

北京大学学报(自然科学版)
Acta Scientiarum Naturalium Universitatis Pekinensis
doi: 10.13209/j.0479-8023.2015.034

基于 MFCCG-PCA 的语音情感识别

陈炜亮 孙晓[†]

合肥工业大学情感计算与人工智能实验室, 合肥 230001; [†]通信作者, E-mail: sunx@hfut.edu.cn

摘要 针对语音情感值具有维度大、难处理的问题, 通过结合 MFCC 改进算法和 PCA 模型, 进行二次优化, 提出一种新的语音情感值提取模型 MFCCG-PCA。多组实验表明, 相比一般的 MFCC 模型, MFCCG-PCA 模型在语音情感识别方面的性能有较大提高。

关键词 语音情感; 梅尔频率倒谱系数; 主成分分析; MFCCG-PCA; 支持向量机; CASIA 汉语情感语料库
中图分类号 TP391

Mandarin Speech Emotion Recognition Based On MFCCG-PCA

CHEN Weiliang, SUN Xiao[†]

Hefei University of Technology, Laboratory of Affective Computing and Advanced Intelligent Machine, Hefei 230001;
[†] Corresponding author, E-mail: sunx@hfut.edu.cn

Abstract To solve the problem that the dimension of the speech emotion characteristic value is big and it is difficult to train, a new speech emotion recognition model, MFCCG-PCA model, is put forward by the combination of the MFCC model and the PCA model. Multiple sets of experiments show that the MFCCG-PCA model has larger performance improvement than general MFCC model in the aspect of speech emotion recognition.

Key words speech emotion recognition; MFCC; PCA; MFCCG-PCA; SVM; CASIA

随着人工智能的新发明和新技术的应用, 人机交互体验的友好度不断提高。如何让计算机、机器人能够和人类进行无障碍交流, 接受、处理、识别、理解人机交互中人类所传达出来的信息, 自适应地给出合适的回答, 成为人工智能发展的更高层次目标。在人类日常交互的过程中, 语音是最直接、最方便, 也是最常用的交流手段。人类语音中除了包含语音信息外, 还常常蕴含着丰富的情感特征。在人机交互领域中, 语音情感识别是情感计算的一个关键的研究方向。计算机通过分析人类语音中含有的情感特征, 理解人类语音中包含的情感信息, 识别并模拟说话人的情感状态, 从而实现人机交互无障碍和友好的目标。

情感计算是近些年才开始的新兴领域, 最早的研究开始于 20 世纪 80 年代。1985 年, 人工智能的创始人, 美国麻省理工学院的 Minsky^[1]提出“计

算机的情感能力”的概念。之后, “情感计算”这个概念在 1997 年被美国麻省理工学院媒体实验室的 Picard^[4]第一次提出。从此情感计算成为计算机科学的一个热门研究领域。日本产业技术研究所(AIST)^[3]发布了一个具有丰富人类表情的女性机器人“HRP-4C”, 能够对接收的语音信号进行反映做出喜怒哀乐等表情。

对语音情感进行研究需要选取合适的语音情感特征模型、基于韵律学的特征以及基于谱的相关性的特征。基于韵律学的特征一定程度上能够反映被提取音频文件的语音特征, 但是与平静的语音相比, 带有强烈感情色彩的语音音频在语音语速、频率和幅值等参数上存在明显波动, 这些特征较难提取, 并且不能全面的反映人类语音中蕴含的情感^[6]。相比之下, 基于谱的相关性的语音特征被认为是声道形状变化与发声运动之间相关性的体现, 目前已

在包括语音识别、说话人识别等在内的语音信号处理领域有成功的运用。

MFCC 是一种基于谱的语音情感特征值,它有别于文本情感特征,具有数量大,维度多,特征值之间关系复杂等特点,单纯使用没有优化处理过的语音情感特征值进行模型训练,不能得到有价值的模型。因此需要对得到的特征值进行进一步优化处理,使特征值中代表语音情感的特征性进一步突出,除去噪音特征。

针对语音情感值维度多、数据量大、难以分析的问题,本文提出对语音情感特征 MFCC 模型的提取算法进行改进进行第一步优化,通过将改进后提取的 MFCC 特征利用 PCA 模型进行第二次优化,用两次优化的方法来获得更好的语音情感值模型,实验表明,通过两次优化得的特征值模型比起一般的特征值在语音情感识别的效果上有较大的提高。

1 相关工作

目前普遍使用的 Mel 频率倒谱系数(Mel Frequency Cepstrum Coefficient, MFCC)是基于人耳听觉系统所提出的参数,它考虑了人类发出声音与接受声音的过程和特点,其频率的增长与人耳的听觉特性一致,并且与语音的实际频率息息相关^[5]。Mel 频率与音频的实际频率具有非线性对应关系,当实际频率在 1 kHz 以下时,是线性分布;当实际频率在 1 kHz 以上时, Mel 频率成对数增长。因此 Mel 具有很好的辨识性能和鲁棒性。

对语音特征值的优化方法有很多,李桂春等^[6]提出在语音特征值的不同维度中添加模糊权值以提高语音特征的属性,庞陟儒等^[7]使用 PCA 模型对语音特征值进行降维,并且对不同说话人的识别结果进行分析,发现降维后大大提高了分类器的处理效率,同时也能保持识别的正确率维持在同一水准,叶庆云等^[8]通过改变 MFCC 提取算法中浮点数以及实数与虚数,来获得更好的 MFCC 参数。庞陟儒等单纯只使用了 PCA 的模型对提取的 MFCC 特征值进行处理,表明 PCA 模型在对 MFCC 数据的处理上是可行的,但同时他只针对单纯说话人进行识别,没有用于语音情感的识别。叶庆云等的方法使 MFCC 特征值识别速度提高近 20 倍,但分辨率下降了 9%。本文针对现有语音情感识别方法存在的问题,通过使用改良的 MFCC 提取算法,修改提取算法中的参数以及 DCT 参数,在提

取 MFCC 模型时对 MFCC 特征值进行第一步优化,之后结合数据降维方法对语音特征值进行第二次优化,进一步减少语音情感特征值的维度同时提高分类器的处理效率,以提高分类器生成模型的质量。

2 MFCCG-PCA 模型

通过对 MFCC 模型进行改进的提取(即第一步优化)后,再通过 PCA 模型进行特征提取(即第二步优化),经过两步优化得到的模型在实验中发现,能够大幅度提升语音情感识别性能,本文称之为 MFCCG-PCA 模型。

2.1 MFCC 的提取

MFCC 模型与实际频率具有以下关系:

$$\text{Mel}(f) = 2595 \times \lg \left(1 + \frac{f}{700} \right). \quad (1)$$

通过 MFCC 模型与实际频率之间的特点,可以用以下的方法提取 MFCC 特征值^{[9][10]}:对输入语音音频进行预处理,主要包括对语音信号的预加重、分帧、加窗等操作;将预处理完的信号每一帧通过快速傅里叶变换(FFT)求倒谱。将完成上一步的信号通过 Mel 滤波器组,即对幅度谱加 Mel 滤波器组;对 Mel 滤波器组输出的所有结果取对数,通过离散余弦变换(DCT)后,得到原始的静态 MFCC 参数;对静态 MFCC 参数进行一阶差分,对得到的一阶差分系数进行二阶差分,得到动态的一阶 MFCC 参数和二阶 MFCC 参数。

2.2 MFCCG-PCA 模型

2.2.1 MFCC 的改良提取

Mel 滤波器组实质上是满足式(2)的一组三角滤波器。

$$\begin{aligned} \text{Mel}[f(m)] - \text{Mel}[f(m-1)] = \\ \text{Mel}[f(m+1)] - \text{Mel}[f(m-1)], \end{aligned} \quad (2)$$

其中 $f(m)=1, 2, 3, \dots, M$ 为三角滤波器的中心频率,各 $f(m)$ 之间的间隔随 m 值减小而缩小,随 m 值的增大而增宽。定义一个具有 24 个滤波器的滤波器组,因为每一个三角滤波器的覆盖范围都近似于人耳的一个临界带宽,因此可以用来模拟人耳的掩蔽效应^[12]。24 个滤波器形成满足公式(3)的滤波器组,通过这个 Mel 滤波器组可以得到经过滤波后的 Mel 信号。

$$H_m(k) = \begin{cases} 0, & k < f(m-1), \\ \frac{2[k - f(m-1)]}{[f(m+1) - f(m-1)][f(m) - f(m-1)]}, & f(m-1) \leq k \leq f(m), \\ \frac{2[f(m+1) - k]}{[f(m+1) - f(m-1)][f(m+1) - f(m)]}, & f(m) \leq k \leq f(m+1), \\ 0, & f(m+1) \leq k. \end{cases} \quad (3)$$

在使用 Mel 滤波器前需要对信号进行 DFT 变换, 同时在滤波后还需用 DCT 对得到的滤波结果进行解卷。

$$\text{dctcoef}(k) = \cos(\cos((2 \times n + 1) \times k \times \pi / (2 \times 24))), \quad (4)$$

式(4)中 k 代表滤波器将得到的 DFT 转换为 DCT 参数, 一般选择 $k=12$, 得到 24 维的 MFCC 特征值, 若将 k 减小一半, 即把原本 24 维 MFCC 降维为 12 维, 实验表明经过 DCT 得到的 MFCC 特征值的维数与 DCT 的参数有关; 如果选取的 k 值为 12。通过改变 DCT 的参数, 改良 MFCC 提取算法, 可以减少通过一段语音得到的 MFCC 参数的维数, 并且能够保持采样不变, 输入的语音信号不会失真, 保持原本语音语料的情感特征值。将算法中 24 个滤波器提取 MFCC 模型改为 12 个滤波器提取, 得到 12 维度的 MFCC 模型, 修改 DCT 参数后再通过 DCT 降维可以得到 6 维的 MFCC 模型, 由于 MFCC 特征在加性噪声的情况下并不稳定, 因此在语音识别系统中通常要对其进行归一化处理以降低噪声的影响^[9]。

$$w(A) = 1 + 6 \times \sin\left(\pi \times \frac{A}{6}\right). \quad (4)$$

通过式(5)的归一化倒谱提升窗口进行归一化处理, A 为输入的矩阵。修改 DCT 参数 k 和 n 的值以及归一化倒谱提升窗口, 使其满足 12 维 MFCC 的提取要求, 最终得到比一般的 MFCC 特征值维数更少的特征值。

实验发现通过这种方法能够使 MFCC 的维度下降一半左右, 并且不减少 MFCC 参数中所含的特性信息。本文将这种方法称为第一次优化, 得到的 MFCC 特征值称为改良的 MFCCG 特征值, 简称 MFCCG。

2.3 数据降维

通过改良 MFCC 提取算法获得优化的 MFCC 参数后, 使用数据降维的方法对提取完成后获得的特征值进行第二次优化。数据降维的基本原理是将高维空间的数据集中的样本点通过空间中的线性或

非线性变换, 映射到一个低维空间的点上, 这种数据降维的方法又叫做特征提取, 常常用在计算机视觉以及图像处理上。本文使用数据降维的方法对 MFCCG 进行优化, 即对 MFCC 模型进行第二次优化, 得到优化后的模型比 MFCCG 模型的维度有进一步的减小。

主成分分析(Principal Component Analysis, PCA)是一种线性的无监督特征提取方法, 本质上是一种多元的统计学分析方法^[13]。PCA 将数据集中的多个多元变量通过线性变换选出较少的几个重要变量, 以达到特征提取的目的。语音的特征值维数很大, 其中可能有很多对语音包含的情感值没有贡献或者贡献较小的特征值, 同时某一个特征值对语音的情感信息贡献很大, 但是这个特征值有可能是其余多个特征值进行线性变化或者组合而成, 这些噪声信息影响提取到 MFCC 特征值。

PCA 的主要过程如下: 先将每个矩阵写成列矩阵的形式, 计算平均特征值 $\varphi = \frac{1}{N} \sum_{i=1}^{i=N} x_i$, 然后计算每个特征值与平均值的差值 d_i , 构建协方差矩阵 $C = \frac{1}{N} \sum_{i=1}^N d_i d_i^T = \frac{1}{N} A A^T$, $A = (d_1, d_2, \dots, d_N)$, 计算协方差矩阵 C 的特征值和特征向量, 构建特征向量空间, 在空间中对特征值和特征向量进行排序选择, 得到降维后的向量。

由于 MFCC 模型模拟人耳听觉系统的特性, 所以 MFCC 特征通常低频部分的语音信息提取的精度较高, 而对高频部分的处理精度较低。研究表明, MFCC 模型往往在语音高频部分所含的情感的有价值的信息较少而在低频部分含有较多的有用信息。通过主成分分析方法, 能够对语音 MFCC 模型复杂的数据结构进行分析, 保留在语音低频部分的有价值特征值, 筛选高频部分的特征信息, 并将高频部分与低频部分结合进行数据提取, 得到维度降低、数量级大大减小但是包含情感特征信息不减少的 MFCC 特征向量。

3 实验与结果分析

3.1 实验准备

实验语料采用 CASIA 汉语情感语料库, 共分为 6 类情感: 愤怒(angry)、害怕(fear)、高兴(happy)、中性(neutral)、悲伤(sad)和惊喜(surprise)。采用说话人 liuchanhg、zhaozuoxiang 和 zhaoquanyin 所说的 6 种情感语音语料每种 300 条。其中选取相同的前 200 条(编号 201~400)作为训练集, 每人共 1200 条语音, 选取后 100 条语音语料(编号 401~500)每人共 600 条语音语料作为测试集。另外, 为了试验相同说话人不同说话内容对模型的影响, 还选用说话人 liuchanhg 的 6 种情后 200 条语音(编号 301~500)作为训练集, 前 100 条语音(编号 201~300)作为测试集。

实验采用 SVM 分类器, 使用台湾大学林智仁副开发 Libsvm 作为工具, Libsvm 能够快速并且有效实现 SVM 所需要的各种功能, 其中使用的核函数为 RBF 核函数^[14]。

使用 Libsvm 提供的 ν -折交叉验证工具进行参数选择, ν -折交叉验证和网络搜索的方法来确定 RBF 核函数中的惩罚因子 c 与参数的最佳取值。 ν -折交叉验证是先将数据集划分同样大小的 ν 个集合, 将其中的一个集合作为未知的, 称为测试集, 其他 $\nu-1$ 子集作已知的训练集, 从“未知”测试集上获得预测准确率, 可以更精确地反映分类器在独立数据集上的效果^[15]。

3.2 模型性能评价

模型可以从分类结果质量、分类的速度和算法的复杂度等方面来评价模型性能, 其中分类结果质量是模型最重要的性能。从分类结果质量可以明显看出哪一种方法和模型对语音情感的分类最好。本实验中采用交叉验证正确率和值两种数值来评价一个语音识别模型。

3.3 实验结果分析

本设计共获得 20 组实验的数据, 可以分成 7 组来研究不同语料情况。使用不同优化方法和不同模型等差别下, 分析系统对语料语音情感的识别情况。

3.3.1 提取 MFCC 模型差分参数对语音情感识别的影响

实验中使用同一组语音语料, 说话人 liuchanhg 所录制的 6 种情感各 200 条(201~400)为训练集, 使

用每种后 100 条(401~500)作为测试集, 对训练集进行静态 MFCC 模型提取、一阶 MFCC 模型提取和二阶 MFCC 模型提取, 对比 3 种模型的语音情感识别结果, 见表 1。

通过表 1 可以发现, 一阶 MFCC 模型对语音情感的识别度有一定的提高, 但提高程度有限。而二阶 MFCC 参数由于多了一倍的参数值, 导致维数增加, 在当前语音特征值维数与语料相比过大的情况下, 维数的提升反而导致了模型识别性能的下降。

3.3.2 使用改进 MFCC 模型算法对语音情感识别的影响

使用与 4.3.1 实验相同的语料, 分别进行一般静态 MFCC 和一阶 MFCC 的提取, 再使用改进的 MFCC 提取算法, 提取到改进的一阶 MFCC 模型, 称为 MFCCG。表 2 为使用这 3 种数据处理方法后, 语音情感识别的性能评价。

实验发现对 MFCC 模型的提取算法进行改进后, 准确率和值比静态 MFCC 模型和一阶 MFCC 模型都有一定的提升, 说明在特定说话人情况下, MFCCG 模型有更好的性能。

表 1 静态、一阶和二阶 MFCC 模型对比

Table 1 Comparison of static, first order and second order MFCC model

数据集	准确率/%	F1 值/%	
liuchanhg	静态 MFCC 模型	28.33	24.83
	一阶 MFCC 模型	32.00	27.50
	二阶 MFCC 模型	24.5	22.17
zhaozuoxiang	静态 MFCC	23.67	20.67
	一阶 MFCC	27.24	29.00
	二阶 MFCC	25.50	18.66

表 2 MFCCG 模型的性能

Table 2 Performance of the MFCCG model

数据集	准确率/%	值/%	
liuchanhg	静态 MFCC 模型	28.33	24.83
	一阶 MFCC 模型	32.00	27.50
	MFCCG 模型	44.25	38.50
zhaozuoxiang	静态 MFCC	23.67	20.67
	一阶 MFCC	27.24	29.00
	MFCCG 模型	40.28	37.46

表 3 数据降维模型对语音情感的影响

Table 3 Influence on speech emotion recognition by data dimensionality reduction

数据集	准确率/%	值/%
MFCCG 模型	44.25	38.50
MFCCG-PCA 模型	66.16	54.33
MFCCG-LDA 模型	41.17	31.00
MFCCG-LE 模型	52.00	35.50

3.3.3 对 MFCC 模型进行数据降维后语音情感识别的影响

对于与 4.3.1 实验相同的语料, 进行改进的 MFCC 模型 MFCCG 的提取, 对提取结果分别通过主成分分析模型(PCA, 数据集称为 MFCCG-PCA)、线性判别式分析模型(LDA, 数据集称为 MFCCG-LDA)和拉普拉斯特征映射(LE, 数据集称为 MFCCG-LE)3 种数据降维方法进行处理, 表 3 为实验结果。

在特定人说话情感识别的情况下, 使用 LE 模型和 PCA 模型都能达到很好的特征提取效果, 并且在改进的 MFCC 模型上再使用 PCA 模型进行特征提取, 可以让理论上模型能达到的最高识别率提升将近 30 个百分点, 接近三分之二, 而实际的识别率也能够超过一半, 模型的语音情感识别性能有了明显的提高。

3.3.4 PCA 模型对 MFCC 模型的具体影响

为了研究 PCA 模型对 MFCC 模型的具体影响, 选取同一组语音语料, 进行静态 MFCC 模型、一阶 MFCC 模型、二阶 MFCC 模型、改进的 MFCC 模型以及二阶改进的 MFCC 模型的提取, 并且使用 PCA 模型分别对它们进行分析, 得到的实验结果如表 4。

使用 PCA 模型对各类 MFCC 模型的分析表明, 不管对于哪一种 MFCC 模型, PCA 模型都可以让模型的性能指标有所提升, 对语音语料的情感的提取和识别都能够有所提高。

3.3.5 在同一说话人不同语料上研究 MFCCG-PCA 模型的识别性能

之前的试验表明, MFCCG-PCA 模型能够让语音语料的情感识别大幅提升。现在通过使用同一说话人不同的说话内容的语料, 来判断 MFCCG-PCA 模型能否让语音情感识别保持在一个相对高的的基准。实验采用同一说话人 liuchanhg 的六种情感前 200 条语音(编号 201~400)作为训练集 1(共 1200

条语音), 对应测试集为 6 种情感后 100 条语音(编号 401~500, 共 600 条语音); 采用 6 种情感后 200 条语音(编号 301~500)作为训练集 2(共 1200 条语音), 对应测试集为 6 种情感的前 100 条语音(编号 201~300, 共 600 条语音), 实验结果如表 5 所示。

实验发现, 在不同语料上, MFCC-PCA 模型能够取得相对平稳、差别不大的性能指标, 说明

表 4 PCA 模型对 MFCC 的影响效果

Table 4 Influence on MFCC model by PCA model

数据集	准确率/%	值/%
静态 MFCC 模型	28.33	24.83
MFCC-PCA	33.17	32.25
一阶 MFCC 模型	32.00	27.50
一阶 MFCC-PCA	36.75	32.50
二阶 MFCC 模型	24.50	22.17
二阶 MFCC-PCA	29.83	23.33
MFCCG 模型	44.25	38.50
MFCCG-PCA	66.16	54.33
二阶 MFCCG 模型	42.17	37.33
二阶 MFCCG-PCA	56.33	45.67

表 5 同一说话人不同语料上 MFCCG-PCA 模型

Table 5 MFCCG-PCA model used in the same speaker but different corpora datasets

数据集	准确率/%	值/%
训练集 1	66.16	54.33
训练集 2	57.50	54.00

MFCC-PCA 模型在特定人不同语音语料的情感信息识别上能够保持一个较高且稳定的识别性能。

3.3.6 不同说话人独立识别上 MFCCG-PCA 模型的识别性能

针对 MFCCG-PCA 模型对不同的人语音情感识别性能的分析, 实验使用 3 个不同说话人 liuchanhg, zhaozuoxiang 和 zhaoquanyin 的语音语料 6 种情感每种前 200 条(编号 201~400, 每人共 1200 条)作为训练集, 用后 100 条作为测试集(编号 401~500, 每人共 600)分析 MFCCG-PCA 在不同说话人情况下的语音情感识别性能。

表 6 可以清楚反映, 在针对非特定人识别的情况下, 使用 MFCCG-PCA 也能够大幅度提升语音情感识别的正确率, 提升模型的情感识别性能。说明 MFCCG-PCA 在语音情感识别方面能够有很好的应用。

3.3.7 不同说话人混合的识别性能

采用与上一实验同样语料, 将两个说话人 liuchanhg 和 zhaozuoxiang 的每种前 200 条语音结合作为训练集 1(共 2300 条语音), 对应的每种后 100 条语音作为测试集 1(共 1200 条语音), 将 3 个说话人 liuchanhg, zhaozuoxiang 和 zhaoquanyin 的语音语料每种前 200 条语音结合作为训练集 2(共 3600 条语音), 对应的每种后 100 条语音作为测试集 2, 分析在混合环境下 MFCCG-PCA 模型的识别性能, 结果见表 7。

表 7 的实验结果表明, 即使在混合环境下, MFCCG-PCA 模型对语音情感的识别也能达到一定程度, 比一般的 MFCC 模型有较大的提高, 并且能够获得较好的识别性能。

表 6 不同说话人 MFCCG-PCA 模型的识别性能

Table 6 MFCCG-PCA model used in the different speaker datasets

数据集		准确率/%	值/%
liuchanhg	liuchanhg 静态 MFCC	28.33	24.83
	liuchanhg MFCCG-PCA 模型	66.16	54.33
zhaozuoxiang	zhaozuoxiang 静态 MFCC	23.67	20.67
	zhaozuoxiang MFCCG-PCA 模型	65.33	58.00
zhaoquanyin	zhaoquanyin 静态 MFCC	29.08	23.50
	zhaoquanyin MFCCG-PCA 模型	62.17	53.50

表 7 不同说话人混合 MFCCG-PCA 模型的性能

Table 7 MFCCG-PCA model used in the Multi-speaker recognition

数据集		准确率/%	值/%
训练集 1	静态 MFCC 模型	31.75	30.42
	MFCCG-PCA 模型	54.00	46.67
训练集 2	静态 MFCC 模型	22.50	20.17
	MFCCG-PCA 模型	43.22	40.56

4 结论

语音情感识别的关键在于特征值的提取和模型的选择, 本文使用对 MFCC 模型的改进算法和 PCA 模型相结合的二次优化方法, 对语音情感特征值进行优化处理, 很大程度上解决了语音情感值维度大、识别难度高的问题。比一般的 MFCC 模型, 提取的 MFCCG-PCA 模型在语音情感识别上有较大的进步, 能够使语音情感识别的成功率显著提高, 并且能够强化识别模型的性能。

参考文献

- [1] Minsk M L. The society of mind. NewYork: Touchstone, 1985: 85—86
- [2] Picard R W. Affective computing. London England: MITPress, 1997: 192—195
- [3] 陈建厦. 语音情感识别综述 // 第一届中国情感计算与智能交互学术会议论文集. 北京: 中国自动化协会, 2003: 179—184
- [4] 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. 电路与系统学报, 2007, 12(4): 76—83
- [5] Ganchev T, Fakotakis N, Kokkinakis G. Comparative evaluation of various MFCC implementations on the speaker verification task // 10th International Conference on Speech and Computer: Proceedings of the SPECOM-2005. Patras, 2005: 191—194
- [6] 李桂春, 郑能恒, 李泰. 基于模糊隶属值加权的 MFCC 特征提取算法 // 第七届和谐人机环境联合学术会议(HHME2011)论文集. 北京, 2011: 40—46
- [7] 庞陟儒, 张雪英, 刘晓峰. 基于 PCA-MFCC 支持向量机快速语音识别. 科学技术与工程, 2013, 13(4): 1034—1036
- [8] 叶庆云, 蒋佳. 基于语音 MFCC 特征的改进算法. 武汉理工大学学报, 2007(5): 150—153
- [9] Tyagi V, Wellekens C. On desensitizing the melcepstrum to spurious spectral components for robust speech recognition // ICASSP '05. Vancouver, 2005: 529—532
- [10] Xu Min, Duan Ling-Yu, Cai Jianfei, et al. HMM-based audio keyword generation. Lecture Notes in Computer Science, 2004, 3333: 566—574
- [11] Sahidullah M, Saha G. Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. Speech Communication, 2012, 4(4): 543—565
- [12] 杜志平. 基于内容的音频指纹提取技术研究[D]. 西安: 西安电子科技大学, 2011
- [13] Pearson K. On lines and planes of closest fit to systems of points in space. Philosophical Magazine, 1901, 2: 559—572
- [14] Chang Chih-Chung, Lin Chih-Jen. LIBSVM: a library for support vector machines [EB/OL]. (2001)[2011—05—20]. <http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf>: Chih-Chung Chang
- [15] Hsu Chih-Wei, Chang Chih-Chung, Lin Chih-Jen. A practical guide to support vector classification [EB/OL]. (2003)[2011—04—15]. <http://www.csie.ntu.edu.tw/~cjlin>: Chih-Chung hang