

Introduction to Systems Engineering

Introduction to Systems Engineering

WUTP

普通高等学校自动化类专业新编系列教材

Introduction to Systems Engineering

Introduction to Systems Engineering

Introduction to Systems Engineering

Introduction to Systems Engineering

Introduction to Systems Engineering

Introduction to Systems Engineering

Introduction to Systems Engineering

Introduction to Systems Engineering

Introduction to Systems Engineering

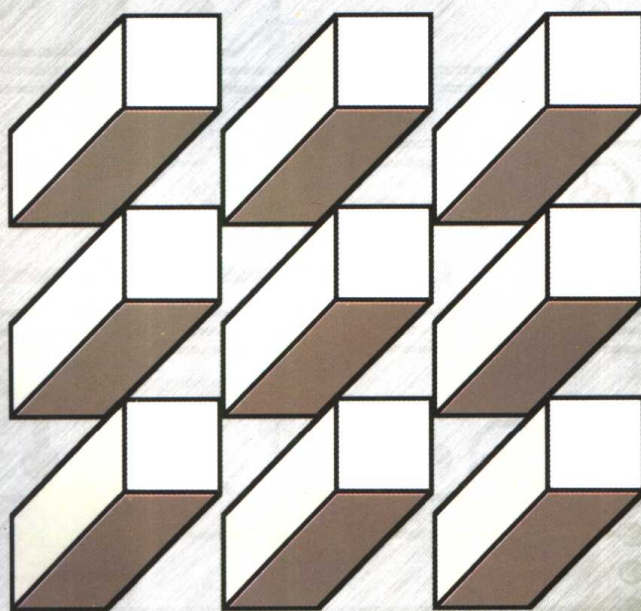


系统工程概论

(第二版)

主 编 杨家本、

副主编 林锦国 黎明森 宋靖雁



武汉理工大学出版社



责任编辑/白立华

封面设计/许伶俐

普通高等学校自动化类专业新编系列教材目录

- | | |
|---------------|--------------------|
| ● 自动化(专业)概论 | ● 电气工程基础 |
| ● 电路分析 | ● 电力电子技术 |
| ● 电路实验指导书 | ● 运动控制系统 |
| ● 模拟电子技术基础 | ● 过程控制系统 |
| ● 数字电子技术基础 | ● 计算机控制系统 |
| ● 电子技术实验 | ● 检测技术与仪表 |
| ● 电磁场与电磁波理论基础 | ● 软件技术基础 |
| ● 信号处理技术 | ● 网络技术基础 |
| ● 自动化专业英语 | ● 系统工程概论 |
| ● 自动控制原理 | ● 智能运输系统——原理、方法及应用 |

项目负责: 武汉理工大学出版社第三工作室

电 话: (027) 87384716

E-mail: huangchun@whut.edu.cn

wutpbook@sohu.com

wutpbailh@163.com

<http://www.techbook.com.cn>

武汉理工大学出版社发行部

地 址: 武汉市洪山区珞狮路122号(430070)

电 话: (027) 87394412

传 真: (027) 87397097

ISBN 978-7-5629-2566-8



9 787562 925668 >

定价: 26.00元

WUJIA

中国出版集团 中国对外翻译出版公司 中国出版集团 中国对外翻译出版公司

中国出版集团 中国对外翻译出版公司

系统工程概论

（第二版）

主编：王仲良
副主编：王仲良 王仲良 王仲良



中国出版集团 中国对外翻译出版公司

普通高等学校自动化类专业新编系列教材

Introduction to Systems Engineering

系统工程概论

(第2版)

主 编 杨家本

副主编 林锦国 黎明森 宋靖雁

武汉理工大学出版社

· 武汉 ·

内 容 提 要

系统工程是运用系统理论和系统方法,借助运筹学、控制论、信息论和计算机等现代科学技术手段,处理各类系统并使其达到总体最优的技术。系统工程对各类工程系统和管理系统具有广泛的适用性。本书介绍了系统概念,系统学基础,运筹学基础,系统建模、优化和决策,系统经济分析,系统可靠性分析等。本书可作为大学工科、管理学科、经济学科等本科生教材,也可作为各类专业技术和管理人员的参考书。

图书在版编目(CIP)数据

系统工程概论(第2版)/杨家本主编. —武汉:武汉理工大学出版社,2007.6

普通高等学校自动化类专业新编系列教材

ISBN 978-7-5629-2566-8

I. 系… II. 杨… III. 系统工程-概论 IV. N94

中国版本图书馆 CIP 数据核字(2007)第 092161 号

出版发行:武汉理工大学出版社

武汉市武昌珞狮路 122 号 邮政编码 430070

<http://www.techbook.com.cn>

E-mail: wutpbook@sohu.com; wutpbook@tom.com; wutpbailh@163.com

经 销 者:各地新华书店

印 刷 者:武汉理工大印刷厂

开 本:787×1092 1/16

印 张:16.25

字 数:416 千字

版 次:2002 年 1 月第 1 版,2007 年 6 月第 2 版

印 次:2007 年 6 月第 5 次印刷

印 数:14001—19000 册

定 价:26.00 元

凡购本书,如有缺页、倒页、脱页等印装质量问题,请向出版社发行部调换。

本社购书热线电话:(027)87397097 87394412

普通高等学校自动化类专业新编系列教材

再版说明

自1998年7月教育部正式颁布实施新的普通高等学校本科专业目录至今,相应专业的培养目标、业务培养要求、主干课程、主要实践性教学环节等都有了不同程度的变化。2001年,武汉理工大学出版社响应教育部学科专业调整的政策,组织国内30余所大学的近100位教授共同编写了《普通高等学校自动化类专业新编系列教材》。本系列教材出版至今,得到了广大自动化专业教育工作者的支持和帮助,尤其要感谢那些参与编写这套教材的专家教授们和那些在使用这套教材并无私提出意见的专家学者们。同时,这套教材在对自动化教育教学方面的一些探索得到了广大读者的肯定和支持,具有一定的社会影响。

时过境迁,5年过去了,科技在发展,社会在进步,教学的手段、方式、理念和指导方针都发生了相应的变化,许多学校的专家教授们也提出了一些修改和修订意见,因此,武汉理工大学出版社决定于2006年进行《普通高等学校自动化类专业新编系列教材》的修订再版工作,以适应教学环境和教学内容变化要求。

过去的5年中,我们在自动化类专业教材的数字化方面做了很多有益的探索,全国近百所大学使用了本套教材及其配套交互式智能型多媒体课件。武汉理工大学出版社从交互式多媒体课件的制作、学习网站的建设到专业仿真软件的二次开发,都积累了相当的经验、取得了初步成果,在自动化专业教育界独树一帜。很多专家教授们都已经通过参加自动化教育年会、控制年会等学术活动,在武汉理工大学出版社的走访以及使用本套教材的过程中了解了自动化系列教材的数字化方面的进展、具体思路和成果。

“用信息化改造传统教学,让传统教学现代化”,“让教学中只有重点,没有难点”是我们的工作目标。我们将高度重视本次修订再版工作,兢兢业业,保证质量,恳请选用本教材的广大老师在使用过程中多提意见和建议,以便我们不断修订、补充、完善全套教材。

新的世纪是知识爆炸的时代,知识经济高速发展的时代,知识的载体、表现形式和传播方式都在发生着革命性的变化,面向新世纪的中国高等教育也正在经历前所未有的变革和发展。武汉理工大学出版社愿与各位专家、读者真诚合作,共同努力,实践和探索高等教育的发展。

武汉理工大学出版社

2007年2月

普通高等学校自动化类专业新编系列教材

出版说明

世纪之交,我国高等学校的人才培养工作正处在一个关键的历史时期。为了适应我国改革开放和社会主义现代化建设特别是社会主义市场经济体制对高等教育人才培养工作的新要求,为了适应世界科学技术发展的新趋势和新特点,原国家教育委员会组织对普通高等学校本科专业目录进行了第四次全面修订,并于1998年7月由教育部正式颁布实施。修订后的专业目录中,自动化类专业的专业面大大拓宽,相应的专业培养目标、业务培养要求、主干学科、主要课程、主要实践性教学环节等都有了不同程度的变化。要适应新的专业培养目标和教学要求,组织一套新的自动化类专业系列教材就成了当务之急。为此,武汉理工大学出版社在广泛调研的基础上,组织国内近30所大学的近100位教授共同编写了这套系列教材。

本套教材定位于普通高等学校自动化专业本科层次,遵照教育部颁发的《普通高等学校本科专业介绍》中所提出的培养目标和培养要求,依据2000年5月全国23所高等院校的70多位专家教授在武汉共同确定的指导思想和编写大纲进行编写,具有如下特点:

观念新——主动适应教学改革的需要和市场经济对人才培养的要求;

内容新——自动化技术在近20年来进展巨大,这套教材尽可能反映了这些内容;

体系新——在以前的基础上重构和重组,而非重建。各门课程及内容的组成、顺序、比例更加优化,避免遗漏和不必要的重复;

与国际接轨——自动化类专业教育要面向世界,面向未来,面向区域经济。在借鉴发达国家高等教育的专业模式和课程设置的同时,适当兼顾当前各地区经济文化发展不平衡的现状;

教学手段现代化——本套教材力求具有网络化、电子化、数字化的特色,大力推进电子讲稿和多媒体课件的出版工作。

“用信息化改造传统教学,让传统教学现代化”,“让教学中只有重点,没有难点”是我们的工作目标,近百所大学使用过的交互式智能型多媒体课件就是最好的证明。我们将高度重视,兢兢业业,保证质量,恳请选用本套教材的广大师生在使用过程中给我们多提意见和建议,以便我们不断修订、补充、完善全套教材。

21世纪已经到来,知识经济的曙光已经初现。面向新世纪的中国高等教育正在经历前所未有的变革和发展,人文与理工相通,科学与技术相融,教学与研究并重,知识与智慧同尊,以培养社会经济文化发展所需要的复合型人才,这是我国建立知识创新体系的重大挑战和空前机遇。我社愿与各位专家、读者真诚合作,共同努力,为新世纪的中国高等教育事业做出更大的贡献。

武汉理工大学出版社

2001年8月

普通高等学校自动化类专业新编系列教材
编审委员会

顾问:

郑大钟 熊有伦 戴冠中 萧德云 陈伯时 周祖德
项国波 席裕庚 褚 健

主任委员:

萧蕴诗 张崇巍 陈大钦 吴 坚 雷绍锋

委员(按姓氏笔画顺序):

马建国	王 辉	王孝武	王明阳	王建华	王俊杰
王 姮	文 方	方康玲	卢京潮	龙 伟	申功璋
叶春生	全书海	吕 锋	刘 泉	刘涤尘	刘京南
刘崇新	李汉强	李磊民	李裕能	宋靖雁	林 都
林 辉	林锦国	杨 波	杨天怡	杨家本	周泽义
胡 超	赵英凯	赵曾贻	侯朝桢	钟 珞	须文波
翁维勤	夏承铨	郭圣权	徐科军	黄席樾	章卫国
彭容修	程耕国	温阳东	曾庆军	谢克明	熊前兴
樊亚东	黎明森	戴文进			

编委会秘书:

黄 春

总责任编辑:

杨学忠 徐秋林

第2版前言

《系统工程概论》自2001年出版以来,至今已整整六年了。这次按照普通高等学校自动化专业系列教材修订计划,我们对《系统工程概论》进行了修订。

系统总体优化的概念和方法越来越被广大科技工作者所重视,掌握并运用系统工程的基本方法是科技工作者所必需的基础知识之一。系统工程作为处理一般系统的技术与方法,具有跨学科性质,它是对所有系统都普遍适用的方法,已广泛应用到工业、农业、国防及经济、社会等各个领域。

对本书的内容及结构,我们考虑了如下几个方面:我们按照系统科学的学科体系,对它的基础理论——系统学,和它的技术基础——运筹学作了基本的介绍,因此在第1章绪论之后,设置了第2章系统学基础和第3章数学规划基础、第4章图与网络;另外,也考虑到系统工程处理系统问题的步骤,设置了第5章系统优化、第6章决策分析、第7章系统分析与建模、第8章网络计划;另外,作为系统分析人员应该对系统可靠性和对工业系统具备基本的经济分析知识,因而设置了第9章系统可靠性分析和第10章工业技术经济;最后给出了第11章系统工程应用实例分析。

本书由清华大学杨家本主编,武汉理工大学黎明森、南京工业大学林锦国、清华大学宋靖雁副主编。具体编写分工为:第1章及各章的提要由杨家本编写;第3章、第4章、第7章、第8章由黎明森编写;第2章、第6章、第9章由林锦国编写;第5章、第10章、第11章由宋靖雁编写。

本次修订过程中,合肥工业大学黄云志老师、广西大学林靖宇老师、南京工业大学梁雪春老师、武汉科技学院罗维平老师等提出了很多具体的修改意见,在此表示感谢!

本书配套电子教案ppt(免费赠予)及含数百个动画的多媒体课件(需另购),会给老师们教学带来更大的方便。

由于我们水平有限,欢迎读者对本书中的错误和不妥之处提出批评意见。

作者

2007年6月

第 1 版前言

根据普通高等学校自动化类专业系列教材编写计划,我们编写了这本《系统工程概论》教材。系统工程是从系统的观点出发,对系统总体给出满意的规划、设计、决策、实现及管理。系统工程作为处理一般系统的技术与方法,具有跨学科性质,它是对所有系统都普遍适用的方法,已广泛地应用到工业、农业、国防及经济、社会等各个领域。我们认为,作为自动化专业,包括各个专业的学生,不仅应该掌握本学科所应具有的专业知识,还应该学习系统工程,具备处理系统问题的能力。

对本书的内容及结构,我们考虑了如下几个方面:我们按照系统科学的学科体系,对它的基础理论——系统学,和它的技术基础——运筹学作了基本的介绍,因此在第 1 章绪论之后,设置了第 2 章系统学基础和第 3 章数学规划基础、第 4 章图与网络;另外,也考虑到系统工程处理系统问题的步骤,设置了第 5 章系统优化、第 6 章决策分析、第 7 章系统分析与建模、第 8 章网络计划;另外,作为系统分析人员应该对系统可靠性和对工业系统具备基本的经济分析知识,因而设置了第 9 章系统可靠性分析和第 10 章工业技术经济;最后给出了第 11 章系统工程应用实例分析。

本书由清华大学杨家本主编,武汉理工大学黎明森、南京工业大学林锦国、清华大学宋靖雁副主编。具体编写分工为:第 1 章及各章的提要由杨家本编写;第 3 章、第 4 章、第 7 章、第 8 章由黎明森编写;第 2 章、第 6 章、第 9 章由林锦国编写;第 5 章、第 10 章、第 11 章由宋靖雁编写。

由于我们水平有限,欢迎读者对本书中的错误和不妥之处提出批评意见。

作 者

2001 年 3 月

目 录

1 绪论	(1)
1.1 系统及系统工程的定义	(1)
1.1.1 系统的定义与分类	(1)
1.1.2 系统工程的定义	(3)
1.2 系统工程在系统科学中的地位	(4)
1.3 系统工程的产生与发展	(4)
1.4 系统工程方法论	(6)
1.5 本书的体系结构	(7)
2 系统学基础	(9)
2.1 引言	(9)
2.2 基本概念	(9)
2.2.1 系统与环境	(9)
2.2.2 系统的状态与涨落	(11)
2.2.3 系统的熵与序	(12)
2.2.4 系统的自组织	(14)
2.3 一般系统论	(15)
2.3.1 一般系统论的形成	(15)
2.3.2 系统的整体性	(16)
2.3.3 系统的层次性	(17)
2.3.4 系统的目的性	(17)
2.3.5 系统的同形性	(17)
2.4 突变论	(19)
2.5 混沌理论	(21)
2.6 耗散结构理论	(24)
2.6.1 耗散结构的概念	(24)
2.6.2 布鲁塞尔模型	(25)
2.6.3 形成耗散结构的条件	(26)
2.7 协同学原理	(27)
2.7.1 协同学基本观	(27)
2.7.2 协同学原理的数学描述	(28)
2.7.3 协同学的应用	(28)
习题	(29)
3 数学规划基础	(30)
3.1 线性规划	(30)

3.1.1	基本概念	(30)
3.1.2	几何图形特征与图解法	(33)
3.1.3	单纯形法	(35)
3.1.4	线性规划的对偶问题	(40)
3.2	目标规划	(44)
3.2.1	目标规划的数学模型	(44)
3.2.2	目标规划的单纯形法	(44)
3.2.3	目标规划的图解法	(48)
3.3	整数规划	(49)
3.3.1	数学模型	(49)
3.3.2	整数规划的几何含义	(50)
3.3.3	分枝定界法	(51)
3.3.4	0—1 规划问题	(54)
3.4	非线性规划	(57)
3.4.1	区间消去法	(57)
3.4.2	Fibonacci 法	(58)
3.4.3	最速下降法(最优梯度法)	(61)
3.4.4	罚函数法	(66)
3.4.5	线性逼近法	(67)
3.5	动态规划	(69)
3.5.1	一个多段决策问题	(69)
3.5.2	动态规划的基本方程,最优性原理	(70)
3.5.3	应用举例	(72)
	习题	(77)
4	图与网络	(78)
4.1	图的基本概念	(78)
4.1.1	图与路	(78)
4.1.2	割集	(80)
4.2	最短路问题	(81)
4.2.1	基本概念	(81)
4.2.2	最短路的标号算法	(82)
4.2.3	从任一点到另外任一点的最短路算法	(84)
4.3	网络最大流	(86)
4.3.1	可行流与最大流	(86)
4.3.2	最大流——最小割定理	(87)
4.3.3	最大流算法	(88)
	习题	(91)
5	系统优化	(92)
5.1	系统优化方法概述	(92)
5.2	启发式算法和现代优化算法	(92)

5.3 遗传算法.....	(95)
习题.....	(104)
6 决策分析	(106)
6.1 决策概述	(106)
6.1.1 引言	(106)
6.1.2 发展与趋势	(106)
6.2 决策的基本要素及分类	(108)
6.3 决策模式与决策过程	(109)
6.3.1 问题分析	(109)
6.3.2 目标确定	(110)
6.3.3 制订方案	(110)
6.3.4 建立系统模型	(110)
6.3.5 优化、排序.....	(110)
6.3.6 决策	(110)
6.4 确定型决策	(111)
6.5 风险型决策	(111)
6.5.1 效用理论	(112)
6.5.2 表格法	(113)
6.5.3 决策树法	(113)
6.6 不确定型决策	(115)
6.6.1 等可能准则(Laplace 准则)	(115)
6.6.2 后悔值准则(Savage 准则)	(115)
6.6.3 悲观准则(max-min 准则)	(116)
6.6.4 乐观准则(max-max 准则)	(116)
6.6.5 折衷准则(Hurwicz 准则)	(116)
6.7 多目标决策	(117)
6.7.1 多目标决策问题的数学描述	(117)
6.7.2 多目标决策方法概述	(118)
6.7.3 效用函数法	(119)
6.7.4 层次分析法	(119)
6.8 决策支持系统	(123)
6.8.1 决策支持系统的定义	(123)
6.8.2 决策支持系统的作用及应用	(124)
习题.....	(124)
7 系统分析与系统建模	(125)
7.1 系统分析概述	(125)
7.1.1 系统分析的基本概念和作用	(125)
7.1.2 系统分析的要素、原则和步骤.....	(126)
7.1.3 系统分析的方法	(127)
7.1.4 系统分析应用举例	(130)

7.2 系统模型概述	(132)
7.2.1 系统模型的定义与特征	(132)
7.2.2 使用系统模型的必要性	(133)
7.2.3 系统模型的分类	(134)
7.2.4 使用数学模型的好处	(135)
7.3 系统建模方法	(135)
7.3.1 对系统建模的要求	(135)
7.3.2 系统建模应遵循的原则	(136)
7.3.3 系统建模的主要方法	(137)
7.4 状态空间法	(137)
7.4.1 系统的状态和状态变量	(137)
7.4.2 连续系统状态空间表达式	(138)
7.5 结构模型解析法(ISM)	(146)
7.5.1 有向连接图、相邻矩阵、可达性矩阵	(146)
7.5.2 可达性矩阵的分解——结构模型的建立	(147)
7.6 小样本问题的建模方法	(151)
7.6.1 对最小二乘估计的分析	(151)
7.6.2 LKL 估计	(152)
7.6.3 LKL 的八个步骤	(153)
习题	(155)
8 网络计划法	(156)
8.1 网络图	(156)
8.1.1 网络图的组成	(157)
8.1.2 网络图作图规则	(158)
8.1.3 绘制网络图的方法	(158)
8.2 计算网络图的时间参数	(158)
8.2.1 作业时间	(158)
8.2.2 节点最早开始时间	(159)
8.2.3 节点最迟完成时间 $T_L(j)$	(159)
8.3 作业的时间参数与计算	(161)
8.3.1 作业的最早开始时间 $t_{ES}(i,j)$	(161)
8.3.2 作业的最早完成时间 $t_{EF}(i,j)$	(161)
8.3.3 作业的最迟开始时间 $t_{LS}(i,j)$	(161)
8.3.4 作业的最迟完成时间 $t_{LF}(i,j)$	(161)
8.3.5 作业的总时间差 $R(i,j)$	(162)
8.3.6 作业的单时差 $r(i,j)$	(162)
8.4 关键路线与时差的关系	(162)
8.5 任务按期完成的概率分析与计算	(164)
8.5.1 任务完成时间近似符合正态分布规律	(164)

8.5.2 任务按期完成的概率计算	(165)
8.6 网络图调整与优化	(167)
8.6.1 缩短网络计划工期	(168)
8.6.2 网络执行过程中的检查与调整	(169)
8.7 网络的时间—费用分析与优化	(172)
8.7.1 直接成本与时间的关系	(172)
8.7.2 经济赶工方法	(173)
8.7.3 考虑间接成本的 CPM	(176)
8.8 图解评审技术——GERT	(176)
8.8.1 PERT 的发展——GERT	(176)
8.8.2 GERT 网络节点特征及其绘制实例	(178)
8.8.3 用 GERT 网络解决系统问题的步骤	(180)
习题	(180)
9 系统可靠性分析	(181)
9.1 可靠性概念及度量指标	(181)
9.1.1 可靠性的基本概念	(181)
9.1.2 可靠性的度量	(183)
9.2 系统可靠性模型	(185)
9.2.1 逻辑框图	(185)
9.2.2 串联系统可靠性模型	(186)
9.2.3 并联系统可靠性模型	(187)
9.2.4 混联系统的可靠性模型	(188)
9.2.5 表决系统可靠性模型	(188)
9.2.6 储备系统可靠性模型	(189)
9.2.7 桥式网络系统模型	(190)
9.3 可靠性计算及预估	(191)
9.3.1 事件树分析法	(191)
9.3.2 故障树分析法	(192)
9.3.3 上、下限法	(192)
9.4 可靠性设计与分配	(193)
9.4.1 代数分配法	(193)
9.4.2 加权分配法	(194)
9.4.3 动态规划法	(195)
9.5 可靠性工程与管理	(196)
9.5.1 提高可靠性的技术措施	(196)
9.5.2 提高人-机系统的可靠性	(197)
9.5.3 可靠性管理	(197)
习题	(198)
10 工业技术经济	(199)

10.1	项目投资效果静态分析法	(200)
10.1.1	投资收益率法	(200)
10.1.2	投资回收期	(201)
10.1.3	投资效果的评价依据	(202)
10.1.4	最小费用法	(203)
10.1.5	投资方案的可比性	(204)
10.2	项目投资效果动态分析法	(207)
10.2.1	投资方案的现金流量	(207)
10.2.2	净现值法	(207)
10.3	盈亏平衡分析	(208)
10.3.1	独立方案盈亏平衡分析	(209)
10.3.2	敏感性分析	(211)
	习题	(214)
11	应用实例分析	(216)
11.1	制造资源管理系统(MRP-Ⅱ)和企业资源计划(ERP)	(216)
11.1.1	MRP 的基本原理	(216)
11.1.2	MRP 的基本结构	(217)
11.1.3	闭环 MRP 和 MRP-Ⅱ	(218)
11.1.4	实施 MRP 系统的注意事项	(219)
11.1.5	企业资源管理(ERP)	(220)
11.2	计算机集成制造系统(CIMS)	(221)
11.2.1	CIMS 的概念	(221)
11.2.2	CIMS 的组成	(222)
11.2.3	实施 CIMS 应注意的问题	(223)
11.2.4	CIMS 应用实例——波音 767-X 并行设计工程	(223)
11.3	智能交通系统研究	(226)
11.3.1	系统集成环境信息流程设计	(228)
11.3.2	信号控制系统	(229)
11.4	城市交通系统建设的优化	(231)
11.4.1	城市交通系统建设优化中项目效益—费用分析的原则	(232)
11.4.2	城市交通系统建设优化寻找最优解的方法	(233)
11.4.3	各种项目排序方法的比较	(235)
11.5	区域规划的系统工程方法	(235)
11.5.1	区域规划的系统工程方法	(236)
11.5.2	区域科学典型模型和方法	(237)
11.5.3	实例分析——北京市城市发展模型及发展战略研究	(240)
11.5.4	区域发展策略优化分析	(243)
	习题	(244)
	参考文献	(245)

1 绪 论

本章提要

在客观世界中,处处存在着系统,包括小规模的系统,如:自动控制系统、企业生产经营管理系统等;大规模系统,如:经济系统、环境系统、城市系统等。因此,什么是系统?什么是系统工程?这是首先应该予以明确的最为基本的概念。系统工程作为处理一般系统总体最优的技术与方法,它在系统科学的学科体系中处于工程技术层次。随着社会的发展,人们所面临处理的系统日益庞大与复杂,在这种背景下,系统工程应运而生并不断发展。在本章中,还介绍了系统工程处理问题的一般方法和处理步骤。

1.1 系统及系统工程的定义

人类在早期并没有系统的概念,对自然界的认识,往往把现象或事物看成是彼此孤立的、割裂的、互不联系的。但是,随着社会和科学的发展,逐步改变了这种自然观,认为自然界是有着内部联系的统一整体,现象或事物之间都是相互依赖、相互联系、相互制约的。人类长期的实践活动和科学总结逐步抽象出系统的概念,产生了系统工程。

1.1.1 系统的定义与分类

系统 System 一词在古希腊就已使用,从词源上讲,它来自拉丁语 Systema,由词头“共同”和词尾“位于”结合而成,表示共同组成的群或是集合的概念。系统是一个涉及面广、内涵丰富的概念。

系统是系统科学、系统工程的最基本的概念,但是关于它的定义尚没有统一定论。萨多夫斯基曾经汇总过近四十种关于系统的定义。下面列举几种关于系统的定义:

《一般系统论》的创始人贝塔朗菲(L. V. Bertalanffy)认为:“系统可以定义为相互关联的元素的集合。”

前苏联学者 A. N. 乌约莫夫关于系统的定义是,“可以把系统定义为客体的集合,在这个集合上实现着带有固定性质的关系。”

日本工业标准对系统的定义是,“许多组成部分保持着有机的序,并向着同一个目标行动,就称作系统。”

韦氏辞典把系统定义为:“系统是有组织的或是组织化了的总体,以及构成总体的各种概念、原理和规则的相互作用及相互依赖诸要素的集合。”

钱学森等学者对系统的定义是,“系统是由相互作用和相互依赖的若干组成部分结合成的、具有特定功能的有机整体。”

对于这些定义,尽管表述不同,但是都共同地指出了系统的三个基本特征,第一,系统是由元素所组成的;第二,元素间相互影响、相互作用、相互依赖所构成的元素关系;第三,由元

素及元素间关系构成的整体具有特定的功能。

依据上述定义可以看出,系统几乎无所不在,我们所处的正是由各种系统所构成的客观世界。

我们可以从不同角度对系统进行分类,下面我们从研究对象、系统形成、系统结构、系统依赖时间变化、系统复杂性、系统不确定性和系统的开放性等角度,对系统分类如下:

(1) 按研究对象分类

系统按研究具体对象的不同分为工程系统、经济系统、教育系统、商业系统、城市系统、军事系统、环境系统、人口系统、社会系统等等。

(2) 从系统形成角度分类

从系统形成角度可以把系统分为自然系统和人造系统。自然系统是依据自然规律,不以人们意志为转移而形成的系统,如生态系统就属于自然系统;人造系统则相反,系统是按人们意志设计与实现的系统,如通信系统就属人造系统。

但是,我们应该看到,在自然系统和人造系统之间还存在大量的既包含自然形成的系统部分,又有人的参与,使得系统既有客观自然规律支配着系统的演化,又包含人的介入与参与并支配着系统的发展。这样的系统一般地可以称为有人参与的复杂系统。城市系统、环境保护系统等都属于这类系统。

(3) 按系统结构分类

按系统的结构不同,系统可分为集中系统、多级递阶系统和分散系统。集中系统指的是构成系统的元素,或称为子系统的行为受集中控制器集中控制,例如个人微型计算机的运行都受 CPU 的控制就属于集中系统,这类系统结构如图 1-1 中的(a)所示;多级递阶系统指的是,一部分子系统受上一级局部控制器控制,另一部分子系统受另外的上一级局部控制器控制,而这些上一级控制器又受更高级控制器的控制与协调,并依此可扩展至多层,这类系统结构称为多级递阶系统,例如行政办公系统就属于多级递阶系统,这类系统结构如图 1-1 中的(b)所示;分散系统指的是,各子系统不受上级控制与协调,而是独立运作,例如没有集中调度管理的城市交通管理系统就属于分散系统,这类系统如图 1-1 中的(c)所示。

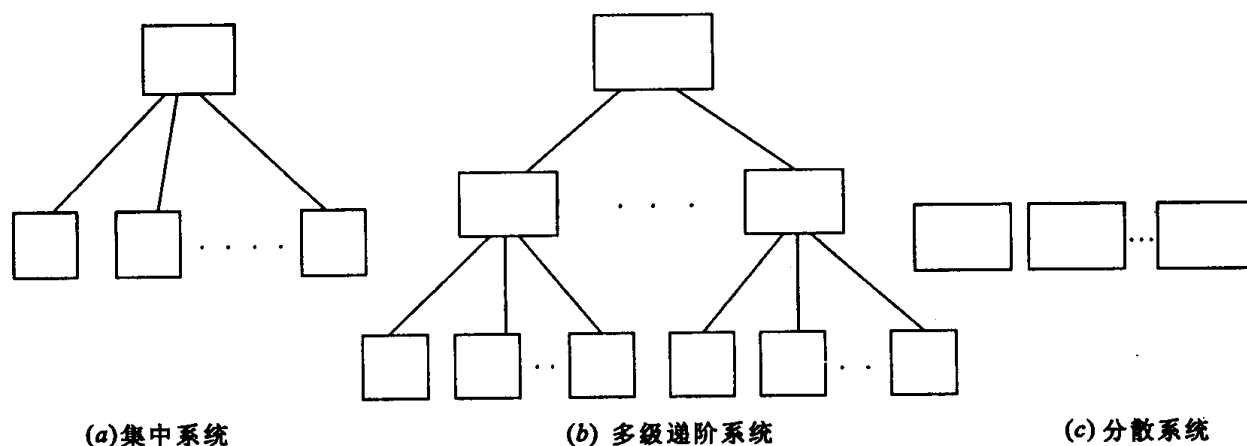


图 1-1 三种系统结构

这三类系统随着结构的不同,在处理系统建模与优化的复杂程度、系统可靠性等方面是不同的,各有优缺点,这方面的详细分析可参考文献[5]。在这三类系统结构中多级递阶系统是经常遇到的系统结构。

(4) 依据系统依赖时间变化分类

依据系统依赖时间演化情况,将系统分为静态系统和动态系统,前者被认为是系统不随时间演化,后者则被认为系统是随时间演化的。显然,绝大多数系统属于后者。

(5) 依据系统复杂性分类

如果组成系统元素数量少,关系不复杂,则称这类系统为简单系统;如组成系统元素数量庞大,关系复杂,则称这类系统为复杂系统。不过需要说明的是,什么是复杂?什么是复杂性?也是至今没有统一的明确概念。

(6) 依据系统是否具有不定性进行分类

如果系统是确定的,不包含不定性,则称其为确定性系统,如电路中电压、电阻、电流之间在一定环境条件下它们的关系是确定的,就属于确定性系统;如系统包含不定性,则称这类系统为不确定性系统,如股票系统就是不确定性系统。

(7) 按系统开放性分类

所谓系统开放性是指系统与环境是否存在物质、能量、信息的交换,如果系统与环境之间存在物质、能量与信息的交换,则称这类系统为开放系统;反之,则称为封闭系统。

开放性的概念是源自热力学,在热力学对系统开放性有更细致的划分,把系统分为孤立系统、封闭系统和开放系统,在此不再赘述。

在客观世界中,绝大部分系统都属于开放系统,如城市系统等。

(8) 根据组成系统的子系统数量、种类及关联复杂程度分类

钱学森根据组成系统的子系统数量和种类的多少,以及它们之间关联的复杂程度,把系统分为简单系统和巨系统两大类。在巨系统中分为简单巨系统和复杂巨系统,由于复杂巨系统又都是开放的,所以又称作开放的复杂巨系统,例如:生物体系统、人脑系统、人体系统、地理系统、星系系统都属于开放的复杂巨系统。在开放的复杂巨系统中,以有意识活动的人作为子系统的社会系统最为复杂,因而又称其为特殊的复杂巨系统。

当然,系统也可以从其他角度进行分类。

1.1.2 系统工程的定义

系统工程在系统科学的学科体系中属于工程技术层次,对于系统工程至今尚没有统一的定义。

1978年钱学森对系统工程的定义是,“系统工程是组织管理系统的规划、研究、设计、制造、试验和使用的科学方法,是一种对所有系统都具有普遍意义的方法。”

1975年美国科学技术辞典对系统工程解释为,“系统工程是研究复杂系统设计的科学,该系统由许多密切联系的元素所组成。设计该复杂系统时,应有明确的预定功能及目标,并协调各元素之间及元素和总体之间的有机联系,以使系统能从总体上达到最优目标。在设计系统时,要同时考虑到参与系统活动的人的因素及其作用。”

1977年日本学者秋山穰和西川智登把系统工程定义为,“系统工程是为了把对象创造出来或者在改善的时候,最优地并且最有效地达到该对象的目的,根据系统的思考方法,把它作为系统而进行开发、设计、制造及运行的思考方法、步骤以及各种方法的综合性的工程体系。”同年,三浦武雄认为,“系统工程的目的是研制一个系统,而系统不仅涉及到工程学的领域,还涉及到社会、经济和政治等领域,所以为了适当地解决这些问题,除了需要某

些纵向技术以外,还要有一种技术从横的方向把它们组织起来,这种横向技术就是系统工程。”

从以上各种对系统工程的定义可以看出,系统工程是以有人参与的复杂大系统为研究对象,按照一定的目的对系统进行分析与管理,以期达到总体效果最优的理论与方法。通俗地,对系统工程可以这样进行概括,就是,所谓“系统”就是对所研究的问题,从“总体”上去考察、分析与研究,所谓“工程”就是“最优”地去处理所研究的系统问题,因此,所谓系统工程就是寻求“总体最优”的理论与方法。再简化概括而又不失本质地去定义系统工程,可以认为系统工程就是处理系统的技术。

1.2 系统工程在系统科学中的地位

系统工程属于系统科学的学科范畴。系统科学研究系统演化的一般规律、系统有序结构的自组织原理和系统复杂性。系统科学是 20 世纪产生的,它的诞生是科学发展上的重大事件之一。

依据系统思想建立的完整科学体系称为系统科学。按照钱学森的观点,系统科学作为完整的学科体系,包含“基础科学、技术科学和工程技术”三个层次,他所建立的系统科学体系如图 1-2 所示。

在他的系统科学学科体系结构中基础科学指的是这个学科中的理论基础,它解释着这个学科中的一般规律,作为系统科学的理论基础就是“系统学”;技术科学指的是这个学科中的技术基础,它沟通着基础理论到实践应用、指导工程技术的实现,作为系统科学的技术基础就是“运筹学”、“控制理论”和“信息理论”;工程技术指的是这个学科中的应用技术,作为系统科学的应用技术就是“系统工程”。所以系统工程在系统科学的学科体系结构中处在工程技术层次。

1.3 系统工程的产生与发展

在中国,早在 2600 多年前就有了《周易》,其后,孔子所作的《易传》,老子创始的道家,庄子的《天运》,都试图对自然的演化、社会的发展作出统一解释,他们都把世界看成是一个由基本要素组成的,动态演化的,多层次的系统整体,主张从整体上把握这个基本要素组织起来的系统世界。

在西方,大体也同样的久远,古希腊泰勒斯(Thales)、赫拉克利特(Herakleitos)探索着组成万物的要素,德谟克利特(Demokritos)提出构成宇宙系统要素的原子论,以及亚里士多德(Aristoteles)的“整体大于各部分的总和”更是系统论的最基本思想。

进入 18 世纪,工业革命有力地推动着社会化大生产,同时也极大地推动着社会的前进。在这种背景下,马克思主义的诞生,标志着人类认识史上的一次伟大飞跃,把系统观作为对世界的总的看法包括在唯物辩证法中。

现代系统思想——以系统科学为标志的诞生,应该说是进入 20 世纪之后的事。随着生产的发展,社会的发展,人们所处理系统的规模越来越大,在这种形势下《系统工程》应运而生。回顾系统工程的产生与发展也是有一个过程:1911 年泰勒(F. W. Taylor)发表了《科学

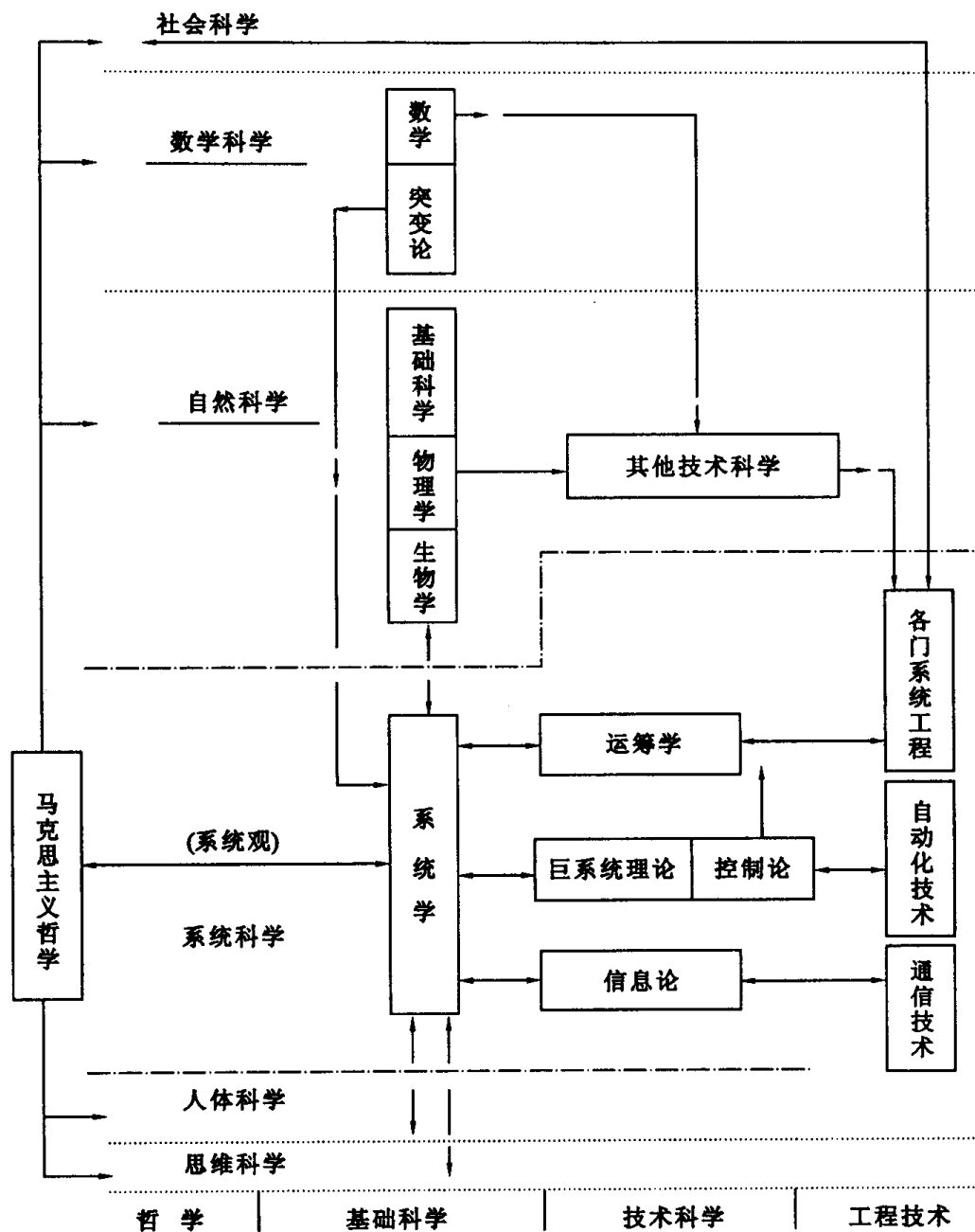


图 1-2 系统科学体系

管理原理》(Principle of Scientific Management)一书,它的基本思想是,一个工厂在原有设备的基础上,如果从整体系统的角度合理地安排生产,可以提高生产率。到了 20 世纪的 20~30 年代,美国贝尔电话公司发现要想发展电话通讯,不能单纯依靠加装电话机,还要从系统角度注意电话网络的发展。第二次世界大战,从战争需要的角度,很大地促进了作为系统工程的技术基础理论——运筹学的发展。战争的需要,不仅仅是武器的研制与生产,还要重视系统研究,如美国道格拉斯公司成立研究发展(Research And Development)部,它就是兰德(RAND)公司的前身。第一次出现“系统工程”(System Engineering)一词,应该说是以 H. H. Goode 发表《System Engineering》一书为标志。其后,1962 年 A. D. Hall 发表《系统工程方法论》一书,对系统工程概念、方法作了系统阐述。到了 1972 年 12 个国家在奥地利成立国际应用系统分析研究所 IIASA(International Institute of Applied Systems Analy-

sis),这个国际性系统分析研究所的建立,标志着人们从应用角度和更广泛的角度,开始关注着全球性系统问题。

在我国,中国科学院于1956年在力学研究所成立“运用(Operation)组”,即后来“运筹组”的前身。到1980年成立“系统科学研究所”,1980年成立“中国系统工程学会”,这些都标志着我国对系统工程研究发展的重视。到了1986年钱学森发表“为什么创立和研究系统学”,又把我国系统工程研究提高到系统工程基础理论,从系统科学体系的高度进行研究。随着社会的发展,人们所处理系统的综合性、复杂性及系统规模都在不断扩大。在这种背景下,系统思想及系统工程方法日益受到重视,它已渗透到各个应用领域,都力图于从总体上更好地把握系统的发展。在当前,不论是系统工程理论,或是系统工程方法论方面,以及跨学科应用方面都在快速发展。

1.4 系统工程方法论

各个学科根据本身的特点,形成了各个学科自己的方法论。作为系统工程方法论指的是,在处理系统问题过程中的一般分析问题和解决问题的途径、手段和方式,或者说是人们处理系统问题所应用的知识、观点去解决问题的行为方式。在系统工程方法论研究发展过程中,逐渐形成处理系统问题的思路及框架,其中1962年霍尔(A. D. Hall)在《系统工程方法论》一书中所提出的“三维结构”最为大家所公认。

霍尔的“三维结构”指的是,对于所处理的系统问题,从所需要专业知识、处理问题的思维步骤和处理问题的工作程序阶段,即相应称之为“知识维”、“逻辑维”和“时间维”的三维结构的系统工作模型,如图1-3所示。

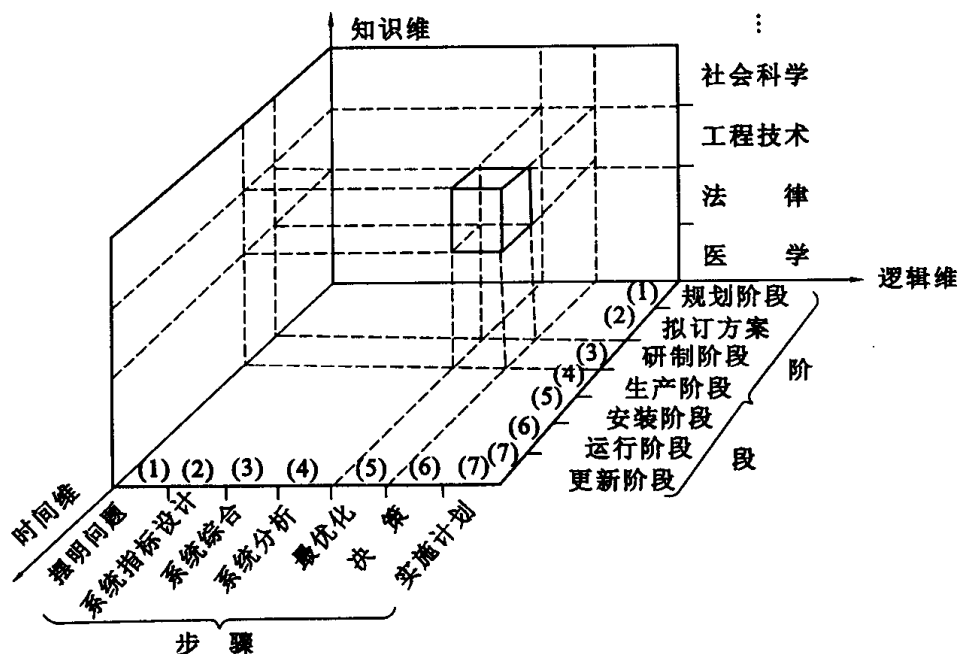


图 1-3 霍尔三维结构

霍尔三维结构中的知识维指的是处理系统问题所需要的知识,应该强调指出的是,由于系统的综合性,除了包含系统的专门知识外,还包括跨学科的知识,如:工程技术、经济、环

境、法律、社会、心理、艺术等多领域知识。

时间维指的分析处理系统的工作进程,一般分为7个阶段,即:① 规划阶段,包括:确定系统目标,分析环境条件及资源约束,制定规划;② 拟订方案,其实所指就是设计阶段,确定系统方案;③ 研制阶段,就是依据设计方案进行系统研制,并作出生产方案;④ 生产阶段,按照设计方案进行系统的零部件生产;⑤ 安装阶段,就是按照系统方案,把零部件组装成系统;⑥ 运行阶段,把组装成的系统投入运行,检验系统性能是否达到预期指标,并进一步提出改进方案;⑦ 更新阶段,系统经过运行,提出对原系统的改进方案,对原有系统进行更新。

逻辑维是指应用系统工程方法在处理各阶段系统问题的步骤,也大体分为7步,即:① 摆明问题(Problem Definition),在系统预调研的基础上,对系统和系统所处理环境,拟解决的系统问题,以及系统的未来发展作出基本分析;② 确定系统目标(Selecting Objectives),选择评价系统功能的指标体系、评价准则,这一步在系统分析中十分重要,因为系统目标一旦确定,系统将按照所确定的目标方向发展;③ 系统综合(System Synthesis),通俗地说,这一步骤就是提出并形成系统的可行方案,一般是多方案的;④ 系统分析(System Analysis),在这一步骤中,就是根据问题进行系统建模,以便分析、推断系统发展的各种可能结果;⑤ 系统优化分析(Selecting the Best System),基于系统模型,进行优化分析,对方案进行比较,判别方案优劣;⑥ 决策分析(Decision Analysis),决策与优化分析的区别在于,在决策中要考虑决策者对价值及风险的偏好,通过决策分析,对优化方案进行排序,寻求满意方案,推荐给决策者,对系统方案作出最终抉择;⑦ 实施计划(System Development),对所抉择的满意方案付诸实施。在进行上述过程中必须强调的是,由于人们的认识不能一次完成,应该进行多次的反复,不断修改系统目标、补充与扩展新的系统方案、完善系统模型、进行系统优化分析与决策分析,以作出更好的系统选择。我国学者钱学森于1989年提出“综合集成法”,是对系统工程方法论研究方面作出的新贡献。从定性到定量的综合集成法指的是,在处理开放复杂巨系统时,将专家群体、数据和各种信息与计算机仿真有机地结合起来,把各种学科的科学理论和人的经验知识结合起来,发挥综合集成系统的整体优势去解决复杂系统问题。复杂系统具有多变量、多层次、结构复杂、机制复杂、跨学科的特点,采用定性和定量相结合,科学理论和人的经验知识结合起来,依据系统思想把多种学科结合起来,把宏观研究与微观研究结合起来,通过系统分析,完成对复杂巨系统认识上从定性到定量的飞跃,所以综合集成法具有解决开放复杂巨系统问题的能力。

1.5 本书的体系结构

本书的内容分为11章,按系统科学体系结构——系统科学的基础理论、技术基础和工程技术三个层次进行安排和组织,它们的关系示于图1-4。

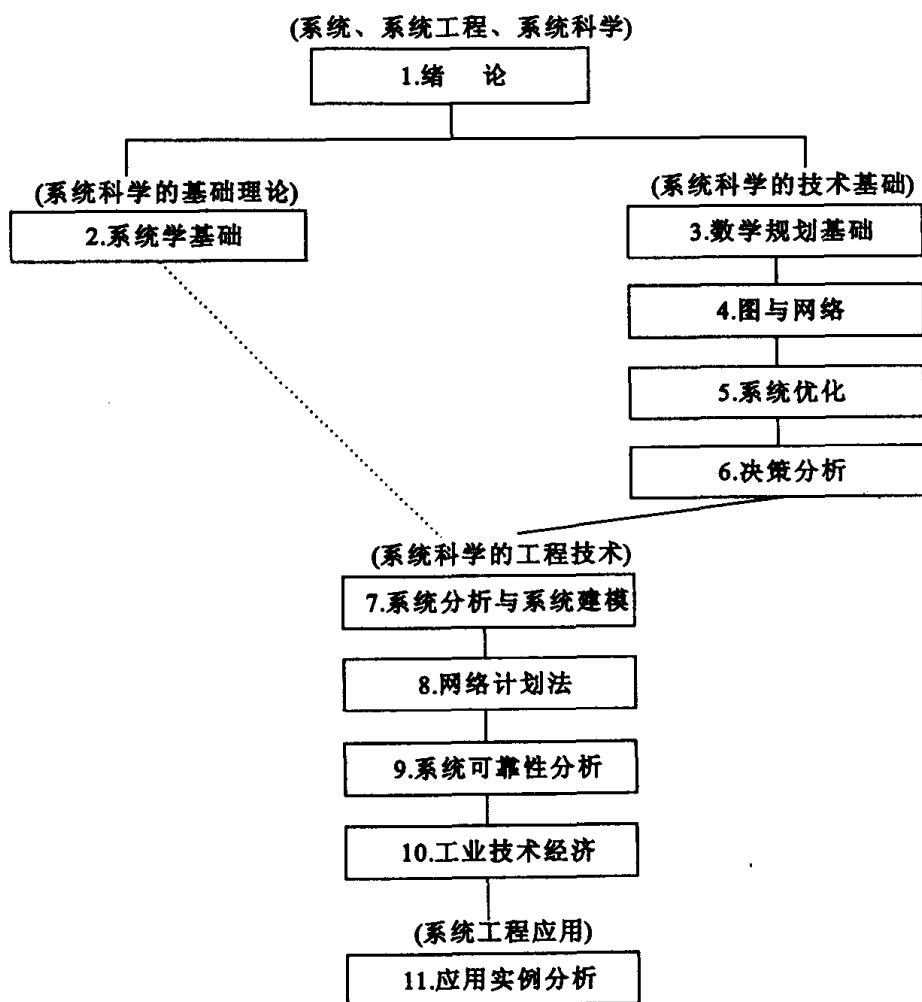


图 1-4 本书的体系结构

2 系统学基础

本章提要

系统科学的基础理论是“系统学”。系统学的研究对象是一般系统,尤其是复杂系统。系统学研究系统的演化规律,即系统是如何从无序演化到有序,或者从有序到新的有序,当然,也包括系统如何从有序到无序。无序与有序、混沌与自组织成为开放的非线性系统和系统复杂性研究的核心。

本章在叙述系统和系统演化有关基本概念的基础上,简要介绍了非线性系统可能发生的混沌,描述系统演化突然发生的数学工具——突变理论及有关系统自组织理论——耗散结构理论和协同学。同时遗憾地指出,系统学发展至今尚不完备,还有待于人们进一步发展与完善。

2.1 引言

系统科学是研究系统组成、结构及其演化与发展一般规律的科学,系统学则是其理论基础;运筹学、控制论、信息论是其技术基础;系统工程则是其具体应用技术。

人类很早就有了系统的思想,认为事物的发展不是孤立的、互不联系的,而应将自然界看成是相互联系的、相互制约的统一整体。古希腊的亚里士多德认为整体大于各部分之和,古代中国的老子也强调自然界的统一性。但作为系统科学的理论基础——系统学的建立仅仅是 20 世纪的事情。奥地利生物学家贝塔朗菲(L. V. Bertalanffy)于 1925 年提出了一般系统论的思想,于 1968 年正式发表著作《一般系统论:基础、发展和应用》;1969 年,比利时科学家普里高津(I. Prigogine)提出“耗散结构理论”;同年,德国科学家哈肯(H. Haken)提出了“协同学”概念,之后于 1977 年发表了著作《协同学导论》,1983 年发表了著作《高等协同学》;德国科学家艾根(M. Eigen)于 1971 年提出“超循环”概念,在 1979 年发表了著作《超循环理论》;另外,法国数学家托姆(R. Thom)于 1972 年发表了《结构稳定性和形态发生学》,提出了突变理论。这些理论和学说是当今系统学的主要内容。在介绍以上各个系统学分支或学说之前,下面引入并简要讨论一下系统学的几个基本概念。

2.2 基本概念

2.2.1 系统与环境

系统无疑是系统学中最基本的概念。迄今为止,关于系统的定义有数十种之多,它们从不同的方面对系统做了定义。如贝塔朗菲认为“系统是处于一定相互联系中的与环境发生关系的各组成部分的总体”;钱学森认为“把极其复杂的研究对象称为系统,即由相互作用和

相互依赖的若干组成部分结合成具有特定功能的有机整体,而且这个系统本身又是它们从属的更大系统的组成部分”。所以可以简要地定义“系统是具有某种功能或属性的若干元素和元素间关系的集合”,即构成系统的各元素的集合可能是具有某种功能的,也可能不具有任何功能仅具有某种属性。这可能是比较一般性的关于系统的定义。

一般地讲,构成某个系统的集合之外的一切,都是该系统的环境;每个系统又可以划分为若干个子系统;同时,每个系统又都是更大的、更高层次系统的子系统。

物质世界与精神世界的一切都可以划分为各种各样的系统,以便于人们对其进行研究,按构成系统元素的多少,可把系统分为大、中、小系统,元素特多的也可称为巨系统;按构成系统诸元素之间的关系可划分为简单系统和复杂系统;按系统的时间特性,可划分为静态系统和动态系统。更为重要的,从热力学角度,按系统与环境的关系,可划分为孤立系统、封闭系统和开放系统,这种划分在系统学理论上有特别的意义,下面即对这样划分的三种系统做一简要介绍。

(1) 孤立系统

如果系统与其环境之间既没有物质的交换,也没有能量的交换,就称其为孤立系统。图 2-1 是孤立系统示意图。S 表示系统, E 代表环境,带箭头的单直线表示物质流,带箭头的双直线表示能量流。在孤立系统中,系统与环境之间是相互隔绝的,系统内部的能量和物质不能传至系统外,系统环境的能量和物质也不能传至系统内。当然,从严格意义上讲也就没有信息的传递。显然,客观世界是不存在这种孤立系统的。

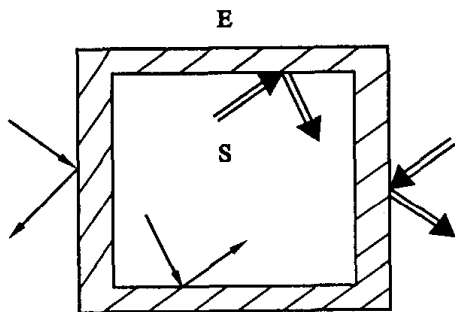


图 2-1 孤立系统

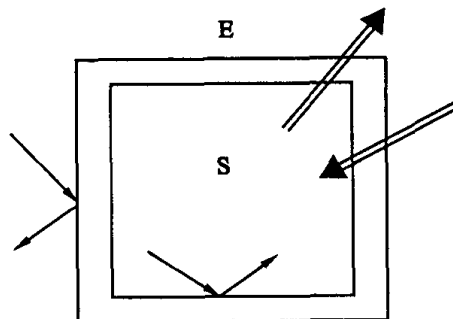


图 2-2 封闭系统

(2) 封闭系统

如果系统与其环境可以交换能量但不可以交换物质,称其为封闭系统。如图 2-2 所示。一个密闭的容器,可以与外界交换能量,但不能交换物质,可看作为封闭系统。

(3) 开放系统

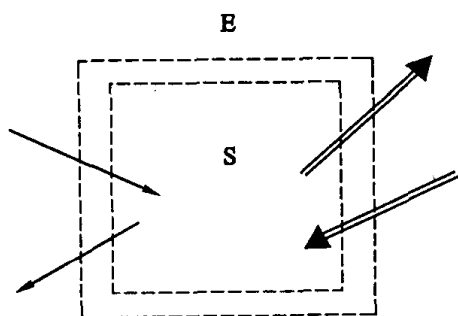


图 2-3 开放系统

如果系统与环境之间既有能量交换,又有物质交换,就称其为开放系统。显然,开放系统是普遍存在的系统,封闭系统和孤立系统都是其特例或近似系统。通常,有物质和能量的交换,就会伴随有信息的交换。小至细胞、原子、分子,大至生物、城市、国家等任何系统每时每刻都与环境进行着物质、能量及信息的交换,都是开放系统,而这种交换的方式、内容和程度则对该系统的运行和演化具有决定性的影响。

2.2.2 系统的状态与涨落

(1) 动力学状态

动力学系统的状态是描述系统所必需的最小一组变量,只要知道了在 $t = t_0$ 时的这组变量和 $t \geq t_0$ 时的输入,那么就能完全确定系统在任何 $t \geq t_0$ 时间的行为,这组变量叫做状态变量。因此,动力学系统在时间 t 的状态是由 t_0 时的状态和 $t \geq t_0$ 时的输入唯一确定的,而与 t_0 前的状态和输入无关。如要描述空间中一个质点的状态,需要该点在三维坐标上的位置和在三个方向上的速度。如系统中有 N 个质点,理论上,需要 $6N$ 个状态变量才能完全描述这个系统的行为。

(2) 热力学状态

由于在标准状态下, 1cm^3 气体含有 2.6875×10^{19} 个分子,如果采用动力学状态的变量来描述,每个分子需要 6 个状态变量, 1cm^3 气体就需要 1.6125×10^{20} 个状态变量。显然,这是难以求解的,也是没有必要的。因此,提出了热力学状态的问题。对于热力学系统,由于大量分子的持续无规则运动,其力学状态无穷多,但可以采用宏观平均统计量来描述热力学系统。如描述一定容积气体的状态,用压力 P 和温度 T 这两个热力学状态变量就可以了。

(3) 热力学平衡态与非平衡态

如果系统的热力学状态变量不再随时间而变化,称系统达到定态;在定态系统中,如果不存在物理量的宏观流动(如热流、粒子流等),则称该系统处于热力学平衡态。不具备上述任何一个条件的系统,则称其处于非平衡态。相应的系统可称为平衡态系统和非平衡态系统。

对于一个孤立系统,在初始时,无论其热力学状态变量值如何,随着时间的变化,最终系统将达到一种不变的定态,而且会没有物理量的宏观流动,是平衡态。所以,孤立系统的定态就是平衡态。

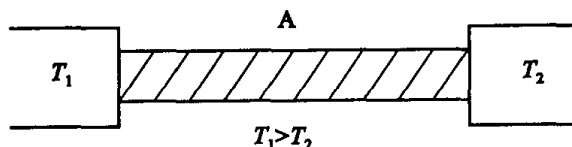


图 2-4 封闭系统达到定态但不是平衡态

对于一个金属棒 A,如图 2-4 所示,与理想的稳定热源 T_1 和 T_2 连接,如经过足够长的时间后,金属棒 A 上温度分布不再发生变化,即达到定态了。但由于从 T_1 到金属棒 A,再到 T_2 的热传递并没有停止,故不是热力学平衡态。

对于一个开放系统液位槽,如图 2-5 所示,输入流量为 Q_i ,输出流量为 Q_o ,且输入物质的流量和温度都恒定不变,因此,一段时间后,液位槽的液位和液体的温度都将保持不变,达到定态。但由于这时仍然有宏观的物流和能量流进出系统,所以不是热力学平衡态系统。可见封闭系统和开放系统的演化强烈地依赖于系统的环境条件,而孤立系统仅以平衡态作为演变的目标。

(4) 涨落

由于组成热力学系统的粒子数非常地大,人们不可能完全控制粒子的运动过程,而描述系统的宏观物理量是数量很少的,如温度、压力等。因此,在任一时刻,系统的实际物理量不

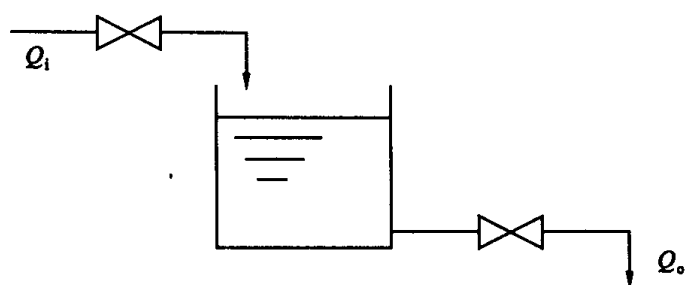


图 2-5 开放系统达到定态但不是平衡态

能够精确地等于现实统计平均量,多少有些偏离于平均量,这种偏离就叫做“涨落”。显然,涨落是杂乱无章的、随机的。它是系统在演化过程中,由于微观运动和环境干扰产生的对原来系统结构或运动轨道的微小偏离。

通常,涨落都是很小的,对需要的稳定性和可靠性影响极小,即使偶尔有较大的涨落,也会被耗散掉,因此,经常被忽略掉。然而,当系统处于临界点附近时,情况就不同了。这时,涨落可能被放大,导致系统的宏观变化,促成系统达到新的平衡状态。

2.2.3 系统的熵与序

(1) 熵

德国物理学家克劳修斯(R. J. E. Clausius)于 1865 年提出了熵(Entropy)这个概念,用符号 S 表示。如果一个物体的绝对温度为 T ,输入该物体的热量为 ΔQ ,则该物体熵的增加量为

$$\Delta S = S_2 - S_1 = \frac{\Delta Q}{T} \quad (2.1)$$

式中 S_1 ——物体输入热量前的熵;

S_2 ——物体输入热量后的熵。

这样定义的熵又称为热力学熵。热力学第二定律指出,一个系统的自发过程总是朝着热平衡的方向变化,它内部的温差越来越小,变化的结果是系统的熵越来越大。

温度描述的是系统冷热程度,而熵是系统的一个状态量,它描述的是系统的紊乱程度或有序程度。

玻尔兹曼(L. Boltzmann)于 1872 年在研究气体分子运动的过程中,对熵提出了微观解释。认为在由大量粒子(分子、原子)组成的系统中,熵表示系统的紊乱程度,系统越“乱”,熵就越大;系统越有序,熵就越小。

例如,设有 10 个粒子,分配到 2 个小室(1[#]室和 2[#]室)中,则各种可能出现的状态及其出现的概率如表 2-1 所示。为此引入配容数概念,当有 N 个粒子,分成 k 组,分到第 i 组的粒子数记为 N_i

$$\sum_{i=1}^k N_i = N$$

则配容数 W 为

$$W = \frac{N!}{\prod_{i=1}^k N_i!} \quad (2.2)$$

本例中, $N = 10, k = 2$, 则每种状态出现的概率为该状态的配容数比总配容数。

表 2-1 各种状态的配容数及出现概率

1# 室中的粒子数	2# 室中的粒子数	配 容 数	出现的概率
10	0	1	0.00098
9	1	10	0.00977
8	2	45	0.04395
7	3	120	0.11719
6	4	210	0.20508
5	5	252	0.24609
4	6	210	0.20508
3	7	120	0.11719
2	8	45	0.04395
1	9	10	0.00977
0	10	1	0.00098

如果两个小室是由一个容器加一隔板构成的, 那么, 拿掉这块隔板, 容器中粒子将趋于概率较大的均匀分布。当粒子数很多时, 如 $N = 6.023 \times 10^{23}$, 即 1mol 气体时, 拿掉隔板, 容器中的粒子将均匀分布于整个容器中。一般认为, 均匀分布是一种无序的状态, 不均匀是一种有序的状态, 一旦分布趋于均匀后, 就不会自动地向不均匀方向变化。

因此, 玻尔兹曼定义熵为

$$S = k \ln W \quad (2.3)$$

式中 k ——玻尔兹曼常数。

根据玻尔兹曼的解释, 热力学定义的熵可以看作为能量在空间分布均匀性的度量, 也就是物质系统中能量衰竭程度的度量。能量分布越均匀, 系统的无序程度越高; 反之, 能量分布越不均匀, 系统的有序程度越高。

(2) 信息熵

早在 1928 年哈特莱(Hartley) 就在“信息技术”一文中提出了一种度量信息的方法。设有一个符号表, 共有 S 种符号, 发信者选择 N 个符号构成一个符号系列。因此, 每选定一个符号, 就否定或排除了若干种可能的序列, 即消除了部分不确定性。显然, S^N 越大, 信息量越大。定义信息量为

$$H = \lg S^N = N \lg S \quad (2.4)$$

1948 年香农(C. E. Shannon) 继承和发展了哈特莱的思想, 并引入玻尔兹曼的熵概念, 将熵作为一个随机事件的不确定性或信息量的量度, 提出了信息熵的概念。设有一个概率实验 A (随机事件) 有 n 个可能的结局事件: $x_1, x_2, \dots, x_n$; 每一个事件的出现概率分别为 $p_1, p_2, \dots, p_n$, 且满足下列条件:

$$0 \leq p_i \leq 1 \quad i = 1, 2, \dots, n \quad (2.5)$$

$$\sum_{i=1}^n p_i = 1 \quad (2.6)$$

定义

$$H = -k \sum_{i=1}^n p_i \ln p_i \quad (2.7)$$

为信息熵或香农熵,是概率实验 A 实验结果的不确定性的度量,式中, k 为玻尔兹曼常数。在实验前, H 是实验结果不确定性的度量;在实验完成之后,它是人们从该实验中所得信息量的度量。

(3) 序

序是描述系统各元素之间关系状况的一个概念。一盒黄砂,没有任何结构构造时,处于最无序状况;当把黄砂构造为各种河流、道路、房屋等模型时,它就具有了较原来为高的组织程度,有序化程度得到提高。一滴墨汁滴入一杯清水中,开始时可见到墨汁的扩散过程,即呈现一定的构造,随着扩散的持续,最终杯中达到完全均匀的状态,是一杯淡淡的黑水,一般认为这时的系统是无序的。这样的扩散过程是不可逆的,因为要把一杯黑水分离为墨汁和清水是不能自动进行的。同样,能量不可能自发地从低温物体向高温物体传输。

从熵的角度看,系统总是自发地向着熵增大的方向,即无序的方向发展。系统要从无序走向有序,必须从外界输入负熵。一块磁铁矿石在被磁化之前,它的各个磁畴磁极呈现的是杂乱无章的状态,是无序的;一旦输入负熵,使之磁化后,各磁畴按磁性有规律地排列,就出现了有序的状态。

2.2.4 系统的自组织

系统中的元素在环境作用下,不依靠外力,发展形成有序结构的过程,称为系统自组织。目前,人们已经发现了许多自组织现象,这对于研究系统如何从无序到有序的演化有重大意义。

19 世纪末化学家利色根发现,将碘化钾溶液加入到含有硝酸银的胶体介质中,在一定的条件下,所形成的碘化银沉淀物会构成一圈圈有规律间隔的环状分布,这种有序的环称为利色根环。

1900 年贝纳德(Bénard)进行了如下流体力学实验:在一个平底容器中装有液体,在容器底部均匀加热,见图 2-6(a),由于液体上下界面温度不同而产生温度梯度,在温度梯度小于某个特征值之前,液体中的传热方式为传导;当温度梯度达到某个特征值时,液体中的传热方式呈现如图 2-6(b)所示情景。液体中自发形成排列整齐的、由微小六角形单元组成的有序结构,称为“贝纳德花纹”。在每个六角形单元中,中心部分的液体上升,周边部分的液体下降,形成宏观对流。这是一种高度有序化的组织系统。

1958 年物理学家贝洛索夫(Belousov)在某些金属离子催化剂存在的条件下进行某些有机酸的溴酸氧化实验时,发现在一定的条件下,某些组分的浓度呈现出周期性的变化,形成时间上的有序结构,如造成反应液不是稳定的单色,而是一会儿红、一会儿蓝,像一个时钟的钟摆一样发生规则的震荡。这类现象被称为“化学钟”。后来化学家萨博廷斯基(Zhabatinski)继续改进这类实验,又发现了在空间上的有序结构,反应浓度分布会变得不均匀,形成许多漂亮的花纹。

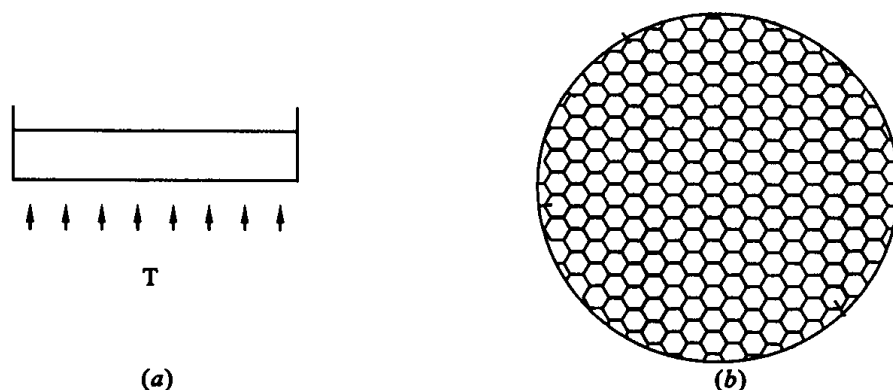


图 2-6 贝纳德花纹

自组织现象广泛地存在于自然界和社会经济系统中。有些已被人们认识并加以利用，如激光的产生就是一个典型的自组织过程；有些则还没有完全弄清楚其机理，如生物的进化也是一个自组织过程，其中的许多谜至今仍然没有解开。

2.3 一般系统论

2.3.1 一般系统论的形成

自 20 世纪 20 年代起，以奥地利生物学家贝塔朗菲 (L. V. Bertalanffy) 为代表的一批多个学科的科学家们致力于系统研究，寻求“不同系统、不同学科之间的共同语言和术语”。1937 年，贝塔朗菲首次明确提出了一般系统论原理，1954 年成立了“一般系统论研究会”。到 60 年代，一般系统论的思想和方法已经产生了较大的影响。1968 年，贝塔朗菲发表了奠基性的专著《一般系统论——基础、发展与应用》。一般系统论原理的提出早于维纳的控制论 (1948 年) 和香农的信息论 (1948 年)，但代表性著作的发表又晚于控制论和信息论，可见，他们的思想观点是相互影响、相互推动、深入发展的。

一般系统论采用微分方程组来描述系统。设 Q_i 是元素 $p_i (i=1, 2, \dots, n)$ 的某个测度，对于有限数量的元素，其测度有如下形式：

$$\left. \begin{aligned} \frac{dQ_1}{dt} &= f_1(Q_1, Q_2, \dots, Q_n) \\ \frac{dQ_2}{dt} &= f_2(Q_1, Q_2, \dots, Q_n) \\ &\vdots \\ \frac{dQ_n}{dt} &= f_n(Q_1, Q_2, \dots, Q_n) \end{aligned} \right\} \quad (2.8)$$

上式表明，任何测度 Q_i 的变化都是所有测度 (从 Q_1 到 Q_n) 的函数，而任何测度 Q_i 的变化又会影响所有测度甚至整个系统的变化。

当系统处于定态时，

$$\frac{dQ_i}{dt} = 0 \quad (2.9)$$

则

$$f_1 = f_2 = \cdots = f_n = 0 \quad (2.10)$$

得方程组的定态解分别为: $Q_1^*, Q_2^*, \cdots, Q_n^*$ 。

这些定态解有些是稳定的,有些可能是不稳定的,引入新变量

$$Q'_i = Q_i^* - Q_i$$

于是方程(2.8)可写成为

$$\left. \begin{aligned} \frac{dQ'_1}{dt} &= f'_1(Q'_1, Q'_2, \cdots, Q'_n) \\ \frac{dQ'_2}{dt} &= f'_2(Q'_1, Q'_2, \cdots, Q'_n) \\ &\vdots \\ \frac{dQ'_n}{dt} &= f'_n(Q'_1, Q'_2, \cdots, Q'_n) \end{aligned} \right\} \quad (2.11)$$

假定这个系统可以用泰勒级数展开, a_{ij} 代表泰勒系数,于是得到

$$\left. \begin{aligned} \frac{dQ'_1}{dt} &= a_{11}Q'_1 + a_{12}Q'_2 + \cdots + a_{1n}Q'_n + a_{111}Q'^2_1 + a_{112}Q'_1Q'_2 + a_{122}Q'^2_2 + \cdots \\ \frac{dQ'_2}{dt} &= a_{21}Q'_1 + a_{22}Q'_2 + \cdots + a_{2n}Q'_n + a_{211}Q'^2_1 + a_{212}Q'_1Q'_2 + a_{222}Q'^2_2 + \cdots \\ &\vdots \\ \frac{dQ'_n}{dt} &= a_{n1}Q'_1 + a_{n2}Q'_2 + \cdots + a_{nn}Q'_n + a_{n11}Q'^2_1 + a_{n12}Q'_1Q'_2 + a_{n22}Q'^2_2 + \cdots \end{aligned} \right\} \quad (2.12)$$

方程(2.12)的一般解为

$$\left. \begin{aligned} Q'_1 &= G_{11}e^{\lambda_1 t} + G_{12}e^{\lambda_2 t} + \cdots + G_{1n}e^{\lambda_n t} + G_{111}e^{2\lambda_1 t} + \cdots \\ Q'_2 &= G_{21}e^{\lambda_1 t} + G_{22}e^{\lambda_2 t} + \cdots + G_{2n}e^{\lambda_n t} + G_{211}e^{2\lambda_1 t} + \cdots \\ &\vdots \\ Q'_n &= G_{n1}e^{\lambda_1 t} + G_{n2}e^{\lambda_2 t} + \cdots + G_{nn}e^{\lambda_n t} + G_{n11}e^{2\lambda_1 t} + \cdots \end{aligned} \right\} \quad (2.13)$$

式中 G 为常数, λ 为特征方程

$$\begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix} = 0$$

的根。根 λ 可以是实数,也可以是虚数。考察方程(2.13)发现,如果所有 λ 为负实数;或者虽然是复数,但其中的实数部分是负数;则 Q'_i 随时间增加而趋向 0。按照定义, $Q_i = Q_i^* - Q'_i$, 得到 Q_i 等于定态值 Q_i^* 。这时,平衡是稳定的。如果有一个 λ 是正数或是 0,平衡就不稳定,系统就要离开平衡态。

贝塔朗菲进一步归纳并提出了一般系统的性质,有整体性、同形性、目的性、层次性和动态性等,下面分别简要介绍。

2.3.2 系统的整体性

贝塔朗菲认为系统是最重要的概念,而系统的首要显著特征是整体性。他借用亚里士多德的一句名言“整体功能大于各部分功能之和”来表达这一原理。有人称之为系统第一定

律。认为系统不是各孤立部分属性的简单叠加,它还具有各孤立部分所没有的新的性质和行为。

系统的整体性质有时也通俗地表达为“ $1+1>2$ ”。但实际情况是复杂的,也有可能等于2或小于2,这取决于系统的结构、各部分的属性及系统内协同作用的强弱。这主要是从系统的效能角度来理解的。

从系统中各部分间的关系考察,有如下三种情况:

(1) 累加性。如果方程(2.12)中所有的系数 a_{ij} ($i \neq j$) 都等于零,则方程组(2.12)退化为每个元素的变化只决定于该元素本身。因此,每个元素都独立于其他元素。此时,整个系统的变化就等于各个元素单独变化之和,称该系统具有累加性质。

(2) 逐步分离。如果系数 a_{ij} ($i \neq j$) 不是常数且随时间逐渐变小,

$$\lim_{t \rightarrow \infty} a_{ij} \rightarrow 0 \quad (2.14)$$

则该系统是逐步分离系统。说明这样的系统将从最初的统一系统,逐步减弱各元素之间的关系,逐步分离,最终各个元素会成为相互独立的部分。如胚胎发育过程就是从等势发展的整体逐渐分化为各个器官。因此,逐步分离系统有时又称为“发育系统”。

(3) 逐渐集中。如果元素 P_i 的系数在每个方程中都比较大,其他元素的系数较小甚至为零,则系统将以 P_i 元素为中心, P_i 称为主导部分。这时, P_i 的微小变化,就会导致整个系统的变化。目前认为基因决定着生物个体的发展和未来目标,可以认为基因就是系统中的主导部分。

2.3.3 系统的层次性

系统存在层次性是显然的,如在微观方向上,系统可以逐级分解为更小的子系统,如物质都由分子组成,分子由原子组成,原子由原子核与电子组成等;在宏观方向上,有动物、植物、区域、国家、地球、太阳系、银河、宇宙等。

高层次系统由低层次系统组成。因此,原来低层次系统的属性,仍然存在于高层次系统中。另外,又由于多个低层次系统的相互作用,产生了新的属性,因此,从系统的角度看, $1+1>2$ 总是成立的。所以,一个系统的层次越高,系统的结构、功能与属性就越复杂。

2.3.4 系统的目的性

目的性也可称为终极性或预决性。贝塔朗菲认为“一个系统的发展方向,取决于系统的预决性”。一个系统的发展趋势,不仅取决于实际系统的状态(必然性),而且还取决于对系统未来的预测(偶然性),即要受将来达到的最后状态的制约,这两者的结合就是系统的目的性。在包括人这样的对系统未来具有预测能力的社会经济系统中,可以见到很多这样的案例;在包括电脑或机器人等人工智能体这样的对系统未来具有预测能力的系统中,也可能具有预决性。

2.3.5 系统的同形性

同形性也称同型性,是指多种不同系统中存在着一般化的发展模式和发展规律,而不管系统的种类和系统的具体组成如何。即本质不同的领域会出现系统结构同一的情形。例如:

(1) 指数增长

当系统只有一种元素时,系统演化方程简化为方程(2.15)

$$\frac{dQ}{dt} = f(Q) \quad (2.15)$$

其泰勒级数展开式为

$$\frac{dQ}{dt} = a_1 Q + a_2 Q^2 + \dots \quad (2.16)$$

只保留级数第一项时,有

$$\frac{dQ}{dt} = a_1 Q \quad (2.17)$$

其解为

$$Q = Q_0 e^{a_1 t} \quad (2.18)$$

式中, $Q_0 = Q(t) |_{t=0}$ 。这是一个指数曲线模型,在社会、经济、生物等领域有广泛的应用。如图 2-7 所示。 $a_1 > 0$, 曲线随时间上升, 对应于马尔萨斯人口模型、经济增长、细菌繁殖等情形; $a_1 = 0$, 系统零增长; $a_1 < 0$, 曲线随时间下降, 对应于物种灭绝、放射性衰变、细菌中毒死亡等情形。

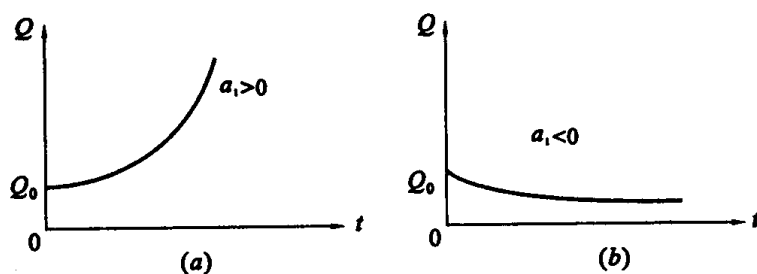


图 2-7 指数曲线

(2) S 型饱和

当公式(2.16)中仅保留前两项时,系统的图形就是所谓的 S 型曲线,也称逻辑斯迪克(Logistic)曲线。方程为

$$\frac{dQ}{dt} = a_1 Q + a_2 Q^2 \quad (2.19)$$

其解为

$$Q = \frac{a_1 C e^{a_1 t}}{1 - a_2 C e^{a_1 t}} \quad (2.20)$$

式中, C 是决定于初始条件的初值。S 型曲线如图 2-8 所示。马尔萨斯人口模型考虑资源限制时,进行修正,成为 S 型曲线;新产品问世至市场饱和通常也呈现为 S 型曲线发展过程。

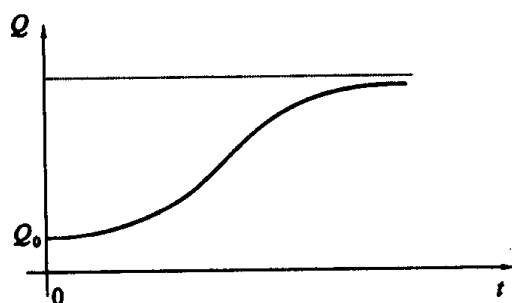


图 2-8 S 型曲线

(3) 依存与竞争

当考虑有两个元素 i 和 j 的系统,且所有系数 $a_{ij} = 0$ ($j \neq i$ 时),则每个元素的增长取决于该元素自身,令 $a_{ii} = a_i$,

得到

$$\left. \begin{aligned} \frac{dQ_1}{dt} &= a_1 Q_1 \\ \frac{dQ_2}{dt} &= a_2 Q_2 \end{aligned} \right\} \quad (2.21)$$

其解为

$$\begin{aligned} Q_1 &= C_1 e^{a_1 t} \\ Q_2 &= C_2 e^{a_2 t} \end{aligned} \quad (2.22)$$

由于两个元素都独自随时间变化,它们之间保持着一定的比例关系或依存关系

$$\frac{Q_1}{C_1} = \left(\frac{Q_2}{C_2} \right)^{\frac{a_1}{a_2}} \quad (2.23)$$

在泰勒展开式中再增加一个相关项时,方程为

$$\left. \begin{aligned} \frac{dQ_1}{dt} &= \frac{r_1}{k_1} Q_1 (k_1 - Q_1 - \alpha Q_2) \\ \frac{dQ_2}{dt} &= \frac{r_2}{k_2} Q_2 (k_2 - Q_2 - \beta Q_1) \end{aligned} \right\} \quad (2.24)$$

例如,在生物系统中,

Q_1, Q_2 表示系统中两个物种的个体数量;

r_1, r_2 为 Q_1, Q_2 的自然增长率;

k_1, k_2 为 Q_1, Q_2 单独增殖时的饱和密度;

α, β 是竞争系数。

在这个生物系统中,

当 $\alpha < \frac{k_1}{k_2}, \beta < \frac{k_2}{k_1}$, Q_1 与 Q_2 共存;

当 $\alpha < \frac{k_1}{k_2}, \beta > \frac{k_2}{k_1}$, 只有 Q_1 生存下来;

当 $\alpha > \frac{k_1}{k_2}, \beta < \frac{k_2}{k_1}$, 只有 Q_2 生存下来。

2.4 突变论

突变论既是系统学的一个分支也是数学的一个分支,它以不连续现象为研究对象。突变现象广泛地存在于自然、社会、经济和生活中。如物理学中的相变、地震的爆发与火山的喷发、楼房的突然倒塌、经济危机、政治危机、生物体的病变等等都是突然发生的现象。由于这些突变,都是事先难以预料和把握的,他们的发生对于原先系统及环境的影响是巨大的,可能带来灾难性的后果,也可能带来积极性的后果。

法国数学家托姆(R. Thom)于1965年创立了突变论。这开拓了人们认识系统突变现象的眼界,使人们科学地预测和恰当地处理突变现象的可能性大大增加。

突变论认为突变现象的本质是系统从一种稳定状态到另一种稳定状态的跃迁。因此,系统的结构稳定性是突变理论的研究重点。突变理论可描述为

若有状态空间 $X, x \in X$; 控制参数空间 $\Lambda, \lambda \in \Lambda$; 用势函数 $V(x)$ 描述系统的突变行为。

则集合

$$M = \{(\lambda, x) \in \Lambda \times X; \text{grad}_x V(\lambda, x) = 0\} \quad (2.25)$$

就确定了一个突变流形;突变集合由 M 在 Λ 空间上投影中的临界点组成。自然界中的一切突变形式都可以依据系统的控制参数空间 Λ 和状态空间 X 的维数来分类,控制参数空间的维数就决定了突变模式的总数。托姆按照系统发生突变时,控制参数在其所在空间上的形状来命名突变模型。状态变量数小于 3,控制参数个数小于 6 的突变模型列于表 2-2。

表 2-2 突变模型

序号	名称	势函数	状态变量数	控制参数个数
1	折叠(fold)	$x^3 + ax$	1	1
2	尖角(cusp)	$\pm x^3 + ax^2 + b$	1	2
3	燕尾(swallow-tail)	$x^5 + ax^3 + bx^2 + cx$	1	3
4	蝴蝶(butterfly)	$\pm x^6 + ax^4 + bx^3 + cx^2 + dx$	1	4
5	印第安人茅舍(wigwam)	$x^7 + ax^5 + bx^4 + cx^3 + dx^2 + ex$	1	5
6	椭圆脐点(elliptic umbilic)	$x_1^3 - x_1 x_2^2 + a(x_1^2 + x_2^2) + b x_1 + c x_2$	2	3
7	双曲脐点(hyperbolic umbilic)	$x_1^3 + x_2^3 + ax_1 x_2 + b x_1 + c x_2$	2	3
8	抛物脐点(parabolic umbilic)	$x_1^4 + x_1^2 x_2 + ax_1^2 + bx_2^2 + cx_1 + d x_2$	2	4
9	符号型脐点(symbolic umbilic)	$x_1^3 \pm x_2^4 + ax_1 x_2^2 + bx_2^2 + cx_1 x_2 + d x_1 + ex_2$	2	5
10	第二椭圆脐点(second elliptic umbilic)	$x_1^2 x_2 - x_2^5 + ax_2^3 + bx_2^2 + cx_1^2 + d x_1 + ex_2$	2	5
11	第二抛物脐点(second parabolic umbilic)	$x_1^2 x_2 + x_2^5 + ax_2^3 + bx_2^2 + c x_1^2 + d x_1 + ex_2$	2	5

对于一维势函数 $V(x)$,其泰勒级数展开式为

$$V(x) = c^{(0)} + c^{(1)}x + c^{(2)}x^2 + \cdots + c^{(m)}x^m + \cdots \quad (2.26)$$

式中

$$c^{(0)} = V(x)|_{x=x_0}$$

$$c^{(1)} = \frac{d}{dx}V(x)|_{x=x_0}$$

$$c^{(2)} = \frac{1}{2!} \frac{d^2}{dx^2}V(x)|_{x=x_0}$$

$$\vdots$$

$$c^{(m)} = \frac{1}{m!} \frac{d^m}{dx^m}V(x)|_{x=x_0}$$

通过移动坐标,可实现

$$V(x)|_{x=x_0} = 0 \quad (2.27)$$

在平衡点有

$$\frac{d}{dx}V(x) = 0 \quad (2.28)$$

当忽略 $V(x)$ 展开式中的 5 阶及 5 阶以上的高阶项时,有势函数

$$V(x) = c^{(0)} + c^{(1)}x + c^{(2)}x^2 + c^{(3)}x^3 + c^{(4)}x^4 \quad (2.29)$$

通过移动 x 轴的原点,可去掉 x^3 项,得到其一般式(2.30)为

$$V(x) = \frac{1}{4}x^4 + \frac{1}{2}ax^2 + bx \quad (2.30)$$

在平衡点 $\frac{d}{dx}V(x)=0$, 则有

$$\frac{d}{dx}V(x)=x^3+ax+b=0 \quad (2.31)$$

将上式画在 (x, a, b) 空间上, 得到图 2-9。

在图 2-9 中可见到下列几个现象:

A 突跳(catastrophe jump), 如 $C_1 \rightarrow D$;

B 分叉(bifurcation), 如 $A \rightarrow B_1, B_2$;

C 不可到达区(inaccessible region), 如上图的折面阴影区;

D 延迟(hysteresis), 如 $CC_1 \rightarrow D$ 和 $DD_1 \rightarrow E$ 。

由于在发生突变时, 控制参数在控制空间上的形状如尖角, 因此, 托姆将其定名为“尖角突变”模型。

当忽略势函数 $V(x)$ 中的 6 阶及 6 阶以上的高阶项时, 得

$$V(x)=\frac{1}{5}x^5+\frac{1}{3}ax^3+\frac{b}{2}x^2+cx \quad (2.32)$$

在平衡点 $\frac{d}{dx}V(x)=0$, 则有

$$\frac{d}{dx}V(x)=x^4+ax^2+bx+c=0 \quad (2.33)$$

该方程的控制参数空间 (a, b, c) 图形如图 2-10 所示, 形如燕尾, 故托姆将其定名为“燕尾型突变”模型。

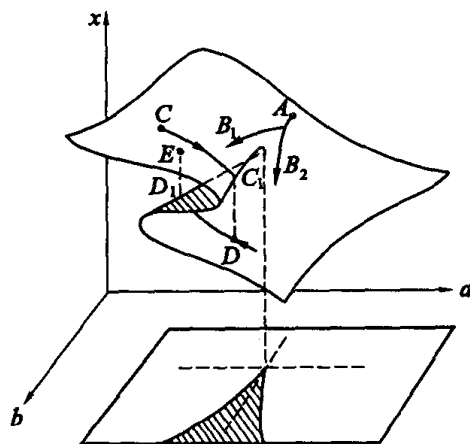


图 2-9 尖角突变模型

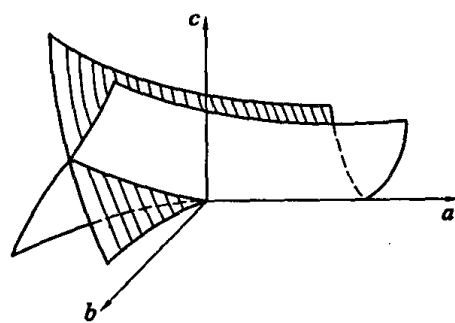


图 2-10 燕尾型突变模型

2.5 混沌理论

混沌现象广泛存在于自然界和社会经济系统中, 人们很早就注意到混沌现象, 许多领域的专家学者从不同的角度研究过混沌现象。但直到 20 世纪 70 年代中期, 混沌理论才初步建立。在科学史上, 决定论与非决定论进行着长期的争论, 混沌则把决定性系统内在随机性表象联系在一起, 使人们在那些以前令人迷惑的甚至望而生畏的混沌现象中, 发现了许多出乎意料的规律性, 为系统科学特别是系统演化提供了新的思想和方法, 有着重要的理论意义和实用价值。

混沌(chaos)是由确定性的发展过程中产生出来的一种随机运动。它不是简单的无序状态, 在“杂乱无章”运动中又包含普适常数, 包含自相似性。由确定性的 Logistic 方程可以看到如何产生混沌现象的, 它是研究生物种群发展中所抽象出来的数学模型, 其方程为

$$x_{n+1} = \lambda x_n (1 - x_n) \quad (2.34)$$

式中, x 是状态变量, 取值范围 $[0, 1]$; λ 是参数, 取值范围 $[0, 4]$; n 是迭代步数。

给定一个状态变量的任意初始值, 当 λ 在 $[0, 3]$ 范围内变化时, 系统最终都会收敛于一个稳定解。如 $\lambda = 2, x_0 = 0.5$ 时, 该系统的解为 $x_1 = x_2 = x_3 = \dots = x_n = x_{n+1} = 0.5$; 当 $\lambda = 2, x_0 = 0.1$ 时, 迭代得, $x_1 = 0.1800, x_2 = 0.2952, x_3 = 0.3654, x_4 = 0.4638, x_5 =$

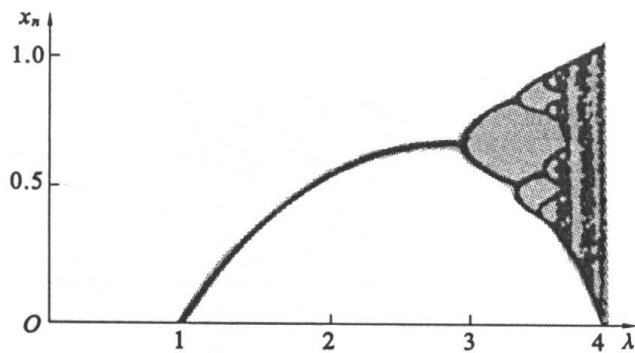


图 2-11 Logistic 混沌

0.4974, $x_6 = 0.5000$, 也收敛于 0.5。

当 $\lambda > 3$ 时, 系统的单一解开始失稳, 出现分叉。在 $3 < \lambda < 3.44949\cdots$ 时, 系统有 2 个稳态解; 在 $3.44949\cdots < \lambda < 3.56637\cdots$ 时, 系统出现 4 个稳态解; 在 $3.56637\cdots < \lambda < 3.57078\cdots$ 时, 系统有 8 个稳态解……; 如此, 随着参数 λ 的增长, 系统以倍频方式很快就进入混沌解状态, 即有无穷多个解。这个过程如图 2-11 所示。由放大的图形可以看到, 在混沌区存在各种周期解的“窗口”和各种宏观的周期性“花纹”。

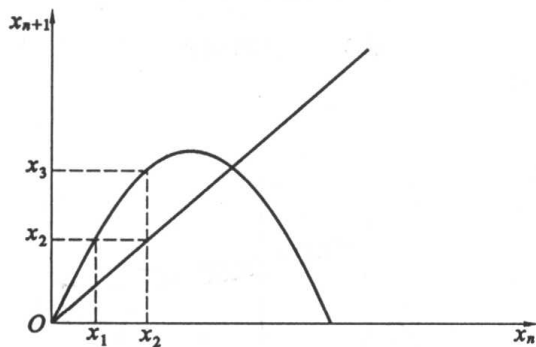


图 2-12 混沌迭代分析方法

像 Logistic 方程这样通过周期分叉进入混沌的现象也可以用几何构图来说明。如图 2-12 所示, 横坐标为 x_n , 纵坐标为 x_{n+1} , 在坐标原点画一根 45° 的斜线。取初值 x_1 , 计算得 x_2 , 标出一点; 由点 (x_1, x_2) 作 x_n 轴的平行线交于 45° 的斜线, 得到 x_n 轴上的 x_2 点; 由 x_2 点, 迭代计算得 x_3 点; …… 如此即可画出几何图形。

像 Logistic 方程这样通过周期分叉进入混沌的现象也可以用几何构图来说明。如图 2-12 所示, 横坐标为 x_n , 纵坐标为 x_{n+1} , 在坐标原点画一根 45° 的斜线。取初值 x_1 , 计算得 x_2 , 标出一点; 由点 (x_1, x_2) 作 x_n 轴的平行线交于 45° 的斜线, 得到 x_n 轴上的 x_2 点; 由 x_2 点, 迭代计算得 x_3 点; …… 如此即可画出几何图形。

对于方程 $x_{n+1} = \lambda x_n(1 - x_n)$,

取 $\lambda = 0.5$ 时, x_n 收敛于 0, 如图 2-13(a);

$\lambda = 1$ 时, x_n 收敛于 0, 如图 2-13(b);

$\lambda = 2$ 时, x_n 收敛于 0.5, 如图 2-13(c);

$\lambda = 3$ 时, x_n 收敛于 $2/3$, 如图 2-13(d);

$\lambda = 3.15$ 时, x_n 约收敛于 0.533 和 0.784, 如图 2-13(e);

$\lambda = 3.53$ 时, x_n 约收敛于 0.368、0.517、0.821 和 0.881, 如图 2-13(f)。

在应用科学领域, 最早从确定性方程发现混沌现象的是美国麻省理工学院的气象学家劳伦兹(E. N. Lorenz)。他在 1963 年偶然发现: 在一特定的方程组中, 极小的误差可能导致灾难性的后果, 其原因在于“蝴蝶效应”。即有些系统极端依赖于初始条件的敏感性。

劳伦兹研究大气环流时, 得到一个模型, 称为 Lorenz 方程, 其表达式为

$$\begin{cases} \dot{x} = \sigma(y - x) \\ \dot{y} = \nu x - y - xz \\ \dot{z} = -\beta z + xy \end{cases} \quad (2.35)$$

式中, x, y, z 是状态变量, σ, β, ν 是参数。该模型在一定参数条件($\sigma = 10, \beta = 28, \nu = 8/3$) 时可以得到一个最典型的混沌吸引子。当用相空间描述混沌运动时, 称其为奇怪吸引子或混沌吸引子。图 2-14 是一个典型的劳伦兹吸引子图, 它像一只展开两个翅膀的蝴蝶或猫头鹰脸, 这个神奇的图形是人们发现来自实践的第一个混沌吸引子, 也是早期混沌学说的标志符号。

在劳伦兹图形中, 轨迹永远不会重复, 是一种混沌现象。如在天气预报上, 天气无法精确地重复自身, 三天以上的预报只能是概率性的。

混沌现象也并不是杂乱无章的, 而应看作一种新的结构和秩序。著名的费根鲍姆

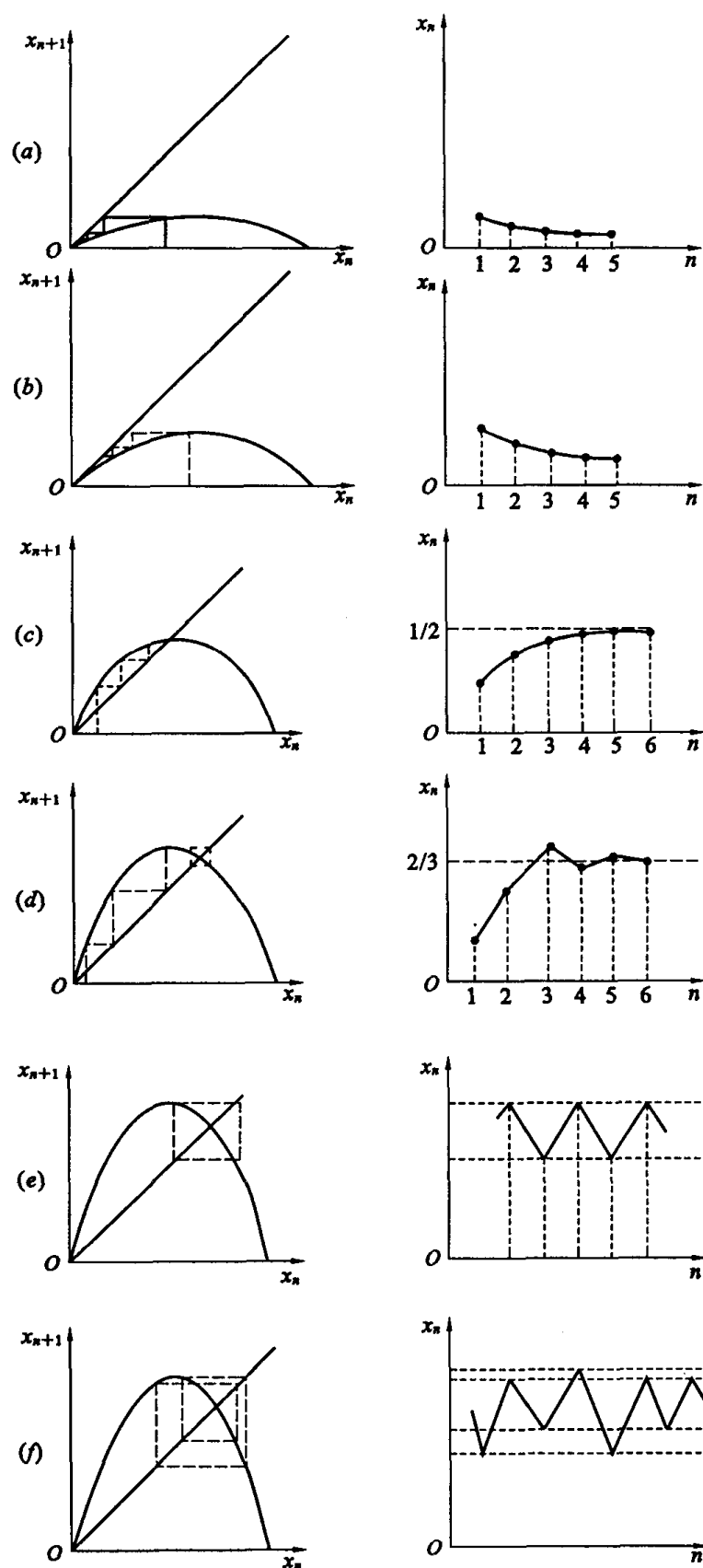


图 2-13 Logistic 方程的迭代分析

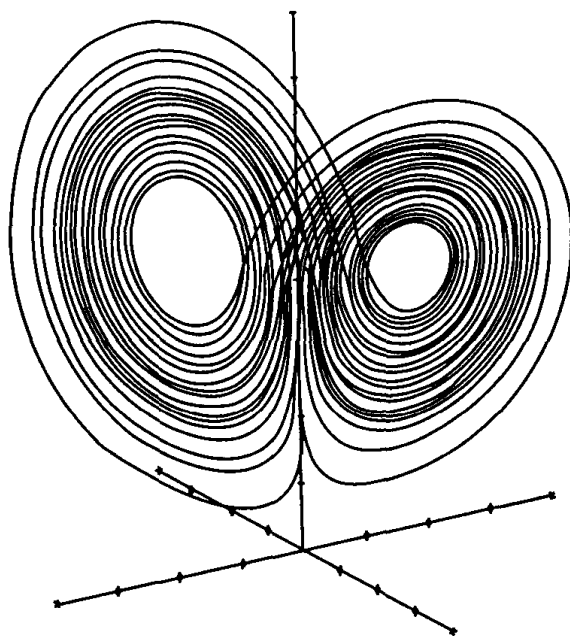


图 2-14 劳伦兹吸引子图

(Feigenbaum)常数描述了一类混沌现象中的普适性。如对于 Logistic 方程,其倍频分叉值 $\lambda_k (k = 1, 2, \dots)$ 如表 2-3 所示,定义下式为费根鲍姆常数 δ 。

$$\delta = \lim_{k \rightarrow \infty} \frac{\lambda_k - \lambda_{k-1}}{\lambda_{k+1} - \lambda_k} = 4.6692016609 \dots \quad (2.36)$$

表 2-3 倍频分叉值

k	分 叉 情 况	λ_k
1	$1 \rightarrow 2$	3
2	$2 \rightarrow 4$	3.449489743...
3	$4 \rightarrow 8$	3.544090359...
4	$8 \rightarrow 16$	3.56407266...
5	$16 \rightarrow 32$	3.56875942...
6	$32 \rightarrow 64$	3.56969161...
7	$64 \rightarrow 128$	3.571935...
$\vdots$	$\vdots$	$\vdots$

通过对混沌现象的研究,深化了人们对系统及其演化方式的认识,有序与无序不是绝对的,是可以互补的,混沌也是一种“正常”状态。哈肯认为,任何系统,当外界的控制不断改变时,大多会经历从无序到有序,从有序到有序,再从有序到混沌的过程。即出现无序(混沌)到有序再到无序的循环。

2.6 耗散结构理论

2.6.1 耗散结构的概念

比利时物理学家普里高津(I. Prigogine)在 1969 年提出耗散结构理论,他认为一个孤立

系统总是朝着均匀和无序的平衡态发展,系统的熵不断增大,直至达到平衡态,此时系统的熵最大。但对于一个开放系统,系统的熵却可能增长、维持或减小。此时,系统的熵增(熵的变化量)为

$$dS = d_i S + d_e S \quad (2.37)$$

式中 dS ——系统总的熵增;

$d_i S$ ——系统内部的熵增;

$d_e S$ ——系统与外部环境的交换熵增。

其相互关系图如图 2-15 所示。如果 dS 大于零,则系统也朝着无序方向发展;如果 dS 小于零,则系统朝着有序化程度提高的方向发展,这时,必有 $d_e S < 0$,且 $|d_e S| > d_i S$,因为 $d_i S$ 对于不可逆过程总是大于零的。

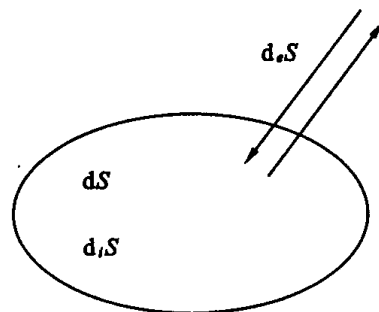


图 2-15 开放系统的熵增

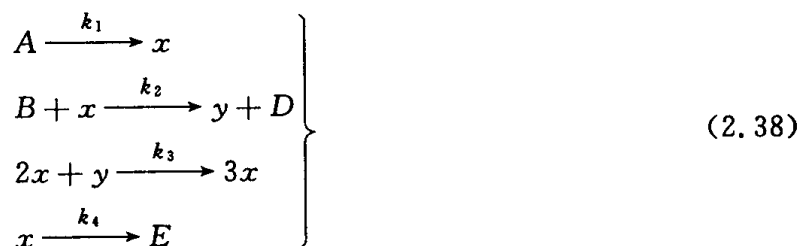
对于非平衡的开放系统,由于存在外力,它会朝着非平衡定态发展。在近平衡区,不可逆力与不可逆流可近似为线性关系,线性关系成立的区域是非平衡线性区。如图 2-4 所示的金属棒传热系统,温度梯度是不可逆力,热量传递是不可逆流。达到非平衡定态时,金属棒内维持不变的温度梯度。此时,如果改变外部条件,使 T_1 连续不断地接近 T_2 ,直到 $T_1 = T_2$,消除非平衡约束。这时,系统的非平衡定态会连续地向平衡态发展,最终达到平衡态。即系统在近平衡区不会发展出新的有序。

当开放的非线性系统远离平衡态时,不可逆力与不可逆流之间的线性关系不再成立。由于 $dS < 0$,即系统与外界交换能量与物质,吸进负熵。当外界条件达到一定阈值时,系统有可能从原来的混乱无序发展为一种在时间上、空间上或功能上的新有序状态或结构,称其为耗散结构。

2.6.2 布鲁塞尔模型

著名的布鲁塞尔(Brussels)模型或称三分子模型是普里高津领导的布鲁塞尔学派构造的一种理论模型。该模型能模拟广泛的宏观自组织行为,给出对应于各种类型耗散结构的解,在非平衡相变研究中得到广泛的应用,它使我们能够更好地理解出现极限环、非均匀定态、空间化学波等现象的条件,对理解和研究各种各样的时空有序结构有重要理论意义。

但布鲁塞尔模型并非实际存在的一种化学反应模型,它是以斯克洛格(Schlögl)模型和洛卡特-伏洛特拉(Lotka-Volterra)模型为原型而构造的。其方程为



式中 A, B, D, E, x, y ——假设的反应物、中间物或生成物;

k_1, k_2, k_3, k_4 ——分别是各方程的反应速率。

其状态变量的微分方程是

$$\left. \begin{aligned} \frac{dx}{dt} &= k_1 A - k_2 Bx + k_3 x^2 y - k_4 x \\ \frac{dy}{dt} &= k_2 Bx - k_3 x^2 y \end{aligned} \right\} \quad (2.39)$$

为便于求解, 设 $k_1 = k_2 = k_3 = k_4 = 1$, 则(2.39)式的定态解为

$$\left. \begin{aligned} x_0 &= A \\ y_0 &= \frac{B}{A} \end{aligned} \right\}$$

在定态点 (x_0, y_0) 附近作线性稳定性分析, 令

$$\left. \begin{aligned} x &= x_0 + \Delta x \\ y &= y_0 + \Delta y \end{aligned} \right\} \quad (2.40)$$

代入方程(2.39), 得

$$\left. \begin{aligned} \frac{d\Delta x}{dt} &= (B-1)\Delta x + A^2 \Delta y \\ \frac{d\Delta y}{dt} &= -B\Delta x - A^2 \Delta y \end{aligned} \right\}$$

其特征方程为

$$\lambda^2 - \omega \lambda + T = 0 \quad (2.41)$$

式中 $\omega = B-1-A^2$; $T = A^2$ 。

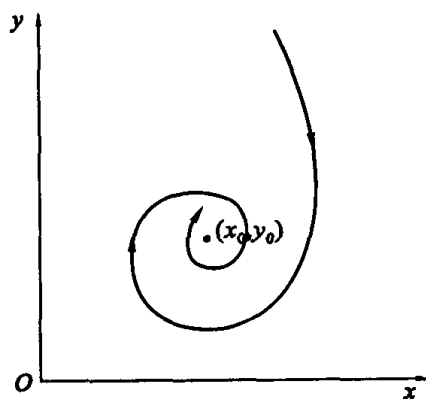


图 2-16 $\omega < 0$ 时的情形

由方程(2.41)看到, 控制参数 A, B 决定着系统的稳定性。

当 $\omega < 0$, 即 $B < 1 + A^2$ 时, 特征方程有两个负实部的共轭复根。这时, 随着时间 t 的增长, $\Delta x, \Delta y$ 均趋于零, 于是定态点 (x_0, y_0) 成为吸引中心, 系统是稳定的。或者说, 任何从偏离的初态开始的运动轨迹, 最终都会回到原始的均匀态。其运动轨迹相平面图表示为图 2-16。

当 $\omega > 0$, 即 $B > 1 + A^2$ 时, 特征方程的根为两个具有正实部的共轭复根。此时方程的解将不稳定, 随着时间 t 的增长, $\Delta x, \Delta y$ 也趋于增长, 但由于非线性项 $x^2 y$ 的存在, $\Delta x, \Delta y$ 最终被限定在一个范围内, 其相平面上出现了稳定的极限环, 如图 2-17 所示。这时相当于产生了化学振荡, 在环上, 浓度 x 和 y 随 t 而产生周期性变化。在布鲁塞尔模型方程(2.38)中, 如再考虑浓度差及扩散项, 在一定条件下, 系统会出现非均匀定态, 也可能出现空间分布的化学波现象, 这里不再详述。

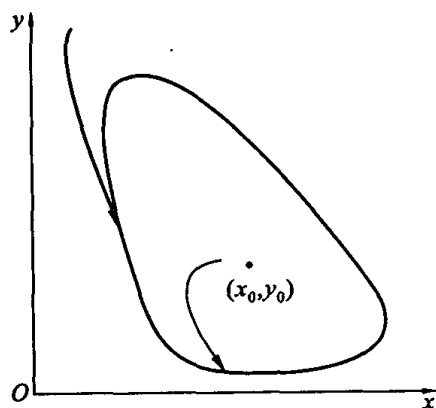


图 2-17 $\omega > 0$ 时的情形

2.6.3 形成耗散结构的条件

首先, 系统必须是远离平衡态的开放系统, 不断与外界大量交换能量与物质, 来维持系统形成新的有序结构。

其次, 系统必须进入远离平衡态的非线性区域。

最后,还要有涨落的触发。涨落是微小的波动或干扰。在线性区,涨落被消耗掉,几乎没有什么作用,而在远离平衡的非线性区临界点附近,微小的随机小扰动会得到“放大”,成为一个“巨涨落”。触发系统跃迁到一个新的稳定的有序状态从而形成耗散结构。

2.7 协同学原理

与耗散结构理论一样,协同学也是研究一个系统如何自发地形成有序结构的。德国物理学家哈肯(H. Haken)在20世纪60年代研究激光理论的基础上,于1969年提出了协同学的微观理论,1977年发表了专著《协同学导论》,1983年发表了《高等协同学》,形成了协同学这一系统理论新的分支。

协同学(synergetics)的通俗词义就是“协同工作”的学问。它研究系统的各个部分如何进行协作,并通过协作导致系统出现空间上、时间上或功能上的有序结构。因此,协同学也是系统的自组织理论。

2.7.1 协同学基本观

协同学认为协同导致有序,但协同是如何产生的,在影响系统行为的若干个变量中,大多数变量在系统受到干扰而波动时,总是力图使系统回到原先的稳定状态,它们起一种阻尼作用,且衰减很快,对系统发生结构性转变的过程没有大的影响,称这类变量为快弛豫变量或快变量。而另外一些为数极少(往往只有一个或几个)的变量,则在系统受干扰而波动时,使系统偏离稳态,走向非稳态,或进入新的稳态,它们始终左右着演化的进程,决定着演化结果的结构与功能,称其为慢弛豫变量或慢变量。哈肯定义慢变量为表示系统有序程度的序参量。如果系统处于完全无序的混沌状态时,其序参量为零;在接近临界区时,序参量迅速增大;进入临界区,序参量达最大值。它主导着系统出现新的有序结构。

哈肯认为,快变量服从慢变量,序参量支配着子系统的行为。这个观点也称为“伺服原理”或“支配原理”。当系统处于阈值时,系统的有序结构形成的速度很快,外界对系统的影响很小可予以忽略。在系统内部,可以用慢变量表示所有的快变量,消去大量的自由度,方程就只包含一个或几个参量的序参量方程,使方程易于求解。

多个序参量之间的关系是既合作又竞争。在不同的外部条件下,竞争与协作的结果是,形成一种有序结构。如贝纳德花纹,在一定的温差时出现六角形蜂窝状;当温差继续增大至一个新的阈值时,出现了新的有序结构,呈现滚动的卷筒状结构。

激光器如果被较弱的外泵激发,就相当于一个普通的灯,各原子相互独立地发射相位无规的杂乱的光波。当泵激作用增强到一定程度时,它能发出单一波长的光波,形成激光。激光在发射之初,也存在多种不同的光波,且相互竞争。光电子本身在加强不同光波时其作用是不同的,对于某种特定的光波,虽然它们往往只是领先一点点,但由于得到雪崩似的增强,很快战胜其他光波并接受它们的能量,迫使所有光电子以共同的节奏振荡。这种新产生的波是决定激光秩序的序参量。激光的产生是一种典型的协同行为。

涨落是系统形成有序结构的内部决定性因素。与耗散系统一样,通常系统的随机涨落对系统的演化基本上不起作用,但在临界区涨落得到放大,触发系统形成新的有序结构,进入新的稳定平衡状态。

2.7.2 协同学原理的数学描述

设所研究系统可用具有 m 个分量的状态向量来描述,即

$$q = (q_1, q_2, \dots, q_m) \quad (2.42)$$

各分量 q_i 既依赖于时间 t , 也依赖于空间 x , 即

$$q_i = q_i(x, t) \quad (2.43)$$

则系统的基本演化方程为

$$\dot{q}(x, t) = \frac{dq}{dt} = N[q(x, t), \nabla, \alpha, x] + F(t) \quad (2.44)$$

式中, ∇ 是微分算子,

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) \quad (2.45)$$

α 是控制参数, 如贝纳德花纹受外界温差的控制等。

$F(t)$ 是系统内部的涨落力, 在许多情况下可忽略, 但有时又起到决定性的作用。

N 是非线性函数向量, 它依赖于(2.44)式中左边出现在同一个或是不同空间上各点的状态向量 q 。

通常, 一般非线性方程(2.44)是无法求解的。通过一系列的假设与简化, 可以进行较为深入的研究。实际上, 在哈肯的《协同学——自然成功的奥秘》一书中就几乎没有使用任何数学表达式。关于序参量及支配原理的推导请见《高等协同学》等参考文献。

2.7.3 协同学的应用

在系统学的诸个分支中, 协同学的观点得到了更为广泛的应用。哈肯本人除了对物理学有深入的研究之外, 还广泛涉猎化学、生物学、经济学和社会学等学科领域, 通过探索多学科的共同点, 提炼出协同学; 同时, 也应用协同学的思想观点分析多个不同领域的问题, 得到许多有益的结论。

激光是哈肯研究得最多的领域之一; 化学钟和贝纳德花纹也是经常提到的经典案例; 这些案例中存在的自组织现象, 或者说多个序参量之间的竞争与合作, 前已述及, 此略。

在社会、经济、文化等系统结构的演化中, 有些现象也可以用协同学的观点来分析。例如, 城市的形成与变迁, 多种因素(如交通、水、气候、食物和安全等)的综合作用导致城市的逐步形成。应该说, 在若干个系统变量中, 起决定性作用的序参量并不多。如果能够分析得较为清楚, 把握准机会, 适度地加以调控, 可以收到好的效果。但是, 在不同的时期及不同的环境中, 决定系统演化的序参量是不同的, 要在实践中做好是非常困难的。

市场的形成与演化也是可以用协同学观点来分析的。按照简单的思维分析, 各种商店及服务企业应该在整个城市中均匀分布, 但实际上, 有很多类型的商店及服务企业却与此相反, 形成了中心商业区和各种各样的特色街: 如美食城或美食一条街、电脑城或电脑一条街等等; 而且, 在许多中心商业区特大型的店家往往不止一个, 这些都是竞争的结果; 同时, 要维持总体的利益与形象, 各店家特别是领头的几个又要保持一定的竞争中的协同。形成这种局面的一种解释是, 客户希望有快捷方便且多种不同的选择以便较快地做出抉择。

产品的区域市场、国内市场和国际市场的分工与分布也都是多种因素竞争与协作的结

果。例如,计算机硬件的高端产品,生产越来越集中于美国等国家的极少数掌握尖端技术的企业,如 Intel 公司等;而计算机硬件的低端产品则正向中国等发展中国家与地区转移。有些表面上很热闹的东西可能是一些快变量,不会持久发挥作用,而有些持续发挥作用的慢变量,即序参量则终将使市场的结构发生实质性的变化。

在上述系统的多参数竞争与合作中,涨落的作用也是显见的。最常见的现象是,某种结果呼之欲出或者很多人都认为应该出现了,但却等待着某个事件甚至是很小的一个事件的触发,才真正出现。就像在物理学中所看到的,冰点以下的水按理早该结冰了,但它却处于一种亚稳态,需要通过一个自发的涨落或外来的一次非常微弱的推动,水才突然结成冰。

习 题

- 2.1 什么是孤立系统、封闭系统和开放系统? 试分别举例加以说明。
- 2.2 什么是系统自组织现象? 试描述一个具体的系统自组织现象。
- 2.3 中国科学家对系统科学与技术有过哪些贡献?
- 2.4 如何全面正确理解系统的整体性和“ $1+1>2$ ”表达式?
- 2.5 耗散结构理论、协同学和混沌理论的主要观点是什么? 有什么共同点与不同点?

3 数学规划基础

本章提要

作为系统科学中技术基础理论之一的运筹学,是从系统总体角度寻求系统最优解的数学工具,它包括数学规划、图与网络、决策分析、排队论、库存论、对策论等。在本书中仅介绍经常应用的数学规划(第3章)、图与网络(第4章)和决策分析(第6章)的基本概念和算法。

在本章介绍数学规划的过程中,首先讨论线性规划,它是把系统简化为线性优化问题的求解方法,在此基础上,介绍了目标规划、整数规划和0-1规划,然后介绍了系统为非线性优化问题的非线性规划。最后介绍处理多阶段优化问题的动态规划。

3.1 线性规划

线性规划是运筹学的一个重要分支。随着科学技术的发展,它在理论和方法上日趋成熟,应用极为广泛。它是研究一定数量的资源条件下,如何科学的运用这些资源,以获得最大效益,或者在一定技术条件下,寻求最优化的设计。这种数学规划的方法,如果用数学语言表达出来,就是在一定约束条件下,寻找目标函数的极值问题。所谓线性规划,是指约束条件为线性等式或线性不等式,且目标函数也为线性函数。

3.1.1 基本概念

线性规划可以用以下数学形式表达,即求一组变量 $x_1, x_2, \dots, x_n$, 在满足约束条件:

$$\left. \begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &\leq b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &\leq b_2 \\ \dots & \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &\leq b_m \\ x_1, x_2, \dots, x_n &\geq 0 \end{aligned} \right\} \quad (3.1)$$

的情况下,使目标函数

$$f = c_1x_1 + c_2x_2 + \dots + c_nx_n \quad (3.2)$$

达到最大值(或最小值)。

其中: a_{ij}, b_i, c_j 均为已知, $i = 1, 2, \dots, m; j = 1, 2, \dots, n$ 。

以上表达式还可简写成:

求 $x_j (j = 1, 2, \dots, n)$, 使之满足约束条件

$$\sum_{j=1}^n a_{ij}x_j \leq b_i (i = 1, 2, \dots, m)$$

且

$$x_j \geq 0, (j = 1, 2, \dots, n)$$

使得目标函数

$$f = \sum_{j=1}^n c_j x_j$$

达到最大值(或最小值)。

满足约束条件式(3.1)的所有解,称为可行解。在可行解中,其中一个可行解能使目标函数达到极值,则该可行解称为最优解,也可称为线性规划的解。线性规划在经济建设、企业管理、技术设计和生产实践方面都有成功的应用。

【例 3.1】 (生产管理问题)某工厂生产 A、B 两种产品,其产量分别为 x_1 、 x_2 个单位,每天的可用资源限制:原料:1575kg;占地面积:1500m²;工时:7 小时(420 分钟)。

资源分配及产值如表 3-1 所示:

表 3-1 资源分配及产值表

名 称	符 号 及 单 位	A	B	
资源 消耗 系数	a_1 m ² /单位产量	4	5	占地面积
	a_2 kg/单位产量	5	3	原料消耗
	a_3 单位/小时	60	30	生产率
产值系数	c 元/单位产量	13	11	产 量

问:每天应生产多少 x_1 、 x_2 ,使 $f(x) = c_1 x_1 + c_2 x_2$ 为最大。

其数学模型,可表示为:

$$\max f(x) = 13x_1 + 11x_2 \quad (3.3)$$

$$\text{s. t.} \quad \left. \begin{aligned} 4x_1 + 5x_2 &\leq 1500 \\ 5x_1 + 3x_2 &\leq 1575 \\ x_1 + 2x_2 &\leq 420 \left(\frac{x_1}{60} + \frac{x_2}{30} \leq 7 \right) \end{aligned} \right\} \quad (3.4)$$

$$x_1, x_2 \geq 0 \quad (3.5)$$

式(3.3)表示工厂所要求的最大利润,即求目标函数的最大值。而式(3.4)、(3.5)表示不等式约束。

【例 3.2】 (运输问题)从两个仓库(发点)运送库存原棉到三个纺织厂(收点),两仓库的库存量、三个纺织厂的需求量,以及每吨原棉从各仓库运到工厂的所需运费如表 3-2 所示。

表 3-2 货源需求与运价

运输单价(元/吨) 仓库 i \ 工厂 j	1#	2#	3#	库存量
1#	2	1	3	50
2#	2	2	4	30
需求量(吨)	40	15	25	

问:在保证各个纺织厂的需求都得到满足的条件下,应采用哪一种运送原棉的方案使运费最少。

这里用 x_{ij} 表示 i 仓库运送到 j 厂的原棉数量,于是总的运费 $f(x)$ 为:

$$f(x) = 2x_{11} + x_{12} + 3x_{13} + 2x_{21} + 2x_{22} + 4x_{23}$$

其目标定在上述条件下使 f 达到最小值。由于从各仓库运出的原棉数量不可能超过它的库存量,所以应有:

$$x_{11} + x_{12} + x_{13} \leq 50$$

$$x_{21} + x_{22} + x_{23} \leq 30$$

同时,还必须保证各纺织厂对原棉的需求量,所以应有:

$$x_{11} + x_{21} = 40$$

$$x_{12} + x_{22} = 15$$

$$x_{13} + x_{23} = 25$$

此外,运送量 x_{ij} 都取非负值:

$$x_{ij} \geq 0 \quad (i = 1, 2, \quad j = 1, 2, 3)$$

于是,该问题可以表达成以下的形式:

$$\min f(x) = 2x_{11} + x_{12} + 3x_{13} + 2x_{21} + 2x_{22} + 4x_{23} \quad (3.6)$$

$$\text{s. t.} \quad \left. \begin{array}{l} x_{11} + x_{12} + x_{13} \leq 50 \\ x_{21} + x_{22} + x_{23} \leq 30 \\ x_{11} + x_{21} = 40 \\ x_{12} + x_{22} = 15 \\ x_{13} + x_{23} = 25 \end{array} \right\} \quad (3.7)$$

$$x_{ij} \geq 0 \quad (3.8)$$

$$(i = 1, 2, \quad j = 1, 2, 3)$$

这类运输问题可以作如下一般叙述:要将 m 个发点的货物运送到 n 个收点去,已知 i 发点有货物 a_i 吨, j 收点需货物 b_j 吨,每单位货物从发点 i 运送到 j 收点的运费为 c_{ij} 元。那么,在保证每个收点对货物需求的情况下,应采用哪种运输方案才能使总运费最省?

设 x_{ij} 为从 i 发点运到 j 收点的货物量,则上述问题可作如下描述:

$$\min f = \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} \quad (3.9)$$

$$\text{s. t.} \quad \left. \begin{array}{l} \sum_{j=1}^n x_{ij} \leq a_i, i = 1, 2, \dots, m \\ \sum_{i=1}^m x_{ij} \leq b_j, j = 1, 2, \dots, n \end{array} \right\} \quad (3.10)$$

$$x_{ij} \geq 0 \quad (i = 1, 2, 3, \dots, m, \quad j = 1, 2, \dots, n) \quad (3.11)$$

在以上例题中,都首先需要确定一组具有明确含义的变量 x ,称之为决策变量,问题的目标是选取这些决策变量的值,使某一个函数取得最大值或最小值,此函数称之为目标函数。同时又利用决策变量把问题的条件表示成等式或不等式,并称这些等式或不等式为约束条件。如果目标函数是决策变量的线性函数,约束条件也是关于决策变量的线性等式或线性不等式,则相应的问题即称为线性规划问题。

显然,应用线性规划在解决国民经济建设中的投资,运输,调度,分配和工作技术设计等优化问题时,必须符合以下特征:

- (1) 所求问题的目标函数可以用最大值或最小值的形式来表达;
- (2) 为寻求达到最优目标,必须有多种方案可供选择;
- (3) 寻求的目标必然有条件约束;
- (4) 目标函数,约束条件都是线性方程式且决策变量具有非负性。

3.1.2 几何图形特征与图解法

为了对求解线性规划问题有一个直观的几何概念,首先应用图解法求解线性规划问题。

【例 3.3】 某工厂制造两种产品 P_1, P_2 ,需用三种原材料 M_1, M_2, M_3 ,制造 1kg 产品 P_1 需用原材料 M_1 9kg, M_2 4kg, M_3 3kg,制造 1kg 产品 P_2 需用 M_1 5kg, M_2 4kg, M_3 10kg; 产品 P_1 每 kg 有利润 700 元,产品 P_2 每 kg 有利润 1200 元;工厂每天能供应的原材料 M_1 360kg, M_2 200kg, M_3 300kg。问在上述条件下,工厂要使利润最高,产品 P_1, P_2 的生产量分别为多少?

【解】 问题所给出的数据列入下表(见表 3-3),首先建立数学模型。

表 3-3

需 要 量 原 材 料	产 品 P_1 (kg)	P_2 (kg)	原材料日使用限额(kg)
M_1	9	4	360
M_2	4	5	200
M_3	3	10	300
利润(百元)	7	12	

设产品 P_1, P_2 的日产量分别为 x_1, x_2 ,则约束条件可表示成:

$$\left. \begin{aligned} 9x_1 + 4x_2 &\leq 360 \\ 4x_1 + 5x_2 &\leq 200 \\ 3x_1 + 10x_2 &\leq 300 \end{aligned} \right\} \quad (3.12)$$

$$x_1 \geq 0, x_2 \geq 0 \quad (3.13)$$

总利润为

$$f = 7x_1 + 12x_2 \quad (3.14)$$

首先将式(3.12)看作等式,并作出图形如图 3-1 所示,直线 FD, GH 和 AE (即图中 1, 2 和 3 直线)满足等式(3.12)中各个不等式, y 点(x_1, x_2)应分别在以 FD, GH 和 AE 为界的平面内,而条件(3.13)式要求点(x_1, x_2)在第一象限内,所以满足全部约束条件式(3.12), (3.13)的点的集合应当是五边形 $OABCD$ 周界上和内部的所有点。

另外,根据式(3.14)的斜率可以获得一组平行等值直线组。当直线向右上方平行移动, f 值随之增大,这里必须找出一条既与凸多边形 $OABCD$ 相交(这意味着在规定的约束条件范围内),而又尽可能赋有较大 $f(x)$ 值的等值线(即目标函数的最大值)。显然与 B 点相切

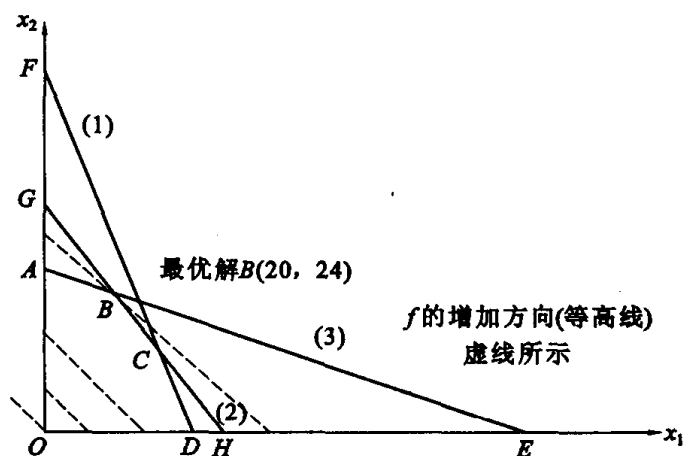


图 3-1 线性规划的几何图形及可行解域

的等值线是符合该要求的,即在 B 点的坐标是在满足约束条件的前提下,使目标函数获得最大值,于是,取得的最优解为:即 P_1 的日产量 20kg, P_2 的日产量 24kg,此时所得的最大利润为:

$$\max f(x_1, x_2) = f(20, 24) = 7 \times 20 + 12 \times 24 = 42800 \text{ 元}$$

从图 3-1 可以看到可行解的集合五边形 $OABCD$ 是凸的,而最优解 B 点是这个凸五边形的一个顶点,可以证明:线性规划问题的所有可行解构成一个凸多边形,而最优解如果存在,则一定可以在凸多边形的顶点找到。

【例 3.4】 已知线性规划的数学模型

约束条件

$$\left. \begin{aligned} -x_1 + x_2 &\geq 1 \cdots \cdots (1) \\ x_1 + x_2 &\leq -2 \cdots \cdots (2) \end{aligned} \right\} \quad (3.15)$$

$$\left. \begin{aligned} x_1 &\geq 0 \cdots \cdots (3) \\ x_2 &\geq 0 \cdots \cdots (4) \end{aligned} \right\} \quad (3.16)$$

求 x_1, x_2 使目标函数为 $f(x) = 2x_1 + 2x_2$ 的值为最小。

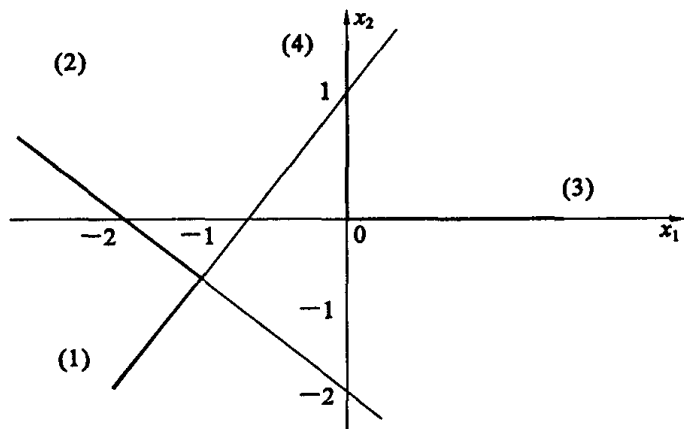


图 3-2 无可行解域

【解】 根据约束条件式(3.15)、(3.16)作几何图形 3-2,从图中可以看出线性规划模型不存在可行域,所以也没有可行解,当然也必然不存在最优解。

从上面例子中可以看到,线性规划问题可以有唯一解,对于有解的情况,可以发现两个

重要特点:

(1) 线性问题的任意两个可行解连线上的点都是可行解。

(2) 线性规划问题的最优解如果存在,必然可以在可行域的某个“顶点”上达到,各顶点的解称为基本可行解。

3.1.3 单纯形法

(1) 基本原理

在线性规划中,图解法只能求解两个变量的问题,求解过程较麻烦,且精确性较差,而单纯形法可以借助计算机,计算成千上百个变量和约束条件的线性规划问题。单纯形法的基本原理与图解法相似,即线性规划问题若存在最优解,则一定可以在可行解域(即凸域)的“顶点”上找到。同时,可行解域不论有多少个顶点只要通过有限次数的迭代计算、比较就一定可以找到最优解,下面通过实例说明单纯形法的原理和求解步骤。

如例 3.1 的生产管理问题。

$$\max f(x) = 13x_1 + 11x_2 \quad (3.17)$$

$$\text{s. t.} \quad \left. \begin{aligned} 4x_1 + 5x_2 &\leq 1500 \cdots \cdots (1) \\ 5x_1 + 3x_2 &\leq 1575 \cdots \cdots (2) \\ x_1 + 2x_2 &\leq 420 \cdots \cdots (3) \end{aligned} \right\} \quad (3.18)$$

$$x_1 \geq 0, x_2 \geq 0 \quad (3.19)$$

第一,引入松弛变量。单纯形法首先要把不等式约束变成等式约束,以形成凸域上的顶点。其办法是增加非负变量 x_3, x_4, x_5 成为补充变量,(一般,当约束条件不等式为“小于”时,增加的补充变量称为松弛变量,当不等式为“大于”时,则补充变量称为剩余变量)同时令式(3.17)的目标函数方程右端为 0,即:

$$\left. \begin{aligned} f(x) - 13x_1 - 11x_2 &= 0 \cdots \cdots (1) \\ 4x_1 + 5x_2 + x_3 + 0 + 0 &= 1500 \cdots \cdots (2) \\ 5x_1 + 3x_2 + 0 + x_4 + 0 &= 1575 \cdots \cdots (3) \\ x_1 + 2x_2 + 0 + 0 + x_5 &= 420 \cdots \cdots (4) \end{aligned} \right\} \quad (3.20)$$

在式(3.20)中的(2)、(3)、(4)式为等式约束,可写成

$$AX = B \quad (3.21)$$

其中

$$A = \begin{bmatrix} 4 & 5 & 1 & 0 & 0 \\ 5 & 3 & 0 & 1 & 0 \\ 1 & 2 & 0 & 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 1500 \\ 1575 \\ 420 \end{bmatrix}$$

这样,问题就变为求非负变量 $x_1, x_2, \dots, x_5$ 的数值,在满足以上等式约束的条件下,使 $f(x)$ 最大。

第二,确定可行域顶点。根据式(3.20)可作图,如图 3-3,此时选原点为初始基础可行解, $x_1 = 0, x_2 = 0$, 则初始解向量 $X^{(0)}$ 为

$$X^{(0)} = \underbrace{[0 \ 0]}_{\text{非基础变量}} \underbrace{[1500 \ 1575 \ 420]}_{\text{基础变量}}^T$$

在初始解中,三个非零变量 x_3, x_4, x_5 作为基础变量。基础变量的个数正好等于约束式数,这三个变量的系数在矩阵 A 中相应的列向量线性无关。零向量 x_1, x_2 称为非基础变

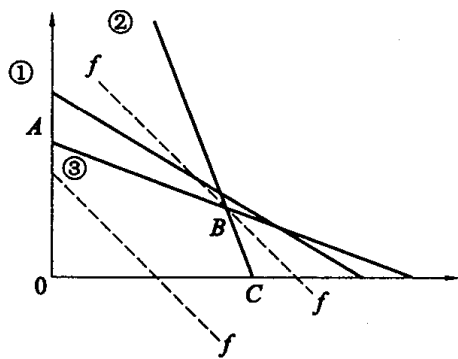


图 3-3

量,这时目标函数根据式(3.20)中的(1)式为: $f[x^{[0]}] = 0$ 。

第三,第一次迭代选取临近顶点,如能使目标函数值增大,则为新的基础可行解,否则原来的解就是最优解。

在更换顶点时,显然要对 x 换元,将原来是零的某元素与非零的某元素互换,即原基础可行解中一个基础变量要换成非基础变量,此变量称为“退出变量”,而另一个非基础变量要换成基础变量称为“引入变量”。

① 选择引入变量 x_k ,其原则是应使目标函数值增加最快。分析目标函数 $f(x) = 13x_1 + 11x_2$,显然 x_1 的系数比 x_2 的系数大,要使 $f(x)$ 增加快,应选 x_1 为引入变量,即使 x_1 由零变量为非零。具体做法是在式(3.20)方程组(1)式 $f(x) - 13x_1 - 11x_2 = 0$ 中选系数最负的一个变量为引入变量,即 $x_k = x_1$ 。

② 选择退出变量 x_r ,从 x_3, x_4, x_5 三个变量中选一个,该变量应随 x_1 增大,率先变为 0。

已知 $x_2 = 0$,由式(3.20)方程组中的(2)式, $x_3 = 1500 - 4x_1 - 5x_2$,若取 $x_3 = 0$,则 $x_1 = \frac{1500}{4} = 375 = \frac{b_1}{a_{11}}$,由式(3.20)方程组(3)式, $x_4 = 1575 - 5x_1 - 3x_2$,若取 $x_4 = 0$,则 $x_1 = \frac{1575}{5} = 315 = \frac{b_2}{a_{12}}$,由式(3.20)方程组(4)式, $x_5 = 420 - x_1 - 2x_2$,若取 $x_5 = 0$,则 $x_1 = 420 = \frac{b_3}{a_{13}}$,由此选 $\theta_{\min} = \min \frac{b_i}{a_{ki}}$ (或引入变量为 x_k)。则本例 $\theta_{\min} = 315$,即 $x_1 = 315$ 时, $x_4 = 0$ 。若 $x_1 > 315$ 时, $x_4 < 0$ 不满足非负约束,所以 x_1 从 0 变到 315 时, x_4 率先为 0,故选 x_4 为退出变量,即 $x_r = x_4 = 0$,所以此时 $x_1 = 315, x_2 = 0, x_4 = 0$ 。

③ 求其余变量 x_3, x_5 的值

用代数中的高斯消元法,使目标函数与约束方程中分别只含一个基础变量,即 x_1, x_3, x_5 以及 $f(x)$, (这里 $f(x)$ 暂时作为基础变量)。对(3.20)式中的(2)、(3)、(4)进行如(3.22)式的运算,并由式(3.22)方程组中的(3)'式计算,所得 $\frac{b_2}{a_{12}}$ 为最小,所以(3)'式是确定引入变量和退出变量的关键,故称该式为关键式,引入变量 x_1 的系数 5 则称为关键系数,则有:

$$\left. \begin{array}{l} (3) \times (-\frac{4}{5}) + (2): \quad \frac{13}{5}x_2 + x_3 - \frac{4}{5}x_4 = 240 \cdots \cdots (2)' \\ \hspace{15em} ((2) \text{ 式中消去 } x_1) \\ (3): \quad x_1 + \frac{3}{5}x_2 + \frac{1}{5}x_4 = 315 \cdots \cdots (3)' \\ (3) \times (-\frac{1}{5}) + (4): \quad \frac{7}{5}x_2 - \frac{1}{5}x_4 + x_5 = 105 \cdots \cdots (4)' \\ \hspace{15em} ((4) \text{ 式中消去 } x_1) \\ (3) \times \frac{13}{5} + (1): \quad f(x) - \frac{16}{5}x_2 + \frac{13}{5}x_4 = 4095 \cdots \cdots (1)' \\ \hspace{15em} ((1) \text{ 式中消去 } x_1) \end{array} \right\} \quad (3.22)$$

在式(3.22)中令 $x_2=0, x_4=0$, 可得 $f(x)$ 的解, 相当于图 3-3 中的 C 点。

以上第一次迭代结果用 $X^{[1]}$ 表示:

$$X^{[1]} = [315 \quad 0 \quad 240 \quad 0 \quad 105]^T$$

再由式(3.22)方程组中(1)'式中得,

$$f(x^{[1]}) = 4095 \text{ 大于初始解 } f(x^0) = 0$$

第四, 第二次迭代, 重复上次步骤, 找新的邻近顶点, 又一次进行换元, 即求新的引入变量 $x_k = x_2$ 为引入变量(实际上本例只选 x_2 为引入变量), 从 x_1, x_3, x_5 中选其中一个为退出变量 x_r :

由(2)'式, 若 $x_3=0$, 则 $x_2=240/13/5, b_1/a_{21}=1200/13$

由(3)'式, 若 $x_1=0$, 则 $x_2=315/3/5, b_2/a_{22}=1575/3=525$

由(4)'式, 若 $x_5=0$, 则 $x_2=105/7/5, b_3/a_{23}=75$ 。

由此选 $\theta_{\min} b_i/a_{ik}$, 其中(4)'中的 b_3/a_{23} 为最小, 故(4)'为关键式。

$\theta_{\min}=75$, 即 x_2 从 0 变到 75 时, x_5 率先为 0, 故 $x_r = x_5$ 。

所以新的非基础变量为 $x_4=0, x_5=0$, 基础变量 $x_2=75$ 。再用上述同样方法确定 x_1, x_3 ; 即保证每个方程中只含一个基础变量 $[f(x) \quad x_1 \quad x_2 \quad x_3]$ 。显然式(3.22)中(4)'为关键式, x_2 的系数 $7/5$ 为关键系数, 所以经过对(3.22)式中的(1)'、(2)'、(3)'、(4)'运算之后, 得到

$$\left. \begin{aligned} (1)' + (4)' \times \frac{16/5}{7/5}: \quad & f(x) + \frac{15}{7}x_4 + \frac{16}{7}x_5 = 4335 \dots\dots\dots(1)'' \\ (2)' + (4)' \times \frac{-13/5}{7/5}: \quad & x_3 - \frac{3}{7}x_4 - \frac{13}{7}x_5 = 45 \dots\dots\dots(2)'' \\ (3)' + (4)' \times \frac{-3/5}{7/5}: \quad & x_1 + \frac{10}{35}x_4 - \frac{3}{7}x_5 = 270 \dots\dots\dots(3)'' \\ (4)' \div \frac{7}{5}: \quad & x_2 - \frac{1}{7}x_4 + \frac{5}{7}x_5 = 75 \dots\dots\dots(4)'' \end{aligned} \right\} \quad (3.23)$$

新的基础变量可行解相当于图 3-3 中的 B 点用 $X^{[2]} = [270 \quad 75 \quad 45 \quad 0 \quad 0]^T$ 表示。

而 $f(X^{[2]}) = 4335 \geq f(X^{[1]})$

第五, 检查式(3.23)中的(1)''式, 其中 x_4, x_5 的系数均为正, 无需再换元, 即只有当 x_4, x_5 均为零时, $f(x)$ 为最大, 故 B 点为最优解。停止迭代。

这时的松弛变量 $x_4 = x_5 = 0$, 而 $x_3 \neq 0$, 表示有关 x_3 的约束方程还有松弛的余地。

将上述运算结果画成系数表格形式, 称为单纯形表。

表 3-4 单纯形表(初始表)

	变 量	x_1 关键列	x_2	x_3	x_4	x_5	
系 数		-13	-11	0	0	0	$f(x)=0$
		4	5	1	0	0	1500
	关键行	⑤	3	0	[1]	0	1575
		1	2	0	0	1	420

表中第三行与关键式(3)相对应称为关键行。 x_1 的系数最负, 作为第一行迭代的引入

变量,与 x_4 互换。 x_1 这一列的系数称为关键列,关键行和关键列上的系数称为关键系数,用圆括号标出。退出变量 x_4 的系数用方括号标出。

初始值

$$x^0 = [0 \ 0 \ 1500 \ 1575 \ 420]^T$$

表 3-5 单纯形表(第一表)

	变 量	x_1	x_2 关键列	x_3	x_4	x_5	
系 数		0	-16/5	0	13/5	0	4095
		0	13/5	1	-4/5	0	240
		1	3/5	0	1/5	0	315
	关键行	0	7/5	0	1/5	[1]	105

第一次迭代后:

$$x^1 = [315 \ 0 \ 240 \ 0 \ 105]^T$$

下一次迭代 x_2 为新的引入变量与退出变量 x_5 互换, x_2 所在的列为关键列,得第二表:

表 3-6 单纯形表(第二表)

	x_1	x_2	x_3	x_4	x_5	
系 数	0	0	0	15/7	16/7	4335
	0	0	1	-3/7	-13/7	45
	1	0	0	2/7	-3/7	270
	0	1	0	-1/7	5/7	75

第二次迭代后:

$$x^2 = [270 \ 75 \ 45 \ 0 \ 0]^T$$

目标函数方程的系数均大于 0,已得最优解,停止计算。

(2) 单纯形法的一般理论

从上述实例中,可归纳出如下规律:

① 求基础可行解

变量 $x = [x_1 \ x_2]^T$

在 $m+n$ 个变量中(包括 m 个松弛变量),任选 n 个等于零。再根据 m 个约束求其他 m 个变量,记 x_1 为 m 维非零向量,称为基础变量,而 x_2 为 n 维非零向量,称为非基础变量。

约束方程的系数矩阵: $A_{m \times (m+n)} = A_1 \underset{m \times m}{\times} A_2 \underset{m \times (m+n)}{\times} A = \underset{m \times m}{[A_1 \ A_2]}$

等式约束

$$AX=B$$

(3.24)

即

$$[A_1 \ A_2] \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = A_1 X_1 + A_2 X_2 = B$$

$$X_1 = A_1^{-1} B - A_1^{-1} A_2 X_2$$

A_1 的秩为 m ,非奇异,其逆 A_1^{-1} 存在。

因 $X_2=0$,故基础解 $X_1=A_1^{-1}B$ 。

若 $X_1 \geq 0$ 又满足 m 个约束方程,则 X_1 即为基础可行解,得 $X=[X_1 \ 0]^T$ 。这样单纯形法将原来求 n 维可行解域中有无穷多个解的问题转变为求维数低于 n 的有限个基础可行

解,使极值问题的搜索范围大大减小了。

② 求最优解

$$f(x) = C^T X = C_1^T X_1 + C_2^T X_2$$

$$C = [C_1 \quad C_2]^T$$

已知 $X_1 = A_1^{-1} B - A_1^{-1} A_2 X_2$, 代入上式得

$$f(x) - C_1^T A_1^{-1} B - (C_2^T - C_1^T A_1^{-1} A_2) X_2 = 0$$

$$X_2 = 0 \quad f(x) = C_1^T A_1^{-1} B$$

若 X_2 的系数为负, 即 $(C_2^T - C_1^T A_1^{-1} A_2) > 0$, 则 X_2 中任一变量成为引入变量。例如向量 X_2 中任一元素 X_{2i} 由零增加, 变为非零的正数, 将使 $f(x)$ 增大, 即目标函数值比以前更好。否则, 若 X_2 的系数非负, $(C_2^T - C_1^T A_1^{-1} A_2) \leq 0$, 说明已为最优解。若 X_{2i} 由零增加, 都会使 $f(x)$ 下降。

相反, 若问题是求极小值, 则 $(C_2^T - C_1^T A_1^{-1} A_2) \geq 0$ 才达最优。所以

$$(C_2^T - C_1^T A_1^{-1} A_2) \leq 0 \quad (3.25)$$

并记

$$\sigma_j = (C_2^T - C_1^T A_1^{-1} A_2)$$

为单纯形求其最优解的判据。因此称 σ_j 为检验数。

由上例的初始表可求得

$$X_1 = [1500 \quad 1575 \quad 420]^T$$

$$X_2 = [0 \quad 0]^T$$

$$A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 4 & 5 \\ 5 & 3 \\ 1 & 2 \end{bmatrix}$$

为便于计算, 一般希望初始基础变量的系数矩阵为单位矩阵。同时

$$C_1 = [0 \quad 0 \quad 0]^T, \quad C_2 = [13 \quad 11]^T$$

代入得:

$$C_2^T - C_1^T A_1^{-1} A_2 = C_2^T = [13 \quad 11] > 0$$

所以未达到最优解

而

$$X_1 = [x_3 \quad x_4 \quad x_5]^T, \quad X_2 = [x_1 \quad x_2]^T = 0$$

代入

$$A_1 X_1 + A_2 X_2 = B$$

得:

$$X_1 = A_1^{-1} (B - A_2 X_2) = B = \begin{bmatrix} 1500 \\ 1575 \\ 420 \end{bmatrix}$$

在第二表中

$$X_1 = [270 \quad 75 \quad 45]^T, \quad X_2 = [0 \quad 0]^T$$

$$C_1 = [0 \quad 0 \quad 0]^T, \quad C_2 = \left[-\frac{15}{7} \quad -\frac{16}{7}\right]$$

$$A_1 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad A_2 = \frac{1}{7} \begin{bmatrix} -3 & -13 \\ 2 & -3 \\ -1 & 5 \end{bmatrix}$$

∴

$$C_2^T - C_1^T A_1^{-1} A_2 = \left[-\frac{15}{7} \quad -\frac{16}{7}\right] < 0$$

即达最优解。

③ 单纯形法要求基础变量对应的系数矩阵为单位矩阵。

(3) 单纯形法小结

① 引入松弛变量。将线性规划模型中不等式约束化为等式约束。

$$\text{求 } \max f(x) = \sum_{j=1}^n c_j x_j, f(x) - \sum_{j=1}^n c_j x_j = 0$$

$$\sum_{j=1}^n a_{ij} x_j = b, \begin{pmatrix} i = 1, 2, \dots, m \\ j = 1, 2, \dots, n \end{pmatrix}$$

选松弛变量为初始基础变量,则初始基础变量可行解为

$$x_i = b_i \quad (i=1, 2, \dots, m)$$

第 i 个基础变量 x_i 为对应的第 i 个约束方程的松弛变量。

② 检查目标函数等式,如果至少有一个非基础变量的系数为负,说明目标函数还可以进一步增加。否则,各系数非负,说明已达到最优,停止计算。

③ 选系数最负的一个基础变量 x_k 为引入变量,作为新的基础变量。若有两个或两个以上的非基础变量系数均为最小,则可任选一个。 x_k 所在的列为关键列。

④ 决定退出变量。计算约束方程每列系数

$$\theta_i = \frac{b_i}{a_{ki}} = \text{基础可行解除以关键列的系数}$$

其中 a_{ki} 为第 i 个方程中 x_k 之系数,即关键列的各系数。

令 $\theta_{\min}$ 所在的行为关键行,位于关键行、关键列上的元素为关键元素,用圆圈标明。

若将上述初始表重画在下表,将会进一步理解关键行、关键列的涵义。

		关键列					
		x_1	x_2	x_3	x_4	x_5	
关键行		-13	-11	0	0	0	0
		4	5	0	0	0	1500
		5	3	0	1	0	1575
		1	2	0	0	1	420
							$\theta_1 = 375$
							$\theta_2 = 315$
							$\theta_3 = 420$

⑤ 为关键元素

$$\theta = [375 \quad 315 \quad 420]^T \text{ 中, 选 } \theta_{\min} = 315$$

取关键行上系数为 1 的基础变量 x_r 为退出变量。

⑥ 用高斯消去法使关键行 x_k 的系数(关键元素)为 1,其余行 x_k 的系数为 0,决定新的基础可行解。

⑦ 检验目标函数方程中,所有变量系数。即检验各系数是否均 ≥ 0 ,若是,则取得最优解,否则,重复 3~5 步,直到没有新的关键列出现。

3.1.4 线性规划的对偶问题

在线性规划问题中,不论从理论方面还是从实际方面,都存在一个有趣的问题:就是对于任何一个求极大值的线性规划问题必有一个求极小值的规划问题与之相对应。反之亦然。而且两者包含有相同的数据,它们的解之间有紧密联系,分别从不同角度反映着同一个

① 对偶问题与原问题的关系

在对偶问题的定义中,可以将 I 和 II 中的任何一种形式作为原问题,则另一种问题就是它的对偶问题;因此,已知一个线性规划问题,为求其对偶,可根据具体情况将原问题转化为 I 或 II,就可直接写出对偶问题,具体方法如下:

如果目标函数为求最大值,就应将约束条件全部转化为“ $\leq$ ”形式。如果原约束条件不等式为“ $\geq$ ”形式,则可在该式两边同乘以“-1”,不等式改变方向;若约束条件为等式约束,则可变为“ $\leq$ ”和“ $\geq$ ”两个不等式,然后将“ $\geq$ ”形式再转为“ $\leq$ ”形式;若原问题中有自由变量,应当用两个非负变量相减代替。

显然,若目标函数是求最小值,就应当将约束条件全部转换为“ $\geq$ ”形式。

【例 3.5】 写出下面的对偶问题

$$\begin{aligned} \min f(x) &= 5x_1 + 8x_2 + 20x_3 \\ \text{s. t.} \quad &\begin{cases} -x_1 + 6x_2 + 12x_3 \geq 1 \\ x_1 + 7x_2 - 9x_3 \geq 2 \\ -x_1 - 3x_2 + 9x_3 \geq -3 \\ -3x_1 - 5x_2 + 9x_3 = 4 \end{cases} \\ &x_2, x_3 \geq 0, x_1 \text{ 为自由变量} \end{aligned} \quad (3.30)$$

【解】 设 $x_1 = x'_1 - x''_1$, 并将第四个约束条件化为不等式,改写原问题为:

$$\begin{aligned} \min f(x) &= 5x'_1 - 5x''_1 + 8x_2 + 20x_3 \\ \text{s. t.} \quad &\begin{cases} -x'_1 + x''_1 + 6x_2 + 12x_3 \geq 1 \\ x'_1 - x''_1 + 7x_2 - 9x_3 \geq 2 \\ -x'_1 + x''_1 - 3x_2 + 9x_3 \geq -3 \\ -3x'_1 + 3x''_1 - 5x_2 + 9x_3 \geq 4 \\ 3x'_1 - 3x''_1 + 5x_2 - 9x_3 \geq -4 \\ x'_1, x''_1, x_2, x_3 \geq 0 \end{cases} \end{aligned} \quad (3.31)$$

对偶问题为

$$\begin{aligned} \max f(x) &= y_1 + 2y_2 - 3y_3 + 4y_4 - 4y_5 \\ \text{s. t.} \quad &\begin{cases} -y_1 + y_2 - y_3 - 3y_4 + 3y_5 \leq 5 \\ y_1 - y_2 + y_3 + 3y_4 - 3y_5 \leq -5 \\ 6y_1 + 7y_2 - 3y_3 - 5y_4 + 5y_5 \leq 8 \\ 12y_1 - 9y_2 + 9y_3 + 9y_4 - 9y_5 \leq 20 \\ y_1, y_2, y_3, y_4, y_5 \geq 0 \end{cases} \end{aligned}$$

② 对偶问题的最优解与原问题最优解的关系

互为对偶的两个问题的最优解之间有着密切的关系:

A 互为对偶的两个问题,求极小问题的任意可行解的目标函数值均不小于求极大问题的最优目标函数值;而求极大问题的任一可行解的目标函数值都不大于求极小问题的最优目标函数值。

B 如果一个线性规划问题有最优解,则其对偶问题也一定有最优解,且两个问题的最优目标函数值相等。

极大值问题的原变量 x_j 的值等于极小值对偶问题的第 j 个剩余变量的检验数。

【例 3.6】 极大值问题

$$\begin{aligned} \max f(x) &= 20x_1 + 6x_2 + 10x_3 \\ \text{s. t.} \quad &\begin{cases} x_1 + \frac{1}{2}x_2 + 2x_3 \leq 6 \\ x_1 + \frac{1}{2}x_2 + x_3 \leq 3 \\ x_1 + \frac{1}{4}x_2 + x_3 \leq 2 \\ x_1, x_2, x_3 \geq 0 \end{cases} \end{aligned}$$

【解】 与原问题的对偶的极小值问题为：

$$\begin{aligned} \min f(y) &= 6y_1 + 3y_2 + 2y_3 \\ \text{s. t.} \quad &\begin{cases} y_1 + y_2 + y_3 \geq 20 \\ \frac{1}{2}y_1 + \frac{1}{2}y_2 + \frac{1}{4}y_3 \geq 6 \\ 2y_1 + y_2 + y_3 \geq 10 \\ y_1, y_2, y_3 \geq 0 \end{cases} \end{aligned}$$

表 3-7 原问题(max)最优单纯形表

	c_i	20	6	10	0	0	0	
c_B	基	x_1	x_2	x_3	x_4	x_5	x_6	b
0	x_4	0	0	1	1	-1	0	3
6	x_2	0	1	0	0	4	-4	4
20	x_1	1	0	1	0	-1	2	1
	$\bar{c}_i$	0	0	-10	0	-4	-16	

表 3-8

	c_j	6	3	2	0	0	0	M	M	M	
c_B	基	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	b
2	y_3	0	0	1	-2	4	0	2	-4	1	16
0	y_6	-1	0	0	-1	0	1	1	0	-1	10
3	y_2	1	1	0	1	-4	0	-1	4	0	4
	$\bar{c}_j$	3	0	2	1	4	0	$M-1$	$M-4$	M	

原问题的最优解为 $x_1=1, x_2=4, x_3=0$, 与此对应的对偶问题的松弛变量 y_4 的检验数为 1, y_5 的检验数为 4, y_6 的检验数为 0。

对偶问题的最优解 $y_1=0, y_2=4, y_3=16$, 与此对应的原问题的松弛变量 x_4 的检验数为 0, x_5 的检验数为 -4, x_6 的检验数为 -16。

对偶问题最优解的关系为我们求解线性规划问题提供了方便。如果一个线性规划问题比较复杂, 而它的对偶问题比较简单时, 就可通过求解对偶问题而得到原问题的最优解。

3.2 目标规划

线性规划虽然得到了广泛成功的应用,但仍存在着局限性,它只限于求解单一的目标,例如最大利润或最小费用等。而在实际中有许多具体问题不可能只存在一个目标,尤其是社会经济系统问题,经常提出一系列的目标。这些目标,既有轻重缓急之别,又有时相互矛盾,而且彼此的量度单位不尽一致。例如某企业不仅要求利润尽可能大,还要求花费的工时(小时)尽量少;再如水库的水资源,既要求多发电,又要求保持较高的水位,以满足旅游和农业的需要等。在这种情况下,如果仍然沿用线性规划的方法,就会有困难。而目标规划法正好可以对线性规划存在的问题予以改进。

3.2.1 目标规划的数学模型

目标规划模型的建立类似线性规划。首先必须确定决策定量,选择约束条件和各个目标函数,构造出多目标的线性规划。在此基础上,根据决策者的意图,对各目标确定相应的目标值,接着在这些目标函数方程中加入偏离系数,组成新的一组约束条件,然后用这些偏离系数产生目标规划的目标函数,再根据各个目标的轻重缓急赋予不同的优先级因子,使目标函数达到最小,这就是目标规划的计算方法。

一般目标规划的数学模型可以表述为:

$$\min F = \sum_{i=1}^m \omega_i (d_i^- + d_i^+) \quad (3.32)$$

$$\text{s. t. } \left. \begin{array}{l} \sum_{j=1}^n a_{ij} x_j + d_i^- - d_i^+ = b_i \quad i=1, 2, \dots, m \\ x_j \geq 0 \quad j=1, 2, \dots, n \\ d_i^- \cdot d_i^+ = 0 \\ b_i, d_i^-, d_i^+ \geq 0 \end{array} \right\} \quad (3.33)$$

其中 d_i^- ——偏离系数,表示达不到目标的偏离程度;

d_i^+ ——偏离系数,表示超过目标的偏离程度;

ω_i ——决策者根据具体情况选定的优先次序的权因子;

$d_i^- \cdot d_i^+ = 0$ ——表示对任何一个目标在能否达到或超过偏离值方面,只可能选取其中之一,如 d_i^- 不为 0,则 d_i^+ 为 0;如 d_i^+ 不为 0,则 d_i^- 为 0。

3.2.2 目标规划的单纯形法

目标规划的求解方法,仍然采用求解线性规划时的单纯形法。下面可以通过例子加以说明:

【例 3.7】 某企业生产两种产品 x_1 和 x_2 ,需消耗两种原材料,每生产 1 个 x_1 需甲原料 1 个单位,乙原料 1 个单位;每生产 1 个 x_2 需 3 个单位甲原料和 1 个单位的乙原料,每个 x_1 产品的利润为 15 元,每个 x_2 产品的利润为 25 元。企业现有甲原料 60 单位,乙原料 40 单位,问如何组织生产才可能获得最大利润?

显然,这是一个线性规划的问题,其数学模型为:

$$\max f(x) = 15x_1 + 25x_2 \quad (3.34)$$

$$\text{s. t.} \quad \begin{cases} x_1 + 3x_2 \leq 60 \\ x_1 + x_2 \leq 40 \\ x_1, x_2 \geq 0 \end{cases} \quad (3.35)$$

用单纯形法求解,得到最优解为 $x_1 = 30, x_2 = 10$,最大利润 $f(x) = 700$ 元,原料正好用完。

现在,企业决策者根据市场实际情况,认为目标达到 700 元的利润不现实,因此把利润目标下调至 600 元左右,并要求重新安排生产计划。由于只有一个利润目标,所以设定 d_1^- , d_1^+ 为利润目标的偏离系数,这样,就形成了一个简单的目标规划问题。

$$\min F = d_1^- + d_1^+ \quad (3.36)$$

$$\text{s. t.} \quad \begin{cases} x_1 + 3x_2 \leq 60 \\ x_1 + x_2 \leq 40 \end{cases} \quad (\text{原料约束}) \quad (3.37)$$

$$15x_1 + 25x_2 + d_1^- - d_1^+ = 600 \quad (\text{利润目标约束})$$

$$x_1, x_2, d_1^-, d_1^+ \geq 0$$

增加资源约束的松弛变量 s_1 和 s_2 ,得以下目标规划:

$$\min F = d_1^- + d_1^+ \quad (3.38)$$

$$\text{s. t.} \quad \begin{cases} x_1 + 3x_2 + s_1 = 60 \\ x_1 + x_2 + s_2 = 40 \\ 15x_1 + 25x_2 + d_1^- - d_1^+ = 600 \\ x_1, x_2, s_1, s_2, d_1^-, d_1^+ \geq 0 \end{cases} \quad (3.39)$$

通过单纯形法求得答案,最优解为 $x_1 = 40, x_2 = 0$,最大利润 $f(x) = 600$ 元,此时甲原料剩余 20 个单位,乙原料不剩余。

这里目标函数 $= d_1^- + d_1^+$ 表示决策者对于达到 600 元的利润目标未给予任何偏向。因此 d_1^+ 和 d_1^- 的权因子均为 1,但是如果决策者认为,对超过 600 元利润的要求更为迫切,那么就应该使目标函数 $F = d_1^- + d_1^+$ 中的 d_1^- 更优先为 0,尽量使 d_1^+ 有某个值,即超过利润目标的某个偏差值。于是,目标函数应该写成: $\min F = \omega_1 d_1^- + \omega_2 d_1^+$,显然权因子 $\omega_1 > \omega_2$ 。

现在,再分析和改进上述模型,决策者根据某种需求,对未来的企业调整提出以下指令:

① 需要保证 600 元的利润目标;② 应该充分利用现有的原材料。于是形成了多个目标的目标规划。即首先要实现 600 元的利润目标,这时应赋予最大的权因子,其次要实现充分利用甲原料和乙原料的目标,所以赋予以比利润目标较小的权因子。显然,这样的决策要求,应用通常的线性规划就难以实现,而应用目标规划的模型就有独到之处。

其目标规划模型为:

$$\min F = \omega_1 (d_1^- + d_1^+) + \omega_2 (d_2^- + d_3^-) \quad (3.40)$$

$$\text{s. t.} \quad \begin{cases} 15x_1 + 25x_2 + (d_2^- - d_2^+) = 600 \\ x_1 + 3x_2 + (d_1^- - d_1^+) = 60 \\ x_1 + x_2 + d_3^- - d_3^+ = 40 \\ x_1, x_2, d_i^-, d_i^+ \geq 0 \quad i=1,2,3 \end{cases} \quad (3.41)$$

其中的 $\omega_1 > \omega_2$,于是依然可以运用单纯形法求解,得到最优解为: $x_1 = 15, x_2 = 15, d_3^- =$

10,这就意味着保证利润值为 600 元,而乙原料尚剩 10 个单位。

以上模型对甲、乙两种原料是“现有”的进行了限制,即指不能超过原有的原料,但如果决策者根据新的信息进行综合分析以后,给出以下要求:

- ① 尽量保证利润值为 750 元的目标。
- ② 充分利用现有的原材料,但允许再购进适量原料。
- ③ 甲原料的价格三倍于乙原料,在材料不够用时,应尽可能少补充甲原料。

这个问题又比以上目标规划增加了第③个约束条件,是在上述原料目标的约束条件中又增加了一个子目标。这时的目标规划的数学模型为:

$$\begin{aligned} \min F &= \omega_1(d_1^- + d_1^+) + \omega_2(d_2^- + d_2^+) + 3\omega_3 d_2^+ + \omega_3 d_3^+ \\ \text{s. t. } &\begin{cases} 15x_1 + 25x_2 + d_1^- - d_1^+ = 750 & (\text{目标利润}) \\ x_1 + 3x_2 + d_1^- - d_2^+ = 60 & (\text{甲原料资源}) \\ x_1 + x_2 + d_2^- - d_3^+ = 40 & (\text{乙原料资源}) \\ x_1, x_2, d_1^-, d_1^+, d_2^-, d_2^+, d_3^-, d_3^+ \geq 0 \end{cases} \end{aligned} \quad (3.42)$$

其中, $\omega_1 > \omega_2 > \omega_3$ 。

现用单纯形法进行求解,具体计算步骤如下:

(1) 目标规划模型的单纯形初始表(表 3-9)

由于目标规划是把所有的变量分成不同的优先级的,所以表 3-9 中的检验数 b ,是一个 $m \times n$ 阶矩阵(在本例中为 3×8),其中 m 是优先级的数目, n 是变量的数目,包括决策变量 x 与偏离系数变量 d 。

但单纯形初始表格的建立与前面的线性规划模型同理。

表 3-9 单纯形初始表

c_j		0	0	ω_1	ω_1	ω_2	$3\omega_3$	ω_2	ω_3	b
C_B	X_B	x_1	x_2	d_1^-	d_1^+	d_2^-	d_2^+	d_3^-	d_3^+	
ω_1	d_1^-	15	25	1	-1	0	0	0	0	750
ω_2	d_2^-	1	③	0	0	1	-1	0	0	60
ω_2	d_3^-	1	1	0	0	0	0	1	-1	40
σ_j	ω_3	0	0	0	0	0	-3	0	-1	0
	ω_2	2	4	0	0	0	-1	0	-1	100
	ω_1	15	25	0	-2	0	0	0	0	750

(2) 决定引入变量和退出变量

目标规划是按优先等级不同而分先后的,首先应考虑满足权因子最大的要求,然后依顺序处理次一级的权因子,所以在表 3-9 的检验数矩阵中首先应满足 ω_1 的要求开始,直到 ω_1 行向量的所有数都小于或等于零时,才能转入 ω_2 行进行检验。按此原则,从表 3-9 中可以知道引入变量,再根据线性规划中所述的可行性检验准则(即迭代比)决定 d_2^- 应是退出变量,于是就得到了进行第一次迭代后的单纯形表(表 3-10)。

如此反复进行,便相应得到第二次迭代后的单纯形表(表 3-11)和第三次迭代后的单纯形表(表 3-12)。最后的最优结果列入表 3-13。

表 3-10 第一次迭代后表

c_j		0	0	ω_1	ω_1	ω_2	$3\omega_3$	ω_2	ω_3	b
C_B	X_B	x_1	x_2	d_1^-	d_1^+	d_2^-	d_2^+	d_3^-	d_3^+	
ω_1	d_1^-	20/3	0	1	-1	-25/3	25/3	0	0	250
0	x_2	1/3	1	0	0	1/3	-1/3	0	0	20
ω_2	d_3^-	2/3	0	0	0	-1/3	1/3	1	-1	20
σ_j	ω_3	0	0	0	0	0	-3	0	-1	0
	ω_2	2/3	0	0	0	-4/3	1/3	0	-1	20
	ω_1	20/3	0	0	-2	-25/3	25/3	0	0	250

表 3-11 第二次迭代后表

c_j		0	0	ω_1	ω_1	ω_2	$3\omega_3$	ω_2	ω_3	b
C_B	X_B	x_1	x_2	d_1^-	d_1^+	d_2^-	d_2^+	d_3^-	d_3^+	
$3\omega_3$	d_2^+	4/5	0	3/25	-3/25	-1	1	0	0	30
0	x_2	9/15	1	1/25	1/25	0	0	0	0	30
ω_2	d_3^-	6/15	0	-1/25	1/25	0	0	1	-1	10
σ_j	ω_3	12/5	0	9/25	-9/25	-3	0	0	-1	90
	ω_2	6/15	0	-1/25	1/25	-1	0	0	-1	10
	ω_1	0	0	-1	-1	0	0	0	0	0

表 3-12 第三次迭代后表

c_j		0	0	ω_1	ω_1	ω_2	$3\omega_3$	ω_2	ω_3	b
C_B	X_B	x_1	x_2	d_1^-	d_1^+	d_2^-	d_2^+	d_3^-	d_3^+	
$3\omega_3$	d_2^+	0	0	1/5	-1/5	-1	1	-2	②	10
0	x_2	0	1	1/10	-1/10	0	0	-3/2	3/2	15
0	x_1	1	0	-1/10	1/10	0	0	0	0	25
σ_j	ω_3	0	0	3/5	-3/5	-3	0	-6	5	30
	ω_2	0	0	0	0	-1	0	-1	0	0
	ω_1	0	0	-1	-1	0	0	0	0	0

表 3-13 最优结果表

c_j		0	0	ω_1	ω_1	ω_2	$3\omega_3$	ω_2	ω_3	b
C_B	X_B	x_1	x_2	d_1^-	d_1^+	d_2^-	d_2^+	d_3^-	d_3^+	
ω_3	d_2^+	0	0	2/5	-2/5	-1/2	1/2	-1	1	5
0	x_2	0	1	-1/2	1/2	3/4	-3/4	0	0	15/2
0	x_1	1	0	9/10	-9/10	-5/4	5/4	0	0	75/2
σ_j	ω_3	0	0	2/5	-2/5	-1/2	-2/5	-1	0	5
	ω_2	0	0	0	0	-1	0	-1	0	0
	ω_1	0	0	-1	-1	0	0	0	0	0

(3) 判断求得最优解的准则

与线性规划单纯形法相类似,对所得可行解必须进行最优性判断,其判断依据为检验数 $\sigma_j, \sigma_j = C_B A - c_j$ 进行检验,在 σ_j 式中 C_B 为利润目标函数约束式中,对应于基础变量的系数阵。同时约束方程(3.33)式中 $A = \sum_{j=1}^n a_{ij}, (i=1,2,\dots,m, j=1,2,\dots,n)$, 在 n 个决策变量中,有 m 个基础变量, $(n-m)$ 个非基础变量,而 B 则是 A 阵中一个 $m \times m$ 维非奇异子矩阵,与基础变量相对应。 c_j 为利润目标约束式中 x_j 的系数。

当求极小值的目标规划时,应符合以下准则,否则不是最优值。

① 检验 $\omega_1, \omega_2 \dots \omega_m$ 各行之 σ_j 值,均为零或小于零。

② 如果检验数 σ_j 的某行中存在有一正数,便要看同列的高一级优先权因子的 σ_j 是否为负数,若是负数,便为最优解。如表3-13中 ω_3 行 d_1^- 列为 $2/5$,但 ω_1 行 d_1^- 列为 -1 ,故为最优解。

最后求得最优解 $x_1 = 75/2, x_2 = 15/2, d_3^+ = 5$ 。故最优化的结果是:利润保证为750元。甲原材料恰好用完,乙原材料不足,需要再购进5个单位。

3.2.3 目标规划的图解法

对于只具有两个决策变量的目标规划的数学模型,用图解法求解简单直观,下面结合例题的求解说明图解法的步骤。

用图解法求解下例:

$$\begin{aligned} \max f(x) &= 4x_1 + 3x_2 \\ \text{s. t.} \quad &\begin{cases} 2x_1 + 3x_2 \leq 24 \\ 3x_1 + 2x_2 \leq 26 \\ x_1, x_2 \geq 0 \end{cases} \end{aligned}$$

【解】 (1) 现在平面直角坐标中做出各约束条件所确定的区域。

资源约束条件的作图与线性规划相同,如图3-4中的阴影部分,即三角形 OAB 所在的区域为可行解域。对目标约束条件,令相应的正、负偏差(d_i^+, d_i^-)均为0,作相应的直线。

(2) 标出目标约束在相应直线上 d_i^+, d_i^- 的方向。

(3) 根据目标函数的优先因子分析求解。

在可行域中,首先考虑 P_1 的实现,即要求 $\min d_1^-$,从图中可见,满足 $d_1^- = 0$ 的区域为三角形 OBC ;接着考虑 P_2 相应的目标,即在区域 OBC 中要实现 $\min(d_2^+ + d_2^-)$,容易看出, $C(24/5, 24/5)$ 点距离直线(2)最近,即可使 $(d_2^+ + d_2^-)$ 取最小,此时 $d_2^+ = 0, d_2^- = 2$;最后,同理考虑 P_3 相应目标的实现,此时只有 C 点, $d_3^- = 0, d_3^+ = 18/5$ 。因此,满意解为 $x_1 = 24/5, x_2 = 24/5$,相应的利润为 $168/5 = 33.6$ 元。

图解法不仅直观表达了 d_i^+ 和 d_i^- 的几何意义,而且可以得出(理论上也可以证明),任何目标规划问题都有满意解的结论。

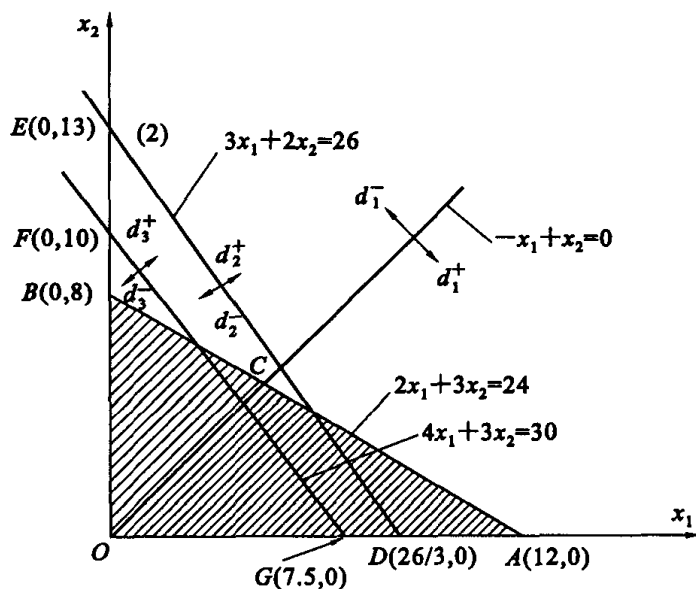


图 3-4

3.3 整数规划

整数规划问题,严格说来属于具有特殊约束条件下的线性规划问题。线性规划对决策变量只限于取非负值的连续型数值,而整数规划则要求取离散的非负整数值。在解决生产实际问题时,求解整数规划是一个十分现实和具有重要意义的问题。因为在很多实际问题中,要求的变量往往是整数值,如运载工具(船、车)装运货物如何充分利用其运载能力;经济规划中的企业布局如何合理;工厂大型设备的安装如何充分有效利用空间等等,这些都是必需整件或整体地予以考虑,都要用整数规划的方法来解决。

当然,在求解整数规划问题时有时也用通常的线性规划的方法,并将获得的最优化解用“舍入取整”的方法化为整数值来近似的。但在这种情况下,其条件是近似值的误差不能太大。有时问题的本身就是选择一些不可分割的项目,决策者要做要么选取,要么淘汰的决策,显然这种“舍入取整”的近似方法就不可奏效,而必须采用整数规划的方法。

3.3.1 数学模型

【例 3.8】 (装载问题) 一运载工具的最大载重量为 b , 现有几种货物可供装载, 设 j 货物每件重量为 a_j , 每件货物收费为 c_j ($j=1, 2, \dots, n$), 问要采用怎样的装载方案才能使运输费收入最大。

【解】 设 x_j 表示运载工具装载货物的件数。于是可以得以下整数规划模型:

$$\max f(x) = \sum_{j=1}^n c_j x_j \quad (3.41)$$

$$\text{s. t. } \sum_{j=1}^n a_j x_j \leq b \quad (3.42)$$

$$x_j \geq 0 \quad \text{且为整数} \quad j = 1, 2, \dots, n$$

其中,约束条件 $\sum_{j=1}^n a_j x_j \leq b$ 表示运载工具装载的总重量不应该超过它允许的最大载重量。

这里必须说明,如果 b 较大,而相对 c, a 都较小,则用线性规划求得的解,并舍入取整,离最优解的误差一般不会太大。但若有关参数不符合这些条件,则必须用整数规划的方法求解。

【例 3.9】 (生产计划问题)某工厂生产 A_1, A_2 两种产品,产品分别由 B_1, B_2 两种部件组装而成,每件产品所用部件的数量和部件的限额,以及产品的利润由表 3-14 给出。要如何安排 A_1, A_2 的生产数量,工厂才能获得最大利润?

表 3-14 生产安排

部 件	B_1	B_2	利 润
产 品			
A_1	6	1	15
A_2	4	3	20
部件最大限额	25	10	

【解】 设 x_1, x_2 分别表示产品 A_1 和 A_2 的产量。依题意可以建立如下模型:

$$\begin{aligned} \max f(x) &= 15x_1 + 20x_2 \\ \text{s. t. } \begin{cases} 6x_1 + 4x_2 \leq 25 \\ x_1 + 3x_2 \leq 10 \\ x_1, x_2 \geq 0 \quad \text{且为整数} \end{cases} \end{aligned}$$

这是一个纯整数规划问题。又因为约束条件和目标函数的系数全为整数,故亦称为整数规划问题。

3.3.2 整数规划的几何含义

为说明整数规划的几何含义和求解整数规划的思路,特以下例说明。

【例 3.10】 求解 $\max f(x) = x_1 + x_2$ (3.43)

$$\text{s. t. } \begin{cases} 2x_1 + x_2 \leq 6 \\ 4x_1 + 5x_2 \leq 20 \\ x_1, x_2 \geq 0 \quad \text{且为整数} \end{cases} \quad (3.44)$$

【解】 根据给出的约束条件,作图 3-5,其具体求解思路为:

① 用求解线性规划的单纯形法在 B 点求其最优解为:

$$x_1 = \frac{5}{3}, x_2 = \frac{8}{3}, f(x) = \frac{13}{3}$$

显然,这个最优解不满足整数要求,可以用对 x_1 或 x_2 取整的办法来加以限制。若取 $x_1 \leq 1$,即在以上问题中增加了一个求解的约束条件。

② 根据上述增加了 $x_1 \leq 1$ 的约束条件后原线性规划问题转换为:

$$\max f(x) = x_1 + x_2 \quad (3.45)$$

$$\text{s. t. } \begin{cases} 2x_1 + x_2 \leq 6 \\ 4x_1 + 5x_2 \leq 20 \\ x_1 \leq 1 \\ x_1, x_2 \geq 0 \quad \text{且为整数} \end{cases} \quad (3.46)$$

在图中的 Q 点求得最优解为 $x_1=1, x_2=16/5, f(x)=21/5$, 其最优解仍未满足整数要求, 为此, 继续增加约束条件, 令 $x_2 \leq 3$ 。

③ 这样, 原线性规划问题转换为:

$$\max f(x) = x_1 + x_2 \quad (3.47)$$

$$\begin{aligned} & 2x_1 + x_2 \leq 6 \\ & 4x_1 + 5x_2 \leq 20 \\ \text{s. t. } & x_1 \leq 1 \\ & x_2 \leq 3 \\ & x_1, x_2 \geq 0 \text{ 且为整数} \end{aligned} \quad (3.48)$$

根据题意, 在图 3-5 中求得 P 点为最优解, 即 $x_1=1, x_2=3$, 得 $f(x)=4$ 并满足整数

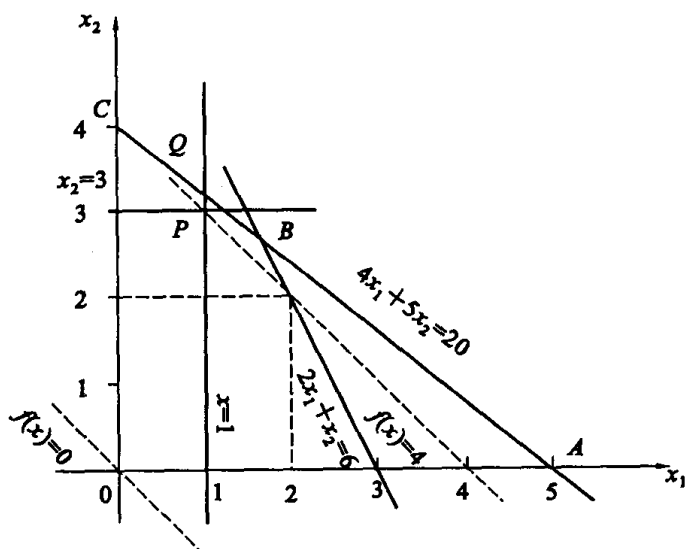


图 3-5 整数规划的单纯形法

要求, 计算结束。

从以上计算中可以发现, 在可行解域内, 寻找有限多的整数极点以求取最优解, 除了已求得 $x_1=1, x_2=3$ 外, 还可取 $x_1 \geq 2, x_2 \geq 2$, 等等, 也可能找到另外的最优解。同时也可从中得知, 整数规划不同于一般线性规划。

3.3.3 分枝定界法

分枝定界法是根据上述的思路, 通过枚举的整数组合, 进行比较, 从而找到整数最优解。其步骤如下:

(1) 松弛: 先不考虑整数约束的因素, 把整数规划中的整数约束条件暂不考虑, 即予以松弛, 使问题变得容易求解。在松弛以后求得的最优解如果满足问题的整数约束条件, 问题就已解决, 如果不满足整数要求, 那这个解的值就是原问题最优解的上界值, 记为 $\bar{f}(x)$ (当求极大时) 或下界值, 记为 $\underline{f}(x)$ (当求极小时)。

(2) 分枝: 即把原问题进行分枝, 分成若干子问题, 并分别补充一个约束条件。如在不满足整数约束条件的变量中任选一个 x_i (设它的值为 $\bar{b}_i$), 则新的约束条件为 $(x_i \leq [\bar{b}_i], x_i \geq [\bar{b}_i] + 1)$ 然后对每一个子问题在求松弛整数约束的情况下求最优解。如果所有子问题的最

优解仍然不是原问题的可行解的话,即仍不满足整数约束条件;则选择其值最大(求极大值时)或最小(求极小值时)的子问题,再进一步分为子问题求解,分枝一直进行下去直到找到一个问题的可行解时为止。

(3) 剪枝:将子问题的边界值与原问题的可行解的值进行比较,对边界值大于(求极小时)或小于(求极大时)可行解的分枝将不再予以考虑,即将这些分枝剪去。

根据这一思路,仍然可以用例 3.10 作为例子,给出分枝定界法的求解步骤(图 3-6)。

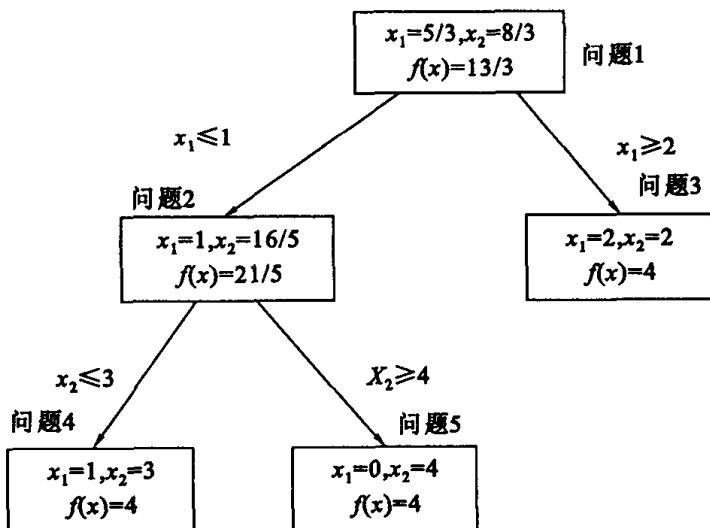


图 3-6 整数规划的分枝定界法

原问题(问题 1)可以分枝成问题 2 和问题 3。问题 3 已具有整数解,而从问题 2 分枝以后的问题 4,问题 5 的边界均为 $f(x)=4$,与原问题保留下来的可行解(问题 3)相等已达到最优化,所以该整数规划的最优解有三组,即:

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$f(x) = 4$$

下面将以本节例 3.9 的生产计划为例说明分枝定界法的方法步骤:

【解】 第一步,先不考虑 x_1, x_2 的整数约束,即把整数约束条件予以松弛。用单纯形法求解最优解 $x_1=2.5, x_2=2.5, f(x)=87.5$ 并取 87.5 为上界值,即 $\bar{f}(x)=87.5$,令下界值 $f(x)=0$ (见图 3-7)。

第二步,将原问题进行分枝,即在 x_1, x_2 两变量中先任选一个(如选 x_1)并对其原问题增加两个约束条件 $x_1 \leq 2$ 和 $x_1 \geq 3$,将原问题 1 分枝成问题 2 和问题 3 两个子问题,可将问题 3 暂放置一旁,求问题 2 的最优解:

$$x_1=2, x_2=2.67, f(x)=83.3$$

显然, x_2 为非整数, $f(x)$ 为非整数值。这时有必要对问题 2 分别增加新的约束条件,从而将问题 2 再分枝成两个子问题,其相应的松弛子问题为问题 4 和问题 5。

第三步,先将问题 5 暂放置待求解,并用上述同样的方法解得问题 4 的最优解 $x_1=2, x_2=2, f(x)=70$ 。由于最优解已满足整数要求,故不再分枝。此时原问题的目标函数值的新下界值为 70,即

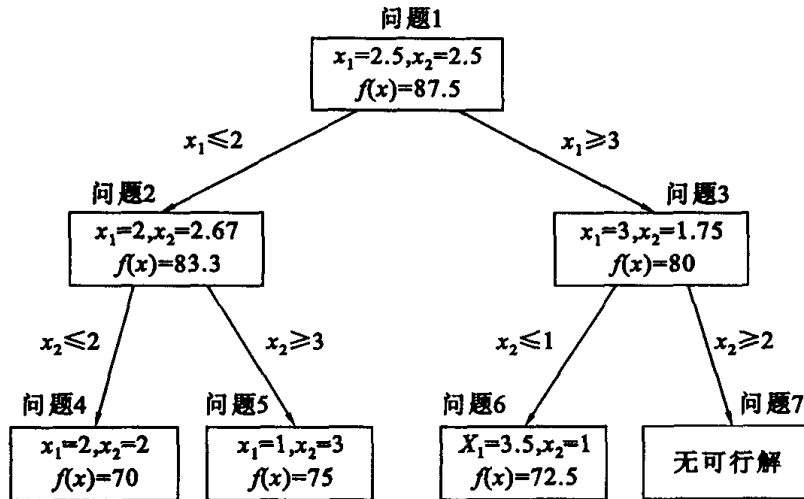


图 3-7 分枝定界法用于生产计划

$$\underline{f}(x) = \max[0, 70] = 70$$

第四步,用同样的方法计算求解的问题 5。得最优解 $x_1=1, x_2=3$, 及 $f(x)=75$; 同样, 由于他们已是整数解, 不必再分枝, 只是修改原问题目标函数的下界值为:

$$\underline{f}(x) = \max[70, 75] = 75$$

第五步, 由问题 2 分枝下来的问题 4、问题 5 的求解即可到此为止。而是进一步求到原放置起来的问题 3, 得最优解

$$x_1=3, x_2=17.5, f(x)=80$$

由于 x_2 不满足整数约束, 且 $f(x)=80 \geq \underline{f}(x)=75$, 所以对问题 3 有必要继续分枝, 其松弛后的子问题为问题 6、问题 7。首先求解问题 6, 得最优解 $x_1=3.5, x_2=1, f(x)=72.5$ 。此时因 $f(x)=72.5 < \underline{f}(x)=75$, 即使继续分枝其最新解也不可能超过当前解的下限值 75, 故不必再分枝。

第六步, 求解问题 7, 由于在 $x_1=3$ 的情况下, 若 $x_2 > 2$ 则超过了可行解域, 故此问题无可行解, 不再分枝。

至此, 所有可分枝的子问题都作了分析和求解, 在所有可行解中, 当前新的下界值所对应的解是现行最优的整数可行解, 也即为原整数规划问题的最优解, 即: $x^*=1, x^*=3$ 最优目标函数值 $f^*(x)=75$ 。

图 3-7 中的每个分枝的地方称为节点。从上述求解过程可知, 再分枝的目标, 借助于不断修改问题目标函数值的上、下界, 从而简化求解过程。分枝定界法也因此而得名。

从上述例子, 不难概括出分枝定界法的五条规律:

(1) 每个松弛问题的最优值均为相应整数规划问题最优值的上界。

(2) 在求解子问题的松弛问题时, 若松弛问题无可行解, 则相应的子问题也无可行解, 此时则不必再分枝, 如图 3-7 的问题 7。

(3) 若松弛问题的解满足整数约束条件, 则此解为相应子问题的最优解。此时不必再分枝, 同时, 若目标函数值大于目前的下界值, 则必须修改下界值。如图 3-7 中的问题 4、问题 5。

(4) 若松弛问题的解不满足整数约束条件, 且目标函数不大于目前的下界值, 则不必再

分枝。如图 3-7 的问题 6。

(5) 若松弛问题的解不满足整数约束条件,且目标函数值大于当前的下界值,则对应的子问题必须进一步分枝,如图 3-7 的问题 3。

必须指出,对上例的子问题 4、5、6,不再分枝,并不是说它们分枝后不可以找到新的整数可行解,而是表明即使找到新的整数可行解,也不会优于目前最优的整数可行解。其目标函数值不会大于目前的下界值。这些可行解已被全部隐含枚举了。所以分枝定界法实质上是一种隐含枚举法。

3.3.4 0—1 规划问题

0—1 规划问题是整数的特殊情况。

【例 3.11】 (装载问题)有一装载设备,它可以载运七类货物,每件的重量和运价不一样,如表 3-15 所示,该运载设备的最大载重量为 35 吨,而每件货物是完整而不可分拆的,所以只能有两种选择,要么装上,要么不装。问应如何装载货物才能使利润最高。

表 3-15 装载问题的数据

货 物	重 量	运 价	运价/单位重量
1	3	12	4
2	4	12	3
3	3	9	3
4	3	15	5
5	15	90	6
6	12	26	2
7	16	112	7

【解】 设各类的情况用 $x_1, x_2, x_3, x_4, x_5, x_6, x_7$, 当 $x_j = 1$ 时,表示货物 i 装上运载设备,当 $x_j = 0$ 时,表示货物 i 未装上运载设备。因此该整数规划属 0—1 规划类型,其数学模型可表示为:

$$\begin{aligned} \max f(x) &= 12x_1 + 12x_2 + 9x_3 + 15x_4 + 90x_5 + 26x_6 + 112x_7 \\ \text{s. t. } &3x_1 + 4x_2 + 3x_3 + 3x_4 + 15x_5 + 12x_6 + 16x_7 \leq 35 \\ &x_j = 0 \text{ 或 } 1 \quad j = 1, 2, \dots, 7 \end{aligned}$$

或表示为:

$$\begin{cases} \max f(x) \\ x \in s_0 \end{cases}$$

如果在上述模型中,把 $x_j = 0$ 或 1 改成:

$$\begin{cases} x_j \leq 1 \\ x_j \geq 0 \end{cases} \quad j = 1, 2, \dots, 7$$

就成为非整数约束的一般线性规划问题,即为原装载问题的一个松弛问题,很容易求解,只要按表 3-15 提供的数据,按“单位重量的价值最大的尽量先取”的原则,显然货物 7 对

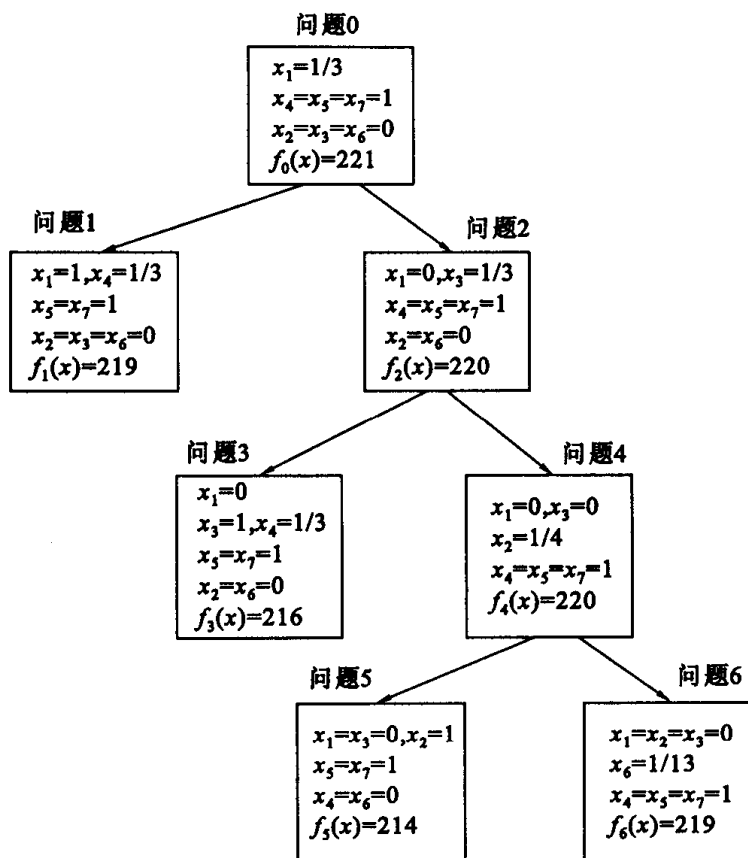


图 3-8 0-1 规划问题

应的单位重量价值最大,所以令 $x_7=1$,接着令 $x_4=x_5=1$,这时已取的货物的重量为:

$$16+15+3=34(\text{吨})$$

还可以装取 1 吨,应取 $x_1=1/3$ (x_1 每件 3 吨),因此这时得解 $x_7=x_5=x_4=1, \alpha_1=1/3, x_2=x_3=x_6=0$ 是该装载问题在松弛条件下的一个最优解。对应的最优值 $f(x)=221$ 。

但因这组解中包含有分数,不全为整数,也就是说,这组解不是问题的可行解。或者说不属于 s_0 ,因此应对 s_0 进行分枝。

分枝可以采取以下的方法。令不是整数的变量 x_1 必须取值 1 或 0,这就把 s_0 化成了两个子集的和,即 $s_0=s_1 \cup s_2$ 。其中 $s_1=\{s_0 \text{ 中满足 } x_1=1 \text{ 的可行解的集合}\}, s_2=\{s_0 \text{ 中满足 } x_1=0 \text{ 的可行解的集合}\}$ 。分枝以后,问题 0 就不再考察了,而改为分别求解问题 1 和问题 2,其方法仍然是求解问题 1 和问题 2 的松弛问题,即求解 $\max_{x \in T_1} f(x)$ 和 $\max_{x \in T_2} f(x)$,从而得问题 1 的解为:

$$\begin{aligned} x_1=x_5=x_7 &=1 \\ x_4=1/3, x_2=x_3=x_6 &=0 \quad f_1(x)=219 \end{aligned}$$

问题 2 的解为:

$$\begin{aligned} x_4=x_5=x_7 &=1, x_3=1/3 \\ x_1=x_2=x_6 &=0, \quad f_2(x)=220 \end{aligned}$$

由于两个问题的最优解都不是原问题的可行解,所以选取其值较大的问题 2,再进一步分为问题 3 和问题 4 两个子问题进行求解,得到图 3-8 的结果,此时仍无整数最优解(问题 3 中 $x_4=1/3$,问题 4 中 $x_2=1/4$)。在枚举树上角可标志最优解属空集 $\emptyset$ 以及最大值为 $-\infty$ 。

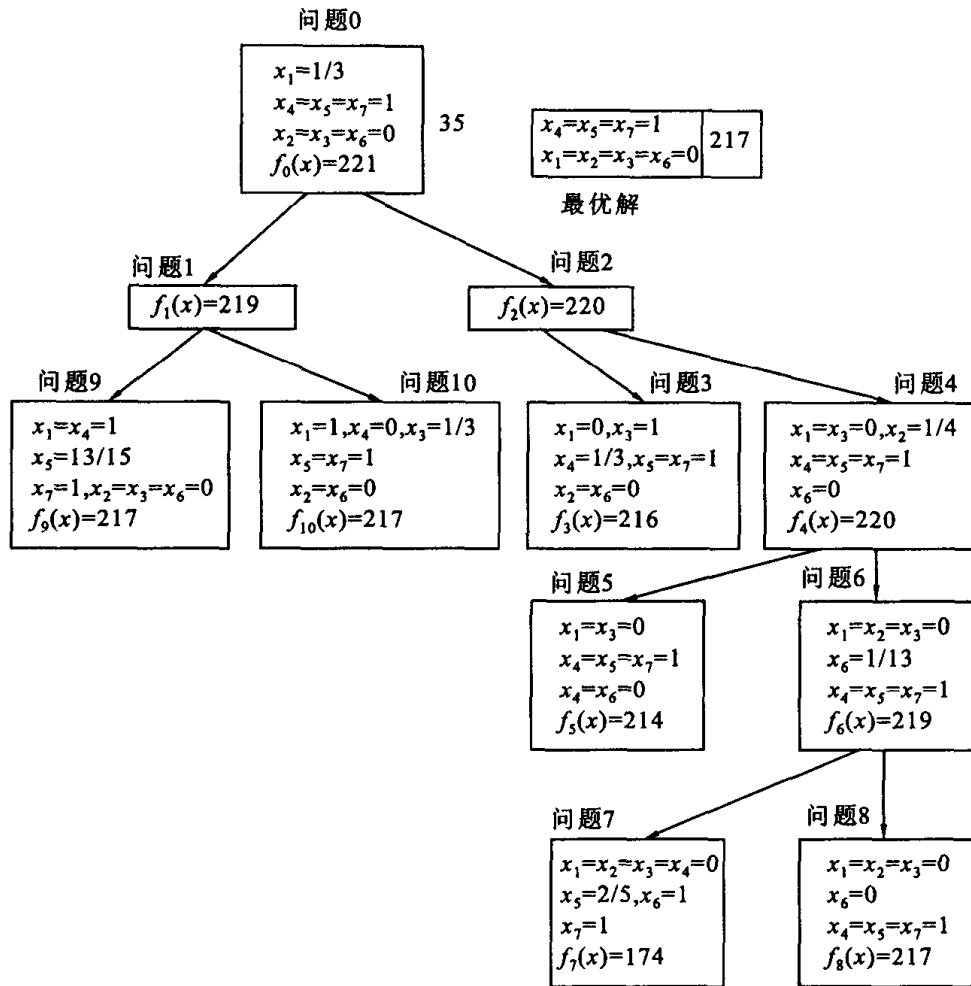


图 3-9

接着再仿照前面的方法,由问题 4 再分成问题 5 和问题 6 两个子问题。并求其松弛问题的最优解,其中问题 5 的最优解 $x_2=x_5=x_7=1, x_1=x_3=x_4=x_6=0, f_5(x)=214$,属于原问题的可行解,在图 3-8 上可以标出具有的最优值及其值,这时把其他未经剪枝的子问题进行一次检查比较,如果发现某个子问题的边界值比现有最优值还小,那么,该子问题可以去掉,即为剪枝。在图 3-8 中,因问题 6、问题 3 和问题 1 所对应的 $f(x)$ 值都大于 214,都不能被剪掉。

接着在上述三个子问题中(6、3、1)取其 $f(x)$ 值较大者(问题 6)继续进行分枝,得到问题 7、问题 8(见图 3-9 所示)。问题 8 的松弛问题的解全是整数,因此是原问题的可行解。得到图 3-9 的现有最优解: $x_4=x_5=x_7=1, x_1=x_2=x_3=x_6=0, f_8(x)=217$,这时问题 3、5、7 可以剪掉,剩下问题 1 的 $f_1(x)$ 值大于问题 8 的 $f_8(x)$ 值,所以继续对问题 1 进行分枝,得到问题 9 和问题 10,它们的松弛问题的解都不是原问题的可行解,并且其最优值都不比问题 8 的大,所以问题 9、10 可以剪掉。至此全部问题都已检查。问题 8 的最优解是原问题的最优解, $f_8(x)$ 为最优值。整个计算枚举树见图 3-9。

3.4 非线性规划

静态规划的数学模型可以是线性的,也可以是非线性的,然而线性是相对而言,非线性才是绝对的,线性规划是非线性规划的特殊情况。

典型的非线性规划,除二次型规划外,还有几何规划。如果目标函数具有变量平方和的形式求极小值,就是最小二乘问题,这也是一种类型的非线性规划。目前比较常用的求解非线性函数极值问题的方法大致分无约束非线性规划和有约束非线性规划。本节对无约束非线性规划问题,将扼要介绍一个最优化的区间消去法与 Fibonacci 法,以及多维最优化的最速下降法。另外将以罚函数法和线性逼近法为例介绍有约束非线性规划的求解。

3.4.1 区间消去法

若目标函数和约束条件不能表示成待定变量的显函数,则需要用数值方法求解问题的最优解。许多工程设计问题具有这一特点。数值法的基本思想是经过一系列的迭代,产生点的序列(即标点列)使之逐步接近最优点,所以这种方法称之为直接搜索法。下面将以区间消去法为例来说明直接搜索法的含义。

设单峰函数 $f(x)$,其最小值点在闭区间 $[a_0, b_0]$ 内,任取两点 λ_1 与 λ_2 ,这时可能出现三种情况(图 3-10)。

① $f(\lambda_1) < f(\lambda_2)$: x^* 必然在 a_0 和 λ_2 区间,因此可将区间 $[\lambda_2, b_0]$ 消去,使搜索范围由 $[a_0, b_0]$ 缩减为 $[a_1, b_1]$,其中 $a_1 = a_0, b_1 = \lambda_2$ (图 3-10a)。

② $f(\lambda_2) < f(\lambda_1)$: x^* 必然在 $[\lambda_1, b_0]$ 闭区间内,而 $[a_0, \lambda_1]$ 区间可消去,其搜索范围缩小成为 $[a_1, b_1]$ 。其中 $a_1 = \lambda_1, b_1 = b_0$ (图 3-10b)。

③ $f(\lambda_2) = f(\lambda_1)$: 此时 x^* 必然在 $[\lambda_1, \lambda_2]$ 区间,而 $[a_0, \lambda_1]$ 及 $[\lambda_2, b_0]$ 可以消去,搜索范围缩小成 $[a_1, b_1]$ 。其中 $a_1 = \lambda_1, b_1 = \lambda_2$ (图 3-10c)。

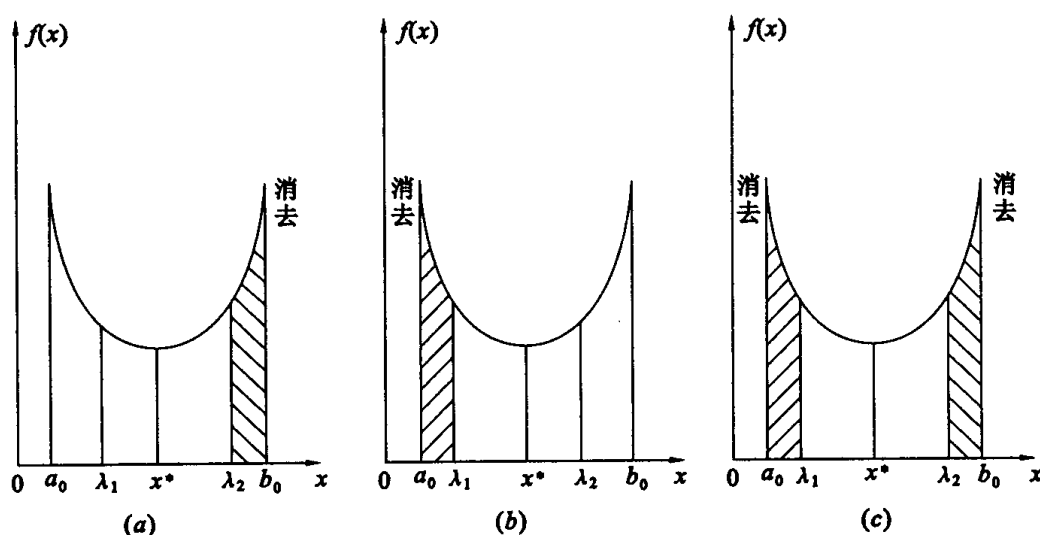


图 3-10

由以上情况可知,在 $[a_1, b_1]$ 中已有一点的函数值计算过,因此只要在 $[a_1, b_1]$ 中再任取一点,计算该点函数值,进行比较,再消去一段区间后得 $[a_2, b_2]$,经过几次迭代,可将搜索区

间减缩为 $[a_n, b_n]$ 。如果 $|b_n - a_n| < \epsilon$, ϵ 为规定误差, 则可得到最优解 $x^* = \frac{a_n + b_n}{2}$ 。

3.4.2 Fibonacci 法

这种方法是按弗波拉契数列的规律进行区间缩减的一种搜索最优点的方法。

下面整数数列称为弗氏数列, 记作:

$$\{F_k\} \quad k = 0, 1, 2, \dots$$

这个数列是由差分方程

$$\begin{cases} F_{k+2} = F_k + F_{k+1} & k \geq 0 \\ F_0 = 1, F_1 = 1 \end{cases} \quad (3.49)$$

的递推关系产生的。因此 $\{F_k\} = 1, 1, 2, 3, 5, 8, 13, 21, \dots$

前后两项之比成为一分数序列

$$1, \frac{1}{2}, \frac{2}{3}, \frac{3}{5}, \frac{5}{8}, \frac{8}{13}, \frac{13}{21}, \dots$$

可以表示成

$$\left\{ \frac{F_k}{F_{k+1}} \right\}$$

从整个数列可以看出从 $k = 2$ 起, 每一项都是前两项之和, 即:

$$\text{或} \quad \begin{cases} F_{k+2} = F_k + F_{k+1} & k > 1 \\ F_{k+1} = F_k + F_{k-1} & k > 1 \end{cases} \quad (3.50)$$

$$\text{或} \quad \frac{F_k}{F_{k+1}} + \frac{F_{k-1}}{F_{k+1}} = 1$$

研究分数列可以发现, 当 $k \rightarrow \infty$ 时, 则有

$$\frac{F_k}{F_{k+1}} = 1 - \frac{F_{k-1}}{F_{k+1}} = 0.618$$

这是一种逐渐逼近 0.618 的规律。

如果我们用费氏数列来搜索一维函数 $f(x)$ 最小值, 其方法是在 $[0, 1]$ 区间选择 λ_1, λ_2 两点, 例如第一次迭代, 取区间两点为 λ_1^1, λ_2^1 , 且令 $(\lambda_1^1, \lambda_2^1)$ 的上标“1”表示第一次迭代)

$$\lambda_1^1 = \frac{F_{k-1}}{F_{k+1}}, \lambda_2^1 = \frac{F_k}{F_{k+1}}$$

显然, 根据上述公式必然有 $\lambda_1^1 + \lambda_2^1 = 1$

同时,

$$\begin{aligned} \lambda_2^1 - \lambda_1^1 &= \frac{F_k - F_{k-1}}{F_{k+1}} = \frac{F_{k-2}}{F_{k+1}} \\ \lambda_1^1 &= 1 - \lambda_2^1 \end{aligned}$$

所以 $[0, \lambda_1^1]$ 与 $[\lambda_2^1, 1]$ 两区间的长度相等, 这时计算 $f(\lambda_1^1)$ 与 $f(\lambda_2^1)$ 数值, 若 $f(\lambda_1^1) \leq f(\lambda_2^1)$ 则可消去 $[\lambda_2^1, 1]$ 区间, 剩下的区间为 $[0, \lambda_2^1]$, 在区间进行第二次迭代(如图 3-11 所示)。

第二次迭代时, 在 $[0, \lambda_2^1]$, 区间取新的两点 λ_1^2, λ_2^2 且令 $\lambda_2^2 = \lambda_1^1 = \frac{F_{k-1}}{F_{k+1}}$, 则 $[\lambda_2^2, \lambda_2^1]$ 区间的

的长度为 $\lambda_2^1 - \lambda_2^2 = \frac{F_k}{F_{k+1}} - \frac{F_{k-1}}{F_{k+1}} = \frac{F_{k-2}}{F_{k+1}}$ 。同时根据同样的对称性的道理, $[0, \lambda_1^2]$ 与 $[\lambda_2^2, \lambda_2^1]$

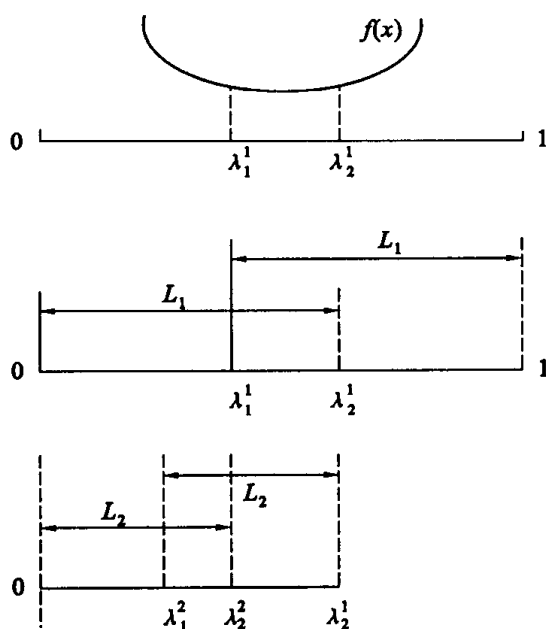


图 3-11 区间迭代

之间应相等。

所以
$$\lambda_1^2 = \frac{F_{k-2}}{F_{k+1}} = \lambda_2^1 - \lambda_2^2$$

$\therefore \lambda_1^1 = \lambda_2^2$

$\therefore f(\lambda_2^2)$ 的数值已计算过, 只需计算 $f(\lambda_1^2)$, 并与 $f(\lambda_2^2)$ 比较值的大小, 使搜索区再次减小, 其长度变为 $L_2 = \frac{F_{k-1}}{F_{k+1}} = \lambda_2^2$, 从而归纳出如下搜索规律:

迭代次数	$k:0$	1	2	$\cdots$	$k-1$	k
搜索区间长度 L_k	1	$\frac{F_k}{F_{k+1}}$	$\frac{F_{k-1}}{F_{k+1}}$	$\cdots$	$\frac{2}{F_{k+1}}$	$\frac{1}{F_{k+1}}$

显然, 若给定进行 k 次迭代以后的搜索区间长度为 δ , 则可计算出所需的迭代的次数 k 。

如: 原区间长度 $L_0 = 1$, 要求 k 次迭代后的搜索区间长度 $\delta = 0.08$, 问迭代次数 k 为多少?

$\therefore L_k = \frac{1}{F_{k+1}} \leq 0.08$

$\therefore F_{k+1} = 1/\delta = 12.5$ 在弗氏数列中 $F_6 = 13$, $\therefore k = 5$

再如: 已知 $F_{20} = 10946$, 求进行第 19 次迭代后的搜索区间的长度。设原区间长度为 1,

则
$$\delta = \frac{1}{F_{20}} = \frac{1}{10946} = 0.000091$$

推广到一般情况, 若原区间为 $[a_0, b_0]$, 则搜索长度为 $L_0 = b_0 - a_0$, 设 n 表示需要计算的函数值的个数, 则 $k = n - 1$ 。应迭代 k 次时, 区间长度 $L_{n-1} = b_{n-1} - a_{n-1}$, 即 $L_{n-1} = L_k = \frac{(b_0 - a_0)}{F_n}$ 。

若在第 k 次迭代时搜索到了最优点, 则必然有:

$$\frac{b_0 - a_0}{F_{k+1}} \leq \epsilon \quad (\epsilon \text{ 为很小的正数})$$

另外,再根据

$$\lambda_1^1 = \frac{F_{k-1}}{F_{k+1}}, \lambda_2^1 = \frac{F_k}{F_{k+1}}$$

则迭代公式为:

$$\lambda_1^1 = a_0 + \frac{F_{n-2}}{F_n}(b_0 - a_0) = b_0 + \frac{F_{n-1}}{F_n}(a_0 - b_0)$$

$$\lambda_2^1 = a_0 + \frac{F_{n-1}}{F_n}(b_0 - a_0) = b_0 + \frac{F_{n-2}}{F_n}(a_0 - b_0)$$

其中

$$\lambda_1^1 + \lambda_2^1 = a_0 + b_0$$

进行几次迭代以后,则

$$\lambda_1^k = b_{k-1} + \frac{F_{n-k}}{F_{n-k+1}}(a_{k-1} - b_{k-1}) \quad (3.51)$$

$$= a_{k-1} + \frac{F_{n-(k+1)}}{F_{n-k+1}}(b_{k-1} - a_{k-1})$$

$$\lambda_2^k = b_{k-1} + \frac{F_{n-k}}{F_{n-k+1}}(a_{k-1} - b_{k-1})$$

$$= a_{k-1} + \frac{F_{n-k}}{F_{n-k+1}}(b_{k-1} - a_{k-1}) \quad (3.52)$$

这种迭代方法的特点是:

- ① 给定试验次数时,可以使区间缩短为最小,所以是最优搜索方法。
- ② 这种方法是序贯试验的方法,它是限定试验次数并每次只做一个实验的最优方法。
- ③ 算法简单,用计算机计算方便,其程序框图如图 3-12。

下例可说明弗氏方法的具体步骤:

【例 3.12】 求 $f(x) = (x-3)^2$ 的极小值点,设区间长度为

$L_0 = [0, 10]$, 要求搜索 3 次, $n = 4$ 。

【解】 由解析法可知极小值点为 $x^* = 3$, 而用弗氏方法迭代, 则为:

第一次迭代: $a_0 = 0, b_0 = 10$, 在 $[0, 10]$ 区间取 λ_1^1, λ_2^1 两点, 则

$$\lambda_1^1 = a_0 + \frac{F_{n-2}}{F_n}(b_0 - a_0) = 4$$

$$\lambda_2^1 = a_0 + \frac{F_{n-1}}{F_n}(b_0 - a_0) = 6$$

其中: 从 F_k 数列可知, F_n, F_{n-1}, F_{n-2} 分别为 5, 3, 2。

则 $f(\lambda_1^1) = 1, f(\lambda_2^1) = 9$, 剩下的区间为 $[0, 6]$, 其长度为 $L_1 = 6$ 。

第二次迭代: $a_1 = 0, b_1 = 6$, 在 $[0, 6]$ 区间取 λ_1^2, λ_2^2 两点, 则有

$$\lambda_1^2 = b_1 + \frac{F_{n-2}}{F_{n-2+1}}(a_1 - b_1) = 2$$

$$\lambda_2^2 = \lambda_1^1 = 4$$

则 $f(\lambda_1^2) = 1, f(\lambda_2^2) = 1$, 两者相等, 若我们取 $a_2 = 2, b_2 = 4$, 即消去 $[0, 2]$ 一段, 这时 $L_2 = 2$, 显然 $2 < x^* < 4$ 。

第三次迭代: 因为 $f(\lambda_1^2) = f(\lambda_2^2)$, 为了使函数值能区别开, 令搜索区间为 $[2, 3.99]$, 则 $f(3.99) = 0.9801$, 显然 $2 < x^* < 3.99$, 这时 $a_2 = 2, b_2 = 3.99, L_2 = 1.99$

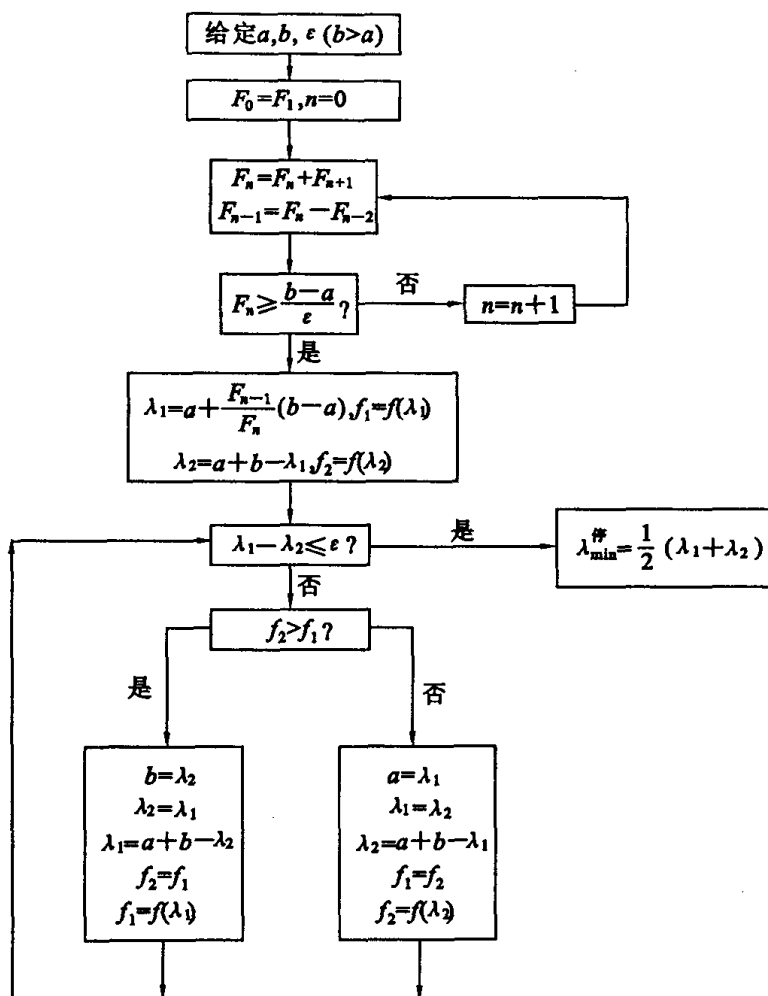


图 3-12 Fibonacci 法框图

$$\lambda_1^3 = b_2 + \frac{F_{n-3}}{F_{n-2}}(a_2 - b_2) = 2.995$$

则

$$\lambda_2^3 = a_2 + \frac{F_{n-3}}{F_{n-2}}(b_2 - a_2) = 2.995$$

∴ 最优点

$$x^* = 2.995 \quad f(x^*) = 0.000025$$

3.4.3 最速下降法(最优梯度法)

最速下降法是应用目标函数的负梯度方向作为每步迭代的搜索方向,在计算时每步都是沿负梯度方向取最优步长,因而这种方法也称为最优梯度法。应用这种方法,使得目标函数只在开头几步下降最快。最速下降法比较简单,但它只是用一阶梯度的信息以确定下一步搜索方向,因此收敛较慢,而且越是接近极值点,收敛越慢,这是它的主要缺点。后来又发展了许多收敛较快的算法。如共轭梯度法、边尺度法等,但最速下降法是这些最优化方法的基础,所以这里做重点介绍。

(1) 梯度和梯度方向

一个非线性函数 $f(X)$ 的梯度定义为

$$\nabla f(X) = \frac{\partial f}{\partial x} = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right]^T \quad (3.53)$$

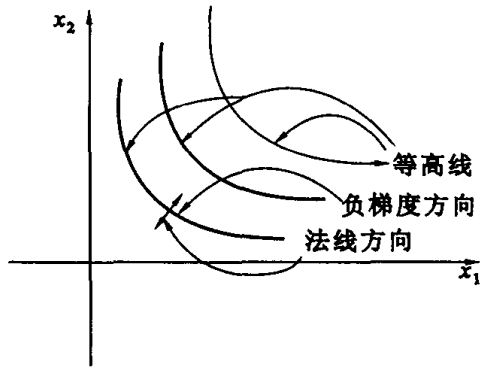


图 3-13 函数的负梯度方向

这是一个 n 维向量。

最速下降法是以负梯度方向作为搜索方向的, 令梯度向量 $P = \nabla f(X)$, 梯度方向的单位向量为

$$S = \frac{\nabla f(X)}{\|\nabla f(X)\|} = \frac{P}{\|\nabla f(X)\|} \quad (3.54)$$

梯度方向 S 也就是函数的法线方向。

图 3-13 的负梯度方向为 $-S$ 。

$\|\nabla f(X)\|$ 为梯度 $\nabla f(X)$ 的范数:

$$\|\nabla f(X)\| = \sqrt{\left(\frac{\partial f}{\partial x_1}\right)^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 + \cdots + \left(\frac{\partial f}{\partial x_n}\right)^2} \quad (3.55)$$

因为梯度与函数的等高线是互相垂直的, 因此沿梯度方向 $f(X)$ 值增长最快, 反之, 沿负梯度方向 $f(X)$ 值下降最快。

【例 3.13】 $f(X) = -2x_1 + x_2 = [-2, 1] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = c^T X$, 求梯度方向。

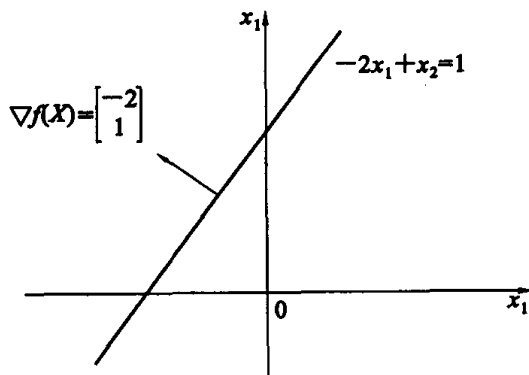
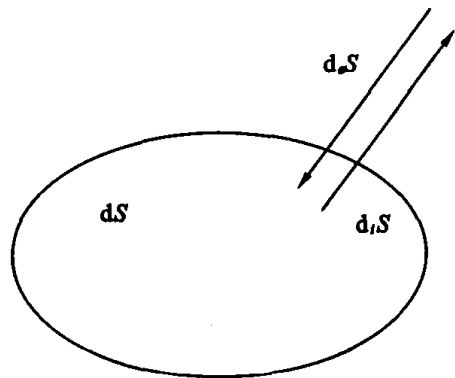
【解】 梯度: $\nabla f(X) = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$, 范数: $\|\nabla f(X)\| = \sqrt{(-2)^2 + 1^2} = \sqrt{5}$, 梯度方向的单位向量

$$S = \frac{\nabla f(X)}{\|\nabla f(X)\|} = \frac{1}{\sqrt{5}} \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

如图 3-14 所示, 梯度垂直于直线函数 $f(X)$ 。

【例 3.14】 求函数 $f(x_1, x_2) = x_1^2 + 2x_1x_2 + 3x_2^2$ 在 $X_0 = [2, 1]^T$ 点的梯度。

【解】 $\nabla f(X) = \begin{bmatrix} 2x_1 + 2x_2 \\ 2x_1 + 6x_2 \end{bmatrix}$, 在 $X^0 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ 点, $\nabla f(X)|_{x^0} = \begin{bmatrix} 6 \\ 10 \end{bmatrix}$, $\|\nabla f(X_0)\| = \sqrt{136}$, 梯度向量 $P = \begin{bmatrix} 6 \\ 10 \end{bmatrix}$ 。图 3-15 中画出了该梯度方向量, 梯度方向单位向量为 $S = \frac{1}{\sqrt{136}} \begin{bmatrix} 6 \\ 10 \end{bmatrix}$ 。

图 3-14 $f(X) = -2x_1 + x_2$ 图 3-15 $f(X) = x_1^2 + 2x_1x_2 + 3x_2^2$

推广到一般非线性函数,某点梯度方向与该点切线相垂直,在 x^k 点,函数梯度为 $\nabla f(X^k)$,该点的单位梯度方向为

$$S^k = \frac{\nabla f(X^k)}{\|\nabla f(X^k)\|} \quad (3.56)$$

不同点的梯度方向是不相同的。

下面列出几种常用函数的梯度公式:

$$f(X) = c^T X : \nabla f(X) = c$$

$$f(X) = X^T X : \nabla f(X) = 2X$$

$$f(X) = \frac{1}{2} X^T Q X$$

Q 为实对称矩阵:

$$\nabla f(X) = QX,$$

$$f(X) = K + C^T X + \frac{1}{2} X^T Q X$$

C 为常向量, K 为常数, Q 为实对称阵:

$$\nabla f(X) = C + QX$$

(2) 最优梯度

如果函数的等高线为圆,则半径与各等高线是垂直的,也即半径方向就是梯度方向,以任一点 X^0 出发(图 3-16)沿负梯度方向,选择合适的步长,一步可达最优点,第 $k+1$ 次迭代时, $X^{k+1} = X^k + \lambda \nabla f(X^k) = X^k + \lambda P^k$,上式表示从 X^k 点出发沿梯度方向,取步长参数为 λ ,下一步可达到 X^{k+1} (求极小值时 P^k 为负,求极大值时 P^k 为正)。 X^{k+1} 式也可写成

$$X^{k+1} = X^k + a \frac{\nabla f(X^k)}{\|\nabla f(X^k)\|} = X^k + a S^k \quad (3.57)$$

步长 $a = \lambda \|\nabla f(X^k)\|$,求极小值, $a S^k$ 为负。

图 3-16 沿圆的负梯度方向一步达最优

S^k 为从 X^k 出发的单位梯度方向,称为搜索方向。

对于任意二次函数,设 $f(X) = x_1^2 + 2x_2^2$,如果选固定步长参数,例如 $\lambda = 4$,从 $X^0 = [4, 4]^T$ 开始,走了 3 步以后目标函数值反而上升(图 3-17),这一例子启发我们,沿负梯度方向搜索函数最优值,当远离最优点时,步长要大,当接近最优点时,步长应当减小。

① 最速下降法(最优梯度法)的基本概念

沿着负梯度方向选择步长时搜索步数为最少的方法称为最速下降法(最优梯度法)。这种方法要求每步计算一次最优步长,使每走一步目标函数下降最多。显然用最优梯度法计算,工作量要大,比固定步长增加了一维寻优的过程。

由(3.57)式可得

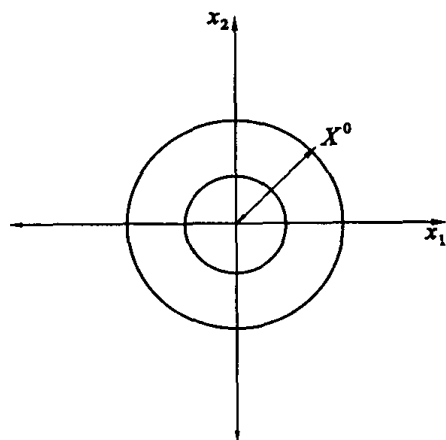
$$f(X^{k+1}) = f(X^k + a S^k) \quad (3.58)$$

沿着 S^k 方向求使 $f(X^{k+1})$ 为最小的步长 a^k 。

$$\min f(X^{k+1}) = \min f(X^k + a S^k) = f(X^k + a^k S^k)$$

$$X^{k+1} - X^k = \Delta X^k = a^k S^k \quad (3.59)$$

求最优步长时,是对函数 $f(X^{k+1})$ 求极小,步长 a 为变量,因此这时一维搜索问题,可以用一



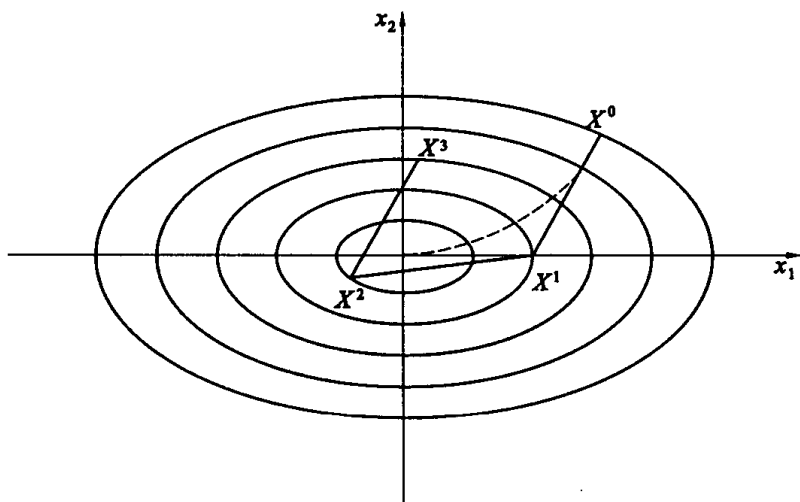


图 3-17 任意二次函数固定步长沿负梯度方向迭代

维搜索方法计算。

【例 3.15】 函数 $f(X) = x_1^2 + 25x_2^2$, 求 $\min f(x)$ 。

【解】 用最速梯度法求最优步长。

$$\begin{aligned} \text{设} \quad X^0 &= [2, 2]^T, \nabla f(X^0) = \begin{bmatrix} 2x_1 \\ 50x_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 100 \end{bmatrix}, \\ f(X^0) &= 104 \quad X^1 = X^0 + aS^0 \\ S^0 &= \frac{\nabla f(X^0)}{\|\nabla f(X^0)\|} = \frac{[4, 100]^T}{\sqrt{4^2 + 100^2}} = [0.03997, 0.999]^T \\ f(X^1) &= (2 + 0.04a)^2 + 25(2 + a)^2 \end{aligned}$$

显见这是单变量函数, 现在用解析法求极小的条件:

$$\frac{df(X^1)}{da} = 0,$$

$$2(2 + 0.04a)0.04 + 50(2 + a) = 0$$

表 3-16 列出迭代三步的计算结果。

表 3-16

	x_1	x_2	$f(X^k)$
初始点	2	2	104
$a_0 = 2.003$	1.92	-0.003	3.69
$a^1 = 1.85$	0.07	0.07	0.13
$a^2 = 0.07$	0.07	0	0.0049

最优步长, 由 $a^* = a^0 = 2.003$, 由 $X^1 = X^0 - a^0 S^0$

可得 $X = [2 - 2.003 \times 0.44, 2 - 2.003 \times 1]^T = [1.92, -0.003]^T$

可见, 从 $[2, 2]^T$ 出发, 沿 $-S^0$ 方向走一步, 选最优步长 $a^0 = 2.003$, 则达到新的点 X^1 , $f(X^1) = 3.69$, 目标函数从 $f(X^0) = 104$ 沿 $-S^0$ 方向走一步降到 $f(X^1) = 3.69$, 下降甚速。

按固定步长 $a = 1$, 则用下降法迭代五步的计算结果如表 3-17:

表 3-17

k	x_1	x_2	$\nabla f(X^k)^T$	$f(X^k)$
0	2	2	[4, 100]	104
1	1.96	1	[3.92, 50]	28.8416
2	1.88	0	[3.76, 0]	3.5344
3	0.88	0	[1.76, 0]	0.7744
4	-0.12	0	[-0.24, 0]	0.0144
5	0.88	0	[1.76, 0]	0.7744

可见按固定步长,若步长 a 取得小,收敛慢,计算次数多,但若 a 取得大,则在极值点附近来回振荡。而最优梯度法收敛较快,三步即接近极值点。

② 最优步长的确定

对于任意函数 $f(X)$ 在点 X^k 展开成泰勒级数,取二次近似可得:

$$f(X^{k+1}) = f(X^k) + \nabla^T f(X^k) \Delta X + \frac{1}{2} \Delta X^T A \Delta X \quad (3.60)$$

$$\Delta X = X^{k+1} - X^k$$

式中:

$A = \nabla^2 f(X^k)$ 为海赛(Hesse)阵

令

$X^{k+1} = X^k + aS^k$, S^k 为单位向量

若

$$\Delta X = aS^k$$

$$\text{则得} \quad f(X^k + aS^k) = f(X^k) + \nabla^T f(X^k) \cdot aS^k + \frac{1}{2} (aS^k)^T A \cdot aS^k$$

上式对 a 求导,并令其为零,即

$$\frac{df(X^k + aS^k)}{da} = 0, \text{得}$$

$$\nabla^T f(X^k) S^k + (S^k)^T A S^k a = 0$$

$$\therefore \text{获得最优步长为:} \quad a^* = a^k = -\frac{\nabla^T f(X^k) S^k}{(S^k)^T A S^k} \quad (3.61)$$

【例 3.16】 $f(X) = \frac{1}{2} X^T Q X$, 求最优步长。

【解】 $\nabla f(X^k) = QX^k$, $\nabla^2 f(X^k) = Q$, $S^k = \frac{QX^k}{\|QX^k\|}$

将上面参数代入(3.61)式,得最优步长:

$$a^k = \frac{-(QX^k)^T (QX^k)}{(QX^k)^T A (QX^k)} \cdot \|QX^k\|$$

或

$$a^k = -\frac{(X^k)^T Q^2 X^k}{(X^k)^T Q^3 X^k} \cdot \|QX^k\|$$

如果在(3.60)式中令 $X^{k+1} = X^k + \lambda P^k$, 代入 $P^k = \nabla f(X^k)$, 则

$$f(X^{k+1}) = f(X^k) + (P^k)^T \lambda P^k + \frac{1}{2} (\lambda P^k)^T A (\lambda P^k)$$

$$\frac{df(X^{k+1})}{d\lambda} = 0, (P^k)^T \lambda P^k = -\lambda^* (P^k)^T A P^k$$

的最优步长参数为:

$$\lambda = \frac{-(P^k)^T P^k}{(P^k)^T A (P^k)} = -\frac{\nabla^T f(X^k) \nabla f(X^k)}{\nabla^T f(X^k) A \nabla f(X^k)}$$

【例 3.17】 $f(X) = x_1^2 + 2x_1x_2 + 2x_2^2 = \frac{1}{2}X^T QX$, $Q = \begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix}$, $X^k = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$, 求 λ^* 。

【解】 $\nabla f(X^k) = \begin{bmatrix} -2 \\ -6 \end{bmatrix}$, 从 X^k 到 X^{k+1} 的最优步长参数为:

$$\lambda^* = \frac{[-2, -6] \begin{bmatrix} -2 \\ -6 \end{bmatrix}}{[-2, -6] \begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} -2 \\ -6 \end{bmatrix}} = -\frac{40}{200} = -0.2$$

3.4.4 罚函数法

罚函数法是应用最广泛的一种求解非线性规划的数值解法。它的实质是将有约束的最优化问题通过罚因子的选择变为一系列求罚函数的极小值,从而将原问题转化为求解一系列无约束极值的问题,所以可称为序贯无约束极小化方法,简称 SUMT。

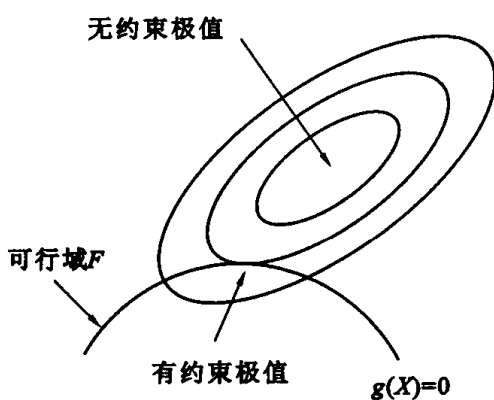


图 3-18 有等式约束的最优解

设最优化问题为: $\min f(X)$

$$g_i(X) = 0, i = 1, 2, \dots, m$$

图 3-18 表示只有一个等式约束时的最优解,可行域就是约束方程 $g(x) = 0$ 。

罚函数是将原目标函数 $f(X)$ 增广为一个新的函数,称为罚函数 P 。即

$$P(X, M_K) = f(X) + M_K \sum_{i=1}^m [g_i(X)]^2 \quad (3.62)$$

其中 M_K 是一个很大的正数,称为罚因子或罚参数;当 $g_i(X) \neq 0$ 即不满足约束时,罚函数的值 $P(X, M_K)$ 将很大,只有当 $g_i(X) = 0$ 时,罚函数 P 才等于原函数 $f(X)$,因此罚因子 M_K 的引入,相当于对不遵守等式约束的惩罚。

(3.62) 式中 $M_K \sum_{i=1}^m [g_i(X)]^2$ 称为罚项,如果对不同的约束方程加不同的权,则(3.62)式中的罚项应改写为

$$\sum_{i=1}^m M_i [g_i(X)]^2$$

式中 M_i 表示对第 i 个约束所加的权因子。(罚因子)

(3.62) 式中 $P(X, M_K)$ 是约束函数与目标函数 $f(X)$ 的组合,故称为组合函数,求 $P(X, M_K)$ 的极小值是一个无约束非线性最优化问题:

$$\min P(X, M_K) = \min \{ f(X) + M_K \sum_{i=1}^m [g_i(X)]^2 \} \quad (3.63)$$

它等价于求原函数 $f(X)$ 在等式约束 $g(X) = 0$ 下的非线性规划问题。 $P(X, M_K)$ 的最优点与 M_K 有关, 记作 $X_K^* = X^*(M_K)$ 。这样, 每一步迭代选择一个罚因子, M_K 表示第 K 步迭代所选的罚因子, 而 X_K^* 则表示第 K 步迭代时所得无约束非线性规划问题 $\min P(X, M_K)$ 的最优解。

显然, M_{K+1} 应大于 M_K , 即随着迭代过程的进展, 罚因子越取越大, 迫使 $g_i(X)$ 趋近于零, 才能使 $P(X, M_K)$ 为极小, 最后 X_K^* 收敛于 $f(X)$ 的最优解, 这时一定满足 $g_i(X) = 0$ 的约束条件。

为了说明罚函数的实质, 可以举一个例子, 并用解析法求解。

【例 3.18】 $\min f(X) = x_1^2 + x_2^2, x \in E^2$

$$g(X) = x_2 - 1 = 0$$

【解】 构造罚函数 $P(X, M_K) = x_1^2 + x_2^2 + M_K(x_2 - 1)^2$

用解析法求上式极小值

$$\frac{\partial P}{\partial x_1} = 0, 2x_1 = 0, x_1^* = 0$$

$$\frac{\partial P}{\partial x_2} = 0, 2x_2 + 2M_K(x_2 - 1) = 0$$

$$x_2^*(M_K) = \frac{M_K}{1 + M_K}$$

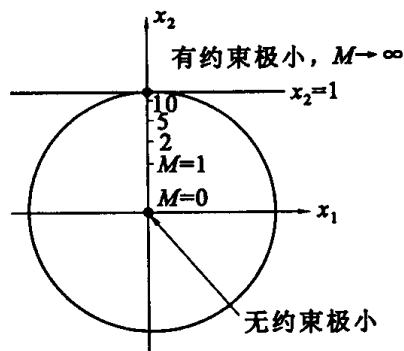


图 3-19 罚函数法

上式随罚因子 M_K 选取不同 $x_2^*(M_K)$ 而不同, 表 3-18 列出了按上式计算的结果: $M_K = 0$ 的情况, 实际上就是求 $f(X)$ 的无约束极值, 当 M_K 增加时, 最优解 x_2^* 以非线性形式向约束方程 $x_2 - 1 = 0$ 移动(图 3-19)。当 $M \rightarrow \infty$ 时 $x_2^* \rightarrow 0$ 即满足等式约束 $x_2 - 1 = 0$, 这时 $f(X^*) = P(X^*) = 1$, 即 $P(X, M_K)$ 收敛 $f(X)$ 在等式约束下的最优点。

表 3-18

M_k	0	1	2	5	10	...	∞
$x_2^*(M_k)$	0	0.5	2/3	0.833	0.91	...	1

另外, 应用罚函数法求解不等式约束的非线性规划问题, 有外点法和内点法, 这里不一一介绍。

3.4.5 线性逼近法

设目标函数是非线性的, 约束方程为线性不等式, 则数学模型为:

$$\left. \begin{array}{l} \min f(X) \quad X \in E^n \\ AX \geq B \quad X \geq 0 \end{array} \right\} \quad (3.64)$$

约束条件构成的可行域 F 是凸多面体, 其顶点为有限个, 最优解可能在边界上。

设 X^0 是某个起始点, 在 $X^0 \in F$ 这一点将 $f(X)$ 用泰勒级数展开取线性近似, 即用线性函数 $L(X)$ 逼近 $f(X)$

$$L(X) = f(X^0) + \nabla^T f(X^0)(X - X^0)$$

$$= f(X^0) - \nabla^T f(X^0) \cdot X^0 + \nabla^T f(X^0) X = K + C^T X \quad (3.65)$$

式中 $K = f(X^0) - \nabla^T f(X^0) \cdot X^0$ 为常数

$$C = \nabla f(X^0)$$

求线性函数 $L(X)$ 的最优解作为 $f(X)$ 的最优解的近似, 这相当于求 $\Delta^T f(X^0) X$ 的最优解, 于是非线性最优化问题在 X^0 这一点变成线性规划问题, 其最优解 Y^0 一定在约束凸多面体的顶点上。

【例 3.19】

$$\min f(X) = 4x_1^2 + (x_2 - 2)^2$$

$$-2 \leq x_1 \leq 2, -1 \leq x_2 \leq 1$$

设初始点

$$X^0 = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$$

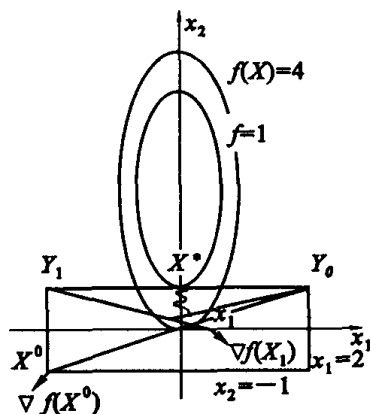


图 3-20 线性逼近法

【解】 ① 过 X^0 做切面平 $\nabla^T f(X^0)(X - X^0)$, 由线性规划定理, 可知本步最优解 Y^0 一定在约束四边形的顶点 $Y^0 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$; 如图 3-20 所示。

② 由下列判别式判别 X^0 是否为非线性函数的极小值,

$$\text{若 } \nabla^T f(X^0)(Y^0 - X^0) = 0$$

$$\text{或 } |\nabla^T f(X^0)(Y^0 - X^0)| \leq \epsilon$$

$$\text{则 } \nabla^T f(X^0)Y^0 = \nabla^T f(X^0)X^0$$

所以 X^0 是非线性规划最优解, 停止计算。

若 $\nabla^T f(X^0)(Y^0 - X^0) \neq 0$, 则一定小于零, 这是因为:

$$\nabla^T f(X^0)Y^0 = \min \nabla^T f(X^0)X, X^0 \text{ 在可行解域内,}$$

∴

$$\nabla^T f(X^0)X^0 > \nabla^T f(X^0)Y^0$$

$$\text{在本例, } Y^0 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, X^0 = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$$

显然

$$\nabla^T f(X^0)(Y^0 - X^0) = [8x_1, 2(x_2 - 2)] \begin{bmatrix} 4 \\ 2 \end{bmatrix}_{x=x^0} = [-16, -6] \begin{bmatrix} 4 \\ 2 \end{bmatrix} < 0$$

③ 从 X^0 沿 $Y^0 - X^0$ 方向走一步, 使 a_0 最优, $a_0 \in [0, 1]$ 。

$$X^1 = X^0 + a_0(Y^0 - X^0) \in F$$

取 a 最优是要在 X^0 和 Y^0 间求 $\min f(X)$, 即

$$\min \varphi(a) = f[X^0 + a(Y^0 - X^0)]$$

在本例中:

$$X^0 + a(Y^0 - X^0) = \begin{bmatrix} -2 \\ -1 \end{bmatrix} + a \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix} - \begin{bmatrix} -2 \\ -1 \end{bmatrix} \right\} = \begin{bmatrix} 4a - 2 \\ 2a - 1 \end{bmatrix}$$

$$\varphi(a) = 4(4a - 2)^2 + (2a - 1)^2$$

$$\frac{d\varphi(a)}{da} = 0 \quad \therefore a_0 = \frac{19}{24} = 0.56, \text{求得}$$

$$X^1 = \begin{bmatrix} -2 \\ -1 \end{bmatrix} + 0.56 \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 0.24 \\ 0.12 \end{bmatrix}$$

④ 在 X^1 点作 $f(X)$ 的线性近似函数,求得第二次线性规划,其最优解一定在凸多面体另一个顶点上,设为 Y^1 ,本例中 $Y^1 = [-2, 1]^T$,即

$$\min \nabla^T f(X^1)X = \nabla^T f(X^1)Y^1$$

再用判别式判别 X^1 是否已达最优。

重复以上步骤,最后总可以收敛到 $X^* = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ 即 $f(X^*) = 1$

由上例可以总结出一般情况下迭代公式为:

$$X^{K+1} = X^K + a_K(Y^K - X^K) \quad (3.66)$$

a^K 取最优解使 $f(X^{K+1})$ 最小。 $f(X^{K+1}) < f(X^K)$

$$\min \nabla^T f(X^K)X = \nabla^T f(X^K)Y^K$$

其中 $f(X^{K+1}) < f(X^K)$

不断迭加,可使 $\nabla^T f(X^K)(Y^K - X^K) \rightarrow 0 \quad X^K \rightarrow X^*$

线性逼近法计算程序为:

① 设允许误差 $\epsilon > 0$,顶点 $X^0 \in F$,取 $K = 0$ 。

② 求线性规划问题 $\min \nabla^T f(X^K)(Y^K - X^K)$ 的最优解 $Y^K \in F$ 。

③ 检验是否满足 $|\nabla^T f(X^K)(Y^K - X^K)| \leq \epsilon$ 。

若满足则已得 $X^K = X^*$,停止计算,否则转 ④。

④ 求单变量极值: $\min f[X^K + a(Y^K - X^K)]$ 的最优解 $a^K, 0 < a^K \leq 1$ 。

⑤ 令 $X^{K+1} = X^K + a_K(Y^K - X^K), K = K + 1$,转 ②。

3.5 动态规划

将系统运行过程分为若干相继的阶段,而在每个阶段都要做出决策的过程,就叫做多段决策过程。多段决策过程的每一段的结束状态,就是下一段的初始状态。因此,如果从某一段看决定的策略是最优,对于下一段未必是有利的,所以多段决策过程的策略选定,要通盘考虑整体规划。选定一个策略使问题的目标值达到最优,这就是多段决策过程的最优问题而且可以表示为数学规划问题,同时用动态规划来解决这类问题往往更有效。动态规划是研究多段决策而提出来的一种数学方法,它的中心思想是所谓的“最优性原理”,这个原理归结为用一个基本递推关系式,从整个过程的终点出发,由后向前,使过程连续地转移,一步一步地推到始点,找到最优解。

3.5.1 一个多段决策问题

首先用一个简单的例子来说明动态规划的基本思想。

从 S 点要铺设一条天然气管道到 F 点,中间必须经过三个中间站(如图 3-21)。第一个中间站可以从 $\{P_1, Q_1\}$ 两点中任选一点;第二站、第三站可以分别在 $\{P_2, Q_2\}, \{P_3, Q_3\}$ 中任选一点。管道相连的两站之间的距离,用图 3-21 上连线边上的数字表示。现在要求选择一条由 S 到 F 的铺管线路,使总的距离为最短。

可以先用穷举法来求解:画出 8 个可能选择的路线,把每两点间的距离都加起来,就算出了每条线路的距离。每条线路要相加 3 次,总共需要相加 24 次。一般说来,如果不是 4 段而

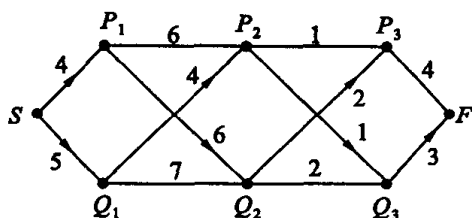


图 3-21 多段决策问题

是 n 段,则需要相加 $(n-1) \cdot 2^{n-1}$ 次。

让我们来考虑另一种寻找从 S 到 F 的最短路线的方法。可以从最后一段开始,在两个点 P_3 与 Q_3 算出从该点到 F 的可能的最小“代价”(即距离),代价用 V 表示。实际上,在本例的最后一段(第 4 段)没有选择余地;从 P_3 点起始的代价,记作 $V_p(4)$,为 4,而 $V_q(4)$ 为 3。

其次,考虑第 3 段的决策,在 P_2 点有两个方案可供选择:

路径 $P_2 P_3 F, V = 1 + V_p(4) = 5$

路径 $P_2 Q_3 F, V = 1 + V_q(4) = 4$

后一条路径更快些,因此计算:从 P_2 到 F 的最小代价为 $V_p(3) = 4$,用同样的方法,可以算出 $V_q(3) = 5$ 。以此类推将各点上的最小代价综合起来,排列如下:

$$V_p(2) = 10, V_p(3) = 4, V_p(4) = 4$$

$$V_q(2) = 8, V_q(3) = 5, V_q(4) = 3$$

最后计算第一段的选择。

路径 $SP_1 F \quad V = 4 + V_p(2) = 14$

路径 $SQ_1 F \quad V = 5 + V_q(2) = 13$

因此选择经过 Q_1 ,即采取 Q 决策,整个问题的最小代价是 $V_1(1) = 13$ 。注意对于每个最初决策来说,我们并不需要知道其后的所有决策,只须知道最初的两个可能决策之后的最优决策的代价就够了。

上述计算,除了一些决策的比较之外只需相加 10 次,对于 N 段来说,则需要相加 $4(N-2) + 2$ 次,如果在 10 段情况下,用穷举法需要相加 4608 次,而上述方法只须相加 34 次,因此这种方法对于处理多段决策过程是很有用的。

3.5.2 动态规划的基本方程,最优性原理

首先介绍以下几个术语:

① 状态变量,用 y_k 表示第 $k+1$ 段开始时的过程状态,称为状态变量。显然,第 $k+1$ 段終了时的状态即第 $k+2$ 段开始时的状态。应记作 y_{k+1} 。对第 $k+1$ 段而言, y_k 称为输入状态, y_{k+1} 为输出状态。

② 决策(或控制)变量:用 u_{k+1} 表示第 $k+1$ 段所采用的决策(或控制),称为决策变量,如上例中的第 $k+1$ 段所用的 P 或 Q 决策。

③ 状态转移方程:就 $k+1$ 段而言,其输出变量 y_{k+1} 显然与输入变量 y_k 及所采取的决策变量 u_{k+1} 值有关。一般情况下,令 u_{k+1} 代表第 $k+1$ 段的可能决策集合(即上例中两个可能决策)中的一个。应该有某种关系式来规定由决策 u_{k+1} 使状态 y_k 下一个状态 y_{k+1} ,这种关系式可记成

$$y_{k+1} = t_{k+1}(y_k, u_{k+1})$$

称为第 $k+1$ 段的转移方程。

④ 阶段效应:第 $k+1$ 段的输入状态 y_k 和决策变量 u_{k+1} 不仅决定了该段的输出状态 y_{k+1} ,而且对整个问题的代价函数值 V 也有影响,即给定了 y_k 与 u_{k+1} 后就有确定的代价值付出。例如上例中在每个点上由于采取决策 P 或 Q 就将增加一段铺管距离,这样对于第

$k+1$ 段

$$\begin{aligned} & y_k \text{ 与 } u_{k+1} \diamond (y_k, u_{k+1}) \\ & = r_{k+1}(y_k, u_{k+1}) + V'_{k+1}(y_{k+1}) \end{aligned} \quad (3.67)$$

其中 $r_{k+1}(y_k, u_{k+1})$ 就是 $k+1$ 段由于 y_k 及 u_{k+1} 而付出的代价, 称为阶段效应。而 $V'_{k+1}(y_{k+1})$ 是 $V_{k+1}(y_{k+1}, u_{k+1})$ 中当 u_{k+2} 为最优决策时的代价函数, 即

$$V'_{k+1}(y_k, u_{k+1}) = V_{k+1}(y_{k+1}, u_{k+2}(\text{最优}))$$

在明确了上述术语的含义后, 将问题的求解过程用图 3-22 表示如下:

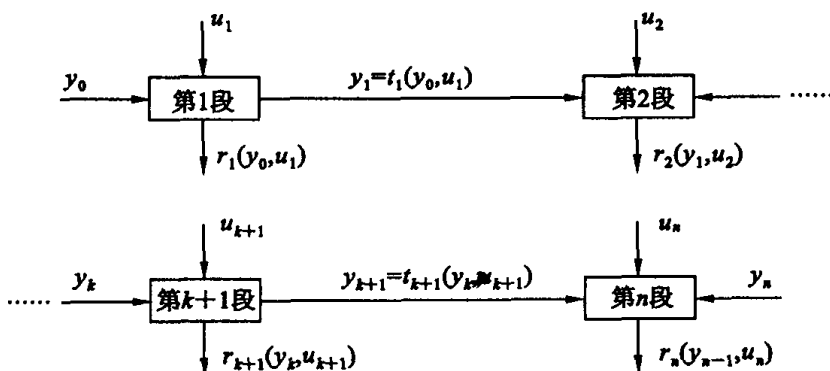


图 3-22

实际上, 是用回代过程进行求解, 即按倒过来的顺序由后面向前推算, 其具体计算步骤为:

第一步: 计算 n 段的最优 $V'_{n-1}(y_{n-1})$, 即

$$\text{opt } V_{n-1} = r_n(y_{n-1}, u_n) \quad (3.68)$$

这里 $V_n(y_n)$ 为零, 而 y_{n-1} 为参数求出 u_n 的最优解及 V_{n-1} 的最优值, 记为

$$u'_n = u_n(y_{n-1}), V'_{n-1} = V_{n-1}(y_{n-1})$$

第二步: 计算 $n-1$ 段的最优 $V'_{n-2}(y_{n-2})$, 即

$$\begin{aligned} \text{opt } V_{n-2} &= r_{n-1}(y_{n-2}, u_{n-1}) + V'_{n-1}(y_{n-1}) \\ &= r_{n-1}(y_{n-2}, u_{n-1}) + V'_{n-1}(t_{n-1}(y_{n-2}, u_{n-1})) \end{aligned} \quad (3.69)$$

其中 y_{n-2} 为参数, 求出 u_{n-1} 的最优解及 V_{n-2} 的最优值:

$$u'_{n-1} = u_{n-1}(y_{n-2}), V'_{n-2} = V_{n-2}(y_{n-2})$$

如此下去, 计算第 k 段的一般公式是

$$\begin{aligned} \text{opt}_{(u_k)} V_{k-1} &= r_k(y_{k-1}, u_k) + V'_k(y_k) \\ &= r_k(y_{k-1}, u_k) + V'_k(t_k(y_{k-1}, u_k)) \end{aligned} \quad (3.70)$$

其中 y_{k-1} 为参数, 求 u_k 的最优解与 V_{k-1} 的最优值:

$$u'_k = u_k(y_{k-1}), V'_{k-1} = V_{k-1}(y_{k-1}) \quad (3.71)$$

最后一步, 第 n 步, 计算第 1 段的公式可以求出

$$u'_1 = u_1(y_0), V'_0 = V_0(y_0)$$

将初始状态 y_0 代入, 求出最优解 $u'_1 V'_0$ 。此时 V'_0 即为问题的最优值, 而 u'_1 为最优决策中的第一个分量, 将 y_0, u'_1 代入 $t_1(y_0, u'_1)$ 求出 y_1 , 再将 y_1 代入 $u_2(y_1)$ 求出 u'_2 , 这是最优决策中第二个分量, 依次下去直到最后求 u'_n , 就得到了最优决策 $(u'_1, u'_2, \dots, u'_n)$ 。

整个计算过程可以表示如下(图 3-23)。

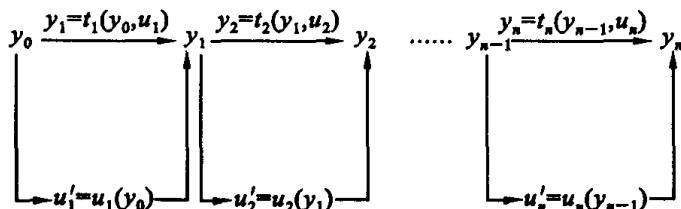


图 3-23

下面取第 k 段作进一步分析。

由 (3.69) 式看：

$$\underset{(u_k)}{\text{opt}} V_{k-1} = r_k(y_{k-1}, u_k) + V'_k(y_k)$$

其中 y_{k-1} 是输入状态, y_k 是输出状态, 这里 y_k 也可以看成是由 y_0 与前面的决策 $u_1, u_2, \dots, u_k$ 所造成的状态, 将 $V'_k(y_k)$ 具体写出来就是:

$$\begin{aligned} V_k(y_k) &= \underset{(u_{k+1})}{\text{opt}} \{r_{k+1}(y_k, u_{k+1}) + V'_{k+1}(y_{k+1})\} \\ &= \underset{(u_{k+1})}{\text{opt}} \{r_{k+1}(y_k, u_{k+1})\} + \underset{(u_{k+2})}{\text{opt}} \{r_{k+2}(y_{k+1}, u_{k+2}) + V'_{k+2}(y_{k+2})\} \\ &= \underset{(u_{k+1})}{\text{opt}} \{r_{k+1}(y_k, u_{k+1}) + \underset{(u_{k+2})}{\text{opt}} \{r_{k+2}(y_{k+1}, u_{k+2})\} + \underset{(u_{k+3})}{\text{opt}} \{\dots\}\} \dots \end{aligned}$$

上式说明不但在各段中 u'_n, u'_{n-1}, u'_{k+1} 分别是 $V_{n-1}, V_{n-2}, \dots, V_{k+1}, V_k$ 的最优解, 而且对于初始状态 y_k 来说, $u'_{k+1}, u'_{k+2}, \dots, u'_n$ 组成相对于代价函数 $V_k(y_k)$ 的最优决策。

所以递推关系式 (3.69) 是动态规划的基本方程, 可简明地概括为如下的 Bellman 最优性原理: “多段决策过程中的最优决策序列具有这样的性质: 不论初始状态如何, 对于前面决策所造成的某一状态而言, 余下的所有决策总构成一个最优决策。” 它在动态规划中起着核心的作用。

3.5.3 应用举例

【例 3.20】 (机器损坏的最优分配问题) 设某工厂 100 台机器, 能加工两种零件, 要安排四周任务, 据以往经验得知, 机器加工第一种零件, 一周后损坏的概率为 $1/3$, 加工第二种零件, 一周后损坏的概率为 $1/10$, 加工第一种零件一周的效益为 10, 加工第二种零件一周的效益为 7, 现在要分四周安排任务, 试问怎样分配机器的任务能使总收益最大?

【解】 为了用动态规划法求解, 可以把问题看作一个四段的最优化问题。

设原有机器数为 $y_0 = 100$, 第一周中有 u_1 台机器加工第一种零件, 余下的 $y_0 - u_1$ 台加工第二种零件, 可以预计到第二周初可使用的机器将只有 $y_1 = y_0 - \frac{1}{3}u_1 - \frac{1}{10}(y_0 - u_1) = \frac{9}{10}y_0 - \frac{7}{30}u_1$ 台, 而第一周内厂方的收益是

$$r_1 = 10u_1 + 7(y_0 - u_1) = 3u_1 + 7y_0$$

以后三周内可使用的机器效率与收益将依这个公式计算。因此, 我们可令每周开始时可用的机器数 y_n 为状态变量, u_n 为决策变量, 收益 r_n 为阶段效应。显然, 转换方程为:

$$y_n = \frac{9}{10}y_{n-1} - \frac{7}{30}u_n \quad (n = 1, 2, 3, 4)$$

阶段效应是:

$$r_n = 3u_n + 7y_{n-1} \quad (n = 1, 2, 3, 4)$$

并且 $u_n \geq 0, y_{n-1} - u_n \geq 0$, 即 $0 \leq u_n \leq y_{n-1}$

其过程如图 3-24 所示:

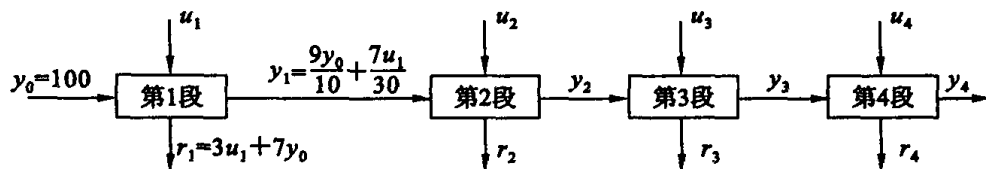


图 3-24

用反向递推求解

$$\text{opt } V_{n-1} = r_n(y_{n-1}, u_n)$$

计算第 4 段有:

$$\max_{0 \leq u_4 \leq y_3} V_3 = 3u_4 + 7y_3$$

这里 y_3 为参数, $u_4(y_3)$ 为决策变量, 所以最优解与最优值是:

$$u_4 = u_4(y_3) = y_3, V'_3(y_3) = 10y_3$$

必须说明, V'_3 为 V_3 段之后最优决策 u_4 时的最优代价函数值, 而第 4 段的收益 $r_4 = 3u_4 + 7y_3$, 因为 $u_4 = y_3$, 所以 $r_4 = 10y_3$ 。

计算第 3 段, 由 (3.68) 式, 有:

$$\begin{aligned} \max_{0 \leq u_3 \leq y_2} V_2 &= V'_3(y_3) + r_3(y_2, u_3) = 10y_3 + 3u_3 + 7y_2 \\ &= 3u_3 + 7y_2 + 10\left(\frac{9}{10}y_2 - \frac{7}{30}u_3\right) \\ &= \frac{2}{3}u_3 + 16y_2 \end{aligned}$$

其最优解及最优值为

$$u_3 = u_3(y_2) = y_2, V'_2(y_2) = \frac{50}{3}y_2$$

计算第 2 段, 求解

$$\begin{aligned} \max_{0 \leq u_2 \leq y_1} V_1 &= 3u_2 + 7y_1 + \frac{50}{3}\left(\frac{9}{10}y_1 - \frac{7}{30}u_2\right) \\ &= 22y_1 - \frac{8}{9}u_2 \end{aligned}$$

其最优解及最优值为 $u_2 = u_2(y_1) = 0, V'_2(y_1) = 22y_1$

类似上述情况求出第 1 段的解: $u_1 = u_1(y_0) = 0, V'_1(y_0) = 26.8y_0$

将 $y_0 = 100$ 代入, 得问题的最优值为 2680, 再依次求出:

$$\begin{aligned} u_1 &= 0, y_1 = 9y_0/10 - 7u_1/30 = 90 \\ u_2 &= 0, y_2 = 9y_1/10 - 7u_2/30 = 81 \\ u_3 &= y_2 = 81, y_3 = 9 \times 81/10 - 7 \times 81/30 = 54 \\ u_4 &= y_3 = 54, y_4 = 9 \times 54/10 - 7 \times 54/30 = 36 \end{aligned}$$

所以最优解为 $u_1 = 0, u_2 = 0, u_3 = 81, u_4 = 54$, 四个阶段以后还余下未损坏的机器 36 台, 总收益为 2680。

【例 3.21】 (生产计划问题) 某机床厂根据合同, 在 1 至 4 月份为客户生产某种机床。工厂每月生产能力为 10 台, 机床可库存, 存储费为每台每月 0.2 万元, 每月需要的数量及每台

机床的生产成本如表 3-19 所示。试确定每月的生产量,要求既能满足各月的需求,又使生产成本和存储费用之和达到最小。

表 3-19 每月的需求量及生产成本

月 份	1	2	3	4
需求(台)	6	7	12	6
生产成本(万元/台)	7	7.2	8	7.6

这是一个 4 阶段的动态规划问题,阶段数 k 取值为 1, 2, 3, 4, 第 k 阶段代表第 k 月份。状态变量和决策变量的含义如下:

x_k ——第 k 阶段开始时的机床台数;

u_k ——第 k 阶段的机床台数。

还规定:

c_k ——第 k 阶段的单位生产成本;

q_k ——第 k 阶段的需求量。

阶段指标函数(阶段效应) 式为: $d_k(x_k, u_k) = c_k \times u_k + 0.2x_k$

状态转移方程为: $x_{k+1} = u_k + x_k - q_k$

指标传递函数(代价函数) 为: $f_k(x_k, u_k) = d_k(x_k, u_k) + f_{k+1}(u_k)$

$$f(x_k, u_k) = d_k(x_k, u_k) + f_{k+1}(u_k)$$

$$k = 3, 2, 1$$

现从第 4 段开始, 逆序进行逐阶段递推计算。

在第 4 阶段, 由于前 3 个月的最大产量为 30 台, 需求量为 25 台, 所以 4 月初的最大存储量等于 5 台, 即 $5 \geq x_4 \geq 0$ 。而在 4 月底必须完成合同规定的生产任务, 又不应有库存, 所以 $u_4 = q_4 - x_4 = 6 - x_4$, 所以可得

$$\begin{aligned} f_4(x_4, u_4) &= c_4(6 - x_4) + 0.2 \times x_4 \\ &= 45.6 - 7.4x_4 = f_4(x_4) \quad (c_4 = 7.6) \end{aligned}$$

在第 3 阶段, 由于 3 月份的需求量是 12 台, 而生产能力只有 10 台, 所以 3 月初至少有 2 台库存, 3 月初的最大库存量可能达到 7 台, (因 3 月初之前的 1、2 月份生产能力为 20 台, 需求量为 $6 + 7 = 13$ 台) 所以 $7 \geq x_3 \geq 2$, $10 \geq u_3 \geq 12 - x_3$, $x_4 = x_3 + u_3 - 12$, 得

$$\begin{aligned} f_3(x_3, u_3) &= \min_{u_3} [d_3(x_3, u_3) + f_4(x_4)] \\ &= \min_{u_3} [c_3 u_3 + 0.2x_3 + 45.6 - 7.4x_4] \\ &= \min_{u_3} [8u_3 + 0.2x_3 + 45.6 - 7.4(x_3 + u_3 - 12)] \end{aligned}$$

当 $u_3 = 12 - x_3$ 时, 上式右端取值最小

$$\begin{aligned} f_3(x_3) &= 0.6(12 - x_3) - 7.2x_3 + 134.4 \\ &= 141.6 - 7.8x_3 \end{aligned}$$

在第 2 阶段, 可得

$$\begin{aligned} 4 &\geq x_2 \geq 0 \\ 10 &\geq u_2 \geq 7 - x_2 \\ x_3 &= x_2 + u_2 - 7 \end{aligned}$$

$$\begin{aligned}
 f(x_2, u_2) &= \min_{u_2} [d_2(x_2, u_2) + f_3(x_3)] \\
 &= \min_{u_2} [7.2u_2 + 0.2x_2 + 141.6 - 7.8x_3] \\
 &= \min_{u_2} [7.2u_2 + 0.2x_2 + 141.6 - 7.8(x_2 + u_2 - 7)] \\
 &= \min_{u_2} (-0.6u_2 - 7.6x_2 + 196.2)
 \end{aligned}$$

当 $u_2 = 10$ 时上式右端取最小值

$$f_2(x_2) = 190.2 - 7.7x_2$$

在第 1 阶段, 可得 $x_1 = 0$

$$\begin{aligned}
 6 &\leq u_1 \leq 10 \\
 x_2 &= u_1 - 6 \\
 f_1(x_1, u_1) &= \min_{u_1} [d_1(x_1, u_1) + f_2(x_2)] \\
 &= \min_{u_1} [7u_1 + 190.2 - 7.6(u_1 - 6)] \\
 &= \min_{u_1} (-0.6u_1 + 235.8)
 \end{aligned}$$

$u_1 = 10$ 时, 上式右端取值最小 $f_1(x_1) = 229.8$

至此, 已求出问题的最优解 $u_1 = 10, u_2 = 10, u_3 = 5, u_4 = 6$, 最小总费用为 229.8 元。

【例 3.22】 (装载问题) 有一只容积为 5 的集装箱, 用于装载 3 种货物, 货物的编号用 k 表示, 第 k 种货物的单件所占体积为 a_k , 单件价格为 c_k , a_k 和 c_k 的取值见表 3-20。问: 各种货物应装多少件, 才能使整箱货物的价值最大?

表 3-20 装载问题取值表

k	a_k	c_k
1	2	65
2	3	80
3	1	30

这个问题可划为 3 个决策阶段, 第 k 阶段对第 k 种货物进行决策。

定义:

状态变量: x_k ——进入第 k 阶段时的剩余载货容积;

决策变量: u_k ——第 k 种货物的装载件数;

阶段指标函数(阶段效应): $d_k(x_k, u_k) = c_k \times u_k$;

状态转移方程: $x_{k+1} = x_k - a_k \times u_k$

指标传递函数(代价函数):
$$\begin{cases} f_k(x_k, u_k) = d_k(x_k, u_k) + f_{k+1}(x_{k+1}, u_{k+1}) & (k = 2, 1) \\ f_3(x_3, u_3) = d_3(x_3, u_3) \end{cases}$$

第 3 阶段:

由于 $0 \leq x_3 \leq 5$ 且 $a_3 = 1$, 所以 $u_3 = 0, 1, 2, 3, 4, 5$

对于给定情况, 计算结果见表 3-21。

表 3-21 第 3 阶段计算结果

x_3	$f_3(x_3, u_4) = 30u_4$						最 优 解	
	$u_3 = 0$	1	2	3	4	5	u_3^*	$f_3(x_3)$
	$c_3 u_3 = 0$	30	60	90	120	150		
0	0						0	0
1	0	30					1	30
2	0	30	60				2	60
3	0	30	60	90			3	90
4	0	30	60	90	120		4	120
5	0	30	60	90	120	150	5	150

第 2 阶段:

由于 $0 \leq x_2 \leq 5$, 且 $a_2 = 3$, 所以 $u_2 = 0.1$. 对于各种情况, 计算结果见表 3-22.

表 3-22 第 2 阶段计算结果

x_2	$f_2(x_2, u_2) = 80u_2 + f_2(x_2 - 3u_2)$		最 优 解	
	$u_2 = 0$	1	u_2	$f_2(x_2)$
	$c_2 u_2 = 0$	80		
0	0 + 0		0	0
1	0 + 30 = 30		0	30
2	0 + 60 = 60		0	60
3	0 + 90 = 90	80 + 0 = 80	0	90
4	0 + 120 = 120	80 + 30 = 110	0	120
5	0 + 150 = 150	80 + 60 = 140	0	150

从表 3-21 中可看出, 无论 x_2 取何值, u_2 的值都为零, 即在整个问题的最优解中 u_2 等于零。

第 1 阶段:

由于 $x_1 = 5$, 且 $a_1 = 2$, 所以 $u_1 = 0, 1, 2$, 计算结果见表 3-23.

表 3-23 第 1 阶段计算结果

x_1	$f_1(x_1, u_1) = 65u_1 + f_1(x_1 - 2u_1)$			最 优 解	
	$u_1 = 0$	1	2	u_1^*	$f_1(x_1)$
	$c_1 u_1 = 0$	65	130		
5	50 + 150 = 200	65 + 90 = 155	130 + 30 = 160	2	160

综合上述计算结果, 得最优解为 $u_1^* = 2, u_2^* = 0, u_3^* = 1$.

总的价值为 160。

习 题

3.1 合理下料问题。要用一批长度为 7.4m 的圆钢做 100 套钢索。每套由长 2.9m, 2.1m 和 1.5m 的圆钢各一根组成, 应如何下料才能使所用原料最省? 试建立此问题的数学模型。

3.2 用分枝定界法求解下列整数规划。

$$\begin{aligned} \max f(x) &= 3x_1 + 2x_2 \\ \text{s. t. } &\begin{cases} 2x_1 + 3x_2 \leq 14 \\ 2x_1 + x_2 \leq 9 \\ x_1, x_2 \geq 0 \text{ 且为整数} \end{cases} \end{aligned}$$

3.3 设线性规划的原问题为:

$$\begin{aligned} \max f(x) &= 3x_1 + x_2 \\ \text{s. t. } &\begin{cases} x_2 \leq 10 \\ 2x_1 + x_2 \leq 12 \\ x_1, x_2 \geq 0 \end{cases} \end{aligned}$$

试列出增加松弛变量以后的线性规划模型和对偶问题模型。

3.4 用图解法求线性规划问题:

$$\begin{aligned} \max f(x) &= x_1 + 2x_2 \\ \text{s. t. } &\begin{cases} 3x_1 + 2x_2 \leq 60 \\ 2x_1 + 6x_2 \leq 90 \\ x_1 + 3x_2 \geq 30 \\ 2x_1 + x_2 \geq 20 \end{cases} \end{aligned}$$

3.5 设 $f(x) = \frac{1}{2} X^T Q X$, $Q = \begin{bmatrix} -1 & 0 \\ 0 & 100 \end{bmatrix}$, 试求最优步长的最优函数值 $f(x^*)$ 。

3.6 求 $\min f(x) = x_1^2 + x_2^2 \quad x \in E^2$

$$\text{s. t. } \begin{cases} g_1(x) = 3x_1 + 2x_2 - 6 \geq 0 \\ g_2(x) = x_1 \geq 0 \\ g_3(x) = x_2 \geq 0 \end{cases}$$

3.7 某厂从所在地 A 运一套大型设备到港口城市 G。有多条路线可供选择, 其路线及费用如图 3-25 所示。要选一条总运费最小的路线。试将该问题用动态规划的方法求最优解。

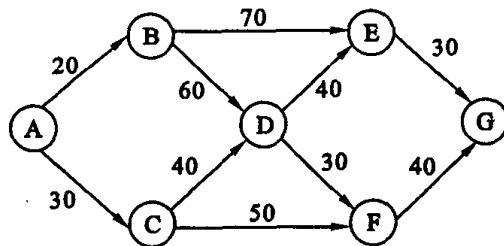


图 3-25

4 图与网络

本章提要

图与网络是运筹学的重要组成部分,是其一个重要分支。图论在运筹学中是从实践抽象出最早的数学问题之一。它研究的基本内容是连通问题,如最早欧拉于1736年提出的“哥尼斯堡七桥问题”,以及后来的“哈密尔顿环问题”和“地图四色问题”等。其后又发展出网络,网络是对“节点”间的“弧”赋予“权”(即“代价”或“价值”)的有向图。图与网络在近数十年在理论及应用方面都得到了发展,对庞大复杂的工程系统和管理都有效地用图与网络进行描述与分析。

本章介绍有关图的基本概念,分析了无向图和有向图,图论对研究系统可达性有重要意义;介绍了有关网络的基本概念,并给出了网络极值问题,即网络最优化的三个基本问题——最短路问题,最大流问题和最小费用——最大流问题的算法。

4.1 图的基本概念

在电路原理中,我们已有了网络的概念。在电网络中它是支路和节点的集合。而研究网络图,实质上是考虑网络结构而不考虑网络中元件的性质。这样去掉了元件符号,而只有节点和支路的网络图,是一个抽象的图。研究图的理论称为图论。现在,图论除了在电路和控制理论中早已得到广泛应用外,还渗透到其他许多学科领域:如信息论、运筹学、运输网络、通信网络、计算机网络等等。

实际上,商业储运、交通运输、管道工程等领域的设计、运行优化问题,都离不开图和网络基础。

4.1.1 图与路

(1) 图

图4-1是一个画在平面上的线图。它是由一些点和一些线段构成的。这些点称为顶点,如 a, b, c, d 。这些线段称为边或支路,如 $1, 2, 3, 4, 5, 6, 7$ 。如果线段上带指方向的箭头,则这个线段称为弧,顶点 a 和顶点 c 之间是线段 1 ,则称边 1 与顶点 a, c 相关联,而边 7 只与顶点 a 相关联。这时,我们可以给“图”一个定义:图 G 是由点集 V 和边集 E (或弧 A)组成,且点与边具有关联关系。记作 $G = (V, E)$ 或 $G = (V, A)$ 。

其中: $V = \{v_1, v_2, \dots, v_p\}$ 是 p 个点的集合

$E = \{e_1, e_2, \dots, e_q\}$ 是 q 个边的集合

或 $A = \{e_1, e_2, \dots, e_q\}$ 且 $e_i = (v_i, v_j) \neq (v_j, v_i), v_i, v_j \in V$ 所以是 q 个弧的集合。

一般, $G = (V, E)$ 称为无向图, $G = (V, A)$ 则称为有向图。

两个顶点间具有两条或两条以上的边称为多重边,若边的两端为同一顶点,则称其为

环,如图 4-1 中的“7”。含多重边的图为多重图;无多重图且无环的图称为简单图。本章讨论的图一般都为简单图。

若图 $G = (V, E)$, $G' = (V', E')$ 且 $V' \subseteq V, E' \subseteq E$, 则称 G' 为 G 的子图(图 4-2(b))。若在同一图中, $V' = V, E' \subseteq E$, 则称 G' 为 G 的支撑子图(如图 4-2(c) 所示)。

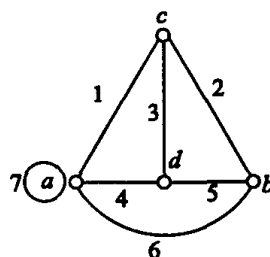


图 4-1

(2) 链、路、树

在图形中,有由 n 条边组成的序列 $(v_1, v_2, \dots, v_n)$, 若除 v_1 和 v_n 外, 每条边的一个端点与前一边相连, 而另一端点与后一边相连接, 而 v_1 的一个端点是这个序列的起点, 它的另一端点与第二条边相连接; 最后一个边 v_n 的一个端点为序列的终点。另一端与前一个边 v_{n-1} 相连接, 则这个由这 n 个边组成的序列 $(v_1, v_2, \dots, v_n)$ 称为链。

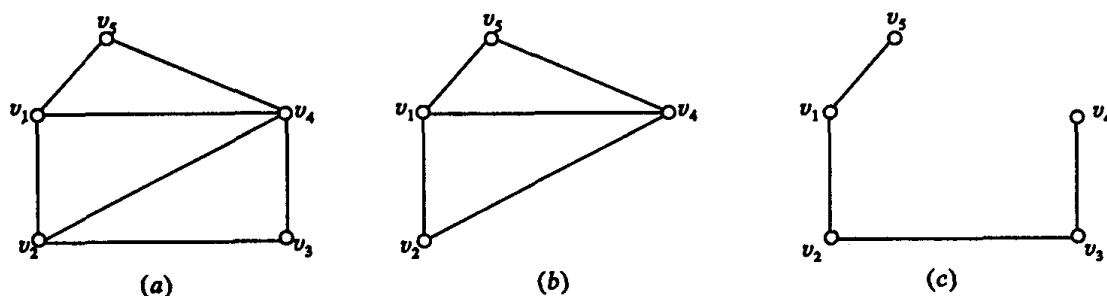


图 4-2

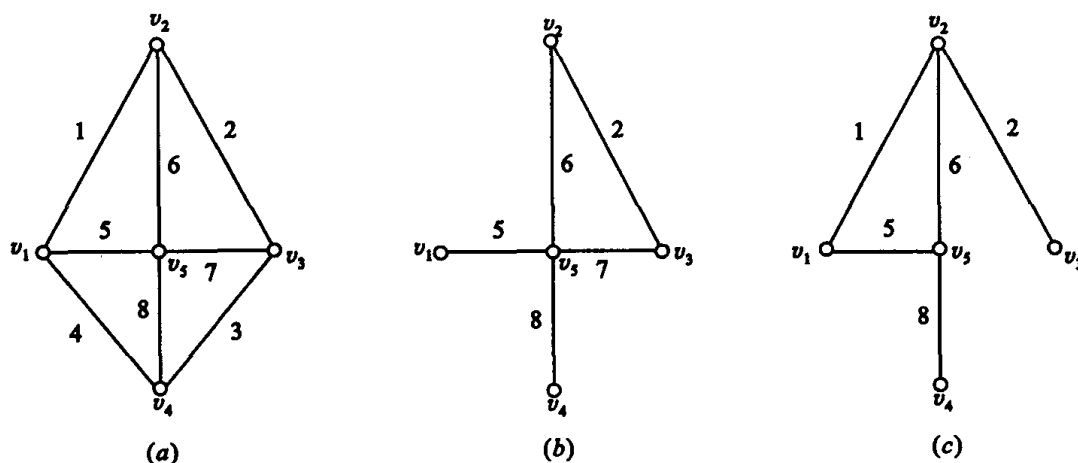


图 4-3 链与路

当一个链的起点与终点不是同一顶点时称为开链, 反之为闭合链。一个没有重复边的链称为单纯链; 一个没有重复顶点(当然不会有重复边)的链称为基本链。以图 4-3 为例, 边序列 $\{5, 6, 2, 7, 8\}$ 为单纯链。一个开的基本链即称为路, 如边序列 $\{1, 6, 7, 3\}$, $\{7, 5, 4\}$, $\{1, 5, 8, 3\}$ 等等, 一个闭合的基本链称为回路, 如图 4-3 图形中的边序列 $\{1, 6, 5\}$, $\{1, 2, 3, 4\}$, $\{4, 1, 6, 7, 3\}$ 等。

在图形中, 若任意两个不相同的顶点之间, 至少有一条路, 则该图形称为连通图。或者说在连通图中两个任意不同顶点之间都能沿着一组边的序列相通。

树在图论中是一个十分重要的概念。

设 G 为连通图, 从 G 中移出一些边后, 得 G 的子图 G_1 。如果 G_1 也是连通的并包括了 G 的所有顶点, 且没有回路, 则 G_1 为 G 的一个生成树, 简称树, 例如图 4-4(a) 即为图 $G(V, E)$, 其

中 $V = \{a, b, c, d\}$, $E = \{1, 2, 3, 4, 5\}$, 而图 4-4(b)、(c)、(d) 即为图 G 的树: $T_1 = \{1, 3, 5\}$, $T_2 = \{2, 3, 5\}$, $T_3 = \{2, 4, 5\}$ 。当然还有很多树。组成树的边叫做树枝。

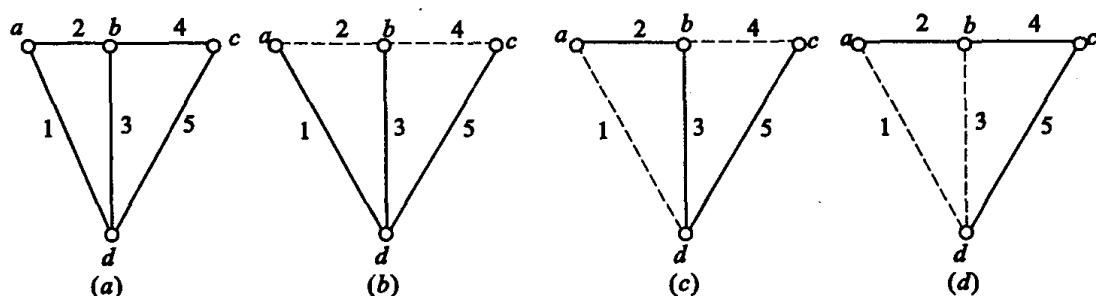


图 4-4 树与树枝

与树互补的子图叫做树的补图, 或简称为补树, 组成补树的边叫做连枝。从图 4-4(c) 可以看出, 补树不一定是连通的, 也不一定包含全部顶点(如图 4-4(b) 和 (d) 所示), 有时也可以形成回路。总之补树不一定是树形, 所以有时为了避免误解, 常将补树称为连枝集。

树具有以下的特点:

- ① 树的任意两顶点之间只有一条链;
- ② 在树的两个不相邻的顶点间添上一条边(即一个连枝)就得到一个回路, 这个回路称作基本回路;
- ③ 在树中去掉任何一条边, 图就不连通;
- ④ 含有 n 个顶点的树, 则有 $n-1$ 条边。

4.1.2 割集

图 G 为连通图, 当移走一组边, 将图分割成两个互相分离的部分而不再连通; 而只要少去掉一个边, 被分开的两部分又是连通的。所以这刚够把这连通图分割为两部分。所需割断的边的集合, 叫做割集。其简单的定义为: 把连通图分割为两个子图所需割断的最少边集。

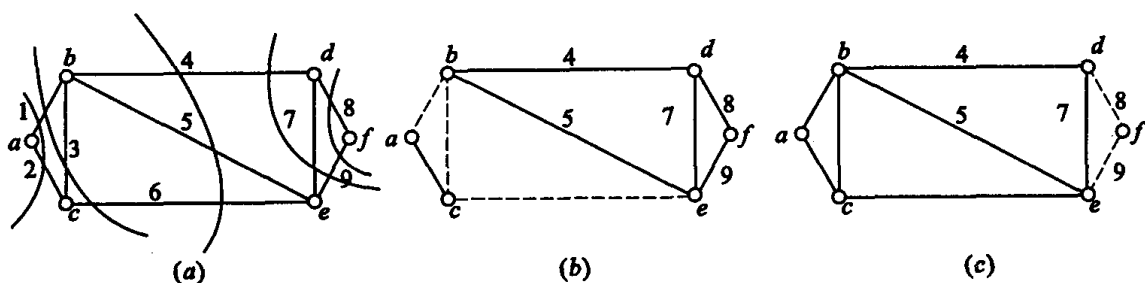


图 4-5 图的割集

如图 4-5(a) 所示为一个连通图。若去掉边集 $\{1, 3, 6\}$, 将使原图分为两个分离部分, 如图 4-5(b) 所示, 只要把边集 $\{1, 3, 6\}$ 中的任意一个边加上去, 则图 4-5(b) 就变成了连通的, 故边集 $\{1, 3, 6\}$ 是一个割集。同理, 边集 $\{8, 9\}$ 也是一个割集, 移去边 $\{8, 9\}$ 便得到图 4-5(c) 其中孤立的 f 点也是一个分离部分。而边集 $\{8, 7, 9\}$ 就不是割集, 因为移去这条边集可以使图分离为两部分, 但若将边 7 再接上, 已分离的图仍不能连通, 不符合割集的定义。

对于一个图有若干个割集, 其中边集最少的割集, 称为最小割集, 简称最小割。

4.2 最短路问题

最短路问题是研究最小费用流的基础,也是网络理论中应用最广的问题之一。值得较深入研究。

4.2.1 基本概念

行车路线、铺设管道、架设电力线路都要考虑在满足要求的情况下,使其最经济、最省力,所以必须选择一个最佳路线,使得走的距离最短,这就是最短路问题。

【例 4.1】 某邮递员投递邮件时,在街道网络上,选择最短路线。(图 4-6)

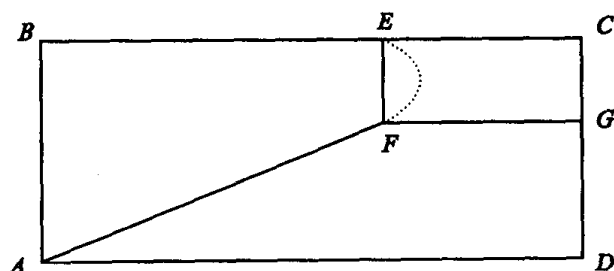


图 4-6 最短路问题

从图中可以发现,邮递员最好能不重复地一次走完规定街道。如果在距离最短的两点间加一条重复路线(如图中虚线 EF),则可沿路线 $ABEFADGCEFG$ 走,除 EF 重复一次外,其余都不重复地一次走完,是一条最短的邮递路线。

【例 4.2】 铺设管道问题,要求水源与各供水点相通,而且各点到水源间路径最短。

【解】 在图 4-7 中, v_1 为水源, $v_2, v_3, \dots, v_{12}$ 为供水点,若 $v_1 - v_2 - v_3 - \dots - v_n$ 是从 v_1 到 v_n 的最短路,则显然, $v_1 - v_2 - v_3 - \dots - v_{n-1}$ 是从 v_1 到 v_{n-1} 的最短路,根据这一道理,可求出水源到各供水点的最短路。

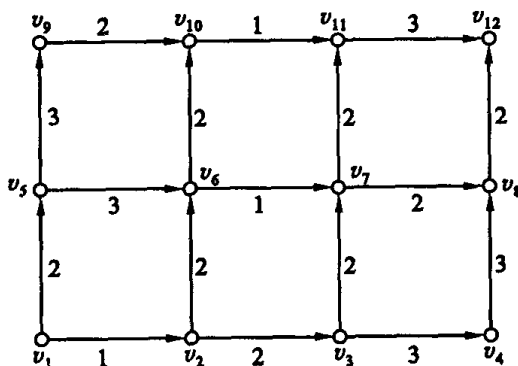


图 4-7 管道铺设问题

(1) 与 v_1 直接相连的点为 v_2, v_5 , 但 $l_{12} = 1, l_{15} = 2, l_{12} < l_{15}$, 所以 $v_1 v_2$ 为最短路。联接 v_1 及 v_2 两点(图 4-8)。

(2) 与 $v_1 v_2$ 直接相连的点为 $v_3, v_5, v_6, l_{15} = 2, l_{13} = 3, l_{16} = 3, \min\{l_{13}, l_{16}, l_{15}\} = 2$, 联接 v_1 和 v_5 。

(3) 与 $v_1 v_2 v_5$ 相连的点为 v_3, v_6, v_9 , 则 $l_{13}, l_{16} = 3, l_{19} = 5, \min\{l_{13}, l_{16}, l_{19}\} = l_{13}, l_{16} =$

从始点 v_1 开始, 给每点一个数, 或称标号, 分 T 标号与 P 标号两种, T 为试探性标号, P 为永久性标号。

给 v_j 点一个 P 标号时, 表示从 v_i 到 v_j 最短路长, 给 v_j 点一个 T 标号, 表示从 v_i 到 v_j 的估计最短路长。凡是没有标上 P 标号的点, 标以 T 标号, 以便用来进一步计算该点的 P 标号, 已得到 P 标号的点不再改变标号, 一旦终点得到 P 标号算法停止。最后经过 $m-1$ 步算法就可以得到从始点 v_1 到终点 v_m 的最短路。

对图 4-9 的最短路算法如下:

① v_1 到 v_1 的距离为 0, 故将 v_1 标上 P 标号为 0, 即 $P(1) = 0$, 其他点 ($v_2 \sim v_6$) 还没有 P 标号, 给它们标上 $T(j) = \infty, j = 2, 3, 4, 5, 6$,

② 分析还没有得到 P 标号的各点 $v_j (j = 2, \dots, 6)$, 比较 $T(j)$ 和 $P(1) + l_{1j}$, l_{1j} 为 v_1 到 v_j 的弧长, 将 v_j 的新标号取较小者, 即 $T(j) = \min[T(j), P(1) + l_{1j}]$, 如果 $[P(1) + l_{1j}] < T(j)$, 则有 $P(1) + l_{1j}$ 代替 $T(j)$ 作为 v_j 点的新的 T 标号。

$$T(2) = \min[T(2), P(1) + l_{12}] = \min[\infty, 0 + 16] = 16$$

$$T(3) = \min[T(3), P(1) + l_{13}] = \min[\infty, 0 + 22] = 22$$

$$T(4) = \min[T(4), P(1) + l_{14}] = \min[\infty, 0 + 30] = 30$$

$$T(5) = \min[T(5), P(1) + l_{15}] = \min[\infty, 0 + 41] = 41$$

$$T(6) = \min[T(6), P(1) + l_{16}] = \min[\infty, 0 + 59] = 59$$

点 v_2 的 T 标号数最小, 令 $P(2) = 16$

③ 以点 v_2 (刚得到 P 标号的点) 作为比较基础, 重复上述过程, 即比较 $T(j)$ 与 $[P(2) + l_{2j}]$, $j = 3, 4, 5, 6$, 取较小的数作为 v_j 的新的 T 标号。

$$T(3) = \min[T(3), P(2) + l_{23}] = \min[22, 16 + 16] = 22$$

$$T(4) = \min[T(4), P(2) + l_{24}] = \min[30, 16 + 22] = 30$$

$$T(5) = \min[T(5), P(2) + l_{25}] = \min[41, 16 + 30] = 41$$

$$T(6) = \min[T(6), P(2) + l_{26}] = \min[59, 16 + 41] = 57$$

显见, 点 v_6 的 T 标号由 59 变为 57, 其余未改变。比较 $T(3), T(4), T(5), T(6)$, $T(3) = 22$ 为最小, 将点 v_3 的 T 标号改为 P 标号: $P(3) = 22$ 。

④ 再重复上述过程, 以点 v_3 为比较基础。

$$T(4) = \min[T(4), P(3) + l_{34}] = \min[30, 22 + 17] = 30$$

$$T(5) = \min[T(5), P(3) + l_{35}] = \min[41, 22 + 23] = 41$$

$$T(6) = \min[T(6), P(3) + l_{36}] = \min[57, 22 + 31] = 53$$

点 v_6 的 T 标号由 57 变为 53, 其余未动。 $T(4) \sim T(6)$ 中 $T(4) = 30$ 为最小, 令 $P(4) = 30$ 。

⑤ 下一步以点 v_4 为比较基础

$$T(5) = \min[T(5), P(4) + l_{45}] = \min[41, 30 + 17] = 41$$

$$T(6) = \min[T(6), P(4) + l_{46}] = \min[53, 30 + 23] = 53$$

v_5, v_6 的 T 标号均未变, $T(5) = 41$ 为最小, 令 $P(5) = 41$

⑥ 最后一步以点 v_5 为比较基础

$$T(6) = \min[T(6), P(5) + l_{56}] = \min[53, 41 + 18] = 53$$

令 $P(6) = 53$ 。

终点 v_6 已标了 P 标号, 程序到此结束。前已指出, $P(j)$ 表示始点 v_1 到 v_j 的最短路长。例

如点 v_1 到 v_3 最短路长为 22, 点 v_1 到 v_5 最短路长为 41, 而点 v_1 到终点最短路长为 $P(6) = 53$ 。

如果只需知道从点 v_1 到点 v_k 的最短路, 则一旦点 v_k 标了 P 标号, 算法即可终止。(编程时可安排在这时停止计算)。

这种算法只适用于弧长为正的情况。在考虑弧长为负的情况时, 上述算法显然不适用了。关于这种特殊情况算法, 我们不作详细讨论。

4.2.3 从任一点到另外任一点的最短路算法

这种算法要求先构造长度矩阵 L , 依次计算:

$$L^{(2)} = L * L, L^{(3)} = L^{(2)} * L \dots$$

定义:

$$A * B = C = [c_{ij}]_{m \times n}, \text{ 其中 } A = [a_{ij}]_{m \times l} \quad B = [b_{ij}]_{l \times n}$$

$$c_{ij} = \min\{a_{i1} * b_{1j}, a_{i2} * b_{2j}, \dots, a_{il} * b_{lj}\} \quad (4.2)$$

若 $a_{ik} * b_{kj} (k = 1, 2, 3, \dots)$ 全为 0, 则 $c_{ij} = 0$

$$a_{ik} * b_{kj} = \begin{cases} 0 & \text{当 } a_{ik} = 0 \text{ 或 } b_{kj} = 0 \\ a_{ik} + b_{kj} & a_{ik} \neq 0, b_{kj} \neq 0 \end{cases} \quad (4.3)$$

【例 4.3】 $A = [1 \ 2 \ 3 \ 4 \ 0 \ 0], B = [0 \ 2 \ 1 \ 1 \ 1 \ 0]^T$, 由 (4.2) 式有:

则 $A * B = [1 \ 2 \ 3 \ 4 \ 0 \ 0] * \begin{bmatrix} 0 \\ 2 \\ 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \min_{\text{非零元素}} \{0, 4, 4, 5, 0, 0\} = 4$

下面我们以图 4-10 说明各长度矩阵的构造法。取没有路时长度为 0。

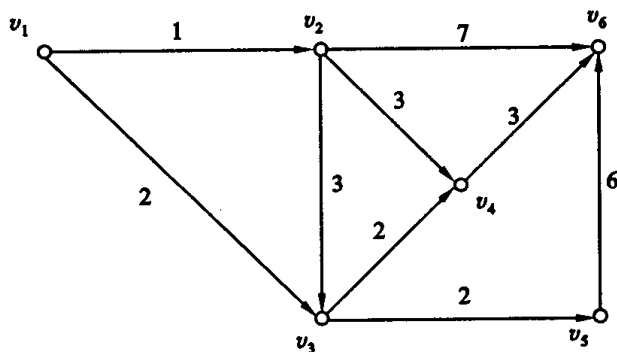


图 4-10 网络举例

$$L^{(2)} = [l_{ij}^{(2)}] = L * L = \begin{matrix} & \begin{matrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 \end{matrix} \\ \begin{matrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{matrix} & \begin{bmatrix} 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 3 & 0 & 7 \\ 0 & 0 & 0 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{matrix}$$

$$= \begin{bmatrix} 0 & 0 & 4 & 4 & 4 & 8 \\ 0 & 0 & 0 & 5 & 5 & 6 \\ 0 & 0 & 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$l_{ij}^{(2)} \neq 0$ 表示从 v_i 到 v_j 走两步的路程中距离最短的一条。

$l_{ij}^{(2)} = 0$ 表示从 v_i 到 v_j 两步到不了。

例如 $l_{14}^{(2)} = 4$ 表示从 v_1 到 v_4 走两步的最短距离为 4。(从 v_1 经达 v_2 或 v_3 到 v_4)。

$l_{36}^{(2)} = 5$, 由图 4-10 可见从 v_3 到 v_6 , 走两步有两条路, $v_3 \rightarrow v_5 \rightarrow v_6$, $l_{36} = 8$ 及 $v_3 \rightarrow v_4 \rightarrow v_6$, $l_{36} = 5$ 。其中最短路为 $v_3 \rightarrow v_4 \rightarrow v_6$, $\therefore l_{36}^{(2)} = 5$ 。

$$L^{(3)} = L^{(2)} * L = \begin{bmatrix} 0 & 0 & 0 & 6 & 6 & 7 \\ 0 & 0 & 0 & 0 & 0 & 8 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = [l_{ij}^{(3)}]$$

同理 $l_{ij}^{(3)}$ 表示从 v_i 走三步到 v_j 距离最短的一条路。例如 $v_2 \rightarrow v_3 \rightarrow v_5 \rightarrow v_6$, $l_{26}^{(3)} = 11$, $v_2 \rightarrow v_3 \rightarrow v_4 \rightarrow v_6$, $l_{26}^{(3)} = 8$, 所以 $l_{26}^{(3)} = \min[8, 11] = 8$

$$L^{(4)} = L^{(3)} * L = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 9 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$L^{(5)} = L^{(4)} * L = 0 \text{ (零矩阵)}$$

为求从 v_i 到 v_j 的最短距离, 定义 $A \oplus B = [c_{ij}]_{m \times n}$

$$A = [a_{ij}]_{m \times n} \quad B = [b_{ij}]_{m \times n}$$

其中:

$$c_{ij} = \begin{cases} 0, & a_{ij} = b_{ij} = 0 \\ a_{ij}, & a_{ij} > 0, b_{ij} = 0 \\ b_{ij}, & a_{ij} = 0, b_{ij} \neq 0 \\ \min(a_{ij}, b_{ij}), & a_{ij}, b_{ij} \neq 0 \end{cases} \quad (4.4)$$

根据上述定义累加得:

$$L_{\min} = L \oplus L^{(2)} \oplus L^{(3)} \oplus L^{(4)} \oplus L^{(5)}$$

$$= \begin{bmatrix} 0 & 1 & 2 & 4 & 4 & 7 \\ 0 & 0 & 3 & 3 & 5 & 6 \\ 0 & 0 & 0 & 2 & 2 & 5 \\ 0 & 0 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = [l_{ij \min}]$$

$l_{ij \min}$ 表示从 v_i 到 v_j 的最短距离。

4.3 网络最大流

许多工程系统的网络包含了流量问题。例如公路系统中有车辆流(车辆数/小时), 控制系统中有信息流, 供水系统中有水流($\text{m}^3/\text{单位时间}$), 金融系统中有现金流等等。网络中每条弧有一定的容量 c , 即最大通过能力有限制。实际流量 f 并不一定等于容量 c 。决定网络最大流量称为网络最大流问题。

4.3.1 可行流与最大流

图 4-11 表示一个输油管道网络。弧为输油管, 弧上的数字(c_{ij})表示从 v_i 到 v_j 每条弧(输油管)的最大输油能力。点 $v_1, v_2, \dots, v_6$ 表示油泵站。

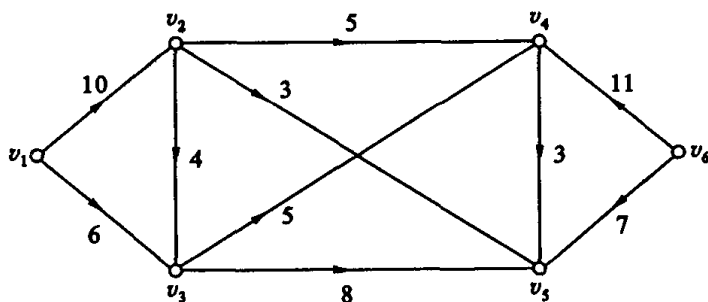


图 4-11 管道网络

设 f_{ij} 为从 v_i 到 v_j 的流量。由图 4-11 可知:

$$\left. \begin{aligned} f_{12} &\leq 10, f_{13} \leq 6, f_{23} \leq 4, f_{24} \leq 5, f_{25} \leq 3 \\ f_{34} &\leq 5, f_{35} \leq 8, f_{46} \leq 11, f_{54} \leq 3, f_{56} \leq 7 \end{aligned} \right\} \quad (4.5)$$

点 v_1, v_6 分别称为发点和收点, 应满足:

$$f_{12} + f_{13} = f_{46} + f_{56} \quad (4.6)$$

最大流问题可用线性规划模型表示:

$$\text{目标} \quad \max F = f_{12} + f_{13} \quad (4.7)$$

约束条件为各中间点的输入油量 = 输出油量

$$\text{即} \quad \left. \begin{aligned} f_{12} - f_{23} - f_{24} - f_{25} &= 0 \\ f_{13} + f_{23} - f_{34} - f_{35} &= 0 \\ f_{24} + f_{34} + f_{54} - f_{46} &= 0 \\ f_{25} + f_{35} - f_{54} - f_{56} &= 0 \\ f_{12} + f_{13} - f_{46} - f_{56} &= 0 \end{aligned} \right\} \quad (4.8)$$

此外流量 f_{ij} 总是非负的, 所以对所有 i 及 j

$$f_{ij} \geq 0 \quad (4.9)$$

并应满足式(4.6)规定的范围。

运输问题是最大流问题的实际应用。在 A 、 B 两城市之间有几个中间站由铁路网络连成一片, 每两站之间铁路通过能力即容量用数 c_{ij} 表示。设 A 为某产品产地, B 为销地, 产品通过铁路网从 A 运到 B , 求从 A 到 B 的最大运输量(即最大流)。每段铁路的运输量当然不能超过该段最大通过能力。而每个中间站总运进量等于总运出量。从 A 城的运出量 F 等于给 B 城的运进量。

最大流问题可以用线性规划问题求得, 也可以用网络模型求解。

设有向图 $G(V, A)$ 。弧 $a_{ij} = (v_i, v_j) \in A$, 弧 a_{ij} 的容量为 c_{ij} , f_{ij} 为 a_{ij} 上的流量。弧的流量函数集合称为网络流。

网络中有发点(或源) v_s 和收点(或汇集点) v_t , 其余点为中间点。

定义: 满足下述条件的流 f 称为可行流:

① 容量限制条件, 对每弧 $a_{ij} = (v_i, v_j) \in A$

$$0 \leq f_{ij} \leq c_{ij} \quad (4.10)$$

这是线性不等式约束条件。

② 平衡条件

对中间点来说流出量 = 流入量

$$\sum f_{ij} - \sum f_{ji} = 0, i \neq s, t \quad (4.11)$$

$$\left. \begin{aligned} \text{发点 } v_s \text{ 净输出 } \sum f_{ij} &= F \\ \text{收点 } v_t \text{ 净输入 } \sum f_{ji} &= F \end{aligned} \right\} \quad (4.12)$$

这是线性等式约束条件。

用网络模型求可行流就是一个总输送量为 F 的运输方案, 相当于线性规划中求一个可行解。如果所有各弧流量为 0, 即 $f = 0$ 这也是一种可行流, 它的流量为 0。

最大流问题为求一个使流量 F 达到最大的可行流 $\{f_{ij}\}$, 使满足

$$\left\{ \begin{aligned} 0 &\leq f_{ij} \leq c_{ij} \\ \sum f_{ij} - \sum f_{ji} &= \begin{cases} F, & i = s \\ 0, & i \neq s, t \\ -F, & i = t \end{cases} \end{aligned} \right. \quad (4.13)$$

4.3.2 最大流 —— 最小割定理

【例 4.4】 图 4-12(a) 为一个交通运输网络 $G(V, A)$ 。假设有一条小河流穿过 (v_b, v_t) 和

(v_i, v_t) 两弧, 能渡河的运输量各为 22 与 10。所以运输量不会超过 32。若小河穿过 (v_i, v_a) , (v_b, v_a) , (v_b, v_c) , (v_b, v_t) 四段路, 则渡河运输量不超过 $20 + 9 + 13 + 22 = 64$, 见图 4-12(b)。

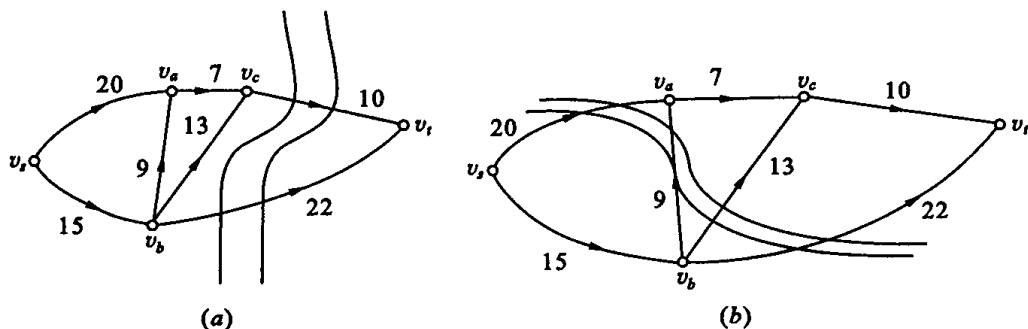


图 4-12 交通运输网络

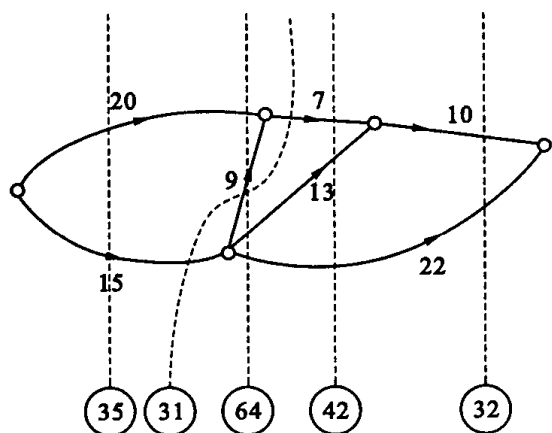


图 4-13 割集

上面所述两条小河分别割断, 使连通图 G 中收点 v_t 和发点 v_s 分别属于两个子图, 成为割集。割集是从 v_s 到 v_t 的必经之路, 显见任何一个可行流的流量 F 不会超过任一割集的容量, 称为割集容量。上述运输网络有 5 个割集 (如图 4-13), 其容量分别为 35, 31, 64, 42, 32。最小的割集容量为 31, 简称为最小割。显然从 v_s 流向 v_t 的运输量 F 最大不能超过 31。因此, 对于我们所关注的网络最大流问题, 可以得到这样的结论: 任一网络 G 中从 v_s 到 v_t 的最大流的流量等于分离 v_s, v_t 的最小割集容量。这种关系称为最大流——最小割定理:

$$\text{最大流的流量} = \text{最小割集容量} \quad (4.14)$$

这是求最大流问题的一个重要定理, 它告诉我们: 若能找到一个可行流 f^* 和一个割集 $(x^*, \bar{x}^*)$ 使 $F(f^*) = C(x^*, \bar{x}^*)$, 则 f^* 就是最大流, 而 $(x^*, \bar{x}^*)$ 是所有割集中容量最小的一个。

定理: 任一网络 G 中, 若所有弧容量为整数, 则必存在整数的最大流。

求网络 (两点间) 的最短距离和求网络的最大流 (或最小割) 是一个对偶问题, 这牵涉到网络的变换问题, 本章不予研究。

4.3.3 最大流算法

为了说明寻找最大流的算法, 先介绍增广链的概念。

定义: 设 f 是一个可行流, μ 是从 v_s 到 v_t 的一条链, 满足下列条件的链称为增广链。

$$\left. \begin{aligned} \text{弧}(v_i, v_j) \in \mu^+, 0 \leq f_{ij} < c_{ij} \\ \text{弧}(v_i, v_j) \in \mu^-, 0 \leq f_{ij} < c_{ij} \end{aligned} \right\} \quad (4.15)$$

弧向与走向一致标为 μ^+ , 相反的时候标为 μ^- 。

【例 4.5】图 4-14 中每弧标有两个数字 (c, f) , 第一位数字 c 为容量, 第二位数字 f 为流量。该网络的流量 $F = 4$, 它不一定是最大流。

取从 v_s 到 v_t 的一条链 (如图 4-14(b)), $v_s \rightarrow v_1 \rightarrow v_2 \rightarrow v_4 \rightarrow v_t$, (v_s, v_1) , (v_2, v_4) , (v_4, v_t)

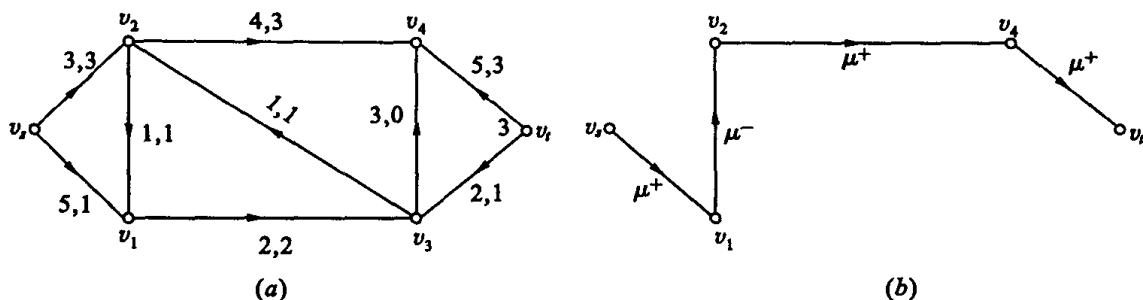


图 4-14 计算最大流的例子

$\in \mu^+, (v_2 v_1) \in \mu^-$, 满足增广链的条件式(4.15)。所以图 4-14(b) 表示的链为增广链。

定理:可行流 f^* 是最大流, 当且仅当不存在关于 f^* 的增广链。

这一定理没有给出最大流, 但它提供了求最大流的搜索方向, 即将最大流问题化为求增广链问题。增广链的实际意义是: 沿着这条链从发点到收点输送一个流, 有潜力可挖。将流量进行调整后可能得到最大流, 这是应用增广链概念的好处。调整各弧的流量后, 各点仍满足平衡条件及容量限制条件。调整方法是: 若调整前流量为 f_{ij} , 设调整后流量为 f'_{ij} 。取 Δ 为适当小的数, 则

$$f'_{ij} = \begin{cases} f_{ij} + \Delta, & (v_i v_j) \in \mu^+ \\ f_{ij} - \Delta, & (v_i v_j) \in \mu^- \\ f_{ij}, & (v_i v_j) \notin \mu \end{cases} \quad (4.16)$$

图 4-14(b), 取 $\Delta = 1$, 则增广链调整前后流量如图 4-15 所示。调整后 f' 满足平衡及容量限制条件, 所以仍是可行流量 $F(f') = 5$ 代替原来的 $F(f) = 4$ 。这里所述的最大流量算法仍是一种标号法。从一个可行流 f 开始, 若网络中没有给定 f , 则可以假设 $f = 0$, 即从零流开始(零流也是可行流)。

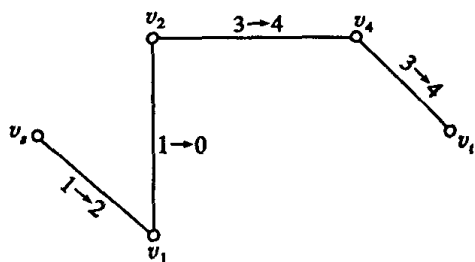


图 4-15 增广链调整前后的流量

然后用标号法求关于可行流 f 的增广链。若增广链存在, 则可以经过调整, 得到一个新的可行流 f' , 其流量 $F(f')$ 比 $F(f)$ 大。下一步重复这一过程, 直到得出某一可行流 f^* , 关于该流的增广链不复存在, 则 f^* 即为最大流。上述求最大流的方法实质上就是通过调整尽量使正向弧的流量等于(或接近于)容量, 而使反向弧流量尽量等于(或接近)0。

算法分两步(见图 4-16):

第一步是标号过程, 用标号法找有无增广链。

第二步是调整过程, 若有增广链, 调整 f 。

现以图 4-14(a) 为例说明算法。取网络已给定的 f 为可行流。

第一步, 标号过程。

(1) 首先给 v_s 标上 $[s, +\infty]$ 。

(2) 检验 v_s , 看有无满足增广链条件的弧, 它有两个相邻点 v_1 及 v_2 , 检查从点 v_s 出来的弧其 f 是否小于 c ?

$f_{s2} = c_{s2} = 3$, 舍弃不用。

$f_{s1} = 1 < c_{s1}$, 在 v_1 点标上 $[+s, 4]$, $+s$ 表示从 v_s 来, 可以由该点向 v_1 增加流量, 4 表示从发点到 v_1 可以增加的流量。

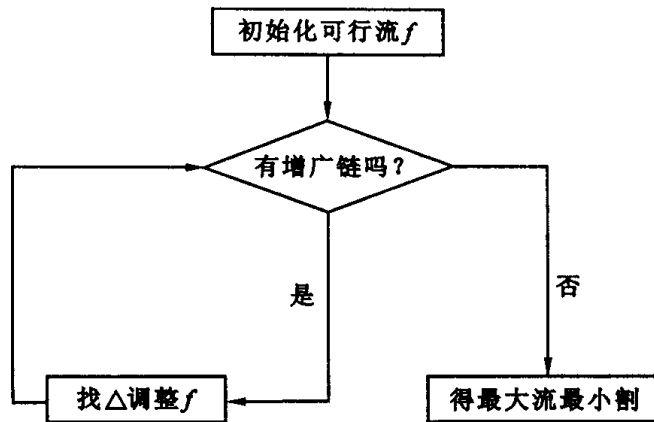


图 4-16 最大流量算法框图

【注】一般情况,对弧 $(v_j v_k)$,若 $0 \leq f_{jk} < c_{jk}$,给 v_k 标上 $[+j, \Delta(k)] = \min[\Delta(l_j), c_{jk} - f_{jk}]$ 。

对弧 $(v_k v_j)$,若 $0 < f_{kj}$,给 v_k 标上 $[-j, \Delta(k)] \Delta k = \min[\Delta(j), f_{kj}]$ 。

(3) 检验 v_1 ,看从 v_1 出去的弧是否满足 $f < c$ 条件,进入 v_1 的弧是否满足 $f > 0$ 的条件。

检验结果:

$(v_1 v_3)$,	$f = c = 2$	舍弃
$(v_1 v_2)$,	v_1 已标号	不考虑
$(v_2 v_1)$,	从 v_2 进入 v_1 ,	$f = 1 > 0$

从 v_2 点标 $[-1, 1]$,其中 -1 表示“到 v_1 去”, 1 表示可以从 f_{21} 减少的流量。

$$\min[4, f_{21}] = \min[4, 1] = 1.$$

(4) 检验 v_2 。 v_2 有 v_3, v_4 两个相邻点。

$f_{24} < c_{24}, f_{32} = 1 > 0$,所以, v_4, v_3 均可标号,分别标以 $[+2, 1] [-2, 1]$ 。

(5) 在 v_4, v_2 两个点中任选一个进行检验。

例如 v_3 其相邻点 v_1 ,未标号,而 $f_{31} < c_{31}$,所以 v_1 可标以 $[+3, 1]$ 。

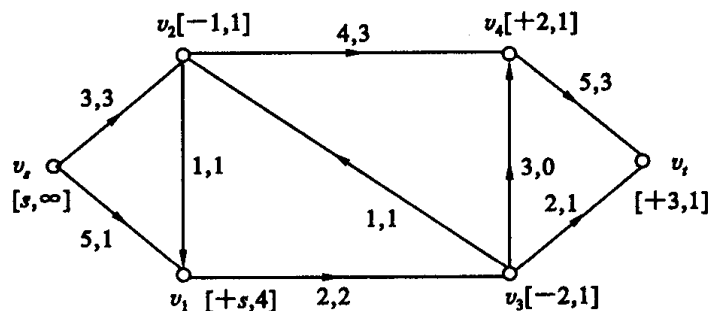


图 4-17 求增广链

至此标号过程完毕,找到一条增广链如图 4-17 所示。其中 $\mu^+ = \{(v_1 v_1), (v_3 v_1)\}$, $\mu^- = \{(v_2 v_1), (v_3 v_2)\}$ 。

第二步,调整过程

取调整量 $\Delta = \min\{4, 1, 1, 1\} = 1$,调整后各弧 f 如图 4-18 得新的可行流。

重复进行标号过程找增广链。

重新给 v_1 标以 $[+s, +\infty]$, v_1 标以 $[+s, 3]$;从 v_1 到 v_3 , $f = c$,从 v_2 到 v_2 , $f = 0$ 。均不符合增广链条件,所以不可能再给其他点标号。

这时 $F(f) = f_{s1} + f_{s2} = f_{3t} + f_{4t} = 5$, f 即为所求的最大流。算法过程至此结束。

用标号法可以很快找到最小割,如图 4-18 所示 f 已是最大流,则该割集容量应为最小,而且最小割 = 最大流 = 5。

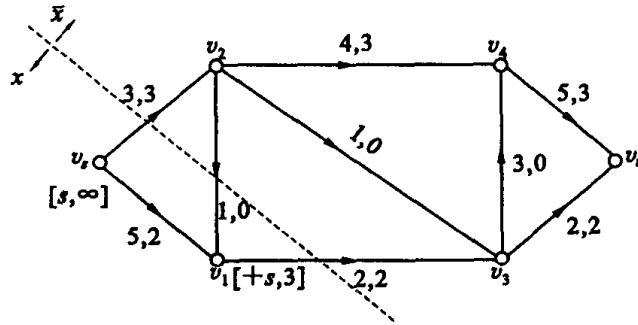


图 4-18 虚线表示最小割

令标号的点组成集合 x , $x = \{v_s, v_1\}$, 未标号的点组成集合 $\bar{x}$, $\bar{x} = \{v_2, v_3, v_4, v_t\}$, 则割集 $(x, \bar{x}) = \{(v_s, v_2), (v_1, v_3)\}$, 弧 (v_2, v_1) 由 $\bar{x}$ 指向 x , 所以不计入割集中。由计算可知 $c(x, \bar{x}) = 3 + 2 = 5$ 。

由上述可见,用标号找增广链以求最大流的结果,同时求得最小割。最小割的意义是:网络中从出发点到收点,有一条必经之路,其容量决定了网络通过能力。最小割则是这条必经之路中的咽喉部分,其容量最小,因此决定了网络的最大通过能力。要提高流量,提高网络运输必须改造这个咽喉部分。

习 题

- 4.1 图 4-19 中数字表示各弧流量,试求最小割与最大流。(提示:先求出所有可能的割集)
- 4.2 用标号法求图 4-20 的最短路。

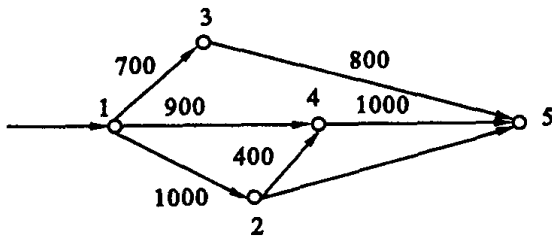


图 4-19

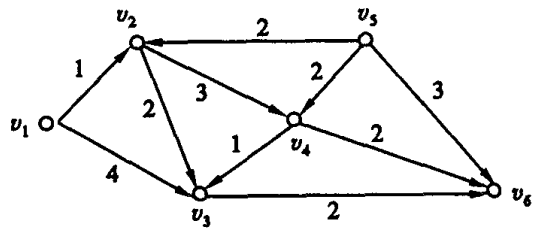


图 4-20

- 4.3 图 4-21 为 v_a 、 v_b 两城市公路网。弧上的数字为每段的距离(里数),求 v_a 、 v_b 间的最短路。

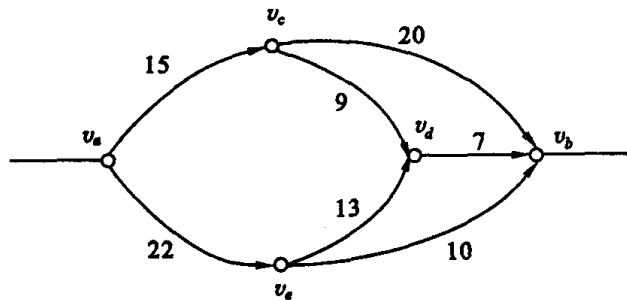


图 4-21

5 系统优化

本章提要

系统优化问题是在制约条件下选择准则函数最优解的问题。前面几章介绍了一些传统系统优化的方法,例如线性规划、非线性规划、动态规划、图论等一系列全局优化的方法等。本章侧重于介绍现代优化算法和大系统的分解与协调。重点解决系统模型过于复杂无法用明确的解析方程来描述的问题。这是一种多学科综合性的解决问题方法,例如启发式算法和遗传算法、神经网络算法等。另一方面,处理大系统优化的思想就是对大系统进行分解使被处理的系统降阶、降维,达到易于处理的目的。大系统分解成若干个子系统之后,优化的困难在于各子系统之间存在的耦合。正是这些关联,在分解后必须在总体目标和约束下进行协调。

5.1 系统优化方法概述

在系统设计和系统规划等许多领域中,人们常常要面临这样的一类问题:工程设计中怎样选择参数,使得设计既能满足要求,又能降低成本;生产计划安排中,怎样选择生产计划方案才能提高产值和利润;在城市建设规划中,怎样安排工厂、学校、机关、商店、住宅和其他公用设施,才能方便群众,使城市更加合理有效的运行。这一类问题归结起来都是:在一定制约条件下,选择最合理的、达到目标(准则)解的最优。这就成为系统的优化问题。

优化问题是一个古老的课题。早在 17 世纪在欧洲就有人提出了各种求最大和最小的问题,以及某些相应的法则。但在其后的二、三百年间,这个方面的发展比较缓慢。直到 20 世纪 40 年代,由于生产和科学研究突飞猛进的发展,也由于军事上的需要,推动了运筹学的发展,从而产生了如线性规划、非线性规划、动态规划、图论等一系列全局优化的方法。其后经过几十年的努力,特别是启发算法和现代优化算法的出现,使得最优化理论和算法也迅速发展起来,形成了一个新的学科,并使优化理论得到了更加广泛的应用。

为了将系统优化的方法应用于具体系统中,传统的优化算法必须定义被优化系统的性能指标和约束条件;必须选择代表优化因素的独立变量;写出表示各种变量之间关系的数学模型。而现代优化算法是要解决优化问题中的难解问题,或者说是系统模型过于复杂无法用明确的解析方程来描述的问题。它是一种多学科综合性的解决问题方法。

5.2 启发式算法和现代优化算法

传统的优化算法对很多特定问题的求解给出了非常好的方法。但随着技术水平的提高和现代信息和网络技术的普及应用,许多更加复杂的问题已经无法用传统的优化方法来解决。随着 20 世纪 80 年代初期模拟退火、遗传算法和人工神经网络算法的兴起,科学工作

者对这些复杂问题的优化问题和各种新的算法进行了深入的分析研究,并将这些算法称为现代优化算法。现代优化算法的主要应用对象是优化问题中的难解问题,或者是系统模型过于复杂无法用明确的解析方程来描述的问题。正是因为许多实际优化问题的难解性和现代优化算法在一些实际优化问题中的成功应用,使得现代优化算法成为解决优化问题的一种有利工具。

现代优化算法涉及多学科的知识,如生物进化、人工智能、数学和物理科学、神经系统和统计力学等概念。它们都是一定的直观基础构造的算法,我们称之为启发式算法。启发式算法是相对于最优化算法提出的。一个问题的最优化算法可以求得该问题每个实例的最优解,而启发式算法是一种技术。这种技术使得在可接受的计算费用内去寻找最好的解,但不一定保证所得解的最优性。甚至在多数情况下,无法阐述所得解同最优解的近似程度,也就是说利用启发式算法所寻找的最好解与最优解的差距,因为这种问题的最优解在大部分情况下是无法明确知道的。

在某些情况下,特别是实际问题中,最优算法的计算时间使人无法忍受或因问题的难度使其计算时间随问题规模的增加以指数速度增加,此时只能通过启发式算法求得问题的一个可行解。

启发式算法的特点是不考虑算法所得解与最优解的偏离程度。如果采用分析最坏情况的误差界限的概念来评价启发式算法,则可以将近似算法定义如下:

定义 5.1 设 A 是一个问题。记问题 A 的任何一个实例 I 的最优解和启发式算法 H 的解的目标值分别为 $Z_{OPT}(I)$ 和 $Z_H(I)$ 。于是对于某个 $\epsilon \geq 0$,称 H 是 A 的 ϵ 近似算法,当且仅当

$$|Z_H(I) - Z_{OPT}(I)| \leq \epsilon |Z_{OPT}(I)|, \forall I \in A \quad (5.1)$$

上面的定义给出启发式算法和近似算法两个概念的联系和区别。启发式算法定义的算法集合包含了近似算法概念定义的算法集合。近似算法强调给出算法最坏情况的误差界限,而启发式算法不需考虑偏差程度。由于很多组合优化问题算法的最坏情况误差估计需要很强的数学基础和较强的技巧,甚至一些问题很难或无法给出最坏情况误差界,而实际问题又迫切要求求解的方法,因此,只能通过启发式算法解决问题。

早在 20 世纪 40 年代末期,由于实际问题的需要,人们已提出一些解决实际问题的快捷有效的启发式算法,有代表性的工作是 1948 年 Polya 的著作。随着 20 世纪 60~70 年代对数学模型及最优解算法的重视,这些算法被称为“快速与丑陋(quick and dirty)”方法。随着 20 世纪 70 年代算法复杂性理论的完善,我们不再强调一定要求到最优解,因此促使 20 世纪 80 年代初兴起的现代优化算法在今天得到了巨大的发展。

让我们通过下面的例子对启发式算法有一定的了解。

【例 5.1】 背包问题的贪婪算法(greedy algorithm)

有一个容积为 b 的背包, n 个体积分别为 $a_i (i=1,2,\dots,n)$,价值分别为 $c_i (i=1,2,\dots,n)$ 的物品,如何以最大的价值装包?这个问题称为背包问题。对背包问题可以构造下面的贪婪算法

STEP1 对物品以价值密度 $\frac{c_i}{a_i}$ 从大到小排列,不妨把排列记成 $\{1,2,\dots,n\}, k=1$;

STEP2 若 $\sum_{i=1}^{k-1} a_i x_i + a_k \leq b$, 则 $x_k = 1$; 否则, $x_k = 0$ 。即 1 为装入, 0 为舍弃。 $k = k+1$;

当 $k = n \pm 1$ 时, 停止; 否则, 重复 STEP2。

$(x_1, x_2, \dots, x_n)$ 为贪婪算法所得解。单位体积价值比越大越先装包是贪婪算法的原则。这样的算法非常直观, 非常容易操作。

【例 5.2】简单的领域搜索(local search) 算法

给定组合优化问题, 假设其领域结构已确定, 设 S 为解集合, f 为 S 上的费用函数, N 为领域结构, 算法为:

STEP1 任选一个初始解 $s_0 \in S$;

STEP2 在 $N(s_0)$ 中按某一规则选一 s ; 若 $f(s) < f(s_0)$, 则 $s_0 = s$; 否则 $N(s_0) = N(s_0) - s$; 若 $N(s_0) = \Phi$, 停止; 否则, 返回 STEP2。

简单的领域搜索从任何一点出发, 达到一个局部最优值点。从算法中可以看出, 算法停止时得到点的性质依赖算法初始解的选取、领域的结构和领域选点的规则。

启发算法能够迅速发展是因它有以下长处: 数学模型本身是实际问题的简化, 或多或少地忽略一些因素; 数据采集具有不精确性; 参数估计具有不准确性; 以上因素可能使最优算法所得解比启发式算法所得解产生更大误差; 有些难的组合优化问题可能还没有找到最优算法, 即使存在, 由算法复杂性理论, 得到它们的计算时间是无法接受或不实际的; 一些启发式算法可以用在最优算法中, 如在分枝定界算法中, 可以用启发式算法估界; 简单易行; 比较直观; 易被使用者接受; 速度快, 在适时管理中非常重要; 多数情况下, 程序简单, 因此易于修改。

虽然有诸多好处, 但它有其短处, 这些不足往往成为争论的焦点: 不能保证求得最优解; 表现不稳定, 启发式算法在同一问题的不同实例计算中会有不同的效果。有些解很好, 而有些则很差。在实际应用中, 这种不稳定性造成计算结果不可信。

启发式算法由多种算法构成, 我们可以将其简单的分类为:

(1) 一步算法

一步算法的一个典型实例是背包问题的贪婪算法。每一步迭代选一物品入包, 直到无法再装, 该算法没有两个可行解之间比较选择, 算法结束时得到一可行解。

(2) 改进算法

改进算法的迭代过程是从一个可行解到另一个可行解, 通常通过对两个解的比较而选择好的解, 进而作为新的起点进行新的迭代, 直到满足一定的要求为止。因此, 也可以称之为迭代算法。

(3) 数学规划算法

数学规划算法主要指用线性规划的方法求解组合优化问题, 其中包括一些启发式规划。这一类方法中, 典型的是线性规划及对偶理论在网流中的应用, 产生标号算法等一系列最优算法。其次是基于整数规划分枝定界的启发式算法, 只搜索一些特殊分支, 或是基于整数规划的割平面的算法。

(4) 现代优化算法

现代优化算法是 20 世纪 80 年代初兴起的启发式算法。这些算法包括禁忌搜索(tabu search), 模拟退火(simulated annealing), 遗传算法(genetic algorithms), 人工神经网络(neural networks)。它们主要用于解决大量的实际应用问题。目前, 这些算法在理论和实际应用方面都得到了较大的发展。虽说启发式算法有诸多优点, 但它的一个缺陷是无法保证得

到最优解,于是,对算法的评价显得尤为重要。一个好的启发式算法可以使其解同最优解尽可能地接近,同时保证有较好的稳定性。评价启发式算法的性能有不同的方法,主要分为对算法计算复杂性的分析和对算法计算效能的分析。可以采用最坏情形分析方法和概率分析方法对问题进行较深入的理论分析。这些理论分析方法要求有很强的数学基础和推演能力。正因为如此,目前研究才只对一些特殊和简单的问题采用上面两种理论分析方法。实际中大量的组合优化问题是采用大规模数值计算方法。简单地说,大规模计算分析就是通过大量的实例计算,评价算法的计算效果。算法的计算效果分成两个方面:一方面是算法的计算复杂性,它的效果通过计算机的中央处理器(CPU)的计算时间表现;另一个方面是计算解的性能,它通过计算停止时输出的解表现。

大规模计算分析可对一个算法进行评价。对一个算法进行评价时,它的计算时间效果往往通过目前的计算机设备能否承受、用户能否接受现有的计算时间来衡量。对它的计算解进行评价时,一个简单的要求是用户是否满意,更深一步需要知道问题的最优解或问题的一个下限(假设目标为最小),这时,可以通过数据的计算比较一个算法的计算效果。

大规模计算分析可对多个算法进行评价,比较分析不同算法的效果。通过大量数据的计算,根据各个算法的计算结果,采用简单或统计的方法比较不同算法的性能。这种比较不需要已知问题的最优解或上下界。

5.3 遗传算法

本节简单介绍一种较新的现代优化算法——遗传算法。遗传算法主要借用生物进化中“适者生存”的规律。“适者生存”揭示了大自然生物进化过程中的一个规律:最适合自然环境的群体往往产生了更大的后代群体。因此有必要在介绍遗传算法之前,先简单了解生物进化的基本过程,见图 5-1。

以这个循环圈的群体(population)为起点,经过竞争后,一部分群体被淘汰而无法进入这个循环圈,而另一部分则成为种群(reproduction)。优胜劣汰在这个过程中起着非常重要的作用,这在自然界显得更加突出。因为自然天气的恶劣和天敌的侵害,大自然中的很多动物的成活率是非常低的。即使在成活群体,还要通过竞争产生种群,种群通过婚配的作用产生子代群体(简称子群)。进化的过程中,可能会因为变异而产生新的个体。综合变异(mutation)的作用,子群成长为新的群体而取代旧群体,在新的一个循环过程中,新的群体将替代旧的群体而成为循环的开始。

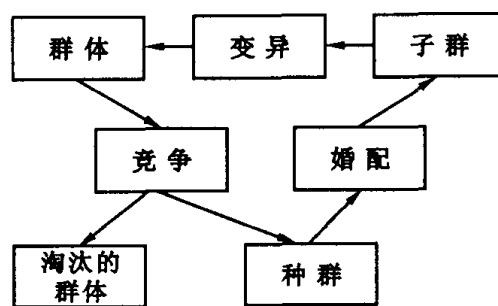


图 5-1 生物进化循环图

以人类进化为例,人类群体在自然竞争、淘汰后,形成了繁殖后代的群体,即成人群体。群体的每一个体是一个人,每个人包含有 46 条染色体(chromosome),组成 23 对同源染色体。男女性结合使得对应的 23 对染色体优胜劣汰再产生 23 对染色体,因而形成一个新的生命。以性别染色体为例,母亲的性染色体中由 $X(1)$ 和 $X(2)$ 基因(gene)组成,父亲的性染色体由 $X(m)$ 和 Y 基因组成。两个染色体的结合是这两对基因竞争的产物,竞争的结局可能是 $\{X(1), X(m)\}$ (女), $\{X(1), Y\}$ (男), $\{X(2), X(m)\}$ (女), $\{X(2), Y\}$ (男), 这一新的结合物

为子代的一对性染色体。每一条染色体是由特定的基因组成。在自然进化中,会出现某些基因的变异,形成一个新的个体。

遗传算法主要借鉴了生物进化的一些特征,其主要特征体现为:

进化发生在解的编码上。这些编码按生物学的术语称为染色体。由于对解进行了编码,优化问题的一切性质都通过编码来研究。编码和解码是遗传算法的一个主题;自然选择规律决定哪些染色体产生超过平均数的后代,遗传算法中,通过优化问题的目标而人为地构造适应函数以达到好的染色体产生超过平均数的后代;当染色体结合时,双亲的遗传基因的结合使子女保持父母的特征;当染色体结合后,随机的变异会造成子代同父代的不同。

遗传算法包含以下的主要处理步骤。首先是对优化问题的解进行编码,此外,我们称一个解的编码为一个染色体,组成编码的元素称为基因。编码的目的主要用于优化问题解的表现形式和利于之后遗传算法中的计算。第二是适应函数的构造和应用。适应函数基本上依据优化问题的目标函数而定。当适应函数确定以后,自然选择规律是以适应函数值的大小决定概率分布来确定哪些染色体适应生存,哪些被淘汰。生存下来的染色体组成种群,形成一个可以繁衍下一代的群体。第三是染色体的结合。双亲的遗传基因结合是通过编码之间的交配(crossover)达到下一代的产生。新一代的产生是一个生殖过程,它产生了一个新解。最后是变异,新解产生过程中可能发生基因变异,变异使某些解的编码发生变化,使解有更大的遍历性。表 5-1 列出了生物遗传基本概念在遗传算法中作用的对应关系。

表 5-1 生物遗传概念在遗传算法中的对应关系

生物遗传概念	在遗传算法中的作用
适者生存	在算法停止时,最优目标值的解有最大的可能被留住
个体(individual)	解
染色体(chromosome)	解的编码(字符串,向量等)
基因(gene)	解中每一分量的特征(如各分量的值)
适应性(fitness)	适应函数值
群体(population)	选定的一组解(其中解的个数为群体的规模)
种群(reproduction)	根据适应函数值选取的一组解
交配(crossover)	通过交配原则产生一组新解的过程
变异(mutation)	编码的某一个分量发生变化的过程

最优化问题的求解过程是从众多的解中选出的最优的解。生物进化的适者生存规律使得最具有生存能力的染色体以最大的可能生存。这样的共同点使得遗传算法可以在优化问题中应用。我们以一个简单的例子来理解表 5-1。

【例 5.3】 用遗传算法求解 $\max f(x) = x^2, 0 \leq x \leq 31, x$ 为整数的最大值。

一个简单的表示解的编码是二进制编码,即 0,1 字符串。由于变量的最大值是 31,因此可以采用 5 位数的二进制码,如

10000 $\rightarrow$ 16 11111 $\rightarrow$ 31 01001 $\rightarrow$ 9 00010 $\rightarrow$ 2,

以上的 5 位字符串称为染色体。每一个分量称为基因,每个基因有两种状态 0 或 1。模拟生物进化,首先要产生一个群体,可以随机取 4 个染色体组成一个群体,如 $x_1 = (00000)$, $x_2 = (11001)$, $x_3 = (01111)$, $x_4 = (01000)$ 。群体有 4 个个体,适应函数可以依据目标函数而定,如适应函数 $\text{fitness}(x) = f(x) = x^2$,于是

$$\begin{aligned}\text{fitness}(x_1) &= 0, \text{fitness}(x_2) = 25^2, \\ \text{fitness}(x_3) &= 15^2, \text{fitness}(x_4) = 8^2.\end{aligned}$$

定义第 i 个个体入选种群的概率为

$$p(x_i) = \frac{\text{fitness}(x_i)}{\sum_j \text{fitness}(x_j)} \quad (5.2)$$

于是,适应函数值大的染色体个体的生存概率自然较大,若群体中选 4 个个体成为种群,则极有可能竞争上的是 $x_2 = (11001)$, $x_2 = (11001)$, $x_3 = (01111)$, $x_4 = (01000)$ 。若它们结合,采用如下的交配方式,称为简单交配

$$\begin{aligned}x_2 = (11 | 001) & \left. \begin{array}{l} \\ \\ \end{array} \right\} \rightarrow y_1 = (11 | 111) \\ x_3 = (01 | 111) & \left. \begin{array}{l} \\ \\ \end{array} \right\} \rightarrow y_2 = (01 | 001) \\ x_2 = (11 | 001) & \left. \begin{array}{l} \\ \\ \end{array} \right\} \rightarrow y_3 = (11 | 000) \\ x_4 = (01 | 000) & \left. \begin{array}{l} \\ \\ \end{array} \right\} \rightarrow y_4 = (01 | 001)\end{aligned}$$

即交换第二个位置以后的基因,得到 y_1, y_2, y_3 和 y_4 。若 y_4 的第一个基因发生变异,则变成 $y_4 = (11001)$ 。

通过例 5.3,我们可以将求解组合优化问题的遗传算法简化地描述为:

遗传算法

STEP1 选择问题的一个编码,给出一个有 N 个染色体的初始群体 $\text{pop}(1)$, $t = 1$;

STEP2 对群体 $\text{pop}(t)$ 中的每一个染色体 $\text{pop}_i(t)$ 计算它的应适应函数

$$f_i = \text{fitness}(\text{pop}_i(t))$$

STEP3 若停止规则满足,则算法停止;否则,计算概率

$$p_i = \frac{f_i}{\sum_{j=1}^N f_j}, i = 1, 2, \dots, N, \quad (5.3)$$

并以概率分布(4.1)从 $\text{pop}(t)$ 中随机选一些染色体构成一个种群

$$\text{newpop}(t+1) = \{\text{pop}_j(t) \mid j = 1, 2, \dots, N\};$$

【注】 $\text{newpop}(t+1)$ 集中可能重复选 $\text{pop}(t)$ 中的一个元素,如例 5.3 中的 x_2 就选取两次。

STEP4 通过交配,交配概率为 p_c ,得到一个有 N 个染色体的 $\text{crosspop}(t+1)$;

STEP5 以一个较小的概率 p ,使得一个染色体的一个基因发生变异,形成 $\text{mutpop}(t+1)$;

$t = t + 1$, 一个新的群体 $\text{pop}(t) = \text{mutpop}(t)$, 返回 STEP2。

种群的选取方式每次都可能不同,在计算过程中无法预测哪一次的选择会获得好的结果,因而在某种程度上称为轮盘赌,以下再通过例 5.4 理解遗传算法。

【例 5.4】 用遗传算法求 $\max f(x) = 1 - x^2, x \in [0, 1]$

由于对连续变量求解,要解决的一个问题是如何编码,假设对解的误差要求是 $1/16$,则可以采用 4 位二进制编码,对应关系为

$$(abcd) \leftrightarrow \frac{a}{2} + \frac{b}{4} + \frac{c}{8} + \frac{d}{16}$$

若计算的一步如表 5-2。

表 5-2 遗传算法中的一步计算

x 值	旧群体 pop(1)	适应函 数值 f	概率分布 P	new pop	交配位	cross pop	是否 变异	mut pop	x 值
1/16	0001	0.996	0.318	0001	00 01	0000	N	0000	0
1/4	0100	0.938	0.299	0100	01 00	0101	Y	1101	13/16
3/16	0011	0.965	0.308	0001	000 1	0001	N	0001	1/16
7/8	1110	0.234	0.075	0011	001 1	0011	N	0011	3/16

平均值 = 0.7833, 交配概率 $p = 1$, 变异概率 $p_m = 0.02$

在表 5-2 中, newpop 只选择 4 个染色体, 交配方法采用随机选择一对染色个体按交配位进行交叉, 如 0001 和 0100 的交配位随机选在位置 2, 经过交配产生 0000 和 0101 两个个体。在算法中, 0001 和 0011 的交配位随机选在位置 3, 经过交配产生 0001 和 0011 两个个体, 这样的交配称为简单交配。交配后的染色体组成 crosspop。在 crosspop 中可能有些染色体会产生变异, 简单的变异是 crosspop 中每一个基因都以同样一个概率变异。如 0101 第一位的基因发生变异。变异后, 得到 mutpop, 也就是新的群体。

简单遗传算法可以理解为: 求解的问题是极大目标函数的优化问题; 采用 0—1 二进制编码; pop(t) 中的染色体个数是一个常数; 初始群体随机选取; 适应函数为目标函数; 按轮盘赌方法选取染色体个数同 pop(t) 相同的种群; 交配按常规交配方法——一对染色体按随机位交换后的基因, 染色体中的每一个基因都以相同的概率变异。

以上只对遗传算法有一个直观的了解, 但对遗传算法的应用来说, 应该侧重于如下主要因素来进行研究。

(1) 解的编码和解码。遗传算法的基础工作之一是解的编码, 只有在编码之后才可能有其他的计算。例 5.3 和例 5.4 采用二进制编码。编码和解码是相对应的。算法的最后一个工作是通过解码得到问题的一个解。

(2) 初始群体的选取和计算中群体的大小。一般采用随机产生初始群体或通过其他方法先构造一个初始群体。通过其他方法构造的初始群体可能会节省进化的代数, 但也可能过早地陷入局部最优群体中。我们称过早地陷入局部最优群体中的现象为早熟 (premature) 现象。群体中个体的个数称为群体的维数。群体的维数越大, 其代表性越广泛, 最终进化到最优解的可能性越大。但维数大的群体势必造成计算时间的增加, 这又是我们不希望的。群体的维数常常采用一个不变的常数, 在一些应用中, 群体的维数可以采用同遗传代数有关的变量, 以使算法更有效。

(3) 适应函数的确定。一般情况, 适应函数同目标函数相关, 以保证较优的解有较大的生存机会。

(4) 三个算子。遗传算法的三个算子是: 种群选取、交配和变异。新群体产生中的一个主

要问题是如何选取种群。上面的遗传算法中已经介绍,种群是以一个概率分布——轮盘赌的形式选择个体而产生的。种群选定后需考虑它的交配规则。交配规划较多,如双亲遗传法,主个体副个体规则,单亲遗传规则等。还需考虑交配位的选取、交配概率及产生后代等一些细节问题,变异是扩大染色体选择范围的一个手段。通常,遗传算法实现变异的方法是赋予每一个基因一个相对比较小的变异概率,通过随机模拟而决定在一个基因是否变异。变异概率过小使解有一定的局限性,遍历性较差。变异概率较大使得进化的随机性增大,也不容易得到稳定的解。因此,不可忽略变异概率如何确定这一过程。

遗传算法的优越性可以简单地归结为以下三条:

(1) 遗传算法适合数值求解那些带有多参数、多变量、多目标和在多区域但连续性较差的 NP-hard 优化问题。对参数、多变量的 NP-hard 优化问题,通过解析求解或是计算求最优的可能性很小,主要依赖于数值求解。遗传算法是一种数值求解的方法,是一个有普适性的方法,对目标函数的性质几乎没有要求,甚至都不一定要写出目标函数,因此用遗传算法求解优化问题不足为奇。遗传算法所具有的特点是记录一个群体,它可以记录多个解而不同于局部搜索、禁忌搜索和模拟退火仅仅是一个解,这多个解的进化过程正好适合于多目标优化问题的求解。

(2) 遗传算法在求解很多组合优化问题时,不需要有很强的技巧和对问题有非常深入的了解。如排序、路线调度问题、布局问题等。遗传算法在给问题的决策变量编码后,其计算过程是比较简单的,且可以较快得到一个满意解。

(3) 遗传算法同求解问题的其他启发式算法有较好的兼容性。如可以用其他的算法求初始解;在每一群体,可以用其他的方法求解下一代新群体。

遗传算法作为一种先进的全局优化算法,已经显示出其强大的生命力,并已经在许多学科领域得到了应用,利用遗传算法求解最优化问题还有很多工作要做。在此只是提供一种初步的介绍。

5.4 大系统的分解与协调

大系统理论是 20 世纪 60 年代末才逐步建立起来的。对“大系统”(Large Scale System)至今并没有统一的定义,系统多大才算“大”?也不能给出准确的回答。按一般理解是,系统的变量、阶次及约束条件的规模大一些就称其为大系统。处理大系统优化问题就称其为“大系统理论”。

处理大系统优化的思路就是对大系统进行分解(Decomposition),使被处理的系统降阶、降维,达到易于处理的目的。分解就是把总系统“切割”成若干个子系统,这样各个子系统的问题规模就变小了。这种“切割”有时是自然划分的,比如进行整个企业的优化,可以按车间划分,在各个车间优化基础上,实现整个企业的优化;当然也可以按照其他原则划分。

大系统分解成若干个子系统之后,优化的困难在于各子系统之间存在的耦合。正是这些关联,在分解后必须在总体目标和约束下进行协调(Coordination)。所以大系统优化中的核心概念是大系统的分解/协调。

为了说明分解/协调的概念和大系统分解/协调的方法,我们用一个例子进行分析。这个例子虽然是线性的,变量不多,但是它却能揭示大系统分解/协调方法的本质。

【例 5.5】 某总公司下属两个分公司,它们分别生产两种产品,第一个子公司产量为 x_1 ,第二个子公司产量为 x_2 ,每种产品单位产量的产值均为 10 个单位,两个子公司的生产能力限制分别为 4 和 3。

问:如何安排生产计划,使全公司的产值最大?

【解】 这个问题的生产优化数学模型如下:

$$\begin{aligned} \max Z &= 10x_1 + 10x_2 \\ \text{s. t. } \quad &\left. \begin{aligned} x_1 &\leq 4 \\ x_2 &\leq 3 \\ x_1, x_2 &\geq 0 \end{aligned} \right\} \end{aligned} \quad (5.4)$$

总公司为了便于领导,将该生产计划分解、分别下达给两个子公司,也是容易实现的,如果全公司的产值为 Z ,分解到两个子公司的产值分别为 Z_1 和 Z_2 ,分解过程如图 5-2 所示。

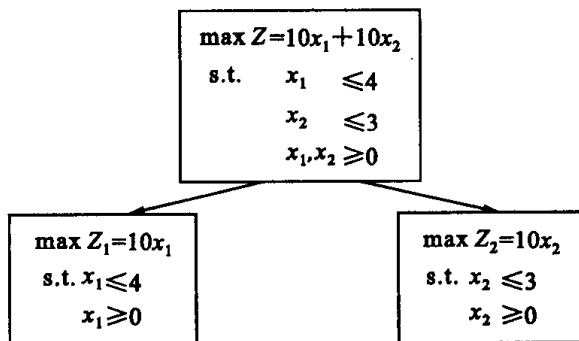


图 5-2 生产计划分解

在这种情况下,系统可以分解为两个独立的子系统,子系统的最优解,就构成整个系统的最优解,即: $x_1 = 4, x_2 = 3$, 总产值 $Z = Z_1 + Z_2 = 70$ 。

再问:如果原材料供应紧张,生产 1 个单位 x_1 消耗 1 个单位的原材料;生产 1 个单位 x_2 消耗 2 个单位的原材料,总原材料限制是 8 个单位,这时又该如何安排生产?

【解】 在增加原材料供应的约束情况下,生产优化数学模型如下:

$$\begin{aligned} \max Z &= 10x_1 + 10x_2 \\ \text{s. t. } \quad &\left. \begin{aligned} x_1 &\leq 4 \\ x_2 &\leq 3 \\ x_1 + 2x_2 &\leq 8 \\ x_1, x_2 &\geq 0 \end{aligned} \right\} \end{aligned} \quad (5.5)$$

在增加原材料约束情况下,不可能分解为两个独立子系统,原因是有耦合约束 $x_1 + 2x_2 \leq 8$ 。

那么,进一步问:系统还能否进行分解,分解为两个子系统?回答是,可以。在这种情况下,分解为子系统特征是什么?因为有关联,有耦合,分解后的两个子系统不独立。再问:分解为两个不独立的子系统,能否找到系统最优解?可以。那么如何找出系统的最优解?回答是,在上级协调下,即上级系统对系统进行协调下获得总体系统的最优解。

下面介绍上级如何对下级子系统进行协调,让我们仍以上例进行说明。

方法 1:对耦合条件进行了解 / 协调

这种方法的思路对耦合条件进行分解与协调。既然总资源约束是 8,那么就把 8 分成 b_1

和 b_2 , 分 b_1 给第一个子公司, 分 b_2 给第二个子公司, 这时总系统就分解为两个“独立”的子系统, 如图 5-3 所示:

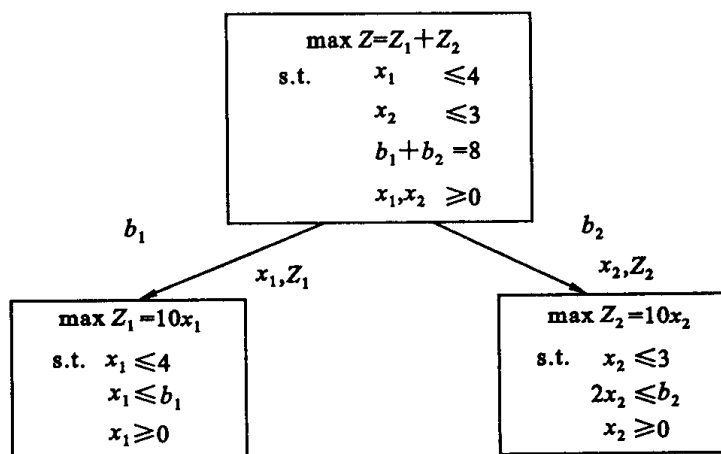


图 5-3 生产计划分解

现在的问题转化为如何确定 b_1, b_2 , 使系统的总目标 $Z \rightarrow \max$ 。它的实质就是资源优化分配问题。让我们分析如下的 b_1, b_2 分配方案:

方案 1 $b_1 = 2 \quad b_2 = 6$

则两个子公司生产优化结果分别是:

$$x_1 = 2, Z_1 = 20$$

$$x_2 = 3, Z_2 = 30$$

$$\text{得: } Z = Z_1 + Z_2 = 50$$

方案 2 $b_1 = 3 \quad b_2 = 5$

则两个子公司生产优化结果分别是:

$$x_1 = 3, Z_1 = 30$$

$$x_2 = 2.5, Z_2 = 25$$

$$\text{得: } Z = Z_1 + Z_2 = 55$$

方案 3 $b_1 = 4 \quad b_2 = 4$

则两个子公司生产优化结果分别是:

$$x_1 = 4, Z_1 = 40$$

$$x_2 = 2, Z_2 = 20$$

$$\text{得: } Z = Z_1 + Z_2 = 60$$

从上述方案中可以看出, 增大 b_1 的配额(相应就减少 b_2 的配额)是会使系统总体收益提高, 那么我们再提高 b_1 配额, 看看会有什么结果。

方案 4 $b_1 = 5 \quad b_2 = 3$

则两个子公司生产优化结果分别是:

$$x_1 = 4, Z_1 = 40$$

$$x_2 = 1.5, Z_2 = 15$$

$$\text{得: } Z = Z_1 + Z_2 = 55$$

从这种方案分析可以看出, 如在方案 3 基础上进一步提高 b_1 (即同时降低 b_2), 则总体系

统收益反而会降低。

通过上述分析,可以得到结论,方案 3 的资源分配方案是最优的方案,即:

$$\begin{cases} \text{资源分配方案: } b_1 = 4, b_2 = 4 \\ \text{生产安排: } x_1 = 4, x_2 = 2 \\ \text{分系统收益: } Z_1 = 40, Z_2 = 20 \\ \text{总系统收益: } Z = Z_1 + Z_2 = 60 \end{cases}$$

方法 2 对目标函数进行分解 / 协调

这种方法不是约束条件进行分解与协调,而是根据耦合关系,对目标函数进行分解 / 协调。其思路是,把耦合约束条件以待定系数 k ,合并到目标函数中,对于这个例子就是:

$$\text{耦合约束条件: } x_1 + 2x_2 \leq 8$$

$$\text{移项得: } 8 - x_1 - 2x_2 \geq 0$$

$$\text{乘待定系数为: } k(8 - x_1 - 2x_2) \geq 0$$

把它合并到目标函数中,得到:

$$\max Z = 10x_1 + 10x_2 + 8k - kx_1 - 2kx_2$$

因为 $8k$ 是常数,所以上式成为

$$\max Z = (10x_1 - kx_1) + (10x_2 - 2kx_2)$$

上述方法把耦合约束条件并到目标函数中,其实就是把约束条件中的耦合约束条件去掉,给分解子系统带来了方便。这时问题转换成为,在满足总的资源条件下,选择待定系数 k ,使子系统达到最优,此时就是总体系统最优解。这个问题分解与协调过程示于图 5-4。

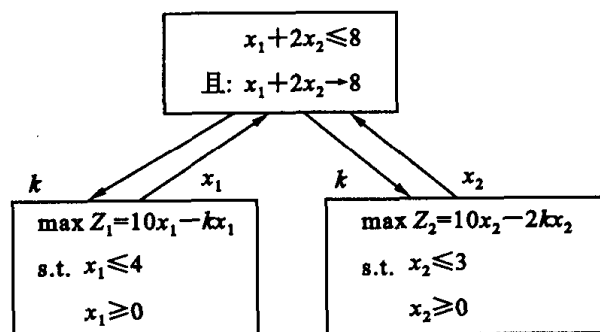


图 5-4 生产计划分解

从上图可见,形式上分解为两个“独立”的子系统,关键是如何选定参数 k ,我们仍是设定不同的方案来进行分析。

方案 1 $k = 4$

对于第一个子系统有:

$$\begin{cases} \max Z_1 = 10x_1 - 4x_1 = 6x_1 \\ \text{s. t. } x_1 \leq 4 \\ x_1 \geq 0 \end{cases} \quad (5.6)$$

得解为: $x_1 = 0$

对于第二个子系统有:

$$\begin{cases} \max Z_2 = 10x_2 - 8x_2 = 2x_2 \\ \text{s. t. } x_2 \leq 3 \\ x_2 \geq 0 \end{cases} \quad (5.7)$$

得解为: $x_1 = 4, x_2 = 3$

检查: $x_1 + 2x_2 = 4 + 2 \times 3 = 10 > 8$

结论: 当选定 $k = 4$, 不能满足约束条件, 即资源不够。

方案 2 $k = 6$

对于第一个子系统有:

$$\begin{cases} \max Z_1 = 10x_1 - 6x_1 = 4x_1 \\ \text{s. t. } x_1 \leq 4 \\ x_1 \geq 0 \end{cases} \quad (5.8)$$

得解为: $x_1 = 4$

对于第二个子系统有:

$$\begin{cases} \max Z_2 = 10x_2 - 12x_2 = -2x_2 \\ \text{s. t. } x_2 \leq 3 \\ x_2 \geq 0 \end{cases} \quad (5.9)$$

得解为: $x_2 = 0$ (由于目标函数为 $-2x_2$, 生产越多越赔钱, 所以第二个子公司只能作出不生产的决定, 即: $x_2 = 0$)

检查: $x_1 + 2x_2 = 4 + 2 \times 0 = 4 < 8$

结论: 当选定 $k = 6$, 满足约束条件, 但是总资源利用为 4, 远远未逼近 8, 说明设定 $k = 6$ 时, 资源未充分利用。

这时全公司总产值是:

$$Z = Z_1 + Z_2 = 4 \times 10 + 0 \times 10 = 40$$

方案 3 $k = 5$

对于第一个子系统有:

$$\begin{cases} \max Z_1 = 10x_1 - 5x_1 = 5x_1 \\ \text{s. t. } x_1 \leq 4 \\ x_1 \geq 0 \end{cases}$$

得解为: $x_1 = 4$

对于第二个子系统有:

$$\begin{cases} \max Z_2 = 10x_2 - 2 \times 5x_2 = 0 \\ \text{s. t. } x_2 \leq 3 \\ x_2 \geq 0 \end{cases}$$

得解为: 在这种情况下, $\max Z_2 = 10x_2 - 2kx_2 = 0$, 而 x_2 取 0, 1, 2, 3 都是可以的。

检查: 由于目标是

$$x_1 + 2x_2 \leq 8$$

且: $x_1 + 2x_2 \xrightarrow{\text{尽量逼近于}} 8$

所以 x_2 应取为 2, 即 $x_2 = 2$

结论:在满足目标约束条件下总体系统最优解是:

$$\begin{cases} x_1 = 4 \\ x_2 = 2 \end{cases}$$

系统总体收益为:

$$Z = Z_1 + Z_2 = 4 \times 10 + 2 \times 10 = 60$$

方法 2 的实质是什么?把约束条件乘以一个待定系数 k 合并到目标函数中,如本例的求解: $Z = 10x_1 + 10x_2 - k(x_1 + 2x_2 - 8) \rightarrow \max$ 的目的是,找到一个 k ,它尽量满足约束条件: $k(x_1 + 2x_2 - 8) \rightarrow 0$,注意: $k(x_1 + 2x_2 - 8) \geq 0$,因此,在上述目标函数中减去的正数为 0,自然就使 $Z \rightarrow \max$ 。

再进一步问: k 的物理意义是什么? k 的物理意义就代表该资源的“影子价格”。在上例中,方案一把 k 设定为 4,相当是把资源价格定得低了,两家公司都争着利用资源,致使总资源使用量达 10,超过了 8 的制约,致使优化结果不可行;相反,方案三,把 k 设定为 6,相当是把资源价格定得过高,两家公司都不爱过多地生产,尤其是第二个子公司,在这种资源定价下,甚至造成生产越多越赔钱,因此最终形成不利于生产的局面;只有当 k 值选定的合适,即方案三, $k = 5$ 时,全公司总体收益达到最大值。

注意:上述方法无论是方法一:“对约束条件进分解 / 协调方法”,或是方法二:“对目标函数进行分解 / 协调方法”,上级系统不仅对下级子系统下达 b_1, b_2 (对于方法一) 或是下达 k (对于方法二),都还有一个反馈协调过程,即下级执行 b_i 或 k 之后,子系统的状态要反馈给上级系统,上级对子系统进行协调。

我们所举的例子虽然十分简单,但是体现了大系统分解 / 协调的本质和大系统分解 / 协调的基本方法。

把上述大系统分解 / 协调,不论是方法一中的 b_i 或是方法二中的 k ,都是上级对下级进行协调的参量,统称为协调变量;下级送至上级的状态量或是参量统称为反馈变量;上级系统起着协调作用,有时又可以把它称为协调器。所以大系统分解 / 协调过程一般性示意图,如图 5-5 所示。

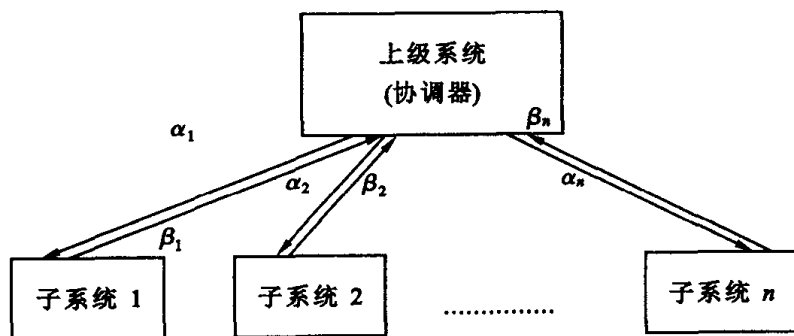


图 5-5 大系统分解 / 协调示意图(α_i 为协调变量; β_i 为反馈变量)

习 题

- 5.1 传统的优化算法和现代优化算法包括哪些内容?两者之间的最大区别是什么?
- 5.2 对感兴趣的组合优化问题,比较遗传算法与其他优化算法的计算结果。

-
- 5.3 遗传算法的主要步骤是什么?在应用遗传算法解决实际优化问题时应该注意的问题有哪些?
- 5.4 用遗传算法求 $\max f(x) = 1 - x^3, x \in [0, 1]$.
- 5.5 系统分解 / 协调的原则是什么?

6 决策分析

本章提要

系统决策分析和系统优化分析的共同点在于都是寻求系统最优解(满意解),但是它们之间又存在着原则性的区别,前者考虑了决策者偏好,包括决策者对多目标问题目标间的权衡和对风险的态度等。

本章在讨论决策的基本要素的基础上,首先介绍了单目标的确定型决策、风险型决策和不确定型决策;继而介绍了更为广泛存在的多目标决策方法;最后简要地介绍了基于计算机的辅助决策工具——决策支持系统。

6.1 决策概述

6.1.1 引言

“决策”是贯穿于人类日常生活、生产、经济、科学实验、政治、军事等一切活动中的思维过程和结果。好的决策就会获得好的结果;反之,则会造成较大的损失。当然,决策者总是希望所作出的决策能花费最小的代价而获得最大的利益。

一般所说的决策多指管理和领导过程中做决定的策略和方法,或已决定好的策略和办法。著名学者西蒙(H. Simon)认为,“管理就是决策”。一般认为,决策就是作决定的过程和作出的决定。狭义地说,决策指的是在几种行为方案中作出选择;广义地讲,决策还包括在作出最后选择前所进行的一切思维活动。

决策都是有目的的,都是对未来而言的,都是建立在对未来预测的基础之上的。而未来都具有不确定性,因而也就有种种可能和选择。对未来进行预测并在此基础上作出选择的过程就是决策过程,选定的策略就是决策结果。决策过程的本质是优化,是一个反复分析、综合并作出抉择的复杂且多次循环的过程。

决策分析包括收集可行性方案、对未来进行预测、建立目标集、优化分析各种可行方案并给出结果等等,每个环节的工作都要依靠决策者和专家的知识、经验和胆识。决策分析的目的不是代替决策者去作决策,而是澄清事物的内在复杂性,协助决策者作出满意的决策。决策的优劣依赖于基本状况和数据的全面性和完整性,也依赖于决策分析人员的技巧,更依赖于决策者高瞻远瞩的能力、眼光与魄力。

6.1.2 发展与趋势

决策及决策分析与人类社会的发展有着同样长的历史,因为人类要正确地行动就必然有决策,没有决策就不能行动,所以,一部人类社会发展史同时也是一部人类重大决策的历史。我国古代的《资治通鉴》、《孙子兵法》等著作都是这方面的重要史料。人类历史上有许

多优秀的决策家,人们常用“运筹于帷幄之中,决胜于千里之外”来赞誉他们。如汉高祖刘邦的谋士张良,汉刘备的丞相诸葛亮等等。一个国家、一个民族如能正确决策,这个国家、这民族就充满希望。

但决策作为科学,作为关于思维、关于行动、关于人们切身利益的科学,仅仅是 20 世纪特别是 20 世纪后半叶的事情。1903 年美国的泰勒(F. W. Taylor)发表了《科学管理原理》,他是现代企业管理科学的创始人。20 世纪中叶,著名的“曼哈顿计划”、“阿波罗飞船计划”等重大系统工程行动极大地促进了管理决策科学的发展。相继形成了运筹学、控制论、信息论、系统工程、系统论等学科。20 世纪后半叶,信息技术迅猛发展,人类社会进入了有史以来最为迅速的发展时期,以信息技术带动的新技术革命已惠及全球,它给人类社会带来的影响是广泛、深入、不可逆转和尚难充分预料的。

在当今高速发展、紧张运转的世界潮流中,人们面临的危机和选择越来越多,需要作出的决策也越来越多,面对大量的、重要的决策问题,决策科学本身也在不断地发展着。当前,可以看到决策科学已基本形成下列几个发展趋势。

(1) 决策的科学化

随着社会进步和决策理论的发展,需要科学化的决策已形成共识。问题是如何进行决策才是科学的,在这个方面有许多不同的看法,我们认为科学化体现在决策科学自身的发展、决策理论及方法的普遍应用和决策的制度化、民主化、国际化和产业化之中。

(2) 决策的民主化

随着人们文化水平和生活水平的提高,越来越多的人将参与到决策特别是重大社会经济、环境问题的决策中来。不同范围的投票选举也是一种群体决策。

(3) 决策的制度化

决策的制度化与决策的民主化之间的关系是,民主化需要制度化的保证,但制度化的决策程序不一定是民主的。我国历史上走过许多弯路,都是决策失误造成的。在进入新世纪的今天,决策的制度化、民主化和科学化已成为潮流,全国人大的立法工作空前繁重。

(4) 决策的国际化

随着人类社会进入信息化时代,地球变得越来越“小”,每个地区和国家的决策都会对其他地区和国家甚至整个世界产生越来越多的程度不同的影响。目前,环境保护、卫生防疫和反恐怖等领域的国际合作干预已得到加强。未来国际化决策将更多地制度化,国际决策的作用将越来越强而有力。

(5) 决策的产业化

常设的专业性决策机构将越来越多,为人们提供科学、就业、投资理财等决策服务,为企业事业单位提供工商管理、税务管理、投资、人力资源管理等决策服务。

(6) 决策的电脑化

随着 PC 机和互联网的普及,越来越多的人可在家里的电脑上咨询升学、就业、投资、购物等问题,越来越多的决策问题可以在网络上进行处理,如电子投票、网上购物等等。

6.2 决策的基本要素及分类

决策的基本要素有:① 决策者;② 决策对象;③ 决策信息;④ 决策目标;⑤ 决策理论和方法;⑥ 决策环境;⑦ 决策工具。其中,决策者、决策对象和决策信息是决策的基本要素。分述如下:

(1) 决策者

通常,决策者就是决策过程的主体,即有理智的人,又称决策人。但对于较高层次的许多决策问题,真正的决策人没有精力也不太可能熟悉决策科学的具体技术方法,而需要委托专家进行研究,提出方案。这时,专家也参与了决策过程,但只是作为决策研究人员,不是真正的决策人。决策人是采用和实施决策方案之主体的利益代表者。特别需要注意的是,决策人的偏好会对决策产生较大的影响。

按决策者分类,决策可分为:

个人决策:可能是决定其个人事务,也可能是决定其任领导职务的部门的事务;

群体决策:多个决策人共同决策,这里是指为了某个决策对象的共同目标。如果是为不同的决策对象,不同的目标,则不属于这种类型。

(2) 决策对象

决策对象是决策问题所研究的对象,是采用或实施最终决策方案的系统。通常,决策对象的负责人或管理层就是决策者。按决策对象分类,可分为宏观规模对象与微观规模对象两大类决策。

微观对象主要是企业及企业以下规模系统的决策;

宏观对象有:

经济决策:决定一个行业、一个地区甚至一个国家的经济发展计划、发展目标、金融政策等等;

政治决策:制定各种法律、法规、外交政策、竞选公职的策略等;

还有文化决策、军事决策、教育决策、环境保护决策等等。

(3) 决策信息

决策信息是决策者对决策对象的认识,广义地讲还应包括决策者对决策环境等诸多方面情况的了解。按决策信息分类,有:

任意决策:有时可用的信息量极少,有时多种方案难分伯仲,没有更多的时间或没有必要做出抉择。这时可采用任意决策,如抽签、摇奖等;

确定型决策:所需信息比较充分,结局常常是唯一的,如最短邮路等;

非确定型决策:信息不完全、不准确、但基本信息已知时所做的决策,结局常常是非唯一的。这是比较普遍的情况,任意决策和确定型决策是其特例。

有时对某些决策问题的处理需要运用模糊数学和灰色系统理论等,因而也就有模糊决策和灰色系统决策等。

(4) 决策目标

决策目标是决策中非常重要的内容,决策目标应尽量具体明确,但不是越多越好,有时目标太多会造成系统僵化。按决策目标分类,有:

单目标决策、多目标决策、局部目标(战术)决策、全局目标(战略)决策。

(5) 决策理论和方法

随着决策理论与方法的不断发展、丰富和完善,相应的决策类型也很多,如按决策的复杂性来分类,有:

简单决策:不需要多少决策理论和方法即可做出决定的决策;

常规决策:应用一般的决策理论和方法进行的决策;

创造性决策:需要采用一些特殊和复杂的理论和方法进行的决策。

(6) 决策环境

决策对象总是处于一定的环境之中的,即它又是一个更大系统的子系统,决策者也是在一定环境条件下进行决策的。因此,决策环境对决策有着重大的影响。图 6-1 表示了决策对象、决策者和决策环境这三者所构成的决策系统。

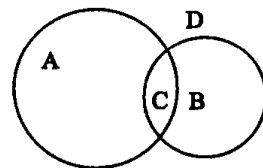


图 6-1 决策系统

图中 A 表示决策对象,B 表示决策者,C 是 A 与 B 的交集,D 是决策环境。A 与 B 相交表明,决策者通常至少部分是决策对象的一部分。如 B 属于 A,则说明是为本系统自身做决策。(B-C)部分是 A 的环境的一部分,它与 A 的关系通常比较密切。决策者 B 中的一部分是真正的决策人,另一部分是决策研究人员。

根据不同的决策环境分类,有:竞争环境决策、敌对环境决策、独立决策、从属决策、紧迫环境决策、从容环境决策等等。

(7) 决策工具

决策工具的发展也极大地影响着决策科学的发展,如古时的重大发明、逻辑关系、数制、数学的发展、现代的重大发明、电脑与网络等都促进了决策科学的发展。按决策工具分类,有:

直接决策:不借助于大脑之外的工具进行决策;

简单工具支持决策:仅借助于计算器和计算机进行演算的决策;

智能网络支持决策:使用计算机网络或决策支持系统帮助进行的决策。

6.3 决策模式与决策过程

决策模式就是决策的方式,决策的程序或决策的一般步骤,它反映了众多决策过程中的一般规律,是各类决策过程的理论概括和抽象,是一般的决策过程。图 6-2 是决策过程或决策模式的流程图。下面对各个部分做一个简介。

6.3.1 问题分析

对于决策问题,首先是要进行问题分析。爱因斯坦认为:“提出一个问题往往比解决一个问题更重要。因为解决一个问题也许仅仅是一个数学上或实验上的技能而已,而提出新的问题、新的可能性,从新的角度看旧的问题,却需要有创造性的想像力,而且标志着科学的真正进步。”决策中所说的问题,是指人们在活动中所面临的困难或困境,是为了某种目的而提出的任务或课题。因此,要恰当地分析问题和表达问题。

6.3.2 目标确定

系统目标的确定是十分重要的。其本质就是确定衡量决策方案优劣的尺度。目标又都是受一定的社会价值体系约束的,价值指标体系应有明确的、规范的和科学的描述。要做出这样的描述,其前提是对系统的未来要有基本正确的预测。“凡事预则立,不预则废”。因此,预测也是做好决策的前提。社会经济领域的许多问题是难于定量预测的,但随着信息技术的发展,定量方法越来越多,预测也越来越准确。

6.3.3 制订方案

根据上面提出的系统目标体系,设计可能的方案。这个环节包括了目标分解、初步论证、模型化设计等内容,并且不是一次就能完成,可能有多次反复的设计。而且,应该保证提出多个不同的方案供下一步的比较排序。

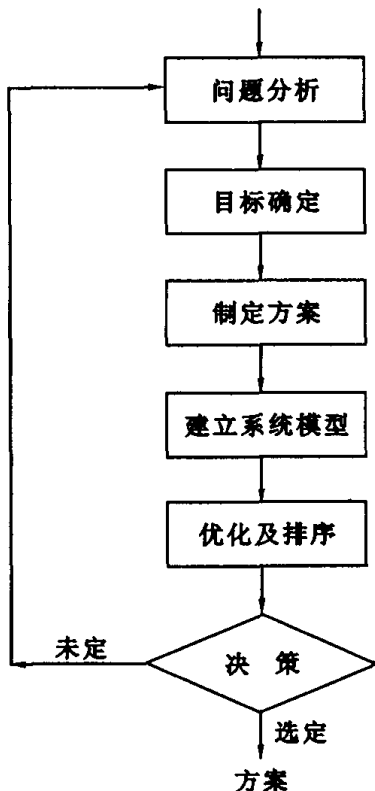


图 6-2 决策流程图

6.3.4 建立系统模型

为了更准确地分析和把握系统的特性,需要建立系统模型。实际上,在目标确定等阶段也需要根据情况建立相应的模型。

6.3.5 优化、排序

优化排序有多种方法,根据不同的决策问题选择适用的方法,进行优化分析、排序及优化方案的鲁棒性分析。如条件许可,应采用多种方法进行优化的排序。

何为鲁棒性分析,其实质就是灵敏度分析。如果排序结果的鲁棒性很差,还不能算是最优方案。

6.3.6 决策

这是决策的最后关头,一旦“决策”形成,就进入方案实施过程,就不可能全部返回。在重大问题的决策中,由于研究人员及专家提供的往往是多个可供选择的方案,排序也可能随着其他因素的变化而变化,最终要由决策人定夺。因此决策人都有“决策”前的焦虑。它是对决策人思维、意志和魄力的一种考验。

总之,决策模式与决策过程是所有决策都应该遵循的一般流程。遵循这个模式,就为决策的科学化和正确性提供了有力的方法保证。在实际运用中,每一个阶段出现问题,都可能要返回到前面的某个环节进行修正。另外,由于决策环境的不完全确定和决策信息的不完全确知,在决策中特别是重大问题的决策中,应尽量做好下列三个方面的工作:

(1) 决策者非唯一

决策问题可委托给多家研究机构研究,这样可得到多种不同的方案和结论;真正的决策者也最好是多人集体决策,可在一定程度上避免个人知识水平、品质修养甚至利益关系等方

面因素带来的局限性。重大问题的决策应大力推行科学化的民主程序决策,以避免重大失误。

(2) 决策方法非唯一

由于决策信息不完全,决策环境不确定,而每一种决策方法都不是万能的。因此,有必要在决策中采用多种理论和方法。

(3) 决策结局非唯一

对于决策信息不完全程度较高、决策环境不确定性较大、方案实施时间很长的情况,最好引入竞争机制,多种决策方案同时在不同的范围供实施,在实施的实践中,进一步检验各种实施方案。“一刀切”的做法,可能把一些有生命力的和有前途的事物也切掉了。而且,既已被“切掉”,就难以证明被切的事物也是有前途和生命力的,也难以判断“切”的正确还是错误。

6.4 确定型决策

在确定型决策中,决策信息是确知的,决策方案引起的结局是确定的。实际上,绝大多数决策问题都因为决策环境和信息等的不确定而不可能是确定型的决策。但许多决策问题经过简化处理,可作为确定型决策问题来解决。

解决确定型决策问题的方法很多,最终都是比较各种方案的价值函数值,描述如下:

$$A_i^* = \max_{A_i \in A} V(A_i) \quad (6.1)$$

式中 A ——各种方案的集合, A_i 为第 i 个方案;

$V(A_i)$ ——方案 A_i 的价值函数值;

A_i^* ——最佳方案。

6.5 风险型决策

在风险型决策问题中,决策环境是不完全确定的,决策信息是不完全确知的,每一种方案对应的结局是非唯一的,但可以知道各种结局的发生概率。风险型决策也称为随机型决策。

风险型决策的基本方法是通过计算比较各个方案的期望效用值,得出最佳方案。设 A 为所有方案的集合, A_i 为第 i 个方案,则第 A_i 个方案的期望效用值为

$$U(A_i) = \sum_{j=1}^m P(C_j) U(A_i, C_j)$$

式中 C_j ——第 j 种结局, $j = 1, 2, \dots, m$;

$P(C_j)$ ——结局 C_j 出现的概率, $\sum_{j=1}^m P(C_j) = 1$;

$U(A_i, C_j)$ ——第 i 种方案在 C_j 结局下的效用值;

则最佳方案为

$$A_i^* = \max_{A_i \in A} U(A_i) \quad (6.2)$$

式中 $U(A_i)$ ——方案 A_i 的期望效用值。

下面在介绍具体的风险型决策方法前,先介绍一下效用理论。

6.5.1 效用理论

“效用”是决策者对在有风险情况下价值的衡量。

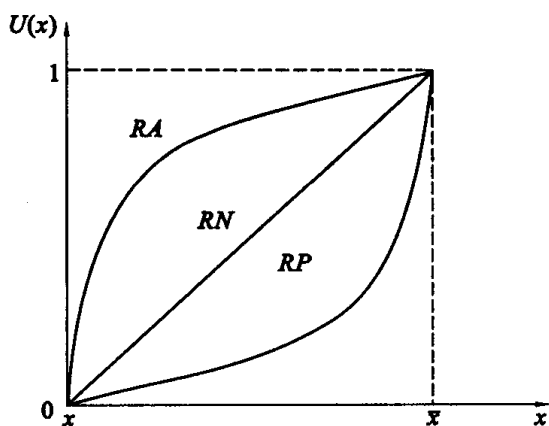


图 6-3 效用函数的类型

对于决策者认为最理想(最好)的事件,记为 $\bar{x}$,规定其效用值 $U(\bar{x})$ 最大,取为 1;最不希望(最坏)的事件,记为 x ,规定其效用值最小,取为 0;中间状态的事件,记为 x ,在 $(0,1)$ 中取效用值 $U(x)$ 。用横坐标表示事件,纵坐标表示效用值,就得“效用曲线”或“效用函数(Utility function)”。有些人怕冒风险,属于避免风险型或保守型,记为 RA(risk aversion);有些人愿意冒风险,属倾向风险型或冒险型,记为 RP(risk proneness);大多数人介于这两者之间,属于风险中性型,记为 RN(risk neutral)。这三种类型的效用函数可表示为图 6-3。

效用函数可以通过与决策者对话获得,对话方式有三种:一是在其他参数确定的情况下,要求决策者回答“确定性等价量” x 的值;二是在其他参数确定的情况下,要求决策者回答事件发生的概率;三是在其他参数确定的情况下,要求决策者回答“最好情况” $\bar{x}$ 的值。

如决策者面临两种可选方案 A_1 和 A_2 。选方案 A_1 ,可能以概率 P 获得收益 x_1 ,也可能以概率 $(1-P)$ 获得收益 x_2 ;选方案 A_2 ,可无风险地稳获收益 x_3 ,且 $x_1 > x_3 > x_2$ 。调整 x_3 的大小至决策者认为 A_1 、 A_2 两个方案相当,即两个方案的效用值相等,其表示式为

$$PU(x_1) + (1-P)U(x_2) = U(x_3) \quad (6.3)$$

【例 6.1】 投资生产某产品,投放市场后的销路可能很好、较好、一般甚至差。设最好收益为 80,效用值 $U(80) = 1$;最差收益为 -20, $U(-20) = 0$ 。

假定一, A_1 方案为:可能以 0.5 的概率取得收益 80,也可能以 $(1-0.5)$ 的概率亏损 20; A_2 方案为:肯定获得收益 x 。求 x 取何值时,决策者认为 A_1 与 A_2 相当。由式(6.3)有

$$\begin{aligned} PU(x_1) + (1-P)U(x_2) &= 0.5 \times U(80) + 0.5 \times U(-20) \\ &= 0.5 \times 1 + 0.5 \times 0 = 0.5 \end{aligned}$$

如决策者回答 x 为 10,则在效用坐标上确定一个点 $(10, 0.5)$,即 $U(10) = 0.5$ 。

假定二, A_1 方案为:可能以 0.5 的概率取得收益 80,也可能以 0.5 的概率获得收益 10; A_2 方案为:肯定获得收益 x 。问 x 取何值时,决策者认为 A_1 与 A_2 相当。由式(6.3)有

$$\begin{aligned} PU(x_1) + (1-P)U(x_2) &= 0.5 \times U(80) + 0.5 \times U(10) = 0.5 \times 1 + 0.5 \times 0.5 = 0.75 \end{aligned}$$

如决策者回答 x 为 30,在效用坐标上又得到一点 $(30, 0.75)$,即 $U(30) = 0.75$ 。

假定三, A_1 方案为:可能以 0.5 的概率取得收益 10,也可能以 $(1-0.5)$ 的概率亏损 20; A_2 方案为:肯定获得收益 x 。求 x 取何值时,决策者认为 A_1 与 A_2 相当。由式(6.3)有

$$\begin{aligned} PU(x_1) + (1-P)U(x_2) &= 0.5 \times U(10) + 0.5 \times U(-20) \\ &= 0.5 \times 0.5 + 0.5 \times 0 = 0.25 \end{aligned}$$

如决策者回答 x 为 -10 , 则在效用坐标上确定一个点 $(-10, 0.25)$, 即 $U(-10) = 0.25$ 。

将这三点标于效用坐标上, 可得到一个粗略的效用函数。(见图 6-4) 该效用函数表明, 该决策者是偏于保守型的。

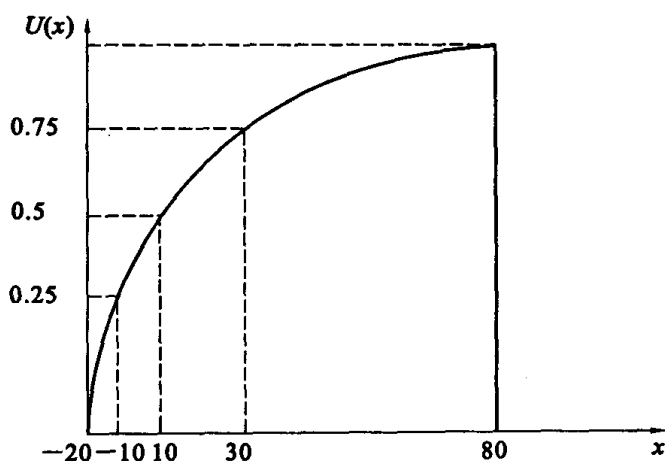


图 6-4 求取效用函数

6.5.2 表格法

表格法也称为矩阵法, 下面结合例题介绍。

【例 6.2】 投资生产某新产品, 有三种方案; A_1 , 建大型装置; A_2 , 建中型装置; A_3 , 建小型装置。根据调查分析, 该产品上市后, 畅销(C_1) 的可能性是 $P(C_1) = 0.3$, 销售一般(C_2) 的可能性是 $P(C_2) = 0.4$, 滞销(C_3) 的可能性是 $P(C_3) = 0.3$ 。设各种方案对应于各种可能性结局的收益值 $U(A_i, C_j)$ 如表 6-1 所示。

表 6-1 决策矩阵收益值表

单位: 万元

方 案 \ 结 局	C_1 $P(C_1) = 0.3$	C_2 $P(C_2) = 0.4$	C_3 $P(C_3) = 0.3$	效用 $U(A_i)$
A_1	2500	800	-200	1010
A_2	1700	600	200	810
A_3	800	400	200	460

$$\text{由 } U(A_i) = \sum_{j=1}^3 P(C_j)U(A_i, C_j)$$

计算得 $U(A_1) = 1010$, $U(A_2) = 810$, $U(A_3) = 460$ 。显然, A_1 为最佳方案。

6.5.3 决策树法

这种方法就是把决策问题用树状图形表示出来。仍以前面的例子来说明决策树法, 得到决策树如图 6-5 所示。方块表示决策节点, 也称为决策点。在决策节点, 由决策者主观作出抉择; 由方块引出的射线代表方案; 射线的终点是圆圈, 称为机会节点或方案节点, 圆圈内标注的是该方案的期望效用值。在机会节点, 决策者的主观行为无能力作出抉择, 而完全依赖于客观规律, 如自然规律或经济规律等作出抉择; 由圆圈引出的射线代表不同的结局, 线上注出不同结局出现的概率, 线的右端标注每种结局的效用值 $U(A_i, C_j)$ 。决策树法仍然用的是

和表格法一样的计算效用方法。但决策树有时更形象些。

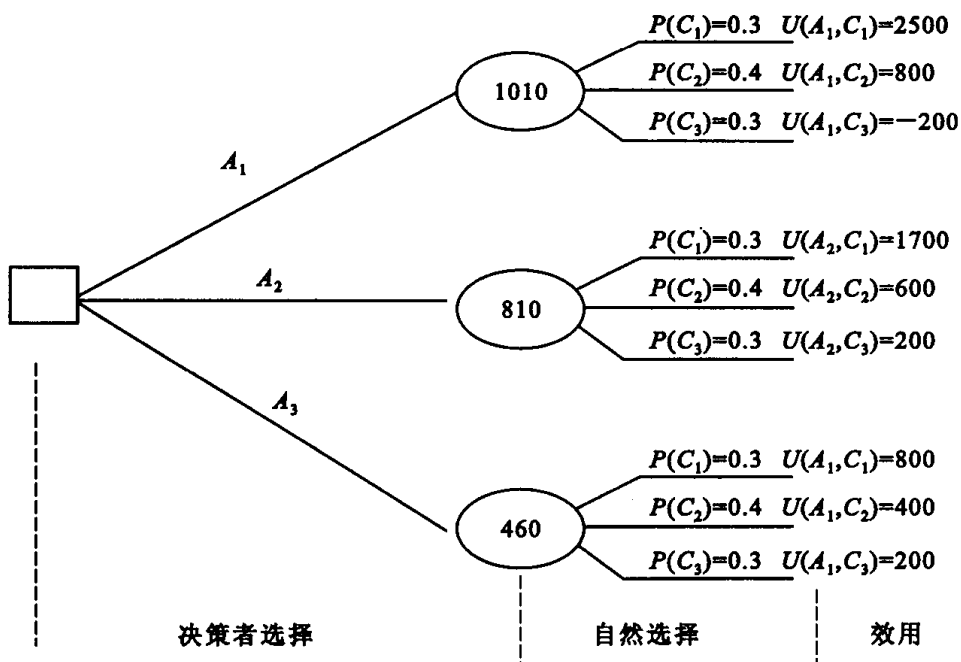


图 6-5 投资生产新产品的决策树

现实的决策问题往往不会像例 6.1 那么简单,有时是连续的多级决策,这时可用多级决策树描述,这样的决策也称为序列决策。如图 6-6 是一个二级决策树。在这里,首先要决定是否研制某种新产品;如研制,研制的实际发生成本可能有三种自然选择状态;在不同的开发成本下,又要选择何种规模投产;投产后,实际的销售又有三种自然选择状态。

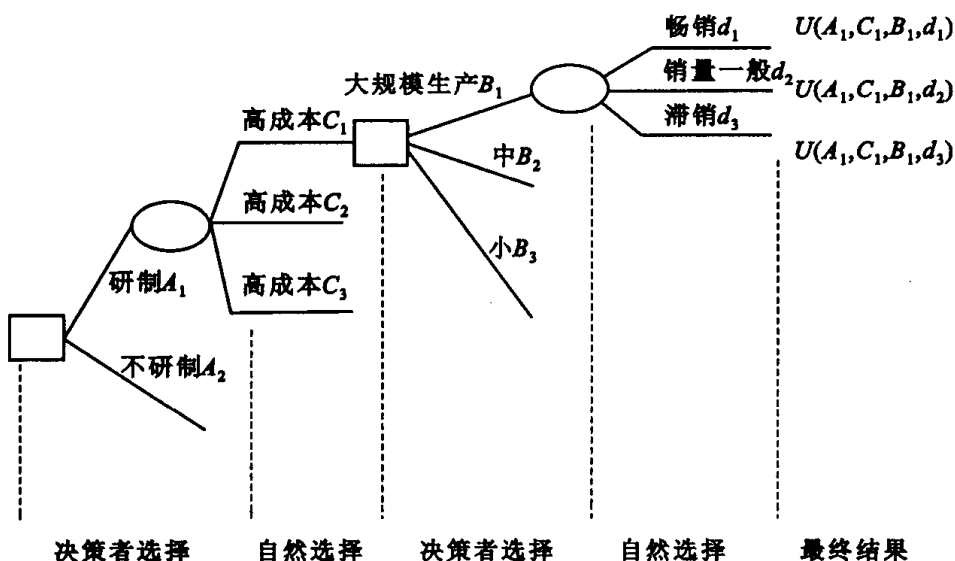


图 6-6 多级决策树

在前面的风险型决策中,都假设事件或结局出现的概率已知,通常需要通过随机实验来确定这些概率。但许多决策问题无法进行随机实验,例如,新产品的研制,投放市场后的销售状况就无法实验,大型水利工程等也无法获取结局的概率。即使有些项目可以进行实验,但成本太高,难以实验或实验次数也不可能多。因此,风险型决策中的结局概率,在无法通过随机实验获取的情况下,就只有靠决策者根据其对结局事件的理解,设定概率。这样取得的概

率称为主观概率,而通过随机实验取得的概率称为客观概率。

6.6 不确定型决策

不确定型决策与风险型决策不同的是,不确定型决策不知道各种结局发生的概率。于是,依据不同的处理问题的倾向,决策者可采用下列不同的决策准则进行决策。

6.6.1 等可能准则(Laplace 准则)

因为不知道每种结局出现的概率,就认为它们出现的可能性相等。

如有 n 个结局状态 $C_i, i = 1, 2, \dots, n$, 则

$$P(C_1) = P(C_2) = \dots = P(C_n) = \frac{1}{n} \quad (6.4)$$

方案 A_i 的期望效用值为

$$U(A_i) = \frac{1}{n} \sum_{j=1}^n U(A_i, C_j) \quad (6.5)$$

最优方案是 A_i^*

$$A_i^* = \max_{A_i \in A} U(A_i) \quad (6.6)$$

【例 6.3】 设某新产品的生产方案有 4 种,分别为 A_1 、 A_2 、 A_3 和 A_4 ,投产后可能的销售状况为好、一般、差 3 种。这 3 种状况出现的概率很难估计,12 种结局的收益值如表 6-2 所示,选择最好方案。

表 6-2 例 4 的收益表

单位:万元

销 售 方 案	C_1 好	C_2 一般	C_3 差	效用 $U(A_i)$
A_1	400	700	500	533
A_2	800	500	400	567
A_3	500	600	400	500
A_4	500	600	600	567

由等可能性准则得

$$P(C_i) = \frac{1}{3}$$

$$U(A_1) = \frac{1}{3}(400 + 700 + 500) = 533$$

$$U(A_2) = 567$$

$$U(A_3) = 500$$

$$U(A_4) = 567$$

A_2 与 A_4 均为最佳方案。

6.6.2 后悔值准则(Savage 准则)

这种方法的目标是把后悔值降为最低。定义每种状态下各方案结局中最好效益值与该

状态下某方案结局的效益值之差为该结局的后悔值。如例 6.3 中状态 C_1 下四种结局中最好效益值为 800, 最差是 400, 中间是 500。则得各结局的后悔值分别为 $R(C_1, A_1) = 400, R(C_1, A_2) = 0, R(C_1, A_3) = R(C_1, A_4) = 300$ 。依此可得后悔值矩阵表 6-3。各方案中最大后悔值为 $R(A_i)$, 可见最小的 $R(A_i)$ 为 $R(A_2) = 200$, 即认为方案二是依该准则得到的最佳方案。

表 6-3 后悔值表

方 案 \ 状 态	C_1 好	C_2 一般	C_3 差	$R(A_i)$
A_1	400	0	100	400
A_2	0	200	200	200
A_3	300	100	200	300
A_4	300	100	0	300

6.6.3 悲观准则(max-min 准则)

也称为保守准则或 Wald 准则。其思路是从最坏处着想, 在每种方案中选择效益值最小的作为最坏的结局, 即对问题持最悲观的态度, 然后再从已选的结局中选最好的作为最佳方案。其表达式为

$$U(A_i^*) = \max_{A_i \in A} \{ \min_{1 \leq j \leq n} U(A_i, C_j) \} \quad (6.7)$$

对于例 6.3 中的情况,

$$U(A_1) = U(A_1, C_1) = 400$$

$$U(A_2) = U(A_2, C_3) = 400$$

$$U(A_3) = U(A_3, C_3) = 400$$

$$U(A_4) = U(A_4, C_1) = 500$$

于是最好方案是 A_4 。

6.6.4 乐观准则(max-max 准则)

这种对不确定型决策问题的处理思想和悲观准则相反, 认为每种方案中选择的结局总是最好的, 即对问题总是持乐观的态度, 然后再从诸结局中选最大值, 其对应的方案是最佳方案。其表达式为

$$U(A_i^*) = \max_{A_i \in A} \{ \max U(A_i, C_j) \} \quad (6.8)$$

仍看例 6.3, 对应的四个效益值为: 700, 800, 600, 600, 其中, 800 最大, 对应的最佳方案为 A_2 。

6.6.5 折衷准则(Hurwicz 准则)

它认为过于乐观, 容易冒进, 会发生大的偏差; 过于悲观, 则太保守, 易失去大好时机, 同样偏差也大。它主张采用折衷的办法, 在乐观与悲观之间取一个适当的折衷点 α , α 即为乐观系数或折衷系数, $0 \leq \alpha \leq 1$ 。 $\alpha = 1$, 表示乐观; $\alpha = 0$, 表示悲观。其表达式为

$$U(A_i) = \alpha \max_{1 \leq j \leq n} U(A_i, C_j) + (1 - \alpha) \min_{1 \leq j \leq n} U(A_i, C_j) \quad (6.9)$$

再从中选最大的

$$U(A_i^*) = \max_{A_i \in A} \{U(A_i)\} \quad (6.10)$$

仍看例 6.3, 设 $\alpha = 0.4$, 得

$$U(A_1) = 0.4 \times 700 + 0.6 \times 400 = 520$$

$$U(A_2) = 560$$

$$U(A_3) = 480$$

$$U(A_4) = 540$$

可见, 最佳方案为对应于 $U(A_2) = 560$ 的方案。

如设 $\alpha = 0.2$, 则

$$U(A_1) = 460 \quad U(A_2) = 480$$

$$U(A_3) = 440 \quad U(A_4) = 520$$

可见, 此时最佳方案为对应于 $U(A_4) = 520$ 的方案 A_4 。

通过对同一决策问题, 按不同的决策准则所做的决策看, 最佳方案有时是不一致的。应该说, 这是由问题的不确定性所带来的结果。在运用中究竟应该采用哪种准则, 没有统一的规律可循, 受决策者意志的影响很大。因此, 在决策过程中, 应根据具体问题选取合适的决策方法并与决策者共同协商做好决策分析。

6.7 多目标决策

对于实际的决策问题, 系统目标往往不止一个, 多个目标反映了系统在不同方面的追求, 有时不同的目标之间又是相互矛盾的。例如, 新建一座工厂要考虑投资、人力、回收期、市场、交通、环境污染等一系列因素。即使是比较简单的决策问题, 目标也可能不止一个。因此单目标决策仅仅是多目标决策的简化和特例。

6.7.1 多目标决策问题的数学描述

多目标决策实际上就是求解多目标最优化问题。可描述如下:

当目标函数向量愈大愈优时, 有

$$\begin{cases} V - \max F(x) \\ x \in R \\ g_i(x) \leq 0, i = 1, 2, \dots, m \end{cases} \quad (6.11)$$

当目标函数向量愈小愈优时, 有

$$\begin{cases} V - \min F(x) \\ x \in R \\ g_i(x) \leq 0, i = 1, 2, \dots, m \end{cases} \quad (6.12)$$

式中 $V - \max$ ——求向量最大值;

$V - \min$ ——求向量最小值;

x ——决策向量, $x = (x_1, x_2, \dots, x_n)^T$;

$F(x)$ ——目标向量, $F(x) = [f_1(x), f_2(x), \dots, f_p(x)]^T$;

$g_i(x)$ ——约束条件函数。

如系统共有 P 个子目标,则能使 P 个目标同时达最优的解是系统的最优解;对于至少有一个目标为优的称为非劣解或有效解;其他的解称为劣解。在实际决策问题中,最优解是很难求取的,很可能是不存在的,而非劣解更有实际意义。

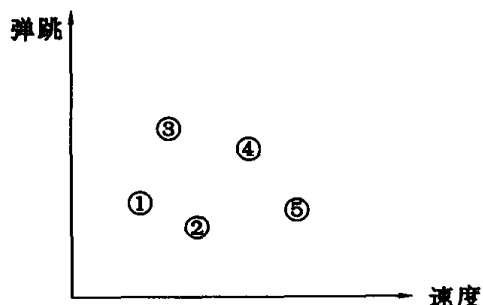


图 6-7 运动数据

例如,选拔一名优秀运动员,身高、体重、年龄、速度、弹跳等需要考虑的因素很多,即以两个因素速度和弹跳而论,情况也是复杂的。如有 5 名候选人的资料如图 6-7 所示,显然①、②运动员速度与弹跳都不太好,属于劣解,可以淘汰;③、④、⑤运动员显然是优于①、②运动员的,属于非劣解。在③、④、⑤运动员之间则较难选择。当考虑 2 个以上因素时,情况更加复杂。而且,除了这些可以具体量化的指标外,还有许多难于量化的因素,如运动员的心理素质、与其他运

动员的关系、决策者的偏好等等。因此,要追求系统的最优解是不现实的。著名经济学家西蒙(H. Simon)就认为,由于决策者认识上的局限性,不可能知道他们所做决策产生的全部结果;又由于受时间、资金和信息等因素的制约,也不可能对所有方案进行很详尽的比较分析。因此,西蒙否定了“经济人”概念和“最大化”行为准则,提出了“管理人”概念和“令人满意的”行为准则。即,决策所追求的不是绝对意义的最优解,而是相对意义的满意解。

6.7.2 多目标决策方法概述

多目标决策方法主要有以下几种:

(1) 化多为少法 由于同时处理多个决策目标很困难,于是将多个决策目标简化为一个或两个决策目标来处理。这种方法实际上是对主要决策目标尽可能优化,其余决策目标只要达到一般要求即可。

(2) 理想点法 找出每个决策目标的最优值,称为“理想点”。“理想点”通常是无法实现的。但知道了理想点的位置,就可以比较容易确定距离“理想点”较近的点作为满意解。

(3) 效用函数法 这是自上而下的方法,首先充分考虑决策者对风险和价值的偏好,然后对多属性效用函数进行综合,最终给出定量的方案排序。这是一种常用的方法。

(4) 目的规划法 对于一组预定的目标,在约束条件下,这组目标有可能实现也可能实现不了。目的规划法就是合理地配置有限的资源,使决策结果尽可能地接近预定目标。它是理想点法的一个特例。

(5) 分层序列法 对于系统目标较多又较复杂的情况,可以对目标按重要程度进行分层分类,首先尽量满足最重要的目标,再处理次重要目标,最后处理最不重要目标。如此分层处理,有利于最终找到较满意决策方案。

(6) 层次分析法 该方法将一个复杂的决策问题分解为不同层次的若干元素,然后依靠决策者对元素进行主观比较评判和赋值,得到判断矩阵。然后,对判断矩阵进行分层处理及综合,得到多种决策方案的排序结果。

下面重点介绍效用函数法和层次分析法这两种常用的方法。

6.7.3 效用函数法

在风险型决策中,已经简要地介绍了效用理论及效用函数。但那里讨论的是单属性效用函数;对于多属性效用函数 $u(x_1, x_2, \dots, x_n)$, 当多目标的属性相互独立时,多目标效用函数可用线性叠加组成,即

$$u(x_1, x_2, \dots, x_n) = \sum_i w_i u_i(x_i) \quad (6.13)$$

式中 x_i ——评价属性, $i = 1, 2, \dots, n$;

$u_i(x_i)$ —— x_i 的单属性效用函数;

w_i —— x_i 属性的权重系数。

如属性间相互不独立,效用函数不能用上式计算,需要确定多属性效用函数。这时,对于方案 A_i ,其效用值可计算如下:

$$U(A_i) = \sum_{j=1}^m p_j u_{ij}(X) \quad (6.14)$$

式中 X ——属性向量, $X = (x_1, x_2, \dots, x_n)^T$;

$u_{ij}(X)$ —— A_i 方案 j 结局的多属性效用函数。

最佳方案 A^* 依据各方案的期望效用值排序获得,即

$$U(A^*) = \max_{A_i \in A} \{U(A_i)\} \quad (6.15)$$

因此,从本质上讲,效用函数法处理多目标决策问题,相当于把多目标寻优问题简化为单一目标求最大值问题,以使用各种单目标数学规划方法寻求最优解,从而得到多目标问题的满意解。但求取效用函数往往是比较困难的。

6.7.4 层次分析法

层次分析法 (Analytical Hierarchy Process) 简称为 AHP 方法,是美国运筹学家 T. L. Saaty 教授于 70 年代提出的。在 AHP 法提出之前,许多人片面地认为,决策就是依靠数学模型解决问题。例如投入大量的人力、物力和时间建立社会经济系统的复杂的数学模型,再用大型计算机求解。结果是,投入越来越多,模型越来越复杂。能够理解和求解模型的人越来越少,实际的效果却不明显。但实际上,问题的关键是,人们无法回避决策过程中决策者的选择和判断所起的决定性作用。而且,许多因素是难于用传统的方法进行量化的。

Saaty 教授通过认真研究决策思维的规律,提出用 AHP 将人的主观判断用数量形式表达出来及进行相应的处理。虽然深入系统地研究 AHP 需要相当的数学基础,但它从本质上讲是一种思维方式。其关键在于将复杂的决策问题表达为不同层次的若干因素的集合,层次之间的因素具有支配关系。通过两两比较确定各因素的重要性程度。最后进行演算,得到各个不同方案的排序。整个过程都体现了人类决策思维的基本特征,即分解、比较、判断与综合。通常,运用 AHP 进行决策分析时,有如下四个步骤:

- A 建立系统的递阶层次结构;
- B 构造比较判断矩阵;
- C 计算权重向量;
- D 计算合成权重,求出总排序。

下面举例进行说明。

(1) 建立递阶层次结构

建立合理实用的 AHP 递阶层次结构的基础是对决策问题的透彻了解。例如,某决策问题有五种可能的决策方案(x_1, x_2, x_3, x_4, x_5),五个方案就构成了 AHP 的最基础的层次。如何评价这些方案呢?如果有 8 项指标($c_1, c_2, \dots, c_8$)可用来评价各个方案,但可能有的方案不一定需要运用全部 8 个指标来加以评价。如果这 8 个指标可以分类为三个大类,如社会效益 B_1 ,经济效益 B_2 ,环境效益 B_3 。这两个层次是 AHP 递阶层次结构的中间层次。有时,可能不止两个层次。最高层次是一个采用不同方案决策的总目标,如这里可用综合效益描述。如此,得到一个 AHP 递阶层次结构如图 6-8 所示。有些复杂的问题,可能要采用更加复杂的循环层次结构或反馈层次结构等来描述。

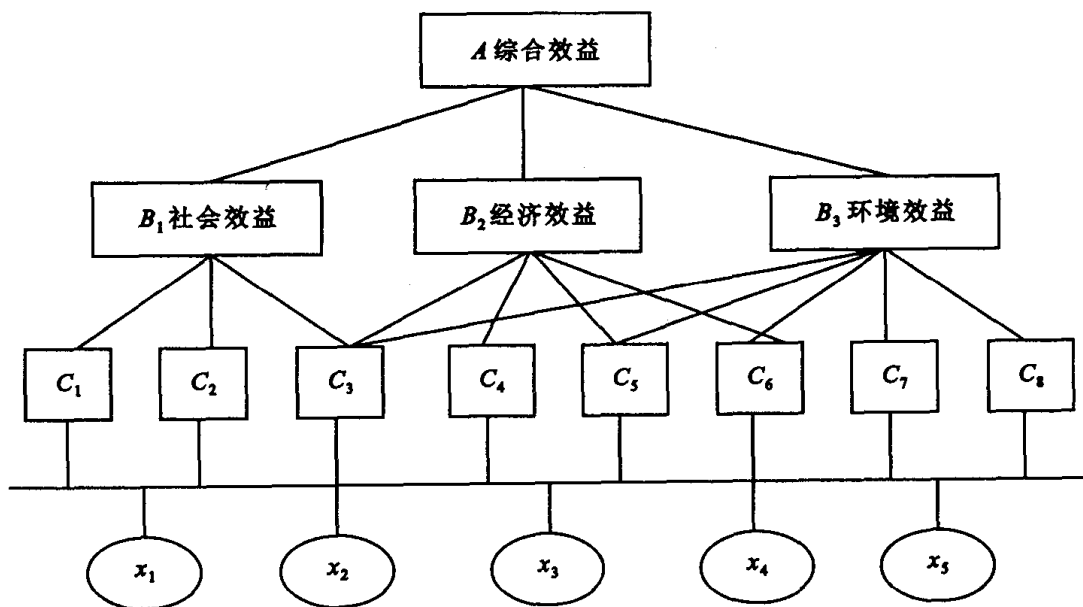


图 6-8 某案例 AHP 递阶层次结构

(2) 构造判断矩阵

决策问题的递阶层次结构确定之后,层次之间的支配关系就确定了,如社会效益 B_1 支配 C_1, C_2 和 C_3 这三个因素,实际上是 B_1 取决于 C_1, C_2 和 C_3 这 3 个因素; B_2 支配 C_3, C_4, C_5 和 C_6 这 4 个因素; B_3 支配 C_3, C_5, C_6, C_7 和 C_8 这 5 个因素。综合效益 A 支配 B_1, B_2 和 B_3 这 3 个因素。每个支配量所支配的各个因素,其重要性如何?权重多少?这些需要采用各因素间的两两比较来确定。进行两两比较,就需要一个比较标准,一个定量的方法,称为比较标度。

如在 A 所支配的三个效益之间,如认为 B_1 比 B_2 重要,可记为 2, B_2 不如 B_1 重要,则可记为 $1/2$; B_1 比 B_3 重要得多,记为 5, B_3 远不如 B_1 重要,可记为 $1/5$; 每个因素与自身相比,是同等重要,记为 1; 这时显然, B_2 比 B_3 重要,可记为 2, B_3 不如 B_2 重要,记为 $1/2$ 。如此,可得到一个比较判断矩阵如下:

$$A = (a_{ij})_{n \times n} = \begin{pmatrix} 1 & 2 & 5 \\ 1/2 & 1 & 2 \\ 1/5 & 1/2 & 1 \end{pmatrix} \quad (6.16)$$

上式矩阵就是对于综合效益 A 的三个效益之间的两两比较判断矩阵, a_{ij} 是比例标度。显然, a_{ij} 有如下性质:

$$a_{ij} = \frac{1}{a_{ji}}, a_{ii} = 1, a_{ij} > 0 \quad (6.17)$$

符合上述条件的矩阵称为正互反矩阵。由于是正互反矩阵,实际工作中,只要获得两两比较判断矩阵的上三角矩阵的各项即可。

如果矩阵 A 的元素有传递性,即

$$a_{lm} \cdot a_{mn} = a_{ln} \quad (6.18)$$

则称判断矩阵 A 为一致性矩阵。式(6.16)中的矩阵不是一致性矩阵。下式则是一个一致性矩阵。

$$(a_{ij}) = \begin{pmatrix} 1 & 2 & 8 \\ 1/2 & 1 & 4 \\ 1/8 & 1/4 & 1 \end{pmatrix} \quad (6.19)$$

用什么样的量来作为比较判断的标度呢?目前已提出的有几十种标度。Saaty 教授特别推崇 $\frac{1}{9}, \frac{1}{8}, \dots, \frac{1}{2}, 1, 2, \dots, 9$ 标度。其含义为: B_1 与 B_2 同等重要,取 1; B_1 比 B_2 重要,取 3; B_1 比 B_2 明显重要,取 5; B_1 比 B_2 重要得多,取 7; B_1 比 B_2 极端重要,取 9。反之,分别取 $1/3, 1/5, 1/7, 1/9$ 。介于这些状态之间的可分别取 2、4、6、8 和 $1/2, 1/4, 1/6, 1/8$ 。这种标度可简称为 1—9 标度,是目前应用最为广泛的一种 AHP 标度。为了适应不同的应用领域,人们已经提出数十种 AHP 标度。表 6-4 列出了其中的 7 种标度。

表 6-4 几种 AHP 标度

标度范围	标度值 / 符号								
$\frac{1}{3} - 1 - 3$	1	1	2	2	2	3	3	3	3
$\frac{1}{5} - 1 - 5$	1	2	2	3	3	4	4	5	5
$\frac{1}{7} - 1 - 7$	1	2	2	3	4	5	6	6	7
$\frac{1}{9} - 1 - 9$	1	2	3	4	5	6	7	8	9
$\frac{1}{17} - 1 - 17$	1	20	30	40	50	60	70	80	90
$\frac{1}{90} - 1 - 90$	1	20	30	40	50	60	70	80	90
$<<- \approx ->>$	$\approx$	$\approx$	$>$	$>$	$>$	$>>$	$>>$	$>>$	$>>$

表中最后一行是一种间接符号标度,它便于设计判断表,供领导层或高层专家使用,经转换后再进行计算和排序。

(3) 权重计算方法

计算权重的方法有和法、根法、特征根法(EM 法)、最小二乘法等,最常用的是特征根法(EM 法),下面简介和法、特征根法(EM 法)及判断矩阵一致性检验方法。

① 和法

有如下三个步骤:

第一步:对 A 的元素按列归一化,得 $\bar{A} = (\bar{a}_{ij})$

$$\bar{a}_{ij} = \frac{a_{ij}}{\sum_{i=1}^n a_{ij}} \quad i, j = 1, 2, \dots, n \quad (6.20)$$

第二步:将归一化后的 $\bar{A}$ 按行相加,得 $\bar{W} = [\bar{w}_1, \bar{w}_2, \dots, \bar{w}_n]^T$

$$\bar{w}_i = \sum_{j=1}^n \bar{a}_{ij} \quad (6.21)$$

第三步:对 $\bar{W}$ 归一化,得 $w = [w_1, w_2, \dots, w_n]^T$

$$w_i = \frac{\bar{w}_i}{\sum_{i=1}^n \bar{w}_i} \quad (6.22)$$

从求得的权重向量 W 可得各方案的排序结果。

例如,有一比较判断矩阵 A 为

$$A = \begin{bmatrix} 1 & 1 & 5 & 7 \\ 1 & 1 & 5 & 7 \\ 1/5 & 1/5 & 1 & 3 \\ 1/7 & 1/7 & 1/3 & 1 \end{bmatrix}$$

用和法求得:

$$\bar{A} = \begin{bmatrix} 0.427 & 0.427 & 0.441 & 0.389 \\ 0.427 & 0.427 & 0.441 & 0.389 \\ 0.085 & 0.085 & 0.088 & 0.167 \\ 0.061 & 0.061 & 0.029 & 0.056 \end{bmatrix} \quad (6.23)$$

$$W = \begin{bmatrix} 0.421 \\ 0.421 \\ 0.106 \\ 0.052 \end{bmatrix} \quad (6.24)$$

② 特征根法

对于判断矩阵 A , 定义 $\lambda_{\max}$ 为 A 的最大特征根, w 是相应的特征向量, 有下列关系

$$Aw = \lambda_{\max} w \quad (6.25)$$

由上式所得到的 w 经归一化可求得权重向量。

(4) 一致性检验

人们总是希望所获得的判断矩阵 A 是一致性矩阵。但在实际的判断中完全一致是很难的, 问题是多大的不一致性是可以接受的, 定义一致性指标 $C. I.$ 为

$$C. I. = \frac{\lambda_{\max} - n}{n - 1} \quad (6.26)$$

通常 $\lambda_{\max} > n$, 故 $C. I.$ 大于 0; 如 A 为完全一致性矩阵, 则 $\lambda_{\max} = n$, $C. I.$ 等于 0, 经过多次取样可计算得到 $C. I.$ 的平均随机一致性指标 $R. I.$ 。对不同的 n , 取足够的样本数, 计算得到 n 为 3 到 11 时的 $R. I.$ 值如表 6-5 所示。

$$\text{如} \quad \frac{C. I.}{R. I.} < 0.1 \quad (6.27)$$

则认为比较判断矩阵 A 的一致性是可以接受的。

表 6-5

矩阵阶数	3	4	5	6	7	8	9	10	11
R. I. 值	0.515	0.893	1.119	1.249	1.345	1.420	1.462	1.487	1.516
最小 C. I. 值	0	0	0.109	0.171	0.357	0.519	0.579	0.857	0.929
最大 C. I. 值	3.220	2.857	2.676	2.596	2.320	2.141	2.122	2.061	2.117

6.8 决策支持系统

决策支持系统(Decision Support Systems)是决策科学与计算机技术相结合的产物,简称 DSS。计算机应用于管理始于计算机出现不久的 20 世纪 50 年代,开始主要用于简单的记账和计算类的数据处理,提出了分离式单独处理某一具体事务的(如财务、人事、销售、库存等)事务处理系统,用于减轻一般管理人员处理日常事务中的原始数据的负担。20 世纪 60 年代 J. D. Gallamber 提出了管理信息系统(MIS)的设想,可以完成一个部门或一个单位大多数日常管理事务的数据处理任务。另外,MIS 中也开始具有一些预测功能,为各层管理人员的工作带来了许多便利,提高了工作效率。

随着 MIS 的进一步发展,人们也对它提出了新的要求,希望 MIS 能够在系统预测和决策中发挥更大的作用,辅助或支持人进行预测和决策。20 世纪 70 年代 S. Morton 提出了决策支持系统 DSS,用来支持决策者进行决策,特别是解决非结构化问题的决策。

6.8.1 决策支持系统的定义

决策支持系统尚没有一致公认的和确切的定义。1982 年 R. H. Sprague 和 E. D. Carlson 综合自己的观点,提出了决策支持系统的如下 5 个基本特征:

- ① 对管理人员面临的非结构化或半结构化决策问题提供支持;
- ② 应用计算机技术,把模型定量分析方法和数据管理方法结合起来并予以实现;
- ③ 对环境变化和决策问题的变化要有一定的柔性;
- ④ 人一机界面友好,便于非计算机专业人员以会话方式使用;
- ⑤ 支持决策者而不是代替决策者。

所谓结构化或非结构化决策问题,是著名系统科学专家西蒙(Simon)提出来的,凡是在决策前能够对决策环境和求解规划准确识别的问题,可以用标准的方法解决,是一种有章可循的程序决策,称为结构化决策;凡是难以在决策前对决策环境和求解规划准确识别的问题,只能凭决策者的经验、直觉做出应变的决策,一般是无章可循的非程序决策,称为非结构化决策;而大多数决策问题是介于这两者之间的情况,称为半结构化决策问题。

有人认为只要能够为决策提供某种支持的计算机系统就叫做决策支持系统。按照这样的定义,所有的 MIS 系统都是决策支持系统或包含决策支持系统。这样定义会模糊了 MIS 和 DSS 的关系,无助于 DSS 的深入研究。通常,可以认为 DSS 是 MIS 的顶层部分,也可以作为单独的系统存在。

决策支持系统的主要作用是为管理决策人员提供人难于或不便收集的信息,提供高效

处理大量信息的模型,支持决策者求解决策问题。在 DSS 中,人工智能是处于非常重要位置的。其中,DSS 更重视模型库和方法库的作用,MIS 的主要成分是数据库。也可以说,MIS 是初级的 DSS,DSS 是高级的 MIS。

6.8.2 决策支持系统的作用及应用

决策支持系统一般由数据库、模型库、方法库、知识库和相应的系统管理软件组成,按用途,DSS 可划分为:

(1) 为宏观系统决策服务的 DSS,如为国家、省、部、市、区、县等决策部门制订发展战略、发展计划、政策分析和管理服务的决策支持系统;

(2) 为微观系统决策服务的 DSS,如为企业、事业单位制订发展战略、发展规划、经管策略和管理服务的决策支持系统。

在宏观系统中应用 DSS 的研究与开发尚处于起步阶段,在我国当前所处的特殊历史时期,各方面的改革不断深入,不确定因素特别多,因此,开发和应用 DSS 的难度都非常大。在信息化时代的今天,国家在一系列的重要文献中都特别提出要以信息化带动工业化,发展以信息技术为带头学科的高科技。这些都为 DSS 的开发和应用创造了良好的环境。

在微观系统应用 DSS 的研究和开发,相对于宏观系统,发展比较好,许多企业结合 CAD、CAM、MRP-II、MIS 等系统的开发,结合控制网络的开发,形成 CIMS 和 CIPS,在石油、化工、冶金、机械等行业的企业中得到较好的应用。

习 题

6.1 什么是决策科学? 你知道哪些决策科学的代表性著作和代表性人物?

6.2 决策科学在人类历史发展过程中是如何发挥作用的? 请举例说明。

6.3 你认为决策科学与技术发展的大趋势将会给未来人类社会带来那些重大影响?

6.4 个人与集体在决策时应该如何应用决策模型,如何避免或减少失误?

6.5 某单位拟在一固定日期举办一次大型活动,有甲、乙、丙三个地点可供选择。在不同地点举办活动,由于天气情况(晴 S_1 、阴 S_2 、雨 S_3)不同,活动可能产生的效益差异较大,如下表所示。从气象部门得知届时各种天气出现的概率分别为 $P_1=0.5$ 、 $P_2=0.3$ 、 $P_3=0.2$,试用决策树法或表格法选择较佳方案;如果 P_1 、 P_2 、 P_3 分别为 0.2、0.4、0.4,结果又如何?

天 气	晴 S_1 $P_1=0.5$	阴 S_2 $P_2=0.3$	雨 S_3 $P_3=0.2$
地 点			
甲	7	6	6
乙	8	8	-2
丙	10	9	-3

6.6 上题中,如果不知道各种天气的出现概率,试用不确定型决策的各种准则来进行决策分析,并将结果加以比较。

6.7 什么是非结构化决策问题? 如何对这类问题进行研究和决策?

6.8 如果有北京、南京、武汉、上海、广州等五个城市可供某投资项目进行地点选择,试设想一个具体的投资计划并用层次分析法进行决策分析。

7 系统分析与系统建模

本章提要

在处理实际工程系统或管理系统时,应该如何入手?应该如何对系统进行分析?这是本章主要解决的问题。首先在介绍系统分析基本概念的基础上,重点介绍了系统分析的方法和步骤。系统分析强调的是应用系统模型对系统进行定量分析,所以在这一章还介绍了系统建模的主要方法,包括:状态空间法、结构模型解析法,还介绍了小样本问题的系统建模方法。

7.1 系统分析概述

7.1.1 系统分析的基本概念和作用

所谓系统分析,就是运用系统理论和方法对系统进行全面综合分析,以求得系统总体优化的科学方法。系统分析包括系统目标分析、确定系统可行方案、系统建模、系统优化和系统评价。系统分析的目的是辅助决策者,给决策者提供最优系统方案和决策依据。

系统分析的目的在于:通过分析比较各种替代方案的费用、效益、功能和可靠性等各项技术经济指标,得出决策者所必需的资料和信息,以便获得最优方案。系统分析的主要工具是电子计算机,其主要方法是系统建模和最优化方法,如规划论、排队论等等。

系统分析在整个系统建立过程中处于非常关键的地位,一般说来,系统建立过程分为三个阶段。(如图 7-1 所示)第一阶段为系统规划阶段。其主要任务是定义系统的概念,明确建立系统应具备的环境条件及估计系统所需的各种制约条件,最后制订系统开发计划书。第二阶段是系统设计阶段。首先对系统进行概要设计,其内容包括各种替代方案的建立,然后进行系统分析,分析项目包括:目的、替代方案、费用和效益、模型及评价基准等,在分析的基础上确定系统设计方案,并进行详细设计。第三阶段是系统制造和运行阶段,首先是对系统中一些关键项目进行试验和试制,然后进行工艺设计、制造、调试和投入运行,当

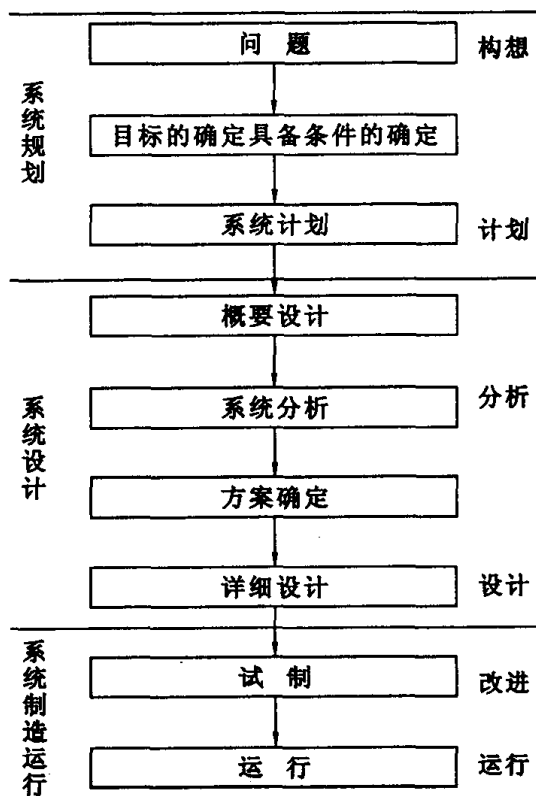


图 7-1 系统建立过程图

然运行前应制订运行、维护方法。

在以上三阶段中,系统分析起着承上启下的作用,它的任务首先是要分析和确定系统规划阶段的有关项目,如对系统概念的定义、分析和确定以及系统目标的分析和确定。然后对概要设计中的有关替代方案进行分析,根据分析结果来确定方案,进行详细设计。可见,系统分析是建立系统必不可少的一环。

7.1.2 系统分析的要素、原则和步骤

系统分析的要素主要有以下五个方面:

(1) 目的

对系统分析人员来说,最初的也是最重要的任务就是要充分了解建立系统的目的和要求。目的和要求既是建立系统的根据,也是系统分析的出发点。只有充分理解和掌握系统的目的和要求,才能进一步分析系统的目的和要求是否确切、完善和合理,才能为系统分析奠定良好的基础。

(2) 替代方案(可行方案)

在概要设计阶段,可以制订出为达到同一目的的各种不同的替代方案。在一般情况下,不同方案各有利弊,究竟何种方案最优,需要对各种方案进行分析、比较。

(3) 费用和效益

建立一个系统,需要有投资,系统建成以后,必然有效益;如果将投资费用和效益都折算成货币形式进行比较,则效益大于费用的设计方案是可取的,反之是不可取的。

(4) 模型

模型是描述实体系统的映像。根据需要建立的模型可以用来预测各替代方案的性能、费用和效益,以利各种方案的分析比较,同时可凭借一定的模型来有效地求得系统设计所需要的参数。

(5) 评价基准

是指确定各种替代方案优先选用顺序的标准。如在评价系统的费用和效益时,可以从三种基准中选用。若以替代方案效益为相同基准,则选择费用最小的方案为最优方案;若以各方案费用相同为基准,则选效益最大的方案为最优方案;若以效用费用比为基准,则选择比值最大的方案为最优方案。

系统分析的原则是:

(1) 内部因素与外部因素相结合

系统的内部因素往往是可控的,而外部因素往往是不可控的,系统的功能或行为不仅受到内部因素的作用,而且受到外部因素的影响和制约,因此对系统进行分析,必须把内外部各种有关因素结合到一起来考虑。通常的处理办法是,把内部因素选为决策变量,把外部因素作为约束条件,用一组联立方程式来反映它们之间的相互关系。

(2) 当前利益与长远利益相结合

选择最优方案,不仅要从目前利益出发,而且还要同时考虑长远利益,要两者兼顾。如果两者发生矛盾,应该坚持当前利益服从长远利益的原则。

(3) 局部效益与总体效益相结合

局部的最优并不意味着总体最优。总体的最优往往要求局部放弃最优而实现次优或次

次优。所以进行系统分析,必须坚持“系统总体效益最优、局部效益服从总体效益”的原则。

(4) 定性分析与定量分析相结合

定量分析是指采用数学模型进行的数量指标的分析,但是一些政治因素与心理因素、社会效果与精神效果目前还无法建立数学模型进行定量分析,只能依靠人的经验和判断力进行定性分析。因此在系统分析中,定性分析不可忽视,必须把定性分析与定量分析结合起来进行综合分析,或者交叉地进行,这样才能达到系统选优的目的。

系统分析的步骤可概括为:

(1) 系统目的分析和确定

分析和确定对象系统的目的和目标,分析和定义系统需要的功能,进而以这些数据作出模型,进行仿真。研讨成功的可能性,借以得到模型化所需要的技术条件。

(2) 模型化

这是系统分析过程中比较重要、工作量比较大的一个步骤。模型化的作用是为对象系统建立所需的各种模型。

(3) 系统最优化

它的作用在于运用最优化的理论和方法,对若干替换模型进行最优化,求出若干个替换解。

(4) 系统评价

根据替换解,考虑前提条件、假设条件和约束条件,在经验和知识的基础上决定满意解,从而为选择系统设计方案提供足够信息。

7.1.3 系统分析的方法

系统分析没有一套特定的普遍适用的技术方法,随着分析的对象不同,分析的问题不同,所使用的具体方法可能很不相同。一般说来,系统分析的各种情况可分为定性和定量的两大类。定量方法适用于系统结构清楚,收集到的信息准确,可建立数学模型等情况。在本书以后各章节中将陆续介绍各种定量的系统分析方法,例如投入产出分析法、效益成本分析法等。如果要解决的问题涉及的系统结构不清,收集到的信息不太准确,或是由于评价者的偏好不一,对于所提方案评价不一致等,难以形成常规的数学模型时,可以采用定性的系统分析方法,例如目标一手段分析法、因果分析法、KJ法等。

(1) 目标一手段分析法

目标一手段分析法,就是将要达到的目标和所需要的手段按照系统展开,一级手段等于二级目标,二级手段等于三级目标,以此类推,便产生了层次分明、相互联系又逐渐具体化的分层目标系统。在分解过程中,要注意使分解的分目标与总目标保持一致,分目标的集合一定要保证总目标的实现。分解过程中,分目标之间可能一致,也可能不一致,甚至是矛盾的,这就需要不断调整,使之在总体上保持协调。将总目标分解为若干个阶层的分目标,需要有很大的创造性和掌握丰富的科学技术知识与实践经验。目标分解需反复地进行,直到认为满意为止。

目标一手段分析法实质是运用效能原理不断进行分析的过程。图 7-2 是发展能源的目标一手段分析图。要发展能源,其手段主要有发展现有能源生产、开发研究新能源和节约能源,而节约能源的主要手段是综合利用能源和开发节能设备。

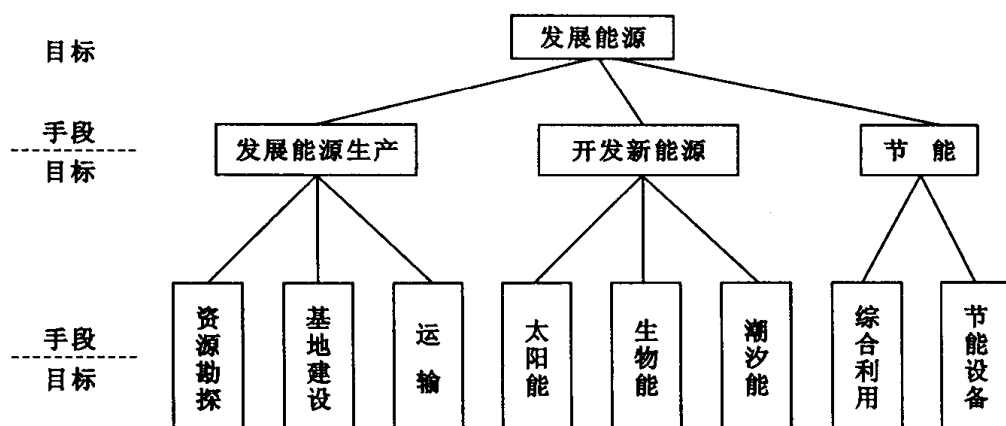


图 7-2 发展能源的目标—手段分析图

(2) 因果分析法

它是利用因果分析图来分析影响系统的因素，并从中找出产生某种结果的主要原因的一种定性分析方法。

系统某一行为(结果)的发生，绝非是一种或两种原因所造成，而往往是由于多种复杂因素的影响所致。为了分析影响系统的重要因素，找出产生某种结果的主要原因，系统分析人员广泛使用了一种简便而有效的定性分析方法——因果分析法。这种方法是在图上用箭头表示原因与结果之间的关系(如图 7-3 所示)，形象简单，一目了然，特别是在分析的问题越复杂时越能发挥其长处，因为它把人们头脑中所想问题的结果与其产生的原因结构图形化、条理化。在许多人集体讨论一个问题时，这种方法便于把各种不同意见加以综合整理，从而使大家对问题的看法逐渐趋于一致。

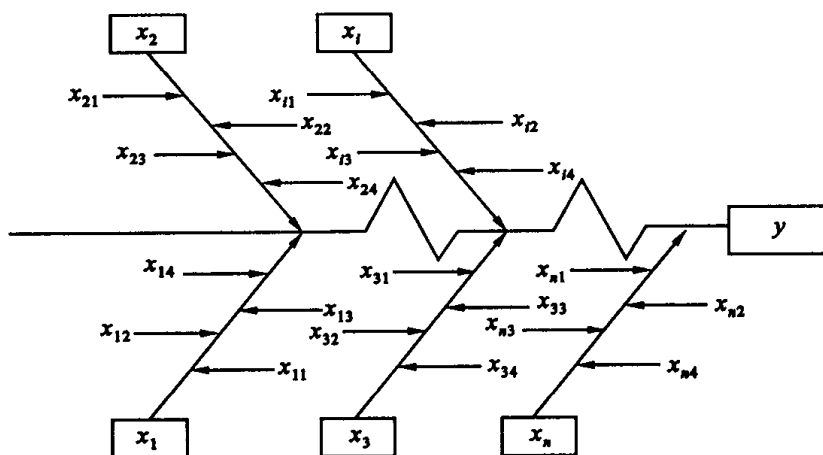


图 7-3 因果分析图

图 7-4 是某工厂在分析产品质量不稳定的原因时曾经用过的一张因果分析图，将其制成幻灯片，投影在工厂会议室的墙壁上，便于参加质量分析会的各方面专家集体会诊，因而比较快地找到了质量不稳定的主要原因及其应该采取的对策。

(3) KJ 法

KJ 法是一种直观的定性分析方法，它是由日本东京工业大学的川喜田二郎(Kauakida Jir)教授开发的。

KJ 法是从很多具体信息中归纳出问题整体含义的一种分析方法。它的基本原理是：把

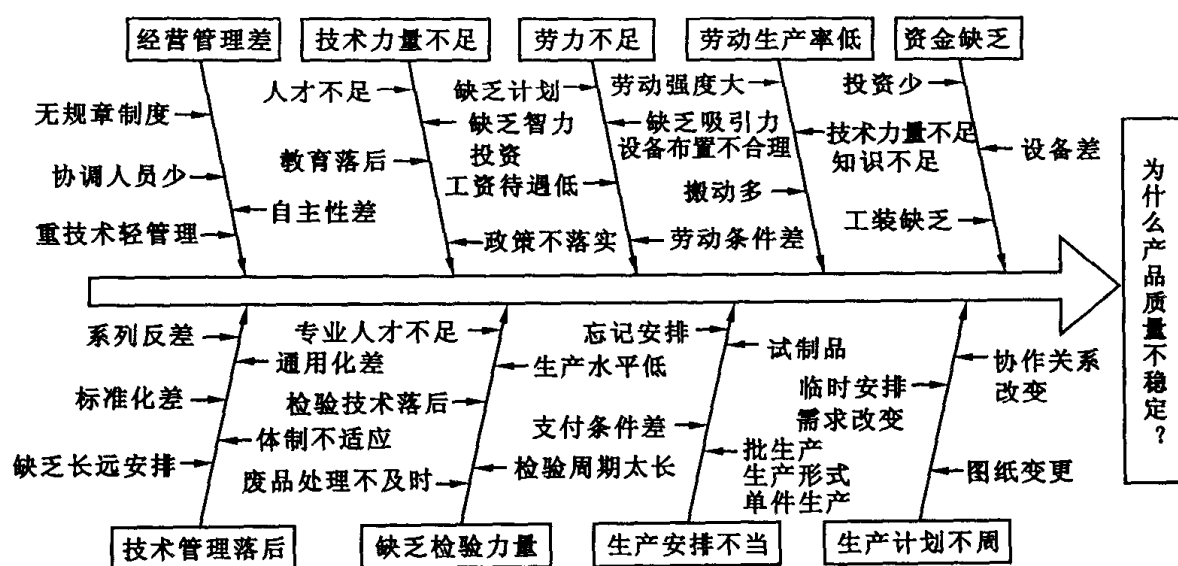


图 7-4 某工厂用于分析产品质量不稳定的因果分析图

一个个信息做成卡片，将这些卡片摊在桌子上观察其全部，把有“亲近性”的卡片集中起来合成为子问题，依次做下去，最后求得问题整体的构成。这种方法把人们对图形的思考功能与直觉的综合能力很好地结合起来，不需要特别的手段和知识，不论是个人或者团体都能简便地实行，因此，是分析复杂问题的一种有效的方法。

KJ 法的实施按下列步骤进行：

- ① 尽量广泛地收集与问题可能有关的信息，并用关键的语句简明地表达出来。
- ② 一个信息做一张卡片，卡片上的标题记载要简明易懂。如果是个体实施，则要在记载前充分协商好内容，以防误解。
- ③ 把卡片摊在桌子上通观全局，充分调动人的直觉能力，把有“亲近性”的卡片集中到一起作为一个小组。
- ④ 给小组取个新名称，其注意事项同步骤①。这个小组是由小项目（卡片）综合起来的，应把它作为子系统来登记。这个步骤不仅要凭直觉，而且还运用综合和分析能力发现小组的意义所在。
- ⑤ 重复步骤③和④，分别形成小组、中组和大组，但对难以编组的卡片不要勉强地编组，可把它们单独放在一边。
- ⑥ 把小组（卡片）放在桌子上进行移动，根据小组间的类似关系、对应关系、从属关系和因果关系等进行排列。
- ⑦ 将排列结果画成图表，即把小组按大小用粗细线框起来，把一个个有关系的框用“有向枝”（带箭头的线段）连接起来，构成一目了然的整体结构图。
- ⑧ 观察结构图，分析它的含义，取得对整个问题的明确认识。

在日本，KJ 法的应用范围很广泛，有很多实例。

近年来，我国许多系统工程人员也已使用这种方法分析社会经济问题。例如，《40 年后我国社会的老龄化问题》就是应用 KJ 法进行分析的。

针对这个问题，组织了由系统工程人员、经济学家、社会学家和计划生育部门领导干部、老中青代表等人组成的研究小组，经过集体的创造性思考，得到了如表 7-1 所示的一些信息。应该指出，实际上产生的看法比这还要多得多，但为了便于 KJ 法的运用，已经进行了

若干次归纳整理,最后合并成 45 条信息。然后按照以上步骤进行工作,得到了如图 7-5 所示的结构模型图。

表 7-1 社会老龄化问题的有关信息

编 号	内 容	编 号	内 容
1	劳动力不足	24	衣着式样不适应
2	家庭不和睦	25	社会道德观念不适应
3	出生率下降	26	体育设施不适应
4	需照顾的老人增多	27	兵源不足
5	平均寿命增加	28	经济缺乏竞争力
6	生活水平下降	29	工资结构不合理
7	老年人孤独	30	生产成本上升
8	思想僵化	31	社会需求改变
9	社会缺乏活力	32	生活环境变化
10	“代沟”加深	33	食物结构改变
11	老年医疗设施不足	34	服务行业增加
12	能源消耗大	35	服务人员不足
13	住房紧张	36	教育设施不适应
14	组织新家庭阻力大	37	娱乐设施不适应
15	交通设施不适应	38	青年人受压抑
16	年轻人提拔难	39	不易产生新思想
17	管理制度不适应	40	不易出研究成果
18	教育设施不适应	41	社会矛盾增加
19	社会福利开支大	42	年轻人缺乏独立性
20	要求退休年龄增大	43	年轻人嫌弃老年人
21	青少年感到孤独	44	经济增长减慢
22	老年人不受尊敬	45	鲜艳色彩减少,生活单调枯燥
23	社会创造力下降		

从得到的结构模型图 7-5,可以对所要分析的问题作如下的解释:40 年后,由于人口出生率的下降和人均寿命的延长,老年人口在总人口中的比重将会增加,据预测,65 岁以上的老人比重将超过 15%,使我国人口出现老龄化问题。对此,如不采取积极有效的措施加以解决,就将对社会经济产生一系列重大的影响:一方面,人口老化,人力资源缺乏,创造力下降,社会负担加重,限制了生产力的发展和经济的增长;另一方面,老年人增多,原有的文教体卫设施已不适应,衣食住行和社会需求发生了很大变化,不仅给医疗保健和社会服务工作增加了压力,而且会产生新的社会矛盾:老年人因要求得不到满足会产生不幸福感,而青年人因晋升和施展才能方面受到压抑会出现不满情绪,这将导致“代沟”和社会矛盾的加深,经济增长减慢与社会矛盾加深交织在一起,相互影响,必将阻碍社会的进步和发展。

7.1.4 系统分析应用举例

【例 7.1】 阿拉斯加原油输送方案的系统分析

这里所分析的是如何由阿拉斯加东北部的普拉德霍湾油田向美国本土运输原油的问题。

(1) 任务和环境

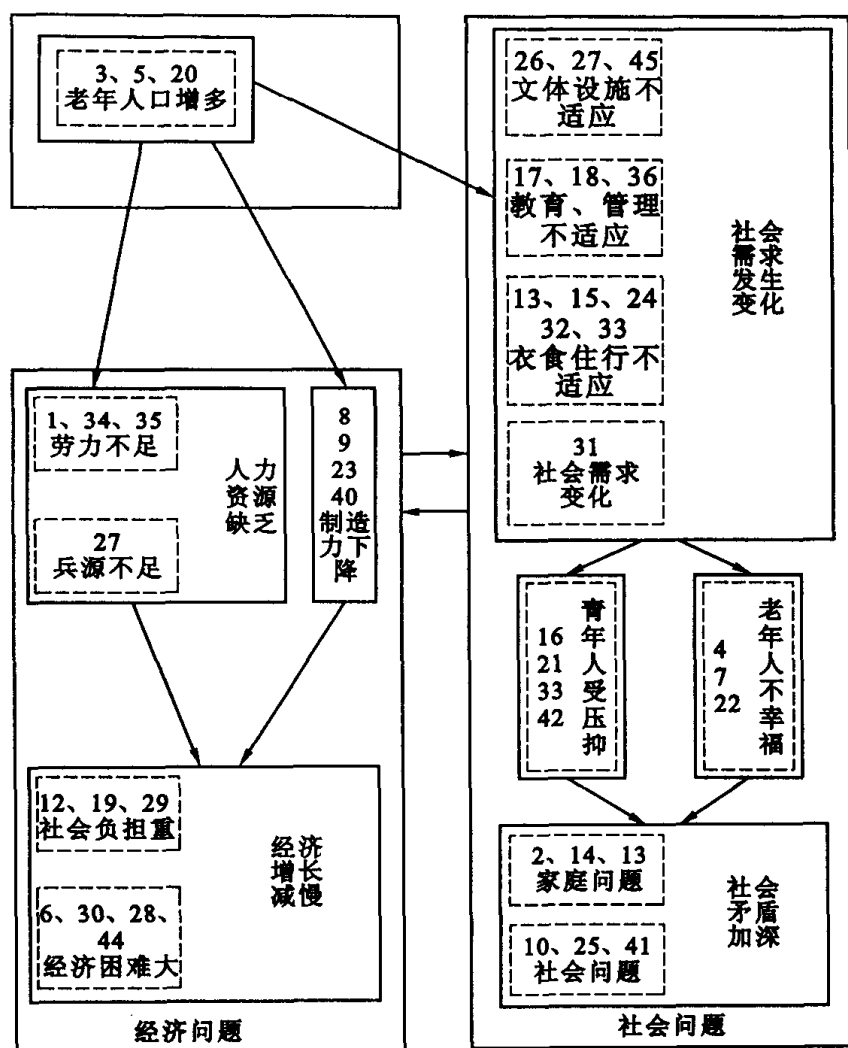


图 7-5 社会老龄化问题的结构模型图

要求每天运送 200 万桶。油田处在北极圈内,海湾长年处于冰封状态,陆地更是常年冰冻,最低气温达零下 50℃。

(2) 提出竞争方案

方案竞争的第一阶段,提出了两个方案:

方案 I: 由海路用油船运输;

方案 II: 用带加温系统的油管输送。

方案 I 的优点是每天仅需四至五艘超级油轮就可满足输送量的要求,似乎比铺设油管省钱,但存在的问题是:第一,要用破冰船引航,既不安全又增加了费用;第二,起点和终点都要建造大型油库,这又是一笔巨额花费,而且考虑到海运可能受到海上风暴的影响,油库的储量应在油田日产量的十倍以上。归纳起来这一方案的主要问题是:不安全、费用大、无保证。

方案 II 的优点是可以利用成熟的管道输油技术,但存在的问题是:第一,要在沿途设加温站,这样一来管理复杂,而且要供给燃料,然而运送燃料本身又是一件相当困难的事情;第二,加温后的输油管不能简单地铺在冻土里,因为冻土层受热溶化后会引引起管道变形,甚至造成断裂。为了避免这种危险,有一半的管道需要用底架支撑和作保温处理,这样架设管道

的成本费用要比铺设地下油管高出三倍。

(3) 决策人员的处理策略

① 考虑到安全和供油的稳定性,暂把方案Ⅱ作为参考方案做进一步的细致的研究,为规划做准备;

② 继续拨出经费,广泛邀请系统分析人员提出竞争的新方案。

(4) 提出了竞争方案Ⅲ

其原理是把含 10%~20%氯化钠的海水加到原油中去,使在低温下的原油成乳状液,仍能畅流,这样就可以用普通的输油管道运送了。这个方案获得了很高的评价,并取得了专利。其实,这一原理早就用于制作汽车的防冻液了,把这一原理运用到这个工程中来,并断定它能解决问题,这是一个有价值的创造。

(5) 提出竞争方案Ⅳ

正当人们在称赞方案Ⅲ的时候,另有人提出了竞争方案Ⅳ。该方案提出者对石油的生成和变化有丰富的知识,他们注意到埋在地下的石油原来是油、气合一的,这时它们的熔点是很低的,经过漫长的年代以后,油气才逐渐分离。他们提出将天然气转换为甲醇以后再加到原油中去,以降低原油的熔点,增加流动性,从而用普通的管道就可以同时输送原油和天然气。与方案Ⅲ相比,不仅不需要运送无用的海水,而且也不必另外铺设输送天然气的管道了。这一方案的出现使得人们赞赏不已。由于采用这一方案,仅管道铺设费就节省了近 60 亿美元,比方案Ⅲ省了一半。

从这个例子我们看到了系统分析的实际价值。如果当初仅在方案Ⅱ、Ⅲ上搞优化,即确定最好的管道直径、壁厚、加压和距离等等,是无论如何也得不到方案Ⅳ所达到的巨大效益的。

这个例子同时也说明了系统分析人员的工作性质和应该具有的知识结构,以及系统分析与专业工程技术工作之间相辅相成的关系。

7.2 系统模型概述

要对大型复杂系统进行有效的分析、研究并得到有效的结果,就必须首先建立系统模型,然后才能借助模型对系统进行定量的、或者定量与定性相结合的分析。因此建立系统模型是系统分析的工具和基础。

7.2.1 系统模型的定义与特征

(1) 定义

系统模型是一个系统某一方面本质属性的描述,它以某种确定的形式(例如文字、符号、图表、实物、数学公式等)提供关于该系统的知识。系统模型一般不是系统对象本身,而是现实系统的描述或抽象。系统是复杂的,系统的属性也是多方面的。对于大多数研究目的而言,没有必要考虑对象的全部属性,因此,系统模型知识系统某一方面本质属性的描述,本质属性的选取完全取决系统工程研究的目的。所以,对同一个系统根据不同的研究目的,可以建立不同的系统模型。例如,图 7-6 所示是一个 RLC 网络系统,如果考虑输入电压 $u(t)$ 与输出电压 $u_c(t)$ 之间的关系,则可得到如下数学模型(应用传递函数的算子阻抗即可得下

式):

$$\frac{u_c(s)}{u(s)} = \frac{1}{LCs^2 + RCs + 1}$$

$$u_c(t) = L^{-1} \left[\frac{1}{LCs^2 + RCs + 1} \right] u(t) \quad (7.1)$$

另一方面,同一种模型也可以代表多个系统,

例如

$$y = kx \quad (k \text{ 为常量})$$

几何上:代表一条通过原点的直线;

代数上:表示比例关系;

设 $k = \pi$, x 表示伸长量,则 y 表示弹簧力大小;

当 $k = a$ 表示加速度, $x = m$ 表示质量,则 y 表示物体所受外力大小等等。

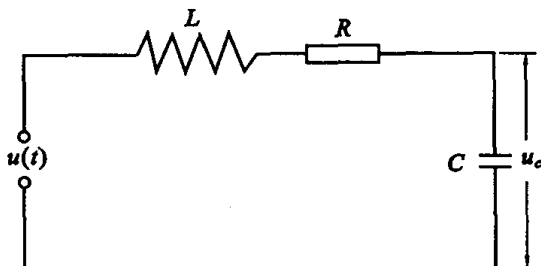


图 7-6 RLC 电路

(2) 特征

系统模型反映着实际系统的主要特征,但它又高于实际系统而具有同类问题的共性。因此,一个实用的系统模型应该具有如下三个特征:

- ① 它是现实系统的抽象或模仿;
- ② 它是由反映系统本质或特征的主要因素构成的;
- ③ 它集中体现了这些因素之间的关系。

7.2.2 使用系统模型的必要性

人类认识和改造客观世界的研究方法,一般来说主要有三种,即实验法、抽象法、模型法。实验法是通过客观事物本身直接进行科学实验来进行研究的,因此局限性比较大。抽象法是把现实系统抽象为一般的理论概念,然后进行推理和判断,因此这种方法缺乏实体感,过于概念化。模型法是在对现实系统进行抽象的基础上,把它们再现为某种事物的、图画的或数学的模型,然后通过模型来对系统进行分析、对比和研究,最终得出结论。由此可见,模型法既避免了实验法的局限性,又避免了抽象法的过于概念化,所以它成为现代工程中一种最常用的研究方法。

在工程系统中,广泛使用系统模型还基于以下五个方面的考虑:

① 系统开发的需要。在开发一个新系统时,由于此时系统尚未建立,无法直接进行实验,只能通过建造系统模型来对系统的性能进行预测,以实现对系统的分析、优化和评价。

② 经济上的考虑。对大型复杂系统直接进行实验,其成本是十分昂贵的,但是使用系统模型就便宜多了。

③ 安全上的考虑。对有些系统(如载人航天飞行器、核电站等)通过直接实验进行分析,往往是很危险的,有时甚至是根本不允许的。

④ 时间上的考虑。社会、经济、生态等系统,由于惯性大、反应周期很长,对其直接进行实验要等若干年之后才能看到结果,这是系统分析和评价所不允许的。而使用系统模型进行分析,很快就可以得到分析结果。

⑤ 系统模型容易操作,分析结果易于理解。有时对现实系统进行直接实验虽然是允许

的,而且也不过分费时、费钱,但此时采用系统模型仍然具有优越性。因为现实系统中包含的因素太多而且复杂,实验得到的结果往往难以直接与其中的某一因素挂钩,因此,直接实验的结果不易理解,而且实验过程中要改变系统参数也相当困难。但是,若使用系统模型情况就不一样了,由于系统模型突出了研究目的所要关注的主要特征,因此容易得到一个清晰的结果,而且在系统模型(尤其是数学模型)上进行参数修正也是非常容易的。

7.2.3 系统模型的分类

系统种类繁多,作为系统的描述——系统模型的种类也是很多的。如表 7-2 列出了按不同原则分类的系统模型,从中我们可以了解系统模型的多样性。

表 7-2 系统模型的一种分类方法

	分 类 原 则	模 型 种 类
1	按建模材料不同	抽象、实物
2	按与实体的关系	形象、类似、数学
3	按模型表征信息的程度	观念性、数学、物理
4	按模型的构造方法	理论、经验、混合
5	按模型的功能	结构、性能、评价、最优化、网络
6	按与时间的依赖关系	静态、动态
7	按是否描述系统内部特征	黑箱、白箱
8	按模型的应用场合	通用、专用
9	数学模型的分类	
	(1) 按变量形式分	确定性、随机性、连续型、离散型
	(2) 按变量之间的关系分	代数方程、微分方程、概率统计、逻辑

下面介绍常用的几种系统模型。一般将系统分为物理模型、文字模型和数学模型三大类,其中物理模型与数学模型又可分为若干种,如图 7-7 所示。

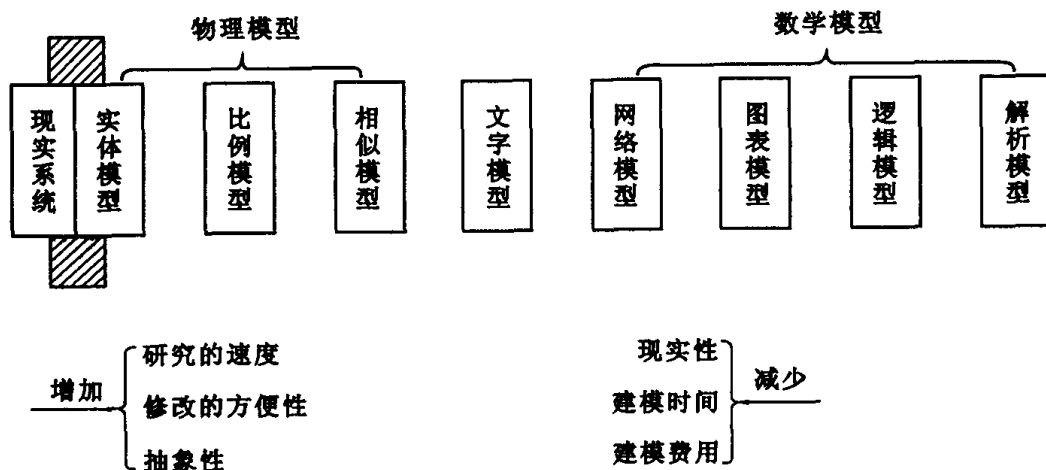


图 7-7 系统模型的分类与特征比较

各种模型的含义解释如下:

① 实体模型。即系统本身,当系统的大小刚好适合研究而又不存在危险时,就可以把系统本身作为模型。实体模型包含抽样模型,例如标准件的生产检验是从总体中抽取一定数量的样本进行的,样本就是实体模型。

② 比例模型。是放大或缩小的系统,使之适合于研究。

③ 相似模型。根据相似原理,利用一种系统去代替另一种系统。例如用电路系统代替机械系统、热力系统进行研究,则电路系统就是后二者的相似模型。

④ 文字模型。如技术报告、说明书等。在物理模型和数学模型都很难建立时,有时不得不用它描述研究结果。

⑤ 网络模型。用网络图来描述系统的组成元素以及元素之间的相互关系(包括逻辑关系与数学关系)。

⑥ 图表模型。用图像和表格描述的模型,它们相互转化,这里说的图像是指坐标系统的曲线、曲面和点等几何图形。

⑦ 逻辑模型。表示逻辑关系的模型,如方框图、程序单、模拟机排队图等。

⑧ 解析模型。用数学方程式表示的模型。

7.2.4 使用数学模型的好处

在系统工程中经常使用数学模型分析问题,其主要原因在于它具有如下好处:

① 它是定量分析的基础。在自然科学和工程技术领域里,数量不准将招致质量低劣,在社会学领域里,没有定量分析会使人心中无数,造成决策失误,引起不必要的混乱。因此,采用数学模型进行定量分析已成为当代自然科学和社会科学进一步发展的共同要求。

② 它是系统预测和决策的工具。可以利用系统已有的数据建立预测模型,用来预测系统的未来状态,为正确决策提供依据。

③ 它可变性好,适应性强,分析问题速度快,省时省钱,而且便于使用计算机。因此,它是所有模型中使用最广泛的一种。我们通常所说的系统建模,大多数情况下都是指建立系统的数学模型。

7.3 系统建模方法

系统建模是系统工程人员的重要工作之一。建立一个简明的适用的系统模型,将为系统的分析、评价和决策提供可靠的依据。建造系统模型,尤其是建造抽象程度很高的系统数学模型,是一种创造性劳动。因此有人讲,系统建模既是一种技术,又是一种“艺术”。

7.3.1 对系统建模的要求

对系统建模的要求可以概括为三条,即现实性、简明性、标准化。

(1) 现实性

即把一定程度上能够较好地反映系统的客观实际,应把系统本质的特征和关系反映进去,而把非本质的东西去掉,但又不影响反映本质的真实程度。也就是说,系统模型应有足够的精度。精度要求不仅与研究对象有关,而且与所处的时间、状态和条件有关。因此,为满足

现实性要求,对同一对象在不同情况下可以提出不同的精度要求。

(2) 简明性

在满足现实性要求的基础上,应尽量使系统模型简单明了,以节约建模的费用和时间。这也就是说,如果一个简单的模型已能使实际问题得到满意的解答,就没有必要去建一个复杂的模型,因为建造一个模型并求解是要付出很高代价的。

(3) 标准化

在建立某系统的模型时,如果已有某种标准模型可供借鉴,则应尽量采用标准化模型,或者对标准化模型加以某些修改,使之适合对象系统。

以上三条要求往往是相互抵触的,容易顾此失彼。例如,现实性和简明性就常常存在矛盾,如果模型复杂一些,虽然满足现实性要求,但建模和求解却相当困难,费时、费钱,同时也可能影响标准化的要求,为此,必须根据对象系统的具体情况妥善处理。一般的处理原则是:力求达到现实性,在现实性的基础上达到简明性,然后尽可能满足标准化。

7.3.2 系统建模应遵循的原则

根据对系统建模提出的三条要求,可以导出系统建模时应该遵循的四项原则是:

(1) 切题

模型只应包括与研究目的有关的方面,而不是对象系统的所有方面。例如,对一个空运指挥调度系统的研究,建模只需考虑飞机的飞行航向而无需考虑其飞行姿势。

(2) 清晰

一个大型复杂系统是由很多联系密切的子系统组成的,因此对应的系统模型也是由许多子模型(或模块)组成的。在子模型与子模型之间,除了保留研究目的所必要的信息联系外,其他的耦合关系要尽可能减少,以保证模型结构尽可能清晰。

(3) 精度要求适当

建立系统模型,应该视研究目的和使用环境不同,选择适当的精度等级,以保证模型切题、实用,而又不致花费太多。例如,一个受外力 F 作用下的物体 M ,其动力学系统的数学模型,在不同使用环境下有不同精度等级,应该适当选择。

当物体的运动速度 v 足够小时,可以忽略空气阻力的影响,其符合精度要求的数学模型为

$$F = M \frac{d^2 v}{dt^2}$$

当速度 v 提高到必须考虑空气阻力的影响时,则其符合精度要求的数学模型为

$$F = M \frac{dv}{dt} + kv^2$$

当物体的运动速度接近于光速 $3 \times 10^8 \text{ m/s}$ 时,按相对论原理,此时 M 将不是常数,因此其符合精度要求的数学模型为

$$F = \frac{d}{dt}(Mv) + kv^2$$

(4) 尽量使用标准模型

在建立一个实际系统的模型时,应该首先大量调阅模型库中的标准模型,如果其中某些可供借鉴,不妨先试用一下。如能满足要求,就应该使用标准模型,或者尽可能向标准模型

靠拢。

7.3.3 系统建模的主要方法

针对不同的系统对象,可以采取不同的方法建造系统模型,其中主要方法有以下五种:

(1) 推理法

对于内部结构和特性已经清楚的系统,即所谓的“白箱”(例如大多数的工程系统),可以利用已知的定律和定理,经过一定的分析和推理,得到系统模型。

(2) 实验法

对于那些内部结构和特性不清楚和不很清楚的系统,即所谓的“黑箱”或“灰箱”系统,如果允许进行实验性观察,则可以通过实验方法测量其输入和输出,然后按照一定的辨识方法,得到系统模型。

(3) 统计分析法

对于那些属于“黑箱”,但又不允许直接进行实验观察的系统(例如非工程系统多数属于此类),可以采用数据收集和统计分析的方法来建造系统模型。

(4) 混合法

大部分系统模型的建造往往是上述几种方法综合运用结果。

(5) 类似法

即建造原系统的类似模型。有的系统,其结构和性质已经清楚,但其模型的数量描述和求解却不好办,这时如果有另一种系统其结构和性质与之相同,因而建造出的模型也类似,但是该模型的建立及处理要简单得多,我们就可以把后一种系统模型看成是原系统的类似模型。利用类似模型,按对应关系就可以很方便地求出原系统的模型。例如很多机械系统、气动力学系统、水力学系统、热力学系统与电路系统之间某些现象彼此类似,特别是通过微分方程描述的动力学方程基本一致,因此可以利用研究的很成熟的电路系统来构造上述这些系统的类似模型。

上面针对不同情况提出了建造系统模型的五种方法(或思路)。应该指出的是,这些方法只能供系统建模者参考,而要真正解决系统建模问题还必须充分开发人的创造力,综合运用各种科学知识,针对不同的系统对象,或者建造新模型,或者巧妙地利用已有的模型,或者改造已有的模型,这样才能创造出更加适用的系统模型。因此,有人把建造系统模型看成是一种艺术,这说明建造系统模型确实需要充分发挥人的创造性,而不可能有现成的模式可以照搬。

7.4 状态空间法

一个大型系统,多属于多输入—多输出系统,且为时变系统。这时,为了分析系统的动态特性,往往以系统的状态变量来进行描述,这样的模型称为状态变量模型,建立这种模型的技术叫做状态空间法。

7.4.1 系统的状态和状态变量

系统的状态是指系统在已知 $t > t_0$ 时所有行为所需的足够变量 $X_1(t_0), X_2(t_0), \dots, X_n(t_0)$ 的最小集合。也就是说,系统的状态至少要用 n 个独立的状态变量 $X_1(t_0), X_2(t_0),$

..., $X_n(t_0)$ 才能充分地表现出来。而状态变量, 就是能完整地确定系统状态所必需的变量。例如图 7-6 所示一个 $R-L-C$ 回路受电压为 $u(t)$ 的电源激励。当初始条件全为零时, 电容作为输出端的电压 u_c 和电源电压 $u(t)$ 有如下关系:

$$LC\ddot{u}_c + RC\dot{u}_c + u_c = u \quad (7.2)$$

其传递函数为:

$$\frac{u_c(s)}{u(s)} = \frac{1}{LCs^2 + RCs + 1} \quad (7.3)$$

若令 $X_1 = u_c, X_2 = \dot{u}_c$, (7.2) 式可以化为:

$$\begin{aligned} \dot{X}_1 &= X_2 \\ \dot{X}_2 &= -\frac{1}{LC}X_1 - \frac{R}{L}X_2 + \frac{1}{LC}u \end{aligned} \quad (7.4)$$

X_1/X_2 即为描述系统的状态变量, 如用矩阵方程表示, 即

$$\dot{X} = \begin{bmatrix} \dot{X}_1 \\ \dot{X}_2 \end{bmatrix} = AX + BU \quad (7.5)$$

式中 $A = \begin{bmatrix} 0 & 1 \\ -\frac{1}{LC} & -\frac{R}{L} \end{bmatrix}$ —— 状态矩阵;

$B = \begin{bmatrix} 0 \\ \frac{1}{LC} \end{bmatrix}^T$ —— 控制矩阵。

如果用电压表测出电容上的电压 u_c 即 X_1 , 则可写出测量仪表的测得值, 即观测值:

$$Y = CX + e_1 \quad (7.6)$$

式中 $C = [C_1 \ 0]$ —— 观测矩阵或输出矩阵;

e_1 —— 测量误差。

式(7.6)称为输出方程。

上例是一个连续系统的状态空间模型, 推广到一般, 它可以从不同的方面来获得。

7.4.2 连续系统状态空间表达式

(1) 由微分方程导出状态空间表达式

设 n 阶线性定常系统的微分方程为

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1} y^{(1)} + a_n y = u$$

若取 $y(t), y^{(1)}(t), \cdots, y^{(n-1)}(t)$ 等 n 个状态变量, 作为变量集合, 且已知其初始条件 $y(0), y^{(1)}(0), \cdots, y^{(n-1)}(0)$ 和 $t \geq 0$ 时的输入 u

则令

$$\begin{aligned} x_1 &= y \\ x_2 &= y^{(1)} \\ &\vdots \\ x_n &= y^{(n-1)} \end{aligned}$$

此时方程可写成

$$\left. \begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= x_3 \\ &\cdots \\ \dot{x}_{n-1} &= x_n \\ \dot{x}_n &= -a_n x_1 - \cdots - a_1 x_n + u \end{aligned} \right\} \quad (7.7)$$

即

$$\dot{X} = AX + BU \quad (7.8)$$

式中

$$\dot{X} = [x_1 \quad x_2 \quad \cdots \quad x_n]^T$$

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix}$$

$$B = [0 \quad 0 \quad \cdots \quad 0 \quad 1]^T$$

输出方程为

$$Y = CX$$

式中 $C = [1, 0, 0, \cdots, 0]$; X —— n 维状态向量; U —— m 维控制向量; Y —— r 维输出向量; A —— $n \times n$ 维状态系数矩阵; B —— $n \times m$ 维控制系数矩阵; C —— $r \times n$ 维输出系数矩阵。

(2) 由传递函数导出状态空间表达式

已知描述函数的微分方程式可以进行拉氏变换, 求其系统的传递函数。相反, 若给出系统传递函数也可以求出其系统的微分方程式, 然后确定系统的状态变量, 确定系统的状态空间表达式。

例如某系统的方块图如图 7-8 所示, 由图可以求系统的闭环传递函数为

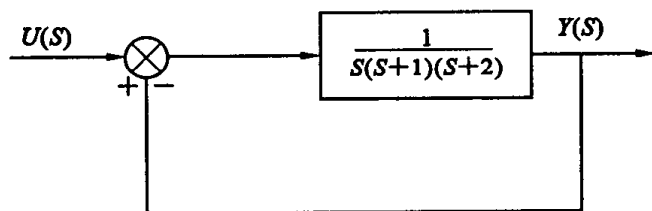


图 7-8 系统方框图

$$\frac{Y(S)}{U(S)} = \frac{1}{S(S+1)(S+2)+1} = \frac{1}{S^3 + 3S^2 + 2S + 1}$$

根据上述闭环传递函数求的系统的微分方程式为:

$$y^{(3)} + 3y^{(2)} + 2y^{(1)} + y = u \quad (7.9)$$

$$x_1 = y$$

选取

$$x_2 = \dot{x}_1 = \dot{y}$$

$$x_3 = \dot{x}_2 = \ddot{y}$$

为系统的状态变量, 则(7.9)式通过状态变量 X_1, X_2 及 X_3 改写成如下状态空间表达式

$$\dot{X} = \begin{bmatrix} \dot{X}_1 \\ \dot{X}_2 \\ \dot{X}_3 \end{bmatrix} = AX + BU$$

$$= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -2 & -3 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} U \quad (7.10)$$

$$Y = CX = [1 \ 0 \ 0] \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$$

(3) 状态变量的非唯一性

一般地说,传递函数法是确定系统的输入—输出特性的方法,而状态变量法是描述系统内在状态的方法。系统的状态变量可以根据不同的情况任意选取,也即从给定的同一个传递函数,可以用不同的状态变量集合来表示,因此状态变量模型不是唯一的。下面将以同一传递函数用三种不同的求解方法,求取状态变量,说明同一系统的状态变量,可以选择不同的集合。

设系统的传递函数 $G(S)$ 为:

$$G(S) = \frac{2S+2}{S^3+9S^2+26S+24}$$

① 采用直接程序法,即化为积分形式来处理

$$\frac{Y(S)}{U(S)} = \frac{2S^{-2} + 2S^{-3}}{1 + 9S^{-1} + 26S^{-2} + 24S^{-3}}$$

$$Y(S) = \frac{2S^{-2} + 2S^{-3}}{1 + 9S^{-1} + 26S^{-2} + 24S^{-3}} \cdot U(S)$$

由此可以画出系统的状态变量图(图 7-9)。

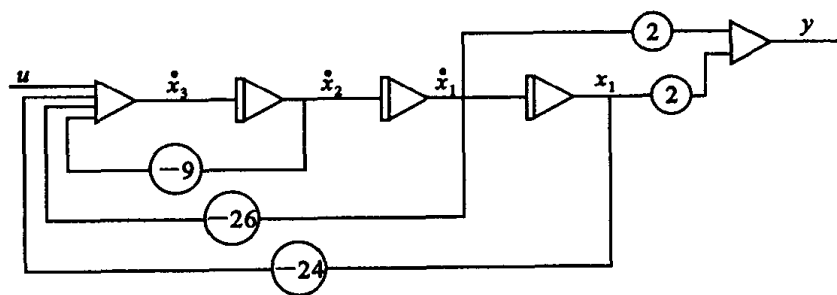


图 7-9 系统直接程序状态变量图

图中所选的状态变量为

$$x_3 = \int_0^t u dt + x_3(0)$$

$$x_2 = \int_0^t x_3 dt + x_2(0)$$

$$x_1 = \int_0^t x_2 dt + x_1(0)$$

∴ 状态方程和输出方程为

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = x_3$$

$$\dot{x}_3 = u - 24x_1 - 26x_2 - 9x_3$$

$$Y = 2x_1 + 2x_2$$

$$\dot{X} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -24 & -26 & -9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u \quad (7.11)$$

即

$$Y = [2 \quad 2 \quad 0] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

② 采用并联程序法, 即化成为部分分式

$$G(S) = \frac{Y(S)}{U(S)} = \frac{a_1}{s + P_1} + \frac{a_2}{S + P_2} + \cdots + \frac{a_n}{S + P_n}$$

$$G(S) = \frac{2S + 2}{S^3 + 9S^2 + 26S + 24} = \frac{2S + 2}{(S + 2)(S + 3)(S + 4)}$$

$$= \frac{-1}{S + 2} + \frac{4}{S + 3} + \frac{-3}{S + 4}$$

由此可得并联程序状态图(图 7-10)。

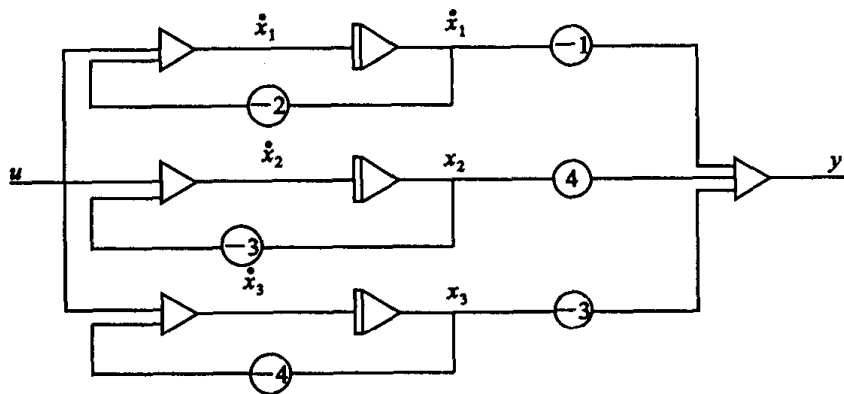


图 7-10 并联程序的状态变量图

由图可得状态方程和输出方程

$$\begin{aligned} \dot{x}_1 &= u - 2x_1 \\ \dot{x}_2 &= u - 3x_2 \quad \text{及} \quad y = -x_1 + 4x_2 - 3x_3 \\ \dot{x}_3 &= u - 4x_3 \end{aligned}$$

$$\dot{X} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} -2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} u \quad (7.12)$$

$$Y = [-1 \quad 4 \quad -3] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

③ 采用迭代程序法即化成连乘形式

$$\frac{Y(S)}{U(S)} = \frac{2S+2}{S^3+9S^2+26S+24} = \frac{1}{S+2} \cdot \frac{2}{S+3} \cdot \frac{S+1}{S+4}$$

由上式得迭代方程的状态向量

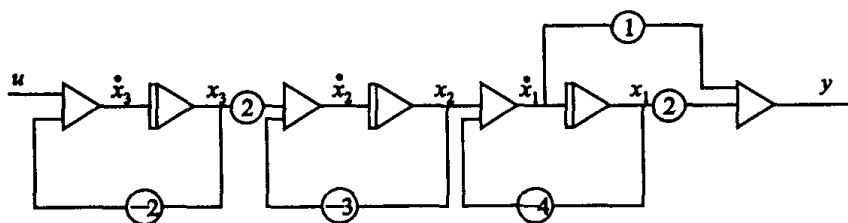


图 7-11 迭代程序的状态向量图

由图 7-11 可得状态方程和输出方程

$$\dot{x}_3 = u - 2x_3$$

$$\dot{x}_2 = -3x_2 + 2x_3$$

$$\dot{x}_1 = -4x_1 + x_2$$

及

$$y = 2x_1 + x_1 = 2x_1 - 4x_1 + x_2 = -2x_1 + x_2$$

或

$$\dot{x} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} -4 & 1 & 0 \\ 0 & -3 & 2 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u$$

$$Y = \begin{bmatrix} -2 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad (7.13)$$

上例说明,所选状态变量不同,所得的状态方程和输出方程也不同,这说明了状态变量模型的非唯一性。同时,从数学意义上看,这些方程都能满足输入—输出关系,都能正确反映系统的行为。从物理意义上看,这些方程都代表了不同的系统结构。当已知系统特性(如传递函数)后,用怎样的结构来实现特性,这就与选择的状态变量有关,一般采用利于测量和计算的参数作为状态变量。

(4) 矩阵的特征值、特征向量、对角化及特征值的不变性

① 矩阵特征值 λ 可以通过求取特征方程 $|\lambda I - A| = 0$ 的根得到。

在上述例子中, A 具有三种不同的矩阵形式,在式(7.11)中

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -24 & -26 & -9 \end{bmatrix} \text{ 代入方程 } |\lambda I - A| = 0 \text{ 中}$$

$$\text{得} \quad \left| \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix} - \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -24 & -26 & -9 \end{bmatrix} \right| = 0$$

$$\begin{aligned} \text{即} \quad & \left| \begin{bmatrix} \lambda & -1 & 0 \\ 0 & \lambda & -1 \\ 24 & 26 & 9+\lambda \end{bmatrix} \right| = \lambda^2(\lambda+9) + 24 + 26\lambda \\ & = (\lambda+2)(\lambda+3)(\lambda+4) = 0 \end{aligned}$$

求得特征值为

$$\lambda_1 = -2, \lambda_2 = -3, \lambda_3 = -4$$

同样在(7.11)中,将 $A = \begin{bmatrix} -2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -4 \end{bmatrix}$ 代入特征方程 $|\lambda I - A| = 0$ 中

得 $\lambda_1 = -2, \lambda_2 = -3, \lambda_3 = -4$

通过(7.13)式中的 A ,也可求得相同的特征值,说明了特征值的不变性,反映系统内在规律的特征方程不变。

② 求特征向量 t 和变换矩阵 T

假定向量 t 使 $At = \lambda t$

$$\text{即 } \lambda t - At = 0 \quad (\lambda I - A)t = 0 \quad (7.14)$$

这时把不为零的矢量 t 叫做 A 的特征向量。

假如,设 $A = \begin{bmatrix} 0 & 1 \\ -6 & -5 \end{bmatrix}$ 代入 $(\lambda I - A)t = 0$ 的方程,

$$\text{得 } \begin{bmatrix} \lambda & -1 \\ 6 & \lambda + 5 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = 0 \quad \text{有 } \det \begin{bmatrix} \lambda & -1 \\ 6 & \lambda + 5 \end{bmatrix} = 0$$

求得 $\lambda_1 = -2 \quad \lambda_2 = -3$

当 $\lambda_1 = -2$ 时,有

$$\begin{bmatrix} -2 & -1 \\ 6 & -2 + 5 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = 0 \quad \therefore 2t_1 = -t_2$$

对应于 λ_1 的特征向量为:

$$t_1 = C_1 \begin{pmatrix} 1 \\ -2 \end{pmatrix} \quad (C_1 \text{ 为任意实数})$$

同理,当 $\lambda_2 = -3$ 时,得

$$\begin{bmatrix} -3 & -1 \\ 6 & -3 + 5 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = 0 \quad \therefore 3t_1 = -t_2$$

对应于 λ_2 的特征向量为

$$t_2 = C_2 \begin{bmatrix} 1 \\ -3 \end{bmatrix} \quad (C_2 \text{ 为任意实数})$$

当 λ 为相异实数时, t 可以由下式来决定,即

$$At_1 = \lambda_1 t_1, At_2 = \lambda_2 t_2$$

用特征向量 $t_1 \ t_2$ 建立如下的变换矩阵 T :

$$T = [t_1 \ t_2] = \begin{bmatrix} 1 & 1 \\ -2 & -3 \end{bmatrix}$$

③ 求对角矩阵 Λ

如果将 $n \times n$ 矩阵 A 变成对角线矩阵 Λ ,那么对角线上各元素就是它的特征值。其对角矩阵的一般形式为

$$\Lambda = T^{-1}AT = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \lambda_n \end{bmatrix} \quad (7.15)$$

根据上例

$$A = \begin{bmatrix} 0 & 1 \\ -6 & -5 \end{bmatrix}, \text{ 则 } T = \begin{bmatrix} 1 & 1 \\ -2 & -3 \end{bmatrix}$$

$$T^{-1} = \frac{\text{adj} T}{|T|} = \frac{\begin{bmatrix} -3 & -1 \\ 2 & 1 \end{bmatrix}}{-1} = \begin{bmatrix} 3 & 1 \\ -2 & -1 \end{bmatrix}$$

式中: $\text{adj} T$ ——矩阵 T 的伴随矩阵

$$\therefore \Lambda = T^{-1}AT = \begin{bmatrix} 3 & 1 \\ -2 & -1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -6 & -5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -2 & -3 \end{bmatrix} = \begin{bmatrix} -2 & 0 \\ 0 & -3 \end{bmatrix}$$

因而得矩阵 A 的特征值为 $\lambda_1 = -2, \lambda_2 = -3$ 。

④ λ, T, A 在经济模型中的应用——对产品需要量的预测

【例 7.2】 某种产品有三种规格, 分别为 X_1, X_2, X_3 。现根据订货统计资料进行分析, 结果发现用户对不同规格产品需要的比例如下表(表 7-3)。

表 7-3 订货统计分析

百分比		目 前 订 货		
		X_1	X_2	X_3
上 年 需 要 情 况	X_1	40%	30%	30%
	X_2	60%	30%	10%
	X_3	60%	10%	30%

表中统计资料表示上年度需要各种产品的数量, 于今年订货时更换其他产品的百分比。例如 X_1 产品上一年需要量中, 今年有 40% 仍订购 X_1 产品, 30% 转订 X_2 , 另外 30% 转订 X_3 产品, 余类推, 现要预测第三年度的订货情况和 10 年后的产品订购趋向。

【解】

第一, 将表 7-3 中的统计资料用状态转移矩阵 A 表示, 则得

$$A = \begin{matrix} & \begin{matrix} X_1 & X_2 & X_3 \end{matrix} \\ \begin{matrix} X_1 \\ X_2 \\ X_3 \end{matrix} & \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.6 & 0.3 & 0.1 \\ 0.6 & 0.1 & 0.3 \end{bmatrix} \end{matrix}$$

第二, 求状态转移矩阵 A 的特征值 λ

由特征方程 $[\lambda I - A] = 0$

$$\text{得} \quad \begin{vmatrix} \lambda - 0.4 & -0.3 & -0.3 \\ -0.6 & \lambda - 0.3 & -0.1 \\ -0.6 & -0.1 & \lambda - 0.3 \end{vmatrix} = 0$$

解得特征值为 $\lambda_1 = 1, \lambda_2 = \frac{1}{5}, \lambda_3 = -\frac{1}{5}$

第三, 求其相应的 t 和 T ;

根据上述方法, 可分别求出 $\lambda = 1, \frac{1}{5}$ 及 $-\frac{1}{5}$ 时的特征向量;

$$T_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad T_2 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} \quad T_3 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

$$\therefore T = [T_1 \quad T_2 \quad T_3] = \begin{bmatrix} 1 & 0 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

第四,求 A

对角矩阵可表示为

$$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \lambda_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/5 & 0 \\ 0 & 0 & -1/5 \end{bmatrix}$$

第五,利用上述计算结果进行预测

第三年度订货情况:

$$\therefore \Lambda = T^{-1}AT$$

$$\text{上式两边左乘 } T: \quad T\Lambda = TT^{-1}AT = AT$$

$$\text{然后两边再右乘 } T^{-1}: \quad T\Lambda T^{-1} = ATT^{-1} = A$$

预测第二年度可得

$$A^2 = (T\Lambda T^{-1}) \cdot (T\Lambda T^{-1}) = T\Lambda^2 T^{-1}$$

同理第三年度

$$A^3 = T\Lambda^3 T^{-1} \quad (7.16)$$

$$A^3 = \begin{bmatrix} 1 & 0 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/5 & 0 \\ 0 & 0 & -1/5 \end{bmatrix}^3 \begin{bmatrix} 1 & 0 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 0.496 & 0.252 & 0.252 \\ 0.504 & 0.252 & 0.244 \\ 0.504 & 0.244 & 0.252 \end{bmatrix}$$

以上结果说明,任选一位上年度订购 X_1 产品的用户,则其第三年度订购 X_1 、 X_2 、 X_3 产品的概率分别为 0.496, 0.252, 0.252;选一位上年度订购 X_2 产品的用户,则其第三年度订购 X_1 、 X_2 、 X_3 产品的概率分别为 0.504, 0.252, 0.244,余可类推。10 年后各产品订货趋向为:

$$\begin{aligned} A^{10} = T\Lambda^{10}T^{-1} &= \begin{bmatrix} 1 & 0 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1^{10} & 0 & 0 \\ 0 & \left(\frac{1}{5}\right)^{10} & 0 \\ 0 & 0 & \left(-\frac{1}{5}\right)^{10} \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 0.5 & 0.25 & 0.25 \\ 0.5 & 0.25 & 0.25 \\ 0.5 & 0.25 & 0.25 \end{bmatrix} \quad \text{注: } \left(\frac{1}{5}\right)^{10} \approx 0 \end{aligned}$$

由以上结果可知,10 年以后,订购 X_1 、 X_2 、 X_3 产品的概率为 0.5, 0.25, 0.25, 也即 10 年以后, X_1 产品占有市场的 $\frac{1}{2}$, X_2 、 X_3 产品各占市场的 $\frac{1}{4}$ 。

7.5 结构模型解析法(ISM)

结构模型属宏观模型,结构模型解析法通过有向连接图和相邻矩阵的有关运算,可以得到可达性矩阵;然后再分解可达性矩阵,从而使复杂的系统分解成多级递阶形式。

通过这种结构模型,可以确定组成复杂系统的大量元素之间存在着怎样的关系(这种关系有大小关系、因果关系、上下关系等),然后通过人机对话,明确这种关系,并使复杂关系分解成条理分明的多级递阶结构形式。此种方法在制定复杂的企业计划和决定政策方面,在区域环境规划方面,在城市规划设计等方面得到逐步应用。

7.5.1 有向连接图、相邻矩阵、可达性矩阵

(1) 有向连接图:设有一系统其元素用点(i)表示,元素之间的关系用带有箭头的边($\rightarrow$)表示,即可构成有向连接图(如图 7-12 所示)。

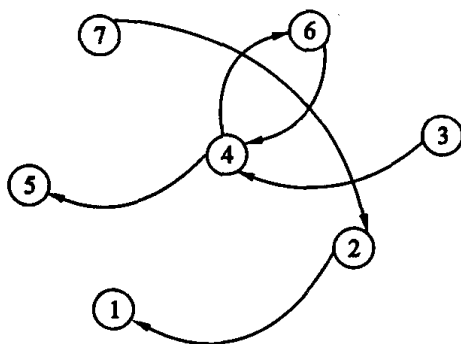


图 7-12 系统有向连接图

(2) 相邻矩阵:用来表示有向连接图中各元素之间连接状态的矩阵叫做相邻矩阵(A),相邻矩阵的元素 a_{ij} 可以定义如下:

$$a_{ij} = \begin{cases} 1 & n_i R n_j \\ 0 & n_i \bar{R} n_j \end{cases} \begin{matrix} R \text{ 表示可以从 } n_i \text{ 到达 } n_j \\ \bar{R} \text{ 表示不能从 } n_i \text{ 到 } n_j \end{matrix}$$

因此图 7-12 所示的相邻矩阵为

$$A = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix} & \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{matrix}$$

(3) 可达性矩阵:可达性矩阵(M)是用矩阵形式来反映有向连接图各节点之间通过一定的路径可以到达的程度。可达矩阵 M 可以用相邻矩阵加上单位矩阵(I) 经过一定运算后求得。即先将 A 加 I , 得一新矩阵 $A_1 = A + I$, A_1 中的元素 a_{ij} 为 1 时,表示从节点 i 到 j 可直接到达。例如图 7-12 中,通过 A 可以求得:

$$A_1 = A + I = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix} & \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix}$$

如上式元素 $a_{21} = 1$ 则说明从节点 2 到 1 有一条直线到达的路径, 但 A_1 还不是可达性矩阵 M , 因此要继续进行运算。将 A_1 平方, 并用布尔代数进行运算, 得

$$(A_1)^2 = (A + I)^2 = A^2 + A + I$$

$$\text{则 } A_2 = (A_1)^2 = A^2 + A + I = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix} & \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix}$$

在矩阵 A_2 中, 如元素为 1, 则表示节点之间可以用多至两条的路径才到达, 依次运算可得:

$$A_1 \neq A_2 \neq A_3 \neq \cdots \neq A_{r-1} = A_r$$

$$\text{也即 } (A + I) \neq (A + I)^2 \neq \cdots \neq (A + I)^{r-1} = (A + I)^r = M$$

此时,

$$(A + I)^{r-1} = (A + I)^r \text{ 时}$$

$$(A + I)^{r-1} = A_{r-1} = M \quad (7.17)$$

称为可达性矩阵, 若该矩阵中元素为 1, 则表示节点间可以用多至 $(r-1)$ 条路径到达。

在以上例子中, 因 $(A + I) \neq (A + I)^2 = (A + I)^3$, 而 $A_1 \neq A_2 = A_3$, 所以 $M = (A + I)^2$,

此时

$$M = [m_{ij}] = (A + I)^2 = A_2$$

结合图 7-12 参看 A_2 矩阵中的各元素可知, 从节点 7 到 1 可以用多至两条的路径到达。

7.5.2 可达性矩阵的分解——结构模型的建立

通过可达性矩阵分解, 可求得系统结构模型, 其分解方法与步骤为: ① 区域分解 (分块对角化), 即把元素分解成几个区域, 不同区域的元素相互之间是没有关系的; ② 级间分解, 即对属于同一区域内的元素进行分级分解; ③ 求解结构模型, 具体步骤如下:

(1) 区域分解

在可达性矩阵中, 可将元素组成可达性集合 $R(n_i)$ (将 n_i 可以到达的元素集合定义为 n_i 的可达集合 $R(n_i)$) 和先行集合 $A(n_i)$ (将 n_i 的先行集定义为能够到达 n_i 的元素的集合 $A(n_i)$), 并定义如下:

$$R(n_i) = \{n_j \in N \mid m_{ij} = 1\} \quad (7.18)$$

$$A(n_i) = \{n_j \in N \mid m_{ji} = 1\}$$

又将共同集合 T 定义如下,

$$T = \{n_i \in N \mid R(n_i) \cap A(n_i) \neq \emptyset\} \quad (7.19)$$

(7.19) 式中, “ $\cap$ ” 表示 $R(n_i)$ 与 $A(n_i)$ 的交, 表示既属 $R(n_i)$ 又属 $A(n_i)$ 的一切组成的集合。

现有属于共同集合 T 的任意两个元素 t_u, t_v , 如果

$$R(t_u) \cap R(t_v) \neq \emptyset \quad (7.20)$$

式中: $\emptyset$ 为空集, 即不包括任何元素的集合, 则元素 t_u 和 t_v 属于同一区域。反之, 如果

$$R(t_u) \cap R(t_v) = \emptyset \quad (7.21)$$

则元素 t_u 和 t_v 属于不同区域。

例如 $R(n_3) \cap R(n_7) = \emptyset$, 所以 n_3 和 n_7 分属两个不同的区域, 经过这样的运算后的集合 N 就叫做区域分解, 可以表示成

$$\prod_1(N) = P_1, P_2, P_3, \dots, P_m \quad (7.22)$$

式中: m ——区域数。

对于上述图 7-12 的可达性矩阵进行区域分解时, 由表 7-4 可知 $T = \{n_1, n_7\}$

$\therefore$

$$R(n_3) \cap R(n_7) = \emptyset$$

$\therefore$

n_3, n_7 属不同区域,

同样, 按以上方法依次将可达性矩阵分解成两个区域, 即

$$\prod = P_1, P_2 = \{n_3, n_4, n_5, n_6\}, \{n_1, n_2, n_7\}$$

据此可将可达性矩阵写成分块对角化的形式:

$$M = \begin{array}{c|cccc|ccc} & 3 & 4 & 5 & 6 & 1 & 2 & 7 \\ \hline 3 & 1 & 1 & 1 & 1 & & & \\ 4 & 0 & 1 & 1 & 1 & & 0 & \\ 5 & 0 & 0 & 1 & 0 & & & \\ 6 & 0 & 1 & 1 & 1 & & & \\ \hline 1 & & & & & 1 & 0 & 0 \\ 2 & & 0 & & & 1 & 1 & 0 \\ 7 & & & & & 1 & 1 & 1 \end{array} \quad (7.23)$$

表 7-4 可达性集合、先行集合和共同集合

i	$R(n_i)$	$A(n_i)$	$R(n_i) \cap A(n_i)$
1	1	1, 2, 7	1
2	1, 2	2, 7	2
3	3, 4, 5, 6	3	3
4	4, 5, 6	3, 4, 6	4, 6
5	5	3, 4, 5, 6	5
6	4, 5, 6	3, 4, 6	4, 6
7	1, 2, 7	7	7

(2) 级间分解

级间分解在每一区间内进行,一般在系统不是十分复杂的情况下,可以采用简易的方法,即直接从可达性矩阵中,找出矩阵元素全部为 1 的某一列,于是将该列和其相应的行抽出,作为第一级;然后缩减了的新矩阵,再在新矩阵中重新找出矩阵元素全部为 1 的新的一列,重复进行上述运算,直到分解完毕为止。

上例按简易分解过程如图 7-13 所示。

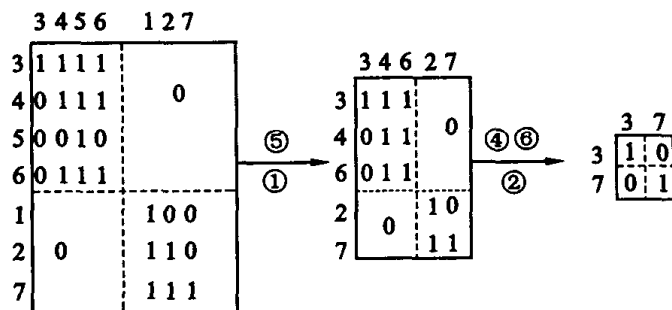


图 7-13 级间简易分解图

如果按区域分解,则同上例有(步骤从略)

$$\prod_2(P_1) = \{n_5\}, \{n_4, n_6\}, \{n_3\}$$

$$\prod_2(P_2) = \{n_1\}, \{n_2\}, \{n_7\}$$

然后将可达性矩阵按级变换可得

$$M = \begin{array}{c|cccc|ccc} & 5 & 4 & 6 & 3 & 1 & 2 & 7 \\ \hline 5 & 1 & 0 & 0 & 0 & & & \\ 4 & 1 & 1 & 1 & 0 & & 0 & \\ 6 & 1 & 1 & 1 & 0 & & & \\ 3 & 1 & 1 & 1 & 1 & & & \\ \hline 1 & & & & & 1 & 0 & 0 \\ 2 & & 0 & & & 1 & 1 & 0 \\ 7 & & & & & 1 & 1 & 1 \end{array} \quad (7.24)$$

从 M 矩阵中可以看到, $\{n_4, n_6\}$ 的相应行和列的矩阵元素完全一样,因此,可以把两者当作一个系统元素看待,从而可削减相应的行和列,得到新的可达矩阵 M' 。 M' 叫做缩减矩阵。上例的缩减矩阵 M' 为:

$$M' = \begin{array}{c|ccc|ccc} & 5 & 4 & 3 & 1 & 2 & 7 \\ \hline 5 & 1 & 0 & 0 & & & \\ 4 & 1 & 1 & 0 & 0 & & \\ 3 & 1 & 1 & 1 & & & \\ \hline 1 & & & & 1 & 0 & 0 \\ 2 & & 0 & & 1 & 1 & 0 \\ 7 & & & & 1 & 1 & 1 \end{array} \quad (7.25)$$

(3) 求解结构模型

求解结构模型,即建立反映系统多级递阶结构的矩阵,此矩阵为结构矩阵(A')。结构矩阵可以从缩减后的可达性矩阵(M')通过一系列的计算求得。其中简化的计算方法是从缩减矩阵 M' 中减去单位矩阵 I ,得到一个新的矩阵 M'' ,再对 M'' 进行分析,从而找出结构矩阵 A' 。

对于上例

$$M'' = M' - I = \begin{array}{c} \begin{array}{ccc|ccc} & 5 & 4 & 3 & 1 & 2 & 7 \\ \hline 5 & 0 & 0 & 0 & & & \\ 4 & 1 & 0 & 0 & & 0 & \\ 3 & 1 & 1 & 0 & & & \\ \hline 1 & & & & 0 & 0 & 0 \\ 2 & & 0 & & 1 & 0 & 0 \\ 7 & & & & 1 & 1 & 0 \end{array} \end{array} \quad (7.26)$$

从矩阵 M'' 中,先找出系统元素的第一级和第二级之间的关系,其中 $m''_{45} = 1$,说明节点 n_4, n_5 间有 $n_4 \rightarrow n_5$ 的关系,然后除去 n_5 的行和列,再找第二级与第三级元素之间的关系,从 M'' 中可得 $m''_{34} = 1$,即在节点 n_3, n_4 中有 $n_3 \rightarrow n_4$ 的关系。

同理在另一区域内 $m''_{21} = 1$,即 $n_2 \rightarrow n_1$; $m''_{72} = 1$ (除去 n_1 的行和列)即 $n_7 \rightarrow n_2$ 。

最后,以上述 $m''_{45} = 1, m''_{34} = 1, m''_{21} = 1, m''_{72} = 1$ 作为结构矩阵的元素,从而可得结构矩阵 A'

$$A' = \begin{array}{c} \begin{array}{ccc|ccc} & 5 & 4 & 3 & 1 & 2 & 7 \\ \hline 5 & 0 & 0 & 0 & & & \\ 4 & 1 & 0 & 0 & & 0 & \\ 3 & 0 & 1 & 0 & & & \\ \hline 1 & & & & 0 & 0 & 0 \\ 2 & & 0 & & 1 & 0 & 0 \\ 7 & & & & 0 & 1 & 0 \end{array} \end{array} \quad (7.27)$$

根据结构矩阵 A ,或者根据图 7-13 的级间分解图,可以绘制系统的多级递阶结构图(图 7-14)。

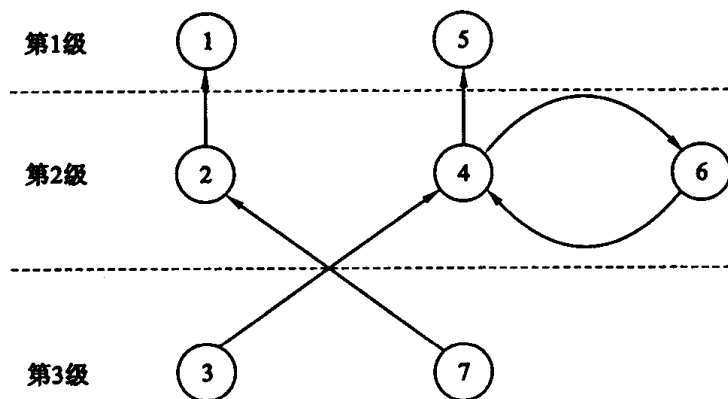


图 7-14 系统多级递阶结构图

7.6 小样本问题的建模方法

系统分析时采用的建模方法,大多数是借用传统统计学和工程系统辨识的方法。在那里,可用以建模的数据是比较多的,所以辨识方法主要的问题是—致性。而只要有充分的数据,这种辨识方法就可最后找到一个真正反映实际系统的模型。至于工程系统的动态辨识,它可以用比较精确的测试工具,把实际系统的数据测试出来,同时还可以找出测试工具本身误差的统计特征。

但对于系统工程的对象,特别是对于社会经济系统,对数据误差的规律是不清楚的。因此,寻求一种能够在小样本数据下,克服较大观测误差的参数估计方法,就是解决社会经济系统建模的重要问题,属于动态建模范畴。本节将扼要介绍 LKL 方法。

7.6.1 对最小二乘估计的分析

设有下列离散时间的动态系统

$$\left. \begin{aligned} A(Z^{-1})X(k) &= B(Z^{-1})u(k) + W(k) \\ y(k) &= X(k) + V(k) \end{aligned} \right\} \quad (7.28)$$

式中 $X(k)$ —— k 时刻的内生变量;

$u(k)$ —— k 时刻的外生变量;

$y(k)$ —— k 时刻的观测值;

$W(k), V(k)$ ——分别为动态系统和观测的正态、独立平稳分布零均值白噪声序列。它们的方差分别为 σ_w^2, σ_v^2 。

$$A(Z^{-1}) = 1 + a_1 Z^{-1} + \cdots + a_n Z^{-n}$$

$$B(Z^{-1}) = b_1 Z^{-1} + \cdots + b_m Z^{-m}$$

令未知参数为

$$\theta = [-a_1, -a_2, \cdots, -a_n \quad b_1, \cdots, b_m]^T$$

建模的问题要从观测到的 N 集数

$$\{y(i), y(i-1), \cdots, y(i-n), u(i-1), \cdots, u(i-m)\} \quad i = 1, 2, \cdots, n$$

中来估计未知参数 θ 。

合并(7.28)式

$$A(Z^{-1})y(k) = B(Z^{-1})u(k) + \zeta(k) \quad (7.29)$$

式中

$$\zeta(k) = A(Z^{-1})V(k) + W(k) \quad (7.30)$$

当 $V(k) = 0$ 时, $\zeta(k) = W(k)$, 只有在这种特殊情况下,最小二乘法估计(LS)才能得到最好的结果。

当 $V(k) \neq 0$ 时,则 $\zeta(k)$ 是一个 n 阶相关的噪声,在这种条件下,如果再用 LS 估计,就会由于下述原因而使估计结果的精度大大降低。

① $V(k) \neq 0$, (7.29) 式可写成

$$y(k) = - \sum_{i=1}^n a_i y(k-i) + \sum_{i=1}^m b_i u(k-i) \zeta(k)$$

这时观测变量 $y(k-i)$ 与 $\zeta(k)$ ($i=1,2,\dots,n$) 是相关的, 因此严重影响了 LS 估计精度。

② $V(k) \neq 0$, 由下式可知其总噪声大大增加

$$D[\zeta(k)] = D\left[\sum_{i=1}^n a_i V(k-i) + W(k)\right] = \sigma_v^2 \sum_{i=1}^n a_i^2 + \sigma_w^2$$

式中 D ——取方差的运算子。

总噪声的增加量 $\sigma_v^2 \sum_{i=1}^n a_i^2$, 原式中阶次愈高, 总噪声愈大, 精度也就愈低。

为此, 研究一种能够解除相关性, 同时又能降低总噪声水平的新估计方法是非常必要的。

如果引入观测数据的条件平均 $E\{V(k) | Y(k)\}$, 其中 $Y(N) = \{y(1), y(2), \dots, y(n)\}$ 是观测数据的样本, 则可用 $V(k) - E\{V(k) | Y(k)\}$ 代替 $V(k)$ 做估计, 得到较好的结果。从 (7.28) 式, 并考虑 (7.29) 式可知

$$\begin{aligned} A(Z^{-1})[y(k) - E\{V(k) | Y(k)\}] \\ = B(Z^{-1})u(k) + A(Z^{-1})[V(k) - \\ E\{V(k) | Y(N)\}] + W(k) \end{aligned}$$

由于,

$$\begin{aligned} E\{X(k) | Y(N)\} &= E\{y(k) - V(k) | Y(N)\} \\ &= y(k) - E\{V(k) | Y(N)\} \end{aligned}$$

$\therefore$

$$A(Z^{-1})\hat{X}(k) = B(Z^{-1})u(k) + A(Z^{-1})\hat{V}(k) + W(k) \quad (7.31)$$

式中

$$\begin{aligned} \hat{X}(k) &= E\{X(k) | Y(N)\} \\ \hat{V}(k) &= V(k) + E\{V(k) | Y(N)\} \end{aligned} \quad (7.32)$$

(7.29) 式、(7.30) 式、(7.31) 式之差, 只在于用内生变量 $X(k)$ 的最佳估计 $\hat{X}(k)$ 代替 $X(k)$, 用 $\hat{V}(k)$ 代替观测噪声 $V(k)$ 。如果能确实求到 $\hat{X}(k)$, 则 (7.31) 式就可用一般最小二乘法估计 LS 得到较好的估计值。用上述代置可以解除相关性, 减少噪声水平。

为了求 $\hat{X}(k)$ 可以先用最小二乘法估计得到一个 $\hat{\theta}_1$, 再用卡尔曼平滑求 $\hat{X}(k)$, 以 $\hat{X}(k)$ 代替 $y(k)$, 再用 LS 法即可求得一个较好的估计 $\hat{\theta}_2$, 这样用三步求到一个较好估计的方法即称为 LKL 估计法。

7.6.2 LKL 估计

LKL 估计的步骤见图 7-15。要做卡尔曼平滑, 先要把 (7.28) 式离散时间方程化为状态空间模型 (这是控制理论中的实现问题), 再选系统的噪声方差 Q 和观测的噪声方差 R , 最后确定初值 $\hat{X}_0$ 和 $P(0)$ 。

(1) LKL 状态空间模型的实现

如取 $\bar{u}(k) = B(Z^{-1})u(k+1)$

则 (7.28) 式可以写成

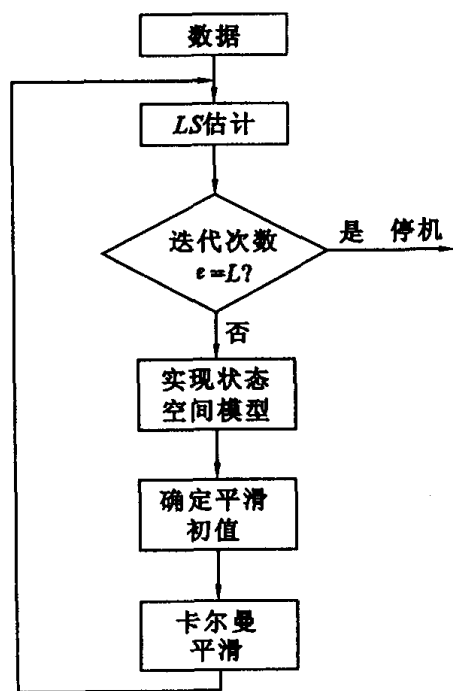


图 7-15 LKL 估计计算步骤框图

$$A(Z^{-1})X(k) = \bar{u}(k-1) + W(k)$$

令

$$x_1(k) = x(k)$$

$$x_2(k) = x(k-1)$$

$$\vdots$$

$$x_n(k) = x(k-n+1)$$

则(7.28)式可以化成

$$x(k+1) = Ax(k) + B\bar{u}(k) + \Gamma W(k+1)$$

$$y(k) = Hx(k) + V(k)$$

(7.33)

式中 $x(k) = [x_1(k), x_2(k), \dots, x_n(k)]^T$

$$B = [1, 0, \dots]^T$$

$$\Gamma = [1, 0, \dots]^T$$

$$H = [1, 0, \dots]$$

$$A = \begin{bmatrix} -a_1 & -a_2 & \cdots & -a_n \\ 1 & 0 & \cdots & 0 \\ \ddots & & \ddots & \\ 0 & 0 & 1 \cdots & 0 \end{bmatrix}$$

(2) 确定卡尔曼平滑中的系统噪声方差 Q , 观测噪声方差 R 和初值

通过仿真实验可知, 取 Q 太大或太小, 都将使 LKL 估计失去调节能力而趋于 LS 估计。为得到好的结果, 可用以下经验公式求取:

$$Q = s^2 / (N - m - n) + \sigma_w^2 \quad (7.34)$$

式中 s^2 ——LS 估计的残差平方和。

若 σ_w^2 难以确定, 则可取 $Q = s^2 / 3(N - m - n)$ 。

观测噪声方差 R 隐含着对观测数据可信度的度量, 如果已知 σ_w^2 , 则 $R = \sigma_w^2$, 但对大多数社会经济系统, σ_w^2 是不知的。但实验证明, LKL 估计对 R 并不敏感, 在 σ_w^2 未知时, 通过对观测数据有信任程度, 可以粗略的给出 R 。一般可取观测数据的 10% 作为 R , 即

$$R = \frac{1}{N} \sum_{k=1}^N [x_0 | x(k)]^2 \quad (7.35)$$

因

$$x_0 = [x_1(0), x_2(0), \dots, x_n(0)]^T$$

$$= [x(0), x(-1), \dots, x(-n+1)]^T$$

故

$$x_0 = [y(0), y(-1), \dots, y(-n+1)]^T$$

通过仿真发现, $P(0)$ 较小时, 估计精度较高, 有时为简单起见, 就取 $P(0) = 0$ 。

7.6.3 LKL 的八个步骤

参阅图 7-15。

第一步 组织原始数据

用 N 个样本

$$\{y(i), y(i-1), \dots, y(i-n), u(i-1), \\ u(i-2), \dots, u(i-m)\} \quad i = 1, 2, \dots, N$$

构成下列向量 Y 和矩阵 Φ

$$Y = [y(1), y(2), \dots, y(N)]^T$$

$$\Phi = \begin{bmatrix} y(0) & y(-1) & \cdots & y(-n+1) & u(0) & u(-1) & \cdots & u(-m+1) \\ y(1) & y(0) & \cdots & y(-n+2) & u(1) & u(0) & \cdots & u(-m+2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ y(N-1) & y(N-2) & \cdots & y(N-n) & u(N-1) & u(N-2) & \cdots & u(N-m) \end{bmatrix}$$

其中将图 7-15 中的 ϵ 设定为 1。

第二步 最小二乘法估计(LS)

参数 $\alpha = [-a_1, -a_2, \dots, -a_n, b_1, b_2, \dots, b_m]^T$ 的最小二乘法估计为

$$\hat{\theta} = (\Phi^T \Phi)^{-1} \Phi^T Y \quad (7.31)$$

这时再检查 ϵ 值, 如 $\epsilon = 3$ (迭代二次以上即可收敛到最后满意的结果) 则停止运算, $\hat{\theta}$ 即为 LKL 估计。否则进至第三步:

第三步 状态空间实现

将系统(7.23) 改成如下状态空间形式

$$x(k+1) = Ax(k) + B\bar{u}(k) + \Gamma W(k+1)$$

$$y(k) = Hx(k) + V(k)$$

式中

$$x(k) = [x_1(k), x_2(k), \dots, x_n(k)]^T$$

$$= [y(k), y(k-1), \dots, y(k-n+1)]$$

$$\bar{u}(k) = \hat{b}_1 u(k) + \hat{b}_2 u(k-1) + \cdots + \hat{b}_m u(k-m+1)$$

$$A = \begin{bmatrix} -\hat{a}_1 & -\hat{a}_2 & \cdots & -\hat{a}_n \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 1 & \cdots & 0 \end{bmatrix}$$

$$B = [1, 0, \dots]^T$$

$$\Gamma = [1, 0, \dots]^T$$

$$H = [1, 0, \dots]$$

第四步 求方差 Q 和 R

先查 ϵ , 如 $\epsilon > 1$ 则进至第六步, 否则作下列计算: 求残差平方和

$$s^2 = (y - \Phi \hat{\theta})^T (y - \Phi \hat{\theta})$$

再按是否已知 σ_w^2 , 用式(7.34) 求 Q

按(7.35) 式求 R

第五步 初值设定

令状态滤波初值为

$$\hat{x}_0 = [y(0), y(-1), \dots, y(-n+1)]^T$$

$$P(0) = [0]_{n \times n}$$

第六步 卡尔曼滤波

用已知的 $Q, \hat{x}_0, R, P(0)$ 按以下递推计算:

$$\begin{aligned}
\hat{x}(k | k-1) &= A\hat{x}(k-1) + B\bar{u}(k-1) \\
P(k | k-1) &= AP(k-1)A^T + \Gamma Q \Gamma^T \\
K(k) &= P(k | k-1)H^T \{HP(k | k-1)H^T + R\}^{-1} \\
\hat{x}(k) &= \hat{x}(k | k-1) + K(k)[y(k) - H\hat{x}(k | k-1)] \\
P(k) &= [I - K(k)H]P(k | k-1) \\
k &= 1, 2, \dots, N
\end{aligned}$$

第七步 卡尔曼平滑

用卡尔曼滤波结果,通过以下递推计算,可得状态平滑值

$$\begin{aligned}
\hat{x}(k | N) &= \hat{x}(k) + W(k)[\hat{x}(k+1 | N) - A\hat{x}(k) - B\bar{u}(k)] \\
W(k) &= P(k)A^T P^{-1}(k+1 | k) \\
k &= N-1, N-2, \dots, n
\end{aligned}$$

这样就求到状态平滑 $\hat{x}(n | N), \hat{x}(n+1 | N) \dots \hat{x}(N | N)$, 取

$$\hat{x}(k | N) = [\hat{x}(k), \hat{x}(k-1), \dots, \hat{x}(1)]^T$$

取 $\hat{x}(k | N)$ 中后面 $n-1$ 元素, 可得 $\hat{x}(k), \hat{x}(k), \dots, \hat{x}(n-1)$ 这样就得到每一个 k 时刻的内生变量估计值 $\hat{x}(k)$ 。

第八步 组织修正数据

用第七步中得到的数据改写第一步中的向量 Y 和矩阵 ϕ

$$\begin{aligned}
Y &= [\hat{X}(1) \ \hat{X}(2) \ \dots \ \hat{X}(N)]^T \\
\phi &= \begin{bmatrix} \hat{X}(0) & \hat{X}(-1) & \dots & \hat{X}(-n+1) & u(0) & \hat{u}(-1) & \dots & u(-m+1) \\ \hat{X}(1) & \hat{X}(0) & \dots & \hat{X}(-n+2) & u(1) & u(0) & \dots & u(-m+2) \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \hat{X}(N-1) & \hat{X}(N-2) & \dots & \hat{X}(N-n) & u(N-1) & u(N-2) & \dots & u(N-m) \end{bmatrix}
\end{aligned}$$

再回到第二步,同时使迭代次数 $\epsilon = \epsilon + 1$ 。

以上八步即可实现 LKL 估计。

习 题

7.1 什么叫系统模型?有哪些类型?如何进行系统分析?

7.2 已知系统可达矩阵为:

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

利用可达性集合 $R(n_j)$ 和先行集合 $A(n_j)$ 进行区域划分和级间划分,建立结构模型。

7.3 试举例采用 LKL 法进行系统建模和仿真。分析样本误差。

8 网络计划法

本章提要

网络计划法是从图与网络基础上发展出的现代科学管理中的一种方法,网络计划是运用网络的基本概念,对工程总体计划和调度进行统筹安排,所以又称网络统筹法。

在这一章中,介绍了关键路线法 CPM(Critical Path Method)、计划评审技术 PERT(Program Evaluation Review Technique)和图解评审技术 GERT(Graph Evaluation Review Technique)。

网络计划法是现代管理科学中一种十分重要的方法。它在经济建设、科学研究、企业管理中显示出越来越大的作用。由于技术的进步、生产的发展,工艺分工细密,作业工程规模不断扩大,指导管理生产的任务繁重,因而要从整体上运用科学的管理方法,最大限度地安排好人力和物力资源,要求决策者要事先做好调查,充分掌握资料,用科学的方法,制定周密计划。其中网络计划法就是一种非常实用有效的方法。

网络计划法又称网络统筹法。该方法中最具代表性的是关键路线法(CPM)和计划评审法(PERT)。CPM法开始于1957年,美国杜邦公司在制定化学工厂建设计划和实施计划中,采用CPM法,大大提高了建设的合理性。而且它在大工程停工大修中,缩短工时37.6%,所以杜邦公司在实施CPM法的第一年,就节约工程费用100万美元。PERT法也始于20世纪50年代末,美国国防部和美国海军特种计划局,在研制导弹和核潜艇计划中,应用该方法,协调近万个工厂企业,大大缩短了研制生产周期,获得了满意的结果。

关键路线法使用网络图反映某项工程(任务)各道工序所需时间以及他们的衔接关系,通过计算各工序有关时间参数,和完成工程(任务)所需的最少时间,从而确定关键工序和关键路线,并在此基础上通过网络分析方法制定出时间、成本和资源优化的网络计划方案。该方法主要应用于有以往类似项目经验的工程上。而计划评审法也同样应用了网络计划与网络分析的方法,但着重于对工程(任务)安排的评价与审查,主要应用于对研究与开发的新项目上。

8.1 网络图

网络图因其形状如网络而得名。它表示一个计划中各道工序(或称计划项目、作业项目、活动、工作等,以后均泛指工序)前后上下的衔接关系和所需时间的图解模型。这种图形是根据各工序间的相互依存的逻辑关系,从总体出发,统筹安排人力资源和设备,绘制而成。所以又称程序网络图,简称网络图。

8.1.1 网络图的组成

由带有箭头,标明走向的有向边组成的网络图,称箭线式网络图。它是由事项(节点)、活动(箭线)、线路组成的。

(1) 事项(节点)

在网络图中,事项用圆圈表示,它是两条或两条以上箭线的交接点(整个网络的起始点和终点除外)故称节点,表示前接工序的完工事项和后续工序的开发事项,也是相邻工序在时间上的分界点。

(2) 活动(箭线)

活动是指每道工序的具体内容,以一条有向边表示,故又称箭线,一般在箭线下面标出工序名称,工序上面标出工序时间,即一道工序所需的时间。

(3) 线路

从网络图的起始事项,顺着箭线方向连续不断地到达终点事项的一条通道称作线路。如图 8-1 所示的网络图,从始点①到终点⑥共有 8 条线路。

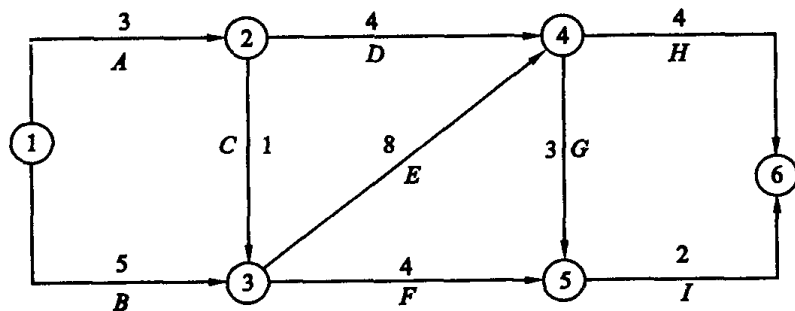
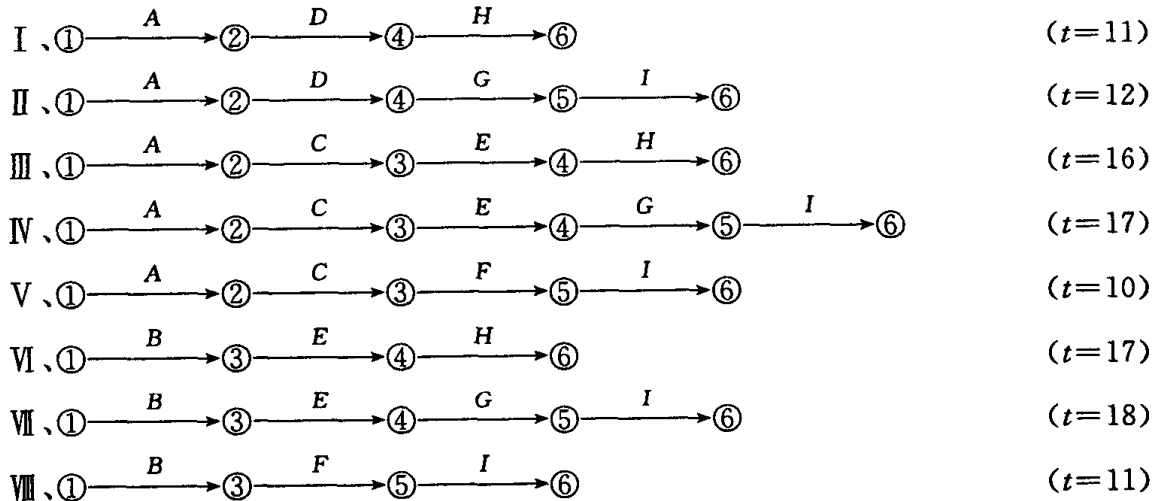


图 8-1 网络图组成

网络图中所需时间最长的线路称为关键线路。如①→③→④→⑤→⑥所需时间最长。如果箭线上的数字表示天数,则共需 18 天。故在网络图中该线路为关键线路。若最长的线路完工了,那么其他线路肯定可完工。所以关键线路代表了工程周期。合理地制定工程周期是企业管理的主要目标之一。

8.1.2 网络图作图规则

绘制网络图所必须遵循的原则为：

(1) 工序的箭线由左向右,相邻两节点间只能有一条箭线相连,节点按时序编号,序号按箭头方向由小到大,出发始点为1。

(2) 网络中不能有缺口或回路,除始点和终点以外,任何节点都应有箭线相连,不能中断。

(3) 网络图中只能有一个始点,一个终点。它们表示工程的开始和结束。

(4) 如果某工序开始或结束同时有两个或两个以上工序,可采用虚工序,增加节点就可按规则作图。所谓虚工序,主要是为了解决网络中逻辑关系的需要,只是起连接作用,它既不消耗资源,又不占用时间,所以称为虚工序。如图8-2所示,图中(a)为错误画法,(b)为正确画法。

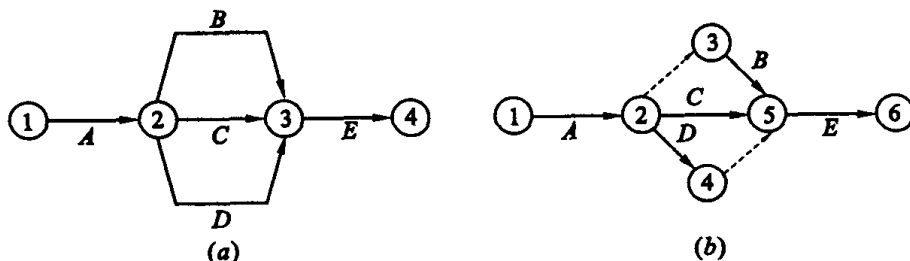


图8-2 工序图

(5) 对绘制的网络图要进行整理,去掉多余的虚工序,同时尽量避免出现交叉箭线;最后要按工序前后和工艺流程以及组织管理的需要重新调整图中节点的序号。

8.1.3 绘制网络图的方法

在绘制一项工程的网络图时,首先要将整个工程分解成若干工序并确定工序的长度(即每工序所需时间),然后根据工序流程确定各工序之间的顺序关系。绘制中还必须注意:

(1) 设计图纸一般有总图和分图等多个层次,因此,对于规模较大的工程,在编网络图时,同样也要分层次、区分网络总图和分图。

(2) 工序是工艺上和生产管理组织上具有相对独立的活动。完成工序需要一定的劳动力、原材料、能源、设备和时间。所以划分工序时要综合各种因素予以考虑,工序之间的顺序关系应根据工艺流程来确定。

(3) 有时为了缩短工期,在条件允许时,有必要交叉作业,对作业时间较长的某些工序,分解成若干小工序以便在某一小工序完工后即可转入紧后工序,形成交叉作业。

8.2 计算网络图的时间参数

8.2.1 作业时间

各工序的作业时间,一般应该是一个确定值,应根据定额或以往的统计资料加以确定,但如果本工程或本工序是一个新开发的工作,没有以往的经验 and 资料可借鉴,则可采用三点

估计法进行计算。所谓三点估计法就是将时间分为三种情况：

- ① 乐观时间：工作最顺利的作业时间，用 a 表示；
- ② 保守时间：工作最不顺利的作业时间，用 b 表示。
- ③ 最可能的时间：是凭经验估计的，在大多数情况下，工作均可在此时间内完成，用 m 表示。

将以上三种情况的数值按概率分布加权。可以设 a 、 b 各占该工序的 $1/6$ 的时间， m 占 $4/6$ ；然后用下式进行平均；即可求出一个该工序较为接近实际的确定值：

$$T_E = \frac{a + 4m + b}{6} \quad (8.1)$$

以后在网络图的各工序中标出的时间，一般都是通过概率估算出的确定值。

如果工程中各工序的时间相互独立，且方差正态分布，则工程完成的时间等于各关键工序时间之和，即工程完工时间及方差为：

$$T_E = \sum_i \frac{a_i + 4m_i + b_i}{6} \quad (8.2)$$

$$\sigma_E^2 = \sum_i \left(\frac{b_i - a_i}{6} \right)^2 \quad (8.3)$$

8.2.2 节点最早开始时间

某任一节点 j 可能最早开工时间简称节点最早时间，记为 $T_E(j)$ 表示从工程开始施工到节点 j 最长路线的各工序之和。所以从始点开始，从左至右逐个计算即可得到所有节点的最早时间。如果同时有几条箭线指向节点 j 时，则取这些箭线的箭尾事项各节点最早时间与工序时间之和的最大值，即：

$$\begin{aligned} T_E(1) &= 0 \\ T_E(j) &= \max\{T_E(i) + t_{ij}\} \quad (j = 2, 3, \dots, n) \end{aligned} \quad (8.4)$$

现将图 8-1 中的各节点的最早时间计算如下：

$$T_E(1) = 0$$

$$T_E(2) = 3$$

$$T_E(3) = \max\{T_E(2) + t_{23}, T_E(1) + t_{13}\} = \max\{4, 5\} = 5$$

$$T_E(4) = \max\{T_E(2) + t_{24}, T_E(3) + t_{34}\} = \max\{7, 13\} = 13$$

$$T_E(5) = \max\{T_E(3) + t_{35}, T_E(4) + t_{45}\} = \max\{9, 16\} = 16$$

$$T_E(6) = \max\{T_E(4) + t_{46}, T_E(5) + t_{56}\} = \max\{17, 18\} = 18$$

其中 $T_E(6) = 18$ ，就是完成工程最少所需时间，即工程总工期。

8.2.3 节点最迟完成时间 $T_L(j)$

某节点 j 的最迟时间 $T_L(j)$ 表示在不误计划工程总工期的前提下，某节点 j 最迟必须完成的时间。显然， $T_L(j)$ 等于计划总工期 T_E 减去节点 j 至终点 k 的最长时间间隔，所以从终点开始，自右向左逐个推算即可得到所有节点的最迟时间，如果节点 j 同是几条箭线的箭尾事项，则节点 j 的最迟时间应是这些箭线的箭头事项各节点最迟完工时间与工序时间之差的最小值，即

$$\begin{aligned}
 T_L(n) &= T_E(n) \\
 T_L(j) &= \min_k \{T_L(k) - t_{jk}\} \quad (j = n-1, n-2, \dots, 1)
 \end{aligned}
 \tag{8.5}$$

现将图 8-1 网络图中各节点的最迟时间计算如下：

$$\begin{aligned}
 T_L(6) &= T_E(6) = 18 \\
 T_L(5) &= T_L(6) - t_{56} = 18 - 2 = 16 \\
 T_L(4) &= \min\{T_L(5) - t_{45}, T_L(6) - t_{46}\} = \min\{13, 14\} = 13 \\
 T_L(3) &= \min\{T_L(4) - t_{34}, T_L(5) - t_{35}\} = \min\{5, 12\} = 5 \\
 T_L(2) &= \min\{T_L(3) - t_{23}, T_L(4) - t_{24}\} = \min\{4, 9\} = 4 \\
 T_L(1) &= \min\{T_L(2) - t_{12}, T_L(3) - t_{13}\} = \min\{1, 0\} = 0
 \end{aligned}$$

8.2.4 确定关键线路

利用节点的最早时间和最迟时间，可以计算相应工序的几个时间参数。

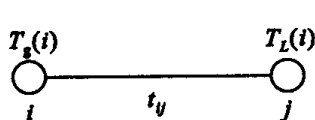


图 8-3

工序 (ij) 的总时差：

$$R(ij) = T_L(j) - T_E(i) - t_{ij}$$

表示在不误工程总工期的前提下，工序 (ij) 的开工时间有多少机动时间。

若 $T_L(j) = T_E(j)$ ，则 j 节点称为关键节点。

若工序 (ij) 的总时差 $R(ij) = T_L(j) - T_E(i) - t_{ij} = 0$ ，则工序 (ij) 称为关键工序。

从网络图的始点 1 到终点 n ，全部由关键工序组成的最长路称为该网络图的关键线路。

以图 8-1 网络图为例求取关键线路，其图算法步骤如下：

第一，图 8-1 中，标注相关的时间参数如图 8-4 所示，首先从始点开始，按计算节点最早时间 $T_E(j)$ 的公式，自左向右分别计算各节点的最早时间并将计算结果记入图中各节点的左下方方框“□”中。（用(8.4)式已计算出结果）

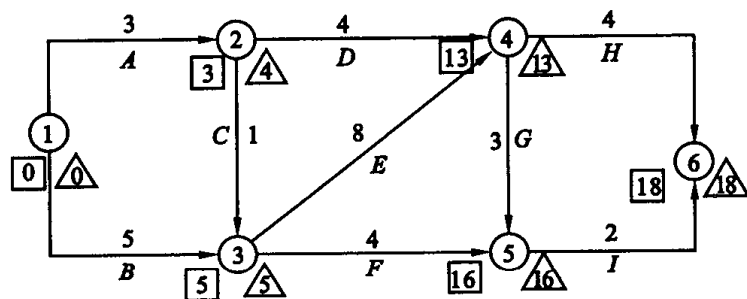


图 8-4

第二，从网络图的终点开始，按计算节点的最迟时间 $T_L(j)$ 的公式，自右向左，分别计算各节点最迟时间，并记入各节点右下方的三角框“△”中。（式(8.5)已计算出结果）

第三，计算 $T_L(j) - T_E(j)$ ， $(j = 1, 2, \dots, n)$ 其结果为 0 的各节点，即为关键节点（事项），由这些节点所确定的工序，即为关键工序，由这些关键工序组成的线路，即为关键线路。

在图 8-1 和图 8-4 网络图中，①，③，④，⑤，⑥ 为关键节点（事项），① $\xrightarrow{B}$ ③，③ $\xrightarrow{E}$ ④，④ $\xrightarrow{G}$ ⑤，⑤ $\xrightarrow{I}$ ⑥ 为关键工序，而由 B—E—G—I 依次组成了关键线路。

8.3 作业的时间参数与计算

8.3.1 作业的最早开始时间 $t_{ES}(i, j)$

一个作业必须等它前面的作业完工以后才能开始,在这之前是不具备开工条件的。这个时间就叫做作业的最早开始时间。其意义是该作业最早是什么时候可以开始,用 $t_{ES}(i, j)$ 表示。

它的计算方法有两种:一是通过它的先行作业的最早开始时间加上作业时间来进行;二是通过节点的最早开始时间来进行。

用第一种办法,其公式为

$$t_{ES}(i, j) = \max_{k < i < j} [t_{ES}(k, i) + t(k, i)] \quad (8.6)$$

式中: $t_{ES}(i, j)$ 为 $i \rightarrow j$ 作业的现行作业的最早开始时间。从左至右直到终点为止。

用第二种办法,其公式为

$$t_{ES}(i, j) = t_E(i) \quad (8.7)$$

即作业的最早开始时间等于箭尾节点的最早开始时间。

8.3.2 作业的最早完成时间 $t_{EF}(i, j)$

一个作业的最早完成时间,就是它的最早开始时间加上本作业所需的时间。其意义是指该作业最早什么时间可以完成,以 $t_{EF}(i, j)$ 表示。它的计算也有两种办法。

第一种办法,其公式是

$$t_{EF}(i, j) = t_{ES}(i, j) + t(i, j) \quad (8.8)$$

第二种办法,其公式是

$$t_{EF}(i, j) = t_E(i) + t(i, j) \quad (8.9)$$

即作业的最早完成时间等于箭尾所触节点的最早开始时间加上作业所需时间。

8.3.3 作业的最迟开始时间 $t_{LS}(i, j)$

一个作业,紧接其后有一个或几个作业,为了不影响后续作业的如期开始,每个作业应有一个最迟必须开始时间,这个时间就叫做作业的最迟开始时间,其意义是该作业最迟应该什么时间开始,以 $t_{LS}(i, j)$ 表示,它的计算也有两种办法。

第一种办法,其公式是

$$t_{LS}(i, j) = \min_{i < j < k} [t_{LS}(j, k) - t(i, j)] \quad (8.10)$$

式中: $t_{LS}(j, k)$ 是 $i \rightarrow j$ 作业的后续作业的最迟开始时间。计算是从右向左直到始点为止。

第二种办法,其公式是

$$t_{LS}(i, j) = t_L(j) - t(i, j) \quad (8.11)$$

即作业的最迟开始时间等于箭头所触节点的最迟完成时间减去作业时间。

8.3.4 作业的最迟完成时间 $t_{LF}(i, j)$

一个作业的最迟完成时间等于它的最迟开始时间加上本作业所需的时间。其意义就是

该作业最迟应该什么时候完成,以 $t_{LF}(i, j)$ 表示。它的计算也有两种办法。

第一种办法,其公式为

$$t_{LF}(i, j) = t_{LS}(i, j) + t(i, j) \quad (8.12)$$

第二种办法,其公式为

$$t_{LF}(i, j) = t_L(j) \quad (8.13)$$

即作业的最迟完成时间等于箭头节点的最迟完成时间。

8.3.5 作业的总时间差 $R(i, j)$

任取一个作业 (i, j) 来分析。如果它在最早开始时间 $t_{ES}(i, j)$ 开始,并且耗费规定工时 $t(i, j)$,则它一定能在最早完成时间 $t_{EF}(i, j)$ 完工;作业 (i, j) 又有一个最迟完成时间 $t_{LF}(i, j)$,它只要不超过 $t_{LF}(i, j)$ 而完工,就不会拖延整个任务的工期。所以,作业的安排具有一定的回旋余地,其回旋范围称为作业 (i, j) 的“总时差”,计算公式如下

$$R(i, j) = t_{LF}(i, j) - t_{EF}(i, j) \quad (8.14)$$

将式(8.8)及式(8.12)代入式(8.14),得到

$$R(i, j) = t_{LS}(i, j) - t_{ES}(i, j) \quad (8.15)$$

将式(8.13)及式(8.9)代入式(8.14),得到

$$R(i, j) = t_L(j) - t_E(i) - t(i, j) \quad (8.16)$$

当 $R(i, j) = 0$ 时,称作业 (i, j) 为“关键作业”。

当 $R(i, j) \neq 0$ 时,作业 (i, j) 可以作如下两种机动安排:或者适当推迟其开工时间(不超过其最迟开始时间),或者适当放慢进度,延长其工时 $t(i, j)$ (延长的幅度不超过 $R(i, j)$ 就行)。当然,也可以将两种机动安排结合使用:即适当推迟一些开时间,又适当放慢一些该作业的进度。这些情况称为“时差的调用”,记调用量为 $\delta(i, j)$,则调用原则可统一表示为 $\delta(i, j) \leq R(i, j)$ 。

$R(i, j)$ 的数值实际上表明作业 (i, j) 具有潜力,调用时差就是挖掘其中的潜力(人力、物力、财力)。

8.3.6 作业的单时差 $r(i, j)$

当 $R(i, j) \neq 0$ 而 $\delta(i, j) > 0$ 时,只要满足 $\delta(i, j) \leq R(i, j)$,整个任务的总工期是不会拖延的,但是对于后续作业 (j, k) 来说,则会出现两种情况:一是后续作业受到干扰,无法在其最早开始时间 $t_{ES}(j, k)$ 开工;一是后续作业不受干扰,仍然可以在 $t_{ES}(j, k)$ 开工,对于后一种情况,我们称作业 (i, j) 具有“单时差”,记为 $r(i, j)$ 。就是说,作业 (i, j) 的单时差 $r(i, j)$ 定义为:在不影响后续作业 (j, k) 的最早开始时间 $t_{ES}(j, k)$ 的前提下,作业 (i, j) 可以自由地利用机动时间范围。用公式表示为

$$r(i, j) = t_{ES}(j, k) - t_{EF}(i, j) \quad (8.17)$$

将式(8.7)及式(8.9)代入式(8.17),得到

$$r(i, j) = t_E(j) - t_E(i) - t(i, j) \quad (8.18)$$

8.4 关键路线与时差的关系

网络图中,时差为零的节点称为关键节点,总时差为零的作业称为关键作业。

在一张网络图中,从始点到终点,沿箭头方向把总时差为零的作业连接起来所形成的线路称为“关键路线”。关键路线在图上用粗线或双线表示。

要想缩短整个任务的工期,必须在关键路线上想办法,即缩短关键路线上的作业时间。反之,若关键路线的工期拖长,则整个计划工期就要拖长。在一张网络图中,关键路线可以有 multiple 条。关键路线越多,表明各项作业的周期都很紧张,要求必须加强管理,严格控制,以保证任务的按期完成。

关键路线与时差的关系可归纳为以下三条定理。

定理 8.1 在 $t_L(n) = t_E(n)$ 的前提下,对于网络图上所有各项作业均有

$$0 \leq r(i, j) \leq R(i, j)$$

且当 j 为关键节点时, $r(i, j) = R(i, j)$, 特别地, 有 $r(i, n) = R(i, n)$ 。

定理 8.2 在关键路线上,全部节点时差均为零,反之不真。

定理 8.3 在关键路线上,全部作业的总时差均为零,反之亦真。

定理 8.3 给出了确定关键路线的充分必要条件。当我们计算了各项作业的总时间差以后,由 $R(i, j) = 0$ 的作业所连成的通路即为关键路线。或者说,关键路线就是关键作业所组成的通路。但要注意:定理 8.2 只是提供了确定关键路线的必要条件,而关键节点组成的通路并不就是关键路线。

由定理 8.1 还可知道,在关键路线上,各项作业的单时差亦为零。

非关键路线上各项作业的时差不全为零,就是说,或多或少有潜力可挖。

系统网络技术的精华就在于根据网络图找出关键路线,重点保证关键路线;利用非关键路线上作业的时差,调用其中的人力、物力、财力去支援关键路线,使得关键作业——从而使得整个任务——能按期或提前完成。

调用时差时,应该调用单时差,因为它是自由时差,对后续作业不发生影响。总时差与单时差的关系可以用图 8-5 表示。

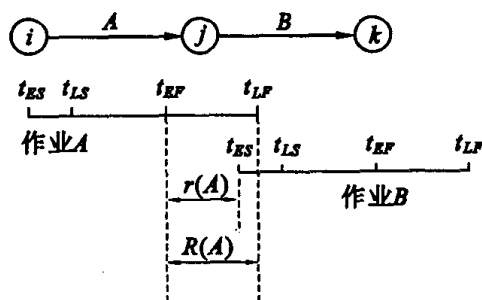


图 8-5

计算 $t_E(j)$ 、 $t_L(i)$ 、 $R(i, j)$ 和 $r(i, j)$, 找出网络图中的关键路线, 这种在网络图上计算作出作业时间的方法称为作图算法, 其计算程序如图 8-6 所示。

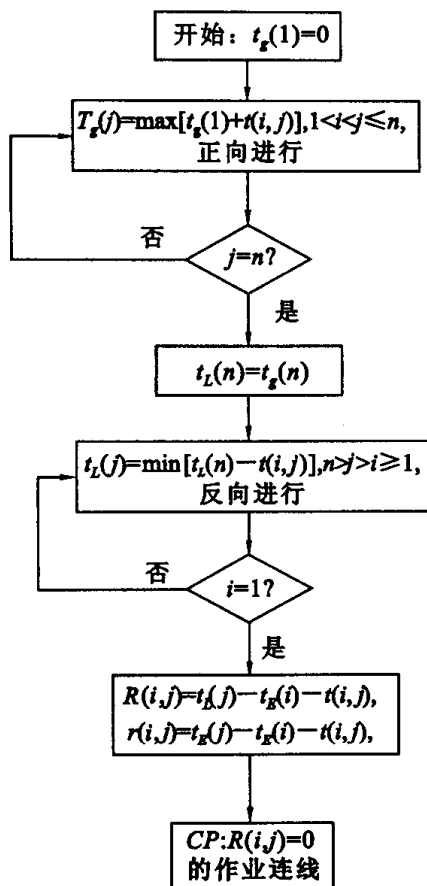


图 8-6 作图计算法过程

8.5 任务按期完成的概率分析与计算

在前面一节里,我们是从确定的情况出发,对网络的有关时间参数进行计算的。这里有一个前提,即每个作业所需时间都是确定的数值,因而算出来的总计划所需时间也是确定的数值。在这种把握比较大的情况下,我们使总预定完成时间等于总的所需时间,即前面讲的 $t_L = t_E$ 。这就是说,我们经过科学的计算,一个计划完成需要多少时间,就给它多少时间,这样找出的关键路线是以总时差为零来确定的。

在实际工程中,完成某项作业或任务所需的时间是不容易确定为某一数值的,往往只能凭经验或过去的试验研究结果来做一定的估计。对于这种估计所带来的一系列问题需要慎重考虑。前面讲过,作业时间是用“三点估计法”求出的平均时间,因而用公式算出的最早开始时间,也必然有某些不确定因素在内,并不是非常准确的时间。所以,我们就要研究由于这些不确定因素而引起的计划是否能按期完成的问题,即计划按期完成的可能性有多大?这就是任务按期完成的概率分析的由来。下面对此做一简单的介绍。

8.5.1 任务完成时间近似符合正态分布规律

根据概率论中有名的“中心极限定理”,我们可以认为,任何事项的完工时间是近似符合正态分布的。有了这样一个假设后,我们只要计算出每个作业预计完工时间的平均值和方差,就可以用下面公式求出各个事项按期完成的概率,就能对整个任务是否按期完成给予概

率评价,并对计划的执行作出预测。

(1) 平均值

前面讲过, $t_E = \frac{a + 4m + b}{6}$ 叫做作业的平均时间,它与三个估计时间参数有关。

(2) 离差与方差

$\sigma = \frac{b-a}{6}$, σ 为作业时间的标准离差,它只与作业的最长时间与最短时间有关,与最可能时间无关,它是作业时间概率分布离散程度的度量。方差为

$$\sigma^2 = \left(\frac{b-a}{6} \right)^2$$

(3) 任务完成时间

根据概率论中有名的“中心极限定理”来分析,任务最后完成时间呈正态分布,以 $t_m = \sum_{i=1}^J \frac{a_i + 4c_i + b_i}{6}$ 为平均值,以 $\sigma = \sqrt{\sum_{i=1}^J \left(\frac{b_i - a_i}{6} \right)^2}$ 为标准离差。

这是当 J 充分大时的一个渐近估计。实际上,我们可用概率曲线来形象地分析。当关键路线上的作业由 10 个以上组成时,从均值左侧伸展 $-3\sigma_\phi$ (σ_ϕ 为关键事项的标准离差),又从右侧伸展 $+3\sigma_\phi$,事项在这一段时间内实现的机会为 99.7%,这可用图 8-7 来说明。

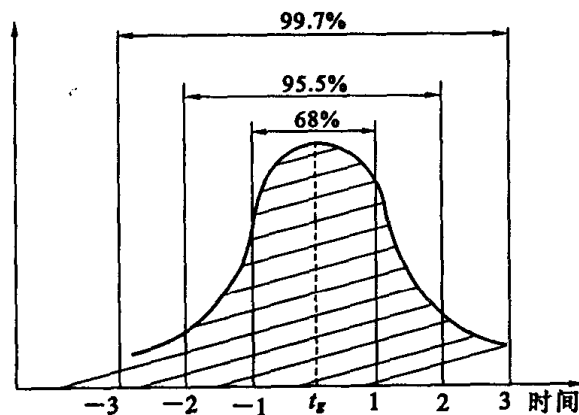


图 8-7 完成任务的概率分布

曲线上任一点的高度用来衡量事件在横轴上相应的时间里实现的概率密度,任意一条垂直线左方的面积代表在相应时间里实现的概率,实际概率曲线从 $-3\sigma_\phi$ 到 $+3\sigma_\phi$ 的水平宽度是对于 t_E 的不确定性的一个度量。 σ_ϕ 越小,水平宽度越窄,则 t_E 越肯定。反之, σ_ϕ 越大,则 t_E 越不肯定。

σ_ϕ 是从一系列 σ^2 值推导出来的,而 σ^2 值又依赖于每项作业的 $\frac{b-a}{6}$ 。为了尽可能地增加终点节点的 t_E 值的确定性,应使 $(b-a)$ 越小越好。但是不大可能在一切情况下都把 $(b-a)$ 限制得很小,可是对那些特别大的 $(b-a)$ 至少应该检查一下,看看是否判断错误。

8.5.2 任务按期完成的概率计算

(1) 计算方法

为计算概率,引入一个概率因子 z ,以便于查正态分布表。概率 $P = P(z)$

$$z = \frac{t_L - t_E}{\sigma_{cp}}$$

式中 t_L 为节点(事项)的最迟完成时间;

t_E 为节点(事项)的最早开始时间;

$t_L - t_E$ 为节点的时差。

σ_{cp} 是关键路线上方差之和的平方根,也叫节点的标准离差。 $\sigma_{cp} = \sqrt{\sum \sigma_i^2}$, σ_i^2 是关键路线上作业的方差, σ_{cp} 是平均值。

若已经规定了计划的完成时间为 t_s , 则 t_L 就用此规定时间。因 t_E 是已知的, σ_{cp} 也是已知的, 这样就可以计算出 z 值。用 z 值查正态分布表即得概率 P 。

(2) 应用举例

【例 8.1】 设某工程的网络图如图 8-8 所示。试计算该工程在 20 天完成的可能性。如果完成的可能性要求达到 94.5%, 则工程的工期应规定为多少天?

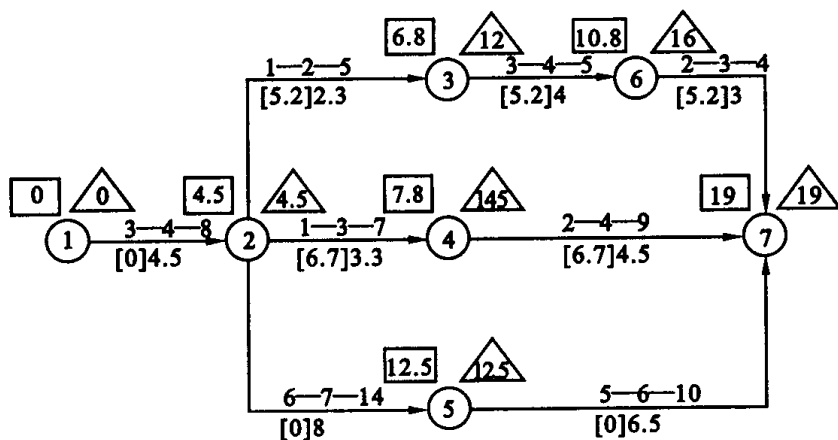


图 8-8 某工程的网络图

【解】 ① 先求出 $t_m(i, j)$ 。 $t_m(i, j) = \frac{a + 4c + b}{6}$ 求得, 标在每个作业线的下面, 时间单位为天。

② 计算 t_E 值。用公式 $t_E(j) = \max[t_E(i) + t_m(i, j)]$ 求得 $t_E(7) = 19$ (天)。

③ 计算 t_L 值。用公式 $t_L(i) = \min[t_L(j) - t_m(i, j)]$ 求得 $t_L(7) = 19$ (天)。

④ 计算 $t_L = 19$ (天) 的完成概率。

$$\text{使用 } z = \frac{t_L(7) - t_E(7)}{\sigma_{cp}} = \frac{19 - 19}{\sigma_{cp}} = 0$$

查表: 由 $z = 0$ 查得 $P = 0.5$, 即工程按时完成的可能性为 50%。

⑤ 当 $t_s = 20$ (天) 时完成任务概率的计算。

关键路线上的方差 σ_i^2 分别为

$$\left(\frac{8-3}{6}\right)^2 = \frac{25}{36}; \quad \left(\frac{14-6}{6}\right)^2 = \frac{64}{36}; \quad \left(\frac{10-5}{6}\right)^2 = \frac{25}{36}$$

$$\text{关键路线上方差之和} \quad \sum \sigma_i^2 = \frac{25}{36} + \frac{64}{36} + \frac{25}{36} \approx 3.2$$

由此可得概率因子

$$z = \frac{t_s - t_E}{\sqrt{\sum \sigma_i^2}} = \frac{20 - 19}{\sqrt{3.2}} = \frac{1}{1.8} \approx 0.56$$

查正态分布表得 $P = 0.71$, 即该工程在 20 天完成的可能性为 71%。

⑥ 如要求按时完成的可能性为 94.5%, 计算需要的天数, 由公式 $z = \frac{t_L - t_E}{\sqrt{\sum \sigma_i^2}}$ 得

$$t_L = t_E + z\sqrt{\sum \sigma_i^2}$$

查表 $P = 0.945$ 时之 $z = 1.6$, 于是 $t_L = 19 + 1.6 \times 1.8 = 22$ (天)。即按时完成可能性为 94.5% 时所需天数为 22 天。

(3) 讨论

由公式 $z = \frac{t_L - t_E}{\sigma_{cp}}$ 来看, $t_L - t_E =$ 时差。

严格地讲, 时差也要用概率来处理, 但这样算起来工作量太大了。由于时差只是作为调整的参考, 所以也就省略了概率计算。

若 $t_L - t_E = 0$, 则完成任务的概率为 50%; 若 $t_L - t_E > 0$, 则完成任务概率就大于 50%; 假设 $t_L - t_E$ 为 σ_{cp} 的三倍, 则完工概率为 99%; 若 $t_L - t_E < 0$, 则完成任务的概率就小于 50%; 如果 $t_L - t_E < -\sigma_{cp}$ 时, 可以判断按时完成任务是成问题的; 如果 $t_L - t_E < -2\sigma_{cp}$, 则肯定是大有问题了; 如果 $t_L - t_E < -3\sigma_{cp}$, 则按时完成任务只有千分之一的可能性。

这样, 管理人员就可根据时差和标准离差来判断按时完成任务的可能性, 也就有了一个标准。这就是置信度的分析与计算。

一般当任务在指定日期完成的概率 $P(z)$ 满足下列条件:

$$0.3 \leq P(z) \leq 0.70$$

则表示按此网络图执行计划, 在指定日期完成是可能的, 比较合适的计划定得既先进又留有充分余地。

(4) 按照上面讨论的几种情况重新评价关键路线的定义

我们是用时间平均值把非确定型化为确定型, 从而找出关键路线。这样做是否合适, 值得重新考虑。化为确定型而算关键路线的方法, 可以看成在以 1/2 的可能性为完成整个任务的条件下, 确定关键路线。确切的提法应该是: 给一个预计完成日期, 在所有的线路中, 依预计日期完成的可能性最小的才是关键路线。而不能只把总时差为零的路线称为关键路线, 即应从时差为负值、零或正值三种情况综合考虑来定关键路线。

8.6 网络图调整与优化

在编制一项工作计划时, 企图在最初的方案中得到完全合理的指标, 一般来说是不大可能的, 采用系统网络技术也如此。通常初始方案制定了以后, 需要根据指标的要求, 对初始方案进行调整和优化, 使计划符合要求。网络图调整和优化的主要内容有: 缩短网络计划的工期, 使其符合规定工期的要求; 工期不变, 使需要的资源(人力、物力、资金)最少; 降低人力使用高峰, 使其符合人力供应能力, 并使各工种人员的使用连续均衡, 且工期最短; 降低资源需要高峰, 使其符合资源供应情况, 且工期最短; 缩短工期并使费用增加最少等等。通过上述

几个方面的调整与优化,不断改善网络的初始方案,使之获得最优的周期、最低的成本和对资源的最有效的利用,并根据时间、资源和成本等方面优化的结果,选择最优方案,作出决策。

8.6.1 缩短网络计划工期

在网络系统中,关键路线决定了总工期。当规定的工期大于关键路线持续时间时,关键路线上各作业的总时差就会出现正值,说明完成该任务的时间比较宽裕,计划内安排的时间还有潜力可挖,必要时可适当延长某些工作的持续时间,以便做到减少资源或节省费用。当任务比较紧急,规定的工期小于关键路线的持续时间时,则需对网络进行调整。调整时可针对超过工期的各条线路上的某些作业,采取组织上和技术上的措施,以缩短它们的工作持续时间,从而使计划符合规定工期的要求。

缩短计划工期,通常可从组织上和技术上采取如下方法:

- (1) 在关键路线上寻找最有利的作业来缩短其作业时间;
- (2) 在可能条件下采取平行作业或交叉作业缩短工期;
- (3) 采取新技术和新工艺,搞技术革新和技术改造,增加人力和设备等,缩短某些工作的作业时间;
- (4) 利用时差,从非关键路线上抽调适当的人力、物力集中于关键路线,以缩短关键路线的持续时间。

采用上述诸方法缩短网络工期,在调整过程中都会引起网络计划的改变,每次改变后都要重新计算网络时间参数和确定关键路线,直到求得最短周期为止,现举例说明如下。

【例 8.2】 图 8-9 为某工程的网络图,网络初始方案计划时间为 19 周,现根据客观情况的要求,上级规定工程的完工期为 16 周,试对网络进行调整。

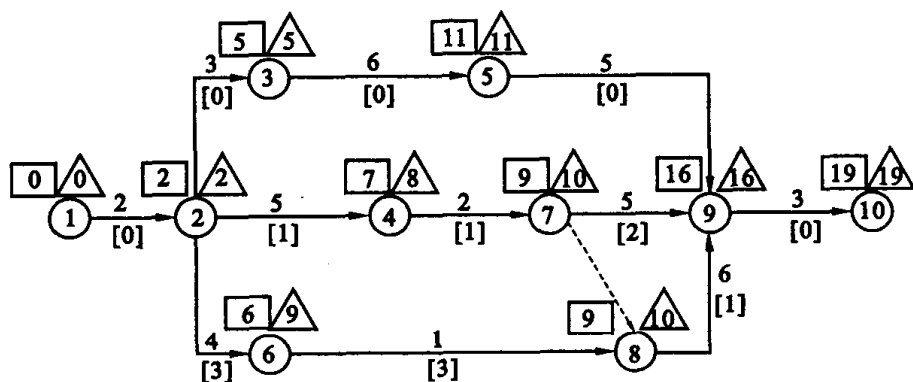


图 8-9 某工程的网络图

【解】 为达到上级规定的工期要求,需要对作业时间重新进行安排,采取措施压缩关键路线的持续时间。图 8-9 的关键路线是:

$$\textcircled{1} \rightarrow \textcircled{2} \rightarrow \textcircled{3} \rightarrow \textcircled{5} \rightarrow \textcircled{9} \rightarrow \textcircled{10}$$

缩短工期的计算:首先将终点 ⑩ 的最迟完成时间定为 16 周,然后采取倒退算法,求出网络中各作业的时差,再确定应该缩短哪条路线上的工期,计算结果如图 8-10 所示。

图 8-10 中箭线下方括号内的数字为总工期 16 周时各作业的总时差,从计算结果可以看出,在原先的关键路线上各作业的总时差为(-3),这意味着在原来的关键路线上应缩短 3

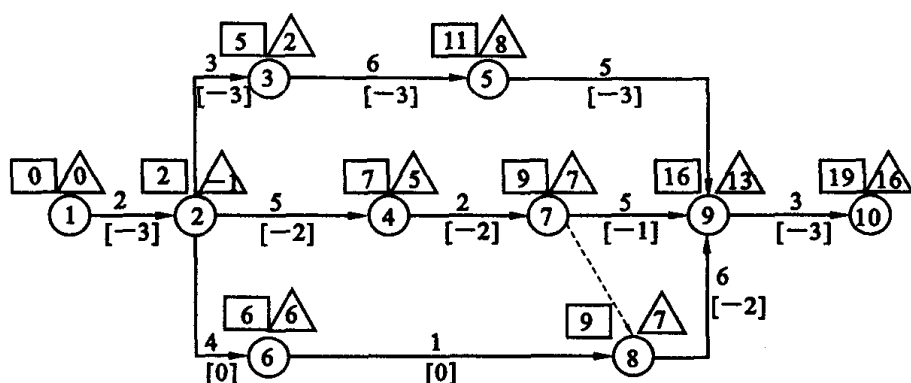


图 8-10 调整工期(1)

周。除原来关键路线是负时差外,其他非关键路线也出现了负时差,即在这些路线上也要进行日期的缩短。需要缩短日期的线路和作业有:

- (1) ① → ② → ③ → ⑤ → ⑨ → ⑩
- (2) ② → ④ → ⑦ → ⑧ → ⑨
- (3) ⑦ → ⑨

上述各线路和作业需要缩短的时间分别为3周、2周、1周。首先考虑在(1)线路上缩短3周,即②→③、③→⑤、⑨→⑩各缩短1周,缩短后进行网络时间参数的计算,其结果如图8-11所示。

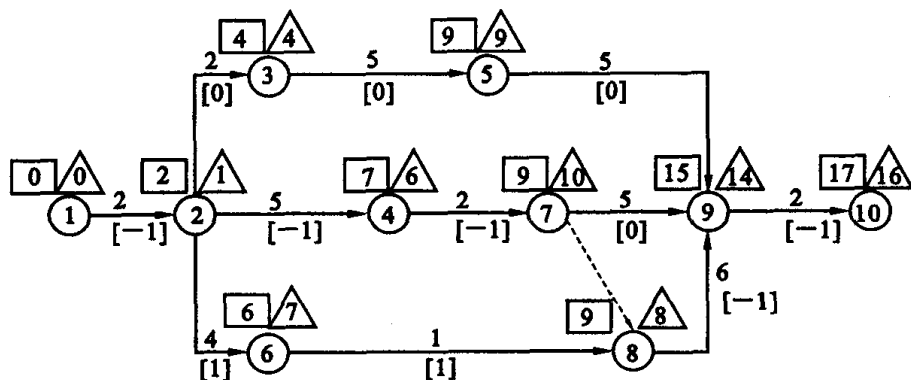


图 8-11 调整工期(2)

从图8-11中可以看出,有负时差的线路为①→②→④→⑦→⑧→⑨→⑩。假设将上述线路中作业②→④缩短1周成为4周,则再通过计算,其结果如图8-12。

此时,网络中已全部消灭了负时差,下面两条线路均是关键路线:

- ① → ② → ④ → ⑦ → ⑧ → ⑨ → ⑩
- ① → ② → ③ → ⑤ → ⑨ → ⑩

工期为16周,符合上级规定的要求。

8.6.2 网络执行过程中的检查与调整

利用网络图对工程进行控制,是系统网络技术的主要功能之一。系统网络技术之所以能被广泛应用,有今日之成就,主要是因为它能以动态规划方法随时对工程进度进行检查、调整与控制。

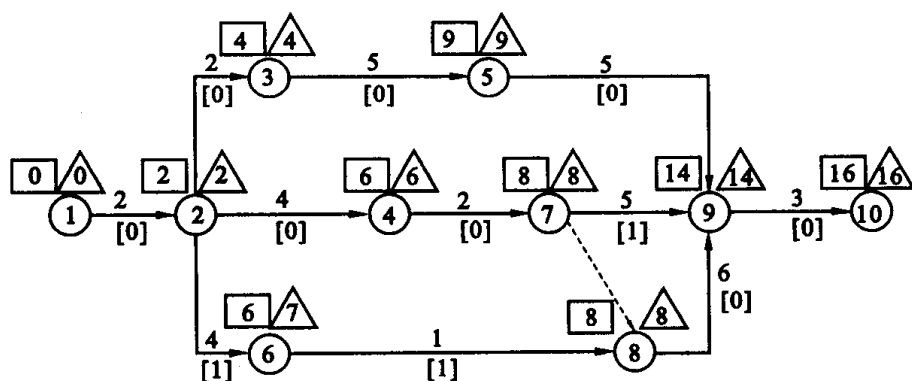


图 8-12 调整工期(3)

网络虽然经初步规划,但实际上并不一定能按原计划顺利实现或完成,因为在实施过程中会遇到许多无法确定的因素以及可能出现的意外情况,因此,对网络进行定期检查与调整,是监督计划执行、控制工程进度的有效方法,也是保证实现预期目标的一项重要工作。

(1) 对关键路线上各作业的检查与调整

关键路线决定总工期,关键路线上任何一个作业(关键作业)的缩短或延长都会影响整个工程的进度,因此,检查与调整的重点是首先分析关键作业。研究这些作业是否有可能提前或拖延,并随时分析原因,采取措施进行调整。例如,当关键路线上某些作业的延续时间缩短或延长了,必将相应改变关键路线上后续各作业和非关键作业的时间参数,因此,需要对这些作业进行必要的调整,并重新计算时间参数,使其符合要求。

(2) 对非关键路线上各作业的检查与调整

当非关键路线上某工作的作业时间延长了,但在作业总时差范围内,这时虽然改变了相应作业的时间参数,但不影响总工期,在一般情况下,网络不需要进行调整。如果非关键作业延长的时间超过了可利用的时差,有可能出现关键路线的转移,产生新的关键路线,这时需检查非关键路线,并对其进行适当的调整,以满足工期的要求。

(3) 检查网络编号,调整网络逻辑关系

当网络计划在执行过程中需要增加某些作业时,则需检查网络编号,重新调整逻辑关系,并要计算出调整后的关键路线和总工期。例如某工程网络,原由 A、B、C、D 四个作业组成,如图 8-13(a) 所示。在执行过程中,发现 A 作业完成后,与 B、C 作业同时进行的还有 E 作业,否则 D 作业不能开始。由于在原网络图编号时考虑了增加作业的可能性,留有备用号,因此增加 E 作业不必改变原来编号,只需增加作业箭线和事项(节点),并补上一个预留空号即可,如图 8-13(b) 所示。

关于执行过程中如何对网络进行检查与调整的问题,现举例说明如下:

图 8-14 为某工程项目的网络图,根据各项时间参数的计算,得出该工程的完工期为 23 周。

开工 10 周后,检查工程进度,即对作业 ② → ③、② → ④、② → ⑤ 进行检查,其结果是:

② → ③ 完成约 33%,剩余日数为 2 周;

② → ④ 完成约 15%,剩余日数为 6 周;

② → ⑤ 完成约 12.5%,剩余日数为 7 周。

将检查结果绘成网络图,则得图 8-15。为使新加入的事项与原有事项有所区别,新事项

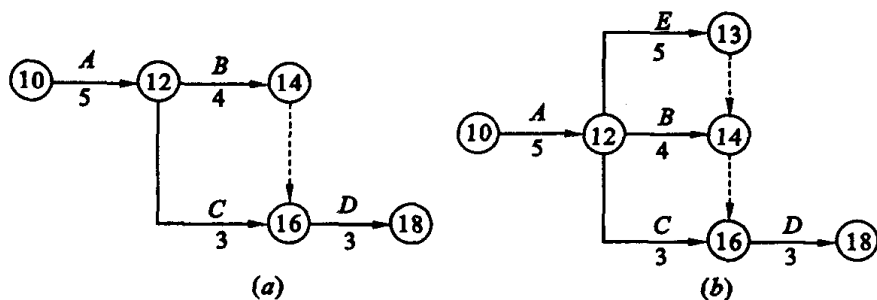


图 8-13 某工程网络图

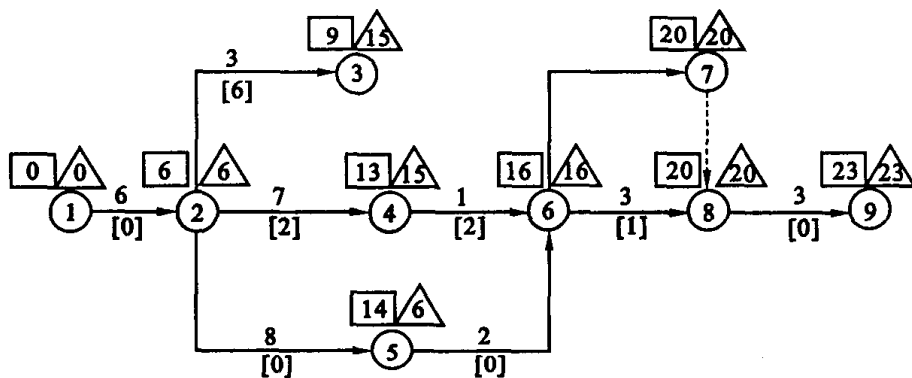


图 8-14 某工程项目网络图

编号用 a 、 b 、 c 、 d 表示。

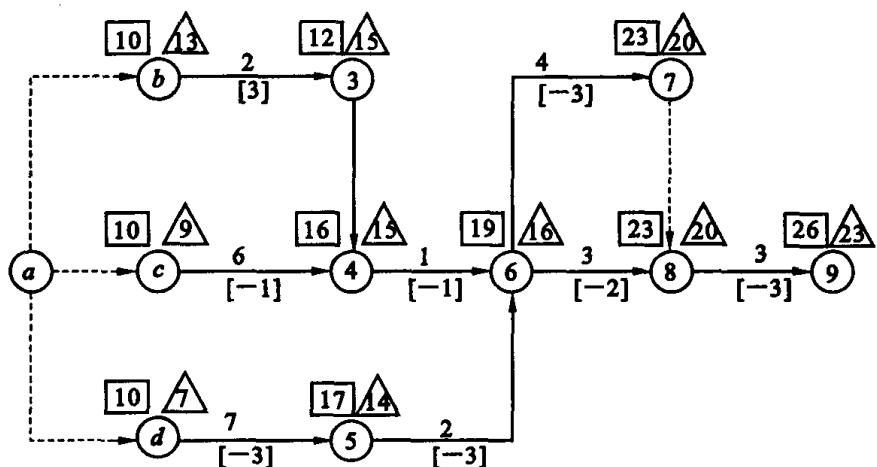


图 8-15 调整工程工期(1)

通过对图 8-15 各时间参数进行计算,得出此工程总完工期为 26 周,比原网络的总完工期超过 3 周,在图 8-15 中出现了负值的作业总时差。为消除负值的总时差,可将作业 ②→⑤ 由原来 7 周作业时间缩短 2 周改为 5 周,将作业 ⑥→⑦ 由 4 周改为 3 周,再经过时间参数的计算,则可达到规定的总工期 23 周,而关键路线也发生了变化,如图 8-16 所示。这时,如要使总工期在 23 周内结束,应按图 8-16 的网络图的时间参数进行工程实施,而由于出现了两条关键路线,工程的进度要比原网络图更加紧迫了。

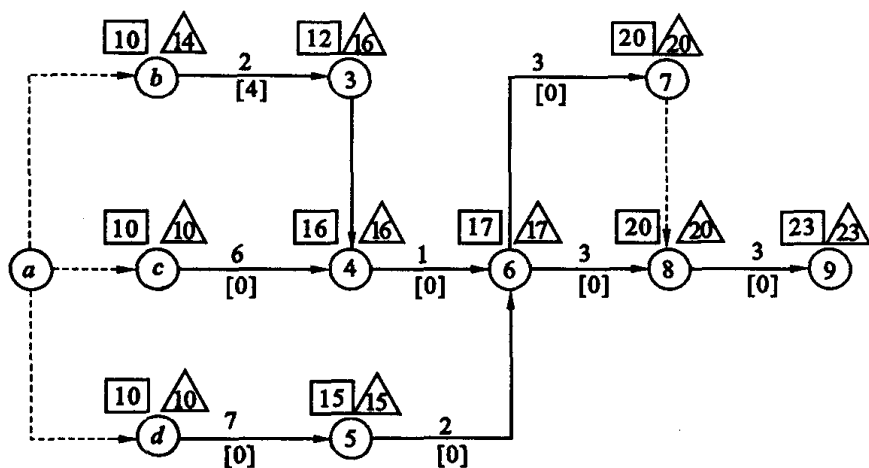


图 8-16 调整工程工期(2)

8.7 网络的时间—费用分析与优化

在一个系统的分析中,我们希望的是在保证性能和效果的前提下,任务完成的时间最短和费用最少,因此需要对时间和费用进行综合分析,基本的 PERT 只谈如何缩短时间问题,而 CPM 却将成本因素作为重点,谋求最少的成本来缩短时间,这样去求网络的最优解,故称网络的费用分析。这里主要介绍以最少的直接成本实现经济赶工的方法。

8.7.1 直接成本与时间的关系

计算完成时间常常受到资源的影响,若增加资源,则可缩短完成时间,但成本也将随之增加。反之,若减少资源,将使完成时间拉长,然而可能降低成本。因此两者可以相互替换。需进行比较。直接成本与时间的关系可用图 8-17 表示。

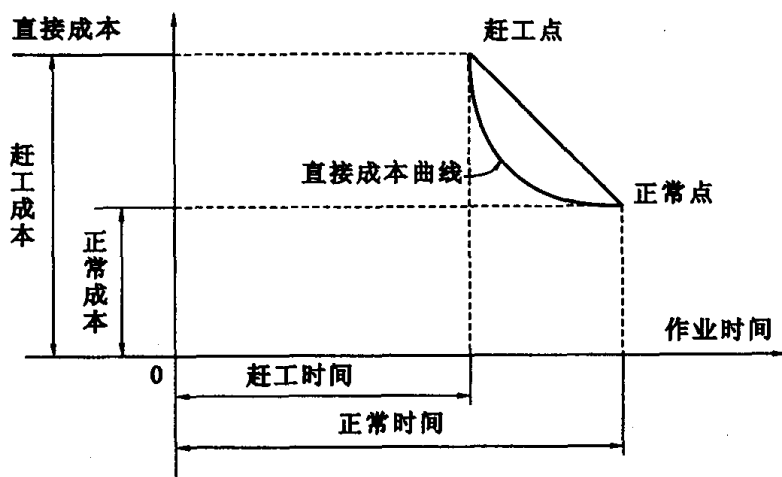


图 8-17 直接成本与时间的关系

赶工时间是指某项工作的作业时间从正常状态缩短到无法再缩短为止的时间,在此时间里所需要的成本为赶工成本。由于将某项作业时间缩短到比正常的时间短,所以,此时所用的成本往往会比正常成本高,达到赶工点时,其成本最高。直接成本与作业时间的关系为一曲线,接近于赶工点与正常点的连线,该连线的斜率,其意义是每缩短一个单位时间所需

增加的费用。用公式表示为

$$\text{成本斜率} = \frac{\text{赶工成本} - \text{正常成本}}{\text{正常时间} - \text{赶工时间}}$$

8.7.2 经济赶工方法

用一个例子来说明,图 8-18 为某一方案的网络图,数字的单位是天,已知的成本斜率资料列于表 8-1。

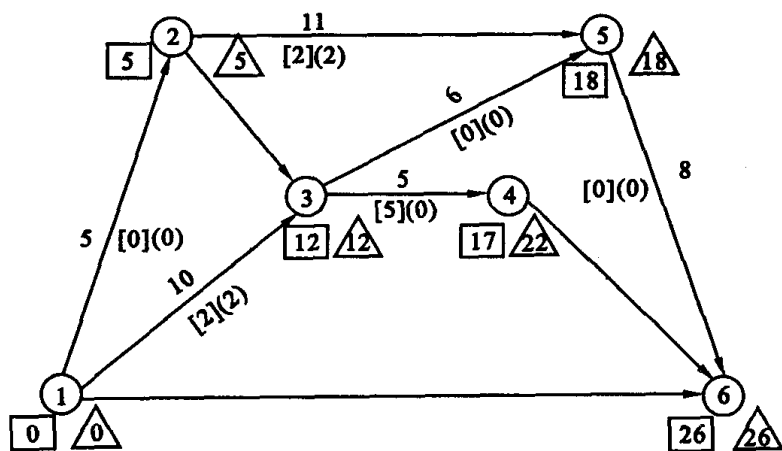


图 8-18 二程网络图(一条关键路线)

表 8-1

A	B		C		D
作 业	正 常		赶 工		成本斜率 (元/天)
	时 间(天)	成 本(元)	时 间(天)	成 本(元)	
①→②	5*	220	4	300	80
①→③	10	480	8	660	90
①→⑥	8	440	6	640	100
②→③	7*	600	5	720	60
②→⑤	11	560	9	640	40
③→④	5	600	2	1200	200
③→⑤	6*	150	5	250	100
④→⑥	4	180	4	180	不能赶工
⑤→⑥	8*	600	7	760	160
	26**	3830	21	5350	

(1) 现考虑如何以最低成本来缩短一天时间,即完成时间缩短为 25 天。思考方向:从关键路线着手,即在关键路线上取成本斜率最小的作业来缩短一天。图 8-18 中的关键路线为①→②→③→⑤→⑥,将关键作业及成本斜率另列于表 8-2 中,取成本斜率最小的②→③作业缩短一天即可。这样关键路线不变,但成本增加 60 元。

表 8-2

作 业	成 本 斜 率
①→②	80
②→③	60
③→⑤	100
⑤→⑥	160

(2) 若将总时间由 25 天缩短到 24 天。同理,可再花 60 元缩短②→③即可,如图 8-19 所示。这样是关键路线从一条增至三条,即:①→②→③→⑤→⑥;①→③→⑤→⑥;①→②→⑤→⑥。

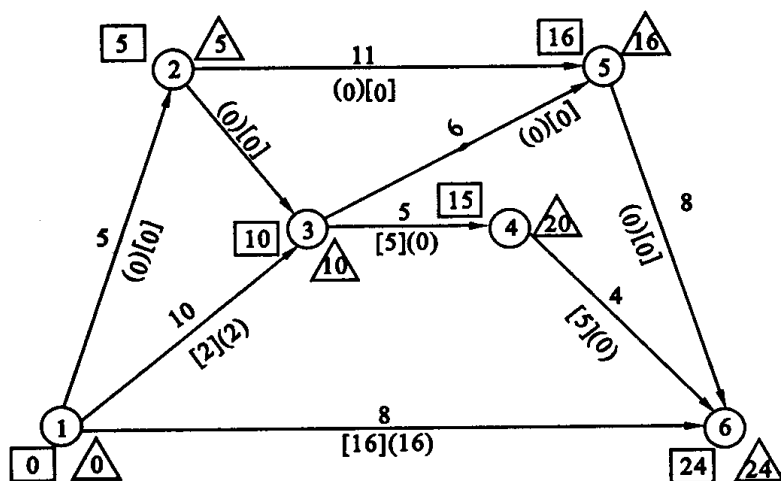


图 8-19 具有三条关键路线(缩短至 24)的网络图

(3) 若再将时间缩短一天,变为 23 天时,如图 8-20 所示。由于关键线路已有三条,因此,必须对每一条关键线路缩短一天才能达到要求。先把三条关键线路和成本斜率加以比较后列于表 8-3 中。

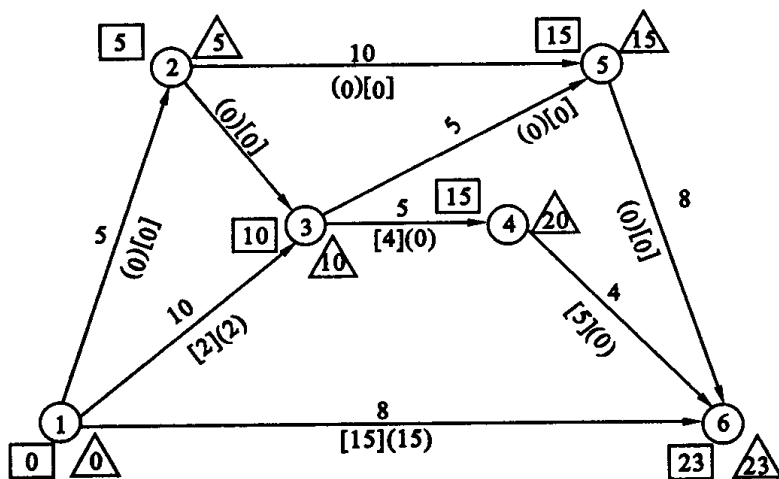


图 8-20 缩短至 23 天的网络图

表 8-3

关键路线 I		关键路线 II		关键路线 III	
作 业	成本斜率	作 业	成本斜率	作 业	成本斜率
①→②	80	①→②	80	①→③	90
②→⑤	40	②→③	60	③→⑤	100
⑤→⑥	160	③→⑤	100	⑤→⑥	160
		⑤→⑥	160		

从各关键路线中选出成本斜率最小的:即 I 中选②→⑤; II 中选①→②(因②→③已不能再减); III 中选①→③。三者加起来为 210 元。我们把三个关键路线各缩短一天的各种组合列成表 8-4。可以看到 B 方案成本最少,即②→⑤的 40 元,加上③→⑤的 100 元,合计 140 元,同时它也满足三条关键路线的需要。再加上②→③作业缩短 2 天所花的费用为 60 元 $\times$ 2=120 元,共需 140 元+120 元=260 元。这就是由 26 天缩短为 23 天所需的赶工费用。

表 8-4

方 案	I	II	III	成本合计
A	②→⑤40 元	①→②80 元	①→③90 元	210 元
B	②→⑤40 元	③→⑤100 元	③→⑤0 元	140 元
C	②→⑤40 元	⑤→⑥160 元	⑤→⑥0 元	200 元
D	①→②80 元	①→②0 元	①→③90 元	170 元
E	①→②80 元	③→⑤100 元	③→⑤0 元	180 元
F	①→②80 元	⑤→⑥160 元	⑤→⑥0 元	240 元
G	⑤→⑥160 元	⑤→⑥0 元	⑤→⑥0 元	100 元

(4) 从表 8-4 还可以看出,E 虽然比 B 不利(180 元 $>$ 140 元),但如果按 E 方案绘制网络图如图 8-21 所示,即 I 的①→②:80 元、II 的③→⑤:100 元;III 与 II 共通③→⑤:0 元,共花 180 元。

这样,可以看出②→③作业已离开关键路线,成了有时差为[1]的作业,因此②→③作业于 6 天完成也不会影响 23 天的总完成时间。所以可以使②→③只减少一天,这样节省一天的费用 60 元。因此:

E 方案的费用为:(60 元 $\times$ 2+180 元)–60 元=240 元

B 方案的费用为:(60 元 $\times$ 2+140 元)=260 元

为什么会出现 E 方案反而比 B 方案好的这种情况? 主要是由于缩短的 3 天是在②→③缩短 2 天的基础上再缩短。因此,这是比较而言的,一定要把各种方案都列出来,然后选择费用最小的方案来缩短时间。

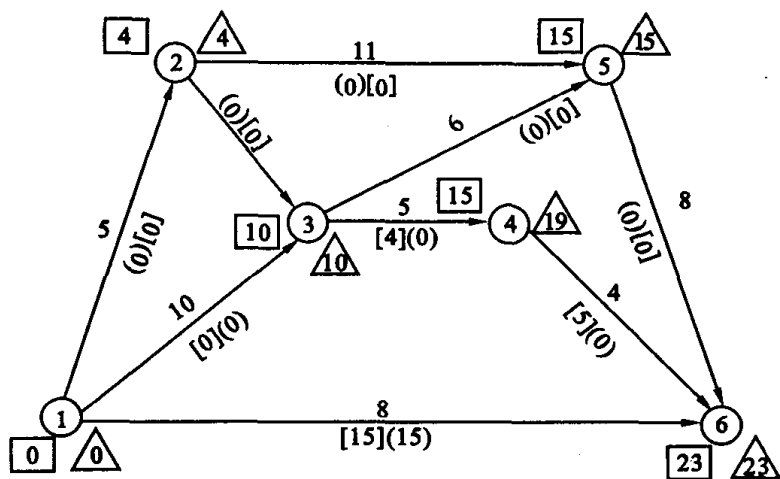


图 8-21

8.7.3 考虑间接成本的 CPM

它是在 PERT 的实践基础上,加入 CPM 的成本,即从实践和成本的综合效益最佳来确定方案的优劣。有关的参数及计算与 PERT 及 CPM 相同,这里就不具体说明了。现用曲线来表示它的基本原则:一个系统的总费用等于直接费用加上间接费用。一般地说,直接费用是一条曲线,间接费用是一条直线,总费用是一条 U 形曲线。所以,总费用曲线的最低点,即是时间费用的最佳值,如图 8-22 所示。

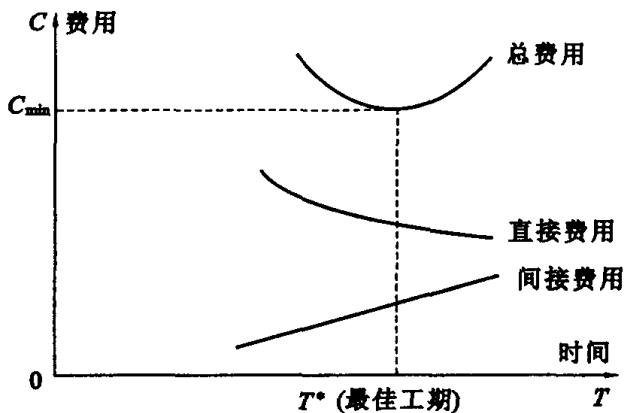


图 8-22 费用曲线

8.8 图解评审技术——GERT

8.8.1 PERT 的发展——GERT

系统网络技术自 20 世纪 50 年代初产生 CPM 以后,描述现实管理系统方便、直观,易于学习掌握。50 年代末又产生了著名的计划评审技术 PERT 相继在许多工程管理中获得很大成功,成为管理中最广泛的应用技术之一。但因其本质上均为确定性网络,仍有较大局限性。表现在:

(1) 它只能协调评审确定性的网络模型。即模型中各个事项都确定要完成,在其输出端只有一种预定发展可能,无法处理概率型事项。

(2) 在它的网络模型中不能出现回路。即事项只能出现一次,不能重复执行。

(3) 整个网络模型中,只有一个终端事项。对于不确定因素较多的开发项目或经济运行系统,尚不能完成管理和决策的职能。

为了克服这些局限性,70年代以来,人们进而开发研究了随机型的网络技术——GERT,即 Graphical Evaluation and Review Technique 的缩写。最先,它是由美国 A. Pritskan 在研究阿波罗系统最终发射时间计算时于 1966 年提出的。随即该技术作为系统网络技术的一个分支,在工程、经济、生产、运输、军事和社会服务等方面得到广泛使用,显示出了强大的生命力。由于网络形式多样,其求解方法也随着计算机日新月异的发展,形成相应的计算机程序。这些程序大体上可归为两类:一类取解析求解方法;一类取仿真算法。如 GERT 为仅有“异或”一类输入逻辑节点的解析性算法求解的计算机程序。而 Q-GERT 则为 GERT 类仿真算法的软件大“家族”中的一种,其功能全面,通用性强。多数程序均以 FORTRANIV 为基础写成,有些程序对用户已形成自己的语言。

为了对 GERT 与 PERT 一类网络图作个对比性分析,下面试举一例来说明 GERT 的产生和形成。例如加工一个部件的过程分成加工工序、检查工序、送检工序、返修工序、再加工工序以及废品处理工序。这样一个简单的过程若用网络形式表示出来,如图 8-23 所示。求解的目的是这一加工过程的成品率和平均加工时间应是多少?

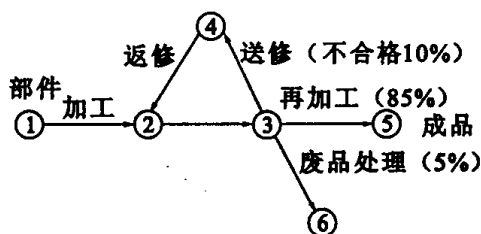


图 8-23 加工部件的 GERT 网络图

图 8-23 虽然在形式上与 CPM 及 PERT 相似,但实际上有下列显著不同:

(1) 检查工序完成后有三种可能:85%的部件经过再加工后成为成品;10%的部件送返修工序进行返修;5%的部件无法返修而成为废品。需要指出的是,这三种可能对同一个部件来说只有一种发生的可能,即它不需经过所有工序。对网络来说,只要网络中一个终点和箭线实现了,网络就被认为实现了。

(2) 检查工序、送检工序和返修工序,构成了一个闭合的加工过程。如果一个部件要经过两次以上的返修,则网络中的环可能循环两次以上。从网络图中清楚地看出,凡经过检查后达到节点的部件,无论它前面加工情况如何,都有 10%的可能性送返修。

(3) 对一个具体部件,检查后究竟要经过哪些工序或到底需返修几次?这完全是个随机问题。能够确定的只是他们经过哪个工序的概率。

从以上分析和图 8-23 可知,系统本身和它的网络模型与 CPM、PERT 以及它们所描述的对象相比都发生了突出的变化,解决问题的目标和计算方法也不相同,这就启发了人们研究和开发 GERT 模型。

8.8.2 GERT 网络节点特征及其绘制实例

与 PERT 出发,对网络中“节点”的输入输出逻辑加以扩展,输入端可以实现多项“活动”的“与”、“或”及“异或”等多种关系,节点输出活动具有一定执行概率或条件约束,活动参数向量化,活动时间分布多样化,并允许出现反馈回路。还可以在网络中引进自学习功能,自行进行网络更改。因此,整个网络中并非所有节点及活动分支都必须执行,部分网络以一定概率值出现或不出现,从而形成了执行的随机性。与 PERT 相比,主要是节点为逻辑节点。整个网络由逻辑节点和有向支路组成。逻辑节点用以说明系统行为状态,每个节点由输入端和输出端组成。输入端规定为三种逻辑符号,输出端规定为两种逻辑符号,此输入端的三种逻辑符号和输出端的两种逻辑符号共组成六种可能形式。详见表 8-5。

表 8-5 GERT 逻辑节点特点特征组合图表

组 合 形 式 输 出 端 \ 输 入 端	异 或	或	与
确定型			
概率型			

表中符号特征与含义:(1)“确定”型,如果该输出端节点实现,则所有从该节点输出的分支均要发生。这表明所有输出支路出现的概率均等于 1;(2)“概率”型,如果该输出端节点实现,则从该节点输出的分支最多只能出现一个;(3)“异或”型,输入端任何一条支路的实现都要导致该节点实现,但在一个时间间隔中,只允许实现一条支路;(4)“或”型,任何一条输入支路实现,都会导致该节点实现,该节点实现时间即为最早实现的支路所需时间;(5)“与”型,该输入节点只能在所有支路实现后方能实现。所以,它的实现时间应按支路中时间最长的支路计算。

在 GERT 网络图中,每条有向支路有两个特征参数,即该支路出现的概率(P)和完成该支路活动所需时间(t)。

GERT 网络图的绘制,就是根据以上逻辑节点的组合形式,针对欲研究的对象,再加上有向支路而描述的。为了准确的使用各种逻辑节点,必须在绘制前掌握研究对象的活动过程和规律。只有全面深入的掌握了研究对象的活动过程,以及每一局部活动所带来的结果的可能性,才能如实的运用各种逻辑节点。这是绘制 GERT 网络图的前提。

下面以某机械厂加工某零件的简单生产过程系统为研究对象,来说明 GERT 的绘制及应用(实际工作中运用 GERT 网络图研究系统时比这要复杂、困难得多)。假设这个厂生产某零件需要经过粗加工和精加工才能形成零件成品最后入库,并且在粗加工后和精加工前,均要进行检验,合格者入库,不合格者报废。

对这个简单的生产过程系统进行分析后,就可以将其分为 1,2,3,...,10 个活动,并有 8 个逻辑节点。这样,便可绘制出如图 8-24 所示的网络图。显然该图有两个终点,每个结点

也因其生产活动的结果不同,而用不同的逻辑节点形式组合表示,如开始加工,输入端为“异或”型,输出端为确定型,所以用 ∇ 表示;反之,在零件第一阶段加工完成进行检验,检验结果必然有两种结果,或是合格或是不合格,所以其节点的输入端为“异或”型,输出端为“概率型”,故用 $\diamond$ 表示。其余以此类推,图 8-24 这类网络模型可以作为更大网络模型的一个组成部分。可见,它能够用来描述“多级”结构系统。

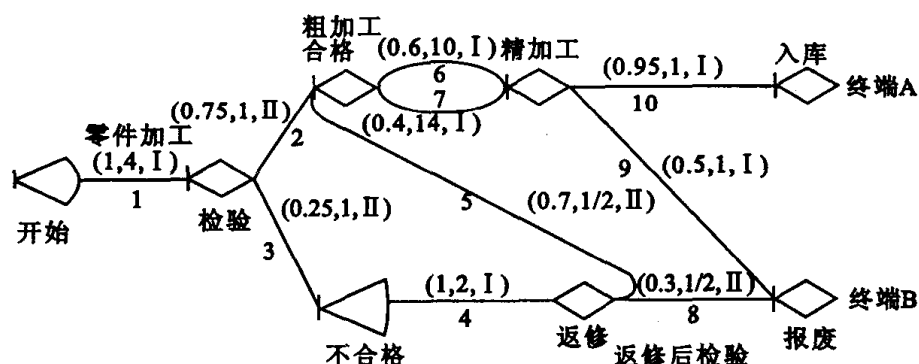


图 8-24 某厂某零件加工过程随机网络图

构成上述随机模型后,就形成了 GERT 网络模型的基本结构模型,然后,就应在此基础上做网络分析。进行网络分析的目的就是要根据等效网络中各支路的概率及需用平均时间,计算出终端事项的概率。这就是确定等效支路的特征参数,称之为网络分析。如图 8-25 即为等效网络示意。

在进行网络分析时,运用概率论及数理统计方面的知识,就可以求出等效网络各支路出现的概率和时间的数学期望值及方差。图中, P 为出现概率, t 为完成该支路时间。现在,我们仍以前面提到的某机械厂生产的零件为例,若已知一个零件从加工开始通过生产线,加工时间为 4 小时,在粗加工前经检验后有 25% 零件不合格需要返修。检验时间(包括等待时间)按负指数分布,期望值为 1,返修 3 小时,而返修后

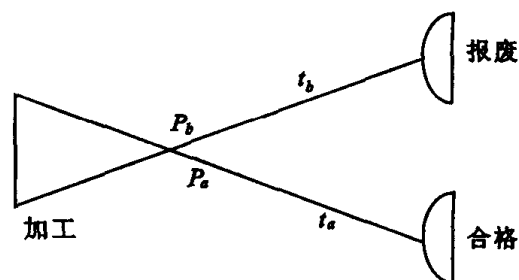


图 8-25 等效网络

另外有 30% 在下次检验后不合格而报废。返修零件检验时间仍按负指数分布,平均时间为半小时,送往精加工零件中有 60% 用 10 小时精加工,40% 用 14 小时,最后检验花 1 小时,且有 95% 零件合格入库。5% 零件报废。根据已有数据,再分别计算出各支路的概率,即可绘制出图 8-24。图中各行动箭线上括号内的三个数字:第一个数字表示该分支行动发生的概率 P_i ,如为 1 则表示确定型;第二个数字表示时间估计值;第三个数字表示行动时间符合哪一种概率分布(图中 I 表示正态分布,II 表示指数分布)。

根据这样的网络图,就可以求出投产的零件中有多少合格出厂,有多少报废,一个零件从投产到出厂平均需要多少时间,由此可见,分析求解复杂的随机网络模型,往往比较困难。所以在大系统中必须借助电子计算机进行数字模拟试验,从而得出等效网络的特征参数。然后有蒙特卡洛(Monte-Carlo)法,对其中一些随机数值,试验足够次数(如 1000 次,即相当于 1000 零件通过该生产过程),就可算出成品数、报废数和制造周期等。据资料介绍 GERT

主要用在时域系统中,如大型科研或工程的组织管理、生产组织管理、存储控制和可靠性设计等领域。

综上所述,GERT网络与CPM、PERT网络相比较,它克服了CPM和PERT网络中的局限性,扩大了适应范围,而且计算结果具有较强的预测性。所以说CPM和PERT网络是GERT网络的特例。

8.8.3 用GERT网络解决系统问题的步骤

用GERT解决系统问题与用CPM、PERT解决系统问题的步骤基本上是一致的,只不过在内容上有些不同。

(1) 对一个系统进行分析,确定要建立图解网络模型后,针对模型的要求对系统进行反复的考虑和剖析,找出其主要因素,反映到模型中,用分解—协调的原理分解工序,估计工期,选择节点符号,最后画出其GERT网络图。

(2) 对估计的工期进行认真的测算,以保证模型参数的准确。若工序的工期不能用常数表示时,可以事先估算一个方差的均值,或者估计出一个范围,然后再选择一个合适的理想分布密度函数表达它。

(3) 对模型进行分析计算,提出算法,编制程序。计算内容要由系统研究的目标来定。若在时间范畴内,主要是对时间、资源、费用进行优化。若在时间范畴外,主要是针对网络流进行分析计算,在理论方面进行深入研究是要涉及到网络拓扑和图论的内容,GERT法不仅要求解网络所消耗的时间、费用和资源,而且还要求得网络中的流。

(4) 最后综合评议和审定计算的结果,作出决策。

总之,在已有软件的支持下,随机网络方法的应用关键在于建模。网络建模较之其他方法为易,随机网络中语汇丰富,建模能力强,适用范围广泛。前面提到的Q—GERT中含有运行错误诊断能力,可以帮助建模人员确保模型结构合理,运行正确。新建网络模型应正确反映现实系统的内在联系,并注意保证网络元素之间有正确的逻辑结构。建模步骤上可以由简到繁,从局部到全局,分布逐级扩展。至于仿真运行次数的选择与问题本身性质及模型结构都有关系,应合理而经济地加以确定。

习 题

- 8.1 什么叫网络技术? PERT和GERT各具有什么特点?
- 8.2 用成本计划管理的(PERT)法来分析调整工厂的工程时间费用。
- 8.3 GERT法较之PERT法在哪些方面改进?

9 系统可靠性分析

本章提要

系统本身及系统所处的环境都存在着不确定性,诸多的不定性都影响着系统的可靠性。对于系统进行可靠性分析的目的在于,尽量提高系统的可靠性,以降低或避免各种不定性给系统带来的风险。

本章在给出系统可靠性定义和系统可靠度概念的基础上,介绍了6种典型的系统可靠性模型;给出了系统可靠性计算及预估方法;针对如何提高系统整体可靠性,讨论了系统可靠性指标合理分配到各子系统的方法;最后,从系统工程的角度介绍了可靠性工程与管理。

9.1 可靠性概念及度量指标

9.1.1 可靠性的基本概念

系统可靠性是评价系统的一个重要性能指标,系统可靠性分析是系统工程的一个重要组成部分。可靠性分析的任务是,在系统规划设计、生产制造、使用与维护的各个环节中,运用各种系统可靠性技术与方法评估系统的可靠性,提高系统的可靠性,更好地发挥系统效能,避免社会、经济和技术风险,获得尽可能大的社会经济效益。

系统的可靠性和可靠性分析的重要性是不言而喻的。只有高可靠性的产品才能赢得市场,获得成功。由于一个元器件的失效或者很微小的系统设计失误而导致失败,甚至发生悲剧的事故很多。虽然几乎所有领域都有可靠性问题,但引起人们关注的仅是民航飞机失事、环境污染和航天飞机爆炸等事件。可靠性研究的早期工作是40年代从德国和美国开始的。在朝鲜战争期间,美国国防部发现可靠性不高的设备维护费用极高,每一美元的电子设备每年需要2美元的维护费用。人们越来越认识到高可靠性元件和高可靠性系统的重要性。美国国防部在1966年就开始要求一切国防合同都必须在系统研制的所有阶段进行系统安全可靠性分析,并逐步形成了一系列关于可靠性的法规。

20世纪40~80年代,可靠性理论及应用获得了快速的发展。今天,它在人类社会经济的发展中仍然具有非常重要的作用,不仅仅是在研究所、实验室和制造业领域,在各行各业的多种类型的系统中都存在可靠性问题,因此,各类研究人员、工程技术人员和管理人员都应该掌握相应的可靠性理论知识和技能。

(1) 系统可靠性的定义

关于系统可靠性的定义主要有:

- A 系统在规定的条件下,规定的时间内,完成规定任务的能力。
- B 一个系统或设备,在预计的时间内,在给定的条件下能满足工作的概率。

这里,“规定的时间”很重要。因为,一幢房子,至少要求它能安全可靠地使用 50 年;一枚火箭只要在发射后的短短数分钟内安全可靠地工作即可。所以,离开时间的概念,可靠性就无从谈起。通常,人们会用平均无故障时间等时间尺度术语来描述系统的可靠性。

“规定的条件”也很重要。这些条件包括使用条件、维护条件、操作条件和环境条件等。对于同一系统,条件不同,会表现出不同的可靠性。如一台机床得到很好的维护保养,使用寿命可增加数倍;由于实验室的条件与现场的条件相差很大,因此,研究可靠性需要指定条件,即可靠性都是指在一定条件下的可靠性。

完成“规定的任务”,指系统能正常发挥其各项功能。如不能完成规定的任务,就称系统产生“故障”而“失效”。在实际应用中,有一系列的指标度量系统的“故障”和“失效”。

衡量“能力”或“满意工作的概率”有许多定量指标。由于系统的多样性,度量系统可靠性“能力”的指标也是多样的。常用的有“可靠度”、“失效率”、“平均寿命”、“平均无故障时间”等。

另外,可靠性与经济性又是紧密地联系在一起的,增加可靠性,必然要增加系统的生产成本或维护费用;另一方面,可靠性较高的产品,日常维护的费用较少,相应的由系统可靠性

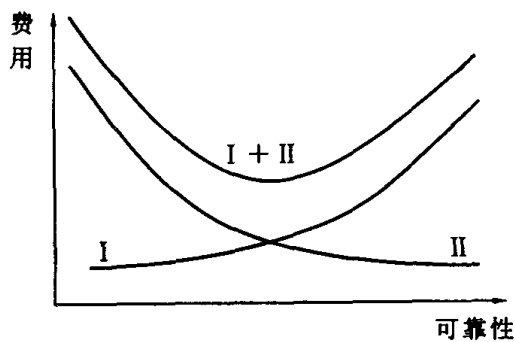


图 9-1 可靠性与经济性的关系

问题带来的损失也较少。因此,需要在这两个方面综合考虑。在图 9-1 中,曲线 I 表示,采购产品时,可靠性越高,价格越高;曲线 II 表示,可靠性越高的产品,使用中因可靠性问题带来的损失越小。曲线 I + II 是这两者的综合。

(2) 系统可靠性的研究内容

对于各种各样的系统,在系统或设备的试制、设计、生产、使用和维护的各个阶段,都存在可靠性问题,都有必要进行可靠性分析,系统的可靠性分析包

括以下几个方面的内容:

A 可靠性设计

在研制和设计一个新的系统时,要认真考虑系统的可靠性问题,合理分配系统各组成部分的可靠性指标,使系统在总体上效能优,费用省。

B 可靠性计算及预测

在系统设计阶段,对系统的每一个元件都要计算其可靠性指标,以利于各个不同设计方案之间的比较。

C 可靠性评估及管理

在系统的生产制造阶段,可靠性管理是非常重要的环节,只有通过严格的可靠性管理,才能保证产品的质量。广义的可靠性管理则包括产品从设计到使用的每个环节。在系统的运行使用阶段,需要按照可靠性的要求,保证系统的条件,监控系统的运行状况,评估系统的可靠性参数。

(3) 失效规律

任何元件及系统都可能在运行一段时间后出现故障,失去原来的效能。因此,研究系统的失效规律非常重要。

系统失效的原因有三个方面,即系统设计上的缺陷、系统运行维护时的不当及系统的自

然损耗和老化。

大量的实验和实践数据表明,系统失效具有明显的规律:即早期失效的可能性较高,中期失效的可能性较小,后期失效的可能性又增高,呈现如图 9-2 所示形状,俗称为浴盆曲线。

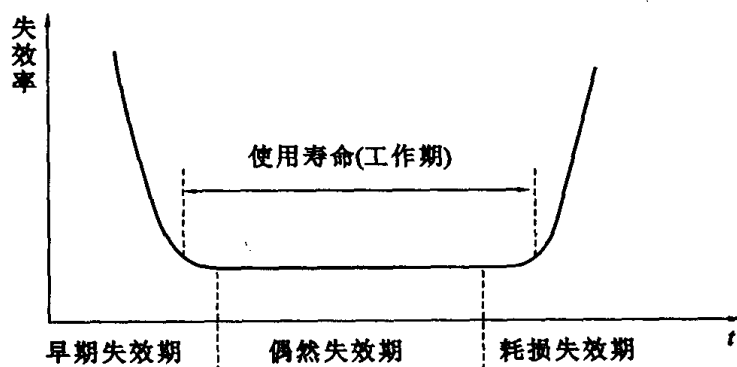


图 9-2 失效曲线

在系统进入稳定的工作期之前,是早期失效期。在这个阶段的开始部分,失效率较高,但失效率迅速下降。失效的主要原因是设计欠周、制造缺陷、使用不当和与环境的磨合等。

当失效率降低并进入稳定状况时,系统进入偶然失效期。其特点是失效率高、稳定且几乎与时间无关。此时的失效是偶然原因引起的,如工艺缺陷、材料弱点、维护失误、操作错误等因素。此时不可能通过延长磨合期和定期更换失效件来加以改善系统的状况。通常,电子设备的使用寿命比机械设备长。

每个系统和每种产品,当它的使用寿命快到时,失效率开始增加,这时就进入了耗损失效期。其特点是失效率随时间的增加而增加,其原因是系统的老化、疲劳和损耗。

(4) 系统的修复

产品或系统投入使用后,一旦失效,就不能修复或不值得修复,称为不可修复系统;如果失效后可修复,则称为可修复系统。可修复系统正常工作期称为工作时间(随机变量 T_i)与失效维修期称为维修时间(随机变量 τ_i)通常是交替出现的。本章内容,未加特别说明的均指不可修复系统。

9.1.2 可靠性的度量

(1) 可靠度

可靠度定义:系统到时刻 t 时无故障的概率称为系统的可靠度,记为 $R(t)$ 。如果用随机变量 T 来表示系统从开始工作到发生故障的连续正常工作时间,则可靠度 $R(t)$ 可表示为

$$R(t) = p\{T > t\} \quad (9.1)$$

相应地,系统从开始工作到 t 时刻内发生故障的概率称为系统的不可靠度,记为 $F(t)$ 。显然

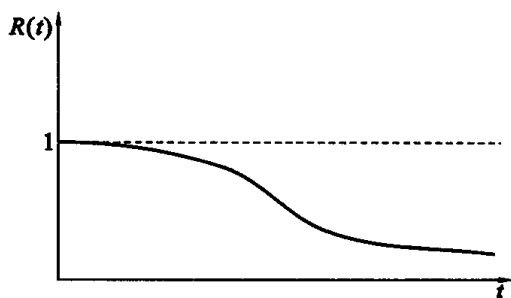
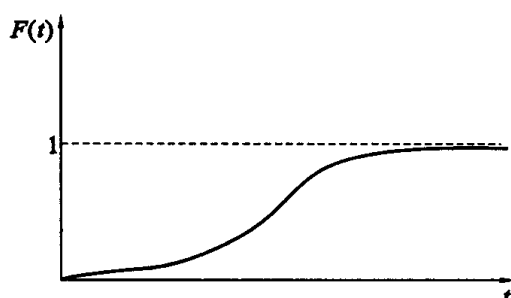
$$F(t) = 1 - R(t) \quad (9.2)$$

可靠度 $R(t)$ 是 t 的非增函数,在 $t = 0$ 时, $R(t) = 1$; 当 $t \rightarrow \infty$, $R(t) \rightarrow 0$, 见图 9-3。不可靠度 $F(t)$ 是 t 的非减函数,在 $t = 0$ 时, $F(t) = 0$; 当 $t \rightarrow \infty$, $F(t) \rightarrow 1$, 见图 9-4。 $f(t) = F'(t)$ 称为系统无故障工作时间的分布密度函数或失效密度函数。

例如,有大量的同类产品 (N_0 个),在 $t = 0$ 时刻开始使用或试验,设 $N_s(t)$ 为到 t 时刻仍正常工作的产品数量, $N_f(t)$ 为到 t 时刻已失效产品的数量,则有

$$N_0 = N_s(t) + N_f(t) \quad (9.3)$$

$$R(t) = \frac{N_s(t)}{N_0} \quad F(t) = \frac{N_f(t)}{N_0} \quad (9.4)$$

图 9-3 可靠度函数 $R(t)$ 图 9-4 不可靠度函数 $F(t)$

(2) 失效密度

设随机变量 T 是连续型的, 可定义失效密度函数 $f(t)$ 为

$$f(t) = \frac{dF(t)}{dt} = -R'(t) \quad (9.5)$$

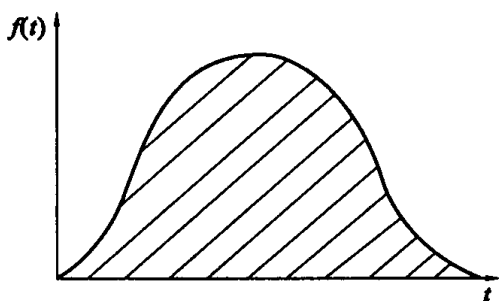


图 9-5 失效密度函数分布

图 9-5 表示了失效密度函数的分布情况, 显然, 图中的阴影面积等于 1。

对于上例中, 由单位时间 Δt 内故障的发生数 $\Delta n(t)$, 可得失效密度函数的近似值为

$$f(t) \approx \frac{\Delta n(t)}{N_0 \Delta t} \quad (9.6)$$

(3) 产品的寿命

对于同一种产品, 将每一个产品投入使用至发生故障的时间间隔 T 记下来, 其平均值就是平均工作

寿命。

A 平均寿命

设 $f(t)$ 是产品的失效密度函数, 则产品的平均寿命 m 就是随机变量 T 的数学期望值:

$$m = E(T) = \int_0^{\infty} t f(t) dt \quad (9.7)$$

它是标志产品平均能正常工作多长时间的可靠性指标之一。也称为平均寿命。

B 失效前平均时间 MTTF

对于可修复系统, 设首先发生失效的时间为 T_1 , 其分布为 $F_1(t)$, 失效密度为 $f_1(t)$, 这时的寿命分布与不可修复系统是完全一样的, 称为失效前平均时间 MTTF (Mean Time To Failure)。

$$\text{MTTF} = E(T_1) = \int_0^{\infty} t f_1(t) dt \quad (9.8)$$

C 平均无故障时间 MTBF

对于可修复系统, 修复后可以重新使用, 如假定各随机变量 T_i 和 τ_i 都是相互独立的, 修复后的性能与新的完全一样。则 T_i 都服从相同的寿命分布 $F(t)$ 。这时系统的平均无故障时间 MTBF (Mean Time Between Failures) 就等于失效前平均时间 MTTF。

D 可靠寿命

使系统可靠度 $R(t)$ 下降至给定值 r 所需的时间 t_r 称为系统的可靠寿命, r 称为可靠水平, 有

$$R(t_r) = r \quad (9.9)$$

另外, 还定义 $r = 0.5$ 时的可靠寿命为中位寿命; $r = e^{-1}$ 时的可靠寿命为特征寿命。

这三种寿命定义的意义是: 系统工作到可靠寿命 t_r 时, 失效的概率为 $100(1-r)\%$; 工作到中位寿命 $t_{0.5}$ 时, 失效的概率为 50% ; 工作到特征寿命 $t_{e^{-1}}$ 时, 失效的概率为 63.2% 。

(4) 失效率

对于已经正常工作到 t 时刻没有发生失误或故障的系统, 在接下来的单位时间内发生故障或失效的概率称为失效率, 记为 $\lambda(t)$

$$\lambda(t) = \frac{\text{在 } [t, t + \Delta t] \text{ 时间内单位时间失效的产品数}}{\text{在 } t \text{ 时刻仍正常工作的产品数}} \quad (9.10)$$

根据定义, $\lambda(t)$ 是一个条件概率,

$$\begin{aligned} \lambda(t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P\{\text{在 } (t, t + \Delta t) \text{ 故障} \mid t \text{ 时间正常}\} \\ &= \lim_{\Delta t} \frac{1}{\Delta t} P\{t < T \leq t + \Delta t \mid T > t\} \\ &= \frac{f(t)}{R(t)} = \frac{1}{R(t)} \cdot \frac{dF(t)}{dt} = -\frac{R'(t)}{R(t)} \end{aligned} \quad (9.11)$$

由式(9.11), 可靠度 $R(t)$ 可描述为

$$R(t) = \exp\left\{-\int_0^t \lambda(t) dt\right\} \quad (9.12)$$

当 $\lambda(t)$ 为常数 λ 时, 得

$$R(t) = e^{-\lambda t}$$

失效率 $\lambda(t)$ 的单位是每小时的百分数, 如

$$\lambda(t) \mid_{t=3000} = 4 \times 10^{-2} h^{-1}$$

表示系统工作 3000 小时后, 每小时出现故障的可能性是 4% , 或者说, 每 100 个产品中大约有 4 个会失效。

9.2 系统可靠性模型

9.2.1 逻辑框图

在研究一个系统特别是一个复杂系统的可靠性时, 必须了解系统各组成部分甚至每个元件的可靠性、各元件之间或各部分之间的相互关系。为此, 需要建立系统的可靠性模型, 逻辑框图就是建立系统可靠性模型的一种方法。

逻辑框图用一个方框表示一个系统、一个部分或者一个元件, 每个方框有一个输入信号和一个输出信号。根据方框的相互关系, 通常系统的可靠性模型有串联、并联和混联三种。构建一个系统的逻辑框图, 首先要确定系统完成预期任务的工作方式, 然后是决定顺利完成预期任务所需要的元部件, 接下来就是分析每个元部件的失效形式及其对预期任务的影响。

要注意系统的结构框图与系统的可靠性框图有时是不一致的。

9.2.2 串联系统可靠性模型

只有所有子系统都正常工作,系统才正常工作的系统,无论结构如何,其系统可靠性模型都是由多个方框以简单串联方式组成,这样的系统称为串联系统。图 9-6 表示电机系统由电源、控制开关、电机等单元组成,是一个简单串联系统,其功能是实现电机正常转动,可靠性模型框图见图 9-7。

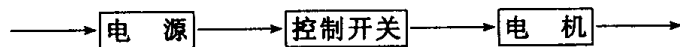


图 9-6 电机系统

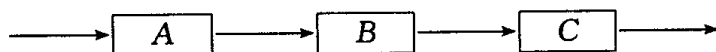


图 9-7 电机系统可靠性模型

但是如图 9-8 所示三个电容的并联结构系统中,考虑短路失效时,任何一个电容失效,系统失效,因此是一个串联系统,其可靠性模型如图 9-9 所示。

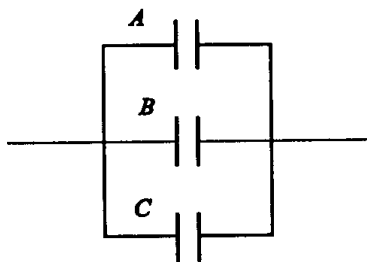


图 9-8 电容分布

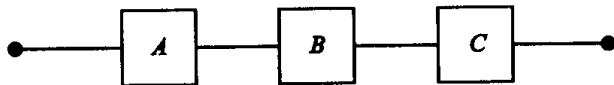


图 9-9 电容系统可靠性模型

为简化问题,设串联系统由 n 个单元组成,整体和各单元都只有正常和失效两种状态,各单元之间相互独立。

设:

A 为系统处于正常工作状态;

$\bar{A}$ 为系统处于失效状态;

A_i 为系统中第 i 个单元处于正常工作状态;

$\bar{A}_i$ 为系统中第 i 个单元处于失效状态;

$R_s(t)$ 为系统在 t 时刻正常工作的概率,即系统的可靠度;

$F_s(t)$ 为系统的不可靠度;

$R_i(t)$ 为第 i 个单元的可靠度;

$F_i(t)$ 为第 i 个单元的不可靠度;

$\lambda_i(t)$ 为第 i 个单元的失效率。

由于是串联系统,因此,只有当各个单元都正常工作时,系统才能正常工作,记为

$$A = A_1 A_2 \cdots A_n$$

又因为各单元相互独立,则有

$$R_s(t) = \prod_{i=1}^n R_i(t)$$

如果各单元的寿命都服从指数分布,且

$$R_i(t) = e^{-\lambda_i t}$$

则各单元的失效率为

$$\lambda_i(t) = -\frac{R'_i(t)}{R_i(t)} = \lambda_i$$

系统的可靠度为

$$R_s(t) = \prod_{i=1}^n R_i(t) = \prod_{i=1}^n e^{-\lambda_i t} = \exp\left(-\sum_{i=1}^n \lambda_i t\right) = e^{-\lambda_s t}$$

式中, $\lambda_s = \lambda_1 + \lambda_2 + \cdots + \lambda_n$, 是系统的失效率。

上式说明, 系统的寿命与各单元一样, 也服从指数分布。

系统的平均寿命 m_s 为

$$m_s = \int_0^{\infty} t f(t) dt = \int_0^{\infty} t [-R'_s(t)] dt = \int_0^{\infty} t \lambda_s e^{-\lambda_s t} dt = \frac{1}{\lambda_s} \quad (9.13)$$

系统的不可靠度为

$$F_s(t) = 1 - R_s(t) = 1 - e^{-\lambda_s t} = \lambda_s t - \frac{1}{2!}(-\lambda_s t)^2 - \frac{1}{3!}(-\lambda_s t)^3 - \cdots$$

如果 $\lambda_s t$ 较小, 如 < 0.1 , 则

$$F_s(t) \approx \lambda_s t = \sum_{i=1}^n \lambda_i t \approx \sum_{i=1}^n F_i(t)$$

可见, 这时的串联系统, 不可靠度近似等于各组成单元的不可靠度之和。

9.2.3 并联系统可靠性模型

对于由 n 个单元组成的系统, 如果只有当 n 个单元都失效时, 系统才失效, 该系统就是并联系统。记为

$$\bar{A} = \bar{A}_1 \cdot \bar{A}_2 \cdot \cdots \cdot \bar{A}_n$$

$$A = A_1 \cup A_2 \cup A_3 \cdots \cup A_n$$

如降落伞上的张伞切具, 一般都由 2 个或 2 个以上的切具并联组成, 以便一个切具失效时, 另一个切具仍可保证降落伞的打开; 多路电源供电系统等都是并联系统。如图 9-10 所示。

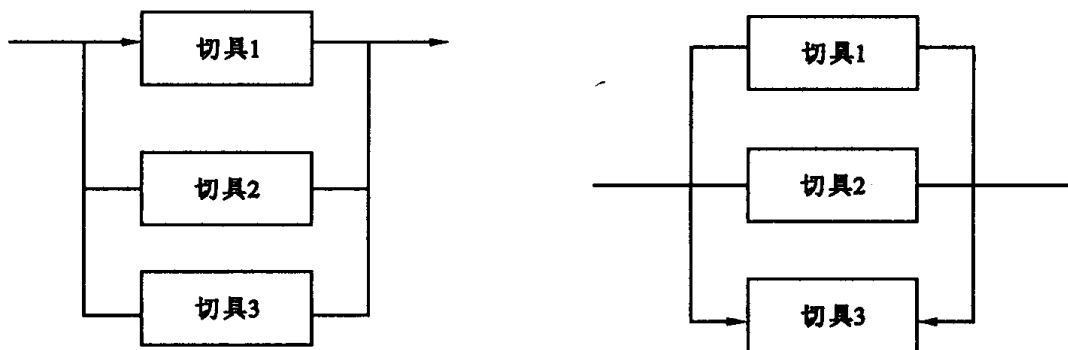


图 9-10 切具系统及其可靠性模型

有些系统结构上是串联的, 但可靠性逻辑框图模型中从功能上看是并联的。如图 9-11(a) 是由一个泵和两个抑制阀门组成的流体系统, 其功能是当泵不工作时, 阻止流体倒流。因此, 只要有任何一个阀门能正常工作即可。这两个结构上串联的阀门, 在系统可靠性模

型中是并联的,见图 9-11(b)。

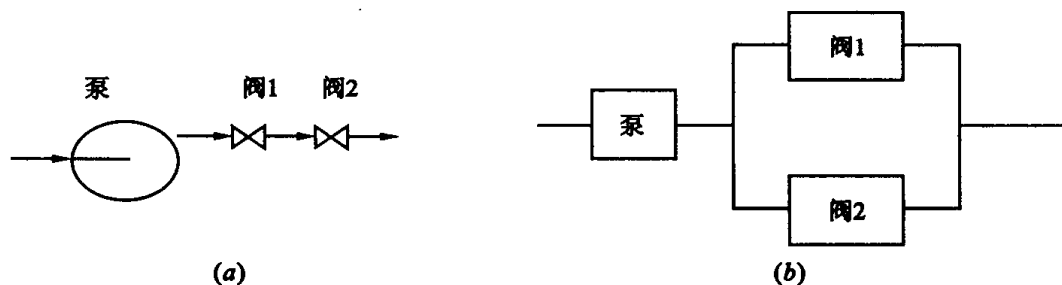


图 9-11 流体系统及其可靠性模型

仍然采用串联系统中设定的符号,并设组成系统的各单元相互独立。则系统的不可靠度 $F_s(t)$ 为

$$F_s(t) = \prod_{i=1}^n F_i(t)$$

设各单元的寿命均服从失效率为 λ_i 的指数分布,则可靠度 $R_s(t)$ 为

$$\begin{aligned} R_s(t) &= 1 - F_s(t) \\ &= 1 - \prod_{i=1}^n [1 - R_i(t)] = 1 - \prod_{i=1}^n (1 - e^{-\lambda_i t}) \end{aligned} \quad (9.14)$$

于是,可推导出系统的平均寿命为

$$m_s = \sum_{i=1}^n \frac{1}{\lambda} - \sum_{1 \leq i < j \leq n} \frac{1}{\lambda_i + \lambda_j} + \cdots + (-1)^{n-1} \frac{1}{\sum_{j=1}^n \lambda_j}$$

特别当 $n = 2$ 时

$$\begin{aligned} R_s(t) &= e^{-\lambda_1 t} + e^{-\lambda_2 t} - e^{-(\lambda_1 + \lambda_2)t} \\ m_s &= \frac{1}{\lambda_1} + \frac{1}{\lambda_2} - \frac{1}{\lambda_1 + \lambda_2} \end{aligned} \quad (9.15)$$

当各单元服从相同参数的指数分布时,即 $\lambda_1 = \lambda_2 = \cdots = \lambda_n = \lambda$

$$\begin{aligned} R_s(t) &= 1 - (1 - e^{-\lambda t})^n \\ m_s &= \sum_{i=1}^n \frac{1}{i\lambda} = \frac{1}{\lambda} + \frac{1}{2\lambda} + \cdots + \frac{1}{n\lambda} \end{aligned} \quad (9.16)$$

可见,并联系统中,随着 n 的增加,系统平均寿命得到增加。

9.2.4 混联系统的可靠性模型

串联系统中含有并联子系统,或并联系统中含有串联子系统的系统称为混联系统。如图 9-12 所示。

9.2.5 表决系统可靠性模型

在 n 个子系统中,至少有 k 个子系统正常,系统才能正常工作的系统,称为 $k/n(G)$ 表决系统。显然,串联系统就是 $n/n(G)$ 表决系统,并联系统就是 $1/n(G)$ 表决系统。 $k/n(G)$ 表决系统的逻辑框图见图 9-13。

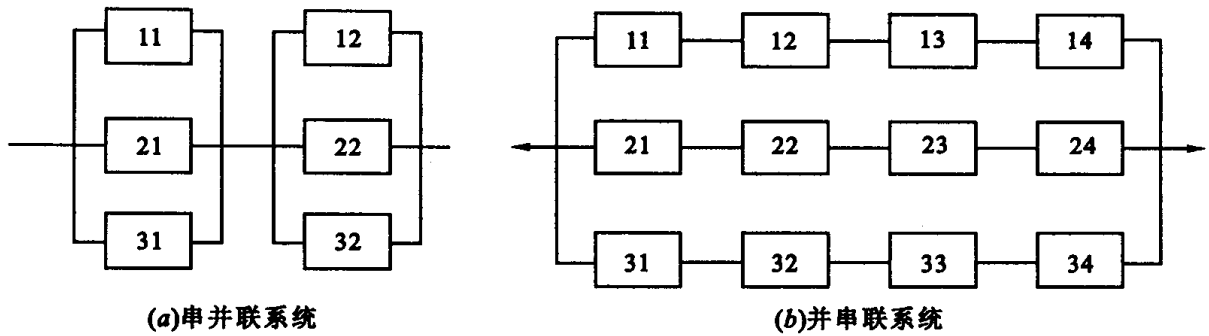


图 9-12 混联系统可靠性模型

设 n 个单元相互独立, 且各单元寿命分布均为指数分布, 可靠度均为 $R(t)$ 。

设 n 个系统中有 k 个正常的概率为

$$P\{n \text{ 个中有 } k \text{ 个正常}\} = C_n^k R^k (1-R)^{n-k}$$

可求得

$$R_{k/n}(t) = \sum_{i=k}^n C_n^i R^i (1-R)^{n-i}$$

当 $R(t) = e^{-\lambda t}$ 时, 有

$$R_{k/n}(t) = \sum_{i=k}^n C_n^i e^{-i\lambda t} (1 - e^{-\lambda t})^{n-i}$$

$$m_s = \sum_{i=k}^n \frac{1}{i\lambda} \quad (9.17)$$

当 $n = 3, k = 2$, 即为 $2/3(G)$ 表决系统, 有

$$m_s = \sum_{i=2}^3 \frac{1}{i\lambda} = \frac{1}{2\lambda} + \frac{1}{3\lambda} = \frac{5}{6\lambda} \quad (9.18)$$

9.2.6 储备系统可靠性模型

为了提高系统的可靠性, 可以另外储备一些具有相同功能的系统, 以便当工作中的系统失效时, 用一个备用系统替代。具有这种功能的系统叫做储备系统, 如图 9-14 所示。储备系统又分为冷储备和热储备两种。在冷储备系统的 n 个子系统中, 有一个是正常工作子系统, 另外 $n-1$ 个子系统作冷储备用, 当工作子系统失效时, 储备的子系统中的某一个就替代工作子系统的工作, 直至所有的子系统都失效, 系统才失效。设各储备子系统在储备期间失效率为零。

在热储备系统中, 储备子系统失效率不为零。

对于有 n 个子系统的冷储备系统, 设各子系统的寿命分别为随机变量 $T_1, T_2, \dots, T_n$, 平均寿命分别为 $m_1, m_2, \dots, m_n$, 则系统寿命的随机变量为

$$T_s = T_1 + T_2 + \dots + T_n$$

系统的可靠度为

$$R_s(t) = R\{T_1 + T_2 + \dots + T_n = T_s > t\} \quad (9.19)$$

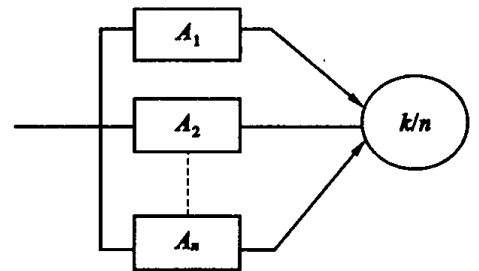
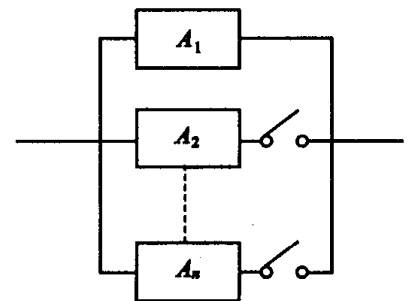
图 9-13 $k/n(G)$ 表决系统

图 9-14 储备系统

由于各子系统相互独立,所以系统的平均寿命为

$$m_s = m_1 + m_2 + \cdots + m_n \quad (9.20)$$

特别当 $n = 2$ 时,且设各子系统服从指数寿命分布,其失效率分别为 λ_1 和 λ_2 ,则

$$f_s(t) = \lambda_1 \lambda_2 \frac{e^{-\lambda_1 t} - e^{-\lambda_2 t}}{\lambda_2 - \lambda_1}$$

$$R_s(t) = \frac{\lambda_2}{\lambda_2 - \lambda_1} e^{-\lambda_1 t} + \frac{\lambda_1}{\lambda_1 - \lambda_2} e^{-\lambda_2 t}$$

平均寿命为

$$m_s = \frac{1}{\lambda_1} + \frac{1}{\lambda_2}$$

更进一步,如 $\lambda_1 = \lambda_2 = \lambda$,则

$$f_s(t) = \lambda^2 t e^{-\lambda t}$$

$$R_s(t) = (1 + \lambda t) e^{-\lambda t}$$

$$m_s = \frac{2}{\lambda}$$

9.2.7 桥式网络系统模型

有些复杂系统,不能用串联或并联系统可靠性模型来描述。如图 9-15 所示系统是一个桥式网络系统,可以用桥式开关网络来建立桥式系统的可靠性模型,如图 9-16 所示。

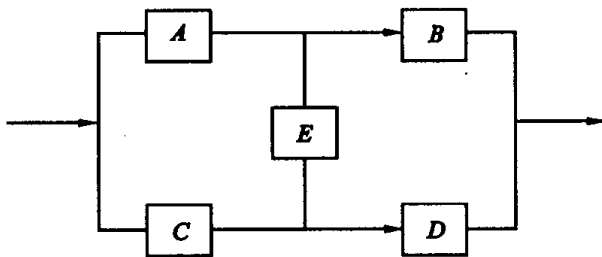


图 9-15 桥式系统

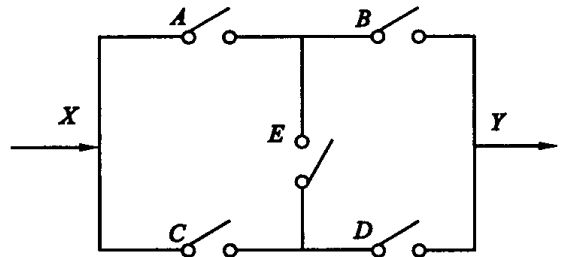


图 9-16 桥式系统的可靠性模型

图 9-16 中,每个开关对应于桥式系统中相应位置的子系统,开关闭合的概率就是相应子系统的可靠度。也就相当于信息从 X 端传递到 Y 端的概率。

该桥式开关网络系统共有 32 种状态,其中有 16 种状态可以将信息从 X 端传递到 Y 端。可整理成表格进行计算,称为布尔代数真值表示法,也叫做状态枚举法。

设该系统的每个开关的可靠度均为 0.9,其中一个状态为 A 和 B 闭合,C、D、E 断开,此时可传递信息,该状态的概率为

$$\begin{aligned} P\{A、B \text{ 通}, C、D、E \text{ 闭}\} &= P_A \cdot P_B \cdot \bar{P}_C \cdot \bar{P}_D \cdot \bar{P}_E \\ &= (0.9)^2 \times (1 - 0.9)^3 \\ &= 0.00081 \end{aligned}$$

系统中共有 16 个可形成通路的状态,其概率之和为 0.97848,即相应的桥式系统的可靠度为 0.97848。

9.3 可靠性计算及预估

可靠性计算及预估的任务就是对系统、每个子系统、每一个子系统的每一个分系统直至每个元件,进行可靠性指标计算,用于评估系统的性能,优化系统的设计。前面介绍的配合可靠性模型分析的概率分析法,当然是有用的,但对于实际的系统,常常显得太繁琐,甚至不可能使用。

可靠性预估又分为单元可靠性预估和系统可靠性预估。有时还要分别计算实验室条件下的基本失效率、现场使用条件下的应用失效率及在使用条件下完成某一项任务的任务失效率。

下面介绍几种比较实用的可靠性计算和预估的方法。

9.3.1 事件树分析法

事件树分析法(Event Tree Analysis)简称ETA法,它是一种利用树形图进行自下而上的顺向分析的方法。它从故障的原因事件开始,按时间进程分析可能导致的各种事件序列,以便定性和定量地评价系统的可靠性。

- ETA法的主要内容有:
- A 确定或寻找可能导致故障的原因事件并进行合理分类;
 - B 建立事件树并进行合理简化;
 - C 定量计算系统的失效率和各种事件的概率。

如图9-17是一个煤气报警连锁系统的事件树,当室内煤气泄漏时,传感器检测到并送出报警连锁信号,该信号用于启动排风装置并关闭煤气。煤气泄漏后,煤气关闭,排风启动是成功的状态(S);煤气关闭和排风启动两件事中只完成一件的是部分成功状态(P);煤气未关闭且排风未启动是失效状态(F),信号未检测到和信号未传好都是失效状态(F),⊗表示系统失败。这三种状态的发生概率分别为

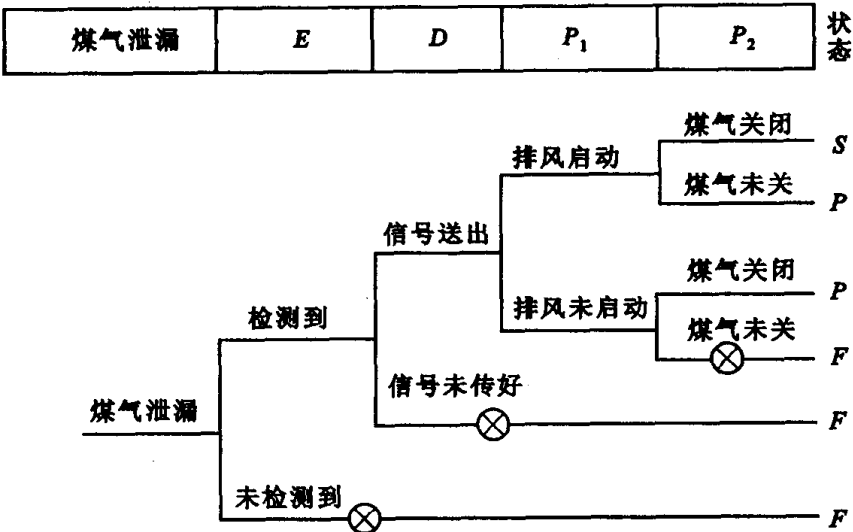


图 9-17 煤气报警连锁系统的事件树

$$P(S) = P(E)P(D)P(P_1)P(P_2)$$

$$P(P) = P(E)P(D)P(P_1)P(\bar{P}_2) + P(E)P(D)P(\bar{P}_1)P(P_2)$$

$$P(F) = P(E)P(D)P(\bar{P}_1)P(\bar{P}_2) + P(E)P(\bar{D}) + P(\bar{E})$$

如果每个事件的发生概率均为 0.988, 则

$$P(S) = 0.988^4 \approx 0.952857$$

$$P(P) = 0.988^3(1 - 0.988) \times 2 \approx 0.023147$$

$$P(F) = 0.988^2(1 - 0.988)^2 + 0.988(1 - 0.988) + (1 - 0.988) \\ \approx 0.023996$$

EAT 法比较直观, 能有效地找出事故发生的种种可能原因。

9.3.2 故障树分析法

故障树分析法(Fault Tree Analysis)简称 FTA 法, 它是一种从结果开始, 自上而下逆向分析的逻辑分析方法。从故障的结果, 即顶事件开始, FTA 把不希望发生的事件(失效状态)作为故障树的顶事件, 并用规定的逻辑符号表示。通过分析故障的原因, 逐步深入, 直至找出故障树的底事件, 底事件也称为基本事件。其特点是逻辑严密, 表达直观, 虽然它基本上是一种定性的、逻辑性的分析方法, 但如果能给出基本事件的概率值, 也能够作出定量的预估。FTA 一般的实施步骤如下:

(1) 熟悉对象, 合理选择顶事件。通常顶事件不宜太多, 但要确切表达。

(2) 故障树的建立与简化。通过对所收集资料的分析及与设计管理人员的沟通, 构建故障树并做适当的简化。其特点是事件和逻辑关系都采用一定的符号来表示。

(3) 定性分析。求取最小割集和最小路集。

最小割集的定义是: 设故障树中有 n 个基本事件 $X_1, X_2, \dots, X_n, C = \{X_{i1}, X_{i2}, \dots, X_{in}\}$ 为部分基本事件的集合。 C 中的全部事件都发生时, 顶事件必然发生, 则称 C 是故障树的一个割集。如果 C 是一个割集, 在 C 中任意去掉一个事件, C 就不是故障树的割集, 则 C 是故障树的一个最小割集。

最小路集的定义是: 设 $D = \{X_{i1}, X_{i2}, \dots, X_{in}\}$ 是部分基本事件的集合。当 D 中每一个事件都不发生时, 顶事件才不发生, 则 D 是故障树的一个路集。如果 D 是一个路集, 在 D 中任意去掉一个事件, 它就不再是路集, 则 D 是故障树的一个最小路集。

(4) 定量分析。计算顶事件的发生概率, 评价各基本事件对顶事件影响的重要性, 进行重要度分析和灵敏度分析。

如图 9-18 所示是某系统的故障树, 逻辑门图形中的“+”表示逻辑关系“或”, “·”表示逻辑关系“与”, X_1, X_2, X_3, X_4, X_5 是基本事件, G_1, G_2, G_3, G_4, G_5 为逻辑门, T 是顶事件。根据最小割集和最小路集的定义, $(X_1), (X_2, X_3), (X_4, X_5)$ 是最小割集; $(X_1, X_2, X_4), (X_1, X_3, X_4), (X_1, X_2, X_5), (X_1, X_3, X_5)$ 是最小路集。

9.3.3 上、下限法

上、下限法也称边值法或界限法。它根据系统各单元的可靠度和失效率, 逐步计算出整个系统越来越精确的可靠度上限和下限, 然后用适当的公式, 计算整个系统可靠度预测值。

在计算上限和下限时, 经常需要根据情况进行一些必要的简化处理和假设, 才能估算出

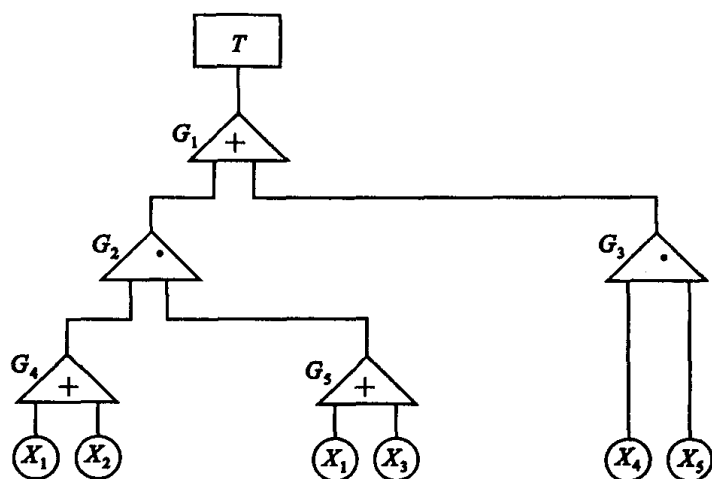


图 9-18 故障树

系统的可靠性上限和下限。

当求出的系统可靠度满足实际精度要求上限和下限后,可采用下式计算系统的可靠度 R_s 。

$$R_s = 1 - \sqrt{(1 - R_{um})(1 - R_{lm})}$$

式中, R_{um} 和 R_{lm} 分别是系统可靠度的上限和下限预估值。

另外,可靠度预估的方法还有数学模型法和蒙特卡罗法。数学模型法是一种根据系统可靠性模型逻辑框图,估算系统可靠性指标的方法。其优点是计算比较精确,缺点是比较繁琐,特别是对于复杂系统计算比较困难。蒙特卡罗法也称仿真法,它通过建立系统的模型来预估系统的可靠性,其特点是可以方便地用计算机进行仿真。

9.4 可靠性设计与分配

可靠性设计的基本任务是根据对系统可靠性要求指标,在最低费用的基础上,综合考虑系统的性能、费用和可靠性,尽量提高系统的固有可靠性。

可靠性分配则是具体地把系统整体的可靠性要求指标合理地分配到各个子系统,它是可靠性设计的重要环节。一个合理的可靠性设计任务需要经过多次的预估、分配、再预估、再分配,才能完成。

下面介绍几种常用的可靠性设计和可靠性分配方法。

9.4.1 代数分配法

(1) 等分配法

等分配法将可靠度相等地分配给各子系统,是一种最简单的可靠度分配法。

设 R_s 是系统的可靠度要求, R_i 是分配给每个子系统的可靠度要求,共有 n 个子系统,则对于串联系统有

$$R_i = R_s^{\frac{1}{n}} \quad i = 1, 2, \dots, n \quad (9.21)$$

对于并联系统有

$$R_i = 1 - (1 - R)^{\frac{1}{n}} \quad i = 1, 2, \dots, n \quad (9.22)$$

(2) 比例分配法

即考虑各子系统现有可靠度状况,按不可靠度成比例地分配不可靠度。即以老系统的状况为参考,以新系统的可靠度要求给新系统的各单元部件分配失效率。

对于串联系统

$$\begin{aligned} \text{可靠度} \quad R &= R_1 R_2 \cdots R_n = (1 - F_1)(1 - F_2) \cdots (1 - F_n) \\ \text{不可靠度} \quad F &= 1 - R \approx F_1 + F_2 + \cdots + F_n \end{aligned} \quad (9.23)$$

根据近似公式(9.23),设计要求系统总的不可靠度调整多少比例,各子系统也调整多少比例,再推得各子系统的可靠度。调整比例按下列约定计算。

$$\text{现有不可靠度 } F_x = \sum_{i=1}^n \text{现有 } F_i$$

$$\text{设计允许的不可靠度 } F_r = 1 - R$$

设有一 $n = 3$ 的串联系统,要求可靠度 $R = 0.96$, 3 个子系统现有可靠度分别为 0.98, 0.96, 0.94, 现按比例分配法推导各子系统可靠度的合理分配值如下:

系统现有可靠度为

$$R = R_1 R_2 R_3 = 0.884352$$

而要求的可靠度为

$$R = 0.96$$

则

$$F_r = 1 - R = 0.04$$

$$F_x = F_1 + F_2 + F_3 = 0.02 + 0.04 + 0.06 = 0.12$$

$$\frac{F_r}{F_x} = \frac{1}{3}$$

则要求

$$R_1 = 1 - \frac{1}{3}F_1 = 0.9933, R_2 \approx 0.9867, R_3 = 0.98$$

如此得

$$R = R_1 R_2 R_3 = 0.964087$$

结果略大于设计要求的 0.96, 方案可行。

9.4.2 加权分配法

加权分配法考虑各子系统对于系统的重要程度,根据子系统发生故障后引起整个系统故障的概率,来分配各子系统的可靠度。

$$\text{定义} \quad W_i = \frac{\text{由于第 } i \text{ 个子系统失效引起的系统故障次数}}{\text{第 } i \text{ 个子系统的总失效次数}} \quad (9.24)$$

为第 i 个子系统的加权因子,也称为重要因子或重要度。

设系统由 n 个子系统串联组成,各子系统的可靠度都服从指数分布。如第 i 个子系统 A_i 的可靠度为

$$R_i = e^{-\lambda_i t_i}$$

第 i 个子系统的预估可靠度为

$$\hat{R}_i = 1 - W_i(1 - R_i)$$

则整个系统的可靠度为

$$\begin{aligned} R_s &= \prod_{i=1}^n \hat{R}_i = \prod_{i=1}^n [1 - W_i(1 - R_i)] \\ &\approx \prod_{i=1}^n (1 - W_i F_i) \approx \prod_{i=1}^n e^{-W_i F_i} \\ &= \exp\left(-\sum_{i=1}^n W_i F_i\right) \end{aligned} \quad (9.25)$$

如要求系统的可靠度指标为 R , 则按等分配法分配到各子系统的可靠度为

$$R_i = (R)^{\frac{1}{n}}$$

考虑加权因子 W 时, 求得

$$\begin{aligned} R_i &= (R)^{\frac{1}{n}} = e^{-W_i F_i} \\ F_i &= -\frac{1}{n} \frac{\ln R}{W_i} \end{aligned} \quad (9.26)$$

如果子系统又是由 m_i 个单元所组成, 系统总单元数为

$$N = \sum_{i=1}^n m_i$$

则

$$F_i = -\frac{m_i}{N} \frac{\ln R}{W_i} \quad i = 1, 2, \dots, n \quad (9.27)$$

式中 m_i/N ——系统的复杂因子。

9.4.3 动态规划法

动态规划是解决多阶段过程最优化问题的有效方法。设系统由 n 个子系统串联组成, 为了提高系统的可靠度, 采用冗余技术, 将子系统设计为并联系统, 于是整个系统成为串并联复合系统。规划的目的就是确定采用冗余技术时, 各系统中的并联元件数, 使系统符合约束条件, 且可靠度最高。

设第 i 个子系统中有 u_i 个并联元件, 其可靠度为 $R_i(u_i)$ 。单个并联元件的成本为 c_i , 重量为 w_i 。系统的总费用不超过 C , 总重量不超过 W 。如此构成一个非线性整数规划问题:

$$\max R = \prod_{i=1}^n R_i(u_i) \quad (9.28)$$

约束条件为

$$\begin{aligned} \sum_{i=1}^n c_i u_i &\leq C \\ \sum_{i=1}^n w_i u_i &\leq W \end{aligned} \quad u_i \geq 0 \text{ 为整数, 且 } i = 1, 2, \dots, n$$

该问题求解很困难, 但用动态规划方法求解则较方便。设:

状态变量 x_k 表示从子系统 k 到 n 所允许具有的费用之和, 则 $x_1 = C$;

状态变量 y_k 表示从子系统 k 到 n 所具有的重量之和, 则 $y_1 = W$;

决策变量 u_k 为子系统 k 中的并联元件数; k 是表示子系统序号的阶段变量。

于是可得状态转移方程为

$$\begin{cases} x_{k+1} = x_k - c_k u_k \\ y_{k+1} = y_k - w_k u_k \end{cases} \quad (9.29)$$

允许决策集合为

$$D_k(x_k, y_k) = \{u_k \mid 0 \leq u_k \leq \min(\frac{x_k}{c_k}, \frac{y_k}{w_k}), u_k \text{ 为整数}\}$$

目标函数为

$$J_k = \prod_{i=k}^n R_i(u_i) \quad (9.30)$$

求 J_k 的极大值。递推方程为

$$\begin{aligned} J_{n+1}^* &= 1 \\ J_k^* &= \max_{u_k \in D_k} \{R_k(u_k), J_{k+1}^*\} \quad k = 1, 2, \dots, n \end{aligned} \quad (9.31)$$

也可以将问题转换为,把可靠度作为约束条件之一,而把某个约束条件作为目标函数,如总费用最低或总收益最高等。

9.5 可靠性工程与管理

一个产品从计划开始,到设计、生产、使用、维修直至报废,要经过若干个阶段和多个部门,其中任何一个环节出问题,都可能导致严重的后果。因此,可靠性管理在可靠性工程中起着枢纽的决定作用。要提高系统或产品的可靠性,必须严格按照可靠性工程与管理的系统工程方法开展工作。

可靠性管理是一个从不完善到逐步完善的过程,通过可靠性管理、监督并保证可靠性技术的实施,提高产品的质量水平。

9.5.1 提高可靠性的技术措施

(1) 精心设计

应对系统的可靠性进行全面的设计和分配,逐步优化,使系统的可靠性得到最好的改善。

(2) 优选元器件

这个环节也很重要,我国在军工领域提出了“七专”,使用“七专”元器件使系统的可靠性提高了一个数量级,这样做还缩短了调试周期,降低了成本。“七专”指:专批、专技、专人、专机、专料、专检、专卡。

(3) 采用冗余技术

如并联系统、表决系统和储备系统等。

(4) 降额使用

即让部分工作条件指标低于额定工作条件指标,在一定的范围内,可以使可靠度大幅度提高。

(5) 改善环境条件

使系统的震动环境、空气环境、温度、压力等环境条件得以改善,也可在一定程度上提高可靠性。

(6) 适当的维修措施

定期预防性维修或定期更换策略;失效更换或随机更换措施;以及保证系统冗余部分保持良好状态的措施。

9.5.2 提高人一机系统的可靠性

在系统的可靠性工程与管理中,要特别重视人的作用,应把人看作为系统的一个组成部分,从人一机系统的角度改善系统的可靠性。

人的工作可能有差错,人工作时间长了肯定会疲劳,不同的工作环境和条件,人的疲劳状况不同。

研究人在什么环境条件下能更好地发挥作用的学科,叫做“工程心理学”。在进行人机系统设计时,研究工作人员如何准确可靠地执行操作任务的学科,叫做“人机工程学”。

操作人员的工作疲劳受到各种因素的影响,如工作制度、工作环境、工作特点、生活质量、生理素质等等。人疲劳了,更容易出差错。人员可靠性分析中著名的 Rook 模型就是用来计算人员操作差错发生概率的模型。

9.5.3 可靠性管理

可靠性管理范围广、层次多、跨度大,要落实到位,难度很大。按管理职能、内容和要求,有这样三个层次的可靠性管理工作:

(1) 国家与政府级可靠性管理

国家与政府级可靠性管理的内容包括:

制定可靠性管理法规和政策、发展规划和计划;

制定可靠性标准和规范等;

实施可靠性认证制度;

组织可靠性宣传、教育培训与技术交流等。

(2) 企业级可靠性管理

企业级可靠性管理的内容包括:

建立企业可靠性管理体制和可靠性目标;

制定可靠性管理计划和可靠性目标、进行可靠性设计;

制定元器件、原材料采购计划;

开展可靠性宣传与教育;

开展可靠性监督与评审等。

(3) 产品级可靠性管理

产品级可靠性管理的内容包括:

市场调查和用户调查;

可靠性设计与分配;

合理选择元器件;

可靠性评审;

可靠性数据的收集等。

习 题

9.1 提高系统可靠性的重要意义是什么？你在工作与生活中遇到过哪些系统可靠性问题？

9.2 系统可靠性的研究有哪些内容？

9.3 度量系统的可靠性有哪些指标？

9.4 试考察人的身体中有哪些结构是为了提高人类生命的可靠性的？

9.5 如果系统的失效密度函数 $f(t)$ 为 Weibull 分布, 即

$$f(t) = \lambda a t^{a-1} e^{-\lambda t^a}$$

试求系统的可靠度 $R(t)$ 。

9.6 列举几个具有不同类型可靠性模型的实际系统, 并分别建立其可靠性模型。

9.7 什么是系统的冷储备和热储备？它们各有何优缺点？

9.8 设某系统由三个相同的子系统组成, 至少要有两个子系统正常工作才能保证系统的正常运行, 如果子系统的 MTBF 为 3.5×10^4 小时, 试计算该系统 100 小时正常运行的可靠度和 10000 小时正常运行的可靠度。

9. 求如图 9-19 混联系统的可靠度。

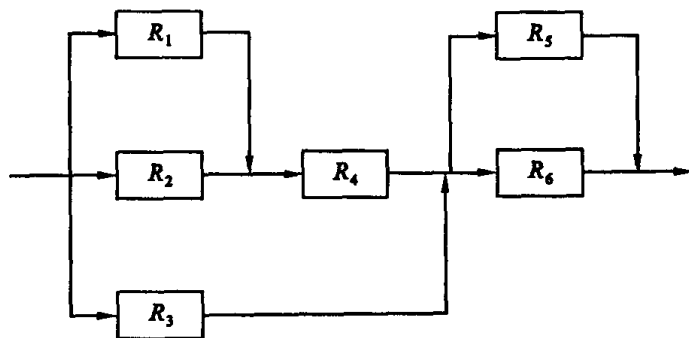


图 9-19

9.10 试对某个建筑物进行火灾的事件树分析。

9.11 为加强安全, 将某重要岗位由一人负责调整为两人共同负责, 问此时可靠性是否一定能得到提高。请举例加以说明。

10 工业技术经济

本章提要

工业技术经济通过对技术政策、技术方案、技术措施的经济效果进行计算、分析、比较和评价,从而选择技术上先进、经济上合理的最优化方法。本章通过对项目投资的静态分析法和动态分析法的介绍,提供一种定量的方法对技术方案进行经济评价,从而构成了技术经济理论研究的主要内容。

现代工业是一个国家重要的物质基础,也是一个区域和全球经济发展的支柱之一。在信息产业高度发展的今天,现代工业已经不再把某一生产过程作为其存在的主要形式,取而代之的是以先进的科学技术为基础,利用具有良好的性能和广泛的用途,物美价廉、经久耐用的新产品、新材料和新设备,极大限度地节约制造和使用过程中的社会劳动,降低产品成本。因而也就提出需要根据市场需要,对产品的设计、制造、使用过程中的诸多要素进行综合分析和论证,使每一项新技术能够在保证使用功能和质量的基础上、尽可能的减少社会劳动消耗;在一定物质消耗和社会劳动消耗的条件下,在保证合理的基础上尽可能提高新技术的功能和质量。这样的研究和分析,综合了经济学、运筹学、优化理论、网络分析和决策分析的知识,可以称为技术经济分析。研究技术经济分析理论与方法的科学,就是技术经济学。

技术经济学是一门连接系统工程和管理学科的新兴综合性学科。是一门介于自然科学和社会科学之间的交叉学科。它是对为达到某种预定目的而可能被采用的各种不同技术政策、技术方案、技术措施的经济效果进行计算、分析、比较和评价,从而选择技术上先进、经济上合理的最优化方案的科学。

技术经济学主要研究的内容归纳起来包括:① 技术经济分析与评价的理论和方法;② 新建项目、技术改造及设备更新的技术经济分析;③ 各种投资项目的技术经济分析;④ 可行性研究与费用—效益分析;⑤ 敏感度分析和不确定性分析,等等。

技术经济学的研究方法主要有两种:第一种方法是系统分析法。这种研究方法虽然在其他研究工作中都能采用,但对技术经济学研究来说有着特别重要的意义。因为技术经济学的核心内容是对各种技术方案的效果进行计算、分析、评价,并在多个可能方案中评选出较优的方案,因而需要各种技术经济的基本原始资料和数据。同时,技术经济学面临的各种问题与社会生产和环境有着密切的关系,这些很难用精确的数学模型和专业实验室内通过实验来取得和解决。因此,调查研究是技术经济学研究工作中必不可少的重要组成部分,是技术经济学研究的重要方法。技术经济学研究的第二种方法是理论研究。它在技术经济学中具有特别重要的意义。因为技术经济学的基本任务就是要寻找解决在生产实践中技术和经济之间普遍存在的矛盾,找出技术和经济之间的合理关系。技术经济学需要运用定量的方法对技术方案进行经济评价。因此研究评价的指标和指标的计算方法也就构成了技术经济理论研究的主要内容。

技术经济研究的工作包含以下几个步骤:

(1) 建立各种可行的技术方案。在全面掌握技术与经济信息和数据的基础上,利用各种专门知识,尽可能的建立各种客观上能够存在的方案。

(2) 分析各种可能的技术方案的优缺点及其影响因素。在调查研究的基础上,从技术、经济、社会、环境等方面对各个技术方案所产生的效果和影响,进行客观的、全面的分析,这是技术经济研究工作中十分重要的一步。

(3) 建立评价的指标和指标体系。评价技术方案的经济效果,需要一定的指标和依据,这就是一般所说的指标。但是任何一个指标只能反映经济效果的一个侧面,要对技术方案的经济效果作出客观全面的评价,需要建立一组指标,从不同角度进行多方面的评价,才能找出整体最优的方案。这些指标构成优化过程的约束集合,形成一个体系,即指标体系。这一指标体系,通常包括生产成果、生产消耗、生产成果与生产消耗的对比三个方面的评价指标。

(4) 评价指标的计算。求解评价指标一般采用两种不同的方法,一种是列表和图解的方法,另一种是数学分析的方法。后一种方法只有当评价指标和参数的函数关系能够完全用数学公式表达的时候,才能采用。

(5) 技术方案的综合技术经济评价。要从技术、经济、社会、生态等方面综合进行分析、论证和评价,才能选出经济上合理有效的技术方案。

以上是技术经济理论研究的一般基本程序,实际上,技术经济的研究方案和程序,应该在实践过程中不断发展和完善。

10.1 项目投资效果静态分析法

对项目投资效果进行经济评价的方法,有静态分析法和动态分析法。动态分析法考虑资金的时间价值,而静态分析法则不考虑资金的时间价值。静态分析法计算比较简单,对若干方案进行粗略评价,或对短期投资项目作经济分析时,此方法简易实用。目前国内外多采用投资收益率法、投资回收期法、追加投资效果评价法、最小费用法等对项目的投资效果进行静态分析。

10.1.1 投资收益率法

投资收益率(或称投资回收率、投资效果系数)是工程项目投产后获得的年净收入与项目总投资额之比。

计算公式:

$$E = \frac{M}{K} \quad (10.1)$$

式中 E ——投资收益率;
 K ——项目的总投资额;
 M ——年净收入。

在计算投资收益率时,由于人们对年净收入和总投资这两个术语有不同的理解,或由于目的不同,对投资收益率有不同的计算方法。对于年净收入,有的以毛利作为年净收入,有的

以净利作为年净收入,有的以净利加折旧费作为年净收入。对于项目的总投资额,有的以固定投资加生产前资金费用和流动资金作为总投资,有的则仅以固定投资作为总投资。因此,同一工程项目,利用各种不同的方法计算出来的投资收益率值是不相同的。

在我国,计算工程项目的投资收益率,可以根据出发点的不同而采用不同的计算方法。从国家的角度出发,企业利润、折旧费和税金之和为年净收入。可按下式计算投资收益

$$\text{投资收益率} = \frac{\text{企业利润} + \text{折旧费} + \text{税金}}{\text{总投资}} \quad (10.2)$$

从企业的角度出发,年净收入只包括企业利润和折旧费。所以投资收益率应按下式计算:

$$\text{投资收益率} = \frac{\text{企业利润} + \text{折旧费}}{\text{总投资}} \quad (10.3)$$

按我国的惯例,折旧费的一部分要上缴,另一部分则由企业留作设备更新改造资金。因此,我国常采用下列两个与投资收益率相类似的指标来估算经济效果:

$$\text{投资利税率} = \frac{\text{企业利润} + \text{税金}}{\text{总投资}} \quad (10.4)$$

$$\text{投资利润率} = \frac{\text{企业利润}}{\text{总投资}} \quad (10.5)$$

投资利税率是从国家的角度出发的衡量指标,而投资利润率则是从企业的角度出发的。以上各式中的总投资,必须包括总固定资产投资和流动资金。

10.1.2 投资回收期

投资回收期也称为投资返本期或投资偿还年限。投资回收期就是一个工程项目从开始投资算起或从开始投入生产算起达到全部投资回收时所经历的时间,通常以年数表示。可以按达产年净收入或按累计净收入进行计算。

(1) 按达产年净收入计算

它是指项目投产后,以达到设计产量后的第一个整年所获得的净收入,来计算收回该项目全部投资所需年数。

计算公式:

$$T = \frac{K}{M} \quad (10.6)$$

式中 T ——投资回收期。

若从投入生产开始算起,则投资回收期正好是投资收益率的倒数。与计算投资收益率一样,若从国家的角度出发,年净收入为企业利润、折旧费和税金之和;若从企业的角度出发,则年净收入为企业利润与折旧费之和。

(2) 按累计净收入计算

它是指项目正式投产之日起,累计提供的净收入的积累总额达到投资总额之日止所需的时间。如表 10-1 所示,从正式投产之日算起,总投资额 1 亿元可在 5 年内全部回收。因此,投资回收期则为 5 年。

表 10-1 每年投资回收情况

单位:万元

年 份		回收金额	年末未收回投资余额累计
建 设 期	第一年	- 3000	3000
	第二年	- 5000	8000
	第三年	- 2000	10000
生 产 期	第四年	1000	9000
	第五年	1000	8000
	第六年	2000	6000
	第七年	3000	3000
	第八年	3000	0

在国外计算投资回收期,一般从开始投资的建设期算起。若这样计算,表 10-1 所列举的投资回收期应为 8 年。

投资回收期从何时算起,除了以上不包括建设期和包括全部建设期这两种计算方法外。还有一种包括部分建设期的计算方法。这种计算方法把全部投资看作一次投入。以建设期末为零点,用求重心的方法求得投资投入的时间。

表 10-1 所列的投资时间计算方法如下:

$$\frac{3000 \times 2.5 + 5000 \times 1.5 + 2000 \times 1.5}{3000 + 5000 + 2000} = 1.6 \text{ 年} \approx 1 \text{ 年 } 7 \text{ 个月}$$

投资投入时间离零点为 1 年 7 个月,若从建设期初算起为 1 年 5 个月。因此表 10-1 所列举的投资时间应定在第二年的第五个月。

包括建设期投资占用因素的投资回收期应为 6 年又 7 个月(1.6 年 + 5 年 = 6.6 年)。

10.1.3 投资效果的评价依据

当项目的投资收益率越大,或者说投资回收期越短,项目的经济效果就越好。那么投资收益率应该多大,投资回收期应该多长时,所选择的项目投资方案才算是比较经济合理呢?这就需要有一个统一的尺度来进行衡量,也就是必须确定标准投资收益率 E_0 和标准投资回收期 T_0 。在国内此统一的尺度应根据目前各部门、各行业反映一般水平的实际投资收益率或实际投资回收期作为基数,根据国家的需要、技术水平、人力、财力、物力等因素综合考虑。

如果考虑利息率时,标准投资收益率必须大于利息率或该行业的平均投资收益率。标准投资回收期不能大于固定资产的折旧年限,否则就没有促进技术发展的作用了。一般情况下,重工业部门的标准投资回收期可定得长些或标准投资收益率定得小些,而轻工业部门则与其相反。按国外经验,一般来说,工业部门的标准投资回收期不大于 3 ~ 7 年,即标准投资收益率不小于 0.15 ~ 0.3。对一些部门,如交通运输和动力工业等部门,标准投资回收期可规定较长些,不超过 10 年,或标准投资收益率不小于 0.1。

项目投资方案的投资回收期应小于或等于标准投资回收期,或项目投资方案的投资收益率大于或等于标准投资收益率时,该方案被认为经济可行。

【例 10.1】 新建某工厂一次性投资总额是 620 万元。建成后,该厂的年销售收入是 1300 万元,年产品经营成本是 1052 万元,试求该厂多少年可偿还项目的总投资。若标准投资

回收期为4年,此项目的经济效益如何?

已知: $K = 620$ 万元, $M = 1300 - 1052 = 248$ 万元

根据公式 $T = \frac{K}{M}$ 可求得该厂偿还项目总投资的年数

$$T = \frac{K}{M} = \frac{620}{248} = 2.5 \text{ 年}$$

由于该厂用2.5年的时间即可偿还项目的总投资,比标准投资回收期4年短得多,因此,经济效益是好的。

投资回收期法和投资收益率法虽然简单易行,能反映投资的经济效果,但这两种方法都没有考虑资金的时间价值,只能用于建设期、使用期较短的项目,而且一般要求投资所获得的年收入,在整个使用年限内大体一致。

10.1.4 最小费用法

若某一工程项目具有多个技术方案可供比较、选择,虽然可采用追加投资回收期法,以一对一,逐一进行比较。在这些方案互相比对的“循环赛”中,不但会造成经济比较工作的复杂化,而且可能得到错误的结论。这时若采用最小费用法,会使计算工作简化,便于选择。

从经济效益的评价原则中,我们已知:如果存在几个创造相同成果的技术方案进行比较、选择时,经济效益大的方案一定是社会劳动消耗最小的方案。技术方案的社会劳动消耗包括投资和年成本这两项不同性质的费用。投资属于一次性支出,非一次性消耗;年成本属于经常性支出,一次性消耗,两者不能简单地相加。怎样才能使两项不同性质的费用,通过简单的计算,在多方案中寻找出社会劳动消耗最小的方案呢?这时可以标准投资回收期为计算期,得到年平均费用或总费用。当方案的年费用或总费用为最小时,其经济效益最好,故这种方法称为最小费用法。

(1) 标准偿还年限内年费用法

计算公式:

$$Z_{\text{年}} = C + \frac{1}{T_0} \times K \quad (10.7)$$

式中 $Z_{\text{年}}$ —— 标准偿还年限内年平均费用;

C —— 年成本;

K —— 投资额;

T_0 —— 标准投资回收期。

公式中 $\frac{1}{T_0} \times K$ 表示把技术方案的一次性投资分摊到标准回收期的每一年费用里,此值与方案的年成本 C 之和,组成了标准偿还年限内年费用。应选择标准偿还年限内年费用最小的方案。

(2) 标准偿还年限内总费用法。

计算公式:

$$Z_{\text{总}} = K + T_0 \times C \quad (10.8)$$

式中 $Z_{\text{总}}$ —— 标准偿还年限内总费用。

【例 10.2】某厂采用标准偿还年限内总费用法,要从五个技术方案中选择出经济效益

最优的方案。

各方案具体数据列于表 10-2(假定标准投资回收期为 5 年)。

表 10-2

单位:万元

技术方案	投资额	年成本	标准偿还年限内年平均费用
第一方案	2000	500	$500 + 2000/5 = 900$
第二方案	2300	430	$430 + 2300/5 = 890$
第三方案	2500	420	$420 + 2500/5 = 920$
第四方案	2700	400	$400 + 2700/5 = 940$
第五方案	3000	360	$360 + 3000/5 = 960$

从表 10-2 计算结果可以看到:第二方案的年率均费用最小,投资效果最好的是五个方案中的最优方案。一般来说,技术方案投资额的增加会带来使日常生产成本下降的结果,但并不意味着投资额越多,相应年成本减少的方案一定是最佳方案。只有当投资额适当增加,而能获得年成本相对多的节约时,方案的经济效果才会显得较好,如第二方案那样。第三、四、五方案虽然投资额不断增加,年成本不断减少,但年成本减少的幅度小于投资额增加的幅度,因此其经济效果随着投资的增加而变差。

10.1.5 投资方案的可比性

一切技术方案一般来说都是以其产品的数量、品种和质量等技术经济指标来满足社会需要的。前面已介绍的几种静态分析法是在假设技术方案的产品数量、品种、质量都在相同的条件下进行讨论的。然而,对某一项目提出的多种方案所引起产品的数量或品种或质量往往存在差异。因此对满足相同需要的不同技术方案进行比较时,首先要求不同方案的产品数量、品种、质量等指标具有可比条件。为了获得方案的可比性,必须对产品数量不同或品种不同,或质量不同的不可比因素进行分析及修正计算,使各个方案基本上处于一个可比的基础上,然后再进行比较和评价。

(1) 产量不同的可比性

如果两个比较方案的产量不同,投资和年成本的绝对值没有可比的基础,不能直接进行比较,必须把投资和年成本的绝对值换算为相对值,即通常所采用的把总投资和年成本化为单位产量投资额及单位产量年成本,然后再进行比较的方法。

换算公式:

$$\begin{aligned} k &= \frac{K}{Q} \\ c &= \frac{C}{Q} \end{aligned} \quad (10.9)$$

式中 k, c ——分别为单位产量投资额和单位产量年成本;

K ——总投资额;

C ——年成本;

Q ——年产量。

如果两个比较方案的总投资额分别为 K_1 、 K_2 ，年成本分别为 C_1 、 C_2 ，年产量分别为 Q_1 、 Q_2 ，那么这两个方案的单位产量投资额和单位产量年成本分别为：

$$\begin{aligned} k_1 &= \frac{K_1}{Q_1} \\ c_1 &= \frac{C_1}{Q_1} \\ k_2 &= \frac{K_2}{Q_2} \\ c_2 &= \frac{C_2}{Q_2} \end{aligned} \quad (10.10)$$

利用求追加投资回收期或求追加投资效果系数的方法，可对此两方案进行比较。

公式

$$T_a = \frac{k_1 - k_2}{c_2 - c_1} = \frac{\frac{K_1}{Q_1} - \frac{K_2}{Q_2}}{\frac{C_2}{Q_2} - \frac{C_1}{Q_1}} \quad (10.11)$$

或

$$E_a = \frac{c_2 - c_1}{k_1 - k_2} = \frac{\frac{C_2}{Q_2} - \frac{C_1}{Q_1}}{\frac{K_1}{Q_1} - \frac{K_2}{Q_2}} \quad (10.12)$$

当 $T_a < T_0$ 或 $E_a > E_0$ 时第一方案优于第二方案。

【例 10.3】某钢铁厂有两个建厂方案，其投资、年成本和产量分别列于表 10-3 中（假设标准投资收益率 $E_0 = 0.15$ ）：

表 10-3

指 标	单 位	第一方案	第二方案
年产量	万吨	600	550
总投资	万元	120 000	105 000
年成本	万元	59 200	55 500
单位产量投资	元 / 吨	200	190.9
单位产量成本	元 / 吨	98.7	100.9

根据公式可得 $E_a = 0.24$ 大于标准投资收益率，所以第一方案经济效益优于第二方案。

把总投资或年成本化为单位产量投资以及单位产量成本，然后两方案进行比较的方法是比较简单的方法。两方案的年产量相差 50 万吨，这是一个相当大的数量。所差的 50 万吨必须要采取另外的措施来向社会提供差额部分的产品。可以采取改建或扩建旧钢厂方案，也可以新建一间年产 50 万吨钢的小型钢厂来达到这一目的。因此，我们必须全面考虑问题；也应考虑所相差的 50 万吨钢的另外投资这一因素。

(2) 质量不同的可比性

产品质量要符合国家规定的质量标准，如果第一方案的产品质量达到了国家规定的质

量标准,第二方案产品质量超过国家规定标准,并且对方案的技术经济效益产生了显著的影响,那么应该对第二方案的投资及成本进行调整,然后再与第一方案进行比较。调整时采用使用效果系数 α 进行修正。

使用效果系数

$$\alpha = \frac{F}{F_0} \quad (10.13)$$

式中 F_0 、 F ——分别为调整前、后产品质量的使用效果。

产品质量的使用效果可用产品的使用寿命、可靠性等指标来表示。

投资额的调整:

$$K_{\text{调}} = K \times \frac{1}{\alpha} \quad (10.14)$$

年成本的调整:

$$C_{\text{调}} = C \times \frac{1}{\alpha} \quad (10.15)$$

当对比方案品种不同时,其调整方法与上述基本相同,使用效果可用材料的节约、工资的节约等表示。

【例 10.4】 已知某工厂年产 1 万台马达,有两个不同的生产方案,具体数据如表 10-4 所示(设标准投资收益率为 0.3)。

如果不进行换算,第二方案肯定不经济,因为此方案投资额与成本都比第一方案大。

表 10-4

方 案	投 资 额	成 本	产品使用寿命
第一方案	1 150 万元	235 元 / 台	4 000 小时
第二方案	1 550 万元	250 元 / 台	5 000 小时

但根据公式可得:

使用效果系数 $A = 5\,000 / 4\,000 = 1.25$

将第二方案的投资和成本进行调整。

根据公式可得:

$$K_{\text{调}} = 1550 \times 1 / 1.25 = 1\,240 \text{ 万元}$$

$$C_{\text{调}} = 250 \times 1 / 1.25 = 200 \text{ 元 / 台}$$

再把单位产品成本换算为年产品成本,第一方案的年产品成本为:

$$C_1 = 235 \times 10\,000 = 2\,350\,000 \text{ 元} = 235 \text{ 万元}$$

第二方案调整后的年产品成本为:

$$C \times 2 = 200 \times 10\,000 = 2\,000\,000 \text{ 元} = 200 \text{ 万元}$$

然后根据公式对两方案的经济效果进行比较,可得:

$$E_A = (235 - 200) / (1\,240 - 1\,150) = 0.4$$

由于 0.4 大于标准投资收益率 0.3,所以第二方案经济效果优于第一方案。

综上所述,静态分析法在对短期投资项目的方案进行粗略评价时,不失为一个比较简便的方法,但对长期投资项目进行经济评价时往往产生假象。特别是它不能确切地表达投资效

果,没有顾及资金收支的时间因素和整个项目存在期间的投资经济效益。因此,这种研究方法只能在某种条件、某种程度上使用,否则会产生使人误解的结果,使用时应特别注意。

10.2 项目投资效果动态分析法

10.2.1 投资方案的现金流量

为了对不同技术方案的投资效果用动态分析法进行比较,首先必须了解投资方案的现金流量。当一个投资方案的年现金流量有收入又有支出时,应该计算出年净现金流量。净现金流量是指同一时期发生的现金收入与现金支出之代数和(正号表示现金收入,负号表示现金支出)。

【例 10.5】 某投资方案的投资额为 130 万元,使用年限为 6 年。假定使用年限终了时固定资产的残值为 10 万元。每年折旧费为 20 万元,每年销售收入为 100 万元,年经营成本为 50 万元,所得税率为 50%,则零年末的净现金流量为 -130 万元,以后一至五年的净现金流量应为:

- (1) 年销售收入(+)100 万元
- (2) 年经营成本(-)50 万元
- (3) 年折旧(-)20 万元
- (4) 年需纳税的收入(+)30 万元(计算方法:(1) - (2) - (3))
- (5) 年税金(-)15 万元(计算方法:(4) × 50%)
- (6) 年折旧(+)20 万元
- (7) 年净利(+)15 万元(计算方法:(4) - (5))
- (8) 年净现金流量(+)35 万元(计算方法:(6) + (7))

第 6 年的净现金流量等于 35 万元加上残值的回收,即 45 万元。求得各年的净现金流量见表 10-5:

表 10-5

单位:万元

年 末	现金支出	现金收入	净现金流量
0	- 130		- 130
1	- 85	120	35
2	- 85	120	35
3	- 85	120	35
4	- 85	120	35
5	- 85	120	35
6	- 85	120	45

10.2.2 净现值法

净现值法是目前国内外评价工程项目经济效益的最普遍、最重要的方法之一。这里所指

的“现值”并不是一定指“现在”的价值。而是以做决策所采用的基准时点(或基准年)为标准,把不同时期发生的现金流量(或净现金流量)按基准收益率折算为基准时点的等值额。求其代数和即得净现值。

净现值的计算首先要确定一个基准收益率。国外称为最小目标收益率,以它作为计算现值的依据。其次要确定一个计算现值的基准年,这个基准年可以选在建设期末,使用期初的那个时点为基准年末,也可以选择建设期初作为基准年末,基准年末一般都用零年末表示,两种确定基准年时的方法都可以用,只要计算时加以说明即可。然后将不同时期发生的净现金流量以选定的基准收益率折算为零年末的现值,这些现值之代数和即为所要求的净现值。评价的标准是以净现值大于或等于零为依据。如果大于零说明除了达到基准收益率以外尚有盈余。如果等于零说明刚刚达到基准收益率水平。如果小于零说明达不到基准收益率的水平。很明显,只有当净现值大于或等于零时投资方案才是经济可行的。

净现值的数学模型可以按现金流量的类型分为两种:

(1) 等额支付系列型的现金流量

$$NPV(i_0) = -P + A(P/A, i_0, n) \quad (10.16)$$

即:

$$NPV(i_0) = -P + A \left[\frac{(1+i_0)^n - 1}{(1+i_0)^n \cdot i_0} \right] \quad (10.17)$$

式中 $NPV(i_0)$ ——基准收益率等于 i_0 时的净现值;

P ——一次支付(常指投资);

A ——等额支付系列的一次支付额;

i_0 ——基准收益率;

n ——使用寿命;

$A \cdot \frac{i_0}{(1+i_0)^n - 1}$ ——反映折旧基金。

(2) 非等额支付系列型的现金流量

$$NPV(i_0) = \sum_{t=0} F_t(P/F, i_0, t) \quad (10.18)$$

式中 F_t ——第 t 年的现金流量。

t ——现金流量发生的年份的序号。

基准收益率是人为确定的收益率,对于某个投资方案来说,现金流量已经确定,而 i_0 没有确定,则根据不同的 i_0 ,可以计算出不同的 $NPV(i_0)$ 值。将 NPV 作为 i_0 的函数绘出的曲线可以为决策者提供有用的知识。

10.3 盈亏平衡分析

作为投资决策依据的技术经济分析是建立在分析人员对未完事件所作的预测与分析基础之上的。由于影响各种方案经济效果的政治、经济形势,资源条件,技术发展情况等因素的未来的变化带有不确定性。加上预测方法和工作条件的局限性,对方案经济效果评价中使用的投资、成本、产量、价格等基础数据的估算与分析结果不可避免地会有误差。这使得方案经济效果的实际值可能偏离预期值,从而给投资者和经营者带来风险。例如,投资超支和建设

工期拖长,生产能力达不到设计要求,原材料价格上涨,劳务费用增加,产品售价波动,市场需求量变化,贷款利率变动等都可能使一个工业投资项目达不到预期的经济效益,甚至发生亏损。

为了尽量避免决策失误,我们需要了解各种外部条件发生变化时对投资方案经济效果的影响程度,需要了解投资方案对各种外部条件变化的承受能力以及对应于可能发生的外部条件的变化投资方案经济效果的概率分布,需要掌握风险条件下正确的决策原则与决策方法。各种不确定因素(如投资、成本、销售量、产品价格等)的变化会影响投资方案的经济效果,当这些因素的变化达到某一临界值时,就会影响方案的取舍。盈亏平衡分析方法就是找出这种临界值,判断投资方案对不确定因素变化的承受力,为决策提供依据。

10.3.1 独立方案盈亏平衡分析

独立方案盈亏平衡分析的目的是通过分析产品产量、成本与方案盈利能力之间的关系找出投资方案盈利与亏损在产量、产品价格、单位产品成本等方面的界限,以判断在各种不确定因素作用下方案的风险情况。

(1) 销售收入、生产成本与产品产量的关系

投资项目的销售收入与产品销售量(如果按销售量组织生产,产品销售量等于产品产量)的关系有两种情况:

第一种情况,该项目的生产销售活动不会明显地影响市场供求状况。假定其他市场条件不变,产品价格不会随该项目的销售量的变化而变化,可以看作一个常数。销售收入与销售量呈线性关系,即

$$B = PQ \quad (10.19)$$

式中 B ——销售收入;
 P ——单位产品价格;
 Q ——产品销售量。

第二种情况,该项目的生产销售活动将明显地影响市场供求状况,随着该项目产品销售量的增加,产品价格有所下降,这时销售收入与销售量之间不再是线性关系,对应于销售量 Q_0 ,销售收入为

$$B = \int_0^{Q_0} P(Q) dQ \quad (10.20)$$

项目投产后,其生产成本可以分为固定成本与变动成本两部分。固定成本指在一定的生产规模限度内不随产量的变动而变动的费用,变动成本指随产品产量的变动而变动的费用。变动成本总额中的大部分与产品产量成正比例关系。也有一部分变动成本与产品产量不成正比例关系,如与生产批量有关的某些消耗性材料费用,工夹模具费及运输费等,这部分变动成本随产量变动的规律一般是呈阶梯形曲线,通常称这部分变动成本为半变动成本。由于半变动成本通常在总成本中所占比例很小,在经济分析中一般可以近似地认为它也随产量成正比例变动。

总成本是固定成本与变动成本之和,它与产品产量的关系也可以近似地认为是线性关系,即

$$C = C_f + C_v Q \quad (10.21)$$

式中 C ——总生产成本;
 C_f ——固定成本;
 C_v ——单位产品变动成本。

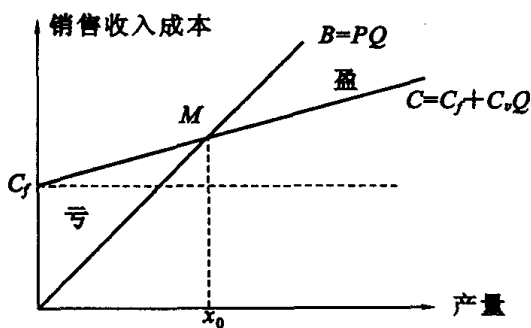


图 10-1 量一本一利分析图

(2) 盈亏平衡点及其确定

将式(10.19)与式(10.21)在同一坐标图上表示出来，可以构成线性量—本—利分析图(图 10-1)。

图 10-1 中纵坐标表示销售收入与产品成本，横坐标表示产品产量。销售收入线 B 与成本线 C 的交点称为盈亏平衡点(Break Even Point, 简称 BEP)，也就是项目盈利与亏损的临界点。在 BEP 的左边，总成本大于销售收入，项目亏损；在 BEP 的右边，销售收入大于总成本，项目盈利；在 BEP 点上，项目不亏不盈。

在销售收入及总成本都与产量呈线性关系的情况下，可以很方便地用解析方法求出以产品产量、生产能力利用率、产品销售价格、单位产品变动成本等表示的盈亏平衡点。

在盈亏平衡点，销售收入 B 等于总成本 C ，设对应于盈亏平衡点的产量为 Q^* ，则有：

$$PQ^* = C_f + C_v Q^* \quad (10.22)$$

盈亏平衡产量

$$Q^* = \frac{C_f}{P - C_v} \quad (10.23)$$

若项目设计生产能力大于 Q^* ，则盈亏平衡生产能力利用率

$$E^* = \frac{Q^*}{Q} \times 100\% = \frac{C_f}{(P - C_v)Q_c} \times 100\% \quad (10.24)$$

若按设计能力进行生产和销售，则盈亏平衡销售价格

$$P^* = \frac{B}{Q_c} = C_v + \frac{C_f}{Q_c} \quad (10.25)$$

若按设计能力进行生产和销售，且销售价格已定，则：盈亏平衡单位产品变动成本

$$C_v^* = P - \frac{C_f}{Q_c} \quad (10.26)$$

【例 10.6】 某工业项目年设计生产能力为生产某种产品 3 万件，单位产品售价 3 000 元，生产总成本为 7 800 万元，其中固定成本 3 000 万元，总变动成本与产品产量成正比例关系。求以产量、生产能力利用率、销售价格、单位产品变动成本表示的盈亏平衡点。

【解】 首先计算单位产品变动成本：

$$C_v = \frac{(7\,800 - 3\,000) \times 10^4}{3 \times 10^4} = 1\,600 (\text{元/件})$$

盈亏平衡产量

$$Q^* = \frac{30\,000\,000}{3\,000 - 1\,600} = 21\,400 (\text{件})$$

盈亏平衡生产能力利用率

$$E^* = \frac{30\,000\,000}{(3\,000 - 1\,600) \times 30\,000} \times 100\% = 71.43\%$$

盈亏平衡销售价格

$$P^* = 1\,600 + \frac{30\,000\,000}{30\,000} = 2\,600 (\text{元/件})$$

盈亏平衡单位产品变动成本

$$C_v^* = 3\,000 - \frac{30\,000\,000}{30\,000} = 2\,000 (\text{元/件})$$

通过计算盈亏平衡点,结合市场预测,可以对投资方案发生亏损的可能性作出大致判断。如果未来的产品销售价格及生产成本与预期值相同,项目不发生亏损的条件是年销售量不低于 21 400 件,生产能力利用率不低于 71.43%,如果按设计能力进行生产并能全部销售,生产成本与预期值相同,项目不发生亏损的条件是产品价格不低于 2 600 元/件;如果销售量、产品价格与预期值相同,项目不发生亏损的条件是单位产品变动成本不高于 2 000 元/件。

10.3.2 敏感性分析

所谓敏感性分析,是通过测定一个或多个不确定因素的变化所导致的决策评价指标的变化幅度,了解各种因素的变化对实现预期目标的影响程度,从而对外部条件发生不利变化时投资方案的承受能力作出判断。敏感性分析是经济决策中常用的一种不确定性分析方法。

(1) 单因素敏感性分析

单因素敏感性分析是就单个不确定因素的变动对方案经济效果的影响所做的分析。在分析方法上类似于数学上多元函数的偏微分,即在计算某个因素的变动对经济效果指标的影响时,假定其他因素均不变。

单因素敏感性分析的步骤与内容如下:

① 选择需要分析的不确定因素,并设定这些因素的变动范围。

影响投资方案经济效果的不确定因素有很多,严格说来,所有影响方案经济效果的因素都在某种程度上带有不确定性。但事实上没有必要对所有的不确定因素都进行敏感性分析,可以根据以下原则选择主要的不确定因素加以分析:第一,预计在可能的变动范围内,该因素的变动将会比较强烈地影响方案的经济效果,第二,对在确定性经济分析中采用的该因素的数据的准确性把握不大。

对于一般的工业投资项目来说,要作敏感性分析的因素通常从下列因素中选定:

A 投资额,包括固定资产投资与流动资金占用。根据需要还可将固定资产投资划分为设备费用、建筑安装费用等。

B 项目建设期限、投产期限、投产时的产出能力及达到设计能力所需时间。

C 产品产量及销售量。

D 产品价格。

E 经营成本,特别是其中的变动成本。

F 项目寿命期。

G 项目寿命期末的资产残值。

H 折现率。

在选择需要分析的不确定因素的过程中,应根据实际情况设定这些因素可能的变动

范围。

② 确定分析指标。

各种经济效果评价指标,如净现值、净年值、内部收益率、投资回收期等,都可以作为敏感性分析的指标。由于敏感性分析是在确定性经济分析的基础上进行的,就一般情况而言,敏感性分析的指标应与确定性经济分析所使用的指标相一致,不应超出确定性分析所用指标的范围另立指标。当确定性经济分析中使用的指标比较多时,敏感性分析可围绕其中一个或若干个最重要的指标进行。

③ 计算各不确定因素在可能的变动范围内发生不同幅度变动所导致的方案经济效果指标的变动结果,建立起一一对应的数量关系,并用图或表的形式表示出来。

④ 确定敏感因素,对方案的风险情况作出判断。

所谓敏感因素就是其数值变动能显著影响方案经济效果的因素。判别敏感因素的方法有两种:第一种是相对测定法,即设定要分析的因素均从确定性经济分析中所采用的数值开始变动,且各因素每次变动的幅度(增或减的百分数)相同,比较在同一变动幅度下各因素的变动对经济效果指标的影响,据此判断方案经济效果对各因素变动的敏感程度。第二种方法是绝对测定法,即设定因素均向对方案不利的方向变动,并取其有可能出现的对方案最不利的数值,据此计算方案的经济效果指标,看其是否可达到使方案无法被接受的程度。如果某因素可能出现的最不利数值能使方案变得不可接受,则表明该因素是方案的敏感因素。方案能否接受的判据是各经济效果指标能否达到临界值。例如,使用净现值指标要看净现值是否大于或等于零;使用内部收益率指标要看内部收益率是否达到基准折现率。绝对测定法的一个变通方式是先设定有关经济效果指标为其临界值,如果净现值等于零,令内部收益率等于基准折现率,然后求待分析因素的最大允许变动幅度,并与其可能出现的最大变动幅度相比较。如果某因素可能出现的变动幅度超过最大允许变动幅度,则表明该因素是方案的敏感因素。

在实践中可以把确定敏感因素的两种方法结合起来使用。

【例 10.7】 有一个生产城市用小型电动汽车的投资方案,用于确定性经济分析的现金流量表见表 10-6,所采用的数据是根据对未来最可能出现的情况的预测估算的。由于对未来影响经济环境的某些因素把握不大,投资额、经营成本和产品价格均有可能在 20% 的范围内变动。设标准折现率为 10%,试分别就上述三个不确定因素作敏感性分析。

表 10-6

年 份	0	1	2 ~ 10	11
投资	15 000			
销售收入			22 000	22 000
经营成本			15 200	15 200
销售税金(销售收入的 10%)			2 200	2 200
期末资产残值				2 000
净现金流量	- 15 000	0	4 600	4 600 + 2 000

【解】 设投资额为 K , 年销售收入为 B , 年经营成本为 C , 年销售税金为 TA , 期末资产

残值为 V_L 。用净现值指标评价本方案的经济效果,计算公式为

$$NPV = -K + (B - TA - C)(P/A, 10\%, 10)(P/F, 10\%, 1) + V_L(P/F, 10\%, 11)$$

$$NPV = -15\,000 + 4\,600 \times 6.144 \times 0.9091 + 2\,000 \times 0.3505 = 11\,394 \text{ 万元}$$

下面用净现值指标分别就投资额、产品价格和经营成本等三个不确定因素作敏感性分析:

设投资额变动的百分比为 x , 分析投资额变动对方案净现值影响的计算公式为:

$$NPV = -K(1+x) + (B - TA - C)(P/A, 10\%, 10) \times (P/F, 10\%, 1) + V_L(P/F, 10\%, 11)$$

设经营成本变动的百分比为 y , 分析经营成本变动对方案净现值影响的计算公式为:

$$NPV = -K + [B - TA - C(1+y)](P/A, 10\%, 10) \times (P/F, 10\%, 1) + V_L(P/F, 10\%, 11)$$

设产品价格变动的百分比为 z , 产品价格的变动将导致销售收入和销售税金的变动, 销售收入和销售税金变动的比例与产品价格变动的比例相同, 故分析产品价格变动对方案净现值影响的计算公式可写成

$$NPV = -K + [(B - TA)(1+z) - C](P/A, 10\%, 10) \times (P/F, 10\%, 1) + V_L(P/F, 10\%, 11)$$

按照上述三个公式, 使用表 10-6 的数据, 分别取不同的 x, y, z 值, 可以计算出各不确定因素在不同变动幅度下方方案的净现值。计算结果见表 10-7。

表 10-7 不确定因素的变动对净现值的影响

单位: 万元

不确定因素 / 变动率	- 20%	- 15%	- 10%	- 5%	0
投资额	14 394	13 644	12 894	12 144	11 394
经营成本	28 374	24 129	19 884	15 639	11 394
产品价格	- 10 725	- 5 195	335	5 864	11 394
不确定因素 / 变动率	+ 20%	+ 15%	+ 10%	+ 5%	
投资额	10 644	9 894	9 144	8 394	
经营成本	7 149	2 904	- 1 341	- 5 586	
产品价格	16 924	22 453	27 983	33 513	

由此可以绘出敏感性分析图(见图 10-2)。

由表 10-7 和图 10-2 可以看出, 在同样的变动率下, 产品价格的变动对方案净现值的影响最大, 经营成本变动的影响次之, 投资额变动的影响最小。

根据上面的分析, 对于本投资方案来说, 产品价格与经营成本都是敏感因素。在作出是否采用本方案的决策之前, 应该对未来的产品价格和经营成本及其可能变动的范围作出更为精确的预测与估算。如果产品价格低于原预期值 10.3% 以上或经营成本高于原预期值 13.4% 以上的可能性较大, 则意味着这笔投资有较大的风险。另外, 经营成本的变动对方案经济效益有较大影响这一分析结果还提醒我们, 如果实施这一方案, 严格控制经营成本将是提高项目经济效益的重要途径。至于投资额, 显然不是本方案的敏感因素, 即使增加 20% 甚至更多一些也不会影响决策结论。

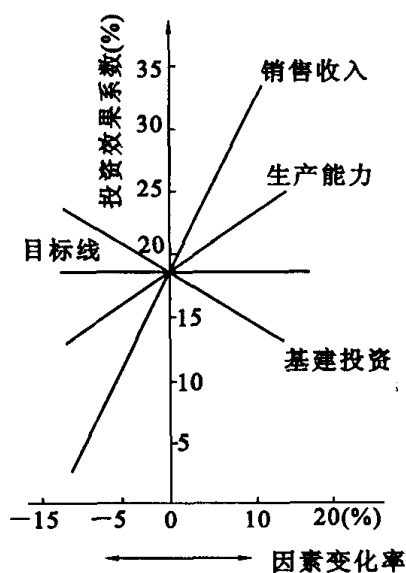


图 10-2 不确定性因素对投资效果的影响

(2) 多因素敏感性分析

在进行单因素敏感性分析的过程中,当计算某特定因素的变动对经济效果指标的影响时,假定其他因素均不变。实际上,许多因素的变动具有相关性,一个因素的变动往往也伴随着其他因素的变动。例如,对于例 10.7 中生产电动汽车这个方案,如果世界市场上石油价格上涨,电动汽车的市场需求量有可能增加,这将导致销售量和产品价格的上升,然而,石油价格上升还会引起其他生产资料价格的上涨,这将导致生产成本的增加。所以,单因素敏感性分析有其局限性。改进的方法是进行多因素敏感性分析,即考察多个因素同时变动对方案经济效果的影响,以判断方案的风险情况。

多因素敏感性分析应考虑可能发生的各种因素不同变动幅度的多种组合,计算起来要比单纯用敏感性分析复杂得多。如果需要分析的不确定因素不超过三个,而投资效果指标的计算比较简单,可以用解析法与敏感性分析法相结合的方法进行分析。敏感性分析在一定程度上对各种不确定因素的变动对方案经济效果的影响作了定量描述,这有助于决策者了解方案的风险情况,有助于确定在决策过程中及方案实施过程中各种需要研究与分析的因素。但是,敏感性分析没有考虑各种不确定因素在未来发生变动的概率,这可能会影响分析结论的准确性。实际上,各种不确定因素在未来发生某一幅度变动的概率一般是有所不同的。可能有这样的情况,通过敏感性分析找出的某一敏感因素未来发生不利变动的概率很小,因而实际上所带来的风险并不大,以至于可以忽略不计。而另一非敏感因素未来发生不利变动的概率却很大,实际上所带来的风险比那个敏感因素更大。这些是敏感性分析所无法解决的,必须借助于概率分析方法。

习 题

10.1 某项目生产某一新产品,已知下列数据:产品售价 12.5 元/件;设备成本费 20 万元;营业成本费 5 万元/年;运行维修费 25 元/运行小时;每千件产品耗时 100 小时;计划生产年数 5 年;具有吸引力的最小收益率 15%;5 年后设备残值为 0。试决定盈亏平衡时的产品年产量。

10.2 设某钢铁厂的建设规模为年产钢材 10 万吨;预测钢材售价 2750 元/吨;估算的单位成本(不含基本折旧)为 1750 元/吨;其中固定费用约占 30%;基本建设投资为 3 亿元,流动资金为销售收入的 25%;

建设期为2年;使用寿命期为15年;投产后上交税金为8%;固定资产残值为原值的10%。试作敏感性分析。

10.3 某项目投资1万元,残值2 000元,年收入5 000元,年支出2 000元,寿命周期5年,基准收益率设为8%。试进行投资和年收入这两个因素的敏感性分析。

11 应用实例分析

本章提要

应用系统工程方法处理实际问题时,人们经常把系统分为“微观系统”和“宏观系统”。这里所说的微观系统指的是,系统规模小于或等于企业级的系统,反之则称其为宏观系统。本章介绍了5个系统工程在不同领域中的应用,属于前者的有:“企业制造资源管理系统(MRP)企业资源计划(ERP)”及“企业计算机集成制造系统(CIMS)”;属于后者的有:“智能交通系统”和“区域规划的系统工程方法”等。

11.1 制造资源管理系统(MRP-Ⅱ)和企业资源计划(ERP)

1970年 Joseph A. Orlicky, George W. Plossl 和 Oliver W. Wight 三人在美国生产库存管理协会(American Production and Inventory Control Society)的学术年会上首先提出了材料需求计划(MRP)的基本框架。并得到该协会的大力支持和推广普及。1981年 Oliver W. Wight 在材料需求计划的基础上将 MRP 的领域由生产、材料和库存管理扩大到营销、财务和人事管理等方面,提出了制造资源计划(Manufacturing Resource Planning 简称 MRP-Ⅱ)。此后在制造资源计划的基础上,发展出了企业资源计划(Enterprise Resource Planning 简称 ERP)。

制造资源管理是制造业基于生产经营模型建立的一整套计算机辅助企业管理信息系统,它涉及到系统工程领域中许多知识,是系统工程在制造和管理领域中的集中体现。现在,MRP-ⅡI 已不只是有效地用于离散加工制造业的生产经营管理系统中,而且已推广到连续型和半连续型生产系统及其他生产和管理部门。MRP-Ⅱ和 ERP 日益成为管理生产、协调加工过程的一种新的思路和有力工具,使企业中的生产、经营、库存、管理中的许多难题正在逐步得到解决。

11.1.1 MRP 的基本原理

MRP 系统是一个数据处理系统,将生产的全过程统一在一个数据信息平台之上。MRP 系统可以根据一个企业的生产规模和所有产品的数量、产品结构、库存数据、在制品数据、材料消耗定额等,计算出为完成生产计划所需的各种零件数量和材料的总需求量,然后从中减去可用库存量和在制品数量得到净需求量。同时,MRP 按零件提前期安排生产作业计划,并对通用件进行合并,分批生产。使用 MRP 的前提是要对整个生产过程有一个充分的了解,对生产工艺进行准确的分解。这些需求都要有一个主生产规划(Master Production Scheduling,简称 MPS)来完成,也就是说,需要有一个关于生产多少、什么产品和什么时间生产出来的权威性指令。MRP 有两种基本实施形式,一种称为全重排式,即计划日程全面重排。其特点是:MPS 中列出来的每一个最终项目需求都要加以分解;每一个有效的文件

都要被访问到;每一个有效的库存状态记录都要经过处理;系统输出大量的报告。另一种称为净改变式,其关键是局部分解。所谓局部分解就是只对当前状态与以往状态的差异进行分解,具体表现在每次运行系统时,都只需要分解 MPS 中的一部分内容,由库存事物处理引起的分解只局限在所分解的库存中某个项目的下属层次上。

11.1.2 MRP 的基本结构

MRP 的基本结构包括 MRP 的输入,MRP 的实行和 MRP 的输出等三个部分。通过相应问题的解决可以确定所需材料的生产和订货日程和进度,保证按生产进度的要求进行生产,同时维持最低的库存水平。

(1) MRP 的输入系统

MRP 的基本输入系统主要由三部分组成:基本生产进度计划,材料清单表和库存状态记录。基本生产进度计划是指在每个时间段根据各种终端物品(一般是最终产品)的需求数量和需求时间,在平衡企业资源和生产能力的基础上制定出的生产进度表;材料清单也称为产品结构表,它表示产品结构和组成单位产品的原材料和零部件的数量。常用的表示方式有两种:一种为树状结构,另一种为表格结构。有些企业根据用途的不同,把材料清单区分为设计材料清单和制造材料清单。制造材料清单通常是指上面所讲的材料清单。设计材料清单是指企业的产品设计人员从产品设计上的考虑,把同一类型或有类似功能的零部件(之间有替代性)作为一个集合所编成的材料清单,通过计算机辅助设计(CAD)与设计材料清单结合起来设计新产品。库存状态记录是指有关材料库存水平的详细记录资料。这些资料包括现有的库存水平,在途库存,交流周期,订货批量,安全库存,材料用途,供应商资料等。这些记录都是动态的记录。完整、准确、动态的库存信息是使 MRP 系统发挥作用的保证。

(2) MRP 的实行过程

MRP 的实行过程就是在材料清单和库存状况记录的基础上,通过计算求得每个时间段上各种材料的净需求数量,同时也确定材料订货的数量,订货时间,订货批量和零部件的加工组装时间等内容。

(3) MRP 的输出

MRP 系统提出的报告分为两种:一种是基本报告;另一种是补充报告。基本报告的内容主要有计划订货日程进度表,进度计划的执行和订货计划的修正及优先次序的变更。其中订货日程进度表包括将来的材料订购数量和订购时间,材料加工的数量和加工时间等。进度计划的执行包括材料品种、规格、数量及到货时间、加工结束时间等规定事项。订货计划的修正及优先次序的变更包括到货日期、订购数量的调整、订单的取消、材料订货优先次序的改变等事项。补充报告的内容主要有成果检验报告、生产能力需求计划报告和例外报告。其中成果检验报告包括物流成本效果,供应商信誉,是否按时到货,材料质量、数量是否符合要求,预测是否准确等;生产能力需求计划报告包括设备和人员的需求预测,工序能力负荷是否满足需求等;例外报告是针对重大事项提出的报告,为高层管理人员提供管理上的参考和借鉴,例如,发生到货时间延后严重影响生产进度造成重大损失时,就到货延期产生的主要原因以及防范应变措施提出的报告。

11.1.3 闭环 MRP 和 MRP-II

一个真正的 MRP 系统并不是一条工作流程直线就能完成,它可能需要多次反复,从而构成 MRP 的闭环。MRP 系统中的所谓闭环,在这里有两层意思:一是指把生产能力需求计划、车间作业计划、采购作业计划等几个环节纳入 MRP,形成一个闭环系统;二是指在执行计划的过程中,必须有反馈信息,并利用反馈信息进行调整平衡。美国著名的 MRP-II 专家 Oliver Wight 给出其基本逻辑图如图 11-1 所示。

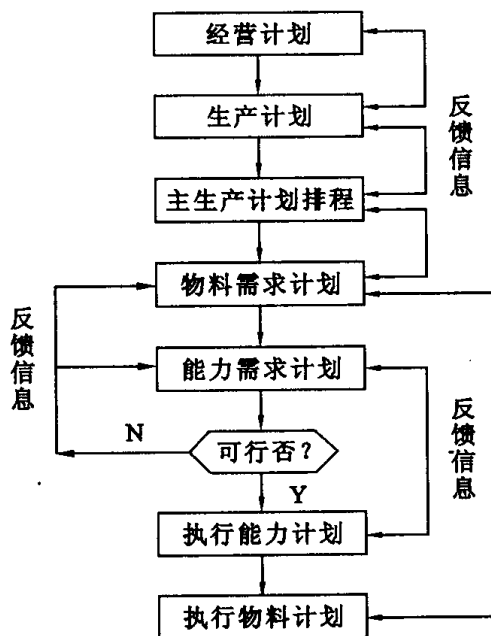


图 11-1 MRP-II 基本逻辑图

MRP-II 是闭环 MRP 的基础上逐渐发展形成的,它从制造业企业的广义资源,如人力、设备、物料、资金、市场、工艺、技术等需求出发,考虑到企业经营决策的战略级、中短期管理调度级和车间作业计划执行的监控级,将系统功能、市场销售、各级计划和控制、物资供应、设备管理、财务管理、成本核算、车间管理以及库存、技术和生产数据管理等广泛领域,并以经营计划、销售计划、主生产排程、物料需求计划、采购计划、生产能力计划、车间作业计划为中心形成一体化(或称集成化)生产、计划、经营管理系统。MRP-II 的特点在于:第一,它是一个一体化系统,可以把企业中的各个子系统有机地结合起来,特别是在财务与生产两个子系统之间,关系应该尤为密切;第二,所有的数据来源于企业的中央数据库,各子系统在统一的数据库环境下协调工作;第三,MRP-II 具有模拟功能,能够根据不同的决策方针模拟未来各种将会发生的结果,因此它也是企业上层管理机构的辅助决策工具。

根据以上所述 MRP-II 的基本原理,结合国内外不同企业的具体情况,对 MRP-II 系统可有多多种划分方法。国外 MRP-II 系统软件的一种简单的子系统划分如图 11-2 所示,它是以物料管理软件为主,辅助以销售订货管理、生产管理、生产成本管理、采购管理和三个财务账目管理软件。物料管理软件以 MRP 为主,兼有 MRP-II 的其他基本功能,根据用户的特殊需求还可以从其他有关软件中选取一个或几个与物料管理软件相结合,以进一步增强整个系统的功能。

国内 MRP-II 软件的一种子系统划分如图 11-3 所示,它是由与中央数据库发生关系的

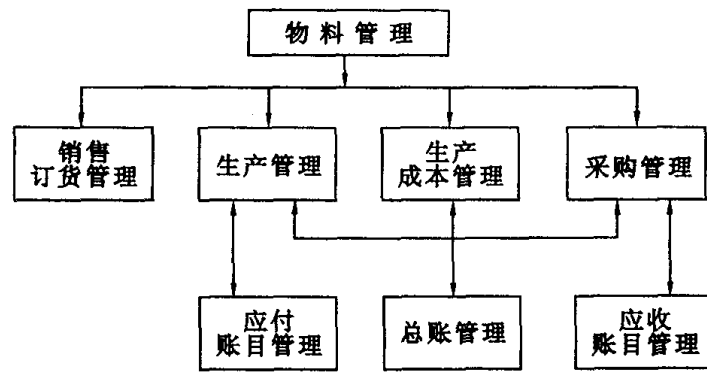


图 11-2 物料管理为主的 MRP-II 系统

13 个子系统组成,这 13 个子系统基本包括了我国制造企业在生产、经营、计划管理中的所有方面。

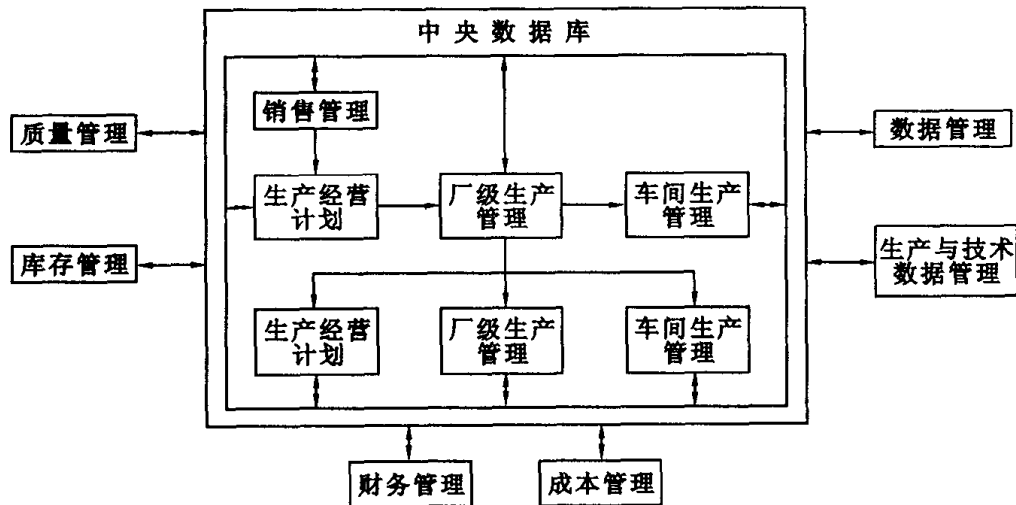


图 11-3 MRP-II 软件的子系统划分

11.1.4 实施 MRP 系统的注意事项

为了使 MRP 系统顺利地运行,需要处理好以下几个主要方面的事项:

A 必须使生产与营销紧密地结合起来,这是确保 MRP 有效的基本条件。因为通过 MRP 系统制定材料需求计划需要花大量的时间和资源。如果在时间段内销售信息经常发生变化,则需要不断修改 MRP。这样不仅会花费大量的时间和资源,而且难以保证 MRP 的有效性。

B 正确及时的库存状况信息是保持 MRP 系统有效的重要条件。对库存状况信息进行及时、准确的更新极为重要。通常采用两种更新方法:一种是定期更新方式;另一种是及时更新方式。定期更新方式是每隔一定的时间更新库存数据(如每隔一周更新一次),根据更新后的结果来对 MRP 进行修正。这种方式成本较低,适用于较为稳定的系统,其最大的缺点是可能导致库存数据与实际库存状况相背离的情况发生,即库存状况和库存记录之间存在着时间差。及时更新方式是一旦库存发生变动的情况马上就更新库存数据的方式,这是一种连续更新方式,它适用于动态型系统。由于库存数据的及时更新往往导致 MRP 的及时更新,它的优点是能及时反映实际情况并做出迅速反应,缺点是这种方式需要投入大量

的资源 and 花费大量的时间,成本较大。

C 材料清单,生产流程,工序能力,交纳周期等基本数据必须准确完备,并根据需求及时更新。只有这些基本数据准确完备才能保证 MRP 系统的顺利运行。

D 必须与其他部门紧密联系,才能保证 MRP 目标的实现。MRP 必须与企业的销售系统,物流系统,采购系统,研究开发系统,财务系统,人事系统等紧密协作,从而有效地实现材料需求计划的基本目标。即在合理利用组织资源保持生产流程畅通的前提下维持最低的库存水平。

E MRP-II 是以生产控制和库存控制为重点。然而,为了满足顾客需求,提高顾客服务水平,同时为了有效利用企业资源,降低库存,减少成本,必须把 MRP 的实施与营销、开发、财务、人事等其他部门有机地结合起来。比如营销部门提供的销售预测是 MRP 最重要的输入数据;研究开发部门应提供准确的产品结构的材料清单;财务部门为采购材料能提供资金支持;人事部门应为 MRP 的有效运行提供人员保证。另外,需要把以工厂为单位的 MRP 事例改为以整个企业为单位的 MRP,这样,就发展出一个以 MRP-II 为核心,加上营销管理,物流管理,财务管理,产品材料数据库管理等管理功能,同时应用现代信息技术的企业资源计划(ERP)。进一步以 ERP 为基础,在供应链水平上开发出供应链计划。

11.1.5 企业资源管理(ERP)

随着全球化经济的形成,社会消费水平、消费结构和消费市场发生了深刻的变革。产品呈现多样化、个性化、系统化和国际化的特征。以面向企业内部信息集成为主,强调单纯的离散制造环境和单纯的流程环境的 MRP-II 系统有时不能满足企业多元化、跨地区、多供应和销售渠道的全球化经营管理的需求。随着网络通讯技术的迅速发展和广泛运用,为了实现柔性制造、快速占领市场,取得高回报率,生产企业必须转换经营管理模式,改变传统的“面向生产经营”的管理方式,转向“面向顾客生产”,注重产品的研究开发、质量控制、市场营销和售后服务等环节,把经营过程的所有参与者如供应商、客户、制造工厂、分销商网络纳入一个紧密的供应链中。

(1) ERP 的核心思想

ERP (Enterprise Resource Planning)就是在 MRP-II 的基础上通过前馈的物流和反馈的信息流和资金流,把客户需求和企业内部的生产活动,以及供应商的制造资源整合在一起,体现完全按用户需求制造的一种供应链管理思想的功能网链结构模式。ERP 管理是一种全新的管理方法,它通过加强企业间的合作,强调对市场需求快速反应、高度柔软性的战略管理以及降低风险成本、实现高收益目标等优势,从集成化的角度管理供应链问题。ERP 的特征包括四个方面:一是超越了 MRP-II 范围和集成功能;二是支持混合方式的制造环境;三是支持动态的监控能力,提高业务绩效;四是支持开放的客户机/服务器计算环境。

在管理技术上,ERP 在对整个供应链的管理过程中更加强调对资金流和信息流的控制,同时通过企业员工的工作和业务流程,促进资金、材料的流动和价值的增殖,并决定了各种流的流量和流速(见图 11-4)。供应链管理还包括以下内容:战略性供应商和用户伙伴关系管理;供应链产品需求预测和计划;全球节点企业的定位设备和生产的集成化计划跟踪与控制;企业内部与企业之间的材料供应与需求管理;基于供应链管理的产品设计—制造管理;基于供应链的用户服务和物流管理;基于 Internet/Intranet 的供应链交互信息管理。

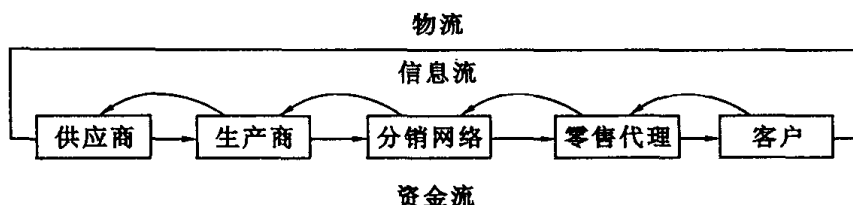


图 11-4 企业 ERP 管理价值链

ERP 已打破了 MRP-Ⅱ 只局限在传统制造业的格局,并把它的触角伸向各行各业,如流程工业、金融业、高科技产业、通讯业、零售业等,从而使 ERP 的应用范围大大扩展。为给企业提供更好的管理模式和管理工具,ERP 还在不断吸收先进的管理技术和 IT 技术。如人工智能、并行工程、Internet/Intranet、数据库等。未来的 ERP 将在动态性、集成性、优化性和广泛性方面得到更大的发展。

(2) ERP 与 MRP 的关系

① MRP 是 ERP 的核心功能

MRP 主要用于订货管理和库存控制,它从产品的结构或材料清单出发,根据需求的优先顺序,在统一的计划指导下,实现企业的产销信息集成,解决了制造业所关心的缺货与超储的矛盾。MRP 作为生产计划与控制模块,是 ERP 系统不可缺少的核心功能。特别是 MRP-II 是 ERP 的重要组成,是以将生产活动中的销售、财务、成本、工程技术等主要环节与闭环 MRP 集成为一个系统,覆盖企业生产制造活动所有领域的一种综合制定计划的工具。MRP-II 通过周密的计划,有效地利用各种制造资源,控制资金占用,缩短生产周期,降低成本,提高生产率,实现企业制造资源的整体优化。MRP-II 较好地运用管理会计的概念,用货币形式说明了执行企业“材料计划”带来的效益,实现材料信息同资金信息集成,保证了资金流同物流的同步,便于实时做出决策。

② ERP 与 MRP(II)的主要区别

ERP 除了传统 MRP-II 系统的制造、销售和财务功能外,还借助于网络通讯技术的应用,使 ERP 系统得以实现供应链管理信息集成,加快了信息传递速度和实时性,扩大了业务的覆盖面和信息的交换量,提高了信息的敏捷通畅,增强了企业的竞争优势,促进了企业业务流程、信息流程和组织结构的变革,推动了 ERP 通过网络信息对内外环境的变化做出能动的反应,为企业进行信息的实时处理和决策提供了极其有利的条件。

11.2 计算机集成制造系统(CIMS)

11.2.1 CIMS 的概念

进入 80 年代,特别是 80 年代中期以来,计算机集成制造系统(Computer Integrated Manufacturing System,简称 CIMS)日渐成为制造工业的热点,世界上很多国家和企业都在制定和实施本国或本企业的 CIMS 发展战略。究其原因在于它符合社会经济和技术的发展趋势,并适应激烈的市场竞争的需要。

电子技术的飞速发展促进了信息产业的形成和迅猛发展,信息时代使世界的市场发生了重大的变化,由过去传统的、相对稳定的市场,演变为动态而多变的市场。激烈的市场竞

争对企业的生存构成严酷的环境,每个企业都面临着求生存争发展的严峻考验。根据美国劳工部公布的美国劳动力在农业、工业、服务业和信息业中的分布统计和预测资料,信息产业所占比例越来越大,其作用也越来越大。

因此,一方面,人们开始探索新的生产方式和经营模式,以提高企业对市场需求的应变能力,加快新产品研制周期,提高产品质量,降低成本,缩短交货期,压缩在制品等等;另一方面,在企业中,随着计算机应用的普及,人们逐步认识到:计算机管理信息系统和数控机床等所谓自动化孤岛,不能充分发挥其应有的效率,需要将它们集成起来,实现资源共享,以发挥更大的能力,促进企业生产、经营的飞跃。很多国家的企业把 CIMS 看作是信息时代的一种工业革命,因为它涉及的不仅仅是制造技术上的进步,而且将涉及到设计、管理乃至相应的体制变革。

对 CIM, CIMS 如何定义? 国内外专家众说纷纭,尚有争议。有人认为, CIMS 就是工厂的计算机集成化;也有人认为, CIMS 就是无人工厂 (Humanless)。站的角度不同,理解的含义也不同,但一致的认识是, CIMS 是 21 世纪企业发展的很有前途的模式。

CIM 是一种概念、一种哲理、是一种用来组织现代工业生产的指导思想,它是由美国学者约瑟夫·哈林顿于 1973 年提出的,哈林顿提出的 CIM 概念有两个基本观点:

(1) 企业生产的各个环节,即从市场分析、产品设计、加工制造、经营管理到售后服务的全部生产活动是一个不可分割的整体,要紧密连接、统一考虑。

(2) 从信息流的角度看,整个生产过程实质上是一个数据的采集、传递和加工处理的过程。

这两个基本观点至今仍被认为是 CIMS 的核心内容,它明确提出 CIMS 的功能包含了一个工厂的全部生产经营活动,比传统的工厂自动化的范围要大得多,是一个复杂的大系统; CIMS 涉及的自动化不是工厂各个环节的自动化和计算机化的简单相加,而是把企业内相互分离的技术(如 CAD, CAM, FMC, MRP-II, ...)和人员(各个部门、各个层次)通过计算机有机地结合起来,实现集成化,使企业内各种活动能高速、灵活、协调地进行,以求在一定条件下,使企业的总体达到最优或准优。

11.2.2 CIMS 的组成

CIMS 是“未来工厂”(Factory of the Future, 简称 FOF)生产经营与自动化的一种新模式,它代表了新一代的生产力,其目标是实现高效率、高柔性、高质量和低成本的生产,以提高制造工业的技术经济效益和增强市场竞争能力。其主要技术手段是集成化、系统化、自动化和信息化。

CIMS 的主要应用对象是离散型制造业,其原因是这类工厂所面临的问题最多、水平最低、生产效率也低,采用 CIMS 后的效果会更大;但 CIMS 的思想及系统方法、集成技术等也同样可用于流程工业和混合型制造业,并已向连续及半连续生产类型的企业扩展。这也是我国“十五”规划和“八六三”研究领域中的一个十分重要的研究课题。

CIMS 目前仍处于发展阶段,还没有固定的发展模式,但其组成主要包括设计、制造、管理三项内容,把这三者所需要的支持集成起来形成一个整体,再加上系统技术及数据库和网络,就形成了 CIMS 工厂的主要功能,其主要功能模块及外部信息输入输出关系叙述如下:

(1) 计算机辅助设计和制造系统(CAD/CAM):该系统包括计算机辅助工程分析 CAE,

产品辅助设计 CAD,成组技术 GT,工艺计划 CAPP,夹具设计,机加工数控编码,模具设计与制造等计算机辅助制造 CAM。

(2) 集成管理信息系统(I-MIS):包括从外部的市场预测、订单及原材料到内部的设备、人力、财务、生产、库存管理及成本、质量、经营计划等方面的信息及其集成。

(3) 柔性制造工程(FME):主要有单元作业计划调度系统,系统运行状态的监控、检测,智能控制和各种设备(加工、刀具、清洗、物料传送、立体仓库)等的物流调度系统。

(4) 数据库和网络:要有一个容量足够大和结构合理的数据库系统;为确保信息的集成和畅通,还要制定一个统一规范及通信协议。

在实施 CIMS 时,要有一套系统分析设计及指导系统最优运行的标准化方法。需要强调的是,CIMS 是一种新型的自动化工厂,但绝不是无人工厂,它追求的是在一定经济条件下系统总体目标的最优,而不是无人化。

11.2.3 实施 CIMS 应注意的问题

CIMS 开发周期长、投资大、涉及面广,因此在实施应用 CIMS 技术、建立 CIMS 企业上,要采取积极而又慎重的态度,要作细致的调研和分析,要注意以下几方面的问题:

(1) 搞好总体规划,是实施 CIMS 的首要一步。要对企业进行全面规划,包括企业的市场需求情况、现有生产技术状况以及对信息技术发展状况的调研,拟出企业 CIMS 的总体目标、它的功能、所需信息数据的采集、传递和处理方式、现有设备的改造、更新、实施周期、经费来源等。总体规划的好坏,将直接影响以后的具体实施,在总体规划指导下,根据企业发展的关键及现有的基础条件,提出切实可行的实施步骤。

(2) 实施过程中,选用正确的技术,采用实用的产品。CIMS 是一个复杂的人—机大系统,需要一种新的思维方式和制造方法来实施,如并行工程等。在实施过程中不应片面追求所谓的“先进”和“时髦”,而应尽量选用实用和相互匹配的技术和产品,以减少不必要的困难和浪费。

(3) 坚持开放式系统体系结构,采用标准化、模块化的结构框架,为以后动态修正和扩展留有充分的余地。由于 CIMS 的开发周期长,又没有形成固定的发展模式,在设计 CIMS 系统时,一定要具有良好的开放性,尽量采用国际标准和公认的准则、约定,以保证系统的兼容性和可扩展性,以及解决信息集成过程的一致性和共享性。

11.2.4 CIMS 应用实例——波音 767-X 并行设计工程

作为 CIMS 应用的具体工程,我们以波音 767-X 并行工程为例讨论 CIMS 应用的一些问题。过去的飞机开发大都沿用传统的设计方法,按专业部门划分设计小组,采用串行的开发流程。大型客机从设计到原型制造多则十几年,少则七到八年。波音公司在 767-X 的开发过程中引入了系统化和集成化的概念,采用了全新的“并行产品定义”的方法,通过优化设计过程集合了最新管理方案,改善设计,提高飞机生产质量,降低成本,改进计划,实现了三年内从设计到一次试飞成功的目标。

波音公司在新型 767-X 飞机的开发中,全面应用 CAD/CAM 系统作为基本设计工具,使得设计人员能够在计算机上设计出所有的零件三维图形,并进行数字化预装配,获得早期的设计反馈,便于及时了解设计的完整性、可靠性、可维修性、可生产性和可操作性。同时,

数字化设计文件可以被后续设计部门共享,从而在制造前获得反馈,减少设计更改。

飞机零件设计采用 CATIA 设计零件的 3D 数字化图形。采用 CATIA 系统设计飞机的零件,可方便地设计 3D 实体模型,并很容易在计算机上进行装配,检查干涉与配合情况,也可利用计算机精确计算重量、平衡、应力等特性。直观的零件图有助于外观设计,并能帮助了解装配后的情况。另外,可以很容易地从实体中得到剖面图;利用数字化设计数据驱动数控机床加工零件;产品插图也能更加容易、精确地建立;用户服务组可利用 CAD 数据编排技术出版用户资料。所有零件设计都只形成唯一的数据集,提供给下游用户。针对用户的特殊要求,只对数据集修改,不对图纸修改。每个零件数据集包括一个 3D 模型和 2D 图,数控过程可用 3D 模型的线架和曲面表示。

数字化整机预装配是在计算机上进行建模和模拟装配的过程,用于检查干涉配合问题,这个过程以设计共享为基础。数字化整机预装配将协调零件设计、系统设计、检查零件的安装和拆卸情况。数字化整机预装配的应用将有效地减少因设计错误或返工而引起的工程更改。随着新一代数字化整机预装配软件工具的不断出现,其功能将包括干涉配合检查,选择最佳精度。数字化整机预装配可以在发图前辅助设计员消除干涉现象。设计员能搜索并进入其他相关设计系统中检查设计协调情况。其他设计小组如工程分析、材料、计划、工装、用户保障等也陆续介入设计范围,并在发图前向设计员提供反馈信息。

并行产品设计是对集成、并行设计及其相关过程的研究(包括设计、制造、保障等)。并行设计要求设计者考虑有关产品的所有因素,包括质量、成本、计划、用户要求等。要充分发挥并行设计的效能,还需以下因素的支持:多方面培养设计人员,合理配置设计制造团队、集成产品设计、制造及保障过程;利用 CAD/CAE/CAM 保障集成设计、协同产品设计、共享产品模型、共享数据库;利用多种分析工具优化产品设计、制造、保障过程。

为什么波音公司在过去的十多年中也采用了 CAD/CAM 系统却没有明显地加快进度、降低费用和提高质量呢?原因是其开发过程和管理还停留在原来的水平上,CAD/CAM 系统的应用能有效地减少更改和设计返工的次数,设计进程也大大加快,由此而带来的效益远比减少更改和返工所带来的直接效益大。波音 767-X 采用全数字化的产品设计,在设计发图前,设计出 767-X 所有零件的三维模型,并在发图前完成所有零件、工装和部件的数字化整机预装配。同时,采用其他的计算机辅助系统,如用于管理零件数据集与发图的 IDM 系统,用于线路图设计的 WIRS 系统,集成化工艺设计系统,以及所有下游的发图和材料清单数据管理系统。由于采用了一些先进的计算机辅助手段,波音公司在 767-X 开发时改进了相应的产品开发过程,如在发图前进行系统设计分析,在 CATIA 上建立三维零件模型,进行数字化预装配,检查干涉配合情况,增加设计过程的反馈次数,减少设计制造之间的大返工。

下面对几个主要的设计过程进行描述。

(1) 工程设计研制过程

设计研制过程起始于 3D 模型的建立,它是一个反复循环过程。设计人员用数字化预装配检查 3D 模型,完善设计,直到所有的零件配合满足要求为止。最后,建立零件图、部装图、总装图模型,2D 图形完成并发图。设计研制过程需要设计制造团队来协调。

表 11-1 波音 767-X 开发方式与传统方式的比较

	767-X 方式	传统方式
工程设计员	在 CATIA 上设计和发图 利用数字化预装配设计管路、线路、舱 利用数字化整机预装配确保满足要求 利用数字化整机预装配检查、解决干涉 利用 CATIA 进行产品插图	在硫酸纸上设计发图 在硫酸纸上设计 利用样机 在生产制造过程处理 利用样件手工绘制
工程分析员	用 CATIA 进行分析 发图前完成设计载荷分析	用图纸分析 鉴定期完成
制造计划员	与设计员并行工作 在 CATIA 上设计工程零件树 用 CATIA 建立插图计划 检查重要特征,辅助软件改型管理	常规顺序 设计—900 零件 建立 mfg. 工程图 无
工装设计员	与设计员并行工作 用 CATIA 设计工装并发图 用 CATIA 允安装检查、解决干涉问题 零件—工装允装配,确保满足要求	常规顺序 用硫酸纸设计 在生产工装时处理 在生产工装时处理
NC 程序员	与设计员并行工作 用 CATIA 生成和检查 NC 过程	常规顺序 用其他系统
用户服务组	与设计员并行工作 用 CATIA 设计所有地面保障设备并发图 技术出版利用工程数据出版资料 零件与地面保障设备预装配,确保满足要求	常规顺序 用硫酸纸设计 手工插图 生成零件/工装
协调人员	设计制造团队	各种机构

(2) 数字化整机预装配过程

数字化预装配利用 CAD/CAM 系统进行有关 3D 飞机零部件模型的装配仿真与干涉检查,确定零件的空间位置,根据需要建立临时装配图。作为对数字化预装配过程的补充,设计员接受工程分析、测试、制造的反馈信息。数字化预装配模型的数据管理是一项庞大、繁重的工作,它需要一个专门的数字化预装配管理小组来完成,确保所有用户能方便进入并在发图前作最后的检查。

(3) 数字化样件设计过程

767-X 利用 CAD/CAM 系统进行数字化预装配,数字化样件设计过程负责每个零件设计和样件安装检查。

(4) 区域设计(AM)

区域设计是飞机区域零件的一个综合设计过程,它利用数字化预装配过程设计飞机区域的各类模型。区域设计不仅包括零件干涉检查,而且包括间隙、零件兼容、包装、系统布置美学、支座、重要特性、设计协调情况等。区域设计由每个设计组或设计制造团队成员负责,各工程师、设计员、计划员、工装设计员都应参与区域设计。区域设计是设计小组或设计制造团队每个成员的任务,它的完成需要设计组、结构室、设计制造团队的通力协作。

(5) 设计制造过程

设计制造团队由各个专业的技术人员组成,在产品设计中起协调作用,最大限度的减少

更改、错误和返工。

(6) 综合设计检查过程

综合设计检查过程用于检查所有设计部件的分析、部件树、工装、数控曲面的正确性。综合设计检查过程涉及到设计制造团队和有关质量控制、材料、用户服务和子承包商,一般在发图阶段进行。有关人员定期检查情况,对不合理的地方提出更改建议。综合设计检查是设计制造团队任务的一部分。

(7) 集成化计划管理过程

集成化计划管理是一个提高联络速度、制定制造工艺计划、测试及飞机交付计划的过程。集成化计划管理过程不但制定一些专用过程计划,而且对整个开发过程的各种计划进行综合。集成化计划的管理,将提高总体方案的能见度。

数字化整机预装配(DPA)是一个计算机模拟装配过程,它根据设计员、分析员、计划员、工装设计员要求,利用各个层次中的零件模型进行预装配。零件是以3D实体形式进行干涉、配合及设计协调情况检查。利用整机预装配过程,全机所有的干涉能被查出,并得到合理解决。

利用数字化预装配过程,工程设计要验证所有设计干涉自由、所有配合良好,这就使过程极少更改。数据集在没有经过最后的审批不能发图,这个最后的检查降低了风险,保障了发图后无零件干涉。

数字化整机预装配是在计算机上进行建模和模拟装配的过程,用于检查干涉配合问题,这个过程以设计共享为基础。数字化整机预装配的应用将有效地减少因设计错误或返工而引起的工程更改。随着新一代数字化整机预装配软件工具的不断出现,其功能将包括干涉配合检查,选择最佳精度。数字化整机预装配可以在发图前辅助设计员消除干涉现象。设计员能搜索并进入其他相关设计系统中检查设计协调情况。其他设计小组如工程分析、材料、计划、工装、用户保障等也陆续介入设计范围,并在发图前向设计员提供反馈信息。

CIMS技术的有效运用会带来以下几方面的效益:提高设计质量,极大地减少了早期生产中的设计更改;缩短产品研制周期,和常规的产品设计相比,并行设计明显地加快了设计进程;降低了制造成本;优化了设计过程,减少了报废和返工率。

11.3 智能交通系统研究

我国是世界上经济持续高速增长的发展中国家,人民生活水平得到了极大的提高,但随之而来的是城市中交通拥挤,堵塞路段、路口逐年增多,交通秩序日趋恶化,交通事故率逐年上升,城市中机动车平均时速逐年下降,出行时间延长,机动车低速及怠速行驶带来更大的燃油消耗和更严重的环境污染,交通拥挤堵塞和交通事故造成的经济损失,每年高达2千亿元以上。造成这种状况的原因是多方面的,解决这些问题也不是单一学科所能解决的。这就需要我们用系统工程的观点全面综合的去分析问题和解决问题。智能交通系统,就是将先进的信息技术、智能技术、数据通讯技术、自动控制技术以及信息处理技术等有效地融合起来,将系统工程的原理运用于整个交通管理系统而建立起来的一种在大范围内,全方位发挥作用的实时、准确、高效的综合智能运输管理系统。

智能交通系统的建立,可以实现城市的信息化和集成化,通过建立基于交通信号控制系

统、交通视频监控系统、综合指挥平台、GPS 移动目标定位系统、地理信息平台、网络通讯平台、车管驾管系统、122 交通事故报警系统、统计分析系统、指挥决策系统及交通诱导系统等系统的集成环境和必要的基础设施,促进高新技术尤其是信息技术在交通管理系统中的应用,实现整个行业的信息和技术资源的共享与优化配置。

智能交通系统涉及多种交通监控、指挥和管理业务,同时又需要综合采用通信、计算机、智能信息处理等多方面的关键技术解决从落后的现实到先进目标的工程实现问题。面对这样复杂的工程,需要通过系统化的总体框架设计,把复杂的大系统抽象成各子系统,从基本构成入手,构造总体结构,分析各子系统之间的相互关系,优化和组织各系统参数,从而为各子系统的设计把握整体方向。

本实例是为我国西北某城市设计的一个典型的智能交通系统。整个系统由以下多个子系统组成,如图 11-5 所示。各个子系统可以完成各自给定的具体工作,各个子系统之间又存在着各种复杂的联系。只有将各个子系统协调在一起,才能完成一个城市复杂的交通控制问题。

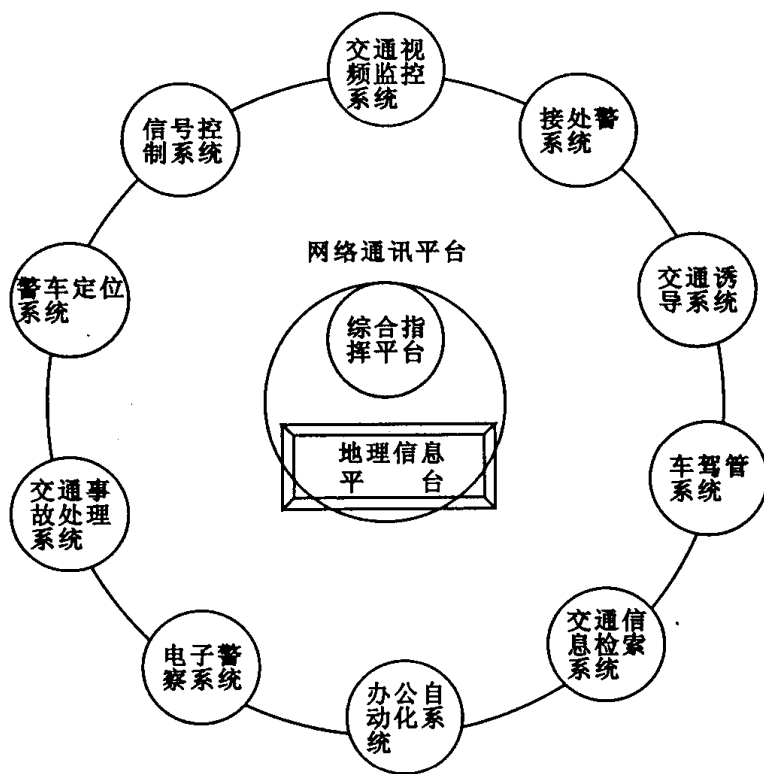


图 11-5 智能交通指挥管理系统总体结构

根据各类交通管理业务的层次特性,我们将智能交通指挥管理系统的分为如下五个层次,如图 11-6 所示。

● 网络通讯层

网络通讯层是整个交通指挥管理系统的信息交换平台,实现各个子系统之间以及相关人员之间的信息共享和交换。网络通讯层主要由“指挥中心局域网”、“城区光纤网”、“电信网”和“无线集群”组成,这些通讯基础设施可以满足不同层次、不同类型的信息采集、交换和命令发布需求。

● 地理信息层

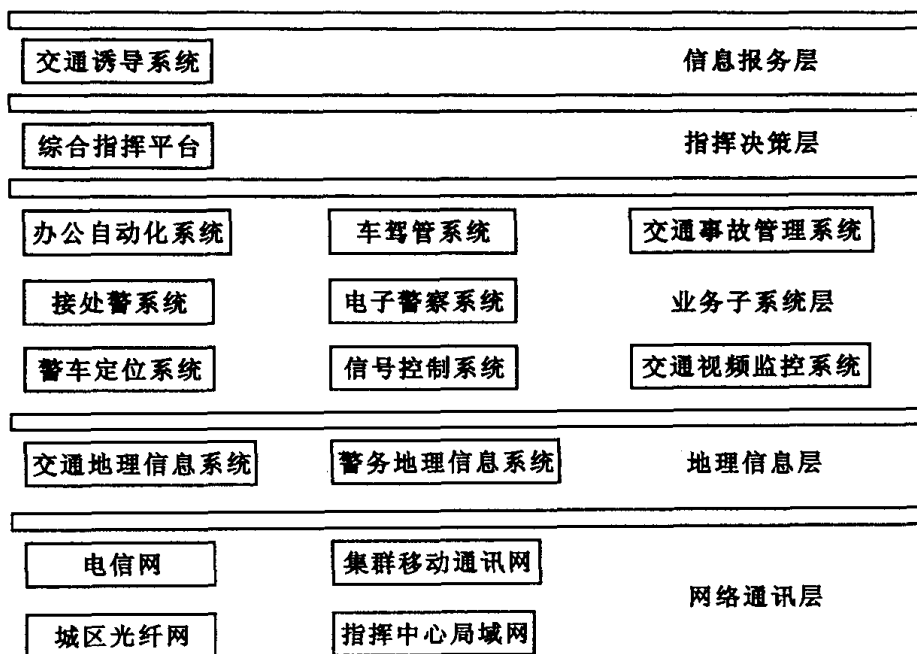


图 11-6 智能交通指挥管理系统层次分析

地理信息层与网络通讯层相同,都作为整个交通指挥管理系统的基础设施,为各类交通指挥管理业务提供基于地图的信息支持,主要包括“交通地理信息系统”和“警务地理信息系统”等。

● 业务子系统层

业务子系统层在网络通讯层与地理信息层之上,交通指挥管理中的主要业务子系统属于这一层,包括“信号控制系统”、“交通视频监控系统”、“122 接处警系统”、“交通事故处理系统”、“警车定位系统”、“电子警察系统”、“车驾管系统”及“办公自动化系统”等。

● 综合指挥决策层

综合指挥调度决策层位于业务子系统层之上,它对来自各个具体的业务子系统的信息进行综合分析,并依据分析的结果作出跨子系统的指挥决策,同时能综合显示当前的交通态势、控制措施及效果,该层包括“综合指挥平台”。

● 信息服务层

信息服务层作为本系统的最高层,主要实现面向公众的交通信息检索和交通诱导,包括“交通诱导系统”。

11.3.1 系统集成环境信息流程设计

智能交通指挥中心的系统集成环境的信息流程可由核心信息流层、子系统间信息流层和外部信息流层三个层次构成。具体流程如图 11-7 所示。

核心信息是指指挥中心内部各个子系统,如信号控制系统、视频监控系统、办公自动化系统等与综合指挥平台之间交换的信息。子系统信息流是指各个相对独立的子系统之间,以及这些相关独立的子系统的中心控制部分与其位于指挥中心之外的现场设备之间交换的信息。外部信息流是指各子系统与外部和操作人员交换的信息。

信息的传输依靠完善的信息网络。交通指挥管理网络是一个覆盖城市的大型综合性通

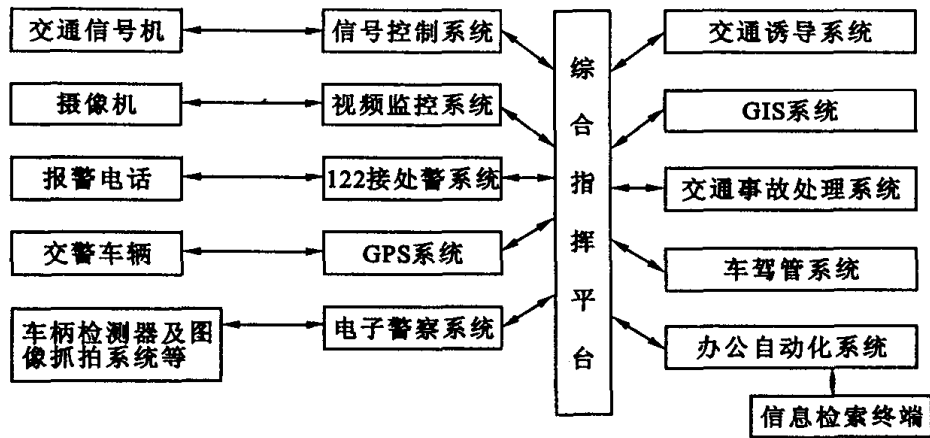


图 11-7 系统集成环境信息流程图

讯网络,应该尽可能的采用各种先进成熟的网络技术。所选购的设备应该具有先进水平,并能够适应未来网络技术发展潮流。各种网络应用软件也应该尽可能先进,保证其在相当长的时间内能够具有技术优势并满足交警部门的需求。同时由于网络技术发展迅速,新技术不断涌现,交通指挥通讯网络也将不断发展。在设计时,应该充分考虑网络的可扩展性,确保在网络进行升级改造时,不需要对网络的整体结构进行大的修改。通讯网络主要是为各交警业务系统和其他网络应用提供可靠的网络通信服务。许多业务,特别是一些实时性业务,对网络的可靠性提出了很高的要求。因此,必须保证整个网络的可靠性,确保网络中的每个节点都有两条及两条以上的线路与网络连接。在确保网络具有高可靠性的同时,还必须兼顾低成本的原则。不仅要保证网络建设的低成本,还要保证整个网络的运行维护的低成本。另外,要求通讯网络在一般情况下保持稳定运行,在各种灾难和突发事件来临时保持连通。

交通综合指挥系统的数据量十分庞大,使得数据的管理显得尤为重要。数据库系统是交通综合指挥和管理的基础,它提供系统所需的所有数据和文档,因而必须包含关系型数据库和文档型数据库。系统集成环境数据管理整体结构如图 11-8 所示。根据公安部的标准,指挥中心数据库选用 Oracle,以 ODBC 方式连接提供数据库通用接口,采用中心和子系统数据库管理系统相结合的混合式数据库管理模式。在一些对数据库的容量和性能要求不高、计算机技术力量不是很强的部门或基层单位,也可以采用一些桌面型数据库,如 Access 等。交通管理业务数据可以存放在不同种类、不同结构和不同类型的数据库中。这虽然实现了因地制宜地使用最符合需要的数据库产品和结构,但同时也是实现信息共享的最大障碍。其中最迫切需要解决的是数据结构定义不统一的异构问题。为此,应在现有应用系统基础上平滑过渡,即遵照相关国家标准,为同一类信息制定统一的数据格式及代码,把现有系统的数据格式逐步转换成统一标准形式,即实现数据标准化。在此基础上,充分利用各数据库系统自身的异种数据库数据迁移工具,实现异种、异构数据库的整合。

11.3.2 信号控制系统

区域交通自适应信号控制系统是随着交通控制理论不断发展,以及通讯、检测、计算机技术在交通控制领域的广泛应用而发展起来的,包括车辆检测、数据采集与传输、信息处理与显示、区域交通信号控制与优化等多个组成部分,可实现单点信号、干线信号和网络信

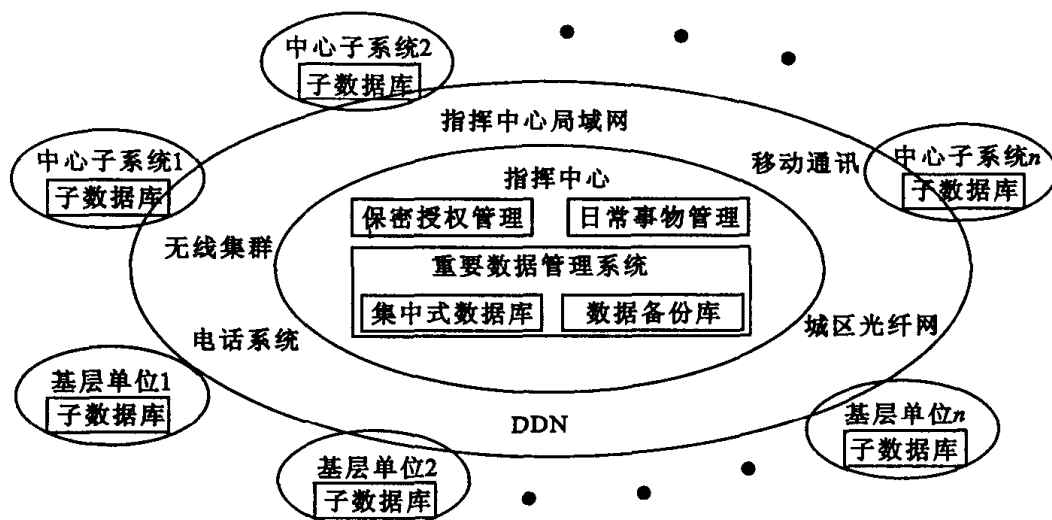


图 11-8 系统集成环境数据库结构图

号等多个层次的信号优化与控制。

信号控制系统是城市交通控制的灵魂。按控制范围分,信号控制系统可分为单点信号控制、干线信号协调控制和网络信号控制三类,即通常所说的“点控”、“线控”和“面控”。按控制方法分类,可分为定时控制和全感应式控制。

当前城市交通信号控制的发展方向是区域信号控制,指的是某个城市或城市某些区域中全部的交通信号控制系统,利用有线或无线进行连接,有机结合成为一个整体的控制系统,同时单点信号、干线信号系统和网络信号系统的综合控制系统。区域交通信号控制系统的控制对象是所有交叉口的交通信号,包括三种基本的信号控制类型:单点、干线和网络控制。其控制的目标是让整个区域内的路网通行能力最大,交通延迟、停车次数以及环境污染减少。

建立区域信号控制系统的优点是:

● 整体监视和控制。整体监视和控制是指能全面、连续地监视和控制交通信号机,即在系统控制区域内无论什么地方发生交通问题或设备故障,都可在较短的时间内检测出来,利用系统自动或人为地主动采取相应的措施,解决问题或排除故障;并能从系统控制的道路网上随时收集交通状态数据。

● 因地制宜地选用合适的控制方法。单点信号、干线信号或网络信号系统都有其有效、经济的适用条件,可根据城区各交叉口的不同情况,选用最合适的控制方法。

● 经济有效地使用设备。在交通信号控制系统区域内,可按选用的各类交通信号控制方法合理地选用必要的设备,达到物尽其用。

交通信号控制系统应提供如下功能:

● 采用区域信号控制的方式,点控、线控、面控相结合,且全部基于感应式控制;

● 信息采集和处理。通过安装于各路口每条车道上的车辆检测器所采集的车辆到达信息,实现路口交通流量信息的自动采集,交通流量统计分析,报警分析,系统监视分析等功能;

● 方案形成与方案选择相结合。通常情况下,通过对采集到的信息的分析处理,形成控制方案,实时地调整绿信比、周期时长及相位差等参数,使之同变化的交通流相适应;在特

定条件下(如与控制中心的通讯联系中断),也可以根据对当前交通状况的分析判断,在预先设定的配时方案中进行选择;

● 主要基于中心控制,即控制方案的形成或选择在指挥中心的信号控制主机上完成。但在中心控制失效时,各信号控制机可“各自为政”,实行自适应感应控制;

● 数据管理和存储。配有专门的数据库服务器,对每台信号控制机所采集到的数据、运行参数、统计分析结果、控制方案、控制效果等进行动态储存和脱机备份;

● 提供图形化的操作界面,方便交通控制操作人员工作;

● 提供交通控制数学模型接口,便于对控制模型进行扩充、改进和更换。

信号控制系统的网络结构如图 11-9 所示。交通信号机与指挥中心之间的通讯视具体情况通过城区光纤广域网或有线电话网实现。

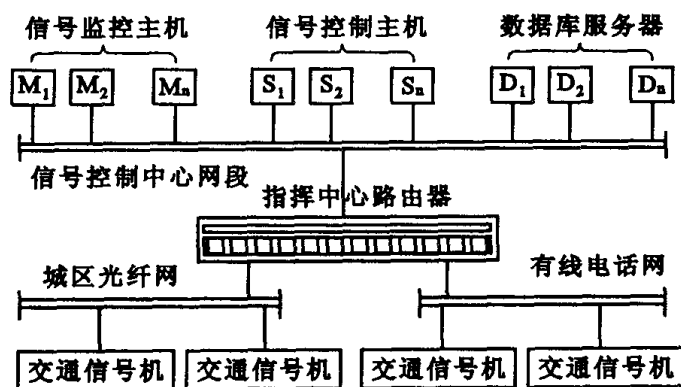


图 11-9 信号控制系统网络结构图

信号控制系统所采集的信息主要是各个路口检测到的车流量、道路占用率、平均车速等,而且检测的周期短、频率高,因而数据量十分庞大。而且,信号控制系统所采集的数据是一笔宝贵的资源,不仅可以用于城市交通流时间及空间分布特征的研究,优化信号控制系统的配置,而且在交通信息服务、统计分析、指挥决策、城市道路及交通规划、交通研究等方面都可发挥重要的作用。

11.4 城市交通系统建设的优化

城市交通系统建设实施方案的优劣可以通过一系列的准则、标准来衡量,一般分为单目标优化和多目标优化。单目标优化和多目标优化不是绝对的,多目标优化中往往将多个目标通过某种方式转化为一个目标以便比较,单一目标评价中如效益—费用分析包括城市交通系统建设项目的可能产生的多种效益,都转化为用货币衡量的经济价值来衡量。

单一目标优化最常见的是效益—费用分析,即将比较方案的效益和费用折合成货币单位,以确定提供最大净效益的方案。一般采用的指标主要有:净现值(NPV)、效益费用比(BCR)、内部收益率(IRR)、投资回收期(N)等。此外还有影响—费用分析、风险排队分析等。典型的多目标优化方法有层次分析法、加权评分法、多目标规划法等。城市交通系统的影响因素较多,应考虑多个目标,如以使用者效益、环境影响、安全等指标为基础的优先指数。例如,西澳大利亚 Pilbara 地区采用的用户和建设部门效益、城市发展、城市环境、服务水平、社会效益五个方面的评价指标如表 11-2 所示。

表 11-2 多目标评价指标实例

类型	用户和建设部门效益	城市 发展	城市 环境	服务 水平	社 会
指标	节约旅行时间	城市和地区经济发展	空气和粉尘污染	全天候服务	安全
	节约车辆运营费用	旅游业发展	水污染	桥涵是否足够	社区可达性
	事故的减少	矿业和能源的开发	植被和动物生态	—	国防
	道路维护费用的减少	—	文化、休憩场地	—	—
	其 他	—	—	—	—

又如,美国芝加哥的郊县采用的一个排序评价系统包括 6 种指标:

- (1) 高峰小时出行的变化;
- (2) 事故的当量财产损失率的变化;
- (3) 平均日路段饱和度(流量/通行能力)的变化;
- (4) 非高峰每日出行时间的变化;
- (5) 大气污染的变化;
- (6) 占用土地的数量。

利用现有的数据资料,应用通常的预测技术,得到各个项目上述指标的数值。用专家意见法识别每项有关的权重系数,将所有的指标加权求和,得出对每个项目的效果指数。这些成本—效果指数可用以排定项目的优先次序,也就是说最有效的计划应按成本—效益递减的次序排列的项目,直到预算的投资用完。由这种方法所识别的项目和决策者最后选择的项目,虽然有某些差异,但还是相似的。

不同城市的性质、特点、发展方向及规划要求不同,需要建立的目标体系和相应的分析方法也不同。但在各种目标体系中,费用—效益分析都是不可缺少的和主要的。

城市交通系统建设优化在项目评估的基础上进行,评估一般采用有无项目法进行效益—费用分析。应该注意的是,在建设优化中的经济评价,是为了用于项目比较而进行的,与建设项目可行性分析或专门的建设项目经济评价不同。后者要求投资、效益等估算的数值与实际尽可能接近,而前者要求尽可能地体现有无项目的区别以及项目之间的区别。因此,有必要研究适于建设优化中项目比较的评价方法。

11.4.1 城市交通系统建设优化中项目效益—费用分析的原则

(1) 采用经济评估

项目评价一般分为财务评估和国民经济评估。城市交通系统建设项目的效益,以社会效益为主,因此采用国民经济评价。

(2) 采用直接费用和效益

项目的效益应包括该项目为国民经济做出的全部贡献。直接效益主要指可用货币形式表现出的项目产出物的经济价值。对应地,费用应考虑直接费用,即可用货币形式表现出的为完成规划项目的建设、管理、运营所付出的经济价值。

(3) 采用有无项目网络比较方法

城市交通系统建设项目的有无比较是将拟建项目建设情况下发生的各种费用和效益与假定拟建项目不实施的情况下的各种费用和效益进行比较,来评估拟建项目的费用与效益。

这是准确衡量项目所产生的净效益的必要方法。

如城市路网建设项目的效益主要产生于项目本身或邻近的平行道路,建成项目上部分运量是从路网上其他路段转移过来的,从而缓和了这些路段的拥挤程度,产生伴生效益。相反,如果项目建设后,使得与其相连的未改建部分的交通量增加,道路拥挤度增大,则会产生负效益。因此,在计算某待建设段的效益时应考虑整个路网,而不仅仅是项目本身发生的费用,即进行有无项目网络费用比较。

(4) 采用统一年限和价格

为了使结果具有可比性,排序中采用统一的年限和价格。如统一采用规划期初价格,计算年限包括建设年限和使用年限,投入使用后的预测年限一般以 20 年计算。

(5) 采用预测的需求 O—D(起始—目标)量进行网络分配

交通量反映了系统使用者的多少,是计算使用者在系统中获得的效益和费用多少的基本数据之一。项目建设后可能会引起系统中的交通量重新分布,也可能引发新的交通出行。变化的交通量可分为转移交通量和诱增交通量两类。但在建设项目实施方案的规划设计中,如果被动地仅考虑是由建设项目引起的交通量变化,将使得交通规划失去主动的指导意义,这并不合理。从另一角度考虑,依据经济发展规划和预测交通量需求,即预测未来的 O—D 出行量进行网络交通分配,而建设项目为需求交通量提供的效益,反映了建设项目满足城市交通需求的程度,或者说是经济发展对于该建设项目需求程度的一种度量,用这种效益作为评价项目优先次序的指标是合理的,这样也使得发展性项目(为满足经济或运输发展扩大的要求)和提高性项目(提高现有路网的效率和服务水平)的可比性增强。

11.4.2 城市交通系统建设优化寻找最优解的方法

城市交通系统项目建设优化决策是序列决策的问题,通常有简单排序法、线性规划方法和遗传算法等。城市交通系统中,每个项目都需要相当长的时间才能建设完成,而影响项目建设的因素较多,过于复杂的算法由于其假设条件太多,反而更加偏离实际,本节研究一种实用性、适应性强并较为可靠的阶段滚动排序方法。

项目排序方法

(1) 全部可能顺序比较法

最佳的道路网建设项目实施方案满足下面的条件:

$$A_2 = \max\{A_{11}, A_{12}, \dots, A_{1n}\}$$

式中 A_1 —— n 个实施方案中建设总效益最大的方案;

A_{1i} ——第 i 个实施方案。

对于短期规划,只有几个建设项目的情况,可以将各种可能的实施顺序一一比较,得到建设项目的实施顺序。但对于规划期较长、建设项目较多的情形,就无法适用了。如果有 n 个路网建设项目,则其可能的实施顺序有 $P_n^n = n!$ 种,也就是说要对 $n!$ 个实施方案作出评价。要对这么多个方案一一加以评价比较,难以实现,因此必须找到一个适用的方法。

(2) 项目一次性比较法

目前经常被采用的一种较简单的方法,是对各建设项目只进行一次评价比较,按这一次评价得到的评价指标从优到劣的顺序确定路网建设项目的实施方案,这样就只要对 n 个项目进行评价。

然而实际上各项目的评价指标并不是静态的,一次比较仅代表一个时刻的情形。低优先次序的项目可能随着时间的推移,在规划期末成为高优先顺序的(由于项目未建而使得交通紧张);而高优先次序的项目也有可能随着时间的推移在规划期末成为低优先顺序的(由于其他平行项目已经建设)。因此,一次性的项目比较方法也只适用于短期规划和项目数目较少的情况。

(3) 项目滚动排序法

滚动排序是一种动态的排序方法,它不像静态的一次性项目比较方法那样仅考虑了一个时刻而不能动态地反映项目建设的实际情况,又不像项目全部可能顺序比较方法那样计算量过大而难以实现。

滚动排序可以看成是一个公路网建设的模拟过程,重复地进行了“比较—建设”的过程:在滚动的某一时刻,将待建项目进行比较,选择最优的项目建设,从而修改了当前的路网和时间(对应不同的交通需求量)到达滚动的下一时刻,直到所有项目建完为止。项目滚动排序法的流程图如图 11-10 所示。

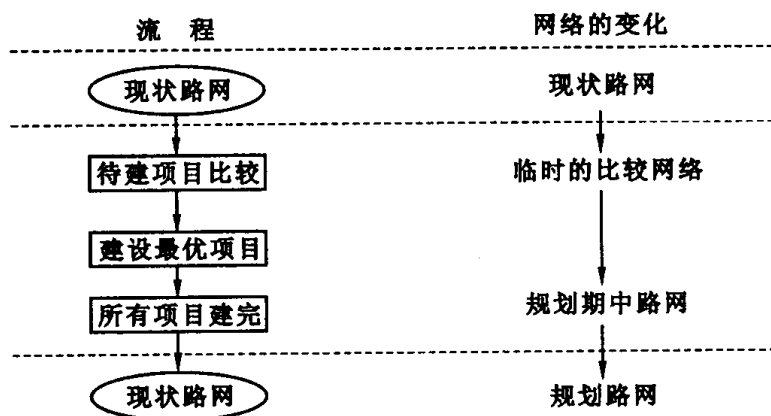


图 11-10 项目滚动排序法的流程图

项目滚动排序法在第 i 次滚动中要比较 $n-i+1$ 个项目,或者说对 $n-i+1$ 个项目进行评价。对 n 个项目,要进行 $n-1$ 次滚动。因此,总计要进行的评价次数为:

$$n + (n-1) + (n-2) + \cdots + 2 = \frac{n^2}{2} - \frac{n}{2} - 1 \text{ 次}$$

显然,滚动排序得到的项目实施顺序,并不是实施方案的最优解,而只是近似最优解。

(4) 阶段滚动排序方法

当项目数较多时,上述滚动排序的计算量仍是较大的。并且在实际的建设过程中,并不是一个项目完全建成后再建下一个项目,而一般是多个项目同时实施,或在一个建设阶段内安排几个项目,这几个项目的具体实施日期是比较灵活的,因此,可以考虑在每次滚动中选出多个项目,而不仅仅是一个项目。每次选出的项目数可以相同,也可以不同。例如将整个建设期分为若干个建设阶段,依次按各建设阶段滚动,对各阶段的建设项目进行排序,而对同一建设阶段内的各个项目,按其评价指标排列。按这种方法得的项目优先次序与项目滚动法得到的结果中能稍有不同,但对建设实施方案的安排影响不大。一般建设期和建设阶段的长短是已知或可以大致确定的:建设期为 5~30 年,建设阶段为 1~5 年,那么所进行的滚动次数一般不超过 30 次,对确定的滚动次数,所需做的评价次数为:

$$n = n_r \frac{N + N^{n_r}}{2}$$

式中 n ——阶段滚动排序方法所需进行的项目评价次数；

N ——项目数；

N^{n_r} ——最后一次(第 n_r 次)滚动评价的建设项目数。

对不同的路网建设规划要求和不同的项目评价方法,所能忍受的最大评价次数也不同。阶段滚动排序的另一个好处在于它可以调整滚动的次数控制评价的次数,以适应不同的评价方法。此外,在路网建设项目的实施过程中,很有可能因资金、施工等多种因素而导致项目的实施建设和完成时间发生变动,阶段滚动排序的同一滚动阶段内,各项目的优先次序不受项目建设先后次序的影响,排序结果仍可作为其他待建项目实施顺序决策的依据。

11.4.3 各种项目排序方法的比较

各种项目排序方法的比较见表 11-3。

表 11-3 各种排序方法计算量和结果准确性的比较

方 法	单项评价次数	结果准确性
全部可能顺序比较法	$n!$	最优解
一次项目比较法	N	静态(某一时刻)解
项目滚动比较法	$\frac{n^2}{2} - \frac{n}{2} - 1$	近似最优解
阶段滚动排序法	$n = n_r \frac{N + N^{n_r}}{2}$	近似最优解

注:表中 n_r 为项目数的滚动次数。

一般来说,结果的准确性总是随着计算量的增加而提高,合理的计算方法应具有适当的准确性和可行的计算量。上述 4 种排序方法的计算量和结果准确性的比较如表 11-3 所示。一般来说,后 2 种方法是较好的方法。阶段滚动法是在其他方法的基础上得到的。取单个项目的建设年限为滚动阶段时,阶段滚动法即为项目滚动法;取滚动次数为 1 时,阶段滚动法又成为静态的一次性项目比较方法。一般可采用阶段滚动排序的方法。

11.5 区域规划的系统工程方法^①

区域科学(Regional Science)是采用定量分析研究和定性经验研究相结合的方式,分析对一个有意义的区域或区域系统的发展产生影响的政治、经济、社会、文化的诸因素,进而研究该区域满意的发展策略和空间分布。区域科学的研究内容非常广泛,比如,发达地区和不发达地区人均国民收入的差别;公民受教育层次的差别及其对工作的影响;不同人群的出行模式;区域间价格和工资的差别;某一工业部门为什么愿意在这个区域“落户”,而不愿意到其他区域;如何分析相互关联的复杂因素,比如在较大的城市区域与较小的城市区域相比,可能会有较好的学校,但同时房租可能会很高,会有交通阻塞等。

① 本节由杨振斌执笔

区域科学在西方国家起始于 20 世纪 40 年代,当时有关的学者们在美国经济学协会(American Economics Association)的年会上举行会议。最早的区域科学研究会成立于 1954 年,在这之后,与区域科学有关的协会和组织如雨后春笋般发展起来。

随着全球经济的发展和城市化水平的提高,区域(小到城镇和乡村,大到国家和大洲)内的可持续发展和协调发展越来越为学者们和各级政府所关注,区域科学和区域发展规划(Regional Development Planning)的研究也越来越广泛,越来越深入。

在国外,美国的 Ira S. Lowry 早在 1964 年就提出了著名的 Lowry 模型,并于 1967 年发表了论文《城市发展的七个模型》,为区域科学的发展做出了重要贡献;美国 W. Isard 教授于 1975 年出版了《区域科学导论》一书;联合国人居中心于 1991 年出版了专为基础设施规划人员和工程人员介绍定量方法基本原理的小册子,提出了 15 种方法和模型;德国的一个政府部就叫做“区域规划、建设和城市发展部”(Federal Ministry for Regional Planning, Building and Urban Development),该部定期出版对全国具有指导意义的《区域规划指南》;英国 P. M. Allen 教授将自组织理论应用于城市和区域规划的研究。

在我国,与区域科学有关的研究主要开始于 20 世纪 80 年代初,规划部门、研究单位和高等学校进行过许多有关宏观经济发展、区域规划等方面的研究。到了 20 世纪 90 年代,在改革开放十几年后,我国经济有了很大的发展,城市化步伐加快,区域科学在我国蓬勃发展起来。在新的世纪,我国的《第十个五年计划纲要》明确提出:“实施西部大开发战略,促进地区协调发展;实施城镇化战略,促进城乡共同进步。”我们有理由相信,区域科学和区域规划的系统工程方法将会得到更广泛的应用,并在应用中得到更大的发展。

11.5.1 区域规划的系统工程方法

由于决策部门的重视和广大科学工作者的努力,我国在区域科学和区域规划研究方面做了大量的研究工作。相当数量的规划研究是成功的,研究成果受到有关领导部门的重视、采纳或参考。同时,对区域科学及区域规划理论和方法论方面的研究也取得不少成绩,推动了学科的发展。

区域规划所研究的是从地区的具体情况出发,对该地区的人口发展、国民经济建设、社会发展、环境质量不断改善等所进行的总体部署及发展战略研究。由于一个县、一个城市、一个地区都是由人口、自然资源、工业、农业、交通运输业、邮电通信业、商业、服务业、基础设施建设、环境质量、科技、教育、文化、卫生等进行综合分析。

区域规划研究首先应该从调查研究开始,分析所处理区域系统的特征、地位、性质和功能,规定系统边界、确定系统范围。在上述基本分析基础上,确定区域发展目标,这项工作十分重要,因为系统目标决定着系统的发展。系统分析的重要手段是应用模型对系统进行定量分析,因此应该建立地区发展综合模型。有了数学模型之后,可以进行发展策略设定,进行政策分析,从而得出满意策略。概括起来,区域规划的研究内容及步骤如图 11-11 所示。

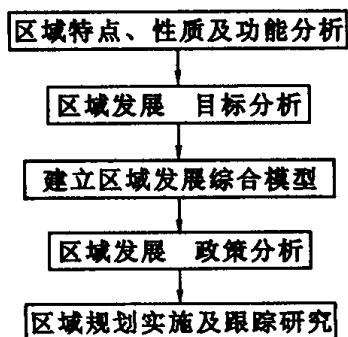


图 11-11 区域规划研究步骤

能,规定系统边界、确定系统范围。在上述基本分析基础上,确定区域发展目标,这项工作十分重要,因为系统目标决定着系统的发展。系统分析的重要手段是应用模型对系统进行定量分析,因此应该建立地区发展综合模型。有了数学模型之后,可以进行发展策略设定,进行政策分析,从而得出满意策略。概括起来,区域规划的研究内容及步骤如图 11-11 所示。

区域规划研究应该采取定量分析和定性分析相结合的方法。定量分析可以帮助决策者从“度”上把握系统的发展;定性分析则是对问题给出更为全面、深刻的“质”的描述,如对区域

发展政策、发展模式、发展战略的表述等等。

11.5.2 区域科学典型模型和方法

1991 年联合国人居中心向从事基础设施投资规划和管理的人员推荐了 15 种定量研究方法和模型,由于基础设施是与一个区域的发展规划紧密相连的,因此这 15 种方法和模型中有很大部分也是与区域科学紧密相连或属于区域科学范围的。限于篇幅,我们只列出这 15 种方法的名称,而不再做进一步的介绍。这 15 种方法是:

- (1) 成本-效益分析(Cost-benefit analyses);
- (2) 效用分析(Utility analyses);
- (3) 综合多目标法(General multi-criteria methods);
- (4) 成本-效率分析(Cost-effectiveness analyses);
- (5) 国家级投入产出分析(National input-output analyses);
- (6) 区域级投入产出分析(Regional input-output analyses);
- (7) 出口主导原理(Export base concept);
- (8) 线性规划(Linear programming);
- (9) 递归规划(Recursive programming);
- (10) 引力模型(Gravity models);
- (11) Lowry 模型(Lowry models);
- (12) 动态仿真模型(Dynamic simulation models);
- (13) 吸引模型(Attraction models);
- (14) 人口迁移模型(Migration models);
- (15) 偏离-份额分析(Shift-share analyses)。

德国经济地理学家 W. Christaller 把地理学的空间观念和经济学的价值观念结合起来,探索城市的数量、规模和分布规律性,从而以抽象演绎的方法创立了主要根据城市向它周围的腹地所提供的服务来解释城镇体系空间结构的理论,称之为中心地理论,其要点是:

(1) 任何城镇都具有向周围地区提供商品和服务的职能,而该职能一般要在它服务区域的相对中心位置来执行,因此这些城镇都可看作是规模不等、级别不同的中心地;

(2) 一个地区会形成一套中心地的等级体系,同等级的中心地有同样大小的服务范围,也称市场区,市场区是六边形的;

(3) 整个中心地及其市场区是由一级套一级的网络相互嵌套而成;

(4) 各级中心地在空间的分布遵循三个原则,即:市场原则、交通原则和行政原则。

这三个原则可以用图 11-12 直观地表达如下:

市场原则表明,一个较高级中心地的市场区正好是下一级市场区面积的 3 倍,因此又称作 $K = 3$ 原则。

交通原则表明,在交通线合理布局的前提下形成的中心地体系,会使较低级的中心地在两个较高级中心地之间的中点发展起来,导致较高级中心地的市场区面积是下一级市场区的 4 倍,又称作 $K = 4$ 原则。

行政原则表明,不同等级的中心地的分布首先受制于行政管理和政治控制的需要,低级中心地及其市场区必须完全被包含在高一级中心地的控制范围之内,以消除行政管理上模

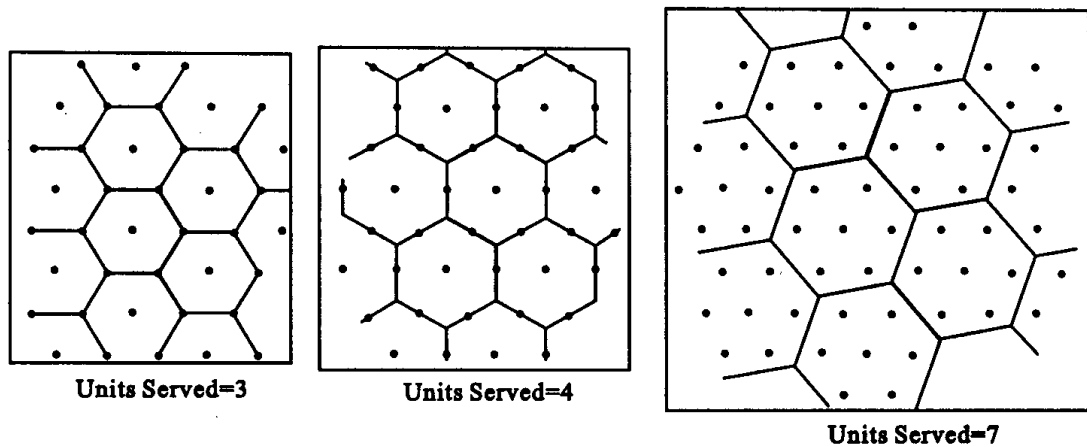


图 11-12 中心地理论的简单图示

棱两可的任何可能,这样,市场区和行政区实际趋于一致。一个较高级中心地供应和管理周围 6 个低一级中心地,包括自身一共 7 个,又称作 $K=7$ 原则。

Cristaller 认为,以上三个原则在一个区域常常会交叉发生作用,但又各有侧重:在经济发达、交通便利的地区,市场原则最占优势;在自给自足和偏僻地区,行政原则居于首位;而在新开发地区,移民拓荒沿交通线推进,则交通原则会占上风。在 20 世纪 60 年代,兰德公司受邀为美国匹兹堡(Pittsburgh)市的未来发展制定土地利用战略分析模型,这个模型就是以后的“Lowry 模型”(Lowry Model),因为其作者为 I. S. Lowry。模型将所研究的区域分成 1 平方英里大小(约合 2.6 平方公里)的“片”(Tracts),通过迭代计算找出居住和服务这些“非基础活动”(non-basic activities)相对于“基础活动”(basic activities)以及上班和到服务点的出行方式之间的最优分布。Lowry 模型不但给出了有关居住和服务用地最佳位置的信息,还可针对不同的设定,比如基础活动的不同方案而给出相应的结果,而且还可提供系统内片与片之间的上班和服务点的出行流的信息。

Lowry 模型后来得到进一步的发展,例如大东京地区公共运输系统就是基于 Lowry 模型而规划的,德国斯图加特大学与上海理工大学合作研究的上海闵行区土地利用和交通模型也使用了 Lowry 模型。

Lowry 模型由 9 个方程和 3 个约束组成:

$$\begin{aligned}
 A_j &= A_j^U + A_j^B + A_j^R + A_j^H & N &= f \sum_{j=1}^n E_j \\
 E^k &= a^k N & N_j &= g \sum_{i=1}^n \frac{E_i}{T_{ij}} \\
 E_j^k &= b^k \left[\left(\sum_{i=1}^n \frac{c^k N_i}{T_{ij}^k} \right) + d^k E_j \right] & N &= \sum_{j=1}^n N_j \\
 E_j &= E_j^B + \sum_{k=1}^m E_j^k & E_j^k &\geq Z^k, \text{ 或 } E_j^k = 0 \\
 E^k &= \sum_{j=1}^n E_j^k & N_j &\leq Z_j^H A_j^H \\
 A_j^R &= \sum_{k=1}^m e^k E_j^k & A_j^R &\leq A_j - A_j^U - A_j^B
 \end{aligned}$$

其中:

变量:

A ——区域土地;

N ——一块区域里的居民数;

T ——系统中片之间的交通“成本”指标;

E ——一块区域里的就业人数;

Z ——约束。

上下标:

B ——基础活动;

H ——住房;

R ——非基础活动;

U ——未使用的土地;

k —— $k = 1, \dots, m$ 非基础活动中的行业;

i, j —— $i, j = 1, \dots, m$ 所研究区域中的片。

参数:

a ——每户中非基础活动的就业人数;

e ——非基础活动中每个就业人的平均土地空间需求;

f ——就业年的倒数;

b, c, d, g ——权重及放大系统。

英国 Cranfield 大学国际生态技术研究中心的 P. M. Allen 教授将 Christaller 的中心地理论、Zipf 定律 (Zipf Law) 与系统自组织理论相结合, 提出了基于系统自组织的城市 and 区域发展研究方法, 该方法既基于中心地理论的经济相互作用, 又动态地描述了系统的深化。Allen 认为, 实质上的, 城市化的过程就是一个在一定区域内长时间的移民和经济投资的空间集中的过程, 这也是一个系统自组织的过程, 持续的不稳定性可使系统产生深化, 而涨落 (fluctuations) 可导致系统走向不同的将来。如图 11-13 所示。

Zipf 定律指出, 在一个区域内中心城市与周围城镇规模的大小基本遵循以下规律:

$$P_n = \frac{P_1}{n^b}$$

式中 P_1 ——中心城市人口数;

P_n ——第 n 级 (Rank) 城镇的人口;

b ——指数。

Verhulst 的人口增长方程为:

$$\frac{dx}{dt} = bx(P - x) - dx$$

式中 x ——人口数。

常数 b 和 d 与出生率和死亡率有关。

P 是该点自然状况的一个综合反映。

与 Christaller 一样, Allen 假设一个区域搭开始时其人口、资源等都是均匀分布的, 涨落可以在任何一点产生, 运动可向任何方向进行。根据 Zipf 定律、中心地理论、Verhulst 的人口增长方程和系统自组织理论, Allen 先导出了简化的理论模型, 在这一简化的模型中, 一个区

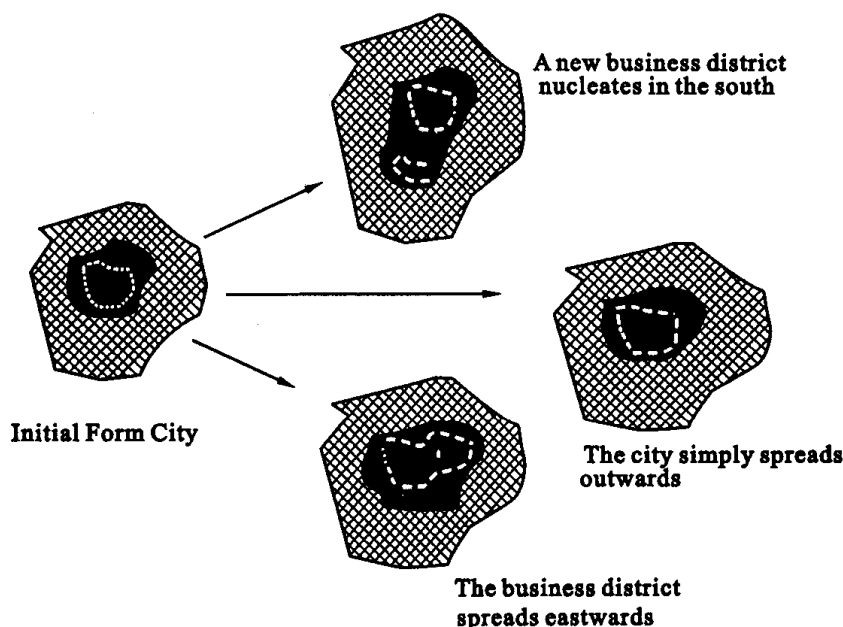


图 11-13 涨落对系统的影响

域的城市化被看成是由于在随机的地点和时间进入系统的经济功能的顺序集成,同时伴随以交通和通讯手段的进化,不同中心的发展及相互间的竞争,以及由此引起的对在空间上不均匀的就业机会做出的响应的人口变迁。

以后 Allen 又逐渐将经济创新,不同中心间的竞争,交通网络的发展,贸易壁垒等因素考虑进来,形成以下的最终方程:

$$\begin{aligned}\frac{dx_i}{dt} &= bx_i(N_i + \sum_k S_i^k - x_i) - mx_i - \tau \sum_j e^{-\beta d_{ij}} (x_i^2 - x_j^2) \\ \frac{dS_i^k}{dt} &= \alpha S_i^k (M_i^k - S_i^k) \\ M_i^k &= \eta^k \sum_j \frac{x_j \varepsilon^k}{(P_i^k + \Phi^k d_{ij})^p} \cdot \frac{A_{ij}^k}{\sum_i A_{ij}^k} \\ A_{ij}^k &= \frac{\left[\delta + \frac{1}{\gamma + \rho^k (x_i - x_h^k)} \right]^I}{(P_i^k + \Phi^k d_{ij})^I}\end{aligned}$$

Allen 得出结论认为,一个区域的城镇结构依赖于其“内部”历史事件的顺序、人口的变化以及撤销或建立经济壁垒的相对时限,还依赖于中心之间相互作用的范围,而这种范围又依赖于在某一给定时间的交通条件。所有这些因素的相互作用影响系统的演化。

11.5.3 实例分析 —— 北京市城市发展模型及发展战略研究

北京市是一个超大型城市,解放前的北京是一个消费城市;解放后,对北京市的发展提出了“变消费城市为生产城市”。用了三十多年的时间,北京市实现了这一目标。到 20 世纪 80 年代初,建起了以冶金、机械、石油化工为主的完整工业体系,农业也有了很大的发展,全市工农业总产值在全国名列前茅。进入 80 年代之后,再次对北京市城市功能及战略发展目标进行分析,认为北京市是我们国家的首都,1983 年,中央对首都的功能及其发展作了明确的指示:“北京是我们伟大社会主义祖国的首都,是全国的政治中心和文化中心。北京的城市建

设和各项事业的发展,都必须服从和充分体现这一城市性质和要求。”所以,北京市又一次面临城市发展目标的调整和转移,系统目标的调整及确定如图 11-14 所示。

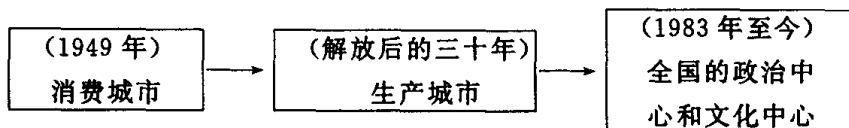


图 11-14 北京城市发展目标

城市系统包括人口、经济、资源、环境、基础设施建设、社会发展等各个方面。应用城市系统综合动态模型进行城市发展战略研究的系统结构图如图 11-15 所示。这里,决策者首先根据战略总目标制定发展政策,城市综合模型根据政策计算出城市今后若干年各个方面可能出现的状态,然后根据评价指标体系进行政策分析,最后从策略集合中选择出满意的方案推荐给决策者。

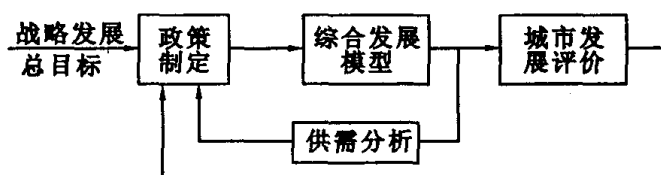


图 11-15 城市系统政策分析总体结构

北京城市系统综合动态综合模型由 6 个子系统模型和一个供需模型组成,6 个子系统模型各自独立存在,它们的联系由供需模型完成。这 6 个子系统模型为人口、经济、资源、环境、基础设施、科教等组成。下面只介绍其中两个子模型的结构:

(1) 经济子系统模型

经济模型形式为

$$x_i(k+1) = A_i x_i(k) + B_i u_i(k) + D \quad (11.1)$$

其中

$i = 1, 2, \dots, 5$ 分别表示工业、农业、建筑业、运输邮电业、商业等五大生产部门;

k 为时间变量,代表第 k 年;

(以下说明中,为了简化,省去下标变量 i)

$x = [x_1, x_2]^T$, x_1, x_2 分别表示第 i 部门固定资产和劳动力变量;

$$u = \begin{bmatrix} u_{10} & u_{11} & u_{12} & u_{13} \\ u_{20} & u_{21} & u_{22} & u_{23} \end{bmatrix}$$

u_{1j}, u_{2j} 分别表示第 i 部门 j ($j = 0, 1, 2, 3$) 年前固定资产和劳动力的投入;

A, B, D 为相应维数的系数矩阵,其中 A, D 为对角阵, B 为满阵。

输出方程为

$$y_i(k) = f[x_i(k)]$$

这里 f 为生产函数,我们选用道格拉斯函数,不记下标 i 时,输出方程的具体形式为

$$y(k) = e^r x_1^a(k) x_2^b(k) \quad (11.2)$$

(2) 人口子模型

人口子模型为

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ x_3(k+1) \end{bmatrix} = \begin{bmatrix} 1-g_1-h_1 & pq(k) & 0 \\ h_1 & 1-g_2-h_2 & \\ 0 & h_2 & 1-g_3-h_3 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{bmatrix} + \begin{bmatrix} b_1 & 0 & 0 \\ 0 & b_2 & 0 \\ 0 & 0 & b_3 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \quad (11.3)$$

或 x_1 ——0—15 岁人口;
 x_2 ——16—60 岁人口;
 x_3 ——60 岁以上人口;
 u_1, u_2, u_3 ——与上述相应年龄段人口的机械增长率;
 g_1, g_2, g_3 ——相应年龄段人口死亡率;
 h_1, h_2, h_3 ——相应年龄段人口转移率;
 p ——16—60 年龄段育龄妇女比例;
 q ——生育率;
 b_1, b_2, b_3 ——相应年龄段人口增减系数。

输出方程为总人口数 y , 有

$$y(k) = [1, 1, 1]x(k) \quad (11.4)$$

关于模型参数的辨识,以经济系统为例,经济系统中许多因素可以认为是线性相关的。不精心选择变量往往得不到正确的辨识结果,考虑到这一因素,可以选用具有明确物理意义的状态变量,并力求状态转移矩阵 A 为对角阵。因此辨识过程可以归结为:①按物理意义,经验或统计方法先确定 A 元素的值。②在 A 已知情况下,用最小二乘法确定 B 。如果对所得结果不满意,可以利用 B 已知的条件,用最小二乘法再一次对 A 估值,如此反复,直到满意为止。下面以工业子系统为例说明具体算法。

工业系统中 A 阵对角元素 a_1, a_2 的物理意义分别是固定资产扣除折旧部分和劳动力中扣除离职部分,所以:

$$a_1 = 1 - \text{折旧率}$$

$$a_2 = 1 - (\text{退休人数} + \text{死亡人数}) / \text{劳动力总数}$$

如设备使用期限为 20 年,故折旧率为 0.05,则 $\hat{a}_1 = 0.95$;又根据多年统计数字可计算出 $\hat{a}_2 = 0.98$ 。

B 阵参数识别方法是,将子系统模型展开为多输出单输出形式:

$$x(k+1) = \hat{a}x(k) + b_1u_1(k) + b_2u_2(k) + \cdots + b_nu_n(k) \quad (11.5)$$

引用

$$z(k+1) = x(k+1) - \hat{a}x(k) = h^T(k+1)b \quad (11.6)$$

其中

$$h^T(k+1) = [u_1(k) \quad u_2(k) \quad \cdots \quad u_n(k)]$$

$$b^T = [b_1 \quad b_2 \quad \cdots \quad b_n]$$

按照最小二乘递推计算公式求得估值为:

$$\hat{b}(N+1) = \hat{b}(N) + K(N+1)[z(N+1) - h^T(N+1)\hat{b}(N)]$$

$$K(N+1) = P(N)h(N+1)[\alpha + h^T(N+1)P(N)h(N+1)]^{-1}$$

$$P(N+1) = \frac{1}{\alpha} [1 - K(N+1)h^T(N+1)]P(N)$$

式中 α ——遗忘因子(考虑到参数的时变性,我们采用了带有遗忘因子的算法);
 N ——年数(从 1949 年到 2000 年)。

11.5.4 区域发展策略优化分析

区域发展策略优化分析是在确定城市发展目标基础上,从实际出发,在可行政策方案集合内,通过决策分析,寻求出满意的策略方案。

策略优化使用的是动态多目标效用分析方法,依据下式对策略方案进行优化分析:

$$v_j(x) = \sum_k \sum_i \omega_i(k) \frac{\Delta v_j(x, k)}{[1 + r_i(k)]^k} \quad (11.7)$$

$$a^* = \max_j v_j(x)$$

式中 x ——属性;
 v ——价值函数;
 ω ——权重系数;
 r ——贴现率;
 i ——属性指标标号, $i = 1, 2, \dots, n$;
 j ——策略方案标号, $j = 1, 2, \dots, J$;
 k ——离散时间变量, $k = 1, 2, \dots, K$;
 a^* ——优化方案。

运用所建立的北京市城市发展宏观综合模型,进行了 9 个首都发展战略方案的系统分析和策略评价,最后比较得出“严格控制人口增长,大力发展城市基础设施,大力整治环境,发展科教事业,抑制工业发展,维持农业生产”的首都发展策略是满意方案,推荐给决策者。在我们的研究成果报告中所分析的 9 个首都发展决策方案均作出详细描述及比较,尤其是对最优策略方案作出更为深入的分析。

研究所得的几项专项成果尤具价值,即在 80 年代中期我们就作出如下预测:

(1) 北京市到 20 世纪 90 年代中期全市人口总量将突破 1000 万,这对于北京市原来一直“规划人口控制在 1000 万以内”是第一次提出人口总量突破的时间,并预计到 21 世纪初期北京市人口总量可能达到 1200 万人。

(2) 水资源将在 21 世纪初期成为制约北京市工农业总产值增长的主要因素,并具体预测出水源的短缺数量和影响强度。

(3) 环境质量将继续恶化,大约到 2015 年前后达到峰值。

(4) 实施大力发展城市基础设施的策略。北京的城市道路密度、城市人口平均居住面积、城区电话普及率将在 10 ~ 15 年内有较大幅度的增长,并给出数量上的预测。

习 题

11.1 MRP 和 ERP 在实现过程中应该注意什么问题,两者的应用对象有何区别?

11.2 复杂系统在分解为不同的子系统进行研究时,如何划分子系统?子系统之间的关系如何把握?
如何确定主导子系统和其在整个系统中的作用?

11.3 复杂系统的评估准则如何确定?如何根据复杂评估准则确定最优方案?

11.4 在区域规划中所遵循的研究步骤包括哪些,各部分的主要内容是什么?

参 考 文 献

- 1 L. V. Bertalanffy 著, 秋同, 袁嘉新译. 一般系统论. 北京: 社会科学文献出版社, 1987
- 2 A. N. 乌约莫夫, 系统方式和一般系统论. 长春: 吉林出版社, 1983
- 3 《日本工业标准》, 运筹学用语“JIS-Z-8121”
- 4 钱学森等. 论系统工程. 长沙: 湖南科学技术出版社, 1982
- 5 陈禹六. 大系统理论及其应用. 北京: 清华大学出版社, 1988
- 6 钱学森, 于景元, 戴汝为. 一个科学新领域——开放的复杂巨系统及其方法论. 自然, No. 1, 1990
- 7 钱学森. 再谈开放的复杂巨系统. 模式识别与人工智能, No. 1, 1991
- 8 于景元, 钱学森. 关于开放的复杂巨系统的研究, 系统工程理论与实践, 第 12 卷 5 期, No. 5, 1992
- 9 夏绍玮, 杨家本, 杨振斌. 系统工程概论. 北京: 清华大学出版社, 1995
- 10 秋山穰, 西川智登著, 高烈夫译. 系统工程. 北京: 机械工业出版社, 1983
- 11 魏宏森, 曾国屏. 系统论——系统科学哲学. 北京: 清华大学出版社, 1995
- 12 A. D. Hall, a Methodology for System Engineering, Van Nostrand, 1962
- 13 程理民, 吴江, 张玉林. 运筹学模型与方法教程. 北京: 清华大学出版社, 2000
- 14 蔡宝三. 最优化与最优控制. 北京: 清华大学出版社, 1983
- 15 汪应洛. 系统工程导论. 北京: 机械工业出版社, 1981
- 16 刘豹, 顾培亮, 张世英. 系统工程概论. 北京: 机械工业出版社, 1987
- 17 魏国华, 傅家良, 周仲良. 实用运筹学. 上海: 复旦大学出版社, 1987
- 18 谭家华. 管理运筹学基础. 上海: 上海交通大学出版社, 1992
- 19 杨山. 线性网络分析. 北京: 高等教育出版社, 1987
- 20 谭跃进, 陈英武, 易进先. 系统工程原理. 长沙: 国防科技大学出版社, 1999
- 21 W. Isard, Introduction to Regional Science, Prentice-Hall, 1975
- 22 I. S. Lowry, A Model of Metropolis, the RAND Corporation, 1964
- 23 I. S. Lowry, Seven Models of Urban Development; A Structural Comparison, the RAND Corporation, 1967
- 24 UN Center for Human Settlements(HABITAT), Methods for the Allocation of Infrastructure Investments, Nairobi, 1991
- 25 Federal Ministry for Regional Planning, Building and Urban Development, Guidelines for Regional Planning-General Principles for Spatial Development in the Federal Republic of Germany, 1993
- 26 P. M. Allen, Cities and Regions as Self-Organizing Systems; Models of Complexity, Cranfield University, 1994
- 27 谢文慧等. 城市经济学. 北京: 清华大学出版社, 1996
- 28 国务院发展研究中心课题组. 中国跨世纪区域协调发展战略. 北京: 经济科学出版社, 1997
- 29 P. Treuner, 张永谦等. 江苏沿江城镇经济发展研究. 江苏经济探讨(增刊), 1994
- 30 W. Christaller, Central Places in Southern Germany, Prentice-Hall, 1966
- 31 Colin Lee, Models in Planning, Pergamon Press, 1973
- 32 <http://www.erc.cims.edu.cn>
- 33 王英. 工业技术经济. 北京: 北京理工大学出版社, 1992
- 34 周裕新, 李国好, 蔡国强. 技术经济学. 广州: 华南理工大学出版社, 1988
- 35 傅家骥, 仝允桓. 工业技术经济. 北京: 清华大学出版社, 1991
- 36 周惠珍. 投资项目经济评价. 北京: 中国审计出版社, 1997
- 37 Singh M. G, Titli A., Systems: Decomposition, Optimization and Control, Pergamon Press

Oxford, 1978

- 38 Man K F, Tang K S, Kwong S and Halang W A. Genetic Algorithms for Control and Signal Processing, Springer-Verlag, 1997
- 39 Goldberg D E. Genetic Algorithms in Search, Optimization and Machine Learning, Addison Wesley Publishing Company, 1989
- 40 邢文训, 谢金星等. 现代优化计算方法. 北京: 清华大学出版社, 1999
- 41 孙增圻等. 智能控制理论与技术. 北京: 清华大学出版社, 1997
- 42 周东华, 叶银忠. 现代故障诊断与容错控制. 北京: 清华大学出版社, 2000
- 43 汪国强. 数学建模优秀案例选编. 广州: 华南理工大学出版社, 1998
- 44 邓先礼. 最优化技术. 重庆: 重庆大学出版社, 1998
- 45 韦鹤平. 最优化技术应用. 上海: 同济大学出版社, 1987
- 46 孙德敏. 工程最优化方法及应用. 合肥: 中国科技大学出版社, 1997
- 47 施光燕, 董加礼. 最优化方法. 高等教育出版社, 1999
- 48 解可新, 韩立兴, 林友联. 最优化方法. 天津: 天津大学出版社, 1997
- 49 蒋慰孙. 系统控制. 上海: 华东化工大学出版社, 1988
- 50 冯. 贝塔朗菲著, 林康义, 魏宏森译. 一般系统论基础发展和应用. 北京: 清华大学出版社, 1987
- 51 H. 哈肯著, 戴鸣钟译. 协同学——自然成功的奥秘. 上海: 上海科学普及出版社, 1988
- 52 H. 哈肯著, 杨家本译. 协同计算机和认知——神经网络的自上而下方法. 北京: 清华大学出版社, 1994
- 53 H. 哈肯著, 高等协同学. 北京: 科学出版社, 1989
- 54 詹姆斯. 格莱克著, 张彦等译. 混沌学. 北京: 社会科学文献出版社, 1991
- 55 成思危. 复杂性科学探索. 北京: 民主与建设出版社, 2000
- 56 伊. 普里戈金, 伊. 斯唐热著, 曾庆宏, 沈小峰译. 从混沌到有序——人与自然的新对话. 上海: 上海译文出版社, 1987
- 57 金观涛. 整体的哲学. 成都: 四川人民出版社, 1987
- 58 魏宏森. 系统理论及其哲学思考. 北京: 清华大学出版社, 1988
- 59 郑维敏. 正反馈. 北京: 清华大学出版社, 1998
- 60 王雨田. 控制论、信息论、系统科学与哲学. 北京: 中国人民大学出版社, 1988
- 61 钟义信. 信息科学原理. 北京: 北京邮电大学出版社, 1996
- 62 吴大进等著. 协同学原理和应用. 武汉: 华中理工大学出版社, 1990
- 63 陈廷. 决策分析. 北京: 科学出版社, 1987
- 64 王金山, 谢家平. 系统工程基础与应用. 北京: 中国地质出版社, 1996
- 65 冯允成等编. 系统工程基础. 北京: 北京航空航天大学出版社, 1995
- 66 承泓良, 林锦国. 灰色系统决策与应用. 南京: 江苏科学技术出版社, 1989
- 67 王莲芬, 许树柏. 层次分析法. 北京: 中国人民大学出版社, 1990
- 68 沈泰昌, 谷宝贵, 谢秉兰. 系统工程基础. 北京: 国防工业出版社, 1981
- 69 亨利 E. J. 等著, 吕应中等译. 可靠性工程与风险分析. 北京: 原子能出版社, 1988
- 70 魏世孝. 兵器系统工程. 北京: 国防工业出版社, 1989
- 71 王玉麟. 动力装置系统可靠性. 北京: 北京航空航天大学出版社, 1995

Images have been losslessly embedded. Information about the original file can be found in PDF attachments. Some stats (more in the PDF attachments):

```
{
  "filename": "MTE5MTM4Mjguemlw",
  "filename_decoded": "11913828.zip",
  "filesize": 22541459,
  "md5": "e006059f41ed13b5378bec7299f4efcd",
  "header_md5": "7d68fcf6c74d1678f1f752747b36c780",
  "sha1": "3486e9cdfbe6d37a32423c424c29cfa8192dea9a",
  "sha256": "5b04c968e000754e49329206a69c75ad62ae8390fe6394fa3945e1c930212acf",
  "crc32": 167525238,
  "zip_password": "julian",
  "uncompressed_size": 23665575,
  "pdg_dir_name": "11913828",
  "pdg_main_pages_found": 246,
  "pdg_main_pages_max": 246,
  "total_pages": 262,
  "total_pixels": 1675856816,
  "pdf_generation_missing_pages": false
}
```