

Xiaoyang Song

New York, NY 10025 | +1 (608)-960-2708 | xs2485@columbia.edu | [Github](#) | [Linkedin](#)

RESEARCH INTERESTS

My goal is to research on the cutting-edge theories and methodologies in the field of computer science, machine learning, statistics, and optimizations and apply them to real-world industries, including transportation, health technology, and conservation, with guarantees in robustness/privacy/fairness and positive social impacts. In addition, I am deeply interested in Federated Learning and Out-of-Distribution (OoD) Learning, where we made ML models broadly applicable by protecting the users' privacy and increasing the robustness of ML models, respectively. Beyond these, with a goal to quantify the reliability of ML models, I am also interested in the development of mathematical/statistical foundation of ML algorithms.

EDUCATION

Columbia University

M.S. in Data Science

New York, NY

Aug 2022 — Aug 2023 est.

University of Michigan, Ann Arbor

B.S. in Computer Science & B.S. in Mathematics, GPA: 4.0/4.0

Ann Arbor, MI

Aug 2020 — May 2022

University of Wisconsin, Madison

Major in Computer Science & Mathematics, GPA: 4.0/4.0

Madison, WI

Aug 2018 — May 2020

RESEARCH EXPERIENCES

Out-of-Distribution (OoD) Learning via Generative Model

Ann Arbor, MI | May 2022 — Present

Research Assistant at University of Michigan Transportation Research Institute (UMTRI)

Supervisors: Dr. Wenbo Sun, Prof. Judy Jin and Prof. Raed Al Kontar from University of Michigan

- Proposed a new Wasserstein-distance-based loss function on GAN to generate/discriminate OoD data given in-distribution (InD) training data and a subsample of OoD data ([poster](#)).
- Implemented the OoD GAN and trained it on MNIST/FashionMNIST and CIFAR-10/SVHN datasets.
- Generated OoD adversarial samples using OoD GAN and trained a Logistic Regression detector on them.
- Our model yielded comparable detection accuracy than the SOTA with extremely limited exposure to OoD data and was more robust to outliers; our work exhibited its generalizability to real-world industries.
- Currently working on statistical theory derivation and targeting at ICML 2023.

Federated Unsupervised Learning on Item Response Theory Model

Ann Arbor, MI | Feb 2021 — Present

Undergraduate Researcher at University of Michigan

Supervisors: Prof. Ziwei Zhu from University of Michigan, Prof. Yuqi Gu from Columbia University

- Reproduced FedSGD/FedAvg algorithms on the MNIST/CIFAR-10 datasets based on CNN architecture in Pytorch.
- Combined Federated Learning algorithms FedAvg, DistributedPCA, and ReBoot with stochastic approximation algorithm Metropolis-Hastings Robbins-Monro (MH-RM) for the MIRT model using R/Python.
- Conducted simulation studies on MIRT, 2PL, and Rasch model and identified the superiority of ReBoot algorithms over other federated/distributed aggregation algorithms in terms of statistical error bounds.
- Resolved the identifiability problem of IRT models by imposing the PLT sparsity constraint during optimization.
- Conducted sensitivity analysis on the IRT latent distribution and identified the higher stability of ReBoot compared to other federated algorithms, especially when the local sample size is not sufficiently large.
- Identified the generalizability of federated learning algorithm ReBoot on IRT models on different data sources.

Deep Canonical Correlation Analysis (DCCA) on Multi-modal Data

Remote | Oct 2022 — Present

Research Assistant at Michigan Engineering, Department of Industrial and Operation Engineering (IOE)

Supervisors: Dr. Wenbo Sun and Prof. Judy Jin from University of Michigan

- Processed CT-scan images through clipping/denoising and locate defects via heuristic contour-finding algorithms.
- Implemented and trained an autoencoder to extract hidden features for processed CT-scan images.

- Reproduced the autoencoder architectures introduced in recent papers ([link](#)) to extract features for spectra data.
- Proposed and implemented a LSTM-based model to process windows of CT-scan images and spectra and an adapted DCCA module that is stacked on top of feature extraction modules.
- Our work assigned high correlations for spectra/images window pairs and made possible the cheaper in situ monitoring using spectra data in 3D printing industry.

Symbolic Music Processing and Generation

Ann Arbor, MI | Jan 2022 — Present

*Undergraduate Researcher at Michigan Engineering, Department of Computer Science (CS)**Supervisor: Prof. Rada Mihalcea at Language and Information Technology (LIT)*

- Proposed new feature representation schemes and tokenization pipelines for music pieces according to music theories: the proposed representation involves information and structure from melody, key, tuplet, and bass.
- Embedded tuplet/key structure explicitly by appending the duration token at the end of the pitch sequence.
- Trained a Transformer-XL and Reformer model to generate symbolic music conditionally (*generated samples*).
- Designed objective evaluation metrics In-Key-Ratio (IKR) to measure the quality of produced samples based on harmonies in music theories; our work exhibited advantages over SOTA on music theory-based objective metrics.

Data-driven Analysis of Marine Sympatric Speciation

Ann Arbor, MI | Sep 2020 — Jun 2021

*UROP Researcher at University of Michigan**Collaborator/Mentor: Dr. James Saulsbury*

- Built a statistical model for measuring living ranges and constructed a database of geographic and phylogenetic data of 500 marine sister species pairs from the *WoRMs* and *OBIS* datasets.
- Designed and implemented a statistical algorithm to calculate the living range overlap of sister species pairs based on the intersection of convex hulls using R and performed rule-based outlier detection on generated convex hulls.
- Conducted regression and correlation analysis and data visualization and identified strong correlations between species age and the likelihood of sympatry; presented at the *UROP Academic Symposium (2021)*.

Statistical Learning Application on Phylogenetics

Madison, WI | Feb 2020 — May 2020

Undergraduate Researcher at University of Wisconsin-Madison

- Explored the Parsimonious Principle of phylogenetic tree construction and its topology and combinatorics.
- Researched on the traditional Maximum-Likelihood methods and neural network models in phylogenetics tree reconstruction and generation from alignments of genetic data.
- Implemented a customized tree Likelihood calculation model with assumption of site independence in Python and conducted small-scale experiments on simulation datasets for testing.

Mathematical Contest of Modeling (MCM)

Madison, WI

*Mathematical Modeling Practitioner at University of Wisconsin-Madison***Louvre Museum Evacuation Route Design**

Jan 2019 — Feb 2019

- Completed project feasibility analysis report based on internal structure of the Louvre Museum.
- Built a mathematical model for Louvre Museum based on Graph Theory and implemented the Dijkstra algorithms using MATLAB to obtain the optimal evacuation routes under different circumstances.

North Sea Pelagic Fishes' Population Pattern Simulation Model Design

Jan 2020 — Feb 2020

- Built a Logistic model and a Lotka-volterra model to predict fish population changes in the North Sea.
- Designed and implemented an algorithm of fish migration pattern calculation using python.
- Our simulation suggested that pelagic fish will generally migrate to the northeast North Sea over the next 50 years.
- Our analysis and visualization were used to propose strategies for fishing companies around the North Sea.

PROJECTS EXPERIENCES**Robust SimCLR via Adversarial Meta-Learning**

New York, NY | Oct 2022 — Present

*Representation Learning Course Project**Instructor: Prof. Carl Vondrick from Columbia University*

- Reproduced self-supervised learning algorithm SimCLR on CIFAR-10 datasets in Pytorch.
- Designed a robust version of SimCLR by adding additional adversarial branches with different types of adversarial attacks and combining it with meta-learning frameworks.
- Experimented the proposed model on CIFAR-10 dataset and identified the robustness of learned representations.
- Plan turning this into a long-term research and continuing experimentation.

Computer Game Design and Development

Ann Arbor, MI | Aug 2021 — Jan 2022

*Undergraduate Computer Science Capstone Project**Instructor: Prof. Austin Yarger from University of Michigan*

- Recreated the original *Nintendo* switch game *Metroid* using game development engine Unity (C#) and explored Unity physics systems to create new levels and mechanisms.
- Led a 4-people team, the *JADECat* studio, to develop an adventure game *Chaos* ([link](#)). Presented our game at *WolverineSoft showcase*. *Chaos* was selected as one of the top 10 best games of the showcase.
- Published the game on *itch.io* ([link](#)) and *GameJolt* ([link](#)).
- Led a 4-people team to develop a 3D puzzle game *Trapped* ([link](#)) with theme of Déjà vu.

Vision Transformer (ViT) Ablation Studies

Ann Arbor, MI | Oct 2021 — Jan 2022

*Computer Vision Course Project**Instructor: Prof. Andrew Owens from University of Michigan*

- Researched on the architecture of GoogLeNet/ResNet/vision transformers (ViT) and their performance bottleneck in image classification tasks. Implemented a *customized ViT* ([link](#)) and trained it on CIFAR-10/ImageNet.
- Performed small-scale hyperparameter tuning on Colab and trained model with fine-tuning.
- Led a team to conduct ablation experiments on key components of ViT, including positional embedding layers, normalization layers, multi-head attention blocks, and residual connections.

ODE-based Mathematical Modeling Projects

Ann Arbor, MI | Sep 2020 — Jan 2021

*Differential Equation Course Projects**Instructor: Dr. Jeff Dunworth from University of Michigan*

- **Two-compartment Pharmacokinetic Model:** Used the two-compartment ODE model in pharmacokinetic to analyze drug distribution in body and proposed a recommendation of how to adjust drug size for children.
- **Model of Structural Rigidity:** Modeled the behavior of three-story structure during earthquake using 6th order ODE systems and discovered that proper dampers can retard the motions of each floor.
- **Delayed Differential Equation (DDE) Model for CO₂ Concentration in Human Blood:** Modeled the carbon dioxide concentration in human blood in high altitude environment using DDE and derived its steady state solution.

Carbon Footprint Generator App Development

Madison, WI | Oct 2019 — Oct 2020

*MadHack (2019) Competition Project**Team Leader & Android App Developer*

- Designed a carbon calculation App for modern people's weak environmental awareness.
- Modified and optimized the calculation algorithm of carbon footprint to accommodate the situation of the U.S..
- Developed an App on Android Studio using Java, including the module of carbon footprint computation, the module of comparing users' carbon emissions with the world's average level, and the module of pushing news of environment and sustainability.

CU COMS 4111: Database SystemAn interactive Web App for restaurants in New York city ([demo](#))**CU STAT GR5702: Data Visualization**RStudio editor theme customization tutorial ([link](#))**CU STAT GR5702: Data Visualization** An analysis with visualization of *Walt Disneyland* waiting time data ([link](#))**UMich EECS 445: Machine Learning** A transfer-learning-based model with data augmentation to classify animals.**UMich STATS 415: Data Mining** A statistical learning model on financial dataset to predict forward return of stock.**UMich EECS 281: Data Structure & Algo**

A solution to TSP problem using MST and B.&B. algorithms.

UW CS 400: Data Structure & Algo

An interactive farming data visualization project using Java GUI.

AWARDS & HONORS

- *Highest Distinction* of class 2022 (*University of Michigan*) **Awarded (Top 0.5%) | 2022**
- James B. Angell Scholar (*University of Michigan*) **Awarded (Top 2%) | 2022**
- 4 × University Honors (*University of Michigan*) **Awarded (Top 5%) | 2020 – 2022**
- 4 × Dean's List (*University of Wisconsin-Madison*) **Awarded (Top 5%) | 2018 – 2020**
- *Mathematical Contest of Modeling (MCM)* **Honorable Mentioned (Top 10-15%) | 2019 & 2020**

COURSE HIGHLIGHTS & SKILLS

Mathematics Skills	Abstract Algebra, Algebraic Combinatorics, Real Analysis, Probability Theory, Stochastic Process, Linear Algebra, Differential Equations, Mathematics for Finance, Calculus, Discrete Mathematics.
Computer Science Cores	<ul style="list-style-type: none"> • Programming/Algorithm/Theory: Advanced Programming, Data Structure and Algorithm, Foundation of Computer Science. • Machine Learning: Computer Vision, Theory of Machine Learning, Natural Language Processing, Data Mining and Statistical Learning, Deep Learning for Computer Vision, Representation Learning. • Computer Engineering(CE)/Others: Computer Organization, Computer Engineering, Database System, Computer Game Design and Development, Exploratory Data Analysis and Visualization.
Programming Languages	Python, C/C++/C#, R, SQL, Java, HTML, JavaScript, CSS, Matlab.
Frameworks & Packages	Pytorch, OpenCV, Numpy, NLTK, scikit-learn, Flask, Postgresql, ggplot.
Languages	Mandarin (Native), English (Full Professional Proficiency).
Other Skills	LaTeX, Github, Unity, Android Studio, Mathematica, SketchUp.

VOLUNTEERING & EXTRACURRICULAR EXPERIENCES

Study Group Mentors

Linear Algebra Mentor

Ann Arbor, MI | Feb 2021 – May 2022

- Gave review mini-lectures on Linear Algebra for my study group to help them in their courses/projects.
- Contents spanned from basic concepts like eigenspaces to its application in computer science algorithms.
- Since W22, I started sharing my knowledge of application of linear algebra in algebraic combinatorics to the group.

Python Programming Mentor

Madison, WI | Sep 2019 – Mar 2020

- Gave bi-weekly mini-lectures on programming/algorithms to non-CS major students (about 10) in the study group.
- Guided students to complete basic data structure projects using Python.
- Introduced to students the advantages of elegant coding styles and taught them why/how to do version control.
- Once a month, led discussion among students in the group to share knowledge/stories in their own major(s).

Volunteer of *University Career Fair* (2019)

Madison, WI | Sep 2019

- Arranged students' admission and guided them to find their favorite companies.
- Assisted in personnel arrangement and coordination of talent demand publicity of each company.

Founder & Organizer of Basketball Team

Madison, WI | Sep 2018 — May 2020

- Founded a basketball team and organized daily training and took in charge of new player recruit and selection.
- Organized the team members to participate in intramural sports competitions and Chinese basketball matches, and won the winning prizes for many times.