Incremental Attribute Reduction Based on Elementary Sets

Feng Hu, Guoyin Wang, Hai Huang, and Yu Wu

Institute of Computer Science and Technology,
Chongqing University of Posts and Telecommunications,
Chongqing 400065, P.R.China
wanggy@ieee.org

Abstract. In the research of knowledge acquisition based on rough sets theory, attribute reduction is a key problem. Many researchers proposed some algorithms for attribute reduction. Unfortunately, most of them are designed for static data processing. However, many real data are generated dynamically. In this paper, an incremental attribute reduction algorithm is proposed. When new objects are added into a decision information system, a new attribute reduction can be got by this method quickly.

1 Introduction

Rough Sets [1] (RS) is a valid mathematical theory to deal with imprecise, uncertain, and vague information. It has been applied in such fields as machine learning, data mining, intelligent data analyzing and control algorithm acquiring, etc, successfully since it was developed by Professor Z. Pawlak in 1982.

Attribute reduction [2] is a key problem in rough sets based knowledge acquisition, and many researchers proposed some algorithms for attribute reduction [3-5]. Unfortunately, most of them are designed for static data processing. However, many real data are generated dynamically. Thus, many researchers suggest that knowledge acquisition algorithms should better be incremental[6-8]. Some incremental rough sets based rule extraction algorithms [9-11] have been developed, but they don't consider attribute reduction problem. An incremental attribute reduction algorithm is developed in paper [12]. It can only process information systems without decision attribute. However, most of real information systems are decision information system. Incremental attribute reduction in decision information system would be more important. In paper [13-15], some methods are proposed for incremental attribute reduction in decision information system. In this paper, we develop an incremental attribute reduction algorithm. It can generate a new reduction result for a decision information system quickly after new objects are added. Now, we are comparing our algorithm with these methods developed in paper [13-15] and testing more difficult data.

The rest of this paper is organized as follows: In section 2, basic notions about rough sets theory are introduced. In section 3, we proposed the principle

D. Ślęzak et al. (Eds.): RSFDGrC 2005, LNAI 3641, pp. 185–193, 2005.

of incremental attribute reduction. In section 4, the incremental attribute reduction algorithm based on elementary set is developed. In section 5, the result of simulation experiments is discussed. At last, we conclude this paper in section 6.

2 Basic Notions in Rough Sets Theory

For the convenience of description, some basic notions of decision information systems are introduced here at first.

Definition 1. *(decision information systems* [16-18]*) A decision information system is defined as $S =< U, R, V, f >$, where U is a non-empty finite set of objects, called universe, R is a non-empty finite set of attributes, $R = C \bigcup D$, where C is the set of condition attributes and D is the set of decision attributes, $D \neq$. $V = \bigcup\limits_{p \in R} V_p$, and V_p is the domain of the attribute p. $f : U \times R \to V$ is a total function such that $f(x_i, p) \in V_p$ for every $p \in R, x_i \in U$.*

Definition 2. *(indiscernibility relation* [16-18]*) Given a decision information system $S =< U, C \bigcup D, V, f >$, each subset $B \subseteq C$ of attribute determines an indiscernibility relation IND(B) as follows: $IND(B) = \{(x, y)|(x, y) \in U \times U, \forall b \in B, (b(x) = b(y))\}$. Equivalent classes of the relation IND(B) will be called B elementary sets in S, we denote it as E_i. The set of all elementary sets will be denoted by U/IND(B).*

Definition 3. *(consistency and inconsistency of elementary sets) Given a decision information system $S =< U, C \bigcup D, V, f >$. An elementary set $E_{\text{i}} \in U/IND(B)$ $(B \subseteq C)$ is consistent iff all its objects have the same decision value. Otherwise, it is inconsistent.*

Definition 4. *(lower-approximation and upper-approximation* [16-18]*) Given an information system $S =< U, R, V, f >$, for any subset $X \subseteq U$ and indiscernibility relation IND(B), the B lower-approximation and upper-approximation of X is defined as: $B\_(X) = \cup\{Y_i|Y_i \in U/IND(B) \wedge Y_i \subseteq X\}$, $B^-(X) = \cup\{Y_i|Y_i \in U/IND(B) \wedge Y_i \cap X \neq \}$.*

Definition 5. *(positive region* [16-18]*) Given a decision information system $S =< U, R, V, f >$, $P \subseteq R$ and $Q \subseteq R$, the* P *positive region of* Q *is defined as: $Pos_P(Q) = \bigcup_{X \in U/Q} P\_(X)$.*

3 Principle of Incremental Attribute Reduction

According to Definition 3, all elementary sets can be divided into two parts: positive elementary set and negative elementary set.

Definition 6. *(positive elementary set and negative elementary set): Given a decision information system $S =< U, C \bigcup D, V, f >$, D=$\{d\}$. All consistent C elementary sets in S construct a set Ps, that is,$\forall_{E_i \in Ps} \forall_{x,y \in E_i} (d(x) = d(y))$.*

Ps is called the positive elementary set. All inconsistent C elementary sets in S construct another set Ns, that is,$\forall_{E_i \in Ns} \exists_{x,y \in E_i}(d(x) \neq d(y))$ *, Ns is called the negative elementary set.*

Definition 7. *(collision elementary set): Given a decision information system* $S =< U, C \bigcup D, V, f >$*, C and D are its condition attribute set and decision attribute set respectively. Ps is its positive elementary set and Ns is its negative elementary set. Given an attribute subset* $B \subseteq C, \forall E_i \in Ps$*, if* E_i *can satisfy one of the following two conditions:*

(1) *There is a elementary set* $E_j \in Ps(E_i \neq E_j)$ *,* E_i *and* E_j *have the same values for the condition attribute subset B, and different values for the decision attribute set D.*

(2) *There is a elementary set* $E_j \in Ns$ *,* E_i *and* E_j *have the same values for the condition attribute subset B.*

Then E_i *is called a collision elementary set on attribute set B in Ps, otherwise* E_i *is called a non-collision elementary set on attribute set B in Ps.*

Proposition 1. *(monotony of collision element set): Assume* E_i *is a non-collision elementary set on attribute set* $B(B \subseteq C)$ *, for any attribute set* $A(B \subseteq A \subseteq C)$ *,* E_i *is also a non-collision elementary set on attribute A. On the contrary, assume* E_i *is a collision elementary set on attribute set* $B(B \subseteq C)$*, for any attribute set* $A(A \subseteq B \subseteq C)$*,* E_i *a is also collision elementary set on attribute set A.*

Proof: according to Definition 7, it is obvious.

Proposition 2. *Given a decision information system* $S =< U, C \bigcup D, V, f >$*, for any attribute set* $B \subseteq C$*,* $Pos_B(D) = Pos_C(D)$ *iff there is no collision elementary set on B in its positive elementary set Ps.*

Proof: It is obvious according to the define of positive region.

In this paper, an incremental attribute reduction algorithm is developed based on Proposition 1 and Proposition 2.

4 Incremental Attribute Reduction Algorithm Based on Elementary Sets

Let *red* be an attribute reduction of a decision information system $S =< U, C \bigcup D, V, f >$, where $D=\{d\}$. Thus, $Pos_C(D) = Pos_{red}(D)$. According to Proposition 2, there is no collision elementary set on attribute set *red* in the positive elementary set *Ps* of *S*. When a new object (we denote it as $record_{new}$) is added into the decision information system, three different cases may happen.

1. $\exists_{E_i \in Ns}(record_{new} \in E_i)$. According to Definition 7, it is obvious that there is no collision element on *red* in *Ps* after $record_{new}$ is added into the decision information systems, thus $Pos_C(D) = Pos_{red}(D)$ still holds after $record_{new}$ is added.

2. $\exists_{E_i \in Ps}(record_{new} \in E_i)$. There will be the following two cases:
   2.1. $\forall_{x \in E_i}(d(x) = d(record_{new})$. According to Definition 7, it is obvious that there is no collision elementary set on *red* in *Ps* after $record_{new}$ is added, thus $Pos_C(D) = Pos_{red}(D)$ still holds.
   2.2. $\forall_{x \in E_i}(d(x) \neq d(record_{new})$. That is, the elementary set E_i will no longer belong to *Ps* after $record_{new}$ is added into E_i. Thus, $Ps = Ps - E_i, Ns = Ns \bigcup E_i$. It is obvious that there is no collision elementary set on *red* in *Ps* after $record_{new}$ is added, thus $Pos_C(D) = Pos_{red}(D)$ still holds.
3. $\forall_{E_i \in Ps \wedge E_i \in Ns} record_{new} \notin E_i$. Thus, a new elementary set E_{new} will be generated for $record_{new}$. E_{new} must belong to the positive elementary set, so $Ps = Ps \bigcup E_{new}$. We compare E_{new} with all other elementary sets in *Ps* and *Ns*:
   3.1. If E_{new} is a non-collision elementary set on *red*, it is obvious that there is no collision elementary set on *red* in *Ps* after $record_{new}$ is added, then $Pos_C(D) = Pos_{red}(D)$ still holds.
   3.2. If E_{new} is a collision elementary set on *red*, that is $Pos_C(D) \neq Pos_{red}(D)$ after $record_{new}$ is added into *Ps*. We might as well assume that E_{new} contradicts E_k on *red*. There must be some condition attributes in attribute set $C - red$ on which E_{new} and E_k have different values. Let $(c_1, c_2, ..., c_k \in (C - red)) \wedge \forall_{x \in E_{new}, y \in E_k}((c_1(x) \neq c_1(y)) \wedge (c_2(x) \neq c_2(y)) \wedge ... \wedge (c_k(x) \neq c_k(y)))$, according to Definition 7, it is obviously that E_{new} won't contradict E_k on $red \cup \{c_i\}(i = 1, ..., k)$. According to Proposition 1, those elementary sets that don't contradict E_{new} on *red* won't contradict on $red \cup \{c_i\}(i = 1, ..., k)$ also. Therefore, there is no collision elementary set on $red \cup \{c_i\}$ in *Ps*, that is $Pos_C(D) = Pos_{red \cup \{c_i\}}(D)$.

In step 3.2, in order to get the attribute reduction result with as less number of condition attributes as possible, we would not choose a condition attribute from $c_i, c_2, ..., c_k$ and add it into *red* immediately in our algorithm when more than 1 objects are added. We could generate a disjunctive formula for each of them, that is $b_j = c_i \vee c_2 \vee, ..., \vee c_k$. After all new objects are added into the original decision information system, we unite all these disjunctive formulas, let it be $F = b_1 \wedge b_2 \wedge ... \wedge b_m$(Suppose m disjunctive formulas are generated). F is a conjunctive formal formula (CNF). We could transform F into a disjunctive normal formula (DNF), that is, $F = q_1 \vee q_2 \vee ... \vee q_i \vee ... \vee q_n (q_i = c_1 \wedge c_2 \wedge ... \wedge c_l)$. Suppose $q_i(q_i = c_1 \wedge c_2 \wedge ... \wedge c_l)$ is the smallest term in *F*, that is the number of attributes in q_i is less than all other terms in *F*. Let $A = \{c_1, c_2, ..., c_l\}$. It is obvious that there is no collision elementary set on $red \cup A$ in *Ps*, that is, $Pos_C(D) = Pos_{red \cup A}(D)$.

In the following, we can get the attribute reduction of the new decision information systems after getting rid of possible redundant attribute in $red \cup A$.

Algorithm1: Incremental attribute reduction algorithm based on element set.

Input: An original decision information system $S = < U, C \bigcup D, V, f >$, one of its attribute reduction *red*, its positive elementary set *Ps* and negative element set *Ns*, and a new object set *Add* to be added.

Output: the attribute reduction result *Red* of the new decision information system $S_1 =< U \bigcup Add, C \bigcup D, V_1, f_1 >$.

1. Choose the first object from Add and denote it as $record_{new}$, let k=1, $Add = Add - \{record_{new}\}$.
2. if $(\exists_{E_i \in Ns}(record_{new} \in E_i))$, go to Step 5.
3. if $(\exists_{E_i \in Ps}(record_{new} \in E_i))$,
   3.1. if $\forall_{x \in E_i}(d(x) = d(record_{new})$, go to Step5.
   3.2. if $\forall_{x \in E_i}(d(x) \neq d(record_{new}), Ps = Ps - E_i, Ns = Ns \bigcup E_i$, go to Step5.
4. Generate a new elementary set E_{new} for $record_{new}$.
   4.1. for i=1 to $|Ps|$ do
      4.1.1. Let E_i be the i-th elementary set in *Ps*.
      4.1.2. If E_{new} contradicts E_i on attribute set *red*, let $c_i, c_2, ..., c_j$ be the attributes in the attribute set $C - red$ on which E_{new} and E_i have different values, and $b_k = c_i \vee c_2 \vee, ..., \vee c_j$, $k = k + 1$.
   4.2. for i=1 to $|Ns|$ do
      4.2.1. (Let E_i be the i-th elementary set in *Ns*.
      4.2.2. If E_{new} contradicts E_i on attribute set *red*, just like Step 4.1, we can get $b_k = c_i \vee c_2 \vee, ..., \vee c_j$, $k = k + 1$.
      4.2.3. 4.3 $Ps = Ps \cup E_{new}$.
5. If $Add = \emptyset$, go to Step6. Otherwise, choose the next object from *Add*, and denote it as $record_{new}$, $Add = Add - \{record_{new}\}$, go to Step 2.
6. Let $F = b_1 \wedge b_2 \wedge ... \wedge b_{k-1}$ and transform F to a disjunctive formula $(F = q_1 \vee q_2 \vee ... \vee q_i \vee ... \vee q_n$). Choose the smallest term $q_j(q_i = c_1 \wedge c_2 \wedge ... \wedge c_l)$ from F. $A = \{c_1, c_2, ..., c_l\}$.
7. $Red = red \cup A$.
8. for i=1 to $|Red|$ do
   8.1. $P = Red$;
   8.2. Let c_i be i-th attribute in Red;
   8.3. $P = P - \{c_i\}$;
   8.4. if $Pos_C(D) = Pos_P(D)$, then $Red = P$.
9. Return *Red*.

5 Experiment Results

In order to test the validity of the our algorithm, three classical algorithms for attribute reduction (attribute reduction algorithm based on information entropy, attribute reduction algorithm based on discernibility matrix, and attribute reduction algorithm based on character choice) in RIDAS system [19] are used. The configuration of the PC in our experiments is P4 2.66G(CPU), 512M(memory), windows2000 (operation system).

5.1 UCI Database Test

We use datasets Heart_c_ls, Pima_India, Crx_bq_ls, Liver_disorder and Abalone from UCI database (These data sets can be downloaded at http://www.ics.uci.edu) as test dataset. The parameters of these five data sets are shown in Table 1. 80% objects of these five data sets are used as the original decision information systems, and the other 20% are used as additive datasets respectively. The whole dataset are used as the new decision information systems. Firstly, we use the three classical algorithms to generate the attribute reductions for each original decision information system. Secondly, based on previous results, we use Algorithm 1 to generate the attribute reductions for each new decision information system. Finally, we use the three classical algorithms to generate the attribute reductions for each new decision information system, and compare them with the attribute reductions generating by Algorithm 1. The experiment results are shown in Table 2. Where, T is running time of algorithm, its unit is second. T=0 means that the running time is less than 1 millisecond. n is the number of condition attributes in reduction results.

From Table 2, we can find that the running time of our incremental algorithm is much less than non-incremental algorithms.

5.2 Test on Inconsistent Dataset

In order to test the validity of Algorithm 1 for processing inconsistent decision information systems, we construct five inconsistent datasets randomly. The

Table 1. Experiment Dataset

Dataset	Number of Condition Attributes	Number of Objects
Heart_c_ls	13	303
Pima_India	8	738
Crx_bq_ls	15	690
Liver_disorder	6	1260
Abalone	8	4177

Table 2. Experiment Results for UCI Databases

Dataset	Character choice				Information entropy				discernibility matrix			
	Non-incremental		incremental		Non-incremental		incremental		Non-incremental		incremental	
	T	n	T	n	T	n	T	n	T	n	T	n
Heart_c_ls	4.704	9	0	9	0.156	9	0	9	0.187	9	0.016	9
Pima_India	36.422	5	0	5	0.406	5	0.016	5	0.5	5	0.016	5
Crx_bq_ls	48.187	13	0.109	6	1.032	6	0.031	6	1.954	6	0.015	6
Liver_disorder	71.64	5	0	5	0.078	5	0.016	5	1.359	5	0	5
Abalone	118.406	7	1.266	6	17.234	6	0.594	6	41.875	6	0.578	6

Table 3. Experiment results of Intolerant Dataset

Dataset	Character choice				Information entropy				discernibility matrix			
	Non-incremental		incremental		Non-incremental		incremental		Non-incremental		incremental	
	T	n	T	n	T	n	T	n	T	n	T	n
DataSet1	109.484	7	0.11	7	2.109	6	0.032	6	7.015	6	0.047	6
DataSet2	783.438	7	0.344	7	8.484	6	0.156	7	19.922	6	0.172	7
DataSet3	2598.39	7	0.875	7	19.219	6	0.453	7	42.875	7	0.484	7
DataSet4	6800.11	7	1.5	7	34.516	6	1.219	7	76.563	6	0.813	7
DataSet5	12727.11	8	1.375	8	54.625	7	1.5	7	122.06	7	1.719	7

number of objects of the five datasets are 1000, 2000, 3000, 4000 and 5000 respectively. The number of condition attributes and decision attribute are 15 and 1. For the former 80% objects, the values of their former 10 condition attributes and decision attribute are generated randomly from 0 to 9, and the other 5 condition attributes are all set to be 0. It is taken as the original decision information system. For the other 20% objects, the values of their former 10 condition attributes and decision attribute are generated randomly from 0 to 9, and the other 5 condition attributes are generated randomly from 0 to 1. It is taken as the dataset to be added. The whole dataset is taken as the new decision information system. Some conflict objects will be generated in this way. The test method is similar to 5.1. The experiment result is shown in Table 3.

From Table 3, we can find that the running time of our incremental algorithm is also much less than non-incremental algorithms when processing inconsistent decision information systems.

5.3 Continuous Incremental Learning Test

In order to simulate the incremental knowledge learning ability of a human brain, we construct another dataset randomly. The number of condition attributes and decision attribute are 15 and 1 respectively. The values of condition attributes and decision attribute are generated randomly from 0 to 9, the number of objects of this dataset is 1000 at first. We use the attribute reduction algorithm based on information entropy to generate its attribute reduction. Then, we add new objects into this dataset and use Algorithm 1 to generate the attribute reduction of the new decision information system continuously. The number of objects are 2000, 5000, 10,000, 20,000, 50,000, 100,000, 200,000, 500,000 and 1,000,000 respectively after new objects are added each time. The experiment result is shown in Table 4. Where, N is the number of objects in new decision information system, T is running time, h means hour, m means minute, s means second, and n is the number of condition attributes in the attribute reduction result.

From Table 4, we can find that continuous incremental attribute reduction could be conducted with Algorithm 1. It could simulate the incremental knowledge learning process of a human brain.

Table 4. Experiment Result of Continuous Incremental Test

N	2,000	5,000	10,000	20,000	50,000	100,000	200,000	500,000	1,000,000
T	0.328s	6s	22s	1m29s	12m34s	45m11s	2h20m42s	15h12m6s	71h46m40s
n	6	7	8	8	9	10	10	11	11

Based on above tests, we could have a conclusion that the Algorithm 1 can generate attribute reduction for dynamic decision information systems quickly.

6 Conclusion

Incremental learning is an important problem in AI research. Attribute reduction is a key problem for rough sets based knowledge acquisition. Some incremental rough sets based algorithms for rule extraction have been developed. Unfortunately, they don't consider the attribute reduction problem. In paper [13-15], some methods are proposed for incremental attribute reduction in decision information system. In this paper, an incremental attribute reduction algorithm based on elementary sets is proposed, it can generate attribute reduction for dynamic decision information systems quickly. Our experiment results illustrate that this algorithm is effective. Now, we are comparing our algorithm with these methods developed in paper [13-15] and testing more difficult data.

Acknowledgements. This paper is partially supported by Program for New Century Excellent Talents in University (NCET), National Natural Science Foundation of P. R. China (No.60373111), Application Science Foundation of Chongqing, and Science & Technology Research Program of the Municipal Education Committee of Chongqing of China (No.040505), Young Teacher Foundation of Chongqing University of Posts and Telecommunications (No. A2004-16).

References

1. Pawlak Z.: Rough Set. International Journal of Computer and Information Sciences, Vol.11 (1982) 341-356
2. Wang G.Y. (ed): Rough Set Theory and Knowledge Acquisition, Xi'an Jiaotong University Press, Xi'an (2001)
3. Bazan J.G.: Dynamic reducts and statistical inference, Sixth International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU'96), Vol.III (1996) 1147–1152
4. Bazan J. G., Skowron A., Synak P.: Reducts as a tool for extracting laws from decision tables, International Symposium on Methodologies for Intelligent Systems (ISMIS), Charlotte, NC, series: Lecture Notes in Artificial Intelligence, Vol.869 (1994) 346-355
5. Wroblewski J.: Theoretical Foundations of Order-Based Genetic Algorithms, Fundamenta Informaticae, Vol.28(3-4) (1996) 423-430

6. Fayyad U.M., Shapiro L.P., Smyth P., Uthurusamy R.: Advances in Knowledge Discovery and Data Mining. Menlo Park (ed.), AAAI Press/The MIT Press, California (1996)
7. Cercone V., Tsuchiya M.: Luesy Editors Introdunction, IEEE Transactions on Knowledge and Data Engineering, Vol.5(6) (1993) 901-902
8. Shapiro L.P., Frawley W.J.: Knowledge Discovery in Database, Menlo Park (ed.), AAAI Press/The MIT Press, California (1991)
9. Zheng Z., Wang G.Y.: RRIA: A Rough Set and Rule Tree Based Incremental knowledge Acquisition Algorithm, Fundamenta Informaticae, Vol.59 (2004) 299-313
10. Synak P.: Adaptation of decomposition tree to extended data, Fifth World Multiconference on Systemics, Cybernetics and Informatics, Vol.VII (2001) 552-556
11. Li Y.L., Xu C.F., Gen W.D.: Incremental rule extraction algorithm based on improved discernibility matrix. Chinese Journal of Computer Science, Vol.30(5.A) (2003) 46-49
12. Liu Z.T.: An Incremental Arithmetic for the Smallest Reduction of Attributes, Chinese Journal of Electronics, Vol.27(11) (1999) 96-98
13. Orlowska M., Orlowski M.: Maintenance of Knowledge in Dynamic Information Systems, In: Slowinski R. (ed.), Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory, Kluwer Academic Publishers, Dordrecht (1992) 315-330
14. Susmaga R.: Experiments in Incremental Computation of Reducts, In: Skowron A., Polkowski L. (eds), Rough Sets in Data Mining and Knowledge Discovery, Springer-Verlag, Berlin (1998)
15. Ziarko W., Shan N.: Data-Based Acquisition and Incremental Modification of Classification Rules, Computational Intelligence, Vol.11(2) (1995) 357-370
16. Hu X.H. (ed), Knowledge Discovery in Databases: An Attribute-Oriented rough set approach, University of Regina Press, Regina (1996)
17. Pawlak Z.: Rough sets – Theoretical aspects of reasoning about data. Kluwer Academic Publishers (1991)
18. Polkowski L.: Rough Sets: Mathematical Foundations, Physica Verlag, Heidelberg (2002)
19. Wang G.Y., Zheng Z., Zhang Y.: RIDAS – A Rough Set Based Intelligent Data Analysis System. The First Int. Conf. on Machine Learning and Cybernetics (2002) 646-649