语音情感识别研究进展*

赵腊生[1,2]，张 强[2]，魏小鹏[1,2]

(1. 大连理工大学 机械工程学院，辽宁 大连 116024；2. 大连大学 先进设计与智能计算省部共建教育部重点实验室，辽宁 大连 116622)

摘 要：首先介绍了语音情感识别系统的组成，重点对情感特征和识别算法的研究现状进行了综述，分析了主要的语音情感特征，阐述了代表性的语音情感识别算法以及混合模型，并对其进行了分析比较。最后，指出了语音情感识别技术的可能发展趋势。

关键词：语音；情感；识别

中图分类号：TP391 **文献标志码**：A **文章编号**：1001-3695(2009)02-0428-05

Survey on speech emotion recognition

ZHAO La-sheng[1,2], ZHANG Qiang[2], WEI Xiao-peng[1,2]

(1. School of Mechanical Engineering, Dalian University of Technology, Dalian Liaoning 116024, China; 2. Key Laboratory of Advanced Design & Intelligent Computing of Ministry of Education, Dalian University, Dalian Liaoning 116622, China)

Abstract: First, introduced the system of speech emotion recognition. Second, detailed the used features. Then, presented comparison and analysis on the classification algorithms. At last, discussed future directions.

Key words: speech; emotion; recognition

语音情感识别是指由计算机自动识别输入语音的情感状态。作为人机语音智能交互的关键技术，语音情感识别技术吸引了越来越多学者的注意；同时，随着该项技术研究的深入，其对计算机发展和社会生活的重要性也日益凸现出来，在诸多领域如互动电影、情感翻译、心理检测、电子游戏和辅助心理治疗等得到了应用。因此，语音情感识别研究具有重要的理论价值和应用前景。

目前有许多关于语音和情感关系的研究，如美国、日本、韩国、欧洲等许多国家的一些研究单位都在进行情感语音处理研究工作。国内也已有多所高校从事语音情感识别的研究，东南大学较早地开始了这方面的研究，中国科学院、清华大学、浙江大学、哈尔滨工业大学、微软亚洲研究院、中国台湾的一些大学和研究所等机构也在这方面做了大量工作。

至今已有数篇综述文献分别从不同的角度总结了语音情感识别的研究成果[1~3]，以上文献主要总结了 2005 年之前的研究概况，本文的目的是对这些文献提供进一步的补充，着重增加 2005 年以后的有关语音情感识别的新进展，以供读者了解语音情感识别目前的发展趋势和方向。

1 语音情感识别的系统概述

语音情感识别系统大致包括三个环节，即预处理、特征提取和情感分类。系统的框架如图 1 所示。通常，输入的语音信号都要进行预处理，预处理过程的好坏在一定程度上也影响系统的识别效果。预处理主要包括采样量化、预加重、端点检测、分帧加窗。当然以上环节根据提取特征的不同略有调整。特征提取的任务是从输入的语音信号中提取能够区分不同情感的参数序列，在提取特征数据的过程中，为了获得最优特征子集，可能还需要特征降维、特征选择等进一步处理。而模式分类的任务则包含了两个方面：a) 在训练时用反映情感特征的参数序列，为每种情感建立相应的情感模型；b) 在测试或识别时根据所得到的待识别语音信号的特征参数序列，由系统对这些参数和已知情感模型之间的相似程度进行评估，并根据评估的结果判断输入语音信号的情感归属。

图 1 语音情感识别系统

2 语音情感特征参数

语音情感的变化通过特征参数的差异来体现，研究者已尝试使用了诸多情感特征。目前语音情感识别系统主要依靠语音的低层次声学特征来进行识别，这些特征大致可分为基于模型的特征和非基于模型的特征。

2.1 基于模型的特征

2.1.1 线性激励源/滤波器语音产生模型特征

这类特征主要表现在语音的频谱结构上，包含了反映声道共振的频谱包络特征信息和反映声带振动等音源特性的频谱细节构造特征信息，具有代表性的特征参数有基音频率和

收稿日期：2008-06-04；修回日期：2008-08-11 基金项目：国家教育部新世纪优秀人才支持计划资助项目(NCET-06-0298)；辽宁省高等学校优秀人才支持计划资助项目(RC-05-07, 2006R06)；辽宁省教育厅科学研究计划资助项目(05L020)；大连市科学技术计划资助项目(005A10GX106)；大连大学智能信息处理重点实验室开放课题资助项目(05-2)

作者简介：赵腊生(1978-)，男，博士研究生，主要研究方向为语音信号处理(goodzls@126.com)；张强(1971-)，男，教授，主要研究方向为智能计算；魏小鹏(1959-)，男，教授，博导，主要研究方向为计算机图形学.

共振峰。浊音的声带振动基本频率称为基音频率,文献[4]对多语种下的情感语音基频参数变化进行了统计分析,统计结果表明情感语音的基频结构特征随情感状态改变有明显的变化,且不同语种下这种结构的变化有较好的一致性。基频对于情感识别的重要作用已成为语音情感研究人员的共识,众多的文献都采用了基频作为获取情感信息的一个重要参数[5~7]。在这些研究中,提取的基音参数一般是一段语音的基频衍生参数,如基频的均值、范围、方差、中值、轮廓变化等。由于基频与人的生理构造密切相关,在不同的个体上表现出较强的相异性和不稳定性,基频本身绝对数值使用较少,更为常用的是基频的统计数值,而且在不同的性别上基频差异更为明显。文献[8]注意到了这种差异,通过分析基频均值、方差、统计分布模型在性别上的差异,对基频参数进行基于性别差异的规整;引入规整后的基频均值和方差以及基频统计分布模型距离作为情感特征参数,实验结果表明应用规整后的参数提高了识别率。声道可以看成是一根具有非均匀截面的声管,在发音时起共鸣器作用。当元音激励进入声道时会引起共振特性,产生一组共振频率即共振峰。共振峰是反映声道特性的一个重要参数,考虑到不同情感的发音可能使声道有不同的变化,而每种声道形状都有一套共振峰频率作为特征,因此,共振峰也是表达情感的特征参数之一。它一般包括共振峰的位置和频带宽度,在文献中应用最广泛的是前三个共振峰峰值及前三个共振峰的带宽。由于共振峰参数存在个体差异,类似于基频特征其应用较多的也是其统计特征[6,9,10]。

2.1.2　非线性激励源/滤波器语音生成模型特征

传统的语音学方法对语音模型的研究将语音的产生假定为线性源—滤波器模型,语音被假设是沿声道方向传播的平面波;但 Teager 等人认为当气流通过声带和伪声带区域会出现气流的分离、附着,进而形成涡流,并与平面波一起构成语音生成的原因。基于对这一非线性问题的考虑,文献[11]提出了 Teager 能量算子(TEO),并给出了离散形式的 TEO 算子运算公式为

$$\psi[x(n)] = x^2(n) - x(n-1)x(n+1) \tag{1}$$

其中:$\psi[x(n)]$ 为 TEO 算子;$x(n)$ 为信号时域采样值。信号 $x(n)$ 在 n 点的 TEO 只与该样本点及其前后各样本点有关。随着 TEO 算子的提出,许多基于 TEO 的特征被用于识别语音中的情感。文献[12]将 TEO 算子分别应用于信号的时域和频域,时域 TEO 变换采用式(1),频域 TEO 变换采用如下公式:

$$\psi[x(f)] = x^2(f) - x(f-1)x(f+1) \tag{2}$$

其中:$x(f)$ 为信号频域采样值。将以上两种变换分别与子带能量特征相结合,提出两种基于 TEO 的非线性特征用于识别语音情感。文献[13]结合小波分析的多分辨率思想将不同形式的 TEO 与美尔频域倒谱系数(MFCC)相结合,提出五种非线性特征用于语音情感识别,结果显示文本有关时,非线性特征性能优于 MFCC。文献[14]提出将基于 TEO 的非线性特征用于带噪语音情感的识别,实验结果证明上述特征具有较高鲁棒性。

2.1.3　语音的全极点模型特征

这类特征主要表现在语音频谱结构随时间的变化上,包含了特征参数的动态特性。代表性的特征参数是倒谱系数,如线性预测倒谱系数(LPCC)和 MFCC。LPCC 是基于语音信号为自回归信号的假设,利用线性预测分析获得倒谱系数。根据同态处理的概念和语音信号产生的模型,语音信号的倒谱等于激励信号的倒谱和声道传输函数的倒谱之和。通过分析激励信号的语音特点及声道传输函数的零极点分布情况可知,激励信号的倒谱分布范围很宽,而声道传输函数的倒谱主要分布于低时域中。考虑到不同情感的发音可能使声道有不同的变化,进而引起声道传输函数倒谱的变化,因而在语音情感识别中语音信号倒谱的低时域系数 LPCC 得到了应用。文献[15]采用 10 阶 LPCC 作为情感特征参数,文献[16]则应用 LPCC 的统计量作为特征参数。然而,LPCC 在所有的频率上是线性逼近语音的,这与人的听觉特性不一致,而且 LPCC 包含了语音高频部分的大部分噪声细节,使其抗噪声性能较差。针对以上的缺陷提出了 MFCC,并在语音情感识别领域得到广泛应用。文献[17~19]表明 MFCC 是一组有效的语音情感特征参数。

2.1.4　正弦语音模型特征

正弦语音模型已在多个语音处理领域获得了应用,近来这一模型在语音情感识别领域得到了研究。在这种模型中,语音信号被假设可以由一组不同频率、幅度和相位的正弦波之和表示,因此这组正弦波的频率、幅度和相位可以作为表达语音情感变化的特征参数。语音帧的正弦模型表示如下:

$$s(n) = \sum_{j=1}^{L} A_j \cos(2n\pi f_j/f_s + \phi_j) \tag{3}$$

其中:$s(n)$ 表示信号时域采样值;A_j 和 ϕ_j 分别表示第 j 个正弦波的幅度和相位;f_s 表示信号 $s(n)$ 的采样频率,$0 \le f_j \le f_s/2$;L 表示正弦模型的阶数。文献[20]基于上述正弦模型分别研究了幅度特征、频率特征以及相位特征与情感的变化特性,仿真结果表明上述三种特征可以有效地刻画语音情感的变化,并且性能优于常用的倒谱特征参数。

2.2　非基于模型的特征

这类特征通常由一帧或一段语音信号的各个时域采样直接计算一个特征矢量,常用的特征参数有语速[21]、短时平均过零率[22]、发音持续时间和能量[23]等。通常认为,欢快、愤怒、惊奇的发音长度和平静发音相比压缩了,而悲伤的发音长度却稍稍伸长了。从语速和情感的关系来看,欢快、愤怒、惊奇和平静发音相比变快了,而悲伤却变慢了。在提取持续时间时应注意包括无声部分,因为无声部分本身对情感是有贡献的。对于汉语而言,一个汉字即为一个音节,所以用总音节数除以持续时间即得到语速[1]。语音作为一种能量有限的信号,能量特征是其最重要的特征之一。从人们的直观感觉中就可感受到语音信号的能量特征与情感具有较强的相关性,如当人们愤怒时,发音的音量往往变大;而当人们悲伤时,往往讲话声音较低。语音帧的短时能量可用如下表达式表示:

$$E_n = \sum_{m=n-N+1}^{n} [x(m)w(n-m)]^2 \tag{4}$$

其中:$w(n)$ 为窗函数;$x(n)$ 为语音信号采样值。能量参数由于受录音设备和个人发音习惯影响较大,在实际运用中通常需要归一化处理。早期的能量特征多集中于原始信号采样的直接计算如式(4),随着小波分解、多带滤波器等子带分解方法的引入,一些新的子带能量分布特征逐渐被提出。如文献[24]基于多滤波器分解方法,提出一种新的短时能量特征称做对数频域能量系数(LFPC),仿真结果证明该特征优于常用的 MFCC 和 LPCC。文献[14]在特征 LFPC 的基础上,通过将 LFPC 减去均值进而生成新的特征参数,相对于原始 LFPC 的性能有了进一步的提高。

3 特征选择和降维方法

综上,从不同的角度理解语音,分别提出了不同的特征参数,但上述的任一类型特征都有各自的侧重点和适用范围,不同的特征之间具有一定的互补性。因此,相当多的文献采用了混合参数构成特征向量。但在特征融合时,并非特征参数越多越好,这是因为多特征之间除存在互补性外,还可能存在相关性,多特征融合时存在一个最佳的特征子集。另外从模式识别的研究也表明,识别率不与特征空间的维数成正比,在高维情况下泛化能力反而减弱,甚至导致维数灾难。现在解决此问题的方法是对高维特征向量进行特征选择或者降维。常用的特征选择方法有序列前向选择(SFS)[6,22,25]、序列后向选择(SBS)[26]、优先选择法(PFS)等。文献[27]针对普通话情感语音特征分别运用了 PFS、SFS、SBS 和逐步判别分析(stepwise discriminant analysis, SDA)进行特征选择,分析了特征个数和特征选择方法对平均准确率的影响,最后进行了特征选择的有效性分析。常用的降维方法有主成分分析法(PCA)[6]、线性判别分析(LDA)[19]等。近年来,关于特征降维又有新的方法,如文献[6]采用遗传算法进行特征选择,该算法的基本原理是模拟生物遗传特点,通过对原始特征集复制、变异等操作,最后在某种准则下获得最优特征子集。

这些方法在进行特征提取时各有优势,如 PCA 提取了最有代表性的特征,可以有效地消除冗余,降低维数,但它没有考虑不同类别数据之间的区分性。而 LDA 则通过最大化数据的类间离散度和最小化类内离散度来选择合适的投影方向,侧重于寻找具有最大分辨力的方向。特征选择方法比特征降维方法理论简单,容易理解,但其工作量繁琐。SFS 法考虑了所选特征与已选定特征之间的相关性,但它的主要缺点是一旦某特征已入选,即使由于后加入的特征使它变得冗余,也无法再将它剔除。SBS 在计算过程中可以估计每除去一个特征所造成的可分性的降低,与 SFS 相比,由于要在较大的变量集上计算可分性判据,其计算量要比 SFS 大。PFS 方法虽然不能得到最优的结果,但它能快速、方便地完成特征选择过程,在一些原始特征数量较大、可分性判据计算复杂的情况下,被普遍使用,在有些情况下它的综合效率比 SFS 和 SBS 都要高。基于智能算法的特征选择方法是一种较新的尝试,需作进一步研究。

4 语音情感识别算法

语音情感识别现在的处理思路仍然是把它作为典型的模式识别问题,所以到目前为止,几乎所有的模式识别算法都被应用其中。在这些方法中,有两大类方法是较为流行的:a)基于概率生成模型的方法如高斯混合模型(GMM)和隐马尔可夫模型(HMM);b)基于判别模型的方法,主要有支持向量机(SVM)和人工神经网络(ANN)。近来,一种新的解决思路是把上述若干模型融合起来,各自取长补短,形成混合模型。

4.1 隐马尔可夫模型(HMM)

HMM 是一种基于转移概率和传输概率的随机模型,由于它既能用短时模型即状态解决声学特性相对稳定段的描述,又能用状态转移规律刻画稳定段之间的时变过程,在基于时序特征的语音情感识别模型中,HMM 已成为研究人员广泛采用的模型。其中 HMM 的结构成为识别研究的重点。应用较多的模型结构有自左向右连续型 HMM 模型[7,28]、状态回跳连续 HMM 模型[13]、各态历经离散 HMM 模型[24]、自左向右半连续型 HMM 模型[29]。从文献研究结果来看,自左向右的状态转移结构适合文本相关的情感识别,各态历经的状态转移结构适合文本无关的情感识别。离散型模型相对简单,但其语音情感特征参数必须经过矢量量化(VQ)处理从而造成一些信息的丢失;另外,VQ 的码本训练和离散 HMM 的训练不是同时进行优化训练,因而很难保证训练的全局优化。连续型 HMM 模型避免了矢量量化的计算,可以直接处理特征参数,但为得到较精确的状态观察值的概率密度分布函数必须使用较多的概率密度函数进行混合,这样造成模型复杂、运算量大,并且需要足够多的训练数据才能得到可靠的模型参数。半连续型模型的特点介于上述两种模型之间。

采用 HMM 对语音进行情感识别,不是孤立地利用语音的时序特征,而是把这些特征和一个状态转移模型联系起来,它的合理性在于把情感的变化看做是语音时序特征动态变化,不同的情感可以由不同的 HMM 模型来表现。基于 HMM 的语音情感识别扩展性好,增加新样本不需要对所有的样本进行训练,只需训练新样本;缺点是模型结构参数的选择仍与待处理的语音数据有关,需由实验确定,并且训练时的计算量较大。

4.2 高斯混合模型(GMM)

GMM 本质上是一种多维概率密度函数,可以用来表示语音特征矢量的概率密度函数。它可以看做一种状态数为 1 的连续分布 HMM。通过对情感特征矢量聚类,把每一类看做是一个多维高斯分布函数;然后求出每一类的均值、协方差矩阵和出现的概率,将此作为每种情感的训练模板。识别时将测试矢量输入每种情感模板,最大后验概率即为识别结果。文献[30]在其情感识别实验中使用 GMM 识别七种情感状态,实验结果表明,GMM 的识别率高于采用短时特征矢量与 HMM 分类器的识别率。传统的 GMM 算法中,通常假设特征矢量之间是统计独立的,而事实上语音在发生过程中,特征矢量之间存在相互的制约关系,而矢量回归模型(VR)则可有效地描述矢量之间的相关性。文献[19]利用 VR 改进传统的 GMM,提出一种称为高斯混合回归模型(GMVAR)的分类器,作者还将 GMVAR 算法与 HMM、K 近邻算法及前向神经网络算法进行实验比较,结果表明 GMVAR 算法的识别效果明显优于其他三种算法。

GMM 的优点是可以平滑地逼近任意形状的概率密度函数,每个密度分布可以表示出基本声学类,并且模型稳定、参数容易处理;但 GMM 阶数和初值较难确定,特别是阶数很难从理论上推导出来,通常根据不同的语音样本由实验确定。

4.3 支持向量机(SVM)

支持向量机是贝尔实验室研究人员 Vapnik 等人在对统计学习理论进行了多年研究的基础上提出的一种全新的机器学习算法,该算法基于结构风险最小化原则,能够较好地解决小样本学习问题。由于 SVM 有统计学习理论作为坚实的数学基础,可以很好地克服维数灾难和过拟合等传统算法所不可避免的问题,近年来已成为一种有效的分类工具,并被广泛地应用于语音情感识别研究当中。文献[31]利用 SVM 把提取的韵律情感特征数据映射到高维空间,从而构建最优分类超平面实现对汉语普通话中生气、高兴、悲伤、惊奇四种主要情感类型的识别。计算机仿真实验结果表明,与已有的多种语音情感

识别方法相比，SVM 对情感识别取得的识别效果优于其他方法。SVM 通过确定类别之间的最优超平面实现分类,如果将以上机制变为寻找同类数据分布的最优超平面,则可获得一种基于 SVM 的新分类方法,即支持向量回归模型(SVR) 。文献[32] 应用 SVR 实现情感识别,此外作者还将 SVR 算法与模糊逻辑分类算法和模糊 K 近邻算法进行实验比较,结果表明 SVR 算法的识别率明显优于其他两种算法。

SVM 良好的分类性能在模式识别中得到了日益广泛的应用,然而,目前在 SVM 的训练和实现上仍然存在一些亟待解决的问题。SVM 中核函数的选择影响分类器的性能,如何根据语音样本数据选择和构造合适的核函数及确定核函数的参数等问题缺乏相应的理论指导,所以在多数文献中采用实验的方法进行确定。另外,虽然多类 SVM 的训练算法已被提出,但用于多分类问题的有效算法及多类 SVM 的优化设计等仍需进一步研究。

4.4　人工神经网络(ANN)

神经网络可视为大量相连的简单处理器(神经元) 构成的大规模并行计算系统,具有学习复杂的非线性输入输出关系的能力,可以利用训练过程来适应数据,对于模型和规则的依赖性较低。对于语音情感识别问题,根据使用的特征和情感分类的不同,可以使用不同的网络拓扑结构。文献[33] 使用了一种称为 all-class-in-one(ACON) 的网络拓扑结构,即为所有情感训练一个网络。他们认为利用两层的网络结构容易实现较为满意的近似映射,因此该网络包含与特征维数相同的输入节点、一个隐含层和与情感类别相同数目的输出节点。对每一个待识别的情感语句,将其特征矢量输入到网络中,再根据网络的输出判断其属于何种情感。文献[34] 使用了一种称为 one-class-in-one(OCON) 的网络拓扑结构,即为每一种情感训练一个子网络,每个子网络是一个多层感知器(MLP) 。将提取出的特征矢量输入到每一个子神经网络中,每个子网输出界于 0 ～1的数值,表示输入的参数矢量可能属于某种情感的似然程度,利用各个子网络的输出进行决策得出情感识别结果。

神经网络的自学习功能非常强大,由于语音样本特征向量与情感的许多规律进行显性的描述是困难的,而神经网络则可以通过反复学习的过程获得对这些规律的隐性表达,其在语音情感识别中具有独特的优势。为充分学习这些隐性规则,神经网络方法一般都采取了含有大量神经元的隐含中间层,从而导致复杂度和计算量较高。

4.5　混合模型

基于概率生成模型的方法能够反映同类数据本身的相似度特性,而判别模型的特点是寻找不同类别之间的最优化分类面来反映异类数据之间的差异。一些研究者将两者结合起来,用混合的识别模型进行情感识别。这种混合模型现已基本形成两类模式,即并联融合和串联融合。并联融合是将单项特征分别进行独立的匹配处理,得到各个匹配分数,通过融合算法将各匹配分数进行综合得到最终决策结果;串联融合是将前面分类器的输出作为后面分类器的输入,最终决策结果由后面分类器决定。文献[35] 提出了 GMM/K 最近邻的方法; 文献[6] 提出了 SVM/K-NN 的方法; 文献[36] 则提出了多分类器融合方法,所用的分类器包括 K 最近邻、加权 K 最近邻(WKNN) 、加权离散 K 最近邻(W-DKNN) 、加权平均 K 最近邻(WCAP) 及 SVM; 文献[37] 提出了 HMM/PNN 的方法。以上文献皆是将各个单独的分类器输出按照一定规则结合,属于并联融合方式。文献[18] 提出了 GMM/SVM 方法,它用 GMM 给出的概率信息作为特征参数,再用 SVM 进行训练与识别,属于串联融合方式。

HMM 和 GMM 是基于概率生成模型的方法,这类模型可以从统计的角度充分表示语音同类情感特征矢量的分布情况,具有较好的鲁棒性。但是概率生成模型只考虑同一类模式内部的相关性,而忽略了不同模式之间的区别,所以对于比较相近的情感,概率生成模型的区分能力较差。ANN 和 SVM 是基于判别模型的方法,这类模型是寻找不同类别之间的最优化分类面,由于它们利用了训练数据的类别标志信息,具有较好的识别性能。但其忽略了同类情感的特征相似性,这会导致识别结果过分依赖于不同情感类中的少数样本特征,进而造成识别错误。因此,两类模型在识别机理上有着很大的互补性。混合模型的优点是能对不同模型取长补短,将会在一定程度上使识别率得到提高;缺点是模型复杂、计算量大,并联融合通常需要实验来确定各分类器的加权系数,串联融合不能同步训练各个模型,因而很难获得全局最优混合模型。

5　结束语

本文对近年来语音情感识别领域的研究成果从情感特征、识别模型两个方面进行了总结。至今，有关语音情感识别的研究已经取得了丰硕成果，就其情感特征提取和识别算法而言,尚有许多问题需要探索和解决。将来的发展和热点可能会集中在以下几个方面:

a) 研究者们已分析了多种类型的特征与情感变化的关系特性,但就各类特征提取而言,不同的提取方法产生不同的特征精度,如基频的提取目前仍是一项开放的研究课题。因此,更加准确的特征提取方法有待进一步研究。

b) 由于语音情感变化引起语音的诸多特征发生变化，将多种特征混合起来可以更全面地表示情感。多类特征组合将是特征获取的一个研究方向。

c) 特征混合带来的最直接的问题是特征维数可能很高。模式识别研究表明,准确率不与特征空间的维数成正比,且在高维情况下分类器的泛化能力反而会减弱,甚至导致维数灾难。对语音情感进行高效识别,必须进行针对性的情感声学特征降维和选择等方法的研究。基于智能算法的特征选择方法作了一些尝试,但研究仍需深入。

d) 不同的训练和测试环境导致语音情感特征参数的变异,也使识别系统的性能明显降低,影响这种变异的因素包括环境、生理、心理、文化背景、语境、语义等。如何充分利用好这些影响情感的因素,有待深入地研究。

e) 高效、稳定的语音情感识别算法仍将是未来研究的热点,而将现有的几种主要算法各取所长、集成使用将有可能是解决该问题的有效途径。在这方面已有部分研究,有待进一步发展。

f) 部分文献的仿真结果虽然取得了较高的识别率,但鲜有文献从识别模型本身进行识别算法优劣的深层次理论分析。为识别模型的优劣提供理论支持有待研究。

参考文献:

[1]　余伶俐,蔡自兴,陈明义. 语音信号的情感特征分析与识别研究综述[J]. 电路与系统学报,2007,12(4):77-84.

[2] 林奕琳, 韦岗, 杨康才. 语音情感识别的研究进展[J]. 电路与系统学报, 2007, 12(1): 90-98.

[3] VERVERIDIS D, KOTROPOULOS C. Emotional speech recognition: resources, features, and methods[J]. Speech Communication, 2006, 48(9): 1162-1181.

[4] 田岚, 姜晓庆, 侯正信. 多语种下情感语音基频参数变化的统计分析[J]. 控制与决策, 2005, 20(11): 1311-1313.

[5] HYUN K H, KIM E H, KWAK Y K. Emotional feature extraction based on phoneme information for speech emotion recognition[C]//Proc of the 16th IEEE International Symposium on Robot & Human Interactive Communication. 2007: 802-806.

[6] MORRISON D, WANG Rui-li, De SILVA L C. Ensemble methods for spoken emotion recognition in call-centres[J]. Speech Communication, 2007, 49(2): 98-112.

[7] LI Xi, TAO Ji-dong, JOHNSON M T, *et al*. Stress and emotion classification using jitter and shimmer features[C]//Proc of IEEE International Conference on Acoustics, Speech, and Signal Processing. 2007: 1081-1084.

[8] 王治平, 赵力, 邹采荣. 基于基音参数规整及统计分布模型距离的语音情感识别[J]. 声学学报, 2006, 31(1): 28-34.

[9] PAO T L, CHEN Y T, YEH J H, *et al*. Mandarin emotional speech recognition based on SVM and NN[C]//Proc of the 18th International Conference on Pattern Recognition. Washington DC: IEEE Computer Society, 2006: 1096-1100.

[10] ZHAO Li, CAO Yu-jia, WANG Zhi-ping, *et al*. Speech emotional recognition using global and time sequence structure features with MMD[C]//Proc of the 1st International Conference on Affective Computing and Intelligent Interaction. Berlin: Springer, 2005: 311-318.

[11] KAISER J F. On a simple algorithm to calculate the energy of a signal[C]//Proc of IEEE International Conference on Acoustics, Speech, and Signal Processing. 1990: 381-384.

[12] NWE T L, FOO S W, DE SILVA L C. Classification of stress in speech using linear and nonlinear features[C]//Proc of IEEE International Conference on Acoustics, Speech, and Signal Processing. 2003: 9-12.

[13] GAO Hui, CHEN Shan-guang, SU Guang-chuan. Emotion classification of mandarin speech based on TEO nonlinear features[C]//Proc of the 8th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing. Washington DC: IEEE Computer Society, 2007: 394-398.

[14] 林奕琳. 基于语音信号的情感识别研究[D]. 广州: 华南理工大学, 2006.

[15] MAO Xia, ZHANG Bing, LUO Yi. Speech emotion recognition based on a hybrid of HMM/ANN[C]//Proc of the 7th WSEAS International Conference on Applied Informatics and Communications. Stevens Point: World Scientific and Engineering Academy and Society, 2007: 367-370.

[16] LIU Jia, CHEN Chun, BU Jia-jun, *et al*. Speech emotion recognition using an enhanced co-training algorithm[C]//Proc of IEEE International Conference on Multimedia and Expo. 2007: 999-1002.

[17] LUENGO I, NAVAS E, HERNAEZ I, *et al*. Automatic emotion recognition using prosodic parameters[C]//Proc of the 9th European Conference on Speech Communication and Technology. 2005: 493-496.

[18] HU Hao, XU Ming-xing, WU Wei. GMM supervector based SVM with spectral features for speech emotion recognition[C]//Proc of IEEE International Conference on Acoustics, Speech, and Signal Processing. 2007: 413-416.

[19] EL AYADI M M H, KAMEL M S, KARRAY F. Speech emotion recognition using Gaussian mixture vector autoregressive models[C]//Proc of IEEE International Conference on Acoustics, Speech, and Signal Processing. 2007: 957-960.

[20] RAMAMOHAN S, DANDAPAT S. Sinusoidal model-based analysis and classification of stressed speech[J]. IEEE Trans on Audio, Speech, and Language Processing, 2006, 14(3): 737-746.

[21] 赵力, 将春辉, 邹采荣, 等. 语音信号中的情感特征分析和识别的研究[J]. 电子学报, 2004, 32(4): 606-609.

[22] TABATABAEI T S, KRISHNAN S, GUERGACHI A. Emotion recognition using novel speech signal features[C]//Proc of IEEE International Symposium on Circuits and Systems. 2007: 345-348.

[23] 詹永照, 曹鹏. 语音情感特征提取和识别的研究与实现[J]. 江苏大学学报: 自然科学版, 2005, 26(1): 72-75.

[24] NWE T L, FOO S W, De SILVA L C. Speech emotion recognition using hidden Markov Models[J]. Speech Communication, 2003, 41(4): 603-623.

[25] LIN Yi-lin, WEI Gang. Speech emotion recognition based on HMM and SVM[C]//Proc of the 4th International Conference on Machine Learning and Cybernetics. 2005: 4898-4901.

[26] KWON O W, CHAN K, HAO J, *et al*. Emotion recognition by speech signals[C]//Proc of the 8th European Conference on Speech Communication and Technology. 2003: 125-128.

[27] 谢波, 陈岭, 陈根才, 等. 普通话语音情感识别的特征选择技术[J]. 浙江大学学报, 2007, 41(11): 1816-1822.

[28] KAMMOUN M, ELLOUZE N. Pitch and energy contribution in emotion and speaking styles recognition enhancement[C]//Proc of Multiconference on Computational Engineering in Systems Applications. 2006: 97-100.

[29] NOGUEIRAS A, MORENO A, BONAFONTE A. Speech emotion recognition using hidden Markov models[C]//Proc of the 7th European Conference on Speech Communication and Technology. 2001: 2679-2682.

[30] SCHULLER B, RIGOLL G, LANG M. Hidden Markov model-based speech emotion recognition[C]//Proc of IEEE International Conference on Acoustics, Speech, and Signal Processing. 2003: 1-4.

[31] 张石清, 赵知劲, 戴育良, 等. 支持向量机应用于语音情感识别的研究[J]. 声学技术, 2008, 27(1): 87-90.

[32] GRIMM M, KROSCHEL K, NARAYANAN S. Support vector regression for automatic recognition of spontaneous emotions in speech[C]//Proc of IEEE International Conference on Acoustics, Speech, and Signal Processing. 2007: 1085-1088.

[33] RAZAK A A, KOMIYA R, ABIDIN M I Z. Comparison between fuzzy and NN method for speech emotion recognition[C]//Proc of the 3rd International Conference on Information Technology and Applications. Washington DC: IEEE Computer Society, 2005: 297-302.

[34] LI Wu, ZHANG Yan-hui, FU Ying-zi. Speech emotion recognition in e-learning system based on affective computing[C]//Proc of the 3rd International Conference on Natural Computation. Washington DC: IEEE Computer Society, 2007: 809-813.

[35] KIM S, GEORGIOU P G, LEE S, *et al*. Real-time emotion detection system using speech: multi-modal fusion of different timescale features[C]//Proc of the 9th IEEE Workshop on Multimedia Signal Processing. 2007: 48-51.

[36] PAO T L, CHIEN C S, CHEN Y T, *et al*. Combination of multiple classifiers for improving emotion recognition in mandarin speech[C]//Proc of the 3rd International Conference on Intelligent Information Hiding and Multimedia Signal Processing. Washington DC: IEEE Computer Society, 2007: 35-38.

[37] 蒋丹宁, 蔡莲红. 基于语音声学特征的情感信息识别[J]. 清华大学学报: 自然科学版, 2006, 46(1): 86-89.