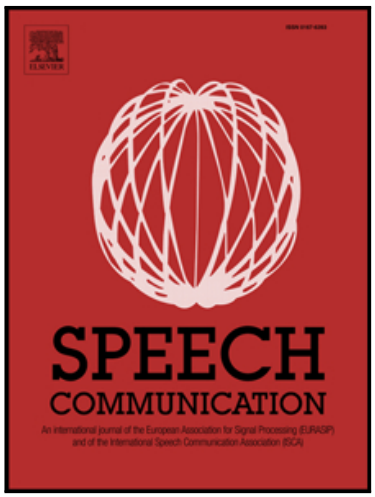
Accepted Manuscript



Learning Emotion-discriminative and Domain-invariant Features for Domain Adaptation in Speech Emotion Recognition

Qirong Mao, Guopeng Xu, Wentao Xue, Jianping Gou, Yongzhao Zhan

PII: S0167-6393(16)30192-3
DOI: 10.1016/j.specom.2017.06.006
Reference: SPECOM 2469

To appear in: *Speech Communication*

Received date: 26 July 2016
Revised date: 20 May 2017
Accepted date: 21 June 2017

Please cite this article as: Qirong Mao, Guopeng Xu, Wentao Xue, Jianping Gou, Yongzhao Zhan, Learning Emotion-discriminative and Domain-invariant Features for Domain Adaptation in Speech Emotion Recognition, *Speech Communication* (2017), doi: 10.1016/j.specom.2017.06.006

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

LEARNING EMOTION-DISCRIMINATIVE AND DOMAIN-INVARIANT FEATURES FOR DOMAIN ADAPTATION IN SPEECH EMOTION RECOGNITION

Qirong Mao, Guopeng Xu, Wentao Xue, Jianping Gou and Yongzhao Zhan

Department of Computer Science and Communication Engineering, Jiangsu University, China

ABSTRACT

Conventional approaches for Speech Emotion Recognition (SER) usually assume that the feature distributions between training and test set are identical. However, this assumption does not hold in many real scenarios. Although many Domain Adaptation (DA) methods have been proposed to solve this problem, the conventional emotion discriminative information is ignored. In this paper, we propose a DA based method called Emotion-discriminative and Domain-invariant Feature Learning Method (EDFLM) for SER, in which both the domain divergence and emotion discrimination are considered to learn emotion-discriminative and domain-invariant features by using emotion label constraint and domain label constraint. Furthermore, to disentangle the emotion-related factors from the emotion-unrelated factors, we introduce an orthogonal term to encourage the input to be disentangled into two blocks: emotion-related and emotion-unrelated features. Our method can learn emotion-discriminative and domain-invariant features through a back propagation network which uses the acoustic features of INTERSPEECH 2009 Emotion Challenge as the input rather than raw speech signals. Experiments on the INTERSPEECH 2009 Emotion Challenge two-class task show that the performance of our method is superior to other state-of-the-arts methods.

***Index Terms*—** Domain adaptation, speech emotion recognition, neural network

1. INTRODUCTION

The problem of automatically predicting the emotional states in speech emotion recognition has been the subject of increasing attention among the speech community. Many conventional state-of-the-art speech emotion recognition methods usually assume that the features of the training and test samples are drawn from the same distribution. This assumption does not hold in many real world applications. This is mainly because the speech signals from different domains are highly dissimilar in terms of speakers, type of emotion, recording situation and degree of spontaneity. A classifier just trained on a specific corpus and then applied directly to another corpus cannot be expected to have excellent performance.

Domain adaptation (DA), proposed by Daumé III [11], has proven to be efficient for this problem. DA is one special type of transfer learning problem. The feature distributions of samples in source and target domains are different, but the tasks of the source and the target remain the same [11, 27]. Based on whether the target domain data is partially unlabeled or completely unlabeled, DA techniques are commonly classified into two categories: semi-supervised DA and unsupervised DA. Unsupervised DA is more challenging and is more in line with the practical situations. So in this paper, we mainly deal with unsupervised DA for Speech Emotion Recognition (SER).

Recently, deep leaning has achieved state-of-the-art performance on many machine learning tasks [2]. The success of deep learning mainly contributes to the ability of extracting abstract hierarchical non-linear features of the input [29, 4, 9]. Meanwhile, deep learning has shown to suit well to DA [3].

Although previous deep-learning based DA methods aim to learn a more powerful feature representation to reduce the discrepancy between the source and target domains, most of them do not take the label information into account in training time. For SER, we are eager to learn an emotion-discriminative and domain-invariant feature representation. So, to meet the above two requirements at the same time, we should find a tradeoff between them. Thus seeking for a saddle point of the loss function is an urgent demand. Ganin et al. [19] propose a feasible approach to find this saddle point by introducing a gradient reversal layer (GRL). It is simple to implement such a layer in any feed-forward models, and the parameters update can be done using the standard stochastic gradient descent (SGD).

In SER, conventional methods aim to learn the emotion-specific features that are robust to the nuisance factors, such as speaker variation and environment distortion [15, 26]. So we really hope to get such a powerful feature representation in the DA method for SER. Just using the emotion label predictor cannot achieve an excellent emotion-discriminative feature representation.

This work is supported by the National Natural Science Foundation of China (No. 61272211, No.61672267 and No.61502208), the Six Talent Peaks Foundation of Jiangsu Province (No.DZXX-027), the Open Project Program of the National Laboratory of Pattern Recognition (NLPR, No.201700022) and the general Financial Grant from the China Postdoctoral Science Foundation (No. 2015M570413).

ACCEPTED MANUSCRIPT

Download English Version:

https://daneshyari.com/en/article/4977791

Download Persian Version:

https://daneshyari.com/article/4977791

Daneshyari.com