A Holistic Architecture for Knowledge Discovery in Multi-Database Environments

Alex G. Büchner[§], Baizheng Yang[†], Sudha Ram[†], David A. Bell[‡], John G. Hughes[§]

[§] Northern Ireland Knowledge Engineering Laboratory, University of Ulster
[†] Department of Management Information Systems, University of Arizona
[‡] School of Information and Software Engineering, University of Ulster
ag.buchner@ulst.ac.uk, byang@bpaosf.bpa.arizona.edu

Introduction

Knowledge discovery in multi-databases, i.e. distributed, heterogeneous and federated environments is twofold: schema integration using knowledge discovery techniques, and the mining of data from multi-databases themselves. Whereas the first problem is concerned with semantic equivalence among different databases, the latter expands existing KDD techniques and methodologies.

We argue that the appropriate combination of those technologies leads to improved integrated schemata, as well as discovered knowledge of higher quality. The objective of our proposed research is to establish the two described components into one holistic architecture. This novel approach allows better interaction between the knowledge discovery component and the integrated schema, and thus improves the quality of both elements.

A Holistic Architecture

Figure 1 depicts our simplified proposed overall holistic architecture for knowledge discovery in multi-database environments. Its key elements are the Global Knowledge Coordinator (GKC) and the Task-driven Integrated Schema (TIS).

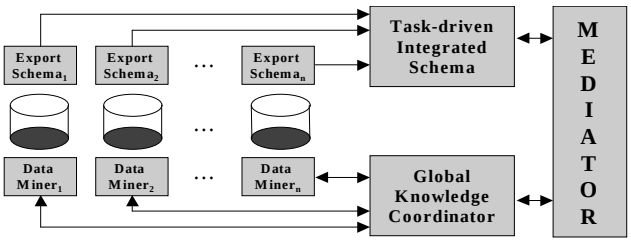

Figure 1: Overall Holistic Architecture

The GKC's rôle is twofold. Firstly, it receives requests from the mediator to discover knowledge including thresholds and posts back discovered rules in an agreed interchange format. Secondly, it initiates local data miners, which have to be able to handle heterogeneity ([Büc96]), to discover knowledge and then combines the results.

The TIS is based on information from the local export schemata and requests from the GKC. The initial TIS is constructed employing traditional schema integration techniques. In addition, the schema integration process can be enhanced by adopting knowledge discovered from local sites, as proposed by [Dao95]. However, the unique characteristic of the schema integration component is that the TIS and the export schemata are constructed with an orientation towards knowledge discovery.

Multi-agents are used as a communication vehicle among different components, which are coordinated by the mediator in order to handle requests, discovered knowledge and schema information. It accepts requests from users, the TIS, as well as the GKC, and responses back to the requesting party. The mediator, which is an extension to [Ram95]'s work, can thus be seen as the supervising agent of the architecture.

Conclusions and Future Work

The proposed architecture provides an appropriate environment for knowledge discovery in multi-databases. Before the implementation of the entire system, various problems have to be tackled. Different integrated schema constructs, such as subclasses, aggregates, or composites (with different cardinalities) require different knowledge combination strategies, which have to be evaluated. Existing knowledge measures, such as support, confidence, or interestingness (based on local and global thresholds), which can lead to refined local knowledge discovery based on a modified TIS, have to be refined ([Rib95]). Schematic information has to be specified appropriately for incorporation as domain knowledge at the local data mining sites. There is a need to explore schema integration improvement methods for utilizing knowledge discovery outputs. Finally, the components which have been modeled as intelligent agents have to accommodate more intentional behavior. These anthropomorphic concepts, such as belief, intention, desire, ability, etc., can then be mapped onto knowledge discovery and schema integration to provide more flexibility in a heterogeneous data habitat.

References

[Büc96] A.G. Büchner, S.S. Anand, D.A. Bell, J.G. Hughes: A Framework for Discovering Knowledge from Distributed and Heterogeneous Databases, in *Proc. IEE Coll. on Knowledge Discovery and Data Mining*, pp. 8/1-8/4, 1996.

[Dao95] S. Dao, D. Perry: Applying a Data Miner to Heterogeneous Schema Integration, in *Proc. of KDD-95*, pp. 63-68, 1995.

[Ram95] S. Ram, V. Ramesh: A Blackboard-Based Cooperative System for Schema Integration, in *IEEE Expert*, 10(3):56-62, 1995.

[Rib95] J.S. Ribeiro, K.A. Kaufman, L. Kerschberg: Knowledge Discovery from Multiple Databases, in *Proc. of KDD-95*, pp. 240-245, 1995.