

# 基于属性重要度的变精度邻域粗糙集属性约简算法

郑文彬<sup>1,2</sup> 李进金<sup>3</sup> 何秋红<sup>1,2</sup>

(闽南师范大学计算机学院 福建漳州 363000)<sup>1</sup>

(福建省粒计算及其应用重点实验室(闽南师范大学) 福建漳州 363000)<sup>2</sup>

(闽南师范大学数学与统计学院 福建漳州 363000)<sup>3</sup>

**摘要** 邻域粗糙集理论主要用于知识发现、属性选择、决策分析和数据挖掘等领域,能够根据数据的特点选择合适的离散化策略,在处理模糊和不确定性知识方面表现良好。但是,传统粗糙集属性约简算法存在难以确保获得约简、约简后的粗糙集属性识别准确率低等不足。对此,文中提出了一种基于属性重要度的属性约简算法。在充分考虑现有条件信息熵多方面不足的基础上,借鉴变精度邻域粗糙集理论对阈值参数进行重选,以新的条件信息熵作为度量基准,根据决策信息系统中的偏好属性推导出偏好决策规则集。对偏好决策规则集进行粗糙规则提取,并通过邻域粒化方法建立了变精度邻域粗糙集模型。该模型在处理大规模粗糙集属性数据时,计算时间较长,冗余属性过多。针对该问题,给出了一种属性重要度评价策略,在此基础上通过融合多叉树理论设计了变精度邻域粗糙集属性约简算法。实验结果表明,与传统方法相比,所提算法的属性识别准确率为 92%,提高了 10% 左右,这充分验证了所设计的属性约简算法具有较强的有效性和较高的应用价值。

**关键词** 属性重要度,变精度邻域,粗糙集,属性约简,多叉树

中图法分类号 TP393

文献标识码 A

DOI 10.11896/jsjx.181102184

## Attribute Reduction Algorithm for Neighborhood Rough Sets with Variable Precision Based on Attribute Importance

ZHENG Wen-bin<sup>1,2</sup> LI Jin-jin<sup>3</sup> HE Qiu-hong<sup>1,2</sup>

(School of Computer Science, Minnan Normal University, Zhangzhou, Fujian 363000, China)<sup>1</sup>

(Lab of Granular Computing, Minnan Normal University, Zhangzhou, Fujian 363000, China)<sup>2</sup>

(School of Mathematics and Statistics, Minnan Normal University, Zhangzhou, Fujian 363000, China)<sup>3</sup>

**Abstract** Neighborhood rough set theory is mainly used for knowledge discovery, attribute selection, decision analysis and data mining, and other fields. It can choose appropriate discretization strategy based on the characteristics of data and perform well in dealing with fuzzy and uncertain knowledge, but the traditional rough set attribute reduction algorithm is difficult to obtain reduction, and the attribute recognition accuracy of reduced rough set is low. Therefore, this paper put forward a kind of attribute reduction algorithm based on attribute importance. Considering the shortcomings of conditional information entropy in many aspects, the threshold parameters are re-selected by using the theory of variable precision neighborhood rough set. Based on the new conditional information entropy as the measurement benchmark, the preference decision rule set is deduced according to the preference attributes in the decision information system. This paper extracted rough rules from preference decision rule set and established a variable precision neighborhood rough set model by using neighborhood granulation method. When dealing with large-scale rough set attribute data, this model takes a long time to calculate and has too many redundant attributes. Aiming at this problem, an evaluation strategy of attribute importance was given. Based on this, a variable precision neighborhood rough set attribute reduction algorithm was theoretically designed by fusing multi-tree. The experimental results show that compared with the traditional method, the accuracy of attribute recognition of the proposed method is 92%, which is improved by 10%. This fully verifies that the proposed attribute reduction algorithm has strong effectiveness and higher application value.

**Keywords** Attribute importance, Variable-precision neighborhood, Rough set, Attribute reduction, Multi-way tree

收稿日期:2018-11-27 返修日期:2019-03-08 本文受国家自然科学基金项目(11871259),国家自然科学基金青年基金项目(11701258),福建省自然科学基金项目(2019J01749)资助。

郑文彬(1971-),男,硕士,高级讲师,主要研究方向为粗糙集、粒计算、数据挖掘、人工智能,E-mail:zwnxzb@126.com(通信作者);李进金(1960-),男,博士,教授,主要研究方向为人工智能、粒计算、拓扑学;何秋红(1977-),女,硕士,讲师,主要研究方向为粗糙集及大数据技术。

## 1 引言

粗糙集理论是由波兰科学家 Pawlak 于 1982 年提出的一种具有信息科学性的数学计算方法。以粗糙集理论为基础,衍生出了邻域粗糙集理论,该理论通过在其邻域空间内创建上下近似来实现对目标信息的获取与近似逼近。邻域粗糙集理论自提出以后,便被人们广泛应用于知识发现、机器学习、模式识别以及归纳推理等领域。在使用粗糙集理论前,需要对连续属性值进行属性约简。而属性约简后的属性值没有全面保留原始邻域粗糙集属性值数据的差异,这将会导致在属性约简过程中产生不同程度的信息丢失。

针对这些问题,相关领域学者进行了研究。吴尚智等<sup>[1]</sup>提出了一种基于遗传粒子群算法的属性约简算法,该算法采用粗糙集属性依赖性计算属性核,并在种群迭代过程中将属性核作为限制条件,动态调节种群适应度函数,最终可达到粗糙集属性约简的目的。该算法对规模较大的粗糙集属性数据进行约简的速度较快,但也存在数据丢失现象。安若铭等<sup>[2]</sup>提出了一种权重计算方法,并将其应用到粗糙集属性数据约简过程中,在计算出粗糙集属性数据最优权值的基础上,给出了最优组合权值的计算方式,进一步将计算结果应用到粗糙集属性特征值的分析过程中,完成粗糙集属性的约简。该方法能够有效减少粗糙集属性特征值的数量,但若权值计算结果不准确,将会导致属性约简达不到理想效果。王映龙等<sup>[3]</sup>提出了一种基于启发式的粗糙集属性约简算法。该算法的约简原则是:将新的粗糙集属性信息增加至决策系统时,利用原信息决策系统的属性约简结果更新粗糙集属性约简结果,并去除更新后的约简结果中的一些冗余属性。该方法的计算效率较高,但约简后的粗糙集属性识别准确率较低。

针对上述方法存在的问题,笔者提出了一种基于属性重要度的变精度邻域粗糙集属性约简算法。

## 2 属性约简算法

### 2.1 基于模糊决策系统的变精度邻域粗糙集

设定知识描述系统是一种多属性决策信息系统,该系统中, $U$  表示论域, $A$  表示非空有限集合的属性集, $A=C \cup Cl$ ,  $C \cap Cl = \Phi$ ,  $C$  和  $Cl$  分别用于描述非空有限集合中的两种属性集, $V_c = \{V_q | q \in C\}$  表示条件属性值集合, $V_a = \{V_d | d \in Cl\}$  表示决策属性值集合。 $V_c = \{V_q | q \in C\}$  与  $V_a = \{V_d | d \in Cl\}$  具有偏好顺序; $f: U \times A \rightarrow V$  为一个粗糙集信息函数,表示对任何一个  $x \in U, a \in A, f(x, a) \in V_a$ 。令  $x \in U, y \in U$ , 对于  $P \subseteq C, q \in P$ , 假设偏好属性值为  $f(y, q) \geq f(x, q)$ , 记作  $yD_P x$ , 即  $D_P = \{\{x, y\} \in U \times U | f(y, q) \geq f(x, q)\}$ , 将这种联系称为优势关联<sup>[4]</sup>。 $D_P^+(x) = \{y | yD_P x\}$  为优势集合,  $D_P^-(x) = \{y | xD_P y\}$  为劣势集合。

根据设定的决策属性集合,把集合解中的论域划分为有限量的决策类别。设定  $Cl = \{Cl_t | t \in Z^+\}, Cl_n > \dots > Cl_1 > \dots > Cl_1$ , 将不同类别的朝上并集、朝下并集表示为:

$$Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s, Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s, t, s \in Z^+ \quad (1)$$

决策者按照原有的综合评价把决策信息中的解移到  $Cl$  的类别中:最不好的解划分到  $Cl_1$  类别,最好的解划分到  $Cl_n$

类别,剩下的解划入  $Cl_t$  类别。按照这个规则,  $t \in Z^+ = \{1, 2, \dots, n\}$  越大,则类别  $Cl_t$  越好。

假设  $A = \{a_1, a_2, \dots, a_{|A|}\}$  为非空有限集合,  $A = C \cup Cl$ ,  $C \cap Cl = \Phi, a \in A, V_a$  是属性  $a$  的值域;  $f(x, a) \in V_a$  是信息函数,对于  $a \in A, x \in U, f(x, a) \in V_a$ , 它代表有限非空集合中的任何一对解的偏好属性值<sup>[5]</sup>。设定信息属性条件集合为  $P \subseteq C, x \in U, Cl_t \subseteq C, t \in \{1, 2, \dots, n\}$ , 以及阈值参数为  $0.5 < \beta \leq 1$ , 则  $Cl_t^{\geq}$  的  $\beta$ -上下近似各自设定如下:

$$\overline{apr_{\beta}^{\geq}(Cl_t^{\geq})} = \bigcup \left\{ x \in U \mid \frac{|D_P^+(x) \cap Cl_t^{\geq}|}{|D_P^+(x)|} \geq \beta \right\} \quad (2)$$

$$\underline{apr_{\beta}^{\geq}(Cl_t^{\geq})} = \bigcup \left\{ x \in U \mid \frac{|D_P^-(x) \cap Cl_t^{\geq}|}{|D_P^-(x)|} \geq 1 - \beta \right\} \quad (3)$$

其中,  $|T|$  代表集合  $T$  中的元素数量。

同理,  $Cl_t^{\leq}$  的  $\beta$ -下上近似分别设置为:

$$\underline{apr_{\beta}^{\leq}(Cl_t^{\leq})} = \bigcup \left\{ x \in U \mid \frac{|D_P^-(x) \cap Cl_t^{\leq}|}{|D_P^-(x)|} \geq \beta \right\} \quad (4)$$

$$\overline{apr_{\beta}^{\leq}(Cl_t^{\leq})} = \bigcup \left\{ x \in U \mid \frac{|D_P^+(x) \cap Cl_t^{\leq}|}{|D_P^+(x)|} \geq 1 - \beta \right\} \quad (5)$$

当满足  $\beta = 1$  时,  $\beta$ -上下近似分别简化为优势关系下的粗糙近似。

设定  $Cl$  的分类质量<sup>[6]</sup>为:

$$\gamma_{\beta}^{\geq}(Cl) = \frac{\bigcup \left\{ \frac{|D_P^+(X) \cap Cl_t^{\geq}|}{|D_P^+(X)|} \geq \beta \right\}}{\bigcup \left\{ \frac{|D_P^-(x) \cap Cl_t^{\leq}|}{|D_P^-(x)|} \geq \beta \right\}} \quad (6)$$

其中,  $Cl$  的分类质量  $\gamma_{\beta}^{\geq}(Cl)$  代表了在  $\beta$  阈值水平情况下信息决策系统中正确分类的属性数量和总属性数量的对比率。 $C$  是邻域粗糙集的最小属性约简集合, 满足  $\gamma_{\beta}^{\geq}(Cl) = \gamma_{\beta}^{\leq}(Cl)$ , 是  $Cl$  的分类质量中的一个  $\beta$ -约简。约简可以将整体完全分解成多个不含有其他属性的最小子集。根据观察了解, 属性间存在依赖关系, 这样就会产生一个最小条件的约简集合, 该集合具有与初始集合一样的分类质量<sup>[7]</sup>。决策系统由一个或者多个约简集合组成, 而决策系统中最重要的核由全部约简的交集构成。因此, 核也可以称为最关键的属性集, 但其也会出现空集的现象<sup>[8]</sup>。

给出属性集合  $P \subseteq C, x \in U, y \subseteq C$ , 在向上合并集合  $Cl_t^{\geq}$  中, 若  $t \in \{1, 2, \dots, n\}$ , 则以下条件有一条成立时, 就称为优势关联的不相容(矛盾)性<sup>[9]</sup>。

1)  $x \in Cl_t^{\geq}$ , 则  $x$  在类别  $Cl_t$  或者比其更优的分类中, 考虑到偏好属性时,  $y$  具有与  $x$  一样或者远超过  $x$  的优势, 则  $x \in Cl_t^{\geq}, D_P^+(x) \cap Cl_{t-1}^{\leq} \neq \Phi$ ;

2)  $x \notin Cl_t^{\geq}$ , 则  $x$  在比  $Cl_t$  差的分类中, 考虑到偏好属性时,  $x$  具有与  $y$  一样或者远超过  $y$  的优势; 但在实际情况中,  $y$  往往存在于一个比  $x$  更具优势的集中。因此,  $x \notin Cl_t^{\geq}, D_P^+(x) \cap Cl_t^{\leq} \neq \Phi$ 。

不考虑属性集的向上合并集合  $Cl_t^{\geq}$  的指向性信息时, 满足  $Cl_t^{\geq} = Cl_t^{\leq}$ , 此时属性集之间的独立性为不可划分关系概念下的独立性。根据此特点, 设定基于优势关系不可划分类别的基准原则。

满足以下条件的任一条件的属性集类别, 可被看作基

于势关系的不可划分类。

1)在等价联系过程中,条件属性为不可分辨的分类;

2)在优势关系下,通过对决策偏好属性子集的划分,了解到条件偏好属性间各子集是互不兼容的。

对模型中各个类别的对象建立一系列完整的规则是尤其重要的,而约简在构建规则的过程中是必不可少的。首先需要得到优势关系下的粗糙约简,其次由决策<sup>[10]</sup>信息系统中的属性推断出决策规则集合。把决策规则集合分成两大类,一类是向上并集合,另一类是向下并集合,则产生的决策语法规则表示如下。

1)最终构建一系列规则

如果满足  $f(x, q_1) \geq r_{q_1} \wedge \dots \wedge f(x, q_p) \geq r_{q_p}$ , 则有  $x \in Cl_i^{\geq}$ 。

这些规则都由向上合并集合解支持<sup>[11-12]</sup>。

2)生产决策语法规则

如果满足  $f(x, q_1) \leq r_{q_1} \wedge \dots \wedge f(x, q_p) \leq r_{q_p}$ , 则有  $x \in Cl_i^{\leq}$ 。

这些决策规则都由向下合并集合解支持。

## 2.2 基于属性重要度的粗糙集属性约简

寻找最小约简解属于 NP 问题。基于属性重要度的搜索算法已经被广泛运用到属性约简中,这些算法能够有效地得到一些次优约简结果,并且只需要付出较少的时间即可完成属性约简。例如,传统的机器学习算法把香农信息熵视为属性重要性的度量。这种建立在构建模糊相似联系下的信息熵不确定性度量,可以有效提高度量属性的分辨能力<sup>[13-14]</sup>。在信息表中,信息熵值越大,粗糙集属性的辨别率就越高,属性的重要度也就越高。下文给出粗糙集属性重要度的设定。

若  $(U, A, V, f)$  表示信息系统,  $B \subseteq A$ , 则  $\forall a \in B$  的属性重要度设定为:

$$SIG(a, B) = H^a(B) - H^a(B - \{a\}) \quad (7)$$

若  $A = C \cup d, B \subseteq A, \forall a \in B \cup d$ , 则  $d$  的属性重要度设定为:

$$SIG(a, B, d) = H^a(d|B - \{a\}) - H^a(d|B) \quad (8)$$

针对粗糙集混合属性,将上述不同的属性重要度视为启发信息,利用向前搜索构建了基于信息熵的邻域粗糙集属性约简算法<sup>[15]</sup>,其包含约简与相对约简两方面,具体描述如算法 1、算法 2 所示。

### 算法 1 基于信息熵的混合属性约简算法

输入:一个完整的信息表集合

输出:该信息表的约简集合

Step1 对于  $\forall a \in C$ , 计算邻域粗糙相似关系为  $M(R) = (r_{ij})_{n \times n}$ 。

Step2 开始邻域粗糙约简集合计算。

Step3 对于  $\forall a \in (C - \text{reduct})$ , 计算任何一个邻域粗糙约简集合的重要度  $S_i = SIG(a, C - \text{reduct})$ 。

Step4 选择约简属性重要度最大的属性,则有  $SIG(a, C - \text{reduct}) = \max_i(S_i)$ 。

Step5 假设  $SIG(a, C - \text{reduct}) > \delta$ , 则把邻域粗糙属性加入到约简集合中,  $a \rightarrow \text{reduct}$ , 返回 Step3。

### 算法 2 基于信息熵的粗糙集混合属性相对约简算法

输入:一个完整的信息决策集合

输出:此决策表的约简集合

Step1 对于  $\forall a \in C$ , 计算邻域粗糙相似关系  $M(R) = (r_{ij})_{n \times n}$ 。

Step2 初始化邻域约简集。

Step3 对于  $\forall a \in (C - \text{reduct})$ , 计算任何一个邻域粗糙属性相对决策属性的重要度,即  $S_i = SIG(a, C - \text{reduct}, d)$ 。

Step4 选择约简属性重要度最大的属性,则有  $SIG(a, C - \text{reduct}, d) = \max_i(S_i)$ 。

Step5 假设  $SIG(a, C - \text{reduct}, d) > \delta$ , 则把邻域粗糙属性加入到约简集合中,  $a \rightarrow \text{reduct}$ , 返回 Step3。

在粗糙属性约简过程中,约简属性集合和出现属性集合的信息熵不一定相同<sup>[16]</sup>。在模糊近似关系下,上述条件不一定成立。对截止条件设定相应的阈值,当满足  $SIG(a, C - \text{reduct}) \leq \delta$  或者  $SIG(a, C - \text{reduct}, d) \leq \delta$  时,约简过程结束。在实际应用过程中,合理的  $\delta$  值可依据原有粗糙集属性的数量和总体的信息熵值进行设定。

## 2.3 融合多叉树的粗糙集属性约简算法

基于属性重要度的粗糙集属性约简算法无法对规模较大的变精度邻域粗糙集进行属性约简,针对此弊端,结合多叉树理论,构建一种融合多叉树的变精度粗糙集属性约简算法。

设置一个知识描述系统  $IS = (U, C, V, f)$ , 其中  $|C| = h$ ,  $|U| = n$ , 它所对应的邻域粗糙集则是高为  $h + 1$  的树。由于邻域粗糙集的最小属性约简集合  $C$  中的属性特征排列顺序不相同,导致多叉树的结构也不相同<sup>[17-18]</sup>。例如,粗糙集系统的属性约简集合的排列顺序为  $\alpha_1, \alpha_2, \dots, \alpha_h$ , 而得出的粗糙集多叉树结果具有以下特征:

1)除了在  $h$  层的叶子层的节点外,处于相同层中的所有节点的取值都是一样的,设定每一个叶子层节点的取值为  $\alpha_1, \alpha_2, \dots, \alpha_h$ ;

2)第一层节点与其他层节点的数量都是由粗糙集属性  $\alpha_1$  的取值数量所决定的,其他非终端节点的数量不同,但至少保证有一个节点,最多则是与粗糙集属性的取值个数相同<sup>[19-21]</sup>;

3)所有叶子节点都处于相同的层,即粗糙集多叉树从  $h + 1$  层开始,同时叶子节点的数量不大于  $n$ , 每一个叶子节点和粗糙集系统的全部条件属性在论域中,划分后形成的等价类必须一一对应。

表 1 为知识表达系统的信息表,该粗糙集系统表格中共有 8 个数据,用于叙述各个数据的  $a, b$  和  $c$  3 个特征属性。

表 1 粗糙集信息系统对应的信息表

Table 1 Information table corresponding to rough set information system

$U$	$a$	$b$	$c$
$x_1$	2	0	0
$x_2$	2	2	4
$x_3$	4	2	4
$x_4$	0	0	2
$x_5$	2	4	2
$x_6$	0	2	4
$x_7$	2	2	0
$x_8$	0	4	2

由表 1 可以分析出多叉树的构建过程。当粗糙集属性序列为  $\{a, b, c\}$  时,将目标  $x_1$  代入到多叉树结果中,在初始阶

段,由于多叉树中没有节点,因此首先需要在这个空树中加入一个根节点,根节点的值为粗糙集属性序列中的首个元素。其次,观测目标  $x_1$  在属性  $a$  上的值。假设其取值为 1,则走分支“1”至第二层节点,并在这个节点上添加节点取值元素  $b$ ,再观测对象在属性  $b$  上的值;假设其取值为 0,则走分支“0”至第三层节点,加入节点取值元素  $x_2$ ,观测对象  $x_2$  在属性  $c$  上的值。当粗糙集属性序列中的各个元素均添加完成时,再加入一个叶子节点元素  $x_1$ ,至此分叉树分支添加结束。对于变精度邻域粗糙集论域中的其他对象,根据上述步骤逐次加入到多叉树中,即可完成信息系统相应多叉树结构的构建。

由上述过程可知,融合于多叉树理论的基于属性重要度的粗糙集属性约简算法的求核思想是:对粗糙集中的全部属性信息逐次执行相同的操作,即将属性信息从粗糙集的全部属性数据集中提取出来,并根据剩余属性信息构成的信息表来创建相应的多叉树;同时需要判定叶子节点数量与全部粗糙集属性信息集合构造的多叉树结构的叶子节点数量是否相同,若相同,则说明这个元素信息对于整个信息系统是冗余的,该属性信息为非核元素,将其除去,否则,该属性信息是必要的,为核元素之一;最后,输出粗糙集内的全部必要属性。

### 3 仿真实验与结果分析

本节通过仿真实验来验证所提属性约简算法的综合性能。仿真实验环境为:Windows7 操作系统,4 GB 内存,Intel Core i5-3470 处理器;采用 MatlabR2012a 软件搭建仿真实验平台。

为了验证所提属性约简算法的优势,首先利用所提约简算法与传统约简算法同时对属性较多的粗糙集数据进行特征约简实验,然后采用某种识别算法计算不同算法进行属性约简前后的识别准确率,最后依据这两种特征约简识别后的准确率来检测这几种属性约简算法的性能。实验所用属性较多的粗糙数据集的属性信息如表 2 所列。

表 2 粗糙数据集的属性信息

粗糙数据集	样本个数/个	属性数/个	数据类别/类
vehicle	1692	36	8
ionosphere	702	68	4
sonar	416	120	4

原邻域粗糙集的邻域约简值基本设定为 0.15 左右,实验中原邻域粗糙集的邻域约简值也设定为 0.15。本文分别利用融合了多叉树理论的基于属性重要度的粗糙集属性约简算法与传统的邻域粗糙集特征约简算法对属性较多的目标粗糙集进行属性特征约简,得到的较高识别准确率对应的约简后的属性数量如表 3 所列,与其相应的属性特征约简后的属性结果分别如表 4 和表 5 所列。

表 3 利用两种算法进行属性特征约简后的属性结果

Table 3 Attribute results after attribute feature reduction using two algorithms

粗糙数据集	vehicle	ionosphere	sonar
原始属性数	36	68	120
传统算法约简后属性数	6	16	36
改进算法约简后属性数	26	28	14

表 4 利用传统算法进行属性特征约简后的属性结果

Table 4 Attribute results after attribute feature reduction using traditional algorithm

粗糙数据集	约简后属性结果
vehicle	[24;20;2]
ionosphere	[10;2;18;28;48;16;52;40]
sonar	[90;72;42;52;46;66;38;48;44;22;24;54;28;68;80;86;92]

表 5 利用改进算法进行属性特征约简后的属性结果

Table 5 Attribute results after attribute feature reduction using improved algorithm

粗糙数据集	约简属性后结果	k 值
vehicle	[22;12;34;30;32;10;26;2;28;36;4;6;14]	1
ionosphere	[10;2;6;56;16;40;12;14;22;20;60;68]	1
sonar	[22;32;72;32;6;10;2]	1

根据表 4 和表 5 所列的约简后的变精度邻域粗糙集属性,采用“一对多”支持向量机的识别算法对原始变精度粗糙集属性约简算法、融合了多叉树理论的基于属性重要度的变精度粗糙集属性约简算法进行约简前后的属性进行识别,利用两种方法进行属性约简后的识别准确率情况如图 1 所示。

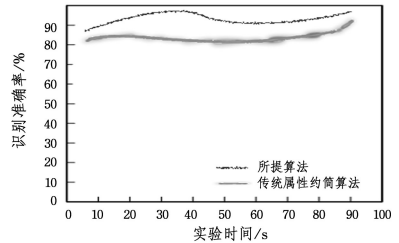


图 1 不同算法进行属性特征约简后的识别准确率

Fig. 1 Recognition accuracy of different algorithms after attribute reduction

由图 1 可知,利用传统变精度邻域粗糙集属性约简算法进行约简后的属性中,大部分相对重要的属性被去除,导致约简后属性的识别准确率有所降低,且会消耗大量的时间。利用融合了多叉树理论的基于属性重要度的变精度邻域粗糙集属性约简算法进行约简后,重要属性均被保留,对变精度邻域粗糙集特征进行有效约简后,降低了特征维数,减少了计算量,缩短了计算时间,保证了识别准确率。

**结束语** 针对传统粗糙集属性约简算法存在的问题,本文提出了一种融合了多叉树理论的变精度邻域粗糙集属性约简算法,并采用属性较多的粗糙集属性数据集对所提算法、传统属性约简算法进行了特征约简实验。通过实验可以获得以下结论:

- 1) 基于属性重要度的粗糙集属性约简算法不仅保证了约简后数据属性识别的准确率,同时也降低了属性特征的维数,提升了粗糙集属性识别的速度;
- 2) 从约简后的丢失率对比情况可以看出,所提方法的丢失率明显低于传统属性约简算法的丢失率,说明所提算法在进行属性约简过程中,能够有效保证数据的完整性。

### 参考文献

[1] WU S Z, LUO Y C, ZHAI J P. A minimum attribute reduction algorithm based on genetic & particle swarm optimization and

- rough sets[J]. Computer Engineering and Science, 2016, 38(5): 1007-1013. (in Chinese)
- 吴尚智, 罗艺纯, 翟敬鹏. 基于遗传粒子群和粗糙集的最小属性约简算法[J]. 计算机工程与科学, 2016, 38(5): 1007-1013.
- [2] AN R M, SUO M L. Application of attributes reduction and weights calculation through neighborhood rough set[J]. Computer Engineering and Applications, 2016, 52(7): 160-165. (in Chinese)
- 安若铭, 索明亮. 邻域粗糙集在属性约简及权重计算中的应用[J]. 计算机工程与应用, 2016, 52(7): 160-165.
- [3] WANG Y L, HUA J J, QIAN W B, et al. Dynamic algorithm of attribute reduction in set-valued decision information system [J]. Computer Engineering and Applications, 2017, 53(17): 60-64. (in Chinese)
- 王映龙, 华佳佳, 钱文彬, 等. 集值决策信息系统的动态属性约简算法[J]. 计算机工程与应用, 2017, 53(17): 60-64.
- [4] YAO S, WANG J, XU F, et al. Uncertainty measurement and attribute reduction in incomplete neighborhood rough set[J]. Journal of Computer Applications, 2018, 38(1): 97-103. (in Chinese)
- 姚晟, 汪杰, 徐风, 等. 不完备邻域粗糙集的不确定性度量及属性约简[J]. 计算机应用, 2018, 38(1): 97-103.
- [5] CHANG H Y, MENG Z Q. New Heuristic Algorithm for Attribute Reduction in Decision-theoretic Rough Set[J]. Computer Science, 2016, 43(6): 218-222. (in Chinese)
- 常红岩, 蒙祖强. 一种新的决策粗糙集启发式属性约简算法[J]. 计算机科学, 2016, 43(6): 218-222.
- [6] HONG H J, YE D Y. An Evolutionary Algorithm for Multi-objective Attribute Reduction Involving Optimization of the Number of Decision Rules[J]. Journal of Chinese Computer Systems, 2016, 37(8): 1707-1711. (in Chinese)
- 洪华剑, 叶东毅. 含规则数优化的多目标属性约简进化算法[J]. 小型微型计算机系统, 2016, 37(8): 1707-1711.
- [7] XIONG F, ZHANG X Y. Regional attribute reduction and their structural heuristic algorithms for variable precision rough sets [J]. Journal of Computer Applications, 2016, 36(11): 2954-2957. (in Chinese)
- 熊方, 张贤勇. 变精度粗糙集的区域属性约简及其结构启发算法[J]. 计算机应用, 2016, 36(11): 2954-2957.
- [8] ZHANG Q N, LI D M, XU J, et al. Reduction algorithm for variable precision rough sets[J]. Fuzzy Systems and Mathematics, 2017(6): 132-135. (in Chinese)
- 张秋娜, 李冬梅, 徐珺, 等. 变精度粗糙集的约简算法[J]. 模糊系统与数学, 2017(6): 132-135.
- [9] WANG Y L, ZENG Q, QIAN W B, et al. Attribute reduction algorithm of the incomplete neighborhood decision system with variable precision [J]. CAAI Transactions on Intelligent Systems, 2017, 12(3): 386-391. (in Chinese)
- 王映龙, 曾淇, 钱文彬, 等. 变精度下不完备邻域决策系统的属性约简算法[J]. 智能系统学报, 2017, 12(3): 386-391.
- [10] ZHU H, DOU H L. Attribute reduction for decision-theory model on the idea of local [J]. Electronic Design Engineering, 2017, 25(21): 64-67. (in Chinese)
- 朱辉, 窦慧莉. 基于决策粗糙集模型的局部属性约简[J]. 电子设计工程, 2017, 25(21): 64-67.
- [11] LIU D J, LI L. Group Evaluation Model of Information Aggregation Based on Attribute Reduction [J]. Computer Simulation, 2016, 33(3): 371-375. (in Chinese)
- 刘东君, 李力. 基于属性约简的群体评价信息集结模型研究[J]. 计算机仿真, 2016, 33(3): 371-375.
- [12] MA F M, CHEN J W, ZHANG T F. Quick attribute reduction algorithm for neighborhood multi-granulation rough set based on double granulate criterion [J]. Control & Decision, 2017, 32(6): 1121-1127.
- [13] LIU Y, XIE H, WANG L, et al. Hyperspectral band selection based on a variable precision neighborhood rough set [J]. Applied Optics, 2016, 55(3): 462-481.
- [14] CHEN D, YANG Y, DONG Z. An incremental algorithm for attribute reduction with variable precision rough sets [J]. Applied Soft Computing, 2016, 45: 129-149.
- [15] CHEN Y, ZENG Z, LU J. Neighborhood rough set reduction with fish swarm algorithm [J]. Soft Computing, 2016, 21(23): 1-12.
- [16] KUMAR S U, INBARANI H H. PSO-based feature selection and neighborhood rough set-based classification for BCI multi-class motor imagery task [J]. Neural Computing & Applications, 2017, 28(11): 3239-3258.
- [17] LIU Y, XIE H, TAN K, et al. Hyperspectral band selection based on consistency-measure of neighborhood rough set theory [J]. Measurement Science & Technology, 2016, 27(5): 550-551.
- [18] PAL U, SHARMISTHA B H, DEBNATH K. R implementation of Bayesian Decision Theoretic Rough Set [J]. Model for Attribute Reduction, 2018, 22: 25-33.
- [19] WANG X Y, SHEN J L, SHEN Y X, et al. Incomplete Weighted Grade Multi-granulation Rough Set and Granular Reduction [J]. Journal of Chinese Computer Systems, 2017, 11: 141-146.
- [20] CHEN Y, YANG D. Attribute Reduction Algorithm Based on Information Entropy and Its Application [J]. Journal of Chongqing University of Technology (Natural Science), 2013, 27(1): 42-46. (in Chinese)
- 陈媛, 杨栋. 基于信息熵的属性约简算法及应用[J]. 重庆理工大学学报(自然科学), 2013, 27(1): 42-46.
- [21] CHEN Y, GOU G L, LU L. Improved Algorithm for Attribute Reduction based on Consistency Criterion [J]. Journal of Chongqing University of Technology (Natural Science), 2014, 28(5): 79-83, 92. (in Chinese)
- 陈媛, 苟光磊, 卢玲. 基于一致性准则的属性约简改进算法[J]. 重庆理工大学学报(自然科学), 2014, 28(5): 79-83, 92.