Poster Abstracts WiDS Central Mass @ WPI
March 4, 2019

Attention-based Approaches to Question Answering

Mili Shah, Kartik Chhapia, Mohit Iyyer

Crassostrea virginica Oysters: From Genomics to Proteomics

K. S. Reed, J. A. F. Robledo, Tracy A. N. Yadavalli, D. Bailey-Bishop

Bivalves, from raw oysters to steamed clams, are popular choices among seafood lovers and once limited to the coastal areas. For this reason, oysters, mussels, scallops, and clams have been the focus of research for improving the production and investigating basic biological and ecological questions. During this decade, an impressive amount of information using high-throughput genomic, transcriptomic and proteomic technologies has been produced in various classes of the Mollusca group. In this study, we use basic research to chart the oyster genome. It is intended that this data will have implications outside of the oyster by providing genetic information that modulate stem cells and cell differentiation, the ability to fight opportunistic and specific pathogens in the absence of adaptive immunity, as source of alternative drugs, mucosal immunity and, microbiome turnover, toxicology, and cancer research.

Graduate Student Aspiring a DS Job

Xiaoying Shi

Identifying When Effect Restoration Will Improve Estimates of Causal Effect

Akanksha Atrey, Hüseyin Oktay, David Jensen

Several methods have been developed that combine multiple models learned on different data sets and then use that combination to reach conclusions that would not have been possible with any one of the models alone. We examine one such method—effect restoration—which was originally developed to mitigate the effects of poorly measured confounding variables in a causal model. We show how effect restoration can be used to combine results from different machine learning models and how the combined model can be used to estimate causal effects that are not identifiable from either of the original studies alone. We characterize the performance of effect restoration by using both theoretical analysis and simulation studies. Specifically, we show how conditional independence tests and common assumptions can help distinguish when effect restoration should and should not be applied, and we use empirical analysis to show the

limited range of conditions under which effect restoration should be applied in practical situations.

Low Resource Supervised Domain Adaptation with Generalization Across Domains

Rheeya Uppaal

Current state of the art methods in Domain Adaptation follow adversarial approaches, making training a challenge. Other non-adversarial methods learn mappings between source and target domains, to achieve reasonable performance. However, even these methods do not focus a key aspect of maintaining performance on the source domain, even after optimizing over the target domain. Additionally, there exist very few methods in low resource supervised domain adaptation. This work proposes a method, LRS-DAG, that aims to solve these current issues in the field. By adding a set of "encoder layers" which map the target domain to the source, and can be removed when dealing directly with the source data, the model learns to perform optimally on both domains. LRS-DAG is unique in the sense that a new algorithm for low resource domain adaptation, which maintains performance over the source, with a new metric for learning mappings has been introduced.

Neural Relation Extraction with Entity Features

Twinkle Tanna

Performance of relation extraction on unstructured data is highly dependent on the domain. This work explains a project in the domain of mergers and acquisitions that performs relation extraction using position and entity embeddings using novel deep learning techniques. It also demonstrates the use of crowdsourcing platforms to work with unlabelled unstructured data.

Using an Agent-Based Model to Estimate Time to Restoration of Storm-Induced Power Outages

Tara Walsh, David Wanik, Thomas Layton, Jonathan Mellor

Extreme weather can cause severe damage and widespread power outages across utility service areas. The restoration process can be long and costly and emergency managers may have limited computational resources to optimize the restoration process. This study takes an agent based modeling (ABM) approach to optimize the utility storm recovery process in Connecticut. The ABM is able to replicate past storm recoveries and can test future case scenarios. We found that parameters such as the number of outages, repair time range and the number of utility crews working can substantially impact the estimated time to restoration (ETR).

Other parameters such as crew starting locations and travel speeds had comparatively minor impacts on the ETR. The ABM can be used to train new emergency managers as well as test strategies for storm restoration optimization.

VTeller: Telling the Values Somewhere, Sometime in a Dynamic Network of Urban Systems

Yan Li

Dynamic networks are very common in urban systems today. As data are acquired, unfortunately, they are rarely complete observations of the whole system. It is important to reliably infer the unobserved attribute values anywhere in the graphs, at certain times—either in the past or in the future. Previous work does not sufficiently capture the correlations inherent with graph topology and with time. We propose a machine learning approach using a novel probabilistic graphical model. We devise a series of algorithms to efficiently group the vertices, to learn the model parameters, and to infer the unobserved values for query processing. Furthermore, we propose a method to incrementally and automatically update the model. Finally, we perform an extensive experimental study using two real-world dynamic graph datasets to evaluate our approach.

Wenqiu Cao

Martha Clump