

DOI: 10.11992/tis.201905052

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20190904.1826.004.html>

## 非完备决策信息系统中的不确定性度量

高文华<sup>1</sup>, 梁吉业<sup>2</sup>, 王宝丽<sup>3</sup>, 庞天杰<sup>1</sup>

(1. 太原师范学院 计算机科学与技术系, 山西 晋中 030619; 2. 山西大学 计算智能与中文信息处理教育部重点实验室, 山西 太原 030006; 3. 运城学院 数学与信息技术学院, 山西 运城 044000)

**摘要:** 针对粗糙集数据分析中的不确定性度量问题。本文首先构造一种新型的考虑条件属性缺失度的目标概念条件熵和决策知识条件熵。在此基础上, 提出基于条件熵的属性权重确定技术和最小条件熵非完备属性取值补充方法, 用以解决属性权重完全未知的非完备多属性决策问题。应用实例分析表明: 该方法能有效结合粗粒度的初步分级信息, 客观地确定决策因素取值, 具有很强的解释意义, 得到的决策结果更为合理有效。

**关键词:** 非完备决策系统; 相容关系; 知识粒度; 不确定性度量; 条件熵; 属性权重; 最小条件熵原则; 多属性决策方法

中图分类号: TP301 文献标志码: A 文章编号: 1673-4785(2019)06-1100-11

中文引用格式: 高文华, 梁吉业, 王宝丽, 等. 非完备决策信息系统中的不确定性度量[J]. 智能系统学报, 2019, 14(6): 1100-1110.

英文引用格式: GAO Wenhua, LIANG Jiye, WANG Baoli, et al. Uncertainty measure in incomplete decision information system[J]. CAAI transactions on intelligent systems, 2019, 14(6): 1100-1110.

## Uncertainty measure in incomplete decision information system

GAO Wenhua<sup>1</sup>, LIANG Jiye<sup>2</sup>, WANG Baoli<sup>3</sup>, PANG Tianjie<sup>1</sup>

(1. Department of Computer Science and Technology, Taiyuan Normal University, Jinzhong 030619, China; 2. Key Laboratory of Ministry of Education for Computational Intelligence and Chinese Information Processing, Taiyuan 030006, China; 3. School of Mathematics and Information Technology, Yuncheng University, Yuncheng 044000, China)

**Abstract:** In order to solve uncertainty measure problem in data analysis of rough set, this study first constructs a new type of conditional entropy of the objective concept and conditional entropy of decision knowledge, with consideration of the degree of missing of conditional attributes, and moreover, proposes the conditional entropy-based attribute weight determination technique and a complementary method for incomplete attributes with minimum conditional entropy, so as to solve a kind of incomplete multi-attribute decision-making problem whose attribute weight is completely unknown. The real practical application shows that the proposed method can effectively combine coarse-grained preliminary classification information to objectively determine the value of decision factors, having strong explanatory significance, and the obtained decision results are more reasonable and effective.

**Keywords:** incomplete decision system; tolerance relation; knowledge granularity; uncertainty measure; conditional entropy; attribute weight; minimum conditional entropy principle; multi-attribute decision-making method

粗糙集理论是由 Pawlak<sup>[1]</sup>提出的一种处理不精确、不确定和模糊信息的数学工具, 目前已被广泛应用于病症诊断、金融风险 and 机器故障分析等问题中<sup>[2-4]</sup>。不确定性度量是粗糙集理论中的一个重要研究问题, 它在属性约简、规则获取中

发挥着重要的作用。

目前, 国内外学者已经对经典粗糙集理论的不确定性度量问题做了系统的研究<sup>[5-13]</sup>。这些研究大多基于完备描述的信息系统。现实问题中, 由于属性度量的高代价性、评价者对问题的判断水平不足或数据输入人员的疏忽问题, 常存在属性值缺失的非完备信息系统。这种非完备信息系统的不确定性度量对于非完备知识获取等相关

收稿日期: 2019-05-27. 网络出版日期: 2019-09-05.

基金项目: 国家自然科学基金项目(61703363); 山西省重点实验室开放课题基金项目(CICIP2018008).

通信作者: 梁吉业. E-mail: [ljy@sxu.edu.cn](mailto:ljy@sxu.edu.cn).

问题的求解也至关重要。近年来,研究者对非完备信息系统或非完备决策系统的不确定度量取得了一系列非常重要且有意义的研究成果<sup>[14-16]</sup>。文献[17]对3种类型的决策表(完备、非完备和最大一致块)计算粗糙集的模糊性和粗略决策。文献[18]在非完备决策系统中定义了一种条件熵。而进一步研究发现,该条件熵对知识粒度不具有单调性,这使得评估不完备决策系统中的不确定性变得不那么合理。文献[19]进一步研究非完备决策系统的不确定性度量,提出对知识粒度变化敏感且具有单调性的条件熵。虽然这些度量值随着知识粒度的变化而改变,但没有充分考虑由于属性值缺失引起的不确定性。因此,构造既具有知识粒度单调性又可体现属性缺失的不确定性度量具有重要的意义。

非完备多属性决策问题是一类重要的多属性决策问题。属性权重的合理设定是获得可信决策结果的保障;属性缺失值填充是有效集结不同属性信息的关键。目前属性权重确定方法大都基于完备描述的多属性决策问题。文献[20]基于熵权法设定属性权重。文献[21]利用优势粗糙集中属性的综合优势度来确定多属性决策中的属性权重。文献[22]基于信息量确定属性权重。上述赋权方法是从信息论的角度来计算的属性权重,不再依赖于数据分布,且具有客观性。属性缺失值的填充方法通常使用统计分析法和最近邻法等填充方法将非完备系统完备化,然而,在特定情况下,这些填充方法填充的结果会与实际数据产生偏离,从而导致不尽合理的决策结果。

本文考虑条件属性值缺失的情形,提出了一种新的条件熵用以刻画非完备决策系统中的知识不确定性程度,同时分析新的条件熵具有有界、单调以及完备可退化的特性。此外,将新条件熵应用于非完备多属性决策问题求解中,提出一种基于条件熵的非完备多属性决策方法。该方法以条件熵为统领,确定属性权重并以最小条件熵为准则选择填充值,以此更加客观地求解现实中存在的非完备多属性决策问题,并获得合理有效的决策结果。最后应用房屋评测实例说明所提方法的有效性与合理性。

## 1 相关概念

### 1.1 基本定义

信息系统  $S = \langle U, V, f, A \rangle$  是一个4元组,其中  $U$  是由对象构成的非空有限集合,称为论域。 $A$  为有限属性集,  $V_a$  为属性  $a$  的值域,  $V$  为属性集  $A$  的

值域,则  $V = \cup_{a \in A} V_a$ ; 信息函数  $f: U \times A \rightarrow V$ , 表示论域中每个对象在每个属性上均对应一个属性值。

若在信息系统  $S$  中存在  $x \in U$ 、 $a \in A$  使得  $f(x, a)$  等于空值,其中空值用\*表示,则称该信息系统为非完备信息系统,记作  $IS$ 。

若  $A = C \cup d$ ,  $C$  为有限的条件属性集,  $d$  为决策属性,  $C \cap d = \emptyset$ 、 $V = \cup_{a \in C} V_a$ 、 $f: U \times (C \cup d) \rightarrow V$ , 则称之为决策信息系统,记作  $DS$ 。若  $* \in V_C$ 、 $* \notin V_d$ , 则称决策信息系统为  $IDS$ 。

在非完备信息系统中,由于一些缺失值的存在,完备信息系统中的等价关系已不适用于辨别任意两个对象间的关系。Kryszkiewicz<sup>[23]</sup> 运用较弱的相容关系刻画非完备信息系统中对象之间的相似性关系,进一步刻画概念的上、下近似。

**定义1<sup>[23]</sup>** 相容关系  $T(P)$  定义为:

$$T(P) = \{(u, v) \in U \times U \mid \forall a \in P, a(u) = a(v) \vee a(u) = * \vee a(v) = *\} \quad (1)$$

$T(P)$  是论域  $U$  上的相容关系,满足自反性和对称性。在非完备信息系统中,相容关系指的是将缺失值看作与任何同属性下的已知值有相等的可能性的一种描述。对象  $u$  在知识  $P$  下与对象  $v$  可能的不可区分的相容类为  $T_P(u) = \{v \in U \mid (u, v) \in T(P)\}$ ,  $u, v \in U$ , 称  $T_P(u)$  为相容关系下的信息粒度。

### 1.2 非完备信息系统中的不确定性度量

非完备信息系统中的信息熵在文献[14]中首次进行了深入的探讨与研究。

**定义2<sup>[14]</sup>** 在非完备信息系统  $IS = \langle U, A, V, f \rangle$  中,  $P \subseteq A$ , 属性  $P$  在论域  $U$  上的信息熵定义为:

$$H(P) = -\frac{1}{U} \sum_{i=1}^{|U|} \log \frac{|T_P(u_i)|}{|U|} \quad (2)$$

式中:  $T_P(u_i)$  是非完备信息系统中在  $U$  上定义的相容关系下的相容类;  $|U|$  表示集合  $U$  的基数。

文献[14]对非完备信息系统的不确定性度量进行了研究,文献[18]进一步考虑了非完备决策信息系统中的不确定性度量,提出了条件信息熵的概念。

**定义3<sup>[18]</sup>**  $IDS = \langle U, C \cup \{d\}, V, f \rangle$  为非完备决策信息系统,  $P, Q \subseteq C \cup \{d\}$ 。属性  $Q$  相对属性  $P$  的条件熵定义为:

$$H_1(Q|P) = -\frac{1}{|U|} \sum_{i=1}^m \log \frac{|T_Q(u_i) \cap T_P(u_i)|}{|T_P(u_i)|} \quad (3)$$

其中  $T_P(u_i)$  与  $T_Q(u_i)$  是非完备信息系统中在  $U$  上定义的相容关系下的相容类。

文献[19]进一步对文献[18]中的条件信息熵进行了改进,提出一种满足单调性的条件熵度量公式。

**定义4<sup>[19]</sup>** 设  $IDS = \langle U, C \cup D, V, f \rangle$ ,  $* \notin V_D$ ,  $* \in V_C$  是非完备决策系统, 其中,  $U = \{u_1, u_2, \dots, u_n\}$ ,  $B \subseteq C$  是一个属性集,  $U/D = \{Y_1, Y_2, \dots, Y_m\}$ 。则决策属性  $D$  相对属性  $B$  的条件熵定义为:

$$H_2(D|B) = - \sum_{i=1}^{|U|} \sum_{j=1}^m \frac{|T_B(u_i) \cap Y_j|}{|U|} \log \frac{|T_B(u_i) \cap Y_j|}{|T_B(u_i)|} \quad (4)$$

其中  $T_B(u_i)$  是非完备信息系统中在  $U$  上定义的相容关系下的相容类。

然而, 定义3及定义4并没有明显地刻画出非完备决策系统的属性值缺失程度的特点, 如果此属性(或属性集)的信息变得更粗糙(即此属性或属性集上有更多的缺失值), 则在该属性集下的不确定性应该更大, 但有时上述不确定性度量值却保持不变, 相关实例见例1。

**例1** 表1与表2是2个给定的决策信息表。论域与条件属性集分别为  $U = \{u_1, u_2, u_3, u_4, u_5, u_6\}$  和  $A_1 = \{a_1, a_2, a_3, a_4\}$ 、 $A_2 = \{b_1, b_2, b_3, b_4\}$ 。二者的决策知识相同, 可表示为  $U/d = \{\{u_1, u_3\}, \{u_2, u_4, u_5, u_6\}\}$ , 表2将表1的部分属性值进行修改, 以获得比表1缺失值更多的决策表。

表1 一个非完备决策表  
Table 1 An incomplete decision table

$U$	$a_1$	$a_2$	$a_3$	$a_4$	$d$
$u_1$	1	2	3	4	×
$u_2$	2	2	3	4	√
$u_3$	2	2	4	2	×
$u_4$	1	*	3	2	√
$u_5$	*	2	3	2	√
$u_6$	2	2	*	*	√

表2 表1部分已知值缺失化后的非完备决策表  
Table 2 An incomplete decision table with partial known values modified in Table1

$U$	$b_1$	$b_2$	$b_3$	$b_4$	$d$
$u_1$	1	2	3	4	×
$u_2$	2	*	3	4	√
$u_3$	*	*	4	2	×
$u_4$	1	*	3	2	√
$u_5$	*	*	3	2	√
$u_6$	2	2	*	*	√

利用定义3和定义4, 分别计算表1和表2两个系统的条件熵。

1) 由相容关系分别计算6个对象在属性  $A_1$ 、 $A_2$  下的相容类:

$$\begin{aligned} S_{A_1}(u_1) &= \{u_1\} \\ S_{A_1}(u_3) &= \{u_3, u_6\} \\ S_{A_1}(u_5) &= \{u_4, u_5, u_6\} \\ S_{A_1}(u_2) &= \{u_2, u_6\} \\ S_{A_1}(u_4) &= \{u_4, u_5\} \\ S_{A_1}(u_6) &= \{u_2, u_3, u_5, u_6\} \\ S_{A_2}(u_1) &= \{u_1\} \\ S_{A_2}(u_3) &= \{u_3, u_6\} \\ S_{A_2}(u_5) &= \{u_4, u_5, u_6\} \\ S_{A_2}(u_2) &= \{u_2, u_6\} \\ S_{A_2}(u_4) &= \{u_4, u_5\} \\ S_{A_2}(u_6) &= \{u_2, u_3, u_5, u_6\} \end{aligned}$$

由此, 可知

$$\begin{aligned} |S_{A_1}(u_1)| &= |S_{A_2}(u_1)|, |S_{A_1}(u_2)| = |S_{A_2}(u_2)| \\ |S_{A_1}(u_3)| &= |S_{A_2}(u_3)|, |S_{A_1}(u_4)| = |S_{A_2}(u_4)| \\ |S_{A_1}(u_5)| &= |S_{A_2}(u_5)|, |S_{A_1}(u_6)| = |S_{A_2}(u_6)| \end{aligned}$$

2) 根据定义3, 计算2个非完备决策系统的条件熵分别为:

$$\begin{aligned} H_1(d|A_1) &= - \frac{1}{|U|} \sum_{i=1}^m \log \frac{|T_d(u_i) \cap T_{A_1}(u_i)|}{|T_{A_1}(u_i)|} = \\ &= - \frac{1}{6} \times \left( 0 + 0 + \log \frac{1}{2} + 0 + 0 + \log \frac{3}{4} \right) = 0.2358 \end{aligned}$$

$$H_1(d|A_2) = - \frac{1}{6} \times \left( 0 + 0 + \log \frac{1}{2} + 0 + 0 + \log \frac{3}{4} \right) = 0.2358$$

3) 根据定义4, 计算2个非完备决策系统的条件熵分别为:

$$\begin{aligned} H_2(d|A_1) &= - \sum_{i=1}^{|U|} \sum_{j=1}^m \frac{|T_{A_1}(u_i) \cap Y_j|}{|U|} \log \frac{|T_{A_1}(u_i) \cap Y_j|}{|T_{A_1}(u_i)|} = \\ &= - \left[ 0 + \frac{2}{6} \log \frac{2}{2} + 2 \times \frac{1}{6} \log \frac{1}{2} + \frac{2}{6} \log \frac{2}{2} + \frac{3}{6} \log \frac{3}{3} + \right. \\ &\quad \left. \left( \frac{1}{6} \log \frac{1}{4} + \frac{3}{6} \log \frac{3}{4} \right) \right] = 0.8742 \\ H_2(d|A_2) &= - \left[ 0 + \frac{2}{6} \log \frac{2}{2} + 2 \times \frac{1}{6} \log \frac{1}{2} + \frac{2}{6} \log \frac{2}{2} + \right. \\ &\quad \left. \frac{3}{6} \log \frac{3}{3} + \left( \frac{1}{6} \log \frac{1}{4} + \frac{3}{6} \log \frac{3}{4} \right) \right] = 0.8742 \end{aligned}$$

从以上计算可以看出表1与表2两个非完备决策信息系统的两种条件熵都相同, 但实际上由于属性值缺失程度不同, 它们所蕴含的信息不同, 所以非完备决策信息系统的确定性也应该不同。

## 2 非完备决策系统中的条件熵

在经典粗糙集理论中, 产生不确定性的原因主要有两个方面: 1) 信息粒度带来的知识不确定

性;2)粗糙集边界引起的集合不确定性。然而对于非完备信息系统,经过分析发现:已有的不确定性度量方法在同一非完备信息系统中,相同属性集下的属性值缺失程度不同所形成的系统会得到相同的度量值,这时,如果直接应用已有的度量公式来度量非完备信息系统的确定性,就存在一定的局限性,因为已有的度量方法只考虑了非完备信息系统中由于知识粒度引起的不确定性以及由粗糙集边界带来的集合不确定性,而没有考虑由于属性值缺失所引起的不确定性。针对以上问题,本文在非完备信息系统中提出一种考虑缺失值程度的不确定性度量,并进一步考虑了非完备决策信息系统中的不确定性度量,提出新的条件熵,讨论了其相关性质。从新的角度验证并说明条件熵能够更加精确地度量非完备决策系统的不确定性。

通过以上分析,将完备信息系统下的不确定性度量公理化定义推广到非完备信息系统中,提出如下非完备信息系统中不确定性度量满足条件。

**定义5** 设  $IS = \langle U, V, f, A \rangle, P \subseteq A$ , 是一个非完备信息系统,  $X \subseteq U$ ,  $\Psi(U)$  是  $U$  上所有覆盖的全体集合,  $\Gamma(U)$  是  $U$  的幂集, 若存在  $\Psi(U) \Gamma(U)$  到实数集  $\mathbf{R}^1$  的映射函数  $H(X|P): \Psi(U) \Gamma(U) \rightarrow \mathbf{R}^1$ , 满足如下条件, 则称它为目标概念  $X$  在覆盖  $U/T(P)$  下的不确定性度量。

1) 非负性:  $\forall P \subseteq A$ , 有  $H(X|P) \geq 0$ 。

2) 不变性: 若  $\forall P, Q \subseteq A$ , 有  $H(X|P) = H(X|Q)$ , 那么存在映射  $f: U/T(P) \rightarrow U/T(Q)$ , 使得对于任意  $i \in \{1, 2, \dots, |U|\}$ , 有  $|T_P(u_i)| = |f(T_P(u_i))|$ , 且在属性集  $P$  与属性集  $Q$  下的缺失值程度相同, 其中,  $U/T(P) = \{T_P(u_1), T_P(u_2), \dots, T_P(u_{|U|})\}$ ,  $U/T(Q) = \{T_Q(u_1), T_Q(u_2), \dots, T_Q(u_{|U|})\}$ 。

3) 单调性:  $\forall P, Q \subseteq A$ , 若  $P \leq Q$ , 且属性集  $P$  比属性集  $Q$  下的缺失值程度小, 则  $H(X|P) < H(X|Q)$ 。

基于上述条件, 构造出非完备信息系统中知识的不确定性度量。

**定义6** 设  $IS = \langle U, V, f, A \rangle, P \subseteq A$ , 是一个非完备信息系统,  $X \subseteq U$ ,  $U/T(P) = \{T_P(u_1), T_P(u_2), \dots, T_P(u_{|U|})\}$ , 则属性集  $P$  在论域  $U$  上的信息熵定义为:

$$H_\alpha(P) = -\frac{2^\alpha}{|U|} \sum_{i=1}^{|U|} \frac{|T_P(u_i)|}{|U|} \log \frac{|T_P(u_i)|}{|U|} \quad (5)$$

式中:  $\alpha = \frac{\sum_{i=1}^{|U|} |Q(u_i)|}{|P|}$ ;  $Q(u_i) = \{a | a \in P \wedge a(u_i) = *\}$ ;  $u_i \in U$ ;

$i = 1, 2, \dots, |U|$ ;  $|Q(u_i)|$  表示对象  $u_i$  在条件属性描述

下缺失值的个数。以下出现的相同符号意义与此相同, 不再赘述。

基于上述分析, 构造出非完备信息系统中目标概念的不确定性度量。

**定义7** 设  $IS = \langle U, V, f, A \rangle, P \subseteq A$ , 是一个非完备信息系统,  $U/T(P) = \{T_P(u_1), T_P(u_2), \dots, T_P(u_{|U|})\}$ ,  $X \subseteq U$ , 则目标概念  $X$  在覆盖  $U/T(P)$  下的不确定性度量为:

$$H_\alpha(X|P) = -\frac{2^\alpha}{|U|} \sum_{i=1}^{|U|} \frac{|T_P(u_i) \cap X|}{|U|} \log \frac{|T_P(u_i) \cap X|}{|T_P(u_i)|} \quad (6)$$

定义7给出了非完备信息系统中目标概念的不确定性度量, 该度量不但反映了非完备信息系统中由于知识粒度及集合引起的不确定性, 还反映了由于属性值缺失所引起的不确定性。

**定理1** 设  $IS = \langle U, V, f, A \rangle, P \subseteq A$ , 是一个非完备信息系统,  $X \subseteq U$ , 那么  $H_\alpha(X|P)$  是定义5下的不确定性度量。

**证明** 式(6)是由2部分因子组成, 其中一个因子为  $H(X|P) = -\sum_{i=1}^{|U|} \frac{|T_P(u_i) \cap X|}{|U|} \log \frac{|T_P(u_i) \cap X|}{|T_P(u_i)|}$ , 另一个因子为  $G(\alpha) = \frac{2^\alpha}{|U|}$ 。

设  $Y_i = T_P(u_i) \cap X$ ,  $Y_i^c = T_P(u_i) \cap X^c$ ,  $x_i = |Y_i| = |T_P(u_i) \cap X|$ ,  $y_i = |Y_i^c| = |T_P(u_i) \cap X^c|$ ,  $X^c = U - X$ 。则

$$T_P(u_i) = Y_i \cup Y_i^c, |T_P(u_i)| = |Y_i \cup Y_i^c| = |Y_i| + |Y_i^c| = x_i + y_i$$

所以

$$H(X|P) = -\sum_{i=1}^{|U|} \frac{|T_P(u_i) \cap X|}{|U|} \log \frac{|T_P(u_i) \cap X|}{|T_P(u_i)|} = -\sum_{i=1}^{|U|} \frac{x_i}{|U|} \log \frac{x_i}{x_i + y_i}$$

易知函数  $f(x, y) = -x \log \frac{x}{x+y}$  在  $x \geq 0, y \geq 0$  范围内单调递增, 且  $f(0, 0) = 0$ , 则  $f(x, y) \geq 0$ , 故  $H(X|P) \geq 0$ 。

1) 非负性: 易知  $\alpha \geq 0$ , 所以  $G(\alpha) > 0$ 。又因为  $H(X|P) \geq 0$ , 得  $H_\alpha(X|P) \geq 0$ ;

对于任意  $T_P(u_i) \in U/T(P)$ , 当且仅当  $T_P(u_i) \cap X = \emptyset$  或  $T_P(u_i) \subseteq X$  时, 有  $H_\alpha(X|P) = 0$ 。

2) 不变性: 若  $\forall P, Q \subseteq A$ , 要使  $H(X|P) = H(X|Q)$ , 则必对覆盖  $U/T(P) = \{T_P(u_1), T_P(u_2), \dots, T_P(u_{|U|})\}$ ,  $U/T(Q) = \{T_Q(u_1), T_Q(u_2), \dots, T_Q(u_{|U|})\}$ , 有  $|T_P(u_i)| = |T_Q(u_i)|$  ( $\forall i \in \{1, 2, \dots, |U|\}$ ), 即

$$H(X|P) = -\sum_{i=1}^{|U|} \frac{|T_P(u_i) \cap X|}{|U|} \log \frac{|T_P(u_i) \cap X|}{|T_P(u_i)|} = -\sum_{i=1}^{|U|} \frac{|T_Q(u_i) \cap X|}{|U|} \log \frac{|T_Q(u_i) \cap X|}{|T_Q(u_i)|} = H(X|Q)$$

又必须在属性集  $P$  与属性集  $Q$  下的缺失值

程度相同, 即  $G(\alpha_P) = G(\alpha_Q)$ 。

则得到

$$H_\alpha(X|P) = G(\alpha_P) H(X|P) = G(\alpha_Q) H(X|Q) = H_\alpha(X|Q)$$

3) 单调性: 因为函数  $f(x, y) = -x \log(x/x+y)$  在区间  $x \geq 0, y \geq 0$  上单调递增, 且  $f(x, y) \geq 0$ , 所以式 (6) 中一个因子  $H(X|P) \geq 0$ , 且单调递增, 另一个因子为函数  $G(\alpha) = 2^\alpha > 0$ , 且单调递增, 所以  $H_\alpha(X|P) \geq 0$ , 且单调递增, 即满足单调性。

由此可知, 式 (6) 是非完备信息系统下的一种不确定性度量。

考虑非完备信息系统中目标概念的不确定程度, 可诱导出非完备决策系统中知识的不确定性度量。

**定义8** 设  $IDS = \langle U, C \cup D, V, f \rangle$ ,  $P \subseteq C, * \notin V_D, * \in V_C$  是一个非完备决策系统,  $U/T(P) = \{T_P(u_1), T_P(u_2), \dots, T_P(u_{|U|})\}$ ,  $X \subseteq U$ ,  $T_P(u_i)$  是在  $U$  上定义的相容关系下的相容类,  $U/D = \{Y_1, Y_2, \dots, Y_m\}$ 。则决策属性  $D$  相对属性  $P$  的条件熵定义为:

$$H(D|P) = -\frac{2^\alpha}{|U|} \sum_{i=1}^{|U|} \sum_{j=1}^m \frac{|T_P(u_i) \cap Y_j|}{|U|} \log \frac{|T_P(u_i) \cap Y_j|}{|T_P(u_i)|} \quad (7)$$

**性质1** 设  $IDS = \langle U, C \cup D, V, f \rangle$ ,  $P, Q \subseteq C, * \notin V_D, * \in V_C$  是一非完备决策系统, 其中  $U/T(P) = \{T_P(u_1), T_P(u_2), \dots, T_P(u_{|U|})\}$ ,  $U/T(Q) = \{T_Q(u_1), \dots, T_Q(u_{|U|})\}$ , 若  $|T_P(u_i)| = |T_Q(u_i)|$ , 且在属性集  $P$  与属性集  $Q$  下的缺失值程度相同, 即  $\alpha_P = \alpha_Q$ , 则  $H(D|P) = H(D|Q)$ 。

性质1表明本文构造的条件熵满足不确定性度量的不变性条件, 即若2个知识  $P, Q \subseteq C$ , 粗细相同, 且缺失值程度相同时, 它们具有相同的不确定性度量值, 这意味着知识粒度粗细程度及属性值缺失程度相同, 其不确定性也相同。

**性质2** 设  $IDS = \langle U, C \cup D, V, f \rangle$ ,  $P, Q \subseteq C, * \notin V_D, * \in V_C$  是一非完备决策系统, 其中  $U/T(P) = \{T_P(u_1), T_P(u_2), \dots, T_P(u_{|U|})\}$ ,  $U/D = \{Y_1, Y_2, \dots, Y_m\}$ 。决策属性  $D$  相对属性  $P$  的条件熵取值范围为  $0 \leq H(D|P) \leq 2^{\alpha_P} \log |U|$ 。

**证明** 1) 当  $T_P(u_i) \subseteq Y_i$  时,  $H(D|P)$  取最小值为0。

由  $T_P(u_i) \subseteq Y_i$  可知, 对任意的  $u_i \in U$ , 有

$$\frac{|T_P(u_i) \cap Y_j|}{|T_P(u_i)|} = \frac{|T_P(u_i)|}{|T_P(u_i)|} = 1$$

则

$$\log \frac{|T_P(u_i) \cap Y_j|}{|T_P(u_i)|} = 0$$

根据定义8得  $H(D|P) = 0$ 。反之, 若  $T_P(u_i) \not\subseteq Y_i$ , 则

$$\sum_{i=1}^{|U|} \sum_{j=1}^m \frac{|T_P(u_i) \cap Y_j|}{|U|} \log \frac{|T_P(u_i) \cap Y_j|}{|T_P(u_i)|} \neq 0$$

又  $2^\alpha$  始终大于零, 所以  $H(D|P) \neq 0$  产生矛盾,

因此  $H(D|P)$  可以取到最小值, 且为0。

2) 当  $T_P(u_i) = U, U/D = \{\{u_1\}, \{u_2\}, \dots, \{u_{|U|}\}\}$ , 且当所有条件属性值缺失时,  $H(D|P)$  达到最大值  $2^{\alpha_P} \log |U|$ 。

当所有条件属性值缺失时, 即  $|Q(u_i)| = |U| \times |P|$ , 得  $\alpha = |U|$ , 又因为  $T_P(u_i) = U, U/D = \{\{u_1\}, \{u_2\}, \dots, \{u_{|U|}\}\}$  所以, 利用式 (7) 得  $H(D|P) = 2^{|U|} \log |U|$ 。

因此, 决策属性  $D$  相对属性  $P$  的条件熵取值范围为  $0 \leq H(D|P) \leq 2^{\alpha_P} \log |U|$ 。

**性质3** 设  $IDS = \langle U, C \cup D, V, f \rangle$ ,  $P, Q \subseteq C, * \notin V_D, * \in V_C$  是一非完备决策系统, 其中  $U/T(P) = \{T_P(u_1), T_P(u_2), \dots, T_P(u_{|U|})\}$ ,  $U/D = \{Y_1, Y_2, \dots, Y_m\}$ 。对于任意  $u_i \in U, a \in C, (i \in 1, 2, \dots, |U|)$ , 将  $f(u_i, a) \neq *$  转化为  $f(u_i, a) = *$ , 得到改变后的非完备决策系统  $IDS'$ 。则改变后的非完备决策系统下决策属性  $D$  相对属性  $P$  的条件熵大于原决策系统下决策属性  $D$  相对属性  $P$  的条件熵, 即  $H'(D|P) > H(D|P)$ 。

**证明** 由于  $\forall u_i \in U, a \in C, (i \in 1, 2, \dots, |U|)$ , 将  $f(u_i, a) \neq *$  转化为  $f(u_i, a) = *$ , 则  $T_P(u_i) \subseteq T_{P'}(u_i)$ 。由定理1可知, 一因子  $H(X|P)$  随着  $T_P(u_i)$  变粗, 单调递增, 则  $H'(Y_j|P) \geq H(Y_j|P), (j = 1, 2, \dots, m)$ , 即

$$-\sum_{i=1}^{|U|} \frac{|T_{P'}(u_i) \cap Y_j|}{|U|} \log \frac{|T_{P'}(u_i) \cap Y_j|}{|T_P(u_i)|} \geq -\sum_{i=1}^{|U|} \frac{|T_P(u_i) \cap Y_j|}{|U|} \log \frac{|T_P(u_i) \cap Y_j|}{|T_P(u_i)|}, (j = 1, 2, \dots, m)$$

则

$$-\sum_{i=1}^{|U|} \sum_{j=1}^m \frac{|T_{P'}(u_i) \cap Y_j|}{|U|} \log \frac{|T_{P'}(u_i) \cap Y_j|}{|T_P(u_i)|} \geq -\sum_{i=1}^{|U|} \sum_{j=1}^m \frac{|T_P(u_i) \cap Y_j|}{|U|} \log \frac{|T_P(u_i) \cap Y_j|}{|T_P(u_i)|}$$

又因为改变后的非完备决策系统  $IDS'$  的缺失值增多, 即  $\alpha' > \alpha$ , 也即

$$\frac{2^{\alpha'}}{|U|} > \frac{2^\alpha}{|U|}$$

所以

$$H'(D|P) > H(D|P)$$

性质3表明由定义8构造的条件熵满足不确定性度量的单调性条件, 即若2个知识  $P, Q \subseteq C, P \leq Q$ , 且  $|Q_P(u_i)| < |Q_Q(u_i)|$ , 则决策属性  $D$  相对属性  $P$  的条件熵小于决策属性  $D$  相对属性  $Q$  的条件熵, 这意味着非完备决策信息系统存在的缺失值越多, 则条件熵越大, 不确定性也越大。

**性质4** 非完备决策系统  $IDS = \langle U, C \cup D, V, f \rangle$ , 退化为完备决策信息系统  $DS = \langle U, C \cup D, V, f \rangle$ , 其中  $P \subseteq C, U/T(P) = \{T_P(u_1), T_P(u_2), \dots, T_P(u_{|U|})\}$ ,  $U/D = \{Y_1, Y_2, \dots, Y_m\}$ 。则完备决策系统中决策属性  $D$  相对属



性  $P$  的条件熵为:

$$H_{\text{Ds}}(D|P) = - \sum_{i=1}^n \left( \frac{|X_i|}{|U|} \right)^2 \sum_{j=1}^m \frac{|X_i \cap Y_j|}{|X_i|} \log \frac{|X_i \cap Y_j|}{|X_i|} \quad (8)$$

**证明** 完备决策系统中无缺失值, 即  $\alpha = 0$ , 相容关系退化为等价关系, 则相容类  $T_P(u_i)$  退化为等价类  $[u_i]_P$ ,  $U/P = \{X_1, X_2, \dots, X_n\}$ , 因此,

$$\begin{aligned} H(D|P) &= - \frac{2^\alpha}{|U|} \sum_{i=1}^{|U|} \sum_{j=1}^m \frac{|T_P(u_i) \cap Y_j|}{|U|} \log \frac{|T_P(u_i) \cap Y_j|}{|T_P(u_i)|} = \\ &= - \frac{1}{|U|} \sum_{i=1}^{|U|} \sum_{j=1}^m \frac{|[u_i]_P \cap Y_j|}{|U|} \log \frac{|[u_i]_P \cap Y_j|}{|[u_i]_P|} = \\ &= - \frac{1}{|U|} \sum_{i=1}^n |X_i| \sum_{j=1}^m \frac{|X_i \cap Y_j|}{|U|} \log \frac{|X_i \cap Y_j|}{|X_i|} = \\ &= - \sum_{i=1}^n \frac{|X_i|}{|U|} \sum_{j=1}^m \frac{|X_i|}{|U|} \frac{|X_i \cap Y_j|}{|X_i|} \log \frac{|X_i \cap Y_j|}{|X_i|} = \\ &= - \sum_{i=1}^n \left( \frac{|X_i|}{|U|} \right)^2 \sum_{j=1}^m \frac{|X_i \cap Y_j|}{|X_i|} \log \frac{|X_i \cap Y_j|}{|X_i|} \end{aligned}$$

完备决策信息系统中属性  $D$  相对属性  $P$  的条件熵  $H_{\text{Ds}}(D|P)$  可以看成是以条件概率  $|X_i \cap Y_j|/|X_i|$  为自变量的函数值的加权平均的平方。

### 3 基于条件熵的多属性决策方法

多属性决策广泛存在于现实生活中, 它主要包括选择、分级及排序。排序是指根据属性信息将系统中的整个对象集形成全序。分级是指根据一定准则将对象划分成若干类, 若分类的个数等同于对象的个数时, 就将分级转化为排序。但在实际情况下, 决策结果多数以分级的形式存在, 因此可将决策系统中的分级看作初步决策, 进而利用条件熵计算决策系统的属性权重, 及对属性缺失值进行填充, 在此基础上, 通过计算加权平均值来获得细粒度的排序, 使得排序结果更合理。

#### 3.1 基于条件熵的属性权重确定方法

**定义 9** 设  $\text{IDS} = \langle U, C \cup \{d\}, V, f \rangle$  是一个非完备决策信息系统, 属性集  $C = \{a_1, a_2, \dots, a_m\}$ , 定义属性  $a_i \in C$  在属性集  $C$  中的重要度为:

$$\text{Sig}(a_i) = 1 - \frac{h_i}{\sum_{i=1}^m h_i} \quad (9)$$

其中  $h_i = H(d|a_i) (i = 1, 2, \dots, m)$ 。

基于条件熵的属性重要度定义具有如下性质:

**性质 5** 设  $\text{IDS} = \langle U, C \cup D, V, f \rangle, P, Q \subseteq C, * \notin V_D, * \in V_C$  是一非完备决策系统, 则对于任意属性  $a_i \in$

$C$ , 都有  $0 \leq \text{Sig}(a_i) \leq 1$ 。

**证明** 1) 由性质 2 可知,  $H(D|P)$  可以取最小值, 且为 0。则若  $\exists h_i = 0, (i = 1, 2, \dots, m)$  则决策属性  $d$  相对该属性  $a_i$  的条件熵为  $h_i = 0$ , 从而  $\text{Sig}(a_i) = 1$ 。

2) 要使  $\text{Sig}(a_i) = 0$ , 则必须  $h_i = \sum_{i=1}^m h_i$ , 即存在一个条件熵  $h_j \neq 0 (j = 1, 2, \dots, m)$ , 其余条件熵  $h_i = 0 (i \neq j)$ , 即属性  $a_i$  的重要度为 0, 则可对该决策系统的条件属性集进行属性约简。

由式 (9) 可知  $h_i \leq \sum_{i=1}^m h_i$ , 所以  $0 \leq \frac{h_i}{\sum_{i=1}^m h_i} \leq 1$ , 则  $0 \leq 1 - \frac{h_i}{\sum_{i=1}^m h_i} \leq 1$ , 即  $0 \leq \text{Sig}(a_i) \leq 1$ 。

**性质 6** 设  $\text{IDS} = \langle U, C \cup D, V, f \rangle, P, Q \subseteq C, * \notin V_D, * \in V_C$  是一非完备决策系统, 则对于  $\forall a_i \in C$ , 属性  $a_i \in C$  是必要的充分必要条件是  $\text{sig}(a_i) > 0$ 。

由以上性质知,  $\text{sig}(a_i)$  的取值越大, 则该属性在属性集  $C$  中越重要。

**定义 10** 设  $\text{IDS} = \langle U, C \cup \{d\}, V, f \rangle, * \notin V_D, * \in V_C$  是一个非完备决策信息系统, 属性集  $C = \{a_1, a_2, \dots, a_m\}$ , 属性  $a_i \in C$  的权重为:

$$\omega_i = \frac{\text{Sig}(a_i)}{\sum_{i=1}^m \text{Sig}(a_i)} \quad (10)$$

#### 3.2 基于条件熵的缺失值填充方法

设  $\text{IDS} = \langle U, C \cup \{d\}, V, f \rangle, * \notin V_D, * \in V_C$  是一个非完备决策信息系统, 其中  $U/d = \{Y_1, Y_2, \dots, Y_m\}$ , 则基于条件熵的非完备决策系统的填充方法叙述如下。

非完备决策系统中缺失值的平均填充值:

$$\text{FL}_{\text{avg}}(u_i, a_j) = \frac{\sum_{u_k \in U_j'} f(u_k, a_j)}{|U_j'|} \quad (11)$$

其中  $U_j' = \{f(u_k, a_j) | d(u_k) = d(u_i)\}$ 。

非完备决策系统中缺失值的众数填充值:

$$\text{FL}_{\text{mod e}}(u_i, a_j) = \text{mode}(V_{a_j}/\{*\}) \quad (12)$$

其中  $\text{mode}(V_{a_j}/\{*\})$  表示全体对象  $u_i$  在属性  $a_j$  下的所有已知值中出现次数最多的属性值。

$$\begin{aligned} \text{FL}_{\text{mod e-d}}(u_i, a_j) &= \text{mode}\{f(u_k, a_j) | d(u_k, a_j) = d(u_i, a_j) \\ &\quad u_k \in U\} \end{aligned} \quad (13)$$

利用  $\text{FI} = \text{FL}_{\text{avg}} \cup (\text{FL}_{\text{mod e}} \cap \text{FL}_{\text{mod e-d}})$  得到缺失值填充  $f(u_i, a_j) = *$  的所有可能值, 将所有可能值依次填充该缺失值, 然后分别计算填充后系统的条件熵, 比较度量值, 取最小度量值对应的填充

值,即

$$v^* = \arg \min_{v \in FI} H(d|C, v) \quad (14)$$

为该缺失值的填充值。若  $v^*$  有多个取值,则该缺失值的填充值为  $v^*$  的平均值。

### 3.3 基于条件熵的多属性决策方法

**算法** 基于条件熵的多属性排序方法

**输入** 非完备决策信息系统  $IDS = \langle U, C \cup \{d\}, V, f \rangle, * \notin V_D, * \in V_C$ , 其中  $U/d = \{Y_1, Y_2, \dots, Y_m\}$ 。

**输出** 全序化结果  $I$ 。

1) 基于相容关系, 计算每个对象在每个属性  $a_i \in C$  下的相容类:

$$T(a) = \{(u_i, u_j) \in U \times U \mid a(u_i) = a(u_j) \vee a(u_i) = * \vee a(u_j) = *, a \in C\}, T_a(u_i) = \{u_j \in U \mid (u_i, u_j) \in T(a), \forall a \in C\}$$

2) 利用条件熵的定义, 计算决策属性  $d$  相对每个属性  $a_i \in C$  的条件熵

$$H(d|a) = -\frac{2^\alpha}{|U|} \sum_{i=1}^{|U|} \sum_{j=1}^m \frac{|T_a(u_i) \cap Y_j|}{|U|} \log \frac{|T_a(u_i) \cap Y_j|}{|T_a(u_i)|}$$

3) 利用基于条件熵的属性权重方法, 根据式 (9)、(10) 计算每个属性  $a_i \in C$  分别在属性集  $C$  中的重要度及权重。

4) 利用式 (11)~(14) 对非完备决策系统进行填充。

5) 在填充后的系统中, 计算每个对象在属性集  $C$  上的加权求和, 即

$$D_C(u_i) = \sum_{j=1}^{|C|} \omega_j f(u_i, a_j) \quad (15)$$

## 4 案例分析

以某鉴定机构对危旧房屋评估受损水平的多属性决策问题为例进行分析。鉴定机构负责人派出 2 位专家根据房屋装修质量  $a_1$ 、房屋安全出口结构安全性  $a_2$ 、房屋结构可靠性  $a_3$  和房屋主体构件耐久性  $a_4$ , 将 10 座房屋进行了初步评定, 决策属性  $d$  表示评测等级,  $U/d = \{\{x_1, x_4, x_8\}, \{x_2\}, \{x_3, x_6, x_9\}, \{x_5, x_7, x_{10}\}\}$ , 见表 3, 表中 "\*" 表示该房屋在某属性下的属性值是存在的, 只是由于数据暂时无法得到而被遗漏了。现根据非完备多属性排序方法的算法将 10 座房屋进行降序排列。

1) 计算 10 个对象在属性  $a_1$  下的相容类:

$$T_{a_1}(x_1) = \{x_1, x_5, x_6, x_8, x_9\}, T_{a_1}(x_2) = \{x_2, x_3, x_5, x_6, x_{10}\}$$

$$T_{a_1}(x_3) = \{x_2, x_3, x_5, x_6, x_{10}\}, T_{a_1}(x_4) = \{x_4, x_5, x_6, x_7\}$$

$$T_{a_1}(x_5) = \{x_1, x_2, \dots, x_{10}\}, T_{a_1}(x_6) = \{x_1, x_2, \dots, x_{10}\}$$

$$T_{a_1}(x_7) = \{x_4, x_5, x_6, x_7\}, T_{a_1}(x_8) = \{x_1, x_5, x_6, x_8, x_9\}$$

$$T_{a_1}(x_9) = \{x_1, x_5, x_6, x_8, x_9\}, T_{a_1}(x_{10}) = \{x_2, x_3, x_5, x_6, x_{10}\}$$

表 3 一个具有非完备评价信息的房屋评测表

Table 3 A house evaluation form with incomplete decision information

房屋	装修质量 $a_1$	结构安全 $a_2$	结构可靠性 $a_3$	构件耐久性 $a_4$	评测等级
$x_1$	2	2	3	3	优
$x_2$	1	1	1	*	差
$x_3$	1	*	2	2	中
$x_4$	3	3	*	1	优
$x_5$	*	2	3	2	良
$x_6$	*	1	1	3	中
$x_7$	3	1	2	*	良
$x_8$	2	3	*	2	优
$x_9$	2	1	1	3	中
$x_{10}$	1	2	3	*	良

类似地, 可得到所有对象在其他 3 个属性  $a_2$ 、 $a_3$ 、 $a_4$  下对应的相容类。

2) 分别计算决策属性  $d$  相对每个属性  $a_1$ 、 $a_2$ 、 $a_3$ 、 $a_4$  的条件熵为

$$|Q(x_i)| = 2, \alpha = \frac{|Q(x_i)|}{|a_1|} = 2, h_1 = H(d|a_1) = 3.8227$$

$$|Q(x_i)| = 1, \alpha = \frac{|Q(x_i)|}{|a_2|} = 1, h_2 = H(d|a_2) = 1.3977$$

$$|Q(x_i)| = 2, \alpha = \frac{|Q(x_i)|}{|a_3|} = 2, h_3 = H(d|a_3) = 3.4921$$

$$|Q(x_i)| = 3, \alpha = \frac{|Q(x_i)|}{|a_4|} = 3, h_4 = H(d|a_4) = 10.3726$$

3) 由式 (9) 分别计算属性  $a_1$ 、 $a_2$ 、 $a_3$ 、 $a_4$  在属性集  $C$  中的属性重要度为

$$\text{Sig}(a_1) = 1 - \frac{h_1}{\sum_{i=1}^4 h_i} = 0.799, \text{Sig}(a_2) = 1 - \frac{h_2}{\sum_{i=1}^4 h_i} = 0.927$$

$$\text{Sig}(a_3) = 1 - \frac{h_3}{\sum_{i=1}^4 h_i} = 0.817, \text{Sig}(a_4) = 1 - \frac{h_4}{\sum_{i=1}^4 h_i} = 0.457$$

由式 (10) 计算各属性权重分别为

$$\omega_{a_1} = 0.266, \omega_{a_2} = 0.309, \omega_{a_3} = 0.272, \omega_{a_4} = 0.152$$

4) 利用条件熵方法对非完备决策系统进行填充:

首先得到房屋评测系统中  $f(x_4, a_3)$ ,  $f(x_8, a_3)$ ,  $\dots$ ,  $f(x_3, a_2)$  的值缺失, 然后由式 (11)~(13) 计算每个缺失值的所有可能值。

$$\text{FL}_{\text{avg}}(u_5, a_1) = \{2\}, \text{FL}_{\text{mod } e-d}(u_5, a_1) = \{1, 3\}$$

$$\text{FL}_{\text{mod } e}(u_5, a_1) = \{1, 2\}$$

由此可得  $Fl(u_5, a_1) = \{1, 2\}$ , 即  $f(x_5, a_1) = 1$  或  $f(x_5, a_1) = 2$ 。

类似地, 可得其他 7 个缺失值填充的所有可能值:

$$f(x_6, a_1) = 1.5 \text{ 或 } f(x_6, a_1) = 1 \text{ 或 } f(x_6, a_1) = 2$$

$$f(x_3, a_2) = 1$$

$$f(x_4, a_3) = 3$$

$$f(x_8, a_3) = 3$$

$$f(x_2, a_4) = 2 \text{ 或 } f(x_2, a_4) = 3$$

$$f(x_7, a_4) = 2$$

$$f(x_{10}, a_4) = 2$$

所以以下对  $f(x_5, a_1)$ 、 $f(x_6, a_1)$  和  $f(x_2, a_4)$ , 利用式 (14) 进行选择。

1) 当  $f(x_5, a_1) = 2$  及  $f(x_5, a_1) = 1$  时, 分别利用式 (7) 计算填充该缺失值后系统的条件熵, 并比较条件熵值。

计算当  $f(x_5, a_1) = 1$  时, 每个对象在属性集  $C$  下的相容类为  $T_C^1(u_i)$ , 当  $f(x_5, a_1) = 2$  时, 每个对象在属性集  $C$  下的相容类为  $T_C^2(u_i)$ , 则  $T_C^1(u_i) \subset T_C^2(u_i) \subset Y_j, i = 5, 10$ , 其余对象的相容类  $T_C^2(u_i) = T_C^1(u_i)$ , 则  $H^1(d|C) = H^2(d|C)$ 。因此, 根据式 (14) 得  $v_1^* = 1, v_2^* = 2$ , 故  $f(x_5, a_1) = 1.5$ 。

2) 类似地, 分别将  $f(x_6, a_1) = 1.5$ 、 $f(x_6, a_1) = 1$  及  $f(x_6, a_1) = 2$  填充该系统, 利用式 (7) 计算填充后系统的条件熵 ( $T_C^1(u_i)$ 、 $T_C^2(u_i)$  和  $T_C^3(u_i)$  分别表示 3 个可能值填充后系统中每个对象在属性集  $C$  下的相容类), 即

$$H^1(d|C) = -\frac{2^a}{|U|} \sum_{i=1}^{|U|} \sum_{j=1}^m \frac{|T_C^1(u_i) \cap Y_j|}{|U|} \log \frac{|T_C^1(u_i) \cap Y_j|}{|T_C^1(u_i)|} = 0$$

$$H^2(d|C) = -\frac{2^a}{|U|} \sum_{i=1}^{|U|} \sum_{j=1}^m \frac{|T_C^2(u_i) \cap Y_j|}{|U|} \log \frac{|T_C^2(u_i) \cap Y_j|}{|T_C^2(u_i)|} = -\frac{2^{1.75}}{10} \times (4 \times \frac{1}{10} \times \log 2(\frac{1}{2})) = 0.1345$$

$$H^3(d|C) = -\frac{2^a}{|U|} \sum_{i=1}^{|U|} \sum_{j=1}^m \frac{|T_C^3(u_i) \cap Y_j|}{|U|} \log \frac{|T_C^3(u_i) \cap Y_j|}{|T_C^3(u_i)|} = 0$$

比较 3 个度量值, 由式 (14) 得  $v_1^* = 1.5, v_2^* = 2$ , 故  $f(x_6, a_1) = 1.75$ 。

3) 当  $f(x_2, a_4) = 2$  及  $f(x_2, a_4) = 3$  时, 利用式 (7) 计算填充该缺失值后系统的条件熵, 并比较度量值。

由相容关系计算得: 当  $f(x_2, a_4) = 2$  时, 每个对象在属性集  $C$  下的相容类为  $T_C^1(u_i)$ ,  $f(x_2, a_4) = 3$  时, 每个对象在属性集  $C$  下的相容类为  $T_C^2(u_i)$ , 得到  $T_C^1(u_i) \subseteq T_C^2(u_i)$ , 则  $H^1(d|C) < H^2(d|C)$ 。因此, 由式 (14) 得到  $f(x_2, a_4) = 2$ 。填充结果见表 4。

表 4 完备化表 3 后的房屋评测表  
Table 4 An evaluation form by complementing Table 3

$U$	$a_1$	$a_2$	$a_3$	$a_4$	$d$
$x_1$	2	2	3	3	优
$x_2$	1	1	1	2	差
$x_3$	1	1	2	2	中
$x_4$	3	3	3	1	优
$x_5$	1.5	2	3	2	良
$x_6$	1.75	1	1	3	中
$x_7$	3	1	2	2	良
$x_8$	2	3	3	2	优
$x_9$	2	1	1	3	中
$x_{10}$	1	2	3	2	良

在填充后的系统中, 根据式 (15) 及 3) 的属性权重, 计算非完备决策信息系统中每个对象的加权平均值, 则取值为:

$$D_C(x_1) = 2.422, D_C(x_2) = 1.151, D_C(x_3) = 1.423$$

$$D_C(x_4) = 2.693, D_C(x_5) = 2.137, D_C(x_6) = 1.503$$

$$D_C(x_7) = 1.955, D_C(x_8) = 2.579, D_C(x_9) = 1.569$$

$$D_C(x_{10}) = 2.004$$

最终全序化结果为:

$$x_4 > x_8 > x_1 > x_5 > x_{10} > x_7 > x_9 > x_6 > x_3 > x_2$$

在初始评测等级中, 优、良、中、差分别对应的分类为  $\{x_1, x_4, x_8\}, \{x_5, x_7, x_{10}\}, \{x_3, x_6, x_9\}, \{x_2\}$ , 根据排序结果可知, 该 4 类按照等级高低依次排出。房屋 4 的初始评测为优, 且在排序结果中排在第 1 个, 说明房屋 4 在 4 个评测指标下的综合评测是处于最优的。本文方法将 10 座不同房屋区分开, 得到合理有效的全序化结果。

为了进一步说明本文基于条件熵的多属性决策方法的可行性, 现将本文所提方法与其他排序方法在 2 个完备信息和 2 个非完备信息的算例中进行比较分析。由于本文需要进行初步决策, 所以将其他排序方法获得的结果按照 1:1:1 的比例分成 3 类或 4 类视为分级结果, 进而利用本文排序方法将分级转化为排序。对比结果如表 5~8 所示。

表 5 本文方法与文献 [24] 方法的排序结果比较  
Table 5 Comparison of the ranking results between the proposed method and literature [24]

排序方法	文献[24]	本文方法
排序结果	$x_1 > x_{12} > x_{10} > x_3 >$	$x_1 > x_{12} > x_{10} > x_4 >$
	$x_5 > x_{15} > x_4 > x_{14} >$	$x_5 > x_{14} > x_3 > x_{11} >$
	$x_2 > x_9 > x_8 > x_{11} >$	$x_{15} > x_9 > x_2 > x_8 >$
	$x_{13} > x_6 > x_7$	$x_{13} > x_6 > x_7$



文献[24]采用熵权法确定属性权重,并基于优势关系的排序方法将具有完备信息的对象进行排序,获得的结果与本文获得的排序结果大致相同,但仍然存在主观因素。条件熵可用于评估属性的重要性,因此本文根据条件熵越小,属性越重要的特征确定属性权重,更客观、更合理地解决排序问题。

表6 本文方法与文献[25]方法的排序结果比较

Table 6 Comparison of the ranking results between the proposed method and literature [25]

排序方法	文献[25]	本文方法
排序结果	$x_6 > x_3 > x_1 > x_2 > x_4 \sim x_5$	$x_6 > x_3 > x_1 > x_2 > x_5 > x_4$

文献[25]提出一种新的排序模型将对象排序,然而利用该模型得到的排序结果存在明显的“并列”现象。本文在获得的排序结果基本一致的情况下,既考虑了条件属性与决策属性之间建立的关联性,又将并列的对象区分开,使得到的排序结果更合理。

表7 本文方法与文献[26]方法的排序结果比较

Table 7 Comparison of the ranking results between the proposed method and literature [26]

排序方法	文献[26]	本文方法
排序结果	$x_4 > x_1 > x_5 > x_3 > x_6 > x_2$	$x_4 > x_1 > x_5 > x_3 > x_6 > x_2$

文献[26]基于 $\alpha$ -先验概率优势关系的对象排序方法获得的排序结果相同,虽然该方法较好地解决了非完备信息系统的排序问题,但是未考虑属性在决策方法中的重要性。因此,本文基于条件熵的多属性决策方法更客观、合理。

表8 本文方法与文献[27]方法的排序结果比较

Table 8 Comparison of the ranking results between the proposed method and literature [27]

排序方法	文献[27]	本文方法
排序结果	$x_1 > x_7 > x_8 > x_4 > x_5 > x_{10} > x_9 > x_6 > x_3 > x_2$	$x_4 > x_8 > x_1 > x_7 > x_5 > x_{10} > x_9 > x_6 > x_3 > x_2$

文献[27]提出基于加权 $\alpha$ 优势关系的排序方法,较好地解决了非完备数据的排序问题,但是由于 $\alpha$ 取值的主观性,得到的排序结果不稳定。本文方法不仅使用客观赋权法确定属性权重避免主观性,还利用使系统条件熵最小原理进行缺失值填充,进而将对象区分开,获得合理、有效的全序化结果。

对比分析:

1) 从排序方法上看,由于条件属性集中每个

属性的重要度不同,本文利用提出的新条件熵计算决策系统的属性权重,能够避免排序问题中赋权方法带来的主观性。对于属性值缺失的情形,本文以最小条件熵为准则对属性缺失值进行填充,并利用基本的加权求和方法获得细粒度的排序,能够避免计算量较大的问题。

2) 从排序结果上看,本文提出基于条件熵的多属性决策排序方法,能够有效地解决由分级结果的粗粒度到排序结果的细粒度的转换,获得合理的排序结果。

## 5 结束语

本文首先针对非完备决策系统中属性值缺失导致系统不确定性的问题,构造一种新型的考虑条件属性缺失度的目标概念条件熵与决策知识条件熵。性质分析表明本文所提出的条件熵在体现不确定性时会更加敏感,是一种合理的不确定性度量。具有当系统缺失值增加时,新的条件熵增大;当知识粒度变细时,新条件熵随之减小的性质。其次,基于本文所提出的新的条件熵,设计了基于条件熵的属性权重确定及最小条件熵非完备属性取值补充方法,以解决属性权重完全未知的非完备多属性决策问题。最后通过与其他排序方法比较分析,说明了该方法能有效利用粗粒度的分级信息,获得更加合理有效的细粒度的排序结果。

## 参考文献:

- [1] PAWLAK Z. Rough sets[J]. International journal of computer and information science, 1982, 11(5): 341-356.
- [2] 苗夺谦, 张清华, 钱宇华, 等. 从人类智能到机器实现模型——粒计算理论与方法[J]. 智能系统学报, 2016, 11(6): 743-757.  
MIAO Duoqian, ZHANG Qinghua, QIAN Yuhua, et al. From human intelligence to machine implementation model: theories and applications based on granular computing[J]. CAAI transactions on intelligent systems, 2016, 11(6): 743-757.
- [3] 梁吉业, 钱宇华, 李德玉, 等. 大数据挖掘的粒计算理论与方法[J]. 中国科学(信息科学), 2015, 45(11): 1355-1369.  
LIANG Jiye, QIAN Yuhua, LI Deyu, et al. Theory and method of granular computing for big data mining[J]. Scientia sinica informationis, 2015, 45(11): 1355-1369.
- [4] 王国胤, 张清华, 马希骛, 等. 知识不确定性问题的粒计算模型[J]. 软件学报, 2011, 22(4): 676-694.  
WANG Guoyin, ZHANG Qinghua, MA Xi'ao, et al. Gran-

- ular computing models for knowledge uncertainty[J]. Journal of software, 2011, 22(4): 676–694.
- [5] PAWLAK Z. Vagueness and uncertainty: a rough set perspective[J]. *Computational intelligence*, 1995, 11(2): 227–232.
- [6] DÜNTSCH I, GEDIGA G. Uncertainty measures of rough set prediction[J]. *Artificial intelligence*, 1998, 106(1): 109–137.
- [7] WIERMAN M J. Measuring uncertainty in rough set theory[J]. *International journal of general systems*, 1999, 28(4/5): 283–297.
- [8] CHAKRABARTY K, BISWAS R, NANDA S. Fuzziness in rough sets[J]. *Fuzzy sets and systems*, 2000, 110(2): 247–251.
- [9] 苗夺谦, 王珏. 粗糙集理论中知识粗糙性与信息熵关系的讨论 [J]. 模式识别与人工智能, 1998, 11(1): 34–40.  
MIAO Duoqian, WANG Jue. On the relationships between information entropy and roughness of knowledge in rough set theory[J]. *Pattern recognition and artificial intelligence*, 1998, 11(1): 34–40.
- [10] LIANG Jiye, WANG Junhong, QIAN Yuhua. A new measure of uncertainty based on knowledge granulation for rough sets[J]. *Information sciences*, 2009, 179(4): 458–470.
- [11] 王国胤, 于洪, 杨大春. 基于条件信息熵的决策表约简 [J]. 计算机学报, 2002, 25(7): 759–766.  
WANG Guoyin, YU Hong, YANG Dachun. Decision table reduction based on conditional information entropy[J]. *Chinese journal of computers*, 2002, 25(7): 759–766.
- [12] 黄国顺, 文翰. 基于严凹函数的粗糙集不确定性度量 [J]. 软件学报, 2018, 29(11): 3484–3499.  
HUANG Guoshun, WEN Han. Uncertainty measures of rough set based on strictly concave functions[J]. *Journal of software*, 2018, 29(11): 3484–3499.
- [13] 黄国顺, 曾凡智, 陈广义, 等. 基于严凸函数的知识粒度与相对粒度 [J]. 模式识别与人工智能, 2013, 26(10): 897–908.  
HUANG Guoshun, ZENG Fanzhi, CHEN Guangyi, et al. Knowledge granularity and relative granularity based on strictly convex function[J]. *Pattern recognition and artificial intelligence*, 2013, 26(10): 897–908.
- [14] LIANG J, SHI Z, LI D, et al. Information entropy, rough entropy and knowledge granulation in incomplete information systems[J]. *International journal of general systems*, 2006, 35(6): 641–654.
- [15] DAI Jianhua, XU Qing. Approximations and uncertainty measures in incomplete information systems[J]. *Information sciences*, 2012, 198: 62–80.
- [16] QIAN Yuhua, LIANG Jiye, WANG Feng. A new method for measuring the uncertainty in incomplete information systems[J]. *International journal of uncertainty, fuzziness and knowledge-based systems*, 2009, 17(6): 855–880.
- [17] QIAN Yuhua, LIANG Jiye, DANG Chuangyin. Consistency measure, inclusion degree and fuzzy measure in decision tables[J]. *Fuzzy sets and systems*, 2008, 159(18): 2353–2377.
- [18] DAI Jianhua, XU Qing, WANG Wentao, et al. Conditional entropy for incomplete decision systems and its application in data mining[J]. *International journal of general systems*, 2012, 41(7): 713–728.
- [19] DAI Jianhua, WANG Wentao, XU Qing. An uncertainty measure for incomplete decision tables and its applications[J]. *IEEE transactions on cybernetics*, 2013, 43(4): 1277–1289.
- [20] YANG Jiping, QIU Wanhua. A measure of risk AND A decision-making model based on expected utility and entropy[J]. *European journal of operational research*, 2005, 164(3): 792–799.
- [21] 吕跃进, 张旭娜, 韦碧鹏. 基于优势关系粗糙集的模糊综合评价的权重确定 [J]. 统计与决策, 2012(20): 44–46.  
LV Yuejin, ZHANG Xuna, WEI Bipeng. Rough set based on dominance relation of fuzzy comprehensive evaluation of the weight[J]. *Control and decision*, 2012(20): 44–46.
- [22] 李佳, 梁吉业, 庞天杰. 一种基于优势粗糙集的多属性决策排序方法 [J]. 南京大学学报(自然科学), 2016, 52(5): 844–852.  
LI Jia, LIANG Jiye, PANG Tianjie. A sorting method of multi-attribute decision making based on dominance rough set theory[J]. *Journal of Nanjing University (Natural Sciences)*, 2016, 52(5): 844–852.
- [23] KRYSZKIEWICZ M. Rough set approach to incomplete information systems[J]. *Information sciences*, 1998, 112(1/2/3/4): 39–49.
- [24] 王利东, 田晓娟, 杨艳冰. 基于熵权与优势关系的教学效果评价方案 [J]. 数学的实践与认识, 2014, 44(10): 8–12.  
WANG Lidong, TIAN Xiaojuan, YANG Yanbing. The comprehensive evaluation of teaching based on entropy weight and dominance relation[J]. *Mathematics in practice and theory*, 2014, 44(10): 8–12.
- [25] 翁世洲, 吕跃进, 莫京兰. 基于优势关系的排序模型及

其保序性约简理论 [J]. 广西师范大学学报 (自然科学版), 2013, 31(3): 37–44.

WENG Shizhou, LV Yuejin, MO Jinglan. Ranking model and its order-preserving reduction theory based on dominance relations[J]. Journal of Guangxi Normal University (Natural Science Edition), 2013, 31(3): 37–44.

- [26] 施玉杰, 杨宏志, 徐久成. 先验概率优势关系下的粗糙集模型研究 [J]. 南京大学学报 (自然科学版), 2016, 52(5): 899–907.

SHI Yujie, YANG Hongzhi, XU Jiucheng. The study of rough set model under  $\alpha$ -prior probability dominance relation[J]. Journal of Nanjing University (Natural Sciences), 2016, 52(5): 899–907.

- [27] 李佳, 梁吉业, 庞天杰. 基于加权  $\alpha$  优势关系的多属性决策排序方法 [J]. 模式识别与人工智能, 2017, 30(8): 761–768.

LI J, LIANG Jiye, PANG Tianjie. Sorting method of multi-attribute decision making based on weighted  $\alpha$  dominance relation[J]. Pattern recognition and artificial intelligence, 2017, 30(8): 761–768.

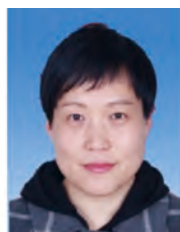
## 作者简介:



高文华, 女, 1994 年生, 硕士研究生, 主要研究方向为智能计算与数据建模。



梁吉业, 男, 1962 年生, 教授, 博士生导师, 主要研究方向为粒计算、数据挖掘与机器学习。主持国家 863 计划项目 2 项、国家自然科学基金项目 7 项 (其中重点项目 2 项)、国家 973 计划前期研究专项 1 项, 获山西省自然科学一等奖 2 项, 2014—2018 入选爱思唯尔中国高被引学者榜单。发表学术论文 200 余篇。



王宝丽, 女, 1982 年生, 副教授, 博士, 主要研究方向为粒计算、智能决策。主持国家自然科学基金项目 1 项、参与国家级项目 2 项、省部级科研及教研项目多项。发表学术论文 20 余篇。