

doi: 10.16104/j.issn.1673-1891.2023.04.011

双论域模糊概率三支决策模型及其应用

卓雪雪, 朱苍璐

(安徽三联学院计算机工程学院, 安徽 合肥 230000)

摘要:双论域信息系统下的模糊概率粗糙集是粗糙集理论的一个重要拓展模型,然而,该模型目前还未有三支决策方面的相关研究。针对这一问题,提出一种双论域信息系统下的模糊概率三支决策模型。首先在双论域模糊概率粗糙集基础上定义了决策动作和决策代价的概念;然后以贝叶斯决策最小化代价为基础,推导出了双论域模糊概率关系下决策区域的三元划分,即三支决策模型;最后提出了双论域信息系统的模糊概率三支决策分类算法。实验分析证明了所提出的三支决策分类算法在双论域信息系统上的有效性和优越性。

关键词:粗糙集;双论域信息系统;模糊概率;三支决策;分类

中图分类号:TP181 **文献标志码:**A **文章编号:**1673-1891(2023)04-0062-07

Two Universes Fuzzy Probability's Three-way Decision Model and Its Application

ZHUO Xuexue, ZHU Canglu

(School of Computer Engineering, Anhui Sanlian University, Hefei, Anhui 230000, Chian)

Abstract:The fuzzy probability rough set under the two universes information system is an important extension model of rough set theory. However, there is currently no research on the three-way decision aspects of this model. To address this issue, a fuzzy probability three-way decision model in the two universes information system is proposed in this paper. Firstly, the concepts of decision actions and decision costs are defined based on the two universes fuzzy probability rough set; then, based on minimizing the cost of Bayesian decision-making, a three-way partition of decision regions under the two universes fuzzy probability relation was derived, which is the three-way decision model; finally, a fuzzy probability three-way decision classification algorithm for the two universes information systems was proposed. The experiments has demonstrated the effectiveness and superiority of the proposed three-way decision classification algorithms on the two universes information systems.

Keywords:rough set; two universes information system; fuzzy probability; three-way decision; classification

0 引言

粗糙集理论^[1]是人工智能领域一种重要的数据分析工具,由Pawlak于1982年首次提出。作为一种新的建模在不完全知识的数学理论,它可以有效地处理复杂系统中数据不精确或信息不精确的问题,目前已成功应用于数据挖掘、智能控制等领域^[2-3]。

然而,目前粗糙集的大多数研究都是在同一论域的假设下进行的。但在现实中,可能存在2个或多个不同的论域情形,针对这一问题,学者们提出

了基于双论域信息系统的粗糙集模型,并受到了学界的广泛关注,对双论域信息系统的粗糙集进行不断推广和改进,如Ugur等^[4]提出了广义的双论域粗糙集模型;Sun等^[5]在双论域信息系统下建立了模糊相似关系,并提出了多粒度的模糊双论域模型;黄心宏等^[6-7]进一步推广,提出了直觉模糊关系的双论域粗糙集模型。三支决策模型是粗糙集理论的重要应用,一经提出便受到了学者们的广泛关注和研究,并成功应用于双论域信息系统,如Li等^[8]在双论域信息系统下提出了三支决策模型;Sun等^[9]在双论

收稿日期:2023-07-19

基金项目:安徽高校科学研究重点项目(2022AH051988、2022AH051994)。

作者简介:卓雪雪(1989—),女,安徽宿州人,讲师,硕士,主要研究方向:人工智能,e-mail:yuanhx456@163.com。

域概率粗糙集下提出了三支决策方法;Xu等^[10]提出了具有排序和参考元组的三支决策双论域模型;袁路妍等^[11]提出了双论域三支决策的增量式更新方法。

由于传统的双论域粗糙集模型在处理模糊含噪声数据方面的不足,Yang等^[12]提出了模糊概率双论域粗糙集模型,使得在实际环境下的应用性能更加优越,受到了研究人员的青睐;Hu等^[13]针对双论域信息系统的动态性,提出了增量式的模糊概率双论域粗糙集模型;刘丹等^[14]在Hu的基础上进行改进,提出了矩阵策略的增量式模糊概率双论域粗糙集;Yang等^[15]将模糊概率双论域粗糙集推广至多粒度空间,提出了多粒度模糊概率双论域粗糙集模型。

然而,当前已有的模糊概率双论域粗糙集没有关于三支决策方面的研究。由于三支决策模型在近似和决策方面的优越性,因此这限制了模糊概率双论域粗糙集的实用化推广。本文将提出模糊概率双论域粗糙集的三支决策模型,并提出一种分类算法的应用。首先介绍了模糊概率双论域粗糙集;其次在其基础上定义了双论域信息系统的决策动作和决策代价,构建了贝叶斯决策模型;再次以最小化决策代价为原理,推导出了双论域信息系统模糊概率关系下决策区域的三元划分,即三支决策模型;最后利用三支决策对论域的区域划分原理,提出了双论域信息系统的模糊概率三支决策分类算法。进行双论域信息系统的实验证明了所提出分类算法的有效性和优越性。

1 基本理论

在粗糙集理论^[1-3]中,信息系统 S 表示为 $S = (U, A)$,其中 U 为全体对象集, A 为全体属性集,属性子集 $I \subseteq A$ 确定的等价关系为 E_A ,对于论域中的对象 $x \in U$ 在 E_A 诱导确定的等价类表示为 $[x]_{E_A}$,考虑近似对象集 $X \subseteq U$,在等价关系 E_A 下确定的上近似 $\bar{E}_A(X)$ 和下近似 $\underline{E}_A(X)$ 分别定义如式(1)和(2)所示。

$$\bar{E}_A(X) = \{x \in U \mid [x]_{E_A} \cap X \neq \emptyset\} \quad (1)$$

$$\underline{E}_A(X) = \{x \in U \mid [x]_{E_A} \subseteq X\} \quad (2)$$

称 $\bar{E}_A(X)$ 和 $\underline{E}_A(X)$ 为对象集 $X \subseteq U$ 在等价关系 E_A 下的粗糙集。同时,称 $P_A(X)$ 、 $B_A(X)$ 和 $N_A(X)$ 为

对象集 $X \subseteq U$ 在等价关系 E_A 下的正区域集、边界域集和负区域集。

若满足 $\underline{E}_A(X) = \bar{E}_A(X)$,也称对象集 $X \subseteq U$ 在等价关系 E_A 下是精确的,否则称之为粗糙的。

在某些实际应用环境,信息系统往往包含了2个论域,这类信息系统称之为双论域信息系统^[4],典型的如个信息化推荐信息系统。

定义1^[4] 设双论域信息系统 D 表示为 $D = (U, V)$,其中 U 和 V 分别为信息系统的2个论域, R 为论域 U 和 V 诱导确定的二元关系, $\forall x \in U$ 在二元关系 R 下确定的相似类定义为 $R(x) = \{y \in V \mid (x, y) \in R\}$ 。

定义2^[4] 双论域信息系统 $D = (U, V)$, R 为论域 U 和 V 诱导确定的二元关系。对象集 $Y \subseteq V$ 关于 R 的下近似集和上近似集分别定义式(3)和(4)所示。

$$\underline{R}(Y) = \{x \in U \mid R(x) \subseteq Y\} \quad (3)$$

$$\bar{R}(Y) = \{x \in U \mid R(x) \cap Y \neq \emptyset\} \quad (4)$$

称 $\bar{R}(Y)$ 和 $\underline{R}(Y)$ 为对象集 $Y \subseteq V$ 在二元关系 R 下的双论域粗糙集。

当 $U = V$,那么双论域粗糙集退化为传统的粗糙集,因此双论域粗糙集是传统粗糙集的进一步推广,传统粗糙集是双论域粗糙集的特例。

2 双论域模糊概率三支决策模型

近年来,学者们对双论域粗糙集模型提出了多种扩展模型,其中双论域模糊概率粗糙集是其研究热点^[12-15]。本节将双论域模糊概率粗糙集进一步扩展,提出一种双论域模糊概率三支决策模型。

定义3^[12] 双论域信息系统, F_λ 为论域 U 和 V 诱导确定的模糊二元关系, $F_\lambda: U \times V \rightarrow [0, 1]$,定义 $\forall x \in U$ 基于 F_λ 的相似类 $F_\lambda(x)$ 定义为

$$F_\lambda(x) = \{y \in V \mid F(x, y) \geq \lambda\},$$

式中: $F(x, y)$ 表示对象 x 与 y 之间的模糊相似度; λ 为模糊阈值,满足 $0 \leq \lambda \leq 1$,下文中在不引起混淆的情形,将省略 λ 标记。

定义4^[12] 双论域信息系统, F 为论域 U 和 V 诱导确定的模糊二元关系,对于 $0 \leq \lambda \leq 1$, $0 \leq \beta < \alpha \leq 1$,对于 $Y \subseteq V$ 关于模糊二元关系 F 的模糊概率下近似集 $\underline{F}_\alpha^\lambda(Y)$ 和模糊概率上近似集 $\bar{F}_\beta^\lambda(Y)$ 分别定义式(5)和(6)所示。

$$\underline{F}_\alpha^\lambda(Y) = \{x \in U \mid P(Y|F(x)) \geq \alpha, F(x) \neq \emptyset\} \quad (5)$$

$$\bar{F}_\beta^\lambda(Y) = \{x \in U | P(Y|F(x)) > \beta, F(x) \neq \emptyset\} \quad (6)$$

式中: $P(Y|F(x)) = \frac{|Y \cap F(x)|}{|F(x)|}$; $\underline{F}_\alpha^\lambda(Y)$ 和 $\bar{F}_\beta^\lambda(Y)$ 称为 $Y \subseteq V$ 关于模糊二元关系 F 基于参数 λ, α 和 β 的模糊概率粗糙集。

在定义 4 中, 参数 λ, α 和 β 可以被视为给定的最小阈值。对于实际应用环境下的决策, 这些参数通常是由决策者预先选择并代表着其需求。

在传统的分类决策模型中, 决策对象的判定结果为满足或不满足决策标准, 然后将对象集划分为 2 个不相交的决策区域, 即满足决策准则对象的正区域集和不满足决策准则对象的负区域集 N 。然而, 这种二元分类方法通常会存在一些分类错误。因此, 这种二元分类方法的主要局限是它们对分类的要求过于严格。

三支决策模型是由学者姚一豫基于概率粗糙集提出的一种新的决策模型。根据一组评估标准, 以三元分类的方式描述了该模型的基本思想^[8]。假设非空有限对象集 U 和非空有限属性集 A , 三支决策模型基于属性 A 将集合 U 划分为 3 个不相交的区域, 即正区域集, 负区域集 N 和边界域集, 分别代表了 3 种不同的分类决策结果。基于该思想, 接下来将提出基于双论域模糊概率粗糙集的三支决策模型。

设双论域信息系统, 对象子集族 $\Omega = \{Y_1, Y_2, \dots, Y_k\}$, 其中 $Y_k \subseteq V$, 代表了论域 V 中的 k 个状态。设 $Z = \{d_1, d_2, \dots, d_m\}$ 表示决策者的 m 种可能动作行为。 $P(Y_j|F(x))$ 是给定状态 Y_j 下特征描述集 $F(x)$ 与对象 $x(x \in U)$ 的条件概率。 $\vartheta(d_i|Y_j)$ 是在状态 Y_j 的情况下做出决策 d_i 的成本或代价。假设决策者对目标 $x(x \in U)$ 的特征描述集 $F(x)$ 采取动作行为 d_i , 则预期决策代价 $C(d_i|F(x))$ 为:

$$C(d_i|F(x)) = \sum_{j=1}^k \vartheta(d_i|Y_j) \times P(Y_j|F(x))。$$

因此, 如果存在 d_i , 使得 $C(d_i|F(x))$ 的决策代价最小, 则 d_i 为对象 $x(x \in U)$ 的最优决策, 为 x 的最小决策代价。

根据三支决策规则, 令 $Y = Y_j, Y^c = \bigcup_{i=1, i \neq j}^k Y_i, d_* \in Z, * = p, b, n$, 定义 d_p 为接受决策动作, d_b 为延迟决策动作, d_n 为拒绝决策动作。那么, 对于任意 $x \in U$, 可以得到双论域信息系统下执行 d_p, d_b 和 d_n

这 3 种决策动作 $C(d_p|F(x)), C(d_b|F(x))$ 和 $C(d_n|F(x))$ 的预期代价分别为:

$$C(d_p|F(x)) = \vartheta(d_p|Y) \times P(Y|F(x)) + \vartheta(d_p|Y^c) \times P(Y^c|F(x)),$$

$$C(d_b|F(x)) = \vartheta(d_b|Y) \times P(Y|F(x)) + \vartheta(d_b|Y^c) \times P(Y^c|F(x)),$$

$$C(d_n|F(x)) = \vartheta(d_n|Y) \times P(Y|F(x)) + \vartheta(d_n|Y^c) \times P(Y^c|F(x))。$$

基于贝叶斯决策过程会得到以下最小代价决策规则:

1) 执行决策 $P(Y)$, 当且仅当

$$C(d_p|F(x)) \leq C(d_b|F(x)), C(d_p|F(x)) \leq C(d_n|F(x));$$

2) 执行决策 $B(Y)$, 当且仅当

$$C(d_b|F(x)) \leq C(d_p|F(x)), C(d_b|F(x)) \leq C(d_n|F(x));$$

3) 执行决策 $N(Y)$, 当且仅当

$$C(d_n|F(x)) \leq C(d_p|F(x)), C(d_n|F(x)) \leq C(d_b|F(x))。$$

通常决策代价满足

$$\vartheta(d_p|Y) \leq \vartheta(d_b|Y) \leq \vartheta(d_n|Y),$$

$$\vartheta(d_n|Y^c) \leq \vartheta(d_b|Y^c) \leq \vartheta(d_p|Y^c)。$$

同时 $P(Y|F(x)) + P(Y^c|F(x)) = 1$, 因此

1) 对于执行决策 $P(Y)$, 即

$$\frac{\vartheta(d_p|Y) \times P(Y|F(x)) + \vartheta(d_p|Y^c) \times (1 - P(Y|F(x))) \Rightarrow P(Y|F(x)) \geq \frac{\vartheta(d_p|Y^c) - \vartheta(d_n|Y^c)}{(\vartheta(d_p|Y^c) - \vartheta(d_n|Y^c)) + (\vartheta(d_n|Y) - \vartheta(d_p|Y))};$$

2) 对于执行决策 $B(Y)$, 即

$$\frac{\vartheta(d_b|Y) \times P(Y|F(x)) + \vartheta(d_b|Y^c) \times (1 - P(Y|F(x))) \Rightarrow P(Y|F(x)) \leq \frac{\vartheta(d_p|Y^c) - \vartheta(d_b|Y^c)}{(\vartheta(d_p|Y^c) - \vartheta(d_b|Y^c)) + (\vartheta(d_b|Y) - \vartheta(d_p|Y))} \Rightarrow \frac{P(Y|F(x)) \geq \frac{\vartheta(d_b|Y^c) - \vartheta(d_n|Y^c)}{(\vartheta(d_b|Y^c) - \vartheta(d_n|Y^c)) + (\vartheta(d_n|Y) - \vartheta(d_b|Y))}。$$

3) 对于执行决策 $N(Y)$, 即

$$\vartheta(d_n|Y) \times P(Y|F(x)) + \vartheta(d_n|Y^c) \times$$

$$(1 - P(Y|F(x))) \Rightarrow P(Y|F(x)) \leq \frac{\vartheta(d_b|Y^c) - \vartheta(d_n|Y^c)}{(\vartheta(d_b|Y^c) - \vartheta(d_n|Y^c)) + (\vartheta(d_n|Y) - \vartheta(d_b|Y))} \Rightarrow P(Y|F(x)) \leq \frac{\vartheta(d_p|Y^c) - \vartheta(d_n|Y^c)}{(\vartheta(d_p|Y^c) - \vartheta(d_n|Y^c)) + (\vartheta(d_n|Y) - \vartheta(d_p|Y))}.$$

这里令

$$\alpha = \frac{\vartheta(d_p|Y^c) - \vartheta(d_b|Y^c)}{(\vartheta(d_p|Y^c) - \vartheta(d_b|Y^c)) + (\vartheta(d_b|Y) - \vartheta(d_p|Y))},$$

$$\beta = \frac{\vartheta(d_b|Y^c) - \vartheta(d_n|Y^c)}{(\vartheta(d_b|Y^c) - \vartheta(d_n|Y^c)) + (\vartheta(d_n|Y) - \vartheta(d_b|Y))},$$

$$\gamma = \frac{\vartheta(d_p|Y^c) - \vartheta(d_n|Y^c)}{(\vartheta(d_p|Y^c) - \vartheta(d_n|Y^c)) + (\vartheta(d_n|Y) - \vartheta(d_p|Y))},$$

那么:

- 1) 当 $P(Y|F(x)) \geq \alpha$ 且 $P(Y|F(x)) \geq \gamma$, 有 $x \in P(Y)$;
- 2) 当 $P(Y|F(x)) \leq \alpha$ 且 $P(Y|F(x)) \geq \beta$, 有 $x \in B(Y)$;
- 3) 当 $P(Y|F(x)) \leq \beta$ 且 $P(Y|F(x)) \leq \gamma$, 有 $x \in N(Y)$.

同时,若代价满足如下关系

$$\frac{\vartheta(d_n|Y) - \vartheta(d_b|Y)}{\vartheta(d_b|Y^c) - \vartheta(d_n|Y^c)} > \frac{\vartheta(d_b|Y) - \vartheta(d_p|Y)}{\vartheta(d_p|Y^c) - \vartheta(d_b|Y^c)},$$

此时有 $0 \leq \beta < \gamma < \alpha \leq 1$, 因此双论域信息系统最小代价决策规则可以简化表示为:

- 1) 当 $P(Y|F(x)) \geq \alpha$, 则 $x \in P(Y)$;
- 2) 当 $\beta < P(Y|F(x)) < \alpha$, 则 $x \in B(Y)$;
- 3) 当 $P(Y|F(x)) \leq \beta$, 则 $x \in N(Y)$.

因此基于双论域模糊概率关系的三支决策模型定义如下。

定义5 双论域信息系统, F 为论域 U 和 V 诱导确定的模糊二元关系, 对于 $0 \leq \lambda \leq 1$, 对于 $\forall x \in U$ 关于 $Y \subseteq V$ 的模糊概率关系的三支决策区域集分别定义为

$$P(Y) = \{x \in U | P(Y|F(x)) \geq \alpha\},$$

$$B(Y) = \{x \in U | \beta < P(Y|F(x)) < \alpha\},$$

$$N(Y) = \{x \in U | P(Y|F(x)) \leq \beta\}.$$

3 双论域模糊概率三支决策分类

学者们将三支决策模型应用于数据的分类, 提出了多种基于三支决策的分类算法^[16], 基于当前的可获取的分类准则, 它将目标对象的分类决策判定

为属于特定类、不属于特定类以及待定, 即当前条件可以确定分类的, 直接将分类结果确定为某个类, 对于那些处于类与类边缘的对象, 暂时进行待定状态, 待后续更多的分类信息被获取则进一步进行分类。受此启发, 本节将提出双论域信息系统下的三支决策分类模型, 具体如算法1所示。

算法1 双论域模糊概率三支决策分类算法。

输入: 双论域信息系统, 阈值 $0 \leq \lambda \leq 1$, 类别 $Y_i \subseteq V (1 \leq i \leq m)$ 的决策代价 $\vartheta(d_*|Y_i)$, $*$ = p, b, n ; 待分类目标对象 $x \in U$ 。

输出: $x \in U$ 的分类类别。

Step1: 计算双论域模糊概率关系 F_λ ;

Step2: 根据决策代价 $\vartheta(d_*|Y_i)$ 计算类别 $Y_i \subseteq V$ 的决策阈值 α_i 和 β_i ;

Step3: 计算 $x \in U$ 基于 F_λ 的相似类 $F_\lambda(x)$;

Step4: 对于每个决策类 $Y_i \subseteq V$, 如果 $P(Y_i|F_\lambda(x)) \geq \alpha$, 那么 $x \in P(Y_i)$;

Step5: 对于每个决策类 $Y_i \subseteq V$, 如果 $\beta < P(Y_i|F_\lambda(x)) < \alpha$, 那么 $x \in B(Y_i)$;

Step6: 对于每个决策类 $Y_i \subseteq V$, 如果 $P(Y_i|F_\lambda(x)) \leq \beta$, 那么 $x \in N(Y_i)$;

Step7: 对于所有 Y_i , 选择 $P(Y_i|F_\lambda(x))$ 最大值对应的 Y 作为最终决策结果, 其中 $Y_i \in \{Y_i \subseteq V | x \in P(Y_i)\}$;

Step8: 返回 $x \in U$ 的分类类别 Y 。

算法1主要计算量集中在双论域模糊概率关系计算流程, 因此整个算法1的时间复杂度为 $O(|U| \times |V|)$ 。

4 实验分析

本节将通过仿真实验验证所提出的双论域模糊概率三支决策分类算法的有效性。实验在 Windows 10 和 Intel (R) Core (TM) i3-7100 CPU 的个人主机上进行, CPU 频率为 3.90 GHz, 内存为 8 GB。实验使用的数据集如表1所示, 其中数据集1和2下载于 <http://groupLens.org/datasets/movielens/>, 数据集3下载于 <http://www.ieor.berkeley.edu/goldberg/jester-data/>, 数据集4下载于 UCI 数据集 <https://archive.ics.uci.edu/ml/index.php>。

为了消除数据集属性量纲带来的影响, 实验前将数据集进行归一化处理, 归一化方法为

表 1 实验数据集

数据集编号	数据集名称	论域 U	论域 V
数据集 1	MovieLens 1M	6 040	3 952
数据集 2	MovieLens 100k	943	1 682
数据集 3	Jester	73 421	100
数据集 4	Wdbc	569	30

$$a'(x) = \frac{a(x) - \min(a)}{\max(a) - \min(a)},$$

式中： $a(x)$ 表示对象 $x \in U$ 在论域 V 下属性 a 的属性值； $\min(a)$ 表示属性 a 下所有对象的最小值； $\max(a)$ 表示属性 a 下所有对象的最大值； $a'(x)$ 为归一化之后的值，满足 $0 \leq a'(x) \leq 1$ 。本实验中各个类别的分类代价 $\vartheta(d_n|Y_i)$ 采用在区间 $[0, 1]$ 进行随机选取的方式，并满足决策代价的基本大小关系。

针对本文提出的三支决策分类方法，本实验提出 4 种模型评估指标，具体如下定义：

$$A_c = \frac{N_{PP} + N_{NN}}{N_{PP} + N_{BP} + N_{NP} + N_{PN} + N_{BN} + N_{NN}},$$

$$P_r = \frac{N_{PP}}{N_{PP} + N_{PN}},$$

$$R_e = \frac{N_{PP}}{N_{PP} + N_{BP} + N_{NP}},$$

式中： A_c 、 P_r 、 R_e 分别表示准确度、精度和召回率； N_{PP} 、 N_{BP} 和 N_{NP} 分别表示原本属于类别 Y 采取动作 d_p 、 d_b 和 d_n 对象的数量； N_{PN} 、 N_{BN} 和 N_{NN} 分别表示原本不属于 Y 采取动作、和对象的数量。

理想情况下，当一个模型具有较高的精度和召回率时，它在分类方面会表现得更好。但是在本文所提出的分类算法中， $0 \leq \lambda \leq 1$ 是一个可变参数，它的取值不同对模型的性能有着很重要的影响，为了选择合适的实验参数，本实验将 λ 在 0 和 1 之间以 0.1 为步长分别进行取值，利用本文的三支决策分类算法进行分类计算，各个数据集的精度和的实验结果如图 1 和图 2 所示。

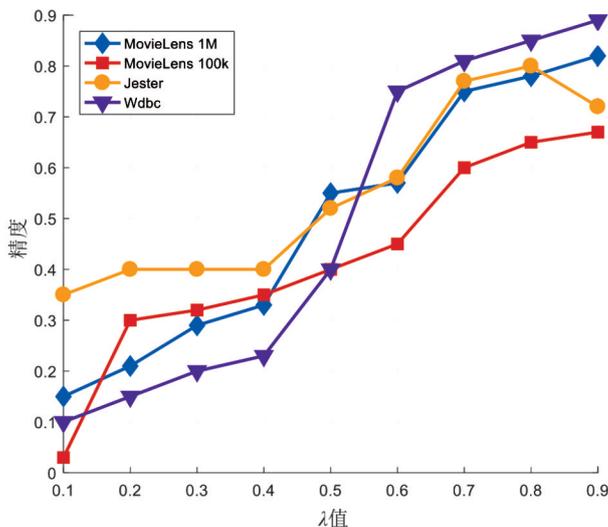


图 1 不同 λ 下精度实验结果

从图 1 和图 2 可以发现，随着 λ 取值的增加，模型分类的精度值也随之增加，但达到 0.7 之后开始趋于稳定。对于，随着 λ 取值的增加，数据集 MovieLens 1M、MovieLens 100k 和 Wdbc 的值逐渐增加，达到 0.7 之后开始趋于稳定，因此综合起来 $\lambda = 0.7$ 是一个较佳的取值。

为了验证本文所提出分类算法的优越性，本实验选择了机器学习领域 3 种常用的分类器进行分类比较，这 3 种分类器分别为支持向量机(SVM)、朴素贝叶斯(NB)和 k 近邻(kNN, $k=3$)。表 2 所示的是各

个数据集在三种对比分类器和本文分类算法下的精度实验结果，表 3 所示的是各个数据集在 3 种对比分类器和本文分类算法下的实验结果。

从表 2 可以发现，本文分类算法在大部分数据集下的分类结果精度值均高于其余 3 种分类算法。因此从分类精度视角可以看出本文分类算法的性能更优。

从表 3 可以发现，数据集 Jester 下本文算法的值低于其余分类算法，其余数据集下本文分类算法的结果最高。这主要是由于本文提出的三支决策分

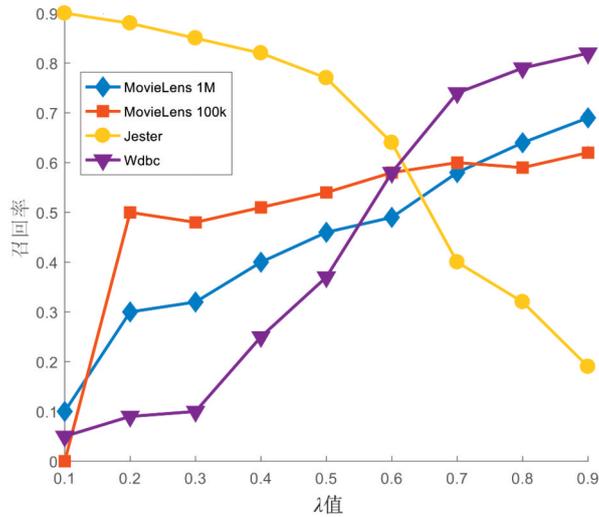


图 2 不同 λ 下实验结果

表 2 各个数据集精度实验结果

数据集名称	SVM	NB	kNN	本文算法
MovieLens 1M	0.782 4	0.804 5	0.790 3	0.815 6
MovieLens 100k	0.655 7	0.676 3	0.623 8	0.680 2
Jester	0.805 8	0.770 1	0.782 2	0.792 6
Wdbc	0.853 7	0.885 1	0.853 9	0.890 0

表 3 各个数据集实验结果

数据集名称	SVM	NB	kNN	本文算法
MovieLens 1M	0.646 1	0.669 2	0.634 9	0.689 9
MovieLens 100k	0.606 8	0.599 1	0.624 9	0.652 1
Jester	0.634 7	0.552 0	0.591 6	0.423 9
Wdbc	0.793 8	0.761 7	0.826 2	0.874 2

类算法引入了分类边界域,即对于不确定类别的对象进行延迟分类,待后续有充分的分类信息时进行进一步分类决策。

由于本文所提出的三支决策分类算法建立在代价的基础上,因此接下来计算各个分类算法的误分类代价结果,误分类代价即分类算法分类错误对象的代价之和,具体结果如表 4 所示。

从表 4 可以发现,各个数据集下本文分类算法的误分类代价是最低的,尤其对于数据集 Jester,其分类精度并不是最高的,但是误分类代价最低,这主要是由于对比的 3 种分类算法对每个对象进行了确定的分类,因此误分类产生的代价会比较高,而本文算法将暂时不能确定分类的对象进行延迟分

表 4 各个分类算法误分类代价

数据集	SVM	NB	kNN	本文算法
MovieLens 1M	1051.4	944.6	1 013.2	851
MovieLens 100k	259.7	244.1	283.8	211.2
Jester	11 406.6	13 503.5	12 792.8	11182
Wdbc	66.5	52.3	66.5	40.2

类,减少了错误分类的情形,因此也降低了误分类的代价,因此本文算法的误分类能力整体更优。

综合所有实验结果,证明了本文所提出的三支决策分类算法在双论域信息系统的分类方面具有较优的分类性能。

5 结束语

当前的双论域模糊概率粗糙集还未有三支决策的相关研究,针对这一问题,文中在模糊概率粗糙集基础上定义了双论域信息系统的决策动作和决策代价,以贝叶斯决策最小化代价为原理,推导出了双论域信息系统模糊概率关系下决策区域的三支决策模型,最后提出了双论域信息系统的模糊概率三支决策分类算法。实验分析证明了所提出分类算法的有效性和优越性。接下来将进一步对双论域信息系统模糊概率三支决策模型进行增量式更新方法的研究,提升现实动态数据处理的有效性。

参考文献:

- [1] PAWLAK Z. Rough sets[J]. *Internal Journal Information Science*, 1982, 11(5):341-356.
- [2] XIAO Jiayu, ZHANG Qinghua, AI Zhihua, et al. A fast neighborhood classifier based on hash bucket with application to medical diagnosis[J]. *International Journal of Approximate Reasoning*, 2022, 148:117-132.
- [3] JIN Chenxia, MI Jusheng, LI Fachao, et al. A novel probabilistic hesitant fuzzy rough set based multi-criteria decision-making method[J]. *Information Sciences* 2022, 608:489-516.
- [4] UGUR A A, DIKER M. Generalized textural rough sets: rough set models over two universes[J]. *Information Sciences*, 2020, 521:398-421.
- [5] SUN Bingzhen, MA Weimin, QIAN Yuhua. Multigranulation fuzzy rough set over two universes and its application to decision making[J]. *Knowledge-Based Systems*, 2017, 123(1):61-74
- [6] 黄心宏,张贤勇.基于双量化的双论域多粒度直觉模糊粗糙集[J].*四川师范大学学报(自然科学版)*,2023(1):125-133.
- [7] 黄心宏,张贤勇,杨霖琳.基于直觉模糊关系的双论域多粒度概率粗糙集[J].*模式识别与人工智能*,2022,35(5):439-450.
- [8] LI Xiaonan, SUN Qianqian, CHEN Hongmei, et al. Three-way decision on two universes [J]. *Information Sciences*, 2020, 515:263-279
- [9] SUN Bingzhen, CHEN Xiangtang, ZHANG Liye, et al. Three-way decision making approach to conflict analysis and resolution using probabilistic rough set over two universes[J]. *Information Sciences*, 2020, 507:809-822.
- [10] XU Wenyan, JIA Bing, LI Xiaonan. A two-universe model of three-way decision with ranking and reference tuple[J]. *Information Sciences*, 2021, 581:808-839.
- [11] 袁路妍,李锋.基于双论域信息系统的三支决策增量式更新算法[J].*计算机应用与软件*,2021,38(3):269-278+313.
- [12] YANG Hailong, LIAO Xiuwu, WANG Shouyang, et al. Fuzzy probabilistic rough set model on two universes and its applications[J]. *International Journal of Approximate Reasoning*, 2013, 54(9):1410-1420.
- [13] HU Jie, LI Tianrui, LUO Chuan, et al. Incremental fuzzy probabilistic rough sets over two universes[J]. *International Journal of Approximate Reasoning*, 2017, 81:28-48
- [14] 刘丹,李敬伟.基于矩阵的双论域模糊概率粗糙集增量更新算法[J].*控制与决策*,2021,36(3):553-564.
- [15] YANG Dan, CAI Mingjie, LI Qingguo, et al. Multigranulation fuzzy probabilistic rough set model on two universes[J]. *International Journal of Approximate Reasoning*, 2022, 145:18-35.
- [16] LI Weiwei, HUANG Zhiqiu, JIA Xiuyi, et al. Neighborhood based decision-theoretic rough set models[J]. *International Journal of Approximate Reasoning*, 2016, 69:1-17.