ARTICLE IN PRESS

Information Sciences xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Information Sciences

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

An entropy-based uncertainty measurement approach in neighborhood systems

Yumin Chen [a,b,*], Keshou Wu [a], Xuhui Chen [a], Chaohui Tang [a], Qingxin Zhu [b]

[a] Department of Computer Science and Technology, Xiamen University of Technology, Xiamen 361024, China
[b] Department of Computer Science and Technology, University of Electronic Science and Technology, Chengdu 611731, China

ARTICLE INFO

Article history:
Received 21 September 2012
Received in revised form 16 March 2014
Accepted 23 March 2014
Available online xxxx

Keywords:
Rough set theory
Neighborhood system
Uncertainty measurement
Information entropy

ABSTRACT

Uncertainty measures can provide us with principled methodologies to analyze uncertain data and unveil the substantive characteristics of the data sets. Accuracy and roughness proposed by Pawlak are two main tools to deal with uncertainty measurement issue in rough set theory. Many uncertainty measure methodologies for discrete-valued information systems or discrete-valued decision systems have been developed. However, there are only limited on the uncertainty measurement for neighborhood systems. In this paper, we address the issues of uncertainty of a neighborhood system and extend the traditional accuracy and roughness measures to deal with neighborhood systems. In particular, a concept called neighborhood entropy is first introduced to evaluate the uncertainty of a neighborhood information system. Consequently, the entropy-based roughness and approximation roughness measures of neighborhood system are presented. Theoretical analysis indicates that entropy-based measures can be used to evaluate the uncertainty in neighborhood systems. Experiments are conducted on artificial data sets and standard UCI data sets to test our proposed methodologies. Results demonstrate that the entropy-based measures are effective and valid for evaluating the uncertainty of neighborhood systems.

1. Introduction

Rough set theory [32,33], proposed by Pawlak, is a valid mathematical tool to deal with uncertain, inaccurate and fuzzy data. Many achievements have been made in rough set theory. For example, Grzymala-Busse [11] developed a rule induction system LERS, which can handle inconsistencies and induce both certain and possible rules. Huang et al. [15] proposed an incremental rule-extraction algorithm to deal with the problems of large database. Dubois and Prade [9] combined rough and fuzzy set theory to define the first fuzzy rough sets, which employed the min and max fuzzy operators to describe the fuzzy lower and upper approximations. Skowron et al. [42] worked on the relation and the combination of rough set theory and granular computing [58]. Lin [25] proposed a granular computing model based on binary relations. Pedrycz [34] proposed granular computing theory for the analysis and design of intelligent systems. Yao [51,52] studied three-way decisions in probabilistic rough set model. Rough set theory has become a useful tool for many application such as feature selection, pattern recognition, image processing, data mining, outlier detection and knowledge discovery from large data sets [5,6,18,27,28,35,38,45,48,57,59–62].

* Corresponding author at: Department of Computer Science and Technology, Xiamen University of Technology, Xiamen 361024, China. Tel.: +86 13799762651.
E-mail address: cym0620@163.com (Y. Chen).

http://dx.doi.org/10.1016/j.ins.2014.03.117
0020-0255/

Please cite this article in press as: Y. Chen et al., An entropy-based uncertainty measurement approach in neighborhood systems, Inform. Sci. (2014), http://dx.doi.org/10.1016/j.ins.2014.03.117

Recently, many measures for uncertainty have been proposed in rough set theory due to the existence of many fuzzy data. Pawlak [33] originally proposes four numerical uncertainty measurement methods, which are accuracy and roughness to evaluate uncertainty of a rough set in information systems, as well as approximation accuracy and approximation roughness of a rough classification in decision systems. Several authors have studied the uncertainty measures of data sets from other viewpoints. Based on information granularity, a measure of uncertainty of a set in an information system and approximation accuracy of a rough classification in a decision table was proposed by Liang et al. [20,22]. Dai et al. [7] have extended the approximate accuracy approach into an incomplete information system. The measurement method of information entropy and its applications were adapted for rough set by many authors [1,2,8,10,19,21,31,36,37,39,44,47,50]. Dntsch and Gediga [10] defined the information entropy and three kinds of conditional entropies in rough sets for predicting a decision attribute. Miao et al. [31,47] proposed two heuristic attribute reduction methods, which are based on information entropy in an information table and mutual entropy in a decision table. Beaubouef et al. [1] proposed a method to measure uncertainty of relation databases based on rough entropy. Zhu et al. [64] developed a new pair of information theoretic entropy and co-entropy functions associated to partitions and approximations. Yao et al. [50,53,56] studied several kinds of information entropy measures for attribute importance in rough set theory. Liang et al. [19] proposed a new method for evaluating both uncertainty and fuzziness. Qian and Liang [36] proposed a combination entropy for evaluating uncertainty of a knowledge from an information system. However, the methods mentioned above are mainly based on discrete information systems or decision systems.

Traditional rough set theory is suitable for discrete data rather than continues data since only equivalence class and equivalence relation are considered. The requirement of equivalence relation is too restrictive for many practical data sets and applications. In view of this, equivalence relation has been extended to similarity relation [43], tolerance relation [4,30,41], and even arbitrary binary relation of the traditional rough sets [17,54,55]. An example extension is covering-based rough set model [3,67], which is based on the relaxation of the partition arising from equivalence relation to a covering. The covering of a universe is used to construct the lower and upper approximations of any subset of the universe. It is well-known that coverings are a fundamental concept in topological spaces and play an important role in the study of topological properties. This motivates the research of covering rough sets from the topology point of view. Zhu and Wang [65,66] examined the topological properties of the lower and upper approximation operations for covering-based rough sets. Wu et al. [46] combined the notion of topological spaces into rough sets and then discussed the properties of topological rough spaces. In fact, neighborhoods, another elementary concept in topology, have been used to define the lower and upper approximations; some properties of approximation operations for this type of covering rough sets have been explored as well. Lin [23,24] pointed out that neighborhood spaces are more general topological spaces than equivalence spaces and introduced neighborhood relation into rough set theory. Yao [49] discussed the properties of neighborhood approximation spaces. Hu et al. [12] constructed a unified theoretical framework for a neighborhood-based classifier using a neighborhood-based rough set model and a forward feature set selection algorithm towards hybrid data. The neighborhood-based rough set model is a powerful tool to attribute reduction, feature selection, classification and reasoning with uncertainty [13,14,16,26,29,63]. So far, however, there are only limited studies on the uncertainty measurement issue for neighborhood systems. In this paper, we address the uncertainty measurement issue in neighborhood systems and propose effective uncertainty measures for neighborhood systems. A similarity relation with a neighborhood parameter of neighborhood system is given, under which the concept of neighborhood entropy is introduced to neighborhood system. Besides, the original approximation accuracy and approximation roughness measures proposed by Pawlak are extended to deal with neighborhood decision systems and the concepts of entropy-based roughness and approximation roughness measures are presented. Experimental results demonstrate that the entropy-based measures are effective and valid for evaluating the uncertainty information of neighborhood systems. Experimental results also indicate that the entropy-based measures outperform the traditional Pawlak measures in the neighborhood systems.

The rest of this paper is organized as follows. Some preliminary concepts and notations in rough set theory are briefly recalled in Section 2. In Section 3, we introduce the concepts of neighborhood relation and information entropy theory to characterize several important measures including neighborhood entropy, neighborhood accuracy and neighborhood roughness. In addition, we employ the uncertainty measures to understand information entropy and information granularity from the viewpoint of neighborhood relations. Simulation experiments are conducted to evaluate the effectiveness of the proposed measures in Section 4. Section 5 concludes this paper with some remarks and discussions.

2. Preliminary concepts in rough sets

In this section, we review some basic concepts in rough set theory, including information system, equivalence relation, approximation regions and uncertainty measures.

2.1. Equivalence relation and approximation regions

The notion of information system has been studied by many authors as a simple knowledge representation method. Formally, an information system is a quadruple $I = (U, A, V, f)$, where: U is a nonempty finite set of objects, A is a nonempty

Please cite this article in press as: Y. Chen et al., An entropy-based uncertainty measurement approach in neighborhood systems, Inform. Sci. (2014), http://dx.doi.org/10.1016/j.ins.2014.03.117