

拟单层覆盖粗糙集中近似集的增量更新算法

吴正江, 张亚宁, 张真, 梅秋雨, 杨天

(河南理工大学 计算机科学与技术学院, 河南 焦作 454003)

摘要: 拟单层覆盖粗糙集与集值信息系统之间存在一一对应的映射关系, 当集值信息系统中的对象集动态添加或删除时, 对应拟单层覆盖粗糙集中的信息单元也会随之改变, 导致拟单层覆盖粗糙集中的近似集发生变化。针对拟单层覆盖粗糙集中近似集的动态更新问题, 将拟单层覆盖粗糙集与增量学习相结合, 提出近似集的增量更新算法。设计拟单层覆盖集中信息单元的更新算法, 以分析信息单元的变化情况, 分别构建近似集中可靠单元和争议单元的相关可靠单元集的更新算法。在此基础上, 设计与可靠单元和争议单元更新算法相对应的增量更新算法, 并且分析其时间复杂度。在UCI数据集上的实验结果表明, 与静态算法相比, 该算法在对象集发生添加和移除情况下的近似集更新效率分别提高21.5和29倍, 能够有效提高近似集的计算效率。

关键词: 粗糙集; 拟单层覆盖; 集值信息系统; 增量学习; 近似集

开放科学(资源服务)标志码(OSID):



中文引用格式: 吴正江, 张亚宁, 张真, 等. 拟单层覆盖粗糙集中近似集的增量更新算法[J]. 计算机工程, 2022, 48(6): 200-206, 212.

英文引用格式: WU Z J, ZHANG Y N, ZHANG Z, et al. Incremental updating algorithm for approximation sets on semi-monolayer cover rough sets[J]. Computer Engineering, 2022, 48(6): 200-206, 212.

Incremental Updating Algorithm for Approximation Sets on Semi-Monolayer Cover Rough Sets

WU Zhengjiang, ZHANG Yaning, ZHANG Zhen, MEI Qiuyu, YANG Tian

(School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan 454003, China)

[Abstract] A one-to-one mapping relationship exists between the semi-monolayer cover rough sets and the set-valued information system. When an object set in a set-valued information system is dynamically added or removed, the information unit that corresponds to the semi-monolayer cover rough sets also changes. This leads to the change in the approximation sets on the semi-monolayer cover rough sets. To address the dynamic updating problem of approximation sets on semi-monolayer cover rough sets, an incremental updating algorithm for approximation sets is proposed by combining semi-monolayer cover rough sets with incremental learning. The updating algorithm, comprising a centralized information unit with semi-monolayer coverage, is designed to analyze the changes in the information unit. The updating algorithms of relevant reliable unit sets of approximation centralized reliable and controversial units are constructed. An incremental updating algorithm that corresponds to the updating algorithms of reliable and controversial units is designed, and then its time complexity is analyzed. The experimental results on the UCI dataset indicate that compared with the static algorithm, the update efficiency of the approximation sets with the addition and removal of the object sets is improved by 21.5 and 29 times, respectively. Therefore, the incremental updating algorithm can effectively improve the computational efficiency of the approximation sets.

[Key words] rough sets; semi-monolayer cover; set-valued information system; incremental learning; approximation sets

DOI: 10.19678/j.issn.1000-3428.0062982

0 概述

经典的粗糙集^[1]在没有任何先验知识的情况下, 通过上下近似集表示某个确定的概念, 能够处理具有不确定性、不一致性等特点的数据集。粗糙集理论广泛应用于数据挖掘^[2]、推荐系统^[3]、工业控制

系统^[4-5]等领域。

信息系统可能伴有部分缺失值或集值。研究人员提出了容差关系^[6]、非对称相似关系^[7]、限制性容差关系^[8]、最大相容类^[9]等二元关系代替不可区分关系, 使得泛化的粗糙集模型能够处理集值信息系统。文献[10-11]提出拟单层覆盖粗糙集界于一般覆盖

基金项目: 国家自然科学基金(61972134, 11601129)。

作者简介: 吴正江(1981—), 男, 副教授、博士, 主研方向为粗糙集、粒计算; 张亚宁、张真、梅秋雨、杨天, 硕士研究生。

收稿日期: 2021-10-18 修回日期: 2021-11-23 E-mail: wuzhengjiang@hpu.edu.cn

与划分之间,是一个特殊的邻域系统。近似质量是衡量一个模型的标准,文献[12]将拟单层覆盖粗糙集应用于集值信息系统中,并在真实数据集上进行实验,证明了该模型在近似质量和计算效率方面均优于容差关系、非对称相似关系、限制性容差关系以及最大相容类,但是该模型无法应用于动态集值信息系统。

随着时间的推移,信息系统也会持续不断地发生变化。求解近似集的效率将直接影响规则提取和属性约简的效率。当信息系统发生改变时,快速获取更新系统中的近似集成为亟待解决的难题。

增量学习是指充分利用已知的信息并且避免从头开始计算,从而达到提升计算效率的目的。将增量学习灵活运用于动态信息系统中近似集的求解可以显著提升其计算效率。信息系统结构的动态变化方式有属性集的变化^[13-15]、属性值的变化^[16-17]以及对象集的变化。关于对象集发生变化的情况,文献[18]基于模糊优势邻域粗糙集提出动态区间值有序数据的增量特征选择方法。根据云平台下的并行模型 MapReduce,文献[19]提出经典粗糙集的并行增量算法,用于更新大规模数据的近似集。当对象批量发生变化时,文献[20]提出一种邻域决策粗糙集模型的增量式更新算法。针对混合型信息系统,文献[21]提出基于邻域决策粗糙集矩阵方法的增量式更新算法。文献[22]提出邻域多粒度粗糙集模型的矩阵化方法并设计了相应的增量更新方法,用于更新正域、负域及其边界域。文献[23]研究了优势粗糙集模型中动态有序数据的增量属性约简方法。

本文提出拟单层覆盖粗糙集中近似集的增量更新算法。当一个对象集添加至原始系统时或一个对象集从原始系统移除时,通过分析拟单层覆盖集中信息单元的变化情况,根据信息单元的变化对各近似集可靠单元和争议单元的相关可靠单元集的影响,设计相应的更新算法。在此基础上,通过计算更新系统中各近似集的最终结果,从而提高近似集的计算效率。

1 基本概念

集值信息系统 $S = (U, A, V, f)$ 是一个四元组。其中: U 是一个非空有限的对象集,称为论域; A 是有限的属性集; V 为属性的值域且 $V = \bigcup_{a \in A} V_a$; $f: U \times A \rightarrow 2^{V_a}$ 是从 $U \times A$ 到 V 的集值映射。

定义 1 令 $S = (U, A, V, f)$ 为集值信息系统,其中 $A = (a_1, a_2, \dots, a_n)$ 。对象 $x \in U$ 的信息解释是一个集值向量。表达式 $\vec{x} = \langle f(x, a_1), f(x, a_2), \dots, f(x, a_n) \rangle$, 其中 $f(x, a_i)$ 是集合。

定义 2 令 $S = (U, A, V, f)$ 为集值信息系统。 $\text{Cell}_x = \{y \in U | \vec{x} = \vec{y}\}$ 是 S 上的一个信息单元,且 $x \in \text{Cell}_x$, $\vec{x} = \overrightarrow{\text{Cell}_x}$ 。如果 $\overrightarrow{\text{Cell}_x}$ 中的任意值均为单值,则该信息单元被称为可靠单元。如果 $\overrightarrow{\text{Cell}_x}$ 中存在集值,则该信息单元被称为争议单元。 Cell_c 的相关可靠单元集记为

$\text{RS}(\text{Cell}_c)$, 其中 $\text{RS}(\text{Cell}_c) = \{\text{Cell}_r \in \text{RC} | \forall a_i \in A, x \in \text{Cell}_x, y \in \text{Cell}_r, f(x, a_i) \subseteq f(y, a_i)\}$ 。

可靠单元和争议单元分别记为 Cell_r 和 Cell_c , 并且所有可靠单元和争议单元的集合分别记为 RC 和 CC 。

定理 1 令 $S = (U, A, V, f)$ 为集值信息系统。 RC 和 CC 分别包含 S 中所有的可靠单元和争议单元。对应任意 $X \subseteq U$, S 上 X 的 DA0 和 DE0 近似集如下:

$$\underline{C}_{\text{DE0}}(X) = (\text{Cell}_r \in \text{RC} | \text{Cell}_r \in \underline{C}_{\text{GC0}}(X)) \cup (\text{Cell}_c \in \text{CC} | \text{RS}(\text{Cell}_c) \cap \underline{C}_{\text{GC0}}(X) \neq \emptyset)$$

$$\bar{C}_{\text{DA0}}(X) = (\text{Cell}_r \in \text{RC} | \text{Cell}_r \in \bar{C}_{\text{GC0}}(X)) \cup (\text{Cell}_c \in \text{CC} | \text{RS}(\text{Cell}_c) \subseteq \bar{C}_{\text{GC0}}(X))$$

$$\bar{C}_{\text{DE0}}(X) = (\text{Cell}_r \in \text{RC} | \text{Cell}_r \in \bar{C}_{\text{GC0}}(X)) \cup (\text{Cell}_c \in \text{CC} | \text{RS}(\text{Cell}_c) \cap \bar{C}_{\text{GC0}}(X) \neq \emptyset)$$

其中: $\underline{C}_{\text{GC0}}(X) = \{\text{Cell}_r \in \text{RC} | \text{Cell}_r \subseteq X\}$; $\bar{C}_{\text{GC0}}(X) = \{\text{Cell}_r \in \text{RC} | \text{Cell}_r \cap X \neq \emptyset\}$ 。

2 近似集的增量更新

当论域发生变化时,拟单层覆盖粗糙集中的近似集也会发生变化。传统的静态方法将从头开始计算近似集,会浪费大量的时间。本文提出当对象集变化时拟单层覆盖粗糙集的增量更新方法,充分利用已知的计算结果,达到提高计算效率的目的。

2.1 对象增加时近似集动态更新

令 $S^{t_0} = (U^{t_0}, A, V, f)$ 为原始集值信息系统, $S^A = (\Delta U, A, V, f)$ 为新增集值信息系统,其中 U^{t_0} 为原始对象集, ΔU 为新增对象集。假设系统 S^{t_0} 和 S^A 中的信息单元分别记为 $\text{Cell}^{t_0} \in \text{CELL}^{t_0}$ 和 $\text{Cell}^A \in \text{CELL}^A$, 系统 S^{t_0} 中争议单元的相关可靠单元集记为 $\text{RS}(\text{Cell}_c^{t_0})$ 。当一个对象集 ΔU 被添加到系统 S^{t_0} 中,形成更新后的集值信息系统 $S^{t_1} = (U^{t_0} \cup \Delta U, A, V, f)$, 分别记系统 S^{t_1} 的信息单元及其相关可靠单元集为 $\text{Cell}^{t_1} \in \text{CELL}^{t_1}$ 和 $\text{RS}(\text{Cell}_c^{t_1})$ 。同时, X 可能增加 ΔX 至 X' , 其中 $\Delta X \subseteq \Delta U$ 。因此,在对象集增加过程中,系统 S^{t_0} 的近似集将由 $\underline{C}_{\text{DA0}}^{t_0}(X)$, $\underline{C}_{\text{DE0}}^{t_0}(X)$, $\bar{C}_{\text{DA0}}^{t_0}(X)$ 和 $\bar{C}_{\text{DE0}}^{t_0}(X)$ 变化为 $\underline{C}_{\text{DA0}}^{t_1}(X')$, $\underline{C}_{\text{DE0}}^{t_1}(X')$, $\bar{C}_{\text{DA0}}^{t_1}(X')$ 和 $\bar{C}_{\text{DE0}}^{t_1}(X')$ 。

定理 2 令 $S^{t_0} = (U^{t_0}, A, V, f)$ 为原始集值信息系统, $S^A = (\Delta U, A, V, f)$ 为新增集值信息系统,其中 U^{t_0} 为原始对象集, ΔU 为新增对象集。对于任意一个 $\text{Cell}^A \in \text{CELL}^A$, 如果存在 $\text{Cell}^{t_0} \in \text{CELL}^{t_0}$, $\text{Cell}^A \in \text{CELL}^A$ 并且 $\overrightarrow{\text{Cell}^A} = \overrightarrow{\text{Cell}^{t_0}}$, 则在更新系统 S^{t_1} 中必然存在 $\overrightarrow{\text{Cell}^{t_1}} = \overrightarrow{\text{Cell}^A} = \overrightarrow{\text{Cell}^{t_0}}$ 且 $\text{Cell}^{t_1} = \text{Cell}^{t_0} \cup \text{Cell}^A$ 。如果存在 $\text{Cell}^{t_0} \in \text{CELL}^{t_0}$, 并且任意 $\text{Cell}^A \in \text{CELL}^A$ 使得 $\overrightarrow{\text{Cell}^{t_0}} \neq \overrightarrow{\text{Cell}^A}$ 成立, 则更新系统 S^{t_1} 中必然存在 $\overrightarrow{\text{Cell}^{t_1}} = \overrightarrow{\text{Cell}^{t_0}}$ 且 $\text{Cell}^{t_1} = \text{Cell}^{t_0}$ 。如果存在 $\text{Cell}^A \in \text{CELL}^A$, 并且不存在 $\text{Cell}^{t_0} \in \text{CELL}^{t_0}$ 使得 $\overrightarrow{\text{Cell}^A} = \overrightarrow{\text{Cell}^{t_0}}$ 成立, 则 $\text{Cell}^{t_1} = \text{Cell}^A$ 。

证明 当 S^{t_0} 和 S^A 中均存在 $\text{Cell}^{t_0} \in \text{CELL}^{t_0}$, $\text{Cell}^A \in \text{CELL}^A$ 使得 $\exists \text{Cell}^{t_1} \in \text{CELL}^{t_1}, \overrightarrow{\text{Cell}^A} = \overrightarrow{\text{Cell}^{t_1}}, \text{Cell}^{t_0} = \overrightarrow{\text{Cell}^{t_1}}$ 成立。根据定义 2 中 Cell_x 的定义, $\text{Cell}^{t_1} = \text{Cell}^{t_0} \cup \text{Cell}^A$; 当仅 S^{t_0} 中存在 $\text{Cell}^{t_0} \in \text{CELL}^{t_0}$ 使得 $\exists \text{Cell}^{t_1} \in \text{CELL}^{t_1}, \overrightarrow{\text{Cell}^A} = \overrightarrow{\text{Cell}^{t_1}}$ 成立, 则 $\text{Cell}^{t_1} = \text{Cell}^{t_0}$; 同理第 3 种情况可证得。

定理 3 令 $S^0 = (U^0, A, V, f)$ 为原始集值信息系统, $S^A = (\Delta U, A, V, f)$ 为新增集值信息系统, 其中 S^0 和 S^A 中的信息单元为 Cell^0 和 Cell^A 。对于 $\forall X$, 当 S^A 中的所有对象被添加至 S^0 中时, 假设 X 增加 ΔX 至 X' , 主要有 6 种情况: 1) 如果 $\text{Cell}^0 \in \underline{C}_{\text{GC0}}^{t_0}(X)$, 且 $\exists \text{Cell}^A$ 使得 $\overrightarrow{\text{Cell}}^A = \overrightarrow{\text{Cell}}^0$ 和 $\text{Cell}^A \subseteq \Delta X$ 成立, 则 $\text{Cell}^0 \cup \text{Cell}^A = \text{Cell}^{t_1} \in \underline{C}_{\text{GC0}}^{t_1}(X')$; 2) 如果 $\text{Cell}^0 \in \underline{C}_{\text{GC0}}^{t_0}(X)$, 且 $\nexists \text{Cell}^A$ 使得 $\overrightarrow{\text{Cell}}^A = \overrightarrow{\text{Cell}}^0$ 成立, 则 $\text{Cell}^0 = \text{Cell}^{t_1} \in \underline{C}_{\text{GC0}}^{t_1}(X')$; 3) 如果 $\text{Cell}^A \in \text{CELL}^A$, $\text{Cell}^A \subseteq \Delta X$, 且 $\nexists \text{Cell}^0$ 使得 $\overrightarrow{\text{Cell}}^A = \overrightarrow{\text{Cell}}^0$ 成立, 则 $\text{Cell}^A = \text{Cell}^{t_1} \in \underline{C}_{\text{GC0}}^{t_1}(X')$; 4) 如果 $\text{Cell}^0 \in \bar{C}_{\text{GC0}}^{t_0}(X)$, 且 $\exists \text{Cell}^A$ 使得 $\overrightarrow{\text{Cell}}^A = \overrightarrow{\text{Cell}}^0$, 则 $\text{Cell}^0 \cup \text{Cell}^A = \text{Cell}^{t_1} \in \bar{C}_{\text{GC0}}^{t_1}(X')$; 5) 如果 $\text{Cell}^0 \in \bar{C}_{\text{GC0}}^{t_0}(X)$, 且 $\nexists \text{Cell}^A$ 使得 $\overrightarrow{\text{Cell}}^A = \overrightarrow{\text{Cell}}^0$ 成立, 则 $\text{Cell}^0 = \text{Cell}^{t_1} \in \bar{C}_{\text{GC0}}^{t_1}(X')$; 6) 如果 $\text{Cell}^A \in \text{CELL}^A$, $\text{Cell}^A \cap \Delta X \neq \emptyset$, 且 $\nexists \text{Cell}^0$ 使得 $\overrightarrow{\text{Cell}}^A = \overrightarrow{\text{Cell}}^0$ 成立, 则 $\text{Cell}^A = \text{Cell}^{t_1} \in \bar{C}_{\text{GC0}}^{t_1}(X')$ 。

证明 关于 S^A 中 $\underline{C}_{\text{GC0}}^{t_1}(X')$ 的计算, 本文对前 3 种情况进行讨论。因为第 1 种情况满足定理 2 中的第 1 种情况且 $\text{Cell}^0 \in \underline{C}_{\text{GC0}}^{t_0}(X)$, 所以 $\text{Cell}^{t_1} = \text{Cell}^0 \cup \text{Cell}^A$, $\text{Cell}^0 \subseteq X$ 。由于 $\text{Cell}^A \subseteq \Delta X$, $X' = X \cup \Delta X$, 因此 $\text{Cell}^{t_1} \subseteq X'$, $\text{Cell}^{t_1} \in \underline{C}_{\text{GC0}}^{t_1}(X')$ 。同理, 第 2 种和第 3 种情况可证。

关于 S^A 中 $\bar{C}_{\text{GC0}}^{t_1}(X')$ 的计算, 本文对后 3 种情况进行讨论。如果 $\text{Cell}^0 \in \bar{C}_{\text{GC0}}^{t_0}(X)$, 则 $\text{Cell}^0 \cap X \neq \emptyset$ 。因为 $\exists \text{Cell}^A$ 使得 $\overrightarrow{\text{Cell}}^A = \overrightarrow{\text{Cell}}^0$ 成立, 所以 $\text{Cell}^{t_1} = \text{Cell}^0 \cup \text{Cell}^A$ 且 $\text{Cell}^{t_1} \cap X' \neq \emptyset$ 。因此 $\text{Cell}^{t_1} \in \bar{C}_{\text{GC0}}^{t_1}(X')$ 。同理, 第 5 种和第 6 种情况可证。

由定理 1 可知, 如果 $\text{Cell}_r \in \underline{C}_{\text{GC0}}^{t_0}(X)$, 则 $\text{Cell}_r \in \underline{C}_{\text{DA0(DE0)}}^{t_0}(X)$; 如果 $\text{Cell}_r \in \bar{C}_{\text{GC0}}^{t_0}(X)$, 则 $\text{Cell}_r \in \bar{C}_{\text{DA0(DE0)}}^{t_0}(X)$ 。因此, 定理 2 解决了可靠单元的更新问题。当集值信息系统中的对象增加时, 由于可靠单元的增加, 因此更新后的系统 S^{t_1} 中的 $\text{RS}(\text{Cell}_c^{t_1})$ 可由原始系统 S^0 中的 $\text{RS}(\text{Cell}_c^0)$ 更新得到。根据定理 2, 本文通过原始系统和新增系统中的信息单元更新获取更新系统的信息单元及其信息解释。

定理 4 令 $\text{RS}(\text{Cell}_c^0)$ 为原始系统 S^0 中争议单元的相关可靠单元集, 且 CELL^A 分为 RC^A 和 CC^A , 其中 $\text{Cell}_r^A \in \text{RC}^A$ 和 $\text{Cell}_c^A \in \text{CC}^A$ 为系统内 S^A 中的可靠单元集和争议单元集。如果 $\exists \text{Cell}_c^0 \in \text{CC}^0$ 且 $\overrightarrow{\text{Cell}}_c^0 = \overrightarrow{\text{Cell}}_c^{t_1}$, 则 $\text{RS}(\text{Cell}_c^{t_1}) = (\text{Cell}_r^{t_1} \in \text{RC}^{t_1} | \text{Cell}_r^{t_1} \in \text{RS}(\text{Cell}_c^{t_1}), \exists \text{Cell}_r^{t_1} \in \text{RC}^{t_1}, \overrightarrow{\text{Cell}}_r^0 = \overrightarrow{\text{Cell}}_r^{t_1}) \cup (\text{Cell}_r^{t_1} \in \text{RC}^{t_1} | \exists \text{Cell}_r^A, \overrightarrow{\text{Cell}}_r^A = \overrightarrow{\text{Cell}}_r^{t_1} \wedge \nexists \text{Cell}_r^0, \overrightarrow{\text{Cell}}_r^0 = \overrightarrow{\text{Cell}}_r^{t_1} \wedge \forall a_i \in A, x \in \text{Cell}_r^A, y \in \text{Cell}_r^{t_1}, f(x, a_i) \subseteq f(y, a_i))$; 如果 $\nexists \text{Cell}_c^0 \in \text{CC}^0$ 且 $\overrightarrow{\text{Cell}}_c^0 = \overrightarrow{\text{Cell}}_c^{t_1}$, 则 $\text{RS}(\text{Cell}_c^{t_1}) = (\text{Cell}_r^{t_1} \in \text{RC}^{t_1} | \forall a_i \in A, a_i \in A, x \in \text{Cell}_r^{t_1}, y \in \text{Cell}_c^{t_1}, f(x, a_i) \subseteq f(y, a_i))$ 。

证明 当 $\exists \text{Cell}_c^0 \in \text{CC}^0$ 且 $\overrightarrow{\text{Cell}}_c^0 = \overrightarrow{\text{Cell}}_c^{t_1}$ 时, 已知 $\text{RS}(\text{Cell}_c^0)$, 即原始系统 S^0 中符合 $\text{RS}(\text{Cell}_c)$ 定义的可靠单元已经得到。本文仅需计算新增系统中符合 $\text{RS}(\text{Cell}_c)$ 定义的可靠单元。由定理 2 可知, 如果 $\exists \text{Cell}_r^A, \overrightarrow{\text{Cell}}_r^A = \overrightarrow{\text{Cell}}_r^{t_1}$ 且 $\nexists \text{Cell}_r^0$ 使得 $\overrightarrow{\text{Cell}}_r^0 = \overrightarrow{\text{Cell}}_r^{t_1}$ 成立, 则 $\text{Cell}_r^A = \text{Cell}_r^{t_1}$ 。因此, $\forall a_i \in A, x \in \text{Cell}_r^A, y \in \text{Cell}_c^{t_1}, f(x, a_i) \subseteq f(y, a_i)$ 等

价于 $\forall a_i \in A, x \in \text{Cell}_r^{t_1}, y \in \text{Cell}_c^{t_1}, f(x, a_i) \subseteq f(y, a_i)$ 。根据 $\text{RS}(\text{Cell}_c)$ 的定义可知, $\text{Cell}_r^{t_1} \in \text{RS}(\text{Cell}_c^{t_1})$ 。当 $\nexists \text{Cell}_c^0 \in \text{CC}^0, \overrightarrow{\text{Cell}}_c^0 = \overrightarrow{\text{Cell}}_c^{t_1}$ 时, 根据定义计算即可。

2.2 对象移除时近似集动态更新

令 $S^0 = (U^0, A, V, f)$ 为原始集值信息系统, $S^A = (\Delta U, A, V, f)$ 为被移除的集值信息系统, 其中 U^0 为原始对象集, ΔU 为被移除的对象集。假设系统 S^0 和 S^A 中的信息单元分别记为 $\text{Cell}^0 \in \text{CELL}^0$ 和 $\text{Cell}^A \in \text{CELL}^A$, S^0 中争议单元的相关可靠单元集记为 $\text{RS}(\text{Cell}_c^0)$ 。当一个对象集 ΔU 从 S^0 中移除时, 更新后的集值信息系统 $S^{t_1} = (U^0 - \Delta U, A, V, f)$ 将会形成, 记 S^{t_1} 的信息单元及其相关可靠单元集分别为 $\text{Cell}^{t_1} \in \text{CELL}^{t_1}$ 和 $\text{RS}(\text{Cell}_c^{t_1})$ 。同时, X 将被移除 ΔX 至 X' , 其中 $\Delta X \subseteq \Delta U$ 。因此, 在对象集移除过程中, S^0 中的近似集将由 $\underline{C}_{\text{DA0}}^{t_0}(X), \underline{C}_{\text{DE0}}^{t_0}(X), \bar{C}_{\text{DA0}}^{t_0}(X)$ 和 $\bar{C}_{\text{DE0}}^{t_0}(X)$ 变化为 $\underline{C}_{\text{DA0}}^{t_1}(X'), \underline{C}_{\text{DE0}}^{t_1}(X'), \bar{C}_{\text{DA0}}^{t_1}(X')$ 和 $\bar{C}_{\text{DE0}}^{t_1}(X')$ 。

定理 5 令 $S^0 = (U^0, A, V, f)$ 为原始集值信息系统, $S^A = (\Delta U, A, V, f)$ 为被移除的集值信息系统, 其中 U^0 为原始对象集, ΔU 为被移除的对象集。对于任意一个 $\text{Cell}^A \in \text{CELL}^A$, 存在 3 种情况: 1) 如果存在 $\text{Cell}^0 \in \text{CELL}^0, \overrightarrow{\text{Cell}}^A = \overrightarrow{\text{Cell}}^0$ 并且 $\text{Cell}^A = \text{Cell}^0$, 则在更新后的系统 S^{t_1} 中必然不存在 $\text{Cell}^{t_1} \in \text{CELL}^{t_1}$, 使得 $\overrightarrow{\text{Cell}}^{t_1} = \overrightarrow{\text{Cell}}^A = \overrightarrow{\text{Cell}}^0$ 成立; 2) 如果存在 $\text{Cell}^0 \in \text{CELL}^0, \overrightarrow{\text{Cell}}^A = \overrightarrow{\text{Cell}}^0$ 并且 $\text{Cell}^A \subsetneq \text{Cell}^0$, 则在更新后的系统 S^{t_1} 中存在 $\text{Cell}^0 - \text{Cell}^A = \text{Cell}^{t_1} \in \text{CELL}^{t_1}$, 使得 $\overrightarrow{\text{Cell}}^{t_1} = \overrightarrow{\text{Cell}}^A = \overrightarrow{\text{Cell}}^0$ 成立; 3) 如果存在 $\text{Cell}^0 \in \text{CELL}^0$ 但 $\nexists \text{Cell}^A \in \text{CELL}^A, \overrightarrow{\text{Cell}}^A = \overrightarrow{\text{Cell}}^0$, 则在更新后的系统 S^{t_1} 中存在 $\text{Cell}^0 = \text{Cell}^{t_1} \in \text{CELL}^{t_1}$, 使得 $\overrightarrow{\text{Cell}}^{t_1} = \overrightarrow{\text{Cell}}^0$ 成立。

证明 由于 Cell^A 是被移除的信息单元, 因此任意 Cell^A 一定存在 $\text{Cell}^0 \in \text{CELL}^0$ 满足 $\overrightarrow{\text{Cell}}^A = \overrightarrow{\text{Cell}}^0$ 。如果 $\text{Cell}^A = \text{Cell}^0$, 则说明该信息单元中的所有对象均被移除。因此该信息单元不存在 S^{t_1} 的 CELL^{t_1} 中。如果 $\text{Cell}^A \subsetneq \text{Cell}^0$, 则说明该信息单元中的部分对象均被移除。因此 $\exists \text{Cell}^{t_1} \in \text{CELL}^{t_1}$ 满足 $\text{Cell}^{t_1} = \text{Cell}^0 - \text{Cell}^A$ 且 $\overrightarrow{\text{Cell}}^{t_1} = \overrightarrow{\text{Cell}}^A = \overrightarrow{\text{Cell}}^0$ 。第 3 种情况说明 Cell^0 中的对象未发生改变, 因此 $\text{Cell}^0 = \text{Cell}^{t_1} \in \text{CELL}^{t_1}$ 且 $\overrightarrow{\text{Cell}}^{t_1} = \overrightarrow{\text{Cell}}^0$ 。

定理 6 令 $S^0 = (U^0, A, V, f)$ 为原始集值信息系统, $S^A = (\Delta U, A, V, f)$ 为被移除的集值信息系统, 其中 S^0 和 S^A 中的信息单元为 Cell^0 和 Cell^A 。对于 $\forall X$, 当 S^A 中的所有对象从 S^0 中移除时, 假设 X 移除 ΔX 至 X' , 存在以下 5 种情况: 1) 如果 $\exists \text{Cell}^A$ 使得 $\overrightarrow{\text{Cell}}^A = \overrightarrow{\text{Cell}}^0$ 且 $\text{Cell}^A = \text{Cell}^0$ 成立, 则 $\nexists \text{Cell}^{t_1} \in \text{CELL}^{t_1}$ 使得 $\text{Cell}^{t_1} \in \underline{C}_{\text{GC0}}^{t_1}(X')$ 和 $\text{Cell}^{t_1} \in \bar{C}_{\text{GC0}}^{t_1}(X')$ 成立; 2) 如果 $\text{Cell}^0 \in \underline{C}_{\text{GC0}}^{t_0}(X)$ 且 $\exists \text{Cell}^A$ 满足 $\overrightarrow{\text{Cell}}^A = \overrightarrow{\text{Cell}}^0$ 和 $\text{Cell}^A \subsetneq \text{Cell}^0$, 则 $\text{Cell}^0 - \text{Cell}^A = \text{Cell}^{t_1} \in \underline{C}_{\text{GC0}}^{t_1}(X')$; 3) 如果 $\text{Cell}^0 \in \underline{C}_{\text{GC0}}^{t_0}(X)$ 但 $\nexists \text{Cell}^A \in \text{CELL}^A, \overrightarrow{\text{Cell}}^A = \overrightarrow{\text{Cell}}^0$, 则 $\text{Cell}^0 = \text{Cell}^{t_1} \in \underline{C}_{\text{GC0}}^{t_1}(X')$; 4) 如果 $\text{Cell}^0 \in \bar{C}_{\text{GC0}}^{t_0}(X)$ 且 $\exists \text{Cell}^A$ 满足 $\overrightarrow{\text{Cell}}^A = \overrightarrow{\text{Cell}}^0$ 和 $\text{Cell}^A \subsetneq \text{Cell}^0$, 则 $\text{Cell}^{t_1} = \text{Cell}^0 - \text{Cell}^A$; 如果 $\text{Cell}^{t_1} \cap X \neq \emptyset$, 则 $\text{Cell}^{t_1} \in \bar{C}_{\text{GC0}}^{t_1}(X')$; 5) 如果 $\text{Cell}^0 \in \bar{C}_{\text{GC0}}^{t_0}(X)$ 但 $\nexists \text{Cell}^A \in \text{CELL}^A, \overrightarrow{\text{Cell}}^A = \overrightarrow{\text{Cell}}^0$ 且 $\text{Cell}^0 \cap X \neq \emptyset$, 则 $\text{Cell}^0 = \text{Cell}^{t_1} \in \bar{C}_{\text{GC0}}^{t_1}(X')$ 。

证明 由定理5可知,如果 $\exists \text{Cell}^A$ 使得 $\overrightarrow{\text{Cell}}^A = \overrightarrow{\text{Cell}}^{t_0}$ 且 $\text{Cell}^A = \text{Cell}^{t_0}$ 成立,则 S^{t_0} 中必然不存在 $\text{Cell}^{t_1} \in \text{CELL}^{t_1}$ 。因此,第1种情况可证得。如果 $\exists \text{Cell}^A$ 使得 $\overrightarrow{\text{Cell}}^A = \overrightarrow{\text{Cell}}^{t_0}$ 和 $\text{Cell}^A \subsetneq \text{Cell}^{t_0}$ 成立,则 $\exists \text{Cell}^{t_1} \in \text{CELL}^{t_1}$, $\text{Cell}^{t_1} = \text{Cell}^{t_0} - \text{Cell}^A$ 且 $\overrightarrow{\text{Cell}}^{t_1} = \overrightarrow{\text{Cell}}^A = \overrightarrow{\text{Cell}}^{t_0}$ 。因为 $\text{Cell}^{t_0} \in \underline{C}_{\text{GC0}}^{t_0}(X)$,所以 $\text{Cell}^{t_0} \subseteq X$ 。由于 $\text{Cell}^{t_0} \cap \Delta X = \text{Cell}^A$,因此 $\text{Cell}^{t_0} - \Delta X = \text{Cell}^{t_0} - \text{Cell}^A \subseteq X - \Delta X$,即 $\text{Cell}^{t_1} \subseteq X'$ 。因此 $\text{Cell}^{t_1} \in \underline{C}_{\text{GC0}}^{t_1}(X')$ 。同理,第3种情况可证。第4种和第5种情况根据定理1中 $\bar{C}_{\text{GC0}}(X)$ 的定义即可证得。

定理7 令 $\text{RS}(\text{Cell}_c^{t_0})$ 为原始系统 S^{t_0} 中争议单元的相关可靠单元集。如果 $\exists \text{Cell}_c^{t_0} \in \text{CC}^{t_0}$, $\overrightarrow{\text{Cell}}_c^{t_0} = \overrightarrow{\text{Cell}}_c^{t_1}$,则 $\text{RS}(\text{Cell}_c^{t_1}) = (\text{Cell}_r^{t_1} | \text{Cell}_r^{t_0} \in \text{RS}(\text{Cell}_c^{t_0}), \exists \text{Cell}_r^{t_1} \in \text{RC}^{t_1}, \overrightarrow{\text{Cell}}_r^{t_1} = \overrightarrow{\text{Cell}}_r^{t_0})$ 。

证明 由定理5可知,一些信息单元可能被移除,所以可靠单元也有可能被移除。如果 $\exists \text{Cell}_c^{t_0} \in \text{CC}^{t_0}$, $\overrightarrow{\text{Cell}}_c^{t_0} = \overrightarrow{\text{Cell}}_c^{t_1}$,则 $\text{Cell}_c^{t_0}$ 移除 ΔX 后得到 $\text{Cell}_c^{t_1}$ 不为空。同时,如果 $\exists \text{Cell}_r^{t_0} \in \text{RS}(\text{Cell}_c^{t_0})$ 且存在 $\text{Cell}_r^{t_1} \in \text{RC}^{t_1}$, $\overrightarrow{\text{Cell}}_r^{t_0} = \overrightarrow{\text{Cell}}_r^{t_1}$,则 $\text{Cell}_r^{t_0}$ 移除 ΔX 后得到 $\text{Cell}_r^{t_1}$ 不为空。根据 $\text{RS}(\text{Cell}_c^{t_0})$ 的定义可知, $\text{Cell}_r^{t_1} \in \text{RS}(\text{Cell}_c^{t_1})$ 。

3 本文算法

本文设计增量更新算法,算法1为静态算法,算法2和算法3分别为对象集增加时和减少时的增量更新算法。

3.1 静态算法

在定义2和定理1的理论基础上,本文设计相应的静态算法。

算法1 计算拟单层覆盖粗糙集中近似集的静态算法

输入 集值信息系统 $S = (U, A, V, f)$ 和目标概念 X

输出 $\underline{C}_{\text{DA0}}(X)$, $\underline{C}_{\text{DE0}}(X)$, $\bar{C}_{\text{DA0}}(X)$ 和 $\bar{C}_{\text{DE0}}(X)$

1.根据定义2构造信息单元并分为可靠单元和争议单元;
2.根据定理1,通过判断可靠单元与 X 之间的关系计算 $\underline{C}_{\text{GC0}}(X)$ 和 $\bar{C}_{\text{GC0}}(X)$;

3.根据 $\text{RS}(\text{Cell}_c)$ 的定义计算争议单元的相关可靠单元集;

4.根据定理1,计算 $\underline{C}_{\text{DA0}}(X)$, $\underline{C}_{\text{DE0}}(X)$, $\bar{C}_{\text{DA0}}(X)$ 和 $\bar{C}_{\text{DE0}}(X)$ 。

算法1主要分为4个步骤:1)构造信息单元的时间复杂度为 $O(|U|)$,其中 $|U|$ 为集值信息系统中包含对象的数目;2)计算 $\underline{C}_{\text{GC0}}(X)$ 和 $\bar{C}_{\text{GC0}}(X)$ 的时间复杂度为 $O(|\text{RC}|)$,其中 $|\text{RC}|$ 为可靠单元的数目;3)计算 $\text{RS}(\text{Cell}_c)$ 的时间复杂度为 $O(|\text{CC}| \times |\text{RC}|)$,其中 $|\text{CC}|$ 为争议单元的数目;4)计算 $\underline{C}_{\text{DA0}}(X)$, $\underline{C}_{\text{DE0}}(X)$, $\bar{C}_{\text{DA0}}(X)$ 和 $\bar{C}_{\text{DE0}}(X)$ 的时间复杂度为 $O(|\text{CELL}|)$,其中 $|\text{CELL}|$ 为信息单元的数目,即可靠单元与争议单元数目之和。由于信息单元是互不相交的对象集,因此其大小与整体对象个数相比可以忽略。算法1的整体时间复杂度为 $O(|U| + |\text{RC}| + |\text{CC}| \times |\text{RC}| + |\text{CELL}|) \approx O(|U| + |\text{CC}| \times |\text{RC}|)$ 。

3.2 对象集添加时近似集的增量更新算法

当原始系统中添加一个对象集时,根据2.1节提出的方法设计相应的增量更新算法。

算法2 添加一个对象集时动态更新拟单层覆盖粗糙集中近似集的增量算法

输入 原始集值信息系统 $S^{t_0} = (U^{t_0}, A, V, f)$ 中的信息单元 $\text{Cell}^{t_0} \in \text{CELL}^{t_0}$, $\underline{C}_{\text{GC0}}^{t_0}$, $\bar{C}_{\text{GC0}}^{t_0}$ 以及 $\text{RS}(\text{Cell}_c^{t_0})$,其中 $\text{Cell}_c^{t_0} \in \text{CC}^{t_0}$,新增集值信息系统 $S^A = (\Delta U, A, V, f)$,被添加的概念 ΔX 以及更新后的概念 X'

输出 $\underline{C}_{\text{DA0}}^{t_1}(X')$, $\underline{C}_{\text{DE0}}^{t_1}(X')$, $\bar{C}_{\text{DA0}}^{t_1}(X')$ 和 $\bar{C}_{\text{DE0}}^{t_1}(X')$

1.使用算法1中构造信息单元的方法计算 $\text{Cell}^A \in \text{CELL}^A$;
2.根据定理2,通过 S^{t_0} 和 S^A 中的信息单元形成更新系统中的信息单元,针对更新系统中的可靠单元,通过定理3判断其是否属于 $\underline{C}_{\text{GC0}}^{t_1}(X')$ 和 $\bar{C}_{\text{GC0}}^{t_1}(X')$;

3.根据定理4,通过判断原始系统中是否存在 $\text{Cell}_c^{t_0}$,使得更新系统中存在 $\text{Cell}_c^{t_1}$ 与其信息解释相等,采取不同的方式计算 $\text{RS}(\text{Cell}_c^{t_1})$;

4.根据定理1,计算 $\underline{C}_{\text{DA0}}^{t_1}(X')$, $\underline{C}_{\text{DE0}}^{t_1}(X')$, $\bar{C}_{\text{DA0}}^{t_1}(X')$ 和 $\bar{C}_{\text{DE0}}^{t_1}(X')$ 。

算法2主要分为4个步骤:1)构造新增系统中信息单元的时间复杂度为 $O(|\Delta U|)$;2)形成更新系统中的信息单元并计算 $\underline{C}_{\text{GC0}}^{t_1}(X')$ 和 $\bar{C}_{\text{GC0}}^{t_1}(X')$,仅需遍历 CELL^A ,其时间复杂度为 $O(|\text{CELL}^A|)$;3)如果原始系统中存在 $\text{Cell}_c^{t_0}$,使得更新系统中存在 $\text{Cell}_c^{t_1}$ 与其信息解释相等,则遍历 RC^A ,其时间复杂度为 $O(|\text{CC}^A| \times |\text{RC}^A|)$,否则需要遍历 RC^{t_1} ,其时间复杂度为 $O(|\text{CC}^A| \times |\text{RC}^{t_1}|)$;4)时间复杂度为 $O(|\text{CELL}^{t_1}|)$ 。由于算法2的整体时间复杂度为 $O(|\Delta U| + |\text{CELL}^A| + |\text{CC}^{t_0}| \times |\text{RC}^A| + |\text{CC}^A| \times |\text{RC}^{t_1}|) \approx O(|\Delta U| + |\text{CC}^{t_0}| \times |\text{RC}^A| + |\text{CC}^A| \times |\text{RC}^{t_1}|) < O(|\Delta U| + (|\text{CC}^{t_0}| + |\text{CC}^A|) \times |\text{RC}^{t_1}|) = O(|U^{t_1}| + |\text{CC}^{t_1}| \times |\text{RC}^{t_1}|)$,因此算法2的时间复杂度低于算法1的时间复杂度。

3.3 对象集移除时近似集的增量更新算法

当原始系统中移除一个对象集时,根据2.2节提出的方法设计相应的增量更新算法。

算法3 移除一个对象集时动态更新拟单层覆盖粗糙集中近似集的增量算法

输入 原始集值信息系统 $S^{t_0} = (U^{t_0}, A, V, f)$ 中的信息单元 $\text{Cell}^{t_0} \in \text{CELL}^{t_0}$, $\underline{C}_{\text{GC0}}^{t_0}$, $\bar{C}_{\text{GC0}}^{t_0}$ 以及 $\text{RS}(\text{Cell}_c^{t_0})$,其中 $\text{Cell}_c^{t_0} \in \text{CC}^{t_0}$,被移除的集值信息系统 $S^A = (\Delta U, A, V, f)$,被移除的概念 ΔX 以及更新后的概念 X'

输出 $\underline{C}_{\text{DA0}}^{t_1}(X')$, $\underline{C}_{\text{DE0}}^{t_1}(X')$, $\bar{C}_{\text{DA0}}^{t_1}(X')$ 和 $\bar{C}_{\text{DE0}}^{t_1}(X')$

1.使用算法1中构造信息单元的方法计算 $\text{Cell}^A \in \text{CELL}^A$;

2.根据定理5,计算更新系统中的信息单元,针对更新系统中的可靠单元,进一步通过定理6判断其是否属于 $\underline{C}_{\text{GC0}}^{t_1}(X')$ 和 $\bar{C}_{\text{GC0}}^{t_1}(X')$;

3.如果原始系统中存在 $\text{Cell}_c^{t_0}$,使得更新系统中存在 $\text{Cell}_c^{t_1}$ 且 $\overrightarrow{\text{Cell}}_c^{t_0} = \overrightarrow{\text{Cell}}_c^{t_1}$,则根据定理7的方法计算 $\text{RS}(\text{Cell}_c^{t_1})$;

4.根据定理1,计算 $\underline{C}_{\text{DA0}}^{t_1}(X')$, $\underline{C}_{\text{DE0}}^{t_1}(X')$, $\bar{C}_{\text{DA0}}^{t_1}(X')$ 和 $\bar{C}_{\text{DE0}}^{t_1}(X')$ 。

在算法3中步骤1的时间复杂度为 $O(|\Delta U|)$ 。步骤2计算更新系统中的信息单元和 $\underline{C}_{\text{GC0}}^{t_1}(X')$, $\bar{C}_{\text{GC0}}^{t_1}(X')$ 需遍历 CELL^A ,因此其时间复杂度为 $O(|\text{CELL}^A|)$ 。根据定理7,算法3计算 $\text{RS}(\text{Cell}_c^{t_1})$ 的时间复杂度为 $O(|\text{CC}^{t_1}|)$ 。算法3的步骤4时间复杂度为 $O(|\text{CELL}^{t_1}|)$ 。由于算法3的时间复杂度为 $O(|\Delta U| + |\text{CELL}^A| + |\text{CC}^{t_1}| + |\text{CELL}^{t_1}|)$ 且小于 $O(|U^{t_1}| + |\text{CC}^{t_1}| \times |\text{RC}^{t_1}|)$,因此算法3的时间复杂度低于算法1的时间复杂度。

本文通过对比静态算法和两个增量算法的时间复杂度可知,无论对象集被添加还是被移除,增量算法的时间复杂度总低于静态算法的时间复杂度。

4 实验结果与分析

本文通过在真实数据集上的一系列实验验证增量算法的有效性。在计算结果保持一致的前提下,本文计算拟单层覆盖粗糙集中近似集所消耗的时间,对比静态算法和增量算法的效率。

本文实验分为对象添加和对象移除两种情况。这两种情况下的对比实验均在UCI数据集上进行。数据集的具体描述如表1所示。

表1 数据集描述			
Table 1 Data sets description			
数据集	对象数	属性数	类
wdbc	569	30	2
Anuran calls	7 195	22	4
Magic Gamma Telescope	19 020	10	2
Letter recognition	20 000	16	26
Sensorless	58 509	48	11
Covertypes	581 012	54	7

本文对数据集进行预处理形成对应的集值信息系统,分别计算每个属性的最小值、最大值、平均数和中

位数,将其从小到大排列产生3个间隔。如果该列的某个属性值位于第1个间隔,则该属性值对应于单值记录。如果其位于第3个间隔,则该属性值对应于与前者不同的单值记录,否则该值对应于前两者组成的集值记录。

实验环境的操作系统为Windows 10,CPU为Intel® Core™ i7-9750H,内存为16 GB。本文使用Java编程语言在IDEA平台上实现静态和增量算法,其中Java虚拟机版本为JVM 1.8。

当对象集发生变化时,本文在保持原始数据集包含对象数不变的前提下,依次增加发生变化(被添加到原始系统或从原始系统中移除)的对象数,对比静态算法和增量算法的计算时间,以验证增量理论和对应算法的有效性。

对于一个对象集被添加至原始集值信息系统的情况,本文对数据集进行如下处理:1)取出前50%的对象作为原始数据;2)将后50%的对象平均划分为10份;3)依次将10%,20%,⋯,100%添加至原始数据。针对算法1和算法2,本文使用以上产生的10组添加对象数不同的数据进行实验。在不同的数据集上,当对象增加时算法1和算法2计算时间的对比如图1所示。

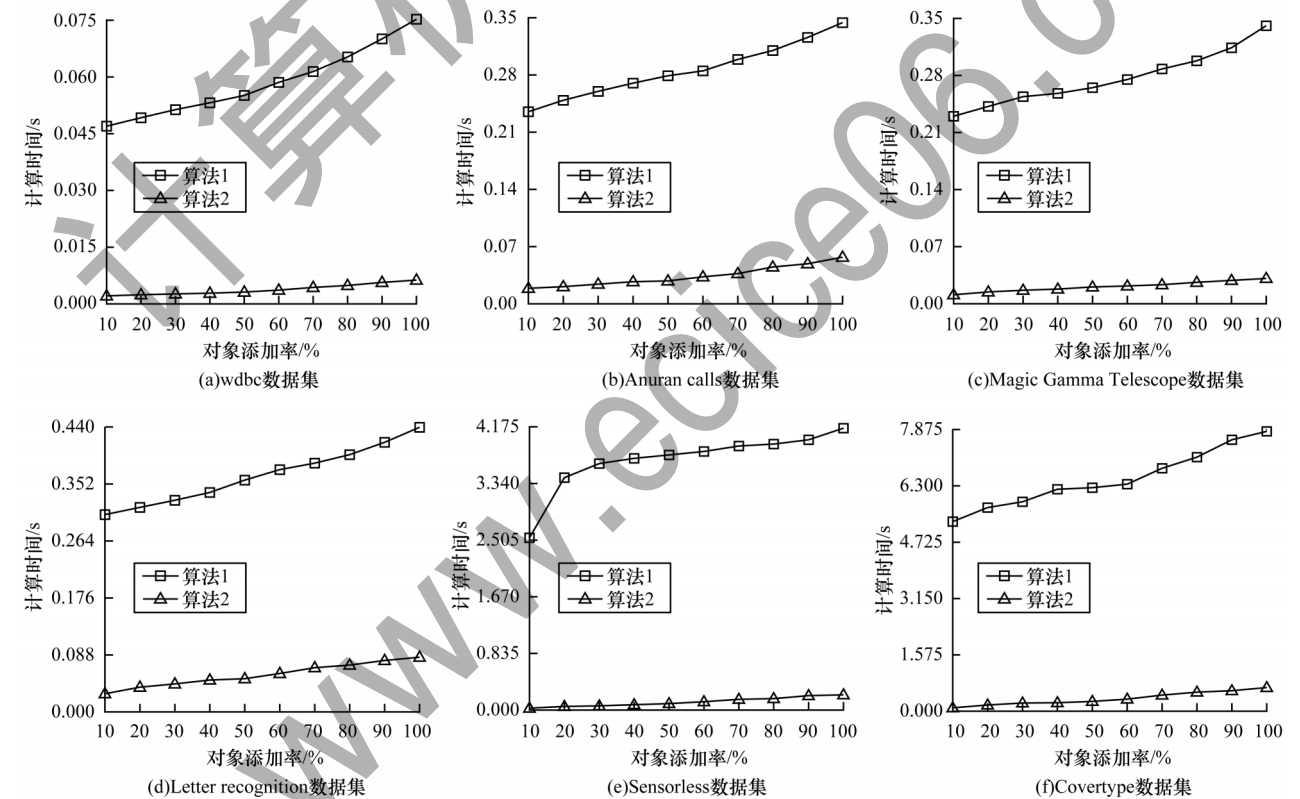


图1 当对象增加时算法1和算法2的计算时间对比

Fig.1 Computation time comparison of algorithm 1 and algorithm 2 when objects are added

从图1可以看出,随着添加至原始数据集中对象数目的增加,算法1和算法2的计算时间都呈增加趋势,但算法1对应曲线的斜率更大,并且计算时间比算法2多。因此,算法1的效率低于算法2。当数据集中包含对象增加时,信息单元(可靠单元和争议单元)的数目也会随之增加。根据算法1和算法2的时间复杂度可知,2个算法随着对象集的增加所需的计算时间也会增加,即图1的结果也与时间复杂度

的分析保持一致。

当对象添加率为10%、50%和100%时,静态算法1和增量算法2计算近似集所需运行时间的比值如表2所示。从表2可以看出,随着对象添加率的增加,算法1和算法运行时间的比值越来越小。在Sensorless数据集上,当对象添加率达到100%时,算法1的执行时间为4.155 s,而算法2仅需0.224 s,前者仍是后者的18倍。

表2 算法1与算法2运行时间的比值

Table 2 Running time ratio of algorithm 1 and algorithm 2

数据集	运行时间比值			平均值
	对象添加率 10%	对象添加率 50%	对象添加率 100%	
wdbc	22.00	17.48	11.93	17.13
Anuran calls	12.36	9.96	6.03	9.45
Magic Gamma Telescope	20.90	12.86	11.00	14.92
Letter recognition	10.82	6.98	5.20	7.66
Sensorless	90.71	39.58	18.00	49.61
Covertypes	55.80	22.81	11.83	30.14

对于一个对象集从原始集值信息系统中移除的情况,本文对数据集进行如下处理:1)将完整的数据集作为原始数据;2)将后50%的对象平均划分为10份;3)依次将10%,20%, \cdots ,100%从原始数据中移除。针对算法1和算法3,本文使用以上产生的10组移除对象数中不同的数据进行实验。在不同的数据集上,当对象移除率增加时算法1和算法3的计算时间对比如图2所示。从图2可以看出,随着移除对象的增加,算法1呈下降趋势,而算法3呈上升趋势,但是算法3的曲线总在算法1对应曲线的下方。因此,当论域中部分对象集移除时算法3的效率更高。

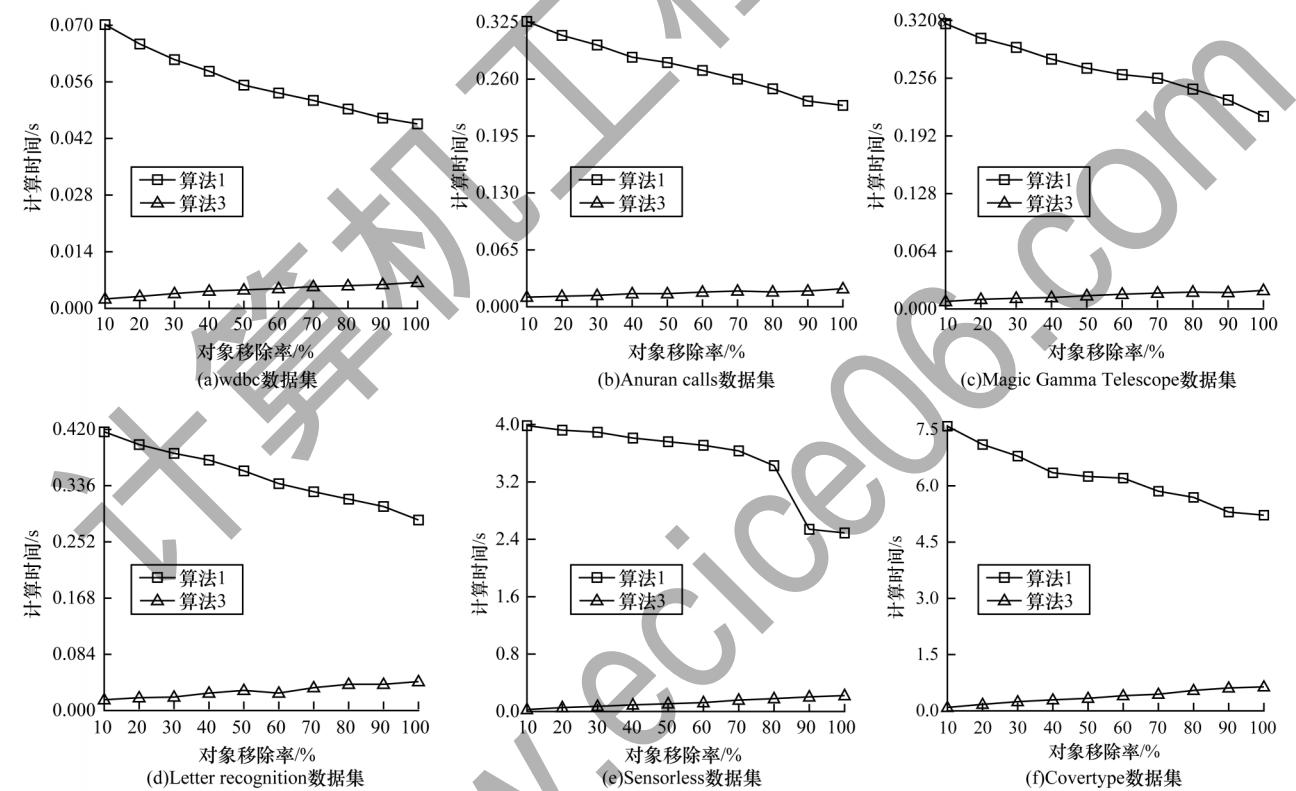


图2 当对象移除时算法1和算法3的计算时间对比

Fig.2 Computation time comparison of algorithm 1 and algorithm 3 when objects are removed

当对象移除率为10%、50%和100%时,静态算法1和增量算法3计算所需运行时间的比值如表3所示。

表3 算法1与算法3运行时间的比值

Table 3 Running time ratio of algorithm 1 and algorithm 3

数据集	不同运行时间比值			平均值
	对象移除率 10%	对象移除率 50%	对象移除率 100%	
wdbc	29.95	12.04	7.11	16.37
Anuran calls	29.63	18.60	11.16	19.80
Magic Gamma Telescope	37.83	18.15	10.39	22.12
Letter recognition	25.87	11.86	6.58	14.77
Sensorless	147.55	35.14	11.26	64.65
Covertypes	82.51	18.65	8.16	36.44

从表3可以看出,当对象移除率增加时,静态算法1和增量算法3计算所需运行时间的比值越来越小。在Sensorless数据集上,当对象移除率达到100%时,算法1的执行时间为2.489 s,而算法3仅需0.221 s,前者是后者的11.26倍。

在不同数据集上,对象移除率为10%和100%时 $|\Delta U|+|CELL^A|+|CC^U|+|CELL^U|$ 如表4所示。当在集值信息系统中移除一个对象集 ΔU 时,由于更新后系统中对象数 $|U^U|$ 和信息单元(可靠单元 RC^U 和争议单元 CC^U)的数目也随之减少,因此算法1的计算时间减少。当被移除对象集 ΔU 增大时, $|\Delta U|+|CELL^A|$ 变大, $|CC^U|+|CELL^U|$ 变小,但前者的增量大于后者的减量,因此在此过程中前者与后者总和是变大的。因此算法3呈上升趋势,并且图2的结果与时间复杂

度的分析保持一致。

表4 在不同对象移除率下 $|\Delta U|+|\text{CELL}^A|+|\text{CC}^C|+|\text{CELL}^L|$ 的结果

Table 4 Result of $|\Delta U|+|\text{CELL}^A|+|\text{CC}^C|+|\text{CELL}^L|$ under different object removal ratio

数据集	$ \Delta U + \text{CELL}^A + \text{CC}^C + \text{CELL}^L $	
	对象移除率为 10%	对象移除率为 100%
wdbc	697	895
Anuran calls	6 048	8 786
Magic Gamma Telescope	6 733	15 096
Letter recognition	14 361	22 496
Sensorless	75 196	94 716
Coverttype	85 769	349 697

5 结束语

因集值信息系统中的对象集随时间的推移而增加或移除,导致拟单层覆盖粗糙集中的近似集发生变化。本文结合增量学习与拟单层覆盖粗糙集,提出近似集的增量更新算法。通过设计信息单元、可靠单元和争议单元的更新方法,以达到提高计算效率的目的。构建与更新算法相对应的增量算法,并分析其时间复杂度。在UCI数据集上进行实验,结果表明,当对象集发生变化时,本文算法相较于静态算法的近似集计算效率高。下一步将拟单层覆盖粗糙集增量更新算法与大数据框架相结合,并对本文增量更新算法的并行化问题进行研究,使其能够实时处理海量数据。

参考文献

- [1] PAWLAK Z. Rough sets [J]. International Journal of Computer & Information Sciences, 1982, 11(5): 341-356.
- [2] 李抒音,刘洋. 权重模糊粗糙集的分类规则挖掘算法[J]. 计算机工程, 2019, 45(9): 211-215.
LI S Y, LIU Y. Classification rule mining algorithm for weighted fuzzy rough sets[J]. Computer Engineering, 2019, 45(9): 211-215. (in Chinese)
- [3] 姜宇,张大方,刁祖龙. 基于点击流的用户矩阵模型相似度个性化推荐[J]. 计算机工程, 2018, 44(1): 219-225.
JIANG Y, ZHANG D F, DIAO Z L. Similarity personalized recommendation of user matrix model based on click stream[J]. Computer Engineering, 2018, 44(1): 219-225. (in Chinese)
- [4] 王建华,陈永乐,张壮壮,等. 基于工控系统功能码特征的同源攻击分析[J]. 计算机工程, 2020, 46(7): 36-42.
WANG J H, CHEN Y L, ZHANG Z Z, et al. Same origin attack analysis based on features of industrial control system function code[J]. Computer Engineering, 2020, 46(7): 36-42. (in Chinese)
- [5] 王运明,陈波,周敏,等. 基于主客观权重判决的异构无线网络接入选择算法[J]. 计算机工程, 2019, 45(11): 107-111, 120.
WANG Y M, CHEN B, ZHOU M, et al. Heterogeneous wireless network access selection algorithm based on subjective and objective weight judgment[J]. Computer Engineering, 2019, 45(11): 107-111, 120. (in Chinese)
- [6] KRYSZKIEWICZ M. Rough set approach to incomplete information systems[J]. Information Sciences, 1998, 112(1): 39-49.
- [7] STEFANOWSKI J, TSOUKIAS A. Incomplete information tables and rough classification[J]. Computational Intelligence, 2001, 17(3): 545-566.
- [8] WANG G Y. Extension of rough set under incomplete information systems[C]//Proceedings of IEEE International Conference on Fuzzy Systems. Washington D. C., USA: IEEE Press, 2002: 1098-1103.
- [9] GUAN Y Y, WANG H K. Set-valued information systems[J]. Information Sciences, 2006, 176(17): 2507-2525.
- [10] WU Z J, CHEN N, GAO Y. Semi-monolayer cover rough set; concept, property and g-ranular algorithm[J]. Information Sciences, 2018, 456: 97-112.
- [11] 吴正江,刘永利,高岩. 拟单层覆盖上的覆盖粗糙集族[J]. 山东大学学报(理学版), 2014, 49(8): 6-14.
WU Z J, LIU Y L, GAO Y. Cover rough sets on a semi-monolayer cover[J]. Journal of Shandong University (Natural Science), 2014, 49(8): 6-14. (in Chinese)
- [12] WU Z J, WANG H, CHEN N, et al. Semi-monolayer covering rough set on set-valued information systems and its efficient computation[J]. International Journal of Approximate Reasoning, 2021, 130: 83-106.
- [13] HUANG Q Q, LI T R, HUANG Y Y, et al. Dynamic dominance rough set approach for processing composite ordered data[J]. Knowledge-Based Systems, 2020, 187: 1-17.
- [14] 吴杰,梁妍,马垣. 基于下确界不可约的概念格属性约简方法[J]. 计算机工程, 2016, 42(11): 213-218.
WU J, LIANG Y, MA Y. Attribute reduction method of concept lattice based on infimum irreducibility [J]. Computer Engineering, 2016, 42(11): 213-218. (in Chinese)
- [15] YANG X, LIU D, Yang X B, et al. Incremental fuzzy probability decision-theoretic approaches to dynamic three-way approximations[J]. Information Sciences, 2021, 550: 71-90.
- [16] HU C X, LIU S X, HUANG X L. Dynamic updating approximations in multigranulation rough sets while refining or coarsening attribute values[J]. Knowledge-Based Systems, 2017, 130: 62-73.
- [17] 刘凤玲,林国平. 动态更新属性值变化时的最优粒度[J]. 小型微型计算机系统, 2020, 41(10): 2063-2067.
LIU F L, LIN G P. Updating optimal scale in multi-scale decision systems under environment of attribute value[J]. Journal of Chinese Computer Systems, 2020, 41(10): 2063-2067. (in Chinese)
- [18] SANG B B, CHEN H M, YANG L, et al. Feature selection for dynamic interval-valued ordered data based on fuzzy dominance neighborhood rough set[J]. Knowledge-Based Systems, 2021, 227: 1-10.
- [19] 张钧波,李天瑞,潘毅,等. 云平台下基于粗糙集的并行增量知识更新算法[J]. 软件学报, 2015, 26(5): 1064-1078.
ZHANG J B, LI T R, PAN Y, et al. Parallel and incremental algorithm for knowledge update based on rough sets in cloud platform[J]. Journal of Software, 2015, 26(5): 1064-1078. (in Chinese)

(下转第212页)

(上接第 206 页)

- [20] 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 H X. 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. (in Chinese)
- [21] 苑红星, 卓雪雪, 竺德, 等. 基于矩阵的混合型邻域决策粗糙集增量式更新算法[J]. 控制与决策, 2022, 37(6): 1621-1631.
YUAN H X, ZHUO X X, ZHU D, et al. Incremental updating algorithms of neighborhood decision-theoretic rough set model for hybrid data based on matrix[J]. Control and Decision, 2022, 37(6): 1621-1631. (in Chinese)
- [22] 徐怡, 孙伟康, 王泉. 邻域多粒度粗糙集知识更新增量算法[J]. 小型微型计算机系统, 2020, 41(5): 908-918.
XU Y, SUN W K, WANG Q. Incremental algorithm for knowledge updating of neighborhood multigranulation rough set[J]. Journal of Chinese Computer Systems, 2020, 41(5): 908-918.
- [23] SANG B B, CHEN H M, YANG L, et al. Incremental attribute reduction approaches for ordered data with time-evolving objects[J]. Knowledge-Based Systems, 2021, 212: 1-13.

编辑 薛晋栋