

Preface

This volume contains papers selected for presentation at PKDD'2000, the Fourth European Conference on Principles and Practice of Knowledge Discovery in Databases. The first meeting was held in Trondheim, Norway, in June 1997, the second in Nantes, France, in September 1998, and the third in Prague, Czech Republic, in September 1999.

PKDD 2000 was organized in Lyon, France, on 13–16 September 2000. The conference was hosted by the Equipe de Recherche en Ingénierie des Connaissances at the Université Lumière Lyon 2. We wish to express our thanks to the sponsors of the Conference, to the University Claude Bernard Lyon 1, the INSA of Lyon, the Conseil général of the Rhône, the Région Rhône Alpes, SPSS France, AFIA, and the University of Lyon 2, for their generous support.

Knowledge discovery in databases (KDD), also known as data mining, provides tools for turning large databases into knowledge that can be used in practice. KDD has been able to grow very rapidly since its emergence a decade ago by drawing its techniques and data mining experiences from a combination of many existing research areas: databases, statistics, mathematical logic, machine learning, automated scientific discovery, inductive logic programming, artificial intelligence, visualization, decision science, knowledge management, and high performance computing. The strength of KDD came initially from the value added by the creative combination of techniques from the contributing areas. But in order to establish its identity, KDD has to create its own theoretical principles and to demonstrate how they stimulate KDD research, facilitate communication, and guide practitioners towards successful applications.

Seeking the principles that can guide and strengthen practical applications has been always a part of the European research tradition. Thus “Principles and Practice of KDD” (PKDD) make a suitable focus for annual meetings of the KDD community in Europe. The main long-term interest is in theoretical principles for the emerging discipline of KDD and in practical applications that demonstrate utility of those principles. Other goals of the PKDD series are to provide a European-based forum for interaction among all theoreticians and practitioners interested in data mining and knowledge discovery as well as to foster new directions in interdisciplinary collaboration.

A Discovery Challenge hosted at PKDD 2000 was conducted for the second year in a row. It followed a successful and very broadly attended Discovery Challenge in Prague. Discovery challenge is a new initiative that promotes cooperative research on original and practically important real-world databases. The Challenge requires a broad and unified view of knowledge and methods of discovery, and emphasizes business problems that require an open-minded search for knowledge in data. Two multi-relational databases, in banking and in medicine, were available to all conference participants. The Challenge was born out of the conviction that knowledge discovery in real-world databases requires

an open-minded discovery process rather than application of one or another tool limited to one form of knowledge. A discoverer should consider a broad scope of techniques that can reach many forms of knowledge. The discovery process cannot be rigid and selection of techniques must be driven by knowledge hidden in the data, so that the most and the best of knowledge can be reached.

The contributed papers were selected from 157 full papers (48% growth over PKDD'99) by the following program committee:

Magnus L. Andersson (AstraZeneca, Sweden),
Petr Berka (U. Economics, Czech Republic),
Jean-Francois Boulicaut (INSA Lyon, France),
Henri Briand (U. Nantes, France),
Leo Carbonara (American Management System, UK),
Luc De Raedt (U. Freiburg, Germany),
Ronen Feldman (Bar Ilan U., Israel),
Arthur Flexer (Austrian Research Inst. for AI, Austria),
Alex Freitas (PUC-PR, Brazil),
Patrick Gallinari (U. Paris 6, France),
Jean Gabriel Ganascia (U. Paris 6, France),
Mohand-Said Hacid (U. Lyon I, INSA Lyon, France),
Petr Hájek (Acad. Science, Czech Republic),
Howard Hamilton (U. Regina, Canada),
David Hand (Imperial Col. London, UK),
Mika Klemettinen (U. Helsinki, Finland),
Willi Kloesgen (GMD, Germany),
Yves Kodratoff (U. Paris-Sud, France),
Jan Komorowski (Norwegian U. Sci. & Tech.),
Jacek Koronacki (Acad. Science, Poland),
Michel Lamure (U. Lyon I, France),
Heikki Manilla (Nokia, Finland),
Stan Matwin (U. Ottawa, Canada),
Hiroshi Motoda (Osaka U., Japan),
Jan Mrazek (Bank of Montreal, Canada),
Witold Pedrycz (U. Alberta, Canada),
Mohamed Quafafou (U. Nantes, France),
Ricco Rakotomalala (F. Lyon II, France),
Jan Rauch (U. Economics, Czech Republic),
Zbigniew Ras (UNC Charlotte, USA),
Gilbert Ritschard (U. Geneva, Switzerland),
Michele Sebag (Ecole Polytech. Paris, France),
Arno Siebes (CWI, Netherlands),
Andrzej Skowron (U. Warsaw, Poland),
Myra Spiliopoulou (Humboldt U. Berlin, Germany),
Nicolas Spyrtos (U. Paris-Sud, France),
Olga Štěpánková (Czech Tech. U., Czech Republic),
Einoshin Suzuki (Yokohama Natl. U., Japan),

Ljupco Todorovski (Josef Stefan Inst., Slovenia),
Shusaku Tsumoto (Shimane Medical U., Japan),
Gilles Venturini (U. Tours, France),
Inkeri Verkamo (U. Helsinki, Finland),
Louis Wehenkel (U. Liege, Belgium),
Gerhard Widmer (U. Vienna, Austria),
Rudiger Wirth (DaimlerChrysler, Germany),
Stefan Wrobel (U. Magdeburg, Germany),
Ning Zhong (Maebashi Inst. Techn., Japan),
Djamel A. Zighed (U. Lyon 2, France), and
Jan Żytkow (UNC Charlotte, USA).

The following colleagues also reviewed papers for the conference and are due our special thanks:

Florence d'Alche buc (U. Paris 6),
Jérôme Darmont (U. Lyon 2),
Stéphane Lallich (U. Lyon 2), and
Christel Vrain (U. Orleans).

Classified according to the first author's nationality, papers submitted to PKDD'2000 came from 34 countries on 5 continents (Europe: 102 papers; Asia: 21; North America: 17; Australia: 3; and South America: 4), including Argentina (1 paper), Australia (3), Austria (3), Belgium (5), Brazil (3), Canada (4), Croatia (1), Czech Republic (4), Finland (3), France (38), Germany (7), Greece (5), Hong Kong (1), Italy (3), Japan (9), Mexico (2), Netherlands (1), Norway (1), Poland (5), Portugal (2), Romania (1), Russia (1), Singapore (4), Slovakia (1), Slovenia (1), South Korea (3), Spain (5), Sweden (1), Switzerland (3), Taiwan (4), Turkey (1), United Kingdom (10), and USA (20).

Many thanks to all who submitted papers for review and for publication in the proceedings. The accepted papers were divided into two categories: 29 oral presentations and 57 poster presentations. In addition to poster sessions each poster paper was available for the web-based discussion forum in the months preceding the conference.

There were three invited presentations:

- Sep.14, **Willi Kloesgen** (GMD, Germany), Multi-relational, statistical, and visualization approaches for spatial knowledge discovery
- Sep.15 **Luc De Raedt** (U.Freiburg, Germany), Data mining in multi-relational databases
- Sep.16 **Arno Siebes** (Utrecht U.), Developing KDD Systems

Six workshops were affiliated with the conference. All were held on 12 September.

- **Workshop 1** Data Mining, Decision Support, Meta-learning, and ILP. Chairs: **Pavel Brazdil and Alipio Jorge, U. Porto, Portugal**;
- **Workshop 2** Temporal, Spatial, and Spatio-temporal Data Mining (TSDM 2000). Chairs: **John F. Roddick, Flinders U., Australia & Kathleen Hornsby, U. Maine, USA**;

- **Workshop 3** Knowledge Discovery in Biology. Chair: **Jan Komorowski, Norwegian U. Sci. & Techn, Norway**;
- **Workshop 4** Machine Learning and Textual Information Access. Chairs: **Hugo Zaragoza, LIP6, U. Paris 6, France** **Patrick Gallinari, LIP6, U. Paris 6, France** **Martin Rajman, EPFL, Switzerland**;
- **Workshop 5** Knowledge Management: Theory and Applications. Chair: **Jean-Louis Ermine, CEA Paris, France**;
- **Workshop 6** Symbolic Data Analysis: Theory, Software, and Applications for Knowledge Mining. Chair: **Edwin Diday, U. Paris Dauphine, France**.

Five tutorials were offered to all conference participants on 13 September:

- **Tutorial 1** An Introduction to Distributed Data Mining, presented by **H. Kargupta, Washington State U., USA**;
- **Tutorial 2** Clustering Techniques for Large Data Sets: From the Past to the Future, presented by **A. Hinneburg and D.A. Keim, U. Halle, Germany**;
- **Tutorial 3** Data Analysis for Web Marketing and Merchandizing Applications, presented by **M. Spiliopoulou, Humboldt U. Berlin, Germany**;
- **Tutorial 4** Database Support for Business Intelligence Applications, presented by **W. Lehner, U. Erlangen-Nürnberg, Germany**;
- **Tutorial 5** Text Mining, presented by **Yves Kodratoff, Djamel Zighed, and Serge Di Palma (France)**.

Members of the PKDD'2000 organizing committee did an enormous amount of work and deserve the special gratitude of all participants: **Willi Kloesgen** – Tutorials Chair, **Jan Rauch** – Workshops Chair, **Arno Siebes and Petr Berka** – Discovery Challenge Chairs, **Leonardo Carbonara** – Industrial Program Chair, **Djamel Zighed** – Local Arrangement Chair, and **Jean Hugues Chauchat, Fabrice Muhlenbach, Laure Tougne, Fadila Bentayeb, Omar Boussaid, Salima Hassas, Stéphane Lallich, Céline Agier, David Coeurjolly, Alfredos Mar, Luminita Firanescu Astrid Varaine, Sabine Rabaséda, Fabien Feschet, and Nicolas Nicoloyannis**.

Special thanks go to **Alfred Hofmann** of Springer-Verlag for his continuous help and support and to **Karin Henzold** for the preparation of the Proceedings.