

### LNAI Series Editors

Randy Goebel

*University of Alberta, Edmonton, Canada*

Yuzuru Tanaka

*Hokkaido University, Sapporo, Japan*

Wolfgang Wahlster

*DFKI and Saarland University, Saarbrücken, Germany*

### LNAI Founding Series Editor

Joerg Siekmann

*DFKI and Saarland University, Saarbrücken, Germany*

More information about this series at <http://www.springer.com/series/1244>

Huiping Cao · Jinyan Li · Ruili Wang (Eds.)

# Trends and Applications in Knowledge Discovery and Data Mining

PAKDD 2016 Workshops, BDM, MLSDA, PACC, WDMBF  
Auckland, New Zealand, April 19, 2016  
Revised Selected Papers

*Editors*

Huiping Cao  
New Mexico State University  
Las Cruces, NM  
USA

Ruili Wang  
Massey University  
Auckland  
New Zealand

Jinyan Li  
University of Technology  
Sydney, NSW  
Australia

ISSN 0302-9743                      ISSN 1611-3349 (electronic)  
Lecture Notes in Artificial Intelligence  
ISBN 978-3-319-42995-3              ISBN 978-3-319-42996-0 (eBook)  
DOI 10.1007/978-3-319-42996-0

Library of Congress Control Number: 2016944916

LNCS Sublibrary: SL7 – Artificial Intelligence

© Springer International Publishing Switzerland 2016

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature  
The registered company is Springer International Publishing AG Switzerland

# Preface

This edited volume contains selected papers from the four workshops that were held on April 19, 2016, in Auckland, New Zealand. These workshops were run in conjunction with the 20th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2016), a leading international conference in the areas of data mining and knowledge discovery. The four workshops are: Workshop on Biologically Inspired Data Mining Techniques (BDM), Workshop on Machine Learning for Sensory Data Analysis (MLSDA), Workshop on Predictive Analytics for Critical Care (PACC), and Workshop on Data Mining in Business and Finance (WDMBF). The aim of these workshops was to provide forums for discussing research topics related to emerging data mining theories and real-life applications, where knowledge discovery was found to be necessary and/or useful.

The PAKDD 2016 workshops received a total of 38 full-length paper submissions. Each submitted paper was rigorously reviewed by at least two Program Committee members. Although many papers were worthy of publication, only 23 regular papers could be accepted for presentation at the workshops and publication in this volume.

The general quality of submissions was high and the competition was tough. We would like to thank all the authors who submitted their papers on many exciting and important research topics to the PAKDD workshops. We thank the workshop organizers for their tremendous effort and valuable time to make the workshops possible. We also thank all the workshop participants and presenters for attending these workshops. It is our hope that the workshops will provide a lasting platform for disseminating the latest research results and practice of data-mining approaches and applications.

These workshops would not have been possible without the help of many colleagues. We would like to thank the Program Committee members for their invaluable review time and comments. Given the extremely tight review schedule, their effort to complete the review reports before the deadline was greatly appreciated. In addition, we found some reviewers' comments were truly excellent, as good as what is usually found in a survey paper—critical, constructive, and comprehensive. These comments were very helpful for us in selecting the papers.

Thank you all and may the papers collected in the volume inspire your thoughts and research.

May 2016

Huiping Cao  
Jinyan Li  
Ruili Wang

# Organization

## Workshop Co-chairs

Huiping Cao

New Mexico State University, USA

Jinyan Li

University of Technology Sydney, Australia

## BDM Workshop (The 5th PAKDD Workshop on Biologically Inspired Data Mining Techniques)

### Workshop Organizers

Shafiq Alam Burki

University of Auckland, New Zealand

Gillian Dobbie

University of Auckland, New Zealand

### Program Committee

Patricia Riddle

University of Auckland, New Zealand

Kamran Shafi

DSARC, UNSW, Australia

Stephen Chen

York University, Canada

Kouroush Neshatian

University of Canterbury, Christchurch, New Zealand

Ganesh Kumar

Missouri University of Science and Technology, USA

Venayagamoorthy

Yanjun Yan

Western Carolina University, USA

Ming Li

Nanjing University, China

Ismail Khalil

Johannes Kepler University, Austria

David Taniar

Monash University, Australia

Redda Alhaj

University of Calgary, Canada

Lean Yu

Chinese Academy of Sciences (CAS), China

Fatos Xhafa

Universitat Politècnica de Catalunya, Spain

Xiao-Zhi Gao

Aalto University, Finland

Emilio Corchado

University of Burgas, Bulgaria

Michela Antonelli

University of Pisa, Italy

Khalid Saeed

AGH Krakow, Poland

Richi Nayek

QUT, Australia

Saeed u Rehman

Unitec, Institute of Technology Auckland,  
New Zealand

Zawar Shah

Whitireia Community Polytechnic, New Zealand

## **MLSDA Workshop (Machine Learning for Sensory Data Analysis)**

### **Workshop Organizers**

Ashfaqur Rahman	CISRO Australia
Bernhard Pfahringer	The University of Waikato, New Zealand
Jiuyong Li	The University of South Australia
Jeremiah D. Deng	University of Otago, New Zealand

### **Program Committee**

Adnan Al-Anbuky	AUT, New Zealand
Weidong Cai	Sydney, Australia
Rachael Cardell-Oliver	UWA, Australia
Paulo De Souza	CSIRO, Australia
Jeremiah Deng	Otago, New Zealand
Alberto Elfes	CSIRO, Australia
Clinton Fookes	QUT, Australia
Jia Hu	Hope, UK
Yuan Jiang	Nanjing, China
Eamonn Keogh	UCR, USA
Irena Koprinska	Sydney, Australia
Daniel Lai	Victoria, Australia
Ickjai Lee	James Cook, Australia
Ivan Lee	South Australia, Australia
Christopher Leckie	Melbourne, Australia
Jiuyong Li	South Australia, Australia
Craig Lindley	CSIRO, Australia
Ann Nowé	VUB, Belgium
Mariusz Nowostawski	Gjøvik University College, Norway
Paul Pang	Unitec, New Zealand
Matthew Parry	Otago, New Zealand
Yonghong Peng	Bradford, UK
Bernhard Pfahringer	Waikato, New Zealand
Martin Purvis	Otago, New Zealand
Ashfaqur Rahman	CISRO, Australia
Daniel Smith	Data61, CSIRO, Australia
James S.C. Tan	(UniSIM), Singapore
Duc A. Tran	Massachusetts, USA
Zhiyong Wang	Sydney, Australia
Brendon Woodford	Otago, New Zealand
Haibo Zhang	Otago, New Zealand
Jun Zhang	Sun Yat-sen, China
Zhi-hua Zhou	Nanjing, China
Xingquan Zhu	UTS, Australia

## **PACC Workshop (The First International Workshop on Predictive Analytics for Critical Care)**

### **Workshop Organizers**

Vaibhav Rajan	Xerox Research Centre, India
Geetha Manjunath	Xerox Research Centre, India

### **Program Committee**

Chiranjib Bhattacharyya	Indian Institute of Science (IISc), India
Sakyajit Bhattacharya	Xerox Research Centre India (XRCI), India
Inderjit Dhillon	University of Texas, Austin, USA
Shuai Huang	University of Washington, USA
Gang Luo	University of Utah, USA
Sriganesh Madhvanath	Palo Alto Research Centre (PARC), USA
Chandan Reddy	Wayne State University, USA
S. Sadagopan	IIIT-Bangalore, India
Jing Zhou	Palo Alto Research Centre (PARC), USA

## **WDMBF Workshop (Workshop on Data Mining in Business and Finance)**

### **Workshop Organizers**

Ling Liu	Southwestern University of Finance and Economics, China
Qing Li	Southwestern University of Finance and Economics, China
Yuanzhu Chen	Memorial University of Newfoundland, Canada
Weidong Huo	Southwestern University of Finance and Economics, China

### **Program Committee**

Ting Hu	Memorial University of Newfoundland, Canada
Sung Hyon Myaeng	KAIST, Korea
Keng Siau	Missouri University of Science and Technology, USA
Sa-kwang Song	KISTI, Korea
Shengyan Sun	IBM, China
Zhangxi Lin	Texas Tech University, USA
Juchen Yang	Tianjin Tech University, China
Weiwei Yuan	Harbin Engineering University, China



# Contents

## Biologically Inspired Data Mining Techniques (BDM)

Towards a New Evolutionary Subsampling Technique for Heuristic Optimisation of Load Disaggregators . . . . .	3
<i>Michael Mayo and Sara Omranian</i>	
Neural Choice by Elimination via Highway Networks . . . . .	15
<i>Truyen Tran, Dinh Phung, and Svetha Venkatesh</i>	
Attribute Selection and Classification of Prostate Cancer Gene Expression Data Using Artificial Neural Networks. . . . .	26
<i>Sreenivas Sremath Tirumala and A. Narayanan</i>	
An Improved Self-Structuring Neural Network . . . . .	35
<i>Rami M. Mohammad, Fadi Thabtah, and Lee McCluskey</i>	
Imbalanced ELM Based on Normal Density Estimation for Binary-Class Classification . . . . .	48
<i>Yulin He, Rana Aamir Raza Ashfaq, Joshua Zhexue Huang, and Xizhao Wang</i>	
Multiple Seeds Based Evolutionary Algorithm for Mining Boolean Association Rules . . . . .	61
<i>Mir Md. Jahangir Kabir, Shuxiang Xu, Byeong Ho Kang, and Zongyuan Zhao</i>	

## Machine Learning for Sensory Data Analysis (MLSDA)

Predicting Phone Usage Behaviors with Sensory Data Using a Hierarchical Generative Model . . . . .	75
<i>Chuankai An and Dan Rockmore</i>	
Comparative Evaluation of Action Recognition Methods via Riemannian Manifolds, Fisher Vectors and GMMs: Ideal and Challenging Conditions . . .	88
<i>Johanna Carvajal, Arnold Willem, Chris McCool, Brian Lovell, and Conrad Sanderson</i>	
Rigidly Self-Expressive Sparse Subspace Clustering . . . . .	101
<i>Linbo Qiao, Bofeng Zhang, Yipin Sun, and Jinshu Su</i>	

Joint Recognition and Segmentation of Actions via Probabilistic Integration of Spatio-Temporal Fisher Vectors . . . . .	115
<i>Johanna Carvajal, Chris McCool, Brian Lovell, and Conrad Sanderson</i>	
Learning Multi-faceted Activities from Heterogeneous Data with the Product Space Hierarchical Dirichlet Processes . . . . .	128
<i>Thanh-Binh Nguyen, Vu Nguyen, Svetha Venkatesh, and Dinh Phung</i>	
Phishing Detection on Twitter Streams. . . . .	141
<i>Se Yeong Jeong, Yun Sing Koh, and Gillian Dobbie</i>	
Image Segmentation with Superpixel Based Covariance Descriptor . . . . .	154
<i>Xianbin Gu and Martin Purvis</i>	

**Predictive Analytics for Critical Care (PACC)**

Normalized Cross-Match: Pattern Discovery Algorithm from Biofeedback Signals . . . . .	169
<i>Xueyuan Gong, Simon Fong, Yain-Whar Si, Robert P. Biuk-Aghai, Raymond K. Wong, and Athanasios V. Vasilakos</i>	
Event Prediction in Healthcare Analytics: Beyond Prediction Accuracy . . . . .	181
<i>Lina Fu, Faming Li, Jing Zhou, Xuejin Wen, Jinhui Yao, and Michael Shepherd</i>	
Clinical Decision Support for Stroke Using Multi-view Learning Based Models for NIHSS Scores . . . . .	190
<i>Vaibhav Rajan, Sakyajit Bhattacharya, Ranjan Shetty, Amith Sitaram, and G. Vivek</i>	

**Data Mining in Business and Finance (WDMBF)**

A Music Recommendation System Based on Acoustic Features and User Personalities . . . . .	203
<i>Rui Cheng and Boyang Tang</i>	
A Social Spam Detection Framework via Semi-supervised Learning . . . . .	214
<i>Xianchao Zhang, Haijun Bai, and Wenxin Liang</i>	
A Hierarchical Beta Process Approach for Financial Time Series Trend Prediction. . . . .	227
<i>Mojgan Ghanavati, Raymond K. Wong, Fang Chen, Yang Wang, and Joe Lee</i>	
Efficient Iris Image Segmentation for ATM Based Approach Through Fuzzy Entropy and Graph Cut . . . . .	238
<i>Shibai Yin, Yibin Wang, and Tao Wang</i>	

Matching Product Offers of E-Shops . . . . .	248
<i>Andrea Horch, Holger Kett, and Anette Weisbecker</i>	
Keystroke Biometric Recognition on Chinese Long Text Input . . . . .	260
<i>Xiaodong Li and Jiafen Liu</i>	
Recommendation Algorithm Design in a Land Exchange Platform . . . . .	272
<i>Xubin Luo and Jiang Duan</i>	
Erratum to: Normalized Cross-Match: Pattern Discovery Algorithm from Biofeedback Signals . . . . .	E1
<i>Xueyuan Gong, Simon Fong, Yain-Whar Si, Robert P. Biuk-Aghai, Raymond K. Wong, and Athanasios V. Vasilakos</i>	
<b>Author Index</b> . . . . .	281