Information Sciences 320 (2015) 190–205

Contents lists available at ScienceDirect

Information Sciences

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

∗-reductions in a knowledge base

Bin Qin

School of Information and Statistics, Guangxi University of Finance and Economics, Nanning, Guangxi 530003, PR China

ARTICLE INFO

Article history:
Received 6 August 2014
Received in revised form 10 May 2015
Accepted 16 May 2015
Available online 22 May 2015

Keywords:
Knowledge base
Reduction
∗-reduction
Relation mapping
Homomorphism

ABSTRACT

One of the advantages of rough set theory is the fact that an unknown target concept can be approximately characterized based on existing knowledge structures in a knowledge base. Knowledge reductions are one of the key problems in the study of rough set theory. In this paper, ∗-reductions and their features in a knowledge base are explored. Moreover, communication between knowledge bases is proposed based on the use of homomorphisms between relation information systems, and several invariant characterizations under homomorphisms are obtained. These results will be significant for establishing a framework for granular computing in knowledge bases.

1. Introduction

Rough set theory, which was first proposed by Pawlak [18] and is an important tool for coping with the fuzziness and uncertainty of knowledge, has become an active branch of information science. With its development over the past decades, rough set theory has been successfully applied to machine learning, intelligent systems, inductive reasoning, pattern recognition, mereology, image processing, signal analysis, knowledge discovery, decision analysis, expert systems and many other fields [19–22].

From the perspective of rough set theory, knowledge is the ability to classify elements [4,10]. In other words, knowledge is a family of classifications of the universe. Rough set theory focuses on equivalence relations regarding the universe, which determine partitions of the universe. We consider not only a single classification (or partition) of the universe but also a family of classifications (or partitions) of the universe. This leads to the generation of knowledge bases, which is an important concept in rough set theory.

For a given knowledge base, one of the tasks of data mining and knowledge discovery is to generate new knowledge based on known knowledge.

It is well known that not all elements in a knowledge base are of the same importance. Some are even redundant. Thus, we often remove the redundant ones under the requirement of retaining the ability of classification. Such a process is known as a knowledge reduction in a knowledge base. In judging whether elements in a knowledge base are redundant, we may operate in accordance with the following procedure: if elements are unrelated, then they are naturally redundant; if the features of two elements are identical, then we may retain either one of them and the other is redundant. Several scholars have studied knowledge reductions in information systems, knowledge reductions in formal contexts and related problems [1,5,7–9,11–17,24–26,29,31–33].

E-mail address: binqin8846@126.com

http://dx.doi.org/10.1016/j.ins.2015.05.030
0020-0255/

Note that Pawlak's reductions in a knowledge base $K = (U, \mathcal{R})$ involve only the special intersection $ind(\mathcal{R}) = \bigcap \mathcal{R}$; thus, they may lose much of the original information and introduce new uncertainty. In view of these factors, Xu and Zhao [30] considered the saturation of a subfamily of the universe and defined a new type of knowledge reduction for abstract knowledge bases, called a saturation reduction. Such saturation reductions have particular significance in the context of knowledge bases. They also provide examples of possible applications for topology and ordered structures.

The purpose of this paper is to investigate $*$-reductions in a knowledge base and their invariant characterizations under homomorphisms. The remainder of this paper is organized as follows: Section 2 briefly recalls basic concepts related to rough sets, knowledge bases and consistent mappings; Section 3 examines the communication between knowledge bases; Section 4 explores $*$-reductions in a knowledge base and discusses their characterizations that are invariant under homomorphisms; and Section 5 provides a summary of the main points of this paper.

2. Preliminaries

Basic concepts related to rough sets, knowledge bases and consistent mappings will be briefly recalled in this section.

Throughout this paper, U denotes a non-empty finite set called the universe, 2^U denotes the set of all subsets of U, $2^{U\times U}$ denotes the set of all binary relations on U, $\mathcal{R}^\star(U)$ denotes the set of all equivalence relations on U, and $\Delta = \{(x,x) : x \in U\}$. All mappings are assumed to be surjective.

For $\mathcal{R} \subseteq 2^{U\times U}$, denote $ind(\mathcal{R}) = \bigcap_{R\in\mathcal{R}} R$.

It is easy to verify that if $\mathcal{R} \subseteq \mathcal{R}^\star(U)$, then $ind(\mathcal{R}) \in \mathcal{R}^\star(U)$.

For $R \in \mathcal{R}^\star(U)$, denote $[x]_R = \{y \in U : xRy\}$.

2.1. Rough sets

Let $R \in \mathcal{R}^\star(U)$. Then, the pair (U, R) is called a Pawlak approximation space. Based on (U, R), one can define the following two rough approximations:

$$\underline{R}(X) = \{x \in U : [x]_R \subseteq X\},$$

$$\overline{R}(X) = \{x \in U : [x]_R \cap X \neq \emptyset\}.$$

Then, $\underline{R}(X)$ and $\overline{R}(X)$ are called the Pawlak lower approximation and the Pawlak upper approximation of X, respectively. The boundary region of X is defined by the difference between these rough approximations, i.e., $bnd_R(X) = \overline{R}(X) - \underline{R}(X)$. A set is rough if its boundary region is not empty, i.e., if $\underline{R}(X) \neq \overline{R}(X)$.

2.2. Knowledge bases

Definition 2.1 [35]. A pair $(U, \mathcal{R})$ is called a knowledge base if $\mathcal{R} \subseteq \mathcal{R}^\star(U)$.

Definition 2.2 [35]. Let $(U, \mathcal{R})$ be a knowledge base, and let $\mathcal{P} \subseteq \mathcal{R}$.

(1) $\mathcal{P}$ is called a coordinate subfamily of $\mathcal{R}$ if $ind(\mathcal{P}) = ind(\mathcal{R})$.
(2) $R \in \mathcal{P}$ is called independent in $\mathcal{P}$ if $ind(\mathcal{P} - \{R\}) \neq ind(\mathcal{P})$; $\mathcal{P}$ is called an independent subfamily of $\mathcal{R}$ if, for each $R \in \mathcal{P}$, R is independent in $\mathcal{P}$.
(3) $\mathcal{P} \subseteq \mathcal{R}$ is called a knowledge reduction (for short, a reduction) of $\mathcal{R}$ if $\mathcal{P}$ is both coordinate and independent.

The set of all coordinate subfamilies and the set of all reductions of $\mathcal{R}$ are denoted by $co(\mathcal{R})$ and $red(\mathcal{R})$, respectively. Obviously,

$$\mathcal{P} \in red(\mathcal{R}) \iff \mathcal{P} \in co(\mathcal{R}) \text{ and for each } \mathcal{Q} \subset \mathcal{P}, \mathcal{Q} \notin co(\mathcal{R}).$$

Definition 2.3. Let $(U, \mathcal{R})$ be a knowledge base. Suppose that

$$core(\mathcal{R}) = \bigcap_{\mathcal{P}\in red(\mathcal{R})} \mathcal{P}.$$

Then, $core(\mathcal{R})$ is called the core of $\mathcal{R}$.

Proposition 2.4 [35]. *Let $(U, \mathcal{R})$ be a knowledge base. The following are equivalent:*