

**本科生毕业论文（设计）**



**题 目 基于BERT的中文电影评论情感分析系统设计与实现**

**学 院**

**专 业**

**学生姓名**  **此处填写学生姓名**

**学 号**  **\*\*\*\*\*\* 年级 2021**

**指导教师**  **此处填写指导教师姓名**

**教务处制表**

**二〇二五 年 五 月 二十 日**

**基于BERT的中文电影评论情感分析系统设计与实现**

学生 □□□ 指导老师 □□□，□□□

**摘要：**随着互联网的普及，在线电影评论已成为观众表达观影体验和意见的重要途径。情感分析技术的进步使得自动化分析用户评论的情感倾向成为可能。本研究基于BERT（Bidirectional Encoder Representations from Transformers）模型，构建了一个中文电影评论情感分析系统。系统涵盖数据采集与预处理、BERT模型微调、后端API开发、前端可视化展示及系统部署等关键环节。实验结果表明，基于BERT的情感分析模型在中文电影评论数据集上的准确率优于传统方法。该系统可广泛应用于电影推荐、观众反馈分析及市场趋势预测等领域。

**主题词：**情感分析；自然语言处理；BERT

**Evading Anti-Malware Engines with Deep Reinforcement Learning**

Cybersecurity

Student： Zhiyang Fang Adviser: □□□

**Abstract：**□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□.

**Key Words：**□□□□□；□□□□□；□□□□□；□□□□；□□□□□□；□□□□□□；□□□□□；□□□□□□。

**目 录**

[1 绪论 2](#_Toc195481703)

[1.1 研究背景 2](#_Toc195481704)

[1.2 国内外研究现状 2](#_Toc195481705)

[1.3 论文主要工作 2](#_Toc195481706)

[1.4 论文组织与结构 2](#_Toc195481707)

[2 背景知识简介 3](#_Toc195481708)

[2.1 知识和技术要点一 3](#_Toc195481709)

[2.2 知识和技术要点二 3](#_Toc195481710)

[3 系统分析与设计 3](#_Toc195481711)

[4 系统详细设计与实现 3](#_Toc195481712)

[5 工作总结和心得体会 3](#_Toc195481713)

[参考文献 4](#_Toc195481714)

[致 谢 7](#_Toc195481715)

# 绪论

* 1. 研究背景

随着国民生活水平不断提高，人民对美好生活的向往愈加浓烈。娱乐方式呈现出多样化趋势，其中走进电影院观影成了这其中必不可少的一种娱乐方式[1]。但是如何创作出吸引人的好电影也就成为了令不少电影导演头疼的问题，同时在众多院线大片中如何选择一部好电影进行观看也是观影人员需要思考的问题。随着互联网和社交媒体的快速发展，电影影评已经成为观众表达观影体验的重要渠道[2, 3]。海量的评论文本数据蕴含着丰富的情感信息，对电影制作方、发行方和观众都具有重要的价值。然后人工分析这些评论不仅耗时费力，而且难以实现大规模处理。因此，利用自然语言处理技术自动分析电影评论的情感倾向成为了研究热点[4, 5]。

目前，许多评论分析系统利用自然语言处理技术进行评论情感分析，以实现自动化的检测大众的情绪及意见。其中，利用预训练语言模型进行文本情感分析是比较先进的技术，可以高效地实现自动化的评论理解和分类。其中BERT（Bidirectional Encoder Representations from Transformers）是一种基于Transformer架构的先进语言表示模型，能够通过上下文感知的深度双向编码，有效捕捉文本中的语义和情感特征，完全能够支持中文电影评论情感分析系统的实现。

本项目旨在设计和实现一个基于BERT的中文电影评论情感分析系统，利用自然语言处理技术对中文电影评论进行分析，以实现自动化的检测观众情绪和反馈。该系统将通过对电影影评数据抓取和电影影评情感分析两个模块实现对电影影评的自动化分析。通过本项目研究，可以提高行业人员分析效率和观众人群选择适合自己电影的概率，为电影从业人员提供更高效、更精准的数据分析服务，为观众人群提供更全面的电影分析结果，帮助他们选择适合自己的电影，有助于提升电影从业人员制作更符合大众口味电影的能力和提升观众的观影体验，从而提升人民生活满足感和幸福感。

* 1. 国内外研究现状

在过去的几十年里，机器学习和自然语言处理领域的迅速发展促进了各种文本分析系统的出现。这些系统采用了一系列先进的技术，例如CNN、LSTM和BERT等，以实现对大量文本进行自动化处理和分析。

国外早期情感分析主要基于监督学习和情感词典方法：Pang et al. (2002)首次将机器学习应用于电影评论情感分类，比较了SVM、朴素贝叶斯和最大熵模型在二元情感分类（正面/负面）上的表现。研究发现，使用unigram特征时SVM表现最佳（准确率约82.9%），为后续文本分类任务奠定了基础[6]。情感词典：Taboada et al. (2011)开发了基于SentiWordNet的词典方法，通过计算文本中情感词的极性得分（正/负/中性）来预测整体情感。该方法在商品评论分析中达到约74%的准确率，但面临否定词处理（如"not good"）和上下文依赖的挑战[7]。

随着深度学习兴起，神经网络显著提升了情感分析性能：Kim (2014)提出使用静态词向量（Word2Vec）和卷积神经网络（CNN）提取文本局部特征。在MR电影评论数据集上，单通道CNN达到81.5%的准确率，多通道CNN提升至82.7%[8]。Tang et al. (2015)设计了基于长短期记忆网络（LSTM）的模型，通过建模词序依赖关系提升长文本分析效果。在Stanford Sentiment Treebank数据集上，LSTM比CNN的准确率提高约3%[9]。Vaswani et al. (2017)提出的Transformer架构通过自注意力机制（Self-Attention）动态计算词间权重，在机器翻译任务中取得突破，为后续BERT等模型奠定基础[10]。

最后在大模型时代，情感分析更是实现了质的飞跃：Devlin et al. (2018)提出的双向Transformer模型通过掩码语言建模（MLM）学习上下文相关表示。在SST-2情感数据集上，BERT-base达到92.7%的准确率，较LSTM提升约10%[11]。Zadeh et al. (2018)构建的CMU-MOSEI数据集包含23,500条YouTube视频片段，涵盖文本、音频和视觉模态。实验表明，多模态融合模型（如TFN）比单模态文本分析的F1值提高15%[12]。

总结来说，各种自然语言处理技术为情感分析系统的发展提供了基础。这些技术的综合应用使得自动化评估电影影评成为可能。随着机器学习和自然语言处理领域的进一步发展，情感分析系统的性能和准确性将得到持续提高。与此同时，中文语料和多模态分析也将在未来的研究中得到更多的关注。

* 1. 论文主要工作

制定研究计划，把自己拟做的毕业设计工作做个简要介绍。

简要说明毕业设计的主要工作内容和工作成绩，以及创新点和特色，反映毕业设计的工作量和完成质量。

项目主要工作是实现了对电影影评数据进行情感分析，构建了基于BERT的中文电影评论情感分析系统。

根据不同的功能需求可以将系统划分为五个功能模块。

1）xx模块：

2）xx模块：

* 1. 论文组织与结构

本论文主体部分由六部分组成，各部分作用分别为：

第一部分：绪论。本章介绍本项目的研究背景，陈述本项目相关的国内外研究现状，最后对本文的主要工作作出介绍。

第二部分：项目涉及的相关技术知识背景。主要介绍了本系统所使用到的算法，对BERT模型进行微调以完成情感分析任务，以及系统开发的其它相关背景知识。

第三部分：系统概要设计。主要介绍了概要情况。

第四部分：详细设计与实现。主要介绍了系统的详细设计流程和实现方法。

第五部分：实验方法。主要介绍了布署微调后的BERT模型以实现对中文电影评论进行情感分析；以及开发Web应用以实现模型的实际应用。

第六部分：讨论和体会。本章写出了项目整个开发过程中的感悟和反思。总结全过程，介绍项目开发过程体会、项目评价、项目成功用于实际意义的阐述以及未来工作。主要总结了本篇论文的工作和主要贡献并对现存问题和未来发展发表了自己的看法。

为了方便读者更进一步了解本文设计的系统项目，也便于有兴趣的读者进行二次开发，本文附录部分给出了项目的需求规格说明书、设计文档、软件测试文档、用户文档说明书以及项目的所有源码。

# 背景知识简介

* 1. 知识和技术要点一

这部分内容放算法基础或背景知识的简要介绍，篇幅不宜过多，不能大量抄袭。应对相关知识加以综合后进行写作。这里的知识都应是和毕业设计内容密切相关，而且后面论文内容会引用或以之为基础的。

本章写作没有一定之规，也不一定要放置背景知识。比如，如果是大家都很熟悉的内容，如某数据库系统设计，你又采取常规方法进行设计开发，完全可以省略这一章。

一般地，不要对开发工具进行详细介绍，如C++、JAVA等，除非你采用了目前最新技术，尤其是现成课本上都没有的知识，才有必要专门写一章告诉读者这些新信息。

* 1. 知识和技术要点二
     1. 表和论文主要格式

# 系统分析与设计

* 1. 代码框
  2. 图片插入和自动题注编号



图 3‑1 测试图



图 3‑2 测试图2

* 1. 表格插入和自动题注编号

表 3‑1 测试表

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 标题1 | 标题2 | | | | | | | | |
| 小项1 | 小项2 | 小项3 | 小项4 | 小项5 | 小项6 | 小项7 | 小项8 | 小项9 |
| A | 16 | 16 | -16 | -8 | 0 | 8 | 8 | 0 | 16 |
| B | -16 | 40 | 16 | 48 | 16 | 40 | 24 | -16 | -16 |
| C | 0 | -8 | 0 | -8 | -8 | 8 | 8 | 8 | 0 |
| D | 16 | 8 | -8 | -8 | 8 | 0 | 8 | 8 | 8 |
| E | 0 | -24 | 40 | -16 | 0 | 0 | 16 | 8 | -16 |
| F | -8 | 0 | 0 | -8 | 8 | 24 | 8 | 24 | 0 |
| 平均 | 1.3 | 5.3 | 5.3 | 0.0 | 4.0 | 13.3 | 12.0 | 5.3 | -1.3 |

表 3‑2 测试表2

|  |  |  |  |
| --- | --- | --- | --- |
| CPU | RAM | GPU | OS |
| Intel ® Core(TM) i7-8750H 2.20GHz | 32.0 GB | NVIDIA RTX 4060 6G | Windows 10 64位操作系统 |

表 3‑3 测试表3

|  |  |  |
| --- | --- | --- |
| **工具名称** | **版本** | **用途** |
| WebStorm | 2024.3 | 前端开发和设计主要IDE |
| MongoDB | 8.0.6 | 数据库 |
| PyCharm | 2024.1 | 后端开发以及网页爬虫主要IDE |
| Python | 3.11.9 | 后端开发主要使用的编程语言 |
| MongoDB Compass | 1.46 | 数据库管理工具 |
| Node.js | 22.12.0 | 前端开发所需环境 |
| TypeScript | 5.8.3 | 前端开发主要使用的编程语言 |
| Docker | 28.0.1 | 容器化部署 |
| Git | 2.41 | 版本控制 |

# 系统详细设计与实现

* 1. 公式和自动编号

例如公式（4-1）

|  |  |  |
| --- | --- | --- |
|  |  | (4-1) |

再添加公式（4-2）

|  |  |  |
| --- | --- | --- |
|  |  | (4-1) |

# 工作总结和心得体会

参考文献

1. 李佳, 实现人民对美好生活向往的路径. 党课参考, 2023(06): 第65-70页.
2. 王宜文与张云, 互联网时代的电影评论. 艺术评论, 2015(08): 第14-22页.
3. 司若, 黄莺与许婉钰, 近年来我国网络影评发展脉络研究. 当代电影, 2020(03): 第137-143页.
4. 张璐, 从Python情感分析看海外读者对中国译介文学的接受和评价：以《三体》英译本为例. 外语研究, 2019. 36(04): 第80-86页.
5. 刘玲玉与邓燕燕, 基于Python情感分析和批评隐喻的网络话语分析——以影片《流浪地球》中美德影评为例. 江苏大学学报(社会科学版), 2022. 24(03): 第76-88页.
6. Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. EMNLP.
7. Taboada, M., Brooke, J., Tofiloski, M., et al. (2011). Lexicon-based methods for sentiment analysis. Computational Linguistics.
8. Kim, Y. (2014). Convolutional neural networks for sentence classification. EMNLP.
9. Tang, D., Qin, B., & Liu, T. (2015). Document modeling with gated recurrent neural network for sentiment classification. EMNLP.
10. Vaswani, A., et al. (2017). Attention is all you need. NeurIPS.
11. Devlin, J., et al. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL.
12. Zadeh, A., et al. (2018). Multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion. ACL.
13. Radford, A., et al. (2018). Improving Language Understanding by Generative Pre-Training. OpenAI.
14. Brown, T., et al. (2020). Language Models are Few-Shot Learners. NeurIPS.
15. Liu, Y., et al. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv.
16. Jurafsky, D., & Martin, J. H. (2021). Speech and Language Processing (3rd ed.). Pearson.
17. Rajpurkar P, et al. SQuAD: 100,000+ questions for machine comprehension of text. EMNLP 2016.
18. Sun C, et al. How to fine-tune BERT for text classification? CCL 2019.
19. Yamada I, et al. LUKE: Deep contextualized entity representations with entity-aware self-attention. EMNLP 2020.
20. Strubell E, et al. Energy and policy considerations for deep learning in NLP. ACL 2019.
21. Gururangan S, et al. Don't stop pretraining: Adapt language models to domains and tasks. ACL 2020.
22. Sanh V, et al. DistilBERT, a distilled version of BERT. NeurIPS 2019.
23. Houlsby N, et al. Parameter-efficient transfer learning for NLP. ICML 2019.
24. Raffel C, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. JMLR 2020.
25. Chodorow K. MongoDB: The Definitive Guide. O'Reilly, 2013.
26. Banker K. MongoDB in Action. Manning, 2011.
27. Sharma R. Mastering MongoDB Aggregation Framework. Packt, 2021.
28. Bradshaw S. MongoDB Indexing Strategies. Apress, 2015.
29. MongoDB Inc. Security Reference Architecture. Technical Report, 2022.
30. Sadalage P. NoSQL Distilled. Addison-Wesley, 2012.

**声 明**

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得四川大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

本学位论文成果是本人在四川大学读书期间在导师指导下取得的，论文成果归四川大学所有，特此声明。

作者签名： 导师签名：

年 月 日

**学位论文使用授权书**

本学位论文作者完全了解四川大学有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或相关机构送交论文的原件、复印件和电子版，允许论文被查阅和借阅。本人授权四川大学将本学位论文的全部或部分内容编入有关数据库进行信息技术服务，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文，并用于学术活动。

（涉密学位论文在解密后适用于本授权书）

作者签名： 导师签名：

年 月 日

致 谢

首先我在这里向养育和培养我的父母表示深深地思念和感谢，再向关心我，帮助我的老师和同学表示感谢！

能顺利圆满地完成我的设计，离不开自身的努力，离不开指导老师于忠德老师的指点，更离不开四川大学这个大环境四年来在学习和生活的熏陶和培养。

大四下学期开始不久，我开始了我的毕业设计。在这几个月里，我得到了导师方老师的关心和教导，方老师在学术上的渊博知识让我在这几个月的学习和研究中深受其益。我向方老师表示深深地谢意；再向和我进行讨论问题，让彼此相互帮助和学习工作的同学表示感谢；向软件学院实验室中，陪着我们度过几多光阴的管理员和老师表示谢意。

设计的完成，为我在川大的学习生涯画上了一个圆满的句号。在四川大学度过的美好时光将使我终生受益。