



# 计算机科学

COMPUTER SCIENCE

## 基于层次聚类的三支决策移动策略

徐怡, 骆帆, 王敏

### 引用本文

徐怡, 骆帆, 王敏. 基于层次聚类的三支决策移动策略[J]. 计算机科学, 2023, 50(6): 92-99.

XU Yi, LUO Fan, WANG Min. [Three-way Decision Movement Strategy Based on Hierarchical Clustering](#) [J]. Computer Science, 2023, 50(6): 92-99.

---

### 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

**Similar articles recommended (Please use Firefox or IE to view the article)**

#### [基于三支聚类的云任务优化调度](#)

Optimal Scheduling of Cloud Task Based on Three-way Clustering

计算机科学, 2022, 49(11A): 211100139-7. <https://doi.org/10.11896/jsjcx.211100139>

#### [一种基于层次聚类 and 模拟退火的选择性集成算法的风控模型研究](#)

Study on Risk Control Model of Selective Ensemble Algorithm Based on Hierarchical Clustering and Simulated Annealing

计算机科学, 2022, 49(11A): 210800105-7. <https://doi.org/10.11896/jsjcx.210800105>

#### [融合多层次信息的海关同义词识别方法](#)

Customs Synonym Recognition Fusing Multi-level Information

计算机科学, 2022, 49(11A): 210800197-5. <https://doi.org/10.11896/jsjcx.210800197>

#### [基于优化YOLO-V4的交通标志检测识别方法](#)

Traffic Sign Detection and Recognition Method Based on Optimized YOLO-V4

计算机科学, 2022, 49(11): 179-184. <https://doi.org/10.11896/jsjcx.220300251>

#### [一种基于正域的三支近似约简](#)

Three-way Approximate Reduction Based on Positive Region

计算机科学, 2022, 49(4): 168-173. <https://doi.org/10.11896/jsjcx.210500067>

# 基于层次聚类的三支决策移动策略

徐怡<sup>1,2</sup> 骆帆<sup>1</sup> 王敏<sup>1</sup>

1 安徽大学计算机科学与技术学院 合肥 230039

2 计算智能与信号处理教育部重点实验室(安徽大学) 合肥 230039

**摘要** 治略是三支决策 TAO 模型中的一个重要步骤,是实现对象移动的重要手段。通过实施策略,促使对象从不利区域移动到有利区域。近年来,对于治略方面的研究,学者们提出了两种移动策略,一种是基于区域的移动,另一种是基于对象的移动。然而,这两种移动策略都是从单层次的角度分析和制定移动策略,并未从多层次上考虑移动策略的制定。因此,为了制定多个层次上的移动策略,文中引入层次聚类,提出了一种基于层次聚类的三支决策移动策略模型。首先,使用层次聚类,将不利区域中的对象划分成不同的层次,每一层次上的聚类结果不同。然后,根据全局属性值频率最高准则,为每个层次中的簇制定一个移动策略,不同的簇有不同的移动策略。此外,文中还利用移动过程中产生的收益和代价,对不同层次上的移动策略进行评估。最后,实验结果证明了所提模型的有效性。

**关键词**: 三支决策; TAO 模型; 层次聚类; 移动策略; 多层次

**中图法分类号** TP301

## Three-way Decision Movement Strategy Based on Hierarchical Clustering

XU Yi<sup>1,2</sup>, LUO Fan<sup>1</sup> and WANG Min<sup>1</sup>

1 College of Computer Science and Technology, Anhui University, Hefei 230039, China

2 Key Laboratory of Intelligent Computing and Signal Processing, Ministry of Education, Anhui University, Hefei 230039, China

**Abstract** Acting is an important step in three-way decision TAO model, and it is also an important method to realize object movement. By implementing the strategies, the object is moved from a disadvantageous area to the advantageous area. In recent years, scholars have proposed two kinds of movement strategies, one is region-based movement, the other is object-based movement. However, the two movement strategies are analyzed and formulated from a single-level perspective, and the formulation of movement strategy is not considered from a multi-level perspective. Therefore, in order to make multi-level movement strategy, this paper introduces hierarchical clustering and proposes a three-way decision movement strategy based on hierarchical clustering. Firstly, it uses hierarchical clustering to divide the objects in the disadvantageous area into different levels, and the clustering results are different at each level. Then, according to the highest frequency of global attribute value criterion, a movement strategy is formulated for clusters in each hierarchy, and different clusters have different movement strategies. In addition, the paper also uses the benefit and cost of the movement process to evaluate the different levels of the movement strategies. Finally, experimental results prove the validity of the proposed model.

**Keywords** Three-way decision, TAO model, Hierarchical clustering, Movement strategies, Multi-level

## 1 引言

三支决策<sup>[1-2]</sup>是处理不确定问题的重要方法之一,它符合现实中人们对未知事物的认识与处理方法。其主要思想是三分而治,将一个整体区域划分成 3 个两两不相交的区域或者部分,并且根据实际的问题场景,对不同的部分分别采取不同的行动或者策略,根据对象在区域间的移动结果,来评价相应

策略的有效性。策略的结果反过来又可以进一步指导决策。近年来,有关三支决策的研究发展迅速,如三支分类<sup>[3-4]</sup>、三支聚类<sup>[5-7]</sup>、三支概念分析<sup>[8-10]</sup>、三支推荐系统<sup>[11-13]</sup>、三支属性约简<sup>[14-15]</sup>、三支前景理论<sup>[16-17]</sup>、序贯三支决策<sup>[18-19]</sup>、三支冲突分析<sup>[20-21]</sup>等。

目前,在三支决策的治略研究方面, Jiang 等<sup>[22]</sup>提出了一种基于区域的移动策略,针对整个不利区域,设计一种移动

到稿日期:2022-09-05 返修日期:2022-12-05

基金项目:国家自然科学基金(62076002);安徽省自然科学基金(2008085MF194)

This work was supported by the National Natural Science Foundation of China(62076002) and Natural Science Foundation of Anhui Province, China(2008085MF194).

通信作者:徐怡(xuyi1023@126.com)

策略,使得不利区域中的对象尽可能地移动到有利区域。该方式移动策略少,移动代价低,但移动成功的对象个数少。Gao 等<sup>[23]</sup>提出了一种基于对象的移动策略,针对不利区域中对象的特征,为每个对象设计一种移动策略。通过该方式移动成功的对象数目多,移动收益高,但移动策略多,移动代价高。

然而,以上两种移动策略都是从单层次的角度为不利区域中的对象分析和制定移动策略,并未考虑从不同层次上为不利区域中的对象制定移动策略。本文针对多层次的移动策略研究,引入层次聚类,提出了一种基于层次聚类的三支决策移动策略模型,可以为不利区域中的对象制定多层次的移动策略。首先,利用层次聚类,将不利区域中的对象划分成不同的层次,每一层次上的聚类簇数不同,得到的聚类结果也不同。然后,针对每个层次的聚类结果,为每个类簇设计一种移动策略,不同类簇的移动策略也不相同,类簇数越多,制定的移动策略就越多。本文根据全局属性值频率最高准则(the Highest Frequency of Global Attribute Value, HFGAV)来设计移动策略,选取类簇中每个属性频率最高的属性值,作为该类簇的参考对象的属性值。移动策略的制定需要在有利区域中随机选取一个对象作为目标对象,在目标对象与参考对象稳定属性值相同的条件下,根据灵活属性值制定相应的移动策略。对象在移动的过程中,无论是否移动成功都会产生收益和代价,因此,我们引入一个效用函数<sup>[22]</sup>,根据移动过程中产生的收益和代价来评估移动策略的有效性。

本文第 2 节介绍了三支决策的相关理论知识;第 3 节提出了一个基于层次聚类的三支决策移动策略模型;第 4 节通过实验验证了该模型的有效性;最后总结全文。

## 2 相关理论

三支决策包括 3 个阶段:分、治、效。“分”是将论域划分为 3 个互不相交的区域;“治”是对 3 个区域施加不同的策略来优化三分的结果;“效”则是衡量三分和治略后的结果是否达到预期的效果。其中,分是基础,治是手段,效是最终的目的。三者相互统一,缺一不可。

**定义 1** 给定一个决策表  $S$

$$S = \{U, AT = A_s \cup A_f \cup \{d\}, \{V_a | a \in AT\}, \{I_a | a \in AT\}\} \quad (1)$$

其中, $U$  是一个有限非空的对象集合, $AT$  是一个有限非空的属性集合, $A_s$  是稳定属性, $A_f$  是灵活属性, $d$  是决策属性, $V_a$  是属性  $a \in AT$  的值的集合, $I_a$  是一个映射函数, $I_a(x) = v$  表示对象  $x$  在属性  $a$  处的取值为  $v$ 。

**定义 2**(给定一个决策表  $S$ ) 假设  $A \subseteq A_s \cup A_f$  是一个属性子集,于属性集  $A$  有一个等价关系,由  $R_A$  表示,其中:

$$R_A = \{(x, y) \in U \times U | \forall a \in A, I_a(x) = I_a(y)\} \quad (2)$$

在粗糙集理论中,设  $U$  是一个有限非空的对象集合,即论域, $R$  是  $U$  上的等价关系, $R$  一般被称为不可分辨关系。关系  $R$  的等价类被称为  $S$  中的基本集。等价类可由以下方式来表示: $U/R_A = \{[x] | x \in U\}$ ,其中  $[x] = \{y | (x, y) \in R_A\}$  为

相对于  $A$  的等价类。

三支决策根据一对阈值  $(\alpha, \beta)$ , 将一个整体划分为 3 个两两不相交区域,即正域、负域和边界域。

**定义 3** 给定一个决策表  $S$  和一对阈值  $(\alpha, \beta)$ ,  $0 \leq \beta < \alpha \leq 1$ ,  $e(x)$  是集合  $U$  上的评估函数, $U$  可以被划分为 3 个两两不相交的区域。

$$\begin{aligned} P_1 &= \{x \in U | e(x) \geq \alpha\} \\ P_2 &= \{x \in U | \beta < e(x) < \alpha\} \\ P_3 &= \{x \in U | e(x) \leq \beta\} \end{aligned} \quad (3)$$

目前,在实际应用中已有多种评估函数,本文选择条件概率<sup>[24]</sup>作为评估函数。区域  $P_1$  由评估值大于或等于  $\alpha$  的对象组成,区域  $P_2$  由评估值大于  $\beta$  且小于  $\alpha$  的对象组成,区域  $P_3$  由评估值小于或等于  $\beta$  的对象组成。3 个区域满足以下两个条件:

$$\begin{aligned} (1) & P_1 \cup P_2 \cup P_3 = U \\ (2) & P_1 \cap P_2 = \emptyset, P_1 \cap P_3 = \emptyset, P_2 \cap P_3 = \emptyset \end{aligned}$$

将一个整体划分为 3 个区域后,针对 3 个区域的不同特点,制定相应的移动策略。对 3 个区域施加策略后,区域中的对象会发生转移,形成一个新的三分区。

**定义 4** 给定一个有限非空集合  $U$ ,  $\pi = \{P_1, P_2, P_3\}$  是  $U$  被划分成的 3 个区域。设  $O = \{o_1, o_2, \dots, o_n\}$  是一组策略集合,对三分区  $\pi = \{P_1, P_2, P_3\}$  施加一个策略  $o \in O$ , 区域中的对象在策略  $o$  的作用下发生转移,形成一个新的三分区  $\pi' = \{P_1', P_2', P_3'\}$ 。我们将 3 个区域在策略  $o$  下的转变记为:  $\pi \rightarrow_o \pi'$ 。

**定义 5** 给定一个决策表  $S$ , 集合  $U$  被划分为 3 个区域  $\pi = \{P_1, P_2, P_3\}$ , 设  $P_i$  表示有利区域,  $P_j (j \neq i)$  表示不利区域, 对于  $\forall x \in P_j$ , 随机选取  $P_i$  中的对象  $y$  作为目标对象, 则将  $P_j$  中的对象  $x$  移动到  $P_i$  中, 移动策略可以表示为:

$$\begin{aligned} r_x \rightarrow r_y: & [\bigwedge_{f \in A_f} I_f(x) \rightarrow I_f(y)] \Rightarrow I_d(x) \rightarrow I_d(y) \\ \text{s. t. } & [\bigwedge_{s \in A_s} I_s(x) = I_s(y)] \end{aligned} \quad (4)$$

其中,  $r_x \rightarrow r_y$  表示对象  $x$  到  $y$  的移动策略,  $\bigwedge_{f \in A_f} I_f(x) \rightarrow I_f(y)$  表示  $x$  的灵活属性值转变为  $y$  的灵活属性值,  $\bigwedge_{s \in A_s} I_s(x) = I_s(y)$  则表示  $x$  和  $y$  的所有稳定属性值都相同。

对不利区域施加移动策略,促使对象从不利区域移动到有利区域,在该移动过程中,无论移动是否可取,都会产生收益和代价。以下是收益和代价的计算方法。

**定义 6**<sup>[23]</sup> 给定一个决策表  $S$ ,  $\Omega = \{X, X^-\}$  分别表示对象  $x$  属于目标概念  $X$  和不属于目标概念  $X$ 。集合  $U$  被划分为 3 个区域  $\pi = \{P_1, P_2, P_3\}$ , 假设  $P_i$  表示有利区域,  $P_j (j \neq i)$  表示不利区域。对于  $P_j$  中的对象  $x$ , 我们想要将其移动到  $P_i$  中, 随机选取  $P_i$  中的  $y$  作为目标对象, 则移动过程中产生的收益为:

$$\begin{aligned} B_{r_x \rightarrow r_y} &= W_j [a\lambda_{j1} + (|[x]| - a)\lambda_{j3}] + W_i [-b\lambda_{i1} - \\ & (|[x]| - b)\lambda_{i3}] \end{aligned} \quad (5)$$

其中,  $W_i$  和  $W_j$  分别表示区域  $P_i$  和  $P_j$  的权重,  $\lambda_{i1}$  和  $\lambda_{j1}$  分别表示  $X$  中的对象移动到  $P_i$  和  $P_j$  的代价,  $\lambda_{i3}$  和  $\lambda_{j3}$  分别表示  $X^-$

中的对象移动到 $P_i$ 和 $P_j$ 的代价,它们的值由表1的代价矩阵给出。 $a$ 表示 $[x]$ 中的对象属于 $X$ 的数量, $b$ 表示变化后的 $[x]$ 中属于 $X'$ 的对象数量, $X'$ 表示变化后的目标概念。

$$a = |X \cap [x]|$$

$$b = |X' \cap [x]|$$

**定义 7**<sup>[23]</sup> 给定一个决策表 $S, \Omega = \{X, X^-\}$ 。集合 $U$ 被划分为3个区域 $\pi = \{P_1, P_2, P_3\}$ ,假设 $P_i$ 表示有利区域, $P_j (j \neq i)$ 表示不利区域。对于 $P_j$ 中的对象 $x$ ,我们想要将其移动到 $P_i$ 中,随机选取 $P_i$ 中的 $y$ 作为目标对象,则移动过程中产生的代价为:

$$C_{r_x \rightarrow r_y} = \sum_{f \in A_f} C_f(I_f(x), I_f(y)) \quad (6)$$

其中, $C_f(I_f(x), I_f(y))$ 表示将对象 $x$ 的灵活属性值设置为对象 $y$ 的灵活属性值所需要的代价, $C_{r_x \rightarrow r_y}$ 表示对象 $x$ 移动的总代价,即改变所有灵活属性值所需的代价总和。

Jiang等<sup>[22]</sup>提出了一种效用函数,根据移动过程中产生的收益和代价来评估移动策略是否有效。

表1 代价矩阵

Table 1 Cost matrix

	$P_1$	$P_2$	$P_3$
$X$	$\lambda_{11}$	$\lambda_{21}$	$\lambda_{31}$
$X^-$	$\lambda_{13}$	$\lambda_{23}$	$\lambda_{33}$

**定义 8**<sup>[22]</sup> 给定一个三分区 $\pi = \{P_1, P_2, P_3\}$ ,对不利区域施加策略,使得对象从不利区域移动到有利区域,形成一个新的三分区 $\pi' = \{P_1', P_2', P_3'\}$ 。移动过程中产生的移动效用为:

$$E(\pi \rightarrow \pi') = B - C \quad (7)$$

其中, $B$ 和 $C$ 分别表示对象移动产生的收益和代价。

根据对象的移动效用,可以得出3种判断:

(1)当 $E(\pi \rightarrow \pi') > 0$ ,代表移动的收益大于代价,即该移动策略是可取的。

(2)当 $E(\pi \rightarrow \pi') = 0$ ,代表移动的收益等于代价,即该移动策略对相应区域内的对象没有影响。

(3)当 $E(\pi \rightarrow \pi') < 0$ ,代表移动的收益小于代价,即该移动策略是不可取的。

### 3 基于层次聚类的三支决策移动策略模型

本节提出了一种基于层次聚类的三支决策移动策略模型,从多个层次上为不利区域中的对象制定移动策略。通过层次聚类,我们得到不同层次上的聚类结果,再进一步对每一层次上的不同类簇制定相应的移动策略。

#### 3.1 基于层次聚类的移动策略

本文采用层次聚类中的凝聚法,在最底层,每个等价类为一个类簇,通过计算类簇之间的欧氏距离,自底向上地合并距离最小的两个类簇,直到所有等价类合并为一个类簇。通过层次聚类,我们可以在不同层次上为不利区域中的对象制定移动策略。

**定义 9** 给定一个决策表 $S$ ,集合 $U$ 被划分为3个区域 $\pi = \{P_1, P_2, P_3\}$ ,设 $P_i$ 表示有利区域, $P_j (j \neq i)$ 表示不利

区域。假设 $P_j$ 中包含 $n$ 个等价类,利用层次聚类对 $P_j$ 中的等价类进行聚类分析,得到不同层次上的聚类结果,所得的层次划分为 $M = \{M_1, M_2, \dots, M_n\}$ 。在第 $t (t = 1, 2, \dots, n)$ 层上包含的类簇为 $M_t = \{M_t^1, M_t^2, \dots, M_t^l\}$ ,每一层次的类簇满足以下关系:

$$(1) M_t^l \cap M_t^k \cap \dots \cap M_t^1 = \emptyset$$

$$(2) M_t^l \cup M_t^k \cup \dots \cup M_t^1 = P_j$$

例1 给定一组对象 $U = \{x_1, x_2, x_3\}$ ,两个条件属性 $AT = \{a_1, a_2\}$ ,具体如表2所列,自底向上地对其进行层次聚类。

表2 例子  
Table 2 Example

	$a_1$	$a_2$
$x_1$	2	1
$x_2$	4	5
$x_3$	3	2

将第三层上的每个等价类看成一个类簇,则:

$$M_3 = \{M_3^1 = \{x_1\}, M_3^2 = \{x_2\}, M_3^3 = \{x_3\}\}$$

在第三层上,类簇 $M_3^1$ 与类簇 $M_3^2$ 和 $M_3^3$ 之间的距离分别为4.47,1.41,类簇 $M_3^2$ 与类簇 $M_3^3$ 之间的距离为3.16。故类簇 $M_3^1$ 与类簇 $M_3^3$ 之间的欧氏距离最小,因此这两个类簇合并成一个类簇,则第二层的类簇为:

$$M_2 = \{M_2^1 = \{x_1, x_3\}, M_2^2 = \{x_2\}\}$$

同理可得,第一层的类簇为:

$$M_1 = \{M_1^1 = \{x_1, x_2, x_3\}\}$$

对不利区域中的等价类进行层次聚类后,不同层次上的聚类簇数不同。在同一层次上,针对每一个类簇,我们根据HFGAV为每个类簇构造参考对象。

**定义 10**(全局属性值频率最高准则, HFGAV) 给定一个决策表 $S$ ,集合 $U$ 被划分为3个区域 $\pi = \{P_1, P_2, P_3\}$ ,设 $P_i$ 表示有利区域, $P_j (j \neq i)$ 表示不利区域,对 $P_j$ 中的等价类进行层次聚类。假设在第 $t (t = 1, 2, \dots, n)$ 层上包含的类簇为 $M_t = \{M_t^1, M_t^2, \dots, M_t^l\}$ ,对于每个类簇 $M_t^l (l = 1, 2, \dots, t)$ , $\forall f \in A_f, V_f = \{v_1, v_2, \dots, v_{|V_f|}\}$ , $f$ 的属性值 $v_i (1 \leq i \leq |V_f|)$ 出现的频率记为 $F_f(v_i)$ ,则簇 $M_t^l$ 的参考对象构造如下:

$$x_t^l: \{I_f(x_t^l) = v_f^l \mid \forall f \in A_f\} \quad (8)$$

其中, $v_f^l = \{v_i \in V_f \mid F_f(v_i) > F_f(v_j), 1 \leq j \leq |V_f|, i \neq j\}$ 。

该准则选取了类簇 $M_t^l$ 中每个灵活属性频率最高的属性值,所构造出的参考对象根据目标对象制定的移动策略适用于大多数对象。然后,我们在有利区域中随机选取一个目标对象,制定相应的移动策略,使得参考对象的灵活属性值转变为目标对象的灵活属性值。

**定义 11** 给定一个决策表 $S$ ,集合 $U$ 被划分为3个区域 $\pi = \{P_1, P_2, P_3\}$ ,设 $P_i$ 表示有利区域, $P_j (j \neq i)$ 表示不利区域,对 $P_j$ 中的等价类进行层次聚类。假设在第 $t (t = 1, 2, \dots, n)$ 层上包含的类簇为 $M_t = \{M_t^1, M_t^2, \dots, M_t^l\}$ ,对于每个类簇 $M_t^l (l = 1, 2, \dots, t)$ ,构造 $x_t^l$ 为参考对象,并在有利区域中随机选取 $y \in P_i$ 为目标对象,则由参考对象到目标对象的移动策略为:

$$r_{x_t^l} \rightarrow r_y: \left[ \bigwedge_{f \in A_f} I_f(x_t^l) \rightarrow I_f(y) \right] \Rightarrow I_d(x_t^l) \rightarrow I_d(y) \quad (9)$$

$$\text{s. t. } \left[ \bigwedge_{s \in A_s} I_s(x_t^l) = I_s(y) \right]$$

其中,  $x_i^l$  表示所构造的簇  $M_i^l$  的参考对象,  $I_f(x_i^l)$  表示参考对象  $x_i^l$  的灵活属性值, 当簇中对象与目标对象的稳定属性值同时, 它们的移动策略相同。

### 3.2 实例分析

我们用一个简单的医疗实例来阐明所提出的基于层次聚类的三支决策移动策略模型。表 3 是一个医疗决策表, 包括 1 个稳定属性、3 个灵活属性和一个决策属性。其中稳定属性“性别”有女性和男性两种属性值, 为了方便描述, 我们用“0”表示女性, “1”表示男性, 灵活属性“咳嗽”的属性值正常、较轻、严重分别用“1”“2”“3”表示, “体温”的属性值正常、较高、高烧也分别用“1”“2”“3”表示, 灵活属性“头痛”和决策属性“感冒”的属性值是和否分别用“1”和“0”表示。

表 3 医疗决策

Table 3 Medical decision

	性别	咳嗽	体温	头痛	感冒
$x_1$	0	1	1	0	0
$x_2$	0	2	3	1	1
$x_3$	0	3	3	0	1
$x_4$	1	2	1	0	0
$x_5$	0	1	2	1	1
$x_6$	1	1	1	1	0
$x_7$	0	1	3	1	1
$x_8$	0	2	2	0	1
$x_9$	0	2	2	0	0

定义  $X = \{x \in U \mid I_{感冒}(x) = 0\}$  为无感冒症状的人的集合。代价矩阵如表 4 所列, 根据一对阈值 ( $\alpha = 0.6, \beta = 0.4$ ) 和条件概率, 可将对象划分到 3 个两两不相交的区域中:

$$P_1 = \{x_1, x_4, x_6\}$$

$$P_2 = \{x_8, x_9\}$$

$$P_3 = \{x_2, x_3, x_5, x_7\}$$

表 4 代价矩阵

Table 4 Cost matrix

	$P_1$	$P_2$	$P_3$
$X$	2	8	14
$X^-$	20	11	7

对于  $P_3$  中的对象, 我们对其进行层次聚类, 每一层次的聚类结果如下:

$$M_1: M_1^1 = \{x_2, x_3, x_5, x_7\}$$

$$M_2: M_2^1 = \{x_3\}, M_2^2 = \{x_2, x_5, x_7\}$$

$$M_3: M_3^1 = \{x_3\}, M_3^2 = \{x_5\}, M_3^3 = \{x_2, x_7\}$$

$$M_4: M_4^1 = \{x_3\}, M_4^2 = \{x_5\}, M_4^3 = \{x_2\}, M_4^4 = \{x_7\}$$

根据 HFGAV, 我们为每一层次的每个类簇制定一个移动策略。当频率最高的属性值有多个时, 随机选取其中一个作为参考对象的属性值。我们随机选取  $P_1$  中的  $x_1$  作为目标对象。

在  $M_1$  层上共有 1 个类簇。根据类簇中每个灵活属性的属性值出现的频率:  $F_{咳嗽}(1) = 2, F_{咳嗽}(2) = 1, F_{咳嗽}(3) = 1, F_{体温}(2) = 1, F_{体温}(3) = 3, F_{头痛}(0) = 1, F_{头痛}(1) = 3$ , 故簇  $M_1^1$  的参考对象  $x_1^1$  为: 咳嗽 = 1, 体温 = 3, 头痛 = 1,

簇  $M_1^1$  的移动策略为:

$$r_{x_1^1} \rightarrow r_{x_1}: \text{体温}: 3 \rightarrow 1 \wedge \text{头痛}: 1 \rightarrow 0 \Rightarrow \text{感冒}: 1 \rightarrow 0$$

在  $M_2$  层上, 簇  $M_2^1$  只包含一个对象, 故  $M_2^1$  的移动策略为:

$$r_{x_3^1} \rightarrow r_{x_1}: \text{咳嗽}: 3 \rightarrow 1 \wedge \text{体温}: 3 \rightarrow 1 \Rightarrow \text{感冒}: 1 \rightarrow 0$$

同理可得, 簇  $M_2^2$  的移动策略为:

$$r_{x_2^2} \rightarrow r_{x_1}: \text{体温}: 3 \rightarrow 1 \wedge \text{头痛}: 1 \rightarrow 0 \Rightarrow \text{感冒}: 1 \rightarrow 0$$

在  $M_3$  层上, 每个簇的移动策略为:

$$r_{x_3^2} \rightarrow r_{x_1}: \text{咳嗽}: 3 \rightarrow 1 \wedge \text{体温}: 3 \rightarrow 1 \Rightarrow \text{感冒}: 1 \rightarrow 0$$

$$r_{x_5^2} \rightarrow r_{x_1}: \text{体温}: 2 \rightarrow 1 \wedge \text{头痛}: 1 \rightarrow 0 \Rightarrow \text{感冒}: 1 \rightarrow 0$$

$$r_{x_7^2} \rightarrow r_{x_1}: \text{体温}: 3 \rightarrow 1 \wedge \text{头痛}: 1 \rightarrow 0 \Rightarrow \text{感冒}: 1 \rightarrow 0$$

在  $M_4$  层上, 每个类簇只包含一个对象, 其移动策略分别为:

$$r_{x_3^3} \rightarrow r_{x_1}: \text{咳嗽}: 3 \rightarrow 1 \wedge \text{体温}: 3 \rightarrow 1 \Rightarrow \text{感冒}: 1 \rightarrow 0$$

$$r_{x_2^4} \rightarrow r_{x_1}: \text{体温}: 2 \rightarrow 1 \wedge \text{头痛}: 1 \rightarrow 0 \Rightarrow \text{感冒}: 1 \rightarrow 0$$

$$r_{x_7^4} \rightarrow r_{x_1}: \text{咳嗽}: 2 \rightarrow 1 \wedge \text{体温}: 3 \rightarrow 1 \wedge \text{头痛}: 1 \rightarrow 0 \Rightarrow \text{感冒}: 1 \rightarrow 0$$

$$r_{x_1^4} \rightarrow r_{x_1}: \text{体温}: 3 \rightarrow 1 \wedge \text{头痛}: 1 \rightarrow 0 \Rightarrow \text{感冒}: 1 \rightarrow 0$$

对不利区域施加策略, 使得对象从不利区域移动到有利区域, 对象的稳定属性不会发生改变, 因此, 对象的移动代价由灵活属性  $I_f(x) \rightarrow I_f(y)$  的改变而产生。我们可以根据移动策略以及式(6)、式(5)计算对象移动产生的代价和收益。假设每个灵活属性的权重为  $\omega_f$ , 则参考对象  $x$  移动到目标对象  $y$  所在区域的移动代价为:

$$C_{r_x \rightarrow r_y} = \sum_{f \in A_f} \omega_f C_f(I_f(x), I_f(y)) \quad (10)$$

对于每一个类簇来说, 该类簇中对象的移动策略都相同, 因此, 每个簇的移动代价是簇中所有对象的移动代价之和。而不利区域的移动代价则是所有类簇的移动代价之和, 可以表示为:

$$C_{r_x \rightarrow r_y} = \sum_{i=1}^l |M_i^l| \sum_{f \in A_f} \omega_f C_f(I_f(x), I_f(y)) \quad (11)$$

令  $\omega_{咳嗽} = \omega_{体温} = \omega_{头痛} = 1$ , 灵活属性咳嗽、体温、头痛的代价函数分别如表 5—表 7 所列。

表 5 “咳嗽”代价函数

Table 5 “Cough” cost function

	1	2	3
1	0	1	2
2	1	0	1
3	2	1	0

表 6 “体温”代价函数

Table 6 “Fever” cost function

	1	2	3
1	0	1	2
2	1	0	1
3	2	1	0

表 7 “头痛”代价函数

Table 7 “Headache” cost function

	0	1
0	0	1
1	1	0

根据表 3 的医疗决策信息表以及层次聚类的结果,可以计算出不同层次上不利区域中对象的移动代价。

在  $M_1$  层上,簇  $M_1^1$  的移动代价为:

$$\begin{aligned} C_{M_1} &= C_{M_1^1} = |M_1^1| \sum_{f \in A_f} \omega_f C_f(I_f(x_1^1), I_f(y)) \\ &= 4 * [\omega_{体温} C_{体温}(3, 1) + \omega_{头痛} C_{头痛}(1, 0)] \\ &= 4 * 3 = 12 \end{aligned}$$

同理,在  $M_2, M_3, M_4$  层上,移动代价为:

$$\begin{aligned} C_{M_2} &= C_{M_2^1} + C_{M_2^2} = 4 + 9 = 13 \\ C_{M_3} &= C_{M_3^1} + C_{M_3^2} + C_{M_3^3} = 4 + 2 + 6 = 12 \\ C_{M_4} &= C_{M_4^1} + C_{M_4^2} + C_{M_4^3} + C_{M_4^4} = 4 + 2 + 4 + 3 = 13 \end{aligned}$$

为对象施加移动策略后,根据移动成功的对象个数,我们可以计算不同层次上对象移动带来的收益。用  $\Delta X$  表示移动成功的对象的集合:

$$\Delta X = P_j \cap P_i' \quad (12)$$

其中,  $P_j$  表示不利区域,  $P_i'$  表示施加策略后形成的新的有利区域。  $\Delta X$  中的对象在移动前有两种状态,一种是属于  $X$ ,另一种是属于  $X^-$ ,两种状态下的对象数量分别为:

$$\begin{aligned} a_1 &= |\Delta X \cap X| \\ b_1 &= |\Delta X \cap X^-| \end{aligned} \quad (13)$$

在移动之后,  $\Delta X$  中的对象同样有两种状态,其数量可根据目标对象  $y$  的等价类  $[y]$  来计算:

$$\begin{aligned} a_2 &= |\Delta X| \frac{|X \cap [y]|}{|[y]|} \\ b_2 &= |\Delta X| \left( 1 - \frac{|X \cap [y]|}{|[y]|} \right) \end{aligned} \quad (14)$$

其中,  $\frac{|X \cap [y]|}{|[y]|}$  表示等价类  $[y]$  中的对象属于  $X$  的概率。

根据  $\Delta X$  中对象移动前后的两种状态的数量以及代价矩阵,可以计算得到不利区域中的对象移动到有利区域所产生的收益。

$$B = W_j [a_1 \lambda_{j1} + b_1 \lambda_{j3}] - W_i [a_2 \lambda_{i1} + b_2 \lambda_{i3}] \quad (15)$$

其中,  $i, j \in \{1, 2, 3\}$ ,  $W_i, W_j$  表示区域的权重,  $a_1 \lambda_{j1} + b_1 \lambda_{j3}$  表示移动成功的对象移动之前在不利区域中的代价,  $a_2 \lambda_{i1} + b_2 \lambda_{i3}$  表示移动成功的对象移动之后在有利区域中的代价。

假设各区域的权重均为 1,则不同层次上的移动收益为:

在  $M_1$  层上,移动成功的对象  $\Delta X = \{x_7\}$ 。

由式(13)可得移动前两种状态的数量:

$$\begin{aligned} a_1 &= |\Delta X \cap X| = 0 \\ b_1 &= |\Delta X \cap X^-| = 1 \end{aligned}$$

由式(14)可得移动后两种状态的数量:

$$\begin{aligned} a_2 &= |\Delta X| \frac{|X \cap [y]|}{|[y]|} = 1 \\ b_2 &= |\Delta X| \left( 1 - \frac{|X \cap [y]|}{|[y]|} \right) = 0 \end{aligned}$$

故移动收益为:

$$\begin{aligned} B_{M_1} &= W_3 [a_1 \lambda_{31} + b_1 \lambda_{33}] - W_1 [a_2 \lambda_{11} + b_2 \lambda_{13}] \\ &= [0 * 14 + 1 * 7] - [1 * 2 + 0 * 20] = 5 \end{aligned}$$

同理,  $M_2, M_3, M_4$  层的移动收益为:

$$B_{M_2} = B_{M_2^1} + B_{M_2^2} = 5 + 5 = 10$$

$$B_{M_3} = B_{M_3^1} + B_{M_3^2} + B_{M_3^3} = 5 + 5 + 5 = 15$$

$$B_{M_4} = B_{M_4^1} + B_{M_4^2} + B_{M_4^3} + B_{M_4^4} = 5 + 5 + 5 + 5 = 20$$

最后,我们根据上述计算得到的移动代价和移动收益,利用式(7)计算不同层次上的移动效用。

$M_1$  层的移动效用为:

$$E_{M_1} = B_{M_1} - C_{M_1} = 5 - 12 = -7$$

同理,其余各层的移动效用为:

$$E_{M_2} = B_{M_2} - C_{M_2} = 10 - 13 = -3$$

$$E_{M_3} = B_{M_3} - C_{M_3} = 15 - 12 = 3$$

$$E_{M_4} = B_{M_4} - C_{M_4} = 20 - 13 = 7$$

由计算结果可知,从多个层次上,为不利区域中的对象制定移动策略,聚类簇数越多,移动成功的对象个数越多,移动效用就越大。

## 4 实验分析

本节使用 UCI 上的数据集,根据 HFGAV 为不同层次上的每个类簇制定相应的移动策略。我们对数据集做了相应的处理,对于缺失值,我们使用该属性下出现频率最高的属性值来代替,并对数据集中连续的属性值进行了(1-5)范围内离散化。各数据集中的“age”“sex”属性以及 CMC 数据集的“number of child”属性被看作稳定属性,其余为灵活属性。数据集的基本信息如表 8 所列。实验环境的具体配置如下:Microsoft Windows10、Inter (R) Core (TM) i3-4170 CPU @ 3.70 GHz,4.0 GB 内存、编程语言为 Python。

表 8 数据集的基本信息

Table 8 Basic information of dataset

编号	数据集	对象	条件属性
1	Caesarian Section Classification Dataset (CSC)	80	5
2	Mammographic Mass(MM)	961	5
3	Abalone	4177	8
4	Blood Transfusion Service Center Dataset(BTSC)	748	4
5	Contraceptive Method Choic(CMC)	1473	9
6	Wireless	2000	7

实验中,代价矩阵在满足一定关系的前提下由计算机随机模拟给出,可以看成是由该领域的专家给出。根据代价矩阵,我们可将数据集划分为互不相交的 3 个区域  $P_1, P_2, P_3$ , 其中  $P_1$  为有利区域,  $P_3$  为不利区域。对不利区域中的对象进行层次聚类,将其划分为不同的层次。在实验过程中,我们选取类簇数  $K=1, 2, 4, 8, \dots, n$  的层次,并为每一层次的每个类簇制定相应的移动策略。类簇数越多,需要制定的移动策略也越多。

对于每个数据集,我们随机从有利区域  $P_1$  中选取一个目标对象。对于不利区域  $P_3$  中对象的移动代价,我们利用代价函数  $C_f(v_i, v_j) = \omega_f |v_i - v_j|$  计算,其中  $\omega_f$  为属性  $f$  的权重,实验中所有属性的权重都为  $1/|A_f|$ 。其移动代价如图 1 所示,  $X$  轴表示不同层次上的聚类簇数,  $Y$  轴表示不利区域中的移动代价。

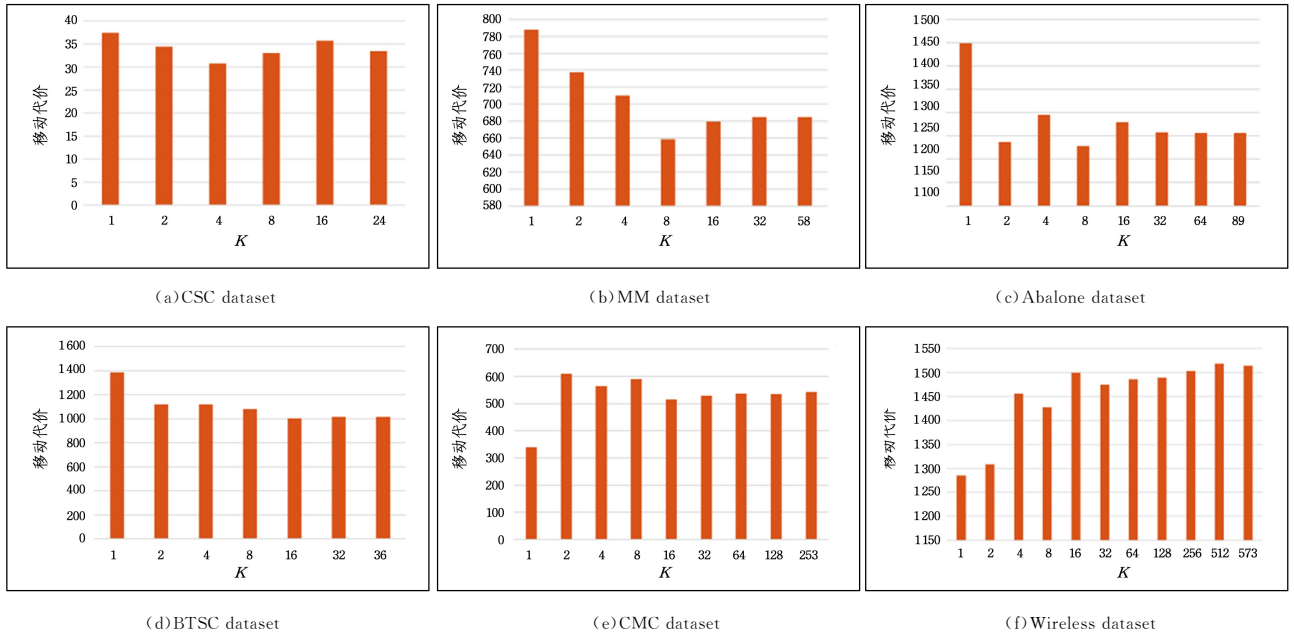


图1 移动代价

Fig. 1 Movement cost

从图中可以看出,在不同层次上,移动代价是变化不定的,这是因为我们根据 HFGAV 来选取每个类簇的参考对象,该参照对象与类簇内对象的属性值频率有关,而与类簇的个数无关。因此,当聚类个数逐渐增加时,其移动的代价并不随之单调变化。

除了移动代价,我们还计算了不同类簇数下移动成功的对象个数,其结果如图 2 所示。随着类簇数的增加,针对每个类簇制定的移动策略会适合更多的对象,能够成功从不利区域移动到有利区域的对象数目也越多。当所有等价类合并为一个类簇时,相当于基于区域的移动,为整个不利区域制定一种移动策略,故移动成功的对象个数

较少;当类簇数为等价类个数时,相当于基于对象的移动,为每个对象制定一种移动策略,因此,移动成功的对象个数最多。

为了进一步评估移动策略是否有效,我们根据式(7),利用移动收益与移动代价的差值来计算移动效用。实验过程中,移动收益是根据移动成功的对象个数来计算的,对于未成功移动的对象,我们假设其移动收益等于移动代价,即未成功移动对象的移动效用为零。如图 3 所示,当类簇数增多时,移动效用逐渐增大。移动效用与移动成功的对象个数有关,类簇数越多,移动策略和移动成功的对象个数越多,移动效用也就越大。

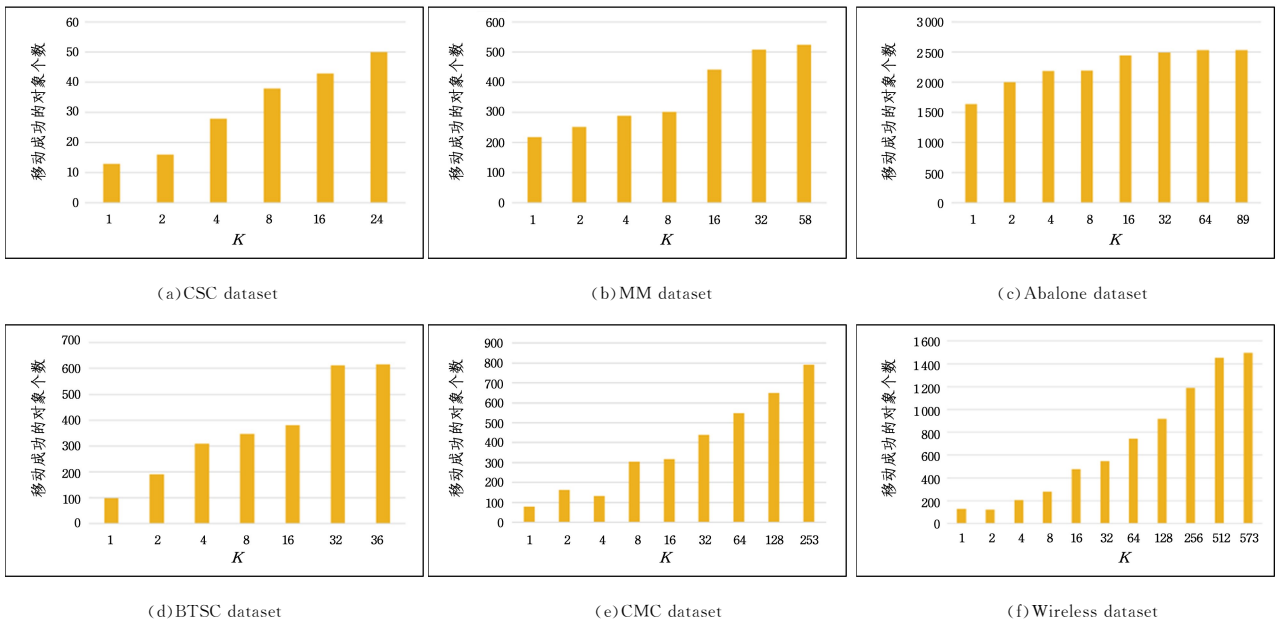


图2 移动成功的对象个数

Fig. 2 Number of successfully moved objects

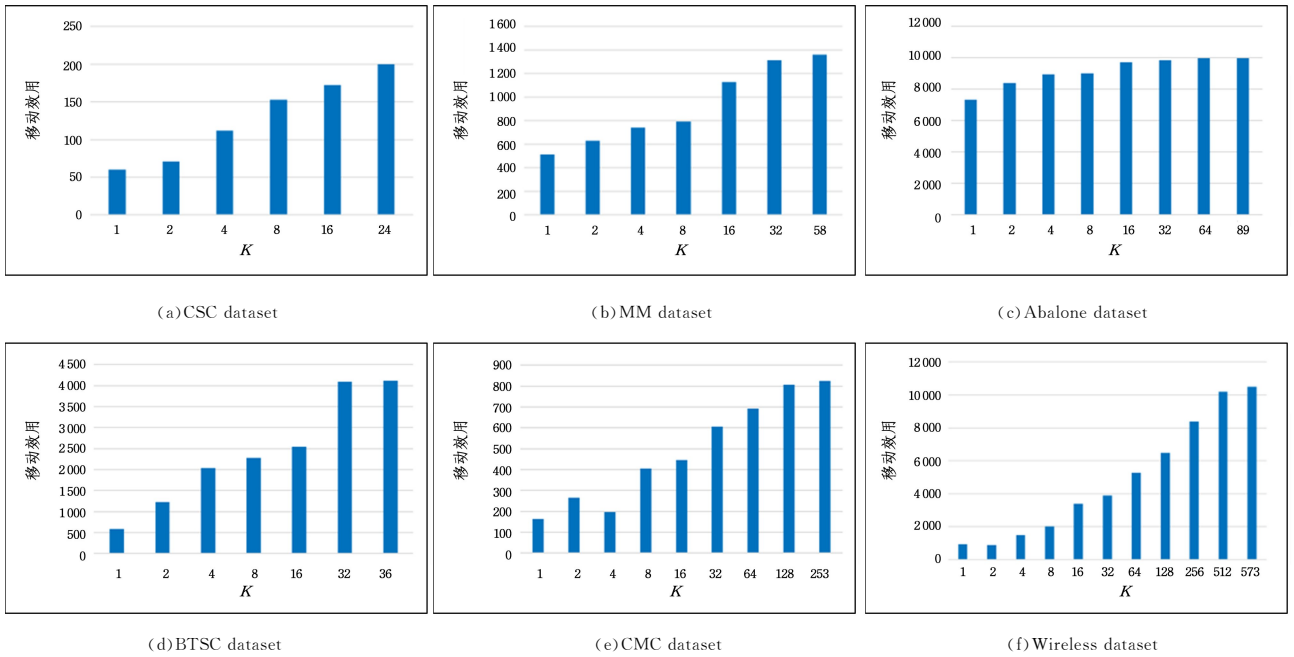


图3 移动效用

Fig. 3 Movement utility

综上所述,我们利用层次聚类,从多层次的角度为不利区域中的对象制定了相应的移动策略。从实验结果来看,不同层次上,聚类结果不同,所制定的移动策略也不同,移动成功的对象个数、移动代价以及移动效用都发生了变化。由实验结果可以看出,本文提出的基于层次聚类的三支决策移动策略模型在制定多个层次的移动策略方面是有效的。

**结束语** 本文提出了一种基于层次聚类的三支决策移动策略模型,根据层次聚类的结果,从多层次的角度,为不利区域中的对象制定移动策略。并且分析了在不同层次上,不利区域中对象的移动代价、移动成功的对象个数以及移动效用的变化。最后,我们通过实验证明了该模型的有效性。

本文只是从多层次的角度分析了移动策略的制定,并未给出在实际应用中,如何从多层次中选取一个最佳的层次,为不利区域中的对象制定相应的移动策略,以满足决策者的需求。

## 参考文献

[1] YAO Y Y. Three-way decisions and cognitive computing[J]. *Cognitive Computation*, 2016, 8(4): 543-554.

[2] YAO Y Y. Three-way decisions with probabilistic rough sets[J]. *Information Sciences*, 2010, 180(3): 341-353.

[3] LIU D, LIANG D C, WANG C C. A novel three-way decision model based on incomplete information system[J]. *Knowledge-Based Systems*, 2016, 91: 32-45.

[4] FANG Y, GAO C, YAO Y Y. Granularity-driven sequential three-way decisions: a cost-sensitive approach to classification[J]. *Information Sciences*, 2020, 507: 644-664.

[5] WANG P X, YAO Y Y. CE3: A Three-way Clustering Method

Based on Mathematical Morphology[J]. *Knowledge-Based Systems*, 2018, 155: 54-65.

[6] YU H, ZHANG C, WANG G Y. A tree-based incremental overlapping clustering method using the three-way decision theory[J]. *Knowledge-Based Systems*, 2016, 91: 189-203.

[7] YU H, WANG X C, WANG G Y, et al. An active three-way clustering method via low-rank matrices for multi-view data[J]. *Information Sciences*, 2020, 507: 823-839.

[8] HUANG C C, LI J H, MEI C L, et al. Three-way concept learning based on cognitive operators: An information fusion viewpoint[J]. *International Journal of Approximate Reasoning*, 2017, 83: 218-242.

[9] QI J J, QIAN T, WEI L. Connections between three-way and classical concept lattices[J]. *Knowledge-Based Systems*, 2016, 91: 143-151.

[10] YU H Y, LI Q G, CAI M J. Characteristics of three-way concept lattices and three-way rough concept lattices[J]. *Knowledge-Based Systems*, 2018, 146: 181-189.

[11] LIU D, YE X Q. A matrix factorization based dynamic granularity recommendation with three-way decisions[J]. *Knowledge-Based Systems*, 2020, 191: 105243.

[12] ZHANG H R, MIN F. Three-way recommender systems based on random forests[J]. *Knowledge-Based Systems*, 2016, 91: 275-286.

[13] ZHANG H R, MIN F, SHI B. Regression-based three-way recommendation[J]. *Information Sciences*, 2017, 378: 444-461.

[14] FANG Y, MIN F. Cost-sensitive approximate attribute reduction with three-way decisions[J]. *International Journal of Approximate Reasoning*, 2019, 104: 148-165.

[15] MA X A, ZHAO X R. Cost-sensitive three-way class-specific at-

- tribute reduction[J]. *International Journal of Approximate Reasoning*, 2019, 105: 153-174.
- [16] LIANG D C, WANG M W, XU Z S. Heterogeneous multi-attribute nonadditivity fusion for behavioral three-way decisions in interval type-2 fuzzy environment [J]. *Information Sciences*, 2019, 496: 242-263.
- [17] WANG T X, LI H X, ZHOU X Z, et al. A prospect theory-based three-way decision model[J]. *Knowledge-Based Systems*, 2020, 203: 106129.
- [18] YANG X, LI T R, FUJITA H, et al. A sequential three-way approach to multi-class decision[J]. *International Journal of Approximate Reasoning*, 2019, 104: 108-125.
- [19] QIAN J, LIU C H, MIAO D Q, et al. Sequential three-way decisions via multi-granularity[J]. *Information Sciences*, 2020, 507: 606-629.
- [20] LANG G M, LUO J F, YAO Y Y. Three-way conflict analysis; a unification of models based on rough sets and formal concept analysis[J]. *Knowledge-Based Systems*, 2020, 194: 105556.
- [21] YAO Y Y. Three-way conflict analysis; reformulations and extensions of the pawlak model[J]. *Knowledge-Based Systems*, 2019, 180: 26-37.
- [22] JIANG C M, YAO Y Y. Effectiveness measures in movement based three-way decisions[J]. *Knowledge-Based Systems*, 2018, 160: 136-143.
- [23] GAO C, YAO Y Y. Actionable strategies in three-way decisions [J]. *Knowledge-Based Systems*, 2017, 133: 141-155.
- [24] PAWLAK Z. Rough sets[J]. *International Journal of Computer and Information Sciences*, 1982, 11(5): 341-356.



**XU Yi**, born in 1981, Ph.D, professor, is a member of China Computer Federation. Her main research interests include intelligent information processing, granular computing and edge computing.

(责任编辑:杨雪敏)