Yaya (Jinya Jiang)

yayajinyajiang@gmail.com   (253)230-3117   San Diego, CA   Linkedin   Github

Education

University of California, San Diego	*San Diego, CA, US*
MS, Computer Science, Machine Learning	Sep 2024 - Mar 2026
BS, Computer Science(Major), Mathematics(Minor)	Sep 2021 - June 2024

Publication

- *Xintong Li, **Jinya Jiang**, Ria Dharmani, Jayanth Srinivasa, Gaowen Liu, Jingbo Shang. 2024. Open-world Multi-label Text Classification with Extremely Weak Supervision. **EMNLP** 2024. https://aclanthology.org/2024.emnlp-main.841.pdf*
- ***Jinya Jiang**, Richa Kafle, Christa Lehr, Simone Wright, Clarissa GuitierrezGodoy, and Christine Alvarado. 2024. Understanding California's Computer Science Transfer Pathways. In Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1 (**SIGCSE** 2024), March 20–23, 2024, Portland, OR, USA. ACM, New York, NY, USA, 7 pages. https://dl.acm.org/doi/10.1145/3626252.3630956*
- Related: https://aclanthology.org/2024.naacl-demo.13.pdf

Research

Jr Applied Scientiest Intern, Amazon   June 2025 - June 2026
- Incoming Applied Scientiest Intern at San Diego, CA.

Machine Learning Researcher, Ebay AWF   Dec 2024 - Present
- Utilized PySpark and Scala to process large-scale data, extracting insights from user search queries and item information to develop and recommend optimized keywords for items in ebay advertising systems(CPC).
- Identified trends and patterns in search queries, enhancing the accuracy and relevance of **keyword recommendation**.

Machine Learning Researcher (Shang Lab) , UC San Diego *EMNLP 2024*   Oct 2023 - Nov 2024
- Helped Pioneer an innovative framework, X-MLClass, for open-world **multi-label text classification** under extremely weak supervision, enhancing label space coverage by 40% over traditional methods.
- Achieved superior performance in zero-shot classification across benchmark datasets by refining label spaces through clustering key phrases from raw text, identifying dominant classes, and and iteratively incorporating rare long-tail labels often missed by traditional models. EMNLP 2024 paper accepted.

Machine Learning Researcher & Web Developer (Shang Lab) , UC San Diego   Sep 2022 - Sep 2023
- Applied Neural Network models (BERT, RoBERTa and LayoutLM) to fine-tune a model for question answering system to find specific information needed from offer letters, streamlining work authorizationfor international students.
- Developed a document processing application for UCSD's International Office (ISPO), automatically highlighting critical fields in offer letters, improving efficient document processing in handling paperwork, deployed in **Flask**.
- Contributed to the development of a platform for annotating PDF documents, training models, and document question-answering, preserving text and layout information for improved document processing.

Project

Deep Learning Developer, UC San Diego   Mar 2023 - Jun 2023
- Developed an ensemble of XGBoost regressors with hyperparameter tuning and implemented a Random Forest model, leveraging feature engineering skills like Haversine distance and time encoding to predict taxi ride durations.
- Built a multi-layer perceptron (MLP) model with embedding layers for encoding and incorporated an ensemble of 6,240 XGBoost regressors, which utilized a validation-based voting mechanism to adjust predictions for extreme cases, contributing to a top-3 finish in a Kaggle predictive modeling competition.

Machine Learning Developer, UC San Diego   Sep 2022 - Dec 2022
- Deployed classification models for book category prediction using bag-of-words and logistic regression to perform data mining on large textual datasets. Processed large datasets with feature engineering techniques.
- Designed and implemented a recommender system for video game reviews on Steam. Applied logistic regression and Jaccard similarity for user-game interaction prediction and time prediction.

Work

Software Engineer Intern, Meituan, *Beijing, CN*   Jun 2024 - Sep 2024
- Developed **unit tests** and automated test scripts using Appium to validate price expressions and ensure correct labeling.
- Designed, and deployed a real-time update website using **React, Redux** boosting productivity by 300% and reducing data update time by 80%. Enhanced cross-department collaboration and improved efficiency.
- Conducted **user research** and engineered a large font size feature for core user journeys, implementing dynamic font scaling, layout adjustments, and CSS variable mapping to ensure accessibility without compromising UI integrity.

CSE Teaching Assitant / Tutor, UC San Diego   Fall 2022, Winter 2024, Winter 2025, Spring 2025
- Instructed DSC30: Data Structures and Algorithms, Mgta415: Analysing Unstructured Data, and tutored CSE8A: Introduction to Programming, CSE30: Computer Organization & Systems Programming, focus on java, python, C. Held lab and office hours over 200+ hours, graded assignments, created and tested homework, resolved 400+ issues.