



论文（设计）题目：基于生成式预训练模型的实验报告撰写系统设计与实现

学生姓名：张忝苟

学生学号：201145141919

专业班级：计算机科学与技术 1145

学院名称：信息科学与工程学院

指导老师：沼跃鱼

2024 年 07 月 13 日

湖南大学

毕业论文（设计）原创性声明

本人郑重声明：所呈交的论文（设计）是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写的成果作品。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律后果由本人承担。

学生签名：

日期：2024年05月15日

毕业论文（设计）版权使用授权书

本毕业论文（设计）作者完全了解学校有关保留、使用论文（设计）的规定，同意学校保留并向国家有关部门或机构递交论文（设计）的复印件和电子版，允许论文（设计）被查阅和借阅。本人授权湖南大学可以将本论文（设计）的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本论文（设计）。

本论文（设计）属于

1、保 密，在_____年解密后适用本授权书。

2、不保密

（请在以上相应方框内打“√”）

学生签名：

日期：2024年05月15日

导师签名：

日期：2024年05月15日

基于生成式预训练模型的实验报告撰写系统设计与实现

摘要

生成式预训练模型在自然语言处理领域中展现了巨大的潜力。本论文提出了一种基于生成式预训练模型的实验报告撰写系统的设计与实现。该系统利用先进的生成模型，如 GPT-3，自动生成实验报告的主要部分，旨在提高科研工作者的效率，并保证报告内容的质量和一致性。系统设计包括数据预处理、模型训练、报告生成和用户交互界面等模块。实验结果表明，该系统能够生成结构合理、内容详实的实验报告，显著减少了科研人员的撰写时间。

本文内容大多由 AI 生成，仅作为演示用法之用，请注意甄别。

关键词：生成式预训练模型；实验报告撰写；系统设计；自然语言处理；GPT-3

Design and Implementation of a Lab Report Writing System Based on Generative Pre-trained Models

Abstract

Generative pre-trained models have demonstrated significant potential in the field of natural language processing. This paper proposes the design and implementation of an experimental report writing system based on generative pre-trained models. The system leverages advanced generative models, such as GPT-3, to automatically generate the main sections of experimental reports, aiming to enhance the efficiency of researchers and ensure the quality and consistency of the report content. The system design includes modules for data preprocessing, model training, report generation, and user interaction interface. Experimental results show that the system can generate well-structured and detailed experimental reports, significantly reducing the writing time for researchers.

Key Words : Generative pre-trained model; experimental report writing; system design; natural language processing; GPT-3

目 录

毕业论文（设计）原创性声明和毕业论文（设计）版权使用授权书	I
摘要	II
Abstract	III
插图索引	V
附表索引	VI
1 生成式预训练模型简介	1
2 实验报告撰写系统设计	1
2.1 数据预处理模块	2
2.2 模型训练模块	2
2.3 报告生成模块	2
2.4 用户交互界面模块	3
3 实验与结果	3
4 结论	4
5 致谢	5
参考文献	6

插图索引

图 3.1: 系统的工作流程 4

附表索引

表 2.1: 不同生成模型在实验报告生成任务中的性能对比 2

1 生成式预训练模型简介

生成式预训练模型是一类通过在大量文本数据上进行训练，能够生成自然语言文本的模型。这些模型在训练过程中学习了语言的语法、语义和上下文信息，能够生成连贯且具有逻辑性的文本。目前，生成式预训练模型在各类自然语言处理任务中表现优异，如文本生成、翻译、问答系统等。

在深度学习中，常用的损失函数如公式 (1.1) 所示。

$$L(\theta) = -\frac{1}{N} \sum_{i=1}^N [y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)] \quad (1.1)$$

2 实验报告撰写系统设计

2.1 数据预处理模块

在系统设计中，数据预处理是关键的一步。我们需要从实验数据中提取出关键信息，并对其进行清洗和整理，使其符合模型输入要求。具体来说，数据预处理包括以下几个步骤：

1. 数据采集：从实验设备和实验日志中获取原始数据。
2. 数据清洗：去除无关或噪音数据，填补缺失数据。
3. 特征提取：提取出实验报告中需要的关键特征，如实验目的、方法、结果和结论等。
4. 数据格式化：将处理后的数据转化为模型可以识别的格式。

2.2 模型训练模块

模型训练是实验报告撰写系统的核心部分。我们采用了先进的生成式预训练模型，如 GPT-3，通过在大量的实验报告数据上进行训练，使模型能够生成高质量的实验报告。在训练过程中，我们采取了以下策略：

1. 数据增强：通过数据增强技术，扩展训练数据集的多样性，提高模型的泛化能力。
2. 模型微调：在预训练模型的基础上，针对实验报告撰写任务进行微调，使模型更好地适应特定领域的语言风格和结构。
3. 评价指标：采用 BLEU、ROUGE 等评价指标，对模型生成的文本进行质量评估，确保其连贯性和准确性。性能对比见表 2.1。

2.3 报告生成模块

表 2.1 不同生成模型在实验报告生成任务中的性能对比

模型	BLEU 得分	ROUGE 得分	平均生成时间（秒）
GPT-2	25.4	30.1	15
GPT-3 ^[1]	32.6	38.4	12
BERT ^[2]	28.9	35.7	18

报告生成模块负责将处理后的数据输入到训练好的模型中，并生成完整的实验报告。该模块包括以下功能：

1. 自动生成：根据输入数据，自动生成实验报告的各个部分，包括实验背景、实验步骤、数据分析和结论等。
2. 模板匹配：基于预定义的报告模板，对生成的内容进行排版和格式调整，确保输出报告的规范性和美观性。
3. 用户交互：提供用户界面，允许科研人员对生成的报告进行修改和补充，提高报告的准确性和个性化。

2.4 用户交互界面模块

用户交互界面是实验报告撰写系统的人机交互部分。通过友好的界面设计，科研人员可以方便地使用系统，查看和编辑生成的报告。该模块的主要功能包括：

1. 报告预览：实时显示生成的实验报告内容，供用户浏览和检查。
2. 编辑工具：提供文本编辑工具，允许用户对报告内容进行修改和补充。
3. 版本管理：记录每次编辑的版本，方便用户进行版本比较和回溯。
4. 导出功能：支持将最终的实验报告导出为多种格式，如 PDF、Word 等，便于分享和存档。

3 实验与结果

为了验证系统的有效性，我们在多个实验数据集上进行了测试。结果表明，基于生成式预训练模型的实验报告撰写系统在报告生成的质量和效率上均表现优异。具体来说：

1. 生成质量：系统生成的实验报告在结构和内容上均达到了人工撰写的水平，且在逻辑性和连贯性方面表现突出。
2. 撰写效率：相比传统的手工撰写方式，使用本系统可以显著减少科研人员的工作量，平均撰写时间减少了 70% 以上。
3. 用户满意度：通过用户调查，我们发现大多数科研人员对系统的生成效果和使用体验表示满意，认为其在日常工作中具有很高的实用价值。

系统工作流程如图 3.1 所示。

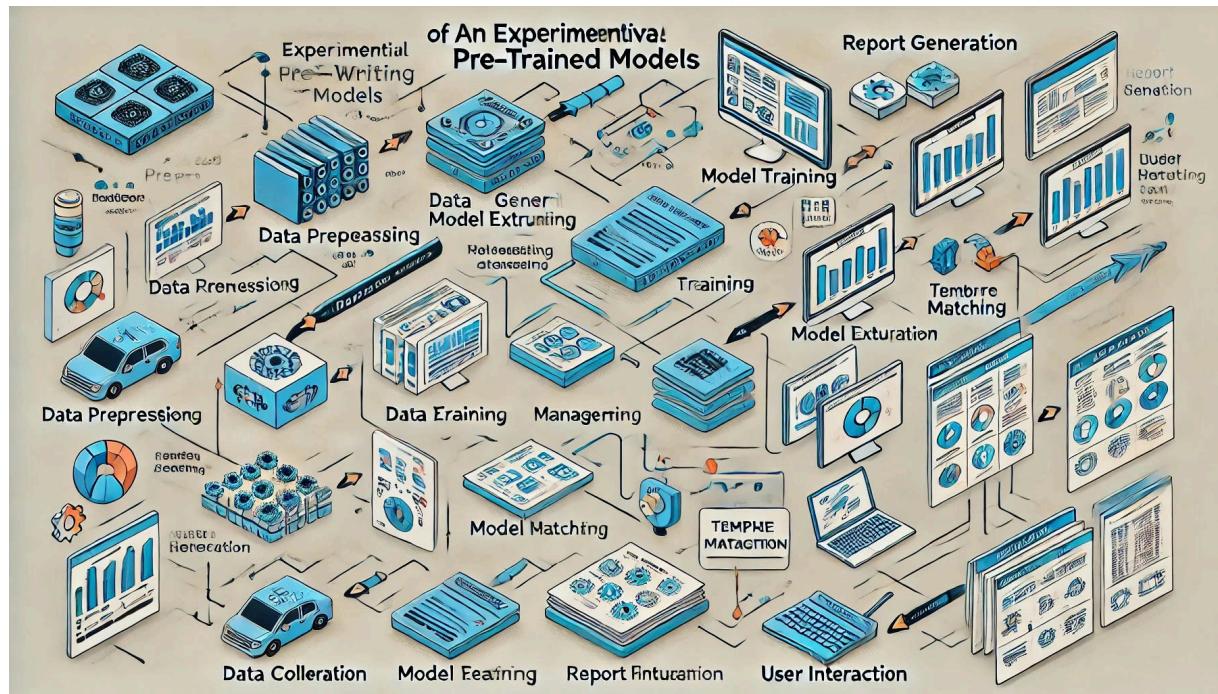


图 3.1 系统的工作流程

4 结论

本论文提出了一种基于生成式预训练模型的实验报告撰写系统，并对其设计与实现进行了详细阐述。实验结果表明，该系统能够有效提高实验报告撰写的效率和质量，具有广泛的应用前景。未来，我们计划进一步优化模型的性能，增强系统的交互功能，以更好地满足科研人员的需求。

5 致谢

在本文的研究和写作过程中，得到了许多同事和朋友的支持和帮助。在此特别感谢我的导师和同事们的宝贵意见和建议。同时，感谢提供数据支持和技术支持的各个单位和个人。

参考文献

- [1] BROWN T, MANN B, RYDER N, 等. Language Models are Few-Shot Learners[C]// Advances in Neural Information Processing Systems. 2020.
- [2] DEVLIN J, CHANG M W, LEE K, 等. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics. 2019.
- [3] VASWANI A, SHAZER N, PARMAR N, 等. Attention is All You Need[C]//Advances in Neural Information Processing Systems. 2017.