第五届全国人机语音通讯学术会议(NCMMSC'98)，134-137，98 年 7 月 26~31 日，哈尔滨

基于模型等价类的快速识别算法

武 健　郑 方　吴文虎　方棣棠
(清华大学计算机科学与技术系　100084)
[jwu,fzheng]@sp.cs.tsinghua.edu.cn

摘要：本文介绍了一种应用在非特定人大词汇表连续语音识别中的快速算法。该算法充分利用了部分识别基元之间较强的相似性，通过定义识别基元间的相似测度，并依据近似基元相似测度矢量的相关性，将识别基元分为若干个等价类，然后根据分组结果给出模型匹配概率近似值的计算方法，从而达到快速识别的目的。该方法经实验证明，在同等计算复杂度的前提下，与常规算法相比，对以音节为识别基元的汉语语音识别具有较好的效果，降低了计算的复杂度，有很高的实用价值，同时也可以推广到其它基元的识别系统中去.

关键词：相似测度　识别基元　中心距离连续概率模型(CDCPM) 模型等价类(EMC) 最小分类错误的声学模型间距离度量

一、前言

目前在已实现的各种非特定人大词汇表连续语音识别系统中，HMM 或其修正模型占据了很重要的地位，在这些不同的系统中，识别的本质实际上都是语音信号序列 s 与 N 个基元模型 $M_i\ (i=1,\cdots,N)$ 相匹配的过程，也即计算出概率 $P(s|M_i)$ 并选取出较大的若干个作为我们的识别候选以供后续处理过程使用。

但是，一般来说，这种常规计算方法存在两个问题影响着识别的效率：一是计算复杂度大，经典的 HMM 中计算 $P(s|M_i)$ 需经过时空复杂度较高的 Viterbi 解码；二是计算冗余度大，许多概率或匹配得分计算出来之后尚没有得到利用即失去了存在意义。

如果能够较好地解决这两个问题，就可以使计算复杂度在识别效果下降并不太明显的前提下，得到大大的降低，从而有利于系统的实时性，并提供通过进一步采用高阶模型来提高系统精度的可能。我们这里将要提出的这种利用识别基元相似测度的快速识别算法实际上就是从尽量避免对所有识别基元计算匹配概率的角度来考虑计算冗余度的降低。

二、声学模型间的相似测度

令 $M_i(i=1,\cdots,N)$ 为待选择的 N 个基元模型。这些基元的发音具有一定程度的相似性，导致了统计模型参数上的相似性。为了刻划这些基元之间的相似性，我们可以定义出声学模型间的距离。Juang 和 Rabiner 提出的基于 HMM 的声学模型间概率测定的相似测度 [1]虽然利用了信息论，较好地描述出两个隐马尔可夫模型之间的关系，但由于要对大量数据进行统计，计算起来相当复杂，为此我们对这个相似测度距离作了一定的改进。

我们的识别系统并不是基于传统的 CHMM，而是 CHMM 的一个变形，即中心距离连续概率模型 CDCPM（Center-Distance Continuous Probabilistic Model）[2] 。该模型与经典 CHMM 相比，去掉了状态转移矩阵 A，且输出观察概率矩阵 B 中各状态的概率密度函数简化成了一个一维的满足中心距离分布的概率密度函数。采用这种模型，可以极大地减小模型存储的空间复杂度和模型训练及计算的时间复杂度。

我们在 CDCPM 的基础上提出了一种基于最小分类错误的声学模型间距离度量[3]。在这个距离空间度量下，模型 M_i 和 M_j 之间的距离按公式 1 进行计算，其中 $D(M_i,M_j)$ 是由 M_i 和 M_j 的分布参数估计出的误识概率。这种距离度量充分利用了 CDCPM 模型的结构特性，同时易于计算，避免了对所有测试数据的识别结果进行统计，通过对模型参数的简单计算即可获得 CDCPM 模型间相互距离的近似值，而且从实际使用效果上来看，它也的确较好地刻划了汉语声学模型之间的相似度。

$$d\left(M_i, M_j\right) \overset{def}{=} 1 - P\left(M_j\text{的样本被误识为}M_i\right) \approx 1 - D\left(M_i, M_j\right) \qquad (1)$$

三、声学模型间相似测度的利用

根据上述的过程，我们可以得到一个 $N \times N$ 的模型间距离矩阵 $D = \left\{d_{ij}\right\}$，其中 $d_{ij} = d\left(M_i, M_j\right)$ 是模型 M_i 和 M_j 之间的相似测度。需要注意的是，由于我们的距离定义是非对称的，因而矩阵 D 也是一个非对称的。但观察该矩阵可以发现，比较相似的基元模型之间具有很强的趋同性，即若两个识别基元模型 M_i 和 M_j 之间的距离 d_{ij} 与 d_{ji} 都较小，则行矢量 $\vec{d}_{i\bullet}$ 与 $\vec{d}_{j\bullet}$ 之间有较大的相关度，这是由两者之间的相似决定的。于是我们可以在矩阵 D 的基础上进一步计算出一个对称矩阵 $E = \left\{e_{ij}\right\}$，其中

$$e_{ij} = 1 - \rho\left(\vec{d}_{i\bullet}, \vec{d}_{j\bullet}\right) \qquad (2)$$

$\rho\left(\vec{d}_{i\bullet}, \vec{d}_{j\bullet}\right)$ 是矢量 $\vec{d}_{i\bullet}$ 与 $\vec{d}_{j\bullet}$ 之间的相关系数，因而当 $\vec{d}_{i\bullet}$ 和 $\vec{d}_{j\bullet}$ 比较相似的时候，e_{ij} 趋近于 0。为了使 e_{ij} 的值更能代表两模型间的关系，在计算时可以先对两个行矢量作归一化处理。

由于 E 矩阵的生成既考虑了两个基元模型间的相互关系，同时也考虑了这两个基元模型与其他基元之间相似测度的一致程度，因而更能准确地反映出基元模型的相似性。从这个对称矩阵我们可以很容易地利用简单聚类方法形成 K 个模型等价类（Equivalent Model Class, EMC） $\mathrm{MC_i}\ (i = 1, \cdots, K)$。每个模型等价类 $\mathrm{MC_i}$ 包括 T_i 个相互间极易混淆的模型 $\mathrm{MC_{ij}}\ (j = 1, \cdots, T_i)$，而且满足：

$$MC_i \cap MC_j = \varnothing \qquad \forall\, i \neq j \qquad (3)$$

$$\bigcup_{i=1}^{K} MC_i = \left\{M_j \mid j = 1, \cdots, N\right\} \qquad (4)$$

根据全概率公式，我们可以通过以下变形来计算概率 $P\left(s \mid M_i\right)$。

$$P\left(s \mid M_i\right) = \sum_{j=1}^{K} P\left(s \mid MC_j, M_i\right) \cdot P\left(MC_j \mid M_i\right) \qquad (5)$$

不妨设 $M_i \in MC_k\ (1 \le k \le K)$，则从前面产生模型等价类 MC_i 的过程可以认定概率 $P\left(MC_j \mid M_i\right)\ (\forall j \neq k)$ 非常小（因为 M_i 与 $MC_j\ (j \neq k)$ 中的基元模型相似程度较小，否则的话，M_i 就不会归入 MC_k，而应归入 MC_j 了），我们可以近似认为

$$P\left(MC_j \mid M_i\right) = \delta(j, k) \qquad (6)$$

综合 (5) (6)，可得 $$P\left(s \mid M_i\right) \approx P\left(s \mid MC_k, M_i\right) \qquad (7)$$

为了进一步计算出概率 $P\left(s \mid M_i\right)$，对每一个模型等价类 MC_i 取一个代表模型 $\overline{MC_i}$，则 (7) 式又可以变化为 $$P\left(s \mid M_i\right) \approx P\left(s \mid \overline{MC_k}\right) \qquad (8)$$

现在我们可以很容易地通过少量的计算得出所有匹配概率的近似值，通过这些置信度较高的近似值，能比较准确地排除掉许多完全没有可能进入候选的模型等价类，然后再在保留下来的若干个候选等价类中，通过进一步计算更精确的匹配概率，选择出所需的候选。

四、识别基元的选择和复杂度分析

汉语的识别基元可以是音素、音节以及介于两者之间的半音节（或声韵），基元越小，模型个数就越少，但训练数据库的标定越困难，基元之间连接的不确定性越大，基元之间的相似程度越小，我们这种快速识别方法的实用效果也就越差，反之亦然。另一方面，从

我们以前实验的结果来看[4]，以汉语的音节来做识别基元是比较合适的选择。

采用这种快速方法能够很大程度上降低复杂度。以完全计算一次匹配概率 $P(s|M_i)$ 所需要的时间空间复杂度为单位 T，传统方法的计算复杂度为 $O_1 = NT$ (9)

快速方法的计算复杂度计算如下：

设 N 个基元模型被分为 K 个等价类，每个等价类平均拥有模型 $\left\lceil \frac{N}{K} \right\rceil$ 个，则第一次粗选须进行计算 $O_{21} = KT$ (10)

然后从中选出 C 个等价类用于第二次的进一步计算，共须进行计算

$$O_{22} = C \cdot \left\lceil \frac{N}{K} \right\rceil \cdot T - CT \quad (11)$$

于是，$$O_2 = O_{21} + O_{22} \approx \left(\frac{C}{K} + \frac{K}{N} - \frac{C}{N} \right) \cdot NT \quad (12)$$

若 $N \approx 8K$，$K \approx 5C$ 则 $O_2 \approx \frac{3}{10} NT = \frac{3}{10} O_1$ (13)

即减少了 2/3 以上的计算量。

五、实验结果及分析

实验所使用的数据库来自于 863-306 课题资助建立的汉语普通话连续语音数据库，包含有 16～45 岁年龄范围内的 76 人的语音信息，其中男女各半；每人完成一组 520 个句子的发音，语料由社会科学院语言所设计，共含有 396 个音节，几乎涵盖所有汉语音节；语速为 3.5 个音节/秒至 6 个音节/秒，以 5 个音节/秒为主；16KHz 采样，16bit 量化。在这个数据库的基础上，我们选取了其中的一部分（共计 70592 个音节）作为测试集，剩下的作为训练集（共计 180065 个音节），足够大的数据量说明了实验结果的可靠性。

实验所使用的模型是基于 NN（Nearest Neighbor）方法的 CDCPM（状态数 N＝6，混和数 M＝16）；特征参数为 16 维 LPC 倒谱系数与其回归系数 ARCEP 的组合；窗长与帧移分别为 32ms 和 16ms。

1、对测试集的识别结果比较（包括相同复杂度条件下两种方法的比较）

EMC（Equivalent Model Class）表示快速计算方法，Normal 表示常规计算方法，K 表示模型等价类数，C 表示第一次粗选保留的等价类数，N 表示 CDCPM 模型的状态数，M 表示 CDCPM 模型的混和数，除特殊指出外，N＝6，M＝16。结论：i) 对于测试集而言，采用同等规模模型的两种计算方法虽有差距，但差别不大，在计算复杂度大大降低的条件下仍具有很强的实用性； ii) 此外，基于较复杂的 CDCPM 模型（M＝16）的快速算法与采用简单模型（M＝8）的常规算法相比，两者计算复杂度相近，而前者的效果明显优于后者，充分说明了前者的优势。由此可见，如果提高快速算法的模型精度，使快速算法与 M＝16 的复杂模型常规算法的计算复杂度相当，那么快速算法的识别性能将大大高于现有的结果.

2. 模型等价类个数的影响：

设定不同的阈值，可以从 E 矩阵获得不同规模的模型等价类。结论：在增加了 1/8 数量的等价类的情况下，对测试集的识别效果并未得到较大的改进，说明目前我们所取的规模比较合适，同一等价类内各个模型之间的相似度较大，无须将其分解成更多的等价类。

3. 粗选时保留候选等价类个数的影响：

在第一步获得每个模型匹配概率的近似值之后，需要对这些等价类进行取舍，取舍的程度不同造成识别效果的差别。结论：在增加了一定候选数目之后，效果有所改进，与常规方法的差别进一步减少，而且计算复杂度并未显著提高。以目前的实验结果来看，$N \approx 8K$，$K \approx 5C$ 是一个较好的比例关系。

表 1 两种计算方法对测试集的识别结果比较

前n名	1	2	3	4	5	6	7	8	9	10
EMC(*K*=56,*C*=10)	0.49	0.65	0.72	0.76	0.79	0.81	0.82	0.83	0.84	0.84
EMC(*K*=63,*C*=10)	0.49	0.65	0.73	0.77	0.79	0.81	0.83	0.84	0.84	0.85
EMC(*K*=56,*C*=15)	0.50	0.67	0.75	0.79	0.82	0.84	0.86	0.87	0.87	0.88
Normal(*M*=8)	0.40	0.56	0.64	0.70	0.74	0.77	0.79	0.81	0.83	0.84
Normal(*M*=16)	0.52	0.70	0.78	0.83	0.86	0.88	0.90	0.91	0.92	0.93

六、结论

本文提出的这种快速计算方法经实验证明有着很强的实用性，它较好地利用了汉语音节间的普遍相似性，实际上是利用了高阶的统计信息，减少了大量无用的计算。

这个方法的应用效果好坏与两个因素紧密相关：一是距离矩阵的估计，即距离度量的定义需要准确反映出识别基元之间的关系；二是代表模型的选取，由于代表模型是计算匹配概率近似值的依据，一定要能够切实起到代表的作用，否则会发生很大程度的误识。

此外，得到的等价类还可以看作是一个新分类器，利用每个模型与同族模型间的关系，我们能够构造出新的拒识方法，以提高识别率，这将是以后值得继续研究的工作之一。

参考文献

1. **Juang, BH. and Rabiner, LR.,** "A probabilistic distance measure for hidden Markov Models", *AT&T Technical Journal*, Feb., 1985, pp.391-408
2. **郑方，吴文虎，方棣棠，**"CDCPM 及其在语音识别中的应用"，*软件学报*，863 专刊，Oct., 1996, pp.69-75
3. **郑方，武健，吴文虎，方棣棠，**"基于最小分类错误的声学模型间距离度量"，第三届全国计算机智能接口和应用学术会议，Aug., 1997
4. **郑方，吴文虎，方棣棠，**"汉语语音听写机中的语音识别基元"，第四届全国人机语音通讯学术会议，Oct., 1996, pp. 32-35