

分类号 _____
UDC _____

学号 02020063
密级 _____

理学硕士学位论文

HMM 的学习问题和解码问题研究

硕士生 姓名 韩 广 军

学 科 专 业 概率论与数理统计

研 究 方 向 应用概率

指 导 教 师 李 兵 教 授

国防科学技术大学研究生院

二零零四年十一月

摘要

隐马氏模型(the Hidden Markov Model, HMM)是一种重要的统计信号模型。这一模型的基本理论由 L. E. Baum 在上世纪六十年代给出,随后这一模型逐渐被应用到很多领域,如语音识别、基因关联分析和基因识别、文字识别、图象处理、目标跟踪和信号处理等。隐马氏模型需要解决三个问题:学习问题、识别问题和解码问题。对这三个问题的回答就构成了隐马氏模型理论。

本论文创新之处:

1 由于 HMM 的传统的参数训练方法容易陷入问题的局部最优解,而且 HMM 的参数训练属于带约束的优化问题,所以一般的优化方法很难解决这种问题。而遗传算法却相对具有搜索全局最优解的能力,而且它处理约束问题的能力也是令人满意的。故本文通过结合一种比较机制并运用遗传算法进行了 HMM 的参数训练,其中的仿真试验充分说明了该方法的可靠性。

2 由于带有连续观察空间的连续时间 HMM 较难处理,所以很少被考虑。本文对之进行了研究,得到了该结构下一类特殊 HMM 的参数最大后验概率(MAP)估计。

3 对于我们接触到的识别方法---传统的 Viterbi 解码算法和简单模型下的基于滤波和内插的方法来说,后者操作起来相对前者较简单,所以为了拓广后者的应用范围,我们将后者进行了推广,即考虑了基于滤波和内插方法的非齐次非平稳隐马氏链的状态识别算法。同时我们还对该方法下的识别误差进行了分析。

4 本文还在第三章第五节总结了隐马氏模型应用在语音降噪处理中的理论基础。有些信号是伴随着较强的噪音发送的,即我们接受到的信号除了系统误差外,还有额外的附加噪声过程。为了准确地估计信号,我们必须先对接收到的信号作初步的处理,使得处理后的信号相对接近真实信号。这样有助于提高我们的解码质量。

关键词: 隐马氏模型 遗传算法 扩散过程 Viterbi 算法 拓广的 EM 算法
Jenson 不等式 滤波

ABSTRACT

The hidden Markov model (HMM) is a sort of statistical signal model. The basal theory of this model was brought forward in 1960s by L. E. Baum, then it has been gradually applied to many fields such as speech recognition, gene relative analysis and gene recognition, character recognition, image processing, target tracking and signal processing etc. There are three problems needed to be solved by HMM, which are training, recognizing and decoding. The answers for those three problems consist of the theory of HMM. Parameter estimation or train is the core problem of the training process. The work of this paper can be generalized as follow.

Innovations of this article.

1. It is easier to get into the trap of local best solutions applying traditional training methods of HMM to training parameters, what's more, the training of parameters belong to optimization problem with restriction, so it's difficult to apply general methods to solving this problem. Luckily Genetic Algorithm has the ability of searching global best solutions relatively, and its ability of dealing with optimization problem with restriction. So this paper uses GA to train the parameters combining a sort of comparison mechanism. The result of simulation embodies the reliability of this method.

2. HMM with continuous time and continuous observation space is rarely considered since it is difficult to deal with it. However it is studied in this article and a sort of MAP estimation of parameter is derived.

3. As for as the recognition methods we have known---traditional Viterbi algorithm and the recognition algorithm of Hidden Markov Model based on file and smoothing are concerned, the latter does better than the former. So we extend this method to inhomogeneous nonstationary HMM in order to extend its applicational range. Meanwhile we also analyze the error of this algorithm.

4. The fifth section of third chapter of this paper generalizes the theory foundations of application of HMM in reducing noises of speech. Some signals are sent together with stronger noises, i.e. besides the error of system, the signals we have received also include additive noises. In order to estimate the signal well and truly, the primary disposal has to be done firstly so that the signal disposed is close to the true signal relatively. So the quality of recognition may be enhanced.

Keywords: HMM Genetic Algorithm Diffusion Process Viterbi algorithm Filter
Extended EM Algorithm Jensen Inequality

独 创 性 声 明

本人声明所呈交的学位论文是我本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表和撰写过的研究成果，也不包含为获得国防科学技术大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文题目： HMM 的学习问题和解码问题研究

学位论文作者签名： _____ 日期： _____ 年 _____ 月 _____ 日

学位论文版权使用授权书

本人完全了解国防科学技术大学有关保留、使用学位论文的规定。本人授权国防科学技术大学可以保留并向国家有关部门或机构送交论文的复印件和电子文档，允许论文被查阅和借阅；可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

(保密学位论文在解密后适用本授权书。)

学位论文题目： HMM 的学习问题和解码问题研究

学位论文作者签名： _____ 日期： _____ 年 _____ 月 _____ 日

作者指导教师签名： 李 兵 日期： 2004 年 12 月 10 日

第一章 绪 论

人类社会已经进入信息时代,通信与网络作为信息时代的两大技术支撑有力地促成了全球的一体化。通信是人类社会发展的基础,而信号是了解通信系统的本质所在,因此对信号进行加工处理和识别就成了一项突出的任务。隐马氏模型是完成此项任务的最突出的技术之一。隐马氏模型是一种非常重要的信号与系统分析工具。它由 L. E. Baum 于上世纪六十年代首先提出,在上世纪七十年代由 Jenik 等人应用到语音识别领域,然后逐步发展成为语音识别领域中最受瞩目、最有效的技术之一^[1,2]。不仅如此,进入九十年代以后,随着人类基因组计划的启动,它在生物统计、基因关联分析和基因识别等方面的成功应用更加光彩夺目和有目共睹^[1,2]。此外,目前它在目标跟踪、图象处理等方面也有广泛应用。

§ 1.1 HMM 相关理论的发展过程

隐马氏模型是一种不完全数据的统计模型,它包括两个随机过程,其中一个过程不能够直接观测并且具有马尔科夫性,称为状态马氏过程;另一个是受此状态马氏过程决定的可观测的随机过程,称为观测过程。隐马氏模型的理论研究实际上是对三个问题:学习问题、识别问题和解码问题的研究。学习问题的目的是通过对样本集的统计计算来调整模型参数使得对每一个模式找到一组最适合样本集的参数。调整模型参数的过程就是参数估计的过程或者说是参数训练的过程。目前来讲,相应的方法也有不少,按它们出现的先后顺序,有 Baum 和 Welch 共同给出的 B-W 公式^[3],该方法适合于离散时间隐马氏模型,大量的试验和应用表明,它是一种不错的方法,不足之处在于,这种方法比较容易陷入局部最优解。另外一种是比较传统的统计计算方法—EM 算法^[4],它首先由 Baum 和 Eagon 等人应用到一般隐马氏模型当中。此算法实际上是一种近似极大似然估计的估计方法,它与 B-W 公式有着共同的缺陷,即容易陷入局部最优解。2000 年,Steffen Michalek, Mirko Wagner, Jens Timmerg 给出了一种新的自回归滑动平均隐马氏模型的似然估计。这也是隐马氏模型在时间序列领域的典型应用。2002 年 2 月, Vikram Krishnamurthy 和 George Gang Yin 给出了隐马氏模型的一种递归估计,讨论了这种估计的收敛和收敛速度以及怎样提高这种速度。与众不同地是,2002 年 5 月, Peter M. Djuric 和 Joon-Hwa Chun 运用马氏链蒙特卡罗抽样方法^[5]将离散时间隐马氏模型的参数训练和状态识别交叉进行,达到了不错

的效果。马氏链蒙特卡罗方法是一种重要的信号处理方法,它有着鲜明的统计特点,其基本思想是通过建立一个平稳分布为某种分布的马氏链来得到这种分布的样本,基于这些样本我们可以作各种统计推断。除此之外还有如今比较流行的软计算技术,如模拟退火算法,遗传算法^[6],以及神经网络方法等。遗传算法的搜索能力固然强大,但对于系统中的反馈信息却无能为力,这往往造成大量的无为迭代,而简单遗传算法与蚂蚁算法的结合^[7]较好地解决了这个问题。不管怎样,单独使用遗传算法,其效果是较差的。软技术之间以及软技术与其它技术的相互结合将是未来的发展趋势,本文的一部分内容将遗传算法,模糊技术和一种比较机制结合起来较好地解决了一类隐马氏模型的参数训练问题的,试验结果表明,该方法相对于传统的 B-W 公式是比较好的(可参见文献[8])。解码问题,即常说的隐状态的识别问题往往是我们解决实际问题的最终目的。相应的识别方法有传统的 Viterbi 解码^[3]算法,这种方法也是针对离散时间隐马氏模型设计的,广泛的应用表明它是一种识别率很高的方法。但它首先是在简单 HMM 下给出的,所以,它所能处理的模型比较单一。以至于人们将其推广到了可应用到非平稳隐马氏模型上的 Viterbi 解码算法。在识别方法的应用方面,1989 年, Yairiv Ephraim, David Malah 和 Biing-Hwang Juang 以隐马氏模型为工具运用最大后验概率方法研究了语音的降噪。1992 年, Yairiv Ephraim 以隐马氏模型为工具运用贝叶斯估计方法研究了语音领域里的语音增强或者说是语音降噪。1996 年,文献[9]给出了基于滤波和内插的齐次 HMM 的状态识别方法。受文献[9]启发,本文研究了基于滤波和内插的带有逗留时间的离散时间非齐次 HMM 的状态识别问题,得到了一种可操作的较好的算法。1998 年 Jason J. Ford 和 John B. Moore 给出了一种基于广义最小二乘估计和递归预测误差方法的状态识别方案。纵观隐马氏模型的应用,大多数是针对离散时间的,这与连续时间 HMM 较难处理有关。而本文则对一类特殊的带有连续时间指标的隐马氏模型进行了研究,得到了求该类隐马氏模型参数的最大后验概率估计的算法。总之,性能更好的参数训练和状态识别方法将是我们永恒的追求。

由于隐马氏模型来自工程模式识别中,具有很强的工程实际背景,在很多情形下(例如在语音识别中考虑时域上的相关性)很难推导出严格的数学公式,另外,很多结果的得出必须要在相当多的条件的假定下才能得到,所以隐马氏模型作为一种新的理论虽然被广泛应用到各个领域,但其仍然是有缺点和局限性的^[1]。这些缺点和不足恰好为研究者提供了广阔的思维空间。

§ 1.2 本文工作

由于运用传统的方法进行参数训练容易陷入问题的局部最优解,而且HMM的参数训练属于带约束的优化问题,所以一般的优化方法很难解决这种问题。而遗传算法却相对具有跳出局部最优解的能力,而且处理约束的能力也是令人满意的。故本文第二章使用带有一种比较机制的DCPM-FGA方法^[10]进行HMM的参数训练。其中的仿真试验充分说明了该技术的可靠性。但它也有不足之处,那就是参数的设定需要特别地谨慎,否则就达不到预期的效果。再者,初始化对结果也有影响。这实际上与我们研究的问题有关,我们研究的问题本质上属于带不等式约束的优化问题,目前来讲,没有一种方法可以解决这种问题,我们所能做的是利用最新的方法解决我们的问题。不仅如此,本章还从估计角度出发,就一类带有连续时间指标的特殊模型进行了研究,得到了可操作的求参数的最大后验概率估计的算法。但这种方法也有不足之处,首先,Riccati方程的解用一般的方法很难求得;再者,参数估计序列仅仅收敛到对数后验概率的稳定点,从而其极限有可能是次优的。本章接着总结了前人给出的带逗留隐马氏模型的参数重估计公式。第三章主要讨论隐马氏模型的状态识别问题。首先,总结回顾了一下传统的状态识别方法—Viterbi解码算法;由于滤波和内插是降噪的有力工具,所以接着我们基于一种鞅差表示定理针对非时齐带逗留时间隐马氏模型的状态识别问题进行了研究,得到了其滤波和内插的近似表达式,其中仿真试验的结果说明该方法还是不错的。在给出了运用滤波和内插方法进行识别的方法后,本章接着在一类特殊隐马氏模型下对这种方法的误差进行了细致的分析,得到了该误差的下界;然后本章介绍了运用一种马氏链蒙特卡罗方法—Gibbs方法进行状态识别的具体步骤。最后讨论了隐马氏模型应用在语音降噪处理中的理论基础。由于有些信号是伴随着较强的噪音发送的,即我们接受到的信号除了系统误差外,还有额外的附加噪声过程。为了准确地估计信号,我们必须先对接收到的信号作初步的处理,使得处理后的信号相对接近真实信号,这样有助于提高我们的解码质量。所以本章最后对隐马氏模型应用在语音降噪处理中的理论基础进行了讨论。

第二章 HMM 的几种新的参数训练方法

在 HMM 理论当中, HMM 的参数训练占有十分重要的地位, 针对之, 相关数学工作者一直在努力地寻求更简单, 更易操作, 更准确的估计求优方法。在应用过程中, 也先后出现了不少的方法。分析这些方法不难发现参数训练都是通过两种途径进行的。一是通过优化, 二是通过估计。本章分别从优化和估计角度针对相应的模型提出了两种方法。

本章的创新之处在于以下两点:

一、隐马氏模型的参数训练就是在区间 $(0,1)$ 内搜索一组与样本最为匹配的参数, 其结果的质量显然依赖于搜索方法的功能。鉴于此我查阅了很多优化方面的资料, 发现遗传算法具有多初值同时搜索功能, 从而相对来讲, 所得结果会更好, 搜索到相同质量解的时间相对较少。因而引入遗传算法进行参数训练。从所作的大量实验中, 我们发现了直接运用标准隐马氏模型进行训练会产生很多问题。而文献[10]给出了一种个体寻优的比较机制, 形成了 DCPM-FGA 方法。这种方法搜索到的解的质量更高, 但也存在另外一个问题, 即早熟现象。为此, 本文又采取了控制措施进行, 有效地解决了过早收敛的现象, 从而使得试验结果更加可信。

二、连续时间隐马氏模型的参数训练问题, 理论上可以讨论, 但实际上做起来比较困难。本文从另外一种角度——估计角度出发, 对一类带有连续时间指标的特殊模型进行了研究, 得到了可操作的求参数的最大后验概率估计的算法。

§ 2.1 基于 DCPM-FGA 的离散时间 HMM 的参数训练

由于传统的方法容易陷入局部最优解的圈套, 所以寻求更好的参数训练方法具有十分重要的意义。令人高兴地是, 新近发展起来的遗传算法在优化方面显示了强大的生命力, 实践证明它是一种有效的方法。计算机的性能越强大, 该方法的能力就越明显。它具有很强的全局搜索功能, 从而克服了传统的梯度下降法容易陷入局部最优解的缺点。特别地, 将遗传算法和其它计算技术结合起来所达到的效果会更好。本节就是将一种比较机制和模糊遗传算法结合进行参数训练的。

2.1.1 遗传算法简介

遗传算法 (Genetic Algorithm, 简称 GA) 是由美国密执安大学的 John Holland 教

授于 1975 年首先提出的一类仿生型优化算法。它是以达尔文的生物进化论“适者生存，优胜劣汰”和孟德尔的遗传变异理论“生物进化主要在染色体上，子代是父代遗传基因在染色体上的有序排列”为基础，模拟生物界进化过程，以寻求问题的全局最优解。具体地，它从一群编码化的可行解（称为种群）出发通过种群的更新与迭代来搜索全局最优解。种群中的每个个体都是问题的一个解，生物学上称之为染色体。染色体是一串符号，比如一个二进制串。这些染色体在后续迭代中不断进化，称作遗传。种群的迭代是通过选择，交叉和变异等具有生物意义的遗传算子来实现的。每一代中用适应值来测量染色体的好坏。生成的下一代染色体称为后代，后代是由前一代染色体经过选择，交叉和变异等遗传操作得到的。在新一代的形成中，正是根据适应值的大小正比地选择了部分后代，淘汰了部分后代。但为防止陷入局部最优解，适应值低的染色体也有可能被选中，如转轮法选择就可达到此目的。详见文献[10]，[12]，[13]。

遗传算法具有如下优点：

- (1) 具有大范围全局搜索的能力，且它与问题的领域无关；
- (2) 从群体出发进行搜索，具有潜在的并行性；可进行多值比较，鲁棒性强；
- (3) 搜索使用评价函数启发，过程简单；
- (4) 使用概率机制进行迭代，具有随机性；
- (5) 具有可扩展性，容易与其它算法结合；

其缺点：对于系统中的反馈信息利用不够，但求解到一定范围时往往做出大量无为的多余迭代，使得求精确解效率低。

2.1.2 HMM 的参数训练

所谓 HMM 的参数训练，就是根据极大似然准则找 λ^* ，使得

$$\lambda^* = \arg \max_{\lambda} P(Y=y|\lambda)$$

其中 Y 为观察， λ 为模式。且过程带有约束： $\forall i, \pi_i \neq 1, a_{ij} \neq 1, b_{jk} \neq 1, \sum_j \pi_j = 1, \forall i, \sum_j a_{ij} = 1,$

$\forall j, \sum_k b_{jk} = 1$ 。因而最大化 $P(Y=y|\lambda)$ 其实就是一个带有约束的优化问题。针对之，近年来

已经提出了诸如拒绝策略，修复策略，改进遗传算子和惩罚策略等用遗传算法满足约束的技术。但大量的实验证明 DCPM-FGA 方法与前面那些技术相比无论是最好解的质量还是解的平均质量都显著地超过了以上技术。

我们准备通过如下方法来计算 $P(Y=y|\lambda)$ 。

首先, 用 T 时刻的状态来分解 $P(Y=y|\lambda)$, 即

$$P(Y=y|\lambda) = \sum_j P(Y(1)=y(1), \dots, Y(T)=y(T), X(T)=j|\lambda)$$

然后, 令 $\alpha_i(i) = P(Y(1)=y(1), \dots, Y(t)=y(t), X(t)=i|\lambda)$, 称之为前向概率, 则

$$\begin{aligned} \alpha_{t+1}(i) &= P(Y(1)=y(1), \dots, Y(t+1)=y(t+1), X(t+1)=i|\lambda) \\ &= \sum_j P(Y(1)=y(1), \dots, Y(t)=y(t), X(t)=j, Y(t+1)=y(t+1), X(t+1)=i|\lambda) \\ &= \sum_j P(Y(1)=y(1), \dots, Y(t)=y(t), X(t)=j|\lambda) \cdot \\ &\quad P(Y(t+1)=y(t+1), X(t+1)=i|Y(1)=y(1), \dots, Y(t)=y(t), X(t)=j, \lambda) \\ &= \sum_j \alpha_t(j) \cdot P(X(t+1)=i|Y(1)=y(1), \dots, Y(t)=y(t), X(t)=j, \lambda) \cdot \\ &\quad P(Y(t+1)=y(t+1)|Y(1)=y(1), \dots, Y(t)=y(t), X(t)=j, X(t+1)=i, \lambda) \\ &= \sum_j \alpha_t(j) \cdot a_{ji} \cdot b_i(y(t+1)) \end{aligned}$$

在 HMM 中, 我们一般假定

$$\begin{aligned} &P(Y(t+1)=y(t+1)|Y(1)=y(1), \dots, Y(t)=y(t), X(t)=j, X(t+1)=i|\lambda) \\ &= P(Y(t+1)=y(t+1)|X(t+1)=i, \lambda) \end{aligned}$$

而 $\alpha_1(i) = \pi_i \cdot b_i(y(1))$. 故我们可通过如下式子容易地计算 $P(Y=y|\lambda)$,

$$P(Y=y|\lambda) = \sum_j \alpha_T(j).$$

2.1.3 DCPM-FGA 方法^[10]

将 DCPM(direct comparison-proportional method: 直接比较比例方法)与 FGA(fuzzy Genetic Algorithm: 模糊遗传算法) 结合即为 DCPM-FGA 方法。它不必显示地利用目标函数的某些性质, 可以采用非处处可微的罚函数形式。同时对于遗传算法所实行的选择方法来说, 适应值所提供的唯一功能就是对给定个体进行比较和排序。当采用竞争选择法时可以完全不必显示地生成新的函数, 而只需通过建立基于适应值函数和罚函数的直接比较机制, 这样就可避免选择罚因子; 为处理约束问题提供了一种全新的思路。DCPM 方法提供了种群中个体之间的比较准则, 并给出在种群中保持一定比例的不可行解的适应性策略。具体如下。

首先, 定义一个刻画个体 X 违反约束程度的量:

$$Viol(X) \triangleq \sum_{j=1}^m f_j(X)$$

其中, $f_j(X)$ 为个体 X 违反其中一个约束的违反程度。然后, 对事先给定的 $\varepsilon > 0$, DCPM 按照如下准则比较两个个体 X 和 X' 的优劣: (以极大化问题为例)

- (1) 当两个个体都可行时, 比较适应值, 适应值大的个体为优;
- (2) 当两个个体都不可行时, 比较罚函数值, 罚值小的个体为优;
- (3) 当其中一个, 比如 X , 可行, 而 X' 不可行时, 若 $Viol(X') \leq \varepsilon$, 则比较适应值, 适应值大的个体为优; 否则, 即 $Viol(X') > \varepsilon$, 则 X 为优。

显然, ε 越大, 群体中不可行解的比例就可能越高, 为了将不可行解的比例控制在一个事先规定的合理水平 p ($p > 0$), 引入如下适应性调整策略。对给定的正整数 K , 从群体中产生可行解的那一代起, 每进化 K 代, 就计算出这 K 代中的每一代中的不可行解在群体中的比例, 并按如下准则将 ε 调整为 ε' :

- (1) 若所有比例都小于 p , 则 $\varepsilon' = 1.2\varepsilon$;
- (2) 若所有比例都大于 p , 则 $\varepsilon' = 0.8\varepsilon$;
- (3) 其它, 则 $\varepsilon' = \varepsilon$ 。

大量的试验证明 DCPM-FGA 与以往那些技术相比, 无论是最好解的质量还是解的平均质量, 都显著地超过了以上技术, 且 DCPM-FGA 处理约束条件并搜索到高性能解的能力也是令人满意的。

2.1.4 基于 DCPM-FGA 的 HMM 的参数训练

运用 DCPM-FGA 进行参数训练的算法流程:

1. 确定实数形式编码;
2. 确定要使用的选择方法、遗传算子及各参数值, 包括 ε , p 和 K 的值;
3. 设置代数 $t = 0$;
4. 随机生成初始化种群 $X(0)$;
5. 计算每个个体的适应值和罚函数值;
6. 判断有无可行解, 有则计算不可行解的比例;
7. 当终止条件不满足时, 令 $t = t + 1$;
8. 根据适应值、罚值和 ε 的值按照比较准则用竞争选择法从 $X(t-1)$ 中选择出 $X'(t)$;
9. 根据交叉概率 p_c 用交叉算子对 $X'(t)$ 中的个体进行交叉操作, 得种群 $X''(t)$;

10. 根据变异概率 p_m 用变异算子对 $X''(t)$ 中的个体进行操作, 得到第 t 代种群 $X(t)$;
11. 计算 $X(t)$ 中个体的适应值和罚值, 判断有无可行解, 若有则计算不可行解的比例;
12. 在群体中出现第一个可行解后, 每隔 K 代, 按照上面所述调整 ε 的值。

遗传策略: 在运用遗传算法处理优化问题时, 较好的遗传策略往往导致能够以较大的概率较快地搜寻到全局最优解。遗传策略包括很多的内容, 简单一点, 它主要涉及七大要素: 编码设计、初始种群设定、种群规模设定、适应值函数设计、遗传算子设计、控制参数设计、终止条件设定。

仿真试验及结果分析: 我们针对一个简单的 HMM 模型在以下遗传策略下运用以上方法用 Matlab 数学软件进行编程并执行了数次。其中状态集设为 $S = \{1, 2, 3\}$, 码本大小分别为 28 和 4, 观察量分别为 130 和 5, 最大代数 $T=2000$, 种群规模 $N=40$ 。

适应函数 在大多数情况下, 适应值函数取为目标函数。此处我们以基于前向概率的 $P(Y=y|\lambda)$ 为适应值函数。

编码选择 一般来说, 应根据实际问题选择编码方式。遗传算法对编码要求并不苛刻。大量试验证明, 对同一优化问题二进制编码和实数编码并不存在显著的差异。这儿, 我们用实数编码。

种群规模设定 它主要取决于个体中所含变量的个数, 种群规模 $N=40$ 。

初始种群生成 若一开始就让解中每个变量在 $[0, 1]$ 中随机产生, 则搜索空间太大, 花费时间很长, 甚至是不可能做到的。所以, 我们就让初始种群中大部分是可行解。

遗传算子设计 我们主要采用: 算术交叉, 单点随机变异, 竞争选择法。

参数设定。 $\varepsilon = 0.0001$, $K = 5$, $p = 0.1$, $p_c = 0.25$, $p_m = 0.012$, 竞争规模为 2, 精度为 0.0001。

终止条件设定 我们用最大代数终止法。

罚函数 采用绝对值罚函数形式, 即

$$Viol(X) = \sum_{j=1}^3 f_j(X)$$

$$\text{其中 } f_1(X) = \left| \sum_{i=1}^N \pi_i - 1 \right|, \quad f_2(X) = \sum_{i=1}^N \left| \sum_{j=1}^N a_{ij} - 1 \right|, \quad f_3(X) = \sum_{j=1}^N \left| \sum_{k=1}^M b_j(K) - 1 \right|.$$

加入控制早熟的因子

$$c_k = \min\{1, (F(X'(k))/F(X(k)))^k\}, \text{ 其中 } F(X(k)) = \max\{f(X_i(k)), 1 \leq i \leq N\}$$

在选择第 $k+1$ 代时, 先产生 $[0, 1]$ 中的一个随机数 ρ , 然后有

1. 若 $\rho \leq c_k$, 则令 $X(k+1) = X'(k)$; 2. 若 $\rho > c_k$, 则令 $X(k+1) = X(k)$.

由于我们采用最大代数终止, 所以每次得到不同的结果是合理的。在多次试验中, 当码本大小为128时, 所得最好结果为 $\log P(Y=y|\lambda) = -45.3477$, 与 Baum-Welch 公式得到的实验结果 $\log P(Y=y|\lambda) = -48.22$ 及 Viterbi 算法的 $\log P(Y=y|\lambda) = -51.55$ 相比明显提高了很多。当码本大小为4时, 最好结果为 $\log P(Y=y|\lambda) = -4.9477$, 而用 B-W 方法得到的最好的结果为 -4.7531 . 需要注意的是前面所得结论都是在转移概率矩阵和特征概率矩阵中的元素都不等于1的前提下得到的。若允许其中的参数可取1, 则 B-W 方法得到的最好的结果和用遗传算法一样为 $P(Y=y|\lambda) = 1$, 但遗传算法的收敛速度要慢一些。值得注意的是选择概率, 交叉概率以及变异概率的选择也是较困难的, 它对结果的影响比较大。在此我们仅仅根据经验对它们进行了选择。所以结果不是很理想。

-4.9477 对应的 λ 为:

$$\pi: (0.2703 \quad 0.5699 \quad 0.1598)$$

$$A: \begin{bmatrix} 0.2424 & 0.0866 & 0.6710 \\ 0.9928 & 0.0029 & 0.0043 \\ 0.2668 & 0.3614 & 0.3718 \end{bmatrix}$$

$$B: \begin{bmatrix} 0.0226 & 0.7916 & 0.0586 & 0.1271 \\ 0.3779 & 0.0826 & 0.2674 & 0.2721 \\ 0.1157 & 0.1478 & 0.5038 & 0.2327 \end{bmatrix}$$

下面是搜寻到解 λ 的途径的图形:

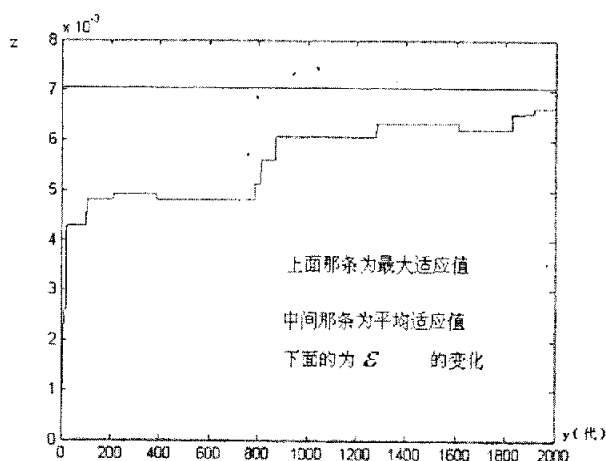


图 2.1.1 适应值函数 (y 为代数, z 为适应值)

由于我们采取的是最大代数终止, 所以我们仍然没有跳出有限搜索的圈子, 所得到的结果也有可能只是局部最优, 我们只是利用它相对于传统方法的优点。但由于我们采取了竞争保留策略, 故理论上它是一定收敛的, 虽然图中没有显示出。明显其平均适应值是在上升的, 这充分说明了解的进化性。

试验证明 DCPM-FGA 对参数 ε 和 K 的不同选择计算结果都差异不大, 不足之处就在于参数的设定需要特别的谨慎, 否则就达不到预期的效果。再者, 初始化对结果也有影响。但是, 由试验可知它却具有比较的优势。尽管如此, 我们的结果也还是不理想, 很多地方有待改进。这实际上与我们研究的问题有关, 我们研究的问题本质上属于带不等式约束的优化问题, 目前来讲, 没有一种方法可以解决这种问题, 我们所能做的是利用最新的方法解决我们的问题。

§2.2 一类扩散 HMM 参数的 MAP 估计

有限状态离散时间 HMM 的参数状态估计问题已经研究得很多, 并且形成了几种成熟的算法, 如 B-W 算法, Viterbi 算法, MCMC 抽样方法, 还有一些软计算技术, 如遗传算法, 模拟退火算法等。但实际当中却也存在着具有连续时间不可列状态的隐过程或信号过程, 为了能够运用 HMM 技术对此类过程相应的模型参数进行估计, 寻求一种针对连续时间不可列状态的 HMM 的参数估计方法将是十分有意义的。

我们具体讨论了一类此种模型, 即一类扩散 HMM 参数的 MAP 估计。基本做法是在滤波和内插的基础上构造参数估计的迭代序列, 然后我们从理论上说明随着迭代的进行, 对数后验概率是单调增的, 且迭代序列收敛到对数后验概率的稳定点。我们以该点作为参数的最终估计。

给定概率空间 (Ω, F, P) , 假设 $X = (X(t), 0 \leq t \leq T)$ 为扩散马氏过程, 不可观测, $Y = (Y(t), 0 \leq t \leq T)$ 为观测过程, 从而有定义于概率空间 (Ω, F, P) 上的扩散 HMM (X, Y) , 且假设它们满足如下信号模型

$$\begin{cases} dX(t) = \theta X(t) dt + dW(t) \\ dY(t) = \alpha X(t) dt + dV(t) \end{cases} \quad (2.2.1)$$

其中 $W(t)$ 与 $V(t)$ 为两个相互独立的布朗运动。从而扩散 HMM 的模型参数就为 $\lambda = (\theta, \alpha)$ 。这里我们假设 α 已知, 而 $\theta \in R$ 为未知参数。假设 $X(0)$ 与 $W(t)$ 独立, 且 $X(0) = 0$,

$\int_0^T X^2(s)ds < \infty, a.s.P, \int_0^T W^2(s)ds < \infty, a.s.P$. 我们的目标是去估计 θ .

下面我们首先说明满足 (2.2.1) 式的马氏过程的确是存在的, 因为这是我们讨论的基础. $\forall x, y \in R$,

$$\begin{aligned} |\theta x - \theta y| &= |\theta| |x - y| \leq (|\theta| + 1) |x - y| \\ |\theta x| + 1 &\leq (1 + |\theta|)(1 + |x|) \end{aligned}$$

从而相应的 Lipschitz 条件和线性增长条件成立. 又由前面的假设 $X(0)$ 与 $W(t), t \leq T$ 独立, 故根据文献[14]中的定理 4.10, $dX(t) = \theta X(t)dt + dW(t)$ 存在唯一的强解, 且解具有马氏性.

令 $Y_0^T = \{Y(s), 0 \leq s \leq T\}$, 则由文献[14]中的定理 7.7 知, 滤波 $\hat{X}(t) \triangleq E_\theta(X(t)|Y_0^T)$ 所满足的方程为

$$d\hat{X}(t) = [\theta - \alpha^2 S(t)]\hat{X}(t)dt + \alpha S(t)dY(t) \quad (2.2.2)$$

其中 $S(t) \triangleq E_\theta(X(t) - \hat{X}(t))^2$ 为滤波的均方误差, 它满足 Riccati 方程

$$\frac{dS(t)}{dt} = 2\theta S(t) - \alpha^2 S^2(t) + 1 \quad (2.2.3)$$

这样一个非随机微分方程.

内插 $\hat{X}(T-r|T) \triangleq E_\theta(X(T-r)|Y_0^T)$ 及其均方误差 $S(m|T) \triangleq E_\theta(X(m) - \hat{X}(m|T))^2$ 满足如下方程^[15]

$$\frac{d\hat{X}(T-r|T)}{dr} = \frac{\hat{X}(T-r)}{S(T-r)} - \left(\theta + \frac{1}{S(T-r)} \right) \hat{X}(T-r|T) \quad (2.2.4)$$

$$\frac{dS(T-r|T)}{dr} = 1 - 2 \left(\theta + \frac{1}{S(T-r)} \right) S(T-r|T) \quad (2.2.5)$$

注:

1. $\hat{X}(T|T) = \hat{X}(T)$.
2. $S(T|T) = S(T)$.

由 (2.2.2), (2.2.3), (2.2.4), (2.2.5) 就可求出或用其它数学方法近似求出滤波, 内插及它们的均方误差. 它们将在估计 θ 时用到.

在已知观测数据 $\{Y(s), 0 \leq s \leq T\}$ 的基础上, 考虑 θ 的估计问题。具体做法为

1. 随机产生一个实数作为 θ 的初始化。

2. 在已知 $\hat{\theta}_p (p \geq 1)$ 的基础上, 利用 (2.2.2), (2.2.3), (2.2.4), (2.2.5) 求出 $\hat{X}_p(T)$, $S_p(T)$, $\hat{X}_p(m|T)$, $S_p(m|T)$, 它们分别为第 p 次迭代时的滤波及其均方误差, 内插及其均方误差。

3. 利用 $\hat{X}_p(T), S_p(T)$, $\hat{X}_p(m|T)$, $S_p(m|T)$ 可求出 $\hat{\theta}_{p+1}$, 这从下面的定理 2.2.1 可看到。

由于 X 和 Y 是带有连续时间指标的过程, 从而在对 θ 作估计时我们考虑由文献[15]提供的推广的 EM 算法。接下来给出该算法的描述及理论基础。

令 (Ω_X, F_X) 和 (Ω_Y, F_Y) 分别为两个波雷尔可测空间, 分别称为状态过程空间和观测过程空间。 θ 为未知参数, 从贝叶斯观点, 可将它看作 $(R, B_R, P_\theta(\theta))$ 上的随机变量。假设 $\forall \theta$, $P_\theta(\cdot)$ 都使得 $(\Omega_{X \times Y}, F_1, P_\theta(\cdot))$ 为完备概率空间, 其中 $(\Omega_{X \times Y}, F_1)$ 为 $HMM(X, Y)$ 的过程空间。我们准备在已知观测 $\{Y(s), 0 \leq s \leq T\}$ 的情况下, 考虑 θ 的 MAP 估计。为此, 我们先作如下假设:

$$\forall \theta \neq \theta', \frac{dP_\theta(X|Y)}{dP_{\theta'}(X|Y)} \neq 1, a.s. dP_{\theta'}(X|Y) \quad (2.2.6)$$

其中 $P_\theta(\cdot|Y)$ 为条件测度。

由前面的假设可知

$$\begin{aligned} \int_0^T (\theta X(s))^2 ds &= \theta^2 \int_0^T \dot{X}^2(s) ds < \infty, a.s. P \\ \int_0^T (\theta W(s))^2 ds &= \theta^2 \int_0^T W^2(s) ds < \infty, a.s. P \end{aligned}$$

从而

$$\begin{aligned} P\left(\int_0^T (\theta X(s))^2 ds < \infty\right) &= 1 \\ P\left(\int_0^T (\theta W(s))^2 ds < \infty\right) &= 1 \end{aligned}$$

故由文献[14]中定理 6.8 可知, $\forall \theta, P_\theta \sim \mu_w$, 从而 $\forall \theta, \theta'$, P_θ 关于 $P_{\theta'}$ 的 Radon-Nikodym 导

数 $\frac{dP_\theta}{dP_{\theta'}}$ 存在, 其中 P_θ 为扩散过程测度, μ_w 为维纳过程测度。

令 P_θ^y 表示 P_θ 在 $\sigma(Y(s), 0 \leq s \leq T)$ 上的限制, 则

$$\frac{dP_\theta^y}{dP_{\theta'}^y} = E_{\theta'} \left(\frac{dP_\theta}{dP_{\theta'}} \middle| Y_0^T \right) \quad (2.2.7)$$

由后验概率的定义可给出如下定义

定义 2.2.1 $\hat{\theta}_{MAP} \triangleq \arg \max \left\{ P_0(\theta) \frac{dP_\theta^y}{dP_\alpha^y} \middle| \theta \in R \right\}, \alpha \in R$

其中 $P_0(\theta)$ 为 θ 的先验分布。

为计算方便, 考虑对数后验概率

$$L_\alpha(\theta) \triangleq \log P_0(\theta) + \log \frac{dP_\theta^y}{dP_\alpha^y}$$

易见 $\forall \theta \neq \theta', L_\alpha(\theta') - L_\alpha(\theta) = \log \frac{dP_{\theta'}^y}{dP_\alpha^y} - \log \frac{dP_\theta^y}{dP_\alpha^y} = \log \frac{dP_{\theta'}^y}{dP_\theta^y}$. 再由 (2.2.7) 可得

$$\forall \theta \neq \theta', L_\alpha(\theta') - L_\alpha(\theta) = \log \frac{dP_{\theta'}^y}{dP_\theta^y} = \log E_\theta \left(\frac{dP_{\theta'}}{dP_\theta} \middle| Y_0^T \right)$$

又 $\frac{dP_{\theta'}}{dP_\theta}$ 为 Radon-Nikodym 导数, 从而一定关于 $\sigma(Y(s), 0 \leq s \leq T)$ σ -可积, 所以由 Jensen

不等式可得

$$L_\alpha(\theta') - L_\alpha(\theta) = \log E_\theta \left(\frac{dP_{\theta'}}{dP_\theta} \middle| Y_0^T \right) \geq E_\theta \left(\log \frac{dP_{\theta'}}{dP_\theta} \middle| Y_0^T \right)$$

由文献[15]知, 上式等号成立等价于 $\frac{dP_{\theta'}(X|Y_0^T)}{dP_\theta(X|Y_0^T)} = 1, a.s. dP_\theta(X|Y_0^T)$, 再由假设 (2.2.6) 可知,

上式等号成立等价于 $\theta = \theta'$.

令 $\Lambda_{\theta\theta'} = \frac{dP_{\theta'}}{dP_\theta}$, 则

$$L_\alpha(\theta') - L_\alpha(\theta) \geq E_\theta(\log \Lambda_{\theta\theta'} | Y_0^T) \quad (2.2.8)$$

(2.2.8) 在证明定理 2.2.2 时要用到。

定义 2.2.2 $\forall \theta, \theta' \in R, Q(\theta, \theta') \triangleq \log P_0(\theta) + E_{\theta'}(\log \Lambda_{\theta\theta'} | Y_0^T)$

定义 2.2.3 $\hat{\theta}_{p+1} \triangleq \arg \max \{ Q(\theta, \hat{\theta}_p) | \theta \in R \}$

对定义 2.2.2 中的先验分布 $P_0(\theta)$, 为便于给出下面的定理, 我们不妨设 θ 服从参数为 μ 和 σ 的正态分布, 假设 μ 和 σ 已知, 从而

$$P_0(\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(\theta-\mu)^2}{2\sigma^2}\right\}, \quad \theta \in R$$

定理 2.2.1 设 $\hat{\theta}_p$ 已知, 则

$$\hat{\theta}_{p+1} = \frac{\frac{1}{2} \left[S_p(T) + \hat{X}_p^2(T) - T \right] + \frac{\mu}{\sigma^2}}{\frac{1}{\sigma^2} + \int_0^T \left(S_p(m|T) + \hat{X}_p^2(m|T) \right) dm}$$

其中 $\hat{X}_p(T), S_p(T), \hat{X}_p(m|T), S_p(m|T)$ 如前所述。

证明 由 (2.2.1) 式知, $X(t)$ 所满足的微分方程为

$$dX(t) = \theta X(t) dt + dW(t) \quad (2.2.9)$$

由前面的假设及文献[14]中定理 6.8 可知 $\forall \theta, P_\theta \sim \mu_w$, 且

$$\frac{d\mu_w}{dP_\theta} \triangleq \frac{d\mu_w}{dP_\theta}(T, X) = \exp\left\{-\int_0^T \theta X(s) dX(s) + \frac{1}{2} \int_0^T \theta^2 X^2(s) ds\right\}$$

其中 P_θ 和 μ_w 如前所述。从而

$$\frac{dP_\theta}{dP_0} = \frac{dP_\theta}{d\mu_w} \times \frac{d\mu_w}{dP_0} = \exp\left\{\int_0^T \theta X(s) dX(s) - \frac{\theta^2}{2} \int_0^T X^2(s) ds\right\}$$

下面计算 $\int_0^T X(s) dX(s)$. 易见

$$X(T) = \int_0^T \theta X(s) ds + \int_0^T 1 dW(s)$$

令 $H(t) = \frac{1}{2} X^2(t)$, 由 Ito 公式, 则有

$$\begin{aligned} H(T) &= \int_0^T X(s) dW(s) + \int_0^T \theta X^2(s) ds + \frac{1}{2} T \\ &= \frac{1}{2} T + \int_0^T X(s) [dX(s) - \theta X(s) ds] + \int_0^T \theta X^2(s) ds \\ &= \frac{1}{2} T + \int_0^T X(s) dX(s) - \int_0^T \theta X^2(s) ds + \int_0^T \theta X^2(s) ds \\ &= \frac{1}{2} T + \int_0^T X(s) dX(s) \end{aligned}$$

所以

$$\int_0^T X(s) dX(s) = \frac{1}{2} X^2(T) - \frac{1}{2} T$$

从而

$$\frac{dP_\theta}{dP_0} = \exp \left\{ \frac{\theta}{2} (X^2(T) - T) - \frac{\theta^2}{2} \int_0^T X^2(s) ds \right\}$$

故

$$\Lambda_{\theta\theta'} = \frac{dP_\theta}{dP_{\theta'}} = \frac{dP_\theta}{dP_0} \times \frac{dP_0}{dP_{\theta'}} = \exp \left\{ \frac{1}{2} (X^2(T) - T) (\theta - \theta') - \frac{1}{2} (\theta^2 - \theta'^2) \int_0^T X^2(s) ds \right\}$$

通过交换积分次序可得

$$E_\theta (\log \Lambda_{\theta\theta'} | Y_0^T) = \frac{\theta - \theta'}{2} \left[E_\theta (X^2(T) | Y_0^T) - T \right] - \frac{\theta^2 - \theta'^2}{2} \int_0^T E_\theta (X^2(s) | Y_0^T) ds$$

故

$$Q(\theta, \theta') = \log \frac{1}{\sqrt{2\pi\sigma}} - \frac{(\theta - \mu)^2}{2\sigma^2} + E_\theta (\log \Lambda_{\theta\theta'} | Y_0^T)$$

令

$$\frac{\partial Q(\theta, \hat{\theta}_p)}{\partial \theta} = \frac{1}{2} \left[E_{\hat{\theta}_p} (X^2(T) | Y_0^T) - T \right] - \theta \int_0^T E_{\hat{\theta}_p} (X^2(s) | Y_0^T) ds + \frac{\mu}{\sigma^2} - \frac{\theta}{\sigma^2} \triangleq 0$$

可得

$$\hat{\theta}_{p+1} = \frac{\frac{1}{2} \left[E_{\hat{\theta}_p} (X^2(T) | Y_0^T) - T \right] + \frac{\mu}{\sigma^2}}{\frac{1}{\sigma^2} + \int_0^T E_{\hat{\theta}_p} (X^2(s) | Y_0^T) ds}$$

又由于 Riccati 方程的解是正的, 故 $\forall \theta$,

$$E_\theta (X^2(s) | Y_0^T) = E_\theta \left((X(s) - \hat{X}(s|T) + \hat{X}(s|T))^2 | Y_0^T \right) = S(s|T) + \hat{X}^2(s|T) > 0$$

从而

$$\hat{\theta}_{p+1} = \frac{\frac{1}{2} \left[S_p(T) + \hat{X}_p^2(T) - T \right] + \frac{\mu}{\sigma^2}}{\frac{1}{\sigma^2} + \int_0^T \left(S_p(m|T) + \hat{X}_p^2(m|T) \right) dm}$$

下面以定理的形式给出对数后验概率随着迭代的进行具有单增性。

定理 2.2.2 $L_\alpha(\hat{\theta}_{p+1}) \geq L_\alpha(\hat{\theta}_p)$, 等号成立等价于 $\hat{\theta}_{p+1} = \hat{\theta}_p$

证明 由 (2.2.8) 可得

$$L_{\alpha}(\hat{\theta}_{p+1}) - L_{\alpha}(\hat{\theta}_p) \geq E_{\hat{\theta}_p}(\log \Lambda_{\hat{\theta}_p, \hat{\theta}_{p+1}} | Y_0^T)$$

再由定义 2.2.3, 可有如下式子成立

$$E_{\hat{\theta}_p}(\log \Lambda_{\hat{\theta}_p, \hat{\theta}_{p+1}} | Y_0^T) \geq E_{\hat{\theta}_p}(\log \Lambda_{\hat{\theta}_p, \hat{\theta}_p} | Y_0^T) = 0$$

从而

$$L_{\alpha}(\hat{\theta}_{p+1}) \geq L_{\alpha}(\hat{\theta}_p)$$

由我们前面的论述可知, 等号成立等价于 $\hat{\theta}_{p+1} = \hat{\theta}_p$.

下面讨论一下迭代序列的收敛性。

下面我们以 $\alpha = 0$ 来讨论。由

$$L_0(\theta) = \log P_0(\theta) + \log E_0\left(\frac{dP_{\theta}}{dP_0} \middle| Y_0^T\right)$$

而

$$\frac{dP_{\theta}}{dP_0} = \exp\left\{\frac{\theta}{2}(X^2(T) - T) - \frac{\theta^2}{2} \int_0^T X^2(s) ds\right\}$$

可知 $L_0(\theta)$ 关于 θ 是可微的, 且 $\frac{dP_{\theta}}{dP_0} \rightarrow 0, |\theta| \rightarrow \infty$, 从而集合 $\{\theta, L_0(\theta) \geq L_0(\theta')\}$ 是相对紧的,

由 $L_0(\theta)$ 关于 θ 的连续性可知它是闭的, 从而为紧集。又明显地, $Q(\theta, \theta')$ 关于 θ 和 θ' 是连续的, 且迭代序列在 R 的内部。故由文献[15]中的定理 2 知, 迭代序列收敛到对数后验概率的稳定点。

当然, 就某一具体问题而言, 先验分布 $P_0(\theta)$ 可能是其它分布, 但做法是一样的。

本文类似于文献[5] (离散时间情形), 都要求参数的先验分布, 且以参数估计序列的极限作为最终的参数估计。从方法上讲, 这种方法是好的, 但不足之处也很明显, 首先, Riccati 方程的解用一般的方法很难求得; 再者, 参数估计序列仅仅收敛到对数后验概率的稳定点, 从而其极限有可能是次优的。

§2.3 一种特殊 HMM 的参数估计

我们已经知道, 用标准 HMM 去描述状态的逗留行为是不合适的^[5]。为此, 人们曾经提

出了一种非常明确的状态逗留模型,但它在为过程的内在动态性或非平稳性(即转移概率与时间有关)建模时又是不够充分的。更加不幸的是,此时大多数的努力都仅仅放在状态的内在平稳性上,例如,可变的状态逗留,以及局部的特征相关等。直到1995年, Bonkee Sin 和 H. Kim 在他们的论文[16]里考察了马氏链的非平稳性,并且用 $A(\tau) = (a_{ij}(\tau))$, τ 为状态的逗留时间,定义了非平稳 HMM。和传统的 HMM 相比,该模型可理解为标准 HMM 和状态逗留 HMM 的一般化^[16]。

HMM 就怎样为非平稳信号源建模给出了一个自然的框架,传统的时间序列分析正是在这种工具下被完成得非常完美^[16]。在 HMM 中,非平稳性是经过隐马氏链来展现的,即所谓的非平稳 HMM 实际上是指隐马氏链的非平稳性。然而标准 HMM 仅提供了一种固定的和较弱的机制去处理相关的演化过程,该过程模拟了动态信号的变化过程。

在 HMM 中,从一个状态产生的观察序列 $Y = (Y(t), Y(t+1), \dots, Y(t+\tau-1))$, $t > 0$ 是一个平稳的离散时间随机过程,常假设观察是独立同分布的,因而有

$$P(Y = y | \lambda) = \prod_{k=t}^{t+\tau-1} P(Y(k) = y(k) | \lambda)$$

而通常的这种假设,又忽略了观察间的非独立性或相关性。因此,这种模型仅对平稳信号有用。而现实世界中的信号都是高度动态且以一种系统的方式连续变化的。用分段平稳信号去近似非平稳信号需要较多的状态,以及大量自由独立参数,这使得参数训练困难增加。正是由于这些缺点,标准 HMM 才经常受到指责。在为时间序列建模方面的大多数努力曾经涉及到引入观察之间的一种暂时的关系。最简单的特征相关模型是 Welleken 的直方图模型(参见文献[17]),在该模型中一个观察 Y_t 仅被允许依靠以前的一个特征 Y_{t-1} 。近来, Ghitza 和 Sondhi 建议在相应于一双因素的每一个非平稳状态中应用动态时包模板。这种模型能被认为是 HMM 和随机片断模型^[18,19]的结合,其中的动态时包替换了线性时包。以上模型均忽略了隐马氏链的非平稳性。

所谓状态逗留是指在转移到另一个状态之前,过程在当前状态停留的时间长度或大小。HMM 具有以下性质:

$$P(X(t) = s_i | X(1) = s_{i_1}, \dots, X(t-1) = s_{i_{t-1}}) = P(X(t) = s_i | X(t-1) = s_{i_{t-1}})$$

从该式看不出逗留的长度,而实际中的信号却允许任一状态逗留任意长时间。在这种标准 HMM 模型中,得到的逗留概率密度 $P_i(d)$ 有如下形式:

$$P_i(d) = a_{ii}^{d-1} \cdot (1 - a_{ii})$$

其中 $a_{ii} = P(X(t+1) = s_i | X(t) = s_i)$. 而 $P_i(d)$ 这种模型对很多物理信号都不合适. Bonke Sin 和 H. Kim 建议了一种时变 HMM, 即转概矩阵是时间的函数.

2.3.1 带逗留时间 HMM

Ferguson 在文献[20]中首先构想出明确的时间逗留模型, 其中得到带逗留时间 HMM 的过程序列的可通过如下步骤:

1. 初始状态 $X(1) = s_i$, 按照初始状态分布 π 来确定.
2. $X(1)$ 的逗留按照状态逗留密度 $P_i(d)$ 来确定. 在实际的操作中逗留时间 d 被允许有一个最大值 D .
3. 观察 $Y(1), Y(2), \dots, Y(d_1)$ 按照联合观察密度 $b_i(y(1), y(2), \dots, y(d_1))$ 被选择. 假设观察独立, 则上式等于 $\prod_{k=1}^{d_1} b_i(y(k))$.
4. 下一个状态 $X(2) = s_j$, 按照状态转移概率 a_{ij} 来选择, 带有约束 $a_{ii} = 0$, 即不可以转回自身.
5. 重复 2—4.

这种模型在描述信号的动态本性方面表现不错, 然而, 在描述马氏链的非平稳性方面又不是很完全. 还有就是它总是将逗留模型建模为几何分布. 这一点不好, 脱离了实际.

注: $b_i(y(1), y(2), \dots, y(d_1))$ 是表示在隐状态为 s_i 的情况下, 观察序列 $Y(1), Y(2), \dots, Y(d_1)$ 的联合分布密度. $b_i(y(k))$ 是表示在隐状态为 s_i 的情况下, 观察 $Y(k)$ 的分布密度.

2.3.2 非平稳 HMM

$$\text{在标准 HMM 中, } P(Y=y|\lambda) = \sum_i \sum_j \alpha_i(i) a_{ij} b_j(y(t+1)) \beta_{i+1}(j) \quad (2.3.1)$$

其中 $Y = Y(1), Y(2), \dots, Y(T)$ 为观察, λ 为模型参数. 而

$$\begin{aligned} \alpha_i(j) &= \sum_i \alpha_{i-1}(i) \cdot a_{ij} \cdot b_j(y(t)) \\ \beta_i(j) &= \sum_k a_{jk} \cdot b_k(y(t+1)) \cdot \beta_{i+1}(k) \end{aligned} \quad (2.3.2)$$

分别为前后向概率.

相似地, 在逗留 HMM (DHMM) 中

$$P(Y=y|\lambda) = \sum_j \sum_{i \neq j} \sum_{\tau} \alpha_{i-\tau}(i) \cdot a_{ij} \cdot d_j(\tau) \cdot \prod_{l=1}^{\tau} b_j(y(t-\tau+1)) \cdot \beta_i(j), 1 \leq \tau \leq D \quad (2.3.3)$$

其中 $d_j(\tau)$ 表示隐马氏链在状态 s_j 逗留时间的概率密度函数。而

$$\begin{aligned} \alpha_i(j) &= \sum_{i \neq j} \sum_{\tau} \alpha_{i-\tau}(i) \cdot a_{ij} \cdot d_j(\tau) \cdot \prod_{l=1}^{\tau} b_j(y(t-\tau+1)) \\ \beta_i(j) &= \sum_{k \neq j} \sum_{\tau} a_{jk} \cdot d_k(\tau) \cdot \prod_{l=1}^{\tau} b_k(y(t+1)) \cdot \beta_{i+1}(k) \end{aligned} \quad (2.3.4)$$

分别为非平稳 HMM 中的前后向概率。在标准模型中, 转移概率 $(a_{ij}), i \neq j$ 是时间的常数函数, 因而, 马氏链被描述为平稳。为了将过程描述得更加细致, 我们考虑马氏链的非平稳性, 即转移概率是时间的非常数函数。

定义 2.3.1^[20] $a_{ij}(\tau) \triangleq P(X(t)=s_j | X(t-1)=\dots X(t-\tau)=s_i)$

其中 τ 为状态 s_i 的逗留时间, 它可以有一个最大值 D 。

在非平稳 HMM 中, 转移过程记录了状态过程在某一状态逗留的时间。而转移到的下一个状态则依赖于上一个状态的逗留时间。下面的图显示了隐马氏链的一部分, 它为从 $t-\delta-\tau+1$ 时刻起, 状态为 s_i , 并一直到 $t-\tau$ 时刻, 然后 $t-\tau+1$ 转到另外一个状态 s_j , 并逗留 τ 个时间单位直到时刻 t 。

下面给出非平稳 HMM 中非平稳马氏链的产生过程:

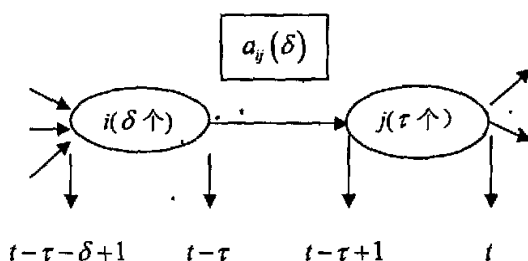


图 2.3.1 非平稳隐马氏链的一部分

1. 在时刻 $t-\tau-\delta+1$, 过程进入或开始于状态 s_i 。

2. 对于接下来的 δ 个时间单位, 按时变概率分布 $a_{ij}(d)$ 来选择, 过程作了 $\delta-1$ 个自转移 $a_{ii}(d), d=1, \dots, \delta-1$ 。如架设相应的观察是独立的, 则有

$$b_i(y(t-\tau-\delta+1)) = \prod_{d=t-\delta-\tau+1}^{t-\tau} b_i(y(d)).$$

3. 根据 $a_k(\delta)$ 找下一个状态 s_j .

4. 同 2.

下面介绍上述三种 HMM 的关系:

在非平稳 HMM 中, 转移概率满足

$$\sum_{j=1}^N a_{ij}(\tau) = 1, i=1, \dots, N. \quad \tau=1, \dots, D$$

当 $D=1$ 或 $a_{ij}(\tau)$ 为常数时, 模型就变为标准 HMM. 当自转移概率为 0 时, 模型就变为状态逗留 HMM.

2.3.3 参数估计

将前后向概率的定义进行推广, 如下

$$\alpha_{t,\tau}(i) = P(Y(1)=y(1), \dots, Y(t)=y(t), X(u)=s_i, t-\tau+1 \leq u \leq t | \lambda)$$

$$\beta_{t,\tau}(i) = P(Y(t+1)=y(t+1), \dots, Y(T)=y(T), X(u)=s_i, t-\tau+1 \leq u \leq t | \lambda)$$

那么通过归纳推理可得

$$\alpha_{t,\tau}(j) = \sum_{i=1, i \neq j}^N \sum_{\delta=1}^{t-1} \alpha_{t-\tau,\delta}(i) \cdot a_{ij}(\delta) \cdot \prod_{k=1}^{\tau-1} a_{ij}(k) \cdot \prod_{l=1}^{\tau} b_j(y(t-\tau+1)) \quad (2.3.5)$$

其中 δ 和 τ 分别是在状态 s_i 和 s_j 逗留的时间长度。

初始化

$$\alpha_{1,1}(j) = \pi_j \cdot b_j(y(1))$$

$$\alpha_{2,1}(j) = \sum_{i \neq j} \alpha_{1,1}(i) \cdot a_{ij}(1) \cdot b_j(y(2))$$

$$\alpha_{2,2}(j) = \pi_j \cdot a_{jj}(1) \cdot \prod_{l=1}^2 b_j(y(3-l))$$

$$\alpha_{3,1}(j) = \sum_{i=1, i \neq j}^N \sum_{\delta=1}^2 \alpha_{2,\delta}(i) \cdot a_{ij}(\delta) \cdot \prod_{l=1}^1 b_j(y(4-l))$$

(2.3.6)

以前的前向概率和现在的前向概率有如下关系

$$\alpha_t(i) = \sum_{\tau} \alpha_{t,\tau}(i) \quad (2.3.7)$$

因此

$$P(Y=y|\lambda) = \sum_{i=1}^N \sum_{\tau} \alpha_{T,\tau}(i) = \sum_{i=1}^{N'} \alpha_{\tau}(i) \quad (2.3.8)$$

同样经过归纳推理, 有

$$\beta_{i,\delta}(i) = \sum_{j \neq i} \sum_{\tau} a_{ij}(\delta) \cdot \prod_{k=1}^{\tau-1} a_{ij}(k) \cdot \prod_{l=1}^{\tau} b_j(y(t+l)) \cdot \beta_{i,\tau\tau}(j) \quad (2.3.9)$$

且同传统的后向概率有如下关系:

$$\beta_i(i) = \sum_{\tau} \beta_{i,\tau}(i) \quad (2.3.10)$$

此时

$$P(Y=y|\lambda) = \sum_j \sum_{i \neq j} \sum_{\tau} \sum_{\delta} \alpha_{i-\tau,\delta}(i) \cdot a_{ij}(\tau) \cdot \prod_{k=1}^{\tau-1} a_{ij}(k) \cdot \prod_{l=1}^{\tau} b_j(y(t-\tau+l)) \cdot \beta_{i,\tau}(j) \quad (2.3.11)$$

由(2.3.6)和图 2.3.1 可得

$$\tau > 1 \text{ 时, } \alpha_{i,\tau}(j) = \alpha_{i-1,\tau-1}(j) \cdot a_{ij}(\tau-1) \cdot b_j(y(t)) \quad (2.3.12)$$

$$\tau = 1 \text{ 时, } \alpha_{i,\tau}(j) = \left[\sum_{i \neq j} \sum_{\delta} \alpha_{i-1,\delta}(i) \cdot a_{ij}(\delta) \right] b_j(y(t)) \quad (2.3.13)$$

关于 λ 最大化(2.3.11) 其实是一个约束优化问题, 对下面的式子运用拉格朗日乘数法

$$Q = \log P(Y=y|\lambda) + k_0 \left(\sum_i \pi_i - 1 \right) + \sum_{i=1}^N k_i \cdot \left(\sum_j a_{ij}(\delta) - 1 \right) + \sum_j l_j \cdot \left(\sum_k b_j(k) - 1 \right) \quad (2.3.14)$$

就可得到如下结果。

定理 2.3.1 非平稳 HMM 的参数估计为

$$\widehat{a_{ij}(d)} = \frac{\sum_{l=1}^{T-1} \sum_{i \neq j} \sum_{\tau > d} \sum_{\delta} \alpha_{i-\tau,\delta}(i) a_{ij}(\delta) \prod_{k=1}^{\tau-1} a_{ij}(k) \prod_{l=1}^{\tau} b_j(y(t-\tau+l)) \beta_{i,\tau}(j)}{\sum_{n=1}^{T-1} \sum_{i \neq n} \sum_{\tau > d} \sum_{\delta} \alpha_{i-\tau,\delta}(i) a_{in}(\delta) \prod_{k=1}^{\tau-1} a_{in}(k) \prod_{l=1}^{\tau} b_n(y(t-\tau+l)) \beta_{i,\tau}(n)} \quad (2.3.15)$$

$$\widehat{b_j(k)} = \frac{\sum_{l=1}^{T-1} \sum_{i \neq j} \sum_{\delta, \tau} \alpha_{i-\tau,\delta}(i) a_{ij}(\delta) \prod_{d=1}^{\tau-1} a_{ij}(d) c_{i-\tau+1,l}(k) \sum_{l=1}^{\tau} b_j(y(t-\tau+l)) \beta_{i,\tau}(j)}{\sum_{m=1}^{T-1} \sum_{i \neq m} \sum_{\delta, \tau} \alpha_{i-\tau,\delta}(i) a_{ij}(\delta) \prod_{d=1}^{\tau-1} a_{ij}(d) c_{i-\tau+1,l}(m) \sum_{l=1}^{\tau} b_j(y(t-\tau+l)) \beta_{i,\tau}(j)} \quad (2.3.16)$$

$$\widehat{\pi_i} = \frac{\sum_l \pi_i \prod_{k=1}^{\tau-1} a_{ii}(k) \prod_{l=1}^{\tau} b_i(y(l)) \beta_{i,\tau}(i)}{p} \quad (2.3.17)$$

其中 $p = P(Y = y|\lambda)$, $c_{t-\tau+1,t}(k)$ 为观察状态 $Y(k) = v_k$ 在观察序列 $Y(t-\tau+1), \dots, Y(t)$ 中出现的频率。

本章小结

本章从带约束优化角度出发, 结合一种比较机制, 对隐马氏模型的参数进行了训练。它相对单独使用遗传算法是较好的, 但它也有不足之处。不足之处就在于参数的设定需要特别的谨慎, 否则就达不到预期的效果。再者, 初始化对结果也有影响。但是, 由试验可知它却具有比较的优势。尽管如此, 我们的结果也还是不理想, 很多地方有待改进。这实际上与我们研究的问题有关, 我们研究的问题本质上属于带不等式约束的优化问题, 目前来讲, 没有一种方法可以解决这种问题, 我们所能做的是利用最新的方法解决我们的问题。

本章还从估计角度出发, 就一类带有连续时间指标的特殊隐马氏模型讨论了其参数的最大后验估计。但这种估计也有不足之处。首先, Riccati 方程的解用一般的方法很难求得; 再者, 参数估计序列仅仅收敛到对数后验概率的稳定点, 从而其极限有可能是次优的。

第三章 HMM 的几种新的识别方法及 HMM 的相关应用

在许多实际问题当中, 识别问题或者说解码问题才是最终的目的。参数训练也是为解码服务的。本章主要讨论隐马氏模型的状态识别问题, 创新有以下两点:

一、由于传统的 Viterbi 算法所处理的模型相对比较单一, 故我们考虑了基于滤波和内插的隐马氏链的状态识别算法, 这种技术所处理的对象比较广泛。我们针对非时齐带逗留时间隐马氏模型, 运用滤波和内插的方法从近似角度给出了一种状态识别方法。其中的仿真试验结果说明该方法还是可以的。

二、文献[21], [22]就 Wiener 和 Kalman 滤波的性能进行了比较, 但由于理论上误差公式的缺乏, 故以前很少对滤波和平滑的性能进行比较, 直到文献[23]提供了一种误差渐近公式, 接着便迎来了滤波和平滑的误差分析的高潮。在这种背景下, 本文在给出了运用滤波和内插方法进行识别的方法后, 还在一类特殊模型下对这种方法的误差进行了细致的分析。

§3.1 Viterbi 解码

3.1.1 标准 HMM 的 Viterbi 解码问题

给定一个观测序列 $Y = (Y(1), Y(2), \dots, Y(T))$ 和一个 HMM 的参数 λ , 我们往往需要知道观测链所对应的隐状态序列是什么, 这一问题的回答取型状态物理意义的理解, 对于同一个观测序列 $(Y(1), Y(2), \dots, Y(T))$, 根据不同的优化原则, 可以得到不同的状态决于对模序列。

记 $y = (y(1), \dots, y(T))$, 令 $\gamma_k(i) = P(X(k) = s_i | Y = y, \lambda)$, 按 HMM 的意义, 显然有如下公式

$$\gamma_k(i) = \frac{P(Y = y, X(k) = s_i | \lambda)}{\sum_{j=1}^N P(Y = y, X(k) = s_j | \lambda)} = \frac{\alpha_k(i)\beta_k(i)}{\sum_{j=1}^N \alpha_k(j)\beta_k(j)}$$

以上 $\gamma_k(i)$ 的具体推导请参考文献[24] [25] [26] [27], 这里就不再给出具体的推导公式了。

下面我们给出两种优化原则。

(1) 原则 1 单点状态最优原则。即在每一时刻, 求最可能的状态: 如果 i^* 满足 $\gamma_k(i^*) = \max_i \gamma_k(i)$, 则取 $X(n) = s_{i^*}$. 单点状态最优的长处在于简单易算。但是它存在显见的缺点, 由于它没有考虑到前后两个状态的制约关系, 因此得到的状态序列完全有可能不存在。基于这种原因, 考虑路径最优的原则就是很自然的了。

(2) 原则 2 路径(轨道)最优原则。这是基于整体性考虑的方法。令

$$\delta_k(i) = \max_{i_1, \dots, i_{k-1}} P(X(k) = s_i, X(1) = s_{i_1}, \dots, X(k-1) = s_{i_{k-1}}, Y(1) = y(1), \dots, Y(k) = y(k) | \lambda)$$

则

$$\delta_{k+1}(i) = b_i(y(k+1)) \max_j (\delta_k(j) a_{ji}).$$

它的实际计算程序可以用如下的 Viterbi 算法。

Viterbi 算法:

(1) 初始化

$$\delta_1(i) = \pi_i b_i(y(1))$$

$$\psi_1(i) = 0$$

(2) 递推

$$\delta_k(j) = b_j(y(k)) \max_i (\delta_{k-1}(i) a_{ij})$$

$$\psi_k(j) = \arg \max_i (\delta_{k-1}(i) a_{ij})$$

(3) 结束

$$p^* = \max_i (\delta_T(i))$$

$$X_T^* = \arg \max_i (\delta_T(i))$$

(4) 路径回溯(即最佳状态链的确定)

$$X_k^* = \psi_{k+1}(X_{k+1}^*)$$

Viterbi 算法不仅可以对观测序列 Y 确定一个最佳状态链, 而且也可以近似地求出模型 λ 产生观测序列 Y 的概率, 即 $P \approx P(Y = y | \lambda)^{[2]}$, 虽然 Viterbi 算法每一步都只选择了一条最优路径, 但最终概率得分计算的近似程度也相当好。

3.1.2 非平稳 HMM 的 Viterbi 解码问题

非平稳 HMM 的状态识别问题可视为非平稳 HMM 中的前后向算法的一种特殊情况。此种

HMM 的 Viterbi 解码算法, 其实就是关于 X 最大化 $P(Y=y, X|\lambda)$. 具体方法可见文献[16].

§3.2 基于滤波的方法

状态识别问题是隐马氏模型理论的主要问题之一, 而常规隐马氏模型在描述状态的逗留时, 始终将逗留模型建模为几何分布^[5], 这在某些应用领域, 如语音识别, 说话人识别等是不合适的. 为克服这种不合理性以及考虑到广泛的应用背景, 我们将考虑更一般的隐马氏模型, 即非时齐带逗留时间隐马氏模型 (有些文献又称带逗留时间隐马氏模型为非平稳隐马氏模型) 的状态识别问题.

本文在这种模型参数已知的假设下, 通过基于内插的识别技术给出了一种状态识别方法. 需要注意的是上面所提到的模型参数可能因具体背景的不同而具有不同的结构, 但最终的决定参数只有三种, 那就是初始状态分布, 转移概率以及特征概率.

为了在每次状态估计中都用到全部观察的信息, 我们将根据内插这一状态估计进行识别而不是根据滤波.

将该技术运用于简单或常规隐马氏模型的例子可参见文献[9], 其中的仿真试验充分说明了该技术的相对优越性.

给定概率空间 (Ω, F, P) , 考虑定义于其上的非时齐带逗留时间隐马氏模型 (X, Y) , 其中 $X = (X(1), \dots, X(T))$ 为隐马氏模型状态序列, $Y = (Y(1), \dots, Y(T))$ 为观察序列, T 为链的长度. 状态空间设为 $\{s_1, s_2, \dots, s_N\}$, 观察空间设为 $\{\gamma_1, \gamma_2, \dots, \gamma_M\}$. 模型参数为初始状态分布 π , 转移概率矩阵族 $\{A(k, d), 1 \leq k \leq T-1, 1 \leq d \leq D\}$, 其中 d 为逗留时间, D 为最大逗留时间, 而

$$A(k, d) = (a_{ij}(k, d)), \quad 1 \leq i, j \leq N \quad (3.2.1)$$

$$a_{ij}(k, d) = P(X(k+1) = s_j | X(k) = s_i, X(k-1) = \dots = X(k-d+1) = s_i) \quad (3.2.2)$$

且满足 $\sum_{j=1}^N a_{ij}(k, d) = 1, 1 \leq k \leq T-1, 1 \leq d \leq D$. 还有特征概率矩阵族 $\{Q(k), 1 \leq k \leq T\}$, 其中

$$Q(k) = (q_{ij}(k)), \quad 1 \leq i \leq M, 1 \leq j \leq N$$

$$q_{ij}(k) = P(Y(k) = \gamma_j | X(k) = s_i), \quad 1 \leq i \leq M, 1 \leq j \leq N \quad (3.2.3)$$

鉴于识别的需要, 假设模型参数已知.

一般我们还假定

$$E(Y(k)|X(1), \dots, X(k), Y(1), \dots, Y(k-1)) = E(Y(k)|X(k))$$

$$E(X(k+1)|X(1), \dots, X(k), Y(1), \dots, Y(k)) = E(X(k+1)|X(k), d(X(k))).$$

不难计算条件期望

$$E(Y(k)|X(k)) = H(k)^T X(k) \quad (3.2.4)$$

$$E(X(k+1)|X(k), d(X(k))) = A(k, d(X(k)))^T X(k), \quad k \geq 1 \quad (3.2.5)$$

其中 $H(k)$ 是从 $Q(k)$ 中去掉最后一行得到的, 从而 $H(k)$ 是一个 $(M-1) \times N$ 矩阵. (3.2.4), (3.2.5) 将在后面用到.

为了处理问题的方便, 我们就不妨假设状态空间为 $\{u_1, u_2, \dots, u_N\}$, 其中

$$u_i^T = (0, \dots, 0, 1, 0, \dots, 0),$$

1 为第 i 个分量, 其余分量全为 0. 即 u_1, \dots, u_N 为 R^N 的一组标准正交基. 否则将状态空间 $\{s_1, s_2, \dots, s_N\}$ 一一对应到集合 $\{u_1, u_2, \dots, u_N\}$, 方式为 s_i 对应 $u_i, i=1, \dots, N$. 同样地, 将 $\{\gamma_1, \gamma_2, \dots, \gamma_M\}$ 一一对应到 $M-1$ 维向量的集合 $\{d_1, d_2, \dots, d_M\}$, 其中

$$d_1^T = (1, 0, \dots, 0), d_2^T = (0, 1, 0, \dots, 0), \dots, d_{M-1}^T = (0, \dots, 0, 1), d_M^T = (0, \dots, 0).$$

此时, 观察过程就成了 $M-1$ 维离散点过程.

为了后面讨论的需要, 我们给出几个定义和符号. 需要说明的是, 以后我们所讨论问题若无特别声明均基于给定的概率空间 $\{\Omega, F, P\}$. 设 $\{F_i, i=1, 2, \dots\}$ 是 F 的一族单调非降子 σ -代数.

定义 3.2.1 称序列 $X = \{X(1), X(2), \dots\}$ 为关于 $\{F_i, i=1, 2, \dots\}$ 的鞅差序列. 若序列 $X = \{X(1), X(2), \dots\}$ 关于 $\{F_i, i=1, 2, \dots\}$ 是适应的, 且对 $\forall i, E(X(i)|F_{i-1}) = 0$.

定义 3.2.2 设 $\{Z(i), i=1, 2, \dots\}, \{V(i), i=1, 2, \dots\}$ 为关于 $\{F_i, i=1, 2, \dots\}$ 的 m 维鞅差序列 ($m \geq 1$), 定义条件方差矩阵序列 (Z, Z) 和条件协方差矩阵序列 (Z, V) 为:

$$(Z, Z)(i) = E(Z(i)Z^T(i)|F_{i-1})$$

$$(Z, V)(i) = E(Z(i)V^T(i)|F_{i-1}).$$

记

$$W(k+1) = X(k+1) - E(X(k+1)|X(1), \dots, X(k), Y(1), \dots, Y(k)) \quad (3.2.6)$$

$$V(k) = Y(k) - E(Y(k)|X(1), \dots, X(k), Y(1), \dots, Y(k-1)) \quad (3.2.7)$$

由(3.2.5), (3.2.6)可得

$$X(k+1) = A(k, d(X(k)))^T X(k) + W(k+1) \quad (3.2.8)$$

由(3.2.4), (3.2.7)得

$$Y(k) = H(k)^T X(k) + V(k) \quad (3.2.9)$$

令

$$\hat{X}(k|k) = E(X(k)|Y(1), Y(2), \dots, Y(k)) \quad (3.2.10)$$

则 $\hat{X}(k|k)$ 为 $X(k)$ 的滤波。

令

$$\begin{aligned} U(k) &= E(X(k)|Y(1), \dots, Y(k)) - E(X(k)|Y(1), \dots, Y(k-1)) \\ &= \hat{X}(k|k) - \hat{X}(k|k-1), \end{aligned} \quad (3.2.11)$$

$$L(k) = Y(k) - E(Y(k)|Y(1), \dots, Y(k-1)) = Y(k) - \hat{Y}(k|k-1), \quad k \geq 1. \quad (3.2.12)$$

其中 $E(Y(k)|Y(1), \dots, Y(k-1)) \triangleq \hat{Y}(k|k-1)$, $k \geq 2$. 由鞅差序列定义, 不难验证 $\{U(k), k=1, 2, \dots, T\}$ 和 $\{L(k), k=1, 2, \dots, T\}$ 都是关于 $\{\sigma(Y(1), Y(2), \dots, Y(k)), k=1, 2, \dots, T\}$ 的鞅差序列。

下面给出一个引理。

引理 3.2.1^[9] 每一个关于 σ -代数序列 $\{\sigma(Y(1), \dots, Y(k)), k=1, 2, \dots\}$ 为鞅差序列的离散时间随机过程 $\{J(k), k=1, 2, \dots\}$, 都可表示为

$$J(k) = G(k; Y(1), \dots, Y(k))L(k) \quad (3.2.13)$$

其中 $G(k; Y(1), \dots, Y(k))$ 是关于 $\{\sigma(Y(1), \dots, Y(k)), k=2, 3, \dots\}$ 可预报的, 即 $G(k; Y(1), \dots, Y(k))$ 是关于 $\sigma(Y(1), \dots, Y(k-1)), k=2, 3, \dots$ 可测的。而

$$L(k) = Y(k) - \hat{Y}(k|k-1).$$

由该引理不难得到滤波的表达式:

定理 3.2.1 滤波

$$\hat{X}(k|k) = \hat{X}(k|k-1) + \left\{ \text{diag} \left\{ \hat{X}(k|k-1) \right\} H(k) - \hat{X}(k|k-1) \hat{Y}^T(k|k-1) \right\} \\ \left(1 - \sum_{j=1}^{M-1} \hat{Y}_j(k|k-1) \right)^{-1} C(k) (Y(k) - \hat{Y}(k|k-1))$$

其中 $\text{diag} \{ \hat{X}(k|k-1) \}$ 表示以向量 $\hat{X}(k|k-1)$ 的分量为主对角线上元素的对角矩阵, $C(k)$ 为一个主对角元为 $\beta_{ii}(k) = \left(\hat{Y}_i(k|k-1) \right)^{-1} \left(1 - \sum_{j \neq i} \hat{Y}_j(k|k-1) \right)$, $1 \leq i \leq M-1, k \geq 1$, 其余元素都为 1 的 $(M-1) \times (M-1)$ 矩阵。

证明 由引理不难得到

$$U(k) = G(k; Y(1), \dots, Y(k-1)) L(k) \quad (3.2.14)$$

由 (3.2.14) 可得

$$\begin{aligned} (U, L)(k) &= E(U(k) L^T(k) | Y(1), \dots, Y(k-1)) \\ &= E(G(k; Y(1), \dots, Y(k-1)) L(k) L^T(k) | Y(1), \dots, Y(k-1)) \\ &= G(k; Y(1), \dots, Y(k-1)) (L, L)(k) \end{aligned}$$

从而

$$G(k; Y(1), \dots, Y(k-1)) = (U, L)(k) [(L, L)(k)]^{-1} \quad (3.2.15)$$

而

$$\begin{aligned} (U, L)(k) &= E(U(k) L^T(k) | Y(1), \dots, Y(k-1)) \\ (L, L)(k) &= E(L(k) L^T(k) | Y(1), \dots, Y(k-1)). \end{aligned} \quad (3.2.16)$$

下面考虑条件协方差阵 $(U, L)(k)$ 和条件方差阵 $(L, L)(k)$ 的位于第 i 行第 j 列的元素。

$$\begin{aligned} &E(U_i(k) L_j(k) | Y(1), \dots, Y(k-1)) \\ &= E\left(\left(\hat{X}_i(k|k) - \hat{X}_i(k|k-1)\right)\left(Y_j(k) - \hat{Y}_j(k|k-1)\right) | Y(1), \dots, Y(k-1)\right) \\ &= E\left(\begin{aligned} &\hat{X}_i(k|k) Y_j(k) - \hat{X}_i(k|k) \hat{Y}_j(k|k-1) - \hat{X}_i(k|k-1) Y_j(k) \\ &+ \hat{X}_i(k|k-1) \hat{Y}_j(k|k-1) \end{aligned} | Y(1), \dots, Y(k-1)\right) \\ &= E\left(\hat{X}_i(k|k) Y_j(k) | Y(1), \dots, Y(k-1)\right) - \hat{X}_i(k|k-1) \hat{Y}_j(k|k-1) \\ &\quad - \hat{X}_i(k|k-1) E(Y_j(k) | Y(1), \dots, Y(k-1)) + \hat{X}_i(k|k-1) \hat{Y}_j(k|k-1) \end{aligned}$$

$$\begin{aligned}
&= E\left(E\left(X_i(k)Y_j(k)|Y(1),\dots,Y(k)\right)|Y(1),\dots,Y(k-1)\right)-\widehat{X}_i(k|k-1)\widehat{Y}_j(k|k-1) \\
&= E\left(X_i(k)Y_j(k)|Y(1),\dots,Y(k-1)\right)-\widehat{X}_i(k|k-1)\widehat{Y}_j(k|k-1) \\
&= E\left(X_i(k)\left(\left(H(k)^T X(k)\right)_j+V_j(k)\right)|Y(1),\dots,Y(k-1)\right)-\widehat{X}_i(k|k-1)\widehat{Y}_j(k|k-1) \\
&= E\left(X_i(k)\left(q_{ji}(k)X_1(k)+\dots+q_{jN}(k)X_N(k)+V_j(k)\right)|Y(1),\dots,Y(k-1)\right) \\
&\quad -\widehat{X}_i(k|k-1)\widehat{Y}_j(k|k-1)
\end{aligned}$$

(由于 $X(k)$ 的取值为 R^N 的标准正交基中的某一个, 故 $X_i(k)$ 取值要么为 0, 要么为 1, 从而有: $\forall k, \forall i, \forall j, X_i^2(k)=X_i(k)$, 而 $X_i(k)X_j(k)=0, i \neq j$)

$$\begin{aligned}
&= E\left(q_{ji}(k)X_i(k)+X_i(k)V_j(k)|Y(1),\dots,Y(k-1)\right)-\widehat{X}_i(k|k-1)\widehat{Y}_j(k|k-1) \\
&= E\left(q_{ji}(k)X_i(k)|Y(1),\dots,Y(k-1)\right)+ \\
&\quad E\left(X_i(k)\left(Y_j(k)-E\left(Y_j(k)|X(1),\dots,X(k),Y(1),\dots,Y(k-1)\right)\right)|Y(1),\dots,Y(k-1)\right) \\
&\quad -\widehat{X}_i(k|k-1)\widehat{Y}_j(k|k-1) \\
&= q_{ji}(k)\widehat{X}_i(k|k-1)+E\left(X_i(k)Y_j(k)|Y(1),\dots,Y(k-1)\right)- \\
&\quad E\left(E\left(X_i(k)Y_j(k)|X(1),\dots,X(k),Y(1),\dots,Y(k-1)\right)|Y(1),\dots,Y(k-1)\right) \\
&\quad -\widehat{X}_i(k|k-1)\widehat{Y}_j(k|k-1) \\
&= q_{ji}(k)\widehat{X}_i(k|k-1)+E\left(X_i(k)Y_j(k)|Y(1),\dots,Y(k-1)\right)- \\
&\quad E\left(X_i(k)Y_j(k)|Y(1),\dots,Y(k-1)\right)-\widehat{X}_i(k|k-1)\widehat{Y}_j(k|k-1) \\
&= q_{ji}(k)\widehat{X}_i(k|k-1)-\widehat{X}_i(k|k-1)\widehat{Y}_j(k|k-1)
\end{aligned}$$

从而

$$E\left(U(k)L^T(k)|Y(1),\dots,Y(k-1)\right)=\text{diag}\left\{\widehat{X}(k|k-1)\right\}H(k)-\widehat{X}(k|k-1)\widehat{Y}^T(k|k-1) \quad (3.2.17)$$

由 (3.2.12)

$$\begin{aligned}
&E\left(L_i(k)L_j(k)|Y(1),\dots,Y(k-1)\right) \\
&= E\left(\left(Y_i(k)-\widehat{Y}_i(k|k-1)\right)\left(Y_j(k)-\widehat{Y}_j(k|k-1)\right)|Y(1),\dots,Y(k-1)\right) \\
&= E\left\{\begin{aligned} &Y_i(k)Y_j(k)-Y_i(k)\widehat{Y}_j(k|k-1)-\widehat{Y}_i(k|k-1)Y_j(k)+ \\ &\widehat{Y}_i(k|k-1)\widehat{Y}_j(k|k-1) \end{aligned} \middle| Y(1),\dots,Y(k-1)\right\} \quad (3.2.18)
\end{aligned}$$

由于 $Y(k) \in \{d_1, \dots, d_M\}$, 故 $Y_i(k)=Y_i^2(k)$, $Y_i(k)Y_j(k)=0, i \neq j$, 故

$$\begin{aligned}
 (3.2.18) &= E(Y_i(k)Y_j(k)|Y(1), \dots, Y(k-1)) - \hat{Y}_i(k|k-1)\hat{Y}_j(k|k-1) \\
 &= \begin{cases} -\hat{Y}_i(k|k-1)\hat{Y}_j(k|k-1), i \neq j \\ \hat{Y}_i(k|k-1) - \hat{Y}_i^2(k|k-1), i = j \end{cases} \quad (3.2.19)
 \end{aligned}$$

从而 $(L, L)_k = (E(L_i(k)L_j(k)|Y(1), \dots, Y(k-1)))$ 已知。

容易验证

$$\begin{aligned}
 (L, L)_k^{-1} &= [E(L(k)L^T(k)|Y(1), \dots, Y(k-1))]^{-1} \\
 &= \left(1 - \sum_{j=1}^{M-1} \hat{Y}_j(k|k-1)\right)^{-1} C(k) \quad (3.2.20)
 \end{aligned}$$

将 (3.2.17), (3.2.20) 代入 (3.2.15) 就可以得到 $G(k; Y_1, \dots, Y_{k-1})$ 的基于已知信息的具体表达式。再由 (3.2.14) 就可以得到

$$\begin{aligned}
 \hat{X}(k|k) &= \hat{X}(k|k-1) + \left\{ \text{diag}\{\hat{X}(k|k-1)\} H(k) - \hat{X}(k|k-1) \hat{Y}^T(k|k-1) \right\} \\
 &\quad \left(1 - \sum_{j=1}^{M-1} \hat{Y}_j(k|k-1)\right)^{-1} C(k) (Y(k) - \hat{Y}(k|k-1)) \quad (3.2.21)
 \end{aligned}$$

易见

$$\hat{X}(1|0) = E(X(1)) = \pi \quad (3.2.22)$$

$$\hat{Y}(k|k-1) = E(Y(k)|Y(1), \dots, Y(k-1)) = H^T(k) \hat{X}(k|k-1), k \geq 1 \quad (3.2.23)$$

$$\begin{aligned}
 \hat{X}(k|k-1) &= E(X(k)|Y(1), \dots, Y(k-1)) \\
 &= E\left(A(k-1, d(X(k-1)))^T X(k-1) | Y(1), \dots, Y(k-1)\right) \\
 &\approx E\left(A(k-1, d(X_{MAP}(k-1|k-1)))^T X(k-1) | Y(1), \dots, Y(k-1)\right) \\
 &= A(k-1, d(X_{MAP}(k-1|k-1)))^T \hat{X}(k-1|k-1), k \geq 2 \quad (3.2.24)
 \end{aligned}$$

其中 $X_{MAP}(k|k)$ 是基于 k 以前的信息运用最大后验概率原则对 $X(k)$ 做出的识别。由上面的递推关系可知 $d(X_{MAP}(k-1|k-1))$ 是已知的，从而 (3.2.24) 是正确的。

由 (3.2.17), (3.2.20), (3.2.21), (3.2.22), (3.2.23), (3.2.24) 可算出近似滤波。

内插 $\hat{X}(k|k)$ 事实上是 $X(k)$ 的最小均方误差估计。我们根据它对状态序列作出识别。

若 $\hat{X}_i(k|k) = \max\{\hat{X}_j(k|k), j=1, \dots, N\}$, 则判断 $X(k)$ 的状态为 u_i . 若有多个 i 满足上式, 则 $X(k)$ 可在对应的 u_i 中任取一个。以上做法的理论依据是最大后验概率原则, 具体地, 由于 $\{u_1, \dots, u_N\}$ 为 R^N 的标准正交基, 故 $X(k)$ 关于 $\sigma(Y(1), \dots, Y(k))$ 的条件期望其实就是向量 (e_1, \dots, e_N) , 其中 $e_i = P(X(k) = u_i | Y(1), \dots, Y(k))$, 即

$$E(X(k) | Y(1), \dots, Y(k)) = \sum_{i=1}^N u_i P(X(k) = u_i | Y_1, \dots, Y_k) = (e_1, \dots, e_N)$$

易见, e_1, \dots, e_N 中哪一个大, $X(k)$ 就应该取哪一个 e_i 对应的 u_i .

基于滤波的状态识别的性能可通过如下的误差计算来反映:

$$\begin{aligned} S(k|k) &\triangleq E\left(\left(X(k) - \hat{X}(k|k)\right) \cdot \left(X(k) - \hat{X}(k|k)\right)^T \middle| Y(1), \dots, Y(k)\right) \\ &= \text{diag}\{\hat{X}(k|k)\} - \hat{X}(k|k) \cdot \hat{X}(k|k)^T, k \geq 1 \end{aligned}$$

我们考虑了一个齐次, 离散时间, 四状态, 带逗留时间马氏链。最大逗留时间为 3, 50 次观察, 3 次执行。初始状态概率向量, 转移概率矩阵, 特征概率矩阵分别为

$$\pi = [0.25 \quad 0.25 \quad 0.25 \quad 0.25]^T$$

$$A(1) = \begin{bmatrix} 1-v & 0 & v & 0 \\ v & 0 & 1-v & 0 \\ 0 & 1-v & 0 & v \\ 0 & v & 0 & 1-v \end{bmatrix},$$

$$A(2) = \begin{bmatrix} 0 & 1-v & v & 0 \\ 0 & v & 1-v & 0 \\ v & 0 & 0 & 1-v \\ 1-v & 0 & 0 & v \end{bmatrix},$$

$$A(3) = \begin{bmatrix} v & 0 & 0 & 1-v \\ 1-v & 0 & 0 & v \\ 0 & 1-v & v & 0 \\ 0 & v & 1-v & 0 \end{bmatrix},$$

$$Q = \begin{bmatrix} q \cdot q & q \cdot (1-p) & q \cdot (1-p) & (1-p) \cdot (1-p) \\ q \cdot (1-q) & q \cdot p & (1-p) \cdot (1-q) & p \cdot (1-p) \\ q \cdot (1-q) & (1-p) \cdot (1-q) & q \cdot p & p \cdot (1-p) \\ (1-q) \cdot (1-q) & p \cdot (1-q) & p \cdot (1-q) & p \cdot p \end{bmatrix}$$

得到了如下试验结果:

	$p=q=0.7$	$p=q=0.95$	$p=q=0.99$	$p=q=0.995$
MAP filtering	6.0E-1	2.8E-1	2.6E-1	2.4E-1

表 3.2.1 $v=0.5$, 基于近似滤波的错误率

	$p=q=0.7$	$p=q=0.9$	$p=q=0.99$
MAP filtering	6.6E-1	2.8E-1	2.7E-1

表 3.2.2 $v=0.4$, 基于近似滤波的错误率

	$p=q=0.7$	$p=q=0.99$	$p=q=0.999$
MAP filtering	6.2E-1	2.7E-1	2.0E-1

表 3.2.3 $v=0.2$, 基于近似滤波的错误率

可以看得出, 该识别技术的效果随着 $A(\cdot)$ 和 Q 中参数的增大而提高, 且当其中的参数比较接近 0.5 时, 由于模糊性的增加而导致了识别效果较差。

由于常规隐马氏模型在状态逗留建模方面的不合理性限制了它的适用范围, 故对离散时间有限状态的非齐次非平稳隐马氏模型进行研究是比较有意义的。我们在模型参数和观察已知的前提下, 给出了运用滤波对隐马氏链的状态序列进行识别的方法, 从上下文可以看到, 我们综合运用了全部已知信息, 且计算不是很复杂, 但也看得出, 我们的结果不是很好, 方法有待进一步改进。

求滤波时仅仅使用了一部分观察信息, 而很多应用中, 为识别出状态序列, 常常还需要知道剩余的信息。如果每次估计都用上全部已知信息, 估计效果会更好。即基于内插作识别会更好一些, 很多文献都说明了这一点。

下面只给出内插的递推公式。

先给出几个符号:

$$\hat{X}(k|T) \triangleq E(X(k)|Y(1), \dots, Y(T))$$

$$O(m) \triangleq H(m) \left[E(L(m)L^T(m)|Y(1), \dots, Y(m-1)) \right]^{-1} L(m), m \geq 1$$

$$P(k;m) \triangleq E(X(k)X^T(m)|Y(1), \dots, Y(m-1)), m \geq k+2$$

$$F(k; m) \triangleq E\{X(k)X^T(m-1)|Y(1), \dots, Y(m-1)\}, m \geq k+2$$

$$\Lambda^{(i)}(k; m) \triangleq (0, \dots, 0, P_i(k; m-1), 0, \dots, 0)$$

其中 $P_i(k; m)$ 为矩阵 $P(k; m)$ 的第 i 列。

由文献[25]可类比地给出 $P(k; m)$ 的初始化:

$$P(k; k+1) = \text{diag}\{\hat{X}(k|k)\} A(k; d(X_{MAP}(k|k))) \quad (3.2.25)$$

下面以定理的形式给出内插的表达式。

$$\text{定理 3.2.2} \quad \hat{X}(k|T) = \hat{X}(k|k) + \sum_{m=k+1}^T \left\{ P(k; m) - \hat{X}(k|m-1) \hat{X}^T(m|m-1) \right\} O(m)$$

证明 令

$$\eta(s) = \hat{X}(k|s) - \hat{X}(k|s-1), k+1 \leq s \leq T \quad (3.2.26)$$

则 $\eta(s)$ 关于 $\sigma(Y(1), \dots, Y(s))$ 可测, 且 $\forall k+1 \leq s \leq T$

$$E(\eta(s)|Y(1), \dots, Y(s-1)) = \hat{X}(k|s-1) - \hat{X}(k|s-1) = 0$$

故 $\{\eta(s), k+1 \leq s \leq T\}$ 为关于 $\{\sigma(Y(1), \dots, Y(s)), k+1 \leq s \leq T\}$ 的鞅差序列。由上述引理得

$$\eta(s) = G(s)L(s),$$

其中 $G(s)$ 为 $G(s; Y(1), \dots, Y(s-1))$ 的简写。同上一部分相似地有

$$\begin{aligned} G(s) &= (\eta, L)(s)(L, L)^{-1}(s) \\ &= E(\eta(s)L^T(s)|Y(1), \dots, Y(s-1)) \left[E(L(s)L^T(s)|Y(1), \dots, Y(s-1)) \right]^{-1} \end{aligned}$$

所以

$$\eta(s) = E(\eta(s)L^T(s)|Y(1), \dots, Y(s-1)) \cdot \left[E(L(s)L^T(s)|Y(1), \dots, Y(s-1)) \right]^{-1} L(s) \quad (3.2.27)$$

由 (3.2.26)

$$\sum_{s=k+1}^T \eta(s) = \hat{X}(k|T) - \hat{X}(k|k)$$

故

$$\begin{aligned} \hat{X}(k|T) - \hat{X}(k|k) &= \sum_{s=k+1}^T \eta(s) = \sum_{s=k+1}^T G(s)L(s) = \sum_{s=k+1}^T E(\eta(s)L^T(s)|Y(1), \dots, Y(s-1)) \cdot \\ &\quad \left[E(L(s)L^T(s)|Y(1), \dots, Y(s-1)) \right]^{-1} \end{aligned} \quad (3.2.28)$$

下面计算一下上式右端的加数因子。首先, $\left[E(L(s)L^T(s)|Y(1), \dots, Y(s-1))\right]^{-1}$ 可以由 (3.2.20) 得到。故只需计算 $E(\eta(s)L^T(s)|Y(1), \dots, Y(s-1))$, 考虑它的位于第 i 行第 j 列的元素:

$$\begin{aligned}
 & E(\eta_i(s)L_j(s)|Y(1), \dots, Y(s-1)) \\
 &= E\left(\left(\hat{X}_i(k|s) - \hat{X}_i(k|s-1)\right)\left(Y_j(s) - \hat{Y}_j(s|s-1)\right)|Y(1), \dots, Y(s-1)\right) \\
 &= E\left(\hat{X}_i(k|s)Y_j(s) - \hat{X}_i(k|s)\hat{Y}_j(s|s-1)|Y(1), \dots, Y(s-1)\right) \\
 &= E\left(\hat{X}_i(k|s)Y_j(s)|Y(1), \dots, Y(s-1)\right) - \hat{X}_i(k|s-1)\hat{Y}_j(s|s-1) \\
 &= E\left(\hat{X}_i(k|s)\left(H(s)^T X(s) + V(s)\right)_j|Y(1), \dots, Y(s-1)\right) \\
 &\quad - \hat{X}_i(k|s-1)E(Y_j(s)|Y(1), \dots, Y(s-1)) \\
 &= E\left(\hat{X}_i(k|s)\left(H(s)^T X(s)\right)_j|Y(1), \dots, Y(s-1)\right) + \\
 &\quad E\left(\hat{X}_i(k|s)\left(Y_j(s) - E(Y_j(s)|X(1), \dots, X(s), Y(1), \dots, Y(s-1)))\right)_j\right) \\
 &\quad - \hat{X}_i(k|s-1)E\left(\left(H(s)^T X(s)\right)_j|Y(1), \dots, Y(s-1)\right) \\
 &= E\left(\hat{X}_i(k|s)\left(\sum_{h=1}^N q_{jh}(s)X_h(s)\right)_j|Y(1), \dots, Y(s-1)\right) + 0 \\
 &\quad - \hat{X}_i(k|s-1)\left[\sum_{h=1}^N q_{jh}(s)E(X_h(s)|Y(1), \dots, Y(s-1))\right] \\
 &= E\left(E\left(X_i(k)\sum_{h=1}^N q_{jh}(s)X_h(s)|Y(1), \dots, Y(s)\right)|Y(1), \dots, Y(s-1)\right) \\
 &\quad - \sum_{h=1}^N q_{jh}(s)E(X_i(s)X_h(s)|Y(1), \dots, Y(s-1)) \\
 &= E\left(\sum_{h=1}^N q_{jh}(s)X_h(s)X_i(k)|Y(1), \dots, Y(s-1)\right) - \\
 &\quad \sum_{h=1}^N q_{jh}(s)E(X_i(s)X_h(s)|Y(1), \dots, Y(s-1))
 \end{aligned}$$

故

$$E(\eta(s)L^T(s)|Y(1), \dots, Y(s-1)) = \begin{bmatrix} E(X(k)X^T(s)|Y(1), \dots, Y(s-1))H^T(s) - \\ \hat{X}(k|s-1)\hat{X}^T(s|s-1)H^T(s) \end{bmatrix}$$

从而

$$\begin{aligned}\widehat{X}(k|T) - \widehat{X}(k|k) &= \sum_{m=k+1}^T \left\{ P(k;m) - \widehat{X}(k|m-1) \widehat{X}^T(m|m-1) \right\} H(m) \cdot \\ &\quad \left[E(L(m)L^T(m)|Y(1), \dots, Y(m-1)) \right]^{-1} \cdot L(m) \\ &= \sum_{m=k+1}^T \left\{ P(k;m) - \widehat{X}(k|m-1) \widehat{X}^T(m|m-1) \right\} \cdot O(m)\end{aligned}$$

由 $P(k;m)$ 的定义易见, 将 $X(m) = A(m-1, d(X(m-1)))^T X(m-1) + W(m)$ 代入 $P(k;m)$ 的表达式, 与 (3.2.24) 同样地可得

$$P(k;m) \approx F(k;m) A(m-1, d(X_{MAP}(m-1|m-1)))^T, m \geq k+2, k \geq 1.$$

接下来, 我们证明一个等式:

定理 3.2.3

$$F_i(k;m) = P_i(k;m-1) + \left\{ \Lambda'(k;m) - P_i(k;m-1) \widehat{X}^T(m-1|m-2) \right\} O(m-1) \quad (3.2.29)$$

其中 $F_i(k;m), P_i(k;m-1)$ 分别为矩阵 $F(k;m)$ 和矩阵 $P(k;m-1)$ 的第 i 列。

证明 令

$$\eta(m-1) = F_i(k;m) - P_i(k;m-1)$$

易见 $\eta(m)$ 为关于 $\{\sigma(Y(1), \dots, Y(m)), m=1, 2, \dots\}$ 的鞅差序列。故由上述引理

$$\eta(m) = G(m; Y(1), \dots, Y(m)) L(m)$$

从而, $G(m; Y(1), \dots, Y(m)) = (\eta, L)(m) [(L, L)(m)]^{-1}$, 由 (3.2.20), $[(L, L)(m)]^{-1}$ 已知。故只需计算 $(\eta, L)(m)$,

$$\begin{aligned}(\eta, L)(m) &= E(\eta(m)L^T(m)|Y(1), \dots, Y(m-1)) \\ &= E\left((F_i(k;m+1) - P_i(k;m))\left(Y^T(m) - \widehat{Y}^T(m|m-1)\right)|Y(1), \dots, Y(m-1)\right) \\ &= E(F_i(k;m+1)Y^T(m) - P_i(k;m)Y^T(m)|Y(1), \dots, Y(m-1)) \\ &= E\left(F_i(k;m+1)\left[H(m)^T X(m) + V(m)\right]^T|Y(1), \dots, Y(m-1)\right) \\ &\quad - P_i(k;m)E\left(\left[H(m)^T X(m) + V(m)\right]^T|Y(1), \dots, Y(m-1)\right) \\ &= E\left(E\left(X(k)X_i(m)\left[H(m)^T X(m) + V(m)\right]^T|Y(1), \dots, Y(m)\right)|Y(1), \dots, Y(m-1)\right)\end{aligned}$$

$$\begin{aligned}
& -P_i(k;m)E\left(\left[H(m)^T X(m)+V(m)\right]^T|Y(1),\dots,Y(m-1)\right) \\
& =E\left(X(k)X_i(m)X^T(m)|Y(1),\dots,Y(m-1)\right)H(m) \\
& \quad -P_i(k;m)\hat{X}^T(m|m-1)H(m) \\
& =\Lambda'(k;m+1)H(m)-P_i(k;m)\hat{X}^T(m|m-1)H(m)
\end{aligned}$$

故(3.2.29)成立。

由于 $P(k;k+1)$ 已知,从而 $F_i(k;k+2)$,进一步 $F(k;k+2)$ 已知,由 $P(k;m)$ 的表达式, $P(k;k+2)$ 已知,如此循环下去,对 $\forall m \geq k+1, P(k;m)$ 已知。而 $\hat{X}(m|m-1)$ 由(3.2.24)可得。 $O(m)$ 也是已知的,所以,下面只需计算 $\hat{X}(k|m-1)$ 。利用和定理3.2.2同样的方法可得:

$$\hat{X}(k|m-1) = \hat{X}(k|m-2) + \{P(k;m-1) - \hat{X}(k|m-2)\hat{X}(m-1|m-2)\}O(m), \quad m \geq k+2, k \geq 1$$

从而, $\hat{X}(k|m), k+1 \leq m \leq T-1$ 已知。代入(3.2.28)就可算出内插。

内插 $\hat{X}(k|T)$ 事实上是 $X(k)$ 的最小均方误差估计。我们根据它对状态序列作出识别。

若 $\hat{X}_i(k|T) = \max\{\hat{X}_j(k|T), j=1, \dots, N\}$, 则判断 $X(k)$ 的状态为 u_i 。若有多个 i 满足上式, 则 $X(k)$ 可在对应的 u_i 中任取一个。以上做法的理论依据是最大后验概率原则, 具体地, 由于 $\{u_1, \dots, u_N\}$ 为 R^N 的标准正交基, 故 $X(k)$ 关于 $\sigma(Y(1), \dots, Y(T))$ 的条件期望其实就是向量 (e_1, \dots, e_N) , 其中 $e_i = P(X(k) = u_i | Y(1), \dots, Y(T))$, 即

$$E(X(k)|Y(1), \dots, Y(T)) = \sum_{i=1}^N u_i P(X(k) = u_i | Y_1, \dots, Y_T) = (e_1, \dots, e_N)$$

易见, e_1, \dots, e_N 中哪一个大, $X(k)$ 就应该取哪一个 e_i 对应的 u_i 。

§ 3.3 滤波和平滑的渐近误差分析

前一节讨论了基于滤波和平滑的HMM的状态识别问题, 这一节我们将运用假设检验思想对一族HMM的滤波和平滑估计的性能做一下分析, 得到了若干有用的结论。简便起见, 我们只讨论平滑的误差, 得到了其上界和下界。对滤波可同样讨论。

文献[21], [22]就Wiener和Kalman滤波的性能进行了比较, 由于理论上误差公式

的缺乏, 故以前很少对滤波和平滑的性能进行比较, 直到文献[23]提供了一种误差渐近公式, 接着便迎来了滤波和平滑的误差分析的高潮。

我们将要研究的是一族 HMM。研究思路是先通过假设检验工具得到平滑误差的上界和下界, 然后运用这个上界和下界得到了平滑误差的表达式。

在我们研究的问题当中, 参数 ε 只出现于转移概率矩阵中, 特征概率矩阵不受其影响。

简单考虑, 我们研究的 HMM 将是一个离散时间, 离散状态的, 相应的隐状态空间和观测状态空间分别仍设为 $\{s_1, s_2, \dots, s_N\}$ 和 $\{\gamma_1, \gamma_2, \dots, \gamma_M\}$, 转移概率为

$$a_{ij} = P(X(k+1) = s_j | X(k) = s_i) = \begin{cases} \varepsilon \lambda_{ij}, & s_i \neq s_j \\ 1 - \varepsilon \lambda_{ii}, & s_i = s_j \end{cases}$$

其中 $\lambda_{ii} = \sum_{j \neq i} \lambda_{ij}$, λ_{ij} 为从状态 s_i 到状态 s_j 的转移率, ε 为决定这种转移发生频率的参数。

特征矩阵不变。除非特别声明, 否则 $a_{ij} > 0, b_{mn} \triangleq b_n(m) > 0$, b_{mn} 表示在隐状态为 s_n 的情况下, 观察为 γ_m 的概率。以 $b_{.i}$ 表示特征矩阵的第 i 列。

假设 $\forall i, j$, Kullback-Leibler 散度

$$K(i, j) = \sum_k b_{ki} \cdot \log \frac{b_{ki}}{b_{kj}} \triangleq E_i \log \frac{b_{.i}}{b_{.j}}$$

存在, 且 $\min_{i,j} K(i, j) > 0$. 其中 E_i 表示关于 $b_{.i}$ 的期望。

我们知道, 滤波

$$\hat{X}(k|k) = \arg \max_i P(X(k) = s_i | Y(0) = y(0), \dots, Y(k) = y(k)), \quad \forall i \in \{s_1, s_2, \dots, s_N\}$$

而带有时滞 $\Delta > 0$ 的平滑为

$$\hat{X}(k|k+\Delta) = \arg \max_i P(X(k) = s_i | Y(0) = y(0), \dots, Y(k+\Delta) = y(k+\Delta)), \quad \forall i \in \{s_1, \dots, s_N\}.$$

在文献[28]中, 也有如下表示

$$\begin{aligned} & P(X(k) = s_i | Y(0) = y(0), \dots, Y(k+\Delta) = y(k+\Delta)) \\ &= \frac{P(\hat{X}(k) = s_i | Y(0) = y(0), \dots, Y(k) = y(k))}{\theta P(X(k) = s_i)} \\ & P(X(k) = s_i | Y(k+1) = y(k+1), \dots, Y(k+\Delta) = y(k+\Delta)) \end{aligned}$$

其中 θ 为标准化常数, 上式的分子有两部分组成: 一个是基于 $T_f = [k-T, k]$ 的前向滤波, 另一个是基于 $T_b = [k+1, k+T]$ 的反时间一步前向预测。

下面首先回顾一下假设检验的知识。

令 $\{G_1, \dots, G_N\}$ 是独立同分布的随机变量序列，于是可有如下的假设检验

H_0 : 数据由密度函数 $p_1(g)$ 产生;

H_1 : 数据由密度函数 $p_2(g)$ 产生。

根据 Neymann-Pearson 引理，我们选用对数似然比检验

$$\frac{1}{N} \sum_{i=1}^N \log \frac{p_2(g_i)}{p_1(g_i)} \begin{cases} > \mu_N, & \text{接受 } H_1 \\ \leq \mu_N, & \text{接受 } H_0 \end{cases}$$

其中 μ_N 为保持检验水平的临界值。

我们将采用如下式子计算平滑误差概率

$$P_N = \alpha_N \cdot P(\text{接受 } H_0) + \beta_N \cdot P(\text{接受 } H_1)$$

其中 $\alpha_N = P(\text{接受 } H_1 | H_0 \text{ 真})$ 为犯第一类错误的概率， $\beta_N = P(\text{接受 } H_0 | H_1 \text{ 真})$ 为犯第二类错误的概率。

我们将要用到的两种假设是：一，马氏链在整个区间 $[1, T]$ 上保持在已知的同一个状态 s_i ；二，在其中某个时刻 τ 转到另一个不同的已知的状态 s_j ，即

$H_0: \{Y(1), \dots, Y(T)\}$ 是独立同分布的随机变量序列，服从分布律

$$b_{.i} = \{P(Y(k) = \cdot | X(k) = s_i)\};$$

$H_1: \{Y(1), \dots, Y(\tau-1)\}$ 是独立同分布的随机变量序列，服从分布律

$$b_{.i} = \{P(Y(k) = \cdot | X(k) = s_i)\}$$

其中 $k \in \{1, \dots, \tau-1\}$; $\{Y(\tau), \dots, Y(T)\}$ 是独立同分布的随机变量序列，服从分布律

$$b_{.j} = \{P(Y(k) = \cdot | X(k) = s_j)\}$$

其中 $k \in \{\tau, \dots, T\}$, τ 服从 $\{1, \dots, T\}$ 上的均匀分布。

我们进一步假设在 $[1, T]$ 上至多仅能发生一个转移，这往往是不可能的。然而可以证明通过选择适当的 T ，可确保过程在 $[1, T]$ 上发生多个转移的概率是很小的，至多是 $O\left(\varepsilon^2 \log\left(\frac{1}{\varepsilon}\right)\right)$, ($\varepsilon \rightarrow 0$) (参见文献[29])。为得到平滑的渐近误差的边界，由文献[23]，数据长度 T 可被定义为

$$T = \frac{1+\delta}{\Lambda} \log \frac{1}{\varepsilon}$$

其中 $\delta > 0$ 是任意的, $\Lambda = \sup_{\theta} \left[-\log E_{\theta} \left(\frac{b_{y_i'}}{b_{y_{k,j}}} \right)^{\theta} \right]$. 此处的 s_i, s_j 是上述假设检验中的两个状态。

我们还假设在区间 $[k-T, k+T]$ 上, 状态转移时刻 t_c 服从均匀分布, 即

$$P(t_c = s) = \frac{1}{2T}.$$

为得到平滑误差的界, 我们需要研究如下一个假设检验问题

$$H_0: k-T < t_c \leq k;$$

$$H_1: k+1 \leq t_c \leq k+T.$$

需要注意的是, 对一个马氏链来说, 在给定 $X(k-T) = s_i, X(k+T) = s_j$ 的情况下从状态 s_i 转移到状态 s_j 的准确时刻的概率分布可如下表示:

令 Z 是在区间 $[k-T, k+T]$ 上仅有一次转移, 则

$$\begin{aligned} & P(t_c = k-T+s | X(k-T) = s_i, X(k+T) = s_j, Z) \\ &= \frac{P(X(k-T) = s_i, \dots, X(k-T+s-1) = s_i, X(k-T+s) = s_j, \dots, X(k+T) = s_j)}{\sum_{t_c=k-T+1}^{k+T} P(X(k-T) = \dots = X(t_c-1) = s_i, X(t_c) = \dots = X(k+T) = s_j)} \\ &= \frac{a_{ii}^{s-1} a_{ij} a_{jj}^{2T-s}}{\sum_{u=1}^{2T} a_{ii}^{u-1} a_{ij} a_{jj}^{2T-u}} = \frac{a_{ii}^{s-1} a_{jj}^{2T-s}}{\sum_{u=1}^{2T} a_{ii}^{u-1} a_{jj}^{2T-u}} \end{aligned}$$

可以证明, 随着 $\varepsilon \rightarrow 0$, 上式的分母是 $2T(1+o(1))$, 由关于 t_c 分布的假设可知, 上式的分子是 $1+o(1)$.

下面给出几个记号和结论。

$$\begin{aligned} E_{[k-T, k+T]}^{i \rightarrow j} &\triangleq \{X(k-T) = \dots = X(t-1) = s_i, X(t) = \dots = X(k+T) = s_j\} \\ E'_{[k-T, k+T]} &\triangleq \{X(k-T) = \dots = X(k+T) = s_i\} \end{aligned}$$

其中 $t \in [k-T, k+T]$. 接下来我们定义对数似然比函数

$$S(t) \triangleq \log \frac{P(y(k-T), \dots, y(k+T) | E_{[k-T, k+T]}^{i \rightarrow j})}{P(y(k-T), \dots, y(k+T) | E'_{[k-T, k+T]})}$$

定理 3.3.1^[29] 考虑独立观测集合 $\{y(k-T), \dots, y(k+T)\}$, 对于 $l=k-T, \dots, t_c-1$, $y(l)$ 的分布律为 $b_{.,l} = (P(Y(l)=y(l)|X(l)=i))$; 对于 $l=t_c, \dots, k+T$, $y(l)$ 的分布律为

$$b_{.,l} = (P(Y(l)=\bullet|X(l)=s_j)).$$

且 $\forall s_i \neq s_j \in \{s_1, \dots, s_N\}, K(i, j) > 0$, 进一步假设 t_c 服从 $[k-T, k+T]$ 上的均匀分布; 那么, 未知的转移时刻 t_c 的假设是

$$H_0: k-T < t_c \leq k; \quad H_1: k+1 \leq t_c \leq k+T.$$

通过选择估计量 \hat{t}_c 去最大化对数似然比函数

$$\begin{aligned} S(t) &= \log \frac{P(Y(k-T), \dots, Y(k+T)|E_{[k-T, k+T]}^{t_i \rightarrow j})}{P(Y(k-T), \dots, Y(k+T)|E_{[k-T, k+T]}^i)} \\ &= \log \frac{b_{y(k-T),i} \cdots b_{y(t-1),i} \cdot b_{y(t),j} \cdots b_{y(k+T),j}}{b_{y(k-T),i} \cdots b_{y(k+T),i}} \\ &= \sum_{l=i}^{k+T} \log \frac{b_{y(l),j}}{b_{y(l),i}} \end{aligned}$$

对于充分大的 T 和某一 $\gamma(i, j) > 0$ 得到的渐近误差为

$$P(C, t_c \in T_f \cup T_b) = \frac{\gamma(i, j)}{2T}$$

其中 C 表事件“得到了不正确的假设”, 而 $\gamma(i, j)$ 依赖于 $b_{.,i}$ 和 $b_{.,j}$.

由该定理可得平滑误差的下界

$$\begin{aligned} P(X(k) \neq \hat{X}(k|k+\Delta)) &\geq \sum_i \sum_{j \neq i} P(X(k-T)=s_i, X(k+T)=s_j, Z) \cdot \\ &\quad \left\{ P(\hat{t}_c \in T_b, t_c \in T_f | X(k-T)=s_i, X(k+T)=s_j, Z) + \right. \\ &\quad \left. P(\hat{t}_c \in T_f, t_c \in T_b | X(k-T)=s_i, X(k+T)=s_j, Z) \right\} \end{aligned}$$

其中 Z 同前。其上界可见文献[29]。

在以上假设下, 有如下结论成立: 随着 $\varepsilon \rightarrow 0$, 渐近平滑误差为

$$P(X(k) \neq \hat{X}(k|k+T)) = \left(\sum_i \pi_i \cdot \sum_{j \neq i} \gamma(i, j) \lambda_{ij} \right) \varepsilon (1+o(1))$$

其中 (π_i) 为马氏链的平稳分布, $\gamma(i, j)$ 仅依赖于 $b_{.,i}$ 和 $b_{.,j}$.

从上式可以看到 $\varepsilon \rightarrow 0$ 时, 平滑误差是 $O(\varepsilon)$, 而滤波的误差是 $O\left(\varepsilon \log \frac{1}{\varepsilon}\right)$, 这意味着平滑相对于滤波提高了 $O\left(\log \frac{1}{\varepsilon}\right)$. 文献[29]中的模拟试验中选用了如下一种模型

$$A = \begin{bmatrix} 1-0.2\varepsilon & 0.2\varepsilon \\ 0.4\varepsilon & 1-0.4\varepsilon \end{bmatrix}, \quad B = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}.$$

其误差概率随 ε 的变化情况可见文献[29]。其中的第一个图充分说明了该渐近表达式的可靠性。

§ 3.4 Gibbs 方法

对于 HMM 的参数训练和隐状态识别, 前面所提到的方法大都将它们分开来讨论。这儿我们将给出一种互为交叉优化将两者同时进行的统计方法—Gibbs 抽样。将这种马氏链蒙特卡罗方案用于简单 HMM 可见文献[30], [31], [32], 用于非平稳 HMM 的可见文献[5]。可以容易地将文献[5]的方法延伸, 即将观察空间拓广为连续的。只需给发散概率密度中的参数选好先验分布即可。

3.4.1 Gibbs 抽样

EM 算法得到的是后验分布的众数, 有时我们希望得到其它一些后验量, 如后验均值和后验方差。而这些后验量都可归结为后验分布的积分计算。对于较简单的后验分布, 我们可以直接计算或利用正态近似, 数值积分, 静态蒙特卡罗方法等。但当后验分布较复杂时这些方法难以实施, 而实际当中, 观测后验分布往往是复杂的, 高维的, 非标准形式的分布。因此, 必须探讨新的计算方法。马氏链蒙特卡罗方法是可以解决这类问题的有效的 Bayes 方法。其基本思想是通过建立一个平稳分布为我们所关心的后验分布的马氏链来得到我们所关心的分布的样本, 基于这些样本我们可以对后验量作各种统计推断。

Gibbs 抽样是最简单, 应用最广泛的一种马氏链蒙特卡罗方法。该方法是由 Geman 最早命名提出的。下面我们给出应用最多的单元素 Gibbs 抽样的具体步骤。

在给出起始点 $x^{(0)} = (x^{(0)}(1), \dots, x^{(0)}(n))$ 后, 假定第 t 次迭代开始时的估计值为 $x^{(t-1)}$, 则第 t 次迭代分为如下 n 步:

(1) 由满条件分布 $\pi(x(1)|x^{(t-1)}(2), \dots, x^{(t-1)}(n))$ 抽取 $x^{(t)}(1)$;

.....

(i) 由满条件分布 $\pi(x(i)|x^{(i)}(1), \dots, x^{(i)}(i-1), x^{(i-1)}(i+1), \dots, x^{(i-1)}(n))$ 抽取 $x^{(i)}(i)$;

.....

(n) 由满条件分布 $\pi(x(n)|x^{(i)}(1), \dots, x^{(i)}(n-1))$ 抽取 $x^{(i)}(n)$.

相关的更多的知识可参见文献[4]。

3.4.2 HMM 参数及隐状态的抽样估计

A 预备

在我们执行 Gibbs 抽样计划之前,我们需要阐述一下 HMM 的所有未知内容的先验分布。一般情况下,初始状态概率和状态转移概率以及发散参数能通过多项分布建模。而当初始概率分布为多项分布时,一个标准的先验分布是狄氏分布^[33],因此我们通过狄氏分布为初始状态概率,转移概率以及发散参数建模。

假设隐状态空间为 $\{s_1, s_2, \dots, s_N\}$, 且对应于状态 s_i 的特征密度为混合高斯密度:

$$b_i(y) = \sum_{i=1}^M c_i \cdot p_i(y)$$

其中 c_i 为混合高斯系数,它满足: $\sum_{i=1}^M c_i = 1$, $c_i \geq 0$. $p_i(y)$ 为均值和方差分别为 u_i 和 σ_i 的正态分布。方便起见,不妨设 $\sigma_i = \sigma, \forall i$.

和初始分布概率一样,设 (c_{i1}, \dots, c_{iM}) 的先验分布假设为狄氏分布,即为

$$(c_{i1}, \dots, c_{iM}) \sim Di(z_{i1}, \dots, z_{ip})$$

而 (u_{i1}, \dots, u_{iM}) 可根据具体的背景来选择先验分布。明显 (c_{i1}, \dots, c_{iM}) , (u_{i1}, \dots, u_{iM}) 和 σ 是独立的,从而 $f_i = (u_{i1}, \dots, u_{iM}, c_{i1}, \dots, c_{iM}, \sigma)$ 的先验分布可以写出。相关参数的先验分布总结如下:

(1) 初始状态概率分布 $\pi = (\pi_1, \dots, \pi_n)$ 的先验分布设为 $p(\pi_1, \dots, \pi_n)$: 它一般为 $n-1$ 维多元狄氏分布,具体为

$$\pi \sim Di(\alpha_1, \dots, \alpha_n)$$

其中 $\alpha_i, i=1, \dots, n$ 为 $n-1$ 维多元狄氏分布的参数。

(2) 状态逗留 d 的分布一般设为截尾泊松分布,即

$$P_i(d) \propto \frac{\varepsilon_i^d \cdot e^{-\varepsilon_i}}{d!}, \quad d=1,2,\dots,D$$

其中 D 为最大逗留时间。 ε_i 是对应状态 s_i 的泊松分布的参数, $\varepsilon_i > 0$, 所有 ε_i 的先验分布一般设为 Gamma 分布, 这是因为此时 Gamma 分布是共轭先验分布。不妨设其参数为 u 和 v 。

(3) 给定一个状态 s_i , 状态转移权重 $\omega_i = (\omega_{i1}, \dots, \omega_{i(i-1)}, \omega_{i(i+1)}, \dots, \omega_{in})$ 的先验分布也建模为狄氏分布。发散概率参数 (c_{i1}, \dots, c_{iM}) 的先验分布也建模为狄氏分布。具体为

$$\begin{aligned} \omega_i &\sim \text{Di}(\eta_{i1}, \dots, \eta_{i(n-1)}) \\ (c_{i1}, \dots, c_{iM}) &\sim \text{Di}(z_{i1}, \dots, z_{ip}) \end{aligned}$$

σ 的先验分布在同等无知的情况下建模为 $(0,2)$ 上的均匀分布。

B Gibbs 抽样程序

运用所选择的先验分布, 可得到在第 k 次迭处有如下后验分布:

(1) 从

$$\pi \sim p(\pi | \varepsilon^{(k-1)}, \omega^{(k-1)}, B^{(k-1)}, x^{(k-1)}, y) = p(\pi | x^{(k-1)}) = \text{Di}(\alpha_1 + \delta_{q_1^{(k-1)}, 1}, \dots, \alpha_n + \delta_{q_n^{(k-1)}, n})$$

中抽取 $\pi^{(k)}$ 。其中 $\delta_{q,j}$ 是 Kronecker 符号。

(2) 从

$$\varepsilon_i \sim p(\varepsilon_i | \pi^{(k)}, \omega^{(k-1)}, B^{(k-1)}, x^{(k-1)}, y) = p(\varepsilon_i | x^{(k-1)}) = \text{Ga}(\tilde{d}_i^{(k-1)} + u, m_i^{(k-1)} + v)$$

抽取 ε_i^k 。其中 $\tilde{d}_i^{(k-1)} = \sum_{t=1}^T \delta_{q_i^{(k-1)}, i}$, 而 $m_i^{(k-1)}$ 是状态 s_i 在第 $k-1$ 次迭代中出现的段数。

(3) 从

$$\omega_i \sim p(\omega_i | \pi^{(k)}, \varepsilon^{(k)}, B^{(k-1)}, x^{(k-1)}, y) = p(\omega_i | x^{(k-1)}) = \text{Di}(\eta_{i1} + n_{i1}^{(k-1)}, \dots, \eta_{i(n-1)} + n_{i(n-1)}^{(k-1)})$$

抽出 $\omega_i^{(k)}$ 。其中 $n_{ij}^{(k-1)}$ 是在第 $k-1$ 次迭代中从状态 s_i 到状态 s_j 的转移数。

(4) $a_{ii}(d)$ 可从如下式子得到

$$a_{ii}(d) = \begin{cases} 1 - P_i(d), & d=1 \\ \frac{1 - \sum_{k=1}^d P_i(k)}{1 - \sum_{k=1}^{d-1} P_i(k)}, & d>1 \end{cases}$$

$a_{ij}^{(k)}(d)$, $i \neq j$ 可从

$$a_{ij}^{(k)}(d) = \begin{cases} \omega_y^{(k)} \cdot P_i(d), & d=1 \\ \frac{\omega_y^{(k)} \cdot P_i(d)}{1 - \sum_{k=1}^{d-1} P_i(k)}, & d>1 \end{cases}$$

得到。

(5) 从

$$\begin{aligned} f_i &\sim p(f_i | \pi^{(k)}, \varepsilon^{(k)}, \omega^{(k)}, x^{(k-1)}, y) = p(f_i | x^{(k-1)}, y) \\ &= p(f_i) \cdot \prod_k (b_i(y(k)))^{n_k} \end{aligned}$$

中抽出 $f_i^{(k)}$. 其中 $p(f_i)$ 是 f_i 的先验分布, n_k 为出现观察 y_k 的次数。

(6) 从

$$\begin{aligned} x(t) &\sim p(x(t) | x^{(k)}(1), \dots, x^{(k)}(t-1), x^{(k)}(t+1), \dots, x^{(k-1)}(T), \pi^{(k)}, \varepsilon^{(k)}, \omega^{(k)}, B^{(k)}, y(t)) \\ &\propto a_{x^{(k)}(t-1), x(t)}(d(x^{(k)}(t-1))) \cdot a_{x(t), x^{(k-1)}(t+1)}(d(x(t))) \cdot \\ &\quad \left(\prod_{l=1}^{t-1} a_{x^{(k-1)}(t+l), x^{(k-1)}(t+l+1)}(d(x^{(k-1)}(t+l))) \right) \cdot b_{x(t)}(y(t)) \end{aligned}$$

中抽出 $x^{(k)}(t)$. 其中 $x^{(k-1)}(t+\tau) \neq x^{(k-1)}(t+\tau-1)$, $x^{(k-1)}(t+\tau-1) = x^{(k-1)}(t+1)$.

相关的收敛性讨论和模拟试验可见文献[4], [5]。

§ 3.5 基于 HMM 的语音的一种初步降噪方法的理论分析

在现实的世界中,有些信号是伴随着较强的噪音发送的,即我们接受到的信号除了系统误差外,还有额外的附加噪声过程。为了准确地估计信号,我们必须先对接收到的信号作初步的处理,使得处理后的信号相对更加真实一些。这样有助于提高我们的解码质量。

3.5.1 模型的建立

设在概率空间 (Ω, F, P) 上,几个定义于其上的过程 $X = (X(t))_{0 \leq t \leq N}$, $Y = (Y(t))_{0 \leq t \leq N}$ 分别为原信号的状态过程和观察过程, $\tilde{X} = (\tilde{X}(t))_{0 \leq t \leq N}$, $V = (V(t))_{0 \leq t \leq N}$ 分别为附加的噪声过程的状态过程和观察过程, $\bar{X} = (\bar{X}(t))_{0 \leq t \leq N}$, $Z = (Z(t))_{0 \leq t \leq N}$ 分别为带有附加噪声的信号过程的

状态过程和观察过程, 即 $\bar{X} = (X, \tilde{X})$ 为合成状态过程。我们的目的是根据 $Z = (Z(t))_{0 \leq t \leq N}$ 来找被降噪的观察过程 $Y = (Y(t))_{0 \leq t \leq N}$, 用 $Y = (Y(t))_{0 \leq t \leq N}$ 去作识别结果要比用 $Z = (Z(t))_{0 \leq t \leq N}$ 进行识别好得多。

设 $(X(t), Y(t))_{0 \leq t \leq N}$ 为一阶 M 状态 HMM, 特征密度为 0 均值的高斯密度。从而有

$$p(y'_0) = \sum_{X'_0} \prod_{\tau=0}^l a_{X_{\tau-1}, X_\tau} \cdot b_{X(\tau)}(y(\tau)) \quad (3.5.1)$$

其中 $a_{X(-1), X(0)} = \pi_{X(0)}$, $y'_0 = (y(0), \dots, y(l))$ 设该概率密度对应的协方差阵为 $S_{X(T)}$ 。

设 $(\tilde{X}(t), \nu(t))_{0 \leq t \leq N}$ 为一阶 \bar{M} 个状态 HMM, 特征密度为 0 均值的高斯密度。从而有

$$p(\nu'_0) = \sum_{\bar{X}'_0} \prod_{\tau=0}^l a_{\bar{X}(\tau-1), \bar{X}(\tau)} \cdot b_{\bar{X}(\tau)}(\nu(\tau)) \quad (3.5.2)$$

其中 $a_{X(-1), X(0)} = \pi_{X(0)}$ 。

由 (3.5.1), (3.5.2) 可得 $p(z'_0)$ 如下:

$$p(z'_0) = \sum_{\bar{X}'_0} \prod_{\tau=0}^l a_{\bar{X}(\tau-1), \bar{X}(\tau)} \cdot b_{\bar{X}(\tau)}(z(\tau)) \quad (3.5.3)$$

其中 $\bar{X}(\tau) = (X(\tau), \tilde{X}(\tau))$, $a_{\bar{X}(\tau-1), \bar{X}(\tau)} = a_{X(\tau-1), X(\tau)} \cdot a_{\tilde{X}(\tau-1), \tilde{X}(\tau)}$ 。

$$b_{\bar{X}(\tau)}(z(\tau)) = \int b_{X(\tau)}(z(\tau) - y(\tau)) \cdot b_{X(\tau)}(y(\tau)) dy(\tau) \quad (3.5.4)$$

该密度的协方差阵是 $S_{X(t)} + S_{\bar{X}(t)}$ 。明显, $(\bar{x}(t), z(t))$ 是一个一阶 $M \times \bar{M}$ 状态 HMM。

给定 $z_0^T = (z(0), \dots, z(T))$, $T \geq t$, $y(t)$ 的条件密度为

$$p(y(t)|z_0^T) = \sum_{\bar{x}(t)} p(\bar{x}(t)|z_0^T) \cdot b_{\{z(t), \bar{x}(t)\}}(y(t)) \quad (3.5.5)$$

其中 $p(\bar{x}(t)|z_0^T)$ 为给定观察 z_0^T 的情况下, 合成状态 $\bar{X}(t)$ 的后验概率密度。而 $b_{\{z(t), \bar{x}(t)\}}(y(t))$ 的条件均值和条件协方差为

$$\begin{aligned} E(Y(t)|Z(t), \bar{X}(t)) &= S_{X(t)} \cdot (S_{X(t)} + S_{\bar{X}(t)})^{-1} \cdot z(t) \triangleq H_{\bar{X}(t)} \cdot z(t) \\ \text{cov}(Y(t)|Z(t), \bar{X}(t)) &= H_{\bar{X}(t)} \cdot S_{\bar{X}(t)} \triangleq \sum_{\bar{X}(t)} \end{aligned} \quad (3.5.6)$$

其中 $H_{\bar{X}(t)}$ 是给定 $z(t)$ 下, 信号 $y(t)$ 的最小均方误差估计 MMSE 估计 (维纳滤波), 即为

$E(Y(t)|Z(t))$, 且 z_t 对应的合成状态为 $(X(t), \tilde{X}(t))$. 而 $\sum_{\tilde{x}(t)}$ 是与该估计有关的 MMSE 估计矩阵。

(3.5.5) 中的后验概率 $p(\bar{x}(t)|z_0^T)$ 和 (3.5.3) 式中的 $p(z_0')$ 可通过 HMM 的前后向公式得到。即对于 $0 \leq t \leq T$, 我们有

$$p(\bar{x}(t)|z_0^T) = \frac{F(\bar{x}(t), z_0') \cdot B(z_{t+1}^T | \bar{x}(t))}{\sum_{\tilde{x}(t)} F(\bar{x}(t), z_0') \cdot B(z_{t+1}^T | \bar{x}(t))} \quad (3.5.7)$$

其中, 带有噪音的信号 $z(t)$ 的前向概率密度是

$$\begin{aligned} F(\bar{x}(0), z(0)) &\triangleq p(\bar{x}(0), z(0)) = \pi_{\tilde{x}(0)} \cdot b(z(0) | \bar{x}(0)) \\ F(\bar{x}(t), z_0') &\triangleq p(\bar{x}(t), z_0') = \sum_{\tilde{x}(t-1)} F(\bar{x}(t-1), z_0'^{-1}) \cdot a_{\tilde{x}(t-1)\tilde{x}(t)} \cdot \\ &\quad b(z(t) | \bar{x}(t)), 0 < t \leq T \end{aligned} \quad (3.5.8)$$

后向概率密度为:

$$\begin{aligned} B(z_{T+1}^T | \bar{x}(T)) &\triangleq 1 \\ B(z_{t+1}^T | \bar{x}(t)) &\triangleq p(z_{t+1}^T | \bar{x}(t)) \\ &= \sum_{\tilde{x}(t+1)} B(z_{t+2}^T | \bar{x}(t+1)) \cdot a_{\tilde{x}(t)\tilde{x}(t+1)} \cdot b(z(t+1) | \bar{x}(t+1)), 0 \leq t < T \end{aligned} \quad (3.5.9)$$

$$\text{由 (3.5.8) 得} \quad p(z_0') = \sum_{\tilde{x}(t)} F(\bar{x}(t), z_0') \quad (3.5.10)$$

由以上几个式子就可以算出 $E(Y(t)|Z_0^T)$, 这个应该叫平滑估计或内插估计, 它为 MMSE 估计。

3.5.2 MMSE 估计和 MAP 估计

给定 z_0^T , $T \geq t$, $y(t)$ 的 MMSE 估计由 (3.5.5) 式可容易得到。我们有

$$\hat{Y}(t) \triangleq E(Y(t) | Z_0^T) = \sum_{\tilde{x}(t)} p(\bar{x}(t) | z_0^T) \cdot E(Y(t) | Z(t), \bar{X}(t)) \quad (3.5.11)$$

其中 $p(\bar{x}(t) | z_0^T)$ 在 (3.5.7) 中有, $E(Y(t) | Z(t), \bar{X}(t))$ 在 (3.5.6) 中有。MMSE 估计在 (3.5.11)

$$\text{中在如下意义下是无偏的, } E(\hat{Y}(t)) = E(E(Y(t) | Z_0^T)) = E(Y(t)) \quad (3.5.12)$$

和该估计相关的 MMSE 收敛到另外一个 MMSE, 即

$$\begin{aligned}\overline{\varepsilon}_t^2 &\triangleq \lim_{k \rightarrow \infty} \frac{1}{k} \text{tr} E \left((Y(t) - \hat{Y}(t)) \cdot (Y(t) - \hat{Y}(t))^T \right) \\ &= \sum_{\bar{x}(t)} p(\bar{x}(t)) \cdot \int_{\mathbf{x}} H_{\bar{x}(t)}(\theta) \cdot S_{\bar{x}(t)}(\theta) \frac{d\theta}{2\pi}\end{aligned}\quad (3.5.13)$$

(迹是误差向量的范数, 实质为误差的大小, k 为该向量的维数, 第二个等号后面是总体的或过程状态的 MMSE) 其中 $H_{\bar{x}(t)}(\theta) \triangleq \frac{S_{x(t)}}{S_{x(t)} + S_{\bar{x}(t)}}$ 是维纳滤波的频率响应。先验概率 $p(\bar{x}(t))$ 可由 (3.5.7) 到 (3.5.9) 以及恒等式 $b(z(t)|\bar{x}(t)) \equiv 1, \forall t$ 得到。以上成立是在如下假设下得到的: 功率谱密度 $\{S_m(\theta), S_n(\theta)\}$ 是一直有界的, 所有的转移概率都是严格正的。

下面讨论一下 MAP 估计。

就 (3.5.5) 关于 $y(t)$ 求最大值就可得到 y_t 的最大后验概率估计。由于 $p(y(t)|z_0^T)$ 的梯度方程是高度非线性以及没有明确的解, 所以我们用 EM 算法来迭代地求 $p(y(t)|z_0^T)$ 的局部极大。根据 (3.5.5) 有 $p(\bar{x}(t), y(t)|z_0^T) = p(\bar{x}(t)|z_0^T) \cdot b(y(t)|z(t), \bar{x}(t))$ 。 (3.5.14)

令 $y'(t)$ 是 $y(t)$ 的当前估计, 由 (3.5.14) 以及 Jensen's 不等式可得

$$\begin{aligned}&\ln p(y(t)|z_0^T) - \ln p(y'(t)|z_0^T) \\ &= \ln \sum_{\bar{x}(t)} \frac{p(\bar{x}(t), y'(t)|z_0^T)}{p(y'(t)|z_0^T)} \cdot \frac{p(\bar{x}(t), y(t)|z_0^T)}{p(\bar{x}(t), y'(t)|z_0^T)} \\ &\geq \sum_{\bar{x}(t)} p(\bar{x}(t)|y'(t), z_0^T) \cdot \ln \frac{b(y(t)|z(t), \bar{x}(t))}{b(y'(t)|z(t), \bar{x}(t))} \\ &\triangleq Q(y(t), y'(t)) - Q(y'(t), y'(t))\end{aligned}\quad (3.5.15)$$

$$\text{即其中 } Q(y_t, y'_t) \triangleq \sum_{\bar{x}_t} p(\bar{x}_t|y'_t, z_0^T) \cdot \ln b(y_t|z_t, \bar{x}_t) = E \left(\ln b(Y_t|Z_t, \bar{X}_t) \middle| Y'_t, Z_0^T \right). \quad (3.5.16)$$

根据 (3.5.15), 关于 $y(t)$ 最大化 $Q(y(t), y'(t))$ 就有 $p(y(t)|z_0^T) \geq p(y'(t)|z_0^T)$, 其中等号成立等价于如下

$$y(t) = y'(t), a.e. P(\bar{x}(t)|y'(t), z_0^T)$$

其中 $P(\bar{x}(t)|y'(t), z_0^T)$ 为条件分布 $p(\bar{x}(t)|y'(t), z_0^T)$ 诱导的概率测度。这样就可得到一具有

单调不减似然值的序列 $y(t, n)$. 其中

$$\begin{aligned} y(t, n+1) &\triangleq \arg \max_{y(t)} Q(y(t), y(t, n)) \\ &= \left[\sum_{\bar{x}(t)} p(\bar{x}(t) | y(t, n), z_0^T) \cdot H_{\bar{x}(t)}^{-1} \right]^{-1} \cdot z(t) \end{aligned} \quad (3.5.17)$$

其中 $p(\bar{x}(t) | y(t, n), z_0^T)$ 使用 (3.5.14) 以及前后向公式 (3.5.7) — (3.5.9) 就可得到。其极限就是 $y(t)$ 的 MAP 估计。

综上所述, 我们得到了无附加噪声情况下的信号的观察的两种估计。一般来说, 当模型较简单时, 运用 MMSE 估计, 当模型较复杂时, 运用 MAP 估计^[21]。

本章小结

本章不仅介绍了传统的和推广的 Viterbi 解码方法, 还引入了基于滤波和内插(或平滑)的非齐次非平稳隐马氏模型的状态识别方法。虽然后者拥有相对较广泛的处理对象, 但从试验结果可以看出, 这种方法的效果不是很好, 还有待进一步改进。本章还针对该方法的误差进行了细致的分析。除此之外本章还介绍了将隐马氏模型的参数训练和状态识别同时进行的一种统计计算方法: Gibbs 方法。本章最后考虑了 HMM 在语音降噪处理中的应用。在其中我们运用 MMSE 和 MAP 估计进行了语音的初步降噪。

致 谢

短暂的三年研究生生活马上就要结束了，在这期间，我得到了老师们的指导，领导们的关怀，同学们的帮助，值此毕业论文完成之际，我衷心地向所有给予我指导、关心和帮助的人们表示感谢。

首先，我要衷心感谢我的导师李兵教授！李老师学识渊博、经验丰富、治学严谨。他独特、科学的思维方式和对学问孜孜不倦的精神永远是我学习的榜样。李老师不仅在学业和学术研究上对我严格要求，细心指导；而且在生活上关心我，在为人处世方面言传身教，使我终身受益。在指导我完成课题与论文期间，李老师时时刻刻关注论文的进展；遇到问题时，他总是仔细分析，耐心指导，独特的分析问题方式和对问题本质的把握能力令我叹服。在三年的时间里，李老师倾注了大量的心血，在人生道路上给了我很多启迪和指导。在此，对李老师致以深深的敬意！

感谢金治明教授！在上课期间，金老师严肃认真的治学态度，严密的思维方式，踏实的作风，开阔的视野使我受益匪浅。

感谢罗强，谢锋、路迎晨等同学，在长期的相处中正是他们真挚的友谊使我拥有一个轻松愉快的学习环境。

最后，我要感谢理学院研究生队的领导和同学们，感谢他们这些年来在学习和生活上对我的关怀和帮助。谢谢！

参考文献

- [1] R. S. Liptser and A. N. Shirayev. Statistics of Random Processes. New York: Springer-Verlag, 1977
- [2] 严加安. 测度论讲义, 北京: 科学出版社, 2000
- [3] 数学模型讲义, 北京大学数学科学学院, 2000
- [4] 苑诗松, 王静龙, 濮晓龙. 高等数理统计, 高等教育出版社, 1998.7
- [5] Peter M.Djuric, Senior Member IEEE and Joon-Hwa Chun. An MCMC Sampling to Estimation of HMMs. IEEE Trans. Signal Processing, Vol.50, no.5, 2002, pp.1113-1123
- [6] Janez Kaiser University of Maribor, Faculty of Electrical Engineering and Computer Science, Smetanova 17 Maribor slovenia. trainin of HMM with GA. www.dsplab.uni-mb.si/Clanki/janez/ast98.pdf
- [7] 丁建立, 陈增强, 袁著祉. 遗传算法与蚂蚁算法的融合. 计算机研究与发展, 2003, 40 卷, 第 9 期
- [8] 韩广军, 罗强, 李兵. 基于 DCPM-FGA 的隐马氏模型的参数训练. 国防科技大学学报, 2004, 第 26 卷, 第 4 期
- [9] Enzo Baccarelli and Robert Cusani. Recursive Kalman-type optimal estimation and detection of Hidden Markov chains. Signal Processing, 1996
- [10] 李敏强等. 遗传算法的基本理论与应用, 科学出版社, 2002
- [11] 易克初, 田斌, 付强. 语音信号处理, 国防工业出版社, 2000.6
- [12] 张文修, 梁怡. 遗传算法的数学基础, 西安交大出版社, 2001
- [13] 云庆夏. 进化算法, 冶金工业出版社, 2000
- [14] 金治明. 随机分析基础及其应用, 北京: 国防工业出版社, 2003.4
- [15] A. Dembo and O. Zeitouni. Parameter Estimation of Partially Observed Continuous Time Stochastic Processing via the EM Algorithm. Stoch. Process. Their Appl., 1986, pp.91-113
- [16] Bongkee Sin and Jin H. Kim, Nonstationary Hidden Markov Model. Signal Processing, Vol. 46, 1995, pp.31-46
- [17] O. Ghitza and M. M. Sondhi. Hidden Markov models with templates as nonstationary states: An application to speech recognition. Computer Speech and Language, Vol.2, 1993, pp.101-119
- [18] M.O. Ostendorf and S. Roukos. A stochastic segment model for phoneme based continuous speech recognition. IEEE Trans. Acoust. Speech Signal Process. Vol.37, 1989, pp.1857-

- [19] S. Roucos and M. O. Dunham. A stochastic segment model for phoneme based continuous speech recognition. Proc. IEEE Internat. Conf. Acoust. Speech Signal Process. Dallas, TX, April, 1987, pp.73-76
- [20] J. D. Ferguson. Variable duration models for speech. Proc. Symp. On the application of Hidden Markov Models to Text and Speech, Princeton, NJ, 1980, pp.143-179
- [21] J. B. Moore and K. L. Teo. Smoothing as an improvement on filtering in high noise. Syst. Contr. Lett., Vol.8, no.1, 1986, pp.51-54
- [22] B. D. O. Anderson and S. Chirarattananon. Smoothing as an improvement on filtering: A universal bound. Electron. Lett., Vol.7, no.18, 1971, pp.524-525
- [23] R. Khasminskii and O. Zeitouni. Asymptotic filtering for finite state Markov chains. Stoch. Process. Their Appl., Vol.63, no.1, 1996, pp.1-10
- [24] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in such speech recognition. Proc. IEEE, vol.77, no.2, 1989, pp.257-285
- [25] 姚天任. 数字语音处理, 华中理工大学出版社, 1992
- [26] 谢贤亚, 王家鼎, 李焕润. EM-HMM 多频率线跟踪算法. 电子学报, 2002, 30(1): 22-26
- [27] 邓明华. 数学模型补充讲义, 北京大学数学科学院, 1999
- [28] L. Shue, B. D. O. Anderson and S. Dey. Exponential Stability of Filters and Smoothers for Hidden Markov Models. IEEE Trans. Signal Processing, Vol.46, 1998, pp.2180-2194
- [29] Louis Shue, Brian D. O. Anderson and Franky De Bruyne. Asymptotic Smoothing Errors for Hidden Markov Models. IEEE Trans. Signal Processing, Vol.48, Processing, Vol.48, no. 12, 2000, pp.3289-3302
- [30] L. R. Rabiner and B. H. Juang. A introduction to hidden Markov models. IEEE ASSP Mag, pp.4-16, Jan, 1986
- [31] G. D. Forney, Jr. The Viterbi algorithm. Proc. IEEE, Vol.61, pp.268-278, Mar. 1973
- [32] 张贤科, 许甫华. 高等代数学, 清华大学出版社, 1998.3
- [33] 郭庆, 柴海新, 吴文虎. 隐马尔可夫模型中状态停留时间的模型化. 清华大学学报自然科学版, 1999, 第 39 卷, 第 5 期
- [34] 谢贤亚, 王家鼎, 李焕润. 基于隐马尔可夫模型的频率线跟踪系统的频率估计精细化 [J] 上海大学学报, 2000, 6 (5): 377-383
- [35] Steffen Michalek, Mirko Wagner, and Jens Timmer. A New Approximate Likelihood Estimation for ARMA-Filtered Hidden Markov Models. IEEE Trans. Signal Processing, Vol.48,

no. 6, 2000, pp.1537-1547

- [36] Louis Shue, Subhrakanti Dey, Brian D. O. Anderson, and Franky De Bruyne. On State-Estimation of a Two-State Hidden Markov Model with Quantization. IEEE Trans. Signal Processing, Vol.49,no. 1, 2001, pp.202-208
- [37] Joonhoon Chun, Joohwan Chun. Nonlinear filtering using the wavelet transform. Signal Processing 80,2000,pp.441-450
- [38] Donald R. Fredkin and John A. Rice. Fast Evaluation of the Likelihood of an HMM: Ion Channel Currents with Filtering and Colored Noise. IEEE Trans. Signal Processing, Vol.49, no.3, 2001, pp.625-633

硕士期间的主要工作

- [1] 韩广军, 罗强, 李兵。基于 DCPM-FGA 的隐马氏模型的参数训练。国防科技大学学报, 2004, 第 26 卷, 第 4 期。
- [2] 韩广军, 李兵。基于近似滤波的非时齐带逗留时间 HMM 的状态识别。计算机工程与应用。已录用。

作者: [韩广军](#)

学位授予单位: [国防科学技术大学研究生院](#)

参考文献(38条)

1. [R S Liptser, A N Shirayev](#) [Statistics of Random Processes](#) 1977
2. [严加安](#) [测度论讲义](#) 2000
3. [数学模型讲义](#) 2000
4. [苑诗松, 王静龙, 濮晓龙](#) [高等数理统计](#) 1998
5. [Peter M Djuric, Senior Member IEEE, Joon-Hwa Chun](#) [An MCMC Sampling toEstimati on of HMMs](#) 2002(05)
6. [Janez Kaiser University of Maribor Faculty of Electrical Engineering and Computer Scienc](#)
[e, Smetanora 17 Maribor slovenia, trainin of HMM with GA](#)
7. [丁建立, 陈增强, 袁著祉](#) [遗传算法与蚂蚁算法的融合](#)[期刊论文]-[计算机研究与发展](#) 2003(9)
8. [韩广军, 罗强, 李兵](#) [基于DCPM-FGA的隐马氏模型的参数训练](#)[期刊论文]-[国防科技大学学报](#) 2004(4)
9. [Enzo Baccarelli, Robert Cusani](#) [Recuursive Kalman-type optimal estimation and detecti on of Hidden](#)
[Markov chains](#) 1996
10. [李敏强, 寇纪淞, 林丹, 李书全](#) [遗传算法的基本理论与应用](#) 2002
11. [易克初, 田斌, 付强](#) [语音信号处理](#) 2000
12. [张文修, 梁怡](#) [遗传算法的数学基础](#) 2001
13. [云庆夏](#) [进化算法](#) 2000
14. [金治明](#) [随机分析基础及其应用](#) 2003
15. [A Dembo, O Zeitouni](#) [Parameter Estimation of Partially Observed Continuous Time Stochastic](#)
[Processing via the EM Algorithm](#) 1986
16. [Bongkee Sin, Jin H Kim](#) [Nonstationary Hidden Markov Model](#)[外文期刊] 1995
17. [O Ghitza, M M Sondhi](#) [Hidden Markov models with templates as nonstationary states:An application](#)
[to speech recognition](#) 1993(02)
18. [M O Ostendorf, S Roukos](#) [A stochastic segment model for phoneme based continuous speech](#)
[recognition](#) 1989
19. [S Roucos, M O Dunham](#) [A stochastic segment model for phoneme based continuous speech recognition](#)
1987
20. [J D Ferguson](#) [Variable duration models for speech](#) 1980
21. [J B Moore, K L Teo](#) [Smoothing as an improvement on filtering in high noise](#) 1986(01)
22. [B D O Anderson, S Chirarattananon](#) [Smoothing as an improvement on filtering:A universal bound](#)
1971(18)
23. [R Khasminskii, O Zeitouni](#) [Asymptotic filtering for fmite state Markov chains](#) 1996(01)
24. [L R Rabiner](#) [A tutorial on hidden Markov models and selected applications in such speech](#)
[recognition](#) 1989(02)

25. [姚天任](#) [数字语音处理](#) 1992
26. [谢贤亚](#), [王家鼎](#), [李焕润](#) [EM-HMM多频率线跟踪算法](#)[期刊论文]-[电子学报](#) 2002(1)
27. [邓明华](#) [数学模型补充讲义](#) 1999
28. [L Shue](#), [B D O Anderson](#), [S Dey](#) [Exponential Stability of Filters and Smoothers for Hidden Markov Models](#)[外文期刊] 1998
29. [Louis Shue](#), [Brian D O Anderson](#), [Franky De Bruyne](#) [Asymptotic Smoothing Errors for Hidden Markov Models](#)[外文期刊] 2000(12)
30. [L R Rabiner](#), [B H Juang](#) [A introduction to hidden Markov models](#) 1986
31. [G D Forney Jr](#) [The Viterbi algorithm](#) 1973
32. [张贤科](#), [许甫华](#) [高等代数学](#) 1998
33. [郭庆](#), [柴海新](#), [吴文虎](#) [隐Markov模型中状态停留时间的模型化](#)[期刊论文]-[清华大学学报（自然科学版）](#) 1999(5)
34. [谢贤亚](#), [王家鼎](#), [李焕润](#) [基于隐马尔可夫模型的频率线跟踪系统的频率估计精细化](#)[期刊论文]-[上海大学学报（自然科学版）](#) 2000(5)
35. [Steffen Michalek](#), [Mirko Wagner](#), [Jens Timmer](#) [A New Approximate Likelihood Estimation for ARMA-Filtered Hidden Markov Models](#) 2000(06)
36. [Louis Shue](#), [Subhrakanti Dey](#), [Brian D O Anderson](#), [Franky De Bruyne](#) [On State-Estimation of a Two-State Hidden Markov Model with Quantization](#) 2001(01)
37. [Joonhoon Chun](#), [Joohwan Chun](#) [Nonlinear filtering using the wavelet transform](#)[外文期刊] 2000
38. [Donald R Fredkin](#), [John A Rice](#) [Fast Evaluation of the Likelihood of an HMM:Ion Channel Currents with Filtering and Colored Noise](#) 2001(03)

本文链接: http://d.g.wanfangdata.com.cn/Thesis_Y788807.aspx