Information Sciences 240 (2013) 72–82

Contents lists available at SciVerse ScienceDirect

Information Sciences

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

Entropy measures and granularity measures for set-valued information systems

Jianhua Dai *, Haowei Tian

College of Computer Science, Zhejiang University, Hangzhou 310027, China

ARTICLE INFO

Article history:
Received 12 October 2011
Received in revised form 30 December 2012
Accepted 25 March 2013
Available online 1 April 2013

Keywords:
Set-valued information system
Entropy
Granularity
Measure
Uncertainty

ABSTRACT

Set-valued information systems are generalized models of single-valued information systems. In this paper, we propose two new relations for set-valued information systems. Based on these two relations, the concepts of knowledge information entropy, knowledge rough entropy, knowledge granulation and knowledge granularity measure are defined in set-valued information systems, and some properties are investigated. Moreover, relationship between knowledge information entropy and knowledge granulation, knowledge rough entropy and knowledge granularity measure are studied. It is also shown that knowledge information entropy and knowledge granularity measure can be used to evaluate the certainty degree of knowledge in set-valued information systems, and knowledge rough entropy and knowledge granulation can be used to evaluate the uncertainty degree of knowledge in set-valued information systems. These results may supply a further understanding the essence of uncertainty and granularity in set-valued information systems.

1. Introduction

Rough set theory is a mathematical tool for dealing with uncertain or imprecise information [35]. It has attracted the attention of many researchers who have studied its theories and its applications during the last decades [1,4–7,10,11,13,17–21,23,29,32,34,44,45,47,48,50,54,56,59,61].

Classical rough set philosophy is based on an assumption that every object in the universe of discourse is associated with some information (knowledge), expressed by means of some attributes used for object description [35]. Objects characterized by the same information are indiscernible. The indiscernibility relation generated in this way forms the mathematical basis for the theory of rough sets. In many practical issues, it may happen that some of the attribute values for an object are set-valued, which are always used to characterize uncertain information and missing information in information systems. Set-valued information systems can be viewed as generalized models of single-valued information systems [15]. Moreover, set-valued information systems can be used to handle incomplete information systems, in which all missing values can be represented by the set of all possible values of each attribute [38,51,55].

Proposed by Shannon [42] to evaluate uncertainty of a system, entropy has been applied in diverse fields as a very useful mechanism for characterizing information contents in various modes. The extension of entropy and its variants were adapted for rough sets in [2,8,9,12,14,24,36,39,46]. For example, Duentsch and Gediga defined the information entropy and three kinds of conditional entropies in rough sets for predicting a decision attribute [14]. Beaubouef et al. [2] proposed a method measuring uncertainty of rough sets and rough relation databases based on rough entropy. Wierman [46] presented the measures of uncertainty and granularity in rough set theory, along with an axiomatic derivation. Liang et al.

* Corresponding author.
E-mail address: jhdai@zju.edu.cn (J. Dai).

0020-0255/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2013.03.045

[24] proposed a new method for evaluating both uncertainty and fuzziness. Qian and Liang [39] proposed a combination entropy for evaluating uncertainty of a knowledge from an information system. All these studies were dedicated to evaluating uncertainty of a set in terms of the partition ability of knowledge. As a powerful mechanism, granulation was introduced by Zadeh [58]. As a recently renewed research topic, granular computing concerns problem solving and information processing at multiple levels of granularity [37]. Rough set model can be viewed as an example of partition based granular computing model [57]. From the viewpoint of granulation, Yao defined a granularity measure in [53]. Several measures on knowledge in an information system were proposed and studied by Liang et al. [25,26]. These measures include granularity measure, information entropy, rough entropy, and knowledge granulation. Qian et al. studied knowledge granulation in a knowledge base [41], and fuzzy information granularity in a binary granular structure [40]. Xu et al. introduced concepts of knowledge granulation, knowledge entropy and knowledge uncertainty measure in ordered information systems [49]. Zhu et al. developed a pair of information-theoretic entropy and co-entropy functions associated to partitions and approximations [60]. For more details on this topic, one can refer to a systematic survey in [57]. So far, however, uncertainty measurement in set-valued information systems has not been reported. In this paper, we aim to address uncertainty measurement issue in set-valued information systems. This paper introduces knowledge entropy, knowledge granulation, and roughness measure into set-valued information systems, and investigates some properties of them. It is shown that these proposed measures provide approaches to evaluate the discernibility ability of different knowledge in set-valued information systems.

The remainder of the paper is organized as follows. In Section 2, we briefly review some information-theoretic measures and granularity measures for rough sets in the literature. We propose two extended similarity relations for set-valued information systems in Section 3. Entropy measures including information entropy and rough entropy are introduced into set-valued information systems in Section 4. In Section 5, granularity measures including knowledge granulation and granularity measure are introduced into set-valued information systems. In Section 6, relationships between entropy measures and granularity measures are investigated. Section 7 concludes the paper.

2. Preliminaries

Shannon entropy has been widely used to measure the structuredness of attributes in databases and the nonspecificity of a finite set [22,52]. There are several information-theoretic measures of uncertainty and granularity for rough sets [2,3,14,25,26,31,46,52,53,60], which are based upon the notion of entropy introduced by Shannon [42].

Definition 2.1. [14,31,46,53,60] Let $\langle U, \pi \rangle$ be an approximation space, where the partition π consists of blocks A_i, $1 \leqslant i \leqslant k$. The Shannon entropy $H(\pi)$ of the partition π is defined by

$$H(\pi) = \sum_{i=1}^{k} \frac{|A_i|}{|U|} log_2 \frac{|A_i|}{|U|} \tag{1}$$

The entropy reaches the maximum value $log_2|U|$ for the finest partition consisting of singleton subsets of U, and it reaches the minimum value 0 for the coarsest partition $\{U\}$. In general, for two partitions with $\pi_1 \preceq \pi_2$, we have $H(\pi_1) \geqslant H(\pi_2)$. That is, the value of the entropy correctly reflects the order of partitions with respect to their granularity [46].

Duntsch and Gediga [14] and Miao et al. [31] used Shannon entropy $H(\pi)$ of a partition as a measure of its roughness or granularity. Wierman explicitly called Shannon entropy of a partition a granularity measure in [46]. In this paper, we call $H(\pi)$ granularity measure of partition π, denoted as $G(\pi)$. Hence, we have

$$G(\pi) = H(\pi) = \sum_{i=1}^{k} \frac{|A_i|}{|U|} log_2 \frac{|A_i|}{|U|} \tag{2}$$

The Shannon entropy in Eq. (1) can be re-expressed as [46,53]:

$$H(\pi) = log_2|U| - \sum_{i=1}^{k} \frac{|A_i|}{|U|} log_2|A_i| \tag{3}$$

Notice that the Hartley measure [16] of uncertainty for a finite set A_i is

$$H(A_i) = log_2|A_i| \tag{4}$$

It measures the amount of uncertainty associated with a finite set of possible alternatives, the nonspecificity inherent in the set.

The first term $log_2|U|$ in Eq. (3) is exactly the Hartley measure of U, which is a constant independent of any partition. The second term of the equation is basically an expectation of granularity with respect to all blocks in the partition. This quantity was first explicitly used by Yao to measure the granularity of a partition in [53]. It should also be noticed that Beaubouef et al. [2] implicitly treated it as a measure of granularity of a partition when defining the uncertainty of a rough set approximation. Liang and Shi [25] called it the rough entropy of π. Bianucci et al. [3] called it co-entropy.