第18卷 第6期
2008年6月
计算机技术与发展
COMPUTER TECHNOLOGY AND DEVELOPMENT
Vol.18 No.6
Jun. 2008

信息熵在粗糙集中衍生的几个概念

纪 滨
(安徽工业大学 计算机学院,安徽 马鞍山 243002)

摘 要:随着对粗糙集理论研究的的深入,基于信息论的信息熵陆续被引入到粗糙集研究中,陆续产生了一些如条件熵、联合熵、知识熵、决策熵、知识粗糙熵、粗集粗糙熵等新的概念,尽管丰富了粗糙集理论和应用,但使用中存在语义不统一的地方,甚至缺乏必要的说明和证明。对这些有价值的新概念作了系统的、严格的、规范的定义及阐述,给出了它们的公式表示,同时,通过相关熵的运算揭示彼此间的关系,最后指出这些熵的应用范畴,以便研究人员在清楚概念的基础上作进一步研究。

关键词:粗糙集;信息熵;知识熵;决策熵;粗糙熵
中图分类号:TP391.72 **文献标识码**:A **文章编号**:1673-629X(2008)06-0073-03

Several Concepts from Information Entropy in Rough Sets

JI Bin
(School of Computer Science, Anhui University of Technology, Maanshan 243002, China)

Abstract: With the development of rough sets theory, entropy based on information theory has been applying to the studying. Some created concepts such as condition entropy, joint entropy, knowledge entropy, decision entropy, knowledge cough entropy and cough-sets cough entropy enrich rough sets' theory, however, the denotations are not clear, and even lacking of the necessary explanations and proofs. These valuable new concepts are explained scientifically and the mutual relations of them are revealed through operating in this paper. Based on it, researchers work conveniently for further studying.

Key words: rough sets; information entropy; knowledge entropy; decision entropy; rough entropy

0 引 言

文中假设读者已经熟悉经典粗糙集理论中的一些基本概念,将不再介绍,如有需要,可以参照文献[1]。传统的粗糙集理论是以代数方法定义的,它将域看成一个集合,而将知识看成是在域上构造分类的能力,即对集合进行构造划分(partition)的能力。信息熵理论是信息论的一个重要内容,它可以量化地对信息量进行分析,知识是从信息中提取出来的,因此它是一种特殊的信息,由此从熵理论研究粗糙集具有理论上的可行性。

由于信息熵与粗糙集的研究对象都是信息系统的不确定性,从文献[2,3]起,已有不少作者将信息熵引用到文章中,陆续加以引申推广产生了一些新的概念,丰富了粗糙集理论和应用。然而,遗憾的是不同作者在使用中存在同名异意的地方,甚至缺乏必要的说明和证明。这或多或少影响了读者对文章的阅读和联想。文中旨在对这些有价值的概念作一系统规范表达及阐述,以便研究人员在清楚概念的基础上方便阅读有关文章,作进一步研究。

文中首先从信息熵出发,分别对条件熵、联合熵、知识熵、决策熵、知识粗糙熵、粗集粗糙熵、互信息进行定义。

1 几个熵的概念

1.1 信息熵

信息熵是1948年香农(Shannon)在论文“通信的数学理论”中引入的,解决了对信息的量化度量问题。他对信息的定义:事物运动状态或存在方式的不确定性的描述。

离散信源的模型:

定义1[4]:设某一概率系统 X 中有 n 个事件(X_1, $X_2, \cdots, X_i, \cdots, X_n$),第 i 个事件 X_i 产生的概率为 p_i($i = 1,2,3,\cdots,n$),当事件 X_i 产生后,给出的信息量称为自信息:$I(X_i) = \log_2 \frac{1}{p_i}$,单位为 bit。自信息的数学

收稿日期:2007-09-01
基金项目:安徽省自然科学基金项目(KJ2007A051)
作者简介:纪 滨(1970-),男,江苏镇江人,讲师,硕士,从事粗糙集、数据挖掘的研究。

期望即平均自信息量，它的值称为信息熵，简记为 $H(X)$：

$$H(X)=E[I(X)]=-\sum_{i=1}^{n}p_i\log_2 p_i \tag{1}$$

引入粗糙集理论中，设 $S=(U,A)$ 是一个信息系统，$P\subseteq A$，令 $U/P=\{X_1,X_2,\cdots,X_n\}$。设 U 为一论域，P 为 U 上的一个等价关系。式(1)代表知识 P 的信息熵，表达为[5]：

$$H(P)=-\sum_{i=1}^{n}\frac{|X_i|}{|U|}\log_2\frac{|X_i|}{|U|} \tag{2}$$

推论 1：信息熵的大小可以用来描述信息系统的平均不确定程度。若某一信息系统中某一知识产生的概率为 1，其他事件产生的概率为 0，由式(1)计算后可知，该系统的信息熵 $H=0$，它就是一个确定系统，不确定度为 0。

推论 2：如果某一信息系统中，其等价类是均匀的，则表示系统中每一知识产生的分类基数相等，该系统的信息熵具有最大值(在相同对象数的情况下)，即该系统的不确定性最大。根据信息熵的定义可知熵值越大，不确定性就越大。那么，搞清楚它所需要的信息量也就越大。

1.2 条件熵

定义 2：在信源 X 输出 X_i 的条件下，信源 Y 再输出 Y_j 所能提供的平均信息量[4]，称为条件熵。记为 $H(Y|X)$。

$$H(Y|X)=E[I(Y|X)]=\sum_{i=1}^{n}p_iH(Y|X=X_i)$$
$$=-\sum_{i=1}^{n}p_i\sum_{j=1}^{m}p_{ij}\log_2 p_{ij} \tag{3}$$

粗糙集中的表达[2]：

$$H(Q|P)=-\sum_{i=1}^{n}\frac{|X_i|}{|U|}\sum_{j=1}^{m}\frac{|X_i\cap Y_j|}{|X_i|}\log_2\frac{|X_i\cap Y_j|}{|X_i|} \tag{4}$$

1.3 联合熵

定义 3[4]：两个互相关联的信源 X 和 Y 的联合信源的信息熵为信源 X 的熵加上在 X 已知条件下信源 Y 的条件熵，称为 XY 的联合熵[4]。符合记为 $H(XY)$ 或 $H(X\cup Y)$。

公式表达：$H(X\cup Y)=H(X)+H(Y|X)$

粗糙集中的表达[2]：

$$H(P\cup Q)=H(P)+H(Q|P) \tag{5}$$

1.4 知识熵

如果 X 是一个有限集合，则其不确定的 Hartley[6] 度量定义为：$H_0(X)=\log_2|X|$，表示确定 $|X|$ 个等概率事件中的一个出现时提供的信息。设 $S=(U,A)$ 是一个信息系统，$P\subseteq A$，令 $U/P=\{X_1,X_2,\cdots,X_n\}$。设 U 为一论域，P 为 U 上的一个等价关系。知识熵度量了信源提供的平均信息量的大小。

根据全概公式有：

$$\sum_{i=1}^{n}\frac{|X_i|}{|U|}=1 \tag{6}$$

$$H(P)=-\sum_{i=1}^{n}\frac{|X_i|}{|U|}\log_2\frac{|X_i|}{|U|}=\sum_{i=1}^{n}\frac{|X_i|}{|U|}\log_2|U|-\sum_{i=1}^{n}\frac{|X_i|}{|U|}\log_2|X_i|=\log_2|U|-\sum_{i=1}^{n}\frac{|X_i|}{|U|}\log_2|X_i|=H_0(U)-\sum_{i=1}^{n}\frac{|X_i|}{|U|}H_0(X_i) \tag{7}$$

定义 4：设 $S=(U,A)$ 是一个信息系统，$P\subseteq A$。P 的知识熵是集合 U 的 Hartley 度量减去知识 P 中等价类的权 Hartley 度量之和。

推论 3：知识熵和信息熵相等，知识熵是信息熵的 Hartley 度量表达形式。

推论 4：$H(P)$ 取最大值为 $\log_2|U|$。

推论 5：$H(P)$ 取最小值 0。即 U 中所有元在 P 下均是不可分辨的。

1.5 决策熵

定义 5[7]：设 $S=(U,A)$ 是一个信息系统，$P,Q\subseteq A$。属性集合 $P\cup Q$ 的信息熵为 $H(P\cup Q)$，而 P 相对于 Q 的条件熵为 $H(P|Q)$，则属性集合 $P\cup Q$ 决策熵 $H(P\to Q)$ 定义为信息熵与条件熵的和，即：

$$H(P\to Q)=H(P\cup Q)+H(P|Q) \tag{8}$$

1.6 知识粗糙熵

关于粗糙熵的定义比较杂乱，有的定义不甚明了。现根据信息论原理及定义 4，给出一种简明形式。集合 U 的 Hartley 度量减去某个知识的信息熵，就得到该知识的粗糙熵，记为 $H_r(X)$。

定义 6：设 $S=(U,A)$ 是一个信息系统，$X\subseteq U$，$P\subseteq A$，关于知识 P 的粗糙熵定义为划分中元素的权 Hartley 度量之和，描述了知识 P 的不精确(粗糙)程度：

$$H_r(P)=\sum_{i=1}^{n}\frac{|X_i|}{|U|}H_0(X_i) \tag{9}$$

上面的知识粗糙熵是利用等价关系对论域的划分进行定义，如果将等价关系放宽为相容关系或相似关系后的一般二元关系，可以利用论域中任一对象在一般二元关系构成的所有对象的邻域中出现的次数来定义知识的粗糙熵。

定义 7：对信息系统 $S=(U,A)$，$U=\{X_1,X_2,\cdots,X_{|U|}\}$，$P\subseteq A$，$nP(X_i)$ 表示在知识 P 下 X_i 的一般二元关系邻域，且 $X_i\in nP(X_i)$，$i=1,2,\cdots,|U|$，$|X_i|_P=|\{(nP(X_j)|X_i\in nP(X_j),1\leqslant j\leqslant|U|\}|$

|,即 X_i 在所有元 $X_j(1 \leqslant j \leqslant |U|)$ 的邻域中出现的次数,则知识 P 的粗糙熵定义为:

$$H_r(nP) = \frac{1}{|U|}\sum_{i=1}^{|U|}\log_2 |X_i|_P \tag{10}$$

若 nP 是在 P 下的等价关系,则 $H_R(nP) = H_R(P)$。证明参见文献[7]。

1.7 粗集粗糙熵

粗糙集的粗糙性可以用粗糙度来度量。

定义 8[1]:设信息系统 $S = (U,A)$,$P \subseteq A$,$X \subseteq U$ 在知识 P 下的粗糙度定义如下:

$\rho_P(X) = 1 - \frac{|P_-(X)|}{|P^-(X)|}$,其中 $P_-(X)$,$P^-(X)$ 分别表示在一般二元关系下关于知识 P 的下、上近似集。

定义 9:设信息系统 $S = (U,A)$,$P \subseteq A$,$X \subseteq U$ 在知识 P 下的粗集粗糙熵定义如下:

$$H'_r(P) = \rho_P(X)H_r(P) \tag{11}$$

由上述定义,粗集的粗糙熵不仅与粗集本身的粗糙度有关,还与论域的知识的不确定性(知识的粗糙熵)有关,是一种比粗糙度 ρ_P 更精确的粗集不确定性度量。

1.8 互信息

定义 10:设 $S = (U,A)$ 是一个信息系统,$P,Q \subseteq A$。知识 P,Q 之间的信息量等于二者的自信息数学期望之差。称作 P,Q 的(平均)互信息,记为 $I(P,Q)$。

$$I(P,Q) = H(P) - H(Q \mid P) = H(Q) - H(P \mid Q) \tag{12}$$

性质[4]:A、$I(P,Q)$ 总是大于 0;B、$I(P,Q) = I(Q,P)$。

2 相关熵的运算

根据定义 4:$H(P) + H_r(P) = H_0(U)$　　(13)

根据式(4),(6) 和(2):

$$H(Q \mid P) = -\sum_{i=1}^{n}\frac{|X_i|}{|U|}\sum_{j=1}^{m}\frac{|X_i \cap Y_j|}{|X_i|}\left(\log_2\frac{|X_i \cap Y_j|}{|U|} - \log_2\frac{|X_i|}{|U|}\right) =$$
$$\sum_{i=1}^{n}\frac{|X_i|}{|U|}\sum_{j=1}^{m}\frac{|X_i \cap| Y_j|}{|X_i|}\log_2\frac{|X_i|}{|U|} -$$
$$\sum_{i=1}^{n}\frac{|X_i|}{|U|}\sum_{j=1}^{m}\frac{|X_i \cap Y_j|}{|X_i|}\log_2\frac{|X_i \cap Y_j|}{|U|} =$$
$$\sum_{i=1}^{n}\frac{|X_i|}{|U|}\log_2\frac{|X_i|}{|U|}\sum_{j=1}^{m}\frac{|X_i \cap Y_j|}{|X_i|} -$$
$$\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{|X_i|}{|U|}\times\frac{|X_i \cap Y_j|}{|X_i|}\log_2\frac{|X_i \cap Y_i|}{|U|} =$$
$$\sum_{i=1}^{n}\frac{|X_i|}{|U|}\log_2\frac{|X_i|}{|U|} -$$
$$\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{|X_i \cap Y_j|}{|U|}\log_2\frac{|X_i \cap Y_j|}{|U|} =$$
$$-H(P) - \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{|X_i \cap Y_j|}{|U|}\log_2\frac{|X_i \cap Y_j|}{|U|} \tag{14}$$

根据式(5) 和(14):

$$H(P \cup Q) = -\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{|X_i \cap Y_j|}{|U|}\log_2\frac{|X_i \cap Y_j|}{|U|} \tag{15}$$

根据式(5) 和(12):

$$I(P,Q) = H(P) + H(Q) - H(P \cup Q) \tag{16}$$

3 结束语

信息熵作为一种衡量信息量的重要工具,将其引入粗糙集作为衡量属性重要性的标志,考虑该属性对于论域中不确定分类子集的影响,使属性重要性这一概念更加完善。知识熵是信息熵的另一种表达形式。

条件熵和互信息可以作为度量属性关联的基本度量,条件熵度量了两个属性集之间的一种方式关联或函数依赖的程度;互信息可以用来评价两个属性集之间两种方式关联的程度。而联合熵是两个属性集作为一个整体的信息量的度量。

决策熵能够在不相容决策表的属性约简中保持"决策能力"[8]不变,而条件熵不能。

知识粗糙熵度量的是知识的粗糙程度。

粗集粗糙熵是知识粗糙熵和粗集粗糙度的结合,比粗糙度更精确。

参考文献:

[1] 王国胤. Rough 集理论与知识获取[M]. 西安:西安交通大学出版社,2001.

[2] Pawlak Z. Rough sets: probabilistic versus deterministic approach[J]. International Journal of Man - Machine Studies, 1998,29:81 - 95.

[3] Gediga G. Uncertainty measures of rough set prediction[J]. Artificial Intelligence,1998,106:109 - 137.

[4] 傅祖芸. 信息论基础理论与应用[M]. 北京:电子工业出版社,2001.

[5] 苗夺谦,王　珏. 粗糙集理论中概念与运算的信息表示[J]. 软件学报,1999,10(2):113 - 116.

[6] Hartley R V L. Transmission of information[J]. The Bell Systems Technical Journal,1928,7(3):535 - 563.

[7] 黄　兵,周献中,史迎春. 基于一般二元关系的知识粗糙熵与粗集粗糙熵[J]. 系统工程理论与实践,2004(1):93 - 95.

[8] 蒋思宇,卢炎生. 两种新的决策表属性约简概念[J]. 小型微型计算机系统,2006,27(3):512 - 516.