浅谈对信息论的认识

摘 要

关键词：

1 香农研究信息学的缘由Equation Chapter (Next) Section 1

1.1 香农其人

克劳德·艾尔伍德·香农（Claude Elwood Shannon, 1916 年 4 月 30 日－2001 年 2 月 26 日），美国数学家、电子工程师和密码学家，被誉为信息论的创始人。香农是密歇根大学学士，麻省理工学院博士。

1948 年，香农发表了划时代的论文——通信的数学原理[1]，奠定了现代信息论的基础。不仅如此，香农还被认为是数字计算机理论和数字电路设计理论的创始人。1937 年，21 岁的香农是麻省理工学院的硕士研究生，他在其硕士论文中提出，将布尔代数应用于电子领域，能构建并解决任何逻辑和数值关系，被誉为有史以来最具水平的硕士论文之一。二战期间，香农为军事领域的密码分析——密码破译和保密通信——做出了很大的贡献。

1.2 香农生平

香农生于密歇根州的 Petoskey。父亲克劳德（1862 - 1934）与他的姓名完全相同，是新泽西州早期移民的后裔，曾自主创业经商，也担任过审核遗嘱的法官。母亲玛贝尔·沃夫·香农（1890 – 1945）是德国移民的女儿，职业是语言学教师，曾长期担任密歇根州 Gaylord 高中的校长。香农人生的前 16 年都是在 Gaylord 度过的，他在那儿接受了公立学校教育，并于 1932 年从 Gaylord 高中毕业。香农对机械和电气电子表现出了极大爱好。他最优秀的学科就是科学和数学，并在家中制作了模型飞机、无线电控制的模型船和一个可与半英里内的朋友家联系的无线电报系统。大一点的时候，他做过西联汇款的投递员。

香农孩提时期仰慕的英雄是托马斯·爱迪生，后来他才知道自己是托马斯·爱迪生的远房亲戚。他们都是约翰·欧格登（John Ogden）的后裔。约翰·欧格登是一个殖民领袖，也是许多杰出人物的先祖。

1.3 布尔理论和二战前的研究

1932 年，香农进入密歇根大学学习，在大学的一门课程中接触到了乔治·布尔的理论。1936 年大学毕业时，香农获得了两个学士学位：电子工程学士和数学学士。不久，香农进入麻省理工学院开始研究生学习，参与了万尼瓦尔·布什的微分分析机（Differential Analyzer）的相关工作。微分分析及是一种模拟计算机，是现代电脑的鼻祖。

在研究微分分析机的自组织（ad hoc）电路时，香农发现引入布尔理论的概念会带来很大的好处。在 1937 年硕士论文的基础上，香农在 1938 年发行的 Transactions of the American Institute of Electrical Engineers 上发表了著名论文“A Symbolic Analysis of Relay and Switching Circuits”。由于这篇文章[2]，香农于 1940 年被授予美国 Alfred Nobel 协会美国工程师奖。哈佛大学的哈沃德·加德纳称香农的硕士论文“可能是本世纪最重要、最著名的硕士学位论文”。

在这篇论文中，香农证明了布尔代数和二进制算术可以简化当时在电话交换系统中广泛应用的机电继电器的设计。然后，香农扩展了这个概念，证明了基于机电继电器的电路

能用于模拟和解决布尔代数问题。

用电子开关模拟布尔逻辑运算是现代电子计算机的基本思路，香农的工作成为数字电路设计的理论基石，完全取代了之前盛行的 ad hoc 方法。Vannevar Bush 建议香农将类似的数学方法应用于孟德尔遗传学，香农接受了这个建议，写出了 An Algebra for Theoretical Genetics。凭此论文，香农于 1940 年获得麻省理工学院博士学位。

1940 年，香农成为普林斯顿高等研究院的研究员。在那里，香农共有很多机会与当时有影响力的科学家和数学家交流，比如阿尔伯特·爱因斯坦、赫尔曼·外尔和约翰·冯·诺依曼，现代信息论的思想逐渐在他脑海中成型。

1.4 二战期间的研究

二战期间，香农加入贝尔实验室，研究火力控制系统和密码学，相关课题直属国防研究委员会领导。

在贝尔实验室，香农遇到了担任数值分析员的 Betty。两人于 1949 年结婚。

1943 年，香农有机会和数学家、密码学家艾伦·图灵合作。图灵被派到华盛顿和美国海军交流破译德国的北大西洋潜艇舰队密码的成果，并在贝尔实验室待了一段时间。香农和图灵在一个自助餐厅见面。图灵向香农介绍了现在被称为“图灵通用机”的概念。香农对此很感兴趣，因为图灵机的概念和香农自己的很多想法相吻合。

1945 年，战争进入尾声，国防研究委员会 NDRC 的使命即将结束。在正式解散之前，NDRC 决定将重要研究成果整理成册，其中有一篇论文“火力控制系统的数据平滑和数据预测”是香农和 Ralph Beebe Blackman、Hendrik Wade Bode 一起写的，他的思路和“通信系统中将信号和噪声相分离”是类似的，也就是说，香农在火力控制系统研究中已经发现了后来成为信息论的基本概念和框架体系。

战时香农在密码学领域的研究于通信领域的关系更加密切。1945 年，香农向贝尔实验室提交了一份备忘录，题目是“密码学的一个数学理论”，之后在 1949 年以“保密系统的通信理论”的标题在 Bell System Technical Journal 正式发表，包含了很多在“通信的一个数学理论”出现的概念和数学公式。香农说，战时对通信理论和密码学的研究使他认识到“两者密不可分”。

还是在贝尔实验室，香农证明了一次性密钥（cryptographic one-time pad）是无法被破译的。香农同时证明了一个无法被破译的密码系统的密钥必须有以下特征：完全随机；不能重复使用；保密；和明文一样长。

1.5 回顾香农的研究历程

从以上简短的香农的研究历程中，可以看出，香农对信息论的研究，起源于对密码学的研究，而从客观历史上面看，密码学在信息论诞生很久很久之前就有了，虽然在古代没有信息论这门科学，不过已有不少加密手段，有了信息论作为基础，可以更加深刻地为很

多行业服务，比如了解信息论之后可以回答如下问题：怎样才能保证通信过程中信息不会丢失、具体如何操作才能最大限度压缩数据、通信过程中信息传输有无速率上限。

而香农的理论，是在非纯理论研究过程中建立的，香农本人也并非是一个纯理论学家，而且是一位杰出的工程师。所以，学习信息论、研究信息论，需要从实际需求与实际问题出发，单纯研究内在的概率论模型是一种无源之水型的学习。本着这种原则，本文给出信息论起源与信息模型的构建的一些解释与看法。

Equation Chapter (Next) Section 1

2 信息量的度量

假设有一台会说话的机器，如果这台机器只会发出一个音节，而且是在相等间隔之内连续进行发音，那么这台机器几乎不会告诉我们任何信息，因为在任何时刻我们听到的东西都是一样的。

然而，如果这台机器会发长音与断音，那就不一样了，长音与短音可以进行很复杂的组合，从而带来很多信息。比如著名的莫尔斯代码，就是用点“.”和划“_”进行信息编码。

再进一步，假如这台机器比较智能，它可以说一个一个的汉字，那就更加复杂了，它说几个汉字音节可能就带来很大的信息量。

令机器一次发出的一个音节的可能结果为随机变量X，X可以看作是信息的载体。

2.1 信息量与随机变量样本空间容量有关

从以上三种情况来看，可以构建一个比较粗陋的信息量模型。那么在第一种情况下，X的样本空间[3]为单元素集合，第二种情况下，X的样本空间为{. , _}，而在第三种情况下X的样本空间为新华字典中所有字构成的集合。

如果把信息看作是一个随机变量，当X只有一个取值时，信息量为0，因为无论机器怎么说，带来的信息都一样；而如果样本空间的元素数目不为1，那就大大不同了，通过排列两个或者多个元素的位置，可以表达丰富的信息。所以，信息与随机变量X的样本空间大小有关。Equation Section (Next)

2.2 信息量与随机变量概率分布有关

当两个机器A，B同时在说话，而且两台机器说话音节的样本空间相同，是否信息量就一样呢。假设机器A发音只发某一个音节，其他音节几乎不发，那么同样的样本量下，发音节比较平均的机器B所带来的信息量比较大。Equation Section (Next)

2.3 信息量数学模型的性质

明确了以上两个需要考量的因素，在建模之前还需要对信息量这个数学模型的性质做规划。

1. 信息模型应该具有可加性，如果说了一段话，之后又说了一段话，那么两段话加起来，信息量不应该变少。当然，如果第二段话仅仅是全部否定了第一段话，也不能认为两段话加起来等于什么都没有说，因为这里只是度量说的音节，而不是逻辑判断。

2. 信息量不应该是一个负数。

3. 如果随机变量X的概率分布仅仅发生了很微小的变化，那么其信息量不能发生很大的变化，换言之，信息量应该是关于$f(X)$的连续函数。

现在观察如下函数：

$$y=\log_2\frac{1}{p(x)} \tag{2.3.1}$$

这个函数中的$p(x)$是当随机变量X取值为x时的概率，所以$0<p(x)\leqslant 1$，假设随机变量X是连续随机变量，那么很显然，这个函数具有以上三个要求的性质，而且如果一个随机变量取某一值的概率特别大，那么从逻辑角度看，这个信息量还是比较小的，因为这种情况比较接近机器只会说一句话的情况。

根据定义，信息熵就是平均意义下发生一件事情我们得到的信息量的大小，所以仅仅在$X=x$下是不够的，从概率论角度可以借鉴一下随机变量的数学期望模型：

$$E(X)=\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x \tag{2.3.2}$$

进而构造如下函数：

$$H(X)=\sum_{i=1}^{\infty}p(x)\log_2\frac{1}{p(x)} \tag{2.3.3}$$

3 参考文献：

[1] SHANNON C E. A Mathematical Theory of Communication [J]. The Bell System Technical Journal, July 1948, 27(3): 379 - 423.

[2] SHANNON C E. A Symbolic Analysis of Relay and Switching Circuits [J]. Electrical Engineering, Dec. 1938, 57(12): 713 - 23.

[3] 赵彦晖. 概率统计 - 2 版 [M]. 北京: 科学出版社, 2015.7.