

基于多粒度的多源决策信息系统的最优选择

石孝安^{1,2*}, 傅丽³, 冶忠林², 王伟杰²

¹青海职业技术大学公共教育学院, 青海 西宁

²青海师范大学计算机学院, 青海 西宁

³青海民族大学数学与统计学院, 青海 西宁

收稿日期: 2025年10月6日; 录用日期: 2025年10月31日; 发布日期: 2025年11月6日

摘要

多粒度思想是近些年来处理复杂数据的常用方法, 受多粒度思维的启发, 本文结合多粒度标记信息系统的概念, 提出了多源决策信息系统的广义决策函数的定义; 然后利用广义决策函数讨论了协调和不协调的多源决策信息系统的最优选择以及局部最优选择, 并讨论了局部相对约简。同时, 给出了相应的例子来验证所提出方法的可行性。

关键词

多粒度, 多源信息系统, 属性约简, 最优选择

Optimal Selection of Multi-Source Decision Information Systems Based on Multi-Granularity

Xiaoan Shi^{1,2*}, Li Fu³, Zhonglin Ye², Weijie Wang²

¹School of Public Education, Qinghai Vocational and Technical University, Xining Qinghai

²The College of Computer, Qinghai Normal University, Xining Qinghai

³School of Mathematics and Statistics, Qinghai Minzu University, Xining Qinghai

Received: October 6, 2025; accepted: October 31, 2025; published: November 6, 2025

Abstract

The multi-granularity thinking is a common method for handling complex data in recent years. Inspired by the multi-granularity thinking, this paper combines the concept of multi-granularity

*通讯作者。

文章引用: 石孝安, 傅丽, 冶忠林, 王伟杰. 基于多粒度的多源决策信息系统的最优选择[J]. 交叉科学快报, 2025, 9(6): 909-918. DOI: 10.12677/isl.2025.96115

labeled information systems and proposes the definition of the generalized decision function of multi-source decision information systems. Then, the generalized decision function was used to discuss the optimal selection and local optimal selection of coordinated and uncoordinated multi-source decision information systems, and the local relative reduction was also discussed. Meanwhile, corresponding examples were given to verify the feasibility of the proposed method.

Keywords

Multi-Granularity, Multi-Source Information System, Attribute Reduction, Optimal Selection

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着大数据时代的快速发展,通信和网络都可为人们提供大量的信息,如何利用已掌握的知识,面对不断变化的环境与条件进行决策等问题是我们面临的重大难题。作为数据库的抽象信息系统在实际生活中大量存在。数据库是非常庞大的,很难直接从中得到我们想要的信息,因此,在海量的信息中,提取知识,识别出正确,新颖,有潜在应用价值的最终可为人们理解的模式,就是我们研究的重点。

粗糙集是1982年英国数学家pawlak提出的理论[1],该理论是一种处理不确定的,不精确信息的数学工具。自提出以来,得到了很多学者的关注和研究,并已应用到了决策分析模式识别数据挖掘等领域[2][3]。

多源数据作为复杂数据之一,如何简化,多源信息系统的信息融合也是近几年研究的一个热点。在文献[4]-[6],分别讨论了多粒度粗糙集与证据理论相结合的信息融合方法、基于矩阵的多粒度融合方法在多源信息中动态更新知识、多源不完全区间值信息系统中信息源和属性变化的动态融合等内容。在文献[7]中,作者研究了多源信息系统中的决策规则挖掘。有时我们得到的信息源有可能不仅是不完备的而且并且具有一定程度的不确定性和模糊性,针对模糊且不完备的信息源如何进行融合,在文献[8],作者将多源不完备信息系统的融合方法应用到模糊多源不完备信息系统中。从中可以看出,由于大数据时代数据的多源异构特点,使得对多源数据的应用开发,合理选择,进行信息融合等问题成了这些年研究的热点之一。

多源数据作为复杂数据之一,是指从同一时间从不同源头获取的信息,对于很多决策者而言,从多方面获取的数据比从单一渠道获得的数据更加全面,实用和可靠。但由于多源数据的多而复杂,无法从中直接获取想要的信息。因此,这些年对多源数据的信息融合研究居多,不同的考虑结果下会有不同的信息融合结果,如文献[9]-[11]分别考虑了焦点信念权重的多源信息融合,一种用于基于方面的情绪分析的新型多信息融合与交互神经网络,多源信息融合中使用非参数合理性Relief算法评估信息源的可靠性。这些都是在不同的背景和数据系统中进行的信息融合,所采用的方法也各不相同。那么对于多源数据及多源信息系统的处理,找到一个简便而有效的信息融合方法就尤为重要。

在文献[12]中,作者基于多粒度思想,讨论了多源数据的知识获取,在此文中作者提出信息融合是将多源信息系统转化为单源信息系统,针对多粒度标记决策信息系统,主要研究的是最优粒度的选择,指依据不同的目标从多粒度标记信息系统中选择一个能达到某种预设要求的信息系统。对此,有很多学者在不同的信息系统做了最优选择的研究,如在文献[13]中,作者讨论了协调的不完备多粒度标记决策系统

的最优粒度选择。文献[14], 作者讨论了多尺度信息系统的粒度。在文献[15]中, 作者假设每个属性粒度级别相同, 研究了多粒度标记决策表的最优粒度选择。文献[16]讨论了多决策的多粒度标记信息系统的最佳选择。基于同样的假设, 学者们在文献[17]-[20]中进一步讨论了, 多粒度粗糙集与证据理论相结合的信息融合方法, 多粒度标记决策信息系统的规则提取方法和知识获取算法, 以及粗糙集的最优近似等问题。这些研究都是基于多粒度粗糙集的知识获取进行了信息融合。在文献[21]-[24]中, 学者们在不同的领域用多粒度思想解决了多源数据及信息系统的处理。

事实上, 最优粒度的选择是指从多个属性取值有规律的信息系统中选择一个满足问题求解的信息系统。多源信息系统的信息融合和多粒度标记决策信息系统的最佳粒度选择, 都是基于特定问题的求解需要, 选择一个或构造一个满足预定要求的信息系统。因此, 我们基于多粒度的思想讨论多源信息系统的最佳信息源选择, 以便更方便, 快捷地得到所需要的信息。

经过研读上述有关多粒度和不同信息系统的相关文献, 发现上述研究还可以对不同的信息系统应用多粒度思想进一步地进行拓展和研究。

下面我们引入相关定理及概念。

2. 基础知识

定义 2.1 [3] 称 (U, A, F) 是一个信息系统, 其中 U 为对象集, 即

$$U = \{x_1, x_2, \dots, x_n\},$$

U 中的每个元素 $x_i (i \leq n)$ 称为一个对象。 A 为属性集, 即

$$A = \{a_1, a_2, \dots, a_m\},$$

A 中的每个元素 $a_l (l \leq m)$ 称为一个属性。

F 为 U 与 A 之间的关系集, 即

$$F = \{f_l : U \rightarrow V_l (l \leq m)\},$$

其中 V_l 为 $a_l (l \leq m)$ 值域。

定义 2.2 [4] 设 $MIS = \{IS_k \mid IS_k = (U, A_k, F_k), k = 1, 2, \dots, s\}$ 是多源信息系统。其中, $IS_k = (U, A_k, F_k)$ 为单源信息系统 F_k 为 U 和 A_k 之间的一个关系集, 即 $F_k = \{f_{lk} : U \rightarrow V_{lk} (l_k \leq |A_k|)\}$, 其中, V_{lk} 是 a_{l_k} 的值域 $a_{l_k} \in A_k$ 。

定义 2.3 [13] 在经典的信息系统 (U, C) 中, 每个对象 x_i 在属性 a_j 上只取一个确定的值, 这就是单粒度信息系统。如果信息系统 (U, C) 中的每个对象在同一个属性上, 根据不同的粒度层面取不同的值, 则信息系统 (U, C) 是一个多粒度信息系统。

定义 2.4 [16] 设 $dIS = (U, C \cup d)$ 为多粒度标记决策信息系统, 对任意的 $C^k \subseteq C, x \in U$, 关于 C^k 的广义决策函数为 $\partial_{C^k}(x) = \{d(y) \mid y \in [x]_{C^k}\}$ 。若 $R_{C^k} \subset R_d$, 则 dIS 协调, 否则不协调; 若进一步满足 $R_{C^k} \subset R_d$, 则称 $S^k = (U, C^k \cup d)$ 协调。当 dIS 不协调时, 若 $\partial_{C^k}(x) = \partial_{C^l}(x)$, 则称 dIS^k 在 dIS 中是协调的。

定义 2.5 [16] 设一个多粒度标记决策信息系统为 $dIS = (U, C \cup d)$, 若 dIS 协调, 则第 k 个粒度层次最优当且仅当 dIS^k 协调且若存在 $k+1 \leq l$ 则 dIS^{k+1} 必不协调; 若 dIS 不协调, 第 k 个水平粒度最优当且仅当 dIS^k 在 dIS 是协调的而且若有 $k+1 \leq l$, dIS^{k+1} 必不协调。

定义 2.6 [17] 在不完备多粒度决策系统 $S = (C, U \cup d)$ 中, 对于给定的 $k (1 \leq k \leq l)$, 若 $S^k = (U, C^k \cup d)$ 是协调的, 而 $S^{k+1} = (U, C^{k+1} \cup d)$ 是不协调的, 则称 k 层粒度是全局最优的粒度。若 $S^l = (U, C^l \cup d)$ 是协调的, 则称 l 层粒度是全局最优粒度。

对于 $x \in U$ ，给定若 $k(1 \leq k \leq I)$ 成立，而 $S_{c^k}(x) \subset [x]_d$ 成立，而 $S_{c^{k+1}}(x) \subset [x]_d$ 不成立，则称 k 层粒度是关于对象 x 的局部最优粒度。若 $S_{c^I}(x) \subset [x]_d$ 成立，则称 I 层粒度关于对象 x 的局部最优粒度。

定义 2.7 [17] 在不协调的不完备多粒度决策系统 $S = (C, U \cup d)$ 中，对于任意的 $k(1 \leq k \leq I)$ ， $S^k = (U, C^k \cup d)$ 都是不协调的。

下面引入对象 x 的广义决策函数 $\partial_{c^k}(x)$ ：

$$\partial_{c^k}(x) = \{d(y) : y \in S_{c^k}(x)\} \quad x \in U$$

对于 $S^k = (U, C^k \cup d)$ ，若 $\forall x \in U$ ，都有 $\partial_{c^k}(x) = \partial_{c^I}(x)$ ，则称 $S^k = (U, C^k \cup d)$ 是广义协调的。对于给定的 $k(1 \leq k \leq I)$ ，若 $S^k = (U, C^k \cup d)$ 是广义协调的，而 $S^{k+1} = (U, C^{k+1} \cup d)$ 不是广义协调的，则称 k 层粒度是广义全局最优粒度；若 $S^I = (U, C^I \cup d)$ 是广义协调的，则称 I 层粒度是广义全局最优粒度。

定义 2.8 [17] 在不完备多粒度决策系统 $S = (U, C \cup \{d\})$ 中，对于 $x \in U$ ，给定 $k(1 \leq k \leq I)$ ，若 $\partial_{c^k}(x) = \partial_{c^I}(x)$ 成立，而 $\partial_{c^{k+1}}(x) = \partial_{c^I}(x)$ 不成立，则称 k 层粒度是关于对象 x 的广义局部最优粒度；若 $\partial_{c^I}(x) = \partial_{c^I}(x)$ 成立，则称 I 层粒度是关于对象 x 的广义局部最优粒度。

3. 多源决策信息系统的最优选择

令 $DIS_k = (U, A_k, F_k)$ 是一个决策信息系统，其中 U 是论域， $A_k = C_k \cup D_k$ ，

C_k 是条件属性集， $D_k = \{d_{k1}, d_{k2}, \dots, d_{ks}\}$ 是决策属性集， V_{kl} 是第 k 个系统中的条件属性的值域，即 $V_{kl} = \{v_{kl} \mid v_{kl} \text{ 是第 } k \text{ 个信息系统中第 } l \text{ 个条件属性 } a_{kl} \text{ 的值}\}$ 。

在下文中，我们讨论所有决策信息系统中条件属性集 a_k 与决策属性集 D_k 都相同，即 $A_k = C \cup D_k, 1, 2, \dots, s$ 。换言之，我们所讨论的多源信息系统是基于相同论域，相同属性，但每个信息系统重的属性值不同。

3.1. 多源决策信息系统的最优信息源选择

定义 3.1.1 设 $MDIS = \{DIS_k \mid DIS_k = (U, C_k \cup \{d\}, F_k), k = 1, 2, \dots, s\}$ 为多源决策信息系统，其中， $DIS_k = (U, C \cup D, F_k)$ 为单源信息系统，则多源决策信息系统是由 S 个单源信息系统构成，即

$$MDIS = DIS_1 \cup DIS_2 \cup \dots \cup DIS_s,$$

其中 $DIS_k = (U, C_k \cup D)$ ， $k \in \{1, 2, 3, \dots, S\}$ 。

定义 3.1.2 设 $MDIS = \{DIS_k \mid DIS_k = (U, C_k \cup \{d\}, F_k), k = 1, 2, \dots, s\}$ 为多源信息系统，对任意的 $x \in U$ ，在 DIS_k 上关于决策集 D 的广义决策函数为

$$\partial_{IS_k}(x) = \{\partial_{IS_1}(x), \partial_{IS_2}(x), \dots, \partial_{IS_s}(x)\},$$

其中 $\partial_{IS_k}(x) = \{d(y) \mid y \in [x]_I, S_k\}$ 。

定义 3.1.3 设 $MDIS = \{DIS_k \mid DIS_k = (U, C_k \cup \{d\}, F_k), k = 1, 2, \dots, s\}$ 为多源决策信息系统，若 $R_{DIS_1} \subseteq R_d$ ，则 $MDIS$ 协调，若进一步满足 $R_{DIS_k} \subseteq R_d$ ，则称 $DIS_k = (U, C_k \cup D)$ 协调，否则不协调。当 $MDIS$ 不协调时，若 $\partial_{c^k}(x) = \partial_{c^I}(x)$ ，则称 DIS_k 在 $MDIS$ 中是协调的。

定义 3.1.4 设 $MDIS = \{DIS_k \mid DIS_k = (U, C_k \cup \{d\}, F_k), k = 1, 2, \dots, s\}$ 为多源决策信息系统，若 $MDIS$ 协调当且仅当 DIS_k 协调且 $k+1 \leq I$ 时， DIS_{k+1} 不协调，则第 k 个信息系统最优；当 $MDIS$ 不协调，第 k 个信息源数据最优当且仅当 DIS_{k+1} 在 $MDIS$ 中是协调的，而且若有 $k+1 \leq I$ ， DIS_{k+1} 在 $MDIS$ 中必不协调。

例 3.1.1 设多源决策信息系统(如图 1 所示)，论域 $U = \{x_1, x_2, x_3, \dots, x_{10}\}$ ，分别从不同的四个信息源得到的属性集 $C = \{a_1, a_2, a_3, a_4\}$ ，决策属性 D 。

U	IS_1				IS_2				IS_3				D
	a_1	a_2	a_3	a_4	a_1	a_2	a_3	a_4	a_1	a_2	a_3	a_4	d
x_1	1	1	3	2	2	1	2	1	1	1	2	2	1
x_2	1	1	3	2	2	1	2	1	1	2	3	2	1
x_3	2	1	2	1	1	1	3	2	1	1	2	2	1
x_4	1	2	2	1	1	2	2	1	1	1	2	2	2
x_5	1	2	2	1	1	2	2	1	1	1	3	2	2
x_6	1	2	1	1	1	2	2	1	2	2	1	2	3
x_7	1	2	2	1	1	2	2	1	2	2	1	2	2
x_8	2	2	1	1	1	2	1	1	1	1	2	3	3
x_9	1	2	1	1	1	2	2	1	2	2	1	2	3
x_{10}	2	2	1	1	1	2	2	1	1	1	2	3	2

Figure 1. Multi-source information system

图 1. 多源信息系统

我们可以求得关于多源决策信息系统的单源信息系统划分如下：

$$U / R_{DIS_1} = \{\{x_1, x_2\}, \{x_3\}, \{x_4, x_5, x_7\}, \{x_6, x_9\}, \{x_8\}, \{x_{10}\}\}$$

$$U / R_{DIS_2} = \{\{x_1, x_2\}, \{x_3\}, \{x_4, x_5, x_7, x_{10}\}, \{x_6, x_8, x_9\}\}$$

$$U / R_{DIS_3} = \{\{x_1, x_3, x_4\}, \{x_2, x_5\}, \{x_6, x_7, x_9\}, \{x_8, x_{10}\}\}$$

$$U / R_{DIS_4} = \{\{x_1, x_3, x_4\}, \{x_4, x_5, x_7, x_{10}\}, \{x_6, x_8, x_9\}\}$$

根据定义 3.1.3, 在 $MDIS = \{DIS_k \mid DIS_k = (U, C_k \cup \{d\}, F_k), k = 1, 2, \dots, s\}$ 多源决策信息系统中, 若 $R_{DIS_1} \subseteq R_d$, 则 $MDIS$ 协调。

$$\text{又 } U / D = \{\{x_1, x_2, x_3\}, \{x_4, x_5, x_7, x_{10}\}, \{x_6, x_8, x_9\}\}$$

根据以上划, 我们可以看到 $U / R_{DIS_1} \subseteq R_d$, 则 $MDIS$ 为协调的多源决策信息系统。

$U / R_{DIS_1} \subseteq U / R_d$, $U / R_{DIS_2} \subseteq U / R_d$ 则 DIS_1, DIS_2 为协调的; $DIS_3 = (U, C_3 \cup d)$ 是不协调的;

所以第 2 个信息系统是最优的信息源。

定理 3.1.1 设 $MDIS = \{DIS_k \mid DIS_k = (U, C_k \cup \{d\}, F_k), k = 1, 2, \dots, s\}$ 为多源决策信息系统, 则 DIS_k 协调当且仅当对任意 $t = \{1, 2, \dots, w\}$, 恒有 $\left| \partial_{C_t^k}(x) \right| = 1$ 。

证明: 由定义 3.1.3, 3.1.4 易证得。

3.2. 协调的多源决策信息系统的最优信息源选择

在多源决策信息系统中, 不同来源的信息系统会有不同的决策规则, 为了使最终提取的规则具有更好的贴合性和准确性, 我们首先根据对象选择局部最优信息源; 然后在保持协调性不变的基础上进行局部相对约简, 最后按“条件→决策”的形式按规则。

从多源信息系统中提取出来。

定义 3.2.1 多源决策信息系统 $MDIS = \{DIS_k \mid DIS_k = (U, C_k \cup \{d\}, F_k), k = 1, 2, \dots, s\}$ 中, 对于 $x \in U$, 给定 $k (1 \leq k \leq I)$, 若 $IS_k(x) \subseteq [x]_d$ 成立, 其中, $DIS_k = (U, C_k \cup \{d\})$, 而 $DIS_{k+1}(x) \subseteq [x]_d$ 不成立, $IS_{k+1} = (U, C_{k+1} \cup \{d\})$, 则称第 k 个信息源是关于对象 x 的局部最优选择, 若 $DIS_1(x) \subseteq [x]_d$ 成立。则称 I

层信息是关于对象 x 的局部最优信息来源。

例 3.2.1 续例 3.1.1 中的多源决策信息系统 $\{(U, C_k \cup \{d\}, F_k), k=1, 2, \dots, s\}$,

关于每个对象 x 局部最优选择如下:

$IS_1(x_1) \subseteq IS_2(x_1) \subseteq [x_1]_d$, 关于 x_1 的局部最优信息来源是 IS_2 信息系统。

$IS_1(x_2) \subseteq IS_2(x_2) \subseteq [x_2]_d$, 关于 x_2 的局部最优信息来源是 IS_2 信息系统。

$IS_1(x_3) \subseteq IS_2(x_3) \subseteq IS_3(x_3) \subseteq [x_3]_d$, 关于 x_3 的局部最优信息来源是 IS_3 信息系统。

$IS_1(x_4) \subseteq IS_2(x_4) \subseteq [x_4]_d$, 关于 x_4 的局部最优信息来源是 IS_2 信息系统。

$IS_1(x_5) \subseteq IS_2(x_5) \subseteq [x_5]_d$, 关于 x_5 的局部最优信息来源是 IS_2 信息系统。

$IS_1(x_6) \subseteq IS_2(x_6) \subseteq IS_3(x_6) \subseteq [x_6]_d$, 关于 x_6 的局部最优信息来源是 IS_3 信息系统。

$IS_1(x_7) \subseteq IS_2(x_7) \subseteq [x_7]_d$, 关于 x_7 的局部最优信息来源是 IS_2 信息系统。

$IS_1(x_8) \subseteq IS_2(x_8) \subseteq IS_3(x_8) \subseteq [x_8]_d$, 关于 x_8 的局部最优信息来源是 IS_3 信息系统。

$IS_1(x_9) \subseteq IS_2(x_9) \subseteq IS_3(x_9) \subseteq [x_9]_d$, 关于 x_9 的局部最优信息来源是 IS_3 信息系统。

$IS_1(x_{10}) \subseteq IS_2(x_{10}) \subseteq IS_3(x_{10}) \subseteq [x_{10}]_d$, 关于 x_{10} 的局部最优信息来源是 IS_3 信息系统。

显然, 在全局最优的第 2 层信息源上, 只有对象 x_1, x_2, x_4, x_5, x_7 达到最优, $x_3, x_6, x_8, x_9, x_{10}$ 的局部最优信息来源为第三个信息系统, 因此不同对象会在不同的信息系统上达到最优。

定义 3.2.2 在多源决策信息系统中 $MDIS = \{DIS_k \mid DIS_k = (U, C_k \cup \{d\}, F_k), k=1, 2, \dots, s\}$ 中, 给定对象 $x \in U$ 的局部最优信息系统 IS_k , 存在 $B_k \subseteq C_k$, 使 $IS_B^k(x) \subseteq [x]_d$ 成立, 其中, $IS_k = (U, C_k \cup \{d\})$ 且对于任意 $b_k \in B_k$ $IS_{B-b}^k(x) \subseteq [x]_d$ 不成立, 则称 B_k 是 C_k 关于对象 x 的局部相对约简。

例 3.2.2 对于例 3.1.1 中的多源决策信息系统 $MDIS = DIS_1 \cup DIS_2 \cup DIS_3$, 由 3.2.1 可知, 关于对象 x_1, x_2, x_4, x_5, x_7 的局部最优信息来源为第二信息系统 DIS_2 , 在 $DIS_2 = (U, IS_2 \cup \{d\})$ 中, x_1, x_2 的相似类为 $\{x_1, x_2\}$, x_4, x_5, x_7 的相似类为 $\{x_4, x_5, x_6, x_7, x_9, x_{10}\}$ 。

则在第二个信息系统中, $\{a_1, a_2\}$ 是 $DIS_2 = (U, IS_2 \cup \{d\})$ 关于对象 x_1, x_2, x_4, x_5, x_7 的一个局部相对约简。

关于对象 $x_3, x_6, x_8, x_9, x_{10}$ 的局部最优信息来源为第三信息系统 DIS_3 , 在 $DIS_3 = (U, IS_3 \cup \{d\})$ 中, x_3 的相似类为 $\{x_1, x_3\}$, x_6, x_9 的相似类为 $\{x_6, x_7, x_9\}$, x_8 的相似类为 $\{x_8\}$, x_{10} 的相似类为 $\{x_{10}\}$ 。则在第三个信息系统中, $\{a_1, a_2, a_3\}$ 是关于对象 $x_3, x_6, x_8, x_9, x_{10}$ 的一个局部相对约简。

例 3.2.3 对于例 3.1.1 中的协调的多源决策信息系统 $MDIS = DIS_1 \cup DIS_2$, 由例 3.2.1 和 3.2.2 可以得到提取局部最优信息来源的规则如下:

$r_1: (a_1, 2) \wedge (a_2, 1) \Rightarrow (d, 1)$, 支持集为 $\{x_1, x_2\}$;

$r_2: (a_1, 1) \wedge (a_2, 2) \Rightarrow (d, 2)$, 支持集为 $\{x_4, x_5, x_7\}$;

$r_3: (a_1, 1) \wedge (a_2, 1) \wedge (a_3, 2) \Rightarrow (d, 2)$, 支持集为 $\{x_3\}$;

$r_4: (a_1, 2) \wedge (a_2, 2) \wedge (a_3, 1) \Rightarrow (d, 3)$, 支持集为 $\{x_6, x_9\}$;

$r_5: (a_1, 1) \wedge (a_2, 1) \wedge (a_3, 2) \Rightarrow (d, 3)$, 支持集为 $\{x_8\}$;

$r_6: (a_1, 2) \wedge (a_2, 1) \wedge (a_3, 1) \Rightarrow (d, 3)$, 支持集为 $\{x_{10}\}$;

3.3. 不协调的多源决策信息系统的局部最优信息来源选择

在不协调的多源决策信息系统 $MDIS = \{DIS_k \mid DIS_k = (U, C_k \cup \{d\}, F_k), k=1, 2, \dots, s\}$ 中, 对于任意的 $k (1 \leq k \leq I)$, $DIS_k = (U, C_k \cup \{d\})$ 都不协调。

定义 3.3.1 设 $MDIS = \{DIS_k \mid DIS_k = (U, C_{IS_k} \cup \{d\}, F_k), k=1, 2, \dots, s\}$ 为一个多源决策信息系统, 对象 x 的广义决策函数

$$\partial_{MDIS}(x) = \{\partial_{IS_1}(x), \partial_{IS_2}(x), \dots, \partial_{IS_s}(x)\}, x \in U$$

其中,

$$\partial_{IS_k}(x) = \{d(y) : y \in IS_k(x)\}, x \in U$$

对于 $DIS_k = (U, IS_k \cup d)$, 若 $\forall x \in U$ 都有 $\partial_{IS_k}(x) = \partial_{IS_1}(x)$, 则称 $DIS_k = (U, C_k \cup d)$ 是广义协调的, 对于给定的 $k (1 \leq k \leq I)$, 若 $DIS_k = (U, C_k \cup d)$ 是广义协调的, 而 $DIS_{k+1} = (U, C_{k+1} \cup d)$ 不是广义协调的, 则称第 k 个信息源是广义全局最优的信息系统。若 $DIS_k = (U, C_k \cup d)$ 是广义协调的, 则称第 I 个信息系统是广义全局最优信息系统。

例 3.3.1 设多源决策信息系统 $MDIS = DIS_1 \cup DIS_2 \cup DIS_3$ (如图 2 所示), 论域 $U = \{x_1, x_2, \dots, x_{10}\}$, 属性集 $C = \{a_1, a_2, a_3, a_4\}$, 决策属性 d 。

U	IS_1				IS_2				IS_3				D
	a_1	a_2	a_3	a_4	a_1	a_2	a_3	a_4	a_1	a_2	a_3	a_4	d
x_1	1	1	2	3	1	1	1	2	1	1	3	4	1
x_2	1	1	2	3	1	1	1	2	2	1	2	4	1
x_3	2	2	2	2	1	1	1	1	3	3	2	3	1
x_4	2	2	2	2	1	1	1	1	3	2	2	3	1
x_5	4	2	2	2	3	1	1	1	1	3	2	3	1
x_6	4	2	2	2	3	1	1	1	1	3	2	3	1
x_7	4	2	2	2	2	1	2	1	1	3	2	3	2
x_8	3	2	3	2	2	1	2	1	4	2	1	3	2
x_9	3	2	1	1	2	1	1	1	3	2	1	2	2
x_{10}	3	2	1	1	2		1	1	3	3	1	1	2

Figure 2. An uncoordinated multi-source information system

图 2. 不协调的多源信息系统

$$U/R_{IS_1} = \{\{x_1, x_2\}, \{x_3, x_4\}, \{x_5, x_6, x_7\}, \{x_8\}, \{x_9, x_{10}\}\}$$

$$U/R_{IS_2} = \{\{x_1, x_2\}, \{x_3, x_4\}, \{x_5, x_6\}, \{x_7, x_8\}, \{x_9, x_{10}\}\}$$

$$U/R_{IS_3} = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5, x_6, x_7\}, \{x_8\}, \{x_9\}, \{x_{10}\}\}$$

$$U/R_{IS_4} = \{\{x_1, x_2, x_3, x_4, x_5, x_6\}, \{x_7, x_8, x_9, x_{10}\}\}$$

$$U/R_{IS_1} \not\subset U/D, \text{ 显然多源信息系统是不协调的。}$$

进而, 由定义 3.3.1. 可求得该多源信息系统中所有对象的广义决策函数值如图 3 所示。

显然, $\forall x \in \{x_1, x_2, x_3, \dots, x_{10}\}$, $\partial_{IS_1}(x) = \partial_{IS_2}(x)$ 成立, 而 $\partial_{IS_2}(x) \neq \partial_{IS_3}(x)$ 。即 $DIS_2 = (U, C_2 \cup d)$ 是广义协调的; $DIS_3 = (U, C_3 \cup d)$ 不是广义协调的。

所以, 第 2 个信息源是广义全局最优的信息系统。

关于每个对象的广义局部最优信息源如下:

当 $x_i \in \{x_1, x_2, x_3, x_4\}$ 时, $\partial_{IS_1}(x) = \partial_{IS_3}(x)$ 成立, 即关于对象 x_1, x_2, x_3, x_4 的广义局部最优信息源为第 3 个信息系统。

当 $x_i \in \{x_5, x_6, x_7\}$ 时, $\partial_{IS_1}(x) = \partial_{IS_2}(x)$ 成立, 而 $\partial_{IS_2}(x) \neq \partial_{IS_3}(x)$, 即关于对象 x_5, x_6, x_7 的广义局部最优信息源为第 2 个信息系统。

当 $x_i \in \{x_8, x_9, x_{10}\}$ 时, $\partial_{IS_1}(x) = \partial_{IS_2}(x)$ 成立, 而 $\partial_{IS_2}(x) \neq \partial_{IS_3}(x)$, 即关于对象 x_8, x_9, x_{10} 的广义局部最优信息源为第 2 个信息系统。

显然, 不同的对象在不同的信息系统上会达到广义局部最优信息。

U	∂_{DIS_1}	∂_{DIS_2}	∂_{DIS_3}
x_1	{1}	{1}	{1}
x_2	{1}	{1}	{1}
x_3	{1}	{1}	{1}
x_4	{1}	{1}	{1}
x_5	{1}	{1}	{1, 2}
x_6	{1}	{1}	{1, 2}
x_7	{2}	{2}	{1, 2}
x_8	{2}	{2}	{2}
x_9	{2}	{2}	{2}
x_{10}	{2}	{2}	{2}

Figure 3. Generalized decision function values of inconsistent multi-source information systems

图 3. 不协调的多源信息系统广义决策函数值

定义 3.3.2 在不协调的多源决策信息系统 $MDIS = \{DIS_k \mid DIS_k = (U, C_k \cup \{d\}, F_k), k = 1, 2, \dots, s\}$ 中, 对于给定的对象 $x \in U$ 的广义局部最优信息源 k 存在 $B_k \subseteq C_k$ 使得 $\partial_{B_k}(x) = \partial_{C_k}(x)$ 成立, 且对于任意 $b_k \in B_k$, $\partial_{B_k - \{b_k\}}(x) = \partial_{C_k}(x)$ 不成立, 则称 B_k 是 C_k 关于对象 x 的广义局部相对约简。

例 3.3.2 对于例 3.3.1 中不协调的多源决策信息系统 $MDIS = DIS_1 \cup DIS_2 \cup DIS_3$, 关于对象 x_1, x_2, x_3, x_4 的广义局部最优信息源为第 3 个信息系统。 x_1, x_2, x_3, x_4 的广义决策函数值为 {1}, $\{a_1, a_2, a_3\}$ 是 $DIS_3 = (U, C_{IS_3} \cup \{d\})$ 关于对象 x_1, x_2, x_3, x_4 的一个广义局部相对约简;

对象 x_5, x_6, x_7 的广义局部最优信息源为第 2 个信息系统。对象 x_5, x_6, x_7 的广义决策函数值为 {1, 2}, $\{a_2, a_3, a_4\}$ 是 $DIS_2 = (U, C_{IS_2} \cup \{d\})$ 关于对象 x_5, x_6, x_7 的一个广义局部相对约简;

对象 x_8, x_9, x_{10} 的广义局部最优信息源为第 3 个信息系统。对象 x_8, x_9, x_{10} 的广义决策函数值为 {2}, $\{a_1, a_2, a_4\}$ 是 $DIS_3 = (U, C_{IS_3} \cup \{d\})$ 关于对象 x_8, x_9, x_{10} 的一个广义局部相对约简。

例 3.3.3 对于例 3.3.1 中不协调的多源决策信息系统 $MDIS = DIS_1 \cup DIS_2 \cup DIS_3$, 由例 3.3.1 和 3.3.2 可以得到关于对象 $x_1, x_2, x_3, \dots, x_{10}$ 的局部最优信息源的规则如下:

- $r_1: (a_1, 1) \wedge (a_2, 1) \wedge (a_3, 1) \Rightarrow (d, 1)$, 支持集为 $\{x_1\}$;
- $r_2: (a_1, 2) \wedge (a_2, 1) \wedge (a_3, 2) \Rightarrow (d, 1)$, 支持集为 $\{x_2\}$;
- $r_3: (a_1, 3) \wedge (a_2, 3) \wedge (a_3, 2) \Rightarrow (d, 1)$, 支持集为 $\{x_3\}$;
- $r_4: (a_1, 3) \wedge (a_2, 2) \wedge (a_3, 2) \Rightarrow (d, 1)$, 支持集为 $\{x_4\}$;
- $r_5: (a_2, 1) \wedge (a_3, 1) \wedge (a_4, 1) \Rightarrow (d, 1)$, 支持集为 $\{x_5, x_6\}$;
- $r_6: (a_2, 1) \wedge (a_3, 1) \wedge (a_4, 1) \Rightarrow (d, 2)$, 支持集为 $\{x_7\}$;
- $r_7: (a_1, 4) \wedge (a_2, 2) \wedge (a_4, 3) \Rightarrow (d, 2)$, 支持集为 $\{x_8\}$;
- $r_8: (a_1, 3) \wedge (a_2, 2) \wedge (a_4, 2) \Rightarrow (d, 2)$, 支持集为 $\{x_9\}$;

$r_9 : (a_1, 3) \wedge (a_2, 3) \wedge (a_4, 1) \Rightarrow (d, 2)$, 支持集为 $\{x_{10}\}$;

4. 小结

本文以大数据时代多源数据处理难题为背景, 立足多粒度思想与粗糙集理论, 结合多粒度标记信息系统概念, 针对现有多源数据研究中“信息融合方法复杂、最优选择缺乏统一标准”的问题, 拓展了多源决策信息系统的分析框架。创新性提出多源决策信息系统的广义决策函数定义。在此基础上, 研究围绕协调与不协调两类多源决策信息系统展开分析探讨: 对于协调系统, 借助广义决策函数确定全局最优信息源, 并针对单个对象筛选局部最优信息源, 进而在局部最优信息源中找到保持对象协调性的局部相对约简。对于不协调系统, 引入广义协调概念, 通过广义决策函数筛选广义全局最优信息源, 同时确定单个对象的广义局部最优信息源及相应的广义局部相对约简, 且均通过实例验证了所提方法的可行性。

该研究首次将多粒度思想与多源决策信息系统的“最优选择”深度结合, 解决了传统信息融合中“方法零散、针对性弱”的问题; 实例证明方法可有效筛选最优信息源、简化属性, 为多源数据的知识提取(如决策规则生成)提供了高效路径, 可应用于数据挖掘、决策分析等领域。后续可进一步拓展两方面内容: 一是深入探索最优信息系统的决策规则提取方法, 提升规则的实用性与泛化性; 二是优化属性约简算法, 针对大规模多源数据场景提升计算效率, 完善多粒度多源信息处理的理论体系。

参考文献

- [1] Pawlak, Z. (1982) Rough Sets. *International Journal of Computer & Information Sciences*, **11**, 341-356. <https://doi.org/10.1007/bf01001956>
- [2] 张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法[M]. 北京: 科学出版社, 2001.
- [3] 张文修, 仇国芳. 基于粗糙集的不确定决策[M]. 北京: 清华大学出版社, 2005.
- [4] Lin, G.P., Liang, J.Y. and Qian, Y.H. (2008) An Information Fusion Approach by Combining Multi-Granulation Rough Sets and Evidence Theory. *Information Sciences*, **49**, 466-477.
- [5] Zhang, X.Y., Huang, X.D. and Xu, W.H. (2023) Matrix-Based Multi-Granulation Fusion Approach for Dynamic Updating of Knowledge in Multi-Source Information. *Knowledge-Based Systems*, **262**, Article 110257.
- [6] Zhang, X.Y., Chen, X.W., Xu, W.H., et al. (2022) Dynamic Information Fusion in Multi-Source Incomplete Interval-Valued Information System with Variation of Information Sources and Attributes. *Information Sciences*, **608**, 1-27.
- [7] 陈辉皇. 多源信息系统中的决策规则挖掘研究[D]: [硕士学位论文]. 漳州: 闽南师范大学, 2018.
- [8] Xu, W., Li, M. and Wang, X. (2017) Information Fusion Based on Information Entropy in Fuzzy Multi-Source Incomplete Information System. *International Journal of Fuzzy Systems*, **19**, 1200-1216. <https://doi.org/10.1007/s40815-016-0230-9>
- [9] Yang, R., Li, H. and Huang, H. (2024) Multisource Information Fusion Considering the Weight of Focal Element's Beliefs: A Gaussian Kernel Similarity Approach. *Measurement Science and Technology*, **35**, Article 025136. <https://doi.org/10.1088/1361-6501/ad0e3b>
- [10] Liu, N., Hu, J. and Liang, W. (2023) MIFINN: A Novel Multi-Information Fusion and Interaction Neural Network for Aspect-Based Sentiment Analysis. *Knowledge-Based Systems*, **280**, Article 110983. <https://doi.org/10.1016/j.knosys.2023.110983>
- [11] Zhang, M., Cui, H., Tian, X., Kang, B. and Huang, L. (2023) Evaluate the Reliability of Information Sources Using the Non-Parametric Plausibility Relief Algorithm for Multi-Source Information Fusion. *Applied Soft Computing*, **148**, Article 110871. <https://doi.org/10.1016/j.asoc.2023.110871>
- [12] 万青, 马盈仓, 魏玲. 基于多粒度的多源数据知识获取[J]. 山东大学学报(理学版), 2020, 55(1): 41-50.
- [13] Wu, W.Z., Chen, Y., Xu, Y.H., et al. (2016) Optimal Granularity Selections in Consistent Incomplete Multi-Granular Labeled Decision Systems. *Pattern Recognition and Artificial Intelligence*, **29**, 108-115.
- [14] Wu, W. and Leung, Y. (2011) Theory and Applications of Granular Labelled Partitions in Multi-Scale Decision Tables. *Information Sciences*, **181**, 3878-3897. <https://doi.org/10.1016/j.ins.2011.04.047>
- [15] Wu, W. and Leung, Y. (2013) Optimal Scale Selection for Multi-Scale Decision Tables. *International Journal of Approximate Reasoning*, **54**, 1107-1129. <https://doi.org/10.1016/j.ijar.2013.03.017>

-
- [16] 刘凤玲, 林国平, 余晓龙. 多决策的多粒度标记系统的最优粒度选择[J]. 重庆理工大学学报(自然科学), 2020, 34(5): 263-270.
- [17] 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.
- [18] Li, F. and Hu, B.Q. (2017) A New Approach of Optimal Scale Selection to Multi-Scale Decision Tables. *Information Sciences*, **381**, 193-208. <https://doi.org/10.1016/j.ins.2016.11.016>
- [19] 吴伟志, 杨丽, 谭安辉, 等. 广义不完备多粒度标记决策系统的粒度选择[J]. 计算机研究与发展, 2018, 5(6): 1263-1272.
- [20] 史进玲, 张倩倩, 徐久成. 多粒度决策系统属性约简的最优粒度选择[J]. 计算机科学, 2018, 45(2): 152-156.
- [21] 吴伟志. 多粒度标记数据的知识表示和知识获取研究[Z]. 浙江海洋学院, 2013-12-30.
- [22] Zhao, H., Mi, J. and Liang, M. (2022) A Multi-Granularity Information Fusion Method Based on Logistic Regression Model and Dempster-Shafer Evidence Theory and Its Application. *International Journal of Machine Learning and Cybernetics*, **13**, 3131-3142. <https://doi.org/10.1007/s13042-022-01584-w>
- [23] Yu, Y., Tang, H., Qian, J., Zhu, Z., Cai, Z. and Lv, J. (2022) Fine-grained Image Recognition via Trusted Multi-Granularity Information Fusion. *International Journal of Machine Learning and Cybernetics*, **14**, 1105-1117. <https://doi.org/10.1007/s13042-022-01685-6>
- [24] Deng, J., Liu, H., Fang, H., Shao, S., Wang, D., Hou, Y., et al. (2023) MgNet: A Fault Diagnosis Approach for Multi-Bearing System Based on Auxiliary Bearing and Multi-Granularity Information Fusion. *Mechanical Systems and Signal Processing*, **193**, Article 110253. <https://doi.org/10.1016/j.ymssp.2023.110253>