Xxxxxx Xxx

+(86) 132xxxxx975 ⋄ xxxxxx@gmail.com
Peking University, Haidian District, Beijing 100871, China

EDUCATION

Peking University *September 2016 - present*
M.S. in Peking University, Institute of Network, GPA:xx(100)

Peking University *September 2012 - June 2016*
B.S. in Peking University, Department of Computer Science
· Total GPA: xx/4.00 Major GPA: xx/4.00

Edinburgh University *June 2015 - September 2015*
Visiting Scholar in School of Informatics

HONORS / AWARDS

NOI – First Prize *July 2011*
ACM/ACPC Tianjin regional competition – Second Prize *September 2012*
Second Prize in ACM Programming Contest (on campus) *May 2013*
Third Prize in ACM Programming Contest (on campus) *May 2014*
Winner in Peking University AI Against Competition *June 2015*
Runner-up in SIGMOD2017 Programming Contest *March 2017*
54-Youth Scholarship *2014*

WORKING EXPERIENCE

Microsoft STC Asia *September 2017 - February 2018*
Internship, Department of Xiaoice
Responsible for exploring new features in QA process and some other system coding tasks in C#.
- Implementation of CNN network in C# which 2 times faster than original version under Intel-MKL Library and 40 times faster than original version under memory cache.
- Finding a new approach to get the matched sentence in an article with its comment.
- Implementing some pipelines deal with databases and ElasticSearch servers.

Bytedance Inc. Beijing *June 2017 - August 2017*
Internship, Department of AI-Lab
Responsible for maintaining the server side of video deduplication under Thrift-Framework in C++.
- Add other features in video duduplication process, such as deep-learning feature, orb feature and k-means based feature.
- Implementation of a web page tool to label the videos manually.
- Tasks of data processing.

Megvii Inc. Beijing *March 2016 - September 2016*
Internship, Team of Text Recognization
Responsible for maintaining the pipeline of using the model to recognize all characters of a picture in C++.
- Implementation of minAreaRect with C++.
- Implementation of text in lines with minimum-cost flow.
- Responsible for adjusting pipeline according to different models, and displaying demo on website.

PROJECTS / RESEARCH EXPERIENCE

Automatic Hyperparameter Tuning *December 2017 - January 2018*
Implementation of a robust Bayesian Optimization framework written in Java. About 2000 lines.
Github: `https://github.com/xxxxxxx`
- Model: Gaussian Process.

- Maximizer: Random Sampling.
- Acquisition Function: EI, LCB, GP-UCB

SIGMOD 2017 Programming Contest *January 2017 – March 2017*

2-people Semi-final Rank(1/40). Final Rank(2/40)

Given the n-grams, documents and many remove and add operations, the task is finding all n-grams in each document. The contest is ranked by time.

- Using Hash Table and Trie to make index.
- Achieving complete parallel to insert, delete and query operations.
- Codes: about 3000 lines in C++.

Alibaba Big Data Contest *October 2016 - December 2016*

Prediction of offline o2o coupon tickets using Ranking in preliminary(1/1500). Ranking in semi-final(19/1500)

Three-people team, responsible for the choice of features and models, data cleaning, etc.

- Learning Ali Shujia in semi-final, writing MR process to extract features with java.
- The model of LR, Random Forest, GBDT and fusion of various models are tried.
- Temporal features were added and AUC raised by 1%.

3D Online Mobile Game *July 2016 – August 2016*

Independent development

A MOBA game on Android, each game supports the maximum of 4 people battling together. The game supports players combating according to capability level, provides user login system, and stores personal information in the server.

- The game logic part uses the Unity3D engine with C#.
- Server side is rented Ali Cloud Server, using Django framework written, Apache deployment.
- Codes: about 7000 lines.

Java Comment Matchment *July 2015 – August 2015*

Exchanging project to Edinburgh University 2 months

Using deep learning to match java code blocks and comment blocks with professor Charles Sutton. The ultimate goal is to achieve code completion automatically

- The dataset is the java project on Github.
- Writing java tokenizer to extract java code token, separating code blocks and comment blocks.
- Making training data: experimenting various rules to match the code blocks with the comment blocks.

SKILLS

- Programming language: proficient in C/C++, C#, java and python, having knowledge of shell, matlab and some functional programming languages like scala and scheme
- Big data: Familiar with Hadoop and Spark.
- Machine learning proficient in numpy, matplotlib, sklearn and xgboost, having knowledge of pandas and tensorflow

OTHERS

- Maths Level: A and A+ in Mathematic Analysis, Advanced Algebra, Discrete Mathematics.
- English Level: CET-6.
- GitHub: https://github.com/xxxxxxx