CZECH TECHNICAL UNIVERSITY IN PRAGUE

Faculty of Electrical Engineering

Department of Cybernetics

MASTER'S THESIS

Michal Náhlík

Sleep-Wake and Sleep Stage Detection from Wrist-Worn Actigraphy

Thesis supervisor: **Ing. Eduard Bakštein, Ph.D.**

January 2019

MASTER'S THESIS ASSIGNMENT

I. Personal and study details

Student's name:	**Náhlík Michal**	Personal ID number:	**434509**
Faculty / Institute:	**Faculty of Electrical Engineering**		
Department / Institute:	**Department of Cybernetics**		
Study program:	**Biomedical Engineering and Informatics**		
Branch of study:	**Biomedical Informatics**		

II. Master's thesis details

Master's thesis title in English:

Sleep-Wake and Sleep Stage Detection from Wrist-Worn Actigraphy

Master's thesis title in Czech:

Detekce cyklu spánek-bdění a spánkových stádií z aktigrafického záznamu

Guidelines:

Wrist-worn actigraphy is a common technique for long-term activity monitoring, used in sleep medicine, analysis of circadian rhythmicity and recently also for management of psychiatric diseases.
For the evaluation of sleep parameters it is necessary to segment the actigraphic recording into sleep and wakefulness periods. Multiple methods exist for detection of sleep segments, but most are provided without rigorous comparison to existing solutions. Moreover, evidence for the ability of actigraphy to detect internal sleep structure is limited.
The aims of this work include:
1. Perform a review of existing methods for sleep/wake detection using acigraphy [1-5].
2. Implement at least 3 of the available solutions and evaluate their performance on real-life actigraphic data.
3. Evaluate the effect of rescoring [6] on the results of sleep detection algorithms.
4. Investigate the possibility to detect sleep stages from 3D actigraphy data.
5. Design an algorithm to detect sleep stages from 3D actigraphy data and compare the results with polysomnography.

Bibliography / sources:

[1] A. Sadeh, K. M. Sharkey, and M. A. Carskadon, "Activity-based sleep-wake identification . An empirical test of Methological issues," Sleep, vol. 17 (3), no. April 1994, pp. 201-207, 1994.
[2] R. J. Cole, D. F. Kripke, W. Gruen, D. J. Mullaney, and J. C. Gillin, "Automatic sleep/wake identification from wrist activity.," Sleep, vol. 15, no. 5, pp. 461-469, 1992.
[3] J. Paquet, A. Kawinska, and J. Carrier, "Wake detection capacity of actigraphy during sleep.," Sleep, vol. 30, no. 10, pp. 1362-9, 2007.
[4] C. Crespo, M. Aboy, F. Jr, and A. Mojón, "Algorithm for Sleep/Wake Identification From Actigraphy," in BIOSIGNAL, 2010, no. 1, pp. 224-228.
[5] J. Tilmanne, J. Urbain, M. V. Kothare, A. Vande Wouwer, and S. V. Kothare, "Algorithms for sleep-wake identification using actigraphy: a comparative study and new results," J. Sleep Res., vol. 18, no. 1, pp. 85-98, 2009.
[6] Webster JB, Kripke OF, Messin S, Mullaney OJ, Wyborney G. ?An activity-based sleep monitor system for ambulatory use?. Sleep 1982;5:389-99.
[7] Yuda, E., Yoshida, Y., Sasanabe, R., Tanaka, H., Shiomi, T., & Hayano, J. (2017). Sleep Stage Classification by a Combination of Actigraphic and Heart Rate Signals. Journal of Low Power Electronics and Applications, 7(4), 28

Name and workplace of master's thesis supervisor:

Ing. Eduard Bakštein, Ph.D.,   Analysis and Interpretation of Biomedical Data,   FEE

Name and workplace of second master's thesis supervisor or consultant:

Date of master's thesis assignment: **18.12.2017**   Deadline for master's thesis submission: **08.01.2019**

Assignment valid until: **30.09.2019**

Ing. Eduard Bakštein, Ph.D.	doc. Ing. Tomáš Svoboda, Ph.D.	prof. Ing. Pavel Ripka, CSc.
Supervisor's signature	Head of department's signature	Dean's signature

III. Assignment receipt

Date of assignment receipt	Student's signature

Author statement

I declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with the methodical instructions for observing the ethical principles in the preparation of university theses.

Prague, date............................   ....

signature

Abstract

The aim of this thesis is to review and evaluate the performance of existing algorithms used to detect sleep-wake periods from a long term actigraphy signal and investigate the possibility to identify different sleep stages that occur during sleep.

Wrist actigraphy has been widely recognized as a suitable low cost and low intrusive alternative to polysomnography for sleep and wake identification. While most of the current applications and algorithms focus mainly on detecting sleep and wakefulness during the night, the main goal of this thesis is to evaluate their ability to identify sleep and wake periods in long-term recordings and compare the results to ground truth obtained from sleep diaries. Several existing algorithms were implemented and evaluated on a data set from a long term actigraphy study. Also a new algorithm based on decision trees and simple features is proposed as well as a new post-processing method which outperforms existing solutions.

In the second part of this work, exploratory data analysis was performed using data obtained from simultaneous actigraphy and polysomnography recordings to investigate the possibility of sleep stages detection during the night. Further on, several machine learning models were developed for automatic detection of sleep stages using actigraphy data or actigraphy in a combination with heart rate signals. However, the results of identifying different sleep stages are not very promising and as other sources suggest, wrist actigraphy even in combination with heart rate signals might not contain the necessary information to do so and further research might be needed.

Keywords: Actigraphy, Machine learning, Sleep monitoring, Sleep stages, Sleep-wake

Abstrakt

Cílem této práce je vyhodnotit existující algoritmy používané k detekci spánku a bdění v dlouhodobých aktigrafických záznamech a prozkoumat možnosti identifikace jednotlivých spánkových stádií které se vyskytují v průběhu spánku.

Aktigrafie je široce uznávaná jako vhodná levná a pohodlná alternativa k polysomnografii pro identifikaci spánku a bdění. Zatímco většina stávajících využití a algoritmů se soustředí hlavně na detekci spánku a bdění v průběhu noci, hlavním cílem této práce je vyhodnotit jejich schopnosti identifikovat období spánku a bdění v dlouhodobých záznamech a porovnat jejich výsledky s daty získanými ze spánkových deníků. Pro ověření bylo implementováno několik existujících algoritmů, které byly následně vyhodnoceny na datech z dlouhodobé aktigrafické studie. Také byl navržen nový algoritmus založený na rozhodovacích stromech a jednoduchých ukazatelích, stejně jako nová metoda následného zpracování, která dosahuje lepších výsledků než stávající řešení.

V druhé části této práce byla provedena explorační analýza dat získaných ze souběžně pořízených záznamů z aktigrafie a polysomnografie pro vyhodnocení možností detekce spánkových stádií během noci. Dále bylo vytvořeno několik modelů k automatické detekci spánkových stádií na základě aktigrafických dat a aktigrafických dat v kombinaci se záznamem srdečního rytmu. Nicméně dosažené výsledky při identifikaci jednotlivých spánkových stádií nejsou příliš povzbudivé, a jak naznačují i další zdroje, aktigrafie i v kombinaci se záznamy srdečního rytmu možná neobsahují potřebné informace a bude tak potřeba dalšího výzkumu.

Klíčová slova: Aktigrafie, Strojové učení, Sledování spánku, Spánková stádia, Spánek-bdění

Contents

List of Figures

List of tables

List of Abbreviations

ACC - Accuracy
ACT - Actigraphy
ASD – Absolute sleep distance
ANN – Artificial neural network
AWD – Absolute wake distance
DT – Decision tree
ECG - Electrocardiography
EEG – Electroencephalography
HRV – Heart rate variability
PSG - Polysomnography
SD – Sleep distance
SEN – Sensitivity
SPE - Specificity
WD – Wake distance

1 Introduction

Humans spent on average a third of their life by sleeping and it's a very important part of it. Proper sleep is essential for a good mental and physical health. The exact purpose of sleep is yet unknown but it has been shown it's critical for many vital physiological functions including development and maintaining synaptic pathways, removing brain waste that builds up while we are awake, modulation of immune responses, memory consolidation and many others. Lack or low quality of sleep can lead to not only physiological but also psychological problems and recent studies suggests that it could be even linked to neurodegenerative diseases like Alzheimer years before the typical symptoms are visible. With up to 20% of population affected by some minor or major sleep problems and with a sleep deprivation as a common part of a current lifestyle the sleep research got a lot of attention in last 20 years.

To better understand the function of sleep and its influence on the human life it's critical to have reliable methods to track its length and evaluate its quality. Those information can then be used as a useful marker of person's current health status, could give us a hint of a future risks and changes in the sleep behavior could provide a warning sign of upcoming problems. While the wearable devices are on the rise, tracking many aspects of our life often including the sleep, in the clinical studies there are three main validated ways how to monitor sleep: polysomnography, sleep diaries and actigraphy.

Sleep is most often defined by the physiological characteristics observed like closed eyes, reduced breathing rates, reduced body movement and responsiveness to external stimuli and specific brain wave activity. These characteristics are used to detect sleep, identify its stages and determine the sleep parameters that are needed to evaluate its quality.

Polysomnography is considered golden standard in sleep evaluation. It is a procedure performed in sleep laboratory where subject is connected to many sensors monitoring brain, heart and muscle activity, eye movement, oxygen levels and other information which are then split into time frames and analyzed by a specialist using standardized set of rules. This provides the most exact data about sleep but is not very comfortable for the patient and it's usually limited to one night. Polysomnography is currently the only reliable method that provides information about the sleep structure and the occurrence of different sleep stages. While this might be important for diagnosing specific sleep disorders or related problems, the complexity of the procedure makes polysomnography not suitable for longer studies and we such a detailed data are not always needed.

Sleep diaries on the other hand represent a very simple approach. Every morning subject fills out a form containing information about last night sleep, including when they went to bed, when they fell asleep, when they woke up, how many time they woke up during the night and so on. This can provide a general overview of the sleep over several weeks, but requires consistency and can be easily affected by individual's errors.

Actigraphy is an accepted alternative to polysomnography for sleep-wake detection. It is a method to monitor person's activity using most often a watch like device worn on the wrist with accelerometer inside that periodically saves the amount of activity recorded. These recordings do not give direct information about sleep but can be later analyzed and used to identify the periods of sleep and wakefulness [18]. There are several different devices and multiple algorithms to automatically analyze the data which makes actigraphy a useful tool to track persons sleep over longer period of time without being intrusive or requiring much human effort. Unfortunately unlike polysomnography there are no standardized set of rules for scoring actigraphic signals and so the sleep-wake detection results can widely vary device by device or algorithm by algorithm. However, actigraphy is considered a long-time established method for evaluation of basic sleep parameters in home environment [19].

As actigraphy is viewed as alternative to polysomnography the current algorithms [1][2][3][4][5][6] focus mainly on detecting sleep in comparison with it. That means that they try to detect sleep and wakefulness occurring during the night of one night examination. While this is a useful as an alternative to polysomnography to detect quality of sleep during the rest period it might not be that useful and accurate in long term studies. The goal of this thesis is to evaluate and compare the current methods on a long term recordings and compare the results to sleep diaries and eventually propose a new method to improve the results. Such method would give us a possibility to analyze and determine person's sleep patterns and circadian rhythms which can be later used for example to monitor status of people with Bipolar disorder where sleep characteristics change during the manic and depression phase [13]. Because the ability to identify sleep stages occurring during the sleep is essential for sleep quality evaluation the second goal of this thesis is to investigate the possibility to do so from actigraphy signal.

2 Methodological background

2.1 Standard methods for sleep monitoring

2.1.1 Polysomnography

Polysomnography is considered a golden standard for monitoring sleep. It is one night non-invasive procedure usually performed in sleep laboratory when the subject is connected to sensors that simultaneously record many different physiological data including brain activity (EEG), heart activity (ECG), eye movement, leg movement, blood oxygen levels etc.
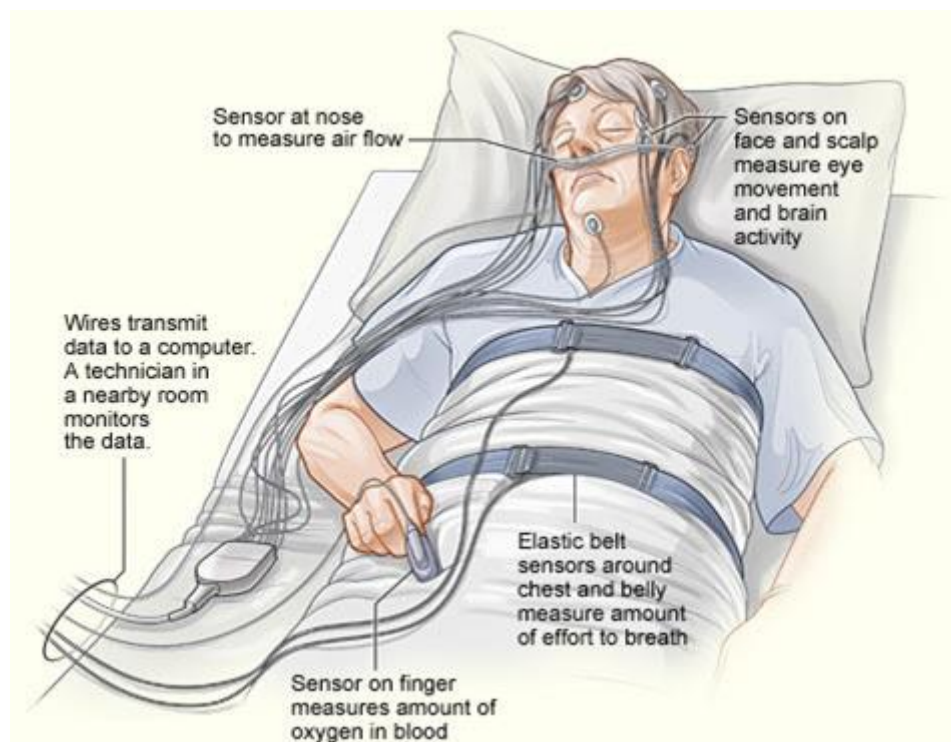
**Figure 1 Patient connected to sensors during the polysomnograph examination.
(Source: https://www.sleep-apnea-guide.com/polysomnogram.html)**

Sleep technician then analyzes the recorded signals in time frames (epochs) and labels them with following information: sleep stage, breathing irregularities, cardiac rhythm abnormalities, leg movements, body position. These information are then used to evaluate the sleep quality based on sleep latency, sleep efficiency and the occurrence of each sleep stage during the night. The evaluation and the recorded data are provided to a sleep medicine physician for interpretation. With consideration of patient's medical history, used drugs and other relevant

3

information the specialist is able to identify possible sleep disorders and offer recommendations for further treatment.

The polysomnography examination provides the most exact information about sleep, its quality and structure during the night. But even though it is non-invasive, the amount of sensors and the need to spend a night in a sleep laboratory may cause some discomfort to patients and negatively influence their sleep. This, the complexity of the examination and the human effort needed to perform and evaluate it limits the possible usage of this method in bigger long term studies. Yet is it currently the only method that provides accurate information about the sleep structure.

2.1.2 Sleep diaries

Sleep diaries represent the simplest approach to track individual's sleep. It is most often a simple paper form that servers as a record about sleep and is filled up by the individual ideally every morning right after waking up. The form might include information like when the person when to bed, when they fell asleep, when they woke up, number of times they woke up during the night, how the person felt during the day and other information that might be relevant for the specific purpose. In general, sleep diaries are considered subjective and systematic shifts have from values measured using polysomnography have been reported [12]. However, it is still the most widely used method for sleep evaluation in home environment.

Figure 2 Example of sleep diary form
(Source: https://sleepfoundation.org/sites/default/files/SleepDiaryv6.pdf)

While very simple it might be a useful tool for monitoring sleep over longer period of time (several weeks) and can be used to extract basic metrics for evaluation of sleep quality like wakefulness after sleep onset, number of awakenings, total sleep time, sleep efficiency, sleep onset latency and others. All these parameters showed high correlation with those obtained by polysomnography [12]. To eliminate the subjectivity that could be introduced by individual's memories, feelings or inconsistencies it is often used in conjunction with some other method like actigraphy.

2.1.3 Actigraphy

In the last 20 years actigraphy gained lots of attention and has been widely recognized as a low intrusive alternative to polysomnography in sleep-wake monitoring. An actigraph device is most often a small watch like wrist worn device that contains an accelerometer measuring motor activity and periodically saves the recorded values to a memory. These recordings are then downloaded and analyzed to identify the periods of rest (sleep) and wakefulness. The main concept of these devices is to be non-intrusive, simple and ideally with a long battery life so they can be used to an uninterrupted monitoring over a long period of time without needing attention.

5

Figure 3 Example of actigraph device (ActiGraph wGT3X-BT)
(Source: https://www.actigraphcorp.com/actigraph-wgt3x-bt/)

This allows to monitor individual's activity and hence sleeping behavior in more natural environment than in case of polysomnography and can serve as more consistent and objective measurement than sleep diaries [18][19]. The limitation of this method is actigraphy provides very specific information and that is just the motor activity. Some devices include a light sensor or button to record events but that might not be enough to obtain accurate results about sleep. For example it is very hard to distinguish between reading a book during the day and taking a nap. Both would show a lower activity but for sleep monitoring only one is relevant. That is why it is recommended to use it in conjunction with sleep diaries that provide additional information that may be then used for sleep evaluation.

2.1.4   Personal health monitors

The use and popularity of personal health monitors and smart wearable devices is increasing over the last decade allowing us to track our activity, heart rate, oxygen levels and other physiological processes in a real time. Many of them even offering options to track aspects of our sleep and analyzing its efficiency. Yet they are rarely used in clinical studies involving sleep due to the lack of validation, standardization or even proper description of measured metrics. Recent studies showed that most of the advanced outputs like sleep efficiency only weakly correlate with the output of polysomnography, while the total sleep time showed strong correlation [14]. The characteristics of the sleep measured and analyzed by these devices should be considered with the lack of validation in mind and possible inaccuracy in mind.

2.2 Current methods for automated evaluation

2.2.1 Sleep-wake detection using actigraphy

As mentioned before actigraphy just provides information about the activity of person in form of periodically recorded measurement by the accelerometer. The activity could represent for example the sum of absolute velocity change measured by accelerometer in a period of time given by the sampling frequency or how many times the change of velocity exceeded some preset threshold in that time frame. Based on the device, the measurement method, number of axis on the accelerator, sampling frequency, sensitivity and many other properties can vary and thus influence the result. Actigraphy unlike polysomnography has no set standards, not just for the devices but also for the algorithms used to identify sleep in actigraphic data. That can be a simple filtering and thresholding, linear regression or more advanced machine learning methods like neural networks. Every manufacture uses different approach and all of them allow to download the raw data so you can create your own classifier and optimize it to your own needs.
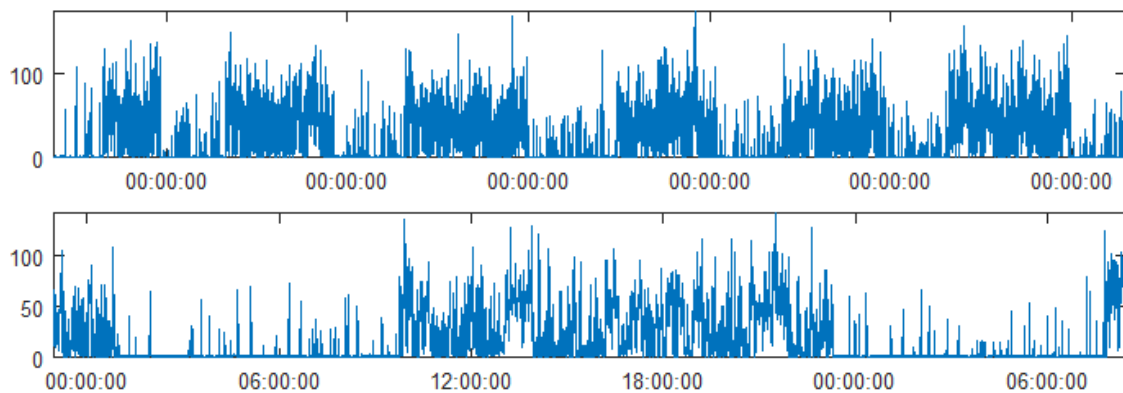
Figure 4 Example of data measured by actigraph. The upper graph shows activity over several days. Lower graph shows activity during a day and surrounding nights.

Several algorithms were considered for the evaluation but only four were selected. One of the first algorithms proposed using features calculated from actigraphy signal and linear regression. Simple algorithm using only values of past activity and logistic regression. And two models from a recent study that uses more advanced machine learning models. This should give us good overview of how different algorithms behave when compared to sleep diaries in long term recordings.

Sadeh et al. [3] developed an algorithm for a specific wrist actigraph in 1994. They calculated sixty-two activity variables for each 1 minute epoch and its surrounding 10 minute window and then used stepwise discriminant analysis to identify four best features. To develop

7

the final scoring algorithm they used discriminant analysis and the four most predictive variables identified before. The final scoring algorithm was:

$$PS = 7.601 - 0.065\ Mean\_w\_5\_min - 1.08\ NAT - 0.056\ SD\_last\_6\_min - 0.703\ LOG\_Act$$

where Mean_W_5_min is the average number of activity counts during the scored epoch and the window of five epochs preceding and following it, SD_last_6_min is the standard deviation of the activity counts during the scored epoch and the five epochs preceding it, NAT is the number of epochs with activity level equal to or higher than 50 but lower than 100 activity counts in a window of 11 minutes that includes the scored epoch and five epochs preceding and following it, LOG_Act is the natural algorithm of the number of activity counts during the scored epoch + 1. PS is the probability of sleep, if the value is zero or greater, then the specific epoch is scored as sleep; if PS is less than zero it is scored as wake.

The algorithm was evaluated on 16 healthy children and adolescents and scored 91.16% accuracy, 94.95% sensitivity and 74.5% specificity.

Sazonov et all. [2] proposed different approach using only activity of current and eight previous epochs in logistic regression. The accelerometer used in their study was sampled at 50HZ and was attached to a diaper of a baby which allowed them to monitor their position in a crib. To get values more similar to traditional actigraph they resamples this position signal into 30 second intervals (epochs) using the maximum value in that time frame. The model looked as follows:

$$\begin{aligned} h = 1.99604 &- 0.1945\ maxACC_0 - 0.09746\ maxACC_{-1} - 0.09975\ maxACC_{-2} \\ &- 0.10194\ maxACC_{-3} - 0.08917\ maxACC_{-4} - 0.08108\ maxACC_{-5} \\ &- 0.07494\ maxACC_{-6} - 0.073\ maxACC_{-7} - 0.10207\ maxACC_{-8} \end{aligned}$$

$$PS(h) = \frac{1}{1 + e^{-h}}$$

where maxAcc$_{-i}$ is the maximum of signal in the interval located *i* epochs before the current one and PS is the probability of sleep.

Sazonov et al. reported their algorithm reached 75.4% accuracy, 93.2% and sensitivity 41%.

Tilmanne etl al. [1] combined features used by Sadeh and Sazonov, added some more which resulted into set of 25 different activity variables. Using Fisher's discriminant analyses

they selected five most discriminant features which were further optimized utilizing Fisher's generalized criterion to find the optimal length of the feature window. The final features were: sum of all the activities of a 37-epoch centered window, activity standard deviation on a 25-epoch centered window, maximum epoch activity on a 19-epoch centered window, number of epochs in a 47-epoch centered window that have an activity superior to 2.025 times the mean activity of the file, logarithm of the current epoch activity increased by one. These features were used as an input for multilayer perceptron classifier with 5 hidden neurons and also for a decision tree model.

Multilayer perceptron model reached 80.5% accuracy, 91.9% sensitivity and 52.2% specificity while decision trees got 81.7% accuracy, 92.1% sensitivity and 57.1% specificity.

Metric Algorithm	Accuracy	Sensitivity	Specificity
Sadeh	91.16	94.95	74.5
Sazonov	75.4	93.2	41
Tilmanne MLP	80.5	91.9	52.2
Tilmanne DT	81.7	92.1	57.1

Table 1 Reported performance (percentage) overview. MLP – multilayer perceptron, DT – decision trees

Other methods considered were for example proposed by Crespo et al. [6] based on sequences of filtering with a mean accuracy of 94.3%. While this method was interesting and different than others the lack of implementation details did not allow us to replicate it. Paquet et al. [5] used two methods, one derived from algorithm proposed by Sadeh et al. [3] and one similar to Sazonov's algorithm [2] with reaching similar results as both mentioned. Cole et al. [4] developed algorithm based on the values in surrounding epochs and reached 87.91% accuracy. These methods were not implemented and evaluated because they are mostly modifications of the already selected ones.

Postprocessing

Webster et al. [7] noticed that the actual wake is more often falsely scored as sleep than sleep as wake which might be caused by the fact that polysomnography data used for training are not balanced by its nature. Polysomnography data usually consists of small period of wakefulness before sleep, sleep with some awakes during the night and a little wake period after sleep. This obviously makes most of the classifiers to optimize towards scoring wake as sleep. To mitigate this issue Webster et al. [7] proposed 5 rescoring rules that became a standard part of many following actigraphy studies. The rules are:

a) after at least 4 min scored wake, the first period of 1 min scored sleep is rescored wake

b) after at least 10 min scored wake, the first 3 min scored sleep are rescored wake

c) after at least 15 min scored wake, the first 4 min scored sleep are rescored wake

d) 6 min or less scored sleep surrounded by at least 10 min (before and after) scored wake are rescored wake

e) 10 min or less scored sleep surrounded by at least 20 min (before and after) scored wake are rescored wake

2.2.2 Sleep stages classification

The only method that currently allows to reliably identify sleep stages is polysomnography. Sleep technician analysis the recorded signals in time frames and assigns them a label based on standardized set of rules. These rules are described for example in The AASM Manual for the Scoring of Sleep and Associated Events [16]. From them we can see that while the polysomnography contains many different signals the key information to detect different sleep stages is EEG. Different types of brain waves appear during the stages which helps the technician to identify them. Further on they take into consideration the information about eye movement, chin EMG and repeated body movements. All this combined helps the specialist to classify the time frames into one of these categories: wake (S0 or sometime W), light non REM sleep (S1,S2), deep non REM sleep (S3, S4), REM Sleep (REM or R)
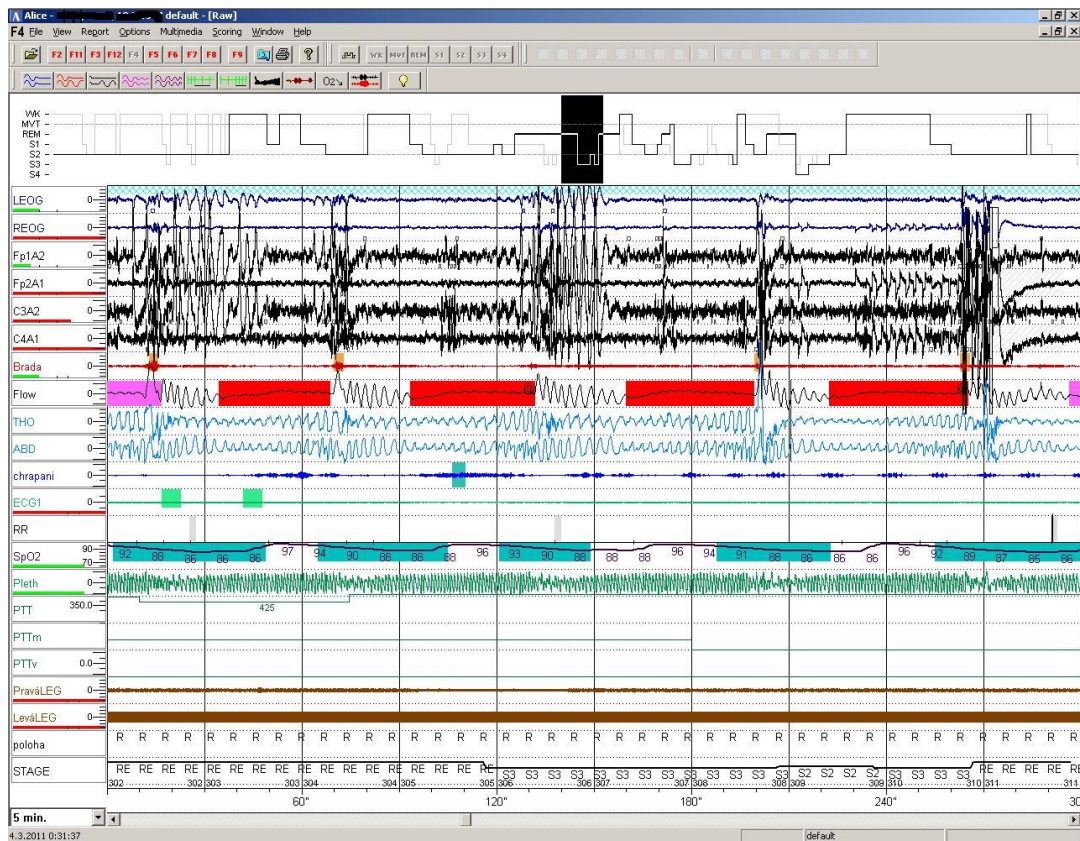
Figure 5 Example of polysomnography data. Top part contains labels assigned by a specialist with different signals that were measured shown below. (Source: http://pfyziolklin.upol.cz)

While the manual scoring is slow and time consuming process that requires a trained specilist there is currently no more precise method that would be used.

With the rise of personal health monitors there is an effort to make sleep stage detection possible with data recorded outside the laboratory settings. Such wearable devices often provide far less information and specifically lack the EEG. Several studies were conducted with the goal to identify at least REM from non REM sleep using a combination of actigraphy data and heart rate signal which can be both obtained from a wrist worn health monitor.

Yuda et al. [8] used the information about average, median, maximum and upper 95% values of body movement with combination of frequency characteristics of heart rate to create a multivariate logistic regression model. The model was able to distinguish NREM from other stages with 75.8% accuracy, 74.5% specificity and 76.9% sensitivity and identify REM from Waking with 74.5% accuracy, 72.3% specificity and 77.2% sensitivity respectively.

Beatie et al. [10] used same features like Yuda and added variables derived from heart rate in time domain like mean heart rate, 10th and 90th percentile of heart rate etc and also some new

actigraphy features like time since the last significant movement. These features were ten used for linear discriminant classifier and trained using labels Wake, Light, Deep, REM for each 30s epoch. The overall per-epoch accuracy of the model was 69%, overall sensitivity in detecting sleep 94.6%, specificity in detecting wake 69.3%, light sleep agreement 69.2%, deep sleep agreement 62.4% and REM sleep agreement 71.6%.

Tripathy et al. [15] reported average accuracy values of 85.51%, 94.03% and 95.71% in the classification of 'sleep vs wake', 'light sleep vs deep sleep' and 'REM vs NREM' sleep stages with features obtained from heart rate and EEG using deep neural network. The accuracy values for individual classes in multiclass classifier were: Wake 83.84%, Light sleep 57.75%, Deep sleep 72.66% and REM sleep 80.11%.

3 Technical background

3.1 Statistical methods

This chapter provides a brief overview of statistical methods used in this thesis. Unless stated otherwise the methods were not implemented by the author but the current implementations in the used software were utilized. For implementation details see the documentation of the used software.

3.1.1 Median filtering

Median filter is a nonlinear filtering technique most often used to remove noise from a signal. Every sample in a signal is replaced by median of its neighboring values where the number of neighbors is given by the size of the filter window. We use this filtering to smooth the classifier output by removing short false classification in the middle of correct ones.

3.1.2 Machine learning

Machine learning is a method of data analysis that focuses on algorithms and techniques allowing the computer to create and optimize a model based on input data. This allows a computer to gain knowledge and treat new unseen data in a similar way as the ones used for training without hard coding the solution. Because our data consists of samples (input) and also labels (desired output) we will use supervised learning methods that optimized the model based on the desired output.

Linear regression is used to model the relationship between a scalar response (dependent variable) and explanatory variables (independent variables). The model has form of following equation

$$Y = b_0 + \sum_{i=1}^{n} (b_i X_i) + \epsilon$$

where Y is continuous response, X_i is i-th independent variable, b_i is i-th parameter of the model (known as regression coefficient) and ε which is called error term (or noise). The goal is to reduce the value of ε over the whole training set of data using a estimation method. Linear regression has been successfully used in many applications and is one of the most used machine learning methods.

While **linear regression** models output as a continuous value the logistic regression is a different generalized linear model using the same formula but with the result representing a probability of a categorical outcome. The following equation is used

$$PS(Y) = \frac{1}{1 + e^{-(b_0 + \sum_{i=1}^{n}(b_i X_i))}}$$

where PS(Y) is probability of even Y (could be for example that Y belongs to class 1).
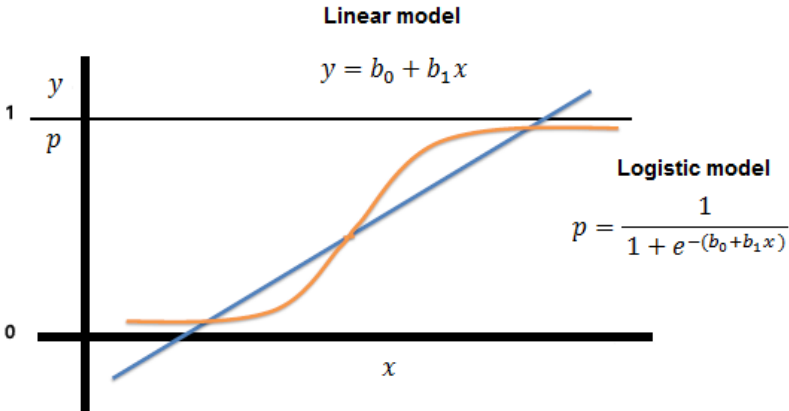
Table 2 Example of linear model outcome (blue) and logistic model outcome (orange)
(Source: https://www.saedsayad.com/logistic_regression.htm)

Decision trees are one of the basic tools in supervised machine learning. As the name suggest they use a set of rules that are applied to an input in a tree like structure. When input comes it's in the root of the tree, each internal node represents a variable of the input, branches represents the decision that was made based on the variable and each leafs is labeled with a class (in case of classification trees) or a value (in case of regression trees). These rules are applied to an input sequentially until it arrives through the tree to a leaf.

14
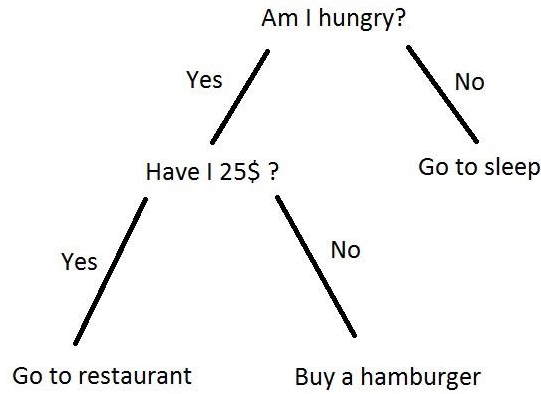
Figure 6 Simple example of decision tree
(Source: https://hackernoon.com/what-is-a-decision-tree-in-machine-learning-15ce51dc445d)

The tree is created in the learning process by splitting the source set into subsets based on the attribute and a value test, that process is then repeated on each derived subset until the splitting no longer improves the prediction. In this process so called split criterion or split function plays the key role, as it controls whether and how subsets will be split.

While simple and easy to interpret they provide a useful method and have been used successfully to solve complex problems. Their simplicity is also their disadvantage as they cannot leverage more complex relations between the attribute due to their sequential nature.

Artificial neural network is a mathematical model inspired by animal brain. It gained lots of attention in recent years as a powerful machine learning method that provides good results in various problems. The basic building unit of the network is a mathematical neuron that takes weighted inputs and if the sum of inputs exceeds a threshold it produces output using an activation functions. The most common type of network consists of one input layer of neurons, one hidden layer and one output layer. All neurons from one lower layer are connected to all from the upper one. This type of network is called feedforward neural network.

15

Figure 7 Detail of mathematical model of neuron (left), structure of multi layer feed forward network (right)

In the most common training method the network is then stimulated with a set of inputs and the weights are modified till the output of the network doesn't agree with the desired one for the given 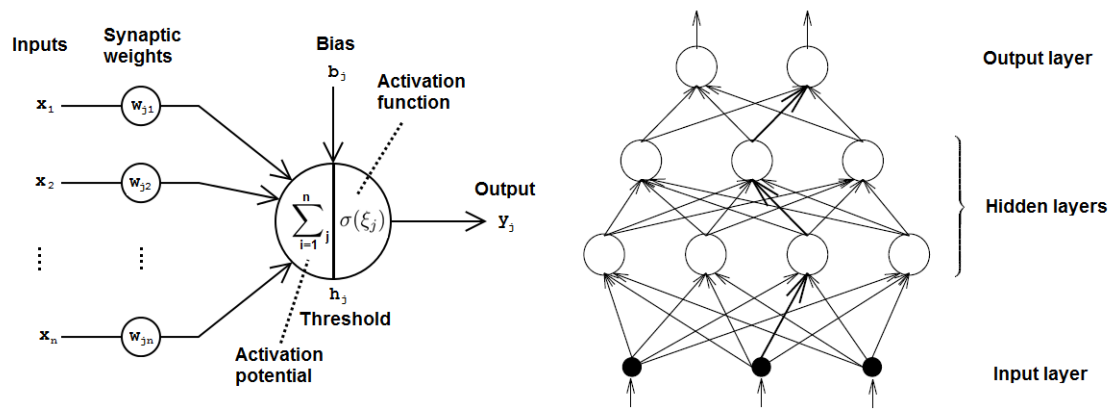input. The most common supervised learning method is backpropagation where the output is compared to the desired one and if there is an error it is back propagated to network and weights are moddified according to it. The training is done in multiple runs (epochs) and it's goal is to minimize the error over the whole training set.

This seemingly simple framework can solve complex problems without the need to implement the exact solution and offeres a wide variaty of possible setups.

3.1.3 Genetic algorithm

Genetic algorithm is an optimization technique based on a natural selection theory that uses mechanism from biological evolution. The process contains of few simple steps: generate randomly initial population, estimate the fitness of each individual in the population, select the best fit individuals for reproduction, create new population from those individuals using crossover and mutation and repeat until a desired solution isn't found. Individuals represent potential solutions to a problem we try to find solution for, population is then a set of different solutions. To be able to determine how fit individuals are we need to define a fitness function which will give us a numerical representation of the solution quality. The best solutions are then combined together in different ways (crossover) and random changes are introduced into them to ensure the populations can evolve and the optimization process doesn't get stuck in the local minimum.

This type of optimization is very useful when looking at a problem that doesn't have only one solution or it cannot be found in reasonable time. While it can be computationally expensive

most times in can converge to a suitable solution in reasonable time, if the population doesn't get better anymore it's easy to start over.

3.2 Metrics

3.2.1 Cross-validation

To increase the chance our model will generalize well and will provide similar results even on unseen independent data we use cross validation. The data set is split into training and testing subset. Training set is used to train the model and test set is used to validate the model, so the model performance is evaluated on data that were not used to optimize the model parameters. This will allow us to easily identify if model is overfitting because the performance would be good when evaluated on training set but poor when the test subset is used. This is done several times with different splits of the data. Multiple rounds of cross validation where the subset is changed reduce the chance that we selected a subset where the model performs well while it would not do so on different ones.

In our case we select the subsets subject-wise so we use never use data from one person for fitting the model and evaluating it at the same time. Subset of subjects is used to train the model and different subjects are used to evaluate the model. We will use the most exhaustive version where all subjects except one are used to train the model and the one is used to validate it. This process will be repeated until all subjects are used for test and the outputs will be used to evaluate the model performance. We will call this leave-one-subject-out cross validation.

3.2.2 Confusion matrix

The most common metrics to evaluate the performance of binary classificatory are based on confusion matrix. Using the predicted values by our model and the labels (correct class of a sample) we can create a confusion matrix. To do so we need to identify samples that were correctly classified as positive class (true positive – TP), correctly classified as negative class (true negative – TN), falsely classified as positive class (false positive – FP) and falsely classified as negative class (false negative – FN). This will allow us to calculate some basic statistical metrics.

		Actual class	
		1	0
Predicted class	1	TP	FP
	0	FN	TN

Table 3 Confusion matrix

In this text, we use the following convention: **Positive class for sleep periods, negative for wake.**

Accuracy is a fraction of predictions that our model classified correctly. In case of imbalanced data set this information might be misleading as classifying everything as the dominant class will result in high accuracy even though we were not able to identify correctly any occurrence of the class with low representation. Thus accuracy alone is not a sufficient metric to evaluate the model performance.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{TP + TN}{P + N}$$

The proportion of correctly classified elements from positive class is called **sensitivity** (also called recall or true positive rate).

$$Sensitivity = \frac{TP}{TP + FN} = \frac{TP}{P}$$

Specificity (also called true negative rate) is the proportion of actual elements from negative class that are correctly identified.

$$Specificity = \frac{TN}{TN + FP} = \frac{TN}{N}$$

3.2.3 Sleep start and sleep end difference

To evaluate how precisely the model identifies the start of the sleep and its end we will calculate additional metrics from the differences of the closest occurrences of the events classified by our model from the actual ones. This will be done by following steps:

1) Find the distances of actual sleep starts/ends according to labels to the closest classified ones by the model.
2) Remove distances that are bigger than 95th percentile. This is done to remove outliers that can occur at the start or end of the data because of missing labels.
3) Sleep distance (SD) /Wake distance (WD) - the mean of distances in minutes
4) Absolute sleep distance (ASD) / Absolute wake distance (AWD) – mean of absolute distance in minutes

3.3   Used software

All the calculations in this thesis were done using Matlab R2015b and following toolboxes: Signal processing toolbox, Statistics and Machine Learning Toolbox, Neural Network Toolbox.

4 Sleep-wake classification

4.1 Problem definition

Sleep-wake classification algorithms based on actigraphy are usually validated in comparison to polysomnography. Such data typically contains of short period of wakefulness, sleep with wake intervals that occur during the night and then short period of wakefulness. These data are recorded simultaneously with PSG and thus might by imbalanced by its nature. Algorithms are then designed and optimized to identify correctly the sleep and ideally to detect the wake that can be detected by polysomnography (which might be even labeled as wake just based on change in brain waves or eye movement and so the reason of its detection might not be visible on actigraphy). Goal of this thesis is to evaluate how these algorithms perform in long-term recordings that span over several days when we don't care that much about very short intervals of wake during night but are more concerned about correctly classifying the overall consistent sleep and wakefulness periods. That kind of information then can be used to analyze circadian rhythms or help with management of psychiatric diseases.

4.2 Dataset description

Data were recorded by National Institute of Mental Health in Prague, Czech Republic. The dataset consist of several days actigraphy recordings for 13 healthy volunteers of different age and gender. Activity was recorder by MotionWatch 8 in 30 seconds intervals with total of 325662 epochs, which is 8.7 days on average of 1 axis continuous signal per person. During that time each person also used sleep diary with following information: laid in the bed time, estimated sleep latency, got out of bed time, number of awakenings during the night. Estimated sleep time was used as a label for the actigraphy signal.

4.2.1 Exploratory data analysis

Dataset contains 325662 samples, with 98365 samples classified as sleep and 227297 as wake. The mean value of sleep samples is 3.89, 50th percentile is 0 and standard deviation is 12.84 for wake samples the values are 36.09, 38 and 25.42 respectively.
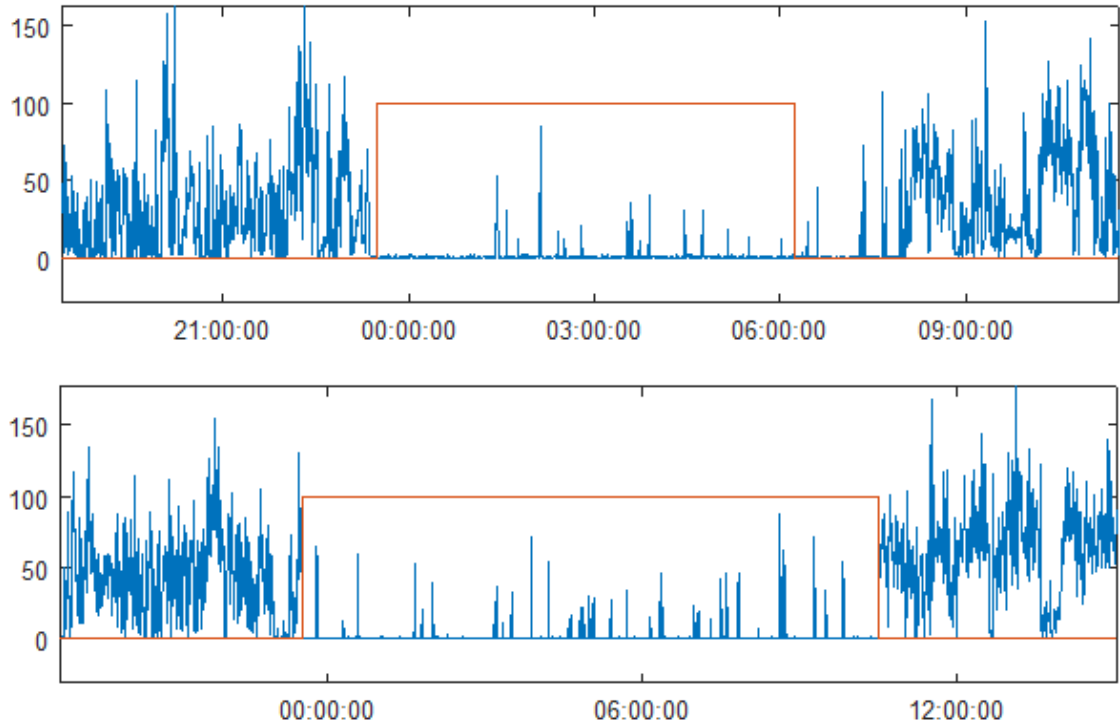
Figure 8 Two examples of actigraphic signal (blue) and label created based on sleep diary (red). Upper graph shows probably incorect label which was innacuratelly recorded in diary. Lower one contains signal where the activity drop is clearly visible inside the labeled sleep.

If we look at the data and labels on Figure 7 we can see that not all labels will probably fit perfectly with the data. That can be caused by several reasons like person wrote down the time of waking up but stayed in bed longer or just forgot to fill the form then did not remember the exact time. These inconsistencies (at least big ones) are luckily rare but they might hurt lower the highest possible accuracy we could reach and they can't be avoided. Small inaccuracies were reported even in other studies comparing actigraphy with sleep diaries [11].

Figure 9 Histogram of sleep and wake values on the left. Box plot of wake and sleep values on the right.

Majority of sleep samples have values lower than 5 while wake samples have values all around the spectrum.
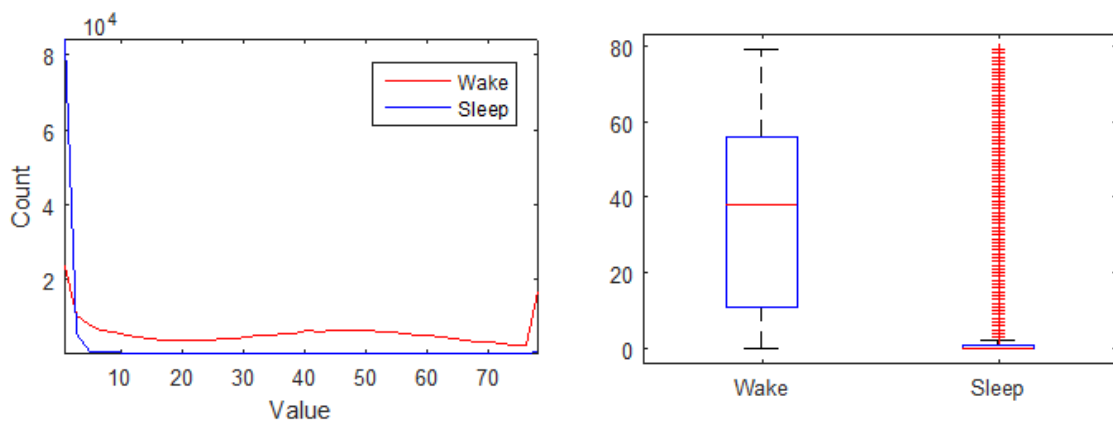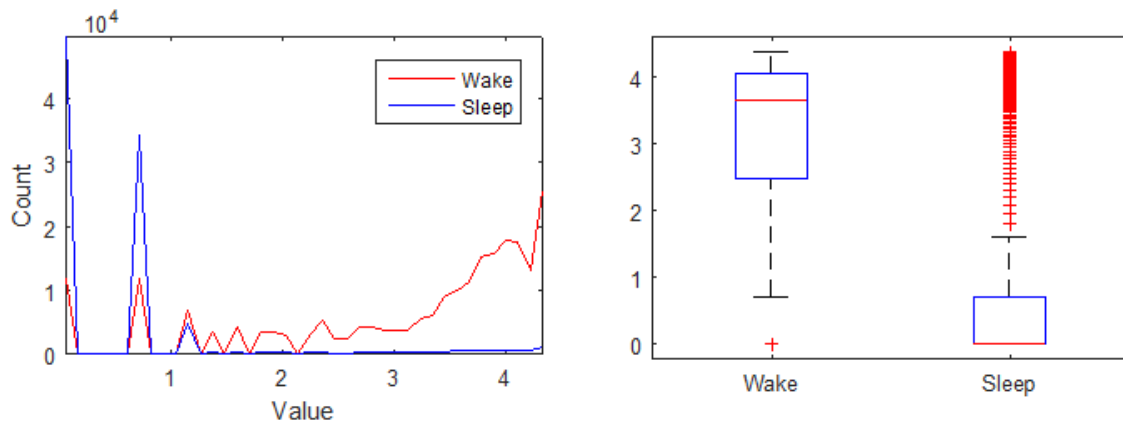

Figure 10 Histogram of sleep and wake logarithmic values on the left. Box plot of wake and sleep logarithmic values on the right.

If we use a logarithm of the actigraphy signal instead of the raw data we can see much more clearly that sleep samples have lower values and wake samples tend to have value higher, which is not surprising.

4.3   Methods

Four different existing methods were evaluated, for shorter we will call them by the name of the first author of the study. Sadeh's and Sazonov's algorithms were evaluated in two different versions, first with the original coefficients (those are marked as orig. in the result tables) and then with new coefficients optimized on our dataset. On top of that Sadeh's algorithm was originally designed for an actigraph using 1 minute epochs while devices from dataset we use have epoch of 30 seconds which was also taken into consideration. Detailed description of evaluated algorithms can be found in Chapter 2.2.1. Tilmanne's algorithms were both retrained with settings as closely as possible based on reported information. We also proposed a new algorithm using decision trees and mean activity in current 2.5 minute epoch, 8 epochs before and 8 epochs after one. This results into 42.5 minute centered window split into 17 epochs we will call this method "Mean in epoch DT"

Training and evaluation was done using leave-one-out subject cross validation described in Chapter 3.2.1 and metrics described in Chapters 3.2.2 and 3.2.3. As the outcome of some models is not binary we used a method to find the best threshold to maximaze the accuracy of the model. After thresholding the metrics were calculated for each fold of cross validation and then we took the mean value across all subsets.

First we evaluated the raw output of the models, then output after the Webster rescoring rules (described in Chapter2.2.1) were applied and lastly after using our proposed rescoring method. As described in Chapter2.2.1 Webster et al. proposed the rescoring rules with a premise that wake is more often falsely classified as sleep than the other way which does not apply in our case.

That's why we developed a new post processing methods with goal to smooth the predicted class in a way that if there are lonely short false prediction (sleep or wake) they get corrected. This approach more corresponds with the idea of looking for a long uninterrupted block of sleep and wakefulness while ignoring short awakenings during the night. This is done by applying a cascade of median filters where the length of the window is increasing. That ensures that small errors are corrected first and then these corrected values help to fix the bigger errors.
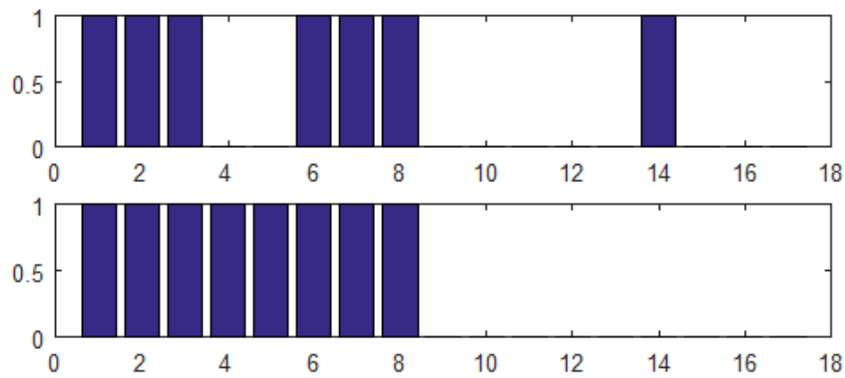
Figure 11 Upper graph shows data with possible miss classification in sleep on positions 4-5 and in wake on position 14. After using a median filter of with window length 5 the miss classifications are corrected as we can see in the secondgraph.

The structure of median filter cascade was optimized for each of the algorithms using a genetic algorithm described in Chapter 3.1.1.

4.4 Results

Figure 12 Raw output of Sadeh algorithm

Figure 13 Raw output of Sazonov algorithm

Figure 14 Raw output of Tilmanne ANN algorithm

Figure 15 Raw output of Tilmanne DT

Figure 16 Raw output of Mean in epoch DT algorithm

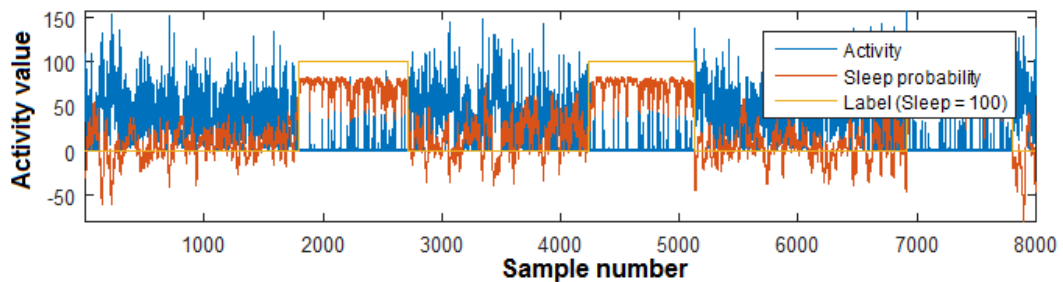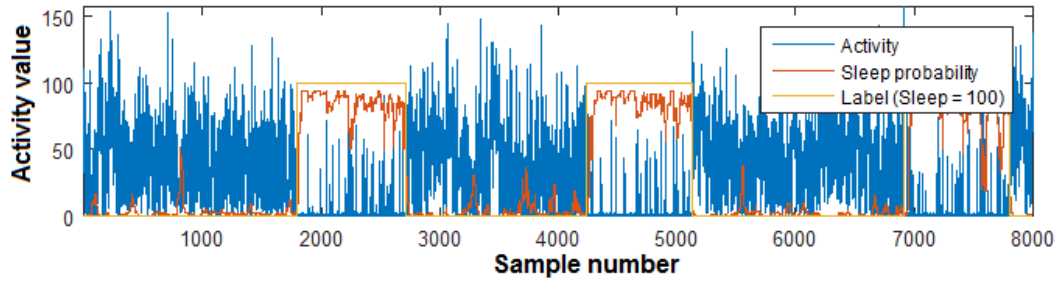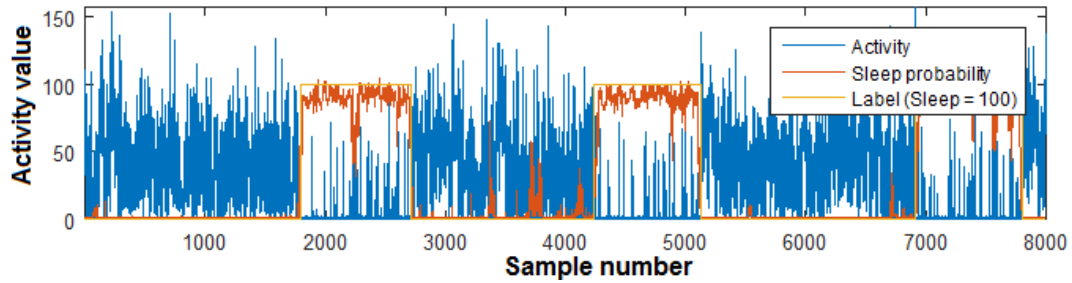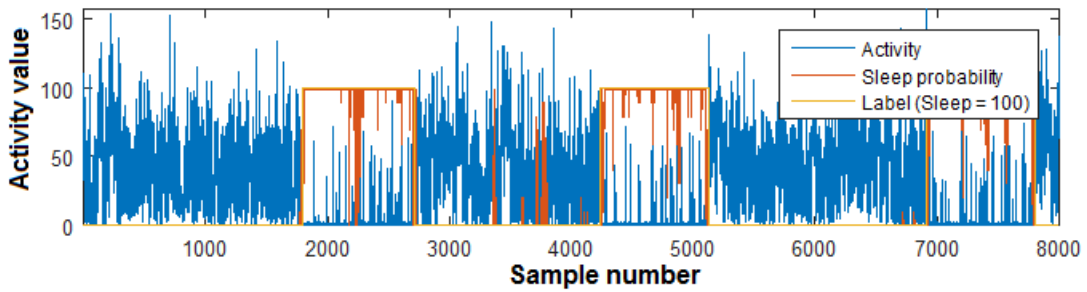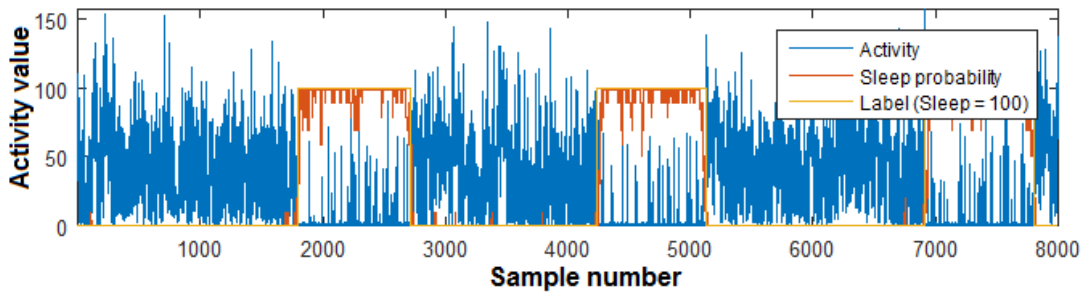| Algorithm	Acc	Sen	Spe	ASD	AWD	SD	WD
Sadeh orig.	0.9042	0.9041	0.9042	5.82	5.13	0.02	-3.18
Sazonov orig.	0.9105	0.9086	0.9093	12.89	8.42	8.16	0.8
Sadeh	0.9207	0.9216	0.9203	5.30	3.6	0.48	-1.72
Sazonov	0.9365	0.9371	0.9363	11.92	7.47	7.03	0.04
Tilmanne DT	0.9443	0.9487	0.9424	5.57	4.17	-0.12	-2.11
Tilmanne ANN	0.9485	0.9490	0.9483	7.52	8.52	1.34	-3.92
Mean in epoch DT	0.9550	0.9617	0.9521	7.69	5.88	-1.53	-1.05

Table 4 Performance of classificators without postprocessing
ACC – Accuracy (%), SEN – Sensitivity (%), SPE – Specificity (%), ASD – Absolute sleep distance (min),
AWD – Absolute wake distance (min), WD – Wake distance (min), SD – Sleep distance (min)

Algorithm	Acc	Sen	Spe	ASD	AWD	SD	WD
Sadeh orig.	0.9193	0.8865	0.9335	7.26	6.28	3.37	-4.46
Sazonov orig.	0.9212	0.9139	0.9128	12.53	8.86	8.71	0.93
Sadeh	0.9333	0.9090	0.9439	6.94	3.78	3.57	-1.92
Sazonov	0.9391	0.9302	0.9429	13.24	7.99	9.83	-0.61
Tilmanne DT	0.9512	0.9379	0.9570	8.12	5.49	2.41	-3.10
Tilmanne ANN	0.9511	0.9427	0.9548	9.81	10.18	4.65	-5.81
Mean in epoch DT	0.9586	0.9554	0.9600	9.06	6.70	0.17	-2.11

Table 5 Performance of classificators with Webster rescoring
ACC – Accuracy (%), SEN – Sensitivity (%), SPE – Specificity (%), ASD – Absolute sleep distance (min),
AWD – Absolute wake distance (min), WD – Wake distance (min), SD – Sleep distance (min)

Algorithm	Acc	Sen	Spe	ASD	AWD	SD	WD
Sadeh orig.	0.9457	0.9506	0.9436	11.24	11.97	2.66	-6.84
Sazonov orig.	0.9449	0.9413	0.9293	14.07	10.81	6.49	0.23
Sadeh	0.9473	0.9673	0.9386	12.63	10.18	-1.02	-2.64
Sazonov	0.9468	0.9470	0.9468	14.75	9.57	7.62	-0.48
Tilmanne DT	0.9561	0.9638	0.9528	11.30	10.90	-2.85	-3.28
Tilmanne ANN	0.9520	0.9517	0.9522	12.65	13.72	1.90	-8.83
Mean in epoch DT	0.9614	0.9689	0.9581	12.50	9.65	-1.94	-1.72

Table 6 Performance of classificators with median filter cascade rescoring
ACC – Accuracy (%), SEN – Sensitivity (%), SPE – Specificity (%), ASD – Absolute sleep distance (min),
AWD – Absolute wake distance (min), WD – Wake distance (min), SD – Sleep distance (min)

4.5   Discussion

As we can see from the results the current algorithms can detect the sleep very well. However they do have some problems during the day where they would detect some false sleep. That might be however caused by the fact that these algorithms were designed to detect wakefulness and sleep during the night so they are optimized that way. That can be seen especially well on Sadeh's algorithm where the sleep during night is detected well and lots of probable awakenings is identified during that period which is what it was designed for. During the day it has problems because it is trying to score even short periods of rest as sleep, which is

again reasonable considering what it was designed for. Other methods don't show that much sensitivity. As expected our proposed algorithm that is designed for identifying long periods of sleep outperforms other but the difference is not that big. Especially after rescoring using the new proposed method that removes short awakenings during sleep and short sleeps during day respectively.

5   Sleep stages classification

5.1   Problem definition

As mentioned in chapter 2.2.2 currently the only way how to get information about sleep structure and sleep stages is polysomnography in sleep laboratory. While those information are fundamental to evaluate the sleep quality the examination itself can be uncomfortable and negatively influence the sleep. On top of that it requires effort of multiple people so it cannot be done very often. Our aim is to investigate whether actigraphy couldn't provide at least some information about the sleep structure so there is an alternative that could be use outside the laboratory settings. That would allow us to monitor the quality of sleep in more natural environment and for a longer period of time.

5.2   Dataset description

The dataset was created by National Institute of Mental Health in Prague, Czech Republic as a part of study focused on sleep, aging and memory. Because the study was focused on senior citizens all the volunteers are over 55 years old. The dataset consist of simultaneous polysomnography and actigraphy recordings from one night examination from 35 volunteers. Sleep technician analyzed the polysomnography recordings and created a hypnogram based on the AASM manual for the scoring of sleep [16] that we will use as labels (ground truth).

5.2.1   Exploratory data analysis

Data set consists of three different parts: polysomnography signals, actigraphy and hypnogram. Hypnogram contains following labels: Wake, S1, S2, S3 and REM. The S4 label is missing as it was scored together with S3. Polysomnograph have 34 channeles sampled on 250 Hz, channels include EEG, EOG, ECG, leg movement sensors, oxygen level sensors and many other data that are not important for this thesis. From those we will use only ECG to extract heart rate. Activity was recorder by activity monitor GT3X with sampling frequency 300 Hz and 3-axial accelerometer.

Before we start we will have to extract HRV data from ECG. The ECG signal is unfortunately very messy so we will not describe the process in detail but just very quickly.
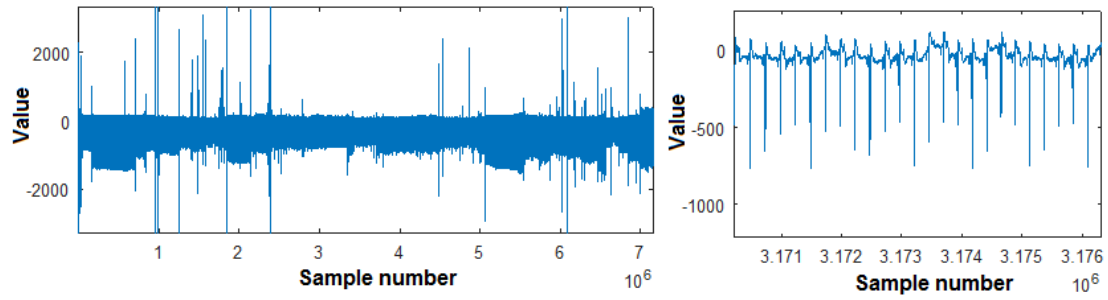
Figure 17 Example of ECG signal, whole signal on the left and detail on the right

As we can see the ECG contains artifacts is not normalized and so on. By removing low frequencies from the signal we can remove some noise (not only noise but also slower processes that influence the signal). Then we will find ouliers to remove the artifact and last we will use window to go through the signal and normalize it part by part. That will leave us with more or less clean ECG signal that we can use to extract RR intervals. With RR intervals extracted we can create HRV signal by interpolating the intervals (which we have to do because the intervals are not uniformly distributed and to extract features in frequency domain we need them to be). Now we will be able to extract features RR and HRV features as described in other studies [8][10][20][21].
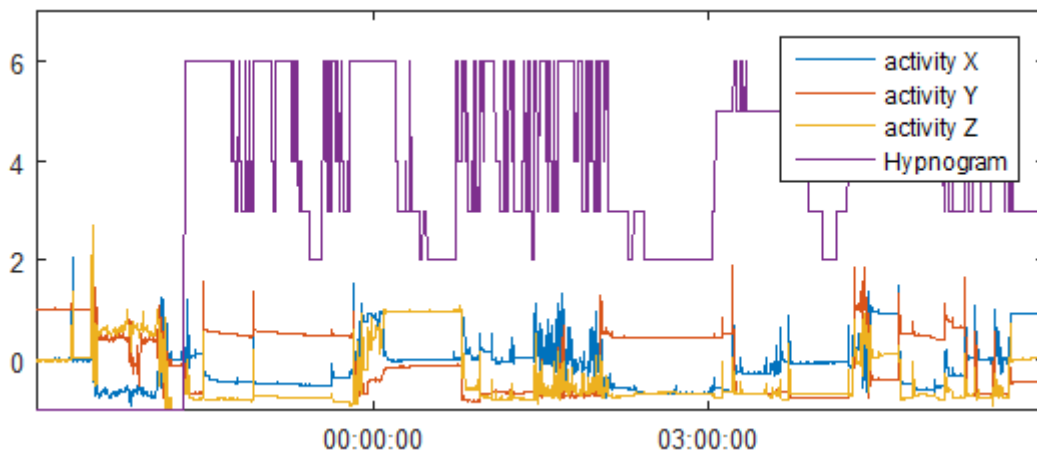
Figure 18 Example of activity and hypnogram, the values of hypnogram match the sleep stages as follow:
Wake (6), REM (5), S1 (4), S2 (3), S3 (2)
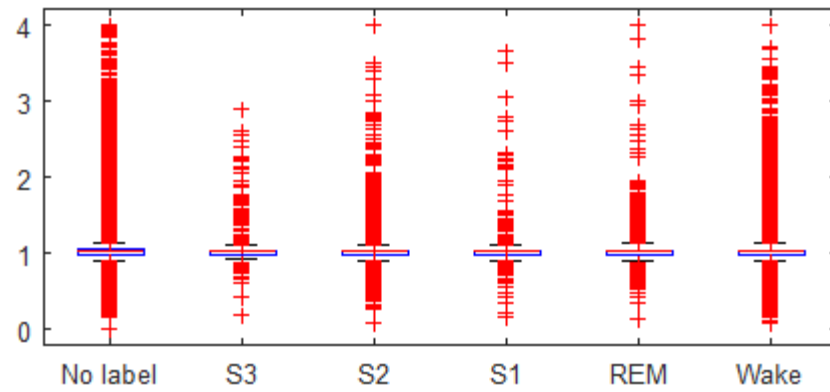
Figure 19 Boxplot of aggregated activity during specific sleep stages
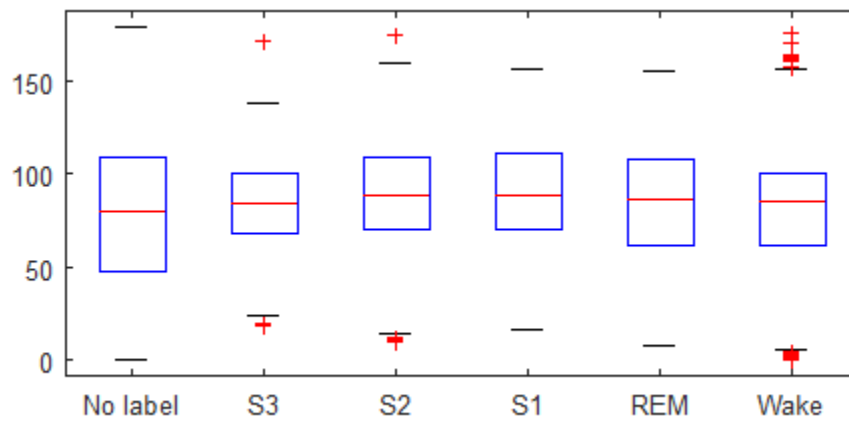
Figure 20 Boxplot of different Z inclination values during sleep stages
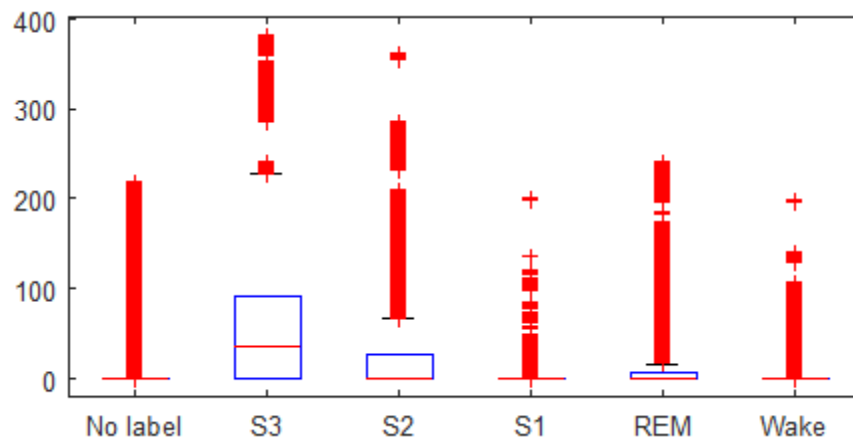
Figure 21 Boxplot of time since last significant movement for different sleep stages

The activity during different sleep stages doesn't seem to significantly vary. The only information that looks promising is the time since last significant movement. That would also agree with rules in the Polysomnography scoring manual [16]. The deep sleep seems to be after

some period without any bigger movement. That might be a way how to obtain at least some useful information about sleep structure.

5.2.2 Methods

We didn't find any existing methods using only actigraphy. All methods mentioned in Chapter 2.2.2 uses actigraphy in combination with heart rate features. We will not replicate those studies but we will try to propose an pure actigraphy algorithm and then a combination with HRV to see the results.

Using the 3-axial actigraphy gave us a chance to get some extra information about the position. So using these equations we got features representing the inclination of the devices

$$\theta_y(t) = \cos^{-1}\left(\frac{y(t)}{\sqrt{x(t)^2 + y(t)^2 + z(t)^2}}\right)$$

$$\theta_z(t) = \cos^{-1}\left(\frac{z(t)}{\sqrt{x(t)^2 + y(t)^2 + z(t)^2}}\right)$$

After that we aggregated the 3 dimensional signal to 1 dimensional so we could extract other features in more common way.

$$d(t) = \sqrt{x(t)^2 + y(t)^2 + z(t)^2}$$

Because the exploratory analysis did not show much of differences in activity during different sleep stages, we decided to calculated all features that were used in studies mentioned in this thesis [1][2][3][8][10][21] which resulted in set of 216 different actigraphy features. We did that for windows of length 6, 20, 40 and 60 getting in total 864 features.

We also extracted heart rate features used by studies in Chapter 2.2.2

As a model we decided to test linear regression, neural networks and decision trees. The last mentioned provided the best results so we will show results only from that.

5.2.3 Results

Class	Accuracy	Sensitivity	Specificity
S3	0.82446	0.41168	0.84675
S2	0.62345	0.4494	0.6834
S1	0.69749	0.06005	0.9363
REM	0.63453	0.15785	0.8626
Wake	0.7623	0.69949	0.769

Table 7 Results of actigraphy classification into 6 stages (Wake, REM, S1, S2, S3, S4)

Class	Accuracy	Sensitivity	Specificity
Deep	0.82259	0.33191	0.85978
Light	0.56999	0.44243	0.64769
REM	0.5834	0.14365	0.87235
Wake	0.74125	0.62728	0.76192

Table 8 Results of actigraphy classification into 4 stages (Wake, REM, Light sleep, Deep sleep)

Class	Accuracy	Sensitivity	Specificity
NREM	0.62301	0.74179	0.53938
REM	0.55211	0.16059	0.87332
Wake	0.74987	0.61429	0.77125

Table 9 Results of actigraphy classification into 3 stages (Wake, REM, NREM)

Class	Accuracy	Sensitivity	Specificity
S3	0. 0.83126	0.44494	0.85697
S2	0.62401	0.46782	0.6834
S1	0.74131	0.060374	0.9363
REM	0.70087	0.20039	0.8626
Wake	0.7438	0.47357	0.769

Table 10 Results of heart rate features classification into 6 stages (Wake, REM, S1, S2, S3, S4)

Class	Accuracy	Sensitivity	Specificity
S3	0.83823	0.50329	0.86626
S2	0.63349	0.47372	0.69655
S1	0.73102	0.071429	0.93984
REM	0.70495	0.18977	0.87369
Wake	0.75398	0.61547	0.77892

Table 11 Results of heart rate + actigraphy features classification into 6 stages (Wake, REM, S1, S2, S3, S4)

Model	Accuracy	Sensitivity	Specificity
Sadeh	0.7548	0.8837	0.4285
Sazonov	0.7164	0.8411	0.3790
Tilmanne DT	0.6907	0.7026	0.6499
Tilmanne ANN	0.6956	0.7228	0.6251

Table 12 Results of existing algorithms to detect sleep-wake during the night

5.2.4 Discussion

The exploratory data analysis suggested that to be able to extract some information about sleep structure it might be more important to focus on events happening during the whole night than using the traditional window limit features like during sleep-wake detection. As we can see from the results of our models while some looks promising overall the sleep stages identification doesn't work very well. No matter if we use features based on actigraphy, heart rate or both in combination. The results are similar to other methods described in Chapter 2.2.2. Further research might be needed.

6 Conclusion

The aim of this thesis was to review current methods used in actigraphy to detect sleep-wake periods and evaluate them on real data from a long term actigraphy study. We selected and implemented four different algorithms that use different types of features and models so we can get an idea how wider variety of methods could behave. Current algorithm are designed and optimized to be an alternative to polysomnography which is a slightly different task than what is being done in long term actigraphy studies. To be comparable to polysomnography they use data from one night where you can presume mostly sleep will appear and are very sensitive to even short periods of wake because that is exactly what polysomnography is monitoring. In long term actigraphy the focus is more on identifying the long periods of sleep and wakefulness rather than detecting small disturbances of sleep during the night, because the goal is to be able to monitor and analyze sleep behavior and circadian rhythms. As we could see in Chapter 4.4 that influenced the results of tested algorithms because they were much more sensitive than it was needed. The new proposed algorithm outperformed all of them in case of evaluation without any post processing. With the newly proposed rescoring method the differences were not that significant as it was designed to remove short sleep or wake periods.

In the second part of this thesis we tried to explore the possibility to identify sleep stages from actigraphy signal. Exploratory data analysis didn't show much variation of activity during different sleep stages which leads me to conclusion that activity alone is not enough to identify them. Maybe analyzing the body movement during whole night and try to identify events and transitions between different stages might be approach worth further research. Not surprisingly the proposed models did not provide very good results. Neither acticraphy nor actigraphy in combination with heart rate signal were enough to be able to accurately identify sleep stages and so the polysomnography will still be the only reliable way to monitor sleep structure.

Bibliography

[1] TILMANNE, JOËLLE, JÉRÔME URBAIN, MAYURESH V. KOTHARE, ALAIN VANDE WOUWER a SANJEEV V. KOTHARE. Algorithms for sleep-wake identification using actigraphy: a comparative study and new results. *Journal of Sleep Research* [online]. 2009, **18**(1), 85-98 [cit. 2018-08-13]. DOI: 10.1111/j.1365-2869.2008.00706.x. ISSN 09621105. Available on: http://doi.wiley.com/10.1111/j.1365-2869.2008.00706.x

[2] SAZONOVA, N.A., E.S. SAZONOV a S.A.C. SCHUCKERS. Activity-based sleep-wake identification in infants. In: *Computers in Cardiology*[online]. IEEE, 2002, s. 525-528 [cit. 2018-08-14]. DOI: 10.1109/CIC.2002.1166825. ISBN 0-7803-7735-4. Available on: http://ieeexplore.ieee.org/document/1166825/

[3] SADEH, Avi, M. SHARKEY a Mary A. CARSKADON. Activity-Based Sleep-Wake Identification: An Empirical Test of Methodological Issues. *Sleep*[online]. 1994, **17**(3), 201-207 [cit. 2018-08-14]. DOI: 10.1093/sleep/17.3.201. ISSN 1550-9109. Available on: https://academic.oup.com/sleep/article-lookup/doi/10.1093/sleep/17.3.201

[4] COLE, Roger J., Daniel F. KRIPKE, William GRUEN, Daniel J. MULLANEY a J. Christian GILLIN. Automatic Sleep/Wake Identification From Wrist Activity. *Sleep* [online]. 1992, **15**(5), 461-469 [cit. 2018-08-14]. DOI: 10.1093/sleep/15.5.461. ISSN 1550-9109. Available on: https://academic.oup.com/sleep/article-lookup/doi/10.1093/sleep/15.5.461

[5] PAQUET, Jean, Anna KAWINSKA a Julie CARRIER. Wake Detection Capacity of Actigraphy During Sleep. *Sleep* [online]. 2007, **30**(10), 1362-1369 [cit. 2018-08-14]. DOI: 10.1093/sleep/30.10.1362. ISSN 1550-9109. Available on: https://academic.oup.com/sleep/article-lookup/doi/10.1093/sleep/30.10.1362

[6] C. CRESPO, M. ABOY, F. Jr, and A. MOJON, "Algorithm for Sleep/Wake Identification From Actigraphy," in BIOSIGNAL, 2010, no. 1, pp. 224-228.

[7] WEBSTER, John B., Daniel F. KRIPKE, Sam MESSIN, Daniel J. MULLANEY a Grant WYBORNEY. An Activity-Based Sleep Monitor System for Ambulatory Use. *Sleep* [online]. 1982, **5**(4), 389-399 [cit. 2018-08-14]. DOI: 10.1093/sleep/5.4.389. ISSN 0161-8105. Available on: https://academic.oup.com/sleep/article-lookup/doi/10.1093/sleep/5.4.389

[8] HAYANO, Junichiro, Emi YUDA a Yutaka YOSHIDA. Sleep stage classification by combination of actigraphic and heart rate signals. In: *2017 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW)* [online]. IEEE, 2017, 2017, s. 387-388 [cit. 2018-08-14]. DOI: 10.1109/ICCE-China.2017.7991158. ISBN 978-1-5090-4017-9. Available on: http://ieeexplore.ieee.org/document/7991158/

[9] RENEVEY, Philippe, Ricard DELGADO-GONZALO, Alia LEMKADDEM, Martin PROENÇA, Mathieu LEMAY, Josep SOLÀ, Adrian TARNICERIU a Mattia BERTSCHI. Optical wrist-worn device for sleep monitoring. ESKOLA, Hannu, Outi VÄISÄNEN, Jari VIIK a Jari HYTTINEN, ed. *EMBEC & NBC 2017* [online]. Singapore: Springer Singapore, 2018, 2018-06-13, s. 615-618 [cit. 2018-08-14]. IFMBE Proceedings. DOI: 10.1007/978-981-10-5122-7_154. ISBN 978-981-10-5121-0. Available on: http://link.springer.com/10.1007/978-981-10-5122-7_154

[10] BEATTIE, Z, Y OYANG, A STATAN, A GHOREYSHI, A PANTELOPOULOS, A RUSSELL a C HENEGHAN. Estimation of sleep

stages in a healthy adult population from optical plethysmography and accelerometer signals. *Physiological Measurement* [online]. 2017, **38**(11), 1968-1979 [cit. 2018-08-14]. DOI: 10.1088/1361-6579/aa9047. ISSN 1361-6579. Available on: http://stacks.iop.org/0967-3334/38/i=11/a=1968?key=crossref.9d3836004c84123c742c124648581e02

[11] KAPLAN, Katherine A, Lisa S TALBOT, June GRUBER a Allison G HARVEY. Evaluating sleep in bipolar disorder: comparison between actigraphy, polysomnography, and sleep diary. *Bipolar Disorders*[online]. 2012, **14**(8), 870-879 [cit. 2018-08-13]. DOI: 10.1111/bdi.12021. ISSN 13985647. Available on: http://doi.wiley.com/10.1111/bdi.12021

[12] VAN DE WATER, ALEXANDER T. M., ALISON HOLMES a DEIRDRE A. HURLEY. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography - a systematic review. *Journal of Sleep Research* [online]. 2011, **20**(1pt2), 183-200 [cit. 2018-09-03]. DOI: 10.1111/j.1365-2869.2009.00814.x. ISSN 09621105. Dostupné z: http://doi.wiley.com/10.1111/j.1365-2869.2009.00814.x

[13] GOLD, Alexandra a Louisa SYLVIA. The role of sleep in bipolar disorder. *Nature and Science of Sleep* [online]. 2016, **8**, 207-214 [cit. 2018-08-13]. DOI: 10.2147/NSS.S85754. ISSN 1179-1608. Available on: https://www.dovepress.com/the-role-of-sleep-in-bipolar-disorder-peer-reviewed-article-NSS

[14] MANTUA, Janna, Nickolas GRAVEL a Rebecca SPENCER. Reliability of Sleep Measures from Four Personal Health Monitoring Devices Compared to Research-Based Actigraphy and Polysomnography. *Sensors* [online]. 2016, **16**(5), 646- [cit. 2018-08-13]. DOI: 10.3390/s16050646. ISSN 1424-8220. Available on: http://www.mdpi.com/1424-8220/16/5/646

[15] TRIPATHY, R.K. a U. RAJENDRA ACHARYA. Use of features from RR-time series and EEG signals for automated classification of sleep stages in deep neural network framework. *Biocybernetics and Biomedical Engineering* [online]. 2018, - [cit. 2018-08-13]. DOI: 10.1016/j.bbe.2018.05.005. ISSN 02085216. Available on: https://linkinghub.elsevier.com/retrieve/pii/S0208521618300949

[16] IBER, Conrad. *The AASM manual for the scoring of sleep and associated events: rules, terminology, and technical specifications*. Westchester, IL: American Academy of Sleep Medicine, c2007.

[17] R. ZIELINSKI, Mark, James T. MCKENNA a Robert W. MCCARLEY. Functions and Mechanisms of Sleep. *AIMS Neuroscience* [online]. 2016, **3**(1), 67-104 [cit. 2019-01-04]. DOI: 10.3934/Neuroscience.2016.1.67. ISSN 2373-7972. Available on: http://www.aimspress.com/article/10.3934/Neuroscience.2016.1.67

[18] ANCOLI-ISRAEL, Sonia, Roger COLE, Cathy ALESSI, Mark CHAMBERS, William MOORCROFT a Charles P. POLLAK. The Role of Actigraphy in the Study of Sleep and Circadian Rhythms. *Sleep* [online]. 2003, **26**(3), 342-392 [cit. 2019-01-08]. DOI: 10.1093/sleep/26.3.342. ISSN 0161-8105. Available on: https://academic.oup.com/sleep/article-lookup/doi/10.1093/sleep/26.3.342

[19] SADEH, Avi a Christine ACEBO. The role of actigraphy in sleep medicine. *Sleep Medicine Reviews* [online]. 2002, **6**(2), 113-124 [cit. 2019-01-08]. DOI: 10.1053/smrv.2001.0182. ISSN 10870792. Available on: http://linkinghub.elsevier.com/retrieve/pii/S1087079201901820

[20] JINDAL, GD, Manasi S SAWANT, Jyoti A PANDE, et al. Heart Rate Variability: Objective Assessment of Autonomic Nervous System. *MGM Journal of Medical Sciences* [online]. 2016, **3**(4), 198-205 [cit. 2019-01-08]. DOI: 10.5005/jp-journals-10036-1123. ISSN 23477946. Available on: http://www.jaypeejournals.com/eJournals/ShowText.aspx?ID=10737&Type=FREE&TYP=TOP&IN=~/eJournals/images/JPLOGO.gif&IID=825&isPDF=YES

[21] SHAFFER, Fred a J. P. GINSBERG. An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health* [online]. 2017, **5** [cit. 2019-01-08]. DOI: 10.3389/fpubh.2017.00258. ISSN 2296-2565. Available on: http://journal.frontiersin.org/article/10.3389/fpubh.2017.00258/full

Appendix A

File structure of attached CD

src/ - source codes used in this thesis

Nahlik-DP-2019-Actigraphy.pdf