

Communications in Computer and Information Science

1127

Commenced Publication in 2007

Founding and Former Series Editors:

Phoebe Chen, Alfredo Cuzzocrea, Xiaoyong Du, Orhun Kara, Ting Liu,
Krishna M. Sivalingam, Dominik Ślęzak, Takashi Washio, Xiaokang Yang,
and Junsong Yuan

Editorial Board Members

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*

Lizhu Zhou

Tsinghua University, Beijing, China

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

Thuc D. Le · Kok-Leong Ong · Yanchang Zhao ·
Warren H. Jin · Sebastien Wong · Lin Liu ·
Graham Williams (Eds.)

Data Mining

17th Australasian Conference, AusDM 2019
Adelaide, SA, Australia, December 2–5, 2019
Proceedings

Editors

Thuc D. Le
School of Information Technology
and Mathematical Sciences
University of South Australia
Adelaide, Australia

Yanchang Zhao
CSIRO Scientific Computing
Canberra, Australia

Sebastien Wong
Consilium Technology
Adelaide, Australia

Graham Williams
Microsoft Proprietary Limited
Singapore, Singapore

Kok-Leong Ong
La Trobe University
Melbourne, Australia

Warren H. Jin
CSIRO Scientific Computing
Canberra, Australia

Lin Liu
School of Information Technology
and Mathematical Sciences
University of South Australia
Adelaide, Australia

ISSN 1865-0929

ISSN 1865-0937 (electronic)

Communications in Computer and Information Science

ISBN 978-981-15-1698-6

ISBN 978-981-15-1699-3 (eBook)

<https://doi.org/10.1007/978-981-15-1699-3>

© 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, expressed 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 17th Australasian Data Mining Conference (AusDM 2019) held at the University of South Australia, Adelaide, during December 2–5, 2019.

The AusDM conference series first started in 2002 as a workshop initiated by Prof. Simeon Simoff, Dr. Graham Williams, and Prof. Markus Hegland. Over the years, AusDM has established itself as the premier Australasian meeting for both practitioners and researchers in data mining. AusDM is devoted to the art and science of intelligent analysis of (usually big) data sets for meaningful (and previously unknown) insights. Since AusDM 2002, the conference series has showcased research in data mining, providing a forum for presenting and discussing the latest research and developments.

Built on this tradition, AusDM 2019 has successfully facilitated the cross-disciplinary exchange of ideas, experiences, and potential research directions, and pushed forward the frontiers of data mining in academia, government, and industry.

AusDM 2019 received altogether 54 valid submissions, with authors from 16 different countries. The top 5 countries, in terms of the number of authors who submitted papers to AusDM 2019, were Australia (71 authors), China (18 authors), India (14 authors), New Zealand (11 authors), and Germany (5 authors). All submissions went through the double-blind review process, and each paper received at least three peer review reports. Additional reviewers were considered for a clear review outcome, if review comments from the initial three reviewers were inconclusive.

Out of these 54 submissions, a total of 20 papers were finally accepted for publication. The overall acceptance rate for AusDM 2019 was 37%. Out of the 34 Research Track submissions, 11 papers (i.e. 32%) were accepted for publication. Out of the 17 submissions in the Application Track, 8 papers (i.e. 47%) were accepted for publication. Out of the 3 submissions in the Industry Showcase Track, 1 paper (i.e. 33%) was accepted for publication. All the papers from the Industry Showcase Track were invited for oral presentation.

The AusDM 2019 Organizing Committee would like to give their special thanks to Prof. Albert Bifet, Prof. Anton van den Hengel, Dr. Dale Lambert, Prof. Kate Smith-Miles, and Prof. Xin Yao for kindly accepting the invitation to give keynote speeches, and to Dr. Mingming Gong, Dr. Sarah Erfani, Dr. Xingjun Ma, and Ehsan Abbasnejad for organizing and presenting the tutorials. The committee would also like to give their sincere thanks to the University of South Australia for providing admin support and the conference venue ensuring the organization of a successful event. The committee would also like to thank Springer CCIS and the Editorial Board for their acceptance to publish AusDM 2019 papers. This will give excellent exposure of the papers accepted for publication. We would also like to give our heartfelt thanks to all student and staff volunteers at the University of South Australia 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 2019.

December 2019

Thuc D. Le
Yanchang Zhao
Sebastien Wong
Warren H. Jin
Kok-Leong Ong
Lin Liu
Graham Williams

Organization

Organization Committee

Conference Chairs

Lin Liu	University of South Australia, Australia
Graham Williams	Microsoft, Singapore

Program Chairs

Thuc Le	University of South Australia, Australia
Yanchang Zhao	Data61, CSIRO, Australia
Warren Jin	Data61, CSIRO, Australia
Sebastien Wong	Consilium Technology, Australia

Proceedings Chair

Kok-Leong Ong	La Trobe University, Australia
---------------	--------------------------------

Publicity Chair

Yee Ling Boo	RMIT University, Australia
--------------	----------------------------

Organizing Chairs

Wolfgang Mayer	University of South Australia, Australia
Cristina Garcia	University of South Australia, Australia

Sponsorship Chair

Michael Bewong	University of South Australia, Australia
----------------	--

Web Master

Vu Viet Hoang Pham	University of South Australia, Australia
--------------------	--

Steering Committee Chairs

Simeon Simoff	University of Western Sydney, Australia
Graham Williams	Microsoft, Australia

Steering Committee Members

Peter Christen	The Australian National University, Australia
Ling Chen	University of Technology Sydney, Australia
Zahid Islam	Charles Sturt University, Australia

Paul Kennedy	University of Technology Sydney, Australia
Yun Sing Koh	The University of Auckland, New Zealand
Jiuyong (John) Li	University of South Australia, Australia
Richi Nayak	Queensland University of Technology, Australia
Kok-Leong Ong	La Trobe University, Australia
Dharmendra Sharma	University of Canberra, Australia
Glenn Stone	Western Sydney University, Australia
Yanchang Zhao	Data61, CSIRO, Australia

Program Committee

Research Track

Xuan-Hong Dang	IBM T.J. Watson, USA
Philippe Fournier-Viger	Harbin Institute of Technology, China
Ashad Kabir	Charles Sturt University, Australia
Yun Sing Koh	The University of Auckland, New Zealand
Gang Li	Deakin University, Australia
Cheng Li	Deakin University, Australia
Kewen Liao	Charles Darwin University, Australia
Brad Malin	Vanderbilt University, USA
Wolfgang Mayer	University of South Australia, Australia
Veelasha Moonsamy	Deakin University, Australia
Muhammad Marwan	Coventry University, UK
Muhammad Fuad	
Quang Vinh Nguyen	Western Sydney University, Australia
Hien Nguyen	La Trobe University, Australia
Xuan-Hoai Nguyen	AI Academy, Vietnam
Dang Nguyen	Deakin University, Australia
Ninh Pham	The University of Auckland, New Zealand
Jianzhong Qi	The University of Melbourne, Australia
Yongrui Qin	University of Huddersfield, UK
Mohammad Saiedur	RMIT University, Australia
Rahaman	
Md Anisur Rahman	Charles Sturt University, Australia
Md Geaur Rahman	Charles Sturt University, Australia
Azizur Rahman	Charles Sturt University, Australia
Jia Rong	Victoria University, Australia
Dharmendra Sharma	University of Canberra, Australia
Xiaohui Tao	University of Southern Queensland, Australia
Dhananjay Thiruvady	Monash University, Australia
Truyen Tran	Deakin University, Australia
Sitalakshmi Venkatraman	Melbourne Polytechnic, Australia
Bay Vo	HUTECH, Vietnam
Kui Yu	Hefei University of Technology, China
Rui Zhang	The University of Melbourne, Australia

Application Track

Hadi Akbarzadeh Khorshidi	The University of Melbourne, Australia
Nathan Brewer	Department of Human Services, Australia
Yonghua Cen	Nanjing University of Science and Technology, China
Adriel Cheng	Defence Science and Technology Group, Australia
Lianhua Chi	La Trobe University, Australia
Yingsong Hu	Department of Social Services, Australia
Ashad Kabir	Charles Sturt University, Australia
Susie Kluth	Department of Social Services, Australia
Dipankar Kundu	Department of Agriculture and Water Resources, Australia
Sherry Li	Australian Trade and Investment Commission, Australian
Jin Li	Geoscience Australia, Australia
Bin Liang	University of Technology Sydney, Australia
Chao Luo	Department of Health, Australia
Simona Adriana Mihaita	University of Technology Sydney, Australia
Khoa Nguyen	CSIRO, Australia
Shirui Pan	University of Technology Sydney, Australia
Clifton Phua	DataRobot, Singapore
Munir Shah	AgResearch, New Zealand
Yanfeng Shu	CSIRO, Australia
Meina Song	Beijing University of Posts and Telecommunications, China
Ronnie Taib	CSIRO, Australia
Dinusha Vatsalan	CSIRO, Australia
Stephen Wan	CSIRO, Australia
Hongzhi Yin	The University of Queensland, Australia
Huaifeng Zhang	Department of Human Services, Australia

Industry Showcase Track

Rohan Baxter	Australian Taxation Office, Australia
Richard Gao	Department of Health, Australia
Warwick Graco	Australian Taxation Office, Australia
Marcus Suresh	Department of Industry, Australia

Additional Reviewers

Bayu Distiawan
Steven Edwards
Atabak Elmi
Warwick Graco
Haripriya Harikumar
Jiayuan He
Peter Hough
Xinting Huang
Yinhao Jiang
Xiaomei Li
Zhaolong Ling
Guanli Liu
Zhigang Lu

Federico Montori
Do-Van Nguyen
Andrew Perrykkad
Su, Yixin Su
Kaibing Wang
Qinyong Wang
Xiaojie Wang
Yuandong Wang
Khin Nandar Win
Theodor Wyeld
Shuai Yang
Yihong Zhang

Contents

Research Track

Improving Clustering via a Fine-Grained Parallel Genetic Algorithm with Information Sharing	3
<i>Storm Bartlett and Md Zahidul Islam</i>	
A Learning Approach for Ill-Posed Optimisation Problems.	16
<i>Jörg Frochte and Stephen Marsland</i>	
Topic Representation using Semantic-Based Patterns	28
<i>Dakshi Kapugama Geeganage, Yue Xu, and Yuefeng Li</i>	
Outlier Detection Based Accurate Geocoding of Historical Addresses	41
<i>Nishadi Kirielle, Peter Christen, and Thilina Ranbaduge</i>	
SPDF: Set Probabilistic Distance Features for Prediction of Population Health Outcomes via Social Media	54
<i>Hung Nguyen, Duc Thanh Nguyen, and Thin Nguyen</i>	
Estimating County Health Indices Using Graph Neural Networks	64
<i>Hung Nguyen, Duc Thanh Nguyen, and Thin Nguyen</i>	
Joint Sequential Data Prediction with Multi-stream Stacked LSTM Network	77
<i>Nguyen Thanh Toan, Orçun Gümüş, Nguyen Thanh Tam, Nguyen Quoc Viet Hung, Rene Hexel, and Jun Jo</i>	
Show Me Your Friends and I'll Tell You Who You Are. Finding Anomalous Time Series by Conspicuous Cluster Transitions	91
<i>Martha Tatusch, Gerhard Klassen, Marcus Bravidor, and Stefan Conrad</i>	
Applying Softmax Classifiers to Open Set	104
<i>Darren Webb</i>	
An Efficient Risk Data Learning with LSTM RNN	116
<i>Ka Yee Wong and Raymond K. Wong</i>	
Using Transfer Learning to Detect Phishing in Countries with a Small Population.	129
<i>Wernsen Wong, Yun Sing Koh, and Gillian Dobbie</i>	

Application Track

Classifying Imbalanced Road Accident Data Using Recurring Concept Drift	143
<i>Robert Anderson, Yun Sing Koh, and Gillian Dobbie</i>	
Interpretability of Machine Learning Solutions in Industrial Decision Engineering.	156
<i>Inna Kolyshkina and Simeon Simoff</i>	
Customer Wallet Share Estimation for Manufacturers Based on Transaction Data.	171
<i>Xiang Li, Ali Shemshadi, Łukasz P. Olech, and Zbigniew Michalewicz</i>	
Curtailing the Tax Leakages by Nabbing Return Defaulters in Taxation System	183
<i>Priya Mehta, Jithin Mathews, Sandeep Kumar, K. Suryamukhi, and Ch Sobhan Babu</i>	
Network Path Estimation in Uncertain Data via Entity Resolution	196
<i>Dean Philp, Naomi Chan, and Wolfgang Mayer</i>	
Interactive Deep Metric Learning for Healthcare Cohort Discovery	208
<i>Yang Wang, Guodong Long, Xueping Peng, Allison Clarke, Robin Stevenson, and Leah Gerrard</i>	
Data Replication Optimization Using Simulated Annealing.	222
<i>Chee Keong Wee and Richi Nayak</i>	
Readiness of Smartphones for Data Collection and Data Mining with an Example Application in Mental Health.	235
<i>Darren Yates and Md. Zahidul Islam</i>	

Industry Showcase

Predictive Analytics for Tertiary Learners in New Zealand Who Are at Risk of Dropping Out of Education	249
<i>Wenyong Xu, Scott Luo, Stephanie Hacksley, Tim Trewinnard, Stuart Cambridge, and Syen Jien Nik</i>	

Author Index	257
------------------------	-----