Formal Concept Analysis Constrained by Attribute-Dependency Formulas

Radim Bělohlávek and Vladimír Sklenář

Dept. Computer Science, Palacký University, Tomkova 40, CZ-779 00 Olomouc, Czech Republic
{radim.belohlavek, vladimir.sklenar}@upol.cz

Abstract. An important topic in formal concept analysis is to cope with a possibly large number of formal concepts extracted from formal context (input data). We propose a method to reduce the number of extracted formal concepts by means of constraints expressed by particular formulas (attribute-dependency formulas, ADF). ADF represent a form of dependencies specified by a user expressing relative importance of attributes. ADF are considered as additional input accompanying the formal context $\langle X, Y, I\rangle$. The reduction consists in considering formal concepts which are compatible with a given set of ADF and leaving out noncompatible concepts. We present basic properties related to ADF, an algorithm for generating the reduced set of formal concepts, and demonstrating examples.

1 Preliminaries and Problem Setting

We refer to [6] (see also [14]) for background information in formal concept analysis (FCA). We denote a formal context by $\langle X, Y, I\rangle$, i.e. $I \subseteq X \times Y$ (object-attribute data table, objects $x \in X$, attributes $y \in Y$); the concept deriving operators by $^{\uparrow}$ and $^{\downarrow}$, i.e. for $A \subseteq X$, $A^{\uparrow} = \{y \in Y \mid \text{ for each } x \in A : \langle x, y\rangle \in I\}$ and dually for $^{\downarrow}$; a concept lattice of $\langle X, Y, I\rangle$ by $\mathcal{B}(X, Y, I)$, i.e. $\mathcal{B}(X, Y, I) = \{\langle A, B\rangle \in 2^X \times 2^Y \mid A^{\uparrow} = B,\, B^{\downarrow} = A\}$.

An important aspect of FCA is a possibly large number of formal concepts in $\mathcal{B}(X, Y, I)$. Very often, the formal concepts contain those which are in a sense not interesting for the expert. In this paper, we present a way to naturally reduce the number of formal concepts extracted from data by taking into account information additionally supplied to the input data table (formal context). We consider a particular form of the additional information, namely, a form of particular attribute dependencies expressed by (logical) formulas that can be supplied by an expert/user. The primary interpretation of the dependencies is to express a kind of relative importance of attributes. We introduce the notion of a formal concept compatible with the attribute dependencies. The main gain of considering only compatible formal concepts and disregarding formal concepts which are not compatible is the reduction of the number of resulting formal concepts. This leads to a more comprehensible structure of formal concepts (clusters) extracted from

B. Ganter and R. Godin (Eds.): ICFCA 2005, LNAI 3403, pp. 176–191, 2005.

the input data. We present basic theoretical results, an algorithm for generating compatible formal concepts, and illustrate our approach by examples.

2 Constraints by Attribute Dependencies

2.1 Motivation

When people categorize objects by means of their attributes, they naturally take into account the importance of attributes. Usually, attributes which are less important are not used to form large categories (clusters, concepts). Rather, less important attributes are used to make a finer categorization within a larger category. For instance, consider a collection of certain products offered on a market, e.g. home appliances. When categorizing home appliances, one may consider several attributes like price, the purpose of the appliance, the intended placement of the appliance (kitchen appliance, bathroom appliance, office appliance, etc.), power consumption, color, etc. Intuitively, when forming appliance categories, one picks the most important attributes and forms the general categories like "kitchen appliances", "office appliances", etc. Then, one may use the less important attributes (like "price $\leq$ \$10", "price between \$15–\$40", "price $>$ \$100", etc.) and form categories like "kitchen appliance with price between \$15–\$40". Within this category, one may further form finer categories distinguished by color. This pattern of forming categories follows the rule that when an attribute y is to belong to a category, the category must contain an attribute which determines a more important characteristic of the attribute (like "kitchen appliance" determines the intended placement of the appliance). This must be true for all the characteristics that are more important than y. In this sense, the category "red appliance" is not well-formed since color is considered less important than price and the category "red appliance" does not contain any information about the price. Which attributes and characteristics are considered more important depends on the particular purpose of categorization. In the above example, it may well be the case that price be considered more important than the intended placement. Therefore, the information about the relative importance of the attributes is to be supplied by an expert (the person who determines the purpose of the categorization). Once the information has been supplied, it serves as a constraint for the formation of categories. In what follows, we propose a formal approach to the treatment of the above-described constraints to formation of categories.

2.2 Constraints by Attribute-Dependency Formulas

Consider a formal context $\langle X, Y, I\rangle$. We consider constraints expressed by formulas of the form

$$y \sqsubseteq y_1 \sqcup \cdots \sqcup y_n. \quad (1)$$

Formulas (1) will be called AD-formulas (attribute-dependency formulas). The set of all AD-formulas will be denoted by ADF. Let now $\mathcal{C} \subseteq ADF$ be a set of AD-formulas.