

张真

求职意向: 自然语言处理 | 算法工程师

✉ zhangzhen82301@gmail.com · ☎ (+86) 131-6141-1563 · 🌐 zhangzhen8230

👤 实习经历

腾讯科技, 北京

2018.8 – 2018.11

医疗诊断预测, 论文实验与撰写

- 负责相关论文的调研与复现, 包括 doctorAI, retain, dipole.
- 负责收集不同数据集, 并进行实验, 包括数据集 MIMICIII, MIMICII, CMS.
- 分析标签编码的结构, 在模型中添加引入层次结构知识, 评价指标 top30 召回率从 56.9% 提升到 62.3 %
- 尝试添加时序信息, 评价指标 top30 召回率从基础的 56.9 %, 提升到 58.2%.

🐱 课题项目

流式文档排版格式的智能化分析与优化方法

2017.9–2018.12

国家自然科学基金, 61672105

- 语料库的收集.** 确定标签集合与语料类别, 编写标注插件, 制定相应的 xml scheme, 最终获得 5000 篇标注语料.
- 文档特征的提取.** 提取格式特征 18 种, 内容特征 6 种以及 doc2vec 编码的语义特征, 将格式特征与内容特征进行特征处理.
- 模型搭建.** 针对文档构件类别不平衡问题, 采用规则与神经网络相结合的方法; 对文档类型多种多样等要求, 构造迁移学习模型. 针对文档构件的上下文关系, 使用 LSTM 学习.
- 系统集成.** 完成文档结构自动识别系统, 整合了上几届的代码.

🏆 竞赛经历

天池对话领域竞赛

2018.6–2018.8

“阿里小蜜机器人跨语言短文本匹配算法竞赛”, 队长

- 针对文本匹配目标, 构建跨语言的短文本匹配模型. 从统计特征, NLP 特征和图特征三个方面构造特征工程
- 其中针对于跨语言问题, 采用 auto-encoder 模型编码句子向量.
- 最后采用 Stack 方式进行模型融合, 最终在 1027 个队伍中, 获得了初赛第 5, 复赛第 8 的成绩.

CCF 情感分析竞赛, 32/1186

2018.2–2018.4

景区口碑评价分值预测竞赛, 队长

- 针对旅游景点的评论, 构建评论打分模型. 使用 word2vec, doc2vec 生成句子表示.
- 使用 lstm, textcnn 等模型学习文档特征表示.
- 最后采用 blend 方式进行模型融合, 最终在 1186 个队伍中, 获得了初赛第 13, 复赛第 32 的成绩.

⚙️ 技能

- 编程语言:** 熟悉 python C#, java.
- 深度学习:** 两年 nlp 经验, 熟悉 tensorflow 与 keras 框架, 熟悉常见深度学习模型原理.
- 英语能力:** CET-6 494 分, 良好的英语阅读能力.
- 组织能力:** 北京市优秀班集体 (申请负责人), 优秀团支部, 班集体 (连续四年, 负责人), 优秀团干部 (3 次)
- 学业情况:** 校级二等奖学金 (2 次), 三等奖学金 (3 次), 三好学生

🎓 教育经历

北京信息科技大学 计算机学院, 硕士

2016.9 – 现在

专业: 数字文化传播, 预计毕业日期: 2019.6

北京信息科技大学 计算机学院, 本科

2012.9 – 2016.6

专业: 软件工程