



## 基于对象变化的邻域决策粗糙集动态更新算法

孙海霞

引用本文:

孙海霞. 基于对象变化的邻域决策粗糙集动态更新算法[J]. 智能系统学报, 2021, 16(4): 746–756.

SUN Haixia. Dynamic updating algorithm of neighborhood decision-theoretic rough set model based on object change[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(4): 746–756.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202010028>

## 您可能感兴趣的其他文章

### 相似度三支决策模糊粗糙集模型的决策代价研究

Decision costs of the similarity three-way decision-theoretic fuzzy rough set model

智能系统学报. 2020, 15(6): 1068–1078 <https://dx.doi.org/10.11992/tis.201909015>

### 三支决策-基于粗糙集与粒计算研究视角

Three-way decisions: research perspectives for rough sets and granular computing

智能系统学报. 2019, 14(6): 1111–1120 <https://dx.doi.org/10.11992/tis.201905039>

### 概率粗糙集三支决策在线快速计算算法研究

Research on a fast online computing algorithm based on three-way decisions with probabilistic rough sets

智能系统学报. 2018, 13(5): 741–750 <https://dx.doi.org/10.11992/tis.201706047>

### 变精度下不完备邻域决策系统的属性约简算法

Attribute reduction algorithm of the incomplete neighborhood decision system with variable precision

智能系统学报. 2017, 12(3): 386–391 <https://dx.doi.org/10.11992/tis.201705027>

### 面向成组对象集的增量式属性约简算法

An incremental attribute reduction algorithm for group objects

智能系统学报. 2016, 11(4): 496–502 <https://dx.doi.org/10.11992/tis.201606005>

### 基于证据理论刻画多粒度覆盖粗糙集的数值属性

Evidence-theory-based numerical characterization of multi-granulation covering rough sets

智能系统学报. 2016, 11(4): 481–486 <https://dx.doi.org/10.11992/tis.201606011>

微信公众平台



关注微信公众号，获取更多资讯信息

DOI: 10.11992/tis.202010028

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.TP.20210409.1403.008.html>

# 基于对象变化的邻域决策粗糙集动态更新算法

孙海霞

(安徽三联学院 计算机工程学院, 安徽 合肥 230601)

**摘 要:** 针对现实环境下数据集不断动态变化的特性, 提出一种邻域决策粗糙集模型的增量式更新算法。采用由简单到复杂的研究思路, 分析了邻域型信息系统论域增加和减少单个对象时, 目标近似集与邻域类之间概率的变化规律, 进一步地利用这种规律来构造单个对象变化时邻域决策粗糙集模型上下近似集的增量式更新, 在单个对象变化的基础上, 通过逐步迭代的方式设计了对象批量变化时的增量式更新算法。实验分析表明, 所提出的算法具有较高的增量式更新性能, 适用于动态数据环境下邻域决策粗糙集模型的动态更新。

**关键词:** 粗糙集; 决策粗糙集; 邻域; 增量式学习; 近似集; 对象; 迭代; 动态

**中图分类号:** TP18    **文献标志码:** A    **文章编号:** 1673-4785(2021)04-0746-11

中文引用格式: 孙海霞. 基于对象变化的邻域决策粗糙集动态更新算法 [J]. 智能系统学报, 2021, 16(4): 746-756.

英文引用格式: SUN Haixia. Dynamic updating algorithm of neighborhood decision-theoretic rough set model based on object change[J]. CAAI transactions on intelligent systems, 2021, 16(4): 746-756.

## Dynamic updating algorithm of neighborhood decision-theoretic rough set model based on object change

SUN Haixia

(College of Computer Engineering, Anhui Sanlian University, Hefei 230601, China)

**Abstract:** In accordance with the dynamic characteristics of a dataset in a real environment, an incremental updating algorithm of the neighborhood decision-theoretic rough set model is proposed after conducting simple to complex research. In this paper, the change law of the probability between the target approximation set and the neighborhood class is first analyzed in the context of the domain of the neighborhood information system increasing or decreasing individual objects, which is then adopted to construct the incremental updating of the upper and lower approximation sets of the neighborhood decision-theoretic rough set model. On the basis of the change of a single object and given the variety of multiple objects, an incremental updating algorithm is designed through step-by-step iteration. Experimental results show that the proposed algorithm has a high incremental updating performance, thereby making it suitable for the dynamic updating of the neighborhood decision-theoretic rough set model in a dynamic data environment.

**Keywords:** rough set; decision-theoretic rough set; neighborhood; incremental learning; approximation set; object; iteration; dynamic

粗糙集理论是当今知识发现领域的一个研究热点, 决策粗糙集模型<sup>[1-2]</sup>是粗糙集理论的重要研究分支, 由于决策粗糙集在处理噪声数据方面具有较好的泛化性能<sup>[2]</sup>, 因此目前已广泛应用于机器学习<sup>[3-4]</sup>、图像处理<sup>[5]</sup>和数据挖掘<sup>[6]</sup>等诸多领域。

收稿日期: 2020-10-26. 网络出版日期: 2021-04-09.

基金项目: 安徽省质量工程项目 (2020ZYRC063, 2020XSKT151); 省教育厅高校自然科学重点项目 (KJ2019A0891).

通信作者: 孙海霞. E-mail: [wangbo19855@163.com](mailto:wangbo19855@163.com).

早期的决策粗糙集建立在完备离散型的信息系统上, 在其基础上, 学者们进一步地提出了多种扩展的粗糙集模型, 例如 Liu 等<sup>[7]</sup>将决策粗糙集推广至不完备信息系统, 提出一种改进的不完备决策粗糙集模型; Sun 等<sup>[8]</sup>在粗糙模糊集下提出了相应的决策粗糙集模型; Dou 等<sup>[9]</sup>基于多个代价的策略, 提出了多代价的决策粗糙集模型; 在数值型的数据集方面, Li 等<sup>[10]</sup>将邻域关系

融入决策粗糙集模型中, 提出了邻域决策粗糙集模型, 该模型进一步提升了决策粗糙集的适用范围。

然而现实环境下的数据不总是静止不变的, 而是时刻处于动态更新之中, 为了提高粗糙集模型在动态数据下的处理效率, 学者们对其提出了多种增量式的模型和算法<sup>[11-13]</sup>。在决策粗糙集等模型中, 学者们同样提出了多种增量式更新算法。例如, Zhang 等<sup>[14]</sup>针对属性值动态变化情形提出了相应的增量式更新算法; 针对流计算环境, Xu 等<sup>[15]</sup>提出了对象在线增加和减少时的增量式更新算法; Chen 等<sup>[16]</sup>利用集合更新的方法设计出了决策粗糙集的增量式更新算法; 赵小龙等<sup>[17]</sup>针对数值型信息系统, 研究了邻域粒化条件熵随对象增加和减少时的增量式更新, 并进一步地提出了对应的增量式属性约简算法; 杨臻等<sup>[18]</sup>研究了混合型信息系统下对象变化时概率的增量式更新, 并进一步地提出了变精度粗糙集模型的增量式更新; Luo 等<sup>[19]</sup>通过矩阵方法构造出了决策粗糙集的增量式更新算法。然而所提出的这些增量式更新算法仅局限于离散数据环境下的决策粗糙集模型, 针对邻域型的决策粗糙集还未有相关研究。

论域中对象的增加和减少是信息系统最为常见的一种变化形式, 本文将针对这类问题进行邻域决策粗糙集的增量式更新研究。邻域决策粗糙集通过邻域关系来对数值型数据进行信息粒化<sup>[10]</sup>, 而对其进行增量式计算时必然涉及邻域类的更新计算, 邻域关系不同于传统的等价关系和容差关系, 其增量式计算的复杂程度要更高。在文献[18]中, 杨臻等通过单个对象逐步迭代的方式去处理混合型信息系统变精度粗糙集模型的增量式更新, 使得问题的处理方式简便高效。由于决策粗糙集与变精度粗糙集有一定的相似性, 因此本文将借鉴文献[18]中关于变精度粗糙集模型的增量式更新研究思路与方法, 构造出邻域决策粗糙集的增量式更新模型。文中首先分别研究论域中增加和减少一个对象时, 近似集与邻域类之间概率的变化规律, 然后根据这种规律来构造单个对象变化时模型上下近似集的增量式更新, 最后在单个对象变化的基础上, 通过逐步迭代的方式设计了对象批量变化时的增量式更新算法。实验分析表明, 所提出的算法具有较高的增量式更新性能, 适用于动态数据环境下邻域决策粗糙集模型的动态更新。

## 1 邻域决策粗糙集模型

粗糙集理论中, 数据集表示为信息系统  $IS = (U, AT)$  的形式, 其中  $U$  为信息系统的对象集, 称为论域。AT 称为信息系统的属性集, 若属性集 AT 可分为条件属性集  $C$  和决策属性集  $D$ , 即  $AT = C \cup D$ , 那么该信息系统又称为决策信息系统。对于  $\forall x \in U$  和  $\forall a \in AT$ ,  $a(x)$  表示对象  $x$  在属性  $a$  下的属性值。当条件属性集  $C$  中的每个属性都为数值型属性时, 此信息系统又称之为邻域型信息系统。

Yao 等<sup>[1-2]</sup>提出的决策粗糙集仅应用于离散型的信息系统, Li 等将邻域关系引入传统的决策粗糙集模型中, 提出了邻域决策粗糙集<sup>[10]</sup>。

**定义 1**<sup>[10]</sup> 设邻域型信息系统  $IS = (U, AT)$ , 对于属性集  $A \subseteq AT$  确定的邻域关系  $N_A^\delta$  定义为

$$N_A^\delta = \{(x, y) \in U \times U \mid \Delta_A(x, y) \leq \delta\}$$

式中:  $\delta$  称为邻域半径, 是一个非负常数;  $\Delta_A(x, y)$  表示对象  $x$  与对象  $y$  之间的闵可夫斯基距离, 定义为

$$\Delta_A(x, y) = \left( \sum_{a \in A} |a(x) - a(y)|^k \right)^{1/k}$$

给定邻域信息系统的邻域关系, 可以进一步得到每个对象在该邻域关系下的邻域类。

**定义 2**<sup>[10]</sup> 设邻域型信息系统  $IS = (U, AT)$ , 属性集  $A \subseteq AT$  确定的邻域关系为  $N_A^\delta$ , 邻域半径为  $\delta$ 。那么对于  $\forall x \in U$  在邻域关系  $N_A^\delta$  下的邻域类定义为

$$\delta_A(x) = \{y \in U \mid (x, y) \in N_A^\delta\}$$

可以看出, 对象  $x$  的邻域类表示与  $x$  相近对象的集合, 其中  $\delta$  体现了相近的程度。基于邻域的思想, Li 等<sup>[10]</sup>提出了邻域决策粗糙集。

在邻域决策粗糙集中, 给定状态集  $\Omega = \{X, X^c\}$ , 其中  $X$  和  $X^c$  分别表示对象  $x$  隶属于  $X$  和  $X^c$  的两种情形, 那么对象  $x$  隶属于  $X$  的程度可通过概率  $P(X|\delta(x)) = \frac{|X \cap \delta(x)|}{|\delta(x)|}$  来表示, 则对象  $x$  隶属于  $X^c$  的程度可表示为  $P(X^c|\delta(x)) = 1 - P(X|\delta(x))$ 。给定对象  $x$  划分入  $X$  和  $X^c$  时的 3 种动作  $\Gamma = \{a_P, a_B, a_N\}$ , 其中  $a_P$ 、 $a_B$  和  $a_N$  分别表示划分入类别的正区域、边界域和负区域。这里令  $\lambda_{*P} (* \in \{P, B, N\})$  分别表示对象  $x$  实际属于  $X$  时采取  $a_P$ 、 $a_B$  和  $a_N$  动作时所产生的决策代价,  $\lambda_{*N} (* \in \{P, B, N\})$  分别表示对象  $x$  实际属于  $X^c$  时采取  $a_P$ 、 $a_B$  和  $a_N$  动作时所产生的决策代价。因此对于对象  $x$  进行不同决策动作时所产生的的代价为:

$$1) R_P = R(a_P|\delta(x)) = \lambda_{PP} \cdot P(X|\delta(x)) + \lambda_{PN} \cdot P(X^c|\delta(x));$$



$$2) R_B = R(a_B|\delta(x)) = \lambda_{BP} \cdot P(X|\delta(x)) + \lambda_{BN} \cdot P(X^c|\delta(x));$$

$$3) R_N = R(a_N|\delta(x)) = \lambda_{NP} \cdot P(X|\delta(x)) + \lambda_{NN} \cdot P(X^c|\delta(x))$$

基于最小化决策代价的原则, 那么有

$$1) R_P \leq R_B \text{ 且 } R_P \leq R_N \text{ 时, } x \in \text{POS}(X);$$

$$2) R_B \leq R_P \text{ 且 } R_B \leq R_N \text{ 时, } x \in \text{BUN}(X);$$

$$3) R_N \leq R_P \text{ 且 } R_N \leq R_B \text{ 时, } x \in \text{NEG}(X)。$$

通常决策代价满足关系  $\lambda_{PP} \leq \lambda_{BP} < \lambda_{NP}$ ,  $\lambda_{NN} \leq \lambda_{BN} < \lambda_{PN}$ , 那么可以得到

$$1) \text{ 当 } P(X|\delta(x)) \geq \alpha \text{ 且 } P(X|\delta(x)) \geq \gamma \text{ 时, } x \in \text{POS}(X);$$

$$2) \text{ 当 } P(X|\delta(x)) \leq \alpha \text{ 且 } P(X|\delta(x)) \geq \beta \text{ 时, } x \in \text{BUN}(X);$$

$$3) \text{ 当 } P(X|\delta(x)) \leq \beta \text{ 且 } P(X|\delta(x)) \leq \gamma \text{ 时, } x \in \text{NEG}(X);$$

这里的  $\alpha$ 、 $\beta$  和  $\gamma$  计算分别为

$$\alpha = \frac{\lambda_{PN} - \lambda_{BN}}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})}$$

$$\beta = \frac{\lambda_{BN} - \lambda_{NN}}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})}$$

$$\gamma = \frac{\lambda_{PN} - \lambda_{NN}}{(\lambda_{PN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{PP})}$$

进一步地, 若  $\frac{\lambda_{NP} - \lambda_{BP}}{\lambda_{BN} - \lambda_{NN}} > \frac{\lambda_{BP} - \lambda_{PP}}{\lambda_{PN} - \lambda_{BN}}$ , 则

$$1) \text{ 当 } P(X|\delta(x)) > \alpha \text{ 时, 那么 } x \in \text{POS}(X);$$

$$2) \text{ 当 } \beta \leq P(X|\delta(x)) \leq \alpha \text{ 时, 那么 } x \in \text{BUN}(X);$$

$$3) \text{ 当 } P(X|\delta(x)) < \beta \text{ 时, } x \in \text{NEG}(X)。$$

因此基于上述决策判别条件, 我们可以进一步得到邻域决策粗糙集的上下近似集的定义。

**定义 3**<sup>[10]</sup> 设邻域型信息系统  $IS = (U, AT)$ , 邻域半径为  $\delta$ , 属性集  $A \subseteq AT$  确定的邻域关系为  $N_A^\delta$ , 为了简便, 这里简单表示为  $N_A$ 。那么对象集  $X \subseteq U$  关于邻域关系  $N_A$  的  $(\alpha, \beta)$  邻域决策粗糙下近似和上近似分别定义为

$$\underline{N}_A^{(\alpha, \beta)}(X) = \{x \in U | P(X|\delta_A(x)) > \alpha\}$$

$$\overline{N}_A^{(\alpha, \beta)}(X) = \{x \in U | P(X|\delta_A(x)) \geq \beta\}$$

式中:  $0 \leq \beta < \alpha \leq 1$ ,  $P(X|\delta_A(x))$  表示对象  $x$  的邻域关于  $X$  的概率, 定义为

$$P(X|\delta_A(x)) = \frac{|\delta_A(x) \cap X|}{|\delta_A(x)|}$$

## 2 邻域决策粗糙集模型的增量式更新

由于现实环境下数据集的动态性, 传统的模型和算法不再有效, 针对该问题, 本节将提出一种论域变化时邻域决策粗糙集模型的增量式更新方法。

定义 3 已经表明, 当信息系统的决策代价已经确定时, 则邻域决策粗糙集中的阈值  $\alpha$  和  $\beta$  也就确定, 因此对于邻域决策粗糙集的研究只需考虑概率  $P(X|\delta(x))$  与阈值  $\alpha$  和  $\beta$  的关系即可。

数据集论域中对象的增加和减少往往都是批

量的, 而这些批量的变化可以分解成对象的一个依次变化, 每次只考虑数据集中一个对象增加或减少时的增量式更新问题, 然后逐步对多个对象进行迭代, 这样可以简化问题的处理<sup>[12, 17-18]</sup>。根据这一思想, 构造了本文模型的增量式更新方法。

### 2.1 论域中对象增加时模型的增量式更新

在邻域决策粗糙集中, 研究增量式更新的关键是上下近似区域的更新问题, 由于阈值  $\alpha$  和  $\beta$  已经确定, 因此主要涉及到对象与对象集之间的概率计算, 首先讨论论域对象增加时概率的增量式更新。

**定理 1**<sup>[18]</sup> 设邻域型信息系统  $IS = (U, AT)$ , 邻域半径为  $\delta$  且属性集  $A \subseteq AT$ 。当信息系统论域  $U$  增加对象  $x^+$ , 新的邻域型信息系统表示为  $IS^+ = (U^+ = U \cup \{x^+\}, AT)$ 。令  $\delta_A^U(x)$  表示对象  $\forall x \in U^+$  在论域  $U$  下的邻域类, 令  $\delta_A^{U^+}(x)$  表示对象  $\forall x \in U^+$  在论域  $U^+$  下的邻域类。对于  $X, Y \subseteq U$ , 若  $X^+, Y^+ \subseteq U^+$  且  $X^+ = X \cup \{x^+\}$ ,  $Y^+ = Y$ 。那么满足:

$$1) \forall x \in \delta_A^{U^+}(x^+) - \{x^+\}, \text{ 有}$$

$$P(X^+|\delta_A^{U^+}(x)) \geq P(X|\delta_A^U(x));$$

$$P(Y^+|\delta_A^{U^+}(x)) < P(Y|\delta_A^U(x)).$$

$$2) \forall x \in U^+ - \delta_A^{U^+}(x^+), \text{ 有}$$

$$P(X^+|\delta_A^{U^+}(x)) = P(X|\delta_A^U(x));$$

$$P(Y^+|\delta_A^{U^+}(x)) = P(Y|\delta_A^U(x)).$$

定理 1 给出信息系统论域  $U$  增加一个对象后, 论域对象在近似集下的概率变化关系, 有了这种变化作为基础, 接下来可以很容易得到邻域决策粗糙集上下近似的增量式更新。

**定理 2** 设邻域型信息系统  $IS = (U, AT)$ , 邻域半径为  $\delta$ 。对于  $X, Y \subseteq U$  关于属性集  $A \subseteq AT$  的邻域决策粗糙下近似分别为  $\underline{N}_A^{(\alpha, \beta)}(X)$  和  $\underline{N}_A^{(\alpha, \beta)}(Y)$ 。当信息系统论域  $U$  增加对象  $x^+$ , 新的邻域型信息系统表示为  $IS^+ = (U^+ = U \cup \{x^+\}, AT)$ 。若  $X^+, Y^+ \subseteq U^+$  且  $X^+ = X \cup \{x^+\}$ ,  $Y^+ = Y$ , 那么  $X^+, Y^+$  关于属性集  $A \subseteq AT$  的邻域决策粗糙下近似增量式更新为

$$1) \underline{N}_A^{(\alpha, \beta)}(X^+) = \underline{N}_A^{(\alpha, \beta)}(X) \cup$$

$$\{x \in \delta_A^{U^+}(x^+) - \underline{N}_A^{(\alpha, \beta)}(X) | P(X^+|\delta_A^{U^+}(x)) > \alpha\}$$

$$2) \underline{N}_A^{(\alpha, \beta)}(Y^+) = \underline{N}_A^{(\alpha, \beta)}(Y) -$$

$$\{x \in \underline{N}_A^{(\alpha, \beta)}(Y) \cap \delta_A^{U^+}(x^+) | P(Y^+|\delta_A^{U^+}(x)) \leq \alpha\} \cup$$

$$\{x \in \{x^+\} | P(Y^+|\delta_A^{U^+}(x)) > \alpha\}$$

**证明** 1) 对于  $\forall x \in \underline{N}_A^{(\alpha, \beta)}(X)$ , 要么  $x \in \delta_A^{U^+}(x^+) - \{x^+\}$ , 要么  $x \in U^+ - \delta_A^{U^+}(x^+)$ 。对于  $\forall x \in \underline{N}_A^{(\alpha, \beta)}(X)$  且满足  $x \in \delta_A^{U^+}(x^+) - \{x^+\}$ , 那么由定理 1 可以得到  $P(X^+|\delta_A^{U^+}(x)) \geq P(X|\delta_A^U(x)) > \alpha$ , 则有  $x \in \underline{N}_A^{(\alpha, \beta)}(X^+)$ 。

对于  $\forall x \in \underline{N}_A^{(\alpha, \beta)}(X)$  且满足  $x \in U^+ - \delta_A^{U^+}(x^+)$ , 那么

由定理1可以得到

$$P(X^+|\delta_A^{U^+}(x)) = P(X|\delta_A^U(x)) > \alpha$$

同样有  $x \in \underline{N}_A^{(\alpha, \beta)}(X^+)$ 。所以对于  $\forall x \in \underline{N}_A^{(\alpha, \beta)}(X)$  都有  $x \in \underline{N}_A^{(\alpha, \beta)}(X^+)$ , 即  $\underline{N}_A^{(\alpha, \beta)}(X) \subseteq \underline{N}_A^{(\alpha, \beta)}(X^+)$ 。

对于  $\forall x \in U - \underline{N}_A^{(\alpha, \beta)}(X)$ , 若  $x \in U^+ - \delta_A^{U^+}(x^+)$ , 由定理1可得  $P(X^+|\delta_A^{U^+}(x)) = P(X|\delta_A^U(x)) \leq \alpha$ 。那么此时  $x \notin \underline{N}_A^{(\alpha, \beta)}(X^+)$ 。

若  $x \in \delta_A^{U^+}(x^+) - \{x^+\}$ , 根据定理1可得

$$P(X^+|\delta_A^{U^+}(x)) \geq P(X|\delta_A^U(x))$$

而  $P(X|\delta_A^U(x)) \leq \alpha$ , 因此  $P(X^+|\delta_A^{U^+}(x))$  与  $\alpha$  的大小无法确定, 只有当  $P(X^+|\delta_A^{U^+}(x)) > \alpha$  时  $x \in \underline{N}_A^{(\alpha, \beta)}(X^+)$ 。也就是说, 只需对  $x \in \delta_A^{U^+}(x^+) - \underline{N}_A^{(\alpha, \beta)}(X)$  中的对象进行具体的计算便可以完成最终的下近似更新, 因此

$$\begin{aligned} \underline{N}_A^{(\alpha, \beta)}(X^+) &= \underline{N}_A^{(\alpha, \beta)}(X) \cup \\ &\{x \in \delta_A^{U^+}(x^+) - \underline{N}_A^{(\alpha, \beta)}(X) | P(X^+|\delta_A^{U^+}(x)) > \alpha\} \end{aligned}$$

即1)成立。

2) 对于  $\forall x \in U - \underline{N}_A^{(\alpha, \beta)}(Y)$ , 都有  $P(Y|\delta_A^U(x)) \leq \alpha$ 。那么根据定理1, 当  $x \in \delta_A^{U^+}(x^+) - \{x^+\}$  时, 有

$$P(Y^+|\delta_A^{U^+}(x)) < P(Y|\delta_A^U(x)) \leq \alpha$$

当  $x \in U^+ - \delta_A^{U^+}(x^+)$  时, 有

$$P(Y^+|\delta_A^{U^+}(x)) = P(Y|\delta_A^U(x)) \leq \alpha$$

因此  $\forall x \in U - \underline{N}_A^{(\alpha, \beta)}(Y)$  都满足  $P(Y^+|\delta_A^{U^+}(x)) \leq \alpha$ , 即  $x \notin \underline{N}_A^{(\alpha, \beta)}(Y^+)$ 。

对于  $\forall x \in \underline{N}_A^{(\alpha, \beta)}(Y)$ , 都有  $P(Y|\delta_A^U(x)) > \alpha$ 。根据定理1, 当  $x \in \delta_A^{U^+}(x^+) - \{x^+\}$  时, 有

$$P(Y^+|\delta_A^{U^+}(x)) < P(Y|\delta_A^U(x)), P(Y|\delta_A^U(x)) > \alpha。$$

因此  $P(Y^+|\delta_A^{U^+}(x))$  与  $\alpha$  的大小无法确定。

当  $x \in U^+ - \delta_A^{U^+}(x^+)$  时, 有

$$P(Y^+|\delta_A^{U^+}(x)) = P(Y|\delta_A^U(x)) > \alpha$$

即  $x \in \underline{N}_A^{(\alpha, \beta)}(Y^+)$ , 则需要对  $x \in \delta_A^{U^+}(x^+) - \{x^+\}$  中的对象计算  $P(Y^+|\delta_A^{U^+}(x))$  可得到新的下近似结果, 即

$$\begin{aligned} \underline{N}_A^{(\alpha, \beta)}(Y^+) &= (\underline{N}_A^{(\alpha, \beta)}(Y) - \\ &\{x \in \underline{N}_A^{(\alpha, \beta)}(Y) \cap \delta_A^{U^+}(x^+) | P(Y^+|\delta_A^{U^+}(x)) \leq \alpha\}) \cup \\ &\{x \in \{x^+\} | P(Y^+|\delta_A^{U^+}(x)) > \alpha\} \end{aligned}$$

定理2证明完毕。

定理2表明, 当信息系统增加一个对象后, 新的邻域决策粗糙下近似集变化呈现一定的规律, 例如从  $\underline{N}_A^{(\alpha, \beta)}(X)$  更新至  $\underline{N}_A^{(\alpha, \beta)}(X^+)$ , 其近似集是增大的, 并且增加的对象集只来源于对象集  $\delta_A^{U^+}(x^+) - \underline{N}_A^{(\alpha, \beta)}(X)$ , 因此只需对这里面的对象进行概率  $P(X^+|\delta_A^{U^+}(x))$  计算, 便可得到最终的更新结果, 这样大幅度减小了计算量。同样地对于  $\underline{N}_A^{(\alpha, \beta)}(Y)$  更新至  $\underline{N}_A^{(\alpha, \beta)}(Y^+)$ , 只需对  $\underline{N}_A^{(\alpha, \beta)}(Y) \cap \delta_A^{U^+}(x^+)$  中的对象进行相关计算便可以完成最终更新, 所以定理2所

示的下近似集增量式更新具有较高的计算效率。

类似于定理2的研究方法, 接下来可以进一步得到上近似集的增量式更新。

**定理3** 设邻域型信息系统  $IS = (U, AT)$ , 邻域半径为  $\delta$ 。对于  $X, Y \subseteq U$  关于属性集  $A \subseteq AT$  的邻域决策粗糙上近似分别为  $\overline{N}_A^{(\alpha, \beta)}(X)$  和  $\overline{N}_A^{(\alpha, \beta)}(Y)$ 。当信息系统论域  $U$  增加对象  $x^+$ , 新的邻域型信息系统表示为  $IS^+ = (U^+ = U \cup \{x^+\}, AT)$ 。若  $X^+, Y^+ \subseteq U^+$  且  $X^+ = X \cup \{x^+\}$ ,  $Y^+ = Y$ , 那么  $X^+, Y^+$  关于属性集  $A \subseteq AT$  的邻域决策粗糙上近似增量式更新为

$$\begin{aligned} 1) \overline{N}_A^{(\alpha, \beta)}(X^+) &= \overline{N}_A^{(\alpha, \beta)}(X) \cup \\ &\{x \in \delta_A^{U^+}(x^+) - \overline{N}_A^{(\alpha, \beta)}(X) | P(X^+|\delta_A^{U^+}(x)) \geq \beta\} \\ 2) \overline{N}_A^{(\alpha, \beta)}(Y^+) &= (\overline{N}_A^{(\alpha, \beta)}(Y) - \\ &\{x \in \overline{N}_A^{(\alpha, \beta)}(Y) \cap \delta_A^{U^+}(x^+) | P(Y^+|\delta_A^{U^+}(x)) < \beta\}) \cup \\ &\{x \in \{x^+\} | P(Y^+|\delta_A^{U^+}(x)) \geq \beta\} \end{aligned}$$

**证明** 证明过程同定理2。

定理3类似于定理2, 更新上近似集时也只需要针对部分对象集进行计算便可完成, 例如从  $\overline{N}_A^{(\alpha, \beta)}(X)$  更新至  $\overline{N}_A^{(\alpha, \beta)}(X^+)$ , 只需计算  $\delta_A^{U^+}(x^+) - \overline{N}_A^{(\alpha, \beta)}(X)$  内的对象, 从  $\overline{N}_A^{(\alpha, \beta)}(Y)$  更新至  $\overline{N}_A^{(\alpha, \beta)}(Y^+)$ , 只需计算  $\overline{N}_A^{(\alpha, \beta)}(Y) \cap \delta_A^{U^+}(x^+)$  中的对象, 因此定理3同样具有很高的更新计算效率。

定理2和定理3分别给出当邻域信息系统增加一个对象时, 邻域决策粗糙上下近似的增量式更新问题, 当信息系统通过增加多个对象, 可以根据定理2和定理3逐步进行迭代, 直至完成最终的更新。

## 2.2 论域中对象减少时模型的增量式更新

类似于第2.1节中研究方法, 本节将进一步提出论域中对象减少时模型的增量式更新, 首先分析对象减少时概率的变化关系。

**定理4**<sup>[18]</sup> 设邻域型信息系统  $IS = (U, AT)$ , 邻域半径为  $\delta$  且属性集  $A \subseteq AT$ 。当信息系统论域  $U$  移除对象  $x^-$ ,  $x^- \in U$ , 新的邻域型信息系统表示为  $IS^- = (U^- = U - \{x^-\}, AT)$ 。令  $\delta_A^U(x)$  表示对象  $\forall x \in U$  在论域  $U$  下的邻域类, 令  $\delta_A^{U^-}(x)$  表示对象  $\forall x \in U$  在论域  $U^-$  下的邻域类。对于  $X, Y \subseteq U$ , 若  $X^-, Y^- \subseteq U^-$  且  $X^- = X - \{x^-\}$ ,  $Y^- = Y$ 。那么满足

$$\begin{aligned} 1) \forall x \in \delta_A^U(x^-) - \{x^-\}, 有 \\ P(X^-|\delta_A^{U^-}(x)) &\leq P(X|\delta_A^U(x)) \\ P(Y^-|\delta_A^{U^-}(x)) &> P(Y|\delta_A^U(x)) \\ 2) \forall x \in U - \delta_A^U(x^-), 有 \\ P(X^-|\delta_A^{U^-}(x)) &= P(X|\delta_A^U(x)) \\ P(Y^-|\delta_A^{U^-}(x)) &= P(Y|\delta_A^U(x)) \end{aligned}$$

定理4给出信息系统论域  $U$  移除一个对象

后, 近似集在不同对象下的概率变化关系, 有了这种变化作为基础, 接下来可以很容易得到邻域决策粗糙集上下近似的增量式更新。

**定理 5** 设邻域型信息系统  $IS = (U, AT)$ , 邻域半径为  $\delta$ 。对于  $X, Y \subseteq U$  关于属性集  $A \subseteq AT$  的邻域决策粗糙集下近似分别为  $\underline{N}_A^{(\alpha, \beta)}(X)$  和  $\underline{N}_A^{(\alpha, \beta)}(Y)$ 。当信息系统论域  $U$  移除对象  $x^-$  且  $x^- \in U$ , 新的邻域型信息系统表示为  $IS^- = (U^- = U - \{x^-\}, AT)$ 。若  $X^-, Y^- \subseteq U^-$  且  $X^- = X - \{x^-\}, Y^- = Y$ , 那么  $X^-, Y^-$  关于属性集  $A \subseteq AT$  的邻域决策粗糙集下近似增量式更新为

$$\begin{aligned} 1) \underline{N}_A^{(\alpha, \beta)}(X^-) &= \underline{N}_A^{(\alpha, \beta)}(X) - \{x \in \underline{N}_A^{(\alpha, \beta)}(X) \cap \\ &\quad (\delta_A^U(x^-) - \{x^-\}) | P(X^- | \delta_A^U(x^-)) \leq \alpha\} \\ 2) \underline{N}_A^{(\alpha, \beta)}(Y^-) &= \underline{N}_A^{(\alpha, \beta)}(Y) \cup \\ &\quad \{x \in \delta_A^U(x^-) - \{x^-\} - \underline{N}_A^{(\alpha, \beta)}(Y) | P(Y^- | \delta_A^U(x^-)) > \alpha\} \end{aligned}$$

**证明** 1) 对于  $\forall x \in \underline{N}_A^{(\alpha, \beta)}(X)$ , 要么  $x \in \delta_A^U(x^-)$ , 要么  $x \in U - \delta_A^U(x^-)$ 。若对象满足  $x \in \delta_A^U(x^-) - \{x^-\}$ , 那么由定理 4 可以得到

$$P(X^- | \delta_A^U(x^-)) \leq P(X | \delta_A^U(x))$$

由于  $P(X | \delta_A^U(x)) > \alpha$ , 则无法直接确定  $P(X^- | \delta_A^U(x^-))$  与  $\alpha$  的关系。即  $\forall x \in \underline{N}_A^{(\alpha, \beta)}(X)$  且  $x \in \delta_A^U(x^-) - \{x^-\}$ , 需要计算  $P(X^- | \delta_A^U(x^-))$  的值。

若对象满足  $x \in U - \delta_A^U(x^-)$ , 那么由定理 4 可以得到  $P(X^- | \delta_A^U(x^-)) = P(X | \delta_A^U(x)) > \alpha$ , 因此  $\forall x \in \underline{N}_A^{(\alpha, \beta)}(X)$  且  $x \in U - \delta_A^U(x^-)$  都有  $x \in \underline{N}_A^{(\alpha, \beta)}(X^-)$ 。

对于  $\forall x \in U - \underline{N}_A^{(\alpha, \beta)}(X)$ , 若对象满足  $x \in \delta_A^U(x^-) - \{x^-\}$ , 那么由定理 4 可以得到

$$P(X^- | \delta_A^U(x^-)) \leq P(X | \delta_A^U(x)) \leq \alpha.$$

因此  $\forall x \in U - \underline{N}_A^{(\alpha, \beta)}(X)$  且  $x \in \delta_A^U(x^-) - \{x^-\}$  都有  $x \notin \underline{N}_A^{(\alpha, \beta)}(X^-)$ 。

若对象满足  $x \in U - \delta_A^U(x^-)$ , 那么由定理 4 可以得到  $P(X^- | \delta_A^U(x^-)) = P(X | \delta_A^U(x)) \leq \alpha$ 。因此  $\forall x \in U - \underline{N}_A^{(\alpha, \beta)}(X)$  且  $x \in U - \delta_A^U(x^-)$  都有  $x \notin \underline{N}_A^{(\alpha, \beta)}(X^-)$ 。综合起来  $\forall x \in U - \underline{N}_A^{(\alpha, \beta)}(X)$  都满足  $x \notin \underline{N}_A^{(\alpha, \beta)}(X^-)$ , 则 (1) 成立。

2) 对于  $\forall x \in \underline{N}_A^{(\alpha, \beta)}(Y)$ , 若对象满足  $x \in \delta_A^U(x^-) - \{x^-\}$ , 那么由定理 4 可以得到

$$P(Y^- | \delta_A^U(x^-)) > P(Y | \delta_A^U(x)) > \alpha$$

若对象满足  $x \in U - \delta_A^U(x^-)$ , 那么由定理 4 可以得到

$$P(Y^- | \delta_A^U(x^-)) = P(Y | \delta_A^U(x)) > \alpha$$

因此  $\forall x \in \underline{N}_A^{(\alpha, \beta)}(Y)$  都有  $x \in \underline{N}_A^{(\alpha, \beta)}(Y^-)$ 。

对于  $\forall x \in U - \underline{N}_A^{(\alpha, \beta)}(Y)$ , 若对象满足  $x \in \delta_A^U(x^-) - \{x^-\}$ , 那么由定理 4 可以得到

$$P(Y^- | \delta_A^U(x^-)) > P(Y | \delta_A^U(x)), P(Y | \delta_A^U(x)) \leq \alpha$$

此时不能直接确定  $P(Y^- | \delta_A^U(x^-))$  与  $\alpha$  之间的关系。

若对象  $x \in U - \delta_A^U(x^-)$ , 那么由定理 4 可得到

$$P(Y^- | \delta_A^U(x^-)) = P(Y | \delta_A^U(x)) \leq \alpha$$

即  $x \notin \underline{N}_A^{(\alpha, \beta)}(Y^-)$ 。

因此只需要计算对象集  $(U - \underline{N}_A^{(\alpha, \beta)}(Y)) \cap (\delta_A^U(x^-) - \{x^-\})$  中的对象便可以完成近似集的增量式更新, 即计算  $x \in \delta_A^U(x^-) - \{x^-\} - \underline{N}_A^{(\alpha, \beta)}(Y)$ , 因此 2) 成立。

定理 5 证明完毕。

定理 5 表明, 当邻域型信息系统移除一个对象  $x^-$  后, 新的邻域决策粗糙集下近似集变化也呈现一定的规律, 例如从  $\underline{N}_A^{(\alpha, \beta)}(X)$  更新至  $\underline{N}_A^{(\alpha, \beta)}(X^-)$ , 其近似集是减小的, 并且减少的对象只来源于一小部分对象集, 因此只需对这里面的对象进行概率  $P(X^- | \delta_A^U(x^-))$  的计算, 便可得到最终的更新结果。同样地对于  $\underline{N}_A^{(\alpha, \beta)}(Y)$  更新至  $\underline{N}_A^{(\alpha, \beta)}(Y^-)$ , 也只需对一小部分对象进行相关计算便可以完成最终更新。所以定理 5 所示的下近似集增量式更新方法具有较高的计算效率。类似于定理 5 的研究方法, 接下来可以进一步得到上近似集的增量式更新。

**定理 6** 设邻域型信息系统  $IS = (U, AT)$ , 邻域半径为  $\delta$ 。对于  $X, Y \subseteq U$  关于属性集  $A \subseteq AT$  的邻域决策粗糙集上近似分别为  $\overline{N}_A^{(\alpha, \beta)}(X)$  和  $\overline{N}_A^{(\alpha, \beta)}(Y)$ 。当信息系统论域  $U$  移除一个对象  $x^-$ ,  $x^- \in U$ , 新的邻域型信息系统表示为  $IS^- = (U^- = U - \{x^-\}, AT)$ 。若  $X^-, Y^- \subseteq U^-$  且  $X^- = X - \{x^-\}, Y^- = Y$ , 那么  $X^-, Y^-$  关于属性集  $A \subseteq AT$  的邻域决策粗糙集上近似增量式更新为

$$\begin{aligned} 1) \overline{N}_A^{(\alpha, \beta)}(X^-) &= \overline{N}_A^{(\alpha, \beta)}(X) - \{x \in \overline{N}_A^{(\alpha, \beta)}(X) \cap \\ &\quad (\delta_A^U(x^-) - \{x^-\}) | P(X^- | \delta_A^U(x^-)) < \beta\} \\ 2) \overline{N}_A^{(\alpha, \beta)}(Y^-) &= \overline{N}_A^{(\alpha, \beta)}(Y) \cup \\ &\quad \{x \in \delta_A^U(x^-) - \{x^-\} - \overline{N}_A^{(\alpha, \beta)}(Y) | P(Y^- | \delta_A^U(x^-)) \geq \beta\} \end{aligned}$$

**证明** 1) 根据定理 5 的证明结果, 对于  $\forall x \in U - \overline{N}_A^{(\alpha, \beta)}(X)$  都满足  $x \notin \overline{N}_A^{(\alpha, \beta)}(X^-)$ 。而对于  $\forall x \in \overline{N}_A^{(\alpha, \beta)}(X)$  且  $x \in U - \delta_A^U(x^-)$  都有  $x \in \overline{N}_A^{(\alpha, \beta)}(X^-)$ , 只有当  $\forall x \in \overline{N}_A^{(\alpha, \beta)}(X)$  且  $x \in \delta_A^U(x^-) - \{x^-\}$  时,  $P(X^- | \delta_A^U(x^-))$  与  $\alpha$  之间的大小关系不能直接确定, 因此定理 6 中的 1) 得到证明。

2) 根据定理 5 的证明结果, 对于  $\forall x \in \overline{N}_A^{(\alpha, \beta)}(Y)$  都有  $x \in \overline{N}_A^{(\alpha, \beta)}(Y^-)$ 。对于  $\forall x \in U - \overline{N}_A^{(\alpha, \beta)}(Y)$  且  $x \in U - \delta_A^U(x^-)$  时, 都有  $x \notin \overline{N}_A^{(\alpha, \beta)}(Y^-)$ , 只有当  $\forall x \in U - \overline{N}_A^{(\alpha, \beta)}(Y)$  且  $x \in \delta_A^U(x^-) - \{x^-\}$  时, 不能直接确定  $P(Y^- | \delta_A^U(x^-))$  与  $\alpha$  之间的大小关系, 因此定理 6 中的 2) 得到证明。

定理 6 与定理 5 类似, 只需要在原先上近似集  $\overline{N}_A^{(\alpha, \beta)}(X)$  和  $\overline{N}_A^{(\alpha, \beta)}(Y)$  的结果上, 进一步对  $\delta_A^U(x^-)$

中的对象计算概率便可完成最终的更新, 因此上近似集的增量式更新同样具有很高的计算效率。

定理5和定理6分别给出当邻域型信息系统移除一个对象时上下近似集的增量式更新问题, 当信息系统同时移除多个对象时, 可以根据定理5和定理6逐步进行迭代, 直至完成最终的更新。

### 3 邻域决策粗糙集更新算法

根据本文所提出的增量式更新方法, 接下来将进一步提出对应的邻域决策粗糙集增量式更新算法, 具体如算法1和算法2所示。

**算法1** 论域增加时邻域决策粗糙集的增量式更新算法

**输入** 1) 邻域型信息系统  $IS = (U, AT)$ , 邻域半径为  $\delta$ , 阈值  $(\alpha, \beta)$ 。近似对象集  $X$  关于属性子集  $A \subseteq AT$  的近似集  $\underline{N}_A^{(\alpha, \beta)}(X)$  和  $\overline{N}_A^{(\alpha, \beta)}(X)$ 。

2) 信息系统增加对象集  $\Delta U^+ = \{x_1^+, x_2^+, \dots, x_s^+\}$ , 新的邻域型信息系统为  $IS^+ = (U^+ = U \cup \Delta U^+, AT)$ , 新的近似对象集为  $X^+$ 。

**输出** 新的近似集  $\underline{N}_A^{(\alpha, \beta)}(X^+)$  和  $\overline{N}_A^{(\alpha, \beta)}(X^+)$ 。

1) 将初始时的信息系统  $IS = (U, AT)$  记为  $IS^{(0)} = (U^{(0)}, AT)$ ,  $X$  记为  $X^{(0)}$ , 近似集  $\underline{N}_A^{(\alpha, \beta)}(X)$  和  $\overline{N}_A^{(\alpha, \beta)}(X)$  分别记为  $\underline{N}_A^{(\alpha, \beta)}(X^{(0)})$  和  $\overline{N}_A^{(\alpha, \beta)}(X^{(0)})$ 。

2) 对于  $x_i^+ \in \Delta U^+ (1 \leq i \leq s)$ , 将  $x_i^+$  加入邻域型信息系统  $IS^{(i-1)} = (U^{(i-1)}, AT)$  中, 新的信息系统表示为  $IS^{(i)} = (U^{(i)}, AT)$ , 其中  $U^{(i)} = U^{(i-1)} \cup \{x_i^+\}$ , 此时新的近似对象集为  $X^{(i)}$ 。

3) 计算对象  $x_i^+$  的邻域类  $\delta_A^{U^{(i)}}(x_i^+)$ , 然后根据定理2和定理3在  $\underline{N}_A^{(\alpha, \beta)}(X^{(i-1)})$  和  $\overline{N}_A^{(\alpha, \beta)}(X^{(i-1)})$  的基础上增量式计算  $\underline{N}_A^{(\alpha, \beta)}(X^{(i)})$  和  $\overline{N}_A^{(\alpha, \beta)}(X^{(i)})$ 。

4) 重复迭代计算步骤2)~3)。

5) 令

$$\begin{aligned}\underline{N}_A^{(\alpha, \beta)}(X^+) &= \underline{N}_A^{(\alpha, \beta)}(X^{(s)}); \\ \overline{N}_A^{(\alpha, \beta)}(X^+) &= \overline{N}_A^{(\alpha, \beta)}(X^{(s)}).\end{aligned}$$

返回  $\underline{N}_A^{(\alpha, \beta)}(X^+)$  和  $\overline{N}_A^{(\alpha, \beta)}(X^+)$ 。

**算法2** 论域减少时邻域决策粗糙集的增量式更新算法

**输入** 1) 邻域型信息系统  $IS = (U, AT)$ , 邻域半径为  $\delta$ , 阈值  $(\alpha, \beta)$ 。近似对象集  $X$  关于属性子集  $A \subseteq AT$  的近似集  $\underline{N}_A^{(\alpha, \beta)}(X)$  和  $\overline{N}_A^{(\alpha, \beta)}(X)$ 。

2) 信息系统减少对象集  $\Delta U^- = \{x_1^-, x_2^-, \dots, x_t^-\}$ , 新的邻域型信息系统为  $IS^- = (U^- = U - \Delta U^-, AT)$ , 新的近似对象集为  $X^-$ 。

**输出** 新的近似集  $\underline{N}_A^{(\alpha, \beta)}(X^-)$  和  $\overline{N}_A^{(\alpha, \beta)}(X^-)$ 。

1) 将初始时的信息系统  $IS = (U, AT)$  记为

$IS^{(0)} = (U^{(0)}, AT)$ ,  $X$  记为  $X^{(0)}$ , 近似集  $\underline{N}_A^{(\alpha, \beta)}(X)$  和  $\overline{N}_A^{(\alpha, \beta)}(X)$  分别记为  $\underline{N}_A^{(\alpha, \beta)}(X^{(0)})$  和  $\overline{N}_A^{(\alpha, \beta)}(X^{(0)})$ 。

2) 对于  $x_i^- \in \Delta U^- (1 \leq i \leq t)$ , 将  $x_i^-$  从邻域型信息系统  $IS^{(i-1)} = (U^{(i-1)}, AT)$  中移除, 新的信息系统表示为  $IS^{(i)} = (U^{(i)}, AT)$ , 其中  $U^{(i)} = U^{(i-1)} - \{x_i^-\}$ , 此时新的近似对象集为  $X^{(i)}$ 。

3) 计算对象  $x_i^-$  的邻域类  $\delta_A^{U^{(i)}}(x_i^-)$ , 然后根据定理5和定理6在  $\underline{N}_A^{(\alpha, \beta)}(X^{(i-1)})$  和  $\overline{N}_A^{(\alpha, \beta)}(X^{(i-1)})$  的基础上增量式计算  $\underline{N}_A^{(\alpha, \beta)}(X^{(i)})$  和  $\overline{N}_A^{(\alpha, \beta)}(X^{(i)})$ 。

4) 重复迭代计算步骤2)~3)

5) 令

$$\begin{aligned}\underline{N}_A^{(\alpha, \beta)}(X^-) &= \underline{N}_A^{(\alpha, \beta)}(X^{(t)}); \\ \overline{N}_A^{(\alpha, \beta)}(X^-) &= \overline{N}_A^{(\alpha, \beta)}(X^{(t)}).\end{aligned}$$

返回  $\underline{N}_A^{(\alpha, \beta)}(X^-)$  和  $\overline{N}_A^{(\alpha, \beta)}(X^-)$ 。

在算法1所示的增量式计算过程中, 每次在更新前信息系统的上下近似基础上进一步计算新的上下近似集, 并且定理2和定理3已经表明, 只需要计算新增对象的邻域类, 便可以完成最终的更新, 而不必去计算其他对象的邻域类, 这样大大减少了重复的计算量, 提高了更新的效率, 因此整个算法1和算法2的时间复杂度可表示为  $O(|A| \cdot |U| \cdot |\Delta U|)$ 。

### 4 实验分析

为了验证所提出增量式更新算法的有效性, 将通过实验比较的方式进行验证。本实验主要将文中所提出的增量式更新算法与传统的非增量更新算法对同一组数据集进行动态更新计算, 通过比较他们的动态更新效率来验证算法的有效性, 其中表1所示的是实验中所使用的数据集, 这些数据集均来源于UCI数据集库, 其中数据集的各个属性均为数值类型。整个实验所运行的硬件环境为Intel Core G4560 3.5 GHz处理器和DDR4 8 GB内存。

表1所示的均为静态的数据集, 为了模拟数据集动态变化的情形, 本实验采用其他学者常用的处理方法<sup>[11-12, 17-18]</sup>, 即让数据集按照论域平均分成多个对象集, 然后通过将这些对象集逐渐进行合并, 达到了数据集论域动态增加的效果, 将原始论域逐渐对这些对象集进行移除, 便达到了数据集论域动态的减少。本实验将论域平均分成9个部分, 这样可以构造出数据集的8次动态更新。实验中将数据集的决策类作为邻域决策粗糙集的近似对象集, 即计算数据集每个决策类的上下近似增量式更新。实验中每个数据集的属性值



均进行归一化处理,并且统一设定邻域半径 $\delta = 0.15$ , 阈值 $\alpha = 0.75, \beta = 0.55$ 。

表 1 实验数据集  
Table 1 Experimental data set

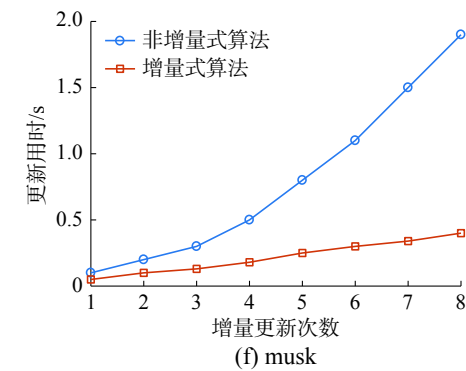
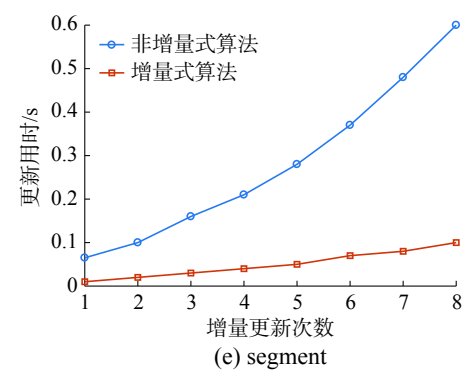
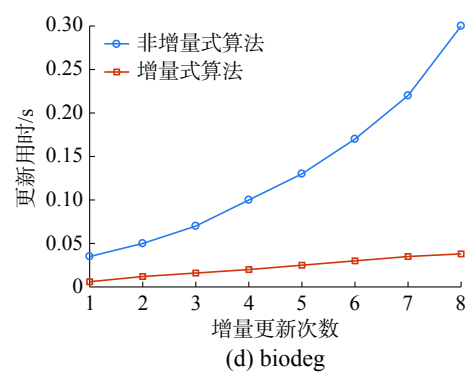
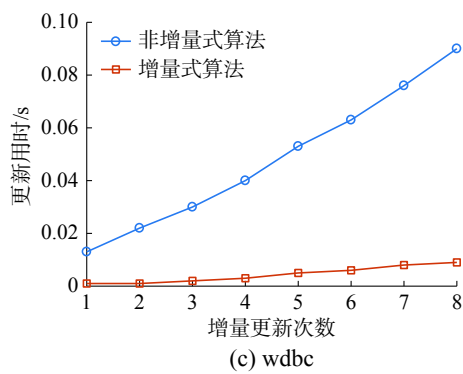
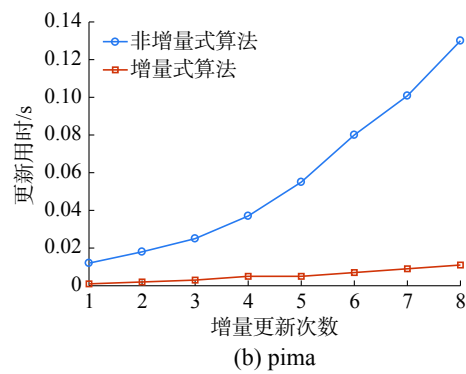
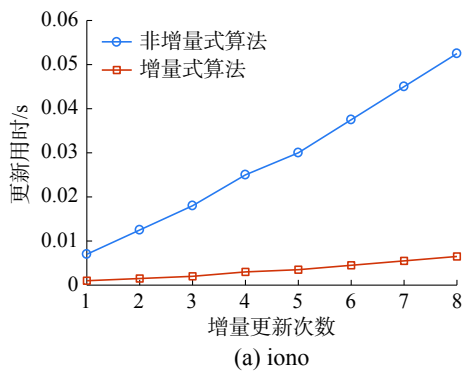
序号	数据集	对象	属性	类别
1	iono	351	34	2
2	pima	768	8	3
3	wdbc	569	31	2
4	biodeg	1055	41	2
5	segment	2310	19	7
6	musk	6598	166	2
7	census income	48 842	14	3
8	music	106 574	518	4

图 1 为各个数据集论域增加时增量式更新算法(算法 1)与非增量式更新算法计算邻域决策粗糙集的时间比较结果,非增量式更新算法采用文

献[10]提出的算法。

在图 1 所示的实验结果中,增量式算法的更新用时大幅度低于非增量式算法,并且随着数据集更新次数的增多,两类算法的差距不断增大。这主要是由于非增量式更新算法在进行模型的更新时,每次均基于完整的论域进行计算,产生的时间会越来越多。对于增量式更新算法,随着数据集论域的增大,更新所需的时间较少且增长的速率较为缓慢,这主要是由于增量式更新算法采用增量式的方法进行更新计算,每次均在前一次更新的结果上进行进一步更新,这样避免了原有对象的重复计算,大幅度提高更新效率,因此增量式算法更加高效。

图 2 展示的是各个数据集论域减少时,增量式更新算法(算法 2)与非增量式更新算法计算邻域决策粗糙集的时间比较结果,非增量式更新算法同样采用文献[10]提出的算法。





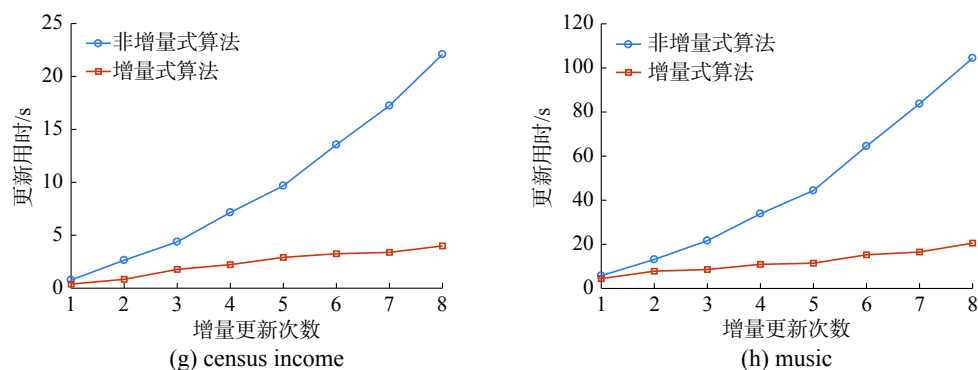
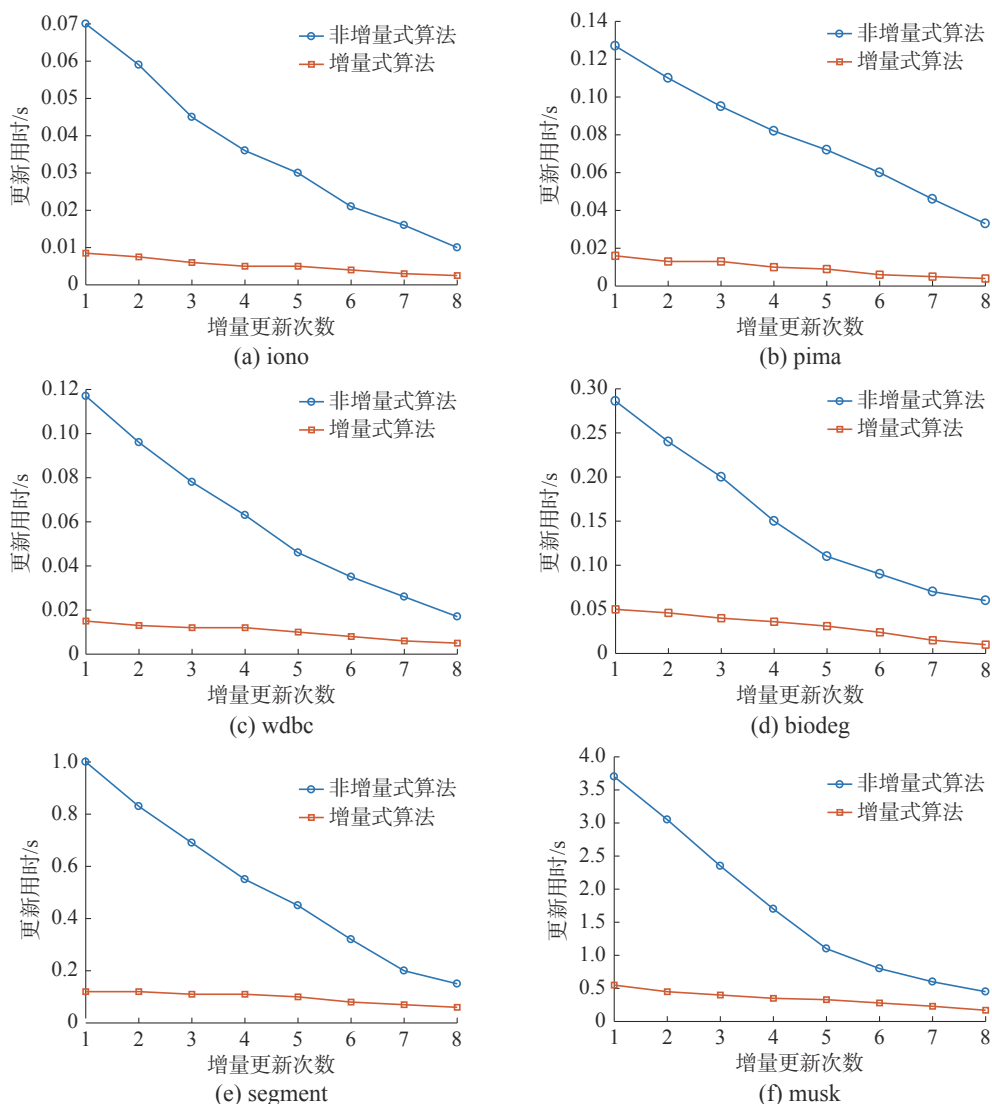


图1 论域增加时两类算法的更新用时比较

Fig. 1 Comparison of update time of two algorithms when universe is added

在图2所示的实验结果中,可以发现增量式更新算法的更新用时同样大幅度低于非增量式更新算法,对于非增量式更新算法,随着数据集论域的逐渐减少,其更新模型的用时也是逐渐减小,这主要是由于非增量式更新算法在进行更新时,对完整论域进行计算,因此随着论域的减少,非增量式算法的计算量也大幅度减小,产生的时间会越来越来少,但是整体还是高于增

量式算法。对于增量式更新算法,随着数据集论域的减小,其整体更新用时始终处于一个较低的水平,并且随着更新次数增加,更新用时也是逐渐减小的。这主要是由于增量式更新算法采用增量式的方法进行更新计算,在前一次更新结果的基础上计算后一次结果,由于论域逐渐减少,则更新时间会更加少,从而效率远高于非增量式算法。



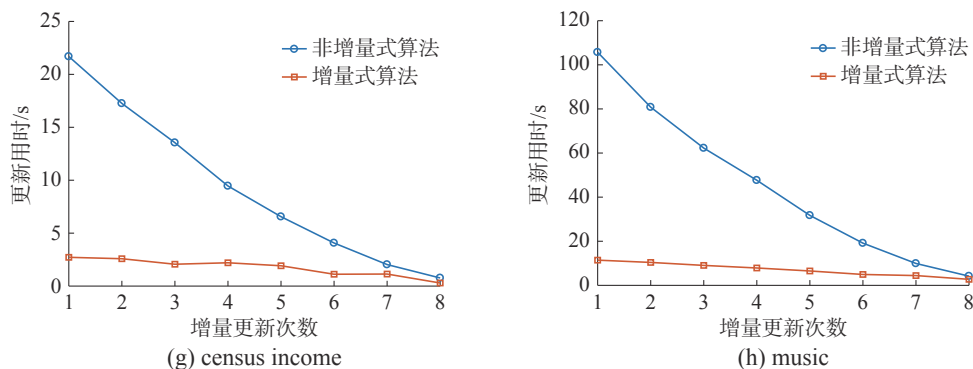


图2 论域减少时两类算法的更新用时比较

Fig. 2 Comparison of update time of two algorithms when universe is reduced

综合图1和图2的实验结果,可以看出本文所提出的增量式更新算法的更新效率均大幅度高于传统的非增量式算法,并且所提出的算法随数据集论域变化的影响较小,这说明了本文所提出的增量式更新算法具有很高的优越性。

另一方面,在本文所提出的邻域决策粗糙集增量式更新算法中,有3个重要的参数,分别为邻域半径 $\delta$ ,和一对阈值 $(\alpha, \beta)$ 。由于阈值 $(\alpha, \beta)$ 可以通过不同的评价方式进行确定,因此阈值 $(\alpha, \beta)$ 可认定为是固定的值,那么接下来将直接探究邻域半径 $\delta$ 对所提算法更新效率的影响。

图3~6所示的是部分数据集在不用邻域半

径 $\delta$ 下增量式更新用时比较结果,其中包含了论域增加和论域减少的两种情形,这里设定邻域半径 $\delta$ 在 $[0.1, 0.28]$ 内以0.02为间隔分别进行取值。

通过图3~6的结果可以看出,随着邻域半径的逐渐增大,增量式更新算法的更新用时是逐渐增大的,这主要是由于本文所提出的增量更新算法,计算新论域下的模型时,需要计算变化对象的邻域类,并基于这些邻域类进行更新计算,而邻域半径的增大无疑会增加邻域类中对象的数量,因此计算量会增加,从而展现出了图3~6的结果,但是对比图1和图2,这种增量式算法的用时仍然大幅度小于非增量式算法。

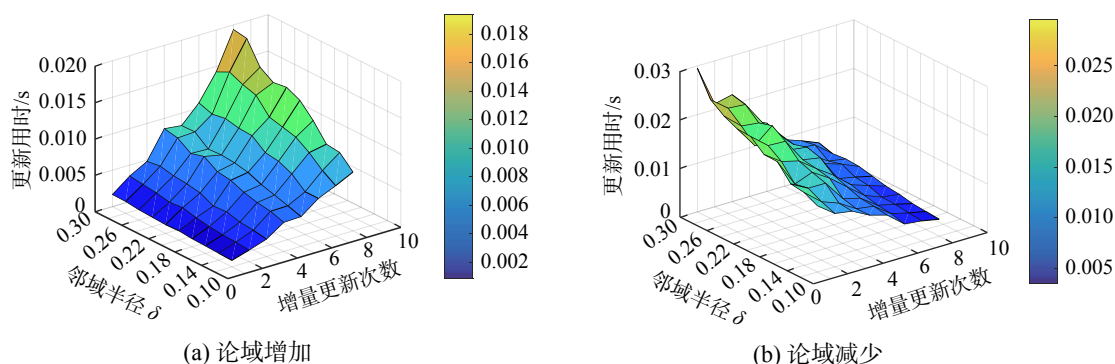


图3 数据集 pima 在不同邻域半径下算法更新用时比较

Fig. 3 Comparison of algorithm updating time of pima data set under different neighborhood radius

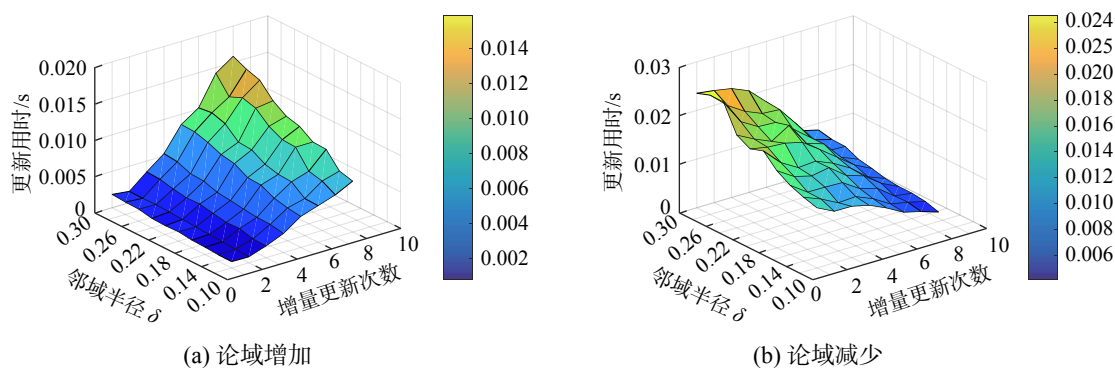


图4 数据集 wdbc 在不同邻域半径下算法更新用时比较

Fig. 4 Comparison of algorithm updating time of wdbc data set under different neighborhood radius

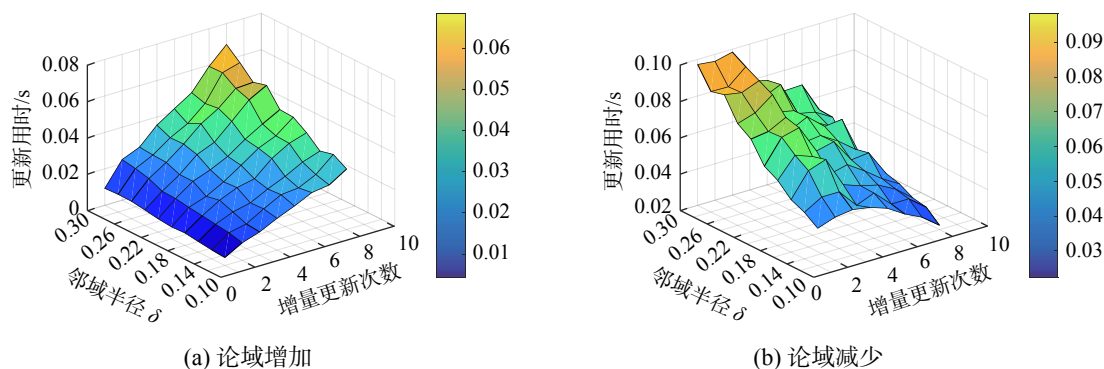


图5 数据集 biodeg 在不同邻域半径下算法更新用时比较

Fig. 5 Comparison of algorithm updating time of biodeg data set under different neighborhood radius

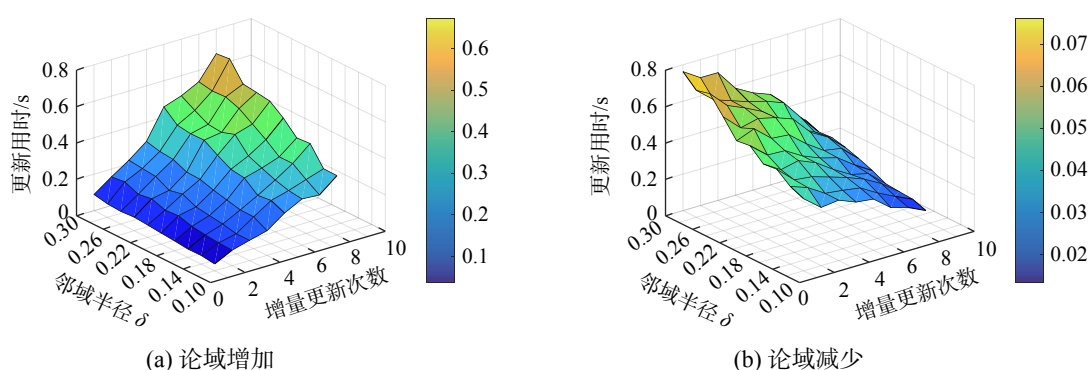


图6 数据集 musk 在不同邻域半径下算法更新用时比较

Fig. 6 Comparison of algorithm updating time of musk data set under different neighborhood radius

## 5 结束语

邻域决策粗糙集是传统决策粗糙集的重要拓展,针对现实环境下数据集的动态性,本文提出一种论域动态变化时的邻域决策粗糙集增量式更新算法。本文首先研究了论域中单个对象变化时,模型的增量式更新问题,然后以单个对象变化为基础,通过迭代方式完成对象批量变化时的增量式更新问题,实验分析表明,所提出的增量式算法在更新动态数据时,其效率大幅度高于非增量式算法,且增量式算法的更新时间受论域对象变化的影响较小,因此说明了所提出的增量式更新算法具有很高的优越性,从而也进一步推动了决策粗糙集在实际环境下的应用。在本文研究成果的基础上,接下来可以进一步在邻域决策粗糙集的增量式属性约简问题上进行探索。

## 参考文献:

- [1] YAO Yiyu. Three-way decisions with probabilistic rough sets[J]. *Information sciences*, 2010, 180(3): 341–353.
- [2] YAO Yiyu. The superiority of three-way decisions in probabilistic rough set models[J]. *Information sciences*, 2011, 181(6): 1080–1096.
- [3] 陈家俊, 徐华丽, 魏贇. 多重代价多粒度决策粗糙集模型研究[J]. *计算机科学与探索*, 2018, 12(5): 839–850.  
CHEN Jiajun, XU Huali, WEI Yun. Multi-cost based multi-granulation decision-theoretic rough set model[J]. *Journal of frontiers of computer science and technology*, 2018, 12(5): 839–850.
- [4] JIA Xiuyi, LI Weiwei, SHANG Lin. A multiphase cost-sensitive learning method based on the multiclass three-way decision-theoretic rough set model[J]. *Information sciences*, 2019, 485: 248–262.
- [5] 张婷, 张红云, 王真. 基于三支决策粗糙集的迭代量化的图像检索算法[J]. *南京大学学报(自然科学版)*, 2018, 54(4): 714–724.  
ZHANG Ting, ZHANG Hongyun, WANG Zhen. image retrieval: Iterative quantization based on saliency detection and three-way decision based rough sets[J]. *Journal of Nanjing University (natural sciences edition)*, 2018, 54(4): 714–724.
- [6] ZHAO Xuerong, HU Baoqing. Three-way decisions with decision-theoretic rough sets in multiset-valued information tables[J]. *Information sciences*, 2020, 507: 684–699.
- [7] LIU Dun, LIANG Decui, WANG Changchun. A novel three-way decision model based on incomplete information system[J]. *Knowledge-based systems*, 2016, 91:

- 32–45.
- [8] SUN Bingzhen, MA Weimin, ZHAO Haiyan. Decision-theoretic rough fuzzy set model and application[J]. *Information sciences*, 2014, 283: 180–196.
- [9] DOU Huili, YANG Xibei, SONG Xiaoning, et al. Decision-theoretic rough set: a multicost strategy[J]. *Knowledge-based systems*, 2016, 91: 71–83.
- [10] LI Weiwei, HUANG Zhiqiu, JIA Xiuyi, et al. Neighborhood based decision-theoretic rough set models[J]. *International journal of approximate reasoning*, 2016, 69: 1–17.
- [11] SHU Wenhao, QIAN Wenbin, XIE Yonghong. Incremental approaches for feature selection from dynamic data with the variation of multiple objects[J]. *Knowledge-based systems*, 2019, 163: 320–331.
- [12] 段海玲, 王光琼. 一种高效的复杂信息系统增量式属性约简 [J]. 华南理工大学学报(自然科学版), 2019, 47(6): 18–30.
- DUAN Hailing, WANG Guangqiong. An efficient incremental attribute reduction for complex information systems[J]. *Journal of South China University of Technology (natural science edition)*, 2019, 47(6): 18–30.
- [13] HU Chengxiang, ZHANG Li, WANG Bangjun, et al. Incremental updating knowledge in neighborhood multi-granulation rough sets under dynamic granular structures[J]. *Knowledge-based systems*, 2019, 163: 811–829.
- [14] ZHANG Qinghua, LV Gongxun, CHEN Yuhong, et al. A dynamic three-way decision model based on the updating of attribute values[J]. *Knowledge-based systems*, 2018, 142: 71–84.
- [15] XU Jianfeng, MIAO Duoqian, ZHANG Yuanjian, et al. A three-way decisions model with probabilistic rough sets for stream computing[J]. *International journal of approximate reasoning*, 2017, 88: 1–22.
- [16] CHEN Hongmei, LI Tianrui, LUO Chuan, et al. A decision-theoretic rough set approach for dynamic data mining[J]. *IEEE transactions on fuzzy systems*, 2015, 23(6): 1958–1970.
- [17] 赵小龙, 杨燕. 基于邻域粒化条件熵的增量式属性约简算法 [J]. 控制与决策, 2019, 34(10): 2061–2072.
- ZHAO Xiaolong, YANG Yan. Incremental attribute reduction algorithm based on neighborhood granulation conditional entropy[J]. *Control and decision*, 2019, 34(10): 2061–2072.
- [18] 杨臻, 邱保志. 混合信息系统的动态变精度粗糙集模型 [J]. 控制与决策, 2020, 35(2): 297–308.
- YANG Zhen, QIU Baozhi. Dynamic variable precision rough set model of mixed information system[J]. *Control and decision*, 2020, 35(2): 297–308.
- [19] LUO Chuan, LI Tianrui, ZHANG Yi, et al. Matrix approach to decision-theoretic rough sets for evolving data[J]. *Knowledge-based systems*, 2016, 99: 123–134.

#### 作者简介:



孙海霞, 副教授, CCF 会员, 主要研究方向为数据挖掘和网络安全。主持安徽省质量工程项目 2 项。发表学术论文 10 余篇。