# 信息论奠基人——香农

聪明如香农者竟然也在信息论应用上自囿。

提到信息科学，有两位学者不会被遗忘：一位是维纳，他是控制论和仿生学的先驱；另一位就是这里要介绍的香农，他是信息论的奠基人。

信息已成为时髦名词，但真正懂得个中奥秘者不多。在香农以前，一般人包括电讯工程师在内都认为，通讯线路中传送的是电信号。这没错，但未能道出通讯的本质。通讯的目的不是传送电信号，而是传送信息。传送信息有多种方式：可以打电话；也可以写信……打电话固然涉及到电信号，但发话人与话筒之间的传送是靠声波而不是电信号，受话人从耳机听到的也是声波；至于写信则完全不靠电信号。所以，通讯的本质不是电信号而是信息。信息是什么？维纳有一句名言：“信息就是信息，不是物质也不是能量。”这个回答很俏皮，只说信息不是什么，而未作正面回答。维纳非常聪明，他知道信息的定义不是容易下的。

香农也绕过这个难题，致力于信息的定量研究。

****

**信息论奠基人香农（Claude E．shannon，1916-2001）**

香农以两篇论文奠定了信息论的基础。第一篇是《开关电路的符号分析》，阐明如何用开关电路进行逻辑运算。虽然香农的动机是解决以电话交换机取代接线员的问题，实际上这篇论文为现代数字计算机的基本原理奠定了基础，计算机中的逻辑电路全由开关组成。开和关代表计算机所采用的二元码中的0和1。任何信息都可以用一串二元码的系列来表示，其重要性自不待言。

香农的第二篇论文是《通讯的数学理论》，被公认是为信息论奠基的经典著作。文中分析了各种信息过程的定量关系，给出了信息量的计算公式。香农还引入了信息量的单位——比特（bit），它是二元码中一个码所包含的信息量，这个单位一直沿用至今。信息能够定量，是香农的一大贡献。

这篇论文的另一个贡献更大，说来近乎神奇。任何通讯线路中都有噪音，噪音的混入使得要传送的信号模糊不清，因而影响通讯的质量。例如，通过质量差的电话线路通话时就听不清楚，容易产生误解。过去一般认为，有噪音时就无法完整无缺地传送信息。香农说：不对！不管有多大的噪音，总有办法完整无缺地传送信息。这出乎一般人意料之外，但细思量确有道理。举一个熟悉的例子：机声隆隆的车间里噪音震耳欲聋，说话完全听不见——根本无法传送信息。香农说有办法！办法之一是提高音量，放声大喊以盖过噪音。办法之二是重复，说一遍听不清，再说一遍，不行再重复下去，直到听清楚为止。“这是常识，有什么神奇？”不错！现在看来是常识。但任何事情都是：不知为神奇，知之为常识。譬如先民不知雷电而敬若神明，现在大家知道雷电乃雨云所带之正、负电荷放电，就成了常识。

香农通讯理论将上述例子中的基本思想推广到一切有噪音的通讯系统，以严格的数学公式给出了存在噪音时能完整无缺地传输信息之最大信息量，这具有重要的实用价值。例如，我们看到从火星传回的照片是如此清晰，如同亲眼目睹那样纤毫毕露，感到科学家神通广大，简直不可思议。其实就是利用了香农通讯理论，才能在无线电波传输中消除噪音的干扰，以保证传送照片的清晰度。又如，从遥远目标反射回的雷达信号极其微弱，往往淹没住比它强成千上万倍的噪音中，找回这样微弱的信号好比大海寻针。利用香农通讯理论，照样能办到。总之，香农通讯理论已成为现代通讯系统的基石。

香农信息理论的应用远远超出通讯领域。

常人对电脑的神通广大感到难以理解。电脑能进行计算，还能进行文字和图像处理；电脑会弈棋，击败过世界冠军；电脑还会逻辑推理，能证明数学定理；至于学校利用电脑教学，医院利用电脑诊断，银行利用电脑理财，……就更不用说了。电脑这些几乎是万能的本领从何而来？香农给出了答案：原来电脑所从事的都是信息处理过程，都包含着一定的信息量。根据香农信息论，无论数字、文件、图像、弈棋的规则和策略，逻辑的规则和语言，数学的公式和定理，以及教学资料，病人病历、银行帐册，等等，不管这些信息的内容是什么，都可以化为0和1的二元码系列。电脑所处理的只是一长串0和1，至于这些0和1代表什么，那是程序师的事，电脑根本不考虑。原来电脑的“万能”来源于它只管信息而不问内容，这不就很好理解了吗？

香农信息理论基于概率统计，按照他的公式，对特定的单个事件而言，发生的概率越小，所具有的信息量就越大。谚云：“狗咬人不是新闻，人咬狗才是新闻。”这完全符合香农信息论：人咬狗极少发生，概率极小，这个事件具有极大的信息量，所以是大新闻。

概率小的单个事件虽然具有最大的信息量，但极少出现，按照香农的信息量公式，它在整个事件系列所包含之总信息量中占有的份额（即平均信息量）却最小。占有最大平均信息量的是具有大慨率的最频繁出现的事件。由此可以得出一些有趣的结论：中文单字的出现概率可以从大量书籍报刊中统计出来，它等于该单字出现的次数除以所有这些资料包含的总字数。你猜什么具有最大的出现概率？原来是标点符号，它比任何中文单字出现的概率都高得多，不信就请在这篇文章中统计一下。按照香农信息论，这意味着标点符号具有最大的平均信息。中国古籍大多不用标点符号，因此而损失了大量的信息？可不是吗？对古文作不同的断句，往往产生迥然不同的含义，有的甚至闹出笑话，就是因为原文缺少标点符号的信息量所致。传说鲁迅的一本书被少算了稿费，他去质问出版商，得到的回答是：“标点符号不计酬。”鲁迅下一次交的书稿就完全没有标点符号，出版商收到后傻了眼。鲁迅的做法符合信息论，他将平均信息量最大的那一部分抽掉，难怪出版商傻眼了。

信息论已成为数学的一个分枝，此外在密码学、物理学及生物学等领域中也有许多重要的应用，以后有机会再谈。

香农的信息论研究工作是在著名的贝尔实验室完成的，后来他去麻省理工学院当教授，致力于研究人工智能。

香农取得这些成就不是偶然的，他学习认真，工作勤奋，具有独到的眼光。他的一位同事说：“香农认为，问题越是困难，发现新东西的机会就越大。”又说：“无论什么事，他总是以喜悦的心情去做，并且会想尽一切办法做好。”香农在贝尔实验室工作时，白天整天把自己关在办公室里，晚上骑着一辆独轮车从山坡上冲下来。香农的办公室的门经常关着，但有人敲门求教时，他总是热情接待，尽量设法帮助。

香农的信息论得到广泛的应用，他当然很高兴，但也有例外。当他听说几位主修英国文学的研究生在论文中试图将信息论用于研究文学时，评论道：“信息论的重要性已经被吹嘘得超过了它的实际成就。”不知这是否是香农的自谦之词，如果是他的真意，我不敢苟同。

依我看，信息论的基本原理是普适的。香农信息论具有一个重要的特点：它只考虑信息的量，而完全不计信息的质。说得具体些：它只管信息的统计特性，完全不管信息的内容——否则电脑的神通就不可能那么广大了。由于这一特点，信息论的应用就特别广泛。信息论的同一数学公式可用于通讯，可用于测量，可用于电脑，还可用于生物遗传，甚年可用于研究人的认知和思维过程。总之，凡是有信息的地方，信息论就可以应用。文学包含信息，信息论当然可以应用。我不明白为什么香农不承认这一点。

实际上不仅文学，信息论也适用于研究美学，由此可以解释许多现象，导出一些有用的结论？我曾经说多：没有考虑信息论的美学是不完整的。

香农已于2001年2月24日逝世。一代宗师，遽归道山。以他命名的信息论以及在亿万架电脑中闪烁飞驰的“比特”滚滚洪流将与世长存。