VOL. 1, NO. 3, JULY 2017 0000000

Sensor Applications

Transfer learning for a multimodal hybrid EEG-fTCD brain-computer interface

Elise Dagois[1*], Aya Khalaf[1], Ervin Sejdic[1] and Murat Akcakaya[1]

[1] Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, PA, 15213, USA

Manuscript received Month xx, 2018; revised Month xx, 2018; accepted Month xx, 2018. Date of publication Month xx, 2018; date of current version Month xx, 2018.

Abstract—Humans can transfer knowledge previously acquired from a specific task to new and unknown ones. Recently, transfer learning has been extensively used in Brain-Computer Interface (BCI) research to reduce the training/calibration requirements. BCI systems have been designed to provide alternative communication or control access through computers to individuals with limited speech and physical abilities (LSPA). These systems generally require a calibration session in order to train the BCI before each usage. Such a calibration session may be burdensome for the individuals with LSPA. In this paper, we introduce a multimodal hybrid BCI based on electroencephalography (EEG) and functional transcranial Doppler ultrasound (fTCD) and present a transfer learning approach to reduce the calibration requirements. In the hybrid BCI, EEG and fTCD are used simultaneously to measure electrical brain activity and cerebral blood velocity respectively in response to motor imagery (MI) tasks. Using the data we collected from 10 healthy individuals, we perform dimensionality reduction utilizing Regularized Discriminant Analysis (RDA). Using the scores from RDA, we learn class conditional probabilistic distributions for each individual. We use these class conditional distributions to perform transfer learning across different participants. More specifically, in order to reduce the calibration requirements for each individual, we choose recorded data from other individuals to augment the training data for that specific individual. We choose the data for augmentation based on the probabilistic similarities between the class conditional distributions. For final classification we use the RDA scores after transfer learning as features input to three different classifiers: Quadratic Discriminant Analysis (QDA), Linear Discriminant Analysis (LDA), and Support Vector Machines (SVM). Using our experimental data, we show that transfer learning decreases the calibration requirements up to $\%87.5$. Also by comparing SVM, LDA and QDA, we observe that SVM provides the best classification performance.

Index Terms—Transfer Learning, Hybrid Brain Computer Interfaces, Electroencephalogram, Functional Transcranial Doppler, Distance Measures, Machine Learning.

I. INTRODUCTION

Humans present the ability to draw knowledge from their previous experiences and exploit such knowledge when learning future and unknown tasks. They are indeed able to recognize similarities between previously and newly encountered tasks, and they utilize these similarities to learn a new task. Therefore, the notion of similarity identification is very important for knowledge transfer and for completion of newly encountered tasks with high accuracy and speed [1]. Conversely, since the acquired knowledge and experience could be specific to a task, if they are poorly related to learning the new task, the human learner may not transfer the existing knowledge and experience to the new task [2]. The same principles also apply to machine learning. Traditional machine learning approaches create models for certain tasks and make predictions about these tasks without applying any knowledge transfer. Such techniques work on a unique and isolated task and are based on labeled data [3]. Transfer learning finds closely related source tasks and draw knowledge from these tasks to predict a new target task with higher performance. Such transfer learning techniques are beneficial especially when the training data for the new task are not sufficient [4]. In this paper, we propose to use transfer learning for multimodal hybrid noninvasive brain-computer interfaces (BCI).

Noninvasive BCIs are designed to provide alternative communication and control access through computers to individuals with limited speech and physical abilities (LSPA). More specifically, such BCIs aim to help individuals with LSPA by allowing them to type on a computer screen or control electronic devices such as prosthetics limbs, wheelchairs or robotic agents using brain signals [5], [6], [7], [8].

Electroencephalogram (EEG) is the most preferred modality to build such BCIs due to its cost effectiveness and portability. However, EEG has usually low signal-to-noise ratio and it presents nonstationarities due to background brain activities. Such an EEG behavior would result in decreased performance for the EEG-based BCIs. Previous studies combined EEG with other type of signals such as electromyography (EMG) [9] [10] or near-infrared spectroscopy (NIRS) [11] to design new approach of BCI called hybrid BCI. In order to overcome the performance degradation due to changing EEG behavior, we have developed a multimodal hybrid BCI that utilizes both EEG and functional transcranial Doppler (fTCD) [12]. In this hybrid BCI, EEG and fTCD are used simultaneously to measure electrical brain activity and cerebral blood velocity respectively in response to MI tasks for classification [13]. Our previous study showed that the hybrid multimodal approach improves both the speed and accuracy of the BCI compared to using EEG or fTCD only for the BCI design [13].

In this paper, we propose to use a novel transfer learning approach to decrease the training requirements for the multimodal hybrid BCI. To assess the proposed approach, we formulated 3 binary classification problems including right arm MI versus baseline, left arm MI versus baseline, right versus left arm MI. In order to apply

* Corresponding author's email: evd16@pitt.edu.

transfer learning, we extract features using Regularized Discriminant Analysis (RDA) from EEG and fTCD data for three different binary classification problems and learn class conditional distributions for these features. Class conditional distributions are then used to probabilistically identify similarities among the data collected from multiple participants and to apply transfer learning. For probabilistic similarities, we utilize various probabilistic measures including Kullback-Leibler divergence (KL divergence), Bhattacharyya, and Hellinger distances. We use the RDA scores for transfer learning and as the final features for the classification. We compare three different classifiers: quadratic discriminant analysis (QDA), linear discriminant analysis (LDA), and linear support vector machine (SVM). Using the data we collected from 10 healthy participants, we demonstrate that the transfer learning can reduce the training dataset by up to %87.5. Also we show that SVM provides the best classification performance after transfer learning.

II. METHODS

In this section, we introduce our experimental setup, pre-processing, and feature extraction methods as well as the similarity measures and the proposed transfer learning algorithm.

A. Experiment Setup

We collected EEG signals using 16 electrodes placed over frontal, central and parietal lobes at the positions Fp1, Fp2, F3, F4, Fz, Fc1, Fc2, Cz, P1, P2, C1, C2, Cp3, Cp4, P5, and P6 with the reference electrode placed over the left mastoid. The signals are amplified using a g.tec EEG system with g. USBamp. Collected data are sent from the amplifiers to a laptop via USB 2.0.

Two 2 MHz transducers of a SONARA TCD system were used to collect the fTCD data. The transducers are positioned on the left and right sides of the transtemporal window located above the zygomatic arch [14] [15]. The fTCD depth was set to the depth of the mid-point of the middle cerebral arteries (50 mm) [14]. In this study, EEG and fTCD are synchronized based on time stamps available from both modalities.

16-channel EEG and two-channels fTCD data were recorded simultaneously (under University of Pittsburgh approved IRB) from 10 healthy right-handed subjects with ages ranging from 23 to 32 years old. Each participant attended one data collection session of 25-min duration. During data acquisition, each participant is observing visual icons corresponding to right and left arm MI mental tasks as well as baseline. In particular, as shown in Figure 1, left arrow represents left arm MI and right arrow represents right arm MI while the fixation cross in the middle represents the baseline. 150 trials are randomly presented per user in total. Assuming equal probability, we approximately have 50 trials per task. During each trial of 10-s duration, a vertical red arrow selects randomly one of the 3 tasks and the participant is instructed to perform the task selected by the vertical arrow.

B. Pre-processing and features extraction

EEG and fTCD data of each trial were segmented. Features corresponding to each trial included power spectrum raw values calculated using Welch method [16]. To reduce the number of features obtained from each power spectrum, the average power over a narrow range of frequencies was computed instead of using all the power spectrum values at all frequency bins. More specifically, the average power over each consecutive 2 Hz of the EEG data was considered as one feature while for the fTCD data, since it has a much higher bandwidth (2.5 kHz compared to 40 Hz for EEG), the average power over each consecutive 50 Hz for the fTCD data was considered as one feature. A single features vector (EEG and fTCD feature vectors of each trial) were obtained for each trial by concatenating the reduced power spectrums corresponding to the 16 EEG segments and to the 2 fTCD segments respectively. Both EEG and fTCD feature vectors were normalized separately across trials using min-max normalization [17].

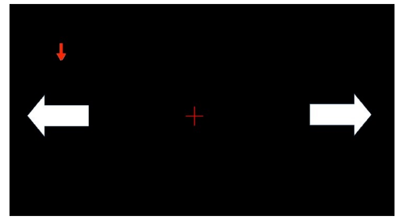

Fig. 1. Stimulus presentation for the hybrid BCI system.

C. RDA Scores and Similarity Measures

To transfer knowledge across participants, we measure the similarity between the dataset under test and the other datasets. Various distance measures are used in this study to measure distance between two probability density functions P and Q including KL divergence, Bhattacharyya, and Hellinger [18]. Instead of calculating similarity using the EEG and fTCD concatenated feature vector corresponding to each trial, we project this vector into a scalar value using RDA.

RDA is a commonly used method to classify objects in low-dimension [19]. RDA is the extension of QDA [20],

In this study, 10-fold cross-validation is used to optimize the RDA parameters λ and γ within a range 0.1 to 1 with a step of 0.1 such that the area under the receiver operating characteristic curve is maximized. This method is performed for the three binary classifications. Using the RDA scores of trials corresponding to each class, we estimate the pdf of each class using kernel density estimation (KDE). During KDE we utilize rbf kernel and the kernel bandwith is computed using Silverman's rule of thumb. The similarities between the pdf of the dataset under test and the pdfs of the other datasets are computed. Five datasets which are most similar to the dataset under test are chosen for transfer learning. In the results section, we compare the performances of different distance measures.

D. Classification

Three classification tasks were formulated to test the performance of the hybrid BCI using the suggested transfer learning algorithm including right arm MI versus baseline (RA vs BL), left arm MI versus baseline (LA vs BL), and right arm MI versus left arm MI (RA vs LA). Linear SVM, LDA, and QDA were used to perform the classification tasks and performance measures obtained from each classifier were compared.

VOL. 1, NO. 3, JULY 2017 0000000

E. Transfer Learning Algorithm

In this paper, we propose to use transfer learning to decrease the training requirements for the hybrid multimodal BCI. The pseudocode for the proposed algorithm is summarized in Algorithm 1. In this algorithm, for each participant, we identify five datasets from the rest of the participants with the most similar EEG and fTCD characteristics for transfer learning. More specifically, we utilize the KDEs of the class conditional distributions of the scores obtained from the RDA projection of EEG and fTCD features. Assuming that N number of trials are collected from each participant, each participant's data are separated into test set of size t trials and training set of size $N - t$ trials. Test and training sets are chosen to include equal number of trials for each class. Then, training dataset and corresponding KDE of RDA values are used to identify five datasets (from the rest of the participants) that are most similar in a probabilistic manner to the training set of each participant. Each participant's training set is then augmented with these most similar datasets (identified from the rest of the participants) for transfer learning. In order to analyze the effect of transfer learning for different training test sizes, t is varied to take different values. Note that for the binary classification problems that are defined above, we have $N = 100$. Accordingly t is varied from 10 to 90 resulting in minimum training size of 10 samples (5 samples for each class).

III. RESULTS

For the 3 binary classification problems, we first analyze the reduction in training size requirements provided by three distance measures used for transfer learning (TL) compared to no transfer learning (NT). In particular, we statistically compared the accuracy vectors of TL with minimum training size and NT using bigger training set sizes than TL. For TL with 10 trials, we performed one-sided Wilcoxon signed rank test [21] between the accuracy vector of TL with 10 trials and NT accuracy vectors obtained at training set sizes ranging from 20 to 90 trials. Same statistical comparison was applied for the 3 distance measures. Using Bhattacharyya distance, at training set size of 10, it was found that the performance of TL is comparable with the performance of NT with 80 trials. In other words, instead of calibrating the system with 80 trials in the NT case, same or higher performance can be achieved when using TL and calibrating the system with 10 from the BCI user. This reflects a reduction in the calibration requirements by %87.50. Using both Hellinger and KL-divergence, the performance of TL at 10 trials was statistically comparable to the performance of NT with maximum of 50 training trials which reflects a maximum possible reduction in calibration requirements by %80.00. By comparing the calibration reduction percentages of the 3 measures, we choose to present the results of Bhattacharyya distance for three binary class classifications Considering the test set size, $t = 90$ trials and training size $N - t = 10$ trials, accuracy, sensitivity and specificity of the classification (averaged across participants) for three different classification methods are presented in Tables 1, 2, and 3 for RA vs. LA, RA vs. BL and LA vs. BL classification problems, respectively. During transfer learning, we also optimize the trial length for each participant and the classification results are presented for these optimum trial lengths (between 1 and 10 seconds).

We observe from Tables 1, 2 and 3 that all classification results achieved using transfer learning (with the augmentation of the top 5 datasets identified as the most similar datasets) provide higher performance than the case without transfer learning. For classification using LDA, accuracy reaches %78.67, %76.11 and %77.11 for RA vs LA, RA vs BL and LA vs BL, respectively for transfer learning. For QDA, %77.67, %74.89 and %76.78 and for SVM %79.33, %77.11 and %77 accuracy were achieved for the same classification problems. It can be noted that SVM achieves higher accuracy compared to QDA and LDA. However, in terms of the accuracy difference between TL and NT, the highest difference is provided by QDA for which the classification is improved by %11.11, %5.56, and %6.11 for RA vs LA, RA vs BL, and LA vs BL respectively compared to %6.23, %2.11, and %2.89 for RA vs LA, RA vs BL, and LA vs BL respectively given by LDA and %6.22, %2.44 and %2.11 given by SVM.

Considering Bhattacharyya as the distance measure to identify top 5 datasets for transfer learning, Figure 2 presents the accuracy values (averaged across participants) for three different classification problems as the training set size changes from 10 to 90 trials. Transfer learning results are compared to the no transfer learning results for each binary classification problem. Results are evaluated using QDA. This figure focuses on the influence of the quantity of information available to train a classifier and effect of the training/calibration data size on the classification accuracy with and without transfer learning. Overall the accuracy values obtained by transfer learning are higher than the case without transfer. Using this figure, we observe that when the training size drops even down to 20 corresponding to $t = 80$ test data size, transfer learning provides between %5 to %7 improvement in the accuracy.

IV. CONCLUSION

In this study, we extend our previous work on hybrid EEG-fTCD BCI employing motor imagery mental tasks. In particular, we developed a transfer learning approach with the aim of reducing the calibration requirements as well as improving the system performance.

TABLE 1. Mean of accuracy among all participants and corresponding sensitivity and specificity for RA vs LA.

Performance measures	LDA		QDA		SVM	
	NT	TL	NT	TL	NT	TL
Accuracy	0.7244	0.7867	0.6656	0.7767	0.7311	0.7933
Specificity	0.8395	0.8191	0.8023	0.7723	0.8140	0.8532
Sensitivity	0.6191	0.7512	0.5404	0.7814	0.6553	0.7279

TABLE 2. Mean of accuracy among all participants and corresponding sensitivity and specificity for RA vs BL.

Performance measures	LDA		QDA		SVM	
	NT	TL	NT	TL	NT	TL
Accuracy	0.74	0.7611	0.6933	0.7489	0.7467	0.7711
Specificity	0.7725	0.7980	0.7075	0.7560	0.78	0.84
Sensitivity	0.7140	0.7150	0.6820	0.74	0.72	0.6850

TABLE 3. Mean of accuracy among all participants and corresponding sensitivity and specificity for LA vs BL.

Performance measures	LDA		QDA		SVM	
	NT	TL	NT	TL	NT	TL
Accuracy	0.7422	0.7711	0.7067	0.7678	0.7489	0.77
Specificity	0.7975	0.8060	0.65	0.7820	0.70	0.7520
Sensitivity	0.6980	0.7275	0.7520	0.75	0.7880	0.7925

0000000 VOL. 1, NO. 3, JULY 2017

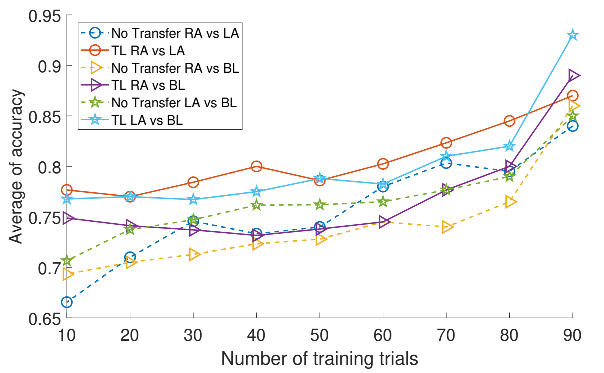

Fig. 2. Accuracy as a function of the number of trials in the training set for each binary problem computed with and without transfer.

for each binary classification problem **do**
  for each distance measure D **do**
    Accuracy=(); Specificity=(); Sensitivity=();
    for $Ts = 1:10$ // Segment length (s) **do**
      for $n = 1:10$ // participants **do**
        N_i = number of trials from class i, $i = 1, 2$;
        N=N_1+N_2;
        Estimate power spectrum features;
        Get RDA scores $s^i = \{s_1, ..., s_{N_i}\}$ for class i;
        Estimate $P_n(s^i|x_i)$ using KDE;
      end
      for $n = 1:10$ **do**
        Get distance between $P_n(s^i|x_i)$ and $P_m(s^i|x_i)$ for $m = \{1, ..., 10\} - \{n\}$;
        Select top similar 5 datasets to participant n;
        $T = [T_E^n, T_R^n,]$;
        T_E^n: testing trials for participant n;
        T_R^n: training trials for participant n;
        for $T_E^n = 10:10:90$ **do**
          Train the classifier using $[T_R^n, T_R^l]$, $l \subset m$ contains indexes corresponding to top similar 5 datasets;
          Test the model using T_E^n;
          Compute performance measures;
        end
      end
    end
  end
end

Algorithm 1: Pseudocode for the proposed transfer learning approach.

To achieve such aim, EEG and fTCD data of each trial were projected into a scalar RDA score. The scores corresponding to each class were used to learn class conditional distributions. Several distance measures were used to identify similarities between class conditional distributions among the data collected from multiple participants. Performance of the proposed approach was evaluated using LDA, QDA, and SVM classifiers with RDA scores used as input features. Experimental results show that the training set size can be reduced by up to %87.5 while achieving reasonable classification accuracy values. The current work identifies the similarities and therefore uses exploitation of the knowledge from the data. Our future work will combine exploitation and exploration to improve the generalization properties of the proposed multimodal hybrid system.

REFERENCES

[1] L. Torrey and J. Shavlik, "Transfer learning," in Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques. IGI Global, 2010, pp. 242–264.

[2] M. T. Rosenstein, Z. Marx, L. P. Kaelbling, and T. G. Dietterich, "To transfer or not to transfer," in NIPS 2005 workshop on transfer learning, vol. 898, 2005, pp. 1–4.

[3] S.-C. Chin, "Knowledge transfer: what, how, and why," 2013.

[4] S. J. Pan, Q. Yang et al., "A survey on transfer learning," IEEE Transactions on knowledge and data engineering, vol. 22, no. 10, pp. 1345–1359, 2010.

[5] T. Davis, H. Wark, D. Hutchinson, D. Warren, K. OâĂŹNeill, T. Scheinblum, G. Clark, R. Normann, and B. Greger, "Restoring motor control and sensory feedback in people with upper extremity amputations using arrays of 96 microelectrodes implanted in the median and ulnar nerves," Journal of neural engineering, vol. 13, no. 3, p. 036001, 2016.

[6] R. Zhang, Y. Li, Y. Yan, H. Zhang, S. Wu, T. Yu, and Z. Gu, "Control of a wheelchair in an indoor environment based on a brain–computer interface and automated navigation," IEEE transactions on neural systems and rehabilitation engineering, vol. 24, no. 1, pp. 128–139, 2016.

[7] J. Faller, B. Z. Allison, C. Brunner, R. Scherer, D. Schmalstieg, G. Pfurtscheller, and C. Neuper, "A feasibility study on ssvep-based interaction with motivating and immersive virtual and augmented reality," arXiv preprint arXiv:1701.03981, 2017.

[8] S. Qiu, Z. Li, W. He, L. Zhang, C. Yang, and C.-Y. Su, "Brain-machine interface and visual compressive sensing-based teleoperation control of an exoskeleton robot." IEEE Trans. Fuzzy Systems, vol. 25, no. 1, pp. 58–69, 2017.

[9] S. Fazli, S. DÃď'hne, W. Samek, F. BieÃ§mann, and K. MÃijller, "Learning from more than one data source: Data fusion techniques for sensorimotor rhythm-based brainâĂŞcomputer interfaces," Proceedings of the IEEE, vol. 103, no. 6, pp. 891–906, June 2015.

[10] R. Leeb, H. Sagha, R. Chavarriaga et al., "Multimodal fusion of muscle and brain signals for a hybrid-bci," in Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE. IEEE, 2010, pp. 4343–4346.

[11] S. Fazli, J. Mehnert, J. Steinbrink, G. Curio, A. Villringer, K.-R. Müller, and B. Blankertz, "Enhanced performance by a hybrid nirs–eeg brain computer interface," Neuroimage, vol. 59, no. 1, pp. 519–529, 2012.

[12] A. Khalaf, M. Sybeldon, E. Sejdic, and M. Akcakaya, "A brain-computer interface based on functional transcranial doppler ultrasound using wavelet transform and support vector machines," Journal of Neuroscience Methods, vol. 293, pp. 174 – 182, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165027017303515

[13] A. Khalaf, E. Sejdic, and M. Akcakaya, "Towards optimal visual presentation design for hybrid eegâĂŤftcd brainâĂŞcomputer interfaces," Journal of Neural Engineering, vol. 15, no. 5, p. 056019, 2018. [Online]. Available: http://stacks.iop.org/1741-2552/15/i=5/a=056019

[14] L. H. Monsein, A. Y. Razumovsky, S. J. Ackerman, H. J. Nauta, and D. F. Hanley, "Validation of transcranial doppler ultrasound with a stereotactic neurosurgical technique," Journal of neurosurgery, vol. 82, no. 6, pp. 972–975, 1995.

[15] N. Stroobant and G. Vingerhoets, "Transcranial doppler ultrasonography monitoring of cerebral hemodynamics during performance of cognitive tasks: a review," Neuropsychology review, vol. 10, no. 4, pp. 213–231, 2000.

[16] P. Welch, "The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," IEEE Transactions on audio and electroacoustics, vol. 15, no. 2, pp. 70–73, 1967.

[17] Y. K. Jain and S. K. Bhandare, "Min max normalization based data perturbation method for privacy protection," International Journal of Computer & Communication Technology, vol. 2, no. 8, pp. 45–50, 2011.

[18] S.-H. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," 2007.

[19] J. H. Friedman, "Regularized discriminant analysis," Journal of the American statistical association, vol. 84, no. 405, pp. 165–175, 1989.

[20] J. Friedman, T. Hastie, and R. Tibshirani, The elements of statistical learning. Springer series in statistics New York, NY, USA:, 2001, vol. 1, no. 10.

[21] R. F. Woolson, Wilcoxon Signed-Rank Test. American Cancer Society, 2008, pp. 1–3. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780471462422.eoct979