EEG Classification with Transformer-Based Models

Jiayao Sun[1,2], Jin Xie[1,2], Huihui Zhou[3]
[1]*Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences*
[2]*University of Chinese Academy of Sciences*
[3]*Peng Cheng Laboratory*
Shenzhen, China
{jy.sun, jin.xie}@siat.ac.cn, zhouhh@pcl.ac.cn

2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech) | 978-1-6654-1875-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/LIFETECH52111.2021.9391844

***Abstract*—Transformer has been widely used in the field of natural language processing (NLP) with its superior ability to handle long-range dependencies in comparison with convolutional neural network (CNN) and recurrent neural network (RNN). This correlation is also important for the recognition of time series signals, such as electroencephalogram (EEG). Currently, commonly used EEG classification models are CNN, RNN, deep believe network (DBN), and hybrid CNN. Transformer has not been used in EEG recognition. In this study, we constructed multiple Transformer-based models for motor imaginary (MI) EEG classification, and obtained superior performances in comparison with the previous state-of-art. We found that the activities of the motor cortex had a great contribution to classification in our model through visualization, and positional embedding (PE) method could improve classification accuracy. These results suggest that the attention mechanism of Transformer combined with CNN might be a powerful model for the recognition of sequence data.**

Keywords— EEG, Transformer, CNN, visualization, brain-computer interface

I. Introduction

Transformer uses multi-head attention instead of recurrent layer or convolutional layer to extract information, which improves the performance of multiple tasks in natural language processing (NLP)[1]. Recently, Transformer-based models have been developed for object detection[2], image classification[3] and protein engineering[4], suggesting its wide applicability.

Compared with convolutional neural network (CNN) and recurrent neural network (RNN), Transformer shows superior ability to deal with long-range dependencies[1], which indicates that it might be a good model for the recognition of sequence data because long-range dependencies is an important feature of time series. So far, there is still a lack of Transformer applications in time series data.

Motor imaginary (MI) task is a frequently used paradigm in brain-computer interface (BCI) based on Electroencephalogram (EEG) signals. So far, the commonly used models in EEG classification are CNN[5][6], RNN[7], and Hybrid CNN, while Transformer has not been used[8]. Transformer also has better interpretability than the above-mentioned deep learning models. In this study, we proposed multiple Transformer-based models and analyzed model behaviors in MI-EEG classification.

II. Dataset

The imagery dataset of the PhysioNet EEG Motor Movement/Imagery Dataset contains 109 subjects with a sampling rate of 160Hz and 64 channels[9][10]. We divided it into 3 data subsets: left fist/ right fist (L/R), left fist/ right fist/ eyes open (L/R/O), and left fist/ right fist/ eyes open/ feet (L/R/O/F) subset. We selected 21 trials from each class of each subject, and used 3s data (480 samples) and 6s data (960 samples). Z-score normalization and random noise were applied to prevent over-fitting.

III. Model Structure

A. Transformer-Based Models

We built five models: Spatial Transformer model with attention heads applied across EEG channels, Temporal Transformer model with attention heads applied across time, CNN + Spatial Transformer model, CNN + Temporal Transformer model, and fusion model that combined the CNN + Spatial and CNN + Temporal Transformer models. In our Transformer model, we used one positional embedding module and three attention modules.

The CNN module of the CNN+Spatial Transformer model includes two convolutional layers with 64 kernels and an average pooling layer to acquire EEG temporal information. In the Temporal CNN + Transformers model, the CNN module includes a convolutional layer with 64 kernels of size 64x1 to extract features across all EEG channels, and an pooling layer with the pooling size of 8 to average features across time.

B. Positional Embedding (PE)

We tested four positional embedding methods: no PE, sinusoidal PE, cosine similarity PE, and learned PE. For the sinusoidal PE, the sine and cosine functions are used to represent relative positions. *pos* is the index of the electrode from 0 to channel number. d_{model} is the dimension of features. i is the feature dimension from 0 to d_{model}.

$$PE_{(pos,2i+1)} = \sin\left(\frac{pos}{10000^{2i/d_{model}}}\right) \quad (1)$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{2i/d_{model}}}\right) \quad (2)$$

In the cosine similarity PE, we chose Cz as the central electrode and calculate the cosine similarities between the central position and other electrode positions. $P_{central}$ is the position of the center electrode, P_k is the position of electrode k, k from 1 to channel number. For the learned PE method, the embedding matrix is a trainable matrix, the parameters are updated during the training process.

$$sim_k = \frac{P_{central} \cdot P_k}{\|P_{central}\| \|P_k\|} \quad (3)$$

978-1-6654-1875-1/21/$31.00 ©2021 IEEE

IV. Results

A. Classificaiton Performance

Table I. shows the accuracies of our models with 5-fold cross-validation. The datasets for different classifications were arranged as in [5]. For comparison, we calculated the performances of Wang's model [6] for L/R/O classification and L/R/O/F classification with 109 subjects. Our Transformer fusion model achieved the highest accuracy of 82.95%, 74.44%, and 64.22% in 2, 3, and 4 class classifications in 3s dataset, respectively. For classifications based on 6s data, the CNN+Temporal Transformer model reaches 87.80%, 78.98%, and 68.54% in 2, 3, and 4 class classifications, respectively, which were better than the previous state-of-art.

B. Effects of Positional Embedding

We tested PE on different positional embedding methods. Table II shows the results from the Spatial Transformer model. Using sinusoidal PE, trained PE and cosine similarity PE can similarly increase the classification accuracy by 0.3 ~ 3%, suggesting that adding spatial information of EEG signals through PE is helpful for classification.

C. Model Visualization

We visualized attention weights of Transformer-based models. The weights of each head of each attentional layer were normalized to 0-1 using min-max normalization. Fig. 1. shows the visualization of Spatial Transformer model using 6s data for L/R classification. Across different attention layers, attention heads were more concentrated on the motor cortex, where the electrodes FC, C, and CP were located, suggesting that the activities of motor cortex had a great contribution to classification in our models.

TABLE I. Classification Accuracy (in %)

Model	480			960		
	L/R	***L/R/O***	***L/R/ O/F***	***L/R***	***L/R/O***	***L/R/ O/F***
Spatial Transformer	81.11	70.25	59.35	87.46	75.41	64.04
Temporal Transformer	80.77	70.31	58.21	86.10	75.24	62.15
CNN+ Spatial Transformer	82.90	72.43	63.07	87.66	76.97	67.97
CNN+ Temporal Transformer	82.56	72.87	63.48	87.80	78.98	68.54
Transformer Fusion	82.95	74.44	64.22	87.26	78.44	67.96
Hauke Dose et al. (2018)[5]	80.38	69.82	58.58	87.98	76.61	65.73
Xiaying Wang et al. (2020) [6]	82.43	75.07 (n=105) 72.33 (n=109)	65.07 (n=105) 63.16 (n=109)	-	-	-

TABLE II. Effects of PE on classificaiton accuracy in Spatial Transformer Model (in %)

PE method	3s (480 samples)			6s (960 samples)		
	L/R	***L/R/O***	***L/R/ O/F***	***L/R***	***L/R/O***	***L/R/ O/F***
no PE	81.13	68.25	57.23	86.83	73.15	61.43
sinosoidal PE	81.11	70.25	59.35	87.46	75.41	64.04
Cosine similarity PE	81.49	69.48	59.47	87.14	75.26	64.05
Trained PE	81.47	70.02	59.08	87.07	75.52	64.06

Fig. 1. Visualization of L/R classification using spatial Transformer model. (*subject 65 is used for illustration*)

V. Conclusions

This research established five new Transformer-based models for EEG classification, which obtained excellent performances in comparison with the state-of-the-art. These results suggest that the attention mechanism within the Transformer combined with CNN might serve as a powerful model for the recognition of time series data.

References

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., "Attention is all you need," in Advances in Neural Information Processing Systems, 2017, vol. 30, pp. 5998-6008.

[2] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with Transformer," unpublished.

[3] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, et al., "An image is worth 16x16 words: Transformer for image recognition at scale," unpublished.

[4] R. Rao, N. Bhattacharya, N. Thomas, Y. Duan, X. Chen, J. Canny, et al., "Evaluating protein transfer learning with TAPE," in Advances in Neural Information Processing Systems, 2019, vol. 32, pp. 9689-9701.

[5] H. Dose, J. S. Møller, H. K. Iversen, and S. Puthusserypady, "An end-to-end deep learning approach to MI-EEG signal classification for BCIs," Expert Systems With Applications, vol. 114, pp. 532-542, December 2018.

[6] X. Wang, M. Hersche, B. Tömekce, B. Kaya, M. Magno and L. Benini, "An accurate EEGNet-based motor-imagery brain–computer interface for low-power edge computing," 2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA), pp. 1-6, 2020.

[7] P. Wang, A. Jiang, X. Liu, J. Shang and L. Zhang, "LSTM-Based EEG Classification in Motor Imagery Tasks," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2018, vol. 26, no. 11, pp. 2086-2095.

[8] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," Journal of Neural Engineering, vol. 16, pp. 031001, April 2019.

[9] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, et al., "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," Circulation, vol. 101, pp. e215-e220, June 2000.

[10] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer and J. R. Wolpaw, "BCI2000: a general-purpose brain-computer interface (BCI) system," in IEEE Transactions on Biomedical Engineering, vol. 51, no. 6, pp. 1034-1043, June 2004.

978-1-6654-1875-1/21/$31.00 ©2021 IEEE