A Generalized Subspace Distribution Adaptation Framework for Cross-Corpus Speech Emotion Recognition

Shaokai Li[1] Peng Song[1*] Liang Ji[14] Yun Jin[2] Wenming Zheng[3]

[1]**Yantai University** [2]Jiangsu Normal University
[3]The Key Laboratory of Child Development and Learning Science (Southeast University)
[4]The State Key Laboratory of Tibetan Intelligent Information Processing and Application

2023.04

CONTENTS

– 01 –

Background

01 Background

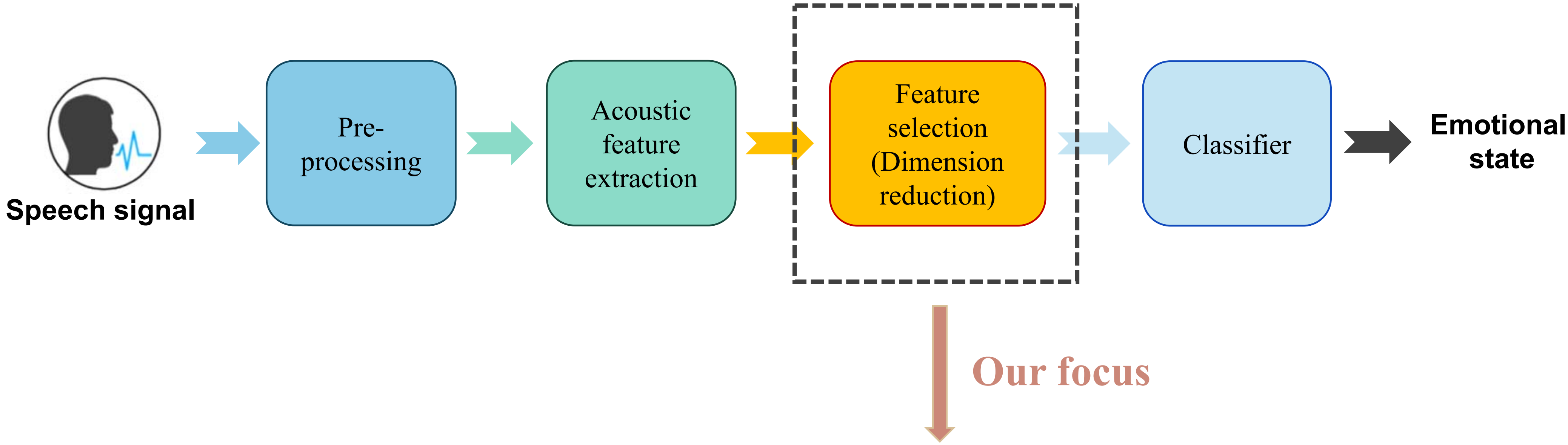
Speech Emotion Recognition (SER) is an important research direction in affective computing, pattern recognition, signal processing and human-machine interaction. The goal of SER is to identify the emotion categories from speech signals, such as fear, anger, sadness, pleasure, and so on.

Driving assist system

Automatic translation

Education

01 The process of SER



Learning a transfer subspace, which can obtain a common subspace to reduce the discrepancy between databases.

01 Traditional SER method

Many classification algorithms have been employed for SER， including：

- Hidden Markov model (HMM)
- Gaussian mixture model (GMM)
- Support vector machine (SVM)
- Neural network (NN)
- Deep neural network (DNN)
- Sparse representation
- Regression algorithms

– 02 –

The challenging problem

02 The challenging problem

The challenging cross-database SER problem

- **Data distribution mismatch problem:** in practical application scenarios, the speaker's gender, language, age and so on are different.

- **Insufficient labels problem:** labeling speech emotion is time-consuming, laborious, and require a large number of professionals.

02 Transfer learning

Transfer learning: The idea of transfer learning is to transfer the knowledge gained from one domain (source domain) to learn the knowledge of related but different domain (target domain).

We take the labeled database as the source domain and the unlabeled database as the target domain. The transfer learning can be used to solve the cross-database SER problem.

02 The related works

Transfer learning for cross-database:

- Joint Distribution Adaptation (JDA) 2013
- Domain-adaptive least-squares regression (DaLSR) 2016
- Joint Geometrical and Statistical Alignment (JGSA) 2017
- Locality Preserving Joint Transfer (LPJT) 2019
- Transfer Sparse Discriminant Subspace Learning (TSDSL) 2019
- Deep Adaptation Networks (DAN) 2015
- Deep Subdomain Adaptation Networks (DSAN) 2021
- Deep Adaptation Regression (DAR) 2021
- Dual-level Adaptive and Discriminative (DLAD) 2022

— 03 —

The proposed method

03 The proposed method

Problem Formulation:

We aim to learn a common projection subspace P by aligning the source and target distributions, where the corpus discrepancy would be well reduced. The objective function of the proposed GSDA can be formulated as follows:

$$\min_{P} \mathcal{F}(P, X) + \mathcal{G}(P, X) + \gamma\, \mathcal{S}(P)$$
$$\text{s.t. } P^T P = I$$

The first item F(P, X) is a generalized subspace learning method, in which the original feature space is projected into a low-dimensional common subspace. The second item G(P, X) is the distance metric learning strategy. The third item S(P) is a sparse regularization term.

03 The proposed method

Distance Metric：

$$\mathcal{G}(P, X) = \alpha\|V \odot S\|_1 - \beta\|V \odot D\|_1$$

Examples of GSDA：

1） GSDA-PCA:

$$\min_{P} -\mathrm{Tr}(P^T X X^T P) + \alpha\|V \odot S\|_1 - \beta\|V \odot D\|_1 + \gamma\|P\|_{2,1}$$
$$\text{s.t. } P^T P = I$$

2） GSDA-LDA:

$$\min_{P} \mathrm{Tr}(P^T (S_w - \mu S_b) P) + \alpha\|V \odot S\|_1 - \beta\|V \odot D\|_1 + \gamma\|P\|_{2,1}$$
$$\text{s.t. } P^T P = I$$

— 04 —

Experiments

04 Experimental setup

- **Databases:** Berlin (B) , IEMOCAP (I), and CVE (C).

We select five common emotional categories, i.e., anger (AN), neutral (NE), happiness (HA), and sadness (SA), in our experiments.

- **Feature Extraction:**

Low-level feature: we use the openSMILE toolkit to extract the feature set of the INTERSPEECH 2010 paralinguistic challenge (1582-dimensional).

Deep feature: we extract the Mel spectrograms to learn 2,048-dimensional deep features by ResNet50. Specifically, given a cross-corpus task, we fine-tune a pre-trained ResNet-50 model on the source corpus, and extract 2048-dimensional deep features of the target corpus using the fine-tuned mode.

- **Classifier:** linear SVM.

- **Evaluation metric:** the weighted average recall (WAR).

04 Experimental results

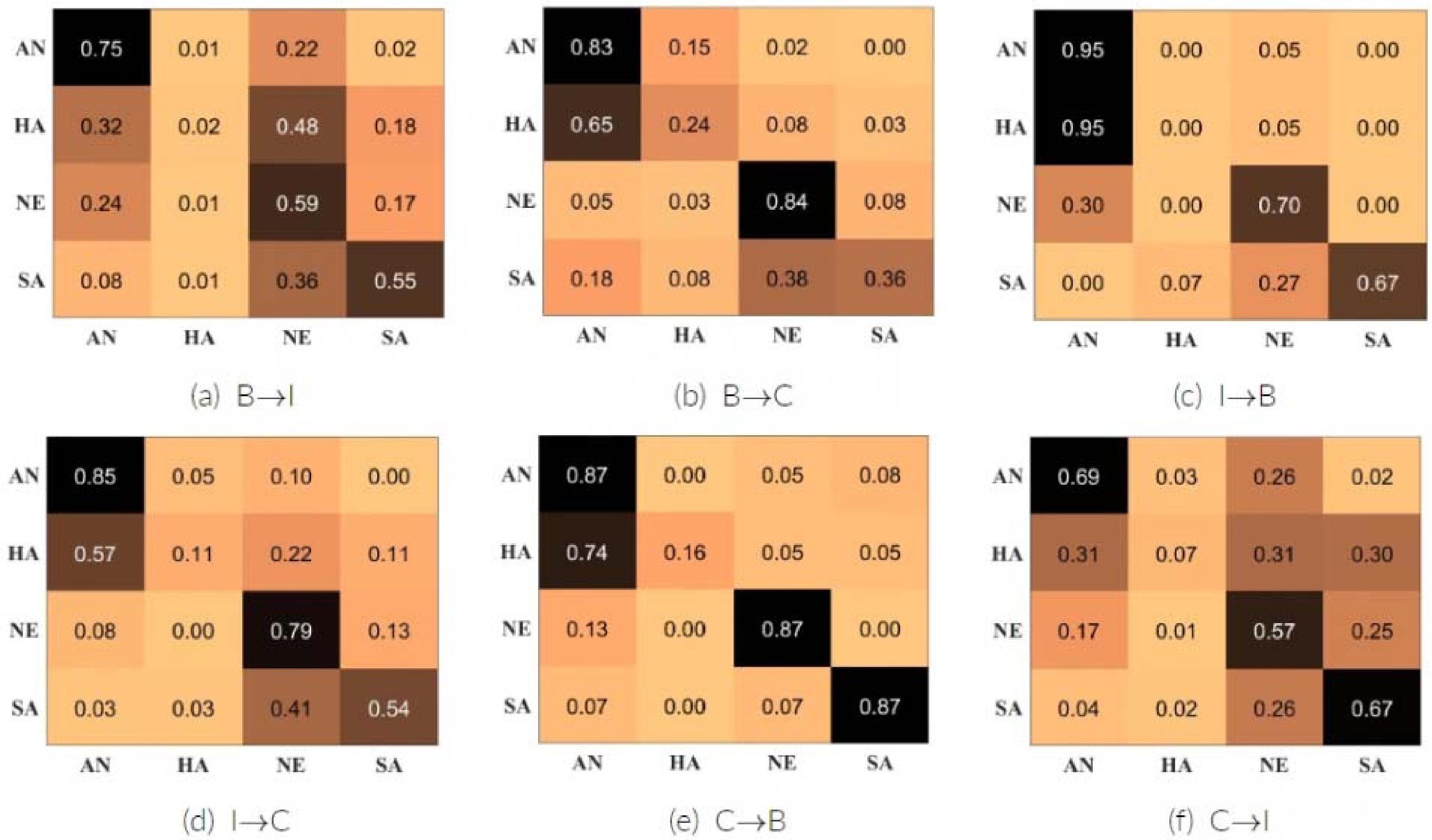
Results for Low-level Feature and Deep Feature

Tasks	Traditional methods		Transfer learning methods						GSDA -PCA	GSDA -LDA
	PCA	LDA	TCA	JDA	DaLSR	JGSA	LPJT	TSDSL		
B→I	44.21	40.11	46.54	46.06	49.19	45.24	45.96	49.25	**50.07**	50.35
B→C	45.74	39.35	49.83	48.16	48.22	49.67	46.87	52.12	**52.74**	51.32
I→B	30.31	40.62	55.85	53.66	49.37	59.70	**60.20**	59.79	57.10	59.18
I→C	35.16	30.32	43.38	47.12	51.61	46.32	53.21	51.19	**53.41**	48.38
C→B	56.54	56.33	59.81	62.62	49.47	58.14	59.18	63.70	63.95	**66.22**
C→I	44.62	32.39	47.13	48.11	49.94	47.78	46.69	46.51	**50.65**	50.45
Average	42.76	39.85	50.42	50.95	49.63	51.14	52.01	53.76	**54.65**	54.31

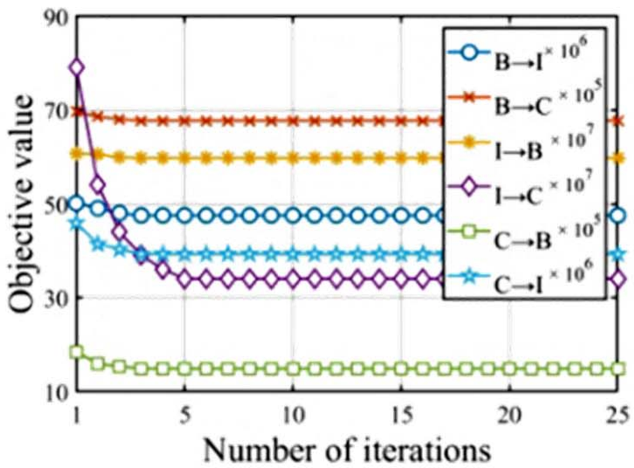
Results for Deep Feature

Tasks	Traditional methods		Transfer learning methods									GSDA -PCA	GSDA -LDA
	PCA	LDA	TCA	JDA	DaLSR	JGSA	LPJT	TSDSL	DAN*	DSAN*	DAR*		
B→I	40.19	37.74	43.61	44.28	43.03	43.74	44.09	44.87	46.96	**47.63**	47.23	44.57	44.94
B→C	41.93	42.87	50.96	49.67	45.16	50.61	53.54	55.06	50.14	46.06	54.93	54.19	**55.80**
I→B	42.70	43.75	59.37	59.53	59.67	63.20	66.66	65.62	55.70	62.96	62.50	**66.75**	65.54
I→C	32.25	26.45	45.80	48.38	49.03	46.80	48.18	49.16	45.74	47.02	46.42	49.09	**49.23**
C→B	61.35	59.37	64.58	65.62	62.38	63.75	**69.79**	67.71	62.16	63.58	62.50	**69.79**	67.70
C→I	35.06	32.42	44.42	43.02	39.37	43.98	43.11	44.02	43.51	40.77	42.19	**45.39**	44.56
Average	42.24	40.43	51.45	51.75	49.77	52.01	54.22	54.40	50.70	51.33	52.62	**54.79**	54.62

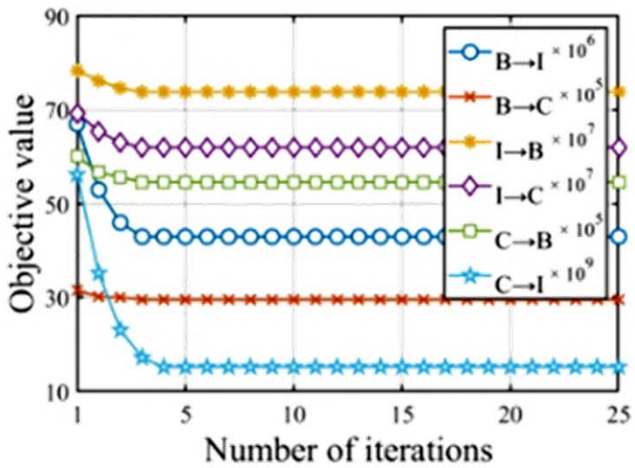
04 Confusion Matrices



04 Convergence Analysis



GSDA-PCA



GSDA-LDA

Conclusions

- We proposed GSDA utilizes a novel distance metric learning strategy to reduce the discrepancy between different corpora
- Extensive experimental results show that the proposed GSDA achieves superior performance than state-of-the-art compared algorithms.
- In the future, we will investigate to develop the deep transfer learning methods using the the proposed strategy to solve the cross-corpus dimensional SER problem.

Thank you!