Online Driver's Drowsiness Estimation Using Domain Adaptation with Model Fusion

Dongrui Wu
DataNova
NY USA
Email: drwu09@gmail.com

Chun-Hsiang Chuang
Institute of Electrical Control Engineering
National Chiao-Tung University, Taiwan
Email: cch.chuang@gmail.com

Chin-Teng Lin
National Chiao-Tung University, Taiwan
University of Technology, Australia
Email: ctlin@mail.nctu.edu.tw

Abstract—**Drowsy driving is a pervasive problem among drivers, and is also an important contributor to motor vehicle accidents. It is very important to be able to estimate a driver's drowsiness level online so that preventative actions could be taken to avoid accidents. However, because of large individual differences, it is very challenging to design an estimation algorithm whose parameters fit all subjects. Some subject-specific calibration data must be used to tailor the algorithm for each new subject. This paper proposes a domain adaptation with model fusion (DAMF) online drowsiness estimation approach using EEG signals. By making use of EEG data from other subjects in a transfer learning framework, DAMF requires very little subject-specific calibration data, which significantly increases its utility in practice. We demonstrate using a simulated driving experiment and 15 subjects that DAMF can achieve much better performance than several other approaches.**

Keywords—*Drowsy driving; EEG; domain adaptation; model fusion; transfer learning*

I. INTRODUCTION

Drowsy driving is a pervasive problem among drivers. A nationally representative telephone survey of 4,010 drivers conducted by the National Highway Traffic Safety Administration (NHTSA) in 2002 [17] found that 37% of drivers had "*nodded off for at least a moment or fallen asleep while driving at least once in their driving career*", including 4% within the past month, 8% within the past six months, and 11% within the past year. AAA Foundation for Traffic Safety conducted a telephone survey to 2,000 drivers in 2010 [18], and reported similar findings: 41.0% of drivers had ever fallen asleep or nodded off while driving, including 3.9% within the past month, 7.1% within the past six months, and 11.0% within the past year. The National Sleep Foundation's 2005 Sleep in America poll [15] showed that drowsy driving may be much more pervasive: 60% of adult drivers said they had driven a vehicle while feeling drowsy in the past year, 37% had actually fallen asleep at the wheel, and 13% said they had done so at least once a month.

As the driver's cognition and reaction abilities are significantly impaired while drowsy, drowsy driving becomes an important cause of motor vehicle accidents. According to the NHTSA [19], between 2005 and 2009, 2.5% of fatal motor vehicle crashes (on average 886 per year in the U.S.) and 2.5% of fatalities (on average 1,004 per year in the U.S.) involved drowsy driving. The National Sleep Foundation poll

[15] showed that 4% of drivers admitted they had had an accident or near accident because they dozed off or were too tired to drive. Some modeling studies [13], [18] even estimated that 15%-33% of fatal crashes might involve drowsy drivers.

As a result, it is very important to be able to recognize a driver's drowsiness level, and allow preventative actions (e.g., 1750Hz tone-burst to arouse the driver from drowsiness [20]) to be taken to avoid accidents. As summarized in [20], generally there are two approaches in the literature and also applications: 1) computer vision based approach [1], [5], which uses cameras to monitor the driver's eye, face and/or nodding activities and then to infer the drowsiness level; and, 2) driving behavior based approach [8], [20], which may use response time (RT), the time interval between the onset of lane deviation and the driver's first response, as an indicator of drowsiness level. RT can be inferred from the driver's EEG signals using a brain-computer interface (BCI) system. The two approaches are complementary to each other. As pointed out in [14], "*neural signals are just another signal modality that can supplement video or voice analysis: less dependent on overt behavior, less susceptible to deception, but requiring more intrusive sensors.*" This paper adopts the second approach.

Although there has been considerable literature on drowsiness classification/estimation from EEG signals [7]–[9], [20], [21], many studies focused on offline recognition, e.g., feature extraction used information from all EEG data. However, a practical drowsiness recognition system needs to recognize a driver's drowsiness level online, in real-time. Additionally, because of large individual differences, it is very challenging, if not impossible, to design a drowsiness recognition algorithm whose parameters fit all subjects. Some subject-specific calibration data must be used to tailor the algorithm for a new subject. To increase the utility of the BCI system, we need to reduce the amount of subject-specific calibration data. Transfer learning (TL) [16], which makes use of data from other existing subjects, can be used for this purpose.

There has not been much work on TL for online drowsiness estimation. To our best knowledge, the only one is [21], which showed that selective TL (STL), which selectively turns TL on or off based the level of session generalizability (more details will be given in Section III-D), can achieve better estimation performance than approaches that always turns TL

on or off. However, it used a long calibration session, and did not show how the learning performance changed with the amount of subject-specific calibration data. This paper proposes a domain adaptation with model fusion (DAMF) online drowsiness estimation approach. By making use of data from other subjects in a TL framework, DAMF requires very little subject-specific calibration data, which significantly increases its ease of use in practice. We also show that DAMF outperforms the TL and STL approaches in [21].

The rest of the paper is organized as follows: Section II introduces the details of the online DAMF algorithm. Section III presents experimental results and performance comparisons of different algorithms. Section IV draws conclusions.

II. ONLINE DOMAIN ADAPTATION WITH MODEL FUSION

This section introduces the online DAMF algorithm. Assume there are data from Z existing (auxiliary) subjects, which could be used to help the learning for a new subject. We apply the online domain adaptation (DA) algorithm for each auxiliary subject separately to obtain Z different models, and then fuse them for a final model.

A. Background

A domain [10], [16] $\mathcal{D}$ in TL consists of a d-dimensional feature space $\mathcal{X}$ and a marginal probability distribution $P(\mathbf{x})$, i.e., $\mathcal{D} = \{\mathcal{X}, P(\mathbf{x})\}$, where $\mathbf{x} \in \mathcal{X}$. Two domains $\mathcal{D}^z$ and $\mathcal{D}^t$ are different means $\mathcal{X}^z \neq \mathcal{X}^t$, and/or $P^z(\mathbf{x}) \neq P^t(\mathbf{x})$.

A task [10], [16] $\mathcal{T}$ in TL consists of a output space $\mathcal{Y}$ and a conditional probability distribution $Q(y|\mathbf{x})$. Two tasks $\mathcal{T}^z$ and $\mathcal{T}^t$ are different means $\mathcal{Y}^z \neq \mathcal{Y}^t$, or $Q^z(y|\mathbf{x}) \neq Q^t(y|\mathbf{x})$.

Given the zth source domain $\mathcal{D}^z$ with n_z samples $\{(\mathbf{x}_i^z, y_i^z)\}_{i=1,\dots,n_z}$, and a target domain $\mathcal{D}^t$ with m calibration samples $\{(\mathbf{x}_j^t, y_j^t)\}_{j=1,\dots,m}$, DA aims to learn a target prediction function $f : \mathbf{x}^t \mapsto y^t$ with low expected error on $\mathcal{D}^t$, under the assumptions $\mathcal{X}^z = \mathcal{X}^t$, $\mathcal{Y}^z = \mathcal{Y}^t$, $P^z(\mathbf{x}) \neq P^t(\mathbf{x})$, and $Q^z(y|\mathbf{x}) \neq Q^t(y|\mathbf{x})$.

For example, in driver's drowsiness estimation from EEG signals, EEG signals from a new subject are in the target domain, while EEG signals from the zth existing subject are in the zth source domain. A single data sample would consist of the feature vector for a single EEG epoch in either domain. Though the features in source and target domains are computed in the same way, generally their marginal and conditional probability distributions are different, i.e., $P^z(\mathbf{x}) \neq P^t(\mathbf{x})$ and $Q^z(y|\mathbf{x}) \neq Q^t(y|\mathbf{x})$, because the two subjects usually have different drowsy neural responses. As a result, data from a source domain cannot represent data in the target domain accurately, and must be integrated with some data in the target domain to induce the target predictive function.

B. DA Using Ridge Regression

Given a feature matrix $\mathbf{X}$ (each row represents a different sample and each column a different feature) and its corre-

sponding output vector $\mathbf{y}$, ridge regression (RR) tries to find a coefficient vector $\boldsymbol{\beta}_{RR}$ such that [6]:

$$\boldsymbol{\beta}_{RR} = \underset{\boldsymbol{\beta}}{\arg\min} \, ||\mathbf{y} - \mathbf{X}\boldsymbol{\beta}||_2^2 + \lambda||\boldsymbol{\beta}||_2^2 \qquad (1)$$

where $\lambda > 0$ is the *ridge parameter*, which has primarily two effects:

1) It improves the conditioning of the problem. This is important because at the beginning of the calibration the number of samples may be smaller than the number of features, and hence regular least squares regression is ill-conditioned, whereas RR can always find a solution.

2) It reduces the variance of the estimates. Although RR usually gives more biased estimates than regular least squares regression, the reduced variance of ridge estimates often result in a smaller root mean squared error (RMSE).

The solution of (1) is:

$$\boldsymbol{\beta}_{RR} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y} \qquad (2)$$

where $\mathbf{X}^T$ is the transpose of $\mathbf{X}$, and $\mathbf{I}$ is the identity matrix.

To use RR in DA, we simply combine features and outputs in the zth source domain and the target domain, i.e.,

$$\mathbf{X}^z = \begin{bmatrix} \mathbf{x}_1^z \\ \vdots \\ \mathbf{x}_{n_z}^z \\ \mathbf{x}_1^t \\ \vdots \\ \mathbf{x}_m^t \end{bmatrix}, \quad \mathbf{y}^z = \begin{bmatrix} \mathbf{y}_1^z \\ \vdots \\ \mathbf{y}_{n_z}^z \\ \mathbf{y}_1^t \\ \vdots \\ \mathbf{y}_m^t \end{bmatrix} \qquad (3)$$

and then use them in (2) to compute $\boldsymbol{\beta}^z$. Note that more sophisticated DA methods [16] could be used; however, we chose the above method for its simplicity and speed, and also because a similar method has demonstrated superior performance in a previous work on classification [23].

C. Model Fusion

After all Z $\boldsymbol{\beta}^z$ have been computed, they can be aggregated by a simple average[1] to give a final RR model, whose coefficient vector is computed as:

$$\boldsymbol{\beta} = \frac{1}{Z} \sum_{z=1}^{Z} \boldsymbol{\beta}^z. \qquad (4)$$

The pseudo-code for the complete online DAMF algorithm is shown in Algorithm 1. It is simple and fast; however, as we will demonstrate in the next section, it can achieve superior performance.

III. EXPERIMENTS AND DISCUSSIONS

Experimental results on simulated driving data are presented in this section to demonstrate the performance of DAMF.

---

[1] We have also investigated more sophisticated aggregation approaches, e.g., giving more weights to models performing better in training, but failed to observe improved performance.

Algorithm 1: The online DAMF algorithm.

Input: Z source domains, where the zth ($z = 1, ..., Z$) domain has n_z samples $\{\mathbf{x}_i^z, y_i^z\}_{i=1,...,n_z}$; m target domain calibration samples, $\{\mathbf{x}_j^t, y_j^t\}_{j=1,...,m}$; Parameters λ in ridge regression.

Output: The DAMF RR model.

for $z = 1, 2, ..., Z$ **do**

    Construct $\mathbf{X}^z$ and $\mathbf{y}^z$ in (3) ;

    Compute $\boldsymbol{\beta}^z$ by (2);

    Compute $\boldsymbol{\beta}$ by (4) ;

end

Return The DAMF RR model with coefficient vector $\boldsymbol{\beta}$.
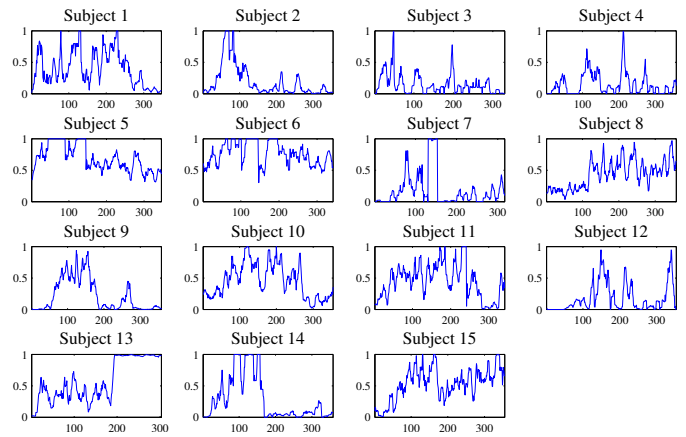
Fig. 1. Drowsiness indices of the 15 subjects.

A. Experiment Setup

This study recruited 16 healthy subjects with normal or corrected to normal vision to participant in a sustained-attention driving experiment [2], [3], which consists of a real vehicle mounted on a motion platform with 6 degrees of freedom immersed in a 360-degree virtual-reality (VR) scene. Each participant read and signed an informed consent form before the experiment began. Each experiment lasted for about 60-90 minutes and was conducted in the afternoon when the circadian rhythm of sleepiness reached its peak. To induce drowsiness in the subjects during driving, the VR scenes simulated monotonous driving at a fixed speed (100 km/h) on a straight and empty highway. During the experiment, lane-departure events randomly appeared every 5-10 seconds, and participants were instructed to steer the vehicle to compensate for these perturbations immediately. Subjects' cognitive states and driving performance were monitored via a surveillance video camera and the vehicle trajectory throughout the experiment. The RT in response to the perturbation was recorded and later converted to drowsiness index. Meanwhile, participants' scalp EEG signals were recorded using a 32-channel Neuroscan system (30-channel EEGs plus 2-channel earlobes) with a sampling rate of 500Hz. We would verify whether the long-RT trial was accompanied with drowsiness by observing slow eye blinks/movements via EEG recordings.

The Institutional Review Board of the Taipei Veterans General Hospital approved the experimental protocol.

B. Preprocessing and Feature Extraction

We used EEGLAB [4] for EEG signal preprocessing. All 30 EEG channels were used in feature extraction. A band-pass filter (1-50 Hz) was applied to remove high-frequency muscle artifacts, line-noise contamination and DC drift. Next the EEG data were downsampled from 500 Hz to 250 Hz and re-referenced to averaged earlobes.

We then defined a function to map the response time τ to a drowsiness index $y \in [0, 1]$:

$$y = \max\left\{0, \frac{1 - e^{-(\tau - \tau_0)}}{1 + e^{-(\tau - \tau_0)}}\right\} \qquad (5)$$

which was first proposed in [21]. $\tau_0 = 1$ was used in this paper.

The 16 subjects had different lengths of experiment, because the disturbances were presented randomly every 5-10 seconds. Data from one subject was not correctly recorded, so we used only 15 subjects. To ensure fair comparison, we used only the first 3,600 seconds data for each subject and sampled them every 10 seconds. The drowsiness indices were then smoothed using a 90-second square moving-average window to reduce variations. This does not reduce the sensitivity of the drowsiness index because the cycle lengths of drowsiness fluctuations are longer than 4 minutes [11]. The smoothed drowsiness indices for the 15 subjects are shown in Fig. 1. Each subject has some drowsiness indices at or close to 1, indicating drowsy driving.

We then epoched 30-second EEG signals right before each sample, and computed the average power spectral density (PSD) in the theta band (4-7 Hz) for each channel using Welch's method [22], as research [12] has shown that theta band spectrum is a strong indicator of drowsiness. Finally, we converted the 30 theta band powers to dBs and used them as our features. To remove noises or bad channel readings, we removed channels whose maximal PSDs are larger than 20. The theta band powers for five selected channels and the corresponding drowsiness index for Subject 1 are shown in Fig. 2. Observe that drowsiness index has strong correlation with the theta band powers.

Fig. 3 shows the mean topoplots of theta band powers across the 15 subjects. To obtain the mean topoplot for the alert state in Fig. 3(a), we first find the mean theta band powers of the 5% smallest drowsiness indices for each of the 15 subjects, and then take their average. To obtain the mean topoplot for the drowsy state in Fig. 3(b), we first find the mean theta band powers of the 5% largest drowsiness indices for each of the 15 subjects, and then take their average. Fig. 3(c) shows the difference between the drowsy state and the alert state. Observe that there is large differences in the topoplots between these two states, which suggests that the theta band power is
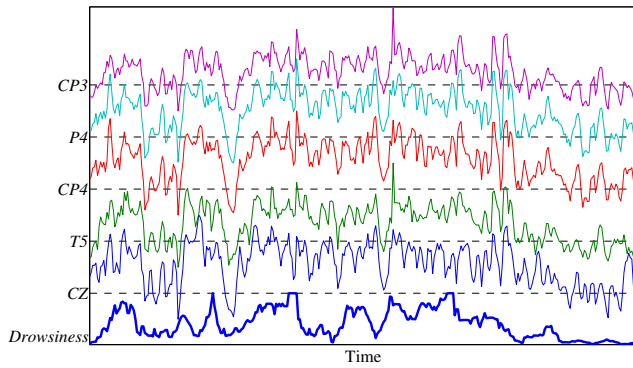
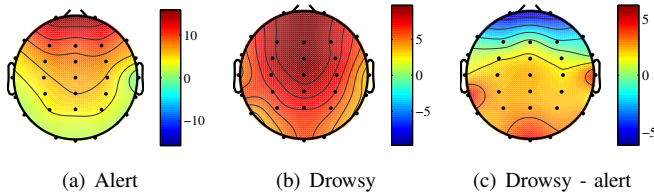Fig. 2. Theta band powers and drowsiness index for Subject 1.

(a) Alert   (b) Drowsy   (c) Drowsy - alert

Fig. 3. Mean topoplots of theta band powers across the 15 subjects.

a good indicator of drowsiness.

C. Evaluation Process and Performance Measures

Among the ∼360 samples for each subject, we reserved 100 samples in a randomly chosen continuous block for calibration (training) and the rest ∼260 samples for testing. Next we use DAMF to illustrate the evaluation process, as shown in Fig. 4.

We start from zero subject-specific calibration data, use DAMF to train a model, evaluate the performance of this model on the ∼260 testing samples[2], acquire and add 5 more subject-specific calibration samples sequentially, re-train the DAMF model, and repeat the process until all 100 calibration samples have been used. We ran this iterative evaluation process 30 times, each time with a randomly chosen 100-sample calibration block, to obtain statistically meaningful results. Finally, we repeated this entire process 15 times so that each subject had the chance to be the "new" subject (target domain) and the remaining 14 to be "existing" subjects (source domains).

The primary performance measured used in this paper is RMSE, which is directly optimized in the object function of RR. The secondary performance measure is the correlation coefficient (CC) between the estimates of the ∼260 testing samples and the true drowsiness indices.

D. Algorithms

We compared the performances of DAMF with five other algorithms:

---

[2]This performance is not available in practice in the online calibration process because the testing data are not available. We evaluate the testing performance in each iteration just to illustrate how it changes with the number of subject-specific calibration samples.
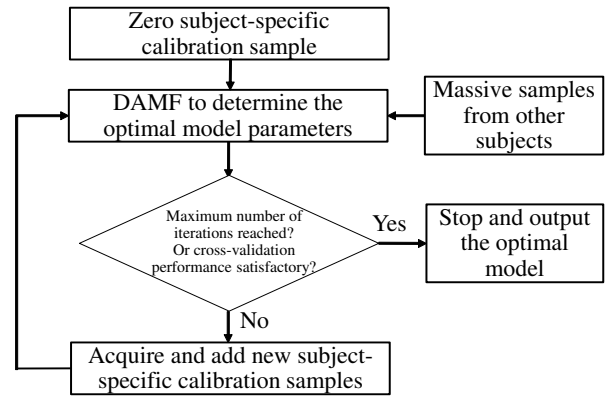

Fig. 4. The iterative online DA algorithm.

1) Baseline 1 (BL1), which combines data from all 14 existing subjects, builds a RR model, and applies it to the new subject. That is, it tries to build a subject-independent model and ignores data from the new subject completely.
2) Baseline 2 (BL2), which builds a RR model using only subject-specific calibration samples from the new subject. That is, it ignores data from existing subjects completely.
3) DAall, which is DAMF without model fusion: in each iteration it combines data from all 14 existing subjects and treats them as data from a single source domain, and builds a DA RR model by combining them with the subject-specific calibration data.
4) TL, which is the transfer learning algorithm in [21]. It builds 14 RR models using data from each auxiliary subject separately, and then the 15th model using the subject-specific data. The final model is the average of these 15 models. TL and DAMF look similar, but there are two important differences: i) TL is the average of 15 models, whereas DAMF is the average of 14 models (DAMF does not include a model built from the subject-specific data only); and, ii) each of the first 14 models in TL is built using only data from an auxiliary subject, whereas the corresponding model in DAMF is built by combining the data from the auxiliary subject with the subject-specific data.
5) STL, which is the selective TL approach in [21]. It first computes a level of session generalizability (LSG) for the new subject, which measures how good the model built from the subject-specific data only can be generalized to other subjects. If the LSG is small, e.g., LSG<1, then the subject may be benefited from TL, and hence the above TL approach is adopted. Otherwise, TL is turned off, and only the model built from the subject-specific data is used.

The ridge parameter $\sigma = 0.01$ was used in all six algorithms. Other σs were also tested, and the results were similar as long as σ is not too large (i.e., $\sigma \leq 0.1$).
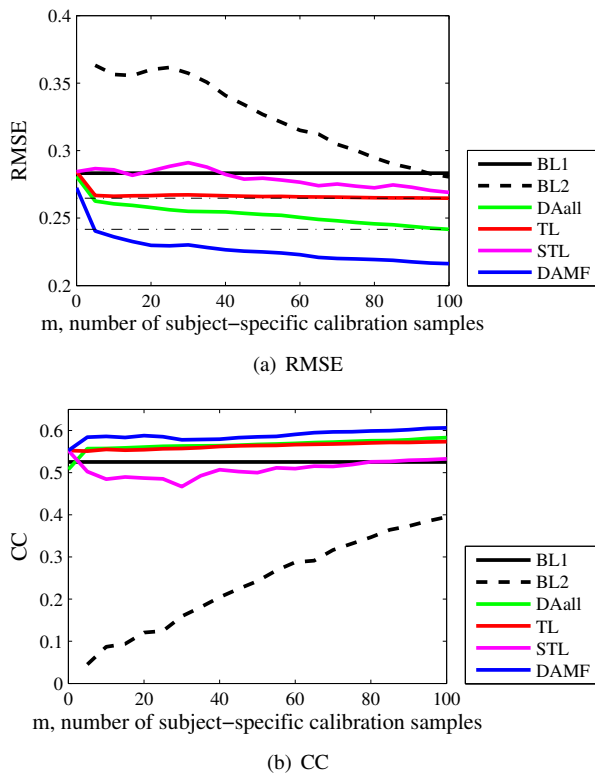
IEEE computer society

(a) RMSE

(b) CC

Fig. 5. Average performances of the five algorithms across the 15 subjects.

E. Experimental Results

The average RMSEs and CCs for the six algorithms across the 15 subjects are shown in Fig. 5, and the RMSEs and CCs for the individual subjects are shown in Fig. 6. Observe that:

1) Except for BL1, whose model does not depend on m, all the other four methods give better models as m increases, which is intuitive.

2) BL1's average RMSE is better than BL2 when m is small, but as m increases, all other models have better RMSEs than BL1, and hence BL1 becomes the worst model. This suggests that there is large individual difference among the subjects, and hence a subject-independent model is not optimal.

3) Because BL2 uses only subject-specific calibration data, it cannot build a model when $m = 0$, i.e., when there is no subject-specific calibration data at all. However, all other five methods can, because they can use data from other subjects. BL2's performance is the worst when m is small, because it cannot get enough training with very few subject-specific calibration samples.

4) On average DAall has better RMSE and CC than BL1 and BL2. This is intuitive, as BL1 does not make use of any subject-specific data, and BL2 does not make use of any auxiliary data. This suggests that even a very simple DA approach can be beneficial.

5) TL achieves better RMSE than BL1, but unlike BL2, DAall, STL and DAMF, its performance does not change

much as m increases. This is because the subject-specific model in TL only has 1/15 weight in the final model, and all other 14 models do not depend on m.

6) STL gives worse RMSE and CC than TL. This is interesting, because [21] showed that STL can result in better performance. There are two reasons: 1) [21] used LSG=1 as the threshold to turn TL on or off, but it may not be optimal for our data; and, 2) [21] used much more subject-specific calibration data, so LSG can be computed more reliably, whereas we have very little calibration data. The results suggest that STL should be tuned for different datasets, and cautions should be paid when the subject-specific calibration dataset is small.

7) On average DAMF gives both the smallest RMSE and the largest CCs, which suggest that overall it is the best algorithm among the six. However, Fig. 6(a) shows that DAMF has worse RMSE than BL1 and DAall for Subjects 1, 2, 9 and 13. This indicates that DAMF still has room for improvement: maybe it is possible to develop a mechanism to switch between DAall and DAMF so that a more appropriate method is chosen according to the characteristics of the new subject, similar to the idea of STL. This is one of our future research directions.

In summary, the four TL or DA based approaches generally have smaller RMSEs than BL2, which do not use auxiliary data. This suggests that TL/DA is indeed beneficial. Moreover, our proposed DAMF achieves best overall performance among the six algorithms.

Finally, we can conclude that given the same amount of subject-specific calibration data, DAMF can achieve much better performance than all the other five approaches. Or, in other words, given a desired RMSE, DAMF needs much less subject-specific calibration data than the other approaches. For example, in Fig. 5(a), the RMSEs for BL2, DAall, TL and STL when $m = 100$ are 0.2806, 0.2416, 0.2647, and 0.2690 respectively, whereas DAMF only needs at most 5 subject-specific calibration samples to achieve them.

IV. CONCLUSIONS

Drowsy driving is a pervasive problem among drivers, and is also an important cause of motor vehicle accidents. As a result, it would be very beneficial to be able to estimate a driver's drowsiness level online so that preventative actions could be taken to avoid accidents. However, because of large individual differences, it is very challenging to design an estimation algorithm whose parameters fit all subjects. Some subject-specific calibration data must be used to tailor the algorithm for each new subject. This paper has proposed an online DAMF drowsiness estimation approach using EEG signals. By making use of EEG data from other subjects in a transfer learning framework, DAMF requires very little subject-specific calibration data, which significantly increases its utility in practice. We demonstrated using a simulated driving experiment and 15 subjects that, given the same amount of subject-specific calibration data, DAMF can achieve much better performance
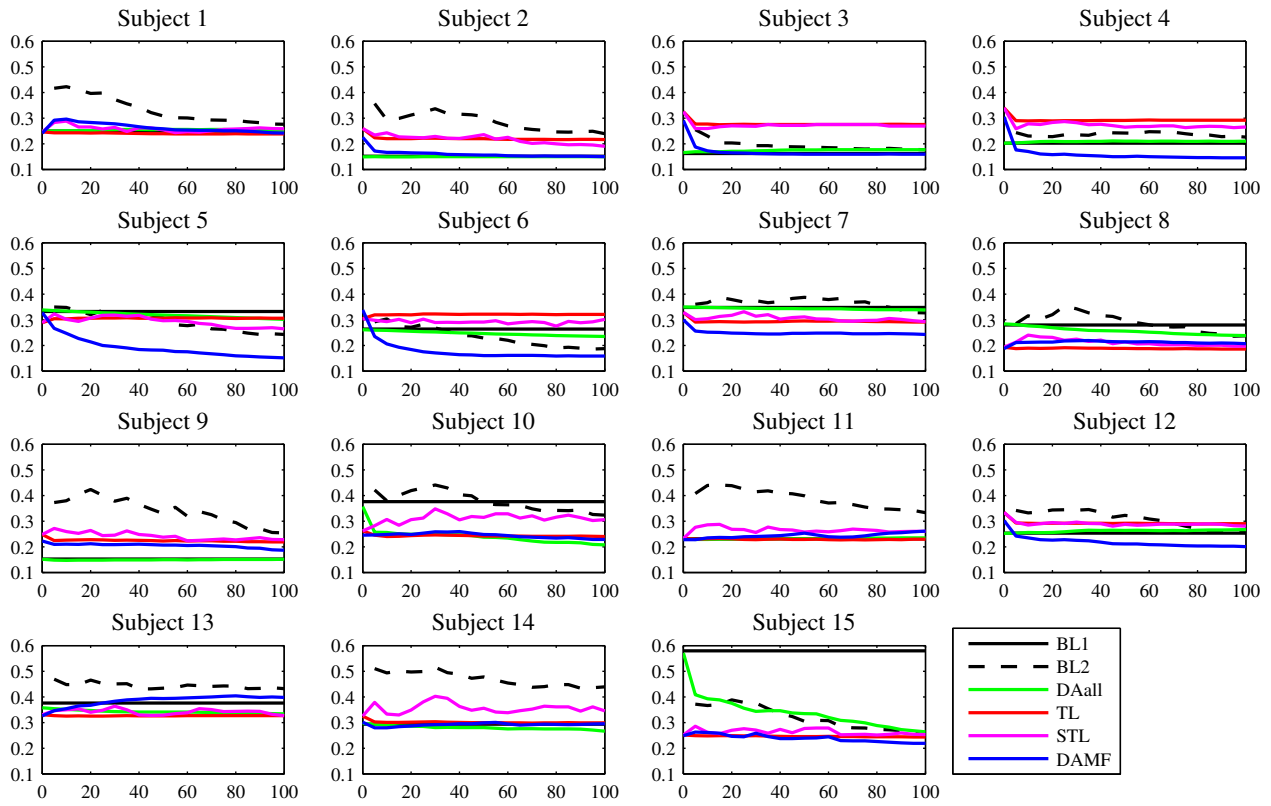
(a) RMSE
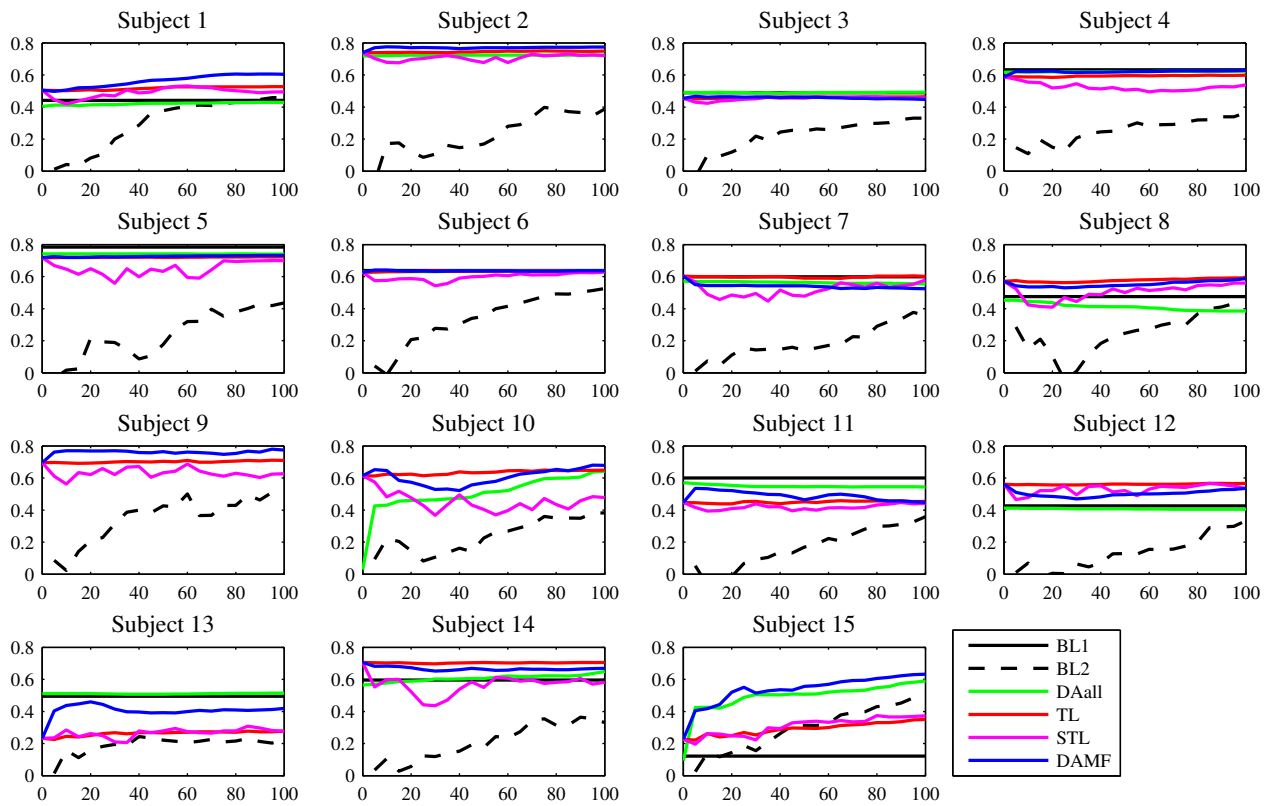
(b) CC

Fig. 6. Performances of the five algorithms for each individual subject. Horizontal axis: m, the number of subject-specific calibration samples.

than several other approaches. Or, in other words, given a desired estimation accuracy, DAMF needs much less subject-specific calibration data.

Although this paper only demonstrated the performance of DAMF for drowsiness estimation, we believe it can also be used in other affective computing applications, e.g., estimating the continuous values of arousal, valence and dominance from speech signals [26]. In addition to domain adaptation for estimation problems, we have also demonstrated its superior performance for classification problems in BCI, e.g., visually-evoked potentials classification [23]–[25]. In summary, domain adaptation, or more generally, transfer learning, is a very valuable machine learning technique in BCI and affective computing.

REFERENCES

[1] L. Bergasa, J. Nuevo, M. Sotelo, R. Barea, and M. Lopez, "Real-time system for monitoring driver vigilance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 63–67, 2006.

[2] C.-H. Chuang, L.-W. Ko, T.-P. Jung, and C.-T. Lin, "Kinesthesia in a sustained-attention driving task," *Neuroimage*, vol. 91, pp. 187–202, 2014.

[3] S.-W. Chuang, L.-W. Ko, Y.-P. Lin, R.-S. Huang, T.-P. Jung, and C.-T. Lin, "Co-modulatory spectral changes in independent brain processes are correlated with task performance," *Neuroimage*, vol. 62, pp. 1469–1477, 2012.

[4] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, pp. 9–21, 2004.

[5] T. DOrazio, M. Leo, C. Guaragnella, and A. Distante, "A visual approach for driver inattention detection," *Pattern Recognition*, vol. 40, no. 8, pp. 2341–2355, 2007.

[6] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2009.

[7] R. 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, 2011.

[8] C.-T. Lin, L.-W. Ko, and T.-K. Shen, "Computational intelligent brain computer interaction and its applications on driving cognition," *IEEE Computational Intelligence Magazine*, vol. 4, no. 4, pp. 32–46, 2009.

[9] C.-T. Lin, L.-W. Ko, I.-F. Chung, T.-Y. Huang, Y.-C. Chen, T.-P. Jung, and S.-F. Liang, "Adaptive EEG-based alertness estimation system by using ICA-based fuzzy neural networks," *IEEE Trans. on Circuits and Systems-I*, vol. 53, no. 11, pp. 2469–2476, 2006.

[10] M. Long, J. Wang, G. Ding, S. J. Pan, and P. S. Yu, "Adaptation regularization: A general framework for transfer learning," *IEEE Trans. on Knowledge and Data Engineering*, vol. 26, no. 5, pp. 1076–1089, 2014.

[11] S. Makeig and M. Inlow, "Lapses in alertness: Coherence of fluctuations in performance and EEG spectrum," *Electroencephalography and Clinical Neurophysiology*, vol. 86, pp. 23–35, 1993.

[12] S. Makeig and T. P. Jung, "Tonic, phasic and transient EEG correlates of auditory awareness in drowsiness," *Cognitive Brain Research*, vol. 4, pp. 12–25, 1996.

[13] S. V. Masten, J. C. Stutts, and C. A. Martell, "Predicting daytime and nighttime drowsy driving crashes based on crash characteristic models," in *Proc. 50th Annual Association of the Association for the Advancement of Automotive Medicine*, Chicago, IL, October 2006.

[14] C. Muhl, B. Allison, A. Nijholt, and G. Chanel, "A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges," *Brain-Computer Interfaces*, vol. 1, no. 2, pp. 66–84, 2014.

[15] "Facts and stats," National Sleep Foundation. [Online]. Available: http://drowsydriving.org/about/facts-and-stats/

[16] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[17] D. Royal, "National survey of distracted and drowsy driving attitudes and behaviors: 2002, vol 1: Findings," National Highway Traffic Safety Administration, Washington, DC, Tech. Rep. DOT HS 809 566, 2003.

[18] B. C. Tefft, "Asleep at the wheel: The prevalence and impact of drowsy driving," AAA Foundation for Traffic Safety, Washington, DC, Tech. Rep., 2010. [Online]. Available: https://www.aaafoundation.org/sites/default/files/2010DrowsyDrivingReport_1.pdf

[19] "Traffic safety facts crash stats: drowsy driving," US Department of Transportation, National Highway Traffic Safety Administration. Washington, DC. [Online]. Available: http://www-nrd.nhtsa.dot.gov/pubs/811449.pdf

[20] Y.-T. Wang, K.-C. Huang, C.-S. Wei, T.-Y. Huang, L.-W. Ko, C.-T. Lin, C.-K. Cheng, and T.-P. Jung, "Developing an EEG-based on-line closed-loop lapse detection and mitigation system," *Frontiers in Neuroscience*, 2014.

[21] C.-S. Wei, Y.-P. Lin, Y.-T. Wang, T.-P. Jung, N. Bigdely-Shamlo, and C.-T. Lin, "Selective transfer learning for EEG-based drowsiness detection," in *Proc. IEEE Int'l Conf. on Systems, Man and Cybernetics*, Hong Kong, October 2015.

[22] 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 Electroacoustics*, vol. 15, pp. 70–73, 1967.

[23] D. Wu, B. J. Lance, and V. J. Lawhern, "Active transfer learning for reducing calibration data in single-trial classification of visually-evoked potentials," in *Proc. IEEE Int'l. Conf. on Systems, Man, and Cybernetics*, San Diego, CA, October 2014.

[24] D. Wu, V. J. Lawhern, and B. J. Lance, "Reducing BCI calibration effort in RSVP tasks using online weighted adaptation regularization with source domain selection," in *Proc. Int'l. Conf. on Affective Computing and Intelligent Interaction*, Xi'an, China, September 2015.

[25] ——, "Reducing offline BCI calibration effort using weighted adaptation regularization with source domain selection," in *Proc. IEEE Int'l. Conf. on Systems, Man and Cybernetics*, Hong Kong, October 2015.

[26] D. Wu, T. D. Parsons, E. Mower, and S. S. Narayanan, "Speech emotion estimation in 3D space," in *Proc. IEEE Int'l Conf. on Multimedia & Expo (ICME)*, Singapore, July 2010, pp. 737–742.

IEEE computer society