Pattern Structures and Their Projections

Bernhard Ganter[1] and Sergei O. Kuznetsov[2]

[1] Institut für Algebra, TU Dresden
D-01062 Dresden, Germany
ganter@math.tu-dresden.de

[2] All-Russia Institute for Scientific and Technical Information (VINITI)
Usievicha 20, 125219 Moscow, Russia
serge@viniti.ru

Abstract. Pattern structures consist of objects with descriptions (called patterns) that allow a semilattice operation on them. Pattern structures arise naturally from ordered data, e.g., from labeled graphs ordered by graph morphisms. It is shown that pattern structures can be reduced to formal contexts, however sometimes processing the former is often more efficient and obvious than processing the latter. Concepts, implications, plausible hypotheses, and classifications are defined for data given by pattern structures. Since computation in pattern structures may be intractable, approximations of patterns by means of projections are introduced. It is shown how concepts, implications, hypotheses, and classifications in projected pattern structures are related to those in original ones.

Introduction

Our investigation is motivated by a basic problem in pharmaceutical research. Suppose we are interested which chemical substances cause a certain effect, and which do not. A simple assumption would be that the effect is triggered by the presence of certain molecular substructures, and that the non-occurence of the effect may also depend on such substructures.

Suppose we have a number of observed cases, some in which the effect does occur and some where it does not; we then would like to form hypotheses on which substructures are responsible for the observed results. This seems to be a simple task, but if we allow for combinations of substructures, then this requires an effective strategy.

Molecular graphs are only one example where such an approach is natural. Another, perhaps even more promising domain is that of *Conceptual Graphs* (CGs) in the sense of Sowa [21] and hence, of logical formulas. CGs can be used to represent knowledge in a form that is close to language. It is therefore of interest to study how hypotheses can be derived from Conceptual Graphs.

A strategy of hypothesis formation has been developed under the name of JSM-method by V. Finn [8] and his co-workers. Recently, the present authors have demonstrated [11] that the approach can neatly be formulated in the language of another method of data analysis: Formal Concept Analysis (FCA) [12].

H. Delugach and G. Stumme (Eds.): ICCS 2001, LNAI 2120, pp. 129–142, 2001.

The theoretical framework provided by FCA does not always suggest the most efficient implementation right away, and there are situations where one would choose other data representation forms. In this paper we show that this can be done in full compliance with FCA theory.

1 Formal Contexts

From every binary relation, a complete lattice can be constructed, using a simple and useful construction. This has been observed by Birkhoff [3] in the 1930s, and is the basis of Formal Concept Analysis, with many applications to data analysis.

The construction can be described as follows: Start with an arbitrary relation between two sets G and M, i.e., let $I \subseteq G \times M$, and define

$$A' := \{m \in M \mid (g,m) \in I \text{ for all } g \in A\} \qquad \text{for } A \subseteq G,$$

$$B' := \{g \in G \mid (g,m) \in I \text{ for all } m \in B\} \qquad \text{for } B \subseteq M.$$

Then the pairs (A, B) satisfying

$$A \subseteq G, B \subseteq M, A' = B, A = B'$$

are called the **formal concepts** of the **formal context** (G, M, I). When ordered by

$$(A_1, B_1) \leq (A_2, B_2) :\iff A_1 \subseteq A_2 \quad (\iff B_2 \subseteq B_1),$$

they form a complete lattice, called the **concept lattice** of (G, M, I).

The name "Formal Concept" reflects the standard interpretation, where the elements of G are viewed as "objects", those of M as "attributes", and where $(g, m) \in I$ encodes that object g has attribute m. It has been demonstrated that the concept lattice indeed gives useful insight in the conceptual structure of such data (see [12] and references there).

That data are given in form of a formal context is a particularly simple case. If other kind of data is to be treated, the usual approach is first to bring it in this standard form by a process called "scaling". Recently, another suggestion was discussed by several authors [14], [15] [16] [17]: to generalize the abovementioned lattice construction to contexts with an additional order structure on G and/or M. This seems quite natural, since the mappings $A \mapsto A'$, $B \mapsto B'$ used in the construction above form a Galois connection between the power sets of G and M. It is well known that a complete lattice can be derived more generally from any Galois connection between two complete lattices.

On the other hand, one may argue that there is no need for such a generalization and that no proper generalization will be achieved, since the basic construction already is as general as possible: it can be shown that every complete lattice is isomorphic to some concept lattice.

Nevertheless, such a more general approach may be worthwhile for reasons of efficiency, and it seems natural as well. Several authors [2], [4], [7] have considered the case where instead of having attributes the objects satisfy certain