Emotion Recognition in Videos From EEG Signals Using Deep Neural Networks (June 2019)

D. Tyrna, T. Marcol, M. Woźnica, K. Włodarski, A. Piepiora, A. Marzec

Abstract **?**

I. Introduction

In 1875 the electrical activity of the brain was observed for the first time. Since that time scientists try to discover how this activity influence us. Nowadays, the EEG signal is used in medicine to diagnose: epilepsy, sleep disorder, coma, brain death etc. This article is about recognition of the emotions evoked by watching videos. Such emotions are detected within the EEG signal recording using deep neural network. In this experiment we would like to classify EEG data (level of arousal and valence) to appropriate class of emotion. The most difficult in this task is analyzing meaningful data from all other data. Reading the electrical activity is used by disabled patient to contact with the people.

II. Literature Review

Literature…

Dlaczego w artykule 1 było lepiej niż w 2, łączymy fakty, oni używali tego , inni tamtego, czym się różniły wyniki

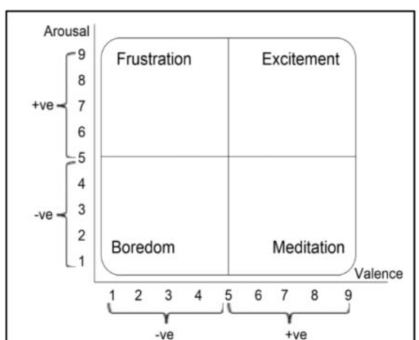
III. Materials

As an input material used for presented project, 'DEAP' dataset was provided.

'DEAP' is a multimodal dataset for the analysis of human affective states. It contains data provided by the electroencephalogram (EEG) and peripheral physiological (using GSR measurement) signals of participants (in presented case: 31 people) as each watched 40 one-minute long videos and was asked about their subjective emotional state (level of valence/arousal, like/dislike, dominance and familiarity). Beside the main sets (which are described in the next part of the work), additional tables were provided, each contains data of for example participant ratings or participant questionnaire. The video list (in which condensed data from all participants in terms of particular videos were presented) was also included. In this project, only the main sets of data were used.

As an input, already preprocessed DEAP dataset was provided. Mentioned preprocessing included bandpass frequency filter (4-40 Hz), downsampling from 512 to 128 Hz and EOG artefacts removal. Each set of data (participant file) contains two arrays with labels (40x4 video/labels) and data (40x40x8064 video/channel/data) respectively. Label array consists of valence, arousal, dominance and liking values computed for each video. In terms of the second array, for this project only first 32 channels were used, as the rest contains data from the GSR measurement, which is not included as an input for dataset processing in this work. Each channel (1-32) refers to particular electrode (EEG technique) and contains data from the experiment provided by the mentioned electrode.

In order to provide labels for validate four-state output of the network, discrete values were extracted from the labels array according to the two-dimensional emotion model approach proposed by Russell, as shown on figure below (as suggested and shown in Al-Nafjan, A., Hosny, M., Al-Wabil, A., & Al-Ohali, Y. (2017). Classification of human emotions from electroencephalogram (EEG) signal using deep neural network. *International Journal of Advanced Computer Science and Applications*, *8*(9).).



Label	State	Mapping
0	Meditation	*Valence >5 ∩ Arousal<5*
1	Boredom	*Valence <5 ∩ Arousal<5*
2	Excitement	*Valence >5 ∩ Arousal>5*
3	Frustration	*Valence <5 ∩ Arousal>5*

Fig. 3. Proposed emotion model and the classification labels.

As the part of final results improvement, additional condition was used. The dataset was split additionally into two groups - ‘strong’ and ‘weak’ in terms of valence/arousal values for each sample. Strong sample refers to one having both valence and arousal values close to the limits of the scale. After some research, lower band was set to 3, and the higher to 7, that provides approximately half-reduced number of samples, but ones having stronger results in terms of presented four-state output, which is thought to be a factor of method improvement.

For being more accurate for processing by neural network, data are transformed to frequency domain and split into four bands: *alpha*, *beta*, *gamma* and *theta*, each representing specified range of frequencies.

To obtain data in form, which can be easily processed by the neural network, the average band power in each channel is mapped into a spatial map (as suggested in Li, Y., Huang, J., Zhou, H., & Zhong, N. (2017). Human emotion recognition with electroencephalographic multidimensional features by hybrid deep neural networks. *Applied Sciences*, *7*(10), 1060.). As a result, each trial is represented by four maps presenting spatial distribution of average band power in each of the four mentioned bands: *alpha*, *beta*, *gamma* and *theta*. Details of the method are described below.

After initial preparation of the data, power spectral density for each channel in each trial completed by each participant is computed. As long as we are not interested in processing each of the frequencies separately, the average band power is calculated for alpha, beta, gamma and theta bands (as suggested in Al-Nafjan, A., Hosny, M., Al-Wabil, A., & Al-Ohali, Y. (2017). Classification of human emotions from electroencephalogram (EEG) signal using deep neural network. *International Journal of Advanced Computer Science and Applications*, *8*(9).). The values computed this way are then mapped into a 9×9 matrix (the dimensions are maximum point numbers between the horizontal or vertical test points). The empty cells of the matrix are filled with average values stored in the surrounding cells, using following formula (from Li, Y., Huang, J., Zhou, H., & Zhong, N. (2017). Human emotion recognition with electroencephalographic multidimensional features by hybrid deep neural networks. *Applied Sciences*, *7*(10), 1060.):

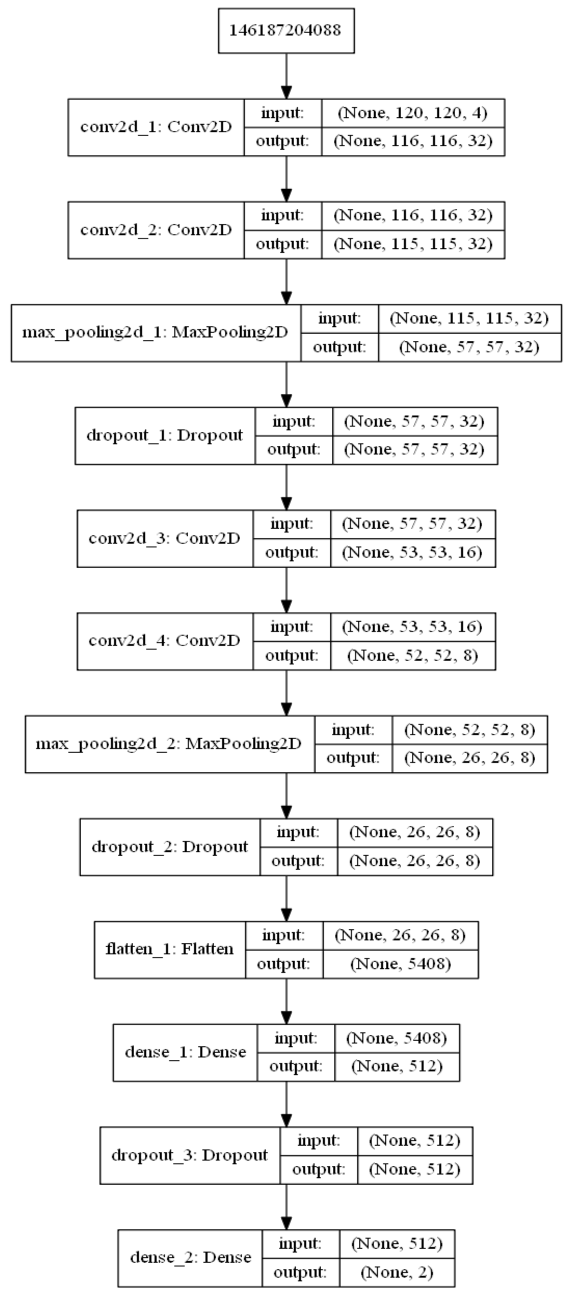
$$V_{(m,n)} = \frac{V'_{(m+1,n)} + V'_{(m-1,n)} + V'_{(m,n+1)} + V'_{(m,n-1)}}{K}, (0 \leq m, n \leq 8, n \in N)$$

where V is the value of the empty cell (corresponding to $P_{(m,n)}$ and V’ is the value of the point surrounding $P_{(m,n)}$. If the index of the surrounding point exceeds the range of 0 and 8, then the value is 0. K is the number of non-zero elements in the numerator, and the default value of K is 1.
After the feature matrix is filled, it is used as a base table to generate data matrix through the interpolation method. The final data matrix has dimensions of 4×120×120 (corresponding to four bands and dimensions of the matrix of average band power obtained through the interpolation).

IV. Methods

The experimental setup of the network is as shown in the schematic figure 1.
The activation functions for all of the convolutional layers are ReLU. The dropout values are 0.5, 0.3 and 0.5.
Compilation parameters were chosen as follows:

- loss function: categorical cross entropy,
- optimizer: adam,
- metric: accuracy.



Training and validation:

Set of samples generated from the original dataset was split into two main parts: train and test set, with the percentage of 70% dedicated for training and 30% for testing. Test set has been chosen randomly through all data.

Two validation methods were proposed for the solution. First of them is the simple approach, which aim is to train data set only once, with the validation percentage set to 30% (could be differentiated across the trials and treated as a parameter of execution). Second one is commonly used K-fold technique, which provide K trainings, where for each one the different validation set is chosen as the package of [nr of samples / K] samples, chosen by random or straight splitting. Number of iterations is one of the most important parameters of execution and could have a strong impact on the results, which is going to be researched.

V. Results

.

Results
Jak dany parameter wpływa, jakie były najlepsze wyniki, dla jakich parametrów itd

VI. Summary

Summary…

Podsumowanie, jakie mieliśmy wyniki I jak wypadło

VII. Appendix

A. Credit – wypisanie kto co robił, może być relacja wiele do wielu, kto co zrobił w raporcie / artykule

B. Code – sensowne fragmenty

References

Basic format for books:

J. K. Author, “Title of chapter in the book,” in *Title of His Published Book, x*th ed. City of Publisher, (only U.S. State), Country: Abbrev. of Publisher, year, ch. *x*, sec. *x*, pp. *xxx–xxx.*

Examples:

[1] G. O. Young, “Synthetic structure of industrial plastics,” in *Plastics*, 2nd ed., vol. 3, J. Peters, Ed. New York, NY, USA: McGraw-Hill, 1964, pp. 15–64.

[2] W.-K. Chen, *Linear Networks and Systems*. Belmont, CA, USA: Wadsworth, 1993, pp. 123–135.