

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: 大规模在线计算、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

2011 – 应用研究工程师, 网易有道, 北京.
主要负责网页搜索的抓取、解析以及购物搜索的排序

技能

编程语言 Java, Python, C, Matlab
操作系统 Linux/Unix, 包括系统命令、开发工具、脚本编写
调优 擅长算法调优、系统调优
大规模数据处理 有道开发的并行计算系统, cwork, odfs 等 (类似 hadhoop, hbase)
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
数学 数学基本功扎实

项目

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 项目负责人及唯一开发人员
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 每天解析量太小, 不能满足大规模抓取需要
时间 2011.4 – 2011.6
职责 项目负责人及唯一开发人员
效果 将底层 dns 解析从同步改为异步, 查询效率提升了一个数量级

集成化 SVM 增量技术及其在大规模垃圾邮件过滤中的应用

介绍 隶属于国家 863 计划智能感知与先进计算专题, 侧重于工程实现

时间 2008 – 2009
职责 项目负责人
效果 负责设计 TGANG 邮件过滤器系统架构并实现了大部分在线分类器, TGANG 系统无论从过滤准确率还是运行效率在国际上都处于领先地位

Terminator

介绍 个人项目, 两层集成垃圾邮件过滤器
项目地址 <https://code.google.com/p/terminator/>
时间 2009 – 2010
效果 集成了 8 个优秀分类器, 在所有数据集上都能得到最优结果, 被某 outlook 插件作为算法核心

经历

2009 校优秀研究生二等奖
2010 使用 Python 独立开发了 nsnb 邮件过滤系统, 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权登记证书
2011 网易有道网页搜索平台组季度标兵

发表论文

- 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)

自我评价

- 学习能力强, 做事认真细心
- 写程序、理论研究两方面都能独当一面
- 具有良好的团队协作能力