Exposé - Identifying the ideal length of time to record Smartphone data using Machine Learning Algorithms, in order to obtain clear clusters to predict Stress

1710601007

FH Salzburg

31st January 2020

Concept

SmartEater is an upcoming mHealth (mobile health) app, which uses smartphone data to predict stress and therefore upcoming eating crises. The app uses the smartphone to record various sensor data. This data includes background noises, time and length of phone calls, screen on time, data volume per unit of time and several other statistics. The recorded situational context data will aid in predicting stress. Through predicting eating crises, the app can provide the user with content-dependent feedback and tips with the goal to avert a craving episode.

Data mining is used to discover patterns and knowledge from data. This includes cleaning data, combining multiple sources, selecting and transforming relevant data and extracting and evaluating data patterns.(page 17, 18) Cluster Analysis is a type of machine learning algorithm known as unsupervised machine learning. It is used to divide data into classes (clusters). Each cluster contains data that is similar to each other, but dissimilar to the data allocated to other clusters. Cluster Analysis can be used to acquire knowledge on the distribution of the data, discover characteristics, detect outliers and reduce noise, or to preprocess data for other algorithms. (page 32, 362, 363, 367)

The sensor data will be recorded for different lengths of time. It is necessary to determine which time period will be most fitting to make accurate predictions in the future. In order to determine which time period is most fitting, this thesis will compare different clusters

Therefore, the goal of this thesis, is to find

This thesis will focus on declaring the most fitting length of time period to log the smartphone data, in order to gain clear clusters.

Why am I clustering data -¿ To predict similarities - similar behaviour which can then be used to create a pattern to predict stress - also to remove outliers (remove noist/clean the data) - curse of dimensionality - clusters that cause stress and clusters that don't

Outline

1. Introduction
2. Theory
   (a) UnsupervisedData Mining
   (b) Cluster Analysis
      i. Overview of Clustering Algorithms (in high dimensions ??, talk about the curse of dimensionality)
      ii. Dimensionality Reduction
      iii. ...
3. Experiment
   (a) K-Means
   (b) Hierarchical
   (c) Comparison of different lengths of time
   (d) ...
4. Conclusions

References

Albanese, Massimiliano, Angelo Chianese, Vincenzo Moscato, and Lucio Sansone. 2004. „A Formal Model for Video Shot Segmentation and its Application via Animate Vision.“ *Multimedia Tools and Applications* 24 (3): 253–272.

Bosch, Martí, Pierre Genevès, and Nabil Layaïida. 2014. „Automated refactoring for size reduction of CSS style sheets,“ 13–16. ISBN: 9781450329491. doi:10.1145/2644866.2644885.

Fried, Carrie B. 2008. „In-class laptop use and its effects on student learning.“ *Computers & Education* 50 (3): 906–914.

McConnell, Steve. 2004. *Code Complete, Second Edition.* Redmond, WA, USA: Microsoft Press. ISBN: 0735619670.

Mulloni, Alessandro, Andreas Dünser, and Dieter Schmalstieg. 2010. „Zooming Interfaces for Augmented Reality Browsers.“ In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services*, 161–170. MobileHCI '10. Lisbon, Portugal: ACM. ISBN: 978-1-60558-835-3. doi:10.1145/1851600.1851629.

Vandevoorde, David, and Nicolai M. Josuttis. 2002. *C++ Templates: The Complete Guide.* Addison-Wesley Professional, November. ISBN: 0201734842.

Schedule

- 31st January 2020 - Hand in this exposé
- February 2020 - Read papers and do research
- 24th February 2020 - Upload the final exposé onto FHSys
- March 2020 - Meet with supervisor, read literature, analyse and experiment with clustering algorithms and write a rough draft
- April 2020 - Meet with supervisor, finish the paper and print and review details
- 10th May 2020 - Submission of the bachelor thesis

Supervisor

FH-Prof. DI Dr. Simon Ginzinger, MSc