苏保君

+86(512)18625240866
freizsu@gmail.com

认真做好每一件事

教育背景

2008 – 2011 **硕士**, 浙江大学, 计算机应用技术.
研究方向: 大规模在线学习、大规模文本分类、推荐系统
2004 – 2008 **学士**, 江苏科技大学, 信息与计算科学.

工作背景

2012.9 – **软件开发工程师**, 微软亚洲搜索技术研发中心, 苏州.
分布式键值数据存储、推荐系统

2011.4 – 2012.8 **应用研究工程师**, 网易有道, 北京.
网页搜索的抓取、解析以及购物搜索的排序

技能

编程语言 Java, Python, C, C#, Matlab, SQL
数据库 SQL Server, Mongodb, Kyoto Cabinet, 分布式键值数据存储
调优 擅长算法调优、系统调优
大规模数据处理 熟悉大规模数据处理技术，包括有道的 CoWork、ODFS、OMap 等以及微软的 Cosmos
数据挖掘 熟悉机器学习、数据挖掘中的常用模型和算法，对在线学习、文本分类和推荐系统方面有深入研究
英语 六级 471 分，具有良好的英文文档阅读及口语表达能力

项目

分布式键值数据存储

介绍 分布式键值数据存储，在线服务 TB 级数据同时具有较小的延时，可插入 SPROC 执行逻辑
时间 2013.10 – 2014.9
职责 系统设计、部分实现、调优
效果 对比之前线上 SQL Server 系统，同样的请求 .95 减小了 3-10 倍，同时提高了灵活性和 SPROC 的可读性

广告推荐系统

介绍 在展示广告系统中为广告主推荐广告位
时间 2012.9 – 2013.4
职责 协同过滤及基于语义的推荐算法设计实现，以及评估算法的设计实现
效果 得到了大部分用户的好评，准确率从 66% 提升到了 73%

时效性网页抓取

介绍 分钟级的延迟抓取新闻、论坛、微博等时效性信息

时间　2011.10 – 2012.4

职责　系统设计、调优

效果　有道的时效性结果从上线前的几乎无结果到上线后后台覆盖率接近百度的水平

DNS 查询效率优化

介绍　优化前 DNS 解析吞吐量太小，不能满足大规模抓取需要

时间　2011.4 – 2011.6

职责　优化 DNS 系统查询效率

效果　通过将底层 DNS 解析从同步 IO 改为异步 IO，查询效率提高了 10 倍

Terminator

介绍　个人项目，两层集成垃圾邮件过滤器

项目地址　`https://github.com/freiz/terminator`

时间　2009 – 2010

效果　实现了自创的在线集成算法，集成了 8 个优秀分类器，在所有公开数据集上都取得最优结果，被某 outlook 插件作为算法核心

经历

2010　使用 Python 独立开发了 NSNB 邮件过滤算法，并以此获得了 SEWM2010 (第八届全国搜索引擎和网上信息挖掘学术研讨会) 大规模垃圾邮件过滤比赛综合第一名，NSNB 项目目前开源，项目地址是: `http://code.google.com/p/nsnb/`

2010　所开发的《基于内容的多分类器多层垃圾邮件实时过滤系统软件》获得了计算机软件著作版权登记证书

发表论文

- 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)