

基于Bayes概率边界域的粗集分类方法及其在高频数据中的应用

2010-12-28 14:39:16

来升强/谢邦昌/朱建平

【内容提要】

作为一种近似处理的工具,粗集主要用于不确定情况下的决策分析,并且不需要任何事先的数据假定。但当前的主流粗集分类方法仍然需要先经过离散化的步骤,这就损失了数值型变量提供的高质量信息。本文对隶属函数重新加以概率定义,并提出了一种基于Bayes概率边界域的粗集分类技术,比较好地解决了当前粗集方法所面临的数值型属性分类的不适应、分类规则不完备等一系列问题。

【关键词】可变精度粗糙集/Bayes边界域/高频数据

一、引言

粗集理论的基本观点是利用上近似、下近似以及两者的差集非常明确地刻画了任何个体的类属[1]。在分类能力不变的前提下,通过知识约简导出分类规则,并用于数据压缩,规则提取以及分类等任务。粗集中的主要研究方向集中在理论扩展、数学性质分析以及算法设计等方面[2-4]。但粗集也面临两个主要的问题:其一,粗集的基础是等价关系,但实际中这一条件比较强,并且考虑到噪声和观测误差,容易出现泛化能力不足的现象[5]。另外,粗集不能直接处理定量数据,必须先对连续数据进行离散化,不可避免地损失了一些分类信息。Ziarko(1993)提出的可变精度粗集(Variable Precision Rough Set)模型比较好地解决了第一个方面的问题[6]。可变精度粗集模型把粗集上下近似域的绝对条件变为相对条件,也就是根据其所考察的属性集的重叠区域的比例来判断近似程度。在属性集不变的条件下, β 越大,其边界越清晰,当 $\beta=1$ 时,就还原为普通的粗集模型。因为采用了概率公式来表达近似程度,所以也称为 β 近似。随后,Ziarko又进一步提出基于样本的 β 近似等价关系,从大数定律的角度进一步探讨了利用 β 近似所得到的规则的完备性、可扩展性和支持度[8]。

可变精度粗集也存在一些不妥之处。由于 β 近似建立在部分属性的等价关系上,组成这些等价关系的属性集所构成的核仅仅是全部属性集的一个子集,根据粗集有关定义,增加一个或多个属性后所形成的增广核,其分类精度至少不应降低。但有研究表明会出现某些增广核的分类准确度反而下降的情况[9],这就直接违背了粗集中约简与核的定义。其次,由于可变精度粗集方法是将落入同一等价类区域中的个体看做是来自同一总体中的样本,直接计算其频率作为近似程度指标。这从统计上来看,并不总是适当的,至少对于有限个体的重复观测情形来说,这个假定完全忽略了个体内在的异质特征。至于改进后的样本 β 近似等价类划分方法[8],则把论域本身看做样本,认为从样本论域中得到的等价关系只是总体论域的一个子集,这就允许在样本论域扩大时出现新的等价关系,比较好地解释了等价关系的不完备性。但这种论域随机化处理并没有解决在实际问题中可变精度粗集泛化能力仍然不足的现象。再次,研究者很少关注 β 值的选择,一般假定是经验判断值或事先给定的,在决策过程作为外生变量处理。而在实际决策过程中, β 初始值选择十分关键,并且伴随着决策过程的持续, β 也应做调整。最后,对于数值型变量的分类,可变精度粗集方法仍然显得力不从心。虽然可变精度粗集已经对隶属函数进行了一定程度的“柔化”,但其隶属函数仍然是由定性数据的隶属关系导出,放弃了数值型数据所提供的高质量分类信息。

本研究针对这些问题,从统计角度提出了一种基于Bayes概率边界域的粗集分类方法。此方法能够直接处理数值型数据,可以充分利用数值型变量的高质量信息。其主要步骤为:首先,通过历史数据获取全属性域的经验分布;其次,利用Bayes后验分布计算各个等价类的隶属概率,并用概率阈值筛选分类规则;最后,通过分类规则的预测准确率,得到最优的等价类划分策略。该方法避免了可变精度粗集中 β 近似的不足和缺陷,能够得出Bayes最优的分类规则,并给出了全域一致的隶属概率,避免了增广约简核的精度下降的不一致情形。

二、基本粗集模型与扩展

其中, $S_{1(k)1} = \sum_{ij \in D_k^{(1)}} x_{ij}$, $S_{1(k)2} = \sum_{ij \in D_k^{(2)}} x_{ij}$, $S_{2(k)1} = \sum_{ij \in D_k^{(1)}} z_{ij}^{(1)}$, $S_{2(k)2} = \sum_{ij \in D_k^{(2)}} z_{ij}^{(2)}$; $D_k^{(1)}$ 为在时间 $y_{(k)}$ 因为第一种风险而发生事件的个体集合, $D_k^{(2)}$ 为在时间 $y_{(k)}$ 因为第二种风险而发生事件的个体集合, D_k 为 $y_{(k)}$ 时刻事件发生以及删失的个体集合; $z_{ij}^{(1)} = (z_{ij1}^{(1)}, \dots, z_{ijq}^{(1)})'$ 为 $q \times 1$ 的指示函数向量, 仅第 i 个分量为 1, 其余分量为 0, $z_{ij}^{(2)} = (z_{ij1}^{(2)}, \dots, z_{ijq}^{(2)})'$ 同样; $v_1 = (v_{11} \ v_{21} \ \dots \ v_{q1})'$, $v_2 = (v_{12} \ v_{22} \ \dots \ v_{q2})'$ 。如果不考虑数据打结和随机效应, 式(15)即为竞争风险下对数部分似然函数^[20]。

令 $\tau = (\beta_1, \beta_2, v_{11}, v_{21}, i=1, \dots, q)'$ 为待估计的参数列向量, 给定 ρ 的初值, τ 的估计的迭代公式为:

$$\hat{\tau}^{(i+1)} = (\hat{\tau}^{(i+1)}) = (\hat{\tau}^{(i)}) + J^{-1} \left(\frac{\partial h^*}{\partial \tau} \right)_{(\tau) = (\hat{\tau}^{(i)})} \quad (16)$$

其中, $J = \frac{\partial^2 h^*}{\partial \tau \partial \tau'}$, 而 J^{-1} 是 $\hat{\tau}$ 的渐近方差-协方差矩阵的估计。对随机效应分布的参数, 可利用 MAPHL (Maximize Adjusted Profile H-Likelihood) 求解, 也就是边际似然的一阶拉普拉斯变换; 由于 $h_{\lambda} = h^1, \tau$ 是关于参数的非线性函数, 需迭代求解:

$$h_{\lambda} = h^* + \frac{1}{2} \log(\det(2\pi J^{-1})), \quad (17)$$

其中 $\frac{1}{2} \log(\det(2\pi J^{-1}))$ 为惩罚项, 迭代过程仍可以采用 Newton-Rapson 方法。

4 模拟研究

下面进行模拟研究, 以验证上述竞争风险下随机效应模型估计的有关结论。为了模拟方便, 有如下假设:

①随机效应 v_{11}, v_{21} 的生成过程为:

$$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{pmatrix} \right)$$

随机效应的数据 v_{11}, v_{21} 由该二元正态分布生成, 标准差均为 1, 相关系数为 -0.1;

②基本危险率 $\lambda_1(T_{ij})$ 和 $\lambda_2(T_{ij})$ 均为 1;

③模型中只包含一个协变量 x , 协变量 x 一半取值为 1, 一半取值为 0;

④个体发生不同危险的时间在给定随机效应的条件下相互独立;

根据上述设定, 协变量通过随机效应模型(1)、(2)影响危险函数, 给定随机效应的情况下, 由参数是 $\exp(X_{ij} \beta_1 + v_{11})$ 和 $\exp(X_{ij} \beta_2 + v_{21})$ 的指数分布分别生成两个事件发生时间。

⑤参数 β_1, β_2 的真值分别是 -1、-2;

之所以将协变量系数设为负数, 是为了方便比较。若将其设为正数, 生成的事件发生时间太小, 很不方便比较。

⑥删失时间来自参数为 $\theta_{ij} = \left(\frac{r}{1-r} \right) \left(\frac{h_{1ij} + h_{2ij}}{2} \right)$ 的指数分布, 其中 r 为删失率, 删失率设为 0.1。

参数		均值		标准差		平均的标准误		包含真值比例	
		h_lik	Cox	h_lik	Cox	h_lik	Cox	h_lik	Cox
b1	模拟数据	-1.0642	-0.7506	0.3594	0.3032	0.3310	0.3036	0.955	0.860
	区间近似	-0.9843	-0.7146	0.3429	0.2973	0.3284	0.3041	0.945	0.825
b2	模拟数据	-2.0325	-1.5250	0.4716	0.3889	0.4469	0.3947	0.960	0.730
	区间近似	-1.9382	-1.4919	0.4536	0.3880	0.4430	0.3951	0.955	0.695
p	模拟数据	0.1971		0.3510					
	区间近似	0.1122		0.3964					
sd1	模拟数据	1.1077		0.5169					
	区间近似	1.1570		0.5157					
sd2	模拟数据	1.0654		0.5612					
	区间近似	1.1400		0.6187					

基于粗集的分类问题, 就是通过个体在其属性集上观测到的取值或状态, 来判断待分类集落入分类知识所形成的区域的情况。特别地, 当 X 中仅能观测到部分属性集, 而不能观测到判别属性时, 粗集的边界判断问题就变成了决策问题①。

X_k

特别地, 当 $\beta=1$ 时, 可变精度粗集就退化为普通的粗集。由于边界柔化, 较之普通的粗集, 可变精度粗集的容错能力和适应性能大大增强, 在此基础上, 派生出一系列基于 β 近似的数据约简[9]以及算法研究[10]等。

三、Bayes 概率边界域划分准则

可变精度粗集的 β 近似考虑的频率主要是指类似于截面数据的多个体单次观测值集合。由于观测误差的存在, β 近似中的频率只能看作是真实概率的一次观测值的实现, 具体的频率值也只是随机变量的一个观测值, 把这样的一次观测频率值作为概率显然是不合适的。虽然可以看作一个初始的先验概率, 并且这个先验信息对于首次决策无疑是有效率的, 但伴随着数据积累和决策经验增长, 不断增加的观测值使得后续决策改进成为可能, 基于概率的粗集方法也应当反映并包含这种由于经验积累所带来的决策改进。因此, 本文认为有必要从统计角度对粗集的边界域近似指标进行重新构建, 使其成为名副其实的近似粗集。

(一) 包含观测误差的隶属度函数

由于观测中无法避免的误差,导致了个体属性值的测量偏差,进而影响分类的准确性②。这种误差首先是出现在单一属性的示性函数上,由示性函数汇总,并表现为隶属度函数的不确定性。

$$F(j, t) = \text{pr}(C=j, Y \leq t), p_j = \text{pr}(C=j) = F(j, \infty), j=1, 2, \dots, L$$

是C的边际分布,满足 $\sum p_j = 1$ 。相应的部分危险率(sub-hazards)函数为

$$h_i(t) = \lim_{\delta \rightarrow 0} \text{pr}(C = j, Y \leq t + \delta | Y > t) = \frac{f(j, t)}{S(t)},$$

$$S(t) = E_i(S(t|v)).$$

6

根据前述有关论述和推理,得到单一属性的示性函数之后,再根据前节推导就可以得到 $x \in X$ 在所有属性变量上的Bayes概率分类规则。

但对于数值型变量而言,等价类划分策略可以千差万别,如何从中选择较优的一组分类策略就成为一个急需解决的问题。在主流的粗集方法中,数值型变量的等价类划分也被称为数值离散化,基本是被作为一个独立的过程进行。但这种处理方式显然疏忽了等价类的划分策略对于最终分类规则的影响。有鉴于此,结合bsyes概率边界域,本文提出了以下一种比较通用的数值型多属性等价类划分和选优步骤。

Y_i

4. 按照阈值 β 筛选出最优的分类策略。

四、Bayes边界粗集分类应用

β_1

从图1可以看出,指标误差

我们假定两个竞争风险情形。设观测数据 $T_{ij}(i=1, 2, \dots, q, j=1, 2, \dots, n_i)$ 表示第*i*个个体的第*j*个重复观测(或者表示第*i*组的第*j*个个体)。用 $\Delta_{ij} = (\delta_{ij1}, \delta_{ij2})$ 表示第*ij*个观测的时间属性,如果 $\delta_{ij1}=1, \delta_{ij2}=0$,则表示第*ij*个个体在 T_{ij} 时刻因第一种风险发生了感兴趣的事件,如失业状态结束或者因提前还贷导致抵押贷款终止;如果 $\delta_{ij1}=0, \delta_{ij2}=1$,则表示第*ij*个个体在 T_{ij} 时刻因第二种风险发生了感兴趣的事件,如失业者退出劳动力市场或者因违约导致抵押贷款终止;如果 $\delta_{ij1}=0, \delta_{ij2}=0$,则表示第*ij*个个体在 T_{ij} 时刻感兴趣的事件还未发生,这称为右删失(right censored)。删失是持续期数据的特点之一,右删失是指研究者确切知道个体进入研究的时间,但是事件发生时间却观测不到的情形。

假设在给定协变量 X 以及不可观测异质性因子 $V=(V_1, V_2)$ 的条件下,两个事件发生时间 T_1, T_2 相互独立;又设随机效应在同一观测组的取值相同,以此来体现个体的组内相关性;第*ij*个个体在不同风险下的危险函数设为半参数Cox比例危险模型:

$$h_1(T_{ij}|X_{ij}, v_{i1}) = \lambda_1(T_{ij}) \exp(X_{ij} \beta_1 + v_{i1}), \quad (1)$$

$$h_2(T_{ij}|X_{ij}, v_{i2}) = \lambda_2(T_{ij}) \exp(X_{ij} \beta_2 + v_{i2}). \quad (2)$$

由于不同风险是否独立难以识别,目前文献中往往对不同风险下的生存时间单独建模,这忽略了不同风险的联系。我们设定两个随机效应变量的联合分布描述危险率之间的联系,因而更符合实际。

Lee和Nelder^[11]在共轭分布的混合广义线性模型中定义了等级似然函数。对不可观测的随机效应 v 和观测数据 y ,关于 (θ, v) 的等级似然定义为:

$$h = \log(L(\theta, v; y, v)) = \log(L(\theta; y)L(\theta, v; v|y)), \quad (3)$$

其中 $L(\theta, v; y, v) = f_{\theta}(y, v)$ 给出了两个数据生成过程的联系:一是随机效应 v ,来自于概率密度函数 $f_{\theta}(v)$;当 v 生成之后,由概率密度函数 $f_{\theta}(y|v)$ 产生随机数据 y 。设 $\hat{\theta}$ 是从边际似然 $L(\theta; y)$ 得到的MLE, $I_m(\hat{\theta})$ 是相应的Fisher信息阵,则

$$I_m^{-1}(\hat{\theta}, \hat{v}) = \begin{pmatrix} I^{11} & I^{12} \\ I^{21} & I^{22} \end{pmatrix} \quad (4)$$

是等级似然 $L(\theta, v; y, v)$ 中关于 $(\hat{\theta}, \hat{v})$ 的信息矩阵的逆,其中 I^{11} 对应 θ 的部分。此时,有如下结论^[12]:

1)关于 θ 最大化边际似然 $L(\theta; y)$ 得到 $\hat{\theta}$,和关于 (θ, v) 一起最大化等级似然 $L(\theta, v; y, v)$ 得到的 $\hat{\theta}$ 是等价的。

2)两种似然估计的信息矩阵中,关于 $\hat{\theta}$ 的部分相同,也就是说: $I_m^{-1} = I^{11}$ 。这意味着从等级似然估计中可以直接得到关于固定参数 θ 的推断。

3) I^{22} 正是 $\text{var}(\hat{v}-v)$ 的估计,如果 $\hat{v} = E(v|y)|_{\theta=\hat{\theta}}$,则有 $\text{var}(\hat{v}-v) \geq E(\text{var}(v|y))$,恰好解释了由估计固定效应参数 θ 带来的不确定性的增加。

在3个等级上的分布形式非常类似,仅均值略有漂移。这表明如果仅对指标的取值区间进行划分,很难准确地判断产品等级准确率(实际上,其余三个单项指标误差分布也有类似的特征)。按等级分组的描述统计进一步说明了这一现象(见表1):各指标在产品等级分组上的标准差相差不大,偏度随着等级上升逐渐左偏,峰度随之降低。如果直接采用粗集分类,由于属性集的等价类大量重叠,这将会产生相当数量的错误分类规则。

由于进行了标准化,观测到的分布可以认为主要是由检测指标误差造成。因此,产品等级划分也将主要取决于每个指标的测量误差的分布。从分布中可以看到同一指标在不同分类等级上都有不同程度的重叠,这意味着,任一单指标的临界值分割方法将会得到非常大比例的次品,同时优良

产品的比例将会非常小。所以要比较准确地判断产品等级，必须要考虑各单项指标误差在不同产品等级下

$$\beta_2$$

$$Y_{i-1}$$

$$\theta_{obs}$$

$$\sigma_{obs}$$

由于核密度函数实际上是未知的，所以最优带宽（式(18)）只能是在某种程度的近似获取。本文选用的是Silverman(1986)提出的拇指规则带宽[12]，

$$z = \text{Min} \left[\frac{1}{i-1} \left(\frac{|Y_{2i} - Y_{2i}|}{\sigma_2} + \frac{|Y_{3i} - Y_{3i}|}{\sigma_3} + \dots + \frac{|Y_{ii} - Y_{ii}|}{\sigma_i} \right) \right] \quad (1)$$

在缺乏关于样本分布先验知识的条件下，为了能够比较精确地拟合出分布特征，笔者将节点数设定为1001个，所有分布拟合的近似均方积分误差(AMISE)均小于0.001④。类似地，在经验分布拟合中，间隔点也被分为1000个。

关于等价类的划分策略，笔者选用了从-0.6（约为10%的分位数）开始，到2.0（约为90%的分位数）之间进行等间隔分类，间隔数依次从5增至20。在规则提取环节，保留了每个等价类的最大概率规则，这就使所有等价类都有明确的类别所属，保证了分类规则的完备性。但考虑到不同分类方法所产生规则的不一致性，统一保留了前100条规则作为优选的规则集。逐个计算其在不同划分策略下的样本内准确率之后，比较其在样本规模和拟合分布类型下的平均准确率。

为了比较分类效率和稳健性，笔者还将原始数据按照批次(Lot)从5批数据逐渐扩展样本容量至所有批次。此外，还将bayes概率边界域划分方法与传统的按照覆盖比例进行划分的粗集方法放在同一框架下进行对比分析。

表2 Bayes概率边界域分类与简单粗集分类准确率的扩展样本对比

nLot	Empirical CDF						Kernel Density Est.					
	5	6	7	8	9	10	5	6	7	8	9	10
simAVG	0.37 (-0.218)	0.367 (-0.223)	0.355 (-0.226)	0.356 (-0.224)	0.358 (-0.224)	0.355 (-0.225)	0.32 (-0.177)	0.317 (-0.178)	0.313 (-0.179)	0.314 (-0.178)	0.314 (-0.179)	0.311 (-0.179)
bayesAVG	0.579 (-0.19)	0.575 (-0.191)	0.559 (-0.196)	0.566 (-0.201)	0.572 (-0.199)	0.566 (-0.199)	0.525 (-0.160)	0.519 (-0.155)	0.514 (-0.159)	0.525 (-0.161)	0.531 (-0.162)	0.524 (-0.163)

注：括号内为标准差。

表2列示了两种分类方法所得到的最优规则集合在不同扩展样本下的平均分类准确率。由于数据分布不具备明显的规律性，这使得基于平滑的核函数估计的整体准确率略低于采用经验分布的结果。从样本规模上看，伴随着样本批次的增加，这两类方法所得分类规则集的平均准确率和对应的标准差几乎不变。这表明两种方法对于样本规模具有很强的稳定性。但在各种情况下，采用bayes概率边界域分类方法不仅其平均准确率显著高于简单粗集分类方法，准确率提高幅度达60%~70%，而且对应的标准差仅为后者的80%~90%。

简单粗集边界判定方法的平均预测准确率（采用经验分布）

Variable	Mean	Standard Error	学生 t 检验		Shapiro-Wilk 检验	
			t value	P value	W value	P value
X1	1853.42	151.70	12.22	<0.000 1	0.903 6	0.108 0
X2	348.91	60.80	5.74	<0.000 1	0.856 4	0.021 4
X3	4840.66	244.75	19.78	<0.000 1	0.842 9	0.013 8

简单粗集边界判定方法的平均预测准确率（采用核密度分布）

$$Y_{ii} = Y_{ii} \pm z \sigma_i \quad (2)$$

图3 简单粗集分类法所得规则的平均分类准确率

图3、图4分别描述了两种粗集分类方法在样本规模和不同划分策略下的平均准确率变化情况。位于图上方的曲线为平均分类准确率，下方为对应的标准差。由于样本扩展一共得到6个不同规模的样本，而在每个样本内验证了16次不同的等价类划分策略，即每个样本本期包含了16个横坐标刻度，因此横坐标刻度为1至96。从图中可以看出，在同一样本期中，伴随着划分策略的细化，准确率的上升呈现了一定的“凸性”，这表明分类准确率与等价类个数的关系存在拐点，这个拐点正是最优的等价类划分策略的所在。数据表明这个最优划分策略是对变量等间隔划分个数为10或11。

bayes边界判定方法的平均预测准确率（采用经验分布函数）

$$\sigma_i$$

bayes边界判定方法的平均预测准确率（采用核密度函数）

$$|Y_{ii} - Y_{ii}|$$

图4 bayes边界域划分法所得规则的平均分类准确率

图4展示了基于bayes概率边界域的粗集分类方法所得到的前100个规则的平均预测准确率变化。与图3简单比较就可以发现bayes方法的明显优势：在每个样本期内，在每个划分策略下的平均分类准确率都要高于简单粗集分类方法，同时分类准确率的标准差及其波动性也都明显低于后者。

五、结论

粗集分类方法的观点就是对于定性数据，特别是不完全的定性数据进行决策分析或者分类。但并不意味着粗集分类就一定要牺牲数值型数据所提供的高质量信息。实际上，在当前计算成本相对较低的条件下，放弃决策信息的精确度已经变得没有必要，而且由此带来的信息损失将会直接误导决策。

本文所提出的基于Bayes概率边界域的粗集分类方法比较好地解决了这一问题。首先从数值型变量中提取充分的Bayes分类信息，然后再根据提取出分类规则的准确率来选择最优等价类。通过隶属概率的方式，充分利用了数值型变量的高质量信息，解决了数值型属性的粗集划分问题。同时

还避免了可变精度粗集中 β 近似的不足和缺陷,以及增广约简核的精度下降的不一致情形。最后本文还对生产实践中的高频检测数据进行了实证分析,所得到的结果也支持了本方法的实践优势。

注释:

①在基于粗集的决策问题中,属性集被分成条件属性集和决策属性集。在条件属性集中寻找恰当的属性集的取值组合并对观测进行划分而得到一个基于值集相等的等价类族,通过考察等价类的属性集的取值频率,提取出“并发概率”较高的值集,并描述成为一条“IF-THEN”规则。对于分类问题,同样也要建立类似的规则,但与决策问题不同,其决策属性集是由分类知识派生出来的类别信息所构成。例如,通过考察一批观测数据后,发现可以将其分为五类。那么,这五个类别(有时还会加入一个无法判断的类别),就构成了后续观测数据的决策属性集。这五个类别并不是由观测数据本身提供,而是根据分析结果推断得出。而对于后续观测数据而言,其决策属性集则是给定的,需要进一步根据其可观测的条件属性集加以判断。所以本文将决策问题作为分类问题的特例看待,不加以区分。

②特别是工业生产中的检测系统,比较容易受到生产环境的影响,检测仪器需要经常性的调整和校准以减小误差。

③在各个等级下的参数分布拟合优度均不显著,故省略。

④通过反复拟合数据,我们发现当节点数 n 大于800时,不同核函数和带宽设定之间的差异微乎其微。

【参考文献】

- [1]Z. Pawlak. Rough Sets[J]. Int'l Journal of Computer and Information Sciences, 1982(11):341-356.
- [2]来升强,朱建平.数据挖掘中高维定性数据的粗糙集聚类分析[J].统计研究,2005,8.
- [3]王钰,苗夺谦,周育健.关于Rough Set理论与应用的综述[J].模式识别与人工智能,1996,9(4):337-344.
- [4]张文修,吴伟志,梁吉业,李德玉.粗糙集理论与方法[M].北京:科学出版社,2001.
- [5]W. Ziarko. Introduction to the special issue on rough sets and knowledge discovery[J]. Communications of ACM, 1995, 38 (II):89-95.
- [6]W. Ziarko. Variable precision rough set model[J]. Journal of Information and System Science, 1993.