Offline EEG-Based Driver Drowsiness Estimation Using Discrepancy Distance Based Deep Adversarial Domain Adaptation

Mengliang Zhu, Dongrui Wu
School of Artificial Intelligence and Automation
Huazhong University of Science and Technology, Wuhan, China
Email: mlzhu@hust.edu.cn, drwu@hust.edu.cn

***Abstract*—Drowsy driving is a major cause of traffic accidents and has become a serious threat to our daily life. Therefore, it is important to make accurate driver drowsiness estimation during driving for safety. This can be achieved by monitoring drivers' electroencephalogram (EEG) signals through a brain-computer interface system. However, due to individual differences, most existing approaches need to collect a large amount of labeled subject-specific data for calibration, which is labor-intensive and user-unfriendly. To solve this problem, domain adaptation (DA) has been proposed to make use of auxiliary subjects' labeled data to train a model for a new subject, with few or even no subject-specific calibration data. This paper proposes a novel deep adversarial DA approach for offline cross-subject drowsiness estimation, without using any labeled subject-specific calibration data. It utilizes a discrepancy distance to measure the distribution discrepancy in regression. A publicly available dataset, SEED-VIG, was used to evaluate the performance of our proposed approach. Experiment results demonstrated that it outperformed two state-of-the-art approaches.**

***Keywords*—Brain-computer interface, EEG, driver drowsiness estimation, deep adversarial DA , discrepancy distance**

I. Introduction

Driving safety is of vital significance with the rapid development and frequent use of the vehicle. However, the high incidence of traffic accidents has been a serious threat to our daily life. Drowsy driving is one of the major causes [1], which only follows to alcohol, speeding and inattention [2]. It is reported that there were at least 396,000 traffic accidents caused by drowsy driving from 2011 to 2015 in the USA [3]. When drivers are in states of high drowsiness, their abilities of quick response and accurate judgement to the road emergencies deteriorate and fall into high risk of car crashes [4]. Therefore, it is essential to make accurate drowsiness estimation during driving for early-warning and safety.

During the past decades, a variety of approaches have been proposed to estimate driver drowsiness level, which can be roughly categorized into two kinds: contactless-based sensor detections and wearable-based sensor detections [5]. The former mainly adopt the computer vision techniques to analyze the drivers' facial activities and then infer the drowsiness level [6]. Some contactless-based sensor sensors such as cameras are applied to monitoring drivers. But these approaches are sensitive to the changes and impacts of surroundings. The later mainly use wearable sensors to collect drivers' physiological signals, e.g., electroencephalogram (EEG) [7], electrocardiography (ECG) [8], electrooculogram (EOG) [9], etc, and then analyze the signals for drowsiness estimation. The whole process is implemented based on the brain-computer interface (BCI) system, which is a kind of communication system that directly builds a way between human brains and computers [10].

BCI-based approaches have be prevailing for drowsiness estimation [11]–[13]. Among the various signals in BCIs, EEG is a kind of frequently used time series signal by monitoring specific brain electromagnetic fields, which is a reliable measurement of human psychophysiological and mental states [14]–[16]. The last several years have witnessed great progress and a variety of works in EEG-based driver drowsiness estimation [6], [17]. Most of the proposed approaches only focus on classification by classifying driver states into some predefined categories [18]. Actually, compared with the arbitrary classification that whether the driver is drowsy or not, it is more meaningful and intuitive if the driving early-warning system can output continuous values for driver drowsiness estimation in practice. In this paper, we consider the offline EEG-based drowsiness estimation and regard it as a standard regression problem.

In BCI systems, EEG-based driver drowsiness estimation suffers from high cross-subject variations, which is one of the major obstacles for real-word BCI applications [19]. Due to the individual differences, performance of models trained on the data from one subject is degraded when applied to another subject. Hence, it is required to collect enough labeled subject-specific calibration data to tune the model parameters [5]. However, collecting labeled EEG data is extremely labor-intensive and user-unfriendly. Many efforts have been tried to reduce the calibration sessions as much as possible. Domain adaptation (DA) serves as a promising solution to solve such a chalilenge, which leverages data and knowledge in some tasks (usually called source domains) to boost the learning performance in a different but related task (usually called target domain) [20]. The cross-domain distribution discrepancy is minimized during training in DA and models learn from more transferable knowledge [21]. Therefore, with a little or even no calibration data of the target domain, models trained on labeled source samples can still perform well on target domain.

In the field of DA, deep DA approaches are more prevailing than traditional DA approaches due to the powerful feature representation of deep networks. The deep DA approaches generally embed DA techniques into deep network training [22] and have gain significant performance improvement. Deep adversarial DA currently is one of the dominant branches in deep DA [23]. Its basic idea is to diminish the distribution discrepancy between the source and target domains in a two-player game. The whole process is implemented by integrating adversarial learning [24], inspired by the principle of generative adversarial networks (GANs) [25]. The deep networks are trained to generate domain-invariant features for better knowledge transfer. However, to our best knowledge, deep adversarial DA approaches for offline cross-subject drowsiness estimation have not been fully explored and there only exist a few works. In [18], Li et al. used two proposed deep adversarial DA approaches, domain adversarial neural network (DANN) [24] and adversarial discriminative domain adaptation (ADDA) [26], to estimate drivers' fatigue level. In [9], Ma et al. extended DANN into domain generalization scenario and then proposed a novel adversarial network structure called domain residual network for driver drowsiness estimation. However, most of the proposed deep adversarial DA approaches are specifically designed for classification, including DANN, ADDA, and etc. While some can be modified for regression, it is not theoretically supportive and often failed in practice.

To adapt to regression problems for offline drowsiness estimation, we propose a novel deep adversarial DA approach based on discrepancy distance. The discrepancy distance can be used for measuring the distribution discrepancy in regression [27]. In our proposed approach, the deep network architecture consists of two regressors and a feature extractor. For the two regressors, one is a task-specific regressor for final estimation, and the other one acts as an adversarial regressor that likes the discriminator in GAN for adversarial learning. During the training process, the adversarial regressor and the feature extractor are trained in an adversarial manner to minimize the discrepancy distance across domains. Then domain-invariant feature representations are generated for the task-specific regressor, which is trained for final predictions. Additionally, it should be noticed that no labeled subject-specific calibration data is required from the target domain during the training process.

The remainder of this paper is organized as follows: Section II introduces related work on EEG-based drowsiness estimation and deep adversarial DA. Section III describes the learning framework and training steps of our proposed approach. Section IV presents the experiments results and evaluates the performance of our proposed approach. Finally, Section V draws the conclusions.

II. Related Work

In this section, we briefly introduce previous works on EEG-based drowsiness estimation and deep adversarial DA.

A. EEG-Based Drowsiness Estimation

Among various physiological signals, EEG is reported to be a reliable measurement of the transition between wakefulness and sleep in many studies because it can directly reflect the human brain state and activity [28]–[30]. Furthermore, according to [31], EEG has natural potential for fatigue detection. It has been proved that the spectral dynamics of EEG from the posterior brain regions are closely related with the decline of drowsiness level [32]. It has also been demonstrated that the EEG signals from this region can be used for drowsiness classification with great accuracy [30].

There were many works that the power spectrum of EEG was used for driver vigilance estimation in many works [33]. In [34], Wu et al. proposed a domain adaptation approach by model fusion for online drowsiness estimation, which only requires little labeled subject-specific calibration data. In [13], by integrating fuzzy sets with domain adaptation, a novel online weighted adaptation regularization algorithm for regression was proposed for drowsiness estimation based on the power spectrum of EEG. In [28], a multimodal approach was proposed to estimate driver drowsiness level using EEG and Forehead EOG. In [5], Cui et al. utilized feature weighting to learn the importance of different features, and adopted episodic training for domain generalization based on the power spectrum of EEG.

In this paper, we establish an offline cross-subject drowsiness estimation model based on the power spectrum of EEG signals without any labeled subject-specific calibration data of target subjects. We consider the drowsiness estimation as a regression problem, which is more meaningful in practice.

B. Deep Adversarial Domain Adaptation

Deep adversarial DA is now one of the most prevailing branches in the field of DA. The basic idea is to match feature distributions by embedding adversarial learning into the training process of deep neural networks, motivated by the idea of GANs [25].

Specifically, DANN first adopted adversarial learning by introducing the domain discriminator to generate domain-invariant feature representations, and the whole training can be easily optimized through a simple gradient reversal layer (GRL) [24]. Based on the adversarial learning framework of DANN, Tzeng et al. [26] designed ADDA approach using the GAN-based adversarial loss. Unlike that source and target domains shared a common feature extractor in DANN, there were two separate feature extractors for source and target domains respectively in ADDA. Lone et al [22] proposed the conditional domain adversarial network, where multilinear conditioning was used to capture the cross-covariance between feature representations and classifier predictions. Saito et al.

[35] proposed a novel adversarial training strategy, i.e. Maximum Classifier Discrepancy (MCD) that maximizes the discrepancy between the predictions of two task-specific classifier and simultaneously trains the feature extractor to minimize the discrepancy. Based on MCD, Li et al. [36] further proposed the joint adversarial domain adaptation approach. In this approach, both the class-wise and domain-wise distributions were jointly matched between source and target domains in a unified adversarial learning framework.

Although there have been plenty of deep adversarial DA approaches, most of them can only be applied to classification. And it is difficult to generalize them to regression due to the lack of theory support. For instance, $\mathcal{H}$-divergence is for the 0-1 loss of classification, which is optimized in DANN . Additionally, the theoretical insight of MCD is the mini-max optimization problem of $\mathcal{H}\Delta\mathcal{H}$-divergence, which is measured as the discrepancy between two classifiers' predictions. To adapt to the regression problem, we propose a novel deep adversarial DA based on discrepancy distance. The discrepancy distance is a kind of distribution discrepancy metric based on arbitrary loss functions, such as mean square error (MSE) loss for regression [27]. Therefore, the discrepancy distance can be used for measuring the distribution discrepancy in regression and the experiment results verified its the effectiveness.

III. Methodology

In this section, we first introduce the problem setting and notation. Then we review the discrepancy distance and define discrepancy loss correspondingly. Finally, we describe the learning framework and training steps of our proposed approach in detail.

A. Problem Setting and Notation

Let $X \in \mathbb{R}^{n\times d}$ denote the input space and $Y \in \mathbb{R}$ denote the output space, where d is the dimension of the input features and n is the number of samples. We consider the unsupervised DA scenario that we have access to labeled samples of the source domain $D_s = \{X_s, Y_s\}$, where $\{X_s, Y_s\} = \{(\mathbf{x_i^s}, y_i^s)\}_{i=1}^{n_s}$, and unlabeled samples of the target domain $D_t = \{X_t\}$, where $\{X_t\} = \{\mathbf{x_i^t}\}_{i=1}^{n_t}$. Typically, there exist difference between distributions of source and target domains, i.e. $P(X_s, Y_s) \neq P(X_t, Y_t)$, $P(X_s) \neq P(X_t)$. And it is the cross-domain distribution discrepancy that leads to performance degradation when the models trained on source domains are applied on target domains. Therefore, our goal is to learn a label function $f : X \to Y$ which can generalize well on target domains by utilizing deep adversarial DA.

Additionally, we denote a loss function $L : Y \times Y \to \mathbb{R}_+$. For any two functions $h, h^{'} : X \to Y$ and any distribution D over X, we denote by $\mathcal{L}_D(h, h^{'})$ the expected loss of $h(x)$ and $h^{'}(x)$ as shown in (1). L can be the MSE loss used for regression.

$$\mathcal{L}_D(h, h^{'}) = \underset{x\sim D}{E}\left[L(h(x), h^{'}(x))\right] \tag{1}$$

B. Discrepancy Distance and Discrepancy Loss

A key problem for DA is how to measure the distribution discrepancy between the source and target domains. Many distribution discrepancy metrics have been proposed for classification with theoretical support. However, most of them cannot be directly used for regression. To adapt to regression problems, we utilize the distribution discrepancy metric proposed in [27], named discrepancy distance. It can be used for measuring the distribution discrepancy in regression. The definition is given below.

Definition 1. Given a hypothesis set $\mathcal{H}$ and a loss function L, the discrepancy distance between two distributions P and Q over X is defined by:

$$disc(P, Q) = \max_{h,h^{'}\in\mathcal{H}} \left|\mathcal{L}_P(h^{'}, h) - \mathcal{L}_Q(h^{'}, h)\right| \tag{2}$$

The discrepancy distance is symmetric by definition and holds the triangle inequality for any loss function L [27]. In particular, L can be the MSE loss, which is commonly used in regression. And it has been proved that $P = Q$ when $disc(P, Q) = 0$ [27].

However, it is generally difficult to calculate the empirical discrepancy distance because it is difficult to find a maximum value over the hypothesis set $\mathcal{H}$. Therefore, we introduce the discrepancy loss to approximate the empirical discrepancy distance. The definition of discrepancy loss is given as follows.

Definition 2. Given two hypotheses h and $h^{'}$ for prediction, the discrepancy loss between them over two distributions P and Q is defined by:

$$\mathcal{L}_{disc}(h, h^{'}) = \left|\mathcal{L}_P(h^{'}, h) - \mathcal{L}_Q(h^{'}, h)\right| \tag{3}$$

The discrepancy distance can be approximated by maximizing the discrepancy loss over the whole hypothesis space $\mathcal{H}$. Therefore, minimizing the discrepancy distance can be equivalently regarded as a min-max optimization problem of the discrepancy loss, which will be applied and solved in our proposed approach by utilizing adversarial learning techniques.

C. Learning Framework

The learning framework of our proposed approach is shown in Fig. 1. In our proposed approach, the network architecture consists of three parts: a feature extractor G_f, an adversarial regressor G_{adv}, and a task-specific regressor G_y. G_{adv} and G_f are trained in an adversarial manner to minimize the discrepancy distance across source and target domains. Motivated by the adversarial training introduced in DANN [24], in our proposed approach, G_{adv} is optimized to maximize the discrepancy loss to approximate the empirical discrepancy distance, while G_f is trained to minimize the discrepancy loss to align the distributions of source and target domains. With the domain-invariant feature representations generated from G_f, G_y is trained for the regression task. In addition, it should be noticed that G_y and G_{adv} have the same network architectures and they are initialized differently to get different regressors at the beginning of training process. The whole training procedure is implemented under the adversarial DA framework.

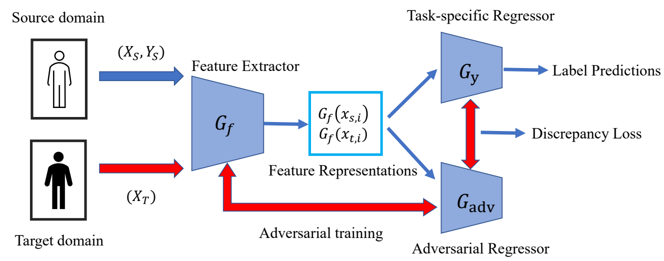


Fig. 1. The learning framework of our proposed approach, where G_f, G_y and G_{adv} are feature generator, task-specific regressor and adversarial regressor respectively. G_{adv} and G_f are trained in an adversarial manner with discrepancy loss. G_y is trained for final label predictions.

We will describe the learning framework of our proposed approach which is established in three steps.

1) Step 1: First, we train both task-specific regressor G_y and feature extractor G_f to perform well on source samples. Here we choose the commonly used MSE loss for optimization of our networks. The optimization objective is given below:

$$\min_{G_y, G_f} \mathcal{L}\left(X_s, Y_s, G_y\right) \tag{4}$$

where:

$$\mathcal{L}(X_s, Y_s, G_y) = \frac{1}{n_s} \sum_{i=1}^{n_s} \left(y_i^s - G_y\left(G_f(x_i^s)\right)\right)^2 \tag{5}$$

2) Step 2: In this step, we first fix the parameters of G_f and G_y, and then update the adversarial regressor G_{adv}. G_{adv} is trained to maximize the discrepancy loss to approximate the empirical discrepancy distance. Given that the empirical discrepancy distance is defined as the supremum over the whole hypothesis space $\mathcal{H}$ according to (2), some bad hypotheses in $\mathcal{H}$ may be chosen during the optimization, which are irrelevant with learning tasks [37], [38]. To alleviate this phenomenon, we add the MSE loss for G_{adv} over the source samples. Therefore, a better G_{adv} will be chosen under the localized hypothesis space $\mathcal{H}$, as pointed in [38]. Additionally, we use the same number of source and target samples to update R_{adv} in every iteration. The optimization objective is given as follows:

$$\min_{G_{adv}} \mathcal{L}\left(X_s, Y_s, G_{adv}\right) - \mathcal{L}_{disc}\left(G_{adv}, G_y\right) \tag{6}$$

where:

$$\mathcal{L}(X_s, Y_s, G_{adv}) = \frac{1}{n_s} \sum_{i=1}^{n_s} \left(y_i^s - G_{adv}\left(G_f(x_i^s)\right)\right)^2 \tag{7}$$

$$\mathcal{L}_{disc}\left(G_{adv}, G_y\right) = \left|\mathcal{L}_{D_s}\left(G_{adv}, G_y\right) - \mathcal{L}_{D_t}\left(G_{adv}, G_y\right)\right| \tag{8}$$

$$\mathcal{L}_{D_s} = \frac{1}{n_s} \sum_{i=1}^{n_s} \left(G_{adv}\left(G_f(x_i^s)\right) - G_y\left(G_f(x_i^s)\right)\right)^2 \tag{9}$$

$$\mathcal{L}_{D_t} = \frac{1}{n_t} \sum_{i=1}^{n_t} \left(G_{adv}\left(G_f(x_i^t)\right) - G_y\left(G_f(x_i^t)\right)\right)^2 \tag{10}$$

3) Step 3: After the previous training step, we fix the parameters of the two regressors G_y, G_{adv}. Then the feature extractor G_f is trained to minimize the discrepancy loss for generating domain-invariant feature representations. The optimization objective is given below:

$$\min_{G_f} \mathcal{L}_{disc}\left(G_{adv}, G_y\right) \tag{11}$$

The above three training steps are repeated in our learning framework using mini-batch gradient descent, and their order is not important. The complete procedure of our proposed approach is shown in Algorithm 1.

Algorithm 1: Our Proposed approach

Input: Labeled source samples $\{X_s, Y_s\}$;
Unlabeled target samples $\{X_t\}$;
Batch size N; Max iteration M.
Output: Feature extractor G_f;
Task-specific regressor G_y;
Adversarial regressor G_{adv}.

G_y and G_{adv} have the same network architecture and they are initialized differently.

while *training* **do**
  for $j = 1$ *to* M **do**
    Sample a batch $\{(\mathbf{x_i^s}, y_i^s)\}_{i=1}^N$ from $\{X_s, Y_s\}$;
    Sample a batch $\{(\mathbf{x_i^t})\}_{i=1}^N$ from $\{X_t\}$;
    Update G_y, G_f in (4) using gradient descent on the batch;
    Update G_{adv} in (6) using gradient descent on the batch;
    Update G_f in (11) using gradient descent on the batch;
  end
end

IV. Experiments

This section presents the experiment results of our proposed approach in offline EEG-based driver drowsiness estimation.

A. The SEED-VIG Dataset

SEED-VIG is a publicly available drowsiness estimation dataset[1] [28] and is used to evaluate the performance of our proposed approach in this paper. According to [28], driving experiments were developed based on a simulated reality-based driving system to collect EEG data. This system consists of a large LCD screen, a real vehicle and a software controller. During the experiments, a four-lane highway scene was shown on the large LCD screen in front of the real vehicle, and subjects were required to drive in the vehicle. Subjects'operations, including steering, throttle controlling and braking, were monitored by the software.

There were 23 subjects (average age is 23.3 years old, 12 females) participating in the experiments in total. They all had normal or corrected-to-normal vision. To induce driver fatigue more easily, all experiments were conducted during early afternoon or late night. All subjects were required to drive for two hours and their EEG signals were recorded using Neuroscan system with a sampling rate of 1000 Hz. Eye tracking glasses were used to obtain the percentage of eye closure (PERCLOSS) [39] and the data was labeled based on PERCLOSS, ranging from 0 (low drowsiness level) to 1 (high drowsiness level). In this paper, we use the EEG signals recorded from the 11-channel posterior site and the 6-channel temporal site for drowsiness estimation.

B. Preprocessing and Feature Extraction

The raw EEG signals from temporal and posterior sites were first filtered by a 1-75 Hz band-pass filter to remove artifacts and noise, and then downsampled to 200 Hz to reduce the computational complexity. Then the signals were segmented by an 8-second non-overlapping time window.

For feature extraction, we computed the average power spectral density (PSD) in five frequency bands: delta (1-4 Hz), theta (4-8 Hz), alpha (8-14 Hz), beta (14-31 Hz), and gamma (31-50 Hz). We adopted Welch's method [40] with Hamming window, 1024 points fast Fourier transform, and 50% overlapping. The PSDs were then converted into dBs and used as our features. Each feature vector had 17×5= 85 dimensions.

C. Performance Metrics and Baseline Approaches

Since we considered the regression problem in this paper, the following two metrics were used to evaluate the estimation results: root mean squared error (RMSE) and Pearson correlation coefficient (CC), which measure the average error and structural correlation between the estimation values and ground truth respectively.

Three algorithms were chosen as baseline approaches and compared with the performance of our proposed approach. We introduce the baseline approaches briefly below.

[1] http://bcmi.sjtu.edu.cn/∼seed/downloads.html

1) Deep Neural Network (DNN): DNN was the basic network in the experiments without any DA techniques. In this paper, DNN was trained only using labeled source samples based on MSE loss, without any information from target domains.

2) Domain Adversarial Neural Network (DANN): DANN was a typical deep adversarial DA approach proposed in [24]. It consists of three components: a feature extractor G_f, a task-specific predictor G_y, and a domain discriminator G_d. G_f and G_d are both trained in an adversarial manner with the domain classification loss. After the adversarial training, domain-invariant feature representations are generated from G_f and then used for training G_y.

The original DANN approach was designed for classification. We modified it for regression by replacing the softmax layer of G_y with a fully connected regression layer with linear activations, and used MSE instead of cross-entropy as the task loss function.

3) Adversarial Discriminative Domain Adaptation (ADDA): ADDA was another typical deep adversarial DA approach proposed in [26], under the basic framework of DANN. In ADDA, two feature extractors G_{fs} and G_{ft} and two task-specific predictor G_{ys} and G_{yt} are separately trained for source and target domains. The training of G_{fs} and G_{ys} relies only on labeled source samples, while G_{ft} and the domain discriminator G_d are optimized with the GAN-based adversarial loss in an adversarial manner. G_{yt} is initialized from the trained G_{ys} for the final prediction.

The original DANN approach was also designed for classification. We generalized G_{yt} and G_{ys} to regression with the same operations that we did for DANN.

D. Evaluation Process and Results

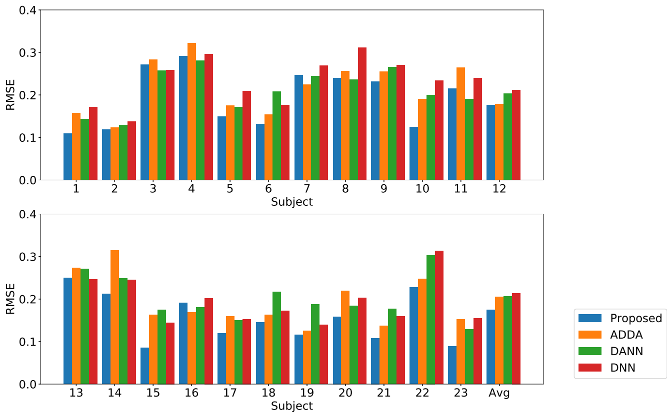
We used leave-one-subject-out cross-validation to evaluate the performance of our proposed approach. In each run, each subject was used as the target domain once and the rest of the subjects were treated as source domains.

All the approaches were trained using mini-batch gradient descent with Adam optimizer, which used batch-size of 128, learning rate of 2×10^{-4}, and weight decay of 5×10^{-5}. We sampled 20% data from the shuffled source samples as validation data in early-stopping to alleviate overfitting. The maximum number of training epochs was set as 200, and the early-stopping patience was 5 epochs. We repeated all the approaches five times for each target subject with the same experiment settings and reported the average performance.

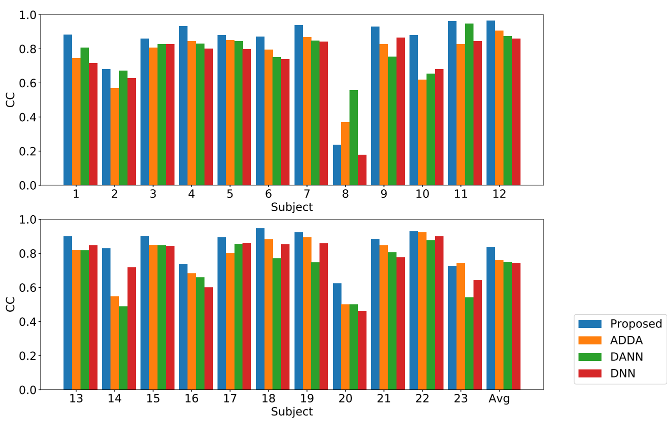
The estimation performance across the 23 subjects is shown in Fig. 2 and the average values are given in Tabel I. It can be observed that:

1) Our proposed approach outpeformed DNN, DANN and ADDA on most of subjects.
2) In traditional deep adversarial DA baseline approaches, DANN and ADDA only performed slightly better than DNN. It is because that they are originally designed for classification and cannot be directly extended to regression for the lack of theoretical support.

3) Our proposed approach improved the RMSE and the CC by 13.78% and 12.70% respectively over DNN, which verified the effectiveness of the discrepancy distance in DA for regression.



(a)



(b)

Fig. 2. Experiment results (a) RMSEs and (b) CCs across the 23 subjects.

TABLE I
AVERAGE RMSES AND CCS ACROSS THE 23 SUBJECTS.

	DNN	DANN	ADDA	Proposed Approach
RMSE	0.2140	0.2067	0.2048	**0.1845**
CC	0.7449	0.7506	0.7617	**0.8395**

V. Conclusion

Accurate driver vigilance estimation is important to safety. In this paper, we consider a more meaningful scenario: drowsiness estimations are provided as continuous values as a standard regression problem. DA serves as a promising approach that can save efforts for new subject calibration in EEG-based BCIs. Therefore, we proposed a novel deep adversarial DA approach based on the discrepancy distance. The discrepancy distance can be used for measuring the distribution discrepancy in regression. Additionally, our proposed approach requires no labeled subject-specific calibration data of new subjects for offline cross-subject drowsiness estimation. Experiment results on a pubic dataset demonstrated the effectiveness of our proposed approach and that the discrepancy distance is useful for regression problems in DA.

Acknowledgement

This research was supported by the IEEE Computational Intelligence Society Graduate Student Research Grant.

References

[1] G. Zhang, K. K. Yau, X. Zhang, and Y. Li, "Traffic accidents involving fatigue driving and their extent of casualties," *Accident Analysis & Prevention*, vol. 87, pp. 34–42, 2016.

[2] F. Sagberg, P. Jackson, H.-P. Krüger, A. Muzet, and A. Williams, *Fatigue, sleepiness and reduced alertness as risk factors in driving*. Institute of Transport Economics Oslo, 2004.

[3] N. H. T. S. Administration, "Drowsy driving 2015," *NHTSA's National Center for Statistics and Analysis: Washington, DC, USA*, 2017.

[4] K. Kozak, J. Pohl, W. Birk, J. Greenberg, B. Artz, M. Blommer, L. Cathey, and R. Curry, "Evaluation of lane departure warnings for drowsy drivers," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 22, pp. 2400–2404, 2006.

[5] Y. Cui, Y. Xu, and D. Wu, "EEG-based driver drowsiness estimation using feature weighted episodic training," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 11, pp. 2263–2273, 2019.

[6] B. Reddy, Y.-H. Kim, S. Yun, C. Seo, and J. Jang, "Real-time driver drowsiness detection for embedded system using model compression of deep neural networks," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, Honolulu, HI, Jul. 2017, pp. 121–128.

[7] Z. Gao, X. Wang, Y. Yang, C. Mu, Q. Cai, W. Dang, and S. Zuo, "EEG-based spatio-temporal convolutional neural network for driver fatigue evaluation," *IEEE Trans. on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2755–2763, 2019.

[8] R. Fu and H. Wang, "Detection of driving fatigue by using noncontact EMG and ECG signals measurement system," *International Journal of Neural Systems*, vol. 24, no. 03, p. 1450006, 2014.

[9] B.-Q. Ma, H. Li, Y. Luo, and B.-L. Lu, "Depersonalized cross-subject vigilance estimation with adversarial domain generalization," in *Proc. Int'l Joint Conf. on Neural Networks (IJCNN)*, Budapest, Hungary, Jul. 2019, pp. 1–8.

[10] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: a general-purpose brain-computer interface (BCI) system," *IEEE Trans. on Biomedical Engineering*, vol. 51, no. 6, pp. 1034–1043, 2004.

[11] D. Wu, V. J. Lawhern, S. Gordon, B. J. Lance, and C.-T. Lin, "Offline EEG-based driver drowsiness estimation using enhanced batch-mode active learning (EBMAL) for regression," in *Proc. IEEE Int'l Conf. on Systems, Man, and Cybernetics (SMC)*, Budapest, Hungary, Oct. 2016, pp. 730–736.

[12] Y. Cui and D. Wu, "EEG-based driver drowsiness estimation using convolutional neural networks," in *Proc. Int'l Conf. on Neural Information Processing*, Guangzhou, China, Nov. 2017, pp. 822–832.

[13] D. Wu, V. J. Lawhern, S. Gordon, B. J. Lance, and C.-T. Lin, "Driver drowsiness estimation from EEG signals using online weighted adaptation regularization for regression (OwARR)," *IEEE Trans. on Fuzzy Systems*, vol. 25, no. 6, pp. 1522–1535, 2016.

[14] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.

[15] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *IEEE Trans. on Affective Computing*, vol. 5, no. 3, pp. 327–339, 2014.

[16] P. Arico, G. Borghini, G. Di Flumeri, N. Sciaraffa, A. Colosimo, and F. Babiloni, "Passive BCI in operational environments: insights, recent advances, and future trends," *IEEE Trans. on Biomedical Engineering*, vol. 64, no. 7, pp. 1431–1436, 2017.

[17] B. Bhowmick and K. C. Kumar, "Detection and classification of eye state in IR camera for driver drowsiness identification," in *Proc. IEEE Int'l Conf. on Signal and Image Processing Applications*, Kuala Lumpur, Malaysia, Nov. 2009, pp. 340–345.

[18] H. Li, W.-L. Zheng, and B.-L. Lu, "Multimodal vigilance estimation with adversarial domain adaptation networks," in *Proc. Int'l Joint Conf. on Neural Networks (IJCNN)*, Rio de Janeiro, Brazil, Jul. 2018, pp. 1–6.

[19] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update," *Journal of Neural Engineering*, vol. 15, no. 3, p. 031005, 2018.

[20] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2010.

[21] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proc. of the IEEE Int'l Conf. on Computer Vision*, Sydney, Australia, Dec. 2013, pp. 2200–2207.

[22] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional adversarial domain adaptation," in *Proc. Advances in Neural Information Processing Systems*, Montréal, Canada, Dec. 2018, pp. 1640–1650.

[23] M. Wang and W. Deng, "Deep visual domain adaptation: A survey," *Neurocomputing*, vol. 312, pp. 135–153, 2018.

[24] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2096–2030, 2016.

[25] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Advances in Neural Information Processing Systems*, Quebec, Canada, Dec. 2014, pp. 2672–2680.

[26] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Honolulu, HI, Jul. 2017, pp. 7167–7176.

[27] C. Cortes and M. Mohri, "Domain adaptation in regression," in *Proc. Int'l Conf. on Algorithmic Learning Theory*, Espoo, Finland, Oct. 2011, pp. 308–323.

[28] W.-L. Zheng and B.-L. Lu, "A multimodal approach to estimating vigilance using EEG and forehead EOG," *Journal of Neural Engineering*, vol. 14, no. 2, p. 026017, 2017.

[29] C.-T. Lin, K.-C. Huang, C.-H. Chuang, L.-W. Ko, and T.-P. Jung, "Can arousing feedback rectify lapses in driving? Prediction from EEG power spectra," *Journal of Neural Engineering*, vol. 10, no. 5, p. 056024, 2013.

[30] C.-T. Lin, C.-H. Chuang, C.-S. Huang, S.-F. Tsai, S.-W. Lu, Y.-H. Chen, and L.-W. Ko, "Wireless and wearable EEG system for evaluating driver vigilance," *IEEE Trans. on Biomedical Circuits and Systems*, vol. 8, no. 2, pp. 165–176, 2014.

[31] P. R. Davidson, R. D. Jones, and M. T. Peiris, "EEG-based lapse detection with high temporal resolution," *IEEE Trans. on Biomedical Engineering*, vol. 54, no. 5, pp. 832–839, 2007.

[32] C.-T. Lin, C.-H. Chuang, Y.-K. Wang, S.-F. Tsai, T.-C. Chiu, and L.-W. Ko, "Neurocognitive characteristics of the driver: A review on drowsiness, distraction, navigation, and motion sickness," *Journal of Neuroscience and Neuroengineering*, vol. 1, no. 1, pp. 61–81, 2012.

[33] R. N. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, "Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm," *IEEE Trans. on Biomedical Engineering*, vol. 58, no. 1, pp. 121–131, 2010.

[34] D. Wu, C.-H. Chuang, and C.-T. Lin, "Online driver's drowsiness estimation using domain adaptation with model fusion," in *Proc. Int'l Conf. on Affective Computing and Intelligent Interaction (ACII)*, Xi'an, China, Sep. 2015, pp. 904–910.

[35] K. Saito, K. Watanabe, Y. Ushiku, and T. Harada, "Maximum classifier discrepancy for unsupervised domain adaptation," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, Jun. 2018, pp. 3723–3732.

[36] S. Li, C. H. Liu, B. Xie, L. Su, Z. Ding, and G. Huang, "Joint adversarial domain adaptation," in *Proc. of the 27th ACM Int'l Conf. on Multimedia*, New York, NY, Oct. 2019, pp. 729–737.

[37] Y. Zhang, T. Liu, M. Long, and M. Jordan, "Bridging theory and algorithm for domain adaptation," in *Proc. of the 36th Int'l Conf. on Machine Learning*, vol. 97, Long Beach, CA, Jun. 2019, pp. 7404–7413.

[38] Y. Zhang, M. Long, J. Wang, and M. I. Jordan, "On localized discrepancy for domain adaptation," *arXiv preprint arXiv:2008.06242*, 2020.

[39] X.-Y. Gao, Y.-F. Zhang, W.-L. Zheng, and B.-L. Lu, "Evaluating driving fatigue detection algorithms using eye tracking glasses," in *Proc. 7th Int'l IEEE/EMBS Conf. on Neural Engineering (NER)*, Montpellier, France, Apr. 2015, pp. 767–770.

[40] 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 Trans. on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.