Real-Time Arousal Detection Based on Skin Conductance and Heart Rate Features

Can Bikcora[1], Denise Klep, Bart Monhemius, Lars van Alem, Viktor Shopov, Denisa Apostol, Reon Smits

[1]corresponding author: can.bikcora@mentechinnovation.eu

January 2020

***Abstract*—From both theory and implementation perspectives, this work combines the most essential features of the electrodermal activity (EDA) and heart rate (HR) in a correct way to arrive at a reasonable arousal detection framework that would work in real-time, with the EDA features (particularly the phasic ones) dominating the selection mechanism as imposed by the theory. This detection framework is further to be restricted to stress detection by eliminating arousal due to physical activities, as well as due to other emotions. The overall results indicate that using the EDA features alone for arousal detection leads to a performance improvement in the range 10-15% when compared to the use of HR features alone (such as 75% for the HR and 88% for the EDA features alone). When they are combined together, an additional increase around 2.5% was encountered, which is unlikely to go over 3-4% because they often detect the same intervals that they are able to detect individually.**

I. Introduction

People suffering from mental disability or dementia often face difficulties in expressing their emotions, which consequently creates challenges for the caregivers to provide appropriate care. To alleviate such problems, sensor-based emotion recognition solutions can contribute to a better understanding of emotions, and therefore enable better care and increased quality of life and happiness. Mentech Innovation aims at such a solution via wearable sensors measuring various physiological signals such as skin conductance, heart rate, skin temperature, etc. Limiting the scope of emotions initially to arousal detection, we elaborate on the first steps required for a reliable stress detection algorithm.

The paper is structured as follows. Section II presents details on the experimental design. Then, Section III covers the essential steps in model development, followed by Section IV that discusses the main findings. Conclusions are provided in Section V.

II. Experimental Design

With the aim of arousal detection under various kinds of situations (along the path of stress detection), this experiment aims to combine arguably the most commonly used and well-established physiological stressors (stimuli). To limit the total experiment duration to 1-1.5 hour, all sessions are carried out while sitting on a chair since this is the most frequent physical posture in daily lives. A model for this posture can therefore be thought as a baseline from which extensions to the other postures and activities can be generated. The used stressors include a math session where a test subject is asked to subtract a two digit number from a four digit one, followed by some additions; two audio sessions in which positive (e.g., nature sounds) and negative (e.g., mechanical noise) sounds, respectively, are played via a headset; a cold pressor test where ice is used as a pain stimulus that acts similarly to a stressor; a typical stroop color test as another cognitive stressor; and several maximal expiration (exhalation) sessions with differing breathing frequencies to mimic stressful situations. Another stressor that is not yet included here would be a fearful game (or an interactive experience) with a virtual reality device. During an experiment, each stressor is followed by a relaxation session, where all sessions vary between 3-5 minutes except for the individual expiration tasks that take 1 minute each. An illustrative plot of how the skin conductance (EDA) changes through different sessions is provided in Fig. 1. Note here that the highlighted calm sessions correspond to the selected portion that is specifically used for model development. Despite relatively long speaking is unlikely to be encountered at care-house residents, it could also be added to the list of activities to distinguish intervals of speaking-induced arousal. Similar to Fig. 1, a desired HR measure-ment is given in Fig. 2.

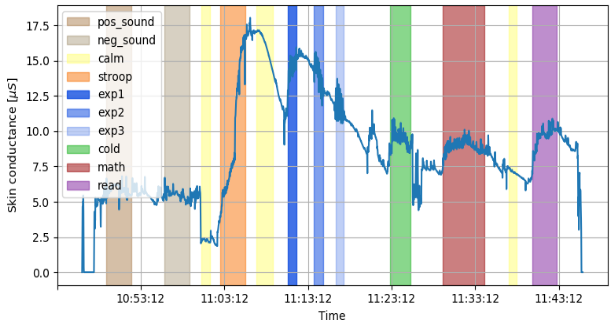

Fig 1. Typical plot of an EDA during various stressor and relaxation sessions.

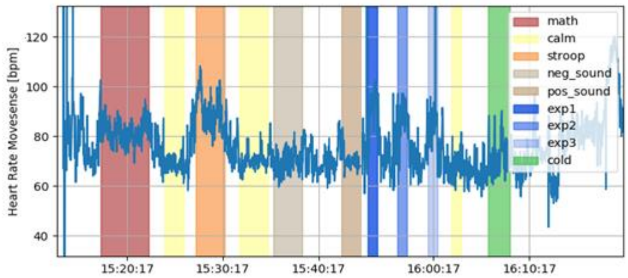

Fig 2. Typical plot of a HR during various stressor and relaxation sessions.

[1] Mentech Innovation, Hurksestraat 19, 5652 AH Eindhoven, The Netherlands

Data labeling: As could be viewed from the highlighted calm sessions in Fig. 1, data labeling is done manually by inspection of distinct characteristics of arousal, such as presence of phasic peaks, steep rise of the EDA (depending on the context), and an increase of the heart rate (decrease of its variability). The stressor sessions exhibiting some of these patterns are labeled as stress to be used in model development. Calm sessions are also included or excluded from the modeling data in a similar manner. If all the intended labels of stress and calm sessions or individual feedbacks from the participants would be taken as the ground truth, this would lead to loss in model's performance. Therefore, proper removal or relabeling of sessions that were not effective is crucial. As an alternative to manual labeling, one can also label stress and calm intervals by using an indicator that is suited more for offline analysis (such as the LF/HF ratio of the heart rate [1], i.e., the ratio of the power in low frequencies to those in high frequencies).

III. Preliminaries On Model Development

A. Data Preparation

After the manual labeling of the calm and stress intervals, the sampling frequency for data extraction is reduced to 1/3 Hz (once per 3 seconds) in order to run the training algorithms faster. The imbalanced dataset with less calm intervals is also turned into a balanced one via the naive random over-sampling (random sampling with replacement) method of the imbalanced-learn [2] library. By default, results are presented with respect to the measured 5 test subjects, but when robustness with respect to heart rate variety is analyzed, 10 more test subjects are artificially (synthetically) added to this list by realistically adding some variety to the existing heart rate signals.

B. Evaluated Classifiers

In general, it is a good idea to not to be limited by a single type of classifier as each one has its own strengths and weaknesses that may become relevant for certain problems. In addition to the default method of the support vector classifier (SVC), we took XGBoost (eXtreme Gradient Boosting) [3] as an alternative. As one of the best in data science competitions, this solver is highly popular and is an implementation of gradient boosted decision trees, where boosting is an ensemble learning technique of building many models sequentially. Despite the fact that nonlinearities are not complicated for our case (as it would have led XGBoost to perform much better than a traditional SVC), the main preference of XGBoost here stems from the fact that it is optimized for speed and performance whereas SVC suffers from scalability and memory issues, making it relatively harder to tune.

C. Performance Evaluation Framework

The train-validation-test splits are carried out via the procedure of nested cross-validation (CV) [4]. In this manner, an inner CV is carried out within an outer CV where the inner loop is used for hyperparameter tuning, based on which the performance is computed on the test set of the outer loop. As a demonstration with 5 people, each chosen as different groups, the [train-validation]-[test] splits (fold) of the outer CV are in the form: [2,3,4,5]-[1], [1,3,4,5]-[2], [1,2,4,5]-[3], [1,2,3,5]-[4], [1,2,3,4]-[5]. Moreover, each fold individually has its own inner CV, such as [2,3,4,5] is separated into [train]-[validation] sets like [3,4,5]-[2], [2,4,5]-[3], [2,3,5]-[4], [2,3,4]-[5]. Since two CVs may take considerable time with large size of data, the test set can also be strictly separated from the train-validation set beforehand which reduces the whole procedure to a single CV (that has not been carried out here).

The nested CV evaluates each outer fold with respect to its unique model derived from the corresponding inner CV where the performance of each outer fold is averaged to arrive at a general measure that estimates an unbiased generalization performance of a model to be developed. Therefore, as the final step, a model is developed by fitting the whole data to the hyperparameters obtained from a single CV. As explained by this procedure, the nested CV is only for arriving at a fair performance measure of a model-to-be-developed whereas the resultant model is built separately from this nested procedure, via another CV.

D. Feature Description and Selection

For real-time analysis, features that are suitable for short window lengths are preferred. At this moment window lengths up to 20 seconds are tested for the HR and EDA features. Therefore, low-frequency components such as LF power that would require minutes of data are omitted for the time being. Traditionally used statistical features are used for both the HR and EDA.

On normalization, all HR features are typically normalized by the corresponding factors of the resting HR or IBI values. Regarding the EDA, since the absolute value of skin conductance varies due to humidity, sweating, individual genetics, etc., some kind of normalization is necessary. Standardization or division via an average magnitude are typical routes that are adopted.

Note that feature combinations are tried manually for now rather than via selection mechanisms (e.g., recursive elimination or more advanced one of sequential floating forward search [5]) since it is not currently essential for a good performance. As the number and complexity of features will increase, then such algorithms will become handier. Furthermore, for hyperparameter tuning, a grid search involving all combinations of C = [0.1, 1, 10, 100, 1000] and gamma = [0.01, 0.1, 1, 10, 100] is used for the SVC, whereas combinations of n_estimators: [5, 10, 50, 100 (for higher dimensional features/data)], max_depth: [3, 5, 7, 9], min_child_weight: [1, 3, 5,…15], gamma: [0.0, 0.1, 0.2, 0.3, 0.5, 1] are tried individually for the XGBoost that does not need any scaler in advance.

IV. Discussion of the Achieved Results

The corresponding results are presented altogether in Table 1, noting again that 10 of 15 test subjects are generated synthetically from the existing 5 real measurements. From this table, the dominance of EDA features is clear, as imposed by the theory, and the gain of combining EDA and HR features is around 2.5%, which is

not expected to go beyond 3-4% with full feature-space search or with more test subjects.

# Test subjects	EDA features	HR features	EDA + HR features
5	88.2	75.6	90.7
15	88.2	72.7	90.4

Table 1: Overall best accuracy results with individual and combined features.

Since a nested CV produces many models to compute an average generalized performance, a final model is obtained by selecting the final hyperparameters through a single CV and then fitting the whole data under this setting. Fig. 3 plots the feature importances produced by a final model where the numbers add up to 1 and the contribution of EDA features to the model is around 75%, indicating that they are 3 times more important than the HR features. Note that individual feature importances within the EDA and HR domains may change from one model to the other, but the fact that EDA features would be similarly or even more performance-wise superior will stay unchanged. Further-more, any contributing influence of the skin temperature that would generalize to different contexts has not yet been observed.

Despite the fact that XGBoost was used as the first model, this does not imply that the SVC is worse for the current data. XGBoost was used more, only due to its relatively more user-friendly tuning especially with respect to its speed.

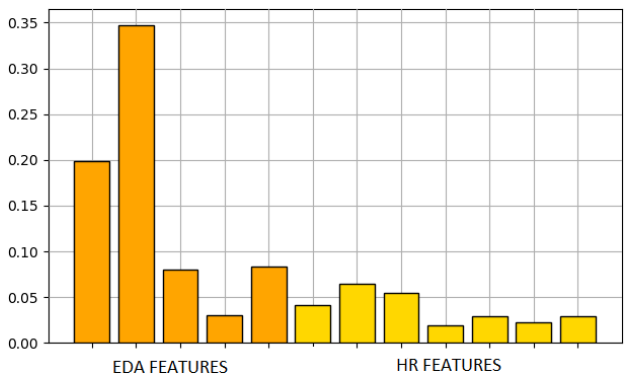

Fig. 3: Feature importances from a typical final fitted model.

V. Conclusions

The designed stress detection experiment combining various stressors while sitting was found to work well, with extra additions and modifications along the way to make it more complete, which enabled the first steps for proper modeling. The achieved results indicated the dominance of EDA features over the HR features, as expected by the theory. Moreover, a performance difference in the range 10-15% was observed when switched from HR to EDA features. When combined together, an additional performance increase around 2.5% was obtained. The superiority of the EDA stems from the fact that it is more suited for real-time detection where short windows are sufficient to capture the phasic peaks. On the other hand, most of the traditional HR features start from minimally 1-minute length windows, thereby making it more suitable for offline analysis in comparison to the EDA. Unless the stress interval is sufficiently long, in real-time, only a limited number of HR features can be efficiently used, as applied in this work. Since the stress intervals are long in the collected data, longer window lengths and relevant features could also be tried, which can be thought of as an extension.

Given the individual and combined performances, a consequent question could be that in the presence of the EDA measurements, where HR features may impact the most. Since EDA is typically associated with arousal, HR features play a role in valence, physical activity, and more detailed emotion classification that will become useful as models develop further. If a biometric identification of individuals will also be necessary, then the ECG data is indispensable.

Moreover, the performance metrics from the collected data should only be interpreted within its own scope. This scope for the current experiment is limited to stressor and relaxation sessions while sitting. Since these accuracies will only partially translate into real-life (where there are no restrictions in terms of stimuli and movement), it is more logical to interpret all performance metrics relatively when comparing various scenarios, models, and features.

Contact

For further information, please contact us:

Address: Hurksestraat 19, 5652 AH Eindhoven, The Netherlands

Email: info@mentechinnovation.eu

Tel: +31 6 28093518

Website: https://www.mentechinnovation.eu/

References

[1] V. Alexandratos, M. Bulut, R. Jasinschi, Mobile real-time arousal detection, Proc. of ICASSP, pp. 4427- 4431, 2014.

[2] https://imbalanced-learn.readthedocs.io/en/stable/index.html, Imbalanced-learn.

[3] XGBoost Documentation, https://xgboost.readthedocs.io/en/latest/

[4] Nested cross validation explained, https://weina.me/nested-cross-validation/

[5] R. Picard, E. Vyzas, and J. Healey, Toward machine emotional intelligence, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 10, pp. 1175–1191, 2001.