Emotion Analysis using Heart Rate Data

Luis Alberto Barradas Chacon, Artem Fedoskin, Ekaterina Shcheglakova, Sutthida Neamsup and Ahmed Rashed

Information Systems and Machine Learning Lab, University of Hildesheim, Germany

{barradas, fedoskin, shchegla, neamsup}@uni-hildesheim.de, ahmedrashed@ismll.de

Abstract. This paper describes the attempt to classify human emotions without including any sensors measuring biosignals, using the DEAP dataset. Basing our research on the original paper [Koelstra et al, 2012], we claim that emotions (Valence and Arousal scores) can be classified with only one pulse detecting sensor with a comparable result to the classification based on the EEG signal. Moreover, we propose the method to classify emotions avoiding any sensors by extracting the pulse of the person from the video based on the head movements. Using Lucas-Kanade algorithm for optical flow, we extract movement signals, filter them, extract principal components, recreate HR signal and then use in the emotion classification. First part of the project was conducted on the dataset containing 32 participants' HR values (1280 activation cases), while the second part was based on the frontal videos of the participants (874 videos). Results support the idea of one-sensor emotion classification and deny the possibility of zero-sensor classification with proposed method.

Keywords: Emotional Classification, DEAP Dataset, Heartrate Estimation, Biosignals.

1 Introduction

Emotional assessment is a common psychology problem where the emotional state of a person is to be deduced. For this, there are two major techniques: a verbal assessment, where the subject must respond to specific questions about their emotional state, and biosignal inference, where the emotional state of the subject is deduced from their physiological responses to stimuli. The second method is usually more reliable when the stimuli presents a heavy cognitive workload to the subject.

When inferring emotional states, the most common biosignals to be used are:

- Blood Volume Pulse (BVP): an optical device measures the amount of blood flowing through an extremity, and records changes in it. With this, the Heart Rate (HR), and other heart related variables, such as the Heart Rate Variability can be inferred.
- Electroencephalography (EEG): Measuring the electrical activity that emanates from the skull can be the most direct way to measure cognitive and emotional changes, but EEC signals are complex and require much processing.

Like with any assessment task, a model of the desired states must be selected. Emotional states vary greatly: Ekman's emotional model [Ekman, 1999] presents six human emotions shared across all cultures through facial expressions. Plutchik's wheel of emotion presents eight common responses to stimuli. One of the most simple models: the affect-activation model is commonly used in text analysis. It allows representing an emotional state in a cartesian plane, where one axis represents the level of activation, and the other the pleasure-ness of such stimuli. The last model was used in this experiment.

In this paper we try to improve over the results from [Koelstra et al, 2012], in which authors used EEG data from DEAP dataset to predict low or high Valence/Arousal scores.

Section 2 presents related work. Section 3 presents methodology. Section 4 presents results. Finally, Section 5 concludes our paper.

2 Related work

One way to study human behavior and emotion recognition, a quantitative way, is to measure various biometric parameters associated with an emotional state or a stress response to the given stimulus. After measuring biosignals such as EEG, heart and breath rates, skin conductance etc. those measures can be used for analysis through Machine Learning, which can help to find hidden insights in data and make it applicable to real-life situations.

In 2012 the group of researchers introduced the DEAP [Koelstra et al, 2012] dataset, which became the emotional classification benchmark. Using EEG signals for the Gaussian Naive Bayes classifier, authors were able to classify arousal and valence ratings with the following accuracy: 62% and 57,6% accordingly. After the release of this work, researchers tried to use different algorithms for emotional recognition to improve the results of the DEAP dataset paper. Bayesian Classification methods for classifying Valence and Arousal from the DEAP dataset into two classes – high and low – were introduced in 2012 and achieved 66.6% and 66.4% accuracy correspondingly [Chung and Yoon, 2012]. In their work, [Hosseini et al, 2010] raise an electroencephalogram (EEG) method in measuring the Biosignal in order to classify emotional stress. He introduces the visual images-based acquisition protocol for recording the EEG together with psychophysiological signals under two categories of emotional stress states, then classify with SVM classifier algorithm. Therefore, the result in this paper suggest that using Support Vector Machine classifier also gives better accuracy to classify categories of states compare among to previous research out in the field. The use of Support Vector Machines as a classifier for the EEG data for the real-time emotion recognition using Higuchi Fractal Dimension Spectrum was introduced in 2014 and produced 53,7% accuracy for classification of 8 emotions [Liu and Sourina, 2014]. In 2015 the effect of the window size on the emotion classification on EEG data using wavelet entropy and SVMs was studied [Candra et al, 2015]. Too short window will not allow the information about emotions to be extracted fully, while too wide window will lead to the information overload. As a result, the following accuracies were received after the experiments for Valence and Arousal scales: 65,13% with 312 seconds window and 65,33% with 310 seconds respectively. In one of the most recent works the Deep and Convolutional Neural Networks approaches were tested and brought the best for the current moment results for the emotional classification: 81,4% for Valence and 73,4% for Arousal using CNN [Tripathi et al, 2017].

However, there also exists the problem of collecting such type of data like EEG for the following analysis. To do so you need to cover a person with different sensors, staying still without any movements, everything should be under strict control. To prevent this, we claim that there is no need in those sensors and with only pulse detector you can classify emotions like with the EEG signal. Moreover, we suggest the method to classify emotions without any sensor based on the video of the person.

Three signals (heart rate, breath rate and skin conductance) are suggested to be linked with physiological activation [Choi, 2010; Lazarus, 1963] along with the heart rate variability (HRV), which is not the pulse, but the heart rate over time, which is also linked to stress [Mohan, 2016]. The work of [Balakrishnan et al, 2013], on which our zero-sensor classification approach is based, presents the novel approach of cardiac monitoring without actual contact-measuring means. The research suggests using the video as then capture the motion of the movement features of the blood flow from the heart to the head frequencies and extract the features by using principal component analysis (PCA) to decompose the trajectories into a set of independent source signals.

3 Methodology

3.1 Dataset Description

In this paper the DEAP dataset is used. It can be separated into two parts: the recorded signals and self-assessment of the participants. Firstly, the electroencephalogram (EEG) and peripheral physiological signals (Pulse, Breath, GSR, Temperature, BVP) were recorded for the 32 participants, which were watching 40 one-minute music videos with most intensive emotional parts (in total 1280 emotional activation cases). For 22 participants also the frontal videos were recorded, on which our zero-sensor classification method is built. Secondly, after each video participants rated the videos based on arousal, valence and dominance scales. Each video has subjective rating of the levels 0-9 of arousal and valence from each participant, which is used as the ground truth for classification in current paper. To make it comparable to the original paper, the scale was also separated into two classes: high (5-9) and low (0-4) for both valence and arousal scales.

Our primary goal was to predict excitement in terms of arousal and valence scales of the participant. Our first approach was to use HR data produced by the sensor (given in dataset) and the second one was to calculate HR of participant using facial recording. Due to the number of participants frontal recordings provided by dataset, 1280 samples for the first approach and 874 activation cases for the second one were used.

3.2 One-sensor Classification

We used heart rate data that was collected using plethysmograph, which was later filtered with Butterworth filter to remove noise. To accomplish that, Scikit-Learn library [Pedregosa et al, 2011] was used to retrieve peak-to-peak measurement. Afterwards, we extracted Beats Per Minute and Heart Rate Variability from filtered heart rate data that will be used as features in our models. 70% of the data was used for train and the remaining 30% served as test set.

We used several models to test our hypothesis, namely - Support Vector Machines, Recurrent Neural Network and Feedforward Neural Network.

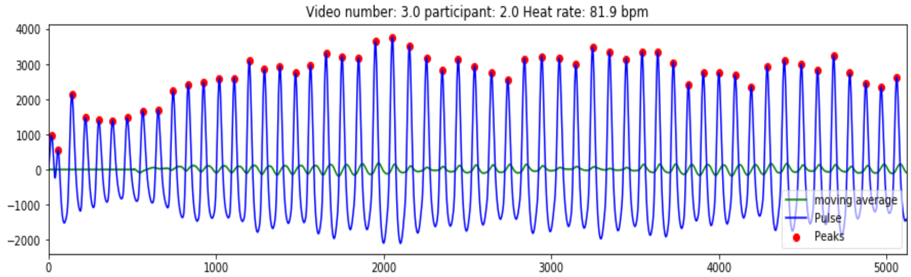


Fig. 1. Plot of HR data sample (from video number 3 and participant number 2)

Our models used several features that were extracted from Heart Rate data:

- Beats Per Minute (BPM) - the number of heart beats per minute.
- Heart rate variability (HRV) - difference between two adjacent peaks measured in milliseconds.
- Pulse was calculated as BPM per every second of the trial

Our first attempt to classify emotional state based on heart rate done used SVM as a classification model. It is well-known model that is famous for its robustness and simplicity thus we used this model only with 2 features - HRV and BPM.

The second model we used was Feedforward Neural Network that had 4 hidden layers each followed by ReLU. This model was designed to handle complex relations as apart from HRV and BPM we used Pulse as a feature. This method proved to be the best among all the models tested.

Our last attempt was done using Recurrent Neural Network that had 4 hidden layers with 100 hidden units in each layer. We used only one feature, namely BPM for 45 seconds of each trial.

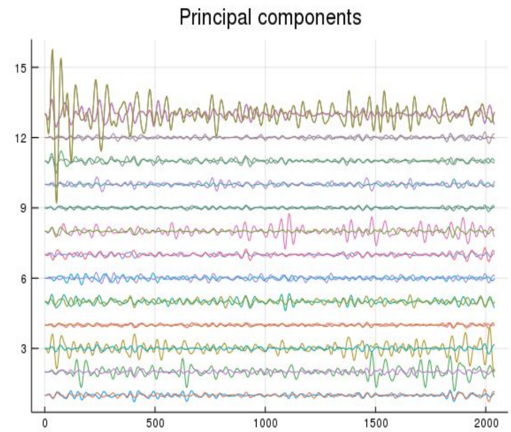
Finally, all experiments with Neural Networks were done with PyTorch and GeForce GTX 1050Ti was used for training the models.

3.3 **Zero-sensor Classification**

In zero-sensor approach our method was to classify participants emotions based solely on the facial video. In order to do that we used HR extraction technique [Balakrishnan et al, 2013] that (1) uses Lucas-Kanade (LK) algorithm to detect optical flow, (2) extracts movement signals, (3) filters them, (4) extracts principal components, (5) recreates HR signals from them.

To accelerate the processing of the LK image flow algorithm, interest points were first selected inside the area of the image that contained the face of the participant. This was done through a fast-corners feature point detection algorithm. This interest points were later the input to a LK algorithm with a 15x15 pixel window, meaning that the flow of the video was only detected in the nearest region. The result were time signals that contained the optical flow inside that image window along the video. These signals were processed through a temporal filter that removed frequencies that do not correspond to heart rate, and later decomposed through a Principal Component Analysis (PCA). The components most similar to a heart rate were then selected to recreate a heart rate signal.

The reconstruction of a heart rate signal takes about 12 minutes for every minute of video, making it unlikely to be used in real time.



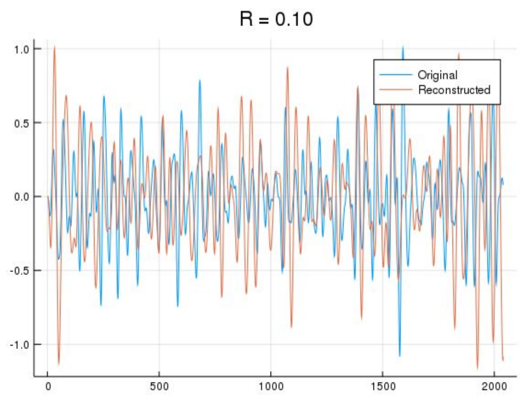


Fig. 2. Up - The plot of all principal components found in the facial video. Down - The original signal and the reconstructed one. R - correlation coefficient

One of the problems we encountered was the choice of principle components; since every interest point was eventually analyzed as a time signal, we could obtain a component for every interest point. We could not select the most similar one to a heartbeat automatically. As a result, the recreated signal, although it had the same or very similar frequency as a heart rate, rarely correlated with the original one. Nevertheless, we decided to try classifying emotions based on the raw reconstructed signal. This signal was used as a single feature for Recurrent Neural Network model, which, as it is shown in results section, did not perform well.

4 Results

DEAP dataset provides Valence and Arousal values for each participant on the 1 to 9 scale. In order to simplify our trials and follow the same procedure as baseline paper did, we converted Valence and Arousal values to binary values with Valence/Arousal of 1-5 representing low values and the remaining ones are labeled as high.

Table 1. Class Balance

Class Balance	Valence	Arousal
High	43%	42%
Low	57%	58%

As it is evident from the results in Table 2 our model performed better than [Koelstra et al, 2012] proving our hypothesis that one can classify Valence and Arousal purely using heart rate data available and get results better than the ones obtained with EEG data.

The combination of pulse signals and HRV with BPM made FNN perform the best in classification of Valence. There is 3% improvement over baseline case given by Gaussian Naive Bayes, which used EEG signals. This shows the capability of using pulse rate in measuring the valence and arousal. Moreover, for Arousal activation, Support Vector Machine classification method produced the best results. Overall, those two models performed better than Recurrent Neural Network classification model.

RNN trained with pulse did not perform the worst with average results, showing that Pulse is a significant feature, but it can perform much better in conjunction with HRV and BPM, which is proven by results of FNN.

Finally, RNN trained on signals obtained within zero-sensor approach did not perform as good as expected because of the problems within the method of HR extraction.

Table 2. Accuracy for executed models (red color denotes best results of our models)

Classification Model	Valence	Arousal
Gaussian Naive Bayes (EEG) [Koelstra et al, 2012]	57.6%	62.0%
SVM (HRV and BPM)	55.5%	*61.6%*
RNN (Pulse)	59.6%	58.5%
FNN (HRV, BPM, Pulse)	*60.6%*	57.0%
RNN (Signal from Video)	57.0%	58.0%

5 Conclusion

Our results show that emotion classification done solely with heart rate data can perform better than the one based on EEG. However, the method that we initially wanted to propose for zero-sensor testing did not meet our expectations as we encountered several problems while implementing and testing it.

Based on our results we see several directions in our future work:

1. Use color differences in the red channel of the video, that has already been tested for the Eckman model.
2. Use infrared camera to enhance biosignals that we are using.
3. Use a CRNN to automatically learn the pulse signal from the video and predict sentiment scores in a multitasking model. This method promises to learn all hyperparameters for previous tasks automatically and poses the best course of action for solving this problem.

References

1. Balakrishnan, G., Durand, F., & Guttag, J. Detecting pulse from head motions in video. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3430-3437 (2013).

2. Candra, H., Yuwono, M., Chai, R., Handojoseno, A., Elamvazuthi, I., Nguyen, H. T., & Su, S. Investigation of window size in classification of EEG-emotion signal with wavelet entropy and support vector machine. In 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 7250-7253. IEEE (2015).
3. Choi, J. & Gutierrez-Osuna, R. Estimating mental stress using a wearable cardio-respiratory sensor. In Sensors, 2010 IEEE, pp. 150-154. IEEE (2010).
4. Chung, S. Y., & Yoon, H. J. Affective classification using Bayesian classifier and supervised learning. In 2012 12th International Conference on Control, Automation and Systems, pp. 1768-1771. IEEE (2012).
5. DEAPdataset, http://www.eecs.qmul.ac.uk/mmv/datasets/deap/, last accessed 2018/04/03
6. Ekman, P. Basic emotions. Handbook of cognition and emotion, 45-60 (1999).
7. Hosseini, S.A., Khalilzadeh, M.A. & Changiz, S. Emotional stress recognition system for affective computing based on bio-signals. Journal of Biological Systems, 18(spec01), pp.101-114 (2010).
8. Kanade, T., Tian, Y., & Cohn, J. F. Comprehensive database for facial expression analysis. In fg(p. 46). IEEE (2000).
9. Koelstra, S., Muhl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., ... & Patras, I. Deap: A database for emotion analysis; using physiological signals. IEEE transactions on affective computing, 3(1), 18-31 (2012).
10. Lazarus, R.S., Speisman, J.C. & Mordkoff, A.M. The relationship between autonomic indicators of psychological stress: Heart rate and skin conductance. Psychosomatic medicine (1963).
11. Liu, Y., & Sourina, O. EEG-based subject-dependent emotion recognition algorithm using fractal dimension. In 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 3166-3171. IEEE (2014).
12. Mohan, P. M., Nagarajan, V., & Das, S. R. Stress measurement from wearable photoplethysmographic sensor using heart rate variability data. In 2016 International Conference on Communication and Signal Processing (ICCSP), pp. 1141-1144. IEEE (2016)
13. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. Scikit-learn: Machine learning in Python. Journal of machine learning research, 12(Oct), 2825-2830 (2011).
14. Tripathi, S., Acharya, S., Sharma, R. D., Mittal, S., & Bhattacharya, S. Using Deep and Convolutional Neural Networks for Accurate Emotion Classification on DEAP Dataset. In Twenty-Ninth IAAI Conference (2017).