基于云模型的决策表规则约简

代 劲 何中市
(重庆大学计算机学院 重庆 400030)

摘 要 通过对决策表的转换,将规则映射成为云向量,用云向量数字特征间的相似度来度量决策表规则间的等价关系。基于此,提出了基于云模型的决策表规则约简算法,不仅解决了粗集基于严格属性匹配的等价关系不能区分相似关系,也克服了基于模糊集等价关系依赖专家先验知识、对属性值随机性分布考虑不足的缺点。实验说明了该算法的高效性。

关键词 粗集,决策表,云模型,规则约简,等价关系
中图法分类号 TP18 **文献标识码** A

Rules Reduction for Decision Table Based on Cloud Model

DAI Jin HE Zhong-shi
(College of Computer Science, Chongqing University, Chongqing 400030, China)

Abstract Through decision table transformation, the rules were mapped to cloud vectors. The equivalence relations between the rules were reckoned using the similarity between the cloud vectors' digital characteristics. On this basis, a rules reduction algorithm for decision table based on cloud model was proposed. It not only resolves the problem that the equivalence relation based on rough set theory doesn't distinguish the similar relationship because it requires matching strictly each attribute, but also overcomes the shortcomings that the equivalence relation based on fuzzy set relies on the priori knowledge and considers inadequate to the random distribution of attributes. Experiments show the algorithm has high performance.

Keywords Rough set, Decision table, Cloud model, Rules reduction, Equivalence relation

粗集(Rough Set, RS)理论[1]由波兰逻辑学家 Z. Pawlak 教授于1982年提出。它能有效地分析和处理不精确、不一致、不完整等各种不完备信息,并能从中揭示潜在的规律,因此近年来在机器学习、数据挖掘等多个领域得到了广泛应用[2,3]。

决策表是粗集进行知识获取的重要工具。而受客观世界及认知程度的影响,要获得完全精确、不包含任何冗余信息的决策表非常困难。因此,对决策表进行规则约简是非常有必要的。对决策表进行规则约简基于规则间的相似关系。而由于决策表属性取值的多样性以及取值可能为连续实值,传统基于严格属性值匹配的粗集等价关系方法难以确定规则间相似关系。文献[4]将粗集中等价关系推广为模糊等价关系,文献[5]提出基于模糊集的 T-模糊相似关系,文献[6]在此基础上提出了决策表的规则约简方法,用隶属度来刻画规则亦此亦彼的程度。但精确的隶属度确定带有主观色彩,常常依靠专家的先验知识给定,并且对属性上的随机性分布考虑不足,约简后规则识别率不高。

云模型是李德毅院士提出的一种定性定量转换模型[7-10],该模型用语言值表示某个定性概念与其定量表示之间的不确定性,已经在智能控制、模糊评测等多个领域得到应用。由于其具有良好的数学性质,可以表示自然科学、社会科学中大量的不确定现象[10]。云模型不需要先验知识,它可以从大量的原始数据中分析其统计规律,实现从定量值向定性概念的转化。

本文第1节介绍有关 Rough 集理论的基本概念;第2节介绍有关云模型的基本概念;第3节提出一种基于云模型的决策表规则约简算法;第4节给出实验结果及其分析;最后总结全文。

1 Rough 集的基本概念

设 $S=\langle U,A,V,f\rangle$ 为一信息系统,$U=\{x_1,x_2,\cdots,x_n\}$ 是论域,A 是属性集合,V 是属性值集合,F 是 $U\times A\to V$ 的映射。若 $A=C\cup D$,$C\cap D=\varphi$,C 称为条件属性集,D 称为决策属性集,则该信息系统称为决策表。

信息系统中每一 x_i 及其对应的属性值称为一个(信息表的)规则。

定义 1(不分明关系[11]) 给定决策表 $S=\langle U,A,V,f\rangle$,对于每个属性子集 $B\subseteq A$,定义一个不分明关系 $IND(B)$,即

$$IND(B)=\{(x,y)\mid(x,y)\in U\times U,\forall b\in B(b(x)=b(y))\}$$

显然不分明关系是一种等价关系。

到稿日期:2009-07-24 返修日期:2009-09-30 本文受国家863计划项目(2007AA01Z423),重庆市自然科学基金项目(2007BB2134)资助。
代 劲(1978-),男,博士生,主要研究方向为智能信息处理、自然语言处理,E-mail:daijin@cqupt.edu.cn;何中市(1965-),男,教授,博士生导师,主要研究方向为模式识别、机器学习。

定义 2(上、下近似集[11]) 给定决策表 $S=\langle U,A,V,f\rangle$，对于每个子集 $X\subseteq U$ 和不分明关系 $IND(B)$，X 的上、下近似集分别可以由 B 的基本集定义如下：

$$B^{-}(X)=\bigcup_{Y_i\in U/IND(B)\wedge Y_i\cap X\neq\phi}Y_i$$

$$B_{-}(X)=\bigcup_{Y_i\in U/IND(B)\wedge Y_i\subseteq X}Y_i$$

定义 3(可省略，独立[11]) 令 R 为一等价关系族，且 $r\in R$，当 $IND(R)=IND(R-\{r\})$ 时，称 r 为 R 中可省略的，否则 r 为 R 中不可省略的。对于任一 $r\in R$，若 r 为 R 中不可省略的，则 R 为独立的。

定义 4(约简[11]) 设 R 为一等价关系族，且 $r\in R$。如果 $P=R-\{r\}$ 是独立的，则 P 是 R 中的一个约简。

约简 P 是能够与 R 表达同样知识的最小等价关系集合，为 R 中的重要部分。P 虽然去除了部分多余的知识，但仍然可以取得与原有的完整知识库一样的分类结果。

规则约简就是在保证信息系统中的规则是一致的前提下，约去等价关系中的冗余规则。

2 云模型理论基本概念

云模型是李德毅院士提出的一种定性定量转换模型[7-10]。该模型用语言值表示某个定性概念与其定量表示之间的不确定性，已经在智能控制、模糊评测等多个领域得到应用。

定义 5(云，云滴[7]) 设 U 是一个用数值表示的定量论域，C 是 U 上定性概念，若定量值 $x\in U$ 是定性概念 C 的一次随机实现，x 对 C 的确定度 $\mu(x)\in[0,1]$ 是有稳定倾向的随机数，$\mu:U\to[0,1]\ \forall x\in U\ x\to\mu(x)$，则 x 在论域 U 上的分布称为云，记为云 $C(X)$。每一个 x 称为一个云滴[7]。如果概念对应的论域是 n 维空间，那么可以拓广至 n 维云。

云用期望 Ex(expected value)、熵 En(entropy)、超熵 He(hyper entropy)这 3 个数字特征来整体表征一个概念。期望 Ex 是云滴在论域空间分布的期望，是最能够代表定性概念的点；熵 En 代表定性概念的可度量粒度，熵越大，概念越宏观，也是定性概念不确定性的度量，由概念的随机性和模糊性共同决定。一方面，En 是定性概念随机性的度量，反映了能够代表这个定性概念的云滴离散程度；另一方面，又是定性概念亦此亦彼性的度量，反映了在论域空间可被概念接受的云滴的取值范围；超熵 He 是熵的不确定性度量，即熵的熵，由熵的随机性和模糊性共同决定。用 3 个数字特征表示的定性概念的整体特征记作 $C(Ex,En,He)$，称为云的特征向量。

逆向云算法是云模型中一个关键算法，实现从定量值到定性概念的转换，将一组定量数据转换为以数字特征 $\{Ex,En,He\}$ 来表示的定性概念。算法如下。

算法 1 逆向云算法[7]

输入：N 个云滴 $\{x_1,x_2,\cdots,x_N\}$

输出：N 个云滴表示的定性概念的期望值 Ex、熵 En、超熵 He。

算法步骤：

(1) 根据 x_i 计算这组数据的样本均值 $\overline{X}=\frac{1}{N}\sum_{i=1}^{N}x_i$，一阶样本绝对中心矩 $\frac{1}{N}\sum_{i=1}^{N}|x_i-\overline{X}|$，样本方差 $S^2=\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\overline{X})^2$

(2) Ex 的估计值为 $\hat{Ex}=\overline{X}$；En 的估计值为 $\hat{En}=\sqrt{\frac{\pi}{2}}\times\frac{1}{N}\sum_{i=1}^{N}|x_i-\hat{Ex}|$；$He$ 的估计值为 $\hat{He}=\sqrt{S^2-\frac{1}{3}\hat{En}^2}$。

3 基于云模型的决策表规则约简算法

对决策表进行规则约简是基于规则间的相似关系的。而由于决策表属性取值多样性以及取值可能为连续实值，传统粗糙集等价关系方法难以确定规则间的相似关系。文献[4-6]引入模糊集隶属度概念来对粗集中等价关系进行刻画，并在此基础上提出了决策表的规则约简方法。但精确的隶属度确定带有主观色彩，常常依靠专家的先验知识给定，并且对属性上的随机性分布考虑不足，约简后规则识别率不高。

利用云模型在定性知识表示以及定性、定量知识转换时的桥梁作用，本文提出一种在知识层面比较决策表规则相似度的方法，不仅克服了传统粗集等价关系严格匹配对象属性的不足，也避免了模糊关系对随机性分布考虑的不足。

3.1 决策表转换算法

将云理论用来刻画规则之间的不确定性关系，必须保证规则在各属性的取值属于同一值域，即：必须保证同一规则在不同属性上的取值都有相同的物理含义(定义 5)。而现有的决策表各属性表示含义不一，取值也千差万别，需要进行转换。现给出决策表转换算法。

算法 2 决策表转换算法

输入：决策表 DT

输出：转换后决策表 DT'

算法步骤：(1)对 D 中每一 D_k 执行以下循环(D 为决策表决策属性集)：

①$S_{jk}=\sum_{i=1}^{m}U_{ij}$，$\overline{D}_{jk}=\frac{1}{m}S_{jk}$($m$ 为决策值为 D_k 的规则数，U_{ij} 为决策值为 D_k 的第 i 条第 j 个属性的规则取值) //S_{jk} 为相同决策规则集在不同属性上的和，$\overline{D}_{jk}$ 为其均值；

②$U_{ij}=|U_{ij}-\overline{D}_{jk}|$ //计算规则在不同属性的波动情况；

③$U_{ij}=U_{ij}/S_{jk}$ //归一化处理。

(2)$DT'=DT$，返回 DT'

经过算法 2 的转换，决策表每一条规则在属性上的取值变成规则基于该属性的取值波动情况，完整地反映了规则的随机分布情况。

3.2 决策表云相似度

决策表经过转换后，其每一规则构成了一云向量，可以用云向量的数字特征 $\{Ex,En,He\}$(可由算法 1 得到)来对规则进行整体刻画。这 3 个数字特征称为规则的特征向量。规则间的相似关系即可转换为特征向量之间的相似度量。

定义 6(云向量相似度) 给定由两个云 i,j 的数字特征组成的向量 $\vec{V}_i$ 和 $\vec{V}_j$，它们之间的余弦夹角称为云 i 和 j 之间的相似度：

$$\mathrm{sim}(i,j)=\cos(\vec{V}_i,\vec{V}_j)=\frac{\vec{V}_i\cdot\vec{V}_j}{||\vec{V}_i||\,||\vec{V}_j||}$$

式中，$\vec{V}_i=(Ex_i,En_i,He_i)$，$\vec{V}_j=(Ex_j,En_j,He_j)$

从上式可以看出，$\mathrm{sim}(i,i)=1$，即一个云与其自身的相似度为 1；$\mathrm{sim}(i,j)=\mathrm{sim}(j,i)$，即云 i 对 j 的相似度等于 j 对 i 的相似度。基于定义 6，给出定义决策表云相似度。

定义 7(决策表云相似度) 给定决策表，在经过算法 2 转换后，两个规则之间特征向量 $\{Ex,En,He\}$ 的余弦夹角，称为规则间的云相似度。

由于转换后决策表规则的特征向量包含了规则对属性的平均取值度、离散度和不确定度，通过计算向量间余弦夹角得

到的相似度中包含了规则的统计特征，因而得到的相似性是合理的。

3.3 基于云模型的决策表规则约简算法

基于云相似度对规则间相似关系的刻画，通过计算规则间的云相似度，可以得到规则间的相似关系度量。通过聚类将最相似的规则归为一簇，取出该簇的中心作为新规则来替换该簇，即完成规则约简。

在聚类方法的选取上，采用了效率比较高的 K-means 动态聚类方法[12]完成对重要特征的提取。然而 K-means 方法存在着需要人工指定聚类个数，容易陷入局部最优解的缺陷。在此分析基础上，提出了逐级 K-means 动态聚类算法。

算法基于如下分析：在聚类时，将特征最为相似的样本数据划分在同一类中，而它们在空间中的相互关系可以按照某些范数度量的大小关系来表征。通过逐步增加层数(即聚类数，含义显然)，层间类与类相互关系随着层数的增加不断变化，可以某种准则来刻画这种变化，一旦达到要求，即表示聚类完成。本文用样本标准差 S_i 来表征这种变化，随着聚类层数的增加，各类聚合程度越来越大，S_i 将会不断减少。在判断聚类是否完成时，采用 $K=n$ 时的平均样本差 $\overline{S}_n$ 作为收敛条件，即 $\min(S_i)<\overline{S}_n$，$(i=1,\cdots,K)$时，聚类完成。

算法 3 决策表规则逐级 K-means 均值聚类算法

输入：规则集云相似矩阵 $D=\{T_i,T_j,\text{sim},Clusterid\}$//sim 为样本 $T_i,T_j(i<j)$云相似度，$Clusterid$ 为聚类编号，初始为 0

输出：聚类后云相似矩阵 $D'=\{T,Clusterid\}$

算法步骤：

(1) 提取 D 中所有不重复的云相似度以升序构成类别向量 $C'=\{\text{sim},Clusterid\}$，其中 $Clusterid$ 为聚类类别编号；

(2) 计算阈值 $e=\overline{S}_n$//聚类结束控制条件；

(3) 初始类别 $K=1$，$v=1$ //v 为循环控制变量；

(4) Do

①构建中心类别表 TC：将 C' 平均分成 $K+1$ 份，取前 K 份的右端点加入 TC，作为 C' 在 K 情况下的初始类别；同时将 C' 各元素 $Clusterid$ 置为 0；

②设定临时循环控制变量 $e_1=0$；

③当 $e_1\neq v$ 时，执行以下循环：//当聚类稳定后，各类的标准差将收敛为一稳定值

a)$e_1=v$；

b)计算 C' 中每个值与 TC 中各类别的距离，将其归并到距离最小类别中；

c)加权平均修正 TC 中各类别中心距离；

d)计算 T 中各类别标准差 S_i，$v=\min(S_i)$

④$K=K+1$

While $(v>e)$ //当 $v\leqslant e$ 时，各类的聚合程度已经比较好，聚类结束

(5)根据 C' 中相似度，更新 D 中 $Clusterid$ //由聚类算法可知 $Clusterid$ 越大，则云相似度越高；

(6) 对 D 按照 T_i 升序、$Clusterid$ 降序方式，进行如下处理：

①依次取 $t_k\in T_i$，$c_{i\max}$ 为 $T_i=t_k$ 时最大类别号，则 t_k 在 D 中最相似样本集为

$T=\{t_k\}\cup\{t_m\mid t_m\in T_j$ 且 $T_i=t_k\wedge Clusterid=c_{i\max}\}$//由聚类算法可知 $Clusterid$ 越大，则云相似度越高；

②$D'=D'\cup\{(T,c_{i\max})\}$ //只保留每簇最相似规则集；

(7) 返回 D'。

经过算法 3 的聚类处理，可以得到决策表在进行云相似关系处理后的规则等价关系，在此基础上提出了基于云模型的决策表规则约简算法。

算法 4 基于云模型的决策表规则约简算法

输入：决策表 DT

输出：规则约简后决策表 DT'

算法步骤：

(1)$DT'=\varphi$；

(2)利用算法 2，将决策表 DT 转换成可进行云相似度量的决策表 DT'；

(3)利用逆向云算法(算法 1)，得到 DT' 中每一规则的特征向量 $\vec{V}_i=(Ex_i,En_i,He_i)$；

(4)计算规则间的云相似度，得到规则集云相似矩阵 $D=\{T_i,T_j,\text{sim},Clusterid\}$；

(5)对 DT 中每一决策值 D_k 执行以下循环：

①利用算法 3，得到规则在决策 D_k 下的等价关系划分 $D_k'=\{T,Clusterid\}$；

②决策表云相似矩阵 $D'=\cup D_k'$；。

(6)根据 D'，DT 被划分成了若干个等价类，取每个类中心规则加入 DT'；

(7)返回 DT'。

4 实验结果及分析

为了考察本文方法的性能，我们采用 UCI 数据集对算法 4 进行了测试。选取 UCI 数据库中的 5 个数据集，采用了 3 交叉法进行测试。各数据集的记录数、条件属性数和决策属性数如表 1 所列。

表 1 选取的 UCI 数据集

数据集	记录数	条件属性数	决策属性数
Iris	150	4	1
Wine	178	13	1
Glass	214	10	1
Abalone	4177	8	1
Artificial Characters	6000	7	1

对于以上数据集，采用未进行规则约简(算法 a)、基于模糊相似关系的规则约简[6](算法 b)、基于算法 4 的规则约简 3 种情况进行了规则获取，并进行样本识别测试。实验过程如下：

先在 3 种情况下对决策表进行处理，然后采用 Nguyen 贪心算法[13]进行离散化，再用基于可辨识矩阵属性约简算法[11]对离散数据集进行属性约简。在值约简过程中，采用基于决策矩阵算法[11]进行值约简处理，最后将得到的规则集对测试数据集进行测试。测试结果如表 2 所列。其中，T 为各算法运行时间(ms)，$Rule$ 为决策表规则数目，Rec(%)为正确识别率。本实验的硬件测试环境是，CPU：Intel P4 2.4GHz，内存：1GB，操作系统：Windows XP，开发工具为 VC++6.0。

表 2 算法对比测试结果

Dataset	Algorithm 4			Algorithm a			Algorithm b (ε=0.1)		
	T	Rule	Rec	T	Rule	Rec	T	Rule	Rec
Iris	158	21	**97.3**	380	150	98.0	120	13	82.0
Wine	181	25	**92.8**	412	178	93.8	135	18	81.5
Glass	202	31	**83.7**	450	214	84.1	172	24	72.0
Abalone	893	388	**81.3**	3251	4177	79.2	2955	315	69.7
Artificial Characters	912	450	**85.4**	3566	6000	81.9	3105	401	75.1

(下转第 277 页)

地发现中国大学生群体存在的问题，及时发出危机预警，以避免受错误方向的诱导和盲目跟从心理导致的弊端。

随着计算机技术的飞速发展以及网络群体带给人们观念的全新影响，对网络群体心理趋势的研究，将有助于更为准确地了解群体的心理状况。通过云计算技术、信息检索技术、中文文字云图技术、结合结构方程模型等现代统计技术，使得对于网络群体心理趋势的分析更为客观、有效。相信，通过信息检索技术注入心理词汇，可为心理学的研究注入新的活力，同时可使群体的心理研究更为科学、客观、准确。

参 考 文 献

[1] 张刚，刘悦，郭嘉丰，等. 一种层次化的检索结果聚类方法[J]. 计算机研究与发展，2008，45(3)：542-547

[2] 姜远，周志华. 基于词频分类器集成的文本分类方法[J]. 计算机研究与发展，2006，43(10)：1681-1687

[3] Westman H. *Finding your way in computer graphics*[J]. ACM SIGGRAPH Computer graphics，2006，40(2)

[4] Hsieh J W，Kuo T W，Chang L P. Efficient Identification of Hot Data for Flash Memory Storage Systems[J]. ACM Transactions on Storage，2006，2(1)：22-40

[5] *Chiang Y-J，Tamassia R. Dynamic algorithms in computational* geometry[R]. CS-91-24. Brown University，1992

[6] An Jun-xiu，Jin Yu-chang. Research on Text Classification Algorithm of Largest Dispersion Based on Term Frequency[C]// International Forum on Computer Science-Technology and Applications(IFCSTA 2009). 2009

[7] Haase C，Guy R. JAVA 动画、图形和极富客户端效果开发/Sun 公司核心技术丛书[M]. 北京：机械工业出版社，2008

[8] Law D. Taligent MVP in interactive statistical graphics [J]. Computational Statistics，2008，23：487-495

[9] Venugopal S. 数据结构从应用到实现(Java 版)[M]. 北京：机械工业出版社，2008

[10] 曹晓华，曹立人. 不规则几何图形识别取样特征的眼动研究[J]. 心理学报，2005，37(6)：748-758

[11] Kathryn E，Van Dam S A. Java 面向对象程序设计图形化方法[M]. 北京：机械工业出版社，2006，7：285-297

[12] 马亮，陈群秀，蔡莲红. 一种改进的自适应文本信息过滤模型[J]. 计算机研究与发展，2005，42(1)：79-84

[13] 王鹏，陈高云，安俊秀，等. 移动搜索引擎原理与实践[M]. 北京：机械工业出社，2009

[14] 王学松. lucene＋nutch 搜索引擎[M]. 北京：人民邮电出版社，2008

[15] 吴明隆. 结构方程模型－APOS 操作与应用[M]. 重庆：重庆大学出版社，2009：15-33

[16] 邱皓政，林碧芳. 结构方程模型的原理与应用[M]. 北京：中国轻工业出版社，2009：10-24

[17] 叶育鑫，欧阳丹彤. 语义 Web 搜索技术研究进展[J]. 计算机科学，2010，37(1)：1-5

[18] 申凡，钟云. 网络粉丝群体心理研究[J]. 南京邮电大学学报：社会科学版，2009，11(4)：27-31

[19] 吴太胜，陈业秀. 网络语言及网民群体的社会文化心理探析[J]. 广西社会科学，2007，9：171-174

[20] 尚得娟，张敏. 信息过滤系统中的混合式过滤算法[J]. 重庆工学院学报：自然科学版，2008，22(1)：118-121

(上接第 267 页)

从表 2 可以看出，当样本数较少时，算法 4 得到的结果与未进行规则约简得到的识别率大致相当，但耗费时间不到其一半；当样本数较多时(Abalone 与 Artificial Characters 数据集)，算法 4 的识别率超过其他两种方法，充分说明了算法的有效性。在时间消耗上，虽然稍大于基于模糊相似关系的规则约简算法，但识别率较后者提高较大。原因在于算法 4 较好地保持了原始决策表的特性，同时对冗余规矩进行了归并处理，在数据集较大时性能提升更为明显。

结束语 决策表是粗集进行知识获取的重要工具。而受客观世界及认知程度的影响，要获得完全精确、不包含任何冗余信息的决策表非常困难，因此在使用决策表进行知识获取前对规则进行约简是非常必要的。这不仅能去掉冗余信息，增强决策表对知识的描述能力，也能有效提高规则提取效率，促进粗集的进一步推广。

参 考 文 献

[1] Pawlak Z. Rough Set [J]. International Journal of Computer and Information Sciences，1982，11：341-356

[2] Pawlak Z，Grzymala-Busse J，Slowinski R，et al. Rough sets [J]. Communications of the ACM，1995，38(11)：89-95

[3] Pawlak Z. Vagueness-a rough set view [A]// Mycielski J，Rozenberg G，Salomaa A，eds. Structures in Logic and Computer Science：A selection of Essays in Honor of A [C]. Ehrenfeucht，Berlin：Springer-Verlag，1997：106-117

[4] Fernandez S J M，Murakami S. Rough set analysis of a general type of fuzzy data using transitive aggregations of fuzzy similarity relations [J]. Fuzzy Sets and Systems，2006，144：1-26

[5] Zhao S X，Wang X Z. Core and reduction from mutual relation view and their fuzzy generalization [A]// IEEE International Conference on Systems，Man & Cybernetics [C]. 2003：2611-2616

[6] 马玉良，杨家强，颜文俊. 基于模糊相似度的实值属性信息系统规则约简[J]. 浙江大学学报：工学版，2006，40(9)：1550-1553

[7] Li D Y. Artificial Intelligence with Uncertainty [M]. Beijing：National Defense Industry Press，2005：171-177

[8] Li D Y，Liu C Y. Study on the universality of the normal cloud model [J]. Engineering Science，2004，6(8)：28-34

[9] Li D Y，Liu C Y，DU Y，Han X. Artificial intelligence with uncertainty [J]. Journal of Software，2004，15(11)：1583-1594

[10] Li D Y. Uncertainty in knowledge representation [J]. Engineering Science，2000，2(10)：73-79

[11] 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

[12] Yuan S M，Cheng X Q. Clustering Method for Mining Quantitative Association Rules [J]. Chinese Journal of Computers，2002，23(8)：866-871

[13] Nguyen S H，Skowron A. Quantization of real value attributes-rough set and Boolean reasoning approach [A]// Proc. of the 2nd Joint Conf. on Information Science [C]. 1995：34-37