Intuitionistic Fuzzy Neighborhood Rough Set Model for Feature Selection

Shivam Shreevastava, Indian Institute of Technology, Banaras Hindu University, Varanasi, India
Anoop Kumar Tiwari, Institute of Science, Banaras Hindu University, Varanasi, India
Tanmoy Som, Indian Institute of Technology, Banaras Hindu University, Varanasi, India

ABSTRACT

Feature selection is one of the widely used pre-processing techniques to deal with large data sets. In this context, rough set theory has been successfully implemented for feature selection of discrete data set but in case of continuous data set it requires discretization, which may cause information loss. Fuzzy rough set theory approaches have also been used successfully to resolve this issue as it can handle continuous data directly. Moreover, almost all feature selection techniques are used to handle homogeneous data set. In this article, the center of attraction is on heterogeneous feature subset reduction. A novel intuitionistic fuzzy neighborhood models have been proposed by combining intuitionistic fuzzy sets and neighborhood rough set models by taking an appropriate pair of lower and upper approximations and generalize it for feature selection, supported with theory and its validation. An appropriate algorithm along with application to a data set has been added.

KEYWORDS

Fuzzy Neighborhood Rough Set, Fuzzy Rough Set, Intuitionistic Fuzzy Rough Set, Intuitionistic Fuzzy Set, Rough Set

1. INTRODUCTION

Every day quintillions of bytes of data are created from various sources like sensor data pertaining to climate information, census information and agricultural information etc., that get posted to social media sites, the digital pictures and videos, sale and purchase transaction records and cell phone GPS signals etc. to name a few. Generation of huge amount of data from various sources may create lots of difficulties in learning by the system using various classifiers because of redundant features available in data sets (Jain, Duin & Mao, 2000). It is required to reduce the dimension of data sets in order to get significant and informative features for decreasing the cost, storage and process time for better classification and prediction. Various preprocessing techniques have been proposed by different authors time to time to tackle the big data but many of these techniques have become inadequate due to their own limitations. Feature selection is one of the well-known techniques in data preprocessing for data mining, machine learning, pattern recognition, bioinformatics and medical image processing (Duda, Hart & Stork, 1973; Zhu, Ong & Dash, 2007; Yu & Liu, 2004; Hall, 1999; Dash & Liu, 2003). Feature selection is the process of selecting those input features that are most predictive of a desired outcome. It is better than other dimensionality reduction techniques as it preserves the real meaning of the features after reduction.

Feature selection methods (Guyon & Elisseeff, 2003; Saeys, Inza & Larrañaga, 2007; Dash & Liu, 1997) can be categorized as filter, embedded, wrapper, unsupervised, semi-supervised and supervised

DOI: 10.4018/IJFSA.2018040104

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
(Bhatt & Gopal, 2005; Zhu, Ong & Dash, 2007; Liu & Yu, 2005), etc. Feature selection techniques can be divided into two categories, firstly, symbolic method and secondly, numerical method. Symbolic methods consider all features as categorical variables and numerical methods take all the features as real valued variables. If there exist any heterogeneous features, symbolic methods(such as Rough set (Pawlak, 1982; Pawlak, 2012) based feature selection) use a discretization approach (Ching, Wong & Chan, 1995) and convert them as symbolic features which may lead to some sort of assumption and cause information loss. Discretization may damage two types of structures, firstly neighborhood structure and secondly ordered structure in real space. Latter problem was handled by fuzzy rough set model, but it was not able to tackle the former problem. Very few researches have been proposed to deal with neighborhood structure of the data sets. 

Dubois and Prade presented fuzzy rough set model (Dubois and Prade, 1992) by combining rough set and fuzzy set (Zadeh, 1965). This model was generalized and successfully implemented by many researchers for feature selection (Jensen & Shen, 2004; Yeung, Chen, Tsang, Lee, & Xizhao, 2005; Tsang, Chen, Yeung, Wang & Lee, 2008; Dougherty, Kohavi & Sahami, 1995; Jensen & Shen, 2009) of real-valued data sets (Jensen & Shen, 2004; Jensen & Shen, 2007) but this model was not able to deal with neighborhood structure of data sets. Qinghua et al proposed neighborhood rough set model (Hu, Yu, Liu & Wu, 2008) and it was successfully implemented to deal with neighborhood structure but was not able to preserve real spaces. Wang et al. (Wang, Shao, He, Qian & Qi, 2016) proposed fuzzy neighborhood rough set model and applied it for feature subset selection. We have extended this model and proposed a novel intuitionistic fuzzy neighborhood rough set model and generalized it for feature selection, which can handle both neighborhood structure and real space structure in more efficient manner. Intuitionistic fuzzy set (Atanassov, 1986) is an extension of Zadeh fuzzy set. Intuitionistic fuzzy sets have its own ability to better narrating and drawing ambiguities of the objective world than the traditional fuzzy sets because it is considered the positive, negative and hesitancy degrees of an object simultaneously (Atanassov, 1989, Atanassov, 1999). Intuitionistic fuzzy sets have been efficiently applied to solve many of the decision problems (Lin, Yuan & Xia, 2007). In recent years, some of the intuitionistic fuzzy rough set models (Chakrabarty, Gedeon & Koczy, 1998; Cornelis, De Cock & Kerre, 2003; Zhang, Zhou & Li, 2012) have been proposed and implemented for feature selection. However, most of the models consider the intuitionistic fuzzy upper and lower approximations of a decision class using a nearest sample, which fails in case of noisy data set. Our proposed model can easily handle this problem as we are considering neighborhood structure of a sample which clearly indicates the uniqueness of our approach.

In this paper, we first define the intuitionistic fuzzy decision of a sample and characterize its intuitionistic fuzzy information granule by introducing parameterized intuitionistic fuzzy relation. The proposed model is the generalization of fuzzy neighborhood rough set concept, which overcomes the drawback of fuzzy rough set approach by introducing a nearest neighbor of a sample to different intuitionistic fuzzy decision classes. We calculate the reduct of an information system by using degree of dependency approach. Furthermore, we present an attribute reduction algorithm for better understanding of our model. Finally, an illustrative example has been given to demonstrate our approach.

2. PRELIMINARIES

In this section, we discuss some basic concepts regarding intuitionistic fuzzy information system.

2.1. Definition 2.1

Let \( \langle u, v \rangle \) be an ordered pair, where \( 0 \leq u, v \leq 1 \) and \( 0 \leq u + v \leq 1 \). Then \( \langle u, v \rangle \) is known as intuitionistic fuzzy value.
Integrated Network Topological Control and Key Management for Securing Wireless Sensor Networks
[www.igi-global.com/article/integrated-network-topological-control-and-key-management-for-securing-wireless-sensor-networks/104158?camid=4v1a](www.igi-global.com/article/integrated-network-topological-control-and-key-management-for-securing-wireless-sensor-networks/104158?camid=4v1a)

Comparative Analysis of Random Forests with Statistical and Machine Learning Methods in Predicting Fault-Prone Classes
[www.igi-global.com/chapter/comparative-analysis-random-forests-statistical/62703?camid=4v1a](www.igi-global.com/chapter/comparative-analysis-random-forests-statistical/62703?camid=4v1a)