Aodong Li

aodongli.github.io | aodongli | Google Scholar | aodongl1@uci.edu | (+1) (646) 895-4765

Summary

- Fifth-year Ph.D. candidate in Machine Learning at UC Irvine;
- Interested in solving machine learning problems from a probabilistic view, especially distribution shifts related topics;
- Experienced in anomaly detection, Bayesian time series, changepoint detection, continual learning, temporal word embeddings, deep reinforcement learning, and neural machine translation;
- Publication records in mainstream machine learning conferences NeurIPS/ICML;

Publications

A. Li, A. Sankararaman, B. Narayanaswamy, **ProHDT: Probabilistic Hash Embeddings for Dynamic and Temporal Tabular Data**, Under submission.

C. Fung, C. Qiu, **A. Li**, M. Rudolph, **Model Selection of Anomaly Detectors in the Absence of Labeled Validation Data**, Under submission.

A. Li*, C. Qiu*, M. Kloft, P. Smyth, S. Mandt, M. Rudolph. **Zero-Shot Anomaly Detection without Foundation Models**. NeurIPS, 2023.

A. Li*, C. Qiu*, P. Smyth, M. Kloft, S. Mandt, M. Rudolph. **Deep Anomaly Detection under Labeling Budget Constraints**. ICML 2023.

C. Qiu*, **A. Li***, M. Kloft, M. Rudolph, S. Mandt. **Latent Outlier Exposure for Anomaly Detection with Contaminated Data**. ICML 2022.

A. Li, A. Boyd, P. Smyth, S. Mandt. **Detecting and Adapting to Irregular Distribution Shifts in Bayesian Online Learning**. NeurIPS 2021.

A. Li, R. Bamler, S. Mandt. **Quantifying Gender Bias Over Time Using Dynamic Word Embeddings**. SoCal Machine Learning Symposium 2020.

B. Foy, **A. Li**, J. McClung, R. Ranganath, J. Higgins. **Data-driven Physiologic Thresholds for Iron Deficiency Associated with Hematologic Decline**. American Journal of Hematology 2019.

A. Li, L. Sun, C. Li, Y. Wang, Y. Liu. **360-degree Video Streaming by Deep Reinforcement Learning**. Technical Report, November, 2018.

A. Li, S. Zhang, D. Wang. **Enhanced Neural Machine Translation by Learning from Draft**. APSIPA '17.

Work Experience

Amazon AWS AI Labs Jun. - Sept. 2023
Applied Scientist Intern, work with Abishek Sankararaman, Murali Narayanaswamy

Research topic and deliverable: probabilistic models of dynamic and temporal tabular data; one paper submitted to AISTATS 2024.

- Propose probabilistic hash embedding (PHE) as a universal interface for data with dynamic dictionaries;
- Propose a temporal model that uses PHE for multi-task temporal tabular data;
- Derive a variational (Kalman filtering) inference algorithm for PHE and the temporal model;
- Our method provides Bayesian online learning capabilities on dynamic and temporal tabular data.

Bosch Center for Artificial Intelligence (BCAI) Jun. 2022 - Sept. 2022
Machine Learning Research Intern, work with Maja Rudolph

Research topic and deliverable: **deep active anomaly detection**; one paper accepted at ICML 2023.

- Determine a set of conditions that generalize the anomaly score rankings from the queried to the unlabeled;
- Based on the conditions, we analyze querying strategies and propose k-means++-based diverse queries;
- Propose a novel semi-supervised anomaly detection loss that exploits both queried data and unlabeled data;
- Our proposed method significantly outperforms baselines on image, video, and tabular benchmarks.

Research topic: **continuous-time latent switching dynamic models**.

Qualcomm Technologies, Inc. Jun. 2021 - Sept. 2021
Software Engineering Intern, work with Hilmi Enes Eğilmez

Work topic: **neural video compression for autonomous driving**.
- Apply state-of-the-art neural video compression to in-car fish-eye videos;
- Analyze the video noise patterns and develop a temporal denoising filter for better compression performance;
- The proposed method reaches the same performance as the best manually-designed video codec HEVC.

Research Experience

University of California, Irvine Sept. 2019 - Present
Graduate Student Researcher, work with Stephan Mandt

Research topic and deliverable: **online changepoint detection and adaptation**; one NeurIPS 2021 paper.
- Consider online learning with distribution shifts that occur at an unknown rate and of unknown intensity.
- Derive a novel Bayesian online learning procedure to jointly infer and adapt to distribution shifts.
- Experiments on real-world data with concept shifts, image data with covariate shifts, and *large-scale* temporal texts show systematic improvements over baselines.

Research topic and deliverable: **deep anomaly detection**; one ICML 2022 paper; one ICML 2023 paper.
- Consider a new, practical problem setup when the training data contains unobserved anomalies;
- Propose a novel unsupervised anomaly detection algorithm inspired by outlier exposure.
- Experiments on three image datasets, 30 tabular data sets, and a video benchmark show consistent and significant improvements over the baselines.

Research topic and deliverable: **quantifying gender bias in texts (NLP)**; one SoCal ML symposium 2020 paper.
- Quantify gender bias evolution through *temporal word embeddings*;
- Analyze three large temporal corpora (100M+ tokens) and automatically select the top gender-biased words;
- Find changes of gender bias align with historical events such as feminist movements.

CILVR Lab at New York University Sept. 2018 - Jun. 2019
Graduate Student Researcher, work with Rajesh Ranganath

Research topic and delivery: **Inference and machine learning for healthcare**; one journal at American Journal of Hematology 2019.
- Consider the problem of estimating red blood cell's age distribution, which contributes the glycation in blood;
- Given very limited number of glucose measurements, infer the age distribution by deconvolving the glycation;
- Propose a data-driven method to re-define the iron-deficiency threshold for different gender groups.

New York University Video Lab Jun. 2018 - Jun. 2019
Graduate Student Researcher, work with Yao Wang

Research topic and delivery: **deep reinforcement learning on video streaming**; one technical report.
- Apply *advantage actor-critic algorithm (A2C)* to manage the playback buffers in 3D VR video streaming;
- Incorporate the user vision information to improve the scheduler;
- Our system improves the video quality of a baseline system by 18%.

Center for Speech and Language Technologies at Tsinghua University May 2016 - June 2017
Research Intern, work with Dong Wang

Research topic and delivery: **neural machine translation (NLP)**; one APSIPA 2017 paper.
- Propose a two-stage drafting and refinement approach for neural machine translatioin;
- The proposed method improves the baseline by 0.9-2.5 BLEU score.

Compose Chapter 1 of the book *Modern Machine Learning Techniques* published by Tsinghua University Press.

Education

2019 - Present Ph.D. in Computer Science at **University of California, Irvine**
2017 - 2019 M.S. in Computer Science at **New York University**
2014 - 2017 M.Eng. in Electronic and Communication Engineering at **BUPT**
2010 - 2014 B.Eng. in Communication Engineering at **BUPT**

Miscellaneous

Teaching Assistants
Fall 2021, 2022 **Deep Generative Models**, UC Irvine
Fall 2019 **Introduction to Machine Learning**, UC Irvine
Spring 2019 **Machine Learning**, NYU
Fall 2018 **Foundamental Algorithms**, NYU

Honors
Fall 2019 **Dean's Award**, ICS@UC Irvine
Fall 2018 **M.S. Thesis/Research Fellowship**, CS@NYU

Academic Services
Reviewer for AISTATS 2021, ICML 2021-2023, NeurIPS 2021-2023, ICLR 2022-2023.

Skills
Python (PyTorch, TensorFlow, Numpy, Theano), LaTeX, Bash, C/C++, JAVA, JavaScript, PHP, HTML, SQL.