Journal of Electronic Research and Application, 2024, Volume 8, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

An Affective EEG Analysis Method Without Feature Engineering

Jian Zhang*, Chunying Fang, Yanghao Wu, Mingjie Chang

Heilongjiang University of Science and Technology, Harbin 150022, Heilongjiang Province, China

Corresponding author: Jian Zhang, zhangmingjian180@qq.com

Abstract: Emotional electroencephalography (EEG) signals are a primary means of recording emotional brain activity. Currently, the most effective methods for analyzing emotional EEG signals involve feature engineering and neural networks. However, neural networks possess a strong ability for automatic feature extraction. Is it possible to discard feature engineering and directly employ neural networks for end-to-end recognition? Based on the characteristics of EEG signals, this paper proposes an end-to-end feature extraction and classification method for a dynamic self-attention network (DySAT). The study reveals significant differences in brain activity patterns associated with different emotions across various experimenters and time periods. The results of this experiment can provide insights into the reasons behind these differences.

Keywords: Dynamic graph classification; Self-attention mechanism; Dynamic self-attention network; SEED dataset

Online publication: January 18, 2024

1. Introduction

The brain, the most complex organ in the human body, has prompted the development of various methods to detect its activity, including electroencephalography (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and more. Among these, EEG has emerged as a primary method for studying brain science due to its high time resolution, cost-effectiveness, and non-invasive characteristics. Emotion, being a conspicuous brain activity, can be objectively reflected through EEG signals. Research on emotion recognition based on EEG signals aims to unveil hidden information within the original EEG data through understanding the relationship between data and tags, thereby discovering the objective law governing brain activity.

Given the high noise and non-stationary characteristics of EEG signals, effective feature extraction methods are crucial for signal classification. Traditional feature extraction methods, such as feature engineering, typically involve preprocessing the data, extracting features (e.g., differential entropy, DE; power spectral density, PSD), and finally employing a simple classifier (e.g., support vector machine, SVM) for classification. Zheng *et al*. utilized DE features and SVM classifiers, achieving an accuracy rate of 83.99% in the full

frequency band using the Shanghai Jiao Tong University (SJTU) Emotion EEG Dataset (SEED) [1]. Additionally, combining DE features with differential asymmetry (DASM), rational asymmetry (RASM), and traditional frequency spectrum (ES) features on symmetrical electrodes proved beneficial [2].

While feature engineering offers good interpretability, directly describing highly abstract features in mathematical language can be challenging. With the recent advancements in deep learning, these features are often extracted automatically with the assistance of neural networks. Zheng *et al*. employed deep belief networks (DBNs), attaining an accuracy of 86.08% across the full frequency band of SEED [1]. Considering the non-Euclidean arrangement of electrode positions in the electrode cap, Song *et al*. applied graph neural networks (GNNs) to emotion recognition, achieving an accuracy of 87.40% by combining graph convolutional neural networks (GCNNs) with DE [3]. Building upon this, the proposed dynamic graph convolutional neural networks (DGCNNs) achieved an accuracy of 90.40%. Furthermore, Liu *et al*. used a three-dimensional convolutional neural network (3D-CNN) based on attention in the time, space, and frequency domains to capture key information between different dimensions of EEG features [4]. Although the accuracy rate reached 96.28%, it is noteworthy that the sample used was 9 seconds, while the aforementioned models typically operated on 1-second intervals.

The hybrid feature extraction method, combining feature engineering and deep learning, has reached a high level. However, given the potent feature extraction capabilities of deep learning, a pertinent question arises: is it possible to set aside feature engineering and directly engage in end-to-end emotion recognition? Kim *et al*. highlighted that EEG signals collected in experimental environments are subject to numerous interference factors, including temporal background, spatial elements, race, language, and culture, resulting in various interference noises in EEG signals [5]. Nevertheless, the end-to-end pattern recognition method is more succinct and offers greater room for development. This holds significant methodological importance for the evolving field of brain interfaces [6].

This paper is based on the dynamic self-attention network (DySAT) proposed by Aravind *et al*. and introduces some enhancements to apply it to emotion recognition in EEG signals [7]. The specific improvements are outlined as follows (https://github.com/zhangmingjian180/DySAT_SEED):

(1) A global pooling layer is added to facilitate a smoother decline in the number of hidden layer units.
(2) The original unidirectional attention in the temporal attention layer is modified to bidirectional attention, enabling the capture of characteristics both before and after.
(3) The sigmoid activation function is replaced with the leaky rectified linear unit (Leaky ReLU) to enhance convergence speed [8].
(4) The residual mechanism is incorporated to stabilize the inter-layer transmission of information [9].

The subsequent sections of this paper unfold as follows: Section 2 briefly introduces the theory related to the model. Section 3 formally presents the DySAT model. Section 4 provides a detailed description of the experiment using DySAT, along with an analysis of the experimental results. Section 5 offers a discussion and a summary of the research in this paper.

2. Related work

The DySAT used in this paper mainly involves two parts: graph attention networks (GATs) and self-attention mechanism [10,11].

2.1. Graph attention network

GAT is essentially a convolution operation, but it is different from the ordinary convolution neural network

(CNN). The latter is based on grid format data, while the former is based on graph, which is a kind of non-Euclidean geometric data. Its formal definition is as follows:

$$G = (V, E) \qquad (1)$$

where V and E represent the set of points and the set of edges. Let the input node eigenvector set be $H = \{h_1, h_2, \ldots, h_n\}$, $h_i \in R^F$, the output node eigenvector set be $H' = \{h'_1, h'_2, \ldots, h'_n\}$, $h'_i \in R^{F'}$, and the parameter set be P, then:

$$H' = \mathrm{GAT}(G, H, P) \qquad (2)$$

2.2. Self-attention mechanism

Self-attention mechanism can be used to extract temporal information from sequences. Previously, this field has been occupied by recurrent neural networks (RNNs). The latest progress of many natural language processing tasks has proved the superiority of the self-attention mechanism in efficiency and performance [11-13]. In principle, GAT is also an example of applying an attention mechanism to graphs.

3. Dynamic graph self-attention network

DySAT is mainly composed of a structural self-attention layer, a sequential self-attention layer, and a global pooling layer. Its overall architecture is shown in **Figure 1**.

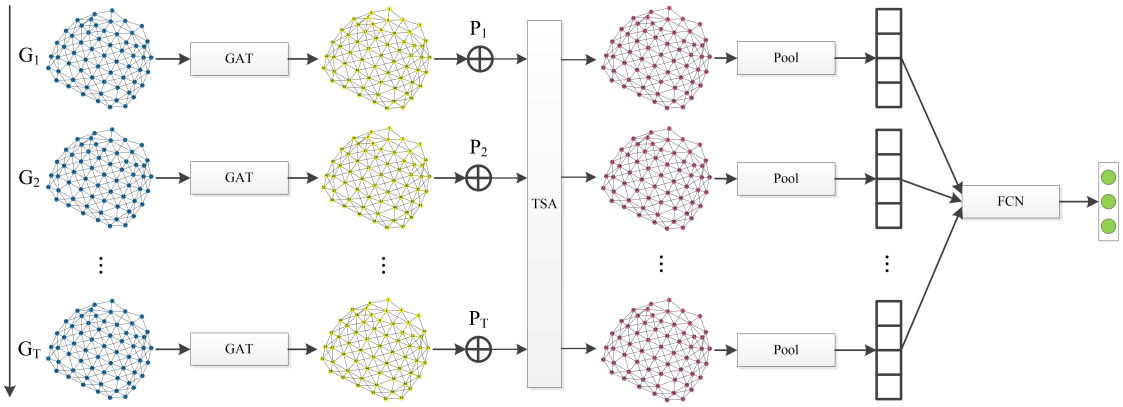

Figure 1. The architecture of DySAT, where GAT represents the graph attention network, TSA represents the tensorized self-attention, and FCN represents the fully connected network

3.1. Model inputs and outputs

The input of the model is a dynamic graph , which is essentially a group of graph sequences, and its formal definition is:

$$D = \{G_1, G_2, \ldots, G_T\} \qquad (3)$$

where T represents the length of the graph sequence. For any $i < j \leqslant T$, meet $G_i.V \subseteq G_i.V$. For a graph G, it can be either a directed graph or an undirected graph. It can also be either a weighted graph or a binary graph.

The output of the model is an n-dimensional vector, and each dimension of the vector represents the corresponding category possibility of the dynamic graph.

3.2. Structural self-attention layer

The structural self-attention layer is used to extract the spatial structure information from dynamic graphs. For the input dynamic graph D, each graph G_i ($1 < i \leqslant T$) in D undergoes the structure self-attention layer pair,

performing the GAT operation respectively. Consequently, the output of this layer is a dynamic graph D'. The specific calculation processes are as follows:

$$z_v = (\sum_{u \partial N_v} \alpha_{uv} W^s x_u) \qquad (4)$$

$$\alpha_{uv} = \frac{\exp(e_{uv})}{\sum_{w \in N_v} \exp(e_{wv})} \qquad (5)$$

$$e_{uv} = Leaky\ ReLU\ (A_{uv} \cdot a^T [W^s x_u \parallel W^s x_v]) , \forall \langle u, v \rangle \in E \qquad (6)$$

where u,v,w indicate the node number and A_{uv} represents the weight of the edge $\langle u,v \rangle$. For a binary graph, its weight is 1. $N_v = \{v \in V: \langle u,v \rangle \in E\}$. Parameters W^s and α are to be learned for neural networks. *ELU* stands for exponential linear units, *LeakyReLU* stands for leaky rectified linear units. x_u, z_v represents the eigenvector of a node u,v.

To better capture the potential features of different types, the multi-head attention mechanism was used, namely:

$$z_v \leftarrow concat(z_v^1, z_v^2, ..., z_v^{HS}), \forall v \in V \qquad (7)$$

where H_s is the number of heads.

3.3. Sequential self-attention layer

The temporal self-attention layer is used to extract the temporal variation characteristics of nodes. For the input dynamic graph D, temporal self-attention performs the following operations on each node v. The input was set as $X_v = [x^1_v, x^2_v, \ldots x^T_v]^T$, where $x^t_v \in R^{D'}$, whereas the output was set as $Z_v = [z^1_v, z^2_v, \ldots z^T_v]^T$, where $z^t_v \in R^{F'}$. The specific calculation processes are as follows:

Initially, add location embedding:

$$Z_v \leftarrow Z_v + P \qquad (8)$$

where p are the parameters learned through the backpropagation algorithm.

Subsequently, add the self-attention process:

$$Z_v = \beta_v X_v W_v \qquad (9)$$

$$\beta_v^{ij} = \frac{\exp(e_v^{ij})}{\sum_{k=1}^{T} \exp(e_v^{ik})} \qquad (10)$$

$$e_v^{ij} = \frac{(X_v W_q)(X_v W_k)^T}{\sqrt{F}} \qquad (11)$$

where W_v, W_α, W_k are the parameters learned by the backpropagation algorithm.

Here, the multi-head attention mechanism was also used to better capture the potential features of different types:

$$Z_v \leftarrow concat(Z^1_v, Z^2_v, \ldots, Z^{HT}_v), \forall_v \in V \qquad (12)$$

where H_T is the number of heads.

3.4. Global pooling layer

The global pooling layer can simply pool the extracted features, thereby reducing the noise. Simultaneously, it can also smoothen the transformation range of the number of neurons between layers. For the input dynamic

graph D, each graph $G_i(1 \leq i \leq T)$ in the graph D undergoes the pool calculation. The output of this layer $O = [o_1, o_1, ..., o_T\}$. The calculation process of the pool is as follows:

$$o_i = mean\left(h_j \middle| \forall vj \in Gi.V\right) \quad (13)$$

3.5. Loss function

The model is a multi-classification model. Therefore, the classical cross-entropy function was used as the loss function. The calculation expression is as follows:

$$L = -\frac{1}{C}\sum_{i=1}^{C} y_1 \log\left(\frac{e^{\hat{y}_i}}{\sum_j^c e^{\hat{y}_j}}\right) \quad (14)$$

wherein, C represents the number of categories, y represents the sample label, and $\hat{y}$ represents the model output value, both of which are in the form of independent hot coding.

4. Experiment

In this section, the relevant experiments were described using DySAT on SEED [1], followed by an analysis of the experimental results.

4.1. Dataset description

SEED is a public dataset created by the team led by Prof. Bao-Liang Lu and Prof. Wei-Long Zheng in the Brain-Like Computing and Machine Intelligence (BCMI) Laboratory of SJTU. The team utilized 62-lead equipment to sample EEG data from 15 subjects three times each. The sampling frequency was 200 Hz, and a band-pass frequency filter of 0.75 Hz was applied. Subjects were required to watch 15 positive, negative, or neutral movie clips during each sampling session. Before each clip, a 5-second silence period occurred, followed by a ~4-minute clip. Subsequently, subjects conducted a self-assessment for about 45 seconds. After a 15-second rest, subjects proceeded to the next movie clip.

4.2. Introduction to the experiment

Song *et al*. achieved a 90.40% accuracy on the SEED using DGCNN combined with DE features, with a sample length of 1 second [3]. In essence, using 1 second as the sample length theoretically allows for information in the sample to reach at least 90.40% accuracy. Song *et al.*'s study was taken as the theoretical basis for designing the experimental scheme for this study.

Subject-dependent experiments were conducted in this study. The EEG data of each trial for each subject serves as the data unit, resulting in 45 experiments. In each experiment, the EEG data of subjects watching the first nine movie clips form the training set, the next three comprise the validation set, and the final three constitute the testing set. For each emotion, the ratio of training set, validation set, and testing set is 9:3:3.

After dividing the dataset, the data was proceeded for slicing with a length of 1 second. To increase the data volume, the step size of the slice window was set to 0.2 seconds, equivalent to 40 points. The sliced sample data cannot be directly input into the model because the model's input data is a dynamic graph. Therefore, the sample data was converted into a dynamic graph, as described in the subsection. Following this conversion, the data was input into the DySAT model for training and testing, ultimately yielding the experimental results.

4.3. Data preprocessing

In this section, the conversion of the sample data from EEG form to dynamic graph form was introduced, as illustrated in **Figure 2**.

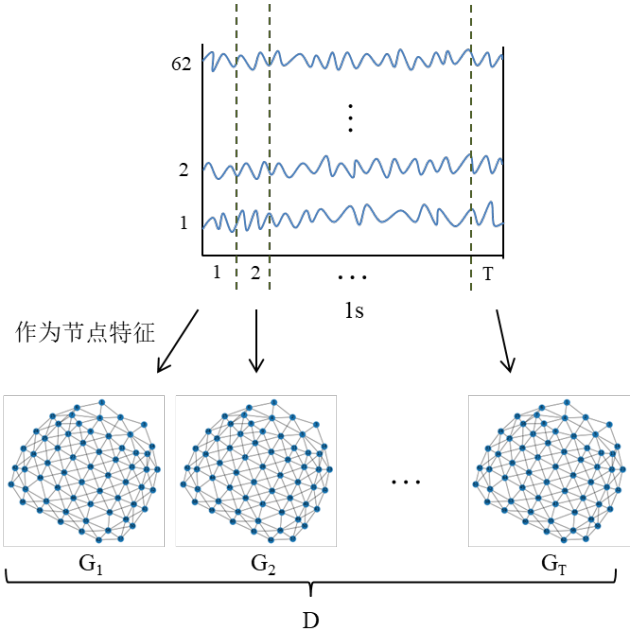

Figure 2. Converting EEG into a dynamic graph

(1) Construction of graphs: This involves creating a binary undirected graph, as depicted in **Figure 3**. In this graph, nodes represent electrodes, and edges indicate adjacency between two electrodes. Consequently, electrode position information is recorded.

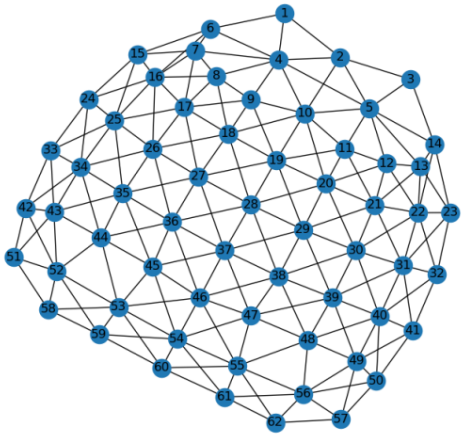

Figure 3. Schematic diagram of the connection between 62 EEG channels, which is used to construct the adjacency matrix of the diagram

(2) Node characteristics: The potential recorded by the electrode was utilized as the node feature. Initially, the sample data was segmented, considering each "small piece" as the state representation of the EEG signal at a specific moment. The reason for this choice, instead of using "every moment" in the real sense, is primarily because EEG signals do not change significantly in a short period of time. Therefore, using a short signal duration better represents the state of the EEG signal. Through repeated experiments, the optimal length for the small signal was determined to be 0.05 seconds, equivalent to

10 data points. This yields a graph sequence length (T) of 20.

(3) Data normalization: Given the substantial difference between the maximum and minimum values of the original potential, direct use as input for the model is impractical, necessitating normalization. According to statistical analysis of EEG signals, the amplitude of EEG signals primarily concentrates between -100 μV to 100 μV. For instance, in the EEG data collected from subject 1 during the first experiment, the highest potential was 14,641.17 μV, while the lowest potential was -12,604.62 μV. The amplitude falls within -100–100 μV, with a ratio of V to V of 0.961. The potential distribution, shown in **Figure 4(a)**, resembles a normal distribution. Similarly, by examining the function image of $f(x) = \tanh(x)$, it was observed that $f(x)$ is approximately linear on (-1,1). To preserve data characteristics to the greatest extent, this paper employs the following formula for data normalization:

$$f(x) = \tanh\left(\frac{x}{100}\right) \quad (15)$$

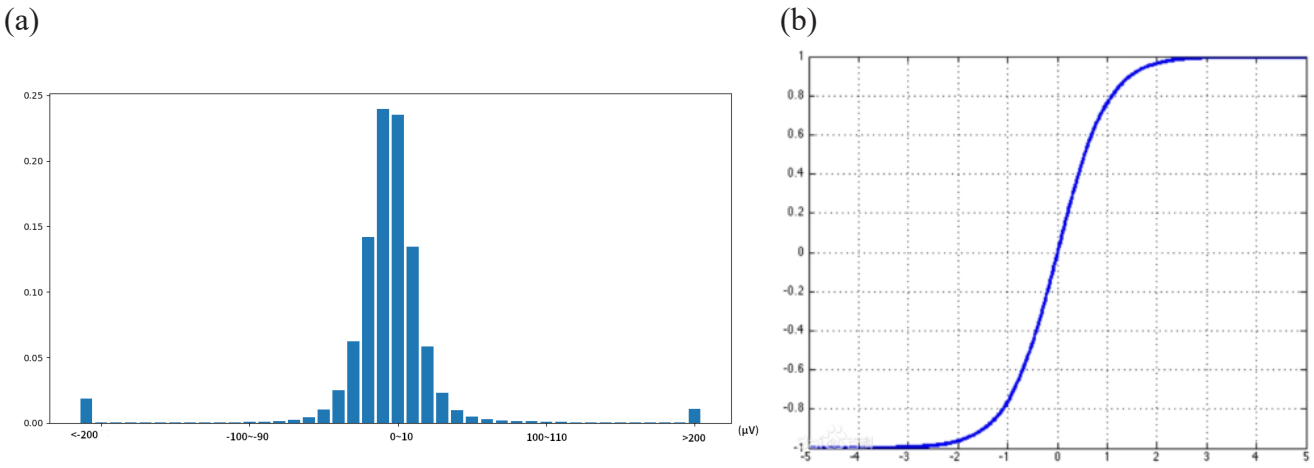

Figure 4. (a) Potential distribution; **(b)** function tanh(x) of subject 1 in the first experiment

4.4. Experimental results and analysis

In deep learning, the configuration of hyperparameters significantly influences model performance. After continuous refinement, a set of approximate optimal hyperparameters consistent with the current experiment was identified. The details are summarized in **Table 1**.

Table 1. Hyperparameters used in the experiment

Name	Value	Explanation
Structural head	2,2	Attention heads in each structural layer
Structural layer	8,8	Units in each structural layer
Temporary head	2,2	Attention heads in each temporary layer
Temporary layer	8,8	Units in each temporary layer
Optimizer	AdamW	Optimizer used in gradient descent
Learning rate	0.001	Initial learning rate for self-attention model
Weight decay	0.000001	This parameter is used in AdamW

Following 45 experiments, the results are presented in **Table 2**. The average accuracy achieved using DySAT as a classifier is 0.6215, markedly higher than random discrimination (with an accuracy of 0.3333 for three classifications). This demonstrates that the end-to-end DySAT, based on a dynamic GNN, can autonomously extract valuable features hidden in EEG data.

Table 2. Experimental results

Subject	Session 1	Session 2	Session 3	AVG	STD
1	0.7224	0.4295	0.4032	0.5184	0.1447
2	0.4041	0.4401	0.5262	0.4568	0.0512
3	0.4960	0.4750	0.3615	0.4442	0.0591
4	0.6796	0.6222	0.4044	0.5687	0.1185
5	0.5202	0.6725	0.6066	0.5998	0.0623
6	0.8655	0.6370	0.7375	0.7467	0.0935
7	0.3432	0.7999	0.3985	0.5138	0.2035
8	0.6544	0.6326	0.8998	0.7289	0.1211
9	0.8034	0.7724	0.5480	0.7080	0.1138
10	0.6240	0.6205	0.5740	0.6062	0.0228
11	0.7224	0.6367	0.8688	0.7426	0.0958
12	0.5398	0.4221	0.7349	0.5656	0.1290
13	0.5380	0.7851	0.8176	0.7136	0.1249
14	0.6559	0.6488	0.5738	0.6262	0.0372
15	0.8274	0.6834	0.8377	0.7828	0.0704
All	All	0.6215	0.1511		

Considering all sessions for all subjects, the standard deviation of the classification results is 0.1511. Across the 45 experiments, the lowest accuracy rate was 0.3432 (Session 1 of subject 7), and the highest accuracy rate was 0.8998 (Session 3 of subject 8). These data reveal significant variations in classification results among different subjects and sessions. Comparing the data in the STD column, it is evident that, except for subject 6, the standard deviation of other subjects is less than the overall average. This indicates that the differences in EEG data between different sessions of the same subject are smaller than those between different sessions of different subjects, emphasizing that subject variation is a crucial factor influencing the characteristics of EEG data.

To align the research in this paper with both domestic and international studies, some data from the experimental results were compared. It is important to note that due to different data preprocessing methods and experimental perspectives (single-mode and multi-mode), the accuracy of experimental results alone cannot conclusively prove the model's performance. Nonetheless, it can offer a reference. **Table 3** indicates that the current effectiveness of emotion recognition algorithms based on feature engineering is not particularly ideal, leaving ample room for improvement. The DySAT model employed in this paper stands out as a superior algorithm among similar approaches.

Table 3. Comparison of similar experiments

Models	Mean	STD	Class	Database
MLP [6]	0.31	0.03	4	AMIGOS
Encoder [6]	0.50	0.01	4	ASCERTAIN
CNN-LSTM [6]	0.37	0.01	4	DECAF
This study	0.62	0.15	3	SEED

5. Summary and discussion

This paper improved DySAT and applied it to emotion recognition in EEG data, demonstrating the model's effectiveness. The experimental results reveal significant variations among different subjects and sessions. It is important to note that these differences are not solely attributable to the model. For instance, in the experiment conducted by Song *et al*. [3], the result was 90.40/8.49, indicating some disparities. Therefore, the primary reasons for the considerable differences in this experiment can be preliminarily attributed to two factors.

Firstly, the inherent nonstationarity of EEG data results in inevitable data differences. However, in theory, appropriately increasing the length of sample data can mitigate this difference. Increasing the sample length theoretically enhances the amount of effective information. Yet, it should be acknowledged that larger samples pose a challenge for the classification method, requiring the handling of a larger volume of data. Balancing these aspects may become a key research direction.

Secondly, differences arise due to limitations within the model itself. As known, the strength of deep learning lies in the "depth" of the model, as deeper models theoretically extract more abstract features. The DySAT proposed in this paper comprises only six layers, which could contribute to the observed difference. It is worth noting that constructing deep GNNs remains an unsolved problem in the field. Anticipated development in GNNs is expected to address this issue.

It is important to recognize that the DySAT proposed in this paper is not the optimal classification model, and there exists a gap between the hybrid feature extraction method, combining feature engineering and deep learning (see Section 2 for details). However, this gap is not fixed, given that deep learning is grounded in the universal approximation theorem [14]. Therefore, the model holds significant potential for improvement. This paper aims to demonstrate the model's effectiveness, leaving the avenue open for substantial performance enhancement in future research.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Zheng W-L, Lu B-L, 2015, Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. IEEE Transactions on Autonomous Mental Development, 7(3): 162–175. https://doi.org/10.1109/TAMD.2015.2431497

[2] Duan R-N, Zhu J-Y, Lu B-L, 2013, 6th International IEEE/EMBS Conference on Neural Engineering (NER), November 6–8, 2013: Differential Entropy Feature for EEG-Based Emotion Classification. IEEE, San Diego, 81–84.

https://doi.org/10.1109/NER.2013.6695876

[3] Song T, Zheng W, Song P, et al., 2020, EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks. IEEE Transactions on Affective Computing, 11(3): 532–541. https://doi.org/10.1109/TAFFC.2018.2817622

[4] Liu J, Zhao Y, Wu H, et al., 2021, 13th Asia Pacific Signal and Information Processing Association (APSIPA) Annual Summit and Conference, December 14–17, 2021: Positional-Spectral-Tempotal Attention in 3D Convolutional Neural Networks for EEG Emotion Recognition. APSIPA, Tokyo, 305–312.

[5] Kim J, André E, 2008, Emotion Recognition Based on Physiological Changes in Music Listening. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(12): 2067–2083. https://doi.org/10.1109/TPAMI.2008.26

[6] Dzieżyc M, Gjoreski M, Kazienko P, et al., 2020, Can We Ditch Feature Engineering? End-to-End Deep Learning for Effect Recognition from Physiological Sensor Data. Sensors (Basel), 20(22): 6535. https://doi.org/10.3390/s20226535

[7] Sankar A, Wu Y, Gou L, et al., 2020, The 13th ACM International Conference on Web Search and Data Mining, February 3–7, 2020: DySAT: Deep Neural Representation Learning on Dynamic Graphs via Self-Attention Networks. Association for Computing Machinery, Houston, 519–527. https://doi.org/10.1145/3336191.3371845

[8] Maas AL, Hannun AY, Ng AY, 2013, Proceedings of the 30th International Conference on Machine Learning, June 17–19, 2013: Rectifier Nonlinearities Improve Neural Network Acoustic Models. JMLR, Atlanta, vol. 28, 3.

[9] He K, Zhang X, Ren S, et al., 2016, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27–30, 2016: Deep Residual Learning for Image Recognition. IEEE, Las Vegas, 770–778. https://doi.org/10.1109/CVPR.2016.90

[10] Veličković P, Cucurull G, Casanova A, et al., 2018, 6th International Conference on Learning Representations, April 30–May 3, 2018: Graph Attention Networks. ICLR, Vancouver. https://doi.org/10.48550/arXiv.1710.10903

[11] Vaswani A, Shazeer N, Parmar N, et al., 2017, 31st Conference on Neural Information Processing Systems (NIPS 2017), December 4–9, 2017: Attention Is All You Need. Neural Information Processing Systems Foundation, Long Island. https://doi.org/10.48550/arXiv.1706.03762

[12] Shen T, Jiang J, Zhou T, et al., 2018, DiSAN: Directional Self-Attention Network for RNN/CNN-Free Language Understanding. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1): 5446–5455. https://doi.org/10.1609/aaai.v32i1.11941

[13] Tan Z, Wang M, Xie J, et al., 2018, Deep Semantic Role Labeling with Self-Attention. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1): 4929–4936. https://doi.org/10.1609/aaai.v32i1.11928

[14] Hornik K, Stinchcombe M, White H, 1989, Multilayer Feedforward Networks Are Universal Approximators. Neural Networks, 2(5): 359–366. https://doi.org/10.1016/0893-6080(89)90020-8