Wenyi (Wesley) Tao

(201) 201-9455 * wt2271@columbia.edu * https://wesleytao.github.io/

EDUCATION

Columbia University *M.A. in Statistics, New York* GPA: 3.91/4.00 Sept.2017 - Jan.2019
Courses: Reinforcement Learning, Machine Learning, Analysis of Algorithm, Bayesian Statistics, Advanced Data Analysis
Fudan University *B.A. in Economics, Shanghai* GPA: 3.53/4.00 (Top 1 final year) Sept.2012 - July.2017
Courses: Applied Statistical Tools, Econometrics, Advanced Mathematics, Data Structures, Intro to Database

SKILLS

Model: Lasso, Ridge, Logistic, Boosting, Bagging, CART, SVM, Clustering, Latent Dirichlet Allocation, EM algorithm
Tools: SQL, Python, R, Git, Scikit-learn, Keras, Pytorch, PySpark, Tableau

MACHINE LEARNING EXPERIENCE AND INDEPENDENT PROJECT

09.2018-present **Machine Learning Engineer Intern,** Pactera OneConnect AI Lab New York

- Built an end-to-end chatbot assistant to facilitate the company's hiring process collaborating with other engineers
- Implemented LSTM, SVM, Tree-based models to upgrade hard-coded dialogue and created a user simulator to generate simulation data for testing
- Created and maintained the SQL databases on AWS for chatbot to access and retrieve relevant information

06.2018-09.2018 **Machine Learning Engineer Intern,** Adatos A.I., New York

- Processed, augmented, and pipelined large batches of images to feed into the end-to-end Deep Unet model
- Decreased the mean absolute error (in around 300 trees) from 8 to 3 by replacing existing Gaussian blob detection with deep-learning powered tree counting model and speeded up the tree counting process from 1.83s per image to 0.1s per image by designing and fine-tuning an Unet Model
- Rescaled, trained, and transformed batches of images to solve the poor performance boundary problem of segmentation models (FCN and Unet)

05.2018-09.2018 **Research Assistant,** Harvard T. H. Chan School of Public Health New York

- Collected and cleaned millions of tweets and designed the corresponding relational database schema
- Discovered that people of Marseille Longchamp have symptoms of asthma 2 hours after local PM2.5 hit 35 μg/m3
- Implemented pagerank algorithm to find the influencer of the retweet network using SQL

PROJECT AND COMPETITION EXPERIENCE

12.2018 **Capstone Project,** Alpha Zero for Board Games using TensorFlow on Google cloud platform New York

- Reimplemented AlphaZero algorithm for board games (Gomoku, Checker and Connect Four)
- Built Monte Carlo tree search (MCTS) algorithm and MCTS self-player API
- Automated data generating process with auto-play between different models among AlphaZero, TD, MC learning

04.2018 **Independent Researcher,** Collaborative Filtering with EM Clustering New York

- Built from scratches a recommender system with Bayesian Clustering Algorithm without using any ML framework
- Tried multiple parametrized distribution within the conjugate family, evaluated different models based on performance
- Use cluster structure robustness, perplexity to tune the hyperparameter and use utility score to evaluate the performance

04.2018 **Group Project Leader** Tweets Sentiment Analysis on Sino-US Trade War New York

- Implemented Latent Dirichlet Allocation to perform topic modeling for millions of tweets related to Trade War
- Use interactive D3 to visualize the retweet relation networks and found those influential opinion leaders on this topic
- Visualize majority opinion's shift across a series of events after performing sentiment analysis on all the comments

01.2018 **Top 5%:** Alibaba Cloud Algorithm Competition- Helping Balloons Navigate the Weather New York

- Planned safe and fastest flight routes for ten unmanned balloons to deliver packages to their destinations
- Creatively implemented the A-Star algorithm in 3-Dimension space to solve the obstacle changing problem

08.2017 **Research Assistant,** Fudan Institute of Data Science Shanghai

- Found people under the non-flat-rate policy use 12 percent more electricity after 21:00 compared to the control group
- Quantified the price elasticity between the two group users, proposed a new pricing policy for the Power Company

INTEREST

- Swimming (2000m nonstop with medley)