Proceedings Paper:
Azab, A.M.M.E., Mihaylova, L. and Arvaneh, M. (2019) Weighted multi-task learning in classification domain for improving brain-computer interface. In: 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 07-10 Oct 2018, Miyazaki, Japan. IEEE , pp. 1093-1098. ISBN 978-1-5386-6650-0

https://doi.org/10.1109/SMC.2018.00193

Weighted Multi-Task Learning in Classification Domain For Improving Brain-computer Interface

Ahmed.M.Azab*, Lyudmila S Mihaylova, and Mahnaz Arvaneh
Automatic Control and Systems Engineering Department
University of Sheffield
United kingdom
Email: *Ammazab1@sheffield.ac.uk

***Abstract*—One of the major limitations of brain computer interface (BCI) is its long calibration time. Due to between sessions/subjects nonstationarity, typically a big amount of training data needs to be collected at the beginning of each session in order to tune the parameters of the system for the target user. In this paper, a number of novel weighted multi-task transfer learning algorithms are proposed in the classification domain to reduce the calibration time without sacrificing the classification accuracy of the BCI system. The proposed algorithms use data from other subjects and combine them to estimate the classifier parameters for the target subject. This combination is done based on how similar the data from each subject is to the few trials available from the target subject. The proposed algorithms are evaluated using dataset 2a from BCI competition IV. According to the results, the proposed algorithms lead to reduce the calibration time by 75% and enhance the average classification accuracy at the same time.**

I. Introduction

A major challenge in brain-computer interface (BCI) is that everyone has unique brain signals [1]. Using machine learning techniques, BCI has to learn the user's brain signals, but this training takes time. Calibration time is the time that a BCI system needs to adapt its parameters to the user's signals in order to accurately classify their thoughts. Generally, this calibration session could take up to 20 - 30 minutes for each new session, which is an exhausting and tiring amount of time that the patient has to undergo before the system is fully functional [2].

Among different brain signals electroencephalograph (EEG) is widely used in BCI. The main advantage of an EEG-based BCI system is providing a non-invasive direct communication between a person's brain and an electronic device without the need for any muscle controls [3]. The reasons for having a long calibration session in EEG based BCI can be as follows: the first reason is the high dimensionality of EEG signals which are very noisy as well. In order to predict correct brain states, features need to be extracted from the training EEG data to calibrate the classifier. The problem here is when there are only few trials available for training, it is hard to estimate probability distributions for high dimensional noisy EEG signals where outliers will have a great negative effects. Second, EEG is highly non-stationary. A lot of factors lead to this non-stationarity such as: the variations of users' mental and psychological states, miss concentration and fatigue; also it may be affected by various measurements circumstances, e.g. changes in the positions of the electrodes when wearing the cap on a new session, and changes in the impedance of the electrodes due to sweating. So, the classifier trained on the features extracted from data of the previous sessions usually performs poorly on a new session data. In order to mitigate the mentioned problem, recent studies try to reduce the calibration time based on different methods while keeping accuracy in an acceptable range [2], [4]–[6].

One promising approach to reduce the calibration time can be transfer learning, where data from other users or sessions are mined and used to compensate the lack of labeled data from the current target subject [7]. Transfer learning aims at learning characteristics that are consistent across sessions and subjects and at the same time adjusting those characteristics to the existing few trails of the target subject. Transfer learning has been successfully applied in different machine learning applications such as: text, image, and human activity classification. In BCI, there are some studies that applied transfer learning-based approaches on raw EEG [8], feature extraction [6], [9], [10] and classification domains [4], [11] and represent some improvements in reduction of calibration time.

Recently, a multi-task learning-based algorithm on the classification domain was proposed to reduce calibration time in BCI for a new subject [11], [12]. Multitask learning is a subfield of transfer learning where multiple classification tasks are learned jointly. In [11], [12], the classification parameters of multiple subjects are learned jointly such that the average errors as well as dissimilarities between the parameters of the different classifiers get minimized. However, the proposed algorithm did not consider the similarity/dissimilarities between the data from the new subjects and the existing data from other subjects during the learning process. To address this problem and improve the BCI classifier trained for a new subject, this paper proposes a novel weighed multi-task learning algorithm, where previously recorded data are mined, processed and reused in a way that higher weights are given to the data that are more similar to the new data and less weights to data that are less similar. Two versions of weighted multitask learning are proposed, namely supervised and unsupervised. The proposed algorithms are evaluated using BCI Competition IV dataset 2a which was recorded from 9 subjects during a motor imagery paradigm. The results show that our proposed algorithms outperform the baseline approaches not only by reducing the calibration time but also by enhancing the classification accuracy for some subjects.

The rest of this paper is structured as follows. Section II introduces the baseline approaches used throughout this work, then the proposed weighted multi-task model is presented. After describing the data set used to evaluate the models in Section III, Section IV covers the results and discussions. Finally, Section IV concludes this work with a short summary and future work.

II. Methodology

A. Baseline Approaches

Two main baseline training approaches will be explained in this subsection. The first approach is the commonly used subject-specific BCI training model where the support vector machines (SVM) classifier is trained independent from other subjects using features extracted from the common spatial patterns (CSP) algorithm for the target subject. The second baseline approach is the standard multi-task learning-based classification algorithm. This approach has two models, the first one is the linear regression-based multi-task linear proposed in [11] and the second one is the logistic regression-based multi-task proposed in [13].

1) Subject Specific Classification (Ss) : In this approach, subject-specific training trials with known labels are used to train an SVM classifier based on CSP features. The classical motor imagery-based BCI subject-specific model, used in this paper, consists of the following parts: bad trials removal, band-pass filtering, common spatial filtering, extraction of log band power features and SVM classifier. These parts are described as follows: First, a threshold test is applied to remove bad trials due to blinks or any unintended motion, then a band pass filter within the band 8 to 35 Hz is used on EEG data to remove brain activities that are out of the range known for motor imagery. Next, CSP, the commonly used spatial filtering algorithm in EEG, is applied for spatial filtering [14], [15]. The importance of spatial filtering arises due to the poor spatial resolution of EEG measurements. CSP linearly transforms the data from the original EEG-channels into new channels to better differentiate between two conditions by maximizing the variance of one condition while minimizing it for the other condition [16]. Thereafter, normalized log band power of CSP filtered EEG signals are extracted as features. Finally, the extracted features are used to train an SVM classifier. This trained classifier is used to predict the labels of the test trials.

2) Multi-Task Learning-based Classification Algorithm-Linear Model (MLLin): Alamgir et al. proposed a framework for multi-task learning in BCI [11]. In this framework, each BCI subject/session was defined as one task. A parametric probabilistic approach that uses shared priors was employed to calculate classification parameters of a new subject by defining the relation between this subject's parameters and shared priors from the available subjects/sessions [11], [12].

This algorithm works as follows: $s \in \{1, ..., S\}$ is the multiple subjects or recording sessions. For each subject/session, the n_s EEG trials are presented as $d_s = (x_s^i, y_s^i)_{i=1}^{n_s}$, where x_i denotes the feature vector extracted from the i^{th} trial of subject s, and y_s^i presents the class label of the i^{th} trial. Thus, $X = \{x^1, ..., x^{n_s}\}$ is the feature matrix for each subject/session with labels presented as $y_s^i \epsilon \{-1, 1\}$.

By assuming the classification model as a linear model with a noise term η which is distributed as $\sim \mathcal{N}(0, \sigma^2)$, the label of any trial can be modelled as

$$y_s^i = w_s^T x_s^i + \eta, \tag{1}$$

where the classification parameters w_s refers to the individual features weights being used to predict the class labels of the trials belonging to the subject/session s. Thus, when a new test trial, x_s^{i+1}, arrives for the subject/session s, the class label can be predicted by

$$y_s^{i+1} = sign(w_s^T x_s^{i+1}). \tag{2}$$

Typically, when there is no prior information available about the distribution of the model's parameters, using the available labelled trials in the data set, the objective is to determine the best w_s that minimizes the prediction error in the data set d_s. The loss function for calculating w_s can be defined using negative log-likelihood as follows:

$$L_1(w_s) = \min_{w_s} \; [1/\sigma^2 \sum_{i=1}^{n_s} (y_s^i - w_s^T x_s^i)^2] \tag{3}$$

When prior information about w_s is available and assumed to be Gaussian distributed with $\mathcal{N}(\mu, \Sigma)$, a regularization term R can be added to the loss function leading to reduce the complexity of the system and hence to prevent over-fitting. Thus, R is defined as:

$$R(w_s; \mu, \Sigma) = (1/2)([(w_s - \mu)^T \Sigma^{-1} (w_s - \mu)] + log|\Sigma|); \tag{4}$$

From this point of view the authors in [12] proposed that for a BCI problem, each subject/session is treated as one task, where the shared structure, μ and Σ can be presented respectively by the mean vector and covariance matrix of W where $W = \{w_1, ..., w_S\}$. This model calculates these shared parameters from all the tasks jointly in a way that the w_s calculated for different subjects reduce the total classification error and also are close together, and this can be achieved by solving the following optimization problem:

$$L_2(W) = \min_{W} \; [1/\sigma^2 \sum_{s} ||(X_s w_s - y_s)||^2 + \sum_{s} R]. \tag{5}$$

Finally, solving this optimization problem with respect to W and holding μ and Σ fixed yields to the following equation:

$$w_s = ((1/\sigma^2)\Sigma X_s^T X_s + I)((1/\sigma^2)\Sigma X_s^T y_s + \mu) \tag{6}$$

For fixed W, solving the optimization problem yields to identify the update equations of μ and Σ as the following equations. Thus, optimum w_s can be calculated in an iterative manner by iteratively updating w_s and (μ^* and Σ^*) until convergence. Finally, σ^2 is calculated using cross validation.

$$\mu^* = (1/S) \sum_{s} w_s \tag{7}$$

$$\Sigma^* = \frac{\sum_s (w_s-\mu)(w_s-\mu)^T}{Tr(\sum_s (w_s-\mu)(w_s-\mu)^T)} + \epsilon I \quad (8)$$

3) Multi-Task Learning-based Classification Algorithm- Logistic Model (MLLog): The authors of [13] modified the previously presented MLLin algorithm by using logistic regression instead of linear regression. Assumptions on the distribution of the dependent variables in logistic regression model could be more suitable for a binary classification problem than those in linear regression.

The MLLog algorithm aims at minimizing the following optimization problem:

$$L_3(W) = \min_W - \sum_s \sum_{i=1}^{n_s} H(w_s, y_i, x_i) + \sum_s R, \quad (9)$$

where H is the point wise cross-error function, and R is the regularization term as defined in (4). By calculating the optimum w_s in (9), the prediction of the labels of a given trial is then calculated as:

$$P(y_s^i|x_s^i) = \frac{1}{1+exp(w_s^T x_s^i)}. \quad (10)$$

Similar to MLLin, L_3 should be minimized with respect to W in order to obtain the parameters of the classifiers across subject. However, unlike the MLLin algorithm, there is no closed form solution for w_s in this optimization problem. However, gradient based optimization procedures [17] could be applied to obtain the optimal w_s given the shared parameters (μ, σ). Following the same steps that were presented in the MLLin algorithm, the shared parameters were calculated using standard Gaussian sample statistics from the optimal weights w_s as in (7,8) respectively. Iterative optimization should be then applied to update w_s and μ and Σ iteratively until convergence.

B. Proposed Weighted multi-task algorithm (WML)

The MLLin and MLLog algorithms treat all the subjects similarly so that the similarities/dissimilarities between the new subject and previous subjects are not considered in the learning process. The proposed WML algorithm addresses this limitation by giving each subject a different weight based on how the features of this subject/session are close to the features of the new subject. Thus, instead of updating shared parameters by giving the same weights to all subjects/sessions, they are weighted by taking into account similarities/dissimilarities of each subject with the new subject.

Fig. 1 presents how the classification parameters of the new subject are calculated in the proposed WML algorithm. As shown in Fig. 1, the proposed WML algorithm consists of two parts. In the first part, the best $W = \{w_1, ...w_s\}$ for the previous subjects are calculated in away that the total classification error is reduced for these subjects and at the same time their classification parameters are close to their weighted average which is calculated by assigning weights to the subjects based on their similarities to the new subject. In the second part, weighted shared priors (μ_w, Σ_w) obtained from the previous part are used with the new subject few trials to obtain this new subject classifier parameters. Optimum w_{new} is calculated in an iterative manner aiming to reduce the classification accuracy error for the new subject while the defined regularization makes it close to the weighted shared priors.

Fig. 1. Weighted multi-task algorithm

Algorithm 1: Proposed weighted multi-task algorithm

part 1
Input : $d = \{d_1, ..., d_S\}$, σ^2, KL weights(α_s)
Output: μ_w, Σ_w
Set $[\mu, \Sigma] = [0, I]$
Repeat
update $W = \{w_1, ...w_s\}$
update μ using weights (15)
update Σ using weights (16:18)
Until convergence
return μ_w, Σ_w weighted shared priors
part 2
Input : d_{new}, σ^2_{new}, μ_w, Σ_w
Output: w_{new}
Set $[\mu, \Sigma] = [\mu_w, \Sigma_w]$
Repeat
calculate w_{new}
Until convergence
return w_{new}

There are two main differences between the proposed algorithms and the baseline multitask algorithms. Firstly, three different methods for covariance matrix calculation are examined, and a comparison between these methods is held to choose the best method based on the best classification accuracy results. The first method to calculate a covariance matrix is referred to as cov1(size) and calculated as below:

$$\Sigma = \frac{\sum_s (w_s-\mu)(w_s-\mu)^T}{size((w_s-\mu)(w_s-\mu)^T)} + \epsilon I. \quad (11)$$

The second method, called cov2 (trace), is calculated as:

$$\Sigma = \frac{\sum_s (w_s-\mu)(w_s-\mu)^T}{Tr((w_s-\mu)(w_s-\mu)^T)} + \epsilon I, \quad (12)$$

and the third method is called cov3 (diagonal) and it's equation is as follows:

$$\Sigma = \frac{diag \sum_s (w_s-\mu)(w_s-\mu)^T}{Trace(\sum_s (w_s-\mu)(w_s-\mu)^T)} + \epsilon I. \quad (13)$$

The second main difference is the weight that is defined for each subject to represent the similarity between this subject and the new subject. Kullback-Leibler (KL) divergence is used to calculate these weights [18]. The KL divergence between two gaussian distributions, presented as $N_0(\mu, \Sigma)$ and $N_1(\bar{\mu}, \bar{\Sigma})$, has a closed-form expression as follows:

$$\begin{aligned} KL[N_0||N_1] = (1/2)[(\bar{\mu}-\mu)^T \bar{\Sigma}^{-1}(\bar{\mu}-\mu) \\ + tr(\bar{\Sigma}^{-1}\Sigma) - \ln\left(\frac{det(\Sigma)}{det(\bar{\Sigma})}\right) - K], \end{aligned} \quad (14)$$

where det and k denote the determinant function and the dimensionality of the data, respectively. Therefore, in the proposed weighted algorithm, (14) is used to calculate the distance between the feature distributions of each subject and the new subject. It is noted that we use CSP features in this study. CSP features are normalized log variance of CSP-filtered EEG data, thus the assumption of having Gaussian distribution can be valid.

If labelled trials from the new subject are available, supervised KL distance is computed for each class and the total distance is the sum of the distances for the two classes. When there are no labelled trials available for the new subject, the KL distance between the two subjects is calculated without considering the class labels and it is called unsupervised KL. Considering these two weighted distances, the proposed algorithms can be supervised weighted multi-task (SMLLin, and SMLLog) and unsupervised weighted multi-task (UMLLin, and UMLLog), where Lin and Log are referring to the applied regression method. The weight between the subject s and the new subject, α_s, is calculated using the following equation:

$$\alpha_s = \frac{(1/KL[d_{new}, d_s])^4}{\sum_{i=1}^{S}(1/KL[d_{new}, d_i])^4}. \tag{15}$$

Based on the obtained weight for each subject, α_s, the new equation to update the weighted μ is:

$$\mu_w = (1/S)\sum_s \alpha_s w_s. \tag{16}$$

Similarly, the weighted Σ is calculated using the following modified equations for cov1 (size), cov2 (trace), and cov3 (diagonal) respectively:

$$\Sigma_w = \frac{\sum_s(\alpha_s w_s - \mu_w)(\alpha_s w_s - \mu_w)^T}{size((\alpha_s w_s - \mu_w)(\alpha_s w_s - \mu_w)^T)} + \epsilon I \tag{17}$$

$$\Sigma_w = \frac{\sum_s(\alpha_s w_s - \mu_w)(\alpha_s w_s - \mu_w)^T}{Tr((\alpha_s w_s - \mu_w)(\alpha_s w_s - \mu_w)^T)} + \epsilon I \tag{18}$$

$$\Sigma_w = \frac{diag\sum_s(\alpha_s w_s - \mu_w)(\alpha_s w_s - \mu_w)^T}{Trace((\alpha_s w_s - \mu_w)(\alpha_s w_s - \mu_w)^T)} + \epsilon I \tag{19}$$

III. Experiments

In order to validate the proposed algorithms and compare them with the baseline algorithms, all the algorithms are applied to data set 2a BCI Competition IV 2008 [19]. This data set consists of EEG data from 9 subjects performing 4 classes of motor imagery task. In this paper only data from right and left hand motor imagery are used. Two sessions on different days were recorded for each subject. Each session is comprised of 6 runs, each run consists of 12 trials for each class.

EEG signal was recorded using 22 electrodes. EEG signals were sampled at 250 Hz, and were bandpass-filtered between 0.5 Hz and 100 Hz. Moreover, a 50 Hz notch filter was applied to remove power line noise. The proposed algorithms and the baseline algorithms are applied only on the trials recorded on

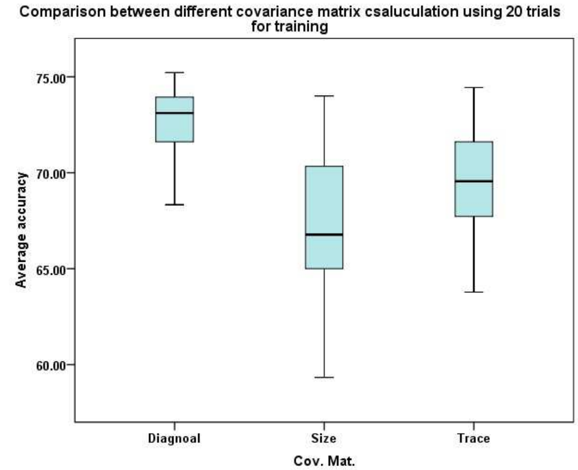

Fig. 2. Comparison between different covariance matrix calculation methods when 20 trials from the new subjects are used for training. The average accuracy calculated include results obtained by MLLin, SMLLin, UMLLin, MLLog, SMLLog, and UMLLog.

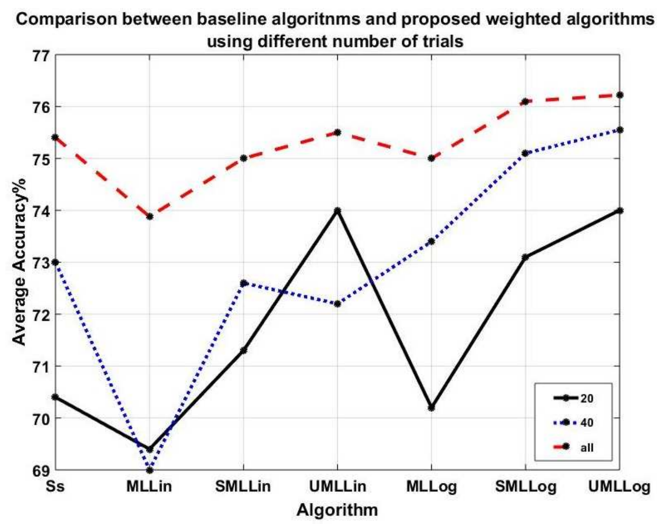

Fig. 3. Comparison between the proposed algorithms (SMLLin, UMLLin, SMLLog, and UMLLog) and the baseline algorithms (Ss, MLLin, and MLLog) using different number of trials for training (20, 40, and all trials) from new subject based on the average accuracy calculated over the nine subjects for each algorithm. UMLLog is the best algorithm when using any number of trials.

the second day by dividing it to two sessions one for training (consists of the first 42 trials recorded per class) and one for testing (consists of the last 30 trials recorded per class). This was done to establish a practical case that new subject data is coming from the same session. For the new subject, different training sizes were examined (i.e. 10, 20 and 42 trials per class). It is note that in each multitask learning algorithm, the train data of each new subject and the other 8 other subjects were used for calculating classification parameters.

IV. Results And Discussion

As mentioned before, in this section the multitask learning algorithms are applied based on three different covariance

TABLE I
CLASSIFICATION ACCURACIES CALCULATED USING THE BASELINE ALGORITHMS (SS, MLLIN, AND MLLOG) AND THE PROPOSED ALGORITHMS (**SMLLIN, UMLLIN, SMLLOG, AND UMLLOG**) FOR EACH INDIVIDUAL SUBJECT WHEN THERE ARE 40 TRIALS AVAILABLE FOR TRAINING FROM THE NEW SUBJECT, SHOWING THAT LOGISTIC ALGORITHMS OUTPERFORM LINEAR ALGORITHMS

Algorithm	sub1	sub2	sub3	sub4	sub5	sub6	sub7	sub8	sub9	Average
Ss	85	53	98	66	55	56	73	86	86	73
MLLin	85	52	97	57	52	55	67	97	60	69
SMLLin	72	58	98	63	55	53	70	98	78	**72.6**
UMLLin	72	57	98	63	55	53	70	95	87	**72.2**
MLLog	90	48	97	67	52	52	75	97	83	73.4
SMLLog	90	50	98	63	58	55	77	98	87	**75.1**
UMLLog	95	50	97	63	58	55	78	97	87	**75.6**

matrix calculation methods and two regression approaches (i.e. Linear and Logistic). All algorithms are evaluated using different number of training trials from new subjects (i.e 20, 40, all 84 trials from both classes).

To identify the most effective method of calculating covarience matrices, first a comparison between the three different covariance matrix calculation methods was held across different number of training trials for new subjects, regression approaches and all the applied multitask learning algorithms. Subsequently, a 3 (Number of trials)×6 (Algorithms)×3 (covariance calculation methods) repeated measure ANOVA test was performed on the results followed by post-hoc analysis.

Fig. 2 compares the classification results obtained by the different methods of calculating covariance matrices using 20 trials from the new subjects. These results include the classification accuracies of all the different multitask learning algorithms in both linear and logistic approaches. According to the average accuracies shown in the Fig. 2, cov3(diagonal) yielded higher classification accuracies than cov1(size) and cov2(trace). Indeed, conducting a repeated ANOVA test revealed that using different covariance matrix calculation methods had a main effect on the classification accuracy results with ($p = 0.009$). Based on the post-hoc analysis cov3(diagonal) significantly outperformed cov1(size) and cov2(trace) with the p values equal to 0.03 and 0.025 respectively. Thus, for the rest of thi paper, all the calculations and comparisons of multitask algorithms will be done using only cov3(diagonal).

Another comparison between the linear regression and the logistic regression approaches was conducted. As shown in Table I, on average the logistic approach outperformed the linear approach in all the considered multitask learning algorithms when 40 trials used from the new subjects for training. Although not presented in the table, the results of using 20 or all the trials from new subjects also showed that the logistic regression approach worked better than the linear one in the majority of the algorithms.

Finally, comprehensive comparisons were conducted based on the classification results of the 7 algorithms (i.e. Ss, MLLin, MLLog, proposed SMLLin, proposed SMLLog, proposed UMLLin, and proposed UMLLog), followed by a 3 (Number of trials)×7 (Algorithms) repeated measure ANOVA test.

Fig. 3 shows that all the proposed weighted multitask learning algorithms outperformed the subject specific algorithm (Ss) when there are only 20 trials available for training the

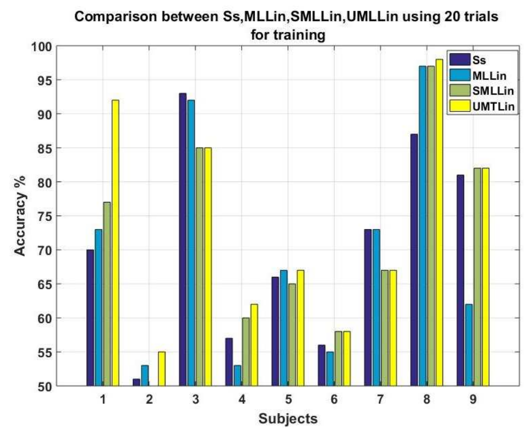

Fig. 4. Comparison between the classification accuracies calculated using the proposed weighted linear multi-task learning algorithms (SMLLin, and UMLLin) and the baseline algorithms (Ss, and MLLin) for all subjects individually when 20 trials are available for training from the new subjects. As can be seen in addition to the calibration time reduction, 7 subjects out of 9 gained an increase in the accuracy when the proposed algorithms are used.

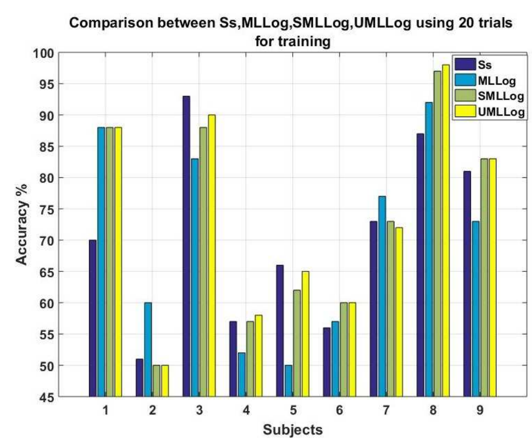

Fig. 5. Comparison between the classification accuracy calculated using the proposed weighted logistic multi-task algorithms (SMLLog, and UMLLog) and the baseline algorithms (Ss, and MLLog) for all subjects individually when 20 trials are available for training from the new subjects. In addition to the calibration time reduction 5 subjects gain an increase in the accuracy when the proposed algorithms are used.

new subjects. When the number of the training trials from the new subject increased to 40 and all, still the majority of the proposed weighted multitask learning algorithms out performed Ss. Besides the proposed algorithms outperformed the baseline linear and logistic multi-task algorithms when using 20, 40, and all trials from the new subjects for training.

Based on the statistical tests, MLLin and MLLog were neither significantly outperformed the state of art Ss algorithm nor any of the proposed algorithms. Importantly, the classification accuracy of the proposed UMLLog algorithm tended to be significantly better than the Ss algorithm results. Moreover, the proposed UMLLog algorithm significantly outperformed the baseline MLog algorithm with the p value of 0.045, whereas SMLLog tended to be significantly better than MLLog with the p value of 0.078. Interestingly, when using diagonal matrix calculation method with the baseline logistic multi-task algorithm, the modified logistic algorithm was significantly better than MLLog with $P = 0.021$. Moreover, statistical tests showed that using different number of trials did not have a main effect on classification results. This finding strengthens the outcome of this work which is reducing the calibration time without altering the overall accuracy of the system.

Fig. 4 and Fig. 5 show the classification results calculated for each subject using the proposed and baseline algorithms for linear and logistic approaches respectively. The results were obtained when when 20 trials were available for training from the new subject. As can be seen, besides reducing the calibration time, the proposed algorithms outperformed the baseline algorithms for 7 subjects out of 9 in linear regression case and for 5 subjects out of 9 in the logistic regression.

In summary, our results suggest that the novel proposed unsupervised weighted logistic multi-task learning algorithm (UMLLog) outperformed all the other algorithms. The proposed UMLLog not only reduced the required calibration time but also enhanced the average classification accuracy.

V. Conclusion

The aim of this work was to develop novel algorithms based on transfer learning to reduce the calibration time for BCI-based systems and at the same time to enhance the overall accuracy of the system. Previous approaches on transfer learning based on multi-task learning in BCI have ignored the similarity/dissimilarities between the data from the new subjects and the existing data from other subjects during the learning process. In this paper, we presented novel weighted multi-task learning algorithms to address this problem. The main finding of this paper suggests that applying the proposed transfer learning algorithms in classification domain leads to reduce the calibration time by 75% and enhance the average accuracy of the BCI-based systems.

The proposed algorithms in classification domain yielded remarkable increase in the classification accuracy of subjects that initially performed BCI with a medium accuracy. However the observed improvement for the subjects with initially poor BCI performance was not pronounced. It seems changing the parameters of classifiers for these subjects is not effective, since their feature spaces for different classes are not separable. These findings suggest that to increase accuracy of subjects that initially perform poor BCI, transfer learning approaches should be applied in a different domain before the classification domain.

References

[1] J. d. R. Millán, R. Rupp, G. R. Müller-Putz, R. Murray-Smith, C. Giugliemma, M. Tangermann, C. Vidaurre, F. Cincotti, A. Kübler, R. Leeb *et al.*, "Combining brain–computer interfaces and assistive technologies: state-of-the-art and challenges," *Frontiers in neuroscience*, vol. 4, 2010.

[2] N. Proesmans, "Reducing calibration time in Motor Imagery Brain-Computer Interfaces using Machine Learning," 2015. [Online]. Available: http://lib.ugent.be/fulltxt/RUG01/002/300/488/RUG01-002300488_2016_0001_AC.pdf

[3] E. A. Curran and M. J. Stokes, "Learning to control brain activity: a review of the production and control of eeg components for driving brain–computer interface (bci) systems," *Brain and cognition*, vol. 51, no. 3, pp. 326–336, 2003.

[4] W. Tu and S. Sun, "A subject transfer framework for EEG classification," *Neurocomputing*, vol. 82, pp. 109–116, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.neucom.2011.10.024

[5] S. Fazli, F. Popescu, M. Danóczy, B. Blankertz, K. R. Müller, and C. Grozea, "Subject-independent mental state classification in single trials," *Neural Networks*, vol. 22, no. 9, pp. 1305–1312, 2009.

[6] F. Lotte and C. Guan, "Learning from other subjects helps reducing brain-computer interface calibration time," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 614–617, 2010.

[7] P. Wang, J. Lu, B. Zhang, and Z. Tang, "A review on transfer learning for brain-computer interface classification," *5th International Conference on Information Science and Technology, ICIST 2015*, pp. 315–322, 2015.

[8] M. Arvaneh, I. Robertson, and T. E. Ward, "Subject-to-subject adaptation to reduce calibration time in motor imagery-based brain-computer interface," in *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug 2014, pp. 6501–6504.

[9] H. Kang, Y. Nam, and S. Choi, "Composite common spatial pattern for subject-to-subject transfer," *IEEE Signal Processing Letters*, vol. 16, no. 8, pp. 683–686, 2009.

[10] W. Samek, F. C. Meinecke, and K.-R. Müller, "Transferring subspaces between subjects in brain–computer interfacing," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 8, pp. 2289–2298, 2013.

[11] M. Alamgir, M. Grosse-Wentrup, and Y. Altun, "Multitask learning for brain-computer interfaces," *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS-10)*, vol. 9, pp. 17–24, 2010.

[12] V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf, and M. Grosse-Wentrup, "Transfer learning in brain-computer interfaces," *IEEE Computational Intelligence Magazine*, vol. 11, no. 1, pp. 20–31, 2016.

[13] K.-H. Fiebig, V. Jayaram, J. Peters, and M. Grosse-Wentrup, "Multi-task logistic regression in brain-computer interfaces," in *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2016, pp. 002 307–002 312.

[14] R. Tomioka and S. Lemm, "Filters for Robust EEG," *IEEE Signal Processing Magazine*, no. January 2008, pp. 41–56, 2008.

[15] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "Bci2000: a general-purpose brain-computer interface (bci) system," *IEEE Transactions on biomedical engineering*, vol. 51, no. 6, pp. 1034–1043, 2004.

[16] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller, "Optimizing spatial filters for robust eeg single-trial analysis," *IEEE Signal processing magazine*, vol. 25, no. 1, pp. 41–56, 2008.

[17] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, 1952, vol. 49, no. 1.

[18] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Eeg data space adaptation to reduce intersession nonstationarity in brain-computer interface," *Neural computation*, vol. 25, no. 8, pp. 2146–2171, 2013.

[19] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "Bci competition 2008–graz data set a."