

简 历

个人资料

姓名：杨鹏
出生日期：1987-01
性别：男
学历：本科 (2006 – 2010, 石家庄经济学院)
专业：计算机科学与技术
电话：15652301070
E-mail: qiwulun@yeah.net

专业能力

熟悉机器学习理论及相关算法，熟悉数据挖掘算法，熟悉NLP相关算法；
对推荐系统有深入研究；
精通c，熟悉c++，精通shell，Python，熟悉R语言；
精通Linux环境编程，熟悉部分Linux内核代码；
熟悉mpi及相关的并行算法；
熟悉Hadoop，对NoSQL数据库有了解，熟悉MongoDB，熟悉Redis源码；

工作经历

2010/03 – 2011/03：
国和信诚信息安全技术服务中心 | 研发部 | 算法工程师
实现以机器学习算法为核心的检测系统，发现网络中僵尸主机，并监控流量。

2011/04 – 2012/03：
北龙中网科技有限责任公司 | 技术部 | 软件工程师
高性能DNS服务器的设计和研发。自动化测试工具开发。

2012/03 – 至今：
58同城 | 数据组 | 算法工程师
数据挖掘，个性化推荐算法研发。

项目

2010/03 – 2011/03：基于机器学习的botnet检测
开发工具：Weka, java, Python
责任描述：使用Weka开源数据挖掘工具对数据分析，选择合适的分类算法，调整数据集及属性提高正确率。部分模块编码。
项目描述：使用开源的入侵检测系统Bro抓取网络中的数据流量，使用分类算法发现网络中的僵尸主机，选择合适的机器学习算法及数据属性尽量提高检测率。

2010/05 – 2010/08：开源库broccoli修改

开发工具: c

责任描述: 修改开源入侵检测客户端库broccoli为web服务, 并增加并发处理功能。

项目描述: Bro是一个开源入侵检测系统, 提供了一个c语言库broccoli请求bro事件, 但是在后续处理过程中用Python开发, 因此需要把broccoli改成一个web服务。当Python程序需要bro的某个事件时, 则向web服务申请事件。web服务收到这个请求时则向bro申请事件。一旦bro产生这个事件, web服务负责转发到客户端。

2011/04 - 2011/08: DNS递归服务器开发

开发工具: c

责任描述: 架构设计, 核心模块编码。

项目描述: 开发类似于opendns, 支持客户自定义规则的dns递归服务器。初期方案是完整开发一个能够满足需求的dns服务器, 由于时间和难度的原因, 最终的方案定为使用bind做后台服务器, 使用redis存储用户数据和规则。

2011/09 - 2012/01: DNS自动测试工具开发

开发工具: Python

责任描述: 设计测试工具框架, 部分模块编码。

项目描述: 使用Python开发dns自动测试工具, 测试包括性能, 稳定性, 正确性等。

2012/03 - : 数据整理和提取

开发工具: Hadoop, HDFS, Python

责任描述: 为数据分析提供各种格式数据。

项目描述: 写mapreduce job从原始数据中提取转化成其它格式, 以满足数据分析的需求。

2012/03 - : 团购推荐系统研发

开发工具: mpi, Hadoop, R

责任描述: 推荐算法研究和推荐系统架构设计。

项目描述: 团购商品周期短导致现有的推荐算法不能取得很好的效果, 经过不断的研究测试, 最后采用youtube论文中提到的item-based cf算法, 以及常用的user-based cf算法作为补充。在实现user-based cf时, 在机器资源有限且计算量很大的情况下, 使用mpi, Hadoop做并行计算, 并使用WAND算法大大缩短计算时间。在最后的过滤阶段使用了ctr预测等机器学习相关技术。

其它

twitter: @yp

weibo: @yphoho