Xiao Han

Email : hanxiao2099@gmail.com

Personal Page | LinkedIn | Github | Google Scholar

Mobile : +1-541-908-8790

Education

- **Utah State University** — Logan, UT
  Ph.D. in Computer Science; GPA: 3.91 — *Aug. 2020 – May 2024*
  Outstanding Graduate Student; Presidential Doctoral Research Fellowship $40,000
- **George Washington University** — Washington, DC
  M.S. in Data Analytics; GPA: 3.76 — *Aug. 2018 – May 2020*
- **Oregon State University** — Corvallis, OR
  M.Eng. in Computer Science; GPA: 3.63 — *Sep. 2014 – Dec. 2016*

Selected Publications

1. **Xiao Han**, Shuhan Yuan, and Mohamed Trabelsi. LogGPT: Log Anomaly Detection via GPT. In 2023 IEEE International Conference on Big Data (**Big Data**). 2023.
2. **Xiao Han**, Lu Zhang, Yongkai Wu, and Shuhan Yuan. On Root Cause Localization and Anomaly Mitigation through Causal Inference. In Proceedings of the 32nd ACM International Conference on Information & Knowledge Management. (**CIKM**). 2023.
3. **Xiao Han**, Lu Zhang, Yongkai Wu, and Shuhan Yuan. Achieving Counterfactual Fairness for Anomaly Detection. In Pacific-Asia Conference on Knowledge Discovery and Data Mining. (**PAKDD**). 2023.
4. **Xiao Han**, Depeng Xu, Shuhan Yuan, and Xintao Wu. Few-shot Anomaly Detection and Classification Through Reinforced Data Selection. In 2022 IEEE International Conference on Data Mining (**ICDM**). 2022.
5. **Xiao Han** and Shuhan Yuan. Unsupervised cross-system log anomaly detection via domain adaptation. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management. (**CIKM**). 2021.

Experience

- **Nokia Bell Labs** — Murray Hill, NJ
  Machine Learning and AI Intern — *Jun. 2023 - Aug. 2023*
  - Conducted a patent application as part of the research team.
  - Performed in-depth research on anomaly detection for log data, utilizing large language models (LLM) and leveraging reinforcement learning techniques, such as Proximal Policy Optimization (PPO) and Advantage Actor-Critic (A2C), to enhance the F1-score across multiple datasets.
- **Utah State University** — Logan, UT
  Research Assistant — *Aug. 2020 - May 2024*
  - **Trustworthy Anomaly Detection**: Developed anomaly detection models for various data types, including tabular, sequential, and time series data, spanning domains such as performance, explainability, and fairness.
  - **Causal Modeling and Inference**: Created open-source causal inference models and applied them to various downstream tasks, including mitigating bias for specific demographic groups to promote fairness and offering recommendations to address undesired outcomes of AI models.

Projects

- **Root Cause Analysis for Time Series** | PyTorch, Causal Discovery — *Aug. 2023 - Feb. 2024*
  Led the creation of the AERCA framework for root cause analysis across diverse systems. Utilized Granger causality and an innovative encoder-decoder architecture to efficiently detect anomalies and analyze their root causes, achieving a 20% improvement in accurately identifying root causes across multiple domains.
- **Language Model-driven Anomaly Detection** | GPT, RAG, PPO, A2C, Model Fusion — *Apr. 2023 - Jul. 2023*
  Developed a cutting-edge Python framework that harnessed the power of Large Language Models (LLMs), such as GPT and T5, and employed reinforcement learning with human feedback (RLHF) to optimize anomaly detection. This approach led to significant enhancements, delivering an impressive 10% improvement in F1 scores on average.
- **Fairness-aware System: CFAD** | Git, Conda, Matplotlib, Scikit-learn — *Aug. 2021 - Dec. 2021*
  Implemented an open-source fairness-aware system using PyTorch, integrating Graph Convolutional Networks to identify causal relationships. Applied adversarial training to reduce biases, achieving a 5.2% improvement in fairness metrics with minimal impact on anomaly detection.
- **Recommendation System** | MongoDB, OpenCV, Pillow (PIL), PCA, NumPy, Pandas — *Aug. 2019 - Dec. 2019*
  Collected data by scraping information from IMDB and YouTube, established a MongoDB setup to store the gathered data, and applied Principal Component Analysis (PCA) to movie synopsis and genres.

Technical Skills

- **Languages**: Python, SQL, C, C++, Haskell, Java, Javascript, PHP, HTML, CSS, R, Idris
- **Tools & Platforms**: Pytorch, Pandas, TensorFlow, Git, Docker, Linux, Unix, MySQL, SQLite, MongoDB, ArangoDB, AWS, Google Cloud Platform, Azure, Databricks, Minitab, Tableau