Weiqi Weng

28 Westland Ave, Apt. 27, Boston, MA 02115
(857) 284-9548 | weng.wei@husky.neu.edu | weiqiweng.github.io

EDUCATION

Northeastern University, Boston, MA
Master of Science in Computer Science GPA: 3.875/4.0 Sept 2015 - Dec 2017
Teaching Assistant of Data Mining Techniques/Unsupervised Learning Sept 2017 - Dec 2017

East China Normal University, Shanghai, China
Bachelor's Degree in Mathematics and Applied Mathematics Sept 2011 - July 2015

TECHNICAL KNOWLEDGE

Language: Python, C/C++, R, Java, JavaScript, Scala
Database: MySQL, MongoDB, PostgreSQL, Vertica
Web: React, AngularJS, Node.js, D3.js, Express, HTML, CSS, jQuery, Bootstrap, PHP
Software: PyCharm, RStudio, Visual Studio, Eclipse, WebStorm, Matlab, SPSS, DBeaver

WORK EXPERIENCE

Wayfair, LLC., Boston, MA Jan 2018 - Present
Data Scientist, Recommender Systems

- Built product clusters through DBSCAN on product visual embeddings and attributes and further merged the clusters with hierarchical approach based on customer behaviors to serve recommendations
- Trained Siamese deep and wide architecture, with Google Inception ResNet V2 and single-layer neural network as the deep and wide part respectively, to learn an embedding space through contrastive loss where cross class products going well together and matching in side-information according to customer interactions are closer to each other
- Boosted training pipeline to feed the model images and side information such as price, rating and product description text embedding by implementing batch iterator which maintains a queue for CPU workers to load, preprocess and push augmented data and GPUs to consume and decreased 11% training time on a NVIDIA P100 GPU machine
- Fine-tuned transfer learning split within Inception ResNet V2 and the model's own layers, optimized model generalization by class-weighted sampling and cyclic learning rate and coded up unit test suite
- Deployed an Airflow DAG generating embedding of Perigold catalog every week to serve cross class recommendations in carousels across the Perigold website through rapid approximate nearest neighbor search facilitated by Hierarchical Navigable Small World graph and gained 6% engagement across session traffic

CARDINAL CONSULTING LLC., Phoenix, AZ Jan 2018 - Present
Full Stack Engineer (part time), Sports Analytic Platform

- Created React components to present NBA team and player game-granular dashboards, gauges and Cardinal's featured reports and implemented Node.js server-side services to fetch statistics like points, mechanical intensity, physio load and Hollinger's PER preprocessed from PostgreSQL database
- Incorporated D3 functionality to visualize team or player's statistics through bar chart, line plot and animated bubble plot and added feature to run linear regression based on ordinary least square

Wayfair, LLC., Boston, MA Jan - July 2017
Data Scientist Co-op, Display Retargeting

- Established an Airflow DAG to fetch Wayfair Display Retargeting and Real Time Bidding cost on Facebook through Facebook Marketing API
- Independently developed a decision-making support software in Python MVC with an Elasticity-based Optimization Model integrated to recommend new bid based on customer segments and DPA tag configuration, and freed the whole display retargeting team from extremely manual operations
- Automated the bidding process for display retargeting team through Jenkins pipeline according to decision-making support software return, ran online tests on 12 DPA tags and lowered ACNR to 17% on average

ACADEMIC PROJECTS

Northeastern University Course Project, Boston, MA

Prediction on Diabetes Mellitus Onset and Patient Readmission Sept - Dec 2016

- Fixed imbalance with SMOTE plus Tomek Link and preprocessed via normalization and missing data imputation
- Ran Grid Search Cross Validation to optimize reduced dimension of PCA, encoding layer size of Auto-encoder and weight decay, Xavier-initialized and trained a Back Propagation Neural Network model with 79.81% testing accuracy to predict Diabetes Mellitus Onset
- Established four different tree-based models, optimized hyper-parameters by Grid Search Cross Validation and picked up Feature-selected Random Forest with 94.14% testing accuracy to predict diabetes patient readmission
- Analyzed model performance with confusion matrix, error bar plot, mean accuracy scatter plot and statistics including sensitivity, specificity, F1 score and precision

Everbridge Co-op Project, Boston, MA

A Regression View of Everbridge Mass Notification Service Sept - Dec 2016

- Cleaned and extracted data of notification configuration, contacts' system setting and their background followed by Exploratory Data Analysis
- Trained a Softmax Regression model to predict whether a notification will be confirmed, confirmed late or not confirmed and then tuned a Multiple Regression model to predict the time for a notification to be confirmed
- Quantified the effect of path type, number of registered contact paths, batch size, contacts' region and notification configuration in terms of final confirmation status and confirmation time to better support notification configuration decision making