

鲍居武 编著



仿真数值分析 方法及其应用

学苑出版社

计算机程序设计语言系列丛书

仿真数值分析方法及其应用

鲍居武 编著
张鸿端 审校

学苑出版社
1993.

(京)新登字 151 号

内 容 提 要

仿真输出数据分析是当前仿真界最活跃的研究领域之一。有效的仿真输出数据分析是对所研究系统做出正确评价和辅助管理决策的基础,它具有重要的理论意义和明显的应用价值。本书系统地讲述了仿真输出数据分析理论及其应用。内容共分四部分,即基于经典统计分析理论的仿真输出分析方法、基于神经网络的仿真输出分析方法、仿真输出数据的分布识别和应用案例分析。本书内容丰富、深入浅出、资料新颖、实用性强,是指导读者学习和应用仿真技术的不可多得的一本重要参考书。

欲购本书的用户,请直接与北京 8721 信箱资料部联系,电话 2562329, 邮码 100080。

计算机程序设计语言系列丛书

仿真数值分析方法及其应用

编 著: 鲍居武

审 校: 张鸿端

责任编辑: 徐建军

出版发行: 学苑出版社 邮政编码: 100032

社 址: 北京市西城区成方街 33 号

印 刷: 北京市朝阳区小红门印刷厂

开 本: 787 × 1092 1/16

印 张: 11.63 字 数: 267 千字

印 数: 1~3000 册

版 次: 1993 年 12 月北京第 1 版第 1 次

ISBN7-5077-0807-1/TP·18

本册定价: 15.00 元

学苑版图书印、装错误可随时退换

前 言

仿真输出数据分析是当前仿真界最活跃的研究领域之一。随着仿真技术的广泛应用，作为整个仿真研究中的关键环节的仿真输出数据分析技术愈来愈受到人们的重视并得到发展。有效的仿真输出数据分析是对所研究系统做出正确评价和辅助管理决策的基础，它具有重要的理论意义和明显的应用价值。

本书对仿真输出数据分析理论与实践进行了系统的研究。本书在回顾与分析传统的仿真输出数据分析技术的基础上，提出从统计理论、人工神经网络和模糊数学等多种技术途径来进行仿真输出数据的分析，以便把多种技术途径结合起来，更好地解决仿真输出分析的问题。本书共分四部分，即基于经典统计分析理论的仿真输出分析理论、基于神经网络的仿真输出分析理论、仿真输出数据的分布识别和应用实例研究。本书的最后还给出了作者开发的经过充分检验的通用分布识别软件系统的源程序和通用的 BP 神经网络的源程序，相信这一切都会对广大读者有所帮助。

总之，本书倾注了作者大量的心血，系作者近年来的研究成果。理论与应用相结合是本书的一大特色。它内容丰富、资料新颖，是指导读者学习和应用计算机仿真技术的重要参考书。

在本书的编写过程中，孙学成、鲍居源、鲍居仕、刘青云、曹军、李桃等参加了编写工作，在此表示衷心地感谢。

由于我们水平有限，书中难免会有缺点和错误，敬请广大读者批评指正，以便进一步修改。

作者于北京
1993年11月

目 录

第一章 概 述	1
1.1 问题提出	1
1.2 国内外研究概况	3
1.3 研究内容概要	6
第二章 稳态仿真输出分析方法的研究	9
2.1 引言	9
2.2 稳态仿真输出分析方法综合评述	10
2.2.1 终态仿真与稳态仿真	10
2.2.2 稳态仿真分析方法	11
2.3 批平均法的改进	15
2.3.1 问题的描述	16
2.3.2 随机序列独立性检验法	18
2.3.3 改进批平均法	19
2.4 间隔抽样法	21
2.4.1 间隔抽样法	22
2.4.2 相关性检验法	23
2.5 实例分析	26
2.6 小结	28
第三章 小样本或单次运行的仿真输出分析方法	29
3.1 引言	29
3.2 正态 Bayes 方法	30
3.3 正态伽玛 Bayes 方法	31
3.4 伽玛 Bayes 方法	32
3.5 Bayes 方法中先验分布的确定方法	34
3.6 实例研究	35
3.7 小结	37
第四章 随机系统仿真中的对偶变量和公共变量法	38
4.1 引言	38
4.2 公共和对偶变量法的数学描述	38
4.2.1 一维情况下的数学描述	39
4.2.2 多维情况下的数学描述	39
4.3 公共和对偶变量法的理论分析	41
4.3.1 具有独立分量的多维随机系统的方差衰减	41

4.3.2	具有相关分量的多维随机系统的方差衰减	47
4.3.3	其它相关类型情况下多维随机系统的方差衰减	48
4.4	公共和对偶变量法的应用分析	49
4.4.1	一般描述	49
4.4.2	在网络和可靠性中的应用	50
4.4.3	排队系统	52
4.5	命题证明	53
4.6	小结	58
第五章	BP 神经网络的研究和改进	59
5.1	引言	59
5.2	MLP-BP 神经网络的研究	60
5.2.1	BP 算法	61
5.3	MLP-BP 神经网络的改进	63
5.3.1	改进算法一: 基于改进的 Sigmoid 型函数的 BP 算法	63
5.3.2	改进算法二: 基于聚类分析的改进 MLP-BP 神经网络	65
5.3.3	改进算法三: 基于变步长的 BP 算法	66
5.3.4	改进算法四: 基于共轭梯度法和一维搜索法的改进 BP 算法	67
5.4	小结	69
第六章	基于神经网络的仿真输出分析	71
6.1	引言	71
6.2	基于神经网络的仿真输出分析的总体框架	72
6.2.1	神经网络作为仿真输出数据分布识别的分类器	72
6.2.2	神经网络代替人来完成仿真输出的后期分析	73
6.2.3	神经网络作为仿真黑箱系统(黑箱辨识)的一种新途径	73
6.3	基于神经网络的仿真输出数据的分布识别	75
6.3.1	问题的提出	75
6.3.2	基于神经网络的分布识别	75
6.3.3	特征提取	76
6.3.4	MLP-BP 分布识别器设计	77
6.3.5	实验检验与实例分析	78
6.4	基于神经网络的智能化的仿真输出后期分析	79
6.4.1	基于神经网络的后期分析系统	79
6.4.2	实例研究	80
6.5	小结	84
第七章	仿真输出数据分布形式的模糊识别	86
7.1	引言	86
7.2	仿真输出数据分布形式的模糊识别模型	86
7.3	特征量的提取与隶属函数确定	87
7.4	样本量对隶属度和识别结果的影响	89

7.5 实验检验与实例分析	91
7.5.1 软件实现	91
7.5.2 实验检验	91
7.5.3 实例分析	93
7.6 小结	93
第八章 仿真输出分析的应用实例研究	94
8.1 实际系统的描述与仿真研究任务分析	94
8.2 实际系统的仿真模型的建立	99
8.2.1 针阀体的生产工序的 SLAM 网络图	99
8.2.2 仿真建模中的问题分析和处理方法	100
8.3 仿真实验与输出结果分析	103
8.3.1 原有生产线的仿真运行与瓶颈分析	103
8.3.2 生产线的调整与瓶颈排除	105
8.3.4 仿真结果的输出分析	108
8.4 小结	116
第九章 结论与展望	117
9.1 本书的主要工作与结论	117
9.2 进一步的设想和展望	118
参考文献	120
附录 A 通用 BP 神经网络算法源程序	133
附录 B 通用分布形式模糊识别软件	147
附录 C 生产线的 SLAM 仿真模型	175

第一章 概 述

1.1 问题提出

问题求解是人们进行科学分析和决策评价的基础。而建模，亦即被研究问题的表示与获取，则是问题求解的关键步骤之一。

当问题的模型关系相当简单时，可用各种数学方法（如代数、微积分或概率论）来精确求解感兴趣的问题，这就是所谓的解析解。但是，由于大多数实际系统太复杂，不允许解析地计算实际模型。这时可以用仿真的方法进行研究，用计算机从数值上估量模型，并根据所得数据来估算模型各种期望的真正特性。

系统仿真是指建立一个系统的数学逻辑模型，并且对该模型在计算机上进行实验处理，通过对系统的动态特性的观测，以研究系统行为的过程。这里包含三方面的含义：(1) 系统仿真的目的是预测系统的动态特性，以研究系统的行为；(2) 系统的建模过程，也即系统的数学逻辑模型的表示，计算机程序的准备等；(3) 系统实验，也即进行设计，系统的仿真运行及输出分析等。

随着计算机技术的发展，计算机仿真作为一种求解复杂问题的有效方法越来越得到非常广泛的应用。特别是在应用实际系统进行实验比较昂贵、比较冒险或是具有很大破坏性时，计算机仿真就更显示出其重要的实用价值。它可以用来对一个新系统的设计进行评估，也可以对一个已有系统的改造进行评估，从而得到较好的方案。在实际工作中，一个好的方案所带来的效益是不容忽视的。

仿真的应用领域非常广泛，下面列出一些应用非常成功的领域：制造系统的设计和分析；计算机系统所需要的软硬件的评估；一种新的武器系统和战术的评估；库存系统定货策略的确定；通讯系统的设计；运输设施（如飞机场、地铁、港口等）的设计；服务系统（如医院、邮局等）的设计评估；经济系统的分析等等。

许多统计资料表明(如 Forgionne (1983)^[87], Harpell (1989)^[110])，仿真仅次于统计分析是当今利用最广泛的实用技术之一。随着计算机功能和仿真软件的改进，仿真的价值与用途将会继续增加。

一般情况下，仿真研究的步骤和内容如下，在 Banks(1984)^[46]和 Law(1991)^[154]中，也有类似的论述。

1. 问题说明：对所研究的系统的一般描述，系统边界的基本定义等。
2. 研究目的：明确仿真研究应解决的问题。
3. 模型建立：确定系统中的实体，每个实体的属性以及从事的活动，系统的可能状态等，明确系统的基本特性，选择和修正系统的基本假设，最终建立一个刻画被研究系统行为，满足研究目标的模型。
4. 数据收集：收集所研究系统的数据，用来估计输入参数以及其概率分布等。

5.程序编码: 把系统模型转换成为可执行的计算程序。

6.程序验证: 确定一个仿真模型能否按预期的要求执行, 也即验证程序与模型的一致性。

7.模型确认: 比较模型所表现的系统行为和实际观察到的系统行为之间的差异, 缩小这些差异以提高模型表现被研究系统的能力。程序验证和模型确认的不断重复, 直到最终建立一个能刻画实际系统, 并能正确执行的仿真模型。

8.试验设计: 确定各种可选方案, 选择主要的输入变量以及试验水平, 给定仿真运行的长度和重复次数。

9.运行与分析: 评价特定输入变量对系统性能的影响, 仿真输出数据的分析和决定参数的优化。

10.仿真报告: 编写仿真输出综合报告, 辅助分析者作出管理决策。

以上十个步骤概括了仿真研究的基本步骤和内容。

仿真研究的最终目的是评价系统行为, 辅助管理部门作出管理决策, 这是仿真研究的本质, 而仿真输出数据的分析是辅助决策的基础和依据, 是仿真研究的核心内容之一。

仿真运行是为了了解所研究系统的行为, 这种了解是对系统进行决策的重要输入量。通常, 系统的行为可归结为一个和多个参数的值, 例如平均等待时间, 资源的平均利用率等等。仿真输出分析的目的是采用适当的统计分析技术, 根据仿真过程中收集到的输出数据对这些参数进行估计。

不幸的是, 许多仿真运行没有对输出数据进行适当的统计分析。

Law(1982)^[149]指出, 在许多仿真研究中, 人们往往把大量的时间和金钱花费在模型开发和编程上, 而很少花费一点精力来合理地分析一下仿真输出数据。实际上通常的作法是作一次任意长的仿真运行, 然后把仿真结果的估计作为模型特性的真值。由于概率分布的随机抽样是典型的推动仿真模型的方法, 因此这些估计仅是随机变量的特定实现, 且具有很大方差。这样, 一次特定仿真运行的估计可能与模型对应的真值有很大不同, 从而导致对所研究系统作出错误的推断。例如分析一个制造系统在两种调度规则下工件的平均延迟时间。对于规则一, 完成 1000 个工件的平均延迟时间分别为 1.2 天; 而对于规则二则是 2.3 天。如果仿真中的所有输入都是确定性的, 则输出必然是确定性的, 那么我们可以说规则一绝对比规则二所产生的平均延迟时间小。但是不幸的是几乎所有仿真运行都带有一些随机输入, 因此输出也是随机的。在这种情况下, 我们可以说在这一特定的情况规则一提供了较优的性能, 但如果没有附加的数据分析, 我们不能作出更一般的结论。这种附加的分析需要估计抽样结果的变化性。变化性通常是通过置信区间来表示的, 例如: “规则一下工件的平均延迟时间是 1.2 ± 0.6 天”, 或者通过一个概率描述表示, 如 “规则一下工件的平均延迟时间比规则二下工件平均延迟时间小的概率是 0.994”。如果不用置信区间或概率描述, 我们不能判断抽样结果是否具有代表性。

历史上, 有几种原因导致输出数据分析不能以合适的方法进行。首先是用户对仿真有一种非常错误的印象, 即认为仿真仅是计算机编程的一种练习, 只不过复杂一点而已。这样仿真研究便开始于模型的建立和编程, 结束于程序的一次运行产生的“解”。但是事实上, 仿真是一种基于计算机的统计抽样实验。因此如果仿真研究的结果要有意义的话, 就必须使用适当的统计技术来设计和分析仿真实验。第二个限制统计分析技术应用的原因是

几乎所有仿真的输出过程都是不一致的和自相关的。这样典型的基于独立同分布 (IID) 观察的统计技术就不能直接应用。直到今天, 仍然有一些输出分析问题没有可完全接受的解决方法, 或者即便有, 也因太复杂以致难以应用。例如稳态仿真输出分析的问题, 小样本或单次仿真运行的输出分析的问题, 复杂系统仿真中的方差衰减问题等等。我们将在下一节进一步分析这方面的问题。第三个限制获得模型特性真值的精确估计的原因是收集必要数量的仿真输出数据的计算机机时成本。特别是在有合适的统计分析方法可以采用的情况下, 收集所需数据的计算机机时成本就显得更为关键。当然这种困难随着计算机的发展将会减小。

另外, 传统的仿真输出分析是基于统计分析的经典理论。由于统计分析方法的局限性, 使得仿真输出分析中的一些问题没有得到很好解决, 例如仿真输出数据的分布识别和仿真输出的后期分析问题。我们是否可以把其他相关理论和技术 (如人工神经网络和模糊数学) 引入仿真输出分析领域中, 同统计技术一起来对仿真输出数据进行有效的分析, 这是十分令人感兴趣的问题。

综上所述, 仿真输出数据的分析是整个仿真过程中极其重要的一环, 如果不能使用合适的分析方法对仿真输出数据进行准确的分析, 不管仿真模型建立的多么精确, 不管仿真的次数多么大, 都是不能达到正确地辅助分析者进行管理决策的最终目的。

八十年代以来, 系统仿真已在国内科技, 生产和管理等领域中逐步得到应用, 但大多停留在仿真单运行的水平上, 从而仿真结果具有很大的偏差, 往往造成科技和管理决策的失误。近年来, 对仿真的统计性质和仿真输出数据的数学分析得到国内外学术界和工业界的广泛关注, 新颖的仿真统计分析理论正在形成。

离散系统仿真输出分析理论将仿真原理与仿真实验设计, 随机过程理论, 数据处理的统计分析方法以及其它相关理论和技术等结合起来, 从而能由仿真结果的样本参数对系统的总体参数进行有效的统计推断, 得到较高的置信度和仿真精度, 以辅助可靠的决策。这一研究将在理论上使系统仿真更加完善。在应用上提高其置信程度。

仿真输出分析的研究成果将能显著地丰富离散系统仿真的理论内涵, 并可支持管理系统仿真的有效应用, 减少决策失误。因此本研究课题既具有重要的学术意义, 又具有明显的实用价值。

1.2 国内外研究概况

本节我们将对仿真及其输出分析理论的发展和研究状况作一简述, 并对本文涉及的相关学科人工神经网络和模糊数学的发展状况也作了简要的论述。以后各章中将作进一步研究。

仿真又称蒙特卡罗方法, 它是通过用随机数做实验求解随机问题的一种技术。这种方法可以追溯到 1773 年法国物理学家 G.L.L. Buffon 为了估计 π 的值所进行的物理实验。然而, 第一个利用这种方法做随机数实验的人也许是美国统计学家 E.L. De Forest, 那是在 1876 年 (参考 Gentle (1985) ^[94], P.612)。比较早而且著名的蒙特卡罗法使用者是 W.S. Gosset。在 1908 年以“Student”为笔名发表论文时, 他使用了蒙特卡罗法来证明他的 t 分布法; 在这之前已经由“theory”发展了 t -分布法, 当然还不是十分精确。尽管蒙特卡

罗法起源于 1876 年,但是直到约 75 年之后 S.Ulam 和 J.Von Neumann 才将它命名为蒙特卡罗法(参考 Ulam (1976)^[196]中的描述)。为什么这么多年过去了它才被命名?其原因是直到数字计算机出现之前,这种方法在许多重要问题上都不能被加以利用。从 1946 年到 1952 年间,数字计算机在一些科研机构得到发展。这些机构有:宾夕法尼亚大学、麻省理工学院、美国国家标准局和国际商用机器公司。现代化的程序储存计算机使冗长的计算成为可能,而这种计算正是蒙特卡罗法所要求的。

与今天的计算机相比,早期的计算机运算速度慢而且几乎不能储存任何东西。比如,在本世纪 60 年代初期,利用计算机进行算术运算,每秒钟少于 10000 次,在 60 年代中期达到约 500000 次,到 70 年代初期达到 20000000 次,而如今的超级计算机每秒钟进行同样的计算可超过 10 亿次。另外,直到 1955 年,程序的运行还是在机器内部由汇编语言完成的,因为当时还没出现诸如 FORTRAN 之类的高级计算机语言;具有特殊目的的仿真语言也只是在约十年之后,即约 1965 年才出现。如今被人们称作为仿真的蒙特卡罗法最初的运用主要是集中在编程技巧上,因为在发展仿真方面,调试和运行程序是最难的工作。早期计算机的局限性往往使得问题过于简化;因为如果不这样简化的话,就不可能在可行的计算时间内以可行的成本完成程序的运行。因此,许多重要的问题,诸如运行哪种程序及怎样分析程序输出等,就被忽略了。

自六十年代中期至七十年代末以来,系统仿真一直处于仿真软件开发和建模的水平上,从而出现徘徊不前的状态。八十年代初,随着计算机技术的发展,仿真技术得到了飞速发展。仿真输出分析成为仿真领域中最活跃的研究方向之一,受到国内外学者的极大关注。特别是一九八三年,美国 Arizona 大学的 Averil M. Law 发表了论文“Statistical Analysis of Simulation Output Data”,这是第一篇系统论述仿真响应统计理论的文章,它为仿真输出分析开辟了一个新纪元。从一九八四年到一九八八年,各国学者和工业界人士进行了多方面的研究和应用,发表了大量论文和研究报告,其中比较有影响的论著和论文有 Kleijnen(1974,1975)^[133-134], Fishman(1978)^[85], Law, Kelton(1982)^[149], Rubinstein(1981)^[176], Bratley, Fox, Schrage(1987)^[54], Lewis, Orav(1989)^[155], Kleijnen(1982)^[136], Law(1983)^[147], Welch(1983)^[200], Law(1991)^[154], Scila(1990)^[193]等等。它们对仿真响应的置信区间,随机数发生器的算法,稳态与终态仿真的仿真精度,仿真响应的相关性等进行了论述和应用,取得显著的进展。八十年代末期美国 Purduc 大学以“Statistical Analysis For Stochastic Modeling and Simulation”为题,进行了专题研究,提出了多响应变量仿真,小样本仿真理论,控制变量方差衰减理论等,对提高系统仿真的精度起了重要的推动作用。九十年代初,美国 Harvard 大学何毓崎教授等提出了离散事件动态系统理论,单次重复运行仿真概念,无穷小扰动分析理论等,从而可用单运行仿真响应估算总体响应,为提高仿真效率作出了贡献。我国自八十年代初期开始研究离散系统仿真的建模与应用,在仿真优化,专家仿真系统和面向对象的仿真等方面进行了初步研究,但在仿真统计分析方面却很少研究,仿真结果置信度很低,机时消耗极大,限制了仿真技术的应用和推广。因此,在具有一定应用规模的条件下,开展系统仿真统计理论的研究和开发相应的软件系统,对于推动仿真应用具有重要意义。

但是仿真输出分析理论的发展还很不完善。仿真输出分析目前仅是基于统计分析理论,而统计分析理论的一些局限性,使得仿真输出分析中的一些问题还没有得到很好地解

决。当前仿真输出分析领域中十分令人关注的几个问题是:

1. 稳态仿真输出分析的问题。由于稳态仿真输出数据是自相关的, 使得经典的基于独立观察数据的统计分析方法不能应用, 因此如何估计稳态输出数据的相关性, 如何对稳态参数进行统计推断是当前一个十分活跃的研究方向。为了解决这个问题提出了许多不同的方法, 如多次重复运行法, 批平均法, 参数建模法, 频谱分析法和标准时间序列法等。每一种方法都有各自的分析观点, 但是许多仿真实验指出 (Law(1983)^[147], Fishman(1978)^[85]等), 上述方法或是难于应用, 或是进行分析时所得的区间估计的置信度偏低, 或是需要凭经验决定。因此这一问题仍是值得研究的方向之一。

2. 小样本或单次仿真运行的问题。在许多实际系统中, 希望在短期内得到某一控制策略(和方案)的响应变化情况。对这些问题, 详细的仿真研究, 由于需要大量时间往往不可行, 从实际需要出发, 对于每种选择, 单次或小样本仿真运行可能正是我们所期望的。因此单次和小样本仿真运行的输出分析有着重要的理论和实际意义。对于这种情况, 传统的统计分析方法(如多次重复运行法等)不能使用, 何毓崎教授等提出用 DEDS(离散事件动态系统)方法解决这一问题。但是寻找适用于单次和小样本仿真运行的统计分析方法依然是仿真输出分析领域中十分关注的问题之一。

3. 复杂系统仿真中的方差衰减技术问题。在复杂的仿真系统中, 获得精确的估计是仿真的重要目的之一。在估计系统输出参数时为了改进精度, 人们使用了方差衰减技术。但是由于方差衰减的效果与所研究的系统直接相关(参见 Law(1982)^[149]), 因而制约了方差衰减技术的发展。目前方差衰减技术的应用依然处于实验水平, 它的内部机理还缺乏全面的深入地理论描述和分析。但是由于方差衰减技术的巨大效益, 使得它从八十年代以来一直是比较活跃的研究方向之一。

4. 仿真输出数据的分布识别问题。样本总体的概率分布形式是描述样本总体特性的一个主要的性能测度。分布识别在仿真输出数据分析中有着重要的理论和实际价值。传统的基于统计分析方法的分布识别, 由于 χ^2 检验和 K-S 检验方法的局限性, 使得对仿真输出数据的分布识别长期以来没有很好解决。而近年来, 模糊数学在模糊识别中的发展和人工神经网络在分类器中的应用, 为解决分布识别问题提供了一些新的技术途径。

5. 仿真输出的后期分析和智能化仿真输出分析的问题。传统的仿真输出分析是基于经典的数理统计理论和人的直观判断。存在的缺陷是仿真输出分析仅限于前期分析, 而后期分析则是由人直观判决。这使仿真输出分析未能构成一个自动的智能化的分析过程。而实现智能化的仿真输出分析则是人们所期望的。

人工神经网络是人类神经系统的某种简化、抽象和模拟, 是一种更接近于人的认知过程的计算模型。它的自组织自学习能力、非线性逼近能力等启发我们是否可用它来代替人来进行仿真输出的后期分析, 从而实现智能化仿真输出分析。这一研究方向显然具有重要的学术意义和明显的实用价值。

以上我们给出了仿真输出分析领域中当前一些主要的研究方向。除此之外, 多响应仿真输出分析的问题、多方案比较的问题等也都是比较重要的研究方向, 这里就不再赘述了。

自 Zadeh 1965 年发表第一篇模糊集论文以来, 模糊理论已取得令人瞩目的成就。模糊数学已在人工智能、知识表示, 聚类分析, 模式识别和控制等领域得到非常成功的应

用。把模糊理论的一些思想和方法吸收到仿真输出分析理论中，以弥补统计分析理论的不足正是我们所期望的。

人工神经网络是当前十分引人注目的交叉学科之一。它以其大规模并行处理、自学习、自组织、非线性、高容错性这些更接近人脑工作方式的计算机制再次引起人们的极大关注。目前，人工神经网络模型还只能模拟人类的一些低层的思维活动，最成功的领域是感知层次上的联想记忆、模式识别等。在众多的神经网络模型中，应用最广泛的是多层感知机神经网络，主要用于分类。分类是一个基本问题，许多问题可归结为分类问题。Hopfield网络被尝试性地应用与优化计算，但缺乏统一的应用方法，如能量函数的设置随问题的不同而不同，另外，也不具备学习算法。而自学习能力是神经网络受到重视的主要原因之一。多层感知机神经网络的研究开始于50年代。Rosenblatt^[21]提出了只有输入、输出层的感知机(Percptron)模型，它只能解决线性可分区域的分类问题。D. E. Rumelhart(1986)^[11]等人提出了误差反向传递学习算法(即BP算法)，解决了引入隐结点后网络的学习问题。Hecht-Nielsen(1989)^[19]对BP网络进行了分析，证明它具有以任意精度逼近任意 L_2 函数的能力。S. I. Gallant(1988)^[15]率先使用多层感知机构成了用于医学诊断的连接机制专家系统。随后出现了许多连接机制专家系统。多层感知机神经网络在应用中存在主要问题是由于隐含结点没有明确的逻辑定义，因而难对计算过程进行解释；多层神经网络对未知情况的推广(Generalization)具有一定的“任意性”，并且缺乏可信度估计；隐含结点如何设置也是一个困难问题。此外，寻找更有效的学习算法则是神经网络当前一个重要的研究方向之一。尽管神经网络本身还存在许多缺点，理论研究还很不成熟，但它已显示出巨大的应用价值和前景。由于神经网络的巨大的优越性，因此把神经网络引入仿真输出分析中是特别引人注目的。

1.3 研究内容概要

本书目的是研究仿真输出数据的分析方法。本书拓宽了传统的仿真输出分析的研究技术途径。从统计理论、人工神经网络、模糊数学等技术途径来分析仿真输出数据，试图把这三种理论结合起来，发挥优点，克服缺点，从而很好地解决仿真输出分析的问题。本书共分四部分，即基于经典统计分析理论的仿真输出分析方法、基于神经网络的仿真输出分析方法、仿真输出数据的分布识别和实例研究。主要研究工作包括：

一、稳态仿真输出分析

稳态仿真输出分析是仿真输出分析领域的一个难点，这是因为稳态仿真输出数据是强自相关的，基于独立假设的经典统计技术不能直接应用。尽管已经提出了多种对稳态仿真输出数据进行分析的方法，但是它们的应用效果不是十分令人满意的，这主要表现在所得的置信区间的置信度往往偏低。本书从分析稳态输出数据的相关性出发，给出了多种计算和检验相关性方法，以此为基础提出了多种新的和改进的输出分析方法。实验检验表明，本书提出的方法大大改进了稳态仿真输出分析方法的性能。

二、小样本或单次仿真输出分析

本书提出用Bayes方法来解决小样本或单次仿真输出分析的问题。给出了多种用于输出分析的Bayes方法。

三、复杂系统仿真中的方差衰减技术

本书对复杂系统仿真中的方差衰减技术，特别是公共随机数和对偶变量法进行了全面的理论分析，给出了各种条件下实现方差衰减的理论和实际结果。

四、神经网络的研究和改进

本书对神经网络，特别是多层感知机进行了全面的分析和评述，并针对多层感知机神经网络存在的问题给出了多种改进学习算法。

五、基于神经网络的仿真输出分析框架

本书提出了基于神经网络的仿真输出分析的概念，并给出了总体分析框架，包括基于神经网络智能仿真输出分析框架、基于神经网络的分布识别框架、基于神经网络的黑箱系统辨识框架。

六、基于神经网络的智能化仿真输出分析

本书提出用神经网络代替人来进行仿真输出分析，以实现智能化的仿真输出分析，给出了基于神经网络的智能仿真输出分析系统的设计与实现，并进行了深入分析和实例研究。

七、基于神经网络的分布识别

分布识别是仿真输出分析中一个长期没有很好解决的问题，本书利用神经网络来解决这一问题。本书对基于神经网络的分布识别系统进行了详细的理论分析、实际设计和实验检验。

八、仿真输出数据分布形式的模糊识别

本书运用模糊识别方法对仿真数据分布形式进行推断和识别，给出了分布识别的模糊识别模型，并开发了通用的概率分布模糊识别软件系统，同时进行实验检验。

九、应用实例研究

本书以一个实际的制造系统为例进行了仿真研究，并运用本书给出的一些方法对仿真输出数据进行了相应的分析，同时对不同的生产作业计划进行了比较分析，从中选出较优的作业方案。因此有效的仿真输出分析将是辅助决策的基础，它在实际应用中具有极其重要的经济效益。

本书各部分的相互关系如图 1-1 所示。

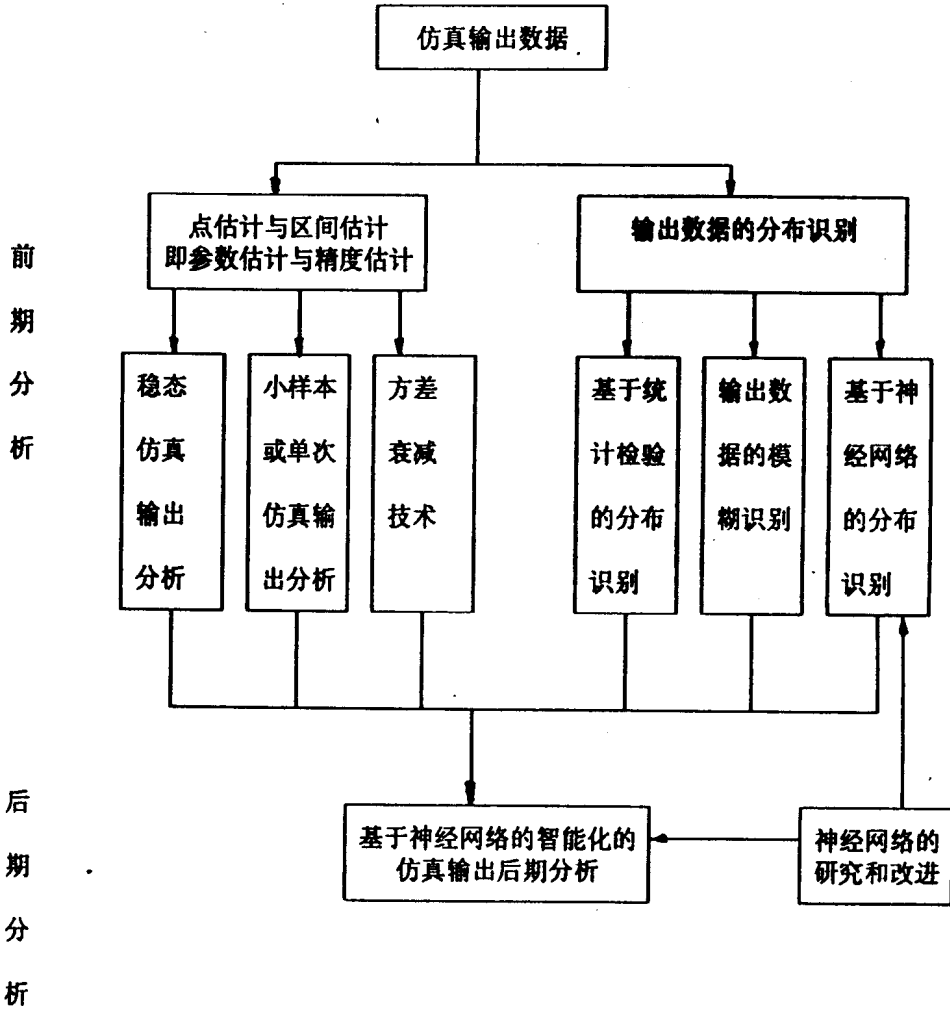


图 1-1 本书各部分的相互关系

第二章 稳态仿真输出分析方法的研究

2.1 引言

仿真建模是一种非常有用的工具，利用它可以给出各种条件下系统的行为。通过均值的辨识可以改进系统性能或更好地了解系统的行为特性，因此在仿真输出数据分析中，一个重要目标是对系统特性进行均值估计。这些均值估计可作为系统特性表征的重要源泉。例如在排队系统中，均值估计可能是实体在系统中或等待线中度过的平均时间，等待线的平均长度，服务设备的平均空闲时间。

仿真输出数据序列通常是由相关随机变量形成的，这就会造成对仿真输出数据使用经典统计推断方法进行参数推断的错误，因为在样本观察值独立性不成立时，这些方法是不可信的。Law(1982)^[149]指出经典估计方法的置信度受到数据相关特性的严重制约。

特别是对于稳态仿真，它的输出数据序列本质上就是一个相关的数据序列，因此如何合理地分析数据序列的相关性，如何从相关的数据序列中得到不相关的数据序列，如何利用统计理论对这些相关的数据序列进行分析，又能克服经典统计理论基于独立观察数据的局限性，则是我们十分关注的。

仿真输出分析是当前一个十分活跃的研究领域。在解决数据相关性问题时，提出了各种不同的方法，这些方法包括重复一删除法（又称独立重复运行法），批平均法，参数建模法（又称自回归分析法），频谱分析法，再生法和标准时间序列法等等。每一种方法都有各自的分析观点，Law(1983)^[147]指出这些方法可以分成四类：

- (1) 用于寻找独立观察数据的方法。
- (2) 用于寻找输出变量相关性估计的方法。
- (3) 用于寻找基本过程的特别随机结构的方法。
- (4) 基于标准化时间序列的方法。

本章我们首先对现有的仿真输出分析方法进行了回顾和比较，在此基础上，针对批平均法存在的问题，给出了一种基于独立性检验的改进批平均法，克服了传统的批平均法缺乏有效的检验方法的缺点，提高了置信区间的置信度。另外，还提出了一种新的基于时间序列分析的分析方法——间隔抽样法。该方法属于第一类方法，它给出一种用于检测数据相关性消失的方法，因而试图从一相关的观察数据序列中收集到本质上不相关或近似不相关的数据，以便在仿真输出数据独立性不成立时尽量减少潜在的错误，从而得到准确的统计推断结果。

2.2 稳态仿真输出分析方法综合评述

2.2.1 终态仿真与稳态仿真

由仿真研究所关心的系统性能测度的不同,人们把仿真问题分为终态和稳态仿真问题。

终态仿真是一种系统性能测度与仿真的时间区间 $[0, T_F]$ 有关的仿真模式。这里 T_F 是指在仿真中特定的仿真终态事件发生的时间。

稳态仿真是在仿真终止时间趋于无限时,定义系统性能测度极限的一种仿真模式。

终态仿真与稳态仿真之间的主要差别在于终态仿真主要研究在规定的的时间区间内的系统行为,仿真终止事件在仿真开始前已经确定,系统的初始状态的选择对系统性能测度有着明显的影响;而稳态仿真主要研究系统的稳态行为,仿真运行的长度要足够长,以便得到好的系统性能测度的估计,系统的初始状态对系统性能测度的影响是无关紧要的,但合适的初始状态的选择可以缩短仿真运行的长度。由于这些差异,导致了仿真输出分析时,终态仿真与稳态仿真在统计方法上的不同。终态仿真与稳态仿真的区分也不是绝对的,对于某些系统究竟采用何种模式进行分析,取决于分析者要达到的目的。

一、终态仿真输出分析

终态仿真的性能测度直接依赖与系统的初始状态。在收集输出数据时,必须使用相同的初始条件进行独立重复仿真运行。如果每次运行产生观察值 Y_i ,那么 n 次重复仿真运行将产生 n 个观察值 Y_1, Y_2, \dots, Y_n 。如果每次观察值将是独立同分布的。这样经典的统计分析技术便可用来建立均值的置信区间。

假设 $\{Y_i\}$ 是独立同分布(IID),且服从正态分布,那么均值

$$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i \quad (2.1)$$

的 $100(1-\alpha)\%$ 置信区间为:

$$\bar{Y} \pm t_{n-1, 1-\alpha/2} \frac{S}{\sqrt{n}} \quad (2.2)$$

这里, S^2 是 Y_1, Y_2, \dots, Y_n 的样本方差,

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (2.3)$$

$t_{n-1, 1-\alpha/2}$ 是自由度为 $n-1$, 置信度为 $100(1-\alpha)\%$ 的 t 分布临界值。

终态仿真中下面三方面问题是关键:

1. 初始条件的选择,必须保证每次仿真运行在相同的初始条件下进行。
2. 随机数流的独立性,以保证各次运行之间相互独立。
3. 假设输出参数服从正态分布,如不服从正态分布,则必须进行大量运行,以遵从大数定理。

由于终态仿真中,人们通过使用独立随机数流很容易保证仿真运行的独立性,从而使