

文章编号 1004-924X(2010)01-0227-07

# 改进的基于决策树的说话人在线聚类

张素敏<sup>1,2</sup>, 苏东林<sup>1</sup>, 王 炜<sup>3</sup>

- (1. 北京航空航天大学 电子信息工程学院, 北京 100083;
2. 中国电子科技集团公司 第 54 研究所, 河北 石家庄 050081;
3. 空军装备研究院, 北京 100085)

**摘要:** 针对采用传统的在线聚类方法时后续判决错误率较高的缺点, 提出了一种改进的基于决策树的在线说话人聚类算法。通过构建一个决策树, 增加判决分支, 对语音段进行判决聚类, 从而有效降低前期错误判决对后续聚类的影响。为了进一步提高算法效率, 缩短运算时间, 还给出了一种决策树剪枝方法, 减少了不合理的判决分支。通过对广播新闻语料进行的说话人聚类实验表明, 相比传统的层次聚类算法, 新算法的平均类纯度和说话人纯度分别提高了 0.9% 和 1.1%, 计算时间减少了 57%。实验结果还表明, 相比手工标注说话人信息, 将该算法的聚类结果应用于说话人自适应可降低系统的误识率。

**关键词:** 说话人聚类; 在线聚类; 决策树; 剪枝算法

**中图分类号:** TP391.9 **文献标识码:** A

## Improved online speaker clustering based on decision tree

ZHANG Su-min<sup>1,2</sup>, SU Dong-lin<sup>1</sup>, WANG Wei<sup>3</sup>

- (1. School of Electronic and Information Engineering, Beihang University, Beijing 100083, China;
2. The 54th Research Institute of Chinese Electron Technology Corporation, Shijiazhuang 050081, China;
3. Institute of Air-force Equipment, Beijing 100085, China)

**Abstract:** Speaker clustering is a key component in many speech processing applications. To solve the problem of error propagating in the posterior clustering caused by the traditional online clustering, an improved online speaker clustering algorithm based on a decision tree is proposed. Unlike typical online clustering approaches, the proposed method constructs a decision tree to increase branches and to distinguish an audio segment clustering to reduce effectively the effect of error distinguishing on the posterior clustering. To shorten the operation time, a pruning strategy for candidate-elimination is also presented. Experiments indicate that the algorithm achieves good performance on both precision and speed. By using this method, the average speaker purity and the average cluster purity have improved by 0.9% and 1.1% respectively, and the time consuming is reduced by 57%. Experiments also show that this method is effective for improving the performance of the unsupervised adaptation as compared with the true speaker-condition.

**Key words:** speaker clustering; online clustering; decision tree; pruning strategy

收稿日期: 2009-10-20; 修订日期: 2009-10-23.

基金项目: 国家 863 高技术研究发展计划资助项目 (No. 2006AA701418)

## 1 引言

说话人自适应(Speaker Adaptation)已被公认为是提高自动语音识别(Automatic Speech Recognition, ASR)系统性能的有效手段,其根据说话人的少量数据调整系统参数,使得模型更偏向于当前说话人,从而提高了语音识别系统的性能<sup>[1-3]</sup>。在实际应用中,由于语音识别系统的输入一般为连续的语音数据,语音的边界及对应的说话人都是未知的,因此系统需要增加语音分段(Speech Segmentation)模块,以及分割后语音段的说话人聚类(Speaker Clustering)模块<sup>[4]</sup>。

说话人聚类是聚类技术在语音识别领域上的一个具体应用,其目的是通过对语音段进行分类,使每一类只包含同一说话人数据,且同一说话人数据都被归并到同一类中,从而获取说话人特定信息。通过说话人的类信息,语音识别系统可进行无监督说话人自适应。在不同识别任务上,无监督说话人自适应方法都取得了 10%~25% 的相对系统误识率<sup>[5]</sup>。因此,说话人聚类算法的好坏将最终影响语音识别的性能。除语音识别领域外,说话人聚类技术还应用于音频信息管理、检索等领域,它有助于在会议、语音邮件、讲座以及新闻广播节目的音频流中实现说话人跟踪,从而实现对音频数据进行结构化的分析、理解和管理<sup>[6-7]</sup>。

K-均值(K-Mean Clustering)算法是一种常用的聚类方法,其通过迭代优化的方法对数据进行划分。由于在语音识别中,说话人实际数目往往是未知的,因此 K-均值算法仅适用于说话人数目已知的场合。层次聚类(Hierarchical Clustering, HC)算法是广泛应用于语音识别领域的一种聚类方法,其通过合并或分裂分层次地融合小类或分割大类。HC 的局限性在于:(1)算法速度较慢;(2)无法进行在线(on-line)聚类,增加样本后需要重新进行全局的聚类<sup>[8]</sup>。在线聚类(On-line Clustering)算法通过对每个到来的语音段进行聚类判决,较好地解决了 HC 方法的局限性,因此近年在实时性要求高的语音识别场合被广泛应用<sup>[9]</sup>。但线聚类在聚类初始阶段存在数据样本不

足的弊端,由此带来的判决错误将不可逆地影响后续的聚类判决<sup>[10]</sup>。针对层次聚类及在线聚类的优缺点,本文提出一种改进的基于决策树的说话人在线聚类算法(Decision-Tree-based Clustering, DTC),在传统的在线聚类算法基础上引入判决树,通过增加判决分支来降低错误判决对后续聚类的影响,并通过剪枝策略减少不合理的判决分支来提高算法效率。

## 2 层次聚类法与在线聚类法

广义上讲,聚类技术是基于数据集内部存在若干“分组”而产生的一种数据描述方法,其每个子集中的数据具有高度的内在相似性,因此聚类算法需要考虑:(1)相似度的定义;(2)“分组”划分好坏的衡量(即准则函数)。相似度度量决定着如何度量样本之间的相似性,不同的度量函数适用于不同的样本,因此选择适合的度量函数是聚类的关键。常用的相似度度量有欧氏距离、Kullback 距离等<sup>[11-12]</sup>。常用的聚类策略是合并最靠近的两类,它体现了最小方差的思想,但这样做过于依赖于所选用的相似度度量,因此更好的做法是通过极值化某一准则函数,通过其选择需合并的类,使得准则函数衡量的是样本集划分的好坏。常用的准则函数有:误差平方和准则(Sum-of-squared-error Criterion)、散布准则等<sup>[11-12]</sup>。

### 2.1 层次聚类法

层次聚类是一种倒树型结构聚类算法,可以通过合并和分裂两种途径实现。合并(自底向上,聚类树的叶节点是底)时,先使得每个样本各成一类,然后通过合并不同的类,来不断减少类别数目。分裂(自顶向下,聚类树的根节点是顶)时,先将所有样本归入一类,然后通过分裂来增加类别数目。本文主要讨论基于合并的层次聚类,图 1 以树图的形式显示了基于合并的层次聚类过程。 $x_1 \sim x_8$  是数据集中的 8 个样本点,图中纵坐标表示类与类的相似度标尺。在  $k=1$  层,所有 8 个点各成一类。在  $k=2$  层,最相似的  $x_6$  和  $x_7$  合并为一类。如此继续,在  $k=8$  层所有的样本点聚为 1 类。

考虑音频段集合  $X = \{x_1, x_2, \dots, x_N\}$ , 每个

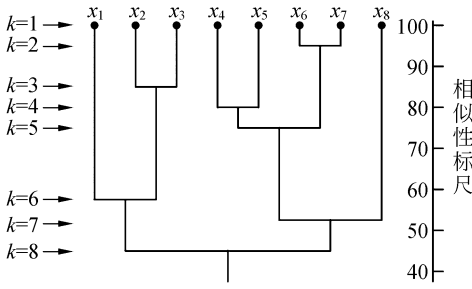


图 1 层次聚类算法的树形示意图

Fig. 1 Construction of hierarchical clustering

$x_n$  代表一个频谱特征序列。层次聚类法可描述如下:

- (1) 初始化聚类数目  $c: c \leftarrow N$ ;
- (2) 计算每两个类之间的距离;
- (3) 在  $c$  类中找到最近的两类:  $x_i$  和  $x_j$ ;
- (4) 将  $x_i$  与  $x_j$  合并成一个新类;
- (5) 更新其它类与新类的距离;
- (6) 计算全局准则函数:  $G(c), c \leftarrow c - 1$ ;
- (7) 重复(3)~(6), 直到  $c = 1$ ;
- (8) 计算说话人数目  $c \leftarrow \underset{c}{\operatorname{argmin}} G(c)$  以及  $c$  个说话人类;

其中  $c$  是指说话人类别的假设数目,  $G(c)$  是全局类内准则函数:

$$G(c) = \left| \sum_{j=1}^c N_j \Sigma_j \right| \sqrt{c}, \quad (1)$$

$N_j$  是类  $j$  的特征参数数目, 而  $\Sigma_j$  是类  $j$  的协方差矩阵,  $|\cdot|$  代表行列式函数。

### 2.2 在线聚类

DSC 算法能够在语音段数目确定不变的情况下, 对语音数据进行说话人聚类的方法。然而在实际应用中, 系统常需要处理随机到来的新的语音数据。下文描述一种混合说话人在线聚类算法 (Hybrid Speaker Clustering, HSC)<sup>[9]</sup>:

- (1) 初始化上下门限值:  $\theta_u$  与  $\theta_l$ , 首个说话人类  $D_1 \leftarrow x$ , 类别数目  $c: c \leftarrow 1$ ;
- (2) 引入新的语音段  $x$ ;
- (3) 计算新段与类间最近的距离:  $d_j \leftarrow \min_j GLR(X, D_j), j \in [1, c]$ ;
- (4) 如果  $d_j < \theta_l$ , 则将语音段  $x$  合并入类  $D_j$ , 并转到(10);
- (5) 如果  $d_j > \theta_u$ , 则添加新类  $D_{c+1} \leftarrow x, c \leftarrow c + 1$ , 并转到(10);
- (6) 将语音段  $x$  合并入类  $D_j$ , 计算  $G1: G1 \leftarrow G(D_j \cup \{x\}, D)$ ;
- (7) 添加新类  $D_{c+1} \leftarrow x$ , 计算  $G2: G2 \leftarrow G(x, D)$ ;
- (8) 如果  $G1 < G2$ , 将语音段  $x$  合并入类  $D_j$ , 转到(10);
- (9) 如果  $G1 > G2$ , 则添加新类  $D_{c+1} \leftarrow x, c \leftarrow c + 1$ , 并转到(10);
- (10) 重复(2)~(9), 直到没有新的语音段;
- (11) 返回  $D_1, D_2 \dots D_c$ 。

其中上下门限值  $\theta_u$  与  $\theta_l$  的设置可以加速聚类的计算速度。当  $\theta_l \rightarrow -\infty$  和  $\theta_u \rightarrow \infty$ , HSC 就等同另一种算法: 基于散布的说话人聚类算法 (Dispersion-based Speaker Clustering, DSC)<sup>[9]</sup>。

### 3 基于决策树的在线说话人聚类算法

在本文中, 类间距离度量函数采用广义概率比 (Generalized Likelihood Ratio, GLR)<sup>[13]</sup>。考虑两个语音段  $x$  和  $y$ 。假设  $L(x; \mu_x, \Sigma_x)$  是  $x$  在高斯模型  $N(\mu_x, \Sigma_x)$  下的似然值,  $L(y; \mu_y, \Sigma_y)$  是  $y$  在高斯模型  $N(\mu_y, \Sigma_y)$  下的似然值。假设  $x$  与  $y$  同属一个说话人,  $z$  是  $x$  与  $y$  的合并语音段,  $L(z; \mu_z, \Sigma_z)$  是  $z$  在高斯模型  $N(\mu_z, \Sigma_z)$  下的似然值。则 GLR 距离值计算如下:

$$GLR(x, y) = -\log \left[ \frac{L(z; \mu_z, \Sigma_z)}{L(x; \mu_x, \Sigma_x) L(y; \mu_y, \Sigma_y)} \right] = -\log \left[ \frac{|\Sigma_x|^\alpha |\Sigma_y|^{1-\alpha}}{|W|} \right]^{\frac{N}{2}} - \log \left[ 1 + \frac{N_x N_y}{N^2} (\mu_x - \mu_y)' W^{-1} (\mu_x - \mu_y) \right]^{-\frac{N}{2}}, \quad (2)$$

其中  $N_x$  和  $N_y$  分别是语音段  $x$  与  $y$  的帧数,  $N = N_x + N_y, \alpha = N_x / N, W$  是  $x$  与  $y$  协方差的加权平均:

$$W = (N_x \Sigma_x + N_y \Sigma_y) / N.$$

GLR 距离和散布准则已经在层次聚类和在

线聚类算法中证明是有效的,本文同样证明了 GLR 距离与散布准则在 DTC 算法中的有效性。DTC 算法主要由决策树和剪枝策略两个部分组成。

### 3.1 决策树

在图 2 中,水平轴代表语音段按照时间顺序到来。其中第 1 个语音段被初始化成首个类{1},并建立根节点(0,1)。第 2 个语音段与第 1 个类合并成新类{1,2},并建立叶节点(1,1);或形成一个新类{1}{2},并建立叶节点(0,1)。以此类推,当第 3 个语音段到来后,可形成 5 个新的叶节点。根据剪枝策略,在每个阶段仅保留  $N(N=2)$  个叶节点。

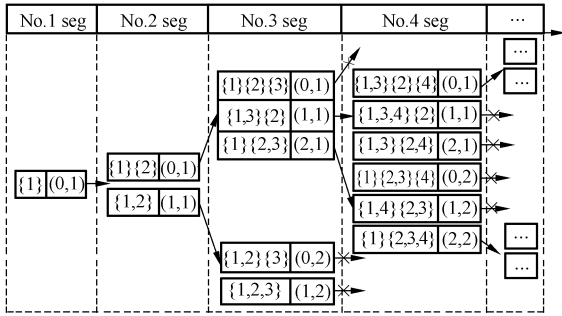


图 2 决策树的构造过程

Fig. 2 Construction procedure of decision tree

通过引入决策树,各种可能的聚类结果被记录下来,有效地提高了聚类算法的准确度,决策树算法可描述如下:

- (1) 初始化  $N$ ;
- (2) 初始化根节点(0,1),由  $D_1 \leftarrow x$  类构成;
- (3) 引入新语音段  $x$ ;
- (4) 取节点  $i$ ,由类集  $\{D_1^i, D_2^i, \dots, D_c^i\}$  组成;
- (5) 计算新语音段与所有类之间的距离:  $d_j^i \leftarrow GLR(x, D_j^i), j \in [1, c]$ ;

(6) 找到  $N$  个最近的距离  $d_{j_1}^i, d_{j_2}^i, \dots, d_{j_N}^i$  ( $d_{j_1}^i < d_{j_2}^i < \dots < d_{j_N}^i$ ),并构建一个节点堆栈(1,  $i$ ), (2,  $i$ ), ..., ( $N$ ,  $i$ );

(7) 假设  $x$  是一个新类,计算叶节点(0,  $i$ )的  $J2[i]$ 值:  $J2[i] \leftarrow J(\{x, D^i\})$ ;

(8) 取下一个叶节点直到没有叶节点为止;

- (9) 根据剪枝算法去除多余的叶节点,仅保留  $N$  个叶节点(剪枝算法将在 3.2 中详细介绍);
- (10) 重复(3)~(9)直到没有新的语音段;
- (11) 从  $N$  个节点中取出  $J$  值最小的叶节点,得出最终聚类结果:  $D_1, D_2, \dots, D_c$ .

其中  $N$  是每层聚类树的节点数,  $J2[i]$  是语音段形成新类的类间散度数组。每当一个新的语音段到来时,一个包含叶节点的节点堆栈将保留可能的聚类结果。通过剪枝策略的,包含  $N$  种概率最大聚类结果,即叶节点被保留在节点堆栈中。显然,对于 DTC 算法,算法复杂度将随着  $N$  的增大而增加。当  $N=1$  时, DTC 将等同 DSC 算法。在下文实验结果中将详细讨论树宽值  $N$  与聚类性能之间的关系。

### 3.2 剪枝策略

图 3 中右部分的堆栈即是上文 3.1 中形成的节点堆栈,由  $N(N=4)$  个节点组成的;图左部分即是  $J$  值数组。

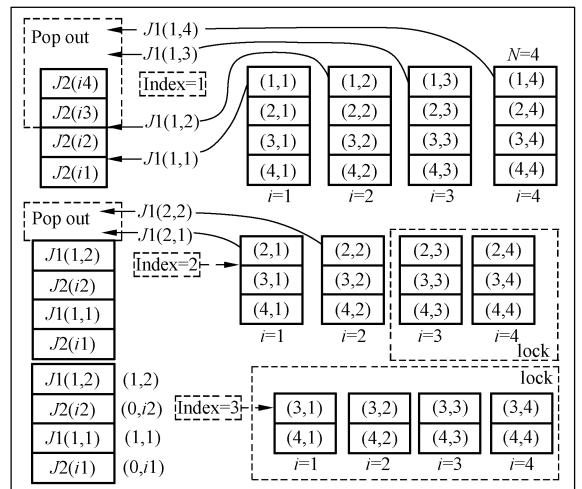


图 3 剪枝过程

Fig. 3 Procedure of pruning

剪枝策略的目的在于保留  $N$  个概率最高的聚类结果,用于后续音频段的聚类。其描述如下:

- (1) 初始化  $J$  值数组  $J2[i], i \in [1, N], index \leftarrow 1$ ;
- (2) 排序数组:  $J2[i_1] < J2[i_2] < \dots < J2[i_N]$ ;

(3) 将  $J1[index, i]$  插入  $J$  值数组,  $i \in [1, N]$ ;

(4) 排序  $J$  值数组, 留下  $N$  个最小的值, 弹出剩余的值;

(5) 锁定弹出  $J1$  值来自的节点堆栈;

(6)  $index \leftarrow index + 1$ ;

(7) 重复(3)~(6)直到没有节点堆栈;

其中  $J1[i, j]$  是节点的  $(i, j)$  类内散布值。

## 4 说话人聚类实验

### 4.1 数据集

实验采用的测试数据来源于 1997 年普通话广播新闻语料(Hub4-NE)数据集, 该语料集包括 CCTV、KAZN 和 VOA 三个频道的广播语料, 其主要内容是新闻报道和访谈节目。语料库中的每个语音文件大约 30 min 长, 每个文件包含 10 个以上的说话人, 语音段从 1 s 到 30 s 不等。语料库的录制环境贴近真实生活, 包含有背景噪声, 短时噪声和随机的背景音乐声。我们选用 5 段 30 min 长的语音数据, 分别标为 voa1、kean1、kean2、cctv1、cctv2, 作为实验的开发数据集, 用于训练算法参数。我们另选取 3 段 30 min 长的语音数据, 分别标为 voa2、kean3、cctv3, 作为实验的测试数据集。实验选取 12 维 MFCC 和 1 维能量作为语音数据的特征。本文的实验机器的 CPU 为奔 4-1.8 G, 内存为 512 M。

### 4.2 评价准则

说话人聚类的评价指标有两个: 类纯度和说话人纯度。类纯度衡量的是同一类是否都来自同一说话人的话语; 而说话人纯度体现的是同一说话人的语音数据被分为多少个类。显然, 当类纯度和说话人纯度均为 100% 时, 聚类算法效果最好。

假设测试集共有  $n$  个说话人, 被分为  $c$  类, 其中  $n_{ij}$  表示第  $i$  类中属于第  $j$  个说话人的语音段的数目, 而  $N_{ij}$  为相应的语音特征帧数。

类纯度和说话人纯度定义分别为:

$$\text{类纯度} = \sum_{i=1}^c \max_{j \in [1, n]} (N_{ij}) / \sum_{i=1}^c \sum_{j=1}^n N_{ij}, \quad (3)$$

$$\text{说话人纯度} = \sum_{j=1}^n \max_{i \in [1, c]} (N_{ij}) / \sum_{i=1}^c \sum_{j=1}^n N_{ij}. \quad (4)$$

## 4.3 实验结果

表 1 显示了不同聚类算法的聚类性能, 由数据可看出 DSC 算法说话人纯度与类纯度均表现最差, 由此可见仅依靠类内散布准则不能有效地进行在线聚类。通过定义合适  $\theta_0$  与  $\theta_1$ , HSC 算法比 DSC 算法有更好的表现。虽然 HSC 算法在某些开发数据集上比 HC 算法表现更好, 但 HC 算法比 HSC 算法更加稳健。当  $N > 2$  时, DTC 算法在绝大多数数据集上表现更好。

表 1 说话人聚类性能分析

Tab.1 Error analysis of speaker clustering

语音源	说话人纯度 (%)					
	HC	HSC	DSC	DTC(N=)		
				2	3	4
voa1	90.3	86.8	83.7	87.5	87.5	87.5
kean1	96.0	97.5	91.5	93.7	93.7	94.8
kean2	82.3	88.7	87.8	92.7	94.1	94.1
cctv1	82.9	73.6	66.4	72.5	78.9	80.6
cctv2	84.4	85.6	79.9	87.3	90.4	92.4
voa2	89.5	84.1	84.7	86.0	86.1	86.1
kean3	95.8	91.2	89.5	95.2	95.9	96.1
cctv3	83.4	72.1	69.4	73.5	78.2	80.1
avg	88.0	84.9	81.6	86.0	88.1	88.9
类纯度 (%)						
	HC	HSC	DSC	DTC(N=)		
				2	3	4
voa1	90.2	87.8	84.6	87.2	88.3	88.3
kean1	95.3	95.5	93.5	94.4	94.4	94.4
kean2	81.4	90.6	89.8	90.6	90.6	90.6
cctv1	85.0	71.7	69.9	75.5	80.7	82.9
cctv2	82.9	86.7	80.0	90.3	92.4	93.4
voa2	88.4	84.2	84.6	86.1	86.3	86.3
kean3	94.5	91.6	89.1	94.3	94.7	94.8
cctv3	82.0	72.5	69.4	70.0	75.3	77.8
avg	87.4	85.0	82.6	86.0	87.8	88.5

图 4 显示了不同算法的时间效率, 从图中可以看出, 随着语音段数的增加, HC 算法时间耗费成几何级增加, 而所有的在线聚类算法耗费成线性增加。

图 5 显示了 DTC 算法与  $N$  值之间的关系, 图中左侧纵轴表示说话人纯度与类纯度值, 右侧纵轴表示算法的时间耗费。由图可看出随着  $N$  的增加, 算法的时间耗费成线性增加, 但说话人纯度与类纯度的提高值却在递减。因此一般根据对

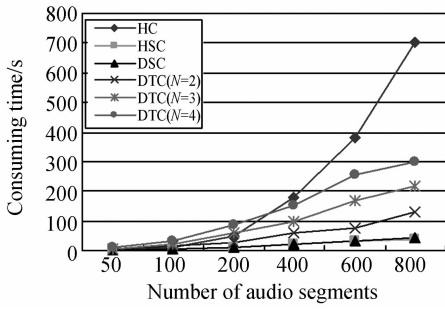


图 4 说话人聚类时间耗费图

Fig. 4 Time consuming for all online speaker clustering algorithms

说话人聚类的要求,  $N$  取值 2 至 4 之间。在对时间要求较高的场合  $N$  取值 2; 在对识别精度要求较高的场合,  $N$  可取值 3 或 4。

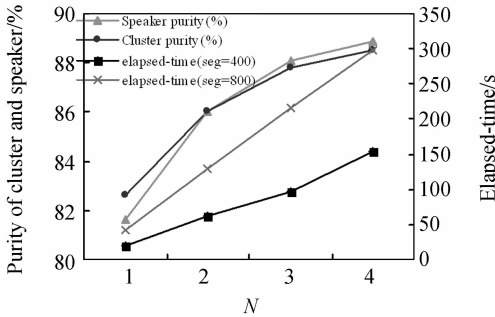


图 5 DTC 算法与  $N$  的关系图

Fig. 5 Relation between performance of DTC and value of  $N$

表 2 显示了 DTC 算法在大词汇量连续语音识别系统 (LVCSR, 1RT) 下的词错误率 (Word Error Rate, WER)。从实验结果可看出使用聚类结果进行说话人自适应的语音识别, 其结果要好于使用真实说话人类别信息进行说话人自适应的语音识别结果。

表 2 大词汇量连续语音识别结果表

Tab. 2 Experimental results on LVCSR

数据集	实验条件	词错误率 (%)	
cctv1	基线系统, 不聚类	29.7	
	真实说话人聚类	24.6	
	HC 聚类	24.1	
	HSC 聚类	26.0	
	DSC 聚类	27.1	
	DTC 聚类	$N=2$	25.6
		$N=3$	24.9
		$N=4$	24.1
	cctv2	基线系统, 不聚类	19.2
		真实说话人聚类	13.1
HC 聚类		12.9	
HSC 聚类		12.8	
DSC 聚类		13.0	
DTC 聚类		$N=2$	12.5
		$N=3$	12.2
		$N=4$	12.0
cctv3		基线系统, 不聚类	24.4
		真实说话人聚类	18.1
	HC 聚类	18.1	
	HSC 聚类	24.1	
	DSC 聚类	25.2	
	DTC 聚类	$N=2$	24.7
		$N=3$	21.7
		$N=4$	18.4

### 5 结 论

本文提出了一种改进的基于决策树的说话人在线聚类算法, 不同于传统的说话人在线聚类算法, 该算法通过决策树与决策队列对语音段进行聚类。通过对广播新闻语料说话人聚类实验的结果表明, 该算法在无需门限调整的情况下能够快速、自动地进行说话人聚类。其中, 平均类纯度和说话人纯度分别提高了 0.9% 和 1.1%, 时间减少了 57%, 保证了聚类结果的类纯度与说话人纯度, 并且有效地提高了大词表连续语音识别系统的性能。

### 参考文献:

[1] PADMANABHAN M, BAHL L, PICHENY M. Speaker clustering and transformation for speaker adaptation in large-vocabulary speech recognition systems [C]. ICASSP, 1996:701-704.

[2] 陈景东, 姚磊, 黄泰翼. 几种高鲁棒性通道及说话人自适应语音识别算法研究[J]. 声学学报, 1998,

23(6):537-544.

CHEN J D, YAO L, HUANG T Y. Study on robust channels and adaptive algorithms for speech recognition[J]. Acta Acustica, 1998, 23(6):537-544. (in Chinese)

[3] 吕萍, 颜永红. 基于回归分析的语音识别快速自适应算法[J]. 声学学报, 2005, 30(3):231-238.

LV P, YAN Y H. An fast adaptive algorithm in

- speech recognition based on regression analysis [J]. *Acta Acustica*, 2005,30(3):231-238. (in Chinese)
- [4] CHEN S, GOPALAKRISHNAN P. Speaker, environment and channel change detection and clustering via the bayesian information criterion [C]. *Proc. of BNTUW-98, Lansdowne*, 1998:127-132.
- [5] JIN H, KUBALA F, SCHWARTZ R. Automatic speaker clustering [C]. *Proc. of DARPA. Speech Recognition Workshop*, 1997:108-111.
- [7] KUBALA F, COLBATH S, LIU D, *et al.*. Integrated technologies for indexing spoken language [J]. *Comm. ACM*, 2000,43(2):48-56.
- [8] 王炜, 吕萍, 颜永红. 一种改进的基于层次聚类的说话人自动聚类算法 [J]. *声学学报*, 2008, 33(1): 9-14.  
WANG W, LÜ P, YAN Y H. An improved hierarchical speaker clustering [J]. *Acta Acustica*, 2008, 33(1):9-14. (in Chinese)
- [9] LIU D, KUBALA F. Online speaker clustering [C]. *ICASSP*, 2004:333-336.
- [10] WANG W, LV P, ZHAO QW, *et al.*. A decision-tree-based online speaker clustering [C]. *LNCS*, 2007: 555-562.
- [11] DUDA R, HART P, STORK D. *Pattern Classification* [M]. 2nd ed. John Wiley & Sons, Inc., 2001.
- [12] 肖述才, 欧智坚, 王作英. 语音识别中的一种说话人聚类算法 [J]. *中文信息学报*, 2005, 19(4):84-88.  
XIAO SH C, OU ZH J, WANG Z Y. A speaker clustering algorithm in speech recognition [J]. *Journal of Chinese Information Processing*, 2005, 19(4):84-88. (in Chinese)
- [13] GISH H, SIU M, ROHICEK R. Segregation of speaker for speech recognition and speaker identification [C]. *ICASSP*, 1991:873-876.

## 作者简介:



张素敏(1970—),女,河北藁城人,博士研究生,高级工程师,1992年于北京大学获得理学学士学位,2004年于西安电子科技大学获得硕士学位,主要研究方向为通信对抗技术、抗干扰理论与应用等。E-mail: zhangsm0202@sina.com



王炜(1980—),男,江西景德镇人,博士,工程师,2001年、2004年于解放军信息工程大学分别获得工学学士和硕士学位,2007年于中科院声学所获得博士学位,主要从事信号与信息处理及信息安全方面的研究。E-mail: wongwere@hotmail.com

## 导师简介:



苏东林(1960—),女,北京人,博士,教授,博士生导师,1983年、1986年、1999年分别于北京航空航天大学获得学士、硕士和博士学位,主要研究方向为电磁兼容理论与应用、抗干扰理论与应用等。E-mail: sdl@buaa.edu.cn