BHATTACHARYYA DISTANCE-BASED TRANSFER LEARNING FOR A HYBRID EEG-FTCD BRAIN-COMPUTER INTERFACE

Elise Dagois, Aya Khalaf, Ervin Sejdic and Murat Akcakaya

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

ABSTRACT

In this paper, we introduce a transfer learning approach for our novel hybrid brain-computer interface in which electroencephalography and functional transcranial Doppler ultrasound are used simultaneously to record brain electrical activity and cerebral blood velocity respectively due to flickering mental rotation and word generation tasks. We reduced each trial into a scalar score using Regularized Discriminant Analysis (RDA). For each individual, class conditional probabilistic distribution of each mental task was estimated using RDA scores of the trials corresponding to that mental task. Similarities between class conditional distributions across individuals were measured using Kullback-Leibler divergence, Bhattacharyya, and Hellinger distances. Classification task was performed using Quadratic Discriminant Analysis (QDA), Linear Discriminant Analysis (LDA), and Support Vector Machines (SVM). We demonstrate that transfer learning can reduce calibration requirements up to $\%87.5$. Moreover, it was found that QDA provides the most significant performance improvement compared to the case when no transfer learning is employed.

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

1. INTRODUCTION

In recent years, transfer learning has been used extensively to develop classification techniques that utilize previously acquired data to train a model that will work on unknown datasets and to improve the generalization performance of classifiers. These methods mimic human memory to generalize the acquired knowledge to perform various tasks [1]. Such transfer learning methods have the capability to increase classification performance when only a small dataset is available to train a classification model [2],[3]. Noninvasive brain computer interface (BCI) design is one application domain that would benefit from such transfer learning approaches.

Non invasive BCIs are designed to help individuals with neurological deficits or with Limited Speech and Physical Abilities (LSPA) to communicate with their caretakers without any surgical interventions. Most noninvasive BCIs are based on electroencephalography (EEG) due to its high temporal resolution, portability and low cost. However, since EEG has low signal-to-noise ratio and EEG demonstrates nonstationarities due to background brain activities, systems built based only on EEG suffer from performance degradations. Hybrid BCI systems are built to overcome the shortcomings of EEG-only BCIs [4]. However the modalities commonly used to complement EEG for hybrid BCI design such as functional near-infrared spectroscopy (fNIRS) lack the speed and accuracy to be used for real-time BCI applications [5],[6]. In our previous work, we have shown that functional transcranial Doppler (fTCD) that measured blood flow in the brain due to different mental activities has temporal resolutions comparable to EEG to complement EEG for hybrid BCI design and fTCD is robust to nonstationarities due to background brain activities [7],[8]. Moreover we have shown that such a hybrid BCI based on EEG and fTCD outperforms BCIs based on EEG and fTCD only [7].

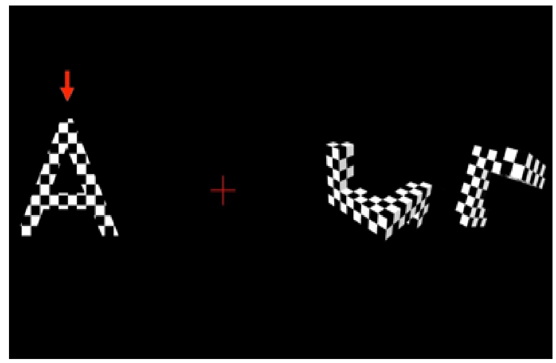

Fig. 1. Stimulus presentation for our flickering MR/WG hybrid BCI

More specifically, the hybrid system we have developed is based on two different paradigms jointly presented to the BCI user to induce changes in EEG and fTCD simultaneously, see Figure 1 [7]. In this figure, there are three visual stimuli: the letter on the left, the geometric shape on the right and a cross sign for baseline. Both the letter and geometric shape images include checkerboards flickering with different frequencies. Therefore, the EEG component of the hybrid BCI is based on steady state visual evoked potentials (SSVEPs) that are recorded in response to these flickering stimuli with two different frequencies [9],[10],[11]. Moreover, the letter on the left and the geometric shape on the right instruct the user to perform word generation (starting with the specific letter presented on the screen) and mental rotation of the presented three dimensional geometric shape to induce changes in the blood flow in two different parts of the brain that are recorded by fTCD. We have utilized such a visual presentation paradigm to induce changes in EEG and fTCD simultaneously.

Using the system presented in Figure 1, we have performed BCI experiments with 11 healthy participants. Through these experiments we have developed 3 different binary classifications including word generation (WG) vs mental rotation (MR), MR vs baseline and WG vs baseline, and we have shown that hybrid BCI increases the classification accuracy by $\%5$ compared to BCI based on EEG only for the same tasks [7]. Even though the hybrid system improves the performance compared to a BCI based on EEG only, such a hybrid system requires long calibration data (which causes fatigue especially in target population) to design user specific classifiers for the BCI operation.

In this paper, we extend our existing system to develop a novel transfer learning algorithm and to decrease calibration requirements. The proposed transfer learning method is based on increasing calibration datasets for each user through identification of probabilistically similar datasets recorded from other users. In the proposed method, EEG and fTCD features are extracted and class conditional distributions (for the above mentioned 3 binary classification problems) are computed using kernel density estimation (KDE). Three different probabilistic distance measures including Bhattacharyya and Hellinger distances and Kullback-Leibler (KL) divergence are compared to identify datasets for most accurate transfer learning. Moreover for final classification (3 binary classification problems), three different classifiers are compared: Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) and support vector machine (SVM). Through our experimental results, we demonstrate that transfer learning can reduce the training dataset by up to $\%87.5$. Also we show that while LDA provides the best classification performance, QDA demonstrates the highest difference between transfer learning and no transfer learning.

2. METHODS

In this section, we explain the details of our experimental setup as well as pre-processing and features extraction stages. Finally, we introduce our proposed transfer learning algorithm.

2.1. Experimental Setup

EEG signals are collected using 16 electrodes located over frontal, central, and parietal lobes at the following positions: Fp1, Fp2, F3, F4, Fz, Fc1, Fc2, Cz, P1, P2, C1, C2, Cp3, Cp4, Cp5, and P6. The reference electrode is placed over the left mastoid. A g.tec EEG system equipped with a g.USBamp amplifier is used for data collection. The data are transferred to a laptop via USB 2.0.

fTCD data are collected using a SONARA TCD system via two 2 MHz transducers. These transducers have to be placed on the left and right sides of the transtemporal window located above the zygomatic arch [12].

It was found that fTCD can successfully differentiate mental rotation (MR) and word generation (WG) tasks [7]. However, since the tasks designed for the hybrid BCI system need to be distinguished by both EEG and fTCD to achieve high performance, with the aim of inducing SSVEPs in EEG, MR and WG tasks are modified such that they flicker with frequencies of 7 and 17 Hz respectively. Moreover, these task are covered with a checkerboard pattern as shown in Figure 1.

16-channel EEG and 2-channel fTCD data are recorded simultaneously and synchronized using time stamps obtained from both modalities. A total of 11 healthy participants with ages ranging from 23 to 32 years old participated in this experiment under University of Pittsburgh approved IRB. Each participant attended one session of 25-min duration. During data collection, each participant is asked to focus on a screen showing visual icons for flickering MR, flickering WG, and baseline represented by a fixation cross as shown in Figure 1. A red vertical arrow randomly selects the task to be performed and points to that task for 10 s (trial length). For flickering WG, the participant has to silently generate words starting with the letter shown on the screen. For flickering MR, the user has to mentally rotate a pair of 3D shapes to decide if they are identical or mirrored. 150 trials are presented per session . Assuming equiprobability, about 50 trials were presented for each task.

2.2. Pre-processing and features extraction

For each individual, EEG and fTCD data of each trial are segmented. For every trial, we consider the power spectrum values estimated using Welch method [13] as features. Instead of considering the power spectrum values over all frequency bins, the number of features is reduced by taking the average power over a small range of frequencies. In particular, features are obtained by taking the average of power over every 2 Hz for EEG data and every 50 Hz for fTCD data as fTCD has much higher bandwidth (2.5 kHz compared to 40 Hz for EEG). EEG and fTCD feature vectors are normalized separately using min-max normalization [14] and the normalized EEG and fTCD feature vectors are concatenated into a single vector that represents each trial.

2.3. RDA Scores and Similarity Measures

To apply transfer learning , similarity is measured between the dataset under test and the other datasets collected from other individuals. KL divergence, Bhattacharyya, and Hellinger [15] are used to measure similarity across individuals. Instead of measuring similarity directly using EEG and fTCD concatenated feature vectors, these feature vectors are reduced using RDA [16] [17] into scalar RDA scores. For each binary classification problem, both RDA parameters, λ controlling the degree of shrinkage and γ the regularization parameter, are optimized via 10-fold cross-validation such that the area under the receiver operating characteristic curve is maximum. λ and γ are ranging from 0.1 to 1 with a step of 0.1. Using the RDA scores of the trials corresponding to each class, the conditional pdfs are estimated. Rbf kernel is used for KDE and the kernel bandwidth is estimated based on 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.

2.4. Classification

Three classification problems are formulated based on the data available from the hybrid BCI system including WG vs MR, WG vs baseline (WG vs BL), and MR vs baseline (MR vs BL). To perform the classification, different classifiers are used including linear discriminant analysis (LDA), quadratic discriminant Analysis (QDA), and linear support vector machine (SVM). The performance of these classifiers is discussed in details in the results section.

2.5. Transfer learning algorithm

In this study, we develop a transfer learning algorithm to reduce the training requirements of the SSVEP hybrid BCI system that utilizes both EEG and fTCD modalities. The pseudocode of the proposed transfer learning approach is given in Algorithm 1.

For each participant, we find the five most similar datasets using the different distance measures mentioned in section 2.3. In particular, to determine the similarity, we use KDEs of class conditional distributions of the scores obtained from the RDA projection of EEG and fTCD features. For each participant, given a number of t test trials and $N-t$ training trials, with N the total number of trials, participant's dataset is divided into testing and training sets of size t and $N-t$ respectively. Then, KDE of RDA scores of training trials for the participant under test is compared with KDEs of other participants to determine the most similar five datasets. Then, each participant's training set is then augmented with the most similar 5 datasets. As we aim to study the effect of transfer learning at different training and test sizes, we apply the proposed algorithm at t ranging from 10 to 90 trials which means the minimum training size is 10 samples.

Table 1. Mean of accuracy among all participants and corresponding sensitivity and specificity for WG vs MR.

Performance measures	LDA		QDA		SVM	
	NT	TL	NT	TL	NT	TL
Accuracy	0.8033	0.8533	0.79	0.8622	0.8178	0.8544
Specificity	0.8698	0.8298	0.8419	0.8234	0.8930	0.8362
Sensitivity	0.7426	0.8791	0.7426	0.9047	0.7489	0.8744

3. RESULTS

For each participant, we first utilize KL divergence, and Bhattacharyya and Hellinger distances to identify the datasets to be used for transfer learning for that specific participant. Through transfer learning (TL) we study the reduction in training size requirements compared to no transfer learning case (NT) for the above mentioned three binary classification problems. More specifically, we perform a one sided Wilcoxon rank test [18] between accuracy obtained with TL with a minimum size of $N-t=10$ training trials and accuracy obtained with NT with a number a training trials in the range 20 to 90 trials. The same test is repeated for each size of NT. An identical statistical comparison is applied for the different distance measures. When TL is performed with 10 trials, the statistical test shows that the performance for TL using Bhattacharyya distance is comparable to the performance for NT at a size of 80 trials. It means that instead of running the algorithm with 80 training trials without transfer, the system gives higher or at least same performance when using only 10 training trials for TL. Similar analysis shows that around 60 trials are needed for TL to obtain identical performance as NT when KL divergence and Hellinger distance are used for dataset identification for TL. Since Bhattacharyya provides the best results for the reduction of training set requirements through TL, in the remaining of this paper we present the classification performances for the three binary classifications provided through Bhattacharrya distance based transfer learning.

Considering a test and a training size of $t=90$ and $N-t=10$ trials respectively, accuracy, specificity, and sensitivity for the three classifiers are given in Tables 1, 2, and 3 with and without transfer learning for MR vs WG, MR vs BL and WG vs BL. The results are presented for an optimal trial length T_s chosen between 1 and 10 s. From Table 1, 2, and 3, we observe that all TL classification results are higher than the NT results for MR vs WG, MR vs BL, and WG vs BL classifications. LDA provides the best classification performance for MR vs BL and WG vs BL classification problems with corresponding accuracies of $\%82.22$ and $\%75.89$ respectively. On the other hand, QDA provides slightly higher performance compared to LDA for the MR vs WG classification. Finally, when the difference between the TL and NT cases are consid-

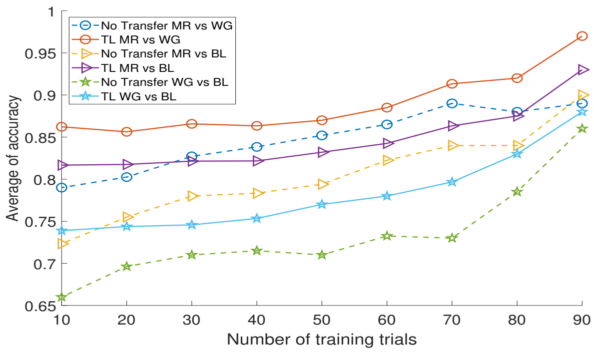

Fig. 2. Accuracy as a function of the number of trials in the training set for each binary problem with and without transfer.

ered, QDA provides the highest improvement for all the classification problems. Moreover, we compare results obtained from TL and NT when QDA is used as a classifier. Figure 2 shows the accuracies (optimal accuracies averaged across individuals) obtained for a number of training trials from 10 to 90 for MR vs WG, MR vs BL, and WG vs BL. The results aim to show the influence of the quantity of information available to train a classifier as well as the effect of the calibration data size on the classification accuracy with and without transfer. For each size of training set, performance obtained with TL is greater than without transfer for all classification problems.

4. CONCLUSION

With the aim of reducing the BCI calibration requirements, we developed a transfer learning algorithm for our novel hybrid BCI system that exploits data acquired from EEG (based on SSVEP) and fTCD (based on WG and MR) modalities simultaneously to infer user intent. To measure similarity across subjects, for every individual, each trial was projected into a scalar RDA score. These scores were used to estimate a class conditional distribution per each mental task. Similarities between class conditional distributions were measured across subjects using 3 different distance measures. It was found that the proposed transfer learning approach not only reduces the calibration requirements but also improves performance of all classifications problems including MR vs WG, MR vs BL, and WG vs BL. In particular, we show that LDA obtained the highest possible performance compared to SVM and QDA. In term of the difference between accuracy achieved using transfer learning and accuracy obtained without transfer learning, the highest difference was provided by QDA. In terms of calibration requirements, experimental results indicate that the size of the training set can be reduced by up to $\%87.5$ using Bhattacharyya distance as a similarity measure.

```
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^n_E, T^n_R,];
        T^n_E: testing trials for participant n;
        T^n_R: training trials for participant n;
        for T^n_E = 10 : 10 : 90 do
          Train the classifier using [T^n_R, T^l_R],
           l ⊂ m contains indexes
           corresponding to top similar 5
           datasets;
          Test the model using T^n_E;
          Compute performance measures;
        end
      end
    end
  end
end
```

Algorithm 1: Pseudocode for the proposed transfer learning approach.

Table 2. Mean of accuracy among all participants and corresponding sensitivity and specificity for MR vs BL.

Performance measures	LDA		QDA		SVM	
	NT	TL	NT	TL	NT	TL
Accuracy	0.7778	0.8222	0.7233	0.8167	0.7856	0.8122
Specificity	0.7750	0.8720	0.74	0.8580	0.7550	0.86
Sensitivity	0.78	0.76	0.71	0.7650	0.81	0.7525

Table 3. Mean of accuracy among all participants and corresponding sensitivity and specificity for WG vs BL.

Performance measures	LDA		QDA		SVM	
	NT	TL	NT	TL	NT	TL
Accuracy	0.72	0.7589	0.66	0.7389	0.7111	0.7511
Specificity	0.6550	0.8120	0.50	0.7820	0.6175	0.8140
Sensitivity	0.7720	0.6925	0.7880	0.6850	0.7860	0.6725

5. REFERENCES

[1] Si-Chi Chin, "Knowledge transfer: what, how, and why," 2013.

[2] Sinno Jialin Pan, Qiang Yang, et al., "A survey on transfer learning," IEEE Transactions on knowledge and data engineering, vol. 22, no. 10, pp. 1345–1359, 2010.

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

[4] Setare Amiri, Reza Fazel-Rezai, and Vahid Asadpour, "A review of hybrid brain-computer interface systems," Advances in Human-Computer Interaction, vol. 2013, pp. 1, 2013.

[5] Byoung-Kyong Min, Matthew J Marzelli, and Seung-Schik Yoo, "Neuroimaging-based approaches in the brain–computer interface," Trends in biotechnology, vol. 28, no. 11, pp. 552–560, 2010.

[6] Noman Naseer and Keum-Shik Hong, "fnirs-based brain-computer interfaces: a review," Frontiers in human neuroscience, vol. 9, pp. 3, 2015.

[7] Aya Khalaf, Ervin Sejdic, and Murat Akcakaya, "Towards optimal visual presentation design for hybrid eeg—ftcd brain–computer interfaces," Journal of Neural Engineering, vol. 15, no. 5, 2018.

[8] Aya Khalaf, Matthew Sybeldon, Ervin Sejdic, and Murat Akcakaya, "An eeg and ftcd based bci for control," in Signals, Systems and Computers, 2016 50th Asilomar Conference on. IEEE, 2016, pp. 1285–1289.

[9] Minjue Wang, Ian Daly, Brendan Z Allison, Jing Jin, Yu Zhang, Lanlan Chen, and Xingyu Wang, "A new hybrid bci paradigm based on p300 and ssvep," Journal of neuroscience methods, vol. 244, pp. 16–25, 2015.

[10] Xiaogang Chen, Yijun Wang, Masaki Nakanishi, Xiaorong Gao, Tzyy-Ping Jung, and Shangkai Gao, "High-speed spelling with a noninvasive brain–computer interface," Proceedings of the national academy of sciences, vol. 112, no. 44, pp. E6058–E6067, 2015.

[11] Masaki Nakanishi, Yijun Wang, Xiaogang Chen, Yu-Te Wang, Xiaorong Gao, and Tzyy-Ping Jung, "Enhancing detection of ssveps for a high-speed brain speller using task-related component analysis," IEEE Transactions on Biomedical Engineering, vol. 65, no. 1, pp. 104–112, 2018.

[12] Lee H Monsein, Alex Y Razumovsky, Stacey J Ackerman, Haring JW Nauta, and Daniel F Hanley, "Validation of transcranial doppler ultrasound with a stereotactic neurosurgical technique," Journal of neurosurgery, vol. 82, no. 6, pp. 972–975, 1995.

[13] Peter 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.

[14] Y Kumar Jain and Santosh Kumar 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.

[15] Sung-Hyuk Cha, "Comprehensive survey on distance/similarity measures between probability density functions," 2007.

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

[17] Jerome Friedman, Trevor Hastie, and Robert Tibshirani, The elements of statistical learning, vol. 1, Springer series in statistics New York, NY, USA:, 2001.

[18] R. F. Woolson, Wilcoxon Signed-Rank Test, pp. 1–3, American Cancer Society, 2008.