Proceedings Paper:

Giles, J., Ang, K.K., Mihaylova, L.S. orcid.org/0000-0001-5856-2223 et al. (1 more author) (2018) Data space adaptation for multiclass motor imagery-based BCI. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 40th International Conference of the IEEE Engineering in Medicine and Biology Society, 18-21 Jul 2018, Honolulu, Hawaii. IEEE , pp. 2004-2007. ISBN 978-1-5386-3646-6

https://doi.org/10.1109/EMBC.2018.8512643

Data Space Adaptation for Multiclass Motor Imagery-based BCI

Joshua Giles,[1] Kai Keng Ang[2], Lyudmila Mihaylova[1], Mahnaz Arvaneh[1]

***Abstract*— Various adaptation techniques have been proposed to address the non-stationarity issue faced by electroencephalogram (EEG)-based brain-computer interfaces (BCIs). However, most of these adaptation techniques are only suitable for binary-class BCIs. This paper proposes a supervised multiclass data space adaptation technique (MDSA) to transform the test data using a linear transformation such that the distribution difference between the multiclass train and test data is minimized. The results of using the proposed MDSA on BCI Competition IV dataset 2a improved the classification accuracy by an average of 4.3% when 20 trials per class were used from the test session to estimate adaptation transformation. The results also showed that the proposed MDSA algorithm outperformed the multi pooled mean linear discrimination (MPMLDA) technique with as few as 10 trials per class used for calculating the transformation matrix. Hence the results showed the effectiveness of the proposed MDSA algorithm in addressing non-stationarity issue for multiclass EEG-based BCI.**

I. Introduction

EEG-based brain-computer interfaces (BCIs) are systems which use the electrical signals generated from the user's brain to allow communication directly between the brain and a computer interface [1]. Among the different types of BCI, motor imagery-based BCI is a rapidly advancing area of research due to its capability of allowing direct communication between the brain and a computer without the need for additional external stimuli [1]. These interfaces have the potential to help a range of people who struggle to communicate with the outside world due to lack of muscular control or damaged neural pathway [2], although, there are currently flaws with BCI systems. High accuracy can be achieved by the majority of users, however for 20 to 25% of people [3] the interface is unable to produce the minimum accuracy of 70% [3]. The people who do find BCI effective then find they require 20-30 minute long calibration sessions [4] where the filters, feature extraction techniques and classifiers are retrained, before each use due to the non-stationary nature of the EEG signals being recorded. After these calibration sessions the BCI can then often be presented with the problem of low information transfer rates (ITR).

In order to improve the accuracy of users facing BCI deficiency, and reduce the calibration time required before each session a lot of research has been done to develop optimal components within the BCI. These range from optimizing the feature extraction through implementing a bank of common spatial pattern filters [5] to adaptive linear discrimination analysis classifiers which updates the global mean [6] referred to as pooled mean linear discrimination analysis. Through years of research the calibration time required has slowly been reduced and the levels of accuracy improved allowing faster communication between the user and the computer [3], [4], [6], [7],however the majority of this research has been focused on binary class BCIs.

Recently there has been a shift from binary class towards multi-class BCI systems, this is due to the opportunity they present of drastically increasing ITR. Multi-class BCIs have the potential to allow faster communication with the user as well as control of complex actuators providing more degrees of freedom. Due to this potential, research is being conducted on multi-class BCIs, with different BCI components such as feature extraction techniques [8] and classifiers [7] being compared and optimized.

A number of adaptation techniques initially developed for binary BCIs have been modified to be applicable in multi-class BCIs to explore their viability when additional classes are present. Pooled mean linear discrimination analysis [6] has been altered to multi-class pooled mean linear discrimination analysis (MPMLDA) [9] allowing it to work within a multi-class setting. Other adaptive classifiers such as the enhanced Bayesian linear discrimination analysis [10] have also been developed as an adaptive classifier for multi-class BCIs. These altered adaptation techniques have proven to be effective at reducing the fall in accuracy caused by the non-stationary nature of EEG, but there is still a lot of room for improvement.

Data space adaptation (DSA) is a method of changing the distribution of data directly before it has gone through feature extraction or classification [11]. This method minimizes the distribution difference between the data used to train the BCI and the data being tested using a linear transform. This means that DSA is not restricted by any particular feature extraction techniques or classifiers. In this paper, DSA is modified so it can be applied to multi-class BCIs. In the case of unsupervised DSA the number of classes does not affect the algorithm however supervised DSA does require altering due to the change in the number of classes. The proposed Multiclass data space adaptation (MDSA) will be evaluated using BCI Competition IV dataset 2a [12]. The proposed MDSA will then be compared to two other adaption methods, unsupervised DSA [11] and MPMLDA [9], providing an evaluation of the algorithms ability to improve multiclass BCI ITRs.

* This work was supported by University of Sheffield, United Kingdom and the Agency for Science, Technology and Research (A*STAR), Singapore. *Asterisk indicates corresponding author.*

[1] J. Giles, L. Mihaylova, and M. Arvaneh are with Automatic control and System Engineering Department, University of Sheffield , Sheffield, United Kingdom (emails: jgiles1, l.s.mihaylova, m.arvaneh@sheffield.ac.uk).

[2] K. K. Ang is with Institute for Infocomm Research,Agency for Science, Technology and Research (A*STAR), Singapore 138632. (email: kkang,@i2r.a-star.edu.sg).

The remainder of this paper is organized as follows. Section II describes the proposed MDSA as well as the data used for evaluation, Section III contains the results collected the algorithms implementation and Section IV concludes the paper.

II. Methodology

A. Proposed Multi-class Data Space Adaptation (MDSA)

The proposed MDSA is an extension of supervised binary DSA [11] allowing the adaptation method to be incorporated into a multi-class BCI. This adaptation method is implemented to alter the test EEG data being tested, after it has been band-pass filtered, so it is as similar to the data used for training as possible through the application of a linear transform. The training data is collected from a separate session and is used to train the feature extraction method and classifier.

Assume the training data is defined as $\overline{\mathrm{D}} = (\overline{\mathrm{x}}_i, \overline{\mathrm{y}}_i)_{i=1}^{N}$ where for each i^{th} trial recorded $\overline{\mathrm{x}}_i \in \overline{\mathrm{X}} \subset R^{n\times t}$ is the recorded data with n being the number of channels and t representing the time sample. $\overline{\mathrm{y}}_i \in \overline{\mathrm{Y}} \subset \mathbb{R}$ represents the corresponding class label. The test data contains a few labelled EEG trials collected from the same user in a second session. In this data $D = (x_i, y_i)_{i=1}^{N_l}$ where $\mathrm{x_i} \in \mathrm{X} \subset R^{n\times t}$ is the i^{th} recorded trial and $y_i \in Y \subset \mathbb{R}$ represents its corresponding class label. The proposed MDSA aims to use a linear transform, $V \subset \mathbb{R}^{n\times n}$, to minimize the distribution difference between the training data and the test data. The ideal goal of V is to have the adapted test data $S(V^T X, Y)$ to have the same distribution of data as the trained data, so that both the feature extraction and classification obtain optimal results.

In order to calculate the optimum V a few characteristics of the training and test data distributions must be known. The normalized co-variance matrix of the EEG data can be estimated using the EEG data x, as shown in (1), where tr is the trace, known as the sum of the diagonal of the matrix; While the mean is zero due to the EEG signal being band-passed. The EEGs data distribution can be modelled as Gaussian based on the maximum entropy principle [13] with zero mean and the co-variance matrix calculated.

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

The difference between the two Gaussian distributions can then be calculated using the Kullback Leibler criteria [13] as shown (2) as they have the same dimension k. The Gaussian distributions used to demonstrate the KL divergence are shown as $N_0(\mu, \Sigma)$ and $N_1(\overline{\mu}, \overline{\Sigma})$ with $\overline{\mu}$ and μ representing the means of the distribution while $\overline{\Sigma}$ and Σ co-variances.

$$\begin{aligned} KL[N_0 \parallel N_1] = \frac{1}{2}[(\overline{\mu}-\mu)^T\overline{\Sigma}^{-1}(\overline{\mu}-\mu) \\ +tr(\overline{\Sigma}^{-1}\Sigma) - ln(\frac{det(\Sigma)}{det(\overline{\Sigma})}) - k], \end{aligned} \tag{2}$$

To find the optimum V for supervised adaptation, the KL divergence is applied on the training and testing data of each class separately. In order to minimize the total loss function from across all the classes the differences are summed before the V is calculated. The transformed test data distribution is defined as $N_t(0, V^T\Sigma_j V)$ and training data distribution as $N_s(0, \overline{\Sigma_j})$ for class j in (3) while m is used to represent the total number of classes in the BCI.

$$L(v) = min\sum_{j=1}^{m}\frac{1}{2}[tr(\overline{\Sigma}_j^{-1}V^T\Sigma_j V) - ln(\frac{det(V^T\Sigma_j V)}{det(\overline{\Sigma})})] \tag{3}$$

To find the optimum V that minimizes L given in (3), the first derivative of L is calculated with respect to v and set to zero, as shown in (4).

$$\frac{dL}{dv} = \sum_{j=1}^{m}\frac{1}{2}[2tr(\overline{\Sigma}_j^{-1}\Sigma_j V) - 2tr(V^{-1})] = 0 \tag{4}$$

$$V = m^{-0.5}\sum_{j=1}^{m}(\overline{\Sigma}_j^{-1}\Sigma_j)^{\dagger 0.5} \tag{5}$$

Using (5) the optimum V is calculated then applied to the test data before it has the features extracted and classified using components previously trained with the training data. In (5) † represents the pseudo-inverse.

B. Adaptation Techniques for Comparison

In order to assess the effectiveness of the proposed MDSA algorithm, the experimental results were compared against the results of two other alternative adaptation techniques, described below:

1) Multi-pooled mean linear discriminant analysis: In this study, MPMLDA [9] is one of the two adaptation methods used for comparison. This method adapts linear discriminate analysis (LDA) classifier used in the multi-class BCI by updating the global mean, $\mu_{i,j}$, of each of the pair-wise LDAs as new trials are classified. The change caused by the new data is weighted by the probability of the new data belonging to a relevant class for the LDA as shown in (6). Here i and j represent the two classes the LDA is classifying, $P_i(x)$ and $P_j(x)$ are the probabilities of the previous trial being that class and is the learning rate, β, set to 0.03 as suggested in [9]. The updated global mean, $\mu'_{i,j}$, is then utilized to recalculate the LDA before the next trial is classified.

$$\mu'_{i,j} = (1-(P_i(x)+P_j(x))\beta)\mu_{i,j} + (P_i(x)+P_j(x))\beta x \tag{6}$$

2) Unsupervised data space adaptation: The unsupervised DSA technique (DSA-US) does not require altering due to the fact it is independent from the classes relying only on the EEG data; as such it is used as a second comparison for the proposed MDSA. This method also uses a linear transform to adapt the test data to the trained data however it does not split the data into its classes. DSA-US uses all the data at once to calculate the optimum linear transform as shown (7), where Σ represents the co-variance.

$$V_{unsupervised} = \overline{\Sigma}^{-0.5}\Sigma^{0.5} \tag{7}$$

C. Experiment

The dataset used for these BCI is the publicly available data set, BCI Competition IV dataset 2a [14]. This data set contains EEG data from nine users who each completed two sessions, each containing six runs, on different days. Each run consists of 48 trials containing 12 trials from each of the four classes making a total of 288 trials from each session. The four classes are all variations of motor imagery with the user imagining the movement of their right hand, left hand, both feet or tongue. To examine the adaptation capabilities of the different techniques the first session was used to train the common spatial patterns (CSP) and LDA which are then used on the second session with the techniques as testing data.

D. Data processing

The EEG data for each user was split into its different sessions, one used to training and the other used for testing the adaptation methods. To allow the adaptation methods to make some progress the first 80 trials of the testing data were set aside for adaptation and not included in the results. The same processing was performed for each of the adaptation methods. After the training data was band pass filtered from 8Hz to 35Hz, a pair wise CSPs was trained for 6 class pairs and then these features were used to create and train 6 pair wise LDAs.

III. Results and Discussion

A. Adaptation accuracy

Fig. 1 shows the increase in accuracy for each of the different adaptation algorithms across different number of trials used for adaptation. As shown in Fig. 1, compared to the base BCI design without any adaptation, all the three examined adaptation algorithms improved the average classification accuracy of the test data. MPMLDA does not require any test trials initially provided to calculate the adaptation parameters as it updates the global mean after every new trial added to the test data. Thus, the MPMLDA accuracy presented in Fig. 1 is fixed across the x-axis.

Initially when only 10 trials per class are used for adaptation, there is a very little difference between the accuracies of the three adaptation algorithms. The limited number of trials may have restricted the accurate estimation of the adaptation parameters in both MDSA and DSA-US as the estimation could be easily distorted by artefact corrupted trials. By increasing the number of trials per class to 15, DSA-US slightly outperformed the MDSA algorithm. In this case estimation of co-variance matrices of test data was based on 60 trials in DSA-US compared to 15 trials in the proposed MDSA. Having more trials for estimating co-variance matrix in DSA-US could have led to a better estimation of adaptation matrix and subsequently better results although using unlabelled data.

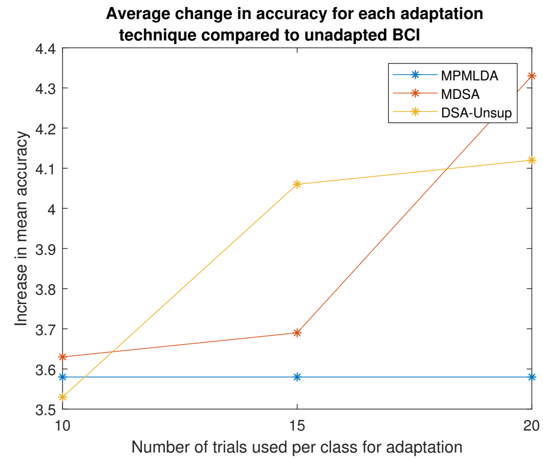

Fig. 1. Average improvement in accuracy for DSA-US, MPMLDA and the proposed MDSA compared to no adaptation, using different number of trials per class for adaption. When 20 trials per class are used for adaptation the proposed MDSA outperforms the other techniques.

In the case of 20 trials per class being provided for adaptation, MDSA outperformed DSA-US, possibly due to each class having enough trials to estimate an accurate distribution of the data. This does highlight one of the faults of MDSA, as it requires more trials than its unsupervised counterpart to produce its best level of accuracy. However, it is also capable of producing higher levels of accuracy when the trials are available. This could be due to the MDSA creating representative data distributions for each class for the adaptation while still being able to recognize changes in the EEG signals relatively quickly unlike DSA-US. The DSA-US algorithm uses the 80 previous trials for each calculation of the linear transform, so if the user's EEG signals start to change, due to fatigue or changes in their mental state, it takes a while to be seen by DSA-US as the change is diluted by the 79 other trials. This problem is not very pronounced in MDSA due to the trials being split by class so the change is only diluted by 19 other trials per class.

B. User comparison

The plots presented in Fig. 2 compare the classification accuracies of the proposed MDSA, DSA-US and MPMLD. As shown in Fig.2, the proposed MDSA and DSA-US both outperformed MPMLDA when implemented with users who were able to achieve levels of accuracy above 70%. Users 1, 3, 7 and 8 all achieved better accuracies when DSA-US or MDSA were applied compared to MPMLDA. The only user who achieved accuracy higher than 70% and performed best with MPMLDA was user 9. Conversely, the users with low levels of accuracy found MPMLDA most effective at improving their accuracy in all cases except for user 5. Fig.2 also displays that the MDSA outperformed the DSA-US in 66% of users, excluding users with less than 1% difference between the two algorithms. Suggesting that although all the adaptation algorithms are capable of improving the average accuracy of the BCI, MDSA and MPMLDA outperformed

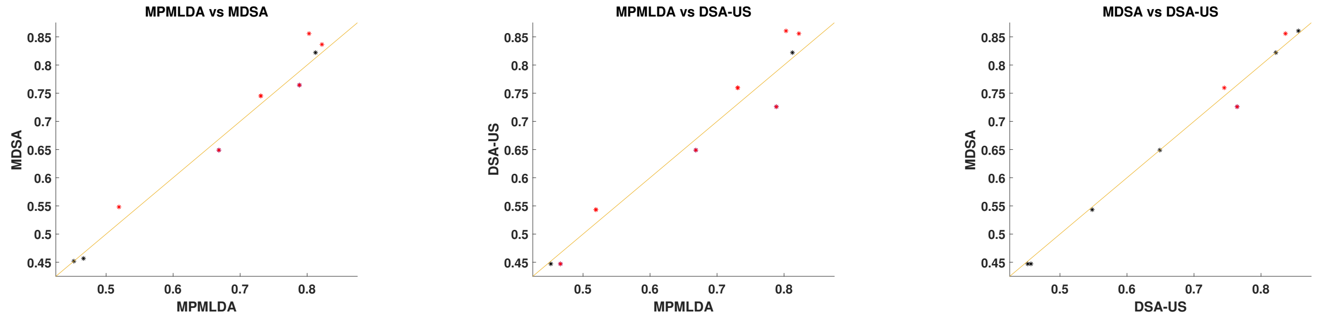

Fig. 2. Scatter plots comparing the classification accuracies of the different adaptation algorithms; Each subject is presented with a dot with black dots being used if the difference between the techniques being less than 1%. Having the dot on the left hand side of the line means the adaptation technique on the y axis works better for the corresponding subject. 20 trials per class were used for estimating transformation matrix for DSA-US and MDSA.

DSA-US when used with users encountering BCI deficiency.

In Table 1 the users were grouped into two groups based on their accuracy without adaptation; i.e. either above 70% accuracy or below 70%. As shown in Table 1, on average MPMLDA and MDSA perform similarly for subjects with accuracies less than 70%, while DSA-US is shown to be less effective for this group. This suggests that MDSA is as useful as MPMLDA when implemented to reduce BCI deficiency. High accuracy users see little improvement from MPMLDA while DSA-US and MDSA both perform equally well. Suggesting that MDSA has a good overall increase in accuracy for users who obtain high levels of accuracy and those encountering BCI deficiency.

TABLE I

AVERAGE ACCURACY IMPROVEMENT FOR EACH TECHNIQUE FOR USERS ENCOUNTERING BCI DEFICIENCY (BELOW 70% WITHOUT ADAPTATION) AND USERS WITH GOOD ACCURACY (ABOVE 70% WITHOUT ADAPTATION)

<70%			>70%		
MPMLDA	DSA-US	MDSA	MPMLDA	DSA-US	MDSA
4.44%	3.96%	4.44%	2.89%	4.23%	4.23%

The average changes were calculated by comparing accuracy of each technique with the accuracy without adaptation for each user.

IV. CONCLUSION

The proposed MDSA has shown to be effective in improving accuracy of multi-class BCIs, capable of outperforming MPMLD and DSA-US when enough data is provided, however the improvement is not statistically significant. Despite the proposed MDSA not showing significant improvement over the other algorithms it did improve accuracy for both users who were proficient with BCIs and users encountering BCI deficiency unlike the DSA-US or MPMLDA. This range of effectiveness suggests that although the overall improvement of accuracy was not statistically significant the MDSA adaptation could be applicable to a wide range of users. It is also important to note that the adaptation occurs in the data space making it independent from the feature selection and classification used by BCI. Thus, combining this adaptation method with a separate technique which focuses on adapting either features or classification could be explored to further improve the BCIs accuracy.

REFERENCES

[1] T. W. Berger, J. K. Chapin, G. A. Gerhardt, D. J. McFarland, J. C. Principe, W. V. Soussou, D. M. Taylor, and P. A. Tresco, *Brain-Computer Interfaces*. Dordrecht: Springer Netherlands, 2008.

[2] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control.," *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–91, 2002.

[3] C. Vidaurre, C. Sannelli, K.-R. Müller, and B. Blankertz, "Co-adaptive calibration to improve BCI efficiency," *Journal of Neural Engineering*, vol. 8, no. 2, p. 025009, 2011.

[4] M. Krauledat, M. Schröder, B. Blankertz, and K.-R. Müller, "Reducing calibration time for brain-computer interfaces: A clustering approach," *Adv. Neural Information Processing Syst.*, vol. 19, p. 753, 2007.

[5] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers in Neuroscience*, vol. 6, no. MAR, pp. 1–9, 2012.

[6] C. Vidaurre, M. Kawanabe, P. Von Bünau, B. Blankertz, and K. R. Müller, "Toward unsupervised adaptation of LDA for brain-computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3 PART 1, pp. 587–597, 2011.

[7] L. F. Nicolas-Alonso, R. Corralejo, J. Gomez-Pilar, D. Álvarez, and R. Hornero, "Adaptive semi-supervised classification to reduce intersession non-stationarity in multiclass motor imagery-based brain-computer interfaces," *Neurocomputing*, vol. 159, no. 1, pp. 186–196, 2015.

[8] W. K. H. Ojrulwkp, Q. W. Ri, F. Ehfrph, D. Y. Wrro, I. R. U. Sdudo, H. G. Sdwlhqwv, S. Pryhphqwv, R. I. Wkh, O. Ru, U. Kdqg, G. Ri, W. K. H. Lqirupdwlrq, and W. Udwh, "One-versus- the-rest (OVR) algorithm: An extension of common spatial patterns (CSP) algorithm to multi-class case," *Engineering In Medicine And Biology*, vol. 3, pp. 2387–2390, 2005.

[9] A. Llera, V. Gómez, and H. J. Kappen, "Adaptive Multiclass Classification for Brain Computer Interfaces," *Neural Computation Massachusetts Institute of Technology*, vol. 26, no. 6, pp. 1108–1127, 2014.

[10] P. Xu, P. Yang, X. Lei, and D. Yao, "An enhanced probabilistic LDA for multi-class brain computer interface," *PLoS ONE*, vol. 6, no. 1, 2011.

[11] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "EEG Data Space Adaptation to Reduce Intersession Nonstationarity in Brain-Computer Interface," *Neural Computation Massachusetts Institute of Technology*, vol. 25, pp. 2146–2171, 2013.

[12] C. Brunner, R. Leeb, G. R. Muller-Putz, A. Schlogl, and G. Pfurtscheller, "BCI Competition 2008 - Graz data set A," tech. rep., Institute for Knowledge Discovery, Graz University of Technology, Austria, 2008.

[13] S. Kullback, *Information theory and statistics*. Courier Corporation, 1997.

[14] M. Tangermann, K. R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. R. Müller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, and B. Blankertz, "Review of the BCI competition IV," 2012.