Yilin Bao

Phone: (310) 853 9757 | Email: yibao@ucsd.edu| git-hub: https://github.com/yilin-bao Website: https://yilin-bao.github.io/

EDUCATION

University of California, San Diego — San Diego, CA
Major in Mathematics and minor in Computer Science — Graduation in September 2021

University of California, San Diego — San Diego, CA
ECE department, Machine Learning and Data Science — Expected Graduation Winter 2023

RELEVANT COURSEWORK

Mathematics & Statistics:
Applied Linear Algebra, Applied Complex Analysis, Real Analysis, Functional Analysis, Random Process, Partial Differential Equation, Abstract Algebra, Actuarial Mathematics, Statistical Methods (R language).

Computer Science & Coding:
Intro Computer Science & Object-Oriented: Java (Java), Basic Data Structure & Object-oriented Design (Java), Software Tools & Techniques (Python, Linux), Advanced Data Structure(C\C++).

Data Science & Machine Learning:
Machine Learning (Pytorch), Deep Learning (Pytorch), Statistical Learning (MATLAB & Python), Virtual Learning (Pytorch); Computer Vision & Image Processing (Matlab & Python) Geometric Computer Vision (Python).

PUBLICATION

[1] Bethel, Brandon J. and Dong, Changming and Zhou, Shuyi and Sun, Wenjin and Bao, Yilin. "Assessing Long Short-Term Memory Network Significant Wave Height Forecast Efficacy in the Caribbean Sea and Atlantic Ocean." SSRN Electrical Journal, July 2022.

SUMMER RESEARCH

LLM RESEARCH

University of California, San Diego, CA — 2023

- Config the environment to fine-tune 7B and 13B generative NLP model
- Commit git repo to solve bug caused by relating package update, old methodology is abandoned, need to implement new function interface
- Construct dataset based on leetcode problem description and solution code for fine-tune training
- Fine-tune 7B/13B FastChat and evaluate results by GPT4 and design the test to evaluate the performance of model fine-tuned with a different approach.

WORK EXPERIENCE

MACHINE LEARNING ENGINEER:

Linkmoo, Charlotte, North Carolina — 01/2023-05/23

- Help the team develop a vector embedding database of PDF based on chatGPT for integrating into SaaS.
- After llama released, fine-tune llama 7B for the private deploy version.
- Using machine learning to study algorithms for reverse prediction of Amazon advertising traffic allocation strategies.

DATA SCIENTIST INTERN: "BrainE: Cognitive & Neural Modality"

University of California, San Diego, NEAT Labs — 02/2022-07/22

- Organized and cleaned raw data, enhancing generalization across various datasets (e.g., human micro-expressions, brain event-related potential signals).
- Evaluated therapy solutions using mathematical observables, applying Fisher's precision probability test to identify the most suitable approach.
- Performed one-tailed T-test comparisons between post and pre-sessions via hypothesis testing, providing robust quantitative evidence for treatment efficacy.
- Scripted tasks for Cognitive/BrainE, ensuring stable data exchange with the Redcap database system, and computed measurement scores (mindfulness, anxiety, empathy, compassion).

MACHINE LEARNING RESEARCH ASSISTANT: "Assessing Long Short-Term Memory (LSTM) Network Significant Wave Height Forecast Efficacy in the Caribbean Sea and Atlantic Ocean"

Nanjing University of Information Science and Technology (Remotely due to Covid) — 05/2020-09/21

- Understand if CS and AO subregions' distinct wave climates would have a strong influence on the forecasting of significant wave height (SWH) using the long short-term memory (LSTM) network.
- Finish data collecting for CS and AO datasets from the National Data Buoy Center from 2010 – 2019.

- Finish data processing and filling the large missing data by Simulating Waves Nearshore
- help group to build and improve the Numerical Wave Model in both datasets.
- Replicate and test various semantic segmentation models and tuning parameters.

MACHINE LEARNING RESEARCH INTERN: "Oceanic Eddy Identification Using AI Scheme"
Nanjing University of Information Science and Technology & UCLA 01/2019 – 05/2019

- Researched the role of oceanic eddies in global energy, material transport, heat and freshwater redistribution, climate change
- Responsible for the collection, sorting, processing of satellite-remote-sensing sea surface height data
- Compared applicability in oceanic eddy detection between AI algorithm and geometry-based methods
- Learned Pyramid Scene Parsing Network (PSPNet) and used it to extract features in multiple layers, research the distribution, size and shapes of oceanic eddies

AI RESEARCH ASSISTANT
Institute of Oceanology, Chinese Academy of Science Qingdao, Shandong
Key Laboratory of Ocean Circulation and Waves 06/2018 – 09/2018

- Design, prototype and build machine learning systems, frameworks, pipelines and tools that process large dataset Competently parallel task fully researched data and enterprise process improvements.
- Build convolutional neural networks based on PyTorch 1.0 that can aid in design of highly nonlinear ocean models and explore the usage of unsupervised learning on ocean area classification.
- Recording the meeting outline and process, playing the role of communication between research group and assistant subteam, holding agile development process in the subteam and making objectives and key results (OKR) table report weekly.

PROJECTS

COMPUTER VISION "Unofficial Pytorch implement for Siamese Masked Autoencoder"

- Reproduce and debug two open-source paper Masked Siamese Networks for Label-Efficient Learning and Masked Autoencoder
- Implement an unofficial code for closed course paper Siamese Maked Autoencoder published by Facebook research group.

MODEL VISUALIZATION "panoramaPy": Powerful Analysis and Network Overview of Robust Adaptive Machine-learning Architecture

- Designed a small interpreter based on ast (abstract synatx tree), encoded machine learning module to multiple visualization formats, no need to downloads dozen GiB .pth model parameters
- Accelerate my speed of reproducing unknown model ten times faster than before

DATA ANALYSIS "Food Insecurity Analysis based on USDA nutrition data"

- Designed and created a user-friendly dashboard by local connection to analyze and present food insecurity trend in the United States over past decade using USDA data via visually appealing and informative graph
- Collaborate with team members to document the data analysis process and conclusions from an analyst's perspective, propose possible policy decisions, and plan effective budgets for the organization

DATA SUPPORTED DECISION SYSTEM "Bank Telecaller Decision Support System"

- Developed a Bank Telecaller Decision Support System using interactive plots and machine learning
- Contributed to the improvement of telecalling efficiency by providing insights into which customers are most likely to make a purchase from data science perspective
- Utilized the XGBoost model and achieved a 90% prediction accuracy for the success of future marketing campaigns

Skills

Programming Tools: Java (intermediate-level), Python (advanced-level), R language (intermediate-level), C (beginner), C++(beginner), MATLAB (advanced-level), SQL (beginner), Golang(beginner);