Expert Systems with Applications 41 (2014) 6748–6754

Contents lists available at ScienceDirect

Expert Systems with Applications

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

An enhancement for heuristic attribute reduction algorithm in rough set

Kai Zheng [a], Jie Hu [a,*], Zhenfei Zhan [b], Jin Ma [a], Jin Qi [a]

[a] Institute of Knowledge Based Engineering, School of Mechanical Engineering, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, PR China
[b] Department of Automobile Engineering, College of Mechanical Engineering, Chongqing University, 174 Shazheng Street, Chongqing 400044, PR China

ARTICLE INFO

Article history:
Available online 14 May 2014

Keywords:
Rough set
Heuristic
Attribute reduction
Enhancement for heuristic attribute reduction

ABSTRACT

Attribute reduction is one of the most important issues in the research of rough set theory. Numerous significance measure based heuristic attribute reduction algorithms have been presented to achieve the optimal reduct. However, how to handle the situation that multiple attributes have equally largest significances is still largely unknown. In this regard, an enhancement for heuristic attribute reduction (EHAR) in rough set is proposed. In some rounds of the process of adding attributes, those that have the same largest significance are not randomly selected, but build attribute combinations and compare their significances. Then the most significant combination rather than a randomly selected single attribute is added into the reduct. With the application of EHAR, two representative heuristic attribute reduction algorithms are improved. Several experiments are used to illustrate the proposed EHAR. The experimental results show that the enhanced algorithms with EHAR have a superior performance in achieving the optimal reduct.

1. Introduction

In machine learning and data mining fields, such databases constructed with a considerable number of attributes are often encountered. It is quite common to find some of these attributes are irrelevant or redundant, which not only occupy extensive computing resources, but also seriously impact the decision making process. For these reasons, it becomes natural to eliminate the irrelevant or redundant information and make the data set compact. Attribute reduction, also called feature selection, which is performed to refine an information system, has been extensively researched (see, e.g., Ahmad & Dey, 2005; Jia, Liao, Tang, & Shang, 2013; Susmaga, 2004; Tahir & Smith, 2010; Thangavel & Pethalakshmib, 2009; Wang, He, Chen, & Hu, 2014).

Rough set theory, first proposed by Pawlak (1982), can serve to deal with data classification problems by adopting the concept of equivalence classes. It provides a powerful tool for acquiring reducts of information system. Such reducts do not contain redundant data, but preserve the same classification ability as the original information system. The role of rough set is to refine an information system through removing redundant attributes. Many attribute reduction methods in rough set theory have been proposed for achieving reducts (see, e.g., Dai, 2013; Pawlak, 1997, 2002; Song & Li, 2013; Yao, 2008; Ye, Chen, & Ma, 2013), and have wide applications in many fields (see, e.g., Chai & Liu, 2014; Chen & Cheng, 2013; Fan, Liu, & Tzeng, 2007; Hu, Lin, & Han, 2004; Inuiguchi & Miyajima, 2007; Li & Wang, 2004; Teng, Wu, Sun, Zhou, & Liu, 2010).

The reducts of information systems usually may be not unique, and all the reducts can be acquired with the top-down attribute selection algorithm (Ma, 2009), or with the discernibility matrix method (see, e.g., Dai, 2013; Skowron & Rauszer, 1992). However, it has been proved that acquiring all the possible reducts or the minimum reduct is a NP-hard problem (Pawlak, 2004). Fortunately, in most of real-world applications, it is unnecessary to find all the reducts and sufficient to have only one of them (Hu et al., 2004). Generally the reduct with the least attributes is selected as the optimal one in the absence of other sources of information (Thangavel & Pethalakshmib, 2009). An increasing number of attribute reduction methods are developed to acquire only one reduct, rather than all (see, e.g., Deng, Yang, & Wang, 2012; Qian, Miao, Zhang, & Li, 2011).

Attribute significance based heuristic attribute reduction in rough set theory are one kind of the most common methods to achieve only one reduct, and have been extensively studied (see, e.g., Dash & Liu, 2003; Hu, Zhao, Xie, & Yu, 2007; Lee & Lee, 2006; Qian, Liang, & Dang, 2008; Song & Li, 2013; Tsai, Cheng, & Chang, 2006). They are usually implemented through a certain measure to evaluate the significance of attributes and a heuristic

* Corresponding author. Tel.: +86 21 3420 6552; fax: +86 21 3420 6313.
E-mail address: hujie@sjtu.edu.cn (J. Hu).

http://dx.doi.org/10.1016/j.eswa.2014.04.042
0957-4174/

searching strategy. They begin with a candidate for the reduct composed of an empty attribute set or an attribute core, and then focus on selecting attributes according to a certain selection criterion until the reduct achieves the same discrimination ability as the entire decision table does. The attribute significance measures can be dependence (see, e.g., Bhatt & Gopal, 2005; Pawlak, 1982) and consistency (see, e.g., Dash & Liu, 2003; Hu et al., 2007). However, such approaches probably converge to a local optimum, and the acquired reducts may be not the minimum ones, but fake optimal reducts.

Actually in real applications the optimal reduct has a special role, while other reducts may make a great deal of difference in the classification ability. In case a fake optimal reduct is treated as the optimal reduct, the complexity of induced rules may increase and improper decisions may be made. Therefore, the issue about how to acquire the optimal reduct of a decision table needs to be investigated.

According to the above analysis, this paper aims at proposing an enhancement for heuristic attribute reduction (EHAR) in rough set to modify the heuristic searching strategy, and finding the optimal reduct more effectively. An effective enhancement for improving the attribute significance based heuristic attribute reduction methods is devised in this paper, providing a means of effectively achieving the optimal reduct for the dependence based and consistency based heuristic reduction algorithms. By incorporating the enhancement into both of the above two representative heuristic algorithms, their improved versions are constructed. Numerical experiments demonstrate that both of the improved algorithms can effectively achieve the optimal reduct without a massive increase in time complexity.

The remainder of this paper is arranged as follows: Section 2 reviews a general background of rough set theory, particularly the above heuristic attribute reduction methods, followed by introducing the limitations of the existing heuristic algorithms in Section 3. The proposed EHAR is presented in Section 4. Several experiments are given to illustrate the validity of the EHAR in Section 5 and conclusion follows in Section 6.

2. Methodological background

2.1. Basic notions

In this subsection, some basic notions on rough set theory will be reviewed.

Definition 1. A decision table can be expressed as a four-tuple $S = (U,A,V,f)$, where U denotes a non-empty finite set of objects, A denotes the entire attribute set including the condition attribute set C and the decision attribute set D, V denotes a non-empty set of attribute values, and the function f is a Cartesian product of A and U into V.

Definition 2. Given a decision table $S = (U,C\bigcup D,V,f)$, an indiscernibility relation regarding to a non-empty attribute subset $B \subseteq C$ can be expressed as:

$$IND(B) = \{(x,y)|f(x,a) = f(y,a), \forall a \in B\} \tag{1}$$

Relation $IND(B)$ divides U into finite subsets, called equivalence class, thus constituting a partition, denoted by $U/IND(B)$. The B equivalence class of $x \in U$ is defined as $[x]_B = \{y \in U|(x,y) \in IND(B)\}$. Similarly, a partition of $U/IND(D) = \{Y_1,Y_2,\ldots,Y_n\}$ regarding to the decision attributes D can be realized.

Definition 3. Given a decision table $S = (U,C\bigcup D,V,f)$, $B \subseteq C$, the lower and upper approximations of $D_i(1 \leqslant i \leqslant n)$ from $IND(D)$ regarding to B can be respectively expressed as:

$$\begin{aligned} &\underline{B}(D_i) = \{x \in U|[x]_B \subseteq D_i\};\\ &\overline{B}(D_i) = \{x \in U|[x]_B \cap D_i \neq \emptyset\} \end{aligned} \tag{2}$$

Lower approximation reflects the set of objects that surely belong to D_i in regard to B; while upper approximation reflects the set of objects that possibly belong to D_i in regard to B.

Definition 4. Given a decision table $S = (U,C\bigcup D,V,f)$, $B \subseteq C$, the B-positive region of D can be expressed as:

$$POS_B(D) = \bigcup_{1\leqslant i\leqslant n} \underline{B}(D_i) \tag{3}$$

Definition 5. A attribute subset can be regarded as the core of a decision table if each attribute in the subset is indispensable and can not be removed without influencing the classification ability, i.e., for each attribute a in the attribute subset, $POS_{C-\{a\}}(D) < POS_C(D)$ should be satisfied.

The calculation of core attributes can be realized by judging whether $POS_{C-\{a\}}(D) < POS_C(D)$, where $a \in C$, if the inequality holds, a is one attribute of the core set, or else it is not. The detailed algorithms for obtaining the core can be referred to Ref. (Qian et al., 2011).

2.2. Representative heuristic attribute reduction algorithms

So far, some scholars have put forward several representative heuristic attribute reduction algorithms with different significance measures of attribute, such as dependency, consistency, information entropy and mutual information. This subsection will mainly describe the first two measures based heuristic attribute reduction algorithms.

2.2.1. Dependence based heuristic attribute reduction method

Definition 6. Given a decision table $S = (U,C\bigcup D,V,f)$, the dependence between D and C can be expressed as (Pawlak, 1991):

$$\gamma_C(D) = \frac{|POS_C(D)|}{|U|} \tag{4}$$

where $|\cdot|$ indicates the cardinality of one set and $0 \leqslant \gamma_C(D) \leqslant 1$. If $\gamma_C(D) = 0$, then D is independent on C, and if $0 < \gamma_C(D) < 1$, D is dependent partially on C in degree $\gamma_C(D)$, or else D is absolutely dependent on C.

Definition 7. Given a decision table $S = (U,C\bigcup D,V,f)$, $B \subseteq C$, the significance of attribute $a \in C - B$ can be expressed as:

$$SGF_1(a,B,D) = \gamma_{B\cup\{a\}}(D) - \gamma_B(D) \tag{5}$$

It can be seen that the larger the dependence changes when an attribute a is added to B, the more significant the added attribute is. Based on these definitions, the corresponding algorithm is shown as follows: