

Room: 115 Beckman Institute

Efficient Estimation of Small Probabilities of Large Cascading Failures in Power Grids

Leyu Yao

Mentor: Konstantin Zuev

Electric power grids are complex networks with heterogeneous distributions of loads, which makes them prone to cascading failures, where an initially small perturbation leads to a sequence of failures that spread through the connections between system components. Thus, it is crucial to estimate and enhance the reliability of power grids under cascading failures. We focus on developing a meaningful way to characterize the mitigating set of nodes in cascading failures, where deliberately switching off the nodes reduces the overall damage caused by the cascading failures. We implement a strategy proposed by Motter to reduce the damage by intentionally turning off nodes with least initial load on the US Western States Power Grids (USWS). A new strategy of assigning weights to links using the betweenness of their endpoints and removing fractions of leaves in the minimum spanning tree (MST) is developed and tested on USWS, which is more effective in most cases. We also aim to implement more realistic models for assessing the reliability of power grids in the presence of cascading failures.

Development of Strategies for Binding DNA Origami to Graphene and Origami Dimer Formation for OASIS (Origami-Assisted Sandwich ImmunoSorbent) Assay

Emily Wu

Mentors: Paul Rothmund and Ashwin Gopinath

The development of efficient and effective detection and quantification methods is crucial for health management, whether in providing the ability to detect diseases, health conditions, and allergens. Our approach to achieve this is via the development of an electronic OASIS (origami-assisted immunosorbent) assay. One component of this development include the ability to bind origami to graphene in order to take advantage of FET-based technology and provide a route for electrical detection and signal production. Several strategies and chemistries are explored, including the binding affinity of poly-T DNA to graphene and an EDC (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride) carbodiimide crosslinker reaction. Another component is to link two DNA origamis together in order to employ a sandwich-type assay approach. This includes investigating a simple linking mechanism that uses complementary DNA strands.

Extension of Continuous Tensor Decomposition Methods to Control Problems

Vivek Bharadwaj

Mentors: Animashree Anandkumar, Rose Yu, Kamyar Azizzadenesheli

Optimal control problems, such as robot motion planning, often require computations that involve high-dimensional state spaces. As the dimension of the state space grows, both the computation time and the storage space necessary to characterize the state space exactly grows exponentially in the dimension, a phenomenon known as the curse of dimensionality. Recent work has shown that high-dimensional tensors (arrays in more than one dimension) can often be approximated using a low-rank tensor train decomposition; likewise, high-dimensional multivariate functions can be compactly expressed using function train (FT) decomposition, and recent work has shown applications of FT-decomposition to stochastic optimal control problems, where the system dynamics are known in advance. The purpose of this project is to demonstrate the application of FT-decomposition to control problems where the system dynamics are learned from exploration of the system environment. Specifically investigate learning the dynamics of systems such as *MountainCar* and *PoleCart* through Gaussian Process (GP) regression models. We then use the dynamics models to learn expected reward functions represented through a high-dimensional function train, which in turn allows planning of the optimal control trajectory. We demonstrate the application of the algorithms by applying our results to control problems and available environments in OpenAI gym.

Deep Learning for Forecasting Quadcopter Dynamics

Akshay R. Vegesna

Mentor: Anima Anandkumar

Forecasting quadcopter dynamics is difficult because such aerial vehicles exhibit high sensitivity and long temporal dependencies in flight. We treat forecasting dynamics here as a regression problem in high dimensions, and use sequential deep learning models for this task. Specifically, we use Tensor-Train Recurrent Neural Networks (TT-RNN) since they have shown success in long-term forecasting of physical systems. From preliminary forecasting results, the TTRNN shows significant forecasting improvement over the untensorized Recurrent Neural Network

baseline over a range of simulated environments. Furthermore since quadcopter flight is a physical dynamical system with constraints, we study the use of modified Neural Network architectures for this specific task. Two ways that this is done is by enforcing physical constraints as a Lagrangian loss term and forecasting the differences in dynamics instead of the dynamics themselves. We aim to apply this improvement of trajectory forecasts to aid quadcopter controllers.

Applications of Virtual Reality in Data Analysis

Su Min Lee

Mentor: George Djorgovski

With the advancement of VR/AR technology, data visualization beyond the traditional 2-dimensional projects is a realistic possibility. Based on many studies, there is a possibility that multidimensional data visualization will make data analysis more efficient and accurate by using virtual reality as a data analytics tool. My project aims to objectively evaluate the advantages of data visualization in VR over traditional data analytics tools such as Excel, Matlab, R, and Python. I have been generating data sets with different properties, such as dimensions, clusters, correlations, outliers, etc. Then, I was internally testing the data sets using both a VR/AR data visualization package and the traditional tools. Eventually, the internal testing will expand to a greater sample of people and the results will be analyzed to assess in which categories and by how much the VR/AR data visualization package is more effective and efficient at analyzing multidimensional data sets. The ability to visualize data in virtual reality may be the next technical revolution that reveal new correlations and meaningful outliers across all fields of science.

Data Analysis in Virtual Reality

Sarah Guenov

Mentor: S. George Djorgovski

Virtual Reality enables useful visualizations of data sets that are extremely challenging if not impossible to visualize using standard visualization tools such as Excel or Python. In particular, Virtual Reality based visualization excels at allowing users to identify the number of clusters present in a data set, a step that is essential to many clustering algorithms. To identify the types of data sets that benefit most from VR visualization, I crafted sets of varying complexity, with differing numbers of clusters and different levels of overlap and discovered that closely grouped clusters separated by angles rather than distance are extremely challenging to understand in 2D but very intuitive in VR. The next phase of this project involves recruiting subjects to analyze a given data set assigned to use either VR or Excel for visualization and comparing the results so as to quantitatively compare results for each tool.

Designing Amplifiable Molecular Gene Expression Classifiers for Disease Diagnostics

Gokul S. Gowri

Mentors: Georg Seelig, Lulu Qian, and Randolph Lopez

For many diseases, diagnosis can be performed through *in silico* classification of gene expression data using support vector machines (SVMs). Although many gene expression data sets have been curated to facilitate the training of such SVMs, this diagnostic approach is currently impractical for widespread clinical use due to the cost and complexity of gene expression profiling. Recently, DNA-based computers have been created to classify gene expression on a molecular level without the need for direct gene expression profiling. However, these molecular SVMs require concentrations of RNA higher than those found in biological samples. In the existing scheme, amplification is not viable, as amplification bias significantly distorts computation. Here, we propose and theoretically explore an alternative method of encoding SVM weights that would allow for a calibration process, after which the classifier can be adjusted to counteract amplification bias. We show that this method can theoretically be used to implement arbitrary SVMs regardless of amplification biases, while using fewer oligonucleotides than the existing scheme. Additionally, we developed a web application that largely automates the design and calibration of amplifiable molecular classifiers.

Understanding How Artists Sketch Based on Stroke Semantics

Nicole Feng

Mentors: Julie Dorsey, Sherry Qiu, and Mathieu Desbrun

Current computer-aided design and sketching systems suffer from lack of true sketch understanding, while current nonphotorealistic rendering (NPR) techniques depend solely on geometric and image-space properties of a particular object. In reality, how a person draws is the product of many mental processes that are not understood well at all. The goal of this work is to understand *stroke semantics* and *artist intent* by using real artist drawings, collecting them in a database that preserves information about how they were drawn. Using an online tracing interface, we collect artist sketches of a variety of photographs, paintings, and images rendered from 3D models. Preliminary analysis of the order, speed, direction, pressure, inter-stroke relationships, and underlying geometry of strokes provide insight into the human process of drawing. This work represents, to the best of our knowledge, the

first rigorous approach to understanding sketch semantics inspired by differential geometry, perceptual psychology, and cognition, and the construction of a sketch database which captures the relevant statistics for reuse in a broad range of future research. Semantic understanding of sketches has applications in NPR; sketch analysis, understanding, and recognition; and sketch-to-3D translation.

Sample Complexity in Human-Robot Interaction

Rohan C. Choudhury

Mentors: Anca Dragan and Yisong Yue

As robots become increasingly capable of complex tasks, training them to collaborate and interact effectively with humans grows more and more important, especially in domains such as self-driving cars. Examples of such collaborative behavior include complicated lane changes, or signaling intent to other drivers at a busy intersection. In this work, we determine the sample complexity of model-based and model free approaches for training autonomous vehicle agents to interact with humans in a range of scenarios. The model-free approaches we consider are constrained policy gradient methods, such as Trust Region Policy Optimization (TRPO) and Proximal Policy Optimization (PPO). The model-based methods we use are based on predicting the human action either through a neural network or through Continuous Inverse Optimal Control (CIOC) and solving the resulting nested optimization. We find that model-based approaches, while less flexible and more sample dependent, have much lower sample complexity than model-free methods. On the other hand, while model-free deep reinforcement learning methods work across a wider range of scenarios, the sample complexity is significantly higher than with model-based planning approaches.

On the Optimization of Protein Folding in the HP Model

Jacob G. Ressler-Criag

Mentors: Pablo A. Moscato and David Van Valen

Proteins are large macromolecules consisting of one or more chains of amino acids. During the formation of a protein, a DNA strand is transcribed into an amino acid chain which then folds into a minimal energy three-dimensional complex. The shape or "fold" of the protein strongly influences both its interactions with other molecules and its function. However, while protein folding occurs rapidly in nature, its computational prediction from amino acid sequence proves NP hard and extremely energy intensive. In our research, we consider the protein folding problem under the Dill HP model in which we search for the maximum possible number of hydrophobic interactions for the protein on a lattice. In this study, we construct several novel heuristics, involving both ad hoc and naive methods as well as a directional folding constraint program for the HP model across all lattices. In addition, we introduce a set-satisfaction optimization program to solve the HP model on the two-dimensional lattice. Our results show significant improvements over the baseline across all studies, and our introduction of two new mathematical programs for this problem anticipates future research areas.

Multi-Variate Rational Regression With Memetic Algorithm

Haoyuan Sun

Mentors: Pablo Moscato and George Djorgovski

This research tries to address the difficulties of symbolic regression, the process of modeling data with algebraic expressions, in handling division. We propose dealing division with a continued fraction representation, from which rational function models can be extracted. A memetic algorithm, which is a paradigm of metaheuristic optimization based on the evolution of solutions by a set of computational agents that have problem domain knowledge, is implemented to generate solutions in this representation. A population of feasible solutions uses local search heuristics to improve model fitting and compete searching for function models that use a smaller number of variables. We utilize a tree-based population structure to improve the program's efficiency. Data sets from real world applications, such as the classification of flower iris, are used to measure the potential our approach and benchmark its performance against an established commercial software.