# Lecture Notes in Artificial Intelligence 13281

## Subseries of Lecture Notes in Computer Science

#### Series Editors

Randy Goebel University of Alberta, Edmonton, Canada

Wolfgang Wahlster DFKI, Berlin, Germany

Zhi-Hua Zhou Nanjing University, Nanjing, China

#### Founding Editor

Jörg Siekmann DFKI and Saarland University, Saarbrücken, Germany More information about this subseries at https://link.springer.com/bookseries/1244

João Gama · Tianrui Li · Yang Yu · Enhong Chen · Yu Zheng · Fei Teng (Eds.)

# Advances in Knowledge Discovery and Data Mining

26th Pacific-Asia Conference, PAKDD 2022 Chengdu, China, May 16–19, 2022 Proceedings, Part II



*Editors* João Gama Laboratory of Artificial Intelligence and Decision Support University of Porto Porto, Portugal

Yang Yu National Key Laboratory for Novel Software Technology Nanjing University Nanjing, China

Yu Zheng JD iCity, JD Technology & JD Intelligent Cities Research Beijing, China Tianrui Li School of Computing and Artificial Intelligence Southwest Jiaotong University Chengdu, China

Enhong Chen School of Computer Science and Technology University of Science and Technology of China Hefei, China

Fei Teng School of Computing and Artificial Intelligence Southwest Jiaotong University Chengdu, China

ISSN 0302-9743 ISSN 1611-3349 (electronic) Lecture Notes in Artificial Intelligence ISBN 978-3-031-05935-3 ISBN 978-3-031-05936-0 (eBook) https://doi.org/10.1007/978-3-031-05936-0

LNCS Sublibrary: SL7 - Artificial Intelligence

© The Editor(s) (if applicable) and The Author(s), under exclusive license

to Springer Nature Switzerland AG 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 Switzerland AG The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

#### **General Chairs' Preface**

On behalf of the Organizing Committee, it is our great pleasure to welcome you to the 26th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD2022), held in Chengdu, China, during May 16–19, 2022. Starting in 1997, PAKDD has long established itself as one of the leading international conferences in data mining and knowledge discovery. PAKDD provides an international forum for researchers and industry practitioners to share their new ideas, original research results, and practical development experiences from all Knowledge Discovery and Data Mining (KDD) related areas. In response to the COVID-19 pandemic and the need for social distancing, PAKDD 2022 was held as a hybrid conference for both online and onsite attendees.

Our gratitude goes first and foremost to the researchers, who submitted their work to PAKDD 2022. We would like to deliver our sincere thanks for their efforts in research, as well as in preparing high-quality presentations. We also thank all the collaborators and sponsors for their trust and cooperation. It is our great honor that three eminent keynote speakers joined the conference: Jian Pei (Simon Fraser University, Canada), Bernhard Schölkopf (Max Planck Institute for Intelligent Systems, Germany) and Ji-Rong Wen (Renmin University, China). They were extremely professional and have high reputations in their respective areas. We enjoyed their participation and talks, which made the conference one of the best academic platforms for knowledge discovery and data mining.

We would like to express our sincere gratitude to the contributions of Steering Committee members, Organizing Committee members, Program Committee members and anonymous reviewers, led by Program Committee Co-chairs: João Gama (University of Porto), Tianrui Li (Southwest Jiaotong University), and Yang Yu (Nanjing University). We are also grateful for the hosting organization Southwest Jiaotong University which is continuously providing institutional and financial support to PAKDD 2022. We feel beholden to the PAKDD Steering Committees for their constant guidance and sponsorship of manuscripts.

Finally, our sincere thanks go to all the participants and volunteers. We hope all of you enjoyed PAKDD 2022.

April 2022

Enhong Chen Yu Zheng

## PC Chairs' Preface

It is our great pleasure to present at the 26th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2022) as the Program Committee Chairs. PAKDD is one of the longest established and leading international conferences in the areas of data mining and knowledge discovery. It provides an international forum for researchers and industry practitioners to share their new ideas, original research results, and practical development experiences from all KDD related areas, including data mining, data warehousing, machine learning, artificial intelligence, databases, statistics, knowledge engineering, big data technologies and foundations.

This year PAKDD received 627 submissions, among which 69 submissions were rejected at a preliminarily stage due to the policy violations. There were 320 Program Committee members and 45 Senior Program Committees members involved in reviewing process. Each submission was reviewed by at least three different reviewers. Over 67% of those submissions were reviewed by four or more reviewers. Eventually, 121 submissions were accepted and recommended to be published, resulting in an acceptance rate of 19.30%. Out of these, 29 submissions were about applications, 4 submissions were related to big data technologies, 46 submissions were on data science and 42 submissions were about foundations. We would like to appreciate all PC members and reviewers, who offered a high-quality program with diligence on PAKDD 2022.

The conference program featured keynote speeches from distinguished researchers in the community, most influential paper talks, cutting-edge workshops and comprehensive tutorials.

We wish to sincerely thank all PC members and reviewers for their invaluable efforts in ensuring a timely, fair, and highly effective PAKDD 2022 program.

April 2022

João Gama Tianrui Li Yang Yu

## **Organization Committee**

## **Honorary Co-chairs**

Dan Yang	Southwest Jiaotong University, China
Zhi-Hua Zhou	Nanjing University, China

## **General Co-chairs**

Yu Zheng

University of Science and Technology of China, China JD.com, China

#### **Program Committee Co-chairs**

Joao Gama	University of Porto, Portugal
Tianrui Li	Southwest Jiaotong University, China
Yang Yu	Nanjing University, China

#### **Workshop Co-chairs**

Gill Dobbie	University of Auckland, New Zealand
Can Wang	Griffith University, Australia

#### **Tutorial Co-chairs**

Gang Li	Deakin University, Australia
Tanmoy Chakraborty	Indraprastha Institute of Information Technology
	Delhi, India

#### **Local Arrangement Co-chairs**

Yan Yang	Southwest Jiaotong University, China
Chuan Luo	Sichuan University, China
Xin Yang	Southwestern University of Finance and
	Economics, China

## **Sponsor Chair**

Xiaobo Zhang	Southwest Jiaotong	University,	China
8		<i>j</i> ,	

### **Publicity Co-chairs**

Xiangnan Ren	Group 42, United Arab Emirates
Hao Wang	Zhejiang Lab, China
Junbo Zhang	JD.com, China
Chongshou Li	Southwest Jiaotong University, China

#### **Proceedings Chair**

Fei Teng	Southwest Jiaotong University, China

#### Web and Content Co-chairs

Xiaole Zhao	Southwest Jiaotong University, China
Zhen Jia	Southwest Jiaotong University, China

## **Registration Chairs**

Hongmei Chen	Southwest Jiaotong University, China
Jie Hu	Southwest Jiaotong University, China
Yanyong Huang	Southwestern University of Finance and
	Economics, China

#### **Steering Committee**

Longbing Cao Ming-Syan Chen David Cheung Gill Dobbie Joao Gama Zhiguo Gong Tu Bao Ho

Joshua Z. Huang

Masaru Kitsuregawa Rao Kotagiri Jae-Gil Lee

University of Technology Sydney, Australia
NTU
University of Hong Kong, China
University of Auckland, New Zealand
University of Porto, Portugal
University of Macau, China
Japan Advanced Institute of Science and
Technology, Japan
Shenzhen Institutes of Advanced Technology,
Chinese Academy of Sciences, China
Tokyo University, Japan
University of Melbourne, Australia
Korea Advanced Institute of Science &
Technology, South Korea

- Ee-Peng Lim Huan Liu Hiroshi Motoda Jian Pei Dinh Phung P. Krishna Reddy
- Kyuseok Shim Jaideep Srivastava Thanaruk Theeramunkong Vincent S. Tseng Takashi Washio Geoff Webb Kyu-Young Whang

Graham Williams Min-Ling Zhang Chengqi Zhang Ning Zhong Zhi-Hua Zhou

#### **Host Institute**

Singapore Management University, Singapore Arizona State University, USA AFOSR/AOARD and Osaka University, Japan Simon Fraser University, Canada Monash University, Australia International Institute of Information Technology, Hyderabad, India Seoul National University, South Korea University of Minnesota, USA Thammasat University, Thailand NCTU Osaka University, Japan Monash University, Australia Korea Advanced Institute of Science & Technology, South Korea Australian National University, Australia Southeast University, China University of Technology Sydney, Australia Maebashi Institute of Technology, Japan Nanjing University, China



南京通大 Southwest Jiaotong University

# **Contents – Part II**

#### Foundations

Text2Chart: A Multi-staged Chart Generator from Natural Language TextMd. Mahinur Rashid, Hasin Kawsar Jahan, Annysha Huzzat,Riyasaat Ahmed Rahul, Tamim Bin Zakir, Farhana Meem,Md. Saddam Hossain Mukta, and Swakkhar Shatabda	3
ENDASh: Embedding Neighbourhood Dissimilarity with Attribute Shuffling for Graph Anomaly Detection	17
Convergence and Applications of ADMM on the Multi-convex Problems Junxiang Wang and Liang Zhao	30
Prototypical Classifier for Robust Class-Imbalanced Learning Tong Wei, Jiang-Xin Shi, Yu-Feng Li, and Min-Ling Zhang	44
Quantum Entanglement Inspired Correlation Learning for Classification Junwei Zhang, Zhao Li, Juan Wang, Yinghui Wang, Shichang Hu, Jie Xiao, and Zhaolin Li	58
Self-paced Safe Co-training for Regression Fan Min, Yu Li, and Liyan Liu	71
Uniform Evaluation of Properties in Activity Recognition Seyed M. R. Modaresi, Aomar Osmani, Mohammadreza Razzazi, and Abdelghani Chibani	83
Effect of Different Encodings and Distance Functions on Quantum Instance-Based Classifiers Alessandro Berti, Anna Bernasconi, Gianna M. Del Corso, and Riccardo Guidotti	96
Attention-to-Embedding Framework for Multi-instance Learning Mei Yang, Yu-Xuan Zhang, Mao Ye, and Fan Min	109
Multi-instance Embedding Learning Through High-level Instance Selection	122

High Average-Utility Itemset Sampling Under Length Constraints	134
Divide and Imitate: Multi-cluster Identification and Mitigation of Selection Bias Katharina Dost, Hamish Duncanson, Ioannis Ziogas, Patricia Riddle, and Jörg Wicker	149
Hypersphere Neighborhood Rough Set for Rapid Attribute Reduction Yu Fang, Xue-Mei Cao, Xin Wang, and Fan Min	161
A Novel Clustering Algorithm with Dynamic Boundary Extraction Strategy Based on Local Gravitation Jiangmei Luo, Qingsheng Zhu, Junnan Li, Dongdong Cheng, and Mingqiang Zhou	174
Modelling Zeros in Blockmodelling Laurence A. F. Park, Mohadeseh Ganji, Emir Demirovic, Jeffrey Chan, Peter Stuckey, James Bailey, Christopher Leckie, and Rao Kotagiri	187
Towards Better Generalization for Neural Network-Based SAT Solvers Chenhao Zhang, Yanjun Zhang, Jeff Mao, Weitong Chen, Lin Yue, Guangdong Bai, and Miao Xu	199
Robust and Provable Guarantees for Sparse Random Embeddings Maciej Skorski, Alessandro Temperoni, and Martin Theobald	211
Transferable Interpolated Adversarial Attack with Random-Layer Mixup Size Ma, Keji Han, Xianzhong Long, and Yun Li	224
Deep Depression Prediction on Longitudinal Data via Joint Anomaly Ranking and Classification Guansong Pang, Ngoc Thien Anh Pham, Emma Baker, Rebecca Bentley, and Anton van den Hengel	236
DeepPAMM: Deep Piecewise Exponential Additive Mixed Models for Complex Hazard Structures in Survival Analysis Philipp Kopper, Simon Wiegrebe, Bernd Bischl, Andreas Bender, and David Rügamer	249
Assessing Classifier Fairness with Collider Bias Zhenlong Xu, Ziqi Xu, Jixue Liu, Debo Cheng, Jiuyong Li, Lin Liu, and Ke Wang	262

Hard Negative Sample Mining for Contrastive Representation in Reinforcement Learning <i>Qihang Chen, Dayang Liang, and Yunlong Liu</i>	277
Reduction of the Position Bias via Multi-level Learning for Activity Recognition Aomar Osmani and Massinissa Hamidi	289
Modeling IsA Relations via Box Structure for Knowledge Graph Embedding Yao Dong, Lei Wang, Ji Xiang, and Kai Liu	303
A GNN-Enhanced Game Bot Detection Model for MMORPGs Xianyang Qi, Jiashu Pu, Shiwei Zhao, Runze Wu, and Jianrong Tao	316
Online Learning with Regularized Knowledge Gradients Donghun Lee and Warren B. Powell	328
Fact Aware Multi-task Learning for Text Coherence Modeling Tushar Abhishek, Daksh Rawat, Manish Gupta, and Vasudeva Varma	340
Open Set Recognition for Time Series Classification Tolga Akar, Thorben Werner, Vijaya Krishna Yalavarthi, and Lars Schmidt-Thieme	354
Discretization Inspired Defence Algorithm Against Adversarial Attacks on Tabular Data	367
Safe Offline Reinforcement Learning Through Hierarchical Policies Shaofan Liu and Shiliang Sun	380
Leveraged Mel Spectrograms Using Harmonic and Percussive Components in Speech Emotion Recognition David Hason Rudd, Huan Huo, and Guandong Xu	392
SelectAug: A Data Augmentation Method for Distracted Driving Detection Yuan Li, Wei Mi, Jingguo Ge, Jingyuan Hu, Hui Li, Daoqing Zhang, and Tong Li	405
Neural Topic Modeling with Gaussian Mixture Model and Householder Flow	417

Dynamic Topic-Noise Models for Social Media Rob Churchill and Lisa Singh	429
Contrastive Attributed Network Anomaly Detection with Data	
Zhiming Xu, Xiao Huang, Yue Zhao, Yushun Dong, and Jundong Li	444
Cross-Lingual Product Retrieval in E-Commerce Search Wenya Zhu, Xiaoyu Lv, Baosong Yang, Yinghua Zhang, Xu Yong, Linlong Xu, Yinfu Feng, Haibo Zhang, Qing Da, Anxiang Zeng, and Ronghua Chen	458
An Adaptable Indexing Pipeline for Enriching Meta Information of Datasets from Heterogeneous Repositories Siamak Farshidi and Zhiming Zhao	472
FLiB: Fair Link Prediction in Bipartite Network Piyush Kansal, Nitish Kumar, Sangam Verma, Karamjit Singh, and Pranav Pouduval	485
SMITH: A Self-supervised Downstream-Aware Framework for Missing Testing Data Handling Chih-Chun Yang, Cheng-Te Li, and Shou-De Lin	499
Tlife-GDN: Detecting and Forecasting Spatio-Temporal Anomalies via Persistent Homology and Geometric Deep Learning	511
Layer Adaptive Deep Neural Networks for Out-of-Distribution Detection Haoliang Wang, Chen Zhao, Xujiang Zhao, and Feng Chen	526
Improving Energy-Based Out-of-Distribution Detection by Sparsity    Regularization    Qichao Chen, Wenjie Jiang, Kuan Li, and Yi Wang	539
Author Index	553