

一种具有强分类能力的离散 HMM 训练算法*

方绍武, 戴蓓倩, 李霄寒

(中国科学技术大学 电子科学与技术系, 安徽 合肥 230026)

E-mail: sw_fang@263.net

http://www.usc.edu.cn

摘要: 提出了一种具有强分类能力的离散 HMM(hidden Markov models)训练算法. 该算法利用矢量量化技术将来自不同话者的训练数据进行混合训练, 以生成包含各个话者特征的话者特征图案. 用该特征图案代替经典的离散 HMM 中的 VQ 码本, 可以提高观察值符号序列的模式辨识能力, 从而提高了离散 HMM 的分类能力. 给出了该方法用于文本有关的话者识别的实验结果, 表明该算法可提高系统的识别性能, 并可降低 HMM 对训练集大小的依赖程度, 且识别时计算量明显小于经典 HMM 训练算法, 具有较大的实用价值.

关键词: 离散 HMM(hidden Markov models); 分类能力; 特征图案; 矢量量化; 鲁棒性

中图分类号: TP391 文献标识码: A

日前, 隐 Markov 模型(hidden Markov models, 简称 HMM)技术是语音信号处理领域中的主流技术, 离散 HMM(DHMM)在训练及识别时计算量较小, 对输入矢量的分布不必作先验假设, 因而在对实时性要求较高的场合得到了较多的应用.

众所周知, HMM 处理语音信号中所包含的统计信息和时序信息的能力非常强. 但是, 由于训练准则和算法的限制, 它对模式的辨识能力较差^[1]. 解决这一问题的方法通常是将不同的信号处理技术结合起来, 相互取长补短. 一种有代表性的方法就是混合 HMM/MLP 方法^[2]. 与 HMM 相比, 神经网络, 特别是 MLP(multi-layer perception), 由于训练算法的特点(基于模式辨别的训练方式)而具有极强的模式辨识分类能力^[3]. 但是, 混合方法的缺点也是明显的, 即增加了算法复杂度及计算量, 且在一般情况下也无助于减小 HMM 对训练数据集的依赖程度. 为此, 本文提出了一种具有强分类能力的离散 HMM 训练算法(后文简称为改进的离散 HMM 训练算法). 该算法在矢量量化这一级采用混合训练方式, 从所有话者的混合训练数据集中提取含有各个话者特征的话者特征图案. 用该特征图案代替经典的离散 HMM 中的 VQ 码本, 可以提高量化后的观察值符号序列的分类能力, 从而相应地提高改进后的离散 HMM 的模式辨识能力. 该算法的性能在与文本有关的话者识别系统中得到了实验验证.

1 基于离散 HMM 的话者识别系统

图 1 为基于离散 HMM 的话者识别系统的原理框图. 图中, 特征参数矢量提取的作用是从说话人的训练语音文本中提取若干帧的话者特征参数, 每一帧参数对应于矢量空间中的一个矢量. 这样, 从一个语音文本中提取的是一个矢量序列. 一般来说, 训练一个话者的离散 HMM 话者模型分

* 收稿日期: 1999-12-28; 修改日期: 2000-05-18

基金项目: 国家自然科学基金资助项目(69872036)

作者简介: 方绍武(1965-), 男, 安徽枞阳人, 博士, 讲师, 主要研究领域为语音信号与信息处理, 模式识别; 戴蓓倩(1941-), 女, 浙江宁波人, 教授, 博士生导师, 主要研究领域为语音信号与信息处理; 李霄寒(1979-), 男, 山东临清人, 博士生, 主要研究领域为语音信号处理.

两步进行:首先用 LBG (linde-buzo-gray) 算法对该话者的训练集中所有训练矢量(帧)训练出 VQ 码本;再用该 VQ 码本将训练集中的各个矢量序列量化为观察值符号序列,最后用 Baum-Welch 算法训练出离散 HMM. 当系统中所有话者的 VQ 码本及离散 HMM(记为 λ) 都训练完毕并存储起来之后,话者识别系统的训练任务就完成了. 识别时,提取待测试语音的特征参数矢量序列,用第 i 个话者的码本量化为观察值符号序列,再用第 i 个话者的模型 λ_i 来计算 Viterbi 评分 $P(O|\lambda_i)$,具有最大 Viterbi 评分的模型 λ_i 所对应的话者身份即为要识别的结果.

2 具有强分类能力的离散 HMM 训练算法

在上述经典的离散 HMM 训练过程中,VQ 的作用仅仅是作为前端处理,将连续的输入参数矢量离散化为观察值符号(VQ 码本标号). 另外,各个话者的 VQ 码本训练是独立进行的,没有利用到其他话者的训练数据. 为了提高离散 HMM 的分类能力,我们利用 VQ 技术对所有话者的训练数据进行混合训练,生成了含有各个话者特征的模式,当以此特征图案代替图 1 中的 VQ 码本时,通过 VQ 产生的对应某个话者的观察符号将只是该特征图案的一个子集,而对应不同话者的各个子集之间所含观察符号及观察符号的顺序均不尽相同,从而增强了观察符号序列中含有的分类信息,并使训练出的离散 HMM 具有较强的模式分类能力. 图 2 为具有强分类能力的离散 HMM 算法框图.

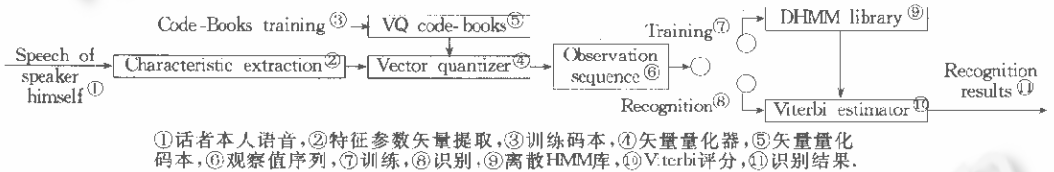


Fig. 1 Speaker recognition system based on discrete HMM
图1 基于离散HMM话者模型的话者识别系统框图

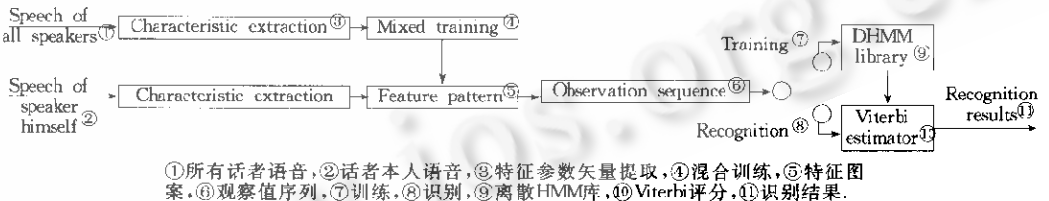


Fig. 2 A DHMM training algorithm with strong classifying ability
图2 一种具有强分类能力的离散HMM训练算法框图

3 实验结果及讨论

为了验证本文提出的算法的有效性,我们进行了基于短语音的、与文本有关的话者识别实验. 短语音文本为“开门”孤立词(约 0.6 秒的发音),训练及识别用的数据来自 14 名男生,每人 100 遍语音.

话者特征参数采用 20 阶的 MFCC (mel-frequency cepstral coefficient) 参数,即 Mel 域倒谱系数. 对应于一个短时帧的 20 阶 MFCC 参数组成 20 维矢量空间的一个特征矢量.

在连续 HMM 及离散 HMM 中,Markov 链均为左-右无跨越的模型,状态数均取为 5;离散 HMM 中 VQ 的码本尺寸均为 256;连续 HMM 中高斯混合概率密度函数的混合数为 3.

表 1 为改进的离散 HMM 训练算法与经典的离散 HMM 的识别性能比较实验结果. 表 2 给出

了改进的离散 HMM 训练算法与连续 HMM(CDHMM)的识别性能比较. 将表 1、表 2 中的数据拟合为曲线, 如图 3 所示.

Table 1 The comparison of recognition performance for improved DHMM training algorithm and conventional DHMM training algorithm

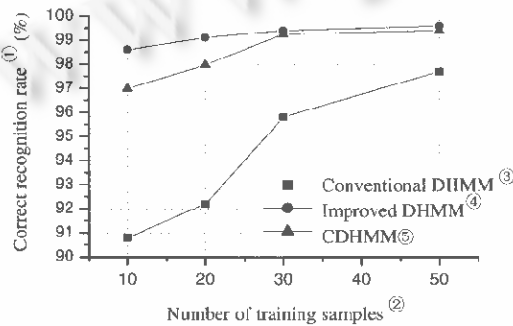
Number of training samples ^①	10	20	30	50
Conventional DHMM ^②	90.8	92.2	95.8	97.7
Improved DHMM ^③	98.57	99.07	99.43	99.57

①训练样本数, ②经典的离散 HMM, ③改进的离散 HMM.

Table 2 The comparison of recognition performance for improved DHMM training algorithm and CDHMM training algorithm

Number of training samples ^①	10	20	30	50
CHMM ^②	96.98	97.97	99.25	99.40
Improved DHMM ^③	98.57	99.07	99.43	99.57

①训练样本数, ②连续 HMM, ③改进的离散 HMM.



①正确识别率, ②训练样本数, ③经典的离散 HMM, ④改进的离散 HMM, ⑤连续 HMM.

Fig. 3 The recognition performance comparison of conventional DHMM, improved DHMM and CDHMM

图 3 经典的离散 HMM, 改进的离散 HMM 及连续 IIMM 的识别性能比较

之用.

在计算复杂度方面, 改进的离散 HMM 训练算法也优于经典的离散 HMM 训练算法, 特别是当待识别的话者较多时. 在进行话者识别时, 如果参与辨识的话者有 N 个, 则经典的离散 HMM 算法需将测试语音分别与 N 个话者的 VQ 码本进行量化, 并分别求出对应 N 个话者的 Viterbi 评分, 而改进的离散 HMM 算法只需将测试语音对共用的话者特征图案量化一次, 再将得到的观察值序列对各个话者的模型求 Viterbi 评分. 可见, 改进的离散 HMM 算法在矢量量化这一阶段的计算量只有经典的离散 HMM 算法的 $1/N$. 为了减小量化误差对离散 HMM 性能的影响, 通常 VQ 码本的尺寸都取得较大, 一般为 $M=256$, 当码本为全搜索码本时, VQ 的计算时间占据了 Viterbi 评分计算的大部分时间. 14 个人的话者识别系统的实验表明, 改进的离散 HMM 算法较之经典的离散 IIMM 算法节省识别时间约 80% 左右. 当话者识别系统的人数较多时 (如 100 人以上), 则改进的离散 HMM 算法对计算量的节省就更明显. 可见, 改进的离散 HMM 算法可以比经典的离散 HMM 算法更进一步地节省计算量, 因而更能满足话者识别系统的实时性要求.

4 结束语

本文提出了具有强分类能力的离散HMM训练算法,较大地提高了离散HMM的模式识别能力,并在很大程度上减小了HMM的性能对较大训练集的依赖,具有较大的实用价值.在14人的文本有关的话者识别实验中,其识别性能明显优于经典的离散HMM,同时也略优于连续HMM,特别是当训练样本较少时,该算法对系统性能的提高更为明显.

该算法适用于话者人数固定的话者识别系统.对于话者人数不固定的系统,则在建立系统时,应事先收集足够多的具有代表性的说话人语音参与话者特征图案的训练,这样,新用户的特征就在已训练好的话者特征图案中得到足够的体现.否则,应利用新用户的数据对话者特征图案进行自适应修正.

References:

- [1] Bourlard, H., Wellekens, C. J. Links between Markov models and multi-layer perceptron. *IEEE Transactions on PAMI*, 1990, 12(12): 1167~1178.
- [2] Cerf, P. L., Maa, W., Compernelle, D. V. Multilayer perceptrons as labelers for hidden markov models. *IEEE Transactions on SAP*, 1994, 2(1): 185~193.
- [3] Visarut, Ahkputra, Somchai, Jitapunkul. A comparison of Thai speech recognition systems using hidden Markov model, neural network, and fuzzy-neural network. In: *Proceedings of the ICSLP'98*. Sydney, Australia, 1998. 283~287.

An Algorithm with Strong Classifying Ability for Discrete HMM Training*

FANG Shao-wu, DAI Bei-qian, LI Xiao-han

(Department of Electronics Science and Technology, University of Science and Technology of China, Hefei 230026, China)

E-mail: sw_fang@263.net

http://www.ustc.edu.cn

Abstract: A discrete-HMM training algorithm which has strong ability of pattern classification is presented in this paper. By VQ (vector quantization) technique, this algorithm trains data from all speakers in mixed mode to generate the speaker characteristic pattern, which includes features of all speakers. By substituting the VQ codebook in conventional discrete HMM with characteristic pattern, the ability of pattern classification for observation symbol sequence is enhanced, therefore the classifying ability of discrete-HMM is improved. The experimental results show that the algorithm can improve the system's recognition performance, and reduce the dependence extent of HMM on the scale of training set. Moreover, the calculation quantum of this algorithm in recognition stage is obviously less than that of conventional HMM training algorithm, therefore it has higher practical value.

Key words: discrete hidden Markov model; classifying ability; characteristic pattern; vector quantization; robustness

* Received December 28, 1999; accepted May 18, 2000

Supported by the National Natural Science Foundation of China under Grant No. 69872036