Deep Learning Based Emotion Classification through EEG Spectrogram Images

Lorenzo Battisti[a], Alessio Ferrato[a], Carla Limongelli[a], Mauro Mezzini[b] and Giuseppe Sansonetti[a]

[a]Department of Engineering, Roma Tre University, Via della Vasca Navale 79, 00146 Rome, Italy
[b]Department of Education, Roma Tre University, Viale del Castro Pretorio 20, 00185 Rome, Italy

Abstract
Emotion modeling for social robotics has the great potential to improve the life quality for the elderly and individuals with disabilities by making communication, care, and interactions more effective. It can help individuals with communication difficulties express their emotions. It can also be used to monitor the emotional well-being of elderly persons living alone and alert caregivers or family members if there are signs of distress. More broadly, emotion modeling is necessary to design robots closer and closer to human beings that can naturally interact with them by understanding their behavior and reactions. Here, we propose a deep learning technique for emotion classification using electroencephalogram (EEG) signals. We aim to recognize valence, arousal, dominance, and likability. Our technique uses the spectrogram from each of the 32 electrodes applied in the skull area. Then, we employ a Resnet101 convolutional neural network to learn a model capable of predicting several emotions. We built and tested our model on the DEAP dataset.

Keywords
Emotion classification, Electroencephalogram, Deep Learning

1. Introduction and Background

Automatic emotion recognition is a vast and complex area of research. It has attracted the attention of scientists in many fields, including psychology, artificial intelligence, neuroscience, and robotics. The main goal of this research is to create systems capable of automatically recognizing and interpreting human emotions.

Emotions are a fundamental part of the human experience. From the pleasant joy of spending time with a loved one to the pain of facing a difficult time in life. Several models have been created to describe emotions, which can be divided into two large groups: *categorical* models that represent the space of all emotions as a finite set, and *dimensional* models that represent emotions through continuous values on multiple axes [1]. Concerning dimensional models, three main components are frequently used to define emotions and affective states: *arousal*, *valence*, and *dominance* [2]. Arousal refers to an individual's level of enthusiasm or activity. High arousal levels are related to emotions of excitement, whereas lower ones are associated with relaxation. The positivity or negativity of an emotional experience is referred to as valence. For example, a negative value of valence is related to sadness whilst a positive value is associated with feelings of happiness. Dominance refers to an individual's sense of control or authority in a specific scenario. A high value of dominance is linked to feelings of control, whereas low levels are linked to feelings of helplessness.

Emotion recognition is crucial for the creation of social robots. The knowledge of these dimensions can be exploited to create a more natural and human-like interaction between robots and humans. By understanding and expressing different emotional states, robots can better understand and respond to the emotional needs of humans, and therefore improve the overall experience. Several techniques for emotion recognition have been proposed in the scientific literature, including the analysis of physiological signals [3], natural language processing [4], and facial expressions [5]. However, there is no single solution for emotion recognition, and research in this field is still under development.

Our proposal focuses on the analysis of physiological signals, in particular, the electroencephalogram (EEG) signals for emotion recognition. This field has already been studied for a decade. In 2014, Wang *et al.* [6] proposed to extract EEG data (power spectrum, wavelet, and nonlinear analysis) from the observation of movie clips, to assess the association between EEG data and emotional states. Using a Support Vector Machine classifier the authors showed that representing the state space model in the form of linear dynamical systems removes the noise not correlated with emotions. This makes the classification of emotions more accurate. In 2018, Dabas

Joint Proceedings of the ACM IUI Workshops 2023, March 2023, Sydney, Australia
lor.battisti5@stud.uniroma3.it (L. Battisti); ale.ferrato@stud.uniroma3.it (A. Ferrato); limongel@dia.uniroma3.it (C. Limongelli); mauro.mezzini@uniroma3.it (M. Mezzini); gsansone@dia.uniroma3.it (G. Sansonetti)
0000-0003-4953-1390 (G. Sansonetti)

et al. [7] proposed a 3D emotional model for classifying the emotions of users watching music videos based on the DEAP dataset [8]. In 2019 Donmez and Ozkurt [9] proposed to classify EEG signals by using a convolutional neural network. They classified three emotions by using brain signals and spectrogram images.

2. Methodology

In this paper, we propose a machine learning [10] technique, more precisely a deep learning [11] technique, for the realization of a predictive model of emotions using the EEG signal. We built and tested this model using one of the best-known online datasets, the DEAP dataset. This dataset contains the EEG signals of 32 individuals that were collected while the subjects watched and listened to music videos taken from YouTube[1]. Each subject was invited to view 40 one-minute videos and then asked to express her emotions on the dimensional model shown in Figure 1. Additionally, a parameter called *likability*

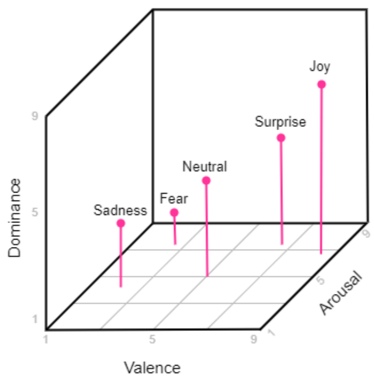

Figure 1: A 3D representation of the emotion dimensional model.

was used to quantify how much the participant liked the stimulus. For each dimension, the participant was asked to rate its intensity on a continuous scale between 1 and 9, where 1 stands for minimum intensity, and 9 for maximum intensity. The EEG signal consists of 32 channels, each corresponding to an electrode that measures the difference in electric potential in the skull area where it is positioned. The proposed methodology provides for the spectral analysis of the signal. This is achieved by applying the Discrete Fourier Transform to the signal, thus obtaining the power of the individual sinusoids that make up the signal. The spectrogram obtained for each EEG channel is a two-dimensional matrix where each cell (f, t) represents the intensity of the sinusoid at the frequency f in the time segment t (for more details please refer to [12]). Figure 2 shows an example of the spectral data of the first two EEG channels of one participant while watching one video. The continuous scale

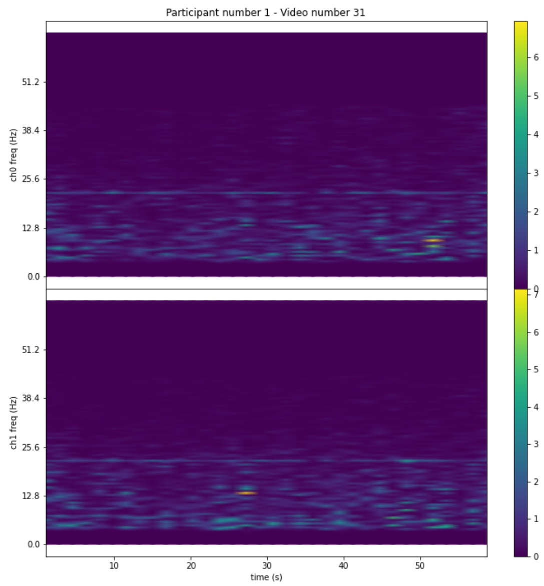

Figure 2: The color plot of the spectral data of the first two EEG channels of one participant while watching one video.

of each emotion e was transformed into a binary value $b(e) \in \{0, 1\}$ so that $b(e) = 0$ if $e < 5$, $b(e) = 1$ if $e \geq 5$. Denoting below the viewing of a video by a subject as an *experiment*, we divided the total number of experiments (i.e., $40 \times 32 = 1280$) reserving 32 experiments for the validation set, 32 for the test set, and the remaining for the training set. The experiments belonging to the validation and test set were arbitrarily chosen, one for each participant relating to different video experiments, so that the sum of positive emotions (in which $b(e) = 1$) and the negative ones (in which $b(e) = 0$) approximately balance each other. We used the ResNet101 [13] convolutional neural network, suitably adapted to take as input a tensor with an arbitrary number of input channels. We empirically tested different hyperparameter configurations through a grid search, obtaining the following optimal values:

- Loss Function: Cross-Entropy
- Optimizer: Stochastic Gradient Descent (SGD)
- Momentum: 0.9
- Weight Decay: 0.0005
- Learning Rate: 3.0e-3

The training set size was limited, so the network tends to overfit after about 200-300 epochs reaching 100% of accuracy on the training set. Therefore, we introduced a data

[1] https://www.youtube.com/

Table 1
Accuracy on 400 epochs without and with augmentation

Emotion	W/O Augmentation	W/ Augmentation
Valence	43%	43%
Arousal	50%	60%
Dominance	50%	63%
Likability	60%	60%

augmentation process that performed a random horizontal offset of the input tensor by up to 20% of the horizontal dimension. The rationale behind this is that a horizontal shift corresponds to a time shift of the signal. It is reasonable that this variability could occur between subject and subject and between experiment and experiment. Table 1 shows the preliminary experimental results.

3. Conclusions and Future Works

In the research literature, it has been largely shown that the knowledge of the user's emotions can make a significant contribution to the creation of increasingly effective human-machine interaction systems. Several aspects can be analyzed to recognize emotions and, more generally, the user's affective state. In this article, we have presented a deep learning approach to EEG signal analysis. Specifically, a ResNet101 convolutional neural network takes the EEG spectrogram as input and returns the values of arousal, valence, dominance, and likability.

Our idea is still evolving, so the possible future developments are manifold. These developments can be methodological or applicative. As regards the former, clearly the data at our disposal are too limited to fully exploit the potential of deep neural networks. We, therefore, need new data, so we are planning to collect it ourselves with the appropriate instrumentation. Another aspect concerns the deep neural network chosen. The ResNet101 is one of the many possibilities that deep learning research makes available today. A further development of our work concerns the data augmentation process, which has been shown to be able to improve the model accuracy. In the system described, the data augmentation concerned only the horizontal shift. Hence, we want to apply new geometric transformations and image processing techniques and verify whether they can further improve the accuracy of the results. As far as application developments are concerned, our idea is to combine physiological data with those related to facial expressions and eye tracking. Our ultimate goal is to improve human-machine interaction, both when the user is dealing with social robots and with recommender systems [14, 15, 16] or multimedia applications [17, 18, 19]. For instance, the information related to the emotions that the user feels when faced with a certain stimulus can be exploited to improve the algorithms for suggesting points of interest to visit (e.g., cultural heritage resources [20, 21, 22] such as museums [23, 24, 25] or restaurants [26, 27]) and itineraries to follow between them [28, 29]. Finally, future development could concern the increase in the stimulus classes to which the user is subjected. In this article, we analyzed the EEG signal collected while the user watches music videos. It would be interesting to collect and subsequently analyze the EGG signal while the user listens to music [30], reads news articles [31], watches a movie [32] or looks at an image [33].

References

[1] A. F. Bulagang, N. G. Weng, J. Mountstephens, J. Teo, A review of recent approaches for emotion classification using electrocardiography and electrodermography signals, Informatics in Medicine Unlocked 20 (2020) 100363.

[2] J. Russell, A circumplex model of affect, Journal of Personality and Social Psychology 39 (1980) 1161–1178.

[3] F. Cavallo, F. Semeraro, L. Fiorini, G. Magyar, P. Sinčák, P. Dario, Emotion modelling for social robotics applications: a review, Journal of Bionic Engineering 15 (2018) 185–203.

[4] J. Deng, F. Ren, A survey of textual emotion recognition and its challenges, IEEE Transactions on Affective Computing (2021).

[5] N. Webb, A. Ruiz-Garcia, M. Elshaw, V. Palade, Emotion recognition from face images in an unconstrained environment for usage on social robots, in: 2020 International Joint Conference on Neural Networks (IJCNN), 2020, pp. 1–8.

[6] X.-W. Wang, D. Nie, B.-L. Lu, Emotional state classification from eeg data using machine learning approach, Neurocomputing 129 (2014) 94–106.

[7] H. Dabas, C. Sethi, C. Dua, M. Dalawat, D. Sethia, Emotion classification using eeg signals, in: Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence, 2018, pp. 380–384.

[8] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras, Deap: A database for emotion analysis using physiological signals, IEEE Transactions on Affective Computing 3 (2012) 18–31.

[9] H. Donmez, N. Ozkurt, Emotion classification from eeg signals in convolutional neural networks, in: 2019 Innovations in Intelligent Systems and Applications Conference (ASYU), 2019, pp. 1–6.

[10] L. Vaccaro, G. Sansonetti, A. Micarelli, An empirical

review of automated machine learning, Computers 10 (2021).
[11] G. Sansonetti, F. Gasparetti, G. D'Aniello, A. Micarelli, Unreliable users detection in social media: Deep learning techniques for automatic detection, IEEE Access 8 (2020) 213154–213167.
[12] M. X. Cohen, Analyzing neural time series data: theory and practice, MIT press, 2014.
[13] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, Los Alamitos, CA, USA, 2016, pp. 770–778. Las Vegas, NV, US, 27–30 June 2016.
[14] F. Gasparetti, G. Sansonetti, A. Micarelli, Community detection in social recommender systems: a survey, Applied Intelligence 51 (2021) 3975–3995.
[15] G. Sansonetti, Point of interest recommendation based on social and linked open data, Personal and Ubiquitous Computing 23 (2019) 199–214.
[16] D. Feltoni Gurini, F. Gasparetti, A. Micarelli, G. Sansonetti, Temporal people-to-people recommendation on social networks with sentiment-based matrix factorization, Future Generation Computer Systems 78 (2018) 430–439.
[17] A. Micarelli, A. Neri, G. Sansonetti, A case-based approach to image recognition, in: Proceedings of the 5th European Workshop on Advances in Case-Based Reasoning, EWCBR '00, Springer-Verlag, Berlin, Heidelberg, 2000, pp. 443–454.
[18] G. Sansonetti, F. Gasparetti, A. Micarelli, Using social media for personalizing the cultural heritage experience, in: Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, UMAP '21, Association for Computing Machinery, New York, NY, USA, 2021, p. 189–193.
[19] L. Xie, Z. Deng, S. Cox, Multimodal joint information processing in human machine interaction: Recent advances, Multimedia Tools Appl. 73 (2014) 267–271.
[20] A. De Angelis, F. Gasparetti, A. Micarelli, G. Sansonetti, A social cultural recommender based on linked open data, in: Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization, UMAP '17, ACM, New York, NY, USA, 2017, pp. 329–332.
[21] G. Sansonetti, F. Gasparetti, A. Micarelli, Cross-domain recommendation for enhancing cultural heritage experience, in: Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization, Association for Computing Machinery, New York, NY, USA, 2019, pp. 413–415.
[22] G. Sansonetti, F. Gasparetti, A. Micarelli, F. Cena, C. Gena, Enhancing cultural recommendations through social and linked open data, User Modeling and User-Adapted Interaction 29 (2019) 121–159.
[23] A. Ferrato, C. Limongelli, M. Mezzini, G. Sansonetti, Using deep learning for collecting data about museum visitor behavior, Applied Sciences 12 (2022).
[24] A. Ferrato, C. Limongelli, M. Mezzini, G. Sansonetti, The meta4rs proposal: Museum emotion and tracking analysis for recommender systems, in: Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization, UMAP '22 Adjunct, Association for Computing Machinery, New York, NY, USA, 2022, pp. 406–409.
[25] M. Mezzini, C. Limongelli, G. Sansonetti, C. De Medio, Tracking museum visitors through convolutional object detectors, in: Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization, UMAP '20 Adjunct, Association for Computing Machinery, New York, NY, USA, 2020, p. 352–355.
[26] C. Biancalana, F. Gasparetti, A. Micarelli, G. Sansonetti, An approach to social recommendation for context-aware mobile services, ACM Trans. Intell. Syst. Technol. 4 (2013) 10:1–10:31.
[27] N. Sardella, C. Biancalana, A. Micarelli, G. Sansonetti, An approach to conversational recommendation of restaurants, in: C. Stephanidis (Ed.), HCI International 2019 - Posters, Springer International Publishing, Cham, 2019, pp. 123–130.
[28] D. D'Agostino, F. Gasparetti, A. Micarelli, G. Sansonetti, A social context-aware recommender of itineraries between relevant points of interest, in: HCI International 2016, volume 618, Springer International Publishing, Cham, 2016, pp. 354–359.
[29] A. Fogli, G. Sansonetti, Exploiting semantics for context-aware itinerary recommendation, Personal and Ubiquitous Computing 23 (2019) 215–231.
[30] M. Onori, A. Micarelli, G. Sansonetti, A comparative analysis of personality-based music recommender systems, in: CEUR Workshop Proceedings, volume 1680, CEUR-WS.org, Aachen, Germany, 2016, pp. 55–59.
[31] S. Caldarelli, D. F. Gurini, A. Micarelli, G. Sansonetti, A signal-based approach to news recommendation, in: CEUR Workshop Proceedings, volume 1618, CEUR-WS.org, Aachen, Germany, 2016, pp. 1–4.
[32] C. Biancalana, F. Gasparetti, A. Micarelli, A. Miola, G. Sansonetti, Context-aware movie recommendation based on signal processing and machine learning, in: Proceedings of the 2nd Challenge on Context-Aware Movie Recommendation, CAMRa '11, ACM, New York, NY, USA, 2011, pp. 5–10.
[33] A. Mensen, W. Marshall, G. Tononi, Eeg differentiation analysis and stimulus set meaningfulness, Frontiers in psychology 8 (2017) 1748.