Ye (Eve) Shen

115 W Squantum St. #1008, yshen61@gatech.edu
Quincy, MA, 02171 608-770-2788

Highlights

- 2nd Year in Master Program in Computer Science of Machine Learning specialization. Already got work authorization. Available Immediately.
- Silver metal in Home Credit Default Risk project in Kaggle competition.
- DCNN image recognition project: disease classification and localization in Chest X-ray images
- Database Project: creation of the GTBay website (similar to Ebay) with Postgre SQL and Java.
- Software development project: developed Cryptogram Game App in Android Studio.
- Summary of computer science and statistics courses and skills:
    - Courses: Big Data Health, Machine Learning, Machine Learning for Trading, Software Development Process, Database System Concepts & Design, Reinforcement Learning and Decision, Introductory and Intermediate Data Analysis with R, Data Visualization Analytics, Statistical Experimental Design.
    - Skills: Python (3 years), R (4 years), Java (2 years), Hadoop, Spark, Scala, Android Studio, Javascript, HTML, d3, css, Tableau, R, SQL (2 year), DOE (2 years), Matlab (2 years), Mathematica (3 years), Origin (5 years), Docker, AWS and Azure(1 year).

Education

Georgia Institute of Technology, Online, GPA: 3.80/4.00 01/2018 - 12/2019(expected)

- Online M.S. in Computer Science and Engineering in Machine Learning specialization

University of Wisconsin-Madison, USA 09/2011 - 10/2016

- Ph.D. in Materials Engineering, GPA: 3.96/4.00

Huazhong University of Science and Technology, China 09/2007 - 06/2011

- B.S. in Materials Science and Engineering, GPA: 3.90/4.00

Selected Computer Science Projects (https://github.com/blackgoldfish)

Chest X-ray image analysis with DCNN: disease classification and localization

- Incorporated the VGG/DenseNet base-model and the additional transitional/pooling/prediction layers to build the DCNN architecture to do multi-label classification and heatmap disease localization within 108,948 ChestXray images by Keras/python on AWS and Azure.
- Achieved AUC up to 0.83. Disease localization gains an accuracy up to 0.6096.

Database Project: Creation of the GTBay website with Postgre SQL and Java

- Created the SQL schemas and SQL queries to insert, update, select, delete and sort data.
- Built the user interface and enabled it to communicate with the database by Postgre SQL and Java.
- Set up the GTBay website which allowed users to sell, bid and buy items like Ebay.

Software design and engineering project: Cryptogram Game App in Android

- Developed an app for multi user to play cryptogram game.
- Performed architecture design, generated use case model, programmed through android studio and tested the functionality through test cases.

Data visualization analytics project: Interactive Map Exploration of the Gerrymandering Effects on Districts

- Preprocessed Big data (10GB) by python and storage by SQLite. Achieved Interactive visualization by Tableau.

Big data analytics in Healthcare

- Used ICU clinical data to predict the mortality of patients in one month after discharge. Obtained descriptive statistics through HIVE and transformed the data through PIG. Trained multiple logistic regression classifiers with Hadoop in parallel on Docker.

Reinforcement Learning project: landing lunar lander rocket with dueling double DQN and prioritized experience replay

- Programmed a space vehicle Lunar Lander to learn to land within a safe zone from a starting point without crashing with the dueling double Deep Q Networking algorithm by Tensorflow/Python.
- Optimized hyper parameters and improving the algorithm with Prioritized Experience Replay: the converged rewards are consistent within 100 trials and higher than the targeted goal value.

Home credit default risk project in Kaggle competition

- Achieved 0.799 ROC AUC score and won the silver metal in the Kaggle competition in classifying whether a client will repay a loan or default with python.
- Improved model performance by: adding hundreds of new features during feature engineering; finding two quite different LightGBM parameter sets to improve cross validation result; adding Xgboost and stacking with those two LightGBM results.

Professional Experience

Software Engineer, Hindsight Imaging 03/2019 – present

Individual project: Development of the Spectral-Imaging Software

- Performed UI design through wxPython and achieved the key functions of data acquisition of the 3d hypter-spectrum, equipment control, equipment calibration, data processing and interactive visualization..

Postdoc Fellow, Harvard University 03/2017 – 03/2018

High-throughput simulations and experiments to develop metallic glasses

- Applied the support vector machine (SVM) method with python to pinpoint the optimal metallic glass forming compositions within more than 100 composition candidates. This machine learning process by Python involves building of the data set, the learning process, model evaluation and model prediction. Made predictions of the good metallic glass formers on the Zr-Cu-Al alloy system, which agreed with the experimental results.

Research Assistant, University of Wisconsin-Madison 09/2011 - 03/2017

A 2_{III}^{4-1} factorial design for thermal analysis of the Aluminum-based metallic glasses

- Designed experiments with four factors, two blocks and two experimental responses.
- Analyzed the data by calculating the standard error and checking the equal variance, normality and independence assumptions through R. Identified the significant factor as "heating rate" and proposed a strategy to enhance the experimental responses.

Crystallization modeling of the Aluminum-based metallic glasses

- Modeled nucleation in aluminum amorphous alloys with R by writing functions to solve differential and integral equations and regression and optimization. Completed one high-impact paper.

Selected Publications

● **Y. Shen**, J.H. Perepezko, "The effect of minor addition of insoluble elements on transformation kinetics in amorphous Al alloys", Journal of Alloys and Compounds, 2015 Five more publications not listed.