基于 Fisher 准则和 Adaboost 的语音情感多分类研究*

邢　尹[1]　刘立龙[1]　程　胜[1]　时满星[1]　李月锋[2]
（1. 桂林理工大学测绘地理信息学院　桂林　541004）（2. 兰州理工大学土木工程学院　兰州　730050）

摘　要　随着社会和科技的快速发展，如何有效识别语音情感已经成为人们关注的一个热点。在众多的分类算法中，Adaboost 多分类算法得到了较好的应用效果。该算法将生气、开心、中性、伤心和害怕 5 种语音情感分为三层，由粗到细，逐层识别。基于柏林情感语音库，将提取的语音情感特征利用 Fisher 准则选择较佳特征作为实验数据，实验结果表明，将相近情感分在一起训练更有利于提升 Adaboost 算法的分类性能。此外，在与传统的 BP 和 SVM 分类模型比较中，Adaboost 多分类算法表现出了优越性。

关键词　Fisher 准则；Adaboost 算法；语音情感识别
中图分类号　TP391.4　　**DOI**：10. 3969/j. issn. 1672-9722. 2018. 11. 007

Reserach on Speech Emotion Classification Based on Fisher Criterion and Adaboost Algorithm

XING Yin[1]　LIU Lilong[1]　CHENG Sheng[1]　SHI Manxing[1]　LI Yuefeng[2]
（1. School of Geomatics and Geoinformation, Guilin University of Technology, Guilin　541004）
（2. School of Civil Engineering, Lanzhou University of Technology, Lanzhou　730050 ）

Abstract　With the rapid development of society and technology, how to effectively identify speech emotion has become a hot topic. In many classification algorithms, Adaboost multi-classification algorithm has a good application effect. The algorithm divides 5 speech emotions (anger, happiness, neutral, sadness and fear) into three layers. It is identified layer by layer from coarse to fine. Based on Berlin emotional speech database, the speech emotion features are extracted, and the better features are selected as the experimental data using Fisher criterion. The experimental results show that it is more advantageous to improve the classification performance of Adaboost algorithm by training similar emotions. In addition, Adaboost multi-classification algorithm shows superiority in comparison with traditional BP and SVM classification models.

Key Words　Fisher criterion, Adaboost algorithm, speech emotion recognition
Class Number　TP391.4

1　引言

语言是人类交流的重要工具，人的语音中不仅包含人的语义信息，也包含情感信息。随着人机交互的越来越紧密，传统的语音识别技术已经不能满足人们的需要。近年来，语音情感识别成为人工智能领域的一个研究热点[1]。在智能的人机交互系统中，通过对操作者的情感进行分析，可以更主动、更准确地完成操作者的指示，实时调整对话方式，使得交流更加友好、更加智能。目前，国内外学者在这方面进行了大量的研究。文献[2]使用高斯混合矢量自回归模型在柏林情感数据集上得到了 76%的识别率；文献[3]使用深度信念网络与 SVM 相结合的算法对语音情感做了深入探讨；文献[4]使用经验模态分解法结合 Teager 能量对情感语音进行处理，得到了较好的效果；文献[5]使用改进的混合蛙跳算法对 SVM 进行优化，为实用语音情感识别提供了新思路。

本文在前人研究的基础上，提出了用 Adaboost 算法对语音情感多分类进行研究。提取了柏林数

* **收稿日期**：2018 年 5 月 7 日，**修回日期**：2018 年 6 月 20 日
作者简介：邢尹，女，硕士研究生，研究方向：机器学习、GNSS 技术及应用。刘立龙，男，教授，博士，研究方向：机器学习、GNSS 技术及应用。程胜，男，硕士研究生，研究方向：机器学习、GNSS 技术及应用。时满星，男，硕士研究生，研究方向：机器学习、GNSS 技术及应用。李月锋，男，硕士研究生，研究方向：机器学习、变形监测。

据集中的生气、开心、中性、伤心和害怕5种情感的140维特征,并采用Fisher准则和分层分类思想有效地实现了情感的分类。在与传统的BP神经网络和SVM模型的分类比较中,结果表明了该算法的有效性。

2 语音情感特征提取

语音情感特征反映了人的情感状态,对最终的情感识别有着重要的影响。经过研究者对心理学和语音语言学的大量研究,目前语音情感特征主要关注在韵律特征和音质特征[6~7]。本文选取了短时能量、基音频率、共振峰、梅尔倒谱系数(MFCC)这4类特征及其衍生参数[8],共构成140维的语音情感特征参数用于识别。在这140维情感特征中,必然存在大量的非重要特征和冗余特征,对特征进一步的处理是必要的。Fisher准则从均值和方差角度对特征进行评价。对 d 个维度,Fisher判别准则可以用式(1)来表示:

$$f(d)=\frac{(\mu_{1d}-\mu_{2d})^2}{\sigma_{1d}^2+\sigma_{2d}^2} \tag{1}$$

其中,μ_{1d}、μ_{2d}、σ_{1d}^2 和 σ_{2d}^2 为第 d 个维度两个类别特征值的均值和方差。Fisher判别准则越大,表明该特征区分这两种类别效果越好。对于多类情况,采用式(2)进行计算:

$$f(d)=\sum_{0<i<j<m}^{m}\frac{(\mu_{id}-\mu_{jd})^2}{\sigma_{id}^2+\sigma_{jd}^2} \tag{2}$$

式中,m 为类别总数。依据Fisher判别准则,对生气、开心、中性、伤心和害怕5种情感选择出的前14个最佳特征见表1。

表1 前14个最佳特征

重要程度排序	特征
1	浊音帧差分基音的均值
2	第二共振峰频率比率的均值
3	8阶MFCC一阶差分的均值
4	基音频率的最小值
5	7阶MFCC一阶差分的均值
6	1阶MFCC一阶差分的均值
7	3阶MFCC一阶差分的均值
8	9阶MFCC一阶差分的均值
9	5阶MFCC的最大值
10	0阶MFCC一阶差分的均值
11	5阶MFCC一阶差分的均值
12	第一共振峰频率的最小值
13	第三共振峰频率的均值
14	2阶MFCC一阶差分的均值

3 Adaboost算法

Adaboost算法是1995年由Freund和Schapire在Boosting算法理论基础上提出来的[9],它的主要思想是合并多个“弱”分类器的输出以产生有效分类。算法主要步骤如下:

Step 1:给定输入输出。输入:训练数据集:$T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$,其中 $x_i\in\boldsymbol{X}\subseteq\boldsymbol{R}^n$,$y_i\in\boldsymbol{Y}=\{-1,+1\}$,$\boldsymbol{X}$ 为实例空间,$\boldsymbol{Y}$ 为标记集合;本文弱分类器采用单层决策树;输出:最终分类器 $G(x)$。

Step 2:初始化训练数据集的权值分布。$D_1=(w_{11},\cdots,w_{1i},\cdots w_{1N})\ w_{1i}=\frac{1}{N},i=1,2,\cdots N$。

Step 3:对 $m=1,2,\cdots M$,使用具有权值分布 D_m 的训练数据集学习,得到基本分类器:$G_m(x):\boldsymbol{X}\to\{-1,+1\}$。

Step 4:计算 $G_m(x)$ 在训练数据集上的分类误差率:

$$e_m=P(G_m(x_i)\neq y_i)=\sum_{i=1}^{N}w_{mi}\boldsymbol{I}(G_m(x_i)\neq y_i) \tag{3}$$

Step 5:计算 $G_m(x)$ 的系数:

$$\alpha_m=\frac{1}{2}\ln(\frac{1-e_m}{e_m}) \tag{4}$$

Step 6:更新训练数据集的权值分布:

$$D_{m+1}=(w_{m+1,1},\cdots,w_{m+1,i},\cdots w_{m+1,N}) \tag{5}$$

$$w_{m+1,i}=\frac{w_{mi}}{Z_m}\exp(-\alpha_m y_i G_m(x_i)),\ i=1,2,\cdots,N \tag{6}$$

其中,Z_m 是规范化因子,为

$$Z_m=\sum_{i=1}^{N}w_{mi}\exp(-\alpha_m y_i G_m(x_i)) \tag{7}$$

Step 7:构建基本分类器的线性组合:

$$f(x)=\sum_{m=1}^{M}\alpha_m G_m(x) \tag{8}$$

得到最终分类器:

$$G(x)=\mathrm{sign}(f(x))=\mathrm{sign}(\sum_{m=1}^{M}\alpha_m G_m(x)) \tag{9}$$

4 语音情感分层分类

Adaboost算法用于多分类问题时,本质上是转化为Adaboost二分类问题[10]。通过将生气、开心、中性、伤心和害怕5种情感进行分层,并逐层处理,最终得到识别结果。本文设置了如下三种二叉树分层方式。图1(a)、(b)顶层为三二情感训练,图1(c)顶层为一一情感训练。其中,图1(a)左边放置

三个具备正负情感于一起，图 1(b)放置三个相近情感于一起。以图 1(a)为例，每个节点设置一个 Adaboost 分类器，这样共需要设置 4 个分类器。将初始情感放置于顶层，其中设置生气、开心、中性训练样本对应标签+1，伤心、害怕训练样本对应标签−1；第二层生气、开心训练样本对应标签+1，中性训练样本对应标签−1，伤心训练样本对应标签+1，害怕训练样本对应标签−1；第三层生气训练样本对应标签+1，开心样本对应标签−1。经过 4 个 Adaboost 分类器训练，得出 4 个训练模型。当未知样本从顶层进入时，采用第一个 Adaboost 模型得出分类结果，进入下一层，由粗及细，逐层处理，并最终得到识别结果。

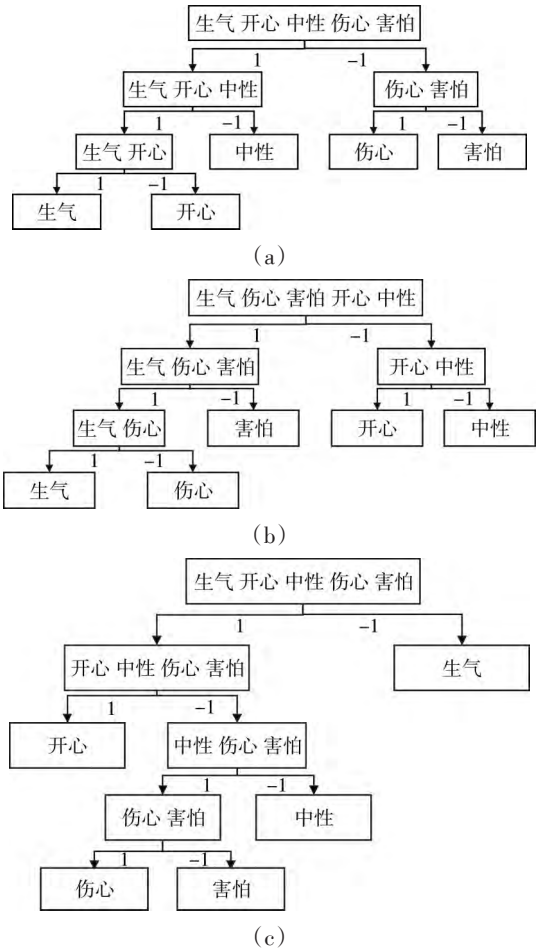

图 1　三种分层方式

5　实验结果及分析

5.1　实验样本

本文实验样本来自于柏林情感语音库[11]。柏林情感语音库是柏林工业大学通过演员表演形式所录制，10 名非专业演员(5 男 5 女)对生气、高兴、中性、伤心、害怕、厌恶、无聊 7 种情感 10 句录音脚本进行演绎，共录制了 800 条情感语句，经过 20 名志愿者试听辨别，最终保留了 535 条语句。本文选取前 5 种语音情感，共 400 条语句构成实验样本，具体为生气 126 条、高兴 68 条、中性 78 条、伤心 62 条和害怕 66 条。数据集是以 16000 采样率，16bit 量化，wav 格式存储，训练样本与测试样本按照 1∶1 分配。

5.2　语音情感识别实验

本实验在 Matlab R2014a 平台上进行编程实现。采用 Fisher 准则进行特征选择，基于 15 个弱分类器的 Adaboost 模型，三种分层方式下情感特征个数与 5 种情感的平均识别率之间的关系如图 2。从图 2 可以看出，采用方式 2 分层可以达到最大的识别率 94%。可能是将相近情感分在一起进行训练，更有利于找出它们之间的内在差异性，从而达到较佳的识别效果。特征选择的原则是在保证识别率的基础上，采用尽可能少的特征。图 2 中，在最佳 14 个特征时，达到了最大的识别率，远高于原始 140 个特征的识别率，说明了 Fisher 准则对语音情感特征选择是有效的，具体的 14 个特征见表 1。为了更好地说明进行特征选择的优越性，方式 2 分层下不同特征数的运行时间如图 3。从图 3 可以看出，特征个数与运行时间基本呈线性增长关系。其中，14 个特征时，运行时间为 0.08109s；140 个特征时，运行时间为 0.7417s。因此，采用 Fisher 准则特征选择处理后，大大缩短了运行时间。

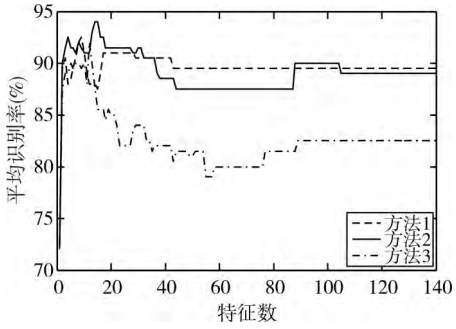

图 2　三种分层方式下不同特征数的识别率

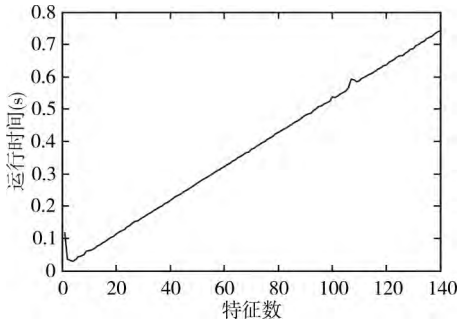

图 3　方式 2 下不同特征数的运行时间

在不同数目的弱分类器下，分层方式 2 下 5 种情感的平均识别率如图 4。从图 4 中可以看出，并

不是采用越多的弱分类器识别效果就越佳。在15个弱分类器时,5种情感的平均识别率达到最大;在19个弱分类器时,平均识别率达到平稳。

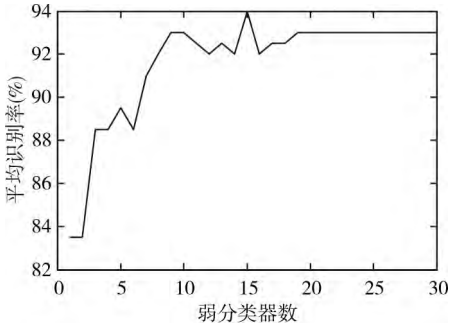

图4 方式2下不同弱分类器数的识别率

为了说明Adaboost算法的优越性,本实验分别建立了BP神经网络和SVM模型对5种情感进行识别。其中,采用的BP网络结构为14-10-5,参数设置:最大迭代次数为20000,目标精度为0.001,学习率为0.1;SVM参数设置:惩罚因子C=2.0,核参数g=1.0,RBF核函数。三种模型识别结果,见表2~4。

表2 Fisher+BP模型5种情感识别结果

测试样本	识别情感/%				
	生气	开心	中性	伤心	害怕
生气	98.41	1.59	0	0	0
开心	0	85.29	8.82	2.94	2.94
中性	10.26	0	89.74	0	0
伤心	3.23	9.68	0	87.10	0
害怕	0	6.06	0	6.06	87.88

表3 Fisher+SVM模型5种情感识别结果

测试样本	识别情感/%				
	生气	开心	中性	伤心	害怕
生气	96.83	3.17	0	0	0
开心	0	91.18	5.88	0	2.94
中性	5.13	0	92.31	2.56	0
伤心	0	9.68	0	90.32	0
害怕	0	9.09	0	3.03	87.88

表4 Fisher+Adaboost模型5种情感识别结果

测试样本	识别情感/%				
	生气	开心	中性	伤心	害怕
生气	96.83	0	0	3.17	0
开心	0	88.24	0	8.82	2.94
中性	5.13	2.56	92.31	0	0
伤心	0	6.45	3.23	90.32	0
害怕	0	0	0	0	100

从表2~4可以得出,Fisher+BP模型的平均识别率为91.00%,Fisher+SVM模型的平均识别率为92.50%,而Fisher+Adaboost模型通过集合15个单层决策树构成的强分类器得到了94.00%的识别率,高于Fisher+BP模型3个百分点,高于Fisher+SVM模型1.5个百分点。特别地,Fisher+Adaboost模型对生气、伤心害怕的平均识别率达到了96.06%,而其他两种模型都只有92.91%。一般来说,负面情感对人的影响最大,如果能有效地识别人的负面情感,将有助于提高个体认知和工作效率。采用Fisher+Adaboost模型对负面情感的感知更为敏感,能更为有效地监测负面情感,具有重要的工程意义。

6 结语

本文提出利用Fisher准则对提取的语音情感特征进行评价,选择其中较佳特征。利用Adaboost算法对生气、开心、中性、伤心和害怕5种情感进行分类,实验结果表明,Fisher准则表现出了良好的特征数降低特性,以及将相近的情感状态分在一起训练达到了更佳的分类效果。此外,在与传统的BP和SVM模型比较中,Adaboost算法表现出了优越性。

参考文献

[1] 张石清,李乐民,赵知劲. 人机交互中的语音情感识别研究进展[J]. 电路与系统学报,2013,18(2):440-451.
ZHANG Shiqing, LI Lemin, ZHAO Zhijin. A survey of speech emotion recognition in human computer interaction [J]. Journal of Circuits and Systems, 2013, 18 (2): 440-451.

[2] El Ayadi M M H, Kamel M S, Karray F. Speech emotion recognition using gaussian mixture vector autoregressive models[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2007, IV-957-IV-960.

[3] Huang C, Gong W, Fu W, et al. A research of speech emotion recognition based on deep belief network and svm [J]. Mathematical Problems in Engineering, 2014 (5): 1-7.

[4] 张卫,张雪英,孙颖. EMD结合Teager能量用于语音情感识别[J]. 科学技术与工程,2013,13(24):7240-7243.
ZHANG Wei, ZHANG Xueying, SUN Ying. EMD combined Teager energy for emotional speech recognition[J]. Science Technology and Engineering, 2013, 13 (24): 7240-7243.

[5] 张潇丹,黄程韦,赵力,等. 应用改进混合蛙跳算法的实用语音情感识别[J]. 声学学报,2014,39(2):271-280.
ZHANG Xiaodan, HUANG Chengwei, ZHAO Li, et al.

(下转第2229页)

land mine detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(6): 577-589.

[7] Allanki Sudheer, Ch. Hima Bindu. Region based Multi-focus Image Fusion using the spectral parameter Variance [C]//IEEE WiSPNET, 2016:1306-1310.

[8] 叶银芳,聂建英. 基于多分辨分析的红外/被动毫米波图像主成分融合[J]. 电光与控制, 2013,20(9):6-9.
YE Yinfang, NIE Jianying. Principal Component Fusion of IR/PMMW Image Based on Multiresolution Analysis [J]. Electronics Optics & Control,2013,20(9):6-9.

[9] 刘军华,雷超阳. 基于B样条梯度的融合图像算法[J]. 吉首大学学报(自然科学版),2012,33(6):61-65.
LIU Junhua, LEI Chaoyang. Gradient Image Fusion Algorithm Based on B-Spline[J]. Journal of Jishou University (Natural Sciences Edition),2012,33(6):61-65.

[10] Wen Guan, Li Li, Weiqi Jin, et al. Research on HDR image fusion algorithm based on Laplace pyramid weight transform with extreme low-light CMOS [J]. AOPC 2015:Image Processing and Analysis,2015,9675:1-10.

[11] Lianhai Wang, Junping Du, Suguo Zhu, et al. New region-based image fusion scheme using the discrete wavelet frame transform[C]//12th World Congress on Intelligent Control and Automation,2016:3066-3070.

[12] J. Malleswara Rao, C. V. Rao, A. Senthil Kumar, et al. Spatiotemporal Data Fusion Using Temporal High-Pass Modulation and Edge Primitives[J]. Ieee Transactions on Geoscience and Remote Sensing, 2015, 53(11): 5853-5860.

[13] 潘赟,赵喜玲. NSST域高斯模糊逻辑的图像融合[J]. 应用激光,2016,36(3):351-356.
PAN Yun,ZHAO Xiling. Fusion of Gaussian Fuzzy Logic on NSST Domain[J]. Applied Laser, 2016, 36(3): 351-356.

[14] 宋村夫,高颖慧,王平. 一种基于最大似然比理论的表决融合算法[J]. 重庆理工大学学报(自然科学版), 2011, 25(6):79-83.
SONG Cunfu, GAO Yinghui, WANG Ping. A Voting Fusion Process based on Generalized Likelihood Ratio Test [J]. Journal of Chongqing University of Technology(Natural Science),2011,25(6):79-83.

[15] Baihong Lin, Xiaoming Tao, Shaoyang Li, et al. Variational Bayesian Image Fusion Based on Combined Sparse Representations[C]//IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016:1432-1436.

[16] 刘坤. 像素级多传感器图像融合的研究[D]. 西安:西北工业大学,2007:1-67.
LIU Kun. Research on Pix based Fusion of Multi-source Images[D]. Xi'an: Northwestern Polytechnical University,2007:1-67.

[17] 张小利. 图像融合及其性能评估若干问题研究[D]. 长春:吉林大学,2016:1-161.
ZHANG Xiaoli. Study on Some Issues of Image Fusion and Performance Evaluation[D]. Changchun: Jilin University,2016:1-161.

(上接第 2200 页)

Recognition of practical speech emotion using improved shuffled frog leaping algorithm[J]. Acta Acustica, 2014, 39(2):271-280.

[6] Huang C, Jin Y, Zhao Y, et al. Recognition of practical emotion from elicited speech[C]// International Conference on Information Science and Engineering. IEEE, 2009:639-642.

[7] Tato R, Santos R, Kompe R, et al. Emotional space improves emotion recognition[C]// International Conference on Spoken Language Processing, Icslp2002-INTERSPEECH 2002, Denver, Colorado, Usa, September. DBLP,2002:2029-2032.

[8] 梁瑞宇, 赵力, 魏昕. 语音信号处理实验教程[M]. 北京:机械工业出版社,2016:231-234.
LIANG Ruiyu, ZHAO Li, WEI Xin. Experimental course on speech signal processing[M]. Beijing: China Machine Press,2016:231-234.

[9] Freund Y, Schapire R E. A decision-theoretic generalization of on-line learning and an application to boosting [C]// European Conference on Computational Learning Theory. Springer-Verlag, 1995:119-139.

[10] 潘虎,陈斌,李全文. 基于二叉树和Adaboost算法的纸币号码识别[J]. 计算机应用,2011,31(2):396-398.
PAN Hu, CHEN Bin, LI Quanwen. Paper currency number recognition based on binary tree and Adaboost algorithm[J]. Journal of Computer Applications, 2011, 31(2):396-398.

[11] Burkhardt F, Paeschke A, Rolfes M, et al. A database of German emotional speech [C]//INTERSPEECH 2005-Eurospeech, European Conference on Speech Communication and Technology, Lisbon, Portugal, September. DBLP,2005:1517-1520.