Student Poster Session Abstracts

March 2, 2020

Title: Entropy of a bacterial stress response is a generalizable predictor for fitness and antibiotic sensitivity.

Author(s): Defne Surujon; Zeyu Zhu; Juan C. Ortiz-Marquez; José Bento; Tim van Opijnen

Abstract: Bacterial pathogens experience various types of stress (e.g. the immune system, antibiotics), upon which they elicit specific responses. Such responses can be captured by transcriptomic screens, resulting in gene expression data that could be used in inferring the fitness of the bacterium. Existing approaches implement models trained with data coming from specific conditions and species. If these models are to be used for antibiotic susceptibility testing, this necessitates the collection of new training data every time the model is applied to a different pathogenic species or antibiotic. In this study we develop more generalizable models for predicting bacterial fitness in a species or condition agnostic way. First, by training a regression model with gene expression data from a variety of conditions and strains of S. pneumoniae, we obtain a model that applies to more than one condition. Second, we use the well-established statistical definition of entropy to quantify the extent of transcriptional perturbation in the organism. We demonstrate that entropy predicts fitness for 19 antibiotic and non-antibiotic conditions, and in 7 species of pathogenic bacteria. We further show that entropy correlates with the minimum inhibitory concentration of an antibiotic, and can therefore be used as a predictor of antibiotic susceptibility.

Title: A Tangible Interaction Technique for Dynamic Data Physicalization

Author(s): Clara Richter, Bridger Herman, Daniel F. Keefe

Abstract: This project proposes a more effective means for researchers to interact with large data sets in a physical as well as a visual way. We hypothesize the combined interaction of vision and touch will be a more effective means for researchers to interact with spatially complex datasets, leading to a better understanding of spatial relationships. Our approach is to combine 3D printed data models with virtual data overlays. To make the resulting hybrid digital+physical displays interactive, we introduce a new technique for sensing touch on 3D models, such as terrains, that can be placed on top of a pressure-sensing mat. The results demonstrate a new ability to interact with a combination of dynamic data and contextual static 3D printed data using physical touch.

Title: Estimating co-occurrences of multiple chronic conditions using maximum entropy

Author(s): Pracheta Amaranath,Peter Haas, Hari Balasubramaniam, Ninad Khargonkar, Prasanna Srinivasan, Roshan Prakash

Abstract: The prevalence of multiple chronic disease conditions among the general population is steadily increasing today. Multiple chronic conditions (MCCs) account for a large portion of the expenditure in healthcare. A baseline model anticipating the probabilities of co-occurrence of these conditions can serve to organize healthcare delivery systems, inform healthcare planning operations and even drive decision support for personalized medicine. In this work, we seek to identify frequently occurring MCC clusters and estimate the prevalence of the occurrences in a population when the data is sparse. We derive key insights into the most common MCC clusters by using market basket analysis on the MEPS (Medical Expenditure Panel Survey) data. Using the principle of maximum entropy, we provide an estimate of the probability distribution of co-occurring chronic conditions and validate our results against a synthetically generated dataset. We present our findings and formulate an algorithm to arrive at this distribution given data from a sample population.

Title: Improving Neural Networks with Robust PCA

Author(s): Marissa Bennett, Randy Paffenroth

Abstract: As machine learning (ML) becomes an increasingly relevant field being incorporated into everyday life, so does the need for consistently high performing models.

With these high expectations, along with potentially limited and/or small data sets, it is crucial to be able to use techniques for machine learning that increase the likelihood of success. Robust Principal Component Analysis (RPCA) not only extracts anomalous data, but also finds correlations among the given features in a data set.

By taking a novel approach to utilizing the output from RPCA, we address the question: Do neural networks perform better when additional information about the data is used along with the original data from the data set?

Title: Imitation Learning From Huamn-Generated Spatio-Temporal Data

Authors(s): Weixiao Huang

Abstract: Methods for imitating an expert through learning from demonstrations have been extensively studied. However, it still remains two challenges: 1) tasks with diverse difficulties require different numbers of demonstrations to learn a quality reward function, so how to determine a proper sample size for inverse reinforcement learning (IRL) to guarantee statistical reliability of the learned reward function; and 2) in many real-world cases, a group of people make decisions either synchronously or asynchronously, so how to aggregate all the cases into one framework to efficiently learn their rewards and policies. This study proposes a novel ReliableIRL algorithm and an innovative framework of asynchronous multi-agent generative adversarial imitation learning to solve the problems.

Title: Placement Optimization in Refugee Resettlement

Author(s): Narges Ahani, Data Science PhD at Worcester Polytechnic Institute, Andrew C. Trapp (Worcester Polytechnic Institute), Alexander Teytelboym (University of Oxford , UK), Alessandro Martinello, Tommy Andersson (Lund University, Sweden)

Abstract: Every year thousands of refugees are resettled to dozens of host countries, and there is growing evidence that the initial placement of refugee families profoundly affects their lifetime outcomes. Our research combines techniques from operations research, machine learning, econometrics, and interactive visualization to create the interactive software tool, Annie MOORE (Matching and Outcome Optimization for Refugee Empowerment). Annie is the first software designed for resettlement agencies to recommend data-driven, optimized matches between refugees and local affiliates while respecting refugee capacities. Initial back-testing indicates that Annie can improve short-run employment outcomes by 22%-37%. Future research directions include dynamic pipeline and quota management, incorporating preferences of refugees and local communities, and further software customization.

Title: A Customized Machine Learning Pipeline To Build State-Of-The-Art Audio Classifiers

Authors(s): Sruthi Kurada

Abstract: Audio classifiers have many real-world applications, from informing medical diagnoses to revealing automobile malfunctions. In this study, I have explored strategies to build an accurate classifier to categorize environmental sounds from the UrbanSound8K dataset. Published classifiers on this ten-class dataset only have 50-79% accuracy. Through engineering a machine learning pipeline, I have built a state-of-the-art classifier with a 99% test-set accuracy on this dataset. In order to examine the general applicability of this pipeline to build reliable classifiers on other audio datasets, I have examined its performance in differentiating four unique heart sounds and found it to be equally effective. The final heart sound classifier achieved a 98% test set accuracy.