

基于决策树的多特征语音情感识别

石 瑛,胡学钢,方 磊

(合肥工业大学 计算机与信息学院,安徽 合肥 230009)

摘 要:数据挖掘技术是近年来计算机领域的重要方向。文中的研究目的就是通过深入分析各种语音情感特征,找出某种特征对语音情感识别的贡献程度,并在数据挖掘技术中寻找适合的模型将有效特征加以利用。分析和研究了多位科学家在进行语音情感分析过程中采用的方法和技术,通过总结和创新建立了语音情感语料库,并成功地提取了相关的语音信号的特征。后研究了基音频率、振幅能量和共振峰等目前常用的情感特征在语音情感识别中的作用,把数据挖掘中常用的决策树分类方法和语音信号的多个特征相结合,建立了语音情感识别模型,对语音情感数据进行了大量的实验,取得了较为满意的识别结果。

关键词:语音情感识别;情感特征;数据挖掘;决策树

中图分类号:TP18

文献标识码:A

文章编号:1673-629X(2009)01-0147-03

Research of Speech Emotion Recognition Based on Decision Tree and Acoustic Features

SHI Ying, HU Xue-gang, FANG Lei

(School of Computer & Information, Hefei University of Technology, Hefei 230009, China)

Abstract:Data mining is one of the important computer technologies. The main goal of this thesis is to search the most useful features with analyzing the features related with emotions, and find a recognition model to make use of these features. Studied the method and technology in the research of the speech emotion recognition. Created the database of the speech emotion recognition and pick-up the features of the speech signal. Then study the effect in emotion-speech recognition from those common features such as pitch, amplitude energy, formant and so on. And also use decision tree with multi-features to recognize speech emotion, the result is perfect.

Key words:speech emotion recognition;emotion features;data mining;decision tree

0 引 言

随着计算机技术的发展,建立一个规范精确的语音交流的平台更是人机交互平台设计的重点。利用这种人与外界的交互作用,应用语音情感识别技术可以制作成交互电影、情感识别游戏和话务中心的情感识别^[1],心理咨询的系统软件等领域。20世纪80年代 Roddy Cowie 在 Activation-Evaluation 空间对语音情感进行了研究分类^[2]。同时科学家们也开始研究人类语音情感识别问题。利用提取的语音特征参数对情感进行分类,是语音情感识别的关键技术。面对人类大量的语音数据,传统的分析方法已经束手无策。从

1989年第11届国际联合人工智能学术会议上首次提出 KDD 这一概念以来,为语音情感识别解决了燃眉之急。语音情感识别与语音情感多个特征相关,而且每种特征参数对各种情感识别的贡献程度也是不尽相同的。而数据挖掘技术中的决策树分类方法的算法特性非常符合多特征语音情感识别的要求,文中就是利用决策树对语音情感识别进行研究。

1 研究现状

1.1 数据挖掘研究现状

数据挖掘是从大量数据中提取或“挖掘”知识。但是这种说法不够确切。有人把数据挖掘视为数据库中的知识发现即 KDD。KDD 是从数据中识别有效的、新颖的并且可能是有价值的可理解的模式的非凡过程。也有人认为数据挖掘是 KDD 过程中的一个阶段,由应用数据分析和发现算法在可接收的计算效率内产生数据的一个特定模式序列组成^[1]。

收稿日期:2008-04-03

基金项目:国家自然科学基金(10605022);安徽省自然科学基金(050420207)

作者简介:石 瑛(1974-),女,江苏溧阳人,黄山学院助理实验师,硕士研究生,研究方向为数据挖掘与语音识别;胡学钢,教授,博士,研究方向为人工智能与数据挖掘。

数据库中的知识发现是近年来随着数据库和人工智能技术的发展而出现的,一般将其分为九个处理阶段^[3]:准备、数据选择、数据预处理、数据缩减、确定 KDD 的目标、确定知识发现算法、数据挖掘、模式解释和知识评价。数据挖掘(Data Mining)是 KDD 中的一个处理步骤。由于数据挖掘在 KDD 中起到很大的作用,人们往往不加区分地使用两者。一般来说,在工程应用领域多称数据挖掘,而在研究领域人们则多称为数据库中的知识发现^[3]。

分类是学习一种函数能将数据项影射到几个预定的类别中的一个。分类的目的是构造一个分类函数或分类模型(分类器),该模型能把数据库中的数据项映射到某一个给定类别。决策树(Decision Tree)是一种最常用的分类知识描述模型,是一棵树结构,每个分支节点对应一个对象集合和一个属性选择,每个分支对应一个测试条件,叶子节点对应一个类别判断。

1.2 语音情感识别研究现状

在情感语音声学分析的基础上,情感语音识别的核心问题变成了寻找一个合适的模式识别的模型。各种模式识别方法,如线性判别分类器(Linear Discriminate Classifier)、支持向量机(Support Vector Machine)、高斯混合模型(Gaussian Mixtures)、人工神经网络(Artificial Neural Networks)和隐马尔可夫模型(Hidden Markov Model)等,都被应用于语音的情感识别,并取得了一定的效果。

Nicholson^[4]所研究的系统整个神经网络由 8 个子网构成,每个子网处理一种特定的情感,这种网络具有很好的可扩展性,增加情感类别时不需要重新训练整个网络。测试发现,负面的情感,如愤怒和悲伤容易识别,但正面的情感,比如喜悦,不易识别。

Dellaert 等^[5]采用了三种分类器,分别是最大似然贝叶斯分类法(Maximum Likelihood Bayes Classification)、核回归(Kernel Regression)和 K-近邻法。从基频轮廓(Pitch Contour)曲线提取特征参数,识别悲伤、愤怒、高兴和害怕,取得了 60%~65% 的识别率。选用了基频、前 2 个共振峰、能量和语速作为特征参数,分析和比较了几

类识别器,取得了大约 70% 的识别率。

Tato 等^[6]致力于研究情感空间,而不是韵律特征参数。他们的实验不仅使用了韵律特征参数(Prosody Features),还使用了音质特征参数(Quality Features)。实验结果表明,音质特征参数的使用能够有效地提高情感语音的识别率。

Yindong Yu^[7]采用 k 最近邻法识别高兴、愤怒、悲痛和自然四类情感。对基频、能量等相关的基本特征作了比较分析,从中发现了新的特征参数。并在此基础上对 k 最近邻法作了改进,使该识别系统所需的特征参数和训练数据较少,提高了其可扩展性。

综合比较各国研究人员的工作可以看出,三类声学特征在语音情感识别方面都起着比较重要的作用,尤其是韵律特征的应用更为广泛。仅仅单独用基频轮廓曲线进行特征提取,有的研究者就取得了 60% 左右的识别率;也有研究者单独应用反映频谱中能量分布的音质特征,取得了可以接受的识别结果。大多数的研究表明,将这三类特征有效地加以结合,能够明显地提高情感语音的识别效果^[8]。

2 基于决策树的多特征语音情感识别

2.1 语音情感识别的特征提取

要进行语音情感识别,提取语音情感特征是关键。图 1 是语音参数提取图,分别提取了语音情感的基音频率、语音强度和共振峰。

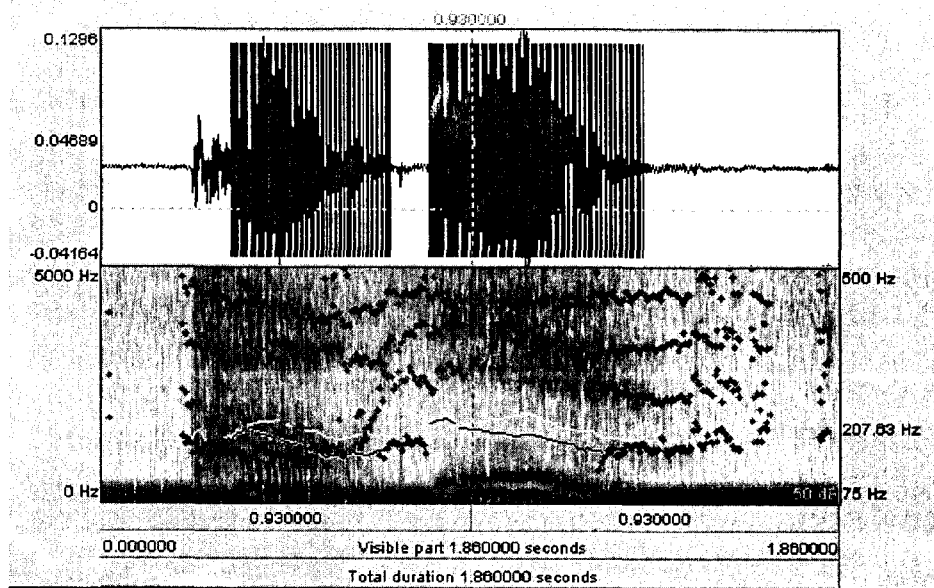


图 1 语音参数提取图

基频在表达情感方面起着非常重要的作用,它决定了音的高低,是声调、重音、语调等特征的声学基础。Cahn 等 MIT Media Lab 的研究人员认为,愤怒、高兴、害怕等情感对应着交感神经的激发,从而使得语音的

基频变高,基频的变化范围变大,变化速度加快;而悲伤、无聊等情感对应着副交感神经的激发,从而使得语音的基频变低,基频的变化范围变小,变化比较平缓^[9]。

声音质量在这里特指由喉部特性引起的声音特色,它通常包括频谱中的能量分布^[10],基频相邻周期之间的不规则变化情况,以及语音中噪音成分的强弱三方面。Tin Lay New 指出,一般地,愤怒、高兴等情感对应着尖锐明亮的声音质量,反映在频谱上为较强的高频成分;而悲伤则对应着较“柔和”的声音质量,在频谱上表现为高频成分的缺乏。害怕的情感语音中常常会出现基频的抖动。同时,在愤怒的语音中也常常会伴随着较强的噪声成分。

调音特征指的是声道对喉部激励信号进行调制的特征,描述了发音的精确程度,主要由共振峰参数来表示。调音特征主要指的是发音的精确程度。愤怒、害怕的情感对应着精确的发音,而悲伤的情感则对应着含糊的发音。在元音部分,调音特征主要由共振峰参数衡量。Schroder^[11]发现,当发音精确时,元音的第 1, 2 共振峰分别向着各元音的共振峰方向发展;而当发音模糊时,则向中央元音集中。Schroder 还发展了度量摩擦音发音精确程度的参数 SB(Spectral Balance),即以各频率成分的能量为权值的频谱频率中心。参数 SB 的值反映了声道中紧缩点的紧缩程度,紧缩程度越小,则 SB 的值越大,发音也就越精确。从表 1 可看出,第三共振峰对于四种语音情感均有一定的区分效果。

表 1 不同情感状态下第三共振峰特征统计

第三共振峰频率(Hz)	愤怒	高兴	悲伤	惊奇
均值	3068	3024	3105	2958
最大值	4069	4035	4259	4092
最小值	2218	2087	2104	1950
动态范围	1851	1948	2156	2142
标准差	345	341	354	416
1/3 中位数	2935	2917	2917	2807
2/3 中位数	3194	3170	3222	3164

2.2 语音情感特征分析

把语音情感的特征分析一下得到表 2。如果把基音频率作为决策树的第一个分类属性,就可以把悲伤这种情感分离出来。接下来依次使用振幅能量参数就可以把愤怒这种情感区分出来,最后再使用第三共振峰标准差这一属性就可以把六种情感成功区分开来。如图 2 所示是语音情感识别决策树的构建过程。

表 2 六种主要情感的声学特征

RID	1	2	3	4
基音频率	较高	较高	较低	较高
振幅能量	较高	非常高	稍低	较高
第三共振峰标准差	较低	较低	较低	较高

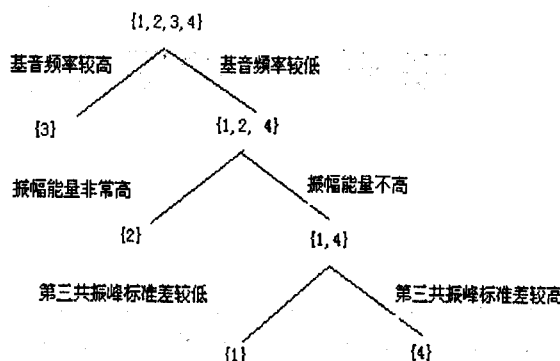


图 2 多特征语音情感识别决策树

2.3 实验结果和分析

按照建立语料库的一般规则,我们邀请了在 20 岁左右的大学学生参加录音(其中男生 5 人、女生 5 人),录音者均具有较高的情感表达能力。所录制语料经过 5 名非录音者进行的听辨实验,去除了其中近 50% 不能确定其情感种类的语料,挑选出了共计 1235 条可用语料,其中包含愤怒、高兴、悲伤、害怕四类情感语料各 300 句左右,组成了录制情感语音数据库。

基于决策树的多特征语音情感识别的实验是用 matlab 实现的。提取语音情感的基音频率、振幅能量和共振峰,运用决策树对提取的特征进行识别研究。识别的过程中,用语音数据的 2/3 作为训练数据,对训练数据的特征进行加权平均,剩下的 1/3 作为测试数据,以训练数据的平均特征值作为决策数判断结点,对四种情感分类。表 3 中是基于决策树的语音情感识别结果,取得了 66.9% 的平均识别率。

表 3 基于决策树的语音情感识别结果

情感类别	愤怒	高兴	悲伤	害怕	识别率
愤怒	68	11	4	20	66.0%
高兴	22	59	2	19	57.8%
悲伤	4	8	86	3	85.1%
害怕	20	19	4	61	58.7%

3 结束语

通过语音情感识别和数据挖掘技术的深入研究,决策树分类技术应用在语音识别中,把语音情感的多个特征作为识别标准,对语音的四种常见的情感进行识别研究。通过实验证实这种基于决策树的语音情感识别方法和传统的识别算法相比,有较高的识别率。

参考文献:

[1] Han Jiawei, Kamber M. 数据挖掘概念与技术[M].北京:机械工业出版社,2000:2-3.
 [2] Cowie R, Cowie E D, Savvidou S, et al. "FEELTRACE": An Instrument for Recording Perceived Emotion in Real Time [C]//ISCA Workshop on Speech & Emotion. Newcastle: [s.

(2) 若 OPEN 为空, 则失败并退出; 若非空, 取 OPEN 的第 1 个节点 n 移入 CLOSE;

(3) 若 n 是目标节点, 在搜索图 G 中回溯找出一条解决方案路径, 成功并退出; 反之, 扩展生成其后继节点集 M 加到 OPEN 中, n 的祖先不能在 M 中。对 M 的成员, 重新定位路径;

(4) 按照评价函数 f 计算每一个新扩展节点的估价值, 按递增 f 值重排 OPEN; 返回(2)。图 3 为最短路径算法; 从 A 点到 B 点的路径轨迹生成如图 4 所示。

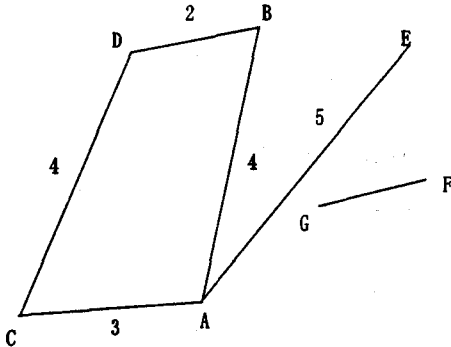


图 3 最短路径算法示意图



图 4 A -> B 路径轨迹自动生成示意图

4 结束语

介绍了将 GPS 与 ArcGIS 技术相结合, 构建多层次架构的城市水电气管理系统。系统采用了 A* 算法实现的最短路径优化算法很好地提高了整个系统的性能, 在实际应用中更加智能化, 更符合实际需求。将 GIS、GPS 技术和城市水电气等公共设施的管理结合起来能够提高市政管理的现代化和智能化的水平。随着城市现代化和信息化的高速发展, 基于 GIS 和 GPS 的技术将会得到越来越广泛的应用。

参考文献:

- [1] 万园, 程雄, 白钢, 等. 基于 GIS、GPS 与 GPRS 结合的物流管理系统设计与实现[J]. 测绘科学, 2006, 31(4): 67-69.
- [2] 田波, 周云轩, 李俊红, 等. 基于 .NET 和 ArcGIS Engine 开发上海市滩涂资源管理系统[J]. 计算机工程与应用, 2007, 43(14): 184-186.
- [3] 李静燕, 陈勇, 王志坚. 基于 Web Service 的分布式 GIS 集成研究[J]. 计算机工程, 2004, 30(1): 59-64.
- [4] 李宋, 王更生. GPS/GIS + GSM 在智能交通系统中的应用[J]. 计算机与现代化, 2004, 12(6): 127-129.
- [5] 荣玮. 基于道路网的最短路径算法的研究与实现[D]. 武汉: 武汉理工大学, 2005: 12-15.
- [6] 马德涛, 刘建忠, 王锐. 基于 ArcGIS Server 的城市道路最短路径分析方法研究[J]. 海洋测绘, 2007, 9(5): 58-61.
- [7] 张书亮, 闻国年, 李秀梅. 网络地理信息系统[M]. 北京: 科学出版社, 2005.
- [8] 张海涛, 程荫杭. 基于 A* 算法的全局路径搜索[J]. 微计算机信息, 2007, 23(6): 238-241.

(上接第 149 页)

- n.], 2000: 19-24.
- [3] Fayyad U M, Shapiro J P, Smyth P. From Data Mining to Knowledge Discovery[C]//Advance in Knowledge Discovery and Data Mining. [s. l.]: AAAI/MIT Press, 1996.
- [4] Nicholson J, Takahashi K, Nakatsu R. , Emotion Recognition in Speech Using Neural Network[J]. Neural Computing & Applications, 2000, 12(9): 290-296.
- [5] Petrushin V A. Emotion Recognition in Speech Signal: Experimental Study, Development, and Application[C]//Sixth International Conference on Spoken Language Processing. Beijing: [s. n.], 2000: 222-225.
- [6] Raquel T, Rocio S, Ralf K, et al. Emotional Space Improves Emotion Recognition[C]//In Proc of ICSIP. Denver, Colorado: [s. n.], 2002: 345-347.
- [7] Yu Yindong, Chang E, Li Cong. Computer Recognition of Emotion in Speech[C]// The 2002 Intel International Science and Engineering Fair. Louisville, Kentucky, United States of America: [s. n.], 2002.
- [8] 郭春宇. 语音情感识别技术的研究[D]. 哈尔滨: 哈尔滨工业大学, 2006.
- [9] Cahn J E. Generating Expression in Synthesized Speech[D]. [s. l.]: MIT Media Lab, 1990.
- [10] Nwe T L, Say Wei Foo, De Silva L C. Speech Emotion Recognition Using Hidden Markov Models[J]. Speech Communication, 2003, 41: 603-623.
- [11] Schroder M, Cowie R, Douglas-Cowie E, et al. Acoustic Correlates of Emotion Dimensions in View of Speech Synthesis [C]//ISCA Workshop on Speech and Emotion. North Ireland: [s. n.], 2000: 87-90.