Baojun Su

✆ +86 18625240866
✉ freizsu@gmail.com

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

2017.2 – now	**Sr. Software Engineer**, *Microsoft*, Suzhou. Campaign management platform, tech lead
2015.2 – 2017.1	**Software Engineer II**, *Microsoft*, Suzhou. Campaign management platform, tech lead
2013.11 – 2015.1	**Software Engineer**, *Microsoft*, Suzhou. Distributed key-value datastore
2012.9 – 2013.10	**Software Engineer**, *Microsoft*, Beijing. Ad recommandation
2011.4 – 2012.8	**Applied Research Engineer**, *Netease Youdao*, Beijing. Crawler, web page parser and analyzer, search ranking.

Technical Ability

Programming Language	Java, Python, C, C#, Matlab, SQL.
Database Technology	Microsoft SQL Server, Mongodb, self implemented distributed key-value datastore.
Machine Learning	Familiar with common models and algorithms of machine learning and data mining, has in-depth study of online learning, text classification, collaborative filtering, and deep learning.
Performance Tuning	Expert in Algorithm performance tuning, service performance turning.
English	471 points in CET-6, good English reading and speaking ability, over 5 years experience using English as primary language at work.

Projects

AdExtension Framework

Introduction	A dynamic framework to speed up developing new extension types for Bing Ads campaign platform.
Timeline	2015.8 – 2016.3

Duty	System design, implementation, tech lead.
Results	We implemented dynamic logic using reflection in every components of the service. With static code generation tool, developing new extension type can be as easy as writing a new C# class. The development time consumption can be reduced from 1/2 year to 1 week.

Distributed Key-Value Datastore

Introduction	Online serving datastore, serving TB data with low latency and can scale out horizontally, with the ability to plugin user logic as SPROC in C#.
Timeline	2013.10 − 2014.9
Duty	System design, implementation, performance tuning.
Results	After migrated the backend of a production service from SQL Server to the new data store, the .95 has been reduced to 1/10 with the same request, and at the same time increased the flexibility and the readability of SPROC.

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, architecture design and implementation.
Results	The solution is we first mine the page library to identify possible seed pages, and calculate an estimated refresh interval per seed page then setup a fast flow to crawl the seed page set according to the refresh interval. The timeliness coverage rate has been catching up with Baidu after this system going online, compared to nearly no timely results before.

Efficiency Optimization of DNS Resolve

Introduction	The DNS lookup latency has been becoming the bottleneck of main crawler's efficiency and we try to speedup it.
Timeline	2011.4 − 2011.6
Duty	Performance optimization.
Results	Upgraded the DNS query logic from multi-thread synchornize call to multi-thread asynchornize call. The DNS lookup job latency has been reduced to less than 1/10 than before.

GoingMerry (afterhours)

Introduction	A strong programming go (baduk) engine after AlphaGo Zero.
Timeline	2018.1 − now
Results	It's ongoing developed, the aim is based on AlphaGo Zero, try to find a lightweight solution to avoid huge reinforcement learning computation while have similar or above strength. The latest strength is around amateur 6D on KGS using 1600 playouts.

Terminator (school)

Introduction	A promising spam filtering library making use of combined machine learnaing algorithms, open sourced at github.
Url	https://github.com/freiz/terminator
Timeline	2009 − 2010

Results Implemented an advanced ensemble algorithm and ensemble 8 machine learning algorithms, can achieve the best results on *All* public datasets, with small memory footprint (default 5 MB) and fast (train/predict hundreds mails in 1 second with 1 thread on normal PC). It won the first place in large-scale spam filtering competition of Eights Symposium of Search Engine and Web Mining and derived two research papers.

Publications

- Congfu Xu, Baojun Su, Yunbiao Cheng, Weike Pan, Li Chen, "An Adaptive Fusion Algorithm for Spam Detection", IEEE Intelligent Systems, vol.29, no. 4, pp. 2-8, July-Aug. 2014, doi:10.1109/MIS.2013.54
- 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.
- 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)