Linxiao Bai

E-mail: lbai2@ur.rochester.edu

Website: https://pagliaccl.github.io/

Tel: 217-305-1947

Education

- **University of Rochester**, Master of Data Science, CS concentration. Sept.2016-May.2017
  GPA: 3.90 (Enrolled with 30% tuition waiver)
- **University of Illinois, Champaign**, BS of Statistics, CS focus, Distinction Honor. June.2013-Dec.2015
  GPA: 3.69 (*John E. Gieseking* Scholarship)

Work Experience

iFlytek, CP BigData Team/ Computational Advertising Team Jan.2016-Aug.2016

Data Scientist/ Software Engineer

Project 1, Anti-Fraud System:

- Designed and developed the Anti-Click-Fraud system for *iFlytek AD Exchange Platform, Demand Supply Platform*.
- Designed and implemented the first IP blacklist rules based on click time pattern analysis and Poisson hypothesis testing.
- Responsible for constructing hardware blacklist based on illegal hardware information recognition.
- Designed a series of online AB testes to validate performances of strategies.
- Composed the first Anti-Fraud system performance evaluation report.
- In the report, proposed two influential indexes that quantify the performance of the Anti-Fraud system. The indexes are still being used in the industry.

Project 2, Data Mining, Machine Learning:

- Contributed as a subject-matter expert to Data/Label mining project of *iFlytek Data Management Platform*.
- Designed GPS trajectory mining strategies with clustering algorithm. The goal is to discover user residential area.
- Responsible for constructing supervised learning algorithms used for user gender prediction and “heavy gamers” prediction.
- Constructed user-label matching strategies using keyword/NLP approaches. Audience include: traveling people, Influenza population, working class, breeding mothers, new car owners, and so on.
- Contributed to the composing of *iFlytek BigData White Book,* a thorough documentation about company’s overall data condition and possible data mining directions supported by the data possessed by the company.

Huishang Bank, Data Analytics Lab June.2015-Aug.2015

Intern Analyst

- Responsible for regular data cleaning, basic statistical analysis using R and exception handling skills.
- Frequently analyze significant data index and compose data analysis report.
- Designed a linear prediction model for recommending new financial products to targeted users.
- Responsible for analyzing factor importance in the linear model. The result of important factors is being used in the industry to help designing financial products and explaining customer interests.

Research Projects

- VisualDX anti-crawler system development. Use RNN to detect session suspiciousness.
- Electricity Abnormal Usage Detection, Feature Engineering of Time Series Data and Supervised Classification.
- Predict NYC Ground Traffic Using Cab Trajectories, A Deep Learning Application.
- American Domestic Airline Network Analysis, An Improvement of the Configuration Model.
- Change Point Detection of Probabilistic Model in Time Series Data, An Application of EM Algorithm.
  See website for paper, source code and visualization!

Technical Skills

- Python, R, Java, MySQL, Matlab, Tensorflow.
- YARN-hdfs based Spark with Scala, Python.
- Spark application optimizing. YARN resource saving.
- Spark SQL fast query, MLlib , and Zeppelin BD visualization.
- Maven project management, GitHub version control.