INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

Unsupervised Transfer Components Learning for Cross-Domain Speech Emotion Recognition

Shenjie Jiang[1], *Peng Song*[1*], *Shaokai Li*[1], *Keke Zhao*[1], *Wenming Zheng*[2]

[1] School of Computer and Control Engineering, Yantai University, China
[2] Key Laboratory of Child Development and Learning Science (Southeast University), Ministry of Education, Research Center for Learning Science, Southeast University, China

pengsong@ytu.edu.cn

Abstract

Cross-domain speech emotion recognition (SER), which utilizes the source domain to recognize the emotions in the target domain, has received significant attention in recent years. In this paper, we propose a novel unsupervised transfer learning method named unsupervised transfer components learning (UTCL) for cross-domain SER. Specifically, we first learn a common projection for the cross-domain data, in which a PCA-like strategy is conducted for the source and target domains separately. Meanwhile, we design a simple strategy to ensure all cross-domain samples share similar manifold structures so that the learned common projection can preserve more transfer components. Furthermore, a novel adaptive structured graph strategy is designed to further narrow the gap between the cross-domain samples. Comprehensive experimental results on several benchmark datasets demonstrate that our method can achieve better performance in comparison with several state-of-the-art methods.

Index Terms: cross-domain, unsupervised transfer learning, speech emotion recognition, structured graph

1. Introduction

Speech emotion recognition (SER) is an important research problem in the fields of pattern recognition and speech signal processing, which has gained widespread attention in many practical application fields [1, 2]. The goal of SER is to automatically identify human beings' emotional states from speech signals, e.g. happiness, surprise, disgust, sadness, anger, and fear [1].

In real scenarios, the gender, ages, languages, and speaking styles of speakers might be different, and the recording scenes and equipment also change over time. These differences would lead to the domain mismatch problem, in which the training data and test data might be from different domains and follow different distributions [3]. In this situation, directly deploying the traditional methods could result in a significant decrease in the recognition performance. Thus, in this paper, we focus on tackling the cross-domain SER problem.

Transfer learning is an effective way to reduce the discrepancy between domains [4]. For instance, in [5], a joint distribution adaptation (JDA) framework is developed for feature transfer. In [6], a balanced distribution adaptation (BDA) approach is proposed for transfer learning, which can adaptively leverage the importance of the marginal and conditional distribution discrepancies. In [3], a feature selection based transfer subspace learning method is presented for cross-corpus SER. In [7], a discriminative transfer feature and label consistency (DTLC) approach is put forward for visual domain adaptation problems.

The above-mentioned methods utilize the label information to improve the performance of transfer learning, which can be categorized as semi-supervised transfer learning [8]. This strategy has achieved satisfactory results, but it relies on label information. In real life, a large amount of data in source domain might be unavailable. Thus, unsupervised learning [9] is an effective strategy to solve this problem. In the past decade, various unsupervised transfer learning algorithms have been presented. For instance, in [10], Pan et al. present a transfer component analysis (TCA) method, which extends PCA to a transferable manner. In [11], Chang et al. propose a novel multi-scale convolutional sparse coding (MSCSC) method for unsupervised transfer learning. In [12], Siddhant et al. propose a faster and simpler unsupervised pre-training method called ELMo-Light (ELMoL) for spoken language understanding. In [13], Noori Saray et al. develop a joint distinct subspace learning and unsupervised transfer classification (JDSC) approach for visual domain adaptation. In [14], Li et al. propose a transferable discriminant linear regression (TDLR) approach for cross-database SER. However, these methods only consider the common information and neglect the domain-specific information, which is vital to the performance of transfer learning [4].

To tackle the above-mentioned problems, in this paper, we propose a novel unsupervised transfer learning method named unsupervised transfer components learning (UTCL) method for cross-domain SER. In our method, we consider both the common and domain-specific principal components in the process of knowledge transfer. In addition, we design an adaptive structured graph as the distance metric, which can efficiently narrow the gap between the source and target domains. For better illustration, we show the flowchart of our method in Fig. 1.

2. Unsupervised Transfer Components Learning

2.1. Notations and symbols

Let $X_s \in \mathbb{R}^{m \times n_s}$ and $X_t \in \mathbb{R}^{m \times n_t}$ be the source and target feature matrices, respectively, where n_s and n_t are the numbers of source and target samples, respectively, and m is the dimensionality of features. Define $X \in \mathbb{R}^{m \times n} = [X_s, X_t]$,

This work was supported by the National Natural Science Foundation of China under Grant U2003207, Jiangsu Frontier Technology Basic Research Project under Grant BK20192004, the Natural Science Foundation of Shandong Province under Grant ZR2022MF314, and the Fundamental Research Funds for the Central Universities under Grants 2242021k30014 and 2242021k30059.
*The corresponding author.

10.21437/Interspeech.2023-274

where $n = n_s + n_t$. Let $P \in \mathbb{R}^{m \times d}$ be the matrix of projection, where d is the dimensionality of the common subspace. $Q_s \in \mathbb{R}^{m \times d}$ and $Q_t \in \mathbb{R}^{m \times d}$ denote the reconstruction matrices of the source and target domains, respectively. For an arbitrary matrix W, the Frobenius norm and the $\ell_{2,1}$-norm of W are defined as $||W||_F$ and $||W||_{2,1}$, respectively, and $\mathrm{Tr}(W)$ denotes the trace of W.

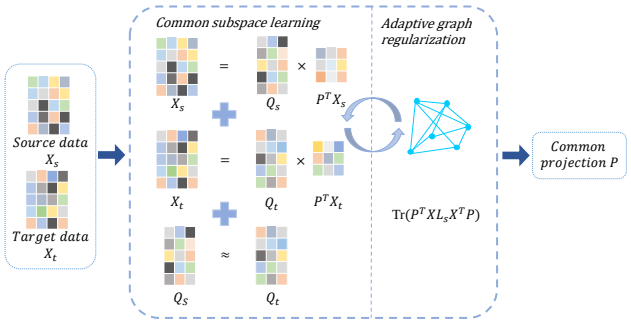


Figure 1: *The flowchart of our method.*

2.2. The proposed method

2.2.1. Common subspace learning

Different from existing transfer learning algorithms [4], which focus on learning the common information across domains, we aim to simultaneously obtain the common and domain-peculiar information. To this end, we learn a common projection and conduct a PCA-like strategy in the source domain and the target domain separately. Thus, the common subspace can preserve more principal components of the source and target domains when performing knowledge transfer. This problem can be formulated as the following equation:

$$\begin{aligned} &\min_{P,Q_s,Q_t} ||X_s - Q_s P^T X_s||_F^2 + ||X_t - Q_t P^T X_t||_F^2 + \beta ||P||_{2,1} \\ &s.t. P^T P = I, Q_s^T Q_s = I, Q_t^T Q_t = I \end{aligned} \tag{1}$$

where P is the common projection that preserves the principal transfer components, Q_s and Q_t represent the reconstruction matrices that map the common subspace into the original source and target domains, respectively. The $P^T P = I$, $Q_s^T Q_s = I$ and $Q_t^T Q_t = I$ are the orthogonal constraints. In addition, an $\ell_{2,1}$-norm is imposed on the matrix P to help select representative features, and β is a regularization parameter to control its sparsity.

By solving Eq. (1), we can obtain a common projection P which preserves the common principal components. Then, we can get the following equation:

$$\left.\begin{aligned} &X_s \approx Q_s P^T X_s \\ &X_t \approx Q_t P^T X_t \\ &Q_s^T Q_s = I \\ &Q_t^T Q_t = I \end{aligned}\right\} \Rightarrow \left.\begin{aligned} &Q_s^T X_s \approx P^T X_s \\ &Q_t^T X_t \approx P^T X_t \end{aligned}\right\} \tag{2}$$

We hope that $P^T X_s$ and $P^T X_t$ share similar feature distributions, i.e., $\phi_s(P^T X_s) \approx \phi_t(P^T X_t)$, in which all the cross-domain samples follow similar manifold structure in the common subspace. According to [15], if two data points are closer, the coupled data points in the projected low-dimensional subspace are also closer. Inspired by this, to eliminate the domain shift, we use a simple but effective strategy, which is expressed as follows:

$$\min_{Q_s,Q_t} ||Q_s - Q_t||_F^2 \tag{3}$$

By minimizing Eq. (3), we can get the following conclusions:

$$Q_s \approx Q_t \Rightarrow \phi_s(Q_s^T X_s) \approx \phi_t(Q_t^T X_t) \Rightarrow \phi_s(P^T X_s) \approx \phi_t(P^T X_t) \tag{4}$$

By integrating Eq. (1) into Eq. (3), we can get

$$\begin{aligned} &\min_{P,Q_s,Q_t} ||X_s - Q_s P^T X_s||_F^2 + ||X_t - Q_t P^T X_t||_F^2 \\ &\quad + \alpha ||Q_s - Q_t||_F^2 + \beta ||P||_{2,1} \\ &s.t. P^T P = I, Q_s^T Q_s = I, Q_t^T Q_t = I \end{aligned} \tag{5}$$

where α is a regularization parameter. By solving Eq. (5), we get a common subspace where all the cross-domain samples share similar manifold structure and can preserve more transfer components.

2.2.2. Adaptive graph regularization

The objective function in Eq. (5) can narrow the gap across domains efficiently. However, it only considers the global similarity across domains. According to [16, 17, 18], the local similarity is also virtual for the transferable performance. Thus, we design an adaptive structured graph to further reduce the distribution divergence across domains.

Given a similarity matrix $S \in \mathbb{R}^{n \times n}$, in which s_{ij} is the element of S. s_{ij} can be regarded as the similarity between x_i and x_j. The closer the two samples are, the larger s_{ij} is. It is obvious that s_{ij} is inversely proportional to the distance between x_i and x_j. For simplicity, the square of the Euclidean distance is used to calculate the distance, which is expressed as $||x_i - x_j||_2^2$. The process of determining s_{ij} can be regarded as solving the following equation:

$$\begin{aligned} &\min \sum_{i,j} ||x_i - x_j||_2^2 s_{ij} + \lambda s_{ij}^2 \\ &s.t. \forall i, s_i^T 1 = 1, 0 \le s_{ij} \le 1 \end{aligned} \tag{6}$$

Through Eq. (6), the exact connected components can be contained by the similarity matrix S [19]. To facilitate the optimization of Eq. (6), we transform it to the following equation:

$$\begin{aligned} &\min \mathrm{Tr}(X L_s X^T) + \lambda \sum_{i,j} s_{ij}^2 \\ &s.t. \forall i, s_i^T 1 = 1, 0 \le s_{ij} \le 1 \end{aligned} \tag{7}$$

where $L_s = D - \frac{S^T + S}{2}$ is the Laplacian matrix, D is a diagonal matrix, in which $\sum_j \frac{(s_{ij} + s_{ji})}{2}$ denotes the i-th item. As Eq. (1), we use the low-dimensional feature representation $P^T X$ to represent the high-dimensional data matrix X. Thus, we reformulate Eq. (7) as the following equation:

$$\begin{aligned} &\min_P \mathrm{Tr}(P^T X L_s X^T P) + \lambda \sum_{i,j} s_{ij}^2 \\ &s.t. \forall i, s_i^T 1 = 1, 0 \le s_{ij} \le 1, P^T P = I \end{aligned} \tag{8}$$

Combining Eq. (5) and Eq. (8), we can obtain the objective function of our proposed method as follows:

$$\begin{aligned}\min_{P,Q_s,Q_t} &||X_s - Q_s P^T X_s||_F^2 + ||X_t - Q_t P^T X_t||_F^2 + \alpha||Q_s - Q_t||_F^2 \\ &+ \beta||P||_{2,1} + \gamma \text{Tr}(P^T X L_s X^T P) + \lambda \sum_{i,j} s_{ij}^2 \\ s.t. &\forall i, s_i^T 1 = 1, 0 \leq s_{ij} \leq 1, P^T P = I, Q_s^T Q_s = I, Q_t^T Q_t = I\end{aligned} \tag{9}$$

where γ is a trade-off parameter.

2.3. Optimization

In this subsection, we develop an alternative optimization algorithm to solve our method. Since it is hard to directly optimize Eq. (9), we rewrite it as follows:

$$\begin{aligned}\mathcal{L} =&||X_s - Q_s P^T X_s||_F^2 + ||X_t - Q_t P^T X_t||_F^2 + \alpha||Q_s - Q_t||_F^2 \\ &+ \beta \text{Tr}(P^T G P) + \gamma \text{Tr}(P^T X L_s X^T P) + \lambda \sum_{i,j} s_{ij}^2 \\ &+ \text{Tr}(\phi(P^T P - I)) + \text{Tr}(\phi(Q_s^T Q_s - I)) + \text{Tr}(\phi(Q_t^T Q_t - I))\end{aligned} \tag{10}$$

where ϕ is a small Lagrange constraint, and $G \in \mathbb{R}^{d \times d}$ is a diagonal matrix [20]. The diagonal element of G can be written as

$$G_{ii} = \frac{1}{2\sqrt{p_i^T p_i + \varepsilon}} \tag{11}$$

where p_i is the i-th row of P, and ε is a very small constant.

We solve Eq. (10) by updating one variable while fixing the others. The steps are given as follows:

1) Fix S, Q_s and Q_t, and update P: We get the items related to P. Taking the partial derivative of $\mathcal{L}$ w.r.t. P, we can get the following equation:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial P} =&(X_s X_s^T P + X_t X_t^T P + \beta G P + \gamma X L_s X^T P + P \\ &- X_s X_s^T Q_s - X_t X_t^T Q_t)\end{aligned} \tag{12}$$

Set $\frac{\partial \mathcal{L}}{\partial P} = 0$, and define $V_s = X_s X_s^T, V_t = X_t X_t^T$, we can get the solution of P as

$$P = (V_s + V_t + \beta G + \gamma X L_s X^T + I)^{-1}(V_s Q_s + V_t Q_t) \tag{13}$$

2) Fix P, Q_s and Q_t, and update S: The problem (10) can be converted into the following equation:

$$\begin{aligned}\min_P &\text{Tr}(P^T X L_s X^T P) + \lambda \sum_{i,j} s_{ij}^2 \\ s.t. &\forall i, s_i^T 1 = 1, 0 \leq s_{ij} \leq 1, P^T P = I\end{aligned} \tag{14}$$

We introduce an auxiliary matrix $M \in R^{n \times n}$, in which $m_{ij} = \left|\left|P^T x_i - P^T x_j\right|\right|_2^2$, Eq. (14) can be converted into the following equation:

$$\min_{s_i^T 1 = 1, 0 \leq s_{ij} \leq 1} \left| |s_i + \frac{1}{2\lambda} m_i \right| |_2^2 \tag{15}$$

where s_i is the vector of matrix S. Then we can update L_s using Eq. (15).

3) Fix P, S and Q_t, and update Q_s: Eq. (10) can be transformed into the following problem:

$$\mathcal{L} = ||X_s - Q_s P^T X_s||_F^2 + ||Q_s - Q_t||_F^2 + \text{Tr}(\phi(Q_s^T Q_s - I)) \tag{16}$$

Taking the partial derivative of $\mathcal{L}$ w.r.t. Q_s, we can obtain the following equation:

$$\frac{\partial \mathcal{L}}{\partial Q_s} = -V_s P + Q_s P^T V_s P + \alpha Q_s - \alpha Q_t \tag{17}$$

Setting $\frac{\partial \mathcal{L}}{\partial Q_s}$ = 0, we can get the solution of Q_s as follows:

$$Q_s = (\alpha Q_t + V_s P)(P^T V_s P + \alpha I)^{-1} \tag{18}$$

4) Fix P, S and Q_s, and update Q_t: Similar to Eq. (18), we can obtain the following solution for Q_t:

$$Q_t = (\alpha Q_s + V_t P)(P^T V_t P + \alpha I)^{-1} \tag{19}$$

We repeat the above four steps until the objective function converges or the maximum iteration number reaches.

3. Experiments

3.1. Experimental settings

In this section, we evaluate the performance of the proposed algorithm for cross-domain SER on three public benchmark datasets, including Berlin [21], CVE [22] and IEMOCAP [23]. Two of the above datasets are randomly selected as the source and target domains, respectively. Thus we can get six types of cross-domain SER tasks (source → target): *C→ B, I→B, B→C, I→C, B→I*, and *C→I*. We choose four common emotion categories for experiments, including anger (AN), happiness (HA), neutral (NE), and sadness (SA). Additionally, each domain is divided into 10 parts, of which 7/10 are used for training and 3/10 are used for test.

We compare the proposed UTCL with several state-of-the-art subspace learning and transfer subspace learning methods, including principal component analysis (PCA), transfer component analysis (TCA) [10], joint distribution adaptation (JDA) [5], transfer joint matching (TJM) [24], balanced distribution adaptation (BDA) [6], transfer linear discriminant analysis (TLDA) [3], discriminative transfer feature and label consistency (DTLC) [7], and joint distinct subspace learning and unsupervised transfer classification (JDSC) [13]. For a fair comparison, we choose the linear SVM as the baseline classifier for all the compared algorithms. Besides, we select the recognition accuracy of the test set for evaluation.

3.2. Results and discussions

Table 1 displays the recognition results of different algorithms on six cross-domain SER tasks. As shown in the table, we have the following observations:

Firstly, compared with all the baseline methods, the proposed UTCL method achieves better recognition performance under all settings. the average classification accuracy of our method is 57.16%, which gains 4.25% improvement in comparison with the second-best method JDSC. This demonstrates that the proposed method can effectively solve the cross-domain SER problem.

Secondly, the recognition results of our method and other transfer learning methods significantly outperform the classical PCA method. This might be attributed to that, PCA does not consider the problem of feature distribution mismatch between source and target samples, whereas the transfer learning methods consider this problem, which can efficiently mitigate the distribution divergence across the domains.

Table 1: *Recognition results (%) of different algorithms on different tasks.*

Settings	Compared methods								UTCL
	PCA	TCA	JDA	TJM	BDA	TLDA	DTLC	JDSC	
C→B	56.54	65.98	60.82	67.01	57.27	59.79	62.71	68.04	**68.20**
I→B	30.31	50.52	53.61	53.61	59.21	56.41	52.73	52.58	**66.13**
B→C	45.74	53.21	51.92	48.08	50.41	55.56	50.76	57.69	**60.38**
I→C	35.16	40.38	51.28	41.03	49.32	**54.49**	44.10	46.17	52.05
B→I	44.21	43.73	37.42	43.21	**48.52**	32.44	40.22	47.29	47.41
C→I	44.62	46.77	46.77	47.29	44.10	50.19	43.23	45.66	**48.82**
Average	42.76	50.10	50.30	50.04	51.48	51.81	48.96	52.91	**57.16**

Thirdly, it is worth noticing that among the six groups of cross-domain tasks, the recognition performance of B→I and C→I is significantly lower than that of other tasks, The reason might be that the number of samples in IEMOCAP is larger than that of the other two datasets. When the IEMOCAP dataset is used as the target domain, the recognition results are relatively lower due to the small number of training samples. This indicates that the adequacy of training samples is vital to the performance of cross-domain SER.

3.3. Ablation analysis

In this subsection, we carry out the ablation study of UTCL by setting the related main parameters to zero. We have the following three cases:

- Setting $\alpha = 0$, the reconstruction matrix alignment term is ignored.
- Setting $\beta = 0$, the sparsity of W is ignored.
- Setting $\gamma = 0$, the adaptive structured graph term is ignored.

From Fig. 2, we can find that these three special cases perform significantly worse than UCTL. This result demonstrates that all three terms play a positive role in our method.

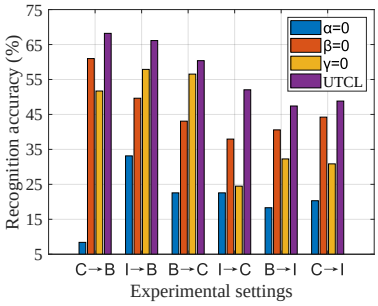


Figure 2: *Ablation study of our method under six different tasks.*

3.4. Convergence analysis

In this subsection, we give the convergence analysis of UTCL under six different tasks. The results are shown in Fig. 3. As can be shown in the figure, we can find that, UTCL can achieve stable satisfactory performance within only a few iterations ($T < 20$), and the values of our objective function decrease monotonously. These results prove the convergence property of our method.

3.5. t-SNE visualization

To better demonstrate the performance of our method, we give the data visualization results of t-SNE [25] in Fig. 4. Here we

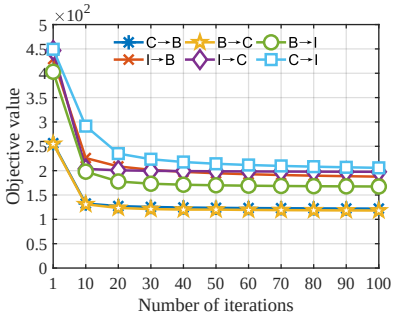
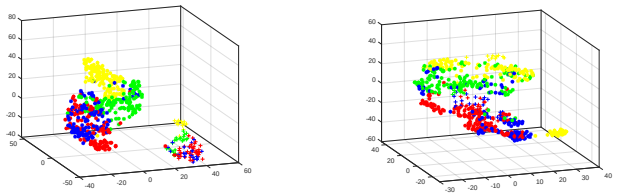


Figure 3: *Convergence analysis of the proposed UTCL.*

take B→C as an example. As shown in Fig. 4 (a), we can find that there exists a distribution discrepancy between source and target domains in the original data space. Fig. 4 (b) shows the results using the proposed UTCL method. From the figure, we can observe that the source and target data follow similar feature distributions after projection, and the cross-domain samples from the same category are close to each other.



(a) Original data (b) UTCL

Figure 4: *t-SNE visualization of data on the C→B task. The + and ∗ represent the data of source and target domains, respectively. Four colors represent four different emotion categories (red: AN, blue: HA, green: NE, and yellow: SA).*

4. Conclusions

In this paper, we propose a novel method called unsupervised transfer components learning (UTCL) for cross-domain SER. Different from current transfer learning algorithms, UTCL can efficiently learn transfer components by preserving the common and domain-specific information. Moreover, we design an adaptive structured graph as the distance metric. Thus, the discrepancy between the two domains can be efficiently reduced. We conduct extensive experiments on three benchmark datasets, and the results demonstrate the superiority of UTCL against several state-of-the-art methods.

5. References

[1] M. B. Akçay and K. Oğuz, "Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers," *Speech Communication*, vol. 116, pp. 56–76, 2020.

[2] S. Latif, R. Rana, S. Khalifa, R. Jurdak, J. Qadir, and B. W. Schuller, "Survey of deep representation learning for speech emotion recognition," *IEEE Transactions on Affective Computing*, 2021.

[3] P. Song and W. Zheng, "Feature selection based transfer subspace learning for speech emotion recognition," *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 373–382, 2020.

[4] L. Zhang and X. Gao, "Transfer adaptation learning: A decade survey," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[5] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 2200–2207.

[6] J. Wang, Y. Chen, S. Hao, W. Feng, and Z. Shen, "Balanced distribution adaptation for transfer learning," in *2017 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2017, pp. 1129–1134.

[7] S. Li, C. H. Liu, L. Su, B. Xie, Z. Ding, C. P. Chen, and D. Wu, "Discriminative transfer feature and label consistency for cross-domain image classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 11, pp. 4842–4856, 2020.

[8] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[9] S. Niu, Y. Liu, J. Wang, and H. Song, "A decade survey of transfer learning (2010–2020)," *IEEE Transactions on Artificial Intelligence*, vol. 1, no. 2, pp. 151–166, 2020.

[10] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2010.

[11] H. Chang, J. Han, C. Zhong, A. M. Snijders, and J.-H. Mao, "Unsupervised transfer learning via multi-scale convolutional sparse coding for biomedical applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 5, pp. 1182–1194, 2017.

[12] A. Siddhant, A. Goyal, and A. Metallinou, "Unsupervised transfer learning for spoken language understanding in intelligent agents," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 4959–4966.

[13] S. Noori Saray and J. Tahmoresnezhad, "Joint distinct subspace learning and unsupervised transfer classification for visual domain adaptation," *Signal, Image and Video Processing*, vol. 15, no. 2, pp. 279–287, 2021.

[14] S. Li, P. Song, and W. Zhang, "Transferable discriminant linear regression for cross-corpus speech emotion recognition," *Applied Acoustics*, vol. 197, p. 108919, 2022.

[15] J. Hamm and D. D. Lee, "Grassmann discriminant analysis: a unifying view on subspace-based learning," in *Proceedings of the 25th International Conference on Machine Learning*, 2008, pp. 376–383.

[16] W. Zhang and P. Song, "Transfer sparse discriminant subspace learning for cross-corpus speech emotion recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 307–318, 2020.

[17] W. Zhang, P. Song, and W. Zheng, "Joint local-global discriminative subspace transfer learning for facial expression recognition," *IEEE Transactions on Affective Computing*, 2022.

[18] J. Li, W. Liu, Y. Zhou, J. Yu, D. Tao, and C. Xu, "Domain-invariant graph for adaptive semi-supervised domain adaptation," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 18, no. 3, pp. 1–18, 2022.

[19] O. R. Oellermann and A. J. Schwenk, "The laplacian spectrum of graphs," *Graph Theory, c, Appl*, vol. 2, pp. 871–898, 1991.

[20] Q. Gu, Z. Li, and J. Han, "Joint feature selection and subspace learning," in *Twenty-second International Joint Conference on Artificial Intelligence*, 2011.

[21] 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.

[22] P. Liu and M. D. Pell, "Recognizing vocal emotions in Mandarin Chinese: A validated database of Chinese vocal emotional stimuli," *Behavior Research Methods*, vol. 44, no. 4, pp. 1042–1051, 2012.

[23] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee, and S. S. Narayanan, "IEMOCAP: Interactive emotional dyadic motion capture database," *Language Resources and Evaluation*, vol. 42, no. 4, pp. 335–359, 2008.

[24] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer joint matching for unsupervised domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1410–1417.

[25] Y.-L. Boureau, F. Bach, Y. LeCun, and J. Ponce, "Learning mid-level features for recognition," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 2559–2566.