

# **SIAM International Conference on Data Mining (SDM21)**

Online  
29 April – 1 May 2021

ISBN: 978-1-7138-3809-8

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2021) by SIAM: Society for Industrial and Applied Mathematics  
All rights reserved.

Printed with permission by Curran Associates, Inc. (2021)

For permission requests, please contact SIAM: Society for Industrial and Applied Mathematics  
at the address below.

SIAM  
3600 Market Street, 6th Floor  
Philadelphia, PA 19104-2688 USA

Phone: (215) 382-9800  
Fax: (215) 386-7999

[siambooks@siam.org](mailto:siambooks@siam.org)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

# TABLE OF CONTENTS

---

---

## Session: CP01

<b>Discovering Reliable Causal Rules</b> .....	1
Kailash Budhathoki, Mario Boley, Jilles Vreeken	
<b>Stochastic Time Series Representation for Interval Pattern Mining via Gaussian Processes</b> .....	10
Fabian Berns, Christian Becks	
<b>Efficient Permutation Testing for Significant Sequential Patterns</b> .....	19
Sam Pinxteren, Toon Calders	
<b>Learning Reliable Rules under Class Imbalance</b> .....	28
Dimitrios I. Diochnos, Theodore B. Trafalis	
<b>GPU-Accelerated Constraint-Based Causal Structure Learning for Discrete Data</b> .....	37
Christopher Hagedorn, Johannes Huegle	

## Session: CP02

<b>Discrete Listwise Collaborative Filtering for Fast Recommendation</b> .....	46
Chenghao Liu, Tao Lu, Zhiyong Cheng, Xin Wang, Jianling Sun, Steven Hoi	
<b>GraphShop: Graph-based Approach for Shop-type Recommendation</b> .....	55
Guo Yuan An, Sung Eui Yoon, Jae Yoon Kim, Lin Wang, Myoung Ho Kim	
<b>Sequence-aware Heterogeneous Graph Neural Collaborative Filtering</b> .....	64
Chen Li, Linmei Hu, Chuan Shi, Guojie Song, Yuanfu Lu	
<b>LDFeRR: A Fuel-efficient Route Recommendation Approach for Long-distance Driving Based on Historical Trajectories</b> .....	73
Min Liu, Zhaohui Peng, Xiaohui Yu, Senzhang Wang, Qiao Song	
<b>Session-based Recommendation with Hypergraph Attention Networks</b> .....	82
Jianling Wang, Kaize Ding, Ziwei Zhu, James Caverlee	

## Session: CP03

<b>Deep Multi-Instance Contrastive Learning with Dual Attention for Anomaly Precursor Detection</b> .....	91
Dongkuan Xu, Wei Cheng, Jingchao Ni, Dongsheng Luo, Masanao Natsumeda, Dongjin Song, Bo Zong, Haifeng Chen, Xiang Zhang	
<b>Noise-Response Analysis of Deep Neural Networks Quantifies Robustness and Fingerprints Structural Malware</b> .....	100
N. Benjamin Erichson, Dane Taylor, Qixuan Wu, Michael W. Mahoney	
<b>Signature-Based Anomaly Detection in Networks</b> .....	109
Charu C. Aggarwal, Yao Li, Philip S. Yu	

<b>A Dimensionality-Driven Approach for Unsupervised Out-of-distribution Detection</b> .....	118
Qizhou Wang, Sarah M. Erfani, Christopher Leckie, Michael E. Houle	
<b>Reconstruction-based Anomaly Detection with Completely Random Forest</b> .....	127
Yi-Xuan Xu, Ming Pang, Ji Feng, Kai Ming Ting, Yuan Jiang, Zhi-Hua Zhou	
<b>Session: CP05</b>	
<b>New Approximation Algorithms for Forest Closeness Centrality – for Individual Vertices and Vertex Groups</b> .....	136
Alexander van der Grinten, Eugenio Angriman, Maria Predari, Henning Meyerhenke	
<b>PRESTO: Simple and Scalable Sampling Techniques for the Rigorous Approximation of Temporal Motif Counts</b> .....	145
Ilie Sarpe, Fabio Vandin	
<b>POTION: Optimizing Graph Structure for Targeted Diffusion</b> .....	154
Sixie Yu, Leo Torres, Scott Alfeld, Tina Eliassi-Rad, Yevgeniy Vorobeychik	
<b>Node Proximity Is All You Need: Unified Structural and Positional Node and Graph Embedding</b> .....	163
Jing Zhu, Xingyu Lu, Mark Heimann, Danai Koutra	
<b>Refining Network Alignment to Improve Matched Neighborhood Consistency</b> .....	172
Mark Heimann, Xiyuan Chen, Fatemeh Vahedian, Danai Koutra	
<b>Session: CP06</b>	
<b>Fairness-aware Agnostic Federated Learning</b> .....	181
Wei Du, Depeng Xu, Xintao Wu, Hanghang Tong	
<b>Equitable Allocation of Healthcare Resources with Fair Survival Models</b> .....	190
Kamrun Naher Keya, Rashidul Islam, Shimei Pan, Ian Stockwell, James Foulds	
<b>Fair Classification Under Strict Unawareness</b> .....	199
Haoyu Wang, Hengtong Zhang, Yaqing Wang, Jing Gao	
<b>Turning Attacks into Protection: Social Media Privacy Protection Using Adversarial Attacks</b> .....	208
Xiaoting Li, Lingwei Chen, Dinghao Wu	
<b>Influencers and the Giant Component: The Fundamental Hardness in Privacy Protection for Socially Contagious Attributes</b> .....	217
Aria Rezaei, Jie Gao, Anand D. Sarwate	
<b>Session: CP07</b>	
<b>A Practical Online Framework for Extracting Running Video Summaries under a Fixed Memory Budget</b> .....	226
Chandrashekhar Lavania, Kai Wei, Rishabh Iyer, Jeff Bilmes	
<b>Adaptive Holding for Online Bottleneck Matching with Delays</b> .....	235
Kaixin Wang, Cheng Long, Yongxin Tong, Jie Zhang, Yi Xu	

<b>Mining Easily Understandable Models from Complex Event Logs</b> .....	244
Boris Wiegand, Dietrich Klakow, Jilles Vreeken	
<b>Deep Neural Network for 3D Surface Segmentation based on Contour Tree Hierarchy</b> .....	253
Wenchong He, Arpan Man Sainju, Zhe Jiang, Da Yan	
<b>Learning Time-series Shapelets via Supervised Feature Selection</b> .....	262
Akihiro Yamaguchi, Ken Ueno	
<b>Session: CP09</b>	
<b>Maximizing Cohesion and Separation in Graph Representation Learning: A Distance-aware Negative Sampling Approach</b> .....	271
M. Maruf, Anuj Karpatne	
<b>DPGS: Degree-Preserving Graph Summarization</b> .....	280
Houquan Zhou, Shenghua Liu, Kyuhan Lee, Kijung Shin, Huawei Shen, Xueqi Cheng	
<b>Density of States Graph Kernels</b> .....	289
Leo Huang, Andrew J. Graven, David Bindel	
<b>SUSAN: The Structural Similarity Random Walk Kernel</b> .....	298
Janis Kalofolias, Pascal Welke, Jilles Vreeken	
<b>Scalable Graph Synthesis with Adj and 1-Adj</b> .....	307
Jinsung Jeon, Jing Liu, Jayoung Kim, Jaehoon Lee, Noseong Park, Jamie Jooyeon Lee, Ozlem Uzunerz, Sushil Jajodiaz	
<b>Session: CP10</b>	
<b>Multi-Armed Bandit Based Feature Selection</b> .....	316
Kunpeng Liu, Haibo Huang, Wei Zhang, Ahmad Hariri, Yanjie Fu, Kien Hua	
<b>UNIANO: Robust and Efficient Anomaly Consensus in Time Series Sensitive to Cross-Correlated Anomaly Profiles</b> .....	324
Leonor Silva, Helena Galhardas, Vasco Manquinho, Rui Henriques	
<b>Random Features Strengthen Graph Neural Networks</b> .....	333
Ryoma Sato, Makoto Yamada, Hisashi Kashima	
<b>AutoGFS: Automated Group-based Feature Selection via Interactive Reinforcement Learning</b> .....	342
Wei Fan, Kunpeng Liu, Hao Liu, Ahmad Hariri, Dejing Dou, Yanjie Fu	
<b>Better Short than Greedy: Interpretable Models through Optimal Rule Boosting</b> .....	351
Mario Boley, Simon Teshuva, Pierre Le Bodic, Geoffrey I. Webb	
<b>Session: CP11</b>	
<b>Complexity-Adaptive Gaussian Process Model Inference for Large-Scale Data</b> .....	360
Fabian Berns, Christian Becks	
<b>Predictive Optimization with Zero-Shot Domain Adaptation</b> .....	369
Tomoya Sakai, Naoto Ohsaka	

<b>Exceptional Model Mining meets Multi-objective Optimization</b> .....	378
Alexandre Millot, Rémy Cazabet, Jean-François Boulicaut	
<b>Estimating Conditional Mutual Information for Discrete-Continuous Mixtures using Multi-Dimensional Adaptive Histograms</b> .....	387
Alexander Marx, Lincen Yang, Matthijs van Leeuwen	
<b>Sign-aware Perturbations Regression</b> .....	396
Zhongnian Li, Tao Zhang, Wei Shao, Songcan Chen, Daoqiang Zhang	
<b>Session: CP13</b>	
<b>Faster Stochastic Second Order Method for Large-scale Machine Learning Models</b> .....	405
Hongchang Gao, Heng Huang	
<b>Maximizing Approximately k-Submodular Functions</b> .....	414
Leqian Zheng, Hau Chan, Grigorios Loukides, Minming Li	
<b>Submodular Maximization via Taylor Series Approximation</b> .....	423
Gözde Özcan, Armin Moharrer, Stratis Ioannidis	
<b>Frank-Wolfe Algorithm for Learning SVM-type Multi-category Classifiers</b> .....	432
Kenya Tajima, Yoshihiro Hirohashi, Esmeraldo Ronnie Rey Zara, Tsuyoshi Kato	
<b>Provable Distributed Stochastic Gradient Descent with Delayed Updates</b> .....	441
Hongchang Gao, Gang Wu, Ryan Rossi	
<b>Session: CP14</b>	
<b>Verifying Tree Ensembles by Reasoning about Potential Instances</b> .....	450
Laurens Devos, Wannes Meert, Jesse Davis	
<b>CiNet: Redesigning Deep Neural Networks for Efficient Mobile-Cloud Collaborative Inference</b> .....	459
Xin Dai, Xiangnan Kong, Tian Guo, Yixian Huang	
<b>"Misc"-Aware Weakly Supervised Aspect Classification</b> .....	468
Peiran Li, Fang Guo, Jingbo Shang	
<b>SUMDocS: Surrounding-aware Unsupervised Multi-Document Summarization</b> .....	477
Qi Zhu, Fang Guo, Jingjing Tian, Yuning Mao, Jiawei Han	
<b>Deep Multi-type Objects Multi-view Multi-instance Multi-label Learning</b> .....	486
Yuanlin Yang, Guoxian Yu, Carlotta Domeniconi, Xiangliang Zhang	
<b>Session: CP15</b>	
<b>Inter-Series Attention Model for COVID-19 Forecasting</b> .....	495
Xiaoyong Jin, Yu-Xiang Wang, Xifeng Yan	
<b>MT-STNets: Multi-Task Spatial-Temporal Networks for Multi-Scale Traffic Prediction</b> .....	504
Senzhang Wang, Meiyue Zhang, Hao Miao, Philip S. Yu	

<b>A Fine-grained Graph-based Spatiotemporal Network for Bike Flow Prediction in Bike-sharing Systems</b> .....	513
Peiyu Yi, Feihu Huang, Jian Peng	

<b>A Deep Adversarial Model for Suffix and Remaining Time Prediction of Event Sequences</b> .....	522
Farbod Taymouri, Marcello La Rosa, Sarah M. Erfani	

<b>Attention-Based Autoregression for Accurate and Efficient Multivariate Time Series Forecasting</b> .....	N/A
Jaemin Yoo, U. Kang	

### **Session: CP17**

<b>Yet Meta Learning Can Adapt Fast, it Can Also Break Easily</b> .....	540
Han Xu, Yaxin Li, Xiaorui Liu, Hui Liu, Jiliang Tang	

<b>Do Winning Tickets Exist before DNN training?</b> .....	549
Kathrin Grosse, Michael Backes	

<b>HALO: Learning to Prune Neural Networks with Shrinkage</b> .....	558
Skyler Seto, Martin T. Wells, Wenyu Zhang	

<b>EduHawkes: A Neural Hawkes Process Approach for Online Study Behavior Modeling</b> .....	567
Lu Jiang, Pengyang Wang, Ke Cheng, Kunpeng Liu, Minghao Yin, Bo Jin, Yanjie Fu	

<b>Contradictory Structure Learning for Semi-supervised Domain Adaptation</b> .....	576
Can Qin, Lichen Wang, Qianqian Ma, Yu Yin, Huan Wang, Yun Fu	

### **Session: CP18**

<b>Learning Robust Patient Representations from Multi-modal Electronic Health Records: A Supervised Deep Learning Approach</b> .....	585
Xianli Zhang, Buyue Qian, Yang Li, Yang Liu, Xi Chen, Chong Guan, Chen Li	

<b>TedPar: Temporally Dependent PARAFAC2 Factorization for Phenotype-based Disease Progression Modeling</b> .....	594
Kejing Yin, William K. Cheung, Benjamin C.M. Fung, Jonathan Poon	

<b>TeX-Graph: Coupled Tensor-matrix Knowledge-graph Embedding for COVID-19 Drug Repurposing</b> .....	603
Charilaos I. Kanatsoulis, Nicholas D. Sidiropoulos	

<b>Physics-Guided Recurrent Graph Model for Predicting Flow and Temperature in River Networks</b> .....	612
Xiaowei Jia, Jacob Zwart, Jeffery Sadler, Alison Appling, Samantha Oliver, Steven Markstrom, Jared Willard, Shaoming Xu, Michael Steinbach, Jordan Read, Vipin Kumar	

<b>Graph-based Reinforcement Learning for Active Learning in Real Time: An Application in Modeling River Networks</b> .....	621
Xiaowei Jia, Beiyu Lin, Jacob Zwart, Jeffery Sadler, Alison Appling, Samantha Oliver, Jordan Read	

## Session: CP19

- A Data-Dependent Algorithm for Querying Earth Mover's Distance with Low Doubling Dimensions** ..... 630  
Hu Ding, Tan Chen, Fan Yang, Mingyue Wang
- PhraseScope: An Effective and Unsupervised Framework for Mining High Quality Phrases** ..... 639  
Omer Anjum, Mohammad Almasri, Jinjun Xiong, Wen-mei Hwu
- Scalable Distributed Approximation of Internal Measures for Clustering Evaluation** ..... 648  
Federico Altieri, Andrea Pietracaprina, Geppino Pucci, Fabio Vandin
- Unsupervised Selective Manifold Regularized Matrix Factorization** ..... 657  
Priya Mani, Carlotta Domeniconi, Igor Griva
- Functional Autoencoders for Functional Data Representation Learning** ..... 666  
Tsung-Yu Hsieh, Yiwei Sun, Suhang Wang, Vasant Honavar

## Session: CP21

- Quadratic Residual Networks: A New Class of Neural Networks for Solving Forward and Inverse Problems in Physics Involving PDEs** ..... 675  
Jie Bu, Anuj Karpatne
- Synthesizing Skeletal Motion and Physiological Signals as a Function of a Virtual Human's Actions and Emotions** ..... 684  
Bonny Banerjee, Masoumeh Heidari Kapourchali, Murchana Baruah, Mousumi Deb, Kenneth Sakauye, Mette Olufsen
- Towards Learning Outcome Prediction via Modeling Question Explanations and Student Responses** ..... 693  
Tianqi Wang, Fenglong Ma, Yaqing Wang, Tang Tang, Longfei Zhang, Jing Gao
- Temporal-Aware Graph Neural Network for Credit Risk Prediction** ..... 702  
Daixin Wang, Zhiqiang Zhang, Jun Zhou, Peng Cui, Jingli Fang, Quanhui Jia, Yanming Fang, Yuan Qi
- Large-Scale Flight Frequency Optimization with Global Convergence in the US Domestic Air Passenger Markets** ..... 711  
Jinsung Jeon, Dongeun Lee, Seunghyun Hwang, Soyoun Kang, Noseong Park, Duanshun Li, Kookjin Lee, Jing Liu

## Session: CP22

- Heterogeneous Graph Neural Networks for Query-focused Summarization** ..... 720  
Jing Ya, Tingwen Liu, Jiangxia Cao, Li Guo
- Dynamic Graph Convolutional Networks Using the Tensor M-Product** ..... 729  
Osman Asif Malik, Shashanka Ubaru, Lior Horesh, Misha E. Kilmer, Haim Avron
- Disentangled Dynamic Graph Deep Generation** ..... 738  
Wenbin Zhang, Liming Zhang, Dieter Pfoser, Liang Zhao
- Robust Dual Recurrent Neural Networks for Financial Time Series Prediction** ..... 747  
Jiayu He, Matloob Khushi, Nguyen H. Tran, Tongliang Liu
- Learning over Families of Sets – Hypergraph Representation Learning for Higher Order Tasks** ..... 756  
Balasubramaniam Srinivasan, Da Zheng, George Karypis