Information Sciences 284 (2014) 60–80

Contents lists available at ScienceDirect

Information Sciences

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

Rough approximation of a fuzzy concept on a hybrid attribute information system and its uncertainty measure

Bingzhen Sun [a,*], Weimin Ma [b,*], Degang Chen [c]

[a] School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou, Gansu 730070, China
[b] School of Economics and Management, Tongji University, Shanghai 200092, China
[c] Department of Mathematics and Physics, North China Electric Power University, Beijing 102206, China

ARTICLE INFO

Article history:
Received 5 January 2014
Received in revised form 26 May 2014
Accepted 26 June 2014
Available online 3 July 2014

Keywords:
Rough set
Fuzzy rough set
Hybrid indiscernibility relation
Hybrid attribute information system
Knowledge granulation

ABSTRACT

One important and valuable topic in rough set theory is the concept of approximation over various generalized information systems. Although rough set models and approaches over hybrid attribute sets have been studied by many researchers recently, the studies focus on the rough approximation of a crisp concept over a hybrid attribute information system. This paper considers the rough approximation of a fuzzy concept over a hybrid attribute information system. We define a hybrid indiscernibility relation by compounding the fuzzy indiscernibility (similarity) relation over numerical attributes with the equivalence relation over symbolic attributes. Then we present the lower and upper approximations of a fuzzy set based on the hybrid indiscernibility relation, i.e., the fuzzy rough set over a hybrid attribute information system. Also, some interesting properties of the hybrid fuzzy rough approximation operators are presented in detail and the relationship between the fuzzy rough set and the existing rough set model over a hybrid attribute information system is established. Meanwhile, based on the definition of the cut set of a fuzzy set of the universe of discourse, the representations of the lower and upper approximations of a fuzzy set are given with respect to hybrid attribute approximation space. Furthermore, the concept of the cut set of the hybrid indiscernibility relation of the hybrid attribute approximation space is defined, then the characterizations of the hybrid fuzzy rough approximation operators are established as well. Two different types of characterization theorems of the hybrid fuzzy rough approximation operators are presented based on the cut set of the hybrid indiscernibility relation over the hybrid attribute information system. At the same time, we study the uncertainty measure of the fuzzy rough set over a hybrid attribute information system by using knowledge granulation. The main contribution of this paper is twofold. One is to extend the existing rough set approach over hybrid attribute information systems to a fuzzy environment. Another is to present a new way to define the rough approximation of a fuzzy concept as well the uncertainty measure over hybrid attribute information systems.

1. Introduction

The approximation of a concept on knowledge space (approximation space) with rough set theory is a principal problem in pattern recognition, fault diagnostics, decision-making under uncertainty, etc. It has induced various generalizations of

* Corresponding author. Tel.: +86 13993196228.
E-mail addresses: bzh-sun@163.com (B. Sun), mawm@tongji.edu.cn (W. Ma).

http://dx.doi.org/10.1016/j.ins.2014.06.036
0020-0255/

the Pawlak rough set [10,27] during the last decade, such as rough fuzzy set [4,5,43,44], fuzzy rough set [2,11,24,25,35,45,51,52,63], intuitionistic fuzzy rough set [36,64,65], vague rough set [7], etc. For a detailed review, readers can consult Ref. [28,29]. The basis of both the Pawlak rough set and its generalized forms is the equivalence relation or arbitrary binary relation on the universe of discourse determined by the attribute set. All existing generalizations have enriched the theoretical basis and also extended the areas of application of the Pawlak rough set. However, there may exist two or more types of attributes that describe the objects on the universe of discourse in practical decision-making. Then some new rough set approaches are needed to effectively deal with the case when different types of attributes appear. So, investigating the rough set approach to hybrid attribute sets is necessary. The aim of this paper is to deal with the fuzzy concept on hybrid attribute set information systems by combining rough set and fuzzy set.

The theory of Pawlak rough set deals with the approximation of an arbitrary subset of a universe by two definable or observable subsets, named lower and upper approximations, based on an indiscernibility relation determined by attributes. In the Pawlak rough set theory, the values of all attributes are assumed to be the nominal (symbolic) attribute, i.e., all attributes are in the same category for classification and then there is defined an indiscernibility relation by the symbol values of the attributes. In the traditional approach of Pawlak rough set theory, it should firstly transform the non-symbolic attributes into symbolic values by means of a discretization process when using the rough set theory to deal with classification decision-making. Though the attributes are in the same category in practical applications, as mentioned in Ref. [19], there could exist both symbolic and real values for the attributes. Then Pawlak rough set theory will encounter difficulty when modeling these kinds of values of attributes. Thus, there is a need for new approaches to handling the symbolic and real value of attributes simultaneously. So, the fuzzy rough set theory is established naturally by combining fuzzy set and rough set [4,5]. Consequently, the fuzzy rough theory has recently attracted much attention and interest of researchers and has also been successfully applied in many fields, such as machine learning, data mining and management decision-making. [3,8,12–17,19–21,33,38,41,42,45,46,48,66].

In many applications, especially management decision-making, however, there may not only be different types of attribute values but also different categories for the attributes which describe the objects of a universe of discourse. For example, let us consider a decision-making problem in management science: a credit card applicant. For all the potential applicants, the characteristics of the applicants could be described by four attributes "Income", "Consumption level", "Education" and "Sex". The attributes "Income" and "Consumption" may be given the values of "high", "mid", and "low" but the values of the attribute "Education" may be "Ph.D", "M.S", and "B.S" and the values of "Sex" are "Male" and "Female". It is easy to see that the values of attributes "Income" and "Consumption" are a fuzzy set and the values of "Education" and "Sex" are symbol-valued. So, the decision-making problem of processing a credit card application includes two different category attributes because "Income" and "Consumption" are numerical attributes (fuzzy set) but "Education" and "Sex" are nominal attributes (symbolic-valued). That is, the decision-making problem for a credit card application has hybrid attributes (or hybrid data; we assume that hybrid attributes and hybrid data are the same in the following sections). As aforementioned, the Pawlak rough set theory is incapable of dealing with the hybrid attribute decision-making problem. Therefore, the generalization of Pawlak rough set theory for handling hybrid data is valuable.

Broadly speaking, there are two kinds of methods for hybrid data processing. One is employing traditional numerical data processing methods, including Principal Component Analysis (PAC) [23], Neural Networks [6,18] and Support Vector Machine (SVM) [26], etc. The hybrid data should be transformed into integral numbers when these traditional approaches are employed. However, there exists an inherent limitation in this method because the transformation for the hybrid data lacks a reasonable semantic interpretation [9]. Another method is using Pawlak rough set directly by means of discretization of numerical data into symbolic data. Likewise, this method could result in some information loss when adapted to the discretization method [21]. So, there exist some inherent limitations and there is a need to develop a new approach to hybrid data processing.

Recently, the rough set approach to hybrid attribute information systems has attracted much attention, and some generalized rough set models have been proposed by researchers [9,10,15,45,47,52]. In order to deal with hybrid attribute information systems, Daniel et al. [52] and Hu et al. [15] propose a fuzzy equivalence relation over the hybrid attribute which includes numerical and fuzzy attributes and construction of fuzzy information granules (i.e., fuzzy equivalence classes). Then they define a new fuzzy rough set model based on a fuzzy equivalence relation for hybrid attribute information systems. Furthermore, they investigate the hybrid attribute reduction by using the fuzzy rough set method based on a fuzzy equivalence relation. In Ref. [47], Wei et al. construct a fuzzy equivalence relation by hybrid attribute which includes numerical and symbolic attributes. Based on the definition of fuzzy equivalence relation derived from numerical and symbolic attributes, they discuss the relationships among the existing fuzzy rough models based on hybrid attribute information systems. Actually, the existing fuzzy rough set models based on fuzzy equivalence relations derived from hybrid attributes approximate a crisp concept of the universe of discourse [10,15]. However, there may also be a fuzzy decision attribute on the hybrid attribute decision-making problem in management science. Just as in the decision-making about the credit card applicant, the evaluation of every applicant also may be a fuzzy set but not a crisp value. So, putting the fuzzy set and rough set together and applying these concepts to hybrid attribute information systems is a natural generalization of the existing fuzzy rough set theory [4,5,14,19,24,38–40,42–44,46,49]. Moreover, it would degenerate into a kind of fuzzy rough set approach to handle hybrid attribute decision making based on the hybrid indiscernibility relation when the decision object is a crisp concept.