ARTICLE IN PRESS

FSS:7078

Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems ••• (••••) •••–•••

FUZZY sets and systems

www.elsevier.com/locate/fss

Fuzzy rough set based incremental attribute reduction from dynamic data with sample arriving

Yanyan Yang [a,c,*], Degang Chen [b], Hui Wang [c], Eric C.C. Tsang [d] Deli Zhang [e]

[a] *School of Control and Computer Engineering, North China Electric Power University, Beijing, PR China*
[b] *Department of Mathematics and Physics, North China Electric Power University, Beijing, PR China*
[c] *School of Computing and Mathematics, University of Ulster, Northern Ireland, UK*
[d] *The Faculty of Information Technology, Macau University of Science and Technology, Taipa, Macau*
[e] *Department of Mathematics, Jilin Provincial Institute of Education, Changchun, PR China*

Received 11 March 2016; received in revised form 17 July 2016; accepted 3 August 2016

Abstract

Attribute reduction with fuzzy rough set is an effective technique for selecting most informative attributes from a given real-valued dataset. However, existing algorithms for attribute reduction with fuzzy rough set have to re-compute a reduct from dynamic data with sample arriving where one sample or multiple samples arrive successively. This is clearly uneconomical from a computational point of view. In order to efficiently find a reduct from such datasets, this paper studies incremental attribute reduction with fuzzy rough sets. At the arrival of one sample or multiple samples, the relative discernibility relation is updated for each attribute. On the basis of the updated relation, an insight into the incremental process of attribute reduction with fuzzy rough sets is gained to reveal how to add new attributes into the current reduct and delete existing attributes from the current reduct. Applying the incremental process, two incremental algorithms for attribute reduction with fuzzy rough sets are presented for one incoming sample and multiple incoming samples, respectively. Experimental comparisons with several non-incremental algorithms and the proposed incremental algorithm for one incoming sample show that our proposed incremental algorithm for multiple incoming samples can efficiently find one reduct with a comparable classification accuracy.

Keywords: Fuzzy rough set; Attribute reduction; Incremental learning; Dynamic data; Relative discernibility relation

1. Introduction

Fuzzy rough set theory [1], as one generalization of rough set theory [2], provides a powerful tool for modeling and handling indiscernibility and fuzziness in datasets with real-valued condition attributes. In such a dataset, there may be an inconsistency between condition attributes and decision labels, i.e., two or more samples have similar or the same condition attribute values but different decision labels. It has been pointed out in [3] that fuzzy rough sets

* Corresponding author at: School of Control and Computer Engineering, North China Electric Power University, Beijing, PR China.
E-mail address: hrsinbba@163.com (Y. Yang).

http://dx.doi.org/10.1016/j.fss.2016.08.001

address this inconsistency by discerning two samples to a certain degree related to decision labels. Attribute reduction with fuzzy rough sets aims to obtain a subset of condition attributes to keep this discerning.

Attribute reduction with fuzzy rough sets has been studied extensively, leading to a large variety of algorithms for finding reducts. A pioneering work on attribute reduction with fuzzy rough sets was proposed in [4] by defining a dependence function to measure the quality of attributes with fuzzy rough sets in [5,6]. An algorithm for computing a reduct was developed in [4] to keep the dependence function unchanged. The existing researches on attribute reduction with fuzzy rough sets mainly pay attention to improve the method in [4]. For example, Bhatt et al. [7] improved the method of [4] by defining the lower approximation operator in [5] on a compact computational domain. In [8], Hu et al. proposed an attribute reduction algorithm by using information entropy to measure the significance of attributes. In [9], a fuzzy discernibility matrix was proposed to search for a reduct. It has been noted in [3], however, these heuristic algorithms cannot find a proper reduct but an over-reduct or under-reduct due to their stop criteria. To find proper reducts, the method of discernibility matrix was developed in [10], by which a discernibility function is constructed and all reducts can be found with its disjunctive form. Although this approach provides a mathematical foundation for the research on attribute reduction with fuzzy rough sets, it requires a heavy computational load. It has been observed in [3] that only minimal elements in the discernibility matrix are sufficient to find reducts. This fact motivates the authors in [3] to develop the algorithms for computing minimal elements and reducts. The experiments in [3] have shown the effectiveness of the proposed algorithms.

The previous algorithms for attribute reduction with fuzzy rough sets perform well in some practical applications, but they are not designed to handle dynamic datasets with sample arriving where one sample or multiple samples arrive successively. At the arrival of new samples, these traditional algorithms have to re-compute a reduct from the whole new dataset including the accumulated samples and new incoming samples, since they have no the explicit scheme of fully utilizing the previous data information from the accumulated samples. As a result, they tend to consume a huge amount of computational time when dealing with dynamic datasets. One alternative solution is to apply the incremental technique to update a reduct dynamically, avoiding some re-computations. The focus of this paper is thus on how to study incremental attribute reduction to improve the time efficiency of the traditional algorithms for attribute reduction with fuzzy rough sets.

There are many research works on incremental attribute reduction with rough sets, which can be categorized along the following three variations: the attribute set [11–13], attribute values [14–16] and samples [17–19]. When adding and deleting attributes, Zeng et al. in [20] analyzed the updating mechanisms of attribute reduction and proposed two incremental algorithms for feature selection with fuzzy rough sets. With dynamically-increasing attributes, Wang et al. [21] presented the updating mechanisms of three measures of information entropy, and developed a dimension incremental strategy for attribute reduction based on the updating mechanisms. When an attribute set is added into and deleted from the incomplete decision systems, Shu et al. [22] proposed two algorithms for updating attribute reduction based on the incremental computation of the positive region in incomplete decision systems. With dynamically varying attribute values, Wang et al. [21] developed an incremental algorithm for attribute reduction based on the incremental computation of three representative measures of information entropy. Based on the incremental computation of the positive region, Shu et al. [23] developed two incremental attribute reduction algorithms for single sample and multiple samples with varying attribute values.

With the variation of samples, several algorithms for attribute reduction were developed in the framework of rough sets. For example, Liu [24] proposed an incremental attribute reduction algorithm to find the minimal reduct. However, it is only applicable for information systems without the decision attribute. For decision tables, two incremental algorithms were proposed in [25,26] respectively, but experimental results in [27] show that both of them have high time complexity. To improve the efficiency of the two algorithms in [25,26], Hu et al. [27] presented an incremental attribute reduction algorithm based on the positive region. Based on the modified discernibility matrix, Hu et al. [28] proposed an incremental algorithm for finding all reducts when adding a new sample into the current dataset. At the arrival of a new sample, Yang [29] proposed an incremental attribute reduction algorithm by updating the discernibility matrix. Guan [30] proposed an incremental algorithm for updating all reducts based on the discernibility matrix. Feng et al. [31] employed the incremental computation for computing attribute core to improve the efficiency of computing a reduct. When a sample set of incomplete decision system is dynamically increasing, Shu et al. [32,33] presented an incremental attribute reduction algorithm to compute one reduct. At the arrival of a new sample, Chen et al. [34] gained an insight into the incremental process of attribute reduction with variable precision rough sets to reveal how to add new attributes into and delete existing attributes from a current reduct. Based on the incremental process, Chen et