

赵臻宇

(+86) 18500234600 shuaihuaiyi
shuaihuaiyi@qq.com zhaozhenyu1996@outlook.com
github.com/shuaihuaiyi 哈尔滨工业大学逸夫楼 111, 150001
1996 年 5 月出生于山西省昔阳县

教育背景

2018 年 9 月 - 2020 年 6 月 硕士研究生在读, 计算机科学与技术专业, 计算机科学与技术学院, 哈尔滨工业大学
机器智能与翻译研究室, 导师: 杨沐昀老师
2014 年 9 月 - 2018 年 6 月 本科, 计算机科学与技术专业, 英才学院, 哈尔滨工业大学

学术论文

> 赵臻宇, 杨沐昀, 赵铁军, 汤步洲 “基于集成学习的医疗文本去隐私系统” CCIR 2018.

命名实体识别 集成学习 规则挖掘 LSTM CRF

相关技能

编程技能: 熟悉 Python
了解 Java, C/C++, Scala, SQL, \LaTeX
机器学习: 方向 自然语言处理
框架 TensorFlow v1.x, PyTorch, Keras
模型 Seq2Seq+Attention, Transformer (特别是 OpenNMT-py 中的实现)
开发工具: PyCharm, Git, Vim, Visual Studio

实习经验

2018 年 8 月 自然语言处理算法实习生, 知乎, 文本生成
2018 年 1 月
> 主要负责基于 Seq2Seq+Attention 的问题、评论、摘要等多种文本生成类任务
> 用 Pointer-Generator Networks 实现文本摘要
> 用 CNN 模型对文本完整性进行分类
> 主要积累了大量工程经验, 包括常见的数据预处理、多线程计算、远程调试等等
> 实习期间曾在 2018 年 DeeCamp 夏令营担任产业导师, 指导两个小组进行回复生成相关任务
自然语言处理 文本生成 Seq2Seq Pointer-Generator Networks Python

项目经历

2018 年 6 月 基于模型融合的医疗文本去隐私化研究, 本科毕业设计, 命名实体识别
2017 年 12 月
> 任务目标: 标记电子病历中的隐私信息
> 基分类器: 规则挖掘、LSTM、CRF
> 集成算法: Stacking
> 性能结果: 创造了新的 state-of-the-art (95.11 \Rightarrow 96.56)
命名实体识别 Stacking 集成 LSTM CRF 规则挖掘

现在 对机器翻译 Transformer 模型的改进和探索, 近期主要工作, 机器翻译
2018 年 12 月
> 令语义相似但语序不同的源语言句子产生相似的输出
> 为 Transformer 添加预调序模块, 自动产生不同顺序的句子
机器翻译 Self-attention Transformer

语言

英语: 阅读 ●●●●● 4 级: 520 (2014 年)
写作 ●●●○○ 6 级: 449 (2015 年)
口语 ●●●○○

奖励与荣誉

2018 年 9 月 哈尔滨工业大学优秀推免硕士生专项奖学金
2018 年 7 月 哈尔滨工业大学百优毕业设计 (论文) 奖
2016-2018 年 哈尔滨工业大学人民奖学金 3 次 (<3%).

(最近更新: 2019 年 6 月)