



赵斌

NLP算法工程师 | 3年

18612032101



zbmain@126.com



个人优势

熟悉深度学习框架：Pytorch、Tensorflow  
了解：LR、HMM、CRF、GBDT、XGBoost等ML算法  
熟悉：Word2vec、Fasttext、RNN变种、Attention、Transformer及Bert等深度学习模型架构  
熟悉数据库：Mysql、Neo4j  
掌握编程语言：Python、JS、AS  
了解Linux开发环境、Shell、Java、C  
熟悉AIR应用开发、游戏开发、了解前端开发及工作流程

项目经历

商城搜索

NLP算法工程师/算法工程师

2019/11 – 2020/10

项目介绍：通过用户输入文本，实现搜索功能推荐关联商品，使用户输入信息可以更加人性化。

- 模型部分架构分为：离线部分、在线部分
- 离线部分
  - 对已有的结构化数据和非结构化数据预处理，抽取商品及描述信息实体
  - 构建关系（[商品-[r:Des]-描述），将抽取的实体信息录入Neo4j图数据库中
- 在线部分
  - 判断用户输入信息是否包含描述实体，并找到描述连接的商品名名称
  - 后端开发调用接口，将匹配的数据库信息展示给前端
- 实现方案
  - 设计实体关系标签表，构建Neo4j原始图数据
  - 使用BiLSTM+CRF实现NER任务，baseline Acc=0.78
  - 构建RNN模型并联合Bert词向量，实现对NE实体合法性的审核
  - 再次扩充3000条样本数据、加入NE审核环节，NER准确率提升到0.92
  - 使用Flask部署接口服务
- 涉及技术
  - BiLSTM+CRF、RNN+Bert中文词向量、Flask、Neo4j

故障咨询客服  
对话机器人

NLP算法工程师/算法工程师

2018/11 – 2019/05

项目介绍：通过对话系统，快速定位汽车故障问题，并通过定位，推荐附近维修点。

- 模型部分架构分为：离线部分、在线部分
- 离线部分
  - 对已有的结构化数据和非结构化数据预处理，抽取故障的描述信息实体
  - 构建实体标签树
    - 故障关系（[故障-[r:Info]-描述）
    - 构建关系（[城市-[r:Has]-维修店）
    - 构建关系（[维修店-[r:Work]-故障）
  - 将抽取的实体信息，添加到图数据库中
- 在线部分
  - 判断用户输入信息的合法性及是否包含描述实体，得到故障实体
  - 通过故障实体，找到该城市擅长领域的维修店列表
  - 前端通过定位系统，找到列表中距离最近的维修店给用户
- 实现方案
  - 设计实体关系标签树，构建原始图数据
  - 使用传统RNN+Bert词向量，实现实体合法性检测。通过回译增强、标签平滑，Acc提升至0.95
  - 使用BiLSTM+CRF实现NER任务，baseline Acc=0.85
  - 经过扩增数据等方式，将NER任务Acc=0.92，F1=0.89
  - 通过Bert预训练模型进行会话主题相关性判定，baseline Acc=0.86
    - 通过字词粒度对比、badcase分析，最终优化Acc=0.97
    - 应用动态量化加快推断速度，Acc=0.91，将单轮响应时间减少45%



- 涉及技术 • BiLSTM+CRF、RNN、Bert、Flask、Neo4j

商城客服辅助接口

NLP算法工程师/项目负责人

- 2018/02 – 2018/07
- 项目介绍：赋能客服部门的聊天系统，根据聊天文本信息，自动录入客户个人信息及感兴趣车型，简化人工繁琐操作，提高效率
- 实现方案
    - 联系方式：电话、微信号，感兴趣车型，有严格规则遵循，使用规则匹配提取
    - 姓名信息二分类模型，对上下文进行检索
      - Fasttext二分类对客服内容判定是否询问姓名。
      - 对下文固定数量的用户会话文本，通过HanLP识别nr实体
    - 将提取的有效信息整合，录入后端数据库
  - 涉及技术 • Numpy、Matpoylib、Fasttext、Hanlp、Django

评论净化系统

NLP算法工程师/算法工程师

- 2017/06 – 2017/12
- 项目介绍：脏话识别过滤系统，过滤社区、商城中用户的输入为脏话的信息，保护社区网络环境
- 制定数据样本格式，方便爬虫工程师收集数据。
  - 基于5000条样本采用“规则+算法”组合方式进行文本判定，对疑似文本进行模型预测。
  - 根据不同模型基于各粒度下的训练指标对比，选择baseline最优的TextCNN
  - 通过扩充数据、badcase分析调整规则:Acc由0.81提升到0.92，召回率：0.90
  - 配合运维部署模型测试与上线
- 涉及技术 • Pytorch、TextCNN、Flask

工作经历

- 2019/05 – 2021/02
- 北京阿里优优科技有限公司 NLP算法工程师
- 2012/03 – 2019/05
- 北京雅邦网络技术有限公司 应用开发&算法工程师
- 2011/02 – 2012/03
- 巨人传众教育科技有限公司 网页游戏开发工程师

教育经历

- 本科
- 北京开放大学 软件工程

自我评价

从事多年互联网开发工作，具备良好的编程逻辑、团队协作、分析解决问题的能力  
为人诚实、上进、有耐心，善于分享  
喜欢跑步、健身，具有良好的综合素质



扫一扫，了解更多