

Lecture Notes in Artificial Intelligence 5433

Edited by R. Goebel, J. Siekmann, and W. Wahlster

Subseries of Lecture Notes in Computer Science

Sanjay Chawla Takashi Washio
Shin-ichi Minato Shusaku Tsumoto
Takashi Onoda Seiji Yamada
Akihiro Inokuchi (Eds.)

New Frontiers in Applied Data Mining

PAKDD 2008 International Workshops
Osaka, Japan, May 20-23, 2008
Revised Selected Papers

Series Editors

Randy Goebel, University of Alberta, Edmonton, Canada
Jörg Siekmann, University of Saarland, Saarbrücken, Germany
Wolfgang Wahlster, DFKI and University of Saarland, Saarbrücken, Germany

Volume Editors

Sanjay Chawla
University of Sydney, NSW, Australia
E-mail: chawla@it.usyd.edu

Takashi Washio
Akihiro Inokuchi
Osaka University, Osaka, Japan
E-mail: {washio,inokuchi}@ar.sanken.osaka-u.ac.jp

Shin-ichi Minato
Hokkaido University, Sapporo, Japan
E-mail: minato@ist.hokudai.ac.jp

Shusaku Tsumoto
Shimane University, Izumo, Shimane, Japan
E-mail: tsumoto@computer.org

Takashi Onoda
Central Research Institute of Electric Power Industry, Tokyo, Japan
E-mail: onoda@criepi.denken.or.jp

Seiji Yamada
National Institute of Informatics, Tokyo, Japan
E-mail: seiji@nii.ac.jp

Library of Congress Control Number: 2009920983

CR Subject Classification (1998): I.2, H.2.7-8, H.3, H.5.1, G.3, J.3

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-642-00398-2 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-00398-1 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.com

© Springer-Verlag Berlin Heidelberg 2009
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12613861 06/3180 5 4 3 2 1 0

Preface

As data mining techniques and tools mature, their application domains extend to previous unchartered territories. The common theme of the workshops organized along with the main 2008 Pacific Asia Conference on Knowledge Discovery and Data Mining (PAKDD) in Osaka, Japan was to extend the application of data mining techniques to new frontiers. Thus the title of the proceedings: “New Frontiers in Application of Data Mining.”

For the 2008 program, three workshops were organized.

1. Algorithms for Large-Scale Information Processing (ALSIP). The focus of the workshop was novel algorithms and data structures to deal with processing of very large data sets.
2. Data Mining for Decision Making and Risk Management (DMDRM), which emphasized applications of risk information derived from data mining techniques on diverse applications ranging from medicine to marketing to chemistry.
3. Interactive Data Mining (IDM), which emphasized the relationship between techniques in data mining and human–computer interaction.

In total 38 papers were submitted to the workshops. After consultation with the workshop Chairs who were asked to rank their submissions, 18 were accepted for publication in this volume. We hope that the published papers propel further interest in the growing field of knowledge discovery in databases (KDD).

The paper selection of the industrial track and the workshops was made by the Program Committee of each organization. Upon the paper selection, the book was edited and managed by the volume editors.

December 2008

Sanjay Chawla
Takashi Washio

Workshop on Algorithms for Large-Scale Information Processing in Knowledge Discovery

The International Workshop on Algorithms for Large-Scale Information Processing in Knowledge Discovery (ALSIP 2008) was held on May 20, 2008 at Hotel Seagull Tempoan Osaka, Japan, in conjunction with the 12th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2008). This workshop was co-organized by MEXT Japan Grant-in-Aid for Science Research and Priority Area: Cyber Infrastructure for the Information-explosion Era, and the JSPS G-COE program of Hokkaido University: Next-Generation Knowledge Information Science for Future Science and Technology. The workshop aimed to exchange fresh ideas on large-scale data processing in problems such as data mining, clustering, machine learning, statistical analysis, and other computational aspect of knowledge discovery problems. Information created by people has increased rapidly since the year 2000, and now we are in a time that we could call the information-explosion era. To cope with such a large-scale information space, novel algorithms and data structures are desired for solving various problems in the area of knowledge discovery. This workshop aimed to exchange fresh ideas on large-scale data processing for various knowledge discovery problems. The topics of the workshop in the call for papers included:

- Machine learning, clustering and statistical methods
- Large-scale itemset mining and associate rule mining
- Graph-based data structures for knowledge representation
- Knowledge data compression and indexing
- Knowledge discovery from text and the Web
- Knowledge discovery from unstructured and multimedia data
- Knowledge discovery from data stream and spatial/temporal data
- Knowledge discovery in network and link data
- Biomedical knowledge discovery, analysis of microarray and gene deletion data
- Information extraction from scientific literature
- Active knowledge discovery
- Data and knowledge visualization
- Other computational aspects of knowledge discovery problems.

ALSIP 2008 was organized as a full-day workshop. All submitted papers were reviewed for quality and originality by the Program Committee. We accepted 11 technical papers out of 16 submissions. We also invited a special talk “The Challenge of Mining Billions of Transactions” by Osmar Zaiane (University of Alberta, Canada). By additional reviewing after the workshop, we selected 10 papers for this proceedings volume.

We are grateful for the great support from the ALSIP 2008 Program Committee members. We are also grateful to the PAKDD 2008 committee for their help,

especially Takashi Washio for his arrangements in conjunction with the PAKDD. We gratefully acknowledge Masaru Kitsuregawa, the leader of the MEXT project of the Information-explosion Era, for giving us an opportunity to organize this workshop, and Hiroki Arimura, the leader of the Hokkaido University GCOE program, for his support of the ALSIP workshop.

Workshop Chair

Shin-ichi Minato Hokkaido University, Japan

Program Committee Members

Masaru Kitsuregawa	University of Tokyo, Japan
Hiroki Arimura	Hokkaido University, Japan
Jean-Francois Boulicaut	INSA Lyon, France
Koichi Hirata	Kyushu Institute of Technology, Japan
Hideki Isozaki	NTT Communication Science Labs., Japan
Kimihito Ito	Research Center for Zoonosis Control, Hokkaido University, Japan
Yoshitaka Kameya	Tokyo Institute of Technology, Japan
Hisashi Kashima	IBM Tokyo Research Lab., Japan
Roni Khardon	Tufts University, USA
Hiroyuki Kitagawa	University of Tsukuba, Japan
YongJin Kwon	Korea Aerospace University, Korea
Seishi Okamoto	Fujitsu Labs. Japan
Jan Poland	ABB Corporate Research Centers, Switzerland
Michele Sebag	LRI, University Paris-Sud, France
Masayuki Takeda	Kyushu University, Japan
Koji Tsuda	Max Planck Institute for Biological Cybernetics., Germany
Shusaku Tsumoto	Shimane University, Japan
Takeaki Uno	National Institute of Informatics, Japan
Takashi Washio	Osaka University, Japan
Akihiro Yamamoto	Kyoto University, Japan
Thomas Zeugmann	Hokkaido University, Japan

Workshop on Data Mining for Decision Making and Risk Management

The organizations and systems in our modern society have become large and complex to provide more advanced services due to the growing variety of social demands. Such organizations and systems are efficient but highly complex, and can cause various unexpected situations. According to this observation, the importance of decision making and risk management of these organizations and systems has been strongly emphasized in recent years. On the other hand, accumulation of a large amount of data on the operations of the organizations and systems has become easier with the introduction of information technology. These data can be used to support decision making or risk management in organizations and systems.

This workshop focused on both data mining and statistical techniques to detect and analyze the risks potentially existing in the organizations and systems and to utilize the risk information for their better management and decision support. The topics covered: (1) data mining and machine learning approaches, (2) statistical approaches, (3) chance discovery, (4) active mining, and (5) application of these techniques to medicine, marketing, security, decision support in business, social activities, human relationships, chemistry and sensor data.

The topics of the workshop in the call for papers included:

- Data mining for decision making/risk management
- Novel statistical approach for decision making/ risk management
- Chance discovery for decision making/risk management
- Active mining for decision making/risk management
- Exploratory data analysis for decision making/risk management
- Machine learning for decision making/risk management
- Other techniques for risk detection, analysis and utilization of risk information
- Applications in the fields (but not limited to) of:
 - Medicine
 - Marketing
 - Security
 - Decision support in business
 - Social activities
 - Human relationships
 - Chemistry
 - Sensor data

Workshop Chairs

Shusaku Tsumoto	Shimane University, Japan
Hiroe Tsubaki	Tsukuba University, Japan
Tzung-Pei Hong	National University of Kaohsiung, Taiwan

Program Committee

Shoji Hirano	Shimane University, Japan
Tony Hu	Drexel University, USA
Genshiro Kitagawa	Institute of Statistical Mathematics, Japan
T.Y. Lin	San Jose State University, USA
Yukio Ohsawa	The University of Tokyo, Japan
Vijay Raghavan	University of Louisiana, USA
Yong Shi	Chinese Academy of Science, China
Andrzej Skowron	Warsaw University, Poland
Guoyin Wang	Southwest Jiaotong University, China
Katsutoshi Yada	Kansai University, Japan
Dirk Van den Poel	Ghent University, Belgium
Ning Zhong	Maebashi Institute of Technology, Japan

Workshop on Interactive Data Mining Overview

The International Workshop on Interactive Data Mining 2008 (IDM08) was held in conjunction with the 12th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2008), Osaka, Japan, May 2008. The workshop aimed at sharing and comparing various fields of interactive data mining research such as interactive information retrieval, information gathering systems, personalization systems, recommendation systems, user interfaces and so on. In summary, the workshop provided a discussion forum for researchers working on interactive data mining where the attendees discussed various aspects of this field. The motivation for the workshop is to discuss recent progress in the research field of interactive data mining, which includes interactions between a human and computers through Web systems, agent systems, sensor systems, or robot systems. Various areas of interactive data mining research have been realized through related technologies including interactive information retrieval, information gathering systems, personalization systems, recommendation systems, user interfaces and so on. Each study and development has been done independently in different research fields such as the information retrieval research field, Web intelligence research field, electric power systems research field, user interfaces research field and so on. However, this situation might discourage us from studying interactive data mining from the unified view of computer-human interaction and making interactive data mining more friendly by applying computational intelligence. Hence, we held the workshop entitled “Interactive Data Mining” in the Pacific-Asia Conference on Knowledge Discovery and Data Mining 2008, to gather together a variety of researchers in diverse fields such as knowledge discovery, information retrieval, Web intelligence, electric power systems, user interfaces and data mining.

The topics of the workshop in the call for papers included, but were not limited to:

- Interactive information retrieval
- Interactive information gathering systems
- Interactive personalization systems
- Interactive Web systems
- Interactive recommendation systems
- Interactive knowledge discovery
- Interactive user interfaces
- Interactive risk mining for electric power systems

All submitted papers were carefully peer reviewed by the Program Chairs. We accepted five papers out of seven submissions. The acceptance rate

is approximately 70%. We would like to thank all the authors who submitted papers to the workshop and participated in the interesting discussions at the workshop.

Program Chairs

Takashi Onoda

Central Research Institute of Electric Power
Industry, Japan

Seiji Yamada

National Institute of Informatics, Japan

Workshop on Interactive Data Mining Overview

The International Workshop on Interactive Data Mining 2008 (IDM08) was held in conjunction with The 12th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2008), Osaka, Japan, May 2008. The workshop aimed at sharing and comparing various interactive data mining researches such as interactive information retrieval, information gathering systems, personalization systems, recommendation systems, user interfaces and so on. In summary, the workshop gave a discussion forum for researchers working on interactive data mining where the attendees discussed various aspects on interactive data mining. The motivation of the workshop is to discuss recent progress in the research field of interactive data mining, which includes interactions between a human and computers through web systems, agent systems, sensor systems, or robot systems. Various interactive data mining researches have been realized through related technologies including interactive information retrieval, information gathering systems, personalization systems, recommendation systems, user interfaces and so on. Each study and development has been done independently in different research fields such as information retrieval research field, web intelligence research field, electric power systems research field, user interfaces research field and so on. However, this situation might discourage us from studying interactive data mining from unified view of computer-human interaction and making interactive data mining more friendly by applying computational intelligence. Hence, at this time, we will hold the workshop entitled "Interactive Data Mining" in the Pacific-Asia Conference on Knowledge Discovery and Data Mining 2008, to gather a variety of researchers in diverse fields like knowledge discovery, information retrieval, web intelligence, electric power systems, user interfaces and data mining.

The topics of the workshop in call for papers included, but are not limited to:

- Interactive Information Retrieval
- Interactive Information Gathering Systems
- Interactive Personalization Systems
- Interactive Web Systems
- Interactive Recommendation Systems
- Interactive Knowledge Discovery
- Interactive User Interfaces
- Interactive Risk Mining for Electric Power Systems, etc

All submitted papers were carefully peer reviewed by program chairs. We accepted 5 papers out of 7 submissions. The acceptance rate is approximately 70%.

We would like to thank all the authors who submitted papers to the workshop and participated in the interesting discussions at the workshop.

Program Chairs

Takashi Onoda	Central Research Institute of Electric Power Industry, JP
Seiji Yamada	National Institute of Informatics, JP

Table of Contents

Workshop of ALSIP 2008

Flexible Framework for Time-Series Pattern Matching over Multi-dimension Data Stream	1
<i>Takuya Kida, Tomoya Saito, and Hiroki Arimura</i>	
An Adaptive Algorithm for Splitting Large Sets of Strings and Its Application to Efficient External Sorting	13
<i>Tatsuya Asai, Seishi Okamoto, and Hiroki Arimura</i>	
Incrementally Mining Recently Repeating Patterns over Data Streams	26
<i>Jia-Ling Koh and Pei-Min Chou</i>	
A Graph-Based Approach for Sentiment Sentence Extraction	38
<i>Kazutaka Shimada, Daigo Hashimoto, and Tsutomu Endo</i>	
Fuzzy Weighted Association Rule Mining with Weighted Support and Confidence Framework	49
<i>Maybin Muyeba, M. Sulaiman Khan, and Frans Coenen</i>	
A Framework for Mining Fuzzy Association Rules from Composite Items	62
<i>Maybin Muyeba, M. Sulaiman Khan, and Frans Coenen</i>	
Mining Mutually Dependent Ordered Subtrees in Tree Databases	75
<i>Tomonobu Ozaki and Takenao Ohkawa</i>	
A Tree Distance Function Based on Multi-sets	87
<i>Arnoldo José Müller-Molina, Kouichi Hirata, and Takeshi Shinohara</i>	
Sibling Distance for Rooted Labeled Trees	99
<i>Taku Aratsu, Kouichi Hirata, and Tetsuji Kuboyama</i>	
Kernel Functions Based on Derivation	111
<i>Koichiro Doi and Akihiro Yamamoto</i>	

Workshop of DMDRM 2008

Dynamic Bayesian Networks for Acquisition Pattern Analysis: A Financial-Services Cross-Sell Application	123
<i>Anita Prinzie and Dirk Van den Poel</i>	

An Automata Based Authorship Identification System	134
<i>Tsau Young Lin and Shangxuan Zhang</i>	
Detection of Risk Factors as Temporal Data Mining	143
<i>Shoji Hirano and Shusaku Tsumoto</i>	
Workshop of IDM 2008	
Two-Phased Active Support Kernel Machine Learning	157
<i>Yasusi Sinohara and Atsuhiko Takasu</i>	
Extracting Topic Maps from Web Pages	169
<i>Motohiro Mase, Seiji Yamada, and Katsumi Nitta</i>	
Interactive Abnormal Condition Sign Discovery for Hydroelectric Power Plants	181
<i>Norihiko Ito, Takashi Onoda, and Hironobu Yamasaki</i>	
Interactive Visualization System for Decision Making Support in Online Shopping	193
<i>Tomoki Kajinami, Takashi Makihara, and Yasufumi Takama</i>	
A Method to Recognize and Count Leaves on the Surface of a River Using User's Knowledge about Color of Leaves	203
<i>Fujio Tsutsumi and Yutaka Tateda</i>	
Author Index	213