2022 SDSU data science symposium

Workshop abstracts

Workshop 1 – Biomarker discovery and pathway enrichment analysis of omics data
Instructors: Ali Rahnavard, George Washington University and Himel Mallick, Merck

Workshop 1 description:

Participants will gain hands-on experience with these analyses using tools for pattern discovery in multi-omics. Interspersed with lecture content, attendees will work through multi-omics analysis exercises with real data. Participants are strongly encouraged to bring their own data and study examples for application. Open to computational biologists, bioinformaticians, principal investigators, and their research teams including advanced Ph.D. students. Basic familiarity with multi-omics upstream bioinformatics tools are recommended. Beginner-level familiarity with R is required. Methodological advancements paired with measured multi-omics data using high-throughput technologies enable capturing comprehensive snapshots of biological activities. In particular, low-cost, culture-independent omics profiling has made metagenomics, metabolomics, and proteomics ("multi-omics") surveys of human health, other hosts, and the environment. The resulting data have stimulated the development of new statistical and computational approaches to analyze and integrate omics data, including human gene expression, microbial gene products, metabolites, and proteins, among others.

Multi-omics data generated from diverse platforms are often fed into generic downstream analysis software without proper appreciation of the inherent data differences, which could result in incorrect interpretations. Further, there are also a large collection of downstream analysis software platforms and appropriately selecting the best tool can be extraneous for untrained researchers.

In this workshop, we will thus present a high-level introduction to computational multi-omics, highlighting the state-of-the-art in the field as well as outstanding challenges geared towards downstream analysis methods. This will include an introduction to the biological goals of typical multi-omics studies and the statistical methods currently available to achieve them.

Outline of workshop 1:
We will begin with an overview of the statistical challenges inherent to analyzing the compositional data arising in multi-omics studies. Introductory lectures will include: The challenges associated with precisely testing for multivariable association testing in population-scale meta-omics studies Challenges and advances in pathway enrichment analyses including techniques and characterization of omics features Meta-analysis of multiple datasets for high-sensitivity discovery and integration with other types of data such as electronic health records and imaging Workshop attendees will gain hands-on experience with these analyses using tools for pattern discovery in multi-omics. Interspersed with lecture content, attendees will work through multi-omics analysis tutorials. Tools will include:

- [Tweedieverse tutorial](#) and [Tweedieverse examples](#): A unified statistical framework for differential analysis of multi-omics data
- [deepath](#): omics pathway enrichment analysis
- [omeClust](#): Omics community detection using multi-resolution clustering
- [IntegratedLearner](#): Integrated machine learning for multi-omics prediction and classification to stratify patients for therapeutic intervention
- Publication-quality figure generation and effective visualization of the results

This workshop will be run by a joint effort between George Washington University and Merck Research Laboratories. Researchers from both industry and academia will come together to share a diverse perspective on the topic both from drug discovery and basic science angles, enabling attendees to achieve a holistic view of multi-omics and clinical data integration through the use of state-of-the-art tools applied to motivating examples and use cases.

Workshop 1: Instructions

1. Install the latest R and RStudio on your local computer
2. Install listed software in the learning objectives
3. Try to run demos of each software
4. Bring your data to apply these techniques

Workshop 2 – Intro to Microsoft PowerBI
Instructor: Negassi Tesfay, Premier Bankcard

Workshop 2 description:

PowerBI is an interactive data visualization and reporting tool that is used in businesses. In this workshop, you will learn how to apply Power BI for best practices that are in line with current businesses and technical requirements for extracting, loading, modeling, visualizing, and analyzing data.

Outline of workshop 2:
Session 1: Introduction to Power BI
Session 2: Get Data in Power BI
Session 3: Harvest, Transform, and Load Data in Power Query
Session 4: Design a Data Model in Power BI
If time allows
Session 5: Creating Measures using DAX in Power BI
Session 6: Optimizing Model Performance
Session 7: Creating Reports

Workshop 2: Instructions
1. No specific prerequisites. Familiarity with data analytics will help.
2. Bring windows laptop (PowerBI does not support Mac machines)
3. Install PowerBI from powerbi.microsoft.com. FREE. No sign up required for this workshop.
4. More information will be sent directly to workshop participants closer to the symposium date

Workshop 3 – Blockchain data analytics
Instructor: Cuneyt Akcora, University of Manitoba, Canada

Workshop 3 description:

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.

In this tutorial, we offer a holistic view on applied Data Science on Blockchains. Starting with the core components of Blockchain, we will detail the state of art in Blockchain data analytics for graph and security domains. Our examples will answer questions, such as, how to parse, extract and clean the data stored in blockchains?, how to store and query Blockchain data? and what features could be computed from blockchains? We will cover use cases such as ransomware and darknet market payment detection.

Tentative outline of the workshop 3:

1. Introduction to the concepts of Blockchain
2. Introduction to Blockchain data mining tools
3. Temporal analysis of Bitcoin transactions
4. Address and transaction embeddings for graph machine learning
5. Darknet market payment detection
6. Ransomware payment detection

Workshop 3 instructions

1. Bring laptops: No specific prerequisites are required.
2. Java and Python 3 is needed. Installation via Anaconda distribution is recommended. Scikit-learn is needed. Jupyter notebook is needed.