Into Paradigm-Independent Brain Computer Interfaces

*Note: Sub-titles are not captured in Xplore and should not be used

Merey Kairgaliyev
Department of Robotics and Mechatronics Engineering
Nazarbayev University
Nur-Sultan, Kazakhstan
merey.kairgaliyev@nu.edu.kz

Maxat Tassemenov
Department of Robotics and Mechatronics Engineering
Nazarbayev University
Nur-Sultan, Kazakhstan
maxat.tassemenov@nu.edu.kz

Aldiyar Bolatov
Department of Computer Science
Nazarbayev University
Nur-Sultan, Kazakhstan
aldiyar.bolatov@nu.edu.kz

Sultan Bauyrzhanuly
Department of Computer Science
Nazarbayev University
Nur-Sultan, Kazakhstan
sultan.bauyrzhanuly@nu.edu.kz

***Abstract*—In this paper, we conducted the experiment and proposed a network, which can classify the paradigm. After the model classified the paradigm of the input data, classic and powerful techniques can be used in the classification of the action inside the paradigm. Results are particularly encouraging, but more research with heavier computational resources can yield more general results.**

***Index Terms*—EEG, Classification**

I. Introduction

Brain-Computer Interface (BCI) is the technology that allows us to record brain signals and transform them into something useful [1]. It might be a more precise medical diagnosis, prosthetic body parts controlled with “thoughts”, etc., all of it are possible due to research in BCI. So, for advancements in this field, BCI requires one of its crucial components, data and it may vary. The recordings of electrical data are achieved by 2 different groups of methods: invasive and more common, noninvasive methods [1]. This report focuses on 3 noninvasive BCI paradigms and corresponding dataset, collected by Lee et al [2]. The electrical data was gathered using Electroencephalography (EEG)-based BCI and paradigms are: The first one is sensory rhythm (SMR)-based BCI, the next paradigm is Event-Related Potential (ERP) and the last one is steady-state visually evoked potential (SSVEP) [2]. During the data collection conducted by Lee et al, for SMR-based BCI, participants of the experiment were asked to perform motor imaging (MI) [2]. They needed to imagine the grasping with the correct hand based on the arrow’s direction on the screen [2]. For the EPR part, participants’ brain signals were recorded during the use of a 6x6 matrix speller. The SSVEP paradigm was implemented also based on people’s gaze as they needed to focus on the highlighted block [2]. In this report, we present a paradigm-independent classifier that is capable of decoding data regardless of which type of paradigm it is out of 3. It was based on, generated, trained, and tested on the dataset created by Lee et al [2]. We also describe and discuss the results and limitations discovered along the way.

II. Methods

A. Motor Imagery

For the implementation of the MI part, the methodology from Lee et al. [2] was followed. For preprocessing, the EEG data was band-pass filtered for frequencies between 8 and 30 and with 5th order Butterworth digital filter used. The data, which was processed, is from 20 electrodes corresponding to the motor cortex region, and they were selected to test the performance. The selected electrodes are C-5/3/1/z/2/4/6, FC-5/3/1/2/4/6, and CP-5/3/1/z/2/4/6. With respect to stimulus onset, the data was divided between 1000 and 3500 ms to form EEG epochs as 250x20x100 (data points x electrodes x trials). Also, in order to maximize the discrimination of the binary class, the Common Spatial Pattern (CSP) was chosen [3]. Next, Log-variance and then the LDA classifier was calculated.

B. SSVEP

For the classification of Steady-State Visual Evoked Potential paradigm, the method mentioned in [2] was used, or namely Canonical Correlation Analysis (CCA).

Initially, the loaded data from each subject’s mat file was preproccessed to have the following format: (100, 62, 4000), where 100 stands for trials, 62 is the number of channels and 4000 is the number of datapoints for each trial corresponding to 4 seconds time intervals of trials. Then, in order to get rid of excess channels, only 10 channels provided in the [2] were chosen. In addition, the datapoints were downsampled from 4000 to 400. Finally, CCA algorithm from sci-kit learn library was performed using for the X values the sample of one trial in form of (10 channels, 400 datapoints) and Y values were calculated according to the formula in Fig. 1 provided in [2].

$$Y_i(t) = \begin{bmatrix} \sin(2\pi f_i t) \\ \cos(2\pi f_i t) \\ \sin(2\pi (2 f_i) t) \\ \cos(2\pi (2 f_i) t) \end{bmatrix}, t = \frac{1}{S}, \frac{2}{S}, \ldots, \frac{T}{S}$$

Fig. 1. Calculation of Y value for CCA

C. ERP

Method described in [2] was chosen for the classification of Event-Related Potential BCI paradigm. The preprocessing part follows the one described in [1]. 32 channels of raw EEG data were bandpass filtered between 0.5-40Hz with 5th order Butterworth filter. Then, elements on the pre-stimulus were picked to perform baseline averaging on the rest. Further, the rest were divided into sub-intervals of 100ms to obtain mean values in each, thus, obtaining the feature vectors. Those are to calculate Linear discriminant analysis (LDA) classifier to perform target/non-target classification.

D. Paradigm Classification

Due to the fact that handcrafted methods are ultimately biased because they rely on inductive assumptions, and therefore, can corrupt or delete discriminative data for classification of the Paradigm, we decided to use all down-sampled data. To make the classificator less biased, we decided to use the most general deep learning model - transformer. The text below was already used in my different work, so I re-used it with appropriate changes (Aldiyar).

The transformer architecture previously surpassed the RNN-based models in natural language task (NLP) and became the standard model for sequence-related problems [4]. This is very helpful with EEG-sequential data.

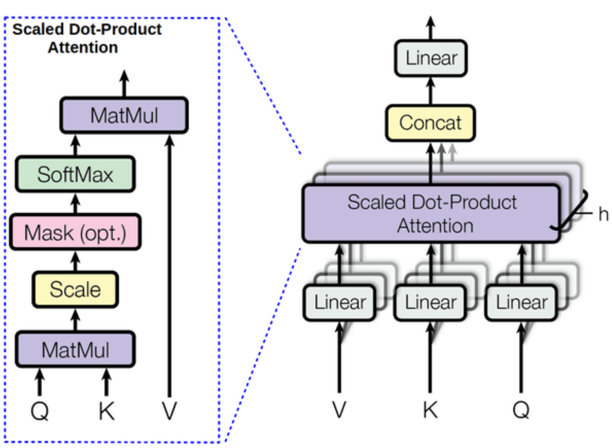

Fig. 2. Multi-head Self-Attention Structure

The overall model architecture can be seen in figure 2. Further, as in BERT [5], a learnable class token is prepended to the EEG - sequence, which output states of the transformer is used for the classification. Also, because of the fact, those transformers are not aware of the spatiality of the information, the data was positionally embedded.

This sequence is then fed into a three-layer transformer network. The transformer network itself consists of two main layers: multi-head self-attention (SA) and feed-forward neural network (FF).

The self-attention function is a process of quantifying the representation of the relative importance of each sequence element. Here, the self-attention mechanism relates each embedded patch with all other patches of the sequence. Then, this new information and relative importance added to the value representation [4]. This is done using the compatibility function. In our work, we used softmax scaled dot-product.

$$f_{scaled}(Q, K) = \frac{Q, K^T}{\sqrt{d_k}} \quad (1)$$

Then new value representation is then computed using:

$$f_{at}(Q, K, V) = V\text{'} = softmax(f_{scaled}(Q, K))V \quad (2)$$

As Vaswani et al. [4] specified, Q, K, V are three different representation of the input sequence, which were acquired using separate learned linear transformation. Usually, the self-attention mechanism is split between heads, which are then concatenated. The purpose is that each head can capture unique features by having individual parameters.

The feed-forward neural network in our work has a form of gated activation linear unit, which is followed by linear projection to the same space as input. There, swish activation function was used [6]. In the formula 3, $LN \in R^(D \times D')$ stands for learned linear projection.

$$f_{ff}(X) = LN_D(LN_{glu1}(X) \times swish(LN_{glu2}(X))) \quad (3)$$

The transformer utilizes Layernorm (LN) and residual connections, which both help to better gradient flow. The overall equations that took place in every block (t) of the transformer are presented below with multi-layer perceptron (MLP), which takes output states of the class token.

$$X_t^{'} = SA(LN(X_{t-1})) + X_{t-1} \quad (4)$$

$$X_t = FF(LN(X_t^{'})) + X_t^{'} \quad (5)$$

$$class = MLP(LN(X_3^0)) \quad (6)$$

III. Results

The results for the classification of the paradigm can be divided into two parts. We took the data from three corresponding paradigm datasets. Then we merged them by trials and we got at the end, three classes. However, here comes two different approaches or parts. First, we can use data from train-data sessions for training and test-data sessions for testing. Or in the second part, we can merge both test and train data in order to get more training data and took a subpart of it (10%) for testing. We did both. The table below shows that by the first approach, the result is pretty poor due to the fact that there is no enough data to train well (and we know that deep learning models especially transformers need a LOT). However, when we merge the data, the results are very high, but it should be considered that by merging we are making a model little bias by merging data. Results for motor imagery and SSVEP are high separately from the method of data separation. But, because the fact, that ERP implementation was unsuccessful, we were not able to conduct a one-for-all classifier, but it is expected to be high for the second part (only 4-5% of data paradigms will be misclassified).

TABLE I
RESULTS

Task	Part1	Part2
Paradigm	54	96.25
MI	97.5	91.3
SSVEP	100	100

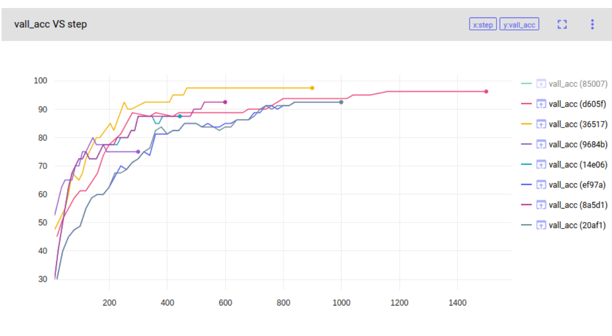

Fig. 3. Validation accuracy for different hyper-parameters

IV. Limits and Potentials

There were several limitations during the working process on this project. Starting from the technical issues, such as incompatibility of toolbox in Matlab, low computational powers and slow internet connection for uploading datasets, and to remote team working conditions. However, judging from the results from successfully implemented parts of the project, the proposed methodology has a significant potential for one-for-all classifier with high accuracy if the arised issues will be solved. Moreover, we have already strong code-backbone, which can make the further research easy.

V. Conclusion

To conclude, the proposed steps regarding all three paradigms were performed according to the methodology provided in [2]. Even though, the attempt to implement ERP was unsuccessful, the results for MI and SSVEP were above expectations by showing accuracies of 97.5% and 100%, respectively for train-data sessions of a single subject. However, implementing the same workflow for other subsets of data combinations, both showed results similar to aforementioned. Due to unsuccessful implementation of ERP paradigm, the one-for-all classification has not been performed. However, the implementation of ERP paradigm would likely result in high-performing general classifier for BCI systems. Using the strong code-backbone, created models, further research into the area can be faster after problems like ERP implementation would be solved.

References

References

[1] R. Leeb, S. Perdikis, L. Tonin, A. Biasiucci, M. Tavella, M. Creatura, A. Molina, A. Al-Khodairy, T. Carlson, and J. dR Millán, "Transferring brain–computer interfaces beyond the laboratory: successful application control for motor-disabled users," *Artificial intelligence in medicine*, vol. 59, no. 2, pp. 121–132, 2013.

[2] M.-H. Lee, O.-Y. Kwon, Y.-J. Kim, H.-K. Kim, Y.-E. Lee, J. Williamson, S. Fazli, and S.-W. Lee, "Eeg dataset and openbmi toolbox for three bci paradigms: an investigation into bci illiteracy," *GigaScience*, vol. 8, no. 5, p. giz002, 2019.

[3] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial eeg during imagined hand movement," *IEEE transactions on rehabilitation engineering*, vol. 8, no. 4, pp. 441–446, 2000.

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[5] I. Tenney, D. Das, and E. Pavlick, "Bert rediscovers the classical nlp pipeline," *arXiv preprint arXiv:1905.05950*, 2019.

[6] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," *arXiv preprint arXiv:1710.05941*, 2017.