Neural Process Lett (2019) 49:555–571
https://doi.org/10.1007/s11063-018-9829-1

EEG Emotion Recognition Based on Graph Regularized Sparse Linear Regression

Yang Li[1,2] · Wenming Zheng[1] · Zhen Cui[3] · Yuan Zong[1] · Sheng Ge[1]

Published online: 25 April 2018

Abstract In this paper, a novel regression model, called graph regularized sparse linear regression (GRSLR), is proposed to deal with EEG emotion recognition problem. GRSLR extends the conventional linear regression method by imposing a graph regularization and a sparse regularization on the transform matrix of linear regression, such that it is able to simultaneously cope with sparse transform matrix learning while preserve the intrinsic manifold of the data samples. To detailed discuss the EEG emotion recognition, we collect a set of 14 subjects EEG emotion data and provide the experiment results on different features. To evaluate the proposed GRSLR model, we conduct experiments on the SEED database and RCLS database. The experimental results show that the proposed algorithm GRSLR is superior to the classic baselines. The RCLS database is made publicly available and other researchers could use it to test their own emotion recognition method.

Keywords EEG · Emotion recognition · Sparse linear regression

1 Introduction

Emotion is a common and basic physiologic phenomenon existing in everyone. Human has the ability to capture the emotion expressed by others but machine can not understand it because of its complexity. Thus emotion recognition, as one important issue in affective computing field, draws a lot of attentions [1].

✉ Wenming Zheng
wenming_zheng@seu.edu.cn

1 Key Laboratory of Child Development and Learning Science of Ministry of Education, Research Center for Learning Science, Southeast University, Nanjing 210096, Jiangsu, China

2 School of Information Science and Engineering, Southeast University, Nanjing 210096, Jiangsu, China

3 School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210096, Jiangsu, China

The responses of emotion can be facial expression, speech and other physiologic signals such as skin conductance response, heart rate, blood pressure, cortisol level, electromyography and respiration rate, etc. In addition to the above peripheral responses, emotion recognition based on brain signals, e.g., electroencephalograph (EEG), had attracted increasing interests among researchers in recent years. From the neuroscience point of view [2], human's emotion is closely related to a variety of brain subregions, such as the orbital frontal cortex, ventral medial prefrontal cortex, amygdala, etc. [3–5]. The technology of EEG can measure the changes of brain electrical activities noninvasively by placing electrodes on the head of participants. Because EEG has high temporal resolution and directly reflects the potential activity of the nerve, it can be used to decode emotions. EEG, as a novel research means for emotion, makes another window for emotion recognition.

As a typical pattern recognition task, EEG emotion recognition problem can be roughly divided into two main steps, *i.e.*, feature extraction and classifier design. Herein, feature extraction is thought more important because designing suitable features for EEG signal according to the mechanism of EEG would contributes to distinguishing different EEG emotions. For this reason, in recent years many researchers focus on investigating EEG features for analyzing the emotions EEG signals convey. Jenke et al. [6] reviewed the existing features of EEG emotion data and divided them into time domain features, frequency domain features, time-frequency domain features and some features calculated from combinations of electrodes. As neurons responsible for emotional process are activated in a special region according to human cognition mechanism, the choice of channels associated with electrodes is often considered in the task of emotion recognition [7,8]. In [8], the effect of different electrode for emotion analysis is elaborately explored on the style of music data. Zheng [9] proposed a group sparse canonical correlation analysis method to select some useful channels with respect to emotion. The latest researches contain that Zhang et al. [10] used empirical mode decomposition (EMD) method to decompose EEG data into a series of intrinsic mode functions (IMFs) and use the first 4 IMFs to calculate sample entropies, and Zheng et al. [11] employed Deep Belief Network (DBN) to try to extract high-level features of EEG emotion signals.

In this paper, we propose a novel EEG emotion recognition method, called graph regularized sparse linear regression (GRSLR), to jointly perform channel selection and feature extraction. GRSLR extends the conventional linear regression method by imposing a graph regularization and a sparse regularization on the transform matrix of linear regression. The sparse regularization is used to cope with channel choice by constraining the space of features of each entire channel with group sparsity, whereas the graph regularization is used to preserve intrinsic manifold structures in the process of data embedding to reduce over-fitting of trained models. To suppress those noises and further extract those representative frequency signals, we divide the raw EEG signals into five frequency bands, i.e., δ, θ, α, β and γ, to feed into our proposed GRSLR model. To evaluate our method, we conduct experiments on the public and prevalent SEED database. Furthermore, in this paper, similar with the protocol of SEED, we collect 14 subjects' EEG emotion data and provide the detailed experiment results with different features. The experimental results show that the proposed GRSLR model is more effective in the EEG based emotion recognition task.

2 GRSLR Model

2.1 Linear Regression

Suppose that we have a training feature matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N] \in \mathbb{R}^{d\times N}$ and its corresponding label matrix is denoted by $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N] \in \mathbb{R}^{c\times N}$ whose ith column $\mathbf{y}_i = [y_{i,1}, \ldots, y_{i,c}]^T$ has a binary entry 1 or 0, where d is the dimension of feature vectors, N is the number of the training samples, c is the class number, and 1 or 0 means whether the ith sample belongs to the jth class. Then, the relationship between training feature matrix and label matrix can be built by using a simple linear regression model which has the following formulation:

$$\min_{\mathbf{B}} \|\mathbf{Y} - \mathbf{B}^T\mathbf{X}\|_F^2, \tag{1}$$

where $\mathbf{B} \in \mathbb{R}^{d\times c}$ are the regression coefficient matrix. $\|.\|_F$ is the Frobenius norm.

2.2 GRSLR

Given one subject, let the matrix $\mathbf{X}_i \in \mathbb{R}^{\tilde{d}\times N}$ denote the feature matrix consisting of the training samples at the i-th ($i = 1, \ldots, K$) channel, where K denotes the EEG channel number and $\tilde{d}$ is the number of EEG features associated with the i-th channel. The K feature matrices associated with all the EEG channels are concatenated to form a larger size feature matrix such that the EEG feature matrix of training data can be written as $\mathbf{X} = [\mathbf{X}_1^T, \ldots, \mathbf{X}_i^T, \ldots, \mathbf{X}_K^T]^T \in \mathbb{R}^{d\times N}$, $d = K\tilde{d}$. Thus Eq. (1) can be reformulated as follows,

$$\min_{\mathbf{B}_i} f_d(\mathbf{B}) = \min_{\mathbf{B}_i} \|\mathbf{Y} - \sum_{i=1}^{K} \mathbf{B}_i^T\mathbf{X}_i\|_F^2, \tag{2}$$

where $\mathbf{B}_i \in \mathbb{R}^{\tilde{d}\times c}$ is a block matrix of coefficients corresponding to the features $\mathbf{X}_i$ of the i-th channel.

In dealing with EEG emotion recognition, it is notable that different EEG channels play different role. Some of the EEG channels contribute more to the emotion recognition results whereas some of them contribute less. Consequently, channel selection is essential to improve the EEG emotion recognition performance. Thus, different from the above Linear Regression in Eq. (2), GRSLR will take into consideration the contribution of different channels to EEG emotion recognition. We add weights s_i on the projecting matrix $\mathbf{B}_i$ as a sparse selection constraint. In this case, the optimization of Eq. (2) can be re-formulated as

$$\min_{\mathbf{B}_i, s_i} f(\mathbf{B}) = \min_{\mathbf{B}_i, s_i} \|\mathbf{Y} - \sum_{i=1}^{K} s_i\mathbf{B}_i^T\mathbf{X}_i\|_F^2 + \mu\sum_{i=1}^{K} s_i, \quad \text{s.t. } s_i \in \{0, 1\} \tag{3}$$

Let $\hat{\mathbf{B}}_i = s_i\mathbf{B}_i$. Then the zero value of s_i means the zero value of $\|\hat{\mathbf{B}}_i\|_F$. From the other side, either $\mathbf{B}_i$ is a zero matrix or s_i is zero when $\|\hat{\mathbf{B}}_i\|_F$ equals to zero. If $\mathbf{B}_i$ is a zero matrix, the feature matrix $\mathbf{X}_i$ has no contribution to the response matrix $\mathbf{Y}$ in the regression model of Eq. (3), which also means that the i-th channel is not important and hence a zero value can be assigned to s_i. Consequently, the value of $\|\hat{\mathbf{B}}_i\|_F$ provides a good surrogation of s_i. For this reason, we replace $\sum_{i=1}^{K} s_i$ in Eq. (3) with $\sum_{i=1}^{K} \|\hat{\mathbf{B}}_i\|_F$ [12]. Furthermore, we introduce

$\sum_{i=1}^{K} \|\hat{\mathbf{B}}_i\|_F^2$ to avoid over-fitting and construct the following group sparse term, formally,

$$f_c\left(\hat{\mathbf{B}}\right) = \sum_{i=1}^{K} \left(\|\hat{\mathbf{B}}_i\|_F + \|\hat{\mathbf{B}}_i\|_F^2\right). \tag{4}$$

The regularized regression optimization problem is obtained:

$$\min_{\hat{\mathbf{B}}_i} f\left(\hat{\mathbf{B}}\right) = \min_{\hat{\mathbf{B}}_i} \|\mathbf{Y} - \sum_{i=1}^{K} \hat{\mathbf{B}}_i^T \mathbf{X}_i\|_F^2 + \mu \sum_{i=1}^{K} \left(\|\hat{\mathbf{B}}_i\|_F + \|\hat{\mathbf{B}}_i\|_F^2\right). \tag{5}$$

To consider the local manifold structures of features as well as the contribution of different channels to EEG emotion recognition, a natural assumption is that if two data points are close to each other in the data distribution, then the representations of these two points should be close. Based on manifold assumption [13], given two feature vectors $\mathbf{v}_j$ and $\mathbf{v}_k$ close to each other in the original feature space, their embedding representations on a new transform space should also preserve their original adjacent relationship, which can benefit for improve the classification accuracy to some extent.

Formally, let G denote the k nearest neighbor graph of N training samples, where each sample is regarded as a vertex, $\mathbf{W}$ be its corresponding weight matrix. If $\mathbf{v}_j$ is among the k-nearest neighbors of $\mathbf{v}_k$ or $\mathbf{v}_k$ is among the k-nearest neighbors of $\mathbf{v}_j$, the element $W_{j,k} = 1$, otherwise $W_{j,k} = 0$, which can be formulated as:

$$W_{j,k} = \begin{cases} 1, & \mathbf{v}_j \in \mathcal{N}(\mathbf{v}_k) \text{ or } \mathbf{v}_k \in \mathcal{N}(\mathbf{v}_j); \\ 0, & \text{otherwise.} \end{cases}$$

To keep the intrinsic data structure, we should embed weighted graph G into the representation space by minimizing the following objective function:

$$f_m\left(\hat{\mathbf{B}}\right) = \frac{1}{2} \sum_{i=1}^{K} \sum_{j=1}^{N} \sum_{k=1}^{N} \left(\hat{\mathbf{b}}_{i,j}^T \mathbf{x}_{i,j} - \hat{\mathbf{b}}_{i,k}^T \mathbf{x}_{i,k}\right)^2 W_{j,k} \tag{6}$$

$$= \sum_{i=1}^{K} \mathrm{tr}\left(\hat{\mathbf{B}}_i^T \mathbf{X}_i \mathbf{L} \mathbf{X}_i^T \hat{\mathbf{B}}_i\right) = \mathrm{tr}(\hat{\mathbf{B}}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \hat{\mathbf{B}}), \tag{7}$$

where $\hat{\mathbf{B}} = [\hat{\mathbf{B}}_1^T, \hat{\mathbf{B}}_2^T, \ldots, \hat{\mathbf{B}}_K^T]^T$, $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the Laplacian matrix, $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_i, \ldots, d_N)$ and $d_i = \sum_1^N W_{j,k}$, and $\mathrm{tr}(\cdot)$ means the trace of a matrix.

By combining the above three equations f_d, f_m, f_c to form the final objective function, i.e., the proposed GRSLR model is,

$$\min_{\mathbf{B}} \|\mathbf{Y} - \sum_{i=1}^{K} \mathbf{B}_i^T \mathbf{X}_i\|_F^2 + \lambda \sum_{i=1}^{K} \mathrm{tr}\left(\mathbf{B}_i^T \mathbf{X}_i \mathbf{L} \mathbf{X}_i^T \mathbf{B}_i\right) + \mu \sum_{i=1}^{K} (\|\mathbf{B}_i\|_F + \|\mathbf{B}_i\|_F^2), \tag{8}$$

where λ and μ are the balance parameters corresponding to two regularization terms. Here $\hat{\mathbf{B}}_i$ is simplified as $\mathbf{B}_i$.

2.3 Optimization

To solve the objective model in Eq. (8), we use the inexact augmented Lagrange multiplier (ALM) strategy [12,14]. Concretely, we introduce an auxiliary variable $\mathbf{C}$ which is expected

to be equal to $\mathbf{B}$. Thus, the problem of Eq. (8) can be converted to the optimization problem with a equation constraint as the following:

$$\begin{aligned}\min_{\mathbf{B},\mathbf{C}} \|\mathbf{Y}-\sum_{i=1}^{K}\mathbf{C}_i^T\mathbf{X}_i\|_F^2+\lambda\sum_{i=1}^{K}\mathrm{tr}\left(\mathbf{B}_i^T\mathbf{X}_i\mathbf{L}\mathbf{X}_i^T\mathbf{B}_i\right)\\ +\mu\left(\sum_{i=1}^{K}\|\mathbf{B}_i\|_F+\sum_{i=1}^{K}\|\mathbf{C}_i\|_F^2\right),\\ \text{s.t. } \mathbf{C}_i=\mathbf{B}_i, i=1,\ldots,K.\end{aligned} \tag{9}$$

Then the augmented Lagrange function can be expressed as:

$$\begin{aligned}L(\mathbf{B}_i,\mathbf{C}_i,\mathbf{T}_i)=&\left\|\mathbf{Y}-\sum_{i=1}^{K}\mathbf{C}_i^T\mathbf{X}_i\right\|_F^2+\sum_{i=1}^{K}\mathrm{tr}[\mathbf{T}_i^T(\mathbf{C}_i-\mathbf{B}_i)]\\ &+\frac{\kappa}{2}\sum_{i=1}^{K}\|\mathbf{C}_i-\mathbf{B}_i\|_F^2+\lambda\sum_{i=1}^{K}\mathrm{tr}\left(\mathbf{C}_i^T\mathbf{X}_i\mathbf{L}\mathbf{X}_i^T\mathbf{C}_i\right)\\ &+\mu\left(\sum_{i=1}^{K}\|\mathbf{B}_i\|_F+\sum_{i=1}^{K}\|\mathbf{C}_i\|_F^2\right),\end{aligned} \tag{10}$$

where κ is a regularization parameter and $\mathbf{T}_i$ is the Lagrange multiplier. To solve Eq. (10), we employ iterate optimization on the three variables, and the optimization steps are summarized as follows:

1. Fix $\mathbf{B}_i$ and $\mathbf{T}_i$, and Update $\mathbf{C}_i$: $\mathbf{C}_i=\arg\min_{\mathbf{C}_i}\|\mathbf{Y}-\sum_{i=1}^{K}\mathbf{C}_i^T\mathbf{X}_i\|_F^2+\sum_{i=1}^{K}\mathrm{tr}[\mathbf{T}_i^T(\mathbf{C}_i-\mathbf{B}_i)]+\lambda\sum_{i=1}^{K}\mathrm{tr}(\mathbf{C}_i^T\mathbf{X}_i\mathbf{L}\mathbf{X}_i^T\mathbf{C}_i)+\frac{\kappa}{2}\sum_{i=1}^{K}\|\mathbf{C}_i-\mathbf{B}_i\|_F^2+\lambda\sum_{i=1}^{K}\|\mathbf{C}_i\|_F^2$;
2. Fix $\mathbf{C}_i$ and $\mathbf{T}_i$, and Update $\mathbf{B}_i$: $\mathbf{B}_i=\arg\min_{\mathbf{B}_i}\frac{\mu}{\kappa}\|\mathbf{B}_i\|_F+\frac{1}{2}\|\mathbf{B}_i-(\mathbf{C}_i+\frac{1}{\kappa}\mathbf{T}_i)\|_F^2$;
3. Update $\mathbf{T}_i$ and κ: $\mathbf{T}_i=\mathbf{T}_i+\kappa(\mathbf{C}_i-\mathbf{B}_i)$ and $\kappa=\min(\rho\kappa,\kappa_{max})$;
4. Check Convergence: $\|\mathbf{C}_i-\mathbf{B}_i\|_\infty\le\epsilon$.

In Step 1, by setting gradient of the objective function as 0, we can obtain a close-form solution of $\mathbf{C}_i$:

$$\mathbf{C}=\left[\frac{2\mathbf{X}(\lambda\mathbf{Y}+\mathbf{I})\mathbf{X}^T}{\kappa}+\left(1+\frac{2\mu}{\kappa}\right)\mathbf{I}\right]^{-1}\left(\frac{2\mathbf{X}\mathbf{Y}^T-\mathbf{T}}{\kappa}+\mathbf{B}\right), \tag{11}$$

where $\mathbf{C}=[\mathbf{C}_1^T,\ldots,\mathbf{C}_K^T]^T$ and $\mathbf{I}$ denotes the identity matrix. As the other steps are similar to Algorithm 1 of the previous work [12], we omit their solutions.

2.4 Testing

After obtaining the optimal solution $\hat{\mathbf{B}}_i$ using the above method, for a given testing sample $\mathbf{x}_t$, we can conveniently calculate its corresponding emotion label vector: $\mathbf{l}_t=\sum_{i=1}^{K}\hat{\mathbf{B}}_i^T(\mathbf{x}_t)_i$. Here $(\mathbf{x}_t)_i$ is the i-th channel of the testing sample x_t. Then, its emotion category can be determined by the following criterion:

$$\text{emotion\_label}=\arg\max_k\{\mathbf{l}_t(k)\}, \tag{12}$$

where $\mathbf{l}_t(k)$ indicates the k-th element of label vector $\mathbf{l}_t$.

Table 1 The objective functions of LDA, LR, GraphSC, MGrSc and GRSLR

Method	Objective function
LDA	$\max_{\mathbf{W}} \frac{tr(\mathbf{W}\mathbf{S}_W\mathbf{W})}{tr(\mathbf{W}\mathbf{S}_B\mathbf{W})}$
LR	$\min_{\mathbf{B}} \|\mathbf{Y} - \mathbf{B}^T\mathbf{X}\|_F^2$
GraphSC	$\min_{\mathbf{W},\mathbf{S}} \|\mathbf{X} - \mathbf{W}\mathbf{S}\|_F^2 + \lambda \mathrm{tr}(\mathbf{S}\mathbf{L}\mathbf{S}^T) + \mu \sum_{i=1}^{N} \|\mathbf{s}_i\|_1$ s.t. $\|\mathbf{w}_i\|^2 \le c, i = 1, 2, \ldots, R$
MGrSc	$\min_{\mathbf{W},\mathbf{S}} \|\mathbf{X} - \mathbf{W}\mathbf{S}\|_F^2 + \lambda \mathrm{tr}(\mathbf{S} \sum_{k=1}^{P} \tau_i \mathbf{L}_i \mathbf{S}^T) + \mu \sum_{i=1}^{N} \|\mathbf{s}_i\|_1 + \gamma\|\tau\|^2$ s.t. $\sum_{k=1}^{P} \tau_i = 1, \tau_i \ge 0, \|\mathbf{w}_i\|^2 \le c, i = 1, 2, \ldots, R$
GRSLR	$\min_{\mathbf{B}} \|\mathbf{Y} - \sum_{i=1}^{K} \mathbf{B}_i^T \mathbf{X}_i\|_F^2 + \lambda \sum_{i=1}^{K} \mathrm{tr}(\mathbf{B}_i^T \mathbf{X}_i \mathbf{L} \mathbf{X}_i^T \mathbf{B}_i) + \mu \sum_{i=1}^{K} (\|\mathbf{B}_i\|_F + \|\mathbf{B}_i\|_F^2)$

Notation and description: $\mathbf{S}_W$ and $\mathbf{S}_B$ are the within-class and between-class scatter. $\mathbf{W}$ and $\mathbf{B}$ are linear transformation matrices. $\mathbf{S}$ is sparse representation matrix. λ, μ and τ_i are balance parameters. $\mathbf{X}$ and $\mathbf{Y}$ are the data matrix and label matrix respectively. P is the number of candidate graphs. τ is the graph weight vector. R is the number of atoms in transformation matrix. N is the number of data points

2.5 The Relationship Among LDA, LR, GraphSC, MGrSc and GRSLR.

The objective functions of Linear Discriminant Analysis (LDA), Linear Regression (LR), Graph regularized Sparse Coding (GraphSC) [15], Multiple Graph regularized Sparse coding (MGrSc) [16] and our GRSLR are listed in Table 1.

LDA is originally one of the most widely-used feature extraction method which aims to learn a transform matrix that maximizes the Fisher's discriminant criterion. Recently, Torre [17] indicated that from the view of regression, LDA is able to be reformulated as a weighted reduced rank regression (WRRR) optimization problem. More specifically, suppose that we have a training feature matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N] \in \mathbb{R}^{d\times N}$ and its corresponding label matrix is denoted by $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N] \in \mathbb{R}^{c\times N}$ whose ith column $\mathbf{y}_i = [y_{i,1}, \ldots, y_{i,c}]^T$ has a binary entry 1 or 0, where d is the dimension of feature vectors, N is the number of the training samples, c is the class number, and 1 or 0 means whether the ith sample belongs to the jth class. Then, according to the work of Torre [17], the optimization problem of LDA can be formulated as follows:

$$\min_{\mathbf{A},\mathbf{B}} \|(\mathbf{Y}\mathbf{Y}^T)^{-1/2}(\mathbf{Y} - \mathbf{A}\mathbf{B}^T\mathbf{X})\|_F^2, \tag{13}$$

where $\mathbf{A} \in \mathbb{R}^{c\times r}$ and $\mathbf{B} \in \mathbb{R}^{d\times c}$ are the regression coefficient matrices and r is their rank.

We can use two feasible operations to reduce the complexity of the novel LDA formulation in advance. Firstly, we assume that the numbers of training samples belonging to each class are approximately equal such that $\mathbf{Y}\mathbf{Y}^T$ can be replaced by an identity matrix. Secondly, we further fix coefficient matrix $\mathbf{A}$ as an identity matrix as well. Thus, we will obtain a least squares regression (LSR) based formulation of LDA as the linear regression problem.

Sparse coding tries to find a dictionary $\mathbf{W}$ and a sparse coefficient matrix $\mathbf{S}$ whose product can best approximate the original data matrix. The column vectors of dictionary $\mathbf{W}$ can be regarded as the basis vectors and each column of sparse coefficient matrix $\mathbf{S}$ is the new representation of each data point in new space. One might further hope that the basis vectors can respect the intrinsic Riemannian structure, rather than ambient Euclidean structure. A natural

assumption here could be that if two data points $\mathbf{x}_i$ and $\mathbf{x}_j$ are close in the intrinsic geometry of the data distribution, then $\mathbf{s}_i$ and $\mathbf{s}_j$, the representations of these two points with respect to the new basis, are also close to each other [15]. GraphSC and MGrSc construct single graph or multiple graphs separately to map the weighted graphs to the sparse representations $\mathbf{S}$.

Our GRSLR model learns a projection matrix $\mathbf{B}$ to bridge the original feature space and the label space from the view of the subspace learning, which seems similar with GraphSC and MGrSc. But they are in fact very different. According to the definition of GraphSC and MGrSc, "sparse" means the learnt codes (original sample in the code space) are sparse while in our proposed model we enforce the projection matrix to be sparse instead of the learnt codes in GraphSC and MGrSc. Concretely, the GraphSC and MGrSc methods aim to learn a dictionary matrix and a sparse coefficient matrix to represent the training EEG data, where an L_1-norm constraint is imposed on the coefficient matrix. Different from the GraphSC and MGrSc, the target of our GRSLR method aims to learn a sparse transform matrix that map the data from feature space to the label space, where the $L_{2,1}$-norm constraint is imposed on the transform matrix, which results in the GROUP SPARSE of the transform matrix.

3 Experiments

3.1 Data Acquisition of RCLS Database

It is necessary to collect EEG emotion data to extend existing EEG databases about emotion and evaluate the performance of our GRSLR. Thus we have collected EEG emotion data of 27 subjects with 1000 Hz sample rate and 14 of these subjects constitute RCLS database (the left 13 subjects' data is invalid in the acquisition process) and used to evaluate our model. These 14 subjects are all right-handed and six of them are males while eight are females. The range ages is 21–26 and mean age is 23.3. None of subjects have a history of mental illness and emotional problems especially. All participants are informed the harmlessness of the equipment and gave written informed consent to participate in the experiment procedures, which were approved by the Ethics Committee of Affiliated Zhongda Hospital, Southeast University.

The following is the detailed acquisition protocol of the data of RCLS.

3.1.1 Stimulus Materials

Videos and audios always bring strong senses to people compared to images and music alternatively. Subjects are always feel strong emotions seeing a movie. The existed studies have been verified stability and reliability of movie clips to elicit the emotion of subjects [18,19]. Visual and auditory stimulation are shown to the participant to induce the brain to produce different emotions. Because of the subjects are Chinese students, we used a set of movie clips that from the Native Chinese Affective Video System [20]. These selected movie clips are departed into three target emotions, i.e. happy, sad and neutral emotions. To make the subjects better understand the movie, we selected the movie clips according to the following two criterions: (1) duration of the clips is relatively short, i.e., four minutes at most. (2) there is only one wanted expressing emotion in the clips. The details of movie clips are shown in Table 2.

3.1.2 Experiment Devices

The software NeuroScan is used as the signal recording software in this experiment. A 64-channel cap which channels are arranged as the extended international 10–20 system is used

Table 2 Details of the movie clips

Number	Movie	Target emotion
1	A big potato	Happy
2	The Eagle Shooting Heroes	Happy
3	Hands Up!	Happy
4	Flirting Scholar	Happy
5	Eat Hot Tofu Slowly	Happy
6	IDE interface repair	Neutral
7	IP encapsulation	Neutral
8	Projector	Neutral
9	Computer repair	Neutral
10	Hardware Conflicts	Neutral
11	ROB-B-HOOD	Sad
12	Butterfly Lovers	Sad
13	My Beloved	Sad
14	Nuan chun	Sad
15	My Sisters and Brothers	Sad

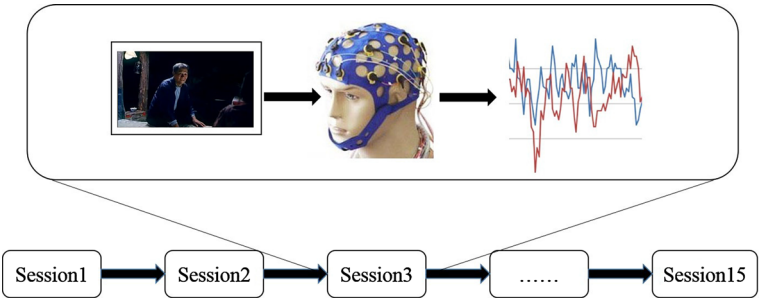

Fig. 1 The acquisition procedure of RCLS database

to receive the EEG signal. The sampling rate is set to 1000 Hz and the EOG and reference voltage are recorded.

3.1.3 Experiment Procedure

The whole procedure of the experiment is shown in Fig. 1 concretely. Subject at each session saw one random movie clip of the fifteen. At the end of each session, there is a rest.

3.2 Experiment on RCLS Database

3.2.1 Data and Experiment Performance Analysis

RCLS database contains 14 subjects and each subject has 15 sessions. After collection, the raw EEG data is acquired. The pure EEG data can be further obtained by employing some preprocessing strategies, such as removing EOG and artifacts. These 14 subject EEG emotion data are preprocessed by 5-order Butterworth filter to obtain 1∼100 Hz processed data. Then these data are cut with a window size of 1s without overlapping to subsequently

generate about 1600 clean samples for each subject. The RCLS database will be released in our website[1].

For EEG emotion signal, different frequency bands can carry with emotion information from different profiles. For this specific emotion recognition task, five bands consisting of delta δ (1–3 Hz), theta θ (4–7 Hz), alpha α (8–13 Hz), beta β (14–30 Hz) and gamma γ (31–50 Hz) are usually considered.

Here we extract some traditional time domain features, frequency domain features and time-frequency domain features to feed into our model and evaluate the performance on some classical machine learning methods. Time domain features include HOC [6,21], FD [22], STATISTICS [6] and HJORTH [23–25]. Higher-Order Crossings (HOC) [21] treats EEG as a zero-mean signal and captures the oscillatory pattern of EEG signal by calculating the number of values crossing zero. By repeating the above procedure, we can get higher order HOC features. In this experiment, we choose 20 orders considering the computation and recognition result. Fractal Dimension (FD) [22] measures the signal's complexity. Here we implement Higuchi algorithm [26] to get FD value on each channel. STATISTICS includes some simply statistical features, i.e., mean, standard deviation, maximum, 1st difference and 2nd difference. HJORTH developed three features of a time series, i.e., activity, mobility and complexity [23–25]. Frequency domain features include ENTROPY [27], PSD and DE [11]. ENTROPY uses the entropy of the power spectrum. Following the same feature extraction step in [11], we try to strictly reproduce the differential entropy (DE) of EEG signals as the features in our database. For a fixed length EEG segment, DE feature is equivalent to the logarithm energy spectrum in a certain frequency band. And in this section, we use moving average to smooth the PSD features and DE features. The results are shown in Table 3. Time-frequency domain feature includes WAVELET. Discrete wavelet transform decomposes the signal in different approximation and detailed levels corresponding to different frequency range while conserving the time information of the signal [6]. Here we use power, mean, variance and entropy of the 4-order wavelet coefficients as the WAVELET features.

In our experiments, we use the leave one session out strategy so that other researchers can reproduce all results of our proposed method. That is to say, one session of a subject is used for testing while the remaining sessions of this subject are used for training, and we loop all sessions for testing and average all testing accuracies as the final performance of this subject.

In the experiments, we use the classification accuracy as evaluation criteria of experimental performance. Suppose N_1 is the amount that the predicted labels are same with the ground-truth labels of test data, and the size of test data is N. The classification accuracy is calculated as $\frac{N_1}{N}$.

Here we compare our GRSLR with the linear SVM [29], which is a standard method and extended to multi-class version through one versus rest strategy, Random Forest (RF) and Canonical Correlation Analysis (CCA) [30], which are also standard methods for EEG emotion recognition, Graph Regularized Sparse Coding (GraphSC) [15] and Group Sparse Canonical Correlation Analysis (GSCCA) [31] based methods. These methods except RF have been used in the classification of EEG signals [9]. The results are shown in Table 3 when the parameters λ and μ in our GRSLR model are simply set to 1 and 1 respectively. To better measure the performance of our model, we search the parameter space of $\lambda \in \{1, 2, \ldots, 10\}$ and $\mu \in \{1, 2, \ldots, 10\}$ and the best result on each subject and the corresponding parameters on HJORTH features are shown in Table 4.

We can obtain two observations from Tabels 3 and 4:

[1] http://aip.seu.edu.cn.

Table 3 The mean accuracies and standard deviations of 14 subjects on RCLS database

ENTROPY						
Model	RF	SVM	CCA	GraphSC [15]	GSCCA [28]	GRSLR
ACC(%)	55.19	66.27	63.89	46.87	62.07	**67.97**
	(13.92)	(11.40)	(11.26)	(15.27)	(11.74)	(**10.44**)
PSD ($\delta+\theta+\alpha+\beta+\gamma$)						
Model	RF	SVM	CCA	GraphSC [15]	GSCCA [28]	GRSLR
ACC(%)	53.78	77.26	74.44	46.80	73.88	**78.21**
	(9.76)	(8.10)	(7.44)	(12.55)	(7.84)	(**6.49**)
DE ($\delta+\theta+\alpha+\beta+\gamma$)						
Model	RF	SVM	CCA	GraphSC [15]	GSCCA [28]	GRSLR
ACC(%)	59.04	77.85	78.32	67.83	78.36	**80.27**
	(9.16)	(9.21)	(8.15)	(11.58)	(8.32)	(**6.86**)
HOC						
Model	RF	SVM	CCA	GraphSC [15]	GSCCA [28]	GRSLR
ACC(%)	48.92	**74.97**	59.04	47.81	59.21	70.79
	(10.84)	(**10.18**)	(6.87)	(17.28)	(7.06)	(7.71)
FD						
Model	RF	SVM	CCA	GraphSC [15]	GSCCA [28]	GRSLR
ACC(%)	56.63	69.86	69.33	65.11	68.57	**72.09**
	(13.40)	(10.28)	(10.71)	(13.14)	(11.05)	(**10.61**)
WAVELET						
Model	RF	SVM	CCA	GraphSC [15]	GSCCA [28]	GRSLR
ACC(%)	52.78	77.88	78.43	47.29	78.15	**79.66**
	(10.99)	(8.73)	(8.04)	(17.58)	(8.17)	(**7.57**)
STATISTICS						
Model	RF	SVM	CCA	GraphSC [15]	GSCCA [28]	GRSLR
ACC(%)	66.01	79.01	81.27	70.13	80.88	**81.99**
	(12.21)	(8.56)	(6.97)	(11.06)	(7.05)	(**7.06**)
HJORTH						
Model	RF	SVM	CCA	GraphSC [15]	GSCCA [28]	GRSLR
ACC(%)	63.97	79.08	79.68	66.75	79.42	**81.13**
	(11.62)	(8.30)	(7.74)	(13.73)	(7.76)	(**7.45**)

The results of GRSLR is measured under $\lambda = 1$ and $\mu = 1$
The best results are highlighted in bold

(1) From Tabel 3, we can see that GRSLR gets the best performance in most of the features except HOC feature even though we simply set λ and μ.
(2) There are some diversities and variances in EEG emotion signals from different subjects. In Table 4, the best results achieved with different parameters, which indicates that the EEG emotion signals are different among different people even collected in the same environment and conditions. This reflects that EEG emotion signals are complicated among different subjects on a side.

Table 4 The parameters in best performance on Hjorth features of RCLS database

Subject	1	2	3	4	5	6	7
λ	1	3	1	1	1	9	1
μ	2	3	9	1	8	7	1
ACC(%)	90.49	76.85	92.03	89.88	80.70	76.92	84.99
Subject	8	9	10	11	12	13	14
λ	2	1	1	1	1	1	1
μ	6	10	1	2	1	9	7
ACC(%)	78.11	90.81	80.74	87.04	70.83	87.14	81.89
Mean/Std(%)	83.46/6.44						

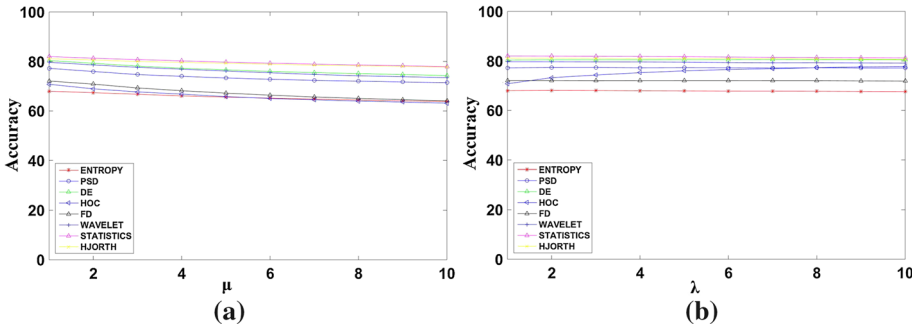

Fig. 2 The accuracy curves of trade-off parameter sensitivity experiments on different features. **a** $\lambda = 1$, **b** $\mu = 1$

3.2.2 Parameter Sensitivity Analysis

The proposed GRSLR has two trade-off parameters λ and μ. The selection of them would affect the performance on EEG emotion recognition. In this experiment, we fix one trade-off parameter as 1 and change the other one from 1 to 10 to conduct experiments to investigate whether the performance of GRSLR would be sensitive to the slight changes of these two parameters. The experiments are shown in Fig. 2. It is clear to see that the accuracy varies slightly with respect to the changes of λ and μ respectively, which can show that our GRSLR is less sensitive to its trade-off parameters.

3.2.3 Confusion Matrix

To see the results of recognizing each emotion class, we depict the corresponding confusion matrices of the 14 subject data in Fig. 3. From this figure, we can obtain two observations:

(a) There exist some similarities in same emotion EEG signals of a subject. From Fig. 3, we can observe that three emotions of each subject have been classified well, which indicates that it is efficient to use the EEG emotion signal of a subject to train a model and predict an unknown EEG emotion signal of this subject.
(b) The positive emotion is much easier to be recognized whereas the negative emotion is relatively difficult to be recognized. We calculate the mean accuracies of three emotions

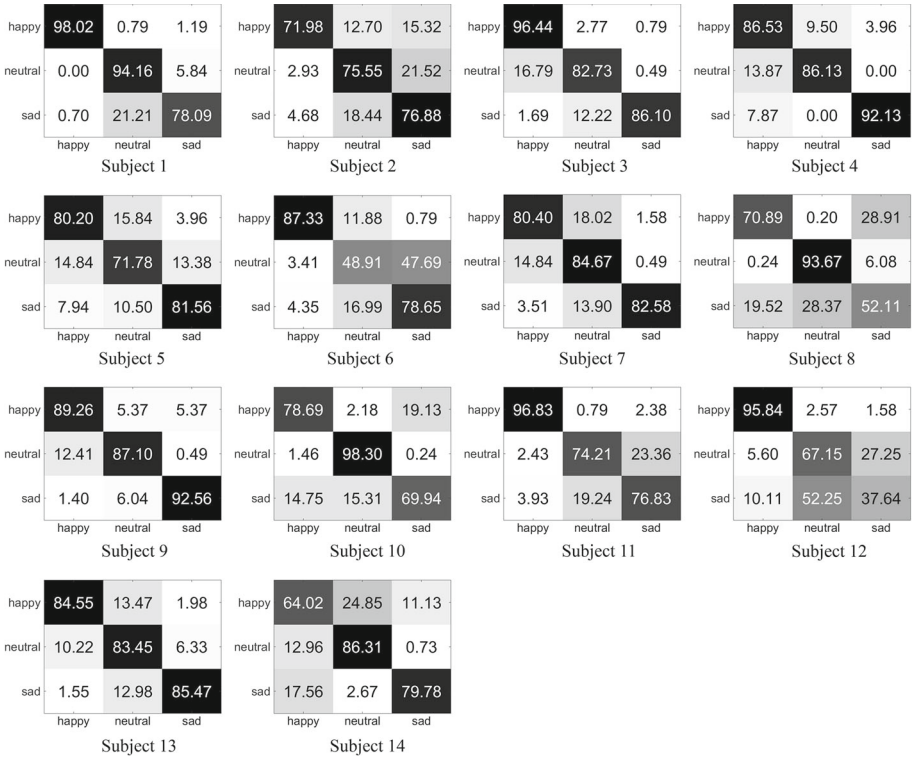

Fig. 3 The confusion matrix of GRSLR on HJORTH features of RCLS database when $\lambda = 1$ and $\mu = 1$

in all 14 subjects from Fig. 3. That is happy 84.34%, neutral 80.96% and sad 76.45%, which indicate that the performance of three emotions, i.e., $happy > neutral > sad$. From confusion matrices of subject 5, 8 and 12 especially, we can observe that the neutral and sad are much easier to be confused.

3.2.4 Brain Electrodes Activity Maps

The activity maps of brain electrodes are shown in Fig. 4, in which we compute the contribution of each electrode to the last classification with our GRSLR on HJORTH features and map them into the corresponding electrodes. In the maps, the contribution of each electrode is calculated by using the 2-norm operation on the corresponding block $\mathbf{B}_i$ of the matrix $\mathbf{B}$.

Here we are interested to obtain two regularities in Fig. 4:

(a) The top brain electrodes are not used mostly for classification in all 14 subjects. This indicates that there is no contribution to emotion expression for top brain area.
(b) Some electrodes have a great contribution to the classification especially those located on lateral temporal according to the sum of $\mathbf{B}$ of all 14 subjects. And there is asymmetry between left-sided and right-sided hemispheric activation. This conforms to the cognition observation in psychology [32].

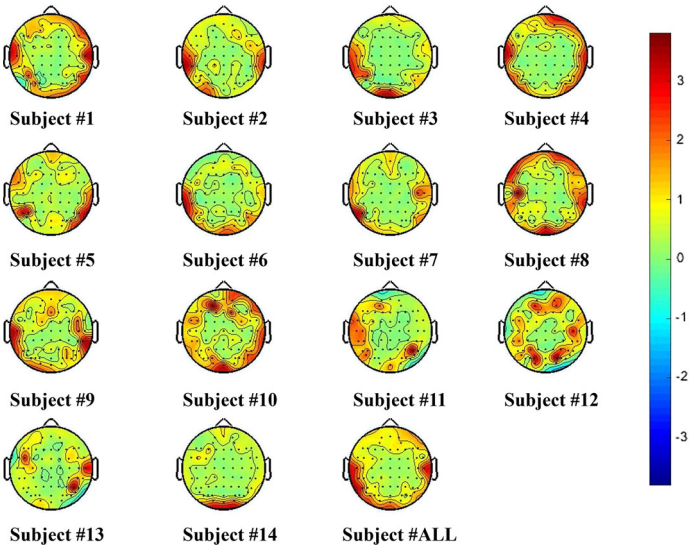

Fig. 4 The Brain electrodes activity maps of GRSLR on HJORTH features of RCLS database when $\lambda = 10$ and $\mu = 1$. The last one is the brain map with the sum of all 14 subjects. Deeper color represents larger contribution in the learnt model **B**

Table 5 The mean accuracies and standard deviations of different frequency bands on SEED database

Band	δ	θ	α	β	γ
GRSLR	**0.5673**	0.6116	**0.6985**	**0.7803**	**0.7884**
	(**0.1181**)	(0.0990)	(**0.1049**)	(**0.0903**)	(**0.1137**)
GSCCA [28]	0.5236	0.6059	0.6552	0.7163	0.7219
	(0.1018)	(0.1213)	(0.1159)	(0.1193)	(0.1306)
GraphSC [15]	0.5412	0.5725	0.5828	0.6856	0.7215
	(0.1091)	(0.0929)	(0.1118)	(0.1242)	(0.1442)
CCA	0.5214	0.6043	0.6520	0.7128	0.7180
	(0.1024)	(0.1194)	(0.1163)	(0.1198)	(0.1304)
SVM	0.5604	**0.6323**	0.6707	0.7495	0.7597
	(0.1121)	(**0.1154**)	(0.0866)	(0.1275)	(0.1218)
RF	0.5128	0.5681	0.5806	0.6592	0.6533
	(0.0933)	(0.1118)	(0.1527)	(0.1330)	(0.1467)

The best results are highlighted in bold

3.3 Experiment on SEED Database

The proposed GRSLR is also verified on the emotion database SEED [11], which contains fifteen subjects across different sessions. Following the same features released in [11], we use differential entropy (DE) of EEG signals as the features to feed into our model.

Here we also use the leave one session out strategy to conduct the experiments. To investigate the influence of different frequency bands for EEG emotion recognition, we conduct the experiments on five frequency bands separately and the mean accuracies of all subjects are shown in Table 5. Furthermore, to investigate the influence of the combination of different

Table 6 The mean accuracies and standard deviations of different combinations on frequency bands on SEED database

Band	$\beta+\gamma$	$\alpha+\beta+\gamma$	$\theta+\alpha+\beta+\gamma$	$\delta+\theta+\alpha+\beta+\gamma$
GRSLR	**0.8263**	**0.8484**	**0.8628**	**0.8841**
	(**0.1023**)	(**0.0926**)	(**0.0811**)	(**0.0821**)
GSCCA [28]	0.7246	0.7414	0.7451	0.7484
	(0.1286)	(0.1295)	(0.1189)	(0.0998)
GraphSC [15]	0.7531	0.7487	0.7786	0.8103
	(0.1020)	(0.1035)	(0.0917)	(0.0850)
CCA	0.7174	0.7310	0.7402	0.7476
	(0.1304)	(0.1326)	(0.1193)	(0.1042)
SVM	0.7843	0.8097	0.8278	0.8523
	(0.1614)	(0.1569)	(0.1575)	(0.1402)
RF	0.6791	0.7038	0.7191	0.7376
	(0.1448)	(0.1216)	(0.1100)	(0.0936)

The best results are highlighted in bold

frequency bands for EEG emotion recognition, we conduct the experiments of using various combinations of five frequency bands' features, which results are shown in Table 6.

Here we also compare with the RF, SVM, CCA, GraphSC and GSCCA based methods. Note that we don't compare with the method of Zheng et.al [11] due to its no reproduction in both coding and data partition. From these results, we can observe that our GRSLR achieves best accuracies at all cases except in the θ band. In addition, we can obtain two external observations from the two tables:

(1) Each frequency band's features do have the discriminative ability to distinguish EEG emotions. Furthermore, high frequency band carries more discriminative emotion information. As shown in Table 5, the performance is improved with the increase of frequency, i.e., $\gamma > \beta > \alpha > \theta > \delta$.
(2) Combination of different frequency bands provide more enhanced discriminative information that is benefit for emotion recognition. As shown in Table 6, the combination of all five frequency bands' features have the most discriminative ability to EEG emotion recognition, which indicates that the five frequency bands are complementary with each other to reflecting human emotion.

To verify the effectiveness of two constraint terms in our model, we especially conduct the experiments with the cases of $\lambda = 0$ or $\mu = 0$, which means removing the manifold constraint or sparse constraint respectively. For this, we also use the leave one session out strategy and the mean accuracies of all the subjects are shown in Table 7. If only considering the sparse constraint, i.e., $f_d(\mathbf{B}) + f_c(\mathbf{B})$ [Eq. (5) and Eq. (4)], the improvement is about 7 percent compared to the original model LSR. After adding the manifold structure preservation term, we can further promote the emotion recognition accuracy more than 4%. Thus it demonstrates that the gains of our method come from the use of two constraint terms to some extent.

Besides, we depict the activity maps of brain electrodes for different bands according to the learnt weight matrix $\mathbf{B}$ in our proposed GRSLR model in Fig. 5. Here we calculate the sum of block $\mathbf{B}_i$ in $\mathbf{B}$ with all the 15 subjects' experiment results. In the maps, the contribution of each electrode is calculated by using the 2-norm operation on the corresponding block $\mathbf{B}_i$ of the matrix $\mathbf{B}$. We can find that the areas of lateral temporal and front brain are heavily activated

Table 7 The performance of two constraints, i.e., sparse and manifold regularization terms

Model	$f_d(\mathbf{B})$	$f_d(\mathbf{B})+ f_c(\mathbf{B})$	$f_d(\mathbf{B})+ f_c(\mathbf{B})+f_m(\mathbf{B})$
GRSLR	0.7718	0.8417	0.8841

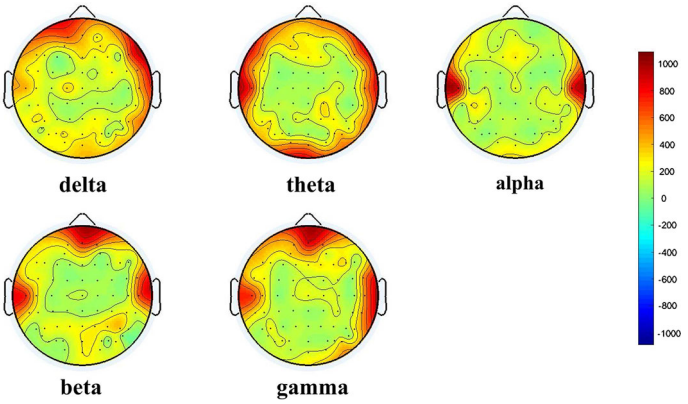

Fig. 5 Brain electrodes activity maps of SEED. Here we calculate the sum of **B** with all the 15 subjects' experiment results. Deeper color represents larger contribution in the learnt model **B**

especially in α, β and γ frequency bands. That means these areas have more contribution to emotion expression, which also conforms to the cognition observation in psychology [32]. For δ and θ bands, their maps seem have no regularity, which provides a possible explanation why their performances are still inferior to other three frequency bands.

4 Conclusion

In this paper, we proposed a novel model called graph regularized sparse linear regression (GRSLR) to deal with EEG emotion recognition and provided an EEG emotion database called RCLS. In GRSLR, two regularization terms, group sparse term and manifold preserving term, are introduced into the linear regression model. As observed from the experiments, we can benefit from three folds: (1) adaptively implement channel selection by using sparse term; (2) reduce overfitting for small database by using manifold embedding; and (3) meanwhile learn more discriminative projecting space by using linear regression. We conduct the experiment with leave one session out strategy. The comparison experiments also demonstrate that our proposed GRSLR is more effective on SEED and RCLS.

Acknowledgements This work was supported by the National Basic Research Program of China under Grant 2015CB351704, the National Natural Science Foundation of China under Grant 61572009, Grant 61772276, and Grant 61602244, and the Jiangsu Provincial Key Research and Development Program under Grant BE2016616.

References

1. Picard RW (1997) Affective computing, vol 252. MIT Press, Cambridge
2. Lotfi E, Akbarzadeh-T M-R (2014) Practical emotional neural networks. Neural Netw 59:61–72

3. Britton JC, Phan KL, Taylor SF, Welsh RC, Berridge KC, Liberzon I (2006) Neural correlates of social and nonsocial emotions: an fMRi study. Neuroimage 31(1):397–409
4. Etkin A, Egner T, Kalisch R (2011) Emotional processing in anterior cingulate and medial prefrontal cortex. Trends Cogn Sci 15(2):85–93
5. Lindquist KA, Barrett LF (2012) A functional architecture of the human brain: emerging insights from the science of emotion. Trends Cogn Sci 16(11):533–540
6. Jenke R, Peer A, Buss M (2014) Feature extraction and selection for emotion recognition from eeg. IEEE Trans Affect Comput 5(3):327–339
7. Li M, Lu B-L (2009) Emotion classification based on gamma-band eeg. In: Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual international conference of the IEEE, 2009, pp 1223–1226
8. Lin Y-P, Wang C-H, Jung T-P, Wu T-L, Jeng S-K, Duann J-R, Chen J-H (2010) Eeg-based emotion recognition in music listening. IEEE Trans Biomed Eng 57(7):1798–1806
9. Zheng W (2017) Multichannel eeg-based emotion recognition via group sparse canonical correlation analysis. IEEE Trans Cogn Dev Syst 9(3):281–290
10. Zhang Y, Ji X, Zhang S (2016) An approach to eeg-based emotion recognition using combined feature extraction method. Neurosci Lett 633:152–157
11. Zheng W-L, Lu B-L (2015) Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. IEEE Trans Auton Ment Dev 7(3):162–175
12. Zheng W (2014) Multi-view facial expression recognition based on group sparse reduced-rank regression. IEEE Trans Affect Comput 5(1):71–85
13. Belkin M, Niyogi P (2001) Laplacian eigenmaps and spectral techniques for embedding and clustering. In: NIPS, vol 14, pp 585–591
14. Lin Z, Chen M, Ma Y (2010) The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices, arXiv preprint arXiv:1009.5055,
15. Zheng M, Bu J, Chen C, Wang C, Zhang L, Qiu G, Cai D (2011) Graph regularized sparse coding for image representation. IEEE Trans Image Process 20(5):1327–1336
16. Jin T, Yu Z, Li L, Li C (2015) Multiple graph regularized sparse coding and multiple hypergraph regularized sparse coding for image representation. Neurocomputing 154:245–256
17. De la Torre F (2012) A least-squares framework for component analysis. IEEE Trans Pattern Anal Mach Intell 34(6):1041–1055
18. Gross JJ, Levenson RW (1995) Emotion elicitation using films. Cogn Emot 9(1):87–108
19. Schaefer A, Nils F, Sanchez X, Philippot P (2010) Assessing the effectiveness of a large database of emotion-eliciting films: a new tool for emotion researchers. Cogn Emot 24(7):1153–1172
20. Xu P, Huang Y, Luo Y (2010) Establishment and assessment of native Chinese affective video system. Chin Mental Health J 24(7):551–554
21. Petrantonakis PC, Hadjileontiadis LJ (2010) Emotion recognition from eeg using higher order crossings. IEEE Trans Inf Technol Biomed 14(2):186–197
22. Sourina O, Liu Y (2011) A fractal-based algorithm of emotion recognition from eeg using arousal-valence model. In: BIOSIGNALS, pp 209–214
23. Hjorth B (1970) Eeg analysis based on time domain properties. Electroencephalogr Clin Neurophysiol 29(3):306–310
24. Ansari-Asl K, Chanel G, Pun T (2007) A channel selection method for eeg classification in emotion assessment based on synchronization likelihood In: Signal processing conference, 15th European, IEEE, 2007, pp 1241–1245
25. Horlings R, Datcu D, Rothkrantz LJ (2008) Emotion recognition using brain activity. In: Proceedings of the 9th international conference on computer systems and technologies and workshop for PhD students in computing, ACM, 2008, p. 6
26. Liu Y, Sourina O (2013) Real-time fractal-based valence level recognition from eeg. In: Transactions on computational science XVIII, Springer, pp 101–120
27. Inouye T, Shinosaki K, Sakamoto H, Toi S, Ukai S, Iyama A, Katsuda Y, Hirano M (1991) Quantification of eeg irregularity by use of the entropy of the power spectrum. Electroencephalogr Clin Neurophysiol 79(3):204–210
28. Lin D, Zhang J, Li J, Calhoun VD, Deng H-W, Wang Y-P (2013) Group sparse canonical correlation analysis for genomic data integration. BMC Bioinformatics 14(1):245
29. Suykens JA, Vandewalle J (1999) Least squares support vector machine classifiers. Neural Process Lett 9(3):293–300
30. Thompson B (2005) Canonical correlation analysis. Encyclopedia of statistics in behavioral science
31. Parkhomenko E, Tritchler D, Beyene J et al (2009) Sparse canonical correlation analysis with application to genomic data integration. Stat Appl Genet Mol Biol 8(1):1–34
32. Mauss IB, Robinson MD (2009) Measures of emotion: a review. Cogn Emot 23(2):209–237