Yilin Bao

Phone: (310) 853 9757 | Email: yibao@ucsd.edu

EDUCATION

University of California, San Diego — San Diego, CA
Major in Mathematics and 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, Mathematical Reasoning, Applied Complex Analysis, Real Analysis, Functional Analysis, Random Process, Partial Differential Equation, Abstract Algebra, Modern 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.

Data Science & Machine Learning:

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

PUBLICATION

[1] Bethel, Brandon J. and Dong, Changming and Zhou, Shuyi and Bao, Yilin, Assessing Long Short-Term Memory Network Significant Wave Height Forecast Efficacy in the Caribbean Sea and Atlantic Ocean. Available at SSRN: https://ssrn.com/abstract=4153300 or http://dx.doi.org/10.2139/ssrn.4153300

WORK EXPERIENCE

MACHINE LEARNING RESEARCH: "Finetune FastChat and Llama based on constructed coding dataset"
University of California, San Diego — 05/2023-09/23

- Code and Debug the FastChat LoRA, learn how to establish fine-tune workflow and contribute high quality pull request to the Github repository
- Designed experiments to validate that models of different sizes (7B and 13B) have varying upper limits on their reasoning chains, and simultaneously designed experiments to verify the hypothesis that "code fine-tuning is more effective at stimulating the model's thinking abilities than human language fine-tuning
- Evaluate the Model fine-tuned by Code solution and Model fine-tuned by Human Language solution, Judge the different alignments outcomes
- Prepare dataset to align language model, one based on leetcode problem description solution and code solution for Fine-Tune training, the other is a pair by physics question and code solution to them
- Quantitative analysis of the performance of different models in several fundamental areas (such as mathematics, reading, and writing), particularly along the logical reasoning chain.

DATA SCIENTIST: "BrainE: Cognitive & Neural Modality and Data Score Analysis"
University of California, San Diego, NEAT Labs — 02/2022-07/22

- Research in a team on human brain and cognitive mechanism such as selective attention and response inhibition, leading to new veterans' mediTMS psychotherapy solution.
- Organize raw data and do cleaning, remove all unreasonable the outliers by cross-checking related features, improve the ability of generalization on multiple datasets, such as human microexpressions images and brain event-related potential (ERP) signal.

- Investigate multiple therapy solutions in mathematical observables and use Fisher's precision probability test to determine the most suitable for our situation.
- Perform one-tailed T test comparison between post and pre session by hypothesis test, providing solid quantitative proof on availability of treatment options from data science view.
- Write scripts for Cognitive/BrainE tasks, build stable data exchange with REDCap database system and calculate the scores of measurements e.g. mindfulness anxiety, empathy, compassion etc.

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.
- Writing and revising articles.

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

Summer Research

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, play the role of communication between research group and assistant subteam, holding agile development process in the subteam and make objectives and key results (OKR) table report weekly.
- Receive assignments in the form of objectives and establishes goals to meet those objectives.
- Collaborated closely with scientists from different field to implement, test and improve machine learning solutions

Projects

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 ANALYSIS "House rent and house price analysis in California"
- Processing on extremely complex data sets with very large parameters, based on appropriate and solid statistical theory, eliminates a large number of noise parameters, greatly improving team work efficiency
- Contributed to a deeper understanding of the real estate market in California and the factors affecting house prices and rents
- Improved decision-making abilities for individuals and organizations looking to invest in California real estate
- Demonstrated proficiency in data science methodologies and their application to real-world problems.

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);