

Numerical Rule Mining for Smart City

1 Environment

The candidate will work in the joined Inria / IRISA research centre located in Rennes. Inria (www.inria.fr) and IRISA (<http://www.irisa.fr/>) are amongst the leading research centres in Computer Sciences in France. The work is supervised by members of the Lacodam Team recognized in the data mining field. This subject will be done in close collaboration with the city of Rennes Métropole, one of the first smart city in France.

2 Supervisors

Luis Galárraga *luis.galarraga-del-prado@inria.fr*

Christine Largouët *christine.largouet@irisa.fr*

Louis Bonneau de Beaufort. *louis.bonneauagrocampus-ouest.fr*

3 Keywords

- artificial intelligence
- data mining
- rule learning
- decision making
- smart city

4 Subject

The collaboration between data-mining and smart city will bring benefits to all stakeholders. Smart cities have to deal with massive data and require efficient computer-based solutions. This work focuses on the area of smart urban planning and more particularly on the relationship between demography and urbanism. The final objective of the internship is to learn accurate association rules from housing data and demography data in order to be able to answer questions such as "How many small flats are required for this neighborhood? " or "Do we have to build a new school within the next two years in this neighborhood? "

Association rule mining is one of the key issues in knowledge discovery for making smart decision. The goal is to learn accurate association rules between attributes such as number of flats, size, type, and ownership, age or position of the inhabitants,... 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]. Dealing with a huge number of rules is the main interpretation problem in decision making. How much a rule is trustworthy is also a matter of discussion. A research track might be the construction of regressions trees providing the user with hybrid rules [3], for example:
$$(\text{numberHomesT1} == X) \wedge (\text{numberHomesT4} == Y) \Rightarrow \text{numberStudents} == 0.9 * X + 3 * Y$$
- The rules should consider the temporal dimension, by focusing, for example, on attributes annual variations Δ :
$$(\Delta\text{HomesT1} == X) \wedge (\text{distMetro} == \text{"low"}) \Rightarrow \Delta\text{Students} == 0.9 * X$$
- Rules quality will have to be evaluated. Spatial and temporal validity will have to be asserted. Furthermore, a way to compare the learned rules with the expert-provided rules should be studied.

To summarize, the internship involves:

- Choosing attributes of interest with the help of urban planning experts of the city of Rennes.
- Developing an experimental framework to learn the more accurate rules from the city data.
- Designing experimental protocols and scenarios to evaluate the learned rules
- Participating in the analysis of the results with experts

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.