An Incremental Learning Algorithm Based on Rough Set Theory^{*}

Yinghong Ma^{**} and Yehong Han

School of Management, Shandong Normal University, Jinan, 250014, P.R. China
yinghongma71@163.com

Abstract. A decision table is a pair $S = (U, A)$, where U and A are nonempty finite sets called the universe and primitive attributes respectively. In order to computing the minimal rule sets on the decision table in which a new instance is added, a classification of the new instances and a criteria for the minimum recalculation are given, an incremental learning algorithm is presented and this algorithm is proved that can be used to the consistent and the inconsistent decision tables. The complexity of this algorithm is also obtained.

Keywords: rough sets, incremental learning, algorithm.

AMS Subject Classifications: TP18.

1 Introduction and Notations

Rough set theory was introduced by Z. Pawlak in 1982, it has been used in a variety of domains [1] [2] [3], such as machine learning, pattern recognition, decision support system and expert systems and so on. The incremental learning is a branch of machine learning. One of the main researches on rough sets theory is how to compute the attribute's reduction on a decision table, but those research is limited by the static data, but in real life, it is not such case, a new instance is allowed to put in S . Computing a minimal rule set on a decision table when the new instance is added, all the data in the decision table will be recalculated by the classical method [4]. In [5], there are three cases when a new instance x is added to S , x confirms the actual knowledge; x contradicts to the actual knowledge; x is completely new. But this category is not a partition of $S \cup \{x\}$. In [6], a new category method based on decision logical was introduced, unfortunately, it isn't including all the cases.

Knowledge representation system is a pair $S = \langle U, A \rangle$, where U and A are all nonempty finite sets called the universe and primitive attributes respectively. Every primitive attribute $a \in A$ is a total function $a : U \rightarrow V_a$, where V_a is the set of values of a , called the domination of a . Let $C, D \subset A$ be the condition and

^{*} This work is carried out under the Taishan Scholar project of Shandong China and the Doctor Foundation of Shandong(2006BS01015).

^{**} Corresponding author.

decision attributes of A respectively. A KR -system with distinguished condition and decision attributes will be viewed as a decision table, denoted by $S = (U, A)$, the decision table often is often called a decision logic. Using the decision logic language(DL-language), the following notations are defined.

The set of formulas in decision logic language is a set satisfying the following conditions: (1) The expression of the forms (a, v) (in short a_v called elementary formulas) are formulas of the DL -language for any $a \in A$ and $v \in V_a$; (2) If θ, ψ are formulas of the DL -language, then $\neg\theta, \theta \vee \psi, \theta \wedge \psi, \theta \rightarrow \psi$ and $\theta \equiv \psi$ are formulas too. If $x \in U$ satisfies formula θ in $S = (U, A)$, denoted by $x \models_S \theta$. If θ is a formula, then the set $|\theta|_S$, defined as $|\theta|_S = \{x \in U : x \models_S \theta\}$, is called the meaning of the formula θ in S . Let $P = \{a_1, a_2, \dots, a_n\}$ and $P \subseteq A$. Then $(a_1, v_1) \wedge (a_2, v_2) \wedge \dots \wedge (a_n, v_n)$ is called a P -basic formula. If θ is a P -basic formula and $R \subseteq P$, we denote the R -basic formula obtained from the formula θ by removing all elementary formulas (a, v_a) such that $a \in P - R$ by θ/R . The other notation and definition not be given here can be find in [5].

2 An Algorithm for Incremental Learning

The Category of New Instances. Denoted the new instance by x and the new decision table by S' , let M', R' be a minimal decision rule-set and a condition attribute reduction set of S' respectively. The new definitions and notations are given.

x is called matching M , if and only if all the rules $\theta \rightarrow \psi$ such that $\theta_x \rightarrow \theta$ in M imply $\psi_x \equiv \psi$, denote the set of thus x by I_m . The set $M_x = \{i | i \in U, \theta_x \rightarrow \theta_i, \psi_x \equiv \psi_i\}$ is called the matching domain of x , where $\theta_i \rightarrow \psi_i$ is the RD -rule reduction of object i . x is called totally contradicts to M if and only if all the rules $\theta \rightarrow \psi$ such that $\theta_x \rightarrow \theta$ in M imply $\psi_x \neq \psi$. x is partially contradicts to M if there exist a rule in M satisfying $\theta_x \rightarrow \theta$ such that $\psi_x \equiv \psi$ and there exist an other rule in M satisfying $\theta_x \rightarrow \theta$ such that $\psi_x \neq \psi$. Denote the set of those x by I_c . The set $C_x = \{i | i \in U, \theta_x \rightarrow \theta_i, \psi_x \neq \psi_i\}$ is called the contradict domination of x . If there is no rule $\theta \rightarrow \psi$ such that $\theta_x \rightarrow \theta$ in M , then x is called completely new in M . Denote the set of thus x by I_n .

By the above definitions, the following property holds.

Proposition 1. *Let I denote all the new instances for a decision table. Then $\{I_m, I_c, I_n\}$ is a partition of I .*

Proposition 2. *Let x be a new instance. If there is no an object $y \in C_x$ such that $y \models_S \theta_x/R$ in S . Then the reduction of condition attributes of S' is R .*

Proof: There is no an object $y \in C_x$ such that $y \models_S \theta_x/R$ in S , therefore, there is no any RD -rule $\theta \rightarrow \psi$ such that $\theta_x/R \equiv \theta$ and $\psi_x \neq \psi$ in S . Set $U' = U \cup \{x\}$ and $S' = (U', A)$. Then $POS_R(D)' = POS_R(D) \cup \{x\}$ and $POS_C(D)' = POS_C(D) \cup \{x\}$, where $POS_R(D)'$ and $POS_C(D)'$ are R -Positive region and C -Positive region of D on U' respectively. Hence $POS_C(D)' = POS_R(D)'$ since $POS_C(D) = POS_R(D)$. Therefor the reduction of condition attributes of S' is R because R is the D -independent subfamily of C in U' . $\square$

Proposition 3. *If there is an object $y \in M_x$, $y \models_S \theta_x$ in decision table S , then the decision table $S' = S$, and its reduction of attributes and minimal set of rules will not change.*

The proof of the above proposition is obvious.

Proposition 4. *If there is an object $y \in C_x$ such that $y \models_S \theta_x$ and there is no other object $i \in C_x (i \neq y)$ such that $i \models_S \theta_x/R$. Then There is a condition attributes reduction R' of S' such that $R' \subseteq R$.*

Proof: If there is unique $y \in C_x$ such that $y \models_S \theta_x$ in S , then $POS_C(D)' = POS_C(D) - \{y | y \models_S \theta_x\}$. There is no any other $i \in C_x$ such that θ_x/R , then $POS_R(D)' = POS_R(D) - \{y | y \models_S \theta_x\}$. Hence, $POS_C(D)' = POS_R(D)'$. Because the size of positive region is reduced when the new instance is add to S , R may be not the D -independent subfamily of C in U' . There is a reduction R' of condition attributes of S' such that $R' \subseteq R$. $\square$

Proposition 5. *If there are two objects $y \in C_x$ satisfying $y \models_S \theta_x$ at least, then R is still a condition attributes reduction of S' .*

Proof: If there are two objects $y_1, y_2 \in C_x$ satisfying $y_i \models_S \theta_x (i = 1, 2)$, then $POS_R(D)' = POS_R(D)$ and $POS_C(D)' = POS_C(D)$. Hence $POS_C(D)' = POS_R(D)'$. The positive region doesn't change in S' , R is the D -independent subfamily of C in U' . So a condition attribute reduction S' is still R . $\square$

Proposition 6. *Let $\theta_i \rightarrow \psi_i$ be a RD -rule reduction of object i in S , and $i \in C_x \cup \{j | \theta'_j \equiv \theta'_h, h \in C_x, j \in M_x\}$ (where θ'_h and θ'_j are C -basic formulas of h and j respectively). Let W be the set of all minimal rule sets of S' . If an attribute reduction R in S is also an attribute reduction in S' . Then there is a minimal rule-set $M' \in W$ such that $\theta_i \rightarrow \psi_i \in M'$.*

Proof: because $i \in C_x \cup \{j | \theta'_j \equiv \theta'_h, h \in C_x, j \in M_x\}$ (where θ'_h and θ'_j are C -basic formulas of h and j respectively). Clearly, there is no $\theta_x \rightarrow \theta_i$ nor $\psi_x \neq \psi_i$. Hence there is a minimal rule-set $M' \in W$ such that $\theta_i \rightarrow \psi_i \in M'$. $\square$

According to the above propositions, the following corollaries can be got.

Corollary 1. *If a new instance x matches to M , then M is also a minimal decision rule-set of S' .*

Corollary 2. *If a new instance x is completely new in M , then R is a condition attributes reduction of S' and a minimal set of rules of S' is $M \cup \{\theta'_x \rightarrow \psi_x | \theta'_x \rightarrow \psi_x \text{ is a reduction of } \theta_x/R \rightarrow \psi_x\}$.*

3 Algorithm for Incremental Learning

The algorithm for incremental learning

Input: $S = (U, A)$, R , M , x and its CD -basic rule $\theta_x \rightarrow \psi_x$.

Output: M' (the minimal rule set of S').

Algorithm: Set $M_1 = M_2 = M_3 = \emptyset$;

$\forall \theta \rightarrow \psi$ in M

If $\theta_x \rightarrow \theta$ is false, then $M_1 = M_1 \cup \{\theta \rightarrow \psi\}$

Else if $\psi_x \equiv \psi$, then $M_2 = M_2 \cup \{\theta \rightarrow \psi\}$

Else $M_3 = M_3 \cup \{\theta \rightarrow \psi\}$.

If $M_2 = M_3 = \emptyset$, then $M' = M \cup \{\theta'_x \rightarrow \psi_x\}$.

Else if $M_3 = \emptyset$, then $M' = M$.

Else compute $C_x = \{i | i \text{ is an object which rule is in } M_3\}$.

If $\exists y \in M_x$ such that $y \models_S \theta_x$, then $M' = M$;

Else if $\nexists y \in C_x$ such that $y \models_S \theta_x / R$, then $R' = \{R\}$

else, $\exists y_i \in C_x$ ($i \geq 2$) such that $y_i \models_S \theta_x$, then $R' = \{R\}$;

if there is unique $y \in C_x$ such that $y \models_S \theta_x$, then compute R' ;

Else compute R' .

If $R \in R'$, then compute $C_{C_x} = \{\theta_j \rightarrow \psi_j | \text{ for all } i \in C_x, j \in M_x, \theta_i \equiv \theta_j\}$ and M'' . Set $M' = M'' \cup M_1 \cup M_2 - C_{C_x}$.

Else simply the rules in S' , then minimal rule set is M' .

End for

OUTPUT M' .

Let m and n be the size of C and U respectively. If x matches to M or is completely new in M , the time-complexity of the algorithm is $O(mn)$. If x contradicts to M and there no any object in C_x such that $y \models_S \theta_x / R$ or there are two objects $y_i \in C_x$ such that $y_i \models_S \theta_x$ ($i = 1, 2$), the time-complexity of the algorithm is $O(mn^2)$. Those two cases are the best cases of the algorithm. But in other situations, the efficiency of this algorithm is $O(2^m n^2)$.

Acknowledgements

The authors are indebted to the anonymous referees for their detailed corrections and suggestions.

References

1. Mollestad T, *A rough set approach to data mining, extracting a logic default rules from data*, Norwegian University of Science and Technology, 1997.
2. Tong Linyun, An Liping, *Incremental learning of decision rules based on rough set theory*, Proceedings of the 4th World Congress on Intelligent Control and Automation, Shanghai, Press of East China University of Science and Technology, 2002: 420-425.
3. Liping An, Jianyong Zhang, Linyun Tong, *Knowledge acquisition, incremental modification and reasoning based on rough set theory in expert systems*, Computer engineering and design, 25(1)(2004): 42-45.
4. Z. Pawlak, *Rough sets and intelligent data analysis*, Information Sciences, 147(2002): 1-12.
5. Z. Pawlak, *ROUGH SETS: Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, 1991.
6. Ning Shan, Wojciech Ziarko, *Data-based acquisition and incremental modification of classification rules*, Computational Intelligence, 11(2)(1995): 357-370.