

## 基于面向属性泛化及信息增益的数据挖掘方法研究

张文宇, 张铭华

(西安邮电学院 管理系, 陕西 西安 710061)

(zwy888459@sina.com)

**摘 要:** 针对面向属性的归纳方法及粗糙集方法对知识粒性连续性的特点, 将两者有机结合, 利用面向属性归纳方法对数据进行泛化, 再用属性的信息增益技术寻找泛化属性之间的数据依赖关系, 能快速地在数据集中挖掘分类规则。将其应用于经典的仿真算例中, 仿真结果合理、可靠。

**关键词:** 数据挖掘; 粗糙集; 属性泛化; 信息增益

**中图分类号:** TP311.13 **文献标识码:** A

## Research on data mining method based on attribute-oriented generalization and information gain

Zhang Wen-yu, Zhang Ming-hua

(Department of Management, Xi'an Post and Telecommunication College, Xi'an Shaanxi 710061, China)

**Abstract:** According to the characteristics of attribute-oriented generalization method and knowledge granule successive based on rough set, after organically combining these two method, the classification rules could be rapidly mined in database. The method of attribute-oriented generalization was firstly used to generalize the data, and then the technology of information gain was utilized to find the dependency relation of data. The algorithm was applied in the classical simulation example and the result is rational and reliable.

**Key words:** data mining; rough set; attribute generalization; information gain

面向属性的归纳方法用于在关系数据库中发现知识, 这种方法没有分析属性数据间的依赖关系, 因而产生的规则不精确, 带有一定的冗余信息。粗糙集方法是用来分析全局属性依赖关系的一种工具, 但由于其计算过程是 NP-hard 问题, 因此不太适合于在实值型数据库中。这两种方法显然是不同的, 但它们的共同之处表现为对象都是由属性及属性值来表征的, 本文主要是寻找面向属性归纳和粗糙集方法中知识粒性的连接性, 即把两种方法的优点结合起来, 提出了一种新的 KDD 方法, 即先利用面向属性归纳方法对数据进行泛化, 然后利用属性的信息增益技术应用于泛化关系中来寻找泛化属性之间的数据依赖关系, 最后使用最小属性集来表示最后的泛化关系, 将最后的泛化关系转换成逻辑规则形式, 从而进行知识的表达和推理。

### 1 关系 DB 学习原理

#### 1.1 面向属性归纳原理

在数据库的知识表示中, 每个属性对应的概念层次树具有如下的特点: 树的根为属性名, 树的所有叶子节点对应所有原始数据, 并表示原始级别的概念; 以树的每个内节点为根的子树覆盖该内节点表示的概念所对应的所有实例。这样, 每个属性所对应的概念层次树就形成了对所有实例所构成的集合的不同精细程度的划分。对较高层次(离树根越近)的划分, 形成的等价类数量就较小, 表示概念的泛化程度也较高。在概念层次树中, 一个节点的父节点表示的概念即是该节点表示的节点的泛化, 一个节点的子节点表示的概念即是该节点表示的概念的一个转化。

面向属性的泛化过程是根据各个属性的概念层次树和属

性阈值进行泛化的, 它规定了关系 DB 在归纳概括后该属性上允许的不同属性值的最大个数, 如果某属性的不同属性值数目大于该属性的属性阈值, 该属性向上归纳概括一层, 如此反复, 直至不同属性值数目小于等于属性阈值, 数据库经过面向属性归纳后, 其中的记录会出现重复, 这就需要删除重复记录, 从而使关系 DB 中的记录数目大量减少, 因此面向属性归纳可以用于数据简化, 减少记录的过程。

此方法的不足之处表现为只考虑原始数据所提供的简单计数信息, 而忽略了抽象的依据来源于实例集的共性, 化简的规则应尽量包含整个实例集的共同特征和信息要素。在提升过程中所被提升的属性对象的选择往往会直接影响最后的结果。即在实际情况中, 用属性阈值作为控制进行数据简化可能造成数据的过度简化, 失去数据简化的意义。主要原因是属性的阈值控制是布尔型的, 某个属性的某个属性值只要有一个记录取到, 就增加该属性的不同属性值计数, 如果有多个记录取到, 也同样的方法处理。也就是说, 统计不同属性值数目时, 仅考虑属性值是否被记录取到, 而未考虑被多少记录取到, 将极少个记录取到的属性值和很多个记录取到的属性值同等考虑。但是在数据库中, 由于属性间存在一些显式或隐式的数据依赖关系, 致使记录在某个属性取某个属性值时, 在另外一个属性上也主要取某几个属性值。又由于记录间也可能存在一些关联, 致使不同的记录在某些属性上的取值相对近似或集中。这样, 数据库的某些取值就有可能存在着属性值和例外属性值之分。

本文同样根据各个属性的概念层次树, 将记录中的详细信息用树中相应的高层信息代替。在进行概念树提升时, 除了每个属性的属性阈值  $k_i (1 \leq i \leq m)$  控制外还增加了每个

属性的记录阈值  $r_i (1 \leq i \leq m)$  控制, 使控制由布尔型变成数量型。如果某属性的某个属性值含有的记录数目大于等于该属性的记录阈值, 增加该属性的属性值计数。如果该属性的主属性值数目大于该属性的属性阈值, 该属性向上归纳概括一层, 如此反复, 直至主属性值数目小于等于属性阈值。同样, 经过提升后的数据库记录会出现重复, 需要删除重复记录, 从而达到简化记录的数据简化目的。其主要的算法思想是对各个属性的概念层次树中的各个属性值从下到上、从左到右进行编号。同时对各个属性值设置一个计数器, 它可以取任意整数, 其绝对值表示该属性值含有的记录数目。如果它为负值, 表示该属性值是主属性值, 否则不是主属性值。每个属性对应一个  $L_i$  和  $A_i$ , 分别用于表示该属性的当前归纳层次和当前归纳层次上该属性的主属性值数目, 并建立一个与待处理数据库结构一致的临时数据库, 让其存放原始数据库中经过变换后的每一个记录。

## 1.2 信息熵原理

令关系型数据库 DB 含有  $m$  个决策类  $C_i (i = 1, \dots, m)$ ,  $s_i$  为属于  $C_i$  决策类的样本个数, 任意一个样本属于决策类  $C_i$  的概率是  $s_i/s$ ,  $s$  为所有样本集合。则一个属性  $A$  具有属性值集合  $\{a_1, a_2, \dots, a_v\}$  可以将所有样本空间按等价关系划分成子集  $\{s_1, s_2, \dots, s_v\}$ , 其中  $s_j$  为样本空间  $s$  中具有属性值  $a_j$  的样本子集。则定义一个属性所含有的期望信息为:

$$I(s_1, s_2, \dots, s_m) = - \sum_{i=1}^m \frac{s_i}{s} \log_2 \frac{s_i}{s}$$

令  $s_{ij}$  为  $s_j$  子集中属于  $C_i$  决策类的样本个数, 则可定义属性  $A$  的信息熵为:

$$E(A) = \sum_{j=1}^v \frac{s_{1j} + \dots + s_{mj}}{s} I(s_{1j}, \dots, s_{mj})$$

则属性  $A$  的信息增益可定义为:

$$\begin{aligned} \text{Gain}(A) &= I(s_1, s_2, \dots, s_m) - E(A) \\ &= - \sum_{i=1}^m \frac{s_i}{s} \log_2 \frac{s_i}{s} - \\ &\quad \sum_{j=1}^v \frac{s_{1j} + \dots + s_{mj}}{s} I(s_{1j}, \dots, s_{mj}) \end{aligned}$$

对于一个属性而言, 可通过计算每个属性的信息增益对不同属性进行重要性和相关性分析。

## 2 规则提取方法

大型关系 DB 通常包含大量显著属性值, 为了对每一决策类获得一个简单而精炼的决策规则, 应先对初始数据集进行泛化, 再进行属性相关性检验, 最后进行规则提取。

### 步骤 1: 面向属性泛化

泛化过程是针对每个属性具有大量的属性值, 可对它进行概念树提升, 即对每个属性值用其高层概念的属性值代替, 泛化的结果是: 不同的对象由于其泛化后是等价的, 因而可划分到同一个等价类里, 对于合并后的记录加一 Count 属性值, Count 表示合并到同一等价类中记录的个数, 反映了归纳过程中的量化信息。

泛化后 Count 值与合并前记录数相同, 注意到根据属性阈值和记录阈值有必要对某些属性进行多次概念树提升以便获得简化的决策类。

### 步骤 2: 基于信息熵的属性约减

利用信息熵进行属性的相关性分析, 给出一个属性重要性阈值, 当某一属性的信息增益小于给定的属性重要性阈值时, 则此属性可认为是冗余属性, 可被删除。其余属性为对决策起重要影响的相关属性。

### 步骤 3: 记录重组与合并

在同一决策类中, 如果两条记录仅有一个条件属性值不同, 则这两条记录可以合并为一条记录。如果被合并的属性值覆盖了这条条件属性的所有概念树的取值范围, 则这一条件属性可以被删除。

### 步骤 4: 逻辑规则转换

将约减关系表转换成逻辑规则。

Count 值是初始 DB 中支持某条规则的数目, 因为初始 DB 中每条记录分布于属性的各个可能值上, 若不考虑量化信息, 则不可能得到一条规则的可信价值。以下是利用量化信息确定规则可信度的方法。

此方法是将约减表中例外及噪声数据, 从规则提取中滤掉, 它通过计算合并约减表中每一条记录的频率, 并使用噪声滤波阈值将例外记录剔除掉 (主要针对没有经过概念树提升的记录而言)。

**定义 1** 有用记录的频率定义为合并后约减表中每条记录的 Count 值与所有记录 Count 值之和的比值。

$$r_{frequency} = \text{Count\_of\_the\_记录 } i / \sum_{i=1}^n (\text{Count\_of\_记录 } i)$$

噪声滤波阈值是用来剔除合并后约减表中噪声规则的最小值。

算法过程:

输入:

- ① 与决策属性相关的条件目标集  $R(A_i) (1 \leq i \leq n)$ ;
- ② 所需属性  $A_i$  的概念等级树  $H_i (i = 1, \dots, m)$ ;
- ③ 属性阈值  $k_i (i = 1, \dots, m)$  与记录阈值  $r_i (i = 1, \dots, m)$ ;
- ④ 决策属性名。

输出:

一套决策逻辑规则。

### 步骤 1: Attribute-oriented induction

begin

初始化  $H_i (i = 1, \dots, m)$  中的各个 Count,  $L_i, A_i$

while ( Tuple = Gettuple( DB,  $H_i (i = 1, \dots, m)$  ) )

For  $I = 1$  to  $m$  do

If tuple-att(i).count < 0

Then tuple-att(i).count = tuple-att(i).count - 1

Else tuple-att(i).count = tuple-att(i).count + 1

If tuple-att(i).count >  $r_i$  then

$A_i = A_i + 1$

tuple-att(i).count = -1

endif

endfor

Put tuple(tuple, temp\_relation)

Endwhile

For  $I = 1$  to  $m$  do

根据  $H_i (i = 1, \dots, m)$ , count,  $A_i, k_i$ , 确定  $L_i$ 。

Endfor

Prime DB = Alloc Prime DB( $H_i, L_i$ )

While( tuple = Get\_tuple( temp\_relation ) )

Insert\_tuple( tuple,  $T_i$ , Prime\_DB )

Endwhile

### 步骤 2: 寻找所需的相关属性最小集

Begin

Compute( attr,  $I(A)$  );

Compute( attr,  $E(A)$  );

Compute( attr,  $\text{Gain}(A)$  );

If  $\text{Gain}(A) > \mu$  then Choose( $A$ )

Else Delete( $A$ )

Endif

end

步骤3:使用属性约减及合并近似项,得到合并约减 DB

步骤4:转换为决策规则。

### 3 算例

Car relation<sup>[1]</sup>有如下属性构成:其中条件属性为 plate#, Make-model, type of fuel system (fuel)、engine displacement (disp), weight, number of cylinders (cyl), power, presence of turbocharge (t), compression ratio (comp), transmitter (tran) and 决策属性 mileage (m) 描述。为了简化决策表,可用以下缩写表示各属性:Pl, M-M, F, D, W, C, P, T, Co, Tr, M。

表1 car relation DB 表

Pl	M-M	F	D	W	C	P	T	Co	Tr	M
BCT89	F escort	E	M	876	6	H	Y	H	A	M
UI89P	D shadow	E	M	1100	6	H	N	M	M	M
P0967	F festival	E	M	1589	6	H	N	H	M	M
LKIPO	Corvette	E	M	987	6	H	N	M	M	M
IUTY5	D stealth	E	M	1096	4	H	N	H	M	M
ERTW3	F probe	E	M	867	6	H	N	M	M	M
TYUR4	F mustang	E	M	1197	4	H	N	H	M	M
0987U	D Daytona	E	M	798	4	H	Y	H	M	H
9876	C Le B	E	M	1056	4	M	N	M	M	M
UYTH	D sprite	E	M	1557	4	H	N	M	M	L
OPLSA	H civic	2B	S	786	4	L	N	H	M	H
KMN89	Fcort	2B	S	1098	4	L	N	H	M	M
LKOPL	Ftemple	2B	S	1187	4	M	N	H	A	M
WEQ54	Tcorolla	E	S	1023	4	L	N	H	M	H
PLMN	M 323	E	M	698	4	M	N	M	M	H
QAS45	D Dayton	E	M	1123	4	M	N	M	M	M
PLMJH	Helude	E	S	1094	4	H	Y	H	M	H
DSA32	To paso	2B	S	1023	4	L	N	M	M	H
GHF6U	C Corsica	E	M	980	4	H	Y	M	M	M
KNM87	C beretta	E	M	1600	6	H	N	M	A	L
IKLO9	C cavalier	E	M	1002	6	H	N	M	A	M
IKHTY	Cle B	E	M	1098	4	H	N	M	A	M
OPL87	M 626	E	S	1039	4	M	N	H	M	H
UYT34	C corsic	E	S	980	4	M	N	H	M	H
UYTB	C lumina	E	S	1000	4	M	N	H	M	M

第一个属性“Plate#”是数据表的主关键词,关系表中每一属性的关键值是有明显区别的,在一个属性的概念树上,若没有更高层次的概念,则属性值不能被泛化,它们应在泛化过程中被移去,同样,其他的候选关键属性及非关键属性也可用同样的条件除去,然后再检查其余的属性。对具有概念树的属性 Make-model 与 Weight 分别设它们的记录阈值和属性阈值为(5,2)及(2,3),在记录阈值和属性阈值的控制下对属性 Make-model 与 Weight 进行概念树提升可得如下归纳表2。

利用属性的信息增益对提升后的决策表进行属性相关性分析,可得到各个属性的重要性系数。

以属性 Make-model 为例,其信息增益的计算过程如下:

$$I(s_1, s_2, s_3) = -\frac{15}{25} \log_2 \frac{15}{25} - \frac{8}{25} \log_2 \frac{8}{25} - \frac{2}{25} \log_2 \frac{2}{25} = 1.26$$

对于 Make-model = “USA”

$$s_{11} = 15, s_{21} = 2, s_{31} = 2 \quad I(s_{11}, s_{21}, s_{31}) = 0.951$$

对于 Make-model = “Japan”

$$s_{12} = 0, s_{22} = 6, s_{32} = 0 \quad I(s_{11}, s_{21}, s_{31}) = 0$$

$$E(\text{Make-model}) = \frac{19}{25} \times I(s_{11}, s_{21}, s_{31}) +$$

$$\frac{6}{25} \times I(s_{11}, s_{21}, s_{31}) = 0.723$$

$$\text{Gain}(\text{Make-model}) = I(s_1, s_2, s_3) - E(\text{Make-model}) = 0.537$$

表2 car relation DB 概念泛化表

M-M	F	D	w	C	P	t	Co	Tt	M	Count
U	E	M	M	6	H	Y	H	M	M	1
U	E	M	M	6	H	N	M	M	M	3
U	E	M	M	6	H	N	M	A	M	1
U	E	M	M	6	M	N	M	M	M	1
U	E	M	M	6	H	N	H	M	M	2
U	E	M	L	4	H	Y	H	M	H	1
U	E	S	M	4	M	N	H	M	H	1
U	E	M	M	6	H	N	H	M	M	2
U	E	M	H	6	H	Y	H	A	L	1
U	E	M	M	6	M	N	M	M	M	1
J	2B	M	H	6	H	N	M	M	L	1
U	2B	S	L	4	L	N	H	M	H	1
U	2B	S	M	4	M	N	H	A	M	1
J	E	S	M	4	L	N	H	M	H	1
J	E	M	L	4	M	N	M	M	H	1
U	E	M	M	4	M	N	M	M	M	1
J	E	S	M	4	H	Y	H	A	H	1
J	2B	S	M	4	L	N	M	M	H	1
U	E	M	M	4	H	Y	M	M	M	1
U	E	M	M	6	H	N	M	M	M	1
J	E	S	M	4	M	N	H	A	H	1

表3 属性约减表

M-M	C	F	Co	W	M	Count
USA	6	E	H	M	M	4
USA	6	F	M	M	M	7
USA	4	E	H	L	H	1
USA	4	E	H	M	H	1
USE	4	E	M	M	M	1
USA	6	E	H	H	L	1
USA	6	E	M	H	L	1
USA	4	2B	H	M	M	1
USA	4	2B	M	M	H	1
Japan	4	2B	H	L	H	1
Japan	4	E	H	M	H	4
Japan	4	E	M	L	H	1
Japan	4	2B	M	M	H	1

表4 合并后的 Car DB 约减表

M-M	C	F	Co	W	M	Count
USA	6	E	※	M	M	11
USA	4	F	H	※	H	2
USA	4	E	M	M	M	1
USA	6	E	※	H	L	2
USE	4	2B	H	M	M	1
USA	4	2B	M	M	H	1
Japan	4	2B	H	L	H	1
Japan	4	E	H	M	H	4
Japan	4	E	M	L	H	1
Japan	4	2B	M	M	H	1

同理可计算出其他条件属性的信息增益,由此可对条件属性进行重要性排序,设属性重要性阈值0.2,删除不相关属

(下转第866页)

对一个不一致决策表(见表2)进行分析如下:

表2 不一致决策表

U	C			D
	a	b	c	d
$u_1$	1	0	2	-1
$u_2$	1	1	0	-1
$u_3$	1	1	0	0
$u_4$	1	0	2	-1
$u_5$	1	1	2	0
$u_6$	1	1	0	-1
$u_7$	0	1	1	1
$u_8$	0	1	0	1
$u_9$	0	1	0	0
$u_{10}$	0	1	1	1

表3 条件信息熵表

Information entropy	value
$H(D/a)$	0.203 0
$H(D/b)$	0.376 0
$H(D/c)$	0.372 2

表4 条件信息熵变化表

Information entropy	value
$H(D/C - \{a\})$	0.229
$H(D/C - \{b\})$	0.166
$H(D/C - \{c\})$	0.218

根据条件熵的定义可得条件信息熵表(见表3)和条件信息熵变化表(表4):

而  $H(D/C) = 0.143$ , 对于 CEBARKNC 算法, 得出任何属性均不可约简, 而对于本文的改进算法可得出约简  $\{b, c\}$ 。

### 3 结语

基于以上可各, 本文提出的算法比 CEBARKNC 算法更优: 从理论上讲, 本文的算法复杂度与 CEBARKNC 算法的复杂度是一致的, 而且也可以证明, 本文的算法包含了 CEBARKNC 算法。但是, 关于该算法对最小约简的完备性问题还需从理论上作进一步的探讨。

#### 参考文献:

- [1] PAWLAK Z. Rough Set-Theoretical Aspect of Reasoning About Data [M]. Kluwer Academic Pub, 1991.
- [2] CHEN B, ZHOU MT. A Pure Mereological Approach to Roughness [A]. 9th International Conference: Rough Sets, Fuzzy Sets, Data Mining And Granular Computing(RSFDGrC) [C]. Chongqing, China, 2003. 430 - 436.
- [3] KOSTEK B, SZCZUKO P, ZWAN P. Processing of Musical Data Employing Rough Sets and Artificial Neural Networks[A]. RSCTC 2004 [C]. Uppsala, Sweden, 2004. 539 - 548.

- [4] LI YC, FANG TJ. Rough Set Methods for Constructing Support Vector Machines [A]. 9th International Conference: Rough Sets, Fuzzy Sets, Data Mining And Granular Computing(RSFDGrC) [C]. Chongqing, China, 2003. 334 - 338.
- [5] LAZARECK L, RAMANNA S. Classification of Swallowing Sound Signals: A Rough Set Approach [A]. 4th International Conference, RSCTC 2004 [C]. Uppsala, Sweden, 2004. 679 - 684.
- [6] BAZAN JG, SZCZUKA MS, WOJNA A, et al. On the Evolution of Rough Set Exploration System [A]. 4th International Conference, RSCTC 2004 [C]. Uppsala, Sweden, 2004. 592 - 601.
- [7] ONG C-S, HUANG J-J, TZEENG G-H. Using Rough Set Theory for Detecting the Interaction Terms in a Generalized Logit Model [A]. 4th International Conference, RSCTC 2004 [C]. Uppsala, Sweden, 2004. 624 - 629.
- [8] LIU XM, HUANG HK, XU WX. A Contribution to Decision Tree Construction Based on Rough Set Theory [A]. 4th International Conference, RSCTC 2004 [C]. Uppsala, Sweden, 2004. 637 - 642.
- [9] XIAO JM, ZHANG TF. New Rough Set Approach to Knowledge Reduction in Decision Table [A]. Proceeding of the International Conference on Machine Learning and Cybernetics [C]. Shanghai, China, 2004. 2208 - 2211.
- [10] WANG J, WANG J. Reduction Algorithms Based on Discernibility Matrix: The Ordered Attributes Method [J]. Journal of Computer Science Technology, 2001, 16(6): 498 - 504.
- [11] 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.
- [12] LI TR, QING KY, YANG N, et al. Study On Reduct and Core Computation in Incompatible Information Systems [A]. 4th International Conference, RSCTC 2004 [C]. Uppsala, Sweden, 2004. 471 - 476.
- [13] CHIN KS, LIANG J, DANG C. Rough Set Data Analysis Algorithms for Incomplete Information Systems [A]. 9th International Conference: Rough Sets, Fuzzy Sets, Data Mining And Granular Computing (RSFDGrC) [C]. Chongqing, China, 2003. 264 - 268.
- [14] 张文修, 吴伟志. 粗糙集理论与方法 [M]. 北京: 科学出版社, 2001.
- [15] WONG SKM, ZIARKO W. On optimal decision rules in decision tables [J]. Bulletin of Polish Academy of Sciences, 1985, 33(11 - 12): 693 - 696.

(上接第 863 页)

性, 选择一个属性的最小约减  $\{ \text{Make\_model}, \text{cyl}, \text{fuel}, \text{comp}, \text{weight} \}$ , 得到如下的约减表。

通过合并近似项, 得到表 4 的合并约减表。

对例外记录进行噪声滤波, 设噪声滤波阈值为 5%, 规则转换可写成如下:

规则 1:

if (make-model = USA)  $\wedge$  (cyl = 6)  $\wedge$  (fuel = EFI)  $\wedge$  (weight = medi) then (mileage = medium) 规则的权重为 44%

规则 2:

if (make-model = USA)  $\wedge$  (cyl = 4)  $\wedge$  (fuel = EFI)  $\wedge$  (weight = heavy) then (mileage = low) 规则的权重为 8%

规则 3:

if (make-model = USA)  $\wedge$  (cyl = 6)  $\wedge$  (fuel = EFI)  $\wedge$  (weight = heavy) then (mileage = low) 规则的权重为 8%

规则 4:

if (make-model = Japan)  $\wedge$  (cyl = 4)  $\wedge$  (fuel = EFI)  $\wedge$  (comp = high)  $\wedge$  (weight = Medi) then (mileage = high) 规则的权重为 16%

#### 参考文献:

- [1] HUA X, CERCONE N, HAN J. An attributed-oriented rough set ap-

proach for knowledge discovery in databases [A]. Rough sets, Fuzzy sets and knowledge discovery [C]. Springer-Verlag, 1993. 90 - 99.

- [2] 陈红梅, 王丽珍. 面向属性的量化归纳 [J]. 计算机研究与发展, 2001, 38(2): 150 - 155.
- [3] 刘明吉, 王秀峰, 李宝林. 基于 GA 的多层次概念归纳学习方法 [J]. 系统工程与电子技术, 2001, 23(4): 76 - 79.
- [4] SLEZAK D. Decomposition and synthese of decision tables with respect to generalized decision functions [A]. Fuzzy and rough set [C]. Springer-Verlag, 2000. 109 - 112.
- [5] DUNTSCH I, GEDIGA G. Statistical evaluation of rough set dependency analysis [J]. International journal of human-computer studies, 1997, 46(11): 589 - 595.
- [6] ZHONG N, DONG JZ, OHSUAGA S, et al. An incremental probabilistic rough set approach to rule discovery [A]. IEEE World Congress on Computational Intelligence (WCCI'98) [C]. IEEE Press, 1998. 933 - 938.
- [7] MROZAK A, SKABEK K. Rough rules in prolog [A]. Proceedings of the First International Conference on Rough Sets and Current Trends in Computing [C]. London, UK: Springer-Verlag, 1998. 458 - 466.
- [8] BROWNE C, DUNTSCH I, GEDIGA G. IRIS revisited: A comparison of discriminant and enhanced rough set data analysis [EB/OL]. <http://www.cosc.brocku.ca/~duentsch/archive/iris.pdf>, 1998.