Lecture Notes in Artificial Intelligence 3755

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Graham J. Williams Simeon J. Simoff (Eds.)

Data Mining

Theory, Methodology, Techniques, and Applications

Springer

Series Editors

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

Volume Editors

Graham J. Williams
Togaware Data Mining
Canberra, Australia
E-mail: graham.williams@togaware.com

Simeon J. Simoff
University of Technology, Faculty of Information Technology
Sydney Broadway PO Box 123, NSW 2007, Australia
E-mail: simeon@it.uts.edu.au

Library of Congress Control Number: 2006920576

CR Subject Classification (1998): I.2, H.2.8, H.2-3, D.3.3, F.1

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-540-32547-6 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-32547-5 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11677437 06/3142 5 4 3 2 1 0

Preface

Data mining has been an area of considerable research and application in Australia and the region for many years. This has resulted in the establishment of a strong tradition of academic and industry scholarship, blended with the pragmatics of practice in the field of data mining and analytics. ID3, See5, RuleQuest.com, MagnumOpus, and WEKA is but a short list of the data mining tools and technologies that have been developed in Australasia. Data mining conferences held in Australia have attracted considerable international interest and involvement.

This book brings together a unique collection of chapters that cover the breadth and depth of data mining today. This volume provides a snapshot of the current state of the art in data mining, presenting it both in terms of technical developments and industry applications. Authors include some of Australia's leading researchers and practitioners in data mining, together with chapters from regional and international authors.

The collection of chapters is based on works presented at the Australasian Data Mining conference series and industry forums. The original papers were initially reviewed for the workshops, conferences and forums. Presenting authors were provided with substantial feedback, both through this initial review process and through editorial feedback from their presentations. A final international peer review process was conducted to include input from potential users of the research, and in particular analytics experts from industry, looking at the impact of reviewed works.

Many people contribute to an effort such as this, starting with the authors! We thank all authors for their contributions, and particularly for making the effort to address two rounds of reviewer comments. Our workshop and conference reviewers provided the first round of helpful feedback for the presentation of the papers to their respective conferences. The authors from a selection of the best papers were then invited to update their contributions for inclusion in this volume. Each submission was then reviewed by at least another two reviewers from our international panel of experts in data mining.

A considerable amount of effort goes into reviewing papers, and reviewers perform an essential task. Reviewers receive no remuneration for all their efforts, but are happy to provide their time and expertise for the benefit of the whole community. We owe a considerable debt to them all and thank them for their enthusiasm and critical efforts.

Bringing this collection together has been quite an effort. We also acknowledge the support of our respective institutions and colleagues who have contributed in many different ways. In particular, Graham would like to thank Togaware (Data Mining and GNU/Linux consultancy) for their ongoing infrastructural support over the years, and the Australian Taxation Office for its

support of data mining and related local conferences through the participation of its staff. Simeon acknowledges the support of the University of Technology, Sydney. The Australian Research Council's Research Network on Data Mining and Knowledge Discovery, under the leadership of Professor John Roddick, Flinders University, has also provided support for the associated conferences, in particular providing financial support to assist student participation in the conferences. Professor Geoffrey Webb, Monash University, has played a supportive role in the development of data mining in Australia and the AusDM series of conferences, and continues to contribute extensively to the conference series.

The book is divided into two parts: (i) state-of-art research and (ii) state-of-art industry applications. The chapters are further grouped around common sub-themes. We are sure you will find that the book provides an interesting and broad update on current research and development in data mining.

November 2005 Graham Williams and Simeon Simoff

Organization

Many colleagues have contributed to the success of the series of data mining workshops and conferences over the years. We list here the primary reviewers who now make up the International Panel of Expert Reviewers.

AusDM Conference Chairs

Simeon J. Simoff, University of Technology, Sydney, Australia
Graham J. Williams, Australian National University, Canberra

PAKDD Industry Chair

Graham J. Williams, Australian National University, Canberra

International Panel of Expert Reviewers

Mihael Ankerst	Boeing Corp., USA
Michael Bain	University of New South Wales, Australia
Rohan Baxter	Australian Taxation Office
Helmut Berger	University of Technology, Sydney, Australia
Michael Bohlen	Free University Bolzano-Bozen, Italy
Jie Chen	CSIRO, Canberra, Australia
Peter Christen	Australian National University
Thanh-Nghi Do	Can Tho University, Vietnam
Vladimir Estivill-Castro	Giffith University, Australia
Hongjian Fan	University of Melbourne, Australia
Eibe Frank	Waikato University, New Zealand
Mohamed Medhat Gaber	Monash University, Australia
Raj Gopalan	Curtin University, Australia
Warwick Graco	Australian Taxation Office
Lifang Gu	Australian Taxation Office
Hongxing He	CSIRO, Canberra, Australia
Robert Hilderman	University of Regina, Canada
Joshua Zhexue Huang	University of Hong Kong, China
Huidong Jin	CSIRO, Canberra, Australia
Paul Kennedy	University of Technology, Sydney, Australia
Weiqiang Lin	Australian Taxation Office
John Maindonald	Australian National University
Mark Norrie	Teradata, NCR, Australia
Peter O'Hanlon	Westpac, Australia

Mehmet Orgun	Macquarie University, Australia
Tom Osborn	Wunderman, NUIX Pty Ltd, Australia
Robert Pearson	Health Insurance Commission, Australia
Francois Poulet	ESIEA-Pole ECD, Laval, France
John Roddick	Flinders University, Australia
Greg Saunders	University of Ballarat, Australia
David Skillicorn	Queen's University, Canada
Geoffrey Webb	Monash University, Australia
John Yearwood	University of Ballarat, Australia
Osmar Zaiane	University of Alberta, Canada

Table of Contents

Part 1: State-of-the-Art in Research

Methodological Advances

Data Linkage

Text Mining

Temporal and Sequence Mining

Part 2: State-of-the-Art in Applications

Health

Finance and Retail