

Editorial Board Members

Joaquim Filipe 

Polytechnic Institute of Setúbal, Setúbal, Portugal

Ashish Ghosh

Indian Statistical Institute, Kolkata, India

Raquel Oliveira Prates 

Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil

Lizhu Zhou

Tsinghua University, Beijing, China

More information about this series at <https://link.springer.com/bookseries/7899>

Yue Xu · Rosalind Wang · Anton Lord ·
Yee Ling Boo · Richi Nayak · Yanchang Zhao ·
Graham Williams (Eds.)

Data Mining

19th Australasian Conference on Data Mining, AusDM
Brisbane, QLD, Australia, December 14–15, 2021
Proceedings

Editors

Yue Xu 
Queensland University of Technology
Brisbane, QLD, Australia


Rosalind Wang 
Western Sydney University
Parramatta, NSW, Australia

Anton Lord 
University of Queensland
Herston, Australia

Yee Ling Boo 
RMIT University
Melbourne, VIC, Australia

Richi Nayak 
Queensland University of Technology
Brisbane, QLD, Australia

Yanchang Zhao 
Data61, CSIRO
Canberra, ACT, Australia

Graham Williams 
Australian National University
Canberra, ACT, Australia

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-981-16-8530-9 ISBN 978-981-16-8531-6 (eBook)
<https://doi.org/10.1007/978-981-16-8531-6>

© Springer Nature Singapore Pte Ltd. 2021, corrected publication 2022

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 19th Australasian Data Mining Conference (AusDM 2021) held at the Queensland University of Technology, Brisbane, during December 14–15, 2021.

The AusDM conference series first started in 2002 as a workshop initiated by Simeon Simoff (Western Sydney University), Graham Williams (Australian National University), and Markus Hegland (Australian National University). Over the years, AusDM has established itself as the premier Australasian meeting for both practitioners and researchers in area of data mining (or nowadays called data analytics or data science). 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 through presentations and discussions on the state-of-art research and development. Built on this tradition, AusDM 2021 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 2021 received 32 valid submissions, with authors from 13 different countries. The top three countries, in terms of the number of authors who submitted papers to AusDM 2021, were Australia (81 authors), India (nine authors), and Germany (eight authors). All submissions went through the double-blind review process, and each paper received at least three peer reviewed reports. Additional reviewers were considered for a clearer review outcome, if review comments from the initial three reviewers were inconclusive.

Out of these 32 submissions, a total of 16 papers were finally accepted for publication. The overall acceptance rate for AusDM 2021 was 50%. To further drill into the composition of papers submitted, there were 17 Research Track submissions and eight papers (i.e. 47%) were accepted for publication. On the other hand, out of the 15 submissions in the Application Track, eight papers (i.e. 53%) were accepted for publication.

The AusDM 2021 Organizing Committee would like to give their special thanks to Jeremy Howard and Sharma Chakravarthy for giving the insightful keynote speeches. The committee were appreciative to have Charles Palmer, Alana Maurushat, and Anton Lord presenting and sharing their industry experiences at the Industry Showcase track. The committee would like to give their sincere thanks to Queensland University of Technology for providing administration support and the conference venue. The committee would also like to thank Springer CCIS and the Editorial Board for their acceptance to publish AusDM 2021 papers. This will give the accepted papers excellent exposure and facilitate the dissemination of knowledge through publication. We would also like to give our heartfelt thanks to all student and staff volunteers at the Queensland University of Technology who did a tremendous job in ensuring a successful conference event.

Last but not least, we would like to give our sincere thanks to all delegates for attending the conference this year virtually or in-person at Queensland University of Technology. We hope that it was a fruitful experience and you enjoyed AusDM 2021!

December 2021

Yue Xu
Rosalind Wang
Anton Lord
Yee Ling Boo
Richi Nayak
Yanchang Zhao
Graham Williams

Organization

Conference Chairs

Richi Nayak	Queensland University of Technology, Australia
Yanchang Zhao	Data61, CSIRO, Australia
Graham Williams	Australian National University, Australia

Program Chairs (Research Track)

Yue Xu	Queensland University of Technology, Australia
Rosalind Wang	Western Sydney University, Australia

Program Chair (Application Track)

Anton Lord	Leap in!, Australia
------------	---------------------

Program Chair (Industry Track)

Warwick Graco	Australian Tax Office, Australia
---------------	----------------------------------

Organizing Chairs

Khanh Luong	Queensland University of Technology, Australia
Thirunavukarasu Balasubramaniam	Queensland University of Technology, Australia

Publicity Chair

Md Abdul Bashar	Queensland University of Technology, Australia
-----------------	--

Proceedings Chair

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

Steering Committee Chairs

Simeon Simoff	Western Sydney University, Australia
Graham Williams	Australian National University, Australia

Steering Committee

Peter Christen	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	University of Auckland, New Zealand
Jiuyong (John) Li	University of South Australia, Australia
Richi Nayak	Queensland University of Technology, Australia
Kok-Leong Ong	RMIT University, Australia
Dharmendra Sharma	University of Canberra, Australia
Glenn Stone	Western Sydney University, Australia
Yanchang Zhao	Data61, CSIRO, Australia

Honorary Advisors

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

Program Committee

Research Track

Xuan-Hong Dang	IBM Thomas J. Watson Research Center, USA
Philippe Fournier-Viger	Shenzhen University, China
Yang Gao	Beijing Institute of Technology, China
Md Zahidul Islam	University of South Australia, Australia
Ashad Kabir	Charles Sturt University, Australia
Selasi Kwashie	Data61, CSIRO, Australia
Gang Li	Deakin University, Australia
Wolfgang Mayer	University of South Australia, Australia
Muhammad Marwan Muhammad Fuad	Coventry University, UK
Dang Nguyen	Deakin University, Australia
Quang Vinh Nguyen	Western Sydney University, Australia
Md Anisur Rahman	Charles Sturt University, Australia
Md Geaur Rahman	Charles Sturt University, Australia
Xiaohui Tao	University of Southern Queensland, Australia
Dhananjay Thiruvady	Deakin University, Australia
Bay Vo	Ho Chi Minh City University of Technology, Vietnam
Jie Yang	University of Wollongong, Australia
Weijia Zhang	Southeast University, China

Application Track

Huma Ameer	National University of Sciences and Technology, Pakistan
Ghazal Bargshady	University of Canberra, Australia
Rohan Baxter	Australian Tax Office, Australia
Rushit Dave	North Carolina Agricultural and Technical State University, USA
Stewart Heitmann	Victor Chang Cardiac Research Institute, Australia
Hafsa Ismail	University of Canberra, Australia
Disha Kalal	Leap in!, Australia
Aunsia Khan	National University of Modern Languages, Pakistan
Apurva Khemani	Queensland University of Technology, Australia
Paul Visendi Muhindira	Queensland University of Technology, Australia
Winnie Nakiyingi	African Institute for Mathematical Sciences, Rwanda
Xueping Peng	University of Technology Sydney, Australia
Sharon Torao Pingi	Queensland University of Technology, Australia
Yanfeng Shu	CSIRO, Australia
Hao Zhu	Australian National University, Australia

Additional Reviewers

Rasoul Amirzadeh
 Fatemah Ansarizadeh
 Saqib Nawaz
 Md Shakil

Contents

Research Track

Parallel Nonlinear Dimensionality Reduction Using GPU Acceleration	3
<i>Yezihalem Tegegne, Zhonglin Qu, Yu Qian, and Quang Vinh Nguyen</i>	
Taking the Confusion Out of Multinomial Confusion Matrices and Imbalanced Classes	16
<i>David Lovell, Bridget McCarron, Brendan Langfield, Khoa Tran, and Andrew P. Bradley</i>	
Sharpshooting Most Beneficial Part of AUC for Detecting Malicious Logs	31
<i>Taishi Nishiyama, Atsutoshi Kumagai, Akinori Fujino, and Kazunori Kamiya</i>	
A Drift Aware Hierarchical Test Based Approach for Combating Social Spammers in Online Social Networks	47
<i>Darshika Koggalahewa, Yue Xu, and Ernest Foo</i>	
Hospital Readmission Prediction Using Semantic Relations Between Medical Codes	62
<i>Sea Jung Im, Yue Xu, and Jason Watson</i>	
HFM++: An Enhanced Holographic Factorization Machine for Recommendation	72
<i>Zhengxin Fang, Mingcheng Qu, Shuai Zhang, Jingyu Zhang, Yi Yuan, Lina Yao, and Shiping Chen</i>	
Deep Learning for Bias Detection: From Inception to Deployment	86
<i>Md Abul Bashar, Richi Nayak, Anjor Kothare, Vishal Sharma, and Kesavan Kandadai</i>	
Exploring Fusion Strategies in Deep Learning Models for Multi-Modal Classification	102
<i>Duoyi Zhang, Richi Nayak, and Md Abul Bashar</i>	

Application Track

Chameleon: A Python Workflow Toolkit for Feature Selection	121
<i>Diviya Thilakeswaran, Simon McManis, and X. Rosalind Wang</i>	

PostMatch: A Framework for Efficient Address Matching	136
<i>Darren Yates, Md Zahidul Islam, Yanchang Zhao, Richi Nayak, Vladimir Estivill-Castro, and Salil Kanhere</i>	
Detection of Classical Cipher Types with Feature-Learning Approaches	152
<i>Ernst Leierzopf, Vasily Mikhalev, Nils Kopal, Bernhard Esslinger, Harald Lampesberger, and Eckehard Hermann</i>	
SOMPS-Net: Attention Based Social Graph Framework for Early Detection of Fake Health News	165
<i>Prasannakumaran Dhanasekaran, Harish Srinivasan, S. Sowmiya Sree, I. Sri Gayathri Devi, Saikrishnan Sankar, and Vineeth Vijayaraghavan</i>	
The Impact of Sentiment in the News Media on Daily and Monthly Stock Market Returns	180
<i>Justin Case and Adam Clements</i>	
Investigation of Topic Modelling Methods for Understanding the Reports of the Mining Projects in Queensland	196
<i>Yasuko Okamoto, Thirunavukarasu Balasubramaniam, and Richi Nayak</i>	
A Semi-automatic Data Extraction System for Heterogeneous Data Sources: a Case Study from Cotton Industry	209
<i>Richi Nayak, Thirunavukarasu Balasubramaniam, Sangeetha Kutty, Sachindra Banduthilaka, and Erin Peterson</i>	
Nonnegative Matrix Factorization to Understand Spatio-Temporal Traffic Pattern Variations During COVID-19: A Case Study	223
<i>Anandkumar Balasubramaniam, Thirunavukarasu Balasubramaniam, Rathinaraja Jeyaraj, Anand Paul, and Richi Nayak</i>	
Correction to: Taking the Confusion Out of Multinomial Confusion Matrices and Imbalanced Classes	C1
<i>David Lovell, Bridget McCarron, Brendan Langfield, Khoa Tran, and Andrew P. Bradley</i>	
Author Index	235