Lecture Notes in Artificial Intelligence 1510

Subseries of Lecture Notes in Computer Science
Edited by J. G. Carbonell and J. Siekmann

Lecture Notes in Computer Science
Edited by G. Goos, J. Hartmanis and J. van Leeuwen

Springer
Berlin
Heidelberg
New York
Barcelona
Budapest
Hong Kong
London
Milan
Paris
Singapore
Tokyo

Jan M. Żytkow Mohamed Quafafou (Eds.)

Principles of Data Mining and Knowledge Discovery

Second European Symposium, PKDD '98
Nantes, France, September 23-26, 1998
Proceedings

Springer

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Jan M. Żytkow
Wichita State University, Department of Computer Science
Wichita, KS 67260-0083, USA
E-mail: zytkow@wise.cs.twsu.edu

Mohamed Quafafou
Université de Nantes, IRIN
2, rue de la Houssinière, F-44322 Nantes Cedex 3, France
E-mail: quafafou@irin.univ-nantes.fr

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Principles of data mining and knowledge discovery : second European symposium ; proceedings / PKDD '98, Nantes, France, September 23 - 26, 1998. Jan M. Żytkow ; Mohamed Quafafou (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Budapest ; Hong Kong ; London ; Milan ; Paris ; Singapore ; Tokyo : Springer, 1998
(Lecture notes in computer science ; Vol. 1510 : Lecture notes in artificial intelligence)
ISBN 3-540-65068-7

CR Subject Classification (1991): I.2, H.3, H.5, G.3, J.1

ISBN 3-540-65068-7 Springer-Verlag Berlin Heidelberg New York

Typesetting: Camera ready by author
SPIN 10692663 06/3142 – 5 4 3 2 1 0 Printed on acid-free paper

Preface

Knowledge Discovery in Databases (KDD), also known as Data Mining, has emerged in the last decade in response to the challenge of turning large and ubiquitous databases into knowledge that can be used in practice.

KDD has been able to grow very rapidly by drawing its techniques and data mining experiences from a combination of many existing research areas: databases, statistics, machine learning, automated scientific discovery, inductive logic programming, artificial intelligence, visualization, decision science, and high performance computing. While each of these areas can contribute in specific ways, the strength of KDD comes from the value that is added by creative combination of techniques from the contributing areas.

The practical successes of KDD in a broad range of application domains have led to very high expectations. In the long run, KDD can meet or fail those expectations. In order to maintain its status of interdisciplinary research of great practical payoff KDD has to establish its own theoretical principles that go beyond each of the contributing areas, and demonstrate how they jointly create a broad and exciting area of research. Such principles will be instrumental in maintaining the identity of KDD research, in effective communication and in guiding the practitioners.

Seeking the principles has always been a part of the European research tradition. Thus "Principles of KDD" (PKDD) make a suitable focus for the annual meetings of the KDD community in Europe. The main long-term interest is in theoretical principles for the emerging discipline of KDD. Another goal of the PKDD series is to provide a European-based forum for interaction among all theoreticians and practitioners interested in data mining and knowledge discovery as well as fostering the interdisciplinary collaboration. The first meeting was held in Trondheim, Norway, in June 1997.

This volume contains papers selected for presentation at the Second European Symposium on Principles of Data Mining and Knowledge Discovery in Databases – PKDD'98, held in Nantes, France, September 23-26, 1998. The University of Nantes hosted the symposium. The symposium was sponsored by the Ville de Nantes, Université de Nantes, MENRT, Conseil Régional des Pays de Loire, Centre de Recherche et Développement de France Télécom (CNET). We wish to express our thanks to the sponsors of the symposium for their generous support. The contributed papers were selected from 73 full draft papers by the following program committee:
Pieter Adriaans (Syllogic, The Netherlands), **Pawel Brazdil** (U. Porto, Portugal), **Henri Briand** (U. Nantes, France), **Leo Carbonara** (British Telecom, UK), **A. Fazel Famili** (IIT-NRC, Canada), **Ronen Feldman** (Bar Ilan, U. Israel), **Patrick Gallinari** (U. Paris VI, France), **Jean Gabriel Ganascia** (U. Paris VI, France), **Attilio Giordana** (U. Torino, Italy), **David Hand** (Open U. UK), **Bob Henery** (U. Strathclyde, UK), **Mikhail Kiselev**

(Megaputer Intelligence, Russia), **Willi Kloesgen** (GMD, Germany), **Yves Kodratoff** (U. Paris VI, France), **Jan Komorowski** (NTNU, Norway), **Nada Lavrac** (Josef Stefan Inst. Slovenia), **Heikki Manilla** (U. Helsinki, Finland), **Steve Muggleton** (Oxford U. UK), **Zdzislaw Pawlak** (Warsaw Technical U. Poland), **Gregory Piatetsky-Shapiro** (Knowledge Stream, Boston, USA), **Lech Polkowski** (U. Warsaw, Poland), **Mohamed Quafafou** (U. Nantes, France), **Zbigniew Ras** (UNC Charlotte, USA), **Lorenza Saitta** (U. Torino, Italy), **Wei-Min Shen** (U. South Calif. USA), **Arno Siebes** (CWI, Netherlands), **Andrzej Skowron** (U. Warsaw, Poland), **Derek Sleeman** (U. Aberdeen, UK), **Nicolas Spyratos** (U. Paris XI, France), **Shusaku Tsumoto** (Tokyo Medical & Dental U. Japan), **Raul Valdes-Perez** (CMU, USA), **Thierry Van de Merckt** (CSC, Belgium), **Rudiger Wirth** (Daimler-Benz, Germany), **Stefan Wrobel** (GTE, Germany), **Ning Zhong** (Yamaguchi U. Japan), **Wojtek Ziarko** (U. Regina, Canada), **Djamel A. Zighed** (U. Lyon II, France), **Jan M. Żytkow** (UNC Charlotte, USA).

PKDD was truly international: papers came from 24 countries, including Australia (2), Belgium (1), Brazil (1), Bulgaria (1), Canada (1), China (1), Cuba (1), Czech Republic (3), Finland (2), France (18), Germany (6), Israel (3), Italy (2), Japan (5), Norway (1), Poland (4), Portugal (2), Russia (1), Singapore (2), Spain (5), Sweden (1), Switzerland (1), United Kingdom (6), United States of America (3). Many thanks to all who submitted papers for review and for publication in the proceedings. The accepted papers were divided into two categories: 26 oral presentations and 30 poster presentations. In addition to poster sessions each poster paper was allocated 4 minutes highlight presentation. All papers were allocated the same number of 9 pages in the proceedings.

Three tutorials were offered to all symposium participants on September 23rd: Scalable, High-Performance Data Mining with Parallel Processing by Alex Alves Freitas, Industrial Applications of Data Mining by Gholamreza Nakhaeizadeh, and Practical Text Mining by Ronen Feldman.

The members of PKDD'98 local organizing committee: Emmanuelle Martienne, Laurent Ughetto, and Abdellatif Saoudi, all of the University of Nantes, did an enormous amount of work and deserve the special gratitude of all the participants. We wish to express our thanks to the following colleagues for their reviewing participation: M.R. Amini, J-F. Boulicaut, V. Corruble, A. Giacometti, A. Jorge, A. Leger, D. Laurent, F. Mitchell, H.S. Nguyen, D. Slezak, N. Valette, H. Zaragoza, J-D. Zucker.

Special thanks are due to Alfred Hofmann of Springer-Verlag for his help and support.

July 1998 Mohamed Quafafou and Jan M. Żytkow

Table of Contents

Communications

Session 1. Rule evaluation

Session 2. Visualization

Session 3. Association rules and text mining

Session 4. Clustering and discretization

Session 5. KDD process and software

Session 6. Tree construction

Session 7. Sequential and spatial data mining

Session 8. Attribute selection

Posters

Tutorials