Third Generation Data Mining: Towards Service-Oriented Knowledge Discovery (SoKD-2010)

Barcelona, 24 September 2010
Workshop held in conjunction with ECML/PKDD-2010
http://cui.unige.ch/~hilario/sokd10/

Background

A major challenge for third generation data mining and knowledge discovery systems is the integration of distributed data/knowledge resources (which are highly diverse in nature in terms of representation and data formats) and computer systems (tools for data integration, data mining and knowledge discovery). First generation data mining systems supported a single algorithm or a small collection of algorithms that are designed to mine attribute-valued data. Today's second generation systems support high performance interfaces to databases and data warehouses, and provide increased scalability and increased functionality; for example, second generation systems can mine larger and more complex data sets and provide increased flexibility by supporting a data mining schema and a data mining query language.

The emerging third generation data mining and knowledge discovery systems should be able to mine distributed and highly heterogeneous data found on intranets/extranets/grid/cloud and integrate efficiently with operational data/knowledge management and data mining systems. The key technologies which will make third generation data mining and knowledge discovery possible should include:

- meta-data (semantic annotations) on different information resources (data, human-coded knowledge, and machine-induced patterns and predictive models), and on the ingredients of data mining and knowledge discovery systems (pattern mining and model discovery tools),
- implementations of data mining and knowledge discovery tools as services available on the web. Such service-oriented knowledge discovery systems will enable meta-level search of data/knowledge resources and mining systems, facilitating the construction of knowledge discovery workflows (representing potentially repeatable sequences of data mining and data integration steps) and resulting in improved pattern and model discovery.

Compared to contemporary search engines which provide a means of locating data on the net, third generation data mining and knowledge discovery systems will provide a means for discovering patterns, associations, changes and anomalies in networked data, where each data source comes with its own structure, semantics, data formats, names, concepts, and access methods. Currently, the burden falls on the user to manually (via programs) convert between the data formats, resolve conflicts, integrate data and interpret results in order to make viable use of this information.

Goals and Target Audience

This workshop intends to gather contributions supporting third generation data mining and knowledge discovery, elaborating a service-oriented approach to information fusion, for the needs of exploratory data analysis in the framework of inductive databases, enriched with ontology information available from the web.

Given the growing amount of information available on the net, this workshop should be of interest to knowledge engineers, as well as students and researchers interested in data mining and advanced methods for knowledge discovery. The workshop will also concern researchers in databases and in software engineering, for whom data mining is an "application area". Finally, it will not fail to attract researchers and practitioners in semantic web technologies, as the ultimate fulfillment of a truly semantic web resides in the possibility of extracting not just readily available information but also deep knowledge in the form of underlying patterns and regularities.

Topics

This workshop intends to gather contributions supporting third generation data mining and knowledge discovery, elaborating a service-oriented approach to information fusion, for the needs of exploratory data analysis in the framework of inductive databases, enriched with ontology information available from the web. The workshop solicits papers on the following topics:

- Theoretical framework for third generation data mining and knowledge discovery
- Inductive databases, constraint-based data mining and inductive queries
- Service-oriented approaches to data mining
- Meta-level annotations and search for data mining services
- Multiple-source learning or learning from heterogeneous data including text & images
- Integrating prior knowledge (probabilities, ontologies) into data mining
- Data mining ontologies, in particular novel ontological representation schemes for handling quantitative data and data streams
- Data mining workflows/scenarios
- Data mining on the grid and cloud computing
- Applications of service-oriented data mining approaches in business, ecological modeling, medicine, health care, e-science, bioinformatics, etc.

Submission

Authors are invited to submit papers related to the topics listed above. Technical papers will be assessed based on relevance, originality, significance, technical soundness and clarity of presentation, by at least 2 Program Committee members.

The maximum length of submissions is 12 pages for technical papers and 6 pages for position papers. All submissions should follow the Springer-Verlag LNCS format. Please submit papers as PDF files to the EasyChair SoKD-10 home page.

Important Dates

Submission deadline: ~~June 21, 2010~~ June 27, 2010
Notification of acceptance: July 12, 2010
Camera-ready papers due: July 21, 2010

Workshop Chairs

Melanie Hilario
Artificial Intelligence Laboratory, University of Geneva

Joost N. Kok
Leiden Institute of Advanced Computer Science, Leiden University

Nada Lavrac
Department of Knowledge Technologies, Jozef Stefan Institute

Vid Podpecan
Department of Knowledge Technologies, Jozef Stefan Institute

Program Committee

Abraham Bernstein, University of Zurich (Switzerlnd)
Michael Berthold, Konstanz University (Germany)
Hendrik Blockeel, Leuven University (Belgium)
Jeroen de Bruin, Leiden University (The Netherlands)
Werner Dubitzky, University of Ulster (UK)
Alexandros Kalousis, University of Geneva (Switzerland)
Igor Mozetic, Jozef Stefan Institute (Slovenia)
Filip Zelezny, Czech Technical University (Czechia)