

Do everything seriously

Educational Background

2008 – 2011 **Master**, *Zhejiang University*, Computer Application Technology.

Research Area: Large scale online learning, text classification, collaborative filtering algorithm.

2004 – 2008 **Bachelor**, *Jiangsu University of Science and Technology*, Information and Computing Sciences.

Work Background

2012.9 – **Software Develop Engineer**, *Microsoft STC*, Beijing.
Campaign recommendation, bidding and budget suggestion.

2011.4 – 2012.8 **Applied Research Engineer**, *Netease Youdao*, Beijing.
Crawler, webpage parser and analyzer, ranking.

Technical Ability

Programming Language Java, Python, C, Scala, C#, Matlab, SQL.

Database Technology SQL Server, MongoDB, Kyoto Cabinet, Self implemented DHT.

Profiling Algorithm profiling, service profiling.

Large-Scale Data Processing Experienced with big data processing technology, familiar with several stacks including Cowork/ODFS/Omap in Youdao and Cosmos in Microsoft.

Data Mining Familiar with common models and algorithms of Data Mining, has in-depth study of online learning, text classification and collaborative filtering algorithms.

English 471 points in CET-6, good English reading and speaking ability.

Math Solid math foundation.

Projects

Product Recommendation Service

Introduction Recommend inventories to advertisers in display advertising business.

Timeline 2012.9 – 2013.4

Duty Experiments and algorithms implementation including content-based and collaborative filtering algorithms, and common utility such as logging and profiling.

Results The feedbacks from users are mostly positive, though lack of instrumentation data.

Timeliness Web Crawler

Introduction Crawl news from portal sites, forums, microblogs with low latency.

Timeline 2011.10 – 2012.4

Duty Project leader with three team members.

Results The timeliness coverage rate has been catching up with Baidu.

Efficiency optimization of DNS Resolve

- Introduction The DNS lookup rate was becoming the bottleneck of main crawler's efficiency and we try to speedup it.
- Timeline 2011.4 – 2011.6
- Duty Project leader and the only programmer.
- Results The rate of DNS resolving has been sped up by an order of magnitude.

The Application of incremental SVMs in Large-Scale Email Spam Filtering

- Introduction National 863 Project
- Timeline 2008 – 2009
- Duty Project leader, I have built the architecture of our system and implemented most classifiers.
- Results The spam filtering system named TGANG is in a leading position internationally both in accuracy and efficiency.

Terminator

- Introduction Individual projects, a two-layer ensemble spam filter.
- url <https://code.google.com/p/terminator/>
- Timeline 2009 – 2010
- Results Ensemble 8 advanced classifiers, can achieve the best results on *All* public datasets.

Experience

Scholarship

- 2009 Second Prize Scholarship of Zhejiang University.

Open Source

- 2010 Wrote the NSNB spam filter in Python, which won the first place in large-scale spam filtering competition of Eights Symposium of Search Engine and Web Mining. Now NSNB is open sourced at <http://code.google.com/p/nsnb>.

Publications

- O Baojun Su, Congfu Xu. Not so naïve online Bayesian spam filter. In: Proceedings of the 21st conference on Innovative Application of Artificial Intelligence (IAAI 2009), July 14-16, 2009, Pasadena, CA, pages 147-152.
- O Congfu Xu, Chunliang Hao, Baojun Su. Research on Markov logic networks. Chinese Journal of Software, 2011, 22(8): 1699-1713. (In Chinese with English abstract)

Self Assessment

I am a quick learner and nice team member, with rich developing experience and good understanding of pattern recognition and machine learning.