

“Stay hungry, Stay foolish.”
- Steve Jobs

Skills

Language	Python = Perl = Shell > Java > C++, 扎实的算法基础.
Machine Learning	熟悉推荐系统架构, 熟悉CF、SVD、LDA、pLSA、LR等常用机器学习算法.
Hadoop	熟练掌握MapReduce编程, 了解Hadoop运行机制及架构, 熟悉HSQL, 了解HBase.
Data Warsehouse	精通数据仓库建模过程, 熟悉使用Pentaho BI套件, 熟练掌握网站分析及数据分析的常见方法.
English	能用英语交流, 熟练阅读英文技术资料.

Work Experience

2010 - 现在 风行视频技术有限公司, 数据挖掘工程师.
数据仓库、推荐系统的开发

Manage Experience

2012 - 2013 敏捷开发、Scrum, 数据BI团队的Scrum master, 负责召开各种需求评审会、迭代会、回顾会等, 带领团队完成BI组各类项目.

2012 - 2013 研发主管, 负责BI组的人员招聘、绩效考核、团队建设、人员培养等工作.

Program Experience

2013 至今 LR模型对推荐结果二次排序, LR模型、二次排序.
采用用户反馈数据(隐式(观看、点击、浏览)+显式(顶、踩、打分等))进行LR建模, 对CF的推荐结果进行二次排序, 优化效果

2013 UGC视频聚类, C++、Python、Openmp(并行库).
采用的pLSA聚类, 该算法主要解决了并行化问题, 之前基于LDA每次训练电影要花时间大约5-10个小时, 现在速度提高到5-10分钟. pLSA主要用来是对小视频的剧情、描述、标题切词, 然后训练每个小视频在主题上的分布, 之后通过相似度计算聚类电影.

2013 标签数据建模, 标签数据、模型融合、Python.
通过抓取外部网站的标签数据, 对媒体的标签数据进行建模, 计算相似度, 用于解决媒体冷启动问题.

2013 通用BI框架, 反射, IOC, Hadoop.
通过反射和IOC技术, 编写通用BI框架. 从而只需要通过前端选择原始日志和各数据字段需要的操作得完成ETL, 并可通过前端选择相关统计按钮得到各类统计结果.

2013 BIEE数据仓库迁移, MapReduce、Kettle、BIEE.
将数据仓库从原来的单机使用脚本及Infobright计算, 迁移至Hadoop使用MapReduce, 以提高系统的可扩展性和稳定性. 同时建模和前端使用BIEE, 数据库采用Oracle. 数据连接及导入使用Kettle, 通过数据分析与管理平台进行任务调度和管理.

- 2012 - 2013 数据分析与管理平台, Hadoop、Hive、PHP、MySQL、Python.
主要使用PHP+Python+MySQL实现一个统一管理风行所有上报数据、ETL数据以及调度任务的系统。目标是通过该系统, 方便业务数据有管理、任务调度, 并能开放整个Hadoop计算资源和数据资源给其他部门, 降低各部门使用数据的门槛。
- 2011 - 2012 Pentaho BI系统, Kettle、Pentaho BI、Infobright.
使用Kettle对原始数据进行ETL, 使用Infobright+Pentaho BI搭建数据仓库, 通过使用Pentaho Schema Workbench进行建模, 统计得到需要的跟踪的指标数据。并研究使用Pentaho Dashboard创建仪表盘, 增强界面展现。

Awards

- 2012 风行月度之星优秀员工.

Education

- 2007 - 2010 硕士, 北京工业大学, 计算机软件与理论.
2003 - 2007 本科, 南昌大学, 计算机科学与技术.