信息熵只是一个数量

许秋雨，2019 年 10 月 26 日

数学的重要任务（如果不算是最重要的任务）之一就是数学表示及量化，使得可量，可算，可比等。比如，对随机变量的表示和量化就是概率论。

要完全刻画一个随机变量，那就需要它的概率分布函数或者密度函数。遗憾的是，概率分布函数或者密度函数是一个函数，它不易被利用，如跟别的随机变量做比较，也不易被刻画。为了应用方便起见，人们就利用它的一些简单化后的数量，如期望，即均值，又如方差，等等。这些都具体到一个数值，好计算，好比，也好用。

信息熵（Information Entropy）也只是关于随机变量的一个数量，一个香农为数字通信发明的数量。一般来说，它就是在平均意义下，要用多少个比特来表示一个随机变量的概念。

对一个正整数 p 来说，如果用人为的二进制来表示它，约需要 log（p）个比特来表示。如果 p 是一个小于 1 的正数，那它的倒数就是一个大于 1 的数，这样就可以用二进制来表

示它的倒数，即用约 log（$1/p$）＝－log（p）个比特来表示 p 的倒数，或者等价于说 p 了。

如果一个随机变量的密度函数是 p_i，i 是非负整数指标，如上面所说，可用约需－log（p_i）个比特来表示概率值 p_i。所以对所有概率值来说，在平均意义下约需要用

$$\Sigma_i \ -\log(p_i)\,p_i$$

个比特来表示它们。这正是这个随机变量的信息熵，也正好确确实实是表示这个随机变量的平均比特数，而哈夫曼编码法就是一个构造性的证明。

您也许会说，上面说的只是离散随机变量。但是正像积分计算是级数计算的推广一样，连续随机变量也只是离散随机变量的推广，从计算的角度来说，没啥大的区别。当然从数学的严谨性来说，也许会有比较大的区别，这里就不多说了。

所以，信息熵就是关于随机变量的一个数量，正如均值或者方差或者中值一样，只是它代表的是平均多少个比特来表示一个随机变量。对于一串数据来说，本来也没有什么物理意义，作为一个数学量，能最少用多少个比特（平均意义下/率）来表示它，即它的信息熵，听上去好像比简单地用它的均值或者方差丰富了很多。但是，就真正的物理意义来说，

信息熵比均值或者方差到底又多了多少含义呢？它们就只是关于一个随机变量的两个数值。

如果要说信息熵的物理意义的话，有人会把信息熵与物理上的熵联系起来，把信息熵与不确定性挂勾，即信息熵大，不确定性就大。信息熵大说明用来表示一个随机变量（或者一组数据）的比特数大，说明此组数据的变化大，这应该没错。其实方差大也表示数据的变化大，或者说不确定性大，不是么？

我个人觉得过去几十年、特别是现在、信息熵或是（香农）信息论已经被泛滥地利用到了很多领域，如生物和智能等等。当然，只要是数据，您都可以用信息熵或者叫信息论的方法运行一下、即试一下，有没有真的好处？谁知道么。但是，我个人觉得（香农）信息论就只是对数字通信（或与数字通信有关的领域）真有用，它是为数字通信而生也很有可能是为数字通信而“死”。

不可否认在不少情况下，跨学科的应用会带来突破性的发展。但是，从另一方面，我又觉得任何一个数学工具的应用都是有限的，不是万能的，大家都可以用它来试一试，但千万不能迷信于任何人或者任何事或者任何工具。