toward the user's active calories burned. The determination of whether or not activity exceeds the intensity of a brisk walk is done through the gyroscopes and accelerometers in the smart watch (the manufacturer asks to make sure that the arms swing freely when doing any exercise). In addition to this, the smart watch uses the data collected from the heart rate sensor and GPS to adjust the active calories burned data so it reflects intesity better.

Figure 5: Active calories burned data.



5 Methodology

Functional data analysis considers each observation as a function that can be evaluated on any interval chosen. In the case of this thesis, the curves are obtained using the non-parametric methodology - the kernel smoothing. This is applicable for both the heart rate and active calories burned data. According to Ferraty and Vieu (2006), the non parametric smoothing of the functional data is given by the smoothing matrix $S = (s_{ij})$. For the chosen method of Nadaraya-Watson estimator, the matrix takes shape:

$$s_j(t_i) = \frac{K(\frac{t_i - t_j}{h})}{\sum_{k=1}^{m} K(\frac{t_i - t_k}{h})}$$

where $K(\cdots)$ is the kernel function. The kernel being used for the analysis is Gaussian:

$$K(t) = exp\left\{-\frac{t^2}{2b^2}\right\}$$

Here b is the bandwidth. The Gaussian kernel is chosen as it is the default smoothing kernel. In addition, the choice of the kernel has a limited impact compared to the choice of bandwidth parameter. The kernel smoothing implementation is provided by Febrero-Bande and Oviedo de la Fuente (2012) in the R software package fda.usc. Said implementation allows to use the generalized cross-validation criterion to pick the optimal bandwidth:

$$Generalized\, cross - validation \ : \ GCV(\nu) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i^{\nu})^2 w_i \Xi(\nu)$$

where $\Xi(\nu)$ denotes the type of penalizing function, and w_i is the weight at point t_i. The implementation calculates 101 GCV values for each curve and chooses the optimal bandwidth value, which corresponds to the lowest GCV value among calculated. The chosen penalization function is the generalized cross-validation:

$$Generalized\, cross - validation \ : \ GCV : \Xi(\nu) = (1 - tr(S_\nu)n^{-1})^{-2}$$

The algorithm chosen for the clustering of the active calories burned data is FunFEM (functional expectation-maximization algorithm) first introduced by Bouveyron and Jacques (2014) and further developed by Bouveyron,Come and Jacques (2015). According to the authors, the algorithm applies 12 different discriminative functional mixture models (hereinafter DFM), which model the data into a single discriminative subspace. This subspace subsequently allows an insightful visualization of the clustered data and eases the comparison of systems regarding the identified patters. A family of 12 models is also proposed by relaxing or constraining the main DFM model, allowing it to handle a wide range of situations. The model selection is performed by the Bayesian information criterion. As this thesis is not analyzing different methods of functional data clustering, the authors elects not to explain the algorithm in-depth, for more details see Bouveyron,Come and Jacques (2015).

The heart rate data is split into the same clusters as active calories burned data according to the date. The new data sets are then checked for outliers and all found outliers are removed. This is done to form "clean" data sets, which would contain no outliers, to have a basis for each group to compare the observations from the test data set to, i.e. a known normal behavior of the heart muscle in each respective group. The algorithm chosen to identify the outliers in the offline detection scenario is implemented in the R software package fda.usc. It is outlier detection in functional data by depth measures, where function weights the data according to depth. The chosen method is described in Febrero-Bande et al. (2008).

According to the authors of the paper, depth and outlyingness are inverse notions, so that if an outlier is in the dataset, the corresponding curve will have a significantly lower depth when compared to the rest of the data set. Therefore, a way to detect the presence of functional outliers is to look for curves with lower depths. The specific method is chosen to align the methodology throughout the analysis, i.e. match to the deepest function methodology.

The cleaned data is then passed to deepest function detection part. The deepest function is considered as a representative for the whole group. The algorithm chosen to identify the deepest function was suggested by Cuevas et al. (2007) and is implemented in the above mentioned R software package fda.usc. The algorithm returns the vectors of sample random projection, and random functional depth values. The random projection depth described in Cuevas et al. (2007) is based on the average univariate depth of one-dimensional projections of the data. According to the authors, the projections are taken randomly as a sample of standard normal d-dimensional random variables, where d stands for the dimensionality of the data. The functional data are projected into the real line in random directions as for the random projection depths. Afterwards, an approximation of the half-space (Tukey) depth based on this limited number of univariate projections is assessed. For more details, see Cuevas et al. (2007).

The last step before the online detection can be tested against the test data set, the L^2-distance is calculated between the deepest function for the group and all the rest functions that fall into the same group. The distance between curves measure chosen for online outlier detection method is the $L^2(X)$ space. $L^2(X)$ denotes the set of square-integrable functions on X. The distance measure was chosen for its easy interpretation, i.e. the L^2-distance can be interpreted as function's energy. In addition, L^2 has been previously used in medicine. For example, von Dobeln,Astrand and Bergstrom (1967) use the L^2-distance in their analysis of age and other factors related to maximum oxygen uptake. The L^2-distance between functions f and g is calculated as follows:

$$d(f,\ g) = \sqrt{\int_X |f - g|^2 d\mu}$$

The 90%, 95%, 97.5% and 99.9% percentile values are calculated for each groups' L^2-distance values. This is used as the thresholds to raise an alarm to a wearable device user that their heart rate shows deviation from the group's norm.

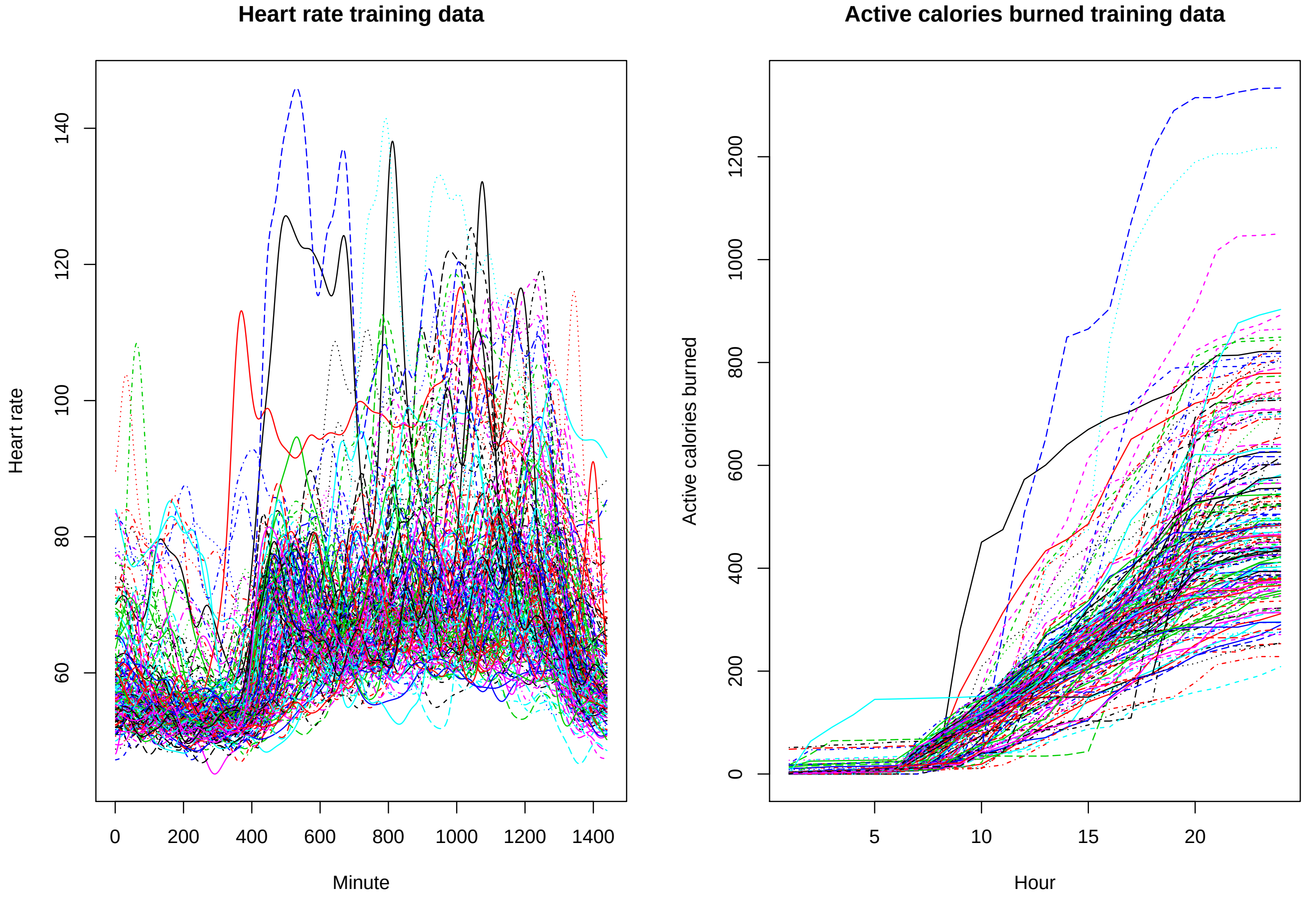
After the groups are formed, cleaned, the deepest functions for each group are identified and the L^2-distances are calculated the online detection algorithm can be setup. Taking the test data set one day at a time, it is assigned to one of the groups identified in the clustering part. This is done using the same methodology described above. Knowing the group that the new day belongs to, the L^2-distance is then calculated between the new function and the group's deepest function. This value is compared to the 90%, 95%, 97.5% and 99.9% percentile values of the relevant group, if the threshold is breached the day is considered as an outlier, if not - the day is added to the group, the deepest function and percentile value are recalculated.

After the online detection algorithm runs through the test data set, conventional outlier detection described above is performed on the test data set split into relevant groups. The performance of the algorithm is evaluated based on the Type I and Type II errors when comparing the online detection algorithm versus the conventional offline method. The final version of the online algorithm uses the threshold value which provides the best results when compared to the offline method.

6 Results

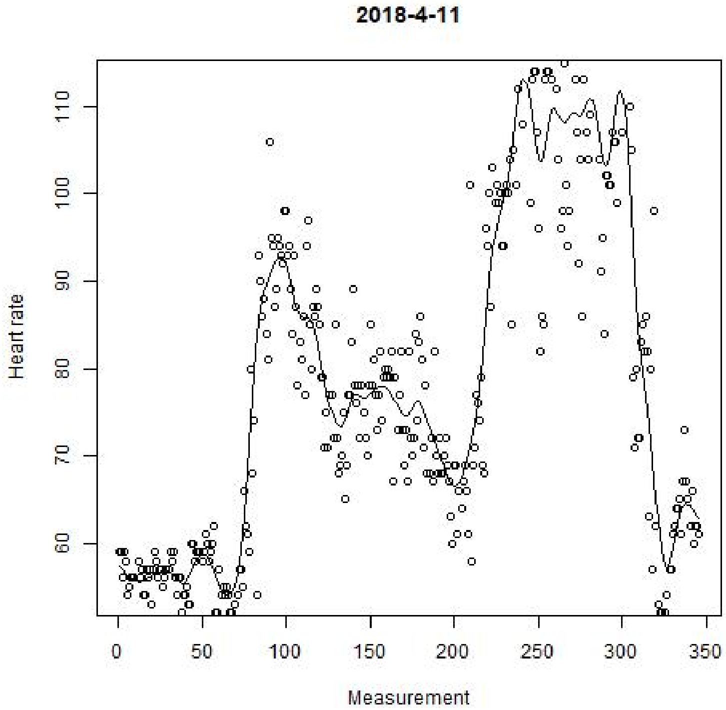
The training sets of hear rate data (see Figure 6) and the active calories burned data are smoothed to X axis of 0 to 1440 (measurement at every minute) and 0 to 24 (measurement at every hour) respectively.

Figure 6: Smoothed heart rate and active calories burned data



An example of a raw data and smoothed curve for one observation is provided in Figure 7. As can be seen the curve follows the trend quite well. The mean squared error for particular example is equal to 69.52, whilst the mean absolute percentage error is 6.4%. These values are quite similar for all the days being considered.

Figure 7: Raw heart rate measurement data and the corresponding smoothed curve.



The smoothed out active calories burned data is passed to the functional EM clustering algorithm with an imposed limitation of forming from 2 to 4 clusters. The limitation is imposed to have interpretable clusters. The optimal cluster number, the same as the type of model defined in the Methodology section, is chosen by the BIC value. The model that fit the analyzed data can be seen in Table 2.

Table 2: Discriminative functional mixture model chosen

Model	$\sum_k$	β_k	No of variance parameters
$DFM_{[\alpha_{kj}\beta_k]}$	Diagonal	Free	$(K-1)(p-K/2)+K^2$

In the case of the analyzed data set, the optimal number of clusters is 4, summary statistics for all four groups are provided in the table below.

Table 3: Summary statistics for the active calories burned data clusters

	Group 1	Group 2	Group 3	Group 4
No. of observations	70	27	39	51
Minimum	281.8	254.4	271.1	208.7
1st Quartile	437.1	380.4	510.7	318.1
Median	470.7	434.6	695.3	377.2
Mean	506.5	462.2	669.1	414.4
3rd Quartile	548.9	499.6	751.0	483.7
Maximum	864.7	849.2	1333.6	740.5

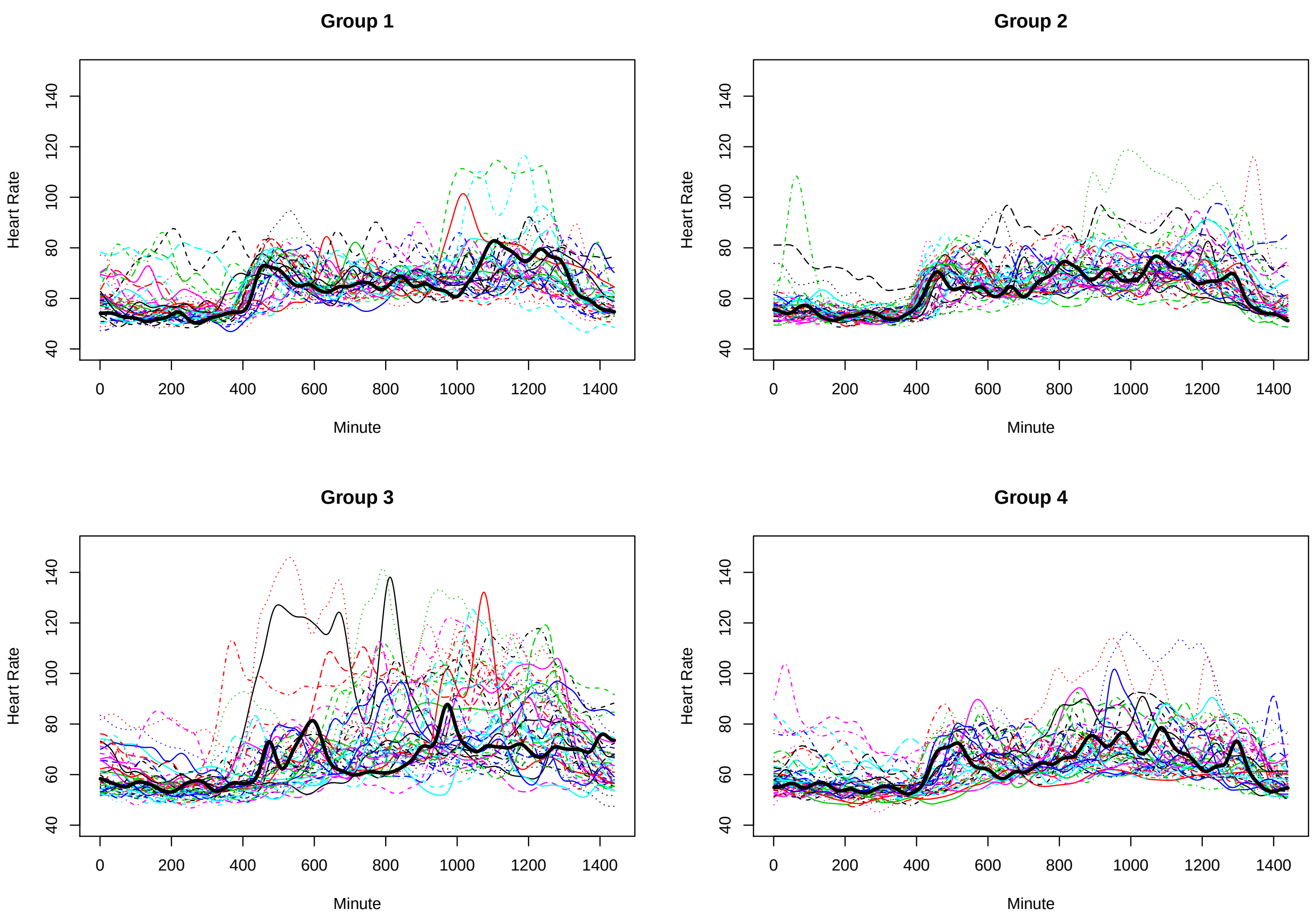
As can be seen from the Table 3, groups 1 and 2 are very similar to each other according to all summary statistics. However, because the number of clusters was determined through iteration and BIC, the similar groups are not joined together and are left as separate for the further analysis. Using the data obtained in the clustering of the active calories burned data, the heart rate data set is split up into the same groups according to the date. The outlier detection is performed for each data set separately and the outliers are removed. Numbers of outliers in each group are provided in the table below.

Table 4: Numbers of outliers found in each group of heart rate data

Group	No. of outliers
1	3
2	4
3	4
4	2

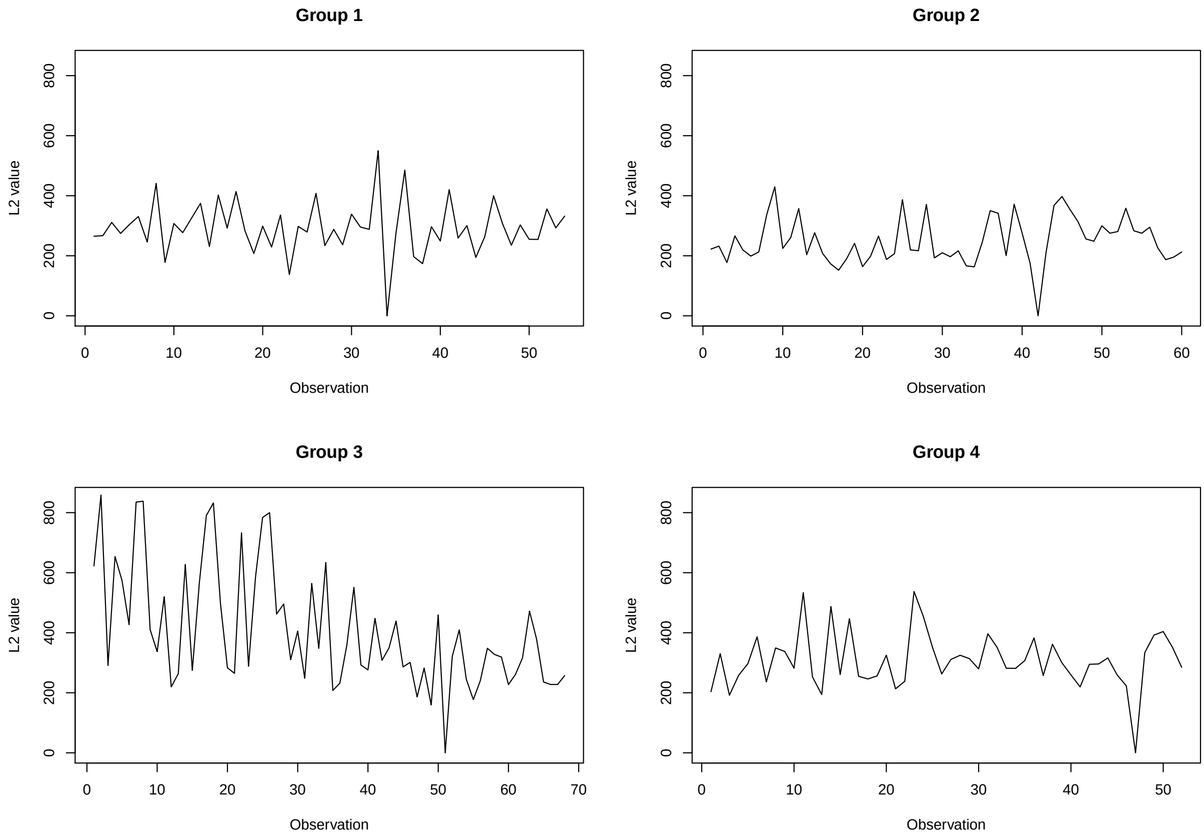
After the outliers have been removed, the deepest function for each group is found (see Figure 8). The thicker black lines in each graph represent the deepest function for each group.

Figure 8: Groups of hear rate data and their deepest functions



The L^2-distances are then calculated for each group (see Figure 9). The downward spikes to zero in each graph identify the location of the deepest function in each group.

Figure 9: L^2-distances for each group



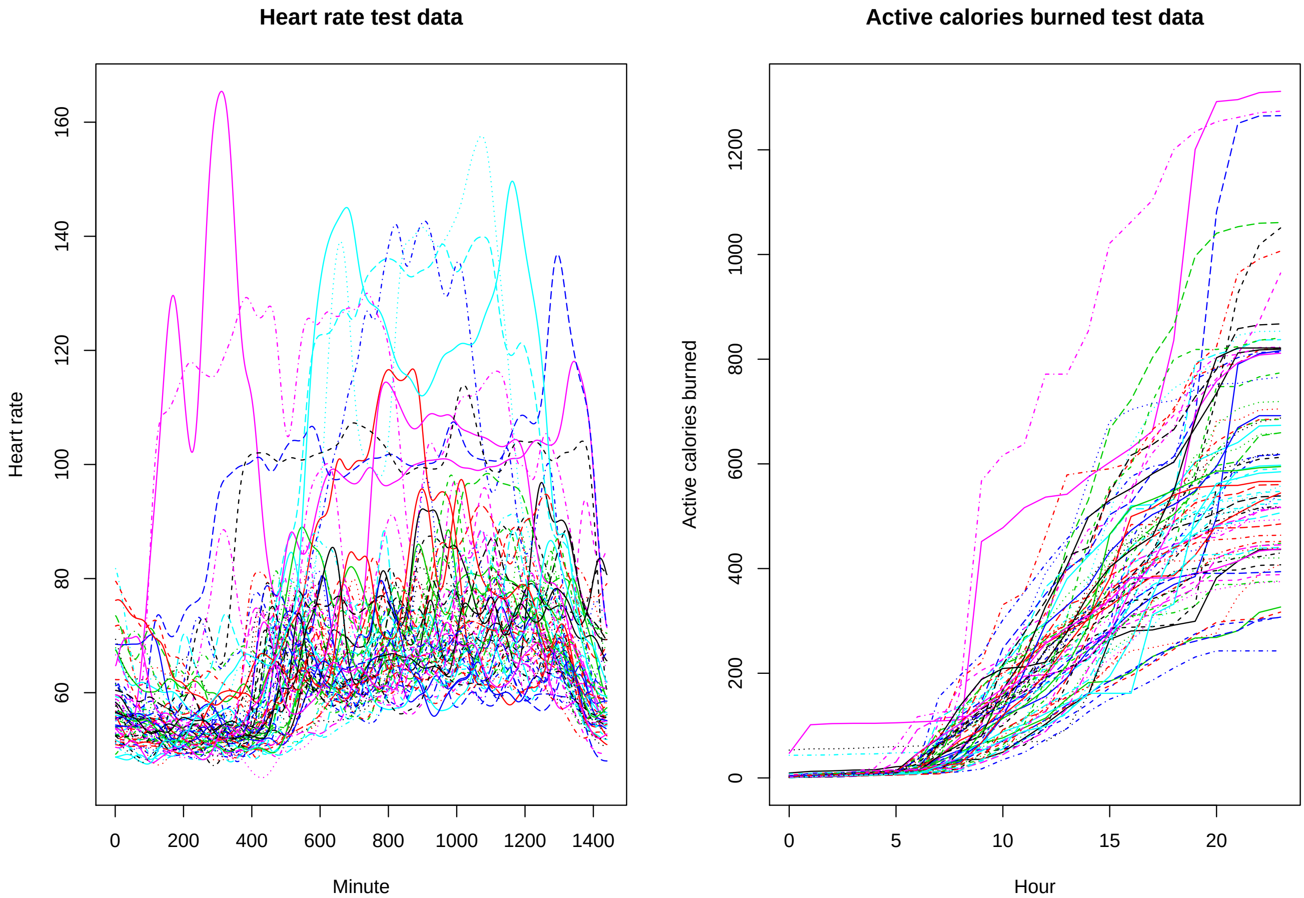
The threshold values for the online detection algorithm are then calculated at 90%, 95%, 97.5% and 99.9% percentile values (see Table 5).

Table 5: Threshold values for the online detection algorithm

	90%	95%	97.5%	99.9%
Group 1	406.71	428.55	471.88	546.47
Group 2	360.12	373.16	392.64	427.70
Group 3	753.21	822.58	836.15	857.34
Group 4	403.93	471.16	521.92	537.16

The test data set is prepared to be plugged in into the online detection algorithm by smoothing both heart rate and active calories burned data (see Figure 10). After obtaining the smoothed data sets, data is plugged in a day at a time into the online detection algorithm.

Figure 10: Smoothed test data sets



The online detection algorithm first assigns the observation it is handling to one of the four identified groups according to the active calories burned data for a specific day. This is done in the same manner that the initial groups were formed (see Methodology). After assigning the observation (t) to a relevant group, the algorithm then picks one of the four group based scenarios. Within the scenario, the algorithm calculates the L^2 metric value comparing the new observation (t) versus the deepest function within the group. Then the L^2 value is compared to four percentile values of the group - 90%, 95%, 97.5% and 99.9%. If the L^2 value is higher then the observation (t) is recorded as an outlier. If the value falls below the percentile then the new observation (t) is joined to the group and both the deepest function and percentile values are re-estimated again. The iteration is completed and the algorithm picks another observation (t+1).

The chosen approach has the advantage of having a zero lead time to outlier detection. Meaning that when an abnormal observation is present it will be detected as soon as that day goes into the algorithm and minimal amount of time passes between the taking of the measurement and outlier detection. The zero lead time of the algorithm has important implications to the use of the algorithm as it ensures that the crucial time between a medical event regarding the heart muscle occurring to the time that the wearable device user is notified of the possible event is minimal, hence, if needed, appropriate action can be taken without a delay.

The online detection algorithm finds 17 outliers with 90% threshold, 16 outliers with 95%, 14 with 97.5% and 13 with 99.9%. For comparison, conventional offline outlier detection algorithm is run for the whole set of the 74 day test data set. The offline method finds 12 outliers within the groups. If we assume that the conventional offline outlier detection method constitutes the ground truth, then he miss-match between the methods contains both Type I and Type II errors (see Table 6). From the table we can see that the number of Type I error decreases when we are increasing the threshold, whilst the Type II error is constant in all four versions, i.e. the algorithm always misses one abnormal observation identified by the offline method.

Table 6: Type I and Type II errors of the algorithm

90%		95%	
Type I error	Type II error	Type I error	Type II error
2019-6-1	2019-7-26	2019-6-7	2019-7-26
2019-6-7		2019-6-15	
2019-6-15		2019-7-1	
2019-7-1		2019-7-12	
2019-7-12		2019-7-24	
2019-7-24			
97.5%		**99.9%**	
Type I error	Type II error	Type I error	Type II error
2019-6-15	2019-7-26	2019-06-15	2019-07-26
2019-7-12		2019-07-24	
2019-07-24			

For the rest of the analysis the version with 99.9% percentile threshold value will be considered as it provides the highest accuracy. All three errors that occurred are investigated deeper. Figure 11 shows the function for the day that alerted as false positive and its group on the left and the relevant group's L^2 values with red dot as value for the false positive. From the left side plot the false positive function seems to be within the range of norm for the majority of the day, only starting to look as an outlier from 16:30. The failed function's close proximity to the others within the group is reflected in the right side plot as well. It shows the L^2 value when comparing the outlier with the deepest function of the group as a red dot. As can be seen from the figure, the red dot is just above the threshold with value of 553.73, whilst the threshold is equal to 537.16.

Figure 11: 2019-06-15 Type I error

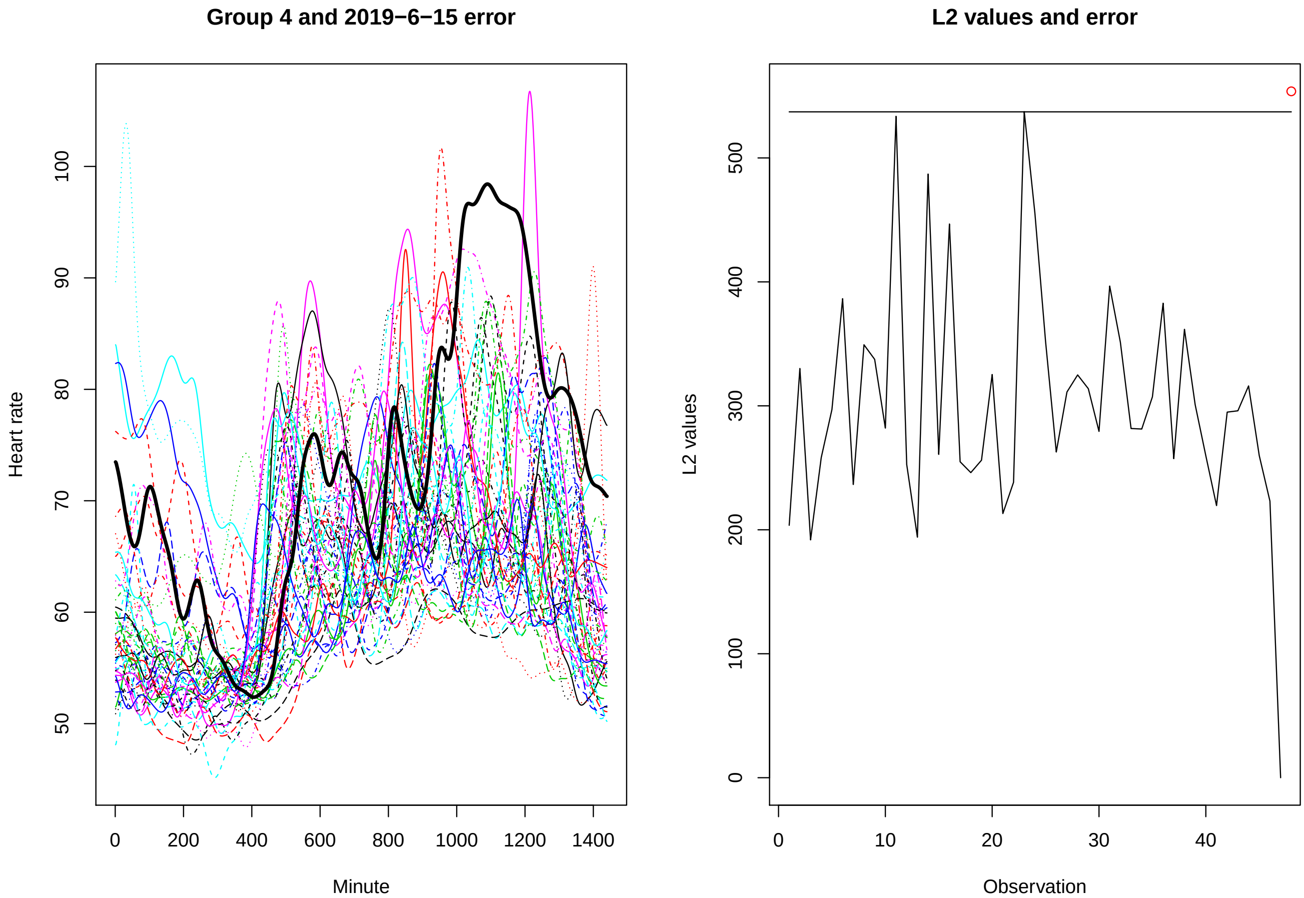
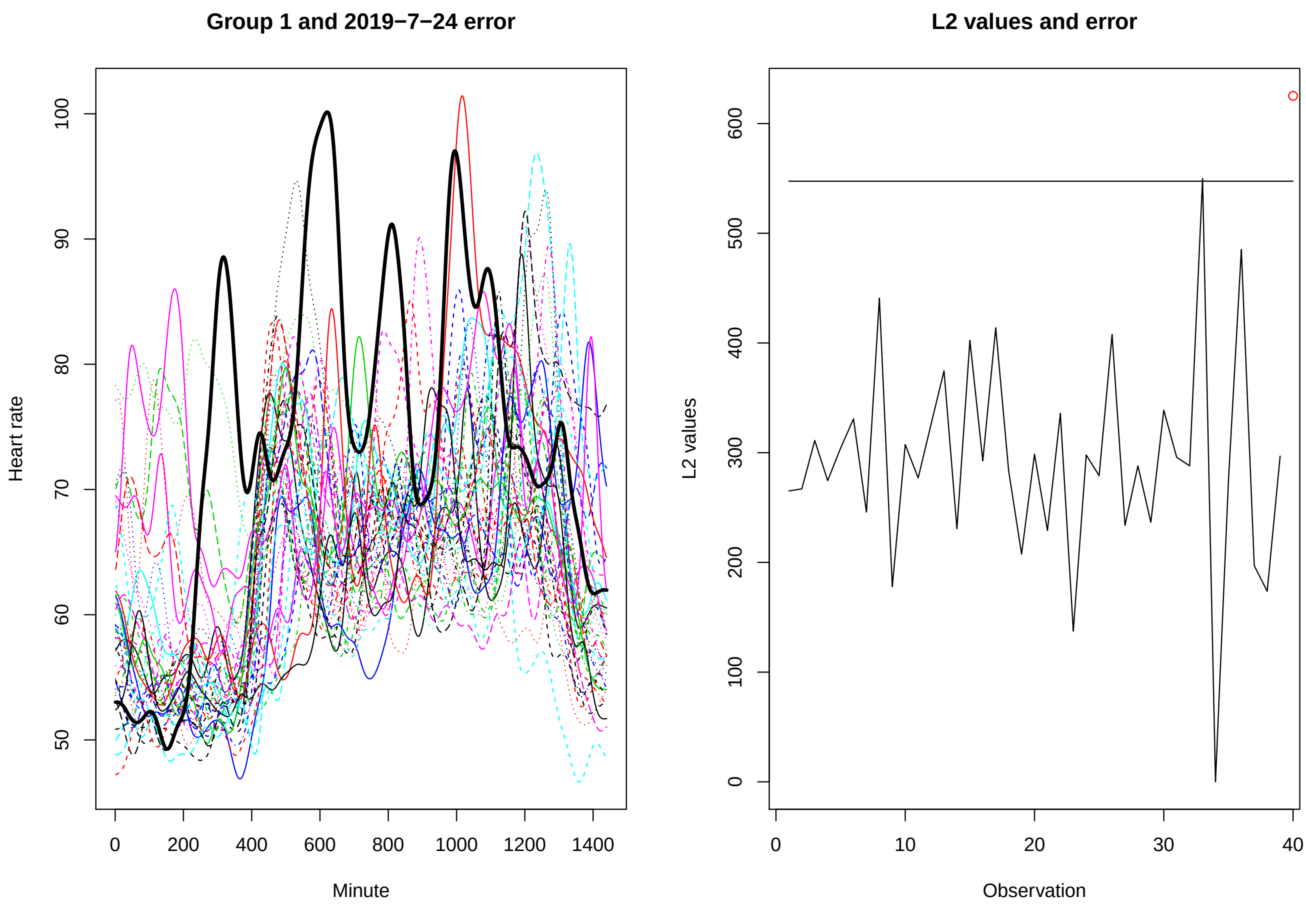


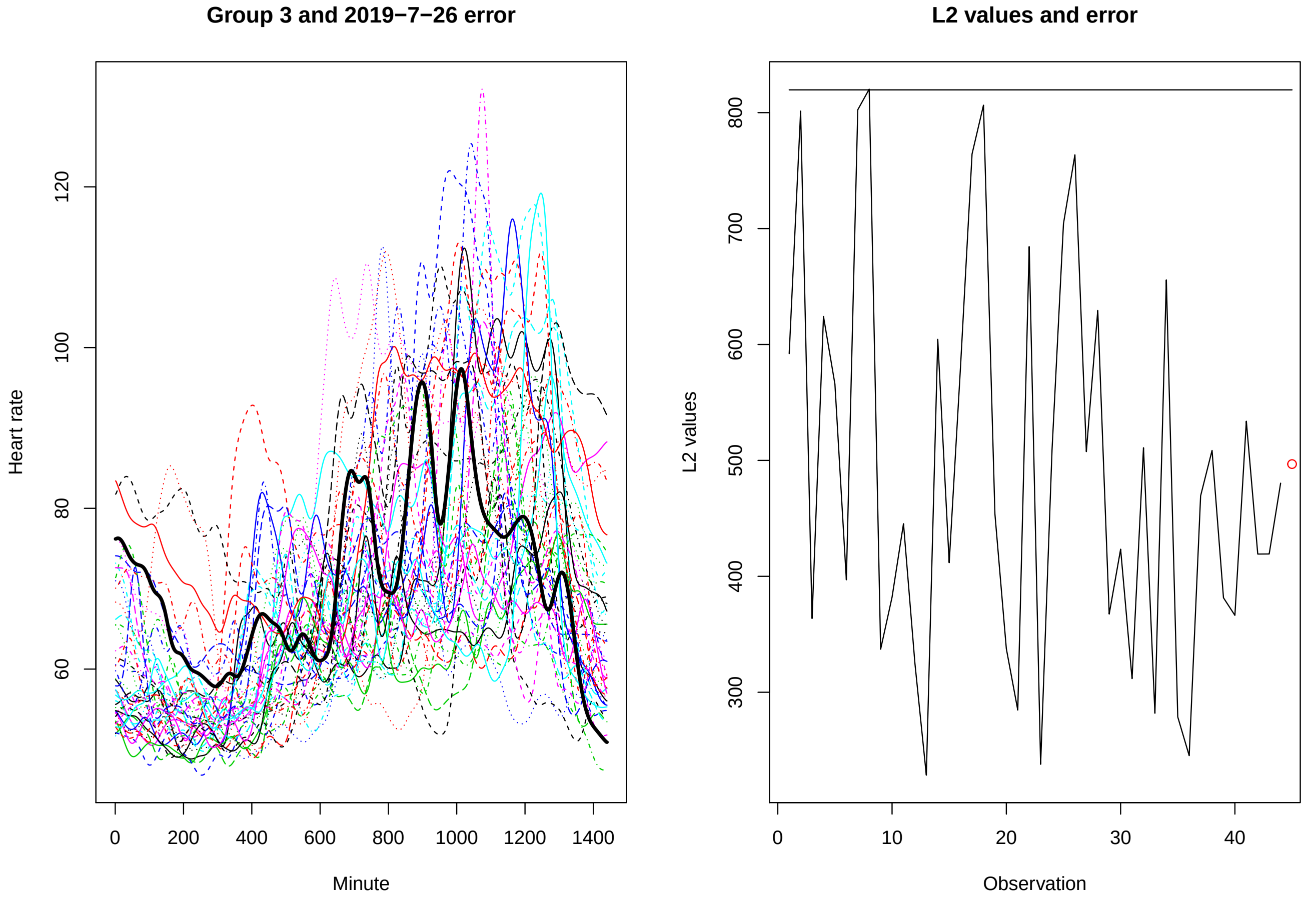
Figure 12 depicts the same figures as described above only for the false positive outlier of the day 2019-07-24. From the left side plot, the false positive outlier looks quite volatile, even though the rest of the group could not be taken as homogeneous. In this particular case, the plot of the right had side shows the outlier being further away from the rest of the group if we compared it to the results of the 2019-06-15 Type I error. The L^2 value for the outlier and the deepest function in the group is 625.26, whilst the threshold is set at 546.47.

Figure 12: 2019-07-24 Type I error



The worst of the three errors is depicted in the Figure 13. In this case, the offline outlier detection method identified this function as an outlier, whilst the online algorithm indicated the function to be normal. This is illustrated in the figure below. In both the left and right hand side plots, the function in question does seem to belong to its specified group. The only obvious region of interest is the night hours depicted in the left hand side plot - they start at a significantly higher point than the majority of the observations and then moves downwards as the hours pass. The L^2value for this function is equal to 496.85, whilst the threshold is set at 819.66.

Figure 13: 2019-07-26 Type II error



As the Type II error is more significant when considering the medical data, i.e. it is better to have a normal observation raised as an alert than to miss an abnormal observation, the error rate for the algorithm comes in at 8.33%. Overall, the result should be improved - with a goal of having zero Type II errors and minimizing the amount of Type I errors. However, this can only be achieve with the further analysis and research into the subject, including the medical professionals to help fine tune the algorithm, decide on definition of medical outlier versus a statistical one and interpretation of the data.

7 Summary

In this master thesis functional data analysis is used to construct an algorithm for the online detection of an abnormal heart rate behavior. The algorithm makes a reasonable assumption that there are at least four different clusters of days according to the activity (measured in the active calories burned) and uses the clusters to form bases of heart rate measurement values using training data. The test data then is assigned to a particular group according to activity levels and checked against the deepest function of the group by calculating the L^2-distance, if it breaches the 99.9% percentile value of the group the day is considered as abnormal. The suggested algorithm reached accuracy of 91.67% when compared to conventional offline methods. The research is planned to be continued further, ingesting all the data that is collected on a smart watch, using activity measures more extensively in creating better detection clusters and analysing the speed of change of the heart rate. In addition, the future research will include medical professionals to help determine the difference between a statistical and a medical outlier.

It is important to note that the biggest strength of the suggested algorithm is the zero lead time to detection, meaning that the algorithm is capable of detecting an abnormal heart rate behavior instantly after the data for a full day is gathered. This has a very important implication to the possible use of the algorithm as it minimizes the time period between a possible medical event and its detection, hence the user can seek medical attention as soon as possible, if that is needed.

References

[1] Apple Support. Get the most accurate measurements using your Apple Watch. *Apple Support*, 2019, URL:https://support.apple.com/en-us/HT207941.

[2] C. Bouveyron and J. Jacques. Adaptive Mixtures of Regressions: Improving Predictive Inference when Population has Changed, *Communications in Statistics - Simulation and Computation*, 2014, p.2570-2592, DOI: 10.1080/03610918.2012.758737

[3] C. Bouveyron, E. Côme and J. Jacques. The Discriminative Functional Mixture Model For a Comparative Analysis of Bike Sharing Systems, *The Annals of Applies Statistics*, 2015, p.1726-1760, DOI:10.1214/15-AOAS861

[4] A. Cuevas, M. Febrero-Bande and R. Fraiman. Robust Estimation and Classification for Functional Data via Projection-based Depth Notions. *Computational Statistics*, 2007, p.481-496, DOI:10.1007/s00180-007-0053-0.

[5] M. Febrero-Bande and M. Oviedo de la Fuente. Statistical Computing in Functional Data Analysis: The R Package fda.usc, The Journal of Statistical Software, 2012, p.1-28, DOI:10.18637/jss.v051.i04

[6] M. Febrero-Bande, P. Galeano, W. Gonzalez-Manteiga. Outlier Detection in Functional Data by Depth Measures, with Application to Identify Abnormal NOx Levels. *Environmetrics*, 2008, p.331–345. DOI:https://doi.org/10.1002/env.878.

[7] F. Ferraty, P. Vieu. Nonparametric Functional Data Analysis. Springer, 2006, p.37-43.

[8] R. Izem, J.G Kingsolver. Variation in Continuous Reaction Norms: Quantifying Directions of Biological Interest, *The American Naturalist*, 2005, p.277-289, DOI:https://doi.org/10.1086/431314

[9] L.S. Leonard. Pathophysiology of Heart Disease: A Collaborative Project of Medical Students and Faculty. Lippincott Williams & Wilkins, 2010, p.74.

[10] M. Matabuena, J. C. Vidal, P. R. Hayes, M. Saavedra and F. Huelin Trillo. Application of functional data analysis for the prediction of maximum heart rate, *IEEE Access*, 2019, DOI: 10.1109/ACCESS.2019.2938466

[11] S.J. Ratcliffe, L.R. Leader, and G.Z. Heller. Functional data analysis with application to periodically stimulated foetal heart rate data. I: Functional regression. *Statistics in Medicine*, 2002, p.1103-1114. DOI:10.1002/sim.1067

[12] W. von Döbeln, I. Åstrand and A. Bergström. An Analysis of Age and Other Factors Related to Maximum Oxygen Uptake, *Physiology*, 1967, p.934-942, DOI:https://doi.org/10.1152/jappl.1967.22.5.934

[13] Y. Zhou and N. Sedransk. Functional data analytic approach of modeling ECG T-wave shape to measure cardiovascular behavior, *The Annals of Applied Statistics*, 2009, p.1382-1402, DOI:10.1214/09-AOAS273