Sida (Star) Li

(510)-847-2494 ◇ listar2000@uchicago.edu ◇ https://listar2000.github.io/

EDUCATION

The University of Chicago Sept 2022 - June 2014
Master of Science, Statistics GPA: 3.92
University of California, Berkeley August 2018 - May 2022
Bachelor of Arts, Statistics & Computer Science GPA: 3.95

RESEARCH INTERESTS

Approximate Bayesian Inference, Probabilistic Machine Learning, Deep Generative Method, AI4Science

RESEARCH EXPERIENCE

The University of Chicago Statistics Department, UChicago, IL
Mentor: Daniel Sanz-Alonso *June 2023 - Present*
Working on accelerating and generalizing Langevin Monte Carlo (LMC) methods for sampling. Experimented and verified how adding a curl matrix into the Langevin SDE accelerates convergence in Gaussian mixture models. Ongoing works include implementing an algorithm to find the optimal matrix parameters and applying the ensemble Kalman sampler to generalize LMC beyond limitations on exact gradient calculations.

Autonomous Empirical Research Group, Brown University, RI
Mentor: Sebastian Musslick *March 2022 - Present*
Researching into symbolic regression (SR) - the ML problem that searches the best-fitting expression for a given dataset. Pioneered the design of an SR method based on Generative Flow Networks and deep learning with SOTA performance under noise. Currently developing a hierarchical Bayesian framework for the SR problem and corresponding inference algorithm to sample from the posterior.

FHL Vive Center for Enhanced Realtiy, UC Berkeley, CA
Mentor: Allen Yang *March 2020 - May 2021*
Developed ROAR, an autonomous racing simulator, and applied model-based deep reinforcement learning algorithms to vehicle controllers for autonomous racing.

Sandrine Dudoit Lab, UC Berkeley, CA
Mentor: Hector Roux de Bezieux and Koen Van den Berge *January - May 2020*
Participated through the Undergraduate Research Apprentice (URAP) program. Investigated how initialization affects unsupervised dimensionality reduction methods such as UMAP and t-SNE for scRNAseq data, with an emphasis on the preservation of global structures in low dimensional space.

WORK EXPERIENCE

Software Engineer Intern, Duolingo, Pittsburgh, PA *May-August 2021*
Implemented internal tools in the ETL data pipeline that support efficient querying and computation on key metrics (e.g. daily bookings, active users); revised the A/B testing framework by enabling auto-correction in confidence intervals for ad-hoc metrics.

Data Consulting Intern, Concha Inc., Berkeley, CA *January-May 2020*
Worked on predicting customer's hearing loss curve based on response data from online testings. Applied and evaluated existing machine learning methods such as regression tree and RNN for the prediction tasks.

PAPERS & REPORTS

Sida Li, Ioana Marinescu, Sebastian Musslick. "GFN-SR: Symbolic Regression with Generative Flow Networks." **NeurIPS 2023 AI4Science Workshop**. [Link] [Web]

Sebastian Musslick, Joshua Hewson, Ben Andrew, **Sida Li**, George Dang, John Gerrard Holland. "Evaluating Computational Discovery in the Behavioral and Brain Sciences." **AAAI 2023 Spring Symposium Series, Computational Approaches to Scientific Discovery**. [Talk Abstract]

Sida Li, Joshua Hewson, Sebastian Musslick. "Hierarchical Bayesian Symbolic Regression." **Work in Progress, 2023**. [Link]

Michael Estrada, **Sida Li**, Xiangyu Cai. "Feedback Linearization of Car Dynamics for Racing via Reinforcement Learning." **Preprint, 2021**. [Link]

SOFTWARES

ROAR Simulator (https://augcog.github.io/ROAR/introduction/)
An open-source platform/API for autonomous driving simulations based on CARLA. Include pre-built algorithms in perception (computer vision), control, planning and visualizations.

OpenArk Atlas (https://wuxiaohua1011.github.io/ATLAS_Annotation_APP)
A 3D model rendering and annotation tool under the FHL Vive Center. Integrate generalized AR/VR algorithms such as point-cloud segmentation, flood-filling, and plane-fitting. Offer a powerful user interface for visualization and annotation.

AWARDS

UChicago M.S. Stat Scholarship (25% tuition remission)	FA22, FA23
UC Berkeley Statistics Department Citation	FA22
UC Berkeley Dean's Honors List (top 10% GPA)	SP19, FA19, SP20, SP21
Upsilon Pi Epsilon (top one third of CS majors)	FA19, SP20, FA20, SP21

SKILLS

Languages: English, Mandarin, Cantonese
Programming: Python, R, C++, Java, Javascript, Ruby, LaTeX

TEACHING

CS 198-097 Robot Autonomous Racing DeCal (Head Instructor)	Fall 2021
CS 198-097 Robot Autonomous Racing DeCal (Instructor)	Fall 2020
STAT 134 Probability Theory (Tutor)	Spring 2020
STAT 134 Probability Theory (Tutor)	Fall 2019
MATH 32 Precalculus (Tutor)	Summer 2019