Physiological signals in affective computing

Pietro Bonazzi
pbonazzi@student.ethz.ch
ETH Zürich (17-200-635)
Zürich, Zürich, Switzerland

ABSTRACT

In this paper, the predictive power of several physiological signals in the domain of affective computing is evaluated. Between electro-cardiograms (ECG), electroencephalograms (EEG), electro-dermalactivities (EDA) and facial landmark trajectories (EMO) , EMO and EEG are the most promising physiological signals to predict emotional state. A random forest classifier has been tested to predict emotional states (in the arousal-valence dimensions) with a leave-one-participant-out validation scheme and a leave-one-clip-out validation scheme.

CCS CONCEPTS

• **Applied computing**; • **Law, social and behavioral sciences**; • **Psychology**; • **Affective computing**;

KEYWORDS

ubiquitous computing, affecting computing, ecg, eda, emo, eeg

ACM Reference Format:
Pietro Bonazzi. 2021. Physiological signals in affective computing. In *Proceedings of Ubiquitous Computing Class 2021: Exercise 2*. ETH Zürich, Zurich, Switzerland, 5 pages.

DISCLAIMER

This paper has been written to report on exercise 2 of the Ubiquitous Computing class. It has been submitted together with a code report. Solutions to the exercises are shared among the two files.

1 INTRODUCTION

Affective computing is a vibrant field of research in ubiquitous computing. If computers can efficiently predict emotions, then people interaction with them will be less alienating, more healthy and productive. In this paper, a series of features will be extracted from electrocardiograms, electroencephalograms, electrodermal and facial landmark trajectories' signals of 44 participants as they watched different affective movie clips from the ASCERTAIN dataset. Next, a random forest classifier (RFC) is evaluated against a baseline classifier to test the capabilities of the former in predicting emotions from physiological signals. Discussions over the obtained results are included in the last section.

Ubiquitous Computing Class 2021: Exercise 2, May 31, 2021, Zurich, CH

2 FEATURE EXTRACTION

In this section, the decisions made in the process of feature extraction will be specified. In accordance with the specifications of the exercise, only the last 50 seconds of each recording was retained for statistical measurements. Valence and arousal were also extracted from the data sets and processed according to specifications.

2.1 Electrocardiograms (ECG)

ECG signals coming from the right and left hand of each participants were collected. In order to detect heart related features, Lead 1 was computed by subtracting measurements from the right to measurements from the left hand. Next, a low pass filter [5] was passed on the data trace. Low-pass filters attenuate only the amplitude of higher frequency components. On ECG signal, they are used to remove high frequency muscle artifact and external interference.

As a complementary task, the central mean of negative values of each ECG trace was computed and removed from the filtered signal to place the ECG signal in the first Cartesian quadrant [8]. Recordings with only negative frequencies were not responsive to this strategy.

Peaks of the QRS complex were detected when the signal hit a prominence of 90 Hz with a minimum distance of 150 milliseconds from other peaks and a width of 3 steps. This strategy excluded the majority of artifacts a priori. `Figure 1` presents the filtered signal in the time domain of the first participant, first clip.

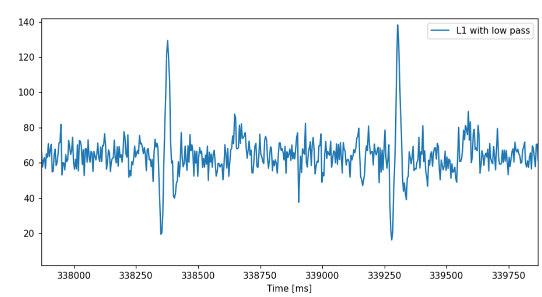

Figure 1: The ECG signal of participant 1, clip 1, after the first steps of data processing.

Further data processing was executed to identify additional artifacts in the signal [3] . `Figure 2` shows one invalid interval that was excluded based on the algorithm for artifacts detection.

Inter-beat intervals (IBI), heart rate (HR) [2] and heart rate variability (HRV) were computed from the valid data. The Power Spectral Density (PSD) was computed with Welch's method. The selected segment length was 15 times the sampling frequency, the number

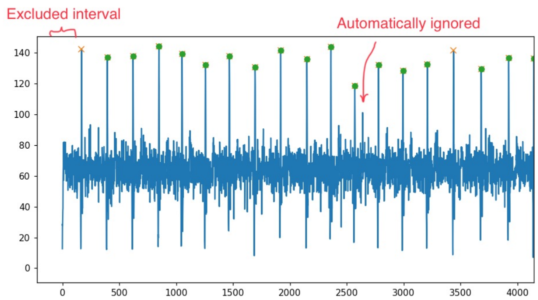

Figure 2: The ECG signal of participant 1, clip 1, after the first steps of data processing.

of points to overlap were 10 times the sampling frequency [6], all data was used to construct the power signal (shown in `Figure 3`).

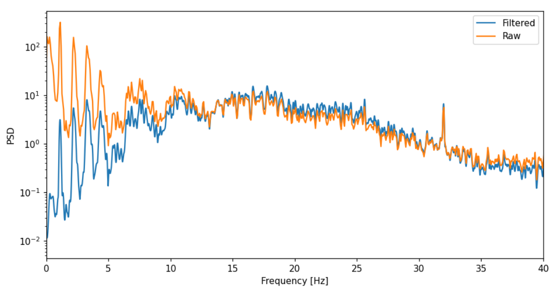

Figure 3: The PSD of ECG data, participant 1, clip 1.

Statistical measurements (band power area, standard deviation, min and max) of the entire 0-0.04Hz band were also collected as additional features, because the integrated power of the four sub-bands in the same frequency domain was every so often unresponsive.

The PSD frequencies band 0-100Hz was divided in 10 sub bands and their band power was integrated to collect additional features. These features were said to contain information about emotion recognition in a previous work [7].

As a personal curiosity, the maximum and minimum measurement in HR, HRV and IBI was extracted to study to what extent the greatest/smallest HR/HRV/IBI can have a role to predict emotional state.

2.2 Electroencephalograms (EEG)

Electroencephalograms features, consisting of the statistical measurements over the channel, average of the first derivative, proportion of negative differential samples,mean number of peaks, mean derivative of the inverse channel signal and average number of peaks in the inverse signal for each channel were computed for each clips and participants. When data was recorded as invalid than the feature was zeroed.

2.3 Electrodermal activities (EDA)

Electrodermal activities was measured as galvanic skin resistance (GSR). Features based on GSR were the average rising time, power density estimates, mean skin resistance, mean of first derivatives of skin resistance, mean of absolute values of first derivatives of skin resistance, mean first derivative for negative values only, percentage of time with negative first derivative, standard deviation and the number of local minima.

`Figure 4` shows the GSR Signal of the second participant watching Clip 1 plotted in the time domain.

The data trace was filtered using a 1st order Butterworth filter with low pass cut off frequency of 5 Hz [1].

A low pass filter is useful in GSR data to remove noises resulting from body movements and surrounding electromagnetic fields.

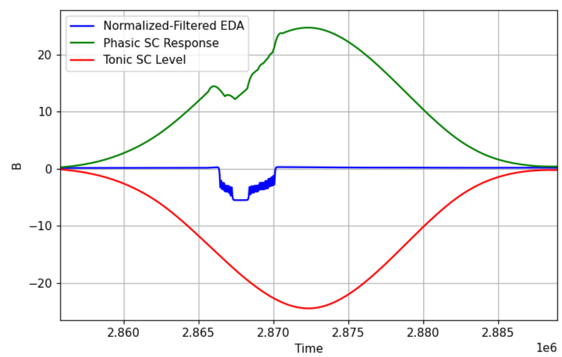

Figure 4: The PSD of EDA data, participant 2, clip 1.

Also conductance's related features were extracted. Conductance is measured as the reciprocal of resistance. The features extracted are the number of local minima, standard deviation, mean of first derivatives, mean of absolute values of first derivatives, mean of absolute values of second derivatives and number of local minima.

`Figure 5` shows the PSD for the EDA signal of the second participant watching Clip 1.

As for ECG, also in EDA PSD were used to capture additional features. Once again, the reason behind is this choice is academical. PSD of EDA were used in related work in emotional state recognition [4]. Several researches have considered statistical aspects (minimum, maximum, and variance signal magnitude area, skewness, kurtosis, harmonics summation) to predict emotional state with EDA signal. Thus, these have been integrated in the array features.

2.4 Facial landmark trajectories (EMO)

Statistical measurements vertical deformation of the upper lip (vdUL), vertical deformation of the lower lip (vdLL), horizontal deformation of the left lip corner (hdLLC), vertical deformation of left lip corner (vdLLC), horizontal deformation of the right lip corner (hdRLC), vertical deformation of the right lip corner (vdRLC), deformation of the right eyebrow (dRE), deformation of the left eyebrow (dLE), deformation of the right cheek (dRC), deformation

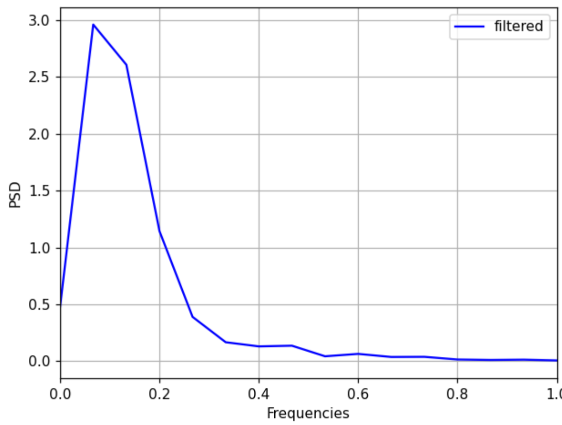

Figure 5: The PSD of EDA data, participant 2, clip 1.

Table 1: Arousal and valance classes

Arousal	Valence	Count
high	high	483
high	low	188
low	high	473
low	low	311

of the left cheek(dLC), deformation of the right lid (dRL) and deformation of the left lid (dLL) were computed for each clips and participants. When data was not available than the feature was zeroed.

2.5 Valence and arousal

Valence and arousal measures are independent to one another,(correlation = -0.01768, and p-value = 0.5), which means that valence is a bad predictor of arousal and vice versa. This finding is not surprising because the two indicators measure two different phenomena (i.e. the sign of the emotion and its intensity). The two labels have been classified in an emotion classification system inspired by dimensional theories of emotions. Arousal levels strictly higher than 3 and valence levels strictly higher than 0 were labeled "high" (1), while the rest was labeled "low" (0).

Positive and negative emotions are equally distributed as well as emotion intensity. This equilibrium is key to efficiently train a machine learning classifier. The `Table 1` reports the data set on the valance and arousal dimension.

Before the classification stage, invalid data has been filtered and removed. Infinite values have been zeroed. Statistical measurements of ECG samples that did not contained any peak in the last 50 seconds were filled with their respective means.

3 RESULTS

A random forest classifier (RFC) with 300 estimators was used and evaluated against a Zero Rule algorithm (ZRC) with majority rule.

Table 2: LOPO, Statistical measurements

Classifier	Accuracy	F1	Precision	Recall
Random Forest	0.57	0.8	0.7	0.95
Zero Rule	0.33	0.6	0.66	0.59

Two validation schemes have been used to evaluate the classifiers : a leave-one-participant-out (LOPO) and a leave-one-clip-out (LOCO). The obtained median accuracy, precision, recall and F1-score for each class is shown in the `Table 2 and 3`. The code report, attached with this paper, illustrates the complete table of accuracy, precision, recall and F1-score data for the RFC in each class. Confusion matrices results will be interpreted in the Discussion's section.

3.1 Leave-one-participant-out (LOPO)

In the LOPO validation scheme, the RFC always outperforms the ZRC.

RFC detects the correct class with 20 % more accuracy and a 36 % greater F1-score than the ZRC . Furthermore, it is 20 % more sensitive and 4 % more precise than what is required by the baseline.

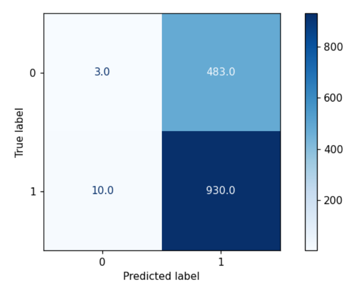

Figure 6: LOPO, Arousal confusion matrix

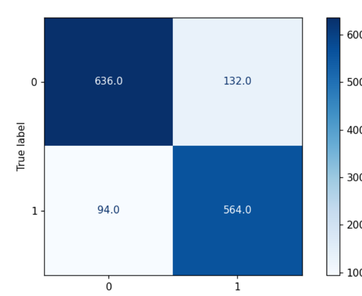

Figure 7: LOPO, Valence confusion matrix

Table 3: LOCO, Statistical measurements

Classifier	Accuracy	F1	Precision	Recall
Random Forest	0.36	0.68	0.62	0.76
Zero Rule	0.21	0.61	0.67	0.54

3.2 Leave-one-clip-out (LOCO)

When one of the last 20 clips is left out, the LOPO validation scheme under-performs the ZRC. Nonetheless, it is true that the RFC detects the correct class with 15 % more accuracy and 7 % greater F1-score. The ZRC is 5 % more precise, but 22 % less sensitive.

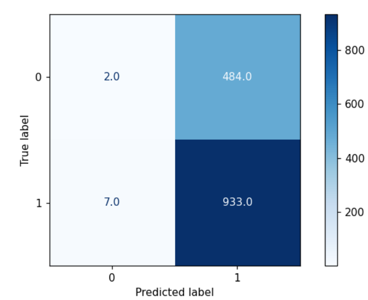

Figure 8: LOCO, Arousal confusion matrix

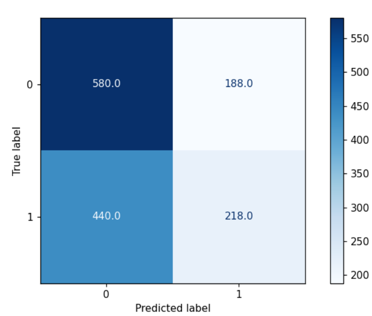

Figure 9: LOCO, Valence confusion matrix

3.3 Best features

`Table 4 and 5` summarize which features and sensor modalities have the greatest absolute predictive power for emotional state classification in arousal / valence classes dimensions. The predictive power is interpreted as the Pearson's correlation coefficient. In a controlled environment, I would select the first five features of each of the two dimensions for emotional classification, because they have the best correlation with the labels. This decision might be different in a in-the-wild context.

Table 4: Correlation with valence classes

Modality	Feature	Correlation
EMO	% vdLL of values under $\mu - \sigma$	0.102161
EMO	standard deviation of hdLLC	0.101611
EDA	PSD 0-0.04Hz bandpower area (add)	0.065592
EMO	mean of dLE	0.065592
EEG	EEG 15	0.059638
EEG	EEG 8	0.058480
EEG	EEG 12	0.056818
EEG	EEG 13	0.056427

Table 5: Correlation with arousal classes

Modality	Feature	Correlation
EEG	EEG 19	0.118785
EMO	% of vdLL values above $\mu + \sigma$	0.099383
EEG	EEG 32	0.089235
EEG	EEG 62	0.071926
EEG	EEG 76	0.071193
EEG	EEG 80	0.070698
EMO	% of vdUL values under $\mu - \sigma$	0.070512
EEG	EEG 77	0.070239

4 DISCUSSION

As said previously, the first five features of each of the two dimensions in `Table 4 and 5` can be considered as the most relevant for emotional classification.

EMO features appear to be the best to classify valence. Recording them is generally unobtrusive, but users always have to face a camera if continuous tracking is desired.

On the other side, EEG features can continuously track real life events (and they do so really well especially for arousal classification), but they are still too obtrusive for daily consumer use, although latest improvements in Brain Computer Interfaces' technology could help lower this barrier.

Finally, EDA electrodes provide continuous tracking, but they are obtrusive and require training. Developments in smart clothing might as well contribute to make this latter modality less obtrusive and wide spread.

FRC generalizes better in LOPO validation scheme than in LOCO validation (see `Table 2 and 3` for reference). Individual differences in the data set are nonetheless remarkable. RFC achieves 95% accuracy when is set to predict the emotional state of participant number 40, but just 35 % accuracy for participant 23. Clip 31 provoked the most predictable (70 %) emotional state, whereas clip 17 was very unpredictable (10%) using the remaining physiological data. Indubitably, removing a clip from this small data set of 44 clips, disrupts the ratio in the arousal and valence classes which lead to worse prediction power.

`Figure 6 to 9` show the confusion matrices of the two validation schemes for both emotion dimensions. RFC classifier performs better for very high arousal level. Low arousal levels are almost never predicted correctly. Contrary to general accuracy seen before, RFC is slightly more precise in LOCO to detect very high arousal level than in LOPO.

RFC classifier performs equally good for high and low valence level in the LOPO validation scheme. In LOCO, RFC classifier is better to predict low valence level. Predictions error are more uniform compared to the arousal's one.

Classification accuracy could be improved by including information about age, gender, height, health history etc. of the participants. In the domain of physiological signals, an additional sensing modality could be sound, in the form of changes in pitch, loudness, timbre of the voice, speech rate, and pauses. Valuable insight could also be gain with measures of the Circadian rhythm and movement. To conclude, I believe continuous feedback with the user is also desirable in real life applications, in order to refine the classifier and account for individual differences.

REFERENCES

[1] Raymond J. Dolan Dominik R. Bacha, Karl J. Fristona. 2013. An improved algorithm for model-based analysis of evoked skin conductance responses. *Biological Psychology* (2013).

[2] Isselbacher KJ Wilson JD Martin JB Kasper DL Hauser SL Longo DL Fauci AS, Braunwald E. 1998. *Electrocardiography.*

[3] Karen Hovsepian. 2015. cStress : towards a gold standard for continuous stress assessment in the mobile environnment. (2015). Proceedings of the 2015 ACM international join conference on pervasive and ubiquitous computing.

[4] Joan Oliver G. C. Nandi Senior Member IEEE Jainendra Shukla, Miguel Barreda-Angeles and Domenec Puig. 2019. Feature Extraction and Selection for Emotion Recognition from Electrodermal Activity. (April 2019).

[5] Dr. Rahul Kher. 2019. Signal Processing Techniques for Removing Noise from ECG Signals. *Journal of Biomedical Engineering and Research* (Feb. 2019).

[6] Seyed Mostafa Kia Paolo Avesani Ioannis Patras Mojtaba Khomami Abadi, Ramanathan Subramanian and Nicu Sebe. 2015. Decaf: Meg-based multimodal database for decoding affective physiolog- ical responses. *IEEE Transactions on Affective Computing* (2015).

[7] Roshan Ragel Theekshana Dissanayake, Yasitha Rajapaksha and Isuru Nawinne. 2019. An Ensemble Learning Approach for Electrocardiogram Sensor Based Human Emotion Recognition. (October 2019).

[8] Maolin Pan Xuanyu Lu and Yang Yu. [n.d.]. QRS Detection Based on Improved Adaptive Threshold. ([n. d.]).