

# 控制与决策

Control and Decision

基于矩阵的双论域模糊概率粗糙集增量更新算法

刘丹, 李敬伟

引用本文:

刘丹, 李敬伟. 基于矩阵的双论域模糊概率粗糙集增量更新算法[J]. 控制与决策, 2021, 36(3): 553–564.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2019.0692>

---

## 您可能感兴趣的其他文章

Articles you may be interested in

[区间粗糙数信息系统的覆盖分类冗余度与属性约简](#)

Coverage classification redundancy and attribute reduction of interval rough number information system

控制与决策. 2021, 36(3): 677–685 <https://doi.org/10.13195/j.kzyjc.2019.0744>

[基于HI-DD-AdaBoost.RT的锂离子动力电池SOH预测](#)

Prediction of Li-ion battery SOH based on HI-DD-AdaBoost.RT

控制与决策. 2021, 36(3): 686–692 <https://doi.org/10.13195/j.kzyjc.2019.0764>

[尺度自适应的多特征融合相关滤波目标跟踪算法](#)

Scale adaptation and multi-feature fusion correlation filtering object tracking algorithm

控制与决策. 2021, 36(2): 429–435 <https://doi.org/10.13195/j.kzyjc.2019.0445>

[基于知识粒度特征的多目标粗糙集属性约简算法](#)

Multi objective rough set attribute reduction algorithm based on characteristics of knowledge granularity

控制与决策. 2021, 36(1): 196–205 <https://doi.org/10.13195/j.kzyjc.2019.0490>

[基于不变网络模型和故障注入的分布式信息系统故障溯源方法](#)

Fault source location algorithm for distributed information system based on invariant network and fault injection

控制与决策. 2020, 35(11): 2723–2732 <https://doi.org/10.13195/j.kzyjc.2019.0214>

# 基于矩阵的双论域模糊概率粗糙集增量更新算法

刘丹<sup>†</sup>, 李敬伟

(1. 河南工学院 计算机科学与技术学院, 河南 新乡 453003;  
2. 河南省生产制造物联大数据工程技术研究中心, 河南 新乡 453003)

**摘要:** 双论域模糊概率粗糙集是针对双论域信息系统的一种新的数据挖掘模型, 现实应用中的双论域信息系统总是处于动态更新中, 针对该问题提出一种基于矩阵的双论域模糊概率粗糙集增量式更新方法. 首先, 通过矩阵方法重新对双论域模糊概率粗糙集进行表示; 然后, 在矩阵表示模型的基础上, 分别研究双论域信息系统两个论域中对象增加和减少时模型的增量式更新机制; 最后, 基于该增量式更新提出相应的增量式更新算法. 实验分析表明: 相比较于非增量式更新算法, 所提出的增量式更新算法可以在很短的时间内完成模型的动态更新, 从而验证算法的有效性; 同时, 与其他相关算法相比, 所提出算法具有一定的优越性.

**关键词:** 粗糙集; 双论域信息系统; 模糊关系; 概率; 增量式学习; 矩阵

中图分类号: TP18 文献标志码: A

DOI: 10.13195/j.kzyjc.2019.0692

开放科学(资源服务)标识码(OSID):



引用格式: 刘丹, 李敬伟. 基于矩阵的双论域模糊概率粗糙集增量更新算法[J]. 控制与决策, 2021, 36(3): 553-564.

## Incremental updating of fuzzy probability rough sets over two universes based on matrix method

LIU Dan<sup>†</sup>, LI Jing-wei

(1. College of Computer Science & Technology, Henan Institute of Technology, Xinxiang 453003, China; 2. Big Data Engineering Research Center of Henan for Production & Manufacturing IoTs, Xinxiang 453003, China)

**Abstract:** The fuzzy probability rough set over two universes is a new data mining model for the two universes information system. However, the two universes information system in real application is always in dynamic updating. In order to solve this problem, a fuzzy probability rough set over two universes incremental updating method based on the matrix is proposed. Firstly, the fuzzy probability rough sets over two universes are expressed by matrix method. Then, on the basis of matrix expression model, the incremental updating mechanism of the model is studied when the objects increase and decrease in two universes of the two universes information system, respectively. And based on this incremental updating, the corresponding incremental updating algorithm is finally proposed. The experimental analysis shows that the proposed incremental updating algorithm can complete dynamic updating of the model in a very short time compared to the non-incremental updating algorithm, which verifies the effectiveness of the proposed algorithm. At the same time, compared with other related algorithms, the proposed algorithm has certain advantages.

**Keywords:** rough set; two universes information system; fuzzy relation; probability; incremental learning; matrix

## 0 引言

粗糙集理论<sup>[1]</sup>是一种新的数据挖掘模型, 尤其在不一致和不确定性的数据中表现出了优秀的处理性能<sup>[2-4]</sup>. 传统的粗糙集适用于标准的离散型信息系统, 目前已推广至其他各类复杂的信息系统中<sup>[5-8]</sup>, 这使得粗糙集理论的应用范围逐渐变广.

在信息处理领域, 双论域信息系统是一种重要的数据形式, 其广泛存在于个性化推荐<sup>[9]</sup>、医疗诊断<sup>[10]</sup>

以及需求预测<sup>[11]</sup>等应用中. 针对这类信息系统, Yao 等<sup>[12]</sup>提出了最早的双论域粗糙集模型, 为双论域信息系统的粗糙集研究奠定了基础. Yao 等提出的模型是基于精确的二元关系, 为了适用于数值和模糊环境下的问题分析, 学者们提出了基于模糊关系的双论域粗糙集模型. 例如, Sun 等<sup>[13]</sup>提出的基于双论域信息系统的模糊粗糙集, Zhang 等<sup>[14]</sup>提出的基于双论域的区间值模糊粗糙集, 郭智莲等<sup>[15]</sup>提出的双论域上的

收稿日期: 2019-05-19; 修回日期: 2019-10-28.

基金项目: 河南省高等学校重点科研项目(19B520005).

责任编辑: 阳春华.

<sup>†</sup>通讯作者. E-mail: liudan\_luckygirl@163.com.

直觉模糊概率粗糙集模型. 另一方面, 针对数据中噪声信息存在, 学者们将双论域粗糙集进行改进, 提出了双论域概率粗糙集的相关模型<sup>[16-17]</sup>. 之后, 为了兼顾模糊和概率两方面的性能, 一种称为双论域模糊概率粗糙集模型被提出<sup>[18]</sup>, 该模型在双论域信息系统的粗糙计算方面具有更强的优越性.

现实环境下的数据总是时刻处于动态更新之中, 而传统模型大多都是针对静态数据而建立, 因此解决这类问题的主要方法是对模型构造增量式学习<sup>[19]</sup>. 在粗糙集理论中, 目前已提出多种模型的增量式更新方法<sup>[19-22]</sup>. Hu等<sup>[23]</sup>基于集合处理的方式提出了一种增量式更新算法, 该方法的增量式更新效率在数据变化的过程中表现出较大的差异, 因此实际运用中不具有较好的稳定性, 受数据集的变化影响较大.

基于粗糙集理论的增量式更新研究中, 很多学者采用矩阵方法来构造模型和算法的增量式学习<sup>[24-28]</sup>. 因此, 本文在其他学者研究的基础上, 提出一种基于矩阵方法的双论域模糊概率粗糙集增量式更新算法. 首先利用矩阵重新对双论域模糊概率粗糙集进行表示, 然后基于矩阵方法分别研究了双论域信息系统中两个论域对象增加和减少时模型的增量式更新. 实验分析证明: 所提出的增量式更新算法具有一定有效性; 同时, 与Hu等学者提出的算法相比, 所提出的增量式更新算法具有更好的稳定性.

## 1 基本理论

**定义1**<sup>[10]</sup> 设双论域信息系统表示为  $S = (U, V)$ , 这里的  $U$  和  $V$  分别为两个非空有限论域, 在  $U$  和  $V$  中可以确定一个二元关系  $R$ , 则称  $(U, V, R)$  为双论域信息系统  $S$  的一个双论域近似空间. 根据二元关系  $R$ , 可以得到两个映射关系

$$R(x) = \{y \in V | (x, y) \in R\}, \quad x \in U, \quad (1)$$

$$R^{-1}(y) = \{x \in U | (x, y) \in R\}, \quad y \in V. \quad (2)$$

**定义2**<sup>[10]</sup> 给定双论域近似空间  $(U, V, R)$ ,  $\forall Y \subseteq V$  关于该近似空间的下近似和上近似分别定义为

$$\underline{R}(Y) = \{x \in U | R(x) \subseteq Y\}, \quad (3)$$

$$\overline{R}(Y) = \{x \in U | R(x) \cap Y \neq \emptyset\}. \quad (4)$$

定义2表明, 传统的双论域粗糙集上下近似的定义较为严格, 对含噪声的数据不具有很好的容忍效果, 并且为了适用于数值或模糊性的数据, Yang等<sup>[18]</sup>又进一步提出了双论域模糊概率粗糙集.

**定义3**<sup>[18]</sup> 设双论域信息系统  $S = (U, V)$ , FR称为  $U$  和  $V$  上的模糊关系, 其中  $FR: U \times V \rightarrow [0, 1]$ . 令  $\lambda \in (0, 1]$ , 则定义:

$$1) FR_\lambda = \{(x, y) \in U \times V | FR(x, y) \geq \lambda\}; \quad (5)$$

$$2) FR_\lambda(x) = \{y \in V | FR(x, y) \geq \lambda\}; \quad (6)$$

$$3) FR_\lambda^{-1}(y) = \{x \in U | FR(x, y) \geq \lambda\}. \quad (7)$$

**定义4**<sup>[18]</sup> 设双论域信息系统  $S = (U, V)$ , FR为  $U$  和  $V$  上的模糊关系. 令  $\lambda \in (0, 1]$ , 且两个阈值  $\alpha$  和  $\beta$  满足  $0 \leq \beta < \alpha \leq 1$ , 则对于  $\forall Y \subseteq V$  关于FR的模糊概率下近似和模糊概率上近似分别定义为

$$\underline{FR}_{\lambda, \alpha}(Y) = \{x \in U | P(Y | FR_\lambda(x)) \geq \alpha, FR_\lambda(x) \neq \emptyset\}, \quad (8)$$

$$\overline{FR}_{\lambda, \beta}(Y) = \{x \in U | P(Y | FR_\lambda(x)) > \beta, FR_\lambda(x) \neq \emptyset\} \cup \{x \in U | FR_\lambda(x) = \emptyset\}. \quad (9)$$

其中  $P(Y | FR_\lambda(x)) = \frac{|Y \cap FR_\lambda(x)|}{|FR_\lambda(x)|}$  为  $FR_\lambda(x)$  包含  $Y$  中对象的概率. 称  $\{\underline{FR}_{\lambda, \alpha}(Y), \overline{FR}_{\lambda, \beta}(Y)\}$  为  $Y$  在模糊关系FR下关于参数  $\lambda, \alpha$  和  $\beta$  的双论域模糊概率粗糙集.

## 2 双论域模糊概率粗糙集的矩阵表示

**定义5** 设双论域信息系统  $S = (U, V)$ , 其中  $U = \{x_1, x_2, \dots, x_n\}, V = \{y_1, y_2, \dots, y_m\}$ . 对于  $\forall Y \subseteq V$  在  $V$  上的特征向量定义为

$$Y = [f_1, f_2, \dots, f_m]^T. \quad (10)$$

这里: 当  $y_i \in Y, 1 \leq i \leq m$  时,  $f_i = 1$ ; 当  $y_i \notin Y, 1 \leq i \leq m$  时,  $f_i = 0$ .

**定义6** 设双论域信息系统  $S = (U, V)$ , 其中  $U = \{x_1, x_2, \dots, x_n\}, V = \{y_1, y_2, \dots, y_m\}$ . FR为  $U$  和  $V$  上的模糊关系,  $\lambda \in (0, 1]$ , 则  $U$  和  $V$  关于模糊关系FR的  $\lambda$  关系矩阵定义为

$$M_{n \times m}^\lambda = [fr_\lambda(i, j)]_{n \times m}, \quad (11)$$

其中  $fr_\lambda(i, j)$  为  $\lambda$  关系矩阵  $M_{n \times m}^\lambda$  的第  $i$  行第  $j$  列元素. 当  $FR(x_i, y_j) \geq \lambda$  时,  $fr_\lambda(i, j) = 1$ , 当  $FR(x_i, y_j) < \lambda$  时,  $fr_\lambda(i, j) = 0$ .

定义6表明, 模糊关系FR的  $\lambda$  关系矩阵是一个  $n$  行  $m$  列且只包含0和1两种元素的矩阵,  $n$  行代表了双论域  $U$  中  $n$  个对象,  $m$  列代表了双论域  $V$  中  $m$  个对象, 对于  $x_i \in U$  和  $y_j \in V$ , 若  $FR(x_i, y_j) \geq \lambda$ , 则此时  $\lambda$  关系矩阵第  $i$  行第  $j$  列元素为1, 否则为0.

**定理1** 设集合  $P, Q \subseteq U, |U| = n$ , 两个集合在  $U$  上的特征向量分别为  $P$  和  $Q$ , 则

$$|P \cap Q| = P^T Q. \quad (12)$$

**证明** 令特征向量  $P = [p_1, p_2, \dots, p_n]^T, Q = [q_1, q_2, \dots, q_n]^T$ . 根据向量的乘法计算规则, 有  $P^T Q = \sum_{i=1}^n p_i q_i$ . 根据定义5, 对于  $\forall x_k \in P \cap Q$  满足  $p_k q_k = 1$ , 而对于  $\forall x_l \notin P \cap Q$  满足  $p_l q_l = 0$ , 因此有  $\sum_{i=1}^n p_i q_i = |P \cap Q|$ .  $\square$

根据定理1,结合定义6可以得到如下定理.

**定理2** 设双论域信息系统  $S = (U, V)$ , 其中  $U = \{x_1, x_2, \dots, x_n\}, V = \{y_1, y_2, \dots, y_m\}$ . FR为  $U$  和  $V$  上的模糊关系,  $\lambda \in (0, 1]$ , 模糊关系FR的  $\lambda$  关系矩阵为  $M_{n \times m}^\lambda$ . 令  $Y \subseteq V$  对应的特征向量为  $\mathbf{Y}$ , 则有

$$M_{n \times m}^\lambda \mathbf{Y} = \begin{bmatrix} |Y \cap FR_\lambda(x_1)| \\ |Y \cap FR_\lambda(x_2)| \\ \vdots \\ |Y \cap FR_\lambda(x_n)| \end{bmatrix}. \quad (13)$$

**证明** 将  $M_{n \times m}^\lambda$  表示成行向量的形式, 即  $M_{n \times m}^\lambda = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n]^T$ , 其中  $\mathbf{v}_i (1 \leq i \leq n)$  均为一行  $m$  列的行向量. 根据定义6可以得到  $\mathbf{v}_i = [fr_\lambda(i, 1), fr_\lambda(i, 2), \dots, fr_\lambda(i, m)]$ . 根据定义3, 向量  $\mathbf{v}_i$  即为对象  $x_i \in U$  在模糊关系FR下  $FR_\lambda(x_i)$  的特征向量. 因此, 根据矩阵的乘法规则和定理1, 定理2成立.  $\square$

**定义7** 设双论域信息系统  $S = (U, V)$ , 其中  $U = \{x_1, x_2, \dots, x_n\}, V = \{y_1, y_2, \dots, y_m\}$ . FR为  $U$  和  $V$  上的模糊关系,  $\lambda \in (0, 1]$ , 模糊关系FR的  $\lambda$  关系矩阵为  $M_{n \times m}^\lambda$ . 定义  $M_{n \times m}^\lambda$  的行基数向量  $\mathbf{C}$  为

$$\mathbf{C} = \left[ \sum_{j=1}^m fr_\lambda(1, j), \sum_{j=1}^m fr_\lambda(2, j), \dots, \sum_{j=1}^m fr_\lambda(n, j) \right]^T. \quad (14)$$

特别地, 当存在  $\sum_{j=1}^m fr_\lambda(\cdot, j) = 0$  时, 将向量  $\mathbf{C}$  对应位置的值记为  $-1$ .

定义7表明, 向量  $\mathbf{C}$  中每个元素即为论域  $U$  中每个对象  $x$  在模糊关系FR下集合  $FR_\lambda(x)$  的基数, 即向量  $\mathbf{C}$  也可以表示为  $\mathbf{C} = [|FR_\lambda(x_1)|, |FR_\lambda(x_2)|, \dots, |FR_\lambda(x_n)|]^T$ .

**定理3** 设双论域信息系统  $S = (U, V)$ , 其中  $U = \{x_1, x_2, \dots, x_n\}, V = \{y_1, y_2, \dots, y_m\}$ . FR为  $U$  和  $V$  上的模糊关系,  $\lambda \in (0, 1]$ , 模糊关系FR的  $\lambda$  关系矩阵为  $M_{n \times m}^\lambda$ , 相应的行基数向量为  $\mathbf{C}$ . 令  $Y \subseteq V$  对应的特征向量为  $\mathbf{Y}$ , 则定义  $\mathbf{Y}$  的概率向量  $\mathbf{P}_Y$  为

$$\mathbf{P}_Y = (M_{n \times m}^\lambda \mathbf{Y}) ./ \mathbf{C} = \begin{bmatrix} |Y \cap FR_\lambda(x_1)| / |FR_\lambda(x_1)| \\ |Y \cap FR_\lambda(x_2)| / |FR_\lambda(x_2)| \\ \vdots \\ |Y \cap FR_\lambda(x_n)| / |FR_\lambda(x_n)| \end{bmatrix} = \begin{bmatrix} P(Y|FR_\lambda(x_1)) \\ P(Y|FR_\lambda(x_2)) \\ \vdots \\ P(Y|FR_\lambda(x_n)) \end{bmatrix}. \quad (15)$$

其中“./”为矩阵的点除运算, 也就是将矩阵对应元素间进行相除.

定理3可根据定理2和定义7直接得到, 此略.

**定理4** 设双论域信息系统  $S = (U, V)$ , 其中  $U = \{x_1, x_2, \dots, x_n\}, V = \{y_1, y_2, \dots, y_m\}$ . FR为  $U$

和  $V$  上的模糊关系,  $\lambda \in (0, 1]$ , 模糊关系FR的  $\lambda$  关系矩阵为  $M_{n \times m}^\lambda$ , 对应的行基数向量为  $\mathbf{C}$ . 令  $Y \subseteq V$  的特征向量为  $\mathbf{Y}$ , 并且  $\mathbf{Y}$  的概率向量为  $\mathbf{P}_Y$ . 对于  $0 \leq \beta < \alpha \leq 1, Y$  关于FR的模糊概率下近似和模糊概率上近似分别表示为  $\underline{FR}_{\lambda, \alpha}(Y)$  和  $\overline{FR}_{\lambda, \beta}(Y)$ , 有

$$\underline{FR}_{\lambda, \alpha}(Y) = \{x_i \in U | [P_Y]_i \geq \alpha \wedge [C]_i \neq -1\}, \quad (16)$$

$$\overline{FR}_{\lambda, \beta}(Y) = \{x_i \in U | ([P_Y]_i > \beta \wedge [C]_i \neq -1) \vee [C]_i = -1\}, \quad (17)$$

其中  $[P_Y]_i$  和  $[C]_i$  分别表示列向量的第  $i$  个元素值.

定理4可根据定义4和定理3直接得到, 此略.

定理4表明, 当得到近似对象集  $Y \subseteq V$  关于模糊关系的概率向量  $\mathbf{P}_Y$  时, 可以直接得到双论域模糊概率粗糙集的上下近似结果.

基于矩阵方法的双论域模糊概率粗糙集上下近似计算算法如下所示.

**算法1** 基于矩阵方法的双论域模糊概率粗糙集计算算法.

输入: 双论域信息系统  $S = (U, V), |U| = n, |V| = m$ ; 模糊关系FR; 近似对象集  $Y \subseteq V$ ; 参数  $\lambda, \alpha$  和  $\beta$ .

输出:  $\underline{FR}_{\lambda, \alpha}(Y), \overline{FR}_{\lambda, \beta}(Y)$ .

step 1: 计算近似对象集的特征向量;

step 2: 对双论域信息系统  $S = (U, V)$  按照模糊关系FR计算  $\lambda$  关系矩阵  $M_{n \times m}^\lambda$ ;

step 3: 计算关系矩阵  $M_{n \times m}^\lambda$  的行基数向量  $\mathbf{C}$ ;

step 4: 计算概率向量  $(M_{n \times m}^\lambda \mathbf{Y}) ./ \mathbf{C}$ ;

step 5: 由定理4导出  $Y$  关于FR的模糊概率的上下近似, 即  $\underline{FR}_{\lambda, \alpha}(Y)$  和  $\overline{FR}_{\lambda, \beta}(Y)$ .

整个算法1的时间复杂度为  $O(4nm + 4m)$ . 对双论域进行矩阵计算与进行集合计算, 其计算量是一样的, 因此双论域模糊概率粗糙集传统的计算时间复杂度与算法1的时间复杂度是一致的.

### 3 基于矩阵的双论域模糊概率粗糙集增量式更新

在现实应用环境下, 其数据时时刻刻处于动态更新之中, 这给传统的数据挖掘模型和算法带来了巨大的挑战. 增量式学习就是为了应对这种挑战而提出的一种新的学习方式<sup>[19-22]</sup>. 现实中的双论域信息系统往往是不断动态更新的, 因此, 在第2节研究的基础上, 提出一种基于矩阵的双论域模糊概率粗糙集增量式更新方法.

在双论域信息系统  $S = (U, V)$  中, 数据的动态更新主要表现为3种形式, 第1种为论域  $U$  中对象的增加和减少, 第2种为论域  $V$  中对象的增加和减少, 第3种为论域  $U$  和  $V$  中对象的同时增加和减少, 而第3种的变化形式可以转化为前两种的变化形式, 因此双论

域信息系统的动态更新本质上只包含第1种变化和  
第2种变化,本文将运用矩阵来研究双论域模糊概率  
粗糙集在这两种变化形式下的增量式更新。

### 3.1 论域 $U$ 对象增加时双论域模糊概率粗糙集的增量式更新

在本小节,主要研究当双论域信息系统 $S = (U, V)$ 中论域 $U$ 的对象发生增加时,双论域模糊概率粗糙集的增量式更新问题。

为了本节叙述的方便,这里统一给出各符号的表示. 设更新前的双论域信息系统 $S = (U, V)$ ,其中论域 $U = \{x_1, x_2, \dots, x_n\}$ . 当论域 $U$ 中的对象发生增加,设增加的对象集为 $U^+$ ,且 $U^+ = \{x_{n+1}, x_{n+2}, \dots, x_{n+n^+}\}$ ,更新后的信息系统表示为 $S' = (U' = U \cup U^+, V)$ ,记 $n' = n + n^+$ . 同时,设更新前双论域信息系统下的模糊关系为FR,更新后的模糊关系为FR'. 根据第2节中关于双论域模糊概率粗糙集的矩阵表示方法,接下来将运用矩阵去增量式更新出新的粗糙集结果。

**定理5** 设双论域信息系统 $S = (U, V)$ ,其中 $U = \{x_1, x_2, \dots, x_n\}, V = \{y_1, y_2, \dots, y_m\}$ . FR为 $U$ 和 $V$ 上的模糊关系, $\lambda \in (0, 1]$ ,模糊关系FR的 $\lambda$ 关系矩阵为 $M_{n \times m}^\lambda$ ,相应的行基数向量为 $C$ . 令 $Y \subseteq V$ 对应的特征向量为 $Y$ ,并且 $Y$ 的概率向量为 $P_Y$ . 当信息系统论域 $U$ 增加对象集 $U^+$ 后,特征向量 $Y$ 对应新的概率向量 $P'_Y$ 增量式计算如下:

$$(P'_Y)_{n' \times m} = \begin{bmatrix} (P'_Y)_{n \times m} \\ (P'_Y)_{n^+ \times m} \end{bmatrix}, n' = n + n^+. \quad (18)$$

其中: $(P'_Y)_{n^+ \times m} = (M_{n^+ \times m}^\lambda Y) ./ C^+, M_{n^+ \times m}^\lambda$ 表示 $U^+$ 与 $V$ 在模糊关系FR'下的 $\lambda$ 关系矩阵, $C^+$ 表示 $M_{n^+ \times m}^\lambda$ 的行基数向量。

**证明** 当论域 $U$ 中对象发生增加,根据定义6有 $M_{n' \times m}^\lambda = [M_{n \times m}^\lambda \ M_{n^+ \times m}^\lambda]^T$ ,那么相应地有 $C' = [C \ C^+]^T$ ,因此

$$P'_Y = (M_{n' \times m}^\lambda Y) ./ C' = [P_Y (M_{n^+ \times m}^\lambda Y) ./ C^+]^T. \quad \square$$

定理4表明,当双论域信息系统论域 $U$ 增加对象集后,只需要对增加的对象集计算出相应的概率向量 $P'_Y$ ,然后将 $P'_Y$ 与 $P_Y$ 进行合并,便可以得到新信息系统下的概率向量 $P'_Y$ ,由于 $P'_Y$ 与 $P_Y$ 是相互独立的,只需要让 $P'_Y$ 按照定理3计算出对应的上下近似集,最后再进行合并便可以得到最终的近似集结果。

### 3.2 论域 $U$ 对象减少时双论域模糊概率粗糙集的增量式更新

类似于3.1节的方法,同样可以得到当论域 $U$ 中对象减少时,双论域模糊概率粗糙集的增量式更新。

设更新前的双论域信息系统 $S = (U, V)$ ,其中论

域 $U = \{x_1, x_2, \dots, x_n\}$ ,当论域 $U$ 减少对象集 $U^- \subseteq U$ ,且 $U^- = \{x_{t_1}, x_{t_2}, \dots, x_{t_{n^-}}\}$ ,更新后的信息系统表示为 $S' = (U' = U - U^-, V)$ ,记 $n' = n - n^-$ . 同时,设更新前双论域信息系统下的模糊关系为FR,更新后的模糊关系为FR'.

同定理5,论域 $U$ 减少后双论域模糊概率粗糙集的动态更新也只需要对变化的对象进行计算更新。

**定理6** 设双论域信息系统 $S = (U, V)$ ,其中 $U = \{x_1, x_2, \dots, x_n\}, V = \{y_1, y_2, \dots, y_m\}$ . FR为 $U$ 和 $V$ 上的模糊关系, $\lambda \in (0, 1]$ ,模糊关系FR的 $\lambda$ 关系矩阵为 $M_{n \times m}^\lambda$ ,相应的行基数向量为 $C$ . 令 $Y \subseteq V$ 对应的特征向量为 $Y$ ,特征向量 $Y$ 的概率向量为 $P_Y$ . 当信息系统论域 $U$ 减少对象集 $U^- = \{x_{t_1}, x_{t_2}, \dots, x_{t_{n^-}}\}$ 后,特征向量 $Y$ 对应新的概率向量 $P'_Y$ 即为删去 $P_Y$ 中第 $t_1, t_2, \dots, t_{n^-}$ 个元素。

**证明** 当论域 $U$ 减少对象集 $U^-$ 时,则根据定义6,新的 $\lambda$ 关系矩阵 $M_{n' \times m}^\lambda$ 即为在原先 $\lambda$ 关系矩阵 $M_{n \times m}^\lambda$ 的基础上删去第 $t_1, t_2, \dots, t_{n^-}$ 行元素. 同样,向量 $C'$ 也是在原先 $C$ 的基础上删去第 $t_1, t_2, \dots, t_{n^-}$ 个元素. 因此根据定理3, $P'_Y$ 即为在 $P_Y$ 中删去第 $t_1, t_2, \dots, t_{n^-}$ 个元素.  $\square$

定理6表明,当论域 $U$ 中对象发生减少时,只需对概率向量中对应位置的元素进行删除便可以更新出新的概率向量,则新的上下近似集也只需要在原来上下近似集的基础上删除对应的对象即可,因此这样的计算具有很高的效率。

### 3.3 论域 $V$ 对象增加时双论域模糊概率粗糙集的增量式更新

在本节,将利用矩阵方法研究当论域 $V$ 中对象增加时,双论域模糊概率粗糙集的增量式更新问题。

设更新前的双论域信息系统 $S = (U, V)$ ,其中论域 $V = \{y_1, y_2, \dots, y_m\}$ ,当论域 $V$ 中的对象发生增加时,设增加的对象集为 $V^+$ ,且 $V^+ = \{y_{m+1}, y_{m+2}, \dots, y_{m+m^+}\}$ ,更新后的双论域信息系统表示为 $S' = (U, V' = V \cup V^+)$ ,记 $m' = m + m^+$ . 同时,设更新前双论域信息系统下的模糊关系为FR,更新后的模糊关系为FR'.

**定理7** 设双论域信息系统 $S = (U, V)$ ,其中 $U = \{x_1, x_2, \dots, x_n\}, V = \{y_1, y_2, \dots, y_m\}$ . FR为 $U$ 和 $V$ 上的模糊关系, $\lambda \in (0, 1]$ ,模糊关系FR的 $\lambda$ 关系矩阵为 $M_{n \times m}^\lambda$ ,相应的行基数向量为 $C$ . 令 $Y \subseteq V$ 对应的特征向量为 $Y$ ,特征向量 $Y$ 的概率向量为 $P_Y$ . 当论域 $V$ 增加对象集 $V^+$ 后,同时 $Y' = Y \cup Y^+$ ,这里的 $Y^+ \subseteq V^+$ ,并且 $Y'$ 对应的特征向量为 $Y'$ ,则 $Y'$ 对应新的概率向量 $P'_Y$ ,增量式计算为

$$P_{Y'} = (P_Y \cdot C + M_{n \times m}^\lambda Y_{V^+}^+) ./ (C + C^+). \quad (19)$$

其中：“·”为点乘运算，即矩阵的对应元素相乘，与点除互为逆运算； $M_{n \times m}^\lambda$ 为论域 $U$ 与对象集 $V^+$ 之间的 $\lambda$ 关系矩阵； $C^+$ 为 $M_{n \times m}^\lambda$ 的行基数向量； $Y_{V^+}^+$ 为 $Y^+$ 在对象集 $V^+$ 下的特征向量，即 $Y_{V^+}^+$ 是一个包含 $m^+$ 个元素的列向量。

**证明** 根据定理3，可以得到

$$P_Y \cdot C = M_{n \times m}^\lambda Y = \begin{bmatrix} |Y \cap FR_\lambda(x_1)| \\ |Y \cap FR_\lambda(x_2)| \\ \vdots \\ |Y \cap FR_\lambda(x_n)| \end{bmatrix}.$$

由于

$$M_{n \times m}^\lambda Y_{V^+}^+ = \begin{bmatrix} |Y^+ \cap FR_\lambda^+(x_1)| \\ |Y^+ \cap FR_\lambda^+(x_2)| \\ \vdots \\ |Y^+ \cap FR_\lambda^+(x_n)| \end{bmatrix},$$

其中 $FR_\lambda^+(x_i) = \{y \in V^+ | FR'(x, y) \geq \lambda\}$ ，有

$$P_Y \cdot C + M_{n \times m}^\lambda Y_{V^+}^+ = \begin{bmatrix} |Y \cap FR_\lambda(x_1)| \\ |Y \cap FR_\lambda(x_2)| \\ \vdots \\ |Y \cap FR_\lambda(x_n)| \end{bmatrix} + \begin{bmatrix} |Y^+ \cap FR_\lambda^+(x_1)| \\ |Y^+ \cap FR_\lambda^+(x_2)| \\ \vdots \\ |Y^+ \cap FR_\lambda^+(x_n)| \end{bmatrix}.$$

由于 $[Y \cap FR_\lambda(x_i)] \cap [Y^+ \cap FR_\lambda^+(x_i)] = \emptyset$ ，有

$$|Y \cap FR_\lambda(x_i)| + |Y^+ \cap FR_\lambda^+(x_i)| = |[Y \cap FR_\lambda(x_i)] \cup [Y^+ \cap FR_\lambda^+(x_i)]|,$$

由于 $Y \cap Y^+ = \emptyset$ ， $FR_\lambda(x_i) \cap FR_\lambda^+(x_i) = \emptyset$ ，有

$$|[Y \cap FR_\lambda(x_i)] \cup [Y^+ \cap FR_\lambda^+(x_i)]| = |(Y \cup Y^+) \cap (FR_\lambda(x_i) \cup FR_\lambda^+(x_i))|,$$

又 $Y \cup Y^+ = Y'$ ， $FR_\lambda(x_i) \cup FR_\lambda^+(x_i) = FR'_\lambda(x_i)$ ，则进一步有

$$|Y \cap FR_\lambda(x_i)| + |Y^+ \cap FR_\lambda^+(x_i)| = |Y' \cap FR'_\lambda(x_i)|,$$

因此

$$P_Y \cdot C + M_{n \times m}^\lambda Y_{V^+}^+ = \begin{bmatrix} |Y' \cap FR'_\lambda(x_1)| \\ |Y' \cap FR'_\lambda(x_2)| \\ \vdots \\ |Y' \cap FR'_\lambda(x_n)| \end{bmatrix}.$$

由于 $C + C^+$ 为 $M_{n \times m}^\lambda$ 的行基数向量，根据定理3可以得到 $Y'$ 对应新的概率向量。□

在新概率向量 $P_{Y'}$ 的基础上，按照定理4便可直接得到新的双论域模糊概率粗糙集上下近似集。

### 3.4 论域 $V$ 对象减少时双论域模糊概率粗糙集的增量式更新

本节利用矩阵方法研究当论域 $V$ 中对象减少时，双论域模糊概率粗糙集的增量式更新。

设更新前的双论域信息系统 $S = (U, V)$ ，其中论域 $V = \{y_1, y_2, \dots, y_m\}$ ，当论域 $V$ 中的对象减少时，设减少的对象集为 $V^-$ ，且 $V^- = \{y_{t_1}, y_{t_2}, \dots, y_{t_{m^-}}\}$ ，更新后的信息系统表示为 $S' = (U, V' = V - V^-)$ ，记 $m' = m - m^-$ 。同时，设更新前双论域信息系统下的模糊关系为 $FR$ ，更新后的模糊关系为 $FR'$ 。

**定理8** 设双论域信息系统 $S = (U, V)$ ，其中 $U = \{x_1, x_2, \dots, x_n\}$ ， $V = \{y_1, y_2, \dots, y_m\}$ 。 $FR$ 为 $U$ 和 $V$ 上的模糊关系， $\lambda \in (0, 1]$ ，模糊关系 $FR$ 的 $\lambda$ 关系矩阵为 $M_{n \times m}^\lambda$ ，相应的行基数向量为 $C$ 。令 $Y \subseteq V$ 对应的特征向量为 $Y$ ，并且 $Y$ 的概率向量为 $P_Y$ 。当信息系统论域 $V$ 减少对象集 $V^-$ 后，此时 $Y' = Y - Y^-$ ，这里的 $Y^- \subseteq V^-$ ，并且 $Y'$ 对应的特征向量为 $Y'$ ，则 $Y'$ 对应新的概率向量 $P_{Y'}$ ，增量式计算为

$$P_{Y'} = (P_Y \cdot C - M_{n \times m}^\lambda Y_{V^-}^-) ./ (C - C^-). \quad (20)$$

其中： $M_{n \times m}^\lambda$ 为 $U$ 与 $V^-$ 之间的 $\lambda$ 关系矩阵； $C^-$ 为 $M_{n \times m}^\lambda$ 的行基数向量； $Y_{V^-}^-$ 为 $Y^-$ 在对象集 $V^-$ 下的特征向量，即 $Y_{V^-}^-$ 是一个包含 $m^-$ 个元素的列向量。

定理8的证明过程类似于定理7，此略。

同样地，在定理8得到新的概率向量 $P_{Y'}$ 的基础上，按照定理4也可直接得到新的双论域模糊概率粗糙集上下近似集。

## 4 双论域模糊概率粗糙集增量式更新算法

在第3节的基础上，本节将进一步提出双论域模糊概率粗糙集的增量式更新算法。

### 4.1 论域 $U$ 对象增加时的增量式更新算法

**算法2** 论域 $U$ 对象增加时基于矩阵的双论域模糊概率粗糙集增量式更新算法。

**输入:** 1) 更新后的双论域信息系统 $S' = (U', V)$ ， $U' = U \cup U^+$ ，参数 $\lambda$ 、 $\alpha$ 和 $\beta$ ；2)  $Y \subseteq V$ 对应的特征向量为 $Y$ ；3)  $Y$ 在原来信息系统下的上下近似集 $\underline{FR}_{\lambda, \alpha}(Y)$ 和 $\overline{FR}_{\lambda, \beta}(Y)$ 。

**输出:**  $Y$ 在新的信息系统下的上下近似集 $\underline{FR}'_{\lambda, \alpha}(Y)$ 和 $\overline{FR}'_{\lambda, \beta}(Y)$ 。

step 1: 计算 $U^+$ 和 $V$ 的 $\lambda$ 关系矩阵 $M_{n^+ \times m}^\lambda$ ；

step 2: 根据 $M_{n^+ \times m}^\lambda$ 计算 $C^+$ ；

step 3: 计算 $[P_Y^+]_{n^+ \times m} = (M_{n^+ \times m}^\lambda Y) ./ C^+$ ；

step 4: 对于 $[P_Y^+]_{n^+ \times m}$ ，按照定理3计算出 $Y$ 的上下近似集，记为 $\underline{FR}'_{\lambda, \alpha}(Y)$ 和 $\overline{FR}'_{\lambda, \beta}(Y)$ ；

step 5:  $Y$ 在新信息系统下的上下近似集为

$$\begin{aligned}\underline{\text{FR}}'_{\lambda,\alpha}(Y) &= \underline{\text{FR}}_{\lambda,\alpha}(Y) \cup \underline{\text{FR}}^+_{\lambda,\alpha}(Y), \\ \overline{\text{FR}}'_{\lambda,\alpha}(Y) &= \overline{\text{FR}}_{\lambda,\alpha}(Y) \cup \overline{\text{FR}}^+_{\lambda,\alpha}(Y).\end{aligned}$$

在算法2中,令 $|U| = n, |V| = m, |U^+| = n^+$ ,则整个算法2的时间复杂度为 $O(3n^+m + 3n^+)$ .

#### 4.2 论域 $U$ 对象减少时的增量式更新算法

**算法3** 论域 $U$ 对象减少时基于矩阵的双论域模糊概率粗糙集增量式更新算法.

**输入:** 1) 更新后的双论域信息系统 $S' = (U', V)$ ,  $U' = U - U^-$ ,  $U^- = x_{t_1}, x_{t_2}, \dots, x_{t_{n^-}}$ ; 2)  $Y$ 在原来信息系统下的上下近似集 $\underline{\text{FR}}_{\lambda,\alpha}(Y)$ 和 $\overline{\text{FR}}_{\lambda,\beta}(Y)$ .

**输出:**  $Y$ 在新的信息系统下的上下近似集 $\underline{\text{FR}}'_{\lambda,\alpha}(Y)$ 和 $\overline{\text{FR}}'_{\lambda,\beta}(Y)$ .

step 1: 分别在 $\underline{\text{FR}}_{\lambda,\alpha}(Y)$ 和 $\overline{\text{FR}}_{\lambda,\beta}(Y)$ 中删除对象 $x_{t_1}, x_{t_2}, \dots, x_{t_{n^-}}$ ;

step 2: 返回 $Y$ 在新信息系统下的上下近似集 $\underline{\text{FR}}'_{\lambda,\alpha}(Y)$ 和 $\overline{\text{FR}}'_{\lambda,\beta}(Y)$ .

算法3只对原先的上下近似集进行相应元素的删除,令 $|U| = n, |V| = m, |U^-| = n^-$ ,因此整个算法3的时间复杂度为 $O(2n^-)$ .

#### 4.3 论域 $V$ 对象增加时的增量式更新算法

**算法4** 论域 $V$ 对象增加时基于矩阵的双论域模糊概率粗糙集增量式更新算法.

**输入:** 1) 更新后的双论域信息系统 $S' = (U, V')$ ,  $V' = V \cup V^+$ ,  $Y' = Y \cup Y^+$ ,以及参数 $\lambda, \alpha$ 和 $\beta$ ; 2)  $\lambda$ 关系矩阵为 $M_{n \times m}^\lambda$ ,行基数向量 $C$ , $Y$ 对应的特征向量 $Y$ ,特征向量 $Y$ 的概率向量 $P_Y$ ; 3)  $Y$ 在原来信息系统下的上下近似集 $\underline{\text{FR}}_{\lambda,\alpha}(Y)$ 和 $\overline{\text{FR}}_{\lambda,\beta}(Y)$ .

**输出:**  $Y'$ 在新的信息系统下的上下近似集 $\underline{\text{FR}}'_{\lambda,\alpha}(Y')$ 和 $\overline{\text{FR}}'_{\lambda,\beta}(Y')$ .

step 1: 计算 $Y^+$ 对应的特征向量 $Y_{V^+}^+$ ;

step 2: 计算 $U$ 和 $V^+$ 的 $\lambda$ 关系矩阵 $M_{n \times m^+}^\lambda$ ;

step 3: 计算 $M_{n \times m^+}^\lambda$ 的行基数向量 $C^+$ ;

step 4: 计算新的概率向量

$$P'_{Y'} = (P_Y * C + M_{n \times m^+}^\lambda Y_{V^+}^+) / (C + C^+);$$

step 5: 根据 $P'_{Y'}$ 计算出 $Y'$ 在新信息系统下的上下近似集 $\underline{\text{FR}}'_{\lambda,\alpha}(Y')$ 和 $\overline{\text{FR}}'_{\lambda,\beta}(Y')$ .

在算法4中,令 $|U| = n, |V| = m, |V^+| = m^+$ ,则整个算法4的时间复杂度为 $O(3nm^+ + 5n + m^+)$ .

#### 4.4 论域 $V$ 对象减少时的增量式更新算法

**算法5** 论域 $V$ 对象减少时基于矩阵的双论域模糊概率粗糙集增量式更新算法.

**输入:** 1) 更新后的双论域信息系统 $S' = (U, V')$ ,  $V' = V - V^-$ ,  $Y' = Y - Y^-$ ,以及参数 $\lambda, \alpha$ 和 $\beta$ ; 2)  $\lambda$ 关系矩阵为 $M_{n \times m}^\lambda$ ,行基数向量 $C$ , $Y$ 对应的特征向量

$Y$ ,特征向量 $Y$ 的概率向量 $P_Y$ ; 3)  $Y$ 在原来信息系统下的上下近似集 $\underline{\text{FR}}_{\lambda,\alpha}(Y)$ 和 $\overline{\text{FR}}_{\lambda,\beta}(Y)$ .

**输出:**  $Y'$ 在新的信息系统下的上下近似集 $\underline{\text{FR}}'_{\lambda,\alpha}(Y')$ 和 $\overline{\text{FR}}'_{\lambda,\beta}(Y')$ .

step 1: 计算 $Y^-$ 对应的特征向量 $Y_{V^-}$ ;

step 2: 计算 $U$ 和 $V^-$ 的 $\lambda$ 关系矩阵 $M_{n \times m^-}^\lambda$ ;

step 3: 计算 $M_{n \times m^-}^\lambda$ 的行基数向量 $C^-$ ;

step 4: 计算新的概率向量

$$P'_{Y'} = (P_Y * C - M_{n \times m^-}^\lambda Y_{V^-}) / (C + C^-);$$

step 5: 根据 $P'_{Y'}$ 计算出 $Y'$ 在新信息系统下的上下近似集 $\underline{\text{FR}}'_{\lambda,\alpha}(Y')$ 和 $\overline{\text{FR}}'_{\lambda,\beta}(Y')$ .

令 $|U| = n, |V| = m, |V^-| = m^-$ ,类似于算法4,整个算法5的时间复杂度为 $O(3nm^- + 5n + m^-)$ .

### 5 双论域模糊概率粗糙集增量式更新算法

本节将通过实验来验证所提出增量式更新算法的有效性和优越性.表1所示的是实验中的数据集,其中序号1~序号3的数据集为推荐系统中关于用户对电影进行评分的公用数据集,序号4~序号6为通过编程工具随机生成的人工数据集.

表1 实验数据集

序号	数据集	简称	论域 $U$	论域 $V$
1	MovieLens 1 M	ML_1M	6 040	3 952
2	MovieLens 100 k	ML_100k	943	1 682
3	Jester	Jester	73 421	100
4	人工数据集1	Data_1	6 000	8 000
5	人工数据集2	Data_2	8 000	11 000
6	人工数据集3	Data_3	12 000	15 000

#### 5.1 实验设置

本实验所要验证的是增量式更新算法的有效性和优越性,而表1所示的均为静态数据集,因此这里需要人为构造数据的动态性.仿照其他更新算法的处理方式<sup>[25-27]</sup>,对于数据集论域对象的动态增加情形,从空集开始,依次从剩余论域中选择10%的对象进行加入;对于数据集论域对象的动态减少情形,依次从论域中选择10%的对象进行删除,即每次完成数据集10%的更新变化量,这样便可构造出数据集的10次更新.下文都是基于该处理方法进行实验.

在实验中,记所提出的算法1为双论域模糊概率粗糙集的非增量式更新算法,简称非增量式更新算法.所提出的算法2~算法5为双论域模糊概率粗糙集的增量式更新算法,简称增量式更新算法.将非增量式更新算法与增量式更新算法进行比较,从而验证增量式更新算法的有效性.同时,将所提出的增量式更新算法与Hu等<sup>[23]</sup>学者提出的同类型增量式更新算法进行实验对比,以此来验证所提算法的优越性.

本实验中设定  $\lambda = 0.6, \alpha = 0.7, \beta = 0.4$ .

### 5.2 与非增量式更新算法对比

图1所示的是当双论域信息系统论域  $U$  逐渐增大时,基于矩阵方法的模糊概率粗糙集非增量式更新

算法(算法1)与增量式更新算法(算法2)在各个数据集下更新上下近似所需的时间比较. 其中每幅图的横坐标表示论域  $U$  的更新比例(%),纵坐标表示的是更新时所需的计算时间(s).

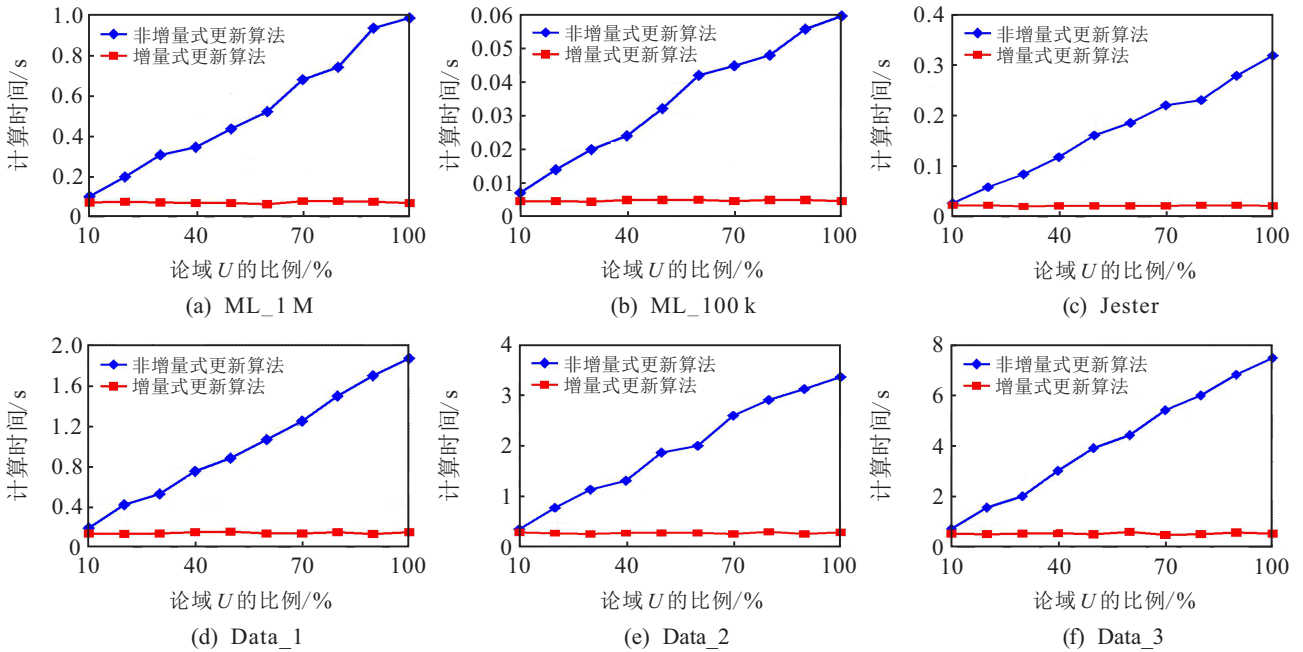


图1 论域  $U$  增加时模型的更新时间比较(与非增量式算法对比)

观察图1的实验结果可以发现,随着论域  $U$  的逐渐增大,非增量式更新算法更新上下近似所需的时间也是逐渐增大的,而增量式更新算法的计算时间保持平稳,并且始终处于一个比较低的时间量上,不随论域  $U$  的变化而变化. 这主要是由于它们计算机制的不同导致的,非增量式更新算法采用的是非增量式计算,随着数据的不断更新增大,其计算过程中矩阵

的规模也是不断增大的,因而消耗了大量的计算时间. 而对于增量式更新算法,该算法采用增量式的方法进行计算,定理4已表明,只需对新加入的数据进行相关的矩阵计算便可以更新出新的上下近似集结果. 由于实验中数据集每次更新时增加的数据量是大致相等的,增量式更新算法每次的计算时间是基本不变的,从而具有更高的动态更新性能.

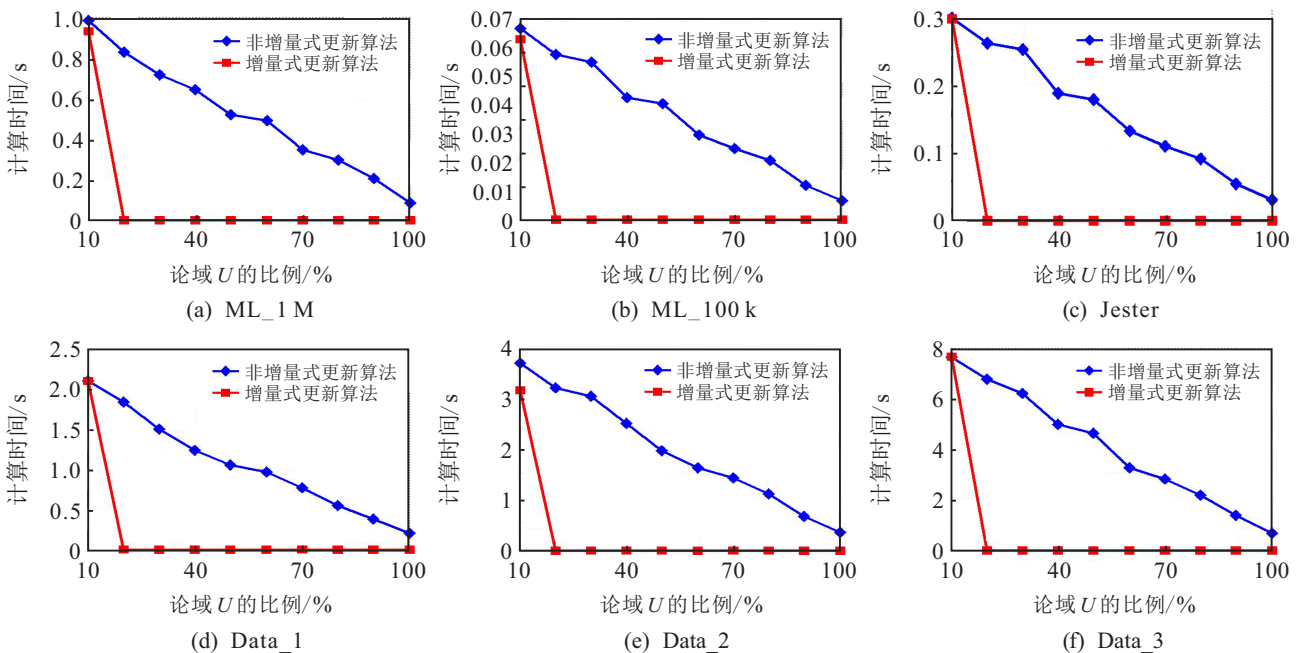


图2 论域  $U$  减少时模型的更新时间比较(与非增量式算法对比)

图2所示的是当双论域信息系统论域 $U$ 逐渐减少时,非增量式更新算法(算法1)与增量式更新算法(算法3)在各个数据集下更新上下近似所需的时间比较.观察图2的实验结果可以发现,在初始整个数据集时,非增量式更新算法与增量式更新算法的计算时间几乎是一样的.当论域 $U$ 中的对象更新减少后,增量式更新算法在已知信息上进行增量计算,因此所需的计算时间大大减小,可以看出计算时间处于一个非常少的水平.定理5已表明,当论域 $U$ 中对象减少后,只需对原来粗糙集上下近似中的相应元素进行删除便可直接更新出新的结果,而删除可以在很短的时间内完成,因此增量式更新算法具有很少的时间消耗,

远远低于非增量式更新算法.

图3所示的是当双论域信息系统论域 $V$ 逐渐增加时,非增量式更新算法(算法1)与增量式更新算法(算法4)在各个数据集下的更新时间结果比较.观察图3可以看出其时间结果与图1的结果类似,即非增量式更新算法的计算时间随着论域 $V$ 的逐渐增大而增大,而增量式更新算法的计算时间保持稳定且处于一个较低的时间水平.主要原因同样是由于增量式更新算法仅对新加入的数据进行相关计算,然后将计算结果与原来数据的模型结果进行合并,从而得到最终新数据的模型结果.因此,该增量式更新算法同样具有很高的动态数据更新性能.

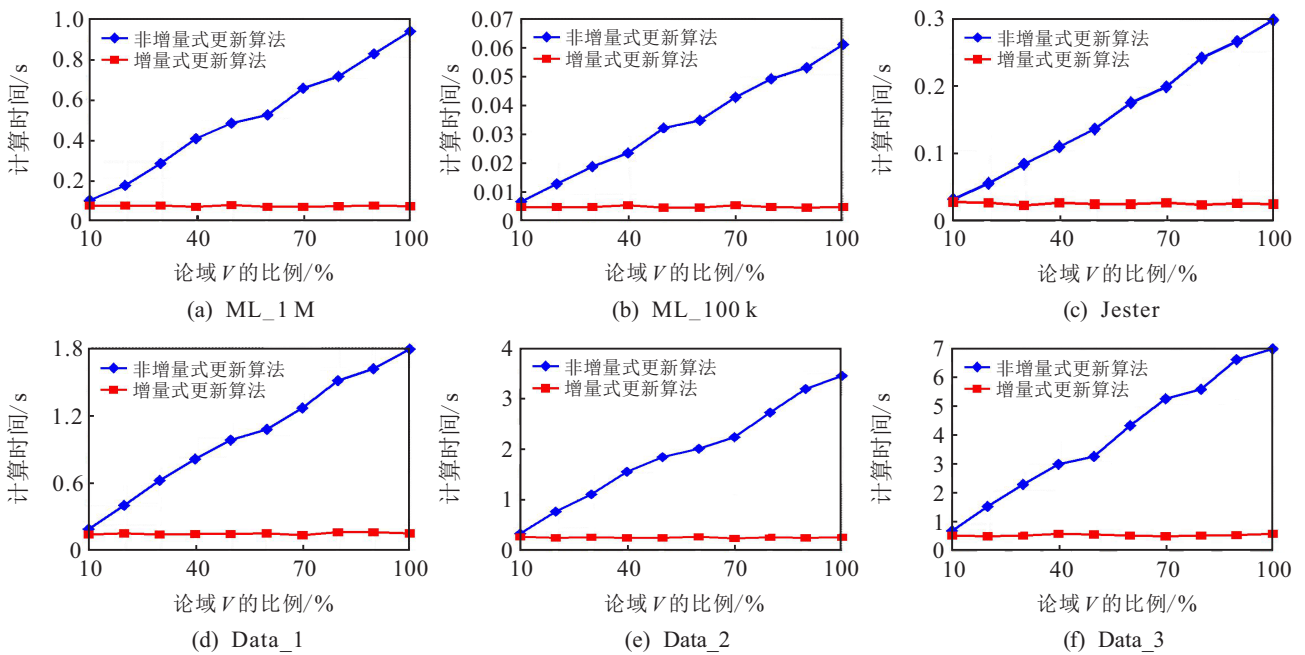


图3 论域 $V$ 增加时模型的更新时间比较(与非增量式算法对比)

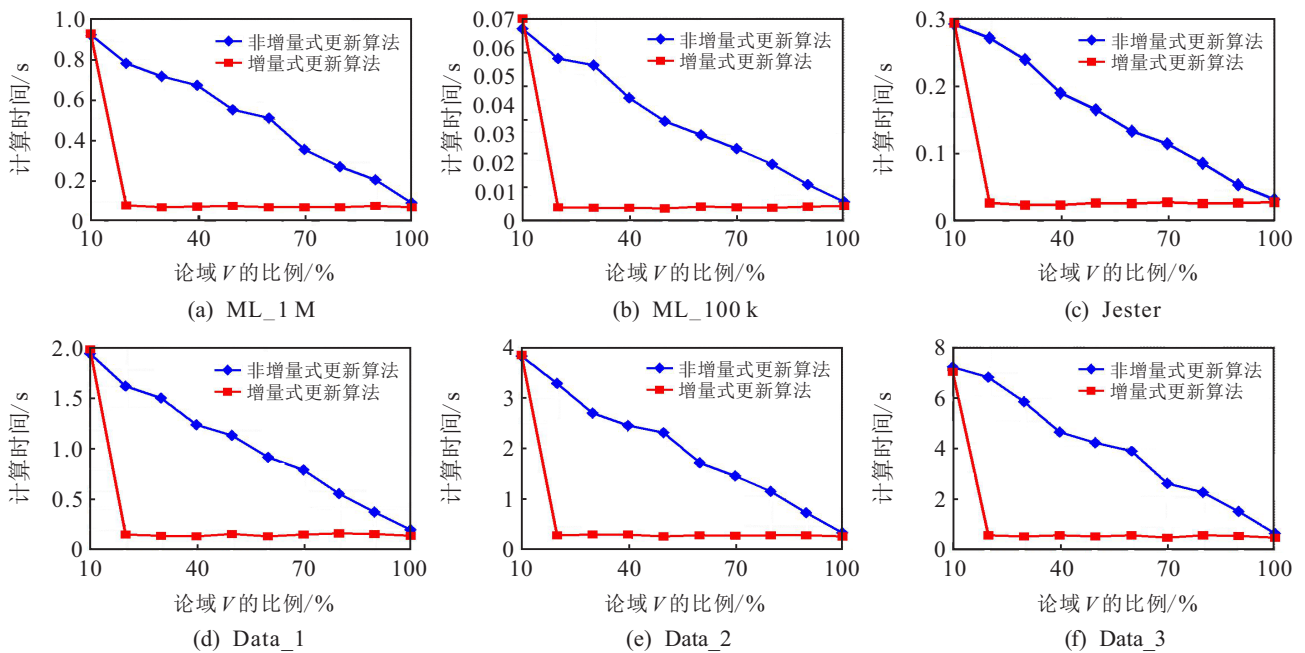


图4 论域 $V$ 减少时模型的更新时间比较(与非增量式算法对比)

图4所示的是当双论域信息系统论域 $V$ 逐渐减少时,非增量式更新算法(算法1)与增量式更新算法(算法5)在各个数据集下更新上下近似所需的时间比较.观察图4的结果可以发现,与图2的结果类似.初始时,非增量式更新算法与增量式更新算法所需的计算时间是一样的.当论域 $V$ 中对象更新减少后,增量式更新算法所需的时间保持平稳且时间量较小,原因也是由于增量式更新只对变化的数据进行计算便可以更新出最终的模型结果,即时间消耗只与变化的数据相关,因此增量式更新算法具有很高的更新效率.可以看出,相比于非增量式更新算法,所提出的增量式更新算法在动态双论域信息系统中具有更高的处

理效率,验证了所提算法的有效性.

### 5.3 与其他同类型增量式算法的对比

本节将本文所提出的增量式更新算法与Hu等<sup>[23]</sup>提出的算法进行实验对比,仿照5.2节的实验过程,对比两类算法分别在论域 $U$ 、 $V$ 动态增加和动态减少时模型的更新效率比较结果.

图5~图8所示的是本文算法与Hu等的算法对所有动态数据集更新上下近似所需时间的比较结果.图中Hu等的算法记为已提出的算法,同时每幅图的横坐标表示论域的更新比例(%),纵坐标表示的是更新时所需的计算时间(s).

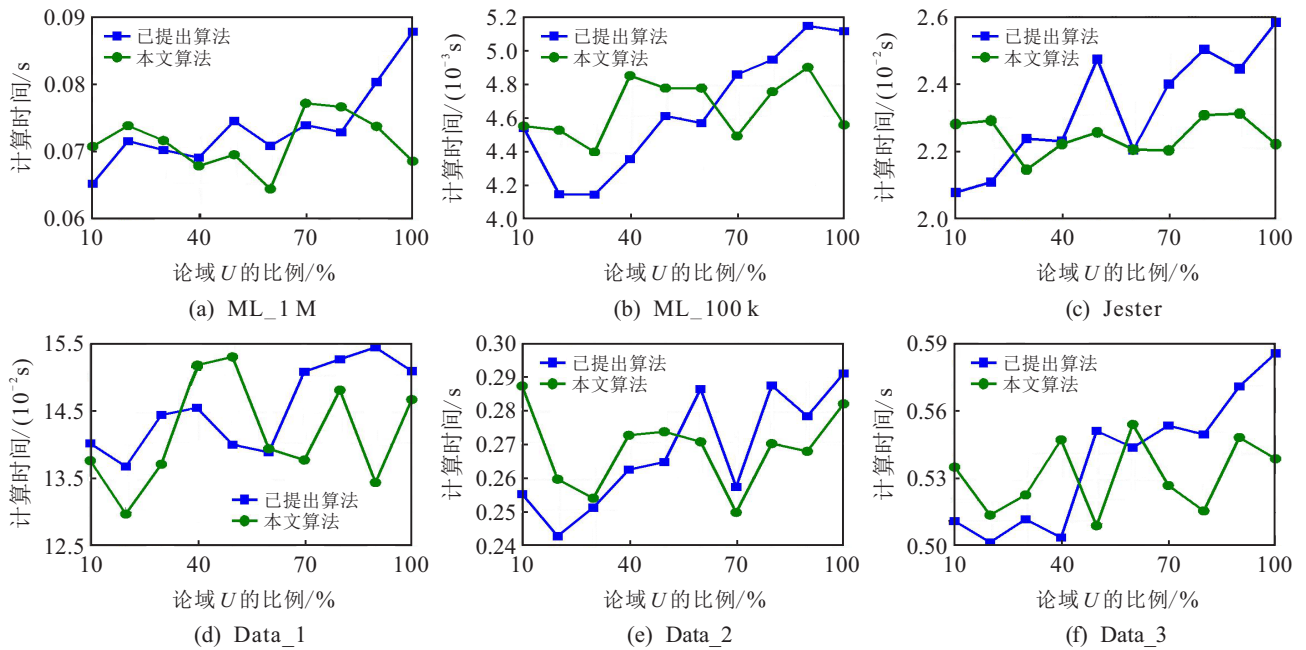


图5 论域 $U$ 增加时模型的更新时间比较(与增量式算法对比)

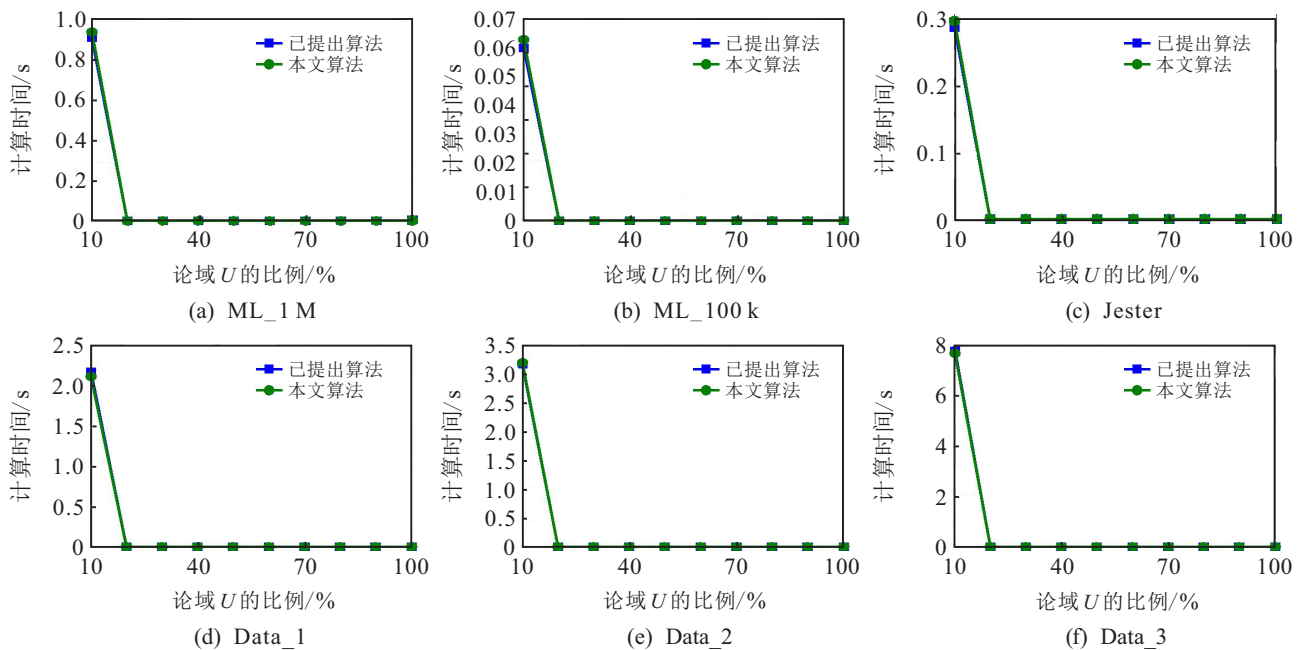
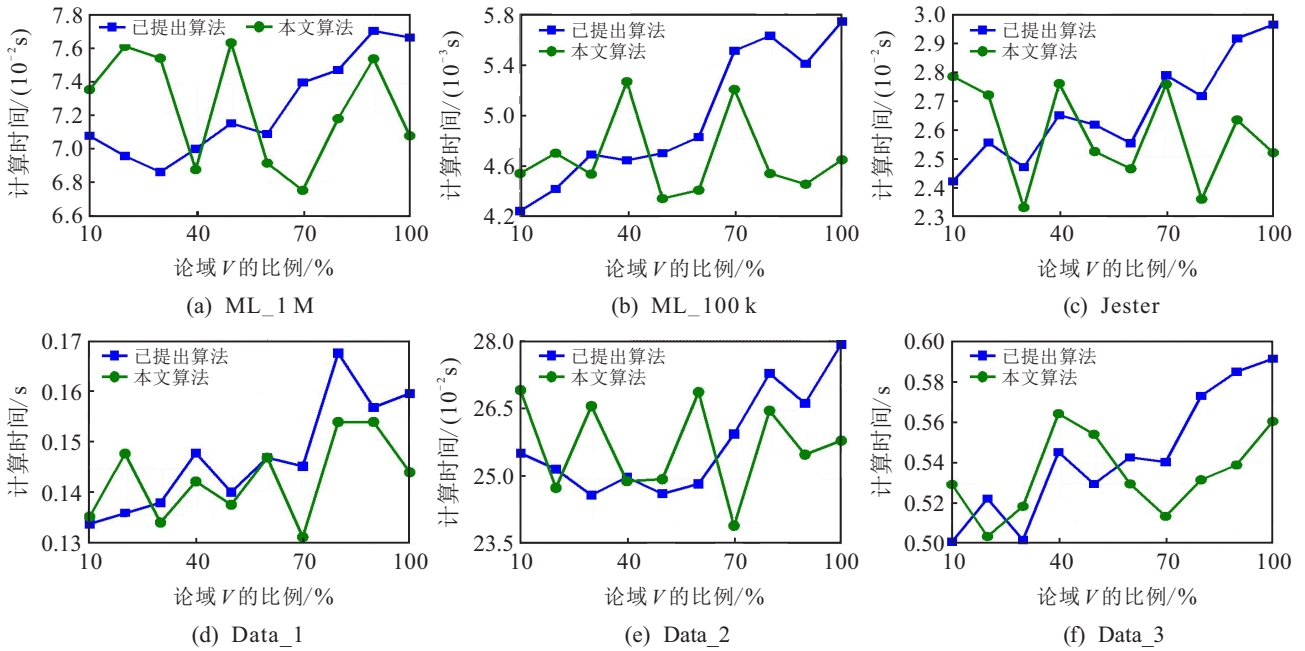
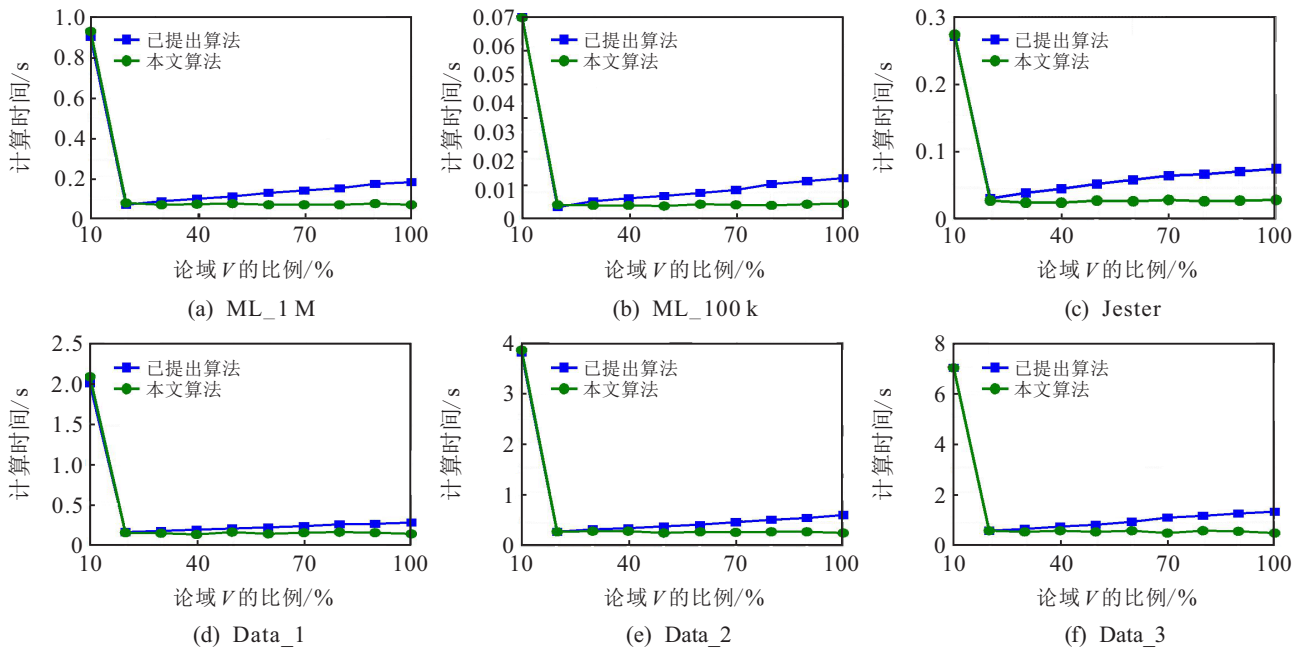


图6 论域 $U$ 减少时模型的更新时间比较(与增量式算法对比)

图7 论域  $V$  增加时模型的更新时间比较(与增量式算法对比)图8 论域  $V$  减少时模型的更新时间比较(与增量式算法对比)

观察图5和图8的实验结果可以看出,当双论域信息系统刚开始动态更新时,Hu等所提出算法的更新时耗要少于本文算法,而随着更新次数的不断增加,其数据集的规模不断增大,Hu等算法的更新耗时将不断增加,并高于本文的更新算法.这主要是由于Hu等的算法是基于集合计算的,随着数据集的变大,其进行的增量计算量也逐渐变大,而本文的算法是基于矩阵方法的计算,由于数据集每次增加的比例是一样的,矩阵方法每次的计算量也是一样的,所以对于论域  $U$  或论域  $V$  的逐渐增大,本文所提出的增量式更新算法的整体效率要更高.

观察图6中论域  $U$  逐渐减小时的实验比较结果,可以看出两种算法的所需时间是一致的.对于论域  $U$  的逐渐减小,Hu等的算法和本文算法都可以很快地完成增量更新.观察图8关于论域  $V$  逐渐减小时的实验比较结果,可以看出随着数据集动态更新次数的增加,本文所提出算法更新数据集所需的时间消耗保持稳定,而Hu等所提出的算法随着更新次数的增多,其时间的消耗逐渐增加且高于本文的算法,主要原因与图5和图8相同,因此对于论域  $V$  的动态减少,本文算法更具优势.

表2所示的是当信息系统变化时,本文算法与Hu

表2 平均时间增长率rate比较

%

数据集	U 增加		U 减少		V 增加		V 减少	
	Hu 等的算法	本文算法	Hu 等的算法	本文算法	Hu 等的算法	本文算法	Hu 等的算法	本文算法
ML_1M	3.5136	0.0124	-0.6445	-0.6445	0.9252	-0.0938	5.885	-1.5476
ML_100k	1.4314	0.1543	-1.1349	-1.1349	3.5179	0.3757	6.574	1.0976
Jester	2.6462	-0.0078	1.8118	1.8118	2.3633	-0.4803	5.178	0.6306
Data_1	0.8928	0.3739	-0.3408	-0.3408	2.1835	1.1001	2.5355	-0.7026
Data_2	0.9024	0.2785	1.9695	1.9695	1.0507	-0.2058	4.545	0.0834
Data_3	2.0499	0.3920	1.6615	1.6615	1.9451	1.2302	5.257	-0.4301

等的算法在各个数据集下更新所需时间的平均增长率结果,其中平均时间增长率定义为

$$rate = \frac{1}{k-1} \cdot \sum_{i=1}^{k-1} \frac{t_{i+1} - t_i}{t_i} \cdot 100\%$$

其中:  $k$  为增量更新次数,  $t_i$  为第  $i$  次更新时的耗时.

观察表2可以发现,对于论域  $U$  增加、 $V$  增加和  $V$  减少的情形,随着数据集的不断更新,Hu 等的算法的平均时间增长率基本上都高于本文提出的算法,并且部分情形具有很高的增长率值,这意味着Hu 等的算法效率受数据集变化的影响较大,不具有很好的稳定性.而本文算法具有较低的平均时间增长率,并且有的数据集下平均增长率接近于0,这说明本文算法的效率受数据集变化的影响较小,稳定性更高.

5.4 实验总结

综合以上实验结果可以看出,本文所提出的增量式更新算法的更新效率远高于非增量式更新算法,说明了本文算法的有效性.同时与已提出的增量式更新算法相比,所提出的算法整体上具有较高效率,说明了本文算法的优越性.因此,本文所提出的算法可以适应于双论域信息系统不断动态更新时的模糊概率粗糙集计算.另一方面,由于推荐系统中的数据集是最直接的双论域表现形式,本文的算法可以高效地运用于实际的个性化推荐研究之中.

6 结论

针对双论域信息系统的动态更新问题,本文提出一种矩阵方法的双论域模糊概率粗糙集增量式更新算法.首先利用矩阵对双论域模糊概率粗糙集进行重新表示,然后在其基础上构造了模型的增量式计算,最后基于该计算方法提出了相应的双论域模糊概率粗糙集增量式更新算法.仿真实验结果表明了所提出的增量式更新算法能够快速地对模型结果进行更新,具有很高的计算性能.接下来,可以将所提算法运用于现实环境下双论域信息系统的增量式更新问题,例如个性化推荐系统和医疗诊断系统等.

参考文献(References)

- [1] Pawlak Z. Rough sets[J]. International Journal Computer Information Science, 1982, 11(5): 341-356.
- [2] 姚晟, 徐风, 吴照玉, 等. 基于邻域粗糙互信息熵的非单调性属性约简[J]. 控制与决策, 2019, 34(2): 353-361. (Yao S, Xu F, Wu Z Y, et al. Non-monotonic attribute reduction based on neighborhood rough mutual information entropy[J]. Control and Decision, 2019, 34(2): 353-361.)
- [3] 蒋瑜. 基于改进差别信息树的粗糙集属性约简算法[J]. 控制与决策, 2019, 34(6): 1253-1258. (Jiang Y. Attribute reduction with rough set based on improved discernibility information tree[J]. Control and Decision, 2019, 34(6): 1253-1258.)
- [4] Dai J H, Wei B J, Zhang X H, et al. Uncertainty measurement for incomplete interval-valued information systems based on  $\alpha$ -weak similarity[J]. Knowledge-Based Systems, 2017, 136(15): 159-171.
- [5] Hu Q H, Yu D R, Liu J F, et al. Neighborhood rough set based heterogeneous feature subset selection[J]. Information Sciences, 2008, 178(18): 3577-3594.
- [6] 史德容, 徐伟华. 区间值模糊决策序信息系统的分布约简[J]. 计算机科学与探索, 2017, 11(4): 652-658. (Shi D R, Xu W H. Distribution reduction in interval-valued fuzzy decision ordered information systems[J]. Journal of Frontiers of Computer Science & Technology, 2017, 11(4): 652-658.)
- [7] Zhang X X, Chen D G, Tsang E C C. Generalized dominance rough set models for the dominance intuitionistic fuzzy information systems[J]. Information Sciences, 2017, 378(1): 1-25.
- [8] Lin Y J, Li Y W, Wang C X, et al. Attribute reduction for multi-label learning with fuzzy rough set[J]. Knowledge-Based Systems, 2018, 152(15): 51-61.
- [9] Zhang H R, Min F. Three-way recommender systems based on random forests[J]. Knowledge-Based Systems, 2016, 91: 275-286.
- [10] Yan R X, Zheng J G, Liu J L, et al. Research on the model of rough set over dual-universes[J]. Knowledge-Based

- Systems, 2010, 23(8): 817-822.
- [11] Sun B Z, Ma W M, Zhao H Y. A fuzzy rough set approach to emergency material demand prediction over two universes[J]. Applied Mathematical Modelling, 2013, 37(10/11): 7062-7070.
- [12] Yao Y Y, Wong S K M, Wang L S. A non-numeric approach to uncertain reasoning[J]. International Journal of General Systems, 1995, 23(4): 343-359.
- [13] Sun B Z, Ma W M. Fuzzy rough set model on two different universes and its application[J]. Applied Mathematical Modelling, 2011, 35(4): 1798-1809.
- [14] Zhang H Y, Zhang W X, Wu W Z. On characterization of generalized interval-valued fuzzy rough sets on two universes of discourse[J]. International Journal of Approximate Reasoning, 2009, 51(1): 56-70.
- [15] 郭智莲, 杨海龙, 王珏. 双论域上的直觉模糊概率粗糙集模型及其应用[J]. 系统工程理论与实践, 2014, 34(7): 1828-1834.  
(Guo Z L, Yang H L, Wang J. Intuitionistic fuzzy probabilistic rough set model on two universes and its applications[J]. Systems Engineering — Theory & Practice, 2014, 34(7): 1828-1834.)
- [16] Ma W M, Sun B Z. Probabilistic rough set over two universes and rough entropy[J]. International Journal of Approximate Reasoning, 2012, 53(4): 608-619.
- [17] 李敬, 王利东. 面向不完备信息系统的双论域决策粗糙集——基于双相对量化信息的角度[J]. 计算机科学与探索, 2018, 12(4): 653-661.  
(Li J, Wang L D. Decision-theoretic rough sets on two universes within incomplete information system: From the view of double relative quantitative information[J]. Journal of Frontiers of Computer Science and Technology, 2018, 12(4): 653-661.)
- [18] Yang H L, Liao X W, Wang S Y, et al. Fuzzy probabilistic rough set model on two universes and its applications[J]. International Journal of Approximate Reasoning, 2013, 54(9): 1410-1420.
- [19] Luo C, Li T R, Yao Y Y. Dynamic probabilistic rough sets with incomplete data[J]. Information Sciences, 2017, 417: 39-54.
- [20] 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 X L, Yang Y. Incremental attribute reduction algorithm based on neighborhood granulation conditional entropy[J]. Control and Decision, 2019, 34(10): 2061-2072.)
- [21] Chen D G, Yang Y Y, Dong Z. An incremental algorithm for attribute reduction with variable precision rough sets[J]. Applied Soft Computing, 2016, 45: 129-149.
- [22] Liang J Y, Wang F, Dang C Y, et al. A group incremental approach to feature selection applying rough set technique[J]. IEEE Transactions on Knowledge & Data Engineering, 2014, 26(2): 294-308.
- [23] Hu J, Li T R, Luo C, et al. Incremental fuzzy probabilistic rough sets over two universes[J]. International Journal of Approximate Reasoning, 2017, 81: 28-48.
- [24] Liu D, Li T R, Zhang J B. A rough set-based incremental approach for learning knowledge in dynamic incomplete information systems[J]. International Journal of Approximate Reasoning, 2014, 55(8): 1764-1786.
- [25] Jing Y G, Li T R, Luo C, et al. An incremental approach for attribute reduction based on knowledge granularity[J]. Knowledge-Based Systems, 2016, 104(15): 24-38.
- [26] Luo C, Li T R, Yi Z, et al. Matrix approach to decision-theoretic rough sets for evolving data[J]. Knowledge-Based Systems, 2016, 99(1): 123-134.
- [27] Huang Y Y, Li T R, Luo C, et al. Matrix-based dynamic updating rough fuzzy approximations for data mining[J]. Knowledge-Based Systems, 2017, 119(1): 273-283.
- [28] 尹林子, 阳春华, 王晓丽, 等. 基于标记可辨识矩阵的增量式属性约简算法[J]. 自动化学报, 2014, 40(3): 397-404.  
(Yin L Z, Yang C H, Wang X L, et al. An incremental algorithm for attribute reduction based on labeled discernibility matrix[J]. Acta Automatica Sinica, 2014, 40(3): 397-404.)

### 作者简介

刘丹(1978—), 女, 副教授, 硕士, 从事数据分析与智能算法、分布式处理等研究, E-mail: liudan\_luckygirl@163.com;

李敬伟(1957—), 男, 讲师, 硕士, 从事大数据和云计算等研究, E-mail: lijingweixx@126.com.

(责任编辑: 齐 霖)