EEG-Based Emotion Recognition Using Convolutional Neural Networks

Maria Mamica, Paulina Kapłon, and Paweł Jemioło

AGH University of Science and Technology, A. Mickiewicza 30, 30-059 Krakow, PL
{mamica,pkaplon}@student.agh.edu.pl, pawljmlo@agh.edu.pl

Abstract. In this day and age, Electroencephalography-based methods for Automated Affect Recognition are becoming more and more popular. Owing to the vast amount of information gathered in EEG signals, such methods provide satisfying results in terms of Affective Computing. In this paper, we replicated and improved the CNN-based method proposed by Li et al. [11]. We tested our model using a Dataset for Emotion Analysis using EEG, Physiological and Video Signals (DEAP) [9]. Performed changes in the data preprocessing and in the model architecture led to an increase in accuracy – 74.37% for valence, 73.74% for arousal.

Keywords: Deep Learning, Convolutional Neural Networks, CNN, Electroencephalography, EEG, Emotion Recognition, Affective Computing

1 Introduction

As technology becomes more advanced and publicly available, the everyday recipients' expectations are growing exponentially fast. The services are supposed to conform to the needs of every customer. The market expands by incorporating Machine Learning to correctly recognize the emotions present while interacting with the services to meet the growing expectations.

If an efficient and highly accurate Emotion Recognition system were implemented, everyday life would be significantly improved. Not only would we be provided with genuinely entertaining products, but also approached more suitably. Thus, there is a big emphasis on studying human psychics and producing new, better methods for Emotion Recognition. The most efficient way is to discuss the findings and discoveries on the broad forum of scientists.

Nevertheless, we are currently witnessing a crisis in psychology. Studies that have been considered credible for decades have been criticized as they contained methodological flaws [3, 14]. Unfortunately, a similar phenomenon also applies to Computer Science. Novel Artificial Intelligence articles are often lacking in description of parameters and architecture. Even the latest articles often overlook such important information [6]. It was an inspiration for the authors of this paper. That is why we decided to replicate and, if possible, extend one of the articles on Emotion Recognition which has been published recently.

We strongly believe science should be replicable, and therefore, we include all the necessary information for anyone wanting to use the proposed methods. What is more, in addition to replicating the described methods, we also

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77977-1_7

introduced some improvements that allowed us to increase accuracy in detecting affective states. We followed the approach introduced in the paper by Li et al. [11] and incorporated a few improvements. We selected the above study as our starting point since we were interested in the presented methodology, especially in processing signals onto the 4-channel images and Convolutional Neural Networks. What is more, the study was relatively recent at that moment and brought promising results.

The rest of the paper is organized as follows. Section 2 gives an overview of Affective Computing and presents the technologies used in this research. In section 3, we described the model [11] and all introduced improvements. In section 4, obtained results are provided and compared. Finally, in section 5, we present the conclusions and plans for future developments related to this research.

2 Background

Affective Computing is a paradigm of Human-Computer Interaction that aims to recognize, interpret, process and simulate human emotions to adapt to a particular user in a specific emotional state. One of Affective Computing core concepts is Affective Loop [15]. In the Loop, emotions are seen as a process based on the interaction. The initial state of the Loop starts as a user begins the communication with a system. Then, the system responds in a way that affects and pulls the user into the following exchanges. The user is more engaged in the cooperation if the system can correctly read reactions in which Emotion Recognition is essential [15]. Nowadays, researchers focus on games as they represent the true potential of Affective Computing [8, 13, 19].

There are many different approaches to Emotion Recognition. It applies to both the computation and the extraction of the data. Among the many ways to gather needed data, the most popular are directly asking about them, using voice or facial expression, analyzing texts or speech [7]. Interpreting physiological signals like electrocardiography or electroencephalography is also growing wide as the technology required to collect the data is slowly becoming more available. This paper focuses on the latter.

EEG (Electroencephalography) is a primarily noninvasive method aiming to record the electrical activity of the brain. The examination is conducted using several electrodes placed directly on the scalp of the subject. These sensors register changes in the electric potential on the surface of the skin caused by the activity of the cerebral cortex. After triggering different emotions, for example, by showing emotion-inflicting images [12] or videos [9], we can gather EEG record corresponding to these emotions [17].

The reading of the electroencephalograph is mainly broken down into frequency, voltage, reactivity, synchrony and distribution [5]. The examination of EEG signals usually involves the analysis of rhythmic activity. Then, such rhythmic activity can be divided into five bands based on the frequency. They consist of consequently: Delta, Theta, Alpha, Beta and Gamma bands, with the Delta band's frequency being the lowest [1].

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77977-1_7

In the paper [11], the most relevant channels are used: Gamma, Beta, Alpha, Theta. According to [1], the Delta waves are the slowest EEG waves and are only detected during deep sleep. Analyzing each bandwidth individually and four of them as a whole helps to map them to the emotions they represent.

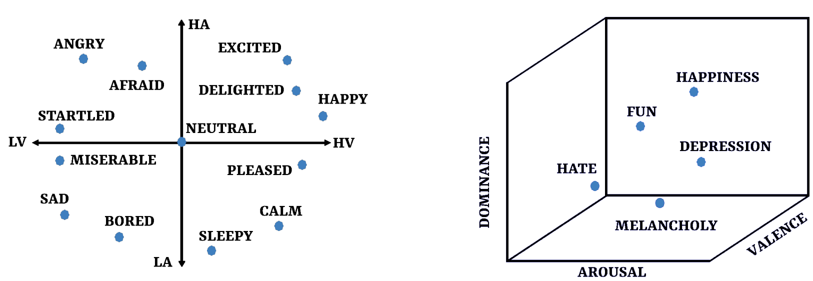
As the human psyche is complex, and it is still unclear how emotions should be represented, one of the most popular approaches has been incorporated. It is called Arousal and Valence Space (AVS) [16]. The AVS allows categorizing the vast abstract, which is emotion, as one of the four classes (see Figure 1(a)). The following parameters determine these classes:

– Arousal – indicates engagement and ranges from inactive to active.
– Valence – ranges from unpleasant to pleasant.

The arousal and valence are then used to define the four classes. For example, according to [10], such states can be represented as:

1. high arousal and high valence [HAxHV], e.g. excitement/delight/happiness
2. high arousal and low valence [HAxLV], e.g. anger/fear/startle
3. low arousal and low valence [LAxLV], e.g. sadness/boredom/misery
4. low arousal and high valence [LAxHV], e.g. pleasure/calmness/sleepiness

The AVS can be extended by adding another parameter – dominance, which could range between feeling powerless and being in control (see Figure 1(b)).



(a) 2D emotion model, based on [10]. (b) 3D emotion model, based on [18].

Fig. 1. Different emotion models visual representations.

3 Presented approach

3.1 Dataset

As we aimed to replicate the paper by Li et al. [11] fully, we decided to develop and evaluate described models using the same reference dataset, DEAP. It consists of the EEG signals gathered from 32 participants while watching 40 one-minute music videos. The trials were chosen manually or with the use of affective tags. Additionally, it involves the ratings provided by volunteers indicating the level of valence, arousal and other dimensions [9].

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77977-1_7

3.2 Prepossessing

The model presented in this paper was inspired by another article [11]. Li et al. proposed not only the preprocessing approach but also the Neural Network architecture, which gave quite satisfying results. Their method is described next.

To retrieve more samples, all EEG signals were segmented using 8 seconds window with 50% overlapping. Furthermore, for each channel, 4 frequency bands were extracted (theta rhythm (4-8 Hz), alpha rhythm (8-13 Hz), beta rhythm (13-30 Hz) and gamma rhythm (30-40 Hz) and the averaged Power Spectral Density was calculated over the bands. In order to project 3D coordinates of EEG electrodes onto 2D space, the Azimuthal Equidistant Projection was performed. Authors [11] chose the Cz sensor as the centre point and computed the azimuth and the distance value – denoted as Θ and ρ respectively – for all the electrodes using the below equations:

$$\rho = arccos(sin\varphi_1 sin\varphi + cos\varphi_1 cos\varphi cos(\lambda - \lambda_0)) \tag{1}$$

$$\Theta = arctan(\frac{cos\varphi sin(\lambda - \lambda_0)}{cos\varphi_1 sin\varphi - sin\varphi_1 cos\varphi cos(\lambda - \lambda_0)}) \tag{2}$$

(a) Low valence. (b) High valence. (c) High arousal.

Fig. 2. Exemplary projections generated using the described method.

Herein, φ denotes latitude, and λ stands for longitude represented in the geographic coordinate system. (φ_1, λ_0) are geographic coordinates of the Cz point and can be found on the website of measuring equipment [4]. Finally, Cartesian coordinates (x,y) were computed using the below set of equations:

$$x = \rho sin\Theta \tag{3}$$

$$y = -\rho cos\Theta \tag{4}$$

To interpolate the calculated values over 32×32 mesh, the Clough-Tocher scheme was applied as the final step of the image generation operation [2]. Exemplary projections are presented in Figure 2.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77977-1_7

3.3 Replicated Model

Li et al. [11] proposed the Convolutional Neural Network (CNN) model. Its graphical representation can be found in the mentioned paper. It was built with three convolutional layers with 3×3 sized filters, each followed by a max-pooling layer with 2×2 sized blocks. Furthermore, the model was equipped with a fully connected layer with 256 units and a 0.5 dropout rate. Finally, 10-fold cross-validation was performed.

Unfortunately, in [11], the number of epochs and the batch size was not provided. Although it was impossible to replicate the model entirely, these parameters might be adjusted empirically without substantially impacting the results. Thus, we decided to set the batch size to 256 and the number of epochs to 200.

3.4 Improvements

In the DEAP dataset [9], participants' ratings are associated with the trials. While segmenting data, each new sample must be labelled with the emotion assigned to the example considered in a particular step. It means that too short segments may not involve specific brain responses associated with the examined emotion [18]. Having that in mind, we decided to increase the window size to 16 seconds. After this operation, the number of samples equals 7680 ($6\times32\times40$).

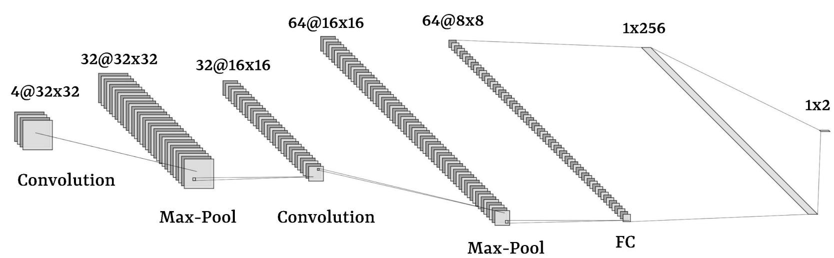


Fig. 3. The proposed Neural Network architecture.

To improve the model, various architectural approaches were tested, and finally, the Neural Network shown in Figure 3 was chosen. It contains two convolutional layers with a 3×3 sized kernel and 1 pixel of padding, each followed by the max-pooling layer having 2×2 filters. The model has one fully connected layer equipped with 256 units and a 0.5 dropout rate. To improve the generalization of the model, an additional dropout with a 0.2 rate was applied directly before the output layer. For the activation, the $ReLU$ function was chosen for all layers, except for the output one, for which the *softmax* was applied.

To measure the performance of the model during the training, the cross-entropy function was used. It is optimized by Adam optimizer with a learning rate set to 0.001. For the training, the number of epochs was adjusted to 200

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77977-1_7

and the batch size to 256. As the number of samples decreased comparing to the article [11], the number of folds used in cross-validation was set to 5 instead of 10. This operation increased the size of the validation set, and more authentic results might have been observed.

4 Preliminary results

The model proposed in this article, and the model [11], which was the starting point of our research, were evaluated using the DEAP dataset [9]. Both methods consider CNN based binary classification for valence and arousal indicators with classes: low and high. Firstly, the method proposed by Li et al. [11] was successfully replicated. However, the researchers did not provide all necessary information. All required hyperparameters, which the authors did not provide, were adjusted empirically to give the best results.

In the article [11], the performance of the model was measured as the average of K-fold cross-validation. We followed this approach and achieved 70.76% for valence and 70.54% for arousal. The performance of the method proposed in this paper was also measured as the average of K-fold cross-validation. Accuracy gained for the valence classification equals 74.37% and for the arousal – 73.74%. The proposed method outperforms the study [11].

The improvement in the performance of the proposed method was, among others, caused by the increase of the segmenting window. Thus, all considered data samples carry more information, and they might be more accurately classified by the model. The application of dropout layers and the K-fold cross-validation technique enabled the generalization of the network. Hence, the model does not overfit and gives quite a high accuracy. Although the CNN architecture was simplified, it gives better results.

5 Conclusions and Future Work

In this paper, we replicated the method proposed in the article [11]. As not all parameters were provided, it was required to find them empirically. We strongly believe that authors should more carefully describe their work to keep the scientific development on a high level. Lack of some required parameters may be confusing for all researchers wanting to contribute or replicate the method. Nevertheless, we find the Li et al. approach promising. Even though some parameters were missing, we could adjust them empirically and fully replicate the model, achieving good results.

In the article, we additionally introduced improvements to the considered model. We observed an increase in the accuracy of the model – 74.37% for valence, 73.74% for arousal. In the future, we want to extend the architecture by applying more fully connected layers and potentially implementing multimodal fusion with other physiological (e.g. ECG, EDA) and behavioural (e.g. photo, video, game logs) data.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77977-1_7

References

1. Abo-Zahhad, M., Ahmed, S., Seha, S.N.: A new eeg acquisition protocol for biometric identification using eye blinking signals. International Journal of Intelligent Systems and Applications (IJISA) **07**, 48–54 (05 2015)
2. Alfeld, P.: A trivariate clough—tocher scheme for tetrahedral data. Computer Aided Geometric Design **1**(2), 169–181 (1984)
3. Bakker, M., Wicherts, J.: The (mis)reporting of statistical results in psychology. Behavior research methods **43**, 666–78 (04 2011)
4. B.V., B.: Biosemi EEG ECG EMG BSPM NEURO amplifier electrodes
5. Duffy, F.H., Iyer, V.G., Surwillo, W.W.: Clinical electroencephalography and topographic brain mapping: Technology and practice. Springer (2012)
6. George, F.P., Mannafee, I., et al.: Recognition of emotional states using eeg signals based on time-frequency analysis and svm classifier. International Journal of Electrical and Computer Engineering (IJECE) **9**, 1012 (04 2019)
7. Imani, M., Montazer, G.A.: A survey of emotion recognition methods with emphasis on e-learning. Journal of Network and Computer Applications (08 2019)
8. Jemioło, P., Giżycka, B., Nalepa, G.J.: Prototypes of arcade games enabling affective interaction. In: International Conference on Artificial Intelligence and Soft Computing. pp. 553–563. Springer (2019)
9. Koelstra, S., Mühl, C., Soleymani, M., Lee, J.S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I.: Deap: A database for emotion analysis using physiological signals. IEEE Transactions on Affective Computing **3**, 18–31 (12 2011)
10. Kollias, D., Tzirakis, P., Nicolaou, M., Papaioannou, A., Zhao, G., Schuller, B., Kotsia, I., Zafeiriou, S.: Deep affect prediction in-the-wild: Aff-wild database and challenge. International Journal of Computer Vision **127** (2019)
11. Li, C., Sun, X., Dong, Y., Ren, F.: Convolutional neural networks on eeg-based emotion recognition. In: Jin, H., Lin, X., Cheng, X., Shi, X., Xiao, N., Huang, Y. (eds.) Big Data. pp. 148–158. Springer Singapore, Singapore (2019)
12. Mikels, J., Fredrickson, B., Samanez-Larkin, G., Lindberg, C., Maglio, S., Reuter-Lorenz, P.: Emotional category data on images from the international affective picture system. Behavior research methods **37**, 626–30 (12 2005)
13. Nalepa, G.J., Kutt, K., Giżycka, B., Jemioło, P., Bobek, S.: Analysis and use of the emotional context with wearable devices for games and intelligent assistants. Sensors **19**(11), 2509 (2019)
14. Nuijten, M., Hartgerink, C., Assen, M., et al.: The prevalence of statistical reporting errors in psychology (1985-2013). Behavior research methods **48** (10 2015)
15. Picard, R.W.: Affective computing. MIT press (2000)
16. Russell, J.: A circumplex model of affect. Journal of Personality and Social Psychology **39**, 1161–1178 (12 1980)
17. SatheeshKumar, J., Bhuvaneswari, P.: Analysis of electroencephalography (eeg) signals and its categorization–a study procedia engineering, volume 38, 2012, pages 2525–2536. Procedia Engineering **38**, 2525–2536 (09 2012)
18. Yang, L., Liu, J.: Eeg-based emotion recognition using temporal cnn. In: Data Driven Control and Learning Systems Conference. pp. 437–442 (2019)
19. Yannakakis, G.N., Martínez, H.P., Jhala, A.: Towards affective camera control in games. User Modeling and User-Adapted Interaction **20**(4), 313–340 (2010)

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77977-1_7