

简 历

基本信息:

姓名: 王涛平

性别: 男

身高: 174cm

体重: 75kg

出生日期: 1990-08-14

居住地址: 昆山花桥

手机: 15221110569

婚姻状况: 已婚

文化程度: 全日制本科

工作年限: 7 年以上

户籍: 安徽宿州

邮箱: wangyitao1989@126.com



个人技能:

专业技能

1. 掌握 Python , java 编程语言
2. 熟练掌握常用机器学习算法, 包括分类 (逻辑回归, SVM 支持向量机, 决策树) 回归 (线性回归, 回归树) 聚类 (Kmeans) 降维 (PCA, LDA) 等, 并可用 sklearn 实现。
3. 熟悉 NLP 相关理论: 中文分词, 词性标注, 命名实体识别, 序列标注、文本向量化及问答系统等 NLP 相关算法, 如: Word2Vec, CBOW, Skip-gram, TF-IDF, TextRank, LDA 并具备 NLP 实际项目经验
4. 熟悉深度学习算法: BP 神经网络, VGG、googleNet、resnet 等 CNN 系列算法, LSTM、GRU、seq2seq 等 RNN 系列算法以及 transform, bert 等相关算法。
5. 熟练使用深度学习框架 pytorch 并实现自己设计的网络。
6. 熟练使用数据库软件 MySQL、Oracle、hive。
7. 熟悉 Linux 操作系统的使用, 掌握 Linux 常用命令, 能够在该平台下进行相关开发工作。
8. 熟悉 hadoop, spark 等大数据开发组件
9. 熟悉 transforms 和 bert 等模型。
10. 了解大模型相关算法: 熟悉 RAG 模式和 lora 等, 对 langchain 有一定了解。
11. 用 vllm、lmdeploy 等模型部署框架部署过大模型。
12. 熟悉 llamafactory 等大模型微调框架。

2024/11-至今 广州久泰汽车服务有限公司上海分公司

1. 主要负责大模型技术的开发和验证。
2. 公司算法的部署和上线。

2022/9-2024/10 江苏蓝创智能科技有限公司 算法工程师

1. 负责公司相关人工智能业务模块的开发。
2. 负责对开发的相关模型效果进行测试和评估。
3. 公司算法模型的迭代和部署。

2021/1-2022/7 上海数鸣人工智能科技有限公司 算法工程师

1. 负责人工智能相关应用在某些领域的一些应用的开发。
2. 公司相关业务的数据进行处理。
3. 公司算法模型的算法优化和上线。

2020/4-2021/1 上海策扑智能科技有限公司 算法工程师

互联网/电子商务 | 0-50人 | 民营企业 | 项目研发

工作描述:

-
1. 负责人工智能相关应用和相关领域的一些应用的开发。
 2. 负责利用自然语言处理和机器学习算法对海量文本数据进行挖掘。
 3. 负责对项目的优化以及生产环境的调试。
 4. 完成领导安排的其他工作

2018-04至2020-02——上海灿谷投资管理咨询服务有限公司

所属行业：互联网金融 企业规模：500-1000

所属岗位：算法工程师

工作职责：

1. 负责公司现有的自然语言处理系统维护和进一步开发
2. 负责问答系统开发，主要是为完成问题分析和答案生成处理
3. 构建出公司第一代命名体识别系统以及智能问答系统
4. 其他 NLP 相关问题
5. 实现机器学习相关算法
6. 待项目完成后，整理相关操作文档。

2016-09 至 2018-03——快钱支付清算信息有限公司

所属行业：互联网/电子商务 企业规模：500-1000

项目经理

1. 反洗钱数据集市业务模型以及处理逻辑熟练掌握(主要是 Mapping 也就是对应关系)
2. 数据集市模块相关技术全部掌握, 相关技术包括 Sqoop、Kettle、Hive、Kylin、Hbase、HDFS、Yarn、Oozie 等的使用
3. 负责大数据平台的系统架构及系统维护方法、流程
4. 快钱投产流程, 系统组日常办公流程全面掌握
5. 与系统组、业务组, 其它部门的沟通协调能力

2015-3 至 2016.09——上海数策软件股份有限公司

所属行业：计算机软件

spark 开发工程师

1. 理解产品需求并提出建议, 同时可根据需求进行部分软件设计
2. 按照产品设计要求进行功能开发, 并按计划要求按时完成编码工作
3. 及时发现编码中的错误, 并加以修正
4. 对软件进行部分测试, 并不断完善

项目经验:

项目名称: 基于 deepseek 的 rag 系统的开发

项目背景: 通过建立数据库以及集合大模型的能力给与用户准确的回答 通过建立企业自有

的知识库，对用户输入的 Query 进行解析，进行语义的混合检索， 将知识库中检索到信息结合 prompt 组成提示词,deepseek 用的 ollama 进行的部署,输入给大模型给用户返回答案,并对语料库进行实时更新。

项目名称：基于广东保险政策的识别

项目背景：通过对打过标签的数据多大模型 qwen2. 5-7b 进行 qlora 微调,通过大模型的微调对相关实体的抽取的验证，这里的微调框架用的 firefly, 最终达到效果使微调后加载的模型可以正确识别相关的实体, 并对识别出的实体做业务上的处理, 达到业务上比较还得效果。

2024/3-2024/10 智慧滨湖项目的开发

项目描述：

该项目主要是为无锡市滨湖区生态环境局做一下平台上面的赋能和实时的监控一些数据。其中模型这边主要做了六个模块分别是污染玫瑰图, 溯源和扩散轨迹, pmf 源解析和天数数据处理、周边传输影响, 颗粒物源解析最近七天的数据、AQI 七天预测。

颗粒物源解析: 本模型基于 AutoGluon 自动机器学习框架开发使用无锡地区颗粒物组分站数据和 PMF 颗粒物源解析数据进行训练, 能够实现对无锡市区域内的颗粒物污染的实时源解析。通过输入颗粒物组分站实时数据, 支持以站点维度对颗粒物进行实时源解析, 源解析类型包括: 燃煤、生物质燃烧、工业、二次硝酸盐、二次硫酸盐、移动源、扬尘源、海盐源, 模型的评价指标为。

通过对东绛站 2024 年 7 月 8 日-14 日的颗粒物组分数据进行源解析分析, 模型综合准确率超过了 90%。

模型的评价指标：

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2}$$

其中：

- n 是数据点的数量
- P_i 是第 i 个预测值
- O_i 是第 i 个实际值

为了让 NRMSE 能够反映预测数据的准确率, 本系统对其计算方法进行了修正, 最终的准确率具体公式如下：

$$\text{预测准确率} = (1 - NRMSE) \times 100\%$$

另外, 针对回归模型, 为了检验数据的拟合程度, 使用回归斜率和 R 平方进行数据回归效果的评估。

回归斜率表示的是预测值相对于实测值的变化率, 越接近 1, 表明数据的拟合程度越好。

R 平方 (R²) 是衡量模型解释变异程度的指标, 范围从 0 到 1, 值越接近 1 表示模型的拟合程度越好, 模型解释变量之间的关系越强。

七天气质预报系统:为了解决目前基于开源数据 API 的城市空气质量预报系统 5 天预报时限问题, 引入 AI 大模型时间序列预测系统, 以 API 返回的预报数据和历史数据为输入, 以大模型输出的预测数据作为延长期预报, 增加模型预测时长, 实现城市级别的七天气质预报。

第一步: 读取输入数据中的城市经纬度, 以及预报因子类型, 调用 Openmeteo API 接口, 输出 5 天预报数据和历史数据, 并进行数据处理。

第二步: 基于 Transformer 架构的大语言模型进行开发和预训练, 能够实现对任意时间序列的零样本预测功能。在对雪浪站 2023 全年臭氧数据的 5 天循环预报测试中, 平均准确率达到 80%, 最高准确率达到 90%以上。详细的模型综合预测准确率。调用 Chronos (示例为 tiny 模型), 将输入数据转换为张量格式, 预测时长设置为 64 (模型限制), 进行预测, 对预测数据进行处理。

模型的评价:

评价方法

本系统使用修正的标准化均方根误差(NRMSE, normalized root-mean-square error)方法, 对预测数据和真实数据进行准确性检验。均方根误差 (RMSE, Root Mean Squared Error) 是一个常用的统计指标, 用于衡量预测值与实际值之间的差异。RMSE 越小, 表示预测值越接近实际值, 模型的预测精度越高。

RMSE 的计算方法如下:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2}$$

其中:

- n 是数据点的数量
- P_i 是第 i 个预测值
- O_i 是第 i 个实际值
- \sum 表示求和操作

标准化均方根误差 NRMSE 是用来衡量预测误差相对于数据范围或其他标准的相对大小。NRMSE 的计算可以有几种不同的方法, 具体取决于标准化的方式。以下是本系统采用的两种标准化方式及其公式:

基于均值的标准化公式:

$$\text{NRMSE} = \frac{\text{RMSE}}{\frac{1}{n} \sum_{i=1}^n O_i}$$

（本公式用于计算 PM10、PM2.5、CO、NO2、SO2、AQI 的预测准确率）

基于数据范围的标准化：

$$\text{NRMSE} = \frac{\text{RMSE}}{\max(O_i) - \min(O_i)}$$

（本公式用于计算 O3 的预测准确率）

为了让 NRMSE 能够反映预测数据的准确率，本系统对其计算方法进行了修正，最终的准确率具体公式如下：

预测准确率 = $(1 - \text{NRMSE}) \times 100\%$

评价流程

收集站点历史数据

使用站点的历史监测数据进行模型评价，数据中的缺测值需要经过插值处理，数据具体格式如下（以雪浪站的臭氧数据为例）：

item_id,timestamp,target	#文件头保持不变
03,2023-01-01 00:00:00,49	#第一列：污染因子名称（英文简写）
...	#第二列：日期时间（yyyy-mm-dd
hh:mm:ss）	
	#第三列：污染物数据值

循环测试准确率

使用历史数据对模型进行循环预报测试，并计算预测结果的 NRMSE，从而得到长时间尺度下的模型预测综合准确率。

考虑到数据具有以天为单位的周期性变化特征，循环预测的时间步长设置为 24 小时。

由于 AI 大模型的特征，模型在相同输入数据的情况下预测结果会出现波动，为消除 AI 大模型的预测波动性，采用多次循环预测测试（至少为 10 次），记录并分析综合预测结果。

预测准确率综合分析

对长时间的模型预测准确率进行分析，并进行可视化分析，综合评价模型的预测结果。可视化类型包括综合预测准确率和平均预测效果：

综合预测准确率：记录全年循环预报的测试结果，计算每天的预报准确率，并计算全年平均准确率。

平均预测效果：将全年的预报数据进行平均化处理，并与对应历史数据的平均值进行比较。

项目描述:

运营那边把抓取的 dpi (特征) 在三个运营商 (移动, 联通, 电信) 的平台上部署, 通过大数据的代码把每天的数据推到我们公司自己的数据库, 我们从数据库中提取相关算法所需要的字段, 对数据进行预处理, 把数据处理成算法需要的格式, 然后放到算法中, 相关算法有, lightgbm, fm, svm, randomforest, gdbt+lr, DCN, deepfm 等等比较每种算法的 auc (这里用 auc 进行评估) 和性能, 最终选择一个最好的, 按阈值或者比例对数据进行输出, 给到运营那边, 进行投放. 通过投放的具体结果, 不断去优化自己的模型。

2020/9-2021/1 智能识别判决书中的争点提取以及双方在同一风险点下判决结果的倾向性

项目描述:

为了更好的服务公司的相关需求, 通过 pyltp 中的句法分析和语义角色标注, 使项目更智能化识别争点 (双方对合同中存在争议的条款即为争点), 和判决书中的判决更倾向于双方哪一方胜。

责任描述:

1. 首先输入一段判决书将判决书中的英文符号转换为中文符号。
2. 将判决书中的每一段内容分别取出来。如【原告诉称】、【经审理查明】、【本院认为】和【判决结果】
3. 利用法务那边整理的一些实体和谓语动词, 并结合 pyltp 里的句法分析和语义角色标注, 提取双方的争点
4. 对判决书中的【买方卖方】和【原被告】的关系。另外可能会使用的称呼为【买方卖方】、【供方需方】、【出卖人, 买受人】【甲方乙方】等可分别从【原告诉称】、【本院查明】、【本院认为】中查找是否存在【原告向被告购买】或【被告向原告销售】等表达 (种类较多) 即确定买卖关系。
5. 当【甲乙原被告】按顺序关系取不出时, 这还是寻找【甲乙方】的买卖关系。针对的文本只是本院查明。依据的主要句式为【甲方向乙方购买】或【甲方向乙方销售】等表达。
6. 通过固定的处理结果, 将判决书中双方的身份及称呼同统一, 现已知称呼有: 【原告】和【被告】、【买方】和【卖方】、【供方】和【需方】、【出卖人】和【买受人】、【甲方】和【乙方】以及【双方真实名称】。
7. 将所有 buy_subject 中的词替换成买方, sal_subject 中的词替换成卖方。
8. 对一段文本中, 一个同义词组包含的词进行替换, 全部替换成特定词。例如存在一个同义词组【A1, A2, A3】, 一个文本为【A2 B4 C1 D3】 (字母表示一个词), 最终替换的结果为【A1 B4 C1 D3】, 即用 A1 替换 A2。
9. 同一风险点下, 判决结论对于双方倾向性的统计结果。

判决结果有四种可能, 分布是:

1. 买方胜: 即判决结果中的每一点更倾向于买方
2. 卖方胜: 即判决结果中的每一点更倾向于卖方
3. 双方护佑胜负: 即判决结果中的每一点既有支持买方的又有支持卖方的
4. 结果未知: 如果判决中有维持原判或原告撤诉等相关表述。

2020/7-2020/9 预案智能匹配

项目描述:

为了更好的服务应急开发需求, 项目管理人员根据客户需求, 提出开发一个能根据部门 id 和事件类别 id 和关键词, 能自动匹配相关的预案。

责任描述:

1. 如果输入一个参数部门 id 的话, 会去查询这个部门是市级部门还是区级部门, 如果是区级部门会判断是那一个区。如果是区级预案返回这个区的所有预案 id 和预案名称, 如果是市级预案则返回所有的市级预案。
2. 如果输入的是两个参数, 一个部门 id, 和一个事件 id 的话, 根据部门 id 返回的相关预案, 根据事件 id 的到事件的名称, 把事件的名称进行关键词的提取, 判断预案名称里有没有相关的关键词, 有多个关键词的话, 按照余弦相似度的大小进行排序, 如果预案名称里没有相关的关键词, 则用 bert 对相关的事件名称和预案名称进行向量化。根据向量化的数据比较相似度, 并按相似度的大小进行排序, 获得预案的索引, 最后根据相似度的大小返回相关预案。
3. 如果输入的是二个参数 一个部门 id 和一个关键词的话, 根据部门 id 返回的相关预案, 把关键词和每个返回的预案用汉明距离公式, 分别计算其距离。根据距离的大小进行排序。根据距离的大小返回相关的预案。

2020/4-2020/6 突发事件属性智能匹配

项目描述:

由于本项目是为应急管理部门服务的, 为了做到识别一些文本更加数字化和智能化, 也为了解决相关人力的成本, 应急关系部门提出相关的需求, 根据客户那边的需求, 所以进行了相关需求的研发。

本项目通过一段突发事件的文本描述, 对这段文本能智能识别一些要提取的信息, 识别的信息包括: 事发时间、接报时间、伤亡人数、接报单位, 时间标题、对相关危化品的识别、事件类别。

责任描述:

1. 对伤亡人数的识别: 读取自定义的词典, 加载自定义的词性, 用 jieba 分词工具进行分词, 并遍历分词后的词性, 截取伤亡人数词性附近的数词, 把数词进行提取, 并按规则进行匹配。
2. 对事发地点的识别: 读取自定义字典的一些地点的名词, 把文本中的相关地点名词按照自定义字典进行提取, 提取到的词用正则匹配过滤掉一些无用的地点名词。把提取到的地点名词 按照自定义字典 (自定义字典中有相关的市、区、镇(街道)、路(街)、小区(办公楼)), 对提取到的地点名词按照行政区划的大小进行排序, 然后进行拼接。
3. 对事发时间和接报时间的识别: 按照相关词性匹配规则对事件进行提取, 如果规则没有提取到事件就把一段文本中词性是时间的词语进行提取, 把上面提取的词语转化为标准化 (格式为 YYYY-MM-DD HH:MM:SS), 把格式化的时间进行排序, 提取相关的时间。
4. 对接报单位的识别: 提取一段文本中第一个接字和第一个接字的索引, 按照索引提取相关的文本, 按照相关表里的字段 对自己提取到的文本进行过滤。

5. 对相关危化品放的识别：把相关的危化品目录加入自定义危化品库，遍历自定义字典库中的相关危化品，按照危化品的长度遍历文本，对遍历的文本和危化品进行相似度的计算（余弦相似度），对相似度大于 0.65 的文本进行危化品类别的提取。

6. 对事件类别的识别：把文本中的进行分词，过滤掉停用词和一些低频词，对得到的相关文本中的词语进行 tfidf 编码，对事件类别进行 onehot 编码，把文本中的数据分成训练集和测试集（90%为训练集，10%为测试集），把训练集的数据输入模型进行训练（这里用的模型是逻辑回归），模型在测试集上的准确率达到 75%左右。

7. 对事件标题的识别：对相关提取到的东西进行拼接，拼接的顺序
地点+“发生一起”+事件类别+“事件”

项目名称:电子病历的命名实体识别

项目描述：项目为某智能医疗项目的一部分，主要采用了词向量，IDCNN，RBF 技术进行特征提取与 51 类命名实体的序列标注，该模型 F1 值达到 91%，并使用 Flask 模块部署至线上。

责任描述：扩充标注数据：利用爬虫，Jieba 分词等技巧，爬取在线诊疗网站上的症状，诊断方式，疾病名，治疗方式等命名实体数据，扩充原有标注数据，该技巧比人工收集，标注数据高效；

数据预处理：利用 iobes 方法对数据进行长度和边界特征信息的标注；构建字典；构建 batch 数据；模型搭建及训练：使用 TensorFlow 模块，基于词向量，IDCNN，RBF 搭建模型。加载预先训练好的字向量，用 IDCNN 对序列进行 4 次特征提取，再拼合 4 次结果作为一个特征，提取序列不同视野范围内的上下文特征，最终用 CRF 算法进行序列标注。训练并调参，最终保存模型和最新的字向量。

模型部署：除了本地离线使用，还使用 Flask 模块将模型数据输入函数映射至线上供远程调用，节省客户计算机算力资源。

项目名称：对公司的 APP 产品增加人脸验证功能

项目描述：

随着公司用户量的增加,这增加了相关工作人员在处理工作上所花费的时间,为给用户增加比较好的体验,以及减少工作人员在业务上所用的时间。所以为公司 APP 增加了一些人脸识别的功能。

前期采用 python+openCV+Dlib 模式，后改进为 openCV+mtcnn+facenet+flask 模式。

1、Dlib 模式：预先存储好相关人员的人脸照片，opencv 调用摄像头获取照片后用 Dlib 对齐人脸，通过与数据库中人脸计算欧式距离检测相距最近的，将判断结果用 ajax 传到前台。

2、mtcnn+facenet 模式：分为注册和登录两部分。注册用 openCV 实现从摄像头读取视频帧，对读取的视频帧采用 mtcnn 方法检测人脸，通过 facenet 对检测的人脸进行 embedding，映射成 128 维度的特征向量，其中标签为人名，存储到数据库 csv 文件中。登录时对新读取的人脸数据与数据库 csv 文件中的 128 维度向量计算欧式距离，返回距离最小值对应的标签给前台。

责任描述:

- 1.项目前期调研,撰写调研报告;完成相关开发文档的书写。
- 2.负责部分项目中人脸验证登录、验证的代码实现。
- 3.完成相关 bug 的修改,以及项目的联调,上线。

项目名称: 客服智能问答系统

项目描述:

随着公司业务的增长,使客服人员服务更加专业化,智能化。借鉴客服人员日常的工作记录和常用的话语描述以及大量的 AMM 等手册,打造公司,信贷,借贷,以及后期服务等问答系统。

责任描述:

- 1, 数据预处理: 处理文本格式,分词,将问答对按字数长度分桶存储,语料较少的桶内数据采用 bootstrap 进行少量采样,尽量保证每个桶内语料的平衡,并建立词汇表,问答对中词与 id 的映射关系。
- 2.算法调研及实现:baseline 模型使用 seq2seq + attention 实现 Encoder 时使用了三层双向的 LSTM,Decoder 时使用三层单向的 LSTM。损失函数使用 Sampled Softmax,降低由于词汇表太大而引起的训练复杂度,模型评价使用了 tri-gram 的 BLEU 评价算法。测试集上 BLEU 为 0.65
- 3.算法升级及代码优化:将 LSTM CELL 替换为 GRU CELL。同时将 Encoder 优化为使用 8 层双向 GRU,Decoder 时使用 4 层单向 GRU,在 Encoder 输入特征加入分词特征、同时在 Decoder 输出增加 beam_search 提高模型在句法及语法上的准确性的同时提高了模型效果。
- 4.超参选择及模型评估:逐步缩小超参范围选择,最终确定 lr 为 1e-3,batch-size 为 128..最终模型在测试集上 BLEU 为 0.82

项目名称: 公司抵押贷款用户逾期预测

项目描述:

我公司是一家汽车抵押贷款服务提供商,通过线下直营门店以及线上 APP 等渠道提供汽车抵押贷款服务,为保证资金流通,运转正常,特对于债务方的还款能力做出预测。

责任描述:

开发工具: numpy、pandas、sklearn.metrics、sklearn.model_selection、matplotlib、time

第一: 特征工程

- 1.数据的加载,使用 pandas 包的 read_csv 进行读取。
- 2.使用 numpy 的 describe、count 方法,查看数据的整体情况。
- 3.缺失值的处理,使用 fillna 将缺失值填为-1。
- 4.对现有的字段进行分析,比如将字符串转化为数值型;将连续型进行分箱,时间类型进行拆分。
- 5.使用 sklearn.preprocessing 的 StandardScaler 对数据进行标准化。
- 6.根据属性的重要性进行属性的筛选,使用随机森林的 feature_importances_以及 numpy 的 argsort 方法,进行属性的重要性的排序和选择。

7.进行自变量和因变量的数据集的分离时，用 `drop (axis=1)` 函数删除 `y` 列为自变量数据集，`y` 为因变量数据集。

第二：数据建模

1.因为数据的因变量为离散值所以选择分类算法进行建模，绘制 ROC 曲线，初步判定模型的好坏，最终筛选模型为随机森林。

2.对随机森林的参数进行调优。使用 `sklearn.model_selection` 中的 `GridSearchCV` 进行模型的参数调优，得到最优参数。

3.使用正确率、召回率，精确度、F1 值等评估方式来判定此模型的好坏。

第三：项目结果

1.模型评估。模型的正确率达到了 91.67%，召回率达到了 94.33%,F1 值达到了 87.28%，该模型可以较准确的将还款能力较低的客户预测出来，基本上公司的资金的正常周转。

项目名称：数据集市接入公司监控平台

项目描述：

项目目的：监控大数据平台的数据，当数据达到一定的阈值时，会通过邮件，短信，电话通知相关工作人员。

项目背景：监控中心不能直接监控大数据平台的数据，只能在生产环境 `mysql` 数据库下，建一个用户 `amldm` 里建一张表 `w_batch_dts_warn`，大数据平台 `t_log` 的数据，实时推送至 `mysql` 数据库下 `amldm` 用户名下的 `w_batch_dts_warn` 表中。监控中心监控 `w_batch_dts_warn` 表。

责任描述：

- 1.完成了相关方案的编写
- 2.完成了项目安全成功的上生产
- 3.完成项目的相关测试

项目名称：智能交通数据分析平台

开发工具：Linux+Spark+hadoop+hive+ookeeper+kafka+flume

项目描述：

智慧交通综合管控平台作为智慧交通建设的核心和数据采集、处理、融合、应用的枢纽，将智能交通建设的各个外场子系统的数据进行实时的采集、转换、处理及存储，并在数据集成、融合、分析挖掘的基础上，汇聚于公安交通的智慧应用中，大大推进了交通业务的智能化和智慧化，提高了交警各部门人员的工作效率

责任描述：

- 1.参与项目的前期分析，搜集到交通的违规数据比如超速，套牌车，遮挡车牌，违法停车以及酒驾等
- 2.用 `flume` 采集的交通数据，放到 `kafka` 实现业务需求
- 3.对采集过来的数据用 `spark` 进行过滤，再进行清洗。然后放到 `kafka` 中
- 4.自定义一些排序的类，自己写的类继承排序类，以实现功能

-
5. 用 spark 分析实时的结果比如各个时间段的超速，黄车，套牌车，遮挡车牌等
 6. 根据分析的数据的结果给交通主管部门提出一些建议。加强对交通的管理

培训经历：

培训机构：北风网

培训地点：上海

培训描述：

1. python
2. 机器学习
3. 深度学习
4. 智能推荐系统
5. 自然语言处理
6. 计算机视觉

教育经历：

2010-09 至 2014-06——黑龙江八一农垦大学 电气工程及其自动化 本科

自我评价：

1. 本人性格开朗、稳重、有活力，待人热情、真诚。工作认真负责，积极主动，能吃苦耐劳
2. 喜欢思考，虚心与人交流,以长补短。有较强的组织能力、实际动手能力和团队协作能力。我不断地完善自己，提高自身素质
3. 我在学好专业课与公共课的基础上，还阅读了大量的课外书籍，不断地增加新知识，陶冶情操、开拓视野
4. 不仅如此，我还积极参加社会实践活动，和各种文体活动，培养了我一定的组织与社交能力