

预测论基础

翁文波 著

石油工业出版社

51.73

预测论基础

翁 文 波 著

石 油 工 业 出 版 社

内 容 简 介

本书试图把自然科学和社会科学中某些预测问题在原则上统一起来,从而扩张了预测的领域,例如从微观粒子的性质到某些全球性工业的盛衰。本书在新的原则上提出超远程预测某些重要现象的可能性,例如预测一个城市在哪星期内下大暴雨或一个区域在哪月前后发生强烈地震的可能性。

预 测 论 基 础

翁 文 波 著

*

石油工业出版社出版

(北京安定门外外馆东后街甲36号)

北京顺义燕华营印刷厂排版

曙光印刷厂印刷

新华书店北京发行所发行

*

850×1168毫米 32开本 45/8印张 79千字 印1,501—5,180

1984年5月北京第1版 1984年8月北京第2次印刷

书号: 15037·2502 精装定价: 1.30元

科技新书目: 77—133

FUNDAMENTALS OF FORECASTING THEORY

Weng Wen-Bo

*(Scientific Research Institute for Petroleum
Exploration and Development, Beijing, China)*

Abstract

This thesis unifies the fundamentals of some forecasting problems in natural and social science, so that the domain of forecasting is extended from the physical property of microscopic particle to the life cycle of some world industry. The new principles point out the possibility of ultra longrange forecasting of some important phenomenon, such as the possibility of heavy downpour in a city within a certain week, or the possibility of some strong earthquake of a region around certain month.

目 录

绪论	1
第一章 预测过程	4
§1 信息	4
§2 信息交流	6
§3 几乎和可能	6
§4 体系和模型	8
§5 映照	9
§6 预测过程	11
§7 反馈	13
第二章 体系的属性分类	15
§1 稳定体系	15
§2 计量体系	16
§3 复合体系	17
§4 突变体系	20
§5 动态体系	23
§6 互逆体系	24
§7 模糊体系	27
§8 不定体系	28
第三章 对称和守恒	30
§1 对称	31

§2 对称多项式	32
§3 投入产出守恒	33
§4 标度模拟	37
§5 类比	39
第四章 整数	41
§1 自然数	42
§2 整数	43
§3 差分	45
§4 可公度信息系	47
§5 有限整数	51
第五章 预知信号	54
§1 乘法和乘法表	54
§2 周期函数多项式	60
§3 互乘表	62
§4 广义定标预测	65
第六章 拟合信号	67
§1 拟合	68
§2 生命旋回	70
§3 正态旋回	71
§4 泊松旋回	79
§5 逻辑斯蒂旋回	86
第七章 回归	90
§1 自回归	90
§2 线性回归	92
§3 同态线性回归	99

§4 多元回归	104
第八章 随机体系	105
§1 分布	106
§2 分布的数字特征	107
§3 正态分布	108
§4 平均法	113
§5 移动平均法	114
§6 指数平滑法	115
§7 高阶移动平均	117
§8 混合移动平均	118
§9 直观随机误差	119
§10 事件预测的置信水平	119
第九章 随机性的否定	123
§1 均匀分布	123
§2 简单随机游动	127
§3 等概率简单随机游动	128
§4 信息的综合	136
附录	142

绪 论

“凡事豫则立，不豫则废”和“人无远虑，必有近忧”等成语都说明预测的重要意义。现代技术预测的范围很广，包括社会预测（如人口等问题）、科技预测（如发展方向）、经济预测（如趋向、市场、管理等）和军事预测（如战略）等方面。

我国有些大专院校也开设技术预测（TF）课，但一般均隶属于社会科学中的技术经济范畴。本文希望在比较广泛的基础上，将技术经济领域内通用的预测技术和自然科学中的外推和预测方法尽可能有系统地汇合起来，组成“预测论”的一个基础概要。只是讨论内容一般是从静止、孤立、片面、单纯和机械的体系出发，还只用到数代数系的模型，所以适用范围还会有很大的局限性。

预测可以区分为：（1）以统计学（STATISTICS）为基础、统计量（如平均值、方差等）为对象的统计预测；（2）以信息学（INFORMATICS）为基础、信号为对象的信息预测。本文主要讨论信息预测。

预测是带有主观性的。有人认为预测中的模式识别

是科学加艺术，所以，预测并不存在唯一的方法。

根据不同问题的性质，预测模型可以区分为确定模型和随机模型。本文主要讨论确定模型，最后二章也涉及随机性和有关问题。

第一章讨论信息的定义、信息过程和它的结果。第二章讨论预测体系。由于预测体系是多种多样的。如：平稳的、不平稳的、线性的、非线性的、趋向的、周期性的、交变性的、均匀的、多重性的、突变或灾患性的、模糊的……等。本文未能作系统的分类和叙述，只是提出几种典型，希望能从典型扩张为类别。

从大量客观事物的观察中，可以归纳出许多规律，它们是对有关体系进行预测的基础。作为典型，本文讨论了几种有相当普遍规律的体系，它们是：对称体系，可数（量子化）体系，周而复始（周期性）体系等。其它信息体系，我们只讨论了两个极端情况，一个极端是事先就预定什么是（所要的）信息组成的体系。另一个极端是事先完全不规定什么是信息，而用相对简单的模型去拟合客观实际体系。在这两个极端之间，当然还有无数个中间类型。实际情况说明：上述典型信息，即：对称性、量子性和周期性等信息，常出现在各种信息体系之中。

回归可以看作是拟合模型中一种比较重要的特款，因为回归分析可以研究多体系的相关，所以可用于所谓

“因果预测”。在技术上,线性回归就是线性方程的最小二乘法拟合。当然,也可以用统计学的原则来定义,如一元正态线性回归,并用统计学的假设检查来判别。所以回归是从确定模型到随机模型的一个接界,也是信息预测和统计预测的一个汇合点。本文第七章有较详细的讨论。

最后,以随机体系的讨论作为一个转折,从统计预测的平均法(平滑法)发展到以否定随机性为原则的信息预测。

本文原来计划在一本有关“初级数据”一文出版以后再写的。但是由于预测技术的迅速发展,只好先写这本小册子。为了避免阅读上过分耗费时间,文字上力求简短,也不作详细的推导和讨论。而且由于预测论是一门新兴的学科,笔者也只能就几个简单的问题作一些定性的叙述,所引用的实例是一些示意性的单项预测,希望能引起读者的兴趣。笔者相信,读者如果有普通高等数学的一般知识和必要的耐心,一定能够理解笔者提出来的预测方法和方向,并且为开拓新的领域和途径作好准备,使将来的预测工作更有成效。

文中引用的实例,虽然原始数据是公开而真实的,但处理方法力求显明、简单,以便达到示意的目的。所以大部分结论带有习题性质,并非实际预测。

本文编写过程中,钱绍新同志提了有益的意见,吕牛顿同志参加了整理定稿工作,特此致谢。

第一章 预测过程

统计学和信息学是两个不同的学科,在某些问题上,它们互相接界、互相汇合。统计预测和信息预测分别是以这两门学科为基础的,所以,根据习惯上的理解,它们各有独特的性质,但也并不排除它们之间的边缘接界或互相汇合。一种学科分类把统计学归属到信息学中,把统计量也作为信息的一类特款,这样,统计预测也就包含在信息预测之中了,那是另外一个问题,不在本文要讨论的范围之内。现在暂把统计预测和信息预测区分开来,并以信息预测作为主要研究对象。为此,首先谈什么是信息。

§1 信 息

什么是信息?信息可以认为是信息体系中的元素、元素集或子体系。那么,什么是信息体系?本文认为信息体系是受人们主观定义约束的秩序类。主观定义的约束可以是:某种理解、信念、设想、定理、法则、规律、法律、契约、编码等。

有许多关于信息的定义。如:“使消息中所描述事件

出现的不定性减少”、“消息中所含的意义，它不随载荷它的物理设备形式的改变而改变”^①；又如：“信息这个名称的内容是我们对外界进行调节并使我们的调节为外界所了解时而与外界交换来的东西”^②（维纳，N. Wiener, 1894~1964），都和上述定义不相抵触。

自从申农（C. Shannon, 1948）提出信息的概念并用概率来定义信息量、信息量的单位（BIT）及演算关系（加法）以来，信息学有了很大的发展。但是，现在不依据概率的信息定义受到非常广泛的注意^③。信息的主观性是被维纳（《控制论》，1948）明确说明的。他把人、动物和机器的控制与通讯过程统一起来。本文以人们的主观定义为依据，动物和机器的控制，只有为人们所认识的条件下，才作为信息来处理。这和维纳的概念有原则上但无实际上的差异。不过本文用的“人们”一词和维纳的“人”有所不同。本文认为：单个人的某种信息思维，只有在传播到其它人并为其他人所共同理解的条件下才成为信息。“人们”之间的传播过程起到扩张、限制和中继等作用。这些作用又使信息超越了时间、空

①《英汉计算技术辞典》，人民邮电出版社，209~210页，1978。

②王鼎昌。“信息论的发展和意义”，自然辩证法通讯（2）1981，35页。

③H. W. Gottinger: “Concepts and measures of information”, CISM, Courses and lectures No. 29, Wien, New York, 1975.

间、形式等的限制。在信息过程中，本文提出“介体”一词，介体包括主观的观点和客观的工具，其中也包括维纳的机器。

§ 2 信息交流

上文提到：人们之间的传播是信息形成的一项必要条件。在传播中，信息受到传播介体的扩张、限制和中继作用。开调查座谈会或专家小组会都是交流和形成信息的形式。在交流中，信息的扩张、限制和中继作用以不同方式表现出来。如有的人“最能说会道”、“最有威望”、“最有说服力”。结果，重要的信息可能在集体互相妥协中失真。为了信息保真，可以采用许多方法。例如“德尔菲 (Delphi) 技术”：由调查人员组织（可能匿名）专家组，用书面形式分别质询各专家的预测意见，再由调查人员综合、分析这些预测和理由，并向专家们提出一系列有关综合意见的问题。这样反复进行多次，可以得到最佳的或满意的结果。如果调查对象并非数目不大的专家而是广大群众，那末，随机抽查、随机蹬点也可能取得相对无偏的信息。

§ 3 几乎和可能

从客观体系中，以主观定义的信息来建立模型，存

在着主客观的矛盾。例如：从模型 X 中得出一个数值 x 和实际体系 Z 中观察到的对应数值 $\hat{x}$ ，一般并不完全相同。

在确定模型中， x 是一个确定值。如果认为差值 $(\hat{x} - x)$ 是观察误差，而且大到不能令人满意，可能重新观察，即舍弃了 $\hat{x}$ ；如果认为差值 $(\hat{x} - x)$ 是模型 X 对体系 Z 的离差程度，而且大到不能令人满意，可能重建模型，即舍弃了 x 。在选择舍取观察位 $\hat{x}$ 或模型值 x 前，有主观决定的临界标准，这个标准在线性规划中称为约束条件，在最优化过程中称为可行域边界。本文用符号 ϵ 表示这个可行临界值。 $|\hat{x} - x| \leq \epsilon$ 称为可行变程条件，它说明模型 X 对于体系 Z 的近似性。

在随机模型中，对应于一个 x ，模型给出一个随机变量的分布。 x 常常是这个分布的数学期望。从随机观点看问题，模型 X 是随机体系 Z 的形象。所以对应于一个 x 值， $\hat{x}$ 应该也是另一个随机变量的分布的数学期望。以上两种情况的随机变量的分布未必相同。对模型 X 或观察值 $\hat{x}$ 的取舍也可以用与某一事先主观决定的置信水平相对应的置信区间作临界标准。如果沿用 $|\hat{x} - x| \leq \epsilon$ 的符号来表示置信区间条件，那末，相应的置信水平 $(1-\alpha)$ 或 $(1-\beta)$ 等可以说明模型 X 形象体系 Z 的可能性（概率）。

在某些事件预测中，如果实际的 $\hat{x}$ 值和确定模型 x 值的近似性合乎主观要求，预测结论可以是：“几乎如此”。如果实际的 $\hat{x}$ 值和随机模型 x 值的可能性合乎主观要

求, 预测结论可以是: “可能如此”。

§4 体系和模型

一个体系可以看作是客观世界中被主观选取的一个局部, 这个体系用符号 Z 来表示。为使主观选取的体系为一群人所共同认识, 就得建立一个大家都能理解的模型, 这个模型用符号 X 来表示。模型的建立是通过介体 Y 所产生的作用 O 而完成的。这个过程可用下式表示:

$$ZOY = X \quad (1)$$

在不同性质的介体 Y 的作用下, 产生两类预测模型。一类是以体系 Z 中各元素的共性为基础的统计模型, 因为体系中各元素的共性包涵在体系之中, 在统计量定义以后, 不再需要其它主观选择。从统计模型所产生的预测称为统计预测。另一类是以体系 Z 中各元素的特性为基础的信息模型。这种特性就是信息的定义。它也有二种基本类别, 对应于两类信息模型。一类是以概率为基础的随机信息模型, 申农提出的信息论就是用概率来定义信息的^①。另一类是确定信息模型, 信息的定义不涉及概率, 而是由主观决定的。如“诧异价值”^②的概念

①Shannon, C E., Weaver, W., "The mathematical theory of Communication" Illinois Univ. Press, 1949.

②Longo, G., "Information theory new trends and open problems", Springer-Verlag, Wien-New York, 12, 1975.

把表示特性的信息定义为体系 Z 中各元素的差别。柯尔莫哥洛夫^①等人提出的信息概念属于后一类。实际预测的信息模型可以同时包括这两类信息。从信息模型所产生的预测称为信息预测。

本文主要讨论信息预测的基础典型。

§ 5 映 照

信息模型中,体系 Z 、介体 Y 和模型 X 的关系可参考映照关系作图解,如图1—1所示。图中直线 Z 代表体系;直线 X 代表模型;点 Y 代表介体;射线 YO 代表 Z 在 X 上的映照。在实际体系中, YO 一般不是单值映照, Z 中的元素和 X 中的元素可以但未必都是一一对应关系。实际上,一一对应关系是存在的,例如:一条消息符号链中的每一符号,和对应的正确代码是一一对应的。射线 YO 就表示代码转换。

显然,完全对应的映照只有在概念中存在,通常认为,在主观确定的可行临界值范围内偏离 YO 是可以被接受的。这一范围在确定模型中称为可行变程;在随机模型中称为置信区间,它是由置信水平决定的。

上文中的映照 YO 已超越了代数中函数的含义。例如

①Kolmogorov A.N, "Logical basis for information theory and probability theory" IEEE Trans. Information Theory, IT-14, 662, 1967.

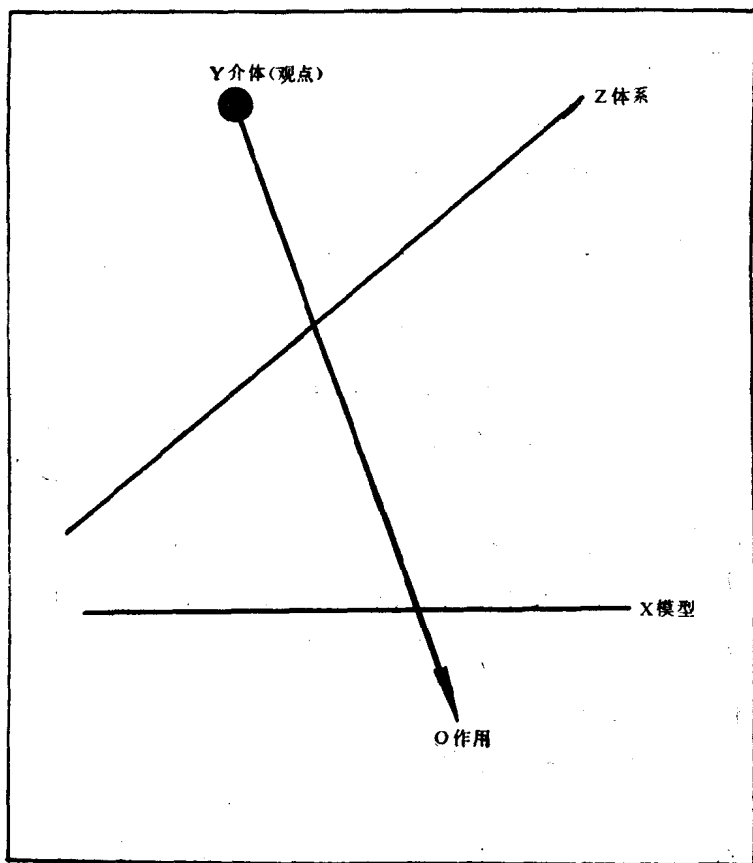


图1-1 信息模型

以地下地质构造为体系 Z, 通过地震勘探工程 Y 的实施 O, 得到磁带上数字序列集的模式 X。从信息过程看, X 中的数字序列集就是地下地质构造 Z 的信息映照。