Artificial Intelligence 174 (2010) 597–618

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Positive approximation: An accelerator for attribute reduction in rough set theory

Yuhua Qian [a,c], Jiye Liang [a,*], Witold Pedrycz [b], Chuangyin Dang [c]

[a] Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Taiyuan, 030006, Shanxi, China
[b] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada
[c] Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 15 July 2009
Received in revised form 6 April 2010
Accepted 7 April 2010
Available online 9 April 2010

Keywords:
Rough set theory
Attribute reduction
Decision table
Positive approximation
Granular computing

ABSTRACT

Feature selection is a challenging problem in areas such as pattern recognition, machine learning and data mining. Considering a consistency measure introduced in rough set theory, the problem of feature selection, also called attribute reduction, aims to retain the discriminatory power of original features. Many heuristic attribute reduction algorithms have been proposed however, quite often, these methods are computationally time-consuming. To overcome this shortcoming, we introduce a theoretic framework based on rough set theory, called positive approximation, which can be used to accelerate a heuristic process of attribute reduction. Based on the proposed accelerator, a general attribute reduction algorithm is designed. Through the use of the accelerator, several representative heuristic attribute reduction algorithms in rough set theory have been enhanced. Note that each of the modified algorithms can choose the same attribute reduct as its original version, and hence possesses the same classification accuracy. Experiments show that these modified algorithms outperform their original counterparts. It is worth noting that the performance of the modified algorithms becomes more visible when dealing with larger data sets.

1. Introduction

Feature selection, also called attribute reduction, is a common problem in pattern recognition, data mining and machine learning. In recent years, we encounter databases in which both the number of objects becomes larger and their dimensionality (number of attributes) gets larger as well. Tens, hundreds, and even thousands of attributes are stored in many real-world application databases [6,12,37]. Attributes that are irrelevant to recognition tasks may deteriorate the performance of learning algorithms [44,45]. In other words, storing and processing all attributes (both relevant and irrelevant) could be computationally very expensive and impractical. To deal with this issue, as was pointed out in [20], some attributes can be omitted, which will not seriously impact the resulting classification (recognition) error, cf. [20]. Therefore, the omission of some attributes could not only be tolerable but even desirable relatively to the costs involved in such cases [32].

In feature selection, we encounter two general strategies, namely wrappers [16] and filters. The former employs a learning algorithm to evaluate the selected attribute subsets, and the latter selects attributes by being guided by some significance measures such as information gain [23,46], consistency [6,41], distance [15], dependency [30], and others. These

* Corresponding author. Tel.: +86 0351 7018176; fax: +86 0351 7018176.
E-mail addresses: jinchengqyh@sxu.edu.cn (Y.H. Qian), ljy@sxu.edu.cn (J.Y. Liang), pedrycz@ee.ualberta.ca (W. Pedrycz), mecdang@cityu.edu.hk (C.Y. Dang).

0004-3702/$ – see front matter
doi:10.1016/j.artint.2010.04.018

measures can be divided into two main categories: distance-based measures and consistency-based measures [20]. Rough set theory by Pawlak [33–36] is a relatively new soft computing tool for the analysis of a vague description of an object, and has become a popular mathematical framework for pattern recognition, image processing, feature selection, neurocomputing, data mining and knowledge discovery from large data sets [7,11,31]. Attribute reduction in rough set theory offers a systematic theoretic framework for consistency-based feature selection, which does not attempt to maximize the class separability but rather attempts to retain the discernible ability of original features for the objects from the universe [13,14,53].

Generally speaking, one always needs to handle two types of data, viz. those that assume numerical values and symbolic values. For numerical values, there are two types of approaches. One relies on fuzzy rough set theory, and the other is concerned with the discretization of numerical attributes. In order to deal with numerical attributes or hybrid attributes, several approaches have been developed in the literature. Pedrycz and Vukovich regarded features as granular rather than numerical [37]. Shen and Jenshen generalized the dependency function in classical rough set framework to the fuzzy case and proposed a fuzzy-rough QUICKREDUCT algorithm [13,14,48]. Bhatt and Gopal provided a concept of fuzzy-rough sets formed on compact computational domain, which is utilized to improve the computational efficiency [3,4]. Hu et al. presented a new entropy to measure of the information quantity in fuzzy sets [21] and applied this particular measure to reduce hybrid data [22]. Data discretization is another important approach to deal with numerical values, in which we usually discretize numerical values into several intervals and associate the intervals with a set of symbolic values, see [5,28]. In the "classical" rough set theory, the attribute reduction method takes all attributes as those which assume symbolic values. Through preprocessing of original data, one can use the classical rough set theory to select a subset of features that is the most suitable for a given recognition problem.

In the last twenty years, many techniques of attribute reduction have been developed in rough set theory. The concept of the β-reduct proposed by Ziarko provides a suite of reduction methods in the variable precision rough set model [60]. An attribute reduction method was proposed for knowledge reduction in random information systems [57]. Five kinds of attribute reducts and their relationships in inconsistent systems were investigated by Kryszkiewicz [18], Li et al. [24] and Mi et al. [29], respectively. By eliminating some rigorous conditions required by the distribution reduct, a maximum distribution reduct was introduced by Mi et al. in [29]. In order to obtain all attribute reducts of a given data set, Skowron [49] proposed a discernibility matrix method, in which any two objects determine one feature subset that can distinguish them. According to the discernibility matrix viewpoint, Qian et al. [42,43] and Shao et al. [47] provided a technique of attribute reduction for interval ordered information systems, set-valued ordered information systems and incomplete ordered information systems, respectively. Kryszkiewicz and Lasek [17] proposed an approach to discovery of minimal sets of attributes functionally determining a decision attribute. The above attribute reduction methods are usually computationally very expensive, which are intolerable for dealing with large-scale data sets with high dimensions. To support efficient attribute reduction, many heuristic attribute reduction methods have been developed in rough set theory, cf. [19,20,22,25,26,39,52, 54–56]. Each of these attribute reduction methods can extract a single reduct from a given decision table.[1] For convenience, from the viewpoint of heuristic functions, we classify these attribute reduction methods into four categories: positive-region reduction, Shannon's entropy reduction, Liang's entropy reduction and combination entropy reduction. Hence, we review only four representative heuristic attribute reduction methods.

(1) Positive-region reduction

The concept of positive region was proposed by Pawlak in [33], which is used to measure the significance of a condition attribute in a decision table. While the idea of attribute reduction using positive region was originated by J.W. Grzymala-Busse in [9] and [10], and the corresponding algorithm ignores the additional computation required for selecting significant attributes. Then, Hu and Cercone [19] proposed a heuristic attribute reduction method, called positive-region reduction, which remains the positive region of target decision unchanged. The literature [20] gave an extension of this positive-region reduction for hybrid attribute reduction in the framework of fuzzy rough set. Owing to the consistency of ideas and strategies of these methods, we regard the method from [19] as their representative. These reduction methods are the first attempt to heuristic attribute reduction algorithms in rough set theory.

(2) Shannon's entropy reduction

The entropy reducts have first been introduced in 1993/1994 by Skowron in his lectures at Warsaw University. Based on the idea, Slezak introduced Shannon's information entropy to search reducts in the classical rough set model [50–52]. Wang et al. [54] used conditional entropy of Shannon's entropy to calculate the relative attribute reduction of a decision information system. In fact, several authors also have used variants of Shannon's entropy or mutual information to measure uncertainty in rough set theory and construct heuristic algorithm of attribute reduction in rough set theory [22,55,56]. Here

[1] The attribute reduct obtained preserves a particular property of a given decision table. However, as Prof. Bazan said, from the viewpoint of stability of attribute reduct, the selected reduct may be of bad quality [1,2]. To overcome this problem, Bazan developed a method for dynamic reducts to get a stable attribute reduct from a decision table. How to accelerate the method for dynamic reducts is an interesting topic in further work.