

Wenyi (Wesley) Tao

Phone: (917) 655-8355 Email: wt2271@columbia.edu Portfolio: <https://wesleytao.github.io/>

EDUCATION

Columbia University *M.A. in Statistics, New York* GPA: 4.00 Expected Dec 2018

Related courses: Reinforcement Learning, Machine Learning, Analysis of Algorithm, Bayesian Statistics, Intro to Database

Fudan University *B.A. in Economics, Shanghai* GPA: 3.53/4.00 (top 1 in final year) Sept.2012 - July.2017

Related courses: Applied Statistical Tools, Econometrics, Advanced Mathematics, Data Structure in C, Ordinary Differential Equation

Model: Lasso Ridge, SVM, AdaBoost, Bagging, K-means, Logistic, Hierarchy Bayesian, EM algorithm, lightGBM

Tools: SQL, Python, R, Git, OpenCV, Scikit-learn, Spacy, Pytorch, PySpark, Tableau

WORK EXPERIENCE

Oneconnect, New York, Machine Learning Engineer Intern Sept.2018 – Present

- Improved the current version of goal-oriented Chatbot which understand the context of dialogue history by embedding the machine learning model into the handcrafted hard-code system and featurizing the slot value and previous intent
- Added a TrueCaseAnnotator component into the data pipeline to improve the entity extraction rate by 30% and reduce fallback rate
- Used bags of word representation to handle multi-intent classification and customize the backend API call to generate proper response
- Replaced the low-level DBAPI with ORM mapper and make the code base easier to maintain and managed

Adatos A.I. , New York, Machine Learning Engineer Intern Jun.2018 – Sept.2018

- Decreased the mean absolute error (in around 300 trees) from 8 to 3 by replacing existing Computer Vision algorithms 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 a Unet Model
- Created production-level code for providing an end-to-end machine learning solution to automatic tree counting
- Solved the poor performance on boundary problem of Deep Learning semantic segmentation models (FCN and UNet)

Harvard T. H. Chan School of Public Health, Research Assistant May.2018 – Sept.2018

- Worked closely with a [Dr. Hwsen](#) on multiple bioinformatics projects leveraging social media data including twitter, google trend
- Discovered that some people of Marseille Longchamp have symptoms of asthma 2 hours after local PM2.5 hit 35 $\mu\text{g}/\text{m}^3$
- Detected a statistical difference that the level of social inequality of Jim-Crow-Law States are higher than others and social capital index served as a mediation effect that accounts for 10% of the total effect

Fudan Institute of Data Science, Research Assistant Aug.2017

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

PROJECT EXPERIENCE

Data Mining for Target Marketing (1st Place Award) – 2018 American Statistical Association DataFest April.2018

- Discovered the dominant features like license-required, job description that can boost around 5%-12% of average clicks per day
- Found job posted before November statistically have more clicks than job posted after November excluded other effects

Sentiment Analysis and Topic Modeling with Tweets on US-China Trade War April. 2018

- 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

Alibaba Cloud Algorithm Competition: Future Challenge Helping Balloons Navigate the Weather Top 5% Jan.2018

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

Recommendation System: Collaborative Filtering Algorithm with MS Website and Movie Score Dataset April.2018

- Implemented from scratches a recommendation algorithm Bayesian Clustering with a users' website browsing history dataset
- Evaluated different algorithms based on a customers' utility metrics which was also implemented from scratches

INTEREST

- Swimming (2000m nonstop with medley)