Jiamian Liu

jiamianliu@outlook.com

EDUCATION

Universiteit van Amsterdam & Vrije Universiteit Amsterdam **Sept. 2018 - May. 2020**
- M.Sc, Computer Science (joint degree)

Southwest Jiaotong University **Sept. 2012 - Jun. 2016**
- B.E., Telecommunication Engineering

RESEARCH EXPERIENCE

Researcher **Course Project- Data Mining** **Apr. 2019 - Present**
- Project 1: Recommended Expedia hotel groups for users based on machine learning
- Created new features (stay duration and season) and reduced destination features by PCA
- Trained model and predicted hotel group by XGBoost
- Evaluated models by cross-validation an achieved 0.41 MAP@5 (Mean Average Precision) score
- Project 2: Built a spam filter for text messages based on NLP tools and machine learning
- Optimized hyperparameters of SVM (Kernel, C, gamma) by grid search and achieved 0.92 F1 score

Researcher **Course project - Experimental Design and Data Analysis** **Mar. 2019 - Apr. 2019**
- Designed experiments and used statistical methods to analyze data (energy drink, drugs, milk, etc), including t-test, ANOVA, linear regression, logistic regression and so on (Based on R language)

Researcher **Course project - Information Visualization** **Mar. 2019 - Apr. 2019**
- Visualized Amsterdam Living Condition, including multiview design and interactive design based on D3.js
- Combined with geo map and 4 subgraphs, including bubble chart, bar chart, line chart and panoramic graph

Researcher **Course project - Developing Services for the cloud** **Jan. 2019 - Feb. 2019**
- Developed an online book searching and sharing web application based on IBM cloud
- Crawled book data from Goodreads and saved the data into Cloudant database
- Implemented back-end by Loopback framework with RESTful APIs
- Implemented front-end by Angular and Bootstrap and deployed the web application on IBM cloud

Researcher **Course project - Web Data Processing Systems** **Nov. 2018 - Dec. 2018**
- Built a web knowledge base by entity linking based on Spark, NLTK and DAS4 clusters
- Extracted raw text from HTML in WARC files by beautifulsoups and pre-processed data by Pandas
- Tokenized text and recognize named entities in the content by using NLTK and Standford NER
- Linked each entity mention to a set of candidate entities in Freebase database using Elasticsearch
- Queried candidate entities' abstract in DBpedia database using SPARQL
- Considered similarities of two entities by computing cosine similarity using Scikit-learn
- Achieved precision 0.0344, recall 0.123 and F1 score 0.0537

Researcher **Course project - Large Scale Data Engineering** **Sep. 2018 - Oct. 2018**
- Analyzed bitcoin blockchain data (120GB) to find the characteristics of criminality and visualization
- Built a high availability distributed cluster by Hadoop, Spark and Zookeeper (2 masters, 5 workers)
- Extracted criminal transaction graph using Scala on DAS4 clusters by Spark
- Visualized the criminal transaction data by dynamic and interactive force-directed graphs based on Echarts
- Classified criminal bitcoin addresses by random forest with 72% accuracy by scikit-learn

WORK EXPERIENCE

Developer **Datang Telecom Technology Co., Ltd.** **Sept. 2016 - Aug. 2018**
- Implemented partial NCT DSP compiler backend (instruction selection and register allocation)
- Supported instruction selection to NCT target-specific SelectionDAG by DAG covering algorithm
- Accelerated runtime of compiled code (3%) in kernel benchmark by implementing Wimmer Linear-Scan algorithm

Tester **Datang Telecom Technology Co., Ltd.** **Jul. 2016 - Aug. 2016**
- Built an automated test tool for the toolchain (compiler, linker and assembler) to ensure software quality
- Parallelized testing benchmarks by refactoring makefiles (reducing test time from 10h to 1h)
- Analyzed benchmarks (8,690 test cases) and auto-generate highly readable testing report with graphs and tables

SKILLS & INTERESTS

- LinkedIn/Github www.linkedin.com/in/jiamianLiu https://github.com/ljm0
- Skills Big Data, NLP, Machine Learning, Web Application, Visualization, Statistics
- Programming Proficient: Python, Shell, Makefile
  Competent: R, Javascript, C/C++; Developing proficiency: JAVA, Scala, SQL
- O/S and Tools Linux, Git, Spark, Hadoop, Pandas, scikit-learn, PyTorch, D3.js, Angular, GDB, LLVM
- Languages English: IELTS6.5, Chinese: Native Speaker
- Hobbies Reading, Cycling, Basketball, Hiking