

Multi-subject neural decoding from intracranial recordings

Joseph H. Zhou

Mentors: Yanan Sui and Glen A. George

Intracranial stereoelectroencephalography (sEEG) neural recording provides high-resolution and important subcortical data, but limited availability of high-quality sEEG data and variations of electrode placement between subjects during experimental data collection make neural decoding models difficult to train with sEEG data. So, we constructed a deep learning model to enable pooling data from multiple subjects to improve sentence onset decoding for a single subject from the Brain Treebank dataset. This way, data can scale and generalize across subjects and improve decoding performance compared to a single-subject within-session model. We first preprocessed data and selected electrodes significant to predicting sentence onset for each subject. Then we used a temporal-spatial transformer architecture called seegnsignificant to classify neural data windows into sentence onset or no-speech events and compared the model to a single-subject-trained model.

Biomechanically driven forward neural encoding: Synthetic EEG generation

Joan Li

Mentors: Yanan Sui and Glen A. George

Electroencephalography (EEG)-based motor execution decoding holds immense promise for closed-loop assistive devices and physical neurorehabilitation. However, training high-capacity decoding architectures remains fundamentally bottlenecked by the high cost, time intensity, and physiological variability of collecting large-scale human EEG datasets. In this work, we propose a physics-grounded data augmentation framework that utilizes biomechanical simulation for neural encoding by using the MS-Human-700 model—a high-fidelity whole-body musculoskeletal simulation of the human body comprising 700 muscle-tendon units. We extract high-dimensional, continuous muscle activation forces and kinematic trajectories during simulated motor execution tasks. These physiological excitation signals are processed through a synthetic forward neural encoder to generate multi-channel, scalp-projected synthetic EEG time-series. By mapping deep musculoskeletal dynamics directly to cortical representations, our pipeline provides a physics-consistent synthetic data generator for neural decoding models. This offers a scalable solution to neural data scarcity in assistive robotics.

Utilizing full memory: An exploration of techniques in catalytic space

Tobjorn L. Nelson

Mentor: Chris M. Umans

In 2014, Buhrman et al. introduced the idea of catalytic logspace. This model differed from traditional logspace by adding an extra polysize work tape initialized with random bits that must be preserved at the end of the computation. In their seminal work, they proved catalytic logspace was stronger than traditional logspace, being able to simulate boolean circuit classes that are not known to be simulatable in logspace. Since then more striking structural results about catalytic logspace have been discovered. These include Koucký et al.'s result that catalytic logspace is equivalent to catalytic nondeterministic logspace, an open problem for its noncatalytic variant, and Alekseev et al.'s result that linear matroid intersection, a problem not believed to be contained in the circuit classes currently simulatable in catalytic logspace. We refine existing proofs of containments in catalytic logspace and work towards proving additional problems are contained in catalytic logspace.

Sparse predict and demote: Optimal scheduling with minimal size advice

Chengyi Liu

Mentors: Adam C. Wierman and Tinashe Handina

Knowledge of job sizes is crucial for optimizing the mean response time in single-server scheduling. However, predictions of job sizes are typically inexact and may be costly to obtain. Thus, a long-standing research question is: what is the minimal information about job sizes required to optimize the mean response time? In this work, we introduce a novel policy, Sparse Predict and Demote (SP&D), and prove that it achieves order-optimal mean response time in heavy traffic while using only one-bit predictions about the size of a vanishing fraction of arriving jobs. SP&D achieves order-optimal mean response time even in situations (i) when predictions are inexact up to any fixed multiplicative error bound, and (ii) when predictions are costly, i.e., the server is used to compute predictions and is thus unavailable to serve jobs while processing prediction queries. Finally, we use numerical experiments to illustrate that the SP&D maintains near-optimal performance beyond the heavy-traffic regime.

FROST: Stochastic prediction-space adaptation for sim-to-real PDE forecasting

Zichang W. Wang

Mentor: Yisong Yue

Simulation pretraining followed by real-data fine-tuning is a natural strategy for adapting learned PDE forecasters, such as neural operators, to physical experiments. Yet deterministic adaptation can leave a structured mismatch between forecasts and experimental futures, whose variability reflects unresolved dynamics and measurement noise. To this end, we introduce FROST (Frozen-Operator Stochastic Transport), which freezes an upstream forecaster during correction and learns the conditional distribution of its residual from real measurements. Conditioned on observed history and the base forecast, FROST corrects systematic bias while modeling structured variability around the forecast. Across five RealPDEBench systems, FROST improves long-horizon ensemble-mean accuracy and fair CRPS over the corresponding fine-tuned FNO on every system, with reductions of up to 61% and 75%, respectively, and matches or outperforms every deterministic and generative baseline in mean rollout accuracy on all five systems. Additionally, we find that when applying FROST, the base FNO need not be fine-tuned at all. Together, these results establish stochastic prediction-space correction as a powerful complementary axis of sim-to-real adaptation for PDE forecasting.

Development of a deterministic adaptive mathematics platform for students transitioning from grade 9 (JHS 3) to grade 10 (SHS 1) in Ghana and Kenya

Ibrahim Adam

Mentors: Franca Hoffmann and Jasmine R. Bryant

This research project studies how an adaptive online mathematics tutor platform can provide unique practice for students transitioning from Grade 9 to Grade 10 in Ghana and Kenya. I have already developed a prototype that organizes mathematical concepts through prerequisite connections, chooses suitable practice questions based on student competency and progress, assesses responses using an online deterministic mathematics assessment system, and uses the resulting evidence to decide subsequent practice questions or advancement for the student. Currently, the system combines verified mathematics questions which covers topics including indices, integer operations, coordinate graphs, ratios and proportions, introductory trigonometry, and similarity. The system supports automatic question rendering, grading and interactive graph-based activities. Some testing that I have done both with the backend on VSCode and the frontend on the live prototype platform has demonstrated successful rendering and grading of both standard and interactive questions, while also revealing compatibility problems in several older question test cases as well as connecting multi-input interactive activities to the adaptive task logic. Currently, I focused my work on resolving these integration problems, adding more questions to the platform, and making the adaptive progression framework stronger before evaluating learner interaction data.

Question and quiz analysis of open STACK questions: Validating diagnostics with real course data

Ernest W. Ting

Mentors: Franca Hoffmann, Jasmine R. Bryant, and Juma Zevick Otieno

STACK, a computer-algebra-based online assessment system deployed within Moodle, generates detailed per-student, per-question, and per-response-tree data on every quiz attempt. This data is rarely turned into actionable insight because existing statistical methods, such as facility and discrimination indices and item response theory, normally require exporting data and analyzing it outside the classroom platform. This project built tools that surface these statistics automatically inside Moodle. First, a plugin reports per-question and per-branch diagnostics directly within a course, and a second uses Moodle's Analytics API to flag struggling students and problematic questions with models trained on indicators such as syntax-error rate and branch usage. These diagnostics were validated using three years of real quiz-response data from a live course. The analysis compared the platform's facility index and a simplified feedback-effect measure against a full item response theory fit and paired significance testing. Additionally, it tracked how diagnostics for the same question shifted across different student cohorts and following instructor revisions. The results showed that the lightweight indicators closely mirrored their statistically fuller counterparts. Furthermore, evidence from revision impacts and instructor interviews confirmed that flagged questions aligned with issues teachers had independently identified. Ultimately, this demonstrated that in-platform diagnostics can reliably guide instructors toward the specific questions and learning gaps most worth addressing.

A manifold constrained Ensemble Kalman Filter algorithm for nonlinear dynamics

Ethan M. Feng

Mentors: Andrew M. Stuart, Bohan Chen, and Sangmin Park

The Ensemble Kalman Filter (EnKF) approximates the filtering distribution with a Gaussian ensemble updated via linear transport. While exact in the linear Gaussian mean-field limit, this approach breaks down for nonlinear dynamics. The ambient affine analysis map transports ensemble members away from the system's intrinsic attractor, causing systematic bias and an inflated posterior covariance.

We propose the Argus Ensemble Filter (AEF), a manifold-constrained variant that restricts the analysis update to evolve along the lower dimensional geometry of the forecast ensemble as opposed to the full ambient space. After a standard EnKF analysis, each member is projected onto a local polynomial chart learned from the forecast dynamics. We examine differing approaches to the local chart construction including a global chart and an adaptive moving least squares, additionally we introduce different projection variants that determine how the point cloud is assembled.

Evaluated on the Lorenz-63 system against a one million particle bootstrap particle filter, AEF significantly reduces the sliced energy distance to the true filtering distribution compared to EnKF. The manifold constraint corrects the systematic bias of the ambient analysis update by preserving inherent local curvature structure. These results suggest embedding EnKF updates within the intrinsic geometry of the forecast leads to a principled route to robust filtering in strongly nonlinear regimes.

Using conflict graphs to determine the best models against adversarial noise

Amudhan S. Gurumoorthy

Mentors: Daniel Cullina and Konstantin Zuev

Deep learning models are very powerful; they can take objects and words and construct strong relations and classifications, which is very useful for identifying images, recognizing speech, and other scenarios which require greater thought. However, one area where these models often struggle is dealing with adversarial data, noise that is slightly and carefully perturbed to fool a classifier. The level of perturbation that is allowed is known as an adversarial budget, and if classifiers cannot deal with these budgets in a proper manner, then they can be exploited, threatening the stability of models in places like fraud detection and medical imaging. In this project, we analyze many different adversarial budgets at the same time from a theoretical lens using matchings and graph theory, leading to greater robustness on larger samples while keeping smaller budgets to maintain accuracy near a sample

point. By representing different samples and different budgets per sample as vertices, we obtain a bipartite conflict graph with special properties that we can exploit. We use extensions and new approaches to algorithms including the Hopcroft-Karp and Hungarian algorithms to calculate and approximate the most number of matchings in this bipartite graph in the least amount of time. This would allow us to determine which adversarial budgets are the best for the classifier to focus on, providing a near-optimal solution to fitting the classifier to be adversarially robust and accurate.