GPS Information and Rate-Tolerance

and its Relationships with Rate Distortion and Complexity Distortions

Chenguang LU[1]

Abstract: I proposed rate tolerance and discussed its relation to rate distortion in my book "A Generalized Information Theory" published in 1993. Recently, I examined the structure function and the complexity distortion based on Kolmogorov's complexity theory. It is my understanding now that complexity-distortion is only a special case of rate tolerance while constraint sets change from fuzzy sets into clear sets that look like balls with the same radius. It is not true that the complexity distortion is generally equivalent to rate distortion as claimed by the researchers of complexity theory. I conclude that a rate distortion function can only be equivalent to a rate tolerance function and both of them can be described by a generalized mutual information formula where $P(Y|X)$ is equal to $P(Y|\text{Tolerance})$. The paper uses GPS as an example to derive generalized information formulae and proves the above conclusions using mathematical analyses and a coding example. The similarity between the formula for measuring GPS information and the formula for rate distortion function can deepen our understanding the generalized information measure.

Keywords: generalized information, semantic information; GPS information; rate distortion; data compression; Kullback formula; Bayesian formula

GPS 信息和限误差信息率

——及其和信息率失真及复杂性失真之间的关系

鲁晨光 [1]

摘　要：在我 1993 年发表的《广义信息论》中，我提出限误差信息率(或者叫: 信息率容差，rate-tolerance)并讨论了它和信息率失真之间的关系。最近我了解到 Kolmogorov 的结构函数及基于其复杂性理论的复杂性失真。我现在知道复杂性失真只是限误差信息率在限制集合由模糊变为清晰并呈球形时的特例; 复杂性理论研究者肯定的复杂性失真和信息率失真之间的一般等价关系是不对的；一个信息率失真函数只能等价于一个特定的限误差信息率函数，两者都能用 $P(Y|X)=|P(Y|$容许误差)时的广义互信息公式描述。文中用 GPS 作为例子推导出广义信息公式，并且通过数学分析和编码举例证明了上述结论。度量 GPS 信息的公式和信息率失真函数表达式的相似性可以加深我们对广义信息测度的理解。

关键词：广义信息；语义信息；GPS 信息；信息率失真；数据压缩；Kullback 公式；Bayes 公式

1 引言

我写这篇文章有两个原因。其一是，我在 1993 年发表的专著《广义信息论》[1]中介绍了如何推广 Shannon 信息论[2]到更一般领域，包括语义信息和感觉信息领域。我在书中介绍信息率失真的两个新版本：限误差信息率 $R(A_J)$（本文记为 $R(T)$，英文写成 rate-tolerance, 直译为：信息率容差）和保精度信息率 $R(G)$，讨论了它们和信息率失真之间的等价关系。后来又发表文章进一步讨论这些问题 [3] [4]。最近，我了解到基于 Kolmogorov 复杂性理论的结构函数[5]和复杂性失真[6]，发现复杂性失真只是限误差信息率的特例，复杂性理论研究者肯定的复杂性失真和信息率失真之间的一般等价关系是不成立的。其二是，我在上书中同时介绍了如何根据人眼色觉分辨率度量色觉信息，最近我发现，GPS 提供信息以类似的方式；并且通过对 GPS 信息的度量，可以加深我们对信息率失真的理解。这篇文章中，我首先介绍如何度量 GPS 提供的信息，然后讨论和数据压缩理论相关的的那些函数的等价关系。

2 GPS 信息——从统计信息到预测信息

2.1 GPS 精度和集合 Bayes 公式

全球定位系统（Global Positioning System）简称 GPS。我们通常假设目标实际位置以 GPS 读数为中心正态分布。

我们假设信源是带有 GPS 跟踪器的被盗小车，为简化起见，其位置用一维离散随机变量 X 表示，实际可能位置是 x_1, x_2, … 它们构成集合 $A=\{x_1, x_2, \ldots\}$。 我们用 Y 表示 GPS 读数， Y 的实际取值是 $y_1, y_2, \cdots$； 它构成集合 $B=\{y_1, y_2, \ldots\}$。 读数 y_j 即 x_j 的估计，通常写为 $\hat{x}_j$。那么在信源等概率假设下，读数 y_j 出现时，x_i 发生的条件概率分布是（如和图 1 所示）：

1. Email: survival99@gmail.com, Homepage: http://survivor99.com/lcg

$$P(x_i \mid y_j) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x_i - x_j)^2}{2\sigma^2}\right] \tag{1}$$

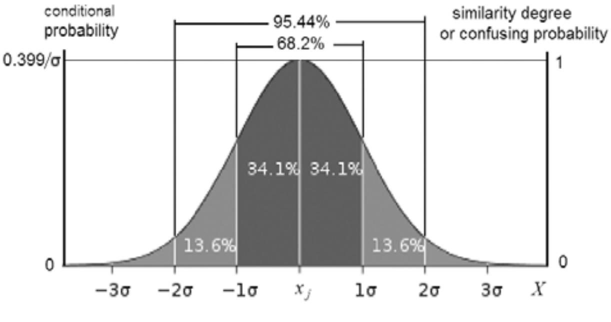

图 1. 反映 X 和 x_j 的混淆概率和相似度的 GPS 的误差函数

GPS 精度常见的表示法是均方根差（root mean square error， 简写为 RMS）表示法[1]。DRMS=10 米就表示 σ=10 米，目标有 68.2%的可能性在以读数 x_j 为中心，半径 10 米的圆圈之类。2DRMS=10 米就表示 2σ=10 米， 目标有 95.44%的可能在半径为 10 米的圆圈之类。常见的还有一种是圆概率误差（CEP）表示法，CEP 是 10 米， 就表示实际位置有 50%可能落在以 GPS 读数为中心半径=10 米的圆圈之内。

请注意， 一般情况下， $P(X)$不是等概率的， 条件概率 $P(X|y_j)$也不会呈正态分布。比如，汽车在公路上的先验概率比较大， 即使 GPS 定位小车在公路附近农田里， 那也并不意味着汽车在农田里概率最大。要表示 GPS 独立于信源的特性，它最好被写为

$$c(x_i, x_j) = \exp\left[-\frac{(x_i - x_j)^2}{2\sigma^2}\right] \tag{2}$$

我们应该把上面曲线理解为 x_i 和 x_j 混淆概率函数或 x_i 和 x_j 的相似度。其最大值等于 1。 而条件概率 $P(x_i \mid y_j)$ 对 x_i 求和等于 1， 对于连续分布其最大值是 0.399/σ。

混淆概率可以来自随机集合的统计[7][8]。假设做很多次分辨实验，每次确定一个 A 中子集，其中元素和 x_j 想混淆。 做 N 次实验就得到 N 个子集，它们的左右边界可能不同，如果有 N_i 个子集包含 x_i, 那么混淆概率就等于：

$$c(x_i, x_j) = \lim_{N\to\infty} N_i / N \tag{3}$$

我们定义模糊糊集合[9] A_j={所有和 x_j 相混淆的 x_i}，那么 x_i 在 A_j 上隶属度，记为 $Q(A_j|x_i)$, 就可以用混淆概率 $c(x_i, x_j)$来定义，即 $Q(A_j|x_i)= c(x_i, x_j)$。

$Q(A_j|x_i)$也可以被理解为 x_i 和 x_j 的相似度，或命题 y_j= “X is about x_j”在 x_i 发生时的逻辑真值或 y_j 的后验逻辑概率。而 y_j 的先验逻辑概率就是 $Q(A_j|x_i)$的平均值：

$$Q(A_j) = \sum_i P(x_i) Q(A_j \mid x_i) \tag{4}$$

[1] http://www.igage.com/mp/GPSAccuracy.htm

它也就是Zedah曾提出的模糊事件的概率[10]。已知 X 在 A_j 中， 求 X 的概率分布， 我们有集合 Bayesian 公式[4]或广义 Bayesian 公式[8]：

$$P(x_i \mid A_j) = P(x_i \mid x_i \in A_j) = P(x_i) Q(A_j \mid x_i) / Q(A_j) \tag{5}$$

后面讨论信息率失真函数和限误差信息率函数时， 我们将看到，上面关于 $Q(A_j)$和 $P(x_i|A_j)$的公式一再出现。当 A_j 是清晰集合时，$P(x_i|A_j)$的图解见图 3。

2.2 从统计信息到预言信息

要度量一个 GPS 读数提供的信息，用 Shannon 熵 $H(X)$ 公式或 Shannon 互信息 $I(X;Y)$公式都是不行的。因为它们是用来度量平均信息的[2]。为此，有人提出用 Shannon 互信息公式中的对数部分度量单个事件提供的信息[11]， 公式是：

$$I(x_i; y_j) = \log \frac{P(x_i \mid y_j)}{P(x_i)} \tag{6}$$

但是因为这个公式能导致负的信息（当 $P(x_i| y_j)<(P(x_i)$ 时)，所以很多人对此有疑问。

我接受这个公式，因为负信息可以理解为因谎言或错误预测而增加的编码长度。但是， 度量 GPS 信息还要考虑事实检验，因为根据常识，定位准确，信息就多， 定位不准， 信息就少，甚至是负的。我且称 GPS 信息是预言信息， 而 Shannon 信息是统计信息。

我们把经典信息公式（6）中的条件”y_j”改为“y_j 为真”，即用 $x_i \in A_j$ 取代 y_j，那么公式(4)就变为：

$$\begin{aligned} I^*(x_i; y_j) &= \log \frac{P(x_i \mid y_j\ is\ \ true)}{P(x_i)} \\ &= \log \frac{P(x_i \mid A_j)}{P(x_i)} = \log \frac{Q(A_j \mid x_i)}{Q(A_j)} \end{aligned} \tag{7}$$

注意：除了需要先验预测 $P(X)$和真值函 $Q(A_j|X)$, 我们还需要用以检验的事实 $X=x_i$.

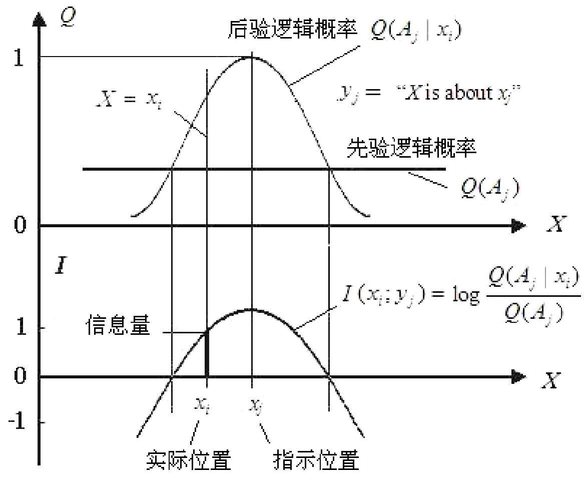

图 2. 预言信息公式图解

由图 3 可见，事实和预测完全一致时，即 x_i=x_j,，信息量最大；随着误差增大， 信息量渐渐变小； 误差大到一定程度信息就是负的。这是符合常理的。

$Q(A_j)$越小， 信息量越大。有两个因素决定 $Q(A_j)$大小， 一是预测精度，即 σ， 它越小，$Q(A_j)$就越小；二是 $Q(A_j|x_i)$覆盖区域的 $P(x_i)$大小，$P(x_i)$越小， $Q(A_j)$就越小。这样我们就有结论：预测精度越高，或者事件越是出乎意外，信息的绝对值就越大。同时犯错误的可能性也越大， 检验也就越严峻。如果经得起检验，即 $Q(A_j|x_i)$较大， 信息量就越大。这一公式非常符合 Popper 关于科学理论进步的信息准则(见 [12], 章节 10-II)。

2.3 理想 GPS 和普遍必然命题的信息

假设有理想 GPS， 它被定义为：

1） 混淆概率仅取 0，1 二值；

2） 声称实际目标 100%落在指定范围内。

一个可能的混淆概率函数如图 2 中的矩形所示。信息量公式变成

$$I^*(x_i;y_j)=\log\frac{Q(A_j|x_i)}{Q(A_j)}=\begin{cases}-\log Q(A_j), & x_i\in A_j\\ -\infty, & x_i\notin A_j\end{cases}\qquad(8)$$

其图解如图 3 所示。

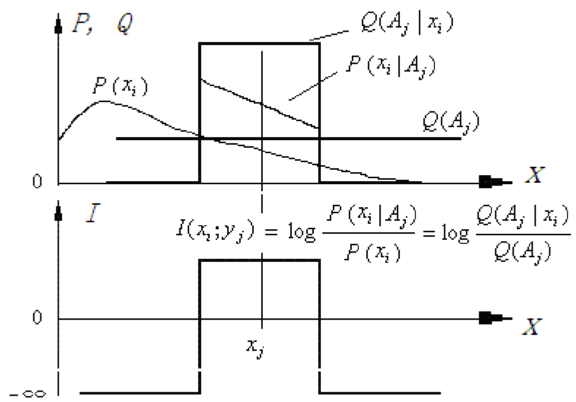

图 3. 理想 GPS 和普遍必然命题提供的信息

不难证明，只要有一个 x_i不在 A_j中，求得的平均信息都是-∞。信息负无穷大就意味着：我们按照语义为分布是 $P(X|A_j)$的信源编码时，只要有一个 x_i不在 A_j中， 则编码长度无论多大，都会失败。

根据 Popper 理论，对于普遍必然命题， 只要有一个反例，该命题就被证伪了[12]。上面公式正好反映 Popper 的这一思想。根据上面公式， 对于永真命题和永假命题，或者永远半真半假的命题，都有 $Q(A_j|x_i)$/ $Q(A_j)$=1， 所以信息量都是 0。这也和 Popper 理论一致。

日常语言中，由于接收信息者总是假定有意外事件的可能，或者说按模糊方式理解所有命题，负无穷大信息极少发生。

2.4 广义 Kullback 公式及其用于预测优化

如何求 y_j提供关于 X 的平均信息？开始我采用 Kullback 公式：

$$I_k(X;y_j)=\sum_i P(x_i|A_j)\log\frac{P(x_i|A_j)}{P(x_i)}\qquad(9)$$

这里假设 $P(X|Y)=P(X|Y$是真的)， 即对于所有 x_i, $P(x_i|y_j)=P(x_i|A_j)$. 但是实际上， 两者未必相同。后来我发现， 保留对数左边的条件概率，采用公式

$$I^*(X;y_j)=\sum_i P(x_i|y_j)\log\frac{P(x_i|A_j)}{P(x_i)}\qquad(10)$$

反而使得公式更有解释力。 这是非常重要的一步，因为这样就能反映信息来自事实检验。这就是广义 Kullback 公式。这里，$P(x_i|y_j)$就是证据， 而 $P(x_i|A_j)$是理论预测，$P(x_i)$是先验知识。Theil 曾经提出的信息差公式与此类似[13]， 但是上面公式含义更丰富。

我们可以把-log$P(x_i)$理解为我们原先为 $P(X)$按最优式编码时 x_i的码长，-log $P(x_i|A_j)$理解为 y_j出现后我们为 $P(x_i|A_j)$按最优方式编码时 x_i的码长。这样，$I(X;y_j)$就是因预测 y_j而节省的平均码长。

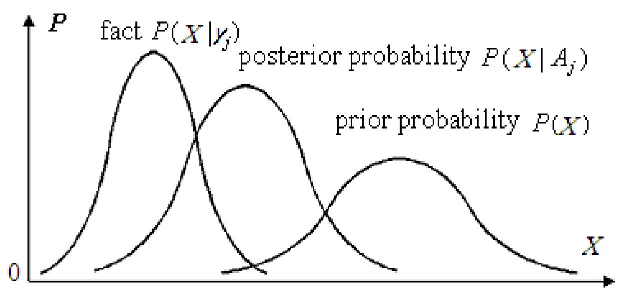

图 4. 图解广义 Kullback 公式

图 4 表明，理论预测 $P(X|A_j)$越是接近事实 $P(X|y_j)$， 信息量越大， 最大值就是 Kullback 信息；信源或先验估计 $P(X)$越是与事实 $P(X|y_j)$不同(表示预测的事情越是出乎意外)，信息量越大。

现在假设 X 是经济指标，我们用两个参数 x_j 和 σ_j 预测经济指标， 即逻辑条件概率是

$$Q(A_j|x_i)=\exp\left[-\frac{(x_i-x_j)^2}{2\sigma_j^{\ 2}}\right]\qquad(11)$$

上面广义 Kullback 公式就可以用于预测的评价和优化。

设 z 是某种客观因素，$\langle y_j,\sigma_j\rangle$表示一种预测， 被预测的经济指标的条件概率是 $P(X|z)$。 选择哪个预测$\langle y_j,\sigma_j\rangle$最好？ 改变 $\langle y_j,\sigma_j\rangle$， 使广义 Kullback 信息

$$I^*(X;y_j)=\sum_i P(x_i|z)\log\frac{Q(A_j|x_i)}{Q(A_j)}\qquad(12)$$

达最大的$\langle y_j,\sigma_j\rangle$就最好。对于 GPS 精度 σ 和读数 y_j的选择， 上面公式同样适用。

如果预测经常不准， 或者发信人有意说谎， 收信人可以总结经验，修改语义， 比如令

$$Q(A_j|x_i)=P(y_j|x_i)/\max_k P(y_k|x_i)\ \text{for all}\ x_i\qquad(13)$$

这样将能使实际收到的信息接近统计信息。

2.5 广义互信息公式

对 $I(X;y_j)$求平均，我们得到广义互信息公式：

$$I^*(X;Y)=\sum_j\sum_i P(y_j)P(x_i\mid y_j)\log\frac{P(x_i\mid A_j)}{P(x_i)} \quad (14)$$
$$=\sum_i P(y_j)P(x_i\mid y_j)\log\frac{Q(A_j\mid x_i)}{Q(A_j)}$$

当预测和事实一致时，$I^*(X;Y)$可以写成

$$I^*(X;Y)=\sum_j\sum_i P(y_j)P(x_i\mid A_j)\log\frac{Q(A_j\mid x_i)}{Q(A_j)} \quad (15)$$

下面讨论中，我们将一再看到类似公式.

从上面公式可见，Shannon 广义互信息是广义互信息在预测总是和事实一致时的特例。它也反映节省的平均码长。Shannon 互信息可以理解为通信成本，而广义互信息可以理解为通信的效用，其上限是 Shannon 互信息。

3 限误差信息率及其和信息率失真及复杂性失真之间的等价关系

3.1 从信息率失真到复杂性失真

在为声音、视频、经济、地理位置…数据编码时，适当的有失真编码将大大提高通信效率。为此 Shannon 提出 Rate distortion 理论[2]，并且得到 T. Berger[14] 等人（[15]，第十章）的大力发展。

假设我们为数字 X 编码，设离散随机变量 $X\in A=\{x_1, x_2, \ldots\}$和 $Y\in B=\{y_1, y_2,\ldots\}$分别表示源码和目的码，信源是前后无关的。假设 y_j 和 x_i 之间的失真是 $d_{ij}=d(x_i, y_j)$, 那么它的平均值就是

$$E(d_{ij})=\sum_j\sum_i P(x_i,y_j)d_{ij} \quad (16)$$

给定信源$P(X)$和$E(d_{ij})$的上限D, 改变$P(Y|X)$, 求 Shannon 互信息的最小值:

$$R(D)=\min_{P(Y|X):E(d_{ij})\le D} I(X;Y) \quad (17)$$

$R(D)$就是信息率失真函数。Shannon 证明了，它反映为每个数字编码的最短平均码长（通过分组编码实现）或最小信息速率（忽略两者之间的微小差别）。

但是，数据压缩实践中，我们需要对每对(x_i, y_j)之间的误差给出限制。为此，Kolmogorov 提出基于其复杂性理论的结构函数[5]，最近又有人提出复杂性失真[6]。复杂性理论把一个字符串的最短编码长度叫做这个字符串的复杂性。有失真编码时，如果

$$d(x_i,y_j)=(x_i-y_j)^2\le D_c, \quad \text{for all } i,\ j \quad (18)$$

成立，那么对于每个 x_i，集合 B 上存在一个以 y_i 为中心的失真球 B_i， 用球中任何一个 y_j 表示 x_i 都可以。同时对于任何 y_j，A 上存在一个以 x_j 为中心失真球 A_j，y_j 可以是 A_j 中任何一个 x_i 的代表. 球的半径都是 $D_c^{0.5}$。给定信源和失真球限制，可以求出最小平均码长或 Shannon 互信息（忽略微小误差），设为 C，它是 D_c 的函数， 即 $C=C(D_c)$. 它就是复杂性失真函数[6]。采用同样的编码，假设对于每个 y_j， 失真球 A_j 中元素一样多， 为 $S=|A_j|$, $h_x(D_c)=\log S$ 就是结构函数。

然而， 我们也需要不同尺寸的容差集合或容差球，用以限制编码误差。比如，随着颜色的明度增加，人眼对色差的敏感性减弱，这意味着，对于较暗的颜色，我们需要较小的容差，对于较亮的颜色，我们需要较大的容差。

3.2 定义限误差信息率并证明复杂性失真是其特例

设 AXB 上存在相似关系，x_i 和 y_j 之间的相似度是 $c_{ij}=c(x_i, y_j)= c(x_i, x_j)\in[0,1]$， 比如，如(2) 所示。对于每个 x_i，B 上存在一个容差集合 B_i，y_j 在 B_i 上的隶属度是 c_{ij}. 对于每个 y_j, A 上存在一个容差集合 A_j，x_i 在 A_j 上的隶属度也是 c_{ij}. 这也就是说，这些容差集合可以是模糊的，并且有

$$Q(B_i\mid y_j)=Q(A_j\mid x_i)=c_{ij} \quad (19)$$

如果对于每个 B_i,

$$P(y_j\mid x_i)\le P(y_j\mid B_i),\ \text{for } y_j\in B \ \text{ as } Q(A_j|x_i)\ne 1 \quad (20)$$

我们称一组集合 $T=\{B_1, B_2,\ldots\}$是对 $x_i, x_2, \ldots$的编码的容差。如果 $B_1, B_2,\ldots$是清晰集合， 即 $c_{ij}\in\{0,1\}$，上面限制变成

$$\sum_j P(y_j\mid x_i)=0,\ \text{for all } y_j\notin B_i \quad (21)$$

现在，在允许我们选择和 x_i 相似的 y_j 代表 x_i 时，给定 $P(X)$和 T, 改变 $P(Y|X)$, 求最小信息速率 $R(T)$,

$$R(T)=\min_{P(Y|X):P(Y|X)\le P(Y|T)\, as Q(A_j|x_i)\ne 1} I(X;Y) \quad (22)$$

就是限误差信息率。很显然，当 B_1, B_2,⋯是分别以 y_1, y_2 为中心的大小一样的球或区域，$R(T)$就变成 $C(D_c)$. 所以 $C(D_c)$是 $R(T)$的特例。

下面假设 $B_1, B_2, \ldots$是清晰集合时，看 $R(T)$和 $R(D)$之间的等价关系。

现在考虑函数 $R(D)$。 我们把 x_i 和 y_j 之间的失真定义为：

$$d_{ij}=d(x_i,y_j)=-\log c_{ij}=\begin{cases}0, & c_{ij}=1,\\ \infty, & c_{ij}=0.\end{cases}\ \text{for all } i,j \quad (23)$$

根据公式(21)和(23)， 限制$E(d_{ij})\le D=0$和限制T等价，即 $R(D=0)=R(T)$。所以 $C(D_c)$也是 $R(D)$ 的特例。

3.3 反映信息率失真和复杂性失真之间等价关系的广义熵

下面证明， 当 T 是一组清晰集合时，$R(D=0)$ 等于一个广义熵。

我们知道， Rate-distortion 函数的参数表示（[16]，316 页）:

$$\begin{cases} D=\sum_i P(x_i)P(y_j)\exp(sd_{ij})\lambda_i d_{ij} \\ R(s)=sD(s)+\sum_i P(x_i)\log\lambda_i \end{cases} \quad (24)$$

其中

$$\lambda_i=1/\sum_j P(y_j)\exp(sd_{ij}) \quad (25)$$

令 $s=-1/(2\sigma^2)$, 我们可以看出 GPS 信息和信息率失真之间的清晰联系。这样就可认为 $\exp(sd_{ij})= Q(B_i|y_j)$, $\lambda_i=1/Q(B_i)$, 于是得到和(15)相似的公式：

$$R(s)=\sum_i\sum_j P(x_i)P(y_j\mid B_i)\log\frac{Q(B_i\mid y_j)}{Q(B_i)} \tag{26}$$

并且

$$D(s)=\sum_i P(x_i)P(y_j\mid B_i)d_{ij} \tag{27}$$

当 T 是一组集合清晰时，exp(-∞)=0, exp(0)=1, D=0 时，R 和 s 无关. 于是有

$$R(s)=H^*(X)=-\sum_i P(x_i)\log Q(B_i) \tag{28}$$

虽然 $Q(B_i)$是 $P(Y)$的函数, 但是并不是任何一个 $H^*(X)$都等于一个 $R(D)$。 改变 $P(Y)$使得 $H^*(X)$达最小， 设为 $H_{P(Y)}{}^*(X)$, 它才等于 $R(D$=0):

$$R(D=0)=H_{P(Y)}{}^*(X)=\min_{P(Y)}[-\sum_i P(x_i)\log Q(B_i)] \tag{29}$$

所以 $R(T)=R(D=0)=H_{P(Y)}{}^*(X)$。

仅当所有 B_i 大小一样呈球形时，才有 $C(D_c)=R(T)=R(D=0)$.

给定 T 时 X 的条件熵是

$$H(X\mid T)=\sum_j\sum_i P(x_i)P(x_i\mid A_j)\log P(x_i\mid A_j) \tag{30}$$

如果在每个失真球 A_j 中元素相等，为 $S=|A_j|$, 并且等概率发生，$H(X|T)$就变成结构函数 $h_x(D_c)$。

现在我们看到，一般情况下，结论 $C(D_c)=R(D_c)$[6] 是不对的。肯定失真−信息率和期望的结构函数之间有一般等价关系[5]，也是不对的。因为在复杂性理论中，那些容差球是清晰集合，而在信息率失真理论中，那些容差球是模糊集合，后者较前者限制更松。下面我们用编码例子说明它们的差别。

3.4 用一个编译码例子说明 $R(D_c)<C(D_c)$

这个例子也将显示如何改变 $P(Y|X)$从而减少 $I(X;Y)$的思路。

Example: A={1,2,3,4}, B=(1,2,3,4), 允许误差 $|X-Y|\le 1$。给定 $P(X)$, 通过适当的编码得到 R.

实际的数据压缩过程是：$P(X)$→信源编码→存储和传输→解码→$P(Y)$, 我们这里不考虑中间环节（中间可能要采用分组码），只考虑用怎样的编码即 $P(Y|X)$ 可以减少 $I(X;Y)$得到 R.

公式（28）提示我们，对于每个 x_i, 逻辑概率 $Q(B_i)$ 越大越好；不同的 B_i 之间相互越重叠越好。这样条件熵 $H(X|Y)$就较大， 信息 $I(X;Y)= H(X)-H(X|Y)$ 就较小。

如果简单用 y_2 表示 x_1 和 x_2；用 y_3 表示 x_3 和 x_4, 平均信息量是 1 比特。 现在我们参看公式（25）和（28），采用优化方法给信源编码。让 X=1 时，Y=2; X=4 时，Y=3；X=2 或 3 时，Y=2 或 3，随机产生。这样，就有表 1 数据。

表 1 为 $C(D_c=1)$寻找 $P(Y|X)$

$P(x_i)$	x_i	y_j	$P(y_j)$	$Q(B_i)$	$-P(x_i)\log Q(B_i)$
0.25	1	1	0	$P(y_2)$=0.5	0.25
0.25	2	2	0.5	$P(y_2)+P(y_3)$=1	0
0.25	3	3	0.5	$P(y_2)+P(y_3)$=1	0
0.25	4	4	0	$P(y_4)$=0.5	0.25

注意，$Q(B_2)=Q(B_3)$=1, 所以 X=2 或 3 时，传递信息是 0. 所以

$R(T)=C(D_c)$ = - 0.25log0.5-0.25log0.5=0.5 比特。

但是，这时候 $E(d_{ij})$=0.75<1. 下面编码可以证明 $R(D_c)$<C(D_c).

现在令 $d_{ij}=(y_j-x_i)^2$，求 $E(d_{ij})\le D=D_c$=1 时的 R。令 s=−0.45, y_2= y_3=0.5, $Q(A_i|y_j)=\exp[s(y_j-x_i)^2]$, $P(y_j|x_i)=P(y_j|A_i)$, 于是有表 2 数据。

表 2 为 $R(D=1)$寻找 $P(Y|X)$

P(x_i)	x_i	$P(y_2\|x_i)$	$P(y_3\|x_i)$	$P(y_j)$	$Q(B_i)$	$-P(x_i)$ $\log Q(B_i)$
0.25	1	0.8	0.2	0	0.395	0.335
0.25	2	0.903	0.097	0.5	0.816	0.073
0.25	3	0.097	0.903	0.5	0.816	0.073
0.25	4	0.2	0.8	0	0.395	0.335

那么，根据(24)，有 D=0.994,

$R(D)=sD+H^*(X)$=-0.447+0.816=0.369 比特。

可见，一般情况下 $R(D_c)<C(D_c)$。

3.5 限误差信息率和信息率失真之间的等价关系

我们假设 T={ B_1, B_2, …}是一组模糊集合，Shannon 互信息是 $I(X;Y)= H(Y) - H(Y|X)$。当容差限制是 T 或不等式(20)时，下式成立时：

$$P(y_j\mid x_i)=P(y_j\mid B_i),\ \text{for all } i,j \tag{31}$$

$P(Y|X)$最分散，所以 $H(Y|X)$最大。这个等式对于 R 是必要的， 但不是充分的。改变 $P(Y)$使 $I(X;Y)$达到最小值，我们就得到限误差信息率函数:

$$R(T)=\min_{P(Y)}\sum_i\sum_j P(x_i)P(y_i\mid B_i)\log\frac{Q(B_i\mid y_j)}{Q(B_i)} \tag{32}$$

假设 $Q(B_i|y_i)=\exp(sd_{ij})$ for all i, j, 参看 3.3 节，我们得到 $R(T)$== $R(D)$. 这意味着 $R(D)$函数是 $R(T)$函数在 $Q(B_i|y_j)=\exp(sd_{ij})$ for all i,j 时的特例。其中 s 就反映预测的精度，s 越大，所需要的信息速率就越大。

显然，广义信息测度和 $R(D)$函数之间存在深刻联系，它们都和误差及语义密切相关。

4 总结

本文以 GPS 为例，推广经典信息公式到广义信息公式， 讨论限误差信息率和信息率失真怎样和广义互信息公式相联系，证明了信息率失真 $R(D)$是限误差信息率 $R(T)$的特例, 而复杂性失真 $C(D_c)$是信息率失真 $R(D)$的特例。文献[1][4]中我还讨论了热力学系统中最大熵原理和限误差信息率之间的关系，提出用广义信息下限 G 取代 D 得到 $R(G)$函数，G/R 就表示通信效率, 包括发信者谎言导致敌人信息损失的效率。关于这些，我将在别处进一步讨论。

参考文献

[1] 鲁晨光，广义信息论，中国科技大学出版社，1993.
[2] C. E. Shannon, “A mathematical theory of communication”， Bell System Technical Journal, Vol.27, pp.379-429 and 623-656, 1948.
[3] 鲁晨光，广义互熵和广义互信息的编码意义，通信学报， Vol. 15, No.6, pp.37-44, 1994..
[4] C. Lu (鲁晨光), “A generalization of Shannon's information theory”, Int. J. of General Systems, Vol. 28, No. 6, pp.453-490, 1999.
[5] P. D. Grunwald and P. B. Vitany, “Kolmogorov Complexity and Information Theory With an Interpretation in Terms of Questions and Answers”, Journal of Logic, Language and Information, pp. 497-529, Dec. 2003.
[6] D. M. Sow, “Complexity Distortion Theory”, IEEE Transactions on Information Theory, Vol. 49, No.. 3, pp. 604-609, March 2003.
[7] P. Z. Wang (汪培庄), “Random sets in Fuzzy Set theory”, In: System & Control Encyclopedia, edited by M. G. Singh, Pergamon Press, pp. 3945-3947, 1987.
[8] J. B. Tenenbaum and T. L. Griffiths, “Generalization, similarity, and Bayesian inference”, Behavioral and Brain Science, Vol. 24, No.4, 629-640, 2001.
[9] L. A. Zadeh, “Fuzzy sets”, Infor. Contr., Vol. 8, pp. 338-353, 1965.
[10] L. A. Zadeh, “Probability measures of fuzzy events”，Journal of mathematical Analyses and Applications. Vol.23, pp. 421-427, 1968..
[11] S. Kullback, Information and Statistics，John Wiley & Sons Inc.，New York，1959.
[12] K. R. Popper, Conjectures and Refutations: The Growth of Scientific Knowledge, Harper & Row, Publishers, New York and Evanston, 1968.
[13] H. Theil, Economics and Information Theory, North-Holland, Amsterdam, 1967.
[14] T. Berger, Rate Distortion Theory: A Mathematical Basis for Data Compression, Englewood Cliffs, NJ: Prentice-Hall, 1971.
[15] T. M. Cover and J. A. Thomas, Elements of Information Theory, Second Edition, John Wiley & Sons, 2006.
[16] 周炯盘，信息理论基础，人民邮电出版社,1983.