DOI: 10.53469/wjimt.2025.08(05).12

Research on the Progress of Deep Learning in Emotion Classification of Electroencephalogram Signals

Haojie Ju

High school student, Aquinas International Academy, Ontario, CA, USA

Abstract: ***Emotion classification of electroencephalogram (EEG) signals is currently a research hotspot in the intersection of brain science and artificial intelligence, which is of great significance for revealing the mechanism of human emotions and achieving human-computer emotional interaction. In recent years, deep learning technology has made remarkable progress in the emotion classification of electroencephalogram (EEG) signals. This paper reviews the current research status of deep learning in the emotion classification of electroencephalogram (EEG) signals, analyzes the application characteristics and advantages of common deep learning models in this field, discusses key technologies such as data preprocessing, feature extraction, and model optimization, and looks forward to future research directions, aiming to provide a reference for the further development of this field.***

Keywords: Deep Learning; EEG; Emotion Classification; Convolutional Neural Networks; Recurrent Neural Networks.

1. INTRODUCTION

Emotions, as an important component of human psychological activities, have a profound impact on an individual's cognition, behavior and health. Electroencephalogram (EEG) signals, as a direct reflection of the electrical activity of brain neurons, contain rich emotional information [1]. Therefore, the research on emotion classification based on electroencephalogram (EEG) signals has significant application value in fields such as psychology, neuroscience, and human-computer interaction. In recent years, with the rapid development of deep learning technology, significant progress has been made in its application in the emotion classification of electroencephalogram (EEG) signals [2].

Traditionally, the emotion classification of electroencephalogram (EEG) signals mainly relies on manually designed feature extraction methods and shallow machine learning models, such as Support Vector Machine (SVM), K-nearest Neighbor (KNN), etc. [3]. However, these methods often have difficulty capturing the complex patterns and temporal dependencies in electroencephalogram (EEG) signals, resulting in limited classification performance. Deep learning technology, by constructing multi-level neural network models, can automatically learn the complex feature representations in electroencephalogram (EEG) signals, providing a new solution for emotion classification [4].

Specifically, convolutional neural networks (CNNS) achieve local feature extraction and global feature integration of electroencephalogram (EEG) signals through the combination of convolutional layers, pooling layers, and fully connected layers [5]. Recurrent neural networks (RNN) and their variants (such as Long Short-Term Memory Networks (LSTM) and gated recurrent units (GRU)) introduce recurrent connections, enabling the networks to process sequential data and capture the dynamic changes of electroencephalogram (EEG) signals [6]. These deep learning models demonstrate powerful feature extraction and sequence modeling capabilities in the emotion classification of electroencephalogram (EEG) signals, significantly improving the accuracy and robustness of the classification.

Furthermore, with the deepening of research, emerging deep learning techniques such as attention mechanisms, graph neural networks (GNN), and generative adversarial networks (GAN) have also been explored and applied in the emotion classification of electroencephalogram (EEG) signals [7]. These techniques have further enhanced the performance and efficiency of emotion classification in electroencephalogram (EEG) signals by introducing new network structures and training strategies.

This paper aims to review the latest research achievements of deep learning in the emotion classification of electroencephalogram (EEG) signals, analyze the advantages and disadvantages of existing methods, and look forward to future research directions. Through a comprehensive analysis of the existing literature, this paper aims to provide valuable references and inspirations for researchers in related fields.

2. THE APPLICATION BASIS OF DEEP LEARNING IN THE EMOTION CLASSIFICATION OF ELECTROENCEPHALOGRAM (EEG) SIGNALS

2.1 Characteristics of Electroencephalogram Signals

Electroencephalogram (EEG), as a bioelectrical signal reflecting the electrical activity of brain neurons, has significant characteristics such as non-stationarity, nonlinearity, low signal-to-noise ratio, high dimensionality and time-varying characteristics. These characteristics make the emotion classification task of EEG signals face many challenges. On the one hand, the dynamic changes and complex generation mechanism of EEG signals make it difficult for traditional manual feature extraction methods to fully capture their intrinsic patterns. On the other hand, EEG signals are vulnerable to contamination by environmental noise (such as power frequency interference) and physiological noise (such as electrooculose and electromyography), and individual differences (such as scalp conductivity and electrode placement errors) may lead to uneven signal quality, thereby affecting the generalization ability of the model. Deep learning technology provides new ideas for solving the above problems by constructing multi-level neural network models. Its core advantage lies in its ability to automatically learn the complex feature representations in EEG signals, such as extracting spatial topological features (such as the adjacency relationship of electrode arrays) through convolutional neural networks (CNN), and capturing temporal dependencies through recurrent neural networks (RNN) and their variants (such as Long Short-Term Memory networks (LSTM) and gated recurrent units (GRU). Or dimensionality reduction and feature representation learning can be achieved through autoencoders (AE) and their variants (such as sparse autoencoders SAE and variational autoencoders VAE).

2.2 Overview of Deep Learning Models

The core of deep learning models lies in their hierarchical structure. By simulating the connection mode of human brain neurons, they achieve layer-by-layer abstraction and feature extraction of input data. In EEG emotion classification, CNN extracts local features (such as spectral patterns) through convolutional kernels and reduces the dimension through pooling layers, which is suitable for processing the spatial distribution of multi-channel EEG signals. LSTM and GRU alleviate the vanishing gradient problem through the gating mechanism and are capable of capturing the dynamic changes of EEG signals. SAE and VAE reconstruct signals through unsupervised learning and extract low-dimensional feature representations. For instance, the research by Craik et al. [8] emphasized the significance of deep feature extraction and transfer learning in cross-subject and cross-device scenarios. Although deep learning has shown significant advantages in EEG emotion classification, its application still faces challenges. The high dimensionality and sparsity of EEG signals may lead to model overfitting, individual differences may limit the generalization ability of the model, and the insufficient interpretability of deep learning models may hinder their clinical application. To this end, researchers have proposed a series of improvement strategies, including feature fusion (such as multimodal fusion, spectral-spatial fusion), transfer learning and domain adaptation (such as cross-subject transfer, cross-device transfer), and attention mechanisms (such as channel attention, temporal attention), to enhance the robustness and generalization ability of the model [9].

3. THE APPLICATION OF CLASSIC DEEP LEARNING MODELS IN THE EMOTION CLASSIFICATION OF ELECTROENCEPHALOGRAM (EEG) SIGNALS

The classification of emotions based on electroencephalogram (EEG) signals represents a crucial research area within the realm of sentiment computing. Traditional deep learning models have demonstrated remarkable strengths in this domain due to their exceptional abilities in feature extraction and pattern recognition. Hereafter, a concise overview is provided on the utilization of convolutional neural networks (CNN), recurrent neural networks (RNN) alongside their derivatives, autoencoders (AE) and their respective variants, graph neural networks (GNN), and deep belief networks (DBN) in the context of EEG-based emotion classification.

3.1 Convolutional Neural Network (CNN)

CNN automatically extracts the spatial features of EEG signals through convolutional layers and pooling layers, and is suitable for the joint analysis of multi-channel EEG signals. Its core advantage lies in the ability to capture the spatial dependence between electrodes. For example, using the two-dimensional convolutional neural network (2D-CNN) to process the time-frequency representation of EEG signals (such as the result of short-time Fourier transform), extracting spectral-spatial features, and significantly improving the accuracy of emotion classification. However, CNN is confronted with the overfitting problem caused by high-dimensional EEG signals. Researchers enhance the model's ability to focus on and generalize key features by combining the attention mechanism or transfer learning techniques.

3.2 Recurrent Neural Network (RNN) and Its Variants

RNN and its variants (such as LSTM, GRU) capture the temporal dependence of EEG signals through cyclic units and are suitable for modeling the dynamic changes of EEG signals. In EEG emotion classification, RNN is often combined with CNN to form convolutional recurrent neural networks (CRNN) to achieve joint modeling of spatial-temporal features. For example, the spatial features of EEG signals are extracted first through CNN, then the temporal dependencies are extracted through LSTM, and finally the emotions are classified through the fully connected layer. Although RNN has a powerful time series modeling ability, its computational complexity is relatively high. Researchers optimize the model performance by combining the attention mechanism or self-supervised learning techniques.

3.3 Autoencoder (AE) and Its Variants

AE and its variants (such as SAE and VAE) achieve dimensionality reduction and feature representation learning of EEG signals through unsupervised learning. AE learns the low-dimensional feature representation of EEG signals by minimizing the reconstruction error, SAE enhances the interpretability of features through sparsity constraints, and VAE learns the probability distribution of EEG signals through variational inference. In EEG emotion classification, AE is often used as a preprocessing step and combined with other classification models to alleviate the influence of individual differences on classification performance. The EEG samples generated by VAE can also be used for data augmentation to enhance the generalization ability of the model in small sample scenarios.

3.4 Deep Belief Network (DBN)

DBN achieves feature extraction and classification of EEG signals through layer-by-layer pre-training and fine-tuning. Its pre-training process can initialize the network weights to avoid the deep neural network falling into local optimum, while the fine-tuning process optimizes the classification performance through the backpropagation algorithm. In EEG emotion classification, DBN is often used for feature dimension reduction and classification. For example, the hidden layer output of DBN is taken as a feature and input into other classification models for classification. Although the training process of DBN is complex, researchers have improved the model performance by combining the contrastive divergence algorithm or regularization techniques.

4. EXPLORATION OF EMERGING DEEP LEARNING TECHNOLOGIES IN EMOTION CLASSIFICATION OF ELECTROENCEPHALOGRAM (EEG) SIGNALS

4.1 Attention Mechanism

The attention mechanism empowers the network to concentrate on crucial information pertinent to the task by allocating varying weights to distinct segments of the input data. When it comes to the classification of emotions in electroencephalogram (EEG) signals, the attention mechanism can steer the network towards focusing on the brain region activities linked to emotions, thereby elevating the precision of classification. To illustrate, Chen and his colleagues put forward an emotion classification model for long short-term memory networks that was grounded in the attention mechanism. The incorporation of attention weights markedly augmented the model's capacity to discern features associated with emotions.

4.2 Figure Neural Network (GNN)

GNN propagates information on the graph through the message-passing mechanism and captures the functional connection patterns of EEG signals. The electrode array of EEG signals naturally forms a graph structure, and GNNS (such as GCN, GAT) can directly process the graph structure data to avoid information loss. For example, the coherence matrix of the EEG signal is taken as the adjacency matrix of the graph, and the functional connection features are extracted by using GCN or GAT to achieve emotion classification. Although the computational complexity of GNN is relatively high, researchers optimize the model efficiency by combining multi-scale graph convolution or graph pooling operations.

4.3 Generative Adversarial Network (GAN)

Generative adversarial networks achieve the approximation and generation of data distribution through adversarial training of generators and discriminators. In the emotion classification of electroencephalogram (EEG) signals, GAN can generate synthetic data similar to real EEG signals, which is used to expand the training set and improve the generalization ability of the model. For instance, Wang et al. proposed an emotion classification model based on Conditional Generative Adversarial Network (CGAN). By generating electroencephalogram (EEG) signals under specific emotions, the model's ability to recognize unseen emotional states was significantly enhanced.

5. RESEARCH TRENDS OF INTERDISCIPLINARY INTEGRATION

5.1 The Integration of Brain-Computer Interfaces (BCI) and Deep Learning

Brain-computer interface technology realizes the control of external devices by directly reading the neural activity signals of the brain. Deep learning technology provides powerful signal processing and pattern recognition capabilities for brain-computer interfaces. In the emotion classification of electroencephalogram (EEG) signals, the combination of BCI and deep learning can achieve more natural and efficient human-computer interaction. For instance, researchers are exploring the use of deep learning models to decode electroencephalogram (EEG) signals in real time, in order to achieve emotional perception and control of smart devices.

5.2 Interdisciplinary Research on Neuroscience and Deep Learning

Neuroscience provides biological inspiration and theoretical basis for deep learning. In the emotion classification of electroencephalogram (EEG) signals, the cross-disciplinary research of neuroscience and deep learning can reveal the neural mechanism of emotion generation and provide guidance for the design and optimization of deep learning models. For instance, researchers are exploring the use of emotion theories in neuroscience, such as fundamental emotion theory and dimensional emotion theory, to guide the feature extraction and classification strategies of deep learning models.

5.3 Multimodal Data Fusion

Electroencephalogram (EEG) signals, as data of a single modality, may have the problem of insufficient information in emotion classification. Multimodal data fusion technology can provide more comprehensive and accurate emotional information by integrating data from different modalities (such as electroencephalogram signals, eye movement signals, facial expressions, etc.). Deep learning technology provides an effective solution for multimodal data fusion. For instance, researchers are exploring the use of deep learning models to jointly model multimodal data in order to achieve more accurate emotion classification.

6. CHALLENGES AND FUTURE RESEARCH DIRECTIONS

Although deep learning has made significant progress in the emotion classification of electroencephalogram (EEG) signals, it still faces many challenges. Firstly, electroencephalogram (EEG) signals have the characteristics of low signal-to-noise ratio and high dimensionality, which makes feature extraction and model training difficult. Secondly, the subjectivity and complexity of emotions lead to subjectivity and inconsistency in the annotation of emotion labels, which affects the generalization ability of the model. Furthermore, the complexity and computational cost of deep learning models also limit their application in real-time emotion classification.

In future research, it is necessary to explore more effective feature extraction methods, such as those based on time-frequency analysis, wavelet transform, empirical mode decomposition and other techniques, in order to improve the signal-to-noise ratio and feature expression ability of electroencephalogram (EEG) signals. Optimize the model structure and design more efficient and robust deep learning model structures, such as lightweight networks, attention mechanisms, graph neural networks, etc., to reduce the complexity and computational cost of the model, and improve the real-time performance and generalization ability of the model. Strengthen multimodal data fusion and explore more effective multimodal data fusion methods, such as joint modeling based on deep learning and cross-modal attention mechanisms, to provide more comprehensive and accurate emotional information. Introduce prior knowledge and incorporate prior knowledge such as the emotion theory in neuroscience and the emotion regulation mechanism in cognitive psychology into deep learning models to guide the design and optimization of the models and enhance their interpretability and reliability. Carry out the construction of large-scale labeled datasets, establish large-scale and high-quality datasets for the emotion classification of electroencephalogram (EEG) signals, provide sufficient training data for deep learning models, and improve the generalization ability and classification accuracy of the models.

7. CONCLUSION

This paper reviews the latest research achievements of deep learning in the emotion classification of electroencephalogram (EEG) signals, analyzes the advantages and disadvantages of existing methods, and looks forward to the future research directions. Deep learning technology, by constructing multi-level neural network models, can automatically learn the complex feature representations in electroencephalogram (EEG) signals, providing effective technical support for emotion classification. However, deep learning still faces many challenges in the emotion classification of electroencephalogram (EEG) signals, such as difficulties in feature extraction, insufficient generalization ability of the model, and high computational costs. Future research can be carried out from aspects such as improving feature extraction methods, optimizing model structures, strengthening multimodal data fusion, introducing prior knowledge, and conducting the construction of large-scale labeled datasets, in order to promote the further development of deep learning in the emotion classification of electroencephalogram (EEG) signals.

REFERENCES

[1] Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2018). EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces. Journal of Neural Engineering, 15(5), 056013.
[2] Li, X., Song, D., Zhang, P., & Yao, L. (2018). Exploring EEG features in cross-subject emotion recognition. Frontiers in Neuroscience, 12, 162.
[3] Lotte, F., Bougrain, L., & Cichocki, A. (2018). A review of classification algorithms for EEG-based brain-computer interfaces: A 10 year update. Journal of Neural Engineering, 15(3), 031005.
[4] Song, T., Wen, W., Liao, W., & Zheng, W. L. (2018). EEG emotion recognition using dynamical graph convolutional neural networks. IEEE Transactions on Affective Computing, 11(3), 532-541.
[5] Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., ... & Ball, T. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. Human Brain Mapping, 38(11), 5391-5420.
[6] 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.
[7] Chen, Z., Liu, Y., & Li, H. (2020). EEG-based emotion recognition using regional-asymmetric spatiotemporal pattern recurrent neural networks. IEEE Transactions on Affective Computing, 11(3), 524-536.
[8] Craik, A., He, Y., & Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: A review. Journal of Neural Engineering, 16(2), 026030.
[9] Wang, Y., Zheng, W. L., Yu, H. J., & Lu, B. L. (2019). Emotion recognition using multimodal residual LSTM networks. IEEE Transactions on Affective Computing, 12(2), 271-283.