

基于 FVQMM 的说话人识别方法

杨彦¹ 赵力²

(1. 江苏盐城工业职业技术学院汽车工程学院, 盐城, 224005; 2. 东南大学信息科学与工程学院, 南京, 210096)

摘要: 为了进一步提高基于高斯混合模型的与文本无关说话人识别系统的识别性能, 本文针对高斯混合模型在建模时需要较多的训练数据的缺陷, 提出了一种新的应用于小样本说话人识别系统的与文本无关说话人识别方法, 该方法综合考虑了模糊集理论、矢量量化和高斯混合模型的优点, 通过用模糊矢量量化误差尺度取代传统高斯混合模型的输出概率函数, 减少了建模时对训练数据量的要求, 提高了模型精度和识别速度。同时由于模糊集理论起到了“数据整形”的作用, 所以增强了目标说话人数据的相似性。实验结果表明该方法针对小样本数据的说话人识别系统, 识别性能优于传统的基于高斯混合模型的说话人识别系统。

关键词: 说话人识别; 模糊集理论; 矢量量化; 高斯混合模型

中图分类号: TN912.34 **文献标志码:** A

Speaker Recognition Based on FVQMM

Yang Yan¹, Zhao Li²

(1. School of Automobile Engineering, Yancheng Institute of Industry Technology, Yancheng, 224005, China;

2. School of Information Science and Engineering, Southeast University, Nanjing, 210096, China)

Abstract: In order to further improve the performance of speaker recognition system based on the GMM independent of text, a new speaker recognition method is applied to the speaker recognition system with small samples and text independent. Aiming at the large quantity demanded of training data during the modeling of the GMM, the advantages of the fuzzy-set theory, vector quantization and the GMM are considered. Then through replacing the output probability function in the traditional GMM with the error scale of the fuzzy VQ, the requirements of the training data amount are reduced while improving the accuracy and recognition speed of the model. Meanwhile as a result of the fuzzy-set theory playing a role of "plastic date", the similarity in the data of the target speakers is enhanced. Experimental results exhibit that the speaker recognition system of the method for the small sample data, achieves a superior recognition performance than the traditional speaker recognition system based on the GMM.

Key words: speaker recognition; fuzzy-set theory; vector quantization; Gaussian mixed model

引 言

自动说话人识别(Automatic speaker recognition, ASR)很久以来就是一个既有吸引力又有相当困难的课题。说话人识别技术按其最终完成的任务可以分为两类:自动说话人确认(Automatic speaker verification, ASV)和自动说话人辨认(Automatic speaker identification, ASI)。本质上它们都是根据说话人所说的测试语句或关键词,从中提取与说话人本人特征有关的信息,再与存储的参考模型比较后做出正确的判断^[1]。另外,说话人识别按其被输入的测试语音来分可以分为与文本有关的说话人识别和与文本无关的说话人识别,而与文本无关的说话人识别在今天无疑有着更广泛的应用前景。

对于与文本无关的说话人识别,由于说话人的个性特征具有长时变动性,而且其发音常常与环境、说话人情绪和说话人健康有密切关系,并且实际过程中还可能引入背景噪声等干扰,这些都是与文本无关说话人识别的识别率得不到进一步提高的主要因素。通常说话人识别的经典方法是基于高斯混合模型(Gaussian mixed model, GMM)的识别方法,由于它作为统计模型能够吸收由不同说话人引起的说话人个性特征的变化,可以得到较好的识别性能。但是由于 GMM 作为统计模型对模型训练数据量有一定的依赖性,所以对于小样本的与文本无关说话人识别系统,要使 GMM 完全吸收由不同说话人引起的语音特征的变化去除话者差别非常困难,所以在实际应用中通常采用话者适应的方法使未知说话人的语音去适应已知标准说话人的语音模型。因此,近年来在说话人识别方法方面,基于高斯混合背景模型(Gaussian mixed model-universal background model, GMM-UBM)方法已成为主流的识别方法^[2-3]。基于 GMM 超向量的支持向量机和因子分析方法^[4-5]则代表 GMM-UBM 方法的新成果。尽管 GMM-UBM 方法作为目前主流的方法已得到共识,但同时一些局部的改进方法和针对不同应用的改进方法也非常有意义^[6-10]。几乎所有成功的语音信号处理方法都是基于统计的、概率的或信息理论的方法。其中较具代表性的方法有矢量量化法(Vector quantization, VQ)和隐马尔可夫模型法(Hidden Markov modeling, HMM),而 GMM 是 HMM 中的一种。VQ 方法是由 Shore 和 Burton 首先提出,并应用于特定人数码识别^[11],其主要优点是无需时间规正或进行动态时间伸缩,同时对于模型训练数据依赖较小。但是,该方法对于由话者差别引起的语音特征变化的模型优化能力较弱。模糊集理论通过隶属度函数引入不确定性思想^[12],实现对硬聚类算法的有效扩展,可以较好地吸收特征参数的变化,起到“数据整形”的作用,对于由话者差别引起的语音特征的变化,模糊集理论可以较好地起到模型优化的作用。模糊矢量量化采用模糊 C 均值聚类算法来实现矢量量化,在同样码本尺寸的情况下,通过模糊 C 均值聚类分析可以减少码本的量化误差,在实际应用中取得过较好的效果。

本文对于 GMM 在建模时需要较多的训练数据的缺陷,针对小样本的说话人识别系统,提出了一种基于模糊矢量量化混合模型(Fuzzy VQ mixed model, FVQMM)与文本无关的说话人识别方法,它综合考虑了 VQ, GMM 方法以及模糊集理论的优点。通过用模糊矢量量化误差尺度取代传统高斯混合模型的输出概率函数,减少了建模时对训练数据量的要求,提高了模型精度和识别速度。同时由于模糊集理论起到了“数据整形”的作用,所以增强了目标说话人数据的相似性。实验结果表明该方法针对小样本数据的说话人识别系统,识别性能优于传统的基于高斯混合模型的说话人识别系统。

1 基于 GMM 的说话人识别方法

为了说明基于 FVQMM 的说话人识别方法,首先必须介绍一下传统的基于 GMM 的说话人识别方法。GMM 是 M 个成员的高斯概率密度的加权和,可以用下式表示

$$p(\mathbf{x} | \lambda) = \sum_{i=1}^M a_i b_i(\mathbf{x}), \quad \sum_{i=1}^M a_i = 1 \quad (1)$$

式中: \mathbf{x} 为 D 维随机向量; $b_i(\mathbf{x})$ ($i=1, 2, \dots, M$) 为每个成员的高斯概率密度函数; a_i ($i=1, 2, \dots, M$) 是

混合权值。完整的 GMM 可表示为 $\lambda_i = \{a_i, \mu_i, \Sigma_i\} (i=1, 2, \dots, M)$ 。每个成员密度函数是一个 D 维变量的高斯分布函数,形式如下

$$b_i(\mathbf{x}) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mu_i)^t - \Sigma_i^{-1}(\mathbf{x} - \mu_i)\right\} \quad (2)$$

对于一个长度为 T 的测试语音时间序列 $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T)$,其 GMM 似然概率可以写作

$$P(\mathbf{X} | \lambda_i) = \prod_{t=1}^T p(\mathbf{x}_t | \lambda_i) \quad (3)$$

或用对数域表示为

$$L(\mathbf{X} | \lambda_i) = \log P(\mathbf{X} | \lambda_i) = \sum_{t=1}^T \log p(\mathbf{x}_t | \lambda_i) \quad (4)$$

识别时运用贝叶斯定理,在 N 个未知话者的模型中,得到的似然概率最大的模型对应的话者即为识别结果

$$i^* = \operatorname{argmax}_{1 \leq i \leq N} L(\mathbf{X} | \lambda_i) \quad (5)$$

2 基于 FVQMM 的说话人识别方法

从式(1)可知,在基于 GMM 的说话人识别方法中,每一帧语音的得分值是通过每个成员的输出概率函数计算出的输出概率值。正确的输出概率函数的估计需要一定数量的训练数据训练得到。在说话人识别系统中,用于各说话人 GMM 训练的语料较少,给 GMM 的学习带来一定困难。而本文提出的基于 FVQMM 的说话人识别方法中,每一帧语音的得分值是每个成员通过用 FVQ 误差尺度取代传统 GMM 的输出概率函数,利用矢量量化误差值取代传统 GMM 的输出概率值得到的。即它的模型参数由混合权值和每个成员的码本组成,对于某个类别的模型,每一帧语音的得分值就是该帧语音与每个成员码本的量化误差值的加权和。识别系统对每个类别都预备一个相应的模型,利用各成员的码本逐帧计算输入序列的量化误差值,并计算所有输入帧的累积误差值。在 N 个未知话者的模型中,得到的累积量化误差值最小的模型对应的话者即为识别结果。

最小累积误差的计算可以通过维特比算法实现。其计算公式为

$$D = \sum_{t=1}^T \sum_{i=1}^M a_i d(\mathbf{x}_t, C_i) \quad (6)$$

式中: $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T$ 表示输入时间序列; $a_i (i=1, 2, \dots, M)$ 是混合权值; C_i 表示与成员 i 相对应的码本; $d(\mathbf{x}_t, C_i)$ 表示 \mathbf{x}_t 和 C_i 间的距离,该距离定义如下

$$d(\mathbf{x}_t, C_i) = \min_j d(\mathbf{x}_t, c_j), \quad c_j \in C_i \quad (7)$$

可以通过维特比算法求取最小累积误差距离

$$\begin{cases} g(t) = \min_i \{g(t-1) + d(\mathbf{x}_t, C_i)\} \\ D = g(T) \end{cases} \quad (8)$$

式中: $g(t)$ 表示输入时间部分序列 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t$ 与模型间的最小累积距离; $t=1, 2, \dots, T$; $i=1, 2, \dots, M$ 。对于一个给定的输入时间序列,识别系统将利用各类别模型逐帧计算该序列的量化误差值,得到最小累积量化误差值的模型所对应的类别即为识别结果。式(7)采用最近邻准则计算误差,也可以采用其他误差准则,如 K 最近邻准则,即

$$d(\mathbf{x}_t, C_i) = \sum_{k=1}^K d_k \quad (9)$$

式中 d_k 表示 \mathbf{x}_t 和 C_i 中所有码字的第 k 个最小距离。也可以采用概率形式

$$-\log P(\mathbf{x}_t | i) = -\log\left(\sum_{k=1}^K e^{-d_k}\right) \quad (10)$$

3 FVQMM 模型参数的估计

在训练 FVQMM 之前,必须首先确定混合模型的成员数 M ,然后利用有效方法把训练数据集对成员数 M 进行分段,再利用归属成员 M 的训练数据集来训练第 M 个成员的 FVQ 码本。本文采用时间规整神经网络来分割训练数据集。整个 FVQMM 模型参数估计过程可以描述如下。

(1) 对于训练数据集 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$,首先利用时间规整神经网络把训练数据集对成员数 M 进行分段。网络的输入层有 N 个节点,对应训练数据集的 N 个输入语音帧矢量,每个节点均有一个与之联系的输入矢量 $\mathbf{x}_k^0 (k=1, 2, \dots, N)$, \mathbf{x}_k^0 是第 k 帧语音信号的特征矢量。进入第一层后,将距离最近的两个顺序矢量以加权平均合并,其余矢量不变,这样第一层就具有 $N-1$ 个节点以及与之联系的 $N-1$ 个矢量 $\mathbf{x}_k^1 (k=1, 2, \dots, N-1)$ 。依次类推,在经过 $N-M$ 步合并后,最终网络的输出层具有 M 个节点以及与之联系的 M 个矢量 $\mathbf{x}_k^{N-M} (k=1, 2, 3, \dots, M)$ 。时间规整网络对特征矢量序列的聚类合并过程,从整体上看又是对输入语音信号的一个分段过程,即从一个输出节点层的 \mathbf{x}_k^{N-M} 回溯到输入节点层就可以找到属于该输出节点的输入语音信号的分段数据集。

具体的时间规整算法描述如下:设 $\mathbf{x}_1^0, \mathbf{x}_2^0, \dots, \mathbf{x}_N^0$ 是输入语音信号的特征矢量。以 f_k^i 表示矢量 \mathbf{x}_k^i 所代表的语音帧数,其中 $i=0, 1, \dots, N-M, k=1, 2, \dots, N-i$ 。特别地,当 $i=0$ 时, $f_k^0=1 (k=1, 2, \dots, N)$ 。并以 d_k^i 表示矢量 \mathbf{x}_k^i 与 \mathbf{x}_{k+1}^i 之间的距离,即 $d_k^i = \|\mathbf{x}_k^i - \mathbf{x}_{k+1}^i\|$ 。从 $i=0$ 时开始重复下述过程直至 $i=N-M-1$ 。

① 计算 $d_k^i (k=1, 2, \dots, N-i-1)$,并找出 j ,使得 $d_j^i < d_k^i$ 对所有的 $k \neq j$ 成立。

$$\textcircled{2} \text{ 计算 } \mathbf{x}_k^{i+1} (k=1, 2, \dots, N-i-1): \mathbf{x}_k^{i+1} = \begin{cases} \mathbf{x}_k^i & k < j \\ \frac{f_k^i \mathbf{x}_k^i + f_{k+1}^i \mathbf{x}_{k+1}^i}{f_k^i + f_{k+1}^i} & k = j \\ \mathbf{x}_{k+1}^i & k > j \end{cases} \quad (11)$$

$$\textcircled{3} \text{ 计算 } f_k^{i+1} (k=1, 2, \dots, N-i-1): f_k^{i+1} = \begin{cases} f_k^i & k < j \\ f_k^i + f_{k+1}^i & k = j \\ f_{k+1}^i & k > j \end{cases} \quad (12)$$

时间规整网络从语音信号的局部入手,依次合并具有最小帧距离的语音分析帧,合并过程有效地将语音过渡段融入各音素段,从而保证了最后的识别率不受说话人发音特点的影响,使得整个识别系统呈现出更强的鲁棒性^[13]。

(2) 每个成员的混合权值等于一个成员划分内的特征矢量数与训练数据集特征总数之比。

(3) 每个成员的码本由成员划分内的特征矢量通过模糊 C 均值聚类算法(Fuzzy c-means algorithm FCM)聚类算法得到。FCM 聚类是在引入模糊 C 划分后,对传统 K 均值聚类算法的模糊推广,它通过隶属度函数引入不确定性思想,实现对硬聚类算法的有效扩展,在实际应用中取得了非常优良的效果^[14]。首先定义 FCM 聚类算法目标函数为

$$J_{\text{FCM}}(\mathbf{x}, u, a) = \sum_{i=1}^n \sum_{k=1}^K u_k^m(\mathbf{x}_i) d(\mathbf{x}_i, a_k) \quad (13)$$

式中: $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ 是对应于某一成员划分的训练矢量序列; $a = \{a_1, a_2, \dots, a_K\}$ 是类聚中心; $u = \{u_1, u_2, \dots, u_K\}$ 是 FCM 隶属度函数,它满足 $0 \leq u_k(\mathbf{x}) \leq 1 (1 \leq k \leq K, \sum_{k=1}^K u_k(\mathbf{x}) = 1)$; $m \in [1, \infty)$ 为模糊度; $d(\mathbf{x}_i, a_k)$ 为距离。由目标函数得到的 FCM 类聚算式如下

$$\begin{cases} a_k = \frac{\sum_{i=1}^n u_k^m(\mathbf{x}_i) \cdot \mathbf{x}_i}{\sum_{i=1}^n u_k^m(\mathbf{x}_i)} & 1 \leq k \leq K \\ u_k(\mathbf{x}_i) = \left(\sum_{j=1}^K \frac{d(\mathbf{x}_i, a_k)^{2(m-1)^{-1}}}{d(\mathbf{x}_i, a_j)^{2(m-1)^{-1}}} \right)^{-1} & 1 \leq k \leq K, 1 \leq i \leq n \end{cases} \quad (14)$$

FCM 算法的收敛性在文献[14]中给出了证明。在迭代计算聚类中心 a_k 及隶属度函数 u_k 直到收敛后,由新的聚类中心组成重估后的新码本。

4 实验结果与分析

实验中采用的语音数据取自在实验室环境下录取 10 个人(5 男、5 女)的语音,其中每人对 10, 20, 30, 40 s 左右时间长度的一段语音各说 3 遍作为训练数据,通过用这些数据进行训练得到每个人的模型参数。另外每人对 30 个不同的字词各说 3 遍共 900 个语音作为测试数据进行测试。

输入语音信号经 12 kHz 采样, $1-0.98z^{-1}$ 的预加重,窗长 21.33 ms(256 点),窗移 10 ms 的汉明窗后,进行 14 阶线性预测编码(Linear prediction coding, LPC)分析,然后从 14 阶 LPC 系数中求出 12 阶的倒谱系数和倒谱的 12 阶线性回归 1 次系数,这些语音参数用于说话人识别实验。实验对 FVQMM 方法和 GMM 方法进行了比较,结果如表 1 和表 2 所示。

表 1 GMM 方法的说话人平均辨识度

Table 1 Average speaker recognition rate of GMM algorithm

训练数据/s	成员个数						
	3	6	10	16	18	20	25
10	奇异阵	奇异阵	奇异阵	0.432	0.689	0.753	0.779
20	奇异阵	奇异阵	奇异阵	0.792	0.821	0.882	0.872
30	奇异阵	奇异阵	0.774	0.845	0.882	0.932	0.882
40	奇异阵	0.782	0.877	0.886	0.920	0.965	0.935

表 2 FVQMM 方法的说话人平均辨识度(码本尺寸为 32)

Table 2 Average speaker recognition rate of FVQMM algorithm (Codebook size is 32)

训练数据/s	成员个数						
	3	6	10	16	18	20	25
10	0.574	0.644	0.674	0.889	0.732	0.753	0.579
20	0.774	0.864	0.894	0.929	0.925	0.925	0.782
30	0.876	0.892	0.913	0.936	0.928	0.920	0.887
40	0.905	0.923	0.937	0.950	0.943	0.932	0.922

从表 1 和表 2 中可以看出,虽然最优结果出现在 GMM,在训练数据为 40s,成员函数总数为 20 时,识别率为 96.5%。而 FVQMM 的最好结果为 95%,没有超过 GMM 的最好结果,但是,当训练数据较少时,FVQMM 的结果明显好于 GMM。而且,当训练数据较少时成员函数混合总数为 3, 6 和 10 时,GMM 训练会出现奇异阵,所以这是基于 EM 算法的 GMM 方法存在的重大缺陷。为了找到最优的 FVQMM 各成员码本的码本尺寸,在成员个数一定的条件下(即取最好识别结果的成员个数为 16,见表 2),测量不同的码本尺寸的识别结果。实验结果如表 3 所示。

表 3 不同码本尺寸时 FVQMM 方法的说话人平均识别率

Table 3 Average speaker recognition rate of FVQMM algorithm under different codebook sizes

训练数据/s	成员个数					
	4	8	16	32	64	128
10	0.474	0.644	0.896	0.889	0.489	0.353
20	0.637	0.864	0.932	0.929	0.782	0.552
30	0.776	0.892	0.944	0.936	0.812	0.686
40	0.805	0.922	0.935	0.950	0.882	0.692

从表 3 可以发现,码本尺寸为 16 时识别结果比较好。这是因为在训练数据有限的情况下,码本分的太细反而会使代表码字的误差变大。从表 1~3 中可知,本文提出的 FVQMM 的识别性能一般优于传统的 GMM,从而证明了这种新方法的有效性。

从另一个角度来分析,在一定的条件下,例如学习数据有限时,过多的模型参数往往不能得到好的学习精度。同时过多的模型参数也会给学习和识别增加运算成本,不利于实时的识别处理。为了分析比较提出的 FVQMM 和传统的 GMM 的参数数量,表 4 计算了取得最好结果时各模型所具有的参数总数。表中 L 表示码本尺寸, $P=24$ 表示语音特征矢量维数, M 表示成员混合数。从表中可知 FVQMM 参数总数比 GMM 要少的多。所以,虽然 FVQMM 是传统 GMM 的特殊形式,由于模型参数较少,在实验中取得了较好的识别结果。另外,本文提出的方法在 FVQMM 参数估计中,通过 FCM 聚类分析进一步减少了码本的量化误差,提高了码本的精度。

表 4 各模型具有的参数总数

Table 4 Total number of parameters in each model

模型类别	模型参数总数	
	计算式	计算值
GMM	$M(P^2 + P + 1)$ ($M=20, P=24$)	12 020
FVQMM	LMP ($L=16, M=16, P=24$)	6 144

5 结束语

本文提出了基于 FVQMM 的说话人识别新方法,作为 GMM 的改进形式,它综合考虑了 VQ 和 GMM 方法的优点。该方法既可以弥补传统 VQ 方法对于由说话人个性特征长时变动性等引起的语音特征的变化吸收能力较弱的缺点,又可以弥补传统 GMM 在建模时需要较多的训练数据的缺点。由于其模型参数数量较传统 GMM 少,模型学习对训练数据量要求不高,所以具有学习收敛速度快、识别速度快、适于实时自适应学习和小样本数据的说话人识别系统等特点。实验证明该方法在模型训练数据较少的情况下可以取得比传统 GMM 和 VQ 方法更好的识别性能。下一步的工作是实验验证 FVQMM 方法是否可以应用到现在主流的 GMM-UBM 说话人识别方法中,并通过主流的 NIST 评测来进行实验验证。

参考文献:

- [1] 赵力. 语音信号处理[M]. 北京:机械工业出版社:2009,236-253.
- [2] Soonil K, Shrikanth N. Robust speaker identification based on selective use of feature vectors[J]. *Pattern Recognition Letters*, 2007 (28): 85-89.
- [3] Dehak N, Dehak R, Kenny P. et al. Comparison between factor analysis and GMM support vector machines for speaker verification[C]// *The Speaker and Language Recognition Workshop (Odyssey 2008)*. Stellenbosch, South Africa: ISCA Ar-

chive, 2008; 21-25.

- [4] Campbell W M, Sturim D E, Reynolds D A, et al. SVM based speaker verification using a GMM supervector kernel and NAP variability compensation[C]// Acoustics, Speech and Signal Processing of ICASSP 2006, IEEE International Conference on. Toulouse, France: IEEE, 2006: 97-100.
- [5] Yin Shou-Chun, Richard R, Patrick K, A joint factor analysis approach to progressive model adaptation in text-independent speaker verification[J]. IEEE Trans on Audio Speech and Language Processing, 2007, 15:1999-2110.
- [6] 何勇军, 孙广路, 付茂国, 等. 基于稀疏编码的鲁棒说话人识别 [J]. 数据采集与处理, 2014, 29(2):198-203.
He Yongjun, Sun Guanglu, Fu Maoguo, et al. Robust speaker recognition based on sparse coding [J]. Journal of Data Acquisition and Processing, 2014, 29(2):198-203.
- [7] 花城, 李辉. 小训练语料下基于均值超向量聚类的说话人确认方法 [J]. 数据采集与处理, 2014, 29(2):238-242.
Hua Cheng, Li Hui. Speaker verification based on supervector clustering with poor corpus [J]. Journal of Data Acquisition and Processing, 2014, 29(2): 238-242.
- [8] 王华朋, 杨军, 许勇. 应用似然比框架的法庭说话人识别[J]. 数据采集与处理, 2013, 28(2):240-243.
Wang Huapeng, Yang Jun, Xu Yong. Forensic speaker recognition in likelihood ratio framework [J]. Journal of Data Acquisition and Processing, 2013, 28(2):240-243.
- [9] 王敏, 赵鹤鸣, 张庆芳. 基于瞬时频率估计和特征映射的汉语耳语音者识别[J]. 数据采集与处理, 2011, 26(2):687-690.
Wang Min, Zhao Heming, Zhang Qingfang. Speaker identification with Chinese whispered speech based on instantaneous frequency estimation and feature mapping[J]. Journal of Data Acquisition and Processing, 2011, 26(2):687-690.
- [10] 奚吉, 赵力, 左加阔. 基于改进多核学习的语音情感识别算法[J]. 数据采集与处理, 2014, 29(5): 730-734.
Xi Ji, Zhao Li, Zuo Jiakuo. Speech emotion recognition based on modified multiple kernel learning algorithm [J]. Journal of Data Acquisition and Processing, 2014, 29(5): 730-734.
- [11] Shore J E, Burton D K Discrete utterance speech recognition without time alignment[J]. IEEE Trans, 1983,29(4):473-491.
- [12] 李忠国, 侯杰, 王凯, 等. 模糊支持向量机在路面识别中的应用[J]. 数据采集与处理, 2014, 29(1): 146-151.
Li Zhongguo, Hou Jie, Wang Kai, et al. Application of fuzzy support vector machine on road type recognition[J]. Journal of Data Acquisition and Processing, 2014, 29(1): 146-151.
- [13] Chen Sin Horng, Chen Wen Yuan. Generalized minimal distortion segmentation for ANN-based speech recognition[J]. IEEE Trans on Speech and Audio Processing, 1995,3(2):141-145
- [14] Bezdek J C. A convergence theorem for the fuzzy ISODATA clustering algorithms[J]. IEEE Trans,1990(2):1-8.

作者简介:



杨彦(1974-),女,副教授,研究方向:信号处理,电子与通讯, E-mail: yfyangyan@126.com。



赵力(1958年-),男,教授,研究方向:信号处理。

