

离散事件系统 建模与仿真

顾启泰 编著

1.9
K



清华大学出版社

<http://www.tup.tsinghua.edu.cn>

离散事件系统建模与仿真

顾启泰 编著

清华大学出版社

(京)新登字 158 号

内 容 简 介

本书系统地介绍了离散事件系统建模与仿真的基本原理和方法及其在管理工程领域中的应用。其内容涉及多种通用和专用仿真语言,还介绍了近年来迅速发展的 Petri 网建模与仿真,重点介绍了排队系统、库存系统、加工系统以及输入、输出数据分析。本书例题丰富,具有较强的实用性。

本书用作工业工程专业研究生教材,也可供机电、控制和管理专业高年级学生和研究生教学参考,对专业技术人员和管理人员更具有参考价值。

版权所有,翻印必究。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

图书在版编目(CIP)数据

离散事件系统建模与仿真/顾启泰编著. —北京:清华大学出版社,1999
ISBN 7-302-03407-9

I. 离… II. 顾… III. ①离散系统(自动化)-建立模型②离散系统(自动化)-仿真
IV. TP271

中国版本图书馆 CIP 数据核字 (1999) 第 07549 号

出版者: 清华大学出版社(北京清华大学校内,邮编 100084)

<http://www.tup.tsinghua.edu.cn>

印刷者: 北京市清华园胶印厂

发行者: 新华书店总店北京发行所

开 本: 787×1092 1/16 印张: 12.5 字数: 292 千字

版 次: 1999 年 8 月第 1 版 1999 年 8 月第 1 次印刷

书 号: ISBN 7-302-03407-9/TP·1847

印 数: 0001~3000

定 价: 13.50 元

前 言

离散事件系统所包含事件的发生在时间和空间上都是离散的,例如交通管理、生产自动线、计算机网络、通信系统和社会经济系统等都是离散事件系统。在这类系统中,各事件以某种顺序或在某种条件下发生,并且大都是随机性的,不能用常规的方法加以研究。特别是如何解决复杂系统性能的管理与控制问题,正面临着越来越大的挑战。

基于模型的数字仿真方法,能在一个系统实现之前预测它的运行性能,比较各种可供选择的方案,可以达到决策优化的目的。

当前,离散事件系统建模与仿真已不再只为专家们所关心,而发展成为广大工程技术人员和管理人员需要学习掌握的一项专门技术,它为各种各样的系统分析、战略研究、运筹规划和预测决策提供了一个强有力的工具。

本书系统地介绍离散事件系统建模与仿真的基本原理和方法及其在管理工程领域中的应用,其特点是:

(1) 全书贯穿了建模与仿真的统一,强调正确进行建模分析、正确设计仿真试验以及正确进行输出分析,从而增强了仿真的实用性。

(2) 内容广泛,涉及多种通用和专用仿真语言:FORTRAN 和 SIMLIB, GASP IV, GPSS, SIMSCRIPT, SIMAN, 并增加了近年来有了迅速发展的 Petri 网建模与仿真的内容。这些方法可用于各种不同的系统,包括排队系统、库存系统、加工系统、运输系统以及计划管理系统等,而不局限于某一类具体系统。

(3) 通过众多的例题,加强对抽象内容的理解,使困难的论题变得直观易懂。具有概率论、基础统计基本知识和计算机程序设计知识的读者都不难掌握本书的内容。

本书共分 10 章:第 1 章绪言,介绍离散事件系统仿真模型的建立,三种程序结构以及仿真研究的步骤。第 2 章排队系统,介绍排队论的基本概念、到达间隔及服务时间的理论分布和经验分布;对单服务台和多服务台排队系统的运行指标进行分析比较;通过一个手工仿真的例子加深对排队系统基本概念和运行指标的理解;阐述如何用 FORTRAN 语言编写简单排队系统的仿真程序,以加深对事件调度算法概念的理解。第 3 章库存系统,介绍单一产品库存系统的基本概念和各种费用的计算;阐述如何建立库存系统的仿真模型和用 FORTRAN 语言进行程序设计;对仿真输出和订货策略进行了讨论。第 4 章加工系统,介绍链式存储分配和 SIMLIB 简单仿真语言;将它用于一个比较复杂的加工系统建模和进行程序设计;最后对仿真输出进行讨论。第 5 章叙述随机数的产生,讨论利用随机数产生各种统计分布随机变量的方法。这是离散事件随机仿真的基础。第 6 章讨论输入数据的收集和分析,包括分布的辨识、参数估计、拟合度检验和相关性分析等内容。因为在实际仿真工作中要花很多精力用于数据的收集和分析,所以必须予以重视。第 7 章仿真的输出分析,介绍了性能测度及其点估计和区间估计,讨论了相关系数对估计的影响,终态和稳态仿真中应用独立重复运行方法的输出分析。仿真的输出分析常常为大多数仿真研究

者所忽视,实际上仿真输出并不一定是可信的。所以,本章内容对于增强仿真的实用性具有指导作用。第8章讨论了根据仿真运行输出数据进行各种可供选择的系统设计方案的比较和评价,还简单介绍了用于估计不同的设计方案效果的统计模型。这是仿真应用的一个很重要方面。第9章简要介绍了GASP IV,GPSS,SIMSCRIPT,SIMAN四种常用的仿真语言及其在排队系统模型仿真中的应用。第10章介绍Petri网的基本概念,条件/事件系统,库所/变迁网,谓词/变迁网,着色网以及带时间的Petri网系统,附有众多的应用实例;简要介绍了其仿真系统的软件实现。它特别适用于复杂的具有并发成分的系统建模与仿真,如FMS(柔性制造系统)。附录A和B为FMS对象库的规格语言语法结构图及其代码。

本书是作者在长期从事教学和科研实践基础上,参阅了众多国内外有关文献写成的。近年来用作工业工程专业研究生教材,取得良好的教学效果。也可供机电、控制和管理专业高年级学生和研究生教学参考,对仿真专业技术人员和管理人员更具有参考价值。

由于作者水平有限,书中难免存在缺点甚至错误之处,恳请读者予以批评指正。

作 者

1997年12月于清华

目 录

第 1 章 绪言	1
1.1 离散事件系统仿真模型的建立	1
1.2 仿真研究的步骤	5
第 2 章 排队系统	7
2.1 排队论的基本概念	7
2.2 到达间隔和服务时间的分布	11
2.3 排队系统的分析	14
2.4 排队系统的仿真	20
2.5 仿真程序设计	23
第 3 章 库存系统	34
3.1 单一产品的库存系统	34
3.2 库存系统模型	35
3.3 仿真程序设计	36
3.4 仿真的输出和讨论	50
第 4 章 加工系统	52
4.1 链式存储分配	52
4.2 SIMLIB 简单仿真语言	55
4.3 加工系统模型	69
4.4 仿真程序设计	70
4.5 仿真的输出和讨论	80
第 5 章 随机变量的产生	82
5.1 随机数的产生	82
5.2 产生随机变量的方法	83
第 6 章 输入数据的分析	90
6.1 数据的收集	90
6.2 分布的辨识	91
6.3 参数估计	92
6.4 拟合度检验	93

6.5 相关性分析.....	93
第7章 仿真的输出分析	98
7.1 性能测度及其估计.....	99
7.2 终态仿真的输出分析	101
7.3 稳态仿真的输出分析	102
第8章 设计方案的比较和评价.....	106
8.1 两个系统设计方案的比较	106
8.2 k 个系统设计方案的比较	109
8.3 用以估计不同的设计方案效果的统计模型	110
第9章 仿真语言.....	115
9.1 GASP IV 仿真语言	116
9.2 GPSS 仿真语言	121
9.3 SIMSCRIPT 仿真语言	125
9.4 SIMAN 仿真语言	130
第10章 Petri 网.....	135
10.1 Petri 网的基本概念	135
10.2 条件/事件系统	138
10.3 库所/变迁网	143
10.4 谓词/变迁网	153
10.5 着色网.....	158
10.6 带时间的 Petri 网系统	162
10.7 仿真系统的软件实现.....	173
附录 A FMS 对象库的规格语言语法结构图	177
附录 B FMS 对象库的代码	180
参考文献.....	189

第1章 绪 言

离散事件系统和连续系统不同,它包含事件的发生在时间和空间上都是离散的,例如交通管理、生产自动线、计算机网络、通信系统和社会经济系统都是离散事件系统。

离散事件系统定义为一组元素(实体)的集合,为了达到某些目的,这些元素以某些规则相互作用、关联而集合在一起。例如制造系统由机器、部件以及操作工人组成,银行系统由出纳员和顾客组成。系统之外称作系统环境,系统经常由于环境发生变化而受到影响,例如上述制造系统中电力供应、原材料供应等属于外部影响,它们直接影响工厂的产量。

我们把在特定时间内和研究目的有关的描述系统所需变量的集合定义为系统的状态。在银行系统中,可能的状态变量是正在工作的出纳员人数,在等待线上排队或正在接受服务的顾客数,以及下一个顾客的到达时间。在制造系统中,机器的忙、闲或故障为可能的状态变量。

离散事件系统状态变量仅仅在可数的一些时间点上才有变化,我们定义可能改变系统状态的瞬间事变为事件。例如在银行系统中,顾客的到达和顾客接受完服务均为事件;在制造系统中故障即为事件。

在离散事件系统中,各事件以某种顺序或在某种条件下发生,并且大都是随机性的,不能用常规的方法加以研究。为了合理地、有效地、经济地生产、管理、销售、服务,对各种各样的离散事件系统进行仿真研究是十分必要的。

在连续系统的数字仿真中,时间通常被分割成均匀的间隔,并以一个基本的时间间隔计时;而离散事件系统的数字仿真则经常是面向事件的,时间并不需要按相同的增量增加。

在连续系统仿真中,系统动力学模型是由表征系统变量之间关系的方程来描述的。仿真的结果为系统变量随时间变化的时间历程;在离散系统仿真中,系统变量是反映系统各部分相互作用的一些事件,系统模型则是反映这些事件状态的数集,仿真结果是产生处理这些事件的时间历程。

1.1 离散事件系统仿真模型的建立

离散事件系统的仿真模型是一个数集,它用以表示系统的状态,现以一个简单的文件处理系统为例来加以说明。

一个工作人员每日的工作是处理文件,他从每天开始工作时即处理文件,直至文件处理完毕或工作结束,每一个文件处理完毕后开始下一个文件的处理,间隔一小时休息一次,但必须是在一个文件处理完毕之后。假设不考虑当天收到的文件,那么文件的数量可以预置,并随着每个作业的完成递减为零。

对这样的一个模型可用表 1.1 来表示。其中: i 行相应于第 i 个文件;第 1 列是文件

号,后 4 列给出了不同的时间值; $t_b(i)$ 为每个文件开始处理的时刻, $t_w(i)$ 为处理文件所需的时间, $t_f(i)$ 为每个文件处理完毕的时刻, $t_c(i)$ 为两次休息之间完成作业的累计时间; F 是一个标记, F 为 1 时表示第 i 个文件处理后休息, F 为 0 表示不休息。表 1.1 的系统模型表示了文件的处理过程。

表 1.1 文件处理系统模型

文件号 i	开始时刻 t_b	工作时间 t_w	完成时刻 t_f	累计时间 t_c	休息标记 F	作业号 N
1	0	45	45	45	0	57
2	45	16	61	61	1	56
3	66	5	71	5	0	55
4	71	29	100	34	0	54
5	100	33	133	67	1	53
6	138	25	163	25	0	52
7	163	21	184	46	0	51

用以描述系统状态的数,称为状态描述字。一些状态描述字具有物理上的意义,如文件数 i 和时间 t ;另一些则只表明条件,如标记 F 。当状态描述字的值发生变化时,则称为产生了一个离散事件。在复杂的系统中,可能同时产生两个不同的事件,只是与该事件相关的那些状态描述字才发生改变。上述模型在计算机上的执行是逐行、从左到右顺序进行的。

更多的更具有实际研究意义的离散事件系统则要复杂得多,不仅系统状态的改变是离散的,而且事件的产生是随机的,常存在排队问题。现以银行系统为例作一说明。

所研究的系统包括出纳员、顾客和对应每个出纳员的排队线。顾客从外部区域(系统环境)进入银行排队或离开银行走入系统环境,两者的边界就是银行的出入口。

所确定的仿真实体及其有关的属性是:

- (1) 顾 客 事务的数目及到达率;
- (2) 出纳员 出纳员的人数,处理事务的原则和速率;
- (3) 排队线 平均长度,每个顾客的平均等待时间。

系统的单个活动是处理顾客提出的一个事务,该活动和顾客及出纳员的属性有关,并影响顾客在排队线上的属性,例如当处理事务的速度慢时,排队线就长。图 1.1 为银行系统的一个子系统的方框图,该银行子系统的仿真源程序可以有面向时间、面向事件和面向进程三种程序结构。图 1.2 和图 1.3 分别为面向时间和面向事件的仿真程序结构图。

比较图 1.2 和图 1.3 可以看出两者的最大差别在于时间 TIME 的处理上。前者程序每做一次循环就增加一个时间单位,不管系统是否有事件发生,每次都要询问一下,计算机做了很多不必要的空操作;后者是根据当前事件的发生时刻进行离散变化的,不存在前者的缺点,但实时性较差。

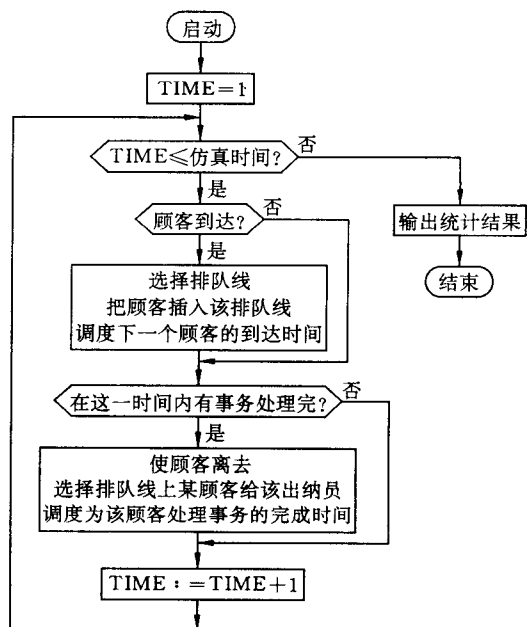
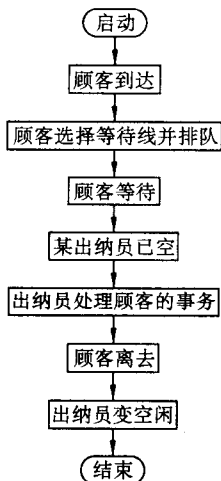


图 1.1 银行子系统

图 1.2 面向时间的仿真程序结构图

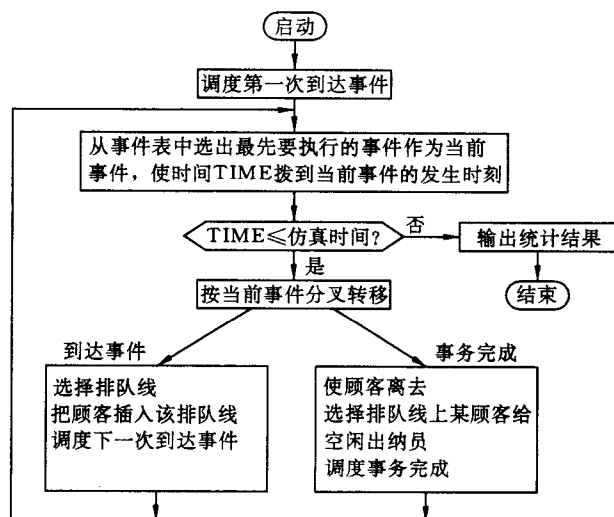


图 1.3 面向事件的仿真程序结构图

另一种仿真程序是面向进程的,它的设计特点是为每一个实体(如银行系统的顾客或出纳员)建立一个进程,该进程反映某一个动态实体从建立开始到结束为止的全部活动。图 1.4 为进程运行的时间图,这里假设出纳员有两个,排队线只有一条的情形。由于顾客的到达时间、出纳员对事务处理的时间均有随机性,在运行中可能出现有多个进程并存

的情形。图中符号△表示一个顾客产生的时刻,也是相应进程*i*开始运行的时刻;符号□表示某顾客离去的时刻,也是相应进程*i*撤销的时刻;符号×表示排队顾客开始接受服务的时刻;虚线表示进程的排队时间;波纹线表示顾客得到服务的时间。图中进程3和4出现虚线,表明第3个和第4个顾客到达银行时,两个出纳员都在服务,故需排队等待的情形。

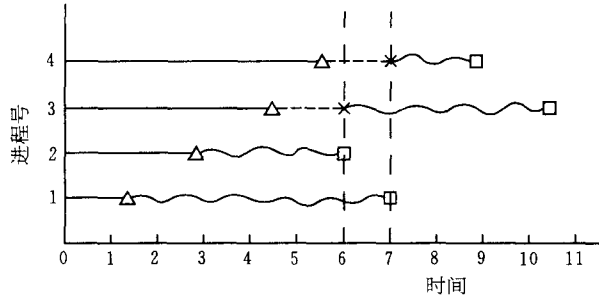


图 1.4 进程运行时间图

在银行系统的例子中,要建立的进程只有两类:一类是动态实体——顾客;另一类是起设备作用的实体——出纳员。图 1.5 和图 1.6 分别为这两类进程运行的程序框图。比较这两个框图,可以看出动态实体的进程较复杂。进程调度的算法和面向事件中事件调度的原理是相似的,通常把满足运行条件和优先级最高的进程投入运行。两者的差别在于事件的调度总是发生在一个事件完成之后,而进程的调度则可以发生在进程运行之中,即在一个进程未完成时调度另一个进程。例如,上述的顾客进程将会由于要排队等待接受服务而挂起,即暂停该进程的运行。

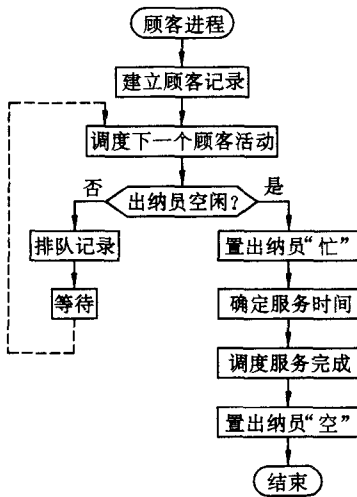


图 1.5 顾客进程

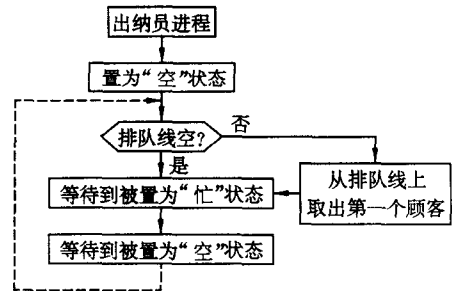


图 1.6 出纳员进程

1.2 仿真研究的步骤

1. 问题的阐述

每一项研究都应从说明问题开始,问题由决策者提供,或由熟悉问题的分析者提供。

2. 设置目标及完整的项目计划

目标表示仿真要回答的问题、系统方案的说明。项目计划包括人数、研究费用以及每一阶段工作所需时间。

3. 建立模型和收集数据

模型和实际系统没有必要一一对应,只需描述实际系统的本质。因此,最好从简单的模型开始,然后建立更复杂的模型。但是,模型的复杂程度要和模型想要达到的研究目标相适应。违背这个原则,只能增加建模和计算机的费用。在很大程度上,研究的目标决定了收集数据的类型。例如在银行系统中,如果要求知道出纳员人数改变时对等待线长度的影响,那么需要的数据类型将是顾客到达间隔时间分布、出纳员服务时间分布以及在不同条件下等待线长度的历史分布。收集的数据将被用于确认仿真模型。

4. 编制程序和验证

建模者必须决定是采用通用语言如FORTRAN还是专用仿真语言编制程序。通用语言需要很长的开发时间,但通常在计算机上执行时比专用语言更快。用专用语言编制程序容易并且很快,所以越来越多的建模者使用专用语言。为判断执行的计算机程序是否正常,需要验证。验证比较简单,通常用常识可完成这一步。如果输入程序的参数以及模型的逻辑结构表达是正确的,则验证就完成了。

5. 确认

确认指确定模型是否精确地代表实际系统。它不是一次完成的,而是比较模型和实际系统特性的差异,不断对模型进行校正的迭代过程。在上述银行的例子中,在当前的条件下收集与等待线长度有关的数据,模型仿真的结果是否重现了该系统的测度,这就是确认模型的标志。

6. 试验设计

确定仿真的方案,涉及初始化周期的长度,仿真运行的长度以及每次运行的重复次数。

7. 生产性运行和分析

通常用于估计被仿真系统设计的性能测度。对仿真中所产生的数据进行分析,其目的

在于预计一个系统的性能,或比较两个或多个不同系统设计的性能,根据分析,确定是否需要进一步运行。

8. 文件清单和报表结果

文件清单中应有必要的程序文件说明,使模型的用户了解程序是如何运行的。所有分析结果应有清晰简要的输出报表,为决策者评价所选方案的最后模式提供某种证实的工具,以增加模型建立的可信度。

9. 实现

成功的实现取决于以上每一步都取得成功,但在整个处理中最关键的一步是步骤 5 (确认),因为一个无效的模型将导致错误的结果或很高的代价。

图 1.7 所示为上述仿真的步骤,以指导建模者进行细致而完整的仿真研究。

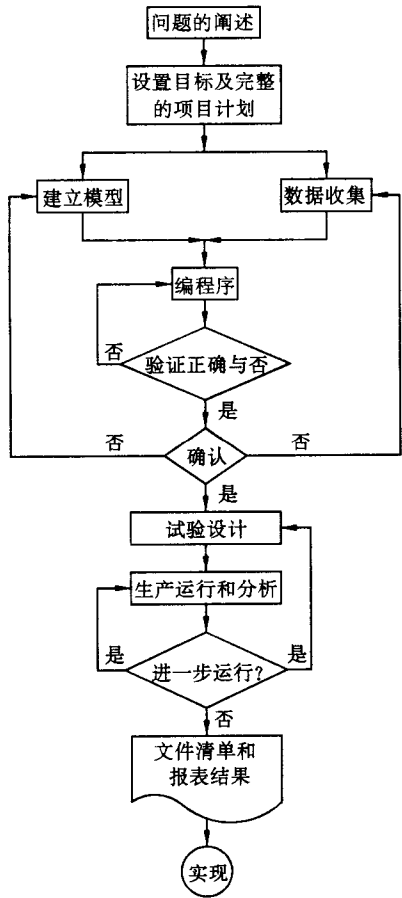


图 1.7 仿真研究的步骤

第2章 排队系统

排队是我们在日常生活中经常遇到的现象,如顾客到商店购买物品,病人到医院看病就常常要排队。一般说来,当某个时刻要求服务的数量超过服务机构的容量时,就会出现排队现象。这种现象远不只在个人日常生活中出现,要求服务的可以是人,也可以是物。例如在计算机网络系统中,要求传输数据的是各个网络结点,这里的服务机构是网络传输机构,而要求服务的就是等待传输数据的网络结点。此外,电话局的占线问题,车站、码头等交通枢纽的车船堵塞和疏导,故障机器的停机待修,水库的存贮调节等都属于排队现象。在各种排队系统中,顾客到达的时刻与接受服务的时间都是不确定的,随着不同时机与条件而变化,因此排队系统在某一时刻的状态也是随机的,排队现象几乎是不可避免的。

如果增添服务设备,就要增加投资或产生空闲浪费;如果服务设备太少,排队现象就会严重,对顾客个人和对社会都会带来不利影响。所以,管理人员必须考虑如何在这两者之间取得平衡,以期提高服务质量,降低成本。

2.1 排队论的基本概念

1. 排队系统的组成

一般的排队系统都有三个基本组成部分:

(1) 到达模式 指动态实体(顾客)按怎样的规律到达,描写实体到达的统计特性。通常假定顾客总体是无限的。

(2) 服务机构 指同一时刻有多少服务设备可以接纳动态实体,它们的服务需要多少时间。它也具有一定的分布特性。通常,假定系统的容量(包括正在服务的人数加上在等待线等待的人数)是无限的。

(3) 排队规则 指对下一个实体服务的选择原则。通用的排队规则包括先进先出(FIFO),后进先出(LIFO),随机服务(SIRO)等。

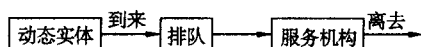


图 2.1 排队系统的基本结构

图 2.1 是排队系统的基本结构。图中的动态实体以一定的到达模式来到服务机构请求服务,由于服务机构的服务员或设备有限,服务要有服务时间,因此某些动态实体在来到服务机构之前不能马上得到服务,需要排队等待,当服务机构完成对某个动态实体的服务之后,就可接纳新的动态实体。排队规则将确定在队列中哪些动态实体可以最先得到服务。在很多实际问题中,动态实体的到达时间是随机的,服务机构的服务时间也是随机的,

这样动态实体排队的长度也会是随机的,最后反映在服务机构处于“忙”或“闲”的时间也是随机的。如何通过已知的到达模式和服务时间的概率分布,来研究排队系统的队列长度和服务机构“忙”或“闲”的程度即服务效率,这就是离散事件仿真所需解决的问题。

2. 到达模式

先介绍将要用到的一些基本概念。

(1) 平均到达间隔时间 T_a 指在考虑模型的总时间 T 中,共到达了 n 个顾客的情况下的比值 T/n 。

(2) 平均到达速率 λ 指单位时间内到达的顾客数

$$\lambda = \frac{1}{T_a} \quad (2.1)$$

(3) 到达间隔分布函数 $A_0(t)$ 指到达间隔时间大于 t 的概率。因为累积分布函数 $F(t)$ 是到达间隔时间小于 t 的概率,所以

$$A_0(t) = 1 - F(t) \quad (2.2)$$

根据定义,函数 $A_0(t)$ 在 $t=0$ 时取最大值为 1。当 t 增加时, $A_0(t)$ 逐渐减小。

(4) 到达时间变化系数 指到达间隔时间的标准差 S_a 与平均到达间隔时间 T_a 的比值 S_a/T_a 。变化系数是个无量纲的值,它描述了数据围绕平均值的分散程度。

以后我们将介绍,指数分布的平均值与标准差相同,其变化系数为 1。所以,如果观测数据的变化系数接近于 1,则假定用指数分布去拟合这些数据是合理的。当变化系数比 1 小很多时,经常应用爱尔朗分布。

到达模式按顾客到来的方式可能是一个一个的,也可能是成批的;按相继到达的间隔时间可以是确定型的,也可以是随机型的;按到达过程可以是平稳的,指描述相继到达的间隔时间分布和所含参数(如期望值、方差等)都是与时间原点无关的,也可以是非平稳的。

3. 服务机构

同到达间隔时间一样,首先定义 T_s 为平均服务时间, μ 为平均服务速率, $S_0(t)$ 为服务时间分布函数,即服务时间大于 t 的概率。

服务机构按机构形式可以没有服务员,也可以有一个或多个服务员。在有多个服务台情形中,它们可以是平行排列(并列)的,可以是前后排列(串列)的,也可以是混合的,如图 2.2 所示。按服务方式可以对单独顾客进行,也可以对成批顾客进行;按服务时间可以是确定型的,也可以是随机型的;按服务过程可以是平稳的,也可以是非平稳的。非平稳情形的数学处理是很困难的,所以同到达过程一样,服务时间的分布也假定是平稳的。

4. 排队规则

顾客依一定的次序和规则接受服务。

(1) 损失制 指顾客到达时,如所有服务台都正被占用,随即离去。

(2) 等待制 指顾客到达时,如所有服务台都正被占用,就排成队伍,等待服务。服务次序可以采用下列各种规则:

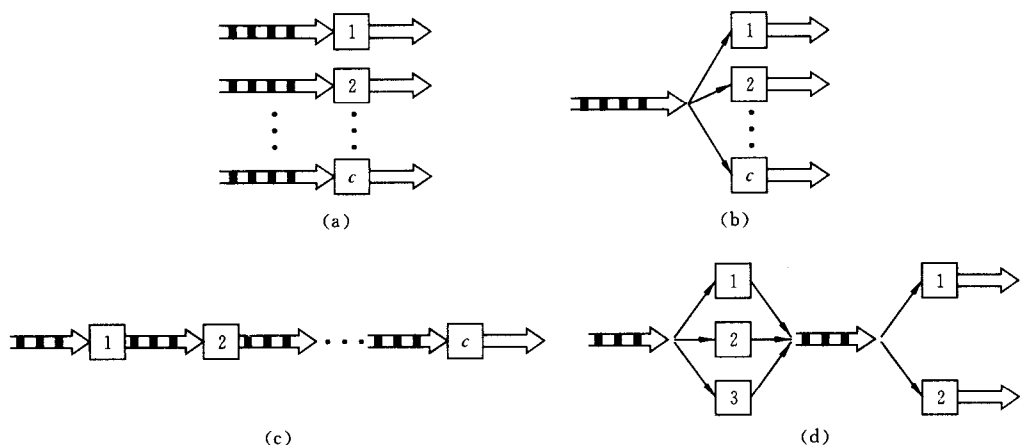


图 2.2 多服务台排队方式

- (a) 多队-多服务台(并列)情形 (b) 单队-多服务台(并列)情形
(c) 多服务台(串列)情形 (d) 多服务台(混合)情形

先到先服务(FIFO) 即按到达次序接受服务,这是最通常的情形。

后到先服务(LIFO) 如乘用电梯的顾客常是后入先出的,仓库中存放的钢板也是如此。在情报系统中,最后到达的信息往往是最有价值的,因而常采用后到先服务的规则。

随机服务(SIRO) 当服务台空时,从等待的顾客中随机地选取一名进行服务,而不管到达的先后,如电话交换台接通呼唤的电话便是如此。

优先权服务(PR) 如医院中急诊病人优先得到治疗。

最短处理时间先服务(SPT) 例如设备选择工件时,首先选择所需加工时间最少的工件进行加工。

对于 n 个服务台的情形,当顾客到达时,可按如下规则在每个服务台前排成一个队,第 $1, n+1, 2n+1, \dots$ 个顾客排入第一个队,第 $2, n+2, 2n+2, \dots$ 个顾客排入第二个队等等。或者所有顾客排成一个公共的队,每当有一个服务台得空闲时,队首的顾客接受服务。也可以这样来排成几个队,当某个顾客到达时,以概率 P_i 排入第 i 个队,且满足

$$\sum_{i=1}^n P_i = 1.$$

(3) 混合制 例如,当排队过长时,后到的顾客会自动离去,此时可定义队长 $q < N$ 时就排入队列;若 $q = N$,则到达的顾客将自动离去。另一种是当等待时间或逗留时间(等待时间与服务时间之和)小于某一时间 T 时,顾客将等待;大于 T 时,顾客将自动离去。

在使用优先权时,必须考虑当一个比现在正在接受服务的实体具有更高优先权级别的实体到达后,系统将作何处理。通常可有两种选择:其一,优先权仅仅决定一个动态实体排队的先后,优先权高的排在队列的前面,而不影响正在接受服务的实体。其二,立即停止当前的服务,为新到的具有更高优先权的实体服务,这种情形称为抢占服务,这时被抢占的实体等待新实体离开后再重新接受服务。

5. 队列的度量

已知平均到达速率 λ 和平均服务速率 μ , 定义业务量强度 u 为

$$u = \frac{\lambda}{\mu} \quad (2.3)$$

在某些场合下, 到达的动态实体并不全都能够得到服务。因此有必要区分实际到达速率以及得到服务的到达速率, 分别用 λ' 和 λ 来表示。此时的业务量强度为

$$u = \frac{\lambda'}{\mu} \quad (2.4)$$

定义服务设备利用率 ρ 为得到服务的动态实体的到达速率与服务速率之比

$$\rho = \frac{\lambda}{\mu} \quad (2.5)$$

在多服务设备系统中

$$\rho = \frac{\lambda}{n\mu} \quad (2.6)$$

式中, n 为服务设备数目, μ 为每个服务设备的平均服务速率, 这里假设每个服务设备的服务速率相同。显然在多服务设备系统中, 服务员人数越少, 服务设备利用率就越高。正常情况是服务设备利用率小于 1, 这样每个动态实体才有希望得到服务。利用率越高, 则动态实体排队等待的时间越长。因此设计系统的设备利用率是一个权衡过程, 可以通过多次的仿真实验加以合理解决。

对于队列的度量, 通常考察两个量: 队列的长度和排队的时间。这两个量都是变量, 不同时间的队列长度是不同的, 不同动态实体的排队时间也是不同的。在仿真实验中, 对这两个量的变化进行统计, 计算出其均值、方差、最大值、最小值等。这些值反映了一个服务系统的重要特征。

6. 排队模型的分类

根据上述三个组成部分中最主要的特征, D. G. Kendall 提出一个分类方法, 现在已被广泛采用了。它只针对并列的服务设备的情形, 用的符号形式是

$$X/Y/Z$$

其中, X 处填写表示相继到达间隔时间的分布; Y 处填写表示服务时间的分布; Z 处填写并列的服务设备的数目。

表示相继到达间隔时间和服务时间的各种分布的符号是:

M——负指数分布(M 是 Markov 的字头, 因为负指数分布具有无记忆性, 即 Markov 性)。

D——确定性(deterministic)。

E_k —— k 阶爱尔朗(Erlang)分布。

GI——一般相互独立(general independent)的随机分布。

G——一般(general)随机分布。

例如, M/M/1 表示相继到达间隔时间为负指数分布, 服务时间为负指数分布, 单服