



Constructing importance measure of attributes in covering decision table



Fachao Li^a, Zan Zhang^{b,*}, Chenxia Jin^a

^a School of Economy and Management, Hebei University of Science and Technology, Shijiazhuang 050018, China

^b College of Management and Economics, Tianjin University, Tianjin 300072, China

ARTICLE INFO

Article history:

Received 9 January 2014

Received in revised form 14 December 2014

Accepted 16 December 2014

Available online 24 December 2014

Keywords:

Importance measure

Covering

Decision table

Rough set theory

Evaluation model

ABSTRACT

In rough set theory, attributes importance measure is a crucial factor in applications of attribute reduction and feature selection. Many importance measure methodologies for discrete-valued information system or decision table have been developed. However, there are only limited studies on importance measurement for numerical-valued information system. In this paper, knowledge change-based importance measure, with the structural characteristics of fuzzy measure, is introduced to evaluate the importance of attributes in covering decision table. We first present the concept of similarity block in attribute space, based on which coverings are induced to construct the lower and upper approximation operators in covering-based rough sets. In particular, the traditional importance measure is extended to deal with covering decision table. Further, an evaluation model based on the knowledge change-based importance measure is constructed. Experiments are conducted on the public data sets from UCI, and a case study on the students' overall evaluation is given finally. Theoretical analysis and experimental results show that the proposed importance measure is effective for evaluating the importance of attributes in covering decision table.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Rough set theory [1], proposed by Pawlak in the early 1980s, is a mathematical tool to deal with uncertainty and incomplete information. It is an important tool in knowledge discovery in information system. Over the last decades, there has been much work on information systems with rough sets. Data mining based on rough sets is employed widely to obtain knowledge from databases. It also has found successful applications in attribute reduction, machine learning, pattern recognition, image processing and decision-making analysis from large data sets [2–7].

In attribute reduction and feature selection methods, attribute importance measure is a crucial factor. To evaluate the importance of attributes in information system, several importance measures have been proposed in rough set theory. Many authors have studied the importance measure of attributes based on positive region, Shannon's conditional entropy, complement conditional entropy, combination conditional entropy and rough entropy. The idea of attribute reduction using positive region was originated by Grzymala-Buss [8,9]. In a classical rough set model, the conditional

entropy of Shannon's information entropy was introduced in [10,11] to find the relative reduct of a decision table. Based on the complement entropy, its conditional entropy was used to measure the importance of attributes in [12]. Qian and Liang [13] presented combination entropy to measure the uncertainty of information systems and used its conditional entropy to obtain an attribute subset. Sun et al. [14] proposed a new rough entropy associated to partitions, and presented the importance measure of attributes on the basis of this entropy. Yao et al. [15–17] studied several kinds of information entropy measures for attribute importance in rough set theory. However, the studies mentioned above are mainly focused on discrete-valued information system or decision systems.

Traditional rough set theory is suitable for discrete data rather than continues data since only partition or equivalence relation are considered. As pointed out by some researchers, partition or equivalence relation in traditional rough set theory is too restrictive for practical applications. Many practical data sets, in which some attributes are numerical, cannot be handled by partitions. To address this issue, some scholars extended the equivalence relations to similarity relation [18], tolerance relation [19,20], and even general binary relations [21,22]. An example extension is neighborhood-based rough set model. Lin [23] pointed out that neighborhood spaces are more general topological spaces than equivalence spaces, and introduced neighborhood relation into

* Corresponding author. Tel.: +86 31181668326; fax: +86 31181668325.

E-mail addresses: lifachao@tsinghua.org.cn (F. Li), zhangzan166@163.com (Z. Zhang), jinchexia2005@126.com (C. Jin).

rough set theory. The properties of neighborhood approximation spaces were discussed in [24]. Hu et al. [25,26] constructed a unified theoretical framework for neighborhood-based rough set model and a feature selection algorithm toward hybrid data. Liu et al. [27] proposed an efficient quick attribute reduction algorithm based on neighborhood rough set model. Another example extension of the traditional rough set model is covering-based rough set model [28–30], which is based on the relaxation of the partition to covering. The covering of a universe is used to define the upper and lower approximations of any subset of the universe. Most of the studies of covering-based rough set model mainly focus on the construction of the upper and lower approximations. In [31–34], covering based approximation operators were discussed according to the element, granule, and subsystem based definitions. However, less effort has been put on the discussions of the construction of covering in numerical-valued information systems. In addition, there are only limited studies on the importance measure for covering decision table. In this paper, we address the issue of attributes importance measurement in covering decision table and propose effective knowledge change-based importance measure for covering decision table. The similarity block induced by attribute is given, under which coverings of the universe with numerical attributes are constructed. Further, the covering approximation operators are presented. In addition, the traditional importance measure is extended to deal with covering decision table, and the knowledge change-based importance measure with the structural characteristics of fuzzy measure is presented. Experimental results demonstrate that the knowledge change-based importance measure is effective and suitable for evaluating the importance of attributes in covering decision table. Experimental results also indicate that the knowledge change-based importance measure outperforms the traditional importance measure in covering decision table. The main contributions of the work are twofold. First, we give the concrete construction strategy of coverings to deal with data with numerical attributes or mixed attributes; second, a knowledge change-based importance measure, which meets the conditions of fuzzy measure, is presented in covering decision table.

The rest of the paper is organized as follows. Section 2 reviews some relative concepts in rough sets. In Section 3, we introduce the concept of similarity block in attribute space to construct coverings of the universe. Further, the approximation operators in covering approximation space are given. Besides, we present the extended importance measure and knowledge change-based importance measure in covering decision table. In Section 4, simulation experiments are carried out to evaluate the validity of the proposed method. The evaluation model with respect to knowledge change-based importance measure is constructed in Sections 5 and 6 shows a case study on students' overall quality evaluation. Section 7 concludes this paper with some remarks and discussions.

2. Preliminaries

In this section, we briefly review some basic concepts related to rough sets and covering.

An information system in rough sets is a quadruple (U, A, ν, f) , where U is a non-empty finite set of objects, A is a non-empty finite set of attributes, ν is the value set of attribute and f is the mapping function: $f: U \times A \rightarrow \nu$. In particular, $(U, A \cup \{d\}, \nu, f)$ is called a decision table, where A denotes the condition attributes and d denotes the decision attribute.

In Pawlak's rough set theory [1], the objects with the same attribute values in terms of attributes B form an equivalence class. Let R be an equivalence relation on U , the quotient set of U by the relation R is denoted by U/R , and $U/R = \{X_1, X_2, \dots, X_m\}$, here X_i

($i = 1, 2, \dots, m$) is an equivalence class of R . The R -lower and R -upper approximation sets of X are defined, respectively, as follows:

$$\underline{R}(X) = \{x | [x]_R \subset X\}, \quad (1)$$

$$\overline{R}(X) = \{x | [x]_R \cap X \neq \emptyset\}, \quad (2)$$

where $[x]_R$ denotes the set of all objects indiscernible with x , i.e., equivalence class determined by x .

Let P and Q be equivalence relations on U , then the positive, negative and boundary regions can be defined as

$$POS_P(Q) = \bigcup_{X \in U/Q} P(X), \quad (3)$$

$$NEG_P(Q) = U - \bigcup_{X \in U/Q} \overline{P}(X), \quad (4)$$

$$BND_P(Q) = \bigcup_{X \in U/Q} \overline{P}(X) - \bigcup_{X \in U/Q} P(X). \quad (5)$$

Given a decision table $(U, A \cup \{d\}, \nu, f)$, $B \subset A$, $\forall a \in B$, the importance of a in B is defined as

$$sig(a, B, d) = \gamma_B(d) - \gamma_{B-\{a\}}(d), \quad (6)$$

where $\gamma_B(d) = \frac{|POS_B(d)|}{|U|}$ is the dependency degree of d to B , $|X|$ is the cardinality of X . $\gamma_B(d)$ reflects the ability of B to approximate d . Since $POS_B(d) \subset U$, we have $0 \leq \gamma_B(d) \leq 1$.

Definition 1 [34]. Let U be a universe and \mathbb{C} be a family of subsets of U . \mathbb{C} is called a covering of U if none subsets in \mathbb{C} is empty, and $\bigcup \{K | K \in \mathbb{C}\} = U$. The order pair (U, \mathbb{C}) is called a covering approximation space if \mathbb{C} is a covering of U .

It is clear that a partition is a covering of U , so the concept of covering is an extension of the concept of a partition.

3. Importance measurement approaches in covering decision table

Traditional importance measure of attributes is mainly suitable for discrete-value information system. In the following, the similarity block is introduced to deal with the numerical attribute in the system and two kinds of importance measures in the covering decision table are given.

3.1. Approximation operators in covering approximation space

In what follows, we refer to the decision table in which the numerical attributes are dealt with based on coverings as covering decision table, and we denote it as $(U, A \cup \{d\}, \nu, f, \mathbb{C})$. Here, the numerical condition attribute values are normalized to interval $[0, 1]$, and the decision attribute is nominal. Based on the covering decision table, we give the following definitions:

Definition 2. Given arbitrary object $x \in U$, and a numerical attribute $a \in A$ to describe it, $f(x, a)$ denotes the value of sample x on attribute a , the similarity block in attribute space a is defined as $S_i^a = \{x | x \in U, f(x, a) \in [\frac{i}{n} - \delta, \frac{i}{n} + \delta]\}$, $0 \leq i \leq n$, where $\delta \in [\frac{1}{2n}, \frac{1}{n})$. Object x_k and object x_j are treated as similar if $x_k, x_j \in S_i^a$, i.e., $x_k SIM^a x_j$.

The similarity block describes the common character of the objects in the same block, and it also reflects the difference of the objects in different blocks. Note that the above definition is given in one-dimensional space. That is, we only define the similarity block in the singleton attribute space. It is clear that an object from a given similarity block may be similar to an object of another

Download English Version:

<https://daneshyari.com/en/article/403537>

Download Persian Version:

<https://daneshyari.com/article/403537>

[Daneshyari.com](https://daneshyari.com)