MSc Internship

Quantitative Partial Order Representations of Preference Data

Antoine Amarilli, Pierre Senellart

Télécom ParisTech

Topic description

Ordered lists of data items are sometimes sorted by a well-defined criterion, such as time, price, name, etc. However, in other cases, the ranking function used is *unknown*. Examples include search engine result pages, lists of post sorted by a proprietary relevance function, and lists of items provided by users to indicate their preferences [3].

A common task is to *integrate* such ordered lists. For instance, we may want to aggregate results from multiple search engines, which is known as *rank aggregation* in the literature [2]. We may want to integrate several views on ordered data from the same source, e.g., to evaluate complex queries not supported by the Web interface that gives access to the results. Last, we may want to integrate conflicting preferences from multiple users to reach a consensus, which relates to the general study of *voting systems*.

In collaboration with M. Lamine Ba (Qatar Computing Research Institute) and Daniel Deutch (Tel Aviv University), we have studied how to represent the uncertain results of combining ordered lists with standard relational algebra operations, using a partial-order-based model [1]. We study in particular the complexity of determining possibility and certainty problems on such representations, and that of evaluating general aggregation queries over them. Our principled approach, however, has some limitations which would make it harder to apply to some use cases.

First, we study a *bag* semantics, i.e., we always retain the multiple occurrences of the same value (e.g., the same Web page in multiple result lists, the same hotel in property lists). This is not suitable for scenarios where we to keep only one single copy of each item, whose order in the integration result should depend on that of all its occurrences in the input data. We have preliminary results about how to lift this limitation, but we do know yet how the result of this "duplicate consolidation" operation can be represented in full generality.

Second, our approach represents *uncertain* orders, namely, it works with partial orders, but without any quantitative assessment of which order is more likely. In practice, however, we may want to know which orders are more probable, and answer queries such as determining the *average* rank of a result among all possible orders, weighted by their probabilities. For now, our model only supports query evaluation in a compositional, world-by-world manner, and does not support integrating information across the possible worlds.

The goal of the internship, at a theoretical level, is to study ways to represent uncertain order on data items with quantitative information, i.e., information about which total orders are more likely. One possibility would be to extend partial orders with *quantitative information*, e.g., probabilistic comparability relations between elements; and study how such representations can be integrated.

A second objective of the internship is to perform practical experiments using the model, to validate its usefulness. For instance, the approach could be evaluated on the CrowdRank dataset [3].

Supervision and Environment

This Master's internship will have a duration of between 4 and 6 months and will be supervised by Pierre Senellart[1], professor at Télécom ParisTech and senior research fellow at the National University of Singapore and Antoine Amarilli[2], third-year PhD candidate at Télécom ParisTech. The internship will be based at Télécom ParisTech, in Paris.

References

[1] A. Amarilli, M. L. Ba, D. Deutch, and P. Senellart. Representing and querying order-incomplete data. Draft: `http://pierre.senellart.com/publications/amarilli2016representing.pdf`, 2015.

[2] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *Proc. WWW*, 2001. `http://www.shibbo.ethz.ch/CDstore/www10/papers/pdf/p577.pdf`.

[3] J. Stoyanovich, M. Jacob, and X. Gong. Analyzing crowd rankings. In *Proc. WebDB*, 2015. `https://www.cs.drexel.edu/~julia/documents/webdb15.pdf`.

[1] `http://pierre.senellart.com/`
[2] `http://a3nm.net/`