Spectral Meta-Learner for Regression (SMLR) Model Aggregation: Towards Calibrationless Brain-Computer Interface (BCI)

Dongrui Wu*, *Senior Member, IEEE*, Vernon J. Lawhern†‡, *Member, IEEE*,
Stephen Gordon§, Brent J. Lance†, *Senior Member, IEEE*, Chin-Teng Lin¶‖, *Fellow, IEEE*
*DataNova, NY USA
†Human Research and Engineering Directorate, U.S. Army Research Laboratory, Aberdeen Proving Ground, MD USA
‡Department of Computer Science, University of Texas at San Antonio, San Antonio, TX USA
§DCS Corp, Alexandria, VA USA
¶Brain Research Center, National Chiao-Tung University, Hsinchu, Taiwan
‖Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia
E-mail: drwu09@gmail.com, vernon.j.lawhern.civ@mail.mil, sgordon@dcscorp.com,
brent.j.lance.civ@mail.mil, ctlin@mail.nctu.edu.tw

***Abstract*—To facilitate the transition of brain-computer interface (BCI) systems from laboratory settings to real-world application, it is very important to minimize or even completely eliminate the subject-specific calibration requirement. There has been active research on calibrationless BCI systems for classification applications, e.g., P300 speller. To our knowledge, there is no literature on calibrationless BCI systems for regression applications, e.g., estimating the continuous drowsiness level of a driver from EEG signals. This paper proposes a novel spectral meta-learner for regression (SMLR) approach, which optimally combines base regression models built from labeled data from auxiliary subjects to label offline EEG data from a new subject. Experiments on driver drowsiness estimation from EEG signals demonstrate that SMLR significantly outperforms three state-of-the-art regression model fusion approaches. Although we introduce SMLR as a regression model fusion in the BCI domain, we believe its applicability is far beyond that.**

***Index Terms*—Brain-computer interface, calibrationless BCI, regression, EEG, ensemble learning, spectral meta-learner**

I. Introduction

EEG-based brain-computer interface (BCI) systems have gained increasing research interest in the last decade, and they have demonstrated promising performance in various applications [24], [25], [30], [33], [40], [41], [44], primarily in laboratory settings. Most BCI systems require a subject-specific calibration session, which could last 5-20 minutes. To facilitate their real-world applications, it is very important to minimize or even completely eliminate this calibration [10], [24], [27].

Lots of approaches have been proposed to minimize the calibration requirement [1], [11], [26], [31], [42], [45]–[52]. Generally they can be partitioned into two groups. The first group focuses on feature extraction: it either extracts more discriminative subject-specific features, e.g., common spatial patterns [2], or extracts more robust and representative features that are less likely to be affected by individual differences, e.g., deep learning [17] and Riemannian geometry [3] features. The second group uses advanced machine learning approaches to achieve high calibration performance with a small amount of subject-specific data, e.g., transfer learning [34], which makes use of auxiliary data from similar/relevant tasks to help the calibration for a new subject, active learning [39], which selects the most informative samples to label, and their combinations [46], [47], [50]. Interestingly, the two groups of approaches are also complementary, and hence they can be combined to further reduce the calibration effort, although a lot of research is needed in this direction.

There has also been considerable literature on completely eliminating subject-specific calibration in BCI [4], [9], [13], [14], [16], [22], [23], [37], which is a very challenging problem due to individual difference and nonstationarity. The approaches here can again be roughly partitioned into two groups: feature extraction and machine learning. For example, [13], [14] performed subject-independent mental state classification by constructing an ensemble of classifiers derived from subject-specific temporal and spatial filters, and then fusing them for a new subject. [4], [9] developed a plug & play BCI system by smartly initializing it using information geometry and then continuously adapting it to the new subject. [22] built a calibrationless P300 speller by combining unsupervised training, transfer learning and language models. [23] performed an online study to verify that zero-training BCI can be achieved through unsupervised learning in an auditory event-related potential paradigm. [16] developed a calibration-free BCI system that allows a user to control an agent to solve a sequential task. It assumes a distribution of possible tasks, and infers the interpretation of EEG signals and the task by selecting the hypothesis that best explains the history of interaction.

This paper focuses on calibrationless BCI in regression

978-1-5090-1897-0/16/$31.00 ©2016 IEEE

problems. To our knowledge, this is the first work in this direction. Our approach shares some similarity with [13], [14], [37], which focus on classification problems: we construct an ensemble of base regression models from relevant auxiliary subjects, and then aggregate them for a new subject. Our main contribution is that we propose a novel spectral meta-learner for regression (SMLR) approach to optimally combine the base regression models. Using a BCI application of driver drowsiness estimation from EEG signals, we show that SMLR can achieve significantly better performance than three other popular regression model combination approaches in the literature.

The remainder of the paper is organized as follows: Section II introduces the details of the SMLR approach. Section III describes experiment setup and performance comparisons of SMLR with four other approaches, and also points out future research directions. Section IV draws conclusions.

II. SMLR for Regression Model Aggregation

This section introduces the proposed SMLR approach for regression model aggregation. We consider regression problems with a continuous value input space $\mathcal{X}$ and a continuous value output space $\mathcal{Y}$. We assume there are n unlabeled samples, $\{\mathbf{x}_j\}_{j=1}^n$, with unknown true outputs $\{y_j\}_{j=1}^n$, and m base regression models, $\{f_i\}_{i=1}^m$. The ith regression model's prediction for $\mathbf{x}_j$ is $f_i(\mathbf{x}_j)$.

Our goal is to accurately estimate y_j by optimally combining $\{f_i(\mathbf{x}_j)\}_{i=1}^m$. The SMLR approach consists of two steps: 1) estimate the accuracy of each base regression model; and, 2) select and combine the strong base regression models.

A. Estimate the Accuracy of the Base Regression Models

The derivation here closely resembles that in [18], which was in turn inspired by [35]. [35] proposed an SML approach for binary classification, which first uses a spectral approach to estimate the accuracies of multiple base binary classifiers from their predictions, and then a meta-learning approach to combine them. [18] proposed *Eigen* and *Eigen-PC*, two approaches to extend SML to prediction scores with arbitrary continuous distributions. Our derivation follows the *Eigen-PC* approach for its simplicity.

First, we normalize $\{f_i(\mathbf{x}_j)\}_{j=1}^n$ to make its mean $\mu_i = 0$ and standard deviation equal 1. We then treat $f_i(\mathbf{x}_j)$ as belonging to a 2-component mixture distribution, where the two components are corresponding to two distinct states $\{0,1\}$ in the output space, e.g., *Non-drowsy* and *Drowsy* in driver drowsiness estimation. Let C_j be an indicator variable of the true state of sample $\mathbf{x}_j$, and $\pi = \mathrm{P}(C_j = 1)$. Then, $f_i(\mathbf{x}_j)$ can be expressed as:

$$f_i(\mathbf{x}_j) = \pi g_{1,i}(\mathbf{x}_j) + (1-\pi) g_{0,i}(\mathbf{x}_j) \tag{1}$$

where $g_{1,i}(\mathbf{x}_j)$ and $g_{0,i}(\mathbf{x}_j)$ are the conditional distributions of $\mathbf{x}_j$ when $\mathbf{x}_j$ belongs to State 1 and State 0, respectively. Let $\mu_{l,i} = E[f_i(\mathbf{x}_j)|C_j = l] = E[g_{l,i}(\mathbf{x})]$, $l = 0, 1$. Then, by taking the expectation on both sides of (1), it follows that

$$\mu_i = \pi\mu_{1,i} + (1-\pi)\mu_{0,i} \tag{2}$$

Because $\mu_i = 0$, we can easily derive

$$\mu_{1,i} = \frac{\pi - 1}{\pi}\mu_{0,i} \tag{3}$$

Let Q be the covariance matrix of $\{f_i(\mathbf{x}_j)\}_{i=1}^m$, i.e., its (i_1, i_2)th element Q_{i_1,i_2} is the population covariance of two base regression models $f_{i_1}(\mathbf{x}_j)$ and $f_{i_2}(\mathbf{x}_j)$:

$$\begin{aligned}
Q_{i_1,i_2} &= Cov(f_{i_1}(\mathbf{x}_j), f_{i_2}(\mathbf{x}_j)) \\
&= \pi E[f_{i_1}(\mathbf{x}_j) f_{i_2}(\mathbf{x}_j)|C_j = 1] \\
&\quad + (1-\pi) E[f_{i_1}(\mathbf{x}_j) f_{i_2}(\mathbf{x}_j)|C_j = 0] \\
&= \pi E[(f_{i_1}(\mathbf{x}_j) - \mu_{1,i_1})(f_{i_2}(\mathbf{x}_j) - \mu_{1,i_2})|C_j = 1] \\
&\quad + \pi\mu_{1,i_1}\mu_{1,i_2} \\
&\quad + (1-\pi) E[(f_{i_1}(\mathbf{x}_j) - \mu_{0,i_1})(f_{i_2}(\mathbf{x}_j) - \mu_{0,i_2})|C_j = 0] \\
&\quad + (1-\pi)\mu_{0,i_1}\mu_{0,i_2} \\
&= \pi Cov(f_{i_1}(\mathbf{x}_j), f_{i_2}(\mathbf{x}_j)|C_j = 1) \\
&\quad + (1-\pi) Cov(f_{i_1}(\mathbf{x}_j), f_{i_2}(\mathbf{x}_j)|C_j = 0) \\
&\quad + \pi\mu_{1,i_1}\mu_{1,i_2} + (1-\pi)\mu_{0,i_1}\mu_{0,i_2}
\end{aligned} \tag{4}$$

Under the assumption that the base regression models are conditionally independent given C_l, we have

$$Cov(f_{i_1}(\mathbf{x}_j), f_{i_2}(\mathbf{x}_j)|C_1 = 1) = 0, \quad i_1 \neq i_2 \tag{5}$$

$$Cov(f_{i_1}(\mathbf{x}_j), f_{i_2}(\mathbf{x}_j)|C_0 = 1) = 0, \quad i_1 \neq i_2 \tag{6}$$

and hence

$$Q_{i_1,i_2} = \pi\mu_{1,i_1}\mu_{1,i_2} + (1-\pi)\mu_{0,i_1}\mu_{0,i_2}, \quad i_1 \neq i_2 \tag{7}$$

Substituting (3) into (7), it follows that

$$\begin{aligned}
Q_{i_1,i_2} &= \pi \cdot \frac{\pi-1}{\pi}\mu_{0,i_1} \cdot \frac{\pi-1}{\pi}\mu_{0,i_2} + (1-\pi)\mu_{0,i_1}\mu_{0,i_2} \\
&= \frac{1-\pi}{\pi}\mu_{0,i_1}\mu_{0,i_2}, \qquad i_1 \neq i_2
\end{aligned} \tag{8}$$

Define a rank-one matrix

$$R = \frac{1-\pi}{\pi}\boldsymbol{\mu}_0\boldsymbol{\mu}_0^T \tag{9}$$

where $\boldsymbol{\mu}_0 = (\mu_{0,1},\ \mu_{0,2},\ \ldots,\ \mu_{0,m})^T$ is the mean vector of the m base regression models given $C_j = 0$. Then, it is easy to see that the off-diagonal entries of the covariance matrix Q are identical to those of R.

From (3) we have

$$\mu_{0,i} - \mu_{1,i} = \frac{1}{\pi}\mu_{0,i}, \tag{10}$$

i.e., $|\mu_{0,i}|$ is a measure on the distance between the two conditional distributions for f_i. This distance may be an indirect indicator of the performance of f_i because, intuitively, a large distance means the two states $\{0,1\}$ can be well separated from the outputs of f_i. So, if we view f_i as a classifier, then it would have good distinguishability, and we expect that this good classification performance also generalizes to good regression performance. However, we must point out that this conclusion is based on intuition, and we do not have a rigorous mathematical proof so far. This is one of our future research directions.

If we know R, then $\boldsymbol{\mu}_0$ can be easily computed as its first leading eigenvector. The question is how to estimate R from $f_i(\mathbf{x}_j)$, $i = 1, ..., m$, $j = 1, ..., n$. Multiple approaches have been proposed in [18], [35]. In this paper we use the simple *Eigen-PC* approach [18], $R \approx Q$, i.e., to approximate R directly by the population covariance matrix Q.

B. Combine Base Regression Models

Once the accuracies of the m base regression models are estimated, a simple weighted average may be used to combine them, i.e.,

$$f(\mathbf{x}_j) = \frac{\sum_{i=1}^{m} \mu_{0,i} f_i(\mathbf{x}_j)}{\sum_{i=1}^{m} \mu_{0,i}} \tag{11}$$

The *Eigen-PC* approach in [18] used a similar idea, but it considered binary classification problems, so a weighted sum instead of weighted average was used.

However, (11) may not be optimal, because:

1) Maybe not all m base regression models are necessary in the final aggregation, because outlier models could significantly deteriorate the ensemble performance. So, it is important to identify and exclude the outliers and maybe also the weak models from the final aggregation.
2) Although $\mu_{0,i}$ is an indirect indicator of the performance of f_i, there is no guarantee that using it directly in (11) will give the best performance. It's possible that some transformation of $\mu_{0,i}$ can serve as a better weight.

In the following we will propose a simple approach to accommodate the first issue. The second one will be considered in our future research.

We first use k-means clustering ($k = 3$) on the absolute values of the elements of $\boldsymbol{\mu}_0$ to partition the m base regression models into three groups:

1) The first group has the smallest centroid, which consists of the outliers.
2) The second group has the median centroid, which consists of the weak models.
3) The third group has the largest centroid, which consists of the strong models.

Clearly, the outliers should be excluded from the final aggregation, and the strong models should be included. The question is whether the weak models should be included or not. Our empirical results show that generally it is beneficial to exclude them, and this approach is used in this paper.

Once the m' strong models $\{f_i\}_{i=1}^{m'}$ are identified, we again use a simple weighted average to aggregate them:

$$f(\mathbf{x}_j) = \frac{\sum_{i=1}^{m'} \mu_{0,i} f_i(\mathbf{x}_j)}{\sum_{i=1}^{m'} \mu_{0,i}} \tag{12}$$

C. The Complete SMLR Algorithm

The complete SMLR algorithm is shown in Algorithm 1. It first uses the spectral approach to estimate the accuracy of the m base regression models, then uses k-means clustering ($k = 3$) to identify the strong models, and finally employs a weighted average to aggregate them.

Algorithm 1: The SMLR algorithm.

Input: n unlabeled samples, $\{\mathbf{x}_j\}_{j=1}^{n}$;
m base regression models, $\{f_i\}_{i=1}^{m}$.
Output: The n estimated outputs, $\{f(\mathbf{x}_j)\}_{j=1}^{n}$.
Compute the covariance matrix Q of $\{f_i\}_{i=1}^{m}$;
Compute the first leading eigenvector, $\boldsymbol{\mu}_0$, of Q;
Perform k-means clustering ($k = 3$) on the absolute values of the elements of $\boldsymbol{\mu}_0$;
Identify the m' strong regression models as those belong to the cluster with the maximum centroid;
Return $\{f(\mathbf{x}_j\}_{j=1}^{n}$ computed by (12).

III. Experiment and Results

This section presents the experiment setup that is used to evaluate the performance of SMLR, the performance comparison of SMLR with four other approaches, and discussions on our future research directions.

A. Experiment Setup

The experimental setup was identical to that in [45]. We recruited 16 healthy subjects with normal/corrected-to-normal vision to participant in a sustained-attention driving experiment [7], [8], consisting of a real vehicle mounted on a motion platform with 6 degrees of freedom immersed in a 360-degree virtual-reality (VR) scene. The Institutional Review Board of the Taipei Veterans General Hospital approved the experimental protocol, and 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 during driving, the VR scenes simulated monotonous driving at a fixed speed (100 km/h) on a straight and empty highway. During the experiment, random lane-departure events were applied every 5-10 seconds, and participants were instructed to steer the vehicle to compensate for them immediately. The response time was recorded and later converted to a drowsiness index, as research has shown that it has strong correlation with fatigue [21]. Participants' scalp EEG signals were recorded using a 500Hz 32-channel Neuroscan system (30-channel EEGs plus 2-channel earlobes), and their cognitive states and driving performance were also monitored via a surveillance video camera and the vehicle trajectory throughout the experiment.

B. Preprocessing and Feature Extraction

The preprocessing and feature extraction methods were almost identical to those in our previous research [45], except that herein we used principal component features instead of the theta band power features for better regression performance [49].

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.

We defined a function [42], [45] to map the response time τ to a drowsiness index $y \in [0, 1]$:

$$y = \max\left\{0, \frac{1 - e^{-(\tau-\tau_0)}}{1 + e^{-(\tau-\tau_0)}}\right\} \tag{13}$$

$\tau_0 = 1$ was used in this paper, as in [45]. 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 [28].

We used EEGLAB [?] for EEG signal preprocessing. A 1-50 Hz band-pass filter was applied to remove high-frequency muscle artifacts, line-noise contamination and direct current drift. Next the EEG data were downsampled from 500 Hz to 250 Hz and re-referenced to averaged earlobes.

We tried to predict the drowsiness index for each subject every 10 seconds. All 30 EEG channels were used in feature extraction. We epoched 30-second EEG signals right before each sample point, and computed the average power spectral density (PSD) in the theta band (4-7.5 Hz) for each channel using Welch's method [43], as research [29] has shown that theta band spectrum is a strong indicator of drowsiness.

Next, we converted the 30 theta band powers to dBs. To remove noises or bad channel readings, we removed channels whose maximum dBs were larger than 20. We then normalized the dBs of each remaining channel to mean zero and standard deviation one, and extracted a few (usually around 10) leading principal components, which accounted for 95% of the variance. The projections of the dBs onto these principal components were then normalized to $[0, 1]$ and used as our features.

C. Evaluation Method and Performance Measures

This work is a step towards calibrationless BCI systems, the goal of which is to design BCI systems without using any labeled subject-specific calibration data. Our complete approach is shown in Fig. 1. Let the 15th subject be a new subject to our BCI system, who has only unlabeled EEG data, and our goal is to map these data to his/her drowsiness indices without asking for any labels. We first use labeled data from the other 14 subjects to build 14 base ridge regression (RR) models, feed the unlabeled data from the 15th subject into them, and then use different model fusion approaches to aggregate the 14 RR models to get the final predictions. Finally we compare the predictions with the true drowsiness indices of the 15th subject and compute the root mean squared error (RMSE) and correlation coefficient (CC) as our performance measures. We repeat this process 15 times so that each subject has a chance to be the "15th" subject.

D. Algorithms

We compare SMLR with a slighted modified *Eigen-PC* approach [18], where the weighted sum is replaced by a weighted average, as stated in Section II-B, and two other

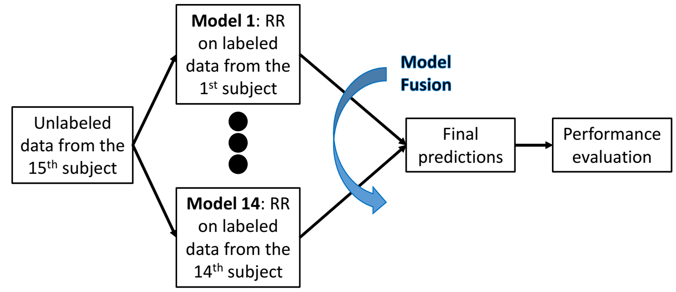


Fig. 1. Illustration of the performance evaluation method.

popular regression model combination approaches in the literature [32], [38]:

1) *Average* [36], which simply takes the average of the m base regression models as the final prediction.
2) *Median* [6], [15], which uses the median of the m base regression models as the final prediction.

Additionally, we also constructed an *Oracle* approach, which assumes that we know the true RMSEs of the m base regression models (which is impossible in practice), uses k-means clustering ($k = 3$) to partition the m models into three groups, and finally employs a weighted average to aggregate the m' strong models from the group with the smallest centroid. The weights were again determined from the 1st leading eigenvector of Q. Generally the *Oracle* approach represents the upper bound of the performance the SMLR approach could reach if the performances of the m base regression models are estimated perfectly. So, it is used as a benchmark to evaluate how much SMLR can be further improved.

Note that there is another popular regression model fusion approach called stacked regression [5], which fits an optimal linear regression model on top of the base regression models to fuse them. However, the objective function in the fitting requires some labeled data from the new subject, which are not available in our application. So, stacked regression is not considered in this paper.

E. Experimental Results and Discussions

The prediction outputs of the five algorithms, along with the groundtruth drowsiness index computed from the response time, are shown in Fig. 2. Observe that without using any labeled data from the new subject, the predictions from the five algorithms all have strong correlations with the groundtruth.

The RMSEs and CCs of the five approaches are shown in Figs. 3(a) and 3(b), respectively, where the last group in each figure shows the average performance across the 15 subjects.

Observe from Fig. 3(a) that *SMLR* achieved smaller RMSE than *Average*, *Median* and *Eigen-PC* for 12 of the 15 subjects, and comparable RMSE for the remaining three subjects (8, 11 and 14). The average RMSE of *SMLR* was also much smaller than those of *Average*, *Median* and *Eigen-PC*. However, *Oracle* achieved smaller RMSE than *SMLR* for most of the subjects, suggesting that *SMLR* can still be improved,

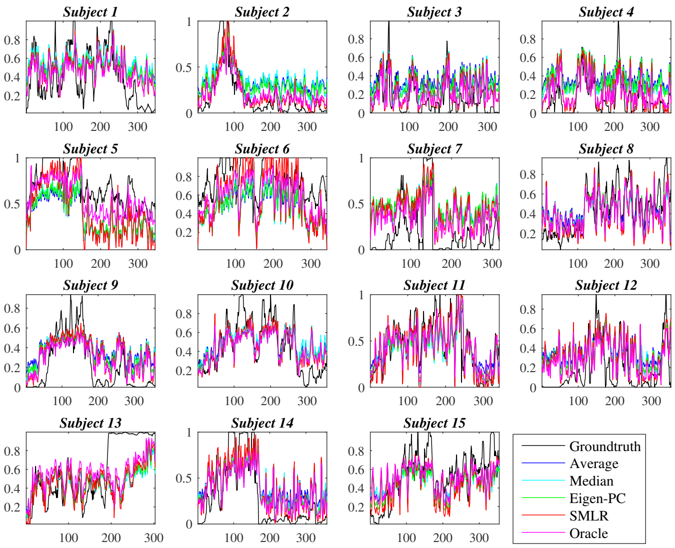
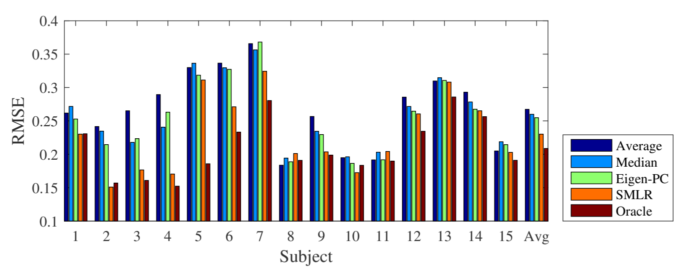


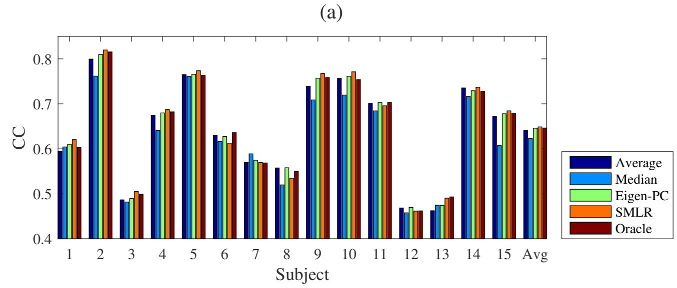
Fig. 2. Predicted drowsiness indices from the five algorithms.

by making better estimations of the performances of the base regression models.

Fig. 3(b) shows that the CC differences among the five approaches were not as significant as the RMSE differences, because our primary objective was to optimize the RMSE instead of the CC. However, *SMLR* still achieved larger CC than *Average*, *Median* and *Eigen-PC* for nine of the 15 subjects, and comparable CC for the remaining six subjects. *SMLR* also had the largest average CC among the five approaches, although it was only slightly better than the CCs of *Average*, *Eigen-PC* and *Oracle*.



(a)



(b)

Fig. 3. (a) RMSEs and (b) CCs of the five approaches on the 15 subjects. The last group in each subfigure shows the average performance across the 15 subjects.

We also performed paired t-tests to check if the RMSE and CC differences between SMLR and the other four approaches were statistically significant, using $\alpha = 0.05$ and Bonferroni correction [12]. The results are shown in Table I, where the statistically significant ones are marked in bold. Observe that the RMSE differences between *SMLR* and three other approaches (*Average*, *Median*, and *Eigen-PC*) were statistically significant, suggesting that *SMLR* significantly outperformed *Average*, *Median* and *Eigen-PC* in terms of RMSE. The RMSE difference between *SMLR* and *Oracle* was marginally statistically significant, indicating that still significant enhancements could be made to *SMLR* to reach its full potential (represented by *Oracle*). The CC differences between *SMLR* and *Median* was also statistically significant, but the CC differences between *SMLR* and the other three approaches were not.

TABLE I
p-VALUES (AFTER BONFERRONI CORRECTION) OF PAIRED t-TESTS BETWEEN *SMLR* AND THE OTHER FOUR APPROACHES ON THE RMSES AND CCS. $\alpha = 0.05$.

Approach	*Average*	*Median*	*Eigen-PC*	*Oracle*
RMSE	**.0110**	**.0022**	**.0283**	.0885
CC	.3031	**.0086**	1.4716	1.6550

Finally, we performed in-depth analysis to study why *SMLR* could significantly outperform *Average*, *Median* and *Eigen-PC* in terms of RMSE. Recall that the two main novelties of *SMLR* are the estimation of the performances of the base regression models, and the identification of the strong models. We studied these two novelties separately.

In Fig. 4 we show the values of the first leading eigenvector of Q (representing the estimated performance of the corresponding base regression models) versus the true RMSEs of the corresponding base regression models on the testing data for each subject. The averages across the 15 subjects are shown in the last subfigure. We sorted the values of the eigenvector in descending order to better visualize the trend. Ideally, a large eigenvector value, which indicates good performance, should corresponds to a small RMSE, and hence monotonically decreasing eigenvector values should correspond to monotonically increasing RMSEs. Observe from Fig. 4, especially the last subfigure, that generally as the value of the eigenvector decreased, the corresponding RMSE increased, although it was not monotonic. This suggests that SMLR can indeed rank the performances of the base regression models, although not perfect.

To check whether the strong models selected by SMLR were really among the best, we marked the selected models in shade in Fig. 4. Observe, especially from the last subfigure, that generally SMLR can indeed identify the strong base regression models, although they were not necessary the best ones.

F. Discussions and Future Research

We have shown that our proposed *SMLR* can significantly outperform *Average*, *Median* and *Eigen-PC* in terms of RMSE, but still there is significant room for improvement until it reaches or exceeds the performance of *Oracle*. We will investigate the following directions in our future research:

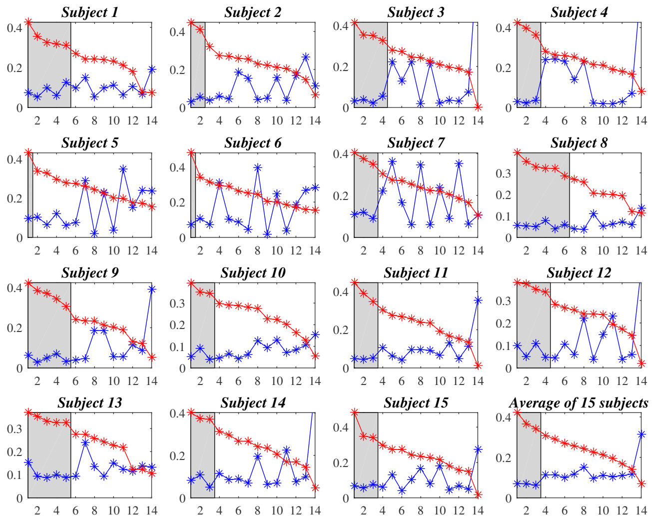


Fig. 4. The values of the leading eigenvector of the covariance matrix (–*–*–; sorted in descending order) versus the testing RMSEs (–*–*–) of the corresponding base regression models. The points in shade are the base models selected by SMLR. The last subfigure shows the averages across the 15 subjects.

1) More accurate estimations of the performances of the base regression models, which better enables us to identify the strongest models. In this paper we used the simplest approach by equating R to Q. Several more sophisticated approaches were proposed in [18], [35]. We implemented the *Eigen* approach in [18] but failed to achieve noticeable performance improvement. In the future we will investigate other possibilities. For example, in [48] we estimated the classification accuracies of base classifiers from their agreement rate (the probability that two classifiers make errors simultaneously), and showed that the resulting model fusion approach achieved better performance than estimating the classification accuracies from the covariance matrix. We will extend that work from classification to regression.
2) As mentioned in Section II-B, we can estimate the accuracies of the base regression models, but there is no guarantee that using them directly as the weights in (11) would give the best performance. Some transformation of $\mu_{0,i}$ could be better. Additionally, it has been shown in [19], [20], [35], [48] that, for classification problems, initializing the weights for the base classifiers using the spectral or the agreement rate approach and then iteratively refining them using a maximum likelihood estimator can improve the performance. It is also interesting to extend this iterative approach from classification to regression.

IV. Conclusions

Today most BCI systems require a calibration session before it can be applied to a new subject, which may hinder their real-world acceptances. There has been active research on calibrationless BCI systems, but so far all of them focused on classification problems. This paper for the first time considers regression problems in calibrationless BCI systems. We proposed a novel SMLR approach to aggregate base regression models built from labeled data from auxiliary subjects, and then apply the fused model to a new subject for offline BCI applications, without requiring any labels from the new subject. Experiments on driver drowsiness estimation from EEG signals demonstrated that SMLR significantly outperformed three state-of-the-art regression model fusion approaches. We believe that SMLR will have broad applications in regression model fusion beyond BCI.

Acknowledgement

Research was sponsored by the U.S. Army Research Laboratory and was accomplished under Cooperative Agreement Numbers W911NF-10-2-0022 and W911NF-10-D-0002/TO 0023. The views and the conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory or the U.S Government.

References

[1] M. Ahn, H. Cho, and S. C. Jun, "Calibration time reduction through source imaging in brain computer interface (BCI)," *Communications in Computer and Information Science*, vol. 174, pp. 269–273, 2011.
[2] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE World Congress on Computational Intelligence*, Hong Kong, June 2008, pp. 2390–2397.
[3] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain-computer interface classification by Riemannian geometry," *IEEE Trans. on Biomedical Engineering*, vol. 59, no. 4, pp. 920–928, 2012.
[4] A. Barachant and M. Congedo, "A plug & play P300 BCI using information geometry," *arXiv: 1409.0107*, 2014.
[5] L. Breiman, "Stacked regressions," *Machine Learning*, vol. 24, pp. 49–64, 1996.
[6] P. Buhlmann, *Bagging, Boosting and Ensemble Methods*. Springer, 2010.
[7] 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.
[8] 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.
[9] M. Congedo, A. Barachant, and A. Andreev, "A new generation of brain-computer interface based on Riemannian geometry," *arXiv: 1310.8115*, 2013.
[10] J. d. R. Millan, R. Rupp, G. R. Muller-Putz, R. Murray-Smith, C. Giugliemma, M. Tangermann, C. Vidaurre, F. Cincotti, A. Kubler, R. Leeb, C. Neuper, K.-R. Muller, and D. Mattia, "Combining brain-computer interfaces and assistive technologies: State-of-the-art and challenges," *Frontiers in Neuroscience*, vol. 4, 2010.
[11] S. DALHOUMI, G. DRAY, and J. MONTMAIN, "Knowledge transfer for reducing calibration time in brain-computer interfacing," in *Proc. IEEE 26th Int'l Conf. on Tools with Artificial Intelligence*, Limassol, Cyprus, November 2014.
[12] O. Dunn, "Multiple comparisons among means," *Journal of the American Statistical Association*, vol. 56, pp. 62–64, 1961.
[13] S. Fazli, C. Grozea, M. Danoczy, F. Popescu, B. Blankertz, and K.-R. Muller, "Subject independent EEG-based BCI decoding," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, Vancouver, Canada, December 2009.
[14] S. Fazli, F. Popescu, M. Danoczy, B. Blankertz, K.-R. Muller, and C. Grozea, "Subject-independent mental state classification in single trials," *Neural Networks*, vol. 22, pp. 1305–1312, 2009.

[15] J. H. Friedman, "Multivariate adaptive regression splines," *Annals of Statistics*, vol. 19, no. 1, pp. 1–141, 1991.

[16] J. Grizou, I. Iturrate, L. Montesano, P.-Y. Oudeyer, and M. Lopes, "Calibration-free BCI based control," in *Proc. 28th AAAI Conf. on Artificial Intelligence*, Quebec City, Quebec, Canada, July 2014.

[17] M. Hajinoroozi, T. Jung, C. Lin, and Y. Huang, "Feature extraction with deep belief networks for driver's cognitive states prediction from EEG data," in *Proc. IEEE China Summit and Int'l. Conf. on Signal and Information Processing*, Chengdu, China, July 2015.

[18] I. Ionita-Laza, K. McCallum, B. Xu, and J. D. Buxbaum, "A spectral approach integrating functional genomic annotations for coding and noncoding variants," *Nature Genetics*, vol. 48, no. 2, pp. 214–220, 2016.

[19] A. Jaffe, E. Fetaya, B. Nadler, T. Jiang, and Y. Kluger, "Unsupervised ensemble learning with dependent classifiers," *arXiv: 1510.05830*, 2015.

[20] A. Jaffe, B. Nadler, , and Y. Kluger, "Estimating the accuracies of multiple classifiers without labeled data," in *Proc. 18th Int'l Conf. on Artificial Intelligence and Statistics (AISTATS)*, San Diego, CA, May 2015.

[21] Q. Ji, Z. Zhu, and P. Lan, "Real-time nonintrusive monitoring and prediction of driver fatigue," *IEEE Trans. on Vehicular Technology*, vol. 53, no. 4, pp. 1052–1068, 2004.

[22] P.-J. Kindermans and B. Schrauwen, "Dynamic stopping in a calibration-less P300 speller," in *Proc. 5th Int'l Brain-Computer Interface Meeting*, Pacific Grove, CA, June 2013.

[23] P.-J. Kindermans, M. Schreuder, B. Schrauwen, K.-R. Muller, and M. Tangermann, "True zero-training brain-computer interfacing an online study," *PLOS One*, vol. 9, no. 7, 2014, e102504.

[24] B. J. Lance, S. E. Kerick, A. J. Ries, K. S. Oie, and K. McDowell, "Brain-computer interface technologies in the coming decades," *Proc. of the IEEE*, vol. 100, no. 3, pp. 1585–1599, 2012.

[25] L.-D. Liao, C.-T. Lin, K. McDowell, A. Wickenden, K. Gramann, T.-P. Jung, L.-W. Ko, and J.-Y. Chang, "Biosensor technologies for augmented brain-computer interfaces in the next decades," *Proc. of the IEEE*, vol. 100, no. 2, pp. 1553–1566, 2012.

[26] F. Lotte and C. Guan, "Learning from other subjects helps reducing brain-computer interface calibration time," in *Proc. IEEE Int'l. Conf. on Acoustics Speech and Signal Processing (ICASSP)*, Dallas, TX, March 2010.

[27] F. Lotte, "Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain-computer interfaces," *Proc. of the IEEE*, vol. 103, no. 6, pp. 871–890, 2015.

[28] 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.

[29] 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.

[30] S. Makeig, C. Kothe, T. Mullen, N. Bigdely-Shamlo, Z. Zhang, and K. Kreutz-Delgado, "Evolving signal processing for brain-computer interfaces," *Proc. of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1567–1584, 2012.

[31] A. Marathe, V. Lawhern, D. Wu, D. Slayback, and B. Lance, "Improved neural signal classification in a rapid serial visual presentation task using active learning," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 3, pp. 333–343, 2016.

[32] J. a. Mendes-Moreira, C. Soares, A. M. Jorge, and J. F. D. Sousa, "Ensemble approaches for regression: A survey," *ACM Computing Surveys*, vol. 45, no. 1, pp. 10:1–10:40, 2012.

[33] 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.

[34] 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.

[35] F. Parisi, F. Strino, B. Nadler, and Y. Kluger, "Ranking and combining multiple predictors without labeled data," *Proc. National Academy of Science (PNAS)*, vol. 111, no. 4, pp. 1253–1258, 2014.

[36] M. P. Perrone and L. N. Cooper, "When networks disagree: Ensemble methods for hybrid neural networks," in *Neural networks for speech and image processing*, R. Mammone, Ed. Chapman and Hall, 1993, pp. 126–142.

[37] A. M. Ray, R. Sitaram, M. Rana, E. Pasqualotto, K. Buyukturkoglu, C. Guan, K. K. Ang, C. Tejos, F. J. Zamorano, F. Aboitiz, N. Birbaumer, and S. Ruiz, "A subject-independent pattern-based brain-computer interface," *Frontiers in Behavioral Neuroscience*, vol. 9, no. 269, 2015.

[38] Y. Ren, L. Zhang, and P. Suganthan, "Ensemble classification and regression–recent developments, applications and future directions," *IEEE Computational intelligence magazine*, vol. 11, no. 1, pp. 41–53, 2016.

[39] B. Settles, "Active learning literature survey," University of Wisconsin–Madison, Computer Sciences Technical Report 1648, 2009.

[40] D. S. Tan and A. Nijholt, Eds., *Brain-Computer Interfaces: Applying our Minds to Human-Computer Interaction*. London: Springer, 2010.

[41] J. van Erp, F. Lotte, and M. Tangermann, "Brain-computer interfaces: Beyond medical applications," *Computer*, vol. 45, no. 4, pp. 26–34, 2012.

[42] 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.

[43] 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.

[44] J. Wolpaw and E. W. Wolpaw, Eds., *Brain-Computer Interfaces: Principles and Practice*. Oxford, UK: Oxford University Press, 2012.

[45] 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*, Xi'an, China, September 2015.

[46] 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.

[47] D. Wu, B. J. Lance, and T. D. Parsons, "Collaborative filtering for brain-computer interaction using transfer learning and active class selection," *PLoS ONE*, 2013.

[48] D. Wu, V. J. Lawhern, S. Gordon, B. J. Lance, and C.-T. Lin, "Agreement rate initialized maximum likelihood estimator (ARIMLE) for ensemble classifier aggregation and its application in brain-computer interface," in *Proc. IEEE Int'l Conf. on Systems, Man and Cybernetics*, Budapest, Hungary, October 2016.

[49] 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*, Budapest, Hungary, October 2016.

[50] D. Wu, V. J. Lawhern, W. D. Hairston, and B. J. Lance, "Switching EEG headsets made easy: Reducing offline calibration effort using active weighted adaptation regularization," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, 2016, in press.

[51] 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.

[52] D. Wu, V. J. Lawhern, and B. J. Lance, "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.