

序号：_____

编码：_____

2023年暨南大学第十届“挑战杯”大学生 课外学术科技作品竞赛 作品申报书

作品名称：_____基于加性注意力和抽取式文本摘要的双路文本情感分析模型

学院名称：_____信息科学技术学院/网络空间安全学院院级

申报者姓名：_____林文豪

类别：

- ☐ 自然科学类学术论文
- ☐ 哲学社会科学类社会调查报告和学术论文
- ☒ 科技发明制作类

申报者情况及作品信息

申报者代表情况	姓名	林文豪		性别	男	出生年月	2003-03-05	
	学校	暨南大学						
	院系	信息科学技术学院/网络空间安全学院院级			专业	网络工程		
	年级	本科二年级			学制	4	入学时间	2021
	作品全称		基于加性注意力和抽取式文本摘要的双路文本情感分析模型					
	毕业论文题目							
	通讯地址	暨南大学番禺校区T5 915房					邮政编码	510632
							宿舍电话	
常住地通讯地址	广东省深圳市宝安区西乡桃源居12区13栋2单元803					邮政编码	518102	
						移动电话	18319380470	
其他作者情况	姓名	性别	年龄	学历	所在学院			
	洪梓晋	男	21	本科三年级	暨南大学伯明翰大学联合学院			

指导老师情况	姓名	性别	年龄	职称	所在单位
	黄斐然	男	35	副教授	暨南大学网络空间安全学院
资格认定	学院学籍 管理部意见	以上作者是否为正式注册在校的全日制非成人教育、非在 职的高等学校专科生、本科生、硕士研究生或博士研究生 <div><input type="checkbox"/>是 <input type="checkbox"/>否</div> <div>(签章) 年 月 日</div>			
	指导老师 意见	本作品是否为课外学术科技或社会实践活动成果 <div><input type="checkbox"/>是 <input type="checkbox"/>否</div> <div>指导老师签名： 年 月 日</div>			

<p>作品简介</p>	<p>我们提出的基于加性注意力和抽取式摘要的双路文本情感分析模型有三层结构。嵌入层基于观点和事件提取对长文本进行摘要提取。特征提取层使用加性注意力机制代替自注意力机制，利用Bi-LSTM和TextCNN构建双路特征提取网络。预测层将两个文本特征融合，用Softmax得出分类结果。模型构建完成后，通过对不同场景的数据进行高质量标注，并使用不同摘要提取方法来进行进一步优化输入，根据实验对模型结构提出更多优化策略</p>
<p>作品的设计、发明的思路和关键技术指标</p>	<p>（一）目的：长文本情感分析包含更丰富的信息，运用更广泛，本模型力求实现对长文本的高效训练和对文本情感倾向的高效识别。</p> <p>（二）基本思路</p> <ol style="list-style-type: none"> 1.词嵌入层中，基于观点提取和事件抽取对长文本摘要，用 Word2vec 词向量工具得到双路文本向量。特征提取层中，用加性注意力机制代替自注意力机制，融合Bi-LSTM与加性注意力机制，结合 TextCNN 用三种特征提取方法提取文本特征并融合。输出预测层中，将提取的两个文本向量进行特征融合，用 Softmax 函数得到最终分类结果 2.模型构建完成后，对不同长文本场景进行数据标注，由预测结果总结出特定场景的显著特征。得到不同数据的结果后，对不同场景的长文本特征运用不同的摘要提取方法，进行模型嵌入部分的优化。确保模型的运行和预测效果，对模型结构参数进行调优 <p>（三）创新点</p> <ol style="list-style-type: none"> 1.基于观点提取和事件抽取对原始文本进行摘要。用加性注意力机制代替自注意力机制，用 Bi-LSTM与加性注意力机制融合，结合TextCNN构造出双路特征提取网络 2.输出预测层中，将提取到的两个文本向量进行特征融合，用 Softmax 函数得到最终分类结果。构建高质量标注的数据集，评估模型在不同场景中的表现。结合特定使用场景，对模型的参数以及结构进行调优 <p>（四）技术关键与主要技术指标：技术关键是模型的构建，不同场景下的摘要提取，高质量的数据标注，以及优化的方案的实施。主要技术指标是F1 score</p>
<p>作品的科学性、先进性</p>	<p>本作品旨在解决目前存在的长文本特征提取精确性问题</p> <ol style="list-style-type: none"> 1.通过抽取式摘要，基于观点提取和事件抽取，减少预处理中信息流失，起到数据增强作用 2.用加性注意力机制代替自注意力机制，让模型在不损失精度情况下，面对长文本训练能有更出色的训练速度、更佳的并行计算效率和泛化能力 3.用Bi-LSTM与加性注意力机制融合作为该模型的单路特征提取网络，选取 TextCNN 作为二路特征提取网络，构造双路特征提取网络。双路网络并行的方式提取特征，能够提高分类准确率 4.输出预测层中，将提取到的两个文本向量进行特征融合，用 Softmax 函数得到最终分类结果，从而得到对于长文本特征更精确的情感分析和预测结果 5.模型搭建完成后，构建高质量标注的数据集，对模型在不同场景的表现上进行效果评估，从而映射到目前不同的情感分析任务当中，得到模型对于不同任务的适配程度以及指标 <p>结合上述特定的使用场景，对模型的参数和结构进行调优，对于不同的场景使用不同的文本摘要抽取方式和不同的网络结构，实现对模型的优化以及改进</p>
<p>作品在何地、何时、何种场合、何种方式、何种平台、何种渠道、何种活动、何种竞赛、何种评比、何种评审、何种展示、何种推广、何种应用、何种成果</p>	<p>无</p>

作品所处阶段	已查阅大量相关文献，复现了部分模型，并基本上搭建出我们的模型，提出了优化策略，标注了部分数据集。
技术转让方式	授权或联合研发。可以考虑与相关公司或研究机构进行合作，共同推进模型的发展和应用。
作品的可展示形式	论文、图片
使用说明书及推广点，作品应推广的说明及经济效益预测	<p>一、使用说明 输入长文本，模型将对其进行情感分类</p> <p>二、技术特点</p> <ol style="list-style-type: none">1. 基于观点提取和事件抽取对文本摘要2. 用Bi-LSTM与加性注意力机制融合作为模型的单路特征提取网络，并用TextCNN 作为二路特征提取网络，构造出双路特征提取网络3. 将双路文本向量进行特征融合，用Softmax函数得到最终分类结果4. 构建高质量标注的数据集，评估模型在不同场景中的表现5. 结合特定使用场景，对模型参数和结构调优 <p>三、技术优势</p> <ol style="list-style-type: none">1. 通过摘要可以去除冗余信息，减少预处理中信息流失，起到数据增强的作用2. 双路特征网络让模型在面对长文本训练能够有较为出色的训练速度、更佳的并行计算效率和泛化能力。3. 标注数据集有利于优化模型 <p>四、适应范围 长文本的情感分析</p> <p>五、推广前景的技术性说明及市场分析和经济效益预测 文本情感分析可以自动分析文本中的情感倾向，具有广泛的应用前景。随着社交媒体和电子商务的普及，对于情感分析的需求日益增加。我们预计本模型的推广可以为企业带来可观的经济效益。</p>

<div>专利申报 情况</div>	<p>暂未提出专利申请，在论文撰写完成后会申报相应专利。</p>
<div>国内课程 研究概述 当前题平</div>	<p>在文本情感分析领域中，目前对长文本的情感倾向分析有许多不同的方法类别，也有着各自不同的进程和发展：</p> <p>首先是基于词典的情感分析，是文本分析将文本任务的一种传统方法。将文本中中出现的情感词的数量和情感词的极性作为衡量文本情感倾向的依据。词典的构建过程需要耗费大量的人力和时间，情感词典往往缺乏上下文信息，难以应对多样化的情感表达方式。</p> <p>其次是基于规则的情感分析，该方法通过制定一系列规则和规则库来判断文本的情感倾向。这种方法可以使用人工制定的规则或使用机器学习和深度学习等技术来自动学习规则。现阶段，该方法普遍运用于特定场景中，针对不同业务制定相应的规则。</p> <p>再者是基于机器学习的情感分析，使用文本中的词汇、词性、词频、语法结构等特征来建立情感分类模型。该方法需要进行特征选择和模型训练等预处理步骤，但可以处理大规模的文本数据。机器学习的方法中，长文本的数据量通常较大，因此需要更多的标注数据，这也增加了数据收集和标注的难度和成本。基于机器学习的方法通常缺乏对情感分析任务的解释性。虽然这些方法可以高效地处理大量数据，但是对于处理结果的解释和理解却存在困难。在如此场景下提高标注数据的可靠性和数据收集的效率，同时结合深度学习等新技术来提高方法的性能和解释性，成为了优化机器学习长文本情感分析的关键路径。</p> <p>最后则是近年来该领域的研究热点，基于深度学习的情感分析，即使用神经网络来建立情感分类模型。这种方法不需要进行特征工程，可以自动学习文本的特征。深度学习方法在长文本情感分析方面表现较好，但是需要大量的计算资源和训练时间。最近的研究尝试使用注意力机制来进一步提高方法的性能，在文本情感分析领域中取得了重大的突破，进而通过深度学习的分析方法成为了主流热点，在一定的算力资源加持下，由于其表现尚佳，也成为了今后在文本情感分析任务中研究优化的首要选择。本项目则正是基于深度学习方法，提出优化方案。</p> <p>总的来说，长文本情感分析面临着一些挑战和缺陷，例如长文本的特征提取、情感表达的多样性、预测的可解释性和语言的多义性等问题，这些问题需要更加精细的建模和深入的研究。本项目通过观点提取，事件提取等摘要方法，结合双路的特征提取神经网络，在数据以及模型两方面着力解决长文本的特征提取问题，通过解决该问题来提高长文本情感分析的任务精度，该方案可行性强，目标明确，是解决问题的良好方案。</p>

学院意见	<div></div> <div>(签章)</div> <div>年 月 日</div>
评审委员会意见	<div></div> <div>(签章)</div> <div>年 月 日</div>
学校领导小组意见	<div></div> <div>(签章)</div> <div>年 月 日</div>