

# 知识粗糙性与知识粒度的关系研究

李 鸿

(宿州学院,安徽 宿州 234000)

**摘 要:**提出了知识的粒数、籽数和粒度的概念,证明了粗糙集理论中知识粗糙性与其所对应的粒度之间的单调关系,从而揭示了知识粗糙性与其所对应的知识的粒数、籽数和粒度之间的密切关系。知识粒度的概念从物理意义上反映了知识库中的知识颗粒状结构的本质。

**关键词:**粗糙集理论;知识粗糙性;知识粒度;知识库

**中图分类号:**TP18

**文献标识码:**A

**文章编号:**1673-629X(2007)08-0117-03

## Research on Relation of Knowledge Roughness and Knowledge Granularity

LI Hong

(Suzhou College, Suzhou 234000, China)

**Abstract:** Proposed the concept of granule amount of knowledge, granule seed and granularity, proved its dull relation between knowledge roughness and its correspondent granularity in rough set theory, declared that knowledge roughness is closely related to granule amount of knowledge, granule seed of knowledge and granularity of knowledge. The concept of knowledge granularity in a sense of physics reflects the nature of granular structure of knowledge in knowledge base.

**Key words:** rough set theory; knowledge roughness; granularity of knowledge; knowledge base

### 0 引言

粗糙集理论<sup>[1,2]</sup>是一种新的处理模糊和不确定知识的数学工具。粗糙集理论从新的视角对知识进行了定义,把知识看作是对对象进行分类的能力,即是关于论域的划分。粗糙集理论认为知识是有粒度的<sup>[1]</sup>,即知识是粗糙的。知识粒度越大,其越粗糙,知识含量越少。并认为知识的不确定性是由于知识粒度太大引起的,知识的粗糙性越大,则其不确定性也越大。文中提出了知识的粒数、籽数和粒度的概念,证明了粗糙集理论中知识粗糙性与其所对应的粒度之间的单调关系,从而揭示了知识粗糙性与其所对应的知识的粒数、籽数和粒度之间的密切关系。知识的粒度的概念从物理意义上反映了知识库中的知识颗粒状结构的本质。

### 1 知识粗糙性

在粗糙集理论中,由于近似空间 $(U, R)$ 中需要对对象进行分类,而把知识看作是关于论域的划分,因此

熵也可用于描述近似空间 $(U, R)$ 中对象的分类情况。近似空间 $(U, R)$ 中基于 Shannon 信息熵的知识不确定性描述可见文献[2,3]。

粗糙集理论认为知识是具有粒度的(granularity),即知识是粗糙的。知识的粒度越大,认为其越粗糙。具体说来,知识的粗糙性部分地来自  $U$  上的不可区分关系,即等价关系。它将  $U$  划分成一些等价类的集合,如果每个等价类包含的对象个数越少,则划分的粗糙性越低,即知识的粒度越低;相反,如果划分的等价类个数越少,即每个等价类包含的对象个数越多,则划分的粗糙性越高,即知识的粒度越大,这时能获得关于  $U$  中对象的知识或信息量就越少,当  $U$  中的所有对象归为一类时,能得到关于  $U$  中对象的信息量降为零。

从不可分辨关系的概念可知,对信息系统  $IS = (U, R)$ ,  $P \subseteq R$  是属性集合的一个子集,不可分辨关系  $IND(P)$  揭示出论域知识的颗粒状结构。并且不可分辨关系是一种等价关系,通过等价关系,可得到决策系统的一个划分,用  $U/IND(P)$  或  $U/P$  来表示。这个划分,称为关于  $U$  的一个知识库,表示为 $(U, P)$ 。

设  $K = (U, R)$  为一个知识库,  $P, Q \subseteq R$  为  $U$  上的两个不可分辨关系。如果  $P \subset Q$ , 则有  $U/IND(P) \supset U/IND(Q)$  成立,记作  $P > Q$ ; 如果  $P \supset Q$ , 则有

收稿日期:2006-10-15

基金项目:安徽高校省级自然科学研究重点项目(kj2007a129zc);宿州学院教授(博士)启动基金(2006jb04)

作者简介:李 鸿(1965-),男,安徽灵璧人,教授,CCF 会员,研究方向为粒计算、KDD 等。



$U/\text{IND}(P) \subset U/\text{IND}(Q)$  成立, 记作  $P < Q$ ; 如果  $P = Q$ , 则有  $U/\text{IND}(P) = U/\text{IND}(Q)$  成立, 仍记作  $P = Q$ 。知识之间的这种关系可以看作是知识库  $K$  上的偏序关系。

设  $K_p = (U, P)$  和  $K_q = (U, Q)$  是两个知识库。如果  $U/\text{IND}(P) \subset U/\text{IND}(Q)$ , 即对任意的  $A \in U/\text{IND}(P)$ , 总存在  $B \in U/\text{IND}(Q)$ , 使得  $A \subset B$  成立, 则知识  $P$  比知识  $Q$  的粗糙性小, 称  $P$  比  $Q$  较细 (finer), 或  $Q$  比  $P$  较粗 (coarser), 记作  $P < Q$ ; 反之, 如果  $U/\text{IND}(P) \supset U/\text{IND}(Q)$ , 则  $P > Q$ 。如果  $U/\text{IND}(P) = U/\text{IND}(Q)$ , 即对于任意的  $A \in U/\text{IND}(P)$ , 总存在  $B \in U/\text{IND}(Q)$ , 使得  $A = B$  成立, 则称知识  $P$  与知识  $Q$  是相等的, 记作  $P \equiv Q$ 。知识之间的这种关系可以看作是知识库之间的偏序关系。因此, 粗糙集理论通过不可分辨关系和集合包含关系定义了知识粗糙性, 其本质是知识之间的依赖关系。

## 2 知识的粒度定义

粗糙集方法处处以知识的颗粒状为其主要特征处理知识, 试图从数据的结构方面着手, 挖掘尽可能多的知识。它对世界的理解, 只是达到了等价关系的一个一个的等价类这种颗粒状的程度, 而对颗粒内的对象是无法分辨的。在粗糙集理论中, 拥有知识  $R$  的智能体 (人, 机器人等) 不能将  $[u]_R$  中的对象与  $u$  分辨。因此,  $[u]_R$  表达了智能体利用知识库中的一部分知识  $R$  所能达到的最高的认知程度, 智能体只能分辨、表达不同颗粒状的等价类。而  $\text{IND}(R)$  表达了知识库  $K = (U, R)$  的最高的分辨、表示程度。

为描述知识的颗粒状, 文献[4]给出了一种知识的粒度定义, 文中从另一角度给出知识粒度定义。

**定义 1** 知识的粒数 (Granule Amount of Knowledge): 设  $U$  为有限对象的集合, 称为论域,  $R$  为  $U$  上的一个不可分辨关系,  $R$  在  $U$  上导出的划分为  $U/\text{IND}(R) = \{R_1, R_2, \dots, R_n\}$ , 称  $R$  在  $U$  上导出的划分中的等价类的个数为知识的粒数, 记为  $\text{GA}(R)$ 。其中的一个一个的等价类称为知识的颗粒。

由定义有:  $\text{GA}(R) = |U/\text{IND}(R)|$ ,  $|U/\text{IND}(R)|$  表示  $R$  在  $U$  上导出的划分中的等价类的个数。

**定义 2** 知识颗粒的籽数 (Granule Seed of Knowledge): 设  $U$  为有限对象的集合, 称为论域,  $R$  为  $U$  上的一个不可分辨关系,  $R$  在  $U$  上导出的划分为  $U/\text{IND}(R) = \{R_1, R_2, \dots, R_n\}$ , 称  $R$  在  $U$  上导出的划分中的每个等价类中的不可分辨对象的个数为知识颗粒的籽数, 记为  $\text{GS}(R_i) (i = 1, 2, \dots, n)$ 。

由定义有:  $\text{GS}(R_i) = |R_i|, |R_i| (i = 1, 2, \dots,$

$n)$  表示  $R$  在  $U$  上导出的划分中的第  $i$  个等价类中的不可分辨对象的个数。由此可知,  $1 \leq \text{GS}(R_i) \leq |U|$ 。

为了更好地描述知识的粗糙性, 考虑一个知识中的平均颗粒籽数 (Average Granule Seed), 记为  $\text{AGS}(R)$ 。 $\text{AGS}(R) = |U| / \text{GA}(R) = |U| / |U/\text{IND}(R)|$ 。用知识的平均颗粒籽数来定义知识的粒度。

由于  $|U| = \sum_{i=1}^n |R_i|$ , 所以有下述命题:

**命题 1** 设  $U$  为有限对象的集合, 称为论域,  $R$  为  $U$  上的一个不可分辨关系,  $R$  在  $U$  上导出的划分为  $U/\text{IND}(R) = \{R_1, R_2, \dots, R_n\}$ , 则

$$\text{AGS}(R) = \sum_{i=1}^n |R_i| / \text{GA}(R) = \sum_{i=1}^n |R_i| / |U/\text{IND}(R)|$$

**定义 3** 知识的粒度 (Granularity of Knowledge): 设  $U$  为有限对象的集合, 称为论域,  $R$  为  $U$  上的一个不可分辨关系,  $R$  在  $U$  上导出的划分为  $U/\text{IND}(R) = \{R_1, R_2, \dots, R_n\}$ , 称知识  $R$  的平均颗粒籽数占  $U$  中对象总数之比为知识的粒度, 记为  $\text{GD}(R)$ 。即:

$$\text{GD}(R) = \text{AGS}(R) / |U| = 1 / \text{GA}(R) = 1 / |U/\text{IND}(R)|$$

规定: 当  $R = \emptyset$  时,  $\text{GD}(\emptyset) = 1$ 。

**命题 2**  $1/|U| \leq \text{GD}(R) \leq 1$ 。当  $R$  为  $U$  上的相等关系时, 知识  $R$  的粒度达到最小值为  $1/|U|$ ; 当  $R$  为  $U$  上的论域关系时, 知识  $R$  的粒度达到最大值为 1。

由知识的粒度定义可以看出, 知识的粒度越小,  $|U/\text{IND}(R)|$  越大, 即  $U$  中对象分类越多, 其分辨能力越强; 反之, 知识的粒度越大,  $|U/\text{IND}(R)|$  越小, 即  $U$  中对象分类越少, 其分辨能力越弱。因此, 知识的粒度可以表示知识的分辨能力,  $\text{GD}(R)$  越小, 分辨能力越强; 否则, 分辨能力越弱。当  $uRv$  时, 表明对象  $u, v$  在  $R$  下不可分辨, 属于  $R$  的同一个等价类; 否则, 它们可分辨, 属于  $R$  不同等价类。因此,  $\text{GD}(R)$  表示在  $U$  中随机选择两个对象, 这两个对象关于  $R$  不可分辨的可能性大小。可能性越大, 即  $\text{GD}(R)$  越大, 表明  $R$  的分辨能力越弱; 否则,  $R$  的分辨能力越强。

由定义 3 可以定义知识  $R$  的分辨度。

**定义 4** 知识的分辨度 (Distinct Degree of Knowledge): 设  $U$  为有限对象的集合, 称为论域,  $R$  为  $U$  上的一个不可分辨关系,  $R$  在  $U$  上导出的划分为  $U/\text{IND}(R) = \{R_1, R_2, \dots, R_n\}$ ,  $\text{GD}(R)$  为知识  $R$  的粒度, 定义知识  $R$  的分辨度  $\text{DD}(R)$  为

$$\text{DD}(R) = 1 - \text{GD}(R)$$

由命题 2 有如下的命题 3:



命题3  $0 \leq DD(R) \leq 1 - 1/|U|$ 。当  $R$  为  $U$  上的相等关系时,知识  $R$  的分辨率达到最大值为  $1 - 1/|U|$ ;当  $R$  为  $U$  上的论域关系时,知识  $R$  的分辨率达到最小值为 0。

### 3 知识粗糙性与知识粒度的关系

为了研究基于知识粒度的知识粗糙性的性质,引入下面的引理:

引理1 设  $U$  为一论域,  $P, Q$  为  $U$  上的两个等价关系集合,则有

$U/IND(P \cup Q) = U/IND(P) \cap U/IND(Q)$  成立。证明:参见文献[5]。

定理1 设  $U$  为一论域,  $P, Q$  为  $U$  上的两个等价关系集合,如果  $P < Q$ ,则  $GD(P) > GD(Q)$ ;如果  $P \equiv Q$ ,则  $GD(P) = GD(Q)$ ;如果  $P > Q$ ,则  $GD(P) < GD(Q)$ 。

证明:

由  $P < Q$ ,则  $P \supset Q$  和  $U/IND(P) \subset U/IND(Q)$  成立,  $|U/IND(P)| > |U/IND(Q)|$ ,而  $GD(P) = 1/|U/IND(P)|$ ,  $GD(Q) = 1/|U/IND(Q)|$ ,从而有  $GD(P) < GD(Q)$ 。

同理有,当  $P \equiv Q$  时,  $GD(P) = GD(Q)$ ;当  $P > Q$  时,  $GD(P) > GD(Q)$ 。

推论1 设  $U$  为一个论域,  $K_p = (U, P)$  和  $K_q = (U, Q)$  是关于  $U$  的两个知识库,如果  $P < Q$ ,则  $GD(P) < GD(Q)$ ;如果  $P \equiv Q$ ,则  $GD(P) = GD(Q)$ ;如果  $P > Q$ ,则  $GD(P) > GD(Q)$ 。

定理2 设  $U$  为一论域,  $P, Q$  为  $U$  上的两个等价关系集合,如果  $P < Q$ ,则  $DD(P) > DD(Q)$ ;如果  $P \equiv Q$ ,则  $DD(P) = DD(Q)$ ;如果  $P > Q$ ,则  $DD(P) < DD(Q)$ 。

证明过程与定理1相似,略。

推论2 设  $U$  为一个论域,  $K_p = (U, P)$  和  $K_q = (U, Q)$  是关于  $U$  的两个知识库,如果  $P < Q$ ,则  $DD(P) > DD(Q)$ ;如果  $P \equiv Q$ ,则  $DD(P) = DD(Q)$ ;如果  $P > Q$ ,则  $DD(P) < DD(Q)$ 。

定理3 设  $U$  为一论域,  $P$  为  $U$  上的一个等价关系集合,且对任意的  $p \in P$ ,  $p$  在  $P$  中是可省略的充分必要条件  $GD(P) = GD(P - \{p\})$ 。

证明:

必要性。设  $p$  为  $P$  中冗余的,即  $U/IND(P - \{p\}) = U/IND(P)$ ,从而有  $|U/IND(P - \{p\})| = |U/IND(P)|$ ,因此,  $GD(P) = GD(P - \{p\})$ 。

充分性。用反证法。

假设  $p$  在  $P$  中是非冗余的,由定理1得:  $P - \{p\} \subset P$ ,  $GD(P) < GD(P - \{p\})$ ,这和条件  $GD(P) = GD(P - \{p\})$  相矛盾,因此,  $p$  在  $P$  中必定是冗余的。

当  $p$  在  $P$  中是可省略时,由引理1得  $U/IND(P) = U/IND(P - \{p\}) \cap U/IND(\{p\}) = U/IND(P - \{p\})$ ,则有  $U/IND(P - \{p\}) \subseteq U/IND(\{p\})$ ,即  $U/IND(P - \{p\})$  中的任一等价类都包含在  $U/IND(\{p\})$  中,这意味着将  $\{p\}$  加到  $P - \{p\}$  中,没有使  $U/IND(P)$  的粒度  $GD(P)$  降低。因此可知,知识库  $(U, R)$  的分类过程是该知识库的粒度不断降低的过程。

例1:设  $U = \{x_1, x_2, \dots, x_8\}$ ,  $U$  上的等价关系  $P, Q, R$  导出的划分分别为:

$U/IND(P) = \{\{x_1, x_4, x_5\}, \{x_2, x_8\}, \{x_3\}, \{x_6, x_7\}\}$ ,  $GD(P) = 1/4 = 0.25$ ;

$U/IND(Q) = \{\{x_1, x_3, x_5\}, \{x_6\}, \{x_2, x_4, x_7, x_8\}\}$ ,  $GD(Q) = 1/3$ ;

$U/IND(R) = \{\{x_1, x_5\}, \{x_2, x_7, x_8\}, \{x_6\}, \{x_3, x_4\}\}$ ,  $GD(R) = 1/4$ ;

设  $S = P \cup Q \cup R$ ,由引理1得  $U/IND(S) = U/IND(P) \cap U/IND(Q) \cap U/IND(R)$

$U/IND(S) = \{\{x_1, x_5\}, \{x_2, x_8\}, \{x_3\}, \{x_4\}, \{x_6\}, \{x_7\}\}$ ,  $GD(S) = 1/6$ ;

$U/IND(S - Q) = \{\{x_1, x_5\}, \{x_2, x_8\}, \{x_3\}, \{x_4\}, \{x_6\}, \{x_7\}\} = U/IND(S)$ ,  $GD(S - Q) = 1/6$ ;

$U/IND(S - R) = \{\{x_1, x_5\}, \{x_2, x_8\}, \{x_3\}, \{x_4\}, \{x_6\}, \{x_7\}\} = U/IND(S)$ ,  $GD(S - R) = 1/6$ ;

$U/IND(S - P) = \{\{x_1, x_5\}, \{x_2, x_7, x_8\}, \{x_3\}, \{x_4\}, \{x_6\}\} \neq U/IND(S)$ ,  $GD(S - P) = 1/5$ ;

因此,  $Q, R$  在  $S$  中是可省略的,各种划分的粒度关系为:

$GD(S) = GD(S - Q) = GD(S - R) < GD(S - P) < GD(P) = GD(R) < GD(Q)$

#### 参考文献:

- [1] Pawlak Z. Rough Sets—Theoretical Aspects of Reasoning about Data[M]. Dordrecht: Kluwer Academic Publisher, 1991.
- [2] 张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法[M]. 北京: 科学出版社, 2001.
- [3] 苗夺谦, 王 珏. 粗糙集理论中知识粗糙性与信息熵关系的讨论[J]. 模式识别与人工智能, 1998, 11(3): 34-40.
- [4] 苗夺谦, 范世栋. 知识的粒度计算及其应用[J]. 系统工程理论与实践, 2002, 22(1): 48-56.
- [5] Guan J, Bell D. Rough computational methods for information systems[J]. Artificial Intelligence, 1998, 105: 77-103.