Self-Weighted Semi-Supervised Classification for Joint EEG-Based Emotion Recognition and Affective Activation Patterns Mining

Yong Peng, *Member, IEEE*, Wanzeng Kong, *Member, IEEE*, Feiwei Qin, Feiping Nie, *Senior Member, IEEE*, Jinglong Fang, Bao-Liang Lu, *Fellow, IEEE*, and Andrzej Cichocki, *Life Fellow, IEEE*

***Abstract*— In electroencephalography (EEG)-based affective brain–computer interfaces (aBCIs), there is a consensus that EEG features extracted from different frequency bands and channels have different abilities in emotion expression. Besides, EEG is so weak and non-stationary that easily causes distribution discrepancies for EEG data collected at different times; therefore, it is necessary to explore the affective activation patterns in cross-session emotion recognition. To address these two problems, we propose a self-weighted semi-supervised classification (SWSC) model in this article for joint EEG-based cross-session emotion recognition and affective activation patterns mining, whose merits include: 1) using both the labeled and unlabeled samples from different sessions for better capturing data characteristics; 2) introducing a self-weighted variable to learn the importance of EEG features adaptively and quantitatively; and 3) mining the activation patterns including the critical EEG frequency bands and channels automatically based on the learned self-weighted variable. Extensive experiments are conducted on the benchmark SEED_IV emotional dataset and SWSC obtained excellent average accuracies of 77.40%, 79.55%, and 81.52% in three cross-session emotion recognition tasks. Moreover, SWSC identifies that the Gamma frequency band contributes the most and the EEG channels in prefrontal, left/right temporal, and (central) parietal lobes are more important for cross-session emotion recognition.**

***Index Terms*— Affective patterns mining, electroencephalography (EEG), emotion recognition, feature self-weighting, semi-supervised classification.**

I. Introduction

EMOTION recognition plays a central role in aBCIs, which aims at assigning machines the ability of accurately recognizing human emotions. In practical applications, there are basically four different data modalities to perform emotion recognition, that is, facial expression, text, speech, and physiological signals [1]. Though emotion recognition is easy to realize by the former three types of data sources, they have two shortcomings. On the one hand, sometimes it is difficult to recognize the true emotional state because these data modalities are easy to disguise (e.g., detecting a smile face does not always indicate happiness because he/she can disguise an expression). On the other hand, automatic emotion recognition is impractical for the disabled. Therefore, it is necessary to develop more objective methods for emotion recognition. Since emotion refers to a state of mind that occurs spontaneously rather than consciously and is usually accompanied by physiological changes in central nervous and periphery, the physiological reactions and the corresponding signals are difficult to control when emotions are excited [2]. Therefore, physiological reactions have been widely used to determine and classify different kinds of emotions. However, peripheral physiological signals such as electrocardiography, galvanic skin response, heart rate variability, and respiration rate usually have slow change rates and therefore are not accurate enough to characterize the essence of emotion. With the rapid progresses in weak signal acquisition and analysis, electroencephalography (EEG) has attracted increasing attention in diverse fields such as BCIs, cerebral disorder, and disease diagnosis. It records the electrical activities of neural cells across the human cerebral cortex

Manuscript received July 28, 2021; revised September 26, 2021; accepted October 18, 2021. Date of publication November 1, 2021; date of current version November 12, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61971173, Grant 61972121, and Grant U20B2074; in part by the National Key Research and Development Program of China for the Intergovernmental International Science and Technology Innovation Cooperation Project under Grant 2017YFE0116800; in part by the Natural Science Foundation of Zhejiang Province under Grant LY21F030005 and Grant LY21F020015; in part by the National Social Science Foundation of China under Grant 19ZDA348; in part by the Fundamental Research Funds for the Provincial Universities of Zhejiang under Grant GK209907299001-008; in part by the China Postdoctoral Science Foundation under Grant 2017M620470; in part by the MoE Key Laboratory of Advanced Perception and Intelligent Control of High-end Equipment (Anhui Polytechnic University) under Grant GDSC202015; and in part by the Guangxi Key Laboratory of Optoelectronic Information Processing (Guilin University of Electronic Technology) under Grant GD21202. The Associate Editor coordinating the review process was Dr. Bardia Yousefi.
(Corresponding author: Jinglong Fang.)

Yong Peng and Wanzeng Kong are with the School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China, and also with the Zhejiang Key Laboratory of Brain-Machine Collaborative Intelligence, Hangzhou 310018, China.

Feiwei Qin and Jinglong Fang are with the School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China (e-mail: fjl@hdu.edu.cn).

Feiping Nie is with the School of Artificial Intelligence, OPtics and ElectroNics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, China.

Bao-Liang Lu is with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China.

Andrzej Cichocki is with the Center for Computational and Data-Intensive Science and Engineering, Skolkovo Institute of Science and Technology, 143026 Moscow, Russia.

Digital Object Identifier 10.1109/TIM.2021.3124056

and has been regarded as the most reliable clue for emotion recognition.

However, two fundamental problems in EEG-based emotion recognition are not sufficiently investigated by existing studies. First, EEG features are usually extracted from multiple frequency bands and channels, whose abilities in emotion expression should be differentiated. Second, due to the weak and non-stationary properties, data discrepancies are commonly found in EEG data collected from different sessions; therefore, it is necessary to explore the stable affective activation patterns over time. To this end, we propose a self-weighted semi-supervised classification (SWSC) model to jointly achieve EEG-based emotion recognition and affective activation patterns mining. SWSC adaptively measures the contributions of different EEG features by a newly introduced self-weighted variable. Moreover, based on the coupling relationship between each feature dimension and each EEG frequency band (channel), we can perform automatic activation patterns mining according to the learned self-weighted variable. That is, more important features corresponding to more important EEG frequency bands and channels contribute more in cross-session emotion recognition, which are considered as the stable affective activation patterns. By conducting extensive experiments on the emotional SEED_IV dataset, these declared superiorities of SWSC are well supported.

Compared with the existing studies, this article has the following contributions.

1) An effective semi-supervised learning model SWSC is proposed to achieve EEG-based cross-session emotion recognition by considering one labeled session and the other unlabeled session. In SWSC, the model variables are jointly optimized with the emotional states to unlabeled EEG samples.
2) SWSC obtains improved recognition performance by employing a self-weighted variable to differentiate the abilities of EEG features in cross-session emotion expression. Then, useful features are encouraged by being adaptively assigned large weights, while less useful ones are suppressed by small weights, leading to automatic EEG feature selection.
3) SWSC automatically identifies the more important EEG frequency bands and channels which are considered as stable affective patterns in cross-session emotion recognition based on the learned self-weighted variable. This is more elegant than the trial-and-error method widely used by existing studies.

The remainder of this article is organized as follows. Section II briefly reviews the background knowledge on EEG-based emotion recognition and some related techniques. In Section III, we provide the SWSC model formulation and optimization. Experiments are conducted to evaluate the effectiveness of SWSC in Section IV. Section V concludes the whole article.

Notations: Throughout the whole article, Greek letters such as λ, γ, and η are used to denote the model variables and parameters. Delta, Theta, Alpha, Beta, and Gamma represent the five EEG frequency bands.

II. Related Works

A. EEG-Based Emotion Recognition

Generally, existing studies on EEG-based emotion recognition can be roughly divided into three stages of data preprocessing, feature extraction, and recognition.

EEG data preprocessing includes sampling, filtering, and artifact removal, in order to provide high-quality data for subsequent analysis [3]. Usually, EEG features can be extracted from time, frequency, time-frequency, and spatial domains. Time-domain features are the most intuitive and the typical ones include the event-related potentials, statistics, energy, power, high-order zero-crossing analysis, instability index, and fractal dimension. Since they cannot reflect the frequency information, we usually transform EEG data from time domain to frequency domain and then decompose the multi-rhythm EEG data into several frequency bands based on which the features such as power spectral density, event-related synchronization/desynchronization, high-order spectrum, and differential entropy (DE) [4] can be calculated. To further explore the time-varying property of frequencies in EEG data, the short-time Fourier transform, wavelet transform, and Hilbert–Huang method are normally used for time–frequency transformation. Due to the multi-channel property of EEG, taking its spatial information into consideration is also beneficial. For example, given frequency-domain features, the spatial-frequency features can be calculated by a common spatial pattern. Differential asymmetry and rational asymmetry, respectively, refer to the difference and ratio of features on the left and right symmetric electrodes. Brain functional connectivity aims to build the connection among electrodes to explore the spatial information [5]. Commonly used feature extraction methods in EEG-based emotion recognition were reviewed in [6] and [7].

Recently, a lot of machine learning models were proposed for EEG-based emotion recognition. Zheng [8] proposed a group sparse canonical correlation analysis for simultaneous EEG channel selection and emotion recognition. Inspired by the joint optimization mode, a semi-supervised random vector functional link network was proposed for emotion classification from EEG signals [9]. In [10], a multiple feature fusion approach was proposed to combine the activation features and connectivity features for emotion recognition [11]. To model the inherent dependencies among EEG and peripheral signals, Restricted Boltzmann machine was employed for multimodal emotion recognition [12]. In [13], the stacked autoencoder was used to handle the linear EEG mixing and the Long Short-Term Memory Recurrent Neural Network was employed for emotion timing. A dynamical graph convolutional neural network was proposed to learn the intrinsic relationship among EEG channels, which facilitates the discriminative EEG feature extraction [14]. Considering that different brain regions play different roles in emotion expression, discriminative spatial–temporal EEG features were learned by deep learning models with regional-to-global mechanism [15]. Sometimes, deep neural networks can perform end-to-end emotion recognition, which directly take raw EEG data as input and output recognition results [16]. Some review studies summarized the recent advances in EEG-based emotion recognition [7], [17].

B. Related Techniques

EEG is typically multi-rhythm and multi-channel. Therefore, features extracted from different frequency bands or channels should correlate differently to the occurrence of mental tasks, for example, emotion and fatigue states. In [18], each EEG feature is weighted by the distance between it and its cluster centroid, which is a rule-based feature weighting technique designed for EEG-based sleep stage identification. To simultaneously search for the optimal feature weights and model parameters, a nature-inspired immune algorithm coupled with SVM was proposed for mental tasks classification [19]. Instead of discarding unreliable features, Mishuhina and Jiang [20] and Gaur *et al.* [21] proposed to use a short-tailed Gaussian function to weight the common spatial pattern features for EEG-based motor imagery. Cui *et al.* [22] proposed a feature weighted episodic training model to eliminate the calibration requirement in EEG-based drowsiness estimation by assigning different weights to different EEG channels.

Since feature discrepancies are commonly appeared in EEG data collected from different sessions [23], [24], it is meaningful to identify which frequency bands and channels the stable features are mainly from, that is, stable affective activation patterns. Zheng *et al.* [25] completed the identification of critical EEG frequency bands and channels in emotion expression by a trial-and-error manner; that is, by inputting a learning model with features extracted from different frequency bands or channels, the critical ones can be identified based on their recognition performance. In [26], the weight distribution of features was obtained by deep belief networks; however, the identification of critical frequency bands and channels was still performed manually and the deep neural networks employed a black-box training mode which has poor interpretation of obtained results [16]. In [27], supervised learning was employed to cross-session emotion recognition, which is less appropriate than semi-supervised paradigm in capturing data properties of both sessions [28].

Here, we specially set aside one paragraph to review the rescaled least-square regression (RLSR) model [29], [30] based on which we will make improvements and propose our model in Section III. RLSR achieves semi-supervised feature selection by quantitatively ranking the importance of features. Mathematically, RLSR introduces a scale factor vector $\boldsymbol{\theta}$ whose jth element θ_j measures the importance of the jth feature and can be adaptively learned from data. By incorporating such scale factor into the least-square regression, RLSR has the following objective function:

$$\min \left\|\mathbf{X}^T\boldsymbol{\Theta}\mathbf{W}+\mathbf{1}\mathbf{b}^T-\mathbf{Y}\right\|_2^2+\gamma\left\|\mathbf{W}\right\|_2^2$$
$$\text{s.t.}\ \ \mathbf{W},\mathbf{b},\quad \boldsymbol{\theta}\geq\mathbf{0},\ \boldsymbol{\theta}^T\mathbf{1}=1,\ \mathbf{Y}_u\geq\mathbf{0},\ \mathbf{Y}_u\mathbf{1}=\mathbf{1}. \quad (1)$$

In (1), $\mathbf{X}=[\mathbf{X}_l,\mathbf{X}_u]$ is the combination of both labeled and unlabeled samples, and $\boldsymbol{\Theta}$ is a diagonal matrix with its ith diagonal element $\boldsymbol{\Theta}_{ii}=(\theta_i)^{1/2}$. $\mathbf{Y}=[\mathbf{Y}_l;\mathbf{Y}_u]$ contains labels, respectively, corresponding to labeled and unlabeled samples, in which $\mathbf{Y}_u$ are relaxed from binary to real values in $[0,1]$. After describing the model formulation and optimization of our proposed SWSC model, we will provide detailed comparisons between SWSC and RLSR in Section III-D.

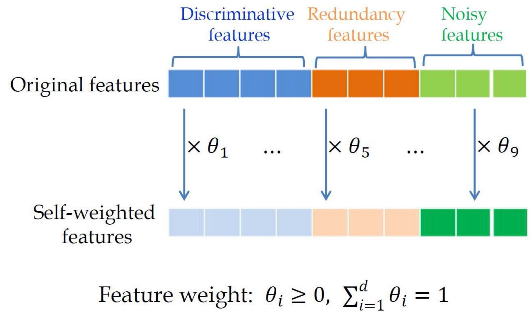

Fig. 1. Feature self-weighting scheme.

III. METHODS

A. Problem Definition

Below we define related terminologies in semi-supervised cross-session emotion recognition from EEG signals. EEG samples from source session are fully labeled, while the ones from the target session are unlabeled. We use $\mathbf{X}_l=[\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_l]\in\mathbb{R}^{d\times l}$ to denote the l labeled EEG samples which are associated with emotional labels $\mathbf{Y}_l=[\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_l]^T\in\mathbb{R}^{l\times c}$ in one-hot encoding. d is the dimensionality of the EEG sample vector and c is the number of emotional states. If $\mathbf{x}_i|_{i=1}^{l}$ is from the jth $(1\leq j\leq c)$ class, then only the jth element in $\mathbf{y}^i$ is 1 and all the others are 0 s. Similarly, we use $\mathbf{X}_u=[\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_u]\in\mathbb{R}^{d\times u}$ to represent the u unlabeled EEG samples, whose labels are defined in $\mathbf{Y}_u\in\mathbb{R}^{u\times c}$. However, $\mathbf{Y}_u$ is unknown and to be estimated.

Our task is twofold. One is to estimate the emotional states of target session EEG samples as accurate as possible; the other is to learn the importance of EEG features, based on which we expect to quantitatively figure out which EEG frequency bands and channels these more important features are from. Similar to [25], we informally name these critical EEG frequency bands (channels) as stable affective activation patterns in cross-session emotion recognition.

B. SWSC Model Formulation

Define $\mathbf{X}=[\mathbf{X}_l,\mathbf{X}_u]\in\mathbb{R}^{d\times n}$ and $\mathbf{Y}=[\mathbf{Y}_l;\mathbf{Y}_u]\in\mathbb{R}^{n\times c}$, where $n=l+u$. In pattern recognition, there is a consensus that different features should have different discriminative capabilities in expressing the semantic information of samples (i.e., emotional states). Roughly, roles of different features can be divided into three groups: discriminative features, redundant features, and noisy features. Inspired by this common view, we mathematically introduce a self-weighted variable $\boldsymbol{\theta}\in\mathbb{R}^d$ such that $\boldsymbol{\theta}\geq0$ and $\boldsymbol{\theta}^T\mathbf{1}=1$, to quantitatively measure the importance of EEG features, as shown in Fig. 1. Therefore, variable $\boldsymbol{\theta}$ is expected to adaptively assign larger weights to discriminative features but smaller weights to noisy features. Then, the discriminative capacity of features can be encouraged, while the negative influence of noisy features can be suppressed.

But how to incorporate $\boldsymbol{\theta}$ into an appropriate model and finish the learning? In this article, the least-square regression-based semi-supervised classification model is used due to its simplicity and effectiveness. Then, the objective function of the proposed SWSC model is formulated as

$$\min_{\boldsymbol{\Theta},\mathbf{V},\mathbf{b},\mathbf{Y}_u} \left\|\mathbf{X}^T\boldsymbol{\Theta}\mathbf{V}+\mathbf{1}\mathbf{b}^T-\mathbf{Y}\right\|_2^2+\lambda\|\mathbf{V}\|_2^2+\gamma\mathcal{R}(\mathbf{Y})$$
$$\text{s.t. } \mathcal{C}(\mathbf{Y}_u),\quad \boldsymbol{\Theta}=\operatorname{diag}\left(\sqrt{\boldsymbol{\theta}}\right),\ \boldsymbol{\theta}\geq\mathbf{0},\ \boldsymbol{\theta}^T\mathbf{1}=1 \tag{2}$$

where $\operatorname{diag}(\boldsymbol{\theta})$ means reshaping vector $(\boldsymbol{\theta})^{1/2} = [(\theta_1)^{1/2},(\theta_2)^{1/2},\ldots,(\theta_d)^{1/2}]$ as a diagonal matrix, λ and γ are two regularization parameters. Below we discuss the concrete forms to the regularizer $\mathcal{R}(\mathbf{Y})$ and the constraint $\mathcal{C}(\mathbf{Y}_u)$.

The local invariance idea said that if two EEG samples are more similar, they should belong to same emotional state with greater probability. If $\mathbf{S}=[s_{ij}]\in\mathbb{R}^{n\times n}$ is a graph adjacency matrix to depict the connection between samples, we define the regularizer $\mathcal{R}(\mathbf{Y})$ as

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\|\mathbf{y}_i-\mathbf{y}_j\|_2^2 s_{ij}=\operatorname{Tr}\left(\mathbf{Y}^T\mathbf{L}\mathbf{Y}\right) \tag{3}$$

where $\mathbf{L}$ is the Laplacian matrix corresponding to $\mathbf{S}$. $\mathbf{L}$ can be calculated by $\mathbf{D}-\mathbf{S}$, and $\mathbf{D}$ is a diagonal matrix whose ith diagonal element d_{ii} is defined by $\sum_{j=1}^{n}s_{ij}$. For simplicity, we employ the '0-1' weighting scheme to construct the graph; specifically, we define

$$s_{ij}=\begin{cases}1, & \text{if } \mathbf{x}_i\in\mathcal{N}_{k1}(\mathbf{x}_j) \text{ or } \mathbf{x}_j\in\mathcal{N}_{k1}(\mathbf{x}_i)\\ 0, & \text{otherwise}\end{cases} \tag{4}$$

where $\mathcal{N}_{k1}(\mathbf{x}_i)$ denotes the $k1$ nearest neighbors of $\mathbf{x}_i$. In experiments, we set $k1$ as 10.

On the constraint $\mathcal{C}(\mathbf{Y}_u)$, by investigating the ground-truth label matrix $\mathbf{Y}_l$ whose each row has only one non-zero element to indicate the class assignment, we constrain $\mathbf{Y}_u$ to be orthogonal, that is, $\mathbf{Y}_u^T\mathbf{Y}_u=\mathbf{I}_u$. Moreover, it is better to enforce the elements in $\mathbf{Y}_u$ to be non-negative rather than mix-signed [31]. Under both constraints, it is expected that only one non-zero element in each row of $\mathbf{Y}_u$ whose desirable value is 1, depicting the emotional state of a certain EEG sample.

Based on the above analysis, by denoting $\mathbf{W}=\boldsymbol{\Theta}\mathbf{V}$, we have $\mathbf{V}=\boldsymbol{\Theta}^{-1}\mathbf{W}$ and then the objective function of SWSC can be rewritten as

$$\min \left\|\mathbf{X}^T\mathbf{W}+\mathbf{1}\mathbf{b}^T-\mathbf{Y}\right\|_2^2+\lambda\|\boldsymbol{\Theta}^{-1}\mathbf{W}\|_2^2+\gamma\operatorname{tr}\left(\mathbf{Y}^T\mathbf{L}\mathbf{Y}\right)$$
$$\text{s.t. } \mathbf{Y}_u\geq\mathbf{0},\quad \mathbf{Y}_u^T\mathbf{Y}_u=\mathbf{I},\ \boldsymbol{\Theta}=\operatorname{diag}\left(\sqrt{\boldsymbol{\theta}}\right),\ \boldsymbol{\theta}\geq\mathbf{0},\ \boldsymbol{\theta}^T\mathbf{1}=1. \tag{5}$$

C. SWSC Model Optimization

There are four variables, $\boldsymbol{\Theta}$, $\mathbf{W}$, $\mathbf{b}$, and $\mathbf{Y}_u$, in the SWSC objective function (5). Accordingly, we propose to solve it by the coordinate blocking method, that is, the alternative method. The detailed updating rule to each variable is derived below.

1) Update $\mathbf{b}$*:* By setting the derivative of (5) w.r.t. $\mathbf{b}$ as zero, we get the optimal solution to $\mathbf{b}$ as

$$\mathbf{b}=\frac{1}{n}\left(\mathbf{Y}^T\mathbf{1}-\mathbf{W}^T\mathbf{X}\mathbf{1}\right). \tag{6}$$

2) Update $\mathbf{W}$*:* By taking its derivative of (5) with respect to $\mathbf{W}$ and setting it to zero, we have

$$\mathbf{W}=\left(\mathbf{X}\mathbf{X}^T+\lambda\boldsymbol{\Theta}^{-2}\right)^{-1}\mathbf{X}\left(\mathbf{Y}-\mathbf{1}\mathbf{b}^T\right). \tag{7}$$

By substituting $\mathbf{b}$ in (7) with (6), we can get the simplified updating rule to $\mathbf{W}$ as

$$\mathbf{W}=\left(\mathbf{X}\mathbf{H}\mathbf{X}^T+\lambda\boldsymbol{\Theta}^{-2}\right)^{-1}\mathbf{X}\mathbf{H}\mathbf{Y} \tag{8}$$

where $\mathbf{H}=\mathbf{I}-(1/n)\mathbf{1}\mathbf{1}^T$ is the centering matrix.

3) Update $\boldsymbol{\Theta}$*:* We find that $\boldsymbol{\Theta}$ is only involved in the second term of objective (5). The corresponding objective $\mathcal{O}(\boldsymbol{\Theta})$ is equivalent to

$$\min_{\boldsymbol{\theta}\geq\mathbf{0},\boldsymbol{\theta}^T\mathbf{1}=1}\left(\theta_1^{-1},\theta_2^{-1},\ldots,\theta_d^{-1}\right)\begin{pmatrix}\|\mathbf{w}^1\|_2^2\\ \|\mathbf{w}^2\|_2^2\\ \cdots\\ \|\mathbf{w}^d\|_2^2\end{pmatrix}=\min_{\boldsymbol{\theta}\geq\mathbf{0},\boldsymbol{\theta}^T\mathbf{1}=1}\sum_{j=1}^{d}\frac{\|\mathbf{w}^j\|_2^2}{\theta_j} \tag{9}$$

where $\mathbf{w}^j$ is the jth row of $\mathbf{W}$. Its Lagrangian function is

$$\mathcal{L}(\boldsymbol{\theta},\alpha,\boldsymbol{\beta})=\sum_{j=1}^{d}\frac{\|\mathbf{w}^j\|_2^2}{\theta_j}+\alpha\left(\sum_{j=1}^{d}\theta_j-1\right)+\boldsymbol{\beta}^T\boldsymbol{\theta} \tag{10}$$

where α and $\boldsymbol{\beta}$ are Lagrangian multipliers in scalar and vector forms, respectively. By taking the derivative of $\mathcal{L}(\boldsymbol{\theta},\alpha,\boldsymbol{\beta})$ w.r.t. θ_j and setting it to zero, we have

$$\frac{\partial\mathcal{L}(\boldsymbol{\theta},\alpha,\boldsymbol{\beta})}{\partial\theta_j}=-\frac{\|\mathbf{w}^j\|_2^2}{\theta_j^2}+\alpha=0 \tag{11}$$

which leads to

$$\theta_j=\left(\frac{1}{\alpha}\|\mathbf{w}^j\|_2^2\right)^{\frac{1}{2}}. \tag{12}$$

We can substitute (12) into the normalization constraint $\boldsymbol{\theta}^T\mathbf{1}=1$ to remove the Lagrangian multiplier α. Then, we get the updating rule to $\boldsymbol{\theta}$ as

$$\theta_j=\frac{\|\mathbf{w}^j\|_2}{\sum_{v=1}^{d}\|\mathbf{w}^v\|_2}. \tag{13}$$

4) Update $\mathbf{Y}_u$*:* We first split the Laplacian matrix $\mathbf{L}$ into four blocks after the lth row and column as $\mathbf{L}=\begin{bmatrix}\mathbf{L}_{ll} & \mathbf{L}_{lu}\\ \mathbf{L}_{ul} & \mathbf{L}_{uu}\end{bmatrix}$. The objective in terms of variable $\mathbf{Y}_u$ is

$$\min_{\mathbf{Y}_u\geq\mathbf{0},\mathbf{Y}_u^T\mathbf{Y}_u=\mathbf{I}}\left\|\mathbf{X}_u^T\mathbf{W}+\mathbf{1}\mathbf{b}^T-\mathbf{Y}_u\right\|_2^2$$
$$+\gamma\operatorname{tr}\left(2\mathbf{Y}_l^T\mathbf{L}_{lu}\mathbf{Y}_u+\mathbf{Y}_u^T\mathbf{L}_{uu}\mathbf{Y}_u\right). \tag{14}$$

By substituting $\mathbf{b}$ in objective (14) with (6) and considering that $\mathbf{H}_u = (1/u)(\mathbf{I}_u - \mathbf{1}\mathbf{1}^T)$ is an idempotent matrix, we have

$$\begin{aligned}&\min_{\mathbf{Y}_u\geq\mathbf{0},\mathbf{Y}_u^T\mathbf{Y}_u=\mathbf{I}} \left\|\mathbf{H}_u\left(\mathbf{X}_u^T\mathbf{W}-\mathbf{Y}_u\right)\right\|_2^2\\ &\qquad+\gamma\,\mathrm{tr}\left(2\mathbf{Y}_l^T\mathbf{L}_{lu}\mathbf{Y}_u+\mathbf{Y}_u^T\mathbf{L}_{uu}\mathbf{Y}_u\right)\\ &\Leftrightarrow \min_{\mathbf{Y}_u\geq\mathbf{0},\mathbf{Y}_u^T\mathbf{Y}_u=\mathbf{I}} \mathrm{tr}\left(\mathbf{Y}_u^T(\mathbf{H}_u+\gamma\,\mathbf{L}_{uu})\mathbf{Y}_u\right)\\ &\qquad-2\mathrm{tr}\left(\left(\mathbf{W}^T\mathbf{X}_u\mathbf{H}_u-\gamma\,\mathbf{Y}_l^T\mathbf{L}_{lu}\right)\mathbf{Y}_u\right)\\ &\triangleq \min_{\mathbf{Y}_u\geq\mathbf{0},\mathbf{Y}_u^T\mathbf{Y}_u=\mathbf{I}} \mathrm{tr}\left(\mathbf{Y}_u^T\mathbf{A}\mathbf{Y}_u\right)-2\mathrm{tr}(\mathbf{B}\mathbf{Y}_u) \end{aligned}\tag{15}$$

where $\mathbf{A} = \mathbf{H}_u + \gamma\,\mathbf{L}_{uu}$ and $\mathbf{B} = \mathbf{W}^T\mathbf{X}_u\mathbf{H}_u - \gamma\,\mathbf{Y}_l^T\mathbf{L}_{lu}$. By relaxing the nonnegative orthogonal constraint, we can rewrite objective (15) as

$$\min_{\mathbf{Y}_u\geq\mathbf{0}} \mathrm{tr}\left(\mathbf{Y}_u^T\mathbf{A}\mathbf{Y}_u\right)-2\mathrm{tr}(\mathbf{B}\mathbf{Y}_u)+\frac{\eta}{2}\left\|\mathbf{Y}_u^T\mathbf{Y}_u-\mathbf{I}\right\|_2^2 \tag{16}$$

where η is a parameter to control the orthogonality. When $\eta \to \infty$, the orthogonality is satisfied. Therefore, in the following experiments, we set it as a large enough value (i.e., 10^6). The Lagrangian function of objective (16) is

$$\min_{\mathbf{Y}_u} \mathrm{tr}\left(\mathbf{Y}_u^T\mathbf{A}\mathbf{Y}_u\right)-2\mathrm{tr}(\mathbf{B}\mathbf{Y}_u)+\frac{\eta}{2}\left\|\mathbf{Y}_u^T\mathbf{Y}_u-\mathbf{I}\right\|_2^2+\mathrm{tr}\left(\mathbf{\Lambda}\mathbf{Y}_u^T\right)$$

where $\mathbf{\Lambda} \geq \mathbf{0}$ is an Lagrangian multiplier. By taking its derivative with respect to $\mathbf{Y}_u$ and setting it to zero, we have

$$2\mathbf{A}\mathbf{Y}_u - 2\mathbf{B} + 2\eta\mathbf{Y}_u\mathbf{Y}_u^T\mathbf{Y}_u + \mathbf{\Lambda} = 2\eta\mathbf{Y}_u. \tag{17}$$

Based on the Karush–Kuhn–Tucker (KKT) condition $\mathbf{\Lambda} \circ \mathbf{Y}_u = \mathbf{0}$ (where $\circ$ is the Hadamard product), we could operate the variable $\mathbf{Y}_u$ on both sides of the above equation via Hadamard product and obtain

$$(\mathbf{Y}_u)_{ij} = (\mathbf{Y}_u)_{ij}\frac{(2\eta\mathbf{Y}_u+2\mathbf{B})_{ij}}{\left(2\mathbf{A}\mathbf{Y}_u+2\eta\mathbf{Y}_u\mathbf{Y}_u^T\mathbf{Y}_u\right)_{ij}}. \tag{18}$$

According to the constraint in objective (5), the matrix $\mathbf{Y}$ needs to be normalized, such that $(\mathbf{Y}_u^T\mathbf{Y}_u)_{ii} = 1, i = 1, 2, \ldots, c$ is satisfied.

Based on the above analysis, we summarize the whole procedure of optimizing the objective function of SWSC model in Algorithm 1.

Algorithm 1 The Optimization to SWSC Objective (5)

Input: Labeled EEG samples from source session $(\mathbf{X}_l, \mathbf{Y}_l) = \{\mathbf{x}_i, y_i\}|_{i=1}^l$, unlabeled EEG samples from target session (subject) $\mathbf{X}_u = \{\mathbf{x}_i\}|_{i=l+1}^n$, parameters λ and γ;

Output: The projection matrix $\mathbf{W}$ and the estimated emotional states $\mathbf{Y}_u$.

1: Initialize $\mathbf{Y}_u = \frac{1}{c}\mathbf{1}\mathbf{1}^T \in \mathbb{R}^{u\times c}$ and $\boldsymbol{\theta} = [\frac{1}{d}, \frac{1}{d}, \ldots, \frac{1}{d}] \in \mathbb{R}^d$;
2: **while** not converged **do**
3: Update variable $\mathbf{W}$ via rule (8);
4: Update variable $\boldsymbol{\theta}$ via rule (13);
5: Update variable $\mathbf{Y}_u$ via rule (18);
6: **end while**
7: Calculate $\boldsymbol{\theta} \in \mathbb{R}^d$ where $\theta_j = \frac{\|\mathbf{w}^j\|_2}{\sum_{v=1}^d \|\mathbf{w}^v\|_2}$;
8: Perform the affective activation patterns mining by the learned self-weighted variable $\boldsymbol{\theta}$.

D. Discussion

Having elaborated the model formulation and optimization of our proposed SWSC model, below we discuss the connections as well as differences between SWSC and RLSR [29], [30].

1) Connections: Different feature dimensions in a sample vector have different contributions in characterizing the sample semantic meaning (i.e., label information). Therefore, different features should have different abilities in recognition tasks. Both SWSC and RLSR aim to adaptively learn feature weights from data within the semi-supervised framework. Technically, our SWSC model is inspired by RLSR and follows the strategy of introducing a scale factor vector to measure the importance of different feature dimensions. After obtaining the importance values, all features can naturally be ranked and feature selection is completed.

2) Differences: There are multiple differences between these two models.

1) The most significant difference is that RLSR is a general machine learning model for semi-supervised feature selection, while our SWSC model is specially designed for cross-session EEG-based emotion recognition and affective activation patterns mining. Chen *et al.* [29] and [30] cared only about the recognition performance based on the selected features. They did not (and essentially had no way to) analyze the underlying meaning of the top ranked features from their used datasets. In this work, we not only expect to improve the emotion recognition performance by adaptively weighting EEG features, but also want to explore the critical EEG frequency bands and channels which generate more powerful features in cross-session emotion expression. This is because each EEG feature can be back traced to a specific frequency band and channel. Therefore, SWSC additionally provides cognitive significance in EEG-based emotion recognition.
2) The objective functions of RLSR and SWSC are different. In SWSC, we introduce a graph regularizer on variable $\mathbf{Y}_u$ to let it conform to the local invariance property, which leads to a new optimization procedure to variable $\mathbf{Y}_u$. In RLSR, $\mathbf{Y}_u$ can be obtained by calculating the Euclidean distance defined on a simplex constraint. In SWSC, it is solved by using the Lagrangian multiplier method together with KKT condition.
3) The constraints defined on $\mathbf{Y}_u$ are different. In RLSR, the combined constraint $\mathbf{Y}_u \geq \mathbf{0}$, $\mathbf{Y}_u\mathbf{1} = \mathbf{1}$ tries to learn a fuzzy indicator matrix [32]. That is, the non-zero elements in each row of $\mathbf{Y}_u$ act as the memberships of a certain sample to different classes. Differently, in current work, we want to uniquely determine the emotional state of each EEG sample by the non-negative and orthogonal constraints.

IV. Experimental Studies

A. Dataset and Experimental Setup

The benchmark SEED_IV emotional dataset http://bcmi.sjtu.edu.cn/~seed/seed-iv.html [33] was used in this article

TABLE I
CROSS-SESSION EMOTION RECOGNITION ACCURACIES (%) OF DIFFERENT MODELS

	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15	avg.
	session1→session2															
m1	37.50	83.89	51.44	54.33	52.88	44.23	71.63	65.38	77.28	41.59	48.80	43.63	57.21	78.61	88.82	59.81
m2	49.04	74.04	53.97	43.51	58.77	53.97	81.61	**78.49**	67.79	60.94	51.68	45.07	58.65	79.21	91.35	63.21
m3	78.13	81.01	75.84	52.64	59.01	41.59	79.93	67.55	63.22	53.37	51.56	42.43	59.62	**87.62**	88.94	65.50
m4	63.46	81.25	67.79	70.79	56.73	60.22	80.17	65.50	66.71	46.51	42.91	67.43	64.06	85.22	**98.56**	67.82
m5	75.48	83.53	**77.88**	68.51	54.33	54.45	83.77	74.76	64.42	47.00	60.82	56.49	61.18	85.22	**98.56**	69.76
m6	**79.21**	**90.26**	**77.88**	**79.57**	**81.85**	**69.23**	**92.31**	75.84	**78.25**	**53.97**	**61.06**	**68.27**	**68.03**	86.66	**98.56**	**77.40**
	session1→session3															
m1	50.85	72.02	41.73	54.01	50.24	84.67	65.45	84.55	59.49	34.79	62.53	30.17	54.62	63.38	84.06	59.50
m2	69.22	63.99	**63.75**	46.23	66.55	85.52	72.75	55.23	**69.10**	57.54	57.79	48.66	52.07	72.87	67.76	63.27
m3	80.90	83.94	53.65	60.58	65.09	75.67	71.05	69.95	49.15	50.85	61.19	35.52	46.84	71.05	83.21	63.91
m4	74.70	**92.21**	47.20	78.47	**75.67**	72.02	87.59	**92.46**	53.89	41.61	75.30	58.64	56.57	77.01	82.48	71.05
m5	80.78	91.00	57.06	80.29	72.51	77.13	80.66	83.21	53.77	41.85	71.65	67.64	60.95	79.44	93.07	72.73
m6	**87.35**	**92.21**	59.37	**83.21**	**75.67**	**90.88**	**91.85**	88.81	65.45	**61.07**	**76.76**	**68.73**	**63.50**	**93.07**	**95.38**	**79.55**
	session2→session3															
m1	66.79	82.12	61.07	62.04	62.04	67.40	81.14	73.97	56.20	67.03	53.65	64.84	50.85	82.12	85.28	67.77
m2	**78.10**	83.94	67.88	78.22	62.04	58.52	82.48	83.33	61.80	58.39	41.85	58.39	**61.80**	92.21	81.75	70.05
m3	66.67	86.01	60.71	73.48	73.72	65.82	87.83	74.21	51.22	60.83	58.64	58.76	62.41	**92.82**	78.71	70.12
m4	55.11	84.31	72.87	83.09	71.53	85.64	86.13	75.67	61.92	70.07	61.44	79.56	53.77	92.46	88.08	74.78
m5	55.96	85.16	74.57	82.85	68.25	85.89	88.93	77.62	69.10	74.45	58.52	77.13	54.38	86.01	88.08	75.13
m6	73.36	**87.47**	**78.71**	**90.27**	**79.44**	**87.10**	**89.78**	**88.44**	**74.70**	**75.91**	**67.88**	**80.78**	61.56	91.97	**95.38**	**81.52**

Note: 's1' to 's15' are the indices of subjects. m1, m2, m3, m4, m5, and m6 respectively represent the compared models of SVM, GFHF, semiSVM, semiLSR, RLSR and SWSC.

since it is suitable for cross-session emotion recognition research. It is a video-evoked dataset. Seventy-two video clips were carefully chosen to elicit four different types of emotional states, that is, sad, fear, happy, and neutral. Each video clip lasts about 2 min. Fifteen subjects were recruited for EEG data collection and for each subject, the EEG data collection experiment was conducted at three different times, corresponding to three sessions. Therefore, there are totally 45 sessions corresponding to the 15 subjects, that is, each subject has three sessions. In each session, there are 24 trials corresponding to the 24 video clips. In each trial, there are three stages, that is, hint of start, video playing, and self-assessment. EEG data was recorded by the ESI NeuroScan system with a 62-channel electrode cap according to the international 10–20 system placement. The sampling rate is 1000 Hz.

After artifact removal and down-sampling, features were extracted from five frequency bands, Delta (1–4 Hz), Theta (4–8 Hz), Alpha (8–14 Hz), Beta (14–31 Hz), and Gamma (32–50 Hz). SEED_IV provides us with multiple features such as the power spectra density (PSD) and DE. In the following experiments, we use the DE feature smoothed by the linear dynamic systems because it has shown excellent performance in many existing studies. By concatenating the 62 points of each of the five frequency bands, the final dimensionality of EEG samples is 310. Each session has approximately 830 samples since the time durations of video clips are slightly different.

To evaluate the effectiveness of SWSC, we compare it with several state-of-the-art semi-supervised classification models including the semi-supervised support vector machine (semiSVM), Gaussian Field and Harmonic Functions (GFHF) [34], RLSR [29], [30], and semi-supervised Linear Square Regression (semiLSR) which is a degenerated version of RLSR by discarding the self-weighted variable. As a baseline method, we also included the supervised SVM in comparison by using the labeled samples from the source session as training and the unlabeled samples from the target session as test. In both SVMs, linear kernel was used. The related parameters (C in SVMs, λ in LSRs, and λ and γ in SWSC) were searched from $\{2^{-10}, 2^{-9}, \ldots, 2^{10}\}$. The graph adjacency matrix in SWSC was built by using the "0–1" weighting scheme and the neighborhood size was set as 10.

B. Emotion Recognition Results and Analysis

In this article, we only consider the following three cross-session tasks in chronological order, "session 1 → session 2" (i.e., samples from session 1 are labeled and samples from session 2 are unlabeled), "session 1 → session 3" and "session 2 → session 3." Table I presents the recognition results of these six algorithms where the best value in each case is highlighted in boldface. From these results, we have the following findings.

1) Obviously, the performance of semi-supervised models is better than that of the supervised model. Concretely, SVM obtained the average accuracies of 59.81%, 59.50%, and 67.77% in the three cross-session recognition tasks. Even the worst semi-supervised model, GFHF, in our experiments, it achieved the average accuracies of 63.21%, 63.27%, and 70.05% in the three tasks which, respectively, made 3.40%, 3.77%, and 2.28% improvements in comparison with SVM. This benefits from involving the unlabeled EEG samples in learning process and then semi-supervised models can better characterize the underlying data properties.
2) By comparing the results of semiLSR and RLSR, the efficacy of feature self-weighting has been fully

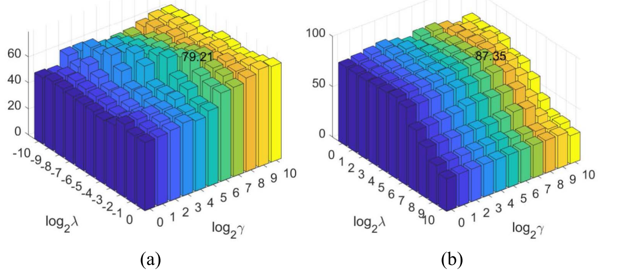

Fig. 2. Recognition accuracy (%) of SWSC in terms of parameters (λ, γ) on subject 1. (a) s1: session 1 $\rightarrow$ session 2. (b) s1: session 1 $\rightarrow$ session 3.

depicted. To be specific, RLSR obtained the average accuracies of 69.76%, 72.73%, and 75.13% in the three tasks, which improves the performance of semiLSR by 1.94%, 1.68%, and 0.35%, respectively. Based on these obtained results, we can conclude that efficiently exploring the different contributions of EEG features in cross-session emotion expression is beneficial for improving the recognition performance.

3) By comparing the results of RLSR and SWSC, it is obvious that SWSC achieved better performance than RLSR in 44 out of the whole 45 cases. Besides, the results generally show the effectiveness of the graph regularizer in preserving the data local invariance. In Fig. 2, it shows the recognition performance of SWSC in terms of different combinations of parameters λ and γ on subject 1, which, respectively, control the row sparsity of the scaled projection matrix and the importance of graph regularizer. We observe that SWSC prefers a moderate value λ and a relatively large value γ; specifically, the best recognition accuracies corresponding to the two cross-session tasks, 79.21% and 87.35%, are, respectively, obtained when (λ, γ)s are at $(2^{-3}, 2^{6})$ and $(2^{5}, 2^{7})$. Similar trends can be found on the remaining subjects.

Confusion matrices shown in Fig. 3 provide us with new insights into the recognition results. That is, we can know: 1) what is the average recognition rate of each emotional state; 2) how many EEG samples from one emotional state are misclassified as the other states; and 3) how much performance improvement SWSC obtained on each of the four emotional states in comparison with the other models. For example, the average recognition rate of RLSR to sad is 73.07%, while it is 79.01% by SWSC. By checking the confusion matrix of SWSC, it is observed that 80.58% of the happy samples were correctly recognized, while 6.75%, 3.31%, and 9.36% of them were, respectively, misclassified as sad, fear, and neutral. From our results, we find that SWSC obtained the best recognition rate to the neutral state but the worst rate to the fear state.

To investigate the statistical difference between SWSC and the other models, we conducted the paired Students' t-test on their recognition results. Here, the hypothesis is "the emotion recognition accuracies obtained by SWSC is better than those obtained by the other model." Each test was run on two accuracy sequences corresponding to the 15 emotion recognition cases in each of the three cross-session settings by our SWSC model and the given model. The results of the statistical tests are reported in Table II. We find that all the elements in this table are ✓s, meaning that the hypothesis is correct (true) with probability 0.95. For example, in the "session 1 $\rightarrow$ session 2" task, the decision "77.40 (SWSC) > 69.76 (RLSR)" (see Table I) is correct with probability 0.95. In summary, the conclusion that "SWSC achieves higher recognition accuracy than the other compared models" is well supported by the results in Table II.

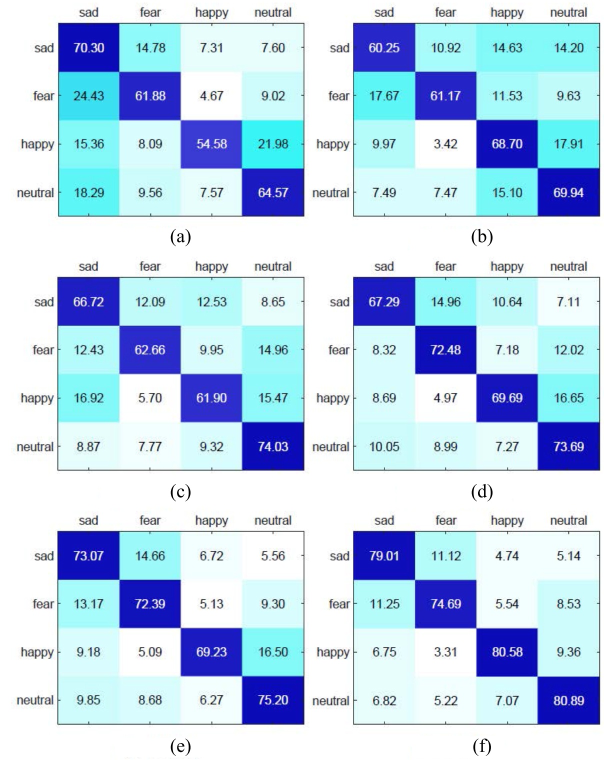

Fig. 3. Recognition results (%) represented by confusion matrices. (a) SVM. (b) GFHF. (c) semiSVM. (d) semiLSR. (e) RLSR. (f) SWSC.

TABLE II

STATISTICAL TEST BETWEEN SWSC AND EACH OF THE OTHER MODELS

t-test	sess.1→sess.2	sess.1→sess.3	sess.2→sess.3
m6>m1	✓	✓	✓
m6>m2	✓	✓	✓
m6>m3	✓	✓	✓
m6>m4	✓	✓	✓
m6>m5	✓	✓	✓

Note: sess.1→sess.2, sess.1→sess.3 and sess.2→sess. respectively denote the three cross-session emotion recognition paradigms.

Besides the t-test, the Nemenyi test is employed to further rank the performance of these six models in all these 45 cross-session emotion recognition cases [35]. Based on the results in Table I, we calculated the average ranks of these six models $[r_{\mathrm{SVM}}, r_{\mathrm{GFHF}}, r_{\mathrm{semiSVM}}, r_{\mathrm{semiLSR}}, r_{\mathrm{RLSR}}, r_{\mathrm{SWSC}}]$ as [4.98, 4.09, 4.18, 3.33, 3.06, 1.32]. In the case of tied ranks, the involved models are enforced to share the average rank. For example, in the "session 1 $\rightarrow$ session 2" task on subject 15, the six models, respectively, obtained accuracies of [88.82%, 91.35%, 88.94%, 98.56%, 98.56%, 98.56%] and accordingly their ranks are [6, 4, 5, 2, 2, 2]. The critical distance (CD) value in

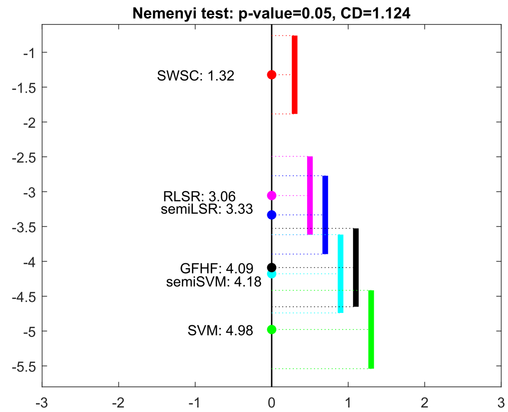

Fig. 4. Nemenyi test for all the used models on EEG-based emotion recognition. The CD is 1.124.

Nemenyi test can be calculated by

$$\mathrm{CD} = q_\alpha \sqrt{\frac{k(k+1)}{6N}} \quad (19)$$

where $k = 6$ is the number of models in comparison, and $N = 45$ is the number of all recognition cases. If the significance level α is set as 0.05 by default, we have $q_\alpha = 2.850$ and then CD = 1.124 [35]. Based on the obtained ranks of all the compared models and the CD value, we draw the test results in Fig. 4. If there exist overlaps between two vertical lines, we declare that the corresponding two models have no significant difference in performance. Otherwise, there exists significant difference. For example, since $r_{\mathrm{SWSC}} = 1.32$ and $r_{\mathrm{RLSR}} = 3.06$, the distance between them is 1.74 which is greater than the CD value 1.124. Therefore, we say that SWSC is significantly better than RLSR based on our obtained results. Similar conclusions can be obtained between SWSC and the remaining other models.

C. Feature Selection

Since the self-weighted variable $\boldsymbol{\theta}$ ranks the importance of different EEG features, it is undoubtedly competent for feature selection. Therefore, we evaluated its feature selection capability by comparing it with two supervised (mRMR [36] and $\ell_{2,1}$-norms [37]) and two semi-supervised (RLSR [29] and PRPC [38]) feature selection models. Linear SVM (with regularization parameter $C = 1$) was used as classifier on the newly formed dataset by selected EEG features. The related parameters in respective models were set as suggested by the original articles. We, respectively, selected 10, 20, 50, 100, and 200 features by different models and the best results in terms of these numbers of selected features are reported in Tables III–V, where the best results are highlighted in bold face. Each number in the brackets corresponds to the dimensionality when the respective model achieved the best performance.

These results reveal some interesting points.

1) SWSC generally achieved the best accuracies in most cases (i.e., 32 out of the total 45 cases), indicating the superiority of the self-weighted variable $\boldsymbol{\theta}$ in feature selection.

TABLE III
FEATURE SELECTION RESULTS OF THE SESSION1->SESSION2 TASK

	session1->session2				
	mRMR	L21	RRPC	RLSR	SWSC
s1	31.01(10)	30.29(100)	68.63(200)	67.07(20)	**71.99**(100)
s2	82.21(50)	80.65(100)	69.95(200)	75.96(100)	**85.58**(200)
s3	72.84(10)	57.81(20)	70.91(200)	71.39(100)	**77.28**(100)
s4	56.73(10)	76.08(10)	56.73(50)	67.55(10)	**81.49**(10)
s5	56.13(50)	54.33(200)	56.61(200)	58.53(200)	**65.38**(50)
s6	39.78(20)	39.90(20)	45.07(200)	46.51(200)	**46.75**(100)
s7	54.09(20)	73.56(200)	62.98(200)	77.88(10)	**79.09**(100)
s8	48.80(200)	68.03(100)	67.19(200)	75.84(10)	**79.33**(50)
s9	76.20(20)	**79.09**(20)	75.72(100)	77.40(100)	78.37(50)
s10	33.05(20)	33.53(200)	43.27(200)	41.11(20)	**45.55**(10)
s11	44.23(20)	45.91(200)	49.40(10)	**52.52**(20)	49.76(200)
s12	37.14(20)	37.14(10)	41.59(10)	50.96(50)	**52.76**(10)
s13	47.48(50)	59.49(20)	62.38(200)	61.90(200)	**69.11**(50)
s14	65.14(200)	72.24(200)	79.69(50)	**85.22**(200)	82.33(100)
s15	88.82(200)	91.23(100)	93.27(200)	95.55(200)	**97.36**(100)

TABLE IV
FEATURE SELECTION RESULTS OF THE SESSION1->SESSION3 TASK

	session1->session3				
	mRMR	L21	RRPC	RLSR	SWSC
s1	59.98(10)	59.12(200)	70.07(200)	85.28(200)	**89.54**(200)
s2	72.75(10)	60.83(100)	79.68(200)	86.74(100)	**87.10**(100)
s3	**56.33**(50)	42.21(20)	42.09(10)	53.28(100)	54.74(50)
s4	68.85(50)	**85.28**(50)	57.42(10)	79.20(10)	74.45(10)
s5	53.16(100)	55.59(50)	59.12(200)	64.36(200)	**69.59**(10)
s6	71.53(50)	72.02(20)	64.36(200)	**79.81**(100)	75.18(50)
s7	52.07(200)	69.59(10)	65.33(200)	72.51(200)	**77.98**(10)
s8	77.49(200)	77.62(200)	65.33(100)	81.63(10)	**84.43**(20)
s9	52.43(100)	67.76(20)	61.19(200)	61.19(200)	**69.95**(100)
s10	45.01(20)	35.04(10)	45.62(10)	49.76(200)	**57.54**(100)
s11	63.75(100)	62.89(200)	**74.09**(100)	65.69(50)	65.69(20)
s12	33.21(200)	39.54(20)	**59.00**(50)	53.76(20)	57.06(50)
s13	48.30(10)	47.32(100)	**65.81**(100)	58.52(200)	54.50(200)
s14	60.22(100)	70.80(20)	66.55(200)	71.53(20)	**83.58**(20)
s15	89.05(200)	87.59(100)	84.79(200)	90.39(50)	**93.07**(50)

TABLE V
FEATURE SELECTION RESULTS OF THE SESSION2->SESSION3 TASK

	session2->session3				
	mRMR	L21	RRPC	RLSR	SWSC
s1	54.87(200)	42.09(10)	62.04(200)	68.13(200)	**72.14**(100)
s2	75.06(50)	75.67(20)	70.07(50)	79.92(20)	**84.43**(20)
s3	51.70(200)	52.07(10)	53.41(50)	58.52(200)	**65.94**(100)
s4	68.00(200)	72.14(200)	56.20(200)	**74.09**(50)	**74.09**(50)
s5	46.84(100)	57.91(10)	49.76(200)	60.95(100)	**66.42**(100)
s6	77.37(50)	83.21(10)	74.21(200)	64.96(100)	**87.10**(20)
s7	**86.62**(50)	83.70(50)	81.39(200)	85.16(100)	83.46(200)
s8	58.64(200)	69.34(50)	71.17(200)	74.82(50)	**82.73**(100)
s9	63.63(10)	61.56(200)	64.60(100)	**70.44**(100)	66.42(100)
s10	67.03(200)	**80.05**(20)	72.26(200)	64.11(50)	77.49(20)
s11	63.26(200)	61.68(50)	57.06(20)	63.50(10)	**67.40**(20)
s12	60.58(50)	61.19(200)	65.69(200)	67.03(200)	**72.26**(50)
s13	46.59(200)	59.00(200)	**63.99**(20)	61.19(100)	63.38(100)
s14	67.40(200)	89.90(100)	88.32(200)	92.34(100)	**90.27**(100)
s15	87.10(20)	88.08(200)	85.04(200)	90.99(10)	**95.01**(50)

2) The best results are not always achieved when the number of selected features is 200, which explicitly demonstrates the necessity of exploring the

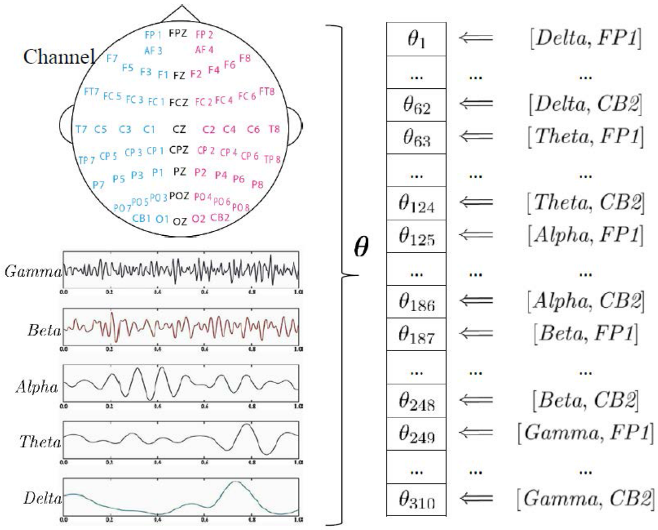

Fig. 5. Correspondence between $\boldsymbol{\theta}$ and the EEG frequency bands and channels in SEED_IV.

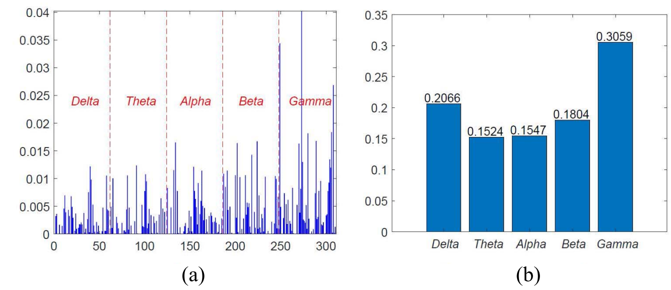

Fig. 6. Importance of different EEG frequency bands by SWSC. (a) Subject 1. (b) Average result.

discriminative abilities of different EEG features in emotion recognition.

For example, the best accuracy of SWSC in the "session 1 → session 2" task on subject 1, 97.36%, is obtained when the number of selected features is 100. From our point of view, the superiority of SWSC is originated from three aspects which include involving unlabeled samples into the learning process for better modeling data properties, jointly estimating the emotional states of unlabeled EEG samples with the other model variables and incorporating the graph regularizer to constrain the regularity of model variables.

D. Affective Activation Patterns Mining

After SWSC was fit by EEG data, we obtain the learned self-weighted variable $\boldsymbol{\theta}$. Before performing the affective activation patterns mining, we establish the correspondence between $\boldsymbol{\theta}$ and the EEG frequency bands (channels). In Fig. 5, we show the correspondence of SEED_IV which has five frequency bands and 62 channels. It is important to point out that such correspondence can be generalized to any number of EEG frequency bands and channels on condition that spectra features are used. Supposing that we have p frequency bands and q channels, the importance of the ith ($1 \leq i \leq p$) frequency band can be calculated as

$$\omega(i) = \theta_{(i-1)*q+1} + \theta_{(i-1)*q+2} + \cdots + \theta_{i*q}. \tag{20}$$

Similarly, the importance of the jth ($1 \leq j \leq q$) channel can be measured by

$$\psi(j) = \theta_j + \theta_{j+q} + \cdots + \theta_{j+(p-1)*q}. \tag{21}$$

From the data-driven perspective, if a certain EEG feature has a greater discriminative ability in cross-session emotion recognition, it should be assigned a larger weight by SWSC. In Fig. 6(a), we show the learned self-weighted variable $\boldsymbol{\theta}$ on subject 1, which is the average of the three cross-session tasks. It is the gamma frequency band which contributes the most in recognizing the emotional states of subject 1. The importance of different frequency bands can be calculated according to rule (20) by setting $p = 5$ and $q = 62$. To remove the individual differences, we provide the average importance measure of each frequency band over all the 15 subjects in Fig. 6(b), which also indicates that the gamma frequency band is the most important one in EEG-based cross-session emotion recognition. This finding coincides with some existing studies [25], [26], [39]. However, in these studies, the conclusion was achieved by a trial-and-error manner, which basically tested each of all the frequency bands and then considered the one as the most important frequency band if it achieved the best recognition accuracy. By comparison, SWSC is more flexible and adaptive, which completely learned the feature importance from data.

Besides the frequency bands, important EEG channels corresponding to critical brain regions can also be identified by rule (21). In Fig. 7(a), we provide the topographical show of the average importance of all the 62 channels over these 45 cases obtained by SWSC. Generally, we conclude that there are four different brain regions, the prefrontal, left/right temporal and (central) parietal lobes, which might be more closely correlated with the video-evoked emotion expression. This finding is also consistent with some existing studies [25], [26], [33], [40]. However, they empirically selected a few EEG channels (e.g., FT7, T7, TP7, FT8, T8, and TP8 were selected in [33]). Though these selected EEG channels achieved comparable performance as with all channels, there is no strong reason to explain its rationality. In Fig. 7(b), we show the top 10 channels selected by SWSC, most of which are located in the prefrontal and left/right temporal lobes. Therefore, we declare that SWSC offers an underlying explanation of the aforementioned studies in channel selection from the data-driven perspective.

Having described both model formulations and experiments of SWSC, now we discuss on the connections as well as differences between the current work and the GFIL [27]. The main connection between them is the utilization of self-weighted variable in measuring feature importance and frequency band (channel) analysis. There are at least three differences between them. First, the learning paradigms are different. GFIL is a supervised model, while SWSC is a semi-supervised one which is more appropriate for cross-session emotion recognition from EEG. Second, the detailed model

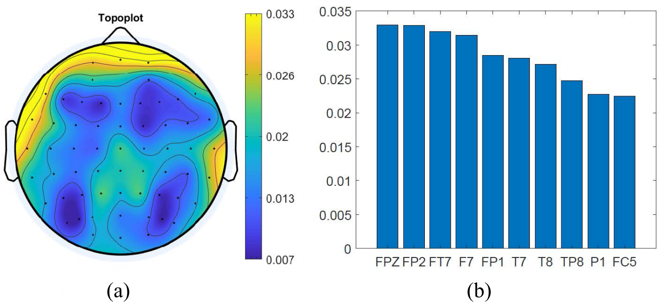

Fig. 7. Importance of different EEG channels based on SWSC. (a) Average result. (b) Top ten channels.

objectives as well as optimization methods are different. Third, SWSC is superior to GFIL in recognition performance. Specifically, SWSC improves the accuracies by 2.07%, 4.53%, and 2.35% in the three cross-session recognition tasks. Besides, the channels in (central) parietal lobes are considered to be more important by the SWSC results.

V. Conclusion

In this article, we proposed an SWSC model to jointly complete EEG-based cross-session emotion recognition and affective activation patterns mining. SWSC was formulated by incorporating a self-weighted variable into a semi-supervised classification model in order to quantitatively and adaptively measure the abilities of different EEG features in cross-session emotion expression. Experimental results demonstrated that this auto-weighting scheme can effectively enhance the emotion recognition performance. Besides, based on the correspondence between EEG features and frequency bands (channels), the affective activation patterns were automatically exploited by the learned self-weighted variable. Based on our results, the gamma band is the most important frequency band; the prefrontal, left/right temporal, and (central) parietal lobes are correlated more to emotion expression.

References

[1] S. K. Khare, V. Bajaj, and G. R. Sinha, "Adaptive tunable Q wavelet transform-based emotion identification," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9609–9617, Dec. 2020.

[2] Z. He *et al.*, "Advances in multimodal emotion recognition based on brain–computer interfaces," *Brain Sci.*, vol. 10, no. 10, p. 687, Sep. 2020.

[3] X. Chen *et al.*, "ReMAE: User-friendly toolbox for removing muscle artifacts from EEG," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 5, pp. 2105–2119, May 2020.

[4] L.-C. Shi, Y.-Y. Jiao, and B.-L. Lu, "Differential entropy feature for EEG-based vigilance estimation," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 6627–6630.

[5] J. Li, N. Thakor, and A. Bezerianos, "Brain functional connectivity in unconstrained walking with and without an exoskeleton," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 3, pp. 730–739, Mar. 2020.

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

[7] M. Yu *et al.*, "A review of EEG features for emotion recognition," *Scientia Sinica Informationis*, vol. 49, no. 9, pp. 1097–1118, Sep. 2019.

[8] W. Zheng, "Multichannel EEG-based emotion recognition via group sparse canonical correlation analysis," *IEEE Trans. Cogn. Devel. Syst.*, vol. 9, no. 3, pp. 281–290, Sep. 2017.

[9] Y. Peng, Q. Li, W. Kong, F. Qin, J. Zhang, and A. Cichocki, "A joint optimization framework to semi-supervised RVFL and ELM networks for efficient data classification," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 106756.

[10] P. Li *et al.*, "EEG based emotion recognition by combining functional connectivity network and local activations," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 10, pp. 2869–2881, Oct. 2019.

[11] Z. Pei, H. Wang, A. Bezerianos, and J. Li, "EEG-based multiclass workload identification using feature fusion and selection," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–8, 2021.

[12] Y. Shu and S. Wang, "Emotion recognition through integrating EEG and peripheral signals," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 2871–2875.

[13] X. Xing, Z. Li, T. Xu, L. Shu, B. Hu, and X. Xu, "SAE+LSTM: A new framework for emotion recognition from multi-channel EEG," *Frontiers Neurorobotics*, vol. 13, p. 37, Jun. 2019.

[14] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Trans. Affect. Comput.*, vol. 11, no. 3, pp. 532–541, Jul./Sep. 2020.

[15] Y. Li, W. Zheng, L. Wang, Y. Zong, and Z. Cui, "From regional to global brain: A novel hierarchical spatial-temporal neural network model for EEG emotion recognition," *IEEE Trans. Affect. Comput.*, early access, Jun. 14, 2019, doi: 10.1109/TAFFC.2019.2922912.

[16] S. Gong, K. Xing, A. Cichocki, and J. Li, "Deep learning in EEG: Advance of the last ten-year critical period," *IEEE Trans. Cognit. Develop. Syst.*, early access, May 13, 2021, doi: 10.1109/TCDS.2021.3079712.

[17] N. S. Suhaimi, J. Mountstephens, and J. Teo, "EEG-based emotion recognition: A state-of-the-art review of current trends and opportunities," *Comput. Intell. Neurosci.*, vol. 2020, pp. 1–19, Sep. 2020.

[18] S. Güneş, K. Polat, and Ş. Yosunkaya, "Efficient sleep stage recognition system based on EEG signal using *k*-means clustering based feature weighting," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 7922–7928, 2010.

[19] L. Guo, Y. Wu, L. Zhao, T. Cao, W. Yan, and X. Shen, "Classification of mental task from EEG signals using immune feature weighted support vector machines," *IEEE Trans. Magn.*, vol. 47, no. 5, pp. 866–869, May 2011.

[20] V. Mishuhina and X. Jiang, "Feature weighting and regularization of common spatial patterns in EEG-based motor imagery BCI," *IEEE Signal Process. Lett.*, vol. 25, no. 6, pp. 783–787, Jun. 2018.

[21] P. Gaur, H. Gupta, A. Chowdhury, K. McCreadie, R. B. Pachori, and H. Wang, "A sliding window common spatial pattern for enhancing motor imagery classification in EEG-BCI," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.

[22] Y. Cui, Y. Xu, and D. Wu, "EEG-based driver drowsiness estimation using feature weighted episodic training," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 11, pp. 2263–2273, Nov. 2019.

[23] J. Li, S. Qiu, C. Du, Y. Wang, and H. He, "Domain adaptation for EEG emotion recognition based on latent representation similarity," *IEEE Trans. Cognit. Develop. Syst.*, vol. 12, no. 2, pp. 344–353, Jun. 2020.

[24] D. Wu, Y. Xu, and B.-L. Lu, "Transfer learning for EEG-based brain-computer interfaces: A review of progress made since 2016," *IEEE Trans. Cognit. Develop. Syst.*, early access, Jul. 7, 2020, doi: 10.1109/TCDS.2020.3007453.

[25] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, "Identifying stable patterns over time for emotion recognition from EEG," *IEEE Trans. Affect. Comput.*, vol. 10, no. 3, pp. 417–429, Jul./Sep. 2017.

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

[27] Y. Peng, F. Qin, W. Kong, Y. Ge, F. Nie, and A. Cichocki, "GFIL: A unified framework for the importance analysis of features, frequency bands and channels in EEG-based emotion recognition," *IEEE Trans. Cognit. Develop. Syst.*, early access, May 21, 2021, doi: 10.1109/TCDS.2021.3082803.

[28] D. Wu, C.-T. Lin, and J. Huang, "Active learning for regression using greedy sampling," *Inf. Sci.*, vol. 474, pp. 90–105, Feb. 2019.

[29] X. Chen, G. Yuan, F. Nie, and J. Z. Huang, "Semi-supervised feature selection via Rescaled linear regression," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 1525–1531.

[30] X. Chen, G. Yuan, F. Nie, and Z. Ming, "Semi-supervised feature selection via sparse rescaled linear square regression," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 1, pp. 165–176, Jan. 2020.

[31] Z. Li, Y. Yang, J. Liu, X. Zhou, and H. Lu, "Unsupervised feature selection using nonnegative spectral analysis," in *Proc. AAAI Conf. Artif. Intell.*, 2012, pp. 1026–1032.

[32] Y. Peng, X. Zhu, F. Nie, W. Kong, and Y. Ge, "Fuzzy graph clustering," *Inf. Sci.*, vol. 571, pp. 38–49, Sep. 2021.

[33] W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, and A. Cichocki, "EmotionMeter: A multimodal framework for recognizing human emotions," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 1110–1122, Mar. 2019.

[34] X. Zhu, Z. Ghahramani, and J. D. Lafferty, "Semi-supervised learning using Gaussian fields and harmonic functions," in *Proc. Int. Conf. Mach. Learn.*, Jan. 2003, pp. 912–919.

[35] Z.-H. Zhou, *Machine Learning*. Beijing, China: Tsinghua Univ. Press, 2016.

[36] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[37] F. Nie, H. Huang, X. Cai, and C. Ding, "Efficient and robust feature selection via joint $\ell_{2,1}$-norms minimization," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 23, 2010, pp. 1813–1821.

[38] J. Xu, B. Tang, H. He, and H. Man, "Semisupervised feature selection based on relevance and redundancy criteria," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 9, pp. 1974–1984, Sep. 2017.

[39] Y. Peng and B.-L. Lu, "Discriminative manifold extreme learning machine and applications to image and EEG signal classification," *Neurocomputing*, vol. 174, pp. 265–277, Jan. 2016.

[40] X.-L. Quan, Z.-G. Zeng, J.-H. Jiang, Y.-Q. Zhang, B.-L. Lu, and D.-R. Wu, "Physiological signals based affective computing: A systematic review," *Acta Autom. Sinica*, vol. 47, no. 8, pp. 1769–1784, 2021.

Yong Peng (Member, IEEE) received the Ph.D. degree from the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2015.

He is currently an Associate Professor with the School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, China, and an Adjunct Post-Doctoral Research Associate with the Center for Optical IMage Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an, China. His main research interests include machine learning, pattern recognition, and EEG-based brain–computer interfaces.

Dr. Peng received the President Prize of Chinese Academy Sciences in 2009 and the Third Prize of the Chinese Institute of Electronics in 2018.

Wanzeng Kong (Member, IEEE) received the Ph.D. degree from the Department of Electrical Engineering, Zhejiang University, Hangzhou, China, in 2008.

He was a Visiting Research Associate with the Department of Biomedical Engineering, University of Minnesota Twin Cities, Minneapolis, MN, USA, from 2012 to 2013. He is currently a Full Professor with the School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou. His current research interests include biomedical signal processing, brain–computer interface, cognitive computing, and pattern recognition.

Feiwei Qin received the Ph.D. degree from the School of Computer Science and Technology, Zhejiang University, Hangzhou, China, in 2014.

He is currently an Associate Professor with the School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou. His research interests include artificial intelligence, computer vision, computer graphics, computer-aided design, and medical image analysis.

Feiping Nie (Senior Member, IEEE) received the Ph.D. degree in computer science from Tsinghua University, Beijing, China, in 2009.

He is currently a Full Professor with Northwestern Polytechnical University, Xi'an, China. His research interests include machine learning and its applications, such as pattern recognition, data mining, computer vision, image processing, and information retrieval. He has published more than 100 papers in top journals and conferences, such as IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, *International Journal of Computer Vision* (IJCV), International Conference on Machine Learning, and *Neural Information Processing Systems*. His articles have been cited more than 10 000 times.

Dr. Nie is an associate editor or a PC member of several prestigious journals and conferences in the related fields.

Jinglong Fang received the Ph.D. degree in control engineering from the Zhejiang University of Technology, Hangzhou, China, in 2012.

He is currently a Professor with the School of Computer Science, Hangzhou Dianzi University, Hangzhou. His research interests include intelligent software engineering, machine learning, and artificial intelligence.

Bao-Liang Lu (Fellow, IEEE) received the Dr.Eng. degree in electrical engineering from Kyoto University, Kyoto, Japan, in 1994.

From 1994 to 1999, he was a Frontier Researcher with the Bio-Mimetic Control Research Center, Institute of Physical and Chemical Research (RIKEN), Nagoya, Japan, and the Research Scientist of the RIKEN Brain Science Institute, Wako, Japan, from 1999 to 2002. Since 2002, he has been a Full Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. His current research interests include brain-like computing, neural networks, machine learning, brain–computer interaction, and affective computing.

Dr. Lu was a Steering Committee Member of IEEE TRANSACTIONS ON AFFECTIVE COMPUTING. He received the IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT Outstanding Paper Award in 2018. He is an Associate Editor of IEEE TRANSACTIONS ON COGNITIVE AND DEVELOPMENTAL SYSTEMS and *Journal of Neural Engineering*.

Andrzej Cichocki (Life Fellow, IEEE) received the M.Sc. (Hons.), Ph.D., and Dr.Sc. (Habilitation) degrees from the Warsaw University of Technology, Warsaw, Poland, in 1972, 1976, and 1982, respectively, all in electrical engineering.

He spent several years at the University of Erlangen, Erlangen, Germany, as an Alexandervon-Humboldt Research Fellow and a Guest Professor. He was the Senior Team Leader and the Head of the Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Wako, Japan. He is currently a Professor with the Skolkovo Institute of Science and Technology, Moscow, Russia, and an Adjunct Professor with Hangzhou Dianzi University, Hangzhou, China. He is the author of more than 500 technical journal articles and six monographs in English (two of them translated to Chinese). His current research interests include multiway blind source separation, tensor decompositions, tensor networks, deep learning, human–robot interactions, and brain–computer interface.

Dr. Cichocki has served as an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON CYBERNETICS, and *Journal of Neuroscience Methods*. He is the Founding Editor-in-Chief of *Computational Intelligence and Neuroscience* journal.