

# 江勇

手机: (+86)15510130056

邮箱: jiangyongplus@163.com

## 教育背景

|                   |          |          |    |          |
|-------------------|----------|----------|----|----------|
| 2013.09 – 2016.07 | 中国科学院软件所 | 软件工程     | 硕士 |          |
| 2009.08 – 2013.07 | 中国科学技术大学 | 计算机科学与技术 | 本科 | 排名: 4/52 |

## 获奖情况

|                      |                                   |
|----------------------|-----------------------------------|
| 2009-2010            | 中国科技大学优秀学生奖学金铜奖                   |
| 2010-2011            | 中国科技大学优秀学生奖学金金奖, 国家励志奖学金, 腾讯奖学金   |
| 2011-2012            | 国家励志奖学金, 中国科技大学 Robogame 机器人大赛优秀奖 |
| 2012-2013            | 中国科技大学三好学生(10%)                   |
| 2013-2014, 2014-2015 | 中国科学院三好学生(10%, 两次)                |

## 项目经验

### 2014.08 – 至今      基于内存数据网格的数据访问加速引擎      核心研发人员

**项目描述:** 基于内存数据网格开源产品 Hazelcast, 实现一款低延时、高扩展性、数据模型灵活的数据访问加速引擎 (EasyCache); 在以 TPC-W 为基准的测试中, EasyCache 相比 Mysql, Oracle 等数据库, 性能提升近 200 倍。

- ◆ **数据加载:** 利用 cglib 动态构造类, 实现关系型数据到 Key-Value 数据的自动转化, 并通过不同粒度并发和自定义数据加载的方式, 进一步提升数据加载的性能。
- ◆ **数据访问:** 自主实现 JDBC 驱动, SQL 解析模块和 SQL 执行模块, 可通过标准 JDBC 接口和 SQL 语句透明访问内存数据网格中的数据; 同时, 在 JDBC 驱动实现查询结果缓存、SQL 解析缓存和序列化优化 (单结点利用深拷贝代替 Java 序列化, 多结点利用 Kryo 序列化), 进一步提升数据访问性能。
- ◆ **数据持久化:** 利用 Java 反射技术, 实现数据从内存数据网格到关系型数据库的持久化, 保障数据的一致性和可靠性, 持久化过程可采用同步、异步和批处理的方式。

**关键技术:** Java, Hazelcast, Cglib, JDBC, JsqlParser, Thread Pool, Distributed Map, Serialization

### 2013.10 – 2014.01      新闻搜索引擎      核心开发人员

**项目描述:** 从 2-3 个新闻网站采集 10 万网页, 通过对这些网站新闻信息的抽取、索引和检索, 实现一个新闻搜索引擎, 查询结果能按相关度进行排序。

**主要贡献:** 网络爬虫模块, 创建倒排索引, 新闻的检索、排序。

**关键技术:** Java, HttpClient, HtmlParser, Jsoup, Ansj, tf, idf, PriorityQueue, 倒排索引, 余弦相似度

## 实习经历

### 2014.03 – 2014.07      机器学习研发实习生      百度网页搜索部 LTR 小组

➤ **通用模型训练平台的设计与开发:**

**主要贡献:** 精简 Ltrtools 工具包, 开发 Python 和 Shell 脚本解析用户配置, 动态执行对应的工具流程。

**关键技术:** Vim, Python, Shell, Linux

➤ **Pimp 学习:**

**项目描述:** 利用机器学习+pairwise 方法学习页面中词项权重。

**主要贡献:** 从 100G 的查询日志中提取训练样本, 通过对特征的选择、增加和组合, 训练 Logistic 回归模型, 页面中词项权重预测的正确率由 66.9% 提升到 88.6%。

**关键技术:** LTR, Hadoop Streaming, Logistic, 样本提取, 特征选择

## 专业技能

- ✧ 熟悉 Java, 掌握常用的数据结构和算法, 了解 C/C++, Python, PHP
- ✧ 对分布式系统, 搜索引擎和常用的机器学习算法有一定了解
- ✧ 熟悉 Eclipse 开发环境, Linux 工作环境, SVN 版本控制
- ✧ 通过英语六级, 具有快速查阅专业英文文献的能力