Yu Liu

Address: 20 Summer St, Malden, MA
Phone: (857) 500-9129
Mail: liu.yu5@outlook.com
Website: yuliu.world

EDUCATION

Master of Science in Electrical and Computer Engineering 2015 - 2017
Northeastern University, Boston, MA GPA: 3.83
Background: Machine Learning, Natural Language Processing, Deep Learning, Parallel Data Processing, Database

Bachelor of Engineering in Automation 2010 - 2014
Harbin Institute of Technology, China

RESEARCH & WORK EXPERIENCES

Goodwill Computing Lab
Sep/2017 – Jan/2018

Research & Teaching Assistant, Northeastern University

- Research Project: **Large-scale Supercomputer System Log Anomaly Detection** (**pySpark, Scikit-learn, Spark MLlib**)
  - Designed and implemented **anomaly events detection** for a **supercomputer** system log (**Sandia National Lab)**
  - Cleaned and preprocessed 80G unstructured **supercomputer** system log using **Spark**, windowed data by time sequence
  - Applied **bag of words** and **word2vec** to generate word vectors for log files
  - Applied machine learning algorithms (**SVM**, **logistic regression** and **decision tree**) on time sequence windows, predicted possible anomaly events
  - Applied **data resampling** methods for data **class imbalance** problem
- **Computer Architecture** Teaching Assistant: Helped more than 50 students with office hour, course doc and grading

Schneider Electric
Feb/2017 – Aug/2017

- **R&D Coop Intern**, Device Intelligent Platform Team (**C, C++, JavaScript, HTML**)
  - Implemented a **IoT** solution for motor speed **data transmission and monitoring** in **Azure** cloud
  - Hosted a **website** in **Azure cloud** to upload and transfer lua files to local machine
  - Deployed **OpenStack** and **Apache Mesos** on multiple virtual machines to construct a distributed system

PROJECTS

Toxic Comment Classification Challenge Feb/2018 – Current
Kaggle Competition (**Python, Keras, Tensorflow, fastText, Pandas, NLTK, NumPy**)

- Developed a toxic comment classifier using **GRU recurrent neural network** and **fastText word embeddings**
- Implemented data pre-processing and text tokenization
- Implemented 10-fold **cross-validation** for model validation
- Achieved **98.38%** classification accuracy and ranked as **top 25%** among all competitors

Signal Rain Attenuation Prediction using Self-Evolving Artificial Neural Network Aug/2017 – Nov/2017
Research Project, Northeastern University & College of William and Mary (**Python, Tensorflow**)

- Optimized an **artificial neural network** architecture based on **evolutionary algorithm**
- Predicted rain attenuation of earth-space communication system signals, achieved 5% performance improvement than previous research work

A Movie Recommendation System Oct/2017 – Dec/2017
Parallel Data Processing Course Project, Northeastern University (**pySpark, Spark MLlib**)

- Applied **collaborative filtering** to build a rating system for movie recommendation based on 21,000,000 movie rating
- Evaluated the performance of parallel data processing by increasing the data partition number from 1 to 20 in **Spark clusters**

SKILLS

Programming Languages	Python, Spark, SQL, Java, C/C++, JavaScript, HTML
Software Tools	PyCharm, VMware, Git, MySQL, Eclipse, MATLAB
Cloud Tech and OS	OpenStack, Apache Mesos, AWS, Ubuntu Linux