FAN YANG

New York • 10026 | (917) 402-4686 | fy2232@columbia.edu | https://github.com/abx67

EDUCATION

Columbia University *Sept. 2017- Feb. 2019*

- M.A. in Statistics GPA: 3.86/4.0 *New York*
- 2nd Place | 2018 American Statistical Association (ASA) DataFest Competition

Xi'an Jiaotong University *Sept. 2013- Jun. 2017*

- B.S. in Statistics GPA: 3.68/4.0 *Xi'an, China*

TECHNICAL SKILLS

- ***Language**:* Python, C++, R, Java, SAS, HTML; ***Skills**:* Deep Learning (CNN/LSTM), ETL pipelines, OCR
- ***Tools**:* SQL, CUDA, Tableau, AWS, Linux; ***Frameworks*:** TensorFlow, Caffe, Gensim, OpenCV

WORK EXPERIENCE

Siemens [C++, C, Python, Linux] *Jun. 2018- Aug. 2018*

- **Software Engineer – Machine Learning** *Suzhou, China*
- Built and modified applications for implementing machine learning algorithms on embedded device
- Implemented yolo v3 deep learning algorithm for real-time vehicle detection and tracking and license plate recognition
- Trained the CNN model based on DARKNET neural network framework and reached recall over 90%
- Created prototype for detecting human key points, gesture and OCR based on TensorFlow framework

Center for National Resource Economic Studies [Python, MySQL, MS office] *Feb. 2017- Apr. 2017*

- **Data Analyst Intern** *Beijing, China*
- Designed and managed data analysis pipeline; built SVM and Random Forest models to analyze economic growth
- Increased accessibility and usability of customer data by redesigning data visualization techniques to include statistical graphs and information graphics and provided data analysis report including data visualization in tableau
- Built regression model to analyze relationships between different economic indicators and achieving 10% more accurate prediction of performance than previous years

LAB EXPERIENCE

DVMM Lab *New York* *Feb.2018- Jun. 2018*

- Conducted event specific (city traffic) multimodal pattern mining, topic modeling and data mining research
- Extracted features from Twitter text and applied Hidden Markov Model, Gaussian Mixture Model and K-means method
- Implemented Word2vec model and N-gram model for semantic interpretation and refined the corpus on twitter data
- Used rank constrained regression and fully convolutional networks methods to understand traffic pattern from large-scale web camera videos with low frame rate and low resolution; conducted feature extraction based on OpenCV

PROJECTS

AlphaZero for Board Games *Oct. 2018- Dec. 2018*

- Implemented AlphaZero algorithm for playing board games (Gomoku, Checker and Connect Four) from self-play training based on TensorFlow framework on google cloud platform
- Responsible for Monte Carlo tree search (MCTS) algorithm and MCTS self-player API
- Created function for auto-play between different models among AlphaZero, TD learning and MC learning

Tumor Detection on Gigapixel Pathology Images *Oct. 2018- Dec. 2018*

- Created deep learning framework to detect and localize tumors in gigapixel microscopy images
- Built CNN architecture based on transfer learning with data resampling and data augmentation
- Developed multi-scale input CNN model and achieved image-level AUC scores above 97%

Movie Recommendation System *Mar. 2018- Apr. 2018*

- Applied memory-based and model-based (EM algorithm) Collaborative Filtering algorithm on *EachMovie* dataset and established demo system recommending the user movies based on the movies feed to the system and their ratings
- Used cosine similarity to find most similar users and SimRank to find most relevant movies and reached 1.026 MAE
- Built demo of deep learning based system with two neural networks, one for candidate generation and one for ranking