ARTICLE IN PRESS
JID: KNOSYS [m5G;February 28, 2018;23:35]
Knowledge-Based Systems 000 (2018) 1–12
Contents lists available at ScienceDirect
Knowledge-Based Systems
journal homepage: www.elsevier.com/locate/knosys

A graph approach for knowledge reduction in formal contexts

Jinkun Chen[a,b], Jusheng Mi[a,*], Yaojin Lin[c,d]

[a] College of Mathematics and Information Science, Hebei Normal University, Shijiazhuang, Hebei 050024, PR China
[b] School of Mathematics and Statistics, Minnan Normal University, Zhangzhou 363000, PR China
[c] School of Computer Science, Minnan Normal University, Zhangzhou 363000, PR China
[d] Key Laboratory of Data Science and Intelligence Application, Fujian Province University, Zhangzhou 363000, PR China

ARTICLE INFO

Article history:
Received 14 October 2017
Revised 25 February 2018
Accepted 27 February 2018
Available online xxx

Keywords:
Formal contexts
Formal concept analysis
Graph theory
Knowledge reduction
Vertex covers

ABSTRACT

Knowledge reduction in formal concept analysis (FCA) is an important procedure for knowledge processing and data analysis. Currently, the knowledge reduction in FCA based on granular computing (GrC) provides another way to analyze and represent the structure of concept lattices. However, the granular reduction method is based on Boolean reasoning and thus is an NP-hard problem. It is therefore natural to develop some heuristic methods to deal with this problem especially for the large data. In this paper, a new framework based on graph theory is used to study the granular reduction in FCA. A graph representation for the granular reduction is first investigated. The results in this paper show that the granular reduction computation in FCA can be translated into a graph optimization problem. Two new heuristic graph-based algorithms for the granular reduction in formal contexts and formal decision contexts are then respectively presented. Furthermore, numerical experiments are conducted to evaluate the effectiveness of the proposed methods.

1. Introduction

Knowledge reduction in FCA is a basic issue for knowledge representation and data analysis [10,55]. In the framework of FCA, the main aim of knowledge reduction is to find a minimal subset of attributes that preserves the predefined properties. Over the past ten years, knowledge reduction in FCA has attracted more and more attention and has been successfully applied in many fields [1,8,11,16–19,21,24–26,30,31,33–37,40,45,49,53,54,58,62].

A variety of reduction methods have been developed to find the knowledge reduction in FCA [14,20,22,23,27,28,43,44,47,57]. In terms of formal contexts, the current studies of knowledge reduction for concept lattices can be classified into two main categories: one is the reduction for formal contexts without decision attributes, and the other is the reduction for formal decision contexts. For the first category, a reduction of a context by deleting rows or columns was first introduced in [10]. In 2005, Zhang et al. [60,61] proposed a new notion of attribute reduction in which the structure and hierarchy of the lattice preserve unchanged. Furthermore, Wei [51] introduced an effective method for obtaining the knowledge reduction of a formal context. Qi [39] proposed a reduction method in formal contexts based on a new discernibility matrix, and it showed that the new method requires less calculation than that in [60,61]. Liu et al. [33] investigated the reduction of object and attribute oriented concept lattices based on rough set theory. The relation of reduction between object and attribute oriented concept lattices was also discussed in [49], and an effective method to obtain the reduction in the object and attribute oriented concept lattices was proposed. Based on the idea of knowledge reduction in rough set theory, Mi et al. [37] formulated a Boolean method to obtain all reducts of a formal context by using the discernibility function. To find the same expression of attribute reduction in rough set theory, Wang [50] presented a new reduction method in a formal context based on congruence relations. It should be noted that most of the above methods for knowledge reduction in formal contexts are based on Boolean reasoning, but this is very time-consuming, certainly for large formal contexts. In fact, as shown in [16], one can use the clarification and reduction method proposed in [10] to obtain the knowledge reduction in formal contexts and it is strictly superior to the methods based on Boolean reasoning. In generally speaking, the similar results have been stated in [4], we can use the covering generalized rough set method to easily calculate all reducts in a formal context. Different from the reduction framework proposed in [10,60,61], Wu et al. [56] introduced the notion of knowledge reduction called granular reduction in formal contexts from the viewpoint of keeping object granule unchanged. Furthermore, an approach to granular reduction in formal contexts was also investigated based on discernibil-

* Corresponding author.
E-mail addresses: cjk99@163.com (J. Chen), mijsh@263.net (J. Mi).

https://doi.org/10.1016/j.knosys.2018.02.039
0950-7051/

Please cite this article as: J. Chen et al., A graph approach for knowledge reduction in formal contexts, Knowledge-Based Systems (2018), https://doi.org/10.1016/j.knosys.2018.02.039

ity matrices and functions [46]. Shao et al. [42] first studied the relations between granular reduct and dominance reduct in formal contexts, and proposed an algorithm for granular reduction based on dominance relation and Boolean reasoning. However, the above methods for granular reduction are time-consuming in terms of computation time.

Formal decision contexts can be regarded as an extension of formal contexts, and the knowledge reduction in formal decision contexts is more complex than that of formal contexts. As concluded in [25], there are many reduction methods for formal decision contexts. For example, Wei et al. [52] introduced the notions of consistent formal decision contexts and the reduction theory. Then, based on the discernibility matrix function, two reduction methods were developed for consistent formal decision contexts. Wang and Zhang [48] presented a different one called generalized consistent formal decision contexts and provided the approaches to knowledge reduction in generalized consistent formal decision context of concept lattices by using the discernibility matrix. Li et al. [19] proposed a rule-acquisition-oriented framework of knowledge reduction in formal decision contexts whose aim is to preserve the number of non-redundant decision rules. Moreover, a heuristic algorithm of searching for a reduct of a consistent formal decision context was designed. Note that the above reduction methods are only suitable for consistent formal decision contexts. To overcome this problem, Li et al. [20] gave a new reduction method which can be suitable for inconsistent formal decision contexts, and they also developed a corresponding reduction method based on the discernibility matrix function. By employing the notion of decision information table, Shao et al. [41] constructed a knowledge-lossless method for complexity reduction in formal decision contexts. Based on the GrC, Wu et al. [56] proposed a new reduction framework for formal decision contexts whose aim is to avoid the redundancy of attributes while preserving granular structure of concept lattices. In addition, they developed an algorithm based on Boolean reasoning to obtain the granular reducts of a formal decision context. It is worth noting that the reduction framework proposed in [56] is different from the other reduction frameworks (see [25] for details). Furthermore, the time complexities of most of the above reduction methods are all exponential because they all depend on Boolean reasoning. To overcome this problem, Huang et al. [14] proposed a heuristic method based on the information entropy to obtain the granular reducts of an inconsistent formal decision context.

As stated in [15,29,38,54,56,57,59], the granular reduction framework based on GrC provides us with another way to analyze, understand and represent the structure of concept lattices. However, the computation of the granular reduct in FCA is very time-consuming. In fact, finding the set of all reducts or finding a minimum reduct (i.e., a reduct with the minimum number of attributes) is an NP-hard problem. To our best knowledge, there is little attention paid to developing the heuristic algorithm for searching the granular reduct in FCA. Although [14] provided a heuristic algorithm for knowledge reduction in FCA, it is also computationally expensive especially for the data set with large attributes, since $\mathbf{O}(|U|^2(|A|^3+|D|))$ operations are necessary to implement, where $|U|$, $|A|$ and $|D|$ are the numbers of objects, conditional attributes and decision attributes, respectively.

The purpose of this paper is to give a new mechanism to study the granular reduction in FCA based on graph theory and design a more efficient algorithm for searching a minimum granular reduct. As mentioned earlier, the purpose of the granular reduct is to find a prime implicant of a discernibility function. In other words, the granular reduct is a process that selects the minimally sized attribute subset which can cover each element of the discernibility matrix. This is similar to what is done in the problem of the vertex cover in graph theory. In fact, a minimal vertex cover of a graph is a minimal subset of vertices that can cover every e.g. [2]. Also, the minimal vertex cover computation can be translated into the calculation of prime implicants of a Boolean function [9,32]. It seems that there is some kind of natural connection between the knowledge reduction in formal contexts and the minimal vertex cover of graphs. The combination of granular reduction and graph theory is seen to offer us another angle to study the knowledge reduction in FCA. But more importantly, this can provide new techniques with low time complexity for knowledge reduction in FCA since the graph-based methods for the vertex cover problem [12,13] are really quite fast and can be applied to the knowledge reduction in FCA. The main contribution of this paper is threefold: 1) A new graph representation for granular reduction in FCA is first established. 2) Two heuristic algorithms for granular reduction in FCA are given. 3) This study can provide us with new methods for the knowledge reduction in FCA since the proposed method can be easily extended to the other reduction framework.

The rest of this paper is organized as follows. In Section 2, some basic notions about FCA and graph theory are reviewed. In Section 3, a graph representation for granular reduction in formal contexts is established. Moreover, a heuristic algorithm for granular reduction based on graph theory is designed. In Section 4, we investigate the granular reduction of formal decision contexts from the viewpoint of the vertex cover problem. Furthermore, a new heuristic algorithm for the granular reduction in formal decision contexts based on graph theory is presented. In Section 5, some experiments are given to show the effectiveness of the proposed method. Finally, some conclusions are drawn in Section 6.

2. Preliminaries

In this section, we review some basic notions related to FCA and graph theory [2,3,5–7,12,13,55,56]. In the following, we always assume that the universe of discourse is a non-empty finite set.

2.1. Knowledge reduction in FCA

Definition 1 [55,56]**.** A formal context is a triplet $F=(U,A,I)$, where $U=\{x_1,x_2,\ldots,x_n\}$ is a set of objects, $A=\{a_1,a_2,\ldots,a_m\}$ is a set of attributes, and I is a binary relation between U and A, where $(x, a)\in I$ means that the object x has the attribute a. For any $C\subseteq A$, we can obtain a formal context called a sub-context of F and is denoted by $F_C=(U,C,I_C)$, where $I_C=I\cap(U\times C)$.

Usually, a formal context can be represented by a table with the values 0 and 1, in which 1 means that the row object has the column attribute. For $X\subseteq U$ and $B\subseteq A$, two operators are defined as [55]:

$$X^{*A}=\{a\in A:\forall x\in X,\ (x,a)\in I\},$$
$$B'^{A}=\{x\in U:\forall a\in B,\ (x,a)\in I\}.$$

X^{*A} is the set of attributes shared by all the objects in X and B'^{A} is the set of objects that possesses all the attributes in B. For simplicity, we write x^{*A} instead of $\{x\}^{*A}$ for all $x\in U$ and write a'^{A} instead of $\{a\}'^{A}$ for all $a\in A$. To simplify notation, when A is clear, we will omit the superscript A.

For the sub-context $F_C=(U,C,I_C)$ and $X\subseteq U$, $B\subseteq C$, the above operators can be rewritten by substituting A into C as follows:

$$X^{*C}=\{a\in C:\forall x\in X,\ (x,a)\in I\},$$
$$B'^{C}=\{x\in U:\forall a\in B,\ (x,a)\in I\}.$$

It is easy to check that $X^{*C}=X^{*A}\cap C=X^{*}\cap C$.

Definition 2 [55]**.** Let $F=(U,A,I)$ be a formal context. For $X\subseteq U$, $B\subseteq A$, a pair (X, B) is called a formal concept of F if $X=B'$ and $X^{*}=B$. We call X the extent and B the intent of (X, B), respectively.

Please cite this article as: J. Chen et al., A graph approach for knowledge reduction in formal contexts, Knowledge-Based Systems (2018), https://doi.org/10.1016/j.knosys.2018.02.039