**开题报告**

# 论文题目：基于图卷积网络的情感依赖模型在方面级别情感分类中的应用

学院：\_\_计算机学院\_\_\_\_学号：\_\_\_\_\_2018302110379\_\_姓名：\_\_陈威\_

1. **毕业论文（设计）题目的来源**

本题目的主要重点在于情感分类，其前缀主要说明情感分类的方法以及方式，所以其来源是指情感分类是在何种环境下诞生的需求。

随着大数据环境以及社交媒体的高速发展，越来越多的人在社交媒体上发布自己的看法，其带来的发展前景与商业价值引来了社会上大部分人们的共同关注，由此情感分类诞生。

1. **毕业论文（设计）选题的目的和意义**

近年来，情感分析成为自然语言处理（ＮＬＰ）领域中最活跃的研究方向之一，在信息检索、数据挖掘、文本挖掘等方面得到广泛研究。文本情感分析是对带有情感色彩的主观性文本进行分析、处理、归纳和推理的过程。随着社交媒体和大数据的快速发展，情感分析已经从计算机科学蔓延到其他学科领域。情感分类是情感分析技术的核心，其任务是判断文本的情感取向，根据文本所表达的含义和情感信息将文本划分成积极或消极２种或多种类别。按照情感的粒度可以分为３种分类问题：二分类（积极和消极）、三分类（积极、消极和中立）和多分类。例如对微博评论进行“开心”、“悲伤”、“愤怒”、“惊讶”的四元情感分类就属于多元分类问题。

按照对文本研究粒度的不同可以分为３种不同级别：篇章级情感分类、句子级情感分类和方面级情感分类。

情感分类完成后，可以对句子的说话人态度进行判断，例如说话人对电影进行评论后经过情感分析可以知道话说人对该电影的态度，由这点可以扩大到整个电影影评，抽离出贬义部分，贬义部分即包含了该电影的缺点而后汇总，即可整理出官方推荐。这只是其中一种用法，带来的收益也是十分明显，在社交软件流行的今天，通过此方法可以得知大部分人对某件事物的看法，再进行分析得到有用的数据。

1. **国内外关于该选题的研究现状和发展趋势**

对于研究领域中主流的情感分类方法，可按照技术路线将其分为２大类：传统方法和深度学习方法。

传统方法又分为基于词典的方法、基于机器学习的方法、基于弱标注信息的方法、词典与机器学习相混合的方法。传统分类方法主要针对篇章级和句子级进行粗粒度的情感分类，只能假定文本只有单一的情感（例如正向和负向），不能对文本的多个方面（即用户从哪个角度评论，或者商品从哪个角度介绍，例如价格、性能、服务等）进行情感识别。

近年来由于神经网络的快速发展，基于深度学习的情感分析取得了良好的效果，该方法的最大优势是不依赖人工定义特征，可以进行端到端的自主学习。当前比较热门有传统深度学习情感分类方法和基于图域的深度学习方法即图神经网络方法（基于图卷积网络）。

图卷积网络ＧＣＮ是一种简单有效 的基于图的卷积神经网络，可以从丰富的关系数据中捕捉相互依赖的信息，在当前基于图神经网络的情感分类方法研究中，大多数研究人员应用了经典的图卷积网络ＧＣＮ提高分类模型的算法性能，表现出很大潜力，而本选题选用的即时图卷积神经网络方法。

在情感分类研究领域，方面级情感分类发展迅速，在单领域和单一语言上不断取得成果，但在一些方面仍面临挑战。例如由于语言的多样性与复杂性，不同领域间的情感表达差别较大。特别是当领域差异过大时，如将商品评论领域训练的情感分析模型应用到新闻领域上，情感分析的效率和性能则会下降。因此，训练高效的通用领域情感分析模型非常值得研究。同时，在近两年开始崭露头角的基于图神经网络的方法所展现出的巨大潜力，预示着在今后的研究中，图神经网络可能成为解决现有问题的一种有效途径，使得方面级情感分类得到进一步提升。

1. **毕业论文（设计）的主要内容、研究方法、技术路线及可行性分析**

该选题需要我们首先完成图卷积网络的情感依赖模型，而关于这点，现有的技术中有许多对于图卷积网络情感模型的优化，本选题将会从中选取一种进行实践。

在方面级情感分类中，我们需要判断一个句子中的相关语义和方面词汇的关系，譬如：在“The price is reasonable though the service is poor”这句话中，对于“price”和“service”这两个方面的词语，情绪的两极分化分别是正面和负面的,这就属于二分类问题，关于在该选题中选用的情感分类粒度问题目前仅属于二分类问题，在后续计划中可能有所更改。

方面级的情感分类可以表示为:给定一个输入上下文由N个单词组成 ， K个方面项。每个方面是句子的子序列，其中包含单词。需要构造一个情绪分类器来预测多个方面术语的情绪极性，而图卷积网络正是其中一种方法。

图卷积网络在现有技术中已有许多新技术将其优化，如下：

ＡＳＧＣＮ：在句子的依赖树上构建图卷积网络（ＧＣＮ），以利用句法信息和单词依赖。

ＳＤＧＣＮ：在位置编码的双向注意力机制上，用 ＧＣＮ 捕捉句子中不同方面之间的情感依赖。

ＳＡ－ＧＣＮ：在依赖树上的ＧＣＮ模型的基础上，使用多头注意力为句子中每个单词选择注意。

ＡＥＧＣＮ：由多头注意力和在句子依赖树上建立的改进图卷积网络组成。

在该选题中，我们选用了ＳＤＧＣＮ来作为情感分析模型的方法。关于SDGCN的总体结构由输入嵌入层、Bi-LSTM、位置编码、双向注意机制、GCN和输出层组成。我们从输入到输出依次使用所有组件。

在实际模型训练部分，目前我们有两个选择，基于glove的模型和基于bert的模型。

基于Glove的模型结合了两个主要模型族的优点:全局矩阵分解和局部上下文窗口方法。模型只训练单词-单词共现矩阵中的非零元素，而不是整个稀疏矩阵或单个上下文窗口的大型语料库，从而有效地利用统计信息。该模型生成了一个具有意义子结构的向量空间，其在最近的一次单词类比任务中的性能为75%。在相似性任务和命名实体识别方面，它也优于相关模型。

基于Bert的模型使用了Transformer的双向编码器表示。与其他语言表示模型不同，BERT旨在通过联合调节所有层中的上下文来预先训练深度双向表示。因此，预训练的BERT表示可以通过一个额外的输出层进行微调，适用于广泛任务的最先进模型的构建，比如问答任务和语言推理，无需针对具体任务做大幅架构修改。

当然还有其他的优秀的模型可以选择，但经过资料查证，这两种模型更适合我们ＳＤＧＣＮ的发挥。

目前来说，我们现在的工作是需要利用数据集来训练出模型，也就是基于glove的模型和基于bert的模型，而后得到模型以及图神经网络的配合的SDGCN-BERT，另一模型也是同理，对比两模型之间的差距，选出优于另外一种的模型，通过后续设计出模型的应用方法与应用试验。

1. **毕业论文（设计）的进度安排**

|  |  |  |  |
| --- | --- | --- | --- |
| **序号** | **毕业论文（设计）各阶段内容** | **时间安排** | **备 注** |
| 1 | 查找资料以及整合资料 | 21.12-22.1 |  |
| 2 | 数据集收集以及预处理 | 22.1-22.1 |  |
| 3 | 模型建立以及调优 | 22.1-22.2 |  |
| 4 | 情感分类应用 | 22.2-22.2 |  |
| 5 | 整合过程撰写报告 | 22.2-22.3 |  |

1. **毕业论文（设计）的主要参考文献**

[1]PinlongZhaoetal.“Modelingsentimentdependencieswithgraphconvolutionalnetworksforaspect-levelsentimentclassification” KnowledgeBasedSystems 193(2020):105443.

[2]李胜旺,杨艺,许云峰,张妍.文本方面级情感分类方法综述[J].河北科技大学学报,2020,41(06):518-527.

1. 王光,李鸿宇,邱云飞,郁博文,柳厅文.基于图卷积记忆网络的方面级情感分类[J].中文信息学报,2021,35(08):98-106.
2. 王婷,杨文忠.文本情感分析方法研究综述[J].计算机工程与应用,2021,57(12):11-24.
3. **其他要求**