Information Sciences 260 (2014) 149–170

Contents lists available at ScienceDirect

Information Sciences

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

Granularity of attributes in formal concept analysis

Radim Belohlavek [a], Bernard De Baets [b], Jan Konecny [a,*]

[a] Dept. of Computer Science, Palacký University, Olomouc, Czech Republic
[b] Dept. of Mathematical Modelling, Statistics and Bioinformatics, Ghent University, Ghent, Belgium

A R T I C L E   I N F O

Article history:
Received 31 January 2013
Received in revised form 31 August 2013
Accepted 19 October 2013
Available online 1 November 2013

Keywords:
Formal concept analysis
Concept lattice
Binary data
Algorithms for data management
Interactive data exploration

A B S T R A C T

We propose a method to control the structure of concept lattices derived from Boolean data. Concept lattices represent the basic structure utilized in formal concept analysis. Their structure is of primary importance for the analysis and understanding of the input data. Our method enables to control the structure of the derived concept lattice by specifying granularity levels of attributes, thus in a sense by focusing the lenses through which we perceive and conceptually carve up the world. The granularity levels are chosen by a user based on his expertise and experimentation with the data. If the resulting formal concepts are too specific and there is a large number of them, the user can choose to use a coarser level of granularity. The resulting formal concepts are then less specific and can be seen as resulting from a zoom-out. In a similar way, one may perform a zoom-in to obtain finer, more specific formal concepts. The paper presents a basic study of this topic. We describe the motivations, the method, a theoretical insight, zoom-in and zoom-out algorithms, and experiments demonstrating the method.

1. Introduction and problem setting

Formal concept analysis (FCA) is a method for data analysis and knowledge discovery with growing popularity across various domains [2,10,16,22]. FCA proved useful for the organization of web search results into a hierarchical structure of concepts based on common topics [5], information retrieval [6,17], hierarchical analysis of software code [7,19,21], visualization in software engineering [9,20], detecting suspects in human trafficking [18], analysis of questionnaire data [4], and mining gene expression data [12], to name just a few. Further references to applications of FCA can be found in [5,11]. Interestingly, the calculus of FCA proved useful in preprocessing of binary data, particularly when removal of redundancy plays a role, see e.g. [3] for the role of FCA in optimal decompositions of binary matrices and [24] for the role of FCA in mining non-redundant association rules. A distinguishing feature of FCA is an inherent integration of three components: discovery of clusters (so-called formal concepts) in data, discovery of data dependencies (so-called attribute implications) in data, and visualization of formal concepts and attribute implications in a single hierarchical diagram (so-called concept lattice).

In the basic setting of FCA, it is assumed that the input data is in the form of a table containing 0s and 1s (Boolean matrix) describing which attributes (columns) apply to which objects (rows). The basic structure derived from the input data is a partially ordered set, called a concept lattice, consisting of particular biclusters which are called formal concepts. The choice of attributes is a fundamental step in FCA because it determines the structure of formal concepts and the concept lattice. In the basic setting of FCA, the attributes are considered fixed.

* Corresponding author.
E-mail addresses: radim.belohlavek@acm.org (R. Belohlavek), bernard.debaets@ugent.be (B. De Baets), jan.konecny@upol.cz (J. Konecny).

0020-0255/$ - see front matter
http://dx.doi.org/10.1016/j.ins.2013.10.021

Suppose now a user wants to use a concept lattice to visualize the structure of car accidents in a certain area. He might want to represent the accidents by objects and use attributes of accidents such as "technical cause", "driver's fault" (describing the cause of the accident), "before noon", "after noon" (describing when the accident happened), etc. The concept lattice over such attributes may not contain interesting formal concepts because the selected attributes are too coarse, resulting in a low level of granularity of concepts. If one uses attributes with a higher level of granularity instead, such as "alcohol", "skid", ..., "early morning", "late evening", etc., the concept lattice may reveal some interesting patterns, such as a formal concept containing "alcohol" and "late evening" among its attributes. If such concept applies to a large number of accidents, it may reveal interesting information. Such a concept may not be detectable when using coarser attributes. On the other hand, too high a level of granularity may result in overly specific formal concepts which may too be of little interest to the user. The capability to change the level of granularity of an attribute in FCA to capture relevant patterns in data is therefore a natural requirement. Note that the importance of granulation and granularity in human reasoning has repeatedly been emphasized by Zadeh [23].

This paper presents a simple framework that enables the user to change the level of granularity of attributes in FCA. The paper is organized as follows. Section 2 contains preliminaries from formal concept analysis. In Section 3, we present our approach and basic results. In Section 4, we present algorithms that enable us to compute from a concept lattice over attributes with given levels of granularity a concept lattice over attributes with new, user-specified levels of granularity. Section 5 contains illustrative examples and an experimental evaluation of the algorithm. In Section 6, we conclude the paper and outline future research.

2. Basic notions from formal concept analysis

For a comprehensive background on FCA, we refer to [11]. The input data to FCA consists of a table with rows and columns representing objects and attributes, respectively. An entry representing object x and attribute y contains $\times$ (or 1) if x has y (y applies to x), otherwise the entry contains a blank (or 0). Formally, such a table is represented by a triplet $\langle X, Y, I\rangle$, called a *formal context*, in which I is a binary relation between X and Y and $\langle x, y\rangle \in I$ means that object x has attribute y. For every set $A \subseteq X$ of objects in X, denote by $A^{\uparrow}$ the subset of attributes in Y defined as

$$A^{\uparrow} = \{y \in Y \mid \text{for each } x \in A: \ \langle x, y\rangle \in I\}.$$

Similarly, for $B \subseteq Y$ denote by $B^{\downarrow}$ the subset of X defined as

$$B^{\downarrow} = \{x \in X \mid \text{for each } y \in B: \ \langle x, y\rangle \in I\}.$$

A *formal concept* of $\langle X, Y, I\rangle$ is a pair $\langle A, B\rangle$ of $A \subseteq X$ and $B \subseteq Y$ satisfying $A^{\uparrow} = B$ and $B^{\downarrow} = A$. That is, a formal concept consists of a set A of objects and a set B of attributes such that A is the set of all objects in X sharing all attributes in B and, conversely, B is the collection of all attributes in Y shared by all objects from A. A and B are called the *extent* and the *intent* of $\langle A, B\rangle$, respectively. This definition formalizes the traditional approach to concepts which is due to Port-Royal logic [13]. As an example, consider the concept *dog*: its extent is the collection of all dogs, while its intent is the collection of all attributes of dogs such as "has four legs", "barks", etc. The set

$$\mathcal{B}(X, Y, I) = \{\langle A, B\rangle \mid A^{\uparrow} = B, B^{\downarrow} = A\}$$

of all formal concepts of $\langle X, Y, I\rangle$, called the *concept lattice of I*, can be equipped with a partial order $\leqslant$ defined by $\langle A_1, B_1\rangle \leqslant \langle A_2, B_2\rangle$ if $A_1 \subseteq A_2$ (or, equivalently, $B_2 \subseteq B_1$). Therefore, $\leqslant$ represents a *subconcept-superconcept hierarchy* due to which *dog* is a subconcept of *mammal*, etc. $\mathcal{B}(X, Y, I)$ happens to be a concept lattice whose structure is described by the so-called basic theorem of concept lattices [11], whose first part is the subject of the following theorem:

Theorem 1. *$\mathcal{B}(X, Y, I)$ equipped with $\leqslant$ is a complete lattice where the infima and suprema are given by*

$$\bigwedge_{j \in J} \langle A_j, B_j\rangle = \left\langle \bigcap_{j \in J} A_j, \left(\bigcup_{j \in J} B_j\right)^{\downarrow\uparrow} \right\rangle,$$

$$\bigvee_{j \in J} \langle A_j, B_j\rangle = \left\langle \left(\bigcup_{j \in J} A_j\right)^{\uparrow\downarrow}, \bigcap_{j \in J} B_j \right\rangle.$$

3. Granularity of attributes

In this section, we present an extension of the basic setting of FCA that enables a user to control the level of granularity of attributes. Granulation is an important phenomenon performed successfully by humans. Basically, granulation consists in