

2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA 2019)

**Washington, DC, USA
5-8 October 2019**



**IEEE Catalog Number: CFP19DSB-POD
ISBN: 978-1-7281-4494-8**

**Copyright © 2019 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP19DSB-POD
ISBN (Print-On-Demand):	978-1-7281-4494-8
ISBN (Online):	978-1-7281-4493-1

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
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA) **DSAA 2019**

Table of Contents

Message from the General/Logistics Chairs	xii
Message from the Program Committee Co-Chairs	xiii
Organizing Committee	xiv
Advisory Committee	xvi
Steering Committee	xvii
Program Committee	xviii
DSAA Keynotes	xxi
Late Breaking Posters	xxiv
Tutorials	xxv

Research and Special Session Tracks

Grade Prediction with Neural Collaborative Filtering	1
<i>Zhiyun Ren (George Mason University), Xia Ning (The Ohio State University), Andrew S. Lan (University of Massachusetts Amherst), and Huzefa Rangwala (George Mason University)</i>	
MARS: Memory Attention-Aware Recommender System	11
<i>Lei Zheng (University of Illinois at Chicago), Chun-Ta Lu (University of Illinois at Chicago), Lifang He (University of Pennsylvania), Sihong Xie (Lehigh University), Huang He (University of Illinois at Chicago), Chaozhuo Li (Beihang University), Vahid Noroozi (University of Illinois at Chicago), Bowen Dong (University of Illinois at Chicago), and Philip S. Yu (University of Illinois at Chicago)</i>	
Variable-Lag Granger Causality for Time Series Analysis	21
<i>Chainarong Amornbunchornvej (National Electronics and Computer Technology Center), Elena Zheleva (University of Illinois at Chicago), and Tanya Y. Berger-Wolf (University of Illinois at Chicago)</i>	
Residual Networks Behave Like Boosting Algorithms	31
<i>Chapman Siu (University of Technology Sydney)</i>	
Hierarchical LSTM Framework for Long-Term Sea Surface Temperature Forecasting	41
<i>Xi Liu (Michigan State University), Tyler Wilson (Michigan State University), Pang-Ning Tan (Michigan State University), and Lifeng Luo (Michigan State University)</i>	
Semantically-Aware Statistical Metrics via Weighting Kernels	51
<i>Stefano Cresci (IIT-CNR, Italy), Roberto Di Pietro (HBKU, Qatar), and Maurizio Tesconi (IIT-CNR, Italy)</i>	

Short Text Embedding for Clustering Based on Word and Topic Semantic Information	61
<i>Ziheng Chen (Sun Yat-Sen University) and Jiangtao Ren (Sun Yat-Sen University)</i>	
A Rademacher Complexity Based Method for Controlling Power and Confidence Level in Adaptive Statistical Analysis	71
<i>Lorenzo De Stefani (Brown University) and Eli Upfal (Brown University)</i>	
SeqScout: Using a Bandit Model to Discover Interesting Subgroups in Labeled Sequences	81
<i>Romain Mathonat (Université de Lyon, ATOS), Diana Nurbakova (Université de Lyon), Jean-François Boulicaut (Université de Lyon), and Mehdi Kaytoue (Université de Lyon, Infologic)</i>	
FSSD - A Fast and Efficient Algorithm for Subgroup Set Discovery	91
<i>Adnene Belfodil (Univ Lyon), Aimene Belfodil (Univ Lyon), Anes Bendimerad (Univ Lyon), Philippe Lamarre (Univ Lyon), Céline Robardet (Univ Lyon), Mehdi Kaytoue (Univ Lyon), and Marc Plantevit (Univ Lyon)</i>	
Biased Resampling Strategies for Imbalanced Spatio-Temporal Forecasting	100
<i>Mariana Oliveira (University of Porto), Nuno Moniz (University of Porto), Luís Torgo (Dalhousie University, University of Porto), and Vítor Santos Costa (University of Porto)</i>	
Explaining the Performance of Black Box Regression Models	110
<i>Inês Areosa (LIAAD, INESC TEC) and Luís Torgo (Dalhousie University)</i>	
Dynamic Graph Embedding via LSTM History Tracking	119
<i>Shima Khoshraftar (York University), Sedigheh Mahdavi (York University), Aijun An (York University), Yonggang Hu (IBM Canada), and Junfeng Liu (IBM Canada)</i>	
Text Summarization by Highlighting Core Content	
<i>Zhixin Li (Guangxi Normal University), Zhi Peng (Guangxi Normal University), Suqin Tan (Guangxi Normal University), and Canlong Zhang (Guangxi Normal University)</i>	
Least Squares and Maximum Likelihood Estimation of Sufficient Reductions in Regressions with Matrix Valued Predictors	135
<i>Ruth Pfeiffer (National Cancer Institute, NIH), Wei Wang (JD AI Research, China), and Efstathia Bura (TU Wien)</i>	
Cross-Media Image-Text Retrieval Combined with Global Similarity and Local Similarity	145
<i>Zhixin Li (Guangxi Normal University), Feng Ling (Guangxi Normal University), and Canlong Zhang (Guangxi Normal University)</i>	
Sensor-Based Human Activity Mining Using Dirichlet Process Mixtures of Directional Statistical Models	154
<i>Lei Fang (University of St Andrews), Juan Ye (University of St Andrews), and Simon Dobson (University of St Andrews)</i>	
Truth Discovery from Multi-Sourced Text Data Based on Ant Colony Optimization	164
<i>Chen Chang (Army Engineer University), Jianjun Cao (National University of Defense Technology), Guojun Lv (Army Engineer University), and Nianfeng Weng (National University of Defense Technology)</i>	
On the Classification Consistency of High-Dimensional Sparse Neural Network	173
<i>Kaixu Yang (Michigan State University) and Taps Maiti (Michigan State University)</i>	

Deep Crowd Counting In Congested Scenes Through Refine Modules	183
<i>Tong Li (Chinese Academy of Sciences), Chuan Wang (Chinese Academy of Sciences), and Xiaochun Cao (Chinese Academy of Sciences)</i>	
A Study on the Impact of Data Characteristics in Imbalanced Regression Tasks	193
<i>Paula Branco (Dalhousie University) and Luis Torgo (Dalhousie University)</i>	
Sign Language Recognition Analysis using Multimodal Data	203
<i>Al Amin Hosain (George Mason University), Panneer Selvam Santhalingam (George Mason University), Parth Pathak (George Mason University), Jana Kořecká (George Mason University), and Huzeifa Rangwala (George Mason University)</i>	
An Incremental Technique for Mining Coverage Patterns in Large Databases	211
<i>Akhil Ralla (International Institute of Information Technology, Hyderabad), P. Krishna Reddy (International Institute of Information Technology, Hyderabad), and Anirban Mondal (Ashoka University)</i>	
Customized Interpretable Conformal Regressors	221
<i>Ulf Johansson (Jönköping University), Cecilia Sönströd (University of Borås), Tuwe Löfström (Jönköping University), and Henrik Boström (KTH Royal Institute of Technology)</i>	
A Novel Multiple Classifier Generation and Combination Framework Based on Fuzzy Clustering and Individualized Ensemble Construction	231
<i>Zhen Gao (University of Texas at San Antonio), Maryam Zand (University of Texas at San Antonio), and Jianhua Ruan (University of Texas at San Antonio)</i>	
VizCertify: A Framework for Secure Visual Data Exploration	241
<i>Lorenzo De Stefani (Brown University), Leonhard F. Spiegelberg (Brown University), Eli Upfal (Brown University), and Tim Kraska (Massachusetts Institute of Technology)</i>	
Joint Selection of Central and Extremal Prototypes Based on Kernel Minimum Enclosing Balls	252
<i>Christian Bauckhage (Fraunhofer Center for Machine Learning) and Rafet Sifa (Fraunhofer Center for Machine Learning)</i>	
A Rapid Prototyping Approach for High Performance Density-Based Clustering	260
<i>Saiyedul Islam (Birla Institute of Technology and Science Pilani), Sundar Balasubramaniam (Birla Institute of Technology and Science Pilani), Poonam Goyal (Birla Institute of Technology and Science Pilani), Ankit Sultana (Birla Institute of Technology and Science Pilani), Lakshit Bhutani (Birla Institute of Technology and Science Pilani), Saurabh Raje (Birla Institute of Technology and Science Pilani), and Navneet Goyal (Birla Institute of Technology and Science Pilani)</i>	
Feature Enhanced Fully Convolutional Networks for Monocular Depth Estimation	270
<i>Chunxiu Shi (Beijing University of Chemical Technology), Jie Chen (Beijing University of Chemical Technology), Juan Chen (Beijing University of Chemical Technology), and Zheng Zhang (Beijing University of Chemical Technology)</i>	

Breast Cancer Classification using Deep Transfer Learning on Structured Healthcare Data	277
<i>Akram Farhadi (University of Georgia, Athens, GA), David Chen (Mayo Clinic, Rochester, MN), Rozalina McCoy (Mayo Clinic, Rochester, MN), Christopher Scott (Mayo Clinic, Rochester, MN), John A. Miller (University of Georgia, Athens, GA), Celine M. Vachon (Mayo Clinic, Rochester, MN), and Che Ngufor (Mayo Clinic, Rochester, MN)</i>	
Shape Constrained Tensor Decompositions	287
<i>Bethany Lusch (Argonne National Laboratory), Eric C. Chi (North Carolina State University), and J. Nathan Kutz (University of Washington)</i>	
Range Analysis and Applications to Root Causing	298
<i>Zurab Khasidashvili (Intel Corporation) and Adam J Norman (Intel Corporation)</i>	
TBOVW : A Temporal Bag of Visual Word Model Based Molecular Subtypes Recognition of Breast Cancer in DCE-MRI with Heterogeneity	
<i>Wei Li (Northeastern University, Ministry of Education), Chaolu Feng (Northeastern University, Ministry of Education), Dazhe Zhao (Northeastern University, Ministry of Education), Yiqing Wang (Northeastern University), Dongjie Wang (Northeastern University), and Xin Min (Northeastern University)</i>	
Unsupervised Drift Detector Ensembles for Data Stream Mining	317
<i>Lukasz Korycki (Virginia Commonwealth University) and Bartosz Krawczyk (Virginia Commonwealth University)</i>	

Application Track

Topology-Based Clusterwise Regression for User Segmentation and Demand Forecasting	326
<i>Rodrigo Rivera-Castro (Skolkovo Institute of Science and Technology), Aleksandr Pletnev (Skolkovo Institute of Science and Technology), Polina Pilyugina (Skolkovo Institute of Science and Technology), Grecia Diaz (Skolkovo Institute of Science and Technology), Ivan Nazarov (Skolkovo Institute of Science and Technology), Wanyi Zhu (Alibaba Cloud Intelligence Business Group), and Evgeny Burnaev (Skolkovo Institute of Science and Technology)</i>	
FARE: Schema-Agnostic Anomaly Detection in Social Event Logs	337
<i>Neil Shah (Snap Inc.)</i>	
SliceNDice: Mining Suspicious Multi-Attribute Entity Groups with Multi-View Graphs	351
<i>Hamed Nilforoshan (Columbia University) and Neil Shah (Snap Inc.)</i>	
Constrained Multi-Objective Optimization for Automated Machine Learning	364
<i>Steven Gardner (SAS Institute Inc.), Oleg Golovidov (SAS Institute Inc.), Joshua Griffin (SAS Institute Inc.), Patrick Koch (SAS Institute Inc.), Wayne Thompson (SAS Institute Inc.), Brett Wujek (SAS Institute Inc.), and Yan Xu (SAS Institute Inc.)</i>	
Detecting Sensitive Content in Spoken Language	374
<i>Rahul Tripathi (Amazon), Balaji Dhamodharaswamy (Amazon), Srinivasan Jagannathan (Amazon), and Abhishek Nandi (Amazon)</i>	

Higher Order Mining for Monitoring District Heating Substations	382
<i>Shahrooz Abghari (Blekinge Institute of Technology, Sweden), Veselka Boeva (Blekinge Institute of Technology, Sweden), Jens Brage (NODA Intelligent Systems AB, Sweden), Christian Johansson (NODA Intelligent Systems AB, Sweden), Håkan Grahn (Blekinge Institute of Technology, Sweden), and Niklas Lavesson (Jönköping University, Sweden)</i>	
Learning Personalized Treatment Rules from Electronic Health Records Using Topic Modeling Feature Extraction	392
<i>Peng Wu (Columbia University), Tianchen Xu (Columbia University), and Yuanjia Wang (Columbia University)</i>	
Lightweight Traffic Flow Prediction Based on Gated Recurrent Unit with Small Dataset	
<i>Hongze Wang (East China Normal University) and Jing Liu (East China Normal University)</i>	
Lightweight and Scalable Particle Tracking and Motion Clustering of 3D Cell Trajectories	412
<i>Mojtaba Sedigh Fazli (The University of Georgia), Rachel V. Stadler (University of Vermont), BahaaEddin Alaila (The University of Georgia), Stephen A. Vella (The University of Georgia), Silvia N. J. Moreno (The University of Georgia), Gary E. Ward (University of Vermont), and Shannon Quinn (The University of Georgia)</i>	
Uplift Modeling for Multiple Treatments with Cost Optimization	422
<i>Zhenyu Zhao (Uber Technologies, Inc.) and Totte Harinen (Uber Technologies, Inc.)</i>	
Chest Tube Management After Lung Resection Surgery using a Classifier	432
<i>William Klement (University of Ottawa, The Ottawa Hospital Research Institute, Faculty of Computer Science, Dalhousie University), Sebastien Gilbert (University of Ottawa, The Ottawa Hospital Research Institute), Donna E Maziak (University of Ottawa, The Ottawa Hospital Research Institute), Andrew J E Seely (University of Ottawa, The Ottawa Hospital Research Institute), Farid M Shamji (University of Ottawa, The Ottawa Hospital Research Institute), Sudhir R Sundaresan (University of Ottawa, The Ottawa Hospital Research Institute), Patrick J Villeneuve (University of Ottawa, The Ottawa Hospital Research Institute), and Nathalie Japkowicz (American University)</i>	
Maximum Relevance and Minimum Redundancy Feature Selection Methods for a Marketing Machine Learning Platform	442
<i>Zhenyu Zhao (Uber Technologies, Inc.), Radhika Anand (Uber Technologies, Inc.), and Mallory Wang (Uber Technologies, Inc.)</i>	
Towards Automated Breast Mass Classification using Deep Learning Framework	453
<i>Pinaki Ranjan Sarkar (Indian Institute of Space Science and Technology, Kerala-695547, India), Priya Prabhakar (Indian Institute of Space Science and Technology, Kerala-695547, India), Deepak Mishra (Indian Institute of Space Science and Technology, Kerala-695547, India), and Gorthi Subrahmanyam (IIT Tirupati, India-517506)</i>	
Matrix Profile XVI: Efficient and Effective Labeling of Massive Time Series Archives	463
<i>Frank Madrid (University of California Riverside), Shailendra Singh (University of California Riverside), Quentin Chesnais (University of California Riverside), Kerry Mauck (University of California Riverside), and Eamonn Keogh (University of California Riverside)</i>	

Attention-Based Network with Prior Knowledge for Domain-Specific Relation Extraction	
<i>Liqi Kang (Beijing Institute of Technology), Chong Feng (Beijing Institute of Technology), Ge Shi (Beijing Institute of Technology), Jun Ma (Beijing Institute of Technology), Yang Wang (Beijing Institute of Technology), and Heyan Huang (Beijing Institute of Technology)</i>	
Colorwall: An Embedded Temporal Display of Bibliographic Data	482
<i>Jing Ming (Shandong University) and Li Zhang (Qilu University of Technology)</i>	
Data-Driven Fault Diagnosis in End-of-Line Testing of Complex Products	492
<i>Vitali Hirsch (Graduate School of Excellence Advanced Manufacturing Engineering), Peter Reimann (Graduate School of Excellence Advanced Manufacturing Engineering), and Bernhard Mitschang (Institute for Parallel and Distributed Systems)</i>	
Towards Automatic Screening of Typical and Atypical Behaviors in Children With Autism	504
<i>Andrew Cook (Keele University), Bappaditya Mandal (Keele University), Donna Berry (Keele University), and Matthew Johnson (North Staffordshire Combined Healthcare NHS Trust, Stoke-On-Trent, UK)</i>	
On Analysing Supply and Demand in Labor Markets: Framework, Model and System	511
<i>Hendrik Santoso Sugiarto (Living Analytics Research Centre, Singapore Management University), Ee-Peng Lim (Living Analytics Research Centre, Singapore Management University), and Ngak-Leng Sim (Living Analytics Research Centre, Singapore Management University)</i>	
Martensite Start Temperature Predictor for Steels Using Ensemble Data Mining	521
<i>Ankit Agrawal (Northwestern University), Abhinav Saboo (QuesTek Innovations), Wei Xiong (University of Pittsburgh), Greg Olson (Northwestern University), and Alok Choudhary (Northwestern University)</i>	
Comparison of Variable Selection Methods for Forecasting from Short Time Series	531
<i>Monnie McGee (Southern Methodist University) and Robert A. Yaffee (New York University)</i>	
A Real-Time Iterative Machine Learning Approach for Temperature Profile Prediction in Additive Manufacturing Processes	541
<i>Arindam Paul (Northwestern University), Mojtaba Mozaffar (Northwestern University), Zijiang Yang (Northwestern University), Wei-keng Liao (Northwestern University), Alok Choudhary (Northwestern University), Jian Cao (Northwestern University), and Ankit Agrawal (Northwestern University)</i>	
Bighead: A Framework-Agnostic, End-to-End Machine Learning Platform	551
<i>Eli Brumbaugh (Airbnb Inc.), Mani Bhushan (Airbnb Inc.), Andrew Cheong (Airbnb Inc.), Michelle Gu-Qian Du (Airbnb Inc.), Jeff Feng (Airbnb Inc.), Nick Handel (Airbnb Inc.), Andrew Hoh (Airbnb Inc.), Jack Hone (Airbnb Inc.), Brad Hunter (Airbnb Inc.), Atul Kale (Airbnb Inc.), Alfredo Luque (Airbnb Inc.), Bahador Nooraei (Airbnb Inc.), John Park (Airbnb Inc.), Krishna Puttaswamy (Airbnb Inc.), Kyle Schiller (Airbnb Inc.), Evgeny Shapiro (Airbnb Inc.), Conglei Shi (Airbnb Inc.), Aaron Siegel (Airbnb Inc.), Nikhil Simha (Airbnb Inc.), Marie Sbrocca (Airbnb Inc.), Shi-Jing Yao (Airbnb Inc.), Patrick Yoon (Airbnb Inc.), Varant Zanoayan (Airbnb Inc.), Xiao-Han T. Zeng (Airbnb Inc.), and Qiang Zhu (Airbnb Inc.)</i>	

Machine Learning for Efficient Integration of Record Systems for Missing US Service Members	561
<i>Julia D. Warnke-Sommer (SNA International) and Franklin E. Damann (Defense POW/MIA Accounting Agency)</i>	
Paths to Empathy: Heterogeneous Effects of Reading Personal Stories Online	570
<i>Mahnaz Roshanaei (Stanford University), Christopher Tran (University of Illinois at Chicago), Sylvia Morelli (University of Illinois at Chicago), Cornelia Caragea (University of Illinois at Chicago), and Elena Zheleva (University of Illinois at Chicago)</i>	
A Novel Record Linkage Interface That Incorporates Group Structure to Rapidly Collect Richer Labels	580
<i>Kayla Frisoli (Carnegie Mellon University), Benjamin LeRoy (Carnegie Mellon University), and Rebecca Nugent (Carnegie Mellon University)</i>	
Analyzing Driving Data using the ADAPT Distributed Analytics Platform for Connected Vehicles	590
<i>Hillol Kargupta (Agnik) and Priyanka Kargupta (UC Berkeley)</i>	

Poster Papers

Augmenting U.S. Census data on industry and occupation of respondents	600
<i>Peter B. Meyer (U.S. Bureau of Labor Statistics) and Kendra Asher (U.S. Bureau of Labor Statistics)</i>	
Data Reduction for real-time bridge vibration data on Edge	602
<i>Anthony Chen (National Taiwan University), Fu-Hsuan Liu (National Taiwan University), and Sheng-De Wang (National Taiwan University)</i>	
Using Machine Learning to Predict High School Student Employability – A Case Study	604
<i>Aarushi Dubey (Academy of Engineering and Technology) and Muthukumara Mani (The World Bank)</i>	
Improving the Personalized Recommendation in the Cold-start Scenarios	606
<i>Peter Gaspar (Slovak University of Technology in Bratislava), Michal Kompan (Slovak University of Technology in Bratislava), Matej Koncal (Slovak University of Technology in Bratislava), and Maria Bielikova (Slovak University of Technology in Bratislava)</i>	
Universal Consistency of Support Tensor Machine	608
<i>Peide Li (Michigan State University) and Taps Maiti (Michigan State University)</i>	
Automating Big Data Analysis Based on Deep Learning Generation by Automatic Service Composition	610
<i>Incheon Paik (University of Aizu) and T.H. Akila S Siriweera (University of Aizu)</i>	
Detection Model of Depression Based on Eye Movement Trajectory	612
<i>Yifang Yuan (Qilu University of Technology) and Qingxiang Wang (Qilu University of Technology)</i>	
Exploring the Relationship Between Conversation Using #MeToo and University Harassment Policies	614
<i>Julianne Zech (Georgetown University), Fransiska Dale (University of Maryland), Lisa Singh (Georgetown University), Jamillah Williams (Georgetown University), and Naomi Mezey (Georgetown University)</i>	
Author Index	617