Printable Program
Printed on: 05/16/2024

Tuesday, June 7

Registration
SDSS Hours
Tue, Jun 7, 7:30 AM - 6:30 PM
Allegheny Grand Ballroom Foyer

SC1 - CANCELED An Introduction to Spatial Statistics
Short Course
Tue, Jun 7, 8:30 AM - 5:30 PM

SC2 - CANCELED Data Visualization, from Zero to Publication
Short Course
Tue, Jun 7, 8:30 AM - 12:30 PM

SC3 - Data Science and Natural Language Processing to Extract Information from Health Care Data
Short Course
Tue, Jun 7, 8:30 AM - 12:30 PM
Allegheny II

Instructor(s): VG Vinod Vydiswaran, University of Michigan

In recent years, many data science and natural language processing (NLP) approaches have gained popularity to derive insights from clinical narratives. Clinical NLP methods based on deep learning models have shown promising results in various information extraction tasks, such as cohort selection, medication extraction, and de-identification of patient-identifiable information. These methods and tools have also been successfully applied to facilitate clinical research, as well as to support healthcare applications.In this mini-course, I will highlight some of the challenges and potential solutions for large-scale processing of clinical narratives. I will introduce the task of clinical information extraction and present data science approaches to identify relevant information from clinical narratives. I will highlight state-of-the-art algorithms and toolkits to identify medical concepts and entities in clinical text. Based on specific examples from multiple clinical domains, I

will lead a hands-on demo for developing an NLP pipeline for clinical information extraction.

SC4 - CANCELED How to Run the DataTrail Program to Expand Access to Data Science Through a Community-Based Education Program

Short Course
Tue, Jun 7, 8:30 AM - 12:30 PM

SC5 - Statistical Data Privacy Techniques for Sharing Sensitive Data

Short Course
Tue, Jun 7, 1:30 PM - 5:30 PM
Allegheny II

Instructor(s): Joshua Snoke, RAND Corporation

The goal of this course is to provide researchers and data practitioners with practical, hands-on experience applying statistical data privacy techniques to produce shareable data sets. Many individuals working in data science applications have data they would like to share but are limited by the need to protect the privacy of the those in the data. We will cover three of the most common approaches that users may wish to implement, (1) variable suppression or recoding, (2) synthetic data, and (3) differentially private techniques. We will cover examples in which practitioners may wish to select one of these methods, give practical guidance on how to evaluate the risk and utility of the protected data, and go through example code (and available packages) for implementing the methods. At the end of the course, participants will have a basic understanding of the statistical data privacy framework, and they will have the necessary tools to start implementing multiple options for producing publicly shareable versions of data.

SC6 - CANCELED Strategies for Success: Early-Stage Collaborating Biostatistics Faculty at Academic Health Centers

Short Course
Tue, Jun 7, 1:30 PM - 5:30 PM

SDSS Opening Mixer & Expo
Social Event
Tue, Jun 7, 5:30 PM - 7:00 PM
Allegheny Grand Ballroom Foyer

Wednesday, June 8

Registration
SDSS Hours
Wed, Jun 8, 8:00 AM - 5:15 PM
Allegheny Grand Ballroom Foyer

SDSS Expo
SDSS Hours
Wed, Jun 8, 8:00 AM - 4:00 PM
Allegheny Grand Ballroom Foyer

GS01 - Welcome and Opening Plenary: Science and Technology Policy Panel
General Session
Wed, Jun 8, 9:00 AM - 10:15 AM
Allegheny Grand Ballroom

Chair(s): Claire Bowen, Urban Institute

CS01 - Using Statistics to Prepare for the Future
Refereed
Wed, Jun 8, 10:30 AM - 12:00 PM
Allegheny Grand Ballroom

Chair(s): Emily Griffith, Data Science Academy, NC State

10:35 AM
Storm-Based Estimation of Design Snow Load on Solar Panels
Kenneth Kin Pomeyie, Utah State University
11:00 AM
Forecasting Weekly Natural Gas Consumption in Residential and Commercial Sectors in the

Northeast Region of the US
Yunwei Cui, Towson University
11:25 AM
Optimal Congestion Control Strategies for Near-Capacity Urban Metros: Informing Intervention via Fundamental Diagrams
Anupriya -, Imperial College London

CS02 - Time Analyses
Refereed
Wed, Jun 8, 10:30 AM - 12:00 PM
Butler

Chair(s): Michael Pokojovy, The University of Texas at El Paso

10:35 AM
Spectral Clustering for Multi-Layer Stochastic Block Models: Theoretical Analysis of Static and Dynamic Settings for Heterophilic Networks
Kevin Lin, University of Pennsylvania
11:00 AM
Forecasting Hierarchical Time Series
Seema Sangari, Kennesaw State University
11:25 AM
A Time-to-Event Framework for Multi-Touch Attribution
Dinah Shender, Google, Inc.

CS03 - Functional Data Analysis
Refereed
Wed, Jun 8, 10:30 AM - 12:00 PM
Cambria

Chair(s): Hasthika Rupasinghe, Appalachian State University

10:35 AM
Deep Neural Network Classifier for Multi-Dimensional Functional Data
Shuoyang Wang, Auburn University
11:00 AM
Optimal Classification for Functional Data
Guanqun Cao, Auburn University
11:25 AM
Forecasting Multivariate Functional Time Series: Multivariate Functional Singular Spectrum Analysis Approaches
Mehdi Maadooliat, Marquette University

CS04 - Financial Data Applications
Refereed
Wed, Jun 8, 1:15 PM - 2:45 PM
Allegheny Grand Ballroom

Chair(s): Faith (Yueqiao) Zhang, University of Massachusetts Amherst

1:20 PM
Realtime Detection of Bitcoin Bubbles and Estimation of Bubble Formation Time
Min Shu, University of Wisconsin-Stout
1:45 PM
Polynomial Quantile Mixture of Hyperbolic Secant Distribution
Mohan Dev Pant, School of Health Professions, Eastern Virginia Medical School
2:10 PM
Early Warning Signals from Early-Exercise Premia
Ricky Rambharat, Office of the Comptroller of the Currency

CS05 - Data Visualization Tools
Refereed
Wed, Jun 8, 1:15 PM - 2:45 PM
Butler

Chair(s): Ali Rahnavard, The George Washington University

1:20 PM
2020 Census County Assessment Tool
Isabel Youngs, Georgetown University
1:45 PM
Exploring Rural Shrink Smart Through Guided Discovery Dashboards
Denise Bradford, University of Nebraska - Lincoln
2:10 PM
Ggdensity: Improved Bivariate Density Visualization in R
James Otto, Baylor University Department of Statistical Science

CS06 - Designing Data Science Curricula
Refereed
Wed, Jun 8, 1:15 PM - 2:45 PM
Cambria

Chair(s): Alicia Lamere, Bryant University

1:20 PM
Teaching Visual Accessibility in the Introductory Data Science Classes: Why, What, When, and How
JooYoung Seo, University of Illinois at Urbana-Champaign
1:45 PM
Evaluation of EDISON's Data Science Framework via Literature Analysis
Karl RB Schmitt, Trinity Christian College

CS07 - Modeling + Non-Parametric Methods
Lightning
Wed, Jun 8, 1:15 PM - 2:45 PM
Fayette

Chair(s): Emily Dodwell, AT&T

1:20 PM
Nonparametric Tests for the Umbrella Alternative in a Mixed Design for a Known Peak
Boampong Adu Asare, United Tribes Technical College
1:25 PM
Skeleton Regression: A Graph-Based Approach to Estimation on Manifold
Zeyu Wei, University of Washington
1:30 PM
Long-Range Dependence in Low-Frequency Earthquake Catalogs
Ariane Ducellier, University of Washington
1:35 PM
Non-parametric identification and estimation of interactions using stochastic intervention target parameters: implications for mixed exposure analysis.
David Brenton McCoy, University of California Berkeley
1:40 PM
Sparse Bayesian Matrix-variate Regression with High-dimensional Data
Hsin-Hsiung Huang, University of Central Florida
1:45 PM
Distribution Free Bootstrap Prediction Intervals After Variable Selection
Lasanthi Watagoda, Appalachian State University
1:50 PM
SMRT: A Structural Model of Latent Ratings and Topics in Text
Desheng Ma, Cornell University
1:55 PM
Alternatives to ANOVA and Regression Amidst Non-normality: Relative Hypothesis Test

Performance
Anthony J. Bishara, College of Charleston
2:00 PM
Oblique and Non-Linear Survival Trees Based on Dipolar Splitting Criteria
Drew Lazar, Ball State University
2:05 PM
Optimisation of relay team selection for various swimming configurations
Gary David Sharp, Nelson Mandela University
2:10 PM
Can a novel human-centered machine learning algorithm predict better than its black-box counterparts? A benchmarking study of transparency-motivated ranked sparsity methods using 66 diverse datasets
Ryan Peterson, Colorado School of Public Health
2:15 PM
A Comparison of Time Series Model Fitting using Traditional Time Series Models vs. Deep Learning Models including RNN and LSTM to Stock Market Data of Big Tech Companies in the US
Benjamin Houghton, Georgetown University

PS01 - Modeling + Non-Parametric Methods, Part 2
Lightning Poster
Wed, Jun 8, 2:45 PM - 3:40 PM
Allegheny I

1
Nonparametric Tests for the Umbrella Alternative in a Mixed Design for a Known Peak
Boampong Adu Asare, United Tribes Technical College
2
Skeleton Regression: A Graph-Based Approach to Estimation on Manifold
Zeyu Wei, University of Washington
3
Long-Range Dependence in Low-Frequency Earthquake Catalogs
Ariane Ducellier, University of Washington
4
Can a novel human-centered machine learning algorithm predict better than its black-box counterparts? A benchmarking study of transparency-motivated ranked sparsity methods using 66 diverse datasets
Ryan Peterson, Colorado School of Public Health
5
Non-parametric identification and estimation of interactions using stochastic intervention target

parameters: implications for mixed exposure analysis.
David Brenton McCoy, University of California Berkeley
6
Sparse Bayesian Matrix-variate Regression with High-dimensional Data
Hsin-Hsiung Huang, University of Central Florida
7
Distribution Free Bootstrap Prediction Intervals After Variable Selection
Lasanthi Watagoda, Appalachian State University
8
Oblique and Non-Linear Survival Trees Based on Dipolar Splitting Criteria
Drew Lazar, Ball State University
9
SMRT: A Structural Model of Latent Ratings and Topics in Text
Desheng Ma, Cornell University
10
Alternatives to ANOVA and Regression Amidst Non-normality: Relative Hypothesis Test Performance
Anthony J. Bishara, College of Charleston
11
Optimisation of relay team selection for various swimming configurations
Gary David Sharp, Nelson Mandela University
12
A Comparison of Time Series Model Fitting using Traditional Time Series Models vs. Deep Learning Models including RNN and LSTM to Stock Market Data of Big Tech Companies in the US
Benjamin Houghton, Georgetown University

CS08 - Classification Methods and Clustering Analysis
Refereed
Wed, Jun 8, 3:45 PM - 5:15 PM
Allegheny Grand Ballroom

Chair(s): Katharine Correia, Amherst College

3:50 PM
A Brief Overview of Explainable and Interpretable AI
William Franz Lamberti, University of Virginia
4:15 PM
K-Means Clustering Applied to the Analysis of Wearables and Biosensors from Clinical Trial Data
Vanja Vlajnic, Colorado State University

CS09 - Bayesian Approaches
Refereed
Wed, Jun 8, 3:45 PM - 5:15 PM

Butler

Chair(s): Guanqun Cao, Auburn University

3:50 PM
Learning Bayesian Networks Through Birkhoff Polytope: A Relaxation Method
Aramayis Dallakyan, Texas A&M University
4:15 PM
Model Selection in Gaussian and Poisson Longitudinal Distributed Lag Models with Variational AICs
Mark J Meyer, Georgetown University
4:40 PM
FROSTY: A High-Dimensional, Scale-Free Bayesian Network Learning Method
Joshua Bang, University of California, Santa Barbara

CS10 - Applications in Social & Behavioral Sciences
Lightning
Wed, Jun 8, 3:45 PM - 5:15 PM
Fayette

Chair(s): Donna LaLonde, American Statistical Association

3:50 PM
Predicting Census Survey Response Rates via Additive Regression with Interactions
Shibal Ibrahim, MIT
3:55 PM
Partial Association Between Mixed Data: Assessing the Impact of COVID-19 on College Student Well-Being
Zhaohu(Jonathan) Fan, University of Cincinnati
4:00 PM
Does the state-based forward guidance change the way policymakers talk about the outlook and the way nancial markets respond to economic news?
Taeyoung Doh, Federal Reserve Bank of Kansas City
4:05 PM
Modeling the Covid effect on Gasoline Price Changes using Latent Markov Models
Rasitha R Jayasekare, Butler University
4:10 PM
Understanding information about COVID-19: how reliability of used sources and level of understanding influence adherence to sanitary measures in Canada
Clémentine Courdi, Université de Montréal
4:15 PM
The Data Mine: Experiential Industry Practicums in Data Science
Margaret Betz, Purdue University - The Data Mine
4:20 PM
Data Science Consulting and Collaboration: A Cooperative Adventure
Mara Blake, NC State University Libraries

4:25 PM
Patterns of Mental Health Problems Among General Population Before and After Easing COVID-19 Restrictions
Depeng Jiang, University of Manitoba
4:30 PM
Multivariate time series analysis and forecasting of US unemployment rate
VIJAYKUMAR RAJARAM REDDIAR, CENTRAL CONNECTICUT STATE UNIVERSITY
4:35 PM
A Fast Initial Response Approach to Real-Time Financial Surveillance
Andrews T. Anum, The University of Texas at El Paso
Thursday, June 9

Registration
SDSS Hours
Thu, Jun 9, 8:00 AM - 5:15 PM
Allegheny Grand Ballroom Foyer

SDSS Expo
SDSS Hours
Thu, Jun 9, 8:00 AM - 4:00 PM
Allegheny Grand Ballroom Foyer

Speed Mentoring (pre-registration required)
Social Event
Thu, Jun 9, 8:00 AM - 8:45 AM
Allegheny I

Are you looking for a quick way to make connections, solicit career advice, and develop professional relationships? Or maybe you want to provide advice and guidance to early-career statisticians and data scientists? Whether you are interested in mentoring or being mentored, you should consider participating in our speed mentoring session. This is a great opportunity for both mentors and mentees to build their professional networks!

Registration Forms: https://ww2.amstat.org/meetings/sdss/2022/events.cfm

GS02 - Plenary: Career Panel
General Session
Thu, Jun 9, 8:45 AM - 9:45 AM
Allegheny Grand Ballroom

Chair(s): Alicia Lamere, Bryant University

PS02 - Applications in Social & Behavioral Sciences, Part 2
Lightning Poster
Thu, Jun 9, 9:50 AM - 10:30 AM
Allegheny I

1
Predicting Census Survey Response Rates via Additive Regression with Interactions
Shibal Ibrahim, MIT
2
Partial Association Between Mixed Data: Assessing the Impact of COVID-19 on College Student Well-Being
Zhaohu(Jonathan) Fan, University of Cincinnati
3
Does the state-based forward guidance change the way policymakers talk about the outlook and the way nancial markets respond to economic news?
Taeyoung Doh, Federal Reserve Bank of Kansas City
4
Modeling the Covid effect on Gasoline Price Changes using Latent Markov Models
Rasitha R Jayasekare, Butler University
5
Group-based trajectories and predictors of adherence to four key sanitary measures during the COVID-19 pandemic
Sahar Ramazan Ali, Université de Montréal
6
The Data Mine: Experiential Industry Practicums in Data Science
Margaret Betz, Purdue University - The Data Mine
7
Data Science Consulting and Collaboration: A Cooperative Adventure
Mara Blake, NC State University Libraries
8
Patterns of Mental Health Problems Among General Population Before and After Easing COVID-19 Restrictions
Depeng Jiang, University of Manitoba
9
Multivariate time series analysis and forecasting of US unemployment rate
VIJAYKUMAR RAJARAM REDDIAR, CENTRAL CONNECTICUT STATE UNIVERSITY
10
A Fast Initial Response Approach to Real-Time Financial Surveillance
Andrews T. Anum, The University of Texas at El Paso

LB - Statistical Approaches for Privacy-Preserving Methodologies
Late Breaking
Thu, Jun 9, 10:30 AM - 12:00 PM
Allegheny Grand Ballroom

Chair(s): Claire Bowen, Urban Institute

10:35 AM
A Formal Privacy Framework for Partially Private Data
Jeremy Seeman, Penn State University
11:00 AM
Developing a Feasible Differentially Private Validation Server for Administrative Tax Data
Joshua Snoke, RAND Corporation
11:25 AM
Differential Privacy on Manifolds
Carlos Soto, Penn State University

CS11 - Data Science in Clinical Contexts
Refereed
Thu, Jun 9, 10:30 AM - 12:00 PM
Butler

Chair(s): Ryan Peterson, Colorado School of Public Health

10:35 AM
When to Initiate Cancer Screening Exam if Sensitivity Is a Function of Sojourn Time?
Dongfeng Wu, University of Louisville
11:00 AM
Active Data Science for Improving Clinical Risk Prediction
Donna Pauler Ankerst, Technical University of Munich
11:25 AM
Investigating Racial Disparities in Assisted Reproductive Technology Utilization and Outcomes: A Case Report on a Complex Missing Data Problem
Katharine Correia, Amherst College

CS12 - High-dimensional Statistics
Refereed
Thu, Jun 9, 10:30 AM - 12:00 PM
Cambria

Chair(s): Xiaoqian Liu, North Carolina State University

10:35 AM
Comparing Methods for Statistical Inference with Model Uncertainty
Anupreet Porwal, University of Washington
11:00 AM
UniCATE: Flexible Predictive Biomarker Discovery

Philippe Boileau, Graduate Group in Biostatistics and Center for Computational Biology, UC Berkeley; Roche Canada
11:25 AM
Generalizable Manifold Learning for Dimensional Reduction
Jungeum Kim, Purdue University

CS13 - Data Mining and Deep Learning
Refereed
Thu, Jun 9, 10:30 AM - 12:00 PM
Fayette

Chair(s): Mingzhao Hu, University of California, Santa Barbara

10:35 AM
Model-Agnostic AI Assurance Scoring Framework
Md Nazmul Kabir Sikder, Virginia Tech
11:00 AM
Malfunction Analysis in Processes Based on Rule Mining
Benoit Vuillemin, IRMA - UMR 7501 - CNRS
11:25 AM
Random Forest to Estimate a Dose-Response Relationship in Quasi-Experimental Student Success Studies
Richard A Levine, San Diego State University

CS14 - Applying and Evaluating Logistic Regression Models
Refereed
Thu, Jun 9, 1:15 PM - 2:45 PM
Allegheny Grand Ballroom

Chair(s): Seema Sangari, Kennesaw State University

1:20 PM
Searching the Web for the Drone Industry: Classifying Websites in Multiple Countries and Languages with a Single Model
Piet J.H. Daas, Statistics Netherlands
1:45 PM
Residuals and Diagnostics for Multinomial Regression Models
Eric Anthony El-Khouri Gerber, California State University, Bakersfield

CS15 - Assessing and Evaluating Data with Visualizations

Refereed
Thu, Jun 9, 1:15 PM - 2:45 PM
Butler

Chair(s): Maya Shen, Carnegie Mellon University

1:20 PM
‘You Draw It’: Implementation of Visually Fitted Trends with R2d3
Emily A. Robinson, University of Nebraska - Lincoln
1:45 PM
Model Diagnostics of Discrete Data Regression: A Unifying Framework Using Functional Residuals
Zewei Lin, University of Cincinnati
2:10 PM
Designing COVIDcast 2.0, Implementing Emergent Data Visualization Designs from the COVID-19 Pandemic
Chris Scott, Google, Inc.

CS16 - Advancements in Machine Learning
Refereed
Thu, Jun 9, 1:15 PM - 2:45 PM
Cambria

Chair(s): Qiyu Wang, Zhejiang University of Finance and Economics

1:20 PM
Combining Capsule Networks and Self-Supervised Learning for Image Classification with Occlusion
Ladyna Wittscher, Friedrich-Schiller-University Jena, Germany
1:45 PM
A Convex-Nonconvex Strategy for Grouped Variable Selection
Xiaoqian Liu, North Carolina State University
2:10 PM
Recalibrating Probability Density Estimates Using Feature-Space Regression
Biprateep Dey, University of Pittsburgh

CS17 - Data-driven Healthcare
Lightning
Thu, Jun 9, 1:15 PM - 2:45 PM
Fayette

Chair(s): Xiaoqing "Ellen" Tan, University of Pittsburgh

1:20 PM
R-Based Clinical Trial Submission to FDA
Ning Leng, Genentech, Inc.
1:25 PM
A Predictive Model for Speech Rehabilitation for Patients with Parkinson's Disease
Ismail EL Moudden, Eastern Virginia Medical School Sentara Healthcare Analytics and Delivery Science Institute
1:30 PM
fiBAG: Functional Integrative Bayesian Analysis of High-dimensional Multiplatform Genomic Data
Rupam Bhattacharyya, University of Michigan
1:35 PM
Defining New Staging Criteria for Metastatic Breast Cancer Using Recursive Partitioning
Samantha M. Thomas, Duke University
1:40 PM
CAMO: A molecular congruence analysis framework for evaluating model organisms
Wei Zong, Department of Biostatistics, University of Pittsburgh
1:45 PM
Comparing Methods for Evaluating the Proportional Hazards Assumption for Time-to-Event Survey Data
John R Pleis, National Center for Health Statistics
1:50 PM
Variable selection in the development of ICU scoring systems
Gary David Sharp, Nelson Mandela University
1:55 PM
Modeling COVID-19 disruptions in longitudinal health registry data: a case study of conventional methods producing incoherent results
Raymond Pomponio, Colorado School of Public Health
2:00 PM
Neural-Network Models for Long-term Care Insurance Insolvency: An Enterprise Risk Management Approach
Sebastain Awondo, University of Alabama
2:05 PM
CircadianPipeline: A pipeline for differential rhythmicity analysis in R/Shiny
Xiangning Xue, University of Pittsburgh
2:10 PM
Delayed Phase Scan Prediction for Multiphase Liver CT Dose Reduction
Bin Chen, Purdue University Fort Wayne

PS03 - Data-driven Healthcare, Part 2
Lightning Poster
Thu, Jun 9, 2:45 PM - 3:40 PM
Allegheny I

1
R-Based Clinical Trial Submission to FDA
Ning Leng, Genentech, Inc.
2
A Predictive Model for Speech Rehabilitation for Patients with Parkinson's Disease
Ismail EL Moudden, Eastern Virginia Medical School Sentara Healthcare Analytics and Delivery Science Institute
3
fiBAG: Functional Integrative Bayesian Analysis of High-dimensional Multiplatform Genomic Data
Rupam Bhattacharyya, University of Michigan
4
Defining New Staging Criteria for Metastatic Breast Cancer Using Recursive Partitioning
Samantha M. Thomas, Duke University
5
CAMO: A molecular congruence analysis framework for evaluating model organisms
Wei Zong, Department of Biostatistics, University of Pittsburgh
6
Comparing Methods for Evaluating the Proportional Hazards Assumption for Time-to-Event Survey Data
John R Pleis, National Center for Health Statistics
7
Variable selection in the development of ICU scoring systems
Gary David Sharp, Nelson Mandela University
8
Modeling COVID-19 disruptions in longitudinal health registry data: a case study of conventional methods producing incoherent results
Raymond Pomponio, Colorado School of Public Health
9
Delayed Phase Scan Prediction for Multiphase Liver CT Dose Reduction
Bin Chen, Purdue University Fort Wayne
10
Neural-Network Models for Long-term Care Insurance Insolvency: An Enterprise Risk Management Approach
Sebastain Awondo, University of Alabama
11
CircadianPipeline: A pipeline for differential rhythmicity analysis in R/Shiny
Xiangning Xue, University of Pittsburgh

CS18 - Improving Algorithms for Big Data
Refereed
Thu, Jun 9, 3:45 PM - 5:15 PM
Allegheny Grand Ballroom

Chair(s): Yu "Wayne" Wang, University of Michigan

3:50 PM
Building the Foundation for More Flexible A/B Testing: Applications of Interim Monitoring to Large-Scale Data
Wenru Zhou, University of Colorado
4:15 PM
Time Series Anomaly Detection in the Age of Big Data: Matching Data Generation Processes with Algorithms
Gorkem Turgut Ozer, University of New Hampshire

CS19 - Contributions to Software and Technology
Refereed
Thu, Jun 9, 3:45 PM - 5:15 PM
Butler

Chair(s): Emily Dodwell, AT&T

3:50 PM
Csurvey: Implementing Order Constraints in Survey Data Analysis
Xiyue Liao, California State University, Long Beach
4:15 PM
WITHDRAWN Practical Target-Based Synchronization Strategies for Immutable Time-Series Data Tables

CS20 - Neural Network Analysis
Refereed
Thu, Jun 9, 3:45 PM - 5:15 PM
Cambria

Chair(s): Ilia Sucholutsky, University of Waterloo

3:50 PM
Predicting Crop-Specific Land Cover Using Transition Probabilities, Deep and Quantum-Inspired Neural Network Models
Luca Sartore, National Institute of Statistical Sciences
4:15 PM
Comprehensive Evaluation of CNNs for Pollen Classification
Predrag Matavulj, BioSense Institute
4:40 PM
A Novel Architecture Combining Central-Peripheral Deviation with Convolutional Neural Networks for Diffusion Tensor Imaging Studies
Soyun Park, University at Buffalo

CS21 - New Models, Methods, and Applications I
Lightning
Thu, Jun 9, 3:45 PM - 5:15 PM
Fayette

Chair(s): Francis Bilson Darku, University of Notre Dame

3:50 PM
CGMM: an algorithm for constrained model-based clustering
Jian Zou, Department of Biostatistics, School of Public Health, University of Pittsburgh
3:55 PM
A Semiparametric Modeling Approach for Analyzing Clinical Biomarkers Restricted to Limits of Detection
Sandipan Dutta, Old Dominion University
4:00 PM
Bayesian Poisson Model with Spatio-temporal Structure for Mortality Projection of Multi-population
Zhen Liu, Department of Mathematics & Statistics, Georgetown University
4:05 PM
Empirically adjusted weighted ordered p-values method for meta-analysis
Sinjini Sikdar, Old Dominion University
4:10 PM
Confidence Intervals for Genetic Correlation via Parametric Bootstrap
Yi-Ting Tsai, Harvard T.H. Chan School of Public Health
4:15 PM
A California Wetland Case Study: A Novel, Predictive Approach to Monitor Estuarine Health
Vedant Janapaty, Silver Creek High School
4:20 PM
Use of Process Crowding in Conditional WGAN for Remaining Process Events Prediction
Yoann Valero, LIST3N, Université de Technologie de Troyes
4:25 PM
Variable Importance Confidence Intervals within Random Forest
Heather Lynn Cook, University of Southern Indiana
4:30 PM
A New, Global Estimate of Biocrust Carbon and Nitrogen Flux from Terrestrial Ecosystems
Shloka V. Janapaty, Columbia University
Friday, June 10

Registration
SDSS Hours
Fri, Jun 10, 8:00 AM - 12:30 PM
Allegheny Grand Ballroom Foyer

SDSS Expo
SDSS Hours
Fri, Jun 10, 8:00 AM - 11:30 AM
Allegheny Grand Ballroom Foyer

PS04 - New Models, Methods, and Applications I, Part 2
Lightning Poster
Fri, Jun 10, 8:15 AM - 9:00 AM
Allegheny I

1
Use of Process Crowding in Conditional WGAN for Remaining Process Events Prediction
Yoann Valero, LIST3N, Université de Technologie de Troyes
2
Variable Importance Confidence Intervals within Random Forest
Heather Lynn Cook, University of Southern Indiana
3
CGMM: an algorithm for constrained model-based clustering
Jian Zou, Department of Biostatistics, School of Public Health, University of Pittsburgh
4
A Semiparametric Modeling Approach for Analyzing Clinical Biomarkers Restricted to Limits of Detection
Sandipan Dutta, Old Dominion University
5
Bayesian Poisson Model with Spatio-temporal Structure for Mortality Projection of Multi-population
Zhen Liu, Department of Mathematics & Statistics, Georgetown University
6
Empirically adjusted weighted ordered p-values method for meta-analysis
Sinjini Sikdar, Old Dominion University
7
Confidence Intervals for Genetic Correlation via Parametric Bootstrap
Yi-Ting Tsai, Harvard T.H. Chan School of Public Health
8
A California Wetland Case Study: A Novel, Predictive Approach to Monitor Estuarine Health
Vedant Janapaty, Silver Creek High School
9
A New, Global Estimate of Biocrust Carbon and Nitrogen Flux from Terrestrial Ecosystems

Shloka V. Janapaty, Columbia University

CS22 - Contributions to Model Methods and Applications
Refereed
Fri, Jun 10, 9:00 AM - 10:30 AM
Allegheny Grand Ballroom

Chair(s): Claire Bowen, Urban Institute

9:05 AM
Mixed Effects State Space Models on Research and Development Performance in the New Energy Vehicle Industry
Mingzhao Hu, University of California, Santa Barbara
9:30 AM
Bayesian Nested Latent Class Models for Cause-of-Death Assignment Using Verbal Autopsies Across Multiple Domains
Zehang Li, University of California, Santa Cruz
9:55 AM
High-Dimensional Causal Mediation Analysis Based on Partial Linear Structural Equation Models
Xizhen Cai, Williams College

CS23 - Non-Parametric Approaches
Refereed
Fri, Jun 10, 9:00 AM - 10:30 AM
Butler

Chair(s): Kevin Lin, University of Pennsylvania

9:05 AM
The AUGUST Two-Sample Test: Powerful, Interpretable, and Fast
Benjamin Lewis Brown, Statistics and Operations Research, UNC Chapel Hill
9:30 AM
A Computational Perspective on Projection Pursuit in High Dimensions: Feasible or Infeasible Feature Extraction
Chunming Zhang, University of Wisconsin-Madison

CS24 - New Models, Methods, and Applications II
Lightning
Fri, Jun 10, 9:00 AM - 10:30 AM
Cambria

Chair(s): Yaomin Xu, Vanderbilt University Medical Center

9:05 AM
Manifold learning analysis suggests novel strategies to align single-cell multi-modal data of neuronal electrophysiology and transcriptomics
Jiawei Huang, University of Cincinnati
9:10 AM
Spatiotemproal Zero-Inflated Bayesian Negative Binomial Regression Nearest Neighbor Gaussian Process Models
Hsin-Hsiung Huang, University of Central Florida
9:15 AM
Imperfect Imputation: Adjusting for the Error Incurred when We Impute
Kyle Frederic Grosser, University of North Carolina at Chapel Hill
9:20 AM
A Case Study with RCT of Varenicline Using Two Machine Learning Approaches
Alondra Cruz, University of California, Los Angeles
9:25 AM
A Brief Review of Blockchain Technology and Impact on Practice of Data Science
David Han, UT San Antonio
9:30 AM
Partial Aggregation Imputation in Geographical Energy Statistics Reporting Networks
Glen Haynes, Energy Information Administration
9:35 AM
On p-value combination of independent and frequent signals
Yusi Fang, Department of Biostatistics, Univertisy of Pittsburgh
9:40 AM
Random Forest Is a Robust Model Choice on Feature Transformed Data for Binary Classification Task
Emma Minasyan, Mimecast
9:45 AM
Ensemble Learning Models for Biomass Estimation and Species Classification of Intertidal Macroalgae Using In-situ and Remote Sensing Spectrometry
Ernst Linder, University of New Hampshire

PS05 - New Models, Methods, and Applications II, Part 2
Lightning Poster
Fri, Jun 10, 10:30 AM - 11:25 AM
Allegheny I

1
Manifold learning analysis suggests novel strategies to align single-cell multi-modal data of neuronal electrophysiology and transcriptomics
Jiawei Huang, University of Cincinnati
2

Spatiotemproal Zero-Inflated Bayesian Negative Binomial Regression Nearest Neighbor Gaussian Process Models
Hsin-Hsiung Huang, University of Central Florida
3
Imperfect Imputation: Adjusting for the Error Incurred when We Impute
Kyle Frederic Grosser, University of North Carolina at Chapel Hill
4
A Case Study with RCT of Varenicline Using Two Machine Learning Approaches
Alondra Cruz, University of California, Los Angeles
5
A Brief Review of Blockchain Technology and Impact on Practice of Data Science
David Han, UT San Antonio
6
Partial Aggregation Imputation in Geographical Energy Statistics Reporting Networks
Glen Haynes, Energy Information Administration
7
On p-value combination of independent and frequent signals
Yusi Fang, Department of Biostatistics, Univertisy of Pittsburgh
8
Random Forest Is a Robust Model Choice on Feature Transformed Data for Binary Classification Task
Emma Minasyan, Mimecast
9
Ensemble Learning Models for Biomass Estimation and Species Classification of Intertidal Macroalgae Using In-situ and Remote Sensing Spectrometry
Ernst Linder, University of New Hampshire

LB - CANCELED Leland Wilkinson: A Graphic Life
Late Breaking
Fri, Jun 10, 11:30 AM - 1:00 PM
Allegheny Grand Ballroom

CS25 - Methods and Studies to Identify Important Variables
Refereed
Fri, Jun 10, 11:30 AM - 1:00 PM
Butler

Chair(s): Donna LaLonde, American Statistical Association

11:35 AM
Analyzing the Impact of Different Countries' Approaches to the COVID-19 Pandemic on Their

Cumulative Infection Curves by Using Nonparametric Density Regression and Clustering Methods
Damian Musk, Stanford OHS
12:00 PM
Water Statistics: A Study Comprehending Several Techniques and Periods at a University
Elisa Henning, Santa Catarina State University
12:25 PM
Partially Constrained Group Variable Selection to Adjust for Complementary Unit Performance in American College Football
Andrey Skripnikov, New College of Florida

CS26 - Cluster and Graphical Analyses
Refereed
Fri, Jun 10, 11:30 AM - 1:00 PM
Cambria

Chair(s): Claire Bowen, Urban Institute

11:35 AM
Conservative Causal Discovery by Use of Supervised Machine Learning
Anne Helby Petersen, University of Copenhagen
12:00 PM
Accounting for Model Misspecification When Using Pseudolikelihood for ERGMs
David R Hunter, Penn State University
12:25 PM
A New Algorithm for Robust Affine-Invariant Clustering
Michael Pokojovy, The University of Texas at El Paso

GS03 - SDSS Wrap-Up & Fireside Chat
General Session
Fri, Jun 10, 1:10 PM - 1:30 PM
Allegheny Grand Ballroom

Chair(s): Emily Dodwell, AT&T