Thesis proposal

Comparing and combining approaches ein formal concept analysis fot relational data

Laboratory: ICube, SDC team, Strasbourg, and IRISA, LACODAM team, Rennes

Thesis advisors: Florence Le Ber (ICube), Sébastien Ferré (IRISA)

Supervisors: Xavier Dolques (ICube), Peggy Cellier (IRISA)

To apply: send CV, transcript of grades, ranking and motivation to florence.leber@icube.unistra.fr Sebastien.Ferre@irisa.fr

Funding: ANR SmartFCA

Context: Many data to be analyzed are relational: spatial data, temporal data, or data describing links between individuals. Few methods are adapted to this type of data, which require specific approaches, including aggregation techniques. Among these approaches, Relational Concept Analysis is based on Formal Concept Analysis (FCA) [1], a mathematical classification method, widely applied on different types of data and in many domains (for example [2,3]). Starting from a table (called context) describing some objects by a set of attributes, it consists in building a concept lattice, i.e. some couples (extent; intent) of closed sets describing the objects and the attributes that define them. Relational concept analysis (RCA) [4] considers two types of contexts, namely object-attribute contexts and object-object contexts describing the relations between objects. RCA extends object-attribute contexts with relational attributes of the form qrC, where q is a quantifier, r is a relation and C is a concept from the codomain of r. The result of RCA is a family of lattices (one per object-attribute context) interlinked by these relational attributes.

In Graph conceptual analysis (Graph-FCA) [5] data are modeled as a multi-hypergraph with objects as nodes, and attributes as hyperedge labels. A unary hyper-arc a(o) corresponds to the description of an object by an attribute, as in FCA. A binary hyper-arc a(o1,o2) corresponds to a relation 'a' from o1 to o2, like relational attributes in RCA. N-ary relations are represented by n-ary hyper-arcs a(o1,...,oN). A graph concept represents a set of object tuples (extension) that can be seen as the exhaustive answers to a conjunctive query (intension), e.g. (x,y) ← a1(x,z), a2(y,z), and where this query expresses everything that these tuples have in common

This thesis part of the ANR SmartFCA, which gathers 5 French teams working in the field of FCA and whose aim is to provide a platform gathering the different variants of this method. Several engineers will be assigned to the development of this platform.

Objectifs de la thèse : The aim of this thesis is *i)* to compare, both theoretically and experimentally the two approaches, RCA and Graph-FCA; *ii)* to propose elements to make the two approaches cooperate ; *iii)* to define a methodological guide for users (data modeling, parameter values, selection of algorithms, etc.). The results, algorithms and methodological guide, will be included into the platform devloped during the ANR SmartFCA project.

Links between the two approaches have already been discussed [6,7,8] and the thesis should deepen these preliminary works. The first step will be to study and compare the two approaches, using existing tools, by testing them on relational datasets provided by the partners of the project. In particular, we will be interested in proposing a declarative model of RCA which is currently defined in an iterative way. We will also be interested in the cooperation between RCA and Graph-FCA by defining the data structures allowing them to be interoperable.

The explosive nature of FCA-based approaches requires the use of algorithms that compute only part of the concepts or lattices: AOC-poset [9], exploratory approaches, neighbouring computation, estimation of results

from parameter choices [10,11] ... These variants will also be studied in order to define a methodological framework for using both RCA and Graph-FCA, including these different options as well as elements to guide their use. This work will be donc in cooperation with an engineer in charge of the platform devlopments.

Expected contributions:

- Theoretical advances on FCA methods
- Methodological Development s
- Experiments on real data

Applicant profile:

- Master 2 in Computer Science or equivalent
- Skills in logic, knowledge representation and programming
- Curiosity, ability to understand different domains and to interact with the experts of these domains

References:

[1] Ganter, B., Wille, R. Formal concept analysis - mathematical foundations. Springer (1999)

[2] Priss, U. Formal concept analysis in information science. ARIST 40(1), 521–543 (2006)

[3] Alam, M., Coulet, A., Napoli, A., Smaïl-Tabbone, M. Formal Concept Analysis Applied to Transcriptomic Data. CLA 2012, Málaga, Spain

[4] Hacene, M.R., Huchard, M., Napoli, A., Valtchev, P. Relational concept analysis: mining concept lattices from multi-relational data. Ann. Math. Artif. Intell. 67(1), 81–108 (2013)

[5] S. Ferré and P. Cellier. Graph-FCA: An extension of formal concept analysis to knowledge graphs. Discrete and Applied Mathematics, 273:81–102 (2020).

[6] C. Nica, A. Braud, and F. Le Ber. RCA-Seq: an Original Approach for Enhancing the Analysis of Sequential Data Based on Hierarchies of Multilevel Closed Partially-Ordered Patterns. Discrete Applied Mathematics, 273:232–251, 2020.

[7] S Ferré, P Cellier. How Hierarchies of Concept Graphs Can Facilitate the Interpretation of RCA Lattices? CLA 2018.

[8] P. Keip, S. Ferré, A. Gutierrez, M. Huchard, P. Silvie, P. Martin. Practical Comparison of FCA Extensions to Model Indeterminate Value of Ternary Data. CLA 2020, 197-208

[9] X. Dolques, F. Le Ber, M. Huchard , C. Grac. Performance-friendly rule extraction in large water data-sets with AOC posets and relational concept analysis, International Journal of General Systems, Taylor & Francis (2016)

[10] Braud, A., Dolques, X., Huchard, M., Le Ber, F. Generalization effect of quantifiers in a classification based on relational concept analysis. Knowledge-Based Systems 160, 119–135 (2018)

[11] A. Ouzerdine , A. Braud, X. Dolques, M. Huchard , F. Le Ber. Adjusting the exploration flow in Relational Concept Analysis -- An experience on a watercourse quality dataset, Advances in Knowledge Discovery and Management, Springer (2022).