

离散事件随机系统的分析和优化 以及对生产自动化的应用*

何毓琦

(美国坎布里奇 哈佛大学 应用科学部)

〔摘要〕 本文扼要阐述二种非概率方法(时域法),它用于研究以随机服务网及加工制造业中的生产系统为典型的离散事件动态系统。本文目的在于把随机服务理论中“运行分析”的新发展同基于控制理论发展起来的生产系统中的某些新技术统一起来。

引言

很多动力学现象可以看作离散事件动态系统(DEDS),象公路网交通流,生产系统材料(另件、夹具)流,通讯网中的讯息以及一般的非标准随机服务系统等与那些我们较熟悉的由常微分方程或差分方程支配的动态系统不同,离散事件系统随多种事件的发生而演变,象一个讯息或顾客的到达,一部机器意外或例行的停机,以及一个任务的完成等。就这些事件的性质而言,可以是随机的,也可以是确定性的。它们通过启动或中止某些动作使系统从一个“状态”运动到另一个“状态”。这类系统分析中那最令人感兴趣但又困难的地方是各事件在整个过程中错综复杂的相互作用。

传统看法认为随机服务理论或马尔可夫链理论是DEDS解析分析方法的典型代表。诚然,在这方面确已有了大量文献,并取得了相当的成就。但是同所有理论一样,其主要缺点是:要理论成立,必须满足一系列的限制性的假设条件。例如,几乎所有随机服务理论都不准有“阻塞”现象,或更一般地说都要求一个服务台的服务速率独立于其他各台及系统的其他部分的状态。在某些DEDS里,离散事件也可能引起含连续变量的动态过程(如指挥控制通讯系统的情况),从而成为一个混合系统,对此,随机服务理论无法单独处理。此外,

DEDS的状态空间也可能会大得无法有效地求解。

仿真是另一种处理CEDS问题的通用手段。若不考虑成本费用,仿真模型可以尽可能需要地做得逼近真实。从现有的GPSS,GA-SP, SIMSCRIPT, SIMULA等离散事件仿真语言看,应公正地说,仿真是蛮力分析离散事件动态系统的主要工具。而该法的主要费用在计算机机时费,特别在涉及描述问题或设计问题时更为突出。因经常含有随机现象,仿真或实验一次便要在计算机上做规模很大的蒙特卡罗运行,所以一个复杂系统在其各参数范围内反复仿真的代价可能很大,甚至是不可行的。

本文提出一种“暴露法”,它可望成为研究DEDS的第三种方法。该方法的出发点是实验。实验(蒙特卡罗或其他)在实际的或仿真的DEDS上进行,该动态系统的某样本路径可观测到,然后进行分析,导出所有被观测量如机器忙时、流量、平均停机时间等必须满足的关系式。这些关系式连同系统基本参数的有关信息一起,用来预报该DEDS沿其他样本路径的性状,而不必再做仿真或实验。这正是观测(实验)——演绎——预报,一种科学上惯用的方法。实验或仿真是对实际系统或近乎实际系统进行的,它避开了随机服务理论中许多假

*译自第八届IFAC世界大会论文集。

设的限制；而通过分析和演绎，又绕过了仿真计算工作量太重的问题。令人惊讶的是这种处理方式事实上在许多场合都可行。而且，在某种意义上讲它比经典随机服务理论简单直观得多，原因是它无概率考虑。在计算机系统性能评价研究中，称这种方法为“运行分析”，并在前些年里取得了相当的成就 (Denning and Buzen, 1977, 1978, 1980)。近来，又把这种方法推广到了自动化系统 (Suri, 1980a)。何及其合作者采用上述分析方法独立解决了串行生产线中积之已久的问题 (Ho, Eyler and Chien, 1979)。现在看来这种解法正是运行分析思想对阻塞为主要关心问题的那种情况的推广。它可应用于非经典随机服务问题。一些初步成果可查文献 (Ho, Eyler, 1980)。最后还有一点，因仿真及运行分析本质是时域法（相对于概率法而言），所以在研究常微分方程描述的动态系统中所建立起来的大量概念都能引用。

本文的目的力图完成仿真与分析的统一与合成，并给正待研究解决的 **DEDS** 问题以启迪。

2. DEDS随机服务系统的运行分析

2.1 单队—台系统

先从最简单的队—台组合作为离散时间动态系统来研究。我们把时间单位选得很小，使每单位时间内至多有一次到达或离开。激励系统的输入过程有两个：到达过程 $a(t)$ ，它的样本实现由用户到达时间集 $\{t_1, t_2, t_3, \dots\}$ 给出；服务过程 $s(t)$ ，假若用户源源不断且服务台永不中断服务，则 $S(\cdot)$ 的样本实现由服务完成时间集 $\{\tau_1, \tau_2, \tau_3, \dots\}$ 给出。若队空，则服务过程被迫暂停。研究这种系统，先引入服务时间变量 $V(t)$ ，使

$$V(t+1) = \begin{cases} V(t) & n(t) = 0 \text{ 且 } t \neq t_i, \\ V(t) + 1 & \text{其余} \end{cases} \quad i = 1, 2, 3, \dots \quad (1)$$

式中 $n(t) \triangleq$ 排在队中的顾客数（包括正被服务的顾客）。

再引入指标函数定义

$$I(V(t) \triangleq K \quad \text{使 } \tau_i \leq V(t) < \tau_{i+1} \\ \triangleq t \text{ 刻以前被服务完的顾客数。}$$

则由简单的流量守恒得

$$n(t+1) = \begin{cases} n(t) + 1 & \text{若 } V(t) > \tau_{IV}, \\ & \text{且 } t = t_i, i = 1, 2, \dots \\ n(t) - 1 & \text{若 } V(t) = \tau_{IV(t)}, \\ & \text{且 } t \neq t_i, \text{ 又 } n(t) > 0 \\ n(t) & \text{其余} \end{cases} \quad (2)$$

方程 (1)、(2) 给出了该单队—台组合的状态空间表示。一旦输入样本 $a(\cdot)$ 、 $s(\cdot)$ 给定，便能据这些方程进行系统仿真。值得注意，该服务台输出仅是 $n(t)$ 和 $s(\cdot)$ 的函数（参看式 (2)），而与其他无关。下面即将阐述的分析方法在向随机服务网推广时，这个事实极为重要。

在区间 $[0, T]$ 上观测此单队—台系统。定义“运行量”或“可观测量” (Denning and Buzen, 1978)，它们是状态变量 $n(t)$ “历史”值的函数。

$A(n)$ —— $n(t) = n$ 时，到达的顾客数，

$C(n)$ —— $n(t) = n$ 时，服务后离台的顾客数，

$T(n)$ —— $n(t) = n$ 时的累计时间，

$$A = \sum_{n=0}^{N-1} A(n) \quad ,$$

$$C = \sum_{n=1}^N C(n) = IV(T),$$

$$T = \sum_{n=0}^N T(n) \quad ,$$

N —— 在 $[0, T]$ 内，队中最大顾客数，

$$N = M_{i, a_x}(n(t)).$$

由这些公式可导出辅助变量：

$$S(n) \triangleq \frac{T(n)}{C(n)} \quad \text{—— } n(t) = n \text{ 时，}$$

平均服务时间，

$$Y(n) \triangleq \frac{A(n)}{T(n)} \quad \text{—— } n(t) = n \text{ 时，顾客到}$$

达率，

$$Y_0 \triangleq \frac{A}{T} \quad \text{—— 总平均到达率，}$$

$B \triangleq \sum_{n=1}^N T(n) \triangleq V(T)$ ——忙时总时间

$S \triangleq \frac{B}{C}$ ——整个过程中, 每个顾客服务时间的总平均,

$$S = \frac{V(T)}{I(V(T))},$$

$X \triangleq \frac{C}{T}$ ——总平均离开率,

$$X = \frac{I(V(T))}{T},$$

$U \triangleq \frac{B}{T}$ ——服务台的利用率,

$W \triangleq \sum_{n=1}^N nT(n)$ = 累计的顾客等待时间 = $\sum_{n=1}^T n(t)$,

$Q \triangleq \frac{W}{T}$ ——平均队长,

$R \triangleq \frac{W}{C}$ ——服务的平均响应 (等待和服务) 时间。

由上述定义直接可得:

$$U = SX \left(\frac{B}{T} = \frac{B}{C} \cdot \frac{C}{T} \right), \quad (3)$$

$$Q = RX \left(\frac{W}{T} = \frac{W}{C} \cdot \frac{C}{T} \right). \quad (4)$$

通常, 分别把它们称为利用定理和 Little 定理*。为进一步推导起见, 此时习惯上引入如下假设:

(A₁) 流量平衡, 即假定 $x = y_0$ 或 等效地假定 $A = C$ 或 $n(0) = n(T)$,

(A₂) 一步性, 即 $n(t)$ 在每时间段内仅能有 ± 1 的变化。

由于时间无限可分, 故总能把时间间隔选得足够小, 使假设 (A₂) 满足。(A₂) 意味着 $a(t)$, $s(t)$ 只能取 1 或 0, 且不能两者同时取 1。(A₁) 实际上并不是什么限制, 它在实验或仿真中经常可以看到。这两条假设直接意味

着

$$A(n) = C(n+1),$$

此称为“状态转移平衡”。原因是, 式 (5) 表明从 $n(t) = n$ 到 $n(t) = n+1$ 的转移次数等于从 $n(t) = n+1$ 到 $n(t) = n$ 的转移次数。在随机服务分析中, 感兴趣的往往是队长 $n(t) = n$ 所占的时间比例, 即

$$P(n) \triangleq \frac{T(n)}{T} = \frac{C(n)}{T(n-1)} \cdot \frac{T(n)}{C(n)} \cdot \frac{T(n-1)}{T},$$

据式 (5), 又可得

$$P(n) = \frac{A(n-1)}{T(n-1)} \cdot \frac{T(n)}{C(n)} \cdot \frac{T(n-1)}{T} = Y(n-1)S(n)P(n-1). \quad (6)$$

式 (6) 给出一个从 $P(0)$ 来确定运行队长分布的递推公式。若假定 $Y(n)$, $S(n)$ 不随观测实验而变, 则式 (6) 及 $\sum_{n=0}^N P(n) = 1$ 能用来预报和计算系统的响应, 如 $Q = \sum_{n=0}^N np(n)$, R 等。事实上, 由于 $Y(n)$, $S(n)$ 很难可靠估计, 再引入如下附加假设**:

(A3a) 均等到达率, 即 $y(0) = y(1) = \dots = y(n-1) = y$ 。

(A3b) 均等服务时间, 即 $s(1) = s(2) = \dots = s(n) = s$ 。

这样, 式 (6) 可写为

$$P(n) = (ys)^n P_0, \quad (7)$$

由式 (7) 和 $\sum_{n=0}^N P(n) = 1$, 可得***

$$P(0) = 1 / \left(1 + \sum_{n=1}^N (ys)^n \right) = \frac{1}{1 + \frac{ys}{1-ys}} = 1 - ys, \quad (8)$$

* 原注: 从运行分析角度看, Little 定理只不过是观察量与导出量的就定义出发而得的一个简单结论。但若从随机角度看 (4) 式, 那就需要相当复杂的推导和假设。

** 原注: 这些假设的合理性在第 2·3 节说明。

*** 译注: 据上文直推, 应有 $P(0) = 1 / (1 + \sum_{n=1}^N (ys)^n)$, 当 $N \rightarrow \infty$ 时, 有式 (8)。

$$Q = \sum_{n=0}^{\infty} np(n) = \frac{ys}{1-ys} \quad (9)$$

倘若从队长分布, 期望队长, 到达时间及平均服务时间的随机含义去解释 $P(n)$ 、 Q 、 y 、 s , 那末式 (8), (9) 正是众所周知的 $M/M/1$ 随机服务系统的结论 (Kleinrock, 1976)。

必须强调指出, 式 (3) — (9) 的推导中没作任何概率假设, 故结论适用于 DEEDS 的输入 $a(\cdot)$, $s(\cdot)$ 的任何一种样本实现。因此, 尽管在平衡——普阿松到达——指数分布服务时间的假设下也可推与式 (9) 一样的结论, 但它们的成立条件、应用范围及解释是不同的。

2.2 简单随机服务网

我们可以研究的最简单的随机服务网是由两个在第 2.1 节中讨论过的队—台组合串联而成, 见如图 1 所示:

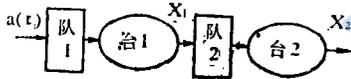


图 1

这称为开随机服务网, “开” 表示该网由已知输入 $a(t)$ 驱动。其状态方程是*

$$\begin{aligned}
 V_1(t+1) &= \begin{cases} V_1(t) & \text{若 } n_1(t) = 0 \text{ 且 } t \neq t_i, i=1,2,\dots \\ V_1(t)+1 & \text{其余} \end{cases} \\
 n_1(t+1) &= \begin{cases} n_1(t)+1 & \text{若 } V_1(t) > \tau_1^i V(t) \text{ 且 } t = t_i, i=1,2,\dots \\ n_1(t)-1 & \text{若 } V_1(t) = \tau_1^i(v_1(t)) \text{ 且 } n_1(t) > 0 \\ n_1(t) & \text{其余} \end{cases} \\
 V_2(t+1) &= \begin{cases} V_2(t) & \text{若 } n_2(t) = 0 \text{ 且 } V_1(t) = \tau_1^i, i=1,2,\dots \\ V_2(t)+1 & \text{其余} \end{cases} \\
 n_2(t+1) &= \begin{cases} n_2(t)+1 & \text{若 } V_2(t) > \tau_2^i(v_2(t)) \text{ 且 } V_1(t) = \tau_1^i, i=1,2,\dots \\ n_2(t)-1 & \text{若 } V_2(t) = \tau_2^i(v_2(t)) \text{ 且 } n_2(t) > 0 \\ n_2(t) & \text{其余} \end{cases}
 \end{aligned} \quad (10)$$

如前所述, 若给定服务台服务过程及队储量, 则该台的输出与其他因素无关。状态变量 $n_1(t), V_1(t)$ 与 $n_2(t), V_2(t)$ 解耦。因此, 可以完全采用前一节的方法分析第一个队—台组合, 确定出 $P(n_1)$ 和 x_1 。现在又是第二个队—台组合的输入。知道了 x_1 , 我们便可进而分析并确定 $P(n_2)$ 及 x_2 。特别在 (A1), (A2), (A3a), (A3b) 成立时**, 有

$$y = x_1 = x_2, \quad (11)$$

$$P(n_1, n_2) = P(n_1)P(n_2) \quad (12)$$

$$P(n_1) = ys_1 P(n_1 - 1) \quad (13)$$

$$P(n_2) = x_1 s_2 P(n_2 - 1) = ys_2 P(n_2 - 1) \quad (14)$$

再由这些量, 可相应算出其他性能指标。

图 1 改画成图 2, 情况就稍复杂些。

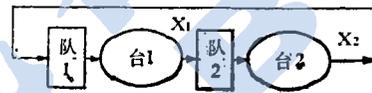


图 2

图 2 称为闭随机服务网。网中流通的顾客数固定为 N 。顾客一旦离开 2 号台, 便排在 1 号队中等待服务。问题是这里不另外给定到达过程。对方程 (10) 仅作如下修改***

$$\begin{aligned}
 n_1(t+1) &= \begin{cases} n_1(t)+1 & \text{若 } V_1(t) > \tau_1^i(v_1(t)) \text{ 且 } V_2(t) = \tau_2^i(v_2(t)) \\ n_1(t)-1 & \text{若 } V_1(t) = \tau_1^i(v_1(t)) \text{ 且 } V_2(t) = \tau_2^i(v_2(t)) \\ n_1(t) & \text{其余, } n_1(t) > 0 \end{cases}
 \end{aligned}$$

其余各式都不变。系统由服务过程 $s_1(\cdot), s_2(\cdot)$, 满足 $n_1(0) + n_2(0) = N$ 的初始条件 $n_1(0), n_2(0)$, 所驱动。

至此, 遇到了两个麻烦: 一个是网的流量 $x = x_1 = x_2$ 未知, 有待确定; 其次是 1 号队、

* 译注: 原文中式 (10) 的 $n_1(t+1)$ 误印为

$$n_1(t+1) = \begin{cases} n_1(t)+1 & \dots \\ n_1(t)-1 & \dots \end{cases}$$

** 译注: 原文为“特别在 (A1), (A2), (A3a), (A4) 成立时”。

*** 译注: 原文为“对方程 (1) 仅作如下修改”

2 号队的长度不再独立, 因为 $n_1 + n_2 = N$ 。而式(6), $p(n) = y(n-1)s(n)p(n-1)$ 对随便哪个队一台组合都成立。在这种网中, 没有理由认为, 在队储量空时, 向队 Q_1 的输入速率与在队储量满时一样。恐怕还值得提一下的是, 方程(12)的结果仍然正确。于是有 (Denning and Buzen, 1977)

$$p(n_1, n_2) = \frac{1}{G} F_1(n_1) F_2(n_2) \quad (15)$$

式中 $F_i(n_i) = s_i F_i(n_i - 1), F_i(0) = 1,$
 $i = 1, 2 \quad (16)$

而归化常数 G 可据

$$1 = \sum_{n_1, n_2 \in S(N, 2)} p(n_1, n_2) \quad (17)$$

或

$$G = \sum_{n_1, n_2 \in S(N, 2)} F_1(n_1) F_2(n_2)$$

求得。其中 $S(N, 2)$ 是该网满足 $n_1 + n_2 = N$ 的所有状态 n_1, n_2 的集合。以后, G 并非通过直接求和获得, 而采用递推算法, 这个方法的一般形式首先由 Buzen 导出(1973)。定义

$$\begin{aligned} G \triangleq g(N, 2) &= \sum_{n_1, n_2 \in S(N, 2)} F_1(n_1) F_2(n_2) \\ &= \sum S_1^{n_1} S_2^{n_2} \\ &= \sum_{\substack{n_1, n_2 \in S(N, 2) \\ n_2 = 0}} S_1^{n_1} S_2^{n_2} + \sum_{\substack{n_1, n_2 \in S(N, 2) \\ n_2 > 0}} S_1^{n_1} S_2^{n_2} \\ &\triangleq g(N, 1) + S_2 g(N-1, 2), \end{aligned} \quad (18)$$

式中第二项基于对求和过程中队 2 始终至少包含一个顾客的认识而来。式(18)的边界条件是:

$$\begin{aligned} g(n, 1) &= S_1^n \quad n = 0, 1, \dots, N \\ g(0, 1) &= 1 \\ g(0, 2) &= 1 \end{aligned} \quad (19)$$

在推广到更一般的网时, 式(18)比式(17)容易计算得多。因为

$$\begin{aligned} x = \text{网流量} = x_1 &= \frac{U_1}{S_1} = \\ &= \frac{1 - p(n_1 = 0, n_2 = N)}{S_1} \end{aligned}$$

稍变化一下, 得

$$x = \frac{g(N-1, 2)}{g(N, 2)}, \quad (20)$$

由它, 其他量都很容易算得。

图 2 有两个极端情况值得一提: 一是, 若 $S_1 > S_2$, 即 2 号台服务较快, 且 $N \rightarrow \infty$, 则 S_1 是“瓶颈”, 队 1 决不会是空, 故 $x = 1/s_1 = \text{MiN}(1/s_1, 1/s_2)$; 另一种情况是, 若 $N = 1 < 2$, 则不存在等待, 网的响应很简单, $R = S_1 + S_2$ 。在更一般的网中也有类似情况。

2.3 一般随机服务网

为把图 1, 图 2 的结果推广到较一般网, 引入路径比。令网有 k 个队一台组合, i 号台服务完的以 r_{ij} % 的比例去 j 号台。假设:

(A4) 发送均等性, 即 r_{ij} 与网络状态无关*, 又引入定义

$$C_{ij} = C_i r_{ij} = i \text{ 号台服务完后去 } j \text{ 号台的顾客数,}$$

其中 $C_i = [0, T]$ 时间内 i 号台服务完的顾客数。

令 A_0 (对应 C_0) 代表某开网的外部到达 (离开外去)。在闭网中, $A_0 = C_0$ 。据流量守恒得

$$A_j = \sum_{i=0}^k C_i r_{ij}。$$

据假设 (A1), (A2) 得

$$A_j = C_j,$$

$$C_j = \sum_{i=0}^k C_i r_{ij}$$

$$\text{或 } X_j = \frac{C_j}{T} = \sum_{i=0}^k x_i r_{ij}, \quad (21)$$

这称为“流量平衡方程”。在开网中, 给定 $X_0 (= A_0)$, 又若 $[r_{ij}]$ 是不可约矩阵**, 则式(21)有唯一解 $x_i, i = 1, 2, \dots, k$ 。一旦知道了各 x_i , 就可以分别象图 1 那样的情况加以分析。在闭网中, x_0 (系统流量) 是未知的。若定义

* 译注: 原文为“(A3)发送均等性”

** 原注: 按定义 $[r_{ij}]$, 是随机矩阵。不可约意味着可以通过某路径从任何一个节点转移到另一个任意节点。

$$\frac{x_i}{x_0} = V_i = \text{访问比率}, \quad (22)$$

则由 $V_0 = 1$, 可唯一解出各 V_i 。事实上, 依照图 2 的方法, 可以记

$$p(n_1, n_2, \dots, n_k) = \left[\prod_{i=1}^k F_i(n_i) \right] / G \quad (23)$$

$$F_i(n_i = n) = (V_i S_i)^n F_i(n_i = n - 1) \quad (24)$$

$$\sum_{n_1, \dots, n_k \in S(N, K)} p(n_1, n_2, \dots, n_k) = 1 \Rightarrow$$

$$G = \sum_{n_1, \dots, n_k \in S(N, K)} \prod_{i=1}^k F_i(n_i), \quad (25)$$

式中 G 可以据归化条件式 (25) 求得。虽然 (23)~(25) 式概念很简单*, 但是式 (3)~(5) 的计算量大得无法求解, 这个问题直到 Buzen 算法发明才解决。这个算法可以式 (16)~(19) 的一般形式叙述为

$$g(n, k) = \sum_{n_1, \dots, n_k \in S(N, k)} \prod_{i=1}^k (V_i S_i)^{n_i}, \quad (26)$$

$$= g(n, k - 1) + (V_k S_k) g(n - 1, k), \quad (27)$$

边界条件

$$\begin{cases} g(n, 1) = (V_1 S_1)^n & n = 0, 1, \dots, N \\ g(0, k) = 1 & k = 1, 2, \dots, K \end{cases} \quad (28)$$

及

$$X_0 = \frac{g(N - 1, K)}{g(N, K)}. \quad (29)$$

在一般网里也有类似的两种极端情况, $N \gg K$ 和 $N \ll K$ 。这时有

$$X_0 = \frac{1}{V_i S_i} \triangleq \text{Min}_i \left\{ \frac{1}{V_i S_i} \right\}$$

和 $R = V_1 S_1 + V_2 S_2 + \dots + V_k S_k$ (Denning and Buzen, 1978)。

以上随机服务网的分析结果可以推广到多类用户, 多服务台, 不同排队规则及随负荷而变的服务时间等更一般的情况 (Baskett and Co-workers, 1975)。在此不再继续讨论。

方程 (23)~(29) 的结果的重要性在于: 只

需知道 $V_i S_i$, $i = 1, 2, \dots, N$, 就可以计算随机服务网的性能指标, 而 $V_i S_i$ 在一次实验中便可观察到**。当然这个结论是否成立还取决于 (A1), (A2), (A3b) 和 (A4) 是否满足。(A3b) 是其中限制最强的一条。它在服务时间按指数分布的服务台稳态时满足*** 可是, 对于在有限时间段的不同 n 、不同实验, $S_i(n)$ 值可能改变。由此出现了结果敏感度问题, 即式 (29) 对 S_i 变化的敏感度。过去十年中积累起来的大量实例表明, 式 (29) 对于假设 (A3b) 的是否满足特别强壮。最近, Suri (1980a) 又论证了这种强壮性的解析基础。他的结果主要有:

(i) 在闭网中, 若除 i 号台以外都满足 (A3b) 且设

$$\left| \frac{\Delta S_i(n)}{S_i} \right| < \delta\%,$$

$$\text{则 } \left| \frac{\Delta X_0}{X_0} \right| < \frac{U_i}{2} \delta\%,$$

式中 U_i 是 i 号台的利用率 (Suri, Thm. 3.1, 1980a)。

(ii) 又若网平衡, 即 $V_i S_i = V_j S_j$, $V_{j,i}$, (实际常常如此), 则对所有可能的

$$\left| \frac{\Delta S_i(n)}{S_i} \right| < \delta\%,$$

$$\text{总有 } \left| \frac{\Delta X_0}{X_0} \right| < \frac{U_i}{K} \delta\%,$$

这里 K 是服务台总数。例如 $N = 7$, $K = 4$, $V_i = 0.7$, 则

$$\left| \frac{\Delta X_0}{X_0} \right| < 0.175\delta\%$$

(Suri, Lemma 4.5, 1980)。

(iii) 在任何情况下, 对所有可行的变化 $\Delta S_i / S_i$, 在平衡网络中有

* 译注: 原文为“虽然 (3)~(5) 式...”。

** 原注: 若路径比 r_i 已知, 则 V_i 可由 (22) 式及 (23) 算出。

*** 原注: 在闭网中, 不需要 (A3) 假设。

$$-\varepsilon_{min} < \frac{\Delta X_0}{X_0} < \varepsilon_{max}$$

在此 $\varepsilon_{min} = \left(2 + \frac{K}{N-1}\right)^{-1}$,

$$\varepsilon_{max} = \left((N+K-1)\left(1 + \frac{K-2}{N-1}\right) - 1\right)^{-1}$$

(Suri, Thm.5.1, 1980a)

2.4 应用于“柔性”机器系统FMS

FMS是传统加工车间的现代形式，它能加工多种零件且批量又可变。FMS由若干通过物料传送装置 MHS 连接的数控机器组成，整个 FMS受中央计算机控制，如图 3 所示。

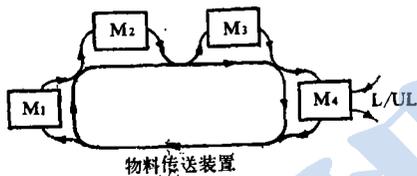


图 3

零件经L/UL收发站M₄进入FMS系统，由传送装置 MHS 送它去“访问”各机器，依次进行各种复杂的机械加工。一般说来，一个零件总要经过若干道工序，而每道工序又有若干台机器供选择。最终，该零件离开FMS，送进零一个新零件。Solberg (1977) 首先证明了，这种系统可以用闭随机服务网有效地模型化。MHS 的行为可看作由多个服务台组成的站，这些服务台的服务时间表现为传送滞后。因为任何时刻系统里的总另件数一般不太多，所以 MHS 的置放空间可用无阻塞的队表示。尽管大多数的服务时间分布是确定性的而非实验性的，但这个随机服务网模型已被许多人广泛验证。Suri 提供了这种模型强壮性的解析说明 (1980a)。详见(Solberg, 1977; Suri, 1980a)。

眼下，FMS 随机服务理论成果的应用 仅限于设计和规划问题。至于确定路线的实时控制和其他运行还没有什么大的研究成果，也许正等待着新一代技术的出现，例如本文下节要阐述的新技术。

非标准随机服务系统和一般离散事件动态系统

3.1 随机服务理论的局限

上节的结果表明，随机服务系统的性能可以用一组 $V_i S_i$ 参数表达，导出了性能指标与参数 $V_i S_i$ 的关系的闭型表达式。这些表达式成功的关键是：服务台的输出仅仅取决于该台的服务过程及台前的队长。从而确认该服务台与网的其余部份解耦程度，使分析简化。然而，这种性质对许多DEDS或随机服务系统来说并不总能满足。例如，两条单行道的交叉口可以模型化为一对队一台组合，如图 4 所示。

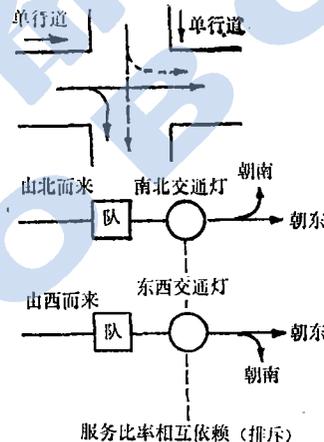


图 4 交通流随机服务系统

在此，交通灯看作为服务比率相互依赖的服务台，两台的输出相互依赖。

另一例子是串行生产线，如图 5 所示：机器 M_i 模型化为服务台，中间贮存库 B_i 模型化为队。因为 $B_i, i = 1, 2, \dots, N$ 的容量有限，所以生产流受“阻塞”制约，即一部机器的故障会影响生产，导致它前面工序的机器停工。在此，输出率取决于系统其他部分的状态。

可以举出许多非标准、非平凡的随机服务系统。对于它们，没有现成的闭型解存在。采

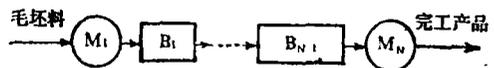


图 5 串行生产线

用粗略的类比，经典随机服务网的闭型解好比线性常微分方程的闭型解。而越出经典范围，就好比跨入了非线性动力学范畴一样。然而，并非前功尽弃。我们必须降低对闭形解的追求，作进一步的类比象处理非线性动力系统那样，先把标称路径附近的解线性化，然后围绕标称路径求摄动解。控制理论中熟知的协态变量及乘子方程都可以用来分析及优化非线性系统的性能。利用这种类比，我们将在下节中对一般离散事件动态系统导出类似的处理方法。

3.2 作为DESE的串行生产线

用图5所画的串行生产线作为模型，着手进行推导。这个问题不仅是第2.2节意义上最简单的非标准随机服务网（有阻塞），而且也是生产研究中久待解决的问题。请参看Buzacott和Hanifin为此课题所列的参考文献(1978)。若机器正常运行周期 CT_i ，在每个周期中失败概率 P_i ，此时于下一周期修复的概率 z_i 。因此可把机器模型化为服务台，它的服务时间分布如图6所示。

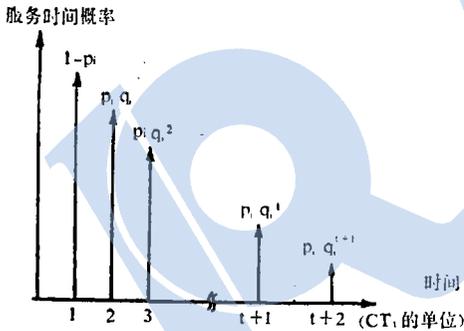


图 6

换言之，我们不去区分机器是运行或停工。而把 M_i 看成这样一种服务台：以 $(1 - P_i)$ 的概率服务时间为 CT_i ；以 P_i 的概率服务时间为 $CT_i + RT_i$ ，其中 RT_i 具有参数 q_i 的几何分布。如果需要服务的顾客（另件）源源不断，那末 M_i 台服务过程的样本路径用其完成服务的时间集 $\{\tau_i^1, \tau_i^2, \dots\}$ 表示。据第2.2节的推导，也把 V_i 定义为 M_i 的服务时间变量， $C_i(t)$ 定义为 M_i 后面的中间贮存库 B_i 的储量*。同时假

定， M_i 前的顾客源和 M_N 后的贮存库都无限大。生产线的效率用流量（即在给定时间内服务完的顾客数）来度量。显而易见，若能使各部机器的阻塞和待料机会最少，则通过量可望增加。

为描述这种DEDS的动态过程，采用与第2节相同的符号。先作如下定义：

- (i) 令 $V_i(t)$ 为 M_i 的服务时间变量， $C_i(t)$ 为 M_i 后随的贮存库 B_i 的储量， $b_i(t)$ 为 B_i 的容量，

又令 $I(V_i(t)) \triangleq K$ 使 $\tau_i^k \leq V_i(t) < \tau_{i+1}^k$ \triangleq 在 t 时 M_i 所服务过的顾客数。

由以上定义和流量守恒，很容易得

$$1 + C_i(t) + I(V_{i+1}(t)) \geq I(V_i(t)) \geq C_i(t) + I(V_{i+1}(t)). \quad (30)$$

- (ii) M_i 的准备输出函数

$$h_i(t) = \begin{cases} 0 & \text{若 } V_i(t) > \tau_i^k(v_i(t)) \\ & \text{或 } I_{v_i(t)} = C_i(t) + \\ & I_{(v_{i+1}(t))} \text{ 且 } V_i(t) = \\ & \tau_{i+1}^k(v_i(t)) \\ 1 & \text{其余} \end{cases} \quad (31)$$

式(31)清楚地表明若 M_i 正在服务或空闲，则它不准备输出。

- (iii) M_i 的准备接受函数

$$g_i(t) = \begin{cases} 0 & \text{若 } V_i(t) > \tau_i^k(v_i(t)) \\ & \text{或 } V_i(t) = \tau_i^k(v_i(t)) \\ & \text{且 } C_i(t) = b_i, g_{i+1}(t) = 0 \\ 1 & \text{其余} \end{cases} \quad (32)$$

对它的解释与式(31)相仿。边界条件 $g_N(t) = 1$ 和 $h_{-1}(t) = 1$ 表示生产线两端有无限源和无限漏。

根据 h_i, g_i ，我们定义

- (iv) 满输出函数(FO)或阻塞函数

* 原注：在生产线中，习惯于把服务台 M_i 与其后的队（中间贮存库 B_i ）当作一个台一队组合。这种记述方法与通常的随机服务理论文献不同，它显得更自然些。

$$f_i(t) = \begin{cases} 1 & \text{若 } C_i(t) = b_i \text{ 且 } h_i(t) = 1, \\ & g_{i+1}(t) = 0 \\ 0 & \text{其余} \end{cases} \quad (33)$$

(V) 零输入函数(NI)或开工函数

$$n_i(t) = \begin{cases} 1 & \text{若 } C_{i-1}(t) = 0 \text{ 且 } g_i(t) = 1 \\ & h_{i-1}(t) = 0 \\ 0 & \text{其余} \end{cases} \quad (34)$$

最后, 状态空间的动态方程是

$$\begin{cases} V_i(t+1) = \begin{cases} V_i(t) & \text{若 } f_i(t) = 1 \\ & \text{或 } n_i(t) = 1 \\ V_i(t) + 1 & \text{其余} \end{cases} \\ C_i(t+1) = \begin{cases} C_i(t) + 1 & \text{若 } 0 \leq C_i(t) < b_i \\ & \text{且 } h_i(t) = 1, \\ & g_{i+1}(t) = 0 \\ C_i(t) - 1 & \text{若 } 0 < C_i(t) \leq b_i \\ & \text{且 } h_i(t) = 0, \\ & g_{i+1}(t) = 1 \\ C_i(t) & \text{其余} \end{cases} \end{cases} \quad (35)$$

可将式(31)~(35)归并在一起给出系统的状态方程,其形如 $[V_i(t+1), C_i(t+1)] = F(V_i(t), C_i(t), V_j)$, 在此 F 是所有状态变量的很复杂的函数, 详情从略(Ho and Cassandras, 1980)。

一个公认的生产线性能度量是

$$J = \frac{V_N(t)}{t}, \quad (36)$$

它用状态变量表示, 很简单。若 M_N 始终满荷, 则 $V_N(t) = t$, M_N 以 100% 的效率工作。因为 M_N 反映了全线的生产, 所以很明显, J 就是生产线的效率或通过量。

主观上总希望通过解动态方程(36)式来算 J , 可是, 正如前面所说, 这种情况下, 一个服务台的输出与系统的所有状态都有关。对这类 DEDES 不再存在。象第 2 节那样的表达式鉴于此, 我们只得试着把雄心勃勃的目标降低些。让我们来观察这种 DEDES 的运行, 即式(35)在 $[0, T]$ 上的解, 总能通过仿真或实际实验做到这一点。假定除了某状态变量 V_i 在 t 刻有小扰动外, 实验能精确重复, 情况又如何呢? 我们尤其希望确定 $\Delta V_N(t)/\Delta V_i(t)$, 当然通过蛮力手段, 也就是通过有特殊扰动的重复仿真或实验, 这个希望总可以实现。但是若要对所有 i 和

许许多多时刻 t 确定 $\Delta V_N^T / \Delta V_i(t)$, 那就难以计算了。在常微分方程中, 我们是通过计算所谓的“乘子方程”或协态变量或转移矩阵来解决这个问题的。因为, 它们把状态变量在 t 时刻的扰动与 T 时刻的状态变量联系起来。在下一节, 我们要导出一个算法, 类似地处理这种 DEDES。

3.3 扰动分析

分析(35)式可揭示出一个显而易见的事实: 状态变量 $V_i(t)$ 随实际时间单调增加, 只有在满输出(FO)或零输入(NI)时才停止增长。换句话说, $V_i(t)$ 上扰动只能影响到 $V_{i+1}(t)$ 或 $V_{i-1}(t)$, 即由于 M_i 造成 M_{i+1} 处的零输入或 M_{i-1} 处的满输出。只要 B_{i-1} 始终不空, B_i 始终不满, M_i 就与系统其他部份解耦。

研究沿标称路径的某时段 $[t_1, t_2, t_3]$ 。

$t_1 \triangleq M_i$ 服务完成时间, $0 < C_i(t) < b_i$, $0 < C_{i-1}(t_1) < b_{i-1}$; $t_2 \triangleq M_{i+1}$ 的零输入的开始时间, 即 $n_{i+1}(t) = 1, t > t_2$ 和 $n_{i+1}(t) = 0, t_1 < t < t_2$; $t_3 \triangleq M_{i+1}$ 的零输入的结束时间, 即 $n_{i+1}(t) = 1, t_3 < t < t_2$ 和 $n_{i+1}(t) = 0, t > t_3$ 。现在假定使 $V_i(t_1)$ 有扰动 ΔV_i 或等价地说有 Δt_i 用“”表示受扰动量, 那末有:

命题 1 $t'_2 = t_2$

证明 设 M_i 在 $[t_1, t_2]$ 内服务过的顾客有 X 个, 且 $C_i(t_1) = y$ 。那末在零输入出现之前, M_{i+1} 服务过的顾客总数为 $x + y$ 。在扰动作用下, $C_i(t'_1) = y$ 且 M_i 在 $[t'_1, t'_2]$ 内仍服务 X 个顾客*。因为 M_{i+1} 不受扰动, 故 $t'_2 = t_2$ 。

命题 2 $t'_3 = t_3 + \Delta t_1$

证明 因为 t_3 完全受 M_i 控制, 所以只要零输入持续时间超过 Δt_1 , 那末 M_{i+1} 将在 t'_3 继续服务, 并因此承受到 M_i 的扰动, 即 $t > t'_3$ 时, $\Delta V_{i+1}(t) = \Delta V_i(t)$

把 M_{i+1} 处的零输入改为 M_{i-1} 处的满输出, 对 M_{i-1} 就有完全类似的一组命题。于是可归纳为: 当 M_i 造成 M_{i+1} 处零输入时, (V_i, t)

* 译注: 原文为“..., $C_i(t'_i) = y$ 且...”。

上的扰动(或等价说为离散事件时间上的扰动)向 M_{i+1} 传播;当 M_i 造成 M_{i-1} 处是满输出时, $V_i(t)$ 上的扰动向 M_{i-1} 传播。

同样,如果 $M_{i+1}(M_{i-1})$ 没有 V 上的扰动,那末倘若 $M_{i+1}(M_{i-1})$ 引起 M_i 处的满输出(零输入)则将使 M_i 上的扰动抵消。更一般情况,表成如下传播律:

若 V_i, V_{i+1} 分别受扰动 $\Delta V_i, \Delta V_{i+1}$ 则只要 M_i 和 M_{i+1} 之间跟着出现的满输出 FO (零输入 NI)持续时间大于 $\Delta V_{i+1} - \Delta V_i$ ($\Delta V_i - \Delta V_{i+1}$),便会在 $FO(NI)$ 消失时引起 $\Delta V_i \leftarrow \Delta V_{i+1}$ ($\Delta V_i \rightarrow \Delta V_i(t_{i+1})$)。

这个规律的简明性不必多说。在标称路径的任意时刻 t ,可在任一 $V_i(t)$ 上引入“虚扰动”。只要对系统演变过程中满输出和零输入的出现(假若有的话)进行简单的观察,便可确定 $V_N(T)$ 上的最终虚扰动, $V_N(T)$ 是该系统整个生产量的重要度量。必做的监视各扰动及其传播的额外辅助工作是最少的。这样,在任何情况下,只要在对标称路径仿真或实验所必须的工作程序上稍加几步便可。

灵敏度或协态变量一经算出,设计一个优化标称路径性能的算法就比较简单了,只要把随便哪一个系统参数扰动与 V_i 联系起来便可。例如,若平均服务时间增加 δ 这就意味着 $V_i(t)$ 在每次服务完成时间上有扰动 $\Delta V_i(t)$,使完成时间滞后了 δ 。令服务完成时间 $t_1^i, t_2^i, \dots, t_k^i$ 上存在扰动,令 $\Delta V_N(T, t_k^i)$ 是在 M_N 上引起的相应的扰动,那末通过量的扰动可算得(用服务完成时间的扰动表示):

$$\Delta TP = \sum_{k=1}^K \Delta V_N(T, t_k^i) \delta \quad (37)$$

其他系统参数的扰动可类似研究。参看 Ho, Eyer 和 Chien(1979); Suri(1980b)。

3.4 向更一般情况推广

第 3.3 节的分析为向非串行生产线的其他情况推广作了准备。

(i) 装配和拆卸机器

装配线与串行生产线不同,它的某些机器

(服务台)需要一个以上的队提供另件(顾客)才能开始服务。如图 7 所示:

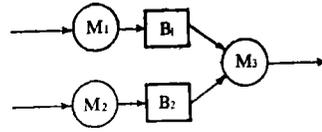


图 7

图中,只要 B_1 或 B_2 有一个是空, M_3 上就出现零输入。在 M_1, M_2 都结束对第 n 个另件加工前, M_3 不能开始这个另件的加工。至于 M_3 的不能开工是由哪部机器(M_1 或 M_2)造成(即零输入 NI 由谁引起)是至关重要的,因为只有这关键机器上的扰动才会传到 M_3 上。当然,这里假定扰动很小,不至于引起关键机器的地位改变。同样, B_1 上的满输出只会把扰动从 M_3 传向 M_1 。关于拆卸机器,可以作对偶的叙述,而图 4 要以相反的零件流向来看。

(ii) 随机服务网

从第 3.3 节分析角度回顾第 2 节的结论是有启发的。在随机服务课题中没有满输出阻塞。扰动向下游服务台传播仅当出现零输入情况(即空队时)。考察第 2.2 节所讨论的那种两服务台的简单闭随机服务网,令 1 号台比 2 号台慢(即 $S_1 > S_2$)。显然对大的顾客数 N ,1 号台前的队绝不会空,因此 1 号台不会出现零输入的情况。这意味着无论什么扰动都传不出去,而事实上将被 2 号台前面队伍中经常出现的零输入所抵消。另一方面,扰动 ΔV_1 几乎肯定地要传到 1 号台。这种分析完全与直观一致,要想使流量增加只能通过修善瓶颈,另一方面,若 $N=1$,则可以肯定两个队都会出现零输入,从而使任何一个服务台上的扰动向另一台传播。这也与 $R = S_1 + S_2$ 相一致(请看第 2.2 节)。

更一般地说,对于任意 N 及任意闭网, $V_i(t)$ 上的扰动依赖于不同队里各“零输入”的相互作用及时间长短。重要的是:随着系统沿标称路径的演变,不难记录虚扰动的传播,从而回答了“若改变…”的问题。当然,对于

严格服从(A1)–(A4)假设的系统来说*,各种性能指标有显式表达,不需要这种扰动分析。但是,若问题涉及阻塞和/或具有其他非标准特征,则可采用本文方法。关于一个应用实例的细节,请参看Ho和Eyer(1980)。

3.5 线性性及统计显著性

扰动分析从根本上讲是线性化思想,也就是参数上的小扰动引起输出的小变化。在比较熟悉的连续变量动态系统里(控制理论中常见的),线性化的正确性很容易理解。从概念上看,这如图8所示。

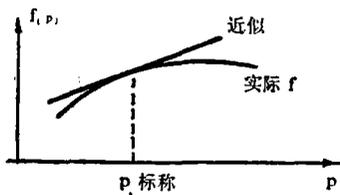


图8

扰动分析的结果只是当标称点处扰动无限小时精确。对有限大小的扰动,这种线性分析只是一种近似,它的精度取决于系统在标称点附近的“曲率”。这种曲率有时容易算,有时却很难。

在生产线或随机服务网这类离散事件动态系统中,情况就非常不同。R.Suri和F.Furtek指出,若采用图8那样示意的话,此时的情况应如图9(Furtek,1980;Buzacott和Hanifin,1978)。

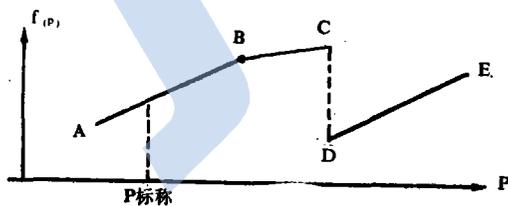


图9

换句话说,对DEDS而言,灵敏度分析结果对标称点处的有限扰动是精确的。对于参数扰动的性能曲线由分段直线段组成,中间可能有间断。对此解释如下:对于某给定的标称路径,全部离散事件诸如服务的完成、顾客的到

达等都可按出现的先后编次序。只要离散事件出现的扰动大小不改变这个次序或事件序列,那必定在图9的严格线性范围AB上。这正是由第3.3节模型的构造所决定。随着扰动的加大,能激发某服务作用的关键事件有可能因此而改变。例如,不是机器等待顾客,现在是顾客等待服务所替代(零输入区间消失了)。此时,工作点从一条线段移到另一条上,即移到图9的BC上(只有比零输入区间短的扰动才能传播)。更一般地说,在任意DEDS中,扰动或许会使零件在网中走完全不同的路径,或许会使事件出现的次序强烈地变化。这形成图9上CD那种间断。弄清这种性态的本质及有关条件正是研究DEDS的最重要部分。

最后,提一下统计显著性问题及这些公式的适用性。要记住:无论是第2节的运行分析还是本节的扰动分析都基本上说的是样本路径。换句话说,我们所建立的是变量或变量扰动间确实存在的关系,这些量与所观察的具体样本路径相联系。当利用这些分析技术或关系式去预报沿新样本路径的性能或性能变动时,显然都作了这样的假定:所讨论的样本路径是“典范的”。在运行分析中,我们利用实验确定了系统参数诸如平均服务时间 $S_i(n)$,访问比率 V_i 。然后假定这些参数值是“典范的”不随样本改变的。在扰动分析中,我们又假定所观察的样本路径(特别是满输出、零输入的次序)是典范的。因此,由某特定样本路径导出的梯度系数 $\Delta V_N(T)/\Delta V_i(t)$ 可以有效地用于其他样本路径以实现最优化。为了回答这类问题,要使用标准的统计技术(参看Ho, Eyer and Chien, 1979; Suri, 1980a)。但是,对用于这类动态系统的通用方法的专门研究还没系统地地进行。有许多课题正待研究。

结 论

离散事件动态系统的理论尚处在萌芽期。

(下转28页)

* 译注:原文为“…服从(A1)–(A5)假设”。

请求或修改参数请求为止。

控制过程首先包括设置速度设定值和采集电动机速度及电流反馈信号,然后执行调节算法,最后修改PWM选通信号。

根据结构程序原理用 PLM/80 语言设计较高级的任务的软件。相反,为了限制处理时间用汇编语言编制了 I/O 驱动程序。这样,总的控制过程的最坏情况执行时间为 4 毫秒。存贮要求包括大约 6K 字节的程序存贮器和 400 字节的读/写存贮器。

3. 实验结果

使用一台工业用功率放大器 and 一台低惯性永磁直流电动机对研制的原型数字控制器性能进行评定。功率放大器特性:

电桥供电电压	±100V
最大峰值输出电流	45A
最大平均输出电流	30A
典型PWM选通信号重复频率	1KHz

电动机(CEMMC17)特性:

T_m	17.2msec	T_e	9msec*
K_1	$0.01\Omega^{-1}$	K_2	$3.5 \times 10^3 \text{rpm/A}$

第一节中概述的调节器综合方法为控制算法参数提供下列值:

K_{pc}	20 Ω	T_{ic}	9msec
K_{po}			0.03rpm/A
T_{io}	100msec	T	4msec

(上接39页)

它不象连续变量动态系统那样,有许多常微分方程方面的知识可供借用。本文回顾了两种很有用的方法。一方面,“运行分析”是分析动态系统集结性能的非概率工具。它处理在某时间段内的平均响应时间、平均队长等整体量,而不强调系统的演变情况。另一方面,我们又用于 DEDS 的“扰动分析”法,它强调的是扰动的传播,这是明确的动态概念。它与“运行分析”一样是时域的,基于样本路径的。实际上,这种扰动分析也许称为“动态运行分析”

不同设定值和负载条件下的电动机速度和电流曲线如图 6、7、8 所示。特别地,图 6、7 的曲线依据的是相同的速度设定值(幅度 $\pm 2000\text{rpm}$,频率 1.5Hz 的方波),但电流限制值不同。显然,正如可在图 7 中注意到的,电动机速度响应由于限流电路的干预变得缓慢。另一方面,图 8 给出了电动机对负载转矩阶跃变化为 $8.5\text{N}\cdot\text{m}$ 时的响应特性。

4. 结 论

实验性试验突出了调节器综合方法的精确性和设计数字控制器所用方法的有效性。尤其值得注意的是由于控制器是以微处理机为基础的,它提供了很大的灵活性,特别就动态选择调节器类型及相应配置参数的可能性而论。

当前工作是朝着全自动方向迈进,特别是为了评价基于微处理机的自适应控制系统的功能,进行并行研究¹⁸。

* 电动机的电气时间常数实际上远小得多。为了减少电流波纹,考虑在电枢电路中有意义串联了引用值为 16mH 的电感。

参 考 文 献 (略)

易镇 译自 Control and Computers,
Vol.10 No.1, 1982,
朱复明 校

更能说明其特征。

上述思想的进一步发展尚需在模型化方面做大量的工作。用常微分方程作为连续变量动力学系统的数学模型,是经过几十年(甚或上百年)多学科的巨大努力之后才凝聚出的结论。为提炼出 DEDS 的数学模型,我们还需要去深入研究更多的具体问题,以便从中获得进一步发展的启示。

(参考文献略)

张志湧 译 薛劲松 校