

# 张越

求职目标: NLP 算法工程师

工作年限: 8 年

联系方式: 177-7109-1249

## 基本资料

姓名: 张越

性别: 男

毕业院校: 华中农业大学

年龄: 29

学历: 211 本科

专业: 生物信息学

期望薪资: 面议

邮箱: 838816070@qq.com

## 工作经历

2021.6-至今                      华为技术有限公司                      AI 解决方案工程师

工作内容:

1. 负责 AI 产品的交付实施, 日常维护, 算法优化, 输出解决方案交付文档, 管理整体项目进度, 整理和提炼同类场景下的解决方案, 客户培训, 后期转维;
2. 交付过的项目有: 铁路故障自动识别项目 (工业质检方向), 红柳林智慧园区项目 (智慧安防方向), 禅城城市大脑项目 (智慧城市方向), 机场航班节点保障项目 (飞机状态自动识别).

2020.11-2021.4                      华融证券股份有限公司                      NLP 算法工程师

工作内容:

负责财经类知识图谱的搭建; 负责金融文档智能审核项目和金融文档智能写作;

2017.8 -2020.9                      温州市群英教育科技有限公司                      NLP 算法工程师

工作内容:

- 1、负责相关教学项目的研发整理, 例如用 BERT 实现学生自动阅卷系统;
- 2、负责教学项目的讲授、辅导、答疑;
- 3、参与相关教学内容题库更新、维护, 以及监考、批改试卷等工作;
- 4、负责一对一辅导教案的设计, 授课教学, 以及新老老师的培训

2015.7-2017.7                      天津诺和致源生物信息有限公司      生物信息工程师 (数据分析)

## 专业技能

- 熟悉常见的深度学习理论 (CNN、RNN) 和框架 (tensorflow、pytorch、keras)
- 熟悉 BERT, fasttext, GPT, DNN, W2V, attention, transformer, seq2seq, textcnn 等常用的深度学习模型;
- 熟悉 Python 和 C++, 了解使用过 perl, R 语言, Matlab 等语言
- 熟悉 Mysql、Neo4j、Redis 等数据库操作
- 掌握 Python 编程基础, 熟悉 Numpy、Pandas、Matplotlib、Seaborn、Scikit-learn 等库的使用, 具备良好的编程规范。
- 熟悉线性回归, 逻辑回归, 决策树, KNN, HMM, CRF, SVM, 随机森林, GBDT, Adaboost, XGboost 等算法, 能够通过 Sklearn 等库中相关函数调用进行建模、预测。

- 掌握基本爬虫技巧,能熟练地抓取自己所需要的网页内容,并使用 python 进行数据分析
- 熟练使用 PyCharm, Jupyter Notebook, Sublime 等开发工具,熟悉 linux 开发环境,能熟练使用 vim 编辑器进行代码编写
- 拥有良好的数学基础,对高数线代概率论都有深刻的理解,理解常见机器学习公式推导
- 算法功底很好,熟悉数据结构的使用,熟悉多种排序查找算法,了解多种常见算法思想,如贪心回溯递归分治,动态规划等算法

## 项目概况

### 1. 项目名称: 群英大脑项目

项目时间: 2020 年 3 月-2020 年 9 月

信息咨询中心辅助系统:

项目描述及实现的功能:

客服人员与客户交流过程中,往往不能及时记录客户信息,导致无法及时联系客户,为此设计信息咨询辅助系统,帮助客服快速提取关键信息.

**项目职责:**客户往往不愿意透露自己全名,只能获取客户的姓氏,因此先采用 fasttext 模型判断客服提问中是否在问姓名,并进行优化,随后再用规则进行匹配;对于手机号,QQ 号和微信号,由于本身具有规则,因此采用正则表达式来匹配;意向学科和意向地区属于有限集,直接使用规则匹配的方式进行提取.

对任务优化措施有:自动超参数调优,字词粒度对比实验,回译法数据增强,领域内领域外词向量迁移,领域内一致性(UDAT)对比实验,其中 UDAT 提升最大,数据增强无明显提升,其余均对准确率有所提升.

**考试中心试卷自动批阅系统:**

项目描述及实现的功能:

为帮助考试中心快速批阅试卷,我们使用 AI 技术帮助改卷.选择题和判断题有明确规则可以直接匹配出答案,而填空和简答题则需要借助模型来判分,本项目只针对填空题展开.

**项目职责:**

分析原始数据,统计文本长度,进行数据清洗,去除异常和无用字符

考试答案与标准答案完全相同时,使用规则进行匹配;不同时,把批阅转换成文本分类问题,按照满分,中间分和零分来做三分类.使用 BERT-Multilingual 预训练模型,再进行迁移学习完成判断.

对模型进行超参数调优,标签平滑优化,对预测错误样本做 badcase 分析

**涉及技术:** pytorch+Fasttext+Bert+Django+Nginx+Supervisor

### 2. 项目名称: 医生智能对话系统

项目时间: 2019 年 12 月- 2020 年 6 月

项目描述及实现的功能:

为帮助病人方便快捷地咨询病情,无需等待医生挂号,我们通过AI技术,基于临床病例数据,构建医疗知识图谱,实现医生在线智能对话,用以辅助诊疗.

**项目职责:**

在线部分数据流:从用户请求开始,通过 werobot 服务,在 werobot 服务内部请求主服务,在主服务中将调用会话管理数据库 redis,调用句子相关模型服务,以及调用图数据库,最后将查询结果输送给对话规则模版或者使用Unit对话API回复.

离线部分数据流:从数据采集开始,将获得结构化和非结构化的数据,对于结构化数据将直接使用实体审核模型进行审核,然后写入图数据库;对于非结构化数据,将使用NER模型进行实体抽取,然后通过实体审核后再写入图数据库.

命名实体识别模型:Bilstm+CRF

命名实体审核模型:bert-Chinese+RNN

句子主题相关性模型:bert-Chinese+微调脚本

**涉及技术:** Bilstm +CRF+Bert+redis+neo4j+werobot+flask

3. **项目名称:** 构建文本评论情感分类模型

**项目时间:** 2019年8月-2019年12月

**项目描述及实现的功能:**

基于学生评论数据,进行老师评价情感分类场景分析,运用Python进行文本清洗及利用词向量转化技术(tf-idf)进行文本向量化转化,并使用逻辑回归模型进行模型训练和调优,准确率90%

**项目职责:**

1. 文本的数据清洗
2. 文本向量化转化
3. 模型的训练好调优

**涉及技术:** Python+词向量转化+机器学习模型开发

4. **项目名称:** 客户流失预警

**项目时间:** 2019年6月-2019年10月

**项目描述及实现的功能:**

随着互联网的发展,企业面临着离客户越来越远,客户流失难以挽回,客户的维系成本不断攀升的问题,造成企业的成本不断提升。通过深入分析客户的特征,为客户提供个性化的营销与服务的手段,进而达到挽留客户的目的,从而降低公司的运营成本。

**项目职责:**

1. 根据内外部数据信息开展数据挖掘,从分析过程中发现问题,提出策略建议。
2. 负责对业务大数据进行采集,识别,特征化,建模,可视化;

3. 为内部同事提供数据支持，参与分析业务运行情况，协助解决业务问题；
4. 根据公司业务需求，进行机器学习方面的研究。

**涉及技术：** pandas+numpy+sklearn+GBDT

## 自我评价

1. 吃苦耐劳，能接受较大的工作强度。
2. 学习能力较强，坚信利用网络可以解决所遇到的问题。并有一套自己的学习方法。
3. 团队意识较强，能较快融入团队的工作。
4. 擅长总结，能较快从所遇问题总结出来内在的联系。