4. 信息和能量间的类比

F. Herrmann and G. Bruno Schmid

【内容摘要】 一个信息贮存系统的总熵可以被分解为几个独立的部分，即分解为几个没有共同的独立变量的函数。它们中的其中之一代表计算机用户所关心的信息(=熵)。这种分解相当于把整个系统分解为几个无不相干的子系统。它类比于将一个系统的总能量分解为几个独立的部分（通常所指的能量形态)。在这二种情况中，人们通常所关心的部分比其他部分要小许多数量级。

一、引 言

一个物理系统的熵 S 可以表示为系统微观态的几率 $p(i)$ 的函数(Reif 1965)：

$$S=-k\sum_{i}p(i)\ln p(i), \tag{1}$$

式中 k 是玻尔兹曼常数。系统的宏观态由几率 $p(i)$ 的分布 $\{p(i)\}$ 决定。

系统信息 I 的 Shannon 表示式(Shannon and Weaver 1949)与 (1) 式有相同的数学形式：

$$I=-f\sum_{i}p(i)\ln p(i). \tag{2}$$

这里 I 是信源每发出一个符号所代表的信息量，$p(i)$ 是发出第 i 个符号的几率，f 是比例常数。I 通常的单位是 bit（比特)。在这单位中，f 的值为(1/ln2)bit.

可以将方程(2)应用于一个热力学系统中，将它理解为观察者所能获得的关于这个系统的信息（假如他将去接收一个能告诉他系统处于什么微观态的信号)。由于这个理由，通常把由 (2) 式所定义的物理学上的熵称为(观察者)“失去的关于系统微观态的信息量”（Rothstein 1951，1952 a，b，Jaynes 1957)。尽管这二个名词——熵和失去的信息——的意义有点含糊，但早已证明，由(1)式定义的熵和 Shannon 的信息表示式 (2)式在物理上是相同的(Tribus and McIrvine 1971)：

$$S=(k/f)I. \tag{3}$$

这一关系表明：

$$1\,\mathrm{bit}=0.96\times10^{-23}\mathrm{JK^{-1}}\simeq10^{-23}\mathrm{JK^{-1}}. \tag{4}$$

现在我们根据这一关系来考虑一个信息贮存装置，例如一台计算机的信息贮存装置。这一宏观系统的熵(=信息)为 $1\mathrm{JK^{-1}}$。根据(4)式，它等于 10^{23}bit.

然而，如果我们去问计算机设计人员或用户：计算机的记忆装置中贮存着多少信息。我们可以料想，回答是“1 Mbit”，即比实际值小 10^{17} 倍。

这一差别的原因是什么？对计算机用户重要的信息量怎么会比其余的信息量小得如此之多？本文的目的就是要用热力学来回答这些问题。在 2、3 二节，我们将证明，一个信息贮存装置的总熵 S 可以被分解为两部分：

$$S=S_{nt}+S_{t}, \tag{5}$$

式中 S_{nt} 为非热熵 (non-thermal entropy)，是贮存着的信息；S_{t} 为热熵(thermal entropy)，是占多数的其余部分。在 4、5 二节，我们将说明，(5) 式类比于一种我们更熟悉的对贮存在系统中的总能量的分解，即把它分解为不同的“能量形态”。

二、1比特计算机记忆单元的状态的确定

考虑一个能用来信息贮存的处于二个可能状态之一的物理系统，例如，可以沿二个可能方向之一磁化的磁铁。沿这些方向的磁化强度是 $+m_0$ 或 $-m_0$。

现在假定系统存在 $m=+m_0$ 的状态（几率为1）。这样，系统 N 个微观态 (i) 的几率分布 $\{p(i,m_0)\}$ $(i=1,2,\cdots,N)$ 就完全确定了。如果后来系统在保持温度和其他宏观变量不变时处于 $m=-m_0$ 的状态，其几率分布 $\{p(i,-m_0)\}$ $(i=1,2,\cdots,N)$ 与前面的相同。

下面我们来考虑一个 Shannon 信息源。这信息源能产生二个信号（用脚标 $j=1$ 或 $j=2$ 表示）之一。发射任一个信号的几率表示为 $p^m(j)$。每当信息源发射一个信号时，前面提到的二进制记忆系统就相应地开启一次。

对于二进制记忆系统，记忆值的意思是微观态的数目从 N 加倍到 $2N$。这样就存在二个类别，每个类别都有 N 个微观态。对于 $j=1$ 类，$m=+m_0$；对于 $j=2$ 类，$m=-m_0$。某一类的每一个微观态与另一类的一个微观态相对应，反过来也是。

某一类的一个微观态 i 仅仅在 m 的值上与另一类相对应的微观态 i 不同。每个微观态由二个数 i 和 j 表征。它出现的几率 $p(i,j)$ 为二个独立的几率 $p^o(i)$ 和 $p^m(j)$ 的乘积：

$$p(i,j)=p^o(i)p^m(j), \tag{6}$$

这里，

$$\sum_{i=1}^{N} p^o(i)=1 \tag{7}$$

和

$$\sum_{j=1}^{N} p^m(j)=1. \tag{8}$$

三、信息贮存系统的熵的分解

将(6)式代入(1)式，对 i 和 j 求和，并应用(7)和(8)式使结果简化，我们得到：

$$S=-k\sum_{i=1}^{N} p^o(i)\ln p^o(i)-k\sum_{j=1}^{N} p^m(j)\ln p^m(j). \tag{9}$$

(9)式右边的第一个求和部分是贮存装置具有的熵（假如 m_0 的每个可能的值的几率为1）。我们把这部分熵称为总熵的热部分，用符号 S_t 表示。(9)式右边的第二个求和部分是总熵中通讯理论家最关心的那部分。我们把这部分熵称为总熵的非热部分，用符号 S_{nt} 表示。因此，

$$S(p^o(i),p^m(j))=S_t(p^o(i))+S_{nt}(p^m(j)). \tag{10}$$

(10)式表示可以将系统的总熵分解为二个独立的部分：第一部分仅仅与 $p^o(i)$ 有关；第二部分仅仅与 $p^m(j)$ 有关。这相当于把系统分解为两个无不相关的支系统。

$p^o(i)$ 和 $p^m(j)$ 的相互独立性等效于这样一个事实：在相应的子系统之间不存在热平衡。实际上，这一事实允许我们把与 $p^m(j)$ 有关的子系统首先用作信息贮存装置。另一方面，由 $p^o(i)$ 表征的子系统将大体上与整个系统的周围环境处于热平衡。S_t 通常仅仅是温度 T 的函数，而 S_{nt} 是磁矩的平均值 $\langle m\rangle$ 的函数。磁矩的平均值

$$\begin{aligned}\langle m\rangle &= p^m(1)m_0+p^m(2)(-m_0)\\ &= p^m(1)m_0+(1-p^m(1))(-m_0)\\ &= (2p^m(1)-1)m_0.\end{aligned} \tag{11}$$

因为给定$\langle m\rangle$时$p^m(1)$(相应地 $p^m(2)=1-p^m(1)$)可由上式确定，所以$S_{nt}=S_{nt}(\langle m\rangle)$，这样就确定了相应于$S_{nt}$的分布。因此，我们可以把(10)式系统的总熵$S$改写为

$$S(T,\langle m\rangle)=S_t(T)+S_{nt}(\langle m\rangle), \tag{12}$$

即用T和$\langle m\rangle$这二个"宏观"变量来表示。

四、系统能量的分解

按照分解系统总熵的方法，对能也已经作了同样的分解(Falk and Ruppel 1976, Falk et al 1983)。考虑一个任意系统的能量E，它作为互相独立的变量$\{x_i\}$的函数的形式是

$$E=E(x_1,x_2,\cdots). \tag{13}$$

对于某种系统，这个函数分为几个由不属于其他项的变量决定的几个项。

为了清楚地说明这一点，我们来考虑一个运动着的电容器。令这个电容器的动量为p、电量为Q、质量为M、电容为C，总能量E可以写为

$$\begin{aligned}E&=E(P,Q)=E_0+E_1(p)+E_2(Q)\\&=E_0+p^2/2M+Q^2/2C.\end{aligned} \tag{14}$$

分解后的每一个独立的项通常有一个不同的名称。在上面这个情况中，E_0叫"静止能量"，$E_1(p)$叫"动能"，$E_2(Q)$叫"电场能"。当然，这样的分解不总是可能的。例如，理想气体的能量作为熵、体积和物质的量的函数就不能作这样的分解。

五、系统熵的分解和能量的分解的比较

有些系统的总熵可以按照(12)式来分解(回顾第三节)。这种系统对数据贮存是有用的。同样，有些系统的总能可以

按照类比于(12)式的关系式(14)式来分解(回顾第四节)。本节将证明，熵和能量间除了这一类比外，对于典型的技术系统分别出现在(14)式和(14)式中的每个独立项的大小的相对数量级方面也存在着类比关系。

对于一台计算机的记忆单元，(12)式中S_t的值为$1JK^{-}$数量级。S_{nt}的典型值是$1Mbit=10^{-17}JK^{-1}$，即比S_t的值小得多。因此，当人们关心这种系统的总熵S时，取S本身的值或取S_t的值是没有关系的。

对于一个重为$10g$，电容为$1\mu F$的电容器，(14)式中的E_0的值为$10^{15}J$的数量级。如果这个电容器以$1ms^{-1}$的速度运动，电压为100V，则(14)式中的E_1和E_2的值都为0.005J。因此，当人们关心这种系统的总能时，取E本身的值或取E_0的值是没有关系的。

S_{nt}和E_1(或E_2)的值分别比S_t和E_0的值小得多，甚至比S_t和E_0随时间的自然涨落小得多。例如，一台计算机记忆单元的温度仅变化 0.01K 就会引起S_t的值以$10^{-3}JK^{-}$的数量级变化，即比S_{nt}的值大14个数量级。同样，对于一辆加速运动的汽车，由于轮胎的磨损所引起的总能的变化(根据$\Delta E=\Delta mc^2$)大于由于动能的增加而引起的总能的变化。

我们再回到引言中所提出的问题："对计算机用户重要的信息量怎么会比其余的信息量小得如此之多？"。这个问题类比于这样一个问题："对汽车驾驶员重要的动能怎么会比汽车的静止能量小得如此之多？"对这些问题的答案可由(12)和(14)式得出。要确定S_{nt}或E_1的值，我们不是去确定熵或能量本身的值，而是要根据分别决定S_{nt}或E_1的其他变量的值：对计算机记忆单元来说，只要确定磁矩m的平均值；对汽

车来说，只要确定动量 p 的值。由这些值，就可以分别计算出 S_{nt} 和 E_1 的值。

六、结 束 语

对计算机用户重要的信息是信息贮存系统的物理熵的很小一部分。因为这部分“非热”熵的值决定于与其余的熵（即”热“的部分)无关的变量，所以它是可以确定的。将系统的总熵分解为几个独立的项(即几个独立变量的函数)，相应于把整个系统分解为互不相干的几个子系统。这种分解类比于把系统的总能分解为几个独立的项。在这二种情况中，人们通常所关心的那个项比其余的要小许多数量级。

参 考 文 献

Falk G, Herrmann F and Schmid G B 1983 Am.J. Phys. 51 1074

Falk and Ruppel W 1976 Energie und Entropie (Berlin: Springer) p 137

Jaynes E T 1957 Phys. Rev. 106 620

Reif F 1965 Fundamentals of Statistical and Thermal Physics (New York: McGraw-Hill) p 219

Rothstein J 1951 Science 114 171

—1952a Phya. Rev. 85 135

—1952b J. Appl. Phys. 23 1281

Shannon C E and Weaver W 1949 The Mathematical Theory of Communication (Chicago: University of Illinois Press)

Tribus M and McIrvine E C 1971 Sci.Am.September 179

(译自 Eur. J. Phys. 7(1986) 174—176)