

傅科达

✉ fukeda_hit@outlook.com · ☎ 188-4615-1787

🎓 教育背景

哈尔滨工业大学	计算机 SCIR 硕士	2016.09 – 2018.07
哈尔滨工业大学	软件工程学士	2012.09 – 2016.07

👥 项目经验

作文论述段生成 (pdf) 2022.5 – 2023.5

- 任务: 输入学生摘要句, 利用 chatglm 微调生成使用显示论证的论述段, 给学生参考
- step1: 论证指示词挖掘。中档分作文, 【A】摘要句召回论述段相似的 6 句话, 【B】tf-idf 模型获得摘要句主题词过滤召回的句子 【C】用 apriori 算法从召回句子挖掘相关指示词
- step2: 构造 instruct。摘要句利用 tf-idf 获得主题词 + 指示词 + 摘要句
- step3: chatglm lora 微调生成论证内容

作文分论点打分 (pdf) 2021.07 – 2022.4

- 任务: < 学生摘要句 1, 参考分论点 123>, 学生摘要句选择最佳匹配, 计算对应得分 0-3 分
- step1: 设计排序、打分 2 个标注标准, 开发标注工具, 标注 2w 条样本
- step2: 选择最佳匹配。挖掘难负样本构建 100w 训练集, 3 层 BERT, simcse 和 pairwise, acc 0.92
- step3: 匹配打分。挖掘 60w 训练样本, 6 层 BERT, pearson 0.86
- step4: 匹配可解释性。结合规则命中, 尝试用积分梯度获得哪些词是得分关键点
- 业务: 结合原有规则, 在作文摘要句匹配任务上, F1 0.37->0.66

作文分论点识别 (pdf) 2021.1 – 2021.06

- 任务: 找出公文论述段的分论点
- step1: 数据增广: 标注数据 1w, 每个摘要句与段落内周围句子交换, 扩增至 20w
- step2: 构造输入:[段 1 前 80, 段 2 前 80, 段 3 前 80, 标题], 输出:[段 1 摘要句位置,..., 段 3...]
- step3: 迁移学习两阶段训练。魔改模型: BERT-SUM+memory+3 层 transformer。F1 0.89
- 业务: 在作文分论点匹配业务中, 召回率不变情况下, acc 0.43->0.80

文章分句 (pdf) 2020.04 – 2020.12

- 任务: 学生答案中存在分点作答、一逗到底的情况, 用户投诉太多
- step1: 针对分点作答, 利用多个正则识别多个批次的【顺序位置】, 将顺序位置划分为 ABC 等级, 归并 K 个有序数组。acc 0.81->0.92, 速度 $\times 10$
- step2: 一逗到底, 模型: BERT+ASR->BERT+剪枝->BiLSTM->IDCNN, F1 0.21->0.78, 速度 $\times 6$
- 业务: 没有再出现因为分句导致的 badcase 投诉

🏢 工作经验

粉笔教育 作文批改 NLP 工程师 2020.04 – 至今

- 利用参考答案对用户提交的试卷进行语义批改

搜狗搜索 相关性组 助理研究员 2018.07 – 2019.03

- 设计新特征用于计算查询词和文档标题的相关性

🏆 竞赛

全国二等奖 第六届大学生数学竞赛 2014.11

⚙️ 个人能力

- 熟练掌握各种 NLP 任务, 在解决方案、模型设计均有创新, 熟悉学术界前沿预训练、对比学习等