知识系统中全粒度粗糙集及概念漂移的研究

邓大勇[1),2),3)] 卢克文[1)] 苗夺谦[3)] 黄厚宽[4)]

[1)](浙江师范大学数理与信息工程学院 浙江 金华 321004)

[2)](浙江师范大学行知学院 浙江 金华 321004)

[3)](同济大学电子与信息工程学院 上海 201804)

[4)](北京交通大学计算机与信息技术学院 北京 100044)

摘 要 概念漂移探测是数据流挖掘的一个研究重点,不确定性分析是粗糙集理论的研究核心之一.大数据、数据流中存在不确定变化和概念漂移现象,但是,除 F-粗糙集外,几乎所有的粗糙集模型都是静态模型或半动态模型,专注于各种不确定性研究,难以处理不确定性变化,也难以探测概念漂移.结合量子计算、数据流、概念漂移和粗糙集、F-粗糙集的基本观点,以上、下近似为工具,定义了知识系统中的全粒度粗糙集和上、下近似概念漂移,上、下近似概念耦合等概念,探讨了全粒度粗糙集的性质,分析了知识系统内概念的全局变化.全粒度粗糙集继承了 Pawlak 粗糙集和 F-粗糙集的基本思想,以上、下近似簇为工具表示了概念在知识系统内的各种可能变化.用嵌套哈斯图表示了概念不同情况下的同一性和差异性:同一层内的表示没有发生概念漂移,不同层内的表示发生了概念漂移.以正区域为工具,定义了决策表中的全粒度正区域和概念漂移、概念耦合等概念,探究了全粒度正区域的性质,分析了决策表内整体概念的全局变化.全粒度正区域表示了决策表中各种可能情况下的正区域,用嵌套哈斯图表示了正区域簇的同一性和差异性:同一层内没有发生相对于正区域的概念漂移,不同层内发生了相对于正区域的概念漂移.在全粒度粗糙集意义下,定义了全粒度绝对约简、全粒度值约简、全粒度 Pawlak 约简等属性约简,并探讨其性质.与大部分的属性约简不同(仅仅与并行约简和多粒度约简类似),全粒度属性约简要求概念的所有可能表示不发生概念漂移.进一步探讨了属性约简的优缺点,属性约简使得概念的表示变得单一,冗余属性的存在增加了概念表示的丰富性、多样性.在认识论方面,以粗糙集和粒计算为工具分析了人类认识世界的局部性与全局性,对人类认识世界的方式进行了进一步探讨.全粒度粗糙集在一定意义下能够表示人类认识的复杂性、不确定性、多样性、层次性和动态性,在量子计算的帮助下能够从一个粒度转跳到另一个粒度并且毫无困难.全粒度粗糙集的研究及其中的概念漂移探测为各种条件下的概念漂移探测和人类智能的模拟提供了有益的启示.

关键词 全粒度粗糙集;概念漂移;偏序关系;概念耦合;上、下近似

中图法分类号 TP18 **DOI 号** 10.11897/SP.J.1016.2019.00085

Study on Entire-Granulation Rough Sets and Concept Drifting in a Knowledge System

DENG Da-Yong[1),2),3)] LU Ke-Wen[1)] MIAO Duo-Qian[3)] HUANG Hou-Kuan[4)]

[1)](*College of Mathematics, Physics and Information Engineering, Zhejiang Normal University, Jinhua, Zhejiang* 321004)

[2)](*Xingzhi College, Zhejiang Normal University, Jinhua, Zhejiang* 321004)

[3)](*School of Electronics and Information Engineering, Tongji University, Shanghai* 201804)

[4)](*School of Computer and Information Technology, Beijing Jiaotong University, Beijing* 100044)

Abstract Concept drifting detection is one of the hot topics in data stream mining, and analysis of uncertainty is dominant in rough set theory. There exist the change of uncertainty and concept drifting in big data and data stream. However, except for F-rough sets, almost all of rough set

收稿日期:2016-03-01;**在线出版日期**:2016-11-29.**本课题得到国家自然科学基金项目**(61473030,61572442,61203247,61273304,61573259,61472166)**和浙江省自然科学基金项目**(LY15F020012)**资助**.**邓大勇**,男,1968 年生,博士,副教授,中国计算机学会(CCF)会员,主要研究方向为粗糙集理论及应用. E-mail: dayongd@163.com. **卢克文**,男,1992 年生,硕士研究生,主要研究方向为数据挖掘. **苗夺谦**,男,1964 年生,博士,教授,主要研究领域为数据挖掘、图像处理. **黄厚宽**,男,1940 年生,教授,主要研究领域为数据挖掘、人工智能.

models are static models or semi-dynamic models, which study on vagueness and uncertainty. It is hard for them to deal with the change of uncertainty, and to detect concept drifting. Combined with the ideas of quantum computing, data stream, concept drifting, rough sets and F-rough sets, a rough set model for entire granulations (called entire-granulation rough sets) is presented, and a lot of concepts, such as concept drifting of upper approximation, concept drifting of lower approximation, coupling of upper approximation and coupling of lower approximation, etc. are defined. The properties of entire-granulation rough sets are investigated, and the change of uncertainty for a concept in a knowledge system is analyzed with these definitions. Entire-granulation rough sets inherit the basic ideas of Pawlak rough sets and F-rough sets, which describe all of the changes of uncertainty for a concept with a family of upper approximations and lower approximations. Embedded Hasse diagram is employed to express the identity and diversity for a concept in different cases: There exists no concept drifting for the same level of concept expressions but exists concept drifting for the different levels of concept expressions. With the positive region, the positive region for entire granulations is defined, and concept drifting, concept coupling are defined in a decision system. The properties of entire-granulation positive region are discussed, and the analysis and measurement for the change of concept uncertainty are conducted. Entire-granulation positive region expresses all of the positive regions in various cases in a decision system. Embedded Hasse diagram is also employed to express the identity and diversity for the family of positive regions: There exists no concept drifting relative to positive region for the same level of concepts, but exists concept drifting relative to positive region for different levels of concepts. In entire-granulation rough sets, entire-granulation absolute reducts, entire-granulation value reducts and entire-granulation Pawlak reducts are defined, and their properties are investigated. Not like most types of attribute reducts (just like parallel reducts and mutil-granulation conditional attribute reducts), entire-granulation conditional attribute reducts ask for no concept drifting for all of concept expressions. The advantages and faults of conditional attribute reduction are further investigated: The unicity of concept expressions is done when condition attribute reduct is conducted, while the redundant conditional attributes can make concept expression more diversified. From the viewpoints of epistemology, the wholeness and locality of human thinking are further analyzed with granular computing and rough sets. To some extent, entire-granulation rough sets can express complexity, uncertainty, diversity, hierarchy and dynamic in the process of human cognition. With the help of quantum computing, the model of entire-granulation rough sets can transform one type of granulation to another fluently. The study on entire-granulation rough sets and concept drifting detection among them can provide heuristic information for various concept drifting detection and simulation of human intelligence.

Keywords entire-granulation rough sets; concept drifting; partial ordering relation; concept coupling; upper and lower approximation

1 引 言

生产和生活中的数据常常随时间推移发生变化,例如,及时通讯数据、监控数据、股票交易数据等,这种随时间变化的数据称为数据流[1].数据流在动态环境下按照时间顺序产生,具有变化速度快、海量、易发生概念漂移等特征[2-7].探测概念漂移,更好地对数据流进行分类(或聚类)是数据流挖掘的重要任务和研究方向.

人们在判定、处理问题、形成概念的时候往往根据部分信息或所掌握的信息.而这些信息往往是变化的、不确定的、甚至是错误的.比如:人们称赞一个女生漂亮,不同的人表达的意义和内涵是不一样

的.有人看重脸庞和肤色,有人注重身材,有人关注气质,还有人重视品德.粒计算[8-9]是人类智能思考和解决不确定性问题的重要方法,对知识的粒化是人类认识主客观世界的重要方式.模糊集理论[10]、粗糙集理论[11-13]、商空间理论[14]和云模型理论[15]等是当前最主要的 4 种粒计算方法.粗糙集理论[11-13]在处理不确定问题时具有客观性、可解释性和可理解性等优点.但是,受限于模型的结构,粗糙集理论(包括经典的粗糙理论和绝大部分扩展的粗糙集模型)在研究和处理动态变化的、增量式的、海量的数据时存在较大的不足,难以刻画数据的动态性质,更不能从整体上把握数据的变化;F-粗糙集[16-17]是第一个完全动态的粗糙集模型,它将粗糙集从单个信息表中的单个上、下近似对推广到多个信息表中的多个上、下近似对,能够把握数据的局部和全局变化,体现了中国古典哲学思想"道生一,一生二,二生三,三生万物".F-粗糙集可以作为研究和处理数据流、概念漂移的有力工具.

用粗糙集理论与方法研究数据流、探测概念漂移的文献不太多见.文献[18-19]用粗糙集最基本的不确定性指标——上、下近似来定义和探测概念漂移;文献[20]把决策子表看成滑动窗口,对多个滑动窗口进行并行约简,删除冗余属性,把不同滑动窗口之间属性重要性差异作为概念漂移探测指标;文献[21]用 F-粗糙集思想研究了单个信息表内的概念漂移现象,从粗糙集、粒计算的角度提出了认识收敛等概念.但是这些文献仅仅研究了单个概念或少数几个概念的变化,还不能完全反映人类认识的复杂性和多样性,也不能从全局上把握知识系统(或信息系统)中的概念漂移现象.人类认识世界的方式极其复杂,我们猜测人类认识世界的方式也许是一个量子计算的过程,所以需要从全局和局部等多方面把握人类思维的变化.

不确定性分析是粗糙集理论最重要的研究方向.结构性指标上、下近似[11-13],依赖性指标属性依赖度、隶属度[22],信息熵指标互信息[23]、条件熵[24-25],扩展信息熵指标粗糙熵、模糊熵[26-27]等不确定性指标能够方便地描述数据的不确定性,并且具有强客观性、无需先验知识等优点.文献[28]对定量的不确定性指标进行了分析、比较.所有粗糙集模型中,上、下近似是最本源的结构性不确定性度量指标.

经典的概念漂移现象是在数据流中由时间变化引起的,但现实的概念漂移或概念的变化不仅仅存在于数据流之中,也不仅仅是时间变化引起的,更多的情况是由空间或条件变化引发的概念漂移.文献[21]扩展了概念漂移的定义和思想,研究了由空间或条件的变化而引发的概念漂移现象.基于文献[21]概念漂移的思想,融入量子计算的基本思想,运用粗糙集和 F-粗糙集的基本方法,本文定义了知识系统(或信息表、决策表)内概念的全粒度度量和表示,揭示了知识系统(或信息表、决策表)内概念的整体变化和概念漂移现象.首先在知识系统内定义了单个概念的上、下近似概念漂移.其次,结合量子计算的思想,定义了全粒度粗糙集,用粗糙集的思想和方法描述概念在知识系统内的全局表示方式,并分析其性质,指出了它们之间的偏序嵌套关系.第三,在决策表内定义了全粒度正区域、概念漂移、概念耦合等概念.据此,分析了全粒度正区域内的偏序嵌套关系和概念漂移、概念耦合.第四,定义了全粒度属性约简,包括全粒度绝对约简、全粒度值约简和全粒度 Pawlak 约简,初步讨论了它们的性质.最后,讨论了全粒度粗糙集的认识论意义.

2　基础知识

假设读者拥有基本的离散数学知识,本节仅简单介绍相关的粗糙集[11-13]基础知识.

假设 $IS=(U,A)$ 是一个知识系统(或信息系统,若含有决策属性则称为决策系统),其中 U 是论域,U 中的元素称为个体,A 是论域 U 上的知识(或条件属性集).任意属性 $a\in A$ 都关联着一个信息函数 $a:U\to V_a$,其中 V_a 为属性 a 的值域.

属性子集 $B\subseteq A$ 和个体 $x\in U$ 关联着如下的一个信息函数:

$$Inf_B(x)=\{(a,a(x)):a\in B\}.$$

B-不分明关系(或称为不可区分关系)定义为

$$IND(B)=\{(x,y):Inf_B(x)=Inf_B(y)\}.$$

满足 $IND(B)$ 关系的 2 个元素 x、y 不能被 B 的任何属性子集区分,x 的 $IND(B)$ 等价类表示为 $[x]_B$.

定义 1.　在知识系统 $IS=(U,A)$ 中,对于任意 $a\in A$,如果 $IND(A-\{a\})=IND(A)$,则称 $a\in A$ 是可约去的,否则称为必不可少的.

定义 2.　在知识系统 $IS=(U,A)$ 中,$B\subseteq A$ 称为 IS 的约简(绝对约简) iff $B\subseteq A$ 满足下列条件:

(1) $IND(B)=IND(A)$;

(2) 对于任意 $S\subset B$,都有 $IND(S)\neq IND(A)$.

知识系统 IS 的所有约简记为 $RED(IS)$，所有约简的交集称为知识系统 IS 的属性核，记为 $\bigcap RED(IS)$. 属性核中的每个元素称为核属性.

对于知识系统 $IS=(U,A)$、属性子集 $B\subseteq A$ 和论域子集 $X\subseteq U$（论域子集 $X\subseteq U$ 通常也称为概念，在决策系统 $DS=(U,A,d)$ 中一般可以表示为 $X=\{x:d(x)=$常数$\}$，条件属性子集 $B\subseteq A$ 可以称为知识，我们可以用它表达决策属性表示的概念），上、下近似、边界域与负区域的个体表示为

$$\overline{B}(X)=\overline{B}(IS,X)=\{x\in U:[x]_B\cap X\neq\varnothing\},$$

$$\underline{B}(X)=\underline{B}(IS,X)=\{x\in U:[x]_B\subseteq X\},$$

$$BN(IS,B,X)=\overline{B}(IS,X)-\underline{B}(IS,X),$$

$$NEG(IS,B,X)=U-\overline{B}(IS,X).$$

上、下近似，边界域及负区域的信息粒表示分别为

$$\overline{B}(X)=\overline{B}(IS,X)=\bigcup\{[x]_B\subseteq U:[x]_B\cap X\neq\varnothing\},$$

$$\underline{B}(X)=\underline{B}(IS,X)=\bigcup\{[x]_B\subseteq U:[x]_B\subseteq X\},$$

$$BN(B,X)=\overline{B}(IS,X)-\underline{B}(IS,X),$$

$$NEG(B,X)=U-\overline{B}(IS,X).$$

在决策系统 $DS=(U,A,d)$ 中，$\{d\}\cap A=\varnothing$，决策属性 d 将论域 U 划分为块，$U/\{d\}=\{Y_1,Y_2,\cdots,Y_p\}$，其中 $Y_i(i=1,2,\cdots,p)$ 是等价类. 决策系统 $DS=(U,A,d)$ 的正区域定义为

$$POS_A(d)=\bigcup_{Y_i\in U/\{d\}}\underline{A}(Y_i).$$

有时决策系统 $DS=(U,A,d)$ 的正区域 $POS_A(d)$ 也记为 $POS_A(DS,d)$ 或 $POS(DS,A,d)$.

定义 3[11-13]. 在决策系统 $DS=(U,A,d)$ 中，$B\subseteq A$ 是 DS 的约简（Pawlak 约简）iff $B\subseteq A$ 满足下列条件：

(1) $POS_B(d)=POS_A(d)$；

(2) 对于任意 $S\subset B$，都有 $POS_S(d)\neq POS_A(d)$.

所有 DS 的约简记为 $RED(DS)$. $\bigcap RED(DS)$ 称为决策系统 DS 的属性核. $\bigcap RED(DS)$ 中的每个元素称为核属性.

根据文献[11-13,29]可知，值约简可定义如下.

定义 4. 在决策系统 $DS=(U,A,d)$ 中，$U/\{d\}=\{Y_1,Y_2,\cdots,Y_p\}$，则 $B\subseteq A$ 是 $Y\in U/\{d\}$ 的值约简 iff $B\subseteq A$ 满足下列条件：

(1) $\underline{B}(Y)=\underline{A}(Y)$；

(2) 对于任意 $S\subset B$ 都有 $\underline{S}(Y)\neq\underline{A}(Y)$.

值约简是保证决策规则中不含冗余条件属性的属性约简. $Y\in U/\{d\}$ 的所有值约简记为 $RED(DS,Y)$，$\bigcap RED(DS,Y)$ 称为值约简的属性核. $\bigcap RED(DS,Y)$ 中的每个元素称为核属性.

3 知识系统中的概念漂移

在文献[21]中我们研究了信息表中的概念漂移，但是它仅仅表示了属性具有包含关系情况下的概念漂移，下面我们将文献[19,21]中概念漂移的相关定义进行改造并运用于知识系统中一般情况下的概念漂移及其表示.

定义 5. 设 $IS=(U,A)$ 是一个知识系统，$X\subseteq U$ 是其中的一个概念，$B_1\subseteq A$ 和 $B_2\subseteq A$ 是两个不同的知识（属性子集），则概念 $X\subseteq U$ 在不同的知识 $B_1\subseteq A$ 和 $B_2\subseteq A$ 表示下的上、下近似漂移分别被定义为：$\overline{\Delta}(B_1,B_2,X)=\{\overline{B_1}(X)-\overline{B_2}(X),\overline{B_2}(X)-\overline{B_1}(X)\}$，$\underline{\Delta}(B_1,B_2,X)=\{\underline{B_1}(X)-\underline{B_2}(X),\underline{B_2}(X)-\underline{B_1}(X)\}$. 概念 $X\subseteq U$ 在知识 $B_1\subseteq A$ 和 $B_2\subseteq A$ 表示下的漂移被定义为 $(\underline{\Delta}(B_1,B_2,X),\overline{\Delta}(B_1,B_2,X))$.

其中，“$-$”为集合减法.

概念 $X\subseteq U$ 的上、下近似漂移表明了概念 $X\subseteq U$ 在不同的知识 $B_1\subseteq A$ 和 $B_2\subseteq A$ 表示下的上、下近似的变化. 概念 $X\subseteq U$ 在知识 $B_1\subseteq A$ 和 $B_2\subseteq A$ 表示下的漂移则由概念 $X\subseteq U$ 的上、下近似漂移的序偶来表示.

定义 6. 概念 $X\subseteq U$ 在不同的知识 $B_1\subseteq A$ 和 $B_2\subseteq A$ 表示下的上、下近似漂移量分别被定义为（条件同定义 5）

$$\overline{\delta}(B_1,B_2,X)=|\bigcup\overline{\Delta}(B_1,B_2,X)|;$$

$$\underline{\delta}(B_1,B_2,X)=|\bigcup\underline{\Delta}(B_1,B_2,X)|.$$

概念 $X\subseteq U$ 在不同的知识 $B_1\subseteq A$ 和 $B_2\subseteq A$ 表示下的上、下近似漂移量是概念 $X\subseteq U$ 在知识 $B_1\subseteq A$ 和 $B_2\subseteq A$ 表示下的上、下近似漂移并集的势，即概念 $X\subseteq U$ 在知识 $B_1\subseteq A$ 和 $B_2\subseteq A$ 表示下上、下近似的对称差的势. 概念 $X\subseteq U$ 在知识 $B_1\subseteq A$ 和 $B_2\subseteq A$ 表示下的上、下近似漂移量反映了概念 $X\subseteq U$ 在不同表示下的变化量.

定义 7. 知识系统 $IS=(U,A)$ 中概念 $X\subseteq U$ 在不同的知识 $B_1\subseteq A$ 和 $B_2\subseteq A$ 表示下的上、下近似漂移度分别被定义为

$$\overline{d}(B_1,B_2,X)=\frac{\overline{\delta}(B_1,B_2,X)}{|\overline{B_1}(X)|+|\overline{B_2}(X)|},$$

$$\underline{d}(B_1,B_2,X)=\frac{\underline{\delta}(B_1,B_2,X)}{|\underline{B_1}(X)|+|\underline{B_2}(X)|}.$$

概念 $X\subseteq U$ 的上、下近似漂移度分别表示在不同的知识 $B_1\subseteq A$ 和 $B_2\subseteq A$ 下的上、下近似的差异程度.

① 注. 当 $|\underline{B_1}(X)|+|\underline{B_2}(X)|=0$ 时，规定 $\underline{d}(B_1,B_2,X)=0$.

定义 8.　知识系统 $IS=(U,A)$ 中概念 $X\subseteq U$ 在不同的知识 $B_1\subseteq A$ 和 $B_2\subseteq A$ 表示下的上、下近似耦合度定义为

$$\overline{c}(B_1,B_2,X)=\frac{2|\overline{B_1}(X)\cap\overline{B_2}(X)|}{|\overline{B_1}(X)|+|\overline{B_2}(X)|}=1-\overline{d}(B_1,B_2,X),$$

$$\underline{c}(B_1,B_2,X)=\frac{2|\underline{B_1}(X)\cap\underline{B_2}(X)|}{|\underline{B_1}(X)|+|\underline{B_2}(X)|}=1-\underline{d}(B_1,B_2,X).$$

概念的上、下近似耦合度表示概念在不同条件或情况下上、下近似的相似度，是和上、下近似漂移度相对立的度量指标.

4　单个概念的全粒度粗糙集与概念漂移

一个概念，它既可用外延表示，也可用内涵表示. 但概念不一定是精确的，所以粗糙集常用上、下近似来表示和逼近一个概念. 但不同的人、不同的时间、不同的地点对于同一个概念的表达意义可能是不同的. 本节我们以上、下近似为工具研究概念在同一个知识系统中的各种可能变化，即概念的全粒度粗糙集表示. 下文规定 $U/\varnothing=\{U\}$，即当没有任何知识(条件属性或决策属性)时，所有的个体都是不可区分的.

定义 9.　在知识系统 $IS=(U,A)$ 中，概念 $X\subseteq U$ 的全粒度上近似、全粒度下近似、全粒度边界区域与全粒度负区域分别定义为

$$\overline{EAPR}(IS,X)=\{\overline{B}(IS,X):B\subseteq A\};$$

$$\underline{EAPR}(IS,X)=\{\underline{B}(IS,X):B\subseteq A\};$$

$$EBN(IS,X)=\{BN(IS,B,X):B\subseteq A\};$$

$$ENEG(IS,X)=\{NEG(IS,B,X):B\subseteq A\}.$$

全粒度上近似、下近似、边界区域与负区域分别是知识系统 $IS=(U,A)$ 中概念 $X\subseteq U$ 相对于所有属性子集的上近似、下近似、边界区域与负区域的集合. 它们分别是概念 $X\subseteq U$ 在知识系统 $IS=(U,A)$ 中所有可能的上近似、下近似、边界区域与负区域的集合，是概念 $X\subseteq U$ 在知识系统中的所有可能的粗糙集表达方式，也是概念 $X\subseteq U$ 在知识系统 $IS=(U,A)$ 中所有可能变化的集合. 序偶 $(\underline{EAPR}(IS,X),\overline{EAPR}(IS,X))$ 称为概念 $X\subseteq U$ 在知识系统 $IS=(U,A)$ 中的全粒度粗糙集.

根据需要，全粒度粗糙集也可以表示为 $\{(\underline{B}(IS,X),\overline{B}(IS,X)):B\subseteq A\}$，$(\underline{EAPR}(IS,A,X),\overline{EAPR}(IS,A,X))$ 或 $\{(\underline{B}(IS,A,X),\overline{B}(IS,A,X)):B\subseteq A\}$.

例如，设知识系统 $IS=\{U,A\}$，其中 $A=\{a,b\}$. $X\subseteq U$ 是知识系统中的一个概念，则

$$\underline{EAPR}(IS,A,X)=\{\underline{\varnothing}(X),\underline{\{a\}}(X),\underline{\{b\}}(X),\underline{A}(X)\}=\{\varnothing,\underline{\{a\}}(X),\underline{\{b\}}(X),\underline{A}(X)\},$$

$$\overline{EAPR}(IS,A,X)=\{\overline{\varnothing}(X),\overline{\{a\}}(X),\overline{\{b\}}(X),\overline{A}(X)\}=\{U,\underline{\{a\}}(X),\underline{\{b\}}(X),\underline{A}(X)\}.$$

分别用图 1、图 2 表示.

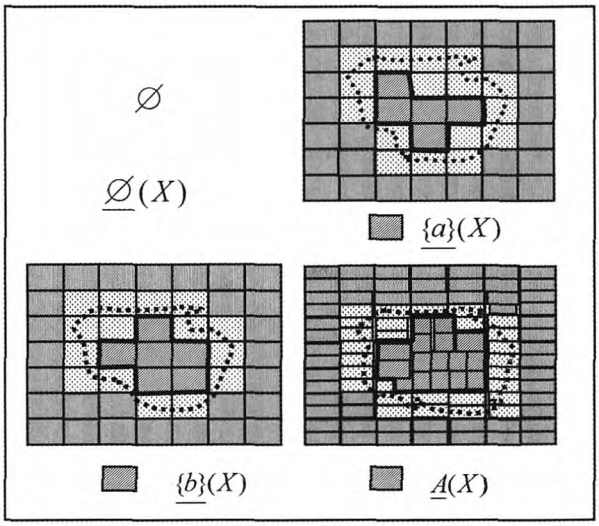

图 1　全粒度下近似 $\underline{EAPR}(IS,A,X)$

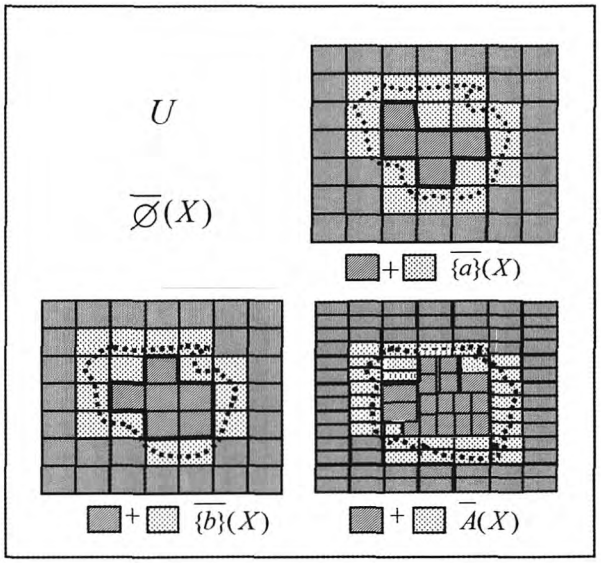

图 2　全粒度上近似 $\overline{EAPR}(IS,A,X)$

与其他粗糙集模型相比，全粒度粗糙集体现了粗糙集动静结合的思想. 除 F-粗糙集外，几乎所有的粗糙集模型从本质上来说都是静态的，虽然不少粗糙集模

型被用于研究动态的、增量式的数据挖掘，但是它们往往把动态的、增量式数据放入静态的模型当中，所以这些粗糙集模型在处理动态的、海量的数据方面显得非常局限和力不从心. F-粗糙集主要研究数据量的变化，不涉及关系(或属性)的变化，全粒度粗糙集则主要研究关系(或属性)的全局与局部变化. 全粒度粗糙集的静主要体现在它的上、下近似，边界区域都可以用一个集合的集合来表示；它的动主要体现在全粒度粗糙集内部，无论是全粒度上近似，还是全粒度下近似的内部都包含概念在不同情况下的所有上、下近似，体现了一种粒度的变化以及概念表示的逼近过程.

全粒度粗糙集定义了概念 $X\subseteq U$ 所有粒度层次上的上、下近似，可以体现量子叠加和纠缠，具有一定的量子计算思想. 全粒度粗糙集具有很强的表示能力，表示了概念在各种情况下的可能变化，能够进行并行计算，但其缺点是时间复杂性和空间复杂性高. 在量子计算的条件下，量子比特具有超强的表达能力和并行计算能力，通过量子计算能够在叠加和纠缠中获得所需的上、下近似，体现人类智能的复杂性、不确定性和多样性，也可以表示人类认识从一个粒度转跳到另一个粒度，转跳自如，毫无困难.

例 1. 设 $DS=(U,A,d)$ 是决策表，如表 1 所示. a、b、c 是条件属性，d 是决策属性. 令 $X=\{x:d(x)=0,x\in U\}$，则 $\overline{\varnothing}(X)=U$；$\overline{\{a\}}(X)=U$；$\overline{\{b\}}(X)=U$；$\overline{\{c\}}(X)=U$；$\overline{\{a,b\}}(X)=\{y_1,y_4,y_5,y_6\}$；$\overline{\{b,c\}}(X)=U$；$\overline{\{c,a\}}(X)=U$；$\overline{\{a,b,c\}}(X)=\{y_1,y_4,y_5,y_6\}$；所以，$\overline{EAPR}(IS,X)=\{\overline{\varnothing}(X),\overline{\{a\}}(X),\overline{\{b\}}(X),\overline{\{c\}}(X),\overline{\{a,b\}}(X),\overline{\{b,c\}}(X),\overline{\{c,a\}}(X),\overline{\{a,b,c\}}(X)\}=\{U,\{y_1,y_4,y_5,y_6\}\}$. $\underline{\varnothing}(X)=\varnothing$；$\underline{\{a\}}(X)=\{y_1,y_4\}$；$\underline{\{b\}}(X)=\varnothing$；$\underline{\{c\}}(X)=\varnothing$；$\underline{\{a,b\}}(X)=\{y_1,y_4\}$；$\underline{\{b,c\}}(X)=\varnothing$，$\underline{\{c,a\}}(X)=\{y_1,y_4\}$；$\underline{\{a,b,c\}}(X)=\{y_1,y_4\}$；所以，$\underline{EAPR}(IS,X)=\{\underline{\varnothing}(X),\underline{\{a\}}(X),\underline{\{b\}}(X),\underline{\{c\}}(X),\underline{\{a,b\}}(X),\underline{\{b,c\}}(X),\underline{\{c,a\}}(X),\underline{\{a,b,c\}}(X)\}=\{\varnothing,\{y_1,y_4\}\}$.

表 1 决策表 *DS*

U	a	b	c	d
y_1	0	1	0	0
y_2	1	1	0	1
y_3	1	1	0	1
y_4	0	1	0	0
y_5	1	2	0	0
y_6	1	2	0	1

下面结合绝对约简、值约简、核属性、概念漂移等，讨论全粒度粗糙集的性质.

定理 1. 在知识系统 $IS=(U,A)$ 中，概念 $X\subseteq U$ 的可能表达方式个数为 $2^{|A|}$；所有可能的概念个数为 $2^{|U|}$(假设 $\varnothing\subseteq U$ 也是一个概念，称为空概念).

证明. 在知识系统 $IS=(U,A)$ 中，属性 A 的可能组合方式为 2^A(A 的幂集)，对于每一种组合概念 $X\subseteq U$ 都有一种表示方式，所以在知识系统 $IS=(U,A)$ 中概念 $X\subseteq U$ 的可能表达方式个数为 $2^{|A|}$. 同理，在知识系统 $IS=(U,A)$ 中所有可能的概念个数为 $2^{|U|}$. 证毕.

根据定理 1，知识系统 $IS=(U,A)$ 中所有可能的概念及其表现形式有 $2^{|U|}\times 2^{|A|}$ 个.

推论 1. 决策系统 $DS=(U,A,d)$ 中所有可能的概念及其表现形式有 $|d|\times 2^{|A|}$ 个.

定理 2[30]. 在一个知识系统 $IS=(U,A)$ 中，对于 $B_1\subseteq B_2\subseteq A$ 和 $X\subseteq U$，有 $\underline{B_1}(IS,X)\subseteq\underline{B_2}(IS,X)\subseteq\overline{B_2}(IS,X)\subseteq\overline{B_1}(IS,X)$.

推论 2. 在一个知识系统中 $IS=(U,A)$ 中，对于 $B_1\subseteq B_2\subseteq A$ 和 $X\subseteq U$，有 $BN_{B_2}(X)\subseteq BN_{B_1}(X)$.

定义 10. 在知识系统 $IS=(U,A)$ 中，$P(A)$ 表示 A 的幂集，$B\subseteq A$ 称为全粒度绝对约简 iff $B\subseteq A$ 满足以下 2 个条件：

(1) 对于任意 $A_1\in P(A)$，存在 $B_1\in P(B)$，使得 $U/A_1=U/B_1$；

(2) 对于任意 $S\subset B$，存在 $A_1\in P(A)$，使得对于任意 $B_1\in P(S)$，都有 $U/A_1\neq U/B_1$.

知识系统中所有全粒度绝对约简的交集称为全粒度绝对约简的属性核，属性核中的属性称为核属性.

全粒度绝对约简是知识系统中保持每一个属性子集对论域的划分都不变的最小属性子集.

显然，知识系统绝对约简的核属性一定是全粒度绝对约简的核属性.

定理 3. 在知识系统 $IS=(U,A)$ 中，若 $B\subseteq A$ 是 $IS=(U,A)$ 的全粒度绝对约简，则对于任意 $X\subseteq U$ 都有 $X\subseteq U$ 关于 $B\subseteq A$ 的全粒度粗糙集等于 $X\subseteq U$ 关于 A 的全粒度粗糙集合. 即

$$(\underline{EAPR}(IS,B,X),\overline{EAPR}(IS,B,X))=(\underline{EAPR}(IS,A,X),\overline{EAPR}(IS,A,X)).$$

证明. 由于 $B\subseteq A$，显然有

$$\underline{EAPR}(IS,B,X)\subseteq\underline{EAPR}(IS,A,X),$$

$$\overline{EAPR}(IS,B,X)\subseteq\overline{EAPR}(IS,A,X).$$

因为 $B\subseteq A$ 是 $IS=(U,A)$ 的全粒度绝对约简，所以对于任意 $A_1\in P(A)$，存在 $B_1\in P(B)$，使得对

于任意 $x\in U$ 都有 $[x]_{B_1}=[x]_{A_1}$，从而有对于任意 $X\subseteq U$，$\overline{B_1}(IS,X)=\overline{A_1}(IS,X)$ 且 $\underline{B_1}(IS,X)=\underline{A_1}(IS,X)$ 成立. 所以有 $\underline{EAPR}(IS,A,X)\subseteq\underline{EAPR}(IS,B,X)$，$\overline{EAPR}(IS,A,X)\subseteq\overline{EAPR}(IS,B,X)$.

于是有，$\underline{EAPR}(IS,B,X)=\underline{EAPR}(IS,A,X)$ 且 $\overline{EAPR}(IS,B,X)=\overline{EAPR}(IS,A,X)$.　　证毕.

推论 3.　在知识系统 $IS=(U,A)$ 中，若 $B\subseteq A$ 是 $IS=(U,A)$ 的全粒度绝对约简，则对于任意 $X\subseteq U$ 和任意 $B_1(B\subseteq B_1\subseteq A)$ 都有 $X\subseteq U$ 关于 $B_1\subseteq A$ 的全粒度粗糙集等于 $X\subseteq U$ 关于 A 的全粒度粗糙集合. 即 $(\underline{EAPR}(IS,B_1,X),\overline{EAPR}(IS,B_1,X))=(\underline{EAPR}(IS,A,X),\overline{EAPR}(IS,A,X))$.

知识系统 $IS=(U,A)$ 的全粒度绝对约简能保持知识系统中任意 $X\subseteq U$ 的全粒度粗糙集不发生改变，保持知识系统中所有的概念不发生概念漂移，包括上近似漂移和下近似漂移.

在知识系统 $IS=(U,A)$ 中约简全粒度非核属性，不会改变知识系统中的分类，也不会改变任何概念的上、下近似，从而不会改变概念的全粒度粗糙集，也不会发生概念漂移. 知识系统 $IS=(U,A)$ 全粒度核属性的约简，一定会改变知识系统的某些分类，从而可能改变概念的上、下近似，可能改变概念的全粒度粗糙集，也可能发生概念漂移. 所以，知识系统 $IS=(U,A)$ 中全粒度绝对约简是保证所有概念的全粒度粗糙集不发生变化且不发生概念漂移的最小属性子集，一旦某个核属性被约简，知识系统中的某些概念就会发生概念漂移，从而它们的全粒度粗糙集发生变化.

虽然从值的角度来看，全粒度绝对约简不会改变概念的全粒度粗糙集，但是冗余属性的存在使得全粒度粗糙集表示得更加丰富多样，而且冗余属性越多，约简的个数和可能性更多. 所以，从来源角度看，全粒度绝对约简同样改变着概念的全粒度粗糙集，它使得全粒度粗糙集的表示变得单一.

然而，对于知识系统 $IS=(U,A)$ 和具体概念 $X\subseteq U$，有着类似却又不完全相同的情况.

定义 11.　在知识系统 $IS=(U,A)$ 中，$P(A)$ 表示 A 的幂集，$X\subseteq U$ 是一个概念，$B\subseteq A$ 称为概念 $X\subseteq U$ 全粒度值约简 iff $B\subseteq A$ 满足以下 2 个条件：

(1) 对于任意 $A_1\in P(A)$，存在 $B_1\in P(B)$，使得 $\underline{B_1}(IS,X)=\underline{A_1}(IS,X)$；

(2) 对于任意 $S\subset B$，存在 $A_1\in P(A)$，使得对于任意 $B_1\in P(S)$，都有 $\underline{B_1}(IS,X)\neq\underline{A_1}(IS,X)$.

全粒度值约简是保持概念所有属性子集下近似不变的最小属性子集.

所有全粒度值约简的交集称为全粒度值约简的属性核，属性核中的属性称为全粒度值约简的核属性.

显然，概念值约简的核属性是全粒度值约简的核属性.

定理 4.　设 $IS=(U,A)$ 是一个知识系统，$B\subseteq A$ 是概念 $X\subseteq U$ 的全粒度值约简，则 $X\subseteq U$ 关于 $B\subseteq A$ 全粒度下近似等于 $X\subseteq U$ 关于 A 的全粒度下近似. 即 $\underline{EAPR}(IS,B,X)=\underline{EAPR}(IS,A,X)$.

证明.　$\underline{EAPR}(IS,B,X)\subseteq\underline{EAPR}(IS,A,X)$ 显然成立. 下面证明：

$$\underline{EAPR}(IS,A,X)\subseteq\underline{EAPR}(IS,B,X).$$

因为 $B\subseteq A$ 是概念 $X\subseteq U$ 的全粒度值约简，所以对于任意 $A_1\in P(A)$，存在 $B_1\in P(B)$ 使得 $\underline{B_1}(IS,X)=\underline{A_1}(IS,X)$ 从而有 $\underline{EAPR}(IS,A,X)\subseteq\underline{EAPR}(IS,B,X)$.　　证毕.

推论 4.　在知识系统 $IS=(U,A)$ 中，$B\subseteq A$ 是概念 $X\subseteq U$ 的全粒度值约简，则对于任意 $B_1(B\subseteq B_1\subseteq A)$ 都有 $\underline{EAPR}(IS,B_1,X)=\underline{EAPR}(IS,A,X)$.

定理 5.　$a\in A$ 是知识系统 $IS=(U,A)$ 中概念 $X\subseteq U$ 值约简的核属性，则 $\underline{EAPR}(IS,B,X)\neq\underline{EAPR}(IS,A,X)$，其中 $B=A-\{a\}$.

证明.　因为 $a\in A$ 是知识系统 $IS=(U,A)$ 中概念 $X\subseteq U$ 值约简的核属性，所以有 $\underline{B}(IS,X)\neq\underline{A}(IS,X)$，从而有 $\underline{A}(IS,X)\notin\underline{EAPR}(IS,B,X)$，故 $\underline{EAPR}(IS,B,X)\neq\underline{EAPR}(IS,A,X)$.　　证毕.

知识系统 $IS=(U,A)$ 中某个概念 $X\subseteq U$ 的全粒度值约简仅能保证概念 $X\subseteq U$ 的全粒度下近似不发生变化，不发生下近似概念漂移；并不能保证概念 $X\subseteq U$ 的全粒度上近似不发生变化，从而可能发生上近似概念漂移. 知识系统 $IS=(U,A)$ 中某个概念 $X\subseteq U$ 值约简的核属性也仅仅是相对于下近似而言的，约简核属性导致下近似发生变化，从而该概念的全粒度下近似发生变化，也发生了下近似概念漂移.

下面讨论知识系统 $IS=(U,A)$ 中全粒度粗糙集内部之间的关系.

定理 6.　在知识系统 $IS=(U,A)$ 中，概念 $X\subseteq U$ 的全粒度上近似、下近似、边界区域与负区域中的元素相对于关系“$\subseteq$”满足自反、反对称、传递，即 $(\underline{EAPR}(IS,X),\subseteq)$、$(\overline{EAPR}(IS,X),\subseteq)$、$(EBN(IS,X),\subseteq)$、$(ENEG(IS,X),\subseteq)$ 均为偏序集.

证明. 根据关系“$\subseteq$”性质,易得. 证毕.

② **注.** 在$\underline{EAPR}(IS,X)$、$\overline{EAPR}(IS,X)$、$EBN(IS,X)$、$ENEG(IS,X)$中将相等元素看成一个或用一个作为代表.

因此,$(\underline{EAPR}(IS,X),\subseteq)$、$(\overline{EAPR}(IS,X),\subseteq)$、$(EBN(IS,X),\subseteq)$、$(ENEG(IS,X),\subseteq)$均可用哈斯图表示.

定理 7. 在知识系统$IS=(U,A)$中,对于概念$X\subseteq U$有

(1) $\underline{A}(X)=\bigcup\underline{EAPR}(IS,X)$;

(2) $U=\bigcup\overline{EAPR}(IS,X)$.

证明. (1) 根据定理 2,对于任意的$B\subseteq A$都有$\underline{B}(X)\subseteq\underline{A}(X)$,所以有$\bigcup\underline{EAPR}(IS,X)\subseteq\underline{A}(X)$;又$\underline{A}(X)\in\underline{EAPR}(IS,X)$,所以有$\underline{A}(X)\subseteq\bigcup\underline{EAPR}(IS,X)$;于是$\underline{A}(X)=\bigcup\underline{EAPR}(IS,X)$.

类似地,我们可以证明(2). 证毕.

定理 8. 在知识系统$IS=(U,A)$中,对于概念$X\subseteq U$,$(\underline{EAPR}(IS,X),=)$、$(\overline{EAPR}(IS,X),=)$、$(EBN(IS,X),=)$、$(ENEG(IS,X),=)$均为等价关系.

证明. 根据关系“$=$”的性质,易得. 证毕.

推论 5. 在知识系统$IS=(U,A)$中,对于概念$X\subseteq U$,$\underline{EAPR}(IS,X)/=$,$\overline{EAPR}(IS,X)/=$,$EBN(IS,X)/=$,$ENEG(IS,X)/=$分别为$\underline{EAPR}(IS,X)$、$\overline{EAPR}(IS,X)$、$EBN(IS,X)$和$ENEG(IS,X)$相对于相等关系的划分.

对于知识系统$IS=(U,A)$中的概念$X\subseteq U$,在全粒度上近似、全粒度下近似、全粒度边界域、全粒度负区域中的等价类是同一个概念的不同表达方式,并没有发生概念漂移;而不同等价类中的表达方式则发生了概念漂移. 定理 1 表明了同一个概念在知识系统$IS=(U,A)$中在不同的知识表示下可能的概念漂移.

③ **注.** 对于知识系统$IS=(U,A)$中概念$X\subseteq U$全粒度上近似等价类、全粒度下近似等价类、全粒度边界域等价类以及全粒度负区域等价类中的每个元素,它们既相等又不相等. 从值的角度看,等价类中的每个元素都相等;从来源角度看,等价类中的每个元素相互不相等. 于是有:

定理 9. 在知识系统$IS=(U,A)$中,对于概念$X\subseteq U$和相等关系对全粒度下近似、全粒度上近似、全粒度边界域及全粒度负区域的划分$\underline{EAPR}(IS,X)/=$,$\overline{EAPR}(IS,X)/=$,$EBN(IS,X)/=$,$ENEG(IS,X)/=$,同一个等价类中的概念$X\subseteq U$的不同表示方式概念漂移度为 0,不同的等价类中概念$X\subseteq U$的不同表示方式概念漂移度大于 0.

证明. 根据相关定义,容易证明该结论. 证毕.

定理 10. 在知识系统$IS=(U,A)$中,对于概念$X\subseteq U$和相等关系对全粒度下近似、全粒度上近似、全粒度边界域及全粒度负区域的划分$\underline{EAPR}(IS,X)/=$,$\overline{EAPR}(IS,X)/=$,$EBN(IS,X)/=$,$ENEG(IS,X)/=$,在同一个等价类中,每个元素的属性子集相对于关系”$\subseteq$”构成偏序关系. 对于每个偏序集,每个极小元素为相应的值约简.

证明. 首先,在每个等价类中,所有元素相应的下近似、上近似、边界域或负区域相等;其次,极小元素意味着没有比其更小的元素. 因此,满足值约简的定义. 即定理成立. 证毕.

全粒度上近似、全粒度下近似、全粒度边界域、全粒度负区域中的元素对于关系“$=$”和“$\subseteq$”可以构成嵌套哈斯图.

例 2. 续例 1. $\overline{EAPR}(IS,X)/=$、$\underline{EAPR}(IS,X)/=$分别如下:

$$\begin{aligned}\overline{EAPR}(IS,X)/=&=\{\{\overline{\varnothing}(X),\overline{\{a\}}(X),\overline{\{b\}}(X),\\&\quad\overline{\{c\}}(X),\overline{\{b,c\}}(X),\overline{\{c,a\}}(X)\},\\&\quad\{\overline{\{a,b\}}(X),\overline{\{a,b,c\}}(X)\}\}.\\\underline{EAPR}(IS,X)/=&=\{\{\underline{\varnothing}(X),\underline{\{a\}}(X),\underline{\{b\}}(X),\\&\quad\underline{\{c\}}(X),\underline{\{b,c\}}(X)\}\},\\&\quad\{\{a\}(X),\underline{\{a,b\}}(X),\\&\quad\underline{\{c,a\}}(X),\underline{\{a,b,c\}}(X)\}\}.\end{aligned}$$

可用嵌套哈斯图表示$\overline{EAPR}(IS,X)/\subseteq$与$\overline{EAPR}(IS,X)/=$内部(如图 3 所示),$\underline{EAPR}(IS,X)/\subseteq$与$\underline{EAPR}(IS,X)/=$内部(如图 4 所示):

$$\overline{\Delta}(\{b,c\},\{a,b\},X)=\{\{y_2,y_3\},\varnothing\};$$

$$\underline{\Delta}(\{b,c\},\{a,b\},X)=\{\varnothing,\{y_1,y_4\}\};$$

$$\overline{\delta}(\{b,c\},\{a,b\},X)=|\bigcup\overline{\Delta}(\{b,c\},\{a,b\},X)|=2;$$

$$\underline{\delta}(\{b,c\},\{a,b\},X)=|\bigcup\underline{\Delta}(\{b,c\},\{a,b\},X)|=2;$$

$$\overline{d}(\{b,c\},\{a,b\},X)=\frac{\overline{\delta}(\{b,c\},\{a,b\},X)}{|\overline{\{b,c\}}(X)|+|\overline{\{a,b\}}(X)|}=\frac{2}{6+4}=\frac{1}{5};$$

$$\underline{d}(\{b,c\},\{a,b\},X)=\frac{\underline{\delta}(\{b,c\},\{a,b\},X)}{|\underline{\{b,c\}}(X)|+|\underline{\{a,b\}}(X)|}=\frac{2}{0+2}=1;$$

$$\overline{c}(\{b,c\},\{a,b\},X)=1-\overline{d}(\{b,c\},\{a,b\},X)=\frac{4}{5};$$

$$\underline{c}(\{b,c\},\{a,b\},X)=1-\underline{d}(\{b,c\},\{a,b\},X)=0.$$

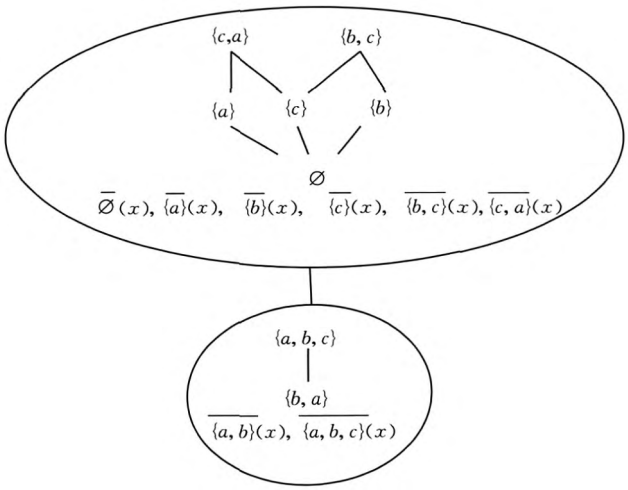

图 3　$\overline{EAPR}(IS,X)/\subseteq$与$\overline{EAPR}(IS,X)/=$内部的嵌套哈斯图

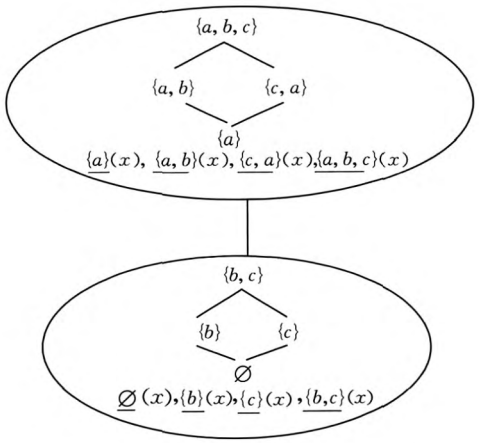

图 4　$\underline{EAPR}(IS,X)/\subseteq$与$\underline{EAPR}(IS,X)/=$内部的嵌套哈斯图

5　决策系统中的全粒度粗糙集及概念漂移

上述讨论是针对知识系统内单个概念的概念漂移、耦合等进行的，对于整个决策表或信息表来说，这些指标显得非常局限，因为一个决策表或信息表中有多个概念，将多个概念放在一起讨论概念漂移、耦合及其度量是本节讨论的内容.

定义 12.　设$DS=(U,A,d)$是一个决策系统，则DS的全粒度正区域定义为$EPOS(DS)=\{POS(DS,B,d):B\subseteq A\}$.

全粒度正区域$EPOS(DS)$可以根据需要记为$EPOS(DS,A)$.

DS的全粒度正区域$EPOS(DS)$是所有条件属性子集正区域的集合. 全粒度正区域具有动静结合、全局与局部相结合的思想. 全粒度正区域是一个全局的集合，一个集合的集合，是一种静态. 但全粒度正区域的内部包含很多正区域，这些不同的正区域是局部的、也是变化的，是一个不断变化的内部过程.

与全粒度粗糙集一样，全粒度正区域的定义也可以体现量子叠加和纠缠，具有一定的量子计算思想. 全粒度正区域表达能力强，能够表示各种粒度情况下信息粒的正区域，适合并行计算，但同样具有时间、空间复杂性高的缺点. 在量子计算情况下，利用量子比特的叠加和纠缠，进行并行计算和表示，通过控制状态的变化获取人们所需的正区域. 所以，全粒度正区域与全粒度粗糙集一样能够体现人类表达和思维的复杂性、多样性和不确定性，同样也能够表示人类认识从一个粒度转跳到另一个粒度，转跳自如，毫无困难.

定理 11[30].　设$DS=(U,A,d)$是一个决策系统，$B_1\subseteq B_2\subseteq A$，则有 $POS_{B_1}(d)\subseteq POS_{B_2}(d)\subseteq POS_A(d)$.

定理 12.　设$DS=(U,A,d)$是一个决策系统，则有 $POS_A(d)=\bigcup EPOS(DS)$.

证明.　对于任何 $POS(DS,B,d)\in EPOS(DS)$，$(B\subseteq A)$都有 $POS(DS,B,d)\subseteq POS_A(d)$，所以，$\bigcup EPOS(DS)\subseteq POS_A(d)$. 又因为 $POS_A(d)\in EPOS(DS)$，所以 $POS_A(d)\subseteq\bigcup EPOS(DS)$. 于是，$POS_A(d)=\bigcup EPOS(DS)$.　　证毕.

根据上述定理将文献[19,21]中的指标进行改造，我们得到下面概念漂移、概念耦合的度量指标.

定义 13.　设$DS=(U,A,d)$是一个决策系统，$B_1\subseteq A$和$B_2\subseteq A$，则决策表中相对于B_1、B_2的概念漂移定义为

$$\Delta(DS,B_1,B_2)=\{POS_{B_2}(d)-POS_{B_1}(d),\ POS_{B_1}(d)-POS_{B_2}(d)\}.$$

决策表的概念漂移表示在不同的知识下决策表中正区域的变化.

定义 14.　设$DS=(U,A,d)$是一个决策系统，$B_1\subseteq A$和$B_2\subseteq A$，则决策表中相对于B_1、B_2的概念耦合度定义为

$$c(DS,B_1,B_2)=\frac{2|POS_{B_1}(d)\cap POS_{B_2}(d)|}{|POS_{B_1}(d)|+|POS_{B_2}(d)|}.$$

决策表的耦合度表示了在不同的知识下正区域的相似度.

定义 15.　设$DS=(U,A,d)$是一个决策系统，$B_1\subseteq A$和$B_2\subseteq A$，则决策表相对于B_1、B_2的概念漂移度定义为

$$d(DS,B_1,B_2)=\frac{|\bigcup\Delta(DS,B_1,B_2)|}{|POS_{B_1}(d)|+|POS_{B_2}(d)|}=1-c(DS,B_1,B_2).$$

④ **注.** 当 $|POS_{B_1}(d)|+|POS_{B_2}(d)|=0$，规定 $d(DS,B_1,B_2)=0$. $d(DS,B_1,B_2)=0$.

决策表的概念漂移度表明了在不同的知识下正区域的变化程度.

下文结合决策表中的属性约简、核属性、概念漂移等概念，讨论全粒度正区域的性质.

定义 16. 在决策系统 $DS=(U,A,d)$ 中，设 $B\subseteq A$ 是全粒度 Pawlak 约简 iff $B\subseteq A$ 满足下面 2 个条件：

(1) $EPOS(DS,B)=EPOS(DS,A)$；

(2) 对于任意 $S\subset B$，都有 $EPOS(DS,S)\neq EPOS(DS,A)$.

所有全粒度 Pawlak 约简的交集称为全粒度 Pawlak 约简的属性核，属性核中的属性称为全粒度 Pawlak 约简的核属性.

定理 13. $B\subseteq A$ 是决策系统 $DS=(U,A,d)$ 的全粒度 Pawlak 约简，则对于任意 $B_1(B\subseteq B_1\subseteq A)$ 都有 $EPOS(DS,B_1)=EPOS(DS,A)$.

证明. 因为 $EPOS(DS,B)=EPOS(DS,A)$，又因为 $EPOS(DS,B)\subseteq EPOS(DS,B_1)$ 且 $EPOS(DS,B_1)\subseteq EPOS(DS,A)$，所以 $EPOS(DS,B_1)=EPOS(DS,A)$. 证毕.

对于全粒度正区域来说，全粒度 Pawlak 约简可以减少冗余属性，而不改变全粒度正区域的值，但是减少了冗余属性的全粒度正区域表示单一，缺乏多样性和灵活性.

定理 14. 设 $a\in A$ 是决策系统 $DS=(U,A,d)$ 的核属性，则 $EPOS(DS,B)\neq EPOS(DS,A)$，其中 $B=A-\{a\}$.

证明. 因为 $a\in A$ 是 $DS=(U,A,d)$ 的核属性，所以有 $POS_B(DS,d)\neq POS_A(DS,d)$，于是，$POS_A(DS,d)\notin EPOS(DS,B)$，因此，$EPOS(DS,B)\neq EPOS(DS,A)$. 证毕.

核属性对于正区域来说是必不可少的，同样核属性对于全粒度正区域来说也是必不可少的.

定理 15. 在决策系统 $DS=(U,A,d)$ 中，关系 $(EPOS(DS),\subseteq)$ 满足自反、反对称和传递，即关系 $(EPOS(DS),\subseteq)$ 是偏序关系.

证明. 根据关系“$\subseteq$”的性质，易得. 证毕.

⑤ **注.** $EPOS(DS)$ 中正域相等的元素既相等又不相等. 从值来说，正域相等的元素完全相等；但从来源来说，因为正域相等的元素对应的属性子集不相等，所以又可认为它们不相等.

定理 16. 在决策系统 $DS=(U,A,d)$ 中，关系 $(EPOS(DS),=)$ 是等价关系，对于 $EPOS(DS)/=$ 中的每一块，其相应的条件属性子集相对于关系“$\subseteq$”构成偏序关系，每个偏序关系的极小元为相应等价类的属性约简.

证明. 根据离散数学知识，容易得到 $EPOS(DS)/=$ 中的每一块，其相应的条件属性子集相对于“$\subseteq$”构成偏序关系，下面简单证明定理的后半部分.

在每个等价类中，所有元素的值都相等，即正域相等；其次，在每个等价类组成的偏序关系中的极小元素(属性子集)没有其它元素比它更小. 所以，每个极小元素都为相应等价类的属性约简. 证毕.

对于 $(EPOS(DS),\subseteq)$ 和 $EPOS(DS)/=$ 块内的属性子集而言，它们构成嵌套哈斯图. 即关系 $(EPOS(DS),\subseteq)$ 构成一个哈斯图，$EPOS(DS)/=$ 块内的属性子集对于关系 $\subseteq$ 来说构成另一个哈斯图.

定理 17. 在决策系统 $DS=(U,A,d)$ 中，对于 $EPOS(DS)/=$ 的等价类，块内的元素之间概念漂移度等于 0；块间元素之间的概念漂移度大于 0.

证明. 根据相关定义，容易证明这个结论. 证毕.

例 3. 条件与例 1 同.

$POS_{\varnothing}(d)=\varnothing$；$POS_{\{a\}}(d)=\{y_1,y_4\}$；$POS_{\{b\}}(d)=\varnothing$；$POS_{\{c\}}(d)=\varnothing$；$POS_{\{a,b\}}(d)=\{y_1,y_2,y_3,y_4\}$；$POS_{\{b,c\}}(d)=\varnothing$；$POS_{\{c,a\}}(d)=\{y_1,y_4\}$；$POS_{\{a,b,c\}}(d)=\{y_1,y_2,y_3,y_4\}$；所以

$EPOS(DS)=\{POS_{\varnothing}(d),POS_{\{a\}}(d),POS_{\{b\}}(d),POS_{\{c\}}(d),POS_{\{a,b\}}(d),POS_{\{b,c\}}(d),POS_{\{c,a\}}(d),POS_{\{a,b,c\}}(d)\}=\{\varnothing,\{y_1,y_4\},\{y_1,y_2,y_3,y_4\}\}$，

$EPOS(DS)/==\{\{POS_{\varnothing}(d),POS_{\{b\}}(d),POS_{\{c\}}(d),POS_{\{b,c\}}(d)\},\{POS_{\{a\}}(d),POS_{\{c,a\}}(d)\},\{POS_{\{a,b\}}(d),POS_{\{a,b,c\}}(d)\}\}$.

$EPOS(DS)/=$ 内部与 $(EPOS(DS),\subseteq)$ 的嵌套哈斯图表示如下(如图 5 所示)：

$\Delta(DS,\{a,b\},\{c,a\})=\{\{y_2,y_3\},\varnothing\}$；

$$d(DS,\{a,b\},\{c,a\})=\frac{|\bigcup\Delta(DS,\{a,b\},\{c,a\})|}{|POS_{\{a,b\}}(d)|+|POS_{\{c,a\}}(d)|}=\frac{2}{4+2}=\frac{1}{3};$$

$$c(DS,\{a,b\},\{c,a\})=1-d(DS,\{a,b\},\{c,a\})=\frac{2}{3}.$$

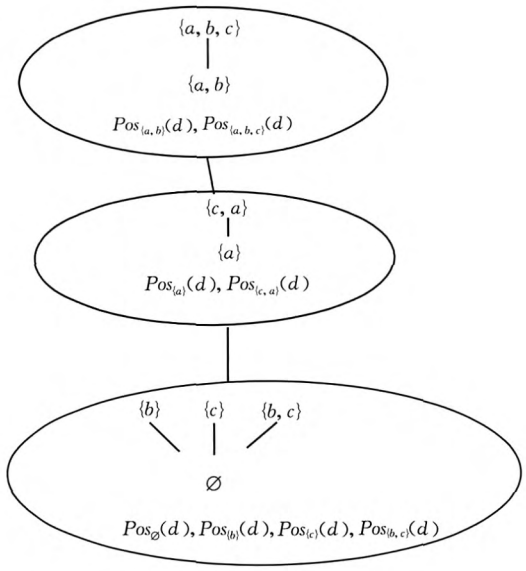

图 5　$EPOS(DS)/=$内部与$(EPOS(DS),\subseteq)$的嵌套哈斯图

6　认识论意义与讨论

粗糙集理论认为“知识就是分类”. 每个人都具有自己的知识体系，但是，不同的情况下人们掌握的信息不一样，有时候掌握的信息全面一些，有时候掌握的信息少些，无论哪种情况，我们都要根据自己的知识体系和所掌握的信息做出判断. 全粒度粗糙集可以对应这种情况，知识系统就是我们所掌握的知识，属性子集就是我们所掌握的信息，我们总是根据我们的知识系统和所掌握的信息做出决策. 通常情况下，成年人掌握的知识体系比较稳定，不太发生变化，但是接受的信息则千变万化，所以做出的决策或选择会有很大的不同.

从认识论和全粒度粗糙集角度看，全粒度属性约简虽然不会导致全粒度粗糙集或全粒度正区域发生变化，不改变分类或认识，但是冗余属性的存在使得认识或表达具有多样性、灵活性和可替换性，因为约简掉冗余属性之后，知识的表达非常单一、非常死板，而且不可替换.

此外，我们在表达一个概念的时候，也许我们只是表示这个概念的某些意义，而不是这个概念的全部意义，所以同一个概念在不同的情况下由不同的人表示出来的意义会有一定的差异，接受概念的意义与表达概念的意义也有一定的差异，这就是概念漂移. 全粒度粗糙集也能够较好地表示这种情况. 另外，同一个概念的同样意义我们可能用不同的方式表示，在全粒度粗糙集中的上、下近似等价类可以描述这种情况.

与现有的粗糙集不一样，全粒度粗糙集表示了知识系统中概念的可能变化，非常具有灵活性，更能够表示和分析概念的不确定性与概念漂移. 全粒度粗糙集可以从一个粒度转跳到另一个粒度，转跳自如，毫无困难，与人类认识世界的方式相吻合. 但是，在全粒度粗糙集中每一个概念都具有 $2^{|A|}$ 种表现形式，在具体情况下如何快速地选择我们需要的一种形式，需要高效的算法或量子算法才能解决. 所以，如何用高效的方式存储或表示全粒度粗糙集并快速地找到我们所需要的表达方式是我们将要进一步研究的内容.

7　结　论

从量子计算、粒计算、粗糙集和数据流、概念漂移的角度观察知识系统，以上、下近似为工具，本文定义了全粒度粗糙集，概念的上、下近似漂移，上、下近似耦合等概念，分析了知识系统内概念的全局变化与局部变化. 从单个概念在知识系统中所有可能的粒度层次上分析和度量了概念漂移和概念耦合. 在决策表中定义了全粒度正区域，分析了决策表中所有概念的可能变化和概念漂移. 结合属性约简、核属性和概念漂移等概念，探究了全粒度粗糙集和全粒度正区域的性质. 讨论和分析了全粒度粗糙集的认识论意义. 全粒度粗糙集和全粒度正区域在一定程度上能够表示人类认识的复杂性、多变性和不确定性，也能够在一定程度上表示人类认识粒度的可变性，从一个粒度转跳到另一个粒度，转跳顺畅，毫无困难.

进一步研究为，在全粒度粗糙集中运用更多的粒计算、粗糙集不确定性分析方法和指标，分析和度量数据流、大数据或知识系统中隐藏的不确定性，并将结果应用于集成分类器与人类智能的模拟.

参　考　文　献

[1] Babcock B, Babu S, Dater M, et al. Models and issues in data stream systems//Proceedings of the 21st ACM SIGACT-SIGMOD-SIGART Symposium on Principles Database Systems. New York, USA, 2002: 1-30

[2] Wang Tao, Li Zhou-Jun, Yan Yue-Jin, et al. A survey of classification of data streams. Journal of Computer Research and Development, 2007, 44(11): 1809-1815(in Chinese)
(王涛，李舟军，颜跃进等. 数据流挖掘分类技术综述. 计算机研究与发展, 2007, 44(11): 1809-1815)

[3] Xu Wen-Hua, Qin Zheng, Chang Yang. Semi-supervised learning based ensemble classifier for stream data. Pattern

Recognition and Artificial Intelligence, 2012, 25(2): 292-299(in Chinese)
(徐文华, 覃征, 常扬. 基于半监督学习的数据流集成分类算法. 模式识别与人工智能, 2012, 25(2): 292-299)

[4] Du L, Song Q, Jia X. Detecting concept drift: An information entropy based method using an adaptive sliding widow. Intelligent Data Analysis, 2014, 18(3): 337-364

[5] Yeon K, Song M S, Kim Y, et al. Model averaging via penalized regression for tracking concept drift. Journal of Computational & Graphical Statistics, 2012, 19(19): 457-473

[6] Mirza B, Lin Z, Liu N. Ensemble of subset online sequential extreme learning machine for class imbalance and concept drift. Neurocomputing, 2015, 149(PA): 316-329

[7] Sun Xue, Li Kun-Lun, Han Lei, et al. Construction of the concept drift detection model based on the information entropy of feature distribution and dynamic weighting algorithm. Acta Eletronica Sinica, 2015, 43(7): 1356-1361(in Chinese)
(孙雪, 李昆仑, 韩蕾等. 基于特征项分布的信息熵及特征动态加权概念漂移检测模型. 电子学报, 2015, 43(7): 1356-1361)

[8] Hobbs J R. Granularity//Proceedings of the 9th International Joint Conference on Artificial Intelligence. Los Angeles, USA, 1985: 432-435

[9] Lin T Y. Granular computing. Announcement of the BASIC Special Interest Group on Granular Computing. California, USA: Berkeley, 1997

[10] Zadel L A. Fuzzy sets. Information and Control, 1965, 8(3): 338-353

[11] Pawlak Z. Rough sets. International Journal of Computer and Information Sciences, 1982, 11(5): 341-356

[12] Pawlak Z. Rough Sets — Theoretical Aspect of Reasoning about Data. Dordrecht, Holland: Kluwer Academic Publishers, 1991

[13] Wang Guo-Yin. Rough Set Theory and Knowledge Acquisition. Xi'an: Xi'an Jiaotong University Press, 2001(in Chinese)
(王国胤. Rough 集理论与知识获取. 西安: 西安交通大学出版社, 2001)

[14] Zhang Bo, Zhang Ling. Theories and Applications for Problem Solving. Beijing: Tsinghua University Press, 1990 (in Chinese)
(张钹, 张铃. 问题求解理论及应用. 北京: 清华大学出版社, 1990)

[15] Li De-Yi, Meng Hai-Jun, Shi Xue-Mei. Membership clouds and Membership cloud generators. Journal of Computer Research and Development, 1995, 32(6): 16-18(in Chinese)
(李德毅, 孟海军, 史雪梅. 隶属云和隶属云发生器. 计算机研究与发展, 1995, 32(6): 16-18)

[16] Deng Da-Yong, Chen Lin. Parallel reducts and F-rough sets//Miao Duo-Qian, Wang Guo-Yin, Yao Yi-Yu, et al, eds. Cloud Model and Granular Computing. Beijing: Science Press, 2012: 210-228(in Chinese)
(邓大勇, 陈林. 并行约简与 F-粗糙集//苗夺谦, 王国胤, 姚一豫等编. 云模型与粒计算. 北京: 科学出版社, 2012: 210-228)

[17] Chen Lin. Parallel Reducts and Decision in Various Levels of Granularity[M. S. dissertation]. Zhejiang Normal University, Jinhua, Zhejiang, 2013(in Chinese)
(陈林. 粗糙集中不同粒度层次下的并行约简及决策[硕士学位论文]. 浙江师范大学, 浙江, 金华, 2013)

[18] Cao Fuyuan, Huang Joshua Zhexue. A concept-drifting detection algorithm for categorical evolving data//Proceedings of the 17th Pacific-Asia Conf on Knowledge Discovery and Data Mining. Berlin, Germany, 2013: 485-496

[19] Deng Da-Yong, Pei Ming-Hua, Huang Hou-Kuan. The F-rough sets approaches to the measures of concept drift. Journal of Zhejiang Normal University: Natural Sciences, 2013, 36(3): 303-308(in Chinese)
(邓大勇, 裴明华, 黄厚宽. F-粗糙集方法对概念漂移的度量. 浙江师范大学学报: 自然科学版, 2013, 36(3): 303-308)

[20] Deng Da-Yong, Xu Xiao-Yu, Huang Hou-Kuan. Concept drifting detection for categorical evolving data based on parallel reducts. Journal of Computer Research and Development, 2015, 52(5): 1071-1079(in Chinese)
(邓大勇, 徐小玉, 黄厚宽. 基于并行约简的概念漂移探测. 计算机研究与发展, 2015, 52(5): 1071-1079)

[21] Deng Da-Yong, Miao Duo-Qian, Huang Hou-Kuan. Analysis of concept drifting and uncertainty in an information system. Journal of Computer Research and Development, 2016, 53(11): 2607-2612(in Chinese)
(邓大勇, 苗夺谦, 黄厚宽. 信息表中概念漂移与不确定性分析. 计算研究与发展, 2016, 53(11): 2607-2612)

[22] Pawlak Z, Skowron A. Rough membership functions//Yager R R, Fedrizzy M, Kacprzyk J eds. Adances in the Dempster Shafer Theory of Evidence. New York, USA: John Wiley, 1994: 251-271

[23] Miao Duo-Qian, Hu Gui-Rong. A heuristic algorithm for reduction of knowledge. Journal of Computer Research and Development, 1999, 36(6): 681-684(in Chinese)
(苗夺谦, 胡桂荣. 知识约简的一种启发式算法. 计算机研究与发展, 1999, 36(6): 681-684)

[24] Wang Guo-Yin, Yu Hong, Yang Da-Chun. Decision table reduction on conditional information entropy. Chinese Journal of Computers, 2002, 25(7): 759-766(in Chinese)
(王国胤, 于洪, 杨大春. 基于条件信息熵的决策表约简. 计算机学报, 2002, 25(7): 759-766)

[25] Yang Ming. Approximate reduction based on conditional information entropy in decision tables. Acta Eletronica Sinica, 2007, 35(11): 2156-2160(in Chinese)
(杨明. 决策表中基于条件信息熵的近似约简. 电子学报, 2007, 35(11): 2156-2160)

[26] Liang J Y, Chin K S, Dang C Y. A new method for measuring uncertainty and fuzziness in rough set theory. International Journal of General Systems, 2002, 31(4): 331-342

[27] Liang Ji-Ye, Li De-Yu. Uncertainty and Knowledge Acquisition in Information Systems. Beijing: Science Press, 2005 (in Chinese)
(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)

[28] Wang Guo-Yin, Zhang Qing-Hua. Uncertanty of rough sets in different knowledge granularities. Chinese Journal of Computers, 2008, 31(9): 1588-1598(in Chinese)
(王国胤，张清华. 不同知识粒度下粗糙集的不确定性研究. 计算机学报，2008，31(9)：1588-1598)

[29] Lin Jia-Yi, Peng Hong, Zheng Qi-Lun. A new algorithm for value reduction based on rough set. Computer Engineering, 2003, 29(4): 70-71(in Chinese)
(林嘉宜，彭宏，郑启伦. 一种新的基于粗糙集的值约简算法. 计算机工程，2003，29(4)：70-71)

[30] Qian Y H, Liang J Y, Pedrycz W, et al. Positive approximation: An accelerator for attribute reduction in rough set theory. Artificial Intelligence, 2010, 174: 597-618

DENG Da-Yong, born in 1968, Ph. D., associate professor. His main research interests include rough sets, granular computing and data mining.

LU Ke-Wen, born in 1992, M. S. candidate. His main research interests include rough sets, data mining.

MIAO Duo-Qian, born in 1964, Ph. D., professor, Ph. D. supervisor. His main research interests include rough sets, granular computing, data mining, computational intelligence and image processing.

HUANG Hou-Kuan, born in 1940, professor, Ph. D. supervisor. His main interests include artificial intelligence, data mining and multi-agent system.

Background

There exist many rough set models, including Pawlak rough sets, various precision rough sets, covering rough sets, three-way decisions, neighborhood rough sets, multi-granulation rough sets and S-rough sets etc. They focus on uncertainty and vagueness. However, it is hard for these models of rough sets to deal with the change of uncertainty. Some researchers try to handle the change of uncertainty with incremental algorithms, but to some extent incremental methods hide the change of uncertainty. F-rough sets are the first dynamical model of rough sets, which have been employed to detect concept drift in data stream and concept drift caused by spaces or conditions. In order to address the change of uncertainty and concept drift, these models of rough sets should be improved, and indexes of uncertainty should also be improved.

This paper presents a new rough set model called entire-granulation rough sets, which extends Pawlak rough set model from a rough set to a family of rough sets in a knowledge system (information system, decision system) from the viewpoints of F-rough sets. Entire-granulation conditional attribute reducts are defined. Concept drifting is detected among entire-granulation rough sets. To some extent, entire-granulation rough sets can express complexity, uncertainty, diversity, hierarchy and dynamic in the process of human cognition. With the help of quantum computing, the model of entire-granulation rough sets can transform one type of granulation to another fluently.