ICDM 2021

21st IEEE International Conference on Data Mining

7 – 10 DECEMBER 2021
AUCKLAND NEW ZEALAND

Contents

Agenda at a glance

Tuesday, December 7 (Pre-conference day)	
14:00-21:00	Workshop 01 - NeuRec: Advanced Neural Algorithms and Theories for Recommender Systems
14:00-21:00	Workshop 02 - SENTIRE: Sentiment Elicitation from Natural Text for Information Retrieval and Extraction
14:00-21:00	Workshop 03 - DMS: Data Mining for Service
14:00-21:00	Workshop 04 - IncrLearn: Incremental classification and clustering, concept drift, novelty detection in big/fast data context
14:00-21:00	Workshop 05 - SSTDM: 16th International Workshop on Spatial and Spatiotemporal Data Mining and (WAIN)/SSTDM: 1st Workshop on AI for Nudging and Personalization
14:00-21:00	Workshop 06 - IAAA: International Workshop on Intelligence-Augmented Anomaly Analytics
14:00- 18:30	PhD Panel
14:00- 17:00	Workshop 07 - UDML: Utility Driven Mining and Learning
14:00- 17:00	Workshop 08 - SFE-TSDM: Systematic Feature Engineering for Time-Series Data Mining
14:00- 17:00	Workshop 09 - DLC: Deep Learning and Clustering
14:00- 17:00	Workshop 10 - OEDM: Optimization Based Techniques for Emerging Data Mining Problems
14:00- 17:00	Workshop 11 - EDMML: Evolutionary Data Mining and Machine Learning
14:00- 17:00	Workshop 12 - LITSA: 2nd Workshop on Large-scale Industrial Time Series Analysis
14:00- 17:00	Workshop 13 - HDM: High Dimensional Data Mining
14:00- 17:00	Workshop 14 - DMBIH: Data Mining in Biomedical Informatics and Healthcare
14:00- 17:00	Workshop 15 - MLLD: Mining and Learning in the Legal Domain
15:00-18:00	Steering Committee Meeting (by invitation only)
16:30-17:30	Networking (Breakout rooms and speed networking)
17:30 - 20:00	Workshop 16 - CLEATED: Continual Learning and Adaptation for Time Evolving Data
17:30- 20:00	Workshop 17 - SDM: Social Data Mining in the Post-pandemic Era
17:30- 20:00	Workshop 18 - DMC: Data Mining and Machine Learning in Cybersecurity

Wednesday, December 8	
14:00- 15:00	Opening Ceremony
15:00- 16:00	Keynote: Towards Robust Machine Learning: Weak Supervision, Noisy Labels, and Beyond
16:00- 18:30	Tutorial 01: Wikimedia Visual Resources and its Application to Neural Visual Recommender Systems
16:00-18:30	Tutorial 02: Complement, Composite and Context: The 3C-Law to Build Multidomain Recommender Systems
16:00-17:30	Session 1A: Data Mining Foundations and Frontiers (Stream A)
16:00-17:30	Session 1B: Graph Mining and Learning (Stream B)
16:00-17:30	Session 1C: Sequences and Time Series 1 (Stream C)
16:00-17:30	Session 1D: Text 1 (Stream D)
16:00-17:30	Session 1E: Traffic Forecasting (Stream E)
17:30-18:00	Networking (Breakout rooms and speed networking)
18:00- 19:30	Session 2A: Clustering (Stream A)
18:00- 19:30	Session 2B: Algorithmic Methods (Stream B)
18:00- 19:30	Session 2C: Sequences and Time Series 2 (Stream C)
18:00- 19:30	Session 2D: Text 2 (Stream D)
18:00- 19:30	Session 2E: Applications (Stream E)
18:30-21:00	Tutorial 03: Roles in Networks - Foundations, Methods and Applications
19:30-20:00	Networking (Breakout rooms and speed networking)
20:00- 21:00	Job Matching Networking Session

Thursday, December 9	
14:00- 15:00	Keynote: Sample Efficient AI with Applications in Health Care and Advanced Manufacturing
15:00-15:30	Networking
15:30- 20:30	Tutorial 04: Automated Taxonomy Discovery and Exploration
15:30 - 17:00	Session 3A: Graph Learning (Stream A)
15:30 - 17:00	Session 3B: Fairness and Interpretability (Stream B)
15:30 - 17:00	Session 3C: Sequences and Time Series 3 (Stream C)
15:30 - 17:00	Session 3D: Deep Learning 1 (Stream D)
15:30 - 17:00	Session 3E: Biological and Physical Systems (Stream E)
15:30 - 17:00	Session 3F: Representation Learning (Stream F)
17:00-17:30	Networking (Breakout rooms and speed networking)
17:30- 19:00	Session 4A: Networks (Stream A)
17:30- 19:00	Session 4B: Causality and Fairness (Stream B)
17:30- 19:00	Session 4C: Theory and Optimization (Stream C)
17:30- 19:00	Session 4D: Deep Learning 2 (Stream D)
17:30- 19:00	Session 4E: Bandits and Online Methods (Stream E)
17:30- 19:00	Session 4F: Meta Learning (Stream F)
19:00-19:30	Networking (Breakout rooms and speed networking)
19:30-20:30	Panel Discussion: Data Mining with Far Fewer Labels: Pretraining, Knowledge and Unsupervised Learning

Friday, December 10	
14:00- 15:00	Keynote: Towards Trustworthy Data Science: Interpretability, Fairness and Marketplaces
15:00-15:30	Networking
15:30- 17:00	Session 5A: Crowdsourcing (Stream A)
15:30- 17:00	Session 5B: Explainability (Stream B)
15:30- 17:00	Session 5C: Graph Neural Networks 1 (Stream C)
15:30- 17:00	Session 5D: Statisticial Techniques (Stream D)
15:30- 17:00	Session 5E: Clinical Data Analysis (Stream E)
15:30- 17:00	Session 5F: Reinforcement Learning (Stream F)
15:30- 17:00	Special Session in Diversity: Women in Data Science and Machine Learning
17:00-17:30	Networking (Breakout rooms and speed networking)
17:30-19:00	Session 6A: Recommender Systems (Stream A)
17:30-19:00	Session 6B: Adversarial Learning (Stream B)
17:30-19:00	Session 6C: Graph Neural Networks 2 (Stream C)
17:30-19:00	Session 6D: Active and Semi-Supervised Learning (Stream D)
17:30-19:00	Session 6E: Matrix and Tensor (Stream E)
19:00-19:30	Networking (Breakout rooms and speed networking)
19:30- 20:30	Awards Ceremony
20:30- 21:30	Closing

All times are in New Zealand time (GMT+13)

Message from the ICDM 2021 General Chairs

On behalf of the organizing committee of the IEEE ICDM 2021 conference and our virtual host Auckland, New Zealand, it is our great pleasure to welcome you to the 2021 IEEE International Conference on Data Mining. Due to the COVID-19 global pandemic, IEEE ICDM 2021 will be hosted virtually for the second time.

Our goal is to run the conference in a way that replicates an in-person conference as best as we can, while leveraging events that can only happen virtually. We have live keynotes, tutorials, and special sessions. We also have speed-networking sessions, online activities, and are trialing a job matching program for the first time. Additionally, this is the first time there have been Reproducibility Chairs, and a Diversity and Inclusion Chair at ICDM. We hope to ultimately encapsulate this IEEE ICDM 2021 conference with our unique New Zealand hospitality.

The organization of a successful conference would not be possible without the dedicated efforts from many individuals. In particular, we would like to express our gratitude to the Program Chairs *Pauli Miettinen*, University of Eastern Finland (Finland), *James Bailey*, The University of Melbourne (Australia); Steering Committee Chair *Xindong Wu*, Mininglamp Academy of Sciences (China); Workshop Chairs *Bing Xue*, Victoria University of Wellington (New Zealand), *Mykola Pechenizkiy*, Eindhoven University of Technology (The Netherlands); Tutorial Chairs *Wei Liu*, University of Technology Sydney (Australia), *Katharina Morik*, TU Dortmund (Germany); Online Experience/Virtual Chair *Heitor Gomes*, University of Waikato (New Zealand); Publicity Chair *Diana Benavides Prado*, The University of Auckland (New Zealand); Finance Chair *Gillian Dobbie*, The University of Auckland (New Zealand); Sponsorship Chairs *Kaiqi Zhao*, The University of Auckland (New Zealand), *Philippe Fournier-Viger*, Harbin Institute of Technology (China), *Eva García-Martín*, Ekkono Solutions (Sweden); PhD Forum Chairs *Sibylle Hess*, Eindhoven University of Technology (The Netherlands), *Meng-Fen Chiang*, The University of Auckland (New Zealand), *Lisi Chen*, Inception Institute of Artificial Intelligence (IIAI) (UAE); Diversity and Inclusion Chair, *Richi Nayak*, Queensland University of Technology (Australia); Job Matching Chairs, *Albert Bifet*, University of Waikato (New Zealand), *Lin Liu*, University of South Australia (Australia); Panel Chairs *Chenliang L*i, Wuhan University (China), *Michael Witbrock*, The University of Auckland (New Zealand); Newcomers Chairs *Huan Liu*, Arizona State University (USA), *Katerina Taskova*, The University of Auckland (New Zealand); Local Arrangement Chair *Joerg Wicker*, The University of Auckland (New Zealand); Reproducibility Chairs *Dragi Kocev*, Jožef Stefan Institute (Slovenia), *Jacob Montiel*, University of Waikato (New Zealand); Award Committee Chair *Xia Ning*, The Ohio State University (USA).

We owe special thanks to our sponsors of the conference, including the US National Science Foundation (NSF), School of Computer Science - The University of Auckland, IEEE Technical Committee on Intelligent Informatics and Google, and Two Sigma. We thank the authors, keynote speakers, special session speakers, panelists, and tutorial speakers for agreeing to present their sessions virtually. We also thank the workshop organizers for running their workshops virtually.

Finally, we thank all researchers, practitioners and students who are working in the field of data mining for their support and promotion of ICDM over the years. We wish you a productive conference with new discoveries, new collaborations and a very enjoyable virtual experience.

Yun Sing Koh and Dacheng Tao
IEEE ICDM 2021 General Co-Chairs

Message from the ICDM 2021 Program Chairs

Since its inception in 2001, the IEEE International Conference on Data Mining (ICDM) has become a premier forum for researchers, users, practitioners, and developers to exchange and disseminate not only original research results but also new research directions in data mining. The 21st IEEE ICDM is being hosted this year in Auckland, New Zealand and run as an entirely virtual conference, due to the COVID-19 pandemic.

It is our great pleasure to welcome you to ICDM 2021 and to present its proceedings to you. The ICDM conference is truly an international forum. During its nineteen-year history, the conference has been held in ten countries around the world. This year's conference continues this global trend: Our organizing and program committee members represent around 36 countries/regions. This year's conference was extremely competitive. We are pleased to announce 990 paper submissions from 46 different countries/regions for review. Best efforts were made to ensure each paper was reviewed by at least three program committee members and the selection was made on the basis of discussion among the reviewers, an area chair, and the program co-chairs. Like previous years, we implemented a triple-blind review process, ensuring that the reviewers do not know the identity of the authors or of the other individuals reviewing the same submission. This process is intended to remove bias during the paper discussions. This year, 98 regular papers (9.9% acceptance rate) and 100 short papers were selected for inclusion in the proceedings and program (giving a total acceptance rate of 20%). Of the papers that were submitted, 573 (57.9%) had student first authors. These authors represent the future of our field.

In keeping with the goal of advancing the state-of-the-art in data mining, paper topics span a range of areas including: novel data mining algorithms in traditional areas, models and algorithms for new, structured, data types; deep learning and its applications; mining sequences and sequential data; mining spatial and temporal datasets; mining textual and unstructured datasets; high performance implementations of data mining algorithms; stream data mining; mining and link analysis in networked settings; data mining in electronic commerce (e.g., recommendation); web search, advertising, and marketing tasks; methodological aspects and the KDD process; and healthcare, epidemic modeling, and clinical research.

In addition to the technical presentations, our program also highlights three outstanding keynotes given by internationally renowned, distinguished scientists Jian Pei (Simon Fraser University), Masashi Sugiyama (RIKEN Center for Advanced Intelligence Project/The University of Tokyo) and Svetha Venkatesh (Deakin University). Four tutorials will be offered and 18 workshops will be run in conjunction with the main conference. A job matching program will also be run as part of the conference. We would like to thank all those who invested their substantial efforts into making this conference what it is, starting with all the authors of the 990 manuscripts for submitting content for the conference.

Reviewing and selecting papers from a large set of submissions required the coordinated effort of many individuals. We want to thank the 55 Area Chairs and 398 Program Committee members who provided insightful feedback to the authors and helped with this selection process.

Organizing the ICDM 2021 program required the time and expertise of numerous contributors. We are very thankful for the outstanding work of Bing Xue and Mykola Pechenizkiy who served as Workshop Co-Chairs, Wei Liu and Katharina Morik who were the Tutorial Co-Chairs, Lisi Chen, Sibylle Hess and and Meng-Fen Chiang who organized the PhD Forum, Chenliang Li and Michael Witbrock who were the Panel Co-Chairs, Diana Benvides Prado who served as Publicity Chair, sponsorship chairs Kaiqi Zhao, Philippe Fournier-Viger and Eva Garcia Martin, newcomers chairs Huan Liu and Katerina Taskova, job matching chairs Albert Bifet and Lin Liu, reproducibility chairs Dragi Kocev and Jacob Montiel, Xia Ning who chaired the Best Paper Award Committee, and Joerg Wicker, Gillian Dobbie, Heitor Gomes who served as Local Arrangement Chair, Finance Chair and Online experience/virtual Chair respectively.

The guidance of the Steering Committee Chair, Xindong Wu, and the General Co-Chairs, Yun Sing Koh and Dacheng Tao, were invaluable throughout each step of the conference organization and we wish to express our appreciation to them for their tireless efforts. We also would like to extend our special thanks to Juzhen Dong for the many hours she put in to maintain and enhance the Cyberchair web system to support the conference.

Finally, we thank the ICDM community for their support of this premier conference. We hope you enjoy the ICDM conference and that you are inspired by the ideas found in these papers.

James Bailey and Pauli Miettinen
ICDM 2021 Program Co-Chairs

Message from the Workshop Chairs

The 21th IEEE International Conference on Data Mining (IEEE ICDM 2021) is a premier and truly international conference for researchers and practitioners in the broad area of data mining. The ICDM Workshops program (IEEE ICDMW) aims to provide a platform for multiple workshops with a range of more focused topics to be discussed and explored, where attendees can present their original results, exchange research ideas, identify limitations, and explore new opportunities on the theoretical development and real-world applications of data mining techniques.

Due to the global COVID-19 pandemic, IEEE ICDMW was held on virtually on 7th December 2021, followed by the IEEE ICDM 2021 conference. This year, we received 24 proposals. After the workshop proposal review, paper review and merging of workshops, the final IEEE ICDMW 2021 program consisted of 18 workshops, including 6 full-day and 12 half-day workshops. Overall, IEEE ICDMW received 266 papers while 131 papers (i.e. 50%) being presented in the final program. All the papers have been through a peer-review process to ensure high-quality papers being presented in the ICDMW 2021 proceedings.

A wide range of research topics in data mining have been included in IEEE ICDMW 2021, covering both theoretical research and real-world applications. Among the more traditional areas of data mining, ICDMW 2021 includes recommendation systems, information retrieval, information extraction, natural text, high-dimensional data mining, multi-source data, incremental learning, continual learning, sentiment analysis, deep learning, clustering, concept drift, novelty detection, feature engineering, time-series data mining, and data optimization. Furthermore, emerging research areas include data mining for bioinformatics, healthcare, engineering service, anomaly analytics, utility-driven data mining and learning, data in legal domain, spatial and spatio-temporal data mining, evolutionary computation based data mining, and cybersecurity.

It is no doubt that the success of ICDMW 2021 is a collective effort from many colleagues. Our particular thanks to all the workshop organizers for their delegate time and effort in preparing, and submitting their workshop proposals, managing their technical program committees, paper review, presentations, and invited talks from experts of their domains. Last but not least, we would like to thank all the reviewers for their thorough evaluations and constructive feedback to the papers, which ensure the high quality of the papers appeared in ICDMW 2021 and the support authors receive in further improving their work.

The PhD forum is featured by IEEE ICDM and continued to this year for the 10th year. As a tradition, the PhD forum aims to provide more opportunities for research students to present their works and communicate with their peers and senior researchers in the areas related to data mining. This forum particularly beneficial for PhD students in the early stage of their doctoral study, and MSc students who are planning to PhD study. The 2021 edition of the PhD forum includes 8 high-quality papers, spanning various topics of data mining and work in related fields including deep learning, crowdsourcing, statistical learning, anomaly detection, recommendation, and reinforcement learning. Many thanks to the program committee members of the PhD forum this year, who helped with the selection of these high quality papers and who provided high quality review comments, under the guidance of the PhD forum Chairs Lisi Chen, Meng-Fen Chiang and Sibylle Hess.

The Workshop Chairs of IEEE ICDM 2021 would like to thank the Steering Committee Chair, Xindong Wu, and the General Chairs, Yun Sing Koh and Dacheng Tao, for their previous advices and leadership. We would also like to thank the Online Experience/ Virtual Chair, Heitor Gomes, for the arrangement of the Workshops day. Furthermore, we also would like to thank the Program Chairs, Pauli Miettinen and James Bailey, for their substantial contribution to the organization of the conference, and the Finance Chair, Gillian Dobbie, for managing the finance and registration of the conference. Furthermore, special thanks to the Publicity Chair, Diana Benavides Prado, for her great effort on managing the conference web and social media, and her valuable help with the call for workshop proposals and call for papers. Last but not least we would like to thank Giuseppe Di Fatta for passing their experience as Workshop Chairs from the past year edition. Thank you all for your great effort in making IEEE ICDMW 2021 a success!

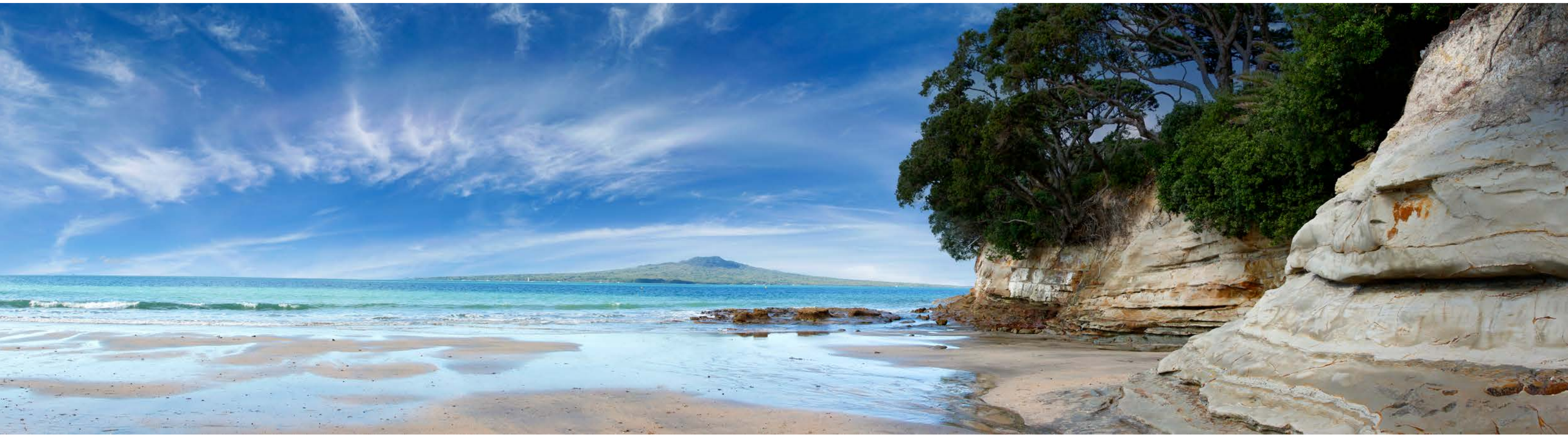
ICDM 2021 Workshop Chairs:
Bing Xue,
Victoria University of Wellington, New Zealand
Mykola Pechenizkiy,
Eindhoven University of Technology, Netherlands

Diversity, Equity, and Inclusiveness Statement

ICDM 2021 encourages the open expression and exchange of ideas, free from all forms of discrimination, retaliation, and harassment. ICDM 2021 is committed to empowering diverse, equitable, and inclusive participation, aligned with the IEEE Diversity Statement.

"IEEE's mission to foster technological innovation and excellence to benefit humanity requires the talents and perspectives of people with different personal, cultural, and disciplinary backgrounds. IEEE is committed to advancing diversity in the technical profession, and to promoting an inclusive and equitable culture in its activities and programs that welcomes, engages and rewards those who contribute to the field without regard to race, religion, gender, disability, age, national origin, sexual orientation, gender identity, or gender expression."

We require all ICDM 2021 conference participants to work towards this commitment to diversity, equity, and inclusion. ICDM 2021 should be welcoming and comfortable to all attendees and those who participate in its activities.



Program Highlights

Keynote Talks

Masashi Sugiyama *(RIKEN Center for Advanced Intelligence Project, The University of Tokyo)*. Towards Robust Machine Learning: Weak Supervision, Noisy Labels, and Beyond. Introduced by Dacheng Tao (The University of Sydney)

Svetha Venkatesh (Deakin University's Applied Artificial Intelligence Institute). Sample Efficient AI with Applications in Health Care and Advanced Manufacturing. Introduced by James Bailey (The University of Melbourne).

Jian Pei (Simon Fraser University). Towards Trustworthy Data Science: Interpretability, Fairness and Marketplaces. Introduced by Pauli Miettinen (University of Eastern Finland).

Workshops

6 full day and 12 half-day workshops in various fields that expand the main conference technical program with presentations and discussions of new directions and applications of data mining.

Panel Session

Data Mining with Far Fewer Labels: Pretraining, Knowledge and Unsupervised Learning.
Panellists: **Min Zhang** *(Tsinghua University),* **Lina Yao** *(University of New South Wales),* **Kevin Knight** *(Didi Chuxing),* **Bowen Zhow** *(JD.com),* **Imed Zitouni** *(Google). Moderated by* **Michael Witbrock** *(The University of Auckland).*

Tutorials

4 tutorials covering a range of advanced topics in data mining research:

Tutorial 1:
Wikimedia Visual Resources and its Application to Neural Visual Recommender Systems.
Organisers: Denis Parra, Antonio Ossa-Guerra, Manuel Cartagena, Patricio Cerda-Mardini, Felipe del Rio, Isidora Palma, Diego Saez-Trumper, and Miriam Redi.

Tutorial 2:
Complement, Composite and Context: The 3C-Law to Build Multidomain Recommender Systems.
Organisers: Liang Hu, Shoujin Wang, Qi Zhang, Dora D. Liu, and Longbing Cao.

Tutorial 3:
Roles in Networks – Foundations, Methods and Applications.
Organisers: Yulong Pei, Pengfei Jiao, Xuan Guo, George Fletcher, Mykola Pechenizkiy.

Tutorial 4:
Automated Taxonomy Discovery and Exploration.
Organisers: Jiaming Shen, Xiaotao Gu, Yu Meng, Jiawei Han.

Special Sessions

Special Session in Diversity: Women in Data Science and Machine Learning. **Kerri Mengersen** (Queensland University of Technology). Let's talk about uncertainty. Introduced by Richi Nayak (Queensland University of Technology).

Women in Data Science and Machine Learning Networking Panel. Moderated by **Richi Nayak** (Queensland University of Technology).

PhD Panel. **Sibo Wang** (The Chinese University of Hong Kong). Managing and Mining with Massive Graphs.

Networking sessions in between technical sessions.

Job Matching Networking Session. Moderated by: **Albert Bifet** (University of Waikato) and **Lin Liu** (University of South Australia)

Presentations by Authors of Peer-Reviewed Papers

- 33 technical sessions in various data mining fields
- 98 long papers
- 99 short papers

Awards

- 2021 IEEE ICDM Research Contributions Award
- 2021 IEEE ICDM Outstanding Service Award
- 2021 IEEE ICDM 10-Year Highest-Impact Paper Award
- 2021 IEEE ICDM Best Paper Awards
- 2021 IEEE ICDM Student Travel Awards

Conference Program

Day 1: Wednesday, December 8

Opening Ceremony **14:00-15:00**			
Keynote: Towards Robust Machine Learning: Weak Supervision, Noisy Labels, and Beyond *Speaker: Prof Masashi Sugiyama. Session Chair: Prof Dacheng Tao* **15:00-16:00**			
Session 1A: Data Mining Foundations and Frontiers **Stream A – 16:00-17:25**			
16:00	**DM455**	Highly Scalable and Provably Accurate Classification in Poincaré Balls *Eli Chien, Chao Pan, Puoya Tabaghi, and Olgica Milenkovic*	★
16:17	**DM942**	Deep Incremental RNN for Learning Sequential Data: A Lyapunov Stable Dynamical System *Ziming Zhang, Guojun Wu, Yun Yue, Yanhua Li, and Xun Zhou*	★
16:34	**DM1055**	Group-Level Cognitive Diagnosis: A Multi-Task Learning Perspective *Jie Huang, Liu Qi, Fei Wang, Zhenya Huang, Songtao Fang, Runze Wu, Chen Enhong, Yu Su, and Shijin Wang*	★
16:51	**DM632**	Isolation Kernel Density Estimation *Kai Ming Ting, Takashi Washio, Jonathan Wells, and Hang Zhang*	★
17:08	**DM947**	FGC-Stream: A novel joint miner for frequent generators and closed itemsets in data streams *Louis-Romain Roux, Tomas Martin, and Petko Valtchev*	★
Session 1B: Graph Mining and Learning **Stream B – 16:00-17:27**			
16:00	**DM462**	Graph Transfer Learning *Andrey Gritsenko, Yuan Guo, Kimia Shayestehfard, Armin Moharrer, Jennifer Dy, and Stratis Ioannidis*	★
16:17	**DM752**	A Regularized Wasserstein Framework for Graph Kernels *Asiri Wijesinghe, Qing Wang, and Stephen Gould*	★
16:34	**DM943**	THyMe+: Temporal Hypergraph Motifs and Fast Algorithms for Exact Counting *Geon Lee and Kijung Shin*	★
16:51	**DM818**	An Effective and Robust Framework by Modeling Correlations of Multiplex Network Embedding *Pengfei Jiao, Ruili Lu, Di Jin, Yinghui Wang, and Huaming Wu*	◆
17:03	**DM825**	Graph Neighborhood Routing and Random Walk for Session-based Recommendation *Zizhuo Zhang and Bang Wang*	◆
17:15	**DM390**	Fair Graph Auto-Encoder for Unbiased Graph Representations with Wasserstein Distance *Wei Fan, Kunpeng Liu, Rui Xie, Hao Liu, Hui Xiong, and Yanjie Fu*	◆
Session 1C: Sequences and Time Series 1 **Stream C – 16:00-17:27**			
16:00	**DM886**	Continual Learning for Multivariate Time Series Tasks with Variable Input Dimensions *Vibhor Gupta, Jyoti Narwariya, Pankaj Malhotra, Lovekesh Vig, and Gautam Shroff*	★
16:17	**DM758**	Towards Generating Real-World Time Series Data *Hengzhi Pei, Kan Ren, Yuqing Yang, Chang Liu, Tao Qin, and Dongsheng Li*	★
16:34	**DM979**	Multi-way Time Series Join on Multi-length Patterns *Md Parvez Mollah, Vinicius M. A. Souza, and Abdullah Mueen*	★
16:51	**DM399**	LIFE: Learning Individual Features for Multivariate Time Series Prediction with Missing Values *Zhao-Yu Zhang, Shao-Qun Zhang, Yuan Jiang, and Zhi-Hua Zhou*	◆
17:03	**DM396**	MERITS: Medication Recommendation for Chronic Disease with Irregular Time-Series *Shuai Zhang, Jianxin Li, Haoyi Zhou, Qishan Zhu, Shanghang Zhang, and Danding Wang*	◆
17:15	**DM1003**	PIETS: Parallelised Irregularity Encoders for Forecasting with Heterogeneous Time-Series *Futoon M. Abushaqra, Hao Xue, Yongli Ren, and Flora D. Salim*	◆

★ **Long** ◆ **short**

Session 1D: Text 1 **Stream D – 16:00-17:27**			
16:00	**DM813**	STAN: Adversarial Network for Cross-domain Question Difficulty Prediction *Ye Huang, Wei Huang, Shiwei Tong, Zhenya Huang, Qi Liu, Enhong Chen, Jianhui Ma, Liang Wan, and Shijin Wang*	★
16:17	**DM848**	Expert Knowledge-Guided Length-Variant Hierarchical Label Generation for Proposal Classification *Meng Xiao, Ziyue Qiao, Yanjie Fu, Yi Du, Pengyang Wang, and Yuanchun Zhou*	★
16:34	**DM921**	Topic-Noise Models: Modeling Topic and Noise Distributions in Social Media Post Collections *Robert Churchill and Lisa Singh*	★
16:51	**DM304**	BioHanBERT: A Hanzi-aware Pre-trained Language Model for Chinese Biomedical Text Mining *Xiaosu Wang, Yun Xiong, Hao Niu, Jingwen Yue, Yangyong Zhu, and Philip S. Yu*	◆
17:03	**DM450**	Aspect-based Sentiment Classification via Reinforcement Learning *Lichen Wang, Bo Zong, Yunyu Liu, Can Qin, Wei Cheng, Wenchao Yu, Xuchao Zhang, Haifeng Chen, and Yun Fu*	◆
17:15	**DM786**	Compressibility of Distributed Document Representations *Blaž Škrlj and Matej Petković*	◆
Session 1E: Traffic Forecasting **Stream E – 16:00-17:29**			
16:00	**DM1205**	Space Meets Time: Local Spacetime Neural Network For Traffic Flow Forecasting *Song Yang, Jiamou Liu, and Kaiqi Zhao*	★
16:17	**DM294**	Trajectory WaveNet: A Trajectory-Based Model for Traffic Forecasting *Bo Hui, Da Yan, Haiquan Chen, and Wei-Shinn Ku*	◆
16:29	**DM545**	Temporal Multi-view Graph Convolutional Networks for Citywide Traffic Volume Inference *Shaojie Dai, Jinshuai Wang, Chao Huang, Yanwei Yu, and Junyu Dong*	◆
16:41	**DM889**	Exploring Reflective Limitation of Behavior Cloning in Autonomous Vehicles *Mohammad Nazeri and Mahdi Bohlouli*	◆
16:53	**DM1023**	DhakaNet: Unstructured Vehicle Detection using Limited Computational Resources *Tarik Reza Toha, Masfiqur Rahaman, Saiful Islam Salim, Mainul Hossain, Arif Mohamin Sadri, and A. B. M. Alim Al Islam*	◆
17:05	**DM543**	C^3-GAN: Complex-Condition-Controlled Urban Traffic Estimation through Generative Adversarial Networks *Yingxue Zhang, Yanhua Li, Xun Zhou, Zhenming Liu, and Jun Luo*	◆
17:17	**DM611**	TEST-GCN: Topologically Enhanced Spatial-Temporal Graph Convolutional Networks for Traffic Forecasting *Muhammad Afif Ali, Suriyanarayanan Venkatesan, Victor Liang, and Hannes Kruppa*	◆

Day 1: Wednesday, December 8 (continued)

Session 2A: Clustering Stream A – 18:00-19:27			
18:00	**DM441**	Triplet Deep Subspace Clustering via Self-Supervised Data Augmentation *Zhao Zhang, Xianzhen Li, Haijun Zhang, Yi Yang, Shuicheng Yan, and Meng Wang*	★
18:17	**DM640**	Outlier-Robust Multi-View Subspace Clustering with Prior Constraints *Mehrnaz Najafi, Lifang He, and Philip S. Yu*	★
18:34	**DM801**	DCF: An Efficient and Robust Density-Based Clustering Method *Joshua Tobin and Mimi Zhang*	★
18:51	**DM330**	K-means for Evolving Data Streams *Arkaitz Bidaurrazaga Barrueta, Aritz Perez, and Marco Capo*	◆
19:03	**DM437**	Density-Based Clustering for Adaptive Density Variation *Li Qian, Claudia Plant, and Christian Böhm*	◆
19:15	**DM829**	Thin Semantics Enhancement via High-Frequency Priori Rule for Thin Structures Segmentation *Yuting He, Rongjun Ge, Jiasong Wu, Jean-Louis Coatrieux, Huazhong Shu, Yang Chen, Guanyu Yang, and Shuo Li*	◆
Session 2B: Algorithmic Methods Stream B – 18:00-19:27			
18:00	**DM883**	USTEP: Unfixed Search Tree for Efficient Log Parsing *Arthur Vervaet, Raja Chiky, and Mar Callau-Zori*	★
18:17	**DM505**	Differentially Private String Sanitization for Frequency-Based Mining Tasks *Huiping Chen, Changyu Dong, Liyue Fan, Grigorios Loukides, Solon Pissis, and Leen Stougie*	★
18:34	**DM936**	Gated Information Bottleneck for Generalization in Sequential Environments *Francesco Alesiani, Shujian Yu, and Xi Yu*	★
18:51	**DM1167**	Source Inference Attacks in Federated Learning *Hongsheng Hu, Zoran Salcic, Lichao Sun, Gillian Dobbie, and Xuyun Zhang*	◆
19:03	**DM601**	A general framework for mining concept-drifting data streams with evolvable features *Jiaqi Peng, Jinxia Guo, Qinli Yang, Jianyun Lu, and Junming Shao*	◆
19:15	**DM1183**	Robust BiPoly-Matching for Multi-Granular Entities *Ween Jiann Lee, Maksim Tkachenko, and Hady Lauw*	◆
Session 2C: Sequences and Time Series 2 Stream C – 18:00-19:27			
18:00	**DM374**	Disentangled Deep Multivariate Hawkes Process for Learning Event Sequences *Xixun Lin, Jiangxia Cao, Peng Zhang, Chuan Zhou, Zhao Li, Jia Wu, and Bin Wang*	★
18:17	**DM719**	Towards Interpretability and Personalization: A Predictive Framework for Clinical Time-series Analysis *Yang Li, Xianli Zhang, Buyue Qian, Zeyu Gao, Chong Guan, Yefeng Zheng, Hansen Zheng, Fenglang Wu, and Chen Li*	★
18:34	**DM938**	CASPITA: Mining Statistically Significant Paths in Time Series Data from an Unknown Network *Andrea Tonon and Fabio Vandin*	★
18:51	**DM343**	Matrix Profile XXIII: Contrast Profile: A Novel Time Series Primitive that Allows Real World Classification *Ryan Mercer, Sara Alaee, Alireza Abdoli, Shailendra Singh, Amy Murillo, and Eamonn Keogh*	◆
19:03	**DM247**	LOGIC: Probabilistic Machine Learning for Time Series Classification *Fabian Berns, Jan Huewel, and Christian Beecks*	◆
19:15	**DM875**	SMATE: Semi-Supervised Spatio-Temporal Representation Learning on Multivariate Time Series *Jingwei Zuo, Karine Zeitouni, and Yehia Taher*	◆

★ Long ◆ short

Session 2D: Text 2 Stream D – 18:00-19:29			
18:00	**DM706**	Powered Hawkes-Dirichlet Process: Challenging Textual Clustering using a Flexible Temporal Prior *Gaël Poux-Médard, Julien Velcin, and Sabine Loudcher*	★
18:17	**DM959**	Semi-Supervised Graph Attention Networks for Event Representation Learning *João Pedro Rodrigues Mattos and Ricardo Marcacini*	◆
18:29	**DM1015**	Learning Dynamic User Interactions for Online Forum Commenting Prediction *Wu-Jiu Sun, Xiao Fan Liu, and Fei Shen*	◆
18:41	**DM1105**	Topic-Attentive Encoder-Decoder with Pre-Trained Language Model for Keyphrase Generation *Cangqi Zhou, Jinling Shang, Jing Zhang, Qianmu Li, and Dianming Hu*	◆
18:53	**DM1125**	TCube: Domain-Agnostic Neural Time-series Narration *Mandar Sharma, John Brownstein, and Naren Ramakrishnan*	◆
19:05	**DM964**	Learning Personal Human Biases and Representations for Subjective Tasks in Natural Language Processing *Jan Kocoń, Marcin Gruza, Julita Bielaniewicz, Damian Grimling, Kamil Kanclerz, Piotr Miłkowski, and Przemysław Kazienko*	◆
19:17	**DM869**	Out-of-Category Document Identification Using Target-Category Names as Weak Supervision *Dongha Lee, Dongmin Hyun, Jiawei Han, and Hwanjo Yu*	◆
Session 2E: Applications Stream E – 18:00-19:27			
18:00	**DM972**	Deep Human-guided Conditional Variational Generative Modeling for Automated Urban Planning *Dongjie Wang, Kunpeng Liu, Pauline Johnson, Leilei Sun, Bowen Du, and Yanjie Fu*	★
18:17	**DM1027**	Conversion Prediction with Delayed Feedback: A Multi-task Learning Approach *Yilin Hou, Guangming Zhao, Chuanren Liu, Zhonglin Zu, and Xiaoqiang Zhu*	★
18:34	**DM1162**	Technological Knowledge Flow Forecasting through A Hierarchical Interactive Graph Neural Network *Liu Huijie, Wu Han, Zhang Le, Yu Runlong, Liu Ye, Liu Qi, and Chen Enhong*	★
18:51	**DM776**	Federated Principal Component Analysis for Genome-Wide Association Studies *Anne Hartebrodt, Reza Nasirigerdeh, David B. Blumenthal, and Richard Röttger*	◆
19:03	**DM290**	PaGAN: Generative Adversarial Network for Patent understanding *Guillaume Guarino, Ahmed Samet, Amir Nafi, and Denis Cavallucci*	◆
19:15	**DM1049**	Addressing Exposure Bias in Uplift Modeling for Large-scale Online Advertising *Wenwei Ke, Chuanren Liu, Xiangfu Shi, Yiqiao Dai, Philip Yu, and Xiaoqiang Zhu*	◆
20:00		**Job Matching Networking Session**	

Day 2: Thursday, December 9

Keynote: Sample Efficient AI with Applications in Health Care and Advanced Manufacturing *Speaker: Prof Svetha Venkatesh* **14:00-15:00**			
Session 3A: Graph Learning **Stream A – 15:30-16:57**			
15:30	**DM540**	GraphANGEL: Adaptive and Structure-Aware Sampling on Graph Neural Networks *Jingshu Peng, Yanyan Shen, and Lei Chen*	★
15:47	**DM629**	Deep Generation of Heterogeneous Networks *Chen Ling, Carl Yang, and Liang Zhao*	★
16:04	**DM1155**	Fast Attributed Graph Embedding via Density of States *Saurabh Sawlani, Lingxiao Zhao, and Leman Akoglu*	★
16:21	**DM509**	Adapting Membership Inference Attacks to GNN for Graph Classification: Approaches and Implications *BANG WU, Xiangwen Yang, Shirui Pan, and Xingliang Yuan*	◆
16:33	**DM577**	Bi-Level Attention Graph Neural Networks *Roshni Iyer, Wei Wang, and Yizhou Sun*	◆
16:45	**DM473**	A new multiple instance algorithm using structural information *Xiaoyan Zhu, Ting Wang, Jiayin Wang, Ying Xu, and Yuqian Liu*	◆
Session 3B: Fairness and Interpretability **Stream B – 15:30-16:50**			
15:30	**DM988**	Cardiac Complication Risk Profiling for Cancer Survivors via Multi-View Multi-Task Learning *Thai-Hoang Pham, Changchang Yin, Laxmi Mehta, Xueru Zhang, and Ping Zhang*	★
15:47	**DM559**	Multi-objective Explanations of GNN Predictions *Yifei Liu, Chao Chen, Yazheng Liu, Xi Zhang, and Sihong Xie*	★
16:04	**DM915**	Precise Bayes Classifier: Summary of Results *Amin Vahedian and Xun Zhou*	★
16:21	**DM798**	Predictive Modeling of Clinical Events with Mutual Enhancement Between Longitudinal Patient Records and Medical Knowledge Graph *Xiao Xu, Xian Xu, Yuyao Sun, Xiaoshuang Liu, Xiang Li, Guotong Xie, and Fei Wang*	★
16:38	**DM710**	Unified Fairness from Data to Learning Algorithm *Yanfu Zhang, Lei Luo, and Heng Huang*	◆
Session 3C: Sequences and Time Series 3 **Stream C – 15:30-16:57**			
15:30	**DM1002**	Attentive Neural Controlled Differential Equations for Time-series Classification and Forecasting *Sheoyon Jhin, Heejoo Shin, Seoyoung Hong, Solhee Park, and Noseong Park*	★
15:47	**DM1006**	SSDNet: State Space Decomposition Neural Network for Time Series Forecasting *Yang Lin, Irena Koprinska, and Mashud Rana*	★
16:04	**DM1104**	Ultra fast warping window optimization for Dynamic Time Warping *Chang Wei Tan, Matthieu Herrmann, and Geoffrey I. Webb*	★
16:21	**DM475**	STING: Self-attention based Time-series Imputation Networks using GAN *Eunkyu Oh, Taehun Kim, Yunhu Ji, and Sushil Khyalia*	◆
16:33	**DM408**	Spikelet: An Adaptive Symbolic Approximation for Finding Higher-Level Structure in Time Series *Makoto Imamura and Takaaki Nakamura*	◆
16:45	**DM656**	Streaming Dynamic Graph Neural Networks for Continuous-Time Temporal Graph Modeling *Sheng Tian, Tao Xiong, and Leilei Shi*	◆

★ **Long** ◆ **short**

Session 3D: Deep Learning 1 **Stream D – 15:30-16:57**			
15:30	**DM286**	Physics Interpretable Shallow-Deep Neural Networks for Physical System Identification with Unobservability *Jingyi Yuan and Yang Weng*	★
15:47	**DM363**	BaT: a Beat-aligned Transformer for Electrocardiogram Classification *Xiaoyu Li, Chen Li, Yuhua Wei, Yuyao Sun, Jishang Wei, Xiang Li, and Buyue Qian*	★
16:04	**DM424**	Robustifying DARTS by Eliminating Information Bypass Leakage via Explicit Sparse Regularization *Jiuling Zhang and Zhiming Ding*	★
16:21	**DM385**	Operation-level Progressive Differentiable Architecture Search *Xunyu Zhu, Jian Li, Yong Liu, and Weiping Wang*	◆
16:33	**DM487**	Improving Deep Forest by Exploiting High-order Interactions *Yi-He Chen, Shen-Huan Lyu, and Yuan Jiang*	◆
16:45	**DM976**	Recurrent Neural Networks Meet Context-Free Grammar: Two Birds with One Stone *Hui Guan, Umang Chaudhary, Yuanchao Xu, Lin Ning, Lijun Zhang, and Xipeng Shen*	◆
Session 3E: Biological and Physical Systems **Stream E – 15:30-16:57**			
15:30	**DM628**	Physics-Guided Machine Learning from Simulation Data: An Application in Modeling Lake and River Systems *Xiaowei Jia, Yiqun Xie, Sheng Li, Shengyu Chen, Jacob Zwart, Jeffrey Sadler, Alison Appling, Samantha Oliver, and Jordan Read*	★
15:47	**DM603**	Partial Differential Equation Driven Dynamic Graph Networks for Predicting Stream Water Temperature *Tianshu Bao, Xiaowei Jia, Jacob Zwart, Jeffrey Sadler, Alison Appling, Samantha Oliver, and Taylor Johnson*	★
16:04	**DM986**	Climate Modeling with Neural Diffusion Equations *Hwangyong Choi, Jeongwhan Choi, Jeehyun Hwang, and Noseong Park*	★
16:21	**DM521**	Pest-YOLO: Deep Image Mining and Multi-Feature Fusion for Real-Time Agriculture Pest Detection *Zhe Tang, Zhengyun Chen, Fang Qi, Lingyan Zhang, and Shuhong Chen*	◆
16:33	**DM743**	DIVINIA: Rare Object Localization and Search in Overhead Imagery *Jonathan Amazon, Khurram Shafique, Zeeshan Rasheed, and Aaron Reite*	◆
16:45	**DM261**	Exploring the Long Short-Term Dependencies to Infer Shot Influence in Badminton Matches *Wei-Yao Wang, Teng-Fong Chan, Hui-Kuo Yang, Chih-Chuan Wang, Yao-Chung Fan, and Wen-Chih Peng*	◆
Session 3F: Representation Learning **Stream F – 15:30-16:57**			
15:30	**DM828**	Efficient Reinforced Feature Selection via Early Stopping Traverse Strategy *Kunpeng Liu, Dongjie Wang, Pengfei Wang, Wan Du, Dapeng Oliver Wu, and Yanjie Fu*	★
15:47	**DM760**	PRGAN: Personalized Recommendation with Conditional Generative Adversarial Networks *Jing Wen, Bi-Yi Chen, Chang-Dong Wang, and Zhihong Tian*	★
16:04	**DM616**	Robust Low-rank Deep Feature Recovery in CNNs: Toward Low Information Loss and Fast Convergence *Jiahuan Ren, Zhao Zhang, Jicong Fan, Haijun Zhang, Mingliang Xu, and Meng Wang*	★
16:21	**DM463**	A Lookahead Algorithm for Robust Subspace Recovery *Guihong Wan and Haim Schweitzer*	◆
16:33	**DM556**	Constrained Non-Affine Alignment of Embeddings *Yuwei Wang, Yan Zheng, Yanqing Peng, Michael Yeh, Zhongfang Zhuang, Das Mahashweta, Bendre Mangesh, Feifei Li, Wei Zhang, and Jeff Phillips*	◆
16:45	**DM695**	Jointly Multi-Similarity Loss for Deep Metric Learning *Li Zhang, Shitian Shen, Lingxiao Li, and Han Wang*	◆

Day 2: Thursday, December 9 (continued)

Session 4A: Networks Stream A – 17:30-18:57			
17:30	**DM987**	Hypergraph Ego-networks and Their Temporal Evolution *Cazamere Comrie and Jon Kleinberg*	★
17:47	**DM872**	Risk-aware Temporal Cascade Reconstruction to Detect Asymptomatic Cases *Hankyu Jang, Shreyas Pai, Bijaya Adhikari, and Sriram Pemmaraju*	★
18:04	**DM619**	Better Prevent than React: Deep Stratified Learning to Predict Hate Intensity of Twitter Reply Chains *Dhruv Sahnan, Snehil Dahiya, Vasu Goel, Anil Bandhakavi, and Tanmoy Chakraborty*	★
18:21	**DM650**	Adaptive Spatio-Temporal Convolutional Network for Traffic Prediction *Mingyang Zhang, Yong Li, Funing Sun, Diansheng Guo, and Pan Hui*	◆
18:33	**DM634**	Generating Structural Node Representations via Higher-order Features and Adversarial Learning *Wang Zhang, Xuan Guo, Ting Pan, Lin Pan, Pengfei Jiao, and Wenjun Wang*	◆
18:45	**DM625**	Heterogeneous Stream-reservoir Graph Networks with Data Assimilation *Shengyu Chen, Alison Appling, Samanth Oliver, Hayley Corson-Dosch, Jordan Read, Jeffrey Sadler, Jacob Zwart, and Xiaowei Jia*	◆
Session 4B: Causality and Fairness Stream B – 17:30-18:57			
17:30	**DM663**	Nonlinear Causal Structure Learning for Mixed Data *Wenjuan Wei and Lu Feng*	★
17:47	**DM762**	A Robust Algorithm to Unifying Offline Causal Inference and Online Multi-armed Bandit Learning *Qiao Tang and Hong Xie*	★
18:04	**DM1067**	Fair Decision-making Under Uncertainty *Wenbin Zhang and Jeremy Weiss*	★
18:21	**DM520**	Joint Scence Network and Attention-Guided for Image Captioning *Dongming Zhou and Jing Yang*	◆
18:33	**DM792**	Promoting Fairness through Hyperparameter Optimization *André Cruz, Pedro Saleiro, Catarina Belém, Carlos Soares, and Pedro Bizarro*	◆
18:45	**DM934**	Causal Discovery with Flow-based Conditional Density Estimation *Shaogang Ren, Haiyan Yin, Mingming Sun, and Ping Li*	◆
Session 4C: Theory and Optimization Stream C – 17:30-18:57			
17:30	**DM452**	LAGA: Lagged AllReduce with Gradient Accumulation for Minimal Idle Time *Ido Hakimi, Rotem Zamir Aviv, Kfir Yehuda Levy, and Assaf Schuster*	★
17:47	**DM1168**	A Linear Primal-Dual Multi-Instance SVM for Big Data Classifications *Lodewijk Brand, Lauren Baker, Carla Ellefsen, Jackson Sargent, and Hua Wang*	★
18:04	**DM881**	Fast computation of distance-generalized cores using sampling *Nikolaj Tatti*	★
18:21	**DM447**	Limited-memory Common-directions Method With Subsampled Newton Directions for Large-scale Linear Classification *Jui-Nan Yen and Chih-Jen Lin*	◆
18:33	**DM465**	Online Testing of Subgroup Treatment Effects Based on Value Difference *Miao Yu, Wenbin Lu, and Rui Song*	◆
18:45	**DM815**	Scalable Pareto Front Approximation for Deep Multi-Objective Learning *Michael Ruchte and Josif Grabocka*	◆

★ Long ◆ short

Session 4D: Deep Learning 2 Stream D – 17:30-18:57			
17:30	**DM995**	Global Convolutional Neural Processes *Xuesong Wang, Lina Yao, Xianzhi Wang, Hye-young Paik, and Sen Wang*	★
17:47	**DM1148**	GANBLR: A Tabular Data Generation Model *Yishuo Zhang, Nayyar Zaidi, Jiahui Zhou, and Gang Li*	★
18:04	**DM1031**	Temporal Clustering with External Memory Network for Disease Progression Modeling *Zicong Zhang, Changchang Yin, and Ping Zhang*	★
18:21	**DM940**	PhyFlow: Physics-Guided Deep Learning for Generating Interpretable 3D Flow Fields *Nikhil Muralidhar, Jie Bu, Ze Cao, Neil Raj, Long He, Naren Ramakrishnan, Danesh Tafti, and Anuj Karpatne*	◆
18:33	**DM971**	PSANet - Subspace attention for personalized compatibility *Meet Taraviya, Anurag Beniwal, Yen-Liang Lin, and Larry Davis*	◆
18:45	**DM878**	Adversarial Regularized Reconstruction for Anomaly Detection and Generation *Angelica Liguori, Giuseppe Manco, Francesco Sergio Pisani, and Ettore Ritacco*	◆
Session 4E: Bandits and Online Methods Stream E – 17:30-18:55			
17:30	**DM673**	Cutting to the Chase with Warm-Start Contextual Bandits *Bastian Oetomo, R. Malinga Perera, Renata Borovica-Gajic, and Benjamin I. P. Rubinstein*	★
17:47	**DM911**	Online Learning in Variable Feature Spaces with Mixed Data *Yi He, Jiaxian Dong, Bo-Jian Hou, Yu Wang, and Fei Wang*	★
18:04	**DM1082**	AutoEmb: Automated Embedding Dimensionality Search in Streaming Recommendations *Xiangyu Zhao, Haochen Liu, Wenqi Fan, Hui Liu, Jiliang Tang, Chong Wang, Ming Chen, Xudong Zheng, Xiaobing Liu, and Xiwang Yang*	★
18:21	**DM486**	Attention-based Feature Interaction for Efficient Online Knowledge Distillation *Tongtong Su, Qiyu Liang, Jinsong Zhang, Zhaoyang Yu, Gang Wang, and Xiaoguang Liu*	★
18:38	**DM360**	Dictionary Pair-based Data-Free Fast Deep Neural Network Compression *Yangcheng Gao, Zhao Zhang, Haijun Zhang, Mingbo Zhao, Yang Yi, and Meng Wang*	★
Session 4F: Meta Learning Stream F – 17:30-18:55			
17:30	**DM1208**	Learning to Reweight Samples with Offline Loss Sequence *Yuhua Wei, Chen Li, Xiaoyu Li, Jishang Wei, and Buyue Qian*	★
17:47	**DM661**	Few-Shot Partial Multi-Label Learning *Yunfeng Zhao, Guoxian Yu, Lei Liu, Zhongmin Yan, Carlotta Domeniconi, and Lizhen Cui*	★
18:04	**DM769**	TRIO: Task-agnostic dataset representation optimized for automatic algorithm selection *Noy Cohen-Shapira and Lior Rokach*	★
18:21	**DM837**	MetaGB: A Gradient Boosting Framework for Efficient Task Adaptive Meta Learning *Manqing Dong, Lina Yao, Xianzhi Wang, Xiwei Xu, and Liming Zhu*	★
18:38	**DM571**	DAC-ML: Domain Adaptable Continuous Meta-Learning for Urban Dynamics Prediction *Xin Zhang, Yanhua Li, Xun Zhou, Oren Mangoubi, Ziming Zhang, Vincent Filardi, and Jun Luo*	★
19:30		Panel Discussion: Data Mining with Far Fewer Labels: Pretraining, Knowledge and Unsupervised Learning. *Moderator: Michael Witbrock Panel: Min Zhang, Lina Yao, Kevin Knight, Bowen Zhou, Imed Zitouni*	

Day 3: Friday, December 10

Keynote: Towards Trustworthy Data Science: Interpretability, Fairness and Marketplaces *Speaker: Prof Jian Pei* **14:00-15:00**			
Session 5A: Crowdsourcing **Stream A – 15:30-16:57**			
15:30	**DM535**	Truth Discovery in Sequence Labels from Crowds *Nasim Sabetpour, Adithya Kulkarni, Sihong Xie, and Qi Li*	★
15:47	**DM1069**	Crowdsourcing with Self-paced Workers *Xiangping Kang, Guoxian Yu, Carlotta Domeniconi, Jun Wang, Wei Guo, Yazhou Ren, and Lizhen Cui*	★
16:04	**DM1197**	Preference-aware Group Task Assignment in Spatial Crowdsourcing: A Mutual Information-based Approach *Yunchuan Li, Yan Zhao, and Kai Zheng*	★
16:21	**DM994**	Practitioner-Centric Approach for Early Incident Detection Using Crowdsourced Data for Emergency Services *Yasas Senarath, Ayan Mukhopadhyay, Sayyed Mohsen Vazirizade, Hemant Purohit, Saideep Nannapaneni, and Abhishek Dubey*	◆
16:33	**DM1007**	Detecting Adversaries in Crowdsourcing *Panagiotis Traganitis and Georgios B. Giannakis*	◆
16:45	**DM1179**	Zero-shot Key Information Extraction from Mixed-Style Tables: Pre-training on Wikipedia *Qingping Yang, Yingpeng Hu, Rongyu Cao, Hongwei Li, and Ping Luo*	◆
Session 5B: Explainability **Stream B – 15:30-16:49**			
15:30	**DM847**	GNES: Learning to Explain Graph Neural Networks *Yuyang Gao, Tong Sun, Rishab Bhatt, Dazhou Yu, Sungsoo Hong, and Liang Zhao*	★
15:37	**DM291**	Generating Explanations for Recommendation Systems via Injective VAE *ZeRui Cai and ZeFeng Cai*	◆
15:49	**DM457**	An Ensemble of Naive Bayes Classifiers for Uncertain Categorical Data *Marcelo Maia, Alexandre Plastino, and Alex Freitas*	◆
16:01	**DM459**	Self-learn to Explain Siamese Networks Robustly *Chao Chen, Yifan Shen, Guixiang Ma, Xiangnan Kong, Srinivas Rangarajan, Xi Zhang, and Sihong Xie*	◆
16:13	**DM624**	Alternative Ruleset Discovery to Support Black-box Model Predictions *Yoichi Sasaki and Yuzuru Okajima*	◆
16:25	**DM950**	A Multi-view Confidence-calibrated Framework for Fair and Stable Graph Representation Learning *Xu Zhang, Liang Zhang, Bo Jin, and Xinjiang Lu*	◆
16:37	**DM1099**	GCN-SE: Attention as Explainability for Node Classification in Dynamic Graphs *Yucai Fan, Yuhang Yao, and Carlee Joe-Wong*	◆
Session 5C: Graph Neural Networks 1 **Stream C – 15:30-16:57**			
15:30	**DM1000**	FRAUDRE: Fraud Detection Dual-Resistant to Graph Inconsistency and Imbalance *Ge Zhang, Jia Wu, Jian Yang, Amin Beheshti, Shan Xue, Chuan Zhou, and Michael Sheng*	★
15:47	**DM435**	Accurate Graph-Based PU Learning without Class Prior *Jaemin Yoo, Junghun Kim, Hoyoung Yoon, Geonsoo Kim, Changwon Jang, and U Kang*	★
16:04	**DM484**	AS-GCN: Adaptive Semantic Architecture of Graph Convolutional Networks for Text-Rich Networks *Zhizhi Yu, Di Jin, Ziyang Liu, Dongxiao He, Xiao Wang, Hanghang Tong, and Jiawei Han*	★
16:21	**DM214**	Dynamic Attributed Graph Prediction with Conditional Normalizing Flows *Daheng Wang, Tong Zhao, Nitesh Chawla, and Meng Jiang*	◆
16:33	**DM1113**	AdaBoosting Clusters on Graph Neural Networks *Li Zheng, Jun Gao, Zhao Li, and Ji Zhang*	◆
16:45	**DM1150**	Heterogeneous Graph Neural Architecture Search *Yang Gao, Peng Zhang, Zhao Li, Chuan Zhou, Hong Yang, Yongchao Liu, and Yue Hu*	◆
Session 5D: Statisticial Techniques **Stream D – 15:30-16:57**			
15:30	**DM904**	Combining Ranking and Point-wise Losses for Training Deep Survival Analysis Models *Lu Wang, Mark Chignell, and Yan Li*	★
15:47	**DM1200**	Spatially and Robustly Hybrid Mixture Regression Model for Inference of Spatial Dependence *Wennan Chang, Pengtao Dang, Changlin Wan, Xiaoyu Lu, Yue Fang, Tong Zhao, Yong Zang, Bo Li, Chi Zhang, and Sha Cao*	★
16:04	**DM461**	A Statistically-Guided Deep Network Transformation and Moderation Framework for Data with Spatial Heterogeneity *Yiqun Xie, Erhu He, Xiaowei Jia, Han Bao, Xun Zhou, Rahul Ghosh, and Praveen Ravirathinam*	★
16:21	**DM802**	Accurately Quantifying under Score Variability *André Maletzke, Denis dos Reis, Waqar Hassan, and Gustavo Batista*	◆
16:33	**DM423**	Gain-Some-Lose-Some: Reliable Quantification Under General Dataset Shift *Benjamin Denham, Edmund Lai, Roopak Sinha, and M. Asif Naeem*	◆
16:45	**DM380**	Boosting Deep Ensemble Performance with Hierarchical Pruning *Yanzhao Wu and Ling Liu*	◆
Session 5E: Clinical Data Analysis **Stream E – 15:30-16:57**			
15:30	**DM580**	Sequential Diagnosis Prediction with Transformer and Ontological Representation *Xueping Peng, Guodong Long, Tao Shen, Sen Wang, and Jing Jiang*	★
15:47	**DM757**	Label Dependent Attention Model for Disease Risk Prediction Using Multimodal Electronic Health Records *Shuai Niu, Qing Yin, Yunya SONG, Yike GUO, and Xian Yang*	★
16:04	**DM817**	SCEHR: Supervised Contrastive Learning for Clinical Risk Prediction using Electronic Health Records *Chengxi Zang and Fei Wang*	★
16:21	**DM588**	SCALP - Supervised Contrastive Learning for Cardiopulmonary Disease Classification and Localization in Chest X-rays using Patient Metadata *Ajay Jaiswal, Tianhao Li, Cyprian Zander, Yan Han, Justin Rousseau, Yifan Peng, and Ying Ding*	◆
16:33	**DM956**	ENGINE: Enhancing Neuroimaging and Genetic Information by Neural Embedding *Wonjun Ko, Wonsik Jung, Eunjin Jeon, Ahmad Wisnu Mulyadi, and Heung-Il Suk*	◆
16:45	**DM1103**	Multi Classification Prediction of Alzheimer's Disease based on Fusing Multi-modal Features *Qiao Pan, Ke Ding, and Dehua Chen*	short
Session 5F: Reinforcement Learning **Stream F – 15:30-16:50**			
15:30	**DM1032**	ACE-HGNN: Adaptive Curvature Exploration Hyperbolic Graph Neural Network *Xingcheng Fu, Jianxin Li, Qingyun Sun, Cheng Ji, Jia Wu, Hao Peng, Senzhang Wang, Jiajun Tan, and Philip S. Yu*	★
15:47	**DM851**	Deep Reinforced Attention Regression for Partial Sketch Based Image Retrieval *Dingrong Wang, Hitesh Sapkota, Xumin Liu, and Qi Yu*	★
16:04	**DM999**	Impression Allocation and Policy Search in Display Advertising *di wu, cheng chen, xiujun chen, junwei pan, xun yang, qing tan, jian xu, and Kuang-Chih lee*	★
16:21	**DM566**	Mcore: Multi-Agent Collaborative Learning for Knowledge-Graph-Enhanced Recommendation *Xujia Li, Yanyan Shen, and Lei Chen*	★
16:38	**DM638**	Multi-Objective Distributional Reinforcement Learning for Large-Scale Order Dispatching *Fan Zhou, Xiaocheng Tang, Chenfan Lu, Fan Zhang, Zhiwei Qin, Jieping Ye, and Hongtu Zhu*	◆

★ Long ◆ short

Day 3: Friday, December 10 (continued)

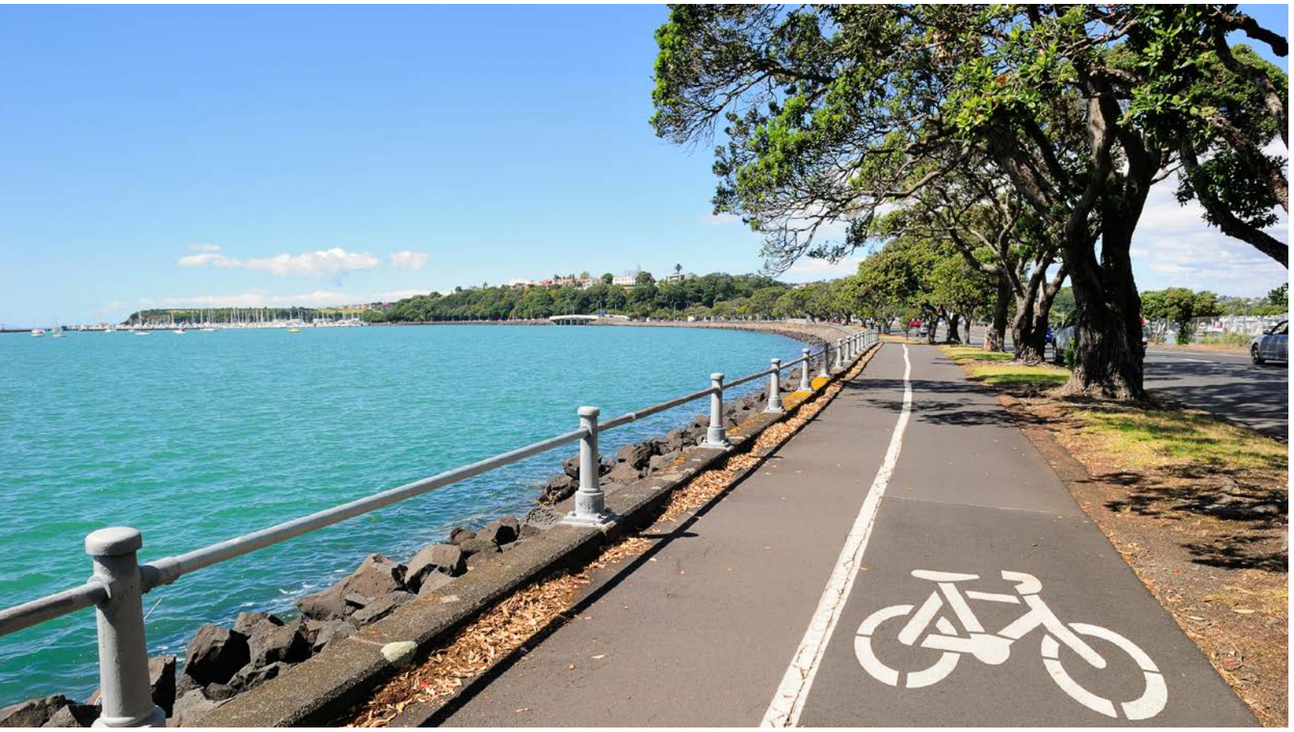
Session 6A: Recommender Systems Stream A – 17:30-18:57			
17:30	**DM468**	Hyper Meta-Path Contrastive Learning for Multi-Behavior Recommendation *Haoran Yang, Hongxu Chen, Lin Li, Philip S. Yu, and Guandong Xu*	★
17:47	**DM980**	Memory Augmented Multi-Instance Contrastive Predictive Coding for Sequential Recommendation *Ruihong Qiu, Zi Huang, and Hongzhi Yin*	★
18:04	**DM1012**	Learning Transferable User Representations with Sequential Behaviors via Contrastive Pre-training *Mingyue Cheng, Fajie Yuan, Liu Qi, Shenyang Ge, Xin Xin, and Chen Enhong*	★
18:21	**DM217**	Composition-Enhanced Graph Collaborative Filtering for Multi-behavior Recommendation *Daqing Wu, Xiao Luo, Zeyu Ma, Chong Chen, Pengfei Wang, Minghua Deng, and Jinwen Ma*	◆
18:33	**DM630**	**Cold Item Integration in Deep Hybrid Recommenders via Tunable Stochastic Gates** *Oren Barkan, Roy Hirsch, Ori Katz, Avi Caciularu, Jonathan Weill, and Noam Koenigstein*	◆
18:45	**DM722**	MetaEDL: Meta Evidential Learning For Uncertainty-Aware Cold-Start Recommendations *Krishna Neupane, Ervine Zheng, and Qi Yu*	◆
Session 6B: Adversarial Learning **Stream B – 17:30-18:52**			
17:30	**DM277**	Structure-Aware Stabilization of Adversarial Robustness with Massive Contrastive Adversaries *Shuo Yang, Zeyu Feng, Pei Du, Bo Du, and Chang Xu*	★
17:47	**DM474**	Graph-based Adversarial Online Kernel Learning with Adaptive Embedding *Peng Yang, Xiaoyun Li, and Ping Li*	★
18:04	**DM626**	Towards Stochastic Neural Network via Feature Distribution Calibration *Hao Yang, Min Wang, Yun Zhou, and Yongxin Yang*	◆
18:16	**DM637**	Adversarial Learning of Balanced Triangles for Accurate Community Detection on Signed Networks *Yoonsuk Kang, Woncheol Lee, Yeon-Chang Lee, Kyungsik Han, and Sang-Wook Kim*	◆
18:28	**DM831**	Detecting and Mitigating Test-time Failure Risks via Model-agnostic Uncertainty Learning *Preethi Lahoti, Krishna Gummadi, and Gerhard Weikum*	◆
18:40	**DM842**	Attacking Similarity-Based Sign Prediction *Michał T. Godziszewski, Marcin Waniek, Yulin Zhu, Kai Zhou, Talal Rahwan, and Tomasz P. Michalak*	◆
Session 6C: Graph Neural Networks 2 **Stream C – 17:30-18:59**			
17:30	**DM834**	Hypergraph Convolutional Network for Group Recommendation *Renqi Jia, Xiaofei Zhou, Linhua Dong, and Shirui Pan*	★
17:47	**DM803**	Heterogeneous Graph Neural Network with Distance Encoding *Houye Ji, Pan Li, Chuan Shi, and Cheng Yang*	◆
17:59	**DM733**	MC-RGCN: A Multi-Channel Recurrent Graph Convolutional Network to Learn High-Order Social Relations for Diffusion Prediction *Ningbo Huang, Gang Zhou, Mengli Zhang, and Meng Zhang*	◆
18:11	**DM854**	HyperTeNet: Hypergraph and Transformer-based Neural Network for Personalized List Continuation *Vijaikumar M, Deepesh Hada, and Shirish Shevade*	◆
18:23	**DM957**	Learnable Structural Semantic Readout for Graph Classification *Dongha Lee, Su Kim, Seonghyeon Lee, Chanyoung Park, and Hwanjo Yu*	◆
18:35	**DM1123**	GQNAS: Graph Q Network for Neural Architecture Search *Yijian Qin, Xin Wang, Peng Cui, and Wenwu Zhu*	◆
18:47	**DM418**	StarGAT: Star-Shaped Hierarchical Graph Attentional Network for Heterogeneous Network Representation Learning *Wen-Zhi Li, Ling Huang, Chang-Dong Wang, and Yuxin Ye*	◆

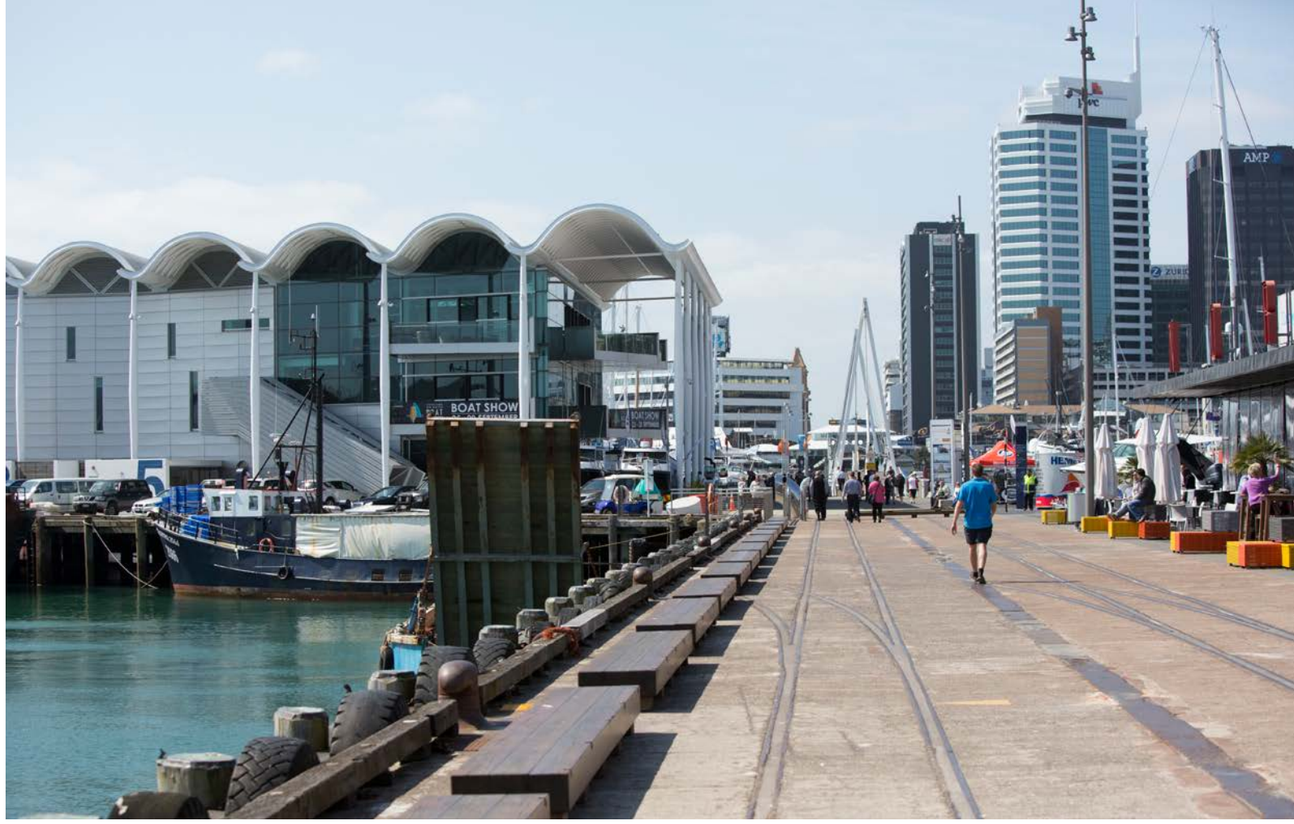
★ **Long** ◆ **short**

Session 6D: Active and Semi-Supervised Learning Stream D – 17:30-18:57			
17:30	**DM389**	Flexible, Robust, Scalable Semi-supervised Learning via Reliability Propagation *Chen Huang, Liangxu Pan, Qinli Yang, Hongliang Wang, and Junming Shao*	★
17:47	**DM843**	PARWiS: Winner determination from Active Pairwise Comparisons under a Shoestring Budget *Dev Sheth and Arun Rajkumar*	★
18:04	**DM1008**	Finding Age Path of Self-Paced Learning *Zhou Zhai, Bin Gu, Li Xiang, and Heng Huang*	★
18:21	**DM1154**	Incomplete Multi-view Multi-label Active Learning *Chuanwei Qu, Kuangmeng Wang, Hong Zhang, Guoxian Yu, and Carlotta Domeniconi*	◆
18:33	**DM298**	Self-supervised Universal Domain Adaptation with Adaptive Memory Separation *Ronghang Zhu and Sheng Li*	◆
18:45	**DM608**	Multimodal N-best List Rescoring with Weakly Supervised Pre-training in Hybrid Speech Recognition *Yuanfeng Song, Xiaoling Huang, Xuefang Zhao, Di Jiang, and Raymond Chi-Wing Wong*	◆
Session 6E: Matrix and Tensor **Stream E – 17:30-18:57**			
17:30	**DM868**	MASCOT: A Quantization Framework for Efficient Matrix Factorization in Recommender Systems *Yunyong Ko, Jae-Seo Yu, Hong-Kyun Bae, Yongjun Park, Dongwon Lee, and Sang-Wook Kim*	★
17:47	**DM544**	Anomaly Detection with Prototype-Guided Discriminative Latent Embeddings *Yuandu Lai, Yahong Han, and Yaowei Wang*	★
18:04	**DM724**	Discriminative Additive Scale Loss for Deep Imbalanced Classification and Embedding *Zhao Zhang, Weiming Jiang, Yang Wang, Qiaolin Ye, Mingbo Zhao, Mingliang Xu, and Meng Wang*	★
18:21	**DM1044**	Overfitting Avoidance in Tensor Train Factorization and Completion: Prior Analysis and Inference *Le Xu, Cheng Lei, Ngai Wong, and Yik-Chung Wu*	◆
18:33	**DM589**	Communication Efficient Tensor Factorization for Decentralized Healthcare Networks *Jing Ma, Qiuchen Zhang, Jian Lou, Li Xiong, Joyce Ho, and Sivasubramanium Bhavani*	◆
18:45	**DM641**	Summarizing User-Item Matrix By Group Utility Maximization *Yongjie Wang, Ke Wang, Cheng Long, and Chunyan Miao*	◆
Special Session in Diversity: Women in Data Science and Machine Learning *Speaker: Prof Kerri Mengersen Session Chair: Prof Richi Nayak* **15:30-17:00**			
15:30		Let's talk about uncertainty Speaker: *Prof Kerri Mengersen Session Chair: Prof Richi Nayak*	
16:30		Women in Data Science and Machine Learning Networking Panel Panel Chair: *Prof Richi Nayak*	

Tutorials Program

Date (NZT)	Title	Organisers
16:00-18:30, Wednesday 8 December 2021	Wikimedia Visual Resources and its Application to Neural Visual Recommender Systems	*Denis Parra (PUC Chile), Antonio Ossa-Guerra (PUC Chile), Manuel Cartagena (PUC Chile), Patricio Cerda-Mardini (PUC Chile and MindsDB), Felipe del Rio (PUC Chile), Isidora Palma (PUC Chile), Diego Saez-Trumper (Wikimedia Foundation), and Miriam Redi (Wikimedia Foundation)*
16:00-18:30, Wednesday 8 December 2021	Complement, Composite and Context: The 3C-Law to Build Multidomain Recommender Systems	*Liang Hu, Shoujin Wang (Macquarie University), Qi Zhang, Dora D. Liu, and Longbing Cao (University of Technology Sydney)*
18:30-21:00, Wednesday 8 December 2021	Roles in Networks - Foundations, Methods and Applications	*Yulong Pei (Eindhoven University of Technology), Pengfei Jiao (Tianjin University), Xuan Guo (Tianjin University), George Fletcher (Eindhoven University of Technology), and Mykola Pechenizkiy (Eindhoven University of Technology)*
15:30-20:30, Thursday 9 December 2021	Automated Taxonomy Discovery and Exploration	*Jiaming Shen (Univ. of Illinois at Urbana-Champaign), Xiaotao Gu (Univ. of Illinois at Urbana-Champaign), Yu Meng (Univ. of Illinois at Urbana-Champaign), and Jiawei Han (Univ. of Illinois at Urbana-Champaign)*
Tutorial Chairs: ***Wei Liu**, University of Sydney, Australia; **Katharina Morik**, TU Dortmund, Germany*		





PhD Forum Program

07/12/2021 (Half-Day Session)		
14:00-14:05	Opening Remarks	
14:05-15:00	Keynote	Speaker: Sibo Wang, The Chinese University of Hong Kong Title: Managing and Mining with Massive Graphs Session Chair: Lisi Chen
15:00-16:30	Students' Presentations	long: 12-15min, short: 7-10min, Compress 5 long + 3 short into 81~105 mins in total before coffee 1hr break at 4:30pm-5:30pm
		(long) Jun Wang, Jinghua Tan, Jiujiu Chen, and Hanlei Jin, "**A Knowledge-aware and Time-sensitive Financial News Recommendation System Based on Firm Relation Derivation**"
		(long) Tianyi LI, Philippe Goupil, Josiane Mothe, and Olivier Teste, "**Early Detection of Atmospheric Turbulence for Civil Aircraft: A Data Driven Approach**"
		(long) Bo Xiao, Wuguannan Yao, Xiang Zhou, and Xianhua Peng, "**Optimal Option Hedging with Policy Gradient**"
		(long) Boris Belyakov and Dmitry Sizykh, "**Deep Reinforcement Learning Task for Portfolio Construction**"
		(long) Timo Spinde, "**An Interdisciplinary Approach for the Automated Detection and Visualization of Media Bias in News Articles**"
		(short) Smilla Hinterreiter, "**A Gamified Approach To Automatically Detect Biased Wording And Train Critical Reading**"
		(short) Pierre Haritz, Lukas Pfahler, Thomas Liebig, and Helena Kotthaus, "**Self-Supervised Source Code Annotation from Related Research Papers**"
		(short) YongKyung Oh and Sungil Kim, "**Multi-channel Convolution Neural Network for Gas Mixture Classification**"
		Session Chair: Sibylle Hess and Meng-Fen Chiang
16:30-18:30	Coffee Break	Gather Town 2hr Coffee Break

Keynote speakers

Masashi Sugiyama
RIKEN Center for Advanced Intelligence Project/
The University of Tokyo

Towards Robust Machine Learning: Weak Supervision, Noisy Labels, and Beyond

Abstract
When machine learning systems are trained and deployed in the real world, we face various types of uncertainty. For example, training data at hand may contain insufficient information, label noise, and bias. In this talk, I will give an overview of our recent advances in robust machine learning.

Bio
Masashi Sugiyama received Doctor of Engineering in Computer Science from Tokyo Institute of Technology, Japan in 2001. Experiencing Assistant Professor and Associate Professor at Tokyo Institute of Technology, he became Professor at the University of Tokyo in 2014. Since 2016, he has been concurrently serving as Director of RIKEN Center for Advanced Intelligence Project. His research interests include theories and algorithms of machine learning. He received the Japan Academy Medal in 2017 for his series of machine learning research. He coauthored Machine Learning in Non-Stationary Environments (MIT Press, 2012), Density Ratio Estimation in Machine Learning (Cambridge University Press, 2012), Statistical Reinforcement Learning (Chapman and Hall, 2015), Introduction to Statistical Machine Learning (Morgan Kaufmann, 2015), and Machine Learning from Weak Supervision (MIT Press, to appear).

Svetha Venkatesh
Deakin University's Applied Artificial Intelligence Institute (A^2I^2)

Sample Efficient AI with Applications in Health Care and Advanced Manufacturing

Abstract
From Guglielmo Marconi who developed the radio telegraph to the Wright brothers who invented flying machines, curiosity driven experimentation has powered human innovation. Such experimental optimisation remains an integral part of the Scientific Method. This time-honoured method needs a step change to accelerate scientific innovation because this iterative method quickly hits limits.

To speed-up innovation, it is imperative to expand the capability of experimental optimisation and improve its efficiency. This talk will demonstrate how sample efficient AI can be used to deliver this acceleration in experimental design. I will discuss how the methods can be applied widely, focusing on health and advanced manufacturing particularly in settings where data is scare and experimentation is expensive. In healthcare, I show how these methods can accelerate the design of clinical/health trials to efficiently determine the optimal strategy. In advanced manufacturing, I will show how it can be applied broadly from inventing new materials and alloys to accelerating industrial processes.

The second part of the talk will focus on the new machine learning innovations that have been formulated and solved to advance experimental design. These include incorporating experimental design constraints such as process constraints, transferring knowledge from pervious experiments or experimenter "hunches", and high dimensional Bayesian optimisation so that the number of experimental control variables can be increased.

Bio
Svetha Venkatesh is an ARC Australian Laureate Fellow, Alfred Deakin Professor and a co- Director of Applied Artificial Intelligence Institute (A2I2) at Deakin University. She was elected a Fellow of the International Association of Pattern Recognition in 2004 for contributions to formulation and extraction of semantics in multimedia data, a Fellow of the Australian Academy of Technological Sciences and Engineering in 2006, and a Fellow of the Australian Academy of Science in 2021 for ground-breaking research and contributions that have had clear impact. In 2017, Professor Venkatesh was appointed an Australian Laureate Fellow, the highest individual award the Australian Research Council can bestow.

Professor Venkatesh and her team have tackled a wide range of problems of societal significance, including the critical areas of autism, security and aged care. The outcomes have impacted the community and evolved into publications, patents, tools and spin-off companies. This includes 650+ publications, three full patents, one start-up company (iCetana) and two significant products (TOBY Playpad, Virtual Observer).

Professor Venkatesh has tackled complex pattern recognition tasks by drawing inspiration and models from widely diverse disciplines, integrating them into rigorous computational models and innovative algorithms. Her main contributions have been in the development of theoretical frameworks and novel applications for analysing large scale, multimedia data. This includes development of several Bayesian parametric and non-parametric models, solving fundamental problems in processing multiple channel, multi-modal temporal and spatial data.

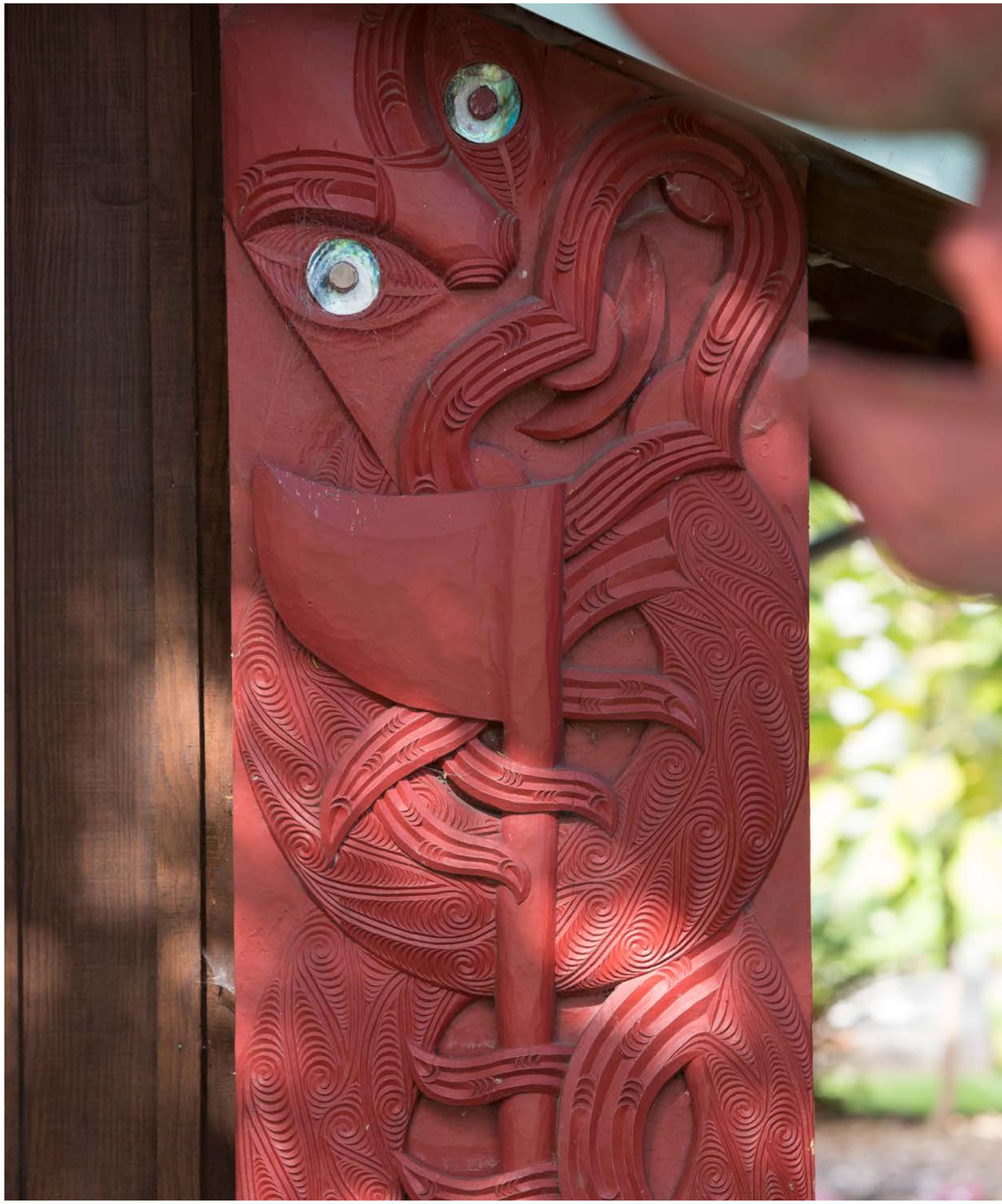
Jian Pei
Simon Fraser University

Towards Trustworthy Data Science: Interpretability, Fairness and Marketplaces

Abstract
We believe data science and AI will change the world. No matter how smart and powerful an AI model we can build, the ultimate testimony of the success of data science and AI is users' trust. How can we build trustworthy data science? At the level of user-model interaction, how can we convince users that a data analytic result is trustworthy? At the level of group-wise collaboration for data science and AI, how can we ensure that the parties and their contributions are recognized fairly, and establish trust between the outcome (e.g., a model built) of the group collaboration and the external users? At the level of data science participant eco-systems, how can we effectively and efficiently connect many participants of various roles and facilitate the connection among supplies and demands of data and models?

In this talk, I will brainstorm possible directions to the above questions in the context of an end-to-end data science pipeline. To strengthen trustworthy interactions between models and users, I will advocate exact and consistent interpretation of machine learning models. Our recent results show that exact and consistent interpretations are not just theoretically feasible, but also practical even for API-based AI services. To build trust in collaboration among multiple participants in coalition, I will review some progress in ensuring fairness in federated learning, including fair assessment of contributions and fairness enforcement in collaboration outcome. Last, to address the need of trustworthy data science eco-systems, I will review some latest efforts in building data and model marketplaces and preserving fairness and privacy. Through reflection I will discuss some challenges and opportunities in building trustworthy data science for possible future work.

Bio
Jian Pei is a Professor in the School of Computing Science at Simon Fraser University. He is a well known leading researcher in the general areas of data science, big data, data mining, and database systems. His expertise is on developing effective and efficient data analysis techniques for novel data intensive applications, and transferring his research results to products and business practice. He is recognized as a Fellow of the Royal Society of Canada (Canada's national academy), the Canadian Academy of Engineering, the Association of Computing Machinery (ACM) and the Institute of Electrical and Electronics Engineers (IEEE). He is one of the most cited authors in data mining, database systems, and information retrieval. Since 2000, he has published one textbook, two monographs and over 300 research papers in refereed journals and conferences, which have been cited extensively by others. His research has generated remarkable impact substantially beyond academia. For example, his algorithms have been adopted by industry in production and popular open source software suites. Jian Pei also demonstrated outstanding professional leadership in many academic organizations and activities. He was the editor-in-chief of the IEEE Transactions of Knowledge and Data Engineering (TKDE) in 2013-16, the chair of the Special Interest Group on Knowledge Discovery in Data (SIGKDD) of the Association for Computing Machinery (ACM) in 2017-2021, and a general co-chair or program committee co-chair of many premier conferences. He maintains a wide spectrum of industry relations with both global and local industry partners. He is an active consultant and coach for industry on enterprise data strategies, healthcare informatics, network security intelligence, computational finance, and smart retail. He received many prestigious awards, including the 2017 ACM SIGKDD Innovation Award, the 2015 ACM SIGKDD Service Award, the 2014 IEEE ICDM Research Contributions Award, the British Columbia Innovation Council 2005 Young Innovator Award, an NSERC 2008 Discovery Accelerator Supplements Award (100 awards cross the whole country), an IBM Faculty Award (2006), a KDD Best Application Paper Award (2008), an ICDE Influential Paper Award (2018), a PAKDD Best Paper Award (2014), a PAKDD Most Influential Paper Award (2009), and an IEEE Outstanding Paper Award (2007).



Panel

Data Mining with Far Fewer Labels: Pretraining, Knowledge and Unsupervised Learning

Min Zhang
Tsinghua University

Dr. Min Zhang is a tenured associate professor in the Department of Computer Science & Technology (DCST), Tsinghua University, and is the vice director of the AI lab, DCST. She specializes in Web search, personalized recommendation, and user modeling. Currently she serves as the Editor-in-Chief of ACM Transaction on Information Systems (TOIS), and ACM SIGIR Executive Committee member, and PC chair or Area Chairs of top conferences such as SIGIR, WSDM, theWebConf, etc. She was awarded IBM Global Faculty Award in 2020. She also owns 12 patents and cooperates with many international and domestic enterprises.

Lina Yao
University of New South Wales (UNSW)

Lina Yao is an Associate Professor with School of Computer Science and Engineering at University of New South Wales (UNSW), Australia. Her research interest lies in machine learning, and its applications in recommender systems, activity recognition, Internet of Things, and Brain Computer Interface. She is serving as the Associate Editor for ACM Transactions on Sensor Networks (ACM TOSN) and Knowledge-based Systems (KNOSYS).

Kevin Knight
Didi Chuxing

Kevin Knight is Chief Scientist for Natural Language Processing at Didi Chuxing. He received a PhD in computer science from Carnegie Mellon University and a bachelor's degree from Harvard University. Prof. Knight's research interests include machine translation, dialog processing, natural language generation, automata theory, and decipherment of historical manuscripts. He taught computer science at USC for 25 years and has authored over 150 NLP research papers. Prof. Knight also co-wrote the widely-adopted textbook "Artificial Intelligence" (McGraw-Hill). In 2001, he co-founded the machine translation company Language Weaver, Inc. Dr. Knight served as President of the Association for Computational Linguistics (ACL) in 2011, as General Chair for ACL conferences in 2005 and 2016, and as Program Chair in 2001 and 2020. He is a Fellow of the ACL and the Association for the Advancement of Artificial Intelligence (AAAI).

Bowen Zhou
JD.com

Dr. Bowen Zhou is a Senior Vice President of JD Group (JD.Com), President of Foundational Technologies of JD Tech & Director of JD AI Research. He has also held a number of top technology executive positions at JD Group including the Chair of JD Technology Committee and President of JD Cloud & AI. Dr. Zhou has decades of experience in natural language technology, machine learning, and artificial intelligence. Prior to joining JD.com in 2017, Dr. Zhou held several key leadership positions during his 15-year tenure at IBM Research's headquarters. He most recently served as Director of the AI Foundations Lab at IBM Research in New York, Chief Scientist of IBM Watson Group, and a Distinguished Engineer of IBM. As an IEEE Fellow, Dr. Zhou has published over 100 peer-reviewed papers at top international conferences and journals, with over 11,200 citations, including a few of his pioneering works were cited more than 1500 times each. He previously served as a member of the IEEE Speech and Language Technical Committee, Associate Editor of IEEE Transactions, ICASSP Area Chair (2011-2015), ACL, and NAACL Area Chair.

Imed Zitouni
Google

Imed Zitouni is a director of engineering at Google leading R&D efforts on NLU to enhance and enable capabilities for Semantic Search and Answers leveraging the power of the Knowledge Graph. Before joining Google in 2019, Imed was at Microsoft leading the NLU and the Conversation Engine effort for the digital assistant Cortana. Prior to joining Microsoft in 2012, Imed was a Senior Researcher at IBM Watson for almost a decade, working on several NLP initiatives including the Watson initiative around informatics extraction, language modeling and automatic machine translation. Prior to IBM, Imed was a research member at Bell Labs, Lucent Technologies, for half a dozen years working on speech recognition, language modeling and spoken dialog systems. Imed received his M.Sc. and Ph.D. from the University-of-Nancy1 and INRIA in France. He also obtained a MEng degree in computer science from ENSI in Tunisia.

Imed is the Editor-in-Chief of ACM Transactions on Asian and Low-Resources Language Processing. He is a senior member of IEEE, served as a board member of the IEEE Speech and Language Processing Technical Committee, and was the associate editor of IEEE Trans. on Audio, Speech and Language Processing. He also served as the Information Officer of the ACL SIG on Semitic-Languages and served as chair as well as reviewing-committee-member of several conferences and journals in the area of machine learning, information retrieval and natural language technologies. Imed is the author/co-author of two books, half-dozen book-chapters as well as more than 120 patents and scientific papers.

Moderator:
Michael Witbrock
University of Auckland

Michael Witbrock is a Professor at the School of Computer Science, The University of Auckland. He is leading the research group, the Strong AI Lab, at the intersection of machine learning, reasoning and natural language understanding, with an additional focus on achieving the best social and civilisational impacts of increasingly powerful AI.

Special Sessions

Special Session in Diversity: Women in Data Science and Machine Learning

Kerri Mengersen
Queensland University of Technology

Let's talk about uncertainty

Abstract
Statistical and machine learning analyses provide a wide array of insights to expand knowledge and enhance data-focused decision making. This has led to a surge in demand for, and indeed understanding of, estimates and predictions based on these analyses, not only by scientists but also policy makers and the public. A concomitant issue that is perhaps less well addressed or understood is the uncertainty of these values. In this presentation, I will touch on the many forms of uncertainty in statistical and machine learning analyses, and discuss by example some of our approaches to quantification and communication of uncertainty. The argument will be made that while we have come a long way, the expression and use of uncertainty in science and decision-making is still in its infancy. Indeed, I will posit that SML uncertainty is so important for all members of the community that it could be considered the '4th R' after reading, writing and arithmetic. I will preface the talk with a brief discussion about women in data science, and I will mention our ambition at QUT to create a Centre of Excellence to amplify our research.

Bio

Kerrie Mengersen graduated in 1985 with a Bachelor of Arts (Honours Class 1), majoring in Mathematics (Statistics) and Computing, and received her PhD in Mathematical Statistics in 1989 from the University of New England, New South Wales. Her PhD thesis was on the topic of ranking and selection under the supervision of Professor Eve Bofinger, one of the pioneer female university researchers in regional Australia.

Following graduation, she was recruited to a commercial statistical consulting company, which provided her with strong experience in a wide range of statistical methods in the context of diverse applied problems. Her career since then has been characterised by a dual focus of engaging with and developing new statistical methodology motivated by, and motivating, challenging statistical applications.

In 2016, QUT awarded the title of Distinguished Professor to Professor Kerrie Mengersen in recognition of her outstanding achievements, both nationally and internationally, in mathematics and statistical research. Distinguished Professor Mengersen is acknowledged to be one of the leading researchers in her discipline.

In 2016 Professor Mengersen also received two more prestigious awards: the Statistical Society of

Australia's Pitman Medal, the highest award presented by the Society and the first time it has been presented to a woman, and the Research Excellence award by the Cooperative Research Centre for Spatial Analysis (CRCSI).

In 2018 Professor Mengersen was elected a Fellow of the Australian Academy of Science (AAS); a Fellow of the Academy of Social Sciences in Australia (ASSA); and an Invited Fellow of the Queensland Academy of Arts and Sciences (QAAS)

Special Session: PhD Forum

Sibo Wang
The Chinese University of Hong Kong

Managing and Mining with Massive Graphs

Abstract
In this talk, I will share some of my research on massive graph data analysis and processing that can be applied to many real scenarios in industries. The topics include personalized PageRank query processing and graph embedding.

Bio

Sibo Wang is an Assistant Professor in the Department of Systems Engineering and Engineering Management at The Chinese University of Hong Kong since Dec 2018. He received his B.E. in Software Engineering in 2011 from Fudan University and his Ph.D. in Computer Science in 2016 from Nanyang Technological University. His main research area is database and data mining, especially big data analytics and processing, graph mining and graph representation learning. Most of his research works have been published in top conferences like SIGMOD, VLDB, and SIGKDD. His professional services include Workshop Chair at ICDE 2022 Conference and Local Organization Chair of ADC 2018, Program Committee for VLDB 2020-2021, SIGKDD 2019-2021, WWW2020-2022, ICDE 2021-2022, IJCAI 2020, AAAI 2021-2022, PAKDD 2018-2022, DASFAA 2019-2022. He also served as a peer reviewer in journals like TODS, TKDE, VLDBJ, TOIS, and TKDD.

Job matching program highlights

The IEEE ICDM 2021 Job Matching Program provides an opportunity for ICDM attendees to apply for data science jobs posted by the sponsors and other attendees of the conference.

A job matching session will be held on Day 1 of the conference, to provide a platform for registered job seekers and recruiters to meet, and for general conference attendees to network.

Please refer to our Job Matching page (https://icdm2021.auckland.ac.nz/job-matching/) to register your interest prior to the conference, or contact icdm2021@auckland.ac.nz for more details or queries.

Virtual social event highlights

We look forward to welcoming you to the IEEE ICDM virtual conference. While we regret that the COVID pandemic prevented us from holding the conference in Auckland, we are excited about the opportunities of holding an engaging virtual conference.

The University of Auckland developed this year's virtual platform. It features live polls, video chat and Q&A, group networking capabilities, and much more.

Based on our experience with previous virtual events, we have designed simple and effective interactions to make the event as engaging as possible.

- Dedicated coffee break rooms after every stream. Attendees can immediately join the room after the stream has finished and continue the discussion.
- Speed networking during the coffee break. Attendees can meet with other researchers without the need to choose a room or approach people directly actively.
- Meeting hub. If you want to catch up with other attendees at any time, you can always join the meeting hub!
- PhD Forum. The attendees of the PhD forum will have access to gather town during the event.

Virtual social event program

On the first day of the conference, we will have a one-hour coffee break. It will include dedicated coffee break rooms and speed networking events.

In the following days of the conference, we will have three coffee breaks of 30 minutes each. Each of these breaks will feature the dedicated coffee break rooms and speed networking events.

Awards

Awards:

- **2021 IEEE ICDM Research Contributions Award**
- **2021 IEEE ICDM Outstanding Service Award**
- **2021 IEEE ICDM 10-Year Highest-Impact Paper Award**
- **2021 IEEE ICDM Best Paper Awards**
- **2021 IEEE ICDM Student Travel Awards**

Awards Committee:

- **Xia Ning,** The Ohio State University (Awards Chair)
- **Pauli Miettinen,** University of Eastern Finland, and **James Bailey,** The University of Melbourne (Program Chairs)
- **Xindong Wu,** Mininglamp Academy of Sciences (Steering Committee Chair)
- **Three established data mining researchers:**
  - **Jian Pei**, Simon Fraser University
  - **Diane Joyce Cook,** Washington State University
  - **Eamonn Keogh,** University of California Riverside

Organising Committee

General Chairs

Yun Sing Koh
The University of Auckland, New Zealand

Dacheng Tao
The University of Sydney, Australia

Program Chairs

Pauli Miettinen
University of Eastern Finland, Finland

James Bailey
The University of Melbourne, Australia

Steering Committee Chair

Xindong Wu
Mininglamp Academy of Sciences, China

Local Arrangement Chair

Joerg Wicker
The University of Auckland, New Zealand

Finance Chair

Gillian Dobbie
The University of Auckland, New Zealand

Sponsorship Chairs

Kaiqi Zhao
The University of Auckland, New Zealand

Philippe Fournier-Viger
Harbin Institute of Technology, China

Eva García-Martín
Ekkono Solutions, Sweden

Tutorial Chairs

Wei Liu
University of Technology Sydney, Australia

Katharina Morik
TU Dortmund, Germany

Worshop Chairs

Bing Xue
Victoria University of Wellington, New Zealand

Mykola Pechenizkiy
Eindhoven University of Technology, Netherlands

Panel Chairs

Chenliang Li
Wuhan University, China

Michael Witbrock
The University of Auckland, New Zealand

Diversity and Inclusion Chair

Richi Nayak

Queensland University of Technology, Australia

PhD Forum Chairs

Lisi Chen
Inception Institute of Artificial Intelligence (IIAI), UAE

Meng-Fen Chiang
The University of Auckland, New Zealand

Sibylle Hess
Eindhoven University of Technology, Netherlands

Online Experience/ Virtual Chair

Heitor Gomes
University of Waikato, New Zealand

Newcomers Chairs

Huan Liu
Arizona State University, US

Katerina Taskova
The University of Auckland, New Zealand

Job Matching Chairs

Albert Bifet
University of Waikato, New Zealand

Lin Liu
University of South Australia, Australia

Reproducibility Chairs

Dragi Kocev
Jožef Stefan Institute, Slovenia

Jacob Montiel
University of Waikato, New Zealand

Award Committee Chair

Xia Ning
The Ohio State University, US

Publicity Chair

Diana Benavides Prado
The University of Auckland, New Zealand

Conference Sponsors

IEEE ICDM 2021 is sponsored by

IEEE

IEEE COMPUTER SOCIETY

Gold Sponsor

TWO SIGMA

Keynote Sponsor

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

SCIENCE
SCHOOL OF COMPUTER SCIENCE

Main Conference Sponsor

Google

Student Attendance Award Sponsor

NSF

Best Student Paper, Best Paper, 10-Year Highest Impact Paper Awards Sponsor

IEEE COMPUTER SOCIETY
TCII
Technical Committee on Intelligent Informatics