Cross-corpus Speech Emotion Recognition Using Transfer Semi-supervised Discriminant Analysis

Peng Song[1], Xinran Zhang[2], Shifeng Ou[1], Jingjing Liu[2], Yanwei Yu[1], Wenming Zheng[2]

[1]Yantai University, [2]Southeast University

pengsongseu@gmail.com

2016.10, Tianjin, China

Outline

- PART1: Introduction
- PART2: Semi-supervised Discriminant Analysis (SDA) algorithm
- PART3: Our proposed method
- PART4: Experimental Results and Discussions
- PART5: Conclusion and Future Work

PART1: 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

Current 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 (2010)
- Deng et al. present an autoencoder-based unsupervised domain adaptation method (2014)
- We introduce a dimension reduction based transfer learning approach (2014, 2015)
- Some adaptation algorithms popuplar in speech/speaker recognition fields
- …

Weakness:

- Most of these methods do not take into account the different distributions of different corpora, and this difference is always very large.
- Our previous dimension reduction based transfer learning algorithms are unsupervised learning algorithms, where the labels are not efficiently utilized.

PART2: Semi-supervised linear discriminant analysis

LDA

Given a set of feature samples $X = [x_1, \ldots, x_n]$, the goal of LDA is

$$\underset{\mathbf{a}}{\operatorname{argmax}} \frac{tr(\mathbf{a}^T S_B \mathbf{a})}{tr(\mathbf{a}^T S_W \mathbf{a})}$$

where $\mathbf{a}$ is the projection vector, $tr(\cdot)$ refers to the trace of a matrix, and S_B and S_W are the between-class and within-class covariance matrices, respectively,

$$S_B = \sum_{i=1}^{c} n_i \left(\mu^{(i)} - \mu\right)\left(\mu^{(i)} - \mu\right)^T$$

$$S_W = \sum_{i=1}^{c} \left(\sum_{k=1}^{n_i} (x_k^{(i)} - \mu^{(i)})(x_k^{(i)} - \mu^{(i)})^T \right)$$

SDA

In SDA (Cai et al. ICCV 2007), the labeled samples are used to maximize the discriminative power between classes, while the unlabeled samples are used to preserve the intrinsic geometric data structure. The objective function can be written as

$$\underset{\mathbf{a}}{\operatorname{argmax}} \frac{tr(\mathbf{a}^T S_B \mathbf{a})}{tr(\mathbf{a}^T (S_W + \alpha X L X^T)\mathbf{a})}$$

where $L = D - W$ is the graph Laplacian, $D = \text{diag}(d_1, \dots, d_n)$, in which d_i is the degree of x_i satisfying $d_i = \sum_{j=1}^{n} w_{ij}$, $W = [w_{ij}]_{i,j=1}^{n}$ is the weight matrix of the graph.

PART3: Our proposed method

Minimizing the distribution divergence

- By using the SDA algorithm, the low dimensional feature representations for the two corpora are obtained. However, the differences between two datasets are still large, so the empirical maximum mean discrepancy (MMD) algorithm is employed for similarity measurement

$$D = \left\| \frac{1}{n_l} \sum_{i=1}^{n_l} \mathbf{a}^T x_i - \frac{1}{n_u} \sum_{j=n_l+1}^{n} \mathbf{a}^T x_j \right\| = tr(\mathbf{a}^T X M X^T \mathbf{a})$$

where $M = [m_{ij}]_{i,j=1}^{N}$ denotes the MMD matrix, and is given by $m_{ij} = \begin{cases} \frac{1}{n_l^2} & v_i, v_j \in V_{src} \\ \frac{1}{n_u^2} & v_i, v_j \in V_{tar} \\ \frac{-1}{n_l n_u} & otherwise \end{cases}$

The transfer SDA (TSDA) method

By regularizing SDA with MMD function, the projection matrix $\mathbf{a}$ is refined, and the distributions of labeled source and unlabeled target data are drawn close under the new feature representations $\mathbf{a}^T X$. The objective function of TSDA is

$$\underset{\mathbf{a}}{\operatorname{argmin}} \frac{tr(\mathbf{a}^T(S_W+\alpha XLX^T)\mathbf{a})}{tr(\mathbf{a}^T S_B \mathbf{a})} + \beta tr(\mathbf{a}^T XMX^T \mathbf{a})$$

$$s.t.\ \ \mathbf{a}^T\mathbf{a} = 1$$

Optimization

The objective function of TSDA is non-linear and it is important to obtain a global optimization solution. The trace ratio of TSDA can be relaxed in a different form as

$$\underset{\mathbf{a}}{\operatorname{argmin}} J(\mathbf{a}) = tr(\mathbf{a}^T(S_W + \alpha XLX^T)\mathbf{a}) - tr(\mathbf{a}^T S_B \mathbf{a}) + \beta tr(\mathbf{a}^T XMX^T \mathbf{a})$$

$$s.t.\ \ \mathbf{a}^T\mathbf{a} = 1$$

By employing the zero gradient condition $\frac{\partial J(\mathbf{a})}{\partial \mathbf{a}} = 0$, the projection matrix $\mathbf{a}$ is given by solving the following generalized eigenvalue problem as

$$(S_W - S_B + G)\mathbf{a} = \lambda \mathbf{a}$$

where $G = X(\alpha L + \beta M)X^T$.

PART4: Experimental Results and Discussions

Experimental setup

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

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

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

- 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

Table 2: The average recognition rates of different methods in *case1*

Methods	Recognition rates (%)					
	Anger	Disgust	Fear	Happiness	Sadness	Average
Automatic	31.51	53.04	16.43	20.02	47.24	34.66
TCA	35.42	73.01	19.05	25.97	69.72	49.62
TNMF	36.15	74.49	19.23	26.58	71.55	51.97
SDA	34.02	58.26	17.36	23.72	57.36	42.13
TLDA	36.20	74.36	18.86	26.14	70.12	50.86
TSDA	**36.28**	**74.69**	**20.01**	**27.12**	**71.63**	**52.25**

Automatic: The classifier trained in source corpus is directly applied to target corpus.

TCA: Transfer component analysis method, one of the typical transfer learning algorithms.

TNMF: Transfer non-negative matrix factorization method.

SDA: Semi-supervised linear discriminant analysis.

TLDA: Transfer linear discriminant analysis, in which α is set to 0.

TSDA: Our proposed transfer SDA method.

Experimental results (cont.)

Table 3: The average recognition rates of different methods in *case2*

Methods	Recognition rates (%)					
	Anger	Disgust	Fear	Happiness	Sadness	Average
Automatic	37.24	19.23	17.97	27.16	28.42	28.92
TCA	50.19	28.89	34.56	45.35	44.03	40.91
TNMF	52.56	29.52	37.63	47.02	44.72	44.03
SDA	47.28	25.82	31.15	45.18	41.87	38.25
TLDA	52.48	29.53	37.59	47.04	45.12	43.96
TSDA	**53.01**	**29.86**	**38.25**	**47.32**	**45.26**	**44.89**

Automatic: The classifier trained in source corpus is directly applied to target corpus.

TCA: Transfer component analysis method, one of the typical transfer learning algorithms.

TNMF: Transfer non-negative matrix factorization method.

SDA: Semi-supervised linear discriminant analysis.

TLDA: Transfer linear discriminant analysis, in which α is set to 0.

TSDA: Our proposed transfer SDA method.

PART5: Conclusion and Future Work

Conclusions

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

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

Discussions

There still exist some problems in current method:

- 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
- How to avoid the negative transfer

THANK YOU!