2023 Summer Undergraduate Research Symposium

August 3, 2023
9:30am-12:00pm
303 Uris Hall

- Aditya Jain, Yun-Tzu Lin, Valerie Rogel (CSUREMM Group 1)

  From Climate Talk to Stock Shock: Modeling Policy Headline Sentiment and Index Fund Performance

  Abstract: This research examines and attempts to model the relationship between climate policy news sentiment and financial metrics for major index funds. We apply the Valence Aware Dictionary and Sentiment Reasoner (VADER) to generate a sentiment score for over 14,000 relevant headlines between January 2014 and June 2023. Employing Granger Causality tests, we find statistically significant evidence that climate policy news sentiment influences certain funds' volatilities and prices over 10-year, 1-year, and quarterly timeframes. To attempt to forecast fund metrics, we turn to standard, univariate autoregressive models (including Autoregressive Integrated Moving Average) and multivariate machine learning (ML) regression models. Our findings indicate that ML models are generally poor approaches to forecasting fund metrics in longer timeframes, with high levels of error in comparison with autoregressive approaches that do not take in climate policy news sentiment. However, the verdict is more mixed in shorter time frames, with one model – the Extreme Gradient Boosted Regressor – generating moderately to strongly accurate predictions in quarterly timeframes.

- Shiruo Liang, Joe Wee Sham, Yiming Song (CSUREMM Group 2)

  Sector-specific forecasting of volatility from Twitter sentiment

  Abstract: Volatility is a highly useful metric that informs investors about risk and potential returns. In this paper, we examine the usefulness of Twitter sentiment data when applied to the prediction of volatility across 11 different market sectors. We first compute the Pearson correlation and test for Granger causality and find statistically significant relationships between market sentiment and volatility for certain sectors, notably the technology and consumer discretionary sectors. We then apply these findings to four different machine learning models. We compare cross-model performance and find that a long short-term memory (LSTM) model outperforms all other models (linear regression, decision tree, XGBoost) when predicting future volatility, with an order-of-magnitude lower mean absolute percentage error, both with and without sentiment data input. The LSTM model is also the only model that has a statistically significant improvement upon inclusion of sentiment data, suggesting that the true underlying relationship between sentiment and future volatility is highly complex and difficult to capture with conventional machine learning models that rely on regression alone.

- Maksym Bondarenko, Justin Haddad, Phebe Lew (CSUREMM Group 3)

  INSIGHT: Intelligent Negotiation Strategy for Information Games with Hidden Tactics

  Abstract: INSIGHT (Intelligent Negotiation Strategy for Information Games with Hidden Tactics) is the culmination of our research in the use of reinforcement learning and game theory to create optimal bidding strategies in environments with hidden information. The broader goal of this research was to devise strategies applicable to a wide variety of real-world situations, such as trade negotiations, financial

markets, and other bidding games. We created a discrete Markov model to simulate bidding episodes in which multiple sellers compete to maximize profit for each buyer. INSIGHT combines different RL and heuristic strategies from existing research to learn over time the most profitable strategies in any given situation. Then, we investigated how fine-tuning of parameters and reward functions within the INSIGHT model affects its performance in different scenarios, competing against opponents with a fixed stochastic strategy, as well as opposing agents with reinforcement learning strategies. Our results show that using this combination of strategies, we can create sellers to outperform any given static opponent, but that two competing Q-Learning sellers which both attempt to adapt to each other fail to cooperate, thus resulting in low pareto efficiency. This mirrors a situation in prisoner's dilemma where both agents play based on the myopic strategy of choosing only to defect, maximizing short term payoffs which eventually lead to a far worse outcome for both agents involved. The INSIGHT model shows the powers and limitations of reinforcement learning for these negotiation situations, and leaves room for a breadth of future research involving other RL algorithms that could learn to cooperate.

- Judah Engel, Milena Harned, Maahi Patel (CSUREMM Group 4)

  A Noise Level Aware Poisson Denoising Algorithm for Fluorescence Microscopy Data Using SwinConv U-Net

  Abstract: Fluorescence microscopy is a powerful imaging technique widely used in biological research to visualize cellular structures and processes. However, the acquisition process is prone to introducing Poisson noise, which can degrade image quality and hinder accurate analysis. Poisson noise, also called shot noise, tends to affect images captured in low light settings and affects the image quality in a signal dependent manner. Many algorithms currently exist to denoise images corrupted by additive Gaussian noise, but fewer address this less common yet equally damaging noise type. To address this challenge, we propose a novel noise level aware Poisson denoising algorithm that leverages the Swin Conv U-Net architecture. The transformer was first applied to computer vision tasks in 2021, and has shown competitive results in a wide range of applications through its multi-headed attention mechanism, allowing it to pick up on long range dependencies. The U-Net's encoder-decoder structure with skip connections allow the model to effectively combine low-level features with high-level context, allowing for effective denoising while preserving important image details. Our algorithm integrates an estimation of the noise parameter with the strengths of Swin Transformers and U-Net to effectively denoise fluorescence microscopy data.

- Yvon Lu, Jordan Shiff, Sophia Tu (CSUREMM Group 5)

  Refining gender bias algorithms using sentiment analysis

  Abstract: Despite efforts to promote diversity and inclusion, the underrepresentation of women in male-dominated fields such as engineering and finance persists, especially in leadership. The hiring process can play an important role in the recruitment and retention of women at companies. One aspect that has gained attention in recent years is the potential influence of the language used in job descriptions on gender imbalances in applicant pools. As new techniques for analyzing language emerge, we sought to counteract existing biases and build more complex models to guide recruiters. Improving on existing models for gender bias checks (Gaucher et al., 2011), we added the inclusion of tone via sentiment analysis to produce a more nuanced understanding of potentially discriminatory language. Using job descriptions scraped from popular job board sites (Indeed, Monster, Careerbuilder), we obtained over 60,000 data points ranging from a wide variety of industries and

career levels. Additionally, we built an RNN model to predict these scores with high accuracy. Post-analysis results on our collected data indicated a statistical significant difference of bias amongst industries, as well as job qualifications. We observed a high negative bias (male leaning) in job descriptions specifying higher educational requirements. Furthermore, we collected job descriptions from the Forbe's 2022 World's Top Female Friendly Companies and ran additional PSM testing to determine if a difference in bias was significant. Our refinement of existing algorithms will hopefully introduce new ways to combat discriminatory language that go beyond the base measurement of gender-coded words.

- Jingtao Li, Jingwen Zhang, Yihan Shen (CSUREMM Group 6)

  Modeling The Dual-Sourcing Inventory Problem – A Reinforcement Learning Approach

  Abstract: Our research investigates the dual sourcing problem to inform better supply chain management. We model the context as an Markov Decision Process and study how well an agent making sequential decisions can maximize long term utilities under two different behavioral policies —— the deterministic, documented TBS approach; and our adaptation of the randomized A2C algorithm. We discuss their tradeoffs and analyze the impact of the parameter of the demand distribution on the performance gap between the two algorithms. Our experiments demonstrate that A2C has a better guarantee of the lower bound on the inventory level than TBS, and when demand distribution is less frequent A2C is able to match up with TBS in performance as well. This study showcases the potential of Reinforcement Learning based algorithms like A2C as a powerful tool for sequential decision-making in stochastic environments such as the dual sourcing problem.

- Zachary Canale, Zoe Curewitz, Shiv Yajnik, Carlos Santiago Calderon, Jacob Daum (CMUSR 1)

  Points of low degree on plane curves

  Abstract: The goal of this project is to study low-degree points on plane curves over a number field. As an application of Falting's theorem, Debarre-Klassen described the geometry of low degree linear series on smooth plane curves. By utilizing further results of Coppens-Kato, we extend some of the results of Debarre-Klassen to the setting of integral curves with mild singularities.

- Zhaocheng Dong, Gabriel Fernandez, Katherine Mekechuk, Xiaohua Wei (CMUSR 2)

  Transcendence Theorems for Arithmetic Gevrey Series

  Abstract: The theory of arithmetic Gevrey series was introduced by Andr'e around 2000, of which the E-functions (e.g., exponential, Bessel function) are a specific type. The transcendence results of E-functions have been worked out in detail, such as the Nesterenko-Shidlovskii theorem on algebraic independence of values of E-functions. We generalize this and other theorems on E-functions to arithmetic Gevrey series of negative integer order.

- Maria Stuebner (Independent Project, Francesco Lin)

  The quotient dimension of hyperbolic three-manifolds

  Abstract: For any finitely generated group Γ, the quotient dimension of Γ refers to the minimum dimension of a complex linear group containing an infinite quotient of Γ. In his work, "How often is $84(g-1)$ achieved?", Michael Larsen shows that the quotient dimension determines the asymptotics of the set of orders of finite quotient groups of Γ; furthermore, he computes the quotient dimension of various groups of hyperbolic surfaces and orbifolds. In this project, we explore further conditions for determining the quotient dimension of finitely generated groups, specifically focusing on the fundamental groups of hyperbolic three-manifolds.