

# 控制与决策

Control and Decision

## 多尺度决策系统中代价敏感的最优尺度组合

张清华, 张雪秋, 庞国弘

引用本文:

张清华, 张雪秋, 庞国弘. 多尺度决策系统中代价敏感的最优尺度组合[J]. *控制与决策*, 2021, 36(10): 2369–2378.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2020.0121>

---

## 您可能感兴趣的其他文章

Articles you may be interested in

### 基于数据分布特性的代价敏感宽度学习系统

Data distribution-based cost-sensitive broad learning system

控制与决策. 2021, 36(7): 1686–1692 <https://doi.org/10.13195/j.kzyjc.2019.1484>

### 基于犹豫度和相似度的专家权重确定方法及其应用

Expert weights determination method and application based on hesitancy degree and similarity measure

控制与决策. 2021, 36(6): 1482–1488 <https://doi.org/10.13195/j.kzyjc.2019.1382>

### 基于知识粒度特征的多目标粗糙集属性约简算法

Multi objective rough set attribute reduction algorithm based on characteristics of knowledge granularity

控制与决策. 2021, 36(1): 196–205 <https://doi.org/10.13195/j.kzyjc.2019.0490>

### 一种高匹配性的多层代价地图生成算法

A high matching layered costmap generation algorithm

控制与决策. 2020, 35(12): 2883–2888 <https://doi.org/10.13195/j.kzyjc.2018.0721>

### 考虑时间序列的动态大群体应急决策方法

Dynamic large group emergency decision-making method considering time series

控制与决策. 2020, 35(11): 2609–2618 <https://doi.org/10.13195/j.kzyjc.2019.0088>

# 多尺度决策系统中代价敏感的最优尺度组合

张清华<sup>†</sup>, 张雪秋, 庞国弘

- (1. 重庆邮电大学 计算机科学与技术学院, 重庆 400065;
2. 重庆邮电大学 计算智能重庆市重点实验室, 重庆 400065)

**摘要:** 最优尺度组合是多尺度决策系统中的研究热点之一, 现有的研究大多是从一致性、不确定性的角度出发, 而没有充分考虑代价信息的影响. 针对该问题, 首先分析最优尺度组合中考虑代价敏感的重要性, 从决策代价的角度提出基于测试代价和延迟代价的多尺度决策系统, 并且定义了尺度代价和属性代价来刻画尺度和属性所产生的代价; 其次, 考虑实际场景中属性代价的影响, 将属性重要度和属性代价结合进行属性排序; 最后, 在属性进行最优尺度选择时, 考虑尺度代价的影响, 建立了一个代价敏感的最优尺度组合选择模型. 实验结果表明, 在现有代价认知场景下, 所提模型能合理地进行最优尺度组合选择, 所得结果更符合实际需求.

**关键词:** 多尺度决策系统; 属性重要度; 代价敏感; 最优尺度组合

中图分类号: TP18

文献标志码: A

DOI: 10.13195/j.kzyjc.2020.0121

开放科学(资源服务)标识码(OSID):



**引用格式:** 张清华, 张雪秋, 庞国弘. 多尺度决策系统中代价敏感的最优尺度组合 [J]. 控制与决策, 2021, 36(10): 2369-2378.

## Cost-sensitive optimal scale combination in multi-scale decision systems

ZHANG Qing-hua<sup>†</sup>, ZHANG Xue-qiu, PANG Guo-hong

- (1. School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;
2. Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

**Abstract:** Optimal scale combination is one of the research hotspots in multi-scale decision systems. Most existing researches start from the perspective of consistency and uncertainty, without fully considering the influence of cost information. In view of this issue, the importance of considering cost-sensitivity in optimal scale combination is analyzed. Firstly, a multi-scale decision system based on test cost and delay cost from the perspective of decision cost is proposed, and the scale cost and attribute cost are defined to describe the cost caused by scale and attribute. Then, considering the effect of attribute cost in the real scene, attribute significance and attribute cost are combined to sort the attributes. Finally, a cost-sensitive optimal scale combination selection model is established by considering the influence of scale cost when selecting the optimal scale of attributes. The experimental results show that the proposed model can reasonably select the optimal scale combination with the existing cost cognition scenario, which is more in line with the actual requirements.

**Keywords:** multi-scale decision systems; attribute significance; cost-sensitivity; optimal scale combination

## 0 引言

近年来, 粒计算 (granular computing, GrC)<sup>[1-2]</sup> 逐渐发展成为解决复杂问题、大规模数据挖掘和模糊信息处理的有效工具. 由 Pawlak<sup>[3]</sup> 提出的粗糙集理论对 GrC 的推动和发展起到了至关重要的作用, 它能够很好地利用等价关系对论域中的粒子进行构造、解释和表示. 在此基础上, 一系列扩展的粗糙集模型被先后提出, 如概率粗糙集<sup>[4-6]</sup>、决策理论粗糙集<sup>[7]</sup>、

多粒度粗糙集<sup>[8-10]</sup> 等. 粗糙集已经成为一种重要的智能信息处理技术<sup>[11-13]</sup>, 越来越受到广大专家学者的青睐. 其中, 层次粗糙集是在不同层次下分析决策系统, 叶明全等<sup>[14]</sup> 针对条件属性具有属性值分类的情况, 结合全子树泛化模式, 提出一种多层次粗糙集模型; Zhang<sup>[15]</sup> 基于层次粗糙集理论, 构建了一个应急规则知识库, 从不同抽象层次挖掘决策规则集. 在这些研究中, 决策系统中对象的属性只选取唯一属性

收稿日期: 2020-02-08; 修回日期: 2020-04-14.

基金项目: 国家自然科学基金项目 (61876201).

责任编委: 刘宝碇.

<sup>†</sup>通讯作者. E-mail: zhangqh@cqupt.edu.cn.

值,即对象的属性信息被固定在某一尺度下的属性值.然而,单一尺度的知识表示及数据处理方法已远不能满足实际应用中用户的需求,对象通常在同一属性下以不同的尺度进行测量会获取到不同的属性值,比如学生的学习成绩,可以用“好”“坏”来评价,也可以用“优”“良”“中”“差”来评价,或者用百分制来评价,不同尺度下做决策可能会导致不同的决策结果.因此,在实际应用中,为了更好地做出决策,人们需要从多个尺度对数据进行表示、分析.

为了解决多粒度所描述的问题, Wu等<sup>[16-17]</sup>将多尺度的观点引入信息表和决策表中,提出了基于多尺度标记划分的粗糙集数据分析方法,并进行了最优尺度选择的研究.在多尺度标记数据模型下,对象的属性可以具有不同的属性值以表示该属性在不同尺度下进行测量,人们可以根据需要在不同尺度的属性信息下处理、分析数据,从而更好地做出决策.近几年,学者们陆续提出在多尺度框架下针对不同类型的粗糙集进行数据分析和处理的方法<sup>[18-27]</sup>.在文献<sup>[18-27]</sup>中,多尺度决策系统都基于一个共同的假设,即系统中所有的条件属性都具有相同的尺度个数,而基于这一假设构建的最优尺度选择模型在实际应用中存在一些局限性.针对以上问题, Li等<sup>[28-29]</sup>提出推广的多尺度数据分析模型,研究了不同属性具有不同尺度个数的最优尺度选择方法.基于此,许多学者也进行了尺度组合的研究<sup>[30-32]</sup>.

目前,代价敏感学习在数据挖掘和机器学习的研究领域中具有重要的地位<sup>[33-34]</sup>,其主要目的是处理决策过程中所产生的各种代价问题.不少学者将代价敏感运用于粗糙集理论中,并且进行了较为深入的研究<sup>[35-36]</sup>.代价大致可分为决策过程代价和决策结果代价,从决策过程代价的角度来看,可看作是获取属性值所付出的代价;从决策结果代价的角度来看,可看作是对象因误分类而产生的风险. Min等<sup>[37]</sup>在测试代价敏感决策系统中引入了分层模型; Yang等<sup>[38]</sup>提出测试代价敏感多粒度粗糙集模型,将测试代价融入到传统决策粗糙集的上、下近似中,给出了新的近似定义方法.这些研究均表明代价敏感学习在决策过程中的重要性.然而,目前在多尺度决策系统的最优尺度组合研究中,没有充分考虑代价敏感的问题,大多最优尺度组合的研究仅考虑了属性或尺度的区分能力,没有充分考虑属性或尺度的代价.属性的尺度越细,划分能力越强,往往伴随着越大的代价.在实际决策过程中,考虑代价信息是非常重要的,因此代价敏感的最优尺度组合是值得研究的,且具有实际意义.

现实中,人们做决策时往往希望精度尽可能高,所付出的代价尽可能低.比如,在对大量供应商进行筛选时,需要对供应商的各个指标进行评判,获取各个属性(指标)的方法可以通过调查问卷这样低成本的方式,也可以通过对供应商内部走访、考察以获取更精确的评价信息,前者的区分能力弱代价低,后者的区分能力强代价高.在实际生活中,人们往往希望能同时考虑属性在尺度选择时的区分能力和代价.为了解决这个问题,在多尺度决策系统中选择出更符合人类认知的最优尺度组合,本文基于代价敏感的角度,首先提出一种基于测试代价和延迟代价的多尺度决策系统,并且定义尺度代价和属性代价来刻画获取不同尺度下的属性值和属性的代价;其次,综合考虑属性重要度和属性代价两个因素,提出加权属性重要度进行属性排序;最后,基于属性排序序列,提出一种代价敏感的最优尺度组合选择模型.实验结果表明,在给定代价场景下,所提模型能合理地进行最优尺度组合选择,且所得到的最优尺度组合结果更符合实际需求.

## 1 相关基本概念

在本节中,为了更清楚地对本文进行阐述,下面给出一些关于粗糙集近似及多尺度决策系统的基本概念.

**定义1(粗糙集近似)**<sup>[3]</sup> 设 $U$ 表示有限非空论域, $R \subseteq U \times U$ 是 $U$ 上的一个等价关系,则 $(U, R)$ 被称为一个Pawlak近似空间.

对于任意 $x, y \in U$ ,如果 $(x, y) \in R$ ,则称 $x$ 和 $y$ 在等价关系 $R$ 下是不可分辨的,可以得到 $U$ 的一个划分 $U/R = \{[x]_R | x \in U\}$ ,其中 $[x]_R = \{y \in U | (x, y) \in R\}$ 表示包含 $x$ 的等价类.

对于任意子集 $X \subseteq U$ , $X$ 的上近似集和下近似集分别被定义为

$$\overline{\text{apr}}(X) = \bigcup \{[x]_R | [x]_R \cap X \neq \emptyset\}, \quad (1)$$

$$\underline{\text{apr}}(X) = \bigcup \{[x]_R | [x]_R \subseteq X\}. \quad (2)$$

此外,论域 $U$ 可以被划分为3个区域,分别为正域、边界域、负域;集合 $\text{BND}_R(X) = \overline{\text{apr}}(X) - \underline{\text{apr}}(X)$ 被称为 $X$ 的边界域, $\text{POS}_R(X) = \underline{\text{apr}}(X)$ 被称为 $X$ 的正域, $\text{NEG}_R(X) = U - \overline{\text{apr}}(X)$ 被称为 $X$ 的负域.

**定义2(多尺度决策系统)**<sup>[16]</sup> 给定一个多尺度决策系统 $S = (U, \text{AT} \cup \{d\})$ ,其中 $U = \{x_1, x_2, \dots, x_n\}$ 是一个非空有限对象集, $\text{AT} = \{a_1, a_2, \dots, a_m\}$ 是一个非空有限条件属性集, $d$ 是决策属性,则一个多尺度决策系统可以表示为

$$S = (U, \{a_j^k | j = 1, 2, \dots, m, k = 1, 2, \dots, s_j\} \cup \{d\}),$$

其中:  $m$  是条件属性个数;  $s_j$  是属性  $a_j$  的尺度个数;  $a_j^k : U \rightarrow V_j^k, V_j^k$  是属性  $a_j^k$  在第  $k$  尺度下的值域, 并且对于  $j = 1, 2, \dots, m, 1 \leq k \leq s_j - 1$ , 存在一个满射  $g_j^{k,k+1} : V_j^k \rightarrow V_j^{k+1}$  使得

$$a_j^{k+1}(x) = g_j^{k,k+1}(a_j^k(x)), x \in U, \quad (3)$$

称  $g_j^{k,k+1}$  为信息粒度变换函数,  $a_j^k(x)$  为对象  $x$  的属性  $a_j$  在第  $k$  尺度下的值.

对于一个属性子集  $B \subseteq AT$ , 可以得到一个等价关系

$$R_B = \{(x, y) \in U \times U | a_j^k(x) = a_j^k(y), \forall a_j^k \in B\}. \quad (4)$$

**定义3**(多尺度决策系统的尺度重要度)<sup>[29]</sup> 给定一个多尺度决策系统  $S = (U, \{a_j^k | j = 1, 2, \dots, m, k = 1, 2, \dots, s_j\} \cup \{d\})$ , 对于属性  $a_j \in AT$  在第  $k$  尺度下的尺度重要度定义为

$$\text{sig}_k(a_j, AT, d) = \gamma_{AT^1 m | \text{jth}=k}(d) - \gamma_{(AT - \{a_j\})^1 m - 1}(d). \quad (5)$$

其中

$$\begin{aligned} 1_{m | \text{jth}=k} &= (1; 1; \dots; 1; k; 1; \dots; 1), \\ AT^{1 m | \text{jth}=k} &= \{a_1^1, a_2^1, \dots, a_{j-1}^1, a_j^k, a_{j+1}^1, \dots, a_m^1\}, \\ (AT - \{a_j\})^{1 m - 1} &= \{a_1^1, a_2^1, \dots, a_{j-1}^1, a_{j+1}^1, \dots, a_m^1\}, \\ \gamma_{AT^1 m | \text{jth}=k}(d) &= |\text{POS}_{R_{AT^1 m | \text{jth}=k}}(d)|/|U|. \end{aligned}$$

**定义4**(多尺度决策系统的属性重要度)<sup>[29]</sup> 给定一个多尺度决策系统  $S = (U, \{a_j^k | j = 1, 2, \dots, m, k = 1, 2, \dots, s_j\} \cup \{d\})$ , 对于属性  $a_j \in AT$  的属性重要度被定义为

$$\text{sig}(a_j, AT, d) = \sum_{k=1}^{s_j} w_k \times \text{sig}_k(a_j, AT, d), \quad (6)$$

其中  $w_k (k = 1, 2, \dots, s_j)$  表示权重, 且  $\sum_{k=1}^{s_j} w_k = 1$ , 假设  $w_k = 1/s_j, k = 1, 2, \dots, s_j$ .

**定义5**(尺度组合)<sup>[28]</sup> 给定一个多尺度决策系统  $S = (U, \{a_j^k | j = 1, 2, \dots, m, k = 1, 2, \dots, s_j\} \cup \{d\})$ ; 属性  $a_1, a_2, \dots, a_m$  分别被限制在第  $l_j$  尺度, 形成一个单尺度决策表  $S^K$ , 其中  $K^m = (l_1; l_2; \dots; l_m)$  被称为多尺度决策表  $S$  中的尺度组合.

## 2 代价敏感的属性重要度计算方法

在实际应用中, 获取属性在不同尺度下的属性值, 会产生不同尺度下的测试代价; 而做决策时, 边界域的对象被延迟决策, 会产生延迟代价. 因此, 考虑到代价信息的影响, 首先提出基于测试代价和延迟代价的多尺度决策系统, 定义尺度代价来刻画某一尺度所产生的代价; 然后基于尺度代价, 定义属性代价来刻画属性所产生的代价; 最后结合属性重要度和属性代价两个因素, 提出加权属性重要度进行属性排序.

**定义6**(基于测试代价和延迟代价的多尺度决策系统) 给定一个多尺度决策系统  $S = (U, \{a_j^k | j = 1, 2, \dots, m, k = 1, 2, \dots, s_j\} \cup \{d\})$ , 令  $\eta = \{\eta_{a_1^1}, \eta_{a_1^2}, \dots, \eta_{a_m^{s_j}}\}, \omega = \{\omega_{a_1^1}, \omega_{a_1^2}, \dots, \omega_{a_j^k}, \dots, \omega_{a_m^{s_j}}\}, \omega_{a_j^k} = \{\omega_{x_1^k}, \omega_{x_2^k}, \dots, \omega_{x_{|U|}^k}\}$ , 其中  $\eta_{a_j^k} (1 \leq j \leq m, 1 \leq k \leq s_j)$  表示属性  $a_j$  在第  $k$  尺度下的测试代价,  $\omega_{a_j^k}$  表示属性  $a_j$  在第  $k$  尺度下对象延迟代价的集合,  $\omega_{x_i^k}$  表示对象  $x_i$  的属性  $a_j$  在第  $k$  尺度下的延迟代价, 则基于测试代价和延迟代价的多尺度决策系统定义为

$$S^* = (U, AT \cup \{d\}, \eta, \omega). \quad (7)$$

**定义7**(尺度代价) 给定一个多尺度决策系统  $S^* = (U, AT \cup \{d\}, \eta, \omega)$ , 对于  $\forall a_j \in AT$ , 第  $k$  个尺度的尺度代价被定义为

$$\text{Cost}_{a_j^k} = |U| \times \eta_{a_j^k} + \sum_{x_i \in \text{BND}_{R_{a_j^k}}(X)} \omega_{x_i^k}. \quad (8)$$

$\text{Cost}_{a_j^k}$  表示属性  $a_j$  在尺度  $k$  下所产生的代价, 该代价由测试代价和延迟代价组成, 分别表示获取属性  $a_j$  在第  $k$  尺度下属性值的测试代价, 以及属性  $a_j$  在第  $k$  尺度下对象所产生的延迟代价.

在多尺度决策系统中, 一个属性具有多个尺度, 属性所产生的代价不仅受单个尺度代价的影响, 而且受各个尺度代价的影响, 如下定义给出了属性代价的定义.

**定义8**(属性代价) 给定一个多尺度决策系统  $S^* = (U, AT \cup \{d\}, \eta, \omega)$ , 对于  $\forall a_j \in AT$ , 其属性代价被定义为

$$\text{Cost}_{a_j} = \frac{1}{s_j} \sum_{k=1}^{s_j} \text{Cost}_{a_j^k}. \quad (9)$$

为了更加清晰地阐述多尺度决策系统中尺度代价和属性代价的意义, 下面给出一个例子详述其计算过程.

**例1** 给定一个多尺度决策系统  $S^* = (U, AT \cup \{d\}, \eta, \omega)$ , 如表1所示. 其中  $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$ , 条件属性  $AT = \{a_1, a_2, a_3\}$ , 测试代价参数  $\eta = \{\eta_{a_1^1}, \eta_{a_1^2}, \eta_{a_1^3}, \eta_{a_2^1}, \eta_{a_2^2}, \eta_{a_2^3}, \eta_{a_3^1}, \eta_{a_3^2}, \eta_{a_3^3}\}$ , 决策属性  $\{d\}$ .

表1 基于测试代价和延迟代价的多尺度决策系统

$U$	$a_1^1$	$a_1^2$	$a_1^3$	$a_2^1$	$a_2^2$	$a_2^3$	$a_3^1$	$a_3^2$	$a_3^3$	$\omega_{x_i^k}^{a_j}$	$d$
$x_1$	A	H	Y	A	H	A	H	Y	Y	$\omega_{x_1^k}^{a_j}$	+
$x_2$	B	S	Y	C	M	B	S	Y	Y	$\omega_{x_2^k}^{a_j}$	-
$x_3$	B	S	Y	C	M	B	S	Y	Y	$\omega_{x_3^k}^{a_j}$	+
$x_4$	C	M	N	B	S	B	S	Y	Y	$\omega_{x_4^k}^{a_j}$	+
$x_5$	E	L	N	B	S	B	S	Y	Y	$\omega_{x_5^k}^{a_j}$	-
$x_6$	E	L	N	E	L	E	L	N	N	$\omega_{x_6^k}^{a_j}$	-
$x_7$	E	L	N	E	L	E	L	N	N	$\omega_{x_7^k}^{a_j}$	-

对于属性 $a_1$ 在不同尺度下的尺度代价分别为

$$\begin{aligned} \text{Cost}_{a_1^1} &= \eta_{a_1^1} \times 7 + (\omega_{x_2^1} + \omega_{x_3^1}), \\ \text{Cost}_{a_1^2} &= \eta_{a_1^2} \times 7 + (\omega_{x_2^2} + \omega_{x_3^2}), \\ \text{Cost}_{a_1^3} &= \eta_{a_1^3} \times 7 + (\omega_{x_1^3} + \omega_{x_2^3} + \omega_{x_3^3} + \\ &\quad \omega_{x_4^3} + \omega_{x_5^3} + \omega_{x_6^3} + \omega_{x_7^3}). \end{aligned}$$

则属性 $a_1$ 的属性代价为

$$\text{Cost}_{a_1} = \frac{\text{Cost}_{a_1^1} + \text{Cost}_{a_1^2} + \text{Cost}_{a_1^3}}{3}.$$

**定理1** 给定一个多尺度决策系统 $S^* = (U, AT \cup \{d\}, \eta, \omega)$ , 对于 $\forall a_j \in AT$ , 如果 $s_j = 1$ , 则多尺度决策表中的尺度代价 $\text{Cost}_{a_j^k}$ 退化为单尺度决策表中属性 $a_j$ 的属性代价.

**证明** 显然, 当 $s_1 = s_2 = \dots = s_m = 1$ 时, 多尺度决策表 $S^*$ 中每个属性的尺度个数均为1, 退化为一个单尺度决策表. 单尺度决策表是本文中一种特殊情况.  $\square$

**定理2** 给定一个多尺度决策系统 $S^* = (U, AT \cup \{d\}, \eta, \omega)$ , 对于 $\forall a_j \in AT$ , 如果 $\text{Cost}_{a_j^1} = \text{Cost}_{a_j^2} = \dots = \text{Cost}_{a_j^{s_j}}$ , 则属性 $a_j$ 的属性代价 $\text{Cost}_{a_j}$ 等于其尺度代价值.

**证明** 对于任意一个属性 $a_j$ , 其属性代价为 $\text{Cost}_{a_j} = \frac{1}{s_j} \sum_{k=1}^{s_j} \text{Cost}_{a_j^k} = (\text{Cost}_{a_j^1} + \text{Cost}_{a_j^2} + \dots + \text{Cost}_{a_j^{s_j}}) / s_j$ , 因为 $\text{Cost}_{a_j^1} = \text{Cost}_{a_j^2} = \dots = \text{Cost}_{a_j^{s_j}}$ , 所以 $\text{Cost}_{a_j} = \frac{s_j \times \text{Cost}_{a_j^1}}{s_j} = \frac{s_j \times \text{Cost}_{a_j^2}}{s_j} = \dots = \frac{s_j \times \text{Cost}_{a_j^{s_j}}}{s_j}$ . 因此,  $\text{Cost}_{a_j} = \text{Cost}_{a_j^1} = \text{Cost}_{a_j^2} = \dots = \text{Cost}_{a_j^{s_j}}$ .  $\square$

在多尺度决策系统, 属性重要度是根据尺度重要度计算得到, 反映的是属性的区分能力; 属性代价是根据尺度代价计算得到, 反映的是属性所产生的代价. 在实际应用中, 人们往往希望做决策时将代价考虑在内, 而属性区分能力和属性代价变化趋势相反且为两个不同的量纲单位, 不能直接进行计算, 为了消除量纲的影响, 分别对属性重要度和属性代价进行归一化处理, 分别为 $p_{\text{sig}(a_j, AT, d)}$ 和 $p_{\text{Cost}_{a_j}}$ , 表示如下:

$$p_{\text{sig}(a_j, AT, d)} = \frac{\text{sig}(a_j, AT, d)}{\sum_{j=1}^m \text{sig}(a_j, AT, d)}, \quad (10)$$

$$p_{\text{Cost}_{a_j}} = 1 - \frac{\text{Cost}_{a_j}}{\sum_{j=1}^m \text{Cost}_{a_j}}. \quad (11)$$

将属性重要度与属性代价归一化后, 综合属性区分能力和属性代价两个因素, 提出加权属性重要度进

行属性排序.

**定义9** (加权属性重要度) 给定一个多尺度决策系统 $S^* = (U, AT \cup \{d\}, \eta, \omega)$ , 对于 $\forall a_j \in AT$ , 加权属性重要度被定义为

$$\text{Sig}(a_j, AT, d) = \alpha p_{\text{sig}(a_j, AT, d)} + (1 - \alpha) p_{\text{Cost}_{a_j}}, \quad (12)$$

其中 $\alpha$ 是属性重要度和属性代价的调节系数.

值得注意的是, 当属性重要度为0时, 该属性是冗余属性, 不对该属性选择最优尺度. 本节提出的尺度代价和属性代价刻画了多尺度决策系统中尺度所产生的代价与属性所产生的代价之间的关系, 并且提出了结合属性重要度和属性代价两个因素的加权属性重要度进行属性排序的方法. 接下来将基于属性排序序列, 建立一个代价敏感的最优尺度组合选择模型.

### 3 代价敏感的最优尺度组合选择模型

在多尺度决策系统中, 基于属性重要度进行最优尺度组合时, 是利用属性的尺度重要度来刻画属性重要度, 通过属性重要度对属性排序, 然后依次对每个属性进行一致性判断来选择最优尺度, 直到所有的属性均被考虑, 从而得到最优尺度组合. 而在实际应用中, 属性的不同尺度不仅具有不同的区分能力, 而且不同尺度所产生的代价也不相同, 因此最优尺度组合选择有必要考虑代价信息的影响.

在第2节中提出了基于属性重要度和属性代价的加权属性重要度进行属性排序, 所得到的属性排序序列是综合考虑属性区分能力和属性代价后的排序, 加权属性重要度越大, 表示属性越重要, 越应该被优先考虑. 进一步地, 由于每个属性具有多个不同的尺度, 需要依次对每个属性进行最优尺度选择. 因此, 本节定义了基于尺度重要度和尺度代价的加权尺度重要度来综合考虑尺度区分能力和尺度代价, 然后建立一个代价敏感的最优尺度组合选择模型. 假设属性排序序列为 $\tau = \{a_{j_1}, a_{j_2}, \dots, a_{j_{t-1}}, a_{j_t}, a_{j_{t+1}}, \dots, a_{j_m}\}$ ,  $a_{j_t}$ 表示第 $t$ 个被选择的属性,  $K_{j_t, k}$ 表示属性 $a_j$ 的尺度选择为 $k$ ,  $K^t$ 表示第 $t$ 个属性进行最优尺度选择后的尺度组合结果.

**定义10** (加权尺度重要度) 给定一个多尺度决策系统 $S^* = (U, AT \cup \{d\}, \eta, \omega)$ , 对于 $\forall a_j \in AT$ , 加权尺度重要度被定义为

$$\text{Sig}_k(a_{j_t}, AT, d) = \beta p_{\text{sig}_k(a_{j_t}, AT, d)} + (1 - \beta) p_{\text{Cost}_{a_{j_t}}}. \quad (13)$$

其中:  $p_{\text{sig}_k(a_{j_t}, AT, d)}$ 和 $p_{\text{Cost}_{a_{j_t}}}$ 分别表示 $\text{sig}_k(a_{j_t}, AT, d)$ 和 $\text{Cost}_{a_{j_t}}$ 的归一化结果, 归一化公式同式(10)

和(11);  $\text{sig}_k(a_{j_t}, \text{AT}, d)$  表示属性  $a_{j_t}$  在第  $k$  尺度下的尺度重要度;  $\text{Cost}_{a_{j_t}^k} = |U| \times \eta_{a_{j_t}^k} + \sum_{x_i \in \text{BND}_{\text{AT}^{K^t}}(X)} w_{x_i}$ ,  $\text{AT}^{K^t}$  表示在第  $t$  个属性被选择时当前尺度组合  $K^t$  的属性子集;  $\beta$  表示属性重要度和属性代价的调节系数.

**定义 11**(代价敏感的最优尺度组合选择模型) 给定一个多尺度决策系统  $S^* = (U, \text{AT} \cup \{d\}, \eta, \omega)$ , 假设属性排序序列为  $\tau = \{a_{j_1}, a_{j_2}, \dots, a_{j_m}\}$ , 起始尺度组合为  $K^0 = (0; 0; \dots; 0)$ , 则每个属性选择最优尺度的目标函数为

$$\max \text{Sig}_k(a_{j_t}, \text{AT}, d), \quad (14)$$

其中  $k$  是模型的尺度变量.

针对以上最优尺度组合选择模型, 根据加权尺度重要度对属性依次进行最优尺度选择. 加权尺度重要度由两个因素组成: 尺度重要度和尺度代价. 值得注意的是, 尺度重要度就是尺度的区分能力, 公式如定义(3)所示, 而尺度代价表示前一个属性选择最优尺度后, 当前属性加入时所产生的尺度代价. 比如在已知属性序列为  $\tau = \{a_{j_1}, a_{j_2}, \dots, a_{j_m}\}$  的情况下, 假设第 1 个属性  $a_{j_1}$  的最优尺度为  $l_1$ , 则对第 1 个属性  $a_{j_1}$  进行最优尺度选择后, 尺度组合为  $K^1 = (0; 0; \dots; l_1; \dots; 0)$ . 对第 2 个属性进行最优尺度选择时, 则是在尺度组合  $K^1 = (0; 0; \dots; l_1; \dots; 0)$  的基础之上得到  $K^2 = (0; 0; \dots; l_1; \dots; l_2; \dots; 0)$ , 按照属性序列顺序, 依次得到各个属性最优尺度选择结果, 从而得到最优尺度组合.

**定义 12**(最优尺度组合) 给定一个多尺度决策系统  $S^* = (U, \text{AT} \cup \{d\}, \eta, \omega)$ . 其中:  $U = \{x_1, x_2, \dots, x_n\}$ ,  $\text{AT} = \{a_1, a_2, \dots, a_m\}$ . 按照属性排序序列为  $\tau = \{a_{j_1}, a_{j_2}, \dots, a_{j_m}\}$ , 对于  $a_{j_t} (t = 1, 2, \dots, m - 1)$  分别被选择为第  $l_j$  尺度, 得到多尺度决策系统  $S^*$  的第  $t - 1$  步最优尺度组合  $K^{t-1}$ ,  $K^m$  是多尺度决策表  $S^*$  通过属性序列  $\tau = \{a_{j_1}, a_{j_2}, \dots, a_{j_m}\}$  所得到的最优尺度组合.

在以上所定义的最优尺度组合中, 对于属性排序序列  $\tau = \{a_{j_1}, a_{j_2}, \dots, a_{j_m}\}$ , 在属性进行最优尺度选择时, 同时考虑了当前尺度的区分能力以及代价来逐步对每个属性选择最优尺度. 对于一个多尺度决策系统  $S^*$ , 代价敏感的最优尺度组合选择算法步骤如下所示, 具体流程如图 1 所示.

**算法 1** 代价敏感的最优尺度组合选择算法.

输入:  $S^* = (U, \text{AT} \cup \{d\}, \eta, \omega)$ , 目标概念  $X$ 、 $\alpha$  以及  $\beta$ ;

输出: 最优尺度组合  $K^m$ .

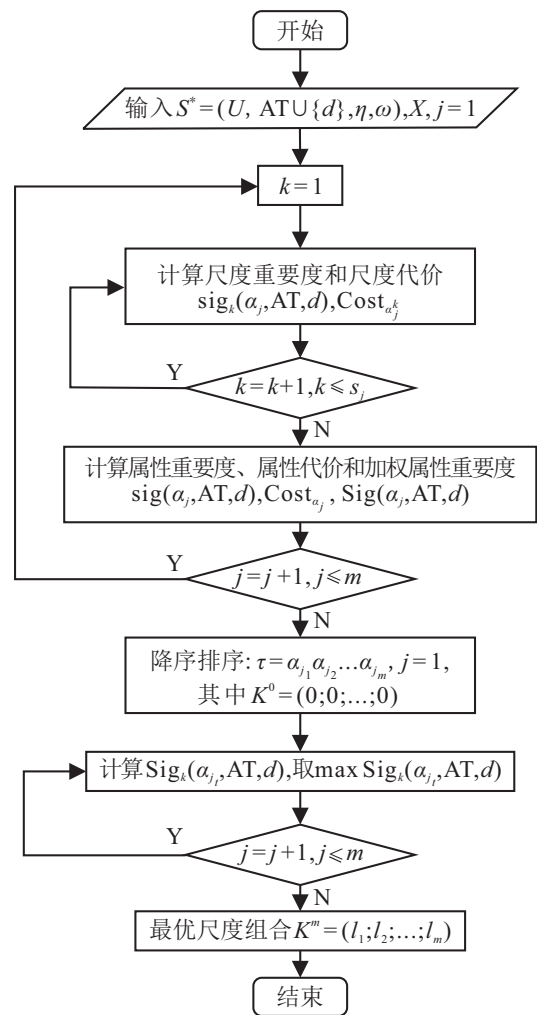


图 1 代价敏感的最优尺度组合选择流程

- step 1: 初始化  $\tau = \{\}$ ,  $j = 1$ ,  $K^0 = (0; 0; \dots; 0)$ ;
- step 2: 判断 AT 是否为空, 如果为空, 则执行 step 6, 否则执行 step 3;
- step 3: for  $j = 1 : |\text{AT}|$   
 for  $k = 1 : s_j$   
 计算  $\text{sig}_k(a_j, \text{AT}, d), \text{Cost}_{a_j^k}$ ;  
 end for  
 end for
- step 4: for  $j = 1 : |\text{AT}|$   
 计算  $\text{sig}(a_j, \text{AT}, d), \text{Cost}_{a_j}, \text{Sig}(a_j, \text{AT}, d)$ ;  
 end for
- step 5: 得到属性序列  $\tau = \{a_{j_1}, a_{j_2}, \dots, a_{j_m}\}$ ;
- step 6: for  $j = 1 : |\text{AT}|$   
 计算  $\text{Sig}_k(a_{j_t}, \text{AT}, d)$ ,  
 取  $\max \text{Sig}_k(a_{j_t}, \text{AT}, d)$ ;  
 end for
- step 7: 得到最优尺度组合  $K^m$ .

下面通过一个实例来说明本算法的计算过程, 具体如下所示.

**例 2** 给定一个多尺度决策系统  $S^* = (U, \text{AT} \cup$

$\{d\}, \eta, \omega)$ , 如表2所示. 其中:  $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}\}$ ,  $U$ 在决策属性 $d$ 划分为 $D_1 = \{x_1, x_2, x_3, x_4, x_5, x_7, x_8\}$ 和 $D_2 = \{x_6, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}\}$ ;  $\eta = \{3, 2, 1, 4, 2, 5, 3, 1\}$ ; 目标概念 $X = D_1$ . 为了便于计算, 假设 $\omega_{a_1^1} = \omega_{a_1^2} = \dots = \omega_{a_m^{s_j}}$ .

表2 基于测试代价和延迟代价的多尺度决策系统

$U$	$a_1^1$	$a_1^2$	$a_1^3$	$a_2^1$	$a_2^2$	$a_3^1$	$a_3^2$	$a_3^3$	$\omega_{a_j^k}$	$d$
$x_1$	1	A	Y	I	Y	2	A	Y	1	+
$x_2$	1	A	Y	I	Y	1	A	Y	4	+
$x_3$	1	A	Y	I	Y	3	B	N	5	+
$x_4$	2	A	Y	III	Y	6	C	N	3	+
$x_5$	2	A	Y	II	Y	7	D	Y	3	+
$x_6$	4	B	Y	VI	N	4	B	Y	2	-
$x_7$	3	B	Y	VI	N	3	B	Y	6	+
$x_8$	6	C	N	III	Y	6	C	Y	5	+
$x_9$	7	D	N	III	Y	8	D	Y	4	-
$x_{10}$	3	A	Y	III	Y	6	C	N	5	-
$x_{11}$	2	A	Y	II	Y	8	D	N	3	-
$x_{12}$	9	E	N	VI	N	9	E	N	2	-
$x_{13}$	9	E	N	VI	N	9	E	N	1	-
$x_{14}$	9	E	N	III	Y	6	B	Y	2	-

最优尺度组合选择的详细过程如下.

1) 通过计算尺度重要度, 得到属性重要度.

对于属性 $a_1$ , 不同尺度下的尺度重要度分别为

$$\begin{aligned} \text{sig}_1(a_1, AT, d) &= \frac{7}{14} - \frac{5}{14} = \frac{2}{14}, \\ \text{sig}_2(a_1, AT, d) &= \frac{1}{14}, \\ \text{sig}_3(a_1, AT, d) &= \frac{1}{14}. \end{aligned}$$

对于属性 $a_2$ , 不同尺度下的尺度重要度分别为

$$\text{sig}_1(a_2, AT, d) = 0, \text{sig}_2(a_2, AT, d) = 0.$$

对于属性 $a_3$ , 不同尺度下的尺度重要度分别为

$$\begin{aligned} \text{sig}_1(a_3, AT, d) &= \frac{1}{14}, \\ \text{sig}_2(a_3, AT, d) &= \frac{1}{14}, \\ \text{sig}_3(a_3, AT, d) &= \frac{1}{14}. \end{aligned}$$

根据定义4, 得到各个属性的属性重要度为

$$\begin{aligned} \text{sig}(a_1, AT, d) &= \frac{2}{21}, \\ \text{sig}(a_2, AT, d) &= 0, \\ \text{sig}(a_3, AT, d) &= \frac{1}{14}. \end{aligned}$$

2) 通过计算尺度代价, 得到属性代价.

对于属性 $a_1$ , 不同尺度下的尺度代价分别为

$$\begin{aligned} \text{Cost}_{a_1^1} &= 14 \times 3 + 20 = 62, \\ \text{Cost}_{a_1^2} &= 60, \text{Cost}_{a_1^3} = 60; \end{aligned}$$

对于属性 $a_2$ , 不同尺度下的尺度代价分别为 $\text{Cost}_{a_2^1} = 92, \text{Cost}_{a_2^2} = 74$ ; 对于属性 $a_3$ , 不同尺度

下的尺度代价分别为 $\text{Cost}_{a_3^1} = 83, \text{Cost}_{a_3^2} = 80, \text{Cost}_{a_3^3} = 60$ . 根据定义8, 得到各个属性的属性代价为

$$\text{Cost}_{a_1} = 60.67, \text{Cost}_{a_2} = 81.5, \text{Cost}_{a_3} = 74.3.$$

3) 基于属性重要度和属性代价得到加权重要度对属性排序.

利用归一化公式(10)和(11)对属性重要度和属性代价进行归一化处理, 结果分别为

$$\begin{aligned} p_{\text{sig}(a_1, AT, d)} &= \frac{4}{7}, p_{\text{sig}(a_2, AT, d)} = 0, p_{\text{sig}(a_3, AT, d)} = \frac{3}{7}; \\ p_{\text{Cost}_{a_1}} &= 0.720, p_{\text{Cost}_{a_2}} = 0.624, p_{\text{Cost}_{a_3}} = 0.657. \end{aligned}$$

为了计算方便, 本文取 $\alpha = 0.5$ , 根据定义9, 加权属性重要度分别为 $\text{Sig}(a_1, AT, d) = 0.646, \text{Sig}(a_2, AT, d) = 0.312, \text{Sig}(a_3, AT, d) = 0.459$ . 因此, 根据加权属性重要度得到一个属性选择顺序 $a_1, a_3, a_2$ , 即 $\tau = \{a_{1_1}, a_{3_2}, a_{2_3}\}$ .

4) 根据属性序列 $\tau = \{a_{1_1}, a_{3_2}, a_{2_3}\}$ , 依次对每个属性进行最优尺度选择, 本文取 $\beta = 0.5$ .

对于属性 $a_1$ , 各个尺度重要度分别为

$$\begin{aligned} \text{sig}_1(a_{1_1}, AT, d) &= \frac{2}{14}, \\ \text{sig}_2(a_{1_1}, AT, d) &= \frac{1}{14}, \\ \text{sig}_3(a_{1_1}, AT, d) &= \frac{1}{14}, \end{aligned}$$

可得归一化结果为

$$\begin{aligned} p_{\text{sig}_1}(a_{1_1}, AT, d) &= \frac{1}{2}, \\ p_{\text{sig}_2}(a_{1_1}, AT, d) &= \frac{1}{4}, \\ p_{\text{sig}_3}(a_{1_1}, AT, d) &= \frac{1}{4}. \end{aligned}$$

各个尺度代价分别为

$$\text{Cost}_{a_{1_1}^1} = 62, \text{Cost}_{a_{1_1}^2} = 60, \text{Cost}_{a_{1_1}^3} = 60,$$

可得归一化结果为

$$p_{\text{Cost}_{a_{1_1}^1}} = 0.341, p_{\text{Cost}_{a_{1_1}^2}} = 0.329, p_{\text{Cost}_{a_{1_1}^3}} = 0.329,$$

则

$$\begin{aligned} \text{Sig}_1(a_{1_1}, AT, d) &= 0.421, \\ \text{Sig}_2(a_{1_1}, AT, d) &= 0.230, \\ \text{Sig}_3(a_{1_1}, AT, d) &= 0.230. \end{aligned}$$

可以得出, 对于属性 $a_1$ , 第1尺度是最优尺度.

对于属性 $a_3$ , 各个尺度重要度分别为

$$\begin{aligned} \text{sig}_1(a_{3_2}, AT, d) &= \frac{1}{14}, \\ \text{sig}_2(a_{3_2}, AT, d) &= \frac{1}{14}, \\ \text{sig}_3(a_{3_2}, AT, d) &= \frac{1}{14}, \end{aligned}$$

可得归一化结果为

$$\begin{aligned}
 p_{\text{sig}_1}(a_{3_2}, AT, d) &= 0.333, \\
 p_{\text{sig}_2}(a_{3_2}, AT, d) &= 0.333, \\
 p_{\text{sig}_3}(a_{3_2}, AT, d) &= 0.333.
 \end{aligned}$$

各个尺度代价分别为

$$\text{Cost}_{a_{3_2}^1} = 112, \text{Cost}_{a_{3_2}^2} = 90, \text{Cost}_{a_{3_2}^3} = 62,$$

可得归一化结果为

$$p_{\text{Cost}_{a_{3_2}^1}} = 0.576, p_{\text{Cost}_{a_{3_2}^2}} = 0.659, p_{\text{Cost}_{a_{3_2}^3}} = 0.765,$$

则

$$\begin{aligned}
 \text{Sig}_1(a_{3_2}, AT, d) &= 0.455, \\
 \text{Sig}_2(a_{3_2}, AT, d) &= 0.496, \\
 \text{Sig}_3(a_{3_2}, AT, d) &= 0.549.
 \end{aligned}$$

可以看出,对于属性  $a_3$ ,第 3 尺度是最优尺度.

对于属性  $a_2$ ,属性重要度即为 0,因此可以判定属性  $a_2$  为冗余属性.

综上所述,最优尺度组合  $K^3$  为 (1; 0; 3).

从例 2 可以看出,从代价敏感的角度出发,可以在给定代价认知场景下,得到综合属性重要度和代价的属性排序序列. 根据这个属性序列,考虑尺度的区分能力和代价,对每个属性逐步选择最优尺度,从而得到最优尺度组合的结果.

### 4 实验及分析

为了验证所提模型的有效性和实用性,本文从 UCI 数据库中随机选取 6 个数据集,并且对每个数据集在给定代价环境下进行实验. 数据集的详细信息如表 3 所示,多尺度决策系统中每个属性的属性值在不同尺度下具有不同的属性值,此处将单尺度决策表转化为多尺度决策表的方法借鉴文献 [28]. 特别地,为了消除数据一致性的影响,本文针对数据集的属性随机删除部分条件属性. 实验环境为: Windows 10, Intel Core (TM) 15-4590 CPU (3.30 GHz), 8 GB RAM, 编程语言为 Matlab 2014 b.

表 3 数据集信息表

序号	数据集	数目	条件属性	随机删除属性
1	Haes-Roth	132	3	name, hobby
2	Breast Cancer	286	5	menopause, nodecaps, breast, irradiat
3	Balance Scale	625	3	Right-Distance
4	Car Evaluation	1 728	4	Persons, lug-boot
5	Nursery	12 960	5	Finance, social, health
6	Chess	28 056	5	White King rank

表 4 是 6 个数据集运用上述算法,通过加权属性重要度得到的属性排序序列. 在进行实验时,由于数据集中没有给定对象的延迟代价以及属性的测试代价,对于对象的延迟代价,假设每一个对象的延迟代

价不一定相等,将测试代价参数  $\eta$  和延迟代价参数  $\omega$  均取一定范围内的随机数,在本次实验过程中,假设  $\omega_{a_1^1} = \omega_{a_1^2} = \dots = \omega_{a_m^{s_j}}$ . 实验过程中,将选取数据集的某一类数据作为目标概念.

表 4 基于加权属性重要度的属性排序序列

数据集	$\eta$	$\omega$	属性选择顺序
Haes-Roth	(1,10)	(1,10)	$(a_3, a_1, a_2)$
Breast Cancer	(1,20)	(1,20)	$(a_4, a_1, a_3, a_5, a_2)$
Balance Scale	(1,20)	(1,30)	$(a_2, a_1, a_3)$
Car Evaluation	(1,10)	(1,20)	$(a_3, a_4, a_1, a_2)$
Nursery	(1,30)	(1,30)	$(a_5, a_4, a_1, a_3, a_2)$
Chess	(1,20)	(1,30)	$(a_3, a_4, a_5, a_1, a_2)$

根据加权属性重要度得到多尺度决策表中的属性排序序列,再按照属性序列依次利用加权尺度重要度对每个属性进行最优尺度选择,直到所有的属性均被选择,得到最优尺度  $K^m$ . 表 5 ~ 表 10 以及图 2 ~ 图 7 分别是 6 个数据集在给定代价场景下的最优尺度组合选择结果.

表 5 Hayes-Roth 实验结果

属性	尺度 1	尺度 2	尺度 3	尺度 4
$a_1$	0.502	0.506	<b>0.508</b>	0.486
$a_2$	<b>0.359</b>	0.335	0.307	/
$a_3$	<b>0.378</b>	0.347	0.274	/

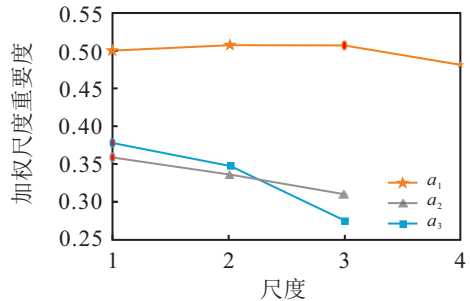


图 2 Hayes-Roth 实验结果

表 6 Breast Cancer 实验结果

属性	尺度 1	尺度 2	尺度 3	尺度 4	尺度 5	尺度 6
$a_1$	0.402	<b>0.431</b>	0.410	0.373	/	/
$a_2$	0.263	<b>0.276</b>	0.221	/	/	/
$a_3$	<b>0.362</b>	0.332	0.354	0.298	/	/
$a_4$	0.328	<b>0.389</b>	0.365	0.297	0.322	0.358
$a_5$	0.372	0.328	<b>0.435</b>	0.429	0.334	/

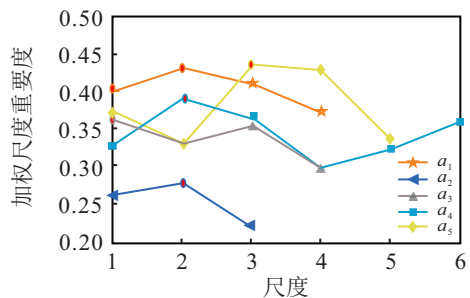


图 3 Breast Cancer 实验结果

表7 Balance Scale实验结果

属性	尺度1	尺度2	尺度3	尺度4	尺度5	尺度6
$a_1$	0.499	<b>0.532</b>	0.521	0.501	/	/
$a_2$	<b>0.482</b>	0.478	0.481	0.476	0.474	0.423
$a_3$	0.395	<b>0.451</b>	0.442	/	/	/

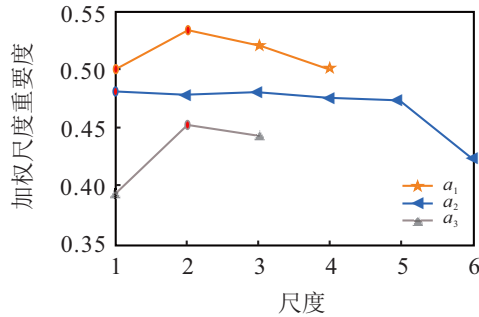


图4 Balance Scale实验结果

表8 Car Evaluation实验结果

属性	尺度1	尺度2	尺度3	尺度4	尺度5
$a_1$	0.423	<b>0.465</b>	0.449	0.455	0.446
$a_2$	0.418	0.415	<b>0.420</b>	/	/
$a_3$	<b>0.477</b>	0.462	0.423	0.456	/
$a_4$	0.501	<b>0.539</b>	0.535	0.552	/

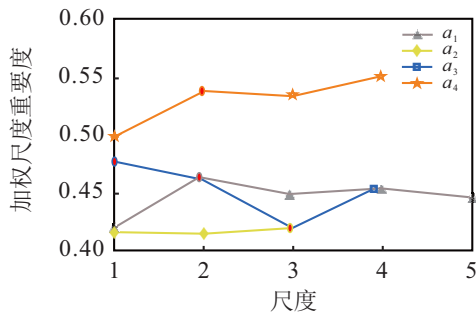


图5 Car Evaluation实验结果

表9 Nursery实验结果

属性	尺度1	尺度2	尺度3	尺度4	尺度5	尺度6
$a_1$	0.486	<b>0.513</b>	0.489	0.509	/	/
$a_2$	0.367	0.370	<b>0.393</b>	0.376	/	/
$a_3$	0.428	0.417	<b>0.432</b>	0.429	/	/
$a_4$	<b>0.382</b>	0.372	0.321	0.327	0.315	/
$a_5$	0.354	<b>0.374</b>	0.335	0.369	0.321	0.366

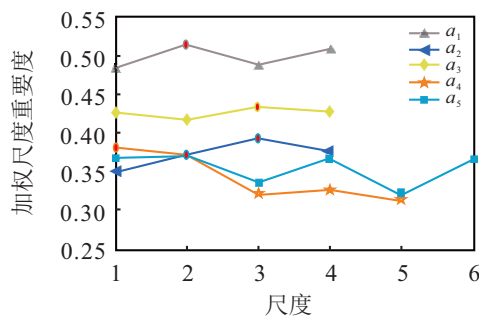


图6 Nursery实验结果

表10 Chess实验结果

属性	尺度1	尺度2	尺度3	尺度4	尺度5	尺度6
$a_1$	0.455	<b>0.460</b>	0.451	0.459	0.465	/
$a_2$	0.516	0.521	<b>0.525</b>	/	/	/
$a_3$	<b>0.517</b>	0.503	0.502	0.479	0.484	/
$a_4$	<b>0.634</b>	0.619	0.632	0.611	/	/
$a_5$	0.428	0.439	0.447	<b>0.452</b>	0.412	0.416

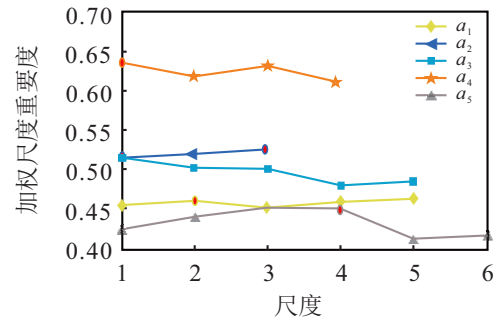


图7 Chess实验结果

利用代价敏感的最优尺度组合选择算法,在多尺度决策系统中进行最优尺度组合选择时,综合考虑了属性重要度和属性代价两个因素来对属性进行选择,以及对属性的尺度进行选择时,综合考虑了尺度重要度和尺度代价这两个因素.具体地,针对Breast Cancer数据集,实验结果如表6和图3所示,其中测试代价参数 $\eta$ 和延迟代价参数 $\omega$ 均为(1, 20)的随机数.首先根据基于属性重要度和属性代价的加权属性重要度进行属性排序,加权属性重要度越大,说明属性越重要;然后根据属性序列依次选择每个属性的最优尺度,直到所有的属性均被选择,得到最优尺度组合.从实验结果可以看出,属性的排序序列为( $a_4, a_1, a_3, a_5, a_2$ ),根据加权尺度重要度,对每个属性进行最优尺度选择,逐步得到属性的最优尺度依次为 $2 \rightarrow 2 \rightarrow 1 \rightarrow 3 \rightarrow 2$ .因此在给定代价场景下, Breast Cancer数据集的最优尺度组合为 $K^5 = (2; 2; 1; 3; 2)$ .

表11是以上6个数据集的实验结果,从实验结果可以看出,在给定代价场景下,可以在多尺度决策系统中得到一个最优尺度组合.属性区分能力强的情况下,属性代价越低的属性越应该被优先考虑;尺度区分能力强的情况下,尺度代价越低的尺度越应该被选择,这与实际应用场景相吻合.本文所提出的代价敏感的最优尺度组合选择模型,在给定代价场景下得到做决策的最优尺度组合,所得到的尺度组合结果不仅考虑了属性的区分能力、代价信息,还考虑了尺度的区分能力、代价,更加符合人类认知以及实际的应用需求.

表 11 最优尺度组合选择结果

数据集	尺度	属性选择顺序	尺度组合
Haes-Roth	$4 \times 3 \times 3$	$(a_3, a_1, a_2)$	(3; 1; 1)
Breast Cancer	$4 \times 3 \times 4 \times 6 \times 5$	$(a_4, a_1, a_3, a_5, a_2)$	(2; 2; 1; 2; 3)
Balance Scale	$4 \times 6 \times 3$	$(a_2, a_1, a_3)$	(2; 1; 2)
Car Evaluation	$5 \times 3 \times 4 \times 4$	$(a_3, a_4, a_1, a_2)$	(2; 3; 1; 2)
Nursery	$4 \times 4 \times 4 \times 5 \times 6$	$(a_5, a_4, a_1, a_3, a_2)$	(2; 3; 3; 1; 2)
Chess	$5 \times 3 \times 5 \times 4 \times 6$	$(a_3, a_4, a_5, a_1, a_2)$	(2; 3; 1; 1; 4)

### 5 结论

多尺度决策系统中的最优尺度组合是当前研究的一个热点,属性在不同尺度下测量具有不同的属性值,如何在决策过程中选择属性的最优尺度是值得研究的.实际应用中,代价信息是现实世界客观存在的,在问题求解中决策者往往要充分考虑到代价这一重要因素.因此,本文首先将代价敏感引入到多尺度决策系统中,综合属性重要度和属性代价两个因素进行属性排序.然后,根据属性排序序列,综合考虑尺度重要度和尺度代价,提出一种最优尺度选择方法依次对每个属性进行最优尺度选择,避免从单一角度进行最优尺度组合选择,能满足用户决策时对于代价的需求.最后,利用 UCI 数据集对所提算法进行验证,实验结果表明了该算法的有效性和实用性.与现有的最优尺度组合选择方法相比,本文的工作有以下优势,主要包括: 1) 将代价敏感引入多尺度决策系统中,使属性选择序列不仅取决于属性的区分能力,还取决于属性的代价信息; 2) 对每个属性进行最优尺度选择时,同时考虑尺度重要度和尺度代价两个因素.本文所提出的代价敏感的最优尺度组合选择模型考虑了实际决策过程中的代价信息,得到了一个综合考虑区分能力和代价的最优尺度组合.在给定价场场景下,能合理地选择出最优尺度组合,所得到的尺度组合结果更符合现实生活的实际需求.

### 参考文献(References)

[1] Bargiela A, Pedrycz W. Granular computing: An introduction[M]. Boston: Kluwer Academic Publishers, 2002: 1-18.

[2] Bargiela A, Pedrycz W. Toward a theory of granular computing for human-centered information processing[J]. IEEE Transactions on Fuzzy Systems, 2008, 16(2): 320-330.

[3] Pawlak Z. Rough sets[J]. International Journal of Computer & Information Sciences, 1982, 11(5): 341-356.

[4] Wei L L, Zhang W X. Probabilistic rough sets characterized by fuzzy sets[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems,

2004,12(1): 47-60.

[5] Wong S K M, Ziarko W. Comparison of the probabilistic approximate classification and the fuzzy set model[J]. Fuzzy Sets and Systems, 1987, 21(3): 357-362.

[6] Wu W Z. Upper and lower probabilities of fuzzy events induced by a fuzzy set-valued mapping[C]. Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing. Berlin: Springer Heidelberg, 2005: 345-353.

[7] Yao Y Y, Zhao Y. Attribute reduction in decision-theoretic rough set models[J]. Information Sciences, 2008, 178(17): 3356-3373.

[8] Hu B Q. Three-way decisions space and three-way decisions[J]. Information Sciences, 2014, 281: 21-52.

[9] Hu B Q. Three-way decision spaces based on partially ordered sets and three-way decisions based on hesitant fuzzy sets[J]. Knowledge-Based Systems, 2016, 91: 16-31.

[10] Hu B Q, Wong H, Yiu K F C. The aggregation of multiple three-way decision spaces[J]. Knowledge-Based Systems, 2016, 98: 241-249.

[11] 张清华, 胡荣德, 姚龙洋, 等. 基于属性重要度的风险决策粗糙集属性约简[J]. 控制与决策, 2016, 31(7): 1199-1205.  
(Zhang Q H, Hu R D, Yao L Y, et al. Risk DTRS attribute reduction based on attribute importance[J]. Control and Decision, 2016, 31(7): 1199-1205.)

[12] Zhang Q H, Yang J J, Yao L Y. Attribute reduction based on rough approximation set in algebra and information views[J]. IEEE Access, 2016, 4: 5399-5407.

[13] 李兵洋, 肖健梅, 王锡淮. 基于多目标邻域差分进化和模糊粗糙集的属性约简算法[J]. 控制与决策, 2019, 34(5): 947-955.  
(Li B Y, Xiao J M, Wang X H. Attribute reduction with fuzzy rough set based on multiobjective neighborhood difference algorithm[J]. Control and Decision, 2019, 34(5): 947-955.)

[14] 叶明全, 胡学钢, 胡东辉, 等. 基于属性值分类的多层次粗糙集模型[J]. 模式识别与人工智能, 2013, 25(5): 481-491.  
(Ye M Q, Hu X G, Hu D H, et al. A multi-level rough set model based on attribute value taxonomies[J]. Pattern Recognition and Artificial Intelligence, 2013, 25(5): 481-491.)

[15] Zhang Y Q. Constructing knowledge base of emergency based on hierarchical rough set theory[J]. Procedia Computer Science, 2018, 131: 699-705.

[16] Wu W Z, Leung Y. Theory and applications of granular labelled partitions in multi-scale decision tables[J]. Information Sciences, 2011, 181(18): 3878-3897.

[17] Wu W Z, Leung Y. Optimal scale selection for multi-scale decision tables[J]. International Journal of Approximate Reasoning, 2013, 54(8): 1107-1129.

[18] 吴伟志, 高仓健, 李同军. 序粒度标记结构及其粗糙近似[J]. 计算机研究与发展, 2014, 51(12): 2623-2632.  
(Wu W Z, Gao C J, Li T J. Ordered granular

- labeled structures and rough approximations[J]. *Journal of Computer Research and Development*, 2014, 51(12): 2623-2632.)
- [19] 吴伟志, 陈颖, 徐优红, 等. 协调的不完备多粒度标记决策系统的最优粒度选择[J]. *模式识别与人工智能*, 2016, 29(2): 108-115.  
(Wu W Z, Chen Y, Xu Y H, et al. Optimal granularity selections in consistent incomplete multi-granular labeled decision systems[J]. *Pattern Recognition and Artificial Intelligence*, 2016, 29(2): 108-115.)
- [20] 吴伟志, 陈超君, 李同军, 等. 不协调多粒度标记决策系统最优粒度的对比[J]. *模式识别与人工智能*, 2016, 29(12): 1095-1103.  
(Wu W Z, Chen C J, Li T J, et al. Comparative study on optimal granularities in inconsistent multi-granular labeled decision systems[J]. *Pattern Recognition and Artificial Intelligence*, 2016, 29(12): 1095-1103.)
- [21] Wu W Z, Qian Y H, Li T J, et al. On rule acquisition in incomplete multi-scale decision tables[J]. *Information Sciences*, 2017, 378: 282-302.
- [22] Gu S M, Wu W Z. On knowledge acquisition in multi-scale decision systems[J]. *International Journal of Machine Learning and Cybernetics*, 2013, 4(5): 477-486.
- [23] Gu S M, Wu W Z. Knowledge acquisition in inconsistent multi-scale decision systems[M]. *Rough Sets and Knowledge Technology*. Berlin: Springer, 2011: 669-678.
- [24] 顾沈明, 顾金燕, 吴伟志, 等. 不完备多粒度决策系统的局部最优粒度选择[J]. *计算机研究与发展*, 2017, 54(7): 1500-1509.  
(Gu S M, Gu J Y, Wu W Z, et al. Local optimal granularity selections in incomplete multi-granular decision systems[J]. *Journal of Computer Research and Development*, 2017, 54(7): 1500-1509.)
- [25] She Y H, Li J H, Yang H L. A local approach to rule induction in multi-scale decision tables[J]. *Knowledge-Based Systems*, 2015, 89: 398-410.
- [26] Xie J P, Yang M H, Li J H, et al. Rule acquisition and optimal scale selection in multi-scale formal decision contexts and their applications to smart city[J]. *Future Generation Computer Systems*, 2018, 83: 564-581.
- [27] Hao C, Li J H, Fan M, et al. Optimal scale selection in dynamic multi-scale decision tables based on sequential three-way decisions[J]. *Information Sciences*, 2017, 415/416: 213-232.
- [28] Li F, Hu B Q. A new approach of optimal scale selection to multi-scale decision tables[J]. *Information Sciences*, 2017, 381: 193-208.
- [29] Li F, Hu B Q, Wang J. Stepwise optimal scale selection for multi-scale decision tables via attribute significance[J]. *Knowledge-Based Systems*, 2017, 129: 4-16.
- [30] 顾沈明, 陆瑾璐, 吴伟志, 等. 广义多尺度决策系统的局部最优粒度选择[J]. *山东大学学报: 理学版*, 2018, 53(8): 1-8.  
(Gu S M, Lu J L, Wu W Z, et al. Local optimal granularity selections in generalized multi-scale decision systems[J]. *Journal of Shandong University: Natural Science*, 2018, 53(8): 1-8.)
- [31] 吴伟志, 庄宇斌, 谭安辉, 等. 不协调广义多尺度决策系统的尺度组合[J]. *模式识别与人工智能*, 2018, 31(6): 485-494.  
(Wu W Z, Zhuang Y B, Tan A H, et al. Scale combinations in inconsistent generalized multi-scale decision systems[J]. *Pattern Recognition and Artificial Intelligence*, 2018, 31(6): 485-494.)
- [32] Xu Y H, Wu W Z, Tan A H. Optimal scale selections in consistent generalized multi-scale decision tables[M]. *Rough Sets Cham: Springer International Publishing*, 2017: 185-198.
- [33] Zhang Y, Zhou Z H. Cost-sensitive face recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(10): 1758-1769.
- [34] 卫东, 郑恩辉, 杨敏, 等. 基于支持向量机的误分类代价敏感模糊推理系统[J]. *控制与决策*, 2010, 25(2): 191-195.  
(Wei D, Zheng E H, Yang M, et al. Misclassification cost-sensitive fuzzy inference system based on support vector machines[J]. *Control and Decision*, 2010, 25(2): 191-195.)
- [35] Liao S J, Zhu Q X, Qian Y H, et al. Multi-granularity feature selection on cost-sensitive data with measurement errors and variable costs[J]. *Knowledge-Based Systems*, 2018, 158: 25-42.
- [36] Yang J, Wang G Y, Zhang Q H, et al. Optimal granularity selection based on cost-sensitive sequential three-way decisions with rough fuzzy sets[J]. *Knowledge-Based Systems*, 2019, 163: 131-144.
- [37] Min F, Liu Q H. A hierarchical model for test-cost-sensitive decision systems[J]. *Information Sciences*, 2009, 179(14): 2442-2452.
- [38] Yang X B, Qi Y S, Song X N, et al. Test cost sensitive multigranulation rough set: Model and minimal cost selection[J]. *Information Sciences*, 2013, 250: 184-199.

### 作者简介

张清华(1974—), 男, 教授, 博士生导师, 从事不确定信息处理、粗糙集与粒计算等研究, E-mail: zhangqh@cqupt.edu.cn;

张雪秋(1993—), 女, 硕士生, 从事不确定信息处理、粗糙集与多尺度的研究, E-mail: 670790845@qq.com;

庞国弘(1994—), 男, 硕士生, 从事不确定信息处理、粗糙集与三支决策的研究, E-mail: 787149254@qq.com.

(责任编辑: 齐 霖)