

职业技能

线性回归、逻辑回归、svm、k-means、KNN等常见机器学习模型 | 熟练

决策树、孤立森林、随机森林、GBDT、XGBoost、LGBM等常见集成学习模型 | 熟练

numpy、pandas、scikit-learn、polars、pyspark等数据处理库 | 熟练

pytorch架构，能够实现自定义模型/复现论文 | 熟练

FNN、CNN、RNN、LSTM深度网络 | 熟练

CV领域的Res-net、U-net、ViT等模型

NLP领域的word vec、TF-IDF、BERT、Transformer | 熟练

多模态领域CLIP、BLIP模型

python编程语言 | 熟练

huggingface

大模型领域的AutoGPT、Toolformer、LLama、SFT、RM

资格证书

BCM烂飞机奖

CET、CET

华为公司python专业级

教育背景

中南大学 985

信息科学与工程学院 · 控制科学与工程 · 硕士
2016-08 至 2019-07



中南大学 985

信息科学与工程学院 · 自动化 · 本科
2012-09 至 2016-07



自我评价

本人本硕毕业于中南大学，毕业后就职于华为，转岗后从事机器学习/深度学习算法开发工作，项目包括异常检测、根因定位、自动标签数据构建流水线、NLP领域的意图识别等。熟悉机器学习/深度学习常见模型；熟悉NLP领域Transformer、Bert、GPT等模型，有相关应用经历；（1）具备快速阅读英文文献能力，并进行复现（pytorch）；（2）具有较强学习能力，完全通过自主学习转换到当前AI赛道；未来仍计划在机器学习/NLP领域深耕！

工作经历

华为技术有限公司

算法工程师 2021-09 至 2023-06

项目经历

意图识别Demo

负责人 2023-06 至 2023-08

项目目标：需要根据用户输入指令，识别用户意图，例如：提升客厅网络优先级；算法输入为用户指令语句，输出为用户意图，当前方案意图覆盖42个意图，覆盖常见用户指令性语句；

利用DIET架构，整体训练流程：输入语句->bert-base-chinese->transformer->Entity loss、Intent loss（模型简化，未考虑mask loss）；模型推理时输出42个意图置信度，置信度最高的意图即为模型结果；

该算法意图识别准确率85%，实现用户指令意图识别功能。

站点识别项目

负责人 2023-04 至 2023-06

项目目标：需要一算法，从脚本信息中提取出站点名称、以及每个网元属于哪个站点；站点识别率达到80%；

算法输入为日志中的网元名称；输出为站点及站点网元关系；

1、网元名称清洗分词后，利用TF-IDF算法对网元名称进行embed操作，获得每个网元对应的特征向量；

2、利用CW算法对网元向量进行聚类，聚类中心即为站点名称，与聚类中心类别相同的网元即为该站点下的网元。其中无向图中两点之间相似度计算时，索引越靠前的乘积项权重越大。

算法站点识别准确率满足80%，相比原有方案提升10%+

网络质差事件识别

主要负责人 2021-09 至 2022-01

项目目标：基于现网数据，识别分钟级网络质差事件，查准率8X%，查全率6X%

1、利用标签数据，通过LGBM获得特征重要性，并结合业务专家经验、特征数据分析筛选特征；

2、对速率特征进行数据平滑、动态阈值处理；

3、利用孤立森林模型获取每条样本数据的异常分数，异常分数超过阈值的即为质差；

4、手动实现孤立森林算法 (python) 支撑算法落地；

基于实验室数据，算法在TOP 长视频应用上，分钟级质差事件识别准确率、召回率均达标。基于局点验证数据，算法识别的质差用户准确率相比规则提升15%。

质差事件根因定位

主要负责人 2022-02 至 2022-06

项目目标：针对网络质差事件，根因定位准确率90%

项目背景：模型输入为质差事件识别算法的异常分数，输出质差根因（从OLT到家庭侧四个网络段中的n个）。

将问题建模为异常分数与四个网络段对应特征的相关关系挖掘，相关关系越大则对应网络段是根因的概率越大。算法主要步骤如下：1、依据专家经验筛选特征；2、利用因果推断中的DML算法，获得每一个特征与异常分数之间的因果系数(CATE)；3、某特征因果系数大于0.1，则该特征所在网络段为根因之一；

基于XX局点上门标签验证，根因与用户故障匹配准确率达标。

质差事件流水线搭建

负责人 2022-07 至 2022-12

项目目标：搭建质差事件流水线，基于家庭网络场景，实现卡顿标签数据全自动生成，覆盖短视频、长视频、直播、会议等典型使用场景

1、轮式机器人室内自主巡航（负责PC与机器人路径规划脚本）；

2、终端设备app操作脚本：前期基于autojs开发抖音、bilibili、腾讯视频操作脚本，后期更改为基于AndroidStudio（与同事合作）开发；脚本功能包括拟人操作、终端设备屏幕录制、音频录制、设备音视频文件上传；

3、KPI数据获取脚本：shell脚本，实现固定周期从服务器获取parquet表格并压缩存储；

4、视频卡顿识别脚本：基于opencv开发帧差法、模板匹配法、像素统计法、音频法，多个子方法融合实现视频的卡顿识别；

5、baseline模型验证模块：基于抖音app，利用流水线自动采集的标签数据，获得XGBoost、LGBM、FNN、CNN、RNN、FNN+CNN、FNN+RNN模型的

baseline结果（多个APP平均查准率75%，查全率65%）；

该项目实现了卡顿标签的自动生成，以及模型训练验证baseline。

城市网络规划布局最优化项目

负责人 2023-01 至 2023-03

项目目标：单城市场景下，目标意图提升5%（例如成本降低5%）

架设城市网络前需要对设备、线缆进行配置与布局，原有方案需专家多次尝试，耗时严重，因此需要设计算法替代人工。

算法输入为设备的配置信息、路网信息；算法输出为该城市下的最优成本。

1、将json格式的路网信息绘制为简单点线图片，利用autoencoder将图片信息编码为向量数字特征；

2、利用历史数据中的设备配置信息、向量形式的路网信息，训练多层神经网络，作为原黑盒模型的代理模型；

3、利用贝叶斯优化过程，获得最优成本；

针对单城市该算法成本比历史最优降低5%，且算法具有通用性，可用于其他算法寻优。

PC快速充电项目

主要负责人 2021-01 至 2021-09

算法相关工作内容:

- 1、基于卡尔曼滤波获取锂电池SOC，结合锂电池电路模型、电路理论获得状态方程以及观测方程，根据实时测量的电压电流估计电池SOC状态，该算法用于支撑其他项目的电路仿真及控制策略制定。
- 2、数据处理、曲线绘制等脚本开发。

竞赛经历

Kaggle-ICR - Identifying Age-Related Conditions 2023-07

基于公开BestScore方案，筛选特征，更改模型融合方案，加权交叉熵损失为0.54，排名50%；

备选方案:

- 1、利用Denoising Autoencoder对原有数据进行特征提取，基于提取后的特征进行有监督训练；
- 2、利用给定数据的metadata进行two-stage训练，先在子任务上训练模型前两层，保留前两层参数基于最终任务进行有监督训练；
- 3、基于pytorch_tabular API的NODE模型、TabNet模型、TabTransformer模型融合方案(bagging\stacking)；

Kaggle-Stable Diffusion-Image to Prompts 2023-02

竞赛背景：根据图片，生成图片描述prompt；

评价标准：竞赛标签通过将标签文本输入到all-MiniLM-L -v 模型，得到标签向量；每张图片经过提交模型也会得到一个向量，比赛评估标准为标签向量与预测向量的平均余弦相似度；

备选方案:

- 1、BLIP+CLIP模型融合方案，平均余弦相似度0.546；
- 2、CLIP微调方案，平均余弦相似度0.546，相比微调前提升0.01；
- 3、CLIP+ViT融合方案，平均余弦相似度0.542；

竞赛最终排名48%

论文期刊

An improved voltage balance strategy for renewable generation energy storage system 2017-10 发表

- 1、基于卡尔曼滤波算法获得超级电容更准确SOC；
 - 2、依据多并联电路中每个超级电容SOC，调整充电策略；
- 该方法有效降低均衡电路切换次数；缩短电容电压均衡时间。
该论文发表在ECCE会议。