

# 基于字符型属性值更新的动态三支决策模型

张清华, 吕功勋, 陈玉洪, 谢 秦  
(重庆邮电大学计算智能重庆市重点实验室, 重庆 400065)

**摘 要:** 在现有基于属性值更新的动态三支决策模型上, 本文充分考虑字符型属性对象在更新过程中属性知识内涵的不确定性以及对象间优异程度的差异, 首先定义字符型属性对象的经验值和经验综合评价价值的概念来初步刻画对象, 再用修正值来表示对象的知识内涵; 通过修正值计算出的基于欧氏距离的最优贴适度作为对象的修正综合评价价值; 然后, 给出了字符型属性对象的动态特征的提取方法, 建立了动态三支决策模型. 最后, 通过大量的仿真实验验证了模型的高效性和适用性.

**关键词:** 智能决策; 三支决策; 动态更新; 粒计算; 特征提取

**中图分类号:** TP181      **文献标识码:** A      **文章编号:** 0372-2112(2019)02-0344-07

**电子学报 URL:** <http://www.ejournal.org.cn>      **DOI:** 10.3969/j.issn.0372-2112.2019.02.013

## A Dynamic Three-Way Decision Model Based on the Updating of Character Attribute Values

ZHANG Qing-hua, LÜ Gong-xun, CHEN Yu-hong, XIE Qin

(Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

**Abstract:** According to the existing dynamic three-way decision model based on the updating of attribute values, in this paper both the uncertainty of attribute knowledge connotation and the difference of excellent degree between objects are fully considered in the process of updating the object. More concretely, both the experience value of character attribute objects and the concept on comprehensive evaluation value of experience are firstly defined to describe the object initially, and the knowledge connotation of the object is expressed by the revised value of character attribute objects. Next, the optimal closeness degree based on the Euclidean distance calculated by the revised value is used as the revised comprehensive evaluation value of the object. Then, the extraction method of dynamic feature on objects with character attributes is presented, and a dynamic three-way decision model is established. Finally, a large number of simulation experiments have been made to validate the efficiency and applicability of the proposed model.

**Key words:** intelligent decision making; three-way decisions; dynamic updating; granular computing; feature extraction

## 1 引言

三支决策<sup>[1]</sup> (Three-way decisions, 3WD) 是由加拿大里贾纳大学的 Yao 教授提出的一种处理不确定性问题的方法. 它的核心思想是将一个论域划分为三个互不相交的区域, 对每一个区域制定相应的决策策略. 三支决策理论将不承诺决策 (延迟决策) 作为信息不足以决定接受或者拒绝时的第三种决策方式.

近年来大量学者在三支决策理论模型及应用上进行了深入的研究. 如 Greco 等考虑了错分类成本, 提出

了优势关系下的三支概率模型<sup>[2]</sup>. Yang 针对目前决策粗糙集模型集中单代理的现状, 利用三支决策思想, 提出了多代理决策粗糙集模型<sup>[3]</sup>. Yao 以动态视角将概率粗糙集的三支决策推广到了一个序列, 采用粒计算的思想, 提出一种三支决策的动态决策思路<sup>[4]</sup>. 他们认为三支决策可视为一序列决策步骤中的一个重要的中间环节, 对其中不承诺选项可再做研究, 通过进一步搜集数据、分析数据而最终获得二支决策结果<sup>[5]</sup>.

在实际应用中, 许多应用场景都需要考虑经济、时间代价与概念规则的不确定性, 并在已知目标需求接

受的对象个数的条件下进行合理的动态决策. 因此, 本文选用更符合动态决策的三支决策模型来处理字符型决策信息系统.

一般地, 当决策对象处于值更新的动态变化时, 属性值更新过程有以下规则:

(1) 属性值较小(大)的对象, 发生更新的提升力度往往更大(小).

(2) 较“劣质”的对象在更新后往往无法达到“优质”对象的优秀程度.

文献[6]充分考虑更新过程中属性值的不确定性和更新成本, 提出了动态三支决策模型. 然而该模型主要解决了属性值为数值型的对象的动态决策问题. 针对字符型属性值决策问题, 尽管可以通过一些等价规则或是权重计算将字符型的属性值转化为数值型, 然后利用常规的数值型决策模型来达到分类的目的<sup>[7-9]</sup>. 但已有转化方法<sup>[10-11]</sup>均有条件限制, 且未考虑此类动态决策过程中, 由于字符型属性值的模糊性导致的错误决策问题, 并且在动态决策中无法直接反映出字符型属性值的更新变化. 因此, 如何合理利用决策过程中属性值的更新策略, 同时考虑属性之间的相互关系, 通过提取字符型属性的特征用于建立相应的动态决策模型是本文的研究重点.

## 2 基础知识

为了更准确地描述模型, 下面给出粗糙集、三支决策理论的相关定义.

**定义 1** 信息系统和等价关系<sup>[12-14]</sup>. 信息系统可以表示为  $S = (U, AT, V, f)$ . 其中,  $U = \{x_1, x_2, \dots, x_n\}$  为非空有限对象集, 称为决策域;  $AT = C \cup D$  是属性全集, 子集  $C$  和  $D$  分别称为条件属性集和决策属性集;  $V_a$  是属性  $a \in AT$  的属性值域;  $f: U \times AT \rightarrow V$  是一个信息函数. 记对象  $x$  在属性  $a$  上的属性值为  $a(x)$ . 则对于任意的条件属性集合  $A \subseteq C$ , 等价关系定义为  $IND(A) = \{(x, y) \in U^2 \mid \forall a \in A, a(x) = a(y)\}$ .

等价关系可形成  $U$  上的一个划分, 记为  $U/IND(A)$ . 给定一个对象  $x \in U$ ,  $[x]_A$  表示在条件属性  $A$  所形成的等价关系下包含  $x$  的等价类, 简记为  $[x]$ .

**定义 2** 择优差<sup>[6]</sup>. 设有数值集合  $A_{n-1} = \{a_1, a_2, \dots, a_{n-1}\}$ ,  $A_n = \{a_1, a_2, \dots, a_{n-1}, a_n\}$ , 则定义  $z_n$  为集合  $A_n$  内的元素  $a_n$  的择优差, 记作:

$$z_n = S_n - S_{n-1} = \sqrt{\frac{1}{n} \sum_{i=1}^n (a_i - \bar{A}_n)^2} - \sqrt{\frac{1}{n-1} \sum_{i=1}^{n-1} (a_i - \bar{A}_{n-1})^2}.$$

其中  $\bar{A}_{n-1} = \frac{1}{n-1} \sum_{i=1}^{n-1} a_i$ ,  $\bar{A}_n = \frac{1}{n} \sum_{i=1}^n a_i$  分别表示集合

$$A_{n-1}, A_n \text{ 的均值, } S_{n-1} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n-1} (a_i - \bar{A}_{n-1})^2}, S_n = \sqrt{\frac{1}{n} \sum_{i=1}^n (a_i - \bar{A}_n)^2} \text{ 表示 } A_{n-1}, A_n \text{ 的标准差.}$$

**定义 3** 属性比率<sup>[6]</sup>. 设  $X = \{x_1, x_2, \dots, x_n\}$  为决策域  $U$  中对象集合,  $V_a$  是属性  $a \in AT$  的属性值域, 若当且仅有唯一属性  $a$  且  $V_a$  为数值集合, 对于  $\forall x_i \in X$ , 对象  $x_i$  在条件属性  $a$  上的属性比率定义为:

$$P_a(x_i) = \begin{cases} \frac{|a(x_i)| + |a_{MIN}(X)|}{|a_{MAX}(X)| + |a_{MIN}(X)|}, & a_{MIN}(X) < 0 \\ \frac{a(x_i)}{a_{MAX}(X)}, & a_{MIN}(X) \geq 0 \end{cases}.$$

其中  $\frac{a_{MIN}(X)}{a_{MAX}(X)} \leq P_a(x_i) \leq 1$ ,  $a_{MAX}(X) = \max_{1 \leq i \leq n} \{a(x_i)\}$ ,  $a_{MIN}(X) = \min_{1 \leq i \leq n} \{a(x_i)\}$ .

**定义 4** 基于属性比率的三支决策模型<sup>[6]</sup>. 设  $S = (U, AT, V, f)$  是一个决策表,  $X \subseteq U$ ,  $0 \leq \beta < \alpha \leq 1$ , 则上下近似分别定义为:

$$\overline{apr}_{(\alpha, \beta)}(X) = \{x \in U \mid P_a(x) > \beta\}, \\ \underline{apr}_{(\alpha, \beta)}(X) = \{x \in U \mid P_a(x) > \alpha\}.$$

给定一对阈值  $(\alpha, \beta)$ , 并且满足  $0 \leq \beta < \alpha \leq 1$ , 上下近似将论域分为 3 个部分: 接受域、延迟域和拒绝域, 其定义分别为:

$$POS_{(\alpha, \beta)}(X) = \underline{apr}_{(\alpha, \beta)}(X) = \{x \in U \mid P_a(x) > \alpha\}, \\ BND_{(\alpha, \beta)}(X) = \overline{apr}_{(\alpha, \beta)}(X) - \underline{apr}_{(\alpha, \beta)}(X) = \{x \in U \mid \beta < P_a(x) \leq \alpha\}, \\ NEG_{(\alpha, \beta)}(X) = U - \overline{apr}_{(\alpha, \beta)}(X) = \{x \in U \mid P_a(x) \leq \beta\}.$$

那么, 在属性比率对对象的刻画下可以得到阈值  $\alpha = a(x_i)/a_{MAX}(X)$  和  $\beta = a(x_j)/a_{MAX}(X)$ . 其中,  $a_{MIN}(X)/a_{MAX}(X) \leq \beta < \alpha \leq 1$ ,  $x_i, x_j \in \{x_1, x_2, \dots, x_n\}$ ,  $a(x)$  为对象  $x$  在属性  $a$  上的属性取值.

## 3 字符型属性值在知识距离度量下的修正方法及特征提取

### 3.1 基于知识距离的字符型属性值修正方法

在实际的决策过程中, 若充分考虑属性之间内在的知识联系并根据原始属性本身的特征来进行分类, 得到的结果往往会更贴近理想解. 针对字符型属性值决策问题, 为了防止因属性的模糊性导致决策错误, 需要通过属性之间形成的知识结构, 尽可能准确地找出可以反映字符对象本身的知识特征, 用群属性之间关联的知识特征和值更新策略来进行决策. 因此, 本文选择属性值之间的知识距离来度量属性之间的关系, 并基于此类关系得到动态特征并利用更新策略来选择最合理的动态决策方案.

借鉴文献[6]中数值型属性的对象利用对象之间优秀程度差异的稳定性作为决策特征的思想,同时考虑到字符型属性值转为数值型时的动态更新状态以及转化方法中存在的主观性.因此,本文采取的策略是将字符型属性对象之间的优秀程度的差异用数值来表示并进行决策.基于字符型属性对象间优异程度的差异,将字符型属性值向数值转化的步骤如下:

**Step1:**将每一个字符型属性值用一个数值表示它们的评价等级.

**Step2:**根据这个相对评价等级与在知识度量下所得的属性权重,计算出每个对象的综合评价值.

**Step3:**计算每个对象的每个属性值与对象的评价值的偏差,继而可得到每个属性的平均偏差.

**Step4:**利用每一个属性的平均偏差修正每一个属性值来得到一个修正值.

由于在相同的属性权重下,每个属性值的评价等与综合评价值的差异是不变的,因此可以用修正值来代表一个接近对象真正知识内涵的值,区分对象的真实差异.

接下来给出经验值和修正值的定义.

**定义 5** 经验值和修正值. 给定一个信息系统  $S = (U, AT, V, f)$ ,  $U = \{x_1, x_2, \dots, x_n\}$  是一个有限论域,  $A = \{a_1, a_2, \dots, a_m\}$  是字符型条件属性集,  $V_A = \{a_i(x_j) \mid i = 1, 2, \dots, m; j = 1, 2, \dots, n\}$  是条件属性值的集合. 若用数值作为字符型属性值  $a_i(x_j)$  的评价等级,则该评价等级称为对象  $x_j$  属性  $a_i$  的经验值,记作  $FuzzyRank_{a_i}(x_j)$ ,简记为  $N$ . 属性值知识的真实评价等级,即表现了对象的知识内涵,称为对象  $x_j$  属性  $a_i$  的修正值,记作  $RealRank_{a_i}(x_j)$ ,简记为  $N$ . 为保证在客观认知层面对象之间的差异并没有改变,修正值可以通过修正属性的经验值得到. 而属性值的经验值可以由字符间的相对排序来制定一个模糊的等级,类似层次分析法中比较因素重要的对比标度法<sup>[15,16]</sup>,其中必须保证相邻等级属性值的差异是一致的.

为了便于区分,通过经验值得到综合评价值可以称为经验综合评价值;通过修正值得到的综合评价值可以称为修正综合评价值.

由于经验值的修正需要对象的经验综合评价值  $d_i$  的参与,为了保证在字符型属性值向数值属性值修正的过程中,对象在属性值之间的差异特征保持不变,这里借鉴文献[17]中基于知识距离的属性重要度和知识权重的概念来刻画字符型属性值之间的差异,用于综合评价值的获取. 显然,  $a_i$  的属性重要度可以表示为:

$$Sig(a_i) = Dis_K(K(R_A), K(R_{A-\{a_i\}})) = \frac{1}{|U|} \sum_{k=1}^{|U|} \left( 1 - \frac{|S_{R_i}(x_k) \cap S_{R_{i-\{a_i\}}}(x_k)|}{|S_{R_i}(x_k) \cup S_{R_{i-\{a_i\}}}(x_k)|} \right).$$

这里,  $R_A$  和  $R_{A-\{a_i\}}$  分别代表属性集合  $A$  和  $A - \{a_i\}$  的等价关系. 因此,  $a_i$  在论域  $U$  的知识权重可以表示为  $W(a_i) = Sig(a_i) / \sum_{i=1}^n Sig(a_i)$ . 则,对于  $\forall x_j$  的经验综合评价值为  $d(x_j) = \sum_{i=1}^n [F_{a_i}(x_j) \cdot W(a_i)]$ . 例 1 详细地介绍了知识权重的计算方法.

**例 1** 设  $S = (U, AT, V, f)$  是一个信息系统, 详细的信息表内容如表 1 所示.

表 1 信息系统表

$U$	$a_1$	$a_2$	$a_3$	$a_4$
$x_1$	Good	Poor	Good	Middle
$x_2$	Middle	Poor	Good	Poor
$x_3$	Middle	Poor	Middle	Middle
$x_4$	Poor	Poor	Poor	Poor
$x_5$	Good	Poor	Good	Poor
$x_6$	Good	Poor	Good	Middle

在不同的等价关系下,知识结构如下所示(如文献[17]中所定义):

$$\begin{aligned} K(R_A) &= \{\{x_1, x_6\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_1, x_6\}\}. \\ K(R_{A-\{a_1\}}) &= \{\{x_1, x_6\}, \{x_2, x_5\}, \{x_3\}, \{x_4\}, \{x_2, x_5\}, \{x_1, x_6\}\}. \\ K(R_{A-\{a_2\}}) &= \{\{x_1, x_6\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_1, x_6\}\}. \\ K(R_{A-\{a_3\}}) &= \{\{x_1, x_6\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_1, x_6\}\}. \\ K(R_{A-\{a_4\}}) &= \{\{x_1, x_5, x_6\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_1, x_5, x_6\}, \{x_1, x_5, x_6\}\}. \end{aligned}$$

每个属性的属性重要度为  $Sig(a_1) = Dis_K(K(R_A), K(R_{A-\{a_1\}})) = 0.167$ , 同理,  $Sig(a_2) = Sig(a_3) = 0$ ,  $Sig(a_4) = 0.222$ . 则在论域  $U$  上, 条件属性  $a_i (i = 1, 2, \dots, n)$  的知识权重分别为  $W(a_1) = 0.429$ ,  $W(a_4) = 0.571$ ,  $W(a_2) = W(a_3) = 0$ .

通过属性重要度,即可从字符型属性的本身获得知识权重,从而加权计算各属性值的经验值从而得到相应经验综合评价值. 基于真实的应用场景,通过评价偏差来将经验值修正为修正值的过程可以理解成:将每个对象  $x_j$  的在该属性  $a_i$  下的经验值  $F_{a_i}(x_j)$  和其综合评价值  $d(x_j)$  的差的平均值作为属性  $a_i$  在这次评价过程中每次评价的平均偏差,再利用每个经验值偏差的比例来修正来得到修正值  $R_{a_i}(x_j)$ . 因此,基于字符型属性对象间优异程度的差异将字符型属性值向数值转化.

**定义 6** 评价偏差. 给定一个信息系统  $S = (U, AT, V, f)$ ,  $U = \{x_1, x_2, \dots, x_n\}$  是有限论域,  $A = \{a_1, a_2, \dots, a_m\}$  是数值型条件属性集,  $V_A = \{a_i(x_j) \mid (i = 1, 2, \dots, m; j = 1, 2, \dots, n)$  是条件属性值的集合,  $C = \{d(x_j) \mid (j = 1, 2, \dots, n)$  是综合评价值的集合, 则对于条件属性  $a_i$ , 对

于综合评价属性  $d$  的评价偏差被定义为  $Dev_d(a_i) = \frac{1}{n} \sum_{j=1}^n \frac{a_i(x_j) - d(x_j)}{d(x_j)}$ .

根据定义 6,若条件属性值  $a_i(x_k)$  被转化为经验值  $F_{a_i}(x_j)$ ,即经验值的评价偏差为  $Dev_d(a_i) = \frac{1}{n} \sum_{j=1}^n \frac{F_{a_i}(x_j) - d(x_j)}{d(x_j)}$ . 因此,在综合评价值为  $d$  的条件下,属性  $a_i$  的评价偏差为  $Dev(a_i)$  的情况下,根据上文字符型属性值向数值转化的步骤,对象  $x$  的修正值为  $R_{a_i}(x_j) = F_{a_i}(x_j) \cdot [1 - Dev_d(a_i)]$ .

### 3.2 字符型属性值的动态特征提取方法

本节利用逼近于理想解排序思想基于欧氏距离<sup>[18-19]</sup>得到了对象的最优贴适度,并作为字符型属性对象的评价值.利用的定义 2 中的数学特征导出对象因进入“优秀”对象集合而产生的影响的思想,提出字符型属性值的动态特征提取方法并将此特征用于之后对象决策.

**定义 7** 最优贴适度. 给定一个信息系统  $S = (U, AT, V, f)$ ,  $U = \{x_1, x_2, \dots, x_n\}$  是一个有限论域,  $A = \{a_1, a_2, \dots, a_m\}$  是条件属性集合,任意对象  $x_k$  属性值的修正值为  $R_{a_i}(x_j)$ ,其中  $i \in \{1, 2, \dots, m\}$ ,  $j \in \{1, 2, \dots, n\}$ ,最优对象的修正值为  $R_{a_i}(x_{OPTI})$ ,则对象基于欧氏距离的最优贴适度为:

$$close_E(x_j, x_{OPTI}) = \sum_{i=1}^m R_{a_i}(x_{OPTI}) - \sqrt{\sum_{i=1}^m (R_{a_i}(x_{OPTI}) - R_{a_i}(x_j))^2}.$$

当对象  $x_j$  取到  $\max_{1 \leq i \leq m} \{ |R_{a_i}(x_j)| \}$  时,有最优对象  $x_{OPTI} = x_j$ ,且  $close_E(x_{OPTI}, x_{OPTI}) = 0$ .  $x_j$  的最优贴适度简记为  $close(x_j)$  或  $c(x_j)$ .

这里,先通过修正值将字符型内涵用数值型表示并基于内涵的知识距离得到了对象的最优贴适度,则该基于欧氏距离的最优贴适度即可作为字符型属性对象的修正综合评价值,并体现了对象间的差异.

字符型属性值的动态特征提取算法具体步骤如下:

**输入:**对象  $X = \{x_1, x_2, \dots, x_n\}$  所构成的降序最优贴适度集合  $C = \{c_1, c_2, \dots, c_n\}$ .

**输出:**对象  $X = \{x_1, x_2, \dots, x_n\}$  的基于最优贴适度的特征集合  $C^* = \{z_1, z_2, \dots, z_n\}$ .

**Step1:**将集合  $A$  进行如下步长为 1 的增量式组合构成集合序列  $B_j$ ,其中  $j = 1, 2, \dots, n$ ,即  $B_1 = \{c_1\}$ ,  $B_2 = \{c_1, c_2\}$ ,  $\dots$ ,  $B_n = \{c_1, c_2, \dots, c_n\}$ .

**Step2:**分别对集合  $B_j$ ,求得均值  $\bar{B}_j = \frac{1}{j} \sum_{i=1}^j c_i$ ,标准差  $S_j = \sqrt{\frac{1}{j} \sum_{i=1}^j (c_i - \bar{B}_j)^2}$ ,其中  $j = 1, 2, \dots, n$ ,即均

值序列:  $\bar{B}_1 = \frac{1}{1} \sum_{i=1}^1 c_i = c_1$ ,  $\bar{B}_2 = \frac{1}{2} \sum_{i=1}^2 c_i, \dots, \bar{B}_n = \frac{1}{n} \sum_{i=1}^n c_i$ ,标准差序列:  $\{S_1, S_2, \dots, S_n\}$ .

**Step3:**计算  $A = \{a_1, a_2, \dots, a_n\}$  中每一个元素  $a_j$  的择优差  $Z_j = S_j - S_{j-1} > 0$ ,其中  $j = 1, 2, \dots, n$ ,即得一个择优差序列:  $C^* = \{z_1, z_2, \dots, z_n\}$ .

## 4 基于字符型属性值的动态三支决策模型

### 4.1 动态三支决策模型的建立

在第 3 节,给出了获取字符型属性对象的属性修正值的方法.在动态的决策过程中,修正值总是会随着每次的变化而更新,对于多属性而言,本节利用最优贴适度作为修正综合评价值来定义了最优贴适度比率,再结合文献[6]中决策模型的思想对字符型属性的对象进行决策,建立基于字符型属性值更新的动态三支决策模型.

首先,列出在给定接受域对象数条件下的三支决策分类方法:

(1)划分一部分肯定成为“理想”情况下的“优秀”对象进入接受域直接接受.

(2)划分一部分可能成为“理想”情况下的“优秀”对象进入延迟域待下一次更新后决策.

(3)划分一部分不可能成为“理想”情况下的“优秀”对象进入拒绝域并不再更新.

重复上述过程,直至接受域对象个数满足  $k$  个.若在实际操作中,更新过程结束时,接受域对象数仍未满足  $k$ ,则考虑从最后一次划分的延迟域中选出若干个属性值较高的对象放入接受域,直至接受域对象数满足需求.

对于字符型的属性而言,为更好地描述决策域与决策规则,本文通过将最优贴适度归一化,定义了最优贴适度比率的概念.

**定义 8** 最优贴适度比率. 设  $X = \{x_1, x_2, \dots, x_n\}$  为决策域  $U$  中对象集合,  $D_a$  是属性  $a \in AT$  的属性值域,若已知任意对象  $x_j$  的最优贴适度为  $c(x_j)$ ,  $j = 1, 2, \dots, n$ . 对于  $\forall x_j \in X$ ,对象  $x_j$  的最优贴适度  $c(x_j)$  上的属性比率

$$\text{定义为: } P_c(x_j) = \frac{c(x_j)}{c_{\max}(X)}.$$

这里  $c_{\max}(X) = \max_{1 \leq i \leq n} \{c(x_i)\}$ ,显然  $P_c(x_j) \geq 0$ .

**定义 9** 基于最优贴适度比率的动态三支决策模型. 设  $S = (U, AT, V, f)$  是一个决策表,  $X \subseteq U$ ,  $0 \leq \beta < \alpha \leq 1$ ,则上下近似可定义为:

$$\overline{apr}_{(\alpha, \beta)}(X) = \{x \in U \mid P_c(x) > \beta\},$$

$$\underline{apr}_{(\alpha, \beta)}(X) = \{x \in U \mid P_c(x) > \alpha\}.$$

上下近似将论域分为 3 个部分:接受域  $POS_{(\alpha, \beta)}$

( $X$ )、延迟域  $BND_{(\alpha,\beta)}(X)$ 、拒绝域  $NEG_{(\alpha,\beta)}(X)$ , 即  $POS_{(\alpha,\beta)}(X) = apr_{(\alpha,\beta)}(X)$ ,  $BND_{(\alpha,\beta)}(X) = \overline{apr_{(\alpha,\beta)}(X)} - apr_{(\alpha,\beta)}(X)$ ,  $NEG_{(\alpha,\beta)}(X) = U - apr_{(\alpha,\beta)}(X)$ .

本文模型的阈值选取方法为:  $k$  视为对象集合  $X$  的划分点,  $X$  被划分成两个不相交的对象子集  $X_1 = \{x_1, x_2, \dots, x_k\}$  和  $X_2 = \{x_{k+1}, x_{k+2}, \dots, x_n\}$ , 其中  $1 < i < k, k+1 < j < n, x_i$  为对象子集  $X_1$  中除  $x_1$  外, 第一个择优差最大的对象,  $x_j$  为对象子集  $X_2$  中第一个择优差最大的对象, 由定义 8 可以得到阈值  $\alpha = c(x_i)/c_{MAX}(X)$  和  $\beta = c(x_j)/c_{MAX}(X)$ .

相比数值型属性对象的划分方法, 字符型属性对象的划分方法进行了优化, 改进规则为  $x_i$  为对象子集  $X_1$  中除  $x_1$  外, 第一个择优差最大的对象, 这是由于定义 7 中  $close_E(x_{opti}, x_{opti}) = \sum_{i=1}^m R_{a_i}(x_{opti})$  可能会导致最优贴适度次大的对象的择优差恒为最大.

## 4.2 模型的动态决策过程与应用

在 4.1 节中, 已给出了基于字符型属性值更新的三支决策模型. 然而, 所有实际问题的决策结果最终都会变为二支结果<sup>[20,21]</sup>.

随着对第一次对象选取过程的阈值的计算, 可以将对象划分至接受域、延迟域、拒绝域. 由于阈值是通过分析字符对象的属性值集合内部的优秀程度的差异来提取特征的, 则当属性值集合内部知识结构发生改变, 相应的决策阈值也会自适应的变化<sup>[22]</sup>. 同理, 当最终接受域对象个数的划分需求改变时, 相应的  $\alpha$  和  $\beta$  也自适应地变化. 因此, 该模型在动态决策过程中具有良好的自适应性.

动态决策过程中字符型属性对象向二支结果转化的过程如下:

**输入:** 给定接受域对象数  $k$ , 对象更新次数  $p-1$ , 初始化  $K_p = k, p = 1$ .

**输出:** 接受域对象集合.

**Step1:** 计算第  $p$  个数据集中  $N_p$  个对象, 将所有对象放入延迟域.

**Step2:** 将延迟域中待决策对象的字符型属性根据刻度标尺法转化成经验值  $F_{a_i}(x_j)$ , 得到待决策对象的经验值表.

**Step3:** 根据待决策对象的字符型属性值计算出每个属性  $a_i$  基于知识距离的属性重要度  $Sig(a_i)$ , 并得到知识权重  $W(a_i)$ , 经验综合评价值和经验值构成的模糊数据表.

**Step4:** 根据经验值表计算属性的评价偏差  $Dev_{a_i}(a_i)$ , 再利用评价偏差得到每个待决策对象属性的修正值  $R_{a_i}(x_j)$ .

**Step5:** 根据待决策对象属性的修正值计算对象最

优贴适度  $c(x_j)$ , 作为待决策对象的修正综合评价价值.

**Step6:** 根据待决策对象的最优贴适度的大小, 降序排列后给对象依次赋予编号 1 至  $N$ .

**Step7:** 将所有待决策对象进行决策划分.

**Step8:** 根据决策结果, 接受了  $k_p$  个对象, 则修改  $K_p = k - \sum_{i=1}^p k_i$ , 若  $p \leq 7$ , 修改决策次数  $p = p + 1$ , 转至 Step1, 否则进入 Step9.

**Step9:** 接受域总对象数若不等于  $k$ , 则从最后次决策的延迟域中选取较高的完成划分, 否则完成决策.

其中,  $k$  代表给定的接受域对象数, 在第  $p$  次动态决策中,  $N_p$  代表本次待决策的对象数,  $k_p$  代表第  $p$  次决策后接受的对象数,  $K_p$  代表接受域需要接受的对象数.

则  $K_p$  的计算公式为  $K_p = k - \sum_{i=1}^p k_{i-1}$  (其中  $k_0 = 0$ ).

## 5 实验对比及分析

### 5.1 实验设计及评价函数

为了验证模型的可行性和有效性, 本节选取了属于 Social Sciences 领域中可以建立出更新联系的 9 个 UCI 数据集 (Adult、Nursery、Census-Income、Drug Consumption、KDD Cup 1999、Census-Income (KDD)、Autism Screening Adult、SPECT Heart、Heart Disease), 参照它们分类结果和更新规则仿真组合出了具有更新状态的字符型属性的仿真数据集并进行了仿真实验. 其中, 仿真数据集的对象属性参考了 Adult、Nursery、Census-Income、KDD Cup 1999、SPECT Heart, 更新规则参考 Adult、Census-Income (KDD)、Drug Consumption、Autism Screening Adult、SPECT Heart. 属性的选定原则为尽可能覆盖具有关联性的数据集, 更新规则的选取为通过 sigmoid 函数拟合对原数据进行了的逻辑回归实现更多对象的仿真. 实验环境为 I5-6700 CPU, 16GB RAM, Windows10 操作系统, 开发环境为 Microsoft Visual Studio 2015. 实验步骤参考 4.2 节中的动态决策过程.

**定义 10** 模型的评价函数<sup>[6]</sup>. 设  $POS(X)$  为本文提出的动态三支决策模型决策后划分出的接受域,  $|\bullet|$  表示集合内的对象数.

$\lambda$  表示三支决策结果与“最优”相似程度;  $\mu$  为三支决策与二支决策的对象属性值总和的差值, 表示三支决策模型较二支的提升值;  $\nu$  为三支决策模型的优化率, 用决策过程中减少的更新对象个数与找寻“最优”结果的决策过程需更新的个数的比值来刻画.

$\lambda$  被定义为  $\lambda = \frac{|POS(X) \cap POS^*(X)|}{|POS^*(X)|}$ . 其中

$POS^*(X)$  代表“最优”结果的对象集合, 即在对象全部更新后, 属性值最高的前  $k$  个对象的集合.

$\mu$  被定义为  $\mu = \frac{|POS(X) \cap POS^*(X)|}{|POS^*(X)|} - \frac{|POS(X) \cap POS^0(X)|}{|POS^0(X)|}$ . 其中  $POS^0(X)$  代表在二支决策下被划分到接受域的对象集合,即在对象在未更新前属性值最高的前  $k$  个对象的集合.

$\nu$  被定义为  $\nu = \frac{\sum_{i=1}^{t+1} (|NEG_i(X)| \cdot (t+1-i))}{N \cdot t}$ . 其中  $N \cdot t$  是找寻“最优”结果需要更新的对象个数,  $|NEG_i(X)|$  是第  $i$  次划分的拒绝域的对象个数,  $t+1-i$  是因第  $i$  次划分到拒绝域而减少的对象更新次数.

本文实验通过控制数据量  $N$ 、更新次数  $t$ 、接受域需要接受的对象数  $k$  来进行对比实验,验证模型的合理性及有效性.

## 5.2 实验结果分析

仿真模型实验选出  $k$  个对象作为最为优秀的—

表 2 字符型属性对象更新 7 次的仿真实验结果

属性个数:5					属性个数:23				
$N$	$k$	$\lambda$	$\mu$	$\nu$	$N$	$k$	$\lambda$	$\mu$	$\nu$
50000	0.1N	95.27%	26.80%	67.82%	50000	0.1N	96.12%	31.68%	73.05%
	0.2N	95.27%	25.32%	66.85%		0.2N	96.47%	32.19%	72.77%
	0.3N	96.78%	26.74%	65.12%		0.3N	97.45%	32.14%	71.87%
	0.4N	96.34%	26.18%	65.02%		0.4N	97.62%	33.01%	70.46%
	0.5N	96.22%	26.24%	65.13%		0.5N	97.45%	32.86%	70.56%
100000	0.1N	96.13%	29.78%	71.82%	100000	0.1N	97.34%	34.64%	74.99%
	0.2N	97.34%	30.29%	71.24%		0.2N	97.35%	35.76%	74.86%
	0.3N	97.34%	32.01%	70.12%		0.3N	97.18%	35.48%	73.75%
	0.4N	97.53%	32.23%	69.64%		0.4N	98.73%	35.12%	73.24%
	0.5N	98.15%	32.75%	68.03%		0.5N	98.97%	35.06%	72.85%
200000	0.1N	98.86%	34.86%	78.99%	200000	0.1N	99.16%	39.59%	83.65%
	0.2N	98.31%	35.98%	78.87%		0.2N	99.12%	40.07%	82.71%
	0.3N	98.68%	36.04%	78.16%		0.3N	99.53%	39.87%	80.75%
	0.4N	98.38%	35.97%	77.47%		0.4N	99.23%	40.88%	79.76%
	0.5N	99.41%	36.24%	77.02%		0.5N	99.23%	40.92%	79.97%

综上所述,仿真实验结果证明了基于字符型属性值更新的动态三支模型较二支决策更优异可行,且模型的决策结果十分接近“最优”结果.

## 6 总结

针对现实应用场景,我们在文献[6]中已经解决了数值型属性对象的决策问题,而对于字符型属性对象,这样的决策过程忽略了值更新过程中字符型属性之间知识内涵的变化带来的错误决策.对此,本文在保证字符型属性在处理中字符型属性对象间优异程度的差异不变的情况下,提出了在知识距离度量下字符型属性对象的修正方法;定义了基于欧氏距离的最优贴适度作为对象的综合评价价值,提出了基于字符型属性值动态特征的提取方法;建立了基于字符型属性值的三支

批次,令  $k \in \{0.1N, 0.2N, 0.3N, 0.4N, 0.5N\}$  来比较模型效果,更新情况为 7 次.

仿真实验结果如表 2 所示.其中  $N$  为仿真决策的对象总个数,  $k$  为最终接受域对象总个数.表中的 3 个评价价值  $\lambda, \mu, \nu$  均通过 50 次实验取平均值.

仿真实验结果显示:三支模型决策结果接近“最优”结果程度  $\lambda$  可以高达 99.53%,说明模型精度较高,具有良好的可行性;优化率  $\nu$  的数据表明了三支决策模型的决策过程相比在对象全更新后去寻找“最优”结果的过程,减少了 52.13% ~ 83.65% 的更新对象个数,说明了模型具有显著的高效性;三支决策结果的提升值  $\mu$  均大于 0,说明了模型在相比二支结果更具优异性.

从对象的属性个数上,可以看出:当对象属性的个数越多,模型相比二支决策的提升越多,且减少的更新对象的个数也越多.

决策模型及阈值的选取办法.通过仿真实验,表明了该模型在处理数值型属性的对象同样可行,且在得到接近“最优”决策结果的同时能减少了决策的更新成本.

## 参考文献

- [1] Yao Yiyu. Three-way decisions with probabilistic rough sets[J]. Information Sciences, 2010, 180(3): 341 - 353.
- [2] Greco S, Matarazzo B, Slowinski R. Parameterized rough set model using rough membership and Bayesian confirmation measures[J]. International Journal of Approximate Reasoning, 2008, 49(2): 285 - 300.
- [3] Yang Xiaoping, Yao Jingtao. Modelling multi-agent three-way decisions with decision-theoretic rough sets[J]. Fun-

- damenta Informaticae, 2012, 115(2-3):157-171.
- [4] Yao Yiyu, Deng Xiaofeng. Sequential three-way decisions with probabilistic rough sets [A]. Proceeding of the 10th IEEE International Conference on Cognitive Informatics and Cognitive Computing [C]. Banff, Alberta, Canada, 2011. 120-125.
- [5] Hao Chen, Li Jinhai, Min Fan, Eric C. C. Tsang. Optimal scale selection in dynamic multi-scale decision tables based on sequential three-way decisions [J]. Information Sciences, 2017, 415:213-232.
- [6] Zhang Qinghua, Lv Gongxun, Chen Yuhong, Wang Guoyin. A dynamic three-way decision model based on the updating of attribute values [J]. Knowledge-Based Systems, 2018, 142:71-84.
- [7] Chen Yufei, Yue Xiaodong, Fujita Hamido, Fu siyuan. Three-way decision support for diagnosis on focal liver lesions [J]. Knowledge-Based Systems, 2017, 127:85-99.
- [8] Li Huaxiong, Zhang Libo, Zhou Xianzhong, Huang Bing. Cost-sensitive sequential three-way decision modeling using a deep neural network [J]. International Journal of Approximate Reasoning, 2017, 85:68-78.
- [9] Li Xiaonan, Yi Huangjian, She Yanhong, Sun Bingzhen. Generalized three-way decision models based on subset evaluation [J]. International Journal of Approximate Reasoning, 2017, 83:142-159.
- [10] Zhang Qinghua, Yang Jingjing, Yao Longyang. Attribute reduction based on rough approximation set in algebra and information views [J]. IEEE Access, 2016, 4: 5399-5407.
- [11] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation [J]. Journal of Machine Learning Research, 2012, 3:993-1002.
- [12] Pawlak Z. Rough sets [J]. International Journal of Computer and Information Sciences. 1982, 11(5):341-356.
- [13] 王国胤. Rough 集理论与知识获取 [M]. 西安:西安交通大学出版社, 2001.  
Wang Guoyin. Rough Set Theory and Knowledge Acquisition [M]. Xi'an: Xi'an Jiaotong University Press, 2001. (in Chinese)
- [14] Yao Yiyu. Probabilistic rough set approximations [J]. International Journal of Approximate Reasoning, 2008, 49: 255-271.
- [15] Wei C C, Chien C F, Wang M J J. An AHP-based approach to ERP system selection [J]. International Journal of Production Economics, 2005, 96(1):47-62.
- [16] Saaty T L. Decision-making with the AHP: Why is the principal eigenvector necessary [J]. European Journal of Operational Research, 2003, 145(1):85-91.
- [17] Qian Yuhua, Liang Jiye, Dang Chuangyin, Xu Wei. Knowledge distance in information systems [J]. Journal of Systems Science & Systems Engineering, 2007, 16(4):434-449.
- [18] P E Danielsson, Euclidean distance mapping [J]. Computer Graphics & Image Processing, 1980, 14(3):227-248.
- [19] Wang Liwei, Zhang Yan, Feng Jufu. On the euclidean distance of images [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2005, 27(8):1334-1339.
- [20] Jiang Chunmao, Wu Junwei, Li zhicong. Generalized three-way decision models based on subset evaluation [J]. International Journal of Approximate Reasoning, 2017, 83:142-159.
- [21] 胡峰, 王蕾, 周耀. 基于三支决策的不平衡数据过采样方法 [J]. 电子学报, 2018, 46(1):135-144.  
Hu Feng, Wang Lei, Zhou Yao. An oversampling method for imbalance sata based on three-way decision model [J]. Acta Electronica Sinica, 2018, 46(1):135-144. (in Chinese)
- [22] 贾修一, 李伟漳, 商琳, 等. 一种自适应求三枝决策中决策阈值的算法 [J]. 电子学报, 2011, 39(11):2520-2525.  
Jia Xiuyi, Li Weiwei, Shang Lin, et al. An adaptive learning parameters algorithm in three-way decision-theoretic rough set model [J]. Acta Electronica Sinica, 2011, 39(11):2520-2525. (in Chinese)

### 作者简介



张清华 男, 1974 年 11 月出生. 教授、博士生导师. 现为重庆邮电大学计算机科学与技术学院执行院长. 主要从事粗糙集、模糊集、粒计算、不确定信息处理等相关研究.  
E-mail: zhangqh@cqupt.edu.cn



吕功勋 男, 1994 年 4 月出生. 重庆邮电大学计算机科学与技术学院硕士研究生. 主要从事三支决策、粗糙集、粒计算等相关研究.  
E-mail: 370303223@qq.com