

Do everything seriously

Educational Background

2008 – 2011 **Master**, *Zhejiang University*, Computer Application Technology.
Research Area: Large scale online learning/text classification/recommender system.
2004 – 2008 **Bachelor**, *Jiangsu University of Science and Technology*, Information and Computing Sciences.

Work background

2011 – **Applied Research Engineer**, *Netease Youdao*, Beijing.
crawler/parser/ranking.

Technical Ability

Programming Language Java, Python, C, Matlab.
Operating System Linux/Unix, including commands, toolchain, scripts.
Profiling Algorithm profiling, system profiling.
Large-Scale Data Processing Youdao's Map-Reduce/Big Table implementation - Cowork, ODFS etc.
Data Mining Familiar with common models and algorithms of Data Mining, has in-depth study of online learning, text classification and recommender system.
English 471 points in CET-6, good English reading and speaking ability.
Math Solid math foundation.

Projects

Timeliness Web Crawler

Introduction Crawl news from portal sites, forums, microblogs with low latency.
Timeline 2011.10 – 2012.4
Duty Project leader and the only programmer.
Results The rate of timeliness coverage is catching up with Baidu.

Efficiency optimization of DNS Resolve

Introduction The DNS refresh rate is not enough.
Timeline 2011.4 – 2011.6
Duty Project leader and the only programmer.
Results DNS resolve is speeding 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 built the architecture of our TGANG system and implemented most classifiers.
Results The TGANG system 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 Eighth Symposium of Search Engine and Web Mining. Now NSNB is open sourced at <http://code.google.com/p/nsnb>.

Publications

- Baojun Su, Congfu Xu. Not so naive 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.
- 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, with rich developing experience and good understanding of pattern recognition and machine learning.