Neurocomputing 123 (2014) 185–196

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Analyzing rough set based attribute reductions by extension rule

Bing Li*, Tommy W.S. Chow, Peng Tang

Department of Electronic Engineering, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong, China

ARTICLE INFO

Article history:
Received 13 June 2012
Received in revised form
25 April 2013
Accepted 14 July 2013
Communicated by N.T. Nguyen
Available online 17 August 2013

Keywords:
Attribute reduction
Extension rule
Reduction distribution
Rough set

ABSTRACT

An improved discernibility function for rough set based attribute reduction is defined to keep discernibility ability and remove redundant attributes without the precondition of the Positive Region. On the basis of discernibility function, the solution of rough set based attribute reduction can be found by satisfiability methods. With extension rule theory, a satisfiability method, the distribution of solutions with different number of attributes is obtained without enumerating all attribute reductions. Then, it is easy to search the attribute reduction with the smallest number of attributes. In addition, the cost of space and time is analyzed to find factors playing role in the computation of the method.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The size of datasets has been increasing dramatically, so it has been an important issue to reduce huge objects and large dimensionality in datasets. Attribute reduction, also called feature selection, finds a subset of attributes to reduce dimensionality. By reducing attributes, it can save the cost of computational time and memory; it is also useful to improve predicative ability as a result of removing redundant and irrelevant attributes [1,2].

There exist two major approaches in attribute reduction: individual evaluation and subset evaluation. Individual evaluation, also known as ranking, assigns each attribute a weight representing its degree of relevance [39,40]. This method is incapable of removing redundancy because of similar weights among redundant attributes. So it is always set as a principal or auxiliary section. According to different mechanisms of reduction, subset evaluation falls into three catalogs: wrapper method, embedded method and filter method. The wrapper method [3,4] utilizes a classifier of interest as a black box to score subsets of attributes according to their predictive power. It can provide a highly predictive subset; however, the bias of classifiers and the setup of experiments play role in the performance of the subset. In addition, large computational cost is also needed. An improved method, the embedded method, incorporates attribute reduction as the part of training process [41,42]. Comparing with the wrapper method, it does not split data into training and validation sets, and it finds solution faster by avoiding retraining the classifier. However, this method is also dependent on classifiers. The filter method selects a subset according to a selection measure. Selection criteria include: distance measure [5,6], dependency measure [7,8], consistency measure [9], and information measure [10–12]. The filter method gives a subset achieving the balance between predicative power and computational cost. Moreover, it is independent on classifiers. So it is more practical comparing with the wrapper and embedded methods.

Most of attribute reduction methods just evaluate the performance of a subset according to predictive accuracy. However, approximate original class distribution is also an important evaluation rule [43]. Rough set based attribute reduction, the filter method with dependency measure, supports both rules. It is proposed by Pawlak [13–17] as a mathematical theory of set approximation, which is used in machine learning [18,19]. Rough set based attribute reduction finds particular subsets of conditional attributes providing the same information for classification purpose with the original set. This selection mechanism keeps the same class distribution with the original set. And its performance of predictive accuracy has been verified to be better or comparable with other methods in large amount of works. Moreover, rough set method has its own advantages. First, it needs no parameters. For general methods, they need take large computational cost to find a super parameter. It is impossible to assess the performance about all values of parameters. Second, it has explicit stopping criterion. The advantages of rough set come from that it deals with data in human-like fashion [44].

The advantages of rough set are obvious; however, its problem is computational complexity, which must be considered. The core issue of rough set theory is "discernibility function" taking $O(n^2)$ time complexity and $O(n^2 \times m)$ memory complexity, where n is the number of objects, and m is the number of attributes. Minimal reduction problem is even NP-hard [21], where the number of attributes is smallest among all possible reductions. Knowledge based methods have been proposed in the area of rough set

* Corresponding author. Tel.: +852 34429963.
E-mail address: bingli5-c@my.cityu.edu.hk (B. Li).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.07.006

[7,8,22–29], and each of them aims at its own requirement. According to their mechanisms, each method just finds a subset of attributes providing the same classification information with the original set, but no one can give a fair evaluation among these methods.

Propositional satisfiability problem (SAT) is one of the most studied NP-complete problems because of its significance in both theoretical research and practical application. Several SAT solvers [30,31,45] are employed in rough set based attribute reduction. However, there are several problems remaining to be done, including building discernibility matrix without any precondition and large addition of computational cost; a complete description of all solutions; analysis of factors playing role in computational cost about space and time. An improved discernibility function reduces the redundant attributes causing by the samples in the both Positive Region and Boundary Region. It just takes $O(n)$ addition of cost to overcome the redundancy. Moreover, the reasons of redundancy are shown by the proof of discernibility function. Extension rule [32,33] is a suitable tool to find solution of rough set reductions, which checks the satisfiability by using inverse of resolution. An important advantage of extension rule is that the combinations of attributes in inverse of resolution are smaller than the reductions, which have been verified in the experiments. So it is useful to save computational cost. By the results of discernibility function extended, the distribution of attribute reductions with different size is found [46]. It provides a new view to analyze attribute reduction based on rough set. And, according to the distribution, it is easy to find the minimal reduction. In this process, computational cost is also analyzed.

The rest of paper is organized as follows. The basic knowledge about attribution reduction using rough set is given in Section 2. Its relationship with SAT is proved in Section 3. The experimental results are presented in Section 4. Finally, the conclusion is drawn in Section 5.

2. Background

In this section, the basic notions [13–17,21,45] related to information system and rough set are shown.

Definition 2.1. Let $IS(U,A)$be an information system, where U is a nonempty finite set of objects and A is a nonempty finite set of attributes so that $f: U \to V_f$ for every $f \in A$. $V_f$is the set of values that f takes. For any $B \subseteq A$, an indiscernible relation $IND(B)$ is

$$IND(B) = \{(x,y) \in U^2 | \forall f \in B, f(x) = f(y)\} \quad (1)$$

Dataset can be seen as an information system, where samples are the objects of U and attributes are the elements of A [34].

Definition 2.2. A partition of U generated by $\{a^*\}$ is defined as

$$U/IND(\{a^*\}) = \{\{x \in U | a^*(x) = i\},\ i \in V_{a^*}\}, \quad (2)$$

where a^* is the decisional attribute.

If$(x,y) \in IND(B)$, x and y are indiscernible according to the subset B. The equivalence class of x on the B-indiscernible relation is denoted by $[x]_B$. If x and y are indiscernible according to the subset B, $y \in [x]_B$. Construct the B-lower approximations and B-upper approximations of X as

$$\underline{BX} = \{x | [x]_B \subseteq X\}, \quad (3)$$

$$\overline{BX} = \{x | [x]_B \cap X \neq \varnothing\}, \quad (4)$$

where (3) is the B-lower approximations, and (4) is the B-upper approximations.

By the definition of the B-lower approximations and B-upper approximations, the objects in U can be partition into three regions which are the Positive Region, the Boundary Region, and the Negative Region.

Definition 2.3. B-Positive Region, B-Boundary Region and B-Negative Region are defined as

$$POS_B(\{a^*\}) = \bigcup_{X \in U/IND(\{a^*\})} \underline{BX}, \quad (5)$$

$$BND_B(\{a^*\}) = \bigcup_{X \in U/IND(\{a^*\})} \overline{BX} - \bigcup_{X \in U/IND(\{a^*\})} \underline{BX}, \quad (6)$$

$$NEG_B(\{a^*\}) = U - \bigcup_{X \in U/IND(\{a^*\})} \overline{BX} \quad (7)$$

The three regions are defined with respect to $\{a^*\}$ which is the set of decisional attribute.

Definition 2.4. In an information system $IS = (U,A)$, an $n \times n$ matrix (c_{ij}) called *discernibility matrix* of IS is defined as

$$c_{ij} = \{a \in A : a(x_i) \neq a(x_j), x_i, x_j \in U\} \quad \text{for} \quad i,j = 1,\ldots,n \quad (8)$$

The discernibility matrix is denoted as $M(IS)$. It is straightforward to find $M(IS)$is symmetric and $c_{ii} = \varnothing$.

Definition 2.5. *Discernibility function* f_{IS} for an information system $IS = (U,A)$ is a Boolean function of m variables $a_1,\ldots,a_m$, defined as

$$f_{IS}(a_1,\ldots,a_m) = \wedge \{\vee (c_{ij}) : 1 \le j < i \le n,\ c_{ij} \neq \varnothing\}, \quad (9)$$

where a_i denotes an attribute in A and $\vee (c_{ij})$ is the disjunction of the variables inc_{ij}.

Example 2.1. A simple example represented in Table 1 is considered to show the discernibility matrix and discernibility function. For information system in Table 1, there are 5 objects and 5 attributes $\{a_1,a_2,a_3,a_4,a^*\}$. Table 2 shows the related discernibility matrix according to Definition 2.4. Then, the discernibility function can be found.

$$\begin{aligned} f_{IS} = {} & (a_1 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3 \vee a_4 \vee a^*) \wedge \\ & \times (a_1 \vee a_2 \vee a^*) \wedge a_2 \wedge (a_2 \vee a_4 \vee a^*) \\ & \times (a_2 \vee a_3 \vee a^*) \wedge (a_4 \vee a^*) \wedge (a_3 \vee a^*) \wedge (a_3 \vee a_4) \end{aligned}$$

3. Extension rule for attribute reductions

In this section, we prove that attribute reduction based on rough set can be solved by SAT with defining a^*-discernibility matrix. By employing the extension rule, the distribution of all

Table 1
Information system of Example 2.1.

U	a_1	a_2	a_3	a_4	a^*
1	0	1	1	1	1
2	1	1	0	1	1
3	1	0	0	1	1
4	1	0	0	0	0
5	1	0	1	1	0

Table 2
Discernibility matrix of Example 2.1.

Object	1	2	3	4	5
1		a_1,a_3	a_1,a_2,a_3	a_1,a_2,a_3,a_4,a^*	a_1,a_2,a^*
2	a_1,a_3		a_2	a_2,a_4,a^*	a_2,a_3,a^*
3	a_1,a_2,a_3	a_2		a_4,a^*	a_3,a^*
4	a_1,a_2,a_3,a_4,a^*	a_2,a_4,a^*	a_4,a^*		a_3,a_4
5	a_1,a_2,a^*	a_2,a_3,a^*	a_3,a^*	a_3,a_4	

Download English Version:

https://daneshyari.com/en/article/407022

Download Persian Version:

https://daneshyari.com/article/407022

Daneshyari.com