Knowledge-Based Systems 75 (2015) 78–86

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Approximate concepts acquisition based on formal contexts

Qing Wan, Ling Wei *

School of Mathematics, Northwest University, Xi'an, Shaanxi 710069, PR China

ARTICLE INFO

Article history:
Received 21 May 2014
Received in revised form 10 October 2014
Accepted 17 November 2014
Available online 25 November 2014

Keywords:
Formal concept
Approximate concept
Concept lattice
k-Grade relation
Rough set

ABSTRACT

Formal concepts are abstraction and formalization of the concepts in philosophy. Acquisition of formal concepts is the base of formal concept analysis. Besides the concrete concepts, uncertain but meaningful concepts sometimes are more appropriate for real life. The paper begins with this practical problem and obtains some interesting knowledge based on formal contexts. First, k-grade relation on the object set of a formal context is defined, and the different values of k in different cases are discussed in detail. Then the algorithms to acquire approximate concepts associated with each object are proposed and these approximate concepts are explained. Second, the relationships between approximate concepts associated with objects and concepts of concept lattice (property oriented concept lattice) are studied. Parallel to the above idea, k-grade relation on the attribute set is proposed dually, and approximate concepts associated with each attribute are obtained and interpreted. Meanwhile, the relationships between approximate concepts associated with attributes and concepts of concept lattice (object oriented concept lattice) are also discussed.

1. Introduction

Formal concept analysis (FCA) was proposed by German mathematician Wille [11,32] in 1982. The basic notions of FCA are *formal context*, *formal concept*, and the corresponding *concept lattice*, which depends on the binary relation between an object set and an attribute set. Concept lattice is the core data structure of FCA, it reveals the generalization and specialization relations among formal concepts. As a useful tool for data analysis, FCA has been applied to many fields. For example, Boucher-Ryan and Bridge [2] applied FCA to the relation between users and items in a collaborative recommender system. Jiang and Chute [13] developed a model for formalizing the normal forms of SNOMED CT expressions using FCA-based model. Li and Wei [16] introduced a FCA method to reduce the image feature data dimension. Xiao et al. [33] proposed a novel framework of image mining for robot vision based on concept lattice theory and cloud model theory. Belohlavek et al. [1] presented a new method of decision tree introduction based on FCA. Yang and Xu [34] put forward the method of decision making with uncertainty information based on the lattice-valued fuzzy concept lattice.

Moreover, the classical formal concept has been extended to other types, such as *property oriented concept* [7], *object oriented concept* [37], *dual concept* [3], *monotone concept* [5], *AFS formal concept* [28], and *mulit-adjoint concept* [18], where, mulit-adjoint concept was introduced as a new general approach to FCA in the framework of fuzzy FCA, and AFS formal concept can be viewed as the generalization and development of monotone concept. Ma and Zhang [19] studied the relationship between concept lattice and dual concept lattice. Yao [35] investigated the relations between property oriented concept lattice and object oriented concept lattice in detail and provided a more systematic examination of concept lattices in rough set theory. In fact, the introduction of property oriented concept and object oriented concept are on the basis of a pair of approximate operators, $\Diamond$ and $\Box$, which are similar to the upper and lower approximations in rough set theory.

Rough set theory (RST) [20,21] was introduced by Pawlak in 1982. RST is also an important tool for data analysis. The basic relation in RST is *an equivalence relation* defined on the object set, based on which, lower and upper approximations of a set are proposed. Now rough set model has been extended to many generalized versions, such as *rough fuzzy sets and fuzzy rough sets* [6], *decision-theoretic rough set* [40], *variable precision rough set* [41], *bayesian rough set* [27], *mulit-granulation rough set* [22], and *probabilistic rough set* [36]. Specially, for probabilistic rough set models, Yao [38] further studied the superiority of three-way decisions in these models. Since FCA and RST have strong connection in data analysis, the relations between FCA and RST are studied in recent years. Refs. [7,37] introduced the notions in concept lattice into rough set theory, such as property oriented concept and object oriented concept defined by Düntsch and Yao. Similarly, Refs.

* Corresponding author. Tel.: +86 29 88308433.
E-mail addresses: wqysbe@163.com (Q. Wan), wl@nwu.edu.cn (L. Wei).

http://dx.doi.org/10.1016/j.knosys.2014.11.020
0950-7051/

[12,25,39] added approximation operators to FCA, which became the new tools of data analysis. Saquer and Deogun [23] presented two different approaches to the concept approximation based on rough set theory and similarity measure. Lai and Zhang [14] provided a comparative study of concept lattices of fuzzy contexts based on FCA and RST. Shi et al. [26] studied relationship between the extents of concept lattice and the equivalence class in RST. Wei and Qi [29] discussed the relationship between concept lattice reduction and rough set reduction based on classical formal contexts. Shao et al. [24] proposed methods of rule acquisition and knowledge reduction in formal contexts, and investigated by employing rough set approaches.

A formal context (G,M,I) consists of an object set G, an attribute set M, and a relation I between G and M. A formal concept of the formal context is defined as a pair (A,B) satisfying some certain conditions, where A is the *extent* of the formal concept (A,B) and B is the *intent*. The formal concept (A,B) means that A is the maximal set of objects that possess all attributes in B and B is the maximal set of attributes shared by all objects in A. It is the abstraction of the concept in philosophy and precise knowledge.

Although precise knowledge is perfect to some extent, in reality, we cannot always achieve our expectation, and sometimes we need not get accurate knowledge. For example, before shopping, people may have many requirements in mind for the things they want to buy, but during the purchase, they have to reduce, adjust, or change their requirements. That is, it is common that there is difference between our expectation and the reality. The above process is similar to choosing an approximate value when the accurate value cannot be obtained. Li [15] focused on the issues of approximate concept construction in incomplete decision contexts. In order to identify different concepts that are semantically close, Formica [8,9] proposed the methods for assessing similarity of concepts in FCA. Ref. [9] improved the theory in [8], in determining the similarity of concept descriptors (attributes) by using the information content approach. Based on the proposal of Ref. [9], Formica [10] suggested a measure for evaluating the similarity of concepts in FFCA (fuzzy formal concept analysis). Therefore, the main aim of this paper is to obtain some meaningful knowledge, approximate concepts, based on a formal context, each approximate concept is denoted by a binary pair (objects, attributes), which has the same expression form as a formal concept, but may not be a formal concept or precise. The approximate concepts are on the basis of the k-grade relation defined on the object set and the attribute set respectively.

The paper is organized as follows. Section 2 reviews the basic definitions in formal concept analysis. Section 3 first introduces the definition of k-grade relation based on an object set, studies its properties, and discusses the range of k in different cases. And then, we give the definition of k-left (right) neighborhood approximate concepts associated with an object and the acquisition algorithms of k-left (right) neighborhood approximate concepts associated with each object, and reveal the relationships between the corresponding k-left (right) neighborhood approximate concepts associated with an object and property oriented concepts (formal concepts). According to the duality principle, Section 4 proposes the definition of k-grade relation based on an attribute set, and describes its properties. Analogously, we define k-left (right) neighborhood approximate concepts associated with an attribute, and propose the corresponding algorithms of approximate concepts acquisition. Then we discuss the relationships between the corresponding k-left (right) neighborhood approximate concepts associated with an attribute and object oriented concepts (formal concepts). Finally, Section 5 concludes this paper.

2. Preliminaries

In this section, we recall some basic notions and properties in formal concept analysis [11,32].

Definition 1 [11]. A formal context (G,M,I) consists of two sets G and M and a relation I between G and M. The elements of G are called the objects and the elements of M are called the attributes of the context. In order to express that an object g is in a relation I with an attribute m, we write gIm or $(g,m) \in I$ and read it as "the object g has the attribute m".

Let (G,M,I) be a formal context. For $A \subseteq G,\ B \subseteq M$, two operators are defined as follows:

$$\begin{aligned} A^* &= \{m \in M \mid (g,m) \in I \text{ for all } g \in A\}, \\ B' &= \{g \in G \mid (g,m) \in I \text{ for all } m \in B\}. \end{aligned} \tag{1}$$

(A,B) is called a formal concept, for short, a concept, if and only if, $A^* = B,\ A = B'$; where, A is called the extent of the formal concept and B is called its intent. The set of all concepts of (G,M,I) is denoted by $L(G,M,I)$.

The concepts of a formal context (G,M,I) are ordered by:

$(A_1,B_1) \leqslant (A_2,B_2) \iff A_1 \subseteq A_2 (\iff B_1 \supseteq B_2).$

In a formal context (G,M,I), for any $A_1,\ A_2,\ A \subseteq G$ and $B_1,\ B_2,\ B \subseteq M$, the following properties hold.

1. $A_1 \subseteq A_2 \Rightarrow A_2^* \subseteq A_1^*$,
   $B_1 \subseteq B_2 \Rightarrow B_2' \subseteq B_1'$.
2. $A \subseteq A^{*\prime}, \quad B \subseteq B^{\prime *}$.
3. $A^* = A^{*\prime *}, \quad B' = B^{\prime * \prime}$.
4. $A \subseteq B' \iff B \subseteq A^*$.
5. $(A_1 \cup A_2)^* = A_1^* \cap A_2^*$,
   $(B_1 \cup B_2)' = B_1' \cap B_2'$.
6. $(A_1 \cap A_2)^* \supseteq A_1^* \cup A_2^*$,
   $(B_1 \cap B_2)' \supseteq B_1' \cup B_2'$.
7. $(A^{*\prime},A^*)$ and $(B',B^{\prime *})$ are both concepts.

For $(A_1,B_1),\ (A_2,B_2) \in L(G,M,I)$, the infimum and supremum are given by:

$$(A_1,B_1) \wedge (A_2,B_2) = (A_1 \cap A_2, (B_1 \cup B_2)^{\prime *}),$$

$$(A_1,B_1) \vee (A_2,B_2) = ((A_1 \cup A_2)^{*\prime}, B_1 \cap B_2).$$

Thus, $L(G,M,I)$ is a complete lattice and is called a concept lattice.

For $A \subseteq G,\ B \subseteq M$, we define $A^{*B} = \{m \mid m \in B,\ \forall g \in A,\ (g,m) \in I\}$, $B^{\prime A} = \{g \mid g \in A,\ \forall m \in B, (g,m) \in I\}$. In this paper, ${}^*M$ and ${}'G$ are shortly denoted by * and $'$.

For convenience, for any $g \in G,\ m \in M,\ \{g\}^*$ and $\{m\}'$ are replaced by g^* and m', respectively. In Ref. [11], Wille called $(g^{*\prime}, g^*)$ an object concept, and $(m', m^{\prime *})$ an attribute concept. If for any $g \in G,\ g^* \neq \emptyset,\ g^* \neq M$, and for any $m \in M,\ m' \neq \emptyset,\ m' \neq G$, then the formal context is called canonical. That is, there are neither full row/column nor empty row/column in a formal context. In this paper, we suppose that all formal contexts are finite and canonical.

The following example illustrates the above definitions.

Example 1 [31]. Table 1 is a formal context (G,M,I), which describes investigation information of a buyer for some buildings. $G = \{1,2,3,4,5,6\}$ is a set of buildings, and $M = \{a,b,c,d,e\}$ is the set of attributes about the buildings. The attributes are defined as follows: a-price of the house, b-traffic situation, c-entertainment instruments, d-estate management, e-architectural quality. $\times$ means that the buyer is satisfied with the item, and if blank