

卢广强

大连 辽宁 | 17604264740 | gqianglu@outlook.com

工作经历

埃森哲大连 **Data Engineer**: 2018.10 – 至今 NewIT BigData

- Accenture HR 推荐系统: 利用Accenture项目数据及员工数据利用NLP技术处理对数据处理, 并利用TF-IDF, LSI构建评估模型, 对Accenture HR系统进行项目推荐及员工推荐, 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MultiOutput利用RFR建模评估, RMSE:13.01(均值:12.90)

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

应用 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51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RandomForest算法分别对银联及异盾三方数据构建模型调优, 利用SparkStreaming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**

东软大连分公司 **算法工程师**: 2016.07 - 2017.08 大数据部

应用 Spark 及 Hadoop 生态系统技术构建机器学习及深度学习算法模型并评估。

- 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KMeans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, RandomForest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%**
- 实时物体检测模型: 利用TensorFlow及Caffe对SSD算法实现实时物体检测

中科软科技有限公司 **软件工程师**: 2015.11 - 2016.06 银行事业部

- 应用Java对百年人寿核心业务系统进行开发和测试, 在核心业务系统上增加及修改相应功能, 熟练应用SQL进行Oracle数据库DML, 对SQL进行优化及多线程应用

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, SparkMLlib, Keras, scikit-learn, Caffe, PySpark, MXNet

大数据处理: SparkSQL, Hive, HBase, Impala, Hadoop, SparkStreaming

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

竞赛: Kaggle 机器学习 **3次** (1次 Top 7%, 1次 Top 9%, 1次 Top 14%)

其他: GitHub 开源项目 **Scikit-Plot 贡献者** (<https://github.com/reiinakano/scikit-plot>)