

# 自组织数据挖掘 与经济预测

贺昌政 著

国家社会科学基金资助

科学出版社

北京

谨以此书献给 GMDH 之父  
——A. G. Ivakhnenko 院士

## 内 容 简 介

本书阐述了自组织数据挖掘与经济系统复杂性研究的关系,以及对经济预测方法完善的贡献,涵盖了作者对自组织数据挖掘的改进与发展工作,同时选入了作者较多的运用自组织数据挖掘方法研究中国经济问题的实例.

本书可供经济、管理、系统科学、应用数学及系统工程等专业的本科生、研究生和教师阅读,也可作为中高层管理人员的参考书.

### 图书在版编目(CIP)数据

---

自组织数据挖掘与经济预测/贺昌政著. —北京:科学出版社,2005

ISBN 7-03-015896-2

I. 自… II. 贺… III. 数据采集-计算机应用-经济预测-中国  
IV. F123.2-39

中国版本图书馆 CIP 数据核字(2005)第 077030 号

---

责任编辑:张 扬 祖翠娥 / 责任校对:李奕莹  
责任印制:钱玉芬 / 封面设计:王 浩

科 学 出 版 社 出 版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

中国科学院印刷厂印刷

科学出版社发行 各地新华书店经销

\*

2005 年 9 月 第 一 版 开本:B5(720×1000)

2005 年 9 月 第一次印刷 印张:12

印数:1—3 000 字数:221 000

定价:30.00 元

(如有印装质量问题,我社负责调换(环伟))

# 前 言

经济系统是演化的复杂系统,我国的社会主义市场经济系统是一个全新的复杂经济系统.经济预测实践提出了新的问题,为了解决这些问题,必须有新的方法.同时,新方法可以使我们重新审视已知问题的解决途径.自组织数据挖掘由乌克兰科学院 A. G. Ivakhnenko 院士创立,建立在“进化—遗传—变异—选择”的进化论原理基础上,其建模方法强烈地体现着由简单到复杂的事物演化过程,是基于复杂系统的前沿学科.站在复杂性科学的高度,用自组织数据挖掘完善经济预测方法,将丰富新形势下的经济预测方法,为我国国民经济实现科学决策做贡献.

相对于计量经济学方法和人工神经网络方法的庞大研究队伍及其丰富的研究成果,自组织数据挖掘的研究工作可以说才开始,在基础理论研究方面还需要加强,方法上还要进一步完善.然而,它运用于经济预测实践的特色吸引我们愿意花力气去探索、去体味.本书以自组织数据挖掘为基础完善经济预测方法,并用之于我国社会主义市场经济预测实践,同时希望给我国经济预测工作者展示一个极富特色、有发展前景、有用武之地的新领域.

对于经济系统,其特征是组织的产生(经济对象的形成与设计、改造、重构等),动态平衡的自动调节(为了达到指定目的而控制经济过程),管理、组织的变革,所以属于自组织系统.为了得到具有要求性质的数学模型,自组织数据挖掘方法中借用了育种学家的大批量育种方式.大规模选择是人类有史以来最古老和最富有成效的启发式研究方法之一,这种途径的思想在于,在生物界要繁殖下一代,就要拥有具有最好特性的但还需要继续完善的那些生物.利用自组织数据挖掘方法建立经济系统动态模型时,根据影响因素的样本及建模者感兴趣的输出变量,在计算机上采用人机对话方式产生大量的竞争模型;用恰当的选择准则(外准则)选出一部分“最有希望”的模型,再利用这些模型产生大批新的竞争模型.遵循这样的方式将模型的结构从简单到复杂逐步改进,最后产生出具有“最优复杂度”的模型.

现代科学的发展扩大了应用大规模育种方法的领域,基于大规模育种的自组织数据挖掘方法可以解决不同知识领域内的新问题,如复杂系统的辨识、模式识别、最优化、人工智能等.使用自组织数据挖掘方法去建立经济过程的模型,不仅扩

大了大规模育种方法的应用领域,而且为解决那些用传统的数理统计方法也无能为力

的问题提供了一条新的思路.

20 世纪 90 年代初,四川大学工商管理学院刘光中教授把自组织数据挖掘方法引入国内,并组织了研讨班.在深入研究中我们对该学科领域产生了强烈的兴趣,并作为《数学建模导论》(成都科技大学出版社,1997 年)的一个重要内容在国内首次对它作了介绍.近年来,我们完善和发展了自组织数据挖掘的理论和方法,并提出了若干新的自组织数据挖掘预测模型,完善了经济预测方法.本专著收入了其中的部分工作成果,如自组织组合预测模型和 GMDH 因果检验模型等.基于自组织经济预测方法,我们对一系列经济热点问题从新的视角提出了新的见解和建议,这部分内容可参见人民币汇率、全国及地区宏观经济的分析预测等实证研究.本专著提出并阐述的如下两个观点是对复杂性科学研究的贡献:①自组织数据挖掘是从定性到定量综合集成的实现技术.②自组织数据挖掘是进行经济系统复杂性研究的一条有效途径.中国正处在经济转轨时期,这个经济系统比其他国家更显复杂性,经济指标数据常显示样本量小且受到干扰的特点.自组织数据挖掘方法在小样本、受到噪声干扰的经济系统建模预测时比计量经济学方法和人工神经网络方法更优越,它是中国社会主义市场经济分析预测的一个值得关注的方法.因此,可以认为我们的研究工作是具有学科前沿并针对中国国情的经济预测方法研究,它给经济预测研究领域开辟了一个新方向.

本书共 7 章.第 1 章从经济是一个复杂的演化系统的观点出发,对现有的预测方法、尤其是定量预测方法进行了评述.在比较分析了现在常用的预测方法后指出:对经济过程的预测,除了使用现有预测方法之外,还应该使用新范式去思维,使用新方法去预测.第 2 章简要介绍了自组织数据挖掘的基础——启发式自组织方法,并对自组织数据挖掘的基本思想、基本原理及它的最终模型——最优复杂度模型进行了详细的探讨.第 3 章讨论了自组织数据挖掘最基本的算法——参数 GMDH 算法基本类型、基本步骤及算法的外准则体系,同时通过与人工神经网络方法和回归分析方法的比较揭示了自组织数据挖掘方法的特点.第 4 章阐述了自组织数据挖掘与经济系统复杂性研究的关系.第 5 章着重论述了自组织数据挖掘对经济预测方法完善的贡献.这两章还讨论了一些有特色的方法:自组织两水平算法、客观系统分析方法(OSA)及 3 种非参数算法(OCA, AC 及 FRI),内容涉及它们的基本原理、算法步骤及实例.这些方法和参数 GMDH 算法在经济预测方法完善中扮演重要角色.第 6、7 两章是自组织数据挖掘方法在经济预测方面的一些实例及

其研究展望.本书的实例主要来自我们的研究实践.

自组织数据挖掘的国际学术权威——乌克兰科学院 A. G. Ivakhnenko 院士和德国学者 J. A. Mueller 教授提供了他们的最新研究资料,同我们进行了数次有益的讨论,使这本书能反映该领域当前国际前沿研究成果,在此谨向他们致以深深的谢意.这本书的问世首先应该提及刘光中教授的贡献.此外,我还要感谢我的同事和研究生们在成书过程中给我的支持和帮助,祝愿他们心想事成.

贺昌政

2005年6月于

四川大学望江校区

# 目 录

<b>第 1 章 经济过程的预测问题</b> .....	<b>1</b>
1.1 经济是一个演化的复杂系统 .....	1
1.1.1 复杂性科学研究 .....	1
1.1.2 经济系统的演化复杂性 .....	5
1.1.3 提高经济预测质量的途径 .....	6
1.2 预测方法评述 .....	8
1.2.1 一般预测方法 .....	8
1.2.2 主要定量预测方法评述 .....	10
1.2.3 预测研究新范式 .....	12
参考文献 .....	15
<b>第 2 章 自组织数据挖掘基础</b> .....	<b>16</b>
2.1 数据挖掘概述.....	16
2.1.1 建模方法:基于理论的方法和基于数据的方法 .....	16
2.1.2 数据挖掘.....	17
2.1.3 数据挖掘与知识提取 .....	19
2.2 启发式自组织方法.....	20
2.2.1 基本原则.....	20
2.2.2 在启发式自组织基础上建模的一般步骤 .....	23
2.2.3 模型结构数 .....	24
2.2.4 关于术语“自组织”的含义.....	26
2.3 自组织数据挖掘.....	27
2.3.1 模型的自动产生 .....	27
2.3.2 自组织数据挖掘建模的基本思想 .....	28
2.3.3 统计学习网络 .....	29
2.3.4 归纳途径——GMDH 算法 .....	31
2.3.5 自组织数据挖掘的基本原理.....	32
2.3.6 最优复杂度模型 .....	36
参考文献 .....	39
<b>第 3 章 自组织数据挖掘算法</b> .....	<b>40</b>
3.1 自组织数据挖掘算法的函数流程.....	40

3.2	参数 GMDH 算法	41
3.2.1	GMDH 算法实现步骤	42
3.2.2	GMDH 的基本算法	43
3.2.3	选择准则	49
3.3	GMDH 与人工神经网络方法比较研究	55
3.3.1	对系统先验知识的应用	55
3.3.2	归纳学习的算法过程	56
3.3.3	模型的推广能力	57
3.3.4	例——系统辨识	57
3.4	GMDH 与回归分析比较研究	58
3.4.1	算法假设的比较	58
3.4.2	算法过程的比较	59
3.4.3	对系统拟合和预测效果的实验比较	60
3.4.4	GMDH 与回归分析方法的联系	61
3.5	算法的内收敛性	61
3.5.1	外准则的标准型	62
3.5.2	输出传播的多层算法	62
3.6	活动神经元	65
	参考文献	66

**第 4 章 自组织数据挖掘与经济系统复杂性研究 68**

4.1	揭示经济对象构成因素及较准确的预测功能	68
4.2	自组织数据挖掘与非物理模型	70
4.3	从定性到定量综合集成方法的具体实现技术	72
4.3.1	专家知识的综合集成;算法的初始输入模型	72
4.3.2	人机对话的语言;外准则	73
4.3.3	专家经验判断与计算机运算相结合;GMDH 算法过程	74
4.4	客观系统分析	74
4.4.1	基本原理	75
4.4.2	基本步骤	75
4.4.3	实例:中国宏观经济模型	77
4.5	客观聚类分析方法	81
4.5.1	基本原理	82
4.5.2	基本步骤	82
4.5.3	实例:基于信息基础设施的国家聚类分析	84
4.6	相似体合成算法	87



4.6.1	基本原理	87
4.6.2	基本步骤	88
4.6.3	实例:中国能源消费预测	90
4.7	自组织模糊规则归纳法	92
4.7.1	工作原理	92
4.7.2	基本步骤	93
4.7.3	实例:四川 GDP 增长的主要影响因素分析	95
	参考文献	97
<b>第 5 章</b>	<b>基于自组织数据挖掘的经济预测方法改进</b>	<b>98</b>
5.1	两水平算法及预测效果改善问题	98
5.1.1	GMDH 两水平算法工作原理	99
5.1.2	GMDH 两水平算法基本步骤	100
5.1.3	两水平算法扩大可预测范围的研究	101
5.2	因果关系检验模型	106
5.2.1	Granger 因果关系检验模型	107
5.2.2	基于 GMDH 的因果关系检验模型	107
5.3	基于 AC 算法的经济预测	110
5.3.1	AC 算法在股市预测中的应用	110
5.3.2	成都市 GDP 的预测	113
5.4	自组织组合预测	114
5.4.1	组合预测方法评述	114
5.4.2	自组织组合预测的步骤	115
5.4.3	应用实例:成都市消费水平预测	116
5.5	基于 DFA 的自组织数据挖掘模型	118
5.5.1	预测的困难性	118
5.5.2	数据预分析对预测的必要性	119
5.5.3	DFA 方法简介	120
5.5.4	实证分析	121
	参考文献	126
<b>第 6 章</b>	<b>自组织数据挖掘在经济预测中的应用</b>	<b>127</b>
6.1	小麦收成预测	127
6.2	成都市居民未来生活用水量预测模型的选择	128
6.2.1	自回归和层次分析供水量预测模型	128
6.2.2	利用自组织数据挖掘方法建立模型	131
6.2.3	预测模型的比较分析	131

6.2.4	成都市居民生活用水量预测 .....	132
6.3	人民币汇率影响因素及预测分析研究 .....	133
6.3.1	人民币实际汇率影响因素的专家见解 .....	134
6.3.2	人民币实际汇率的自组织模型 .....	136
6.3.3	结果分析 .....	138
6.4	自组织数据挖掘方法对电力需求的预测 .....	140
6.4.1	问题的提出 .....	140
6.4.2	城市电力需求静态预测模型 .....	141
6.4.3	成都市电力需求动态预测模型 .....	143
6.5	基于 AC-FRI 的经济预警体系 .....	143
6.5.1	AC-FRI 经济预警体系工作原理 .....	144
6.5.2	AC-FRI 经济预警体系基本步骤 .....	145
6.5.3	四川省经济预警系统 .....	145
6.5.4	结论 .....	150
6.6	新股上市定价的自组织模型 .....	151
6.6.1	问题的提出 .....	151
6.6.2	模型设计 .....	151
6.6.3	模型结果分析 .....	153
6.7	FRI 方法在市场调查分析中的应用研究 .....	155
6.7.1	FRI 方法在市场调查分析中的基本步骤 .....	156
6.7.2	仅用 FRI 的市场调查分析实证研究 .....	157
6.7.3	聚类分析方法与 FRI 结合进行市场调查分析的实证研究 .....	162
	参考文献 .....	165
<b>第 7 章</b>	<b>研究展望 .....</b>	<b>167</b>
7.1	自组织数据挖掘的方法论 .....	167
7.2	自组织数据挖掘方法改进综述 .....	167
7.2.1	对自组织数据挖掘本身的改进研究 .....	168
7.2.2	自组织数据挖掘与其他方法的结合 .....	171
7.3	值得进一步研究的问题 .....	173
	参考文献 .....	175
<b>索引 .....</b>		<b>178</b>

# 第 1 章 经济过程的预测问题

复杂性科学的观点认为，经济系统是一个复杂的演化系统。对经济过程的预测，除了使用传统预测方法之外，还应该使用新范式去思维，使用新方法去预测。

## 1.1 经济是一个演化的复杂系统

### 1.1.1 复杂性科学研究

复杂性科学在 20 世纪 80 年代中期兴起，是一门主要研究复杂系统和复杂性的科学。它是还原论、经验论及“纯科学”为基础的经典科学吸收系统论、理性论和人文精神而发展成的一门新学科，被誉为是“21 世纪的科学”。

什么是复杂性？

Arthur 等人<sup>[1]</sup>指出，狭义复杂性包括以下 6 个特性：

- (1) 在仅发生局部作用的异类作用者之间存在离散的相互作用；
- (2) 在经济系统中没有能够利用所有机会和相互作用的全局控制者，尽管可能存在某些弱的全局相互作用；
- (3) 纵横交错的多层次组织结构，它们之间有混乱的交互作用；
- (4) 通过学习和演化能连续地适应环境；
- (5) 随着新市场、新技术、新行为和新制度的出现，不断创造出“生态”系统中新的生物，从而保持系统永远新颖；
- (6) 具有远离均衡，或者没有均衡，或者有许多均衡的动力学特性，并且系统不太可能达到全局最优。

Michael R. Lissack<sup>[2]</sup>认为复杂性体现在以下诸方面：

- (1) 非线性，即因果之间不成比例；
- (2) 分形性，分形的测算是标度依赖的，概念是模糊的；
- (3) 递归性，标度水平之间是递归的；
- (4) 对初始条件的敏感依赖性；
- (5) 充满反馈和可能的分岔点，以及服从涌现。

Casti<sup>[3]</sup>把涌现 (emergence) 定义为系统的整体行为，该行为产生于众多参与者的相互作用，而且利用系统中任一成员孤立状态下个体表现的知识不能对其

进行预测甚至想像。

钱学森认为,开放的复杂巨系统的以下动力学特性称为复杂性、巨型性、内在差异性、层次性、开放性和动态性<sup>[4]</sup>。

成思危<sup>[5]</sup>指出,复杂性主要表现为5个方面:各单元间网状联系,多层次、多功能结构,会学习能自我重组及完善,向环境开放并与之相互作用,动态性并有某种程度的自我预测能力(广义地说是具有某种智能)。综上可知,系统复杂性有如下体现:

(1) 系统各单元之间的联系广泛而紧密,构成一个网络。因此每一单元的变化都会受到其他单元变化的影响,并会引起其他单元的变化。

(2) 系统具有多层次、多功能的结构,每一层次均成为构筑其上一层次的单元,同时也有助于系统的某一功能的实现。

(3) 系统在发展过程中能够不断地学习并对其层次结构与功能结构进行重组及完善。

(4) 系统是开放的,它与环境有密切的联系,能与环境相互作用,并能不断向更好的适应环境的方向变化。

(5) 系统是动态的,它不断处于发展变化之中,而且系统本身对未来的发展变化有一定的预测能力。

在几乎所有的研究领域中都能举出复杂系统的例子:工程中,一个有很多变量并且变量之间相互制约的系统被称为复杂系统;数学家则列出一组非线性微分方程;物理学家对为何雪花能形成美丽的六角形仍是解释不清;生命世界的多样性则使理论生物学家陷入了困境;经济学家对股票市场的解释亦不能使人信服……。

SFI认为<sup>[6]</sup>,复杂系统是由大量相互作用的单元构成的系统。

成思危指出<sup>[5]</sup>,复杂系统最本质的特征是其组分具有某种程度的智能,即具有了解其所处的环境,预测其变化,并按预定目标采取行动的能力。

Richard Gallagher 和 Tim Appenzeller 指出<sup>[7]</sup>,通过对一个系统分量部分(子系统)的了解,不能对系统的性质作出完全的解释,这样的系统称为复杂系统。即对于复杂系统,整体的性质不等于部分性质的和,或系统整体与部分之间的关系不是一种线性关系。

钱学森<sup>[8]</sup>提出了开放复杂巨系统的概念,这种系统具有开放性、复杂性和动态特征。

(1) 开放性的含义指:系统不可避免地会受外界的影响。如铁磁体的磁性随外界温度的变化而产生变化,同时系统具有主动适应性和进化性。以经济系统为例,在某种程度上,经济系统可以和生命进化系统相类比。区别在于经济系统中的个体具有一定的“预见性”,这种预见性来源于个体间的交互(如通过学习获

取知识)及个体和外部环境的交互(这方面最明显的例子就是股票市场)。由于个体的“预见性”,个体可以根据外部环境的变化,主动地、适应地改变自己的决策方法和行为。子系统之间的关系不仅复杂而且还随时间及情况有极大的易变性。

(2) 开放复杂巨系统中“复杂”的含义指:子系统的种类繁多,系统的层次复杂,子系统之间相互联系与作用很强。

(3) 开放复杂巨系统动力学中,“动力学特性”指:一方面系统的宏观特性连续地随时间变化,如吸引子位置的改变,是量变的过程;另一方面,系统的宏观特性随系统进化的过程而变化,如系统从无序状态形成有组织的行为或者从有序状态走向混沌,是质变的过程。

20 世纪的科学研究,如物理学,其研究方法是“还原论(reduction)”,即研究单一个体。例如,核物理的研究是以原子核为对象;DNA 的研究,是以细胞为对象;力学的研究,以受力体为对象,等等。对于复杂系统,这样做不行。如金融股市,不能研究个别股民,而要研究股民的总体行为;对思维过程,不能研究个别神经元,而要研究神经网络;而胚胎的形成,不能研究单个细胞等。这就要从方法论上进行创新。还原论在研究一些系统的行为时是有效的,但在处理复杂系统时,它却有着以下的局限性:

(1) 单元的行为难以分析:还原论要求独立地分析每个单元的行为,而在复杂系统中,由于单元间的关系很复杂,无法将某个单元与其他单元分离开独立地分析。

(2) 单元间的关系或相互作用难以明确:在研究人的大脑智能产生机理时,我们知道神经元与神经元间依靠突触相互作用,而且这种作用和智能的产生有密切的关系。但一个困难的问题是,我们无法在大脑进行智能活动时观察突触的作用,只能通过某些方法(如离体检测等)间接地了解突触的作用。

1984 年,美国有几位诺贝尔奖金获得者,认识到复杂系统的重要意义,聚集了一批物理、经济、生物、计算机等方面的学者,在美国新墨西哥州 Santa Fe 组织了一个松散的研究团体,这就是著名的桑塔费研究所(Santa Fe Institute, SFI),由 G. A. Cowan 负责。其主要研究目标即是对复杂性的研究。他们分别在所从事的领域中开展对不同复杂系统的研究,寄希望于通过多学科交叉研究,进一步形成有关复杂系统的一般性理论,即复杂性理论。

SFI 的科学家们对复杂性用了各种名词加以描述,并将研究复杂系统的学科称为复杂性科学(complexity science)。从 1984 年 SFI 成立以来,在复杂性方面的研究取得较大的进展,其成就主要包括:一是建立独特的、有关复杂性研究的学术体系;二是为克服还原论的局限性,从多学科研究复杂系统得到的结果出发,概括总结了分析复杂系统的一套方法,即自动机网络(automata network,

AN) 方法. AN 方法的特点是:

(1) 不认为在了解系统的宏观行为时, 每个单元的细节是重要的. 因此, 可以对复杂系统的单元进行简化, 采用充分简单的单元模型.

(2) 认为主要是单元的数目, 即系统的规模对系统的宏观特性有决定性的影响. 因此, 可采用充分简化的模型模拟单元间的相互作用.

(3) 着重于观察与研究系统的宏观特性.

早期 SFI 的主要学术观点可概括为: 复杂系统是由大量相互作用的单元构成的系统. 复杂性的研究内容则是研究复杂系统如何在一定的规则下产生有组织的行为. SFI 的科学家对复杂性的说法恰与混沌理论研究内容相反 (系统根据简单规则产生混沌行为), 所以 SFI 又将其复杂性研究称为反混沌. 近年来, SFI 的一些科学家, 如 Kauffman 等拓宽了复杂性的研究内容, 把兴趣逐步转移到对混沌边界的研究上. 在我国, 钱学森教授于 20 世纪 80 年代组织领导了复杂系统的研讨班. 钱学森教授是战略科学家, 以其深刻的洞察力预见到复杂系统的意义及发展. 他提出了“开放的复杂巨系统”的概念, 并于 1992 年对复杂系统的研究方法提出“从定性到定量的综合集成研讨体系”的设想. 戴汝为、于景元教授是研讨班的主要成员, 并在以后继承和领导了这个研究方向. 成思危教授领导了在管理科学方面复杂系统的研究. 复杂性研究走向高潮的主要表现如下:

(1) 复杂性研究目前已遍及所有发达国家及中国等发展中国家, 成为一种具有世界规模的科学思潮, 一种文化运动.

(2) 目前的复杂性研究称得上学派林立、观点纷呈、新见迭出, 已经有大量著作问世, 有关文献在加速增长.

(3) 当复杂性研究仅由那些从事新型科学、交叉科学、应用科学的学者倡导和推进时, 它的档次总显得不够高, 引不起主流科学界的注意, 对科学的总体影响有限. 这种局面在逐步改变, 有 3 个动向特别有意义. 首先是诺贝尔奖得主 I. Prigogine 和 M. Eigen、有世界性影响的科学家 H. Haken 等人先行介入, 基于现代物理学和生物学成就来探索复杂性, 引起广泛关注, 标志着基础科学层次开始了复杂性探索. 接着是另外三位影响更大的诺贝尔奖得主 Gellman、Anderson 和 Arrow 大力推动建立桑塔费研究所, 被称为“老师倒戈”, 给世界科学界带来的震动非同凡响. 1999 年 Science 杂志的专集《复杂系统》, 精心组织一批正在物理、化学、生物、经济、生态、地理环境、气象、神经科学等主流科学主战场前沿工作的著名学者探讨各自领域的复杂性, 有力地表明在世纪之交主流科学界对复杂性研究的明确关注和认可, 并直接参与进来, 这是极有意义的进展.

从研究范畴看, 目前的复杂性科学研究可从物理、生物和社会等 3 个层次开展研究. 同时, 还需要开展这 3 个层次之间的共性研究, 即关于复杂系统与复杂性的理论与方法的研究.

### 1.1.2 经济系统的演化复杂性

20 世纪 80 年代末 90 年代初, 几位诺贝尔奖得主和数学大师对经济分析提出了一个崭新的思路, 即经济是一个演化的复杂系统. 1988 年, 诺贝尔物理学奖得主 P. Anderson 和诺贝尔经济学奖得主 K. J. Arrow 组织了一个专题讨论会, 主题即经济可看作一个演化着的复杂系统. 参加讨论的有数学、物理、经济、生命科学和计算机等多方面的专家. P. Anderson 和 K. J. Arrow 还给出了一个对演化的经济进行描述的基本思路<sup>[9]</sup>: 设想经济系统可能存在内在的核心动力机制, 并且这种机制可以由少维变量和参量的子系统表示且支配整个经济的发展演化行为.

经济是一个演化的复杂系统. 经济系统包含了上千个变量和参量, 它们之间相互联系、相互作用, 构成了一幅非线性的图像. 这是一个高维的系统, 要对这样多的变量和参量进行分析和计算, 不仅在实际上行不通, 而且其计算的结果也很难检验. 所以, 一个好的经济理论或模型通常是将实际的经济投影到一个恰当的子空间上去, 这个子空间具有较低的维数, 但反映了所讨论的经济问题的本质特性. 能否在经济的动态演化行为中也找到一个恰当的抽象, 这是对人类智慧的一个挑战. P. Anderson 和 K. J. Arrow 已经预言了这种少维空间动力学抽象的可能性. 在其他领域中, 很多非线性动力学模型已经为各种具体的复杂对象找到了其动力学规律以作为参考, 但经济系统本身的动态抽象还有待去揭示.

经济系统的复杂性也体现在它的层次结构. 经济系统有很多的层次, 每一个层次都有其自身的结构, 如整体的经济、各部门经济及所属的企业、工厂等. 每一个经济单位, 不论是整体、部门或是小到一个企业、工厂单位, 都按其经济结构的性质实现其自身的功能. 在这些经济结构中最基本的一项功能就是实现其利益的最大化. 但是对于一个多层次的经济结构, 各个层次之间的利益协调就成为经济系统复杂性的本质问题之一.

经济信息的不完备性和不确定性, 以及经济过程中人的参与, 使得经济学家从来没有对经济系统的运行规律、运行机制及构成原因形成统一的理论, 而是各种学派并存, 仁者见仁, 智者见智. 例如, 经济学界关于我国近年来通货紧缩的成因有三种不同观点: 其一是供需失衡论, 强调通货紧缩是由实体经济因素所致, 包括生产过剩论和有效需求不足论. 其二是货币紧缩论, 认为通货紧缩是一种货币现象, 是由紧缩性货币政策导致. 其三是债务紧缩论, 认为通货紧缩是由债务收缩引起的总需求下降而导致的. 哪种成因观点是正确的, 是其中之一, 还是各种观点的综合? 经济系统的复杂性研究是现实而重要的问题.

Arthur 指出<sup>[10]</sup>, 经济是一个极其复杂的自组织系统, 不同的经济领域要受不同的规律支配. 例如, 高技术几乎可以被定义为“凝结的知识”, 它的边际成

本几乎为零，加之用户具有群体使用标准化产品的倾向，致使该领域服从报酬递增率，高科技市场的赢者能占有整个市场，从而向新古典经济学所描述的“经济是和谐、稳定和均衡的，遵从负反馈或报酬递减率”的观点发出了强有力的挑战。

对经济规律的把握与政策考虑，是各层次经济管理人员关注的焦点。对于研究工作来说，把经济建立在科学分析的基础上是各学科学者们的共同愿望。解决系统的演化规律不仅使人们对经济系统的性质有本质上的突破和认识提高，而且对于其相关的学科，如数学、物理、系统科学、计算机科学等也有实质上的促进。

### 1.1.3 提高经济预测质量的途径

经济预测是指在经济领域应用预测的理论和方法探索研究经济现象未来的发展情况。它以经济理论为指导，以统计调查资料和经济信息为依据，分析研究经济现象的发展变化规律，运用哲学方法、统计学方法、数学方法、逻辑方法、计算机技术等科学方法，对未来不确定的经济现象的发展前景作出某种推断，是一门在综合经济学、哲学、统计学、数学、数量经济学和技术经济学以及工程技术等各学科研究成果的基础上逐步形成、发展起来的边缘学科。

市场经济条件下，市场机制（包括供求机制、竞争机制、价格机制、风险机制等）在资源配置中发挥着基础性作用，无论微观经济还是宏观经济，都需要研究这些机制作用的力度和方向，找到其中的规律，并运用规律指导行动。因此，现今无论是政府机关还是政策研究部门，或是个体预测机构，对经济变量（如GDP、失业、消费、投资、价格、利率等）的预测都已成为一项例行工作。经济预测贯穿于经营主体的整个决策、经营与计划过程中。只有对经济发展前景作出科学准确的预测，才能根据预测结果和对实际情况的分析作出正确决策，为制定经济政策、编制计划及检查政策、计划执行情况提供科学依据。也只有对经济发展前景作出科学的预测，才能提高企业的经济效益和经营管理水平。

历史信息量的规模对经济系统预测的质量有很大的影响，可以从两个方面提高预测的精确度和置信度：其一，采用经济过程比较深刻的历史信息；其二，选择正确描述经济系统发展过程的预测方法及模型类。

第一方面，采用经济过程比较深刻的历史信息。但是，经济过程的特点在于样本量很小，这给增加历史信息的深度带来困难。由于经济信息量太大，指标各具特征，使得即使搜集过去15年的信息也是很复杂的事情。即使可能，经济政策的变化、经济结构的波动及经济周期的转折点等因素也导致了超过15年的信息是没有代表性的。国际货币基金组织研究局世界经济研究处经济学家 Paula R. Demasi 在文献[11]中提出了这样的观点：“原则上说，《世界经济展望》



(WEO) 中的计划与大多数正式的预测一样, 所依据的假设都是预测期间的经济政策不变. 然而, 这样一个假设是难以维持的, 因为随时可以得到的大量市场信息(包括利率、汇率和经济预期指标)都表明, 政策变量将来会发生变化.”在我国, 自 1978 年实行改革开放的政策以来, 计划经济向社会主义市场经济转变, 使得改革开放前的经济信息无助于反映现在的经济变化规律. 此外, 历史信息的价值随着过去时间越长而越来越小, 预测能力不断下降, 产生了信息的“贴现”, 所以沿着这个方向提高预测的精确性有困难.

第二方面, 选择预测方法及模型类. 自预测方法问世以来, 目前在文献中可以见到的各式各样的预测方法已在 200 种以上. 然而, 现有的预测方法在进行经济预测时却显示了某种程度的不适应(见 1.2.2 节). 例如, 传统的统计方法需要预先知道模型结构, 而模拟建模方法需要具备被模拟对象高质量的、非经验的大量信息, 以达到对被建模对象的深刻认识. 但在经济系统中却不容易达到这些要求. 又如人工神经网络模型难以给出结果的经济解释, 且容易产生过拟合. 经济是一个复杂系统, 它的多层次性及层次结构间的关联性对预测方法提出了更高的要求.

复杂性科学把涌现性定义为系统的整体行为, 该行为产生于众多参与者的相互作用, 而且仅利用系统中的成员在孤立状态下个体表现的知识不能对其进行预测甚至想像. 利用这个思想可以从本质上改进经济系统动态建模的精度. 当经济系统中的技术经济指标的观测数据为小样本时, 其本质乃是: 信息量, 指标间的因果关系, 预测背景(对经济规律及经济因素间联系的认识), 以及对指标约束条件的考虑和被研究过程的涌现性. 在规划中, 利用经济过程的全部信息提高了进行系统辨识及预测的能力. 经济指标体系的一般规律既隐含在它的动态变化中, 也存在于其固有的约束及相互联系之中. 在经济系统建模中考虑的影响经济过程的本质因素越多, 能够揭示一般规律性的要素就越多. 换句话说, 历史信息的相对不足可以用考虑的因素的丰富及因素间的联系(预测背景)来补偿.

由此可以得出结论: 离开相互紧密联系的指标体系, 仅对单个指标进行预测不能解决所研究经济过程(其指标是相互联系、互为条件的)的建模问题. 社会经济过程的预测正是着眼于这一点而与其他的过程预测(其中, 指标是相互独立的)相区别. 此外, 为了预测经济系统的单个指标, 运用其他一些指标却未考虑反馈联系, 或估计模型仅用历史信息而未包含与之相互联系的指标体系, 则这样的方法仅能描述现象表面的外部联系, 不能反映其隐藏很深的相互关系. 这样的方法对于短期预测有很好的效果, 但不适于较长期预测. 较长期预测的效益准则是保持指标动态的现实规律, 顾及给定的约束、平衡及预测过程的其他特点.

由于预测背景的所有要求都只在过去时段上满足, 即使预测模型对历史数据有高度精确的逼近, 甚至它的参数估计是用最现代化的方法求出的, 都不能保证

这种预测本身是可靠的。此外，在长期预测中不能仅预测单个指标，这是因为单个指标的预测结果与预测方法密切相关，不同的预测方法将会有不同的预测结果。

## 1.2 预测方法评述

### 1.2.1 一般预测方法

预测，就是根据事物的运动规律推断它的未来。预测作为一项探索未来的活动，有着 3000~4000 年的历史。“凡事预则立，不预则废”，早已成为人们熟悉的道理。而进行系统、有科学依据的技术和经济过程的预测始于 20 世纪 50 年代初期。

对于预测的概念，从理论上讲，就是根据已知事件去推测未知事件。它有广义和狭义之分。广义的预测，不仅包括对目前还没有发生的事件进行推测，还包括对现在已经发生但我们尚未观察到的事件进行预测。而狭义的预测仅仅是对目前尚未发生的事件或对事件的未来情况进行预测。预测是在错综复杂的因素中研究探索事物发展的未来，它超越了专业知识的范畴，突破了自然科学和社会科学的界限，因此是一门综合性学科。将预测的理论和应用于社会、人口、经济、政治、军事、科学技术、气象、地震和环境等领域，结合各领域的特点和发展状况，就会形成社会预测、人口预测、经济预测、政治预测、军事预测、科学技术预测、气象预测、地震预测和环境预测等。

经济预测是将预测的理论和应用于经济系统，将未来学与经济学相结合而产生的一门交叉学科。

预测过程分为几个阶段，每一个阶段解决一个确定的问题：

- (1) 确定预测目标。明确预测对象，确定问题的目的和任务；
- (2) 分析预测对象。揭示对象的结构，分离出本质的因素，建立它们间的隶属关系及其层次和相互联系；
- (3) 收集资料和数据。确定资料和数据来源，对它们进行整理、加工及分析；
- (4) 选择预测方法及其算法。由已知数据选择最适合的预测方法，设计出相应的算法，估计预测的精度；
- (5) 建模。根据观测样本数据建模，进行预测检验；
- (6) 分析预测结果。从经济活动的合理性和经济理论进行分析，以得到正确的结果。还可以对不同的预测结果进行比较分析。

从方法论角度看，在上述各预测阶段中最重要和最难的是第(4)阶段；从逻

辑上看,“选择预测方法”阶段位于分析预测对象、研究其背景之后.为了选择最有效的方法必须预先做以下工作:预测方法的分类,弄清每种方法的特点,列出观测数据表及对预测背景的要求,预测对象数据特征的分类.

根据国内外学者的估计,现在已有 200 种以上的预测方法.最简单的预测方法分类是按预测方法属性分为定性经济预测和定量经济预测.定性经济预测,是指预测者通过调查研究,了解实际情况,凭自己的实践经验和理论、业务水平,对经济现象发展前景的性质、方向和程度作出判断来进行预测,也称为判断预测或调研预测.定性经济预测的准确程度,主要取决于预测者的经验、理论、业务水平及掌握的情况和分析判断能力.定量经济预测,是指根据准确、及时、系统、全面的调查统计资料和经济信息,运用统计方法和数学模型,对经济现象未来发展的规模、水平、速度和比例关系进行测定.由于定量预测和统计资料、统计方法有密切关系,所以也称为统计预测.它包括时间序列预测和因果预测等.进行经济预测时要注意定性分析与定量分析的结合.只有定性描述,对系统行为特性的把握就难以深入、准确.纯粹的定性预测方法往往是在不易应用定量预测方法的情况下使用,如经济现象难以量化;或缺乏必要数据;或有数据,但不易作数学处理.但定性描述是定量描述的基础,定性认识不正确,无论定量描述多么精确漂亮都没有用,甚至会把认识引向歧途.另一方面,借助定量描述能使定性描述深刻化、精确化.定性描述与定量描述的结合,是系统研究的基本方法论之一.许多优秀的预测研究成果发表在计量经济学和统计学杂志上,如 *Journal of Business and Economic Statistics*、*Journal of Applied Econometrics*.“预测方法和技术的应用研究”课题组(中国管理科学研究院管理功效研究所)所做的“中国预测技术发展研究”<sup>[12]</sup>指出我国目前常用的预测方法如下:

- (1) 定性预测方法,包括 Delphi 法和目标分解预测法;
- (2) 时间序列模型,包括移动平均模型和指数平滑法;
- (3) 因果模型,包括分解预测法、博克斯-詹金斯模型 (ARIMA)、回归分析、数量经济模型以及投入产出模型;
- (4) 预测的综合方法.

Delphi 法是美国 RAND 公司于 20 世纪中期提出和使用的一种专家调查法.其特点是匿名性、轮间反馈性和预测结果的统计特性.它主要应用于下列情况的预测:预测对象缺少足够的有代表和可信的统计特征;那些容易受到新发明强烈影响的部门,其运行环境具有很大的不确定性;用于预测的时间不够或者预测问题处于极端情况. Delphi 法用于预测科技进步、社会和政治领域的主要发展方向是有成效的,但对于复杂系统的预测问题,还没有成功的先例.如对全球气候变化的趋势、贝加尔湖地区生态系统保护的预测,各方面的专家各执一词,众说纷纭,没有一致的看法.仅依靠直观的途径,靠人的定性推断与想像是无法解决复

杂经济系统预测问题的。

### 1.2.2 主要定量预测方法评述

常用的定量预测方法包括：传统的统计分析方法、模拟模型法和数据挖掘方法。

#### 1.2.2.1 统计分析方法

传统的统计分析方法是指利用数理统计手段对被预测对象的数量信息进行加工的方法，包括相关分析、方差分析、因素分析、主元分析、回归分析等。但是由于以下原因，传统的统计分析方法不能完全满足描述国民经济指标和因素之间随时间和空间改变的因果关系的需要，用于经济预测时往往误差很大。

(1) 正确地选择模型结构对于预测的成功至关重要。传统的统计模型方法主要用于估计未知参数，而无法确定模型的结构。建模者在确定指标之间的联系时，只考虑时滞和外部因素的作用而硬性规定模型结构，带有很大的主观性，可能将具有深刻本质和关键影响的因素（暂时还不能认识的）排除在模型结构之外，而且这些方法都没有明确的正式标准去检验模型结构的质量。

(2) 在参数估计过程中，传统的统计模型方法不能同时考虑对经济指标的各种约束和平衡关系，只能在建模完成后再考虑对这些约束和关系进行调整，这必然造成指标间的不平衡，出现圈套的误差。

(3) 传统的统计模型方法对过程的信息要求太多，这对于经济过程是很难满足的，容易造成经济预测模型的参数估计偏差太大，置信度不高。

(4) 传统的统计模型方法在估计参数时对数据的质量要求很高，例如，余项 $\eta$ 是随机变量且数学期望为零，独立变量和余项没有联系，变量间不存在多重共线性问题等。但是经济统计数据很难既满足这些近乎苛刻的要求，又要面临增大样本容量的困境。

#### 1.2.2.2 模拟模型法

模拟模型法是建立在演绎法的基础之上的，它需要被模拟对象高质量的、非经验的大量信息，数据干扰水平低，样本十分庞大，以达到对被建模对象的深刻认识。但在经济系统中却难以达到这些要求。

系统动力学(system dynamics, SD)方法<sup>[13]</sup>是研究复杂系统动态行为的一种计算机模拟方法。模型的支持系统不是数据，而是对系统元素间的相互关系构造，这“软化”了数据要求。它将现实的系统看作是多重反馈回路的组合，其模型的状态变量是时变的。只要系统的结构不发生变化，所构造的模型既可以回顾系统的历史行为，又可以预测系统的未来行为。系统动力学模型十分容易引入状

态变量的期望值和反馈控制的调节原理，因而许多不便于做破坏实验与风险实验的社会经济问题，可以通过系统动力学模型作政策性实验，所以常把系统动力学模型称作政策设计工具或政策实验室。

系统动力学是一种方法论。它对因果关系的确立是基于建模者对系统各元素之间关系的充分了解。也就是说，对系统建立 SD 模型时需要专家的指导。然而，对于复杂系统，如经济系统和社会生态系统等，寻找系统各元素之间的因果关系难度很大。如果在建模时忽略了某个关系(或变量)，而这种关系(或变量)恰恰是至关重要的，那么 SD 模型的结果很可能与现实不符，甚至毫无意义。这种情况对建模初学者不可避免，甚至像 SD 大师 Forrester 也不可避免。

复杂适应系统 (complex adaptive system, CAS) 理论由 J. Holland<sup>[14]</sup> 于 1994 年提出，是现代系统科学的一个新的研究方向。它突破了把系统元素看成“死”的、被动的对象的观念，引进具有适应能力的主体概念，从主体和环境的互动作用去认识和描述复杂系统行为，从而开辟了系统研究的新视野。适应性指主体能与环境及其他主体进行交流，在交流过程中“学习”或“积累经验”，并且根据学到的经验改变自身的结构和行为方式。其基本思想是适应产生复杂性：

- (1) 系统的个体称为主体(agent)；
- (2) 主体具有自身目的性、主动性与有活力(active)；
- (3) 主体与环境(包括其他主体)的相互影响、相互作用是系统演变和进化的主要动力；
- (4) 把宏观和微观有机地结合起来。

SFI 开发了 SWARM 模拟工具集来分析复杂适应系统，以免研究人员将过多的精力用在编程上。SWARM 系统是一个通用工具平台，提供了一套标准、灵活、可靠的软件工具。研究人员可以自由地定制 SWARM 中的多种对象，将精力集中在感兴趣的特定系统中，而不必被数据处理、用户界面及其他纯软件工程及编程等方面的问题所困惑。SWARM 是一个设备精良的软件实验室，帮助人们集中精力于研究工作而非制造工具，它支持的是一种“自下而上”或称“基于过程”的经济建模方法。SWARM 提供了真正理解发展和演化的可能性，改变了简单化的、只考虑量变的、线性外推的预测方法，引进了更加符合客观现实的、量变和质变相联系的思维方式。SWARM 基于 GNU/General Public License 的自由软件，也就是说它是开放源代码的，人人都可以免费使用，并且都可以参与到该软件的升级开发中。但 SWARM 不是一个现成的应用软件，它是一套开发工具，需要使用 Object C 或 Java 编程，其学习曲线是陡峭的，当然回报也是可观的。

由于 CAS 方法和计算机技术的先天的、内在的联系，它具有可操作性，从而为研究人员提供了越来越方便的研究工具和软件平台，使得人类对复杂系统规

律的研究和认识进入了一个新的阶段。

然而，基于 CAS 的经济系统建模必须描述主体决策的细节，如对信息的评价、期望、学习过程和决策等。建模者必须清楚了解系统主体的结构，明确主体之间及主体与环境之间的关系，即假定建模者对经济系统的要素及其结构关系已十分了解。然而，面对复杂的经济对象，这是一个难题。

### 1.2.2.3 人工神经网络建模

人工神经网络 (artificial neural network, ANN) 是一种重要的数据挖掘方法。它具有良好的非线性品质，灵活而有效的学习方式，完全分布式的存贮结构。神经网络模型是一种隐式模型，相当于一个黑箱，它将系统的结构隐含于网络的权值中。无论系统模型是何种类型，用于表达或描述这些模型的神经网络结构是不变的，正是由于这些特性，神经网络才拥有了强大的学习能力和适应能力。

Jim Huffman 说过：“我们正处在思维的新时代的初期。神经网络的重要性已超出我们是不是建造一个计算机的问题，更重要的吸引力是弄清大脑是如何工作的。” ANN 简便、学习能力强的特点在预测应用中尽情施展，因此受到不少预测人士的青睞，是预测研究领域的一支生力军。人工神经网络预测对于提高预测精度和适应能力有相当大的作用。然而，它用于复杂系统模拟预测仍存在以下问题<sup>[15]</sup>：

(1) ANN 难以解释结果的实际意义，特别是它无法回答“why”和“how”等问题。它从数据中提取出的信息仍隐含在网络结构中，难以结合经济理论来解释建模结果，并进行灵敏度分析和边际分析；

(2) ANN 没有一种系统的构造网络结构的方法，其结构由建模者通过试探性的反复实验来确定。由于没有一个标准的步骤来指导 ANN 模型的开发工作，与统计模型相比，它常需花更多的时间用于训练计算；

(3) ANN 模型常发生“过拟合”现象。ANN 的优点在于不断学习和训练，但却不能辨识异常值。当异常值影响强烈时，ANN 仍能给出拟合精度高的估计，即 ANN 模型常发生“过拟合”现象。此结果用于预测显然会产生较大的偏差。

### 1.2.3 预测研究新范式

纵观国内外几十年的预测实践，预测失败不乏先例。如 20 世纪 60 年代西方的石油危机、股票市场预测等。诸多事实表明，预测不是万能的，在许多情况下可以说预测是困难的，并且受多因素影响。

(1) 任何事件的发展都具有一定的偶然性，也正是偶然性才构成事物发展的

必然规律。必然规律是事物可预测的基础，而偶然因素又带来了预测的困难。

(2) 被预测事件自身发展规律发生了变化。过去与现有的预测均基于预测期间模型的结构与样本区间模型结构无大的差别，这在一定意义上是合理的。然而，由于社会经济系统本身是复杂的，且随着科学技术的发展及人们认识水平的提高，社会系统处于不断调整中，尤其是转轨期的中国经济系统的激烈变化，给预测带来了困难。

(3) 突发事件的影响。今日世界处于冲突多变的环境中，突发事件时有发生。面对如此多变的环境，如何进行有效预测是广大预测者关注的一项课题。

(4) 对预测对象机理的掌握程度。事件之所以是可预测的，其实质在于预测机理本身。如果刻意追求数学形式的完美和统计意义的合理而忽视对事件机理的研究，则将导致预测方法与模型的滥用与误用。

由于经济系统的复杂性，使预测者无法确定应该用什么理论或按照什么规则综合哪些理论建立预测模型。

**例** 一家企业希望建立产品需求的预测模型。模型的输出为产品需求  $y$ ，它包含产品销售数量  $y_1$ ，销售额  $y_2$  和销售利润  $y_3$ 。根据市场营销理论，可能影响产品需求的因素（输入）有广告费用  $x_1$ ，产品流通数量  $x_2$ ，存货数量  $x_3$ ，新顾客数  $x_4$ ，消费信心指数  $x_5$ ，消费者收入  $x_6$  和通货膨胀率  $x_7$ 。因此，这是一个 7 输入-3 输出模型：

$$y(y_1, y_2, y_3) = f(x_1, x_2, x_3, x_4, x_5, x_6, x_7).$$

预测人员要解决的首要问题是究竟应该用哪种理论或哪些理论的组合解决这一具体问题？如果要综合不同理论，那么综合的规则又是什么？进行预测时，还须考虑宏观、微观的政治、经济环境和社会消费心理等因素。预测人员不可能准确知道用什么理论解决这个问题，只能不断地假设、猜想、试探。那么，随之而来的问题是应该由谁根据什么来判断这些假设、猜想、试探的优劣。

刘洪在文献[16]中提出了转变研究范式的新观点。他指出，从科学发展史看，每一种科学理论都有自己的解释范围。迟早会遇到它的概念体系无法说明的现象。但是，在一种科学范式下被当作不可理解或判定的对象，当范式转换后往往就成为可以理解的了。当今，整个科学的发展正处在从传统范式向系统范式转换的时期，预测学作为自然科学与社会科学交叉的学科，其发展也需要从非线性系统理论等现代科学发展的成果中吸取新的“营养”，寻求新范式(paradigm)，并在新范式基础上发展预测的理论与方法。

非线性系统理论的发展为预测范式的转换提供了理论基础和契机，该理论对于预测学的影响从它们发展的一开始就得到了科学家们的重视。从一定程度上讲正是预测领域尚存在的问题促进了非线性系统理论的发展。普里高津(Ilya Prigogine)指出，我们正处在线性科学可能到达的顶端，但我们也正处在一个新科

学——非线性科学的始端，后者为我们展望世界的未来打开了一扇新的窗户。混沌理论家 Crutchfield 等人的研究<sup>[17]</sup>表明，混沌一方面指出了原本认为不可预测的复杂事物具有可预测性，另一方面也指出了对原本认为可预测的简单事物的预测具有局限性。预测理论家 Gordon 指出<sup>[18]</sup>，混沌理论开辟了预测研究的新领域，为原来被认为不可预测的复杂系统的预测提供了新的理论与方法途径。英国皇家科学院曾于 1994 年 3 月专门举行了“混沌与预测”研讨会并出版了会议论文集《混沌与预测》(Chaos and Forecasting)，来自不同领域的专家达成共识，混沌学在社会经济系统方面所取得的研究进展与在物理学等自然科学中的进展相比显得极不平衡，尤其在预测学方面的研究相当落后，但前景非常广阔。Albert 在《混沌与社会》(Chaos and Society)一书的前言中指出，复杂性科学不是时髦的辞藻，而是为我们理解所处的这个世界提供了方法论。Presti 则认为<sup>[19]</sup>，包括混沌理论在内的非线性系统理论对于预测学的影响将是革命性的，它们不是对现有理论与模型的简单修正，而是对其进行根本性的变革。

应用非线性系统理论研究经济技术预测范式的转换及预测的方法论属于自然科学与社会科学交叉的学科问题，也是管理科学的前沿性研究课题。刘洪在文献[16]中认为，目前虽然国内外有相关的研究，但研究成果零星分散，且大多局限于个别问题的定性分析、理论阐述或方法本身的技巧处理，就事论事。从预测理论与方法的基本假定、观念、思维框架等范式角度和建模原理角度进行系统研究的并不多，关于预测理论与方法的评价和创新也还没有摆脱传统范式的束缚，理论上的深度、体系上的广度均有待拓展。

国内外相关研究所取得的主要结论可归结如下：

(1) 如果系统是稳定的，那么趋势外推是有效的；如果系统是不稳定的，或者稳定的系统演化到了临近分叉结点的附近，那么无论预测模型对历史数据拟合得如何好，外推总是不可靠的。

(2) 复杂系统能够产生“伪信息”，有些似乎强相关的因素之间并不存在任何直接的联系。“蝴蝶效应”则表明，小的不起眼的原因也会造成惊人的结果。这些都要求重新考虑对预测模型中解释变量的选取，不能仅仅依赖于相关系数的分析，还需要寻求理论上的依据。

(3) 系统的随机性态既可以来自系统外部的扰动，也可以由系统内在决定。因此，关于误差的解释需要首先从系统内部找原因，然后再从系统外部找原因。

(4) 复杂系统长期精确预测是不可能的，因而对于经济、技术、社会这样的复杂系统不必投入太多的成本来建立复杂的预测模型。另一方面，由于短期预测是可行的，那么，对于初始条件和系统结构较为准确的刻画则可以改善短期预测的准确性。长期的定性预测是可行的，因此，在还没有认清复杂系统规律的情况下，依靠专家的经验、学识和智慧，通过类比、分析和综合的直观悟觉的预测方



法, 要比定量预测方法更为实用。

(5) 混沌理论的有关混沌吸引子概念、分形概念、相空间重构方法、数字生态模拟法(data ecology)等都可以用来指导预测的建模, 从而开发出混沌动力学预测法、混沌情景预测法、混沌唯象预测法和分形预测法等新的预测方法论。

### 参 考 文 献

- 1 Arthur W B, David A L, Steven N D. The Economy as an Evolving Complex System II (Santa Fe Institute Studies in the Sciences of Complexity S). Perseus Books Group, 1997
- 2 Lissack M R, Gunz H P. Managing Complexity in Organizations: A View in Many Directions. Greenwood Publishing Group, 1999
- 3 Casti J. Complexion: Explaining a Paradoxical World through the Science of Surprise. Harper-Collins, 1994
- 4 苗东升. 系统科学精要. 北京: 中国人民大学出版社, 1998
- 5 成思危. 复杂性科学探索. 北京: 民主与建设出版社, 1998
- 6 Arthur W B. Complexity in Economic and Financial Markets. Complexity, 1995, 1 (1): 20~38
- 7 Gallagher R P, Appenzeller T. Beyond Reductionism. Science, 1999, 284 (5411): 79
- 8 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. 自然杂志, 1990, 13 (1): 3~10
- 9 Anderson P, Arrow K J, Pines D (eds.). The Economy as an Evolving Complex System. Addison-Wesley, 1989
- 10 Arthur W B. Increasing Returns and Path Dependence in the Economy. Ann Arbor: University of Michigan Press, 1994
- 11 Paula R D. 困难的经济预测艺术. 预测, 1997, 16 (4): 69~72
- 12 “预测方法和技术的应用研究”课题组. 中国预测技术发展研究. 预测, 1991, 10 (5): 10~17
- 13 王其藩. 系统动力学 (第二版). 北京: 清华大学出版社, 1994
- 14 Holland J H. Hidden order: how adaptation builds complexity. Reading, Mass., Addison-Wesley, 1995
- 15 司昕. 预测中的神经网络模型. 预测, 1998, 17 (2): 32~35
- 16 刘洪. 非线性系统理论与预测研究新范式. 预测, 1999, 18 (2): 1~6
- 17 Crutchfield J P et al. Chaos. Scientific American, 1986, 254: 46~57
- 18 Gordon T J. Integration of Forecasting Methods and the Frontiers of Futures Research. United Nations University, 1994
- 19 Lo Presti A. Futures Research and Complexity: a critical analysis from the perspective of social science. Future, 1996, 28 (10): 891~902

## 第 2 章 自组织数据挖掘基础

建模方法分为基于理论的方法和基于数据的方法两类，它们各有特色、互为补充，根据实际建模问题而侧重运用。自组织数据挖掘是一种数据挖掘方法，它是在启发式自组织方法基础上发展起来的，详见 2.2 节。自组织数据挖掘的基本思想、基本原理以及关于最终模型——最优复杂度模型的分析见 2.3 节。

### 2.1 数据挖掘概述

#### 2.1.1 建模方法：基于理论的方法和基于数据的方法

模型是指为了某个特定目的将现实系统的某一部分信息缩减、提炼而构造的真实系统替代物(substitute)。经济预测不能没有定量分析，而模型又是经济预测定量分析的主要手段。模型主要用于：

- (1) 系统辨识：识别系统的功能和结构。模型比原对象更易于分析，因而更能深入认识原对象；
- (2) 系统控制：寻求合适的方式对被研究对象施以积极的影响；
- (3) 系统预测：预测被研究对象未来的发展变化规律。运用模型做实验可以回答“如果—那么”的问题。

已有的建模方法很多，可以对建模方法从不同角度分类。J. A. Mueller 将建模方法分为基于理论的方法和基于数据的方法两类<sup>[1]</sup>。

基于理论的方法或理论系统分析方法以建模对象的已知理论为基础，其模型是按物理原理得来的。运用该方法时，建模对象的因果关系需要由一些式子来表达，例如，用数学方程式、代数式、基于计算技术或类似方法得出的式子等。虽然模型给出了真实系统实际过程行为的一个描述，但是在基于理论的建模中，建模者必须知道关于系统本身更多的先验信息：他/她必须能描述被研究对象的基本结构，至少对于建模目的所必需的部分是如此的。另外必须知道原系统的内部影响因素的结构，同时也要明白环境的影响因素。

**例**(参看 1.2.3 节) 一个公司想建立产品需求的预测模型。已知对产品需求有影响的变量可能是：广告费用、流通产品数、库存的产品数、新的消费者人数、消费者信心指数、消费者收入和通货膨胀率。作为系统的输出变量包含产品销售量、利润率和总利润。按照基于理论的建模方法，公司的研究机构必须用定

性方式表达：

- (1) 所有 3 个输出和 7 个输入变量之间的互相依赖结构关系；
- (2) 为每个变量确定关联动态性(时间延迟)。

这意味着公司必须决定产品销售量是否受广告费用或流通产品数的影响，或者同时受它们的影响，或者受其他可能因素的影响。同时公司的研究机构也必须要知道 1 个月、2 个月或 6 个月以前消费者的消费信心指数是否对下个月的产品销售有影响。然而，最重要的在于问题是由如下哪几个方面组成的：全球的，宏观和微观经济的，心理的和政治的等。因此，最终的问题是：在给定的阶段应关注什么理论，并怎样将这些理论组合起来？连接不同理论的规则是什么？这里，市场理论可能是比较有用的。然而，通常这些规则不是完全知道的。但是基于理论的建模方法依赖于此，研究机构人员只能不断地假设、猜想、试探。每次他们改变假设，就可能得到完全不同的结果，这当然是合理的。但是，由什么或由谁来决定什么是正确的假设呢？这是一个问题。

对于先验知识丰富的情况，基于理论的方法就足够了。但这种方法却不能解决具有下列特点的系统，如经济系统的建模问题：

- (1) 关于系统的先验信息知之甚少；
- (2) 大量不可测的变量；
- (3) 数据样本少且含有噪声；
- (4) 具有模糊特征的不确定系统。

基于数据的方法则正好相反，它不需要有足够的先验信息和理论，而着重于从数据或实验中得到分析结果。基于数据的方法从实际系统的观察数据样本中得到一个关于系统行为的描述，它指出在不同的条件(输入)下系统(输出)怎样运作。当被研究对象信息不完全时，基于理论的方法在分析复杂问题时不能很好地确定研究对象(what variables)、研究方法(what methods)和选出恰当的模型(what models)，而基于数据的方法更适合解决这些问题。数据挖掘是典型的数据驱动方法。

## 2.1.2 数据挖掘

1987 年，密歇根大学(University of Michigan)的研究生 Fayyad 在通用汽车公司作暑期工，处理小车维修大型数据库，萌生了一个念头：只询问车子的型号、发动机的容量等，通用汽车维修部的人员就能很快地从数据库中知道需维修的部位所在。听起来很直观，但 Fayyad 回忆道：“有数百万条数据记录，没有人能玩转它。”由 Fayyad 设计的解决该问题的特征识别算法成为他本人 1991 年的博士学位论文，这就是数据挖掘的起源。

随着信息技术的飞速发展，如今几乎人类活动中每一个领域都充斥着大量

的统计观测数据。信息的增长呈现超指数上升，据估计，全世界的信息量不到每 20 个月就增加一倍。通常，在各种企业、商业领域的交易记录与财务报表、科学研究领域收集的数据(例如，气象卫星传回的气象图像)中，其数据量经常为数十兆字节，甚至成百上千兆字节。虽然现代计算机技术与数据库技术已可以支持存储并快速检索如此规模的数据，但是无论在时间意义上还是在空间意义上，传统的数据分析手段还是难以应付，人们无法理解并有效地使用这些数据。如曾经广受赞誉的管理信息系统 MIS 和决策支持系统 DSS 在如山的数据面前也日益显露其局限性，因为它们不能从数据中提炼出足够的直接可用的知识信息。由此导致越来越严重的“数据灾难”，数据分析与数据产生之间出现了越来越大的距离。另一方面，在许多领域，如经济、政治与军事领域等，往往需要分析大量数据之后才能作出正确决策。例如，Wal-Mart 公司每天要处理 2000 万个事务；美国航天局 1999 年发射的地球观测系统每小时要产生 50Gb 的图像数据等。毫无疑问，这些庞大的数据库及其中的海量数据是极其丰富的信息源。但是仅仅依靠传统的数据检索机制和统计分析方法已经远远不能满足需要了。因此，“We are drowning in information, but starving for knowledge (我们淹没在信息的海洋中，却渴望知识)”的现象普遍存在。传统方法(大多为基于理论的建模方法)失效和“数据灾难”是数据挖掘产生的原因。

数据挖掘(data mining)从庞大的数据库中提取模式，如趋势、特征及相关性。2001 年，数据挖掘在《技术评论》(The Technology Review)杂志预测未来技术的十大方向中居第三位。数据挖掘相关技术有<sup>[2]</sup>：市场购物篮分析、决策树、遗传算法、聚类、连结分析、在线分析处理、类神经网络、判别分析、Logistic 回归分析等。其功能大致有：分类、聚类、估计与预测、关联和序列发现、描述等，这些功能大都可以用成熟的计量及统计分析方法来实现。

尽管数据挖掘方法众多，但绝大多数由以下三要素构成：

- (1) 模型表示，即在怎样的前提假设下用什么描述方式；
- (2) 模型评价准则，即对模型及其参数满足目标函数程度的定量描述；
- (3) 搜索方法，包括模型搜索和参数搜索。确定模型表示方法和模型评价准则后，数据挖掘就成为优化问题：从模型集中选择能使模型评价准则最优化的模型和参数。

数据挖掘是一个相对新的领域，目前已有许多不同的数据挖掘方法，而一种好的数据挖掘方法首先应该自动完成数据挖掘过程，以排除人们主观认识对建模结果的影响。另一方面人们对数据挖掘的兴趣通常只在于最后的结果，因此一种好的数据挖掘方法不应该要求用户精通数学、控制论和统计学等理论，也不应要求用户花大量时间去处理一个对话框接一个对话框的复杂数学建模软件。

### 2.1.3 数据挖掘与知识提取

与数据挖掘密切联系的另一个概念是知识提取. 知识提取(knowledge discovery in database, KDD)是从数据中寻找有用知识的全过程, 数据挖掘是实现KDD的其中一个步骤.

KDD一词是1989年在美国底特律市召开的KDD专题讨论会上正式提出的, 含义是基于数据库的知识发现, 指的是从大型数据库或数据仓库中提取人们感兴趣的、事先未知的、潜在有用的、易被理解的模式的非平凡过程. KDD的这一定义包含以下信息:

- (1) 数据是一组事实的集合;
- (2) KDD是一个包含多个步骤的过程: 一方面包括数据收集、模式识别(数据挖掘)、模型评价及修正等不同步骤, 另一方面这些步骤可能要重复进行;
- (3) 非平凡(nontrivial)表明KDD不是像计算数据的平均值一样的简单计算, 而是包含了研究与推断的过程.

不同领域的建模者可以从不同角度定义知识, 同一领域的不同建模者对知识也会有不同理解. KDD中知识的含义是只要提取的模型、模式或其他任何分析结构在一定程度上满足了研究者的兴趣, 那么这种分析结构就可以成为知识. 这种知识一定要是新的、可以理解的. KDD的一般步骤<sup>[3]</sup>包括(图2.1):

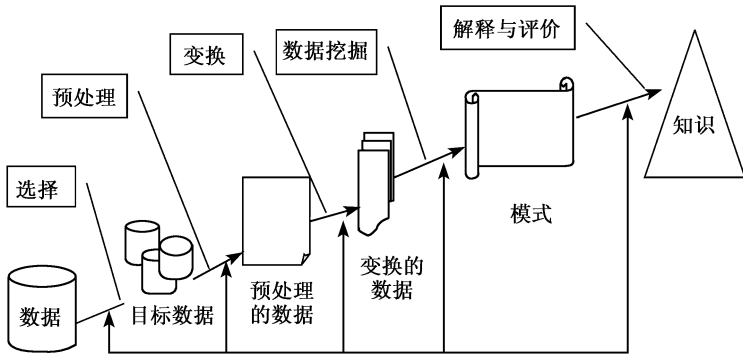


图 2.1 数据的知识提取步骤

- (1) 加深对所涉及的领域及其相关理论知识的理解, 弄清终端用户的目的;
- (2) 创建一个目标数据集——数据选择;
- (3) 数据整理和预处理. 包括消除“噪声”或去掉无用的数据, 决定用于弥补遗漏数据的策略, 说明时间序列信息和已知的变化等;
- (4) 数据降维和转换;
- (5) 数据挖掘(DM): 包括数据挖掘任务的选择, 数据挖掘算法的选择, 数

据挖掘的实施：

① 解释挖掘出的模式；

② 评价 KDD 的结果：将发现的知识以用户能了解的方式呈现给用户。

对 KDD 结果的评价可能导致前面任意一步的改变，甚至导致整个 KDD 的重新开始。所以 KDD 是一个需要不断重复，以最终获得有价值的知识的过程。

## 2.2 启发式自组织方法

A. G. Ivakhnenko 在 1967 年提出了数据分组处理算法 (group method of data handling, GMDH)，并在此基础上逐步形成了自组织数据挖掘的理论与方法，而 GMDH 的基本思想来自启发式自组织方法。

### 2.2.1 基本原则

自组织是一种过程，复杂动态系统的组织在这个过程中产生、繁殖和完善。自组织过程仅出现在复杂程度很高的系统中，这种系统包含大量元素，且元素之间具有非刚性的随机特性关系。自组织是靠变革系统元素间旧的联系，建立系统元素间的新联系来实现的。

自组织有下面的类型：

(1) 组织自生型。由对象集合产生新的具有特殊规律的系统时出现组织自生型，这种类型表现出从量变到质变的辩证规律。

(2) 组织水平保持型。在这种类型的过程中，系统同环境相互作用而保持一定的组织水平，即在运行的内外条件改变时保持系统的完整性。按负反馈原则作用的稳态调节系统通常归于这类系统。

(3) 自学习系统型。这类系统能积累和利用过去的经验以保证其运行的可靠性，并具有不断完善和发展的特点。

对于经济系统，其特点是组织的产生(经济对象的形成、设计、重组及重构等)、动态平衡的自动调节(为达到预定的目的而控制经济过程)、完善(控制、组织等)，所以属于自组织系统。因此，基于经济数学模型的自组织原则来研究经济系统符合经济现象自身的性质。

适应性是系统、特别是自组织系统的重要性质，它表现复杂环境中有所目的的适应行为：系统的主体能与环境及其他主体进行交流，在交流过程中“学习”或“积累经验”，并且根据学到的经验改变自身的结构和行为方式。系统的个体称为主体(agent)，主体具有自身目的性，主动性并有活力(active)，主体与环境(包括其他主体)的相互影响、相互作用是系统演变和进化的主要动力。系统的适应性程度是由适应新条件过程中系统结构变化的水平确定的，在研究自组织系统建