

Communications in Computer and Information Science

996

Commenced Publication in 2007

Founding and Former Series Editors:

Phoebe Chen, Alfredo Cuzzocrea, Xiaoyong Du, Orhun Kara, Ting Liu,
Dominik Ślęzak, and Xiaokang Yang

Editorial Board

Simone Diniz Junqueira Barbosa

*Pontifical Catholic University of Rio de Janeiro (PUC-Rio),
Rio de Janeiro, Brazil*

Joaquim Filipe

Polytechnic Institute of Setúbal, Setúbal, Portugal

Ashish Ghosh

Indian Statistical Institute, Kolkata, India

Igor Kotenko

*St. Petersburg Institute for Informatics and Automation of the Russian
Academy of Sciences, St. Petersburg, Russia*

Krishna M. Sivalingam

Indian Institute of Technology Madras, Chennai, India

Takashi Washio

Osaka University, Osaka, Japan

Junsong Yuan

University at Buffalo, The State University of New York, Buffalo, USA

Lizhu Zhou

Tsinghua University, Beijing, China

More information about this series at <http://www.springer.com/series/7899>


Rafiqul Islam · Yun Sing Koh
Yanchang Zhao · Graco Warwick
David Stirling · Chang-Tsun Li
Zahidul Islam (Eds.)

Data Mining

16th Australasian Conference, AusDM 2018
Bahrurst, NSW, Australia, November 28–30, 2018
Revised Selected Papers


Editors

Rafiqul Islam 
School of Computing and Mathematics
Charles Sturt University
Albury, NSW, Australia

Yun Sing Koh 
University of Auckland
Auckland, New Zealand

Yanchang Zhao
CSIRO Scientific Computing
Canberra, Australia

Graco Warwick
Data Science and Engineering
Australian Taxation Office
Canberra, Australia

David Stirling 
Department of Information Technology
University of Wollongong
Wollongong, NSW, Australia

Chang-Tsun Li
School of Computing and Mathematics
Charles Sturt University
Wagga Wagga, Australia

Zahidul Islam 
School of Computing and Mathematics
Charles Sturt University
Bathurst, Australia

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-981-13-6660-4 ISBN 978-981-13-6661-1 (eBook)
<https://doi.org/10.1007/978-981-13-6661-1>

Library of Congress Control Number: 2019931946

© Springer Nature Singapore Pte Ltd. 2019

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. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

It is our great pleasure to present the proceedings of the 16th Australasian Data Mining Conference (AusDM 2018) held at Charles Sturt University, Bathurst, during November 28–30, 2018. The Australasian Data Mining (AusDM) Conference series first started in 2002 as a workshop that was initiated by Dr. Simeon Simoff (then Associate Professor, University of Technology, Sydney), Dr. Graham Williams (then Principal Data Miner, Australian Taxation Office, and Adjunct Professor, University of Canberra), and Dr. Markus Hegland (Australian National University). The conference series has grown significantly since it first started and is continuing to grow under the leadership of Professors Simoff and Williams as the chairs of the Steering Committee.

The Australasian Data Mining Conference brings together exciting and novel research contributions and their applications for solving real-life problems through intelligent data analysis of (usually large) data sets. The conference focuses on novel contributions on data collection, cleansing, pre-processing, knowledge discovery, making sense of data, knowledge presentation, future prediction, applications of data mining, and various issues related to data mining such as privacy and security. It welcomes researchers and industry practitioners to come together to discuss current problems and share ideas on the possible ways data scientists can contribute to solving the problems through amazing data mining algorithms and their applications. The conference series also gives a wonderful opportunity for postgraduate students in data mining to present their research innovations and network with other students, leading researchers, and industry experts. The conference series has been serving the data mining community with all these aims and achievements for the past 16 years without a miss.

In the past few years, AusDM operated in dual tracks covering both research and application sides of data science. This year AusDM further expanded with a number of main tracks and special tracks. The three main tracks were Research, Application, and Industry Showcase. The Research track also presented four special tracks: Image Data Mining, Identification Through Data Mining, Mobile and Sensor Network Data Mining, and Statistics in Data Science. The Research track and its four special tracks were meant for academic submissions whereas the Application track and Industry Showcase track were for industry submissions.

All academic submissions were reviewed following the same review process to ensure a fair review and all accepted papers were published in the same proceedings, regardless of the categories and tracks. Papers accepted from a special track were grouped together in the same session of the conference.

After partnering with various conferences for the past few years through co-location, this year AusDM was again located on its own as an independent conference on data mining. The conference maintained a well-designed website from almost a year before the actual conference starting date with clear instructions and information on the tracks, submission process, proceedings, author guidelines, keynote speakers, and tutorials.

The inclusion of a new proceedings chair role in the Organizing Committee emphasizes AusDM's commitment to timely publication of the accepted papers. The accepted papers were published by Springer in their CCIS series. The inclusion of several special tracks showed AusDM's impetus to reach out to research communities that are closely related to data mining and bring all parties on the common AusDM platform. The conference was well announced and publicized through various forums including social network sites such as Facebook, LinkedIn, and Twitter. Members of the Organizing Committee and the Steering Committee contributed greatly to make the conference successful. All these efforts have resulted in a significant increase in the number of submissions to AusDM 2018.

This year AusDM received 98 submissions. Of these, 18 submissions were disqualified for not meeting submission requirements and the remaining 80 submissions went through a double-blind review process. Academic submissions and Application track papers received at least three peer review reports and Industry Showcase track papers received two peer review reports. Additional reviewers were considered for a clear review outcome, if review comments from the initial three reviewers were unclear. As a result, some papers received up to six review reports.

Out of these 80 submissions, a total of 30 papers were finally accepted for publication. The overall acceptance rate for this year was 37.5%. Out of 58 academic submissions, 22 papers (i.e., 38%) were accepted for publication. Out of 18 submissions in the Application track, seven papers (i.e., 39%) were accepted for publication. Out of four submissions in the Industry Showcase track, one paper (i.e., 25%) was accepted for publication. All papers from the Industry Showcase track were invited for oral presentation. This year, authors from 19 different countries submitted their papers. The top five countries, in terms of the number of authors who submitted papers to AusDM 2018, were Australia (122 authors), China (17 authors), USA (11 authors), New Zealand (9 authors), and India and the UK (7 authors each).

The AusDM 2018 Organizing Committee would like to give their special thanks to Professors Geoffrey Webb, Junbin Gao, Graham Williams, and Yanchang Zhao for kindly accepting the invitation to give keynote speeches and tutorials. The committee would also like to give their sincere thanks to Charles Sturt University for providing administrative support (including the employment of a new temporary admin officer) and the conference venue that contributed to a successful event. The committee would also like to thank Springer CCIS and the Editorial Board for their acceptance to publish AusDM 2018 papers. This will give excellent exposure of the papers accepted for publication. We would also like to give our heartfelt thanks to all student volunteers at Charles Sturt University, who did a tremendous job. Last but not least, we would like to give our sincere thanks to all delegates for attending the conference this year and we hope you enjoyed AusDM 2018.

Organization

Conference Chairs

Md Zahidul Islam Charles Sturt University, Australia
Chang-Tsun Li Charles Sturt University, Australia

Program Chairs (Research Track)

Md. Rafiqul Islam Charles Sturt University, Australia
Yun Sing Koh University of Auckland, New Zealand

Program Chairs (Application Track)

Yanchang Zhao CSIRO, Sydney, Australia
David Stirling University of Wollongong, Australia

Program Chair (Industry Showcase)

Warwick Graco Australian Taxation Office, Australia

Proceedings Chair

Kok-Leong Ong La Trobe University, Australia

Tutorial Chairs

Jixue Liu University of South Australia, Australia
Yee Ling Boo RMIT University, Australia

Web Chairs

Michael Furner Charles Sturt University, Australia
Ling Chen University of Technology Sydney, Australia

Competition Chairs

Tony Nolan Australian Taxation Office, Australia
Paul Kennedy University of Technology Sydney, Australia

Publicity Chairs

Ji Zhang University of Southern Queensland, Australia
Ashad Kabir Charles Sturt University, Australia

Special Track Chairs (Image Data Mining)

Lihong Zheng Charles Sturt University, Australia
Xufeng Lin Charles Sturt University, Australia

Special Track Chairs (Statistics in Data Science)

Azizur Rahman Charles Sturt University, Australia
Ryan Ip Charles Sturt University, Australia
Hien Nguyen La Trobe University, Melbourne, Australia

Special Track Chairs (Sensor Data Mining)

Quazi Mamun Charles Sturt University, Australia
Sabih Rehman Charles Sturt University, Australia

Special Track Chairs (Identification Through Data Mining)

Xufeng Lin Charles Sturt University, Australia
Xingjie Wei University of Bath, UK

Steering Committee Chairs

Simeon Simoff University of Western Sydney, Australia
Graham Williams Microsoft

Steering Committee Members

Peter Christen The Australian National University, Canberra, Australia
Ling Chen University of Technology, Sydney, Australia
Zahid Islam Charles Sturt University, Australia
Paul Kennedy University of Technology, Sydney, Australia
Jiuyong (John) Li University of South Australia, Adelaide, Australia
Richi Nayak Queensland University of Technology, Brisbane, Australia
Kok-Leong Ong La Trobe University, Melbourne, Australia
Dharmendra Sharma University of Canberra, Australia
Glenn Stone Western Sydney University, Australia
Andrew Stranieri Federation University Australia, Mount Helen, Australia
Yanchang Zhao CSIRO, Sydney, Australia

Honorary Advisors

John Roddick Flinders University, Australia
 Geoff Webb Monash University, Australia

Program Committee

Research Track

Cheng Li	Deakin University, Australia
Yonghua Cen	Nanjing University of Science and Technology
Kewen Liao	Swinburne University of Technology, Australia
Zheng Lihong	Charles Sturt University, Australia
Ailian Jiang	Taiyuan University of Technology, China
Xuan-Hong Dang	IBM T.J.Watson
Turki Turki	King Abdulaziz University, Saudi Arabia
Philippe Fournier-Viger	Harbin Institute of Technology (Shenzhen), China
Pradnya Kulkarni	MIT World Peace University, India
Wei Shao	RMIT University, Australia
Jianzhong Qi	The University of Melbourne, Australia
Chayapol Moemeng	Assumption University
Imdadullah Khan	Lahore University of Management Sciences, India
Md Anisur Rahman	Charles Sturt University, Australia
Ashad Kabir	Charles Sturt University, Australia
Azizur Rahman	Charles Sturt University, Australia
Richard Gao	Commonwealth Australia, Australia
Saiful Islam	Griffith University, Australia
Md Nasim Adnan	Charles Sturt University, Australia
Xiaohui Tao	University of Southern Queensland, Australia
Ahmed Mohiuddin	Canberra Institute of Technology, Australia
David Huang	The University of Auckland, New Zealand
Hien Nguyen	La Trobe University, Australia
Adil Bagirov	University of Ballarat, Australia
R. Uday Kiran	The University of Tokyo, Japan
Yongsheng Gao	Griffith University, Australia
Lei Wang	University of Wollongong, Australia
Alan Wee-Chung Liew	Griffith University, Australia
Veelasha Moonsamy	Utrecht University
Flora Dilys Salim	RMIT University, Australia
Joy Wang	Union College
Rui Zhang	The University of Melbourne, Australia
Louie Qin	University of Huddersfield
Quang Vinh Nguyen	Western Sydney University, Australia
Mendis Champake	Charles Sturt University, Australia

Sitalakshmi Venkatraman	Melbourne Polytechnic, Australia
Mohammad Saiedur Rahaman	RMIT University, Australia
Xiangjun Dong	School of Information, Qilu University of Technology, China
Dhananjay Thiruvady	Monash University, Australia
Kiheiji Nishida	Hyogo University of Health Sciences, General Education Center, Japan
Md Geaur Rahman	Bangladesh Agricultural University
Zahirul Hoque	UAE University
Fulong Chen	Anhui Normal University, China
Guandong Xu	University of Technology, Sydney, Australia
Xingjie Wei	University of Cambridge, UK
Dharmendra Sharma	University of Canberra, Australia
Truyen Tran	Deakin University, Australia
Qinxue Meng	University of Technology, Sydney, Australia
Gang Li	Deakin University, Australia
Dinusha Vatsalan	CSIRO, Australian National University, Australia
Grace Rumantir	Monash University, Australia
Ji Zhang	The University of Southern Queensland, Australia
Muhammad Marwan Muhammad Fuad	Technical University of Denmark
Kok-Leong Ong	La Trobe University, Australia
Chang-Tsun Li	Charles Sturt University, Australia
Paul Grant	Charles Sturt University, Australia
Nectarios Costadopoulos	Charles Sturt University, Australia
Darren Yates	Charles Sturt University, Australia
Khondker Jahid Reza	Charles Sturt University, Australia
Diana Benavides Prado	University of Auckland, New Zealand
Robert Anderson	University of Auckland, New Zealand
Kylie Chen	University of Auckland, New Zealand

Application Track

Rohan Baxter	Australian Tax Office, Australia
Nathan Brewer	Department of Human Services, Australia
Neil Brittliff	University of Canberra, Australia
Adriel Cheng	Defence Science and Technology Group, Australia
Hoa Dam	University of Wollongong, Australia
Klaus Felsche	C21 Directions, Australia
Markus Hagenbuchner	University of Wollongong, Australia
Edward Kang	Australian Passport Office, Australia

Stefan Keller	Teradata, Australia
Sebastian Kurscheid	Australian National University, Australia
Fangfang Li	oOh! Media, Australia
Jin Li	Geoscience Australia, Australia
Chao Luo	Department of Health, Australia
Atif Mahmood	Department of Home Affairs, Australia
Kee Siong Ng	Australian Government, Australia
Tom Osborn	Epifini.ai, Australia
Martin Rennhackkamp	PBT Group Australia, Australia
Goce Ristanoski	CSIRO Data61, Australia
Chao Sun	The University of Sydney, Australia
Tilly Tan	PwC Australia, Australia
Kerry Taylor	Australian National University and University of Surrey, Australia
Jie Yang	University of Wollongong, Australia
Ting Yu	Commonwealth Bank of Australia, Australia

Contents

Classification Tasks

Misogynistic Tweet Detection: Modelling CNN with Small Datasets	3
<i>Md Abul Bashar, Richi Nayak, Nicolas Suzor, and Bridget Weir</i>	
Multiple Support Vector Machines for Binary Text Classification Based on Sliding Window Technique	17
<i>Aisha Rashed Albqmi, Yuefeng Li, and Yue Xu</i>	
High-Dimensional Limited-Sample Biomedical Data Classification Using Variational Autoencoder.	30
<i>Mohammad Sultan Mahmud, Xianghua Fu, Joshua Zhexue Huang, and Md. Abdul Masud</i>	
SPAARC: A Fast Decision Tree Algorithm	43
<i>Darren Yates, Md Zahidul Islam, and Junbin Gao</i>	
LCS Based Diversity Maintenance in Adaptive Genetic Algorithms.	56
<i>Ryoma Ohira, Md. Saiful Islam, Jun Jo, and Bela Stantic</i>	
Categorical Features Transformation with Compact One-Hot Encoder for Fraud Detection in Distributed Environment	69
<i>Ikram Ul Haq, Iqbal Gondal, Peter Vamplew, and Simon Brown</i>	

Transport, Environment and Energy

Combining Machine Learning and Statistical Disclosure Control to Promote Open Data	83
<i>Nasca Peng</i>	
Predicting Air Quality from Low-Cost Sensor Measurements	94
<i>Hamish Huggard, Yun Sing Koh, Patricia Riddle, and Gustavo Olivares</i>	
An Approach to Compress and Represents Time Series Data and Its Application in Electric Power Utilities.	107
<i>Chee Keong Wee and Richi Nayak</i>	

Applied Data Mining

Applying Data Mining Methods to Generate Formative Feedback in Team Projects.	123
<i>Rimmel Nadeem, X. Rosalind Wang, and Caslon Chua</i>	

A Hybrid Missing Data Imputation Method for Constructing City Mobility Indices 135
Sanaz Nikfalazar, Chung-Hsing Yeh, Susan Bedingfield, and Hadi Akbarzadeh Khorshidi

A Novel Learning-to-Rank Method for Automated Camera Movement Control in E-Sports Spectating 149
Hendi Lie, Darren Lukas, Jonathan Liebig, and Richi Nayak

Homogeneous Feature Transfer and Heterogeneous Location Fine-Tuning for Cross-City Property Appraisal Framework. 161
Yihan Guo, Shan Lin, Xiao Ma, Jay Bal, and Chang-tsun Li

Statistical Models of Dengue Fever 175
Hamilton Link, Samuel N. Richter, Vitus J. Leung, Randy C. Brost, Cynthia A. Phillips, and Andrea Staid

Privacy and Clustering

Reference Values Based Hardening for Bloom Filters Based Privacy-Preserving Record Linkage 189
Sirintra Vaiwsri, Thilina Ranbaduge, and Peter Christen

Canopy-Based Private Blocking 203
Yanfeng Shu, Stephen Hardy, and Brian Thorne

Community Aware Personalized Hashtag Recommendation in Social Networks 216
Areej Alsini, Amitava Datta, Du Q. Huynh, and Jianxin Li

Behavioral Analysis of Users for Spammer Detection in a Multiplex Social Network. 228
Tahereh Pourhabibi, Yee Ling Boo, Kok-Leong Ong, Booi Kam, and Xiuzhen Zhang

Taxonomy-Augmented Features for Document Clustering 241
Sattar Seifollahi, Massimo Piccardi, Ehsan Zare Borzeshi, and Bernie Kruger

A Modularity-Based Measure for Cluster Selection from Clustering Hierarchies 253
Francisco de Assis Rodrigues dos Anjos, Jadson Castro Gertrudes, Jörg Sander, and Ricardo J. G. B. Campello

Statistics in Data Science

Positive Data Kernel Density Estimation via the LogKDE Package for R. . . . 269
Hien D. Nguyen, Andrew T. Jones, and Geoffrey J. McLachlan

The Role of Statistics Education in the Big Data Era. 281
Ryan H. L. Ip

Hierarchical Word Mover Distance for Collaboration
 Recommender System 289
Chao Sun, King Tao Jason Ng, Philip Henville, and Roman Marchant

Health, Software and Smart Phone

An Alternating Least Square Based Algorithm for Predicting Patient
 Survivability. 305
Qiming Hu, Jie Yang, Khin Than Win, and Xufeng Huang

Interpreting Intermittent Bugs in Mozilla Applications Using
 Change Angle. 318
David Tse Jung Huang, Yun Sing Koh, and Gillian Dobbie

Implementation and Performance Analysis of Data-Mining Classification
 Algorithms on Smartphones 331
Darren Yates, Md. Zahidul Islam, and Junbin Gao

Image Data Mining

Retinal Blood Vessels Extraction of Challenging Images 347
*Toufique Ahmed Soomro, Junbin Gao, Zheng Lihong, Ahmed J. Afifi,
 Shafiullah Soomro, and Manoranjan Paul*

Sequential Deep Learning for Action Recognition with Synthetic
 Multi-view Data from Depth Maps 360
Bin Liang, Lihong Zheng, and Xinying Li

Provenance Analysis for Instagram Photos 372
Yijun Quan, Xufeng Lin, and Chang-Tsun Li

Industry Showcase

How Learning Analytics Becomes a Bridge for Non-expert Data Miners:
 Impact on Higher Education Online Teaching. 387
Katherine Herbert and Ian Holder

Author Index 397