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.

夏忠林

求职意向: 机器学习 | 广告算法

tinymindx@hotmail.com · (+86) 186-1126-9795 · http://frozenxia.github.io/

教育背景

武汉大学 硕士，计算机应用与技术 2012 – 2015

武汉大学 学士，计算机科学与技术 2008 – 2012

项目经历

小米 互联网商业部 / 金融策略组 2019.02 – 2019.10

研发针对互联网金融用户的人群定向模型，提高激活率和注册率；为金融应用研发召回策略算法，根据不同广告位的联合回测结果, 通过随即初始启发式搜索方法，在额定 ROI 下最优化平台收益, 该人群定向模型和召回策略算法已经投入使用。

小米 云平台 / 采风质量预警系统 2017.10 – 2019.01

以售后、客服以及销售数据为基础研发预警算法，对产品质量问题进行快速预警；利用爬虫获取的用户评论数据，建立产品满意度指数。

- 实现了多维序列预测算法，该算法与 LSTM 模型以及 Prophet 相比, 对产品销量和故障数据的预测结果误差和方差更小。
- 实现了异常检测算法，该算法比 AnomalyDetection(https://github.com/twitter/AnomalyDetection) 具有更好的检测效果。
- 实现了多种情感分类算法 (基于概率图模型和基于神经网络模型)，对用户评论的情感分类准确率超过 0.92。
- 该项目获得了 2018 小米质量奖三等奖。

小米 云平台 / 用户画像表维护 2017.08 – 2017.10

维护公司用户画像宽表，梳理上下游数据链路，提高数据质量和可用性。经过优化之后的数据可用性从 90% 提高到 99.8%，数据生成时间减少 65%。

小米 云平台 / Faas 系统 (https://open.cloud.mi.com) 2017.03 – 2017.08

为小米开放云设计并实现了基于 Kubernetes 的 Serverless 框架 Faas，该系统支撑了小爱同学的语音 skill 调用。

- Faas 支持多种编程语言，包括 python、java 和 js 等。
- Faas 使用消息队列实现组件解耦，以 etcd 作为分布式锁进行全局控制，利用预加载策略提高响应速度。

小米 云平台 / 小米生态云 (https://cloud.mi.com) 2015.7 – 2017.03

为小米生态云设计并实现了用户管理模块，应用自动扩容模块，RDS 和 Cache 的 Service broker 模块以及计量与计费模块，并参与了生态云的资源定价设计。

- 用户管理模块将用户组作为资源的唯一关联实体，与 RBAC 控制策略相结合，避免了用户变更带来的资源和权限管理问题。
- 自动扩容模块为平台上的 App 提供了多种扩容方案，包括基于 metrics 的扩容策略和基于 trigger 的扩容策略，提高了用户资源的使用效率。

个人能力

- 熟悉 C++、Scala、Java、Python, 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、概率图模型领域基本算法和原理。