Relational Data Exploration by Relational Concept Analysis

Xavier Dolques, Marianne Huchard, Florence Le Ber, Clémentine Nebut

Relational Data Exploration by Relational Concept Analysis*

Xavier Dolques[1], Marianne Huchard[2], Florence Le Ber[3], and Clémentine Nebut[2]

[1] INRIA, Centre Inria Rennes - Bretagne Atlantique, Campus universitaire de Beaulieu, 35042 Rennes, France, xavier.dolques@inria.fr
[2] LIRMM, Université de Montpellier 2 et CNRS, Montpellier, France, first.last@lirmm.fr
[3] LHYGES, Université de Strasbourg/ENGEES, CNRS, Strasbourg, France florence.leber@engees.unistra.fr

Abstract. Relational Concept Analysis [4] is an extension to FCA considering several contexts with relations between them. Often used to extend the knowledge that can be learned with FCA, RCA also meets the issue of combinatorial explosion. The initial specification of RCA implies a monotonic growth of the number of concepts and an exhaustiveness of all the concepts that can be obtained when a fixed point is reached. In this position paper we propose a different specification of RCA that permits an interactive exploration of the data by letting the choice of the user for each step. This change will permit to handle richer relational data in a more flexible way by restraining the relations explored at each step hence reducing the number of created concepts.

1 Introduction

Relational Concept Analysis (RCA) [4] is based on iterative use of the classical Formal Concept Analysis algorithm to handle relational data: formal objects are described with formal attributes, and with their relationships with formal objects. Because RCA groups formal objects using relationships to formal objects at any distance, it often comes with a combinatorial explosion, and patterns of interest are difficult to extract from the huge set of built concepts. Various strategies can be used to cope with this complexity, including separating the initial formal object sets into smallest ones after a first analysis, or introducing queries [1]. Here we focus on the use of RCA to interactively explore data by letting the user choosing at each step of the iteration of FCA which contexts (formal and relational) he or she would like to use.

The context of this research is the FRESQUEAU project[4] which aims at developing new methods for studying, comparing and exploiting all the parameters available concerning streams and water areas. In this project, different

* This work was partly funded by french contract ANR11_MONU14.
[4] http://engees-fresqueau.unistra.fr/

approaches of knowledge discovery (including FCA) are tested and combined in order to better assess the ecological functioning of such hydrosystems.

In this paper we first outline the RCA process to highlight potential variation points that would promote exploration. Then we conclude with a short discussion.

2 The RCA algorithm

Algorithm 1 outlines the main steps followed by RCA to build groups of objects by considering attributes and object-object relations [4]. The input of RCA is a Relational Context Family $RCF = (K, R)$ composed of n object-attribute contexts $\mathcal{K}_i = (O_i, A_i, I_i)$, i in $1..n$, and m object-object contexts $\mathcal{R}_j$, j in $1..m$.

1: `proc` MULTI-FCA(`In:` $(\mathbf{K}, \mathbf{R})$ `a RCF,`
2: `Out:` $\mathbf{L}$ `array` $[1..n]$ `of lattices)`
3: $p \leftarrow 0$; `halt` $\leftarrow$ `false`
4: `for` i `from` 1 `to` n `do`
5:   $\mathbf{L}^0[i] \leftarrow$ BUILD-LATTICE($\mathcal{K}_i^0$)
6: `while not halt do`
7:   $p{+}{+}$
8:   `for` i `from` 1 `to` n `do`
9:     $\mathcal{K}_i^p \leftarrow$ EXTEND-REL($\mathcal{K}_i^{p-1}$, $\mathbf{L}^{p-1}$)
10:     $\mathbf{L}^p[i] \leftarrow$ UPDATE-LATTICE($\mathcal{K}_i^p$, $\mathbf{L}^{p-1}[i]$)
11:   `halt` $\leftarrow \bigwedge_{i=1,n}$ ISOMORPHIC($\mathbf{L}^p[i], \mathbf{L}^{p-1}[i]$)

Algorithm 1: The RCA process

For $\mathcal{R}_j \subseteq O_i \times O_j$, we call O_i the domain and O_j the range. The **initialization step** (Lines 4-5) consists in building, for all i in $1..n$, the lattice $\mathbf{L}^0[i]$ associated with the context $\mathcal{K}_i$.

At step p:

- **EXTEND-REL** appends to $\mathcal{K}_i$ the relations obtained by scaling object-object relations for which $\mathcal{K}_i$ is the domain. The scaling consists in including the object-object relations as *relational* attributes. They are obtained using the concepts of the lattices of step $p-1$ and a scaling operator (*i.e.* $\exists$, $\forall$). For example, if the scaling operator $\exists$ is chosen for scaling a given relation $\mathcal{R}_j$, $\mathcal{R}_j$ columns are replaced by attributes of the form $\exists \mathcal{R}_j : C$, where C is a concept in the lattice built upon objects of the range of $\mathcal{R}_j$ at step $p-1$. An object o of the domain of $\mathcal{R}_j$ **owns** $\exists \mathcal{R}_j : C$ if $\mathcal{R}_j(o) \cap Extent(C) \neq \emptyset$.
- **UPDATE–LATTICE** updates the lattices of step $p-1$ in order to produce, for i in $1..n$, the lattice $\mathbf{L}^p[i]$, associated with $\mathcal{K}_i$ concatenated to all scaled object-object contexts with $\mathcal{K}_i$ domain.

The algorithm stops when a fix-point is obtained: a lattice family isomorphic to the lattice family obtained at the previous step is obtained and leaves unchanged concept extents.

The advantage of such a process is that the obtained concepts have in their intent relations to other concepts in addition to classical attributes. Those relations permit the extraction of patterns built from several interconnected concepts as shown in [2] and [3] that could not be easily obtained with the classical process of Formal Concept Analysis.

However, one problem of such a process is the potential difficulty to apprehend the result. In past work in the domain of Model Driven Engineering, data extracted from models of medium size have been easily handled by RCA. Nevertheless in a context of data mining the data are of a different scale. Especially when only small patterns are needed while many relations connect the objects and these relations form a cyclic entity-relationship diagram, the result will appear hard to understand by a human due to the number of concepts to consider simultaneously and the computation time will be considered as a handicap. In such cases, we think it will be more practical to have a kind-of exploratory approach.

Table 1 shows main possible variations on the algorithm to go towards an exploratory approach. We have enumerated the variation points of the algorithm that could affect the result by changing the contexts considered at each step. We have proposed for each variation point an alternative scenario from the process previously described that involves the user by asking him or her to perform selections. All those variations or only a subset of them can be applied depending on the granularity needed.

Table 1. Variations for the exploratory approach

initialization step, L4-5	Build lattices for selected object-attribute contexts concatenated to selected object-object contexts.
Extend-Rel, L9	Rather than using all relations and scaling all object-object relations, select a subset of the RCF and scaling operators for each selected object-object context. Note: lattices for ranges of the selected object-object relations should have been calculated in a previous step (not necessarily $p-1$). At this step, object-attribute contexts can also be selected and the corresponding lattice can be built.
Update–Lattice, L10	Only the lattices for the selected relations are updated.
halt, L11	The decision is left to the expert when to stop (or the fix-point is obtained)

3 Conclusion and discussion

In this position paper, we have outlined an exploratory approach for assisting the use of Relational Concept Analysis in a way that would better fit a data mining process. We have several motivations for disturbing the original RCA process: to go faster to a relevant result by calculating less lattices (preferably lattices of interest), to cope with the inherent complexity of mining relational data, or to let the expert guiding the discovery process based on his/her intuition and the knowledge patterns that appear on-the-fly. In our current approach, the data are given by experts, so we don't use exploration in the sense of [5], unless the data exploration.

Many questions are raised by this way of extracting concepts from relational data. Initialization of the process has an impact for the later discovered structures. It can accelerate the process, if the selected object-object relations contain the main information for the expert, or reversely, it can discard the expert from the relevant information. Nevertheless, the most serious problem comes from the fact that going step-by-step leads to a non-monotonic concept construction and one could build several cases where the process diverges (iterates between recurrent configurations). In the original RCA process, when the fix-point is attained, lattices of the two last steps are isomorphic, thus when a concept references another through a relational attribute, the latter can be found in the same step appropriate lattice. But in the exploratory process we propose, when a concept references another through a relational attribute, the latter is in a lattice of a previous step and may itself reference a concept in a previous step, etc. We should find solutions for presenting the expert information easy to interpret these situations. Nevertheless, we think that such an exploratory approach should be more practical than the "brute force" that iterates until the fix-point and gives results that an expert will hardly understand.

References

1. Azmeh, Z., Huchard, M., Napoli, A., Hacene, M.R., Valtchev, P.: Querying relational concept lattices. In: Proc. of the 8th Intl. Conf. on Concept Lattices and their Applications (CLA'11). pp. 377–392 (2011)
2. Dolques, X., Huchard, M., Nebut, C.: From transformation traces to transformation rules: Assisting model driven engineering approach with formal concept analysis. In: Supplementary Proceedings of ICCS'09. pp. 15–29 (2009)
3. Dolques, X., Huchard, M., Nebut, C., Reitz, P.: Fixing generalization defects in UML use case diagrams. In: CLA'10: 7th International Conference on Concept Lattices and Their Applications. pp. 247–258 (2010)
4. Huchard, M., Hacène, M.R., Roume, C., Valtchev, P.: Relational concept discovery in structured datasets. Ann. Math. Artif. Intell. 49(1-4), 39–76 (2007)
5. Rudolph, S.: Relational exploration: combining description logics and formal concept analysis for knowledge specification. Ph.D. thesis, Dresden University of Technology 2006 (2006)