

三支决策与决策粗糙集融合模型^①

孟超, 余建坤

(云南财经大学 信息学院, 昆明 650221)

摘要: Paw lak 粗糙集模型没有对正域、边界域和负域赋予语义, 不能进行再决策, 而三支决策对边界域赋予了新的语义, 可以对边界域做出进一步刻画, 对于边界域的进一步划分, 依据属性的重要性, 使满足条件的样本划入再决策域, 不满足条件的样本继续保留在边界域中, 降低了边界域样本处理的失误率. 本文在对概率粗糙集模型、三支决策粗糙集的理论、贝叶斯理论的决策过程和决策粗糙集模型进行研究的基础上, 提出了一种三支决策与决策粗糙集融合模型, 与 Paw lak-三支决策模型相比, 其划分损失更小, 处理结果更优. 该模型运用三支决策理论对决策粗糙集的边界域赋予延迟决策的语义, 对于延迟决策再运用三支决策理论进行迭代操作, 对边界域样本进一步处理. 在迭代的过程中, 依据属性的重要程度将属性排序, 从而客观的得到迭代过程中每次优先依据哪个属性进行划分. 实验结果表明, 该模型比单一运用决策粗糙集模型进行决策代价小, 三支决策通过迭代对边界域处理的正确率有所提高, 这为准确决策提供了一种新的方法.

关键词: 三支决策; 决策粗糙集; 边界域样本处理

Merging Three-Way Decisions with Decision-Theoretic Rough Sets

MENG Chao, YU Jian-Kun

(College of Information, Yunnan University of Finance and Economics, Kunming 650221, China)

Abstract: Paw lak rough set model was lacking in giving semantic to positive regions, negative regions and boundary regions. The boundary could not make decision again. But three-way decisions gave a new semantic to boundary regions and we can deal with samples in boundary regions. Based on importance of attribute, samples which meet the conditions were delimit to decision region and others would be maintained in the boundary regions in order to reduce false positives when deal with samples in boundary regions. Based on the study of probabilistic rough set model, three-way decisions-theoretic rough set, Bayesian decision-making process and decisions-theoretic rough set model, this paper presents Three-way Decision mix Decision-Theoretic rough set model(TmD). Compared with the new model with Paw lak-three way decisions model, the loss of division of this model is smaller and the result is more reasonable. This model gives the boundary regions semantic which is delaying decisions. Using three-way decisions makes iterative operation for delaying decisions. In the process of iteration, attributes will be ordered based on the importance of the attribute, thus objectively get the attribute of priority being used in the process of iteration. Experimental results show that the model has a smaller decision cost than only using decision-theoretic rough sets and three-way decisions with iterative operation have a higher accuracy when deal with samples in boundary regions. This paper provides a new method for accurate decision-making.

Key words: three-way decisions; decision-theoretic rough sets; dealing with samples in boundary regions

粗糙集理论是由波兰学者 Z. Paw lak 在 1982 年提出的^[1], 它是一种能够处理不完整性和不确定性的

新型数学工具, 能有效地处理诸如不精确、不一致和不完整等各种不完备的信息, 目前, 该理论已在众多

^① 基金项目: 云南省高校商务智能科技创新团队

收稿时间: 2015-08-05; 收到修改稿时间: 2015-09-28

领域取得较好的应用. 传统的 Paw lak 理论通过运用等价关系将论域区分成多个等价类, 引入上近似、下近似等概念来描述知识的不确定性, 通过这些等价关系, 论域被划分成正域、边界域和负域三个不相交类, 然而, 实际应用中, 这样的划分过于严格, 将无效的属性直接剔除, 划分结果代价较高.

就此问题, Paw lak, Wong 和 Ziarko 提出了 0.5-概率粗糙集模型^[3], Ziarko 提出了变精度粗糙集模型^[2], Paw lak 和 Skowron 提出了粗糙隶属函数的概念^[4], Skowron 和 Stepaniuk 等人提出了参数化粗糙集模型^[5], Slezak 提出了贝叶斯粗糙集模型^[6]等, 这些模型都是用两个阈值 α 、 β 来划分正域、负域和边界域的, 但是阈值的选取都需要人为主观的确定. Yao, Wong 提出了决策粗糙集模型^[7,8], 描述了在贝叶斯期望风险最小时阈值 α 、 β 如何自动获取.

其派生出的三支决策思想赋予了粗糙集一种新的语义, 将边界域作为延迟决策从而可以进行再决策. 粗糙集是直接剔除无用属性, 而三支决策是选取最重要的属性作为决策依据, 决策代价更小. 这种模型在对于边界域样本处理上, 还需要相关知识的补充, 使得边界域样本不能有效的作为决策的正确依据.

本文提出的三支决策融合决策粗糙集模型, 将边界域中的样本依据属性的重要程度进行划分, 从而客观的提高边界域样本处理的正确率. 本文在对概率粗糙集模型、三支决策粗糙集理论、贝叶斯理论的决策过程和决策粗糙集模型进行研究的基础上, 提出了三支决策融合决策粗糙集模型, 该模型将三支决策理论与决策粗糙集相融合, 首先运用决策粗糙集模型对整个论域进行第一次划分, 分别得到正域、边界域、负域; 再将边界域作为延迟决策, 运用三支决策理论进行第二次划分, 把每个属性的重要性作为划分依据; 迭代使用三支决策理论, 将重要度次之的属性作为划分依据, 直至延迟决策域尽可能的转化成可决策样本. 这种新的模型有效减少决策规则的代价损失, 并为客观划分论域提供了客观的依据, 减少失误决策. 本文采用汽车评估数据进行验证, 对该数据集 6 种属性的 1728 个对象进行分类, 从实验结果可以看出, 这种新的三支决策融合决策粗糙集模型与决策粗糙集模型相比, 代价损失更低, 划分依据不需要主观的人为设定, 结果更客观贴近实际. 这对于制造厂商在生产何种车型, 保养费用高低, 车门数量选择, 承载人数选择, 安

全系数等方面的决策提供了重要的依据, 验证了该模型在解决实际问题中, 具有更精确的分类效果.

关键词库的结合大大提高了信息抽取算法的准确性和通用性, 基于 Web 信息抽取的混合交通出行方案生成与表示系统的成功实验也证明了本文提出的 Web 信息抽取算法的实用性.

1 决策粗糙集理论

决策粗糙集^[3]是基于贝叶斯决策思想的, 是一种具有合理语义解释的概率粗糙集. 通过考虑在不同状态下采取不同行动的损失或代价, 将论域进行划分.

1.1 概率粗糙集模型

相在实际应用中, 我们将知识和属性集放入一个决策表(一种重要的知识表达系统)中. 一个决策表描述了所有对象以及它们的条件属性和决策属性.

定义 1. 设 $S=(U, A, V, f)$ 是一个决策表, U 是论域, A 是属性集, $A=C \cup D, C \cap D=\emptyset$, V 是值域, f 是一个映射函数, C 是条件属性集, D 是决策属性集.

定义两个对象 x 和 y 是等价的, 只有当它们在 R 中具有相同的属性值. 称这个等价关系是 U 上的一个划分.

设 U 中的一个子集为 X , 上近似、下近似可以定义为^[1]:

$$\overline{apr}(x) = \{x \subset U \mid [x] \cap X \neq \emptyset\} \quad (1)$$

$$\underline{apr}(x) = \{x \subset U \mid [x] \subseteq X\}$$

基于上近似和下近似的定义, 任一子集 $C \subseteq U$, 可以将 U 进行如下划分:

$$\begin{aligned} POS(C) &= \underline{apr}(C) \\ NEG(C) &= U - \overline{apr}(C) \\ BND(C) &= \overline{apr}(C) - \underline{apr}(C) \end{aligned} \quad (2)$$

其中, $POS(C)$ 、 $NEG(C)$ 、 $BND(C)$ 被称为 C 的正域、负域和边界域. 正域是由完全确定属于某个集合的等价类所构成; 负域是由肯定不属于某个集合的所有的等价类构成; 边界域是由可能属于某个集合的等价类构成.

但是, Paw lak 模型中进入正域的必须是完全确定的规则, 这就需要引进概率粗糙集的相关理论.

概率粗糙集模型^[9]是通过引入一对阈值 α 、 β 来划分正域、负域和边界域的. 其取值的不同将产生不同的分类结果.

定义 2. 设 $S=(U, A, V, f)$ 是一个信息表, 其中

$0 \leq \beta \leq \alpha \leq 1$, $X \subset U$, 则定义基于 α 、 β 的上近似、下近似为^[10]:

$$\begin{aligned} \overline{apr}_{(\alpha,\beta)}(x) &= \{x \in U \mid \Pr(X \mid [x]) > \beta\} \\ \underline{apr}_{(\alpha,\beta)}(x) &= \{x \in U \mid \Pr(X \mid [x]) \geq \alpha\} \end{aligned} \quad (3)$$

在经典的粗糙集理论中, 仅将阈值的取值取为 0 和 1. 因此, 给出合理的 α 、 β 的取值, 是十分重要的. 将论域划分成 3 个区域, 通过引入贝叶斯决策理论, 来降低决策风险, 于是, 决策粗糙集相继被提出^[11].

1.2 贝叶斯理论决策过程

定义 3. 给定对象 x , 设一个信息表为 $S = (U, A, V, f)$, $\delta = \{w_1, w_2, \dots, w_m\}$ 是 X 的 m 个状态集, $\varphi = \{a_1, a_2, \dots, a_n\}$ 是 X 的 m 个行动集. $\Pr(w_i \mid x)$ 记做 X 在 w_i 状态下的条件概率. $\lambda(a_j \mid w_i)$ 记做 X 在 w_i 状态下执行 a_j 动作的损失. 则对于 x 的相关的期望损失^[13], 也就是条件风险可以记做:

$$R(a_j \mid x) = \sum_{i=1}^m \lambda(a_j \mid w_i) \cdot \Pr(w_i \mid x) \quad (4)$$

根据以上描述, 我们可以求出每个动作的损失函数, 并获得在不同状态下采取所有动作的期望损失, 按照风险最小的动作进行划分.

1.3 决策粗糙集模型

在 Paw lak 粗糙集理论中, 由于实际问题都受到噪声等因素的影响, 而决策粗糙集可以较好的解决这个问题, 以下对决策粗糙集模型的理论内容进行描述.

为了描述清楚, 利用两个状态集和三个决策集来描述决策过程^[12]. 设状态集 $\delta = \{X, -X\}$ 分别表示属于状态集 X 和不属于状态集 X , 设行动集 $\varphi = \{a_p, a_B, a_N\}$ 分别表示接受事件的行动、延迟决策事件的行动和拒绝执行事件的行动. 由于采取不同的行动会产生相应的损失, 用 $\lambda_{PP}, \lambda_{BP}, \lambda_{NP}$ 分别表示在 a_p, a_B, a_N 行动下的损失. 由此, 三种行动的相应期望损失可以描述为:

$$\begin{aligned} R(a_p \mid [x]) &= \lambda_{PP} \Pr(X \mid [x]) + \lambda_{PN} \Pr(-X \mid [x]) \\ R(a_B \mid [x]) &= \lambda_{BP} \Pr(X \mid [x]) + \lambda_{BN} \Pr(-X \mid [x]) \\ R(a_N \mid [x]) &= \lambda_{NP} \Pr(X \mid [x]) + \lambda_{NN} \Pr(-X \mid [x]) \end{aligned} \quad (5)$$

由贝叶斯公式, 通过选择损失最小的行动可以得到以下 3 个决策规则:

(P): 当 $R(a_p \mid [x]) \leq R(a_B \mid [x])$ 且 $R(a_p \mid [x]) \leq R(a_N \mid [x])$ 时, 执行决策 $x \in POS(C)$;

(B): 当 $R(a_B \mid [x]) \leq R(a_p \mid [x])$ 且 $R(a_B \mid [x]) \leq R(a_N \mid [x])$ 时, 执行决策 $x \in BND(C)$;

(N): 当 $R(a_N \mid [x]) \leq R(a_p \mid [x])$ 且 $R(a_N \mid [x]) \leq R(a_B \mid [x])$ 时, 执行决策 $x \in NEG(C)$;

其中, 它们损失代价的大小, 可以看出有: $\lambda_{PP} \leq \lambda_{BP} \leq \lambda_{NP}, \lambda_{NN} \leq \lambda_{BN} \leq \lambda_{PN}$, 而且 $\Pr(X \mid [x]_R) + \Pr(-X \mid [x]_R) = 1$, 再令:

$$\begin{aligned} \alpha &= \frac{\lambda_{PN} - \lambda_{BN}}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})} \\ \beta &= \frac{\lambda_{PN} - \lambda_{NN}}{(\lambda_{PN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})} \\ \lambda &= \frac{\lambda_{BN} - \lambda_{NN}}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})} \end{aligned} \quad (6)$$

则上述规则可以描述为:

当 $\Pr(X \mid [x]_R) \geq \lambda$ 且 $\Pr(X \mid [x]_R) \geq \alpha$ 时, 执行决策 $x \in POS(C)$;

当 $\Pr(X \mid [x]_R) \geq \beta$ 且 $\Pr(X \mid [x]_R) \leq \alpha$ 时, 执行决策 $x \in BND(C)$;

当 $\Pr(X \mid [x]_R) \leq \beta$ 且 $\Pr(X \mid [x]_R) \leq \lambda$ 时, 执行决策 $x \in NEG(C)$;

我们加入一个条件在损失函数中: $(\lambda_{PN} - \lambda_{NN}) \cdot (\lambda_{NP} - \lambda_{BP}) > (\lambda_{BP} - \lambda_{PP}) \cdot (\lambda_{BN} - \lambda_{NN})$, 通过证明即有 $\alpha > \lambda > \beta$. 只用 α, β 来定义决策规则, 可得:

(P): 当 $R(C \mid [x]) \geq \alpha$ 时, 执行决策 $x \in POS(C)$;

(B): 当 $\beta < R(C \mid [x]) < \alpha$ 时, 执行决策 $x \in BND(C)$;

(N): 当 $R(C \mid [x]) \leq \beta$ 时, 执行决策 $x \in NEG(C)$;

决策粗糙集不但引入了概率统计的概念, 而且可以用一种易于理解的计算方法来确定阈值. 因此, 决策粗糙集在解决数据分类问题中具有更重要的可操作价值.

2 三支决策模型语义

在经典的 Paw lak 粗糙集理论中, 由于正域是只包含完全确定的规则, 边界域包含的是可能性规则, 导致了负域恒为空从而失去了实际意义. 但是, 按照决策粗糙集的理论, 通过两个阈值 α 、 β 将论域分成 3 个区域, 分别是决策正域、决策负域和延迟决策边界域. 在语义上来说, 这样划分出来的三个区域是根据三个确定的规则得来的, 具有更完整的语义.

三支决策粗糙集的语义^[13]可以描述为: 当 $[x]_R$ 中 X 发生的可能性大于 α 时, 就把 $[x]_R$ 划分到正域中并立即执行这个决策; 当 $[x]_R$ 在 X 发生的可能性小于 β , 就把 $[x]_R$ 划分到负域中并拒绝执行这个决策; 当

$[x]_R$ 在 X 发生的可能性在 α 和 β 之间, 就把 $[x]_R$ 划分到边界域中并延迟做出决策.

Yao 提出了一种新的决策语义^[14], 即把三支决策粗糙集中的绝对阈值函数用相对阈值函数取代. 为选取适当的阈值 α 、 β 提供了依据.

3 三支决策融合决策粗糙集模型

决策粗糙集理论依据贝叶斯期望损失来确定阈值 α 、 β , 可以非常合理、客观的对论域进行划分. 将三支决策赋予决策粗糙集新的语义, 可以对边界域样本进行再处理. 为了提高边界域样本处理的正确性, 将对延迟决策域运用三支决策理论进一步划分.

其基本思想是, 将论域运用决策粗糙集理论进行划分, 得到正域、边界域和负域, 正域和负域中的规则放入最终的二支决策中. 运用三支决策理论进行迭代操作. 由于第一次划分是非常客观的, 阈值 α 、 β 可以合理的通过贝叶斯决策确定出来; 第二次为了避免用主观经验和试凑的方式确定阈值, 本文运用属性的重要性对边界域进行在处理.

基于文献[12], 从犯错损失的大小角度, 证明在条件:

$$\lambda_{pp} < \lambda_{BP} < \lambda_{NP} \text{ 且 } \lambda_{NN} < \lambda_{BN} < \lambda_{PN} \quad (C_1)$$

$$\frac{\lambda_{BP} - \lambda_{PP}}{\lambda_{PN} - \lambda_{BN}} < \frac{\lambda_{NP} - \lambda_{BP}}{\lambda_{BN} - \lambda_{NN}} \quad (C_2)$$

(C₁)、(C₂)同时成立的情况下, 三支决策融合决策粗糙集模型决策方式比 Paw lak-三支决策损失更小, 在条件(C₁)、(C₂)下更优.

证明: 假设

- $a_1 = \{x \in U \mid \Pr(X|[x])=1\}$
- $a_2 = \{x \in U \mid x \in X, \alpha \leq \Pr(X|[x]) < 1\}$
- $a_3 = \{x \in U \mid x \in X, \beta < \Pr(X|[x]) < \alpha\}$
- $a_4 = \{x \in U \mid x \in X, 0 < \Pr(X|[x]) \leq \beta\}$
- $b_1 = \{x \in U \mid \Pr(X|[x])=0\}$
- $b_2 = \{x \in U \mid x \in -X, 0 < \Pr(X|[x]) \leq \beta\}$
- $b_3 = \{x \in U \mid x \in -X, \beta < \Pr(X|[x]) < \alpha\}$
- $b_4 = \{x \in U \mid x \in -X, \alpha \leq \Pr(X|[x]) \leq 1\}$.

由上述定义, $a_2 + b_4 = a_4 + b_2 = |U|$, 则 $\frac{a_2}{a_2 + b_4} \geq \alpha$, $\frac{a_4}{a_4 + b_2} \geq \beta$. 由式(4)(5)有: $\lambda_{BN} \cdot a_2 \geq (\lambda_{PN} - \lambda_{BN})b_4$, $(\lambda_{NN} - \lambda_{BP})a_4 \leq \lambda_{BN} \cdot b_2$.

新模型与 Paw lak-三支决策方式在 $(0, \beta]$ 和 $[\alpha, 1)$

区间内不同, 新模型在这两个区间内的总损失为:

$$Cost_1 = \lambda_{NP} \cdot a_4 + \lambda_{PN} \cdot b_4, \text{ Paw lak-三支决策在这两个区间内的总损失为: } Cost_2 = \lambda_{BP} \cdot (a_2 + a_4) + \lambda_{BN} \cdot (b_2 + b_4).$$

两者总损失相减: $Cost_2 - Cost_1 \geq 0$.

因此, Paw lak-三支决策方式比三支决策融合决策粗糙集模型决策方式总损失更大, 新模型在条件(C₁)、(C₂)下更优, 更合理.

3.1 三支决策与决策粗糙集融合模型结构

本首先依据决策粗糙集理论, 在等价关系 R 下, 利用阈值 α 、 β , 对数据集 C 进行第一次划分:

(P): 当 $R(C|[x]) \geq \alpha$ 时, 执行决策 $x \in POS_{(\alpha, \beta)}(C)$;

(B): 当 $\beta < R(C|[x]) < \alpha$ 时, 执行决策 $x \in BND_{(\alpha, \beta)}(C)$;

(N): 当 $R(C|[x]) \leq \beta$ 时, 执行决策 $x \in NEG_{(\alpha, \beta)}(C)$;

定义 4. 属性的重要程度. C 是条件属性集, D 是决策属性集. C 的子集 $C' \subseteq C$ 对于 D 的重要程度定义为:

$$\delta_{CD}(C') = \gamma_C - \gamma_{C-(C')}(D) \quad (9)$$

当 $C' = \{a\}$ 时, 属性 $a \in C$ 对于 D 的重要程度为:

$$\delta_{CD}(a) = \gamma_C - \gamma_{C-(a)}(D)$$

将 $BND_{(\alpha, \beta)}(C)$ 作为延迟决策域, 利用三支决策理论对延迟决策进行迭代操作, 在每次的迭代操作过程中, 属性的选取依据属性的重要程度, 选择属性中比重最大的属性作为依据. 属性 A_1 对于决策属性集 D 的重要程度定义为:

$$\delta_{CD}(A_1) = \gamma_C - \gamma_{C-(A_1)}(D) \quad (10)$$

属性粒度信息选择如图 2. 先依据重要性最大的属性 A_1 对决策域进行划分, 将划分结果中的正域和负域用来支持决策, 结果中的边界域继续作为延迟决策域, 依据属性 A_2 进行划分, 迭代的重复运用三支决策进行划分, 直至不能再进行划分为止.

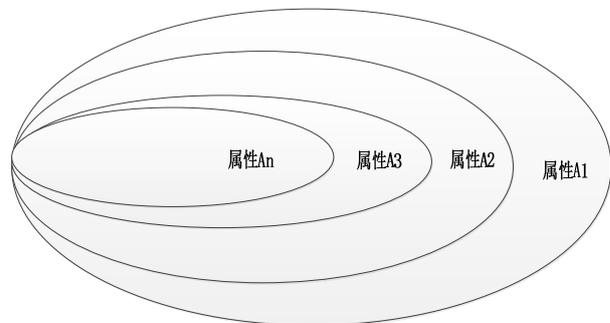


图 2 属性粒度信息选择

3.2 三支决策与决策粗糙集融合模型算法

给定一个信息表 $S=(U, A, V, f)$, U 是论域, A 是属性集, $A=C \cup D, C \cap D = \emptyset$, V 是值域, f 是一个映射函数, C 是条件属性集, D 是决策属性集. 则三支决策融合决策粗糙集算法 TmD(Three-way Decision mix Decision-Theoretic rough set)如下:

输入: 决策表 $S=(U, A, V, f)$, 待划分对象 x

输出: 正域 P , 负域 N , 最终延迟决策域 B

Step1: $P=\{\}, B=\{\}, P'=\{\}, B'=\{\}$;

Step2: for $i = 1, \dots, m$;

Step3: 由(5), (6)式确定阈值 α, β ;

Step4: 由(8)式, 遍历 x 并将其划分到 $P=\{\}, B=\{\}, N=\{\}$ 中;

Step5: 由(9)式, 按照重要程度将所有属性 A_i 从大到小排列, 依次选取 A_i 作为划分依据对 B 进行划分, 将不满足条件的样本继续保留在延迟决策域 B 中;

Step6: 当所有属性迭代完成后, 三支决策理论将延迟决策域中所有对象划分到正域、负域中, 或者延迟决策域已经不能再进行划分的时候, 得到最终结果正域 P , 负域 N 和边界域 B ;

TmD 算法首先将全体对象进行划分, 需要计算每个对象被划分到 3 个域中的期望损失, 然后按照最小的期望损失将论域划分成 3 个部分, 复杂度为 $O(3n)$. 之后每次在对延迟决策域进行划分, 复杂度同样是 $O(3n)$, 有 t 个待决策属性, 复杂度为 $O(3tm)$.

4 实验结果与分析

根据以上所述, 本文提出的 TmD 模型可以降低决策代价, 提高最终二支决策的正确性. 为了证明这一点, 我们采用了一个对比试验来说明. 该实验的数据来自于 UCI 数据库(<http://archive.ics.uci.edu/ml/datasets/Car+Evaluation>). 实验采用汽车评估的数据, 共计 1728 个对象, 每条数据有 6 个属性. 基于粗糙集工具 Rosetta 完成, 因为 Rosetta 所能处理的信息系统的条件属性的个数有限, 所以本文选用了条件属性的个数较少的 CarEvaluation 数据集.

实验环境为 64 位 Window 7 操作系统, Intel 3.20GHz 主频, 8.00 GB 内存. 首先将数据在 Excel 中进行分列, 导入到工具软件 Rosetta 中. 由于 6 个属性都是离散的数据, 只需将数据补全即可. 然后将数据进行属性约简, 结果如图 3. 生成规则 Generate rules 结

果如图 4. 迭代运用三支决策模型, 遍历本题库求得每一个属性的比重值, 最后调用 Approximate decision class, 按照属性比重的排序依次选取属性作为决策属性进行划分, 最终结果统计如表 1.

	Reduct	Suppor	Length
1	{buying, maint, doors,	100	6

图 3 属性约简结果

	Rule	L	R	RHS	LHS Cov	RHS Cov	RH	L	RH
1	buying (vhigh) AND maint (vhigh) AND doors	1	1	1.0	0.000579	0.000826	1.0	6	1
2	buying (vhigh) AND maint (vhigh) AND doors	1	1	1.0	0.000579	0.000826	1.0	6	1
3	buying (vhigh) AND maint (vhigh) AND doors	1	1	1.0	0.000579	0.000826	1.0	6	1
4	buying (vhigh) AND maint (vhigh) AND doors	1	1	1.0	0.000579	0.000826	1.0	6	1
5	buying (vhigh) AND maint (vhigh) AND doors	1	1	1.0	0.000579	0.000826	1.0	6	1
6	buying (vhigh) AND maint (vhigh) AND doors	1	1	1.0	0.000579	0.000826	1.0	6	1

图 4 生成规则结果

表 1 实验结果对比

	POS	BND	NEG
决策粗糙集	474	722	532
TmD	742	52	868

根据结果可以得出, 属性的重要程度依次为价格、发动机性能、后备箱、安全性、车门数和乘车人数. 因为市场上私家车的可乘人数基本是固定的, 乘车人数又决定了车门数, 所以这两个属性重要程度就很低. TmD 模型可以有效将粗糙集模型中的边界域进行再决策, 降低了决策代价, 客观的降低了失误决策的几率. 为生产厂家来生产何种车型和配置提供了重要的决策依据.

由决策粗糙集与 TmD 模型实验结果对比表中可以看出, TmD 模型有效的将延迟决策域进行再决策. 这种新的模型可以进一步减少决策代价, 并提高决策的准确性, 为最终的二支决策提供了有力的支持, 这样对于决策者做出正确的决策来说更为有利. 但是再对边界域进行再处理的时候, 仍不能完全划分成二支决策, 需要补充更多的相关知识才可以进行划分. 目前对边界域样本有两种处理方案, 一种是对边界域样本进行全部处理, 提出了万有引力原则、距边界最近原则、距中心最近原则 3 种方法, 这些方法虽然对边界域样本可以完全划分, 但对边界域样本处理的正确率较低; 另一种是对边界域样本进行部分处理, 即作进一步的划分边界域样本满足一定的条件, 这样边界域中剩余的样本是不满足特定条件的, 提高了边界域样本处理的正确率, 但不能对边界域进行完全处理.

因此对于三支决策融合决策粗糙集模型还需做进一步研究.

5 结语

本文针对 Paw lak 粗糙集理论决策代价较高, 其派生出的三支决策对于边界域样本处理的正确率较低问题, 在对概率粗糙集模型、三支决策粗糙集的理论、贝叶斯理论的决策过程和决策粗糙集模型进行研究的基础上, 提出了三支决策与决策粗糙集融合模型, 实验结果表明, 这种新的三支决策与决策粗糙集融合模型与传统的决策粗糙集模型相比, 延迟决策域样本处理的正确率提高, 决策代价降低, 说明该方法是可行的, 对拓展决策粗糙集理论的应用也有一定的参考价值.

参考文献

- 1 Pawlak Z. Rough sets. *International Journal of Computer and Information Sciences*, 1982, 11(5): 341–356.
- 2 Ziarko W. Variable precision rough set model. *Journal of Computer and System Sciences*, 1993, 46(1): 39–59.
- 3 Paw lak Z, Wong SKM, Ziarko W. Rough sets: Probabilistic versus deterministic approach. *International Journal of Man-Machine Studies*, 1988, 29(1): 81–95.
- 4 Paw lak S. *Rough Membership Functions*. New York: John Wiley and Sons, 1994: 251–271.
- 5 Skow Ron S. Tolerance approximation spaces. *Fundamental Information*, 1996, 27(2): 245–253.
- 6 Slezak. Rough sets and bayes factor. *Lecture Notes in Computer Science*, 2005: 314–324.
- 7 Yao YY, Wong. A decision theoretic framework for approximating concepts. *Inter. Journal of Man-Machine Studies*, 1992, 37(6): 793–809.
- 8 Yao YY. Decision-theoretic rough set models. *Lecture Notes in Artificial Intelligence*, 2007, 4481: 1–12.
- 9 Yao YY. Probabilistic approaches to rough sets. *Expert Systems*, 2003, 20(5): 287–297.
- 10 李健,常太华,杨婷婷.变精度粗糙集模型属性约简分析. *计算机工程与应用*, 2012, 48(13): 130–132.
- 11 Yao YY, Zhao Y. Attribute reduction in decision theoretic rough set model. *Information Sciences*, 2008, 49(2): 255–271.
- 12 刘盾,姚一豫,李天瑞.三支决策粗糙集. *计算机科学*, 2011, 38(1): 246–250.
- 13 李华雄,刘盾,周献中.决策粗糙集模型研究综述. *重庆邮电大学学报(自然科学版)*, 2010, 22(5): 624–630.
- 14 Yao YY. Three-way decisions with probabilistic rough sets. *Information Sciences*, 2010, 180(3): 341–356.
- 15 Baader F, Horrocks I, Sattler U. *Description logics as ontology languages for the semantic web*. *Lecture Notes in Artificial Intelligence*, 2003.