Numerical rule mining for city housing planning

1 Internship environment

1.1 Research team and supervisors

Luis Galárraga, Christine Largouet, Louis Bonneau de Beaufort, Irisa LACODAM team, Rennes Beaulieu.
https://team.inria.fr/lacodam/

1.2 Framework

Multi-year collaboration between Irisa lab and Rennes Métropole, with a broad goal : construction of a systemic model of the evolution of the city.

1.3 Numerical rule mining

The internship will focus on an important issue for the metropolis: eliciting the relationship (over time) between housing and population. The prospects of this work are to propose, from a set of valid rules, a predictive model to help decision-making in urban planning.

The goal is to learn accurate association rules from a database of housing information consisting of attributes of the houses such as number of flats, size, type, and ownership, as well as data about the inhabitants (e.g., number, age).

2 Planned activities

2.1 Choosing attributes of interest

In advance of the internship, urban planing experts will propose rules that will help guide the mining task, and thus reduce the exploratory dimension of the problem.

If needed, new attributes of interest can still be added, with the support of the Rennes Metropole technical services. Examples of attributes of interest are those derived from the spatial dimensions of the data.

2.2 Learning rules

The problem of rule-mining is not new [1], however our needs require to pay attention to the following requirements:

- The rules must be human-readable [2]. A research track might be the construction of regressions trees providing the user with hybrid rules [3], for example:

  `(numberHomesT1 == X)` $\wedge$ `(numberHomesT4 == Y) => numberStudents == 0.9 * X + 3 * Y`

- The rules should consider the temporal dimension, by focusing, for example, on attributes' annual variations Δ:

  `(`Δ`HomesT1 == X)` $\wedge$ `(distMetro == "low") =>` Δ`Students == 0.9 * X`

- Rules quality will have to be evaluated. Spatial and temporal validity will have to be asserted. Furthermore, a mean of comparison between learned rules and expert-provided rules is desirable.

References

[1] R. Agrawal, T. Imielinski and A. Swami. *Mining association rules between sets of items in large databases*. In SIGMOD, 1993.

[2] L. Galárraga, F. M. Suchanek. *Towards a Numerical Rule Mining Language*. Automated Knowledge Base Construction Workshop, 2014.

[3] O. Pelgrin. *Numerical rule mining for prediction of wheat and vine diseases*. Master research internship report, Université Rennes 1, 2018.