

信息表中概念漂移与不确定性分析

邓大勇^{1,2,4} 苗夺谦² 黄厚宽³

¹(浙江师范大学数理与信息工程学院 浙江金华 321004)

²(同济大学电子与信息工程学院 上海 201804)

³(北京交通大学计算机与信息技术学院 北京 100044)

⁴(浙江师范大学行知学院 浙江金华 321004)

(dayongd@163.com)

Analysis of Concept Drifting and Uncertainty in an Information Table

Deng Dayong^{1,2,4}, Miao Duoqian², and Huang Houkuan³

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

²(School of Electronics and Information, Tongji University, Shanghai 201804)

³(School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044)

⁴(Xingzhi College, Zhejiang Normal University, Jinhua, Zhejiang 321004)

Abstract Concept drifting detection is one of hot topics in data stream mining, and analysis of uncertainty is dominant in rough set theory. Combined with the ideas of data stream, concept drifting, rough sets and F-rough sets, a lot of concepts such as concept drifting of upper approximation, concept drifting of lower approximation, concept coupling of upper approximation and concept coupling of lower approximation etc are defined. The change of concepts in an information system is analyzed with these definitions. With the positive region, integral concept drifting, integral concept coupling are defined. The analysis and measurement for the change of concept uncertainty are conducted. From the view of epistemology, the concept of cognition convergence is defined from the ways of idealism and realism. It provides heuristic information for realizing the world of human beings from the viewpoints of granular computing and rough sets.

Key words rough sets; concept drift; attribute reduction; concept coupling; upper and lower approximation

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。结合数据流、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,定义了上下近似概念漂移、上下近似概念耦合等概念,据此分析了信息表内概念随着属性而变化的特点。以正区域为工具,定义了决策表内概念漂移、概念耦合等概念,分析了决策表内整体概念随属性变化而变化。在认识论方面,从理想和现实2方面定义了认识收敛,从粒计算、粗糙集的角度对人类认识世界的方式进行了探讨。

关键词 粗糙集;概念漂移;属性约简;概念耦合;上下近似

中图法分类号 TP18

收稿日期:2015-09-06;修回日期:2016-02-05

基金项目:国家自然科学基金项目(61473030,61572442,61203247,61273304,61573259,61472166);浙江省自然科学基金项目(LY15F020012, LY13F020016)

This work was supported by the National Natural Science Foundation of China (61473030, 61572442, 61203247, 61273304, 61573259, 61472166) and the Natural Science Foundation of Zhejiang Province of China (LY15F020012, LY13F020016).

现实中的数据往往随着时间的变化而变化,例如证券交易数据、微博、视频、传感器数据等,这种类型的数据称为数据流^[1].数据流具有按照时间顺序排列、快速变化、海量甚至无限并且可能出现概念漂移现象等特征^[2-3].数据流挖掘是当前数据挖掘研究的一个热点,数据流分类和概念漂移探测是数据流挖掘的主要研究方向.

粒计算^[4-5]是人类智能处理问题的思维方式,也是处理不确定性问题的方法.粒计算的主要方法有模糊集^[6]、粗糙集^[7-9]、商空间^[10]和云模型^[11]等.粗糙集理论^[7-9]是一种处理不精确、不完全、含糊数据的有效数学工具,是数据挖掘和分类的重要方法.传统的粗糙集理论不太适合研究海量的、动态变化的数据,也不太适合研究数据流;F-粗糙集^[12-13]将粗糙集理论从单个信息表或决策表推广到多个,比较适合研究动态变化的数据,能够研究数据流和概念漂移.

利用粗糙集理论研究数据流和概念漂移比较少见.文献^[14-15]利用粗糙集的上下近似的变化去度量概念漂移;文献^[16]把每个滑动窗口看成是一个决策子表,利用并行约简的方法整体删除冗余属性,通过比较不同子表之间的属性重要性变化探测概念漂移.

粗糙集一个非常大的优势在于不确定性分析.研究者们提出了上下近似^[7-9]、隶属度^[17]、信息熵^[18]、条件熵^[19-20]、粗糙熵、模糊熵^[21-22]等不确定性度量指标来刻画和描述数据的不确定性.其中最原始、最本质、最核心的不确定性分析和度量指标是上下近似.

本文结合数据流、概念漂移和粗糙集、F-粗糙集的基本观点、基本方法,分析了信息表内信息粒度的概念变化和整个信息表内概念的整体变化.首先利用上下近似分析了信息表内单个概念的变化,定义了上下近似概念漂移与概念耦合等指标,在信息粒度变化的基础上度量概念随着属性而变化的特点;其次,在整个信息表内,用正区域的变化度量整个信息表或决策表内整体概念随属性而变化的特性;最后,分别定义了决策表内和信息表内认识收敛的概念,分析了概念漂移、概念耦合等指标的认识论意义.

1 基础知识

本节简单介绍粗糙集^[7-9]的相关基本知识.

$IS=(U,A)$ 是一个信息系统,其中 U 是论域, A 是论域 U 上的条件属性集.对于每个属性 $a \in A$ 都对应着一个函数 $a:U \rightarrow V_a, V_a$ 称为属性 a 的值域, U 中每个元素称为个体、对象或行.

对于每一个属性子集 $B \subseteq A$ 和任何个体 $x \in U$ 都对应着一个信息函数:

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

B -不分明关系(或称为不可区分关系)定义为

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

任何满足关系 $IND(B)$ 的2个元素 x, y 都不能由属性子集 B 区分, $[x]_B$ 表示由 x 引导的 $IND(B)$ 等价类.

对于信息系统 $IS=(U,A)$ 、属性子集 $B \subseteq A$ 和论域子集 $X \subseteq U$,上下近似与边界线的个体表示为

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

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

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

上下近似及边界线的信息粒度表示为

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

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

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

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

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

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 1^[7-9]. 在决策系统 $DS=(U,A,d)$ 中, $B \subseteq A$ 是 DS 的约简 iff $B \subseteq A$ 满足2个条件:

- 1) $POS_B(d) = POS_A(d)$;
- 2) 对于任意 $S \subset B$,有 $POS_S(d) \neq POS_B(d)$.

定义 2^[23]. 在决策系统 $DS=(U,A,d)$ 中, $U/\{d\} = \{Y_1, Y_2, \dots, Y_p\}$,对于任意 $a \in A$,有 $U/\{a\} = \{[x_1]_{(a)}, [x_2]_{(a)}, \dots, [x_n]_{(a)}\}$,则 $B \subseteq A$ 是 $Y \in U/\{d\}$ 的值约简 iff $B \subseteq A$ 满足2个条件:

- 1) $\bigcup_{b \in B} U/\{b\} \subseteq Y$;
- 2) 对于任意 $s \in B$ 有 $\bigcup_{b \in B - \{s\}} U/\{b\} \not\subseteq Y$.

2 信息粒度的概念漂移与不确定性分析

一个概念,它既可能用外延表示,也可能用内涵表示.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. 本节我们将研究概念的上下近似在同一个信息表中的变化,即概念漂移与概念耦合.

定理 1^[24]. 在一个信息表 $IS=(U, A)$ 中, 对于 $\emptyset \subset B_1 \subseteq B_2 \subseteq A$ 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 1. 在一个信息表 $IS=(U, A)$ 中, 对于 $\emptyset \subset B_1 \subseteq B_2 \subseteq A$ 和 $X \subseteq U$, 有 $BN_{B_2}(X) \subseteq BN_{B_1}(X)$.

我们将文献[15]中的度量概念漂移、概念耦合等指标进行改进,使之能更好地度量信息系统中概念的变化.

定义 3. 设信息表 $IS=(U, A)$ 中, $\emptyset \subset B_1 \subseteq B_2 \subseteq A$ 和 $X \subseteq U$, 则概念 X 相对于 B_1, B_2 的上下近似漂移定义为

$$\begin{aligned} \underline{\Delta}_{1,2}(X) &= \overline{B_1}(IS, X) - \overline{B_2}(IS, X), \\ \overline{\Delta}_{1,2}(X) &= \underline{B_2}(IS, X) - \underline{B_1}(IS, X), \end{aligned}$$

其中, “-”为集合减法; $|\underline{\Delta}_{1,2}(X)|, |\overline{\Delta}_{1,2}(X)|$ 分别称为概念 X 相对于 B_1, B_2 的上下近似漂移量.

定义 4. 设信息表 $IS=(U, A)$ 中, $\emptyset \subset B_1 \subseteq B_2 \subseteq A$ 和 $X \subseteq U$, 则概念 X 相对于 B_1, B_2 的上下近似耦合度定义为

$$\begin{aligned} \underline{c}_{1,2}(X) &= \frac{|\overline{B_2}(IS, X)|}{|\overline{B_1}(IS, X)|}, \\ \overline{c}_{1,2}(X) &= \frac{|\underline{B_1}(IS, X)|}{|\underline{B_2}(IS, X)|}, \end{aligned}$$

其中, $|\cdot|$ 表示“ \cdot ”的势.

概念 X 相对于 B_1, B_2 的上下近似漂移度定义为

$$\begin{aligned} \underline{d}_{1,2}(X) &= 1 - \overline{c}_{1,2}(X), \\ \overline{d}_{1,2}(X) &= 1 - \underline{c}_{1,2}(X). \end{aligned}$$

定理 2. 设 $DS=(U, A, d)$ 是一个决策系统, $B_1 \subseteq A$ 是一个约简, 则对于任意的 $B_1 \subseteq B_2 \subseteq A$ 和任意概念 $X=\{x|d(x)=d_1 \wedge x \in U\}$ (其中 d_1 为常数), 有:

$$\begin{aligned} \underline{\Delta}_{1,2}(X) &= \underline{B_2}(IS, X) - \underline{B_1}(IS, X) = \emptyset, \\ \overline{c}_{1,2}(X) &= \frac{|\underline{B_1}(IS, X)|}{|\underline{B_2}(IS, X)|} = 1, \\ \underline{d}_{1,2}(X) &= 1 - \overline{c}_{1,2}(X) = 0. \end{aligned}$$

证明. 我们只证明第 1 个公式, 后面 2 个公式由第 1 个公式立即推得.

反设存在概念 $X=\{x|d(x)=d_1 \wedge x \in U\}$ 使得 $\underline{\Delta}_{1,2}(X) = \underline{B_2}(IS, X) - \underline{B_1}(IS, X) \neq \emptyset$, 根据定理 1 有, $POS_{B_1}(d) \neq POS_A(d)$, 与 B_1 是 DS 的约简矛盾.

证毕.

定理 3. 设 $DS=(U, A, d)$ 是一个决策系统, $B_1 \subseteq A$ 是一个约简, 则对于任意的 $B_0 \subset B_1 \subseteq A$, 则存在概念 $X=\{x|d(x)=d_1 \wedge x \in U\}$ 使得:

$$\begin{aligned} \underline{\Delta}_{0,1}(X) &= \underline{B_1}(IS, X) - \underline{B_0}(IS, X) \neq \emptyset, \\ \overline{c}_{0,1}(X) &= \frac{|\underline{B_0}(IS, X)|}{|\underline{B_1}(IS, X)|} < 1, \\ \underline{d}_{0,1}(X) &= 1 - \overline{c}_{0,1}(X) > 0. \end{aligned}$$

证明. 与定理 2 的证明方式类似, 我们只证明第 1 个公式, 后面 2 个公式由第 1 个公式立即推得.

反设对所有的概念 $X=\{x|d(x)=d_1 \wedge x \in U\}$ 都有 $\underline{\Delta}_{0,1}(X) = \underline{B_1}(IS, X) - \underline{B_0}(IS, X) = \emptyset$, 则有 $POS_{B_0}(d) = POS_{B_1}(d) = POS_A(d)$, 根据约简的定义, B_0 是 DS 约简的超集, 与 B_1 是 DS 的约简矛盾.

证毕.

定理 4. 设 $DS=(U, A, d)$ 是一个决策系统, $X=\{x|d(x)=d_1 \wedge x \in U\}$ 是一个概念, $B_1 \subseteq A$ 是 X 的一个值约简, 则:

$$\begin{aligned} \underline{\Delta}_{1,A}(X) &= \underline{A}(IS, X) - \underline{B_1}(IS, X) = \emptyset, \\ \overline{c}_{1,A}(X) &= \frac{|\underline{B_1}(IS, X)|}{|\underline{A}(IS, X)|} = 1, \\ \underline{d}_{1,A}(X) &= 1 - \overline{c}_{1,A}(X) = 0. \end{aligned}$$

证明. 根据值约简的定义以及相应的概念漂移、概念耦合等定义, 立得上述结论.

证毕.

例 1. 设 $DS_1=(U, A, d)$ 是决策表, 如表 1 所示. 其中, a, b, c 是条件属性, d 是决策属性.

Table 1 A Decision System DS_1

表 1 决策表 DS_1

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

令 $X=\{x|d(x)=0, x \in U\}, B_0=\{a\}, B_1=\{a, b\}$, 则:

$$\begin{aligned} \underline{\Delta}_{0,1}(X) &= \underline{B_1}(X) - \underline{B_0}(X) = \\ & \{y_1, y_4\} - \{y_1, y_4\} = \emptyset; \\ \overline{\Delta}_{0,1}(X) &= \overline{B_0}(X) - \overline{B_1}(X) = \\ & \{y_1, y_2, y_3, y_4, y_5, y_6\} - \{y_1, y_4, y_5, y_6\} = \{y_2, y_3\}; \\ \overline{c}_{0,1}(X) &= \frac{|\underline{B_0}(X)|}{|\underline{B_1}(X)|} = \frac{2}{2} = 1; \\ \underline{c}_{0,1}(X) &= \frac{|\overline{B_1}(X)|}{|\overline{B_0}(X)|} = \frac{4}{6} = \frac{2}{3}; \end{aligned}$$

$$\begin{aligned} \frac{d_{0,1}(X)}{d_{0,1}(X)} &= 1 - \frac{c_{0,1}(X)}{c_{0,1}(X)} = 0; \\ \frac{d_{0,1}(X)}{d_{0,1}(X)} &= 1 - \frac{c_{0,1}(X)}{c_{0,1}(X)} = \frac{1}{3}; \end{aligned}$$

容易看出, B_0 是决策表 DS_1 的约简, 定理 1, 2, 3, 4 都成立.

3 决策表整体的概念漂移与不确定性分析

第 2 节讨论是针对信息表或决策表内单个概念的概念漂移与耦合, 对于整个决策表或信息表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.

定理 5^[24]. 设 $DS=(U, A, d)$ 是一个决策系统, $\emptyset \subset B_1 \subseteq B_2 \subseteq A$, 则有 $POS_{B_1}(d) \subseteq POS_{B_2}(d) \subseteq POS_A(d)$.

根据定理 5 将文献[15]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 5. 设 $DS=(U, A, d)$ 是一个决策系统, $\emptyset \subset B_1 \subseteq B_2 \subseteq A$, 则决策表中相对于 B_1, B_2 的概念漂移定义为

$$\Delta_{1,2} = POS_{B_2}(d) - POS_{B_1}(d).$$

定义 6. 设 $DS=(U, A, d)$ 是一个决策系统, $\emptyset \subset B_1 \subseteq B_2 \subseteq A$, 则决策表中相对于 B_1, B_2 的概念耦合度定义为

$$c_{1,2} = \frac{|POS_{B_1}(d)|}{|POS_{B_2}(d)|}.$$

定义 7. 设 $DS=(U, A, d)$ 是一个决策系统, $\emptyset \subset B_1 \subseteq B_2 \subseteq A$, 则决策表中相对于 B_1, B_2 的概念漂移度定义为

$$d_{1,2} = \frac{|POS_{B_2}(d) - POS_{B_1}(d)|}{|POS_{B_2}(d)|} = 1 - c_{1,2}.$$

定理 6. 设 $DS=(U, A, d)$ 是一个决策系统, $B_1 \subseteq A$ 是一个约简, 则对于任意的 $B_1 \subseteq B_2 \subseteq A$, 有:

$$\Delta_{1,2} = POS_{B_2}(d) - POS_{B_1}(d) = \emptyset,$$

$$c_{1,2} = \frac{|POS_{B_1}(d)|}{|POS_{B_2}(d)|} = 1,$$

$$d_{1,2} = \frac{|POS_{B_2}(d) - POS_{B_1}(d)|}{|POS_{B_2}(d)|} = 1 - c_{1,2} = 0.$$

证明. 根据定理 5、定义 5~7 及属性约简的定义, 立得上述结论. 证毕.

定理 7. 设 $DS=(U, A, d)$ 是一个决策系统, $B_1 \subseteq A$ 是一个约简, 则对于任意的 $B_0 \subset B_1 \subseteq A$,

$$\Delta_{0,1} = POS_{B_1}(d) - POS_{B_0}(d) \neq \emptyset,$$

$$c_{0,1} = \frac{|POS_{B_0}(d)|}{|POS_{B_1}(d)|} < 1,$$

$$d_{0,1} = \frac{|POS_{B_1}(d) - POS_{B_0}(d)|}{|POS_{B_1}(d)|} = 1 - c_{0,1} > 0.$$

证明. 根据定理 5、定义 5~7 及属性约简的定义, 立得上述结论. 证毕.

例 2. 设 $DS_2=(U, A, d)$ 是决策表, 如表 2 所示. 其中, a, b, c 是条件属性, d 是决策属性.

Table 2 A Decision System DS_2

表 2 决策表 DS_2

U	a	b	c	d
x_1	0	0	1	1
x_2	1	1	0	1
x_3	0	1	0	0
x_4	1	1	0	1

令 $B_1 = \{a\}, B_2 = \{a, b\}$, 则:

$$POS_{B_1}(d) = \{x_2, x_4\};$$

$$POS_{B_2}(d) = \{x_1, x_2, x_3, x_4\};$$

$$\Delta_{1,2} = POS_{B_2}(d) - POS_{B_1}(d) = \{x_1, x_3\};$$

$$c_{1,2} = \frac{|POS_{B_1}(d)|}{|POS_{B_2}(d)|} = \frac{2}{4} = \frac{1}{2};$$

$$d_{1,2} = 1 - c_{1,2} = \frac{1}{2};$$

在决策表 DS_2 中, B_2 是它的一个约简, 容易看出, 定理 5~7 都成立.

4 概念漂移与不确定性分析的认识论意义

粗糙集理论认为“知识就是分类”, 区分不同的物体是人类知识的体现, 在此过程中也需要知识, 也就是说知识需要知识来表达. 但不同的知识表达不一样, 人类的认识是一个过程, 在这个认识过程中选取什么特征来表达知识? 选取多少特征? 到什么时候为止? 这些问题对于人类来说目前都是一个自发的直觉过程, 缺乏理性思考. 继承文献[25]的思想, 下面我们利用粗糙集和概念耦合的思想来努力回答这些问题.

在决策系统 $DS=(U, A, d)$ 中, 假定条件属性集 $A_1 \subset A_2 \subset \dots \subset A_n \subset \dots = A$ 是一个不断变化的过程, 也是一个认识不断深入的过程, 我们通过 A 来表达和认识 d .

认识收敛有 2 条标准:

- 1) $POS_A(d) = U$;

$$2) c_{n-1,n} = \frac{|POS_{A_{n-1}}(d)|}{|POS_{A_n}(d)|} = 1.$$

这 2 条标准也是认识收敛的定义. 标准 1 表明决策系统 $DS = (U, A, d)$ 是一致的, 所有概念的边界区域为空, 它们的上近似等于下近似; 标准 2 表明增加的属性 $A_n - A_{n-1}$ 对于区分表中的个体不起作用. 标准 1 是理想的标准, 标准 2 是现实的标准, 这是因为现实世界中并不是每个概念都是精确的, 很多概念都是含糊不清、边界线不为空的.

在信息系统 $IS = (U, A)$ 中, 因为没有决策属性 d 的约束, 认识收敛的标准定义如下:

3) 对于任意的概念 $X \subseteq U$, 有:

$$\overline{c_{n-1,n}}(X) = \frac{|\overline{A_n}(IS, X)|}{|\overline{A_{n-1}}(IS, X)|} = 1$$

且

$$c_{n-1,n}(X) = \frac{|A_{n-1}(IS, X)|}{|A_n(IS, X)|} = 1.$$

标准 3 表明从 A_{n-1} 到 A_n 对于信息系统 $IS = (U, A)$ 中的每一个概念都不会发生变化.

标准 1 是一个理想的标准, 在这个标准中所有的对象都被清晰地区分, 没有不确定、没有模糊、也没有粗糙; 标准 2 和标准 3 只是一个局部收敛的标准, 随着认识的进一步深入, 比如从 A_n 到 A_{n+1} 这些指标值也许不等于 1, 这时认识达到了一个新的高度. 例如: 长期以来人们识别某个人, 一般是根据相貌、体态、步态、声音等, 这种识别方式虽然有一定的误差, 但基本稳定. 只有到了近年来, 使用 DNA 技术才能彻底区分不同的人.

当然, 我们可以等价地用概念漂移度来定义和度量认识收敛, 这里不再赘述.

5 结论与展望

从粒计算、粗糙集和数据流、概念漂移的角度观察信息表, 以上下近似为工具, 本文定义了概念的上下近似漂移、上下近似耦合等概念, 分析了信息表内概念随属性的变化而变化的特性. 从单个概念和整个信息表或决策表 2 种不同的粒度层次上分析和度量了概念漂移和概念耦合. 从信息表和决策表的角度定义了认识收敛的概念, 指出其认识论意义.

进一步的研究有: 运用更多的粒计算、粗糙集不确定性分析方法和指标, 分析和度量数据流或信息表中隐藏的不确定性, 并将结果应用于集成分类器和数据流分类.

参 考 文 献

- [1] Babcock B, Babu S, Dater M, et al. Models and issues in data stream systems [C] //Proc of the 21st ACM SIGACT-SIGMOD-SIGART Symp on Principles Database Systems. New York: ACM, 2002: 1-30
- [2] Wang Tao, Li Zhoujun, Yan Yuejin, et al. A survey of classification of data streams [J]. 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[J]. 计算机研究与发展, 2007, 44(11): 1809-1815)
- [3] Xu Wenhua, Qin Zheng, Chang Yang. Semi-supervised learning based ensemble classifier for stream data [J]. 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[J]. 模式识别与人工智能, 2012, 25(2): 292-299)
- [4] Hobbs J R. Granularity [C] //Proc of the 9th Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 1985: 432-435
- [5] Lin T Y. Granular Computing [M]. Announcement of the BASIC Special Interest Group on Granular Computing, 1997
- [6] Zadeh L A. Fuzzy sets [J]. Information and Control, 1965, 8(3): 338-353
- [7] Pawlak Z. Rough sets [J]. International Journal of Computer and Information Sciences, 1982, 11(5): 341-356
- [8] Pawlak Z. Rough Sets—Theoretical Aspect of Reasoning about Data [M]. Dordrecht, the Netherland: Kluwer Academic Publishers, 1991
- [9] Wang Guoyin. Rough Set Theory and Knowledge Acquisition [M]. Xi'an: Xi'an Jiaotong University Press, 2001 (in Chinese)
(王国胤. 粗糙集理论与知识获取[M]. 西安: 西安交通大学出版社, 2001)
- [10] Zhang Bo, Zhang Ling. Theories and Applications for Problem Solving [M]. Beijing: Tsinghua University Press, 1990 (in Chinese)
(张钹, 张铃. 问题求解理论及应用[M]. 北京: 清华大学出版社, 1990)
- [11] Li Deyi, Meng Haijun, Shi Xuemei. Membership clouds and membership cloud generators [J]. 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[J]. 计算机研究与发展, 1995, 32(6): 16-18)
- [12] Deng Dayong, Chen Lin. Parallel Reducts and F-rough Sets [M] //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[M] //云模型与粒计算. 北京: 科学出版社, 2012: 210-228)

- [13] Chen Lin. Parallel reducts and decision in various levels of granularity [D]. Jinhua, Zhejiang: Zhejiang Normal University, 2013 (in Chinese)
(陈林. 粗糙集中不同粒度层次下的并行约简及决策[D]. 浙江金华: 浙江师范大学, 2013)
- [14] Cao Fuyuan, Huang Joshua Zhexue. A concept-drifting detection algorithm for categorical evolving data [G] // LNAI 7819; Proc of the 17th Pacific-Asia Conf on Knowledge Discovery and Data Mining. Berlin: Springer, 2013; 485-496
- [15] Deng Dayong, Pei Minghua, Huang Houkuan. The F-rough sets approaches to the measures of concept drift [J]. 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[J]. 浙江师范大学学报: 自然科学版, 2013, 36(3): 303-308)
- [16] Deng Dayong, Xu Xiaoyu, Huang Houkuan. Concept drifting detection for categorical evolving data based on parallel reducts [J]. 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[J]. 计算机研究与发展, 2015, 52(5): 1071-1079)
- [17] Pawlak Z, Skowron A. Rough membership functions [C] // Advances in the Dempster Shafer Theory of Evidence. New York: John Wiley, 1994; 251-271
- [18] Miao Duoqian, Hu Guirong. A heuristic algorithm for reduction of knowledge [J]. 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[J]. 计算机研究与发展, 1999, 36(6): 681-684)
- [19] Wang Guoyin, Yu Hong, Yang Dachun. Decision table reduction on conditional information entropy [J]. 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[J]. 计算机学报, 2002, 25(7): 759-766)
- [20] Yang Ming. Approximate reduction based on conditional information entropy in decision tables [J]. Acta Electronica 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[J]. 电子学报, 2007, 35(11): 2156-2160)
- [21] Liang J Y, Chin K S, Dang C Y. A new method for measuring uncertainty and fuzziness in rough set theory [J]. International Journal of General Systems, 2002, 31(4): 331-342
- [22] Liang Jiye, Li Deyu. Uncertainty and Knowledge Acquisition in Information Systems [M]. 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[M]. 北京: 科学出版社, 2005)
- [23] Lin Jiayi, Peng Hong, Zheng Qilun. A new algorithm for value reduction based on rough set [J]. 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[J]. 计算机工程, 2003, 29(4): 70-71)
- [24] Qian Y H, Liang J Y, Pedrycz W, et al. Positive approximation: An accelerator for attribute reduction in rough set theory [J]. Artificial Intelligence, 2010, 174: 597-618
- [25] Deng Dayong, Jiang Feng, Liu Qing. Data reduction & machine learning based on rough set approach [J]. Computer and Modernization, 2002 (1): 21-23 (in Chinese)
(邓大勇, 江峰, 刘清. 基于 Rough 集方法的数据约简与机器学习[J]. 计算机与现代化, 2002 (1): 21-23)



Deng Dayong, born in 1968. PhD and associate professor. His main research interests include rough sets, granular computing and data mining.



Miao Duoqian, born in 1964. PhD, Professor and PhD supervisor. His main research interests include rough sets, granular computing, data mining, computational intelligence and image processing.



Huang Houkuan, born in 1940. Professor and PhD supervisor. His main research interests include computational intelligence, data mining and multi-agent system.