Improved Decoding of EEG-Based Motor Imagery Using Convolutional Neural Network and Data Space Adaptation

Shawn Chua
National University of Singapore
Singapore
a0129612@u.nus.edu

Yang Tao
Institute for Infocomm Research
*Agency for Science, Technology and Research (A*STAR)*
Singapore
tyang@i2r.a-star.edu.sg

Rosa Q. So
Institute for Infocomm Research
*Agency for Science, Technology and Research (A*STAR)*
Singapore
rosa-so@i2r.a-star.edu.sg

***Abstract*—One challenge in decoding EEG-based motor imagery for brain computer interface (BCI) is the inter-session non-stationarities due to factors like subject fatigue or electrode placements. This non stationarity leads to lower decoding accuracy, especially if data from the training and testing sessions are collected on separate days. An existing technique termed data space adaptation (DSA) transforms EEG data such that the distribution of test and training data are aligned. Decoding accuracy has been shown to improve when transformed data was used for classification instead of the original data. In this study, space transformed data from six subjects were fed into a Convolutional Neural Network (CNN) and a Filter Bank Common Spatial Pattern (FBCSP) model. The results show that when no adaptation was applied, there was on average a 2.8% improvement in decoding accuracy when CNN was used compared to FBCSP. When DSA was used in conjunction with FBCSP, the majority of the subjects had improved decoding accuracy. However, DSA together with CNN on average yielded a lower decoding accuracy compared to CNN alone. Hence the results suggest that when decoding non-stationary EEG data for motor imagery, DSA should be used only in conjunction with FBCSP, but not when CNN is used as a decoding model.**

I. Introduction

In EGG-based brain-computer interface (BCI), the electrical activity of the brain is recorded, which allows the brain to communicate directly with a computer [1]. Within the field of BCI, motor imagery-based BCI research has expanded rapidly due to its potential for direct communication between the brain and a computer without additional external stimuli [1]. Interfaces like these have the potential for motor rehabilitation and control.

One challenge in EEG motor-imagery based BCI is non-stationarities in the recorded EEG signals [2], which can be due to subject fatigue, difference in electrode placements and impedance [3]. Models trained using data from one experimental session cannot decode data from another session accurately, especially if collected on separate days [3]. Many papers have proposed new algorithms to deal with this issue. To improve inter-session stability, Coyle, Prasad, and McGinnity [4] used a time-series-prediction technique which uses neural networks to perform one-timestep-ahead prediction on the EEG signals to produce new signals which then have their features extracted and classified. Chai et al. [3] found common features between the training and testing data by combining auto-encoders with the subspace alignment solution to align the source space (from the training data) to the target space (from the testing data). This study used a data space adaptation (DSA) technique proposed by Arvaneh, Guan, and Ang [2]. The DSA technique transforms the EEG data from test sessions so that the distribution is more similar to that from the training sessions.

In addition, we compared the performance of a Convolution Neural Network (CNN) to the commonly used filter bank common spatial pattern (FBCSP) model. CNN has received recognition for its outstanding performance in image classication and is increasingly being used in EEG-based decoding as it allows end-to-end learning without a need for feature extraction [5]. Tayeb et al. [5] developed a spectogram-based CNN and a recurrent CNN to decode raw motor-imagery based EEG signals without having to perform manual feature extraction. Zhang, Yan, and Gong [6] pre-processes the EEG signal by first transforming them into time-frequency images using Short-Time Fourier Transform which are then fed into a CNN. The CNN used in this paper was developed by Schirrmeister et al. [7].

II. Methodology

A. Dataset

In this paper, data from six healthy human subjects were used, all with SingHealth IRB approval. Data from each subject was collected from two different experimental sessions conducted on separate days. For each experimental session, the subjects performed between 80-120 trials. For each trial, the subjects were instructed to perform between 4-10 seconds of motor imagery (right vs left for subject 1, right vs idle for the remaining subjects). Only signals from the first 4 seconds of

This work is partially supported by BMRC-EDB JCO DP grants IAF311022 from the Agency for Science, Technology and Research Singapore.

each trial were used in subsequent analysis. For all subjects, data from the training sessions was used to train a CNN or FBCSP model while data from the testing sessions was used for adaptation and decoding. All decoding was assessed based on 2-class classification.

B. Data Pre-Processing

For each of the adaptation method, data from each subject was processed by bandpassing between 4 Hz to 40 Hz.

C. Data Space Adaptation

The DSA technique used in this paper was developed by Arvaneh, Guan, and Ang [2] and its details are provided as follow. For each subject k, the training data is denoted by $\overline{D_k} = (\overline{x_i}, \overline{y_i})_{i=1}^{\overline{N_k}}$ where $\overline{N_k}$ is the total number of trials recorded for the subject in the training session and each i^{th} trial recorded is $\overline{x_i} \in \overline{X} \subset \mathbb{R}^{n\times t}$ with n being the number of channels and t the number of timepoints for each trial. $\overline{y} \in \overline{Y} = \{0,1\}$ is the corresponding class label for each trial. $\overline{Y} = \{0,1\}$ as each subject has only two classes.

The testing data contains labelled EEG trials obtained from another session and are denoted by $D_k = (x_i, y_i)_{i=1}^{N_k}$. $x_i \in X \subset \mathbb{R}^{n\times t}$ is the i^{th} recorded trial and $y_i \in Y = \{0,1\}$ is its corresponding class label.

The data space adaptation technique aims to find an optimal transformation matrix $V \in \mathbb{R}^{n\times n}$ to transform each EEG trial to $z = V^T x$ so that the distribution difference between the training data and testing data is minimised. The transformed test data can then be used instead of the original for decoding so that higher accuracy can be obtained.

As each EEG trial was being bandpassed, they have approximately zero mean and so the average distribution for a group of EEG trials can be approximated by a Gaussian distribution with zero mean and co-variance matrix given by

$$\Sigma = \frac{1}{N}\sum_{i=1}^{N}\frac{x_i x_i^T}{tr(x_i x_i^T)} \tag{1}$$

which is the average co-variance matrix over N EEG trials, where $tr(x)$ gives the sum of the diagonal elements of x.

The average distribution of EEG trials for class j in the source space j is denoted by $N(0,\overline{\Sigma}_j)$ where $\overline{\Sigma}_j$ is calculated using equation (1). From the target space, the average distribution of the transformed EEG trials for class j is estimated to be $N(0, V^T\Sigma_j V)$ where V is the optimal linear transformation matrix to be calculated. The Kullback-Leibler (KL) criteria was used to calculate the difference between two Gaussian distributions with the same dimension and the goal is to minimise the sum of the KL divergence between average distributions of the source and target space belonging to the same class. The equation for optimal V was calculated in [2] and is given by

$$V^* = \sqrt{2}(\overline{\Sigma}_1^{-1}\Sigma_1 + \overline{\Sigma}_2^{-1}\Sigma_2)^{-0.5} \tag{2}$$

Using the calculated V, each trial in the target space was optimally transformed using equation using $z = V^T x$ and then applied on any classification models trained using data from the source space. Ideally, the transformed trials achieve better accuracy than non-transformed trials as they are more similar to the source space.

D. Single and Continuous adaptation

Two modes were used for DSA: single adaptation and continuous adaptation. In single adaptation, the first t trials per class from the testing sessions were used to calculate V, which was used to transform all trials used for testing. In continuous adaptation, a different V was calculated for each testing trial using the immediate past t trials. Thus, the first t trials per class from the testing session were used only for adaptation. When no adaptation was performed, the same t trials per class were removed so as to provide a fair comparison. Classification was thus performed only on the remainder trials. In this paper, t ranges from 5 to 20.

E. Convolutional Neural Network

The CNN model used in this paper was developed by Schirrmeister et al. [7]. A CNN model was trained using data from the source space and subsequently it was used to decode the transformed EEG trials from the target space.

F. Models for Comparison

To evaluate the effectiveness of DSA on CNN, the same technique was used on the FBCSP model which performs feature extraction and classification. The FBCSP algorithm [8] consists of four progressive stages: multiple bandpass filters applied on EEG signals, spatial filtering using the CSP algorithm, CSP feature selection and classification using the selected features. In this paper, DSA was done prior to the first step of the FBCSP algorithm.

III. Results

A. Baseline accuracy for no adaptation

Subject	CNN	FBCSP	Better Performance
1	60.4 ± 0.3	54.2 ± 0.1	CNN
2	61.4 ± 0.7	52.7 ± 1.5	CNN
3	80.9 ± 0.6	71.5 ± 1.0	CNN
4	77.2 ± 1.2	74.8 ± 0.8	CNN
5	86.0 ± 0.9	89.4 ± 0.6	FBCSP
6	79.3 ± 1.5	86.2 ± 1.1	FBCSP

Table I – Average decoding accuracy for no adaptation per subject, across the number of trials (5-20), for both CNN and FBCSP. The second and third column states the mean and standard deviations of the decoding accuracy. A paired t-test was conducted at a p-value of 0.05 to determine whether there is a significant change when switched from FBCSP to CNN, and the model which performed significantly better is listed in the fourth column.

As shown in Table I, when no adaptation was applied, CNN resulted in significantly higher decoding accuracy in four out of six subjects when compared to FBCSP. The remaining two subjects achieved a decoding accuracy rate of more than 85% with FBCSP. This provided us with a baseline for comparison between the CNN and FBCSP model.

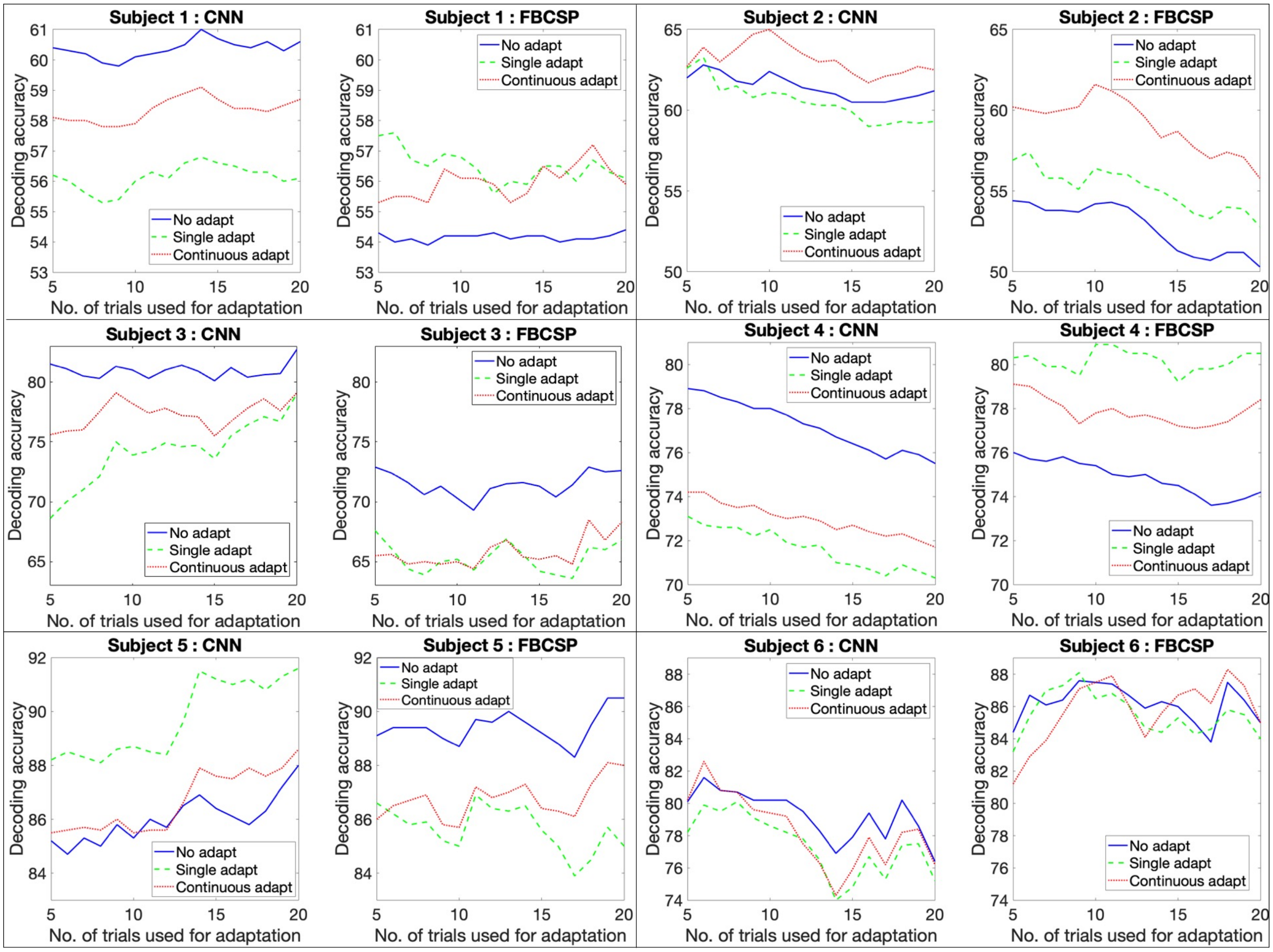

Figure 1 – Decoding accuracy for each of the six subject, using different number of trials per class for adaptation. For each subject, decoding accuracy using CNN is on the left while that of FBCSP is on the right. Solid blue lines are for no adaptation, green dashed lines are for single adaptation and red dotted lines are for continuous adaptation.

B. DSA on decoding accuracy

For decoding using FBCSP, Figure 1 shows that there is an increase in decoding accuracy after implementing DSA (both single and continuous) for the majority of the subjects (subjects 1, 2, 4). Subject 6 saw an improvement only when using a specific number of trials for adaptation, while subjects 3 and 5 performed worse across all trials.
For CNN, subjects 1, 3 and 4 performed worse when using the DSA technique, and saw a general improvement in subjects 2 and 5. There was only an improvement when using 6 trials for single adaptation in subject 6.
Only subject 3 saw a decrease in decoding accuracy for both CNN and FBCSP, while the rest of the subjects had a mix of performance between the two models.

When the relative changes in decoding accuracy was examined (Figure 2), it can be seen that FBCSP outperformed CNN when the DSA technique is applied. Both single and continuous adaptation improved decoding accuracy by an

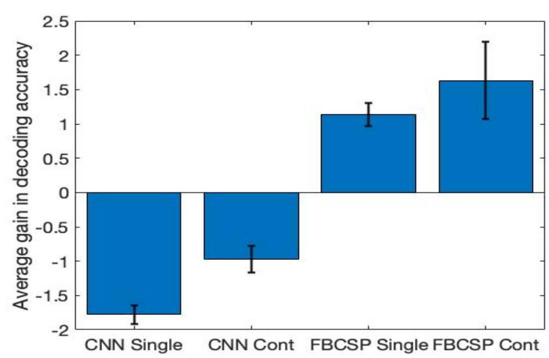

Figure 2 – Across the number of trials used for adaptation, the average gain in accuracy (from no adaptation to adaptation) across all subjects, when using single and continuous adaptation.

average of more than 1% for FBCSP while both decreased by an average of more than 0.5% when CNN was used.

IV. Discussion

In this paper, we presented a comparison of various methods (CNN / FBCSP in combination with DSA) to address the issue of non-stationarity of EEG data. Baseline decoding accuracy when no adaptation was applied suggests that CNN is able to achieve a significantly better performance than FBCSP in the majority of the subjects, especially for subjects with relatively low decoding accuracy performance using FBCSP. When FBCSP already achieves high decoding accuracy, CNN was not able to improve the results any further.

When DSA was applied, results show that it improves FBCSP decoding but not decoding with CNN. Average decoding accuracy across all subjects increased when both single and continuous adaptation was used in conjunction with FBCSP, while it decreased when CNN was used. This suggests that the DSA technique is not suitable to be used with all decoding models. The decrease in accuracy when DSA was used together with CNN could be due to CNN's ability to learn features which are stable across sessions from the raw data, without having to subject the data to DSA. Distribution difference between the source and target space might have little effect on the decoding accuracy when using CNN.

In conclusion, CNN was more effective than FBCSP in decoding EEG-based motor imagery, though not for subjects who already achieved high decoding accuracy using FBCSP. The DSA technique was also shown to improve decoding using the FBCSP model but not the CNN model. These results suggests that DSA may be suitable for use only in conjunction with certain decoding models. The current analysis was performed using 2-class decoding. Future work will include exploring multi-class DSA technique proposed by Giles, Ang, Mihaylova, and Arvaneh [1] to determine the viability of CNN and adaptation technique when more classes are added.

References

[1] Joshua Giles, Kai Keng Ang, Lyudmila Mihaylova, and Mahnaz Arvaneh. Data space adaptation for multiclass motor imagery-based BCI. In *IEEE Engineering in Medicine and Biology Society*, Honolulu, HI, USA, 2018. IEEE.

[2] Mahnaz Arvaneh, Cuntai Guan, and Kai Keng Ang. EEG data space adaptation to reduce intersession nonstationarity in brain-computer interface. *Neural Computation Massachusetts Institute of Technology*, 25:2146–2171, 2013.

[3] Xin Chai, Qisong Wang, Yongping Zhao, Xin Liu, Ou Bai, and Yongqiang Li. Unsupervised domain adaptaion techniques based on auto-encoder for non-stationarity EEG-based emotion recognition. *Computers in Biology and Medicine*, 79:205–214, 2016.

[4] Damien Coyle, Girijesh Prasad, and Thomas M McGinnity. Improving signal separability and inter-session stability for a brain-computer interface by yime-series-prediction-processing. In *IEEE Engineering in Medicine and Biology Society*, Shanghai, China, 2005.

[5] Zied Tayeb, Juri Fedjaev, Nejla Ghaboosi, Christoph Richter, Lukas Everding, Xingwei Qu, Yingyu Wu, Gordon Cheng, and Jörg Conradt. Validating deep neural networks for online decoding of motor imagery movements from EEG signals. *MDPI, Sensors*, 19(1):210, 19.

[6] Jin Zhang, Chungang Yan, and Xiaoliang Gong. Deep convolutional neural network for decoding motor-imagery based brain computer interface. In *IEEE International Conference on Signal Processing, Communications and Computing*, Xiamen, China, 2017. doi: https://doi.org/10.1109/ICSPCC.2017.8242581.

[7] Robin Tibor Schirrmeister, Jost Tobias Springberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*, 38:5391–5420, 2017.

[8] Kai Keng Ang, Zheng Yang Chin, Chuanchu Wang, Cuntai Guan, and Haihong Zhang. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. *Frontiers in Neuroscience*, 6:1–9, 2012.