Updating Approximations Dynamically in Dominance-based Rough Sets*

Shaoyong Li, Tianrui Li, Junbo Zhang

School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China
Email: meterer@163.com, trli@swjtu.edu.cn, JunboZhang86@163.com

Abstract

Dominance-based rough set (DBRS) approach is dedicated to analysis of data with ordinal properties by the dominance relation. Incrementally updating knowledge is a hot topic in data mining and granular computing. This paper presents algorithms for updating approximations dynamically in DBRS when adding or deleting objects in the universe. Experimental evaluation is employed to compare the efficiency of the proposed approaches with its traditional counterpart.

1. Introduction

Rough set theory [1][2][3][4], introduced by Pawlak in 1982, is the extension of classical set theory to process inconsistent information. It partitions the universe by an equivalence relation into equivalence classes. By the lower and upper approximations in rough set model, knowledge hidden in information system may be unraveled. However, original rough set can’t adapt to the case such that attributes with preference-ordered domains. To enable rough set model to solve those problems, Greco et al [5] proposed the dominance based rough set (DBRS) model. The innovation of DBRS is based on substitution of an equivalence relation by a dominance relation.

In recent years, many scholars had achieved significant results in DBRS theory and its applications. Fan [6] introduced the concept of a dominance degree by a fuzzy dominance relation. Yang et al [7][8][9] proposed several improved DBRS approaches to process incomplete information systems. Chan et al [10] developed a method that makes index blocks as granules to compute approximations in DBRS. Inuiguchi et al [11] introduced a variable-precision DBRS model. Dembczynski et al [12] proposed the second-order rough approximations in DBRS with respect to imprecise evaluation and assignment of objects on criteria. The DBRS model has been successfully applied in multiple criteria classification with imprecise evaluations and assignments [13] and customer behavior in the airline market [14].

In many real cases, information systems are always changing, including addition or deletion of objects, attributes and variation of attributes values. As we known, obtaining knowledge in dynamic information systems

*This work is supported by the National Science Foundation of China (No. 60873108).
based on rough set model has received much attention recently. Asharaf et al [15] proposed an incremental clustering approach based on rough set model. Blaszczynski et al [16] introduced an incremental approach of inducing decision rules from approximations in DBRS. Li et al [17] [18] studied the knowledge maintenance by using an incremental method in the characteristics relation based rough set model. Liu et al [19] studied the obtaining knowledge in dynamic information systems by an incremental method. Fan et al [20] studied rules induction by using an incremental approach based on rough sets. Li et al [21] introduced an incremental approach for dynamically updating set approximations in DBRS. This paper aims to evaluate the efficiency of this approach.

The remainder of the paper is organized as follows. Section 2 reviews the incremental approach for dynamically updating approximations in DBRS. Algorithms of the approaches are introduced in Section 3. Section 4 is the experimental evaluation. The paper ends with conclusions and further research topics in Section 5.

2. Preliminaries

Let us consider a decision table, \( T = (U, C \cup \{d\}, V, f) \), including a finite universe of objects \( U \) evaluated on a finite set of condition attributes \( C = \{c_1, \cdots, c_n\} \), and on a single decision attribute \( d \). Let \( A = C \cup \{d\} \). \( V = \cup_{a \in A} V_a \) is regarded as the domain of all attributes. The set of the indices of attributes is denoted by \( I = \{1, \cdots, n\} \). Without loss of generality, \( f_i : U \times A \rightarrow V \) for each \( i \in I \), and, for all objects \( x, y \in U \), \( f_i(x) \geq f_i(y) \) means that \( x \) is at least as good as \( y \) with respect to attribute \( i \), which is denoted by \( x \succeq_i y \). Therefore, it is supposed that \( \succeq_i \) is a complete preorder, \( i.e., \) a strongly complete and transitive binary relation, defined on \( U \). Furthermore, decision attribute \( d \) makes a partition of \( U \) into a finite number of decision classes, \( Cl = \{Cl_1, \cdots, Cl_m\} \), and not \( Cl \rightarrow U \), reads "\( \in \)". Let us remark that \( Cl \subseteq U \). \( Cl \supseteq U \). \( Cl \supseteq Cl \). Furthermore, for \( t = 2, \cdots, m \), \( Cl_{t-1} \subseteq U - Cl_t \). \( x \) dominates \( y \) with respect to set of attributes \( P \subseteq C \) (shortly, \( x P \)-dominates \( y \)), denoted by \( xDPy \), if for every attribute \( c_i \in P \), \( f_i(x) \geq f_i(y) \). The relation of \( P \)-dominance is reflexive and transitive, \( i.e., \) it is a partial preorder.

Given a set of attributes \( P \subseteq C \) and \( x \in U \), the granules of knowledge used for approximations in DBRS are:

- a set of objects dominating \( x \), called a \( P \)-dominating set,

\[
D^+_P(x) = \{y \in U; yDPx\}.
\]
• a set of objects dominated by \( x \), called a \( P \)-dominated set,
\[
D_P^+(x) = \{ y \in U; xD_P y \}. \tag{2}
\]

The \( P \)-lower approximation of \( C_l^+ \), denoted by \( P(C_l^+) \), and the \( P \)-upper approximation of \( C_l^+ \), denoted by \( \overline{P}(C_l^+) \), are defined as:
\[
P(C_l^+) = \{ x \in U; D_P^+(x) \subseteq C_l^+ \}, \tag{3}
\]
\[
\overline{P}(C_l^+) = \{ x \in U; D_P^+(x) \cap C_l^+ \neq \emptyset \}. \tag{4}
\]

Analogously, one can define the \( P \)-lower approximation and the \( P \)-upper approximation of \( C_l^- \) as:
\[
P(C_l^-) = \{ x \in U; D_P^-(x) \subseteq C_l^- \}, \tag{5}
\]
\[
\overline{P}(C_l^-) = \{ x \in U; D_P^-(x) \cap C_l^- \neq \emptyset \}. \tag{6}
\]

### 2.1. Addition of a new object

Suppose a single object \( x' \) enters the decision table \( T \). It will change the basic information granules, the upward union and downward union of decision classes and their lower, upper approximations [21].

**Theorem 2.1** Let \( x \in C_l^+ \) and \( d(x') = d(x) \), then
\[
C_l^+_{i'} = \begin{cases} C_l^+ \cup \{ x' \} : i \leq t \\ C_l^+_{i} : i > t \end{cases} \tag{7}
\]
\[
C_l^-_{i'} = \begin{cases} C_l^- \cup \{ x' \} : i \geq t \\ C_l^-_{i} : i < t \end{cases} \tag{8}
\]

where \( i, t \in T, T = \{ 1, \cdots, m \} \). \( C_l^+_{i'} \) and \( C_l^-_{i'} \) are respectively the upward and downward unions of the updated decision class \( C_l^+_{i} \).

**Theorem 2.2** For all \( x \in U \), then
\[
D_P^+(x) = \begin{cases} D_P^+(x) : x \not\in D_P^+(x') \\ D_P^+(x) \cup \{ x' \} : x \in D_P^+(x') \end{cases} \tag{9}
\]
\[
D_P^-(x) = \begin{cases} D_P^-(x) : x \not\in D_P^-(x') \\ D_P^-(x) \cup \{ x' \} : x \in D_P^-(x') \end{cases} \tag{10}
\]

where \( D_P^+(x) \) and \( D_P^-(x) \) denote the updated \( P \)-dominating and \( P \)-dominated sets of the object \( x \) respectively.

**Definition 2.1** Given a function \( \partial_P : U \to f(V_d), P \subseteq C \), that is to say
\[
\partial_P^+(x) = \{ i | i = d(y) \land y \in D_P^+(x) \}, \tag{11}
\]
\[
\partial_P^-(x) = \{ i | i = d(y) \land y \in D_P^-(x) \}. \tag{12}
\]
\( \partial_P^+ \) will be called as the decision domain of \( P \)-dominated set, and \( \partial_P^- \) will be called as the decision domain of \( P \)-dominating set.

**Theorem 2.3** Let \( i \in T, T = \{ 1, \cdots, m \} \) and \( P \subseteq C \), there are
\[
P(C_l^+_{i'} - \Omega = \{ x \in C_l^+ : i > \max(\partial_P^+(x')) \}, \tag{13}
\]
\[
\overline{P}(C_l^+_{i'}) = \{ x \in C_l^+ : i \leq \min(\partial_P^+(x')) \}, \tag{14}
\]
\[
\overline{P}(C_l^-_{i'}) = \{ x \in C_l^- : i > i_{\min} \}, \tag{15}
\]
\[
\overline{P}(C_l^-_{i'}) = \{ x \in C_l^- : i > i_{\max} \}, \tag{16}
\]

where \( \Omega = \{ y \in P(C_l^+_{i'}) \cap D_P^+(x') \land \max(\partial_P^+(y)) > i \}, \Omega = \{ y \in P(C_l^-_{i'}) \land D_P^+(x') \land \min(\partial_P^+(y)) < i \}, \max(\partial_P^+(\cdot)) \) and \( \min(\partial_P^+(\cdot)) \) are respectively the maximum and minimum of the decision domain. \( C_l^+_{i'} \) and \( C_l^-_{i'} \) are respectively the upward and downward unions of the updated decision class \( C_l^+_{i} \).

### 2.2. Deletion of an object

Suppose a single object \( \hat{x} \) is deleted from the decision table \( T \). It will change the basic information granules, the upward union and downward union of decision classes and their lower, upper approximations [21].

**Theorem 2.4** Suppose \( \hat{x} \in C_l^+ \). After the object \( \hat{x} \) is deleted from the universe \( U \), we have
\[
C_l^+_{i'} = \begin{cases} C_l^+ - \{ \hat{x} \} : i \leq t \\ C_l^+_{i} : i > t \end{cases} \tag{17}
\]
\[ C_T^i = \begin{cases} C_T^i - \{z\} & : i \geq t \\ C_T^i & : i < t \end{cases} \]  
(18)

where \( i, t \in T, T = \{1, \ldots, m\} \).

\( C_T^i \) and \( C_T^\leq \) are respectively the upward and downward unions of decision classes in the updated universe \( U' \).

**Theorem 2.5** Suppose \( U' = U - \{z\} \).
For all \( x_i \in U' \),

\[ D_P^i(x_i) = \begin{cases} D_P^i(x_i) : x_i \notin D_P^i(\bar{z}) \\ D_P^i(x_i) - \{z\} : x_i \in D_P^i(\bar{z}) \end{cases} \]  
(19)

\[ D_P^i(x_i) = \begin{cases} D_P^i(x_i) : x_i \notin D_P^i(\bar{z}) \\ D_P^i(x_i) - \{z\} : x_i \in D_P^i(\bar{z}) \end{cases} \]  
(20)

where \( i \in T, T = \{1, \ldots, m\} \).

**Theorem 2.6** Let \( i \in T, T = \{1, \ldots, m\} \) and \( P \subseteq C \), there are

\[ P(C_T^\geq) = \begin{cases} P(C_T^\geq) \cup B : i < \max(\partial_P(\bar{x})) \\ P(C_T^\geq) - \{z\} : i \geq \max(\partial_P(\bar{x})) \end{cases} \]  
(21)

\[ P(C_T^\leq) = \begin{cases} P(C_T^\leq) - \{z\} : i \notin \partial_P^+(\bar{x}) \\ P(C_T^\leq) : i \in \partial_P^+(\bar{x}) \end{cases} \]  
(22)

\[ P(C_T^\geq) = \begin{cases} P(C_T^\geq) \cup B : i > \min(\partial_P(\bar{x})) \\ P(C_T^\geq) - \{z\} : i \leq \min(\partial_P(\bar{x})) \end{cases} \]  
(23)

\[ P(C_T^\leq) = \begin{cases} P(C_T^\leq) - \{z\} : i \notin \partial_P^-(\bar{x}) \\ P(C_T^\leq) : i \in \partial_P^-(\bar{x}) \end{cases} \]  
(24)

where \( B = \{y | y \in P(C_T^\geq) \cap D_P^+(\bar{x}) \land y \notin P(C_T^\leq) \land \max(\partial_P(y)) \leq i\} \), \( B = \{y | y \in P(C_T^\geq) \cap D_P^+(\bar{x}) \land y \notin P(C_T^\leq) \land \min(\partial_P(y)) \geq i\} \).

3. **Algorithms for incrementally updating approximations in DBRS**

Base on the above definition and theorems, we design algorithms for incrementally updating approximations in DBRS while adding or deleting an object. (See Algorithms 1 and 2.)

---

**Algorithm 1** Updating approximations incrementally after the new object being added.

**Require:**
The decision table, \((U, C \cup D)\);
The decision class family, \(C\);
The approximations of decision classes, \(\overline{P}(C_{\bar{t}})\) and \(P(C_{\bar{t}})\), where \(i \in T, T = \{1, 2, \ldots, n\}\), \(T\) is the subscript family of decision classes.
The adding object, \(x'\).

**Ensure:**
The updated decision table, \((U, C \cup D)\);
The updated decision class family, \(C\);
The updated approximations of decision classes, \(\overline{P}(C_{\bar{t}})\) and \(P(C_{\bar{t}})\);
1: \(U \leftarrow U \cup \{x'\} \); // Update the universe
2: find the decision class \(C_{\bar{t}}\) which \(x'\) belongs to;
3: for each \(i \in T\) do
4: if \(i > t\) then
5: \(C_T^\leq \leftarrow C_T^\leq \cup \{x'\}\)
6: else
7: if \(i = t\) then
8: \(C_T^\leq \leftarrow C_T^\leq \cup \{x'\}\)
9: \(C_T^\geq \leftarrow C_T^\geq \cup \{x'\}\)
10: else
11: \(C_T^\leq \leftarrow C_T^\leq \cup \{x'\}\)
12: compute \(D_P^+(x')\) and \(D_P^-(x')\); // Dominated and dominating sets of \(x'\) with respect to \(P\), where \(P \subseteq C\)
13: compute \(\partial_P^+(x')\) and \(\partial_P^-(x')\); // Decision value domains of \(D_P^+(x')\) and \(D_P^-(x')\)
14: for each \(i \in T\) do
15: if \(i \geq \max(\partial_P(x'))\) then
16: \(P(C_T^\geq) \leftarrow P(C_T^\geq) \cup \{x'\}\)
17: if \(i \in \partial_P^+(x')\) then
18: \(P(C_T^\leq) \leftarrow P(C_T^\leq) \cup \{x'\}\)
19: if \(i \leq \min(\partial_P(x'))\) then
20: \(P(C_T^\geq) \leftarrow P(C_T^\geq) \cup \{x'\}\)
21: if \(i \in \partial_P^-(x')\) then
22: \(P(C_T^\leq) \leftarrow P(C_T^\leq) \cup \{x'\}\)
23: end.
Algorithm 2 Updating approximations incrementally after the object being deleted.

Require:
The decision table, \((U, C \cup D)\);
The decision class family, \(Cl\);
The approximations of decision classes, \(\overline{P}(Cl_i)\) and \(\underline{P}(Cl_i)\), where \(i \in T, T = \{1, 2, \ldots, n\}\), \(T\) is the subscript family of decision classes.
The number of the deleting object, \(M\);

Ensure:
The updated decision table, \((U, C \cup D)\);
The updated decision class family, \(Cl\);
The updated approximations of decision classes, \(\overline{P}(Cl_i)\) and \(\underline{P}(Cl_i)\);

1: find the deleting object \(x\) according to \(M\);
2: \(U \leftarrow U - \{x\}; \quad // Update the universe
3: find the decision class \(Cl_i\) which \(x\) belongs to;
4: for each \(i \in T\) do
5:   if \(i > t\) then
6:     \(Cl_i^x \leftarrow Cl_i^x - \{x\}\)
7:   else
8:     if \(i = t\) then
9:       \(Cl_i^x \leftarrow Cl_i^x - \{x\}\)
10:      \(Cl_i^x \leftarrow Cl_i^x - \{x\}\)
11:     else
12:       \(Cl_i^x \leftarrow Cl_i^x - \{x\}\)
13: compute \(D_P^+(\hat{x})\) and \(D_P^-(\hat{x})\);
14:end.

\[ D_P^+(\hat{x}) \]

4. Experimental evaluation

The aim of this work is to compare the efficiency of incremental method for updating set approximation in DBRS with its traditional counterpart. The data used in the experiments are outlined in Table 1, which are downloaded from UCI Repository of machine learning databases.

These algorithms are run on a personal computer with Windows XP and Inter(R) Core(TM)2 Duo E7500 2.93GHz and 3.21 GB memory. The software being used is Microsoft Visual C++ 6.0.

The strategies of the experiments are to add a case to the data set and delete a case from the data set. Specifically, we randomly choose a case in different size of data sets. In those experiments, we adopt high performance count time API, in order to avoid the error produced by CPU’s abnormal frequency, and adopt means of several experiments’ results as evaluators.

Figures 1 and 2 show the trendline of each of two algorithms with the size of data sets becoming increasing. Similarly, in each of these sub-figures (a)-(b), the x-coordinate pertains to the size of the data set (the 5 data sets starting from the smallest one), while the y-coordinate is the computing time. From Figures 1 and 2, it is easy to see that the incremental methods are consistently faster than its original counterpart. Furthermore the differences
are becoming larger when the size of the data set increases. It is validated that the incremental methods for updating approximations of unions of decision classes.

5. Conclusions

In this paper, we compared the efficiency of approaches for dynamically updating approximations in DBRS with its traditional counterpart. The experiments’ results validate incremental methods for updating approximations of unions of decision classes. However, when a decision table changes frequently, it is a question whether this method is efficient or not. Our future work will further investigate the incremental approaches for updating approximations in DBRS when multi-object changing at the same time.

References


