

The Inductive Database framework: a long-term perspective on query languages for data mining.

Jean-François Boulicaut

Institut National des Sciences Appliquées de Lyon
LISI Bâtiment Blaise Pascal
F-69621 Villeurbanne cedex, France
`Jean-Francois.Boulicaut@insa-lyon.fr`

We consider the challenge of supporting KDD processes by means of queries. We introduce a research perspective, the so-called inductive database framework [4, 1, 3, 2]. An inductive database is a database that in addition to data contains intensionnaly defined generalizations about the data. Specifying and implementing inductive databases is a long-term perspective. However, more or less specialized data mining query languages have been proposed, e.g., MSQL [7] or MINE RULE [5]. It is interesting to discuss whether they are good candidates for inductive database querying. Furthermore, the optimization of inductive queries is clearly related to constraint-based mining (see e.g., [6]). Considering the association rule mining task, we can figure out what should be done in a near future in order to implement inductive databases for that simple but useful kind of pattern. However, applying the framework and studying the robustness of the identified concepts for other data mining tasks remain open. Considering these issues is the main goal of the European project cInQ IST-2000-26469 (May 2001-May 2004), funded by the Future and Emerging Technologies arm of the IST Programme.

References

1. J.-F. Boulicaut, M. Klemettinen, and H. Mannila. Modeling KDD Processes within the Inductive Database Framework. In *Proc. DaWak'99*, pages 293 – 202, September 1999. Springer-Verlag LNCS 1676.
2. L. De Raedt. A Logical Database Mining Query Language. In *Proc. ILP 2000*, pages 78 – 92, July 2000, Springer LNCS 1866.
3. F. Giannotti, and G. Manco. Querying Inductive Databases via Logic-Based User-Defined Aggregates. In *Proc. PKDD'99*, pages 125 – 135, September 1999, Springer LNCS 1704.
4. T. Imielinski and H. Mannila. A Database Perspective on Knowledge Discovery. *CACM*, 39(11):58 – 64, November 1996.
5. R. Meo, G. Psaila, and S. Ceri. A new SQL-like Operator for Mining Association Rules. In *Proc. VLDB'96*, pages 122 – 133, September 1996. Morgan Kaufmann.
6. R. Ng, L. Lakshmanan, J. Han, and A. Pang. Exploratory Mining and Pruning Optimizations of Constrained Associations Rules. In *Proc. SIGMOD'98*, pages 13 – 24, 1998. ACM Press.
7. A. Virmani. Second Generation Data Mining: Concepts and Implementation. PhD thesis, Rutgers University (USA), April 1998.