Time series anomaly detection methods for smartphone sensor-based mental health care

Riku Tikkanen

School of Science

Bachelor's thesis
Espoo 5.9.2021

Supervisor

Prof. Eero Hyvönen

Advisor

Dr. Talayeh Aledavood

Aalto University
School of Science

Aalto University
School of Science

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi
Abstract of the bachelor's thesis

Author Riku Tikkanen

Title Time series anomaly detection methods for smartphone sensor-based mental health care

Degree programme Science and Technology

Major Computer Science **Code of major** SCI3027

Teacher in charge Prof. Eero Hyvönen

Advisor Dr. Talayeh Aledavood

Date 5.9.2021 **Number of pages** 26 **Language** English

Abstract

Wearable health-monitoring systems are becoming increasingly popular as many diseases including mental disorders can be avoided and properly managed through continuous monitoring.

Mental disorders cause changes in patients' mental states which is often followed by changes in physical activity, sleep and social interaction. These parameters can be measured using time series data produced by sensors available in modern smartphones such as GPS, accelerometer and gyroscope. The key in mental health monitoring is to detect the changes in patient's mental states. The solution is to apply anomaly detection methods to detect anomalies in the time series data which translate to changes in mental states.

This bachelor's thesis is a literature review that deals with finding different time series anomaly detection methods that can be used for mental health care purposes using smartphone sensors. Anomalies are abrupt variations in time series data and detecting such variations is useful in many application areas such as medical condition monitoring. The thesis presents common anomaly detection methods available and discusses how they can be used to monitor and detect mental disorders.

Keywords time series, anomaly detection , change point detection , smartphone, health care

Aalto University
School of Science

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi
Tekniikan kandidaatintyön tiivistelmä

Tekijä Riku Tikkanen	
Työn nimi Menetelmiä älypuhelinten sensoreihin perustuvaan mielenterveyshuoltoon	
Koulutusohjelma Teknistieteellinen kandidaattiohjelma	
Pääaine Tietotekniikka	**Pääaineen koodi** SCI3027
Vastuuopettaja Prof. Eero Hyvönen	
Työn ohjaaja Tri. Talayeh Aledavood	
Päivämäärä 5.9.2021	**Sivumäärä** 26 **Kieli** Englanti

Tiivistelmä

Tämä kandidaatin tutkielma on kirjallisuuskatsaus, joka käsittelee erilaisia aikasarjapoikkeamien havaitsemismenetelmiä, joita voidaan käyttää älypuhelinten sensoridataan perustuvassa mielenterveydenhuollossa. Aineisto koostuu tutkimusraporteista, jotka käsittelevät aikasarjapoikkeamia sekä älypuhelinpohjaisia terveydenhuoltoratkaisuja. Työssä osoitetaan, että on paljon todistusaineistoa siitä, kuinka mielenterveyttä voidaan monitoroida ja diagnosoida erilaisten nykyaikaisten algoritmien sekä älypuhelinten sensoridatan avulla.

Mielenterveyshäiriöt aiheuttavat vaihteluja niistä kärsivien henkilöiden mielentilassa, ja niiden havaitseminen on erittäin keskeistä mielenterveydenhuollossa. Mielentilan vaihtumista seuraavat usein erilaiset muutokset henkilöiden käyttäytymisessä kuten fyysinen aktiviteetti, nukkumisrytmi sekä sosiaalinen kanssakäyminen. Näitä käyttäytymisparametreja voidaan seurata nykyaikaisten älypuhelinten sensorien kuten GPS:n, kiihtyvyysmittarin ja gyroskoopin avulla.

Käyttäytymistä mittaavat sensorit tuottavat aikasarjadataa, jossa esiintyvät poikkeamat viittaavat käyttäytymisessä tapahtuviin muutoksiin ja siten voivat myös viitata mielentilan muutoksiin. Poikkeamat ovat äkillisiä tai yllättäviä vaihteluja aikasarjadatassa, ja niiden etsimiseen kehitettyjä menetelmiä kutsutaan aikasarjapoikkeamien havaitsemismenetelmiksi. Menetelmät ovat hyödyllisiä monessa sovelluskohteessa kuten esimerkiksi terveydenhuollossa, sillä poikkeamien havaitseminen mahdollistaa aikaisen hoidon sekä jatko-ongelmien ennaltaehkäisyn.

Avainsanat aikasarjadata, poikkeamien havaitseminen, älypuhelin, terveydenhuolto

Contents

1 Introduction

Mental health problems are common and the world population is aging rapidly thanks to advances in modern medicine and technology[1]. Both mental health patients and old people require a lot of resources in the form of nursing homes, hospitals and personnel. Health care costs in general are rising steadily which leads to the problem of finding cost-efficient alternatives.

Many mental health problems can be prevented, detected and properly managed by continuous monitoring [2]. This was previously only possible under the direct observation of medical personnel in appropriate facilities. However, through the advances in smartphone and wireless communication technology patients can remain connected with their health care providers while remaining in their personal environment [3].

Mental disorders cause changes in patients' mental states which are often followed by changes in physical activity, sleep and social interaction [2]. Such behavioural parameters can be measured using sensors available in modern smartphones such as GPS, accelerometer and gyroscope [4]. The GPS (Global Positioning Sensor) sensor of a smartphone uses a satellite-based navigation system consisting of a network of satellites to provide position, velocity and timing information [5]. The accelerometer of a smartphone measures the change of velocity (acceleration) of the smartphone in its own rest frame, and a gyroscope measures orientation and angular velocity [6]. All of these sensors capture changes at a single point in time and therefore generate time series data.

The key to remote mental health monitoring is to detect changes in patients' behaviour to detect if a change in mental state has occurred [5]. This requires finding changes in the behavioural time series data generated by smartphone sensors. Changes in time series data are better known as anomalies, outliers, exceptions or contaminants depending on the context [7]. The problem of finding anomalies is known as anomaly detection, and different methods have been studied since the 19th century. Anomaly detection methods are increasingly popular in many fields including mental health monitoring because anomalies often imply that a significant or critical event such as a change in mental state is happening.

The aim of this study is to review time series anomaly detection methods and suggest how they can be applied to monitor and detect mental health disorders from passive data created by smartphone sensors. This thesis focuses on a selected collection of papers to find common methods and algorithms used for anomaly detection in time series data.

The content of this thesis is organized as follows. The second chapter presents background information of this thesis and the third chapter discusses which methods are used for research. The fourth and fifth chapter present different anomaly detection methods and algorithms used for mental health care. The sixth chapter brings further insight to monitoring and detecting common mental disorders with smartphone sensor data and chapter seven concludes the findings of this thesis.

2 Background

This chapter presents background information relevant to this thesis. The current state of detecting mental disorders with smartphones is explained and the concept of time series and anomalies are introduced.

2.1 Mental health monitoring with smartphones

Through advances in computing and communication technology smartphones have become useful measuring devices. Smartphones are widely available and their number is growing steadily[3]. Moreover, modern smartphones feature several built-in sensors such as image sensor (camera), global positioning system (GPS) sensor, accelerometer, gyroscope, magnetometer, ambient light sensor, microphone and fingerprint sensor, most of which can be useful in symptom monitoring and diagnosis [2]. Previous studies show that smartphone sensor data can be used for monitoring Schizophrenia [8, 9, 10], Depression [11, 12], Anxiety [13] and bipolarity [2].

The number of people worldwide suffering from mental disorders was estimated to be 792 million in 2017 [1]. The most common disorder being anxiety (284 million) followed by depression (264 million), alcohol use (107 million) and bipolarity (46 million) [1]. In addition to the rise of mental disorders the age of the world population is growing as well. It is expected that by 2050 the number of people over the age of 65 is outnumbering the number of children under the age of 14 [2]. Nursing homes and hospitals are expensive and many elderly people prefer to stay at home. This supports researching mobile monitoring systems for both physical and mental health problems.

Conventionally the diagnosis of mental disorders is based on self-report questionnaires or surveys, which are deployed through smartphones. In clinical practice symptoms of schizophrenia are evaluated using the Brief Psychiatric Rating Scale (BPRS) [8], symptoms of Bipolarity are evaluated using the Bipolar Spectrum Diagnostic Scale (BSDS) [14] and symptoms of Depression are evaluated using the Patient Health Questionnaire-9 (PHQ-9) [5]. Such surveys are unreliable and prone to biases as the key feature in mental disorders is the variation of mood over time, signifying that the outcome may differ depending on subject's mood [5]. This further emphasizes the need for more passive methods.

Smartphone sensor data can be passively aquired without the need for self-report questionnaires. Such data can be used to monitor physiological variables such as heart rate and electromyography and behavioral patterns including sleep duration and social activities [15]. These parameters can be further translated into biomarkers of mental dissorders such as depression, anxiety and schizophrenia, enabling monitoring and early intervention [2]. Furthermore, the results from smartphone sensors might increase the chances of treatment success by focusing on creating personalized treatment plans [2]

2.2 Time series

Time series is a sequence of successive equally spaced data points in time, or a collection of observations made chronologically [16]. In addition to its continuous nature, time series data is characterized by its numerical values. For example we might consider considered a sequence of variables x_1 , x_2, x_3..., where x_1 corresponds to the first value which the series has taken at the first point in time. The variable x_2 corresponds to the second value and so on.

Time series data points created by smartphone sensors could be oberved at any continuous point in time and as such can be treated as *continuous time series*. Furthermore the approximation of such time series will be mostly discrete because of restrictions caused by the method of data collection.

2.3 Anomaly detection

Unexpected changes or patterns in time series data can happen due to internal or external events. Such patterns are considered anomalies, outliers, exceptions or contaminants depending on the context [7]. While malfunctions and errors can also be classified as anomalies, this thesis instead focuses on finding anomalous patterns which might refer to medical symptoms. The term anomaly detection refers to the problem of finding these patterns through various methods, many of which are specificly developed for certain application domains, however also generic techniques exist [7].

Anomaly detection methods assume two things: (1) the occurence of anomalies is rare and (2) anomalies are different from normal data in one way or another [17]. Point anomalies refer to data instances which are different with respect to the rest of the data [18]. An example of a point anomaly is an anomaly in daily behaviour such as lying on the floor in the kitchen. A contextual anomaly on the other hand can be viewed as an unusual sequence of behaviour such as preparing a meal –> reading a book –> having a nap indicating that the subject forgot to eat the meal [18].

Anomalous data points can be considered local when the anomality exists in respect to nearby data points or global when the anomality exists in respect to the whole dataset. One study used local K-nearest neighbor distances to detect local network anomalies such as outages and network-scans [19]. The study built a local anomaly detection model with the goal to improve results when compared to a model using global K-nearest neighbor distances. The existence of both local and global anomalies creates a challenge since it is impossible to determine which class an anomaly belongs to, however it also acts as a reminder that anomalies can differ not only from the data but from each other as well.

Anomaly detection methods are useful in many applications because such anomalies often imply that a significant or critical event is happening. An anomaly in time series data can for example relate to a change in mood or state in mental patients such as transitioning from beeing depressed to not beeing depressed. Detecting behavioral anomalies in real-time can also help to prevent relapse into schizophrenia by signaling the need of an intervention [9]

3 Methods of research

This thesis is a literature review which focuses on studies about time series anomaly detection methods and studies about applying these methods to smartphone data for the purpose of monitoring and detecting mental disorders. The methods of finding and extracting material are explained in this chapter.

3.1 Search strategy

The material used in this thesis was found by using Google Scholar [20], a search engine optimized for finding scientific publications, and the Scopus database [21] which contains references from over 15,000 journals. A combination of different keywords was employed to search for studies which focus on time series anomaly detection methods and on using smartphone sensor data for mental health care purposes. The search was limited to only include article titles, abstracts and keywords and to only include studies written in english.

Keywords used for research include: "smartphone", "smartphone-based", "wearable", "passive data", "time series", "anomaly", "outlier", "change-point", "detection", "monitoring", "diagnosis", "mental health", "health", "health care", "sensor", "depression", "anxiety".

Different queries were established using the boolean operators "AND" and "OR". Below is a query used for searching relevant material from Google Scholar and Scopus database.

("anomal*" OR "change point*" OR "change-point*") AND ("analys*" OR "detect*" OR "recogni*" OR "identif*") AND ("smartphone*" OR "wearable" OR "passive data") AND ("mental health*" OR "digital phenotyping")

The query returned 9 results from the Scopus database and 135 results from Google Scholar.

3.2 Screening

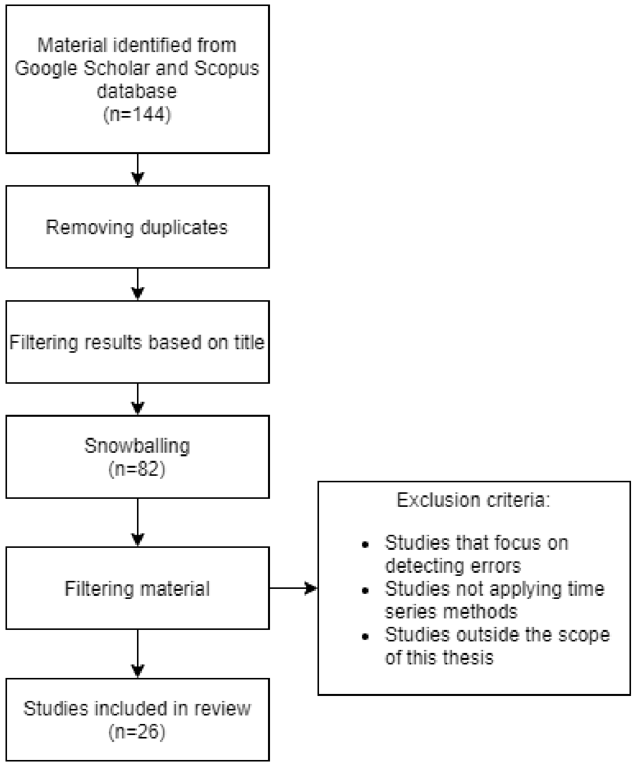

Figure 1: Methodology of research

After removing duplicates from the search results were filtered based on the title and abstract. New literature was found through snowballing, i.e. finding new material in the reference lists of previous studies. As a result a collection of 82 studies was recognized as potentially useful for this thesis. The search results were further delimited by removing studies that focus on detecting errors in data caused by faulty sensors rather than detecting and classifying symptoms and diseases. Next, studies not applying time series methods were removed. Finally studies that were outside the scope of this thesis, i.e. not focusing solely on health care purposes were removed. Eventually the filtering led to 26 studies being selected for the final review.

4 Change Point Detection

Many mental dissorders cause mood swings in patients [2]. The key to successful monitoring and intervention is to find when these changes happen. Change Point Detection (CPD) is a statistical method for discovering the points in time where the behavior of time series suddenly changes [22], making it highly suitable for monitoring purposes.

Half of the selected studies mention change point detection methods being useful for health monitoring purposes [23, 24, 13, 25, 26, 27, 28, 29, 30, 31, 32, 33] and six of these studies directly present applications of change point detection in mental health monitoring [23, 24, 13, 25, 26, 27].

This chapter focuses on presenting change point detection methods as suggested by reviewed studies.

4.1 Supervised methods

Change point detection techniques include supervised methods which can be further separated into multi class classifiers, binary classifiers and virtual classifiers [22]. The latter is not mentioned in any reviewed studies and will therefore not be further analyzed. Supervised methods require labeled datasets, i.e. a training set to learn a function which can be used after training to predict the labels for unlabeled datapoints in new datasets. Supervised methods are commonly used in health care for predicting further problems and decision making [34].

4.1.1 Multi class classifiers

The training of multi class classifiers requires that all possible states are known [22]. Furthermore a large and diverse dataset representing all the different states and all the possible transitions between them must be available. In change point applications multiclass classifiers are used to detect possible changes between two data points by running a predefined sliding window through the data [27]. Common multi-class classifiers used for change point detection include support vector machine (SVM), hidden Markov model (HMM) and Gaussian mixture model (GMM) [25, 22].

4.1.1.1 Support Vector Machines

Support vector machines (SVM) can be used for both classification and regression problems, however for change point detection the focus is on the first [22]. Originally SVM's were made for binary classification, classifying data points as either 1 or 0, however multiclass classification is also possible [22].

For simplifying the concept behind SVM a binary classification example may be considered using a two-dimensional dataset. A SVM finds a hyperplane to divide the dataset into two classes by maximizing the distance (margin) from the hyperplane to the nearest two points. In this example the result would be a line dividing the two dimensional data into two classes. The two datapoints nearest to the hyperplane are considered the most critical ones as their removal would have the greatest impact on the classification result, and are hence given the name support vector.

Mathematically the hyperplane dividing the points is represented by the equation

$$H : w^T \phi(x) + b = 0$$

and the distance between any point and the hyperplane is measured by

$$d = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}$$

Similarly, the distance of the hyperplane from any given point is written as

$$d_H(\phi(x_o)) = \frac{|w^T(\phi(x_0)) + b|}{||w||_2}$$

, where $||w||_2$ is the Euclidean norm form for w.

SVM initially becomes an optimization problem to maximize the margin, i.e. finding the "maximum minimum" distance

$$w^* = \arg\max_w [\min_n d_H(\phi(x_n))]$$

.

While making a prediction, subsituting any positive group point in the hyperplane equation will result in a value higher than zero, $w^T \phi(x) + b > 0$ and vice versa substituting any negative group value will result in a value lower than zero $w^T \phi(x) + b < 0$.

This example can be considered a perfectly separable dataset and therefore the optimal equation becomes

$$w^* = \arg\max_w \left[\min_n \frac{y_n|w^T(\phi(x_n)) + b|}{||w||_2} \right]$$

.

While the example above focused on binary classification, the multiclass problem is not far from it. The multiclass problem consists of multiple binary classification cases used to separate one class at a time from all the other classes [22].

4.1.1.2 Hidden Markov Model

Hidden Markov Model (HMM) is a commonly used method for temporal pattern recognition and has been used for detecting change-points in the field of bioinformatics, brain imaging and speech recognition [35]. As the name suggests, HMM is based on a Markov process which is used to compute the probability for a sequence of observable events [36]. Often the events of interest are however not directly observable, i.e. they are hidden. In contrast to a Markov process a HMM is useful for finding both observed and hidden events [36]. The probability of a state's occurrence are computed by calculating a histogram of the probabilities of successor states.

Change point detection can be initally seen as a HMM with observations being represented by the data and hidden states by the unknown segmentation [35]. Consequently, by identifying the observations where the likelihood of a switch in a hidden state is most likely, HMM adaptations can detect change points [35].

A two layer HMM has been used for change point detection to monitor Activities of Daily Livings (ADLs) in home environments [37]. The model combined low-level sensor data and binary sensor data to map the location of a person with a possible activity such as sleeping in the bedroom or brushing teeth in the bathroom.

4.1.1.3 Gaussian Mixture Model

Gaussian Mixture Models (GMM) are usually employed in an unsupervised fashion for pattern recognition in data, however the regonition accuracy of unsupervised GMM is not always optimal, suggesting a supervised model could be used instead [38]. GMM are based on the assumption that the input data are a combination of gaussian distributions and the fact that a probability density function $f(x)$ can be approximated with the Gaussian probability density basis functions to any degree of accuracy [34]. To initiate a A GMM some parameters are required as the first step after which the model is re-estimated based on the data and calculations [34].

Previous studies show that supervised GMM has been used to detect change points for activity recognition purposes [39]. The study used accelerometer data sampled at 50 Hz as input for a Gaussian Mixture Classifier to distinquish between different acitivities. Additionally microphone data was used to classify between transportation activities and GPS and Wi-Fi was used for validation.

Another study focusing on suicide prevention proved a Gaussian-Bernoulli mixture model to be useful for finding indicators of behavioral shifts in patients [25]. The study focused on collecting passive data from mobile phone sensors including bluetooth and GPS to monitor variables representing daily circadian profiles of a patient. In the study the Gaussian-Bernoulli mixture model was applied to project high-dimensional patient data into a more interpretable discrete representation.

The studies mentioned in this section show proof that GMM can indeed be used in a supervised manner. GMM are able to detect unseen information but the computation time required to construct the model is high and therefore using GMM in real-time frameworks is not suggested [34].

4.1.2 Binary Classifiers

In comparsion to multiclass classifiers, binary classifiers only need to learn two classes; one class representing all of the possible change point sequences and another class representing the sequences within the same state. The low number of classes can be misleading since a binary classification problem becomes quite complex if the number of transitions is large [22].

Binary classifiers have also been used for change point detection for activity monitoring by representing activity recognition as a two-class problem consisting of activity and transition [40]. A range of supervised algorithms such as Naïve Bayes, Logistic Regression, AdaBoost and Decision Trees were used to map sensor data to activities.

As mentioned in the previous chapter SVM's can be used for moth multiclass and binary class problems. One example proves detecting changes in human activities by using a binary class SVM. A high-dimensional hypersphere and the distribution of radii of the spheres [41] was used to classify data points. The high and low values in the distribution of radii corresponded to change points, i.e. changes in activity.

4.2 Unsupervised methods

Supervised methods can be very effective, however they are less useful for a variety of data[30] suggesting that unsupervised methods might be more suitable even for mental health care. The problem with supervised methods is that the labeling process grows with the amount of sensors involved and number of activities measured. Moreover the precise labeling of a supervised training set for long-term continuous data requires observing mental patients in a controlled clinical unit thus increasing costs [7]. It is also possible that the closely supervised conditions used for training purposes might not reflect normal living conditions.

The goal for unsupervised algorithms is to find hidden patterns in a dataset containing only data instances. In contrast to supervised methods, unsupervised algorithms do not require labeled datasets, i.e. training data, which makes them more attractive and cost-efficient for many purposes especially when dealing with large datasets. Many of the selected studies prefer unsupervised methods for this reason [29, 30, 32].

In the field of change point detection unsupervised algorithms are used to segment time series data based on statistical features [27]. A variety of unsupervised algorithms have been applied for change point detection [22], however this chapter will only focus on the most popular methods as suggested by reviewed studies. The most popular unsupervised algorithm is Bayesian which was mentioned in six studies [22, 27, 31, 25, 36, 32], followed by Gaussian Process [22, 27, 13, 36] and CUSUM [22, 27, 28, 30] suggesting that probabilistic methods and likelihood methods might be most succesfull.

4.2.1 Bayesian

Probabilistic methods such as Bayesian and the Gaussian Process (GP) try to estimate probability distributions for each window by using all the observed data since the previous change point in the calculation [27]. The probability of an event happening after taking into consideration all the previous information available is known as the posterior probability [31]. Therefore the Bayesian methods are based on calculating the posterior probability of each data point being a change point.

Early Bayesian change point detection methods were offline however one of the first online method BCBD (Bayesian Change Point Detection) was developed by Adams and McKay [42]. The algorithm assumes that for any run length r_t the predictive distribution conditional can be calculated. Essentially the algorithm estimates how much time has passed since the previous change point. The run length distribution is represented by:

$$P(r_t|x_{1:t}) = \frac{\sum_{r_{t-1}} P(r_t|r_{t-1}))P(x_t|r_{t-1}, x_t^{(r)})P(r_{t-1}, x_{1:t-1})}{\sum_{r_t} P(r_t, x_{1:t})}$$

where $P(r_t|r_{t-1})$ is the prior component, $P(x_t|r_{t-1}, x_t^{(r)})$ the likelihood component and $P(r_{t-1}, x_{1:t-1})$ the recursive component of the equation. The conditional prior gives the algorithm its computational efficiency since it has only two outcomes with nonzero values [42]. Either the run length r_t continues to grow and r_t becomes $r_{t-1} + 1$ or a change point is found and $r_t = 0$.

$$P(r_t|r_{t-1}) = \begin{cases} H(r_{t-1} + 1) & if r_t = 0 \\ 1 - H(r_{t-1} + 1) & if r_t = r_{t-1} + 1 \\ 0 & otherwise \end{cases}$$

A Bayesian change point detection model was used previously in clinical decision support [31]. The study compared several algorithms and found that a Bayesian model is useful within some limits. The strength of a Bayesian method was its ability to provide a probability distribution for each point being a change point and therefore giving information about the number, location and degree of confidence of change points in data. The study suggested that a Bayesian model is best used for detecting point and mean-shift anomalies.

Another study presented a method for suicide attempt prevention using Bayesian change point detection [25]. The study assumed that daily circadian profiles of patients can be divided into non-overlapping subsequences representing behaviors using change points as separators. Furthermore it was assumed that each behavior has a surrogate generative distribution whose unknown parameters could be found using a Bayesian method. The outcome of the study proved that the tool built can be used for suicide prevention and can be further personalized for even better results.

4.2.2 Gaussian Process

As mentioned before, a Gaussian Process is another probabilistic method for time series analysis. A GP is defined as a probability distribution over a set of functions $p(f)$, where the functions are defined by a collection random variables $f = f_1, f_2, ..., f_N$. Any finite linear combination of these variables has a joint (zero mean) Gaussian distribution[22]. The functions used in GP are specified by the mean function $\mu(x)$ and covariance function $k(x, x')$:

$$f(x) = GP(m(x), k(x, x'))$$
$$\mu(x) = \mathbb{E}[f(x)]$$
$$k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))][36]$$

For change point detection a GP is used to estimate the predictive distribution at a time t by using the available observations through time$(t-1)$. For any observation a p-value is calculated and a threshold is used to decide whether the observation follows the predictive distribution or not. By comparing an observation to the predictive distribution the algorithm determines if a change of state has occurred and if the observation is indeed a change point[22].

A GP is more complicated than the BCBD algorithm presented in the previous section, since it uses all possible observations through time $t-1$ instead of only the previous state. This however, makes the GP algorithm also more accurate [22].

4.2.3 CUSUM

Methods based on likelihood ratio are another popular section of unsupervised algorithms used for CPD [22]. These methods are used for in CPD by analyzing the probability distributions of data before and after a possible change point has occurred [22]. Moreover the logarithm of the likelihood ratio between two successive states is monitored [43].

The cumulative sum (CUSUM) algorithm is one of the earliest density ratio based methods [27]. It works by gathering deviations relative to a specified target of measurements and detects change points by comparing the cumulative sum to a threshold value.

The CUSUM algorithm is defined by $T^C = inf\{n \geq 1 | C_n \geq h\}$, where h is the threshold and the statistic C_n is the CUSUM statistic

$$C_n = \max_{1 \leq v \leq n} \{\prod_{j=v}^{n} \frac{g(Y_j)}{f(Y_j)}\}$$

, representing the maximum likelihood ratio over all time windows[28].

The CUSUM algorithm has been previously to monitor cardiac surgery outcomes [44]. The study applied the CUSUM algorithm to identify change points in prior signals.

5 Anomaly Detection

The previous chapter presents common CPD methods as suggested by the selection of reviewed studies. Anomaly detection (AD) is a similar concept used to detect patterns which differ significantly from the remaining data. As mentioned before, detecting anomalies in daily human acitivity may be the key to successful monitoring and intervention in mental health problems.

Nearly half of the selected studies mention the use of AD methods for health monitoring purposes [45, 18, 46, 47, 11, 12, 9, 10, 48]. As do CPD methods, AD methods also rely on machine learning to detect anomalies in time series data and can be employed in a supervised or unsupervised manner, however unsupervised methods seem to be prefered.

The algorithms mostly mentioned for anomaly detection are SVM [47, 12, 45, 18, 46], followed by k-means [11, 48, 45, 47], Principal Component Analysis (PCA) for dimensionality reduction [46, 47, 45, 49], HMM [47, 18, 48] and Bayesian Network [47, 18, 46].

This chapter focuses on presenting AD methods as suggested by reviewed studies while minimizing the repetition of previously mentioned algorithms.

5.1 K-means

K-means is a popular clustering algorithm used to group data points into k clusters such that similar data points are near to each other and data points in different clusters are farther apart. K-mean measures the similiarity of two data points by calculating the distance between them. The distance between two data points can be measured in many ways, however the Euclidean distance $(a, b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$ is commonly used [7].

Put simply the K-means algorithm using the Euclidean distance works as follows:

1. Randomly select a number from 1 to K.
2. For each cluster K calculate the cluster centroid.
3. Calculate the distance of each data point to each centroid.
4. Assign the data points to the nearest cluster.
5. Calculate the new centroid for each cluster by taking the mean of all data points in that cluster.
6. Repeat steps 3, 4 and 5 until convergence.

The K-means algorithm is relatively fast and the results are easy to interpret. Furthermore it quarantees convergence, however the number of clusters k must be pre-determined which can be a difficult task.

One study used K-means with Wasserstein distance to create a unsupervised framework for anomaly detection in seqential data for health care purposes [45].

The study used a Variational Recurrent Autoencoder for unsupervised learning and K-means to separate anomalous and normal data into different clusters. The model was tested on data produced by an electrocardiogram and the results prove that their model was able to perform anomaly detection without labels. Furthermore, the approach was able to compete with conventional SVM based models and outperformed models proposed by recent works mentioned in the study.

Another study based on predicting depressive episodes using smartphone sensors applied the K-means algorithm for anomaly detection [11]. The motivation of the study was to verify sleep patterns by using the GPS sensor available in smartphones. By applying the K-means algorithm a users location could be determined as home, work or school for example.

5.2 PCA

Principal Component Analysis (PCA) is commonly used in conjunction with machine learning to reduce the dimensionality of data. The goal of reducing the dimensionality of the original data is to reduce computational complexity and speed up the process. In Anomaly Detection PCA is often applied to multivariate problems[46]. PCA works by comparing the correlation among monitored parameters and representing a p-dimensional data into a new subspace of k principal components where $k \leq p$. The new principal components each represent a linear comibination of the original p parameters [46]. PCA reduces dimensionality while retaining the maximum variability of the original data, therefore returning a result nearly as good as the original p-dimensional data.

One of the reviewed studies used PCA for detecting and diagnosing anomalous physiological measurements produced by sensors of patients[46]. PCA was applied to lower the complexity of data to enable real time anomaly detection using resource-constrained devices ie. smartphones. The study applied different versions of PCA algorithms such as classical PCA, ROBust PCA (ROBPCA), Spherical Principal Component (SPC) and PCA with robust covariance estimation (PCACOV). The results show that ROBPCA projected the original 14-dimensional dataset into a 6-dimensional dataset while retaining 100% of the variability.

A second study compared Variational Autoencoders to PCA for unsupervised anomaly detection [49]. The study found that the anomaly detection method using the reconstructional probability from a variational autoencoder is much better compared to reconstruction error of an autoencoder and PCA based method.

5.3 Bayesian Network

A Bayesian network (BN) is a probabilistic graphical model that captures both conditionally dependent and conditionally independent relationships between random variables [7]. The network consists of nodes which represent random variables and edges between nodes representing the relationship of those variables. All missing edges in the graph define the conditional independencies in the model. The goal of a BN is to model the posterior conditional probability distribution of variables after observing given evidence [7]. BN are attractive since they can be either constructed manually by experts or can be constructed automatically from data and then used to estimate the probabilities of events happening.

A BN was used previously to develop a smart assisted-living system for elderly care [18]. The study constructed a framework based on complex activity recognition using dynamic Bayesian network modeling. A total of eight difference body activites categorized into stationary and motional activites were monitored. Additionally five different hand gestures were tracked. The study applied semi-supervised learning methods to teach the model using recordings of normal living activites. Experiments in a mock environment proved that the proposed framework detects behavioural anomalies in real time effectively.

Bayesian networks were also used in a medical body sensor network which collects physiological signs to monitor the health of patients [50]. The study formalized a BN to describe the body sensor network considering both the spatial and temporal correlation between physiological parameters measured by different sensors. Experiments were carried out on medical datasets from the PhysioNet database and the results prove that errors were reduced by 60% using the BN based anomaly detection model.

6 Applications in mental health care

The previous chapters present common time series anomaly and change point detection methods which are used in health care as suggested by reviewed studies. Many of these studies apply anomaly and change point detection methods to smartphone sensor data to monitor behavioural and physiological changes in subjects. Behavioural and physiological parameters are important for mental health monitoring because variations in data such as location, movement and sleep duration are all associated with mental disorders [51, 4, 2].

This chapter focuses on presenting information about common mental disorders and how they are being monitored using smartphone sensors as suggested by reviewed studies. The goal of this chapter is to bring more insight into why previously mentioned algorithms can be applied to smartphone sensor data to monitor and detect common mental disorders.

6.1 Depression

Depression is one of the most common mental disorders and is characterized by the absense of positivity, low mood, loss of interest in activities and experiences and a range of associated symptoms [4]. The symptoms can range from concentration difficulties and fatique to feeling of anxiety and emptiness [11]. In worst case scenarios depression can lead to suicidal attempts.

Depression can be detected by focusing on behavioural parameters such as physical activity and location data which can indicate a decrease in social interaction with others [48]. Furthermore abnormalities in sleep such as increased sleep latency, decreased total sleep time and sleep efficiency are all common indicators of major depressive disorder [51]. Additionally conversational frequency can be monitored and voice data can be analyzed to measure pitch, tempo and loudness, which all might refer to depression [4]. Light intensity might also be a plausible indicator, however it is unlikely to be predictive of depressed mood [4]. All of these parameters can be measured using commonly available sensors in smartphones such as GPS, gyroscope, accelerator and microphone.

6.2 Bipolar disorder

Bipolar disorder is characterized by successive episodes of manic and depressed mood states and is a major cause of disability worldwide [4]. As the state of a patient changes from manic state to depressive state, so does the physical and social activity of the patient [36]. It is not unusual that patients in the depressive state have a decreased desire to socially interact with others while patients in the manic state have an increased desire to do so. The physical movement and travel patterns are also affected depending on the state of a patient [52].

One study [25] suggests that changes in mental states can be detected by modelling the daily circadian profiles of a patient and applying change point detection to identify transitions. Other studies found significant correlations between the activity levels and bipolar states of individuals by monitoring the activity level using a smartphone's accelerometer [2]. Further studies [52, 14] used accelerometer data in conjunction with GPS data to detect the mental state and change of state in patients suffering from bipolar disorder. The first of which reported detecting changes in states with a precision of 96% , proving that smartphone sensor data can be used for detecting bipolar disorders as well.

6.3 Schizophrenia

Schizophrenia is a neuropsychiatric syndrome which causes psychotic, negative and cognitive symptoms in patients [4]. Psychotic symptoms include hallucinations and delusions, negative symptoms include the loss of motivation and cognitive syptoms include difficulties in paying attention, decreased working memory and decreased verbal fluency [36]. Schizophrenia is not a fixed state but rather the majority of people suffering from schizophrenia fluctuate between states of partial remission and symptomatic relapse [8]. A relapse can cause severe difficulties in patients resulting in high individual and societal costs. However, schizophrenia related disorders can be properly managed and a relapse can be detectected by the continuous monitoring of subjects [8].

One of the reviewed studies proposed a CrossCheck symptom prediction system which combines smartphone sensors with patient questionnaires to detect symptoms of schizophrenia [8]. A convergence of sensors were used to monitor the movement, activity and sleep patterns of patients. The experiments of the study show that the system performs well using passive sensing with or without the addition of self-reports. The study concludes that a system purely based on passive sensing opens the way for continuous assesment of symptoms and risks.

Another study focusing on relapse prediction in schizophrenia through digital phenotyping used a similar combination of self-reports and sensors [9]. The study collected passive data from GPS and accelerometer and anonymized call and text message logs. Additionally screen on/off time and phone charging status was monitored. The results of the study prove that smartphones offer opportunities to explore psychiatric symptoms and functioning by combining statistical and computational methods such as anomaly detection.

7 Summary and concluding remarks

The objective of this thesis was to find different time series anomaly detection methods that can be applied to smartphone sensor data to monitor and detect mental disorders. A variety of methods and algorithms were introduced based on a collection of studies selected for this literature review. The algorithms fall into two categories, namely Change Point Detection (CPD) and Anomaly Detection (AD).

CPD is a statistical method for discovering points in time where the behaviour of time series data suddenly changes. CPD was mentioned in many of the reviewed studies in correlation with health care purposes including mental health. AD refers to finding unexpected changes or patterns in time series data and was also used for similar purposes.

Both CPD and AD methods rely on machine learning algorithms to find changes in time series data generated by smartphone sensors. GPS, accelerometer and gyroscope were amongst the popular sensors used because they measure behavioural parameters such as activity, sleep patterns and social interaction. Behavioural parameters are of interest in mental health care because changes in behaviour may refer to changes in mental states.

Unsupervised machine learning algorithms were prefered due to the fact that they require no training and labeling. The results show that similar algorithms were used for both CPD and AD methods, namely SVM, HMM and Bayesian Network. SVM's appear to be popular because they can be applied to both multiclass and binary classification problems.

This thesis proves that time series anomaly detection methods can be applied to smartphone sensor data for mental health care purposes. Many of the reviewed studies rely on combining passive sensor data with self-reports of subjects' well being to detect changes in mental states. However, there is also proof that a shift into completely passive monitoring is on its way.

8 References

[1] H. Ritchie and M. Roser, "Mental health," *Our World in Data*, 2018. https://ourworldindata.org/mental-health.

[2] S. Majumder and M. Deen, "Smartphone sensors for health monitoring and diagnosis," *Sensors*, vol. 19, no. 9, 2019.

[3] A. Pantelopoulos and N. Bourbakis, "A survey on wearable sensor-based systems for health monitoring and prognosis," *IEEE Transactions on Systems*, vol. 40, no. 1, p. 1–12, 2010.

[4] Q. M. Sheikh M and Kyriacou, "Wearable, environmental, and smartphone-based passive sensing for mental health monitoring," *Front. Digit. Health*, vol. 3, 2021. https://ourworldindata.org/mental-health.

[5] D. Ben-Zeev, "Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health," *Psychiatric Rehabi-litation Journal*, vol. 38, no. 3, p. 218–226.

[6] B. Bar-Erdene and J. Saver, "Automatic acute stroke symptom detection and emergency medical systems alerting by mobile health technologies: A re-view," *Journal of Stroke and Cerebrovascular Diseases*, vol. 30, no. 7, 2021.

[7] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv*, vol. 41, no. 3, Article 15, p. 58, 2009. DOI = 10.1145/1541880.1541882.

[8] R. Wang, "Predicting symptom trajectories of schizophrenia using mobile sensing," *PACM Interact. Mob. Wearable Ubiquitous Technol*, vol. 1, no. 3, 2017.

[9] T. Barnett, S. Staples, and J.-P. O. Keshavan, "Relapse prediction in schizophrenia through digital phenotyping: A pilot study," vol. 43, no. 8, pp. 1660–1666, 2018.

[10] D. Henson and V. et al., "Anomaly detection to predict relapse risk in schizophrenia," vol. 11, no. 28, 2021.

[11] T. Jeong, D. Klabjan, and J. Starren, "Predictive analytics using smartphone sensors for depressive episodes." Available at:.

[12] K. Opoku Asare, A. Visuri, and D. S. T. Ferreira, "Towards early detection of depression through smartphone sensing," (New York, NY, USA), Association for Computing Machinery, 2019.

[13] J. Rubin, H. Eldardiry, R. Abreu, S. Ahern, H. Du, A. Pattekar, and D. G. Bobrow, "Towards a mobile and wearable system for predicting panic attacks," (New York, NY, USA), Association for Computing Machinery, 2015.

[14] V. Osmani, "Smartphones in mental health: Detecting depressive and manic episodes," *IEEE Pervasive Computing*, vol. 14, no. 3, pp. 10–13, 2015.

[15] T. Aledavood, A. M. T. Hoyos, T. Alakörkkö, K. Kaski, J. Saramäki, E. Isometsä, and R. K. Darst, "Data collection for mental health studies through digital platforms: requirements and design of a prototype," *JMIR research protocols*, vol. 6, no. 6, p. e110, 2017.

[16] T. Fu, "A review on time series data mining," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 1, pp. 164–181.

[17] E. e. a. Šabić, "Healthcare and anomaly detection: using machine learning to predict anomalies in heart rate data," *AI and Soc*, vol. 36, pp. 149–158.

[18] C. Zhu, W. Sheng, and M. Liu, "Wearable sensor-based behavioral anomaly detection in smart assisted living systems," *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 4, pp. 1225–1234, 2015.

[19] V. Saligrama and M. Zhao, "Local anomaly detection," in *Artificial Intelligence and Statistics*, pp. 969–983, PMLR, 2012.

[20] "Google scholar." https://scholar.google.com/. Accessed: 2021-08-31.

[21] "Scopus." https://www.scopus.com/. Accessed: 2021-08-31.

[22] S. Aminikhanghahi and D. Cook, "A survey of methods for time series change point detection," *Knowl Inf Syst*, vol. 51, no. 2, p. 339–367, 2017.

[23] J. Cabrieto, J. Adolf, F. Tuerlinckx, P. Kuppens, and E. Ceulemans, "An objective, comprehensive and flexible statistical framework for detecting early warning signs of mental health problems.," *Psychotherapy and psychosomatics*, vol. 88, no. 3, pp. 184–186, 2019.

[24] M. I. Hehlmann, B. Schwartz, T. Lutz, J. M. Gómez Penedo, J. A. Rubel, and W. Lutz, "The use of digitally assessed stress levels to model change processes in cbt-a feasibility study on seven case examples," *Frontiers in Psychiatry*, vol. 12, p. 258, 2021.

[25] P. Moreno-Muñoz, L. Romero-Medrano, Á. Moreno, J. Herrera-López, E. Baca-García, and A. Artés-Rodríguez, "Passive detection of behavioral shifts for suicide attempt prevention," *arXiv preprint arXiv:2011.09848*, 2020.

[26] L. Cruz, J. Rubin, R. Abreu, S. Ahern, H. Eldardiry, and D. G. Bobrow, "A wearable and mobile intervention delivery system for individuals with panic disorder," in *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*, pp. 175–182, 2015.

[27] S. Aminikhanghahi, T. Wang, and D. J. Cook, "Real-time change point detection with application to smart home time series data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 5, pp. 1010–1023, 2019.

[28] S. W. Han, *Efficient change detection methods for bio and healthcare surveillance*. Georgia Institute of Technology, 2010.

[29] S. Aminikhanghahi and D. J. Cook, "Enhancing activity recognition using cpd-based activity segmentation," *Pervasive and Mobile Computing*, vol. 53, pp. 75–89, 2019.

[30] C. Culman, S. Aminikhanghahi, and D. J Cook, "Easing power consumption of wearable activity monitoring with change point detection," *Sensors*, vol. 20, no. 1, p. 310, 2020.

[31] S. Ray, D. S. McEvoy, S. Aaron, T.-T. Hickman, and A. Wright, "Using statistical anomaly detection models to find clinical decision support malfunctions," *Journal of the American Medical Informatics Association*, vol. 25, pp. 862–871, 05 2018.

[32] E.-S. Apostol, C.-O. Truică, F. Pop, and C. Esposito, "Change point enhanced anomaly detection for iot time series data," *Water*, vol. 13, no. 12, p. 1633, 2021.

[33] H. Dette and D. Wied, "Detecting relevant changes in time series models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 78, no. 2, pp. 371–394, 2016.

[34] H. Banaee, M. U. Ahmed, and A. Loutfi, "Data mining for wearable sensors in health monitoring systems: A review of recent trends and challenges," *Sensors*, vol. 13, no. 12, pp. 17472–17500, 2013.

[35] T. M. Luong, V. Perduca, and G. Nuel, "Hidden markov model applications in change-point analysis," 2012.

[36] M. Osipov, "Towards automated symptoms assessment in mental health," *arXiv preprint arXiv:1908.06013*, 2019.

[37] M. H. Kabir, M. R. Hoque, K. Thapa, and S.-H. Yang, "Two-layer hidden markov model for human activity recognition in home environments," *International Journal of Distributed Sensor Networks*, vol. 12, no. 1, p. 4560365, 2016.

[38] J. Ma and W. Gao, "The supervised learning gaussian mixture model," *Journal of Computer Science and Technology*, vol. 13, no. 5, pp. 471–474, 1998.

[39] I. Cleland, M. Han, C. Nugent, H. Lee, S. McClean, S. Zhang, and S. Lee, "Evaluation of prompted annotation of activity data recorded from a smart phone," *Sensors*, vol. 14, no. 9, pp. 15861–15879, 2014.

[40] K. D. Feuz, D. J. Cook, C. Rosasco, K. Robertson, and M. Schmitter-Edgecombe, "Automated detection of activity transitions for prompting," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 5, pp. 575–585, 2015.

[41] R. Vlasveld, “Temporal segmentation using support vector machines in the context of human activity recognition,” Master’s thesis, 2014.

[42] R. P. Adams and D. J. MacKay, “Bayesian online changepoint detection,” *arXiv preprint arXiv:0710.3742*, 2007.

[43] Y. Kawahara and M. Sugiyama, “Sequential change-point detection based on direct density-ratio estimation,” *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 5, no. 2, pp. 114–127, 2012.

[44] H. Assareh, I. Smith, and K. Mengersen, “Bayesian change point detection in monitoring cardiac surgery outcomes,” *Quality Management in Healthcare*, vol. 20, no. 3, pp. 207–222, 2011.

[45] J. Pereira and M. Silveira, “Learning representations from healthcare time series data for unsupervised anomaly detection,” in *2019 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp. 1–7, 2019.

[46] L. Ben Amor, I. Lahyani, M. Jmaiel, and K. Drira, “Anomaly detection and diagnosis scheme for mobile health applications,” in *2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA)*, pp. 777–784, 2018.

[47] R. A. A. Habeeb, F. Nasaruddin, A. Gani, I. A. T. Hashem, E. Ahmed, and M. Imran, “Real-time big data processing for anomaly detection: A survey,” *International Journal of Information Management*, vol. 45, pp. 289–307, 2019.

[48] D. Mohr, “Harnessing context sensing to develop a mobile intervention for depression,” *Journal of Medical Internet Research*, vol. 13, no. 3.

[49] J. An and S. Cho, “Variational autoencoder based anomaly detection using reconstruction probability,” *Special Lecture on IE*, vol. 2, no. 1, pp. 1–18, 2015.

[50] H. Zhang, J. Liu, and R. Li, “Fault detection for medical body sensor networks under bayesian network model,” in *2015 11th International Conference on Mobile Ad-hoc and Sensor Networks (MSN)*, pp. 37–42, 2015.

[51] T. Aledavood, J. Torous, A. M. T. Hoyos, J. A. Naslund, J.-P. Onnela, and M. Keshavan, “Smartphone-based tracking of sleep in depression, anxiety, and psychotic disorders,” *Current psychiatry reports*, vol. 21, no. 7, pp. 1–9, 2019.

[52] A. Grünerbl, A. Muaremi, V. Osmani, G. Bahle, S. Oehler, G. Tröster, O. Mayora, C. Haring, and P. Lukowicz, “Smartphone-based recognition of states and state changes in bipolar disorder patients,” *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 1, pp. 140–148, 2014.