

面向非特定人语音情感识别的 PCA 特征选择方法

罗宪华¹ 杨大利¹ 徐明星² 徐 露²

(北京信息科技大学计算机学院 北京 100101)¹

(清华大学计算机科学与技术系智能技术与系统国家重点实验室

清华信息科学与技术国家实验室 北京 100084)²

摘 要 在语音情感识别中,如何选取有效的情感特征是识别过程的重要环节。迄今为止,一些常用的特征选择算法虽然能够帮助提高识别性能,但也存在理论性不强、随机性高、计算量大的缺点。因此提出了一种基于主成分分析(PCA)的特征选择方法,亦即对原始特征集合先进行 PCA 变换,再利用变换矩阵分析出原始特征进行变换时各自的权重,最后根据权重的大小对原始特征进行选择。实验结果表明,选择出的特征对识别率具有较大的贡献,属于重要特征。

关键词 情感识别,特征选择,主成分分析

中图分类号 TP391.4 **文献标识码** A

PCA Based Feature Selection Algorithm on Speaker-independent Speech Emotion Recognition

LUO Xian-hua¹ YANG Da-li¹ XU Ming-xing² XU Lu²

(Department of Computer Science and Technology, Beijing Information Science and Technology University, Beijing 100101, China)¹

(State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory of Information Science and Technology,

Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)²

Abstract A very important part of emotion recognition is how to select effective emotional features. Until now, some feature selection algorithms, which are usually used, can help boost recognition accuracy. But some defects, such as less robustness in theory, a higher randomness, more computation, still exist. For these reasons, a new feature selection algorithm based on PCA (principal component analysis) was proposed. First the original feature set was transformed by PCA, then analyzing the weights of these features using the transforming matrix and finally, choosing the important features according to their weights. The experiment result shows that features, which are selected by this method, make a high contribution to the recognition accuracy and they are important.

keywords Emotion recognition, Feature selection, PCA

1 引言

随着人工智能技术的飞速发展,人们已经不仅仅满足于计算机能够播放多媒体信息,而更希望它能够拥有“感知”人们情感的能力。因此,针对语音情感计算的研究日益被人们所重视。如何选取有效的语音情感特征是情感识别过程中的重要环节。语音情感特征包括全局统计特征和局部频谱特征。在全局统计特征方面,基音频率、能量、共振峰、语音停顿等多种情感特征已经被广泛应用于语音情感识别领域,加上它们的统计学特征,其数量已经达到 200 余种^[1,2]。但直接使用所有的情感特征会造成计算效率低、建模精度差、特征相互制约等一系列问题。因此,需要研究特征选择算法,从原始特征集合中选出一个子集。

常用的特征选择方法有以下几类:

1)对原始数据进行随机的试探性的特征选择算法。这类算法包括单独最优特征组合法、顺序前进选择法(SFS)、增

减 r 法、顺序前进浮动选择法(SFFS)^[3]等,其理论性不强、随机性大。

2)对原始数据进行数学变换的特征选择算法。这类算法主要包括主成分分析法、对应分析法、赛德尔线性映射降维法等。它们经过数学变换虽然能够对原始特征降维,但由于变换后的特征空间已经改变,因此无法对原始特征进行选择。

3)对原始数据用分类器进行特征选择的算法。这类算法包括基于支持向量机(SVM)的特征加权法^[4]、基于神经网络的特征预测法^[5]等。它们引入了分类器,使得运算量加大,影响运行的速度。

因此很多研究者在原有的特征选择算法基础上进行改进。Ververidis 提出了一个快速的 SFFS 特征选择算法,其运算时间比传统的 SFFS 缩短了 50%^[6];Ali Hassan 对 SFS 算法进行改进,其中在 Berlin 情感语音数据库上的识别效果高达 90%^[7]。

本文提出了一种基于 PCA 的特征选择算法:首先对原始

特征集合进行 PCA 变换,然后通过分析变换矩阵得出原始特征在进行变换时各自的权重,最后根据权重由大至小的排序确定最终选择的特征子集。

2 基于 PCA 的特征选择方法

现有的很多文献采用 PCA 方法来对原始特征进行降维。虽然经过正交变换后得到的变换域特征子集与原始特征集不在同一空间中,但利用变换矩阵可以将两个集合联系在一起,从而能够由变换后的特征集推导出原始特征集中每个特征变量各自的变换权重,这种权重反映了原始特征的重要性,因此可以选择变换权重较大的特征构成最终选择的特征子集。

1)以识别率为准则,考查不同累积方差贡献率对语音情感识别精度的影响。根据累积方差贡献率与主成分个数的关系,确定主成分数目和变换矩阵:

① 根据主成分数目随累积方差贡献率变化的情况,确定 1 组主成分数目集合 $\{m_1, m_2, \dots, m_n\}$;

② 以上一步确定的主成分数目集合为目标,对原始数据分别做 PCA 变换,得到 1 组变换后的特征集合 $\{F_1, F_2, \dots, F_n\}$ 和 1 组变换矩阵集合 $\{T_1, T_2, \dots, T_n\}$;

③ 对 $\{F_1, F_2, \dots, F_n\}$ 训练和识别,得到 1 组识别率;

④ 描绘识别率曲线,确定识别率的饱和点,即识别率曲线的拐点;

⑤ 根据累积方差贡献率,得到识别率最高点对应的主成分数目 m 和变换矩阵 T 。

2)根据变换矩阵计算原始特征在变换时各自的权重,将权重按照降序排列,即可得到原始特征重要性的排列:

① 对变换矩阵 T 求权重;

$$\omega_i = \sum_{j=1}^m t_{ij} \tag{1}$$

式中, $1 \leq i \leq N, N$ 为原始特征维数;

② 将权重值按降序排列得到 $\{\omega_i'\}$,其中 $1 \leq i \leq N$ 。

3)从 $\{\omega_i'\}$ 中找出变换权重较高的那些特征构成最终选择的特征子集:

① 取 $\{\omega_i'\}$ 中前 m 位的特征向量,得到原始特征集中重要特征的权重排序;

② $\{\omega_i'\}$ 中的 i 组成的集合,即为特征选择的结果。

整个特征选择算法的流程图如图 1 所示。

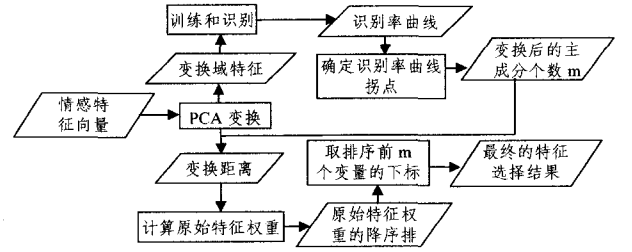


图 1 基于 PCA 的特征选择算法的流程图

3 实验结果与分析

3.1 实验设置

本文选用了 3 个情感语音数据库:CESD 数据库^[8]、EMR^[3]数据库以及 Berlin 语音数据库^[10]。其中,Berlin 情感语音数据库(5 男 5 女),共 535 句话,包括生气、害怕、高兴、伤心、厌恶、烦躁、中性 7 种情感,德语;EMR 数据库(4 男 4

女),共 1309 句话,包括生气、害怕、高兴、中性、伤心 5 种情感,中文。CESD 数据库(25 男 25 女),经过挑选处理,共 1717 句话,包括生气和中性 2 种情感,中文。

3.2 特征提取

本文利用 Praat 语音学软件,从情感语音中提取了全局统计特征:基音频率、能量、共振峰、信噪比、语音停顿、基频微扰和振幅微扰 7 类原始特征。计算了它们的统计学特征,包括平均值、中位值、标准差、最小值、最大值、变化范围、斜率以及我们自己定义的 3 种特征 R_1, R_2, R_3 。

$$R_1 = \frac{\text{均值} - \text{最小值}}{\text{变化范围}} \tag{2}$$

$$R_2 = \frac{\text{均值} - \text{最小值}}{\text{中值} - \text{最小值}} \tag{3}$$

$$R_3 = \frac{\text{均值}}{\text{中值}} \tag{4}$$

所有提取的全局统计特征如表 1 所列。

表 1 85 维全局统计特征

声学特征类别	全局统计量	特征数量
基音频率	平均值、中位值、标准差、最小值、最大值、变化范围、斜率、 R_1, R_2, R_3 、及其一阶差分的相关统计量	21
强度	平均值、中位值、标准差、最小值、最大值、变化范围、斜率、 R_1, R_2, R_3 、及其一阶差分的相关统计量	19
共振峰	前四个共振峰频率的平均值、中位值、标准差、最小值、最大值、变化范围、 R_1, R_2, R_3	36
语音停顿	平均值、标准差	2
信噪比	平均值、标准差、变化范围	3
基频微扰	平均值、标准差	2
振幅微扰	平均值、标准差	2

3.3 确定合理的主成分个数

本文在 EMR 数据库和 Berlin 数据库上分别进行了与说话人无关的 8 交叉实验和 10 交叉实验。对每组实验用训练数据做 PCA,根据主成分数目随累积方差贡献率变化的情况,选择合适的阈值,确定需要保留的主成分数目,最终确定变换后得到 1 维、10 维、20 维、30 维、45 维、55 维、65 维、85 维的特征集。这些取值有助于准确地确定曲线的拐点,识别率取交叉验证的平均值。实验结果如图 2 所示。

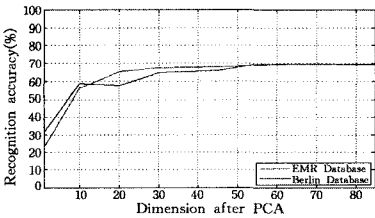


图 2 两组识别率曲线

从图 2 可见,当主成分的维数达到 30 维时,识别率基本饱和,并且这个取值与语种是无关的。因此,本文认为主成分的个数 $m=30$ 的取值最为合理,其对应的变换矩阵为 T_{30} 。

3.4 确定特征子集

对于 EMR 数据库,由每组实验产生的 T_{30} 得到 8 组特征变换权重由大至小的降序排列,取每组前 30 个特征,即可得到这 8 组实验特征选择的结果。计算这些特征在 8 组选择的特征中出现的次数,出现 4 次以上的特征,本文认为是相对于

(下转第 256 页)

- [2] Rieser C J. Biologically Inspired Cognitive Radio Engine Model Utilizing Distributed Genetic Algorithms for Secure and Robust Wireless Communications and Networking [D]. Blacksburg, Virginia, Dept of Electrical Engineering in Virginia Tech, 2004
- [3] Zhang X Q, Huang Y Q, et al. Design of cognitive radio node engine based on genetic algorithm [C]// WASE, ICIE'09, IEEE, Taiyuan, China, 2009: 22-25
- [4] Dorigo M, Birattari M, Stützle T. Ant colony optimization [J]. IEEE Computational Intelligence Magazine, 2006, 1(4): 28-39

- [5] 史恒亮, 白光一等. 基于蚁群优化算法的云数据库动态路径规划 [J]. 计算机科学, 2010, 37(5): 143-145, 213
- [6] Zhao N, Li S Y Z L. Cognitive Radio Engine Design Based on Ant Colony Optimization [J]. Wireless Personal Communications, 2011, 1(10): 1-10
- [7] Newman T R, B A B, Wyglinski A M. Cognitive engine implementation for wireless multicarrier transceivers [J]. Wireless Communications Mobile Computing, 2007, 7(9): 1129-1142
- [8] 王越, 许全文, 黄丽丰. 基于改进遗传算法的连续函数优化 [J]. 重庆理工大学学报: 自然科学版, 2011, 25(2): 62-67

(上接第 213 页)

该数据库的重要特征。同理, 对于 Berlin 数据库, 在最终选择出的 10 组特征集中, 出现 5 次以上的特征被认为是该数据库的重要特征。这样就得到了两组针对不同数据库选择的特征集合。为了找到一组鲁棒性强的特征集, 将这两组重要特征取交集, 得到的特征子集包含 9 种特征。EMR 数据库和 Berlin 数据库各自重要的特征, 以及取交集后得到的稳定特征集, 如表 2 所列。

表 2 两个数据库的特征选择结果及它们的交集

数据库	特征种类
两数据库 共同特征	基音频率: 最小值、 R_3 能量: 均值 第二共振峰: 均值、中值 第三共振峰: 中值、最大值、 R_2 第四共振峰: R_2
Berlin 数据库特征	基音频率: 标准差、最小值、变化范围、 R_3 、一阶差分标准差 能量: 均值、一阶差分中值 第一共振峰: R_2 第二共振峰: 均值、中值、最大值 第三共振峰: 最大值、中值、最小值、 R_2 第四共振峰: R_2 、变化范围
EMR 数据库特征	基频: 最小值、 R_3 能量: 均值、变化范围、 R_1 、 R_2 、 R_3 、最大正斜率、正斜率标准差、负斜率标准差、正斜率所占比率 第一共振峰: 中值、最大值 第二共振峰: 最大值、均值、中值、变化范围 第三共振峰: 最大值、中值、变化范围、 R_2 第四共振峰: 变化范围、 R_2 、 R_3 信噪比: 均值

3.5 特征子集的贡献

针对以上两个数据库, 分别提取上节中取交集得到的这 9 维情感特征, 以 SVM 为分类器, 分别作 8 交叉验证和 10 交叉验证, 平均识别结果如表 3 所列。相比较 85 维特征全部使用所得识别率, 可以看到这 9 维情感特征对识别率确实起了很大作用。

表 3 9 维情感特征与 85 维情感特征识别率比较

	9 维	85 维
EMR 数据库	57.58%	69.23%
Berlin 情感语音数据库	47.39%	69.39%

3.6 9 维情感特征的鲁棒性检验

本文在 CESD 数据库上, 选用 BP 神经网络 (BP-ANN)、高斯混合模型 (GMM) 和 SVM 3 种不同的分类器, 进行了男女性别的交叉实验, 实验结果如表 4 所列。从实验结果中看到, 利用 PCA 方法选择出的 9 维情感语音特征, 对于不同识别器识别性能差异不大, 这说明 9 维情感特征对识别器

是稳定的, 而且与 85 维特征全部使用后的识别结果相当, 说明了利用 PCA 进行特征选择的方法是可行的。

表 4 9 维特征选用不同分类器的识别结果比较

训练	测试	PCA 特征选择算法			无特征选择
		BP-ANN	GMM	SVM	SVM
女	男	66.73%	67.79%	67.50%	67.70%
男	女	65.06%	63.53%	63.99%	66.49%

结束语 本文利用 PCA 方法对语音情感识别的声学特征进行选择。实验结果表明, 最终选择出来的特征子集对识别精度具有较大的贡献, 证明该特征选择方法是有效的、可行的。多个数据库上的交叉实验表明本文利用的基于 PCA 的特征选择方法选择出的特征集对不同人、不同分类器具备鲁棒性。另一方面, 本文所提取的 85 维全局统计特征与迄今发现的情感特征总数相比并不十分全面, 所涉及到的情感种类也比较有限。因此, 今后的研究工作将在更大的特征集、更多的情感种类、更多的语音种类上来研究本方案的有效性。

参 考 文 献

- [1] Clavel C, Vasilescu I, Devillers L, et al. Fear-type emotion recognition for future audio-based surveillance systems [J]. Speech Communication, 2008, 50: 487-503
- [2] Anagnostopoulos C N. Feature selection in acted speech for the creation of an emotion recognition personalization service [C]// Third International Workshop on Semantic Media Adaptation and Personalization, 2008: 116-121
- [3] 扈浩. 汉语语音情感识别研究 [D]. 北京: 清华大学, 2007
- [4] 毛勇. 基于支持向量机的特征选择方法的研究与应用 [D]. 杭州: 浙江大学, 2006
- [5] 郑发泰. 基于神经网络故障诊断主特征量提取的应用研究 [J]. 煤矿机械, 2007, 28: 188-190
- [6] Ververidis D, Kotropoulos C. Fast and accurate sequential floating forward feature selection with the Bayes classifier applied to speech emotion recognition [J]. Signal Processing, 2008, 88: 2956-2970
- [7] Hassan A, Damper R I. Emotion Recognition from Speech using Extended Feature Selection and a Simple Classifier [C]// Interspeech. 2009: 2043-2046
- [8] 徐露, 徐明星, 杨大利. 面向情感变化检测的汉语情感语音数据库 [J]. 北京: 清华大学学报: 自然科学版, 2009: 1413-1418
- [9] Burkhart F, Paeschke A, Rolfes M, et al. A Database of German Emotional Speech [C]// Interspeech. 2005: 1517-1520