



**The 23rd Pacific–Asia Conference
on Knowledge
Discovery and Data Mining**

**14–17 April 2019
Macau SAR, China**

**CONFERENCE
PROGRAM BOOKLET**

Hosted by



澳門大學
UNIVERSIDADE DE MACAU
UNIVERSITY OF MACAU



The 23rd Pacific-Asia Conference on Knowledge Discovery and Data Mining

14-17 April 2019

The Parisian Macao Macau SAR, China



Table of Contents

General Chairs' Preface.....	02
Program Chairs' Preface.....	03
Organizing Committee.....	04
Conference Program.....	07
Keynote Speakers.....	21
Most Influential Paper Award Presentation.....	25
Tutorials.....	26
Workshops.....	38
General Information.....	41
Floorplan.....	43

General Chairs' Preface

On behalf of the Organizing Committee, it is our great pleasure to welcome you to Macau, China for the 23rd Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2019). Since its first edition in 1997, PAKDD has well established as one of the leading international conferences in the areas of data mining and knowledge discovery. This year, after its four previous editions in Beijing (1999), Hong Kong (2001), Nanjing (2007), and Shenzhen (2011), PAKDD was held in China for the fifth time in the fascinating city of Macau, during April 14–17, 2019.

First of all, we are very grateful to the many authors who submitted their work to the PAKDD 2019 main conference, satellite workshops, and data mining contest. We were delighted to feature three outstanding keynote speakers: Dr. Jennifer Neville from Purdue University, Professor Hui Xiong from Baidu Inc., and Professor Josep Domingo-Ferrer from Universitat Rovira i Virgili. The conference program was further enriched with six high-quality tutorials, five workshops on cutting-edge topics, and one data mining contest on AutoML for lifelong machine learning.

We would like to express our gratitude to the contributions of the SPC members, PC members, and external reviewers, led by the PC Co-Chairs, Zhiguo Gong and Min-Ling Zhang. We are also very thankful to the other Organizing Committee members: Workshop Co-Chairs, Hady W. Lauw and Leong Hou U., Tutorial Co-Chairs, Bob Durrant and Yang Yu, Contest Co-Chairs, Hugo Jair Escalante and Wei-Wei Tu, Publicity Co-Chairs, Yi Cai, Xiangnan Kong, Gang Li, and Yasuo Tabei, Proceedings Chair, Sheng-Jun Huang, and Local Arrangements Chair, Andrew Jiang.

We wish to extend our special thanks to Honorary Co-Chairs, Hiroshi Motoda and Lionel M. Ni, for their enlightening support and advice throughout the conference organization.

We appreciate the hosting organization University of Macau, and our sponsors Macao Convention & Exhibition Association, Intel, Baidu, for their institutional and financial support of PAKDD 2019. We also appreciate the Fourth Paradigm Inc., ChaLearn, Microsoft, and Amazon for sponsoring the PAKDD 2019 data mining contest. We feel indebted to the PAKDD Steering Committee for its continuing guidance and sponsorship of the paper award and student travel awards.

Last but not least, our sincere thanks go to all the participants and volunteers of PAKDD 2019—there would be no conference without you. We hope you enjoy PAKDD 2019 and your time in Macau, China.

Qiang Yang
Zhi-Hua Zhou

Program Chairs' Preface

It is our great pleasure to introduce the proceedings of the 23rd Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2019). The conference 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, visualization, decision-making systems, and the emerging applications.

We received 567 submissions to PAKDD 2019 from 46 countries and regions all over the world, noticeably with submissions from North America, South America, Europe, and Africa. The large number of submissions and high diversity of submission demographics witness the significant influence and reputation of PAKDD. A rigorous double-blind reviewing procedure was ensured via the joint efforts of the entire Program Committee consisting of 55 Senior Program Committee (SPC) members and 379 Program Committee (PC) members.

The PC Co-Chairs performed an initial screening of all the submissions, among which 25 submissions were desk rejected due to the violation of submission guidelines. For submissions entering the double-blind review process, each one received at least three quality reviews from PC members or in a few cases from external reviewers (with 78.5% of them receiving four or more reviews). Furthermore, each valid submission received one meta-review from the assigned SPC member who also led the discussion with the PC members. The PC Co-Chairs then considered the recommendations and meta-reviews from SPC members, and looked into each submission as well as its reviews and PC discussions to make the final decision. For borderline papers, additional reviews were further requested and thorough discussions were conducted before final decisions.

As a result, 137 out of 567 submissions were accepted, yielding an acceptance rate of 24.1%. We aim to be strict with the acceptance rate, and all the accepted papers are presented in a total of 20 technical sessions. Each paper was allocated 15 minutes for oral presentation and 2 minutes for Q/A. The conference program also featured three keynote speeches from distinguished data mining researchers, five cutting-edge workshops, six comprehensive tutorials, and one dedicated data mining contest session.

We wish to sincerely thank all SPC members, PC members and external reviewers for their invaluable efforts in ensuring a timely, fair, and highly effective paper review and selection procedure. We hope that readers of the proceedings will find that the PAKDD 2019 technical program was both interesting and rewarding.

Zhiguo Gong
Min-Ling Zhang

Honorary Co-Chairs

- Hiroshi Motoda (Osaka University, Japan)
 - Lionel M. Ni (University of Macau, China)
-

General Co-Chairs

- Qiang Yang (Hong Kong University of Science and Technology, China)
 - Zhi-Hua Zhou (Nanjing University, China)
-

Program Committee Co-Chairs

- Zhiguo Gong (University of Macau, China)
 - Min-Ling Zhang (Southeast University, China)
-

Workshop Co-Chairs

- Hady W. Lauw (Singapore Management University, Singapore)
 - Leong Hou U (University of Macau, China)
-

Tutorial Co-Chairs

- Bob Durrant (University of Waikato, New Zealand)
- Yang Yu (Nanjing University, China)

Contest Co-Chairs

- Hugo Jair Escalante (INAOE, Mexico)
 - Wei-Wei Tu (The Fourth Paradigm Inc., China)
-

Publicity Co-Chairs

- Yi Cai (South China University of Technology, China)
 - Xiangnan Kong (Worcester Polytechnic Institute, USA)
 - Gang Li (Deakin University, Australia)
 - Yasuo Tabei (RIKEN, Japan)
-

Proceedings Chair

- Sheng-Jun Huang (Nanjing University of Aeronautics and Astronautics, China)
-

Local Arrangement Chair

- Andrew Jiang (Macao Convention & Exhibition Association, China)

Conference Venue : Level 5 Parisian Grand Ballroom, The Parisian Macao

Presentation format : Each paper will have an oral presentation in 17 minutes (including 2-minutes Q/A)

April 14, 2019 (Sunday)				
8:00-17:00	Registration@Foyer			
Room	7204	7304	7404-7405	7504-7505
8:30-10:00	Workshop-PAISI 14th Pacific Asia Workshop on Intelligence and Security Informatics	Workshop - BDM 8th Workshop on Biologically-inspired Techniques for Knowledge Discovery and Data Mining	Tutorial 2 IoT Big Data Stream Mining	Tutorial 4 Statistical Machine Learning of Large, Sparse, and Multi-source Data
10:00-10:30	Coffee Break			
10:30-12:00	Workshop-PAISI 14th Pacific Asia Workshop on Intelligence and Security Informatics	Workshop - BDM 8th Workshop on Biologically-inspired Techniques for Knowledge Discovery and Data Mining	Tutorial 2 IoT Big Data Stream Mining	Tutorial 4 Statistical Machine Learning of Large, Sparse, and Multi-source Data
12:00-13:30	Lunch @ Level 1, Le Buffet, the Parisian Macao			
13:30-15:30	Workshop-LDRC Learning Data Representation for Clustering	Workshop- WeL Weakly Supervised Learning: Progress and Future	Tutorial 5 Building and Evaluating a Production-ready Recommendation System	Tutorial 6 Blockchain Data Analytics
15:30-16:00	Coffee Break			
16:00-18:00	Workshop - DLKT 1st Pacific-Asia Workshop on Deep Learning for Knowledge Transfer		Tutorial 5 Building and Evaluating a Production-ready Recommendation System	Tutorial 6 Blockchain Data Analytics
18:00-20:00	Reception (Belon, Level 31 @Banyan Tree Macao Hotel, Galaxy Macau)			

April 15, 2019 (Monday)				
8:00-17:00	Registration@Foyer			
8:30-9:00	Conference Opening (7201A-7303)			
9:00-10:00	Keynote Speech Towards Relational AI – the good, the bad, and the ugly of learning over networks (by Dr. Jennifer L. Neville) <i>Chair: Zhiguo Gong (7201A-7303)</i>			
10:00-10:20	Coffee Break			
Room	7204	7304	7404-7405	7504-7505
10:20-12:20	Session 1A Classification and Supervised Learning <i>Chair: Joao Gama</i>	Session 1B Text and Opinion Mining (I) <i>Chair: Cheng-Wei Wu</i>	Session 1C Deep Learning Models and Applications (I) <i>Chair: Tianrui Li</i>	Session 1D Spatio-Temporal and Stream Data Mining <i>Chair: Leong Hou U</i>
12:20-13:30	Lunch @ Level 1, Le Buffet, the Parisian Macao			
13:30-15:30	Session 2A Healthcare, Bioinformatics and Related Topics <i>Chair: Ke-Jia Chen</i>	Session 2B Clustering and Anomaly Detection <i>Chair: Lazhar Labiod</i>	Tutorial 1 Causally Regularized Machine Learning	Tutorial 3 Knowledge Graph Embedding and Applications
15:30-15:50	Coffee Break			
15:50-17:40	Session 3A Factor and Tensor Analysis <i>Chair: Khoa Nguyen</i>	Session 3B Sequential Pattern Mining <i>Chair: Riccardo Guidotti</i>	Tutorial 1 Causally Regularized Machine Learning	Tutorial 3 Knowledge Graph Embedding and Applications
April 16, 2019 (Tuesday)				
8:00-17:00	Registration@Foyer			
9:00-10:00	Keynote Speech Talent Analytics: Prospects and Opportunities (by Prof. Hui Xiong) <i>Chair: Ee-Peng Lim (7201A-7303)</i>			
10:00-10:20	Coffee Break			

Room	7204	7304	7404-7405	7504-7505
10:20-12:20	Session 4A Representation Learning and Embedding (I) <i>Chair: Senzhang Wang</i>	Session 4B Weakly Supervised Learning <i>Chair: Kuan-Ting Lai</i>	Session 4C Recommender System <i>Chair: Murat Kantarcioglu</i>	Session 4D Social Network and Graph Mining <i>Chair: Philippe Fournier-Viger</i>
12:20-13:30	Lunch@ Level 1, Le Buffet, the Parisian Macao			
13:30-18:00	Excursion (Gathering Point: Foyer Reception Counter)			
18:00-21:00	Banquet (7401-7503, Parisian Grand Ballroom, The Parisian Macao)			
April 17, 2019 (Wednesday)				
8:00-17:00	Registration@Foyer			
9:00-10:00	Keynote Speech Empowering Subjects, Users and Controllers when Anonymizing Big Data for Knowledge Discovery and Data Mining (by Prof. Josep Domingo-Ferrer) <i>Chair: Takashi Washio (7201A-7303)</i>			
10:00-10:20	Coffee Break			
Room	7204	7304	7404-7405	7504-7505
10:20-12:20	Session 5A Data Pre-Processing and Feature Selection <i>Chair: Zhidong Li</i>	Session 5B Text and Opinion Mining (II) <i>Chair: Sheng-Jun Huang</i>	Session 5C Mining Unstructured and Semi-structured Data <i>Chair: Hady Wirawan Lauw</i>	Session 5D Behavioral Data Mining <i>Chair: Cuneyt Gurcan Akcore</i>
12:20-13:30	Lunch @ Level 1, Le Buffet, the Parisian Macao			
13:30-14:10	PAKDD Most Influential Paper Award Presentation On Extreme Support Vector Machine (by Prof. Qing He) <i>Chair: Min-Ling Zhang (7201A-7303)</i>			
14:10-15:10	PAKDD 2019 Challenge Presentation and Award <i>Chair: Wei-Wei Tu (7201A-7303)</i>			
15:10-15:30	Coffee Break			
15:30-17:30	Session 6A Deep Learning Models and Applications (II) <i>Chair: Min-Ling Zhang</i>	Session 6B Visual Data Mining <i>Chair: Yu-Feng Li</i>	Session 6C Representation Learning and Embedding (II) <i>Chair: Xiangliang Zhang</i>	Session 6D Knowledge Graph and Interpretable Data Mining <i>Chair: Katerina Hlavackova-Schindler</i>
17:30-17:40	Conference Closing			

Session 1A - Classification and Supervised Learning

Chair: Joao Gama

Multitask Learning for Sparse Failure Prediction

Simon Luo, Victor W. Chu, Zhidong Li, Yang Wang, Jianlong Zhou, Fang Chen, Raymond K. Wong

Cost Sensitive Learning in the Presence of Symmetric Label Noise

Sandhya Tripathi, N. Hemachandra

Semantic Explanations in Ensemble Learning

Md Zahidul Islam, Jixue Liu, Lin Liu, Jiuyong Li, Wei Kang

Latent Gaussian-Multinomial Generative Model for Annotated Data

Shuoran Jiang, Yarui Chen, Zhifei Qin, Jucheng Yang, Tingting Zhao, Chuanlei Zhang

Investigating Neighborhood Generation Methods for Explanations of Obscure Image Classifiers

Riccardo Guidotti, Anna Monreale, Leonardo Cariaggi

On Calibration of Nested Dichotomies

Tim Leathart, Eibe Frank, Bernhard Pfahringer, Geoffrey Holmes

Ensembles of Nested Dichotomies with Multiple Subset Evaluation

Tim Leathart, Eibe Frank, Bernhard Pfahringer, Geoffrey Holmes

Session 1B - Text and Opinion Mining (I)

Chair: Cheng-Wei Wu

Topic-Level Bursty Study for Bursty Topic Detection in Microblogs

Yakun Wang, Zhongbao Zhang, Sen Su, Muhammad Azam Zia

Adaptively Transfer Category-classifier for Handwritten Chinese Character

Recognition

Yongchun Zhu, Fuzhen Zhuang, Jingyuan Yang, Xi Yang, Qing He

Syntax-aware Representation for Aspect Term Extraction

Jingyuan Zhang, Guangluan Xu, Xinyi Wang, Xian Sun, Tinglei Huang

Short Text Similarity Measurement Based on Coupled Semantic Relation and Strong Classification Features

Huifang Ma, Wen Liu, Zhixin Li, Xianghong Lin

A Novel Hybrid Sequential Model for Review-based Rating Prediction

Yuanquan Lu, Wei Zhang, Pan Lu, Jianyong Wang

Integrating Topic Model and Heterogeneous Information Network for Aspect Mining

with Rating Bias

Yugang Ji, Chuan Shi, Fuzhen Zhuang, Philip S. Yu

Dependency-Aware Attention Model for Emotion Analysis for Online News

Xue Zhao, Ying Zhang, Xiaojie Yuan

Session 1C - Deep Learning Models and Applications (I)

Chair: Tianrui Li

Semi-interactive Attention Network for Answer Understanding in Reverse-QA

Qing Yin, Guan Luo, Xiaodong Zhu, Qinghua Hu, Ou Wu

Neural Network based Popularity Prediction by Linking Online Content with Knowledge Bases

Wayne Xin Zhao, Hongjian Dou, Yuanpei Zhao, Daxiang Dong, Ji-Rong Wen

Passenger Demand Forecasting with Multi-Task Convolutional Recurrent Neural Networks

Lei Bai, Lina Yao, Salil S. Kanhere, Zheng Yang, Jing Chu, Xianzhi Wang

Accurate Identification of Electrical Equipment from Power Load Profiles

Ziyi Wang, Chun Li, Lin Shang

Similarity-aware Deep Attentive Model for Clickbait Detection

Manqing Dong, Lina Yao, Xianzhi Wang, Boualem Benatallah, Chaoran Huang

Topic Attentional Neural Network for Abstractive Document Summarization

Hao Liu, Hai-Tao Zheng, Wei Wang

Parameter Transfer Unit for Deep Neural Networks

Yinghua Zhang, Yu Zhang, Qiang Yang

Session 1D - Spatio-Temporal and Stream Data Mining

Chair: Leong Hou U

FGST: Fine-Grained Spatial-Temporal based Regression for Stationless Bike Traffic Prediction

Hao Chen, Senzhang Wang, Zengde Deng, Xiaoming Zhang, Zhoujun Li

Customer Segmentation Based on Transactional Data Using Stream Clustering

Matthias Carnein, Heike Trautmann

Spatio-Temporal Event Detection from Multiple Data Sources

Aman Ahuja, Ashish Baghudana, Wei Lu, Edward A. Fox, Chandan K. Reddy

Discovering All-chain Set in Streaming Time Series

Shaopeng Wang, Ye Yuan, Hua Li

Hawkes Process with Stochastic Triggering Kernel

Feng Zhou, Yixuan Zhang, Zhidong Li, Xuhui Fan, Yang Wang, Arcot Sowmya, Fang Chen

Concept Drift Based Multi-Dimensional Data Streams Sampling Method

Ling Lin, Xiaolong Qi, Zhirui Zhu, Yang Gao

Spatial-temporal Multi-Task Learning for Within-field Cotton Yield Prediction

Long H. Nguyen, Jiazhen Zhu, Zhe Lin, Hanxiang Du, Zhou Yang, Wenxuan Guo, Fang Jin

Session 2A - Healthcare, Bioinformatics and Related Topics

Chair: Ke-Jia Chen

Time-dependent Survival Neural Network for Remaining Useful Life Prediction

Jianfei Zhang, Shengrui Wang, Lifei Chen, Gongde Guo, Rongbo Chen, Alain Vanasse

ACNet: Aggregated Channels Network for Automated Mitosis Detection

Kaili Cheng, Jiarui Sun, Xuesong Chen, Yanbo Ma, Mengjie Bai, Yong Zhao

Attention-based Hierarchical Recurrent Neural Network for Phenotype Classification

Nan Xu, Yanyan Shen, Yanmin Zhu

Identifying Mobility of Drug Addicts with Multilevel Spatial-Temporal Convolutional Neural Network

Canghong Jin, Haoqiang Liang, Dongkai Chen, Zhiwei Lin, Minghui Wu

MC-eLDA: Towards Pathogenesis Analysis in Traditional Chinese Medicine by Multi-Content Embedding LDA

Ying Zhang, Wendi Ji, Haofen Wang, Xiaoling Wang, Jin Chen

Enhancing the Healthcare Retrieval with a Self-adaptive Saturated Density Function

Yang Song, Wenxin Hu, Liang He, Liang Dou

CytoFA: Automated Gating of Mass Cytometry Data via Robust Skew Factor Analyzers

Sharon X. Lee

Session 2B - Clustering and Anomaly Detection

Chair: Lazhar Labiod

Consensus Graph Learning for Incomplete Multi-view Clustering

Wei Zhou, Hao Wang, Yan Yang

Beyond Outliers and on to Micro-clusters: Vision-guided Anomaly Detection

Wenjie Feng, Shenghua Liu, Christos Faloutsos, Bryan Hooi, Huawei Shen, Xueqi Cheng

Clustering of Mixed-type Data Considering Concept Hierarchies

Sahar Behzadi, Nikola S. Müller, Claudia Plant, Christian Böhm

DMNAED: A Novel Framework Based on Dynamic Memory Network for Abnormal Event Detection in Enterprise Networks

Xueshuang Ren, Liming Wang

NeoLOD: A Novel Generalized Coupled Local Outlier Detection Model Embedded Non-IID Similarity Metric

Fan Meng, Yang Gao, Jing Huo, Xiaolong Qi, Shichao Yi

Dynamic Anomaly Detection Using Vector Autoregressive Model

Yuemeng Li, Aidong Lu, Xintao Wu, Shuhan Yuan

A Convergent Differentially Private k-Means Clustering Algorithm

Zhigang Lu, Hong Shen

Session 3A - Factor and Tensor Analysis

Chair: Khoa Nguyen

Online Data Fusion Using Incremental Tensor Learning

Nguyen Lu Dang Khoa, Hongda Tian, Yang Wang, Fang Chen

Co-clustering from Tensor Data

Rafika Boutalbi, Lazhar Labiod, Mohamed Nadif

A Data-Aware Latent Factor Model for Web Service QoS Prediction

Di Wu, Xin Luo, Mingsheng Shang, Yi He, Guoyin Wang, Xindong Wu

Keyword Extraction with Character-level Convolutional Neural Tensor Networks

Zhe-Li Lin, Chuan-Ju Wang

Neural Variational Matrix Factorization with Side Information for Collaborative Filtering

Teng Xiao, Hong Shen

Variational Deep Collaborative Matrix Factorization for Social Recommendation

Teng Xiao, Hui Tian, Hong Shen

Session 3B - Sequential Pattern Mining

Chair: Riccardo Guidotti

Efficiently Finding High Utility-Frequent Itemsets using Cutoff and Suffix Utility

R. Uday Kiran, T. Yashwanth Reddy, Philippe Fournier-Viger, Masashi Toyoda, P. Krishna Reddy, Masaru Kitsuregawa

How Much Can A Retailer Sell? Sales Forecasting on Tmall

Chaochao Chen, Ziqi Liu, Jun Zhou, Xiaolong Li, Yuan Qi, Yujing Jiao, Xingyu Zhong
Hierarchical LSTM: Modeling Temporal Dynamics and Taxonomy in Location-Based Mobile Check-Ins

Chun-Hao Liu, Da-Cheng Juan, Xuan-An Tseng, Wei Wei, Yu-Ting Chen, Jia-Yu Pan, Shih-Chieh Chang

Recovering DTW Distance between Noise Superposed NHPP

Yongzhe Chang, Zhidong Li, Bang Zhang, Ling Luo, Arcot Sowmya, Yang Wang, Fang Chen

ATNet: Answering Cloze-Style Questions via Intra-attention and Inter-attention

Chengzhen Fu, Yuntao Li, Yan Zhang

Parallel Mining of Top-k High Utility Itemsets in Spark In-Memory Computing Architecture

Chun-Han Lin, Cheng-Wei Wu, JianTao Huang, Vincent S. Tseng

Session 4A - Representation Learning and Embedding (I)

Chair: Senzhang Wang

AAANE: Attention-based Adversarial Autoencoder for Multi-scale Network Embedding

Lei Sang, Min Xu, Shengsheng Qian, Xindong Wu

NEAR: Normalized Network Embedding with Autoencoder for Top-K Item Recommendation

Dedong Li, Aimin Zhou, Chuan Shi

Ranking Network Embedding via Adversarial Learning

Quanyu Dai, Qiang Li, Liang Zhang, Dan Wang

Selective Training: A Strategy for Fast Backpropagation on Sentence Embeddings

Jan Neerbek, Peter Dolog, Ira Assent

Extracting Keyphrases from Research Papers Using Word Embeddings

Wei Fan, Huan Liu, Suge Wang, Yuxiang Zhang, Yaocheng Chang

Sequential Embedding Induced Text Clustering, a Non-parametric Bayesian Approach

Tiehang Duan, Qi Lou, Sargur N. Srihari, Xiaohui Xie

SSNE: Status Signed Network Embedding

Chunyu Lu, Pengfei Jiao, Hongtao Liu, Yaping Wang, Hongyan Xu, Wenjun Wang

Session 4B - Weakly Supervised Learning

Chair: Kuan-Ting Lai

Robust Semi-Supervised Multi-Label Learning by Triple Low-Rank Regularization

Lijuan Sun, Songhe Feng, Gengyu Lyu, Congyan Lang

Multi-class Semi-supervised Logistic I-RELIEF Feature Selection Based on Nearest Neighbor

Baige Tang, Li Zhang

Effort-Aware Tri-Training for Semi-Supervised Just-in-Time Defect Prediction

Wenzhou Zhang, Weiwei Li, Xiuyi Jia

One Shot Learning with Margin

Xianchao Zhang, Jinlong Nie, Linlin Zong, Hong Yu, Wenxin Liang

DeepReview: Automatic Code Review using Deep Multi-Instance Learning

Heng-Yi Li, Shu-Ting Shi, Ferdian Thung, Xuan Huo, Bowen Xu, Ming Li, David Lo

Multi-label Active Learning with Error Correcting Output Codes

Ningzhao Sun, Jincheng Shan, Chenping Hou

Dynamically Weighted Multi-View Semi-Supervised Learning for CAPTCHA

Congqing He, Li Peng, Yuquan Le, Jiawei He

Session 4C - Recommender System

Chair: Murat Kantarcioglu

A Novel Top-N Recommendation Approach Based on Conditional Variational Auto-Encoder

Bo Pang, Min Yang, Chongjun Wang

Jaccard Coefficient-based Bi-clustering and Fusion Recommender System for Solving Data Sparsity

Jiangfei Cheng, Li Zhang

A Novel KNN Approach for Session-based Recommendation

Huifeng Guo, Ruiming Tang, Yunming Ye, Feng Liu, Yuzhou Zhang

A Contextual Bandit Approach to Personalized Online Recommendation via Sparse Interactions

Chenyu Zhang, Hao Wang, Shangdong Yang, Yang Gao

Heterogeneous Item Recommendation for the Air Travel Industry

Zhicheng He, Jie Liu, Guanghui Xu, Yalou Huang

A Minimax Game for Generative and Discriminative Sample Models for Recommendation

Zongwei Wang, Min Gao, Xinyi Wang, Junliang Yu, Junhao Wen, Qingyu Xiong

RNE: A Scalable Network Embedding for Billion-scale Recommendation

Jianbin Lin, Daixin Wang, Lu Guan, Yin Zhao, Binqiang Zhao, Jun Zhou, Xiaolong Li, Yuan Qi

Session 4D - Social Network and Graph Mining

Chair: Philippe Fournier-Viger

Graph Compression with Stars

Faming Li, Zhaonian Zou, Jianzhong Li, Yingshu Li

Neighbor-based Link Prediction with Edge Uncertainty

Chi Zhang, Osmar Zaiane

Inferring Social Bridges that Diffuse Information Across Communities

Pei Zhang, Ke-Jia Chen, Tong Wu

Learning Pretopological Spaces to Extract Ego-centered Communities

Gaetan Caillaut, Guillaume Cleuziou, Nicolas Dugué

EigenPulse: Detecting Surges in Large Streaming Graphs with Row Augmentation

Jiabao Zhang, Shenghua Liu, Wenjian Yu, Wenjie Feng, Xueqi Cheng

TPLP: Two-Phase Selection Link Prediction for Vertex in Graph Streams

Yang Xiao, Hong Huang, Feng Zhao, Hai Jin

Robust Temporal Graph Clustering for Group Record Linkage

Charini Nanayakkara, Peter Christen, Thilina Ranbaduge

Session 5A - Data Pre-Processing and Feature Selection

Chair: Zhidong Li

Learning Diversified Features for Object Detection via Multi-region Occlusion Example Generating

Junsheng Liang, Zhiqiang Li, Hongchen Guo

HATDC: A Holistic Approach for Time Series Data Repairing

Xiaojie Liu, Guangxuan Song, Xiaoling Wang

Double Weighted Low-Rank Representation and Its Efficient Implementation

Jianwei Zheng, Kechen Lou, Ping Yang, Wanjun Chen, Wanliang Wang

Exploring Dual-Triangular Structure for Efficient R-initiated Tall-skinny QR on GPGPU

Nai-Yun Cheng, Ming-Syan Chen

Efficient Autotuning of Hyperparameters in Approximate Nearest Neighbor Search

Elias Jaasaari, Ville Hyvonen, Teemu Roos

An Accelerator of Feature Selection Applying a General Fuzzy Rough Model

Peng Ni, Suyun Zhao, Hong Chen, Cuiping Li

Text Feature Extraction and Selection Based on Attention Mechanism

Longxuan Ma, Lei Zhang

Session 5B - Text and Opinion Mining (II)

Chair: Sheng-Jun Huang

Multi-task Learning for Target-dependent Sentiment Classification

Divam Gupta, Kushagra Singh, Soumen Chakrabarti, Tanmoy Chakraborty

SC-NER: a Sequence-to-Sequence Model with Sentence Classification for Named Entity Recognition

Yu Wang, Yun Li, Ziyi Zhu, Bin Xia, Zheng Liu

BAB-QA: A New Neural Model for Emotion Detection in Multi-Party Dialogue

Zilong Wang, Zhaohong Wan, Xiaojun Wan

Unsupervised User Behavior Representation for Fraud Review Detection with Cold-Start Problem

Qian Li, Qiang Wu, Chengzhang Zhu, Jian Zhang, Wentao Zhao

Gated Convolutional Encoder-Decoder for Semi-Supervised Affect Prediction

Kushal Chawla, Sopan Khosla, Niyati Chhaya

Complaint Classification Using Hybrid-Attention GRU Neural Network

Shuyang Wang, Bin Wu, Bai Wang, Xuesong Tong

Session 5C - Mining Unstructured and Semi-structured Data

Chair: Hady Wirawan Lauw

Context-Aware Dual-Attention Network for Natural Language Inference

Kun Zhang, Guangyi Lv, Enhong Chen, Le Wu, Qi Liu, C. L. Philip Chen

Best from Top k versus Top 1: Improving Distant Supervision Relation Extraction with Deep Reinforcement Learning

Yaocheng Gui, Qian Liu, Tingming Lu, Zhiqiang Gao

Towards One Reusable Model for Various Software Defect Mining Tasks

Heng-Yi Li, Ming Li, Zhi-Hua Zhou

User Preference-Aware Review Generation

Wei Wang, Hai-Tao Zheng, Hao Liu

Mining Cluster Patterns in XML Corpora via Latent Topic Models of Content and Structure

Gianni Costa, Riccardo Ortale

A Large-scale Repository of Deterministic Regular Expression Patterns and Its Applications

Haiming Chen, Yeting Li, Chunmei Dong, Xinyu Chu, Xiaoying Mou, Weidong Min

Determining the Impact of Missing Values on Blocking in Record Linkage

Imrul Chowdhury Anindya, Murat Kantarcioglu, Bradley Malin

Session 5D - Behavioral Data Mining

Chair: Cuneyt Gurcan Akcore

Bridging the Gap between Research and Production with CODE

Yiping Jin, Dittaya Wanvarie, Phu T.V. Le

Distance2Pre: Personalized Spatial Preference for Next Point-of-Interest Prediction

Qiang Cui, Yuyuan Tang, Shu Wu, Liang Wang

Using Multi-objective Optimization to Solve the Long Tail Problem in Recommender System

Jiaona Pang, Jun Guo, Wei Zhang

Event2Vec: Learning Event Representations Using Spatial-Temporal Information for Recommendation

Yan Wang, Jie Tang

Maximizing Gain Over Flexible Attributes in Peer to Peer Marketplaces

Abolfazl Asudeh, Azade Nazi, Nick Koudas, Gautam Das

An Attentive Spatio-Temporal Neural Model for Successive Point of Interest Recommendation

Khoa D. Doan, Guolei Yang, Chandan K. Reddy

Mentor Pattern Identification from Product Usage Logs

Ankur Garg, Aman Kharb, Yash H. Malviya, J. P. Sagar, Atanu R. Sinha, Iftikhar Ahamath Burhanuddin, Sunav Choudhary

Session 6A - Deep Learning Models and Applications (II)

Chair: Min-Ling Zhang

EFCNN: A Restricted Convolutional Neural Network for Expert Finding

Yifeng Zhao, Jie Tang, Zhengxiao Du

CRESA: A Deep Learning Approach to Competing Risks, Recurrent Event Survival Analysis

Garima Gupta, Vishal Sunder, Ranjitha Prasad, Gautam Shroff

Long-Term Traffic Time Prediction Using Deep Learning with Integration of Weather Effect

Chih-Hsin Chou, Yu Huang, Chian-Yun Huang, Vincent S. Tseng

Arrhythmias Classification by Integrating Stacked Bidirectional LSTM and Two-dimensional CNN

Fan Liu, Xingshe Zhou, Jinli Cao, Zhu Wang, Hua Wang, Yanchun Zhang

An Efficient and Resource-Aware Hashtag Recommendation Using Deep Neural Networks

David Kao, Kuan-Ting Lai, Ming-Syan Chen

Dynamic Student Classification on Memory Networks for Knowledge Tracing

Sein Minn, Michel C. Desmarais, Feida Zhu, Jing Xiao, Jianzong Wang

Targeted Knowledge Transfer for Learning Traffic Signal Plans

Nan Xu, Guanjie Zheng, Kai Xu, Yanmin Zhu, Zhenhui Li

Session 6B - Visual Data Mining

Chair: Yu-Feng Li

AggregationNet: Identifying Multiple Changes Based on Convolutional Neural Network in Bitemporal Optical Remote Sensing Images

Qiankun Ye, Xiankai Lu, Hong Huo, Lihong Wan, Yiyao Guo, Tao Fang

Detecting Micro-expression Intensity Changes from Videos Based on Hybrid Deep CNN

Selvarajah Thuseethan, Sutharshan Rajasegarar, John Yearwood

A Multi-Scale Recalibrated Approach for 3D Human Pose Estimation

Ziwei Xie, Hailun Xia, Chunyan Feng

Gossiping the Videos: An Embedding-based Generative Adversarial Framework for Time-sync Comments Generation

Guangyi Lv, Tong Xu, Qi Liu, Enhong Chen, Weidong He, Mingxiao An, Zhongming Chen

Self-paced Robust Deep Face Recognition with Label Noise

Pengfei Zhu, Wenya Ma, Qinghua Hu

Multi-Constraints-Based Enhanced Class-specific Dictionary Learning for Image Classification

Ze Tian, Ming Yang

Discovering Senile Dementia from Brain MRI Using Ra-DenseNet

Xiaobo Zhang, Yan Yang, Tianrui Li, Hao Wang, Ziqing He

Session 6C - Representation Learning and Embedding (II)

Chair: Xiangliang Zhang

On the Network Embedding in Sparse Signed Networks

Ayan Kumar Bhowmick, Koushik Meneni, Bivas Mitra

MSNE: A Novel Markov Chain Sampling Strategy for Network Embedding

Ran Wang, Yang Song, Xinyu Dai

Auto-Encoder Based Co-Training Multi-View Representation Learning

Run-kun Lu, Jian-wei Liu, Yuan-fang Wang, Hao-jie Xie, Xin Zuo

Robust Semi-Supervised Representation Learning for Graph-Structured Data

Lan-Zhe Guo, Tao Han, Yu-Feng Li

Characterizing the SOM Feature Detectors under Various Input Conditions

Macario O. Cordel II, Arnulfo P. Azcarraga

PCANE: Preserving Context Attributes for Network Embedding

Danhao Zhu, Xinyu Dai, Kaijia Yang, Jiajun Chen, Yong He

A Novel Framework for Node/Edge Attributed Graph Embedding

Guolei Sun, Xiangliang Zhang

Session 6D - Knowledge Graph and Interpretable Data Mining

Chair: Katerina Hlavackova-Schindle

Granger Causality for Heterogeneous Processes

Sahar Behzadi, Katerina Hlaváková-Schindler, Claudia Plant

Knowledge Graph Embedding with Order Information of Triplets

Jun Yuan, Neng Gao, Ji Xiang, Chenyang Tu, Jingquan Ge

Knowledge Graph Rule Mining via Transfer Learning

Pouya Ghiasnezhad Omran, Zhe Wang, Kewen Wang

Knowledge Base Completion by Inference from Both Relational and Literal Facts

Zhichun Wang, Yong Huang

EMT: A Tail-Oriented Method for Specific Domain Knowledge Graph Completion

Yi Zhang, Zhijuan Du, Xiaofeng Meng

An Interpretable Neural Model with Interactive Stepwise Influence

Yin Zhang, Ninghao Liu, Shuiwang Ji, James Caverlee, Xia Hu

Multivariate Time Series Early Classification with Interpretability Using Deep Learning and Attention Mechanism

En-Yu Hsu, Chien-Liang Liu, Vincent S. Tseng

Dr. Jennifer Neville (Purdue University)

Title : Towards Relational AI -- the good, the bad, and the ugly of learning over networks

Chair: Zhiguo Gong

Abstract: In the last 20 years, there has been a great deal of research on machine learning methods for graphs, networks, and other types of relational data. By moving beyond the independence assumptions of more traditional ML methods, relational models are now able to successfully exploit the additional information that is often observed in relationships among entities. Specifically, network models are able to use relational information to improve predictions about user interests, behavior, and interactions, particularly when individual data is sparse. The tradeoff however, is that the heterogeneity, partial-observability, and interdependence of large-scale network data can make it difficult to develop efficient and unbiased methods, due to several algorithmic and statistical challenges. In this talk, I will discuss these issues while surveying several general approaches used for relational learning in large-scale social and information networks. In addition, to reflect on the movement toward pervasive use of the models in personalized online systems, I will discuss potential implications for privacy, polarization of communities, and spread of misinformation.



Bio: Jennifer Neville is the Miller Family Chair Associate Professor of Computer Science and Statistics at Purdue University. She received her PhD from the University of Massachusetts Amherst in 2006. She is currently PC chair of the 19th SIAM International Conference on Data Mining. She was an elected member of the AAAI Executive Council from 2015-2018. In 2016 she was PC chair of the 9th ACM International Conference on Web Search and Data. In 2012, she was awarded an NSF Career Award, in 2008 she was chosen by IEEE as one of "AI's 10 to watch", and in 2007 was selected as a member of the DARPA Computer Science Study Group. Her work, which includes 100+ peer-reviewed publications with more than 7500 citations, focuses on developing machine learning and AI methods for complex relational domains, including social, information, and physical networks.

Prof. Hui Xiong

Title : Talent Analytics: Prospects and Opportunities

Chair: Ee-Peng Lim

Abstract: The big data trend has made its way to talent management. Indeed, the availability of large-scale human resources (HR) data provide unparalleled opportunities for business leaders to understand talent behaviors and generate useful talent knowledge, which in turn deliver intelligence for real-time decision making and effective people management at work. In this talk, we introduce the powerful set of innovative big data techniques developed for intelligent human resource management, such as recruiting, performance evaluation, talent retention, talent development, job matching,



team management, leadership development, and organization culture analysis. In addition, we will also demonstrate how the results of talent analytics can be used for other business applications, such as market trend analysis and financial investment.

Bio: Hui Xiong is a Professor at Rutgers University, and is currently on leave and serving as head of Business Intelligence Lab and Talent Intelligence Center at Baidu Inc. His research interests include data mining, mobile computing, and their applications in business. He has authored over 200 research articles, and co-edited or coauthored 4 books including the widely used Encyclopedia of GIS, which has been recognized as the Top 10 Most Popular Computer Science Book authored by Chinese scholars at Springer. Dr. Xiong has served as chair/co-chair for many international conferences in data mining, including a Program Co-Chair (2013) and a General Co-Chair (2015) for the IEEE International Conference on Data Mining (ICDM), and a Program Co-Chair of the Research Track (2018) and the Industry Track (2012) for the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD). Dr. Xiong's research has generated substantive impact beyond academia. He is an ACM distinguished scientist and has been honored by the 2018 Ram Charan Management Practice Award as the Grand Prix winner from the Harvard Business Review, the 2017 IEEE ICDM Outstanding Service Award, and the ICDM-2011 Best Research Paper Award.

Prof. Josep Domingo-Ferrer

Title: Empowering Subjects, Users and Controllers when
Anonymizing Big Data for Knowledge Discovery and
Data Mining



Chair: Takashi Washio

Abstract: Big data are analyzed to reveal patterns, trends and associations, especially relating to human behavior and interactions.

However, according to the European General Data Protection Regulation (GDPR), which is becoming a de facto global data protection standard, any intended uses on personally identifiable information (PII) must be clearly specified and explicitly accepted by the data subjects. Furthermore, PII cannot be accumulated for secondary use. Thus, can exploratory data uses on PII be GDPR-compliant? Hardly so.

Resorting to anonymized data sets instead of PII is a natural way around, because anonymized data fall outside the scope of GDPR. The problem is that anonymization techniques, based on statistical disclosure control and privacy models, use algorithms and assumptions from the time of small data that must be thoroughly revised, updated or even replaced to deal with big data. The following questions arise:

- How can the subject keep control of her data and the way they are anonymized in the current landscape featuring a crowd of data collectors, controllers and processors, most of them not trusted?
- Can anonymized data sets be merged to obtain big data sets? Can these be usefully explored?
- Do the current privacy models share underlying principles that can be exploited to upgrade them for big data?
- Are there universal or near-universal masking approaches that can satisfy any privacy

model?

– How to compare privacy and utility under different privacy models and masking methods?

This keynote talk will give an overview of the current state of the previous questions. It will also identify open research lines aimed at producing anonymized big data sets that can be exploited by users for exploratory analyses, while giving subjects control on their data and making the job of controllers easier.

Bio: Josep Domingo-Ferrer is a Distinguished Professor of Computer Science and an ICREA-Acadèmia Researcher at Universitat Rovira i Virgili, Tarragona, Catalonia, where he holds the UNESCO Chair in Data Privacy. He has founded and leads CYBERCAT-Center for Cybersecurity Research of Catalonia, headquartered in Tarragona and Barcelona.

His research interests are in data privacy, anonymization, statistical disclosure control, security and cryptographic protocols, with a focus on the conciliation of privacy, security and functionality. In these fields he has authored over 418 publications and 5 patents, and he has coordinated major research projects (including the European H2020 "CLARUS" project).

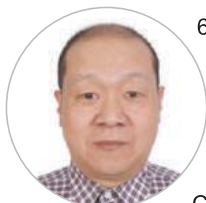
He is an IEEE Fellow, an ACM Distinguished Scientist, an Elected Member of Academia Europaea and an Elected Member of the International Statistical Institute. He has won several other distinctions, including a Google Faculty Research Award and several research prizes.

Most Influential Paper Award Presentation

Title: On Extreme Support Vector Machine

Abstract: I will introduce our series work about the paper-Extreme Support Vector Machine. The paper owned the Most Influential Paper Award in PAKDD 2018. The contents include in the following:

1. Extreme Support Vector Machine
2. A Parallel Incremental Extreme SVM Classifier
3. Clustering in extreme learning machine feature space
4. Parallel Extreme Learning Machine for Regression based on MapReduce
5. Extreme Learning Machine Ensemble Classifier for Large-scale Data
6. Learning Deep Representations via Extreme Learning Machine
7. A Fast Algorithm for Sparse Extreme Learning Machine



Bio: Qing He is a Professor at the Institute of Computing Technology, Chinese Academy of Sciences (CAS), and he is a Professor of University of Chinese Academy of Sciences (UCAS). He is also the Vice Secretary of Chinese Association for Artificial Intelligence, the Member of China Computer Federation Artificial Intelligence and Pattern Recognition Committee, the Member of Chinese Institute of Electronics and Clouding Computing and Big Data Experts Committee. He received the B.S. degree from Hebei Normal University, Shijiazhuang, P. R. C., in 1985, and the M.S. degree from Zhengzhou University, Zhengzhou, P. R. C., in 1987, both in Mathematics. He received the Ph.D. degree in 2000 from Beijing Normal University in Fuzzy Mathematics and Artificial Intelligence, Beijing, P. R. C. He had been teaching at Hebei University of Science and Technology Since 1987 to 1997. He is currently a doctoral supervisor at the Institute of Computing and Technology, CAS. His interests include data mining, machine learning, cloud computing, big data. A series of achievements gained in fuzzy information processing, fuzzy clustering, knowledge representation, text information processing, and in big data mining based on cloud computing. More than 100 papers published in journals, 40 of which are SCI Indexed, cited 3093 times by GoogleScholar, the single paper about PKmeans algorithm cited more than 600 times by GoogleScholar. Qing

He led his machine learning and data mining team (<http://mldm.ict.ac.cn/Home.html>), developed the first cloud-based parallel data mining platform in China by the end of 2008 for mining TB level actual data and achieving high-performance, low-cost data mining. He was invited made a serious technical reports in the second, third, sixth China Cloud Computing Conference. He has excellently completed a number of relevant data mining project supported by the National Natural Science Foundation and 863 projects. He proposed a series of effective data mining algorithms and multiple parallel machine learning algorithms. Multiple big data mining software such as PDMiner, COMS, CWMS and WMCS developed by his team have gotten the software copyright and practical applying to telecommunications, electricity, information security, environmental protection, the financial insurance, and dozens of companies, enterprises with the considerable economic and social benefits. He owned Wenjun Wu Artificial Intelligence Innovation Award in 2015.

Tutorial 1

Title: Causally Regularized Machine Learning

Speakers: Peng Cui, Kun Kuang, Bo Li

Abstract: Owing to the popularity of Big Data, abundant data are accumulated in various domains such as health-care and advertising. At the same time, many machine learning methods are proposed to exploit these data for prediction, aiming to estimate the future outcome in the application of interest. These methods have been proved to be successful in prediction-oriented applications. However, the lack of interpretability of most predictive algorithms makes them less attractive in many settings, especially those requiring decision making. How to improve the interpretability of learning algorithms is of paramount

importance for both academic research and real applications.

Causal inference, which refers to the process of drawing a conclusion about a causal connection based on the conditions of the occurrence of an effect, is a powerful statistical modeling tool for explanatory analysis. In this tutorial, we focus on causally regularized machine learning, where we aim to explore causal knowledge from observational data to improve the explainability and stability of machine learning algorithms. First, we will give some examples on how machine learning algorithms today focus on correlation analysis and prediction, and why those methods are not insufficient for decision making questions like What if. Then, we will give introduction to causal inference and introduce some recent data-driven approaches to explore causal knowledge or make causal inference from observational data, especially in high dimensional setting. Aiming to bridge the gap between causal inference and machine learning, we will introduce some recently causally regularized machine learning algorithms for improving the stability and interpretability of prediction. Finally, we will discussing future directions of the landscape of open research and challenges in machine learning with causal inference.

The goal of this tutorial is to bring machine learning practitioners closer to the vast field of causal inference as practiced by statisticians, epidemiologists and economists. We want to bridge the gap between causal inference and machine learning, and hope the machine learning methods in the future will be more powerful and interpretable under the help of causal inference.

Bio: Peng Cui is an Associate Professor in Tsinghua University. He got his PhD degree from Tsinghua University in 2010. His research interests include network representation learning, social dynamics modeling and human behavioral modeling. He has published more than 60 papers in prestigious conferences and journals in data mining and multimedia. His recent research won the ICDM 2015 Best Student Paper Award, SIGKDD 2014 Best Paper Finalist, IEEE ICME 2014 Best Paper Award, ACM MM12 Grand Challenge Multimodal Award, and MMM13 Best Paper Award. He is the Area Chair of ICDM 2016, ACM MM 2014-2015, IEEE ICME 2014-2015, ICASSP 2013, Associate Editor of IEEE TKDE, ACM TOMM, Elsevier Journal on Neurocomputing. He was the recipient of ACM China Rising Star Award in 2015.

Kun Kuang received the B.E. degree from the Department of Computer Science and technology of Beijing Institute of Technology in 2014. He is a fifthyear Ph.D. candidate in the Department of Computer Science and Technology of Tsinghua University. His main research interests including data mining, high dimensional inference and data driven causal model. He has published several papers on data-driven causal inference and high dimensional inference in top data mining and machine learning conferences/journals of the relevant field such as SIGKDD, AAAI, and ICDM etc.

Bo Li received a Ph.D degree in Statistics from the University of California, Berkeley, and a bachelor's degree in Mathematics from Peking University. He is an Associate Professor at the School of Economics and Management, Tsinghua University. His research interests are statistical methods for high-dimensional data, statistical causal inference and data-driven decision making. He has published widely in academic journals across a range of fields including statistics, management science and economics.

Tutorial 2

Title: IoT Big Data Stream Mining

Speakers: Joao Gama, Albert Bifet , Latifur Khan

Abstract: The challenge of deriving insights from the Internet of Things (IoT) has been recognized as one of the most exciting and key opportunities for both academia and industry. Advanced analysis of big data streams from sensors and devices is bound to become a key area of data mining research as the number of applications requiring such processing increases. Dealing with the evolution over time of such data streams, i.e., with concepts that drift or change completely, is one of the core issues in IoT stream mining. This tutorial is a gentle introduction to mining IoT big data streams. The first part introduces data stream learners for classification, regression, clustering, and frequent pattern mining. The second part deals with scalability issues inherent in IoT applications, and discusses how to mine data streams on distributed engines such as Spark, Flink, Storm, and

Samza.

Bio: Joao Gama received, in 2000, his Ph.D. degree in Computer Science from the Faculty of Sciences of the University of Porto, Portugal. He joined the Faculty of Economy where he holds the position of Associate Professor. He is also a senior researcher and vice-director of LIAAD, a group belonging to INESC TEC. He has worked in several National and European projects on Incremental and Adaptive learning systems, Ubiquitous Knowledge Discovery, Learning from Massive, and Structured Data, etc. He served as Co-Program chair of ECML'2005, DS'2009, ADMA'2009, IDA' 2011, and ECM-PKDD'2015. He served as track chair on Data Streams with ACM SAC from 2007 till 2016. He organized a series of Workshops on Knowledge Discovery from Data Streams with ECMLPKDD conferences and Knowledge Discovery from Sensor Data with ACM SIGKDD. He is author of several books in Data Mining (in Portuguese) and authored a monograph on Knowledge Discovery from Data Streams. He authored more than 250 peer-reviewed papers in areas related to machine learning, data mining, and data streams. He is a member of the editorial board of international journals ML, DMKD, TKDE, IDA, NGC, and KAIS.

Albert Bifet is Full Professor at Telecom Paris Tech and Honorary Research Associate at the WEKA Machine Learning Group at University of Waikato. Previously he worked at Huawei Noah's Ark Lab in Hong Kong, Yahoo Labs in Barcelona, University of Waikato and UPC BarcelonaTech. He is the author of a book on Adaptive Stream Mining and Pattern Learning and Mining from Evolving Data Streams. He is one of the leaders of MOA and Apache SAMOA software environments for implementing algorithms and running experiments for online learning from evolving data streams. He is serving as Co-Chair of the Industrial track of IEEE MDM 2016, ECML PKDD 2015, and as Co-Chair of BigMine (2015, 2014, 2013, 2012), and ACM SAC Data Streams Track (2016, 2015, 2014, 2013, 2012).

Latifur Khan is a full Professor (tenured) in the Computer Science department at the University of Texas at Dallas where he has been teaching and conducting research since September 2000. He received his Ph.D. and M.S. degrees in Computer Science from the University of Southern California in August of 2000, and December of 1996 respectively. He has received prestigious awards including the IEEE Technical Achievement Award for

Intelligence and Security Informatics. Dr. Khan is an ACM Distinguished Scientist and a Senior Member of IEEE. He has chaired several conferences and serves (or has served) as associate editor on multiple editorial boards including IEEE Transactions on Knowledge and Data Engineering (TKDE) journal. He has conducted tutorial sessions in prominent conferences such as ACM WWW 2005, MIS2005, DASFAA 2007, and WI 2008 ("Matching Words and Pictures Problems, Applications, and Progress") and PAKDD 2011 ("Data Stream Mining Challenges and Techniques").

Tutorial 3

Title: Knowledge Graph Embedding and Applications

Speakers: Guandong Xu, Shaowu Liu, Zili Zhou, Qian Li

Abstract: Knowledge Graph (KG), a large-scale semantic web including entities and relations between entities, formalizes the real-world facts into graph structure storage. KG Embedding is to extract the structural information of KG into a continuous vector space, so as to manipulate the KG with latent semantic information of KG components. The KG Embedding techniques can be used in the completion task of current KG which is collected from multiple World Wide Web resources but still far from complete. KG Embedding techniques can also support the decision in several Out-of-KG applications such as Relation Extraction (RE), Question Answering (QA) System and Recommender system (RS).

In this tutorial, we will systematically address the concepts and techniques of KG Construction, KG Embedding, and KG Application. Specifically, the resources and construction methods of KG will be introduced. Then, we will introduce the state-of-the-art KG Embedding approaches, e.g., Translational Distance Models, Semantic Matching Models, and Relation Path-based Models. After that, we will focus on recent advances in KG completion with Embedding. Then we discuss the start-of-the-art algorithms of applying KG in other related systems, such as applications in RE, QA, and RS. Finally, we conclude and present open research for the future.

This tutorial targets at the audiences who are conducting researches or pursuing research degrees in related areas, and who are working in the engineering domains of knowledge discovery, knowledge representation, knowledge inference, knowledge application and so on. The whole tutorial expects to last for three hours.

Bio: Guandong Xu holds the Ph.D. degree in Computer Science and is a Professor in School of Software and the Advanced Analytics Institute at University of Technology Sydney. He has authored three monographs with the Springer and the CRC Press, and 100+ journal and conference papers. His current research interests include data science and data analytics, Web data mining, behaviour analytics, recommender systems, predictive analytics, social network analysis. His research has gained grant funding from Australian and Chinese governments, and projects funded by industries. In last decade, he has had over 200+ publications including TOIS, TNNLS, TIFS, TSC, Inf Sci, IEEE-IS, IJCAI, AAAI, WWW, ICDE, ICDM, and CIKM. He is the Assistant EiC of WWW Journal, serving in the Editorial Board or as Guest Editor for several international journals.

Shaowu Liu received his doctorate from Deakin University in the field of machine learning. He is currently a postdoctoral research fellow at University of Technology Sydney. Since 2012, he has published 25+ papers in the arena of data mining and machine learning, including Machine Learning Journal (MLJ), Future Generation Computer Systems (FGCS), IEEE Transactions on SMC(C), Enterprise Information Systems (EIS), and AAAI. For the community, he has served as co-chair of ES 2016, IIP 2016, KSEM 2017, and KSEM 2019.

Zili Zhou is currently a PhD candidate of School of Software, Faculty of Engineering & IT at University of Technology Sydney. His research interests include knowledge graph representation learning, knowledge inference and knowledge graph application. Since 2017, he has published several papers in area of data mining, machine learning and knowledge graph, including international joint conference on neural networks (IJCNN), The Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) and AAAI.

Qian Li received her doctorate from Chinese Academy of Science. She is currently a postdoctoral research fellow at University of Technology Sydney. She is interested in optimization algorithms for machine learning, topological data analysis and statistical causal analysis. She has published several papers in high-impact machine learning conferences such as AAAI, CVPR and CIKM etc.

Qian Li received her doctorate from Chinese Academy of Science. She is currently a postdoctoral research fellow at University of Technology Sydney. She is interested in optimization algorithms for machine learning, topological data analysis and statistical causal analysis. She has published several papers in high-impact machine learning conferences such as AAAI, CVPR and CIKM etc.

Tutorial 4

Title: Statistical Machine Learning of Large, Sparse, and Multi-source Data

Speakers: Trong Dinh Thac Do, Longbing Cao

Abstract: With the explosion of data on the internet, social networks, finance, and e-commerce websites, modelling large and sparse datasets is highly in demand yet challenging. However, traditional methods face problems in handling these real-life datasets because of the intensive mathematical computation required. Although several statistical methods are proposed to handle a large amount of data, they are still inefficient for sparse data. It is because they perform computation on the whole datasets. However, in many real-life applications, a large amount of data are missing (i.e., sparse data). For example, in recommender systems, e.g., the Netflix data has 98.8% of the matrix entries missing. In this tutorial, we summarize various statistical methods that are effective and efficient in handling large and sparse datasets.

On the other hand, combining observable data; e.g., users' ratings on items in recommender systems, user friendship or item relations, and user/item metadata; helps to deal with cold-start problems where we have no or limited preliminary knowledge about one specific element. Accordingly, we focus on introducing our series of design for tackling these challenges on large, sparse and multi-source data. These designs perform the computation only on non-missing data and combine with efficient Bayesian inference methods.

We will show that in-depth knowledge of statistical methods for large, sparse and multi-source data creates new opportunities, directions, and means for, learning and analysis of complex and practical machine learning problems.

Bio: Trong Dinh Thac Do is currently a Ph.D. candidate in the Advanced Analytics Institute (AAI) at the University of Technology Sydney (UTS). Before joining AAI-UTS, he received his Master of Philosophy degree in Information Technology in a joint program from the Grenoble Institute of Technology (INPG) and Joseph Fourier University (UJF), France. His research interests include machine learning, statistical models and Bayesian nonparametrics. He has been working in applied machine learning applications since he was a CTO and software engineer. He has published several papers in refereed conferences and journals, such as NIPS, AAAI, IJCAI, ICDM, and International Journal of Knowledge and Information Systems (KAIS). He has served the community as program committee member of AAAI 2019, ICDM 2018, ICDM 2017, PAKDD 2019, PAKDD 2018, and reviewer of IEEE Intelligent Systems, International Journal of Knowledge and Information Systems (KAIS), and International Journal of Data Science and Analytics (JDSA).

Professor Longbing Cao holds a Ph.D. in Pattern Recognition and Intelligent Systems in Chinese Academy of Sciences and another Ph.D. in Computing Science at UTS. He has published some 300 publications, four monographs, and four edited books in recent 15 years. He has been working on data science and analytics research, education, development, and enterprise applications since he was a CTO and then joined UTS. Motivated by real-world significant and common challenges, he has been leading the team to develop theories, tools and applications for new areas including non-IID learning, actionable knowledge discovery, behaviour informatics, and complex intelligent systems, in addition to issues generally concerned in artificial intelligence, knowledge discovery, machine learning, and their enterprise applications. In data science and analytics, he initiated the Data Science and Knowledge Discovery lab at UTS in 2007, the Advanced Analytics Institute in 2011, the degrees Master of Analytics (Research) and PhD in Analytics in 2011 which are recognized as the world first degrees in data science, the IEEE Task Force on Data Science and Advanced Analytics (DSAA) and IEEE Task Force on Behavior, Economic and Soci-cultural Computing in 2013, the IEEE Conference on Data Science and Advanced Analytics (DSAA), the ACM SIGKDD Australia and New Zealand Chapter in 2014, and the International Journal of Data Science and Analytics with Springer in 2015. He served as program and general chairs of conferences such as KDD2015. In

enterprise data science innovation, his team has successfully delivered many large projects for government and business organizations in over 10 domains including finance/capital markets, banking, health and car insurance, health, telco, recommendation, online business, education, and the public sector including ATO, DFS, DHS, DIBP and IP Australia, resulting in billions of dollar savings and mentions in government, industry, media and OECD reports. In 2013, AAI was the only organization specially mentioned in the Governments first big data paper: Big Data Strategy Issues Paper. He has delivered invited and keynote speeches to over 20 conferences, guest lectures, and seminars to many universities, and tutorials to conferences including AAAI, IJCAI, and KDD.

Tutorial 5

Title: Building and evaluating a production-ready recommendation system

Speakers: Le Zhang, Graham Williams, Tao Wu, Miguel Gonzalez-Fierro, Nikhil Joglekar

Abstract: Recent decades witnessed grand proliferation of recommendation systems. The technology has brought tremendous profits to business in the verticals of retail, entertainment, etc. Research in the field has been heated from the earlier algorithm such as similarity based collaborative filtering, to the latest deep neural network based method, recommendation technologies have evolved dramatically, which, to some extent, makes it challenging to fresh practitioners to select and then customize the optimal algorithms for a specific business scenario. In addition, it is often observed that auxiliary operations such as data preprocessing, model evaluating, etc., which often play an equally significant role in the whole lifecycle of developing a recommendation system, should be but not attached with enough importance by developers.

Based on the extensive experience in productizing real-world recommendation systems, in this tutorial, the authors review and lecture key tasks in building recommendation systems, with best-practice examples to democratize the technology to every organization. An open

source Github repository, where the aforementioned topics are shaped into Jupyter notebooks and utility function codebase, will be used for hands-on practice. Several recommendation algorithms will be walked through to provide an in-depth understanding of the techniques. In general, the best-practice examples shared in the repository help developers / scientists / researchers to both quickly build production-ready recommendation system and prototype novel ideas with the provided utility functions.

Bio: Le Zhang is Data Scientist with Microsoft Cloud and Artificial Intelligence. He has extensive experience on applying the cutting-edge machine learning and artificial intelligence technology to accelerate digital transformation for enterprises and start-ups. He has helped numerous corporations to develop and build enterprise-grade scalable advanced data analytical system, for scenarios of recommendation system, smart manufacturing, predictive maintenance, financial services, e-commerce, human resource analytics, etc. He specializes in artificial intelligence and machine learning.

Graham Williams is Microsoft's Director of Data Science for Asia based in Singapore. He joined Microsoft after over 30 years of research, development, practice and teaching in Artificial Intelligence, Machine Learning, Data Mining, Analytics and Data Science. He has worked extensively in the open source ecosystem as a regular contributor to numerous open source software projects including Linux and R. Graham has authored a number of books, papers, internet resources and software packages for data scientists.

Tao Wu is a Principal Data Scientist manager with Microsoft Cloud & AI team – he leads a team with special on recommendation system technology. Tao has been part of a team that has won contests in speech recognition in China three years in a row.

Miguel Gonzalez-Fierro is a Senior Data Scientist with Microsoft Cloud & AI team, where has worked with several Microsoft enterprise customers on deep learning and recommendation projects over the last two years. Miguel is an active deep learning blogger and has founded two startups.

Nikhil Joglekar is a Program Manager with Microsoft Cloud & AI team. Nikhil focuses on accelerating customer journeys and investments in AI and machine learning projects, with a focus on recommender systems.

Tutorial 6

Title: Blockchain Data Analytics

Speakers: Cuneyt Gurcan Akcora, Yulia R. Gel, Murat Kantarcioglu

Abstract: Over the last couple of years, Bitcoin cryptocurrency and the Blockchain technology that forms the basis of Bitcoin have witnessed an unprecedented attention.

Designed to facilitate a secure distributed platform without central regulation, Blockchain is heralded as a novel paradigm that will be as powerful as Big Data, Cloud Computing, and Machine Learning.

The Blockchain technology garners an ever-increasing interest of researchers in various domains that benefit from scalable cooperation among trust-less parties. Some of these fields, such as graph analytics, have started analyzing Blockchain by using existing tools and algorithms, but have also offered novel approaches that are specifically tailored for Blockchain data. As Blockchain data analytics further proliferates, a need to glean successful approaches and to disseminate them among a diverse body of data scientists became a critical task. As an inter-disciplinary team of researchers, our aim is to fill this vital role.

We offer a holistic view on Blockchain Data Analytics. Starting with the core components of Blockchain, we will detail the state of art in Blockchain data analytics for graph, security and finance domains. Beyond the cryptocurrency aspects of Blockchain, we will outline the frontier research approaches for data analyses from Blockchain platforms, such as Ethereum, Waves and Omni.

We will share tutorial notes, collected meta-information and further reading pointers on our tutorial website.

Bio: Cuneyt Gurcan Akcora is a Postdoctoral Fellow in the Departments of Statistics and Computer Science at the University of Texas at Dallas. He received his Ph.D. from University of Insubria, Italy and his M.S. from State University of New York at Buffalo, USA.

His primary research interests are Data Science on complex networks and large scale graph analysis, with applications in social, biological, IoT and Blockchain networks. He is a Fulbright Scholarship recipient, and his research works have been published in leading conferences and journals including VLDB, ICDM and ICDE.

Yulia R. Gel is Professor in the Department of Mathematical Science at the University of Texas at Dallas. Her research interests include statistical foundation of Data Science, inference for random graphs and complex networks, time series analysis, and predictive analytics. She holds a Ph.D in Mathematics, followed by a postdoctoral position in Statistics at the University of Washington. Prior to joining UT Dallas, she was a tenured faculty member at the University of Waterloo, Canada. She also held visiting positions at Johns Hopkins University, University of California, Berkeley, and the Isaac Newton Institute for Mathematical Sciences, Cambridge University, UK. She served as a Vice President of the International Society on Business and Industrial Statistics (ISBIS), and is a Fellow of the American Statistical Association.

Murat Kantarcioglu is a Professor in the Computer Science Department and Director of the UTD Data Security and Privacy Lab at the University of Texas at Dallas and a visiting scholar at Harvard University Data Privacy Lab. He is a recipient of NSF CAREER award, and Purdue CERIAS Diamond Award for Academic excellence. His research focuses on creating technologies that can efficiently extract useful information from any data without sacrificing privacy or security. Over the years, his research has been supported by grants from NSF, AFOSR, ONR, NSA, and NIH. In addition, he has published over 160 peer reviewed papers related to data security, privacy and privacy-preserving data mining. Some of his research work has been covered by the media outlets, such as Boston Globe, ABC News, and has received three best paper awards.

Workshops

PAISI – The 14th Pacific Asia Workshop on Intelligence and Security Informatics

URL: <http://www.business.hku.hk/paisi/2019/>

Organisers

- Michael Chau (The University of Hong Kong, China)
- G. Alan Wang (Virginia Tech, United States)
- Hsinchun Chen (The University of Arizona, United States)

Contact: mchau@business.hku.hk

Intelligence and Security Informatics (ISI) is concerned with the study of the development and use of advanced information technologies and systems for national, international, and societal security-related applications. Submissions may include systems, methodology, testbed, modeling, evaluation, and policy papers. Research should be relevant to both informatics and national/international security. Topics include but are not limited to: Information Sharing and Big Data Analytics, Infrastructure Protection and Emergency Responses, Cybercrime and Terrorism Informatics and Analytics, and Enterprise Risk Management, IS Security, and Social Media Analytics. PAISI 2019 will be held in conjunction with PAKDD and will provide a stimulating forum for ISI researchers in Pacific Asia and other regions of the world to exchange ideas and report research progress. Selected PAISI 2019 papers will be published in Springer's Lecture Notes in Artificial Intelligence (LNAI) series, which is indexed by EI Compindex, ISI Proceedings, and Scopus.

WeL - PAKDD 2019 Workshop on Weakly Supervised Learning: Progress and Future

URL: <http://lamda.nju.edu.cn/conf/wel19/>

Organisers

- Yu-Feng Li (Nanjing University, China)
- Sheng-Jun Huang (Nanjing University of Aeronautics and Astronautics, China)

Contact: liyf@nju.edu.cn; huangsj@nuaa.edu.cn

The aim of the workshop is to highlight the current research related to weakly supervised learning techniques in different types of weak supervision and their applications in real

problems. The workshop will also emphasize a discussion for the major challenges for the future of weakly supervised learning and provide an opportunity to researchers for related fields such as optimization, statistical learning to get a feedback from other community.

LDRC - Learning Data Representation for Clustering

URL: <https://sites.google.com/view/pakdd-workshop-ldrc2019>

Organisers

- Lazhar Labiod (University of Paris Descartes, France)
- Mohamed Nadif (University of Paris Descartes, France)

Contact: lazhar.labiod@parisdescartes.fr

This workshop aims at discovering the recent advanced on data representation for clustering under different approaches. Thereby, the LDRC workshop is an opportunity to:

- present the recent advances in data representation based clustering algorithms,
- outline potential applications that could inspire new data representation approaches for clustering,
- explore benchmark data to better evaluate and study data representation based clustering models.

BDM – The 8th Workshop on Biologically-inspired Techniques for Knowledge Discovery and Data Mining

URL: <https://bdm19.blogs.auckland.ac.nz/>

Organisers

- Shafiq Alam (University of Auckland, New Zealand)
- Gillian Dobbie (University of Auckland, New Zealand)

Contact: shafiq@cs.auckland.ac.nz

BDM to highlight the current research related to biologically inspired techniques in different data mining domains and their implementation in real life data mining problems. The workshop will also give an opportunity to the researcher from computational intelligence and evolutionary computation to get a feedback on their work from data mining community, machine learning, and computational intelligence and other KDD community.

DLKT – The 1st Pacific Asia Workshop on Deep Learning for Knowledge Transfer

URL: <http://www.intsci.ac.cn/users/zhuangfuzhen/DLKT/2019/>

Organisers

- Fuzhen Zhuang (Institute of Computing Technology, Chinese Academy of Sciences, China)
- Deqing Wang (Beihang University, China)
- Pengpeng Zhao (Soochow University, China)

Contact: zhuangfuzhen@ict.ac.cn; dqwang@buaa.edu.cn; ppzhao@suda.edu.cn

Previous supervised learning algorithms mainly assume that there are plenty of i.i.d. sampled labeled data to train a good model for test data. However, this assumption does not always hold in real-world applications, since labeling data is time consuming and labor tedious. Furthermore, the test data are usually sampled the distribution which is different from the one of training data. The advanced algorithms based on knowledge transfer or sharing provide an effective way to handle this issue, e.g., transfer learning, multi-task learning and multi-view learning, since they either try to handle the distribution mismatch problem or the shortage of labeled data. In recent years, deep learning has been proved to have the ability to learn powerful representations for various kinds of tasks. On the one hand, although there are large amount of previous works based on knowledge transfer or sharing, there are only small amount of them applying deep learning techniques. In this workshop, we aim to bring researchers and practitioners who work on various aspects of advanced knowledge transfer algorithms based on deep learning techniques, to discuss on the state-of-the-art and open problems, to share their expertise and exchange the ideas, and to offer them an opportunity to identify new promising research directions.

Registration Desk

The PAKDD2019 registration desk will be located in the foyer outside Parisian Grand Ballroom, Level 5 of the Parisian Macao. The desk will be open at the following times.

Saturday 13 April 12:00-18:00

Sunday 14 April 08:00-17:00

Monday 15 April 08:00-17:00

Tuesday 16 April 08:00-17:00

Wednesday 17 April 08:00-17:00

Catering

Lunch @ Level 1, Le Buffet, the Parisian Macao

Social Events

Welcome Reception

Date and Time : Sunday 14 April 2019(18:00-20:00)

Place: Belon, Level 31@Banyan Tree Hotel, Galaxy Macao

Pick-up Point @Registration Counter

Transportation Information

Departure Time	Itinerary	Shuttle Bus
17:30	the Parisian Macao to Banyan Tree Hotel	Bus×2
17:45		Bus×1
18:00		Bus×1
18:15		Bus×1
19:30	Banyan Tree Hotel to the Parisian Macao	Bus×1
19:45		Bus×1
20:00		Bus×3

Conference Banquet

Date and Time: Tuesday 16 April 2019 (18:00-20:00)

Place: 7401-7503, Parisian Grand Ballroom, the Parisian Macao

Wifi Information

SSID: Sands Resorts

Password: Free

Dress Code

The conference dress code is smart casual.

Duplication/recording

Unauthorized photography, audio taping, video recording, digital taping or any other form of duplication is prohibited in the conference.

Lost and Found

Any found items may be turned into the registration counter located in the foyer. Enquires about lost items can be directed to the registration desk also.

Name Badges

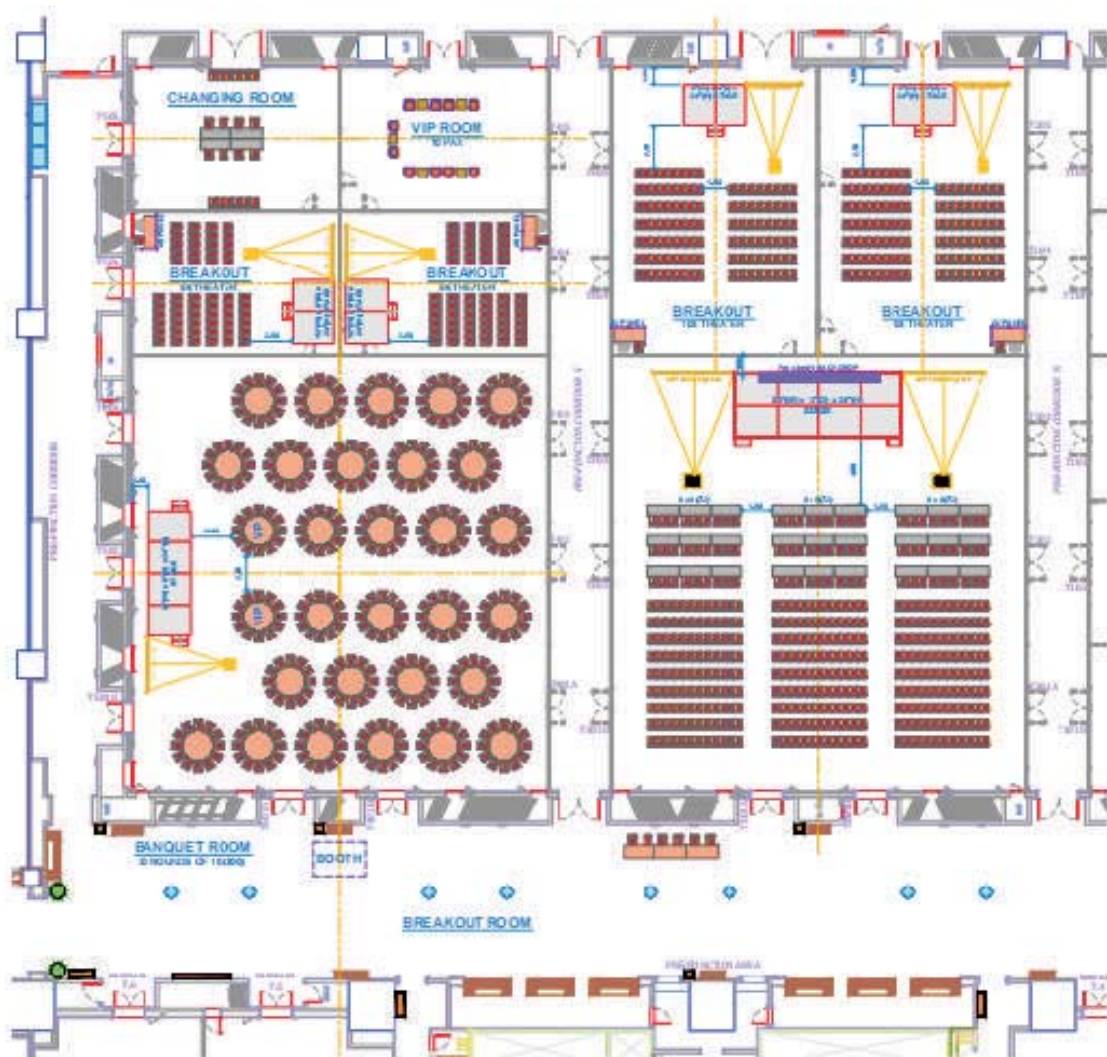
For security purposes, delegates, speakers, exhibitors and staff are required to wear their name badge to all sessions and social functions. Entrance into sessions is restricted to registered delegates only. If you misplace your name badge, please see the staff at the registration desk to arrange a replacement.

Security

Please ensure that you take all items of value with you at all times when leaving a room. Do not leave bags or laptop computers unattended.

Smoking

Smoking is not permitted indoors. Smokers must always remain at least 10m from any doorway when smoking. Fines can be imposed for smoking in prohibited places.

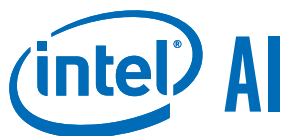




Main Sponsor



Platinum Sponsor



Name Badge Lanyard Sponsor



Publication Sponsor



Competition Sponsors

