

个人简历

姓 名	王召磊	出生年月	1990.01
民 族	汉族	性 别	男
电 话	15632877484	政治面貌	中共党员
邮 箱	wzl410@126.com	毕业院校	中国海洋大学
籍 贯	河南平顶山	学 历	硕士



求职意向

目标职位： NLP 算法工程师

工作性质： 全职

目标地点： 南京

专业技能

- 1.熟练掌握词嵌入模型，如 N-gram, CBOW, Negative Sampling, word2vec;
- 2.熟练掌握深度学习模型，如 MLP, CNN, RNN, LSTM, Seq2Seq, Attention, Transformer, bert 等;
- 3.熟悉中英文分词、文本过滤、词性标注等常用文本挖掘方法；掌握相似度计算、命名实体识别、情感分析等技术；
- 4.掌握线性回归、SVM、KNN、K-means 系列、朴素贝叶斯模型，对 DT、RF、GBDT、XGBoost 等模型有一定了解；
- 5.熟练掌握使用 Tensorflow 框架进行开发，实现自己设计的网络；
- 6.熟练使用 Python 语言，熟练使用 Numpy、Pandas 等数值计算库；
- 7.熟悉特征工程，包含数据清洗，特征提取，特征选择，数据降维，模型评估等；
- 8.了解网络爬虫技术，如 requests, BeautifulSoup、lxml 动态和静态网页爬取方法等；
- 9.了解 java 基础，熟悉 scala 语言，熟练使用 spark 框架，可以使用 spark 进行分布式训练和分布式部署；
- 10.了解 Linux 操作系统及 MySQL 数据库常用命令，熟练使用 docker 进行环境搭建与部署。

工作经历

时间：2020.03—至今 **普恒物联网智能科技（南京）有限公司** **职位：算法工程师**

工作内容：

- 1、构建环保（污水处理）行业数据集；
- 2、行业文本分类算法模型训练；
- 3、行业智能问答系统构建；
- 4、数据中台产品数据开发模块算子开发；

时间：2019.09—2019.12 **上海北风教育培训有限公司** **职位：人工智能技术顾问**

工作内容：

- 1、开展循环课，基础模型算法理论讲解与疑难解答；
- 2、相关项目实现讲解，主要有词向量生成、中文分词、中英文情感分析和聊天机器人；
- 3、阶段考试试卷出题、批阅，作业批阅，阶段项目实践；

时间：2017.10—2019.08 **上海友极网络科技有限公司** **职位：深度学习算法工程师**

工作内容：

- 1、负责对商品评价模块内容进行情感分析；

2、评价内容命名实体识别；

3、负责算法验证、优化与测试

项目经验

项目一：电商平台情感分析项目

项目描述： 顾客评论内容分析对于了解客户心声、精准营销定位、改善产品服务至关重要。使用 CNN+LSTM 的方式提取文本内容的高阶特征数据，然后使用全连接方式分类输出正面评价和负面评价，改善商家服务、提升用户粘度。

责任描述：

1. 对顾客评价文本数据进行分类标注，分为正面评价和负面评价；
2. 数据预处理，采用字向量、分词信息、词性为特征；
3. 搭建开发框架，对比 TextRNN 选择使用 TextCNN+双向 3 层 LSTM 方式提取高阶特征；
4. 获取的高阶向量采用全连接方式进行分类输出；
5. 调节模型的学习率、优化器等，提高模型的准确率；

项目结果： 通过对模型的优化调整，最终使得模型的准确率达到 98%，取得不错的效果。

项目二：评论文本内容命名实体识别项目

项目描述： 顾客评论内容是电商平台的巨大财富，隐含大量的信息，反映出顾客对产品质量、物流及卖家服务等的态度以及平台购物体验，有效挖掘评价数据信息，对于改善卖家服务，将优质产品推荐给潜在消费者，提升平台竞争力有重大意义。

责任描述：

1. 数据预处理：处理文本格式，通过字典匹配及人工辅助的方式为数据打标签，将文本中的数字全部替换为 0，建立字典、字和 id 的映射关系。

2. 算法调研及实现：通过查阅大量资料选取 baseline 模型为 IDCNN 和双向 LSTM 对比选择。验证集上 F1 分别为 84.66 和 86.52%

3. 算法确定以及优化：选择 IDCNN+CRF。减小 padding 噪音，加入分词特征，并采用残差思想改进网络。

4. 超参选择以及模型评估：在训练过程中采用随机采样，逐步缩小超参范围选择，最终确定 lr 为 $1e-3$ ，batch_size 为 128

项目结果： 构建出的命名体识别算法最终模型在测试集上 F1 为 92.13%

项目三：文本分类和情感分析

项目描述： 对从行业网站、微博等途径获取的行业文本信息进行分类汇总，行业发布的政策文件、地方新规等发现后推送给项目负责人，同时应急指挥相应指示灯闪亮。爬虫获取的微博、今日头条等社交网站的信息进行感情色彩标注，数值化后训练情感分类模型，结合点赞和转发等热度据，将重大正面和重大负面信息推送到页面。

责任描述：

1. 文本数值化：分词、清洗，并建立词汇表以及词与 id 的映射关系，文本转为数字 id 并进行填充。
2. 算法调研及实现：模型使用 textCNN+双向 LSTM 训练分类模型；
3. 针对所有数据，使用训练好的模型，将文本转换为向量，并将向量保存到数据库中；
4. 超参选择及模型评估：最终确定 lr 为 $1e-3$ ，batch_size 为 128

项目结果： 通过分类模型将获取的信息分类归整并推送给决策者及时查阅，提高了时效性。

项目四：检索式智能问答系统

项目描述： 污水处理过程中出现设置问题和水质问题时会指派设备运维工程师和水质工艺工程师现场解决，行业领域

内的问题和解决方案较为固定，对于有经验的运维师傅容易解决，缺少经验或不太确定的时候可以咨询专家，构建专家知识库，建立算法模型并训练，基于文本相似度匹配最接近的问题和解决方案。

责任描述：1. 数据预处理：处理文本格式数据，分词、清洗，并建立词汇表以及词与 id 的映射关系，将问答对分词结果转为数字 id 并进行填充。

2. 算法调研及实现：模型使用双向 LSTM 训练高阶向量提取模型，构建三元组损失并训练优化；

3. 针对所有数据，使用训练好的模型，将文本转换为向量，并将向量保存到数据库中；

4. 超参选择及模型评估：使用递减学习率，batch_size 为 128

项目结果：检索匹配结合点击选择方式匹配最接近的问题及答案，提升运维人员的工作效率、

项目五：智慧云舆情监控系统

项目描述：网络舆论的不成熟加上官方回应的不及时，环境问题引起的社会公众的不满情绪持续发酵，事后尽管有关部门及时澄清，也难以消除因此带来的负面影响和政府的形象。因此，环保舆情发生后，及时处理和解决现实问题才是重中之重。整合互联网信息采集技术及信息智能处理技术通过对互联网海量信息自动抓取，基于自然语言处理技术（NLP）进行文本分类，对相关热点问题持续追踪，实时把控舆论动向，及时通知到相关人员，从而第一时间应急响应，为正确舆论导向及收集群众意见提供帮助。

责任描述：1. 原始数据获取：从新浪微博、今日头条等自媒体平台及水务、环保等门户网站爬取相关文本数据；

2. 数据预处理：数据清洗和统一格式转换；

3. 算法调研及实现：使用 bert 提取文本句向量，下游 bi-lstm 模型进行分类任务；

4. 超参选择及模型评估：使用递减学习率，batch_size 为 64

项目结果：舆情监控模块是智慧云产品的核心模块，重要结果直接进入智慧云应急指挥模块，为领导层做出决策提供依据。

教育背景

- 2013.09-2016.06 中国海洋大学 捕捞学（硕士）
- 2009.09-2013.06 浙江海洋大学 海洋渔业科学与技术（本科）

自我评价

- 1.硕士学历，NLP 的相关技能，同时有较好的数学基础；
- 2.善于思考，善于钻研，具有较强的学习能力和逻辑思维；
- 3.对工作充满热情，抗压能力强，具有较强的责任心和团队精神，敢于迎接挑战。