中国科学 E 辑: 信息科学 2008 年 第 38 卷 第 12 期: 2048 ~ 2065
www.scichina.com info.scichina.com
《中国科学》杂志社
SCIENCE IN CHINA PRESS

信息系统中的信息粒与熵理论

梁吉业[①②*], 钱宇华[①②]

① 山西大学计算智能与中文信息处理省部共建教育部重点实验室, 太原 030006;
② 山西大学计算机与信息技术学院, 太原 030006
* E-mail: ljy@sxu.edu.cn

收稿日期: 2007-08-06; 接受日期: 2007-12-26
国家高技术研究发展计划项目(批准号: 2007AA012165)、国家自然科学基金(批准号: 60773133)和国家重点基础研究规划(批准号: 2007CB311002)资助项目

摘要 信息粒度与熵理论是两种有效进行信息系统中不确定性研究的重要工具, 已有许多成功的应用范例. 文中研究了不同二元关系下信息系统中信息粒的刻画和表示, 给出了信息系统中信息粒度的公理化定义, 证实了一些已有的信息粒度度量都是其特殊形式. 发展了信息系统中的熵理论, 证明了这些熵度量的粒化单调性. 同时, 在信息系统意义下, 建立了信息粒度与熵之间的互补关系. 这些研究统一了完备信息系统与非完备信息系统中不确定性度量的相关结果.

关键词
信息系统
信息粒
信息粒度
熵
粗糙集

粒度计算, 是 Zadeh 于 1996 年首先提出的[1]. 他认为, 人类认知主要基于 3 个主要概念, 即粒度(包括将全体分解为部分)、组织(包括从部分集成整体)和因果(包括因果的关联). 而粒度计算是一把大伞, 它覆盖了所有有关粒度的理论、方法论、技术和工具的研究[2~4]. 作为复杂问题求解的有力工具, 粒度计算在粗糙集理论、概念格、知识工程、数据挖掘、人工智能、机器学习等领域有着潜在的应用[5~12], 已成为信息科学的研究热点之一.

粒度计算理论的研究主要有 3 个比较成熟的模型: 1) 词计算理论[1,2]、2) 粗糙集理论[13]、3) 商空间理论[14~16].

信息粒度主要研究信息系统中信息或知识的不确定性问题. Wierman 提出了粒度度量的概念来度量信息的不确定性, 在其提出的公理化约束下该定义和 Shannon 熵具有同样的形式[17]. 梁吉业等研究了完备信息系统和非完备信息系统中的信息粒度, 并将其成功应用于属性重要性度量、特征选择与决策规则获取等方面[18~20]. 钱宇华等则引入了具有直观知识含量特征的组合粒度来度量信息系统中的粒度大小[21].

熵的概念起源于经典热力学, 用来度量系统的无序程度, Shannon 借用熵的称谓, 定义了

量化一个离散型随机变量的随机性大小的量度[22], 称之为信息熵或随机熵. 今天, 熵被广泛用于不确定性度量. Düntsch 运用 Shannon 熵对粗糙集理论中的规则进行度量[23]. 文献[24,25]应用 Shannon 熵的变形研究了粗糙集和粗糙关系数据库的不确定性信息度量. 1968 年 Zadeh 首先提出了模糊熵的概念, 用来量化模糊集合的模糊性. 1972 年 de Luca 和 Termini[26] 引入了一个模糊集熵的公理化构造方法, 并给出了一个模糊集模糊性的距离度量、相似性度量以及这些度量之间的关系. 文献[27]给出了一种新的模糊熵并有效地应用到基于划分的模糊粗糙集的模糊性度量. 梁吉业等提出了一种信息增益具有补特征的信息熵, 并给出其条件熵和互信息, 指出这个熵也是一种模糊熵, 并应用于度量粗糙集和粗糙分类的模糊性[28]. 钱宇华等则在非完备信息系统中引入组合熵的概念, 其信息增益函数具有可能知识含量的特性, 并用于度量非完备信息系统的不确定性[21].

本文统一了完备信息系统与非完备信息系统中的不确定性度量方法, 通过引入一个偏序关系 $\preceq'$, 给出了信息系统中信息粒度的公理化定义, 指出了一些已有的信息粒度度量都是其特殊形式. 进一步发展了非完备信息系统中的熵理论, 证实了信息系统中熵的粒化单调性, 并建立了信息粒度与熵之间的互补关系. 这些结果为信息系统中不确定性研究提供了有效工具.

1 信息系统中的信息粒

1.1 完备信息系统

形式上, $S=(U,A)$ 是一个信息系统, U 表示对象的非空有限集合, A 表示属性的非空有限集合, 任意属性 $a\in A$, a: $U\to V_a$ 是一个映射, 其中 V_a 称为 a 的值集.

对于给定信息系统 $S=(U,A)$, 若任意属性 $a\in A$, V_a 不含空值, 则称 S 是一个完备信息系统[19,20,29].

设 $P\subseteq A$, 定义等价关系:

$$\mathrm{IND}(P)=\{(u,v)\in U\times U \mid \forall a\in P, a(u)=a(v)\}.$$

显然, $\mathrm{IND}(P)=\bigcap_{a\in P}\mathrm{IND}(\{a\})$.

这里, $U/\mathrm{IND}(P)$ 构成 U 的一个划分, 称其为 U 上的一个信息, 其中每个等价类称为一个信息粒. 信息粒度则是对关系(或属性集)P 下信息粒(等价类)大小的平均度量.

特别地, 如果 $U/\mathrm{IND}(P)=\omega=\{X\mid X=\{u\}, u\in U\}$, 称其为恒等关系; 如果 $U/\mathrm{IND}(P)=\delta=\{X\mid X=\{U\}\}$, 称其为全域关系.

1.2 非完备信息系统

在一些情形下, 对一个对象而言, 一些属性值可能是缺损的. 例如, 在医学信息系统中, 可能存在这样一组病人, 它们不能执行所有要求的检查. 一种解决方法是这些缺损值可以通过对应属性的所有可能值的集合来表示. 为了表示这种情形, 一个区分值(即空值)通常安排给这些属性.

在信息系统 $S=(U,A)$ 中，如果至少存在一个属性 $a\in A$, V_a 包含空值，则称 S 是一个非完备信息系统，用*表示空值[19,20,29].

设 $P\subseteq A$, 定义相容关系[29]:

$$\mathrm{SIM}(P)=\{(u,v)\in U\times U\mid \forall a\in P, a(u)=a(v)\ \text{ or }\ a(u)=*\ \text{ or }\ a(v)=*\}.$$

显然，$\mathrm{SIM}(P)=\bigcap_{a\in P}\mathrm{SIM}(\{a\})$.

令 $S_P(u)$ 表示对象集 $\{v\in U\mid (u,v)\in \mathrm{SIM}(P)\}$. $S_P(u)$ 是与 u 不可区分的对象的最大集合(相对 P 而言).

令 $\tilde{P}$ 表示分类，即一族集合 $\{S_P(u)\mid u\in U\}$. $\tilde{P}$ 构成 U 的一个覆盖，即对于每一个 $u\in U$ 有 $S_P(u)\neq\varnothing$, 且 $\bigcup_{u\in U}S_P(u)=U$. 这里，称 $\tilde{P}$ 为 U 上的一个信息，其中每个相容类称为一个信息粒. 信息粒度则是对属性集 P 下信息粒(相容类)大小的平均度量.

特别地，如果 $\tilde{P}=\omega=\{S_P(u)\mid S_P(u)=\{u\}, u\in U\}$, 称其为恒等关系；如果 $\tilde{P}=\delta=\{S_P(u)\mid S_P(u)=\{U\}, u\in U\}$, 称其为全域关系.

例 1 考虑表 1 中几个小汽车的描述[20,29].

表 1 一个关于小汽车的非完备信息系统

Car	Price	Mileage	Size	Max-Speed
u_1	High	Low	Full	Low
u_2	Low	*	Full	Low
u_3	*	*	Compact	Low
u_4	High	*	Full	High
u_5	*	*	Full	High
u_6	Low	High	Full	*

这是一个非完备信息系统,其中 $U=\{u_1,u_2,u_3,u_4,u_5,u_6\}$, $A=\{a_1,a_2,a_3,a_4\}$, 这里 a_1, a_2, a_3, a_4 表示 Price, Mileage, Size, Max-Speed. 通过计算，可得

$$\tilde{A}=\{S_A(u_1),S_A(u_2),S_A(u_3),S_A(u_4),S_A(u_5),S_A(u_6)\},$$

其中，$S_A(u_1)=\{u_1\}$, $S_A(u_2)=\{u_2,u_6\}$, $S_A(u_3)=\{u_3\}$, $S_A(u_4)=\{u_4,u_5\}$, $S_A(u_5)=\{u_4,u_5,u_6\}$, $S_A(u_6)=\{u_2,u_5,u_6\}$.

令 $S=(U,A)$ 是一个完备信息系统，$U/\mathrm{IND}(A)=\{X_1,X_2,\ldots,X_m\}$. 如果我们把它看作一个属性取值不包含空值的非完备信息系统，则 $\tilde{A}=\{S_A(u_1), S_A(u_2), \ldots, S_A(u_{|U|})\}$[20,21]. 不妨令集合 $X_i=\{u_{i1},u_{i2},\ldots,u_{is_i}\}$, 其中 $|X_i|=s_i$, $\sum_{i=1}^{m}s_i=|U|$, 则存在以下关系

$$X_i=S_A(u_{i1})=S_A(u_{i2})=\cdots=S_A(u_{is_i}),\quad |X_i|^2=\sum_{k=1}^{s_i}|S_A(u_{ik})|.$$

因此，如果不需要区分完备信息系统和非完备信息系统，可以把信息系统中的信息统一表示为向量 $K(A)=(S_A(u_1), S_A(u_2), \ldots, S_A(u_{|U|}))$[20,21]. 特别地，恒等关系可以表示为 $|U|$ 维向量

$\tilde{P}=\omega=(\{u_1\},\{u_2\},\ldots,\{u_{|U|}\})$; 同理, 全域关系也可以表示为 $|U|$ 维向量 $\tilde{P}=\delta=(\{U\},\{U\},\ldots,\{U\})$.

令 $S=(U,A)$ 是一个完备信息系统, $P,Q\subseteq A$, $U/\mathrm{IND}(P)=\{X_1,X_2,\ldots,X_m\}$, $U/\mathrm{IND}(Q)=\{Y_1,Y_2,\ldots,Y_n\}$, $K(P)=(S_P(u_1),S_P(u_2),\ldots,S_P(u_{|U|}))$, $K(Q)=(S_Q(u_1),S_Q(u_2),\ldots,S_Q(u_{|U|}))$, 其中 $X_i=\{u_{i1},u_{i2},\ldots,u_{is_i}\}$, $Y_j=\{u_{j1},u_{j2},\ldots,u_{jt_j}\}$, 则 $K(P)$ 与 $U/\mathrm{IND}(P)$ 中的元素之间存在以下关系

$$X_i=S_P(u_{i1})=S_P(u_{i2})=\cdots=S_P(u_{is_i}),\quad |X_i|^2=\sum_{k=1}^{s_i}|S_P(u_{ik})|.$$

同理, $K(Q)$ 与 $U/\mathrm{IND}(Q)$ 中的元素之间存在以下关系

$$Y_j=S_Q(u_{j1})=S_Q(u_{j2})=\cdots=S_Q(u_{jt_j}),\quad |Y_j|^2=\sum_{k=1}^{t_j}|S_Q(u_{jk})|.$$

在本文以下的许多定理证明中将多次使用这种表达方式.

2 信息粒度的公理化方法

2.1 偏序关系“$\preceq'$”

设 $S=(U,A)$ 是一个信息系统, $P,Q\subseteq A$, $K(P)=(S_P(u_1),S_P(u_2),\ldots,S_P(u_{|U|}))$, $K(Q)=(S_Q(u_1), S_Q(u_2),\ldots,S_Q(u_{|U|}))$. 我们定义一个二元关系 $\preceq'$:

$K(P)\preceq' K(Q)\Leftrightarrow$ 如果任意的 $i\in\{1,2,\ldots,|U|\}$, 都有 $S_P(u_i)\subseteq S_Q(u_i)$, 其中 $S_P(u_i)\in K(P)$, $S_Q(u_i)\in K(Q)$. 这里简记为 $P\preceq' Q$.

进一步, $K(P)=K(Q)\Leftrightarrow$ 任意的 $i\in\{1,2,\ldots,|U|\}$, 都有 $S_P(u_i)=S_Q(u_i)$, 简记为 $P\approx Q$; $K(P)\prec' K(Q)\Leftrightarrow K(P)\preceq' K(Q)$ 且 $K(P)\neq K(Q)$, 简记为 $P\prec' Q$.

令 $S=(U,A)$ 是一个信息系统, 记 $K=\{K(P)\mid P\subseteq A\}$.

定理 1　$(K,\preceq')$ 是一个偏序集.

证明　令 $P,Q,R\subseteq A$, $K(P)=(S_P(u_1),S_P(u_2),\ldots,S_P(u_{|U|}))$, $K(Q)=(S_Q(u_1),S_Q(u_2),\ldots,S_Q(u_{|u|}))$, $K(R)=(S_R(u_1),S_R(u_2),\ldots,S_R(u_{|U|}))$.

1) 任意的 $u\in U$, 都有 $|S_P(u)|=|S_P(u)|$ 成立, 因此 $P\preceq' P$.

2) 假设 $P\preceq' Q$ 和 $Q\preceq' P$. 从上面的定义, 我们能够得到:

$P\preceq' Q\Leftrightarrow$ 任意的 $i\in\{1,2,\ldots,|U|\}$, 都使得 $S_P(u_i)\subseteq S_Q(u_i)$, 其中 $S_P(u_i)\in K(P)$, $S_Q(u_i)\in K(Q)$;

$Q \preceq' P \Leftrightarrow$ 任意的 $i \in \{1,2,\ldots,|U|\}$, 都使得 $S_Q(u_i) \subseteq S_P(u_i)$, 其中, $S_Q(u_i) \in K(Q)$, $S_P(u_i) \in K(P)$.

因此, 我们有 $S_P(u_i) \subseteq S_Q(u_i) \subseteq S_P(u_i)$, 即 $S_P(u_i) = S_Q(u_i)$.

所以, 任意的 u_i 都有 $S_P(u_i) = S_Q(u_i)$, 即 $P \approx Q$.

3) 假设 $P \preceq' Q$ 和 $Q \preceq' R$. 从上面的定义, 我们能够得到:

$P \preceq' Q \Leftrightarrow$ 任意的 $i \in \{1,2,\ldots,|U|\}$, 都使得 $S_P(u_i) \subseteq S_Q(u_i)$, 其中 $S_P(u_i) \in K(P)$, $S_Q(u_i) \in K(Q)$;

$Q \preceq' R \Leftrightarrow$ 任意的 $i \in \{1,2,\ldots,|U|\}$, 都使得 $S_Q(u_i) \subseteq S_R(u_i)$, 其中 $S_Q(u_i) \in K(Q)$, $S_R(u_i) \in K(R)$.

因此, 任意的 $i \in \{1,2,\ldots,|U|\}$, 我们有 $S_P(u_i) \subseteq S_Q(u_i) \subseteq S_R(u_i)$, 即 $S_P(u_i) \subseteq S_R(u_i)$, 所以 $P \preceq' R$.

由上可得, $(K, \preceq')$ 是一个偏序集. 证毕.

2.2 信息粒度的公理化定义及其性质

定义 1 设 $S = (U, A)$ 是一个信息系统, 若 $\forall P \subseteq A$ 有实数 $G(P)$ 对应, 且满足:

1) $G(P) \geqslant 0$ (非负性);

2) $\forall P, Q \subseteq A$, 若 $P \approx Q$, 有 $G(P) = G(Q)$ (不变性);

3) $\forall P, Q \subseteq A$, 若 $P \prec' Q$ 时, 有 $G(P) < G(Q)$ (单调性);

则称 G 为信息系统 $S = (U, A)$ 上的信息粒度.

信息粒度其实是对信息细化的不同层次的平均度量.

定理 2(极值性) 设 $S = (U, A)$ 是一个信息系统, $P \subseteq A$, 则当 $K(P) = \omega$ 时, $G(P)$ 取得最小值; 当 $K(P) = \delta$ 时, $G(P)$ 取得最大值.

证明 令 $K(\omega) = (\{u_1\}, \{u_2\}, \ldots, \{u_{|U|}\})$, $K(\delta) = (\{U\}, \{U\}, \ldots, \{U\})$. 因此, 任意的 $R \subseteq A$, $K(R) = (S_R(u_1), S_R(u_2), \ldots, S_R(u_{|U|}))$, 我们有 $\{u_i\} \subseteq S_R(u_i)$ $(u_i \in U)$, 即 $\omega \preceq' R$. 同理, 我们有 $S_R(u_i) \subseteq U$ $(u_i \in U)$, 即 $R \preceq' \delta$.

从定义 1 中的 2)与 3)可知, $G(\omega) \leqslant G(R) \leqslant G(\delta)$. 证毕.

下面一些定义是信息粒度的不同特殊情形.

定义 2[18] 设 $S = (U, A)$ 是一个完备信息系统, $U / \mathrm{IND}(A) = \{X_1, X_2, \ldots, X_m\}$, 则 A 的信息粒度定义为

$$\mathrm{GK}(A) = \frac{1}{|U|^2} \sum_{i=1}^{m} |X_i|^2, \tag{1}$$

其中 $\sum_{i=1}^{m}|X_i|^2$ 是由 $\bigcup_{i=1}^{m}(X_i\times X_i)$ 决定的等价关系中元素的个数.

如果 $U/\mathrm{IND}(A)=\omega$, 则 A 的信息粒度达到最小值 $|U|/|U|^2=1/|U|$.

如果 $U/\mathrm{IND}(A)=\delta$, 则 A 的信息粒度达到最大值 $|U|^2/|U|^2=1$.

定理 3　定义 2 中的 GK 是定义 1 意义下的一个信息粒度.

证明　1) 显然,非负性成立.

2) 令 $P,Q\subseteq A$, 则完备信息系统下的信息 $U/\mathrm{IND}(P)=\{X_1,X_2,\ldots,X_m\}$ 和信息 $U/\mathrm{IND}(Q)=\{Y_1,Y_2,\ldots,Y_n\}$ 可表示为信息 $K(P)=(S_P(u_1),S_P(u_2),\ldots,S_P(u_{|U|}))$ 和信息 $K(Q)=(S_Q(u_1),\ S_Q(u_2),\ldots,S_Q(u_{|U|}))$.

如果 $P\approx Q$, 则任意的 $i\in\{1,2,\ldots,|U|\}$, 都有 $S_P(u_i)=S_Q(u_i)$, 即 $|S_P(u_i)|=|S_Q(u_i)|$. 因此,我们有

$$\begin{aligned}\mathrm{GK}(P)&=\frac{1}{|U|^2}\sum_{i=1}^{m}|X_i|^2=\frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_P(u_i)|\\&=\frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_Q(u_i)|=\frac{1}{|U|^2}\sum_{j=1}^{n}|Y_j|^2\\&=\mathrm{GK}(Q).\end{aligned}$$

3) 令 $P,Q\subseteq A$, 且 $P\prec' Q$, 则任意的对象 $u_i\in U$, 我们都有 $S_P(u_i)\subseteq S_Q(u_i)$, $|S_P(u_i)|\leqslant|S_Q(u_i)|$, 并且存在某个 $u_0\in U$ 使得 $S_P(u_0)\subset S_Q(u_0)$, $|S_P(u_0)|<|S_Q(u_0)|$. 所以

$$\begin{aligned}\mathrm{GK}(P)&=\frac{1}{|U|^2}\sum_{i=1}^{m}|X_i|^2=\frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_P(u_i)|\\&=\frac{1}{|U|^2}\left(\sum_{i=1,u_i\neq u_0}^{|U|}|S_P(u_i)|+|S_P(u_0)|\right)\\&<\frac{1}{|U|^2}\left(\sum_{i=1,u_i\neq u_0}^{|U|}|S_Q(u_i)|+|S_Q(u_0)|\right)\\&=\frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_Q(u_i)|=\frac{1}{|U|^2}\sum_{j=1}^{n}|Y_j|^2\\&=\mathrm{GK}(Q).\end{aligned}$$

由此可得, 定义 2 中的 *GK* 是定义 1 意义下的一个信息粒度. 证毕.

定义 3[20]　设 $S=(U,A)$ 是一个非完备信息系统, $K(A)=(S_A(u_1),S_A(u_2),\ldots,S_A(u_{|U|}))$, 则 A 的信息粒度定义为

$$\mathrm{GK}(A)=\frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_A(u_i)|.\tag{2}$$

若 $K(A)=\omega$, 则 A 的信息粒度可取得最小值 $|U|/|U|^2=1/|U|$.

若 $K(A)=\delta$, 则 A 的信息粒度可取得最大值 $|U|^2/|U|^2=1$.

定理 4 定义 3 中的 GK 是定义 1 意义下的一个信息粒度.

证明 1) 显然, 非负性成立.

2) 令 $P,Q\subseteq A$, $K(P)=(S_P(u_1),S_P(u_2),\ldots,S_P(u_{|U|}))$ 且 $K(Q)=(S_Q(u_1),S_Q(u_2),\ldots,S_Q(u_{|U|}))$.

如果 $P\approx Q$, 则任意的 $i\in\{1,2,\ldots,|U|\}$, 都有 $S_P(u_i)=S_Q(u_i)$, 即 $|S_P(u_i)|=|S_Q(u_i)|$. 因此,我们有

$$\mathrm{GK}(P)=\frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_P(u_i)|=\frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_Q(u_i)|=\mathrm{GK}(Q).$$

3) 令 $P,Q\subseteq A$, 且 $P\prec' Q$, 则任意的对象 $u_i\in U$, 我们都有 $S_P(u_i)\subseteq S_Q(u_i)$, $|S_P(u_i)|\leqslant|S_Q(u_i)|$, 并且存在某个 $u_0\in U$ 使得 $S_P(u_0)\subset S_Q(u_0)$, $|S_P(u_0)|<|S_Q(u_0)|$. 所以

$$\begin{aligned}\mathrm{GK}(P)&=\frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_P(u_i)|\\&=\frac{1}{|U|^2}\left(\sum_{i=1,u_i\neq u_0}^{|U|}|S_P(u_i)|+|S_P(u_0)|\right)\\&<\frac{1}{|U|^2}\left(\sum_{i=1,u_i\neq u_0}^{|U|}|S_Q(u_i)|+|S_Q(u_0)|\right)\\&=\frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_Q(u_i)|\\&=\mathrm{GK}(Q).\end{aligned}$$

由此可得, 定义 3 中的 GK 是定义 1 意义下的一个信息粒度. 证毕.

定义 4[21] 令 $S=(U,A)$ 是一个完备信息系统, $U/\mathrm{IND}(A)=\{X_1,X_2,\ldots,X_m\}$, 则 A 的组合粒度定义为

$$\mathrm{CG}(A)=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\frac{C_{|X_i|}^2}{C_{|U|}^2},\tag{3}$$

其中 $|X_i|/|U|$ 表示等价类 X_i 在论域 U 中的比率; $C_{|X_i|}^2/C_{|U|}^2$ 表示等价类 X_i 中的元素对数在论域 U 中所有元素对数中所占的比例.

如果 $U/\mathrm{IND}(A)=\omega$, 那么 A 的组合粒度达到最小值 $0/C_{|U|}^2=0$.

如果 $U/\mathrm{IND}(A)=\delta$, 那么 A 的组合粒度达到最大值 $C_{|U|}^2/C_{|U|}^2=1$.

定理 5 定义 4 中的 CG 是定义 1 意义下的一个信息粒度.

证明 1) 显然, 它是非负的.

2) 令 $P,Q\subseteq A$, 则完备信息系统下的信息 $U/\mathrm{IND}(P)=\{X_1,X_2,\ldots,X_m\}$ 和信息 $U/\mathrm{IND}(Q)=\{Y_1,Y_2,\ldots,Y_n\}$ 可表示为信息 $K(P)=(S_P(u_1),S_P(u_2),\ldots,S_P(u_{|U|}))$ 和信息 $K(Q)=$

$(S_Q(u_1),\ S_Q(u_2),\dots,S_Q(u_{|U|}))$. 不妨设 $X_i=\{u_{i1},u_{i2},\dots,u_{is_i}\}$, $i\in\{1,2,\dots,m\}$, 其中 $|X_i|=s_i$, 且 $\sum_{i=1}^{m}s_i=|U|$, 则 $K(P)$ 与 $U/\mathrm{IND}(P)$ 中的元素之间存在下面的对应关系:

$$X_i=S_P(u_{i1})=S_P(u_{i2})=\cdots=S_P(u_{is_i}),$$
$$|X_i|=|S_P(u_{i1})|=|S_P(u_{i2})|=\cdots=|S_P(u_{is_i})|=s_i.$$

同理, $|Y_j|=|S_Q(u_{j1})|=|S_Q(u_{j2})|=\cdots=|S_Q(u_{jt_j})|=t_j$. 所以, 我们有

$$|X_i|\times\frac{C_{|X_i|}^2}{C_{|U|}^2}=s_i\times\frac{C_{|X_i|}^2}{C_{|U|}^2}=\sum_{k=1}^{s_i}\frac{C_{|S_P(u_{ik})|}^2}{C_{|U|}^2},$$
$$|Y_j|\times\frac{C_{|Y_j|}^2}{C_{|U|}^2}=t_j\times\frac{C_{|Y_j|}^2}{C_{|U|}^2}=\sum_{k=1}^{t_j}\frac{C_{|S_Q(u_{jk})|}^2}{C_{|U|}^2}.$$

如果 $P\approx Q$, 则任意的 $i\in\{1,2,\dots,|U|\}$, 都有 $S_P(u_i)=S_Q(u_i)$, 即 $|S_P(u_i)|=|S_Q(u_i)|$. 因此,我们有

$$\begin{aligned}\mathrm{CG}(P)&=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\frac{C_{|X_i|}^2}{C_{|U|}^2}=\sum_{i=1}^{m}s_i\times\frac{C_{|X_i|}^2}{C_{|U|}^2}\\&=\frac{1}{|U|}\sum_{i=1}^{m}\sum_{k=1}^{s_i}\frac{C_{|S_P(u_{ik})|}^2}{C_{|U|}^2}=\sum_{i=1}^{|U|}\frac{1}{|U|}\frac{C_{|S_P(u_i)|}^2}{C_{|U|}^2}\\&=\sum_{i=1}^{|U|}\frac{1}{|U|}\frac{C_{|S_Q(u_i)|}^2}{C_{|U|}^2}=\sum_{j=1}^{n}\sum_{k=1}^{t_j}\frac{C_{|S_Q(u_{jk})|}^2}{C_{|U|}^2}\\&=\sum_{j=1}^{n}\frac{|Y_j|}{|U|}\frac{C_{|Y_j|}^2}{C_{|U|}^2}=\mathrm{CG}(Q).\end{aligned}$$

3) 令 $P,Q\subseteq A$, 且 $P\prec' Q$, 则任意的对象 $u_i\in U$, 我们都有 $S_P(u_i)\subseteq S_Q(u_i)$, $|S_P(u_i)|\leqslant|S_Q(u_i)|$, 并且存在某个 $u_0\in U$ 使得 $S_P(u_0)\subset S_Q(u_0)$, $|S_P(u_0)|<|S_Q(u_0)|$. 所以

$$\begin{aligned}\mathrm{CG}(P)&=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\frac{C_{|X_i|}^2}{C_{|U|}^2}=\sum_{i=1}^{|U|}\frac{1}{|U|}\frac{C_{|S_P(u_i)|}^2}{C_{|U|}^2}\\&=\sum_{i=1,u_i\neq u_0}^{|U|}\frac{1}{|U|}\frac{C_{|S_P(u_i)|}^2}{C_{|U|}^2}+\frac{1}{|U|}\frac{C_{|S_P(u_0)|}^2}{C_{|U|}^2}\\&<\sum_{i=1,u_i\neq u_0}^{|U|}\frac{1}{|U|}\frac{C_{|S_Q(u_i)|}^2}{C_{|U|}^2}+\frac{1}{|U|}\frac{C_{|S_Q(u_0)|}^2}{C_{|U|}^2}\\&=\sum_{i=1}^{|U|}\frac{1}{|U|}\frac{C_{|S_Q(u_i)|}^2}{C_{|U|}^2}=\sum_{j=1}^{n}\frac{|Y_j|}{|U|}\frac{C_{|Y_j|}^2}{C_{|U|}^2}\\&=\mathrm{CG}(Q).\end{aligned}$$

由此可得，定义 4 中的 CG 是定义 1 意义下的一个信息粒度. 证毕.

定义 5[21] 令 $S=(U,A)$ 是一个非完备信息系统，$K(A)=(S_A(u_1),S_A(u_2),\dots,S_A(u_{|U|}))$，则 A 的组合粒度定义为

$$\mathrm{CG}(A)=\frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C_{|S_A(u_i)|}^2}{C_{|U|}^2}, \tag{4}$$

其中 $C_{|S_A(u_i)|}^2/C_{|U|}^2$ 表示相容类 $S_A(u_i)$ 中的元素对数在论域 U 中所有元素对数中所占的比例.

如果 $K(A)=\omega$，那么 A 的组合粒度达到最小值 $|U|\times 0/|U|=0$.

如果 $K(A)=\delta$，那么 A 的组合粒度达到最大值 $|U|/|U|=1$.

定理 6 定义 5 中的 CG 是定义 1 意义下的信息粒度.

证明 1) 显然，它是非负的.

2) 令 $P,Q\subseteq A$，$K(P)=(S_P(u_1),S_P(u_2),\dots,S_P(u_{|U|}))$，且 $K(Q)=(S_Q(u_1),S_Q(u_2),\dots,S_Q(u_{|U|}))$. 如果 $P\approx Q$，则任意的 $i\in\{1,2,\dots,|U|\}$，都有 $S_P(u_i)=S_Q(u_i)$，即 $|S_P(u_i)|=|S_Q(u_i)|$. 因此,我们有

$$\mathrm{CG}(P)=\frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C_{|S_P(u_i)|}^2}{C_{|U|}^2}=\frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C_{|S_Q(u_i)|}^2}{C_{|U|}^2}=\mathrm{CG}(Q).$$

3) 令 $P,Q\subseteq A$，且 $P\prec' Q$，则任意的对象 $u_i\in U$，我们都有 $S_P(u_i)\subseteq S_Q(u_i)$，$|S_P(u_i)|\leqslant|S_Q(u_i)|$，并且存在某个 $u_0\in U$ 使得 $S_P(u_0)\subset S_Q(u_0)$，$|S_P(u_0)|<|S_Q(u_0)|$. 所以

$$\begin{aligned}\mathrm{CG}(P)&=\frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C_{|S_P(u_i)|}^2}{C_{|U|}^2}=\sum_{i=1,u_i\neq u_0}^{|U|}\frac{C_{|S_P(u_i)|}^2}{C_{|U|}^2}+\frac{C_{|S_P(u_0)|}^2}{C_{|U|}^2}\\&<\sum_{i=1,u_i\neq u_0}^{|U|}\frac{C_{|S_Q(u_i)|}^2}{C_{|U|}^2}+\frac{C_{|S_Q(u_0)|}^2}{C_{|U|}^2}=\frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C_{|S_Q(u_i)|}^2}{C_{|U|}^2}\\&=\mathrm{CG}(Q).\end{aligned}$$

由此可得，定义 5 中的 CG 是定义 1 意义下的一个信息粒度. 证毕.

定义 6[18,20] 设 $S=(U,A)$ 是一个完备信息系统，$U/\mathrm{IND}(A)=\{X_1,X_2,\dots,X_m\}$，则 A 的粗糙熵定义为

$$E_r(A)=-\sum_{i=1}^{m}\frac{|X_i|}{|U|}\log_2\frac{1}{|X_i|}. \tag{5}$$

如果 $U/\mathrm{IND}(A)=\omega$，那么 A 的粗糙熵达到最小值 0.

如果 $U/\mathrm{IND}(A)=\delta$，那么 A 的粗糙熵达到最大值 $\log_2|U|$.

定理 7 定义 6 中的 E_r 是定义 1 意义下的一个信息粒度.

证明 1) 显然，它是非负的.

2) 设 $P,Q\subseteq A$，$U/\mathrm{IND}(P)=\{X_1,X_2,\dots,X_m\}$，$U/\mathrm{IND}(Q)=\{Y_1,Y_2,\dots,Y_n\}$，$K(P)=(S_P(u_1),$

$S_P(u_2),\ldots,S_P(u_{|U|}))$ 且 $K(Q)=(S_Q(u_1),S_Q(u_2),\ldots,S_Q(u_{|U|}))$. 如果 $P\approx Q$, 则任意的 $i\in\{1,2,\ldots,|U|\}$, 都有 $S_P(u_i)=S_Q(u_i)$, 即 $|S_P(u_i)|=|S_Q(u_i)|$. 因此

$$\begin{aligned}E_r(P)&=-\sum_{i=1}^{m}\frac{|X_i|}{|U|}\log_2\frac{1}{|X_i|}=-\sum_{i=1}^{m}\sum_{k=1}^{s_i}\frac{1}{|U|}\log_2\frac{1}{|S_P(u_{ik})|}\\&=-\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{1}{|S_P(u_i)|}=-\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{1}{|S_Q(u_i)|}\\&=-\sum_{j=1}^{n}\frac{|Y_j|}{|U|}\log_2\frac{1}{|Y_j|}=E_r(Q).\end{aligned}$$

3) 令 $P,Q\subseteq A$, 且 $P\prec' Q$, 则任意的对象 $u_i\in U$, 我们都有 $S_P(u_i)\subseteq S_Q(u_i)$, $1\leqslant|S_P(u_i)|\leqslant|S_Q(u_i)|$, 并且存在某个 $u_0\in U$ 使得 $S_P(u_0)\subset S_Q(u_0)$, $1\leqslant|S_P(u_0)|<|S_Q(u_0)|$, 所以

$$\begin{aligned}E_r(P)&=-\sum_{i=1}^{m}\frac{|X_i|}{|U|}\log_2\frac{1}{|X_i|}\\&=-\sum_{i=1}^{m}\left(\frac{1}{|U|}\log_2\frac{1}{|S_P(u_{i1})|}+\frac{1}{|U|}\log_2\frac{1}{|S_P(u_{i2})|}+\cdots+\frac{1}{|U|}\log_2\frac{1}{|S_P(u_{is_i})|}\right)\\&=-\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{1}{|S_P(u_i)|}=\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2|S_P(u_i)|\\&=\frac{1}{|U|}\sum_{i=1,u_i\neq u_0}^{|U|}\log_2|S_P(u_i)|+\frac{1}{|U|}\log_2|S_P(u_0)|\\&=\frac{1}{|U|}\log_2\prod_{i=1,u_i\neq u_0}^{|U|}|S_P(u_i)|+\frac{1}{|U|}\log_2|S_P(u_0)|\\&<\frac{1}{|U|}\log_2\prod_{i=1,u_i\neq u_0}^{|U|}|S_Q(u_i)|+\frac{1}{|U|}\log_2|S_Q(u_0)|\\&=\frac{1}{|U|}\log_2\prod_{i=1}^{|U|}|S_Q(u_i)|=\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2|S_Q(u_i)|\\&=-\sum_{j=1}^{n}\sum_{k=1}^{t_j}\frac{1}{|U|}\log_2\frac{1}{|S_P(u_{jk})|}=-\sum_{j=1}^{n}\frac{|Y_j|}{|U|}\log_2\frac{1}{|Y_j|}\\&=E_r(Q),\end{aligned}$$

即 $E_r(P)<E_r(Q)$.

由此可得, 定义 6 中的 E_r 是定义 1 意义下的一个信息粒度. 证毕.

定义 7[19]　设 $S=(U,A)$ 是一个非完备信息系统, $K(A)=(S_A(u_1),S_A(u_2),\ldots,S_A(u_{|U|}))$, 则 A 的粗糙熵定义为

$$E_r(A)=-\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{1}{|S_A(u_i)|}. \tag{6}$$

如果 $K(A)=\omega$，那么 A 的粗糙熵达到最小值 0.

如果 $K(A)=\delta$，那么 A 的粗糙熵达到最大值 $\log_2|U|$.

定理 8　设 $S=(U,A)$ 是一个完备信息系统，$U/\mathrm{IND}(A)=\{X_1,X_2,\ldots,X_m\}$，则 A 的粗糙熵退化为

$$E_r(A)=-\sum_{i=1}^{m}\frac{|X_i|}{|U|}\log_2\frac{1}{|X_i|}.$$

证明　假设 $K(A)=(S_A(u_1),S_A(u_2),\ldots,S_A(u_{|U|}))$，$X_i=\{u_{i1},u_{i2},\ldots,u_{is_i}\}$，其中 $|X_i|=s_i$，$\sum_{i=1}^{m}s_i=|U|$ 则存在以下关系：$X_i=S_A(u_{i1})=S_A(u_{i2})=\cdots=S_A(u_{is_i})$. 于是

$$\begin{aligned}
&-\sum_{i=1}^{m}\frac{|X_i|}{|U|}\log_2\frac{1}{|X_i|}\\
&=-\sum_{i=1}^{m}\left(\frac{1}{|U|}\log_2\frac{1}{|S_A(u_{i1})|}+\frac{1}{|U|}\log_2\frac{1}{|S_A(u_{i2})|}+\cdots+\frac{1}{|U|}\log_2\frac{1}{|S_A(u_{is_i})|}\right)\\
&=-\left(\frac{1}{|U|}\log_2\frac{1}{|S_A(u_1)|}+\frac{1}{|U|}\log_2\frac{1}{|S_A(u_2)|}+\cdots+\frac{1}{|U|}\log_2\frac{1}{|S_A(u_{|U|})|}\right)\\
&=-\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{1}{|S_A(u_i)|}=E_r(A).
\end{aligned}$$

证毕.

定理 9　定义 7 中的 E_r 是定义 1 意义下的一个信息粒度.

证明　相似于定理 7 可以得到证明.

3　信息系统中的熵理论

熵是系统无序的度量，熵越大，无序越高. Shannon 将物理学中的熵概念引入到了信息论中，用来度量信息系统结构中的不确定性.

定义 8[22]　设 $S=(U,A)$ 是一个完备信息系统，$U/\mathrm{IND}(A)=\{X_1,X_2,\ldots,X_m\}$，其上有概率分布 $p_i=|X_i|/|U|$，则称

$$H(A)=-\sum_{i=1}^{m}p_i\log_2 p_i \tag{7}$$

为信息系统 S 的信息熵. 当某个 p_i 为 0 时，理解为 $0\cdot\log_2 0=0$.

定理 10　设 $S=(U,A)$ 是一个完备信息系统，$P,Q\subseteq A$，如果 $P\prec' Q$，则 $H(Q)<H(P)$.

证明　令 $U/\mathrm{IND}(P)=\{X_1,X_2,\ldots,X_m\}$，$U/\mathrm{IND}(Q)=\{Y_1,Y_2,\ldots,Y_n\}$.

设 $X_i=\{u_{i1},u_{i2},\ldots,u_{is_i}\}$，$Y_j=\{u_{j1},u_{j2},\ldots,u_{jt_j}\}$，其中 $|X_i|=s_i$，$|Y_j|=t_j$ 且 $\sum_{i=1}^{m}s_i=|U|$，$\sum_{j=1}^{n}t_j=|U|$. 它们在信息系统中的统一表示形式为 $K(P)=(S_P(u_1),S_P(u_2),\ldots,S_P(u_{|U|})\}$ 与

$K(Q)=\{S_Q(u_1),\ S_Q(u_2),\ldots,S_Q(u_{|U|})\}$. $K(P)$ 中的元素与 $U/\text{IND}(P)$ 中的元素之间存在下面的对应关系

$$X_i=S_P(u_{i1})=S_P(u_{i2})=\cdots=S_P(u_{is_i}),$$
$$|X_i|=|S_P(u_{i1})|=|S_P(u_{i2})|=\cdots=|S_P(u_{is_i})|=s_i.$$

同理, 我们有 $|Y_j|=|S_Q(u_{j1})|=|S_Q(u_{j2})|=\cdots=|S_Q(u_{jt_j})|=t_j$. 所以

$$\frac{|X_i|}{|U|}\log_2\frac{|X_i|}{|U|}=\frac{1}{|U|}\log_2\frac{|S_P(u_{i1})|}{|U|}+\frac{1}{|U|}\log_2\frac{|S_P(u_{i2})|}{|U|}+\cdots+\frac{1}{|U|}\log_2\frac{|S_P(u_{is_i})|}{|U|},$$
$$\frac{|Y_j|}{|U|}\log_2\frac{|Y_j|}{|U|}=\frac{1}{|U|}\log_2\frac{|S_Q(u_{j1})|}{|U|}+\frac{1}{|U|}\log_2\frac{|S_Q(u_{j2})|}{|U|}+\cdots+\frac{1}{|U|}\log_2\frac{|S_Q(u_{jt_j})|}{|U|}.$$

又由于 $P\prec' Q$, 于是任意的 $u_i\in U$, 都有 $S_P(u_i)\subseteq S_Q(u_i)$, $1\leqslant|S_P(u_i)|\leqslant|S_Q(u_i)|$, 并且存在某个 $u_0\in U$ 使得 $S_P(u_0)\subset S_Q(u_0)$, $1\leqslant|S_P(u_0)|<|S_Q(u_0)|$. 因此

$$\begin{aligned}H(P)&=-\sum_{i=1}^{m}\frac{|X_i|}{|U|}\log_2\frac{|X_i|}{|U|}=-\sum_{i=1}^{m}\sum_{k=1}^{s_i}\frac{1}{|U|}\log_2\frac{|S_P(u_{ik})|}{|U|}=-\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{|S_P(u_i)|}{|U|}\\&=-\sum_{i=1,u_i\neq u_0}^{|U|}\frac{1}{|U|}\log_2\frac{|S_P(u_i)|}{|U|}-\frac{1}{|U|}\log_2\frac{|S_P(u_0)|}{|U|}\\&>-\sum_{i=1,u_i\neq u_0}^{|U|}\frac{1}{|U|}\log_2\frac{|S_Q(u_i)|}{|U|}-\frac{1}{|U|}\log_2\frac{|S_Q(u_0)|}{|U|}\\&=-\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{|S_Q(u_i)|}{|U|}=-\sum_{j=1}^{n}\sum_{k=1}^{t_j}\frac{1}{|U|}\log_2\frac{|S_Q(\mu_{jk})|}{|U|}\\&=-\sum_{j=1}^{n}\frac{|Y_j|}{|U|}\log_2\frac{|Y_j|}{|U|}=H(Q).\end{aligned}$$

显然, $H(Q)<H(P)$. 证毕.

方便地, 称由偏序关系 $\preceq'$ 导致的熵值的单调性为粒化单调性.

定义 9　设 $S=(U,A)$ 是一个非完备信息系统, $K(A)=(S_A(u_1),S_A(u_2),\ldots,S_A(u_{|U|}))$, 则 A 的信息熵定义为

$$H(A)=-\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{|S_A(u_i)|}{|U|},\tag{8}$$

其中 $H:A\to[0,\infty)$.

如果 $K(A)=\omega$, 那么 A 的信息熵达到最大值 $\log_2|U|$.

如果 $K(A)=\delta$, 那么 A 的信息熵达到最小值 0.

定理 11　设 $S=(U,A)$ 是一个非完备信息系统, $P,Q\subseteq A$, 如果 $P\prec' Q$, 则 $H(Q)<H(P)$.

证明　相似于定理 10 可以得到证明.

作者在完备信息系统下提出了一种新的信息熵，它可以用来度量粗糙集理论中的不确定性和模糊性.

定义 10[28] 设 $S=(U,A)$ 是一个完备信息系统，$U/\mathrm{IND}(A)=\{X_1,X_2,\dots,X_m\}$. 完备信息系统的一种新的信息熵定义为

$$E(A)=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\frac{|X_i^c|}{|U|}=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\left(1-\frac{|X_i|}{|U|}\right), \tag{9}$$

其中 X_i^c 表示 X_i 的补集，即 $X_i^c=U-X_i$，$|X_i|/|U|$ 表示 X_i 在论域 U 中的概率，$|X_i^c|/|U|$ 表示 X_i 的补集在论域 U 中的概率.

在此基础上，作者给出了这个熵的条件熵

$$E(Q|P)=\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{|X_i\cap Y_j|}{|U|}\frac{|Y_j^c-X_i^c|}{|U|}$$

和互信息

$$E(P;Q)=\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{|X_i\cap Y_j|}{|U|}\frac{|Y_j^c\cap X_i^c|}{|U|},$$

它们之间具有关系 $E(P;Q)=E(Q)-E(Q|P)$[28].

区别于 Shannon 熵，这个熵不仅可以用于度量信息系统中的不确定性，同时也可以用于粗糙集和粗糙分类的模糊性度量.

定理 12 设 $S=(U,A)$ 是一个完备信息系统，$P,Q\subseteq A$，如果 $P\prec' Q$，则 $E(Q)<E(P)$.

证明 相似于定理 10，可以得到

$$\frac{|X_i|}{|U|}\left(1-\frac{|X_i|}{|U|}\right)=\frac{1}{|U|}\left(1-\frac{|S_P(u_{i1})|}{|U|}\right)+\frac{1}{|U|}\left(1-\frac{|S_P(u_{i2})|}{|U|}\right)+\cdots+\frac{1}{|U|}\left(1-\frac{|S_P(u_{is_i})|}{|U|}\right),$$

$$\frac{|Y_j|}{|U|}\left(1-\frac{|Y_j|}{|U|}\right)=\frac{1}{|U|}\left(1-\frac{|S_P(u_{j1})|}{|U|}\right)+\frac{1}{|U|}\left(1-\frac{|S_P(u_{j2})|}{|U|}\right)+\cdots+\frac{1}{|U|}\left(1-\frac{|S_P(u_{jt_j})|}{|U|}\right).$$

由于 $P\prec' Q$，于是任意的 $u_i\in U$，都有 $S_P(u_i)\subseteq S_Q(u_i)$，$|S_P(u_i)|\leqslant|S_Q(u_i)|$，并且存在某个 $u_0\in U$ 使得 $S_P(u_0)\subset S_Q(u_0)$, $|S_P(u_0)|<|S_Q(u_0)|$. 因此

$$\begin{aligned}E(P)&=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\left(1-\frac{|X_i|}{|U|}\right)=\sum_{i=1}^{m}\sum_{k=1}^{s_i}\frac{1}{|U|}\left(1-\frac{|S_P(u_{ik})|}{|U|}\right)\\&=\sum_{i=1}^{m}\left(\frac{1}{|U|}\left(1-\frac{|S_P(u_{i1})|}{|U|}\right)+\frac{1}{|U|}\left(1-\frac{|S_P(u_{i2})|}{|U|}\right)+\cdots+\frac{1}{|U|}\left(1-\frac{|S_P(u_{is_i})|}{|U|}\right)\right)\\&=\sum_{i=1}^{|U|}\frac{1}{|U|}\left(1-\frac{|S_P(u_i)|}{|U|}\right)=\sum_{i=1,u_i\neq u_0}^{|U|}\frac{1}{|U|}\left(1-\frac{|S_P(u_i)|}{|U|}\right)+\frac{1}{|U|}\left(1-\frac{|S_P(u_0)|}{|U|}\right)\\&>\sum_{i=1,u_i\neq u_0}^{|U|}\frac{1}{|U|}\left(1-\frac{|S_Q(u_i)|}{|U|}\right)+\frac{1}{|U|}\left(1-\frac{|S_Q(u_0)|}{|U|}\right)\end{aligned}$$

$$=\sum_{i=1}^{|U|}\frac{1}{|U|}\left(1-\frac{|S_Q(u_i)|}{|U|}\right)=\sum_{i=1}^{|U|}\frac{1}{|U|}\left(1-\frac{|S_Q(u_i)|}{|U|}\right)=\sum_{j=1}^{n}\sum_{k=1}^{t_j}\frac{1}{|U|}\left(1-\frac{|S_Q(u_{jk})|}{|U|}\right)$$

$$=\sum_{j=1}^{n}\frac{|Y_j|}{|U|}\left(1-\frac{|Y_j|}{|U|}\right)=E(Q).$$

显然, $E(Q)<E(P)$. 证毕.

定义 11　设 $S=(U,A)$ 是一个非完备信息系统, $K(A)=(S_A(u_1),S_A(u_2),\ldots,S_A(u_{|U|}))$, 则 A 的信息熵定义为

$$E(A)=\sum_{i=1}^{|U|}\frac{1}{|U|}\left(1-\frac{|S_A(u_i)|}{|U|}\right). \tag{10}$$

若 $K(A)=\omega$, 则 A 的信息熵可得到最大值 $1-1/|U|$.

若 $K(A)=\delta$, 则 A 的信息熵可得到最小值 0.

定理 13　设 $S=(U,A)$ 是一个完备信息统, $U/\mathrm{IND}(A)=\{X_1,X_2,\ldots,X_m\}$, $K(A)=(S_A(u_1),S_A(u_2),\ldots,S_A(u_{|U|}))$, 则 A 的信息熵退化为

$$E(A)=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\left(1-\frac{|X_i|}{|U|}\right),$$

即

$$E(A)=\sum_{i=1}^{|U|}\frac{1}{|U|}\left(1-\frac{|S_A(u_i)|}{|U|}\right)=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\left(1-\frac{|X_i|}{|U|}\right).$$

证明　相似于定理 12 可以得到证明.

由定理 13 可知, 完备信息系统中的信息熵是非完备信息系统中的信息熵一种特殊情形.

定理 14　设 $S=(U,A)$ 是一个非完备信息系统, $P,Q\subseteq A$, 如果 $P\prec' Q$, 则 $E(Q)<E(P)$.

证明　相似于定理 12 可以得到证明.

定义 12[21]　设 $S=(U,A)$ 是一个非完备信息系统, $K(A)=(S_A(u_1),S_A(u_2),\ldots,S_A(u_{|U|}))$, 则 A 的组合熵定义为

$$\mathrm{CE}(A)=\frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C_{|U|}^2-C_{|S_A(u_i)|}^2}{C_{|U|}^2}, \tag{11}$$

其中 $(C_{|U|}^2-C_{|S_A(u_i)|}^2)/C_{|U|}^2$ 表示可能相互区分的元素对数在论域 U 中所有元素对数中所占的比例.

若 $K(A)=\omega$, 则 A 的组合熵可得到最大值 1.

若 $K(A)=\delta$, 则 A 的组合熵可得到最小值 0.

定理 15　设 $S=(U,A)$ 是一个非完备信息系统, $P,Q\subseteq A$, 如果 $P\prec' Q$, 则 $\mathrm{CE}(Q)<\mathrm{CE}(P)$.

证明　不妨设 $K(P)=(S_P(u_1),S_P(u_2),\ldots,S_P(u_{|U|}))$ 和 $K(Q)=(S_Q(u_1),S_Q(u_2),\ldots,S_Q(u_{|U|}))$. 由于 $P\prec' Q$, 于是任意的 $u_i\in U$, 都有 $S_P(u_i)\subseteq S_Q(u_i)$, $|S_P(u_i)|\leqslant|S_Q(u_i)|$, 并且存在某个

$u_0 \in U$ 使得 $S_P(u_0) \subset S_Q(u_0)$, $|S_P(u_0)| < |S_Q(u_0)|$. 因此

$$\begin{aligned}\mathrm{CE}(P) &= \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C_{|U|}^2 - C_{|S_P(u_i)|}^2}{C_{|U|}^2} = 1 - \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C_{|S_P(u_i)|}^2}{C_{|U|}^2} = 1 - \frac{1}{|U|}\sum_{i=1,u_i \neq u_0}^{|U|}\frac{C_{|S_P(u_i)|}^2}{C_{|U|}^2} - \frac{1}{|U|}\frac{C_{|S_P(u_0)|}^2}{C_{|U|}^2} \\ &> 1 - \frac{1}{|U|}\sum_{i=1,u_i \neq u_0}^{|U|}\frac{C_{|S_Q(u_i)|}^2}{C_{|U|}^2} - \frac{1}{|U|}\frac{C_{|S_Q(u_0)|}^2}{C_{|U|}^2} = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C_{|U|}^2 - C_{|S_Q(u_i)|}^2}{C_{|U|}^2} = \mathrm{CE}(Q).\end{aligned}$$

显然, $\mathrm{CE}(Q) < \mathrm{CE}(P)$. 证毕.

定理 16 设 $S=(U,A)$ 是一个完备信息系统, $U/\mathrm{IND}(A)=\{X_1, X_2, \ldots, X_m\}$, 则 A 的组合熵退化为

$$\mathrm{CE}(A) = \sum_{i=1}^{m}\frac{|X_i|}{|U|}\frac{C_{|U|}^2 - C_{|X_i|}^2}{C_{|U|}^2}, \tag{12}$$

其中 $(C_{|U|}^2 - C_{|X_i|}^2)/C_{|U|}^2$ 表示能够相互区分的元素对数在论域 U 中所有元素对数中所占的比例.

证明 不妨设 $U/\mathrm{IND}(A)=\{X_1, X_2, \ldots, X_m\}$, $X_i = \{u_{i1}, u_{i2}, \ldots, u_{is_i}\}$, $i \in \{1, 2, \ldots, m\}$, 其中 $|X_i| = s_i$, 且 $\sum_{i=1}^{m} s_i = |U|$. 令 $K(A) = (S_A(u_1), S_A(u_2), \ldots, S_A(u_{|U|}))$, 则 $K(A)$ 中的元素与 $U/\mathrm{IND}(A)$ 中的元素之间存在下面的对应关系

$$X_i = S_P(u_{i1}) = S_P(u_{i2}) = \cdots = S_P(u_{is_i}),$$

即

$$|X_i| = |S_P(u_{i1})| = |S_P(u_{i2})| = \cdots = |S_P(u_{is_i})|.$$

所以, 我们有

$$|X_i| \times \frac{C_{|X_i|}^2}{C_{|U|}^2} = s_i \times \frac{C_{|X_i|}^2}{C_{|U|}^2} = \sum_{k=1}^{s_i}\frac{C_{|S_A(u_{ik})|}^2}{C_{|U|}^2}.$$

因此

$$\begin{aligned}\mathrm{CE}(A) &= \sum_{i=1}^{m}\frac{|X_i|}{|U|}\frac{C_{|U|}^2 - C_{|X_i|}^2}{C_{|U|}^2} = 1 - \frac{1}{|U|}\sum_{i=1}^{m}|X_i| \times \frac{C_{|X_i|}^2}{C_{|U|}^2} \\ &= 1 - \frac{1}{|U|}\sum_{i=1}^{m} s_i \times \frac{C_{|X_i|}^2}{C_{|U|}^2} = 1 - \frac{1}{|U|}\sum_{i=1}^{m}\sum_{k=1}^{s_i}\frac{C_{|S_A(u_{ik})|}^2}{C_{|U|}^2} \\ &= 1 - \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C_{|S_A(u_i)|}^2}{C_{|U|}^2} = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C_{|U|}^2 - C_{|S_A(u_i)|}^2}{C_{|U|}^2}.\end{aligned}$$

证毕.

4 信息粒度与熵之间的关系

信息粒度其实是对信息细化的不同层次的平均度量. 由此得知, 熵和信息粒度之间存在

某种意义上的互补关系. 也就是说, 熵越大, 信息粒度越小; 熵越小, 信息粒度越大.

4.1　完备信息系统下信息粒度与熵之间的关系

定理 17[18]　设 $S=(U,A)$ 是一个完备信息系统, $U/\mathrm{IND}(A)=\{X_1,X_2,\ldots,X_m\}$, 则信息熵 $E(A)$ 与信息粒度 $\mathrm{GK}(A)$ 具有下面的关系:

$$E(A)+\mathrm{GK}(A)=1. \tag{13}$$

定理 18[18]　设 $S=(U,A)$ 是一个完备信息系统, $U/\mathrm{IND}(A)=\{X_1,X_2,\ldots,X_m\}$, 则 Shannon 熵 $H(A)$ 与粗糙熵 $E_r(A)$ 具有下面的关系:

$$H(A)+E_r(A)=\log_2|U|. \tag{14}$$

定理 19　设 $S=(U,A)$ 是一个完备信息系统, $U/\mathrm{IND}(A)=\{X_1,X_2,\ldots,X_m\}$, 则组合熵 $\mathrm{CE}(A)$ 与组合粒度 $\mathrm{CG}(A)$ 具有下面的关系:

$$\mathrm{CE}(A)+\mathrm{CG}(A)=1. \tag{15}$$

证明　根据它们的定义, 我们有

$$\mathrm{CE}(A)=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\frac{C_{|U|}^2-C_{|X_i|}^2}{C_{|U|}^2}=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\left(1-\frac{C_{|X_i|}^2}{C_{|U|}^2}\right)$$

$$=1-\sum_{i=1}^{m}\frac{|X_i|}{|U|}\frac{C_{|X_i|}^2}{C_{|U|}^2}=1-\mathrm{CG}(A).$$

显然, $\mathrm{CE}(A)+\mathrm{CG}(A)=1$. 证毕.

4.2　非完备信息系统下信息粒度与熵之间的关系

定理 20　设 $S=(U,A)$ 是一个非完备信息系统, $K(A)=(S_A(u_1),S_A(u_2),\ldots,S_A(u_{|U|}))$, 则 A 的信息熵和信息粒度之间具有下面的关系:

$$E(A)+\mathrm{GK}(A)=1. \tag{16}$$

证明　设 $K(A)=(S_A(u_1),S_A(u_2),\ldots,S_A(u_{|U|}))$, 则

$$E(A)=\sum_{i=1}^{|U|}\frac{1}{|U|}\left(1-\frac{|S_A(u_i)|}{|U|}\right)=\sum_{i=1}^{|U|}\frac{1}{|U|}-\sum_{i=1}^{|U|}\frac{|S_A(u_i)|}{|U|^2}=1-\mathrm{GK}(A).$$

即 $E(A)+\mathrm{GK}(A)=1$. 证毕.

定理 21　设 $S=(U,A)$ 是一个非完备信息系统, $K(A)=(S_A(u_1),S_A(u_2),\ldots,S_A(u_{|U|}))$, 则 A 的信息熵 $H(A)$ 与粗糙熵 $E_r(A)$ 具有下面的关系:

$$H(A)+E_r(A)=\log_2|U|. \tag{17}$$

证明　由定义 7 和 9, 可以得到

$$H(A)=-\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{|S_A(u_i)|}{|U|}$$

$$
\begin{aligned}
&= -\sum_{i=1}^{|U|} \frac{1}{|U|} (\log_2 |S_A(u_i)| - \log_2 |U|) \\
&= -\left(-\sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 \frac{1}{|S_A(u_i)|} \right) + \log_2 |U| \sum_{i=1}^{|U|} \frac{1}{|U|} \\
&= -E_r(A) + \log_2 |U|.
\end{aligned}
$$

即 $H(A) + E_r(A) = \log_2 |U|$. 证毕.

定理22 设 $S = (U, A)$ 是一个非完备信息系统, $K(A) = (S_A(u_1), S_A(u_2), \ldots, S_A(u_{|U|}))$, 则 A 的组合熵和组合粒度之间的关系为

$$
\mathrm{CE}(A) + \mathrm{CG}(A) = 1. \tag{18}
$$

证明 根据它们的定义, 可以得到

$$
\mathrm{CE}(A) = \frac{1}{|U|} \sum_{i=1}^{|U|} \frac{C_{|U|}^2 - C_{|S_A(u_i)|}^2}{C_{|U|}^2} = 1 - \sum_{i=1}^{|U|} \frac{C_{|S_A(u_i)|}^2}{C_{|U|}^2} = 1 - \mathrm{CG}(A).
$$

显然, $\mathrm{CE}(A) + \mathrm{CG}(A) = 1$. 证毕.

5 结论

信息粒度与熵理论是两种有效进行信息系统中不确定性研究的重要工具, 已有许多成功的应用范例. 本文分析了完备信息系统、非完备信息系统中信息粒的刻画、表示方法, 提出了信息系统中信息粒度的公理化方法, 指出了一些已有的信息粒度都是其特殊形式. 对我们原有的熵理论结果作了进一步的扩充, 性质表明信息系统中的熵度量满足粒化单调性, 建立了信息粒度与熵之间的互补关系. 这些结论统一了完备信息系统与非完备信息系统中不确定性度量的相关结果. 本文研究的信息系统是不包含模糊信息的. 下一步, 我们将对包含模糊信息的模糊信息系统进行模糊信息粒度公理化研究.

参考文献

1 Zadeh L A. Fuzzy logic=computing with words. IEEE T Fuzzy Syst, 1996, 4(1): 103—111

2 Zadeh L A. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Set Syst, 1997, 19(1): 111—127

3 Zadeh L A. Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent system. Soft Comput, 1998, 2(1): 23—25

4 Zadeh L A. Fuzzy sets and information granularity. In: Gupta N, Ragade R, Yager R, eds. Advances in Fuzzy Set Theory and Application. Amsterdam: North-Holland, 1979. 3—18

5 Zhang W X, Wei L, Qi J J. Attribute reduction theory and approach to concept lattice. Sci China Ser F-Inf Sci, 2005, 48(6): 713—726

6 Zhang W X, Qiu G F, Wu W Z. A general approach to attribute reduction in rough set theory. Sci China Ser F-Inf Sci, 2007, 50(2): 188—197

7 Pawlak Z. Granularity of knowledge, indiscernibility and rough sets. In: Proceedings of 1998 IEEE International Conference on Fuzzy Systems. Heideberg: Physica-Verlag, 1998. 106—110

8 Lin T Y. Granular computing on binary relations I: data mining and neighborhood systems, II: rough sets representations and belief functions. In: Polkowski L, Skowron A, eds. Rough Sets in Knowledge Discovery 1. Heidelberg: Physica-Verlag, 1998. 107—140

9 Klir G J. Basic issues of computing with granular probabilities. In: Proceedings of 1998 IEEE International Conference on Fuzzy Systems. Heideberg: Physica-Verlag, 1998. 101—105

10 Liang J Y, Qian Y H, Chu C Y, et al. Rough set approximation based on dynamic granulation. Lect Notes Artif Intel, 2005, 3641: 701—708

11 张讲社, 梁怡, 徐宗本. 基于视觉系统的聚类算法. 计算机学报, 2001, 24(5): 496—501

12 史忠植. 高级人工智能. 北京: 科学出版社, 2006. 1—40

13 Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning about Data. Dordrecht: Kluwer Academic Publishers, 1991. 42—140

14 张钹, 张铃. 问题求解理论及应用. 北京: 清华大学出版社, 1990. 52—73

15 张铃, 张钹. 模糊商空间理论(模糊粒度计算方法). 软件学报, 2003, 14(4): 770—776

16 Zhang L, Zhang B. Fuzzy reasoning model under quotient space structure. Inform Sciences, 2005, 173: 353—364

17 Wierman M J. Measuring uncertainty in rough set theory. Int J Gen Syst, 1999, 28(4): 283—297

18 Liang J Y, Shi Z Z. The information entropy, rough entropy and knowledge granulation in rough set theory. Int J Uncertain Fuzz, 2004, 12(1): 37—46

19 Liang J Y, Xu Z B. The algorithm on knowledge reduction in incomplete information systems. Int J Uncertain Fuzz, 2002, 24(1): 95—103

20 梁吉业, 李德玉. 信息系统中的不确定性与知识获取. 北京: 科学出版社, 2005. 13—113

21 Qian Y H, Liang J Y. Combination entropy and combination granulation in incomplete information system. Lect Notes Artif Intel, 2006, 4062: 184—190

22 Shannon C E. The mathematical theory of communication. Bell Syst Technol J, 1948, 27(3, 4): 373—423, 623—656

23 Düntsch I, Gediga G. Uncertainty measures of rough set prediction. Artif Intel, 1998, 106: 283—297

24 Beaubouef T, Petry F E. Fuzzy rough set techniques for uncertainty processing in a relational database. Int J Intel Syst, 2000, 15(5): 389—424

25 Beaubouef T, Petry F E, Arora G. Information-theoretic measures of uncertainty for rough sets and rough relational databases. Inform Sciences, 1998, 109(1-4): 185—195

26 de Luca A, Termini S. A definition of nonprobilistic entropy in the setting of fuzzy sets theory. Inform Contr, 1972, 20: 301—312

27 Mi J S, Leung Y, Wu W Z. An uncertainty measure in partition-based fuzzy rough sets. Int J Gen Syst, 2005, 34(1): 77—90

28 Liang J Y, Chin K S, Dang C Y, et al. A new method for measuring uncertainty and fuzziness in rough set theory. Int J Gen Syst, 2002, 31(4): 331—342

29 Kryszkiewicz M. Rough set approach to incomplete information systems. Inform Sciences, 1998, 112: 39—49