

ZHIRUI(JERRY) WANG

3333 Broadway, C22E, New York, NY, 10031 • zw2389@columbia.edu • (571) 317-8414

<https://www.linkedin.com/in/zhiruiwang> • <https://github.com/zhiruiwang>

SUMMARY

An interdisciplinary constant learner with experience in research, analytics, coding, modeling and machine learning. Data science problem solver and forward-looking trailblazer with curiosity, research stamina, critical thinking and “Can do” attitude. Storyteller with excellent skills in data visualization. Works well within a matrix team environment.

EDUCATIONS

Columbia University | School of Engineering and Applied Science **New York, NY**

- M.S. in Data Science, GPA:3.85

2016 - 2017

Relevant Courses: Machine Learning, Natural Language Processing, Data Visualization, Big Data Analytics, Probabilistic Programming, Time Series Modelling, Stochastic Methods in Finance

Georgetown University | Graduate School of Arts and Sciences

Washington, D.C.

- M.A. in Applied Economics, GPA:3.87

2015 - 2016

Relevant Courses: Macroeconometrics, Microeconometrics, Behavioral Economics & Finance, Causal Inference

Fudan University

Shanghai, China

- B.S. in Biological Sciences, Minor in Economics, Major GPA: 3.38, Cumulative GPA: 3.17

2011-2015

Relevant Courses: Biostatistics, Bioinformatics, Computational Structure Biology

Awards: Excellent Graduation Thesis, Third Prize of Undergraduate Excellent Student Scholarship (twice)

COMPUTER LANGUAGES AND SOFTWARE SKILLS

- R (tidyverse, shiny, leaflet, plotly, h2o, caret, rmarkdown, parallel), Python (numpy, pandas, sklearn, xgboost, tensorflow, keras, multiprocessing), STATA, SQL, Hadoop, Spark
- Tableau, Alteryx, LaTeX, MS Office, UNIX, Git, AWS

WORKING EXPERIENCE

The Boston Consulting Group

Boston, MA

Summer Analytics Associate | BCG Gamma team (Advanced Analytics & Data Science)

Jun 2017 – Aug 2017

Analytics Associate

Mar 2018 Onward

- Been on a team working for an oil industry company, using NLP techniques such as POS tagging and entity recognition (in Python SpaCy package) to analyze the maintenance text record of oil field equipment and identify potential problems, built a shiny app for visualizing an interactive word network for the entities and problems
- Did predictive maintenance on the oil field equipment based on the sensor data and equipment own features using xgboost in Python, also built a shiny app for interactive failure pattern visualization
- Been on another team working for a video rental company, using xgboost to predictive the transaction volume based on locality and movie features, integrated with colleagues' optimization algorithm to provide a video allocation system

PROJECT EXPERIENCE

Predictive Maintenance with Bayesian Recurrent Neural Networks

Sep 2017 – Dec 2017

Columbia University

New York, NY

- Built Bayesian recurrent neural networks with different number of layers, recurrent units, and priors using Tensorflow and Edward modules in Python, tried different inference methods such as variational inference and Bayes by Backprop, applied these networks on Turbofan Engine Degradation Simulation Data Set provided by NASA, and reached better prediction results than non-Bayesian recurrent neural networks
- Took care of the build-up and fine-tuning of the Bayesian recurrent neural networks and corresponding report write-up

Neural Machine Translation

May 2017

Udacity Deep Learning Nanodegree

- Used Tensorflow in Python to train a Seq2Seq recurrent neural network model with bidirectional LSTM on a dataset of English and French sentences that can translate new sentences from English to French

Exploring the Power of Yelp Review for Business Popularity

Sep 2016 – Dec 2016

Columbia University

New York, NY

- Utilized PySpark and R H2O.ai API to analyze Yelp's review and business datasets, used n-gram model and open source word library to extract positive/negative words from the review data, visualized them as word clouds, and applied these information as features to predict review rating and business popularity using Gradient Boosting and Random Forest model
- Responsible for data cleaning, data visualization in PySpark and machine learning model building in R h2o package

Cross Country and Cross Time Variation in the Anchoring of Inflation Expectation

Jun 2016 – Aug 2016

Georgetown University

Washington, D.C.

- Collected data from FRED Economic Data and OECD Data, implemented nonlinear autoregressive model with exogenous inputs to capture the anchoring of inflation expectation, did cross country and cross time comparison on 41 countries from 1967 to 2016 using rolling window analysis, found key factors for central banks to keep well anchored inflation expectation
- Explored functional programming and parallel computing in R, improved the efficiency of rolling window computation