Applied Acoustics 197 (2022) 108919

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Transferable discriminant linear regression for cross-corpus speech emotion recognition

Shaokai Li, Peng Song *, Wenjing Zhang

School of Computer and Control Engineering, Yantai University, Yantai 264005, China

ARTICLE INFO

Article history:
Received 14 February 2022
Received in revised form 24 May 2022
Accepted 2 July 2022

Keywords:
Linear regression
Speech emotion recognition
Category space
Transfer learning

ABSTRACT

Speech emotion recognition (SER) has attracted much interest recently due to its wide applications. However, it should be noted that most SER methods are conducted on the assumption that the training and testing data are from the same database. In real applications, this assumption does not hold, and the recognition performance will be significantly degraded. To solve this problem, we present a novel transferable discriminant linear regression (TDLR) approach for cross-corpus SER. Specifically, first, we introduce a non-negative label relaxation linear regression on source corpus to help learn transferable feature representations. Second, we propose a simple but effective strategy to keep the linear relationship between the labels of source and target corpora. Meanwhile, we utilize the discriminative maximum mean discrepancy (MMD) as the distance metric between two databases. Furthermore, we use the graph Laplacian to preserve the geometric structure of samples, which can further reduce the distribution gap between the two databases. Additionally, to better obtain the intrinsic properties of data and make the model robust, we impose an $\ell_{2,1}$-norm on the transformation matrices. Extensive experiments have been carried out on several standard databases, and the results show that TDLR can obtain better recognition performance than several state-of-the-art algorithms.

1. Introduction

As an important research direction in speech signal processing, speech emotion recognition (SER) has aroused extensive attention due to its potential applications, e.g., diagnostic tools for the therapist, machine interactions, and board systems in smart cars [1,2]. The goal of SER is to identify the emotion categories from speech signals, such as fear, anger, sadness, embarrassment, pleasure, and so on [3,4].

With the rapid development of pattern recognition techniques, many classification methods are used for SER, e.g., support vector machine (SVM), Gaussian mixture model (GMM), deep neural network (DNN), multi-task learning, and regression based algorithms [5–9]. These methods assume that the training and testing data follow the same probability distribution. However, in practical application scenarios, the probability distribution of data often changes along with the changing of timeline or scenes, resulting in a great discrepancy between the training corpus and the testing corpus, which would significantly influence the recognition performance. Thus, in this work, we consider the cross-corpus SER problem.

To address the problem mentioned above, it is necessary to rebuild the model [10]. However, for these traditional classification methods, this strategy might be time-consuming and undesirable in practice due to the large variety of utterances. Thus, how to deal with the cross-corpus SER problem is challengeable. Transfer learning might be a promising strategy, which can effectively transfer the knowledge learned from one domain to a new related domain [11]. Many transfer learning algorithms are developed to solve the cross-corpus recognition problems, in which the transferable feature representations [12–14] or classifiers [15] are learned. Therefore, in this work, we focus on investigating to develop a new transfer learning method to solve the cross-corpus SER problem.

Over the past decade, many efforts have been made to solve the challenging cross-corpus SER problem. For example, in [16], Hassan et al. propose an importance weighted SVM method to reduce the feature distribution mismatch between the training and testing samples, in which three transfer learning algorithms, i.e., Kullback-Leiber importance estimation procedure (KLIEP), kernel mean matching (KMM), and unconstrained least-squares importance fitting (uLSIF), are employed. In [17,18], Deng et al. present an adaptive auto-encoder-based networks method to learn a transferable feature representation for the cross-corpus SER problem. In [19,20], Zong et al. develop a domain-adaptive least square regression method for cross-corpus SER. More recently, for the

* Corresponding author.
E-mail address: pengsong@ytu.edu.cn (P. Song).

https://doi.org/10.1016/j.apacoust.2022.108919
0003-682X/

cross-corpus SER problem, Song et al. present a subspace learning-based transfer learning approach [21,22]. These algorithms can obtain competitive results in cross-corpus SER tasks. However, they do not fully take into account the label information for knowledge transfer. In other words, the category space of source and target domains should have certain compact correlations. In addition, they often assume that the learned labels are binary, which is too strict in practice. In this paper, inspired by recent progress in linear regression [23] and transfer learning [24], we present a novel transfer learning approach, called transferable discriminant linear regression (TDLR), which generalizes the traditional linear regression to a transfer learning manner. In TDLR, we follow the basic idea of transfer learning that transfers the knowledge from the source domain to the target domain. To this end, we utilize a dual distance metric, i.e., a class compactness graph and discriminative MMD, to measure the discrepancy between two fields. Meanwhile, we use the source data to train a linear regression model, which can well describe the relationship between features and labels. Furthermore, to make our model discriminative, we consider the information of category spaces, which can describe the relationships between the labels of source domain and the virtual labels of target domain. The flowchart of TDLR is illustrated in Fig. 1.

To sum up, several aspects of the main contributions of this work are listed as follows:

- We present a novel transfer learning framework for cross-corpus SER, which can extend the traditional regression to a transferable manner. Thorough experiments show that the proposed method achieves better performance than state-of-the-art transfer learning algorithms.
- We explore the category spaces of different fields, in which the linear relationship between source and target domains is discovered. Thus, the proposed method can preserve more discriminative power and obtain better transfer performance.
- We introduce a dual metric strategy to measure the distance between different fields, in which a discriminative MMD is used as the global metric and a graph Laplacian is used as the local metric. Thus, the distribution divergence between different fields can be efficiently alleviated.

The remainder of this paper is structured as follows. In Section 2, we review the related work about linear regression and transfer learning. Section 3 introduces the proposed TDLR and the optimizing algorithm, respectively. We give the experimental results and analysis in Section 4. Finally, we conclude our work in Section 5.

2. Related work

In this section, we review the related work, i.e., linear regression and transfer learning, in detail. Then, we discuss the main differences between the related work and our method.

2.1. Linear regression

Linear regression has been widely applied to classification problems [25–27]. The conventional linear regression model can be written as follows:

$$\min_{P} \|Y - XP\|_F^2 \tag{1}$$

where $X \in \mathbf{R}^{n\times d}$ is the feature matrix, $Y \in \mathbf{R}^{n\times c}$is the label matrix, and $P \in \mathbf{R}^{d\times c}$ is the projection matrix.

In practice, to make the model more robust, we often impose an $\ell_{2,1}$-norm or nuclear norm on the projection matrix. For example, in [28], Cai et al. have proved that the linear regression in linear discriminant analysis (LDA) subspace is equivalent to the low-rank regression model. In [26], Xiang et al. introduce an ε-dragging technique to force the distance between classes for least squares regression. Thus, the general discriminant linear regression can be written as

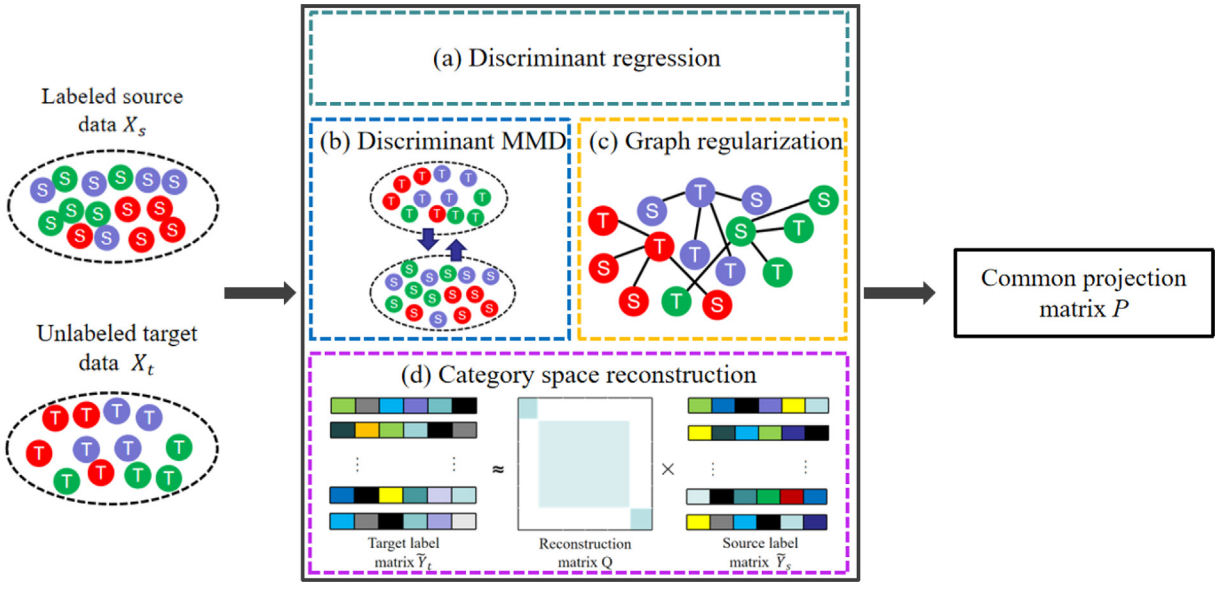

Fig. 1. The framework of TDLR. Different colors mean different emotion categories. Our framework consists of four main parts, i.e., discriminant regression, discriminant MMD, graph regularization, and category space reconstruction. By utilizing the distance metric, i.e., discriminant MMD and graph regularization, we can reduce the divergence between the source and target domains. Meanwhile, the source label matrix $\tilde{Y}_s$ and the target label matrix $\tilde{Y}_t$ keep a close linear relationship by using a reconstruction matrix Q. Finally, we can learn a common projection matrix P.

$$\min_{P,M} \|XP - (Y + B \odot M)\|_F^2 + \gamma \|P\|_{2,1} \quad \text{s.t.} M \geqslant 0 \tag{2}$$

where $\gamma$is a positive regularization parameter, $M \in \mathbf{R}^{n \times c}$is a non-negative label relaxation matrix, and $B \in \mathbf{R}^{n \times c}$ is a constant matrix, which can be calculated as

$$B_{ij} = \begin{cases} +1, & \text{if} Y_{ij} = 1 \\ -1, & \text{if} Y_{ij} = 0 \end{cases} \tag{3}$$

Over the past decades, many variants of linear regression have been presented for classification. For example, in [27], Zhang et al. present a retargeted least squares regression (ReLSR) method, which directly learns the regression targets from data. In [29], Lu et al. propose a robust manifold discriminant regression learning (RMDRL) approach for image classification. In [23], Fang et al. propose a robust label relaxation linear regression (RRLR) method. In [30], Zhang et al. develop an elastic-net regularized linear regression (ENLR) framework for image classification. Recently, in [31], Wen et al. have introduced the inter-class sparsity into least squares regression framework, which obtains good performance for multi-class classification problems. In [32], Han et al. present a double relaxed regression approach for image classification. However, all these methods assume that the training and testing data follow the same probability distributions, However, this assumption does not hold in real situations.

2.2. Transfer learning

For traditional pattern recognition algorithms, an ideal scenario is that the labeled training samples are abundant, and have the same feature distribution as the testing samples. Unfortunately, in real situations, collecting sufficient labeled training data is often expensive and even unrealistic, and unlabeled data is also difficult to collect. Inspired by the ability of humans to learn cross-domain knowledge transfer, the transfer learning techniques have been widely studied, which can use knowledge in the related field to learn knowledge in the target domain with fewer samples. A comprehensive overview of transfer learning algorithms can be referred to Refs. [10,24,33]. As shown in Table 1, transfer learning can be grouped into three types, i.e., unsupervised transfer learning [34], transductive transfer learning [35], and inductive transfer learning [36]. In cross-corpus SER tasks, the target corpus is unlabeled while the source corpus is labeled. Meanwhile, the domains are different but the tasks are the same. Thus, the proposed method falls into the category of transductive transfer learning.

In transfer learning, one of the most important problems is to find a suitable distance measurement to reduce the discrepancy between different fields. Over the past decades, as the most popular metric, MMD is widely used in many transfer learning algorithms [37], e.g., transfer component analysis (TCA) [38], joint distribution adaptation (JDA) [12], domain-invariant projection (DIP) and transfer independently together (TIT). Besides, many other measurements, e.g., Kullback–Leibler divergence [39], Jensen-Shannon divergence [40], and Bregman divergence [41], Wasserstein distance [42], have been employed to measure the distribution divergence. However, they might have the following limitations, a) they do not fully take into account the information of category spaces, which is important for discriminative feature transfer learning and facilitate the following classification tasks; b) they do not take the local geometric information into account, which has been proven very useful for transfer learning [22]; c) the feature transfer learning and label prediction are two independent processes. Thus, in this work, we take into account the category space, and present a novel transferable discriminant linear regression framework, which simultaneously conducts feature transfer learning and linear regression.

Table 1
Taxonomy of transfer learning approaches.

Categories	Source and target domains	Source and target tasks
Unsupervised transfer learning	Different but related	Different but related
Transductive transfer learning	Different but related	The same
Inductive transfer learning	The same	Different but related

3. The proposed method

3.1. Preliminary

We begin with a brief introduction of the symbols used in this article, which are shown in Table 2. For cross-corpus SER tasks, we have a labeled source database X_s and an unlabeled target database X_t. n_s and n_t are the numbers of samples of source and target databases, respectively. Let $X = [X_s, X_t] \in \mathbf{R}^{n \times d}$, where $n = n_s + n_t$, d is the dimension of features, $Y_s = \left[y_{s,1}, y_{s,2}, \ldots, y_{s,n_s}\right] \in \mathbf{R}^{n_s \times c}$ represents the source label vector, where c is the number of categories. We aim at learning a common projection matrix $P \in \mathbf{R}^{d \times c}$ and a label relaxation matrix $M \in \mathbf{R}^{n_s \times c}$ to build a regression model. Given a matrix W, $\|W\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{d}|W_{ij}|^2}$ denotes the Frobenius norm of W, $\|W\|_{2,1} = \sum_{i=1}^{d}\sqrt{\sum_{j=1}^{c}W_{ij}^2}$ denotes the $\ell_{2,1}$-norm of W, and $\mathrm{tr}(W) = \sum_i W_{ii}$ denotes the trace of W.

3.2. The framework of linear regression for transfer learning

One major goal of transfer learning is to reduce the distribution discrepancy by minimizing the appropriate distance measures. In this work, we utilize the discriminative MMD as the distance metric [12], which can alleviate the differences in marginal distribution and conditional distribution.

3.2.1. The discriminant regression

As shown in Eq. (2), we train a classifier in the common feature representation in the source database by minimizing the following regression problem:

$$\arg\min_{P,M} \|X_s P - (Y_s + B \odot M)\|_F^2 \quad \text{s.t.} M \geqslant 0 \tag{4}$$

where $\odot$ means the Hadamad product operator of matrices. Here we give an example below to show the relaxed binary label matrix. Let

Table 2
Notations used in this paper.

Notations	Descriptions
$X_s \in \mathbf{R}^{n_s \times d}$	Labeled source feature matrix
$X_t \in \mathbf{R}^{n_t \times d}$	Unlabeled target feature matrix
$P \in \mathbf{R}^{d \times c}$	Common projection matrix
$Y_s \in \mathbf{R}^{n_s \times c}$	Binary label matrix of source corpus
$\tilde{Y}_s \in \mathbf{R}^{n_s \times c}$	Slack label matrix of source corpus
$\tilde{Y}_t \in \mathbf{R}^{n_t \times c}$	Virtual label matrix of target corpus
$Q \in \mathbf{R}^{n_t \times n_s}$	Sparse reconstruction matrix
$B \in \mathbf{R}^{n_s \times c}$	Luxury matrix
$M \in \mathbf{R}^{n_s \times c}$	Non-negative label relaxation matrix
$M_0, M_c, M_d \in \mathbf{R}^{n \times n}$	MMD matrices
$L \in \mathbf{R}^{n \times n}$	Laplacian matrix

$$Y = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \in \mathbf{R}^{3\times 2} \tag{5}$$

be a binary label matrix, we use a slack variable matrix to replace Y, which is defined as

$$Y\prime = \begin{bmatrix} -m_{11} & 1+m_{12} \\ 1+m_{21} & -m_{22} \\ -m_{31} & 1+m_{32} \end{bmatrix}, \text{s.t.} m_{ij} \geqslant 0. \tag{6}$$

The constraint $m_{ij} \geqslant 0$ can enlarge the distance of samples from different categories.

3.2.2. The marginal distribution adaptation

We first calculate the distance between the means of samples from source and target databases, which can reduce the discrepancy between the marginal distribution of two databases. The objective function is written as follows:

$$\begin{aligned} MMD(X_s, X_t)_m &= \arg\min_P \| \frac{1}{n_s} \sum_{i=1}^{n_s} P^T x_i - \frac{1}{n_t} \sum_{j=n_s+1}^{n_s+n_t} P^T x_j \|^2 \\ &= \arg\min_P \operatorname{tr}\left(P^T X^T M_0 X P\right) \end{aligned} \tag{7}$$

where M_0 can be calculated by

$$(M_0)_{ij} = \begin{cases} \frac{1}{n_s n_s}, & \text{if} x_i, x_j \in D_s \\ \frac{1}{n_t n_t}, & \text{if} x_i, x_j \in D_t \\ \frac{-1}{n_s n_t}, & \text{otherwise} \end{cases} \tag{8}$$

3.2.3. The conditional distribution adaptation

We further introduce MMD to calculate the distance between the conditional distribution of source database $Q_s(x_s|y_s = c)$and that of target database $Q_t(x_t|y_t = c)$ as below:

$$\begin{aligned} MMD(X_s, X_t)_c &= \arg\min_P \| \frac{1}{n_s^{(c)}} \sum_{x_i \in D_s^{(c)}} P^T x_i - \frac{1}{n_t^{(c)}} \sum_{x_j \in D_t^{(c)}} P^T x_j \|^2 \\ &= \arg\min_P \operatorname{tr}\left(P^T X^T M_c X P\right) \end{aligned} \tag{9}$$

where $D_s^{(c)}$and $D_t^{(c)}$are the samples belonging to the c-th class, and $n_s^{(c)} = |D_s^{(c)}|, n_t^{(c)} = |D_t^{(c)}|$. We incorporate Eqs. (7) and (9) into a joint function as

$$\begin{aligned} MMD(X_s, X_t)_d &= MMD(X_s, X_t)_m + MMD(X_s, X_t)_c \\ &= \arg\min_P \sum_{i=0}^{c} \operatorname{tr}\left(P^T X^T M_i X P\right) \\ &= \arg\min_P \operatorname{tr}\left(P^T X^T M_d X P\right) \end{aligned} \tag{10}$$

The discriminative MMD matrix of M_d can be calculated as

$$(M_d)_{ij} = \begin{cases} \frac{1}{n_s^{(c)} n_s^{(c)}}, & \text{if} x_i, x_j \in D_s^{(c)} \\ \frac{1}{n_t^{(c)} n_t^{(c)}}, & \text{if} x_i, x_j \in D_t^{(c)} \\ \frac{-1}{n_s^{(c)} n_t^{(c)}}, & \text{if} \begin{cases} x_i \in D_s^{(c)}, x_j \in D_t^{(c)} \\ x_j \in D_s^{(c)}, x_i \in D_t^{(c)} \end{cases} \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

By minimizing Eq. (10), the marginal and conditional distributions between two databases are dragged close. Meanwhile, we can learn a common projection matrix P. Note that M_d requires the guidance of real labels, and the target database is unlabeled. Thus, in this work, we explore the pseudo-labels by applying the common projection matrix P on the unlabeled target data. For example, given $X_t P \in \mathbf{R}^{n_t \times c} = [l_1; l_2; \ldots; l_{n_t}]$, where $l_i = x_i P$, the pseudo-label of sample x_i can be obtained by the following equation:

$$h = \arg\max_j \{l_{ij}\} \tag{12}$$

where j is the number of elements in l_i, and the sample x_i is labeled as the h-th class.

To make our regression model be more discriminant, we constrain the projection matrix P by using an $\ell_{2,1}$-norm. Then, combining Eq. (10) and Eq. (4), we can obtain the general framework as follows:

$$\begin{aligned} &\arg\min_{P,M} \|X_s P - \widetilde{Y}_s\|_F^2 + \lambda \operatorname{tr}\left(P^T X^T M_d X P\right) + \gamma \|P\|_{2,1} \\ &\text{s.t.} M \geqslant 0 \end{aligned} \tag{13}$$

where $\widetilde{Y}_s = Y_s + B \odot M$, and λ is a positive balancing parameter.

3.3. The graph regularization

Note that the learning of common feature representation is carried out in the Euclidean space, which does not consider the inherent geometric information of data. According to the idea of manifold learning [43], the local geometric structure of data in high-dimensional space should also be maintained in low-dimensional subspace. Given a nearest neighbor graph, let $W = [w_{ij}] \in \mathbf{R}^{n\times n}$ be a weight matrix of the graph, for two data points x_i, x_j that are close to each other, if x_j is among the k nearest neighbors of $x_i, w_{ij} = 1$, otherwise, $w_{ij} = 0$. Considering the problem of mapping the graph to the low-dimensional feature representations, a good mapping can be defined as the following objective function [44]:

$$\begin{aligned} & \arg\min_z \sum_{i,j=1} \|z_i - z_j\|^2 W_{ij} \\ = \ & \arg\min_P \sum_{i=1}^{n} (x_i P)^T (x_i P) d_{ii} - \sum_{i,j=1}^{n} (x_i P)^T (x_i P) w_{ii} \\ = \ & \arg\min_P \operatorname{tr}\left(P^T X^T D X P\right) - \operatorname{tr}\left(P^T X^T W X P\right) \\ = \ & \arg\min_P \operatorname{tr}\left(P^T X^T L X P\right) \end{aligned} \tag{14}$$

where D is a diagonal matrix and $D_{ii} = \sum_j w_{ij}, L = D - W$ is the Laplacian matrix, and $z_i = x_i P$ denotes the low-dimensional representation of x_i. By minimizing Eq. (14), we expect that z_i and z_j are also close in the low-dimensional transformed space. Combining Eq. (13) and Eq. (14), we can obtain a discriminant regression with a dual metric regularization, which can be written as

$$\begin{aligned} &\arg\min_{P,M} \|X_s P - \widetilde{Y}_s\|_F^2 + \gamma \|P\|_{2,1} + \operatorname{tr}\left(P^T X^T (\lambda M_d + \alpha L) X P\right) \\ &\text{s.t.} M \geqslant 0 \end{aligned} \tag{15}$$

By minimizing the above equation, we can learn a common feature projection matrix P, which can reduce the probability distribution divergence across domains and transform the feature space into the category space. In addition, we can learn an optimal relaxation matrix M.

3.4. Linear reconstruction of category space

It is well-known that if the feature distributions of source and target databases are similar in the common subspace, there should be a close linear relationship between the cross-corpus data. Linear reconstruction is usually used to ensure the linear relationship between data in the common subspace to explore the similarity between domains [45,46]. However, we should not only explore

the correlations of features, but also consider the correlations of category space. The linear reconstruction of category space can effectively reflect the intrinsic information between the source and target category space. Thus, we can apply this linear reconstruction strategy to the source and target category space to explore their compact correlations.

To keep the linear relationship between the category spaces of source and target databases, we use a reconstruction matrix to linearly reconstruct the target category space from the source category space, which is expressed as

$$\begin{aligned} &\arg\min_{Q,M}\|\widetilde{Y}_t - Q\widetilde{Y}_s\|_F^2 + \mu_2\|Q^T\|_{2,1} \\ &\text{s.t.} M \geqslant 0 \end{aligned} \tag{16}$$

where $Q \in \mathbf{R}^{n_t \times n_s}$ is the sparse reconstruction matrix, μ_2 is the regularization parameter, and $\widetilde{Y_t}$ is the virtual label matrix of target database predicted by X_tP. We expect the columns of source category space be sparse by adding an $\ell_{2,1}$-norm constraint on Q. By minimizing Eq. (16), the target category space can be linearly represented by the source category space, and the compact correlations are preserved in the matrix Q.

Note that we cannot ensure the target virtual labels be close to the ground truth. To tackle this problem, we adopt an iterative solution to refine the virtual labels. Finally, combining Eqs. (15) and (16), we can get the objective function of TDLR as

$$\begin{aligned} &\arg\min_{P,M,Q}\|X_sP - \widetilde{Y}_s\|_F^2 + \mu_1\|\widetilde{Y}_t - Q\widetilde{Y}_s\|_F^2 + \mu_2\|Q^T\|_{2,1} \\ &\qquad +\text{tr}\left(P^TX^T(\lambda M_d + \alpha L)XP\right) + \gamma\|P\|_{2,1} \\ &\text{s.t.} M \geqslant 0 \end{aligned} \tag{17}$$

where μ_1 and λ are positive regularization parameters, and $\widetilde{Y}_t = X_tP, \widetilde{Y}_s = Y_s + B \odot M$.

3.5. Optimization

The objective function involves the $\ell_{2,1}$-norm, which is a non-smooth problem and difficult to be directly optimized [47]. Thus, we develop an iterative update algorithm to solve the objective function. We first rewrite Eq. (17) as

$$\begin{aligned} \mathcal{L} = &\|X_sP - \widetilde{Y}_s\|_F^2 + \mu_1\|X_tP - Q\widetilde{Y}_s\|_F^2 + \mu_2\|Q^T\|_{2,1} \\ &+\text{tr}\left(P^TX^T(\lambda M_d + \alpha L)XP\right) + \gamma\|P\|_{2,1} \end{aligned} \tag{18}$$

The detailed optimization procedures are given as below.

1) Update P**:** Fixing the other variables. Note that directly optimizing $\|P\|_{2,1}$ is too complicated, thus we compute its sub-gradient G [48], and its i-th diagonal element can be calculated as

$$G_{ii} = \begin{cases} 0, & \text{if} p^i = 0 \\ \frac{1}{2\|p^i\|}, & \text{otherwise} \end{cases} \tag{19}$$

where p^i is the i-th row of P.

Algorithm 1 The TDLR algorithm

Input: The labeled source data X_s and unlabeled target data X_t; the label matrix Y_s of source corpus; the regularization parameters α, λ, β, μ_1 and μ_2; and the number of nearest neighbors k.

Output: The common projection matrix P and the reconstruction matrix Q.

Initialize: $M = 1_{n_s \times c}$; $G = I$; $Q = I$; compute the MMD matrix M_d; construct the k nearest neighbor graph; and set $t = 0$.

repeat

1. Update P by solving Eq. (21);
2. Update M by solving Eq. (23);
3. Update Q by solving Eq. (26);
4. Update the pseudo-labels in target database by solving Eq. (12);
5. Update the discriminative MMD matrix M_d by solving Eq. (11);
6. Update the graph Laplacian L;
7. $t = t + 1$;

until Convergence criterion is satisfied.

We set the derivation of $\mathcal{L}$ with respect to P to zero, and can get the following equation:

$$\left(X_s^TX_s + \mu_1X_t^TX_t + X^T(\lambda M_d + \alpha L)X + \gamma G\right)P - X_s^T\widetilde{Y}_s - \mu_1X_t^TQ\widetilde{Y}_s = 0 \tag{20}$$

By solving the above equation, we can obtain

$$P = K^{-1}\left(X_s^T\widetilde{Y}_s + \mu_1X_t^TQ\widetilde{Y}_s\right) \tag{21}$$

where $K = X_s^TX_s + \mu_1X_t^TX_t + X^T(\lambda M_d + \alpha L)X + \gamma G$.

2) Update M**:** Update M by fixing the other variables. Let $X_sP - Y_s = E$, we can obtain

$$\begin{aligned} &\arg\min_M\|E - B \odot M\|_F^2 \\ &\text{s.t.} M \geqslant 0 \end{aligned} \tag{22}$$

According to [26], we can compute M by

$$M = \max(B \odot E, 0) \tag{23}$$

3) Update Q**:** Update Q by fixing the other variables. As step 1), we also compute the sub-gradient of $\|Q^T\|_{2,1}$ as Z, and its i-th diagonal element is calculated by

$$Z_{ii} = \begin{cases} 0, & \text{if} q^i = 0 \\ \frac{1}{2\|q^i\|}, & \text{otherwise} \end{cases} \tag{24}$$

where q^i is the i-th column of Q.

By setting the derivation of $\mathcal{L}$ with respect to Q to zero, we have

$$Q\left(\mu_1\widetilde{Y}_s\widetilde{Y}_s^T - \mu_2Z\right) - \mu_1X_tP\widetilde{Y}_s^T = 0 \tag{25}$$

By solving the above equation, we can obtain

$$Q = \mu_1X_tP\widetilde{Y}_s^T\left(\mu_1\widetilde{Y}_s\widetilde{Y}_s^T - \mu_2Z\right)^{-1} \tag{26}$$

The procedures of TDLR are described in detail in Algorithm 1. The convergence criterion for our algorithm is that the maximum iteration number reaches 50 or $\frac{|\Omega^{(t)} - \Omega^{(t-1)}|}{\Omega^{(t-1)}} < 0.01$, where $\Omega^{(t)}$ is the objective value of the t-th iteration.

3.6. Computational complexity analysis

In this subsection, we give the computational complexity of the proposed method. For computing P, according to Eq. (21), the computational complexity is $\mathcal{O}\left(d^2n_s + d^2n_t + dn^2 + d^3 + dn_sc + dn_sn_t + n_tn_sc\right)$. For computing M, according to Eq. (23), the computational complexity is $\mathcal{O}(n_sc)$. For computing Q, according to Eq. (26), the computational complexity is $\mathcal{O}(n_tdc + n_sdc + n_sn_tc + n_t^3 + n_s^2c)$. To sum up, the total computational complexity is $\mathcal{O}\left(T\left(d\left(n^2 + n_sn_t + n_sc + n_tc\right) + d^2(n_s + n_t) + d^3 + n_sn_tc + n_s^2c + n_t^3\right)\right)$, where T is the number of iterations.

4. Experiments

In this section, we conduct experiments to evaluate the efficiency of the proposed TDLR method and several state-of-the-art representative approaches on cross-corpus SER tasks. In Section 4.1, we describe the details of experiment setting, including databases, experimental tasks, baseline methods, evaluation metric, feature set, and parameter setting. In Section 4.2, we give the experimental results and discussions. In Section 4.3, we analyze the parameter sensitivity and effectiveness of our method. In Section 4.4, we give the convergence analysis of our method. Finally, Section 4.5 gives the analysis of the data visualization results.

4.1. Experiment setting

4.1.1. Databases

We elaborate the databases used in our experiments as below:

- EMO-DB (E) [49]: 10 amateur performers (five male and five female) are employed to simulate the emotions, i.e., neutral (NE), anger (AN), fear (FE), happiness (HA), sadness (SA), disgust (DI), and boredom (BO), totally 494 German utterances are collected.
- eNTERFACE'05 (e) [50]: 42 subjects (81% are men and 19% are women) from 14 different nationalities are invited to respond in English under six different situations, each of them is elicited as one of the following emotions: sadness (SA), anger (AN), surprise (SU), fear (FE), happiness (HA) and disgust (DI). The numbers of clips of these six emotions are 195, 200, 190, 187, 205, and 189, respectively.
- BAUM-1a (B) [51]: 31 subjects (18 male and 13 female) are employed to watch a stimuli video, and then express their feelings in Turkish. The age range of subjects is from 19 to 65. This database contains eight basic expressions, i.e., happiness (HA), anger (AN), sadness (SA), fear (FE), disgust (DI), boredom (BO), interest (IN), and unsure (UN). The numbers of clips of these expressions are 27, 43, 38, 36, 35, 27, 29, and 38, respectively.

The basic information of these databases is described in Table 3.

Table 3
The details of several databases.

Databases	EMO-DB	eNTERFACE'05	BAUM-1a
Language	German	English	Turkish
Size	494	1287	273
Classes	7	6	8
Modalities	audio	video	video
Dimension	1582	1582	1582

4.1.2. Experimental tasks

In this section, we design six groups of cross-corpus SER experiments, which are described as follows:

- e→E: The training corpus is eNTERFACE'05 while the testing corpus is EMO-DB.
- B→E: The training corpus is BAUM-1a while the testing corpus is EMO-DB.
- E→e: The training corpus is EMO-DB while the testing corpus is eNTERFACE'05.
- B→e: The training corpus is BAUM-1a while the testing corpus is eNTERFACE'05.
- E→B: The training corpus is EMO-DB while the testing corpus is BAUM-1a.
- e→B: The training corpus is eNTERFACE'05 while the testing corpus is BAUM-1a.

In the experiments, we select the common five categories from these three databases, including DI, AN, FE, SA, and HA, for evaluation. Each database is randomly divided into 10 parts, among which random seven parts of the target database and the entire source database are selected for training, and the rest three parts of the target database are used for testing. Our experiments are repeated 10 times to cover the possibilities of different cases and the average results are given.

4.1.3. Baseline methods and evaluation metric

We compare the proposed TDLR with several traditional and state-of-the-art transfer learning methods including the following:

- Principal component analysis (PCA) [52]: PCA is a classic unsupervised subspace learning method.
- Linear discriminant analysis (LDA) [53]: LDA is a classic supervised subspace learning method.
- Regularized label relaxation linear regression (RLR) [23]: RLR is a regression method, which introduces a relaxation variable matrix and the graph Laplacian term to avoid overfitting.
- Transfer component analysis (TCA) [38]: TCA is a transfer subspace learning method, which uses MMD to reduce the feature distribution divergence across two domains.
- Joint distribution adaptation (JDA) [12]: JDA is a transfer learning method, which uses MMD to reduce the marginal and conditional feature distribution divergence across two domains.
- Transfer joint matching (TJM) [54]: TJM is a transfer subspace learning method, which simultaneously performs feature matching and instance reweighting.
- Transfer linear discriminant analysis (TLDA) [22]: TLDA is a transfer subspace learning method, which extends LDA to a transfer learning manner by utilizing the MMD.
- Domain regeneration in the label space (DRLS) [55]: DRLS is a transfer linear regression method, which extends the linear regression to a transfer learning manner by utilizing the MMD.
- Joint transfer subspace learning and regression (JSTLR) [56]: JSTLR is a transfer learning method, which combines the linear regression and transfer subspace learning into a joint framework.

For evaluation, SVM is used as the baseline classifier for all compared algorithms except for JSTLR and DRLS. For the evaluation metric, we use the weighted average recall (WAR) as the recognition accuracy.

4.1.4. Feature set and parameter setting

In our experiments, we leverage the feature set of the INTERSPEECH 2010 paralinguistic challenge [57] as the emotional features, which is extracted by the openSMILE toolkit [58]. Totally

Table 4
The details of LLDs used in our experiments.

Descriptors	Number of features
MFCC [0–14]	630
LSP frequency [0–7]	336
Log mel freq band [0–7]	336
Voicing prob	42
Loudness	42
F0 envelope	42
F0	38
Shimmer	38
Jitter	38
Jitter consecutive frame pairs	38
F0 number of onesets	1
Turn duration	1

1582 dimensional low-level descriptors (LLDs) are used in our experiments. The detailed information of these LLDs is listed in Table 4.

Under the experimental setup, since the training set and the testing set follow different probability distributions, we cannot directly adopt the cross-validation [54] to determine the parameters. Thus we search the optimal parameter values in the parameter space to evaluate all the algorithms. Specifically, the MMD parameter λ is chosen by searching $\{1, 10, 10^2, 10^3\}$, the graph parameter α is set by searching $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10\}$, the regularization parameter γ is set by searching $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2\}$, and μ_1 and μ_2 are chosen from $\{10^{-3}, 10^{-2}, 10^{-1}, 1\}$. Additionally, we choose the number of nearest neighbors k by searching $3 \sim 8$.

For PCA, we reduce the feature dimension by preserving 98% energy. For the transfer learning methods, i.e., JTSLR and DRLS, we select the parameters from $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$. For TCA, TJM and JDA, we set their parameters by searching $\{10^{-2}, 10^{-1}, 1, 10\}$. For TLDA and RLR, we choose the parameters from $\{10^{-1}, 1, 10, 10^2, 10^3\}$. In addition, since the solutions of these methods are different, the subspace learning based methods and regression based methods will be reduced to different dimensions. For PCA, JDA, TCA and TJM, we reduce the dimension of features to 100, while for RLR, LDA, DRLS, TLDA and JTSLR, the dimension is reduced to 5. Note that, in our experiments, all the compared methods are conducted using the optimal parameters.

4.2. Results and discussions

In this section, we compare TDLR with the baseline methods mentioned in Section 4.1.3. Note that each task is repeated ten times and the average results are reported. The average recognition accuracy with standard deviation is given in Table 5. In particular, the golden standard represents the recognition results in single-corpus tasks. From the table, we have the observations as follows.

Firstly, we can find that TDLR obtains higher average recognition accuracy than the baseline methods. The average recognition accuracy reaches 45.08%, with an improvement of 0.69% compared to the best baseline JTSLR. Also, our TDLR algorithm achieves the highest recognition accuracy in three cases, i.e., B→E, E→e, and B→e. These experimental results verify that the TDLR algorithm can learn a good transferable linear regression model to solve the cross-corpus SER problem.

Secondly, we can find that the recognition accuracy of all the transfer learning methods are higher than that of the traditional methods. The reason might be that, these traditional methods assume the training and testing data follow the same probability distributions. However, for cross-corpus SER, this assumption does not hold. Thus, the transfer learning approaches that consider the knowledge transfer could obtain better recognition performance.

Table 5
Recognition accuracy (%) with standard deviation of different methods under six tasks.

Tasks	Golden standard	Traditional methods			Transfer learning methods						TDLR
		PCA	LDA	RLR	TCA	JDA	TJM	TLDA	DRLS	JTSLR	
e→E	79.61 ± 1.23	44.17 ± 0.69	38.76 ± 0.20	37.16 ± 0.53	36.03 ± 2.61	45.58 ± 1.72	45.82 ± 0.28	46.60 ± 0.76	48.00 ± 0.83	**50.98** ± 1.16	48.67 ± 1.07
B→E		35.29 ± 1.35	43.67 ± 0.62	38.93 ± 0.86	44.11 ± 0.83	45.58 ± 1.72	46.60 ± 0.14	42.06 ± 0.95	49.56 ± 1.62	48.04 ± 0.75	**52.21** ± 1.25
E→e	65.47 ± 1.16	35.71 ± 1.06	40.51 ± 0.23	40.07 ± 1.31	38.89 ± 1.11	41.56 ± 0.79	42.06 ± 0.39	37.56 ± 1.17	38.89 ± 0.94	42.33 ± 0.52	**46.42** ± 1.05
B→e		30.69 ± 1.49	30.49 ± 1.89	31.25 ± 1.72	31.02 ± 1.42	31.90 ± 1.03	33.72 ± 0.45	34.05 ± 0.82	34.29 ± 0.65	36.61 ± 2.10	**36.68** ± 0.73
E→B	67.30 ± 1.15	34.61 ± 1.35	35.21 ± 1.05	36.53 ± 1.41	33.58 ± 1.05	34.90 ± 0.90	37.28 ± 1.31	37.56 ± 1.31	44.25 ± 1.29	**46.15** ± 1.35	44.23 ± 1.03
e→B		27.92 ± 1.76	30.24 ± 2.04	32.69 ± 1.73	38.43 ± 1.42	38.57 ± 1.90	41.65 ± 0.75	38.57 ± 1.03	**42.31** ± 1.80	42.23 ± 2.60	42.30 ± 1.15
Average	70.79	34.73	36.48	36.10	37.01	39.68	40.31	41.23	42.88	44.39	**45.08**

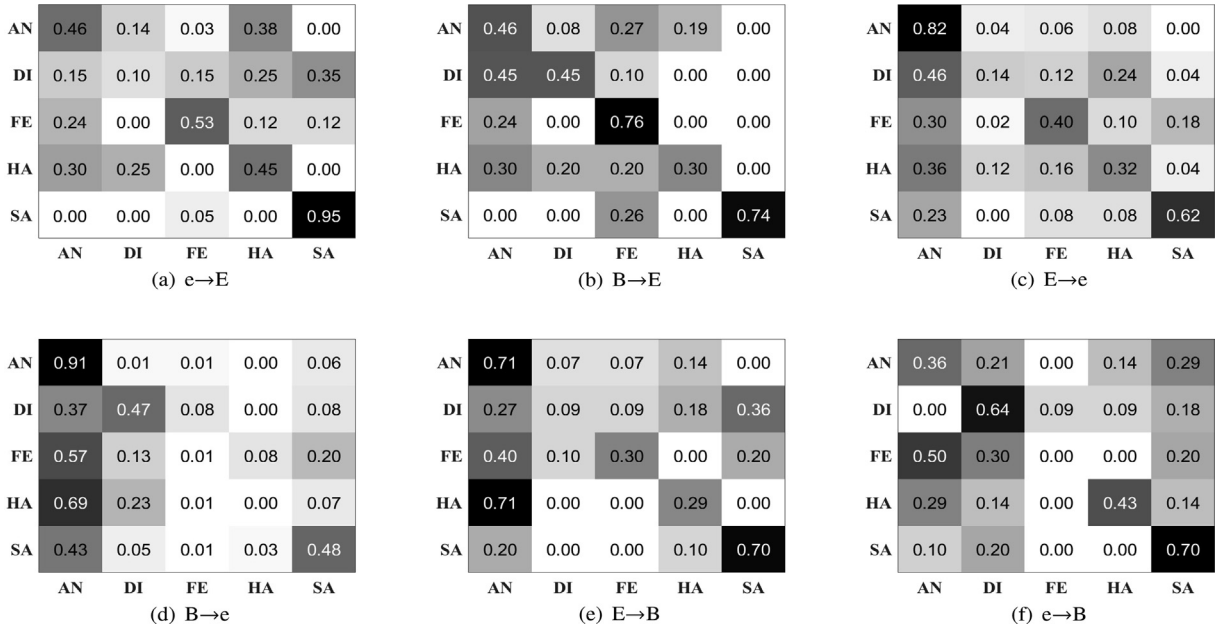

Fig. 2. Confusion matrices of TDLR under six experimental tasks.

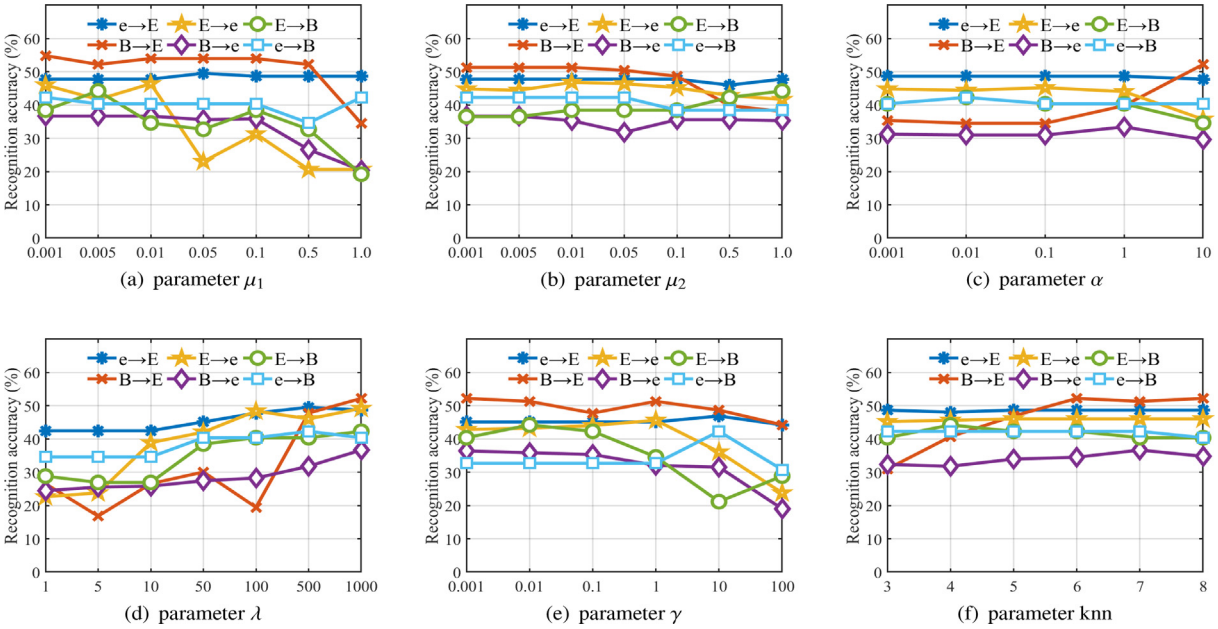

Fig. 3. Parameter sensitivity of TDLR under six experimental tasks.

Thirdly, for transfer learning algorithms, we can observe that the linear regression based algorithms perform better than the subspace learning based algorithms. The possible reason might be that, the regression based methods take into account the procedures of label prediction and feature transfer learning into a joint framework, while the subspace learning based methods conduct these two learning procedures independently, and may obtain sub-optimal results.

Fourthly, it is worth noticing that DRLS and JTSLR also take into account the category space to improve the performance. However, the proposed TDLR leverages the discriminant MMD and graph Laplacian as the distance metric, which can learn a better and more transferable regression model. Meanwhile, we introduce a label relaxation linear regression and impose an $\ell_{2,1}$-norm on the trans-

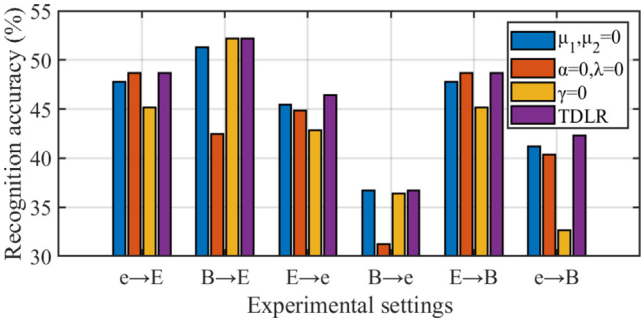

Fig. 4. Effectiveness verification of TDLR under six experimental tasks.

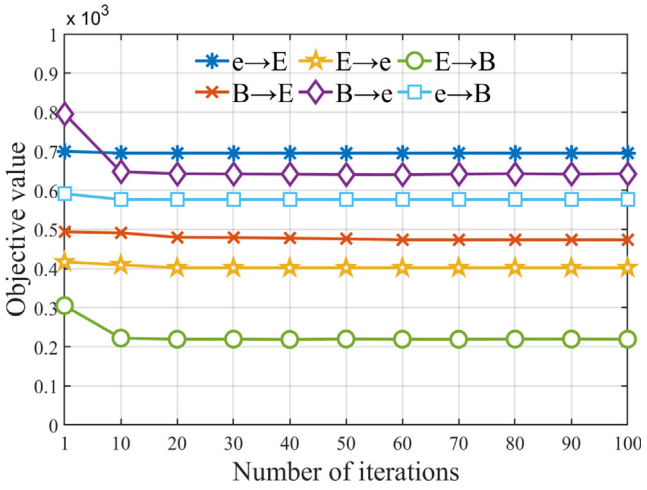

Fig. 5. Convergence analysis of TDLR.

formation matrix to ensure the learned regression model be discriminative. Lastly, we can see that when the number of samples in the target database is less than that in the source database, we can obtain a relatively high recognition accuracy. Specifically, compared with eNTERFACE'05, EMO-DB has fewer samples, and the recognition accuracy of our method in E→e task decreases by 2.25% compared with that in e→E task. We can also find that the recognition accuracy in B→e task decreases by 5.62% compared with e→B task, and the recognition accuracy in E→B task decreases by 7.98% compared with that in B→E task. These results show that, in cross-domain tasks, whether the labeled training samples are sufficient would affect the recognition accuracy.

Fig. 2 gives the confusion matrices of the TDLR algorithm under different tasks. Firstly, it can be seen from the figure that the recognition results of anger (AN) and sadness (SA) in all cases are generally better than those of the other emotions, which indicates that the AN and SA are easier to be recognized. Secondly, the recognition results of disgust (DI) and fear (FE) are also satisfactory. However, from Figures (d) and (f), we can find that FE performs poorly in both B→e and e→B tasks. And from Figures (a), (c) and (e), we can find that DI performs poorly in e→E, E→e, and E→B. These results indicate that, for a given emotion category, it is difficult to ensure that TDLR performs well on all experimental tasks. Thirdly, the overall performance of happiness (HA) is the worst. This indicates that the HA is easier to be confused in comparison with the other emotion categories.

4.3. Parameter sensitivity and effectiveness verification

In this section, we investigate the parameter sensitivity on the six major parameters, i.e., $\mu_1, \mu_2, \alpha, \lambda, \gamma$, and k. The results are depicted in Fig. 3. From the figure, we can have the following findings. Firstly, as shown in Figures (a) and (b), we can find that the algorithm is sensitive to μ_1. The values of μ_1 can be adjusted within the range of $0.001 \sim 0.1$. The optimal value of μ_1 is set as 0.01. The recognition accuracy changes slightly with the varying values of μ_2. This indicates that TDLR is less sensitive to the feature selection term of the linear reconstruction matrix. The optimal value of μ_2 is fine-tuned in the range of $0.01 \sim 0.1$. Secondly, from Figures (c) and (d), we can see that TDLR is highly sensitive to the parameters α and λ. The corresponding values of α and λ should be adjusted over a wide range to obtain the best performance. For our experiments, we choose λ and α as 1000 and 1, respectively.

Thirdly, from Figure (e), we can find that the recognition accuracy changes with the varying values of γ. As the values change, the recognition rates of TDLR fluctuate, but generally increase or decrease within a certain range. Thus, we choose the final value of γ as 0.001.

Finally, from Figure (f), we can observe that TDLR is less sensitive to the number of nearest neighbors. Finally, the value of k is optimally set as 6.

Fig. 4 shows the average recognition results of different components of TDLR. From the figure, we can have the observations as follows. Firstly, when μ_1 and μ_2 are set to zero, the recognition results are lower than those of TDLR, which proves that the linear reconstruction of category space plays a positive role in our model. Secondly, when α and λ are set to zero, the average recognition accuracy is lower than that of TDLR. This indicates that employing MMD and graph Laplacian as a dual metric can well improve the transfer performance. Lastly, when $\gamma = 0$, the average recognition accuracy decreases, which demonstrates that the discriminative feature representation benefits the transfer learning of features.

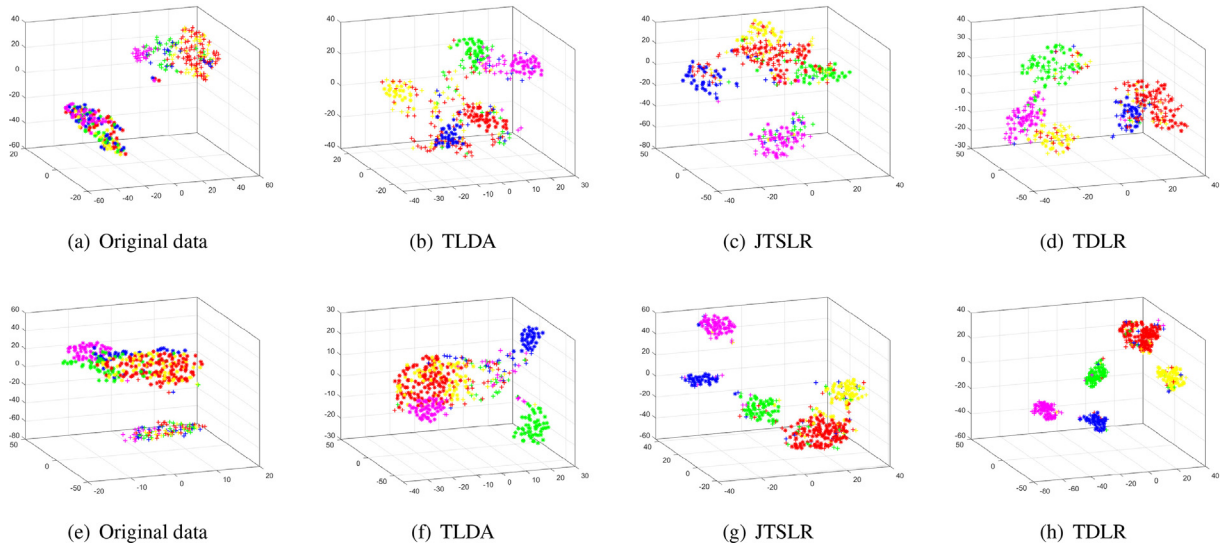

Fig. 6. The t-SNE data visualization results. The $*$ and + represent the source and target samples, respectively, and different colors represent different emotion categories (red: AN, blue: DI, green: FE, yellow: HA, and magenta: SA).

4.4. Convergence analysis

The convergence curves of the TDLR algorithm under six experimental tasks are depicted in Fig. 5. As can be seen from the figure, we can find that the values of the objective function decrease monotonously, and the curves remain stable after 10 iterations. These results prove that the obtained recognition results using 10 iterations are credible.

4.5. t-SNE visualization

To better show the effectiveness of the proposed TDLR, we give the data visualization results using the t-SNE algorithm [59]. Fig. 6 illustrates the t-SNE visualization of features for the tasks B→E (the upper part) and E→B (the lower part). Since TLDA and JTSLR are the most related works with our method, we choose these two methods for comparison. From the results, we have several observations. Firstly, compared with the original data, the divergence between the source and target corpora can be significantly narrowed by utilizing these three transfer learning algorithms, in which the samples follow similar data distributions.

Secondly, from the visualization results of Figures (b), (d), (f), and (h), we can see that although the subspace learned by TLDA effectively narrows the data discrepancy between domains, but the distribution of similar samples is scattered. The reason might be that the subspace learned by TLDA ignores the local information by using the matrix rank as the criterion of linear transformation. By contrast, our method utilizes the graph Laplacian to make the similar samples closer to each other. Thirdly, from the visualization results of Figures (c), (d), (g), and (h), we can see that compared with the best baseline method JTSLR, the subspace learned by our method is more discriminative. Although JTSLR considers the similarity of inter-corpus and intra-corpus to explore the discriminant information, but ignores the compact correlations between the category space of corpora. By contrast, our method develops a linear reconstruction strategy to explore the compact correlations between the source and target category space.

Finally, we can observe that the data distribution of HA is similar to that of AN. As a result, the samples of these two emotions are easily confused. Also, by observing the confusion matrices shown in Fig. 2, it can be found that more than half of HA samples are confused with AN emotion, which is one of the main reasons affecting the recognition performance.

5. Conclusion

In this paper, we propose a novel transfer learning algorithm for cross-corpus SER, named TDLR. Specifically, we develop a transferable discriminant linear regression to learn a transferable feature representation, in which a dual metric is integrated into the relaxed regression framework. We further consider the category spaces of both databases, and develop a linear reconstruction strategy to keep the close linear relationship between the category spaces. In addition, to make the learned feature representations be more discriminative and selective, we choose an $\ell_{2,1}$-norm on the transformation matrices. Extensive experiments on several benchmarks show the superior performance of the TDLR approach.

CRediT authorship contribution statement

Shaokai Li: Methodology, Writing - original draft. **Peng Song:** Supervision, Methodology, Writing - review & editing, Funding acquisition. **Wenjing Zhang:** Methodology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 61703360, the Fundamental Research Funds for the Central Universities under Grants 2242021k30014 and 2242021k30059, and the Graduate Innovation Foundation of Yantai University (GIFYTU).

References

[1] El Ayadi M, Kamel MS, Karray F. Survey on speech emotion recognition: Features, classification schemes, and databases. Pattern Recogn 2011;44(3):572–87.
[2] Fahad MS, Ranjan A, Yadav J, Deepak A. A survey of speech emotion recognition in natural environment. Digital Signal Process 2021;110:102951.
[3] Swain M, Routray A, Kabisatpathy P. Databases, features and classifiers for speech emotion recognition: a review. Int J Speech Technol 2018;21(1):93–120.
[4] Akçay MB, Oğuz K. Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. Speech Commun 2020;116:56–76.
[5] Cen L, Yu ZL, Dong MH. Speech emotion recognition system based on l1 regularized linear regression and decision fusion. In: International Conference on Affective Computing and Intelligent Interaction, Springer. p. 332–40.
[6] H. Hu, M.-X. Xu, W. Wu, GMM supervector based SVM with spectral features for speech emotion recognition, in: 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07, Vol. 4, IEEE, 2007, pp. IV–413.
[7] Atila O, Şengür A. Attention guided 3d cnn-lstm model for accurate speech based emotion recognition. Appl Acoust 2021;182:108260.
[8] Gangeh MJ, Fewzee P, Ghodsi A, Kamel MS, Karray F. Multiview supervised dictionary learning in speech emotion recognition. IEEE/ACM Trans Audio, Speech, Language Process 2014;22(6):1056–68.
[9] Zheng W, Xin M, Wang X, Wang B. A novel speech emotion recognition method via incomplete sparse least square regression. IEEE Signal Process Lett 2014;21(5):569–72.
[10] Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2009;22(10):1345–59.
[11] W.M. Kouw, M. Loog, An introduction to domain adaptation and transfer learning, ArXiv Preprint ArXiv:1812.11806.
[12] Long M, Wang J, Ding G, Sun J, Yu PS. Transfer feature learning with joint distribution adaptation. In: Proceedings of the IEEE International Conference on Computer Vision. p. 2200–7.
[13] Zhou T, Fu H, Gong C, Shen J, Shao L, Porikli F. Multi-mutual consistency induced transfer subspace learning for human motion segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. p. 10277–86.
[14] Taneja A, Arora A. Cross domain recommendation using multidimensional tensor factorization. Expert Syst Appl 2018;92:304–16.
[15] Long M, Wang J, Ding G, Pan SJ, Philip SY. Adaptation regularization: A general framework for transfer learning. IEEE Trans Knowl Data Eng 2013;26(5):1076–89.
[16] Hassan A, Damper R, Niranjan M. On acoustic emotion recognition: compensating for covariate shift. IEEE Trans Audio, Speech, Language Process 2013;21(7):1458–68.
[17] Deng J, Zhang Z, Marchi E, Schuller B. Sparse autoencoder-based feature transfer learning for speech emotion recognition. In: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, IEEE. p. 511–6.
[18] Deng J, Xu X, Zhang Z, Frühholz S, Schuller B. Universum autoencoder-based domain adaptation for speech emotion recognition. IEEE Signal Process Lett 2017;24(4):500–4.
[19] Zong Y, Zheng W, Zhang T, Huang X. Cross-corpus speech emotion recognition based on domain-adaptive least-squares regression. IEEE Signal Process Lett 2016;23(5):585–9.
[20] Zong Y, Zheng W, Huang X, Yan K, Yan J, Zhang T. Emotion recognition in the wild via sparse transductive transfer linear discriminant analysis. J Multimodal User Interfaces 2016;10(2):163–72.
[21] Song P, Zheng W. Feature selection based transfer subspace learning for speech emotion recognition. IEEE Trans Affective Comput 2018;11(3):373–82.
[22] Song P. Transfer linear subspace learning for cross-corpus speech emotion recognition. IEEE Trans Affective Comput 2019;10(2):265–75.
[23] Fang X, Xu Y, Li X, Lai Z, Wong WK, Fang B. Regularized label relaxation linear regression. IEEE Trans Neural Networks Learn Syst 2017;29(4):1006–18.

[24] Zhang J, Li W, Ogunbona P, Xu D. Recent advances in transfer learning for cross-dataset visual recognition: A problem-oriented perspective. ACM Computing Surveys (CSUR) 2019;52(1):1–38.
[25] Li G. Robust regression. Exploring Data Tables, Trends, and Shapes 1985;281:U340.
[26] Xiang S, Nie F, Meng G, Pan C, Zhang C. Discriminative least squares regression for multiclass classification and feature selection. IEEE Trans Neural Networks Learn Syst 2012;23(11):1738–54.
[27] Zhang X-Y, Wang L, Xiang S, Liu C-L. Retargeted least squares regression algorithm. IEEE Trans Neural Networks Learn Syst 2014;26:2206–13.
[28] Cai X, Ding C, Nie F, Huang H. On the equivalent of low-rank linear regressions and linear discriminant analysis based regressions. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. p. 1124–32.
[29] Lu Y, Lai Z, Fan Z, Cui J, Zhu Q. Manifold discriminant regression learning for image classification. Neurocomputing 2015;166:475–86.
[30] Zhang Z, Lai Z, Xu Y, Shao L, Wu J, Xie G-S. Discriminative elastic-net regularized linear regression. IEEE Trans Image Process 2017;26(3):1466–81.
[31] Wen J, Xu Y, Li Z, Ma Z, Xu Y. Inter-class sparsity based discriminative least square regression. Neural Networks 2018;102:36–47.
[32] Han N, Wu J, Fang X, Wong WK, Xu Y, Yang J, Li X. Double relaxed regression for image classification. IEEE Trans Circuits Syst Video Technol 2019;30(2):307–19.
[33] Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, Xiong H, He Q. A comprehensive survey on transfer learning. Proc IEEE 2020;109(1):43–76.
[34] Z. Wang, Y. Song, C. Zhang, Transferred dimensionality reduction, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2008, pp. 550–565.
[35] Daume III H, Marcu D. Domain adaptation for statistical classifiers. J Artif Intell Res 2006;26:101–26.
[36] Raina R, Battle A, Lee H, Packer B, Ng AY. Self-taught learning: transfer learning from unlabeled data. In: Proceedings of the 24th International Conference on Machine Learning. p. 759–66.
[37] Borgwardt KM, Gretton A, Rasch MJ, Kriegel H-P, Schölkopf B, Smola AJ. Integrating structured biological data by kernel maximum mean discrepancy. Bioinformatics 2006;22(14):e49–57.
[38] Pan SJ, Tsang IW, Kwok JT, Yang Q. Domain adaptation via transfer component analysis. IEEE Trans Neural Networks 2010;22(2):199–210.
[39] Kullback S, Leibler RA. On information and sufficiency. Ann Math Stat 1951;22(1):79–86.
[40] I. Dagan, L. Lee, F. Pereira, Similarity-based methods for word sense disambiguation, ArXiv Preprint Cmp-lg/9708010.
[41] Bregman LM. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. USSR Computational Mathematics and Mathematical Physics 1967;7(3):200–17.
[42] J. Shen, Y. Qu, W. Zhang, Y. Yu, Wasserstein distance guided representation learning for domain adaptation, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32, 2018, pp. 4058–4065.
[43] M. Belkin, P. Niyogi, V. Sindhwani, Manifold regularization: A geometric framework for learning from labeled and unlabeled examples, J. Mach. Learn. Res. 7 (11).
[44] M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering., in: NIPS, Vol. 14, 2001, pp. 585–591.
[45] Zhang L, Fu J, Wang S, Zhang D, Dong Z, Chen CP. Guide subspace learning for unsupervised domain adaptation. IEEE Trans Neural Networks Learn Syst 2019;31(9):3374–88.
[46] Wang S, Zhang L, Zuo W, Zhang B. Class-specific reconstruction transfer learning for visual recognition across domains. IEEE Trans Image Process 2019;29:2424–38.
[47] Nie F, Huang H, Cai X, Ding C. Efficient and robust feature selection via joint ℓ_2,1)norms minimization. Adv Neural Inform Process Syst 2010;23:1813–21.
[48] Q. Gu, Z. Li, J. Han, et al., Joint feature selection and subspace learning, in: IJCAI Proceedings-International Joint Conference on Artificial Intelligence, Vol. 22, Citeseer, 2011, p. 1294.
[49] F. Burkhardt, A. Paeschke, M. Rolfes, W.F. Sendlmeier, B. Weiss, et al., A database of german emotional speech., in: Interspeech, Vol. 5, 2005, pp. 1517–1520.
[50] Martin O, Kotsia I, Macq B, Pitas I. The enterface'05 audio-visual emotion database. In: 22nd International Conference on Data Engineering Workshops (ICDEW'06), IEEE. p. 8.
[51] Zhalehpour S, Onder O, Akhtar Z, Erdem CE. Baum-1: A spontaneous audio-visual face database of affective and mental states. IEEE Trans Affective Comput 2016;8(3):300–13.
[52] Bishop CM. Pattern recognition and machine learning. Springer; 2006.
[53] Yan S, Xu D, Zhang B, Zhang H-J, Yang Q, Lin S. Graph embedding and extensions: A general framework for dimensionality reduction. IEEE Trans Pattern Anal Mach Intell 2006;29(1):40–51.
[54] Long M, Wang J, Ding G, Sun J, Yu PS. Transfer joint matching for unsupervised domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. p. 1410–7.
[55] Zong Y, Zheng W, Huang X, Shi J, Cui Z, Zhao G. Domain regeneration for cross-database micro-expression recognition. IEEE Trans Image Process 2018;27(5):2484–98.
[56] W. Zhang, P. Song, D. Chen, C. Sheng, W. Zhang, Cross-corpus speech emotion recognition based on joint transfer subspace learning and regression, IEEE Transactions on Cognitive and Developmental Systems.
[57] B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. Müller, S. Narayanan, The interspeech 2010 paralinguistic challenge, in: Proc. INTERSPEECH 2010, Makuhari, Japan, 2010, pp. 2794–2797.
[58] Eyben F, Wöllmer M, Schuller B. Opensmile: the munich versatile and fast open-source audio feature extractor. In: Proceedings of the 18th ACM International Conference on Multimedia. p. 1459–62.
[59] Boureau Y-L, Bach F, LeCun Y, Ponce J. Learning mid-level features for recognition. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE. p. 2559–66.