Computers in Industry 97 (2018) 111–116

Contents lists available at ScienceDirect

Computers in Industry

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

A novel feature selection method using fuzzy rough sets

Sheeja T.K. [a,*], A. Sunny Kuriakose [b]

[a] Department of Mathematics, T. M. Jacob Memorial Govt. College, Manimalakunnu, Kerala, India
[b] Federal Institute of Science and Technology, Angamaly, Kerala, India

ARTICLE INFO

Article history:
Received 22 May 2017
Received in revised form 6 November 2017
Accepted 31 January 2018
Available online xxx

Keywords:
Information systems
Approximations
Rough set
Divergence measure
Fuzzy rough set
Feature selection

ABSTRACT

The fuzzy set theory and the rough set theory are two distinct but complementary theories that deal with uncertainty in data. The salient features of both the theories are encompassed in the domain of the fuzzy rough set theory so as to cope with the problems of vagueness and indiscernibility in real world data. This hybrid theory has been found to be a potential tool for data mining, particularly useful for feature selection. Most of the existing approaches to fuzzy rough sets are based on fuzzy relations. In this paper, a new definition for fuzzy rough sets in an information system based on the divergence measure of fuzzy sets is introduced. The properties of the fuzzy rough approximations are explored. Moreover, an algorithm for feature selection using the proposed approximations is presented and experimented using real data sets.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

One of the greatest challenges faced by the present information age is the extraction of useful knowledge from a vast amount of raw data. The data mining tools and techniques are of great help in finding and describing the structural patterns in data. The conventional data analysis techniques such as regression analysis, time series analysis, cluster analysis, stochastic models, etc. deal with extracting quantitative data characteristics [1]. They cannot handle qualitative or imperfect data or produce a qualitative description of the dependencies in the data. Fuzzy set theory [2] and rough set theory [3] are two formal mathematical tools which can address the problem of vagueness, imperfection or incompleteness in data. They handle two different aspects of uncertainty namely vagueness and indiscernibility. The successful applications of fuzzy sets and rough sets in data mining have lead to a hybrid theory namely fuzzy rough set theory which can manage both types of uncertainty.

Fuzzy rough set theory is found to be a very effective tool for feature selection [4]. The task of feature selection is to remove the irrelevant and redundant features from a data set and choose only those features that facilitate extraction of useful knowledge. Using fuzzy rough feature selection it is possible to reduce discrete or real valued noisy data without providing any additional information. There are many different approaches to fuzzy rough sets available in the literature. Most of the definitions are based on fuzzy relations [5–9]. In this paper the fuzzy rough lower and upper approximations of a fuzzy set in a fuzzy information system are defined using the divergence measure of fuzzy sets. Divergence measures are fuzzy measures that express the extent to which two fuzzy sets differ from each other. The properties of the proposed approximations are explored. Further, an algorithm for feature selection using the fuzzy positive region is presented and experimented with real data sets.

The rest of the paper is organized as follows: Section 2 gives a brief review of the existing fuzzy rough set models and applications to feature selection. The concept of divergence based fuzzy rough sets in a fuzzy information system are introduced in Section 3 and their properties are studied. Section 4 provides a feature selection technique using the fuzzy positive region in the proposed approach and the process is illustrated with an example. The results obtained on experimentation of the proposed method with real data sets are discussed in Section 5. The conclusion and future work are given in Section 6.

2. Related work

In this section, some preliminary definitions and a brief review of the different approaches to fuzzy rough set theory existing in the

* Corresponding author.
E-mail addresses: sheejatk@fisat.ac.in (S. T.K.), asunnyk@fisat.ac.in (A.S. Kuriakose).

https://doi.org/10.1016/j.compind.2018.01.014
0166-3615/© 2018 Elsevier B.V. All rights reserved.

literature are presented. The basic notions of fuzzy set theory and rough set theory as described in [10,12] respectively, are followed throughout this paper.

2.1. Divergence measure

Let $\mathcal{F}(U)$ be the family of all fuzzy sets on U. Then a function $\delta : \mathcal{F}(U) \times \mathcal{F}(U) \to R$ is a *divergence measure* [11] if and only if $\forall A, B \in \mathcal{F}(U)$

1. $\delta(A, B)=\delta(B, A)$
2. $\delta(A, A)=0$
3. $\max\{\delta(A\cup C, B\cup C), \delta(A\cap C, B\cap C)\}\leq\delta(A, B)$

2.2. Fuzzy rough sets

Fuzzy rough sets encapsulate the related but distinct concepts of vagueness and indiscernibility. A fuzzy rough set consists of a pair of fuzzy membership functions which correspond to the fuzzy lower and upper approximations of a fuzzy set in a fuzzy approximation space. A fuzzy approximation space is a pair (U, R), where U is a non-empty set of objects and R is a fuzzy equivalence relation. The first attempt to define fuzzy rough sets in a fuzzy approximation space was done by Nakamura [5]. He defined the lower and upper approximations of a fuzzy set A on U as the fuzzy sets on U given by

$$\mu_{\underline{R}(A)}(x) = \inf_{R(x,y)\geq\alpha} \mu_A(y) \tag{1}$$

$$\mu_{\bar{R}(A)}(x) = \sup_{R(x,y)\geq\alpha} \mu_A(y) \tag{2}$$

respectively. Dubois and Prade [6] defined fuzzy rough approximations as

$$\mu_{\underline{R}(A)}(x) = \inf_{y\in U} \max(1 - R(x,y), \mu_A(y)) \tag{3}$$

$$\mu_{\bar{R}(A)}(x) = \sup_{y\in U} \min(R(x,y), \mu_A(y)) \tag{4}$$

Since then, intensive studies have been conducted on fuzzy rough sets both in theoretical and application point of view and several extensions and generalizations are proposed [7–9]. A detailed study of the different approaches to fuzzy rough set was done by D'eer et al [14].

2.3. Feature selection using fuzzy rough sets

Since its inception, several researchers have attempted to apply fuzzy rough set theory to feature selection. Many of them used the dependency degree, which is a measure of how well an attribute set can discern between elements as a criteria for feature selection. Whichever definition of fuzzy rough set is used, the dependency degree is defined using the fuzzy cardinality of the fuzzy positive region. A number of papers were authored by Jensen and Shen [15–19] in which the development of a fuzzy rough quick reduct algorithm was outlined. The attribute set having the maximum dependency degree is selected. Another approach to fuzzy-rough feature selection is to use fuzzy entropy as a criteria for feature selection [20]. Algorithms based on discernibility matrix to compute the attribute reducts are also proposed by many authors [19,21]. Fuzzy boundary region based feature selection methods are also available in the literature [19,22].

3. Divergence based fuzzy rough sets

Let (U, C, D) be a fuzzy information system, where U is a non-empty set of objects, C is the set of conditional attributes each of which is a fuzzy set on U and D is the set of decision attributes. Each $x\in U$ can be associated with a fuzzy set on $P\subseteq C$, with membership function

$$\mu_x^P(a) = \begin{cases} a(a), & \text{if } a\in P, \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

Unless there is no ambiguity, the above fuzzy set may be denoted by μ_x.

Let $\delta(A, B)$ be a normalized divergence measure of fuzzy sets. Then, $\delta(\mu_x, \mu_y)$ measure the extent to which the fuzzy set μ_x on P differ from the fuzzy set μ_y on P. In other words, $\delta(\mu_x, \mu_y)$ expresses the dissimilarity between the objects x and y, with respect to the given set of fuzzy conditional attributes, whereas $R(x, y)$ expresses the indiscernibility between x and y. Thus a natural procedure to define divergence based fuzzy rough approximations is to replace $R(x, y)$ by $1-\delta(\mu_x, \mu_y)$.

Definition 3.1. The fuzzy rough lower and upper approximations of a fuzzy set A on U with respect to the divergence measure δ are defined $\forall x\in U$ as

$$\mu_{\underline{\delta}(A)}(x) = \inf_{y\in U} \max[\delta(\mu_x, \mu_y), \mu_A(y)] \tag{6}$$

$$\mu_{\bar{\delta}(A)}(x) = \sup_{y\in U} \min[1 - \delta(\mu_x, \mu_y), \mu_A(y)] \tag{7}$$

respectively.

The following proposition asserts that the above defined approximations are fuzzy sets on U.

Proposition 3.1. *The divergence based fuzzy rough lower and upper approximations of a fuzzy set on an information system are fuzzy sets on U.*

Proof. Both $\mu_A(y)$ and $\delta(\mu_x, \mu_y)\in[0, 1]$, $\forall x, y\in U$. Hence, $\max[\delta(\mu_x, \mu_y), \mu_A(y)]\in[0, 1]$. Using Eq. (6), $\mu_{\underline{\delta}(A)}(x) \in [0, 1]$, $\forall x \in U$. Similarly $\mu_{\bar{\delta}(A)}(x) \in [0, 1]$, $\forall x \in U$.

The properties of the proposed approximations are expressed in the next two theorems.

Theorem 3.1. *The general properties of the fuzzy rough lower and upper approximations with respect to δ are as follows:*

1. $\underline{\delta}(\phi) = \phi = \bar{\delta}(\phi)$
2. $\underline{\delta}(U) = U = \bar{\delta}(U)$
3. $\underline{\delta}(A) \subseteq A \subseteq \bar{\delta}(A)$, $\forall A \in \mathcal{F}(U)$
4. $A \subseteq B \Rightarrow \underline{\delta}(A) \subseteq \underline{\delta}(B)$ *and* $\bar{\delta}(A) \subseteq \bar{\delta}(B)$
5. $\underline{\delta}(\hat{\alpha}) = \hat{\alpha} = \bar{\delta}(\hat{\alpha})$, $\forall\alpha\in[0, 1]$
6. $(\underline{\delta}(A^C))^C = \bar{\delta}(A)$, $\forall A \in \mathcal{F}(U)$
7. $(\bar{\delta}(A^C))^C = \underline{\delta}(A)$, $\forall A \in \mathcal{F}(U)$

Proof.

1. $\mu_\phi(x) = 0$, $\delta(\mu_x, \mu_x) = 0$, $\forall x \in U \Rightarrow \max[\delta(\mu_x, \mu_x), \mu_\phi(x)] = 0$
   $\Rightarrow \mu_{\underline{\delta}(\phi)}(x) = \inf_{y\in U}\max[\delta(\mu_x, \mu_y), \mu_\phi(y)] = 0.$
   Also, $\min[1 - \delta(\mu_x, \mu_x), \mu_\phi(x)] = 0$, $\forall x \in U$.
   So, $\mu_{\bar{\delta}(\phi)}(x) = \sup_{y\in U}\min[\delta(\mu_x, \mu_y), \mu_\phi(y)] = 0.$
   Thus, $\underline{\delta}(\phi) = \phi = \bar{\delta}(\phi)$
2. $\mu_U(x) = 1$, $\forall x \in U \Rightarrow \max[\delta(\mu_x, \mu_y), \mu_U(y)] = 1$, $\forall y \in U$.
   Therefore, $\mu_{\underline{\delta}(U)}(x) = \inf_{y\in U}\max[\delta(\mu_x, \mu_y), \mu_U(y)] = 1.$
   Also, $\min[1 - \delta(\mu_x, \mu_y), \mu_U(y)] = 1 - \delta(\mu_x, \mu_y)$.

Download English Version:

https://daneshyari.com/en/article/6923885

Download Persian Version:

https://daneshyari.com/article/6923885

Daneshyari.com