

# Low-Dimensional Subject Representation-Based Transfer Learning in EEG Decoding

Po-Yuan Jeng<sup>1</sup>, Chun-Shu Wei<sup>1</sup>, *Member, IEEE*, Tzyy-Ping Jung<sup>2</sup>, *Fellow, IEEE*,  
and Li-Chun Wang<sup>1</sup>, *Fellow, IEEE*

**Abstract**—Recently, the advances in passive brain-computer interfaces (BCIs) based on electroencephalogram (EEG) have shed light on real-world neuromonitoring technologies. However, human variability in the EEG activities hinders the development of practical applications of EEG-based BCI. To tackle this problem, many transfer-learning techniques perform supervised calibration. This kind of calibration approach requires task-relevant data, which is impractical in real-life scenarios such as drowsiness during driving. This study presents a transfer-learning framework for EEG decoding based on the low-dimensional representations of subjects learned from the pre-trial EEG. Tensor decomposition was applied to the pre-trial EEG of subjects to extract the underlying characteristics in subject, spatial, and spectral domains. Then, the proposed framework assessed the characteristics to obtain the low-dimensional subject representations such that the subjects with similar brain dynamics can be identified. This method can leverage the existing data from other users, and a small number of data from a rapid, non-task, unsupervised calibration from a new user to build an accurate BCI. Our results demonstrated that, in terms of prediction accuracy, the proposed low-dimensional subject representation-based transfer learning (LDSR-TL) framework outperformed the random selection, and the Riemannian manifold approach in cognitive-state tracking, while requiring fewer training data. The results can greatly improve the practicality, and usability of EEG-based BCI in the real world.

**Index Terms**—Brain-computer interface, transfer learning and tensor decomposition.

Manuscript received May 3, 2020; revised August 12, 2020 and September 11, 2020; accepted September 13, 2020. Date of publication September 22, 2020; date of current version June 4, 2021. This work was supported in part by the Exchange Program of the Ministry of Education (MOE), Taiwan, under Contract 108DR012, in part by the Higher Education Sprout Project of the National Chiao Tung University and Ministry of Education (MOE), in part by the Ministry of Science and Technology (MOST), Taiwan, under Contracts 107-2917-I-009-008, 108-2634-F-009-006, and 109-2222-E-009-006-MY3, in part by the Army Research Laboratory under Grants W911NF-10-2-0022 and W911NF2020088, and in part by the US National Science Foundation (NSF) under Grant CBET-1935860. (*Corresponding author: Li-Chun Wang.*)

Po-Yuan Jeng and Li-Chun Wang are with the Department of Electrical, and Computing Engineering, National Chiao Tung University, Hsinchu 30010, Taiwan (e-mail: pyjeng.eed03g@nctu.edu.tw; lichun@g2.nctu.edu.tw).

Chun-Shu Wei is with the Institute of Education, and the Department of Computer Science, National Chiao Tung University, Hsinchu 30010, Taiwan (e-mail: cswei.tw@gmail.com).

Tzyy-Ping Jung is with the Department of Bioengineering, and Swartz Center for Computational Neuroscience, University of California, San Diego, CA 92093 USA (e-mail: tpjung@ucsd.edu).

Digital Object Identifier 10.1109/JBHI.2020.3025865

## I. INTRODUCTION

RECENTLY, passive brain-computer interfaces (BCIs) based on electroencephalogram (EEG) decoding of brain activities have enabled many non-invasive neuromonitoring applications including emotion recognition [1], [2], psychiatric diagnosis [3], cognitive load estimation [4], alertness tracking [5], etc. A neuromonitoring BCI usually involves a data-driven EEG decoding model that characterizes EEG patterns associated with brain states of interest [6]. However, pervasive and elusive variability in the EEG data poses an inevitable challenge to deploying a neuromonitoring BCI into practical uses. Conventional BCIs rely on individual user's calibration data prior to each usage that can train a model for a new user. In the training phase, a sufficient amount of EEG data relevant to a specific BCI task was collected in a programmed process. Such a calibration procedure could be time-consuming or impractical, depending on the content and the type of a BCI task. Further, some of the brain states cannot be easily available or reproducible. The attempt of applying a neuromonitoring BCI to a new user could easily fail due to the unavailability of individualized calibration data related to specific brain states.

While deploying a neuromonitoring BCI to routine uses might suffer from the notorious calibration process before each use, efforts have been made to seek alternatives that bypass the individualized calibration and/or obviate the human variability in EEG data. Transfer-learning techniques have been applied to tackle cross-domain problems in a variety of EEG-based BCI systems. In brief, transfer learning is a sub-field of machine learning that includes approaches of transferring knowledge from the source domain to the target domain, where the source domain contains sufficient labeled data, and the target domain often lacks labeled data [7], [8]. Often the variability of EEG data across users deteriorates the robustness of EEG decoding, and meanwhile performing individualized calibration for a new user may not be practical if ever possible. Transfer learning herein serves as an alternative solution that leverages the knowledge learned from other users to facilitate EEG decoding for a new user with zero/minimal calibration.

Recently, subject-transfer approaches have been developed to tackle the issues in cross-subject learning in BCI [5], [9]–[12]. These approaches aim at transferring the labeled training data from existing subjects to a new subject who has limited labeled data. The underlying assumption is that subjects might share a trained model with a minor loss of accuracy if they exhibit

comparable brain responses to the same conditions/events. The similarity in the brain responses among subjects refers to “subject similarity” in this study. Previous studies have shown that the subject similarity could facilitate cross-subject transfer learning, resulting in improved BCI performance [5], [9].

The measurement of the subject similarity is an important step, but it is challenging because of the high dimensionality of the EEG space. A straightforward approach is to calculate the power spectral density (PSD) of each channel and then convert them into a two-dimensional array. Then, the Euclidean distance can be used to calculate the similarity. However, this approach has two major problems. First, the vectors ( $\#channel \times \#frequency$ ) falls onto a very high dimensional space, where the concepts of distance and proximity become meaningless [13]. This phenomenon is the *curse of dimensionality* that has been reported as a major problem in many machine-learning studies [14], [15]. Secondly, the two-dimensional representation is the flattened result of the  $channel \times frequency$  data, which omits the information of the spatial-frequency interaction.

This study proposes a low-dimensional subject-representation (LDSR) approach based on tensor decomposition [16], [17]. Although tensor decomposition is a well-known dimensionality-reduction technique for multi-way data, applying tensor decomposition in the field of EEG decoding has less seen in the literature. In [18], the authors provided a comprehensive survey of using tensor decomposition for EEG analysis. However, most of the works focused on discovering the characteristics of EEG but neglected that the outcomes of decomposition can be used for obtaining subject similarity. To our best knowledge, our proposed method, LDSR-Transfer Learning (or LDSR-TL) is the first work to apply tensor decomposition for subject transfer. We apply tensor decomposition to reduce the dimensions of the subject representation while preserving the spatial-frequency interaction. We first form the pre-trial EEG of all subjects into a three-mode tensor ( $subject \times channel \times frequency$ ), and then perform tensor decomposition to obtain low-dimensional vectors as representations of subjects. As a result, the data are reduced to a low-dimensional space, and common distance matrices can be applied to assess the subject similarity. The contributions of this paper are two folds:

- A low-dimensional subject representation-based transfer learning approach is developed for EEG categorization, where tensor decomposition was used to solve the curse of dimensionality. Thus, labeled data of existing subjects can be reused on decoding the EEG of a new user.
- A working subject-transfer system is implemented for driver drowsiness detection. This system is used to evaluate the performance of the three considered methods. We learn that the proposed method improves the performance of accuracy up to 8% and requires 40% fewer training data to reach the performance plateau.

This study implemented the proposed representation learning model under a subject-transfer framework to evaluate the drowsiness-detection accuracy. The dataset for evaluation was collected from a sustained lane-keeping driving task. Twenty-five subjects participated in a total of 54 simulated

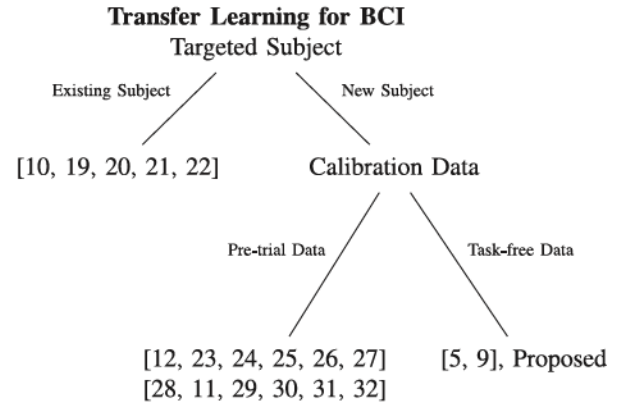


Fig. 1. Related works categorization in a binary tree.

driving sessions, in which simultaneous driving behaviors and 30-channel EEG recordings were measured, synchronized, and recorded. The proposed framework was evaluated by the performance of drowsiness detection based on EEG decoding.

The remaining parts of this paper are organized as follows. Section II introduces the related works of transfer learning on BCI. This section provides an overview of current methods and shows how our approach stands out from others. Section III gives a background of the proposed method. Section IV presents the proposed method and the subject-transfer framework in detail. Section V shows the results and discussions of the leave-one-subject-out (LOSO) cross-validation on a real dataset. Section VII presents our conclusion of this study.

## II. RELATED WORKS

This section reviews the related works on transfer learning techniques for EEG-based BCI. Although some research aimed at intra-subject-cross session transfer [33], this study will mainly focus on the review of inter-subject transfer research. Fig. 1 categorizes the proposed framework and its related works. Table I contrasts the proposed method with the pre-existing works.

In the early stage of development, researchers achieved transfer learning by finding an invariant feature set that is robust across subjects. This type of approach is called the *feature-representation-learning* in the field of transfer learning. Kang *et al.* developed a weighted common spatial filter (CSF) that can be applied to all subjects [34]. Tu and Sun proposed a method that can extract both robust CSFs for all the subjects and adaptive CSFs for a single subject [10]. Recently, Özdenizci *et al.* aimed at discovering the subject-independent features across subjects by using convolutional neural networks and adversarial training [21]. Zhang *et al.* and Jeon *et al.* were developing methods for learning the invariant data representations with deep learning approach [29], [31].

Later, a *parameter-transfer* approach was developed to transfer pre-trained models to the target one. Parameter transfer includes two phases: 1) Train models with parameters learned from all subjects; 2) Adjust a subset of parameters that are subject-dependent arising from individual differences. In [19],

**TABLE I**  
COMPARISON BETWEEN TRANSFER LEARNING METHODS ON BCI

| Target Domain    | Calibration Data       | Paper                   | Year | Transfer Learning Approach                | Brief Description                                                             |
|------------------|------------------------|-------------------------|------|-------------------------------------------|-------------------------------------------------------------------------------|
| N/A              | N/A                    | Kang et al. [33]        | 2009 | Feature-representation                    | Weighted CSF                                                                  |
| Existing Subject | Whole Data             | Tu and Sun [10]         | 2012 | Feature-representation                    | Adaptive CSF                                                                  |
|                  |                        | Özdenizci et al. [21]   | 2020 | Feature-representation                    | Invariant feature by convolutional neural network adversarial training        |
|                  |                        | Alamgir et al. [19]     | 2010 | Parameter                                 | Learning shared prior distribution                                            |
|                  |                        | Jayaram et al. [20]     | 2016 | Parameter                                 | Learning shared structure of data                                             |
|                  |                        | Zhang et al. [22]       | 2015 | Distribution-matching                     | Feature transformation by transfer components analysis                        |
|                  |                        |                         |      |                                           |                                                                               |
| New Subject      | Task-relevant Data     | Reuderink et al. [24]   | 2011 | Baseline-aligning                         | Reference removal                                                             |
|                  |                        | Zanini et al. [25]      | 2018 | Baseline-aligning                         | Data alignment in Riemannian manifold                                         |
|                  |                        | He and Wu [28]          | 2019 | Baseline-aligning                         | Data alignment in Euclidean space                                             |
|                  |                        | Rodrigues et al. [26]   | 2018 | Distribution-matching                     | Feature transformation by Riemannian geometry                                 |
|                  |                        | Dai et al. [27]         | 2019 | Distribution-matching                     | Feature transformation based on distributions                                 |
|                  |                        | Zhang and Wu [23]       | 2019 | Baseline-aligning & Distribution-matching | Data alignment & feature transformation                                       |
|                  |                        | Li et al. [12]          | 2019 | Distribution-matching                     | Feature transformation by affine mapping                                      |
|                  |                        | Dagois et al. [11]      | 2019 | Instance                                  | Subject-transfer by similarity of data distribution                           |
|                  |                        | Zhang et al. [30]       | 2019 | Instance                                  | Instance transfer by the similarity of the data distribution                  |
|                  |                        | Zhang et al. [29]       | 2019 | Feature-representation                    | Invariant feature recurrent attention network                                 |
|                  |                        | Jeon et al. [31]        | 2019 | Feature-representation                    | Invariant feature by DNN with mutual information maximization                 |
|                  |                        | Li et al. [32]          | 2019 | Distribution-matching                     | Feature transformation by both marginal and conditional distribution matching |
|                  |                        |                         |      |                                           |                                                                               |
|                  | Pre-trial/Resting Data | Bolagh and Clifford [9] | 2017 | Instance                                  | Subject-transfer by Riemannian geometry                                       |
|                  |                        | Wei et al. [5]          | 2018 | Parameter                                 | Subject-transfer by multiple distance metrics                                 |
|                  |                        | Proposed                | 2020 | Instance                                  | Subject-transfer based on subject representation                              |

a learning framework was developed to learn a set of shared parameters from the entire group while tuning subject-dependent parameters individually. Jayaram *et al.* demonstrated a general transfer-learning framework for spatiotemporal feature learning that adapts pre-trained models to a new subject/session based on newly collected calibration data [20].

Recently, *feature-transformation* developed an approach to transform the features of the source and/or target and reduce the differences. There are two types of feature transformations for BCI, namely *baseline-aligning-based* and *distribution-matching-based*. First, the *baseline-aligning-based* transformation method aligns the features of data by removing subject-specific baselines so that the model can be trained from all the aligned data [24], [25], [28]. The *Distribution-matching-based* transformation method aims to transform the features of the domain into a latent subspace, where the differences of the feature distributions are small [12], [22], [26], [27], [32]. Also, both approaches can be integrated together to reduce the differences [23].

The *instance-transfer* approach weights the existing labeled data according to their values to the new subject. Dagois *et al.* and

Zhang *et al.* evaluated the values by measuring the differences in data distributions between the two subjects [11], [30].

The above-mentioned works assumed that the data of each label are available during the learning and are used as a prior knowledge to link two domains, which is a plausible assumption in the laboratory environment. However, it is not practical in the real-world BCIs. In the real world, transfer learning needs to be applied to detecting rare events of unseen users, such as the driver's drowsiness and the seizure onset of an outpatient. These are rare life-threatening events that might be difficult, if ever possible, to collect sufficient samples from each user/patient. Therefore, we proposed an instance transfer-based method, which learns the representations of subjects from the pre-trial space (e.g., data collected when the drivers are alert) and uses the subject-transfer approach to transfer the data.

### III. MAPPING EEG ONTO LOW-DIMENSIONAL SPACES

#### A. Background for Tensor and Tensor Decomposition

Tensors are a higher-order generalization of vectors and matrices, where the *order* represents the number of axes of the

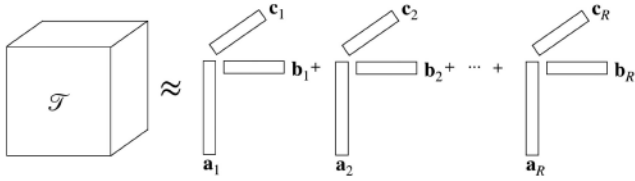


Fig. 2. Illustration of a CP decomposition. We approximated the tensor with a rank- $R$  tensor, where a rank- $R$  tensor refers to a tensor that is the summation of  $R$  rank-one tensor.

data, also known as the *multi-way arrays*. Tensors can store data in various domains as real-world data commonly present a multi-way nature [35]. For instance, tensor comes in handy when we store the EEG of a group of subjects. The PSD of EEG can form a three-order tensor, where each element represents the power of a frequency at a channel of a subject. In this case, the three axes are frequency, channel, and subject. Tensors provide an intuitive way to store and index these high-dimensional data.

Tensor decomposition is the factorization of a tensor into smaller matrices. Each matrix is considered a latent factor, which corresponds to the context of an axis in the original tensor, and the row vectors are the low-dimensional representations of the items along this axis. The following section introduces several types of tensor decomposition.

*Notations:* To be consistent with the literature, a tensor is represented in boldface caligraphic letters (e.g.,  $\mathcal{T}$ ). A matrix is represented by boldface uppercase letters (e.g.,  $\mathbf{A}$ ), and the  $r$ -th column vector of the matrix is represented by its boldface lowercase letter with  $r$  as subscript (e.g.,  $\mathbf{a}_r$ ). The  $j$ -th column vector of a matrix  $\mathbf{A}$  is represented as  $\mathbf{A}_{j:}$ . A scalar is represented by lowercase letters (e.g.,  $c$ ). We use  $\mathcal{T}_{i,j,k}$  to refer to the  $(i, j, k)$  entry of a third-order tensor  $\mathcal{T}$ .

To deal with different kinds of data, tensor decomposition and algorithms come in many forms. Among them, canonical polyadic decomposition (CPD) is one of the most common methods due to easy interpretation. Fig. 2 illustrates the CPD decomposition. Given a third-order tensor  $\mathcal{T} \in \mathbb{R}^{I \times J \times K}$ , we can decompose it into three latent-factor matrices  $\mathbf{A}, \mathbf{B}, \mathbf{C}$  by minimizing the loss function:

$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{C}} \left\| \mathcal{T} - \sum_r \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \right\|_F^2, \quad (1)$$

where  $R$  is the specified rank number and  $\circ$  represents the operation of the outer-product. Matrix  $\mathbf{A} \in \mathbb{R}^{I \times R}$  is interpreted as a latent factor corresponding to the first data mode, and the same concept applies to  $\mathbf{B}$  and  $\mathbf{C}$  as well. The row vectors in the matrices represent objects in the tensor axis. Finally, we can perform vector mathematical manipulation for comparing the objects with some metrics, such as distance metrics that can be applied to the vectors to obtain their similarity.

Besides getting the latent factors of tensor data, tensor decomposition has three other advantages. First, by applying tensor decomposition, large tensor data are compressed into compact latent factors, which implicitly keeps the main underlying components and eliminates the noise by a certain degree. Secondly, constraints may be imposed on the latent factors for model

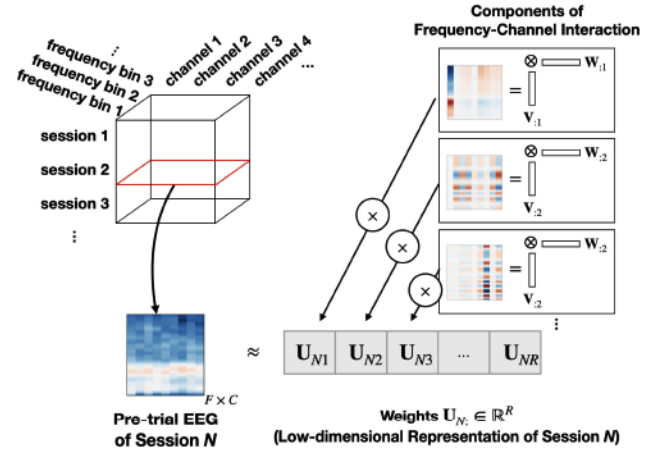


Fig. 3. Tensor decomposition decompose a session into weights of frequency-channel interactions that are the components of composing other sessions.

complexity control such as the sparsity of latent factors [36]. Lastly, tensor decomposition allows us to explore multi-way data without flattening the data and losing information. In other words, tensor decomposition can capture multi-way interactions in the data, which provides more information than the standard pairwise analysis [37]. We will provide a visualized explanation about the tensor decomposition on EEG in Section IV.

## B. Tensor Decomposition on EEG

This section demonstrates the effect of applying CP decomposition to a real EEG dataset. The dataset is composed of 54 sessions collected from 25 subjects, forming a  $54 \times 60 \times 30$  tensor, where 54 stands for the number of sessions, 60 stands for the number of frequency bins and 30 stands for the number of channels. We applied a CP decomposition of rank 10 on the tensor and obtained the latent factors as we described above. Take the latent factor that corresponds to the session axis as an example, the  $n$ -th row vector represents the  $n$ -th session in the latent space. The elements of the vector are the weights of components of the EEG, where the components are the patterns of channel-frequency interactions, as shown in Fig. 3. Every session is composed of many components with different weights. The same interpretation also applies to the latent factors of other modes.

To conduct a sanity check, we calculated the pairwise Pearson correlation on the subjects' latent factors, and visualize the results through a heat map Fig. 4. The brighter color indicates higher correlations between the two sessions. The average correlation of the latent factors is 0.45 for within subject, and -0.02 for cross subject. Generally speaking, the sessions from the same subject are more similar to each other and resulting brighter blocks along the diagonal line. This implies that the intra-subject variability is usually smaller than inter-subject variability, which is consistent with other literature [25]. However, this does not hold for every session. Some sessions of the same subjects have lower similarity, S5-2 and S5-3 for example, and some sessions across subjects have high similarity, 22-4 and 4-1 for example.

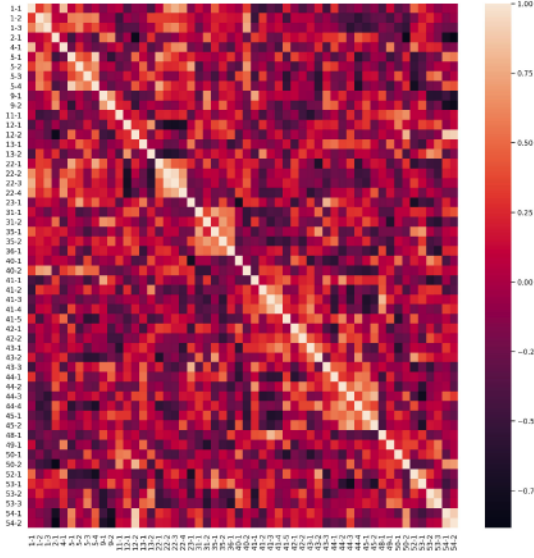


Fig. 4. Correlation of latent representations of sessions/subjects visualized on a heat map. The brighter color represents higher correlations between two sessions. The sessions from the same subject are more similar to each other.

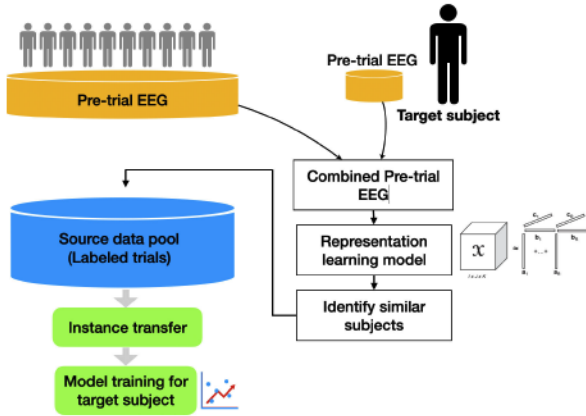


Fig. 5. The flow diagram of the proposed subject-transfer framework.

#### IV. REPRESENTATION-BASED SUBJECT-TRANSFER FRAMEWORK

The proposed subject-transfer framework transfers the labeled data from pre-existing subjects to the new subject (the target) based on their similarities. Assuming that a group of similar subjects can share a set of labeled data for model training, the proposed framework discovers similar subjects by first learning the representations of subjects and then calculating the similarity with the new subject based on the representations. Finally, it can transfer suitable labeled data for the new subject based on the similarity.

Fig. 5 shows the workflow diagram of the subject-transfer framework.

- First, the pre-trial EEG data are collected from each subject in the source pool and target.

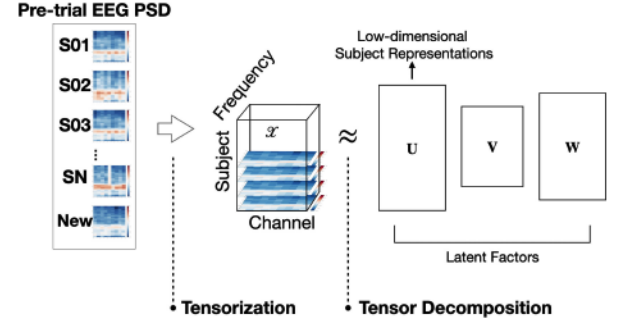


Fig. 6. The pre-trial EEG are stacked into a tensor and then decomposed to obtain low-dimensional subject representations. First, the pre-trial EEG PSD data are tensorized. The three modes of the tensor are the subjects, channels and frequency bins. An element of  $\mathcal{X}_{ijk}$  is the Log-power measurement of the  $k$ th frequency bin at  $j$ th channel of the  $i$ th subject. Then, tensor decomposition is applied to the tensor for learning the low-dimensional representations of subjects.

- Then, the pre-trial EEG data are processed into low-dimensional representations in our framework to learn the similarities among subjects.
- After obtaining the similarities of subjects, the framework performs data transfer based on the results, and train a prediction model for the new subject with these labeled data.

This section provides detailed information about each step of this framework.

##### A. Pre-trial EEG Tensorization

Pre-trial EEG is adopted to measure subject similarity in a recent study [5]. By “pre-trial EEG,” this study refers to the EEG data collected while the subjects were not performing tasks. We adopted it as the baseline for predicting individual differences and further select the best set of subjects for transfer learning. According to our previous study, within-subject variability exists in the dataset, indicating that the sessions within a subject might have different similarity to the target. Hence, each session in the source pool has a baseline for the similarity calculation. There are two advantages of using pre-trial EEG as a baseline. First, it can be collected without performing any tasks. Second, it only takes a short period to collect them. In our simulated driving task, collecting the pre-trial data only took two minutes for each new subject.

After obtaining the pre-trial EEGs, they are averaged and processed into logarithmic PSD to form a matrix for each subject, where the axes are channels and frequency bins. By stacking the matrices of existing subjects and the new subject all together, a three-mode tensor is formed, where the modes correspond to subjects, channels, frequency bins. The pre-trial EEG tensor is denoted by  $\mathcal{X} \in \mathbb{R}^{(N+1) \times E \times F}$ , where  $N$  is the number of existing subjects,  $E$  is the number of channels and  $F$  is the number of frequency bins. In other words, an element  $\mathcal{X}_{ijk}$  is the Log-power measurement of the  $k$ th frequency bin at the  $j$ th channel of the  $i$ th subject. The one added dimension on the subject axis is contributed by the new subject. Fig. 6 shows the tensorized EEG PSD data.

## B. Tensor-Decomposition-based Model for Learning Subject Representation

We applied tensor decomposition to the tensor that was obtained from the pre-trial EEG for learning subject representations, as shown in Fig. 6. Several research works have applied tensor decomposition to multi-way EEG data to obtain the features of prediction tasks [18], [38]. However, none of the previous works have used the latent factor to represent subjects and calculate their similarities. In fact, latent factors from matrix/tensor decomposition are heavily used in recommendation systems to discover users with similar behaviors [39], [40]. We apply this idea to our transfer learning framework for BCI to identify similar subjects who exhibit similar brain responses under the same conditions.

After forming a pre-trial EEG tensor, we now formulate the tensor decomposition model to obtain the low-dimensional representations of subjects. The objective function of our model is:

$$\mathcal{L}(\mathbf{U}, \mathbf{V}, \mathbf{W}) = \frac{1}{2} \|\mathcal{X} - [\mathbf{U}, \mathbf{V}, \mathbf{W}]\|^2 + \frac{\lambda_1}{2} (\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2 + \|\mathbf{W}\|_F^2), \quad (2)$$

where  $[\mathbf{U}, \mathbf{V}, \mathbf{W}]$  denotes the tensor reconstructed by the matrices  $\mathbf{U}, \mathbf{V}, \mathbf{W}$ , and subject latent factors  $\mathbf{U} \in \mathbb{R}^{(N+1) \times R}$ , channel latent factor  $\mathbf{V} \in \mathbb{R}^{E \times R}$  and frequency factor  $\mathbf{W} \in \mathbb{R}^{F \times R}$  with CP form. The first term in the objective function makes the reconstructed tensor and the original tensor as close as possible. The second term is a regularization term that regularizes the complexity of the model.

Minimizing this objective function is a non-convex optimization problem. We implemented an iterative gradient descent algorithm to solve it. The algorithm will iterate over a finite number of epochs to update the variables. In each epoch, the variables are updated by their gradients. We obtain partial gradients for all variables:

$$\nabla_{\mathbf{U}} \mathcal{L} = -(\mathcal{X}_{(1)} - [\mathbf{U}, \mathbf{V}, \mathbf{W}]_{(1)} \cdot (\mathbf{W} \odot \mathbf{V})) + \lambda_1 \mathbf{U} \quad (3)$$

$$\nabla_{\mathbf{V}} \mathcal{L} = -(\mathcal{X}_{(2)} - [\mathbf{U}, \mathbf{V}, \mathbf{W}]_{(2)} \cdot (\mathbf{W} \odot \mathbf{U})) + \lambda_1 \mathbf{V} \quad (4)$$

$$\nabla_{\mathbf{W}} \mathcal{L} = -(\mathcal{X}_{(3)} - [\mathbf{U}, \mathbf{V}, \mathbf{W}]_{(3)} \cdot (\mathbf{V} \odot \mathbf{U})) + \lambda_1 \mathbf{W}. \quad (5)$$

The partial gradients for CANDECOMP/PARAFAC (CP) decomposition term can be obtained from [41]. After obtaining the gradients, we implemented an iterative algorithm to approach the solution, as shown in Algorithm 1. The iterative algorithm has been studied and applied for solving tensor models in recent works [42]–[44]. Basically, the algorithm updates the latent factor matrices according to a proportion of their gradient in each iteration (Line 5). The proportion is referred to as the learning rate. This study implemented a decaying learning rate that was reduced by half after every *step\_size* iterations, which allowed the algorithm to approach the minimum in smaller steps. At the end of each iteration, the error between the original tensor and the tensor constructed with the latent factors was calculated by the Frobenius norm. The iteration was repeated until the error improvement was smaller than the tolerance  $\epsilon$ , so-called *converge*.

---

### Algorithm 1: CP Decomposition.

---

**input** : Baseline data  $\mathcal{X} \in \mathbb{R}^{(N+1) \times E \times F}$ , initial learning rate  $\eta_0$ , step size *step\_size*, tolerance  $\epsilon$ , regularization parameters  $\lambda_1$ , rank number  $R$   
**output**: Latent factors  $\mathbf{U} \in \mathbb{R}^{(N+1) \times R}$ ,  $\mathbf{V} \in \mathbb{R}^{E \times R}$  and  $\mathbf{W} \in \mathbb{R}^{F \times R}$

```

1 begin
2   initialize  $\mathbf{U}, \mathbf{V}, \mathbf{W}$  with SVD ;
3    $Lost_0 \leftarrow$  calculate by Eq. (2) ;
4    $cnt \leftarrow 0$  ;
5   while  $Lost_t - Lost_{t+1} > \epsilon$  do
6     calculate  $\nabla_{\mathbf{U}} \mathcal{L}, \nabla_{\mathbf{V}} \mathcal{L}$  and  $\mathbf{W}$  by Eq. (3), Eq. (4) and Eq. (5) ;
7      $\eta \leftarrow \eta_0 \times 0.5^{\lfloor \frac{cnt}{step\_size} \rfloor}$  ;
8      $\mathbf{U} = \mathbf{U} - \eta \nabla_{\mathbf{U}} \mathcal{L}$  ;
9      $\mathbf{V} = \mathbf{V} - \eta \nabla_{\mathbf{V}} \mathcal{L}$  ;
10     $\mathbf{W} = \mathbf{W} - \eta \nabla_{\mathbf{W}} \mathcal{L}$  ;
11     $Lost_{t+1} \leftarrow$  calculate by Eq. (2) ;
12  end
13 end
```

---

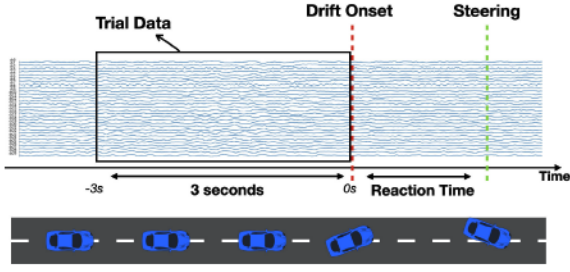
Finally, the proposed model learns the low-dimensional representations  $\mathbf{U}$  of the subjects from the original high dimensional data. Now we can use these representations to discover similar subjects of the target. The next section introduces the method for similarity calculation and subject selection.

## C. Subject Selection

When performing the same task, subjects with similar characteristics in a pre-trial EEG may have similar EEG characteristics. The low-dimensional representations  $\mathbf{U}$  are learned from the pre-trial EEG, so we can discover the subjects similar to the new subject by checking  $\mathbf{U}$ , and transfer their data to train the new subject model. In our framework, Pearson correlation is adopted to calculate the similarity between two latent factors. We calculate the correlation between the new subject and every other subject in the source pool. Then, the sessions in the pool are sorted by their similarities in descending order. After calculating the similarity ranking from  $\mathbf{U}$ , we start to transfer training data from the similar sessions to the new subject. First, we select the  $m$  most similar subjects by looking up the ranking, where  $m$  is an integer parameter specified by the experts and is referred to as the selection number in the following. Then, the labeled data of the selected subjects are collected to form a training dataset for the new subject to train a prediction model. The process of selecting subjects, gathering the data and training model is called *subject-transfer*.

## V. PERFORMANCE EVALUATION

This section evaluates the proposed method with a LOSO test, and compares the results with the baseline performance and the latest research. First, we introduce the dataset used for the evaluation. Second, we will briefly introduce the baseline and state-of-the-art methods. Finally, we will present the details of the experimental settings and their numerical results. The



**Fig. 7.** Overview of the driving experiment. Participants must drive the car and keep it in-lane by steering the steering wheel. Multiple drift events occur randomly, and the subject must steer the car back to the cruising position. The time elapsed between drift onset and subject steering back is called the reaction time. The EEG in the three-seconds window will be recorded with the reaction time as a trial.

evaluation framework and the proposed and baseline methods were implemented by Python 3.6. The results were visualized by Jupyter Notebook.

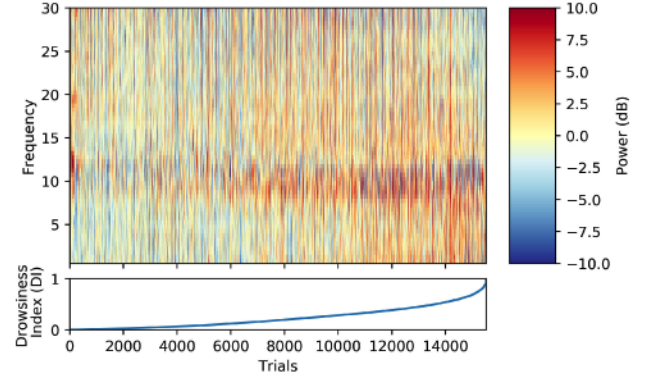
### A. Dataset and Preprocessing

This study exploits a public EEG dataset of a lane-keeping experiment in a simulated driving environment [45]. In this experiment, participants must drive the car and keep it in-lane by steering the steering wheel. During the experiment, multiple drift events occur randomly, and the subject must steer the car back to the cruising position. The time elapsed between drift onset and subject steering back is called the reaction time, which is related to the drivers' drowsiness level. We further transform the reaction time to *Drowsiness Index* (DI), which is a standardization indicator of the drowsiness level. Throughout the experiment, a Quick-30 headset recorded the subject's EEG at a sampling rate of 128 Hz. We segment the EEG in the three-second window before the car starting to drift. Fig. 7 shows an overview of the driving experiment. The task is to predict the reaction time by EEG before the drift events occur. In real-world applications, the system continuously monitors and processes the user's EEG data to predict the driver's drowsiness level to mitigate drowsiness and prevent accidents.

**Definitions of Terms:** A trial refers to a three-second window of EEG data before a car-drift event and the corresponding reaction time. Each subject was allowed to participate in this experiment multiple times, so there are multiple sessions from the same subject in the dataset.

The dataset contains 79 sessions collected from 37 healthy subjects. To evaluate the proposed method by a prediction task, some of the sessions which have insufficient trials were eliminated before the experiment. The requirements were as below: (1) The session has to contain at least 10 drowsy trials (i.e. the react time exceeds  $1.5 \times \mu_0$ , where  $\mu_0$  denotes the median react time of the first 10 trials). (2) The first 10 trials (the baseline) are all not drowsy trials (i.e. the react time does NOT exceed  $1.5 \times \mu_0$ ). As a result, the trimmed dataset includes 54 sessions from 25 different subjects, and a total of 15,516 3 s trials.

The dataset was preprocessed by band-passed filtering and PSD estimation. First, the dataset is epoched into 3 s trials prior to drift onsets. Each trial contains 384 data points because the



**Fig. 8.** The spectrogram of sorted PSD trials, showing that there exists a relation between PSD and reaction time in some but not all of the trials.

sampling rate is 128 Hz. Second, the EEG trials were converted into PSD, where PSD features are the power magnitude at each frequency bin. To obtain PSD, we adopted an open-source neurophysiological data analysis library - MNE [46], [47]. Third, we normalized each trial by subtracting the median PSD of the first 15 trials of the session. Fourth, we took the first 10 trials of a session as the pre-trial baseline, which were used for the subject-similarity calculation. To visualize the trials after pre-processing, Fig. 8 shows the spectrogram of all the trials sorted by their reaction time. This figure shows systematic changes in the PSD as a function of reaction time. Finally, all of the trials in the sessions that were selected by the subject-similarity test were used to train the drowsiness-prediction model (i.e. instance transfer).

### B. Procedures to Validate the Proposed Subject Transfer

We conducted a LOSO cross-validation to evaluate the performance of the proposed method and the competitors. In this evaluation study, each subject was designated as the target subject once, and other subjects were added to the source pool. The pre-trial EEG of the target was taken out as the baseline for finding similar subjects/sessions, and the same baseline was used for testing all methods. It is worth mentioning that when a session was selected as the target, all of the sessions from the same subject were excluded from the source pool. That is, the sessions would not be used as the training data. Fig. 9 shows an overview of the validation procedure.

To investigate the impacts of the number of sessions on the BCI performance using transfer learning, this study systematically varied the numbers of sessions to be selected from the source pool based on the subject similarity. This study compared the proposed method with two methods briefly introduced below:

**1) Random selection:** Random selection is the baseline method of this comparison. It randomly selects a specified number of sessions as training data. For each target, the selection test process was repeated 20 times and then averaged out as a result.

**2) Bolagh and Clifford [9]:** This study adopted a similar instance-transfer approach to select existing labeled data for new subjects. Bolagh and Clifford [9] mapped the subjects'

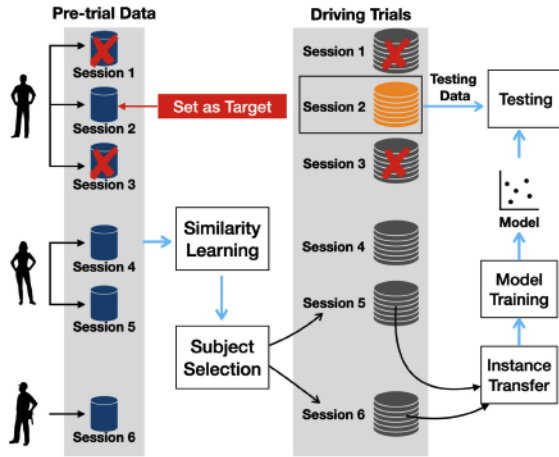


Fig. 9. The overview of our LOSO cross-validation. Each subject was designated as the target once, and other subjects were added to the source pool. If multiple sessions were recorded from the same subject, each session was set as the target once and other sessions from the same subject were excluded from the source pool.

covariance onto the Riemannian manifold to obtain subjects representation and applied Riemannian geometry to calculate subject similarity. Then, the labeled data from similar subjects were transferred for model training. Appendix A presents more information about covariance and Riemannian manifold.

## VI. RESULTS AND DISCUSSIONS

This section presents the experimental results of drowsiness prediction obtained by the proposed subject-transfer method. We compare the LDSR-TL method with the other methods by using the common source pool. The drowsiness-prediction performance was measured by the Pearson correlation coefficient of the predicted DIs ( $\hat{Y}$ ) and the actual DIs ( $Y$ ), where  $\text{corr}(Y, \hat{Y}) = \frac{\text{cov}(Y, \hat{Y})}{\sigma_Y \sigma_{\hat{Y}}}$ . The Pearson correlation coefficient can measure the normalized linear relationship between two numerical sequences, which can reflect how close the predicted DI sequence is to the actual DI sequence. We then compare the averaged Pearson correlation coefficient across all subjects for the Riemannian manifold method and random selection method.

The paired t-test was adopted in this study for testing the statistical significance of the results. The results of a method were the accuracies that were obtained from the LOSO experiment, which contains 54 measurements since each session was selected as target once. In each test, the results of two methods were paired together to test if the proposed method has significantly improved the prediction performance.

### A. Overall Performance Comparison

The overall performance of the three methods was assessed by the averaged correlation coefficients over all subjects based on LOSO cross validation. Fig. 10 plots the results of each method as a function of selection numbers. The results were obtained with using Bayesian Regression as regressor. The proposed method significantly outperformed the random selection (baseline) case at every selection number ( $p < 0.05$ ), indicating

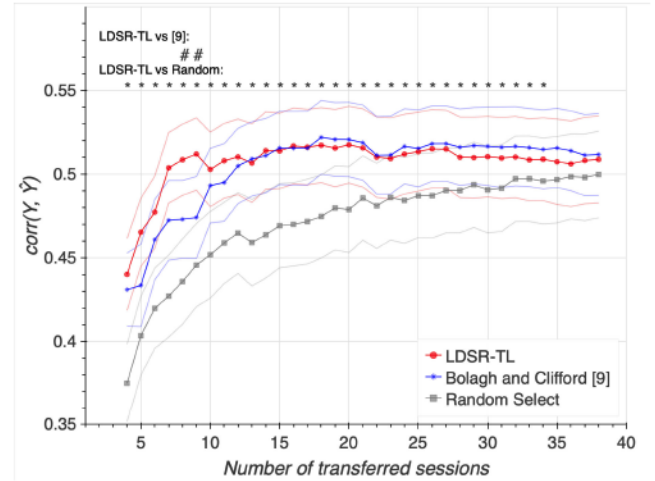


Fig. 10. Performance comparison of the proposed method and two baselines under using Bayesian Regression as regressor. These curves show the overall performance in the DI prediction against the number of transferred sessions that are selected by the proposed method and others. In general, the performance of all three approaches grows with more sessions being selected and transferred. Significant differences ( $p < 0.05$ ) in the overall performance assessed by paired t-test were marked by “\*” (LDSR-TL VS random) and “##” (LDSR-TL vs [9]). Our method showed its superiority when the training data were limited (8 selected sessions).

that the proposed method can improve the performance in most settings. The prediction performance obtained by the proposed method can be 17% higher than that of the random selection baseline when the number of transferred sessions was equal to eight. Comparing with the Riemannian manifold method, the proposed method performed significantly better with limited training data ( $p < 0.05$ ), where the selection numbers were eight and nine. The prediction performance can be improved up to 8% from the results obtained by the Riemannian manifold method when the selection number of transferred sessions was equal to eight. This result suggests that when the source data are limited, the proposed subject-transfer method works better than other approaches on selecting highly informative training data for efficient transfer learning.

Moreover, the proposed method reached the performance plateau rapidly as the training samples increased because the informative data have been all selected. The proposed method took the least number of selected sessions to reach the plateau, indicating it only required a small portion of the source pool to learn a good model for a new subject, and the additional data will not improve the prediction performance. Fig. 10 shows the proposed method reached  $\text{corr}(Y, \hat{Y}) = 0.51$  when the selected number of transferred sessions equaled nine, whereas the Riemannian manifold method reached  $\text{corr}(Y, \hat{Y}) = 0.51$  at the selected number of transferred sessions reached 15. This result implied that the proposed LDSR-TL can reduce 40% of training data compared to the Riemannian manifold method. Therefore, the emerging machine learning techniques, such as deep learning, can benefit from LDSR-TL because they are time- and computation-consuming at the training stage. When a larger dataset is available in the future, we will further investigate the effect of the size of the source pool. The performance decreased

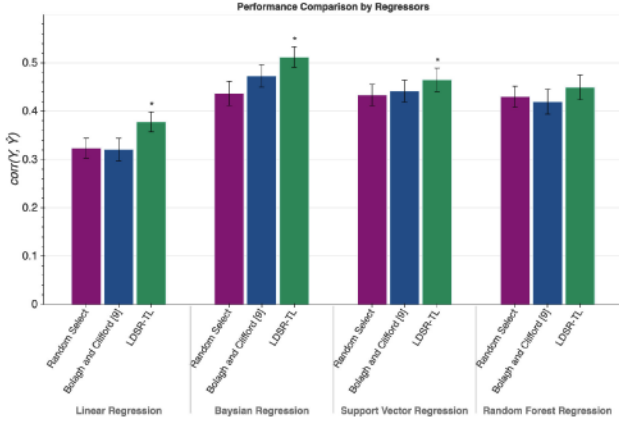


Fig. 11. A comparison of the overall drowsiness-prediction performance across different regressors. The number of selected sessions was fixed at eight. Significant differences ( $p < 0.05$ ) in the overall performance were assessed by paired t-test.

as the selected sessions continued to increase, implying that there are no more good sessions left in the source pool. Hence, the proposed method was forced to select the bad sessions. As the selection number continued to increase, the prediction performance of the three methods converged because the number of selected sessions was close to the size of the source pool. Note that corrections for multiple comparisons was not made because the goal of the experiment in this paper was to investigate the variation of prediction performance with respect to the number of transferred sessions. As the number of transferred sessions increases, the results of the paired t-test in this experiment become less significant because the size of the source pool was limited as mentioned above.

### B. Performance Stability

There are many kinds of learning algorithms that can learn models with different approaches. It is important to understand whether the transferred data are informative for the learning algorithms. Fig. 11 shows the stability of the proposed methods. In this test, we used the three methods to select labeled data from eight similar subjects, and then trained the drowsiness-detection models with four different regressors, including Linear Regression, Bayesian Regression, Support Vector Regression and Random Forest Regression. The results showed that Bayesian Regression had the best overall prediction accuracy, followed by Support Vector Regression, Random Forest Regression and then Linear Regression. In the Bayesian Regression and Support Vector Regression tests, the proposed method performed best and the random selection performed worst. In the Random Forest Regression test, the performance differences were not obvious, but the proposed method had a better average prediction accuracy. In the Linear Regression test, the overall performance was relatively low, and the random selection and the Riemannian manifold methods were unable to distinguish from each other. The proposed method still outperformed the other two methods. Overall, the proposed method outperformed the other two methods in each case ( $p < 0.05$ ), indicating that the proposed method can transfer informative data for any kind of regressors.

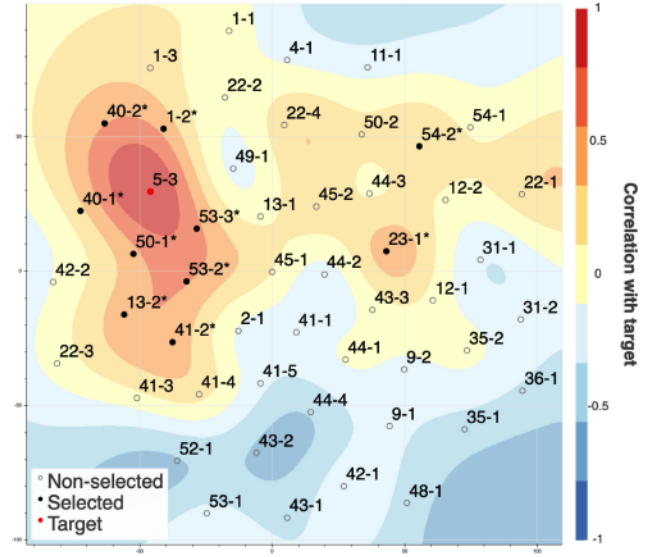


Fig. 12. Subject selection for session S5-3. The low-dimensional representations of subjects are visualized by t-SNE and the selected sessions are marked by an asterisk. The background is the similarity calculated by the proposed method. The figure shows that most of the selected sessions are close to S5-3, which indicates the proposed method is able to select the sessions with smaller variability with the new subject.

### C. Low-Dimensional Representation Visualization

To elucidate the effectiveness of the proposed subject-transfer framework, we take the session S5-3 as an example. By decomposing the pre-trial tensor, we can get the latent factors of the subjects. We select 10 sessions for subject transfer in this example. We can map the latent factors onto a 2-D space by t-distributed stochastic neighbor embedding (t-SNE) visualization [48], as shown in Fig. 12. The labels are the session indices and the selected sessions are marked by asterisks. The background is the similarity of subject representation between the target session and others. We can find that the representations of sessions from the same subject tend to be close to each other, indicating that intra-subject variability is smaller than inter-subject variability for some of the subjects. The distribution of the subject representations shows that the low-dimensional representation learned by tensor decomposition can preserve the similarity of EEG responses among subjects. The red point represents the new session S5-3. The figure shows that most of the selected sessions are close to S5-3, which indicates that the proposed method is able to select the sessions that are smaller to the new subject because they are proximate to S5-3 in the figure.

### D. Demonstration

Fig. 13 shows the drowsiness-prediction results obtained by the transfer-learning method for S5-3. The figure also plots the ground truth and the results from the other two methods for comparison. The DI increased and then decreased twice in this session as shown in the figure. The proposed method estimates the DI quite well because the time courses of the estimated DI matched closely with the measured DI (ground truth). The results

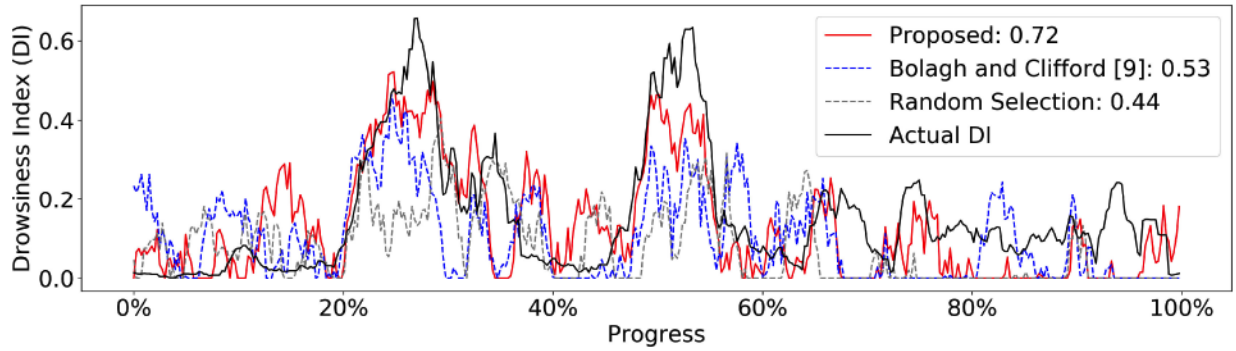


Fig. 13. Prediction results of proposed and baseline methods for the target session, S5-3. The red line is the predicted DI by the proposed method. The blue line is predicted by [9]. The gray line is predicted by the random selection method. The black line is the ground truth, the DI derived from the subject's behavioral responses. The correlations between the actual and predicted DIs are also listed. The proposed method selected the subjects who have similar EEG characteristics for transfer learning, so that the trained model was able to recognize drowsy episodes of the new subject.

showed that the proposed method had a 0.72 correlation with the ground truth, whereas [9] and random selection had 0.53 and 0.44, respectively. The proposed method selected the subjects who have similar EEG characteristics for transfer learning, so that the trained model could recognize drowsy events of the new subject.

In summary, the proposed method outperformed others in selecting sessions that could improve transfer learning on new subjects. Furthermore, fewer sessions were required to train an accurate prediction model for a new subject, thereby avoiding the expensive cost of collecting a large amount of labeled data and reducing model-training time.

## VII. CONCLUSION

This study proposed a low-dimensional subject representation-based approach for transfer learning (LDSR-TL) in EEG decoding. The proposed LDSR-TL method transferred the existing labeled data based on the similarity among subjects, which is estimated according to the distance between the subject representations extracted by tensor decomposition of the pre-trial EEG data. Tensor decomposition can factorize a subject's EEG data into the weights of channel-frequency components. As such, we designed a subject-transfer framework to leverage the low-dimensional representations in calculating the similarity between the new subject with the existing subject. Therefore, a new BCI user can benefit from the existing data collected from other users and a small number of data from a rapid, non-task, and unsupervised calibration. This study showed that the proposed LDSR-TL method outperformed the state-of-the-art Riemannian manifold method and the random selection baseline method. The results show that LDSR-TL can improve the prediction accuracy by up to 8% compared to the Riemannian manifold method.

The proposed method is a promising tool for both eliminating the requirements of supervised, task-relevant calibration data for a new BCI user and improving BCI performance. The future work includes applying the proposed LDSR-TL approach to other biomedical data analysis and applications, such as the electrocardiography (ECG) analysis that suffers from human variability.

## APPENDIX A

### CHANNEL COVARIANCE AND RIEMANNIAN GEOMETRY

Channel covariance is a feature which measures the correlation of channels which is a way of representing EEG signals. It has been reported as a good feature for classification tasks [49]–[52]. A trial of EEG is represented by a covariance matrix, where each element represents the correlation of two channels. The matrices live on the Riemannian manifold where Riemannian geometry is applied. Therefore, the distance between two data points is calculated by Riemannian distance instead of Euclidean distance. The distance is further used for classifying brain activities. In the following, we will briefly introduce how to obtain covariance matrices and calculate Riemannian distance.

Given a trial of EEG data  $X \in \mathbb{R}^{E \times n}$ , where  $E$  represents the number of electrodes and  $n$  represents the number of data points in the corresponding trial. The covariance matrix  $C$  of the trial is calculated by:

$$C = \frac{1}{n-1} XX^T. \quad (6)$$

To calculate the distance between trials on the Riemannian manifold, Riemannian distance is adopted. The distance between two trials (two covariance matrices  $C_1, C_2$ ) is calculated by:

$$\delta(C_1, C_2) = \sqrt{\sum_i \log^2 \lambda_i(C_1^{-1}C_2)}, \quad (7)$$

where  $\lambda_i(\cdot)$  represents the  $i$ th eigenvalue of  $\cdot$ . Note that all covariance matrices are symmetric positive definite (SPD) matrices, and the inverse of SPD matrix is also a SPD matrix. In addition, the product of two SPD matrices does not have negative eigenvalues. Therefore, no errors will ever be induced by the  $\log(\cdot)$  function.

## REFERENCES

- [1] Y.-P. Lin *et al.*, "EEG-based emotion recognition in music listening," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1798–1806, Jul. 2010.
- [2] S. M. Alarcão and M. J. Fonseca, "Emotions recognition using EEG signals: A survey," *IEEE Trans. Affect. Comput.*, vol. 10, no. 3, pp. 374–393, Jul/Sep. 2019.
- [3] W. Wu *et al.*, "An electroencephalographic signature predicts antidepressant response in major depression," *Nat. Biotechnol.*, vol. 38, no. 4, pp. 439–447, 2020.

- [4] A.-M. Brouwer, M. A. Hogervorst, J. B. Van Erp, T. Heffelaar, P. H. Zimmerman, and R. Oostenveld, "Estimating workload using EEG spectral power and ERPs in the n-back task," *J. Neural Eng.*, vol. 9, no. 4, 2012, Art. no. 045008.
- [5] C.-S. Wei, Y.-P. Lin, Y.-T. Wang, C.-T. Lin, and T.-P. Jung, "A subject-transfer framework for obviating inter- and intra-subject variability in EEG-based drowsiness detection," *Neuroimage*, vol. 174, pp. 407–419, Jul. 2018.
- [6] F. Lotte *et al.*, "A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update," *J. Neural Eng.*, vol. 15, no. 3, Jun. 2018, Art. no. 031005.
- [7] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
- [8] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, pp. 1–40, May 2016.
- [9] S. N. G. Bolagh and G. D. Clifford, "Subject selection on a Riemannian manifold for unsupervised cross-subject seizure detection," Dec. 2017, *arXiv:1712.00465*.
- [10] W. Tu and S. Sun, "A subject transfer framework for EEG classification," *Neurocomputing*, vol. 82, pp. 109–116, Apr. 2012.
- [11] E. Dagois, A. Khalaf, E. Sejdic, and M. Akcakaya, "Transfer learning for a multimodal hybrid EEG-fTCD brain-computer interface," *IEEE Sensors Lett.*, vol. 3, no. 1, pp. 1–4, Jan. 2019.
- [12] J. Li, S. Qiu, Y.-Y. Shen, C.-L. Liu, and H. He, "Multisource transfer learning for cross-subject EEG emotion recognition," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3281–3293, Jul. 2020.
- [13] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, "On the surprising behavior of distance metrics in high dimensional spaces," in *Proc. Int. Conf. Database Theory*, 2001, pp. 420–434.
- [14] M. Verleysen and D. François, "The curse of dimensionality in data mining and time series prediction," in *Proc. Int. Conf. Artif. Neural Netw.: Comput. Intell. Bioinspired Syst.*, 2005, pp. 758–770.
- [15] T. Poggio, H. Mhaskar, L. Rosasco, B. Miranda, and Q. Liao, "Why and when can deep-but not shallow-networks avoid the curse of dimensionality: A review," *Int. J. Autom. Comput.*, vol. 14, no. 5, pp. 503–519, Oct. 2017.
- [16] T. Kolda and B. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, Aug. 2009.
- [17] E. Acar and B. Yener, "Unsupervised multiway data analysis: A literature survey," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 1, pp. 6–20, Jan. 2009.
- [18] F. Cong, Q.-H. Lin, L.-D. Kuang, X.-F. Gong, P. Astikainen, and T. Ristaniemi, "Tensor decomposition of EEG signals: A brief review," *J. Neurosci. Methods*, vol. 248, pp. 59–69, Jun. 2015.
- [19] M. Alamgir, M. Grosse-Wentrup, and Y. Altun, "Multitask learning for brain-computer interfaces," in *Proc. Int. Conf. Artif. Intell. Stat.*, 2010, pp. 17–24.
- [20] V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf, and M. Grosse-Wentrup, "Transfer learning in brain-computer interfaces," *IEEE Comput. Intell. Mag.*, vol. 11, no. 1, pp. 20–31, Feb. 2016.
- [21] O. Özdenizci, Y. Wang, T. Koike-Akino, and D. Erdoğan, "Learning invariant representations from EEG via adversarial inference," *IEEE Access*, vol. 8, pp. 27 074–27 085, 2020.
- [22] Y.-Q. Zhang, W.-L. Zheng, and B.-L. Lu, "Transfer components between subjects for EEG-based driving fatigue detection," *Neural Inf. Process.*, pp. 61–68, 2015.
- [23] W. Zhang and D. Wu, "Manifold embedded knowledge transfer for brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 5, pp. 1117–1127, May 2020.
- [24] B. Reuderink, J. Farquhar, M. Poel, and A. Nijholt, "A subject-independent brain-computer interface based on smoothed, second-order baselining," in *Proc. Int. Conf. Eng. Med. Biol. Soc.*, Aug. 2011, pp. 4600–4604.
- [25] P. Zanini, M. Congedo, C. Jutten, S. Said, and Y. Berthoumieu, "Transfer learning: A Riemannian geometry framework with applications to brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 5, pp. 1107–1116, May 2018.
- [26] P. L. C. Rodrigues, C. Jutten, and M. Congedo, "Riemannian procrustes analysis: Transfer learning for brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 8, pp. 2390–2401, Aug. 2019.
- [27] M. Dai, S. Wang, D. Zheng, R. Na, and S. Zhang, "Domain transfer multiple kernel boosting for classification of EEG motor imagery signals," *IEEE Access*, vol. 7, pp. 49 951–49 960, 2019.
- [28] H. He and D. Wu, "Transfer learning for brain-computer interfaces: A Euclidean space data alignment approach," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 2, pp. 399–410, Feb. 2020.
- [29] D. Zhang, L. Yao, K. Chen, and J. Monaghan, "A convolutional recurrent attention model for subject-independent EEG signal analysis," *IEEE Signal Process. Lett.*, vol. 26, no. 5, pp. 715–719, May 2019.
- [30] X. Zhang *et al.*, "Individual similarity guided transfer modeling for EEG-based emotion recognition," in *Proc. IEEE Int. Conf. Bioinf. Biomed.*, Nov. 2019, pp. 1156–1161.
- [31] E. Jeon, W. Ko, J. S. Yoon, and H.-I. Suk, "Mutual information-driven subject-invariant and class-relevant deep representation learning in BCI," Oct. 2019, *arXiv:1910.07747*.
- [32] J. Li, S. Qiu, C. Du, Y. Wang, and H. He, "Domain adaptation for EEG emotion recognition based on latent representation similarity," *IEEE Trans. Cogn. Develop. Syst.*, vol. 12, no. 2, pp. 344–353, Jun. 2020.
- [33] J. He *et al.*, "Boosting transfer learning improves performance of driving drowsiness classification using EEG," in *Proc. Int. Workshop Pattern Recognit. Neuroimage*, Jun. 2018, pp. 1–4.
- [34] H. Kang, Y. Nam, and S. Choi, "Composite common spatial pattern for subject-to-subject transfer," *IEEE Signal Process. Lett.*, vol. 16, no. 8, pp. 683–686, Aug. 2009.
- [35] E. E. Papalexakis, C. Faloutsos, and N. D. Sidiropoulos, "Tensors for data mining and data fusion: Models, applications, and scalable algorithms," *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 2, pp. 16:1–16:44, Oct. 2016.
- [36] N. D. Sidiropoulos, L. D. Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3551–3582, Jul. 2017.
- [37] A. Cichocki *et al.*, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, Mar. 2015.
- [38] A. Onishi, A. H. Phan, K. Matsuoka, and A. Cichocki, "Tensor classification for p300-based brain computer interface," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Mar. 2012, pp. 581–584.
- [39] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, Aug. 2009.
- [40] C. Lin, L. Wang, and K. Tsai, "Hybrid real-time matrix factorization for implicit feedback recommendation systems," *IEEE Access*, vol. 6, pp. 21 369–21 380, 2018.
- [41] E. Acar, T. G. Kolda, and D. M. Dunlavy, "All-at-once optimization for coupled matrix and tensor factorizations," in *Proc. Mining and Learning with Graphs*, Aug. 2011.
- [42] S. Wang, L. He, L. Stenneth, P. S. Yu, and Z. Li, "Citywide traffic congestion estimation with social media," in *Proc. ACM SIGSPATIAL*, 2015, pp. 1–10.
- [43] W. Hu *et al.*, "MUSA: Wi-Fi AP-assisted video prefetching via tensor learning," in *Proc. IEEE/ACM Int. Symp. Qual. Service*, 2017, pp. 1–6.
- [44] H.-H. Shuai *et al.*, "Mining online social data for detecting social network mental disorders," in *Proc. Int. Conf. World Wide Web*, 2016, pp. 275–285.
- [45] Z. Cao, C.-H. Chuang, J.-K. King, and C.-T. Lin, "Multi-channel EEG recordings during a sustained-attention driving task," *Sci. Data*, vol. 6, no. 1, pp. 1–8, Apr. 2019.
- [46] A. Gramfort *et al.*, "MEG and EEG data analysis with MNE-Python," *Front. Neurosci.*, vol. 7, pp. 1–13, Dec. 2013.
- [47] A. Gramfort *et al.*, "MNE software for processing MEG and EEG data," *Neuroimage*, vol. 86, pp. 446–460, Feb. 2014.
- [48] L. v. d. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.
- [49] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain-computer interface classification by Riemannian geometry," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 4, pp. 920–928, Apr. 2012.
- [50] F. Yger, M. Berar, and F. Lotte, "Riemannian approaches in brain-computer interfaces: A review," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1753–1762, Oct. 2017.
- [51] M. Congedo, A. Barachant, and R. Bhatia, "Riemannian geometry for EEG-based brain-computer interfaces: a primer and a review," *Brain-Comput. Interfaces*, vol. 4, no. 3, pp. 155–174, Jul. 2017.
- [52] D. Wu, B. J. Lance, V. J. Lawhern, S. Gordon, T.-P. Jung, and C.-T. Lin, "EEG-Based user reaction time estimation using Riemannian geometry features," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 2157–2168, Nov. 2017.