计算机应用
Journal of Computer Applications
ISSN 1001-9081,CN 51-1307/TP

《计算机应用》网络首发论文

题目：基于决策边界优化域自适应的跨库语音情感识别
作者：汪洋，傅洪亮，陶华伟，杨静，谢跃，赵力
收稿日期：2021-12-08
网络首发日期：2022-06-15
引用格式：汪洋，傅洪亮，陶华伟，杨静，谢跃，赵力．基于决策边界优化域自适应的跨库语音情感识别[J/OL]．计算机应用．
https://kns.cnki.net/kcms/detail/51.1307.TP.20220613.1726.004.html

Journal of Computer Applications　　ISSN 1001-9081
计算机应用　　CODEN JYIIDU　　http://www.joca.cn

doi:10.11772/j.issn.1001-9081.2021122043

基于决策边界优化域自适应的跨库语音情感识别

汪洋[1]，傅洪亮[1]，陶华伟[1*]，杨静[1]，谢跃[2]，赵力[3]

(1.河南工业大学 粮食信息处理与控制教育部重点实验室，郑州 450001；　2. 南京工程学院 信息与通信工程学院，南京 211167；
3.东南大学 信息科学与工程学院，南京 210018)

(*通信作者电子邮箱 thw@haut.edu.cn)

摘 要: 域自适应算法被广泛应用于跨库语音情感识别中。然而，许多域自适应算法在追求减小域差异的同时，丧失了目标域样本的鉴别性，导致其以高密度的形式存在于模型决策边界处，降低了模型的性能。基于此，提出一种基于决策边界优化域自适应（decision boundary optimized domain adaptation, DBODA）的跨库语音情感识别方法。首先利用卷积神经网络进行特征处理，随后将特征送入最大化核范数及均值差异（maximum n-norm and mean discrepancy, MNMD）模块，在减小域间差异的同时，最大化目标域情感预测概率矩阵的核范数，以提升目标域样本的鉴别性，优化决策边界。在以 Berlin, eNTERFACE, CASIA 语音库为基准库设立的六组跨库实验中，所提方法的平均识别精度领先于其他算法 1.28%~11.01%，说明模型有效降低了决策边界的样本密度，提升了预测的准确性。

关键词： 跨库语音情感识别；卷积神经网络；决策边界优化；域自适应；特征分布差异

中图分类号： TP391.4　　**文献标志码：** A

Cross-corpus speech emotion recognition based on decision boundary optimized domain adaptation

WANG Yang[1], FU Hongliang[1], TAO Huawei[1*], YANG Jing[1], XIE Yue[2], ZHAO Li[3]

(1. Key Laboratory of Grain Information Processing and Control (Henan University of Technology), Ministry of Education, Zhengzhou Henan 450001, China;
2. School of Information and Communication Engineering, Nanjing Institute Technology, Nanjing Jiangsu 211167, China;
3. School of Information Science and Engineering, Southeast University, Nanjing Jiangsu 210018, China)

Abstract: Domain adaptation algorithms are widely used for cross-corpus speech emotion recognition. However, many domain adaptation algorithms lose the discrimination of target domain samples while pursuing the minimization of domain discrepancy, resulting in their presence at the decision boundary of the model in a dense form, which degrades the performance of the model. Based on the above problem, a decision boundary optimized domain adaptation (DBODA) based cross-corpus speech emotion recognition method is proposed. The features are firstly processed using convolutional neural networks, and then fed into the maximize n-norm and mean discrepancy (MNMD) module to maximize the nuclear-norm of the sentiment prediction probability matrix of the target domain while reducing the inter-domain discrepancy to enhance the discrimination of the target domain samples and optimize the decision boundary. In six sets of cross-corpus experiments set up with Berlin, eNTERFACE, and CASIA speech datasets, the average recognition accuracy of the proposed method is 1.28%~11.01% ahead of other algorithms, indicating that the model effectively reduces the sample density around the decision boundary and improves the prediction accuracy.

Keywords: cross-corpus speech emotion recognition; convolutional neural network; decision boundary optimization; domain adaptation; feature distribution discrepancy

0　引言

情感是人类智能的重要组成部分，赋予计算机从人类的语音信号中识别情感状态的能力，是当前人工智能、模式识别、认知科学等领域的研究热点[1]。目前大多数语音情感识别

收稿日期: 2021-12-08;　**修回日期:** 2022-04-27;　**录用日期:** 2022-05-11。
基金项目: 国家自然科学基金资助项目(62001215)；河南省教育厅自然科学项目(21A120003);河南工业大学高层次人才启动项目(2018BS037)
作者简介: 汪洋(1999-)，男，河南信阳人，硕士研究生，CCF 学生会员（H0653G），主要研究方向为语音信号处理；傅洪亮(1965-)，男，河南安阳人，教授，硕导，博士，主要研究方向为通信与信息系统；陶华伟(1987-)，男（通信作者），河南郑州人，讲师，博士，CCF 会员，主要研究方向为语音情感识别(thw@haut.edu.cn)；杨静(1983-)，女，河南商丘人，副教授，硕导，博士，主要研究方向为通信信号处理；谢跃(1991-)，男，博士，主要研究方向为人工智能、情感计算；赵力(1958-)，男，江苏南京人，教授，博导，博士，主要研究方向为语音信号处理、情感信息处理.

方法都是在单一语音库上进行，然而在许多实际应用中，测试语音数据的语种、发音风格、录制环境等，往往与训练语音数据存在极大的差异，导致训练过的模型在测试数据上识别性能下降[2]，这是典型的跨库语音情感识别问题。因此，开发更具鲁棒性的，能够更好适应测试数据变化的语音情感识别系统至关重要。

近年来，研究者们从特征处理以及特征分布对齐角度，提出了许多跨库语音情感识别算法, Zhang W 等[3]提出一种迁移稀疏判别子空间学习（Transfer Sparse Discriminant Subspace Learning, TSDSL）方法，引入判别性学习和范数惩罚，学习不同语音库间的域不变特征，并利用最近邻图以减小域间差异。Luo 等[4]介绍了一种基于非负矩阵分解（Non-Negative Matrix Factorization, NMF）的跨库语音情感识别方法，使用最大均值差异（MMD）同时最小化两个语料库的边际分布和条件分布差异。Zhang J 等[5]提出了一种联合分布自适应回归（Joint Distribution Adaptive Regression, JDAR）方法，联合考虑训练和测试语音数据之间的边际和条件概率分布来学习回归矩阵，减轻不同库之间的特征分布偏差。随着深度学习的发展，相关方法被提出用于学习源域和目标域间的可鉴别特征。Deng 等[6]提出了半监督自编码器进行共性情感特征学习，以提升跨库语音情感识别性能。Gideon 等[7]使用对抗域自适应的方法，让模型在不同数据集中学到的表征相近，提高模型的泛化能力。Lee[8]提出一个基于三联体网络的新框架来学习跨多个语料库的更广义的特征。Mohammed 等[9]使用对抗性多任务训练来提取训练域和测试域之间的共同表示。Liu 等[10]基于深度卷积神经网络的特征提取模型和最大均值差异（MMD）算法提取更具鲁棒性的语音特征，以获得更好的跨语料库识别性能。上述方法虽取得了一定的效果，但仍存在部分问题。在传统降维方法中，对于情感变化缓慢的语音信号，该类方法易于丢失情感信息，而深度域自适应方法则会导致无标签的目标域语音库样本可鉴别性降低，致使模型决策边界数据密度大，降低识别性能。

通过对以上问题的分析，本文提出了一种基于决策边界优化域自适应（DBODA）的跨库语音情感识别方法。首先，在特征处理阶段，使用一维卷积神经网络（1D-CNN）作为特征处理网络，在保留特征原有情感信息的同时，深入挖掘相邻情感特征之间的潜在相关性，提升特征表征能力；其次，提出一种基于最大化核范数及均值差异（MNMD）的域自适应算法，在减小域间差异的同时，可以有效的缓解深度域自适应方法面临的决策边界数据密度较大的问题，增强无标签数据的可鉴别性，继而提升跨库语音情感识别性能。

1 基于决策边界优化域自适应的跨库语音情感识别

1.1 跨库语音情感识别模型

基于决策边界优化域自适应（Decision Boundary Optimized Domain Adaptation, DBODA）的跨库语音情感识别模型整体框图如图 1 所示。使用卷积神经网络进行特征处理，经过 softmax 层获得样本属于各个类别的概率，利用源域分类损失反向传播训练模型。为了让模型从源域迁移到目标域，减小域间差异，将经过卷积神经网络处理的源域特征和目标域特征，送入最大化核范数及均值差异（MNMD）模块，执行特征分布对齐操作，最后利用源域分类损失和特征分布对齐损失联合回传，对模型进行优化，在 1.2 节和 1.3 节对特征处理和最大化核范数及均值差异进行详细介绍。

1.2 特征处理

现有研究[11,12] 显示，相比于传统降维方法或 DNN，卷积神经网络在保留特征原有情感信息的同时，能够有效提升特征表征能力，因此本文采用一维卷积神经网络对语音特征进行处理，网络模型如图 2 所示。

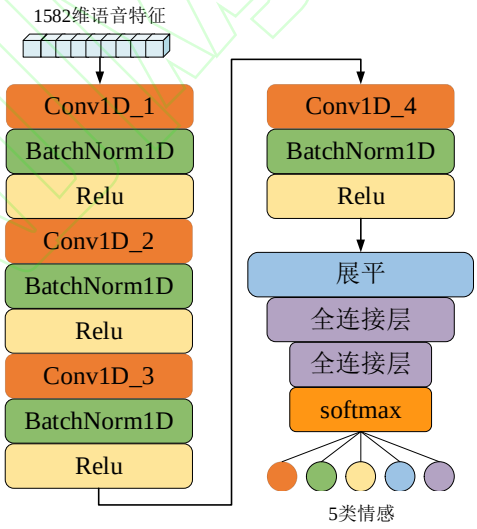


图 2 一维卷积神经网络框架图

Fig.2 Framework diagram of 1D-CNN

使用四个一维卷积层构成前端特征处理网络，由于降采样层的使用会存在特征信息丢失的风险，因此，仅在每个一维卷积层之后加入 BatchNorm 层，将源域语音库和目标域语音库的特征分布归一化，防止网络过拟合的同时，能够提升特征表征的泛化性，使用 ReLU 激活函数，在简单的网络结构设置下进一步提升特征处理速度。与传统降维方法或 DNN 相比，卷积神经网络对全局特征进行处理，且单个卷积层上的多卷积核，提取了多个局部表示，深入挖掘相邻特征间的关联性，更好的保留了情感信息。网络中各层的参数如表 1 所示；经全连接层将特征维度映射为情感类别后，应用 softmax 层将五类情感的预测输出为[0, 1]的概率，将源域的分类结果与标签做交叉熵，得到源域的分类损失为：

$$L_{cls} = -\frac{1}{B}\sum_{i=1}^{B}\sum_{c=1}^{5} y_{ic}\log\left(\hat{y}_{ic}\right) \qquad (1)$$

其中 B 表示训练过程中的批次大小，y_{ic} 取值为 1 或 0，当样

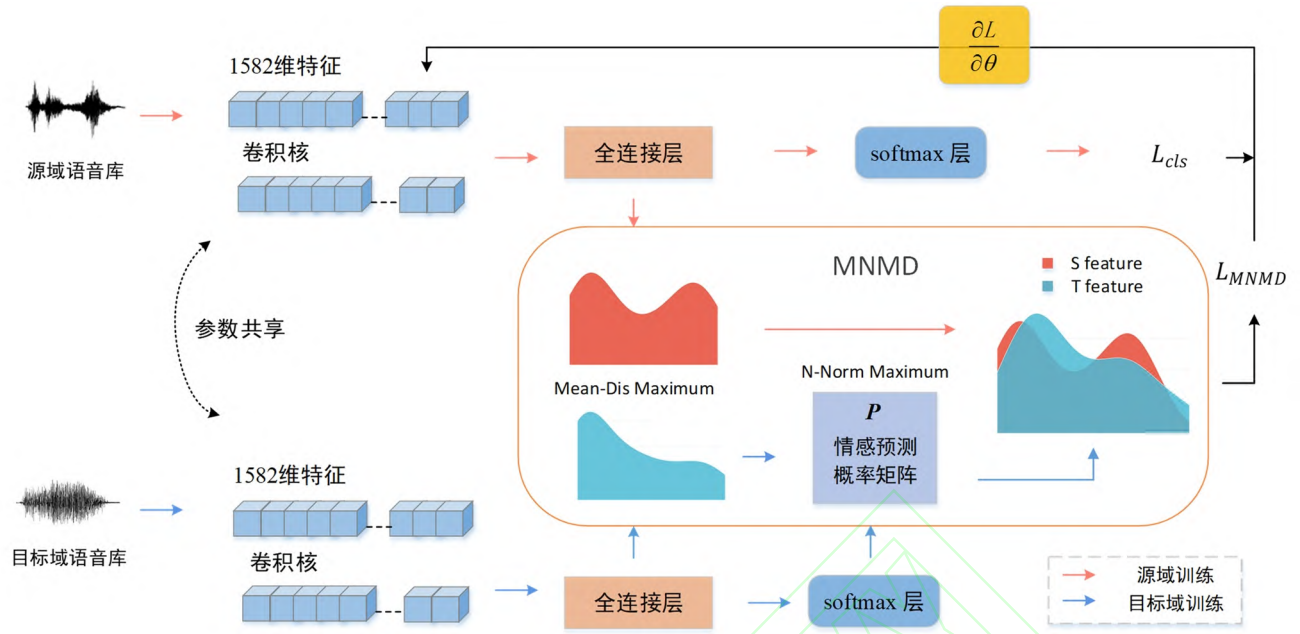


图 1　基于决策边界优化域自适应的跨库语音情感识别框架

Fig. 1 Cross-corpus speech emotion recognition based on decision boundary optimized domain adaptation

本属于第 c 类情感类型则 y_{ic} 取 1，否则取 0，$\hat{y}_{ic}$ 表示样本属于第 c 类情感类型的预测概率。

表 1　1D-CNN 模型参数

Tab. 1　Model parameters of 1D-CNN

网络层	卷积核 n×k×s	输出尺寸 b×(n)×f
Conv1D	16×9×2	16×16×791
Conv1D	32×9×2	16×32×396
Conv1D	64×9×2	16×64×198
Conv1D	128×9×2	16×128×99
展平	—	16×12 672
全连接层	—	16×2048
全连接层	—	16×5
softmax	分类器	16×5
n:卷积核数　k:卷积核尺寸　s:步长　b:批次　f:特征维度		

1.3　最大化核范数及均值差异

经过有效的特征处理，跨库语音情感识别仍面临一个核心问题，即减小源域语音库和目标域语音库间的特征分布差异，在相关研究[4,5,10]中，最大均值差异（MMD）方法已被广泛用于域间差异度量，将源域和目标域特征映射到样本空间上的连续函数，求两个特征分布映射后的函数值均值，作差得到两个分布对应函数的均值差异，可表达为如下形式：

$$MMD^2=\left\|\frac{1}{n}\sum_{i=1}^{n}\phi(D_s)-\frac{1}{m}\sum_{j=1}^{m}\phi(D_t)\right\|_H^2 \tag{2}$$

其中 H 为再生核希尔伯特空间，其中 ϕ 为映射函数，D_s 代表源域的特征分布，D_t 代表目标域的特征分布。

然而最新研究[14]表明在利用 MMD 进行域级特征分布对齐时，会使得特征一般化，丢失类间特性，大量的目标域样本在经过特征分布对齐后，聚集在模型的决策边界上，导致目标域特征的可鉴别性下降。为了提升目标域特征的鉴别性，受批核范数最大化[15]工作的启发，本文提出了最大化核范数及均值差异（MNMD），改进后的损失函数可以表示为：

$$L_{MNMD}=\left\|\frac{1}{n}\sum_{i=1}^{n}\phi(\min F(x_s))-\frac{1}{m}\sum_{j=1}^{m}\phi\left(\max\|F(x_t)\|_*\right)\right\|_H^2 \tag{3}$$

其中 F 表示特征处理网络，x_s 和 x_t 分别表示源域和目标域样本特征，$\|\cdot\|_*$ 代表求解核范数。

将两个域的特征映射函数 ϕ 相减，得到的特征分布差异回传，促进模型从源域迁移至目标域。在此过程中，对于有标签的源域语音库，最小化其分类损失，以优化特征处理网络。在无标签的目标域语音库上，最大化情感预测概率矩阵的核范数，优化模型决策边界。目标域情感预测概率矩阵 $\boldsymbol{P}\in\mathbb{R}^{B\times C}$ 可以表示为如下形式：

$$\boldsymbol{P}_{i,j}\in[0,1]\quad \forall i\in 1\mathbf{:}B, j\in 1\mathbf{:}C \tag{4}$$

其中 $\boldsymbol{P}_{i,j}$ 为将样本 i 预测为情感类型 j 的概率，B 为批次大小，C 为情感类别数，B 和 C 也分别代表了预测概率矩阵 $\boldsymbol{P}$ 的行数和列数。

MNMD 通过最大化 $\boldsymbol{P}$ 的核范数，约束其 Frobenius 范数[16]，以使得其香农熵减小，消除预测不确定性，提升目标域样本可鉴别性，其约束关系如下：

$$\frac{\|\boldsymbol{P}\|}{\sqrt{\min(B,C)}}\le\|\boldsymbol{P}\|_F\le\|\boldsymbol{P}\|_*$$
$$\max\|\boldsymbol{P}\|_F\Leftrightarrow\min H(\boldsymbol{P})\quad \text{s.t. } \boldsymbol{P}_{i,j}\to 0 \text{ or } \boldsymbol{P}_{i,j}\to 1 \tag{5}$$

其中 $\|\boldsymbol{P}\|_*,\|\boldsymbol{P}\|_F,H(\boldsymbol{P})$ 分别表示情感预测概率矩阵的核范数、Frobenius 范数和香农熵。最大化核范数时，可以降低香农熵，

使得情感预测概率 $P_{i,j}$ 趋近于 0 或 1 时，则预测的不确定性下降，模型决策边界得到优化。

此外，MNMD 能够在提升目标域情感特征鉴别性的同时保证预测的多样性，情感预测概率矩阵的秩可以近似为其预测类别数，其核范数为矩阵秩的凸包络[15]，则最大化其核范数可以有效的保证情感预测的多样性，避免了熵最小化导致的模型优化偏移。因此 MNMD 很好地缓解了模型从源域语音库迁移到目标域语音库过程中，低鉴别性的目标域样本高密度堆积于决策边界上的问题。

2 实验设置及结果分析

2.1 语音情感库及语音特征提取

2.1.1 语音情感库

为了评估所提模型的性能，选用 Berlin 语音情感库[17]，eNTERFACE 语音情感库[18]和 CASIA 汉语语音情感库[19]进行了大量的实验。Berlin 库是由柏林工业大学录制的德语情感语音库，也是语音情感识别中使用最为广泛的语音库之一，由 10 位演员对 10 个语句进行 7 种情感的模拟得到，经过听辨测试后保留了 535 条最为有效的语音。eNTERFACE 库是一个视听情感数据集，包含 6 种情感，由来自 14 个国家的 42 位受试者用英语进行录制，共有 1 287 条语音。CASIA 汉语情感语料库由中国科学院自动化所录制，共包括四个专业发音人，1 200 条公开语音，6 种情感。

2.1.2 语音特征提取

参考现有研究[3,5]的实验设置，选取 IS10 情感挑战赛的规定特征集[20]作为模型输入，其中共有 1 582 维特征，包含 34 个基本的低级描述符（low-level descriptors，LLDs），即梅尔频率倒谱系数（Mel-frequency cepstral coefficients，MFCC），线谱对（line spectrum pair，LSP）和 34 个相应的 delta 系数，基于这些低级描述符，应用 21 个统计函数得到 1 428 维特征，此外，对 4 个基于音高的低级描述符与其相应的 delta 系数，应用 19 个统计函数，得到 152 维特征，将音高的开始与持续时间作为最后两个特征，构成 1 582 维语音特征。为了保持和其他研究者的一致性以及实验的可复现性，本文使用 openSMILE 开源工具[21]对原始语音进行特征提取。

2.2 实验设置及评价指标选取

实验根据三个语音情感库设计了六组跨库语音情感识别任务，每组跨库语音情感识别任务选取训练语音库和测试语音库的共同情感进行评估，具体任务设置如表 2 所示。

在六个任务中，将 e2B、B2e、C2e、e2C、C2B、B2C 的学习率和 batchsize 分别设置为{0.001, 0.01, 0.01, 0.01, 0.01, 0.001}与{16,16,16,16,16,16}，迭代轮次设置为 2000 轮。采用非加权平均召回率（UAR）作为评价指标，对不同模型的识别效果进行评估。

表 2 跨库语音情感识别任务设置

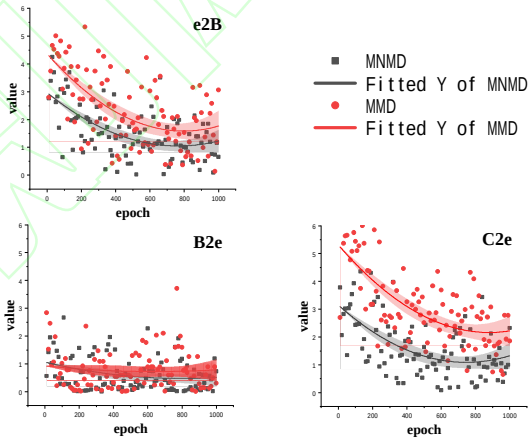
Tab. 2 Cross-corpus speech emotion recognition task settings

源域	目标域	共有情感类型
eNTERFACE(e)	Berlin(B)	愤怒、厌恶、恐
Berlin(B)	eNTERFACE(e)	惧、快乐、悲伤
CASIA(C)	eNTERFACE(e)	愤怒、恐惧、快
eNTERFACE(e)	CASIA(C)	乐、悲伤、惊讶
CASIA(C)	Berlin(B)	愤怒、恐惧、快
Berlin(B)	CASIA(C)	乐、中立、悲伤

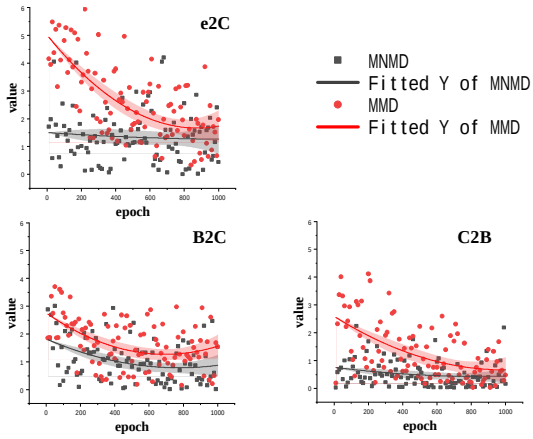
2.3 实验结果及分析

2.3.1 香农熵验证实验

为了验证模型是否降低了预测概率矩阵的香农熵，有效提升预测样本鉴别性，将 MMD 和 MNMD 在六组跨库识别任务中迭代训练 1 000 轮的熵值变化绘制如图 3。



（a）香农熵在 e2B，B2e，C2e 任务中的迭代变化



（b）香农熵在 e2C,B2C,C2B 任务中的迭代变化

图 3 MMD 和 MNMD 迭代训练香农熵变化比较

Fig.3 Comparison of the Shannon entropy of MMD and MNMD

图中阴影部分和曲线分别表示表示熵值变化的 95%置信区间与其拟合曲线。从图中可以看出，在六组跨库识别任

务中，相较于 MMD，MNMD 都有效的降低了预测概率矩阵的香农熵，特别是在 e2B、C2e 和 B2C 任务中，极大地提升了目标域样本的鉴别性，降低了预测的不确定度，证实了最大化核范数能有效缓解决策边界目标域样本密度高的问题。

2.3.2 消融实验

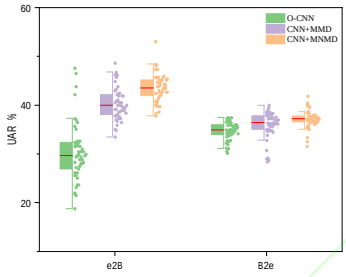
为了进一步验证模型的合理性，以清晰地观察所提域自适应方法的效果和对 MMD 改进后的提升，实验设置了消融模型进行对比，分别为：

a）O-CNN（only CNN）：不使用任何域自适应手段，直接将源域训练后的模型应用于目标域。

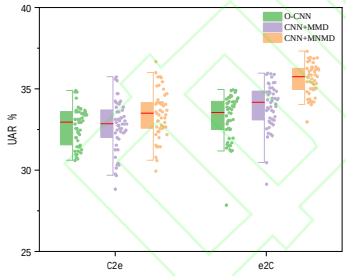
b）CNN+MMD：使用 1D-CNN 和原始的 MMD 分别进行特征提取和源域目标域的特征分布对齐。

c）CNN+MNMD：即所提模型 DBODA。

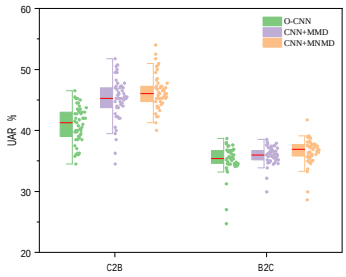
将各个模型在实验中迭代训练得到的准确率绘制成箱形图，如图 4 所示。



(a) e2B 和 B2e 任务中各模型的箱形图



(b) C2e 和 e2C 任务中各模型的箱形图

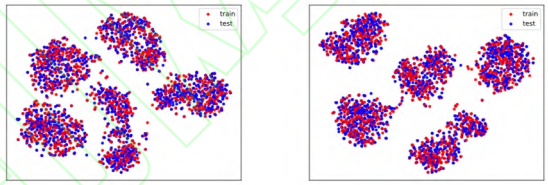


(c) C2B 和 B2C 任务中各模型的箱形图

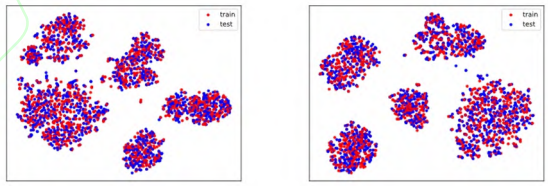
图 4 不同任务中各模型的箱形图

Fig. 4 Box diagram of each model in different tasks

通过图 4 中对各消融实验模型在不同任务中的识别率分析，可以看出，在全部的六个跨库语音情感识别任务中，相较于原始的深度学习方法，使用 MMD 减小域间差异，都获得了一定的性能提升。同时，所提 MNMD 经过对 MMD 的改进在各任务上都获得了最优的识别结果，在 e2B 任务中获得了最大的识别率均值提升，并在 e2B、B2e 和 C2B 任务中显著提升了模型识别的稳定性。将消融实验中各模型的准确率列于表 3，可以看出，所提模型的识别性能在六个跨库识别任务中均获得了最优表现，平均识别率相较于其他消融实验模型分别领先 5.42%和 4.29%，消融实验结果证实了 DBODA 的合理性，为进一步说明所提 MNMD 在对 MMD 进行优化后，能够有效降低决策边界上的数据密度，在 e2B 和 B2e 任务中，将 CNN+MMD 模型和 DBODA 模型训练后的特征使用 t-SNE 绘制为特征分布图如图 5，可以看出 DBODA 模型处理后的特征获得了更好的特征分布对齐效果，类间数据密度低，实现了对决策边界的优化。



（a）e2B 任务特征分布图



（b）B2e 任务特征分布图

图 5 e2B 和 B2e 任务中的特征分布图（左：CNN+MMD，右：DBODA）

Fig. 5 Distribution of features in tasks e2B and B2e (left: CNN+MMD, right: DBODA)

表 3 消融实验各模型的 UAR

Tab. 3 UAR of each model of ablation experiment

任务	O-CNN	CNN+MMD	DBODA
e2B	47.55	48.13	**52.99**
B2e	37.34	38.93	**41.33**
C2e	34.72	35.56	**36.67**
e2C	34.93	35.95	**37.29**
C2B	50.71	52.26	**54.25**
B2C	38.64	39.28	**43.04**
Average	38.84	39.97	**44.26**

2.3.3 与其他算法对比

为了验证所提模型在跨库语音情感识别领域的先进性，将其与基线及最新算法进行性能对比，这几种算法分别为：a）支持向量机（SVM）。选择线性核函数，C 值设置为 0.1。

b）迁移稀疏判别子空间学习（transfer sparse discriminant subspace learning , TSDSL）[3] 。引入鉴别性学习和 $\mathbf{l}_{1,2}$ 范数正则化，学习鉴别性特征并构造了最近邻图作为距离度量手段，以提升源域和目标域的相似度。

c）联合分布自适应回归（joint distribution adaptive regression, JDAR）[5]。通过联合考虑训练语音与测试语音间的边际概率分布和条件概率分布来学习回归矩阵，缓解特征分布偏差。

d）域对抗神经网络（domain adversarial neural network, DANN）[9]。特征提取器采用了四层隐层 DNN 结构，情感分类器和域鉴别器均使用两层隐层 DNN 结构。

e）深度域自适应卷积神经网络（deep domain-adaptive convolutional neural network, DDACNN）[10]。模型采用经典 LeNet 架构，尝试在不同全连接层使用 MMD 以对齐特征分布，最终在第一层全连接层纳入 MMD 获得了最优识别结果。

f）深度自编码器子域自适应（depth autoencoder subdomain adaptive, DASA）[22] 。使用自编码器进行特征处理，在编码和解码阶段均使用五层隐层 DNN 结构，并结合子域自适应实现细粒度的特征分布对齐。

将与传统算法及特征降维算法的识别精度（UAR）对比列于表 4，与深度域自适应算法的识别精度（UAR）对比列于表 5。在全部的六个任务中，相较于传统算法及特征降维算法，所提模型在 e2B、B2e、C2e、e2C 和 B2C 任务上的识别率分别领先了 4.25%~20.99% 、 3.19%~8.86% 、3.42%~10.98%、4.79%~9.89%、4.44%~5.64%，平均识别率领先 3.8%~11.01%，展现出了卷积神经网络良好的特征处理能力。相比于深度域自适应算法，所提模型在 e2B、B2e、C2e、e2C 和 B2C 任务上的识别率分别领先了 0.32%~3.06%、1.22%~6.82%、4.58%~7.5%、0.69%~5.39%、0.15%~4.94%，平均识别率领先 1.28%~5.48%，体现了所提算法经过对原有域自适应算法改进后，跨库语音情感识别模型的泛化性得到了提升。但在 C2B 任务中识别率低于最先进算法，从图 3.b 中也可看出，使用 MMD 进行域对齐，也能在该任务上有效的降低香农熵，实现与 MNMD 相近的效果，说明 MNMD 的普适性需进一步优化。总体而言，所提决策边界优化域自适应模型在对齐源域和目标域特征分布的同时，缓解了使用 MMD 进行域对齐带来的鉴别性丧失问题，提升了目标域样本的鉴别性，优化了模型决策边界，提升了模型识别性能。

表 4 与传统算法及特征降维算法的 UAR 对比

Tab. 4 Comparison of UAR with traditional and feature reduction algorithm

任务	SVM	TSDSL[3]	JDAR[5]	DBODA
e2B	32.00	47.41	48.74	**52.99**
B2e	32.47	35.44	38.14	**41.33**
C2e	25.69	33.25	28.43	**36.67**
e2C	27.40	32.50	30.30	**37.29**
C2B	44.12	**56.74**	49.58	54.25
B2C	37.80	37.40	38.60	**43.04**
Average	33.25	40.46	38.97	**44.26**

表 5 与深度域自适应算法的 UAR 对比

Tab. 5 Comparison with UAR of deep domain-adaptive algorithm

任务	DANN[9]	DDACNN[10]	DASA[22]	DBODA
e2B	52.67	49.93	52.35	**52.99**
B2e	36.53	34.51	40.11	**41.33**
C2e	29.17	31.59	32.09	**36.67**
e2C	36.60	31.90	36.10	**37.29**
C2B	**57.64**	46.62	51.47	54.25
B2C	42.89	38.10	41.40	**43.04**
Average	42.58	38.78	42.25	**44.26**

3 结语

为了解决跨库语音情感识别问题， 本文提出一种新的基于决策边界优化域自适应（DBODA）模型，旨在将源域语音库学习到的知识转移到目标域语音库，新的域自适应方法 MNMD 在进行源域与目标域特征分布对齐的同时，考虑了目标域样本的鉴别性和预测多样性，在三个基准数据集上进行的实验验证了模型的性能提升。在后续的研究中，将针对域自适应导致目标域样本鉴别性下降的问题，进一步改进域自适应算法，增强泛化性，将模型应用于更多的语音情感库中。

参考文献

[1] 李海峰,陈婧,马琳,薄洪健,徐聪,李洪伟.维度语音情感识别研究综述 [J].软件学报,2020,31(08):2465-2491. (LI H F, CHEN J, MA L, et al. Dimensional speech emotion recognition review [J]. Journal of Software, 2020,31(8):2465−2491.)

[2] LUO H, HAN J Q. Nonnegative matrix factorization based transfer subspace learning for cross-corpus speech emotion recognition [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020, 28: 2047-2060.

[3] ZHANG W J, SONG P. Transfer sparse discriminant subspace learning for cross-corpus speech emotion recognition [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2019, 28: 307-318.

[4] LUO H, HAN J Q. Cross-Corpus Speech Emotion Recognition Using Semi-Supervised Transfer Non-Negative Matrix Factorization with Adaptation Regularization [C]//INTERSPEECH. 2019: 3247-3251.

[5] ZHANG J C, JIANG L, ZONG Y, et al. Cross-Corpus Speech Emotion Recognition Using Joint Distribution Adaptive Regression [C] // ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021: 3790-3794.

[6] DENG J, XU X Z, ZHANG Z X, et al. Semisupervised autoencoders for speech emotion recognition [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017, 26(1): 31-43.

[7] GIDEON J, MCLNNIS M G, PROVOST E M. Improving cross-corpus speech emotion recognition with adversarial discriminative domain generalization (addog) [J]. IEEE Transactions on Affective Computing, 2019, 12(4): 1055-1068.

[8] LEE S. Domain Generalization with Triplet Network for Cross-Corpus Speech Emotion Recognition [C]//2021 IEEE Spoken Language Technology Workshop (SLT). IEEE, 2021: 389-396.

[9] ABDELWAHAB M, BUSSO C. Domain adversarial for acoustic emotion recognition [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2018, 26(12): 2423-2435.

[10] LIU J T, ZHENG W M, ZONG Y, et al. Cross-corpus speech emotion recognition based on deep domain-adaptive convolutional neural

network [J]. IEICE TRANSACTIONS on Information and Systems, 2020, 103(2): 459-463.
[11] KWON S. MLT-DNet: Speech emotion recognition using 1D dilated CNN based on multi-learning trick approach [J]. Expert Systems with Applications, 2021, 167: 114177.
[12] ZHAO J F, MAO X, CHEN L J. Speech emotion recognition using deep 1D & 2D CNN LSTM networks [J]. Biomedical Signal Processing and Control, 2019, 47: 312-323.
[13] SONG P, ZHENG W M, OU S F, et al. Cross-corpus speech emotion recognition based on transfer non-negative matrix factorization [J]. Speech Communication, 2016, 83: 34-41.
[14] WANG W, LI H J, DING Z M, et al. Rethinking Maximum Mean Discrepancy for Visual Domain Adaptation [J/OL]. IEEE Transactions on Neural Networks and Learning Systems. (2021-07-09)[2021-10-23]. https://ieeexplore.ieee.org/document/9478936.
[15] CUI S H, WANG S H, ZHUO J B, et al. Towards discriminability and diversity: Batch nuclear-norm maximization under label insufficient situations [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 3941-3950.
[16] RECHT B, FAZEL M, PARRILO P A. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization [J]. SIAM review, 2010, 52(3): 471-501.
[17] BURKHARDT F, PAESCHKE A, ROLFES M, et al. A database of German emotional speech [C]//Interspeech. 2005, 5: 1517-1520.
[18] MARTIN O, KOTSIA I, MACQ B, et al. The eNTERFACE'05 audio-visual emotion database [C]//22nd International Conference on Data Engineering Workshops (ICDEW'06). IEEE, 2006: 8-8.
[19] TAO J H, LIU F Z, ZHANG M, et al. Design of speech corpus for mandarin text to speech [C]//The Blizzard Challenge 2008 workshop. 2008.
[20] SCHULLER B, STEIDL S, BATLINER A, et al. The INTERSPEECH 2010 paralinguistic challenge [C]//Proc. INTERSPEECH 2010, Makuhari, Japan. 2010: 2794-2797.
[21] EYBEN F, WöLLMER M, SCHULLER B. Opensmile: the munich versatile and fast open-source audio feature extractor [C]//Proceedings of the 18th ACM international conference on Multimedia. 2010: 1459-1462.
[22] 庄志豪,傅洪亮,陶华伟,等.基于深度自编码器子域自适应的跨库语音情感识别 [J/OL].计算机应用研究:1-5, (2021-07-15) [2021-09-23]. https://doi.org/10.19734/j.issn.1001-3695.2021.04.0149. (ZHUANG Z H, FU H L, TAO H W, et al. Cross-corpus speech emotion recognition based on deep autoencoder subdomain adaptation [J/OL]. Application Research of Computers:1-5, (2021-07-15) [2021-09-23] https://doi.org/10.19734/j.issn.1001-3695.2021.04.0149.)

This work is partially supported by National Natural Science Foundation of China (62001215), Natural Science Project of Henan Education Department (21A120003), Start-up Fund for High-level Talents of Henan University of Technology (2018BS037).

WANG Yang, born in 1999, M. S. candidate. His research interests include speech signal processing.
FU Hongliang, born in 1965, Ph. D., professor. His research interests include communication and information systems.
TAO Huawei, born in 1987, Ph. D., lecturer. His research interests include speech emotion recognition.
YANG Jing, born in 1983, Ph. D., associate professor. Her research interests include communication signal processing.
XIE Yue, born in 1991, Ph. D., lecturer. His research interests include artificial intelligence and affective computing.
ZHAO Li, born in 1958, Ph. D., professor. His research interests include speech signal processing, emotional information processing.