

# 罗周杨

✉ zhouyang.luo@gmail.com · ☎ (+86) 133-8153-6323 · 🌐 <https://github.com/luozhouyang>

## 🎓 教育背景

上海大学

2013.09 – 2017.07

本科, 学士 通信学院电子信息工程

## 💼 工作经历

e 成科技 | <https://ifchange.com>

NLP 算法工程师

2018.12 – 2020.06

项目: JD 和 CV 的文本相关性

- 使用千万级别的 JD 和 CV 的文本内容, 经过分句和筛选处理, 整理得到预训练数据。把 BERT 预训练的 NSP 任务替换成 JD 和 CV 是否匹配的二分类任务。然后进行预训练, 匹配任务的训练集准确率达到 0.94, 测试集准确率 0.92, 比之前线上的效果提升 3 个百分点左右。

项目: JD 和 CV 的关键词抽取

- 从上述预训练 BERT 模型中, 抽取出 Attention 权重作为词语的权重, 特殊 Token 的权重不小, 并且不同层的 Attention 权重差别较大。对比之后选择最后一层, 并且去除特殊 Token。
- 使用基于 Sentence Embedding 的 EmbedRank 模型进行关键词抽取, 线性组合两者权重, 对比之前线上效果, 无意义词语显著减少, 重要词语的权重区分更明显。

项目: 细分领域的职能识别

- 在上述预训练的 BERT 模型之上, 构建 NER 模型进行微调。解决大部分细分领域 (通用的 NER 覆盖不到) 的职能识别不准的问题, 显著提高了 CV 的召回, 以及 JD 和 CV 之间的推荐度。

项目: JD 和 CV 的推荐工程

- 优化推荐工程性能, 并行请求 ES 和处理 CV, 响应时间提升 5 倍。优化所涉及的基础算法服务。

51job | <https://51job.com>

软件开发工程师

2017.07 – 2018.12

项目: 名片地址文本纠错

- 为了解决 jieba 对道路等文本分词效果不佳的问题, 实现 BiLSTM + CRF 的神经网络分词模型, 对于地址文本的分词效果提升明显。 <https://github.com/luozhouyang/deepseg>
- 使用基于 RNN 的 Seq2Seq + Attention 的生成式模型来实现纠错, 测试集准确率达到 0.86。使用 Tensorflow Serving + Docker 部署模型, 平均推断时间 50ms (GPU)。

## </> 开源项目

- luozhouyang/python-string-similarity: 多种字符串相似度的衡量算法
- luozhouyang/transformers-keras: 基于 Transformer 的多个模型实现
- luozhouyang/embedrank: 基于 Sentence Embedding 的无监督关键词抽取

## ⚙️ 个人技能

- 编程语言: 3 年 Python 使用经验, 作为脚本和构建深度学习模型, 有使用 Flask、FastAPI 构建 Web 服务的经验; 4 年 Java 使用经验, 熟悉多线程和常见的设计模式, 有 Spring Boot 构建 Web 服务经验; 2 年 Android 应用开发经验 (大学期间, Java 语言)
- 工具相关: 常年在 Linux 环境下工作, 熟悉 Docker 使用, 熟悉 Git 及其协作

## 📄 其他

- GitHub: <https://github.com/luozhouyang>
- 博客: <https://luozhouyang.github.io>