Acta Scientific MEDICAL SCIENCES (ISSN: 2582-0931)

Volume 7 Issue 6 June 2023

Research Article

Emotion Recognition from EEG Signals Based on CapsNet Neural Network

Maryam Omri* and Wided Lejouad Chaari

LARIA Laboratory, National School of Computer Sciences (ENSI), University of Manouba, Tunisia

***Corresponding Author:** Maryam Omri, LARIA Laboratory, National School of Computer Sciences (ENSI), University of Manouba, Tunisia.

Received: September 28, 2022

Published: May 25, 2023

Abstract

Brain Computer Interfaces (BCI) are computer systems that capture and analyze brain signals. Out of the several techniques used for capturing brain signals, electroencephalogram (EEG) has been regarded as one of the most promising techniques. A lot of models have been investigated for extracting different emotions based on EEG signals. However, these models suffer from ignorance of spatial dimensions which indicate the location of each channel such as the asymmetry between the electrode pairs and other salient information related to emotional states. In this paper, a deep learning approach based on Fast Fourier Transform (FFT), Power Spectral Density (PSD) and a Capsule Network (CapsNet) is proposed. In order to represent the EEG signals better, we firstly apply FFT to extract the EEG frequency bands Delta, Thêta, Alpha, Bêta and Gamma. Then, Power Spectral Density (PSD) is introduced for describing the activation level of an EEG signal and improving the efficiency of emotion classification. Finally, features extracted are fed into CapsNet for the classification of emotional states. Experiments on the DEAP dataset show that the proposed method achieved 95% accuracy which improves the efficiency of capsule neural networks.

Keywords: EEG Signal; Power Spectral Density; Deep Learning; CapsNet; Emotion Recognition

Introduction

An emotional state is a psycho-physiological process triggered by the conscious and/or unconscious perception of an object or situation and is often associated with personality and motivation [1]. It is crucial to detect human emotions which are the source of communication between people to improve decision making and social interactions. The study of emotional recognition has major significance in medical fields by monitoring health status, indicating some neurological disorders, and providing the appropriate treatment [2]. Another considerable role which is in the company, perceiving the current psychological state of the employees is useful for crisis management and the enhancement of the throughput. Emotion recognition has deepened and become a significant research line in many fields, such as cognitive science, psychology, neuroscience, and artificial intelligence. Research methods of emotion recognition are divided into two types. The non-physiological signals such as gesture [3], speech [4] and facial expressions [5] and the physiological signals such as electroencephalograph EEG [6], magnetoencephalogram MEG [7] and electrocardiogram ECG [8]. The first category is subjective, controlled by a human. Therefore, the use of the second one guarantees the authenticity of emotions like EEG signals which are directly extracted from the cerebral cortex, so it can directly detect changes in human emotions.

In recent years, because of their objectivity and high accuracy of classification, EEG signals have been mostly used in the field of emotion recognition.

For this purpose, we applied in this paper a deep learning technique called the capsule neural network. The EEG features are extracted using FFT and PSD according to the electrode positions and then are fed into CapsNet to accomplish the classification tasks. The data is obtained from the DEAP international dataset [9].

The TensorFlow framework is employed to construct the model which is compared with other classification algorithms in terms of performance.

The content of this paper is organized as follows, we firstly made an overview of the studies applied to identify emotion states from EEG signals. Next, we introduced the DEAP dataset and described the different combinations of EEG features. Then, our approach based on the capsule neural network is detailed to get finally the experimental results compared with the state-of-the-art classifiers.

Related work

Various EEG studies have been proposed for this purpose using different methods of machine learning. [10] classified two different emotional states using EEG data recorded by an EEG headset, they used an artifact removal algorithm based on Stationary Wavelet Transform (SWT) and SVM classifier to achieve an average accuracy of 92%. [11] extracted features using Time Domain Features (TDF) and introduced them for three differents classifiers RIPPER, J4.8 (decision tree algorithm) and SVM to identify rules of emotion classification and the accuracies obtained were respectively 68.79%, 68.19%, and 60%. [12] aims to differentiate between the emotional states of calm, exciting positive and exciting negative using advanced machine learning techniques such as Random Forest and KNN that yielded respectively to an accuracy of 75.25% and 71.49%.

Other studies have applied deep learning techniques and they have proved their efficiency in emotion states identification. [13] proposed an hybrid model of deep learning called Parallel Convolutional Recurrent Neural Network. They Combined CNN and RNN to extract spatial and temporal features respectively and they achieved an average accuracy of 90.91%. [14] extracted the mean, median, maximum, minimum, standard deviation, variance, etc. values from participant readings of the DEAP dataset. They compared two different neural networks for the recognition of 2 emotion classes. They used firstly the DNN to obtain an accuracy of 74.53%. On the other hand, the CNN model was introduced and yielded to 77.38%. [15] applied the Pearson correlation coefficient to represent connectivity information between different channels of EEG by evaluating relevant information between two electrodes. In parallel, CNN was used, and the combination of the two parts was the input of a Softmax function for the final classification to obtain an average accuracy of 75.48%. In [16] the authors applied CapsNet on DEAP dataset to implement binary classification in addition to Bayesian Optimization for hyperparameters' settings which has led to an accuracy of 85.39%.

A comparison between all of these approaches is shown in table 1.

Few studies were interested in spatial dimensions of EEG channels, which may include salient information. Further, few papers have been limited to the electrode pairs asymmetry [17].

As a result, concentrating on the spatial dimension regarding the EEG signals and defining how to integrate it is a key problem. For this reason, we applied the capsules neural network CapsNet [18] that is responsible for training and learning the features to finally recognize the emotional states of the subject.

The contribution of this work consists in two main points: A combination of FFT, PSD, and CapsNet for emotion recognition based on EEG signals. Specifically, PSD is calculated for each frequency band and mapped for each electrode position. Thus, it offers significant information about emotion states in frequency and spatial dimensions.

Moreover, we deliver a new emotion representation which is more exhaustive than other studies giving 8 classes of emotions using the three axes of valence, arousal, and dominance.

Dataset

DEAP dataset

DEAP [9] is the international standard dataset for EEG-based emotion recognition. This dataset contains, for each of the 32 participants (16 male and 16 female subjects), 40 videos of 63 seconds each one with affective labels to match specific emotions. During each video, EEG signals were collected, then a self

Author	Dataset	Methods	Accuracy
Jalilifard., *et al.* [10]	Taken from experience	SWT + SVM	92%
Pane., *et al.* [11]	DEAP	RIPPER	68.79%
		Decision tree	68.19%
		SVM	60%
Giannakaki., *et al.* [12]	eNTERFACE Workshop 2006 + IAPS	Random Forest	75.25%
		KNN	71.49%
Yang., *et al.* [13]	DEAP	CNN + RNN	90.91%
Tripathi., *et al.* [14]	DEAP	DNN	74.53%
		CNN	77.38%
Wen., *et al.* [15]	DEAP	Pearson Correlation Coefficient + CNN	75.48%
Jana., *et al.* [16]	DEAP	CapsNet + Bayesian Optimization	85.39%

Table 1: Comparison of emotion recognition results between different studies.

reported emotional state based on self-assessment manikins [19] using a space of valence, arousal, and dominance was evaluated immediately after the video. Original data are sampled to 512 Hz. In this paper, pre-processed physiological data records (under-sampled to 128 Hz, with EOG deletion, etc.) are used.

The data matrix contains 8064 (128Hz * 63 seconds) physiological signals/EEG data from 40 different channels for each of the 40 experiments for each of the 32 participants. Thus, we have 412,876,800 massive readings to train our algorithm (8064 data × 40 channels × 40 experiments × 32 subjects).

Emotion recognition using DEAP dataset

The valence scale varies from sad to happy. The arousal scale varies from calm or bored to excited. The scale of dominance ranges from uncontrolled to dominant. To reduce subjectivity interference, we transformed classification tasks into binary classification using the following rules: If the emotion state value is less than or equal to 4.5, the label will be defined as "low" and will have the value 0. In contrast, if the value of the emotional state is greater than 4.5, the emotion state label will be set to "high" and will have the value 1.

Thus we obtained the following 6 groups: HV (High Valence), LV (Low Valence), HA (High Arousal), LA (Low Arousal), HD (High Dominance) and LD (Low Dominance).

According to the three-dimensional space model shown in figure 1 that maps different emotions of the 3 axes of valence, arousal, and dominance [20], we can divide the cube into 8 groups by dividing each axis into two parts. Therefore, we have the following 8 classes which are actually ordered from 0 to 7 according to their binary values : 'LVLALD', 'LVLAHD', 'LVHALD', 'LVHAHD', 'HVLALD', 'HVLAHD', 'HVHALD' and 'HVHAHD'.

Methodology

In this paper, we applied an approach which is an adapted CNN algorithm composed of features extraction using FFT and PSD and the classification based on the capsule neural network.

Features extraction

FFT The measured EEG signal can be analyzed simultaneously in the time dimension, frequency dimension and both, but the most widely used is the frequency one [21]. After preprocessing, the data were filtered using the Fast Fourier.

FFT [22] to extract 5 frequency bands Delta (under 4Hz), Thêta (4 - 7Hz), Alpha (8 - 15Hz), Bêta (16 - 31Hz) and Gamma (above 32Hz).

Figure 1: Emotions projection on the space model [20].

PSD EEG is classified as a nonstationary signal. The PSD unit is a very useful tool to know the frequencies and amplitudes of oscillatory signals in the chronological data [23]. We used the method of Welch to calculate the PSD [17]. The PSD is commonly used to describe the activation level of an EEG signal. For each subject, we calculate the PSD of each frequency band from the EEG signal of each electrode. A total of 200 PSD characteristics (40 channels × 5 frequency bands) is obtained.

In order to use the PSD values as an input for the CapsNets, they are transformed into a 40 × 40 topography based on the location of the EEG electrodes. Therefore, the PSD values are assigned to the electrode locations. Finally, the EEG signals of a single subject are presented in the form of 40×40×5 matrices.

Capsule neural network

Deep learning techniques perform better than other methods applied for extracting emotional states based on EEG signals [24]. Specifically, CNN that has been commonly used in recent years, presents the best classification results.

However, CNN does not decode the position and orientation of the object in its predictions. This internal data is discarded during the pooling layer.

To cover the shortfalls of CNN, a new neural network has been applied called capsule neural network. CapsNet is composed of capsules. A capsule is a set of artificial neurons representing the properties of a component and whose activity vector has two characteristics [25]. The length of this vector represents the presence probability of the object and the orientation forms its instantiation parameters (position, size, etc.). In case the object is shifted or resized the orientation of the capsule vector is changed but the length remains the same, unlike ConvNets. Each capsule determines a single component in the object, and all capsules form a tree with the hierarchy of detected features.

Convolution layer It is the first layer in this neural network, it aims to the detection of important features in the input data. Deeper layers detect simple features while upper layers focus on more complex features. The input of the convolutional layer is the signal matrices of PSD densities, the kernel is 9×9, it is larger than the convolution kernel of the CNN neural network, for the purpose of learning more information using fewer network layers, and the stride is 1 so that we use all the values of the input matrices when creating the feature maps.

Primary capsule layer The output of the convolution layer is sent to the primary capsule layer which detects the presence and orientation of a component from the input. At this level, the scalars of the feature maps are converted into vectors that represent the primary capsules. Thus, compared to CNN's scalar inputs and outputs, the vector inputs and outputs of the capsule networks may represent more features.

Convolution capsule layer It is the third layer that consists of capsules called routing capsules, which are used to detect larger and more complex objects. The convolution capsule layer updates the propagation of features based on the output vector of the primary capsule layer according to the dynamic routing algorithm between the primary capsules and the CapsNet convolution capsules. This algorithm, detailed in figure 2, replaces the pooling operation in a traditional CNN and embodies a degree of freedom in the learning mechanism of the neural network, giving the network the option to learn how to gather features without data loss. Each primary capsule has an activity vector u_i.

The length of u_i represents the probability of the existence of its corresponding component, and the orientation encodes various properties (e.g., size, position) of its corresponding component.

Table 2 describes the detailed process of the dynamic routing algorithm.

If u_i is the output of the lower-level capsule i, the prediction for the parent capsule j is calculated as follows :

$\hat{u}j|i = W_{ij}\, u_i$

In the above formula, W_{ij} is a weight matrix (also called a transformation matrix). The coupling coefficients c_{ij} (the real routing weights) are calculated using the Softmax function:

$$c_{ij} = \frac{\exp b_{ij}}{\sum_k \exp b_{ik}}$$

Figure 2: Architecture of the dynamic routing algorithm.

In the above formula, b_{ij} is the probability a priori that the capsule i is coupled to the parent capsule j, and that its initial value is set to 0.

The input vector of the parent capsule j is therefore calculated as follows:

$s_j = P_i\, c_{ij}\, \hat{u}j|i$

Finally, the real output v_j is calculated using a squash function as follows:

$$v_j = \frac{||s_j||^2}{1+||s_j||^2} \frac{s_j}{||s_j||}$$

With this squash function, as we need a probability, the capsule output is reduced to 0 when the capsule output vector is low, and increased to 1 when the capsule output vector is high. The probabilities b_{ij} are updated using n routing iterations using the following equation: $b_{ij} = b_{ij} + \hat{u}j|i.v_j$

According to length and orientation of $\hat{u}_{j|i}$, when the capsules of the lower level and the capsules of the higher level conform to their predictions, the value of c_{ij} increases. Thus, the sub-capsule outputs are sent to the correct parent.

In the above formula, L_k is the loss of each class k. If the input data matches the label, the value T_k is set to 1 and 0 otherwise. According to the experiences the best values of the hyperparameters m^+ and m^- are set to 0.9 and 0.1, respectively. λ is used to reduce the effect of loss on labels that do not belong to the correct class and is equal to 0.5.

Procedure ROUTING ($\hat{u}_j$|i,r,l)
For each capsule i in layer l and capsule j in layer (l +1) : $b_{ij} \leftarrow 0$
For r iterations do
For each capsule i in layer l :
$c_i \leftarrow$ softmax(b_i)
For each capsule j in layer (l +1) :
sj ←Pi cij uˆj|i ; ˆuj|i = Wij ui
For each capsule j in layer (l +1) : $v_j \leftarrow$ squash(s_j)
For each capsule i in layer l and capsule j in layer (l +1) :
bij ← bij + uˆj|i .vj
Return v_j

Table 2: Dynamic routing algorithm capsules. Finally, the loss function of CapsNet (the marginal loss) is defined as: $L_k = T_k \max(0,m^+ - ||v_k||)^2 + \lambda(1 - T_k) \max(0,||v_k|| - m^-)^2$.

Decoder It is the last layer in this neural network which is similar to the fully connected layer in CNN responsible for the classification of the input data, it is composed of two Relu layers that are easier to calculate so that the model can take less time for training and test, and is followed by a Sigmoid layer which exists between 0 and 1 and is therefore particularly used for models in which we need to predict the probability as an output.

The detailed architecture of the approach applied in this paper is presented in figure 3.

Performance of CapsNet

Capsule neural network resolves the problem of position information loss as for CNN in the process of pooling layer thanks to the routing algorithm. CapsNet represents the relative location states and relationships between all components of an object. This particularity is essential in our work since we need the location source of each signal because each electrode emits well-defined frequency bands and different regions of the cerebral cortex have corresponding responses. In addition, this neural network forms a tree that presents the hierarchy of detected.

Figure 3: Architecture of the approach used in this paper.

Features by encoding the relative connections between local parts and whole objects. Therefore, using CapsNet for classification reduces the error rate obtained with CNN.

Results and Discussions

This paper uses the DEAP dataset that contains a large quantity of data. In order to allow our neural model to train effectively and quickly on such massive data, we reduced the dimensionality of our data by dividing it into 40 batches. We trained 3 routings of the algorithm to achieve an emotional classification accuracy of 95% with a confidence interval of 0.00012 and a margin loss of 0.01.

In order to evaluate the performance of the CapsNet, we compared this network to common classifiers such as SVM, CNN, KNN, etc. All of the EEG data used in the different experiments are derived from the same DEAP dataset. The results are presented in figure 4.

As shown in figure 4, CapsNet's average emotional classification accuracy is better than other classifiers. It can be seen that the network can better recognize the emotional characteristics of EEG signals, thereby identifying different emotional states. Compared to CapsNet, CNN, K-NN, SVM, etc. could not preserve essential spatial information between all channels and bands. Therefore, their accuracies were lower. The results show that spatial differences between channels in different frequency bands could provide considerable information on emotional states, and therefore we have retained and exploited them.

Figure 4: Comparison of average emotional classification accuracy percentage between CapsNet classifier and other classifiers using the same dataset.

In addition, the other papers that used the DEAP dataset, used only the axes of valence and arousal to have 4 classes HVHA, HVLA, LVHA, LVLA, whereas, in this paper, we used the 3 axes to have 8 classes and therefore a more accurate representation of emotions.

Conclusion

In this paper, the capsule neural network accompanied by the routing algorithm, this too recent approach is proposed to extract different emotions from EEG signals. To evaluate the performance and feasibility of our method, several experiments were conducted to achieve finally an accuracy of 95%. The results show that capsule neural network can effectively recognize emotional information better than other methods by maintaining spatial information. At the same time, the proposed combination in this paper improves the classification to achieve better results.

In the future work, we will test our method on other datasets, other physiological signals such as MEG and ECG to verify its effectiveness, we can adjust and define the better parameters and hyperparameters to obtain higher accuracy of emotion recognition.

Data Availability Statement

The data that support the findings of this study are available upon reasonable request on http://www.eecs.qmul.ac.uk/mmv/datasets/deap/ due to privacy.

Bibliography

1. Nancy L Stein and Linda J Levine. "Thinking about feelings: The development and organization of emotional knowledge". In Aptitude, learning, and instruction (2021): 165-198.
2. Sharmeen M., *et al*. "Multimodal emotion recognition using deep learning". *Journal of Applied Science and Technology Trends* 2.2 (2021): 52-58.
3. Patrick D Ross., *et al*. "Developmental changes in emotion recognition from full-light and point-light displays of body movement". *PLoS One* 7.9 (2012): e44815.
4. Monita Chatterjee., *et al*. "Voice emotion recognition by cochlear-implanted children and their normally-hearing peers". *Hearing Research* 322 (2015): 151-162.
5. Ralph Adolphs., *et al*. "Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala". *Nature* 372.6507 (1994): 669-672.

6. Zhong Yin., *et al*. "Recognition of emotions using multimodal physiological signals and an ensemble deep learning model". *Computer Methods and Programs in Biomedicine* 140 (2017): 93-110.

7. Mojtaba Khomami Abadi., *et al*. "Decoding affect in videos employing the meg brain signal". In 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG) (2013): 1-6.

8. Foteini Agrafioti., *et al*. "Ecg pattern analysis for emotion detection". *IEEE Transactions on Affective Computing* 3.1 (2011): 102-115.

9. Sander Koelstra., *et al*. "Deap: A database for emotion analysis; using physiological signals". *IEEE Transactions on Affective Computing* 3.1 (2011): 18-31.

10. Amir Jalilifard., *et al*. "Emotion classification using single-channel scalp-eeg recording". In 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (2016): 845-849.

11. Evi Septiana Pane., *et al*. "Identifying rules for electroencephalograph (eeg) emotion recognition and classification". In 2017 5th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME) (2017): 167-172.

12. Katerina Giannakaki., *et al*. "Emotional state recognition using advanced machine learning techniques on eeg data". In 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS) (2017): 337-342.

13. Yilong Yang., *et al*. "Emotion recognition from multi-channel eeg through parallel convolutional recurrent neural network". In 2018 International Joint Conference on Neural Networks (IJCNN) (2018): 1-7.

14. Samarth Tripathi., *et al*. "Using deep and convolutional neural networks for accurate emotion classification on deep dataset". In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (2017): 4746-4752.

15. Zhiyuan Wen., *et al*. "A novel convolutional neural networks for emotion recognition based on eeg signal". In 2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC) (2017): 672-677.

16. Gopal Chandra Jana., *et al*. "Capsule neural networks on spatio-temporal eeg frames for cross-subject emotion recognition". *Biomedical Signal Processing and Control* 72 (2022): 103361.

17. Nattapong Thammasan., *et al*. "Continuous music-emotion recognition based on electroencephalogram". *IEICE TRANSACTIONS on Information and Systems* 99.4 (2016): 1234-1241.

18. Sara Sabour., *et al*. "Dynamic routing between capsules". In Advances in neural information processing systems (2017): 3856-3866.

19. Jon D Morris. "Observations: Sam: the self-assessment manikin; an efficient cross-cultural measurement of emotional response". *Journal of Advertising Research* 35.6 (1995): 63-68.

20. Gyanendra K Verma and Uma Shanker Tiwary. "Affect representation and recognition in 3d continuous valence-arousal-dominance space". *Multimedia Tools and Applications* 76.2 (2017): 2159-2183.

21. Robert Jenke., *et al*. "Feature extraction and selection for emotion recognition from eeg". *IEEE Transactions on Affective Computing* 5.3 (2014): 327-339.

22. Zhong Yin., *et al*. "Cross subject eeg feature selection for emotion recognition using transfer recursive feature elimination". *Frontiers in Neurorobotics* 11 (2017): 19.

23. Wei-Long Zheng., *et al*. "Eeg-based emotion classification using deep belief networks". In 2014 IEEE International Conference on Multimedia and Expo (ICME) (2014): 1-6.

24. Ghulam Muhammad., *et al*. "Automatic seizure detection in a mobile multimedia framework". *IEEE Access* 6 (2018): 45372-45383.

25. GUO Jinliang., *et al*. "Eeg emotion recognition based on granger causality and capsnet neural network". In 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS) (2018): 47-52.