Context-Aware Stress Detection

Olivier Janssens[1], Elena Smets[2,3], Giuseppina Schiavone[4], Emmanuel Rios Velazquez[4], Femke Ongenae[1], Walter De Raedt[3], Chris Van Hoof[2,3,4] and Sofie Van Hoecke[1]

***Abstract*— Whereas research on stress detection typically uses physiological signals, we focus on context-aware stress detection. To do so, a large data set of subjects was created containing contextual information and XGBoost was used to create a hybrid stress detection approach. The resulting user lift of 4.98 % confirms the potential of context-aware stress detection.**

I. Introduction

Stress is a growing problem in our society. More than half of Americans reported personal health problems (53 %) due to stress [1]. To actively deal with stress in daily life, stress has to be measurable continuously. Therefore, research in stress detection is being done using physiological signals such as blood pressure, skin temperature, heart-rate variability and pupil diameter [2]. Nevertheless, acquiring these signals requires the use of unique and expensive sensors, or the use of wearables, such as the Apple watch or Fitbit. Additionally, a lot of readily available contextual information can also be acquired. This contextual information can potentially improve the stress-level detection.

II. Related work

Alberdi et al. presented -among others- a review on context-aware stress detection research, which uses information such as the person's location, their social interactions, smartphone interactions and social activities [2]. However, such studies often only have limited number of participants and focus on logging data for a long period of time. As most studies deal with work-related stress, weekends are not included in the data. Within this study, we focus on a large set of subjects (> 500) for whom data was captured during a short span of time (5 days, weekend incl.) to develop a hybrid stress detection approach. This hybrid stress detection approach uses both data from the individual that is being predicted for as well as data from all other subjects.

III. Methodology

The data set contains data from 524 subjects captured over a span of five days (Thursday till Monday), containing data from both workdays as well as weekends. At the start of the study, information such as age, gender, marital status, and number of children, were collected. Also, lifestyle behavior was asked at the start: whether the subjects have hobbies and exercise, whether they smoke, drink alcoholic beverages as well as dietary habits, medication usage and chronic diseases. Additionally, at the start of the day information on the experienced sleep quality was asked. Similarly, at the end of the day gastro-intestinal symptoms were recorded. Apart from the morning and evening survey, twelve times a day, the participants indicated their stress level (no stress (0) to extremely stressed (4)) -which were considered as the ground truth-, activities, and consumptions of the past hour. Finally, also the participants' accelerations, GPS locations and audio features were recorded by the smartphone.

As the resulting data set contains a mix of continuous, categorical and ordinal features and about 20.55 % of missing values, XGBoost [3] was chosen as it is the state-of-the-art extension of gradient boosted classification trees and can deal with this kind of data.

We propose a hybrid stress detection approach wherein data from all subjects is used in the training set. However, during testing, predictions are done per subject. In fact, predictions are done per sample for each individual subject, i.e. leave-one-out cross-validation, by combining all data (temporally) captured before the sample in the test set.

To deal with the imbalanced data set, weights per sample –inversely proportional to the training set's class distribution– are provided to the XGBoost algorithm to ensure that samples annotated with high stress levels have more impact during training as they are more difficult to detect. Also training samples from the subject for which a sample is currently in the test set have their weights increased 10 times.

IV. Evaluation

Due to the imbalance in the dataset, metrics such as accuracy are not used. Alternatively, user lift [4] is used as it indicates if an algorithm is making better predictions than simply guessing an individuals state. For the presented data set and XGBoost approach, the user lift is 4.98 %. Which means that our approach can predict stress 4.98 % more accurately than always predicting the majority class per subject. We can thus conclude that context information has the potential to help determine a user's stress level, and to complement the physiological-based stress-level detection.

References

[1] A. P. Association *et al.*, "Stress in america," *APA*, 2015.
[2] A. Alberdi, A. Aztiria, and A. Basarab, "Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review," vol. 59, pp. 49–75, 2016.
[3] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of KDD*. ACM, 2016, pp. 785–794.
[4] O. DeMasi, K. Kording, and B. Recht, "Meaningless comparisons lead to false optimism in medical machine learning," *PloS one*, vol. 12, no. 9, 2017.

[1]IDLab, Ghent University - imec, Technologiepark-Zwijnaarde 19, B-9052 Ghent, Belgium olivier.janssens@imec.be
[2]Electrical Engineering-ESAT, University of Leuven, Belgium
[3]imec, Leuven, Belgium
[4]imec, Holst Centre, The Netherlands