Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 258 (2015) 61–78

FUZZY sets and systems

www.elsevier.com/locate/fss

Fuzzy-rough feature selection accelerator ☆

Yuhua Qian [a,b,*], Qi Wang [a], Honghong Cheng [a], Jiye Liang [a], Chuangyin Dang [b]

[a] *Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan, 030006, Shanxi, China*

[b] *Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Hong Kong*

Received 6 August 2013; received in revised form 8 April 2014; accepted 10 April 2014

Available online 22 May 2014

Abstract

Fuzzy rough set method provides an effective approach to data mining and knowledge discovery from hybrid data including categorical values and numerical values. However, its time-consumption is very intolerable to analyze data sets with large scale and high dimensionality. Many heuristic fuzzy-rough feature selection algorithms have been developed however, quite often, these methods are still computationally time-consuming. For further improvement, we propose an accelerator, called forward approximation, which combines sample reduction and dimensionality reduction together. The strategy can be used to accelerate a heuristic process of fuzzy-rough feature selection. Based on the proposed accelerator, an improved algorithm is designed. Through the use of the accelerator, three representative heuristic fuzzy-rough feature selection algorithms have been enhanced. Experiments show that these modified algorithms are much faster than their original counterparts. It is worth noting that the performance of the modified algorithms becomes more visible when dealing with larger data sets.

Keywords: Rough sets; Fuzzy rough sets; Feature selection; Forward approximation; Accelerator; Granular computing

1. Introduction

There are many factors that motivate the inclusion of a feature selection step in a variety of fields, such as data mining, machine learning and pattern recognition, which addresses the problem of selecting those input features that are most predictive of a given outcome [30,33,34,41]. Databases expand quickly not only in the rows (objects) but also in the columns (features) nowadays [3]. In recent several years, big data analysis has become a new hot topic. For a task of data analysis, a given data set is called big data if it cannot be efficiently processed via existing methods. In some tasks of data analysis, some of features are irrelevant to the learning or problem solving. It is likely that the

☆ This is an extended version of the paper presented at 2011 International Conference of Rough Sets and Knowledge Technology, Banff, 2011, Canada.

* Corresponding author at: Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan, 030006, Shanxi, China. Tel./fax: +86 0351 7010566.

E-mail addresses: jinchengqyh@126.com (Y. Qian), wqking@163.com (Q. Wang), chhsxdx@163.com (H. Cheng), ljy@sxu.edu.cn (J. Liang), mecdang@cityu.edu.hk (C. Dang).

http://dx.doi.org/10.1016/j.fss.2014.04.029
0165-0114/

omission of some features will not seriously increase the error probability. In such cases, the loss of optimality may not only be tolerable but even desirable relative to the costs involved.

In the framework of rough set theory, feature selection is also called attribute reduction [8,43,44], which preserves the original meaning of the features after reduction [45]. The classical rough set model, proposed by Pawlak [22,23], is based on crisp equivalence relations and crisp equivalence classes. It is only applicable to categorical attribute reduction and knowledge discovery. In order to deal with numerical and categorical data (or a mixture of both) in data sets, fuzzy rough set model was first proposed by Dübois and Prade [5], which combines rough set and fuzzy set together. The lower/upper approximation in these fuzzy rough set models tries to give a membership function of each object to a set. As Dübois and Prade defined, if a fuzzy set is approached by a family of crisp sets in the same universe, then the corresponding lower/upper approximation pair is called a rough fuzzy set; and if a crisp/fuzzy set is approached by a family of fuzzy sets in the same universe, then the corresponding lower/upper approximation pair is called a fuzzy rough set. To widely apply the fuzzy rough set method, many extended versions and relative applications have been developed, cf. [11,12,14,20,21,24,25,31,35,36,39,40,42,46–48]. In particular, to keep the same form as classical rough set by Pawlak, Hu et al. [11] proposed a novel fuzzy rough model with a crisp lower/upper approximation. In fact, in the new model, the lower approximation and the upper approximation can be seen as the 1-cut/strong 0-cut of original counterparts in Dübois's model, respectively. Taking the same idea into account, Wang et al. [36] developed a generalized fuzzy rough model in which a β-cut is used to define its lower/upper approximation. These two methods have a consistent form with Pawlak's rough set, and their lower/upper approximations induced by a given cut are crisp approximations rather than fuzzy approximations. According to Dübois and Prade's definition, each of these rough set models is a fuzzy rough set.

Attribute reduction using fuzzy rough sets is often called fuzzy-rough feature selection. To support efficient feature selection, many heuristic algorithms have been developed in fuzzy rough set theory, cf. [2,4,10,11,13,15–17,36]. Each of these feature selection methods can extract a single reduct from a given decision table. For convenience, from the viewpoint of heuristic functions, we classify these feature selection methods into two categories: fuzzy positive region reduction and fuzzy information entropy reduction. Hence, we only review two kinds of representative heuristic fuzzy-rough feature selection methods.

(1) Fuzzy positive region reduction

The concept of positive region was proposed by Pawlak in [22], which is used to measure the significance of a condition attribute in a decision table. Then, Hu and Cercone [9] proposed a heuristic attribute reduction method, called positive region reduction, which remains the positive region of target decision unchanged. Under Dübois's fuzzy rough set model, Jensen and Shen [15–17] developed a series of heuristic fuzzy-rough feature selection algorithms based on fuzzy positive region. Bhatt and Gopal [2] proposed a modified version to improve computational efficiency. Under Hu's fuzzy rough set model and Wang's fuzzy rough set model, Hu et al. [11] extended the method from the literature [9] to select a feature subset from hybrid data. Owing to the consistency of ideas and strategies of these methods, we regard the method from [11] as their representative.

(2) Fuzzy information entropy reduction

The entropy reducts have first been introduced in 1993/1994 by Skowron in his lectures at Warsaw University. Wang et al. [37] used conditional entropy of Shannon's entropy to calculate the relative attribute reduction of a decision information system. Hu et al. extended the entropy to measure the information quantity in fuzzy sets and applied its conditional entropy to feature selection from hybrid data [13]. This reduction method remains the conditional entropy of a target decision unchanged. The fuzzy information entropy is an important approach to characterizing the uncertainty of a fuzzy binary relation, which can be used to select a feature subset from a given big data set [13,14].

Each of these above methods preserves a particular property of a given decision table. However, these above methods are still computationally very expensive, which are intolerable for dealing with large-scale data sets with high dimensions. So, this kind of attribute reduction problems can be regarded as data analysis of big data. The objective of this study is to focus on how to improve the time efficiency of a heuristic fuzzy-rough feature selection algorithm.