Ing. Jiří Anýž
National Institute of Mental Health
Topolová 748
250 67 Klecany

Thesis advisor's review of a diploma thesis

Titile (EN): **Analysis of Sleep as Android Application Users' Sleep**
Title (CZ): **Analýza spánkových vzorů uživatelů aplikace Sleep as Android**
Author: Bc. Miroslav Domankuš
Thesis advisors: Ing. Jiří Anýž, Ing. Eduard Bakštein, Ph.D.

Bc. Miroslav Domankuš in his diploma thesis "Analysis of Sleep as Android Application Users' Sleep" provides a comprehensive analysis of the data from the application Sleep as Android. The aims of the work according to the assignment were to 1) review the approaches to analyse sleep in scientific studies and the methods for clustering of the time series data, 2) find an appropriate methods and environments to work with the Sleep as Android data and perform the exploratory data analysis, 3) find an appropriate representation of the data for the purposes of the time series clustering; 4) perform the clustering and 5) validate the results.

In the section 'Introduction', the author provides relevant information about sleep in humans. The basic knowledge of sleep as a human behaviour and its function is presented alongside with various factors affecting the parameters of sleep. The section is well referenced in scientific literature and the author references several works with similar type of data and critically reviews their results.

The section 'Methods' provides an extensive list of methods applied for the data analysis (both exploratory and time series clustering). As in the case of the section 'Introduction', the methods are the state of the art and are well referenced. The author had found appropriate means to perform the main goal of the thesis – the time series clustering. The thesis lists several possible approaches to clustering of time series data and classifies them according several criteria. The author also discuses several measures of the quality of the resulting clustering. These are very important results, because the quality of clustering and its assessment is a difficult problem, stemming from clustering being an unsupervised method without known correct solution.

In the section 'Exploratory data analysis', the data set is described by many exploratory visualizations and tables, which help to understand the data and identify the weak points of the data and subsequent analyses. Finally, the section Analysis of the Sleep as Android data set provides results of several analyses performed on the Sleep as Android data set. In the first part of this section the author shows results of testing of several hypotheses about sleep and factors that affect it. These results are interesting from the point of the validity of our data and this part of the analysis shows that the author was able to observe the same relationships in the data to the results presented in the review of scientific literature. The next part of the analysis section deals with the innovative use of Sleep as Android data to analyse the effects of global impact events of the Brexit vote and the presidential election in the US on the sleep of the application users. These results show an interesting relationship between these events and sleep. The last part of the analysis section presents the main findings of the thesis, the results of the time series clustering. The author applied several clustering methods for the time series clustering on a subset of the data and compared the results by several measures of the clustering quality. Utilizing this rigorous approach for the comparison of several approaches to clustering of time series data, the author was able to choose the best method for this particular data set. The author also interprets the clustering results and finds a relationship between the clustering and the classification of the users into evening and morning chronotype.

In summary, Bc. Miroslav Domankuš showed that he is able to independently perform a complicated analysis of big data form the application Sleep as Android in a rigorous way and

present the results in form of a diploma thesis that is concise and exactly describes the performed analyses. The author's contribution rests in providing observations that show the validity of the Sleep as Android data by comparing the author's results with the relevant scientific literature. The author also provides a rigorous framework for clustering of time series, which can be applied to further analyses on the Sleep as Android data set as well in other contexts. The author fulfilled all the aims of the thesis as stated in the assignment. I recommend the thesis for defence and give the author grade 'A – excellent'.

Questions to the author:
1) The presented approach to the time series clustering works with data of equal length, what would have to be done in order to cluster data of unequal lengths?

2) The author uses several established methods to extract features for model-based clustering. Several of the clustered time series show very specific patterns of different values for weekdays and weekends. Is any of the presented methods better at describing this pattern? Is there a better method to describe this pattern?

3) The author concludes that several of the found relationships are rather small and are not of practical use, what causes that such a relationships with miniscule effect size show as statistically significant?

Prague 15.6.2018