

Shusaku Tsumoto Takahira Yamaguchi
Masayuki Numao Hiroshi Motoda (Eds.)

Active Mining

Second International Workshop, AM 2003
Maebashi, Japan, October 28, 2003
Revised Selected Papers

Series Editors

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

Volume Editors

Shusaku Tsumoto
Shimane University, School of Medicine
Department of Medical Informatics
89-1 Enya-cho, Izumo, Shimane 693-8501, Japan
E-mail: tsumoto@computer.org

Takahira Yamaguchi
Keio University, Faculty of Science and Technology
Department of Administration Engineering
3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan
E-mail: yamaguti@ae.keio.ac.jp

Masayuki Numao
Osaka University, The Institute of Scientific and Industrial Research
Division of Intelligent Systems Science, Dept. of Architecture for Intelligence
8-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan
E-mail: numao@ai.sanken.osaka-u.ac.jp

Hiroshi Motoda
Osaka University, The Institute of Scientific and Industrial Research
Division of Intelligent Systems Science, Dept. of Advance Reasoning
8-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan
E-mail: motoda@ar.sanken.osaka-u.ac.jp

Library of Congress Control Number: 2005926624

CR Subject Classification (1998): H.2.8, I.2, F.2.2, J.3

ISSN 0302-9743
ISBN-10 3-540-26157-5 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-26157-5 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media

springeronline.com

© Springer-Verlag Berlin Heidelberg 2005
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11423270 06/3142 5 4 3 2 1 0

Foreword

This volume contains the papers selected for presentation at the 2nd International Workshop on Active Mining (AM 2003) which was organized in conjunction with the 14th International Symposium on Methodologies for Intelligent Systems (ISMIS 2003), held in Maebashi City, Japan, 28–31 October, 2003. The workshop was organized by the Maebashi Institute of Technology in cooperation with the Japanese Society for Artificial Intelligence. It was sponsored by the Maebashi Institute of Technology, the Maebashi Convention Bureau, the Maebashi City Government, the Gunma Prefecture Government, JSAI SIGKBS (Japanese Artificial Intelligence Society, Special Interest Group on Knowledge-Based Systems), a Grant-in-Aid for Scientific Research on Priority Areas (No. 759) “Implementation of Active Mining in the Era of Information Flood,” US AFOSR/AOARD, the Web Intelligence Consortium (Japan), the Gunma Information Service Industry Association, and Ryomo Systems Co., Ltd.

ISMIS is a conference series that was started in 1986 in Knoxville, Tennessee. Since then it has been held in Charlotte (North Carolina), Knoxville (Tennessee), Torin (Italy), Trondheim (Norway), Warsaw (Poland), Zakopane (Poland), and Lyon (France).

The objective of this workshop was to gather researchers as well as practitioners who are working on various research fields of active mining, share hard-learned experiences, and shed light on the future development of active mining. This workshop addressed many aspects of active mining, ranging from theories, methodologies, and algorithms to their applications. We believe that it also produced a contemporary overview of modern solutions and it created a synergy among different branches but with a similar goal — facilitating data collection, processing, and knowledge discovery via active mining.

We express our appreciation to the sponsors of the symposium and to all who submitted papers for presentation and publication in the proceedings. Our sincere thanks especially go to the Organizing Committee of AM 2003: Shusaku Tsumoto, Hiroshi Motoda, Masayuki Numao, and Takahira Yamaguchi. Also, our thanks are due to Alfred Hofmann of Springer for his continuous support.

March 2005

Ning Zhong, Zbigniew W. Raś
ISMIS'03 Symposium Chairs

Preface

This volume contains the papers based on the tutorials of ISMIS 2003 and the papers selected from the regular papers presented at the 2nd International Workshop on Active Mining (AM 2003) held as part of ISMIS 2003, held at Maebashi, Gunma, October 28, 2003. (URL: http://www.med.shimane-u.ac.jp/med_info/am2003)

There were 38 paper submissions from Japan, US, Korea, China and Vietnam for AM 2003. Papers went through a rigorous reviewing process. Each paper was reviewed by at least three Program Committee members. When all the reviews of a paper were in conflict, another PC member was asked to review this paper again. Finally, 20 submissions were selected by the Program Committee members for presentation.

The PC members who attended this workshop reviewed all the papers and the presentations during the workshops and decided that 12 of them could be accepted with minor revisions, and 8 of them could be accepted with major revisions for the publication in the postproceedings. The authors were asked to submit the revised versions of their papers and, again, the PC members reviewed them for about four months. Through this process, we accepted 16 out of 20 papers. Thus, of 38 papers submitted, 16 were accepted for this postproceedings, corresponding to an acceptance ratio of only 42.1%.

AM 2003 provided a forum for exchanging ideas among many researchers in various areas of decision support, data mining, machine learning and information retrieval and served as a stimulus for mutual understanding and cooperation.

The papers contributed to this volume reflect advances in active mining as well as complementary research efforts in the following areas:

- Discovery of new information sources
- Active collection of information
- Tools for information collection
- Information filtering
- Information retrieval, collection, and integration on the WWW for data mining
- Data mining processes
- Inspection and validation of mined pieces of knowledge
- Description languages for discovery evaluation and accountability
- Interactive mining design and deployment of customer response models in CRM
- Adaptive modeling in data mining
- Selection, transformation, and construction of features
- Selection and construction of instances
- Exception/deviation discovery visualization
- Spatial data mining

- Text mining
- Graph mining
- Success/failure stories in data mining and lessons learned
- Data mining for evidence-based medicine
- Distributed data mining
- Data mining for knowledge management
- Active learning
- Meta learning
- Active sampling
- Usability of mined pieces of knowledge
- User interface for data mining

We wish to express our gratitude to Profs. Zbigniew W. Raś and Ning Zhong, who accepted our proposal on this workshop and helped us to publish the post-proceedings. Without their sincere help, we could not have published this volume.

We wish to express our thanks to all the PC members, who reviewed the papers at least twice, for the workshop and the postproceedings. Without their contributions, we could not have selected high-quality papers with high confidence.

We also want to thank all the authors who submitted valuable papers to AM 2003 and all the workshop attendees.

This time, all the submissions and reviews were made through the CyberChair system (URL: <http://www.cyberchair.org/>). We wish to thank the CyberChair system development team. Without this system, we could not have edited this volume in such a speedy way.

We also extend our special thanks to Dr. Shoji Hirano, who launched the CyberChair system for AM 2003 and contributed to editing this volume.

Finally, we wish to express our thanks to Alfred Hofmann at Springer for his support and cooperation.

March 2005

Shusaku Tsumoto
Takahira Yamaguchi
Masayuki Numao
Hiroshi Motoda

AM 2003 Conference Committee

Organizing Committee

Hiroshi Motoda (Osaka University, Japan)
Masayuki Numao (Osaka University, Japan)
Takahira Yamaguchi (Keio University, Japan)
Shusaku Tsumoto (Shimane University, Japan)

Program Committee

Hiroki Arimura (Hokkaido University, Japan)
Stephen D. Bay (Stanford University, USA)
Wesley Chu (UCLA, USA)
Saso Dzeroski (Jozef Stefan Institute, Slovenia)
Shoji Hirano (Shimane University, Japan)
Tu Bao Ho (JAIST, Japan)
Robert H.P. Engels (CognIT, Norway)
Ryutaro Ichise (NII, Japan)
Akihiro Inokuchi (IBM Japan, Japan)
Hiroyuki Kawano (Nanzan University, Japan)
Yasuhiko Kitamura (Kwansei Gakuin University, Japan)
Marzena Kryszkiewicz (Warsaw University of Technology, Poland)
T.Y. Lin (San Jose State University, USA)
Bing Liu (University of Illinois at Chicago, USA)
Huan Liu (Arizona State University, USA)
Tsuyoshi Murata (NII, Japan)
Masayuki Numao (Osaka University, Japan)
Miho Ohsaki (Doshisha University, Japan)
Takashi Onoda (CRIPEL, Japan)
Luc de Raedt (University of Freiburg, Germany)
Zbigniew Raś (University of North Carolina, USA)
Henryk Rybinski (Warsaw University of Technology)
Masashi Shimbo (NAIST, Japan)
Einoshin Suzuki (Yokohama National University, Japan)
Masahiro Terabe (MRI, Japan)
Ljupico Todorovski (Jozef Stefan Institute, Slovenia)
Seiji Yamada (NII, Japan)
Yiyu Yao (University of Regina, Canada)
Kenichi Yoshida (University of Tsukuba, Japan)
Tetsuya Yoshida (Hokkaido University, Japan)
Stefan Wrobel (University of Magdeburg, Germany)
Ning Zhong (Maebashi Institute of Technology, Japan)

Table of Contents

Overview

Active Mining Project: Overview <i>Shusaku Tsumoto, Takahira Yamaguchi, Masayuki Numao, Hiroshi Motoda</i>	1
---	---

Tutorial Papers

Computational and Statistical Methods in Bioinformatics <i>Tatsuya Akutsu</i>	11
Indexing and Mining Audiovisual Data <i>Pierre Morizet-Mahoudeaux, Bruno Bachimont</i>	34

Active Information Collection

Relevance Feedback Document Retrieval Using Support Vector Machines <i>Takashi Onoda, Hiroshi Murata, Seiji Yamada</i>	59
Micro View and Macro View Approaches to Discovered Rule Filtering <i>Yasuhiko Kitamura, Akira Iida, Keunsik Park</i>	74
Mining Chemical Compound Structure Data Using Inductive Logic Programming <i>Cholwich Nattee, Sukree Sinthupinyo, Masayuki Numao, Takashi Okada</i>	92
First-Order Rule Mining by Using Graphs Created from Temporal Medical Data <i>Ryutaro Ichise, Masayuki Numao</i>	112

Active Data Mining

Extracting Diagnostic Knowledge from Hepatitis Dataset by Decision Tree Graph-Based Induction <i>Warodom Geamsakul, Tetsuya Yoshida, Kouzou Ohara, Hiroshi Motoda, Takashi Washio, Hideto Yokoi, Katsuhiko Takabayashi</i>	126
--	-----

Data Mining Oriented CRM Systems Based on MUSASHI: C-MUSASHI <i>Katsutoshi Yada, Yukinobu Hamuro, Naoki Katoh, Takashi Washio, Issey Fusamoto, Daisuke Fujishima, Takaya Ikeda</i>	152
Investigation of Rule Interestingness in Medical Data Mining <i>Miho Ohsaki, Shinya Kitaguchi, Hideto Yokoi, Takahira Yamaguchi</i> . .	174
Experimental Evaluation of Time-Series Decision Tree <i>Yuu Yamada, Einoshin Suzuki, Hideto Yokoi, Katsuhiko Takabayashi</i>	190
Spiral Multi-aspect Hepatitis Data Mining <i>Muneaki Ohshima, Tomohiro Okuno, Yasuo Fujita, Ning Zhong, Juzhen Dong, Hideto Yokoi</i>	210
Sentence Role Identification in Medline Abstracts: Training Classifier with Structured Abstracts <i>Masashi Shimbo, Takahiro Yamasaki, Yuji Matsumoto</i>	236
<i>CHASE</i> ₂ – Rule Based Chase Algorithm for Information Systems of Type λ <i>Agnieszka Dardzińska, Zbigniew W. Raś</i>	255
 Active User Reaction	
Empirical Comparison of Clustering Methods for Long Time-Series Databases <i>Shoji Hirano, Shusaku Tsumoto</i>	268
Spiral Mining Using Attributes from 3D Molecular Structures <i>Takashi Okada, Masumi Yamakawa, Hirotaka Niitsuma</i>	287
Classification of Pharmacological Activity of Drugs Using Support Vector Machine <i>Yoshimasa Takahashi, Katsumi Nishikoori, Satoshi Fujishima</i>	303
Cooperative Scenario Mining from Blood Test Data of Hepatitis B and C <i>Yukio Ohsawa, Hajime Fujie, Akio Saiura, Naoaki Okazaki, Naohiro Matsumura</i>	312
Integrated Mining for Cancer Incidence Factors from Healthcare Data <i>Xiaolong Zhang, Tetsuo Narita</i>	336
Author Index	349