Available online at www.sciencedirect.com
ScienceDirect
Information Sciences 178 (2008) 1968–1985
INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL
www.elsevier.com/locate/ins

Rough set theory for the interval-valued fuzzy information systems ☆

Zengtai Gong [a,*], Bingzhen Sun [b], Degang Chen [c]

[a] College of Mathematics and Information Science, Northwest Normal University, Lanzhou 730070, PR China
[b] School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, PR China
[c] Department of Mathematics and Physics, North China Electric Power University (Beijing), Beijing 102206, PR China

Received 9 May 2006; received in revised form 12 December 2007; accepted 13 December 2007

Abstract

The notion of a rough set was originally proposed by Pawlak [Z. Pawlak, Rough sets, International Journal of Computer and Information Sciences 11 (5) (1982) 341–356]. Later on, Dubois and Prade [D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, International Journal of General System 17 (2–3) (1990) 191–209] introduced rough fuzzy sets and fuzzy rough sets as a generalization of rough sets. This paper deals with an interval-valued fuzzy information system by means of integrating the classical Pawlak rough set theory with the interval-valued fuzzy set theory and discusses the basic rough set theory for the interval-valued fuzzy information systems. In this paper we firstly define the rough approximation of an interval-valued fuzzy set on the universe U in the classical Pawlak approximation space and the generalized approximation space respectively, i.e., the space on which the interval-valued rough fuzzy set model is built. Secondly several interesting properties of the approximation operators are examined, and the interrelationships of the interval-valued rough fuzzy set models in the classical Pawlak approximation space and the generalized approximation space are investigated. Thirdly we discuss the attribute reduction of the interval-valued fuzzy information systems. Finally, the methods of the knowledge discovery for the interval-valued fuzzy information systems are presented with an example.

Keywords: Fuzzy sets; Rough sets; Interval-valued fuzzy sets; Interval-valued fuzzy information systems

1. Introduction

Modeling imprecise and incomplete information is one of the main research topics in the area of knowledge representation. Many of the existing approaches are based on some extensions of classical set theory: fuzzy sets theory and rough sets theory, among others.

☆ The work is supported by the National Natural Science Fund of China (10771171) and the Scientific Fund of Gansu Province Education Office (0601-20).
* Corresponding author. Tel.: +86 931 7971845.
E-mail addresses: zt-gong@163.com, gongzt@nwnu.edu.cn (Z. Gong).

0020-0255/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2007.12.005

The concept of rough set was originally proposed by Pawlak [18,20] as a mathematical approach to handle imprecision, vagueness and uncertainty in data analysis. This theory has been amply demonstrated to have usefulness and versatility by successful application in a variety of problems [27,28,39]. In articles [21–23], Pawlak pointed out some research directions and applications based on rough sets. He also discussed two selected topics, namely, extensions of the rough set approach and the combination of rough sets and boolean reasoning with applications in pattern recognition, machine learning, data mining and conflict analysis. The theory of rough set deals with the approximation of an arbitrary subset of a universe by two definable or observable subsets called lower and upper approximations. By using the concept of lower and upper approximation in rough sets theory, knowledge hidden in information systems may be unravelled and expressed in the form of decision rules [5,6,17,20,30,31]. Another particular use of rough set theory is the attribute reduction of databases. Given databases with discretized attribute values, it is possible to find a subset of the original attribute that are the most informative. This leads to the concept of attribute reduction which can be viewed as the strongest characteristic in rough set theory to distinguish it from other theories. However, as mentioned in [26], in the existing databases the values of attributes could be both symbolic and real-valued. The classical rough set theory has difficulty in handling such values. There is a need for some method which firstly has the capability of utilizing set approximations and attribute reduction for crisp and real-value attribute databases, and secondly to make use of the degree of similarity of values. This could be accomplished by combining fuzzy sets and rough sets [1,12–14,19,24,34–36], i.e., rough fuzzy sets and fuzzy rough sets. In the meanwhile, Dubois and Prade pointed out that the rough fuzzy set is a special case of the fuzzy rough set [3].

Theories of fuzzy sets and rough sets are generalizations of classical set theories for modeling vagueness and fuzziness respectively. It is generally accepted that these two theories are related, but distinct and complementary, to each other [2,3,10,11,15,29,36–38]. Fuzzy rough sets encapsulate the related, but distinct, concepts of fuzziness and indiscernibility. These occur as a result of uncertainty in knowledge or data. A method employing fuzzy rough sets should be adopted to handle this uncertainty.

There are at least two approaches for the development of the fuzzy rough sets theory: the constructive and axiomatic approaches [34,36]. In the constructive approach, the relation to the universe is the primitive notion. The lower and upper approximation operators are constructed by means of this notion. Dubois and Prade were the first researchers to propose the concepts of rough fuzzy sets and fuzzy rough sets using the constructive approach. They constructed a pair of upper and lower approximation operators for fuzzy sets by respecting the fuzzy similarity relation using the t-norm min and its dual conorm max. It should be noticed that the operators of the minimum and maximum are a special t-norm and conorm. Radzikowska and Kerre [25] presented a more general approach to the fuzziness of rough sets. On the one hand, they define a broad family of fuzzy rough sets by respecting the fuzzy similarity relation that is determined by an implicator and a t-norm. On the other hand, their axiomatic approach takes the lower and upper approximation operators as primitive notions. In this approach, a set of axioms is used to characterize approximation operators. Another valuable generalization of rough set theory to fuzzy case is that the fuzzy neighborhood system can be viewed as a generalized approximation theory of fuzzy sets [7–9].

Fuzzy rough sets have been applied to solve practical problems. These include being used in neural networks, medical time series, case generation [16], mining stock pricing [33], and descriptive dimensionality reduction [26].

As a generalization of Zadeh fuzzy sets, the notion of interval-valued fuzzy set was suggested for the first time by Gorzalkczany [4] and Turksen [32]. It applied to the fields of approximate inference, signal transmission and control etc. In this paper we combine the classical Pawlak rough set theory with the interval-valued fuzzy set theory and define the interval-valued fuzzy information systems. We then discuss the rough set theory of the interval-valued fuzzy information systems, and build the interval-valued rough fuzzy sets models. The interrelationships of the interval-valued rough set models in the classical Pawlak approximation space and the generalized approximation space are also investigated. Finally, the methods of the knowledge discovery on the interval-valued fuzzy information systems are presented.

The structure of this paper is as follows. Section 2 briefly introduces necessary notions of rough sets, fuzzy sets and interval-valued fuzzy sets. In Section 3, we define the approximation operators of the interval-valued fuzzy sets in the classical Pawlak approximation and generalized approximation space and study their properties. In Section 4, the knowledge discovery of the interval-valued fuzzy information systems is studied, and a