EEG-based Human Decision Confidence Measurement Using Graph Neural Networks

Le-Dian Liu[1], Rui Li[1], Yu-Zhong Liu[6], Hua-Liang Li[6], and Bao-Liang Lu[1,2,3,4,5(✉)]

[1] Center for Brain-like Computing and Machine Intelligence, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
{liuledian, realee, bllu}@sjtu.edu.cn
[2] Center for Brain-Machine Interface and Neuromodulation, RuiJin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200020, China
[3] RuiJin-Mihoyo Laboratory, RuiJin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200020, China
[4] Key Laboratory of Shanghai Commission for Intelligent Interaction and Cognitive Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[5] Qing Yuan Research Institute, Shanghai Jiao Tong University, Shanghai 200240, China
[6] Key Laboratory of Occupational Health and Safety of Guangdong, Power Grid Co., Ltd, Electric Power Research Institute of Guangdong Power Grid Co., Ltd.

Abstract. Most of the studies on decision confidence are from the fields of neuroscience and cognitive science, and existing studies based on deep neural networks do not exploit the topology of multi-channel EEG signals. In this paper, we propose an attentive simple graph convolutional network (ASGC) for EEG-based human decision confidence measurement. ASGC captures both coarse-grained and fine-grained inter-channel relationship by learning a shared adjacency matrix and utilizing self-attention mechanism, respectively. In addition, we propose a confidence distribution learning (CDL) loss based on a natural intuition to alleviate two problems: lack of training samples and label ambiguity. We conduct experiments on a dataset built for the confidence measurement in a visual perception task. The experimental results demonstrate advanced performance of our model, achieving an accuracy of 68.83% and F1-score of 66.9%. Finally, we investigate the critical channels for decision confidence measurement with the attention matrix of EEG channels.

Keywords: EEG · Graph Neural Network · Decision Confidence.

1 Introduction

Nowadays, with the development of deep learning algorithms, human participation is no longer needed in some well-structured problems. Nevertheless, professionals are still indispensable in complex tasks with high risk, such as business decision making and military remote sensing images interpreting. Unfortunately, it is impossible to always make decisions with honesty and certainty. Hence, developing an objective and stable method to measure human decision confidence is of great practical value.

Human decision confidence is defined as the probability of an overt or covert decision, given the evidence, being correct [11]. Most of the studies on decision confidence are from the fields of neuroscience and cognitive science. They can be broadly divided into two categories depending on the techniques used to acquire data from brains: functional magnetic resonance imaging (fMRI) and electroencephalography (EEG). The studies based on fMRI [1, 5, 10] indicate that anterior cingulate cortex, prefrontal cortex, posterior parietal cortex, superior parietal lobule, and ventral striatum may be the brain regions closely related to human decision confidence. On the other hand, there are some works that use EEG and event-related potentials (ERPs) to investigate neural patterns of human decision confidence [4, 2]. These works have shown that EEG signals recorded in decision-making process can be used to discriminate different degrees of human decision confidence. However, ERP experiments require rapid presentation of the stimulus in a laboratory environment, so they are not feasible to real-world one-trial applications.

In the field of machine learning, there have been some attempts to handle the strict requirements of ERP experiments and improve discrimination performance [8, 7]. They leveraged the power of deep learning models to measure human decision confidence from multi-channel EEG recorded in decision-making process. Their experimental results show that EEG signals are capable of measuring different levels of decision confidence and the differential entropy (DE) feature using all 5 bands achieves the best performance. However, their approaches ignore the topological structure of multi-channel EEG. Moreover, their approaches suffer from the limited number of training samples and the label ambiguity of 5-level confidence categorization caused by the huge costs of collecting EEG data and the difficulty of acquiring accurate labels.

Graph neural networks (GNNs) are deep learning approaches applied in the graph domain. In GNNs, convolution operations are the most popular propagation operations, which aims to generalize the classical signal processing operation to the graph domain. Kipf and Welling [6] proposed graph convolutional network (GCN), which simplifies the convolution operation and handles gradient exploding/vanishing problem by introducing a renormalization trick. To reduce the excess complexity of GCNs, SGC [13] removes nonlinearities and merges weight matrices between consecutive layers. All of the models mentioned before use the fixed original adjacency matrices to represent relations between nodes. To capture implicit relations between nodes, Li et al. [9] proposed an Adaptive Graph Convolution Network(AGCN). The residual graph Laplacian is learned by AGCN and added to the original Laplacian matrix. In addition, many advanced works [14, 15] in affective computing have successfully adopted GNN to exploit the topology of EEG signals.

In this paper, inspired by biological topology of human brains, we propose an attentive simple graph convolutional network (ASGC) to capture both the coarse-grained and fine-grained topological structure of multi-channel EEG. Specifically, inspired by [12], we use a learnable adjacency matrix to capture coarse-grained inter-channel relations shared among all samples. Furthermore, we use an attention score matrix to capture fine-grained inter-channel relations for each sample in runtime. In addition, we propose a confidence distribution learning (CDL) loss to solve two problems: inadequate training samples and ambiguous labels. Inspired by [3], our CDL loss enforce our model to learn a discrete class distribution rather than a single class for each sample, leading to a better performance of the trained model. We conduct experiments on the dataset built for the confidence measurement in a visual perception task developed in [8]. Our experimental results demonstrate the superior performance of our proposed ASGC compared to other baseline models. Finally, critical channels are explored for decision confidence measurement in the visual perception task.

2 Methodology

In this section, we formulate the confidence classification problem and our attentive simple graph convolutional network (ASGC). Then, we detail the confidence distribution loss (CDL) designed for this specific problem.

2.1 Attentive Simple Graph Convolutional Network

We consider each EEG channel as a graph node, so the input can be represented by a feature matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$, where n denotes the number of channels and d denotes the feature dimension of each channel. For each training sample, a label $\mathbf{Y} \in \{1, 2, ..., C\}$ is given, where C denotes the number of categories.

The overall architecture of ASGC is illustrated in Fig. 1. The SGC is used to exploit the coarse-grained inter-channel relationship. Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be the learnable adjacency matrix in SGC. We define $\mathbf{S}$ as follows

$$\mathbf{S} = \tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}}, \tag{1}$$

where $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_n$, and $\tilde{\mathbf{D}}_{ii} = \sum_j \tilde{\mathbf{A}}_{ij}$. This is a renormalization trick introduced by [6] to solve the exploding and vanishing gradient problem. Then, the simple graph convolution network (SGC) can be formulated as follows

$$\mathbf{Z} = \mathbf{S}^K \mathbf{X} \mathbf{W} = \tilde{\mathbf{X}} \mathbf{W}, \tag{2}$$

where $\mathbf{W} \in \mathbb{R}^d \times h$, and h denotes the hidden size. In the graph convolution operation defined in (2), K implies that each node can aggregate information from the nodes that are $K-$hops away. Although $\tilde{\mathbf{X}}$ is still in the same feature space as $\mathbf{X}$, it incorporates the topological information of EEG channels. The final output of SGC $\mathbf{Z}$, linearly transformed from $\tilde{\mathbf{X}}$, is a feature matrix in a high-dimensional feature space. L1 regularization is applied on $\mathbf{A}$ to improve the sparsity of the adjacency matrix.

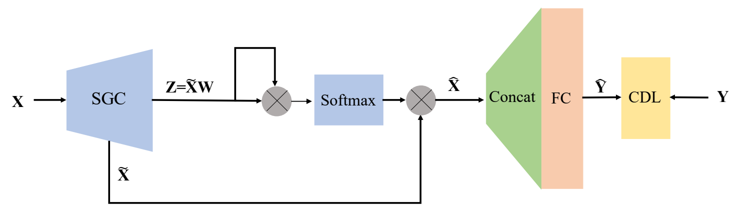

Fig. 1. The overall architecture of our ASGC model.

To further capture the fine-grained topological structure of EEG channels in runtime, we introduce a self-attention module as follows

$$\mathbf{M} = (m_{ij}) = softmax(\mathbf{Z}\mathbf{Z}^T), \tag{3}$$

$$\hat{\mathbf{X}} = \mathbf{M}\tilde{\mathbf{X}}, \tag{4}$$

where $\mathbf{M} \in \mathbb{R}^{n \times n}$ is the attention matrix, and softmax is employed in self-attention to normalize the attention matrix so that $\sum_j m_{ij} = 1$. Essentially, self-attention mechanism can be viewed as the refinement of the input feature using a linear combination of self-values. It worth noting that attention matrix is dynamically calculated for each sample, compared to the fixed adjacency matrix in the SGC.

Finally, rather than use global pooling, we concatenate the feature of all nodes into a vector. Then, the vector is fed into a fully-connected layer with softmax activation function. The output distribution over all classes is computed as follows

$$\hat{\mathbf{Y}} = softmax(\hat{\mathbf{X}}\mathbf{W}^o), \tag{5}$$

where $\mathbf{W}^o \in \mathbb{R}^{h \times C}$, and $\hat{\mathbf{Y}} \in \mathbb{R}^C$.

2.2 Confidence Distribution Loss

The works in [8, 7] simply use the one-hot encoding of 5-level confidence labels to calculate the training loss. However, the distance between each pair of classes should not be considered as equal due to the intuition that the confidence level is a continuous state. For example, the distance between level 1 and level 5 should be greater than the distance between level 1 and level 2. Inspired by [3], we propose a confidence distribution loss to address this problem. Specifically, we convert each label $\mathbf{Y} \in \{1, 2, .., C\}$ to a distribution $\tilde{\mathbf{Y}} \in \mathbb{R}^C$. We assume that the distribution should concentrate around the ground-truth label $\mathbf{Y}$ and the distance between the adjacent confidence levels is equal. Thus, it is nature to use the normal distribution with $\mu = \mathbf{Y}$ and σ to construct the target distribution:

$$\tilde{\mathbf{Y}}_c = \frac{p(c|\mu, \sigma)}{\sum_{l=1}^{C} p(l|\mu, \sigma))}, \tag{6}$$

$$p(l|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(l-\mu)^2}{2\sigma^2}\right), \tag{7}$$

where $c = 1, 2, ..., C$, σ is a hyper-parameter that can be tuned, and the distribution requirement $\sum_{l=1}^{C} \tilde{\mathbf{Y}}_l = 1$ is satisfied.

Then, the confidence distribution loss can be calculated as the Kullback-Leibler (KL) divergence between the predicted distribution $\hat{\mathbf{Y}}$ and target distribution $\tilde{\mathbf{Y}}$:

$$CDL = -\tilde{\mathbf{Y}}_c \sum_{l=1}^{C} \hat{\mathbf{Y}}_c \tag{8}$$

3 Experiments

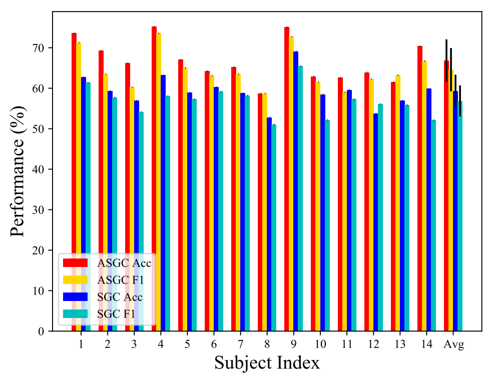

Fig. 2. The accuracy and F1-score (%) of ASGC and SGC for 14 subjects and their average.

3.1 Dataset

We conduct experiments on the dataset developed in [8]. When participants were performing visual perceptual decision-making task, the EEG data is recorded in 62 channels using an active AgCl electrode cap at a sampling rate of 1000 Hz. The dataset comprises 14 subjects and each subject has 135 trials, where each trial corresponds to one decision process. For data preprocessing, a bandpass filter between 0.3 and 50 Hz is applied to each channel to filter the noise and linear dynamic system (LDS) method is adopted to smooth feature. We use the differential entropy (DE) feature on all five bands, since it achieves the best performance in [8, 7]. We follow the subject-dependent classification setting in their works and train a model for each subject.

Table 1. The classification accuracy and F1-score (%) (mean/std) of SVM, DNNS, SGC, SGC with CDL and our ASGC model.

	Model				
Metric	SVM [8]	DNNS [8]	SGC	SGC+CDL	ASGC
Accuracy	40.93/ -	49.14/-	59.32/4.03	63.41/5.46	**68.83**/5.25
F1-score	37.43/ -	45.07/-	56.87/3.84	59.85/5.24	**66.90**/5.33

3.2 Performance Evaluations

The mean accuracy and F1-scores of SVM, DNNS, SGC, SGC with CDL and our final model ASGC are presented in Table 3.2. The performance of SVM and DNNS are quoted from [8]. We can find that SGC performs better than DNNS and SVM, which justifies the idea of capturing inter-channel topological relationship. Moreover, SGC with CDL performs better than SGC only, indicating the effectiveness of our proposed CDL. Finally, our ASGC model achieves the best accuracy and F1-score compared to all other baselines, showing the superior performance in this task.

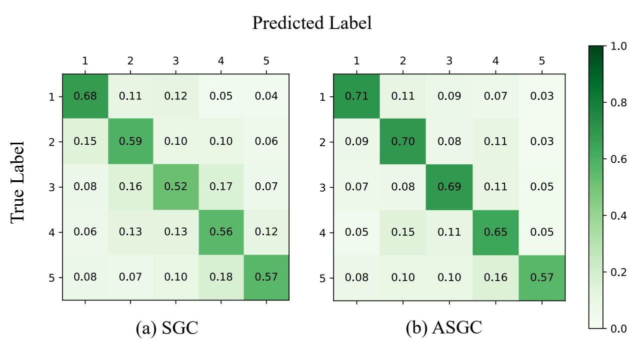

Fig. 3. The confusion matrices of SGC and ASGC. The rows of the confusion matrix represent the target class and the columns represent the predicted class.

The classification results of each subject (14 in total) and their average are shown in Fig. 2. We can find that our ASGC model achieves better accuracy and F1-score than SGC for all subjects. It shows that our ASGC model has consistent performance across different subjects, indicating the robustness of ASGC. The confusion matrices of SGC and ASGC averaged on all 14 subjects are shown in Fig. 3. We can find that both SGC and ASGC perform relatively better on the lowest confidence level, and that the intermediate levels are more difficult to classify. It may indicate that participants in the lowest confidence level may have similar EEG patterns. In addition, our model gets relatively

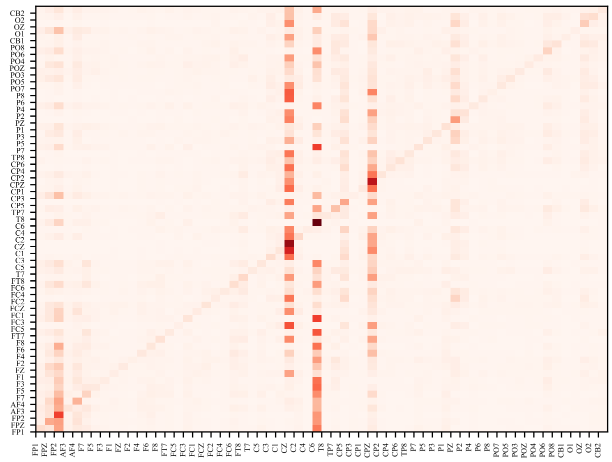

Fig. 4. The heatmap visualization of the attention matrix averaged on all subjects in the dataset.

worse at recognizing the highest confidence level, but still performs better than SGC does. Moreover, ASGC is better than SGC in discriminating intermediate confidence levels, showing a more fine-grained discrimination ability.

3.3 Analysis of Critical Channels

We further investigate the critical channels associated with confidence measurement in the visual perception decision-making task. Fig. 4 visualizes the attention matrix averaged on all subjects in the dataset. Note that the sum of each row is equal 1 due to the softmax operation along each row. The diagonal values represent the attention paid to itself by each channel. We can find that the attention matrix is sparse and some of the channels are of greater importance than others. It is clear form Fig. 4 that FP2, CZ, C6, and CPZ may be the important channels to discriminate the decision confidence in the visual perception task.

4 Conclusion

In this paper, we have proposed an attentive simple graph convolutional network (ASGC) for capturing fine-grained topology structure of EEG channels compared to SGC. In addition, we used a concatenate operation instead of global pooling to preserve the structure of channels. Moreover, we have proposed a confidence distribution loss based on the intuition that samples with closer confidence levels are more similar, alleviating the problems of lacking training samples and label ambiguity. The experimental results demonstrate the superior performance of our ASGC model compared to other baseline models, and the effectiveness of the CDL designed for the confidence level classification problem. Finally, the analysis on the attention matrix suggests that FP2, CZ, C6, and CPZ may be the important channels for measuring decision confidence in the visual perception task.

References

1. Bang, D., Fleming, S.M.: Distinct encoding of decision confidence in human medial prefrontal cortex. Proceedings of the National Academy of Sciences **115**(23), 6082–6087 (2018)
2. Boldt, A., Schiffer, A.M., Waszak, F., Yeung, N.: Confidence predictions affect performance confidence and neural preparation in perceptual decision making. Scientific reports **9**(1), 1–17 (2019)
3. Gao, B.B., Xing, C., Xie, C.W., Wu, J., Geng, X.: Deep label distribution learning with label ambiguity. IEEE Transactions on Image Processing **26**(6), 2825–2838 (2017)
4. Gherman, S., Philiastides, M.G.: Neural representations of confidence emerge from the process of decision formation during perceptual choices. Neuroimage **106**, 134–143 (2015)
5. Hebart, M.N., Schriever, Y., Donner, T.H., Haynes, J.D.: The relationship between perceptual decision variables and confidence in the human brain. Cerebral cortex **26**(1), 118–130 (2016)
6. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
7. Li, R., Liu, L.D., Lu, B.L.: Measuring human decision confidence from EEG signals in an object detection task. In: 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER). pp. 942–945. IEEE (2021)
8. Li, R., Liu, L.D., Lu, B.L.: Discrimination of decision confidence levels from EEG signals. In: 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER). pp. 946–949. IEEE (2021)
9. Li, R., Wang, S., Zhu, F., Huang, J.: Adaptive graph convolutional neural networks. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 32 (2018)
10. Molenberghs, P., Trautwein, F.M., Böckler, A., Singer, T., Kanske, P.: Neural correlates of metacognitive ability and of feeling confident: a large-scale fmri study. Social cognitive and affective neuroscience **11**(12), 1942–1951 (2016)
11. Pouget, A., Drugowitsch, J., Kepecs, A.: Confidence and certainty: distinct probabilistic quantities for different goals. Nature neuroscience **19**(3), 366 (2016)
12. Song, T., Zheng, W., Song, P., Cui, Z.: EEG emotion recognition using dynamical graph convolutional neural networks. IEEE Transactions on Affective Computing **11**(3), 532–541 (2018)
13. Wu, F., Souza, A., Zhang, T., Fifty, C., Yu, T., Weinberger, K.: Simplifying graph convolutional networks. In: International conference on machine learning. pp. 6861–6871. PMLR (2019)
14. Zhang, G., Yu, M., Liu, Y.J., Zhao, G., Zhang, D., Zheng, W.: Sparsedgcnn: Recognizing emotion from multichannel EEG signals. IEEE Transactions on Affective Computing (2021)
15. Zhong, P., Wang, D., Miao, C.: EEG-based emotion recognition using regularized graph neural networks. IEEE Transactions on Affective Computing (2020)