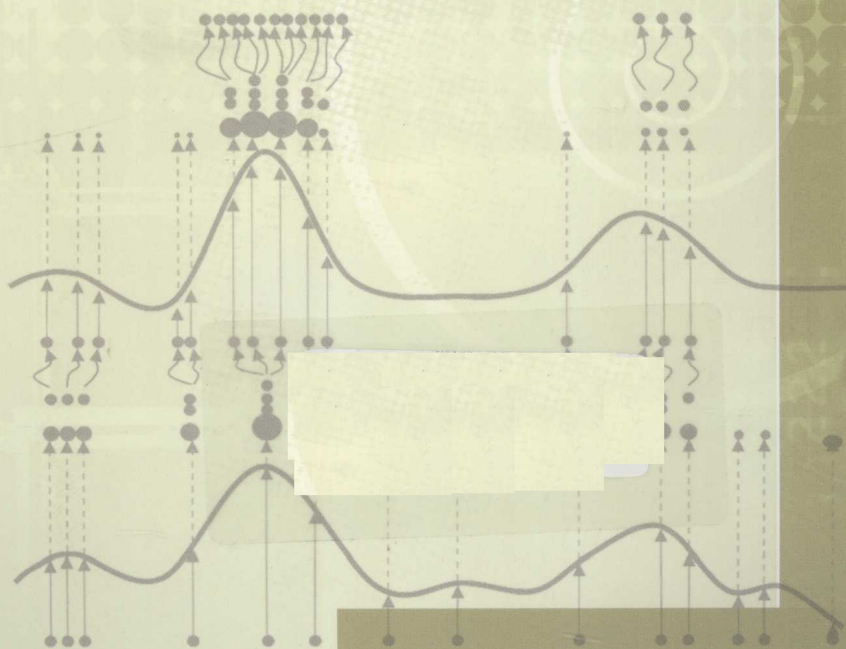



胡士强 敬忠良 著

粒子滤波 原理及其应用



 科学出版社
www.sciencep.com

粒子滤波原理及其应用

胡士强 敬忠良 著

科学出版社

北京

内 容 简 介

如何提高非线性非高斯动态系统的状态估计和预测的精度是系统辨识、适应控制、模式识别、无线传感网络、通信、经济等领域中都会遇到的问题。粒子滤波提供了解决这一问题的采样递推方法。本书结合作者自身的相关研究工作,全面系统地介绍了粒子滤波的主要概念、基本原理、典型算法、应用技术以及国际上有关研究的新成果和新动向。全书可分为理论篇(包含第1~4章)和应用篇(包含第5~7章)。第1章为绪论。第2章给出了动态系统常用的几种滤波方法,主要分析滤波方法的思想,不对算法的适用性进行讨论。第3章介绍了随机采样与粒子滤波方法,这是一种基于采样滤波思想的 Monte-Carlo 贝叶斯估计算法,本章还重点介绍算法如何利用序列重采样实现状态递推估计。第4章讨论了粒子滤波算法的改进算法,主要包括针对重采样过程改进算法和针对似然函数选取的改进算法。第5章利用粒子滤波算法解决了机动目标跟踪非线性问题。第6章讨论了场景监视中的多视频目标检测,表明这类问题可以用粒子滤波算法来解决。第7章用粒子滤波算法研究非线性系统辨识与故障诊断问题。

本书可作为高等院校信息工程、控制科学与工程、电子科学与工程、应用数学、管理工程和计算机科学与应用等学科有关研究的专题阅读材料,也可从事电子对抗、雷达、红外、激光和声呐等声光电传感器跟踪,导航与定位、多传感器信息融合及机器人视觉、智能监视与图像理解、无线传感网络与通信、空中交通管制和金融数据分析等领域研究工作的科学工作者和工程技术人员参考。

图书在版编目(CIP)数据

粒子滤波原理及其应用/胡士强,敬忠良著. —北京:科学出版社,2010
ISBN 978-7-03-028370-2

I. 粒… II. ①胡…②敬… III. 非线性控制系统 IV. O231.2

中国版本图书馆 CIP 数据核字(2010)第 140757 号

责任编辑:姚庆爽 / 责任校对:桂伟利
责任印制:赵 博 / 封面设计:鑫联必升

科学出版社 出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

西陈印刷厂印刷

科学出版社发行 各地新华书店经销

*

2010年8月第一版 开本:B5(720×1000)
2010年8月第一次印刷 印张:10 3/4
印数:1—3 000 字数:217 000

定价:38.00元

(如有印装质量问题,我社负责调换)

前 言

粒子滤波算法作为现代非线性滤波最为关注的一类滤波方法,其近年来的研究取得了引人注目的进展,这些进展必将推动控制科学、信息科学、航空航天和电子科学的快速发展,对包括人工智能、机器学习、信息处理、模式识别、无线传感网络、金融预测等学科的发展产生重要影响。同时对国防、遥感、机器人、智能交通、医学成像、安全检查和工业过程等众多应用领域的非线性滤波水平的提高也会产生很大的作用。

鉴于此,我们选择现代非线性滤波中的研究热点——粒子滤波编写本书,在理论方面,我们对粒子滤波算法的理论基础、重要性函数选择、重采样技术和滤波性能评估进行逐步叙述,初步形成粒子滤波的数学理论架构,并在这一理论框架下,从实际系统的应用和实现问题出发,给出了不同应用领域的范例。本书的宗旨是力图将我们在这方面的理解和研究成果进行逻辑概括,试图用尽量通俗的数学语言和观点,论述粒子滤波的基本内容,并且采用统一的一套术语和符号系统,以便读者阅读和理解。

本书基于作者从事非线性系统状态估计的个人体会和最新研究成果。在编写过程的结构安排上,我们力求注意以下几点:

(1) 全书分为理论与应用两部分,理论部分着重阐述粒子滤波的数学理论基础和实现方法、步骤,应用部分则介绍粒子滤波在实际系统中的具体应用;

(2) 在本书的开始部分,对粒子滤波的基本结构和相关知识作简要介绍,以保证本书的系统性,但取材着重于基本概念、基本理论和基本方法;

(3) 收集和总结了粒子滤波方法的最新动向、最新课题和最新成果,将重点置于粒子滤波的前沿研究及一些有潜力的新思想、新方法和新技术;

(4) 按粒子滤波算法的发展过程设立章节,使对粒子滤波了解较少的读者也能逐步认识和理解粒子滤波的原理和工作过程;

(5) 在应用部分给出了粒子滤波在不同应用领域的研究成果,提供了学习模仿的蓝本,以便读者能自己模仿实践,加深对理论的理解,体会粒子滤波算法的有效性和实用性。

由于粒子滤波算法涉及许多概率论和数理统计的知识,而本书又不是普通的教材,因此在阅读过程中需要读者具备概率论和最优化方法的基本知识。

本书的主要素材依赖于作者及其科研团队多年从事非线性系统状态融合与估计的研究积累,包含胡洪涛、李安平、张洪建、田宏伟、张晓宇等多名博士生和硕士

生的学位论文,并得到了课题组全体教师、研究生以及信息与控制研究所其他教师的关心和支持,在此一并向他们表示衷心的感谢!

本书工作先后得到国家自然科学基金“多传感器协同融合跟踪方法与应用(60674107)”、国家自然科学基金“基于 SNF 网络与多尺度变换的鲁棒多源动态图像融合研究(60375008)”,国家自然科学基金“基于视觉的近距离/超近距离空间非合作目标快速稳健识别跟踪研究(60775022)”国防基础研究项目、河北省自然科学基金重点项目“不确定性多源信息融合方法(F200600343)”、河北省首批百人计划项目“多源信息融合理论与应用”和中国博士后科学基金“多传感器协同融合跟踪与控制(B2004510)”等科研项目的共同资助。

由于粒子滤波算法仍在不断发展,书中内容表述和认识难免存在不足,殷切希望广大读者批评指正。

胡士强、敬忠良

2010年6月16日于上海交通大学

通信地址(邮编):上海市东川路800号上海交通大学航空航天学院(200240)

电话:(021)34206646

传真:(021)34206646

电子邮件:sqhu@sjtu.edu.cn,zljing@sjtu.edu.cn

主页:http://icaa.sjtu.edu.cn

目 录

前言

第 1 章 绪论	1
1.1 粒子滤波算法简介	1
1.2 粒子滤波算法的应用	7
1.3 粒子滤波算法的特点与发展.....	10
参考文献	13
第 2 章 动态系统状态模型与估计算法	16
2.1 动态系统状态模型.....	16
2.2 非线性系统估计算法.....	18
参考文献	20
第 3 章 随机采样与粒子滤波	21
3.1 贝叶斯估计理论.....	21
3.2 蒙特卡罗积分.....	21
3.3 重要性采样.....	22
3.4 序列重要性采样.....	23
3.5 粒子匮乏与重采样.....	24
3.6 次优重要性采样方法.....	26
3.7 随机采样滤波算法.....	27
3.8 算例仿真.....	29
参考文献	29
第 4 章 粒子滤波改进算法	31
4.1 重采样过程问题.....	31
4.2 粒子匮乏问题.....	49
4.3 计算复杂性问题.....	51
参考文献	52
第 5 章 目标定位与跟踪	53
5.1 机动目标跟踪中的非线性问题.....	53
5.2 纯角度跟踪粒子滤波算法.....	53
5.3 雷达非高斯观测噪声粒子滤波跟踪算法.....	64
5.4 闪烁噪声下 Rao-Blackwellised 化粒子滤波算法	72

5.5	可变多目标粒子滤波跟踪算法	76
5.6	本章小结	100
	参考文献	100
第6章	视频运动目标检测与跟踪	102
6.1	视频目标检测与跟踪问题	102
6.2	视频单目标多特征融合粒子滤波跟踪算法	105
6.3	概率模型视频粒子滤波跟踪算法	116
6.4	概率视频多目标跟踪算法	130
	参考文献	145
第7章	系统辨识与故障诊断	148
7.1	非线性系统辨识与变化检测	148
7.2	系统参数辨识	156
7.3	评标决策	158
	参考文献	164

第 1 章 绪 论

1.1 粒子滤波算法简介

近几年来,众多科研人员在机动目标跟踪、机器感知与导航、信号传输与压缩、金融领域数据分析、复杂工业过程故障诊断、人体轮廓跟踪与行为异常分析等研究方向几乎同时提出了多种基于贝叶斯估计思想的非线性滤波新方法,如: Bootstrap Filters^[1]、Condensation Tracks^[2]、Particle Filter^[3]、Sequential Monte Carlo Method^[4]、Interacting Particle Approximations^[5]等,这些算法名称不同,但都是基于贝叶斯抽样估计的序列重要抽样(sequential importance sample, SIS)滤波思想。早在 20 世纪 50 年代末,Hammersley 和 Morton 等就提出了基本的 SIS 方法^[6],并在六七十年代的自动控制等领域得到继续研究和发展^[7,8],但由于这些早期的研究主要集中在对重要性抽样的清晰化描述和重要性函数的选择上,而对于迭代过程中的粒子数匮乏现象和计算量制约等问题则始终没能解决,因此该方法并未引起人们太多的重视。1993 年由 Gordon 和 Salmond^[1]提出了一种新的基于 SIS 方法的 Bootstrap 非线性滤波方法,从此奠定了粒子滤波(particle filter)算法的基础,并引发了 SIS 的新一轮研究热潮。

所谓粒子滤波是指:通过寻找一组在状态空间中传播的随机样本,对概率密度函数 $p(x_k | z_k)$ 进行近似,以样本均值代替积分运算,从而获得状态的最小方差估计的过程,这些样本被形象地称为“粒子”,粒子滤波也由此得名。用数学语言可描述为,针对平稳的动态时变系统,假定 $k-1$ 时刻系统的后验概率密度为 $p(x_{k-1} | z_{k-1})$,依据一定原则选取 n 个随机样本点, k 时刻获得测量信息后,经过状态更新和时间更新过程, n 个粒子的后验概率密度近似为 $p(x_k | z_k)$,随着粒子数目的增加,粒子的概率密度函数就逐渐逼近状态的概率密度函数,粒子滤波器的估计就达到了最优贝叶斯估计的效果^[4]。

由于粒子滤波算法概念清晰,摆脱了解决非线性滤波问题所采用的传统扩展卡尔曼滤波算法的随机量必须满足高斯分布的制约条件,并在一定程度上解决了粒子数样本匮乏问题。随着计算机性能的快速提高,复杂算法计算量不再成为主要障碍,实时运算也变成了可能。因此,短短几年时间,粒子滤波算法便获得了很大的发展,并在许多领域得到成功应用。目前,“国际演化计算会议”、“国际信息融合会议”及众多网络讨论组都将粒子滤波作为专题进行深入讨论和学术

交流^[9~14]。

由于粒子滤波来源于贝叶斯抽样估计的 SIS 方法,因此先给出经典最优贝叶斯滤波算法的基本框架,进而再引入 SIS 方法和粒子滤波。

1. 最优贝叶斯滤波

假定动态时变系统的数学模型可以由式(1.1)和式(1.2)描述,系统的状态 $\{x_k, k \in \mathbf{N}\}$ 更新方程由式(1.1)描述

$$x_k = f_k(x_{k-1}, v_{k-1}) \quad (1.1)$$

其中, $f_x: \mathbf{R}^{n_x} \times \mathbf{R}^{n_v} \rightarrow \mathbf{R}^{n_x}$ 是系统状态 x_{k-1} 的非线性(或线性)函数; $\{v_{k-1}, k \in \mathbf{N}\}$ 是平稳的噪声序列; n_x 、 n_v 分别是状态和过程噪声的维数; N 为采样的次数,并且状态转移符合一阶 Markov 过程,即系统的当前状态仅依赖于状态的历史数据。

系统的测量方程为

$$z_k = h_k(x_k, n_k) \quad (1.2)$$

其中, $h_x: \mathbf{R}^{n_x} \times \mathbf{R}^{n_n} \rightarrow \mathbf{R}^{n_z}$ 是系统状态 x_k 的非线性(或线性)函数; $\{n_k, k \in \mathbf{N}\}$ 是平稳噪声序列; n_z 、 n_n 分别是测量值和测量噪声的维数。

从式(1.1)、式(1.2)可以看出,状态转移函数和观测函数没有做线性、非线性的假设条件,并且也没有对过程噪声和测量噪声作高斯白噪声的假设,系统状态估计的目的就是如何利用含有噪声的测量信息估计系统的状态 x_k ,即从式(1.2)描述的测量方程中结合式(1.1)递推估计系统状态 x_k 。

最优贝叶斯滤波算法的基本思想是:若已知状态的初始概率密度函数 $p(x_0)$,则可利用所有的测量信息 $Z_{1:k}$ 构造状态的后验概率密度函数 $p(x_k | Z_{1:k})$,从而可得到在任何准则下的理论最优估计和滤波值。采用数学语言可描述为

若已知状态的初始概率密度函数为 $p(x_0 | Z_0) = p(x_0)$,则

状态预测方程为

$$p(x_k | z_{1:k-1}) = \int p(x_k | x_{k-1}) p(x_{k-1} | z_{1:k-1}) dx_{k-1} \quad (1.3)$$

状态更新方程为

$$p(x_k | z_{1:k}) = \frac{p(z_k | x_k) p(x_k | z_{1:k-1})}{p(z_k | z_{1:k-1})} \quad (1.4)$$

其中

$$p(z_k | z_{1:k-1}) = \int p(z_k | x_k) p(x_k | z_{1:k-1}) dx_k \quad (1.5)$$

由式(1.3)看出,状态的一步预测概率密度(先验信息)可通过式(1.1)的状态转移函数获得,而通过式(1.2)的测量模型可获得似然函数 $p(z_k | x_k)$,则在式

(1.4)中可利用测量值修正一步预测的先验概率来得到当前状态所需要的后验概率。

所以式(1.3)、式(1.4)构成了最优贝叶斯估计的基本框架。式(1.3)中存在的积分运算,使得贝叶斯估计方法的具体实现遇到了很大困难,仅对于满足某些概率分布条件的动态系统状态估计问题,可以获得问题的解析解,如对于时变线性高斯分布的动态系统,利用贝叶斯估计的基本框架可以验证,卡尔曼滤波器就是贝叶斯估计的解析解,所以卡尔曼滤波器是最优的滤波估计器^[15],而对于非高斯分布的非线性动态系统的状态估计问题,贝叶斯估计的思想仅在理论分析和定性推理等方面得到应用。由于不能得到解析解,人们多采用近似线性化的扩展卡尔曼滤波等次优方法,对于具有较强非线性的时变动态系统,这些方法滤波存在精度较低和发散等问题,而传统的 SIS 方法受到计算量的制约,很不实用,所以对于非线性非高斯的动态系统,始终没有找到较好的解决办法。直到 Gordon 和 Salmond 提出了基于 SIS 思想的 bootstrap 自举滤波算法^[1],SIS 滤波方法才又开始引起人们的广泛关注。

虽然贝叶斯估计算法中的积分运算制约了其解析解的求取,但基于随机抽样运算的经典蒙特卡罗(Monte Carlo)方法可将积分运算变为有限样本点的求和运算。蒙特卡罗方法又称为随机抽样法或统计检验方法,其思想是假设能够独立从状态 $x_{0,k}$ 的概率分布密度函数 $p(x_{0,k} | Z_{1,k})$ 中抽取 N 个样本 $\{x_{0,k}^i, i = 1, 2, \dots, N\}$ ($x_{0,k} = \{x_0, \dots, x_k\}$), 则可以用如下经验概率分布来近似表述状态概率密度分布:

$$\hat{p}(x_{0,k} | Z_{1,k}) = \frac{1}{N} \sum_{i=1}^N \delta_{x_{0,k}^i} (dx_{0,k}) \quad (1.6)$$

其中, $\delta(\cdot)$ 为狄拉克函数。则 $f(x_{0,k})$ 的数学期望为

$$E(f(x_{0,k})) = \int f(x_{0,k}) p(x_{0,k} | z_{1,k}) dx_{0,k} \quad (1.7)$$

式(1.7)的积分运算可以用如下形式的估计值来逼近:

$$\overline{E(f(x_{0,k}))} = \frac{1}{N} \sum_{i=1}^N f(x_{0,k}^i) \quad (1.8)$$

式(1.8)的收敛性可由大数定律得到充分保证,并且估计误差近似服从均值为 0、方差为 σ^2 的正态分布。

2. 重要性抽样

经典蒙特卡罗方法的核心是将式(1.3)中的积分问题转化为有限样本点的概率转移累加过程,但在实际中由于 $p(x_k | z_{1,k})$ 可能是多变量、非标准的,通常不太可能写成典型的解析分布函数的组合形式,抽样过程比较困难。它的实现必须借

助一些抽样性算法,如重要性函数抽样方法、拒绝抽样方法^[16]、分层抽样方法^[17]等。所谓重要性函数就是指概率分布与 $p(x_k | z_{1:k})$ 相同,概率密度分布 $q(x_{0:k} | z_{1:k})$ 已知且容易从中抽样的分布函数,则式(1.7)可以变换为

$$\begin{aligned} E(f(x_{0,k})) &= \int f(x_{0,k}) \frac{p(x_{0,k} | z_{1,k})}{q(x_{0,k} | z_{1,k})} q(x_{0,k} | z_{1,k}) dx_{0,k} \\ &= \int f(x_{0,k}) \frac{p(z_{1,k} | x_{0,k}) p(x_{0,k})}{p(z_{1,k}) q(x_{0,k} | z_{1,k})} q(x_{0,k} | z_{1,k}) dx_{0,k} \\ &= \int f(x_{0,k}) \frac{w_k(x_{0,k})}{p(z_{1,k})} q(x_{0,k} | z_{1,k}) dx_{0,k} \end{aligned} \quad (1.9)$$

$$\overline{E(f(x_{0,k}))} = \frac{\frac{1}{N} \sum_{i=1}^N f(x_{0,k}^i) w_k^i(x_{0,k}^i)}{\frac{1}{N} \sum_{i=1}^N w_k(x_{0,k}^i)} = \sum_{i=1}^N f(x_{0,k}^i) \tilde{w}_k(x_{0,k}^i) \quad (1.10)$$

其中, $w_k(x_{0,k})$ 为未经过归一化的重要性权值; $\tilde{w}_k = \tilde{w}_k(x_{0,k}^i)$ 为归一化后重要性权值; $x_{0,k}^i$ 是 $q(x_{0,k} | z_{1,k})$ 中的独立采样点。 $w_k(x_{0,k})$ 可表示为

$$w_k(x_{0,k}) = \frac{p(z_{1,k} | x_{0,k}) p(x_{0,k})}{q(x_{0,k} | z_{1,k})} \quad (1.11)$$

重要性抽样是一种常用的蒙特卡罗积分方法,但是从算法实现过程中可以看出,在估计 $p(x_k | z_{1,k})$ 之前,需要得到 k 时刻以前所有的观测数据。并且在每一次接收到新的观测数据后,都需要重新计算整个状态序列的重要性权值,计算量随着时间增加而无限增加。因此这种简单的重要性抽样形式,不利于进行递归估计,在实际中更多的是采用能实现递归估计的序列重要性抽样算法。至于拒绝抽样方法和分层抽样方法^[3],由于这两种算法不能事先确定要获得个抽样需要迭代的次数,不利于实际的递推在线估计,因此使用较少。

3. 序列重要性抽样算法

序列重要性抽样(SIS)方法的核心是实现重要性抽样的递归估计。若将重要性函数 $q(x_{0,k} | z_{1,k})$ 写成以下连乘积形式:

$$q(x_{0,k} | z_{1,k}) = q(x_0) \prod_{j=1}^k q(x_j | x_{0,j-1}, z_{1,j}) \quad (1.12)$$

假设状态符合马尔可夫过程,在给定状态下,观测量条件独立,则可以得出

$$p(x_{0,k}) = p(x_0) \prod_{j=1}^k p(x_j | x_{j-1}) \text{ 和 } p(z_{1,k} | x_{0,k}) = \prod_{j=1}^k p(z_j | x_j) \quad (1.13)$$

将式(1.12)和式(1.13)代入式(1.11)得到权值递推公式

$$w_k = \frac{p(z_{1,k} | x_{0,k}) p(x_{0,k})}{q(x_k | x_{0,k-1}, z_{1,k}) q(x_{0,k-1} | z_{1,k})}$$

$$\begin{aligned}
 &= w_{k-1} \frac{p(z_{1,k} | x_{0,k})}{p(z_{1,k-1} | x_{0,k-1})} \frac{p(x_{0,k})}{p(x_{0,k-1})} \frac{1}{q(x_k | x_{0,k-1}, z_{1,k})} \\
 &= w_{k-1} \frac{p(z_k | x_k) p(x_k | x_{k-1})}{q(x_k | x_{0,k-1}, z_{1,k})} \quad (1.14)
 \end{aligned}$$

重新考虑式(1.3)、式(1.4)描述的贝叶斯估计问题,利用序列重抽样思想,可得到完整的递推 SIS。

假定状态量 x_k 初始时刻的概率密度函数满足 $P(x_0 | z_0) \equiv P(x_0)$, $k-1$ 时刻的概率密度函数为 $P(x_{k-1} | z_{1,k-1})$, 在时刻 k 获得测量信息 z_k , 则通过预测和更新的递归过程,即可构造概率密度函数 $P(x_k | z_{1,k})$, 利用 Chapman-Kolmogorov 等式有

$$P(x_k | z_{1,k-1}) = \int P(x_k | x_{k-1}) P(x_{k-1} | z_{1,k-1}) dx_{k-1} \quad (1.15)$$

从概率密度函数 $P(x_{k-1} | z_{1,k-1})$ 利用重抽样方法可以得到 N 个随机样本点 $\{x_{k-1}^i\}_{i=1}^N$, 若样本数量足够多,根据式(1.6),则概率密度函数可由式(1.16)表示为

$$P(x_{k-1} | z_{1,k-1}) = \sum_{i=1}^N w_{k-1}^i \delta(x_{k-1} - x_{k-1}^i) \quad (1.16)$$

根据贝叶斯估计理论,在 k 时刻状态 x 的概率密度函数为

$$P(x_k | z_{1,k}) \propto P(z_k | x_k) P(x_k | z_{1,k-1}) \quad (1.17)$$

将式(1.16)代入式(1.15),并利用式(1.10)可得

$$P(x_k | z_{1,k-1}) \approx \sum_{i=1}^N w_{k-1}^i P(x_k | x_{k-1}^i) \quad (1.18)$$

更新概率密度函数为

$$P(x_k | z_{1,k}) \approx \sum_{i=1}^N w_k^i \delta(x_k - x_k^i) \quad (1.19)$$

其中

$$w_k^i = w_{k-1}^i \frac{p(z_k | x_k^i) p(x_k^i | x_{k-1}^i)}{q(x_k^i | x_{k-1}^i, z_k)} \quad (1.20)$$

而样本点 $\{x_k^i\}_{i=1}^N$ 是由在 $k-1$ 时刻的样本点 $\{x_{k-1}^i\}_{i=1}^N$ 通过式(1.1)获得。

SIS 方法的实现步骤可总结如下:

- (1) 从 $q(x_k | x_{k-1}, z_k)$ 中随机抽取 n 个有限样本;
- (2) 逐点计算对应的 $p(x_k | x_{k-1})$ 和 $p(z_k | x_k)$;
- (3) 利用式(1.20)计算对应样本的重要性权系数;
- (4) 对权进行归一化处理如式(1.21)所示

$$w_k^i = \frac{w_k^i}{\sum_{j=1}^n w_k^j} \quad (1.21)$$

(5) 利用式(1.19)对 $p(x_k | z_{1:k})$ 进行估计。

许多粒子滤波算法都可以看做是 SIS 方法的特例,主要区别只是重要性函数的选择抽样策略的不同,重要性抽样的实质是蒙特卡罗积分的方差缩减方法,是对概率密度函数的加权近似。

对 SIS 算法而言,一个重要缺陷是算法存在粒子数匮乏现象。所谓粒子数匮乏是指:随着迭代次数增加,只有一个粒子的权重非常大(接近 1),而其他粒子权重变得非常小(接近 0),失去了粒子的多样性,从而很容易使 SIS 方法出现发散现象。Kong^[18]和 Doucet^[19]证明了在形如式(1.12)给出的重要性函数的对应权重的协方差是发散的结论,这从理论上说明 SIS 算法出现粒子数匮乏现象的必然性。正是匮乏现象的存在,使得 SIS 方法的发展至少停滞了 20 年。Liu 和 Chen 在 1998 年给出了一种衡量粒子数匮乏程度的方法^[20],定义了“有效粒子数” N_{eff} 为

$$N_{\text{eff}} = \frac{N}{1 + \text{Var}(w_k^{*i})} \quad (1.22)$$

其中, $w_k^{*i} = p(x_k^i | z_{1:k})/q(x_k^i | x_{k-1}^i, z_k)$, 并且 $N_{\text{eff}} \leq N$, 在实际中,经常使用

$$\hat{N}_{\text{eff}} = \frac{1}{\sum_{i=1}^N (w_k^i)^2} \quad (1.23)$$

其中, w_k^i 是由式(1.14)定义的归一化权值。显然为了提高滤波精度,可以选择很大数目的粒子数,但由此产生的计算量问题又制约了粒子滤波的实时性。在实际中为了降低粒子匮乏现象的影响,最有效的方法是选择重要性函数和采用重抽样方法而不是一味增加采样粒子数。

从式(1.22)可以看出,选取的最优原则就是使得重要性权重的方差最小。Liu 等^[21]证明了最优重要性函数为 $q(x_k | x_{0:k-1}, z_{1:k}) = p(x_k | x_{0:k-1}, z_{1:k})$ 。但采用最优重要性函数需要从 $p(x_k | x_{0:k-1}, z_{1:k})$ 抽样并计算积分。从应用角度看,除少数情况外^[18],多数重要性函数都采用了次优的 $q(x_k | x_{0:k-1}, z_{1:k}) = p(x_k | x_{k-1})$, 由于未能利用最新的测量信息,以该函数为重要性函数进行抽样所产生的方差比后验概率 $p(x_k | x_{0:k-1}, z_{1:k})$ 产生的方差大,但由于其容易实现,在粒子滤波算法中仍然得到广泛应用^[1,20,22]。

重抽样算法是降低粒子匮乏现象的另一种方法,其思想是通过粒子数和相应权表示的概率密度函数重新抽样,减少权值小的粒子数,增加权值较大的粒子数。最常用的重抽样方法是随机抽样方法^[20]。

随机抽样的过程是:首先产生 n 个在 $[0, 1]$ 上均匀分布的随机数 $\{\mu_l: l = 1, \dots, n\}$, 然后通过搜索算法找到满足以下条件的整数 m , 使得

$$\sum_{j=0}^{m-1} \tilde{w}_j < \mu_l \leq \sum_{j=0}^m \tilde{w}_j \quad (1.24)$$

记录样本 $x_k^{(m)}$, 并将其作为新样本集中的抽样, 将区间 $[0, 1]$ 按式(1.25)分成 n 个小区间, 当随机数 μ 落在第 m 个区间 $I_m = (\lambda_{m-1}, \lambda]$ 时, 对样本 $x_k^{(m)}$ 进行复制。

$$\lambda_j = \sum_{i=0}^j \tilde{w}_i, \quad i = 1, \dots, n \quad (1.25)$$

显然权值较大的样本将被多次复制 (\tilde{w}_m 大, 则区间 I_m 宽, $x_k^{(m)}$ 复制的次数多), 在抽样总数仍保持为 n 的情况下, 部分权值小的样本将被丢弃, 从而实现重抽样过程。重抽样过程以牺牲计算量和鲁棒性换取降低粒子数匮乏。在实际应用中, 许多算法采用如下简便重抽样方法, 利用式(1.23)衡量粒子匮乏程度, 当有效粒子数小于设定的指定门限时, 启动重抽样过程。

典型的 SIR 方法的实现步骤可总结如下:

(1) 初始化: 从 $p(x_0)$ 中随机抽取 n 个样本, $x_0^i \sim N(\hat{x}_{0|0}, P_{0|0}), \omega_0^i = \frac{1}{N}, i = 1, \dots, N;$

(2) 递推计算: $k = 1, 2, \dots;$

(3) 重要性抽样: 从 $q(x_k | x_{k-1}, z_k)$ 中随机抽取 n 个样本, 利用式(1.20)计算权值, 利用式(1.21)对权进行归一化处理;

(4) 重新抽样: 利用式(1.23)计算有效粒子数, 并与设定门限比较, 当有效粒子数小于门限值时, 启动重抽样过程, 令样本权值为 $\omega_k^i = \frac{1}{N};$

(5) 输出: 利用式(1.19)对 $p(x_k | z_{1:k})$ 进行估计。

1.2 粒子滤波算法的应用

粒子滤波的应用主要涉及机动目标跟踪、机器视觉、机器人导航与定位、通信数据压缩、语音信号处理、金融数据分析、复杂工业过程监视与控制、人体行为异常分析等, 本部分综述粒子滤波算法的主要应用范例。

1.2.1 机动目标跟踪

对目标进行定位和跟踪是典型的动态系统状态估计问题, 由于其在军事和民用上都得到广泛应用而成为当今研究的热点, 从本质上看, 目标跟踪最终可归结为系统状态方程和量测方程描述的目标运动模型的跟踪滤波器设计过程。在模型满足高斯线性条件下, 卡尔曼滤波器可获得最好的跟踪效果。然而, 在诸如纯角度跟踪及红外序列图像小目标检测等系统描述的运动模型中, 线性、高斯假设条件常常不能满足, 采用卡尔曼滤波器跟踪性能将下降。针对系统的非线性问题, 扩展卡尔曼滤波算法、伪线性滤波算法、修正的极坐标算法、Unscented 卡尔曼滤波等成为

最常用的近似方法,但采用这些改进算法引起的高阶截断误差和非高斯噪声都会影响算法收敛速度和跟踪精度。Gordon 等^[1]提出的粒子滤波方法应用到纯角度跟踪问题中,获得了较好的跟踪精度,随后针对目标跟踪问题出现了许多基于粒子滤波的研究成果,如纯角度跟踪的非线性滤波方法比较^[23,24]、多目标跟踪与传感器管理^[25]、基于粒子滤波的定位、导航与跟踪^[26]、采用粒子滤波方法的跟踪性能分析^[27]、多目标跟踪与数据融合^[28]、基于粒子滤波的群目标跟踪^[29]、基于 IMM-DF 的粒子机动目标跟踪算法^[30,31]等。

采用粒子滤波进行目标跟踪的研究成果表明,粒子滤波能较好地处理目标机动、测量断续、密集群目标跟踪、雷达跟踪多路径现象等传统难题,使得粒子滤波成为目标跟踪中非线性问题的最热门的数学支撑工具之一。虽然粒子滤波能有效解决跟踪中的非高斯、非线性系统的滤波问题,但由于其算法运算的实时性问题、状态初始概率选取问题,粒子滤波距离工程应用尚有一定差距,但将粒子滤波与其他方法相结合,选择合适的重要性函数和重采样方法,避免非线性系统的近似线性化,并减小非线性系统线性化后的高阶截断误差和非高斯噪声带来的影响,是跟踪问题中采用粒子滤波方法努力研究的方向。

1.2.2 金融领域数据分析

将粒子滤波思想应用于金融领域的数据分析和建模研究尚处于初级探索阶段^[32~34]。金融领域的许多时间序列数据都可归结为模糊随机时变系统,如证券投资走向分析、国家经济发展金融风险防范、风险投资最优决策、银行贷款风险和最佳收益率等问题无一不依赖于积累的各种历史数据。文献[32]针对常用来描述金融市场波动的随机波动模型(stochastic volatility model)的特点——模型具有明显的非线性和非高斯的“厚尾”特性,采用粒子滤波方法计算了 SV 模型参数的似然函数和边界似然函数,并对真实历史数据进行了对比分析,说明了粒子滤波算法在计算 SV 模型参数方面的优势;文献[33]则从滤波和跟踪角度对随机波动模型进行了分析:首先将随机波动模型转变为状态和测量模型的滤波形式,然后采用交互式粒子滤波方法对由随机波动模型描述的股票交易数据进行预测分析,通过仿真和对比真实历史数据看出,预测数据与真实数据非常接近;文献[34]采用粒子滤波方法研究了风险估值问题(value at risk),文章采用分位数重要性重抽样方法给出了利率的最佳估计,进而可确定最合理的投资组合,同时给出了在 VaR 中协方差阵的收敛性估计值。虽然采用粒子滤波方法处理金融领域中的估值问题刚开始,但由文献[32]~[34]可以看出,粒子滤波在处理基于随机波动模型和风险估值方面的金融数据有着良好的应用前景。

1.2.3 计算机视觉

近年来,基于序列图像信息的智能信息处理技术——机器视觉的研究不断升

温,其中对感兴趣区域进行持续跟踪由于在智能机器人导航、视频安全监控、医学图像分析和计算机人-机接口等领域的潜在应用价值而引起人们的广泛关注。目前图像跟踪的难点在于:① 由于复杂背景的影响,在目标运动区域提取(目标的出现与消失)方面缺乏有效的手段;② 由于目标被障碍物遮挡的影响,测量数据出现断续现象,基于扩展卡尔曼滤波等方法出现失跟等现象;③ 密集人群的线路交叉跟踪问题。对于这些难点问题,从建模角度可认为系统属于非高斯、非线性的时变系统,只有采用对非线性系统处理能力较好的滤波方法才能有效克服这些困难。文献[35]率先在计算机视觉领域引入了粒子滤波算法,其典型应用是如何跟踪一片晃动的树叶,随后文献[36]较系统地研究了采用粒子滤波算法如何实现目标的轮廓跟踪,其研究重点是如何选择重要性函数和避免粒子数的匮乏现象,文献[37]~[39]则在抽样方法上进行了改进,力图使初始的概率密度更接近状态的真实概率密度分布,文献[40]、[41]则构造了新的目标检测的动态模型并与粒子滤波相结合,解决在目标运动区域提取方面遇到的问题,文献[42]将曲线拟合思想引入粒子滤波方法-提出了主动粒子滤波方法,重点解决粒子数目众多的问题。文献[43]较系统地研究了粒子滤波方法在计算机视觉中的应用,从中可系统了解粒子滤波在计算机视觉中的应用成果和最新发展趋势。

1.2.4 状态监视与故障诊断

对复杂过程的状态进行实时监视在工业过程控制、航空航天器运行状态估计和故障诊断、汽车防撞、人体移动路线异常分析等的研究意义非常重大。对状态进行监视的实质就是状态的在线估计过程,故障诊断过程就是在一定准则下的状态异常检测过程。由于有些动态过程比较复杂,若仍采用基于线性高斯白噪声假设条件,则不能准确描述动态过程的状态真实变化过程,采用非高斯非线性假设条件,常规滤波方法的状态估计精度存在问题,影响故障诊断的效果。文献[44]研究了采用粒子滤波方法对复杂工业过程中的热交换器特征参数结合 EM(expect maximum)算法进行辨识,并进行实时控制。实际控制效果曲线显示:采用同一种控制算法,进行粒子滤波特征参数估计的控制效果明显优于未进行参数估计的算法。文献[45]给出了粒子滤波在汽车辅助避碰系统中的应用思想。为了确切描述汽车行驶过程中周围的目标的状态,需要对目标位置和速度进行跟踪和估计,但由于测量噪声的影响,基于双峰高斯假设更能反映其特性。文章同时给出了基于卡尔曼滤波和基于粒子滤波两种跟踪算法,结果显示采用粒子滤波算法更容易实现辅助避碰决策的确定。文献[46]则研究了不完全量测信息下非线性模型人体行为异常检测问题,利用粒子滤波方法建立了状态“缓慢”变化判定的统计量,并以机场停机坪旁行人的异常行走路线为仿真算例,说明了采用粒子滤波方法处理判定状态“缓慢”变化的有效性。

将粒子滤波算法引入状态监视和故障诊断领域目前进展较为缓慢^[47],其主要原因在于如何根据复杂的过程建立合理的数学模型,并在选择合适的重要性函数和似然函数时缺乏理论依据。粒子滤波在故障诊断中的应用充满了机遇和挑战。

1.3 粒子滤波算法的特点与发展

随着粒子滤波方法研究的不断深入和取得众多研究应用成果,许多人错误地认为:在解决所有状态估计的滤波问题时,获得滤波性能最好的办法就是采用粒子滤波,甚至优于采用卡尔曼滤波方法。其实和其他非线性滤波方法一样,粒子滤波作为处理非线性系统状态估计问题的一种方法,也存在着算法自身的适应性问题 and 估计精度问题。从贝叶斯最优估计的思想很容易看出,卡尔曼滤波方法是线性高斯假设条件下贝叶斯估计的解析形式,是线性高斯动态系统的最优滤波器,而粒子滤波作为抽样贝叶斯估计算法,只是随着抽样粒子数的不断增大,逐渐趋向状态的后验概率密度,所以就理论基础而言,粒子滤波算法与其他非线性滤波方法一样(如扩展卡尔曼滤波、Unscented 卡尔曼滤波等)也是一种次优的滤波方法。卡尔曼滤波方法是在回归假设下由测量信息根据模型递推得到的状态估计的最优解析解,而粒子滤波是依据抽样思想由状态预测测量信息根据贝叶斯估计递推得到的状态估计的后验概率。通过与扩展卡尔曼滤波方法、Unscented 卡尔曼滤波方法的滤波思想的比较,可看出粒子滤波的适应范围和解决非高斯分布系统时的独特优势。

1.3.1 扩展卡尔曼滤波

扩展卡尔曼滤波(extended Kalman filter, EKF)方法处理非线性系统的思想是将非线性函数在估计点附近进行泰勒级数展开,并用一个等价于常规卡尔曼滤波方程的近似矩阵来代替非线性函数。对于由式(1.1)和式(1.2)描述的噪声满足高斯分布的线性系统,卡尔曼滤波可以准确计算最优预测值,但对于非线性高斯系统,扩展卡尔曼滤波的近似对应值为

$$\hat{x}_k \approx f(\hat{x}_{k-1}, \bar{v}), \quad \hat{z}_k \approx h(\hat{x}_k, \bar{n}), \quad \kappa_k \approx \hat{P}_{x_k x_k} \hat{P}_{z_k z_k}^{-1} \quad (1.26)$$

其中, $\bar{v} = E[v]$, $\bar{n} = E[n]$ 分别表示过程噪声和量测噪声的数学期望(即平均值,通常为0),式(1.26)的含义指:目标状态 x_k 的预测值可近似认为是满足状态方程的上一时刻估计的平均值的解;目标观测值 z_k 的预测值可近似认为是满足量测方程的上一时刻状态估计的平均值的解;协方差阵由近似线性化状态方程和量测方程确定。扩展卡尔曼滤波方法也可以理解为:将非线性方程近似认为是叠加白噪声后的一阶泰勒近似的线性化后的标准卡尔曼滤波方法。但是,这种一阶近似方