Information Sciences 326 (2016) 297–314

Contents lists available at ScienceDirect

Information Sciences

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

Efficient attribute reduction from the viewpoint of discernibility

Shu-Hua Teng [a], Min Lu [a], A-Feng Yang [a], Jun Zhang [a], Yongjian Nian [b], Mi He [b,*]

[a] *Science and Technology on Automatic Target Recognition Laboratory, National University of Defense Technology, Changsha 410073, China*
[b] *School of Biomedical Engineering, Third Military Medical University, Chongqing 400038, China*

ARTICLE INFO

Article history:
Received 23 May 2013
Revised 30 June 2015
Accepted 29 July 2015
Available online 8 August 2015

Keywords:
Rough set
Discernibility viewpoint
Attribute reduction
Attribute significance

ABSTRACT

Attribute reduction is an important preprocessing step in pattern recognition, machine learning and data mining. As an effective method for attribute reduction, rough set theory offers a useful and formal methodology. It retains the discernibility power of the original datasets; thus, attribute reduction has been extensively studied in rough set theory. However, the inefficiency of the existing attribute reduction algorithms limits the application of rough sets. In this paper, we first analyse the limitations of existing attribute reduction algorithms. Then, a novel measure of attribute quality, called the relative discernibility degree, is proposed based on the discernibility. Theoretical analysis shows that this measure can find relative dispensable attributes and remain unchanged after removing the relative dispensable attributes and redundant objects in the process of selecting attributes. This property can be used to reduce the search space and accelerate the heuristic process of attribute reduction. Consequently, a new attribute reduction algorithm is proposed from the viewpoint of discernibility. Furthermore, the relationships among the reduction definitions of the algebra view, information view and discernibility view are derived. Some non-equivalent relationships among these views of rough set theory in inconsistent decision tables are discovered. A set of numerical experiments was conducted on UCI datasets. Experimental results show that the proposed algorithm is effective and efficient and is applicable to the case of large-scale datasets.

1. Introduction

In recent years, the size of datasets has grown dramatically in terms of both the number of instances and the number of features, such as medical data, text classification, etc. Extracting useful information from the rapidly expanding amount of data is an important and challenging issue. As a preprocessing step in data mining, attribute reduction is very effective in reducing the dimensionality of a feature space, increasing predictive accuracy and improving comprehensibility of the induced models. Therefore, it has been extensively investigated in past decades [1–3].

Rough set theory (RST) [4] is a valid mathematic tool to handle inexact and uncertain information. It provides a formal framework for data mining. One of the most focused on applications of RST is attribute reduction, which is considered an important branch of feature selection methods [1,5–10]. In the framework of rough sets, a reduct is defined as a minimal attribute subset that has the same discernibility as the entire attribute set. As an effective method of feature selection, attribute reduction can select the features that encode the most significant information in a dataset without transforming the data. Meanwhile, it minimizes the loss of information during the selection process. Specifically, it retains the semantics of the original datasets [11].

* Corresponding author. Tel.: +86 15808032997.
E-mail address: hmcherry@126.com (M. He).

http://dx.doi.org/10.1016/j.ins.2015.07.052
0020-0255/

Several methods have been proposed for finding the set of all reducts or a single reduct [12–18]. Skowron and Rauszer [19] proposed an attribute reduction algorithm to find all reducts using a discernibility matrix. This method provides a mathematical foundation to investigate attribute reduction [20,21]. To compress the discernibility function of a decision table, Chen et al. [22] proposed a sample pair selection algorithm (SPSA) to find all reducts or one reduct. The algorithm can find a proper reduct and is effective as a preprocessing step to find reducts. It is known that finding all the reducts from a decision table is computationally expensive. In contrast, a single reduct is sufficient in most applications [23]. Thus, many heuristic methods for finding one reduct have also been proposed [22,24–27]. Shen et al. [28,29] developed a quick reduct algorithm (QRA) to compute a reduct by keeping the dependency function invariant. It is an efficient algorithm and widely used in real applications. However, the dependency function cannot well reflect the attribute importance [1,30,31], which leads to more randomness in the process of selecting attributes. Therefore, in most cases, QRA may not yield a reduct but instead a super-reduct, which contains a reduct as a subset of it. Based on this observation, an improved quick reduct algorithm (IQRA) [30] was proposed by introducing variable precision rough sets (VPRS) in the process of attribute selection. However, the search process based on VPRS in IQRA increases the time of iteration, and the problem of QRA still exists in IQRA. Both QRA and IQRA are reduct methods from the viewpoint of algebra; some researchers have also investigated attribute reduction from the viewpoint of information theory [24,27,32]. The relationship between the definitions of attribute reduction in the algebra view and information view was presented in [33]. The information view-based method is suitable for small-scale data sets, but it is very time-consuming when tested on high-dimensional data sets.

Finding reducts in large datasets is very challenging in RST [34,35]. The above attribute reduction methods are commonly computationally expensive, and they are not acceptable in the case of large-scale datasets. To improve the efficiency of reduct algorithms, attribute reduction algorithms under the algebra and information viewpoints in RST have been enhanced by filtering out redundant objects [36]. However, the enhanced algorithms in [36] only reduce the computation time to a certain extent. Because they choose the same attribute reduct as the original version, the aforementioned problem still exists. Moreover, the enhanced algorithms only aim to reduce redundant objects in the attribute selection process. They do not reduce the redundancy attributes. It has been observed that the number of attributes in datasets can also significantly affect the efficiency of attribute reduction. Motivated by these observations, this paper further improves the performance of heuristic attribute reduction methods by gradually reducing both the size of the universe and the number of attributes in each iteration of attribute reduction. The computational complexity of finding reducts can therefore be reduced.

This paper further improves the efficiency of heuristic reduct algorithms. This paper has three major contributions, which are summarized as follows. First, the limitations of existing attribute reduction algorithms are analysed. We observed that the attribute reduction algorithms from both algebra and information viewpoints select attributes in a random manner, which leads to high computational complexity when tested on large-scale data sets. Second, a novel measure for attribute quality is proposed. Using this measure, we can either remove the relative dispensable attributes or filter out the redundant objects in the process of reduct computation. Therefore, the search space is significantly reduced. Subsequently, a novel attribute reduction algorithm is proposed from the viewpoint of discernibility, which overcomes the limitations of existing reduction algorithms. Moreover, the relationships among the concepts of reducts from the viewpoints of algebra, information and discernibility are discussed. Third, the performance of our algorithm is compared with that of the discernibility matrix, algebra and information viewpoints on the UCI datasets. Numerical experiments show that the proposed algorithm obtained the smallest number of selected attributes in the shortest time in most cases. It achieved higher classification accuracy and can be applied to large-scale data sets with large numbers of attributes or objects.

The rest of this paper is organized as follows. Basic concepts from the viewpoints of algebra and information in RST are briefly reviewed in Section 2. The limitations of the existing attribute reduction algorithms are discussed in Section 3. The relative discernibility degree and its main properties are discussed in Section 4, together with a novel heuristic attribute reduction algorithm from the viewpoint of discernibility. In Section 5, we further study the relationships among the attribute reductions from the algebra viewpoint, information viewpoint and discernibility viewpoint of RST. Experimental analysis is presented in Section 6. Section 7 concludes this paper.

2. Preliminaries

2.1. Preliminary concepts of RST

An information system can be represented by $S = (U, A)$, where $U = \{x_1, x_2, \ldots, x_{|U|}\}$ is a non-empty finite set of objects ($|\cdot|$ denotes the cardinality of the set) and $A = \{a_1, a_2, \ldots, a_{|A|}\}$ is a non-empty finite set of attributes such that $a_j{:}a_j \to V_{a_j}$ for each $a_j \in A$. The set V_{a_j} is called the value set of a_j.

Each subset of attributes $P \subseteq A$ determines a binary indiscernibility relation $\mathrm{IND}(P)$, as follows:

$$\mathrm{IND}(P) = \{(x, y) \in U \times U | \forall a \in P, f(x, a) = f(y, a)\}.$$

Obviously, $\mathrm{IND}(P)$ is an equivalence relation. If $(x, y) \in \mathrm{IND}(P)$, then x and y are indiscernible by P. The partition generated by $\mathrm{IND}(P)$ is denoted by $U/\mathrm{IND}(P)$, which is further abbreviated as U/P. $U/P = \{P_1, P_2, \ldots, P_m\}$ denotes knowledge associated with the equivalence relation $\mathrm{IND}(P)$, where P_i is the equivalence class, $1 \le i \le m$, $1 \le m \le |U|$. Each equivalence class is called an information granule. The attribute set P is therefore, called knowledge P. The equivalence class determined by x with respect to (wrt) attribute set P is denoted by $[x]_P = \{y \in U | (x, y) \in \mathrm{IND}(P)\}$. If $x \in P_i$, then $[x]_P = P_i$.