第三届全国人机语音通讯会议论文集*(NCMMSC-94)*，pp.238-241，1994年10月，重庆

连续距离密度的分段概率模型ＣＤＤ－ＳＰＭ

郑　方　杨红波　吴文虎　方棣棠
(清华大学计算机系　100084)

摘　　要

本文提出了一种新的非特定人语音识别方法　——　连续距离密度的分段概率模型(CDD-SPM)。使用这种模型，其计算量大大低于连续隐式马尔可夫模型(CHMM)，而性能与之相差无几；这种模型避免了分段概率模型(SPM)中的矢量量化所带来的量化误差。这是一种介乎连续隐马尔可夫模型(CHMM)和离散的隐马尔可夫模型(DHMM)之间的模型。

关键词　　连续距离密度，分段概率模型

Continuous Distance Density Segmental Probablistic Model

Zheng Fang, Yang Hongbo, Wu Wenhu, Fang Ditang
(Tsinghua University)

Abstract

A new model for speaker-independent speech recognition is proposed in this paper, which is called Continuous Distance Density Segmental Probabilistic Model (CDD-SPM). The computation of this kind of model is much less than that of the continuous hidden Markov model, yet the performance is almost the same. The new model can eliminate the VQ distortion that is caused in the SPM. So this is a model whose performance and computation are between continuous HMM and discrete HMM.

Key words　continuous distance density, segmental probabilistic model

一. 问题的提出

ＨＭＭ模型在语音识别中占有相当重要的地位，取得了很好的效果。目前，许多语音识别研究者又在ＨＭＭ模型的基础上提出了改进的模型，其中包括连续密度分布的ＨＭＭ模型(CHMM)[1]、半连续的ＨＭＭ模型(SCHMM)[2]以及分段概率模型(SPM)[3]。

连续密度分布的ＨＭＭ模型假定一段语音的特征向量是按某一种概率密度分布的，训练的任务是用训练集中的大量样本对预先假定的概率分布进行参数估计；识别时按照估计的参数对待识样本进行概率计算，将具有最大概率的模型作为识别结果。这种模型性能的好坏取决于假定的概率分布是否符合实际情况。一般地讲，一些常用的概率分布如正态分布并不能

精确描述其分布情况。于是有的研究者用几个中心不同、离散度不同的正态分布的组合来逼近实际的特征向量的分布[4]。

连续密度分布的ＨＭＭ模型出于计算量的考虑和模型的简化，往往把协方差矩阵假定为对角阵，这无疑会降低性能。于是[5]对全协方差矩阵的模型进行了尝试。

ＳＰＭ模型对ＨＭＭ模型进行了很大的简化，在良好的时域分段前提下，ＳＰＭ模型不需要初始概率矩阵和转移概率矩阵，而只要概率输出矩阵。实验结果表明，这种模型对中字表非特定人语音识别颇为有效。

但是不能否认：ＳＰＭ模型需要矢量量化，这必定会导致一些误差；而连续密度的ＨＭＭ模型则会因为大量的计算而影响识别速度。

本文提出一种基于ＳＰＭ模型的新模型，但不必进行矢量量化；它也具有连续密度分布的特征，却没有那么大的计算量。实验表明，它不失为一种较好的实用模型。这就是连续距离密度的分段概率模型ＣＤＤ－ＳＰＭ。

二. 基本思路

假定随机变量ξ符合正态分布Ｎ（０，σ），其概率密度公式为：

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-x^2/2\sigma^2} \tag{1}$$

另有随机变量η＝｜ξ｜，我们把η所服从的概率分布称为ＡＮ（σ），则η的概率密度公式为：

$$g(y)=\begin{cases}0, y<0\\ \dfrac{2}{\sqrt{2\pi}\sigma}e^{-y^2/2\sigma^2}, y\geq 0\end{cases} \tag{2}$$

η的数学期望为：

$$E[\eta]=\int_{\infty}^{0} yg(y)dy=\frac{2\sigma}{\sqrt{2\pi}} \tag{3}$$

ＣＤＤ－ＳＰＭ模型假定某一段语音的所有特征向量离开特征向量中心Ｃ的欧氏距离符合ＡＮ（σ），公式(3)中的ｙ就是特征向量偏离矢量中心Ｃ的欧氏距离。有了这样的假设，训练的任务就是计算某段语音的ＣＤＤ－ＳＰＭ模型中的两个参数：矢量中心Ｃ，以及概率密度参数σ。矢量中心Ｃ和参数σ就可以用来表征这段语音。识别算法比较简单：计算出待识语音段的每个特征向量离开中心Ｃ的距离，进而估计每个向量的出现概率以及整个段的概率。某个词所有段综合起来的最大概率者为识别结果。

三．ＣＤＤ－ＳＰＭ模型的训练算法

1. 对语音库中Ｎ个词的Ｍ个语音样本按帧计算出特征向量（Ｍ＞＞Ｎ）。
2. 从ｎ＝１到Ｎ，对每个词Ｗn的模型进行训练。
   2.1 把属于词Ｗｎ的所有Ｑ个样本中的每一个样本，都非线性分块为Ｊ段；
   2.2 从ｊ＝１到Ｊ按下面方法计算每一段的参数：
      2.2.1 从ｑ＝１到Ｑ统计第ｊ段的矢量中心Ｃnj；
      2.2.2 从ｑ＝１到Ｑ计算第ｊ段中每个矢量偏离Ｃnj的距离Ｄnjq的均值Ｄnj；
      2.2.3 使用公式⑶计算参数σnj。

用上面的步骤得到的Ｃnj和σnj是第ｎ个词Ｗn第j段的模型参数。如令λn为词Ｗn的模型，则：

λn ＝ （Ｃn，σn）

其中

Ｃn ＝ ｛Ｃn1，Ｃn1…ＣnJ｝，σn ＝ ｛σn1，σn1…σnJ｝

四．ＣＤＤ－ＳＰＭ模型的识别算法

1. 计算待识语音Ｗx的特征向量，将其非线性分块为Ｊ段，设每段的特征向量数为mj。
2. 从ｎ＝１到Ｎ计算Ｗx与每个词模型λn＝（Ｃn，σn）匹配概率：
   2.1 从ｊ＝１到Ｊ：
      2.1.1 计算待识语音第ｊ段mj个特征向量偏离Ｃnj的距离Ｄnjk，（１≤ｋ≤mj）
      2.1.2 利用公式⑵计算mj个特征向量出现在λn模型中的概率ｇ（Ｄnjk）。
      2.1.3 用下面的公式计算第ｊ段mj个特征向量出现的总概率：

$$P_{nj}=\prod_{k=1}^{m_j} g(D_{njk}) \tag{4}$$

   2.2 用下面公式计算待识语音与模型λn的匹配概率：

$$P_n=\prod_{j=1}^{J} P_{nj} \tag{5}$$

3. ｎ＝１到Ｎ检查Ｐn，将最大者的模型所对应的词做为识别结果：

$$x=\arg\max_n P_n \tag{6}$$

五．实验结果以及实际的考虑

对上述算法进行实验验证，结果是令人满意的。所用的训练集是20个人的□208个词的发音。识别集是另外20个人的208个词的发音。统计的结果是：训练集识别率为95.0%，识别集

识别率为92.0%。该识别率比全矩阵连续HMM模型的要低4个百分点左右，但识别时间却节省十倍左右。

在实际的计算中，小于 1 的数的连乘会导致溢出，而且指数的计算也比较费时间。解决的方法是把概率的相乘转换为概率对数的相加。

另外一个问题是，实际的语音并不是按我们假定的那样服从正态分布。[4]中采用混合密度的连续ＨＭＭ模型，即用几个正态分布去逼近实际的概率分布。在距离模型中，这个方法仍然可以采用。假定每个语音段都是由几个（个数可以不等）正态分布来逼近的，比如说是Ｎ个。用诸如聚类这样的方法在训练时得到这Ｎ个正态分布的矢量中心，然后再估计每个密度的σ参数。在识别时用几个概率密度函数的迭加作为其概率值。

参考文献

[1] L.R. Bahl, P.F. Brown, P.V. de Souza & R.L. Mercer
Speech Recognition with Continuous-parameter Hidden Markov Models,
Readings in Speech Recognition, 332-339, edited by Alex Waibel & Kai-Fu Lee, 1990

[2] X.D. Huang & M.A. Jack
Semi-continuous Hidden Markov Models for Speech Signals,
Computer Speech and Language (1989), 3:239-251

[3] Jiang Li, Wu Wenhu, Cai Lianhong, Fang Ditang
A Read-time Speaker-independent Speech Recognition System Based on SPM
for 208 Chinese Words, Proceedings of ICSP'90, 473-476

[4a]L.R. Rabiner, B.G. Juang, S.E. Levinson & M.M. Sondhi
Recognition of isolated digits using hidden Markov models with continuous
mixture densities, AT&T Technical Journal, 64, 1211-1234, 1985

[4b]B.H. Juang & L.R. Rabiner
Mixture autoregressive hidden Markov models for speech signals, IEEE Transactions
on Acoustic, Speech and Signal Processing, ASSP-33, 1404-1413, 1985

[5] 方棣棠、余　华、李树青
基于正态分布假设的非特定人语音识别，
中文电脑国际会议’94（新加坡），1994年6月