基于 SVM 的语音情感识别算法①

朱菊霞，吴小培，吕 钊

(安徽大学 计算智能与信号处理教育部重点实验室，合肥 230039)

摘 要：为有效提高语音情感识别系统的识别正确率，提出一种基于 SVM 的语音情感识别算法。该算法提取语音信号的能量、基音频率及共振峰等参数作为情感特征，采用 SVM(Support Vector Machine，支持向量机)方法对情感信号进行建模与识别。在仿真环境下的情感识别实验中，所提算法相比较人工神经网络的 ACON(All Class in one Network，“一对多”)和 OCON（One class in one network，“一对一”）方法识别正确率分别提高了 7.06%和 7.21%。实验结果表明基于 SVM 的语音情感识别算法能够对语音情感信号进行较好地识别。

关键词：SVM（支持向量机）；情感识别；语音信号；情感特征；人工神经网络

Speech Emotion Recognition Algorithm Based on SVM

ZHU Ju-Xia, WU Xiao-Pei, LV Zhao

(Key Laboratory of Intelligent Computing & Signal Processing, Anhui University, Hefei 230039, China)

Abstract: In order to improve recognition accuracy of the speech emotion recognition system effectively, a speech emotion recognition algorithm based on SVM is proposed. In the proposed algorithm, some parameters extracted from speech signals, such as: energy, pitch frequency and formant, are used as emotional features. Furthermore, an emotion recognition model is established with SVM method. Simulation environment experiential results reveal that the recognition ratio of the proposed algorithm obtains the relative increasing of 7.06% and 7.21% compared with artificial neural networks such as ACON (All Class in one Network, “one to many”) and OCON (One class in one network, “one to one”) methods. The result of the experiment shows that the speech emotion recognition algorithm based on SVM can improve the performance of the emotion recognition system effectively.

Keywords: SVM; emotion recognition; speech signal; emotional features; artificial neural network

1 引言

随着人机交互技术的进一步发展，人们已不满足当前计算机具备的智能能力。在这种强烈的需求推动下，情感识别成为当前国内外一个新的研究热点[1]。语音是人类传递情感的一种最直接的媒介[2]。因此，在让计算机识别人类情感状态的众多方法中，基于语音信号的情感识别是一种十分有效的途径[3]。基于语音信号的情感识别因涉及到不同语种之间的差异，发展不尽相同。国外的情感识别起步较早，也取得了一定的成果，但国内对情感识别的研究还处在刚刚起步的阶段。情感信息的重要特点之一就是状况依存性，各国民族习惯以及表达情感的方式不尽相同。因此，对情感信息的分析研究必定要结合特定的语言来展开。

用于语音信号情感识别的方法很多，比如：混合高斯模型法(GMM)[4,5]、隐马尔科夫模型法(HMM)[4,5]以及人工神经网络方法(ANN)[6]等。不同的情感识别方法各有特色。本文提出了基于 SVM 的语音情感识别算

① 基金项目:安徽省自然科学基金(090412261X);博士点基金(200803570002)
收稿时间:2010-09-02;收到修改稿时间:2010-11-26

法，该算法以语音信号的统计特征为参数。通过仿真实验可以得出该方法比 ANN 方法(ACON、OCON)的情感识别率高并且识别结果更稳定。

2　传统情感识别算法

语音情感识别也是一种模式识别。传统的情感识别方法很多，基本上可以分为两大类：一类是以时序特征为基础的，如 HMM 和 GMM 方法；另一类是以统计特征为基础的，如 ANN(ACON，OCON)方法。这两类方法的显著区别是基于统计特征的方法在情感模型训练上花费的时间要比基于时序特征的要少得多。

HMM 和 GMM 方法的基本原理相同，不同的是 GMM 方法比 HMM 方法简单。GMM 可以看作是单状态的 HMM。GMM 和 HMM 方法在各类情感模型的训练过程中需要大量的情感语音样本，同时模型训练的时间花费很大。因此，HMM 方法在情感识别使用方面有一定的局限性。ANN 方法以语音信号的统计特征为参数。它通过模拟人类大脑的机制从而具有线性网络没有的学习和理解能力。因此，ANN 方法在语音情感识别中得到广泛应用。然而，ANN 中各隐层节点数计算目前没有统一标准的方法，一般用的都是经验值，只有通过不断尝试来确定。但是由于尝试的次数有限，故最终网络的隐层节点数不一定是最佳的。隐层节点数越多网络的结构越复杂，这在一定程度上影响了整个网络的性能。ANN 方法中由于网络中隐层节点数等不确定性因素，限制了网络的鲁棒性和情感识别正确率的进一步提高。

本文提出的基于 SVM 的语音情感识别方法可以有效克服上述识别方法的不足。SVM 是一种性能优越的分类算法，具有性能稳定，抗干扰性强等特性。该算法以情感语音信号的统计特征为基础。具有比 ANN 方法更高的时效性，更稳定的情感识别结果。

3　基于SVM的情感识别算法

基于 SVM 的语音情感识别算法主要包括情感特征提取与情感模式识别两部分，原理框图如图 1 所示。

其中情感特征提取模块主要用来提取输入语音信号的情感特征，包括数字化预处理、端点检测以及情感特征计算三个步骤。情感模式识别模块主要由情感识别模型训练和情感识别两部分构成，其中情感识别模型训练部分主要采用 SVM 方法建立所需识别情感的模型，情感识别部分依据所建立的情感识别模型对测试语音信号进行情感分类以实现对说话人的情感识别。

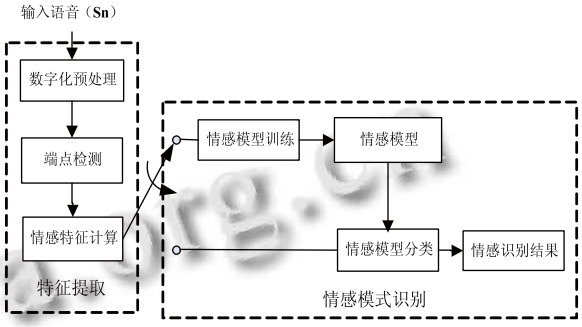

图 1　语音信号情感识别原理框图

3.1 情感特征提取

本文进行的是愤怒、高兴、平静、悲伤四种情感的分类，提取的情感特征为能量、基音频率、共振峰三类特征。

3.1.1 能量特征

语音信号的能量特征与情感的表达具有较强相关性[7,8]。语音信号能量通常有短时能量和短时平均幅度能量两类。由于短时能量计算量较大且对高电平敏感，本文采用短时平均幅度函数。

假设第 n 帧语音信号 $x_n(m)$的短时平均幅度函数为 M_n，则 M_n 的估计表达式为[7]：

$$M_n = \sum_{m=0}^{N-1} |x_n(m)| \tag{1}$$

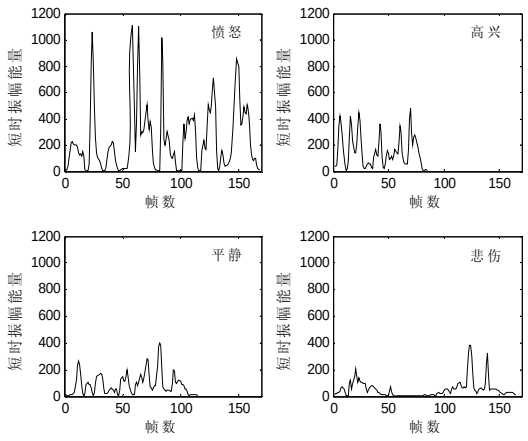

图 2　四种情感的短时能量曲线

其中 N 帧长。图 2 中四个子图分别展示同一文本在四种情感状态下的短时振幅能量曲线，其横坐标为语音的帧数，纵坐标为对应帧的短时振幅能量。

分析大量情感语音的能量信息可知愤怒和高兴情感的能量比悲伤的高。因此，通过能量特征可有效地区分愤怒、高兴与悲伤情感。

3.1.2 基音频率特征

发浊音时由声带振动而引起的周期性称为基音，声带振动频率称为基音频率。基音频率特征是反映语音情感信息[7]的重要特征。本文采用自相关法(ACF)[7]来提取基频特征。图 3 展示的是同一文本在四种情感状态下的基频曲线，其横坐标为语音信号帧数，纵坐标为对应帧的基频。

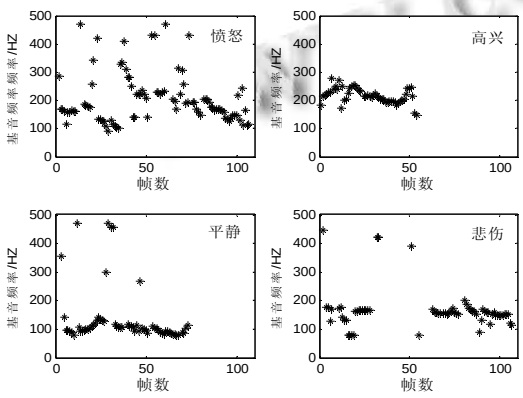

图 3 四种情感的基音频率曲线

观察图 3 可以发现，提取的基频存在一定数量的野点。本文实验中所用的基频特征是平滑操后的基频值。分析大量情感语音信号的基频特征得出愤怒和高兴情感基音频率的均值比平静和悲伤的大。因此，基音频率对本文四种情感的区分有贡献。

基音特性是浊音所特有的，故清浊音判决在基频提取中显得尤为重要。清浊音信号在能量上的差异，本文采用能量门限进行清浊音判决。能量门限 M 估计表达式为：

$$M=[\max(energy)-\min(energy)]/10+\min(energy) \quad (2)$$

其中，向量 energy 为一句语音的逐帧短时平均幅度能量，其各分量由式（1）估计。对某一帧语音而言，当其能量小于能量门限 M 时，判定为清音，直接令其基频为零。当其能量大于能量门限 M 则用自相关法计算其基频。图 4 展示了一句语音的原始波形以及经平滑后的基频值。

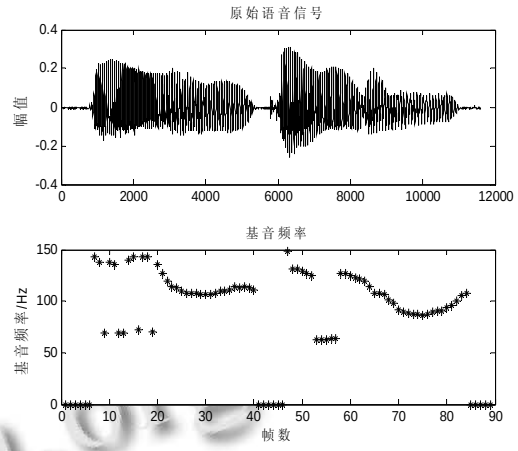

图 4 语音信号的基音频率曲线

3.1.3 共振峰特征

共振峰是反映声道特性的重要参数。不同情感在发音过程中声道呈现出不同的状态。不同情感发音的共振峰位置不同。故可以将共振峰特征作为情感分类的又一基本特征。图 5 展现的是同一文本在四种情感状态下的第一共振峰曲线。分析大量情感语音信号的第一共振峰数据得出与平静情感的语音相比，欢快和愤怒的第一共振峰值有所升高，悲伤情感的第一共振峰值有明显的下降趋势。

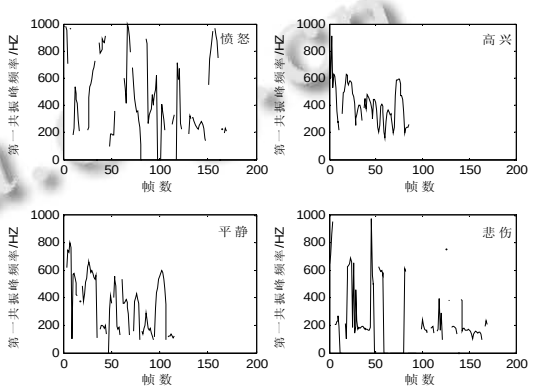

图 5 四种情感第一共振峰曲线

3.2 **SVM 情感识别算法**

SVM 解决非线性可分问题的方法是将低维输入特征向量非线性映射到一个高维特征向量空间中，从而将低维线性不可分问题转化成高维线性可分问题。非线性可分问题解决的关键仍是构建最优分类超平面[9]。最优超平面的构造最终转化为最优权值和偏置的计算。设训

练样本集合为：

$\{(X^1,d^1),(X^2,d^2),...(X^p,d^p),...(X^P,d^P)\}$,最小化权值 W 和松弛变量 ε_p 的代价函数为[9]：

$$\Phi(W,\varepsilon)=\frac{1}{2}W^TW+C\sum_{p=1}^{P}\varepsilon_p \quad (3)$$

其限制条件为：$d^p(W^TX^p+b)\geq 1-\varepsilon_p, p=1,2,...P$，其中 C 为惩罚因子。式（3）中引入松弛变量 $\varepsilon_p\geq 0, p=1,2,...,P$ 用于度量一个样本点相对于线性可分理想条件的偏离程度。由上述样本集合决定的最优权值和偏置的计算可转为其对偶问题[6]来解决。在特征空间中构造最优超平面时，仅使用特征空间中的内积，其核函数定义为：

$$K(X,X^p)=\Phi^T(X)\Phi(X)=\sum_{j=1}^{M}\phi_j(X)\phi_j(X^p), p=1,2,...,P \quad (4)$$

最优分类决策函数为：

$$f(x)=\mathrm{sgn}\left[\sum_{p=1}^{P}a_{0p}d^pK(X^p,X)+b_0\right] \quad (5)$$

SVM 应用于模式分类[9,10]问题一般解决方案有一对多和一对一[10]两种。本文采用一对一的分类方式来进行四种情感（愤怒、高兴、平静、悲伤）的分类，即选取四种情感中的任意两种情感特征训练构成一个 SVM 子分类器。因此，训练阶段共需构成 C_6^2 个情感识别子分类器，分别使用高兴与愤怒情感特征、高兴与平静情感特征、高兴与悲伤情感特征、愤怒与平静情感特征、愤怒与悲伤情感特征、平静与悲伤情感特征训练而成。识别阶段用训练好的各个情感识别子分类器识别未知情感状态的语音信号，每个子分类器都对其情感状态进行判别，最后将得到票数最多的情感作为待识别语音信号的情感状态。

SVM 分类器训练和识别之前均需为每句情感语音信号设计一个标签，用以表示该句情感语音信号所属的情感类别。标签的类型必须设为 double 型。分类器训练中惩罚因子 C 和核函数中的参数 g 的设定可以通过对训练集合的交叉验证来确定。但是需要注意采用这种方式确定的参数 C 和 g 仅是对训练集合识别效果最好，并不一定对测试集合识别效果也是最好的。通过反复实验测试，本文中参数 C 和 g 最终的值分别设为 6 和 0.0714。

4　情感识别试验及其结果

本文语音库为免费的柏林情感语音库[11]，其采样频率为 16KHZ，16bit 量化。该语音库共有 500 句情感语音信号，分别由十名专业演员(5 男，5 女)在不同情感状态下（高兴、愤怒、平静、悲伤、害怕、厌烦、憎恨）朗读十句不同文本的德语组成。本文选取其中的部分情感（高兴、愤怒、平静、悲伤）加以识别。仿真实验环境为 MATLAB7.0。选取的情感特征为与能量、基音频率、共振峰相关的特征，分别为：能量的最大值、最小值、均值、变化率、变化率的变化率；基音频率的最小值、变化范围、均值、方差、变化率、差分的方差；第一共振峰的均值、变化率、差分。为了降低不同人在表达不同情感时的个人差异造成的影响，本文实验过程中将提取的情感特征进行归一化处理。归一化采取将同一个人的四种情感语音信号的情感特征放在一起归一化处理，并将归一化后的情感特征作为 SVM 分类器的训练样本和测试样本。

通过随机选取的方式产生十个训练样本集合以及相应的测试样本集合。分别用每个训练集合训练一个 SVM 分类器模型，再用训练好的 SVM 对相应的测试集合测试。与训练集合相对应的十个测试集合的情感识别正确率如图 6 所示。为了验证 SVM 在情感识别应用中的有效性，本文采用基于 BP 算法的 ACON 网络和 OCON 网络作为对比试验。

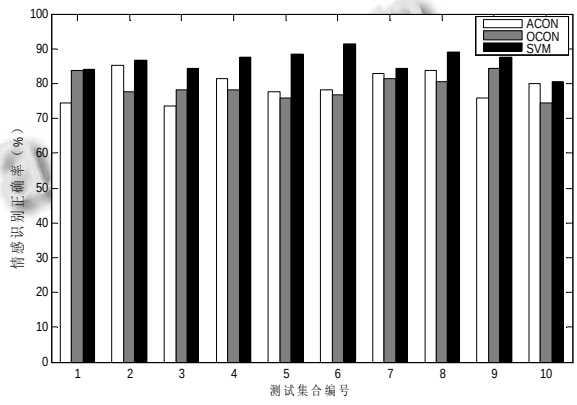

图 6　情感识别正确率

图 6 中横坐标表示的是测试样本集合编号，纵坐标为对应的测试样本集合的情感识别率，其中 ACON、OCON、SVM 的平均情感识别率分别为 79.30%、79.15%、86.36%。观察图 6 中 ACON、OCON 以及 SVM 的识别率可看出，SVM 分类模型比其它两种分类模型的识别率高而且更稳定，每次识别率可以达到

80%以上，其中最高的识别率可以达到 93.02%(高兴识别率：88.89%，平静识别率：94.74%，愤怒识别率：90.20%，悲伤识别率 100%)。这主要是由于 SVM 算法将问题转化为凸二次优化问题，得到的解为全局最优解，而基于 BP 算法的 ACON 网络和 OCON 网络得到的可能是局部最优解，而并非是全局最优解。

5 小结

本文采用 SVM 作为分类器实现了基于语音信号的四种情感(高兴、愤怒、平静、悲伤)识别，达到了 86.36%的平均识别率。对比 ACON 网络、OCON 网络，SVM 方法识别正确率分别提升了 7.06 %和 7.21 %。通过对比试验可知用 SVM 作为情感识别的分类器非常有效。但仍存在不足，比如高兴情感相对其他三种情感的识别率显得较低，主要是将高兴情感误判为愤怒情感。这可能是由于本文计算的情感特征对于区分高兴情感并不一定是最有效的。如从能量特征这方面来讲，高兴情感和愤怒情感的语音信号能量都比较大，这就不利于高兴情感和愤怒情感的区分。因此，提取有效的情感特征是情感识别正确的前提和基础，这有待于以后进一步研究。

参考文献

1 张石清,赵知劲,戴育良,杨广映,等.支持向量机应用于语音情感识别的研究.声学技术,2008,27(1):88–90.

2 Schroder M. Experimental study of affect bursts. Speech Communication, 2003,40(1-2):99–116.

3 韩纪庆,邵艳秋.基于语音信号的情感处理研究进展.语音技术,2006,5:58–62.

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

5 尤鸣宇.语音情感识别的关键技术研究[博士学位论文].杭州:浙江大学,2007.

6 Khanchandani KB, Hussain MA. Emotion recognition using multilayer perceptron and gengealized feed forward neural network. Journal of Scientific & Industrial Research, 2009, 68:367–371.

7 赵力.语音信号处理.北京:机械工业出版社,2008.36-80.

8 姜晓庆,田岚,崔国辉.多语种情感语音的韵律特征分析和情感识别研究.声学学报,2006,31(3):217–221.

9 边肇祺,张学工.模式识别.第 2 版.北京:清华大学出版社,2000.296–303.

10 Wang ZP, Zhao L, Zou CR.Support vector machines for emotion recognition in chinese speech. Journal of Southeast University, 2003,19(4):307–310.

11 Burkhardt F, Kienast M, Paeschke A, Weiss B. Berlin Database of Emotional Speech (Technical University, Institute for Speech and Communication, Department of Communication Science, Berlin). http://pascal.kgw.tu-berlin.de/emodb/

（上接第 68 页）

用了通用安全用户接口，使得这些入侵检测系统之间以及入侵检测系统和其他安全组件之间如何交换信息，共同协作来发现攻击、作出响应并阻止攻击。

4) 较低的成本[8]。基于云计算框架下的入侵检测系统通过云计算数据中心进行控制检测，并不需要在各种各样的主机上进行安装，大大减少了安全和管理的复杂性。

参考文献

1 张为民,唐剑峰.云计算深刻改变未来.北京:科学出版社,2010.30–31.

2 王鹏.云计算.北京:电子工业出版社,2010.5–6.

3 Ganame AK, Bourgeois J, Bidou R, Spies F. A global security architecture for intrusion detection on computer networks. Computers & Security, March 2008, 27: 30–47.

4 Liu XL. Research and application on Hierarchical Intrusion Detection [MS Thesis]. Shanghai Jiaotong University. January, 2008.

5 Sun Y, Huang Hao. Hybrid network intrusion detection system. Computer Engineering, 2008, 34(9).

6 Zhang LJ, Zhou Q. CCOA: Cloud Computing Open Architecture. IBM T.J. Watson Research Center, New York, USA, 2009 IEEE International Conference on Web Services, November 2009.

7 Muttik I, Barton C. Cloud security technologies. information security technical report. 2009 Elsevier Ltd All rights reserved. April 2009.

8 Christodorescu M, Sailer R, Schales DL. Cloud Security Is Not(Just) Virtualization Security. IBMT. J. Watson Research, September 2009.