

李卓君

18845636557 | zhuojunlee@foxmail.com



教育经历

- 哈尔滨工业大学 985 2022年09月 - 2024年12月
计算机科学与技术 硕士 计算学部 哈尔滨
荣誉奖项：研究生特等学业奖学金
- 哈尔滨工业大学 985 2018年09月 - 2022年06月
计算机科学与技术 本科 英才学院 哈尔滨
GPA: 91.19/100
荣誉奖项：一等人民奖学金5次、二等人民奖学金2次、优秀学生（2019-2020）

项目经历

面向新闻文本的立场检测系统

本项目属于针对新闻文本的抽取与分类任务，目标是从海量新闻文本中抽取含有<立场持有者，立场表达词，立场目标>的三元组结构的句子，并且对于立场极性进行分类。

方案：

- 参考ACL 2021工作，使用BART和指针网络构建pipeline模型，在MPQA数据集上达到与SOTA可比水平；
- 复现SemEval 2022 Task10结构化情感分析评测第一名方案，微调roberta-large构建稀疏矩阵表示图结构；
- 从零构建出一套针对中文新闻文本的立场检测解决方案，制订了相关数据标注方案并组织数据标注逾10000条，在语义角色标注和指代消解等技术的基础上使用RoBERTa-wwm进行微调，构造指令数据集并使用Bloom 7B进行指令微调；

比赛经历

“人民网”2021人工智能算法挑战赛—虚假信息检测赛道亚军

比赛提供虚假信息检测数据集，数据集输入为待核查真实性的短文本，数据标签根据虚假程度分为6个类别，分别为极度虚假、虚假、大部分虚假、半真半假、大部分真实、真实，另外有一个事实审查数据集作为补充，包含对于短文本的真假性分析，但是与待核查数据集间缺少映射关系。

方案：我们的系统主要是将该问题建模为选项式阅读理解问题。将待核查文本视为问题，从事实审查数据中使用BM25和模糊匹配检索匹配的证据篇章，然后再将<待核查文本，证据篇章>作为输入，将6个类别视为选项，从而把原问题转化为阅读理解问题。最终通过微调DeBERTa、T5可以接受长文本输入的模型在最终榜单排名第二。

AI创意派：百度认知AI创意赛-创意开发组冠军

利用百度提供的文心ERNIE 3.0中的命名实体识别、关系抽取和事件抽取功能实现了对文言文中的人名、官职、地名、人物关系和重要历史事件一键解析，并且使用爬虫技术将识别实体链接到百度百科进行展示，便于读者更好理解文言文。

论文发表

HIT-SCIR at WASSA 2023: Empathy and Emotion Analysis at the Utterance-Level and the Essay-Level, ACL 2023 WASSA Workshop, 共同第一作者

本工作来源于ACL 2023 WASSA Workshop，任务为对于志愿者阅读新闻文章后所写的读后感进行共情和情绪分析，设置有5个Track，我负责共情预测回顾任务和情感分类任务，使用RoBERTa微调和ChatGPT数据增强，最终分别取得5/9, 3/13的名次。

其他

- 技能：Python, PyTorch
- 语言：英语（CET-6）