

SIAM International Conference on Data Mining (SDM'25)

Alexandria, Virginia, USA
1-3 May 2025

ISBN: 979-8-3313-1997-7

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© (2025) by SIAM: Society for Industrial and Applied Mathematics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2025)

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

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
Web: www.proceedings.com

TABLE OF CONTENTS

Thursday AM

CP1: Embeddings and Features

Automated Learning of Semantic Embedding Representations for Diffusion Models	1
--	----------

Limai Jiang and Yunpeng Cai

Task Aware Modulation Using Representation Learning: An Approach for Few Shot Learning in Environmental Systems	11
--	-----------

Arvind Renganathan, Rahul Ghosh, Ankush Khandelwal and Vipin Kumar

OpenFE++: Efficient Automated Feature Generation via Feature Interaction	21
---	-----------

Lei Wang, Yu Shi, Yifei Jin and Jian Li

Acceleration in Low-Rank Tensor Completion	31
---	-----------

Yifan Kang, Mengyuan Zhang and Kai Liu

CP2: Anomaly Detection

Convergence-Guaranteed Elastic Net Graphical Model Estimation with Applications to Anomaly Localization	41
--	-----------

Dzung T. Phan, Matt Menickelly, Tsuyoshi Ide and Jayant Kalagnanam

Anomaly Detection via Graph Contrastive Learning	51
---	-----------

Emre Sefer

Federated Koopman-Reservoir Learning for Large-Scale Multivariate Time-Series Anomaly Detection	61
--	-----------

Long Tan Le, Tung-Anh Nguyen, Han Shu[†] Suranga Seneviratne, Choong Seon Hong and Nguyen H. Tran

Trajectory Anomaly Detection with By-Design Complementary Detectors	71
--	-----------

Shurui Cao and Leman Akoglu

CP3: Learning Under Uncertainty

Learning Confident Classifiers in the Presence of Label Noise	81
--	-----------

Asma Ahmed Hashmi, Aigerim Zhumabayeva, Nikita Kotelevskii, Artem Agafonov, Mohammad Yaqub, Maxim Panov and Martin Takáč

TADAM: Learning Timed Automata from Noisy Observations	114
<i>Lénaïg Cornanguer and Pierre-François Gimenez</i>	
$\ell_{1,\infty}$-Mixed Norm Promoted Row Sparsity for Fast Online CUR Decomposition Learning in Varying Feature Spaces	124
<i>Zhong Chen, Yi He, Di Wu, Wenbin Zhang and Zhiqiang Deng</i>	
Label Shift Estimation with Incremental Prior Update	134
<i>Yunrui Zhang, Gustavo Batista and Salil S. Kanhere</i>	
Constraint-Focused Training for Multistate Survival Analysis with Neural Networks	143
<i>Ryosuke Takayama and Masanao Natsumeda</i>	
Thursday PM	
CP4: Graph Models and Algorithms	
Defense Against Shortest Path Attacks	152
<i>Benjamin A. Miller, Zohair Shafi, Wheeler Ruml, Yevgeniy Vorobeychik, Tina Eliassi-Rad and Scott Alfeld</i>	
Conformal Edge-Weight Prediction in Latent Space	161
<i>Akash Choudhuri, Yongjian Zhong, Mehrdad Moharrami, Christine Klymko, Mark Heimann, Jayaraman J. Thiagarajan and Bijaya Adhikari</i>	
Efficient Sampling of Temporal Networks with Preserved Causality Structure	171
<i>Felix I. Stamm, Mehdi Naima and Michael T. Schaub</i>	
StarRec: A Hypergraph-Based Framework with Star-Expansion for Multi-Behavior Recommendation	181
<i>Wenhan Zhang, Zijian Song, Yihuan Wu, Lifang Deng, Jiandong Zhang, Kaigui Bian and Bin Cui</i>	
CP5: Recommender Systems	
A Look into News Avoidance Through AWRS: An Avoidance-Aware Recommender System	192
<i>Igor L.R. Azevedo, Toyotaro Suzumura and Yuichiro Yasui</i>	
Beyond Models! Data Valuation and Metric Adaption for Recommendation	203
<i>Renqi Jia, Xiaokun Zhang, Bowei He, Qiannan Zhu, Weitao Xu, Jiehao Chen and Chen Ma</i>	
CDSRNP: Cross-Domain Sequential Recommendation via Neural Process	213
<i>Haipeng Li, Jiangxia Cao, Yiwen Gao, Yunhuai Liu and Shuchao Pang</i>	

Ranking with Confidence for Large Scale Comparison Data	223
--	------------

Filipa Valdeira and Cláudia Soares

Friday AM

CP6: Clustering

Differentially Private Associative Co-Clustering	233
---	------------

Elena Battaglia and Ruggero G. Pensa

Hierarchical Superpixel Segmentation via Structural Information Theory	242
---	------------

Minhui Xie, Hao Peng, Pu Li, Guangjie Zeng, Shuhai Wang, Jia Wu, Peng Li and Philip S. Yu

DynHAC: Fully Dynamic Approximate Hierarchical Agglomerative Clustering	252
--	------------

Shangdi Yu, Laxman Dhulipala, Jakub Łqcki and Nikos Parotsidis

DMDHC: Discovery of Multi-Density Hierarchical Cluster Structures	261
--	------------

Walid Durani, Dominik Mautz, Claudia Plant and Christian Böhm

Multi-View Spectral Clustering for Graphs with Multiple View Structures	270
--	------------

Yorgos Tsitsikas and Evangelos E. Papalexakis

CP7: Applications

Inter-Well Active Magnetic Ranging with Temporal and Interaction Network	279
---	------------

Zelong Hao, Haitao Zhang, Yang Che and Wang Liu

Unanticipated Replenishment: Online Policy for Dynamic Service Composition in Manufacturing Cloud	288
--	------------

Yang Hu, Feng Wu, Xin Li and Yu Yang

Domain-Adaptive Continual Meta-Learning for Modeling Dynamical Systems: An Application in Environmental Ecosystems	297
---	------------

Yiming Sun, Runlong Yu, Runxue Bao, Yiqun Xie, Ye Ye and Xiaowei Jia

Bridging Numbers and Narratives: Enhancing Financial Market Risk Predictions Through Numerical Information from Financial Documents	307
--	------------

Yu Qin, Chengshang Zhang and Wei Xu

Context-Aware Frequency-Embedding Networks for Spatio-Temporal Portfolio Selection	317
---	------------

Ruirui Liu, Huichou Huang, Johannes Ruf, Haoxian Liu and Qingyao Wu

CP8: Blue Sky 1: Advancing AI and Machine Learning

Blue Sky: Expert-in-the-Loop Representation Learning Framework for Audio Anti-Spoofing: Multimodal, Multilingual, Multi-Speaker, Multi-Attack (4M) Scenarios	327
---	------------

Zahra Khanjani, Vandana P. Janeja, Christine Mallinson and Sanjay Purushotham

Heterogeneous Multi-Agent Framework for Dynamic Generalized Category Discovery	331
---	------------

Fatimah Alotaibi, Adithya Kulkarni and Dawei Zhou

Blue Sky: Reducing Performance Gap Between Commercial and Open-Source LLMs	335
---	------------

Adithya Kulkarni and Mohna Chakraborty

Explainable AI for Real-Time Video Anomaly Anticipation	339
--	------------

David C. Anastasiu

Evaluating Time Series Models with Knowledge Discovery	343
---	------------

Li Zhang

Friday PM

CP9: Learning on Graphs

Optimizing Transit Network Expansion with Gated Attentive Graph Reinforcement Learning	347
---	------------

Fanglan Chen, Dongjie Wang, Jianfeng He, Shuo Lei and Chang-Tien Lu

Feature Deviation Embedding Improves Graph Structure Learning for Spatial Interpolation	356
--	------------

Chaofan Li, Till Riedel and Michael Beigl

Equipping Graph Autoencoders: Revisiting Masking Strategies from a Robustness Perspective	366
--	------------

Shuhan Song, Ping Li, Ming Dun, Yuan Zhang, Huawei Cao and Xiaochun Ye

REGE: A Method for Incorporating Uncertainty in Graph Embeddings	376
---	------------

Zohair Shafi, Germans Savcisens and Tina Eliassi-Rad

CP10: Large Language Models

Protecting Privacy Against Membership Inference Attack with LLM Fine-Tuning Through Flatness	386
---	------------

Tiejun Chen, Longchao Da, Huixue Zhou, Pingzhi Li, Kaixiong Zhou, Tianlong Chen and Hua Wei

Language Models Are Explorers for Join Discovery on Data Lakes	398
---	------------

Yaohua Wang, Bolin Ding, Rong Zhu, Haibin Wang, Zhijian Ma and Jingren Zhou

CoMAL: Collaborative Multi-Agent Large Language Models for Mixed-Autonomy Traffic	409
<i>Huaiyuan Yao, Longchao Da, Vishnu Nandam, Justin Turnau, Zhiwei Liu, Linsey Pang and Hua Wei</i>	
CoT-Decoding: Complex Reasoning via Chain-of-Thought Decoding	419
<i>Guoquan Lu, Lin Peng and Li Li</i>	
CP11: Blue Sky 2: AI for Scientific and Societal Impact	
Optimizing External and Internal Knowledge of Foundation Models for Scientific Discovery	431
<i>Sikun Guo, Guangzhi Xiong and Aidong Zhang</i>	
Data Mining the Functional Architecture of the Brain’s Circuitry	435
<i>Adam S. Charles</i>	
What We Talk About When We Talk About AI for Science	439
<i>Runlong Yu, Yiqun Xie and Xiaowei Jia</i>	
Better AI for Understanding Life on Earth: Predict First, Design Later	443
<i>Yana Bromberg and Amarda Shehu</i>	
Saturday AM	
CP12: Machine Learning for Healthcare	
AnchorDrug: A System for Drug-Induced Gene Expression Prediction in New Contexts Through Active Learning	447
<i>Han Meng, Ruoqiao Chen, Bin Chen and Jiayu Zhou</i>	
Accurately Estimating Unreported Infections Using Information Theory	457
<i>Jiaming Cui, Bijaya Adhikari, Arash Haddadan, A S M Ahsan-Ul Haque, Jilles Vreeken, Anil Vullikanti and B. Aditya Prakash</i>	
MEXA-CTP: Mode Experts Cross-Attention for Clinical Trial Outcome Prediction	467
<i>Yiqing Zhang, Xiaozhong Liu and Fabricio Murai</i>	
Domain Knowledge Augmented Contrastive Learning on Dynamic Hypergraphs for Improved Health Risk Prediction	476
<i>Akash Choudhuri, Hieu Vu, Kishlay Jha and Bijaya Adhikari</i>	
Spatially-Delineated Domain Adapted AI Classification: An Application for Oncology Data	487
<i>Majid Farhadloo, Arun Sharma, Alexey Leontovich, Svetomir N. Markovic and Shashi Shekhar</i>	

CP13: Data Integration and Fusion

VisTabNet: Adapting Vision Transformers for Tabular Data	497
---	------------

Witold Wydmański, Ulvi Movsum-zada, Jacek Tabor and Marek Śmieja

Emergence of Cooperation in Multi-Agent Reinforcement Learning via Coalition Labeling and Structural Entropy	507
---	------------

Dingli Su, Hao Peng, Guangjie Zeng, Pu Li, Angsheng Li and Yicheng Pan

Parameter-Efficient Interventions for Enhanced Model Merging	516
---	------------

Marcin Osial, Daniel Marczak and Bartosz Zieliński

Metrics for Inter-Dataset Similarity with Example Applications in Synthetic Data and Feature Selection Evaluation	527
--	------------

Muhammad Rajabinasab, Anton Lautrup and Arthur Zimek

CP14: Time-Series Analysis and Forecasting

AutoSTDiff: Autoregressive Spatio-Temporal Denoising Diffusion Model for Asynchronous Trajectory Generation	538
--	------------

Rongchao Xu, Zhiqing Hong and Guang Wang

AVATAR: Adversarial Autoencoders with Autoregressive Refinement for Time Series Generation	548
---	------------

MohammadReza EskandariNasab, Shah Muhammad Hamdi and Soukaina Filali Boubrahimi

Fine-Grained Spatio-Temporal Event Prediction with Self-Adaptive Anchor Graph	558
--	------------

Wang-Tao Zhou, Zhao Kang, Sicong Liu, Lizong Zhang and Ling Tian

End-to-End Self-Tuning Self-Supervised Time Series Anomaly Detection	568
---	------------

Boje Deforce, Meng-Chieh Lee, Bart Baesens, Estefanía Serral Asensio, Jaemin Yoo and Leman Akoglu

Saturday PM

CP15: Graph Neural Networks

Staleness-Alleviated Distributed GNN Training via Online Dynamic-Embedding Prediction	578
--	------------

Guangji Bai, Ziyang Yu, Zheng Chai, Yue Cheng and Liang Zhao

Evidence-Based Out-of-Distribution Detection on Multi-Label Graphs	588
---	------------

Ruomeng Ding, Xujiang Zhao, Chen Zhao, Minglai Shao, Zhengzhang Chen and Haifeng Chen

FedGrAINS: Personalized SubGraph Federated Learning with Adaptive Neighbor Sampling	598
--	------------

Emir Ceyani, Han Xie, Baturalp Buyukates, Carl Yang and Salman Avestimehr

Unveiling the Impact of Local Homophily on GNN Fairness: In-Depth Analysis and New Benchmarks	608
--	------------

Donald Loveland and Danaï Koutra

GAIM: Attacking Graph Neural Networks via Adversarial Influence Maximization	618
---	------------

Xiaodong Yang, Xiaoting Li, Huiyuan Chen and Yiwei Cai

CP16: Learning and optimization

Hybrid Bayesian Optimization with DIRECT	627
---	------------

Hongsheng Liu and Dzung T. Phan

An Interpretable Measure for Quantifying Predictive Dependence Between Continuous Random Variables	637
---	------------

Renato Assunção, Flávio Figueiredo, Francisco N. Tinoco Junior, Leo M. de Sá-Freire and Fábio Silva

Approximating Splits for Decision Trees Quickly in Sparse Data Streams	647
---	------------

Nikolaj Tatti

Meta-Learning of Class Knowledge in Zero-Shot Learning	656
---	------------

Yuta Nambu, Masahiro Kohjima, Tomoharu Iwata and Ryuji Yamamoto