INTERSPEECH 2023

20th-24th 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]Yantai University

[2]Southeast University

2023.08

# CONTENTS

01 Background

02 The challenging problem

03 The proposed method

04 Experiments

# Background

# 01 Background

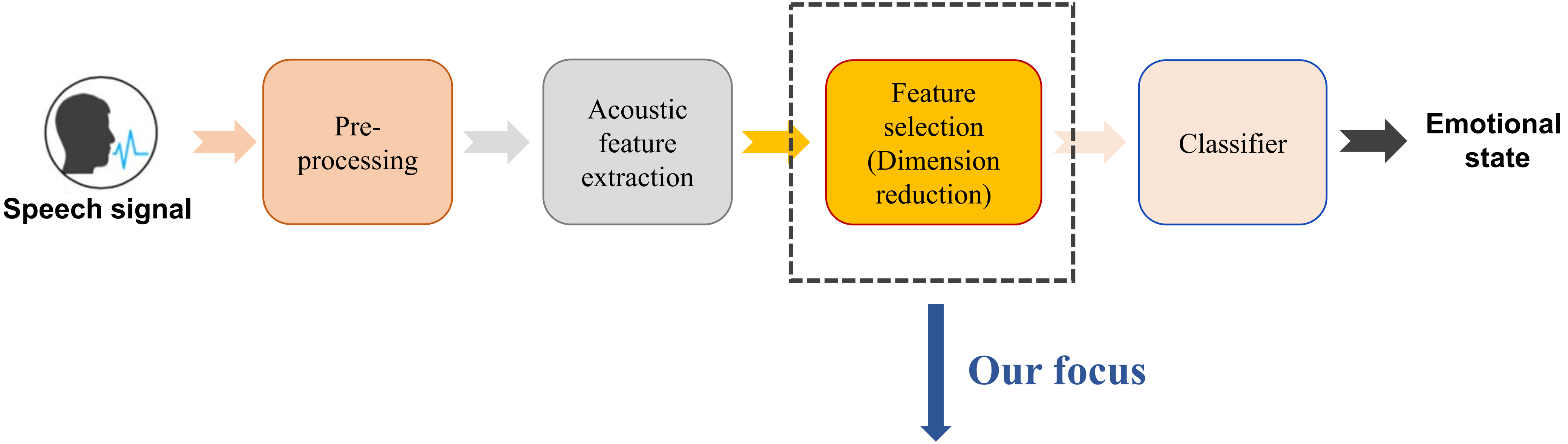
The main purpose of **Speech Emotion Recognition (SER)** is to classify speech signals according to different emotions, such as anger, disgust, fear, happiness, and sadness. It is widely used in various popular fields such as affective computing, pattern recognition, signal processing and human-computer interaction.

Driving assist system

Automatic translation

Robot interaction

# 01 The process of SER



Speech signal

Pre-processing

Acoustic feature extraction

Feature selection (Dimension reduction)

Classifier

Emotional state

Our focus

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 problem of SER

- **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-domain SER problem.

# 02 The related works

## Transfer learning for cross-domain SER:

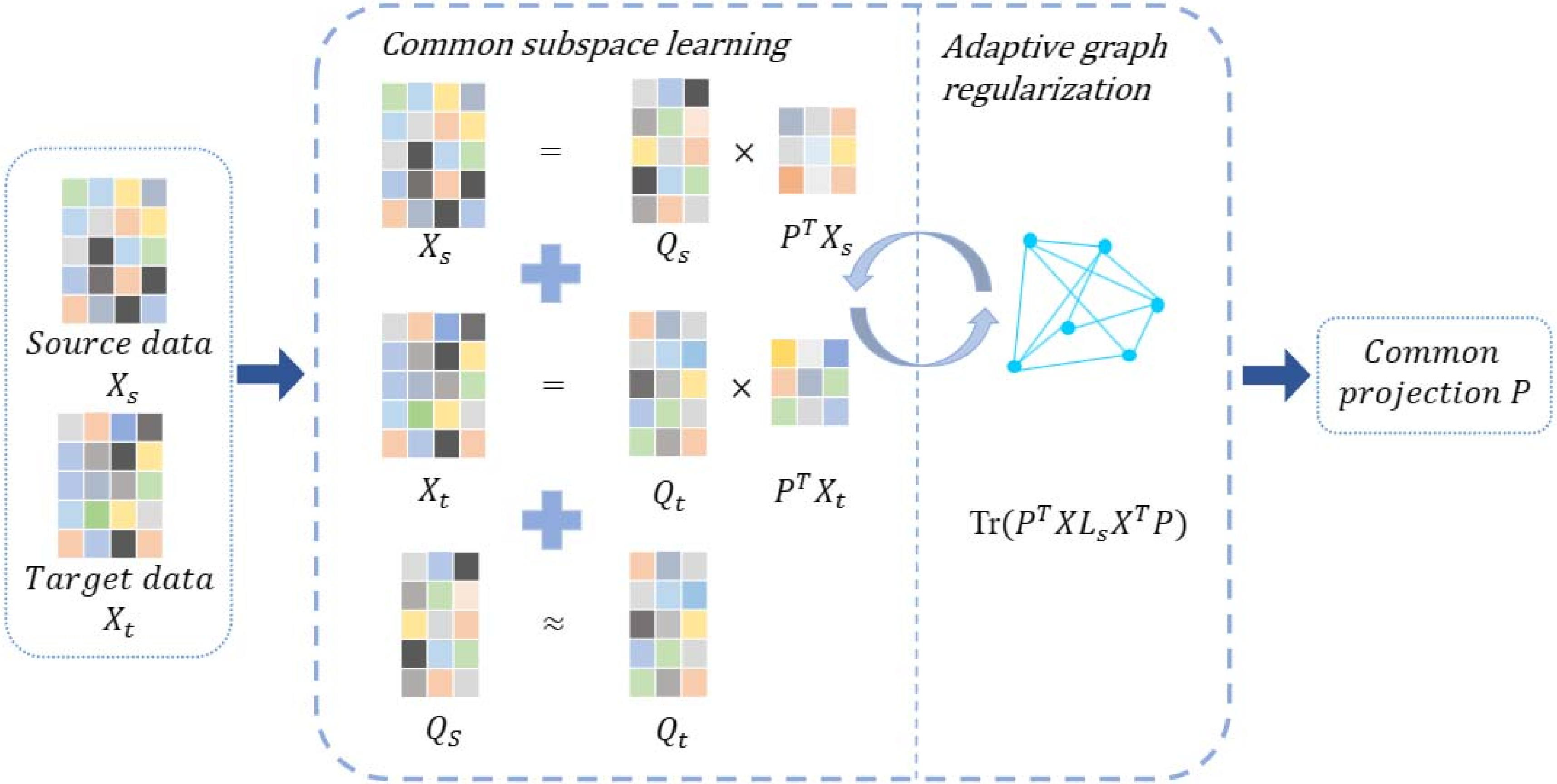
- transfer component analysis (TCA) 2010
- joint distribution adaptation (JDA)  2013
- transfer joint matching (TJM) 2014
- balanced distribution adaptation (BDA) 2017
- transfer linear discriminant analysis (TLDA) 2018
- scriminative transfer feature and label consistency (DTLC) 2020
- joint distinct subspace learning and unsupervised transfer classification (JDSC) 2021
- generalized subspace distribution adaptation(GDSA) 2023

# 03

# The proposed method

# 03 The proposed method

## 1.Our method framework:



# 03 The proposed method

## 1.Common subspace learning

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. In addition, we use a simple but effective strategy to eliminate the domain Shift. This problem can be formulated as the following equation:

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

# 03 The proposed method

## 2. Adaptive graph regularization

Because the local similarity is also virtual for the transferable performance, we design an adaptive structured graph to further reduce the distribution divergence across domains.

$$\min_{P} \mathrm{Tr}(P^{T} X L_{s} X^{T} P) + \lambda \sum_{i,j} s_{ij}^{2}$$

$$s.t. \forall i, s_{i}^{T} 1 = 1, 0 \leq s_{ij} \leq 1, P^{T} P = I$$

# 03 The proposed method

## 2. Our method UTCL

Combining the above two equations, we can obtain the objective function of our proposed method as follows:

common subspace learning

domain distribution alignment

$$\min_{P,Q_s,Q_t} \boxed{||X_s - Q_sP^TX_s||_F^2 + ||X_t - Q_tP^TX_t||_F^2} + \boxed{\alpha||Q_s - Q_t||_F^2}$$

$$+ \boxed{\beta||P||_{2,1}} + \boxed{\gamma \mathrm{Tr}(P^TXL_sX^TP) + \lambda\sum_{i,j} s_{ij}^2}$$

$$s.t. \forall i, s_i^T1 = 1, 0 \leq s_{ij} \leq 1, P^TP = I, Q_s^TQ_s = I, Q_t^TQ_t = I$$

sparse constraint

adaptive graph regularization

# 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.

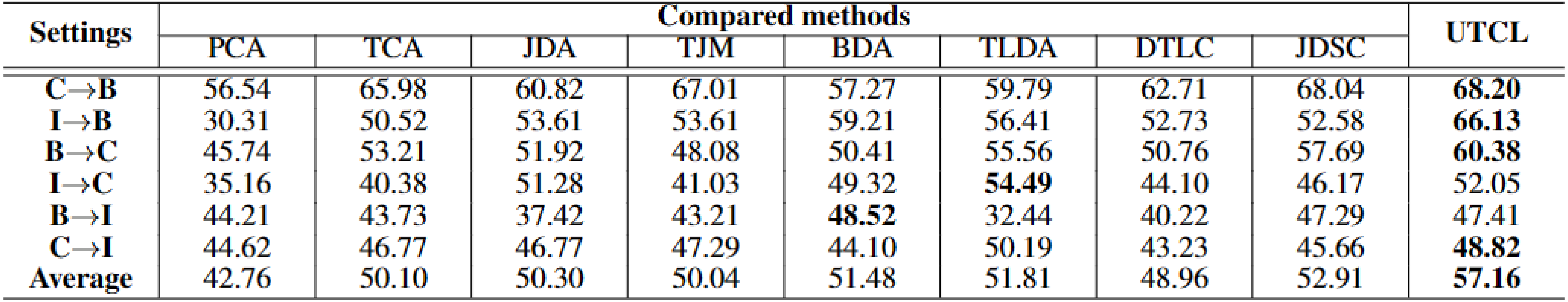
- **Feature Extraction:**

We use the openSMILE toolkit to extract the feature set of the INTERSPEECH 2010 paralinguistic challenge (1582-dimensional).

- **Classifier:** linear SVM.

# 04 Experimental results

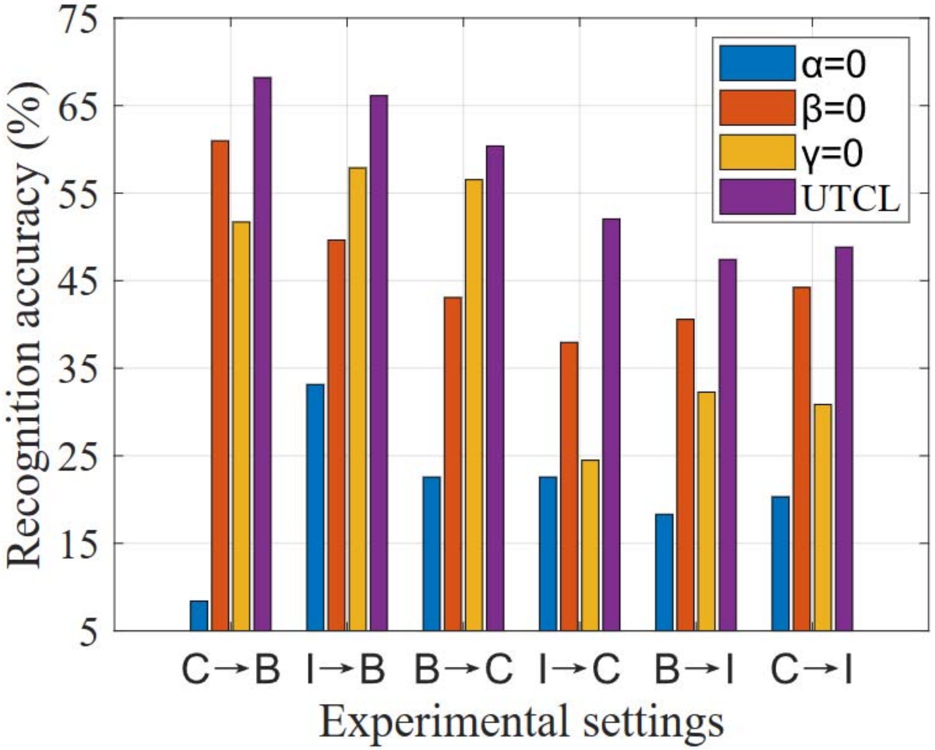
## UTCL results:



| **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** |

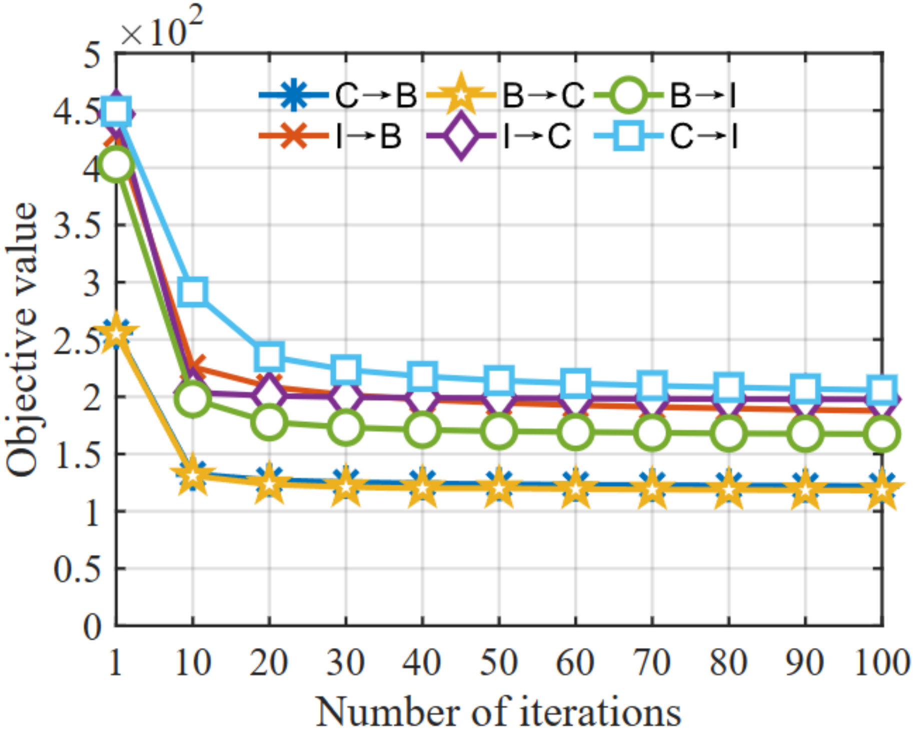
# 04 Experimental results

## Ablation analysis:



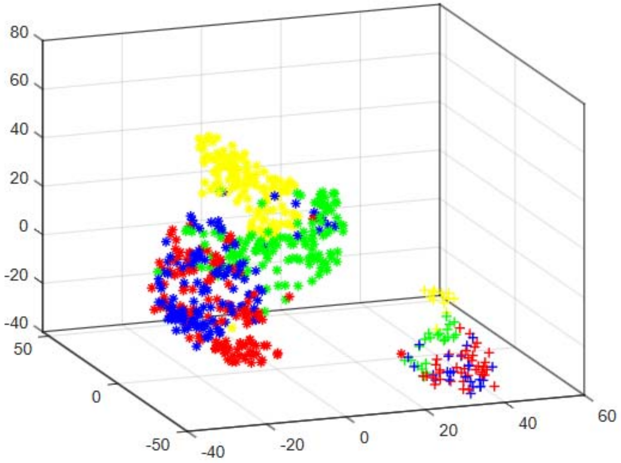
# 04 Experimental results

## Convergence analysis :

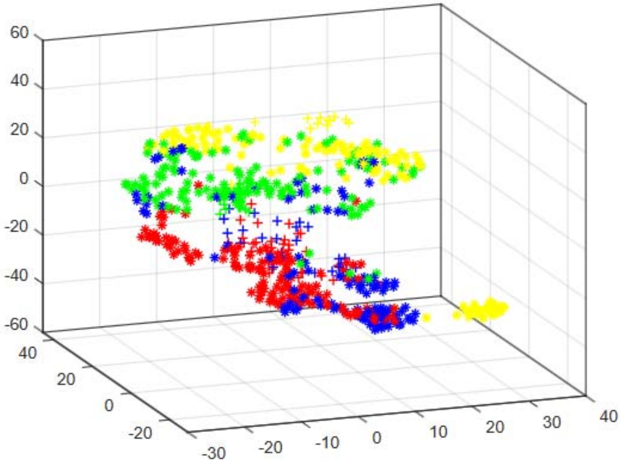


# 04 Experimental results

## t-SNE visualization:



(a) Original data



(b) UTCL

# Conclusions

## In our method UTCL:

- We consider both the common and domain-specific principal components in the process of knowledge transfer.

- We design an adaptive structured graph as the distance metric, which can efficiently narrow the gap between the source and target domains.

- 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!