Boosting Transfer Learning Improves Performance of Driving Drowsiness Classification Using EEG

Jiancong He and Guoxu Zhou
Faculty of Automation
Guangdong University of Technology
Guangzhou, China

Hongtao Wang
School of Information Engineering
Wuyi University
Jiangmen, China

Evangelos Sigalas, Nitish Thakor
Anastasios Bezerianos and Junhua Li*
Singapore Institute for Neurotechnology (SINAPSE)
National University of Singapore
Singapore
juhalee.bcmi@gmail.com

***Abstract*—Drowsy driving poses considerable risk not only to drivers themselves but also to other people on the road. It has been demonstrated that information contained in electroencephalography (EEG) signal can be used to identify driving drowsiness. To date, most of work focused on the detection of drowsiness within a session. This hampers the generalization of the trained model to a following session conducted after a few days. As we know, EEG is non-stationary and changes dramatically across sessions, which leads to a great challenge how to establish a model that has a good performance across sessions. In this study, we combined boosting strategy and transfer learning method to establish a model for identifying driving drowsiness states from alertness states based on the features of power spectral density (PSD). The model trained using the data collected a few days ago (session1) was tuned using very small portion of the data collected in the current session can achieve a good performance as tested in the current session (session2). The results demonstrated that the proposed boosting transfer learning method significantly outperformed the support vector machine (SVM) and AdaBoost methods. The proposed method could promote practical use of drowsiness detection system in a real vehicle due to its good cross-session performance.**

I. INTRODUCTION

Mental drowsiness could result in unreliable judgment, slow reaction, and increasing risk of improper operation. It is well known that driving drowsiness is a significant cause of fatal traffic accident, numerous damage and possible loss of life, as well as people on the road such as motorbike riders and pedestrians [1]. A drowsiness detection system can be used to warn drivers and prevent them from drowsy driving so that drowsiness-related accidents can be eliminated.

To date, diverse indicators of drowsiness have been investigated. One of them is the driving behavior. For instance, drowsiness can be detected through steering actions and lane keeping performance. It can also be detected from video. The sign of drowsiness on the face can be captured by a real-time video processing system based on facial feature tracking technology [2]. Drowsiness can also be assessed through a questionnaire. However, this method is very subjective and its success largely depends on drivers, which is susceptible to drivers' mental status (e.g., memory and consciousness). Literature [3]–[5] has illustrated that drowsiness detection based on self-report is unauthentic. In contrast, EEG-based methods have been shown to be more objective for drowsiness detection and EEG signal is of an adequate temporal resolution [6]. Besides, EEG signal is also widely utilized in clinical application, health care, and laboratory researches. The most challenging problem for the use of EEG signal is its non-stationarity, which is accounted for physiological and instrumental interferences. Therefore, the distribution of EEG varies across sessions which are recorded in different days. In other words, a model trained by the data collected several days ago may not well classify the data currently collected. This issue can be released by transfer learning method. A model trained using the data from past session and a small portion of current session can perform well due to the advantage of intrinsic information transferring. Another method, namely boosting strategy, is able to automatically adjust the weights of training samples in order to focus on the informative instances. In this study, we considered the advantages of transfer learning and boosting methods and combined them for driving drowsiness classification in order to have a good performance in the cross-session classification.

AdaBoost is an adaptive framework which carefully tweaks the weights in favor of those misclassified samples with informative data. As an extension to AdaBoost, transfer AdaBoost learning framework, namely TrAdaBoost [7], adds a mechanism to increase the weights of source samples similar to target data in order to strengthen the linkage between source and target. If a model is trained using multiple sources which are different from the target, the TrAdaBoost is called MultiSourceTrAdaBoost [8]. In this study, data collected a few days ago are considered as source and data collected currently are considered as target. The source data are used as a whole for the TrAdaBoost and source data are divided into several groups as multiple sources for MultiSourceTrAdaBoost. More details can be found in Section II. The results are presented in section III and the conclusion is finally drawn in section IV.

II. MATERIALS AND METHODS

A. Experiment and data preprocessing

The experiment protocol was reviewed and approved by the Institutional Review Board of the National University of Singapore. Nine healthy subjects aged between 20 and 27 years (two of them are female) participated in the study.

* indicates the corresponding author (email: juhalee.bcmi@gmail.com).

978-1-5386-6859-7/18$31.00 © 2018 IEEE

All subjects have normal or corrected-to-normal vision and no history of substance addiction or mental disorders. They were required to have a full night sleep (>7h) prior to the experiment and to refrain from consuming caffeine or alcohol on the day of the experiment. They were instructed to drive virtual car by the Logitech G27 Racing Wheel set to follow a guiding car and avoid collision as possible as they can. Each subject performed two identical driving sessions (one hour and a half for each session), which were implemented approximately one week apart from each other. 24 EEG channels and 2 EOG channels were utilized to record signal by Cognionics headset (Cognionics, Inc., USA) at the sampling rate of 250 Hz.

The EEG data were preprocessed using EEGLAB toolbox [9] through the following steps: common average reference, poor contact channel interpolation, data removal of the last 5-min recording (because of the change of driving mode), band-pass filtering (0.5 $\sim$ 45 Hz), partitioning continuous recording into 2-sec long epochs, abnormal epoch reject, ICA artifacts removal. After the preprocessing, the epochs within the first 15 minutes were considered as samples of alertness while the epochs within the last 15 minutes were considered as samples of drowsiness. The number of alertness samples is 405 ± 27 (mean$\pm$standard deviation) and the number of drowsiness samples is $359{\pm}41$ in session 1. There are $416{\pm}30$ alertness samples and 327 ± 59 drowsiness samples in session 2.

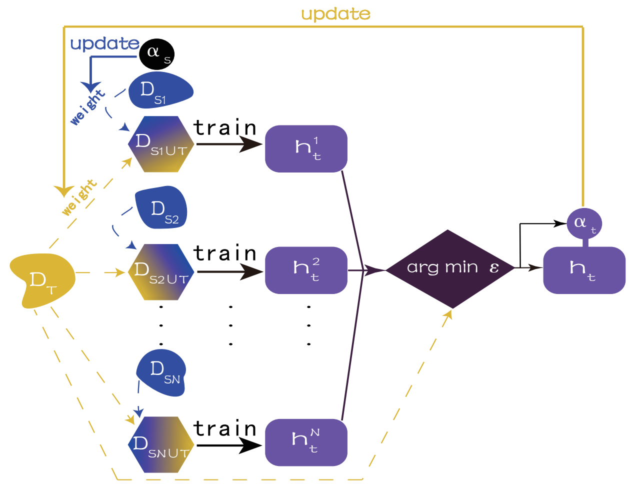

Fig. 1: Illustration of the MultiSourceTrAdaBoost. The yellow block D_T represents extremely limited training data from target domain. N blue blocks $D_{S1}, D_{S2}, ..., D_{SN}$ represent the training data from N source domains. N hexagonal blocks stand for the training data from both D_T and D_S after weighting samples. N weak classifier candidates are trained using N combined training data, respectively. The best one with minimum error rate as tested on a small portion of training data from D_T is selected as a weak classifier h_t. All dashed lines represent data weighting. The weights of training samples are updated according to the formulas (1) and (2).

B. Feature extraction

Fourier transform was used to convert temporal EEG data into spectral data. In this study, five typical frequency bands (i.e., delta: 1 $\sim$ 3 Hz, theta: 4 $\sim$ 7 Hz, alpha: 8 $\sim$ 13 Hz, beta: 14 $\sim$ 30 Hz and gamma: 31 $\sim$ 40 Hz) were extracted to be features, resulting in a feature vector of 120 dimensions (24 channels $\times$ 5 frequency bands). In order to reduce the variation of band powers across samples, band powers were normalized to the total power of five bands. In addition, principal components analysis (PCA) was employed to reduce the number of dimensions in the feature space.

C. Transfer learning

A domain notated as $\mathcal{D} = \{\mathcal{X}, P(X)\}$ contains two parts, a feature space $\mathcal{X}$ and a marginal probability distribution $P(X)$, where $X = \{x_1, x_2, ..., x_n\}$ and $x_i \in \mathcal{X}$. A task notated as $\mathcal{T} = \{\mathcal{Y}, f(x)\}$ also contains two parts, a label space $\mathcal{Y} = \{+1, -1\}$ (alertness state and drowsiness state) and a boolean function $f(x)$ that returns the label for sample x. The traditional machine learning algorithms assume that source domain $\mathcal{D}_S$ is exactly same to target domain $\mathcal{D}_T$. However, they are different in our case because EEG signals are non-stationary and vary from one session to another session. Therefore, traditional machine learning algorithms cannot well classify mental states across sessions. Transfer learning methods may tackle this problem. The invariant information learned from the data collected in a past session (e.g., a few days ago) can be transferred to benefit the classification of the current session by shifting learning attention on the samples which are similar to the samples of the current session. Specifically, all samples from $\mathcal{D}_S$ and a small portion of samples randomly selected from $\mathcal{D}_T$ were used to train a model. The samples from $\mathcal{D}_S$ that are more similar to the samples from $\mathcal{D}_T$ are assigned larger weights so that the model learn more from these samples. Therefore, the model could better classify the current session as its learning focuses on the samples similar to the sample of the current session.

D. Boosting transfer learning

Domain adaptation is a kind of policy in transfer learning that attempts to make the distribution of $\mathcal{D}_S$ closer to $\mathcal{D}_T$ from the source task $\mathcal{T}_S$ to the target task $\mathcal{T}_T$. It was proposed to let classifiers learn with a limited or even no labeled samples from $\mathcal{D}_T$ by leveraging a large number of labeled training data from $\mathcal{D}_S$. However, it doesn't work satisfactorily when $\mathcal{D}_S$ transfer uninformative knowledge. In this case, the brute-force transfer to $\mathcal{D}_T$ is harmful to classification performance in $\mathcal{T}_T$, which is well-known as negative transfer. Thus, knowledge selection is a key procedure in domain adaptation.

We recall AdaBoost, a famous boosting algorithm aiming to improve classification performance by adjusting the weights of training data at each iteration in order to emphasize on misclassified samples. Inspired by this, the informative samples from $\mathcal{D}_S$ are emphasized through increasing their weights. They are believed to be similar to target samples and supposed to increase influence on weak classifiers at the next iteration. In

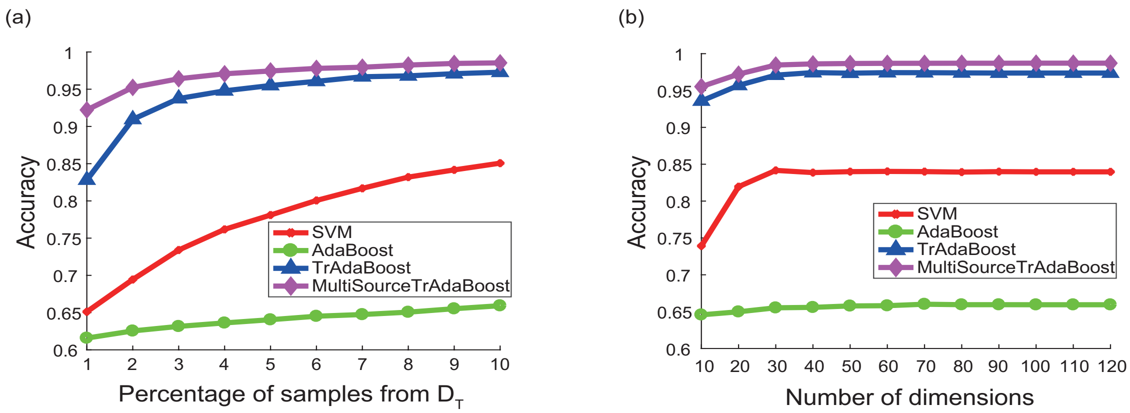

Fig. 2: Average accuracy comparison among four methods. (a) The change of accuracies averaged across all subjects along with different percentages of samples from $\mathcal{D}_T$ used for training when 30 dimensions are retained after the PCA dimension reduction. (b) The change of accuracies averaged across all subjects along with different number of dimensions retained after the PCA dimension reduction when 9 % of samples from $\mathcal{D}_T$ were used for training.

contrast, the weights of the less informative samples from $\mathcal{D}_S$ should be decreased so as to weaken their impacts at the next iteration. More precisely, the updating rule of the weights of training samples from $\mathcal{D}_S$ is

$$\omega_i^{t+1} = \omega_i^t e^{-\alpha_S |h_t(x_i) - y_i|}, i = 1, 2, ..., M_S \tag{1}$$

where M_S is the number of training samples from $\mathcal{D}_S$, ω_i^{t+1} and ω_i^t are the weights of the i^{th} sample in the current iteration and the next iteration, respectively. $h_t(x_i)$ is the weak classifier's prediction of the i^{th} sample after current training iteration and y_i is the ground truth of the i^{th} sample. α_S is a predefined coefficient related to the number of source samples n_s and the maximum number of iterations M according to the formulas (3). The updating rule of the weights of training samples from $\mathcal{D}_T$ is similar to that of AdaBoost as follows

$$\omega_i^{t+1} = \omega_i^t e^{\alpha_T^t |h_t(x_i) - y_i|}, i = 1, 2, ..., M_T \tag{2}$$

$$\alpha_S = \frac{1}{2} \ln (1 + \sqrt{2 \ln \frac{n_S}{M}}) \tag{3}$$

where M_T is the number of training samples from $\mathcal{D}_T$, α_T^t is a coefficient corresponding to error rate of the weak classifier over target domain. This algorithm is an extension of the AdaBoost, known as TrAdaBoost [7].

TrAdaBoost might lead to negative transfer since TrAdaBoost transfers knowledge only from one source. The performance of TrAdaBoost only depends on the relationship between a unique source and target. If there are more sources available and each of them is different from target, the knowledge can be transferred from multiple sources so that the risk of negative transfer can be greatly reduced. This strategy is known as MultiSourceTrAdaBoost [8]. In our case, samples from the source session are randomly partitioned into N groups, which are considered as N sources. It is worth noting that each source should include both alertness and drowsiness samples. Samples from N sources are respectively merged with the samples from $\mathcal{D}_T$ to form N sets of training samples. These N sets of training samples are then used to train N weak classifier candidates, respectively. Subsequently, the best weak classifier with minimum error rate as tested on a small portion of training data from $\mathcal{D}_T$ is chosen. The α_T^t mentioned above represents the coefficient of the weak classifier chosen at the t^{th} iteration,

$$\alpha_T^t = \frac{1}{2} \ln \frac{1 - \varepsilon_t}{\epsilon_t} \tag{4}$$

where ε_t is the minimum error rate at the t^{th} iteration. Assuming that the number of iterations is M, we can finally obtain M weak classifiers. The final decision function $f(x)$ is a linear combination of M weak classifiers as follow,

$$f(x) = sign(\sum_{t=1}^{M} \alpha_T^t h_t(x)) \tag{5}$$

An illustration of the MultiSourceTrAdaBoost is shown in Fig. 1.

III. Results

As support vector machine (SVM) is a powerful supervised learning algorithm which has been widely and frequently used for a number of classification tasks, we therefore compared the boosting transfer leaning methods (i.e., TrAdaBoost and MultiSourceTrAdaBoost) to the SVM. The boosting transfer leaning methods were also compared to the AdaBoost because of its popularity in the category of boosting method. In the comparisons, training data and testing data were kept identical for all methods. For the SVM, a 'grid-search' recommended in [10] was utilized to seek optimal parameters ($\mathcal{C}$ in the linear kernel, $\mathcal{C}$ and γ in the radial basis function (RBF) kernel) in order to avoid the deflation in the performance. We found that the SVM with the RBF kernel generally performed better than that with the linear kernel in our study. Therefore, the reported accuracies for the SVM method were obtained using the RBF kernel. For the methods of AdaBoost, TrAdaBoost and MultiSourceTrAdaBoost, linear SVM was used as weak classifier. For the MultisourceTrAdaBoost method, the number of sources N was optimized at a range of 1 $\sim$ 60 with an incremental step of 1, and the best results were reported. Because accuracy was not consistent due to random selection of samples, we repeated the evaluation 10 times and compared average accuracies between methods.

The performance was affected by the percentage of samples from $\mathcal{D}_T$ used for training and the number of retained dimensions after the PCA dimension reduction. We therefore compared the performance between methods using different settings. The results were shown in Fig. 2. The left panel in Fig. 2 shows that the classification performance was monotonously increased as the increasing in the number of samples from $\mathcal{D}_T$. The accuracy was quickly elevated at the beginning of the increasing of the samples from $\mathcal{D}_T$ and then tended to slight improvement for the boosting transfer learning methods. The accuracy of SVM was gradually increased with the increasing of the samples from $\mathcal{D}_T$. The improvement was very less for the AdaBoost. In the comparison of classification performance, the boosting transfer learning methods consistently outperformed the SVM and AdaBoost. The MultiSourceTrAdaBoost is better than the TrAdaBoost, which might be due to much lower probability of negative transfer. The right panel in Fig. 2 shows that the accuracies of all methods were first increased along with the increment of the dimensions retained after the PCA, and then were not changed too much after 30 dimensions.

Fig. 3 shows the performance comparison of the four methods when 9 % of samples from $\mathcal{D}_T$ were used for training and the number of dimensions retained after the PCA dimension reduction was 30. According to the two-tailed t-test, the boosting transfer learning methods significantly outperformed the AdaBoost and SVM methods. The highest performance (i.e., %) was achieved by MultiSourceTrAdaBoost method. The standard deviations of the boosting transfer learning methods were greatly less than that of the AdaBoost and SVM, which reflects that the boosting transfer learning methods are more robust for driving drowsiness classification across subjects.

IV. Conclusion

In the study, we combined boosting strategy and transfer learning method (i.e., boosting transfer learning) for driving drowsiness classification. The results demonstrated that the performance of driving drowsiness classification was significantly improved compared to the AdaBoost and SVM methods. MultiSourceTrAdaBoost method achieved the highest performance due to the advantage of multiple sources. TrAdaBoost method was worse than MultiSourceTrAdaBoost method in the classification performance, but was much better than AdaBoost and SVM methods. Taken together, it can be seen that the combination of boosting strategy and transfer learning method can improve the driving drowsiness classification. This suggests that the boosting transfer learning might be a good choice when developing a drowsiness detection system in a real vehicle.

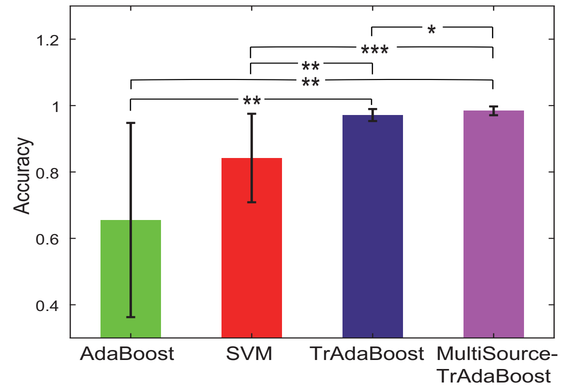

Fig. 3: Performance comparison of the four methods when 9 % of samples from $\mathcal{D}_T$ were used for training and the number of dimensions retained after the PCA dimension reduction was 30. The bars stand for average accuracies while the error bars represent standard deviations. Asterisks indicate the significance levels of accuracy differences between methods (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$).

Acknowledgment

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant 61673124 and the Guangdong Province Natural Science Foundation under Grant 2014A030308009. This work was also supported by the Ministry of Education of Singapore under the grant MOE2014-T2-1-115.

References

[1] G. Matthews, P. A. Desmond, C. Neubauer, and P. A. Hancock, "The handbook of operator fatigue," *Ergonomics*, vol. 56, no. 9, pp. 1486–1486, 2012.

[2] H. Gu and Q. Ji, "An automated face reader for fatigue detection," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 111–116.

[3] S. M. Belz, G. S. Robinson, and J. G. Casali, "Temporal separation and self-rating of alertness as indicators of driver fatigue in commercial motor vehicle operators." *Human Factors*, vol. 46, no. 1, pp. 154–169, 2004.

[4] H. J. Moller, L. Kayumov, E. L. Bulmash, J. Nhan, and C. M. Shapiro, "Simulator performance, microsleep episodes, and subjective sleepiness: normative data using convergent methodologies to assess driver drowsiness." *Journal of Psychosomatic Research*, vol. 61, no. 3, p. 335, 2006.

[5] E. A. Schmidt, M. Schrauf, M. Simon, M. Fritzsche, A. Buchner, and W. E. Kincses, "Drivers misjudgement of vigilance state during prolonged monotonous daytime driving." *Accid Anal Prev*, vol. 41, no. 5, pp. 1087–1093, 2009.

[6] M. Simon, E. A. Schmidt, W. E. Kincses, M. Fritzsche, A. Bruns, C. Aufmuth, M. Bogdan, W. Rosenstiel, and M. Schrauf, "Eeg alpha spindle measures as indicators of driver fatigue under real traffic conditions," *Clinical Neurophysiology*, vol. 122, no. 6, pp. 1168–1178, 2011.

[7] W. Dai, Q. Yang, G. R. Xue, and Y. Yu, "Boosting for transfer learning," in *International Conference on Machine Learning*, 2007, pp. 193–200.

[8] Y. Yao and G. Doretto, "Boosting for transfer learning with multiple sources," in *Computer Vision and Pattern Recognition*, 2010, pp. 1855–1862.

[9] A. Delorme and S. Makeig, "Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis." *J Neurosci Methods*, vol. 134, no. 1, pp. 9–21, 2004.

[10] C. J. Lin, "A practical guide to support vector classification," 2003, p. 012004.