

Enhao Liu

Research Interests

My current research develops the machine learning models including rigorously optimal classification tree applied to cybersecurity. Specifically, (1) the cost-effective cybersecurity maintenance policy development based on (Partially Observable) Markov decision processes; (2) cyber vulnerability risk prediction with social media indications based on supervised machine learning models. **I am especially interested in a challenging position that uses my machine learning, operations research, statistics, and coding skills.**

Methodology: (1) Machine Learning: Optimal Classification Tree, Ensemble Learning, Hyperparameters Optimization; (2) Reinforcement Learning: Partially Observable Markov Decision Process (POMDP), Monte Carlo-Bayesian Reinforcement Learning (MC-BRL), Deep Q-Learning; (3) Time Series Modeling: Autoregressive Distributed Lag Models, Control Charts

Applications: Cyber vulnerabilities maintenance policies management, software vulnerabilities prioritization, anomaly detection in the oil & gas pipelines, COVID-19 comprehensive surveillance

Education

- Aug 2017–May 2021 **Ph.D. Candidate**, *The Ohio State University*, Columbus, OH, USA.
Major: Operations Research; Minor: Machine Learning, Cybersecurity; Advisor: Theodore Allen.
- Aug 2015–May 2017 **Master of Science**, *The Ohio State University*, Columbus, OH, USA.
Major: Operations Research; Thesis Advisor: Theodore Allen.
- Aug 2011–May 2015 **Bachelor of Engineering**, *Jinan University*, Guangzhou, China.
Major: Electric Engineering and Automation. Senior thesis advised by Weihua Li.

Skills

Programming	Python, R, SQL, Matlab, GAMS
Main Packages	Python: pyomo, Gurobi API, Cplex API, scikit-learn, imbalanced-learn, xgboost, hyperopt, numpy, pandas, matplotlib, plotly, Flask, sqlalchemy; R: caret, DMwR, MDPtoolbox, pomdp, ompr, CVXR, dLagM, rsatscan, tidyverse, leaflet, visNetwork, odbc, shiny, shinyjs.
Software	Minitab, ARENA, Simio
Others	LaTeX, Microsoft Office Suite
Coursework	Operation Research Models and Methods; Linear Programming; Non-linear Programming; Mixed Integer Programming; Convex Optimization Algorithm; Stochastic Optimization; Stochastic Process; Decomposition Techniques in Mathematical Programming; Statistical Machine Learning; Mathematical Analysis of Algorithms; Intermediate Data Analysis I & II.
Languages	Chinese (native), English (fluent)

Awards

- 2019 Runner-up, Student Paper Competition, The Social Media Analytics Section of INFORMS.
- 2014 Second Prize of Scholarship for Excellent Students, Jinan University.
- 2013 Third Prize of Scholarship for Excellent Students, Jinan University.
- 2013 Yihai Kerry Scholarship for Innovative Undergraduates, Jinan University.
- 2013 Meritorious Winner, National College Mathematical Contest in Modeling.
- 2012 First Prize, China Undergraduate Mathematical Contest in Modeling.

Employment

- May 2020–present **Graduate Research Assistant, The Ohio State University, Columbus, OH, USA.**
Department: College of Public Health; Supervisor: Elisabeth Root, Ayaz Hyder
- Led the development & deployment (including data architecture, software developments, translating requirements, infrastructure design, and implementation) of COVID surveillance web tools for the Ohio Department of Health and school districts in the State of Ohio.
- Aug 2016–present **Graduate Research Assistant, The Ohio State University, Columbus, OH, USA.**
Department: Integrated Systems Engineering; Lab: Security and Efficiency Through Analytics Laboratory; Supervisor: Theodore T. Allen
- Led the project “Advanced AI in Pipeline Anomaly Classification” funded by ROSEN Group starting from Sep 2020.
 - Led the project “A Framework For Economical Cyber Security Inspection and Assurance” funded by National Science Foundation (NSF Grant #1912166) from May 2019 to Mar 2020.
 - Led and conducted the project “Data-Driven Cyber Vulnerability Maintenance” funded by National Science Foundation (NSF Grant #1409214) from Aug 2016 to Nov 2018.
 - Developed the cyber vulnerability risk prediction model with social media indications based on machine learning models. Worked with Pardee Bob and Geoffery Shoupp from Office of the Chief Information Officer (OCIO) Enterprise Security at the Ohio State University to deploy the proposed system as an API product.
 - Reviewed several papers from *Computer & Industrial Engineering*.
 - Worked with Professor Theodore T. Allen, a Ph.D. candidate Sayak Roychowdhury and a graduate student Akshay K. Murali to publish two journal articles and two conference papers.
 - Worked with Professor Theodore T. Allen to write a book about machine learning applications in operations optimization and supply chain optimization.
 - Worked with Professor Theodore T. Allen to design a new syllabus and materials for a course “Introduction to Operations Analytics” which was offered in Spring 2020. The new version of this course is aimed to help undergraduate and graduate students learn how to implement industrial applications with machine learning methods using Python.

Recent Research Experience

- Sep 2020 – present **Advanced AI in Pipeline Anomaly Classification, Funded by ROSEN Group.**
Team Leader: Enhao Liu; Supervisor: Dr. Theodore Allen.
- Leveraged multiple image processing methods including wavelet transformation, saliency detection, contrast calculation, and blob detection methods to generate over 20 new features for the scanning data of oil & gas pipelines.
 - Achieved over 95% precision & recall for unseen testing data in the proposed machine learning model with hyperparameters optimization.
- Apr 2020 – present **COVID-19 Comprehensive Monitoring Team at OSU.**
Lead developer: Enhao Liu; Supervisor: Dr. Elisabeth Root, Dr. Ayaz Hyder.
- Led the full infrastructure of the R Shiny-based dashboard (over 40,000 code lines) of the COVID-19 surveillance for the Ohio Department of Health, including data architecture, software developments, translating requirements, infrastructure design, and implementation. Public-facing site is available: [link](#).
 - Developed a school-level COVID-19 surveillance tool called COVID-19 Analytics and Targeted Surveillance System (CATS) for over 20 school districts to facilitate superintendents to make learning mode schedules for over 300 schools, 180,000 students, and 20,000 staffs. One of the public-facing dashboards is available on the website of the Hilliard City School District: [link](#)
- May 2019–present **Explainable AI Study: Optimal Classification, NSF Grant #1912166.**
Team Leader: Enhao Liu; Supervisor: Dr. Theodore Allen.
- Developed the Python & R package for the optimal classification tree (OCT) algorithm based on mixed-integer linear programming, which supports Gurobi, CPLEX, and GLPK solvers.

- Applied OCT to the defects classification in pipelines and showed that OCT is comparable to the ensemble learning tree models (such as XGBoost) in terms of model accuracy (96% precision & recall).

Jul 2019 – Mar 2020 **Twitter-based Detection for Exploited Vulnerabilities**, NSF Grant #1912166.

Team Leader: Enhao Liu; Supervisor: Dr. Theodore Allen.

- Wrote scraping scripts using Python to collect recent five-year Twitter data (millions of tweets) and other online resources about 70,000 distinct software vulnerabilities spanned over 9,000 different vendors.
- Developed feature generation and classification models to early detect exploited vulnerabilities based on XGBoost using scikit-learn framework.

Aug 2016–Nov 2018 **Cyber Vulnerabilities Maintenance Management**, NSF Grant #1409214.

Team Leader: Enhao Liu; Supervisor: Dr. Theodore Allen.

- Created a series of procedures using R to process and statistically analyze 2.3 million scanned data about vulnerabilities from the university, which facilitated the development of data-driven cyber vulnerability maintenance based on MDP. This part generated my Master's thesis.
- Constructed a comprehensive framework based on POMDPs to derive cyber vulnerability maintenance policies that address the incomplete inspection issue on multi-type hosts.
- Worked with Dr. Theodore Allen and a Ph.D. candidate Sayak Roychowdhury to construct a new Monte Carlo-Bayesian Reinforcement Learning (MC-BRL) framework to address parameters uncertainty in estimating transition probabilities and expected costs of MDPs in which multiple identical-systems have been considered as well.
- Collaborated with Dr. Theodore Allen and a graduate student Akshay K. Murali to build simulation models in ARENA to evaluate the cyber maintenance policies under different scenarios in terms of maintenance costs and security hardening.

Publications

Journal Articles

- ASMBI 2019 **Liu, E.**, Allen, T. T., & Roychowdhury, S. (2019). Cyber vulnerability maintenance policies that address the incomplete nature of inspection. *Applied Stochastic Models in Business and Industry*, 35(6), 1390-1410.
- CAIE 2018 Allen, T. T., Roychowdhury, S., & **Liu, E.** (2018). Reward-based Monte Carlo-Bayesian reinforcement learning for cyber preventive maintenance. *Computers & Industrial Engineering*, 126, 578-594.

Conference Papers

- WSC 2019 Murali, A. K., **Liu, E.**, & Allen, T. T. (2019, December). Discrete event simulation of cyber maintenance policies according to nested birth and death processes. In *2019 Winter Simulation Conference (WSC)* (pp. 774-785). IEEE.
- WSC 2018 Allen, T. T., & **Liu, E.** (2018, December). Forecasting cyber maintenance costs with improved scan analytics using simulation. In *2018 Winter Simulation Conference (WSC)* (pp. 1218-1225). IEEE.

Dissertations

- Master Thesis 2017 **Enhao Liu**. "Logistic Regression Model for Predicting Warning "Incident" Rates and Implications for the Common Vulnerability Scoring System." Master Thesis, The Ohio State University, 2017.

Manuscripts in Progress

- Ph.D. Thesis 2021 **Enhao Liu**. "Machine Learning Methods Including Rigorously Optimal Classification Trees Applied to Cybersecurity." Ph.D. Thesis, The Ohio State University, 2021.
- Journal 2021 **Enhao Liu** and Theodore T. Allen. "Twitter-based Cyber Security Warning and Prioritization System." 2021.
- Journal 2021 **Enhao Liu** and Theodore T. Allen. "A Sequential Sampling Procedure for Learning Optimal Classification Trees." 2021.

- Book 2021 Theodore T. Allen and **Enhao Liu**. “Introduction to Machine Learning For Supply Chain Optimization - *Python, Statistics, and AI with Operations Analytics Industrial Datasets*.” *Springer*, 2021.

Speaking

Invited Research Talks

- 2020 “Monitoring COVID-19 in Ohio & at Ohio State”, Infectious Diseases Institute Virtual COVID-19 Symposium, Columbus, OH (Dec 03, 2020).
- 2019 “Twitter-based Cyber Security Warning and Prioritization System”, INFORMS Annual Meeting – Session of Information Operations in Social Networks, Seattle, WA (Oct 23, 2019).

Poster Sessions

- 2018 “Cyber Vulnerability Maintenance and Optimal Learning”, Research Poster Session in the Department of Integrated Systems Engineering, OSU, Columbus, OH (Nov 2018).
- 2016 “Data-Driven Cyber Vulnerability Maintenance”, Research Poster Session in the Department of Integrated Systems Engineering, OSU, Columbus, OH (Sep 2016).