International Journal of Information Technology, Vol. 12 No.4 2006

An Approximate Approach to Attribute Reduction

Xiaobing Pei , YuanZhen Wang

Department of computer science,
HuaZhong University of Science & Technology,
wuhan, Hubei, 430074, China

xiaobingp@tom.com

Abstract

The attribute reduction is one of key processes for knowledge acquisition. This paper deals with approximate approach to attribute reduction. The concept of minimal discernible attributes set is introduced and a calculation method for it is investigated. And then, the judgment theorem with respect to keeping positive region invariability is obtained, from which an approximate approach to attribute reduction aimed at the above processes is proposed. In addition, its time complexity and some proofs were presented in detail. Finally, the experimental results show that this algorithm is effective and efficient.
Keywords: Rough set, Attribute reduction, Approximate approach

I. Introduction

Pawlak [1] first proposed the rough set theory (RS) in 1982, which was a formal framework for the management of uncertainty and vagueness in data and the automated transformation of data into knowledge. The main advantage of the RS is that although one can put additional background or additional preliminary information into the computation, it is not required. Recently, RS has been applied to many fields such as data mining [2][3], decision support systems [4][5], knowledge acquisition [6] etc.

One fundamental aspect of rough set theory for knowledge acquisition involves the searching for some particular subsets of condition attributes. By such one subset the information for classification purposes provided is the same as the condition attribute set done. Such subsets are called reductions. It is well known that an information system or a decision table may usually have irrelevant and superfluous knowledge, from which it is inconvenient for us to get concise and meaningful decision. To acquire brief decision rules from inconsistent decision systems, knowledge reduction is needed. Research has been done and some algorithms have been proposed for attribute reduction [7-11]. Although these proposed algorithms are relatively simple, it is usually difficult to achieve an optimal attribute reduction.

The purpose of this paper is to present an approach to solving the complex problems to achieve optimal attribute reduction by constructing effective and efficient algorithms. The basic notions of rough set theory are reviewed in section 2; In section 3, a new concept of minimal discernible attributes set is introduced and an approach to minimal discernible attributes set is presented. Then the judgment theorem with respect to keeping positive region invariability is obtained and its proofs were presented in detail; In section 4, an approximate approach to attribute reduction is proposed and its time complexity were analyzed in detail. The experimental results are also given. Then brief conclusions are included in section 5.

II. Basic Concepts of Rough Set Theory

By an information system we mean a pair S=(U, A), where U and A are finite, non-empty sets called the universe and a set of attributes respectively. With every attribute $a \in A$, we associate a set V_a of its values, called the domain of a.

Any subset B of A determines a binary relation ind(B) on U, which will be called an indiscernibility relation and is defined as follows:
$(x,y) \in ind(B)$ if and only if f (x, a)=f(y, a) for every $a \in B$, where f(x,a) denotes the value of attribute a for element x.

It can be seen that ind(B) is an equivalence relation. The family of all equivalence classes of ind(B), i.e. the partition determined by B, will be denoted by U/ind(B), and U/ind(B) is defined as follows: $U/ind(B)=\{[x]_B : x \in U\}$, where $[x]_B=\{y: (x,y) \in ind(B)\}$ is a equivalence class for an example x with respect to concept B.

The indiscernibility relation will be used next to define two basic operations in rough set theory as follows:

$$B_*(X)=\bigcup\{Y \in U/ind(B) \mid Y \subseteq X\},$$

$$B^*(X)=\bigcup\{Y \in U/ind(B) \mid Y \cap X \neq \Phi\}$$

which are called the B-lower and the B-upper approximation of X , respectively.
The set $bnd_B(X)=B^*(X)-B_*(X)$ will be referred to as the B-boundary region of X.

If the boundary region of X is the empty set, i.e. $bnd_B(X)=\Phi$, then the set X is crisp with respect to B***;*** otherwise, if $bnd_B(X) \neq \Phi$, the set X is referred to as rough with respect to B.
It is important in data analysis to find out if there is dependency between attributes.

Let C and D be subsets of A, such that $D \cap C=\Phi$ and $D \cup C=A$.
We say that D depends on C in degree k, denoted by $C \Rightarrow_k D$, if

$$k=\gamma(C,D)=\frac{|POS_C(U,D)|}{|U|}$$

Where $POS_C(U, D)=\bigcup_{X \in U/ind(D)} C_*(X)$ called a positive region of the partition U/ind(D) with respect to C, is the set of all elements of U that can be uniquely classified into blocks of the partition U/ind(D), by means of C.

Obviously k= γ (C,D)= $\dfrac{|\bigcup_{X \in U/ind(D)} C_*(X)|}{|U|}$

If k = 1 we say that D depends totally on C, and if k<1, we say that D depends partially (in a degree k) on C.

The coefficient k expresses the ratio of all elements of universe which can be properly classified into a block of the partition U/ind(D) employing attributes C and will be called the degree of the dependency.

We often face a question whether we can remove some data from a database while preserving its basic properties, that is, whether a table contains some superfluous data. Let C, D $\subseteq$ A be sets of condition and decision attributes, respectively.

We say that C' $\subseteq$ C is a D-reduction (reduction with respect to D) of C, if C' is a minimal subset of C such that γ(C' , D)= γ (C, D). We denote that RED_D(U, C) is all D-reductions of C on U.

Hence any reduction enables us to reduce condition attributes in such a way that the degree of dependency between condition and decision attributes is preserved.

III. An Approximate Approach to Attribute Reduction

A. *The Foundation of Theory*

In this section, an approach to the minimal discernible attributes set is presented. Then the judgment theorem with respect to keeping positive region invariability is obtained and its proofs were presented in detail.

Definition 1: Let S=<U, A> be an information system, Z = {B_1, B_2, ., B_n}, Z' ={E_1,E_2, .,E_m}, where $B_i \subseteq$ A, $E_j \subseteq$ A. It can be said that Z' is the prime set of set Z if the following conditions are satisfied:

1) For all E∈Z′, there exist B∈Z such that E = B ;

2) For all B∈Z ,if there exist B′∈Z such that B′⊂B, then B$\notin$Z′.

Example: Let R={{ac},{a},{bc},{d},{de}}, then {{a},{bc},{d}} is the prime set of set R.

K-prime set of set Z, denoted by K- Z′, defined as follows:

1) | K- Z′| = K; 2) For any E$\in$K- Z′, |E|$\leq$| E′|, where E′$\in$Z /K- Z′;

Example: Let R={{ac},{a},{bc},{d},{de}}, then {{a},{d}} is 2-prime set of R.

All appearance, the time complexity of calculation the prime set of set R is O(|A|m^2), where m denotes cardinality of set R.

Definition 2: Let S=<U, A> be an information system, $x_0 \in$U, R$\subseteq 2^P$, V$\subseteq$U, V={x_1,…,$x_{|V|}$}. We denote

$$M(x_0,V,R)=(\vee_m\{\wedge_t\{b_t : b_t \in R_m\}\}) \wedge (\wedge_j\{\vee\{a_k : a_k \in D_j\}\})$$

where D_j={a$\in$P | f(x$_0$， a)$\neq$f(x$_j$， a)， x$_j$ $\in$V} is referred to as discernible attributes set of x$_0$ and x$_j$ (x$_j$ $\in$V), $R_m \in R$.

Then $M(x_0,V,R)$ is referred to as discernible function of x$_0$ and V with respect to set R .

Definition 3: Let S=<U, A> be an information system, x$_0$ $\in$U, V$\subseteq$U, V={x$_1$,…,x$_{|V|}$ }, R$\subseteq 2^P$, where 2^C is power set of C. The minimal disjunctive normal form of discernible function $M(x_0,V,R)$ is $M = \bigvee_{k=1}^{r}(\bigwedge_{s=1}^{q_k} a_{ks})$, denoted by R_k ={ a_{ks} ： s=1,…,q$_k$ }. We say that R_k is a minimal discernible attributes set of {x$_0$} and V with respect to set R. The set of all minimal discernible attributes set, denoted by N(x$_0$,V, R), is defined as follows: N(x$_0$,V, R)={ R_k ： k=1,…,r}.

Example: Given an information system in Table 1, U={#1,#2,#3} is universe, P={a, b, c} and Q={d} are subset of attributes respectively:

U	a	b	c	d
#1	2	2	0	1
#2	0	0	0	0
#3	1	0	1	0

Table 1. Information system

Let R={{ab}, {ac}}, x$_0$=#1, V={#2, #3}, then it can be easily calculated that $M(x_0,V,R)$= ((a$\wedge$b)$\vee$(a$\wedge$c))$\wedge$(a$\vee$b)$\wedge$(a$\vee$b$\vee$c);
Thus we have N(x$_0$,V, R)={a$\wedge$b, a$\wedge$c}.

Algorithm 1: The algorithm for minimal discernible attributes set

Input: Given an information system S=<U, A >, P$\subseteq$A, R$\subseteq 2^P$, x$\in$U, V$\subseteq$U;

Output: N(x, V, R)

1. N=$\bigvee_m(\bigwedge_t(b_t : b_t \in R_m))$; N′=Φ ;

2. For all y$\in$V do:

 2.1. For all $\eta \in$N do:

  N′=N′$\bigcup$ {{ $\eta \wedge$ {a} | a$\in$P and f(y, a)$\neq$f(x,a)}};

 2.2. **N**=prime set of set N′; N′=Φ ;

3. N(x, V, R)=N;

The time complexity is determined by step 2 in algorithm 1. Thus it can be easily concluded that the time complexity of algorithm 1 T=O(|A|K^2|V|), where K=Max{ *Card* (N(x,V′,R)), V′ $\subseteq$V}.

Theorem 1: Let S=<U, A > be an information system, P and Q are subset of attributes set A respectively, $x \in U$, U/ind(D)={D_1,D_2,…,D_m}. Let U_0=U-POS_P(U,Q),$U_i = P_*(Q_i)$($1 \le i \le m$), then for any $\gamma \in Red$ (U-{x}), there exist only U_k ($0 \le k \le m$) such that $[x]_\gamma$-{x}$\subseteq U_k$.

Proof: Suppose that there exist U_k,U_t ($k \neq t$) such that ($[x]_\gamma$-{x})$\bigcap U_k \neq \Phi$ and ($[x]_\gamma$-{x})$\bigcap U_t \neq \Phi$, then there exist $y_1 \in U_k$,$y_2 \in U_t$ ($k \neq t$, $x \neq y_1$, $y \neq y_2$) such that f(y_1,γ)=f(y_2,γ), from which we have y_1,$y_2 \in$ U-POS_P(U,Q), which contradicts $\gamma \in Red$ (U-{x}).
Therefore there exist no U_k, U_t ($k \neq t$) such that ($[x]_\gamma$-{x})$\bigcap U_k \neq \Phi$ and ($[x]_\gamma$-{x})$\bigcap U_t \neq \Phi$.
That is that there exist only U_k such that $[x]_\gamma$-{x}$\subseteq U_k$.

Remark 1: *f(x, γ) = (f(x, r_1),f(x, r_2),…, f(x, $r_{|\gamma|}$)), where γ ={r_1,r_2,…,$r_{|\gamma|}$}.*

Theorem 2: Let S=<U, A> be an information system, P and Q are subset of attributes set A respectively, $x_0 \in POS_P$(U,Q), $\lambda \in RED_Q$(U-{x_0}, P), T=N(x_0, V, λ), where V=$[x_0]_\lambda$-{x_0}. Then POS_γ(U, Q) = POS_P(U, Q) for any $\gamma \in$T.

Proof: By the theorem 1, for any $\gamma \in$T, we have $[x_0]_\lambda$-{x_0}$\subseteq$U-POS_P(U, Q) or $[x_0]_\lambda$-{x_0}$\subseteq POS_P$(U, Q) and |($[x_0]_\lambda$-{x_0})/ind(Q)|=1.
If $[x_0]_\lambda$-{x_0}$\subseteq$U-POS_P(U, Q), then we have f(x, γ)$\neq$f(x_0, γ) for any $x \in$U-POS_P(U, Q). Thus we have POS_γ(U, Q) = POS_P(U, Q).
If $[x_0]_\lambda$-{x_0}$\subseteq POS_P$(U, Q) and |($[x_0]_\lambda$-{x_0})/ind(Q)|=1,then there exist $Y \in$U/ind(Q) such that $[x_0]_\lambda$-{x_0}$\subseteq$Y.
If $x_0 \in$Y, then it is easily concluded that POS_γ(U, Q) = POS_P(U, Q), If $x_0 \notin$Y, then we have f(x, γ)$\neq$f(x_0, γ) for any $x \in$Y.
Thus it can be concluded POS_γ(U, Q) = POS_P(U, Q).
Thus we have POS_λ(U, Q) = POS_P(U, Q) for any $\gamma \in$T.

B. An algorithm for attribute reduction

From the above discussion, we will propose an approximate algorithm for an optimal attribute reduction as follows:

Algorithm 2: An approach algorithm for an attribute reduction

Input: Given an information system S=<U, A >, P$\subseteq$A, Q$\subseteq$A, P={p_1, …, $p_{|P|}$};

Output: a Q-reductions of P.

1. RED ={{p_1}, …, {$p_{|P|}$}}; M=Φ; T=Φ; T_RED=Φ;

2. Calculate U/ind(Q)={Q_1、Q_2、…Q_m}, {B_0、B_1、B_2、…B_m},where B_0=U-POS$_P$(Q), $B_i = B_*(Q_i)(1 \le i \le m)$; M= U-POS$_P$(Q);

3. For all $x \in$ POS$_P$(Q) do:

3.1. For all $\gamma \in$ RED do:

3.1.1. Calculate B_k ($k \le j$) such that $[x]_\gamma$-{x}$\subseteq B_k$, where $[x]_\gamma$-{x}$\subseteq$M;

3.1.2. if $x \in B_k$ then T = T$\cup${γ}; goto 3.1;

3.1.3. T_RED = FunK_N(x, B_k, {γ});

3.1.4. T = T$\cup$T_RED ; goto 3.1;

3.2. RED = K-prime set of set T; T=Φ; M=M$\cup${x}; goto 3;

4. Output 1-prime set of set RED.

Function FunK_N(x, B_k, RED) is defined as follows:

FunK_N(x, B_k, RED)

{

1. For all $y \in B_k$ do:

1.1. For all $\lambda \in$ RED do:

If f(y, λ)$\neq$f(x, λ) then T=T$\cup${γ};

Else T=T$\cup${$\gamma \cup$ p | p$\in$P and f(y, p)$\neq$f(x, p)};

1.2. RED= K-prime set of set T; T=Φ;

2. return RED;

}

The time complexity of step 3.1.1 and 3.1.2 is O(|A||U|) and the time complexity of step 3.1.3 is O(|A|K^2|U|), So the time complexity of 3.1 is O(|A|K^3|U|). The time complexity of 3.2 is O(|A|K^3). Thus we have that the time complexity of algorithm 2 is O(|A|K^3|U|^2).Let K=|A| and K=|1|, the time complexity of algorithm 2 are O(|A|^4|U|^2) and O(|A||U|^2) .

C. *Experimentations*

Now, in order to prove the effectiveness and efficiency of algorithm 2 in this paper, we used the algorithm 2 (let K=|C|, where |C| denotes the cardinality of condition attributes set C) on a lot of databases in UCI. Experimental results were obtained in table 2. According to experimental results, the algorithm 2 is effective and efficient.

Database Name	Number of instances	Number of condition attributes	Number of decision attributes	Attributes number of optimal reduction	Algorithm 2	
					Running time (s)	Attributes number of reduction
Monk's (1)	432	6	1	6	0.156	6
Monk's (2)	432	6	1	6	1.469	6
Monk's (1)	432	6	1	6	0.14	6
Balance	625	4	1	4	0.798	4
Breast	699	7	1	2	2.703	2
BUPA	345	6	1	3	0.531	3
Car	1728	6	1	6	15.41	6
Flare	1099	10	1	9	6.11	9
Pageblocks	5473	10	1	3	56.12	3
Tic	960	9	1	8	8.12	8
Krkopt2	12000	6	1	6	842.78	6

Table 2. Experimental results

IV. Conclusions

In this paper, we are concerned with approximate approach to an optimal attribute reduction. The judgment theorem with respect to keeping positive region invariability is obtained, from which an approximate algorithm for attribute reduction is proposed. Finally, in order to prove the effectiveness and efficiency of this approach, we used the approach on a lot of databases in UCI.

References

[1] Pawlak Z. Rough sets. *International Journal of Computer and Information Science*, 1982, 11(5): 341-356

[2] P. J. Lingras, “Data mining using extensions of the rough set model,” J.Amer. Soc. *Inf. Sci.*, vol. 49, pp. 415–422, May 1998.

[3] N.O. Attoh-Okine, “Rough set application to data mining principles in pavement management database,” *J. Comput. Civil Eng.*, vol. 11, pp.231–237, Apr. 1997.

[4] Z. Pawlak, “Rough set approach to knowledge-based decision support,” *Eur. J. Oper. Res.*, vol. 99, pp. 48–75, Jan. 1997.

[5] Liu Qin. Rough Set and Rough Reasoning. Beijing: Science Press, 2001

[6] L. K. Terje, “Rough modeling—Extracting compact models from large databases,” *Masters thesis, Knowledge Systems Group*, Norwegian Univ. Sci. Technol., Trondheim, Norway, 1999.

[7] Guan J W, Bell D A. Rough computational methods for information systems. *Artificial Intelligences*, 1998, 105（1-2）: 77~103

[8] Wang Jue, Wang Ren, Miao Duo-Qian et al. Data Enriching based on rough set theory. *Chinese Journal of Computer*, 1998, 21(5): 393-395

[9] Dan Pan, Qi-Lun Zheng, An Zeng, Jin-Song Hu. A Novel Self-Optimizing Approach for Knowledge Acquisition. *IEEE Transactions on systems,man, and cybernetics—part A: Systems and Humans*, Vol. 32, No. 4, 2002:505～514

[10] Wang Jue, Wang Ju. Reduction algorithm based on discernibility matrix: The ordered attributes method. *Journal of Computer Science & Technology*, 2001, 16(6): 489~504

[11] A. Skowron and C. Rauszer, "The discernibility matrices and functions in information systems," in Intelligent Decision Support—*Handbook of Applications and Advances of the Rough Sets Theory*, R. Slowinski,Ed. Norwell, MA: Kluwer, 1992, pp. 331–362.

[12] Wong S K M, Ziarko W. On optimal decision rules in decision tables. *Bulletin of Polish Academy of Sciences*,1985（33）: 693-696

Xiaobing Pei is a PhD student in the Department of Computer Science at the HuaZhong University of Science & Technology. He received his MSc. in System Engineer and Applied Mathematics from the Chongqing University in 1997. His current research interests are Data mining, network and network security.