Journal of Ambient Intelligence and Humanized Computing (2023) 14:5529–5540
https://doi.org/10.1007/s12652-020-02620-9

ORIGINAL RESEARCH

Transfer discriminative dictionary learning with label consistency for classification of EEG signals of epilepsy

Tongguang Ni[1] · Xiaoqing Gu[1] · Yizhang Jiang[2]

Received: 7 July 2020 / Accepted: 12 October 2020 / Published online: 21 October 2020

Abstract

EEG signal classification play an important role in recognition of epilepsy. Recently, dictionary learning algorithms have shown the effectiveness in this field. When designing dictionaries, due to highly non-stationary of EEG signals, and collecting signals existing in different stimulus and drug modes, training and testing scenarios may be different. Thus, the performance of classical dictionary learning algorithms is unsatisfactory. In this paper, a transfer discriminative dictionary learning with label consistency (called TDDLLC) algorithm is proposed for EEG signal classification. Since each EEG signal can be represented as a linear combination of dictionary atoms, and some atoms are dataset independent, two dictionaries are learned simultaneously in source domain (SD) and target domain (TD) respectively where the discrepancy between two dictionaries is minimized. Meanwhile, utilizing the label information of samples in SD and a small number of labeled samples in TD, these dictionaries are learned with the aim of achieving discriminative abilities. To avoid the NP-hard problem, ℓ_1-norm regularization term is used in TDDLLC, and objective function is solved by block-coordinate descent method. Extensive experiments have been performed on Bonn dataset and show the validity of the TDDLLC algorithm.

Keywords Dictionary learning · Transfer learning · Label consistency · EEG signal classification · Epilepsy

1 Introduction

Epilepsy is a transient brain dysfunction caused by a brain injury. One of its main characters is to have recurring seizures. Epilepsy is one of the most common diseases in the human brain and is extremely harmful to human health (Mendoza et al. 2019; Wang et al. 2018; Zhang et al. 2018a). Electroencephalogram (EEG) signals from patients with epilepsy contain a large amount of physiological and pathological information in the brain, so the intelligent identification of EEG is very important for detection of epilepsy (Hu et al. 2019; Kevric et al. 2017; Sridhar et al. 2019). Meanwhile, more and more researches are focus on the automatic recognition of EEG signals of epilepsy. The EEG epilepsy recognizer can be considered as a classifier in pattern recognition. For classical epilepsy recognition, feature extraction and classification are two core components. The goal of feature extraction is to characterize distinctive EEG patterns, and the feature representation directly affects the performance of epilepsy classifier. Thus, various feature extraction algorithms are employed to represent EEG signal, such as principal component analysis (PCA), wavelet packet decomposition (WPD) and local binary patterns (LBPs) (Jiang et al. 2017a; Gu et al. 2017; Seo et al. 2019). On the other hand, the end-to-end classifiers have received a lot of attention since they do not need the process of feature extraction, such as sparse representation based classification (SRC) and deep learning (Amin et al. 2020; Kundishora et al. 2017).

By coding under a dictionary, sparse representation (SR) can represent data signals adaptively with a few atoms under some orthogonal transformations. In the EEG signals processing and classification domain, due to the advantage of sparsifying dictionary, Abolghasemi et al. (2015) subtracted ballistocardiogram from the original EEG signals and retained the clean EEG is obtained based on dictionary

✉ Yizhang Jiang
yzjiang@jiangnan.edu.cn

Tongguang Ni
hbxtntg-12@163.com

Xiaoqing Gu
czxqgu@163.com

[1] School of Information Science and Engineering, Changzhou University, Changzhou 213164, China

[2] School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China

learning framework. Ameri et al. (2016) proposed a dictionary pair learning algorithm and learned the sparse coefficients by a simple linear projection. Akhavan et al. (2018) proposed a multi-channel discriminative sparse representation model integrated with the discriminative spatial filter. Sreej et al. (2019a) developed a weighted SRC using the dissimilarity information between the training and testing samples. Then the sparse coefficients for the testing samples are obtained on the weighted dictionary. To relief the calibration burden during data acquisition in the brain–computer interface, Sreej et al. (2019b) proposed a sparse group representation model, which exploited the intersubject information and used the constraints of within-group and group-wise sparse. The experimental results showed that this model performed efficiently in real-time applications. SR represents data signals with as few atoms as possible from a dictionary. Dictionary learning is the crucial process of SR. Dictionary learning for classification usually utilizes the discriminative information, including class specific dictionary (learning specific dictionary for each class) and class shared dictionary (learning a shared for all classes). The representative algorithms of the later dictionary learning are discriminative K-singular value decomposition (KSVD) (Zhang et al. 2010) and Label Consistent KSVD (LC-KSVD) (Jiang et al. 2013) KSVD algorithm directly incorporates discriminative terms to minimize the reconstruction error. LC-KSVD incorporates a label consistent constraint into KSVD and guarantees the discrimination of learned representation. The advantage of class shared dictionary is that it can well represent certain common information shared by samples from different classes (Zhang et al. 2018b).

However, classification of EEG signals of epilepsy is still a challenging problem in a new scenario. One of the main challenges is different stimulation paradigm and medication, the change of EEG acquisition equipment et al. under different tasks. Second, EEG signals are highly non-stationary and its statistical features change over time in different patients and even the same patient. Third, often few labeled samples are available in new scenarios. Therefore, traditional classification methods are weak in robustness, generalization ability and accuracy. To solve these problems, one of the recent trends is transfer learning method. Transfer learning method utilizes other related datasets to assist the learning task in the new scenario without recollecting a lot of labeled samples. In the idea of transfer learning, apart from a small number of data related to the current scenario, a large number of data from other domain can be used as auxiliary data for training, and these data are not necessarily drawn from the same distribution. The related scenario is called as source domain (SD) and the current scenario is called as target domain (TD). For instance, the case of classification of EEG signals of epilepsy, only healthy signals and signals during seizures are used for training. However, the test signals can contain the condition that is not presented in the training signals such as signals collected from seizure free intervals (Jiang et al. 2019; Shao et al. 2020).

In this work, we try to solve the EEG epileptic seizure classification using transfer discriminative dictionary learning. The motivation of our idea is that EEG signals in different domains share certain common knowledge that is potentially independent of the domains. The dictionary for reconstructed EEG signals in different domains also will share certain common knowledge, and dictionary learned in the SD can help to improve the quality of dictionary in the TD. To improve the discriminative ability of learned dictionary in TD, we use label consistent constraint in dictionary learning, and propose a transfer discriminative dictionary learning with label consistent (called TDDLLC), with the goal of achieving high performance in cross domain EEG epileptic seizure classification. Our work mainly focuses on threefold. (1) By minimizing the discrepancy between dictionaries in the SD and TD, the learned dictionaries have the ability to transfer the sparse codes from the related scenarios to current scenario. (2) To avoid the NP-hard problem, ℓ_1-norm regularization term (Shao et al. 2020) is used in TDDLLC, instead of the commonly used L_p-norm sparsity constraint. (3) TDDLLC is optimized by adopting the block-coordinate descent (BCD) (Liu et al. 2014) method, which is parameter-free and no need to tune the learning rate, so that TDDLLC can find optimal solution easily. (4) We apply the proposed algorithm on Bonn EEG dataset (Andrzejak et al. 2001), and the experimental results on several transfer learning tasks demonstrate that TDDLLC can achieve satisfactory performance in epileptic seizure classification.

The rest of the paper is organized as follows. In Sect. 2, we first review the existing works on dictionary learning and LC-KSVD. Then we briefly introduce the transfer learning algorithms that are used for epileptic seizure classification. In Sect. 3, we present transfer discriminative dictionary learning with label consistent method. In Sect. 4, we perform the experiment on EEG data. Finally, we conclude our study in Sect. 5.

2 Related work

2.1 Dictionary learning

In a sparse representation framework, a sample $\mathbf{x} \in \mathbb{R}^{d\times 1}$ with d dimension is represented on the dictionary $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_K] \in \mathbb{R}^{d\times K}$ through sparse coding coefficients $\boldsymbol{s}$ as:

$$\mathbf{x} \approx \mathbf{Ds}, \tag{1}$$

where vector $\mathbf{s} = [s_1, s_2, \ldots, s_K] \in \mathbb{R}^{K\times 1}$, and K is the number of atoms in $\boldsymbol{D}$. The coefficient optimization problem is expressed using the L_1 norm as:

$$\arg\min_{\mathbf{s}} \|\mathbf{s}\|_1, \quad s.t.\ \|\mathbf{x} - \mathbf{Ds}\|_2^2 < \varepsilon, \tag{2}$$

where ε is the error tolerance, and $\|\mathbf{s}\|_1 = \sum |s_i|$. Then Eq. (2) can be rewritten in terms of a cost function:

$$\arg\min_{\mathbf{s}} \|\mathbf{x} - \mathbf{Ds}\|_2^2 + \lambda \|\mathbf{s}\|_p, \tag{3}$$

where λ is the regularization parameter. Equation (3) minimizes reconstruction errors, while maintains the sparse constraints. The $\|.\|_p$ is the L_p-norm of a matrix (vector), where $p=0, 1, 2$.

2.2 LC-KSVD

In the classification task, learning a dictionary with discriminative information helps to promote the discrimination ability of sparse models. A well known SRC method is LC-KSVD. LC-KSVD associates class label of signals with each dictionary item, and incorporate errors of classification, reconstruction and discriminative sparse-code into a sparse representation model. Using sparsity constant with L_0-norm, the objective function of LC-KSVD is presented as:

$$\arg\min_{\mathbf{D,S,A,W}} \|\mathbf{X} - \mathbf{DS}\|_F^2 + \beta\|\mathbf{H} - \mathbf{WS}\|_F^2 + \gamma\|\mathbf{Q} - \mathbf{AS}\|_F^2, \quad s.t.\ \forall i\ \|\mathbf{s}_i\|_0 \le T, \tag{4}$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n] \in \mathbb{R}^{d\times n}$ involves n input data and c categories of signals. $\mathbf{S} \in \mathbb{R}^{K\times n}$ is the sparse codes of $\boldsymbol{X}$. $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_K] \in \mathbb{R}^{K\times n}$ is the discriminative sparse code matrix. Each element $\mathbf{q}_i = [q_i^1, \ldots, q_i^K]^T \in \mathbb{R}^K$ in $\boldsymbol{Q}$ contains zero values and nonzero values, where nonzero values indicate $\mathbf{x}_i$ and the dictionary item $\mathbf{d}_k$ belong to the same class. $\mathbf{A} \in \mathbb{R}^{K\times K}$ is a linear transformation matrix. $\mathbf{H} \in \mathbb{R}^{c\times n}$ is the class label matrix that corresponds to $\boldsymbol{X}$. $\mathbf{W} \in \mathbb{R}^{c\times K}$ is the classifier parameter matrix. T is the sparsity constraint factor. γ and β are two positive parameters.

Using the KSVD algorithm (Jiang et al. 2013), Eq. (4) can be solved by the following problem:

$$\arg\min_{\mathbf{D,A,S,W}} \left\| \begin{pmatrix} \mathbf{X} \\ \sqrt{\gamma}\mathbf{Q} \\ \sqrt{\beta}\mathbf{H} \end{pmatrix} - \begin{pmatrix} \mathbf{D} \\ \sqrt{\gamma}\mathbf{A} \\ \sqrt{\beta}\mathbf{W} \end{pmatrix} \mathbf{S} \right\|_2^2, \quad s.t.\ \forall i\ \|\mathbf{s}_i\|_0 \le T. \tag{5}$$

For a testing signal $\boldsymbol{x}$', its sparse representation s' can be computed with the optimal $\boldsymbol{D}$ by:

$$\arg\min_{\mathbf{a}'} \|\mathbf{x}' - \mathbf{Ds}'\|_2^2, \quad s.t.\ \|\mathbf{s}'\|_0 \le T. \tag{6}$$

Then LC-KSVD obtains the class label c of $\boldsymbol{x}$' by finding the largest element of linear classifier $\mathbf{Ws}'$, i.e.,

$$c = \arg\max_c(\mathbf{Ws}'). \tag{7}$$

LC-KSVD simultaneously learns all parameters $\{\boldsymbol{D}, \boldsymbol{S}, \boldsymbol{A}, \boldsymbol{W}\}$. Sine LC-KSVD optimizes the objective function by Orthogonal Matching Pursuit (OMP) method, which is easy to fall into the NP-hard problem and tends to obtain the suboptimal sparse solution (Shao et al. 2020).

2.3 Transfer learning

At present, machine learning methods are most commonly used in EEG epileptic seizure classification. These classification algorithms can struggle with differences between EEG signals, because they are often in a scenario where the distribution of training and test data is different. Transfer learning provides a general strategy to incorporate sample data drawn from different distributions. The types of transferred knowledge in transfer learning can be categorized into (Pan et al. 2010; Zhang et al. 2019): instance transfer, feature representation transfer and parameter transfer. In this study, we use parameter transfer, considering it is universal and flexibility to apply to existing excellent algorithms. Deng et al. (2018) and Yang et al. (2014) integrated transfer learning into the framework of Takagi–Sugeno–Kang (TSK) fuzzy system to reduce the discrepancy in data distribution between SD and TD. Jiang et al. (2017b) combined γ-least squares regression with transfer learning, and transferred both model parameter and label space from SD to TD. Xia et al. (2020) developed a cross-domain classification model, which utilizes both the data global structure of labeled samples in SD and soft clustering regularization term of unlabeled samples in TD, and reduces the distribution difference between SD and TD in the kernel space.

However, feature extraction is the key process in these algorithms. Feature extraction algorithms used in EEG signal processing are commonly divided into three main categories, including time domain analysis, frequency domain analysis, and time–frequency analysis. As different feature extraction methods stem from their different basis, classification results will be different. This brings us to propose a transfer dictionary learning to directly handle with original EEG signals without feature extraction. Recently, several transfer learning based dictionary learning algorithms have been proposed. Chen et al. (2012) proposed a domain adaptation dictionary for image denoising, which transfers the SD dictionary to TD by a regularization term in the energy function. Zhu et al. (2016) proposed a domain-adaptive dictionary pair and combined it into an AdaBoost classifier to perform image classification and human action recognition. Zheng et al. (2016) developed a transferable dictionary for across view scenario. This algorithm learned a common dictionary and view-specific dictionaries, and the former modeled the view-shared features and the latter modeled the features correspond to one view. Although these algorithms are effective in their respective application, transfer dictionary learning is still an open problem for EEG epileptic seizure classification.

3 Transfer discriminative dictionary learning with label consistency

3.1 Objective function of TDDLLC

For a transfer classification problem, given two domain data $\mathbf{T}_s$ and $\mathbf{T}_t$, the source domain $\mathbf{T}_s$ is denoted as $\mathbf{T}_s = \{(\mathbf{x}_{s,1}, h_{s,1}), (\mathbf{x}_{s,2}, h_{s,2}), \ldots, (\mathbf{x}_{s,n}, h_{s,n})\}$ and the target domain $\mathbf{T}_t$ is denoted as $\mathbf{T}_t = \{(\mathbf{x}_{t,1}, h_{t,1}), \ldots, (\mathbf{x}_{t,m}, h_{t,m})\}$. Each sample in $\mathbf{T}_s$ and $\mathbf{T}_t$ is with d-dimensional features, and n, m are the size of $\mathbf{T}_s$ and $\mathbf{T}_t$, respectively ($m < n$).

In view of dictionary learning, dictionary is a set of substrates in the feature space, thus, to reveal the relationship between tasks in SD and TD, TDDLLC learns two dictionaries in two domains respectively, and simultaneously minimizes the discrepancy of dictionary $\boldsymbol{D}$ between SD and TD. In addition, we assume the obtained classifiers are similar in SD and TD, thus TDDLLC minimizes the diversity of parameter $\boldsymbol{W}$ to enhance the similarity of two classifiers. Eventually, our goal is classify the labels of the testing samples in TD. The framework of TDDLLC can be formulated as follows:

$$\min\ J(\mathbf{D}_s, \mathbf{S}_s, \mathbf{W}_s, \mathbf{A}_s) + J(\mathbf{D}_t, \mathbf{S}_t, \mathbf{W}_t, \mathbf{A}_t) + \mu_1 f(\mathbf{D}_s, \mathbf{D}_t) + \mu_2 f(\mathbf{W}_s, \mathbf{W}_t), \tag{8}$$

where $\mathbf{D}_s(\mathbf{D}_t)$, $\mathbf{W}_s(\mathbf{W}_t)$, and $\mathbf{A}_s(\mathbf{A}_t)$ are dictionary, classifier parameter matrix and sparse matrix in SD (TD). The positive constant μ_1 and μ_2 are similarity parameters of dictionaries and classifier parameters, respectively. The function $J()$ is the objective function of dictionary learning, and the first two terms are dictionary learning in data of SD and TD, respectively. The function $f()$ is used to enhance the similarity of $\mathbf{D}_s$ and $\mathbf{D}_t$, and is also used to quantify the discrepancy between $\mathbf{W}_s$ and $\mathbf{W}_t$ in the SD and TD.

Considering the discriminative abilities of label consistent constraint, the function $J()$ can be represented as:

$$J(\mathbf{D}, \mathbf{S}, \mathbf{W}, \mathbf{A}) = \min\ \|\mathbf{X} - \mathbf{DS}\|_F^2 + \beta\|\mathbf{H} - \mathbf{WS}\|_F^2 + \gamma\|\mathbf{Q} - \mathbf{AS}\|_F^2 + 2\varepsilon\|\mathbf{S}\|_{\ell_1}$$

$$s.t.\ \forall k\|\mathbf{d}_k\|_2^2 \le 1,\ \|\mathbf{a}_k\|_2^2 \le 1,\ \|\mathbf{w}_k\|_2^2 \le 1, \tag{9}$$

where $\mathbf{d}_k$, $\mathbf{a}_k$ and $\mathbf{w}_k$ are the kth atom of dictionary $\mathbf{D}$ and sparse matrix $\boldsymbol{X}$ and classifier parameter matrix $\boldsymbol{W}$, respectively. Three constraints $\left\|\mathbf{d}_k\right\|_2^2 \le 1$, $\left\|\mathbf{a}_k\right\|_2^2 \le 1$ and $\left\|\mathbf{w}_k\right\|_2^2 \le 1$ used to make the computation stable. To avoid the NP-hard problem, ℓ_1-norm is used to represent the sparse constraint. Since L_0-norm sparsity constraint makes LC-KSVD to find the optimal sparse solution unless using the good initialized values. Compared with L_0-norm, ℓ_1-norm sparse representation has shown the better performance and is helpful to find the optimal sparse solution more easily (Shao et al. 2020). In this study, we adopt the common used regularization term for function $f()$, i.e.

$$f(\mathbf{D}_s, \mathbf{D}_t) = \|\mathbf{D}_s - \mathbf{D}_t\|^2, \tag{10}$$

$$f(\mathbf{W}_s, \mathbf{W}_t) = \|\mathbf{W}_s - \mathbf{W}_t\|^2. \tag{11}$$

Thus, the objective function of TDDLLC can be represented as:

$$\begin{aligned}
&[\mathbf{D}_s, \mathbf{D}_t, \mathbf{S}_s, \mathbf{S}_t, \mathbf{W}_s, \mathbf{W}_t, \mathbf{A}_s, \mathbf{A}_t] \\
&= \min\ \|\mathbf{X}_s - \mathbf{D}_s\mathbf{S}_s\|_F^2 + \beta\|\mathbf{H}_s - \mathbf{W}_s\mathbf{S}_s\|_F^2 \\
&+ \gamma\|\mathbf{Q}_s - \mathbf{A}_s\mathbf{S}_s\|_F^2 + 2\varepsilon\|\mathbf{S}_s\|_{\ell_1} + \|\mathbf{X}_t - \mathbf{D}_t\mathbf{S}_t\|_F^2 \\
&+ \beta\|\mathbf{H}_t - \mathbf{W}_t\mathbf{S}_t\|_F^2 + \gamma\|\mathbf{Q}_t - \mathbf{A}_t\mathbf{S}_t\|_F^2 \\
&+ 2\varepsilon\|\mathbf{S}_t\|_{\ell_1}^2 + \mu_1\|\mathbf{D}_s - \mathbf{D}_s\|_F^2 + \mu_2\|\mathbf{W}_s - \mathbf{W}_t\|^2,
\end{aligned}$$

$$\begin{aligned}
s.t.\ &\forall k\|\mathbf{d}_{s,k}\|^2 \le 1, \|\mathbf{d}_{t,k}\|^2 \le 1, \\
&\forall k\|\mathbf{a}_{s,k}\|_2^2 \le 1, \|\mathbf{a}_{t,k}\|_2^2 \le 1, \\
&\forall k\|\mathbf{w}_{s,k}\|^2 \le 1, \|\mathbf{w}_{t,k}\|_2^2 \le 1,
\end{aligned} \tag{12}$$

where $\mathbf{d}_{s,k}(\mathbf{d}_{t,k})$,$\mathbf{a}_{s,k}(\mathbf{a}_{t,k})$ and $\mathbf{w}_{s,k}(\mathbf{w}_{t,k})$ are the kth atom of $\mathbf{D}_s(\mathbf{D}_t)$, $\mathbf{A}_s(\mathbf{A}_t)$ and $\mathbf{W}_s(\mathbf{W}_t)$, respectively.

3.2 Optimization of objective function

The optimization of TDDLLC is solved in this section. The optimization problem of Eq. (12) is four separately convex problems for each parameter, i.e., $\boldsymbol{S}$ (with $\boldsymbol{D}$, $\boldsymbol{W}$, $\boldsymbol{A}$ fixed), $\boldsymbol{D}$ (with $\boldsymbol{S}$, $\boldsymbol{W}$, $\boldsymbol{A}$ fixed), $\boldsymbol{W}$ (with $\boldsymbol{S}$, $\boldsymbol{D}$, $\boldsymbol{A}$ fixed) or $\boldsymbol{A}$ (with $\boldsymbol{S}$, $\boldsymbol{D}$, $\boldsymbol{W}$ fixed). We solve the TDDLLC by adopting the alternating optimization strategy.

By introducing an auxiliary linear equality constrained $\mathbf{Z}_s = \mathbf{S}_s$ and $\mathbf{Z}_t = \mathbf{S}_t$, we replace $\|\mathbf{S}_s\|_{\ell_1}$ to $\|\mathbf{Z}_s\|_{\ell_1}$ and $\|\mathbf{S}_t\|_{\ell_1}$ to $\|\mathbf{Z}_t\|_{\ell_1}$ in Eq. (12) and obtain the new objective function as follows:

$$\begin{aligned}
\left[\mathbf{D}_s, \mathbf{D}_t, \mathbf{S}_s, \mathbf{S}_t, \mathbf{W}_s, \mathbf{W}_t, \mathbf{A}_s, \mathbf{A}_t, \mathbf{Z}_s, \mathbf{Z}_t\right] = \min\ & \|\mathbf{X}_s - \mathbf{D}_s\mathbf{S}_s\|_F^2 + \beta\|\mathbf{H}_s - \mathbf{W}_s\mathbf{S}_s\|_F^2 + \gamma\|\mathbf{Q}_s - \mathbf{A}_s\mathbf{S}_s\|_F^2 \\
+2\varepsilon_s\|\mathbf{Z}_s\|_{\ell_1} + 2\mathbf{L}_s^T(\mathbf{S}_s - \mathbf{Z}_s) + \rho_s\|\mathbf{S}_s - \mathbf{Z}_s\|_F^2 &+ \|\mathbf{X}_t - \mathbf{D}_t\mathbf{S}_t\|_F^2 + \beta\|\mathbf{H}_t - \mathbf{W}_t\mathbf{S}_t\|_F^2 + \gamma\|\mathbf{Q}_t - \mathbf{A}_t\mathbf{S}_t\|_F^2 \\
+2\varepsilon_t\|\mathbf{Z}_t\|_{\ell_1}^2 + 2\mathbf{L}_t^T(\mathbf{S}_t - \mathbf{Z}_t) + \rho_t\|\mathbf{S}_t - \mathbf{Z}_t\|_F^2 &+ \mu_1\|\mathbf{D}_t - \mathbf{D}_s\|_F^2 + \mu_2\|\mathbf{W}_s - \mathbf{W}_t\|^2,
\end{aligned}$$

$$\begin{aligned}
s.t.\ & \forall k\|\mathbf{d}_{s,k}\|^2 \le 1, \|\mathbf{d}_{t,k}\|^2 \le 1, \\
& \forall k\|\mathbf{a}_{s,k}\|_2^2 \le 1, \|\mathbf{a}_{t,k}\|_2^2 \le 1, \\
& \forall k\|\mathbf{w}_{s,k}\|^2 \le 1, \|\mathbf{w}_{t,k}\|_2^2 \le 1,
\end{aligned} \tag{13}$$

where $\mathbf{L}_s = [\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_n] \in \mathbb{R}^{K\times n}$ and $\mathbf{L}_t = [\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_m] \in \mathbb{R}^{K\times m}$ are the augmented Lagrange multipliers. ρ_s and ρ_t are the penalty parameters.

1. Update step for $\mathbf{S}_s$ and $\mathbf{S}_t$. With $\mathbf{D}_s, \mathbf{D}_t, \mathbf{S}_t, \mathbf{W}_s, \mathbf{W}_t, \mathbf{A}_s, \mathbf{A}_t$ fixed, Eq. (13) can be reduced to:

$$[\mathbf{S}_s, \mathbf{Z}_s, \mathbf{L}_s] = \|\mathbf{X}_s - \mathbf{D}_s\mathbf{S}_s\|_F^2 + \beta\|\mathbf{H}_s - \mathbf{W}_s\mathbf{S}_s\|_F^2 + \gamma\|\mathbf{Q}_s - \mathbf{A}_s\mathbf{S}_s\|_F^2 + 2\varepsilon_s\|\mathbf{Z}_s\|_{\ell_1} + 2\mathbf{L}_s^T(\mathbf{S}_s - \mathbf{Z}_s) + \rho_s\|\mathbf{S}_s - \mathbf{Z}_s\|_F^2. \tag{14}$$

We can obtain the closed form solution of $\mathbf{S}_s^*$ as:

$$\mathbf{S}_s^* = (\mathbf{D}_s^T\mathbf{D}_s + \beta\mathbf{W}_s^T\mathbf{W}_s + \gamma\mathbf{A}_s^T\mathbf{A}_s + \rho_s\mathbf{I}_n)^{-1}(\mathbf{D}_s^T\mathbf{X}_s + \beta\mathbf{W}_s^T\mathbf{H}_s + \gamma\mathbf{A}_s^T\mathbf{Q}_s + \rho_s\mathbf{Z}_s - \mathbf{L}_s). \tag{15}$$

Then with $\mathbf{D}_s, \mathbf{W}_s, \mathbf{A}_s, \mathbf{S}_s$ fixed, we can obtain the closed form solution of $\mathbf{Z}_s$ as

$$\mathbf{Z}_s^* = \max\left\{\mathbf{S}_s^* + \frac{\mathbf{L}_s}{\rho_s} - \frac{\varepsilon_s}{\rho_s}\mathbf{I}_n, \mathbf{0}_n\right\} + \min\left\{\mathbf{S}_s^* + \frac{\mathbf{L}_s}{\rho_s} + \frac{\varepsilon_s}{\rho_s}\mathbf{I}_n, \mathbf{0}_n\right\}, \tag{16}$$

where $\mathbf{I}_n$ is the *1* matrix and $\mathbf{0}_n$ is the zero matrix.

Updating the Lagrange multiplier $\mathbf{L}_s$ as:

$$\mathbf{L}_s^* = \mathbf{L}_s + \theta(\mathbf{S}_s^* - \mathbf{Z}_s^*), \tag{17}$$

where parameter θ is the gradient of gradient descent method.

Using the similar method, we can update $\mathbf{S}_t^*$ as:

$$\begin{cases}
\mathbf{S}_t^* = (\mathbf{D}_t^T\mathbf{D}_t + \beta\mathbf{W}_t^T\mathbf{W}_t + \gamma\mathbf{A}_t^T\mathbf{A}_t + \rho_t\mathbf{I}_m)^{-1}(\mathbf{D}_t^T\mathbf{X}_t + \beta\mathbf{W}_t^T\mathbf{H}_t + \gamma\mathbf{A}_t^T\mathbf{Q}_t + \rho_t\mathbf{Z}_t - \mathbf{L}_t), \\
\mathbf{Z}_t^* = \max\left\{\mathbf{S}_t^* + \frac{\mathbf{L}_t}{\rho_t} - \frac{\varepsilon_t}{\rho_t}\mathbf{I}_m, \mathbf{0}_m\right\} + \min\left\{\mathbf{S}_t^* + \frac{\mathbf{L}_t}{\rho_t} + \frac{\varepsilon_t}{\rho_t}\mathbf{I}_m, \mathbf{0}_m\right\}, \\
\mathbf{L}_t^* = \mathbf{L}_t + \theta(\mathbf{S}_t^* - \mathbf{Z}_t^*).
\end{cases} \tag{18}$$

where $\mathbf{I}_m$ and $\mathbf{0}_m$ are the *1* and zero matrixes, respectively.

2. Update step for $\mathbf{D}_s$ and $\mathbf{D}_t$: with $\mathbf{S}_s, \mathbf{S}_t, \mathbf{W}_s, \mathbf{W}_t, \mathbf{A}_s, \mathbf{A}_t, \mathbf{Z}_s, \mathbf{Z}_t$ fixed, the objective function Eq. (13) is reduced to:

$$[\mathbf{D}_s, \mathbf{D}_t] = \|\mathbf{X}_s - \mathbf{D}_s\mathbf{S}_s\|_F^2 + \|\mathbf{X}_t - \mathbf{D}_t\mathbf{S}_t\|_F^2 + \mu_1\|\mathbf{D}_t - \mathbf{D}_s\|_F^2,$$

$$s.t.\ \forall k\|\mathbf{d}_{s,k}\|^2 \le 1,\ \|\mathbf{d}_{t,k}\|^2 \le 1. \tag{19}$$

The gradients of $\mathbf{D}_s$ and $\mathbf{D}_t$ are as follows:

$$\nabla_{\mathbf{D}_s} = -\mu(\mathbf{D}_t - \mathbf{D}_s) - (\mathbf{X}_s - \mathbf{D}_s\mathbf{S}_s)\mathbf{S}_s^T, \tag{20}$$

$$\nabla_{\mathbf{D}_t} = \mu(\mathbf{D}_t - \mathbf{D}_s) - (\mathbf{X}_t - \mathbf{D}_t\mathbf{S}_t)\mathbf{S}_t^T. \tag{21}$$

We can update each column of $\mathbf{D}_s$ by:

$$\mathbf{\Theta}_{s,k} = \mathbf{d}_{s,k} + \frac{1}{[\mathbf{S}_s\mathbf{S}_s^T]_{kk}}\left(-\mu(\mathbf{d}_{t,k} - \mathbf{d}_{s,k}) - [\mathbf{X}_s\mathbf{S}_s^T]_k + \mathbf{D}_s[\mathbf{S}_s\mathbf{S}_s^T]_k,\right. \tag{22}$$

$$\mathbf{d}_{s,k} = \frac{1}{\max(\|\mathbf{\Theta}_{s,k}\|_2, 1)}\mathbf{\Theta}_{s,k}. \tag{23}$$

And we can update each column of $\mathbf{D}_t$ by:

$$\mathbf{\Theta}_{t,k} = \mathbf{d}_{t,k} + \frac{1}{[\mathbf{S}_t\mathbf{S}_t^T]_{kk}}\left(\mu(\mathbf{d}_{t,k} - \mathbf{d}_{s,k}) - [\mathbf{X}_t\mathbf{S}_t^T]_k + \mathbf{D}_t[\mathbf{S}_t\mathbf{S}_t^T]_k\right), \tag{24}$$

$$\mathbf{d}_{t,k} = \frac{1}{\max(\|\mathbf{\Theta}_{t,k}\|_2, 1)}\mathbf{\Theta}_{t,k}. \tag{25}$$

3. Update step for $\mathbf{W}_s$ and $\mathbf{W}_t$: with $\mathbf{S}_s, \mathbf{S}_t, \mathbf{D}_s, \mathbf{D}_t, \mathbf{A}_s, \mathbf{A}_t, \mathbf{Z}_s, \mathbf{Z}_t$ fixed, the objective function Eq. (13) is reduced to:

$$[\mathbf{W}_s, \mathbf{W}_t] = \beta\|\mathbf{H}_s - \mathbf{W}_s\mathbf{S}_s\|_F^2 + \beta\|\mathbf{H}_t - \mathbf{W}_t\mathbf{S}_t\|_F^2 + \mu_2\|\mathbf{W}_s - \mathbf{W}_t\|^2,$$

$$s.t.\ \ \forall k\left\|\mathbf{w}_{s,k}\right\|^2 \leq 1, \left\|\mathbf{w}_{t,k}\right\|_2^2 \leq 1. \tag{26}$$

The gradient of $\mathbf{W}_s$ and $\mathbf{W}_t$ are as follows:

$$\nabla_{\mathbf{W}_s} = -\beta(\mathbf{H}_s - \mathbf{W}_s\mathbf{S}_s)\mathbf{S}_s^T + \mu_2(\mathbf{W}_s - \mathbf{W}_t), \tag{27}$$

$$\nabla_{\mathbf{W}_t} = -\beta(\mathbf{H}_t - \mathbf{W}_t\mathbf{S}_t)\mathbf{S}_t^T - \mu_2(\mathbf{W}_s - \mathbf{W}_t). \tag{28}$$

We can update each column of $\mathbf{W}_s$ by

$$\boldsymbol{\Delta}_{s,k} = \mathbf{w}_{s,k} + \frac{1}{[\mathbf{W}_s\mathbf{W}_s^T]_{kk}}(\mu_2(\mathbf{w}_{s,k} - \mathbf{w}_{t,k}) - \beta(\mathbf{H}_s - \mathbf{W}_s\mathbf{S}_s)\mathbf{S}_s^T), \tag{29}$$

$$\mathbf{w}_{s,k} = \frac{1}{\max(\left\|\boldsymbol{\Delta}_{s,k}\right\|_2, 1)}\boldsymbol{\Delta}_{s,k}. \tag{30}$$

And we can update each column of $\mathbf{W}_t$ by:

$$\boldsymbol{\Delta}_{t,k} = \mathbf{w}_{t,k} + \frac{1}{[\mathbf{W}_t\mathbf{W}_t^T]_{kk}}(\mu_2(\mathbf{w}_{t,k} - \mathbf{w}_{s,k}) - \beta(\mathbf{H}_t - \mathbf{W}_t\mathbf{S}_t)\mathbf{S}_t^T), \tag{31}$$

$$\mathbf{w}_{t,k} = \frac{1}{\max(\left\|\boldsymbol{\Delta}_{t,k}\right\|_2, 1)}\boldsymbol{\Delta}_{t,k}. \tag{32}$$

4. Update step for $\mathbf{A}_s$ and $\mathbf{A}_t$ with $\mathbf{S}_s, \mathbf{S}_t, \mathbf{D}_s, \mathbf{D}_t, \mathbf{W}_s, \mathbf{W}_t, \mathbf{Z}_s, \mathbf{Z}_t$ fixed, the objective function Eq. (13) is reduced to:

$$\begin{aligned} &\min\ \|\mathbf{Q}_s - \mathbf{A}_s\mathbf{S}_s\|_F^2 + \|\mathbf{Q}_t - \mathbf{A}_t\mathbf{S}_t\|_F^2, \\ &s.t.\ \forall k\left\|\mathbf{a}_{s,k}\right\|_2^2 \leq 1, \left\|\mathbf{a}_{t,k}\right\|_2^2 \leq 1, \end{aligned} \tag{33}$$

The Lagrange dual function can be used to solve this problem. We have

$$[\mathbf{A}_s] = \ \|\mathbf{Q}_s - \mathbf{A}_s\mathbf{S}_s\|_F^2 + \lambda_s\|\mathbf{A}_s\|_F^2, \tag{34}$$

where λ_s is the penalty parameter.

The gradient of $\mathbf{A}_s$ is as follows:

$$\nabla_{\mathbf{A}_s} = -(\mathbf{Q}_s - \mathbf{A}_s\mathbf{S}_s)\mathbf{S}_s^T + \lambda_s\mathbf{A}_s. \tag{35}$$

Then we can obtain the closed form solution of $\mathbf{A}_s$ as:

$$\mathbf{A}_s^* = \mathbf{Q}_s\mathbf{S}_s^T(\mathbf{S}_s\mathbf{S}_s^T + \lambda_s\mathbf{I})^{-1}. \tag{36}$$

Similarly, we obtain the closed form solution of $\mathbf{A}_t$ as:

$$\mathbf{A}_t^* = \mathbf{Q}_t\mathbf{S}_t^T(\mathbf{S}_t\mathbf{S}_t^T + \lambda_t\mathbf{I})^{-1}, \tag{37}$$

where λ_t is the penalty parameter.

For a testing signal $\boldsymbol{x}$', TDDLLC uses the obtained optimal $\mathbf{D}_t$ to compute its sparse representation $\mathbf{s}'$ by

$$\arg\min_{\mathbf{s}'}\left\|\mathbf{x}' - \mathbf{D}_t\mathbf{s}'\right\|_2^2 + \varepsilon\left\|\mathbf{s}'\right\|_{\ell_1},\ . \tag{38}$$

After obtaining its sparse representation vector $\mathbf{s}'$, the class label j of $\boldsymbol{x}$' is the index of the largest element in $\mathbf{Ws}'$, i.e.,

$$j = \arg\ \max_j\ (\mathbf{Ws}'). \tag{39}$$

3.3 The algorithm of TDDLLC

Based on the analysis above, Algorithm 1 shows the summary of the proposed TDDLLC.

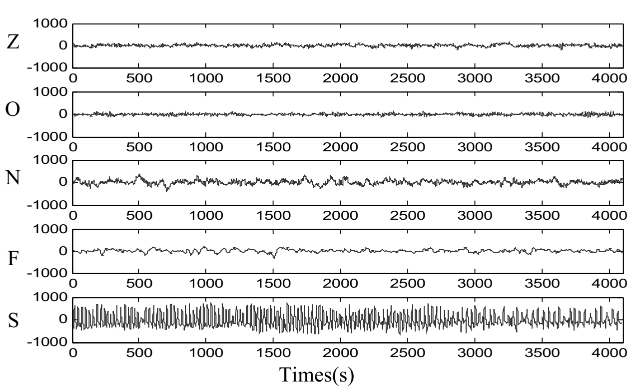

Fig. 1 Original EEG signals in groups Z, O, N, F and S

Table 1 Non-noise transfer tasks in the experiments

Tasks	SD (70 samples each in corresponding group)	TD (30 samples each in corresponding group)	Number of classes
T1	Z and S	O and S	2
T2	N and S	F and S	2
T3	{Z, N} and S	{O, F} and S	2
T4	{O, N} and S	{Z, F} and S	2
T5	{Z, F} and S	{O, N} and S	2
T6	Z, N and S	O, F and S	3
T7	Z, F and S	O, N and S	3
T8	O, N and S	Z, F and S	3
T9	O, F and S	Z, N and S	3
T10	{Z, O}, N and S	{Z, O}, F and S	3

Algorithm 1 TDDLLC

Input: Source domain $\boldsymbol{T}_s$ and target domain $\boldsymbol{T}_t$;

Output: The class label of test data.

1. Initialize parameters $\{\boldsymbol{S}_s=\boldsymbol{0}_{K\times n}, \boldsymbol{S}_t=\boldsymbol{0}_{K\times n}, \boldsymbol{L}_s=\boldsymbol{0}_{K\times n}, \boldsymbol{L}_t=\boldsymbol{0}_{K\times m}, \boldsymbol{D}_s=rand(d,K), \boldsymbol{D}_t=rand(d,K),$
$\boldsymbol{A}_s=rand(d,K),\ \boldsymbol{A}_t=rand(d,K), \boldsymbol{W}_s=rand(d,K),\ \boldsymbol{W}_t=rand(d,K)\}$;

For t=1,2,...,T

2. Update $\boldsymbol{S}_s$ and $\boldsymbol{S}_t$ by Eqs.(15)-(18);
3. Update $\boldsymbol{D}_s$ and $\boldsymbol{D}_t$ by Eqs.(22)-(25);
4. Update $\boldsymbol{W}_s$ and $\boldsymbol{W}_t$ by Eqs.(29)-(32);
5. Update $\boldsymbol{A}_s$ and $\boldsymbol{A}_t$ by Eqs.(36)-(37);
6. If converged, go to step 7; else go to step 2;
   t=t+1;

end

7. Obtain the optimal parameters $\{\boldsymbol{D}_s, \boldsymbol{D}_t, \boldsymbol{S}_s, \boldsymbol{S}_t, \boldsymbol{W}_s, \boldsymbol{W}_t, \boldsymbol{A}_s, \boldsymbol{A}_t, \boldsymbol{Z}_s, \boldsymbol{Z}_t\}$
8. Compute the sparse representation vector $\boldsymbol{s}'$ of test signal $\boldsymbol{x}'$ by Eq.(38);
   Predicted the class label of $\boldsymbol{x}'$ by Eq.(39).

4 Experiment

4.1 EEG dataset

The EEG data used for epileptic seizure classification in the study is provided by Bonn University (Andrzejak et al. 2001). The complete dataset consists of five different groups (called as Z, O, N, F and S). The total number of EEG signals is 500 (each set containing 100 single-channel segments). Each segment has a sampling rate of 173.6 Hz for 23.6 s, thus it contains 4096 samples. Figure 1 shows the original EEG signals in groups Z, O, N, F and S. The signals in groups Z are collected from the surface EEG records of five healthy volunteers who relax their eyes opened, and signals in groups O are collected from these volunteers with closed eyes under the wakefulness. The signals in groups N are collected from intermittent epileptic signals, which were derived from contralateral hemispheric hippocampal structures. The signals in groups F are intermittent epileptic signals derived within epileptogenic zone. The group S contains the epileptic activity in intracranial epileptogenic zone. Therefore, the dataset can be divided into three types: healthy signals, intermittent and seizures. The episodic EEG signals are characterized by abnormal epileptiform discharges, including spike slow wave and spike wave.

4.2 Experimental settings

We evaluate TDDLLC for transfer EEG epileptic seizure classification on two sets of tasks. Two groups of transfer EEG epileptic seizure classification tasks are constructed in the experiment. The first set is ten noise-free transfer tasks, and all involved transfer tasks have original signals with different distributions between TD and SD. The second set is five transfer tasks in the noise scenarios. We add white Gaussian noise to the signals in the TD or SD so that the domain discrepancy occurs in different domains. For each classification task, we construct the TD and SD by using the signals in corresponding groups. Then the training data is formed by using the all labeled signals in SD and a small number of labeled signals in TD, and the rest of signals in TD are used as test data. For fair comparison, we conduct each task for ten times to avoid the bias. In the experiments, we use three evaluation criteria sensitivity, specificity and accuracy (Gu et al. 2020; Wang et al. 2019a, b) for comparing the classification performance.

The performance of TDDLLC is compared with five classifiers, including two dictionary learning algorithms KSVD, LC-KSVD and three transfer learning algorithms LMPROJ (Quanz et al. 2009), DDTML (Ni et al. 2018a) and TSVM-GP (Ni et al. 2018b). For DDTML, a three layers of neural network is trained and its nodes are given as 200→200→100, and the parameters α, β, τ and λ are set as 10^{-1}, 10, 3 and 0.3, respectively. TSVM-GP and LMPROJ extract signal features by using the wavelet packet decomposition (WPD) method (Jiang et al. 2017a) and obtain the features of dimension 6. For TSVM-GP, λ and kernel parameter are determined in the grid $\{10^{-3}, 10^{-2},\ldots, 10^{4}\}$, and the group size is set to be 6. For LMPROJ, the eigenspectrum damping factor is determined in the grid {1.1, 1.5, 2.0} and trade-off parameter is set to be 10. For TSVM-GP, LMPROJ and TDDLLC, the regularization parameters are determined in the grid $\{10^{-3}, 10^{-2},\ldots, 10^{4}\}$. The number of dictionary atoms in K-SVD, LC-KSVD and TDDLLC are determined in the grid {80, 100, 120, 140, 160}. For TDDLLC, the parameters $\varepsilon_s = \varepsilon_t = \rho_s = \rho_t = 1$, β and γ are determined in the grid $\{10^{-2}, 10^{-1}, 1, 2, 4, 10\}$, μ_1 and μ_2 are determined in the grid $\{1, 10, 10^2, 10^3\}$. All the methods are implemented in MATLAB, and the environment that we used in the experiments is a computer with Intel Core i5-3317U 1.70 GHz CPU, 16 GB RAM.

4.3 Noiseless transfer task

In this subsection, we perform ten noiseless transfer tasks, and the information of these ten tasks is shown in Table 1. For example, the task T1 is a binary classification problem, which contains the 70 signals each in groups Z and S denoted as SD and 30 signals from {O, S} denoted as TD. The task T6 is a multi-class classification problem, which

Table 2 The average specificity of the six algorithms on ten noiseless transfer tasks

	KSVD	LC-KSVD	LMPROJ	TSVM-GP	DDTML	TDDLLC
T1	0.9165	0.9279	0.9611	0.9685	0.9740	**0.9828**
T2	0.9135	0.9249	0.9539	0.9600	0.9652	**0.9746**
T3	0.9304	0.9528	0.9672	0.9747	0.9776	**0.9814**
T4	0.9391	0.9540	0.9677	0.9749	0.9782	**0.9837**
T5	0.9264	0.9321	0.9387	0.9578	0.9687	**0.9770**
T6	0.9210	0.9418	0.9578	0.9600	0.9638	**0.9748**
T7	0.9190	0.9353	0.9554	0.9642	0.9674	**0.9714**
T8	0.9181	0.9264	0.9481	0.9557	0.9587	**0.9647**
T9	0.9056	0.9258	0.9436	0.9493	0.9560	**0.9618**
T10	0.9105	0.9252	0.9428	0.9443	0.9509	**0.9662**

Bold values represent the best classification performances in the tasks

Table 3 The average sensitivity of the six algorithms on ten noiseless transfer tasks

	KSVD	LC-KSVD	LMPROJ	TSVM-GP	DDTML	TDDLLC
T1	0.9136	0.9307	0.9488	0.9644	0.9697	**0.9806**
T2	0.9119	0.9264	0.9305	0.9581	0.9597	**0.9733**
T3	0.9320	0.9547	0.9612	0.9686	0.9743	**0.9798**
T4	0.9366	0.9521	0.9641	0.9732	0.9758	**0.9821**
T5	0.9212	0.9282	0.9339	0.9546	0.9680	**0.9767**
T6	0.9144	0.9407	0.9508	0.9465	0.9602	**0.9702**
T7	0.9194	0.9318	0.9517	0.9603	0.9629	**0.9713**
T8	0.9126	0.9186	0.9468	0.9525	0.9553	**0.9643**
T9	0.9053	0.9208	0.9411	0.9469	0.9553	**0.9611**
T10	0.8991	0.9162	0.9412	0.9408	0.9491	**0.9656**

Bold values represent the best classification performances in the tasks

Table 4 The average accuracy of the six algorithms on ten noiseless transfer tasks

	KSVD	LC-KSVD	LMPROJ	TSVM-GP	DDTML	TDDLLC
T1	0.9180	0.9295	0.9524	0.9648	0.9744	**0.9814**
T2	0.9128	0.9259	0.9421	0.9567	0.9644	**0.9741**
T3	0.9303	0.9559	0.9627	0.9689	0.9734	**0.9805**
T4	0.9377	0.9513	0.9640	0.9760	0.9761	**0.9823**
T5	0.9226	0.9300	0.9352	0.9581	0.9675	**0.9763**
T6	0.9138	0.9407	0.9513	0.9549	0.9624	**0.9731**
T7	0.9192	0.9312	0.9574	0.9624	0.9688	**0.9713**
T8	0.9159	0.9206	0.9476	0.9525	0.9562	**0.9651**
T9	0.9089	0.9245	0.9412	0.9494	0.9576	**0.9617**
T10	0.9054	0.9207	0.9432	0.9401	0.9504	**0.9681**

Bold values represent the best classification performances in the tasks contains the 140 signals from Z, N and S, which are denoted as SD and 100 signals from O, F and S, which are denoted as TD. The training data consists of all samples in SD and random 10 samples in TD and the remaining samples in TD are used for testing. Thus, 80% of the entire data is used as training samples and 20% data is used as testing samples. In binary classification, the seizure signals are denoted as positive samples, and the non-seizure signals are denoted as negative samples. For multi-class classification, the true positive (negative) is the average value of true positive (negative) for each class. We record ten classification results and shown their average sensitivity, specificity and accuracy in Tables 2, 3, 4, respectively.

(1) In general, four transfer learning algorithms obtain better sensitivity, specificity and accuracy performance than those of KSVD and LC-KSVD. Specially, KSVD, LC-KSVD and our algorithm TDDLLC are end-to end classifiers in the experiment. It indicates that when the domain discrepancy exists between training and test samples, transfer learning is a helpful strategy to improve the performance.
(2) The average specificity of all algorithms is better than average sensitivity. The reason is that the group S exists both in SD and TD, thus these signals are domain consistent in classification tasks. In converse, the signals in other groups exist either in SD or in TD. The results indicate that the domain discrepancy will severely disrupt the EEG epileptic seizure classification.
(3) We can see from the results in Tables 2, 3, 4 that the proposed TDDLLC is superior to all comparison algorithms in all transfer tasks. The results show that the learned dictionary of TDDLLC using auxiliary samples across domains is discriminative. Although a small number of target task samples can not provide enough information to train a classifier, its useful information can efficiently promote the performance of TDDLLC. In addition, TDDLLC does not need feature exploiting, thus it can enable the model to learn more discriminative information for dictionary. For DDTML, TSVM-GP and LMPROJ, the classification performance is heavily dependent on the appropriately extracted features. Thus, TDDLLC has stronger ability of transfer knowledge across domains.

4.4 Transfer tasks in noise scenario

In this subsection, we perform transfer tasks in noisy scenarios to further evaluate the proposed TDDLLC. Details of these tasks are shown in Table 5. For each task, we choose the fixed number of signals from corresponding group and add zero mean white Gaussian noise to the signals in the TD (or SD). Thus the domain discrepancy occurs between SD and TD due to the presence of noise. For example, the training data in the task S1 contains the 70 signals from {Z, S} with SNR {− 10, −4, 0, 10}dB white Gaussian noise denoted as SD, and clean 10 signals from {O, S} denoted as TD. The rest of 20 signals in the TD are used for testing. SNR is defined as:

$$\mathrm{SNR(dB)} = 10log_{10}(\mathrm{signal}/\mathrm{noise\ power})$$

where the noise power can be obtained by varying the variance of Gaussian distribution.

(1) The accuracy results of all algorithms are shown in Fig. 2. From the experimental results, we can clearly see that the increasing Gaussian noise can decrease the classification of each algorithm in each task. But in contrast to five comparisons algorithms, the result of TDDLLC goes down slower and delivers higher accuracy performance than others. This again demonstrates that the cross domain dictionaries learned with label consistency is more discriminative. In addition, the ℓ_1 -norm regularization term in TDDLLC and block-coor-

Table 5 Transfer tasks in noisy scenario in the experiments

Noisy scenarios	SD (70 samples each in corresponding group)	TD (30 samples each in corresponding group)	Number of classes
S1	Z and S + white Gaussian noise of SNR {− 10, −4, 0, 10}, respectively	Z and S	2
S2	F and S	N and S + white Gaussian noise of SNR {− 10, −4, 0, 10}, respectively	2
S3	Z, N and S + white Gaussian noise of SNR {− 10, −4, 0, 10}, respectively	Z, N and S	3
S4	Z, F and S	O, N and S + white Gaussian noise of SNR {− 10, −4, 0, 10}, respectively	3

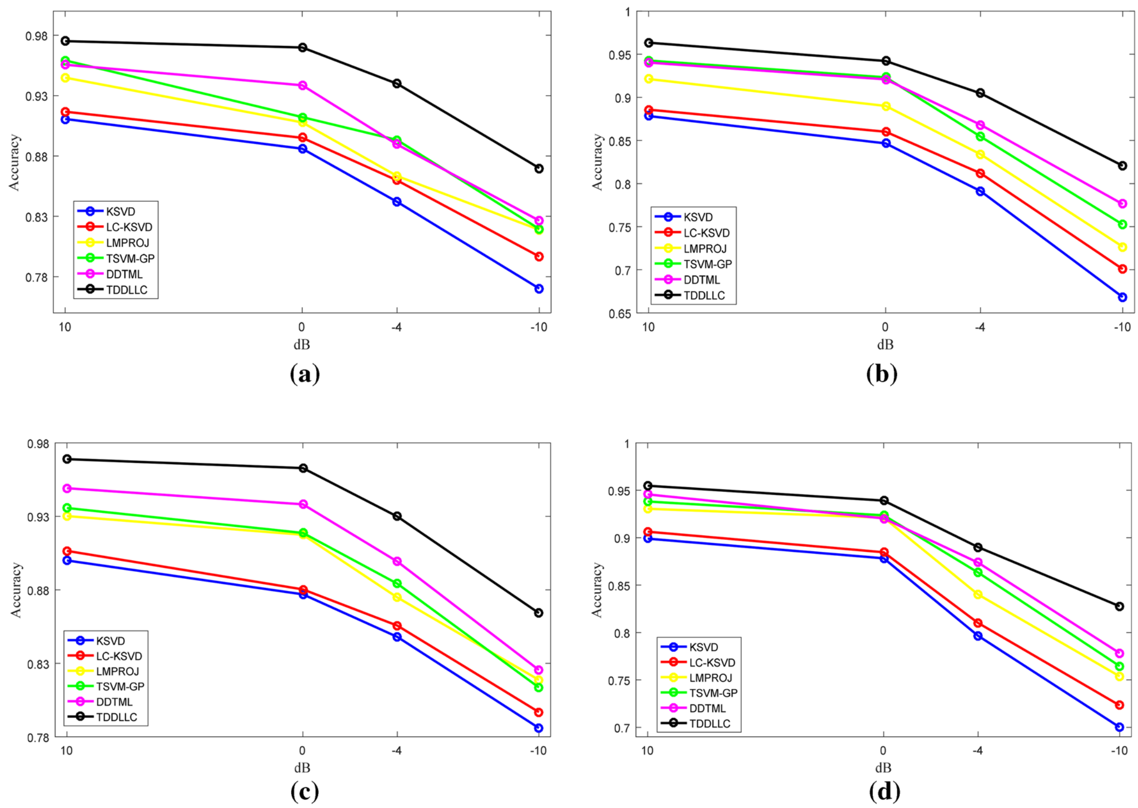

Fig. 2 The average accuracy of the six algorithms on noisy transfer tasks, (**a**) S1, (**b**) S2, (**c**) S3, (**d**) S4

Table 6 Rankings of the Friedman test

Algorithms	Ranking
KSVD	6
LC-KSVD	5
LMPROJ	3.8846
TSVM-GP	2.8846
DDTML	2.2308
TDDLLC	1

Table 7 Post Hoc comparison test for $\alpha = 0.05$

i	Algorithms	z	p	Holm	Hypothesis
5	KSVD	9.636241	0	0.01	Rejected
4	LC-KSVD	7.708993	0	0.0125	Rejected
3	LMPROJ	5.55937	0	0.016667	Rejected
2	TSVM-GP	3.632122	0.000281	0.025	Rejected
1	DDTML	2.371998	0.017692	0.05	Rejected

dinate descent (BCD) method ensures the optimization of the solution and the precision of optimum solution.

(2) When the noise power increases, the accuracy performance difference between TDDLLC and other three transfer learning algorithms increases. For example, in S1, the average accuracy difference between TDDLLC and LC-KSVD is 5.85% in the case of 10 dB. However, in the case of − 10 dB, the difference is 7.32%. This indicates that the TDDLLC algorithm is more robust in the noise scenario.

4.5 Statistical analysis

In this subsection, we evaluate the proposed TDDLLC via statistical analysis. The nonparametric Friedman test is adopted to evaluate whether the difference in classification accuracy is significant among all algorithms in the experiment. In the test, a rank is computed for each algorithm, and the algorithm with the highest rank is view as the best one. The result of the Friedman test is shown in Table 6. It is known that the lower the ranking in Friedman test, the better the algorithm. We can easily see from the Table 6 that TDDLLC has the best classification accuracy in 26 transfer EEG epileptic seizure classification tasks, including 10 non-noise transfer tasks and 16 transfer tasks in noisy scenario. Then the post-doc test is used to further analyze the difference between the best algorithm and the others. The significance level α is set to 0.05. The results of post-hoc test are shown in Table 7. If the p-value is smaller than the Holm value, the null hypothesis is rejected. From the results in Table 7, we can see that TDDLLC is superior to all the other algorithms. Therefore, we conclude that TDDLLC can enhance the transfer learning ability for transfer EEG epileptic seizure classification.

5 Conclusion

We propose a transfer discriminative dictionary learning with label consistency TDDLLC for cross-domain epilepsy EEG signal classification. To handle with highly non-stationary EEG signals, TDDLLC does not need feature extraction process which is common in the existing algorithms. Considering the independence of the domain in some dictionary atoms and utilizing the label consistency information in SD and minimizing the discrepancy between dictionaries of SD and TD, we learn the domain-specific dictionary for two domains, so as to transfer the sparse codes from SD to TD. In addition, we use ℓ_1-norm regularization term to find optimal solution easily. Although TDDLLC has demonstrated its distinctive effectiveness in cross-domain scenarios, there are some potential areas of future works. For example, how to speed up the training time and select the optimal parameters is still an open problem. In addition, we will apply the proposed algorithm to other EEG signal and health-related classification tasks. We will also study the transfer scenario where no labeled samples are available in TD.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China under Grant 61806026, by the Natural Science Foundation of Jiangsu Province under Grant BK 20180956, by the 2018 Six Talent Peaks Project of Jiangsu Province under Grant XYDXX-127.

Compliance with ethical standards

Conflict of interest The author declares no conflict of interest.

References

Abolghasemi V, Ferdowsi S (2015) EEG-fMRI: dictionary learning for removal of ballistocardiogram artifact from EEG. Biomed Signal Process Control 18(4):186–194

Akhavan A, Moradi MH (2018) Detection of concealed information using multichannel discriminative dictionary and spatial filter learning. IEEE Trans Inform Forensics Secur 13(10):2616–2627

Ameri R, Pouyan A, Abolghasemi V (2016) Projective dictionary pair learning for EEG signal classification in brain computer interface applications. Neurocomputing 218(12):382–389

Amin HU, Yusoff MZ, Ahmad RF (2020) A novel approach based on wavelet analysis and arithmetic coding for automated detection and diagnosis of epileptic seizure in EEG signals using machine learning techniques. Biomed Signal Process Control Biomed Signal Process Control 56(2):1–10

Andrzejak RG, Lehnertz K, Rieke C, Mormann F, David P, Elger CE (2001) Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. Phys Rev E 64:061907. https://doi.org/10.1103/PhysRevE.64.061907

Chen G, Xiong C, Corso J J (2012). Dictionary transfer for image denoising via domain adaptation, in Proceedings of 2012 19th IEEE International Conference on Image Processing ICIP 2012, Orlando, USA, 1189–1192, doi: https://doi.org/10.1109/ICIP.2012.6467078

Deng Z, Xu P, Xie L, Choi KS, Wang S (2018) Transductive joint-knowledge-transfer TSK FS for recognition of epileptic EEG signals. IEEE Trans Neural Syst Rehab Eng 26(8):1481–1494

Gu X, Chung F, Wang S (2017) Bayesian Takagi–Sugeno–Kang fuzzy classifier. IEEE Trans Fuzzy Syst 25(6):1655–1671

Gu X, Zhang C, Ni T (2020) A hierarchical discriminative sparse representation classifier for EEG signal detection. IEEE/ACM Trans Comput Biol Bioinform. https://doi.org/10.1109/TCBB.2020.2973978

Hu W, Cao J, Lai X, Liu J (2019) Mean amplitude spectrum based epileptic state classification for seizure prediction using convolutional neural networks. J Ambient Intell Hum Comput. https://doi.org/10.1007/s12652-019-01220-6

Jiang Z, Lin Z, Davis LS (2013) Label consistent K-SVD: learning a discriminative dictionary for recognition. IEEE Trans Pattern Anal Mach Intell 35(11):2651–2664

Jiang Y, Deng Z, Chung F, Wang G, Qian P, Choi KS, Wang S (2017) Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system. IEEE Trans Fuzzy Syst 25(1):3–20

Jiang Y, Wu D, Deng Z, Qian P, Wang J, Wang G, Chung F, Sze CK, Wang S (2017) Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system. IEEE Trans Neural Syst Rehab Eng 25(12):2270–2284

Jiang Z, Chung F, Wang S (2019) Recognition of multiclass epileptic EEG signals based on knowledge and label space inductive transfer. IEEE Trans Neural Syst Rehab Eng 27(4):630–642

Kevric J, Subasi A (2017) Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system. Biomed Signal Process Control 31(1):398–406

Kundishora AJ, Gummadavelli A, Ma C, Liu M, Mccafferty C, Schiff ND, Willie JT, Gross RE, Gerrard J, Hal B (2017) Restoring conscious arousal during focal limbic seizures with deep brain stimulation. Cereb Cortex 27(3):1964–1975

Liu B D, Wang Y X, Shen B, Zhang Y J, Wang Y J (2014). Blockwise coordinate descent schemes for sparse representation, in Proceedings of 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 5267–5271. doi:https://doi.org/10.1109/ICASSP.2014.6854608

Mendoza PF, Menezes ML, Santanna A, Ortiz BM, Samara A, Galway L (2019) Affective recognition from EEG signals: an integrated data-mining approach. J. Ambient Intell. Hum. Comput. 10:3955–3974

Ni TG, Gu XQ, Wang HY, Zhang Z, Chen SB, Jin C (2018) Discriminative deep transfer metric learning for cross-scenario person reidentification. J Electron Imaging 27(4):043026. https://doi.org/10.1117/1.JEI.27.4.043026

Ni TG, Gu XQ, Wang J, Zheng YH, Wang HY (2018) Scalable transfer support vector machine with group probabilities. Neurocomputing 273(17):570–582

Pan SJ, Yang Q (2010) A survey on transfer learning. IEEE Trans Knowl Data Eng 22(10):1345–1359

Quanz B, Huan, J (2009) Large margin transductive transfer learning, in Proceedings of the 18th ACM conference on Information and knowledge management, Hong Kong, China: 1327–1336. doi: https://doi.org/10.1145/1645953.1646121

Seo J, Laine TH, Sohn K (2019) Machine learning approaches for boredom classification using EEG. J Ambient Intell Hum Comput 10:3831–3846

Shao S, Xu R, Liu W, Di LB, Wang YJ (2020) Label embedded dictionary learning for image classification. Neurocomputing 385(14):122–131

Sreej SR, Samanta D (2019) Classification of multiclass motor imagery EEG signal using sparsity approach. Neurocomputing 368(11):133–145

Sreej SR, Samanta D, Sarma M (2019a) Weighted sparse representation for classification of motor imagery EEG signals, in Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany: 23–27. doi: https://doi.org/10.1109/EMBC.2019.8857496

Sridhar KP, Baskar S, Shakeel PM, Dhulipala VR (2019) Developing brain abnormality recognize system using multi-objective pattern producing neural network. J Ambient Intell Hum Comput 10:3287–3295

Wang SH, Zhang Y, Li YJ, Jia WJ, Liu FY, Yang MM, Zhang YD (2018) Single slice based detection for Alzheimer's disease via wavelet entropy and multilayer perceptron trained by biogeography-based optimization. Multimed Tools Appl 77(5):10393–10417

Wang SH, Zhang YD, Yang M, Liu B, Ramirez JG, Juan M (2019) Unilateral sensorineural hearing loss identification based on double-density dual-tree complex wavelet transform and multinomial logistic regression. Integr Comput-Aided Eng 26(4):411–426

Wang SH, Xie SP, Guttery DS, Tang CS, Sun JD (2019) Zhang Y D (2019b) Alcoholism identification based on an AlexNet transfer learning model. Front Psychiatry 10:205. https://doi.org/10.3389/fpsyt.2019.00205

Xia K, Ni T, Yin H, Chen B (2020) Cross-domain classification model with knowledge utilization maximization for recognition of epileptic EEG signals. IEEE/ACM Trans Comput Biol Bioinform. https://doi.org/10.1109/TCBB.2020.2973978

Yang C, Deng Z, Choi KS, Jiang Y, Wang S (2014) Transductive domain adaptive learning for epileptic electroencephalogram recognition. Artif Intell Med 62(3):165–177

Zhang Q, Li B (2010) Discriminative K-SVD for dictionary learning in face recognition, in Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, USA: 2691–2698. Doi: https://doi.org/10.1109/CVPR.2010.5539989

Zhang YD, Wang SH, Sui YX, Yang M, Liu B, Cheng H, Sun JD, Jia WJ, Phillips P, Gorriz JM (2018) Multivariate approach for Alzheimer's disease detection using stationary wavelet entropy and predator-prey particle swarm optimization. J. Alzheimer's Dis 65(3):855–869

Zhang Z, Jiang W, Qin J, Zhang L, Yan S (2018) Jointly learning structured analysis discriminative dictionary and analysis multiclass classifier. IEEE Trans Neural Netw Learn Syst 29(8):3798–3814

Zhang YD, Govindaraj VV, Tang CS, Zhu WG, Sun JD (2019) High performance multiple sclerosis classification by data augmentation and AlexNet transfer learning model. J Med Imaging Health Inform 9(9):2012–2021

Zheng J, Jiang Z, Chellappa R (2016) Cross-view action recognition via transferable dictionary learning. IEEE Trans Image Process 25(6):2542–2556

Zhu F, Shao L, Fang Y (2016) Boosted cross-domain dictionary learning for visual categorization. IEEE Intell Syst 31(3):6–18