Jianpeng Hou

Machine Learning; Software Development

✉ houjp1992@gmail.com · ☎ (+86) 152-0000-0000 · ⚲ https://github.com/houjp/

🎓 Education

Institute of Computer Technology, Chinese Academy of Sciences   Sep. 2014 – Jul. 2017

Master student in Computer Software & Theory, Rank: top 5%

University of Science & Technology Beijing   Sep. 2010 – Jul. 2014

Bachelor student in Computer Science & Technology, Rank: 4th/124

👥 Project Experience

Big Data Analysis Platform (http://159.226.40.104:18080)   Oct. 2015 – Mar. 2016

- Developed distributed algorithms(CART/GBDT/GBRT/RF) on Spark.
- Finished data mining components(Feature-Indexing/Feature-Merging/Feature-Normalization/Scoring).

China Telecom Big Data Application Contest (1st/1112; Team Leader)   Dec. 2015 – Mar. 2016

The goal of this contest is to predict views of users with ten sites succeeding, according to four hundred million user-behavior historical records(25.38G).

- Proposed and implemented a multi-target regression algorithm on Spark. Optimized F1-score 0.8%.
- Designed and developed a probability ranking model for user classification. Optimized F1-score 0.6%.

SIGHAN-2015 Chinese Spelling Check Task (1st Place)   Mar. 2015 – May. 2015

The goal of this task is to detect and correct spelling errors on Chinese essays.

- Handled this task with a unified framework which consisted of candidate generating, two stage candidates re-ranking and global decision making.
- Finished candidate generating model and two stage candidates re-ranking model. Optimized F1-score 18%.

🗠 Intern Experience

DiDi Research Institute   Jul. 2016 – Sep. 2016

- Developed taxi dispatching system which used to balance the supply and demand between urban areas.
- Reconstructed the log system of dispatching based on Kafka and MySQL.

Baidu Online Network Technology(Beijing) Co.,Ltd   Dec. 2013 – May. 2014

- Completed the multi-threaded development of LiveWDBBroom, improved the efficiency by two times.
- Completed the development of the Problems-Tracing Platform which was put into service.
- Passed the examination of Code Master(C++), certified as Good Coder.

🏆 Academic Competitions

1st Place Awarded in China Telecom Big Data Application Contest	Mar. 2016
1st Place Awarded in SIGHAN-2015 Chinese Spelling Check Task	Jun. 2015
1st Place Awarded in China College Students Computer Games Competition	Nov. 2013
1st Prize Awarded in China College Students Computer Application Contest	Nov. 2013
1st Prize Awarded in "LanQiao Cup" Software Development Contest(Beijing Division)	Apr. 2012
Silver Medal (17th / 200) Awarded in ACM-ICPC Asia Beijing Regional Contest	Nov. 2015

⚙ Skills

- Skilled in C++, Scala, Shell. Familiar with data structures and algorithms and had good programming style.
- Experienced in development of distributed machine learning algorithms.
- Strong theoretic knowledge on data mining and machine learning.

<h1 align="center">侯建鹏</h1>

<p align="center">求职意向: 机器学习 | 软件研发</p>

<p align="center">✉ houjp1992@gmail.com · ☎ (+86) 152-0000-0000 · ⚭ https://github.com/houjp/</p>

🎓 教育背景

中国科学院计算技术研究所 (保送)　硕士, 计算机软件与理论, 排名: 5%　　2014 – 2017

北京科技大学　学士, 计算机科学与技术, 排名: 4th/124　　2010 – 2014

👥 项目经历

BDA 大数据分析系统 (http://159.226.40.104:18080)　　2015.10 – 2016.03

- 作为 BDA Lib 核心开发者, 在分布式计算框架 Spark 上实现了 CART/GBDT/GBRT/RF 算法的分布式版本, 同时完成了这些算法单机版本的开发。
- 实现了 BDA Studio 中需要的数据挖掘基本组件 (包括特征索引、特征合并、特征选择、特征值归一化、模型输出结果评分等功能), 并构建真实场景的应用实例。

中国电信大数据竞赛 (1st/1112; CNY 200,000; 队长)　　2015.12 – 2016.03

根据海量用户行为数据 (大于四亿条,25.38G), 预测用户接下来一周对十个视频网站的访问量。

- 基于 Spark 分布式计算框架, 提出并实现了 MTGBM 多目标回归算法, 使得 F1 值提升 0.8%。
- 提出并实现了用于用户分类的概率排序模型, 使得 F1 值提升 0.6%。

SIGHAN-2015 Chinese Spelling Check Task　　2015.03 – 2015.05

该评测任务是对繁体中文进行拼写检查, 给出正确的拼写结果。

- 实现候选生成排序模型: 根据同音、近音、形近字, 为句中每一个繁体字生成候选, 并打分排序。
- 实现两轮候选重排序模型: 采用简单特征进行预排序, 在第一次排序的基础上结合复杂特征进行第二次排序, 以此提高排序效率。最终取得第一名, F 值超出第二名 18%。

🔧 实习经历

滴滴研究院 智能平台部 / 算法工程师　　2016.07 – 2016.09

- 完成专车调度系统开发、测试并上线, 实现对城市各区域专车供需不平衡状态的动态调整。
- 基于 Kafka 及 MySQL 完成专车调度日志系统重构, 解耦子模块间相互依赖, 提高系统可用性。

百度 网页搜索部 / 研发工程师　　2013.12 – 2014.05

- 完成 livewdbbroom 工具多线程开发, 通过测试并上线, 效率提升 2 倍以上。
- 完成离线时效性日志体系建设及离线时效性问题追查平台建设, 并上线。
- 通过 Code Master(C++) 考试, 获得 Good Coder 认证。

🏆 学术竞赛

冠军 (1st/1112; CNY 200,000; 队长) 中国电信大数据算法应用大赛	2016.03
冠军 SIGHAN-2015 Chinese Spelling Check Task	2015.06
冠军 全国大学生计算机博弈大赛	2013.08
银牌 (17th/200) ACM-ICPC 国际大学生程序设计竞赛亚洲区域赛北京赛区	2015.11
四强 (4th/2293) 阿里巴巴大数据竞赛 "新浪微博互动预测大赛"	2015.12
一等奖 华北五省及港澳台大学生计算机应用大赛	2013.11
一等奖 全国软件专业人才设计与创业大赛北京赛区	2012.04

⚙ 个人能力

- 熟悉 Scala、C++、Shell, 熟悉基本数据结构和算法, 有良好的编程风格。
- 有丰富的基于 Spark 分布式计算框架的机器学习算法开发经验。
- 熟悉数据挖掘、机器学习领域基本算法。