A Proposal for Combining Formal Concept Analysis and Description Logics for Mining Relational Data

Mohamed Hacene Rouane[1], Marianne Huchard[2], Amedeo Napoli[3], and Petko Valtchev[1,4]

[1] DIRO, Université de Montréal, C.P. 6128, Succ. CV, Montréal, Canada, H3C 3J7
[2] LIRMM, 161, rue Ada, F-34392 Montpellier Cedex 5
[3] LORIA, B.P. 239, F-54506 Vandœuvre-lès-Nancy
[4] Dépt. informatique, UQAM, C.P. 8888, Succ. CV, Montréal, Canada H3C 3P8

Abstract. Recent advances in data and knowledge engineering have emphasized the need for formal concept analysis (FCA) tools taking into account structured data. There are a few adaptations of the classical FCA methodology for handling contexts holding on complex data formats, e.g. graph-based or relational data. In this paper, relational concept analysis (RCA) is proposed, as an adaptation of FCA for analyzing objects described both by binary and relational attributes. The RCA process takes as input a collection of contexts and of inter-context relations, and yields a set of lattices, one per context, whose concepts are linked by relations. Moreover, a way of representing the concepts and relations extracted with RCA is proposed in the framework of a description logic. The RCA process has been implemented within the GALICIA platform, offering new and efficient tools for knowledge and software engineering.

1 Introduction

Formal concept analysis (FCA) has been successfully applied to a range of knowledge engineering problems [22,24]. Nevertheless, FCA methods and tools aimed at directly processing data for producing knowledge units represented within a knowledge representation language –based on description logics (DL) [1] such as OWL DL– are still under study. One key difficulty lies in the presence and management of relational attributes or links in the data, such as "spouse", "reference", and "part-of". For example, a target group for a marketing campaign may be the class of "spouses of Master Gold credit card holders", that involves both binary and relational attributes.

Current FCA methods and tools have no capabilities for taking into account relational attributes. This is a rather hard problem to solve, since relational attributes introduce dependencies and even cycles between the data items. A standard way for producing DL-like concept descriptions from a formal context including binary and relational attributes remains to be designed. Accordingly, one of the objectives of this paper is to present a methodology for taking into

S.O. Kuznetsov and S. Schmidt (Eds.): ICFCA 2007, LNAI 4390, pp. 51–65, 2007.
© Springer-Verlag Berlin Heidelberg 2007

account relational attributes within FCA, leading to what could be called "relational concept analysis" or RCA.

The introduction of relational information, e.g. relational attributes, in the data formats for FCA has been studied for almost a decade now, leading to three main categories of research lines: (i) the relational attributes remain within the formal objects [10,11,12], (ii) relational attributes are considered as first-class citizens and organized into an independent lattice, separated from the standard concept lattice [14] (just like relation types are represented within the conceptual graph formalism [18]), (iii) relations between concepts are established independently from concept construction, on a manual or semi-automated basis [15]. Although these three approaches successfully deal with relational attributes for solving a specific task, they are still not general enough and do not allow to combine and process binary and relational attributes as object descriptors at the concept formation step. Such a need arises in various practical situations, for example in model engineering for software development or in ontology learning from data.

A first introduction of relational concept analysis (RCA) has been proposed in [9]. The data structure on which is based the relational concept analysis process is called a "relational context family" (RCF): it is composed of a collection of contexts and inter-context relations, the latter being binary relations between pairs of object sets lying in two different contexts. The objective is to build a set of lattices whose concepts are related by relational attributes, similar to DL roles or to UML associations. In addition, there are needs for associating restrictions with relational attributes for describing specific characteristics. RCA has been initially motivated by an application on the engineering of UML static models (see [5]) with an emphasis on expressiveness and algorithmic aspects. Meanwhile, the need for processing complex data such as relational data has become an important problem, especially in the field of knowledge discovery in databases [7], and calls for a formalization of RCA.

In this paper, we propose a global and declarative description of the relational structure within the RCA approach, based on a set of lattices resulting from the processing of the contexts that are successively considered. One feature of the relational structure is that an object lying in the extent of an RCA concept can be connected with another object lying in the extent of another RCA concept, through a set of relational attributes or links. The inter-concept links can be nested leading to a relational structure of an arbitrary depth. An auxiliary graph structure is defined for covering these inter-object links.

Moreover, as experiences with UML model analysis reveal, the complexity of the final concept descriptions calls for a knowledge representation formalism, for managing and taking into account the semantics of the inter-concept links, e.g. classifying links and checking their consistency. In the second part of the paper, it is shown how concepts and relations from RCA can be mapped into a knowledge base (KB) represented within a DL of the $\mathcal{ALE}$ family (more precisely $\mathcal{FL}^{-}\mathcal{E}$). The connection between the structure of the original data mapped into