

杨华

🔗 https://blog.csdn.net/stranger_man ☎ (+86) 17610075871 ✉ 17610075871@163.com
📍 北京市怀柔区中国科学院大学
📅 1994 年 8 月 3 日出生于山东省泰安市



目标：算法工程师

🎓 教育背景

- 2017 年 9 月 - 2020 年 6 月 **工学硕士**，软件工程专业，计算机与控制学院，**中国科学院大学**，北京。
硕士毕业课题，“基于图神经网络的国网台区设备配置合理性分析与研究”
Rank: TOP5
- 2013 年 9 月 - 2017 年 6 月 **本科**，软件工程专业，计算机科学与技术学院，**山东建筑大学**，济南。
毕业设计题目，“基于 Android 的智能家居控制系统的设计与实现”
Rank: TOP5

☰ 相关技能

- 机器学习：** 熟悉逻辑回归、SVM、决策树、XGBOOST、聚类算法
熟悉 CNN、RNN
了解自然语言处理中 Word2Vector、分词、关键词提取等算法
- 平台：** 了解 hadoop、spark 大数据存储运算系统，tensorflow 和 scikit-learn
- 编程语言：** 熟悉 Python、Scala、Java 等编程语言
- 开发平台：** 熟练使用 IntelliJ IDEA、Pycharm 等开发工具
- 阅读书籍：** 《统计学习方法》(李航)、《凸优化》(P.Bertsekas)、《DEEPLARNING》(IanGoodfellow)
《PATTERN RECOGNITION AND MACHINE LEARNING》(M.BISHOP)
《深度学习优化与识别》(焦李成)
《机器学习实战》(PeterHarrington)、《Hadoop 权威指南》(TomWhite)
《统计自然语言处理》(宗成庆)、《Python 自然语言处理实战》(涂铭)

🔗 实习/项目经历

- 2019 年 4 月至今 **多触点归因模型, Multi-touch attribution, 京东广告数据部**
- 代码重构。将串行业务逻辑拆分为并行结构 -> 业务时间缩短 30%，减小代码的耦合度，提高代码复用性；
 - 改进特征类型转换方式。将原来使用 Python 进行 Parquet 到 TFRecord 转换，改进为使用 Spark 进行分布式转换 -> 将转换时间缩短 60%；
 - 增加曝光模型。原项目主要针对用户的点击进行广告位归因，增加了基于广告位曝光的模型；
- biRNN Spark Tensorflow-Keras Scala Python

广告主预算分配, Budget Allocation, 京东广告数据部

- 根据原始技术文档编写初始版本的模型系统，并进行模拟测试；
- 改进初始模型的 ROI (投资回报率) 预测的线性回归模型。使用 Xgboost 算法进行该子模型的改进 -> 将陈天奇开发的 Xgboost-Spark 编译，以免部署的方式运行在 Spark3.0 集群上，使得模型的 rmse 下降了 351。并进一步将 ROI 的预测从天级别提升到小时级别；
- 使用 Maven 将 Xgboost 算法依赖到 JD 的依赖库中，并进行高层的封装和框架的设计。简化了组内复用 Xgboost4Spark 算法的成本，并在此基础上，先后封装了 GDBT、LinearRegression 等常用算法；
- 改进初始模型的 CAP (广告位花销能力) 预测。使用内部 AutoML 框架将原有的基于规则的 CAP 预测升级为稳定可靠的模型；正在进行 CAP 实时预测工作；
- 针对 SparkML 对可视化支持比较差的问题，结合 ClickHouse 和 DarshBoard 搭建了模型结果可视化的 Pipeline；

XGBoost Word2Vector LinearRegression SparkML SparkSQL SparkStreaming AutoML

2017 年 10 月 风机叶片开裂预警, CCF 工业大数据竞赛, 实验室

- 负责完成数据清洗工作，包括缺失值处理、规范化操作以及对数据不平衡问题的处理；
- 负责特征的选择工作，通过分析各个特征间的相关系数、方差、信息系数等来选择最终特征；
- 负责部分模型的构建工作，包括编写 LR 算法完成模型的构建，通过网格搜索来进行模型的调参；

特征工程 逻辑回归 网格搜索

2018 年 9 月

工业控制网络异常行为识别, 国家电网, 实验室项目

- 负责构建网络流量数据包获取, 包括网络数据包深度解析, 主要通过交错时间窗的方式统计流量数据包五元组;
- 负责构建语义特征, 使用 Skip-gram 将数据包转化为向量表示、并通过 PV-DM 和 PV-DBOW 结合的方式建立行为特征向量;
- 负责模型的构建工作, 包括编写 OCSVM 算法完成模型构建, 通过模型调参选择最优参数;

Skip-gram

OCSVM

2014 年 10 月

基于 Android 智能家居房间控制系统, 山东建筑大学, 山东建筑大学机器人与人工智能实验室

- 参与了整个系统的设计;
- 负责智能调配中心的设计与开发工作;
- 负责基于 WIFI 的室内定位算法的设计和开发工作;

软件开发

数据库

JAVA

Android

专利和论文

- 杨华, 李喜旺, 张艳升, “基于卷积神经网络的工控网络异常流量识别”

卷积神经网络

104 规约

- Hua Yang, Xiwang Li, “Research on Identification and Location of IT Equipment Based on Improved Faster-rcnn” (在投)

faster-rcnn

线性插值

注意力机制

语言

英语: 阅读 ●●●●● 4 级: 552

听力 ●●●●○

口语 ●●●○○

奖励与荣誉

荣誉称号 本科期间多次获得优秀班干部、优秀班干部标兵, 优秀团员等荣誉称号;

研究生期间获得中国科学院大学优秀研究生等荣誉称号;

学生工作 本科期间担任班长, 社团创始人等职务; 研究生期间担任实验室负责人等职务;

兴趣爱好

体育: 羽毛球, 篮球, 登山

艺术: 电影, 相声

其他: 旅行, 量化投资

自我评价

学习组织能力强: 本科及研究生期间一直名列前茅, 多次获得奖学金; 曾担任班级班长、社团主席等职务, 踏实肯干, 工作认真, 责任心强, 多次优秀组织班级和社团活动。

团队合作沟通能力强: 本科及研究生期间多次参加科研项目、开发系统、学术竞赛, 深刻理解了团队的概念, 培养了较强的团队意识与人际交往表达能力; 培养了积极的工作态度, 能够适应各种工作压力和工作节奏。