

上海交通大学

本科生毕业设计（论文）任务书

课题名称：舆情文本多分类与敏感性检测技术研究与实践

执行时间：2019 年 12 月 至 2020 年 6 月

教师姓名：唐新怀 职称：副研究员

学生姓名：周笑 学号：516030910370

专业名称：软件工程

学院(系)：电子信息与电气工程学院

毕业设计（论文）基本内容和要求：

内容：结合 BERT 预训练模型对实验室舆情分析平台中的舆情文本内容进行多分类。研究如何结合 BERT 技术检测中文舆情文本敏感性并提升中文舆情内容多分类准确率。将相关研究成果应用于舆情事件监控领域。要求：

（1）

研究基于评价指标准确率（accuracy）的中文舆情内容长文本分类，基于 THUCNEWS 中文新闻长文本标注数据集（http://106.13.187.75:8003/download/?dataset_name=thucnews）的准确率不低于 95.35 （2） 研究基于评价指标准确率（accuracy）的中文舆情内容短文本分类，基于今日头条中文新闻短文本分类数据集（http://106.13.187.75:8003/download/?dataset_name=tnews）的准确率不低于 89.78 （3） 研究基于中文舆情文本关键词、语义的敏感内容检测技术。 （4） 在现有舆情事件分析系统中集成上述方法。

毕业设计（论文）进度安排：			
序号	毕业设计（论文）各阶段内容	时间安排	备 注
1	文献阅读，需求分析	2019 年 12 月 -2020 年 2 月	
2	系统设计，关键技术研究	2020 年 2 月 -2020 年 3 月	
3	系统实现、调试	2020 年 3 月 --2020 年 5 月	
4	系统测试，上线运行，论文撰写	2020 年 5 月 -2020 年 6 月	
5	论文撰写、修改、答辩	2020 年 6 月	

课题信息：

课题性质：设计 ☒ 论文 ☐

课题来源*：国家级 ☐ 省部级 ☐ 校级 ☐ 横向 ☐ 预研 ☒

项目编号 _____

其他_____

指导教师签名：____唐新怀_____

2020 年 3 月 20 日

学院（系）意见：

审核通过

院长（系主任）签名：步丰林

2020 年 3 月 20 日

学生签名：周笑

2020 年 3 月 20 日