模糊 C-均值 (FCM) 聚类法与矢量量化法相结合用于说话人识别 [1]

吴晓娟 韩先花 聂开宝

(山东大学信息科学与技术学院 济南 250100)

摘 要 该文提出了一种将模糊 C-均值聚类法与矢量量化法相结合进行说话人识别的方法。该算法将从语音信号中提取的 12 阶 LPC(线性预测编码) 倒谱系数作为待分类样本的 12 个指标，先用矢量量化法求出每个说话人表征特征参数的码书，作为模糊聚类算法的聚类中心，最后将待识别的特征矢量以得到的码书为聚类中心，进行聚类识别。该算法所使用的特征参数较少，计算比较简单，但识别率较矢量量化法高。

关键词 模糊聚类，矢量量化，说话人识别，语音特征

中图号 TP391.42, TN912.3

1 引 言

说话人识别就是通过一段语音识别讲话者是谁。随着现代科技的发展，对说话人识别方法的研究得到了极大的关注，已有不少学者对此做出了卓有成效的贡献。

现代说话者识别技术作为具有语音识别与理解功能的智能人-机接口，是新一代计算机的重要组成部分，其应用领域正在不断扩大。它使人员进入系统的身份认证成为可能，即通过声音进行进入控制。此技术实现与身份认证有关的各种服务包括：电话预约服务——以用户的声音实现汇款、转账、余额通知、转款、股票行情信息咨询等；特定人声音实现机密保管场所的出入人员检查；声控电子密码锁；法庭认证以及医学方面的应用。

说话人识别系统性能好坏的关键之一是其所选择的识别方法。本文所使用的方法是先对训练资料进行矢量量化，得到用于识别的码书作为各类别的聚类中心，再用模糊 C-均值聚类方法进行说话人的识别。

模糊聚类分析的思想首先是由 Bellman 提出来的；1973 年，Dunn 发展了 FCM 算法。目前有关模糊聚类的许多成果都是对 C-均值聚类算法的推广和改进。本文采用此算法对语音特征参数进行分类，将同一人的某些语音特征聚为同一类，从而有效的识别出说话人。此方法计算量较小，算法不很复杂，但识别效果较好。

2 语音特征参数 [1,2]

2.1 语音信号预处理 首先根据语音信号的短时能量和短时平均过零率可确定语音信号的有无。当短时能量和过零率都很小时，可判定无语音信号。其次必须对语音信号进行预加重处理。本文以 8kHz 采样的语音信号 $s(n)$ 经一阶数字滤波器加重成为 $\check{s}(n)$, $\check{s}(n)=s(n)-0.95s(n-1)$ 。再就是分帧及加窗。进行 LPC 以及倒谱分析时，将语音信号分割成 20ms，前后相邻帧之间各有 5ms 重迭。我们采用汉明 (Hamming) 窗来消除由于分帧引起的信号边缘蜕变。

2.2 LPC 倒谱系数 倒谱特征是说话人个性特征表征和说话人识别的最有效的特征之一。对语音信号进行倒谱分析可使其声道频率特性和激励信号源有效的分离，从而提取出和声道有关的表征不同说话人的个性特征。在语音信号处理中，语音产生模型通常用全极点模型。通过它得到 LPC，LPCC(LPC 倒谱) 参数。它为语音谱的包络提供了很好的近似，比直接由 FFT(离散傅里叶变换) 得到的语音谱平稳。

语音的 LPC 参数分析即是用语音信号对过去 P 个时刻的采样值的线性组合以最小预测误差预测下一时刻的信号采样值，其时域模型表示式为

$$s(n)=\sum a_i s(n-i)+Gu(n),\quad i=1,2,\cdots,P$$

[1] 2000-12-06 收到，2001-04-30 定稿
中科院沈阳自动化所开放实验室基金和省自然科学基金 (Y2001G04) 资助

其中 a_i 是 LPC 参数， $s(n)$ 是原始语音， $u(n)$ 是激励源信号，本文中 P 取 12 。

LPC 系数直接反应了声道发声参数，它是发音内容与声道共同作用的结果 (包括了个性特征和语音特征) 。由倒谱参数得到的谱包络比由 LPC 参数得到谱包络要平滑、稳定，更多的表示了说话的个性特征，因此识别率比较高 [3] 。

3 模糊 C-均值聚类算法

令 $X=\{x_1,x_2,\cdots,x_N\}\subset R^n$ 为 n 维实数空间 R^n 中的一个有限样本数据子集； $x_k=(x_{k1},x_{k2},\cdots,x_{kn})\in R^n$ 称为特征矢量或模式矢量， x_{kj} 为模式矢量 x_k 的第 j 个特征。模式聚类 (分类) 是将 X 分成 C 个子域： $K_1,K_2,\cdots,K_C$ 使其满足： K_i 交 K_j 为空集 (当 $i\neq j$) ，它们所有的并集为 X 。

对任意 X ，所有可能的模糊 C-划分组成的集合记为

$M_{fc}=\{U\in V_{C\times N}|u_{ik}\in[0,1],\forall I,\ k;\ \sum u_{ik}=1(i=1,2,\cdots,C\ \text{求和}),\forall k;\ 0\leq\sum u_{ik}\leq N,\forall i\}$ 其中 U 矩阵中的每一行元素 $\{u_{ik},1\leq k\leq n\}$ 定义了 X 中的一个模糊聚类 $\hat{C}_i$; u_{ik} 为第 k 个样本 (模式) 对第 i 个类的模糊隶属度， $V_{C\times N}$ 是 $C\times N$ 维的矩阵集合。

$$U=\begin{bmatrix}U_1\\U_2\\\vdots\\U_C\end{bmatrix}=\begin{bmatrix}u_{11}&u_{12}&\cdots&u_{1N}\\u_{21}&u_{22}&\cdots&u_{2N}\\\vdots&\vdots&\vdots&\vdots\\u_{C1}&u_{C2}&\cdots&u_{CN}\end{bmatrix}$$

$$\hat{C}_i=u_{i1}/x_1+u_{i2}/x_2+\cdots+u_{iN}/x_N=\sum u_{ik}/x_k$$

这里， $\sum$ 不是求和，而是表示各个元素与隶属度函数对应关系的一个总括。

设 L_C 表示 X 张成的 C 重迪卡儿积，模糊聚类的目标函数可定义如下： $J_m:M_{fc}\times L_C\to R$

$$J_m(U,v)=\sum\sum(u_{ik})\hat{m}\|x_k-v_i\|^2,\qquad 1\leq m<\infty$$

上式中， $\|\bullet\|$ 为 R^n 上任意一种内积范数； $v=(v_1,v_2,\cdots,v_C)\in L_C$, $v_j\in R^n$ $(1\leq j\leq C)$，这个 v_j 可看作上式中 U 所定义的 C 个聚类的原型或聚类中心， $m\in[0,\infty]$ 是一个模糊加权指数。

FCM 算法来自对以下公式的优化求解步骤，即通过下式的一阶必要条件间的迭代来实现。

$$\underset{M_{fc}\times I_C}{\text{Min}}\ \{J_m(U,v)\}$$

在 R^n 空间中， d_{mk} 表示 X_m 和 X_k 间的距离，定义为

$$d_{mk}^2=\|X_m-X_k\|^2=\sum w_m(x_{mj}-x_{kj})^2,\qquad 0\leq w_i\leq 1,\quad 0\leq\sum w_i\leq 1$$

w_j 为第 j 个特征的权重数。上式对于变量 (U,v) 的一阶必要条件分别如下：

$$u_{ik}=1,\exists x_k=v_i;\ u_{jk}=0,\ j\neq i;\quad 1\leq i\leq C,\quad 1\leq k\leq N$$

$$u_{ik}=1/\sum_{j=0}^{C}\left(\|x_k-v_i\|^2/\|x_k-v_j\|^2\right)^{(1/(m-1))}$$

$$v_i=\sum_{k=1}^{n}(u_{ik})^m x_k/\sum_{k=1}^{n}(u_{ik})^m,\qquad (1\leq i\leq C)$$

FCM 算法能最终收敛到一个局部极小点或鞍点，得到 X 的一个模糊 C 划分 [4-6]。

4 模糊聚类法与矢量量化法相结合的说话人识别 [7]

4.1 *矢量量化算法*　矢量量化 (VQ) 是一种高效的数据压缩技术，是标量量化的自然发展。它将 n 维欧式空间中的模拟矢量 x 依某种准则用 n 维空间的有限个点 $\{y_i|i=1,2,\cdots,K\}$ 表示。我们采用的是依失真度最小为准则进行矢量量化的 LBG 算法，这是一种最优的矢量量化器设计方法，识别率较高。

本实验中，对每一个说话人 $n(n=1,2,\cdots,N)$ 用 LBG 算法建立一个大小为 M 的码本 $B^n=\{b_j^n|j=1,2,\cdots,M\}$，即根据每一说话人所发训练语音的特征矢量，通过 VQ 聚类得到该说话人的码本，VQ 过程中所采用的特征矢量 $\boldsymbol{a}$ 与特征矢量 $\boldsymbol{b}$ 之间的距离记为 $d(\boldsymbol{a},\boldsymbol{b})$。本文选择的码本大小 M 为 10，一般其值在实验过程中根据识别率而定。

为提高识别率，还须对原始量化码书重新排序，即进行码号变化。其遵循的原则是：若空间的各量化码字与某一量化码字距离相等，则应使它们与该码字的码号距离尽量相等。

4.2 *设计特征参数的码书*　我们所使用的实验数据是 5 个人 (3 男 2 女) 所发的 10 个阿拉伯数字 (每个发 10 遍)，其中 4 遍作为训练数据，提取码书，从每一个数字语音中提取出它的倒谱系数，作为特征向量，用向量量化法设计每一说话人的码书。本实验中对每一说话人设计出 20 个码书，同一说话人的码书标为同一类 L_i(其中 i 为 $1,2,\cdots,5$, 表明共有 5 个类别，分别代表说话人 1 到说话人 5)，并将其用于后面的模糊聚类法的聚类中心 (共 100 个聚类中心)。

4.3 *特征参数的规格化*　由于各个样本的 M 个指标的量纲和数量级不一定相同，直接利用原始数据进行计算，就可能突出某些数量级特别大的特性指标对分类的作用，而降低甚至排斥某些数量级较小的特性的作用，导致一个指标只要改变一下单位，也会改变分类结果。所以必须对原始数据进行无量纲化处理，使每一指标值统一于某种共同的数据特性范围。

我们采用均值规格化对特征数据进行处理。此方法不改变原有数据的变异程度，也不易受个别极端值的影响。

4.4 *模糊聚类法识别*　待识别的数据为 5 个人发的 6 遍 10 个阿拉伯数字，对每个数字语音中提取的特征矢量进行预处理后，作为聚类样本。以各个说话人的码书作为聚类中心，根据上面介绍的模糊聚类公式对各帧语音进行分类识别，即将得到的某语音的各特征矢量以各个说话人中的码书为聚类中心进行聚类。在计算中，若一段语音的其中一个特征矢量聚类到某一类 L_i(其中 i 为 $1,2,\cdots,5$) 的码书上，我们便将此特征矢量标明是这一类码本所表示的说话人所录语音的特征矢量，依此类推，得到此段语音的各个特征矢量聚类到码本类别，即得到它们所属的说话人，这样再对此语音的所有特征矢量所属类别进行统计，找出这些特征矢量归为最多的某一类，则可断定这段语音为该类所属的说话人所发。

基于 VQ 与模糊聚类法相结合的说话人识别系统结构图如图 1 所示。

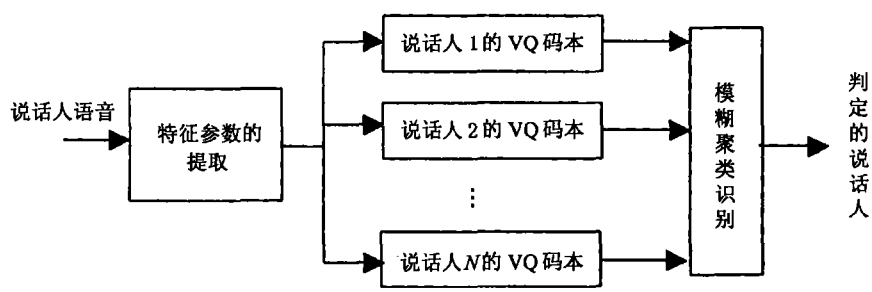

图 1　基于 VQ 与模糊聚类法相结合的说话人识别系统结构图

5 实验结果

在实验中，系统用于训练的语音数据是 5 名说话人 (男 3 名，女 2 名) 相隔半年时间在普通实验室的两次录音，每次对每 10 个阿拉伯数字录 5 次音，共得到 10 个阿拉伯数字的 10 次发音。我们分别对每次录音的各个阿拉伯数字取两次发音，即对 4 次发音提取 VQ 码本，使其作为模糊聚类的聚类中心，其余 6 次发音用于识别测试，这样每人有 60 个数据用于识别，即总共对 5 个人的 300 个数据进行了测试。结果表明：本方法的识别性能较高，比单独使用矢量量化法的识别效果要好得多。表 1 是对仅用矢量量化法进行识别和用本文提出的算法进行识别的识别率比较。本文算法的实验结果示于表 2 中。

由表 2 可见，用本文提出的进行说话人识别的新方法——基于矢量量化与模糊聚类法相结合的说话人识别方法，能将某一说话人的语音中提取的百分之九十以上的特征矢量都聚类到此说话人的码本上，从而能正确地判别出某一段语音为哪位说话人所说。此方法和仅用矢量量化法进行说话人识别比较，识别性能有了明显的改善，是一种行之有效的说话人识别方法。

表 1 两种算法的误识率比较

误识率	说话人 1 码本	说话人 2 码本	说话人 3 码本	说话人 4 码本	说话人 5 码本	所有码本
矢量量化 (VQ) 方法 (%)	10.3	12.2	12.6	11.1	9.8	11.2
矢量量化与模糊 C-均值聚类结合 (%)	5.2	6.1	5.8	4.9	5.0	5.4

表 2 矢量量化法与模糊聚类法相结合的说话人识别的实验结果

聚类百分比	说话人 1 码本	说话人 2 码本	说话人 3 码本	说话人 4 码本	说话人 5 码本
说话人 1 的语音 (%)	95.3	1.07	0.86	1.25	1.52
说话人 2 的语音 (%)	0.99	94.1	1.18	1.80	1.93
说话人 3 的语音 (%)	0.72	0.81	96.2	1.21	1.06
说话人 4 的语音 (%)	1.24	2.03	1.89	93.8	1.04
说话人 5 的语音 (%)	0.71	1.71	1.52	0.96	95.1

参 考 文 献

[1] 朱民维，计算机语音技术，北京，北京航空航天大学出版社，1991, 39–86.

[2] 胡光锐，语音处理与识别，上海，上海科学技术文献出版社，1994, 200–297.

[3] 马卡尔着，娄乃英译，语音信号线性预测，北京，中国铁道出版社，1997, 第一章.

[4] Yu Dantong, Zhang Aidong, ACD: An automatioc clustering and querying approach for large image database[C], In: ACM Multimedia'99 Proc., Orlanda, Florida, USA, 1999, 95–98.

[5] B. S. Everit, Cluster Analysis, 3rd.ED., New York, Halsted Press, part1~part3, 1993.

[6] 刘增良，模糊技术与神经网络技术选编，北京，北京航天航空大学出版社，1995, 120–157.

[7] S. B. Davis, P. Mermelstein, Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences, IEEE Trans. on ASSP, 1980, 28(4), 357–366.

SPEAKER RECOGNITION USING FUZZY C-MEAN CLUSTERING ALGORITHM AND VECTOR-QUANTIZATION(VQ) ALGORITHM

Wu Xiaojuan Han Xianhua Nie Kaibao

(*School of Information Science and Engineering, Shandong Univeristy, Jinan* 250100, *China*)

Abstract In this paper, an efficient method for speaker recognition—the combination of VQ (Vector-Quantization) algorithm with fuzzy C-mean clustering algorithm is proposed. This algorithm extracts 12th order LPC cepstrum coefficients from speech signals and makes them the marker of those samples, which will be classified. At first, codebooks which can represent those feature parameters of each speaker are figured out, and used as the clustering centers of speaker recognition. Finally, all speakers' feature parameters are identified from each other with fuzzy C-mean clustering algorithm in which the clustering centers are these codebooks which have been obtained using VQ algorithm. With relatively less feature parmeters and simpler computation, the proposed algorithm has a higher recognition rate compared with VQ algorithm.

Key words Fuzzy clustering, Vector-Quantization(VQ), Speaker identification, Speech characteristic

吴晓娟: 女, 1944 年生, 教授, 研究方向: 智能测量与控制、信号处理和模式识别.
韩先花: 女, 1976 年生, 硕士生, 研究方向: 信号处理和模式识别.
聂开宝: 男, 1966 年生, 副教授, 研究方向: 生物医学信号处理、图象和语音信息处理、模式识别等.