Information Sciences 236 (2013) 66–82

Contents lists available at SciVerse ScienceDirect

Information Sciences

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

Formal concept analysis based on the topology for attributes of a formal context

Zheng Pei [a,*], Da Ruan [b], Dan Meng [c], Zhicai Liu [a]

[a] Center for Radio Administration & Technology Development, Xihua University, Chengdu, Sichuan 610039, China
[b] Belgian Nuclear Research Centre (SCK•CEN), Boeretang 200, B-2400 Mol, & Dept. of Applied Math. & Computer Science, Ghent University, B-9000 Gent, Belgium
[c] School of Economics Information Engineering, Southwestern University of Finance and Economics, Chengdu 610074, China

ARTICLE INFO

Article history:
Received 27 October 2008
Received in revised form 24 January 2013
Accepted 10 February 2013
Available online 26 February 2013

Keywords:
Formal concept analysis
The lexicographic tree approach
Set-valued mapping
Approximation space
Topological space
KDD

ABSTRACT

Formal concept analysis (FCA) concerns the hierarchical structures induced by a binary relation between a pair of sets, it is widely applied in data analysis, information retrieval, and knowledge discovery. Within the framework of FCA, computing all formal concepts is the main challenge due to its exponential complexity. It has be proved that all formal concepts is a closure system on objects, hence, the closure operator on objects are general used to generate all formal concepts of a formal context. In the lexicographic tree approach, a base on objects is used to generate extensions of all formal concepts and construct the formal concept lattice. Inspired from the approach, we concentrate a base on attributes and generate intensions of all formal concepts in this paper. To this end, we firstly analyze an example, the lexicographic tree approach is used to generate all formal concepts, from the time complexity point of view, we present that it is not trivial to generate the base on attributes by a simple symmetrical way from the methods based on objects. Then, we deduce a set-valued mapping from attributes to the power set of attributes in a formal context and define a binary relation on attributes by the set-valued mapping. Using the binary relation on attributes, we construct an approximation space and a topology for attributes, respectively, and obtain a base for the topology. We prove that intensions of all formal concepts are included in the topology for attributes, this means that the base can be used to generate intensions of all formal concepts of the formal context and construct the formal concept lattice. More general, our results represent relationships and the hierarchical structures among attributes of the formal context, we present some typical applications, in which the topology for attributes and the base for the topology are applied for association rules discovery from a formal context and linguistic concept analysis.

1. Introduction

Knowledge representation is one of the hallmarks of artificial intelligence systems. According to Davis et al.'s knowledge representation principle [6,7], a knowledge representation owns that (1) It is most fundamentally a surrogate, a substitute for the thing itself, used to enable an entity to determine consequences by thinking rather than acting; (2) It is a set of ontological commitments; (3) It is a fragmentary theory of intelligent reasoning; (4) It is a medium for pragmatically efficient computation; (5) It is a medium of human expression.

* Corresponding author.
E-mail addresses: zhengpei@mail.xhu.edu.cn (Z. Pei), druan@sckcen.be (D. Ruan).

0020-0255/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2013.02.027

Formal concept analysis (FCA) presented by Wille [44–46] is a discipline that studies the hierarchical structures induced by a binary relation between the set of objects and attributes. As the core of a mathematical theory of FCA, all formal concepts or the formal concept lattice can be used to represent relationships between objects and attributes or conceptual hierarchies, which are inherent in data. FCA starts with a formal context T, more general, called (conceptual) information systems, which are a tabular form of an object-attribute value relationship [2,8,9,13–15,32,40]. A formal concept of T is formed by a pair of the extension (a subset of objects) and the intension (a subset of attributes), in which, all elements of the extension possess attributes of the intension and all elements of the intension are possessed by elements of the extension. The set of all formal concepts of T together with the order relation between two formal concepts (inclusion relation on all extensions or intensions) is a complete lattice. As a formalized method of knowledge representation, FCA is now considered as the mathematical backbone of conceptual knowledge processing (CKP), a theory located in computer science, having as tasks to provide methods and tools for human oriented, concept-based knowledge processing [35]. In fact, human (or machine) reasoning cannot directly deal with objects in the world, but only with an internal substitute, in FCA, formal contexts, formal concepts, hierarchical structures, association rules, etc. are surrogates which can be applied for reason. Moreover, combining FCA with description logics or conceptual graphs leads to the development of contextual logic [2,8,36,42,43]. Compared with other formalized knowledge representations, the formal concept lattice often supports a more structured representation and facilitates retrieval of the stored information [11,16,25,34].

Recently, fuzzy concepts and fuzzy concept lattice as generalization of FCA have been widely studied [1,10,22–24,50]. In knowledge discovery in databases (KDD), FCA has been applied for implication and association rules discovery and reduction, and improving the response times of algorithms to mine association rules [4,5,37,39,51]. In both theoretical analysis and applications, computing the formal concept lattice is the most basic issue that has been investigated by many researchers [20,21,26,37,38,40,41]. In this paper, we present an alternative method to generate the formal concept lattice, i.e., the topology for attributes of a formal context can be used to obtain intensions of all formal concepts and construct the formal concept lattice. More general, we provide some examples, in which the topology for attributes and the base for the topology are applied for association rules discovery and linguistic concept analysis.

Let $T=(G,M,I)$ be a formal context, G and M be non-empty finite sets, i.e., sets of objects and attributes, respectively, $I: G\times M\to\{0,1\}$ be a binary relation. For any $g\in G$ and $m\in M, I(g,m)=1$ means that the object g possesses the attribute m. The following two set-valued mappings are used to define a formal concept

$$\begin{aligned}
&\uparrow:\mathcal{P}(G)\to\mathcal{P}(M),\\
&\quad X\mapsto\ \uparrow(X)\triangleq X^{\uparrow}=\{m\in M|\ \forall g\in X, I(g,m)=1\}, && (1)\\
&\downarrow:\mathcal{P}(M)\to\mathcal{P}(G),\\
&\quad Y\mapsto\ \downarrow(Y)\triangleq Y^{\downarrow}=\{g\in G|\ \forall m\in Y, I(g,m)=1\}. && (2)
\end{aligned}$$

where $\mathcal{P}(G)$ and $\mathcal{P}(M)$ are power sets of G and M, respectively. In $T=(G,M,I)$, a formal concept is a pair $(A,B)\in\mathcal{P}(G)\times\mathcal{P}(M)$ such that $A^{\uparrow}=B$ and $B^{\downarrow}=A$. The sets A and B are the extension and intension of (A,B), respectively. If L_T is the set of all formal concepts, and $\forall(A_1,B_1),(A_2,B_2)\in L_T, (A_1,B_1)\vee(A_2,B_2)=((A_1\cup A_2)^{\uparrow\downarrow},B_1\cap B_2),(A_1,B_1)\wedge(A_2,B_2)=(A_1\cap A_2,(B_1\cup B_2)^{\downarrow\uparrow})$, then $(L_T,\vee,\wedge)$ is a lattice, i.e., the formal concept lattice of $T=(G,M,I)$. It is a complete lattice due to $\forall(A_i,B_i)\in L_T,\ i=1,\ldots,n$,

$$\bigvee_{i=1}^{n}(A_i,B_i)=\left(\left(\bigcup_{i=1}^{n}A_i\right)^{\uparrow\downarrow},\bigcap_{i=1}^{n}B_i\right),\quad \bigwedge_{i=1}^{n}(A_i,B_i)=\left(\bigcap_{i=1}^{n}A_i,\left(\bigcup_{i=1}^{n}B_i\right)^{\downarrow\uparrow}\right). \tag{3}$$

Formally, the formal concept lattice $(L_T,\vee,\wedge)$ of $T=(G,M,I)$ is performed in the set $C=\{(A,B)|A\subseteq G,B\subseteq M\}$ [13], in which, (A,B) is a formal concept if and only if $A^{\uparrow}=B$ and $B^{\downarrow}=A$, in the worst case, i.e., $T=(\{1,\ldots,n\},\{1,\ldots,n\},\neq),\ (L_T,\vee,\wedge)$ has 2^n formal concepts [21,37], hence, computing all formal concepts has exponential worst-case complexity, a better algorithm for computing all formal concepts evidently becomes the main challenge in FCA. A formal concept (A,B) of $T=(G,M,I)$ satisfies $A^{\uparrow\downarrow}=B^{\downarrow}=A$ and $B^{\downarrow\uparrow}=A^{\uparrow}=B$, this means that all formal concepts can be generated by closure operators $\uparrow\downarrow$ or $\downarrow\uparrow$, in which, $\downarrow\uparrow$ can be used to generate intensions of all formal concepts $C=\{B\subseteq M|B^{\downarrow\uparrow}=B\}\subseteq\mathcal{P}(M)$, i.e., C is a closure system on M defined by $h:\mathcal{P}(M)\to\mathcal{P}(M)$ and $B\prime\mapsto\bigcap_{B\in C,B\prime\subseteq B}B$, by using the weight function s on $\mathcal{P}(M)$ and the closure operator h, the TITANIC algorithm for computing (iceberg) concept lattices is proposed in [37]. By using a base on objects of a formal context, the lexicographic tree approach is proposed in [26], the method generates extensions of all formal concepts (A,B) according to the base $\mathcal{B}=\{G-m^{\downarrow}|m\in M\}$, its advantage is that the formal concept lattice is isomorphic to the family generated by the base $\mathcal{B}$.

Compared with objects of the formal context, attributes are relatively fixed and less than objects. More general, in concept-based knowledge processing, human rather deal with relationship among attributes than objects in practice, e.g., association rules are to determine association relationship among attributes of objectors. All of these mean that considering relationship among attributes and generating formal concepts based on attributes are more realistic. Inspired from the lexicographic tree approach, our attention concentrates on generating relationship among attributes and intensions of all formal concepts by a base on attributes of a formal context in this paper. Based on our initial works [17,31,33], we firstly construct the topology for attributes, then obtain a base for the topology for attributes. We prove that intensions of all formal concepts are included in the topology for attributes. Furthermore, the topology for attributes provides us more knowledge about attributes than intensions of all formal concepts, from the binary relation point of view, intensions of all formal concepts are decided by the equivalent relation on the set of attributes, but the topology for attributes is decided by a reflexive and transitive relation on the set