

UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA

BACHELOR THESIS



论文题目 时域积分方程时间步进算法及其快速算法

学科专业 无线电物理

学 号 201421040223

作者姓名 王稳

指导老师 赖生建 副教授

学 院 物理电子学院

摘 要

为了适应日益增长的宽带信号和非线性系统的工程应用，用于分析瞬态电磁散射问题的时域积分方程方法研究日趋活跃。本文以时域积分方程时间步进算法及其快速算法为研究课题，重点研究了时间步进算法的数值实现技术、后时稳定性问题以及两层平面波算法加速计算等，主要研究内容分为四部分。

.....

关键词：时域电磁散射，时域积分方程，时间步进算法，后时不稳定性，时域平面波算法

ABSTRACT

With the widespread engineering applications ranging from broadband signals and non-linear systems, time-domain integral equations (TDIE) methods for analyzing transient electromagnetic scattering problems are becoming widely used nowadays. TDIE-based marching-on-in-time (MOT) scheme and its fast algorithm are researched in this dissertation, including the numerical techniques of MOT scheme, late-time stability of MOT scheme, and two-level PWTD-enhanced MOT scheme. The contents are divided into four parts shown as follows.

Keywords: Time-domain Electromagnetic Scattering, Time-domain Integral Equation, Marching-on In-time (MOT) Scheme, Late-time Instability, Plane Wave Time-domain (PWTD) Algorithm

目 录

第一章 引 言.....	1
1.1 背景介绍.....	2
1.1.1 大语言模型（LLMs）的兴起.....	2
1.1.2 LLMs 的局限性	2
1.2 检索增强生成（RAG）技术的诞生	2
1.3 RALMs 的研究意义和应用前景	2
第二章 RAG 技术详解.....	3
2.1 RAG 的核心范式	3
2.1.1 检索增强生成（RAG）	3
2.1.2 检索增强理解（RAU）	3
2.2 RAG 的组成要素	3
2.2.1 检索器	3
2.2.2 语言模型.....	3
2.2.3 增强手段.....	3
2.3 RAG 的关键技术	3
2.3.1 检索器类型	3
2.3.2 语言模型的选择	3
2.3.3 性能提升策略.....	3
第三章 研究进展.....	4
3.1 检索技术的改进.....	4
3.2 语言模型的优化.....	4
3.3 性能提升策略的创新	4
第四章 未来研究方向	5
4.1 多语言和多模态能力的提升.....	5
4.2 外部知识的质量控制	5
4.3 计算效率的优化.....	5
第五章 结论.....	6
5.1 RALMs 的重要性的影响.....	6
5.2 RALMs 的未来展望	6
致 谢	7

附录 A 中心极限定理的证明	8
A.1 高斯分布和伯努利实验	8
参考文献	9
攻读本科学位期间取得的成果	12
外文资料原文	13
外文资料译文	14

第一章 引言

综述参考文献列举如下，仅供参考

这篇文章 [2] 是一篇关于检索增强型大型语言模型（Retrieval-Augmented Large Language Models, 简称 RA-LLMs）的综合综述，标题为“A Survey on RAG Meeting LLMs: Towards Retrieval-Augmented Large Language Models”。文章主要研究了以下几个方面：

1. **** 检索增强生成（RAG）技术 ****：介绍了 RAG 技术如何为 AI 生成的内容（AIGC）提供可靠和最新的外部知识，以及它如何辅助现有的生成型 AI 产生高质量的输出。

2. **** 大型语言模型（LLMs）****：讨论了 LLMs 在语言理解和生成方面展示的革命性能力，以及它们面临的固有局限性，例如幻觉（hallucinations）和过时的内部知识。

3. ****RA-LLMs 的架构 ****：综述了现有的 RA-LLMs 研究，覆盖了三个主要技术视角：架构、训练策略和应用。

4. ****RA-LLMs 的训练 ****：探讨了现有的 RAG 方法，包括无需训练的方法和基于训练的方法，如独立训练、顺序训练和联合训练。

5. ****RA-LLMs 的应用 ****：介绍了 RA-LLMs 在自然语言处理（NLP）应用、下游任务和特定领域应用中的使用情况。

6. **** 未来研究方向 ****：讨论了 RA-LLMs 当前研究的局限性和未来可能的研究方向，包括提高模型的可靠性、多语言能力和多模态能力。

根据您正在撰写的综述的结构，这篇文章可以归类到以下几个部分：

****II. RAG 技术详解 **** - 文章详细介绍了 RAG 的核心范式、组成要素以及关键技术，可以作为这一部分的理论基础和技术细节。

****III. 研究进展 **** - 文章中关于 RA-LLMs 在不同任务中的应用和性能提升策略的创新可以归类为研究进展。

****IV. 未来研究方向 **** - 文章最后一部分讨论了未来挑战和机遇，可以为您的综述的未来研究方向提供参考。

****V. 结论 **** - 文章的结论部分可以强化您综述中关于 RALMs 的重要性和影响的讨论，并提供对未来展望的参考。

这篇文章提供了对 RA-LLMs 的全面概述，包括其架构、训练方法、应用领域以及未来发展方向的深入分析，可以为您的综述提供丰富的信息和视角。

1.1 背景介绍

这是一个测试文本，用于模拟真实内容的效果。它可以帮助设计师和开发者在没有实际内容的情况下进行排版和布局测试。这个文本没有实际意义，只是为了展示字体、段落和页面布局的效果。Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

这是一个引用的例子 [1–41]

1.1.1 大语言模型（LLMs）的兴起

这是一个测试文本，用于模拟真实内容的效果。它可以帮助设计师和开发者在没有实际内容的情况下进行排版和布局测试。这个文本没有实际意义，只是为了展示字体、段落和页面布局的效果。

1.1.2 LLMs 的局限性

这是一个测试文本，用于模拟真实内容的效果。它可以帮助设计师和开发者在没有实际内容的情况下进行排版和布局测试。这个文本没有实际意义，只是为了展示字体、段落和页面布局的效果。

1.2 检索增强生成（RAG）技术的诞生

这是一个测试文本，用于模拟真实内容的效果。它可以帮助设计师和开发者在没有实际内容的情况下进行排版和布局测试。这个文本没有实际意义，只是为了展示字体、段落和页面布局的效果。

1.3 RALMs 的研究意义和应用前景

这是一个测试文本，用于模拟真实内容的效果。它可以帮助设计师和开发者在没有实际内容的情况下进行排版和布局测试。这个文本没有实际意义，只是为了展示字体、段落和页面布局的效果。

第二章 RAG 技术详解

这篇文章可以作为“检索增强生成（RAG）”技术的安全性和隐私性讨论的案例研究。特别是，它提供了关于 LLMs 在安全性方面的应用和潜在风险的深入分析，这与 RAG 技术中的“检索增强理解（RAU）”有关联。[1]

2.1 RAG 的核心范式

2.1.1 检索增强生成（RAG）

2.1.2 检索增强理解（RAU）

2.2 RAG 的组成要素

2.2.1 检索器

2.2.2 语言模型

2.2.3 增强手段

2.3 RAG 的关键技术

2.3.1 检索器类型

2.3.1.1 稀疏检索

2.3.1.2 密集检索

2.3.1.3 互联网检索

2.3.1.4 混合检索

2.3.2 语言模型的选择

2.3.2.1 自编码器语言模型

2.3.2.2 自回归语言模型

2.3.2.3 编码器-解码器模型

2.3.3 性能提升策略

2.3.3.1 检索精准性控制

2.3.3.2 知识蒸馏

2.3.3.3 互联网检索的即插即用

第三章 研究进展

文章中关于 LLMs 在安全测试、恶意代码检测和数据隐私保护方面的应用可以归类为“检索技术的改进”和“语言模型的优化”。[1]

这篇文章是一篇介绍如何通过反射性标签（Reflective Tags）来改进检索增强型生成（Retrieval-Augmented Generation，简称 RAG）在大型语言模型（LLMs）中的应用的研究文章。[3]

文章中 [4] 关于 RAG 技术在教育和商业 IT 文档中的应用可以归类为研究进展。这篇文章 [5] 详细介绍了 RAG 技术在特定应用（法律问题回答）中的进展，特别是通过结合 CBR 来提升 LLM 性能的方法。它展示了如何通过结构化检索和案例重用来增强模型的输出质量。

文章中 [7] 提出的端到端系统设计和实验方法可以作为方法论部分的案例研究，展示了如何结合 RAG 和 LLMs 来处理复杂的查询任务。

这篇文章 [6] 详细介绍了 LLM-RAG 在医疗领域的应用进展，特别是在术前医学的定制化知识方面的研究。

文章中 [8] 提出的三种策略（变异、注入和扩展）以及评估方法可以作为方法论部分的案例研究，展示了如何结合 LLMs 和 RAG 来解决实际问题。

3.1 检索技术的改进

这是一个测试文本，用于模拟真实内容的效果。它可以帮助设计师和开发者在没有实际内容的情况下进行排版和布局测试。这个文本没有实际意义，只是为了展示字体、段落和页面布局的效果。

3.2 语言模型的优化

这是一个测试文本，用于模拟真实内容的效果。它可以帮助设计师和开发者在没有实际内容的情况下进行排版和布局测试。这个文本没有实际意义，只是为了展示字体、段落和页面布局的效果。

3.3 性能提升策略的创新

这是一个测试文本，用于模拟真实内容的效果。它可以帮助设计师和开发者在没有实际内容的情况下进行排版和布局测试。这个文本没有实际意义，只是为了展示字体、段落和页面布局的效果。

第四章 未来研究方向

文章提出的未来研究方向，如模型和参数提取攻击的研究、安全指令调整等，可以归类为“未来研究方向”中的“计算效率的优化”和“外部知识的质量控制”。
[1]

文章 [5] 提出的 CBR-RAG 模型和评估结果为未来在类似领域的研究提供了新的方向，特别是在探索如何进一步提升 LLM 的输出质量和可靠性方面。

文章 [6] 研究结果和结论为未来在医疗领域的 LLM-RAG 研究提供了新的方向，尤其是在提高模型性能和应用可扩展性方面。

4.1 多语言和多模态能力的提升

这是一个测试文本，用于模拟真实内容的效果。它可以帮助设计师和开发者在没有实际内容的情况下进行排版和布局测试。这个文本没有实际意义，只是为了展示字体、段落和页面布局的效果。

4.2 外部知识的质量控制

这是一个测试文本，用于模拟真实内容的效果。它可以帮助设计师和开发者在没有实际内容的情况下进行排版和布局测试。这个文本没有实际意义，只是为了展示字体、段落和页面布局的效果。

4.3 计算效率的优化

这是一个测试文本，用于模拟真实内容的效果。它可以帮助设计师和开发者在没有实际内容的情况下进行排版和布局测试。这个文本没有实际意义，只是为了展示字体、段落和页面布局的效果。

第五章 结论

文章的结论部分可以强化您综述中关于“RALMs 的重要性的影响”以及“RALMs 的未来展望”的讨论，因为它提供了对 LLMs 在安全和隐私方面潜力的全面评估。[1]

5.1 RALMs 的重要性的影响

这是一个测试文本，用于模拟真实内容的效果。它可以帮助设计师和开发者在没有实际内容的情况下进行排版和布局测试。这个文本没有实际意义，只是为了展示字体、段落和页面布局的效果。

5.2 RALMs 的未来展望

这是一个测试文本，用于模拟真实内容的效果。它可以帮助设计师和开发者在没有实际内容的情况下进行排版和布局测试。这个文本没有实际意义，只是为了展示字体、段落和页面布局的效果。

致 谢

在攻读博士学位期间，首先衷心感谢我的导师 XXX 教授

附录 A 中心极限定理的证明

A.1 高斯分布和伯努利实验

参考文献

- [1] Yao Y, Duan J, Xu K, et al. A survey on large language model (llm) security and privacy: The good, the bad, and the ugly, 2024.
- [2] Fan W, Ding Y, Ning L, et al. A survey on rag meeting llms: Towards retrieval-augmented large language models[C]. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2024: 6491-6501.
- [3] Yang C, Fujita S. Adaptive control of retrieval-augmented generation for llms through reflective tags[J]. , 2024.
- [4] Posedaru B-S, Pantelimon F-V, Dulgheru M-N, et al. Artificial intelligence text processing using retrieval-augmented generation: Applications in business and education fields[J]. Proceedings of the International Conference on Business Excellence, 2024, 18: 209-222.
- [5] Wiratunga N, Abeyratne R, Jayawardena L, et al. Cbr-rag: Case-based reasoning for retrieval augmented generation in llms for legal question answering[J]. , 2024.
- [6] Ke Y, Jin L, Elangovan K, et al. Development and testing of retrieval augmented generation in large language models – a case study report[J]. CoRR abs/2402.01733, 2024.
- [7] Li J, Yuan Y, Zhang Z. Enhancing llm factual accuracy with rag to counter hallucinations: A case study on domain-specific queries in private knowledge-bases[J]. CoRR abs/2403.10446, 2024.
- [8] Daneshvar S S, Nong Y, Yang X, et al. Exploring rag-based vulnerability augmentation with llms[J]. , 2024.
- [9] He X, Tian Y, Sun Y, et al. G-retriever: Retrieval-augmented generation for textual graph understanding and question answering[J]. , 2024.
- [10] Siriwardhana S, Weerasekera R, Wen E, et al. Improving the domain adaptation of retrieval augmented generation (rag) models for open domain question answering[J]. CoRR abs/2210.02627, 2022.
- [11] Miao J, Thongprayoon C, Suppadungsuk S, et al. Integrating retrieval-augmented generation with large language models in nephrology: Advancing practical applications, 2024.
- [12] Zhang R, Du H, Liu Y, et al. Interactive ai with retrieval-augmented generation for next generation networking[J]. , 2024.
- [13] Li S, Shang H, Wei D, et al. La-rag:enhancing llm-based asr accuracy with retrieval-augmented generation[J]. CoRR abs/2409.08597, 2024.

- [14] Wu M, Cao S. Llm-augmented retrieval: Enhancing retrieval models through language models and doc-level embedding[J]. , 2024.
- [15] Wang K, Duan F, Li P, et al. Llms know what they need: Leveraging a missing information guided framework to empower retrieval-augmented generation[J]. CoRR abs/2404.14043, 2024.
- [16] Kuppa A, Nicholls J, Le-Khac N A. Manipulating prompts and retrieval-augmented generation for llm service providers[C]. Proceedings of the International Conference on Security and Cryptography, 2024: 777-785.
- [17] Giuffrè M, Kresevic S, Pugliese N, et al. Optimizing large language models in digestive disease: strategies and challenges to improve clinical outcomes[J]. Liver International, 2024, 44(9): 2114-2124.
- [18] Hu Z, Wang C, Shu Y, et al. Prompt perturbation in retrieval-augmented generation based large language models[J]. CoRR abs/2402.07179, 2024.
- [19] Hu Y, Lu Y. Rag and rau: A survey on retrieval-augmented language model in natural language processing[J]. CoRR abs/2404.19543, 2024.
- [20] Stefano G D, Schönherr L, Pellegrino G. Rag and roll: An end-to-end evaluation of indirect prompt manipulations in llm-based application frameworks[J]. CoRR abs/2408.05025, 2024.
- [21] Phan H, Acharya A, Chaturvedi S, et al. Rag vs. long context: Examining frontier large language models for environmental review document comprehension[J]. , 2024.
- [22] Zhao S, Yang Y, Wang Z, et al. Retrieval augmented generation (rag) and beyond: A comprehensive survey on how to make your llms use external data more wisely[J]. , 2024.
- [23] Li Z, Li C, Zhang M, et al. Retrieval augmented generation or long-context llms? a comprehensive study and hybrid approach[J]. CoRR abs/2407.16833, 2024.
- [24] Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks[J]. , 2020.
- [25] Xu Z, Cruz M J, Guevara M, et al. Retrieval-augmented generation with knowledge graphs for customer service question answering[C]. SIGIR 2024 - Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2024: 2905-2909.
- [26] Ahn Y, Lee S G, Shim J, et al. Retrieval-augmented response generation for knowledge-grounded conversation in the wild[J]. IEEE Access, 2022, 10: 131374-131385.
- [27] Zhang P, Xiao S, Liu Z, et al. Retrieve anything to augment large language models[J]. , 2023.

- [28] Glass M, Rossiello G, Chowdhury F M, et al. Robust retrieval augmented generation for zero-shot slot filling, 1949.
- [29] Chan C-M, Xu C, Yuan R, et al. Rq-rag: Learning to refine queries for retrieval augmented generation[J]. , 2024.
- [30] Laban P, Fabbri A R, Xiong C, et al. Summary of a haystack: A challenge to long-context llms and rag systems[J]. , 2024.
- [31] Fatehkia M, Lucas J K, Chawla S. T-rag: Lessons from the llm trenches[J]. , 2024.
- [32] Yilma G M, Ayala-Romero J A, Garcia-Saavedra A, et al. Telecomrag: Taming telecom standards with retrieval augmented generation and llms[J]. , 2024.
- [33] Zeng S, Zhang J, He P, et al. The good and the bad: Exploring privacy issues in retrieval-augmented generation (rag)[J]. , 2024.
- [34] Cuconasu F, Trappolini G, Siciliano F, et al. The power of noise: Redefining retrieval for rag systems[C]. SIGIR 2024 - Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2024: 719-729.
- [35] Cheng P, Ding Y, Ju T, et al. Trojanrag: Retrieval-augmented generation can be backdoor driver in large language models[J]. , 2024.
- [36] Nam D, MacVean A, Hellendoorn V, et al. Using an llm to help with code understanding[C]. Proceedings - International Conference on Software Engineering, 2024: 1184-1196.
- [37] Kreimeyer K, Canzoniero J V, Fatteh M, et al. Using retrieval-augmented generation to capture molecularly-driven treatment relationships for precision oncology[J]. Studies in health technology and informatics, 2024, 316: 983-987.
- [38] Du X, Zheng G, Wang K, et al. Vul-rag: Enhancing llm-based vulnerability detection via knowledge-level rag[J]. CoRR abs/2406.11147, 2024.
- [39] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]. ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 30 (NIPS 2017), 10010 NORTH TORREY PINES RD, LA JOLLA, CALIFORNIA 92037 USA, 2017: .
- [40] Guu K, Lee K, Tung Z, et al. Realm: Retrieval-augmented language model pre-training[J]. CoRR abs/2002.08909, 2020.
- [41] Devlin J, Chang M-W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[C]. 2019 CONFERENCE OF THE NORTH AMERICAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: HUMAN LANGUAGE TECHNOLOGIES (NAACL HLT 2019), VOL. 1, 2019: 4171-4186.

攻读本科学位期间取得的成果

- [1] Li J Y, **Zhang S**, Zhao Y W, et al. New memory method of impedance elements for marching-on-in-time solution of time-domain integral equation[J]. Electromagnetics, 2010, 30(5): 448-462.
- [2] 李四, **张三**. 时间步进算法中阻抗矩阵的高效存储新方法 [J]. 电波科学学报, 2010, 25(4): 624-631.
- [3] 李四, **张三**. 时域磁场积分方程时间步进算法稳定性研究 [J]. 物理学报, 2013, 62(9): 090206-1-090206-6.
- [4] **张三**, 李四. 时域磁场积分方程时间步进算法后时稳定性研究. 电子科技大学学报 [J] (已录用, 待刊).
- [5] **Zhang S**. Parameters discussion in two-level plane wave time-domain algorithm[C]. 2012 IEEE International Workshop on Electromagnetics, Chengdu, 2012: 38-39.
- [6] **张三**, 李四. 时域积分方程时间步进算法研究 [C]. 电子科技大学电子科学技术研究院第四届学术交流会, 成都, 2008: 164-168.
- [7] **张三** (4). 人工介质雷达罩技术研究. 国防科技进步二等奖, 2008 年.
- [8] XXX, XXX, XXX, XXX, 王升. XXX 的陶瓷研究. 四川省科技进步三等奖, 2003 年 12 月.

外文资料原文

1 The OFDM Model of Multiple Carrier Waves

外文资料译文

1 基于多载波索引键控的正交频分多路复用系统模型