

# 基于等价关系的最小乐观概念格生成算法



温馨<sup>1</sup> 闫心怡<sup>2</sup> 陈泽华<sup>1</sup>

1 太原理工大学大数据学院 太原 030024

2 太原理工大学电气与动力工程学院 太原 030024

(wx20126106@163.com)

**摘要** 决策信息系统的规则提取是数据挖掘的研究内容之一,概念格理论与粒计算理论是该领域研究的主要数学工具。文中通过探究这两大理论间的关系,利用等价关系定义了最小乐观概念格及其结构,最小乐观概念区别于传统经典概念,但是具有格的结构。在此基础上,提出了一种决策信息系统的规则提取算法,该算法引入了粒度思想,通过求取每一粒层中的最小乐观概念,并根据最小乐观概念的外延与决策属性等价类间的蕴含关系进行决策规则提取,通过设置算法的终止条件来加快其收敛速度,以达到针对决策信息系统知识约简的目的。最小乐观概念的定义比经典概念的定义更宽泛,其生成过程也更简单。最后,通过理论证明、实实验证以及数值实验对比验证了该方法的正确性与优越性。

**关键词:** 决策信息系统;概念格理论;粒计算;最小乐观概念;规则提取

**中图法分类号** TP181

## Minimal Optimistic Concept Generation Algorithm Based on Equivalent Relations

WEN Xin<sup>1</sup>, YAN Xin-yi<sup>2</sup> and CHEN Ze-hua<sup>1</sup>

1 College of Big Data Science, Taiyuan University of Technology, Taiyuan 030024, China

2 College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan 030024, China

**Abstract** Rule extraction of decision information system is an important topic in the field of data mining. Concept lattice theory and rough set theory are both theoretical tool for data analysis. This paper explores the relationship between these two theories, and uses the equivalent relationship to define the minimal optimistic concept lattice and its structure. The minimal optimistic concept is different from the traditional classic concept, but has a lattice structure, and a rule extraction algorithm for decision table is proposed. Based on granular computing, the algorithm computes the concepts in each layer from coarse to fine granularity space, and extracts decision rules according to the relationship between minimal optimistic concepts and decision equivalence classes. In order to achieve the purpose of knowledge reduction for decision information systems, the algorithm accelerates its convergent speed by setting the termination conditions. The definition of minimal optimistic concept is broader than classical concept, and the generation algorithm is simpler. The correctness and effectiveness of the new algorithm are verified by theorem proving and case analysis. Finally, the experimental results based on different data sets demonstrate that the proposed algorithm is more effective for rule extraction in most cases than other algorithms.

**Keywords** Decision information system, Concept lattice theory, Granular computing, Minimal optimistic concept, Rule extraction

## 1 引言

信息系统是一种描述感兴趣对象相关定性与定量信息的标准形式,常被作为数据分析工具与理论的标准输入,结果的输出被称作知识,知识指一些基于信息的数学结构。知识的表示与获取是人工智能领域的关键技术,其中规则挖掘是一项重要的研究内容。粗糙集理论(Rough Set Theory, RST)<sup>[1]</sup>是一种处理不确定信息以及进行数据处理的数学工具,很多

学者利用该理论进行数据分析<sup>[2-4]</sup>。Meng等<sup>[2]</sup>针对异构不完备信息系统的数据处理,提出了参数化公差粒化模型,并基于此提出了基于分辨函数与启发式函数的优化规则集提取算法;Junzo等<sup>[3]</sup>针对利用粗糙集方法获取的规则过多的问题,提出了一种通过合并决策规则与减少条件属性的个数来降低决策规则数量的方法,并利用该方法对时间序列数据进行分析,从而获取预测知识。文献[4]通过定义粒关系矩阵,寻找矩阵中的“0”元素,来实现在不同粒度下获取决策信息系统规

到稿日期:2020-01-07 返修日期:2020-05-20 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61703299);国家重点研发计划资助(2018YFB1404500)

This work was supported by the National Natural Science Foundation of China (61703299) and National Key R&D Program of China (2018YFB1404500).

通信作者:陈泽华(zehuachen@163.com)

则,该算法的正确识别率较高且规则泛化能力强,但算法的时间复杂度和空间复杂度依然很高。

形式概念分析(Formal Concept Analysis, FCA)<sup>[5]</sup>是由 Wille 于 20 世纪 80 年代提出的同样能够挖掘数据潜在信息的工具。针对信息系统的知识获取问题,一些学者多年来致力于此<sup>[6-8]</sup>。Zhi 等<sup>[6]</sup>提出了一种新的知识发现的粒度描述方法并生成三支概念,通过对粒度更加简洁的描述挖掘出近似决策规则,与当前已有方法相比,该算法只需要密切相关的粒度描述,大大降低了挖掘近似决策规则的时间复杂度。由于三支概念格相比传统概念格具有更加丰富的信息,因此 Wei 等<sup>[7]</sup>利用这种特点,从三支概念格出发提出了基于决策形式背景的规则提取方法,并传统概念格进行规则提取的算法进行了对比。Li 等<sup>[8]</sup>提出了一种从决策形式背景生成概念,进而推导无冗余规则的算法,该算法在一定程度上降低了算法的复杂度,在某些情况下,获取的决策规则仍然存在冗余属性。同时,针对不完备决策信息系统的研究也较多<sup>[9-10]</sup>。而本文基于完备决策信息系统利用等价关系定义了最小乐观概念格及其结构。最小乐观概念有别于传统形式概念,其生成过程更加简单,同时也可推广到不完备信息系统的规则提取中。

FCA 与 RST 理论虽在数据处理的方法与方式上存在差异,但由于它们的研究目标相同,因此在某种意义上具有一定的共性。José 等<sup>[11]</sup>利用 FCA, RST 以及它们的泛化理论,针对属性约简问题,结合 RST 中的约简思想,提出了一种模糊 FCA 中属性约简的新方法,并对其主要性质进行了研究。这种方法使得两种理论间的关系更加密切。Xiao 等<sup>[12]</sup>在形式背景中定义了基于格论算子的逼近概念,并讨论了它们的性质,同时为了弥补已有的两种生成近似概念方法的不足,提出了基于集论算子的近似方法,该研究有助于进一步理解粗糙集理论和形式概念分析在数据分析中的应用。Li 等<sup>[13]</sup>比较了通过 FCA 和 RST 两种理论进行规则提取的过程,用两种规则提取的算法比较了两种方式之间的异同,并对两种理论的发展进行了展望,但是其没有探讨两种理论之间的对应关系。Zhu 等<sup>[14]</sup>同时将 FCA 和 RST 理论用于决策表的多知识获取,提出了一种新的知识获取算法,该算法降低了传统概念格规则提取算法的复杂度。

本文通过分析 FCA 与 RST 两者之间的联系,构造出了最小乐观概念格,并据此提出了基于决策信息系统的规则提取算法。该算法的主要任务是依据选取的最小乐观概念,探究其外延与决策属性等价类间的蕴含关系,最终实现决策规则的提取。最小乐观概念提出的意义在于其简化了传统求取形式概念的计算复杂性,并且省去了决策规则中去除冗余属性的过程。

## 2 基本知识

本节将介绍本文主要用到的 FCA 和 RST 的基本概念。

### 2.1 概念格

定义 1<sup>[5]</sup> 形式背景可以用一个三元组  $T = \langle U, A, V, f \rangle$  来表示,其中  $U$  表示非空有限对象集,称为论域; $A$  表示非空有限属性集; $I$  满足  $I \subseteq U \times A$ ,表示形式背景的一种二元关

系,  $(u, a) \in I$  (其中,  $u \in U, a \in A$ ) 表示对象  $u$  拥有属性  $a$ ,  $(u, a) \notin I$  表示  $u$  没有属性  $a$ 。

在形式背景  $T = \langle U, A, V, f \rangle$  中,令  $2^U$  和  $2^A$  分别为对象集  $U$  和属性集  $A$  的幂集。对于任意对象集合  $P (P \subseteq U)$  和任意属性集合  $Q (Q \subseteq A)$ , Wille 定义了两个映射  $P^\uparrow : 2^U \rightarrow 2^A$  和  $Q^\downarrow : 2^A \rightarrow 2^U$  [5]:

$$P^\uparrow = \{a \in A \mid \forall u \in P, uIa\} \quad (1)$$

$$Q^\downarrow = \{u \in U \mid \forall a \in Q, uIa\} \quad (2)$$

易知,  $(\uparrow, \downarrow)$  为  $2U$  和  $2A$  之间的伽罗瓦(Galois)连接<sup>[8]</sup>。

定义 2<sup>[5]</sup> 三元组  $T = \langle U, A, I \rangle$  为一个形式背景,令  $P \subseteq U, Q \subseteq A$ ,若  $P^\uparrow = Q$  且  $Q^\downarrow = P$ ,则称二元组  $(P, Q)$  为一个形式概念,其中  $P$  被称为这个概念的外延,  $Q$  被称为这个概念的内涵。

为了方便叙述,将形式背景  $T$  下的所有概念存入  $\Psi(T)$  中。记  $\Psi(T)$  中的两个概念为  $(P_1, Q_1), (P_2, Q_2)$ ,若满足条件  $P_1 \subseteq P_2$  (或  $Q_1 \subseteq Q_2$ ),则称概念  $(P_1, Q_1)$  是  $(P_2, Q_2)$  的子概念,概念  $(P_2, Q_2)$  是  $(P_1, Q_1)$  的父概念。

### 2.2 等价类

定义 3<sup>[1]</sup> 决策信息系统可以用一个四元组  $T = \langle U, A, V, f \rangle$  来表示,其中  $U$  为对象集,  $A$  表示属性集,其中  $A = C \cup D$ ,且  $C \cap D = \emptyset$ ,同时  $C$  为非空有限条件属性集,  $D$  为非空有限决策属性集。  $V$  为全部属性  $A$  的值域;  $f: U \times A \rightarrow V$  是信息函数,它指定  $U$  中每个对象的属性值。对于任意的属性集合  $B \subseteq A$ ,一种不可分辨关系被定义为  $R_B = \{(x, y) \in U \times U \mid f(x, a) = f(y, a), \forall a \in B\}$ ,而  $U/R_B = \{[x]_B \mid x \in U\}$  (可表示为  $U/B$ ) 则可以表示为  $U$  关于  $R_B$  的划分。其中,  $[x]_B = \{y \in U \mid (x, y) \in R_B\}$ ,表示论域对象  $x$  在属性  $B$  下的等价类。

定义 4<sup>[4]</sup> 在决策信息系统中,令  $1 \leq \omega \leq n$ ,  $\omega$  表征当前系统的粒度,  $n$  为条件属性个数。这样的系统对应有  $n$  种粒度,并且粒度  $\omega$  越小,系统的知识粒度就越粗。

## 3 乐观概念与最小乐观概念

大量文献<sup>[15-16]</sup>表明,根据定义 2 的形式概念进行规则提取是存在冗余的,其冗余体现在概念的内涵存在冗余属性,因此本文在概念定义的基础上提出了最小概念。

定义 5 形式背景  $T = \langle U, A, I \rangle$  中,对于集合  $P \subseteq U, Q \subseteq A$ ,满足  $P^\uparrow = Q$  且  $Q^\downarrow = P$ 。若存在  $Q_1 \subseteq Q$ ,满足  $Q_1^\downarrow \subseteq Q^\downarrow$ ,且不存在  $Q_0 \subseteq Q_1$  满足  $Q_0^\downarrow \subseteq Q_1^\downarrow$ ,则定义  $(P, Q_1)$  为最小概念。

性质 1 最小概念不是概念。

证明:由定义 5 可知  $P^\uparrow = Q_1$ ,不满足定义 2,因此最小概念不是概念。

可以通过提取形式背景中的最小概念,来提取信息系统的非冗余规则,从而在规则提取时简化冗余规则的判断过程,降低算法的复杂度。

定义 6 在决策信息系统  $T = \langle U, A, V, f \rangle$  中,设集合  $B \subseteq A, P \subseteq U$ ,若将  $T$  转化为形式背景,属性集  $B$  需扩展的所有属性值为  $\{b_1, b_2, \dots\}$ 。若存在  $B' \subseteq \{b_1, b_2, \dots\}$  使得  $B'^\downarrow = P$ ,则定义  $(P, B)$  为一个乐观概念,称  $P$  为乐观概念的外延,  $B$  为乐观概念的内涵。

定理 1 在决策信息系统  $T = \langle U, A, V, f \rangle$  中,若  $U$  关于

属性子集  $B(B \subseteq A)$  产生的某一个等价类为  $U_i$ , 则集合  $(U_i, B)$  为一个乐观概念。

证明:1) 设集合  $B$  中只有一个属性  $a$ , 即  $B = \{a\}$ , 且属性  $a$  下有多个属性值  $\{a_1, a_2, \dots, a_i, \dots\}$ 。设由属性值  $a_i$  所产生的等价类为  $U_i = \{x \in U | f(x, a) = a_i\}$ 。在决策信息系统转化成的单值背景  $(U, A', I)$  中, 单个属性下满足  $f(x, a) = a_i \Rightarrow (x, a_i) \in I$ , 而根据定义 1 得,  $a_i^\downarrow = \{x \in U | (x, a) \in I\} = \{x \in U | f(x, a) = a_i\} = U_i$ , 即存在  $a_i \in \{a_1, a_2, \dots, a_i, \dots\}$  使得  $a_i^\downarrow = U_i$ , 由定义 5 可知  $(U_i, B)$  为一个乐观概念。

2) 设集合  $B$  中有两个属性  $a$  和  $b$ , 即  $B = \{a, b\}$ , 其中  $a = \{a_1, a_2, \dots\}$ ,  $b = \{b_1, b_2, \dots\}$ 。若有集合  $B' = \{a_i, b_k\}$ , 则产生等价类  $U_i = \{x \in U | f(x, a) = a_i \wedge f(x, b) = b_k\}$ 。在转化的单值背景中,  $B'^\downarrow = \{x \in U | \forall a \in B', (x, a) \in I\} = \{x \in U | (x, a_i) \in I \wedge (x, b_k) \in I\} = U_i$ , 因此  $(U_i, B)$  为一个乐观概念。

3) 设集合  $B$  中有多个属性, 即  $B = \{a, b, c, \dots\}$ , 可同样设集合  $B' = \{a_i, b_k, c_l, \dots\}$ , 与 2) 同理, 可求得  $(U_i, B)$  同样为一个乐观概念。

综上所述, 无论集合  $B$  中有多少属性, 皆可得到  $(U_i, B)$  为一个乐观概念, 定理 1 得证。

决策信息系统中所有属性子集产生的关于论域的所有等价类之间可能存在重复, 由此便得到外延相同、内涵不同的多个乐观概念。在这种情况下, 本文定义了最小乐观概念以提取不含冗余属性的最简规则。

**定义 7** 在决策信息系统  $T = \{U, A, V, f\}$  中, 可设属性集  $B(B \subseteq A)$  关于  $U$  的任一个等价类为  $U_B$ , 若不存在  $B' \subseteq B$  使得  $B'$  关于  $U$  的等价类为  $U_{B'}$  满足  $U_{B'} = U_B$ , 则称乐观概念  $(U_B, B)$  为最小乐观概念。

在相同的粒度下, 任意给定两个属性子集为  $B_1$  和  $B_2$ , 满足  $|B_1| = |B_2|$ , 当  $B_1$  和  $B_2$  产生相同的乐观概念时, 为了后续提取规则, 我们设第一个求得的乐观概念为最小乐观概念, 即  $(U_i, B)$  为最小乐观概念。

**定理 2** 在决策信息系统  $T = \{U, A, V, f\}$  中, 由所有等价类产生的最小乐观概念组成的集合  $BT$  是一个完备格<sup>[8]</sup>。

证明: 从以下两部分进行证明。

1)  $BT$  是一个格

对于任意两个乐观概念  $(U_1, B_1), (U_2, B_2)$ , 需证明  $(U_1 \cup U_2, B_1 \cap B_2), (U_1 \cap U_2, B_1 \cup B_2)$  均为最小乐观概念。

由定理 1 可知,  $U_1$  和  $U_2$  分别为论域  $U$  关于属性  $B_1$  和  $B_2$  的一种等价类, 则根据粗糙集的基础理论<sup>[2]</sup>,  $U_1 \cup U_2$  和  $U_1 \cap U_2$  分别为  $U$  关于属性  $B_1 \cap B_2$  和  $B_1 \cup B_2$  的一种等价类, 由定理 1 以及  $(U_1, B_1)$  和  $(U_2, B_2)$  均为最小乐观概念可知, 上述两个乐观概念为最小乐观概念。  $BT$  是一个格得证。

2)  $BT$  是完备的

对于任意两个乐观概念  $(U_1, B_1)$  和  $(U_2, B_2)$ , 根据定义可知必满足  $(U_1, B_1) \cup (U_2, B_2) = (U_1 \cup U_2, B_1 \cap B_2)$  和  $(U_1, B_1) \cap (U_2, B_2) = (U_1 \cap U_2, B_1 \cup B_2)$ 。因此, 对于任意  $S$  个乐观概念, 必满足  $\bigcup_{i=1}^S (U_i, B_i) = (\bigcup_{i=1}^S U_i, \bigcap_{i=1}^S B_i)$  和  $\bigcap_{i=1}^S (U_i, B_i) = (\bigcap_{i=1}^S U_i, \bigcup_{i=1}^S B_i)$ , 故完备性成立。  $BT$  是完备的得证。

综上所述, 可知  $BT$  是一个完备格。

## 4 规则提取算法

### 4.1 算法描述

基于前文定义的最小乐观概念, 本节针对决策信息系统, 提出了基于最小乐观概念格的决策信息系统最简规则提取算法。基于最小乐观概念的决策信息系统规则提取算法如算法 1 所示。

**算法 1** 基于最小乐观概念的决策信息系统规则提取算法

输入: 决策信息系统  $T = \{U, A, V, f\}$

输出: 所有最简规则 rules

1. 计算决策属性  $D$  下所有等价类, 初始化粒度  $\omega = 1$ 、已覆盖论域元素  $U_n = \emptyset$ 。
2. 在当前粒度  $\omega$  下计算属性子集  $B$  关于论域的等价类并生成乐观概念, 选取所有最小乐观概念。
3. 对于任意的最小乐观概念  $(P, B)$ , 判断其外延是否为某个决策等价类的子集。
4. 若  $P$  为决策等价类  $D_i$  的子集, 即  $P \subseteq D_i$ , 则得到规则  $B \rightarrow D_i$ , 并且将规则存入 rules, 并更新  $U_n: U_n := U_n \cup P$ 。
5. 若  $P$  不包含于任何决策等价类, 则转至步骤 3, 对下一乐观概念进行判断, 直至所有概念判断结束转至步骤 6。
6. 判断  $U_n$  是否等于  $U$ , 若相等, 则转步骤 8; 否则, 转步骤 7。
7. 若  $\omega < |C|$ , 则  $\omega = \omega + 1$ , 返回步骤 2 继续计算; 否则转步骤 8。
8. 输出 rules 中的逻辑规则。

### 4.2 实例解析

例 1 表 1 列出了一个完备决策信息系统, 其中论域为  $U = \{x_1, x_2, \dots, x_8\}$ , 条件属性集为  $C = \{a, b, c\}$ , 决策属性集为  $D = \{d\}$ 。

表 1 决策信息系统

Table 1 Decision information system

$U$	$a$	$b$	$C$	$d$
1	0	0	0	1
2	0	0	1	0
3	0	1	2	1
4	0	1	1	1
5	1	0	2	0
6	1	0	1	1

首先需求得决策属性下所有等价类:  $U/d = \{\{1, 3, 4, 6\}, \{2, 5\}\}$ 。

在粒度  $\omega = 1$  的情况下, 求所有条件属性子集关于  $U$  的划分, 可得到 7 个最小乐观概念, 如表 2 所列。判断每个概念的外延是否为决策等价类的子集, 可知只有概念  $(3, 4, b_1)$  和  $(1, c_0)$  满足条件, 据此可得到两条决策规则: 1) 根据决策信息系统中论域  $\{3, 4\}$  在属性  $b$  下的取值, 得到  $rule_1 = \{b = 1 \rightarrow d = 1\}$ ; 2) 由论域  $\{1\}$  在属性  $c$  下的取值得到  $rule_2 = \{c = 0 \rightarrow d = 1\}$ 。此时,  $U_n = \{3, 4\} \cup \{1\} = \{1, 3, 4\} \neq U$ , 需要继续计算。

当  $\omega = 2$  时, 求得的所有最小乐观概念如表 3 所列。其中加粗部分表示同一粒度下外延相同、内涵不同的最小乐观概念, 这些概念会重复提取决策信息系统的规则, 在这种情况下, 只保留第一个生成的最小乐观概念, 以保证格结构的最简化(见定义 7)。因此, 在该粒度下, 可得到 7 个最小乐观概念, 判断每个概念的外延是否为决策等价类的子集, 由结果可知只有概念  $(5, a_1 c_2)$  和  $(6, a_1 c_1)$  满足条件, 据此可得到两条决策规则: 1) 根据决策信息系统中论域  $\{5\}$  在属性  $a$  和  $c$  下的

取值,得到  $rule_3 = \{a=1 \wedge c=2 \rightarrow d=0\}$ ; 2) 由论域  $\{6\}$  在属性  $a$  和  $c$  下的取值得到  $rule_4 = \{a=1 \wedge c=1 \rightarrow d=1\}$ 。此时,  $Un = Un \cup \{5, 6\} = \{1, 3, 5, 6\} \neq U$ , 需要继续计算。

当  $\omega=3$  时, 求得所有最小乐观概念如表 4 所列。在该粒度下, 可得到 1 个最小乐观概念  $(2, abc)$ , 可得规则  $rule_5 = \{a=0 \wedge b=0 \wedge c=1 \rightarrow d=0\}$ , 此时  $Un = \{1, 2, 3, 4, 5, 6\} = U$ , 计算结束。最终, 决策信息系统  $T$  规则提取的结果为  $rule_1 \sim rule_5$ 。图 1 给出了决策信息系统  $T$  的最小乐观概念格结构。

表 2 当  $\omega=1$  时最小乐观概念的计算过程

Table 2 Computation process of minimal optimistic concepts when  $\omega=1$

粒度	等价类	较粗粒度下 已存在等价类	最小乐观 概念
	$U/a = \{\{1, 2, 3, 4\}, \{5, 6\}\}$	$\emptyset$	$(1234, a_0)(56, a_1)$
1	$U/b = \{\{1, 2, 5, 6\}, \{3, 4\}\}$	$\emptyset$	$(1256, b_0)(34, b_1)$
	$U/c = \{\{1\}, \{2, 4, 6\}, \{3, 5\}\}$	$\emptyset$	$(1, c_0)(246, c_1), (35, c_2)$

表 3 当  $\omega=2$  时最小乐观概念的计算过程

Table 3 Computation process of minimal optimistic concepts when  $\omega=2$

粒度	等价类	较粗粒度下 已存在等价类	最小乐观概念
	$U/ab = \{\{1, 2\}, \{3, 4\}, \{5, 6\}\}$	$\{3, 4\}, \{5, 6\}$	$(12, a_0 b_0)$
	$U/ac = \{\{1\}, \{2, 4\}, \{3\}, \{5\}, \{6\}\}$	$\{1\}$	$(24, a_0 c_1), (3, a_0 c_2), (5, a_1 c_2), (6, a_1 c_1)$
2	$U/bc = \{\{1\}, \{2, 6\}, \{3\}, \{4\}, \{5\}\}$	$\{1\}$	$(26, b_0 c_1), (3, b_1 c_2), (4, b_1 c_1), (5, b_0 c_2)$

表 4 当  $\omega=3$  时最小乐观概念的计算过程

Table 4 Computation process of minimal optimistic concepts when  $\omega=3$

粒度	等价类	较粗粒度下 已存在等价类	最小乐观概念
3	$U/abc = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$	$\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$	$(2, a_0 b_0 c_1)$

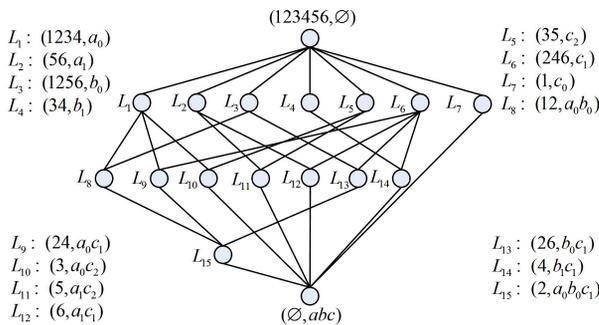


图 1 最小乐观概念格结构

Fig. 1 Minimal optimistic concept lattice structure

## 5 实验与分析

### 5.1 实验测试

为了进一步验证本文算法的有效性, 本文从 UCI 数据集中选取 8 组常用数据集, 如表 5 所列, 用于测试比较本文算法、文献[4]中的算法和文献[15]中的算法的性能, 并分别对比了每组实验规则个数、平均规则长度、正确识别率、运行时间等测试指标。图 2—图 5 显示了实验结果和对比效果。

表 5 测试数据集信息

Table 5 Information of data sets

数据集	论域个数	条件属性个数	决策类个数
Wine	178	13	3
Glass	214	9	7
Ecoli	336	7	8
Liver	345	6	2
Voting	435	16	3
Balance	625	4	5
German	1000	20	2
car	1728	6	4

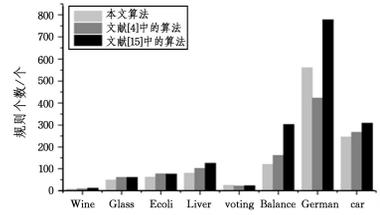


图 2 规则个数对比

Fig. 2 Comparison of number of rules

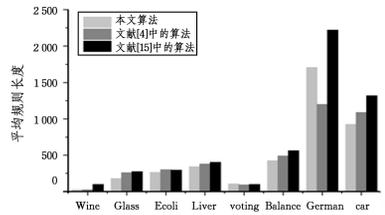


图 3 平均规则长度对比

Fig. 3 Comparison of length of rules

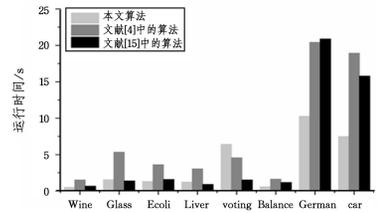


图 4 运行时间对比

Fig. 4 Comparison of time

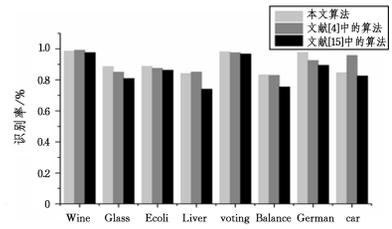


图 5 识别率对比

Fig. 5 Comparison of recognition rate

首先, 从规则个数来看, 本文算法较文献[4]中的算法与文献[15]中的算法具有更少的规则数, 其相应的规则长度也更少; 其次, 如图 5 所示, 从识别率角度观察, 本文算法在多数数据集上的表现良好, 在部分数据集上, 文献[4]中的算法略胜一筹; 最后, 在运行时间方面, 本文算法较文献[4]中的算法、文献[15]中的算法时间更短, 由于文献[4]中的算法通过计算粒矩阵与粒关系矩阵的方式来获取知识, 因此其运行时间更长, 而本文算法求取最小乐观概念的过程, 在本质上是对各属性集等价类进行计算的过程, 因此运行时间相对较短。

综上所述,本文算法具有较少的规则和较高的识别率,说明该算法更具有代表性。

## 5.2 算法复杂度分析

本文算法可分为两大步骤:1)计算决策信息系统中的所有决策等价类以及各个粒度层下的最小乐观概念;2)对所生成的最小乐观概念中的外延与决策等价类进行关系判断,进而确定最简决策规则。算法1中的步骤1-步骤3为第一大步,其核心在于求决策等价类及最小乐观概念,假设属性个数为 $|C|$ 、论域元素个数为 $|U|$ ,在粒度为 $\omega$ 时,求取该层的所有最小乐观概念的复杂度为 $C_{\omega}^{|C|}$ 。因此,在最坏的情况下,计算所有的最小乐观概念的复杂度为 $O\{C_{\omega}^{|C|} + C_{\omega-1}^{|C|} + \dots + C_1^{|C|}\} = O(2^{|C|})$ ,计算决策等价类的算法复杂度为 $O(1)$ ,故第一大步中的算法复杂度为 $O(2^{|C|}) = O(1)$ 。算法1中的步骤4-步骤8为规则提取部分,其计算核心为判断每个最小乐观概念是否为决策等价类的子集,故时间复杂度为 $O(1)$ 。因此,算法1的总时间复杂度为 $O(2^{|C|})$ 。但这只是特殊情况,由于该算法只需满足终止条件即可结束计算,即当算法1中集合 $U_n$ 的元素提前覆盖论域,下一粒层的最小乐观概念可以不用再计算与判断。因此,一般情况下,算法1的复杂度小于 $O(2^{|C|})$ 。该算法与文献[4]中的算法均引入了粒度思想,均在由粗到细的粒度空间下进行规则提取,但该算法的时间复杂度小于或等于文献[4]中的算法,并且空间复杂度也要小于文献[4]中的算法。

**结束语** 本文考虑了FCA与RST之间的联系,在不同的粒度空间下,通过求取本文定义的乐观概念和最小乐观概念,利用最小乐观概念的外延与决策信息系统中决策属性等价类之间的蕴含关系提取最简规则。最后,通过实例验证与数值实验进一步说明该方法的正确性与有效性。本文有以下贡献:1)提出了一种新的最小乐观概念格形式,由其特点可知,所得到的决策规则是最简规则,无需增加去除冗余属性的过程;2)利用最小乐观概念的外延与决策属性等价类之间的联系进行规则提取,简化了规则的判断,降低了算法的复杂度;3)本文避免了传统概念格规则提取算法需要扩展决策信息系统的过程,降低了算法的空间和时间复杂度;4)通过设定最终算法的结束条件,即提取完规则的最小乐观概念的外延并集是否等于论域,从而加快算法的收敛速度,降低计算成本;5)从决策信息系统直接生成概念和概念之间的关系,丰富了决策信息系统知识的含义。本文算法的不足之处在于,最小乐观概念的求取过程较传统形式概念更简单,在一定程度上降低了算法的复杂度,但其时间复杂度依然呈指数阶,在对大规模数据集进行测试时会显示出一定的劣势;此外,决策信息系统各属性子集下的最小乐观概念的计算是相互独立的,本文算法并没有对其实现并行计算。因此,对于大规模的决策信息系统的并行化规则提取,相关研究工作仍在继续。

## 参考文献

- [1] WANG G Y. Rough Set Theory and Knowledge Acquisition [M]. Xi'an: Xi'an Jiaotong University Press, 2001: 147-152.
- [2] MENG Z Q, SHI Z Z. On rule acquisition methods for data classification in heterogeneous incomplete decision systems [J]. Knowledge-Based Systems, 2020, 193: 1-22.
- [3] MATSUMOTO Y, WATADA J. Knowledge Acquisition from

Rough Sets Using Merged Decision Rules [J]. Advanced Computational Intelligence and Intelligent Informatics, 2018, 22 (3): 404-410.

- [4] CHEN Z H, ZHANG Y, XIE G. Mining algorithm for concise decision rules based on granular computing [J]. Control & Decision, 2015, 30(1): 143-148.
- [5] WILLE R. Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts [M]. Springer Netherlands, 1982: 445-470.
- [6] ZHI H L, LI J H. Granule description based knowledge discovery from incomplete formal contexts via necessary attribute analysis [J]. Information Science, 2019, 485: 347-361.
- [7] WEI L, LIU L, QI J J, et al. Rules acquisition of formal decision contexts based on three-way concept lattices [J]. Information Science, 2020, 516: 529-544.
- [8] LI J H, MEI C L, KUMAR C A, et al. On Rule Acquisition in Decision Formal Contexts [J]. International Journal of Machine Learning & Cybernetics, 2013, 4(6): 721-731.
- [9] THUY N N, WONGTHANAVASU S. An efficient stripped cover-based accelerator for reduction of attributes in incomplete decision tables [J]. Expert Systems With Applications, 2020, 143: 113076. 1-113076. 15.
- [10] 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.
- [11] JOSÉ BENÍTEZ-CABALLERO M, MEDINA J, RAMÍREZ-POUSSA E, et al. Rough-set-driven approach for attribute reduction in fuzzy formal concept analysis [J]. Fuzzy Sets and Systems, 2019, 391: 117-138.
- [12] XIAO Q M, LANG G M, LIU W J, et al. The approximations in rough concept lattice and approximable concept lattice [J]. Journal of Intelligent & Fuzzy Systems, 2017, 33(6): 3459-3467.
- [13] LI J H, REN Y, MEI C L, et al. A comparative study of multi-granulation rough sets and concept lattices via rule acquisition [J]. Knowledge-Based Systems, 2016, 91: 152-164.
- [14] ZHU Z Q, LI H, DAI G Y, et al. A rough set multi-knowledge extraction algorithm and its formal concept analysis [C] // International Conference on Intelligent Systems Design and Applications, 2015: 25-29.
- [15] LI J H, MEI C L, ZHANG H Y. Attribute Reduction Method for Formal Decision Contexts Based on Genetic Algorithm and its Application to Decision-making Analysis [J]. Journal of Chinese Computer Systems, 2015, 36(8): 1803-1808.
- [16] KANG X P, MIAO D Q. A study on information granularity in formal concept analysis based on concept-bases [J]. Knowledge-Based Systems, 2016, 105: 147-159.



**WEN Xin**, born in 1993, postgraduate. Her main research interests include granular computing and knowledge engineering.



**CHEN Ze-hua**, born in 1974, professor, is a senior member of China Computer Federation. Her main research interests include data mining and knowledge discovery.