Expert Systems With Applications 101 (2018) 205–212

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Tolerance-based intuitionistic fuzzy-rough set approach for attribute reduction

Anoop Kumar Tiwari[a], Shivam Shreevastava[b,*], Tanmoy Som[b], K.K. Shukla[c]

[a] Department of Computer Science, Institute of Science, BHU, Varanasi, 221005, India
[b] Department of Mathematical Sciences, IIT (BHU), Varanasi, 221005, India
[c] Department of Computer Science and Engineering, IIT (BHU), Varanasi, 221005, India

ARTICLE INFO

Article history:
Received 16 February 2017
Revised 12 January 2018
Accepted 4 February 2018
Available online 7 February 2018

Keywords:
Attribute selection
Similarity measure
Rough set
Fuzzy rough set
Intuitionistic fuzzy rough set
Degree of dependency

ABSTRACT

Due to technological advancement and the explosive growth of electrically stored information, automated methods are required to aid users in maintaining and processing this huge amount of information. Experts, as well as machine learning processes on large volumes of data, are the main sources of knowledge. Knowledge extraction is an important step in framing expert and intelligent systems. However, the knowledge extraction phase is very slow or even impossible due to noise and large size of data. To enhance the productivity of machine learning algorithms, feature selection or attribute reduction plays a key role in the selection of relevant and non-redundant features to improve the performance of classifiers and interpretability of data. Many areas like machine learning, image processing, data mining, natural language processing and Bioinformatics, etc., which have high relevancy to expert and intelligent systems, are applications of feature selection.

Rough set theory has been successfully applied for attribute reduction, but this theory is inadequate in the case of attribute reduction of real-valued data set as it may lose some information during the discretization process. Fuzzy and rough set theories have been combined and various attribute selection techniques were proposed, which can easily handle the real-valued data. An intuitionistic fuzzy set possesses a strong ability to represent information and better describing the uncertainty when compared to the classical fuzzy set theory as it considers positive, negative and hesitancy degree simultaneously for an object to belong to a set. This paper proposes a novel mechanism of attribute selection using tolerance-based intuitionistic fuzzy rough set theory. For this, we present tolerance-based intuitionistic fuzzy lower and upper approximations and formulate a degree of dependency of decision features over the set of conditional features. Moreover, the basic results on lower and upper approximations based on rough sets are extended for intuitionistic fuzzy rough sets and analogous results are established. In the end, the proposed algorithm is applied to an example data set and the comparison between tolerance-based fuzzy rough and intuitionistic fuzzy rough sets approaches for feature selection is presented. The proposed concept is found to be better performing in the form of selected attributes.

1. Introduction

Attribute selection (Dash & Liu, 1997; Duda, Hart, & Stork, 1973; Langley, 1994) has been an interesting area of research in machine learning, pattern recognition, data mining, statistics, and bioinformatics etc. Due to generation of huge amount of real-valued data, some dimensionality reduction techniques are necessary in order to get a smaller set of informative attributes to improve the performance of fast storage system and the accuracy of the prediction algorithms. Attribute selection is one of the dimensionality reduction techniques, which preserves the salient characteristics of the information system. It reduces highly correlated attributes, which may lead to lower accuracy. Attribute selection techniques focus on more relevant and less redundant attributes. It has been applied in many areas like web classification and text categorization (Jensen & Shen, 2008).

Rough set (proposed by Pawlak, 1982, 2012; Pawlak & Skowron, 2007) based approach is one of the most important techniques of attribute selection that acquires information from the dataset itself. Rough set theory does not need any external information for attribute selection. It handles the vagueness in the information system. However, this method can be applied to discrete data only.

* Corresponding author.
E-mail addresses: shivam.rs.apm@itbhu.ac.in (S. Shreevastava), tsom.apm@itbhu.ac.in (T. Som), kkshukla.cse@itbhu.ac.in (K.K. Shukla).

https://doi.org/10.1016/j.eswa.2018.02.009
0957-4174/

Therefore, discretization methods are applied in order to tackle the real-valued information system before attribute selection and this may lead to loss of some information.

In order to cope with this problem, fuzzy rough set (proposed by Dubois & Prade, 1990, 1992) based approach is presented to resolve both uncertainty and vagueness available in the dataset. Combining Zadeh's fuzzy set (Zadeh, 1965) and rough set gives a key route in reasoning with uncertainty for real-valued data. Fuzzy rough set concept has been implemented to surpass the deficiencies of the classical rough set approach in various aspects. Fuzzy set theory is a powerful tool to deal with uncertainty, but it has certain limitations also.

Firstly, fuzzy set theory is not capable of handling many decision-making problems, for example, in a voting problem, where, a panel of 10 experts is voting for 1 item, suppose 5 experts gave the conclusions "agree", three of them "opponent" and the rest of the two experts "abstain" . This scenario can be effectively handled by adding a non-membership degree for "opponent" and hesitancy degree for "abstain" . Secondly, in the many real-world applications, such as medical diagnosis and sensor information, etc., vaguely specified data values are very common. Fuzzy set theory is implemented to handle such vagueness by generalizing the concept of membership in a set. In a fuzzy set, the membership degree of the element in a universe always takes a single value between 0 and 1 but those single values may not completely define about the lack of knowledge as the uncertainty is not found only in judgment but also in the identification. Therefore, some extensions of fuzzy sets are required to handle the latter uncertainty.

In order to tackle more uncertain and complex information system, many concepts have been presented. A vague set and intuitionistic fuzzy set are two well-known extensions of fuzzy set. In a vague set, interval-based membership is used. Interval-based membership is more expressive in capturing vagueness available in the data. Bustince and Burillo (1996) investigated about vague set and intuitionistic fuzzy set (Atanassov, 1986, 1999) and showed that they are equivalent. Intuitionistic fuzzy set-based approaches are widely used to handle uncertainty, in which membership, non-membership and hesitancy functions are considered simultaneously. Thereby, it can handle uncertainty in a much better way when compared to fuzzy approaches. So, it has much stronger ability to deal with information system and draw a better glimpse of fragile ambiguities of the objective world.

Intuitionistic fuzzy set is the suitable choice in the situation when a representation of non-membership degree is found to be simpler than membership degree. Therefore, it is anticipated that the human decision-making process and activities requiring human expertise and knowledge which are inevitably imprecise or not totally reliable could be used to simulate by using intuitionistic fuzzy set concept. Intuitionistic fuzzy set-based approach has been successfully implemented in decision making concept and pattern recognition (Castiñeira, Torres-Blanc, & Cubillo, 2011; Deschrijver & Kerre, 2007; Dymova & Sevastjanov, 2010; Mukherjee & Basu, 2012; Pei & Zheng, 2012; Vlachos & Sergiadis, 2007; Xu, 2007; Zhang & Liu, 2011; Zhang, Shu, & Liao, 2014; Zhang, Xiong, & Ma, 2016).

There are several benefits of intuitionistic fuzzy sets over fuzzy sets are available in the literature. A vague pattern classification can be transformed into a precise and well-defined optimization problem by using intuitionistic fuzzy set approaches. Unlike fuzzy sets, intuitionistic fuzzy sets preserve a precise degree of the uncertainty.

Hereby, we propose a new type of intuitionistic fuzzy lower and upper approximations by applying a tolerance degree on the similarity between two objects and give a novel method along with a suitable algorithm to compute the reduct set of an intuitionistic fuzzy decision system. Although some of the researchers have presented tolerance-based approach (Jensen & Shen, 2007b; Skowron & Stepaniuk, 1996) for feature selection, but none of them have considered the tolerance-based intuitionistic fuzzy rough set assisted approach. Our proposed approach can handle uncertainty, vagueness, and imprecision in a very effective manner as we propose an intuitionistic fuzzy rough set model by combining two effective tools to handle uncertainty, i.e. intuitionistic fuzzy set and rough set and further it is generalized for feature selection. Apart from all the above-mentioned advantages, we have presented the degree of dependency approach for feature selection based on an intuitionistic fuzzy rough set model. We justify our proposed method by using propositions of lower and upper approximations analogous to rough set theory. Finally, we compare our method with the tolerance-based fuzzy rough set approach for an arbitrary information system and show that our approach gives a better result.

The rest of the work is organized as follows: Section 2 presents the literature survey. Section 3 contains a brief introduction of an intuitionistic fuzzy information system. In Section 4, a tolerance-based fuzzy rough set approach for attribute selection has been presented. In Section 5, we propose a tolerance-based intuitionistic fuzzy rough set concept for attribute reduction and establish some theoretical results supporting the approach. In Section 6, we present an algorithm for attribute selection based on our proposed method. In Section 7, we apply our algorithm on an arbitrary dataset. In Section 8, we conclude our entire work.

2. Literature survey

In spite of the fact that rough sets and intuitionistic fuzzy sets both capture specific aspects of the same idea-imprecision, the combination of intuitionistic fuzzy set theory and rough set theory are rarely discussed by the researchers. Jena, Ghosh, and Tripathy (2002), Nanda and Majumdar (1992) and Chakrabarty, Gedeon, and Koczy (1998) demonstrated that lower and upper approximations of intuitionistic fuzzy rough sets are again intuitionistic fuzzy sets. Samanta and Mondal (2001) presented a similar idea. Liu and Lin (2015) presented a novel intuitionistic fuzzy rough set model by using distance concept to resolve many real-life conflict problems. In the last few years, some of the intuitionistic fuzzy rough set models have been established by many researchers (Cornelis, De Cock, & Kerre, 2003; Wu, Gu, Li, & Xu, 2014; Wu, Shao, & Wang, 2017). Çoker (1998); Yin, Li, and Jun (2012); Zhang, Zhou, & Li (2012); Zhou and Wu (2008, 2011) established a relationship between the rough set and intuitionistic fuzzy set and revealed the fact that the fuzzy rough set is admittedly an intuitionistic L-fuzzy set. Nowadays, the intuitionistic fuzzy rough set theory is emerging as an effective and powerful tool to deal with uncertainty and applied for decision making to solve many real life problems (Gong & Zhang, 2014; Huang, Guo, Li, Feng, & Zhou, 2016; Huang, Guo, Zhuang, Li, & Zhou, 2014; Huang, Zhuang, Li, & Wei, 2013; Liang & Liu, 2015; Liu, Lin, & Zhao, 2015; Sun, Ma, & Liu, 2013; Xu, Liu, & Li, 2013; Zhang, 2016; Zhang et al., 2016; Liang, Xu, & Liu, 2017; Zhang, Peng, Wang, & Wang, 2017).

However, very few research works have been introduced in the area of feature selection based on intuitionistic fuzzy rough set. Lu, Lei, and Hua (2009) proposed the g enetic algorithm for attribute reduction of IFIS. Chen and Yang (2011) combined intuitionistic fuzzy rough set with information entropy and introduced a new attribute reduction algorithm. Esmail, Maryam, and Habibolla (2013) studied the structure of the intuitionistic fuzzy rough set model and its properties and presented their method of attribute reduction along with rule extraction. Huang, Li, and Wei (2012) designed an intuitionistic fuzzy rough set based attribute reduction model by using distance function. Zhang (2016) presented an attribute reduction method by using