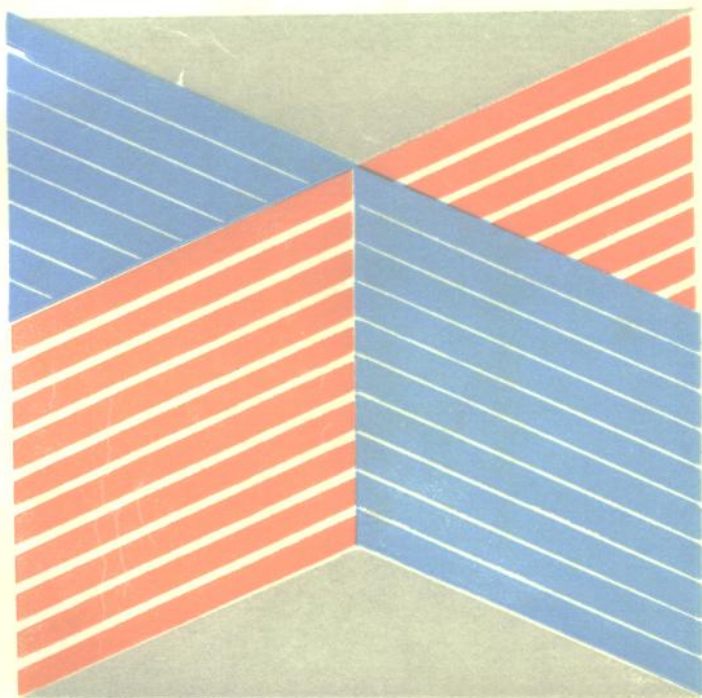


广义信息论

鲁晨光 著



中国科学技术大学出版社

0236

374054

L52

广义信息论

鲁晨光 著



中国科学技术大学出版社

1993·合肥

DY68/27

内 容 提 要

本书回顾了信息和熵理论的历史,介绍了 Shannon 信息论及其局限性;提出了广义通信模型和可度量语义信息、感觉信息及测量信号信息的广义信息测度;讨论了预测和检测的信息准则和优化理论;提出了限误差信息率函数和保质信息率函数——经典信息率失真函数的改进形式,及相应的通信数据压缩理论;介绍了新的信息测度在气象预报、图象通信等领域的应用;分析了信息熵和统计物理熵之间的关系;把新的通信模型和信息测度推广到控制领域,使得信息测度可以用来评价控制效果,而且通信优化方法可用于控制优化;还讨论了有关的经济学、生物学、美学及哲学问题。新理论贯彻和深化了 K. R. Popper 的科学进化论及马克思主义的实践检验真理思想。

本书可供通信、预测、检测、模式识别和人工智能、自然辩证法、哲学等领域的研究人员及大专学生阅读,亦可供语言学、控制、经济学、统计物理等方面的学者参考。

(皖)新登字 08 号

广 义 信 息 论

鲁晨光 著

*

中国科学技术大学出版社出版

(安徽省合肥市金寨路 96 号, 邮编: 230026)

中国科学技术大学印刷厂印刷

安徽省新华书店发行

*

开本: 850×1168/32 印张: 6.5 字数: 168 千

1983 年 10 月第一版 1983 年 10 月第一次印刷

印数: 1—3000 册

ISBN 7-312-00501-2/TN·20 定价: 5.50 元

(凡购买中国科大版图书, 如有白页、缺页、倒页者, 由本社发行部负责调换)

序

1985年,本书作者带了个色觉机制物理模型来到贵阳第二届全国模糊数学会议上,象是带了个稀有动物,引起不少人的兴趣。五年后,他作为汪培庄教授和罗承忠教授的访问学者来到北师大,又带来了他的广义信息理论,大有催枯拉朽重起炉灶之势。很少有象他这样的杂家:当过插队知青、钢铁工人,文凭来自航空,职业是教计算机;他感兴趣的领域是——我们还是看看发表过他文章的杂志吧:《潜科学》,《自然杂志》,《光学学报》,《通信学报》,《模糊系统和数学》,《机器人》,《心理学动态》,《哲学动态》,《现代哲学》,《科学技术与辩证法》,《预测》,……如果他的另外一些文章幸运的话,还会有逻辑学、气象学、美学等杂志加入这一行列。一个人在十年之内涉猎了这么多领域,而且搞出了那么多名堂,令人惊异。

“你干嘛打一枪换个地方?你可以攻其一点,不尽其余,而后再多多益善;比如你可以好好地整理和推销你的色觉理论,也好得个科学进步奖。”我说。

“人生短暂,有那么多的事要做,我没有时间。得奖者成千上万,我决不希罕成为他们其中的一个!”

好一个舍我其谁!

“那你写书干什么,还不是为了得到大家的承认?”

“我看到有人推销的洋货不怎么样,不服气。”

“你了解外国人这方面的所有研究吗?”

“我查过英国前不久出的系统论控制论百科全书,如果还有什么更好的信息理论我不知道的话,这套书就有问题了。”

于是我仔细看了他的国货。

这位自产自销者思维少有的清晰，远不象他初看那样异想天开。这是一本科技书，但是它更有哲学味道。除了通信，它还涉及控制论、物理学、经济学、生物学、逻辑学、美学、行为科学、认识论以及本体论。其中内容的新颖性和思想的深刻性远非一般。

关于广义信息度量，作者得到了或通过数学公式得到了许多简单而又符合常识的结论。比如：语义信息量等于命题逻辑概率和谓词逻辑概率之比的对数；越是把偶然预测为必然并且预测正确，信息量越大；谎言提供的信息是负的；听信算命先生信口开河会减少我们已有的信息；语义清晰是冒险的，语义模糊是保守的，往往太清晰不好太模糊也不好；对于感觉信息，在信号中有噪声存在时，主观分辨率也是如此，太高了不好太低了也不好…最令人兴奋的是，简单的通信模型和数学公式——与仙农理论中的公式形式上差不了多少——还蕴含了马克思主义的实践检验真理思想和波普尔的科学进化思想。作者太幸运了，为什么我就没有想到？

作者提出：要用广义信息量代替通信和控制系统中常用的人为定义的失真量作为通信或系统状态估计的质量指标；用作者的话来说就是，在通信和控制领域也要打破“无过便是德的用人标准”，要给提供更多主观信息的信号以更高的估价，而不是仅仅看信号失真与否。联系天气预报，这种想法的合理性不难理解。如果这种做法确实有效，其实践意义将不同凡响。

关于控制系统和通信系统的异同，书中写到：“通信是认识世界的手段，控制是改造世界的手段。通信或认识要求主观认识和客观事实符合，而控制或改造要求客观事实和主观理想符合，这就是通信和控制的根本差异。但是两者又都存在符合与不符合的检验问题。信息测度便是检验尺度，信息论中的一些方法和结论完全可以推广到控制论中来。”读者将看到，广义信息测度居然可以用来评价火炮控制效果；通信优化方法居然也可以用来优化控制。本书使控制论和信息论之间的关系更加紧密，更加微妙了。

作者对哲学有独到的研究,他声称自己研究的不是科学的哲学问题,而是哲学的科学问题。人的色觉、美感、认识、欲望、…是怎样从达尔文的单调世界甚至康德的原始星云演化而来的,这是当代有头脑的哲学家不会不思考的问题。作者的美感和色觉理论以及关于“另一种目的”的理论令人耳目一新。我感到这本书在影响我的世界观。

这是一本信息理论书,也是一本信息内容极其丰富的书。这本书将会产生怎样的影响,请大家拭目以待。

李洪兴

1992年4月于北京师范大学

目 次

序	I
第一章 引论	1
1.1 背景	1
1.2 熵的历史回顾	4
1.3 Shannon 信息论的诞生	6
1.4 广义信息论溯源	7
1.5 信息和熵的范畴及本书讨论范围	10
1.6 信息的本质——被反映的特殊性	11
参考文献	13
第二章 Shannon 信息论述评	16
2.1 基于频率解释的概率论	16
2.2 经典通信模型	18
2.3 经典信息量公式和 Kullback 信息公式	19
2.4 Shannon 互信息公式	21
2.5 连续信源的熵和交互熵	22
2.6 信道容量和信道编码	23
2.7 Shannon 熵的编码意义	25
2.8 限失真编码和信息率失真论	26
2.9 Shannon 信息论的局限性	30
参考文献	33
第三章 广义信息论的数学基础	34
3.1 集合 Bayes 公式 随机集落影	34
3.2 客观概率 主观概率 逻辑概率	35
3.3 预测熵及其性质	38

3.4 映射 集合划分 Shannon 熵	39
3.5 广义映射 集合覆盖 广义熵	40
3.6 相似关系 相似度 不相似熵	43
3.7 预测熵和广义熵的编码意义	44
3.8 模糊映射 模糊相似关系 广义交互熵	46
3.9 隶属度确定 模糊分辨率函数	50
3.10 隶属度的柏拉图解释	53
参考文献	53
第四章 广义通信模型和广义信息测度	55
4.1 广义通信模型和 Popper 的科学进化论	55
4.2 信源和信道可变时的概率预测信息	59
4.3 单个事件之间的广义信息量公式	62
4.4 广义 Kullback 信息公式及最佳预言选择	65
4.5 广义互信息公式及其性质	66
4.6 模式识别和天气预报的信息评价	68
4.7 广义自信息 连续信源和信宿的广义互信息	73
4.8 自然语言通信的优点	74
4.9 由广义信息测度得出的科学理论进步标准	75
参考文献	76
第五章 通信优化	78
5.1 信源预测无失真编码定理	78
5.2 用广义信息准则代替预测的均方误差准则	79
5.3 检测和估计的广义信息准则	82
5.4 广义信道容量	85
5.5 限误差信息率及其和信息率失真的关系	86
5.6 保质信息率论——信息率失真论改造	89
5.7 具有相似关系的二元信源保质信息率函数	91
5.8 保质信息率和信源量化等级及主观分辨率的关系	93
5.9 图象视觉信息及通信优化	96

5.10 信息价值 保价值信息率	97
参考文献	99
第六章 信息和统计	100
6.1 信息和 Boltzmann 熵的关系	100
6.2 最大熵原理和限误差信息率的关系	102
6.3 量子统计物理熵的 Shannon 信息论解释	103
6.4 建立在相似关系上的物理学熵公式	105
6.5 有序性 多样性 生物学统计	106
6.6 互信息减少定律——熵增大定律修正	110
参考文献	111
第七章 信息 控制 生产	113
7.1 控制和通信的异同	113
7.2 熵 $H(X)$ 的开关控制意义	116
7.3 理想的集合表示及限误差信息率的控制论意义	117
7.4 信道容量和保质信息率的控制论推广	118
7.5 自由能公式 信息价值公式 生产价值公式	120
7.6 局域平衡自由能公式和投入产出价值公式类比	122
7.7 Maxwell 妖的剩余价值解释	124
参考文献	125
第八章 信息论 进化论 哲学	126
8.1 关于另一种目的的哲学	126
8.2 崇高感和美感的信总论控制论意义	129
8.3 从信息论的角度看色觉机制的进化	130
8.4 颠倒色觉的逻辑可能性和感觉相对论	132
8.5 反映论改造	134
8.6 有待进一步研究的问题	136
参考文献	138
附录论文	140
B 模糊集合代数及其用于语义信息和颜色视觉	140

译码模型:一个对称的色觉机制模型	151
试析达尔文留下的香甜美难题	
——历史唯物主义原理的生物学推广.....	165
论美是促进喜爱情绪的反馈信号	171
色觉所反映的客观内容究竟是什么?	180
由颠倒色觉的逻辑可能性澄清实指定义.....	189
后记.....	194

第一章 引 论

这本书之所以叫广义信息论，是因为：

1. 其中基本理论是 C. E. Shannon(仙农)信息论^[1]的自然推广；
2. 所讨论的是更广范围内的信息：语义信息、感觉信息、测量信号信息、甚至手段和目的之间的信息；
3. 它还涉及控制论、物理学、经济学、生物学和哲学。

1.1 背景

在我们这个新理论层出不穷，同时也是理论和见解没有多少差别的时代，本书所写的好象注定要成为这些新理论中的一个，不过不同的是，作者建立理论并不是因为在研究一些问题之前就立志要建立什么，而是因为作者首先被一些问题所吸引，新理论只是解决这些问题之后的总结。这些问题是：怎样度量一系列颜色或图象实际给予的信息？感觉分辨率怎样影响主观信息量？怎样度量语言、预言和谎言的信息？将有不确定事件发生时，选择怎样的语句可以提供最多信息或信息价值？给定通信的主观效果要求时，客观信息率或平均码长可能压缩到多少？信源和信道可变时信息量如何计算？集合中元素相互类似时，如何度量集合的熵、模糊性及有序性？怎样衡量知识的多少和科学理论的进步？怎样评价控制的复杂性和控制结果的合目的性…？

几乎每个领域都有把已有的正统理论奉为圣经，从而拒斥一切“异端邪说”的卫道士。下面我们向经典理论提几个问题，广义信息论的基本思想也由此看出：

1. 人(或动物)能否接收信息? 人收到信息和更具体更正确地了解事实是不是一回事?

2. 通常人并不知道确切的事实发生的概率和条件概率,而只能根据经验、知识、语言、感觉或测量信号作主观预测,这时信息量如何求法?

3. 常识告诉我们,主观预测与事实相符且精确,所得信息就多,反之,信息就少,信息论如何与常识一致?

4. 实际的通信系统通常是开放的,人对于事实(即信源和信道)的认识总是处于进化之中,这时信息量如何求法?

人们常说,我们的时代正走向信息化时代;然而信息论落后于时代在今日是再严重不过了。虽然 Shannon 理论声名卓著,然而它所涉及的信息仅仅是日常语言所说的信息的一小部分,即被减小的随机不确定性。而对于语义信息、感觉信息、偶然事件及测量数据提供的信息,Shannon 理论无能为力;甚至在经典通信范围内,信源或信道可变时,信息也不可度量。主要由于上述原因,从事信息科学的人大多属于下面两类:一类是信息论工作者,他们研究电子通信理论和应用;其基础理论是 Shannon 信息论。另一类是从事计算机科学、人工智能、情报检索及各种用计算机管理的人;在他们眼里,信息无非是被处理过的有用数据,象某些教科书所定义的,而和 Shannon 信息论毫不相干。他们通常称自己研究的信息为广义信息。有的地方还干脆把计算机系称作信息科学系。除了上述两类,也有少数人试图用 Shannon 理论解决广义信息(如图象视觉信息、语音信息)问题;虽然他们的可贵努力取得了不少可喜成果,但是仍然遇到难以克服的困难,比如,关于视觉分辨率(是主观的,并且往往是模糊的)怎样影响图象视觉信息的问题以及图象数据压缩问题,在理论上就未能很好解决。

时至今日,两极分化日益严重。如果你来到某个大学计算机系问某个信息处理专家:“你们的工作要不要信息论指导?”回答很可能是:“信息论研究的是狭义信息,而我们研究的是广义信息…”

那种自信会让你觉得研究广义信息可以不需要度量广义信息的数学理论。另一方面,搞经典信息论的人往往认为只有自己搞的才是真正的信息科学,而视“信息”的其他理解为浅薄。要有一个统一的信息科学,我们需要一个关于广义信息的数学理论。

作者不是不知道下面的事实:Shannon 理论诞生不久,许多人把它用到日常信息交流场合,于是漏洞百出。为此,Shannon 非常愤慨,并措辞严厉地予以指责;要求人们对他的理论的应用不要超出随机统计领域。这说明什么呢?说明 Shannon 有自知之明。然而,有人却得出可笑结论:Shannon 反对信息论研究语义信息(泛指涉及意义的信息)。似乎为了保卫 Shannon 理论的纯洁性,国际权威信息论杂志 IEEE TRAN. ON INFORMATION THEORY 就拒绝发表研究语义信息的论文。

实际上,任何一种通信都不能完全排除信号的意义问题。Shannon 理论中讨论的信息率失真问题就和信号的意义密切相关。没有意义,哪来失真?由于避免考虑意义,经典的信息率失真理论就注定了残缺不全。失实际上就是主观信息损失,因为不考虑意义,它就只能来自人为定义,而不是由统计确定;强调统计却反而导致统计的忽视。控制系统中的预测质量本来就应该用预测提供的信息作为评价标准,因为排除意义,预测信息就无法度量。

其实,曾和 Shannon 合著《通信的数学理论》一书的 W. Weaver 在该书的一篇论文中就提出通信的三个水平:^[2]水平 A——通信的技术问题,如 Shannon 理论研究的;水平 B——考虑到语义问题;水平 C——考虑到效用或价值问题。他还指出:通讯的语义问题与工程技术方面无关,“这并不意味着通讯的工程技术方面必然地与语义无关。”后面我们将看到,只有考虑意义或语义,Shannon 提出的保真度信息率问题才能得到根本解决。

如果说本书所研究的信息超出随机统计领域,那么它超出的也只是经典的随机统计领域。近年来,汪培庄教授等人建立的随机

集落影理论把模糊集合论建立在随机集合统计的基础上^[3], 不仅把模糊性还原为随机性, 而且也推广了经典的随机统计领域。本书正是基于这种更广意义的集合和统计理论。

捍卫一种理论的最好方法就是发展它, 使旧的内核在新的体系中永生, 而不是为了它的纯洁性而限制它的发展和应用。后面我们将看到, Shannon 诸公式本身就有巨大潜力, 只是形式上作一点点甚至是不易察觉的改动, 它马上就威力倍增——不只可以度量语义信息, 而且可以度量控制结果的合目的性。Shannon 本人未必不愿看到这一点。

1.2 熵的历史回顾

物理学中, 熵指的是热力学系统的某种状态函数, 它是对系统紊乱程度的度量。熵的概念和热力学第二定律密不可分。

1850 年, 德国人 Clausius (克劳修斯) 提出了热力学第二定律: 不可能把热量自低温物体传到高温物体而不引起其他变化^[4]。

1851 年, 英国的 Kelvin (开尔文) 勋爵给出第二定律的另一种表示: 不可能从单一热源吸取热量, 使之变为有用功, 而不产生其他影响。

1864 年, 克氏导出了不等式

$$\delta S \geq \frac{\delta Q}{T} \quad (1.2.1)$$

其意是: 熵 S 的变分大于 (过程不可逆时) 或等于 (过程可逆时) 热量 Q 的变分除以温度 T 。系统封闭时,

$$\delta S \geq 0 \quad (1.2.2)$$

表示封闭系统的熵只增不减; 这也就是热力学第二定律的数学表示。

关于熵和热力学第二定律, 还可以追溯到以研究热机效率闻名的法国人 Carnot (卡洛), 他的可逆循环和不可逆循环概念就隐

含了熵增大思想的萌芽。

熵一开始只有相对值而无绝对值,只有到了 Boltzmann(玻尔兹曼),熵的绝对值才有了定义。1882年,玻氏从分子运动论导出了和连续信源的 Shannon 相对熵相近的熵函数

$$H = \int f \ln f dv \quad (1.2.3)$$

并证明它只减不增。和 Shannon 连续信源相对熵相比,两者差一符号,并且这里的 f 是分布密度而不是概率密度。后来,玻氏又提出了一个更有意义的公式,即通过系统微观状态数 W 表示的熵的绝对值公式。

$$S_b = k \ln W \quad (1.2.4)$$

其中 k 是玻氏常数(等于 $1.38066 \times 10^{-23} \text{JK}^{-1}$; J :焦耳, K :绝对温度)。这一熵公式也可以用于非热力学统计。值得一提的是,这里所说的绝对值也只是形式上的,因为微观状态数的划分总是相对的,和假设的分辨率有关。

Bose(玻色)和 Einstein(爱因斯坦)、Fermi(费米)和 Dirac(狄拉克)分别确定了适于不同微观粒子的量子统计^[5],使玻氏开创的统计物理熵理论得到相当的完善。

熵增大定律是否不可抗拒,与 Boltzmann 同时代的 Maxwell(麦克斯韦)——因其电磁场方程而闻名——提出了著名的 Maxwell 妖问题^[6]。这一问题促进了人们对自然界有序根源的探索。

到了本世纪中叶,Prigogine(普里高津)等人研究了非平衡系统的熵产生问题,得出了著名结论:非平衡导致有序^[7]。

时至今日,熵的问题还没有完全解决,关于负熵、有序性、复杂性等概念还存在很不一致的见解。作者相信,从信息论的角度研究熵可使问题更加清楚。

1.3 Shannon 信息论的诞生

1948 年,美国工程师 Shannon 在贝尔实验室杂志上发表了长篇论文:《通信的数学理论》^[1],虽然 Winer(维纳)等人几乎同时提出了关于信息的熵公式;但是作为电子通信理论基础的互信息公式则是 Shannon 独自提出的;因而说 Shannon 是经典信息论的创始人。

信息论的诞生有两个来源,一是来源于物理学的熵理论。Boltzmann 在讨论熵问题时就说过:熵是对失去的信息的度量。信息论中的熵 $H(X)$ 和玻氏熵 S_b 存在某种等价关系(见 6.1 节),这说明了两者有血缘关系。

信息论的另一个来源是早期人们对电报通信的研究。在 16 世纪, Gilbert 等人就研究了电报电码^[9],这一研究的著名产物是 Morse 电报电码。使用该电码可以用较少的电报符号传递较长的电文。而 Shannon 熵正反映了使用最佳电码时,符号可以缩短的极限。1928 年, Hartley 提出用对数函数度量信息^[9],即用

$$I = \log N \quad (1.3.1)$$

表示确定随机发生的 N 个元素中的一个时得到的信息量。若对数以 2 为底,信息单位是 bit(比特);若对数以自然数 e 为底,信息单位是 nat(奈特)。之所以用对数函数,是因为这样得出的信息量反映了传递相同消息所需的符号长度(使用等长码时),比如,要用十进制数字串表示 N 个电话用户,一个用户号码要用不少于 $\log_{10} N$ 个数字来表示。这样,信息量就有了客观意义。与 Hartley 同时代的 Nyquist 对信道的研究^[10]也是 Shannon 思想的来源之一。Winer 用失真量评价系统状态的估计的做法也对 Shannon 保真度信息率论的提出起了一定作用^[11]。

40 多年来,以 Shannon 理论为核心的经典信息理论在编码、检测等方面取得了不少进展;然而,它远不能满足许多领域的需

要。

1.4 广义信息论溯源

意识到 Shannon 理论的局限性并试图推广它,在 Shannon 理论刚诞生不久就开始了。Weaver 不仅提出了通信的三个水平,还提出了广义通信模型^[2];虽然他没有提出具体的数学公式,但是他的基本思想为后人起了引路作用。

为了解决语义信息问题,R. Carnap(卡尔纳普)等人提出用逻辑概率代替普通(或随机性)概率,然后用 Shannon 熵度量语义信息^[12]。然而这种方法有许多限制要求:可选择语句必须是互不相容的;语句的真假必须是严格分明的;语句的使用必须是正确无误的。由于这些限制,这种方法也就少有价值。不过,用逻辑概率代替普通概率的做法也部分地被作者所采用。

K. Popper(波普尔)似乎是与广义信息理论无关的人,但是他一再强调,科学理论的进步标准就是理论所能提供的信息的多少;用作预测的命题,其先验逻辑概率越小越好^[13]。后面我们将看到,本书提出的通信模型和 Popper 的科学进化模型极为一致,Popper 的基本思想在新的通信模型和信息测度公式中得到了充分体现。

50 年代末,Kullback 提出了 Kullback 信息公式^[14]

$$I_K = \sum_i P^*(x_i) \log \frac{P^*(x_i)}{P(x_i)} \quad (1.4.1)$$

虽然它是 Shannon 互信息公式的特例(见 2.4 和 2.5 节),但是如果把 $P^*(x_i)$ 看作是主观预测的可能性测度,则它度量的就是广义信息。不过是非常特殊情况下的广义信息(见 4.4 节)。

如果先验概率不同,有 $P_1(x_i)$ 和 $P_2(x_i)$, 则两种情况下的 Kullback 信息之差就是

$$I_K = \sum_i P^*(x_i) \log \frac{P_2(x_i)}{P_1(x_i)} \quad (1.4.2)$$