

三支决策空间下的区间参数优化模型及应用

李明霞 刘保相 张春英

(华北理工大学理学院 唐山 063009) (河北省数据科学与应用重点实验室 唐山 063009)

摘要 区间概念格理论是在区间参数基础上提出的一种新的对象挖掘方法,它能够更精确地处理不确定信息。区间参数 $[\alpha, \beta]$ 决定了区间概念和格结构,进而对提取的决策准则产生影响。为了解决区间参数优化问题,首先将区间概念格与三支决策粗糙集理论相结合,提出了三支决策空间理论;然后在此基础上将区间概念外延划分为3个域:正域、边界域和负域,并给出了基于区间概念格的三支决策规则与决策损失函数,通过调整区间参数寻找更加可靠的决策准则,从而优化区间参数;最后通过实例对模型进行验证。

关键词 区间概念格,三支决策空间,损失函数,参数优化

中图分类号 TP18 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.01.016

Interval Parameters Optimization Model under Three-way Decisions Space and Its Application

LI Ming-xia LIU Bao-xiang ZHANG Chun-ying

(College of Science, North China University of Science and Technology, Tangshan 063009, China)

(Key Laboratory of Data Science and Application of Hebei Province, Tangshan 063009, China)

Abstract The interval concept lattice theory is a new method of mining objects based on interval parameters. It can more accurately deal with uncertain information. Interval parameters α and β can determine interval concepts and lattice structure, and then affect extracted decision rules. In order to solve the optimization problem of interval parameters, firstly we combined the theories of interval concept lattice and three-way decision-theoretic rough set, and then put forward three-way decision space theory. Secondly, according to the theory, the extension of interval concept was divided into positive region, Negative region and boundary region, moreover, the three-way decision rules and decision loss function were proposed. Through adjusting interval parameters, we can find more credible decision rules, and then optimize interval parameters. Finally we verified the model by an example.

Keywords Interval concept lattice, Three-way decision space, Loss function, Parameters optimization

1 引言

三支决策理论是传统二支决策理论的扩展。二支决策只考虑接受与拒绝两种选择,但是在实际应用中,由于信息的不精确性或不完全性,常常无法做到接受或拒绝,因此在这两种极端情况下引入了第三种情况,即不承诺选择。为了给概率粗糙集和决策粗糙集的3个域提供一个合理的语义解释,姚一豫提出了三支决策的概念^[1]。

对于决策粗糙集三支决策定量模型^[2-5],在决策粗糙集模型中,正域、边界域和负域是由一对阈值 $[\alpha, \beta]$ 决定的,与Pawlak的3个域相比, α 和 β 分别取代了0和1,它们能够容忍一定程度上的错误,阈值给出了一种基于错误分类率的解释。这两个阈值参数由损失函数 λ 决定,但这个损失参数的大小却是由实验或者专家的意见给出,以及通过人类的经验获得,具有很大的主观性。因此,阈值 $[\alpha, \beta]$ 的解释与计算以及概率正域、边界域和负域的解释和应用成为了目前需要解决的两大问题。概率粗糙集的主要结果是在决策粗糙集模型的基础上提出并研究的,概率粗糙集模型^[6,7]的概率阈值 $[\alpha, \beta]$ 利用对象分类正确还是错误的代价来决定不同的域的大

小。由此可以看出,三支决策的理论基础就是基于贝叶斯决策理论的决策粗糙集模型。通过进一步研究发现,三支决策不局限于粗糙集,而是一种更一般的、有效的决策和信息处理模式。

区间概念格理论是在区间参数 $[\alpha, \beta]$ 的基础上提出的一种新的对象挖掘方法,它能够更精确地处理不确定信息。区间参数 $[\alpha, \beta]$ 可将对象论域 U 按照满足条件属性的情况分为3个区域,这种划分与概率粗糙集域的划分有很大的相似之处,通过构建区间概念格可以为处理边界样本和降低决策损失^[6,7]提供依据。区间参数的变化能够改变概念和格结构,进而对决策依据产生影响,因此在三支决策空间下研究区间参数的优化问题是有意义的。

基于以上分析,借鉴概率粗糙集的三支决策理论,并结合区间概念格特性,定义了基于区间概念的决策损失函数,进而给出区间三支决策概念,并构建区间三支决策空间;当区间参数等步长变化时,三支决策空间中的决策概念也会随之变化,最终对用户的决策准则产生影响,从而进一步帮助优化区间参数;最后通过实例演示了区间参数对决策准则的影响和优化区间参数的过程。

到稿日期:2015-09-14 返修日期:2015-11-22 本文受国家自然科学基金(61370168, 61472340),河北省科技厅条件建设项目(14960112D)资助。

李明霞(1992—),女,硕士,主要研究方向为概念格、数据挖掘等,E-mail: nicemingxia@sina.cn; 刘保相(1957—),男,教授,主要研究方向为模糊控制、概念格、数据挖掘等,E-mail: liubx5888@126.com(通信作者); 张春英(1969—),女,博士,教授,主要研究方向为概念格、人工智能、多关系数据挖掘等,E-mail: zchunying@ncst.edu.cn。

2 理论基础

2.1 三支决策与粗糙集

设 U 是有限实体集,关系 $E \subseteq U \times U$ 是实体集 U 上的等价关系,即 E 是自反的、对称的和传递的。包含实体 x 的 E 的等价类记为 $[x]_E = [x] = \{y \in U | xEy\}$ 。所有等价类的集合记为 $U/E = \{[x]_E | x \in U\}$,称为商集,也称作对实体的划分^[8]。

定义 1 设阈值 $[\alpha, \beta]$ 满足条件 $0 \leq \alpha < \beta \leq 1$,则在 $[\alpha, \beta]$ -概率粗糙集中,集合 X 的下近似和上近似可以表示为:

$$\begin{aligned} \underline{apr}_{(\alpha, \beta)}(X) &= \bigcup \{[x] \in U/E | Pr(X|[x]) \geq \beta\} \\ \overline{apr}_{(\alpha, \beta)}(X) &= \bigcup \{[x] \in U/E | Pr(X|[x]) \geq \alpha\} \end{aligned}$$

其中, $X \subseteq U$ 是实体集的一个子集, $Pr(X|[x])$ 表示实体 y 属于等价类 $[x]$ 时, y 属于集合 X 的条件概率,换言之,其表示该类实体属于外延 X 的置信度。概率粗糙集的主要结果是在决策粗糙集模型的基础上提出并研究的。

定义 2 根据上、下近似,概率粗糙集正、负和边界域分别定义为:

$$\begin{aligned} POS_{(\alpha, \beta)}(X) &= \underline{apr}_{(\alpha, \beta)}(X) \\ &= \bigcup \{x \in U/E | Pr(X|[x]) \geq \beta\} \\ NEG_{(\alpha, \beta)}(X) &= (\overline{apr}_{(\alpha, \beta)}(X))^c \\ &= \{x \in U/E | Pr(X|[x]) < \alpha\} \\ BND_{(\alpha, \beta)}(X) &= (POS_{(\alpha, \beta)}(X) \cup NEG_{(\alpha, \beta)}(X))^c \\ &= \{x \in U | \alpha \leq Pr(X|[x]) < \beta\} \end{aligned}$$

其中,上近似 $\overline{apr}_{(\alpha, \beta)}(X)$ 的补表示为 $(\overline{apr}_{(\alpha, \beta)}(X))^c = U - \overline{apr}_{(\alpha, \beta)}(X)$ 。3 个概率区间是两两不相交的,且并集是实体集 U 。

定义 3 设 $Pr(X|[x]) \geq \beta$ 时选择接受,当接受 $[x]$ 的所有实体时,可能导致错误,那么正域的错误接受率为: $IAE(POS_{(\alpha, \beta)}(X), X) = \frac{|POS_{(\alpha, \beta)}(X) \cap X^c|}{|POS_{(\alpha, \beta)}(X)|}$

设 $Pr(X|[x]) \leq \alpha$ 时,选择拒绝,当拒绝 $[x]$ 的所有实体时,可能导致错误,那么负域的错误拒绝率为: $IRE(NEG_{(\alpha, \beta)}(X), X) = \frac{|NEG_{(\alpha, \beta)}(X) \cap X|}{|NEG_{(\alpha, \beta)}(X)|}$

当置信度太低而不足以支持接受或者太高而不足以支持拒绝时,则选择不承诺。这可能导致两类错误,即正例不承诺和负例不承诺:

$$\begin{aligned} NPE(NEG_{(\alpha, \beta)}(X), X) &= \frac{|BND_{(\alpha, \beta)}(X) \cap X|}{|BND_{(\alpha, \beta)}(X)|} \\ NNE(POS_{(\alpha, \beta)}(X), X) &= \frac{|BND_{(\alpha, \beta)}(X) \cap X^c|}{|BND_{(\alpha, \beta)}(X)|} \end{aligned}$$

正因为概率粗糙集模型允许一定程度的错误,因此其边界域可能相对较小。

2.2 区间概念格

定义 4^[9,10] 对于形式背景 (U, A, R) , $L(U, A, R)$ 是其诱导的经典概念格, $RL(M, N, Y)$ 是粗糙概念格。设有区间 $[\alpha, \beta]$, $0 \leq \alpha < \beta \leq 1$, 则

$$\begin{aligned} \alpha \text{ 上界外延 } M^\alpha & \text{ 为:} \\ M^\alpha &= \{x | x \in U, |f(x) \cap Y| / |Y| \geq \alpha, 0 \leq \alpha \leq 1\} \\ \beta \text{ 下界外延 } M^\beta & \text{ 为:} \\ M^\beta &= \{x | x \in U, |f(x) \cap Y| / |Y| \geq \beta, 0 \leq \alpha < \beta \leq 1\} \end{aligned}$$

其中, Y 是概念的内涵。 $|Y|$ 是集合 Y 中包含元素的个数,即基数。 M^α 表示可能被 Y 中至少 $\alpha \times |Y|$ 个内涵属性覆盖的对象, M^β 表示可能被 Y 中至少 $\beta \times |Y|$ 个内涵属性所覆盖的对象。

定义 5^[9,10] 设形式背景 (U, A, R) , 三元序偶 (M^α, M^β, Y) 被称为区间概念,其中, Y 为内涵,是概念的描述; M^α 为 α 上界外延; M^β 为 β 下界外延。

定义 6^[9,10] 用 $L_\beta^p(U, A, R)$ 表示形式背景 (U, A, R) 的全体 $[\alpha, \beta]$ 区间概念, $(M_1^\alpha, M_1^\beta, Y_1) \leq (M_2^\alpha, M_2^\beta, Y_2) \Leftrightarrow Y_1 \subseteq Y_2$, 则“ $\leq$ ”是 $L_\beta^p(U, A, R)$ 上的偏序关系,满足该偏序关系的所有概念构成了形式背景 (U, A, R) 下的区间概念格 $L_\beta^p(U, A, R)$ 。

定义 7 设形式背景 $(U, C \cup D, R)$ 在区间 $[\alpha, \beta]$ 上决定的区间概念格为 $L_\beta^p(U, C \cup D, R)$, $C = (M^\alpha, M^\beta, Y)$ 是格中的一个区间概念,区间概念的上、下界外延将论域 U 分为 3 个域:

$$\begin{aligned} POS_\alpha^p(X) &= M^\beta = \{x | x \in U, |f(x) \cap Y| / |Y| \geq \beta\} \\ BND_\alpha^p(X) &= M^\alpha - M^\beta \\ &= \{x | x \in U, \alpha \leq |f(x) \cap Y| / |Y| < \beta\} \\ NEG_\alpha^p(X) &= U - M^\beta = \{x | x \in U, |f(x) \cap Y| / |Y| < \alpha\} \end{aligned}$$

其中, $X \subseteq U$ 是实体集的一个子集。若 $x \in POS_\alpha^p(X)$, 则对 x 进行接收决策;若 $x \in NEG_\alpha^p(X)$, 则对 x 进行拒绝决策;否则做出不承诺决策。

2.3 区间三支决策空间

定义 8 当对象 x 属于 X 时, $x \in X$, 令 $\lambda_{PP}, \lambda_{NP}, \lambda_{BP}$ 代表将一个对象分别划分到 $POS_\alpha^p(X), NEG_\alpha^p(X)$ 和 $BND_\alpha^p(X)$ 的代价函数;反之,当对象 x 不属于 X 时, $x \notin X$, 令 $\lambda_{PN}, \lambda_{NN}, \lambda_{BN}$ 代表将一个对象分别划分到 $POS_\alpha^p(\bar{X}), NEG_\alpha^p(\bar{X})$ 和 $BND_\alpha^p(\bar{X})$ 的代价函数。

设 $\zeta = \{a_P, a_B, a_N\}$ 分别表示当前对象属于某属性集的 3 个区域中的可能状态,不同状态下,对象 x 采取不同划分方案的风险代价^[11]如表 1 所列。

表 1 不同决策方案在不同状态下的风险代价

	a_P	a_B	a_N
X	λ_{PP}	λ_{BP}	λ_{NP}
X^c	λ_{PN}	λ_{BN}	λ_{NN}

通常,代价函数满足 $\lambda_{PP} \leq \lambda_{BP} < \lambda_{NP}$ 和 $\lambda_{NN} \leq \lambda_{BN} < \lambda_{PN}$, 解释为:对于一个属于 X 的对象 x , 将其划分到 $POS_\alpha^p(X)$ 的风险代价要小于或等于将其划分到 $BND_\alpha^p(X)$ 的风险代价,同时这两者的风险代价都小于将其划分到 $NEG_\alpha^p(X)$ 的风险代价。同理,对不属于 X 的对象 x , 将其划分到 $NEG_\alpha^p(\bar{X})$ 的风险代价要小于或等于将其划分到 $BND_\alpha^p(\bar{X})$ 的风险代价,同时这两者的风险代价都小于将其划分到 $POS_\alpha^p(\bar{X})$ 的风险代价^[8]。

定义 9 采取 a_P, a_B, a_N 3 种决策行动下的期望损失函数^[12-14]分别表示为:

$$\begin{aligned} R(a_P) &= \lambda_{PP} (|X \cap M^\beta| / |M^\beta|) + \lambda_{PN} (|\bar{X} \cap M^\beta| / |M^\beta|) \\ R(a_B) &= \lambda_{BP} (|X \cap (M^\alpha - M^\beta)| / |M^\alpha - M^\beta|) + \lambda_{BN} (|\bar{X} \cap (M^\alpha - M^\beta)| / |M^\alpha - M^\beta|) \\ R(a_N) &= \lambda_{NP} (|X \cap (U - M^\alpha)| / |U - M^\alpha|) + \lambda_{NN} (|\bar{X} \cap (U - M^\alpha)| / |U - M^\alpha|) \end{aligned}$$

定义 10 设形式背景 $(U, C \cup D, R)$ 决定的区间概念格 $L_\beta^p(U, C \cup D, R)$, $C = (M^\alpha - M^\beta, Y)$ 是格中的区间概念, 则称 $\tilde{C} = (M^\alpha - M^\beta, Y; R(a_P), R(a_B), R(a_N))$ 为三支决策概念。

定义 11 设形式背景 $(U, C \cup D, R)$, 由区间三支决策概念及父子关系构成的区间三支决策空间为 $\tilde{L}_\beta^p(U, C \cup D, R)$ 。

则对于一个新实体 x , 由区间三支决策概念得到的决策

由决策动作 a_i (i 表示接受动作 P , 不承诺动作 B 或拒绝动作 N) 及对应的决策损失 R 共同构成, 记为 $J = (a_i, R(a_i))$ 。

从形式背景生成区间概念格的过程实质上是概念聚类过程, 区间概念内涵之间的包含关系决定了格结构中的父子关系。先验的形式背景生成的区间概念格构成了三支决策空间, 格中区间概念划分的 3 个决策域可作为新对象的决策依据, 决策空间中的父子关系可以决定下一步的动作, 以降低决策的损失。

3 三支决策空间下的区间参数优化

3.1 给定区间参数下的决策寻优算法

在决策空间中, 某一对象 x 可依据多个区间三支决策概念做出不同的决策, 若得到了决策准则 J 为不承诺, 对冒险主义者而言, 他们更愿意做出损失相对较小的接受或拒绝决策 J' 。设 J' 对应的区间三支决策概念为 $\tilde{C}$, 当做出损失相对较小的接受或拒绝决策后, 可以根据 $\tilde{C}$ 的子概念划分的决策区域来降低对实体采取接受或拒绝决策的损失。

算法 1 定参数下决策寻优算法 GPOA

输入: 决策形式背景 $(U, C \cup D, R)$, 区间参数 $[\alpha_0, \beta_0]$, 新的对象 x 及其条件属性集合 $A (A \subseteq C)$

输出: 对象 x 的决策准则

- Step1 根据给定的决策形式背景 $(U, C \cup D, R)$, 由区间三支决策概念 $\tilde{C} = (M^0, M^0, Y; R(a_P), R(a_B), R(a_N))$ 及父子关系构建区间三支决策空间 $L_{\alpha_0}^{\beta_0}(U, C \cup D, R)$ (定义 9—定义 11);
- Step2 在区间三支决策概念 $\tilde{C}$ 中寻找内涵 Y 。若 $Y - (Y \cap D) = A$, 转 Step3; 若 $Y - (Y \cap D) \neq A$ 且 $(Y - (Y \cap D)) \cap A \neq \emptyset$, 转 Step 6;
- Step 3 若存在 n 个概念 $\tilde{C} = (M^0, M^0, A; R(a_P), R(a_B), R(a_N))$, 则可得 n 个决策准则, 即 $J_1, J_2, J_3, \dots, J_n$, 构成 n 维决策空间 $JS = \{J_1, J_2, J_3, \dots, J_n\}$;
- Step 4 若有唯一 J_k 满足 $R_{Pk} = \min(R_{P1}, R_{P2}, \dots, R_{Pn})$, 则将决策 J_k 作为对 x 最终的决策准则, 即将非不承诺 (接受或拒绝) 损失最小的决策作为最终决策;
- Step 5 若有多个决策 J_i, J_m 满足 $R_{Pi} = R_{Pm} = \min(R_{P1}, R_{P2}, \dots, R_{Pn})$, 则搜索 $\tilde{C}_i$ 和 $\tilde{C}_m$ 的子概念, 根据这些子概念的内涵为 x 增加衡量属性, 直至得到满足条件的唯一决策;
- Step 6 依据三支决策空间中的父子关系, 搜索内涵为 Y' 的概念, Y' 满足条件 $A \subseteq Y'$ 且 $|Y' - A| = 1$, 转 Step3, 否则转 Step7;
- Step 7 搜索内涵为 Y'' 的概念, Y'' 满足条件 $A \subseteq Y''$ 且 $|Y'' - A| = 2$, 转 Step3 继续决策, 直至内涵为 $\emptyset$ 的末端结点。

算法结束。

3.2 变区间参数的三支决策空间更新

3.1 节已经给出了在固定区间参数的基础上寻找新对象 x 的决策准则算法, 本节将给出变参数下三支决策空间更新算法。区间参数 $[\alpha, \beta]$ 的变化首先会导致区间三支决策概念外延发生变化, 由此分为 4 种情况来说明区间三支决策空间更新问题。

参数由 $[\alpha_0, \beta_0]$ 变为 $[\alpha_1, \beta_1]$ 时会有 4 种情况: 1) $\alpha_1 > \alpha_0$; 2) $\alpha_1 < \alpha_0$; 3) $\beta_1 > \beta_0$; 4) $\beta_1 < \beta_0$ 。前两种变化要更新区间三支决策概念的上界外延 M^0 到 M^1 。后两者要更新区间三支决策概念的下界外延 M^0 到 M^1 。为此, 首先给出 4 个函数, 分别实现这 4 个参数变化时区间三支决策概念的更新。

(1) 函数: $DCL1(\tilde{C}, \alpha_0, \alpha_1) // \tilde{C}$ 是区间三支决策概念格中的任一结点, α_0, α_1 为调整前后的参数, $\alpha_1 > \alpha_0$

$DCL1(\tilde{C}, \alpha_0, \alpha_1)$

{

$Ma = \{\emptyset\}$

对 $\tilde{C}$ 的上界外延 M^0 中的每个对象 x :

If $\frac{|f(x) \cap Y|}{|Y|} \geq \alpha_1$ then

$Ma = Ma \cup x$

$M^{*1} = Ma$

相应的决策损失函数 $R_{\alpha_0}^{\beta_0}(a_B), R_{\alpha_0}^{\beta_0}(a_N)$ 变为 $R_{\alpha_1}^{\beta_0}(a_B), R_{\alpha_1}^{\beta_0}(a_N)$

}

(2) 函数: $DCL2(\tilde{C}, \alpha_0, \alpha_1) // \tilde{C}$ 是区间三支决策概念格中的任一结点, α_0, α_1 为调整前后的参数, $\alpha_1 < \alpha_0$

$DCL2(\tilde{C}, \alpha_0, \alpha_1)$

{

$Ma = M^0$

对 $\tilde{C}$ 的每个父结点 CF 的上界外延 Maf :

{

令 $maf1 = Maf - M^0$

对 $\forall x \in maf1$:

If $\frac{|f(x) \cap Y|}{|Y|} \geq \alpha_1$ then // Y 是 $\tilde{C}$ 的内涵集合

$Ma = Ma \cup x$

}

$M^{*1} = ma$

相应的决策损失函数 $R_{\alpha_0}^{\beta_0}(a_B), R_{\alpha_0}^{\beta_0}(a_N)$ 变为 $R_{\alpha_1}^{\beta_0}(a_B), R_{\alpha_1}^{\beta_0}(a_N)$

}

(3) 函数: $DCL3(\tilde{C}, \beta_0, \beta_1) // \tilde{C}$ 是区间三支决策概念格中的任一结点, β_0, β_1 为调整前后的参数, $\beta_1 > \beta_0$

$DCL3(\tilde{C}, \beta_0, \beta_1)$

{

$Mb = \{\emptyset\}$

对 $\tilde{C}$ 的下界外延 M^0 中的每个对象 x :

If $\frac{|f(x) \cap Y|}{|Y|} \geq \beta_1$ then

$Mb = Mb \cup x$

$M^{\beta_1} = Mb$

相应的决策损失函数 $R_{\alpha_0}^{\beta_0}(a_P), R_{\alpha_0}^{\beta_0}(a_B)$ 变为 $R_{\alpha_0}^{\beta_1}(a_P), R_{\alpha_0}^{\beta_1}(a_B)$

}

(4) 函数: $DCL4(\tilde{C}, \beta_0, \beta_1) // \tilde{C}$ 是区间三支决策概念格中的任一结点, β_0, β_1 为调整前后的参数, $\beta_1 < \beta_0$

$DCL4(\tilde{C}, \beta_0, \beta_1)$

{

$Mb = M^0$

对 $\tilde{C}$ 的每个父结点 CF 的上界外延 Mbf :

{

令 $mbf1 = Mbf - M^0$

对 $\forall x \in mbf1$:

If $\frac{|f(x) \cap Y|}{|Y|} \geq \beta_1$ then // Y 是 $\tilde{C}$ 的内涵集合

$Mb = Mb \cup x$

}

$M^{\beta_1} = Mb$

相应的决策损失函数 $R_{\alpha_0}^{\beta_0}(a_P), R_{\alpha_0}^{\beta_0}(a_B)$ 变为 $R_{\alpha_0}^{\beta_1}(a_P), R_{\alpha_0}^{\beta_1}(a_B)$

}

基于这 4 个函数, 当参数发生变化时, 对区间三支决策空

间中的每个概念结点进行访问判断,根据不同情况对空间中的结点进行更新调整,对新产生的冗余概念和空概念进行判断并将其从三支决策空间中删除。

算法 2 基于参数变化的区间三支决策空间更新算法 SPDA

输入:决策形式背景 $(U, C \cup D, R)$, $\widetilde{L}_{\alpha_0}^{\beta_0}(U, C \cup D, R)$, 区间参数 (α_1, β_1)

输出: $\widetilde{L}_{\alpha_1}^{\beta_1}(U, C \cup D, R)$

Step1 $\widetilde{C}_1 = (M^{\alpha_0}, M^{\beta_0}, Y, R_{\alpha_0}^{\beta_0}(a_P), R_{\alpha_0}^{\beta_0}(a_B), R_{\alpha_0}^{\beta_0}(a_N))$ 为 $\widetilde{L}_{\alpha_0}^{\beta_0}(U, C \cup D, R)$ 的根结点, 若 $Y = \emptyset$, 则 $\widetilde{C}_1$ 不变; 若 $Y \neq \emptyset$, 则如果 $\alpha_1 \geq \alpha_0$, 则调用函数 $DCL1(\widetilde{C}, \alpha_0, \alpha_1)$, 否则调用函数 $DCL2(\widetilde{C}, \alpha_0, \alpha_1)$, 以此更新 M^{α_0} 为 M^{α_1} , $R_{\alpha_0}^{\beta_0}(a_B)$ 为 $R_{\alpha_1}^{\beta_0}(a_B)$, $R_{\alpha_0}^{\beta_0}(a_N)$ 为 $R_{\alpha_1}^{\beta_0}(a_N)$; 同样, 如果 $\beta_1 > \beta_0$, 则调用函数 $DCL3(\widetilde{C}, \beta_0, \beta_1)$, 否则调用函数 $DCL4(\widetilde{C}, \beta_0, \beta_1)$, 以此更新 M^{β_0} 为 M^{β_1} , $R_{\alpha_0}^{\beta_0}(a_P)$ 为 $R_{\alpha_0}^{\beta_1}(a_P)$, $R_{\alpha_0}^{\beta_0}(a_B)$ 为 $R_{\alpha_0}^{\beta_1}(a_B)$ 。则 $\widetilde{C}_1$ 更新为 $(M^{\alpha_1}, M^{\beta_1}, Y, R_{\alpha_1}^{\beta_1}(a_P), R_{\alpha_1}^{\beta_1}(a_B), R_{\alpha_1}^{\beta_1}(a_N))$ 。

Step2 访问 $\widetilde{C}_1$ 的每个子结点 $\widetilde{C}_i$ 。

Step3 设 $\widetilde{C}_i = (M_i^{\alpha_0}, M_i^{\beta_0}, Y_i, R_{\alpha_0}^{\beta_0}(a_P), R_{\alpha_0}^{\beta_0}(a_B), R_{\alpha_0}^{\beta_0}(a_N))$, 如果 $\alpha_1 > \alpha_0$, 则调用函数 $DCL1(\widetilde{C}, \alpha_0, \alpha_1)$, 否则调用函数 $DCL2(\widetilde{C}, \alpha_0, \alpha_1)$, 以此更新 $M_i^{\alpha_0}$ 为 $M_i^{\alpha_1}$, $R_{\alpha_0}^{\beta_0}(a_B)$ 为 $R_{\alpha_1}^{\beta_0}(a_B)$, $R_{\alpha_0}^{\beta_0}(a_N)$ 为 $R_{\alpha_1}^{\beta_0}(a_N)$; 同样, 如果 $M_i^{\alpha_1} = \emptyset$, 则删除结点 $\widetilde{C}_i$, 否则继续更新下界外延; 如果 $\beta_1 > \beta_0$, 则调用函数 $DCL3(\widetilde{C}, \beta_0, \beta_1)$, 否则调用函数 $DCL4(\widetilde{C}, \beta_0, \beta_1)$, 以此更新 $M_i^{\beta_0}$ 为 $M_i^{\beta_1}$, $R_{\alpha_0}^{\beta_0}(a_P)$ 为 $R_{\alpha_0}^{\beta_1}(a_P)$, $R_{\alpha_0}^{\beta_0}(a_B)$ 为 $R_{\alpha_0}^{\beta_1}(a_B)$, 则整个结点 $\widetilde{C}_i$ 更新为 $(M_i^{\alpha_1}, M_i^{\beta_1}, Y_i, R_{\alpha_1}^{\beta_1}(a_P), R_{\alpha_1}^{\beta_1}(a_B), R_{\alpha_1}^{\beta_1}(a_N))$ 。

Step4 对 $\widetilde{C}_i$ 的每个父节点 $\widetilde{C}_j' = \widetilde{C}_i \rightarrow \text{Parent}$, 且 $\widetilde{C}_j' = (M_j^{\alpha_1}, M_j^{\beta_1}, Y_j, R_{\alpha_1}^{\beta_1}(a_P), R_{\alpha_1}^{\beta_1}(a_B), R_{\alpha_1}^{\beta_1}(a_N))$, 若 $M_j^{\alpha_1} = M_j^{\alpha_1}$, $M_j^{\beta_1} = M_j^{\beta_1}$, $R_{\alpha_1}^{\beta_1}(a_P) = R_{\alpha_1}^{\beta_1}(a_P)$, $R_{\alpha_1}^{\beta_1}(a_B) = R_{\alpha_1}^{\beta_1}(a_B)$ 且 $R_{\alpha_1}^{\beta_1}(a_N) = R_{\alpha_1}^{\beta_1}(a_N)$, 则 $\widetilde{C}_i \rightarrow \text{Parent} = \widetilde{C}_j' \rightarrow \text{parent}$, 即将 $\widetilde{C}_i'$ 从格中删除。

Step5 对 $\widetilde{C}_i$ 的每个子结点, $\widetilde{C}_i = \widetilde{C}_i \rightarrow \text{Children}$, 转至 Step3 执行, 直到 $\widetilde{L}_{\alpha_0}^{\beta_0}(U, C \cup D, R)$ 中的最后一个结点为止。

Step6 输出新的概念格结构。

算法结束。

在原有区间三支决策空间的基础上,当参数发生改变时,空间中局部结点的外延和决策损失函数也发生相应的变化,更新算法通过遍历原有空间的每个结点,保留或更新结点外延和决策损失函数,获得新的区间三支决策空间。相比重建,更新算法在时间复杂性上要小得多。

3.3 三支决策空间的区间参数优化

(1) 基本思想

根据给定的决策形式背景,该算法通过改变区间参数 $[\alpha, \beta]$ 来得到各参数下的决策准则。在此依据决策形式背景中的属性个数 n 对区间参数 α 等步长划分,设步长 $\lambda = \frac{1}{n}$, 则 α_i 取值为 $\frac{i}{n} (i=1, 2, 3, \dots, n)$ 。由于已经对区间参数的取值进行了初步探究,发现当取得中间值时保证了格结构的稳定性,因此首先设定区间参数的初值为 $\alpha_0 = \frac{1}{2}, \beta_0 = 1$, 以此来构建区

间三支决策空间,然后进一步挖掘决策准则;当区间参数等步长变化时,在原有三支决策空间的基础上进行更新,相继得到新的概念、格结构以及决策准则;最终找到针对新对象 x 的最优决策准则和该决策下的最优区间参数。

(2) 算法设计

算法 3 三支决策空间的区间参数优化算法 SPOA

输入:决策形式背景 $(U, C \cup D, R)$, 新对象 x 及其条件属性集合 $A(A \subseteq C)$

输出:对象 x 的决策准则,最优决策下的区间参数

Step1 从决策形式背景中获得属性个数 n , 设置步长 $\lambda = \frac{1}{n}$;

Step2 初始化区间参数 $\alpha = \frac{1}{2}, \beta = 1$, 并构造区间三支决策空间

$\widetilde{L}_{\alpha}^{\beta}(U, C \cup D, R)$;

Step3 将 $\widetilde{L}_{\alpha}^{\beta}(U, C \cup D, R)$ 作为算法 1(GPOA)的输入;

Step4 若接受损失为 0, 拒绝损失为 0, 转 Step5;

Step5 令 $\alpha = \alpha + \lambda; \beta = \beta - \lambda$; 依据 3.2 节的 SPDA 更新算法进行三支决策格结构的更新;

Step6 重复 Step3, Step4;

Step7 比较针对 x 的多个决策准则, 输出最佳决策以及区间参数。

算法结束。

通过区间参数的等步长变化,不断更新原有的三支决策空间,每更新一次,针对新对象 x 的决策准则就会发生改变,但是这种决策准则的变动不是突然的,而是逐渐向做出接受和拒绝的决策靠拢,冒险者就可以从中做出决策而非采取不承诺。直到用户做出相对准确的或是符合情况的决策,此时区间参数值可以被认定为该决策形式背景下的最优参数。

4 案例分析

为了简便起见,固定参数 β 的值为 1, 只探讨参数 α 对决策准则的影响规律。下面以故事书推荐为例,给出描述故事书特征的一些属性,例如条件属性集合为 $\{a, b, c, d\}$, 其中属性 a 表示彩本有插图, b 表示厚读本, c 表示英文对照, d 表示童话;决策属性为 ef , 表示某类故事书可以推荐的读者为 e, f 或同时推荐给 ef 读者。决策形式背景如表 2 所列。预处理该形式背景,使决策属性与条件属性的表示法相统一,得到转化后的形式背景,如表 3 所列。

表 2 决策形式背景

对象	a	b	c	d	ef
1	1	0	1	1	f
2	0	1	0	0	ef
3	0	0	1	0	0
4	0	1	0	1	ef
5	1	1	1	0	ef
6	1	0	0	1	ef
7	1	0	1	1	ef
8	0	0	1	1	f
9	1	1	1	0	ef
10	0	1	0	1	f

表 3 转化后的形式背景

对象	a	b	c	d	e	f
1	1	0	1	1	0	1
2	0	1	0	0	1	1
3	0	0	1	0	0	0
4	0	1	0	1	1	1
5	1	1	1	0	1	1
6	1	0	0	1	1	1
7	1	0	1	1	1	1
8	0	0	1	1	0	1
9	1	1	1	0	1	1
10	0	1	0	1	0	1

4.1 模型验证

根据表 2 给出的形式背景(U, A, R),其中属性 A 的个数为 6,对象 U 的个数为 10,新对象 x 为彩本且有插图的故事书。设定 $\lambda_{PP}=0, \lambda_{BP}=9, \lambda_{NP}=15, \lambda_{NP}=17, \lambda_{BN}=2, \lambda_{NN}=0$ 。按照模型中的算法得到相应区间的三支决策概念,如表 4—表 7 所列。

表 4 $\alpha=\frac{3}{6}, \beta=1$ 时的三支决策概念

概念 $\tilde{C}$	上界外延 M^a	下界外延 M^b	内涵 Y	接受损失 $R(a_P)$	不承诺损失 $R(a_B)$	拒绝损失 $R(a_N)$
$\tilde{C}_1$	{12345679}	{5679}	ae	0	5.5	0
$\tilde{C}_2$	U	{15679}	af	0	7.6	0
$\tilde{C}_3$	{245679}	{59}	abe	0	9	0
$\tilde{C}_4$	{135679}	{579}	ace	0	4.33	7.5
$\tilde{C}_5$	{1356789}	{1579}	acf	0	6.67	15
$\tilde{C}_6$	{1345679}	{67}	ade	0	6.2	5
$\tilde{C}_7$	{1345678910}	{167}	adf	0	7.67	15
$\tilde{C}_8$	{123456789}	{59}	abce	0	6	0
$\tilde{C}_9$	U	{59}	abcf	0	8.125	0
$\tilde{C}_{10}$	{1234567910}	$\emptyset$	abde	0	6.67	0
$\tilde{C}_{11}$	{1345678910}	{17}	acdf	0	8	15
$\tilde{C}_{12}$	{13456789}	$\emptyset$	abcde	0	6.375	7.5
$\tilde{C}_{13}$	{1345678910}	$\emptyset$	abcdf	0	8.22	15

当区间参数 $\alpha=\frac{3}{6}, \beta=1$ 时,根据三支决策概念 $\tilde{C}_1$ 和 $\tilde{C}_2$ 的决策损失值,可以看出采取不承诺时会有一定的损失,然而接受或拒绝决策都是没有损失的,这样的决策准则是没有参考意义的;但是增加新对象的考量属性 c 后,根据三支决策概念 $\tilde{C}_4$ 和 $\tilde{C}_5$ 的决策损失值,可以得出结论:若拒绝将新故事书 x 推荐给 e 读者,带来的损失值为 7.5;若拒绝将新故事书 x 推荐给 f 读者,带来的损失值为 15。因此应该将 x 推荐给 f 读者。同理,增加考量属性 d 后,也应将该故事书推荐给 f 读者。

表 5 $\alpha=\frac{4}{6}, \beta=1$ 时的三支决策概念

概念 $\tilde{C}$	上界外延 M^a	下界外延 M^b	内涵 Y	接受损失 $R(a_P)$	不承诺损失 $R(a_B)$	拒绝损失 $R(a_N)$
$\tilde{C}_1$	{12345679}	{5679}	ae	0	5.5	0
$\tilde{C}_2$	{15679}	{15679}	af	0	0	12
$\tilde{C}_3$	{245679}	{59}	abe	0	9	0
$\tilde{C}_4$	{124567910}	{58}	abf	0	9	7.5
$\tilde{C}_5$	{135679}	{579}	ace	0	4.33	7.5
$\tilde{C}_6$	{1356789}	{1579}	acf	0	6.67	15
$\tilde{C}_7$	{1345679}	{67}	ade	0	6.2	5
$\tilde{C}_8$	{1345678910}	{167}	adf	0	7.83	15
$\tilde{C}_9$	{579}	{5}	abce	0	5	6.43
$\tilde{C}_{10}$	{1579}	{59}	abcf	0	7.83	12.5
$\tilde{C}_{11}$	{14567910}	$\emptyset$	abdf	0	9	10
$\tilde{C}_{12}$	{135679}	{7}	acde	0	6.2	7.5
$\tilde{C}_{13}$	{1356789}	{17}	acdf	0	7.6	15
$\tilde{C}_{14}$	{579}	$\emptyset$	abcde	0	5	6.43
$\tilde{C}_{15}$	{1579}	$\emptyset$	abcdf	0	7.83	12.5

当区间参数 $\alpha=\frac{4}{6}, \beta=1$ 时,根据三支决策概念 $\tilde{C}_1$ 和 $\tilde{C}_2$ 的决策损失值,可以看出将 x 推荐给 e 读者或不推荐给 e 读者都不会造成损失,保持不承诺反而会带来一定的损失,这样的决策准则没有指导性意义;相反,将 x 推荐给 f 读者不会造成损失,拒绝推荐给 f 读者反而会造成一定的损失。综上,

当区间参数为 $[\frac{4}{6}, 1]$ 时,无需增加考量属性来推进决断,可由直接相关的三支决策概念得到指导性的决策准则,提高了决策的效率和准确度。

表 6 $\alpha=\frac{5}{6}, \beta=1$ 时的三支决策概念

概念 $\tilde{C}$	上界外延 M^a	下界外延 M^b	内涵 Y	接受损失 $R(a_P)$	不承诺损失 $R(a_B)$	拒绝损失 $R(a_N)$
$\tilde{C}_1$	{15679}	{15679}	af	0	0	12
$\tilde{C}_2$	{579}	{579}	ace	0	0	6.43
$\tilde{C}_3$	{1579}	{1579}	acf	0	0	12.5
$\tilde{C}_4$	{167}	{167}	adf	0	0	12.86

表 7 $\alpha=\frac{6}{6}, \beta=1$ 时的三支决策概念

概念 $\tilde{C}$	上界外延 M^a	下界外延 M^b	内涵 Y	接受损失 $R(a_P)$	不承诺损失 $R(a_B)$	拒绝损失 $R(a_N)$
$\tilde{C}_1$	{15679}	{15679}	af	0	0	12
$\tilde{C}_2$	{1579}	{1579}	acf	0	0	12.5
$\tilde{C}_3$	{167}	{167}	adf	0	0	12.86
$\tilde{C}_4$	{17}	{17}	acdf	0	0	13.125

随着区间参数 α 的增大,将 x 推荐给 f 读者的优势体现得越加明显,即接收(推荐给 f)的损失为 0,拒绝(拒绝推荐给 f)的损失为较大值,进一步验证了当区间参数 $\alpha=\frac{4}{6}, \beta=1$ 时做出的决策的正确性。

4.2 模型对比说明

为进一步说明该模型得到的结论,给出变参数下的各决策损失走势图。假设新对象 x 只有属性 a ,当 α 等步长变化时,对新对象 x 做出关于 e 决策(接受 e ,不承诺,拒绝 e)的走势,如图 1 所示。当 α 等步长变化时,对新对象 x 做出关于 f 决策(接受 f ,不承诺,拒绝 f)的走势,如图 2 所示。

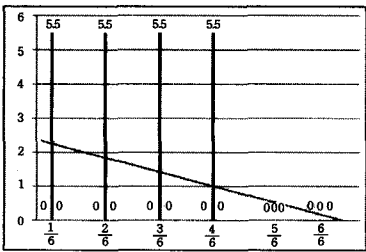


图 1 变 α 下 e 决策的走势

在 $[\frac{1}{6}, 1]$ 三支决策空间中查找内涵中条件属性为 $\{a\}$ 的区间三支决策概念,得到 $\tilde{C}_2$ 和 $\tilde{C}_3$,表明接受和拒绝决策 e 和 f 的损失均为 0,采取不承诺决策反而会带来决策损失,这样的决策准则显然没有什么实际意义,因此 $\alpha=\frac{1}{6}, \beta=1$ 的区间参数数值应当舍弃;等步长增加 α 的数值,得到相应三支决策概念下的接受、不承诺、拒绝损失。当区间参数为 $[\frac{2}{6}, 1]$ 时,查找内涵中条件属性为 $\{a\}$ 的区间三支决策概念,得到 $\tilde{C}_2$ 和 $\tilde{C}_3$,仍旧只有不承诺损失值。由于参数偏小,导致考虑到不确定信息的影响程度偏小,决策一般不会有明显的偏向。因此当 α 取得小于中间值的数值时,往往不能得到指导性的决策准则。

当 $\alpha=\frac{4}{6}$ 时,新对象 x 可以做出接受 f 的决策,因为此时

拒绝 f 决策的损失值为 12, 接受 f 的损失值为 0, 同时关于决策 e 仍然是保持不承诺, 所以当 $\alpha = \frac{4}{6}$ 时最快做出了决策而非不承诺。

当 α 继续增大时, 发现 C_1 中的不承诺损失值为 0, 拒绝损失值为 12, 显然应当采取接受决策 f , 并且内涵为 $\{ae\}$ 的三支决策概念不存在, 因此将故事书 x 推荐给读者 f 是比较明智的决策, 而做出该决策时的区间参数为 $[\frac{4}{6}, 1]$ 。

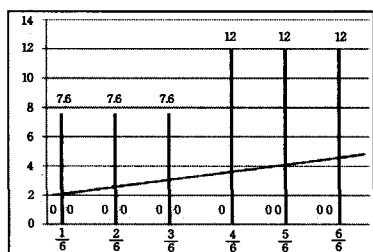


图2 变 α 下 f 决策的走势图

需要补充说明的是, 按照算法 1 增加对象 x 的考量属性 c 后, 当 $\alpha = \frac{3}{6}$ 时, 根据相关的三支决策概念就可以得出将故事书 x 推荐给 f 读者的决策。同样地, 继续增加对象 x 的考量属性 d , 使得新对象 x 具有属性 $\{a, c, d\}$, 当 $\alpha = \frac{2}{6}$ 时, 由相关的三支决策概念可以得出将故事书 x 推荐给 f 读者的决策。因此适当增加对象的考量属性也有助于提高决策的效率。

结束语 来源于区间概念格的参数 $[\alpha, \beta]$ 影响着由决策形式背景生成的概念以及格结构, 然后影响着最终的决策。文章立足于三支决策空间, 将决策损失函数值作为进行决策的准则, 通过改变区间参数得到不同的三支决策概念格, 基于此对新对象 x 及其条件属性集合进行决策。显然, 不同的区间参数下对同一新对象 x 的决策准则是不同的, 但是随着区间参数等步长变化, 决策准则却越加明朗, 直到做出接受或拒绝的决策, 由此便找到了最佳的区间参数, 其略大于中间值。这之前建立的区间概念格的参数优化模型^[16]得到的最优参数大致相同。二者虽然在解决区间参数问题上的立足点不同, 但是最终都得到了大致相同的结论, 为区间概念格在各领域应用中的区间参数选取问题提供了可靠依据。后续将从信息熵的角度继续进行区间参数优化问题的研究。

参考文献

- [1] YAO Y Y. An outline of a theory of three-way decision[C]// Proceedings of the 8th International RSCTC Conference. 2012: 1-17.
- [2] YAO Y Y, WONG S K M, LINGRAS P. A decision-theoretic rough set model[C]// The 5th International Symposium on Methodologies for Intelligent Systems. 1990.
- [3] YAO Y Y, WONG S K. A decision-theoretic framework for approximating concepts[J]. International Journal of Man-Machine Studies, 1992, 37(6): 793-809.
- [4] YAO Y Y. Decision-theoretical rough set models [C]// Proceedings of Rough Sets Knowledge Technology(RSKT'07). 2007: 1-12.
- [5] 李华雄, 周献中, 李天瑞, 等. 决策粗糙集理论及其研究进展[M]. 北京: 科学出版社, 2011.
- [6] YAO J T, YAO Y Y, ZIARKO W. Probabilistic rough sets: approximations, decision-makings, and applications [J]. International Journal of Approximate Reasoning, 2008, 49: 253-254.
- [7] YAO Y Y. Probabilistic rough sets approximation[J]. International Journal of Approximate Reasoning, 2008, 49: 255-271.
- [8] 贾修一, 商琳, 周献中. 三支决策理论与应用[M]. 南京: 南京大学出版社, 2010.
- [9] LIU B X, ZHANG C Y. A new concept lattice structure: Interval concept lattice[J]. Computer Science, 2012, 39(8): 273-277. (in Chinese)
- [10] 刘保相, 张春英. 一种新的概念格结构——区间概念格[J]. 计算机科学, 2012, 39(8): 273-277.
- [11] YAN H C, WANG H F, LIU B X. The structure characteristics and application of interval concept lattice [J]. Microcomputer & Its Applications, 2014, 33(9): 98-100. (in Chinese)
- [12] 阎红灿, 王会芳, 刘保相. 区间概念格的结构特性与应用[J]. 微型机与应用, 2014, 33(9): 98-100.
- [13] LI H X, LIU D, ZHOU X Z. A survey on decision-theoretic rough set [J]. Journal of Chongqing University of Post and Telecommunications(Natural Science Edition), 2010, 22(5): 624-630. (in Chinese)
- [14] 李华雄, 刘盾, 周献中. 决策粗糙集模型研究综述[J]. 重庆邮电大学学报(自然科学版), 2010, 22(5): 624-630.
- [15] JIA X Y, LI W, SHANG L, et al. An adaptive learning parameters algorithm in three-way decision-theoretic rough set model [J]. Acta Electronica Sinica, 2011, 39(11): 2520-2525. (in Chinese)
- [16] 贾修一, 李伟, 商琳, 等. 一种自适应三支决策中决策阈值的算法[J]. 电子学报, 2011, 39(11): 2520-2525.
- [17] JIA X Y, SHANG L. A simulated annealing algorithm for learning thresholds in three-way decision-theoretic rough set model [J]. Journal of Chinese Computer System, 2013, 34(11): 2604-2606. (in Chinese)
- [18] 贾修一, 商琳. 一种求三支决策阈值的模拟退火算法[J]. 小型微型计算机系统, 2013, 34(11): 2604-2606.
- [19] CHEN G, LIU B Q, WU Y. The new algorithm of optimal thresholds for three-way decisions [J]. Computer Application, 2012, 32(8): 2212-2215. (in Chinese)
- [20] 陈刚, 刘秉权, 吴岩. 求三支决策最优阈值的新算法[J]. 计算机应用, 2012, 32(8): 2212-2215.
- [21] ZHANG C Y, WANG L Y. Incremental construction algorithm based on attribute power set for interval concept lattice[J]. Application Research of Computers, 2014(3): 731-734. (in Chinese)
- [22] 张春英, 王立亚. 基于属性集合幂集的区间概念格渐进式生成算法[J]. 计算机应用研究, 2014(3): 731-734.
- [23] LI M X, ZHANG C Y, WANG L Y, et al. Parameters Optimization and Interval Concept Lattice Update with Change of Parameters [J]. ICIC Express Letters, 2016, 10(2): 339-346.