ICASSP·2016
Shanghai, China

SPEECH EMOTION RECOGNITION USING TRANSFER NON-NEGATIVE MATRIX FACTORIZATION

Peng Song
School of Computer and Control Engineering, Yantai University
pengsongseu@gmail.com

2016.3.25

Outline

PARTI: Introduction

PART2: Baseline NMF algorithm

PART3: Our proposed method

PART4: Experimental Results and Discussions

PART5: Conclusion and Future Work

PARTI: Introduction

Speech Emotion Recognition

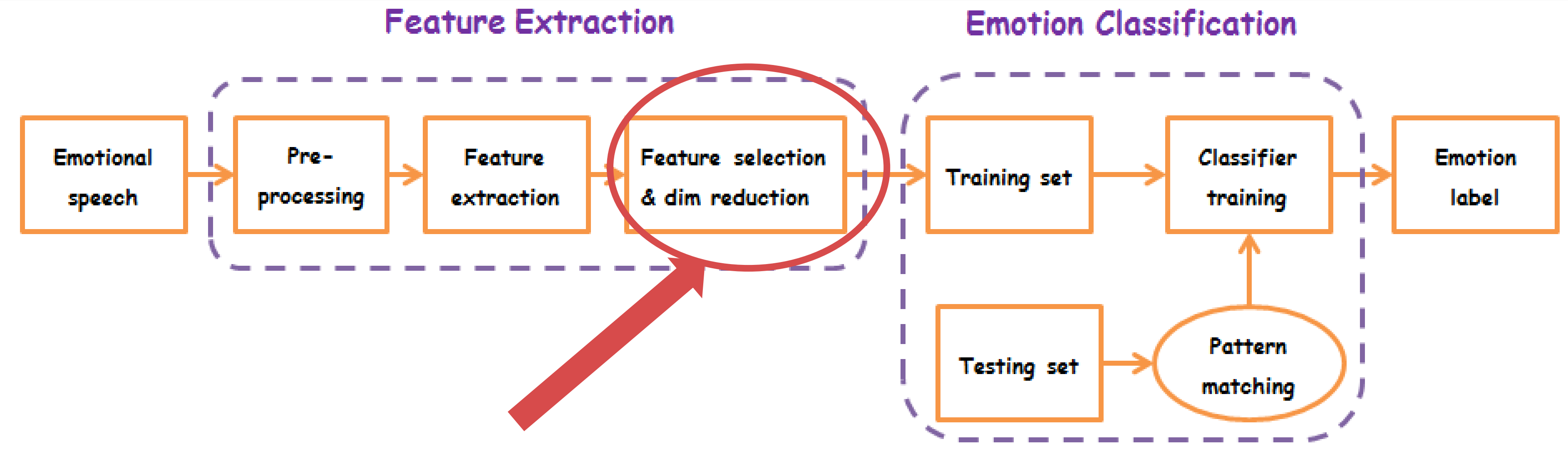
Definition:

- As a hot research topic in affective computing and speech signal processing fields, the goal of speech emotion recognition is to automatically recognize emotions from speech, e.g., anger, happiness, sadness, surprise.

Application:

- ■ Intelligent transportation systems
- ■ Healthcare field
- ■ Call Centers
- ■ Many other HCI fields

Framework of Speech Emotion Recognition



Our focus in this paper

Recognition Methods

All kinds of classification methods popular in pattern recognition and machine learning fields, are employed for emotional label classification or prediction including:

- support vector machine (SVM)
- hidden Markov model (HMM)
- Gaussian mixture model (GMM)
- neural network (NN)
- some regression methods
- deep neural network (DNN)
- extreme learning machine (ELM)

Weakness:

- They are carried out and evaluated on single corpus. In practice, it is too hard to collect a large emotional speech dataset, and **the training data and testing data are often collected from different devices and environments,** this discrepancy will obviously influence the recognition performance.

Recognition Methods(Cont.)

To realize the cross-corpus speech emotion recognition, some efforts have been taken in recent years.

- Schuller et al. conduct preliminary cross-corpus experiments on six different datasets (2011)
- Deng et al. present an autoencoder-based unsupervised domain adaptation method (2014)
- We introduce a dimension reduction based transfer learning approach (2014)
- …

Weakness:

- ■ Most of these methods do not take into account the different distributions of different corpora, and the difference is always very large.
- ■ Our previous dimension reduction based transfer learning algorithm conducts transfer learning or dimension reduction separately.

PART2: Baseline NMF algorithm

NMF

- NMF (non-negative matrix factorization) is a well-known algorithm that can obtain a low dimensional representation of the non-negative data (Lee & Seung, 1999) . It aims at finding two non-negative matrices to well approximate the original matrix data as follows

$$\min_{U,V} \|X - UV\|_F^2, \quad \text{s.t. } U,\ V \geq 0$$

- It is a non-convex problem to optimize U and V together, and can be solved via an iterative algorithm (Lee & Seung, 2001) as

$$u_{ik} \leftarrow u_{ik} \frac{(XV)_{ik}}{(UV^TV)_{ik}}, v_{kj} \leftarrow v_{kj} \frac{(X^TU)_{kj}}{(VU^TU)_{kj}}$$

Graph NMF

- Many previous studies have demonstrated that the naturally occurring data may usually reside on or close to a low dimensional submanifold embedded in a high dimensional space, so Cai et al. (2011) present a graph NMF algorithm, which is written as

$$\min_{U,V} \|X - UV\|_F^2 + \lambda Tr(VLV^T)$$
$$\text{s.t. } U,\ V \geq 0$$

where $L = D - W$ is the graph laplacian, in which $D = [d_{jj}] \in R^{N*N}$, $d_{jj} = \sum_l w_{il}$, and

$$w_{ij} = \begin{cases} 1 & \text{if } x_j \in N_p(x_i) \text{ or } x_i \in N_p(x_j) \\ 0 & \text{otherwise} \end{cases}$$

PART3: Our proposed method

Minimizing the distribution divergence

- By using the GNMF algorithm, the latent coding vectors can be obtained for the two corpus are obtained. However, the differences between the distributions of coding vectors are still large, so the empirical maximum mean discrepancy (MMD) algorithm is employed

$$
\begin{aligned}
D(V_{src}, V_{tar}) &= \left\| \frac{1}{n_l} \sum_{i=1}^{n_l} v_i - \frac{1}{n_u} \sum_{j=n_l+1}^{n_l+n_u} v_j \right\|^2 \\
&= \sum_{i,j=1}^{N} v_i^T v_j m_{ij} \\
&= Tr(VMV^T)
\end{aligned}
$$

where $m_{ij} = \begin{cases} \frac{1}{n_l^2} & v_i, v_j \in V_{src} \\ \frac{1}{n_u^2} & v_i, v_j \in V_{tar} \\ \frac{-1}{n_l n_u} & \text{otherwise} \end{cases}$

The transfer NMF method

- By integrating the GNMF function with the MMD algorithm, the objective function of the transfer NMF can be written as

$$\min_{U,V} \|X - UV\|_F^2 + \boxed{\lambda Tr(VLV^T)} + \boxed{\gamma Tr(VMV^T)}$$
$$\text{s.t. } U,\ V \geq 0$$

- Let $T = \lambda L + \gamma M$, the above equation can be rewritten as

$$\min_{U,V} \|X - UV\|_F^2 + Tr(VTV^T)$$
$$\text{s.t. } U,\ V \geq 0$$

The transfer NMF method (Cont.)

- As NMF, the above Equation is not convex when optimizing *U* and *V* together, so the iterative algorithm is also employed, and the updating functions can be rewritten as

$$u_{ik} \leftarrow u_{ik} \frac{(XV)_{ik}}{(UVV^T)_{ik}}$$

$$v_{kj} \leftarrow v_{kj} \frac{(U^T X + VT^-)_{kj}}{(VU^T U + VT^+)_{kj}}$$

where T^+ and T^- are the positive and negative parts of T.

PART4: Experimental Results and Discussions

Experimental setup

- **Datasets:** Berlin (EMO-DB) dataset, eNTERFACE dataset

- **Strategies:**
  - **The 1st case:** the lableled Berlin dataset is chosen for training, and the unlabeled eNTERFACE dataset is used for testing.
  - **The 2nd case:** the labeled eNTERFACE dataset is chosen for training, and the unlabeled Berlin dataset is used for testing.

- **Emotion Categories:** anger, disgust, fear, happiness, sadness and surprise

- **Features:**
  - Extracted by the openSMILE toolkit
  - The 1582 dimensional feature set of Interspeech 2010 Paralinguistic challenge is adopted

Experimental setup (Cont.)

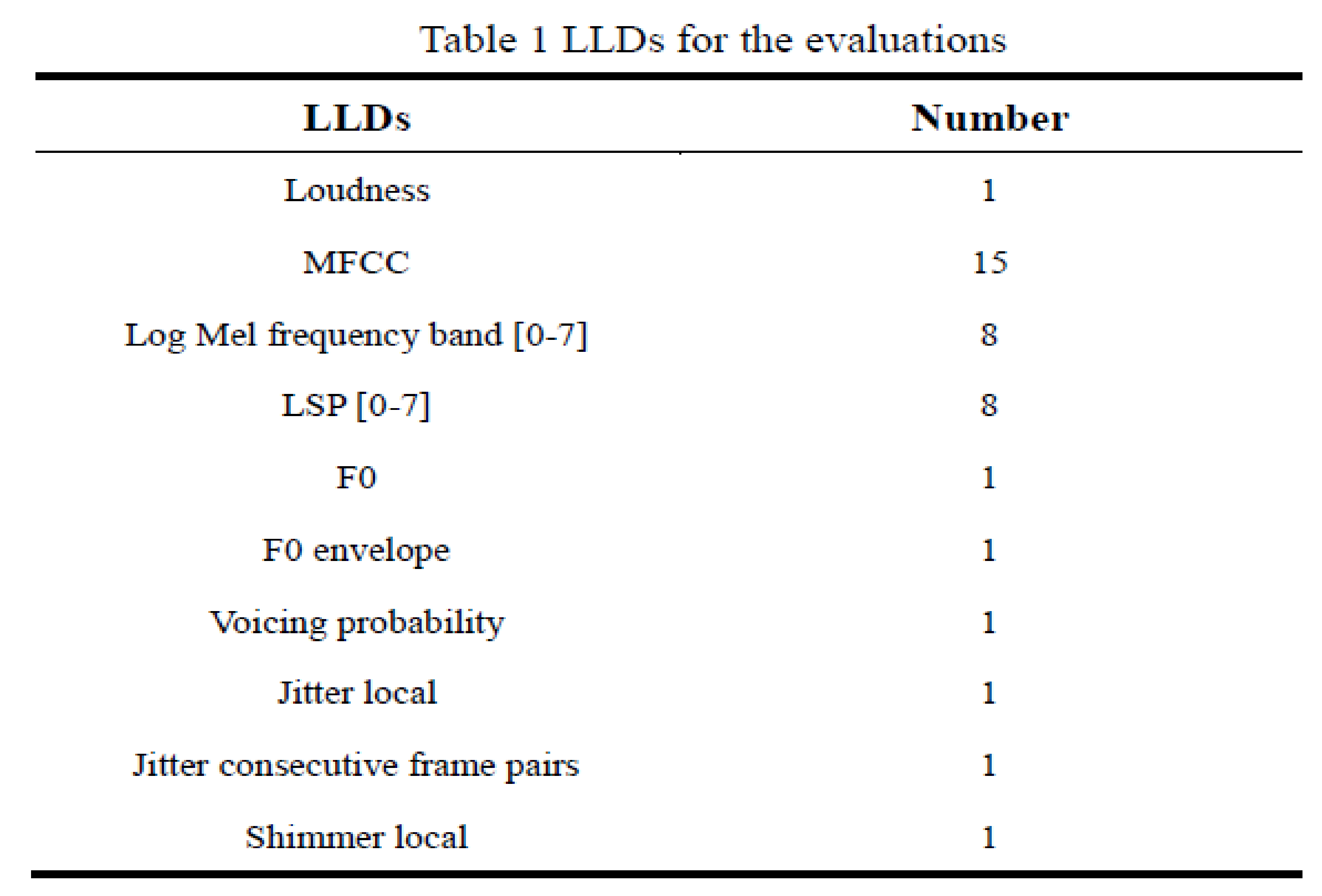
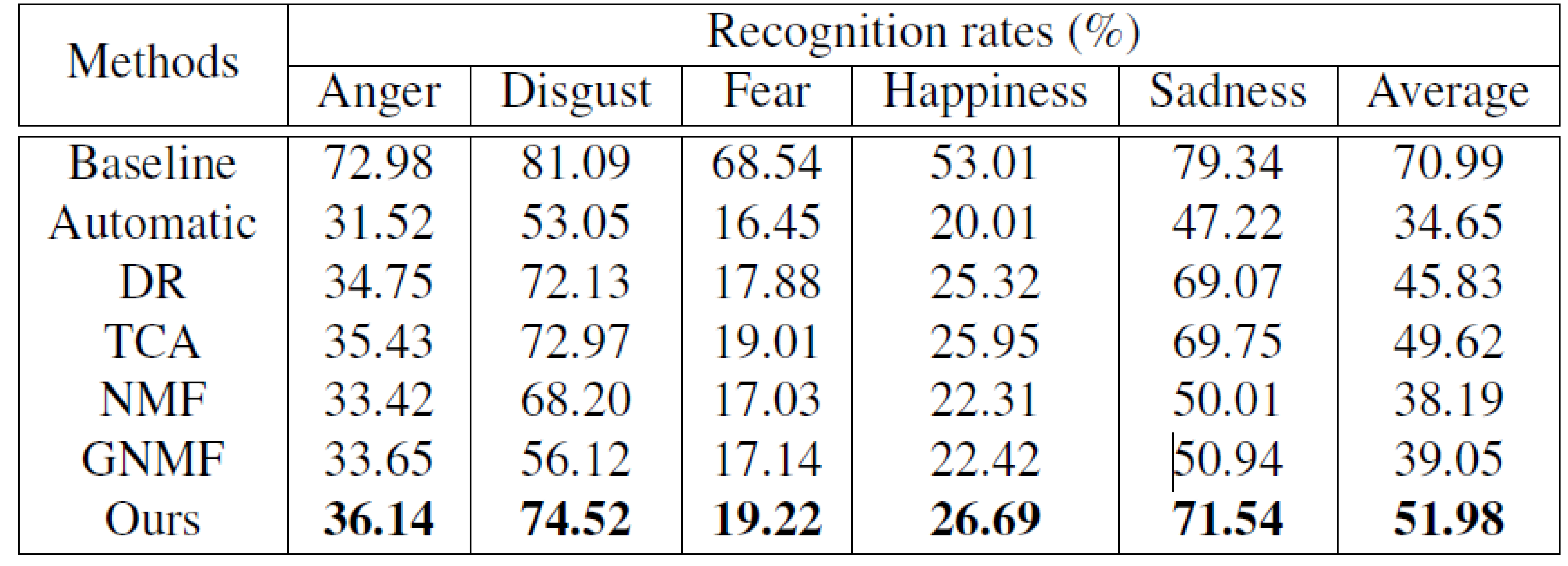


Table 1 LLDs for the evaluations

LLDs	Number
Loudness	1
MFCC	15
Log Mel frequency band [0-7]	8
LSP [0-7]	8
F0	1
F0 envelope	1
Voicing probability	1
Jitter local	1
Jitter consecutive frame pairs	1
Shimmer local	1

Experimental results

Recognition results in *case*1 (eNTERFACE dataset for training, Berlin dataset for testing)



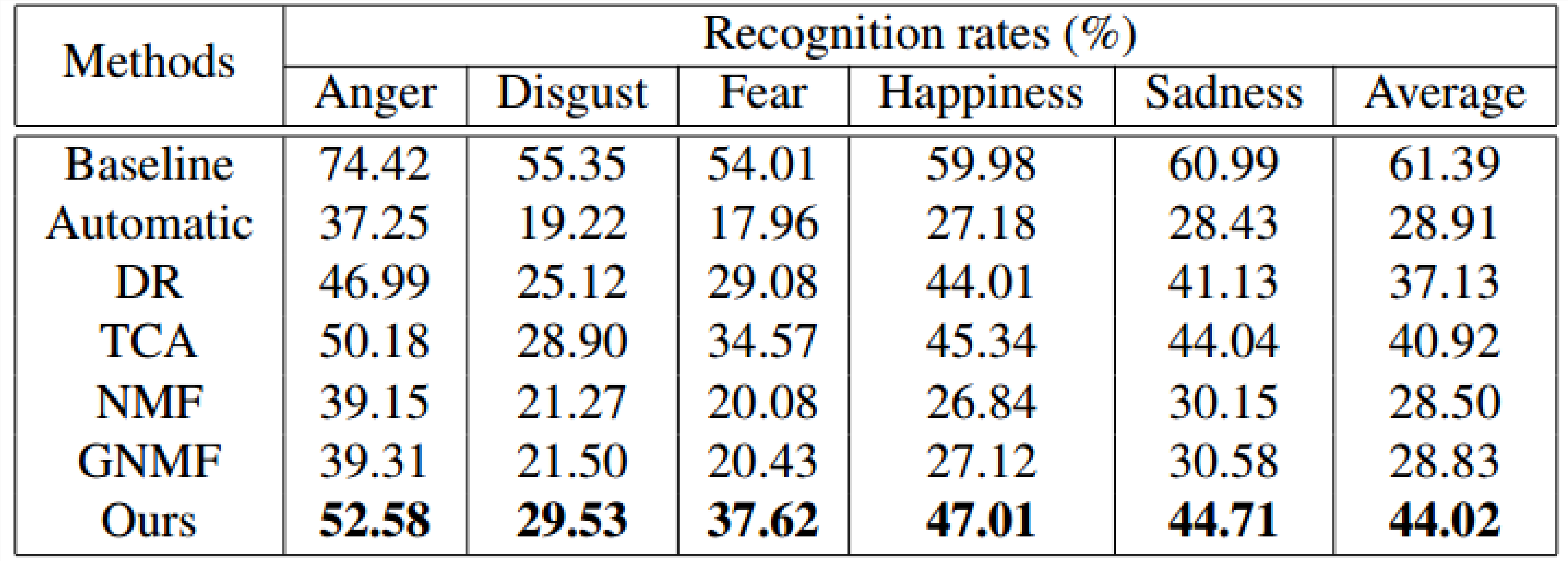
Methods	Recognition rates (%)					
	Anger	Disgust	Fear	Happiness	Sadness	Average
Baseline	72.98	81.09	68.54	53.01	79.34	70.99
Automatic	31.52	53.05	16.45	20.01	47.22	34.65
DR	34.75	72.13	17.88	25.32	69.07	45.83
TCA	35.43	72.97	19.01	25.95	69.75	49.62
NMF	33.42	68.20	17.03	22.31	50.01	38.19
GNMF	33.65	56.12	17.14	22.42	50.94	39.05
Ours	**36.14**	**74.52**	**19.22**	**26.69**	**71.54**	**51.98**

the dimension reduction based transfer learning method (DR)
the transfer component analysis method (TCA)
the NMF method (NMF)
the graph NMF method (GNMF)

the proposed TNMF method (Ours)

Experimental results (Cont.)

Recognition results in *case*2 (Berlin dataset for training, eNTERFACE dataset for testing)



Methods	Recognition rates (%)					
	Anger	Disgust	Fear	Happiness	Sadness	Average
Baseline	74.42	55.35	54.01	59.98	60.99	61.39
Automatic	37.25	19.22	17.96	27.18	28.43	28.91
DR	46.99	25.12	29.08	44.01	41.13	37.13
TCA	50.18	28.90	34.57	45.34	44.04	40.92
NMF	39.15	21.27	20.08	26.84	30.15	28.50
GNMF	39.31	21.50	20.43	27.12	30.58	28.83
Ours	**52.58**	**29.53**	**37.62**	**47.01**	**44.71**	**44.02**

the dimension reduction based transfer learning method (DR)
the transfer component analysis method (TCA)
the NMF method (NMF)
the graph NMF method (GNMF)

the proposed TNMF method (Ours)

PART5: Conclusion and Future Work

Conclusions

In this paper, a new cross-corpus speech emotion recognition method using transfer NMF is presented.

- The NMF approach is proposed for dimension reduction and feature representation
- The MMD algorithm is employed for similarity measurement
- The NMF and MMD are jointly optimized

Discussions

There still exist some problems in current method:

- The classifier is trained only using the labeled features of source dataset, without considering the unlabeled information from the target dataset
- Learning common feature representations may lessen the class discrimination of each corpus
- More datasets should be involved to evaluate the performance of our proposed method

Thank You!