基于遗传优化的多级 SVM 语音情感识别*

谈利芳 刘 蓉 黄 刚 张 雄
(华中师范大学物理科学与技术学院 武汉 430079)

摘 要：针对语音情感识别中特征维数高、识别率较低的问题，提出利用遗传算法进行特征降维，并构建二叉树结构的多级支持向量机(SVM)分类器进行语音多类情感识别的方案。首先对语音信号预处理后提取常用的情感特征，由于涉及特征较多，存在数据的冗余，采用遗传算法对提取的特征进行优化筛选；然后使用选出的最具情感区分能力的特征训练二叉树结构的多级 SVM 分类模型。在包含 7 种情感的柏林情感语料库上进行实验，结果证明提出的语音情感识别方案的有效性。

关键词：语音情感识别；遗传算法；特征降维；支持向量机
中图分类号：TP391.4；TN912.34 **文献标识码**：A **国家标准学科分类代码**：520.2040

Multi-layer SVM speech emotion recognition based on genetic optimization

Tan Lifang Liu Rong Huang Gang Zhang Xiong
(College of Physical Science and Technology, Central China Normal University, Wuhan 430079, China)

Abstract: Aiming at the problem of high feature dimension and low recognition rate in speech emotion recognition, In this paper, we propose a genetic algorithm for feature dimension reduction and construct a multi-layer SVM classifier based on binary tree structure for the recognition of speech emotion. First, the common emotional features are extracted after preprocessing the speech signal. As there are many features and redundant data, the genetic algorithm is used to optimize the extracted features. Then, the hierarchical SVM classification model of the binary tree structure is trained by using the most discriminative features. The experimental results demonstrate the effectiveness of the proposed speech emotion recognition scheme on the Berlin emotion corpus containing 7 emotions.

Keywords: speech emotion recognition; genetic algorithm(GA); feature dimension reduction; SVM

0 引 言

语言是人类交流信息的主要媒介，不仅包含丰富的语义信息，而且承载丰富的情感信息。听者可以通过语调高低、语速快慢等感受说话人的情感变化。如何使计算机自动并正确的从语音信号中识别出说话人的情感状态，是实现自然人机交互的关键，具有很大的研究和应用价值。

目前，越来越多的学者开展了语音情感识别的研究，并取得了一定的进展。如蒋庆斌等人[1]使用改进的高斯混合模型(GMM)，在自己录制的情感数据集上得到了 75.75% 的识别率；Milton 等人[2]使用 MFCC 参数构建 3 层 SVM，在柏林数据集上得到了 68% 的识别率；Li 等人[3]结合深度神经网络(DNN)和隐马尔可夫模型(HMM)，对柏林数据集中的 6 种情感进行识别，取得了 77.92% 的识别率；任浩等人[4]通过将 PCA 与 SVM 相结合构建多级 SVM 分类器，在柏林数据集上取得了 63.74% 的识别率；Chiou 等人[5]对提取的情感特征多次降维后，使用 SVM 进行语音情感识别，在柏林数据集上得到了 80.2% 的平均识别率。

语音情感识别通过对语音信号预处理后提取不同的情感特征，然后构建分类器进行情感分类。SVM 在处理非线性、小样本数据以及高维模式的分类问题中表现出特有的优势[6-8]，因而在语音情感识别领域受到广泛关注，但 SVM 不适合对数据建模。语音情感的变化是通过特征的差异来体现的，语音情感参数众多，选取有效的特征参数是语音情感识别的关键，本文采用遗传算法进行特征优化筛选，即从提取的若干特征中选出一些最有效的特征。SVM 属于二类分类器，本文通过构建二叉树结构的多级 SVM 将多类情感的分类分解为若干两类问题，不仅包含 SVM 高效分类的优势，还发挥了二叉树结构高效计算的特点。

1 情感特征提取

语音情感识别首先提取语音信号中的情感特征参数，

收稿日期：2017-03
* 基金项目：国家科技支持计划课题(2015BAK33B03)、湖北省科技馆特色创新展品展项的设计和建设项目资助

然后再进行分类模型的训练。因此,语音情感特征的提取是非常重要的环节。当前,用于语音情感识别的特征归纳为 3 类:韵律学特征、基于谱的相关特征和音质特征[9-10]。本文提取了其中常用的短时能量和基音频率(韵律特征)、梅尔倒谱系数 MFCC(基于谱的相关特征)以及共振峰(音质特征)。

1.1 短时能量

语音信号的能量与情感的表达有较强的相关性。假设语音信号经过窗函数分帧处理后得到的第 i 帧语音信号 $x_i(n)$ 的短时能量为 E_i,则 E_i 的估计表达式为:

$$E_i = \sum_{n=0}^{N-1} x_i^2(n) \tag{1}$$

式中:N 为帧长。

1.2 基音频率

当说话者发浊音时,声带会产生周期性的震动,震动的频率即为基频。基频描述语音激励源、反映语音情感的一个重要特征。本文采用短时平均幅度差函数法提取基频。第 i 帧语音信号 $x_i(n)$ 的短时平均幅度差函数定义为:

$$D_i(k) = \frac{1}{N}\sum_{n=1}^{N-k-1} x_i(n+k) - x_i(n) \tag{2}$$

式中:N 为帧长,k 为时间延迟量。

1.3 共振峰

人类发声的过程中,当声门处准周期脉冲激励进入声道时会产生一组共振频率,简称为共振峰。共振峰是反映声道特性的一个重要参数,也是反映语音情感的重要特征之一。本文采用线性预测法提取共振峰。首先计算出预测系数,再经过 FFT 运算求得功率谱,最后用峰值检测法检测出共振峰频率。

1.4 MFCC

MFCC 是一种充分利用人耳听觉感知特性的参数,具有较好的识别性能和抗噪能力。线性频率 f 的转换关系是:

$$Mel(f) = 2\ 595\lg\left(1+\frac{f}{700}\right) \tag{3}$$

这些特征以帧为单位进行提取,然后以全局特征统计值的形式参与建模和测试。本文提取出每个语音信号的短时能量,基因频率,第一、第二、第三共振峰,12 阶 MFCC 以及 MFCC 的一阶差分之后,计算出相应的统计特征值,包括最大值、最小值、均值、方差和标准差。这样,每个语音信号的特征向量为 145 维。

2 基于遗传算法的特征降维

语音情感识别中,提取的特征维数太高会导致特征匹配过于复杂,容易出现过拟合现象,不仅建模时间长,而且识别精度低。因此,在建立模型之前,有必要对输入的情感特征进行优化选择,将冗余的一些特征去掉,选择最能区分语音情感的特征参与建模。

特征降维方法有多种,常用的有主成分分析(PCA)、核主成分分析(KPCA)、线性判别分析(LDA)等,这些都是将高维度的特征经过某个函数映射至低维度作为新的特征,降维前后特征值被改变[11-13];而基于遗传算法的特征降维是从一组特征中选出一些最有效的特征,不改变所选特征的数值,能构造出较好的模型。

遗传算法(GA)是模拟生物进化过程的自然选择和遗传学机理的搜索全局最优解的方法[14-15]。遗传算法从经过编码的一个初始种群出发,采用基于适应度大小的选择策略模拟自然"优胜劣汰"法则选择优良个体,并借助于遗传算子进行组合交叉和变异操作产生新个体,使群体进化到搜索空间中越来越好的区域。种群不断繁衍进化,直到满足期望的终止条件,对末代种群的最优个体进行解码,便求得全局最优解。

基于遗传算法的语音情感特征降维流程图如图 1 所示。

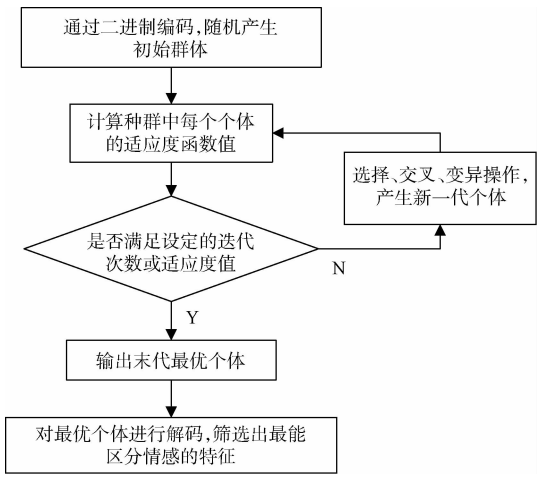

图 1 基于遗传算法的语音情感特征降维流程

利用遗传算法进行特征降维,首先需要将解空间映射到编码空间,本文采用二进制编码。个体的长度设计为语音样本的特征维数 145,每一位对应一维特征,每一位的取值是"1"或者"0",若某一位值为"1",表示该位对应的那一维特征参与建模与测试;反之,若某一位值为"0",则表示该位对应的那一维特征不参与建模与测试。用遗传算法进行语音情感特征降维的具体步骤如下:

步骤 1: 产生初始种群。按上述方法进行二进制编码,随机产生 N 个初始串结构数据构成一个种群,每个串结构数据即为一个个体。

步骤 2: 计算适应度函数值。对种群中的个体进行解码,求得参与建模的特征编号并对提取的情感特征进行降维,把降维后的训练集作为原始数据集采用三层交叉验证的方法获得语音的平均分类准确率,将此识别率便作为对应个体的适应度函数值。

步骤 3: 选择操作。本文采用基于相对适应度的

比例选择策略进行选择操作，即个体被选中的概率与个体的适应度值成正比。第 i 个个体的相对适应度 P_i 表达为：

$$P_i = \frac{f(i)}{\sum_{i=1}^{N} f(i)} \tag{4}$$

式中：N 为种群的大小，$f(i)$ 为第 i 个个体的适应度函数值。

步骤 4：　交叉操作。对群体中个体进行两两随机配对，对于每一对个体，以交叉概率交换它们某基因座之后的部分染色体。

步骤 5：　变异操作。对种群中的个体，以变异概率改变某基因座上的基因值，即“1”变为“0”或“0”变为“1”。

步骤 6：　重复步骤 2，计算新一代种群中每个个体的适应度函数值。

步骤 7：　判断是否满足终止条件，否，则进入步骤 3；是，则输出末代种群中的最优个体并进行解码，筛选出最具代表性的情感特征参与最终的建模与测试。

3　多级 SVM 识别模型

SVM 是一种二元分类器，其目标是找到一个分类最佳的平面，即使得属于两个不同类的数据点间隔最大的超平面。这是一个受限的凸优化问题，表达式如下：

$$\min\left(\frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\xi_i\right) \tag{5}$$

约束条件：

$$y_i(\boldsymbol{w}\cdot\boldsymbol{x}_i + b) + \xi_i \geqslant 1; \xi_i \geqslant 0; i = 1,2,\cdots,N$$

其中，$\boldsymbol{w}$ 为权重向量，b 为阈值；ξ 为松弛变量，C 为惩罚因子；N 为训练数据的个数，y_i 为样本 $\boldsymbol{x}_i$ 的标签，取值为 1 或 -1，$\boldsymbol{w}\cdot\boldsymbol{x}+b=0$ 为推倒出的超平面。最终解得分类超平面的判决函数为：

$$f(\boldsymbol{x}) = \mathrm{sgn}\left\{\sum_{i=1}^{N} y_i\alpha_i K(\boldsymbol{x}_i,\boldsymbol{x}) + b\right\} \tag{6}$$

式中：$\boldsymbol{\alpha}$ 为拉格朗日因子，$K(\ ,\ )$ 为核函数。

对于多类问题，可将其分解为若干 SVM 可直接求解的两类问题。在多类情感分类中，有些情感比较相似，而有些情感容易区分，所以采用“先易后难”原则，先将容易区分的情感分开，进行粗分类，然后对容易混淆的情感进行细分类，通过逐级分类，最终实现对所有情感的分类。本文通过构建二叉树结构的多级 SVM 进行多类情感的分类。

在二叉树的每个分支，情感标签被分割成两类，直到每个类只含有一种情感标签。为先将容易区分的情感分开，而将相似情感归为一组，本文采用距离法，即利用各类情感特征之间的欧式距离进行 K 均值聚类($K=2$)。多级 SVM 情感分类模型构建方法如下：

1)对属于同一种情感的样本特征，计算出其均值向量，这样，7 种情感分别对应于 7 个均值向量；

2)用 K 均值聚类算法把 7 种情感分成两大类，作为根节点下的两个分支；

3)逐个判断分支是否只包含一种情感，是，则结束此分支的分组；否，则继续对属于这一分支的情感进行 K 均值聚类；

4)重复步骤 3)，直到每个分支只含有一种情感。

最终的二叉树结构如图 2 所示，不难发现，情感标签在分支中被分成两类时，较相近的情感被分到了同一组，如第二层中 anger 和 happy 两种亢奋的情感被分到了一组，第四层中 disgust 和 boredom 两种相近情感被分到了一组，然后再对这些容易混淆的情感进行细分类，可大大降低分类误差。该结构对于 N 类问题，只需构造 $N-1$ 个 SVM，分类时也无需遍历所有的 $N-1$ 个 SVM 便能得到分类结果，识别速度较快。

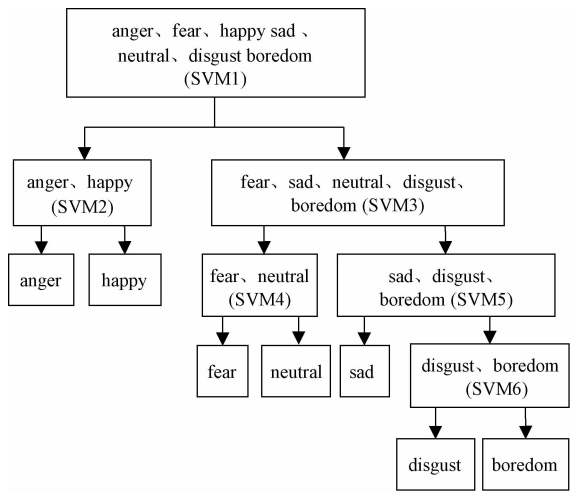

图 2　二叉树结构的多级 SVM

对于给定的二叉决策树，每个节点均进行 SVM 训练，其中只有 SVM1 用到了所有的训练数据，随着层次的增加，分类器所使用的训练数据越来越少。通过对比得知，使用线性核函数时提供了最高的分类精度，被用于所有的 SVM。在这种方式中，构建了 6 个不同的超平面，这 6 个超平面将特征空间划分为 7 个部分。

对未知样本 x 的分类过程为：从根节点开始，根据对应的决策函数大小来判定 x 的流向；对 x 所遍历的每个节点都依此进行判决，直到 x 被分类到叶节点中的某一种情感。此过程中，对未知样本 x 并不需要计算所有的决策函数值，从而节省了测试时间。

4　实验数据与结果分析

4.1　实验数据

本文实验采用柏林技术大学录制的柏林德语情感语料库。它包含生气(anger)、害怕(fear)、高兴(happy)、厌烦(boredom)、悲伤(sad)、平静(neutral)、厌恶(disgust)7 种

情感,共有 535 条情感语句。实验在 MATLABR2008a 环境下完成。

首先对输入的情感语音信号进行预加重、加窗分帧、端点检测等预处理,然后提取出语音信号的短时能量,基因频率,第一、第二、第三共振峰,12 阶 MFCC 以及 MFCC 的一阶差分,并计算出相应的统计特征值:最大值、最小值、均值、方差和标准差。由于生成的特征矩阵存在数据的冗余,为缩短建模时间,提高模型精度,使用遗传算法进行特征降维。之后随机将数据分成训练集和测试集,训练集用于训练多级 SVM 分类模型,然后用训练好的模型对测试集进行情感的识别。

由于特征向量中各维元素单位不统一,对 SVM 影响较大,通过对[0,1]和[−1,1]两种归一化进行对比后,选择在特征降维前对数据进行[−1,1]归一化。

4.2 实验结果与分析

本文实验都采用 10 折交叉验证,即重复实验 10 次,每一次实验随机的将数据分成训练集和测试集,但每一次实验将 80%的数据作为训练集,剩下的 20%作为测试集,最后将 10 次实验的平均值记为实验结果。其中,遗传算法的参数设置如下:迭代次数为 100,初始种群大小为 30,交叉概率和变异概率分别为 0.7 和 0.35。

测试样本在本文方法下的详细识别结果如表 1 所示。从表中可以看出,7 种情感的平均识别率为 78.5%,每种情感的识别率都处于合理的范围。其中,anger、fear、neutral、disgust 和 boredom 的识别率超过 80%;happy 的识别率最低,原因是 happy 容易被识别成与其相似的情感 anger。

表 1 测试样本详细识别结果

测试样本	识别结果								
	anger	fear	happy	sad	neutral	disgust	boredom	rate/%	average/%
anger	27	1	3	0	1	0	0	84.36	
fear	1	10	1	0	0	0	0	83.33	
happy	5	1	10	0	0	0	0	62.5	
sad	0	1	0	11	0	1	1	78.57	78.5
neutral	0	0	0	0	12	0	2	85.71	
disgust	0	0	0	0	0	7	1	87.5	
boredom	0	0	0	0	1	1	9	81.82	

将本文实验结果与相关研究结果进行对比,在柏林数据集下,文献[2]使用 MFCC 参数构建 3 层 SVM,得到了 68%的识别率;文献[3]结合 DNN 和 HMM,对柏林数据集中的 6 种情感进行识别,取得了 77.92%的识别率;文献[4]通过将 PCA 与 SVM 相结合构建多级 SVM,取得了 63.74%的识别率;文献[5]对提取的情感特征多次降维后,使用 SVM 进行语音情感识别,得到了 80.2%的平均识别率。相比之下,本文识别率处于较高水平,表明本文的语音情感识别方法是有效的。其中,文献[5]的识别率高于本文结果,但其经过减少三角回归特征的个数、减少特征组的个数、减少统计特征的个数、用 PCA 算法进行特征降维四个步骤逐步对提取的 6552 维参数进行降维,计算量庞大,过程复杂。为了有更强的对比和更深入的研究,对两者进行详细对比,结果如表 2 所示。

由表 2 可知,两种方法对于 7 种情感的识别率互有高低,本文害怕(fear)、高兴(happy)和平静(neutral)的识别率高于文献[5],两者都是高兴情感的识别率最低,但本文中高兴情感的识别率比文献[5]高 3.45%,原因是本文采取了将高兴和生气两种相似情感先归为一类后再进行细分类的方法,对彼此之间的误判有改善作用。

表 2 识别结果详细对比

情感	识别率	
	本文/%	文献[5]/%
anger	84.36	88.98
fear	83.33	76.81
happy	62.5	59.15
sad	78.57	80.64
neutral	85.71	78.48
disgust	87.5	89.13
boredom	81.82	83.95

5 结 论

针对语音情感识别中特征维数高、识别率较低的问题,本文提出利用遗传算法进行特征降维,并构建二叉树结构的多级 SVM 分类器进行语音多类情感识别的方案。在柏林情感语料库上进行的实验结果表明,本文提出的语音情感识别方案是有效的。但高兴情感的识别率仍不够理想,在今后的研究工作中,可以针对此问题作进一步的研究,降低相似情感之间的误判率,以达到更高的识别率。

参考文献

[1] 蒋庆斌，包永强，王浩，等. 基于改进 GMM 的耳语语音情感识别方法研究[J]. 计算机应用与软件，2012，29(11)：73-74.

[2] MILTON A，SHARMY R S，TAMIL S S. SVM scheme for speech emotion recognition using MFCC feature [J]. International Journal of Computer Applications，2013，69(9)：34-39.

[3] LI L，ZHAO Y，JIANG D，et al. Hybrid deep neural network--hidden markov model (dnn-hmm) based speech emotion recognition [C]. Affective Computing and Intelligent Interaction，IEEE，2013：312-317.

[4] 任浩，叶亮，李月，等. 基于多级 SVM 分类的语音情感识别算法[J]. 计算机应用研究，2017，34(6)：1-4.

[5] CHIOU B C，CHEN C P. Feature space dimension reduction in speech emotion recognition using support vector machine[C]. Signal and Information Processing Association Summit and Conference，IEEE，2013：1-6.

[6] 张一凡，余小清，安炫东. 基于改进 SVM 的纳税评估和预测[J]. 电子测量技术，2016，39(8)：79-84.

[7] 洪翠，杨华锋，卢国仪，等. 基于振动信号 SVM 分类的配变故障识别方法[J]. 仪器仪表学报，2016，37(6)：1299-1308.

[8] 何静，刘林凡，张昌凡，等. 参数优化的支持向量机机车车轮状态检测[J]. 电子测量与仪器学报，2016，30(11)：1709-1717.

[9] 韩文静，李海峰，阮华斌，等. 语音情感识别研究进展综述[J]. 软件学报，2014，25(1)：37-50.

[10] 赵力，黄程韦. 实用语音情感识别中的若干关键技术[J]. 数据采集与处理，2014，29(2)：157-170.

[11] MAO Q R，ZHAO X L，HUANG Z W，et al. Speaker-independent speech emotion recognition by fusion of functional and accompanying paralanguage features[J]. Frontiers of Information Technology & Electronic Engineering，2013，14(7)：573-582.

[12] JIANG J，WU Z，XU M，et al. Comparing feature dimension reduction algorithms for GMM-SVM based speech emotion recognition [C]. Signal and Information Processing Association Summit and Conference，IEEE，2013：1-4.

[13] 翟旭平，杨兵兵，孟田. 基于 PCA 和混合核函数 QPSO_SVM 频谱感知算法[J]. 电子测量技术，2016，39(9)：87-90.

[14] 秦勇，梁旭. 基于混合遗传算法的并行测试任务调度研究[J]. 国外电子测量技术，2016，35(9)：72-75.

[15] 刘浩然，赵翠香，李轩，等. 一种基于改进遗传算法的神经网络优化算法研究[J]. 仪器仪表学报，2016，37(7)：1573-1580.

作者简介

谈利芳，1990 年出生，硕士研究生，主要研究方向为语音情感识别。

E-mail：18086418168@163.com

刘蓉(通讯作者)，1969 年出生，副教授，硕士生导师，主要研究方向为智能信息处理、模式识别等。

E-mail：506111976@qq.com