

Summary

Data Scientist and developer with experience in design, implement and evaluate machine learning solutions for both academic projects and industry productions. Deep Learning practitioner, believer of neural network architecture engineering is the new feature engineering.

Experience

IBM Watson Health – Data Scientist

Cambridge, MA since Apr. 2017

† Trained domain specific word vectors which feed into recurrent networks for genetic mutation effect classification based on medical literatures.

† Built knowledge graphs and implemented graph traversal strategies to facilitate validation and inference on the entity relationships extracted from natural language models.

† Full-stack developer for internal dashboard monitoring status and metrics of continuous integration and delivery pipeline with REST API data collection, MongoDB data storage, Angular.JS frontend and D3 for plotting.

† Conducted propensity score matching analysis for patient cohorts for comparing the effectiveness of medicines.

Fidelity Investments – Data Science Intern

Smithfield, RI Jun. 2016 - Sept. 2016

† Worked on projects for the AI lab including face recognition, voice assistant and text classification with neural networks.

Zhenjiang Library – Database Developer

Zhenjiang, China Nov. 2010 - Feb. 2015

† Implemented a collaborative filtering book recommendation system used to generate customized email newsletters.

† Developed and maintained PL/SQL package for analytic use cases and report generation.

Projects

† Jigsaw Toxic Comment Classification (on going): Experimenting text classification using GloVe/fastText embeddings, VDCNN, DPCNN and stacked bi-directional LSTM with Max K Pooling.

† TensorFlow Speech Recognition Challenge: Ranked 48 out of 1300 teams in Google Brain's competition on small foot-print keyword spotting. Ensemble of a collection of deep learning models includes ResNet, VGG 1D and 2D CNN, stacked stateful LSTM recurrent networks, Convolutional RNN and depth separable CNN.

† Safe Driver Prediction: Ranked 56 out of 5169 teams in Kaggle competition on ranking insurance claim holder on their risk of filing a claim in the next year. Constructed a two-layer stacking model of gradient boosting machines, regularized greedy forests, factorization machines and neural networks with entity embedding.

Education

Bentley University – Master of Science in Business Analytics - GPA 4.0

Dec. 2016

† Research Assistant: Worked with Prof. Babaian on developing D3 interactive data visualization for graph dataset.

† Coursework: Programming, Databases, Hadoop, Time Series, Optimization, Statistical Learning, Data Mining

Publication

† Tamara B., Ren Z., Wendy L. *DTMi – a New Interface for Informed Navigation*, in proceedings HCI International 2017

† Lei Z., Ren Z. Prithwish C. Et al. *Comparative Effectiveness of Edoxaban and Warfarin in Prevention of Stroke and Systemic Embolism in Non-valvular Atrial Fibrillation Using Observational Healthcare Data*, in ISPOR 2018

Skill Set

† Deep Learning, Machine Learning, Statistical Modeling, Optimization, Data Analysis, Software Development, Algorithms

† Languages: Python, R, Java, Scala, Matlab, Javascript

† Libraries: Pandas, Numpy, Scipy, Sklearn, Tensorflow, Keras, PyTorch, NLTK, Gensim, Angular, D3, Seaborn, Networkx

† Systems: Linux, MongoDB, SQL, Hadoop, Spark, Map-Reduce, Hive, Pig

† Tools: Shell scripts, Git, Jenkins, Chef, Urban Code, Docker