Thesis proposal

Relational data analysis on a DBMS – agregation based on relational algebra

Laboratory : Laboratoire ICube, UMR CNRS – Université de Strasbourg – INSA – ENGEES, SDC team (https://sdc.icube.unistra.fr/en/index.php/Home)

Thesis advisor : Florence Le Ber (PhD, Accreditation to Supervise Research, ICube)

Supervisors: Agnès Braud (Assistant professor, ICube), Xavier Dolques (Assistant professor, ICube)

To apply: send CV, transcript of grades, ranking and motivation to florence.leber@engees.unistra.fr, agnes.braud@unistra.fr, dolques@unistra.fr

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.

Thesis objectives: FCA/RCA offer tools which have proven to be useful in different projects, and produce results easily understandable by experts trying to analyse data in their domain. Improvements are however expected both to deal with larger volumes of data [5] and to simplify the process. This thesis will focus on spatiotemporal databases in order to propose a full process for analysing such data. A particular focus will be placed on the quantifiers used in RCA, in order to propose specific quantifiers for spatial and temporal data. For that purpose, we will rely on existing quantifiers (existential, universal or counting [6]) and on qualitative models of space and time [7].

Usually we define a relational schema for RCA by selecting tables (extracted from the database) and by computing the relations between the objects, which is a preprocessing step for the RCA process. For this thesis, we will define a more flexible relational schema, based on instantiable constraints, in order to enable the modification of the dataset considered without having to restart the process from the preprocessing step.

Let us consider the example of a geographical database, in which we want to study the relations between crop plots having certain spatial relationships (*connected, near*, etc.). Currently we extract the tables object-attribute

(plot, crop) and the computed object-object contexts (for example crops are near if their distance is lower than 500 m). We can then create relational attributes exists-near.C by using the existential quantifier. We would like to be able to use a specific quantifier exists-near that can be applied on a more general object-object context including the numeric distances between plots. This quantifier has to be instantiable according to different schemas, in which the threshold distance could vary. Using relational algebras also allows to consider quantifiers at different levels of generality, e.g. it-exists-part-of can be specified as it-exists-part-of-tangential.

This type of quantifier could also be instantiable differently if we consider the objects of a concept or the concept as a whole, enabling it to render the specific relations between groups (as aggregation criteria do in hierarchical classification).

Finally, the work should lead to the implementation of a RCA process that can be applied directly on a database, enabling it to handle more data in a more user-friendly environment for an analyst. Another expected result is the theoretical contribution on spatial and temporal quantifiers and on the notion of instantiable relational schema.

Expected contributions :

- Theoretical advances on quantifiers and the relational schema
- Development of a tool implementing a full analysis framework
- Experiments and validation on real data (geographical and temporal databases)

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

As required by the Doctoral School of the University of Strasbourg, the candidate must have obtained all his/her Master's (and Bachelor's) semesters or equivalent with a grade above 12/20. He/she must also be ranked among the top 20% of graduates of his/her Master promotion.

Références:

[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, Oct 2012, Fuengirola (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] Braud, A., Dolques, X., Gutierrez, A., Huchard, M., Keip, P., Le Ber, F., Martin, P., Nica, C., Silvie, P. Dealing with large volumes of complex relational data using RCA. In CDA_FCA, Rokia Missaoui, Léonard Kwuida, Talel Abdessalem (Eds.), Springer (2022)

[6] 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)

[7] J-F. Condotta , F. Le Ber, G. Ligozat , L. Travé-Massuyès. Qualitative reasoning, A Guided Tour of Artificial Intelligence Research, Henri Prade, Odile Papini, Pierre Marquis (Eds.), Springer, Volume 1, 2020.