

DOI: 10.11992/tis.201905048

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20190909.1712.008.html>

面向混合数据的多伴随三支决策

赵天娜^{1,2}, 苗夺谦^{1,2}, 米据生³, 张远健^{1,2}

(1. 同济大学电子与信息工程学院, 上海 201804; 2. 同济大学嵌入式系统与服务计算教育部重点实验室, 上海 201804; 3. 河北师范大学数学与信息科学学院, 河北 石家庄 050024)

摘要: 针对混合数据的知识表示和分类的问题, 在思考混合数据的有效表示时, 提出代价敏感多伴随模糊粗糙集模型, 在解决混合数据的分类问题上, 引入三支决策思想, 同时在多伴随模型基础上做了两点改进: 1) 提出贴近代价敏感多伴随模糊粗糙集模型特点的概率定义; 2) 借助双量化延迟代价目标函数的思想, 构造面向混合数据的新型三支决策模型。该模型具有如下特点: 1) 引入多个伴随对, 模拟了数值型属性和符号型属性之间异构互补的关系; 2) 定义多伴随算子, 充分表达了不同类型属性之间的偏好; 3) 结合模糊粗糙集, 克服了分类问题的不确定性; 4) 考虑获取不同类型属性的代价, 提高了应用到实际生活的可能性。最后用实例验证了此模型的有效性。

关键词: 混合数据; 模糊粗糙集; 三支决策; 多伴随; 代价敏感; 知识表示; 分类

中图分类号: TP391 **文献标志码:** A **文章编号:** 1673-4785(2019)06-1092-08

中文引用格式: 赵天娜, 苗夺谦, 米据生, 等. 面向混合数据的多伴随三支决策 [J]. 智能系统学报, 2019, 14(6): 1092-1099.

英文引用格式: ZHAO Tianna, MIAO Duoqian, MI Jusheng, et al. Multi-adjoint three-way decisions on heterogeneous data[J]. CAAI transactions on intelligent systems, 2019, 14(6): 1092-1099.

Multi-adjoint three-way decisions on heterogeneous data

ZHAO Tianna^{1,2}, MIAO Duoqian^{1,2}, MI Jusheng³, ZHANG Yuanjian^{1,2}

(1. College of Computer Science and Technology, Tongji University, Shanghai 201804, China; 2. Key Laboratory of Embedded System and Service Computing of Ministry of Education, Tongji University, Shanghai 201804, China; 3. College of Mathematics and Information Science, Hebei Normal University, Shijiazhuang 050024, China)

Abstract: Considering the problem of knowledge representation and classification relating to heterogeneous data, a cost-sensitive multi-adjoint fuzzy rough set model is proposed for the effective representation of heterogeneous data and in order to solve the classification problem of heterogeneous data, the idea of three-way decisions is introduced. Moreover, two improvements are made on the basis of the multi-adjoint model: 1) A revised probability definition is presented to approximately characterize the cost-sensitive fuzzy rough set model. 2) Based on the idea of the dual quantization delay cost objective function, a novel three-way decisions model is constructed for heterogeneous data. This model has the following characteristics: 1) Multiple adjoint pairs are introduced to simulate the relationship of heterogeneous complementarity between numerical attribute and categorical attribute. 2) The multi-adjoint operator is defined to fully express the preference among different attributes. 3) A fuzzy rough set is combined to overcome the uncertainty of the classification problem. 4) The cost of acquiring both numerical and categorical attributes is considered to improve the possibility of application to real life. The effectiveness of the model is verified in the heterogeneous dataset.

Keywords: heterogeneous data; fuzzy rough set; three-way decisions; multi-adjoint; cost-sensitive; knowledge representation; classification

收稿日期: 2019-05-24. 网络出版日期: 2019-09-10.

基金项目: 国家重点研发项目(213); 国家自然科学基金项目(61673301, 61573127, 61763031); 河北省自然科学基金项目(A2018210120); 公安部重大专项项目(20170004).

通信作者: 赵天娜. E-mail: 1810375@tongji.edu.cn.

混合数据^[1]是非结构化的、互补的、超高维的, 包含大量冗余信息, 研究如何有效表示, 特征选择和融合混合数据有重要的实际意义。尤其是

在数字化时代,多通道传感器广泛化产生,混合数据的分类任务普遍存在于实际生活中,如文本情感分类,医学影像分析更多复杂的语义以混合属性值形式表示。数值型和符号型数据的相互作用,导致概念往往具有不一致性和模糊性。如何利用混合数据的上述性质降低此类数据的不确定性是一个值得研究的课题。

模糊粗糙集^[2]是一种处理不确定性问题的有效数学工具,然而现有的模糊粗糙集模型不能处理混合数据分类问题。因此,研究模糊粗糙集扩展模型,并应用于混合属性分类具有重要的理论创新价值。Mi等^[3]用对偶三角模构造逻辑算子,发展了粗糙近似算子的构造方法。Feng等^[4]用概率方法研究并构造了信任函数与似然函数。目前,模糊粗糙近似算子的构造大多使用一对三角模与蕴含算子作为逻辑算子,没有考虑对象或者属性之间的差异性。用多伴随对构造模糊 Galois 连接,模糊粗糙近似算子,可以体现数值型数据和符号型属性之间的偏好。近几年来,Medina等^[5]研究了多伴随理论,并探究其在多伴随概念格和多伴随粗糙集等概念上的应用,但混合数据的表示方面需进一步深入研究。不同类型属性获取所需的代价问题,是现实生活中数据处理时必不可少要考虑的因素。Tan^[6]提出代价敏感学习,并将其应用到许多方面,如遗传算法^[7]、粗糙集^[8-9]等。国内学者也研究进展颇丰,如Fan等^[8]针对测试代价敏感决策系统,构造数据模型。Ju等^[10]提出了带有多粒度指标的代价敏感粗糙集模型。一系列学者都将获取数据的代价作为数据处理时的重要因素。受上述工作的启发,本文引入代价敏感多伴随模糊粗糙集表示数值型属性和符号型属性之间的异构互补关系,使得伴随对与不同类型属性之间产生关联;在利用多伴随算子表达不同类型属性之间偏好的同时,兼顾了获取不同类型属性值的代价。

Yao^[11-12]提出的三支决策是处理数据分类任务时基于人类认知的高效决策模式,已有许多专家将其与其他模型结合,大大提高了数据分类能力。Liang等^[13-14]研究了处理多阶段信息和双犹豫模糊信息的基于直觉模糊粗糙集的三支决策模型。Qiao^[15]将三支决策引入到推荐系统时,考虑了代价和用户的偏好,极大改进了推荐质量。徐健锋^[16]以混淆矩阵为工具,多目标优化三支决策阈值求解模型,同时发现了三支决策的代价与决策的条件概率之间存在非单调现象^[17]。混合数据分类模

型结合三支决策模型的研究已有成果,AL-Hmouz^[18]运用三支决策的思想做多模态生物识别系统的开发。赵天娜^[19]初步研究了多伴随直觉模糊粗糙集与三支决策模型的简单结合。鉴于模型与三支决策组合的优越性,本文将进一步深入地研究用三支决策模型处理混合数据分类问题更优化的方式,代价敏感多伴随模糊粗糙集模型赋予三支语义,并在决策代价最小化意义下优化混合数据三支分类结果,扩展三支决策理论的应用范围。

本文的贡献可从以下几个方面体现:混合数据的非结构化,互补以及大量冗余无关信息等特点造成数据的高度不确定性,因此数据分类不一致性和模糊性使得挖掘混合数据蕴含的内部信息有难度。找出数值型和符号型属性之间的各种关系,降低分类的不确定性是混合数据挖掘中亟待解决的问题。针对上述问题,本文用多个伴随对模拟不同类型属性之间的关系,多伴随算子可充分表达不同类型属性对对象之间的偏好,同时考虑了获取不同类型属性值的代价,提出代价敏感多伴随模糊粗糙集模型。给出充分体现此模型特点的概率公式,并优化了损失函数。改造三支决策模型,挖掘出不同类型属性的关系特点。针对混合数据的不确定性特点,提出此模型。此模型是数据特点驱动的,从数据本质出发,分析混合数据的特点,找出数值型属性和符号型属性之间的各种关系。

1 基本概念

1.1 多伴随模糊粗糙集理论

Cornelis等^[20]提出的多伴随模糊粗糙集理论研究了多伴随对与模糊粗糙集理论结合的基本定义和简单性质。

定义 1^[20] 给定偏序集 (P_1, \leq_1) 、 (P_2, \leq_2) 、 (P_3, \leq_3) 和映射 $\&: P_1 \times P_2 \rightarrow P_3$, $\swarrow: P_3 \times P_2 \rightarrow P_1$, $\nwarrow: P_3 \times P_1 \rightarrow P_2$, 称 $(\&, \swarrow, \nwarrow)$ 是一个伴随三元组, 如果满足:

- 1) 对于 $x \in P_1, y \in P_2, z \in P_3$, 都有 $x \leq_1 z \swarrow y \Leftrightarrow x \& y \leq_3 z \Leftrightarrow y \leq_2 z \nwarrow x$;
- 2) $\&$ 在 P_1 、 P_2 、 P_3 上是保序映射;
- 3) \swarrow, \nwarrow 在第 1 和第 2 论域上是保序映射。

定义 2^[20] 设 $(L_1, L_2, P, \&, \swarrow, \nwarrow)$ 是多伴随框架, 称 (A, U, R, τ) 为多伴随背景, 其中 A 是非空对象集, U 是非空属性集, $R: A \times U \rightarrow P$ 是 P -模糊关系, 满足自反性和对称性。 $\tau: A \times A \rightarrow \{1, 2, \dots, n\}$ 是将 $A \times A$ 中元素对映射到多伴随框架上一个特定的伴随三元组。

定义 3^[20] 在模糊决策信息系统 $(U, A \cup \{d\})$ 中, 给定属性子集 $B \subseteq A$, 和 P -模糊关系 $R_a: X \times X \rightarrow P, \forall a \in A$, 对于 $x, y \in U$, 模糊关系 $R_B: U \times U \rightarrow P$ 定义为:

$$R_B(x, y) = @(\phi_B^{xy}(a_1), \dots, \phi_B^{xy}(a_m)) \quad (1)$$

其中 $@: P^m \rightarrow P$ 是一个聚合算子, 对于 $a \in A$, 有

$$\phi_B^{xy}(a) = \begin{cases} R_a(x, y), & a \in B \\ T_P, & a \notin B \end{cases}$$

其中 T_P 是 P 的最大值。

定义 4^[20] 在多伴随面向属性框架 $(P, L, L, \&_1, \dots, \&_n)$ 下, 考虑形式背景 (U, U, R_B, τ) , 对于 $x, y \in U$, 定义似然算子 $\uparrow_\tau: L^U \rightarrow L^U$ 和必然算子 $\downarrow^\tau: L^U \rightarrow L^U$

$$g^{\uparrow_\tau}(x) = \sup\{R_B(x, y) \&_{\tau(x, y)} g(y) | y \in U\} \quad (2)$$

$$f^{\downarrow^\tau}(y) = \inf\{f(x) \frown_{\tau(x, y)} R_B(x, y) | x \in U\} \quad (3)$$

那么 g^{\uparrow_τ} 是 g 的上近似, f^{\downarrow^τ} 是 f 的下近似。其中 L^U 表示所有由 U 映射到 L 的函数的集合。

定义 5^[20] 对于模糊子集 $h \in L^U$, $(h^{\uparrow_\tau}, h^{\downarrow^\tau})$ 称为多伴随模糊粗糙集。

1.2 基于决策粗糙集的三支决策

决策粗糙集模型用状态集和行动集来描述决策过程。状态集 $\Omega = \{C, \neg C\}$ 中的 2 个状态分别表示某事件属于 C 和不属于 C , 行动集 $\mathcal{A} = \{a_P, a_B, a_N\}$ 分别表示接受某事件、延迟决策和拒绝某事件 3 种行动, 用 λ_{PP} 、 λ_{BP} 、 λ_{NP} ; λ_{PN} 、 λ_{BN} 、 λ_{NN} 表示采取不同的行动会产生不同的损失, 分别表示当 x 属于 C 或 x 不属于 C 时采取行动 $\mathcal{A} = \{a_P, a_B, a_N\}$ 的损失。采取 3 种行动的期望损失函数由如下式子表示:

$$R(a_P|[x]) = \lambda_{PP}P(C|[x]) + \lambda_{PN}P(\neg C|[x]) \quad (4)$$

$$R(a_B|[x]) = \lambda_{BP}P(C|[x]) + \lambda_{BN}P(\neg C|[x]) \quad (5)$$

$$R(a_N|[x]) = \lambda_{NP}P(C|[x]) + \lambda_{NN}P(\neg C|[x]) \quad (6)$$

式中: $P(C|[x]) = \frac{|C \cap [x]|}{|[x]|}$; $|\cdot|$ 是基数; $[x]$ 表示 x 关于 R 的等价类。

根据贝叶斯决策规则, 以期望损失函数最小为原则选择最佳行动方案, 决策规则如下:

(P) 若 $R(a_P|[x]) \leq R(a_B|[x])$ 且 $R(a_P|[x]) \leq R(a_N|[x])$, 则 $x \in POS(C)$;

(B) 若 $R(a_B|[x]) \leq R(a_P|[x])$ 且 $R(a_B|[x]) \leq R(a_N|[x])$, 则 $x \in BND(C)$;

(N) 若 $R(a_N|[x]) \leq R(a_P|[x])$ 且 $R(a_N|[x]) \leq R(a_B|[x])$, 则 $x \in NEG(C)$ 。

2 代价敏感多伴随模糊粗糙集

混合数据的数值型和符号型属性是由多通道

传感器产生的, 获取途径差异很大, 采集不同类型属性付出的代价均不相同。在现实生活中, 考虑经济因素, 研究人员希望用几个代价较小的属性, 提取同样有价值的信息, 因此不同类型的属性代价是必不可少要考虑的因素。本文希望用更少的代价获取更有价值的信息。

定义 6 设 $A^* = (U, A \cup D)$ 是混合数据决策系统, U 是非空论域, $A \cap D = \emptyset$, $A = A_S \cup A_R$, $A_S = \{a_1, a_2, \dots, a_i, \dots, a_m\}$ 和 $A_R = \{a_{m+1}, a_{m+2}, \dots, a_{m+i}, \dots, a_{m+n}\}$ 分别是符号和数值型条件属性的集合。获取对象 x 分别关于不同模态属性 A_S 和 A_R 的信息付出的代价为 $Cost_S$ 、 $Cost_R$ 。考虑模态属性代价得到的权重 w_i^C 定义如下:

1) 对于 $a_i \in A_S, i \in \{1, 2, \dots, m\}$, 每个属性的权重定义为:

$$w_S^C = \frac{1}{|A_S|} \frac{Cost_S}{Cost_S + Cost_R} \quad (7)$$

2) 对于 $a_i \in A_R, i \in \{m+1, m+2, \dots, m+n\}$, 每个属性的权重定义为:

$$w_R^C = \frac{1}{|A_R|} \frac{Cost_R}{Cost_S + Cost_R} \quad (8)$$

定义 7 设 $A^* = (U, A \cup D)$ 是混合数据决策系统, U 是非空论域, $A \cap D = \emptyset$, 对于 $a \in A$, a -模糊关系 $R_a: U \times U \rightarrow [0, 1]$ 定义如下:

$$R_a(x, y) = 1 - |\bar{a}(x) - \bar{a}(y)|$$

$\forall B \subseteq A$, B -模糊关系 R_B^C 定义如下:

$$R_B^C(x, y) = w_S^C \phi_B^{xy}(a_1) + \dots + w_S^C \phi_B^{xy}(a_m) + w_R^C \phi_B^{xy}(a_{m+1}) + \dots + w_R^C \phi_B^{xy}(a_{m+n}) \quad (9)$$

其中 $\phi_B^{xy}: A \rightarrow [0, 1]$ 定义如下:

$$\phi_B^{xy}(a) = \begin{cases} R_a(x, y), & a \in B \\ 1, & a \notin B \end{cases}$$

性质 1 R_B^C 满足自反性和对称性。

证明 1) 由 R_B^C 的定义可知, $R_a(x, x) = 1$, 进而推出 $R_B^C(x, x) = 1$ 。因此, R_B^C 满足自反性。

2) 因为 $R_a(x, y) = 1 - |\bar{a}(x) - \bar{a}(y)|$, 并且 $R_a(y, x) = 1 - |\bar{a}(y) - \bar{a}(x)|$, 所以 $R_a(x, y) = R_a(y, x)$ 。由 R_B^C 的定义可知, $R_B^C(x, y) = R_B^C(y, x)$ 。因此, R_B^C 满足对称性。

性质 2 R_B^C 满足单调性, 即 $B_1 \subseteq B_2 \Rightarrow R_{B_1}^C \leq R_{B_2}^C$ 。

证明 1) 若 $B_1 = B_2$, 则显然 $R_{B_1}^C = R_{B_2}^C$ 。

2) 若 $B_1 \subset B_2$, 则必存在 $a_i \in B_2$, 但是 $a_i \notin B_1$ 。由 $\phi_{B_1}^{(x, y)}(a_i) = 1, \phi_{B_2}^{(x, y)}(a_i) = R_{a_i}(x, y)$, 则 $\phi_{B_2}^{(x, y)}(a_i) \leq \phi_{B_1}^{(x, y)}(a_i)$ 。对于其他属性 $a_j \in B_1 \cap B_2 = B_1$, 有 $\phi_{B_2}^{(x, y)}(a_j) = \phi_{B_1}^{(x, y)}(a_j)$ 成立。由 R_B^C 的定义可知, $R_{B_2}^C \leq R_{B_1}^C$ 。

定义 8 在多伴随面向属性框架 $(P, L, L, \&_1, \dots, \&_n)$ 下, 考虑有代价的形式背景 (U, U, R_B^C, τ) , 对于

$m \in L^U$, 定义似然算子 $\uparrow_\pi: L^U \rightarrow L^U$ 和必然算子 $\downarrow_N: L^U \rightarrow L^U$: 对于 $x, y \in L$

$$m_C^{\uparrow}(x) = \sup\{R_B^C(x, y) \&_{\tau(x, y)} m(y) | y \in U\} \quad (10)$$

$$m_C^{\downarrow}(y) = \inf\{m(x) \frown_{\tau(x, y)} R_B^C(x, y) | x \in U\} \quad (11)$$

其中, m_C^{\uparrow} 是 m 的上近似, m_C^{\downarrow} 是 m 的下近似。

对于模糊子集 $m \in L^X$, 称 $(m_C^{\downarrow}, m_C^{\uparrow})$ 为代价敏感多伴随模糊粗糙集。

经过对代价敏感多伴随模糊粗糙集的研究, 本文发现, 添加一个简单条件后, 它仍满足经典的粗糙集中上下近似的性质。由性质 1 和性质 2 可知, 本文定义的关系 R_B^C 满足文献 [13] 中关系 R_B 的性质, 因此, 本文中代价敏感多伴随模糊粗糙集的上下近似需添加的条件和文献 [7] 中 3.2 中命题 1 的条件一致。

性质 3 对于 $m \in L^U$,

1) 若满足 $m(x) \frown_{\tau(x, x)} 1 \leq m(x)$, 则 $m_C^{\downarrow}(y) \leq m(y)$ 。

2) 若满足 $m(x) \leq 1 \&_{\tau(x, x)} m(x)$, 则 $m(y) \leq m_C^{\uparrow}(y)$ 。

证明过程可参照文献 [12] 中 3.2 中命题 1 的证明。

3 基于代价敏感多伴随模糊粗糙集的新型三支决策模型

基于决策粗糙集的经典三支决策模型利用状态集 $\Omega = \{C, \neg C\}$ 和行动集 $\mathcal{A} = \{a_P, a_B, a_N\}$ 描述决策过程。采取不同的行动产生不同的损失函数矩阵如表 1 所示, 这些损失函数满足 $\lambda_{PP} < \lambda_{BP} < \lambda_{NP}$; $\lambda_{NN} < \lambda_{BN} < \lambda_{PN}$, 其中 $\lambda_{..} \in (0, 1)$ 。

表 1 经典模糊损失函数矩阵

Table 1 Classical fuzzy loss function matrix

λ	C	$\neg C$
a_P	λ_{PP}	λ_{PN}
a_B	λ_{BP}	λ_{BN}
a_N	λ_{NP}	λ_{NN}

由经验可知, 对象属于状态集的概率直接影响决策过程。在三支决策模型中, 体现为概率是定义决策目标函数(期望损失函数)的关键因素。因此, 给出体现模型特点的概率定义对精确决策至关重要。

在经典的三支决策模型中, 代价与决策的条件概率是单调的线性关系, 如图 1, 但在实际生活中, 三支决策的代价与决策的条件概率之间存在着非单调情况^[17], 如图 2, 针对这种现象, 徐健锋等^[17]以延迟代价目标函数为研究对象, 提出了一

种双延迟代价目标函数的策略。本文借鉴这一思想, 在代价敏感多伴随模糊粗糙集模型上, 给出充分贴近此模型特点的概率定义, 并优化做延迟决策 a_B 时的非线性损失函数, 期望得到更准确处理分类任务的三支决策模型。

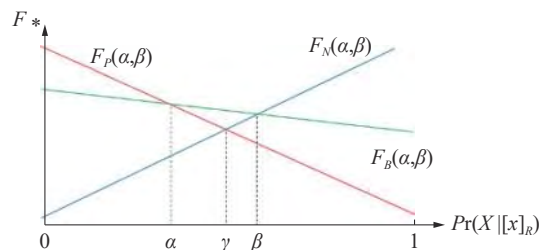


图 1 经典三支决策 F_N 、 F_P 、 F_B 的线性关系

Fig. 1 Classical linear relation diagram of three decisions

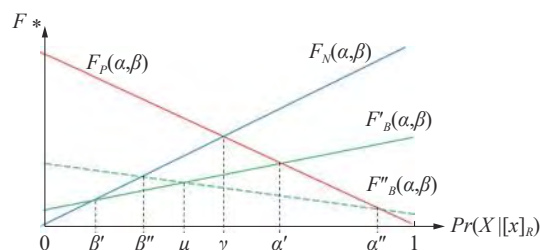


图 2 新型三支决策 F_N 、 F_P 、 F_B 、 $F_{B'}$ 的非线性关系

Fig. 2 Novel linear relation diagram of three decisions

3.1 改进的三支决策模型

改进的三支决策模型, 与传统三支决策模型相比, λ_{PP} 、 λ_{PN} 、 λ_{NP} 、 λ_{NN} 定义不变, 只改进 λ'_{BP} 、 λ'_{BN} 和 λ''_{BP} 、 λ''_{BN} , 如表 2, 它们的含义为: 分别从正域和负域视角出发, 实际为正域元素和负域元素但决策时被划分为延迟域的 2 组代价函数。正域决策代价目标函数 F_P 、负域决策代价目标函数 F_N 不变, 只改进双延迟决策代价目标函数, 记为 F'_B 、 F''_B 。由于非线性函数 F_B 有很多种定义方式, 不同的定义方式导致不同的非线性函数 F_B , 为了简化 and 一般性, 本文选取最简单的非线性模型进行说明, 由这 2 个线性函数组成一个非线性函数, 并假设 $\lambda_{NP} - \lambda_{NN} > 0 > \lambda_{PP} - \lambda_{PN}$ 和 $\lambda_{NP} - \lambda_{NN} > \lambda'_{BP} - \lambda'_{BN} > \lambda''_{BP} - \lambda''_{BN} > \lambda_{PP} - \lambda_{PN}$, 本文只讨论图 2 这种非线性的情况。其他情况类似, 在此不再赘述。

表 2 新型模糊损失函数矩阵

Table 2 New fuzzy loss function matrix

λ	C	$\neg C$
a_P	λ_{PP}	λ_{PN}
a_B	λ'_{BP}	λ'_{BN}
	λ''_{BP}	λ''_{BN}
a_N	λ_{NP}	λ_{NN}

$$F_P = \lambda_{PP} \Pr(X|[x]_R) + \lambda_{PN} (1 - \Pr(X|[x]_R))$$

$$F_N = \lambda_{NP} \Pr(X|[x]_R) + \lambda_{NN} (1 - \Pr(X|[x]_R))$$

$$F'_B = \lambda'_{BP} \Pr(X|[x]_R) + \lambda'_{BN} (1 - \Pr(X|[x]_R))$$

$$F''_B = \lambda''_{BP} \Pr(X|[x]_R) + \lambda''_{BN} (1 - \Pr(X|[x]_R))$$

令 $F'_B = F''_B$, 解得 $\Pr(X|[x]_R) = \mu$ 。

$$\text{所以 } F_B = \begin{cases} F''_B, & \Pr(X|[x]_R) \geq \mu \\ F'_B, & \Pr(X|[x]_R) < \mu \end{cases}$$

多伴随模糊粗糙集充分考虑了数值型和符号型属性之间的互补性、关联性、大量无关冗余信息的特点,用此模型从混合数据中提取的蕴含信息高效体现对象属于状态集的可能性程度。本文用3种策略描述对象属于状态集的概率。

定义9 (平均概率) 对象 x 属于状态集 C 的平均概率定义为:

$$\bar{P}(C|x) = \frac{1}{2} (m^{\uparrow\pi}(x) + m^{\downarrow\pi}(x)) \quad (12)$$

定义10 对象 x 分别采取 $\mathcal{A} = \{a_P, a_B, a_N\}$ 3种行动的平均模糊期望损失函数计算公式为:

$$\overline{F(a_P|x)} = \lambda_{PP} \bar{P}(C|x) + \lambda_{PN} (1 - \bar{P}(C|x)) \quad (13)$$

$$\overline{F(a_B|x)} = \begin{cases} \lambda''_{BP} \bar{P}(C|x) + \lambda''_{BN} (1 - \bar{P}(C|x)), & \bar{P}(C|x) \geq \mu \\ \lambda'_{BP} \bar{P}(C|x) + \lambda'_{BN} (1 - \bar{P}(C|x)), & \bar{P}(C|x) < \mu \end{cases} \quad (14)$$

$$\overline{F(a_N|x)} = \lambda_{NP} \bar{P}(C|x) + \lambda_{NN} (1 - \bar{P}(C|x)) \quad (15)$$

比较 $\overline{F(a_P|x)}$ 、 $\overline{F(a_B|x)}$ 、 $\overline{F(a_N|x)}$ 的大小, 确定行动 $a_i (i \in \{P, B, N\})$ 使平均模糊期望损失函数 \bar{F} 最小。

定义11 (悲观概率) 对象 x 属于状态集 C 的悲观概率定义为:

$$P^P(C|x) = m^{\downarrow\pi}(x) \quad (16)$$

定义12 对象 x 分别采取 $\mathcal{A} = \{a_P, a_B, a_N\}$ 3种行动的悲观模糊期望损失函数计算公式为:

$$F(a_P|x)^P = \lambda_{PP} P^P(C|x) + \lambda_{PN} (1 - P^P(C|x)) \quad (17)$$

$$F(a_B|x)^P = \begin{cases} \lambda''_{BP} P^P(C|x) + \lambda''_{BN} (1 - P^P(C|x)), & P^P(C|x) \geq \mu \\ \lambda'_{BP} P^P(C|x) + \lambda'_{BN} (1 - P^P(C|x)), & P^P(C|x) < \mu \end{cases} \quad (18)$$

$$F(a_N|x)^P = \lambda_{NP} P^P(C|x) + \lambda_{NN} (1 - P^P(C|x)) \quad (19)$$

比较 $F(a_P|x)^P$ 、 $F(a_B|x)^P$ 、 $F(a_N|x)^P$ 的大小, 确定行动 $a_i (i \in \{P, B, N\})$ 使悲观模糊期望损失函数 F^P 最小。

定义13 (乐观概率)

$$P^O(C|x) = m^{\uparrow\pi}(x) \quad (20)$$

定义14 对象 x 采取 $\mathcal{A} = \{a_P, a_B, a_N\}$ 3种行动的乐观模糊期望损失函数计算公式为:

$$F(a_P|x)^O = \lambda_{PP} P^O(C|x) + \lambda_{PN} (1 - P^O(C|x)) \quad (21)$$

$$F(a_B|x)^O = \begin{cases} \lambda''_{BP} P^O(C|x) + \lambda''_{BN} (1 - P^O(C|x)), & P^O(C|x) \geq \mu \\ \lambda'_{BP} P^O(C|x) + \lambda'_{BN} (1 - P^O(C|x)), & P^O(C|x) < \mu \end{cases} \quad (22)$$

$$F(a_N|x)^O = \lambda_{NP} P^O(C|x) + \lambda_{NN} (1 - P^O(C|x)) \quad (23)$$

4 基于代价敏感多伴随模糊粗糙集的三支决策算法

本文提出的基于代价敏感多伴随模糊粗糙集的三支决策算法, 是处理具有偏好的模糊信息的不确定性决策理论, 以最小期望损失为目标, 得到损失最小的决策。

本文仅以乐观概率为例给出混合数据分类的基于代价敏感多伴随模糊粗糙集的三支决策算法。整个算法流程图, 如图3所示。

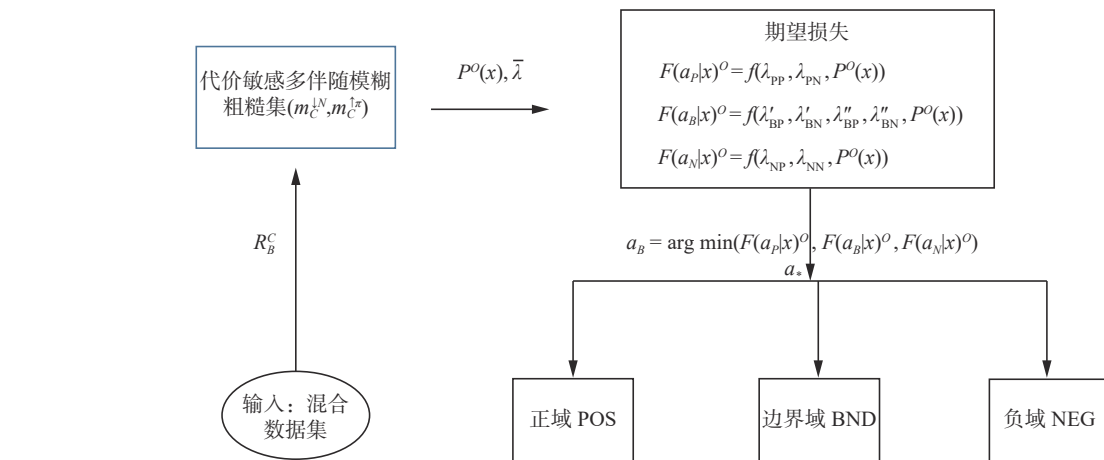


图3 三支决策算法的流程

Fig. 3 Flow chart of three decision algorithms

比较 $F(a_P|x)^O$ 、 $F(a_B|x)^O$ 、 $F(a_N|x)^O$ 的大小, 确定行动 $a_i (i \in \{P, B, N\})$ 使乐观模糊期望损失函数

F^O 最小。

算法1 基于代价敏感多伴随模糊粗糙集的

乐观三支决策算法

输入 混合数据决策表 $(U, A \cup D)$, 属性代价

Cost_S , Cost_R , 模糊集 m , 损失函数矩阵 λ 。

输出 行动集 \vec{a}_i , 分类精度 Acc 。

1) 根据定义 7, 计算模糊关系 R_B^C ;

2) For 每个对象 $x \in U$

根据定义 8, 计算代价敏感多伴随模糊粗糙集 $(m_C^{\downarrow N}, m_C^{\uparrow \pi})$;

根据定义 13, 计算乐观概率 $P^O(C|x) = m^{\uparrow \pi}(x)$;

根据定义 14, 计算采取 3 种行动 $\mathcal{A} = \{a_P, a_B, a_N\}$ 的期望损失 $F_{a_P}, F_{a_B}, F_{a_N}$;

比较期望损失函数 $F_{a_i}, i \in \{P, B, N\}$, 选择期望损失 F_{a_i} 最小的行动集 a_i 作为最佳行动方案;

$$\text{计算分类精度 } \text{Acc} = 1 - \frac{\sum(|\vec{a} - \vec{d}|)}{|U|};$$

End For

3) Return 行动集 \vec{a}_i , 分类精度 Acc 。

5 实例验证

设混合数据的决策表 $(U, A \cup D)$ 如表 3, 获取符号型属性和数值型属性的代价比为 $\text{Cost}_S : \text{Cost}_R = 1$, 则由定义 6 得每个符号型属性和数值型属性的权重分别为 $w_S^C = \frac{1}{4}$ 、 $w_R^C = \frac{1}{2}$, 由定义 7 得到模糊关系 R_B^C 。如表 4 所示。

表 3 混合数据的决策表

Table 3 Decision table of heterogeneous data

U	a_1	a_2	a_3	d
x_1	1	0.25	1	1
x_2	0	0.63	0	0
x_3	0	0.13	0	0
x_4	1	0.35	1	1
x_5	0	0.36	1	1
x_6	1	1	1	1
x_7	1	0.43	0	0

设每个对象对应的 $m(x)$ 的值分别为: 0.76、0.17、0.15、0.76、0.53、1、0, 由定义 8, 利用 Godel t 模和相应的蕴含算子构造的伴随对 $(\&_G, \swarrow_G, \searrow_G)$, Product t 模和相应的蕴含算子构造的伴随对 $(\&_P, \swarrow_P, \searrow_P)$, 其中 $x \&_G y = \min\{x, y\}$; $z \searrow_G x = \begin{cases} 1, & x \leq z \\ z, & \text{else} \end{cases}$; $\swarrow_G = \searrow_G$; $x \&_P y = x * y$; $z \searrow_P x = \min\{1, z/x\}$; $\swarrow_P = \searrow_{P \circ}$ 。
 $\tau(x_i, x_j) = \begin{cases} G, & j = 1, 3, 5, 7, i \in |U| \\ P, & j = 2, 4, 6, i \in |U| \end{cases}$, 得到代价敏感多伴随模糊粗糙集为:

$$\begin{aligned} m_C^{\downarrow N}(x_1) &= 0.50, m_C^{\downarrow N}(x_2) = 0.17, m_C^{\downarrow N}(x_3) = 0.15, m_C^{\downarrow N}(x_4) = \\ &0.50, m_C^{\downarrow N}(x_5) = 0.50, m_C^{\downarrow N}(x_6) = 0.50, m_C^{\downarrow N}(x_7) = 0; m_C^{\uparrow \pi}(x_1) = \\ &0.76, m_C^{\uparrow \pi}(x_2) = 0.17, m_C^{\uparrow \pi}(x_3) = 0.15, m_C^{\uparrow \pi}(x_4) = 0.76 \\ &m_C^{\uparrow \pi}(x_5) = 0.53, m_C^{\uparrow \pi}(x_6) = 1, m_C^{\uparrow \pi}(x_7) = 0 \end{aligned}$$

由对象属于状态集的概率的定义 11, 得 $P(x_1) = 0.76$ 、 $P(x_2) = 0.17$ 、 $P(x_3) = 0.15$ 、 $P(x_4) = 0.76$ 、 $P(x_5) = 0.53$ 、 $P(x_6) = 1$ 、 $P(x_7) = 0$ 。

参照徐健锋^[16]中双延迟代价双量化三支决策模型第 4 章模型 1 的参数:

表 4 模糊关系 R_B^C Table 4 Fuzzy relation R_B^C

R_B^C	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	0.5	0.375	0.375	0.5	0.375	0.5	0.5
x_2	0.375	0.5	0.5	0.375	0.5	0.375	0.375
x_3	0.375	0.5	0.5	0.375	0.5	0.375	0.375
x_4	0.5	0.375	0.375	0.5	0.375	0.5	0.5
x_5	0.375	0.5	0.5	0.375	0.5	0.375	0.375
x_6	0.5	0.375	0.375	0.5	0.375	0.5	0.5
x_7	0.5	0.375	0.375	0.5	0.375	0.5	0.5

$$\lambda_{PP} = 0.00, \lambda_{PN} = 2.00, \lambda'_{BP} = 0.25, \lambda'_{BN} = 1.00$$

$$\lambda''_{BP} = 1.00, \lambda''_{BN} = 0.25, \lambda_{NP} = 2.00, \lambda_{NN} = 0.00$$

以代价风险函数最小为原则, 得到数据分类规则为:

$$a = \begin{cases} a_P, & P(x) < 0.36 \\ a_B, & 0.36 \leq P(x) \leq 0.65 \\ a_N, & P(x) < 0.65 \end{cases}$$

得到数据分类如表 5 所示, 由此可知分类精度为 85.7%。

表 5 是分别采用乐观概率, 悲观概率, 平均概率时, 对每个对象的采取决策的结果, 从结果可知, 依据 3 种概率得到的期望损失函数的判断结果是一致的。

表5 乐观、悲观、平均概率情况下数据分类结果

Table 5 Data classification results in positive, negative, average cases

U	x_1	x_2	x_3	x_4	x_5	x_6	x_7
乐观	a_P	a_N	a_N	a_P	a_B	a_P	a_N
悲观	a_B	a_N	a_N	a_B	a_B	a_B	a_N
平均	a_P	a_N	a_N	a_B	a_B	a_P	a_N

6 结束语

混合数据的分类问题是研究热点。本文从混合数据的模糊性和符号型、数据型属性的不一致性等特点出发,利用多伴随,模糊粗糙集给出不同类型属性之间的不同度量,考虑不同类型属性的代价,期望花费最小的代价得到最有价值的信息。在延迟代价双量化的基础上,研究了多种不同的模糊概率策略,构造了混合数据分类过程中的三支决策模型。该模型既能反应出数据本身的客观不确定性,也可以表达决策者主观上对不同类型属性的偏好及数值型和符号型属性数据获得难易程度的认识。未来可继续研究更多的模糊粗糙集的扩展模型,期望在超大规模数据集上刻画混合数据的精确表示,挖掘出更多数据内部蕴含的信息。

参考文献:

- [1] YANG Rong, WANG Zhenyuan, HENG P A, et al. Classification of heterogeneous fuzzy data by choquet integral with fuzzy-valued integrand[J]. *IEEE transactions on fuzzy systems*, 2007, 15(5): 931–942.
- [2] HU Qinghua, ZHANG Lingjun, ZHOU Yucan, et al. Large-scale multimodality attribute reduction with multi-kernel fuzzy rough sets[J]. *IEEE transactions on fuzzy systems*, 2018, 26(1): 226–238.
- [3] MI Jusheng, LEUNG Y, ZHAO Huiyin, et al. Generalized fuzzy rough sets determined by a triangular norm[J]. *Information sciences*, 2008, 178: 3203–3213.
- [4] FENG Tao, MI Jusheng, ZHANG Shaopu. Belief functions on general intuitionistic fuzzy information systems[J]. *Information sciences*, 2014, 271: 143–158.
- [5] MEDINA J, OJEDA-ACIEGO M, VALVERDE A, et al. Towards biresiduated multi-adjoint logic programming[M]//CONEJO R, URRETAVIZCAYA M, PÉREZ-DE-LA-CRUZ J L. Current Topics in Artificial Intelligence. Berlin, Heidelberg: Springer, 2004: 608–617.
- [6] TAN Ming. Cost-sensitive learning of classification knowledge and its applications in robotics[J]. *Machine learning*, 1993, 13(1): 7–33.
- [7] SUSMAGA R. Computation of minimal cost reducts[M]//RAŚ Z W, SKOWRON A. Foundations of Intelligent Systems. Berlin, Heidelberg: Springer, 1999: 448–456.
- [8] FAN Min, HE Huaping, QIAN Yuhua, et al. Test-cost-sensitive attribute reduction[J]. *Information sciences*, 2011, 181(22): 4928–4942.
- [9] TURNEY P D. Cost-sensitive classification: empirical evaluation of a hybrid genetic decision tree induction algorithm[J]. *Journal of artificial intelligence research*, 1995, 2(1): 369–409.
- [10] JU Hengrong, LI Huaxiong, YANG Xibei, et al. Cost-sensitive rough set: a multi-granulation approach[J]. *Knowledge-based systems*, 2017, 300: 137–153.
- [11] YAO Yiyu. Three-way decisions with probabilistic rough sets[J]. *Information sciences*, 2017, 180(3): 341–353.
- [12] YAO Y Y, WONG S K M. A decision theoretic framework for approximating concepts[J]. *International journal of man-machine studies*, 1992, 37(6): 793–809.
- [13] LIANG Decui, LIU Dun. Deriving three-way decisions from intuitionistic fuzzy decision-theoretic rough sets[J]. *Information sciences*, 2015, 300: 28–48.
- [14] LIANG Decui, XU Zeshui, LIU Dun. Three-way decisions based on decision-theoretic rough sets with dual hesitant fuzzy information[J]. *Information sciences*, 2017, 396: 127–143.
- [15] QIAO Junsheng, HU Baoqing. On transformations from semi-three-way decision spaces to three-way decision spaces based on triangular norms and triangular conorms[J]. *Information sciences*, 2018, 432: 22–51.
- [16] 徐健锋, 苗夺谦, 张远健. 基于混淆矩阵的多目标优化三支决策模型[J]. *模式识别与人工智能*, 2017, 30(9): 859–864.
- [17] 徐健锋, 苗夺谦, 张远健. 延迟代价双量化三支决策[J]. *小型微型计算机系统*, 2019, 40(2): 260–266.
- [18] XU Jianfeng, MIAO Duoqian, ZHANG Yuanjian. Three-way decisions model for multi-object optimization based on confusion matrix[J]. *Pattern recognition and artificial intelligence*, 2017, 30(9): 859–864.
- [19] 徐健锋, 苗夺谦, 张远健. 延迟代价双量化三支决策[J]. *小型微型计算机系统*, 2019, 40(2): 260–266.
- [20] XU Jianfeng, MIAO Duoqian, ZHANG Yuanjian. Double-quantification of cost function with deferment for three-way decisions[J]. *Journal of Chinese computer systems*, 2019, 40(2): 260–266.
- [21] AL-HMOUZ R, PEDRYCZ W, DAQROUQ K, et al. Development of multimodal biometric systems with three-way and fuzzy set-based decision mechanisms[J]. *International journal of fuzzy systems*, 2018, 20(1): 128–140.
- [22] 赵天娜, 米据生, 解滨, 等. 基于多伴随直觉模糊粗糙集的三支决策[J]. *南京大学学报(自然科学)*, 2017, 53(6): 1081–1090.
- [23] ZHAO Tianna, MI Jusheng, XIE Bin, et al. Three-way de-

cisions with multi-adjoint intuitionistic fuzzy rough sets[J]. *Journal of Nanjing University (Natural Science)*, 2017, 53(6): 1081–1090.

- [20] CORNELIS C, MEDINA J, VERBIEST N. Multi-adjoint fuzzy rough sets: definition, properties and attribute selection[J]. *International journal of approximate reasoning*, 2014, 55(1): 412–426.

作者简介:



赵天娜, 女, 1992 年生, 博士研究生, 主要研究方向为模糊粗糙集、多伴随理论、人工智能、机器学习。



苗夺谦, 男, 1964 年生, 教授, 博士生导师, 主要研究方向为人工智能、机器学习、大数据分析、粒度计算。主持完成国家自然科学基金项目 6 项, 在研项目有国家重点研发计划课题和公安部重点计划项目。荣获 CAAI 吴文俊人工智能自然科学奖二等奖、国家教学成果二等奖, 授权专利 12 项。出版教材和学术著作 10 部。发表学术论文 100 余篇。



米据生, 男, 1966 年生, 教授, 博士生导师, 主要研究方向为粗糙集、粒计算、概念格、数据挖掘与近似推理。主持国家自然科学基金项目 3 项, 教育部博士点基金项目 1 项。获得省级自然科学奖 3 项。发表学术论文 130 余篇。

International Conference on Robotics and Automation (ICRA 2020)

ICRA is the IEEE Robotics and Automation Society's flagship conference and the premier international forum for robotics researchers to present and discuss their work. The conference will include plenary sessions, contributed paper sessions, workshops and tutorial sessions, forums, videos, exhibitions, and robot challenges.

会议日期: May 31 to June 4, 2020

会议地点: Palais des Congrès de Paris, Paris, France

网站: <http://icra2020.org/>