

A Bisection Method for Information System Knowledge Reduction

Zhi Huilai

School of Computer Science and Technology, Henan Polytechnic University,
Jiaozuo Henan P.R.China,
E-mail: zhihuilai@126.com

Abstract

In rough set theory, attribute reduction aims to retain the discernability of the original attribute set, and many attribute reduction algorithms have been proposed in literatures. However, these methods are computationally time-consuming for large scale datasets. We develop a bisection method for attribute reduction and the main opinion is to partition the universe into smaller ones by using partition core attributes to reduce the complexity. Experiments and analysis show that, compared with the traditional un-bisection reduction algorithm, the developed bisection algorithm can significantly reduce computational time while maintaining their results as same as before.

Keywords: information system, knowledge reduction, bisection method

Copyright © 2014 Institute of Advanced Engineering and Science. All rights reserved.

1. Introduction

In applications such as image processing, bioinformatics, astronomy, finance, the number of objects is very large and the dimension (the number of attributes) is very high as well [1-2]. Attributes irrelevant to recognition tasks may deteriorate the performance of learning algorithms [3]. Feature selection plays an important role in the preprocessing step in these applications. It provides techniques to reduce knowledge in database, by which the irrelevant or superfluous knowledge (attributes) can be eliminated according to the learning task without losing essential information about the original data in the databases. In information system, in contrast to the feature selection which usually keeps useful knowledge, attribute reduction is process of get rid of irrelevant or superfluous knowledge. To this extent, attribute reduction and feature selection are the facets of a same problem.

Feature selection can be divided into two main categories: distance-based and consistency-based [3]. For consistency-based feature selection, attribute reduction is regarded as a special form of feature selection in rough set theory and offers a systematic theoretic framework, which does not attempt to maximize the class separability but rather to retain the discernible ability of original attribute sets for the objects from the universe [4-6].

In the last two decades, many reduction algorithms have been proposed and can be divided into two types: finding all reducts (or an optimal reduct) and finding one reduct [7]. The well known algorithm which can find all reducts is using a discernibility matrix and proposed by Skowron [8-9]. However, nearly 10 years before it has been proved to be an NP-hard problem to find all reducts [10]. Heuristic knowledge [11], various type of information entropies [12-14] are used in order to reduce its complexity.

Although endless efforts has been made, the question whether there is a room for further improvement still plagues us.

Fortunately, bisection method, which is ignored and seldom mentioned in this problem before, brings hope to us. In computer science bisection method is successfully used in searching a finite sorted array [15], which is called a binary search or half-interval search. A binary search is a dichotomy divide and conquer search algorithm and halves the number of items to check with each iteration, so locating an item or determining its absence takes logarithmic time. Therefore, in this paper we will use this promising method in information system attribute reduction.

The rest of study is organized as follows. Relative basic concept is in Section 2. In Section 3, we develop a bisection method for attribute reduction. In Section 4, algorithm analysis and experiments are carried out to verify its efficiency. Then, conclusion come in Section 5.

2. Preliminaries

An information system, as a basic concept in rough set theory [16-17], provides a convenient framework for the representation of objects in terms of their attribute values. The following definitions from Definition 1 to Definition 5 originated from [5-7].

Definition 1: An information system is a quadruple $S = (U, A, V, f)$, where U is a finite nonempty set of objects and is called the universe and A is a finite nonempty set of attributes, $V = \bigcup_{a \in A} V_a$ with V_a being the domain of a , and $f: U \times A \rightarrow V$ is an information function with $f(x, a) \in V_a$ for each $a \in A$ and $x \in U$. The system S can often be simplified as $S = (U, A)$.

Definition 2: Let $S = (U, A)$ be an information system and each nonempty subset $B \subseteq A$, an indiscernibility relation is defined as:

$$R_B = \{(x, y) \in U \times U \mid f(x, a) = f(y, a), \forall a \in B\}.$$

It is easily to prove that is R_B a equivalence relation, and it partitions U into some equivalence classes given by $U / R_B = \{[x]_B \mid x \in U\}$, where $[x]_B$ denotes the equivalence class determined by x with respect to B , i.e., $[x]_B = \{y \in U \mid (x, y) \in R_B\}$.

Proposition 1: Let $S = (U, A)$ be an information system and an nonempty subset $B \subseteq A$. If $|B| = k$, then $|U / R_B| = 2^k$, including empty equivalence classes.

Definition 3: Let $S = (U, A)$ be an information system and $B \subseteq A$. Then, B is defined as the partition consistent set of S if $R_B = R_A$. Moreover, B is defined as the partition reduction set of S if any proper subset of B is not a partition consistent set of S .

Definition 4: Let $S = (U, A)$ be an information system. Then, partition discernibility set of $[x_i]_A$ and $[x_j]_A$ is defined as: $D([x_i]_A, [x_j]_A) = \{a_l \in A \mid f_l(x_i) \neq f_l(x_j)\}$.

Definition 5: Let $S = (U, A)$ be an information system and $\{B_k \mid (k \leq r)\}$ be all the partition reduction sets of S , and $C = \bigcap_{k \leq r} B_k$, $K = \bigcap_{k \leq r} B_k - C$, $I = A - \bigcup_{k \leq r} B_k$. Then any element of C is called a partition core attribute, and C is called partition core set; any element of K is called a partition necessary attribute, and K is called partition necessary attribute set; any element of I is called a partition unnecessary attribute, and I is called partition unnecessary attribute set.

3. A Bisection Approach for Information System Knowledge Reduction

Definition 6: Let $\Gamma = \{F_1, F_2, \dots, F_n\}$ be a set family. Then, set H is called a minimal cover of Γ , if $\forall F_i \in \Gamma$, $H \cap F_i \neq \Phi$ and $\forall S \subset H$, $\exists F_i \in \Gamma$, $S \cap F_i = \Phi$. Moreover, $Mc(\Gamma) = \{H \mid H \text{ is a minimal cover of } \Gamma\}$ is called the minimal cover family of Γ .

For example, $\Gamma = \{\{a, b, c\}, \{a, b\}, \{b, c\}\}$, according to Definition 6, $\{a, c\}$ is a minimal cover of Γ , and $\{b\}$ is also a minimal cover of Γ .

Theorem 1: Let $\Gamma = \{F_1, F_2, \dots, F_n\}$. $\forall F_i, F_j \in \Gamma$, if $F_i \subseteq F_j$, then $Mc(\Gamma) = Mc(\Gamma - F_j)$.

Converting a formula from CNF(conjunctive normal form) to DNF(disjunctive normal form) is equivalent to calculating the minimal cover of a given set family. For example, we have $(a \vee b \vee c) \wedge (a \vee b) \wedge (b \vee c) = (a \wedge c) \vee b$ and $Mc(\{\{a,b,c\},\{a,b\},\{b,c\}\}) = \{\{a,c\},\{b\}\}$.

Theorem 2: Let $U = \{U_1, U_2, \dots, U_s\}$ be a set family, and for any $i \neq j$, $U_i \cap U_j = \Phi$. If D_1, D_2, \dots, D_s are the discernibility sets of U_1, U_2, \dots, U_s respectively, then a minimal discernibility set of U is H , H is a minimal cover of $D = \{D_1, D_2, \dots, D_s\}$.

Proof: As H is a minimal cover of $D = \{D_1, D_2, \dots, D_s\}$, according to the definition of minimal cover, given any $D_i \in D$, $H \cap D_i \neq \Phi$, which means that H contains the attributes that can distinguish every objects of U from each other. Moreover, $\forall S \subset H$, $\exists D_i \in D$, $S \cap D_i = \Phi$, and this means that H is the minimal set that could distinguish every objects of U from each other. So, the theorem holds.

Theorem 3 [8]: Let $S = (U, A)$ be an information system and $a \in A$, the following propositions are equivalent:

- (1) a is a partition core attribute;
- (2) there must be a pair of objects $x_i \in U$ and $x_j \in U$ make $D([x_i]_A, [x_j]_A) = \{a\}$;
- (3) $R_{A-\{a\}} \neq R_A$.

For any object $u \in U$, we let $f(u) = \{a \in A \mid u \text{ has attribute } a\}$.

Corollary 1: For any object $x, y \in U$, if $f(x) \subset f(y)$ and $|f(y)| - |f(x)| = 1$, then the attribute $d = f(y) - f(x)$ must be a partition core attribute of S .

Theorem 4: Let $S = (U, A)$ be an information system and C be its partition core set. Let $V = \{[x]_C \mid [x]_C > 1, x \in U\} = \{V_t\}_{t \in T}$ (T is the index set), and D_t is the discernibility set of V_t for each $t \in T$. Then a minimal discernibility set of U is $C \cup H$, H is a minimal cover of $D = \{D_t\}_{t \in T}$.

Proof: As V is formed by equivalent classes, then for any $i, j \in T$, if $i \neq j$, $V_i \cap V_j = \Phi$. By Theorem 2, we know H is a minimal discernibility set of V . As the elements contained in $U - V$ are distinguished from each other by C , so a minimal discernibility set of U is the union of C and H .

Let $S = (U, A)$ be an information system, based on Theorem 4, we can get a bisection method for information system attribute reduction as shown in Algorithm 1:

Algorithm 1: a bisection method for information system attribute reduction

Input: an information system $S = (U, A)$

Output: partition reduction set of S

Step 1: find partition core set C of S ;

1-1 sort objects of U into a sequence L by ascending order according to their number of attributes;

1-2 Initialize $C = \Phi$;

For any two object x, y in neighborhood in sequence L do

if x, y satisfy that $f(x) \subset f(y)$ and $|f(y)| - |f(x)| = 1$, then add attribute $d = f(y) - f(x)$ into C ;

Endfor;

Step 2: bisection the universe U by using each attribute in C ;

2-1 Let $k = |C|$, $b = 2$, $index = 1$, set an array $s[]$ and initialize $s[1] = U$;

2-2: For $i = 0$ to $k - 1$ do

While $index < b$ do

{ $left = index * 2$; $right = left + 1$;

By using one attribute in C , divide $s[index]$ into two groups, one group is stored in $s[left]$, and the other group is stored in $s[right]$;

```

index++;
}Endwhile;
b = b*2;
Endfor;

```

Step 3: compute the partition discernibility set of the un- distinguished objects, and denote the results as $D = \{D_i \mid i \leq r\}$;

Step 4: compute the minimal cover $MC(D)$ of the set family D ;

Step 5: get partition reduction set of S by the union of partition core set C and $MC(D)$

Example 1: Consider descriptions of several objects in Table 1. This is an information system, where $U = \{1, 2, 3, 4, 5, 6, 7, 8\}$, and $A = \{a, b, c, d, e, f, g, h, i\}$.

Table 1. Object Descriptions

	a	b	c	d	e	f	g	h	i
1	*	*					*		
2	*	*					*	*	
3	*	*	*				*	*	
4	*	*	*				*	*	*
5	*	*		*		*			
6	*	*	*	*		*			
7	*		*	*	*				
8	*		*	*		*			

Step 1: get partition core set C .

As $f(2) - f(1) = h$, $f(6) - f(5) = c$, $f(6) - f(8) = b$, so $C = \{h, c, b\}$

Step 2: bisection the universe U by using every attribute of C , the process is shown in Figure 1, and finally we get set family $\{\{1,5\},\{6\},\{7,8\},\{3\},\{4\}\}$.

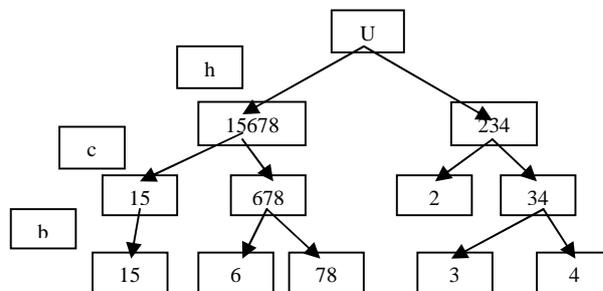


Figure 1. Bisection Process of the Universe

Step 3: compute the partition discernibility set of the un- distinguished objects $\{1,5\}$ and $\{7,8\}$, and we get $D(\{1,5\}) = \{d, f, g\}$, and $D(\{7,8\}) = \{e, f\}$;

Step 4: compute the minimal cover of the set family $D = \{D(\{1,5\}), D(\{7,8\})\}$, and we get $Mc(D) = \{\{d, e\}, \{e, g\}, \{f\}\}$;

Step 5: union partition core set C and $Mc(D)$, and we get the partition reduction set $\{h, c, b, d, e\}$, $\{h, c, b, e, g\}$ and $\{h, c, b, f\}$.

4. Algorithm Analysis and Experiment

The tradition method as well as its improved ones of information system attribute reduction is based on discernibility function, which appears as:

$$f = \wedge \{ \vee m_{ij} \mid 0 \leq j < i \leq n, m_{ij} \neq \Phi \}, m_{ij} \text{ is the partition discernibility set of objects } x_i \text{ and } x_j.$$

Apparently, after converting f form CNF to DNF, the conjunctive clauses connected by the symbol “ \wedge ” are the partition reductions of the information system. So the complexity of CNF determines the efficiency of information system attribute reduction, i.e. the more disjunction clauses contained in CNF, the lower the efficiency becomes, and vice versa.

Based on the previous discussion in section 3, we know the calculation of minimal cover family of a given set is equivalent to a conversion form CNF to DNF. So we can estimate algorithm efficiency by using the number of disjunction clauses contained in the CNF formula.

Let $S = (U, A)$ be a information and $|U| = n$. In the traditional un-bisection method, there are $n^*(n-1)/2$ disjunction clauses contained in the initial discernibility function.

Now we compare our method with the traditional un-bisection method.

(1) Worst case: in step 2, if in each time by using one partition core attribute only one objects can be apart from the others, then after running k times bisection on U , there will be $(n-k)$ undistinguished objects left, so there will be $(n-k)(n-k-1)/2$ disjunction clauses in the family set D .

$$\text{The ratio of our method to the traditional un-bisection method is } \frac{(n-k)*(n-k-1)}{n*(n-1)}.$$

(2) Best case: in step 2, if in each time by using one partition core attribute the universe is divided into two parts equally, then after running bisection process k times, U is divided into 2^k classes, and each class has $n/2^k$ undisguised objects. Computing the partition discernibility set of all these undisguised objects, there will be $2^k * \frac{n}{2^k} * (\frac{n}{2^k} - 1) / 2 = n * (n / \frac{n}{2^k} - 1) / 2$ disjunction clauses in the family set D .

$$\text{The ratio of our method to the traditional un-bisection method is } \frac{n/2^k - 1}{n-1} \approx \frac{1}{2^k}.$$

(3) Average case: we carry this analysis by programming in Java. In initialization, we randomly generate the universe and the number of its objects varies at the range of 100 to 1000. In step 3, we record the number of partition core attributes and the number of disjunction clauses contained in the family set D .

After repeating the experiment 10000 times, based on the number of disjunction clauses, the average ratio of our method to the traditional un-bisection method is derived, which has a strong relationship with the number of partition core attributes. When the number of partition core attributes increases, the ratio decreases dramatically as seen in Table 2.

Table 2. Experiment on Artificial Data Set

number of partition core attributes	1	2	3	4	5	6	7	8
ratio of our method to un-bisection method	61.4%	33.7%	23.3%	14.5%	8.9%	5.6%	3.5%	2.3%

Experiment on UCI database: we select data sets monks, cancer and monkey, and construct information system by using their conditional attributes. And then we program in Java, and run it on a computer with AMD1.4HZ processor and 2G memory. The experiment result is shown in Table 3.

Table 3. Experiment on UCI Data Sets

	number of samples	number of attributes	un-bisection method	our method
monks	432	6	6.1471s	1.7192s
cancer	682	9	53.7854s	8.1753s
monkey	556	17	71.0963s	51.1848s

5. Conclusion

This paper developed a bisection reduction algorithm for information system. Theoretical analysis and experimental results have shown that, compared with the traditional non-bisection reduction algorithm, the proposed algorithm is effective and efficient. It is our wish this study provides new views and thoughts on dealing with knowledge reduction for information systems for all the data types, such as interval number, real number, enumeration and so on.

The work presented in this paper is supported by National Science Foundation of China (Grant No. 60975033) and Doctorial Foundation of He'nan Polytechnic University (Grant No. B2011-102).

References

- [1] Pawlak Z. *Rough Sets: Theoretical Aspects of Reasoning about Data*. Boston: Kluwer Academic Publishers. 1991.
- [2] Guyon I, Elisseeff A. An introduction to variable feature selection. *Journal of Machine Learning Research*. 2003; 3(1): 1157-1182.
- [3] Hu QH, Xie ZX, Yu DR. Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation. *Pattern Recognition*. 2007; 40(2): 3509-3521.
- [4] Babu VS, Viswanath P. Rough-fuzzy weighted k-nearest leader classifier for large datasets. *Pattern Recognition*. 2009; 42(9): 1719-1731.
- [5] Jensen R, Shen Q. Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches. *IEEE Transactions on Knowledge and Data Engineering*. 2004; 16(12): 1457-1471.
- [6] Swiniarski R, Skowron, A. Rough set methods in feature selection and recognition. *Pattern Recognition Letters*. 2003; 24(6): 833-849.
- [7] Qian YH, Liang JY. Combination entropy and combination granulation in rough set theory. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. 2008; 16(2): 179-193.
- [8] Skowron A, Rauszer C. *The discernibility matrices and functions in information systems*. Roman Slowinski. *Intelligent Decision Support-Handbook of Applications and Advances of the Rough Sets Theory*. Dordrecht, Netherland:Kluwer Academic Publishers. 1992: 331-362.
- [9] Skowron A. Boolean Reasoning for Decision Rules Generation. 7th International Symposium on Methodologies for Intelligence Systems. Trondheim, Norway. 1993: 295-305.
- [10] Slezak D. Approximate entropy reducts. *Fundamenta Informaticae*. 2002; 53(3-4): 365-390.
- [11] Hu X H, Cercone N. Learning in relational databases: a rough set approach. *International Journal of Computational Intelligence*. 1995; 11(2): 323-338.
- [12] Wang GY, Zhao J, An JJ. A comparative study of algebra viewpoint and information viewpoint in attribute reduction. *Fundamenta Informaticae*. 2005; 68(3): 289-301.
- [13] Liang JY, Chin KS, Dang CY, et al. A new method for measuring uncertainty and fuzziness in rough set theory. *International of Journal General Systems*. 2002; 31(4): 331-342.
- [14] Liang JY, Xu ZB. The algorithm on knowledge reduction in incomplete information tables. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. 2002; 10(1): 95-103.
- [15] Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, Clifford Stein. *Introduction to Algorithms*. 1st edition. cambridge: MIT Press. 1990.
- [16] Zhang Yi. Research of Strategic Alliance Stable Decision-making Model Based on Rough Set and DEA. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013;11(12): 7269-7276.
- [17] Yue Hou. Rough Set Extension Model of Incomplete Information System. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2014; 12(4): 2843-2849.