

# 家庭保健监测系统中环境声音事件的识别\*

李玲俐

(广东司法警官职业学院 信息管理系, 广州 510520)

**摘要:**用于音乐和语音的识别方法不适用于非结构化环境声音事件的识别。提出一种基于隐马尔可夫模型(HMM)和支持向量机(SVM)的二级分类策略,对家庭保健监测系统中的语音、警报声、电话铃声、笑声、尖叫声和咳嗽声等6种声音事件进行识别。首先,提取Mel频率倒谱系数(MFCCs)来分析环境声音信号。其次,以提取的MFCC特征为输入,依次采用HMM和SVM构造二级分类模型,通过识别和确认两个过程对家庭保健监测系统的环境声音事件进行识别。实验证明,该技术能提高家庭保健监测系统中易混淆环境声音的识别率。

**关键词:**环境声音;隐马尔可夫模型;支持向量机;Mel频率倒谱系数

**中图分类号:**TP393.09

**文献标志码:**A

**文章编号:**1672-6693(2016)04-0118-05

声音作为人类感知环境信息的来源之一,能根据当前发生的事件获得一些关于声音分类的重要特征。这里指的声音不只是语音,也包括其他各类环境声音事件。所谓环境声音事件,是指在特定的环境中,一些物体不规则震动,引起对人类具有警示意义的声音<sup>[1]</sup>,如开/关门声、电话铃声、冲水声和咳嗽声等。近年来,随着医疗保健、预测生态环境变化、安全监控等应用的增强,环境声音的识别越来越重要。环境声音的特征提取和识别率决定了系统的性能,研究者对环境声音事件的识别技术展开了深入细致的研究,并取得丰富的成果。Barry S J等人<sup>[2]</sup>将提取的特征谱系数作为输入,采用概率神经网络(Probabilistic neural network, PNN)作为分类器,对于咳嗽声信号的识别,获得了良好的分类效果。Chen J等人<sup>[3]</sup>提出一种浴室活动自动监控系统,采用Mel频率倒谱系数(Mel frequency cepstrum coefficients, MFCC)特征和隐马尔可夫模型(Hidden Markov model, HMM)作为分类器,对浴室声音事件进行分类识别。Diment A等人<sup>[4]</sup>提出了一个基于HMM分类器的系统,该系统对AASP挑战数据集中提供的声学场景和事件进行检测和分类<sup>[5]</sup>,实验结果显示,准确率、召回率和F-Score分别为19.84%,45.28%和19.8%。在另一项研究中,Roma G等人<sup>[6]</sup>设计了一个同样使用AASP挑战数据集的基于MFCC训练一个支持向量机<sup>[7-9]</sup>(Support vector machine, SVM)分类器进行评估的系统。实验报告显示,基于事件的准确率、召回率和F测试值分别为6.71%,31.43%和11.06%。文献<sup>[10]</sup>针对办公环境中的抽屜声、键盘、鼠标和打印机等16种音频事件,使用MFCC特征和SVM分类器,采用5折交叉验证方法进行实验,得到较高的精准率。

虽然取得了一些研究成果,但由于环境声音事件的多样性和变异性,用于音乐和语音信号的识别方法不适用于环境声音的识别。而且,不同环境下出现的声音复杂多样,发声声源种类多样,多种声音会叠加、反射等<sup>[1]</sup>。因此,仍然需要针对各种不同的环境声音事件,对其特征提取与识别技术进行深入研究。最近的研究表明,环境声音事件的识别可以通过适当的机器学习方法改进<sup>[6,11-13]</sup>。家庭保健监测系统主要用于对高龄人群进行日常生活监测,通过各种环境声音的判断了解独自在家的老年人生活情况,该系统成本低、安全、高效。本文从准确性、复杂性和鲁棒性3个方面进行综合考虑,提出了一种采用MFCC特征,利用HMM和SVM二级分类器,实现对家庭保健监测系统中普遍存在的6类环境声音(语音、警报声、电话铃声、笑声、尖叫声、咳嗽声)进行快速准确的识别。

## 1 系统总体框架

本文对家庭保健检测系统中常见的6种环境声音进行识别,系统总体框架如图1所示,主要包括预处理、特

\* 收稿日期:2015-12-31 网络出版时间:2016-07-07 16:34

作者简介:李玲俐,女,副教授,研究方向为数据挖掘与模式识别,E-mail:mmonly@163.com

网络出版地址:<http://www.cnki.net/kcms/detail/50.1165.N.20160707.1634.042.html>

征提取和识别分类 3 个部分。

### 1.1 预处理

通常采用数字滤波器进行预加重处理。这里选用数字滤波器传递函数  $H(z) = 1 - uz^{-1}$ , 其中参数  $u$  取值为 0.937 5。声音信号经过数字滤波器后会减少低频噪声影响, 调整高频信号。

### 1.2 MFCC 特征提取

表征声音内容的特征参数的选取和分类器的实现是环境声音识别技术研究两个主要内容, 其中, 特征信息的提取是对环境声音进行分类的关键问题, 特征参数的好坏直接影响到分类精度。MFCC 声音信号特征是采用一种非线性 Mel 频率尺度的实对数短时能量的余弦变换, 不同频率的声音在耳内基础膜的不同位置振动时, 人耳就可以很快速分辨出各种声音。MFCC 比其他方法具有更好的抗噪性, 是因为其特有的非线性变换; 其也时常作为有关距离的度量, 所以, MFCC 在预测系统和声音的识别系统中都有着良好的应用前景<sup>[14]</sup>。

MFCCs 的计算是以“bark”为频率基准, Mel 频率与线性频率  $f$  (单位为 Hz) 的关系为:

$$Mel(f) = 2595 \lg \left( 1 + \frac{f}{700} \right)。$$

滤波器通带的分布主要考虑的是 Mel 频率对人耳听觉特性的描述, 不是只对语音, 而是对所有例如语音、音乐、环境声音等可听声音。MFCCs 也是非语音音频识别中常用的特征<sup>[15-16]</sup>。

MFCCs 参数的计算流程如下:

1) 短时间内的环境声音信号才是平稳信号, 因此, 需要先将声音信号划分为一个个的短时段, 每个短时段为 1 帧。本文对每个片段加 Hamming 窗形成帧, 每一帧包含  $N = 256$  个采样点, 且每 2 个相邻帧之间设置一段重叠区域以避免丢失音频信号动态变化信息, 该重叠区域长度设置为 172。

2) 假设环境声音数据时域信号为  $x(n)$ , 第  $i$  帧音频信号  $x_i(n)$  可以表示成  $x_i(n) = x(i * N + n)w(n)$ ,  $0 \leq n \leq N - 1$ , 这里  $w(n)$  是 Hamming 窗。  $w(n) = 0.54 - 0.46 \cos \left[ \frac{2\pi n}{N-1} \right]$ ,  $0 \leq n \leq N$ 。

对每帧信号  $x(n)$  做快速傅立叶变换 (FFT) 后, 对其取模的平方得到离散功率谱  $P(n)$ 。

3) 在语音的频率范围内设置若干个重叠的三角形频带组成的滤波器组:  $H_m(n)$ ,  $m = 0, 1, \dots, M - 1$ ,  $n = 0, 1, \dots, \frac{N}{2} - 1$ , 其中,  $m$  为滤波器的个数,  $N$  为一帧语音信号的点数, 这里为了计算方便, 取  $N = 256$ 。三角形滤波器的中心频率为  $f_m$ , 每个滤波器的跨度在 Mel 频率轴上相等。

4) 计算  $P(n)$  和  $H_m(n)$  在各离散频率点上的乘积之和, 也就是  $P(n)$  通过  $M$  个  $H_m(n)$  后得到的功率值, 即  $M$  个参数  $P_m$ ,  $m = 0, 1, \dots, M - 1$ 。

5) 计算出  $P_m$  自然对数, 再进一步做离散余弦变换 (DCT) 即可得到一组 MFCC 系数。

## 2 分类器设计

隐马尔可夫模型 (HMM) 作为语音信号的一种统计模型, 对动态时间序列有很强的建模能力, 选择参数、结构和训练方法时非常灵活, 基于 HMM 的方法主要用于大词汇量的语音识别, 适合处理连续信号, 在语音处理的各领域中得到广泛的应用。在其他识别领域, HMM 通常只考虑系统处于当前时刻的状态, 会错误地判断一部分正确的样本, 其识别结果反映了同类样本的相似度, 所以存在一定的误差, 类别区分能力较弱, 识别率很难达到很高。

支持向量机 (SVM) 建立在统计学习理论上, 基于结构风险最小化原理, 是为了解决分类和回归问题提出的一种机器学习方法。SVM 是一个凸二次优化问题, 所求的局部最优解就是全局最优解, 其输出结果体现了异类样本间的差异。SVM 对小样本数据具有良好的泛化性和较好的分类精度, 能够有效地解决学习问题。但

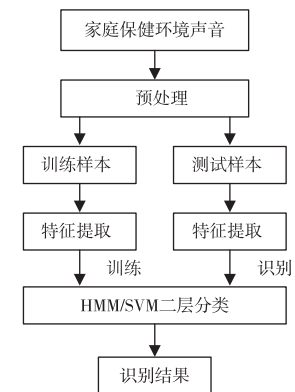


图 1 医疗环境声音识别系统总体框架

Fig. 1 General framework of environmental sounds identification system for family health monitoring system

是当样本容量较大的时候,SVM 会导致训练和识别时间长,识别率下降。

基于 HMM 和 SVM 的优缺点,本文将二者结合构成两级分类器,以提高环境声音的识别率,如图 2 所示。

### 2.1 HMM 识别器的设计

采集的环境声音信号输入后,首先作预处理和特征提取,本仿真实验中,设计了 6 个 HMM 模型,分别对应语音、警报声、电话铃声、笑声、尖叫声和咳嗽声,将训练集的每一种环境声音训练样本数据的 MFCC 特征向量数据集倒入 6 个 HMM 模型,记为  $\lambda_i, i=1, \dots, 6$ ,分别对 HMM 分类器进行训练。基于 HMM 模型的语音识别器训练过程如图 3 所示<sup>[17]</sup>,每个分类器都输出一个待分类的环境声音属于该类的概率给定一个判别声音信号的样本,利用前向算法求出每一个观察序列  $O$  在各个 HMM 模型中  $\lambda_i$  的输出概率  $P(O|\lambda_i)$ ,选出可能性最大的两类,放入对应的 SVM 分类器进一步确定声音所属的类别。

### 2.2 SVM 识别器的设计

分类识别的第二层,SVM 对 HMM 识别器输出结果中不可靠的易混淆的环境声音事件进行再次识别,这是一个典型的多分类问题。众所周知,SVM 是个二(类)分类器,但通过使用不同的技术可使它能应用于多(类)分类问题。传统 SVM 只能对两类样本进行分类,不能处理多分类问题。SVM 在处理多分类问题上的通常采用“一对一”的判别策略,具有更好的分类推广能力,是类别数较少情况下的最佳选择。在判别策略中,任意两类间都需要训练一个 SVM 分类器,因此  $N$  类样本中,共需要  $N(N-1)/2$  个分类器。如图 2 所示,6 种声音事件通过 6 个 HMM 分类器后,得到特征分别属于 6 种声音事件的概率,从中选取概率最大值和次大值的两个,用相应的 SVM 分类器进行确认。本实验需要 15 个 SVM 分类器,以对任意两种声音事件进行分类。所有的分类器训练完成后,每一对都可以被正确区分。

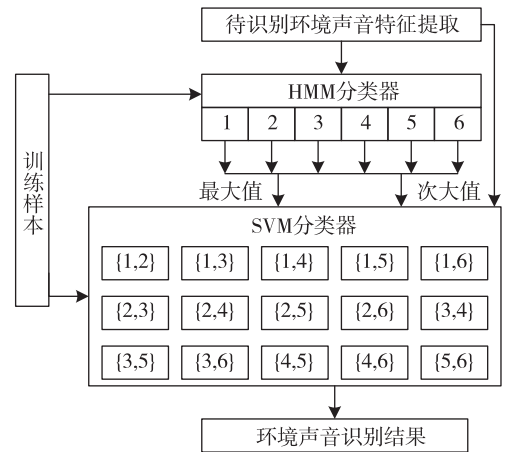


图 2 两级分类结构图

Fig. 2 Two-level classification structure diagram

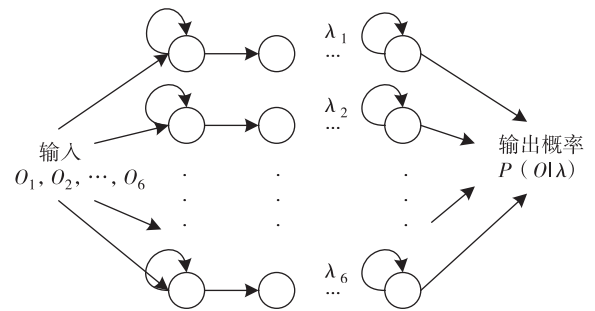


图 3 基于隐马尔可夫模型的语音识别器训练过程

Fig. 3 Speech recognition training process based on HMM

## 3 仿真实验与结果分析

### 3.1 环境声音事件采样

本文针对 6 种环境声音事件,对每种声音的特征进行仿真实验识别。实验选取平台和参数为:1) 软件平台:Matlab 2014;2) 采样率:8 kHz;3) 每种声音事件包括 20 个采样,每个采样包括目标声音和一些接近静音的背景声音,持续时间如表 1 所示。并对每段声音进行移动分帧处理,帧长 256,帧移 172。抽取其中的 2/3 采样作为训练集,1/3 的采样作为测试集。

### 3.2 实验结果

识别的准确度采用识别精度(Accuracy)来衡量:

$$\text{正确识别率} = \frac{\text{识别正确的样本数}}{\text{所有参与识别的样本数}} \times 100\%$$

仿真试验中,使用 MFCCs 特征,单独采用 SVM 或者 HMM 分类器对 6 种环境声音进行识别,可以很准确地区分如警报、电话铃声、笑声、尖叫声,但咳嗽声往往被错误地认为是语音,其原因在于 MFCCs 的低鲁棒性,导致实验对一些声音,如咳嗽声音的识别不够准确。而本文提出的二级分类方法正好弥补这一缺陷。

表 1 环境声音事件采样

Tab. 1 Environmental sound event sampling

类名	持续时间/s
语音	60
警报	40
电话铃声	125
笑声	30
尖叫声	4
咳嗽声	23
总计	342

识别结果如表2所示,系统的第一级,使用 MFCCs 和 HMM 分类,警报、电话铃声、笑声、尖叫声能达到较高的识别率<sup>[18-19]</sup>。第二级,在第一级容易混淆的声音也被区分开来,咳嗽声的识别率提高了10多个百分点,表明其具有良好的识别准确性。结果说明,HMM 和 SVM 构成的两级分类器比单独采用 SVM 或 HMM 作为分类器的识别准确率高,性能更好。

## 4 结语

本文对家庭保健监测系统的环境声音事件识别技术展开研究,提出了一种 MFCC 特征来提取参数,采用 HMM 模型和 SVM 分类器相结合的两级分类识别方法,并与 HMM 或 SVM 分别作为分类器进行比较,实验结果表明,本文提出的分类方法结构简单、具有较好的鲁棒性且识别准确率高,在环境声音识别系统中得到令人满意的结果。模型的训练是影响该系统识别性能最关键的一步,当然,环境声音识别是一个涉及面十分广泛的领域,该技术还有很多需要改善的地方,数据库建立和特征提取技术两个方面开展研究也将是后续进一步要做的工作。

### 参考文献:

- [1] 刘波霞,陈建峰. 基于特征分析的环境声音事件识别算法[J]. 计算机工程,2011,37(22): 261-263.  
Liu B X, Chen J F. An environmental sound event recognition algorithm based on feature analysis[J]. Computer Engineering, 2011, 37 (22): 261-263.
- [2] Barry S J, Dane A D, Morice A H, et al. The automatic recognition and counting of cough[J]. Cough (London, England), 2006, 2(9): 8-15.
- [3] Chen J, Kam A H, Zhang J, et al. Bathroom activity monitoring based on sound[C]//Proceedings of the third international conference on pervasive computing. Munich, Germany: Springer-Verlag Berlin Heidelberg, 2005: 47-61.
- [4] Diment A, Heittola T, Yirtanen T. Sound event detection for office live and office synthetic aasp challenge[J]. IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events, Technical Report, Tampere University of Technology, 2013(3): 23-30.
- [5] Giannoulis D, Benetos E, Stowell D, et al. IEEE AASP challenge on detection and classification of acoustic scenes and events-development dataset for event detection task, sub-task 1-0 L[J]. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 2012, 10(4): 1-4.
- [6] Roma G, Nogueira W, Herrera P. Recurrence quantification analysis features for environmental sound recognition[J]. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 2013: 1-4.
- [7] 张兢,杨超,曾建梅,等. 基于遗传算法与支持向量机的 EMD 改进算法[J]. 重庆理工大学学报:自然科学版, 2015(11): 101-105.  
Zhang J, Yang C, Zeng J M, et al. Improved EMD method based on genetic algorithm and support vector machine[J]. Journal of Chongqing University of Technology: Natural Science, 2015(11): 101-105.
- [8] 王华秋,王斌. 优化的邻近支持向量机在图像检索中的应用[J]. 重庆理工大学学报:自然科学版, 2014(9): 66-71.  
Wang H Q, Wang B. Application of optimized proximal support vector machine in image retrieval[J]. Journal of Chongqing University of Technology: Natural Science, 2014(9): 66-71.
- [9] 周哲,杨志霞. 基于凹凸过程的双支持向量机[J]. 重庆理工大学学报:自然科学版, 2014(10): 90-95.  
Zhou Z, Yang Z X. The concave-convex procedure of the twin support vector machine[J]. Journal of Chongqing University of Technology: Natural Science, 2014(10): 90-95.
- [10] Kucukbay S E, Sert M. Audio-based event detection in office live environments using optimized MFCC-SVM approach[C]//IEEE transactions on semantic computing. Anaheim, CA: IEEE, 2015(2): 475-480.
- [11] Yang Y, Nie F, Xu D, et al. A multimedia retrieval framework based on semi-supervised ranking and relevance feedback[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(4): 723-742.
- [12] Zha Z J, Wang M, Zheng Y T, et al. Interactive video indexing with statistical active learning[J]. IEEE Transactions on Multimedia, 2012, 14(1): 17-27.
- [13] Okuyucu C, Sert M, Yazlcl A. Audio feature and classifier analysis for efficient recognition of environmental sounds

- [J]. IEEE International Symposium on Multimedia (ISM), 2013, 12: 125-132.
- [14] 常飞, 乔欣, 张申, 等. 基于 MFCC 特征提取的故障预测与评价方法[J]. 计算机应用研究, 2015, 32(6): 1716-1719.  
Chang F, Qiao X, Zhang S, et al. Method of failure prediction and evaluation based on MFCC feature extraction[J]. Application Research of Computers, 2015, 32(6): 1716-1719.
- [15] 李明, 李应. 基于遗传算法优化匹配追踪的自然环境声音分类[J]. 福州大学学报: 自然科学版, 2012, 40(6): 719-725.  
Li M, Li Y. Natural environmental sound events classification based on matching pursuit optimized by genetic algorithm[J]. Journal of Fuzhou University: Natural Science, 2012, 40(6): 719-725.
- [16] Chu S, Narayanan S, Kuo C C J. Environmental sound recognition with time-frequency audio features[J]. Proc Audio Speech & Language Processing, 2009, 17(6): 1142-1158.
- [17] 胡明辉. 基于支持向量机和 HMM 的音频信号分类算法研究[D]. 长春: 长春工业大学, 2015.  
Hu M H. Automatic audio stream classification based on hidden Markov model and support vector machine[D]. Changchun: Changchun University of Technology, 2015.
- [18] 左玲云, 张晴晴, 黎塔, 等. 电话交谈语音识别中基于 LSTM-DNN 语音模型的重评估方法研究[J]. 重庆邮电大学学报: 自然科学版, 2016, 28(2): 180-186.  
Zuo L Y, Zhang Q Q, Li T, et al. Reevaluation based on LSTM-DNN Conguage model in telephone conversation speech recognition[J]. Journal of Chongqing University of Posts and Telecommunications: Natural Science Edition, 2016, 28(2): 180-186.
- [19] 罗元, 吴承军, 张毅, 等. Mel 频率下基于 LPC 的语音信号深度特征提取算法[J]. 重庆邮电大学学报: 自然科学版, 2016, 28(2): 174-179.  
Luo Y, Wu C J, Zhang Y, et al. A further speech signal features txtraction algorithm based on LPC Mel frequency scale[J]. Journal of Chongqing University of Posts and Telecommunications: Natural Science Edition, 2016, 28(2): 174-179.

## Environmental Sound Recognition for a Family Health Monitoring System

LI Lingli

(Department of Information Management, Guangdong Justice Police Vocational College, Guangzhou 510520, China)

**Abstract:** The recognition methods applied for music and speech data are not appropriate for the unstructured environmental sounds. This paper presents a recognition method which uses HMM (Hidden Markov Model) and SVM (Support Vector Machine) as two-level classifier, it realizes classification of six environmental sounds, namely speech, alert, phone ringing, laughter, screaming and cough in a family health monitoring system. Firstly, Mel-Frequency Cepstral Coefficients (MFCCs) was used to analyze natural sounds for their feature extraction. In addition, In the first level of environmental sound recognition, the MFCC are used as the inputs of HMM for classification, determine the most possible category which each environmental sound belongs to. Then, SVM classifier is used to make further judgment in the second step. The experiment results show that the technology can be used to recognition for quickly and accurately. Experimental results show that the two-layer recognition technology can improve recognition rate of the easy to be confused environmental sounds in a family health monitoring system.

**Key words:** environmental sound; HMM; SVM; MFCC

(责任编辑 游中胜)