NODE-WISE DOMAIN ADAPTATION BASED ON TRANSFERABLE ATTENTION FOR RECOGNIZING ROAD RAGE VIA EEG

Xueqi Gao\textsuperscript{1}, Chao Xu\textsuperscript{1,2,*}, Yihang Song\textsuperscript{1}, Jing Hu\textsuperscript{1,2}, Jian Xiao\textsuperscript{1}, Zhaopeng Meng\textsuperscript{1,3}

\textsuperscript{1}College of Intelligence and Computing, Tianjin University, Tianjin, China
\textsuperscript{2}Higher Research Institute, University of Electronic Science and Technology of China, Shenzhen, China
\textsuperscript{3}Tianjin University of Traditional Chinese Medicine, Tianjin, China
\{gxq,000110,xuchao\}@tju.edu.cn

ABSTRACT
Road rage is a social problem that deserves attention, but few research has been done so far. In this paper, based on the biological topology of multi-channel electroencephalogram (EEG) signals, we propose a model which combines transferable attention (TA) and regularized graph neural network (RGNN). First, topology-aware information aggregation is performed on EEG signals, and complex relationships between channels are dynamically learned. Then, the transferability of each channel is quantified based on the results of the node-wise domain classifier, which is embedded into the emotion classifier as attention score. Importantly, we recruited 10 subjects and collected their EEG signals in pleasure and rage states in simulated driving conditions. We verify the effectiveness of our method on this dataset and compare it with other methods. The results indicate that our method is simple and efficient, with 85.63% accuracy in cross-subject experiments. It can be used to identify road rage.

Index Terms— Road Rage, Emotion Recognition, Electroencephalogram, Transferable Attention, Regularized Graph Neural Network

1. INTRODUCTION
Rage is very common while driving, which causes drivers to change lanes, forcibly overtake or even attack other vehicles. It is considered to be one of the main reasons for traffic accidents. Usually, drivers are able to detect their emotional changes, but when a negative emotion is strong, it can be difficult to shake off its effects. Therefore, it is necessary to detect the driver’s emotional state, which can provide guidance for subsequent emotional interventions, such as music soothing or safety warning [1].

As the research object of emotion recognition, EEG signal has the advantages of high temporal resolution and direct reflection of brain state, therefore received extensive attention [2]. With the rapid development of wearable devices, wireless EEG headsets provide signal quality comparable to gel electrodes, and some automobile manufacturers have developed biometric seat prototypes using physiological signals to monitor driver’s condition [3]. With these foundations, we focus on the identification of road rage based on EEG signals.

The biological topology of multi-channel EEG signal is very critical for emotion recognition, and the graph neural network (GNN) [4] can be used to mine the rich information. For example, [5] applied graph convolutional neural network (GCNN) to extract graph domain features and long short-term memory neural network (LSTM) to extract temporal correlations. [6] proposed a dynamic graph convolutional network (DGCNN), which first initialized the adjacency matrix with a distance function, and then dynamically learned it in the network. However, the spatial location of channels can not represent their functional dependence or degree of correlation [7]. In this study, we compute mutual information (MI) to characterize functional connectivity between channels and use it as the initial value of adjacency matrix.

On the other hand, the distribution of EEG signals of different subjects varies greatly which reduces the generalization ability of the model. Many methods have been proposed. [8] applied the Domain Adversarial Network (DANN). [9] proposed two regularizers NodeDAT and EmotionDL to better handle this problem. However, existing methods mainly align representations extracted from the whole EEG signal across domains, failing to take into account that channels’ inter-domain differences are not same. Inspired by the idea which proposed on the image classification problem [10], we use an entropy function to quantify the transferability of channels based on the output of the node-wise domain classifier, and embed it into the emotion classification task.

Overall, our contributions include: (1) The EEG signals of drivers in pleasure and rage states are collected, which constitute a valuable dataset. (2) To characterize functional connectivity between channels, the adjacency matrix initialized with MI and dynamically learned in the network. (3) Combining transferable attention with node-wise domain adversarial network, which makes the model focus on domain-invariant representations and improves accuracy.

This work is supported by the National key RD program of China under Grant No. 2021YFF0901203, No. 2018YFB1701701.
2. METHODOLOGY

In this study, our framework including graph modeling, information aggregation and domain adaptation, as shown in Fig 1.

2.1. Graph Modeling

Different from images and texts, multi-channel EEG signals belong to non-European domain data, so graph structure modeling is chosen. In this study, the graph is denoted as \( G = (V, E) \), \( V \) is the set of vertices, represented by the feature matrix \( X \in R^{n \times d} \), \( n \) denotes the number of channels, \( d \) denotes the number of features per channel. \( E \) is the set of edges, represented by an adjacency matrix \( A \in R^{n \times n} \).

We divide the signal of each channel into five frequency bands (\( \delta:0.5\text{~}4 \)hz, \( \theta:4\text{~}8 \)hz, \( \alpha:8\text{~}13 \)hz, \( \beta:13\text{~}32 \)hz, \( \gamma:32\text{~}50 \)hz). Then, the differential entropy (DE) [11] is extracted to form the feature matrix \( X \), and the calculation formula as follows:

\[
DE = -\int_0^\infty \frac{1}{2\pi\sigma_x^2} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} \log \left( \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} \right) dx
\]

\[
= \frac{1}{2} \log (2\pi e\sigma_x^2)
\]

where \( x \) denotes an EEG signal of a certain length which approximately obeys Gaussian distribution \( N (\mu, \sigma_x^2) \), \( e \) is Euler’s constant.

Emotional responses in the brain involve the cooperation of multiple brain regions. We compute the mutual information (MI) between channels as functional connection strength [12]. We average over all samples and normalize them for the initial value of the adjacency matrix \( A \), as follows:

\[
MI (X, Y) = \sum_{x,y} P_{XY} (x,y) \log_2 \frac{P_{XY} (x,y)}{P_X (x) P_Y (y)}
\]

\[
NormalizedMI (X, Y) = \frac{2MI (X,Y)}{H (X) + H (Y)}
\]

where \( P_{XY} (x,y) \) denotes the joint probability distribution of the signals \( x, y \). \( P (*) \) and \( H (*) \) are the probability distributions and entropy of the signal, respectively.

Negative emotions can activate the right frontal, temporal, and parietal lobes, while positive emotions can activate the left region, forming the spatial characteristics of EEG [13–15]. To exploit this information asymmetry, we add several global connections to the adjacency matrix \( A \) based on the method proposed in [9], setting their initial values as \( A_{ij} = A_{ij} - 1 \in [-1, 0] \), where (\( i, j \)) denotes a globally connected pair, including (AF3,AF4), (FC5,FC6), (P7,P8) and (O1,O2).

2.2. Topology-Aware Information Aggregation

We select Simple Graph Convolution (SGC) [16] to perform topology-aware information aggregation for node features. For a given feature matrix \( X \in R^{n \times d} \), information aggregation can be expressed as:

\[
Z = S^{L}XW
\]

where \( S = D^{-\frac{1}{2}}AD^{-\frac{1}{2}} \in R^{n \times n} \), \( D \) is the degree matrix of \( A \), i.e.\( D_{ii} = \sum_j A_{ij} \). \( W \in R^{d \times d} \) denotes the weight matrix, and \( L \) denotes the number of layers. \( Z \in R^{n \times d} \) is the output, \( d \) denotes the output feature dimension.

2.3. Focus on Domain-Invariant Representations

Pass \( Z \) into a gradient reversion layer (GRL) [17]. Then, according to each node representation, the domain classifier \( G_d \) gives the output \( d_i^k = G_d^k (reversalZ_i^k) \), representing the probability that the \( k \)-th node of the \( i \)-th sample belongs to the source domain. If the domain classifier still cannot distinguish its domain after a certain training, that is, the value of \( d_i^k \) is around 0.5, which means it is a domain-invariant representation. In other words, the \( k \)-th node can be transferred across domains, which should produce a larger attention value.

The entropy function is defined as \( H (p) = -\sum_j p_j \cdot \log (p_j) \), which can be used to quantify the transferability of nodes in our study. Specifically, the more transferable the node, the larger \( H (d_i^k) \), and vice versa. According to [18], the effect of false attention can be mitigated by adding residual connections. Therefore, before emotion recognition, the node representation is transformed into:

\[
f_i^k = \left( 1 + H (d_i^k) \right) Z_i^k
\]
In this way, larger weight is given to nodes with small differences between domains, so that they are paid more attention in emotion classification, and the accuracy of the model in cross-subject experiments is improved.

2.4. Objective Function

Our model aims to minimize the following objective function:

\[
obj = L_{\text{cls}} + l_1 - \text{norm} - \lambda L_{\text{domain}}
\]

\[
= \frac{1}{n_s} \sum_{x_i \in D_s} L_{\text{cls}} (G_y (f_i), y_i) + \alpha \|A\|_1
\]

\[
- \frac{\lambda}{n_s + n_t} \sum_{k=1}^{n} \sum_{x_i \in D_s \cup D_t} L_{\text{domain}} (G_d (\text{reverse} \, Z^s_i), d^t_i)
\]

where \(D_s = \{(x^s_i, y^s_i)\}_{i=1}^{n_s}, D_t = \{(x^t_i)\}_{i=1}^{n_t}\) denote the source domain and the target domain, respectively. \(x_i\) is a sample, \(y_i\) is the corresponding label. \(G_y\) denotes emotion classifier and \(G_d\) denotes domain classifier. \(\|\|_1\) is the \(l_1\)-norm, \(\alpha\) and \(\lambda\) are hyper-parameters.

3. EXPERIMENTS

To investigate effective representations of pleasure and rage states, we carefully design a driving experiment.

**Subjects:** We recruit 10 right-handed college students (6 males, 4 females) aged 21 to 27 (mean: 23.90, standard deviation: 1.72) with practical driving experience for the experiment. Subjects did not have any mental illness, nor take drugs, alcohol or caffeine, and ensured reasonable rest and a stable state of mind before the experiment.

**Experiment Protocol:** For safety reasons, we choose to use the City Car Driving software to simulate in the laboratory environment, and set the steering wheel, joystick, pedals and other equipment to restore the driving experience as much as possible. Previous studies have shown that virtual reality-based scenarios are effective in inducing driving emotions [19], that is, our simulation experiments are simple, safe, and able to achieve the desired effect.

We set up pleasure and rage scenarios in the virtual driving software. Among them, the pleasure scenario simulates a wide and smooth suburban road with only a small number of cars, and no emergency situation occurs. The rage scenario simulates the streets of a commercial area, where traffic is congested, frequently waits for traffic lights, and is also affected by the incorrect behavior of surrounding vehicles and pedestrians interference, as shown in Fig2. In addition, all scenarios are set on a summer sunny day, in a Porsche Panamera Turbo and the driving route is free.

The experiment consists of four steps, as shown in Fig3. Since the subjects had no experience with driving simulation equipment, they were asked to practice driving before experiment. Then, the EEG signals were collected while driving under the specified scenario. The experiment was set up to alternately between pleasure and rage scenarios, a total of 4 groups. After driving, the subjects filled out a subjective assessment questionnaire which includes ratings based on how they actually feel. The experiment was performed in an isolated and quiet room.

**Annotations of Emotions:** We use a combination of self-report and experimental scenario to give emotional labels, which can improve confidence in the labels. Specifically, subjects fill out subjective assessment questionnaires that include ratings of seven discrete emotions, based on how they actually feel. The seven discrete emotions are pleasure, excitement, surprise, rage, fear, nervous and frustration. Take the rage scenario, where rage is the target emotion and the other six are non-target emotions. If the following two points are satisfied, we believe that the target emotion is successfully and discretely induced, and can be used as the emotional label.

1. The highest rated emotion, i.e. the emotion actually induced, is the same as the target emotion.
2. The dispersion degree of target emotion induced by scenario is not less than 0.83. The dispersion degree is calculated by \(d_s = \frac{\sum_{s \in S} d_s}{m}\), \(n\) denotes the number of non-target emotions, \(m\) denotes the number of instances which the target emotion was at least one point greater than the nontarget emotions.

**Data Acquisition and Preprocessing:** In order to better fit the proposed application scenario, the Emotiv EPOC+ portable EEG acquisition device is selected for the experiment, and its performance is comparable to professional equipment [20]. Before acquisition, hydrated the electrode sensors with saline to increase conductivity. Subjects were advised to limit unnecessary physical movement as much as possible during the process of data acquisition. The experimental setup is shown in Fig4.
After data acquisition, the sampling rate was reduced to 128 Hz, and EEG signals were band-pass filtered at 0.5–50 Hz. Then, we visually inspected EEG signals to remove the heavily contaminated fragments, and divided the remaining EEG signals into the same length 1s with 0.5s overlap. We implemented the FORCe method [21] which can remove the EOG and EMG artifacts. The main steps as follow: (1) Perform wavelet decomposition on each channel of EEG signals. (2) Perform ICA on the approximate coefficients, and exclude ICs containing artifacts according to the differences in AMI, peak value, kurtosis, power spectral density and other characteristics between artifacts and clean EEG signals. (3) Calculate the peak coefficients and perform soft threshold processing. (4) Reconstruct to obtain clean EEG signals.

4. RESULTS

In this section, we introduce the results of TA-RGNN model on the experimental dataset.

Table 1. Comparison with existing work on the experimental dataset.

| model    | acc    | std    | recall | specificity | F1    |
|----------|--------|--------|--------|-------------|-------|
| DAN [22] | 69.34  | 11.73  | 0.693  | 0.694       | 0.554 |
| DANN [8] | 76.68  | 09.54  | 0.754  | 0.771       | 0.639 |
| RGNN [9] | 82.12  | 08.21  | 0.792  | 0.832       | 0.708 |
| TA-RGNN  | 85.63  | 08.79  | 0.808  | 0.875       | 0.755 |

Performance comparison: We evaluate model performance and compare with previous methods. Table1 summarizes our experimental results. [22] minimizes the maximum mean discrepancy (MMD) between deep feature representations. [8] aligns representations extracted from the entire EEG signals across domains. [9] implements a more fine-grained cross-domain transfer from the node wise. On this basis, we further quantify nodes transferability and embed it into the following emotion recognition task as attention score. From Table1, we can see that TA-RGNN model achieves a higher accuracy and tends to converge in fewer epochs, which has practical application prospects.

In addition, to verify the effectiveness of each module, according to [9,23], we initialize the adjacency matrix $A$ with the inter-electrode distance function. The emotion recognition accuracies are 83.37% and 84.05%, respectively. It shows that multiple brain regions are related to each other when produces emotions, but the spatial location of electrodes cannot reflect this connection well.

Sensitivity Analysis: We set multiple values for hyperparameters, as shown in Fig5. When $\lambda \in [0, 1]$, the accuracy gradually increases, indicating that the confrontation between the feature extractor and the domain discriminator enables the model to learn emotion-related, domain-irrelevant features. As $\lambda$ continues to increase, it will degrade performance because the learned features lose emotional discrimination. However, model performance does not change much as the hyper-parameter $\alpha$ changes. We set $\alpha = 0.01$ to avoid strong regularization.

We plot the confusion matrix as shown in Fig5. Compared with pleasure, the model correctly recognizes a higher proportion of rage, indicating that the EEG signals generated by subjects in rage state are more similar.

5. CONCLUSION

In this study, we propose a model which combines transferable attention and node-wise domain adversarial networks, and demonstrate its effectiveness in cross-subject emotion recognition task. To the best of our knowledge, this is the first study to identify road rage with EEG signals using deep learning. Our method shows that computing MI to initialize the adjacency matrix can better represent the intrinsic relationships between EEG channels. A domain adaptive model based on transferable attention can focus on domain-invariant representations and improve the accuracy of cross-subject experiments with fewer epochs. In future work, we will explore how to reduce the interference of head shaking on the EEG signal, which will help the model generalize to practical.
6. REFERENCES

[1] C. Bankar, A. Bhide, A. Kulkarni, C. Ghube, and M. Be- dekar, “Driving control using emotion analysis via eeg,” in 2018 ieee punecon. IEEE, 2018, pp. 1–7.

[2] P. J. Bota, C. Wang, A. L. Fred, and H. P. Da Silva, “A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals,” IEEE Access, vol. 7, pp. 140990–141020, 2019.

[3] J. Izquierdo-Reyes, R. A. Ramirez-Mendoza, M. R. Bustamante-Bello, S. Navarro-Tuch, and R. Avila- Vazquez, “Advanced driver monitoring for assistance system (admas),” International Journal on Interactive Design and Manufacturing (IJIDeM), vol. 12, no. 1, pp. 187–197, 2018.

[4] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, “The graph neural network model,” IEEE transactions on neural networks, vol. 20, no. 1, pp. 61–80, 2008.

[5] Y. Yin, X. Zheng, B. Hu, Y. Zhang, and X. Cui, “Eeg emotion recognition using fusion model of graph convolutional neural networks and lstm,” Applied Soft Computing, vol. 100, p. 106954, 2021.

[6] T. Song, W. Zheng, P. Song, and Z. Cui, “Eeg emotion recognition using dynamical graph convolutional neural networks,” IEEE Transactions on Affective Computing, vol. 11, no. 3, pp. 532–541, 2018.

[7] A. Fornito, A. Zalesky, and E. Bullmore, Fundamentals of brain network analysis. Academic Press, 2016.

[8] H. Li, Y.-M. Jin, W.-L. Zheng, and B.-L. Lu, “Cross-subject emotion recognition using deep adaptation networks,” in International conference on neural information processing. Springer, 2018, pp. 403–413.

[9] P. Zhong, D. Wang, and C. Miao, “Eeg-based emotion recognition using regularized graph neural networks,” IEEE Transactions on Affective Computing, 2020.

[10] X. Wang, L. Li, W. Ye, M. Long, and J. Wang, “Transferable attention for domain adaptation,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, no. 01, 2019, pp. 5345–5352.

[11] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, “Differential entropy feature for eeg-based emotion classification,” in 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER). IEEE, 2013, pp. 81–84.

[12] A. Schloegl, C. Neuper, and G. Pfurtscheller, “Estimating the mutual information of an eeg-based brain-computer interface,” 2002.