

基于改进型 SVM 算法的语音情感识别

李书玲, 刘 蓉*, 张鏊钦, 刘 红

(华中师范大学 物理科学与技术学院, 武汉 430079)

(* 通信作者电子邮箱 liurong@phy.ccnu.edu.cn)

摘 要:为有效提高语音情感识别系统的识别率,研究分析了一种改进型的支持向量机(SVM)算法。该算法首先利用遗传算法对 SVM 参数惩罚因子和核函数中参数进行优化,然后用优化后的参数进行语音情感的建模与识别。在柏林数据集上进行 7 种和常用 5 种情感识别实验,取得了 91.03% 和 96.59% 的识别率,在汉语情感数据集上,取得了 97.67% 的识别率。实验结果表明该算法能够有效识别语音情感。

关键词:支持向量机;语音情感识别;语音信号;参数优化;遗传算法

中图分类号: TP18 **文献标志码:** A

Speech emotion recognition algorithm based on modified SVM

LI Shuling, LIU Rong*, ZHANG Liuqin, LIU Hong

(College of Physical Science and Technology, Central China Normal University, Wuhan Hubei 430079, China)

Abstract: In order to effectively improve the recognition accuracy of the speech emotion recognition system, an improved speech emotion recognition algorithm based on Support Vector Machine (SVM) was proposed. In the proposed algorithm, the SVM parameters, penalty factor and nuclear function parameter, were optimized with genetic algorithm. Furthermore, an emotion recognition model was established with SVM method. The performance of this algorithm was assessed by computer simulations, and 91.03% and 96.59% recognition rates were achieved respectively in seven-emotion recognition experiments and common five-emotion recognition experiments on the Berlin database. When the Chinese emotional database was used, the rate increased to 97.67%. The obtained results of the simulations demonstrate the validity of the proposed algorithm.

Key words: Support Vector Machine (SVM); speech emotion recognition; speech signal; parameter optimization; Genetic Algorithm (GA)

0 引言

语音是人们交流的主要方式,语音信号不仅传递语义信息,同时承载了说话人的情感状态。情感因素的引入能使人机交互变得更加自然和谐。因此,语音信号的情感识别成为近年来智能人机交互领域的研究热点^[1]。语音情感识别是让计算机通过语音信号识别说话者的情感状态,最终实现自然、友好、生动的人机交互。目前国内外学者在这方面进行了大量研究。如美国麻省理工学院(Massachusetts Institute of Technology, MIT)媒体实验室研究的情感机器人^[2],IBM 公司的“蓝眼计划”以及美国卡内基梅隆大学(Carnegie Mellon University, CMU)可穿戴计算机的研究开发,这些研究都为情感计算提供了一个较好的研究平台。国内的高校如哈尔滨工业大学、中国科学院计算技术研究所以及中国科学院自动化研究所等,也都在进行人机交互、情感识别方面的研究。

语音信号的情感识别方法很多,常用的情感分析方法有混合高斯分布模型(Gaussian Mixture Model, GMM)法、隐马尔可夫模型(Hidden Markov Model, HMM)、人工神经网络(Artificial Neural Network, ANN)、支持向量机(Support Vector Machine, SVM)以及在这几种识别方法上的改进和组合^[3]。Vlasenko 等^[4]在柏林数据集上采用 GMM 进行语音情感的识别,取得了 89.9% 的识别率。但 GMM 需要大量的情感语音

样本,且在情感模型的训练上花费时间较长。文献[5]采用改进的蛙跳算法(Shuffled Frog Leaping Algorithm, SFLA)神经网络进行语音情感的识别研究,得到了 84.2% 的识别率。但 ANN 中隐藏节点如何选取不确定,并且隐藏节点数目越多,网络的结构就会越复杂。文献[6]结合 HMM 和 ANN 两种算法在自制语料库上识别 6 种语音情感(平静、高兴、惊奇、愤怒、悲伤、恐惧),其识别率在 69.1% ~ 94.8%。文献[7]进行基于 SVM 的语音情感识别,在柏林数据集上获得了 86.36% 的识别率。SVM 情感分析方法在解决非线性、小样本以及高维模式识别表现出特有的优势,因而受到广泛的关注。但 SVM 核函数及其参数的选择,目前国际上还没有统一标准,一般是多次尝试取其经验值^[8]。

本文采用 SVM 算法进行语音信号情感识别实验。首先采用遗传算法(Genetic Algorithm, GA)对分类器训练中惩罚因子 c 和核函数中参数 g 进行优化,然后对语音情感信号建模与识别。在柏林数据集和汉语情感数据集进行情感识别,实验表明本文算法能有效识别语音情感。

1 语音情感特征提取

语音情感识别通过提取样本空间中能反映个人情感信息的一些特征参数,并用这些参数训练相应的模型以识别语音情感状态。其中,特征参数的提取起到关键作用。语音情感研

收稿日期:2013-01-07;修回日期:2013-02-15。

基金项目:国家社会科学基金资助项目(12BTQ038);国家自然科学基金资助项目(61202470)。

作者简介:李书玲(1988-),女,山东禹城人,硕士研究生,主要研究方向:语音情感识别;刘蓉(1969-),女,湖南安化人,副教授,主要研究方向:智能信息处理、模式识别;张鏊钦(1986-),女,湖北武汉人,硕士研究生,主要研究方向:模式识别;刘红(1988-),女,湖北孝感人,硕士研究生,主要研究方向:模式识别。

究提取的特征主要为韵律特征和音质特征。与韵律相关的特征有:基音周期、振幅能量、语速、持续时间等及其统计值^[9]。音质特征有:共振峰、梅尔频率倒谱系数(Mel-Frequency Cepstral Coefficient, MFCC)、线性预测系数(Linear Prediction Coefficient, LPC)等及其统计值^[10]。

在语音情感分类识别研究中,提取的特征无法简单地找出信号的特征规律,通常情况下需进行相应的数理统计。本文为振幅能量、基音频率、共振峰、MFCC 四个方面的情感特征参数,共计算了 5 个统计学变化参数,包括最大值、最小值、均值、方差和标准差。

1.1 能量特征

语音信号的能量与情感语音的表达有较强关联性,语音信号能量通常有短时能量和短时平均幅度能量两种。本文采用短时平均幅度法来提取振幅能量。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:

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

其中 N 为帧长。

1.2 基音频率特征

基音是指发语音时,由声带振动所引起的周期性。声带振动频率称为基频。基音频率反映了语音情感的重要特征。本文采用循环平均幅度差函数(Cycle Average Magnitude Difference Function, CAMDF)进行基音频率的提取。采用类似于循环卷积方式定义的 CAMDF 表达式为:

$$D(k) = \sum_{n=0}^{N-1} |sw(\text{mod}(n+k, N)) - sw(n)|; \quad k = 0, 1, \dots, N-1 \quad (2)$$

基音周期估计表达式为:

$$TP = \arg \min_k \left(\frac{TP_{\max}}{TP_{\min}}(D(k)) \right) \quad (3)$$

根据 CAMDF 函数的对称性和谷值点的函数值依次递增的性质,在 0 到 $N/2$ 点的范围里,搜索全局最小值即可对基音周期进行估计。

1.3 共振峰特征

共振峰是反映声道信息的重要参数,代表了发音信息最直接的来源。本文采用线性预测法求取了第一共振峰。首先估计出预测系数,然后估计声道的功率谱,再用峰值检测法检测出各共振峰的频率。

1.4 MFCC 特征

MFCC 是语音情感识别中常用的一种特征参数。它根据人耳的听觉特性,将频谱最终转化为倒谱域上的系数。MFCC 具有较好的识别性能和抗噪能力。

2 遗传算法分组优化 SVM 参数

2.1 SVM 算法原理

对于非线性可分的问题,支持向量机的基本思想是:利用核函数将输入特征的样本点映射到高维特征空间中,数据被超平面分割,在高维空间中变得线性可分,核函数以及超平面的构造是解决非线性可分问题的关键。核函数的形式及其参数的确定决定分类器类型和复杂程度。最优分类超平面保证不同类别能够正确分类,以使经验风险最小;保证分类间隔达到最大,以使置信度最小。找到一个具有最大间隔的分割平面是训练一个支持向量机的目标^[11]。

非线性可分的支持向量机优化问题转化为:

$$\min \left(\frac{1}{2} \|\omega\|^2 + c \sum_{i=1}^L \xi_i \right) \quad (4)$$

约束条件:

$$y_i(\omega \cdot x_i + b) + \xi_i \geq 1; \quad i = 1, 2, \dots, L, \xi_i \geq 0$$

其中: ω 为权系数向量, b 为分类阈值。 c 是错误惩罚因子,它控制着对错分样本的惩罚程度,影响对数据的拟合程度和模型的泛化能力。 ξ_i 是松弛变量,控制着离群点的进一步处理。数据空间样本点 x_i 和 x_j , 使用数据空间到特征空间的映射函数 Φ , 并应用核函数变换等式: $(x_i, x_j) \rightarrow K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$, 最终得分类超平面函数为:

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^L a_i y_i K(x_i, x) + b \right\} \quad (5)$$

其中 a_i 是拉格朗日因子。根据函数判断,输出分类结果。

核函数是支持向量机的关键,目前,一般常用的核函数有线性核函数、多项式核函数、径向基函数和多层感知机核函数等。选择核函数的理论依据较少,直接选用核函数或结合具体实例问题通过实验比较选取。本文在语音情感识别实验中,通过对比,选取了情感识别率最高的多项式核函数和线性核函数作为 SVM 核函数。

如何选择惩罚因子 c 、核函数类型及其参数 g , 成为训练一个 SVM 分类器的首要问题。

2.2 遗传算法参数优化设计

采用 SVM 进行语音情感识别的研究,惩罚因子 c 和核函数参数 g 一般通过实验或经验来进行确定。并且在分类训练时,通常是选用固定的参数值。样本集的大小、实验次数等对参数的选择有一定的影响且不同的训练集 SVM 参数也应不同。

遗传算法是一种将生物进化过程中的适者生存与群体内部染色体的随机信息交换机制相结合的高效全局寻优搜索算法。遗传算法参数优化是将需要优化的参数进行相应编码组成染色体,随机产生初始群体。遗传进化中,采用基于适应度函数的选择策略模拟生存法则“优胜劣汰”选择个体,使用交叉和变异来产生下一代种群,种群不断优化直到满足期望的终止条件。最后的一代染色体就作为全局最优解,经解码得到优化参数。

本文对于不同的训练集采用遗传算法进行参数优化以找到属于训练集的最优参数,然后进行 SVM 模型的训练与识别。图 1 为遗传算法优化 SVM 参数的语音情感识别实验流程图。

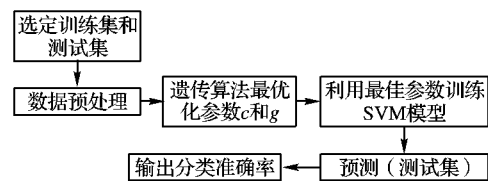


图1 遗传算法优化 SVM 参数流程

本文采用遗传算法进行 SVM 参数优化具体步骤如下:

步骤 1 参数初始化,并对 SVM 分类模型参数 c 和 g 进行二进制编码,每个变量用 20 个二进制位表示,然后随机生成初始种群。

步骤 2 对 c 和 g 解码,并代入 SVM 算法函数中,把训练得到的分类识别率作为适应度值。适应度越高的个体遗传到下一代的概率就较大,适应度越低的个体遗传到下一代的概率就较小。

步骤 3 选择操作,通过个体的适应度值在每一代进化中模拟“优胜劣汰”,从群体中选择优良个体,作为父代再产生新的群体。

步骤 4 交叉操作,挑选选择操作后的个体,按交叉概率来产生新个体。

在柏林数据集进行 5 种情感识别实验,实验结果如表 3。

从表 3 中可以知道,平均识别率为 96.59%,并且 5 种情感的识别率都很高。

表 3 柏林语音数据集上 5 种情感识别结果

测试样本	识别结果					
	angry	boredom	fear	happy	sad	rate/%
angry	126	1	0	0	0	99.21
boredom	1	80	0	0	0	98.77
fear	0	3	57	7	2	82.61
happy	0	0	0	71	0	100
sad	0	0	0	0	62	100
Precision/%	99.21	95.24	100	91.03	96.88	
Average rate/%	96.59					

同样的数据集,5 种情感识别率比 7 种情感的高。原因可能是,选择的 5 种情感是达尔文理论 6 种情感(愤怒、厌恶、恐惧、高兴、悲伤和惊讶)中的 5 种,人们日常生活中也易于表达这几种情感,采集的信息比较真实。这 5 种情感与生命体的基本特征息息相关,而且文化传统间的差异最小。因此,研究此类情感模型优于其他的模型。

在汉语情感数据集上进行相同的语音情感识别实验,结果如表 4 所示。

从表 4 中可以看到平均识别率为 97.67%,其中生气(angry)、惊讶(surprise)情感分类完全正确。

表 4 汉语情感数据集上情感识别结果

测试样本	识别结果						
	angry	fear	happy	neutral	sad	surprise	rate/%
angry	200	0	0	0	0	0	100
fear	6	190	4	0	0	0	95.00
happy	0	2	198	0	0	0	99.00
neutral	0	0	12	188	0	0	94.47
sad	0	0	0	1	196	3	98.00
surprise	0	0	0	0	0	200	100
precision/%	97.09	98.96	92.52	99.47	100	98.52	
Average rate/%	97.67						

3.4 结果分析

本文进行了基于遗传算法的分组优化 SVM 参数的语音情感识别实验。在柏林情感数据集上进行测试,7 种情感的平均识别率为 91.03%,5 种常见情感的平均识别率为 96.59%。对比国内外相关研究成果,文献[14]也在柏林数据集上进行,采用 Tree grammar inference 分类方法,7 种情感识别率为 78.58%,相比之下,本文提高了 12.45%,文献[15]在柏林情感数据集上进行多维输出 GMM 与 SVM 的语音情感识别,识别率为 80.4%。在汉语情感数据集上进行测试,本文取得了 96.59% 的识别结果。文献[16]与本文采用相同的汉语情感数据集进行了实验,识别结果只有 51.72%。分析原因是由于每组训练集都采用遗传算法找到最优参数,再采用 SVM 进行模型的训练和语音情感信号的识别。

4 结语

本文构建了有参数优化的 SVM 模型,在柏林数据集和汉语情感数据集上进行了语音情感识别实验,并与国内外的相关成果进行了比较。结果表明,本文通过遗传算法优化 SVM 参数惩罚因子 c 和核函数参数 g 进行语音情感识别,可以得

到较好的识别结果,充分证明了此种方法的有效性。但仍存在不足,由于每组训练集都要进行各自的遗传算法参数优化实验,会花费一定的时间。在进一步的工作中,缩短优化参数时间,以提高语音情感识别效率。

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): 76-84.
- [2] BANZIGER T, SCHERER K R. Using actor portrayals to systematically study multimodal emotion expression: the GEMEP corpus [C]//PAIVA A, PRADA R, PICARD R W. Proceedings of the 2nd International Conference on Affective Computing and Intelligent Interaction. Lisbon, Portugal: [s. n.], 2007: 476-487.
- [3] el AYADI M, KAMEL M S, KARRAY FAKHRI. Survey on speech emotion recognition: features, classification schemes and databases [J]. Pattern Recognition, 2011, 44(3): 572-587.
- [4] VLASENKO B, SCHULLER B, WENDEMUTH A, et al. Combining frame and turn-level information for robust recognition of emotions within speech [C]// Proceedings of International Conference on Spoken Language Processing. Antwerp, Belgium: [s. n.], 2007: 2225-2228.
- [5] 余华,黄程韦,金赞,等. 基于改进的蛙跳算法的神经网络在语音情感识别中的研究[J]. 信号处理, 2010, 26(9): 1294-1299.
- [6] 胡洋,蒲南江,吴黎慧,等. 基于 HMM 和 ANN 的语音情感识别研究[J]. 电子测试, 2011(8): 33-36.
- [7] 朱菊霞,吴小培,吕钊. 基于 SVM 的语音情感识别算法[J]. 计算机应用, 2011, 20(5): 87-91.
- [8] 史峰,王小川,郁磊,等. matlab 神经网络 30 个案例分析[M]. 北京: 航空航天大学出版社, 2010: 122-128.
- [9] HAN Z Y, LUN S X, WANG J. A study on speech emotion recognition based on CCBC and neural network [C]// Proceedings of the 2012 IEEE International Conference on Computer Science and Electronics Engineering. Washington, DC: IEEE Computer Society, 2012: 144-147.
- [10] JIA R, LI G, CHEN Y-P. Acoustic feature selection for automatic emotion recognition from speech [J]. Information Processing & Management, 2009, 45(3): 315-328.
- [11] DUDA R O, HART P E, STORK D G. 模式分类[M]. 2 版. 李宏东,姚天翔,译. 北京: 机械工业出版社, 中信出版社, 2010: 211-215.
- [12] BURKHARDT F, KIENAST M, PAESCHKE A, et al. Berlin database of emotional speech (Technical University, Institute for Speech and Communication, Department of Communication Science, Berlin) [DB/OL]. [2012-07-15]. <http://pascal.kgw.tu-berlin.de/emodb/>.
- [13] 中科院自动化研究所人机语音交互课题组. 汉语情感语料库 [DB/OL]. [2012-09-22]. <http://www.datatang.com/data/39277>.
- [14] SIDOROVA J. Speech emotion recognition with TGI+.2 classifier [C]// Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop. Stroudsburg: Association for Computational Linguistics, 2009: 54-60.
- [15] DONG F, ZHANG G B, HUANG Y M, et al. Speech emotion recognition based on multi-output GMM and SVM [C]// Chinese Conference on Pattern Recognition. Chongqing: [s. n.], 2010: 1-4.
- [16] 何琨,王国胤,杨勇. 基于选择性集成的情感识别[J]. 重庆邮电大学学报: 自然科学版, 2007, 19(4): 413-416.