

卢广强

大连 辽宁 | 17604264740 | gqianglu@outlook.com

工作经历

2018.10 – 至今 埃森哲大连 CIO Big Data AI Engineer

- 自动机器学习框架**: 基于Sklearn及TensorFlow独立实现自动数据加载、特征工程、模型构建、模型训练、模型选择、模型评估及模型RESTful API调用, 并支持谷歌云GCS或本地文件及内存对象作为训练及测试数据, **3行代码**即可进行全部Pipeline实现, 已开源至PyPI([automl-engine](#)), 可利用pip安装跨平台调用
- 埃森哲HR 推荐系统**: 利用埃森哲项目数据及员工数据利用NLP对数据处理, 并利用TF-IDF, LSI构建评估模型, 对埃森哲HR系统进行项目推荐及员工推荐, 并利用Spark将训练时间从5个小时缩短至1个小时, 现已上线北美地区
- 多输出回归模型**: 对费列罗不同经销商针未来7天销量预测, 针对7天Multi-Output利用RFR建模评估, RMSE:13.01(均值:12.90)

2017.09 – 2018.09 微神马大连有限公司 大数据部 算法工程师

利用 Spark, TensorFlow, Keras, Scikit-learn 等框架构建机器学习及深度学习模型, 利用公司及三方数据对数据挖掘分析并构建模型, 评估资产端及个人还款能力及还款意愿。从特征工程、模型构建、模型评估、模型分析及模型上线独立完成

- 用户授信评分模型**: 利用协同过滤ALS算法对用户淘宝购买信息补全, 针对13类别**多标签**数据应用Stacked LSTM算法准确率: **98.87%**
- 淘宝订单分类模型**: 应用Word2Vec对淘宝订单非结构化文本数据转词向量并利用TF-IDF加权, 应用LSTM, CNN, ResNet, Wide & Deep深度学习算法实现构建模型并调参, ResNet准确率: **92.20%**
- 银联异盾模型**: 利用LR及Random Forest算法分别对银联及异盾三方数据构建模型调优, 利用Spark Streaming**实时**用户评分
- 个人信用评分模型**: 针对样本不均衡问题开发SMOTE算法, 提出并开发负样本中心点样本生成算法, 模型AUC提高**31.15%**, KS提高**65.71%**
- 资产端信用评分模型**: 利用Spark及XGBoost算法针对不同资产端数据构建模型及调优, 模型KS:**0.44**

2016.07 - 2017.08 东软大连分公司 大数据部 算法工程师

- 大连理工教师聚类模型**: 对用户上网及短信信息数据构建特征变量并构建特征工程, 利用K-Means算法对用户进行聚类, 聚类准确度: **93.00%**
- 江西联通月收入预测模型**: 对江西联通数据构建特征工程并数据处理, 利用BP神经网络对月收入回归实现, 模型MSE: **0.45**
- 辽宁联通用户异常检测模型**: 对辽宁联通用户数据构建特征变量, 利用LR+PCA, Random Forest及SVM等算法进行模型构建调参并交叉验证, 模型准确度: **73.50%**
- 监控视频人流量统计模型**: 利用CNN对视频照片特征提取及SVM分类人流量统计, 模型准确度: **85.10%**

教育经历

东北大学 | 数学与应用数学 | 本科

专业技能

机器学习框架: TensorFlow, Spark MLlib, Keras, scikit-learn, PyTorch, PySpark, MXNet, Spacy, Gensim

大数据处理: Spark, Hive, Redis, Power BI, Google Cloud Platform

编程语言: Python, Scala, Java

机器学习及深度学习算法: LR, SVM, RF, DT, GBDT, NB, KMeans, DNN, CNN, RNN, LSTM

认证: **Azure Data Scientist Associate**

Azure AI Engineer Associate

其他: GitHub 开源项目 [automl-engine](#) 创建者, 开源项目 [Scikit-Plot](#) 贡献者

英语: 托业 860(990)