



基本信息

姓名: 赵煜	出生年月: 1997.01
电话: -----	学历: 硕士研究生
邮箱: zhaoyu@sniper97.cn	博客地址: http://www.sniper97.cn
Github : https://github.com/Sniper970119	岗位意向: NLP 算法工程师

教育背景

2019.09-2022.06	大连理工大学	软件学院	软件工程 (硕士)
2015.09-2019.06	大连交通大学	软件学院	软件工程 (本科)

项目经历

2021.08-2022.03 ----- 食品安全问题检测

> 检测商家与顾客对话中的食品安全问题
 >> 由于对话中食品安全问题数量较少,标签不均及领域外对话较多, 容易造成任务转移(如模型可能把“识别食品安全问题”转换为“识别食品问题”)因此通过打伪标签, 构建领域内负样本等方式, 构建较为均衡的数据从而在一定程度上缓解了由于任务转移造成的问题(acc+4%), 解决了模型可用性;同时通过使用 F1 loss+ 交叉熵(acc+0.5%, F1+0.02),EMA(acc+1%,F1+0.002),FGM(acc+0.8%,F1+0.006),最终 acc 从 83.5%→89.8%, F1 从 0.68→0.8558

2021.06-2021.08 SMP SMP-ECISA-2021

> 隐性情感分析
 >> 通过使用多个 trick, 包括: 融合上下文, 回译, 伪标签, R-Drop, FGM, EMA, random mask, whole word mask, (bert)最后两层求和, 损失函数中添加 F1 等方式; 并尝试使用 Bert, Roberta, Albert, Nezha, Wobert, Roformer 等模型进行模型融合。最终 A 榜第 1, B 榜第 4, 获三等奖。多模 Top2%

2021.05-至今 实验室 论文评级 (AAPR)

> 目前学术论文打分任务方向论文较少, 该任务通过对论文质量进行综合分析, 判断论文是否会被录取 (独立完成)
 >> 由于该方向的开山之作(ACL2018)的 Github 开源代码缺少关键模型结构文件不可运行, 因此需要对模型结构进行复现。最终:通过改进原论文的数据处理方式, 在不增加字段的情况下使模型能够获取到更多的论文信息(acc+≈0.5%); 同时通过对模型的 CNN 结构添加注意力和残差(acc+≈1%); 替换 embedding, 最终超越原文 acc 2%以上。同时发现, 由于数据集是在 arxiv 上进行的采集, 相对质量较低(录用结果不准确), 目前在构建更准确的数据集

2021.01-2021-02 天池 全球人工智能技术创新大赛【赛道三】

> 使用脱敏数据对文本进行相似度计算 (独立参赛)
 >> 通过数据增强与将脱敏数据与开源数据集匹配, 从而以更小的代价在脱敏数据上使用 Bert。单模 Top10%

2020.11-2020.12 DCLab 中移集成 OneCity 编程大赛

> 智慧政务, 对公文进行多分类 (独立参赛)
 >> 通过数据格式适配, 优化数据解析, 数据增强等方式, 使用 Bert 对文本进行分类。单模 Top10%

实习经历

2022.06-至今 某外卖/-----/-----组 NLP 算法工程师

> 负责某外卖平台文本智能相关工作

2021.06-2022.03 某外卖/-----/-----组 NLP 算法工程师 (实习)

> 负责某外卖平台客服辱骂识别, 食品安全问题检测, 差评申诉等

其他经历

有一个自己的技术博客(<http://www.sniper97.cn>; 月访客量过万, 月点击量数万), 用于日常总结学习总结, 包括深度学习, java 虚拟机, 设计模式等。目前深度学习相关的笔记达 70 余篇, 总文章数近 300 篇 (2022.07)

阅读过 transformers(hugging face)和 bert4keras 的源码, 为 transformers 的 contributor, 并对 bert4keras 进行详细技术文档编写 (已开源)

工程能力强, 有较多代码编写经验, 包括 python java 等, 具备良好的代码风格与习惯。Github 最多项目获得 420 余 star 和 80 余 fork. <https://github.com/Sniper970119> (2022.07)

技能证书

IBM 高级软件工程师

两次 (第八届第九届) 蓝桥杯算法设计大赛省一等奖; 一次 (第八届) 国家二等奖

