



华南理工大学

South China University of Technology

专业学位硕士学位论文

面向垂直领域的检索增强

对话生成研究及应用

学位类别 电子信息硕士(软件工程)

所在学院 软件学院

论文提交日期 2024 年 3 月 15 日

摘 要

垂直领域对话生成技术是构建智能对话系统中重要的基础技术，该技术旨在为用户解决特定领域的问题、提供专业解答。目前该技术已经广泛应用在医疗保健、金融、法律、科技等领域的服务助手、答疑机器人等应用中，具有很高的研究价值。

目前主流的面向垂直领域的对话生成方法，都是基于检索增强的自回归模型设计的。然而，现有面向垂直领域的检索增强对话生成存在以下难点：1) 专有领域背景知识丰富，且逻辑相对复杂；2) 用户问题形式多样，意图难以理解。针对以上存在的两个问题，本文采取的研究方案如下：

1) 针对专有领域背景知识丰富、逻辑相对复杂的问题，提出了基于内外部知识对齐的检索增强对话生成方法。该方法利用一个语义切分模块提取知识文档的文档级信息和实体级信息，并将提取出来的知识分别用于构建外部知识库和内部知识注入，实现垂直领域对话模型内外部知识对齐。本文将该方法应用于金融分析领域，在真实股票价格预测任务和金融问答任务上验证了该方法的有效性。

2) 针对用户问题形式多样、意图难理解的问题，提出了基于人类偏好对齐的检索增强对话生成方法。该方法通过采集人类对真实场景对话样本的偏好，利用大型语言模型的理解与分析能力进行问题优化，并训练单独的问题优化语言模型，实现了与模型无关的、可解释、效果稳定的人类偏好对齐。本文分别在两个不同的基准测试机上与目前主流的语言模型对齐方法进行实验比较，证明了该方法能同时提升知识文档召回准确率和模型理解与用户意图的一致性。

关键词：对话系统；人类偏好对齐；检索增强生成；自回归语言模型

Abstract

Vertical domain question-answering dialogue generation tasks are helpful to improve user experience, solve domain-specific problems and provide personalized services. This paper studies the generation of question-answering dialogue for vertical fields, aiming to provide users with effective and accurate answers to vertical questions, and provide support for technical service assistants and answering robots in medical care, finance, law, education, science and technology and other fields. At present, the mainstream method is to use retrieval enhanced large language model to generate dialogue. However, the dialogue generation for vertical domains has the following difficulties: 1) the background knowledge of the proprietary domain is rich and the logic is relatively complex; 2) User questions have various forms and their intentions are difficult to understand. In view of the above two problems, the research scheme adopted in this paper is as follows:

1) Aiming at the problem that the background knowledge of the proprietary domain is rich and the logic is relatively complex, a vertical domain dialogue generation method based on internal and external knowledge alignment is proposed. This method uses a semantic segmentation module to extract the document-level information and entity-level information of the knowledge document, and uses the extracted knowledge to build an external knowledge base and an internal knowledge injection, respectively, to realize the alignment of internal and external knowledge of the vertical domain dialogue model. In this paper, the method is applied to the field of financial analysis, and its effectiveness is verified on a real stock price prediction task and a financial question answering task.

2) Aiming at the problem that users' questions have various forms and their intentions are difficult to understand, a dialogue generation alignment method based on human preferences was proposed. This method achieves model-independent, interpretable, and stable human preference alignment by collecting human preferences on real-world dialogue samples, leveraging the understanding and analysis capabilities of large language models for question optimization, and training a separate question optimization language model. Compared with the current mainstream language model alignment methods on two different benchmark machines, the experimental results show that the proposed method can simultaneously improve the recall

accuracy of knowledge documents and the consistency between model understanding and user intention.

Keywords: Open-Domain Dialogue; Retrieval Augmented Generation

目 录

摘 要	I
Abstract	II
第一章 绪论	1
1.1 研究背景和意义	1
1.2 国内外研究现状	2
1.2.1 面向垂直领域的对话生成研究现状	2
1.2.2 基于序列到序列模型的对话生成	2
1.2.3 基于检索增强生成技术的对话生成	3
1.2.4 基于人类偏好对齐的对话生成	4
1.3 本文主要研究内容	4
1.4 本文组织结构	5
第二章 相关研究技术	7
2.1 引言	7
2.2 自回归语言模型	7
2.2.1 自注意力机制	7
2.2.2 Transformer 架构	8
2.2.3 自回归语言模型	10
2.3 语义检索模型	10
2.3.1 BERT 模型	11
2.3.2 Bi-Encoder 与 Cross-Encoder 模型	12
2.4 本章小结	13
第三章 基于内外部知识对齐的检索增强对话生成	14
3.1 引言	14
3.2 基于内外部知识对齐的检索增强对话生成	14
3.2.1 算法总体框架	15
3.2.2 知识库构建	15
3.2.3 知识检索与回答生成	16
3.2.4 数据集预处理	18

3.2.5	两阶段模型微调	21
3.3	实验	21
3.3.1	实验数据和评价指标	21
3.3.2	实验细节	24
3.3.3	与现有方法的性能比较	24
3.4	消融实验	25
3.4.1	多级检索模块的有效性验证	25
3.4.2	不同数据集对性能的影响	26
3.5	本章小结	31
第四章	基于人类偏好对齐的检索增强对话生成	32
4.1	引言	32
4.2	基于人类偏好对齐的检索增强对话生成	32
4.3	总体方案	33
4.3.1	偏好数据采集阶段	33
4.3.2	优化提示词构建阶段	34
4.3.3	提示词有效性验证阶段	34
4.3.4	提示提优化器训练阶段	34
4.4	实验	36
4.4.1	实验数据和评价指标	36
4.4.2	实现细节	38
4.4.3	与现有方法的性能比较	38
4.4.4	提示词有效性验证模块的有效性	40
4.4.5	迭代优化对算法性能的影响	42
4.4.6	人类反馈对算法性能的影响	44
4.5	本章小结	45
	总结与展望	46
4.6	全文总结	46
4.7	未来展望	46
	攻读博士/硕士学位期间取得的研究成果	48
	参考文献	48

致 谢	49
-----------	----

第一章 绪论

1.1 研究背景和意义

在各行各业，垂直领域检索增强对话生成任务有助于提高用户体验、解决特定领域的问题和提供个性化的服务。它可以应用于医疗保健、金融、法律、教育、科技等领域，为用户提供更加专业、全面和个性化的信息交流和服务。因此，垂直领域检索增强对话生成任务对于满足用户需求、提高工作效率、提供个性化服务等方面具有重要意义。

垂直领域检索增强对话生成目前主要的挑战是领域知识丰富，用户问题多种多样且抽象。针对领域知识丰富的挑战，研究人员致力于构建更加智能和灵活的对话生成模型，能够充分利用领域内的丰富知识资源，包括专业词汇、行业规范、学术研究成果等，以更好地应对用户的专业性问题。这包括基于知识图谱和预训练语言模型的技术，以及定制化的领域知识处理方法。针对用户问题多样性和抽象性的挑战，研究人员在探索如何构建更加灵活和多样化的对话生成模型，能够理解和回答各种类型的问题，包括事实性问题、推理性问题、情绪化问题等。此外，研究人员也致力于开发更加智能、个性化的对话交互方式，以满足用户多样化的沟通需求。同时，深度学习技术在对话生成任务中的应用也在不断演进。例如，针对领域知识丰富和用户问题多样性的挑战，研究人员正在探索如何通过多模态融合（如文本、图像、语音）、增强学习、迁移学习等技术手段，提高对话生成模型的适应性和泛化能力。

然而，相对于开放域对话生成，垂直领域对话内容更复杂、背景知识要求更多，生成准确、真实、可靠回复的技术挑战性更高，使垂直领域检索增强对话生成存在以下几个难点：

- 难点 1：垂直领域背景知识丰富，且逻辑相对复杂。
- 难点 2：用户问题形式多样，且意图难以理解。

近年来，越来越多研究人员投身面向垂直领域的检索增强对话生成研究。来自斯坦福大学、加利福尼亚大学、清华大学等国内外高校和企业研究机构的学者在该领域开展了大量研究工作。国际上的一些主流学术会议和学术期刊也将垂直领域检索增强对话生成作为一个研究热点，如 ACL、EMNLP 等国际会议。综上，面向垂直领域的检索增强对话生成研究是当前人工智能领域的重点与热点，不仅具有重要的理论价值，而且具有丰富的实际应用价值。

1.2 国内外研究现状

人机对话系统一直是人工智能领域的重要研究方向，其旨在模拟人类并与人类形成连贯通顺的对话。对话系统主要分为三类：问答型对话、任务型对话、闲聊型对话。问答型对话多是一问一答，用户输入问题，系统查找相关知识返回答案或直接返回问题的解答，如语音客服，技术答疑助手等；任务型对话有明确的对话主题和任务目标，用于精确识别用户的意图并执行相应的动作，如智能订票助手等；闲聊型对话系统没有固定的领域和话题，其目标是产生有趣且富有信息量的回复，为用户提供情绪价值，如微软小冰等。

1.2.1 面向垂直领域的对话生成研究现状

垂直领域的对话内容相较于开放域对话具有更强的知识性与复杂性，直接使用开放域对话生成的方法无法得到准确可靠的回复。为此，许多研究者对面向垂直领域的对话生成展开研究。[] 等人提出 DukeNet，构建了一个双向知识交互网络，实现面向垂直领域知识的对话生成。[] 等人提出扁鹊模型，通过构建高质量的医学领域对话数据集，得到预训练模型在下游对话任务上的良好性能。[] 等人提出 KdConv，即中文多领域知识驱动对话数据集，通过在数据集上训练将知识注入模型中。

1.2.2 基于序列到序列模型的对话生成

近年来，随着 GPT，BERT，T5 等 Seq2Seq 预训练模型的兴起，开放域对话系统、问答型对话系统的研究重心也从基于规则的流水线式方法和检索式方法转向数据驱动的端到端生成式方法。到 2020 年，基于预训练模型的端到端方法已经成为对话系统的主流，使用这类方法的对话系统包括 Blender、DialoGPT、Meena 和 PLATO 等。

2022 年 11 月，随着 ChatGPT 的发布，国内外的大型语言模型如雨后春笋般出现，如 ChatGLM、文心一言、通义千问、LLaMA、百川等。从此，基于自回归式大型语言模型的对话生成成为了对话系统的主要研究方向。

基于大型语言模型的对话生成可分为两类：（1）基于提示词工程的方法；（2）基于监督数据微调的方法。

[] 等人提出 Zero-shot 提示方法，利用大型语言模型进行范式迁移，无需训练即可让预训练模型完成下游任务，但该方法得到的输出可能不够准确或不符合预期。为此，[] 等人提出思维链（Chain-of-Thought, CoT）提示方法，通过在模型输入的结尾引导模型一步步思考，并在给出回复前先对问题进行拆解和分析，使得模型最终输出的答案

更加准确。[] 等人提出 Few-shot 提示方法，[] 等人引入自洽性（Self-Consistency）解码策略，[] 等人提出思维树（Tree-of-Thought, ToT）提示方法，[] 等人提出思维图（Graph-of-Thought, GoT）提示方法，。这类方法不需要重新训练模型即可提升模型性能，但提升受限于预训练模型自身的能力上限。

[] 等人提出指令微调（Instruction Tuning）方法，通过监督训练，让语言模型学会按照指令要求完成任务，从而具备遵循指令的能力，即使面对训练中未曾见过的任务，模型也能够生成合适的回复。[] 等人提出 Prefix-Tuning 方法，在模型输入首端增加一个连续的、任务相关的嵌入向量来进行训练，在显著减少训练参数量的情况下提升模型在自然语言生成任务上的性能。与 Prefix-Tuning 思想相似，[] 等人提出提示学习（Prompting Tuning）方法，通过在输入中插入一段任务特定的、可被训练的离散提示词元，获得了与微调相近的效果，同时超过了人工设计提示词的性能。[] 等人在 Prefix-Tuning 的基础上，进一步提出 P-Tuning v2 方法，在模型的每一层上都加上了可训练的层级提示词元，且对于不同难度的任务使用不同的提示长度。[] 等人提出了低秩自适应（Low-Rank Adaptation, LoRA）方法，通过使用低维结构来近似大模型的高维结构，以降低模型训练的复杂度和计算开销。总体来说，基于监督数据微调的方法性能优于基于提示词工程的方法，但存在高质量标注数据难获取的问题。

1.2.3 基于检索增强生成技术的对话生成

大型语言模型具有良好的自然语言理解和自然语言生成能力，但往往面临幻觉问题，即回复内容不符合事实，甚至胡编乱造。这可能是因为模型在预训练阶段记忆了错误的知识，或是推理时的输入是预训练阶段没有遇到过的长尾知识。针对后者，[] 等人提出检索增强生成（Retrieval Augmented Generation, RAG）技术，通过构建本地知识库，在对话阶段从知识库中召回与用户问题相关的文档，作为外部知识辅助语言模型给出回复，很好地缓解了大型语言模型的幻觉问题和实时性不足的问题。[] 等人提出 Query2doc 方法，利用大型语言模型对用户问题生成伪文档，以提升知识库的召回准确度，以减少无关文档对语言模型回复产生噪声干扰。[] 等人在 Query2doc 方法的基础上提出思维链技术与伪相关反馈（Pseudo-Relevance Feedback, PRF）算法相结合的方法，在多个基准数据集上获得了超过 Query2doc 方法的效果。[] 等人的研究表明，当相关信息出现在模型输入上下文的开头或结尾时，模型的性能最好，相关信息出现在中间位置时模型表现最差，且随着输入上下文的增长，模型性能显著下降，表明模型很难从长输入上下文中检索和使用相关信息。因此召回的知识文档数量及其在模型输入中的位置

对模型性能至关重要。[] 等人提出 Self-RAG 方法，生成模型通过检索召回多个相关文档，并通过并行处理和排序选择最合适的回复。[] 等人提出 ChatLaw 中文法律大模型，在 RAG 的基础上。融入法律意图识别、法律关键词提取等模块，满足法律相关领域的应用需求。这类方法在背景知识丰富且逻辑相对复杂的专业领域上表现不佳。

1.2.4 基于人类偏好对齐的对话生成

大型语言模型为促进通用人工智能的实现制定了蓝图，其主要目标是成为对人类有帮助的、忠诚的、无害的助手。[] 等人提出基于人类反馈的强化学习（Reinforcement Learning with Human Feedback, RLHF）算法，其工作过程包括采集高质量监督数据集并微调语言模型、收集人类偏好排名数据集并训练奖励模型、执行近端策略优化（Proximal Policy Optimization, PPO）强化学习。该算法能够很好地帮助模型生成符合人类偏好的回复，同时减少生成式模型中的偏见。然而，RLHF 算法在 PPO 强化学习训练阶段需要同时使用四个大型语言模型，导致训练计算资源开销大，同时 PPO 强化学习过程不稳定，导致模型难以训练。为此，许多研究者提出其他的人类偏好对齐算法替代 RLHF 算法。[] 等人提出 RAIN 可回滚自回归推理算法，利用语言模型评估自己生成的结果，并用评估结果来指导语言模型输出，以确保输出符合人类偏好，无需微调即可实现语言模型与人类偏好的对齐。[] 等人针对 PPO 训练不稳定的问题，通过实验探索了 PPO 训练中最关键的技巧，并用 PPO-max 表示这套最佳实现方式。[] 等人提出 RAFT 方法，。[] 等人提出 RRHF 算法，。[] 等人提出 RLAIIF 方法，用语言模型代替人类进行偏好的标记工作。[] 等人提出 BPO 黑盒 Prompt 优化算法，。[] 等人提出 DPO 算法，。[] 等人提出 OPRO 算法，。[] 等人提出 RSO 算法，。此外，[] 等人提出 ReMax 算法，通过削减 PPO 中冗余庞大的计算开销，节省 RLHF 算法 50% 的内存消耗，并加快 2 倍训练速度。现有方法一方面需要高昂的人力成本完成偏好数据标注，另一方面存在训练计算资源开销大、难训练的问题。

1.3 本文主要研究内容

如前所述，面向垂直领域的检索增强对话生成的关键在于解决以下挑战：1）垂直领域背景知识丰富，且逻辑相对复杂；2）用户问题形式多样，且意图难以理解。

为解决上述挑战，本文进行了两个方面的研究，研究思路与研究内容如图 1-1 所示。包括基于内外部只是对齐的垂直领域对话生成和基于人类偏好的对话生成对齐。具体研究内容如下：

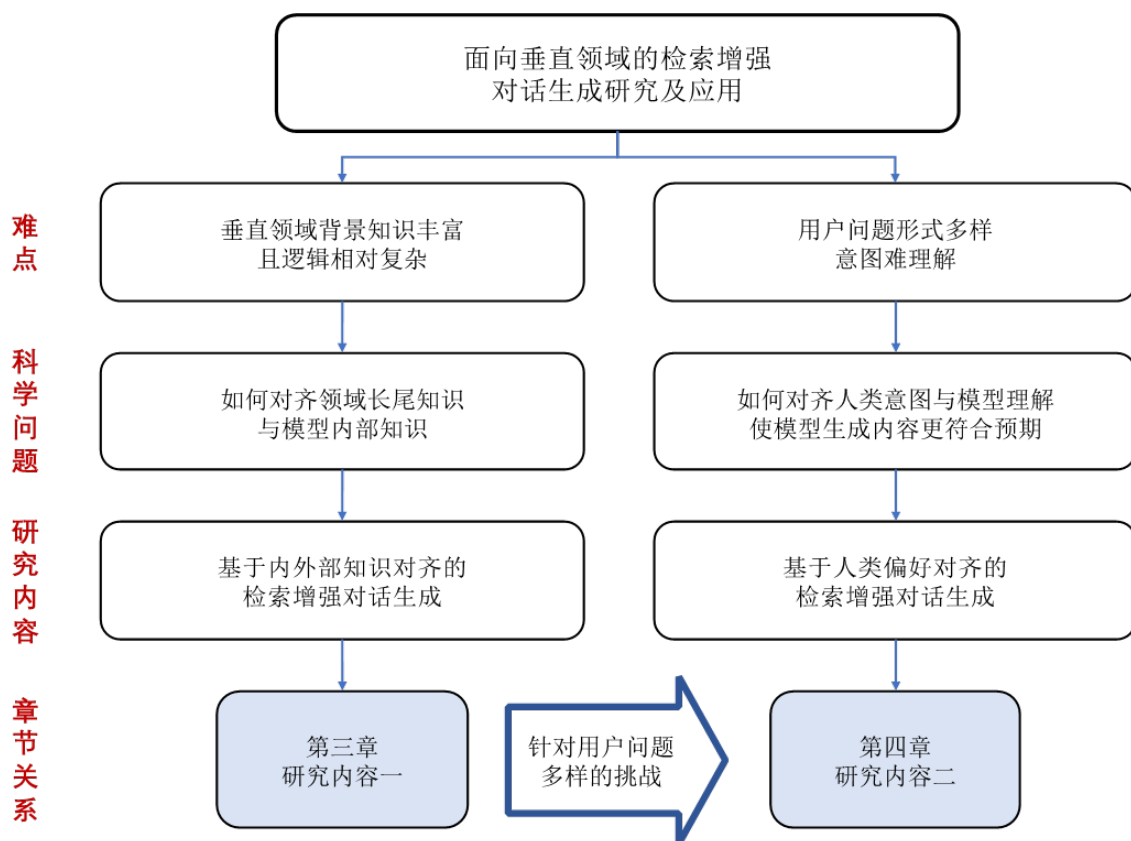


图 1-1 本文的研究思路与研究内容。

1) 基于内外部知识对齐的检索增强对话生成

大模型在预训练中没见过垂直领域的长尾知识，因此检索增强所补充的知识文档可能未能与模型内部知识完全对齐。为此，本文对内外部知识的对齐开展研究，利用一个语义切分模块提取知识文档的文档级信息和实体级信息，并将提取出来的知识分别用于构建外部知识库和内部知识注入，实现垂直领域对话模型内外部知识对齐。

2) 基于人类偏好对齐的检索增强对话生成

弥合人类意图和 LLM 之间的对齐差距是提升对话模型回复质量的关键。然而，主流的对齐方法需要使用强化学习算法，使得训练成本高、难度大。为此，本文研究低成本、高效的人类偏好对齐方法，通过采集人类对真实场景对话样本的偏好，利用大型语言模型的理解与分析能力进行问题优化，并训练单独的问题优化语言模型，实现了与模型无关的、可解释、效果稳定的人类偏好对齐。

1.4 本文组织结构

本文的组织结构和章节关系安排如下：

第一章是绪论部分，介绍了本文的研究背景与意义，分析了该研究方向的国内外研

究现状，最后阐释了本文主要的研究内容和贡献。

第二章是相关研究技术，介绍了本文在算法设计中使用到的相关技术，其中包括 GPT 模型与 BERT 模型、低秩自适应微调。

第三章提出了一种基于内外部知识对齐的垂直领域对话生成方法，利用一个语义切分模块提取知识文档的文档级信息和实体级信息，并将提取出来的知识分别用于构建外部知识库和内部知识注入，实现垂直领域对话模型内外部知识对齐。相关研究成果已经发表于自然语言处理和计算语言学领域的顶级会议 COLING。

第四章提出了一种基于人类偏好的对话生成对齐方法，通过采集人类对真实场景对话样本的偏好，利用大型语言模型的理解与分析能力进行问题优化，并训练单独的问题优化语言模型，实现与模型无关的、可解释、效果稳定的人类偏好对齐。

第五章对全文研究工作进行了总结，并对领域未来研究方向进行了展望。

第二章 相关研究技术

2.1 引言

在本章中，本文将简要介绍本文研究用到的相关技术。由于本文将采用自回归语言模型来设计对话生成模型，同时使用语义检索模型设计检索增强框架，本文将首先介绍自回归语言模型中所使用的自注意力机制与 Transformer 模型架构，然后介绍 GPT 架构的自回归语言模型的基础结构。其次，介绍语义检索模型中的 BERT 模型架构，最后介绍 Bi-Encoder 和 Cross-Encoder 结构的语义检索模型。

2.2 自回归语言模型

2.2.1 自注意力机制

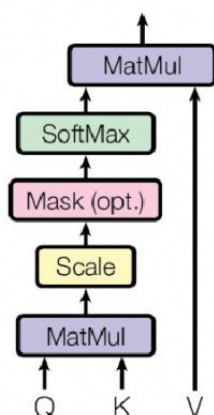


图 2-1 缩放点积注意力

自注意力机制（Self-attention）是一种将输入序列中不同位置的词关联起来并计算输入序列特征表示的注意力机制 [48]，而 Vaswani 等人 [49] 在该自注意力机制的基础上提出了一种缩放点积的自注意力机制（Scaled Dot-Product Attention），并使用该机制构建出著名的 Transformer 模型。该自注意力机制的结构如图 2-6 所示。如图 2-6 所示，其中的 Q、K、V 为三个向量矩阵，其中的行代表输入序列中的词向量表示。其中 Q 与 K 先进行矩阵相乘操作（MatMul）后进行缩放（Scale），然后经过一个可选的 Mask 操作，最后经过一个 Softmax 层得到权重矩阵 A。其中的 Mask 操作用以将权重矩阵 A 中的某些位置的权重值置为 0。权重矩阵 A 中第 i 行代表第 i 个词相对于输入序列中所有词的权重值。最后将 A 与 V 两个矩阵相乘得到输出。当利用该自注意力机制对输入序列进行编码时，Q、K、V 三者为同一个向量矩阵，该机制捕捉的是输入序列与其自身

的依赖信息，以得到整个序列的编码向量；当利用该自注意力机制在解码器中计算解码序列与编码特征的注意力时，K 和 V 相同而 Q 与 K、V 不同，该机制捕捉的是 Q 与 K 之间的依赖信息。

2.2.2 Transformer 架构

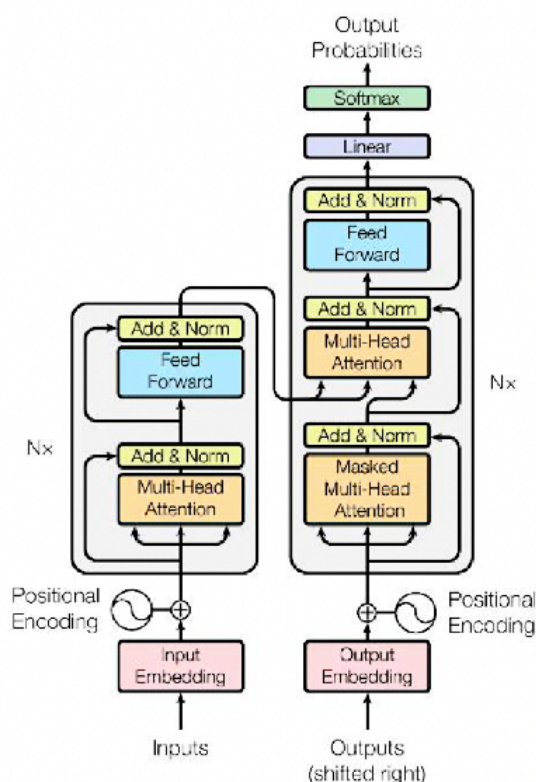


图 2-2 Transformer 模型架构

Transformer 模型是基于编码器-解码器框架的。如图 2-7 所示，与传统 Seq2Seq 使用 RNN 构建解码器和编码器不同，Transformer 模型使用自注意力机制设计编码器和解码器。RNN 在处理序列信息时是按顺序依次处理的，每一时刻的运算是基于上一时刻的。RNN 的串行处理特征导致了其只能进行串行运算，在序列长度很长，数据量很大时，将需要非常大的训练时间。而 Transformer 模型直接抛弃 RNN 的序列处理方式，而是使用自注意力机制直接对整个序列进行编码以及学习输入输出的全局依赖关系。Transformer 模型结构如图 2-7 所示。其中图中的 Inputs 和 Outputs 分别为编码器和解码器的输入序列；位置编码（Positional Encoding）层是为了帮助自注意力机制编码输入序列中的顺序信息。具体做法是将输入序列中的每一个位置，计算一个位置向量，该向量的计算公式如下：

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{model}}) \quad (2-1)$$

$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{model}}) \quad (2-2)$$

其中公式中的 pos 为在序列中的位置编号， i 为位置向量的维度下标， $modeld$ 为向量维度。如公式 (2-7) 和 (2-8) 所示，对于位置向量中的偶数位置和奇数位置的值分别使用正弦、余弦的函数进行计算得到。另外，如图 2-7 所示，在 Transformer 编码器中，输入序列经过一个向量层后得到输入序列的向量矩阵，然后与位置编码向量相加后输入到多头自注意力（Multi-HeadAttention）层中，最后经过残差连接和层级正则化层（Add&Norm）后经过一个前馈神经网络（Feed Forward）层得到输入序列的特征向量。在 Transformer 解码器中，需要先使用 Masked 多头自注意层对已生成的输出序列进行编码，然后使用另外一个多头自注意层来计算与输入序列各个位置的注意力值，最后经过一个线性层（Linear）和 Softmax 层，得到当前解码时刻的输出概率分布（Output Probabilities）

在 Transformer 模型中，最核心的是多头自注意力层。该多头自注意力层是将多个不同的自注意力层结合在一起构成的，其结构如图 2-8 所示。如图 2-8 所示，多头自注意力层将输入的 Q、K、V 矩阵依次经过 h 个自注意力层（Scaled Dot-Product Attention）后拼接在一起，最后经过一个全连接层产生输出。其公式表示如下：

$$MultiHead(Q, K, V) = \text{concat}(h_1, \dots, h_h) \dots W^O \quad (2-3)$$

$$h_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (2-4)$$

在公式 (2-19) 和公式 (2-20) 中，Q、K、V 中每一行的维度为 $model_d$ 。由此可知，分别使 h 个不同的全连接层将 Q、K、V 三个输入矩阵分别映射到长度为 kd 的特征空间中，然后分别经过 h 个自注意力层运算后重新拼接起来。结合位置编码和多头自注意力层，Transformer 模型可以学习输入序列中不同位置、不同维度的信息。

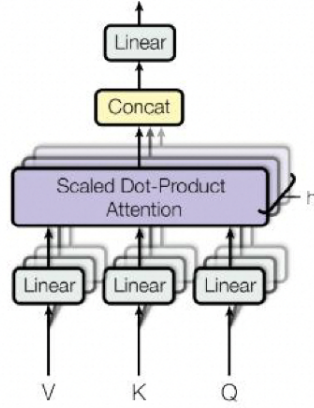


图 2-3 多头自注意力机制

2.2.3 自回归语言模型

GPT 模型采用了 Transformer 架构，它由多个堆叠的 Transformer 解码器组成，每个解码器包含多个注意力头和前向神经网络。这使得模型可以从多个抽象层次对文本进行建模，从而更好地捕捉文本的语义信息。同时，GPT 模型使用了一种称为“生成式预训练”（Generative Pre-Training, GPT）的技术。预训练分为两个阶段：预训练和微调（fine-tuning）。在预训练阶段，GPT 模型使用了大量的无标注文本数据集，例如维基百科和网页文本等。通过最大化预训练数据集上的 log-likelihood 来训练模型参数。在微调阶段，GPT-1 将预训练模型的参数用于特定的自然语言处理任务，如文本分类和问答系统等。通过微调 GPT 模型，可以针对特定的任务进行优化，例如文本生成、机器翻译和对话系统等。

自回归模型应用线性回归，其输出中的滞后变量取自先前的步骤。与线性回归不同，自回归模型除了先前预测的结果外，不使用其他自变量。

$$p(x) = \prod_{i=1}^n p(x_i | x_1, x_2, \dots, x_{i-1}) = \prod_{i=1}^n p(x_i | x_{<i}) \quad (2-5)$$

2.3 语义检索模型

语义检索模型被广泛应用于搜索、推荐、数据挖掘等重要领域，将自然形式的数据样本（如语言、代码、图片、音视频）转化为向量（即连续的数字序列），并用向量间的“距离”衡量数据样本之间的“相关性”。从结构上，主流的语义检索模型是 BERT 模型。从方法上，语义检索模型可以分为 Bi-Encoder 型和 Cross-Encoder 型。

2.3.1 BERT 模型

BERT 模型由 Transformer 架构的编码器组成，一个 transformer 的 encoder 单元由一个 multi-head-Attention + Layer Normalization + feedforward + Layer Normalization 叠加产生，BERT 的每一层由一个这样的 encoder 单元构成。在比较大的 BERT 模型中，有 24 层 encoder，每层中有 16 个 Attention，词向量的维度是 1024。在比较小的 BERT 模型中，有 12 层 encoder，每层有 12 个 Attention，词向量维度是 768。在所有情况下，将 feed-forward/filter 的大小设置为 $4H$ （ H 为词向量的维度），即 $H = 768$ 时为 3072， $H = 1024$ 时为 4096。

这种 transformer 的结构可以使用上下文来预测 mask 的 token，从而捕捉双向关系。

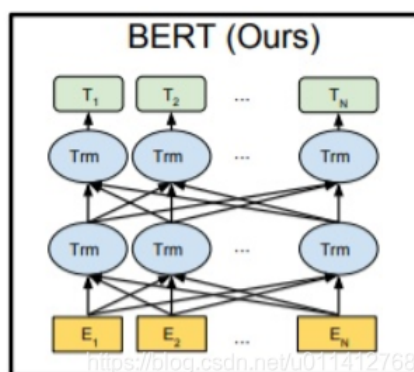


图 2-4 BERT 模型网络结构图

在 BERT 中，输入的向量是由三种不同的 embedding 求和而成，分别是：

1) wordpiece embedding: 单词本身的向量表示。WordPiece 是指将单词划分成一组有限的公共子词单元，能在单词的有效性和字符的灵活性之间取得一个折中的平衡。

2) position embedding: 将单词的位置信息编码成特征向量。因为我们的网络结构没有 RNN 或者 LSTM，因此我们无法得到序列的位置信息，所以需要构建一个 position embedding。构建 position embedding 有两种方法：BERT 是初始化一个 position embedding，然后通过训练将其学出来；而 Transformer 是通过制定规则来构建一个 position embedding。

3) segment embedding: 用于区分两个句子的向量表示。这个在问答等非对称句子中是用区别的。

最终的模型输入通过三者相加得到，wordpiece token embedding + segment embedding + position embedding。

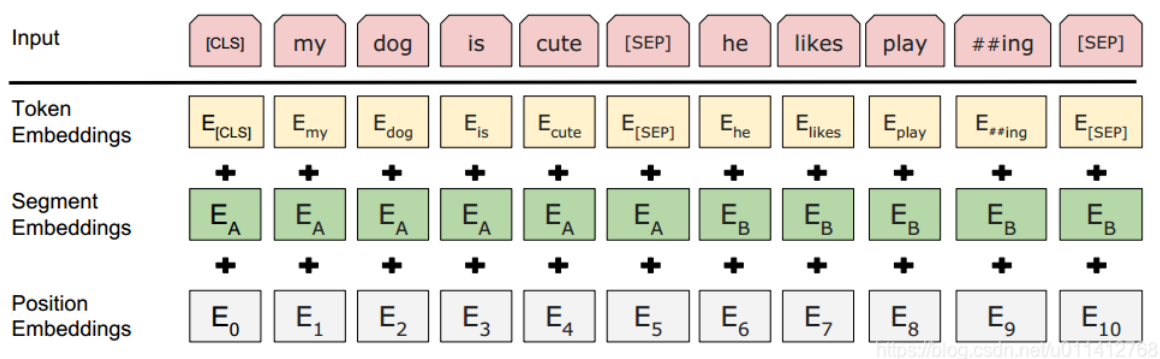


图 2-5 bert 模型输入组成

2.3.2 Bi-Encoder 与 Cross-Encoder 模型

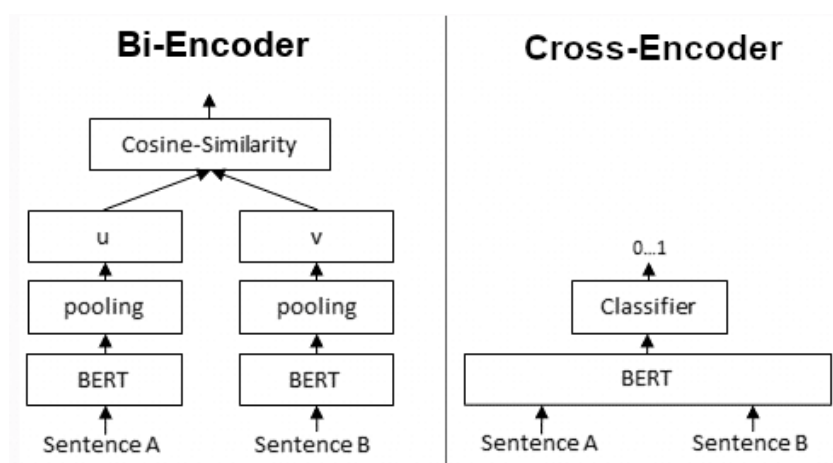


图 2-6 Bi-Encoder 模型与 Cross-Encoder 模型结构示意图

如图所示，左边的模型表示 Bi-Encoder 模型，右边的模型表示 Cross-Encoder 模型。

对于 Bi-Encoder 模型，给定句子生成句子的 embedding。我们将句子 A 和 B 独立地传递给 BERT，从而产生句子嵌入 u 和 v ，然后可以使用余弦相似度比较。

对于 Cross-Encoder 模型，同时将两个句子传递给 Transformer 网络。它产生一个介于 0 和 1 之间的输出值，表示输入句对的相似性。不产生句子的 embedding。并且，无法将单个句子传递给 Cross-Encoder。

Cross-Encoder 只要有一组要评分的预定义句子对，就可以用。例如，有 100 个句子对，想获得这 100 个句子对的相似度分数。当需要在向量空间中句子的 embedding 进行有效比较时。Cross-Encoders 在这些应用上将是错误选择：使用 Cross-Encoders 聚类 10,000 个句子需要计算大约 5000 万个句子组合的相似度分数，这需要大约 65 小时。使用 Bi-Encoder，您可以计算每个句子的嵌入，只需 5 秒。然后您可以执行聚类。

Cross-Encoder 比 Bi-Encoder 具有更高的性能，但是，对于大型数据集，它们不能很

好地扩展。在这里，结合交叉编码器和双编码器是有意义的，例如在信息检索/语义搜索场景中：首先，您使用高效的双编码器来检索例如前 100 个最相似的查询语句。然后，您使用交叉编码器通过计算每个（查询、命中）组合的分数来重新排列这 100 个命中。

2.4 本章小结

本章首先介绍了自回归语言模型中的自注意力机制与 Transformer 模型架构，然后介绍 GPT 架构的自回归语言模型。其次，介绍语义检索模型中的 BERT 模型架构，最后介绍语义检索模型中的 Bi-Encoder 和 Cross-Encoder 结构。基于 GPT 架构的自回归语言模型是目前垂直领域对话生成任务中主流的基础模型。基于 BERT 架构的语义检索模型是目前向量检索技术中主流的基础模型。本文提出的方法也是基于这两个基础模型进行设计的。

第三章 基于内外部知识对齐的检索增强对话生成

3.1 引言

基于大语言模型的对话生成存在事实性、实时性不足的问题。现有方法主要通过检索增强生成方法进行垂直领域对话生成，以输入的形式将垂直领域知识引入对话模型中。然而，面对垂直领域中较难的问题时，模型往往不能很好理解知识文档中的复杂信息，导致检索增强效果不佳。大模型在预训练中没见过垂直领域的长尾知识，因此检索增强所补充的知识文档可能未能与模型内部知识完全对齐。为此，本章研究如何对齐外部知识文档的垂直领域长尾知识和模型内部知识，进而使模型兼具事实性与垂直领域推理能力。同时，以金融领域问答为应用场景对算法进行验证。

3.2 基于内外部知识对齐的检索增强对话生成

问题描述：

对于股票趋势预测任务，给定一组公司 $C = c_i (i = 1)^N$ 以及对应的知识文档 $D = d_j (j = 1)^M$ ，对话系统给出该股票的未来趋势预测

$$Pred_i = \pi(c_i, d_j), Pred_i \in up, down \quad (3-1)$$

其中， π 表示股票预测系统， d_j 是检索得到的与公司 c_i 相关的知识文档。目标是选择出一批被预测股价会上涨的公司。

$$C_{chosen} = c_i \mid c_i \in C \wedge Pred_i = up \quad (3-2)$$

对于金融问答任务，我们将一个多轮对话视为两个对话者之间的诸多“问题-回复”对。令 Q_t 和 R_t 表示在第 t 轮对话时的用户问题和系统回复， $H_t = [Q_0, R_0, \dots, Q_{(t-1)}, R_{(t-1)}]$ 作为对话历史。我们将金融问答任务的形式定义为，给定对话历史、用户问题和检索到的相关文档，对话系统 π 能够给出相应的回复：

$$R_t = \pi(d_k, H_t, Q_t) \quad (3-3)$$

其中, d_k 表示检索到的与 Q_t 相关的知识文档。

3.2.1 算法总体框架

我们采用检索增强生成技术来提升 LLM 的 QA 能力, 该

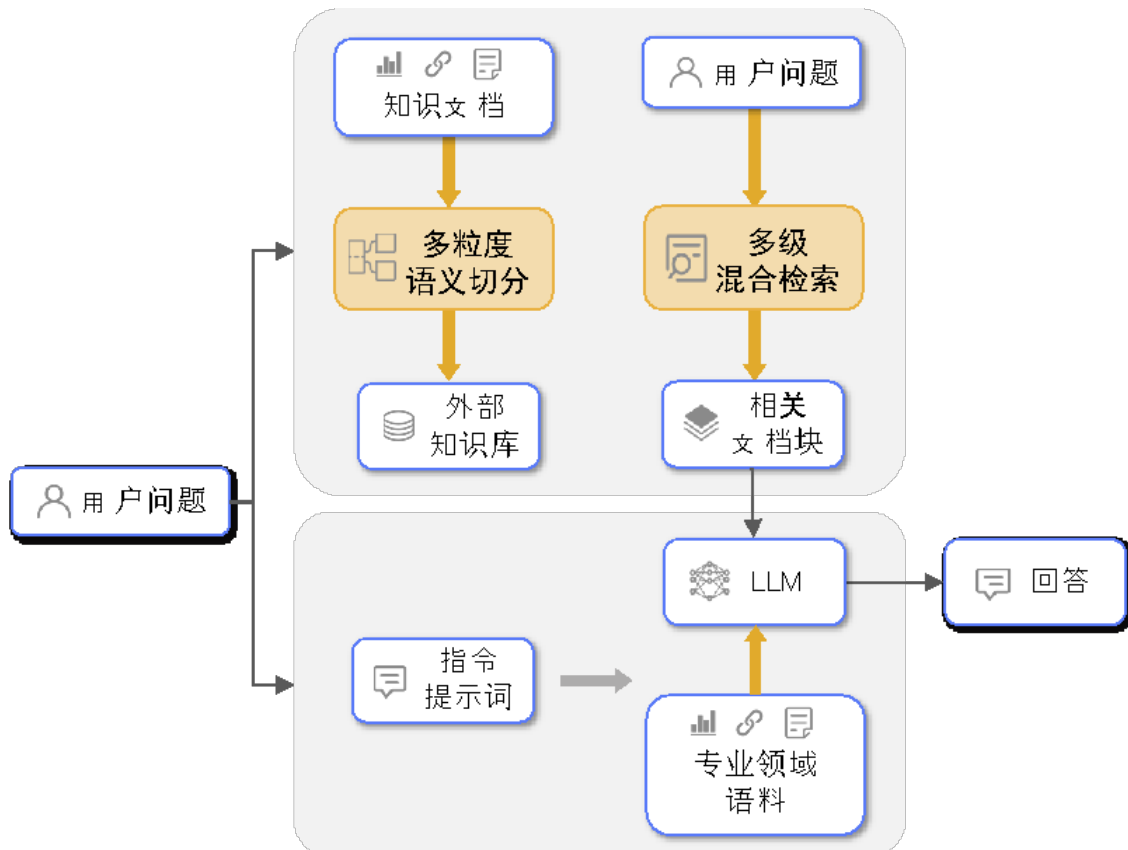


图 3-1 本章所提出的垂直领域对话框架示意图。

3.2.2 知识库构建

如图 3-1 所示, 外部知识库的构建是检索增强生成中的的重要组成部分, 用于高效存储和检索相关知识文档。

为了提升知识库召回文档的准确度和效率, 本章节对原始知识文档进行多粒度语义切分, 从文档中提取出关键信息。如图 3-2 所示, 我们采用两种切分策略: 粗粒度文档级总结和细粒度实体级对话生成。其中, 语义切分过程使用大型语言模型 (如 ChatGPT 模型) 通过设计相应的提示词完成。对于文档 d_k , 其语义切分过程如下:

$$s_k = LLM_{sum}(d_k) \quad (3-4)$$

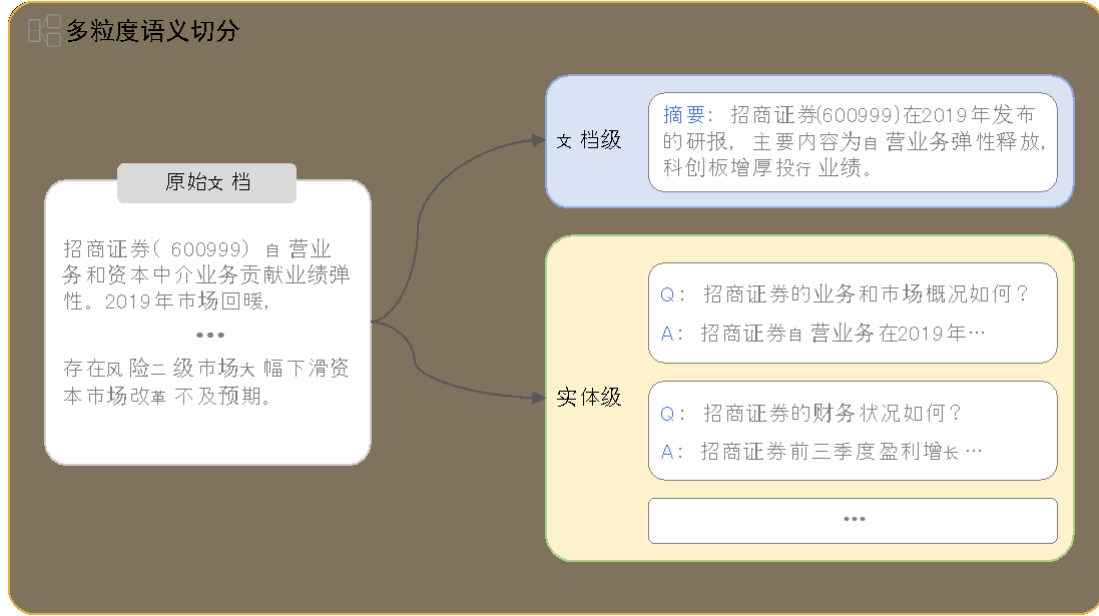


图 3-2 对话框中的多粒度语义切分模块示意图。

$$(q_k0, a_k0), (q_k1, a_k1), \dots = LLM_q(d_k) \quad (3-5)$$

其中, s_k 表示文档 d_k 的摘要, $(q(k_), a(k_))$ 是所生成对话的“问题-回答”二元组。例如, 假设 d_k 是与“股票 k 线”相关的文档, $q(k_)$ 则可能是“股票 k 线是什么?”。

我们以文档级摘要切分策略为例, 给定文档摘要 s_k , 我们通过句子嵌入模型得到其嵌入向量 e_{sk} 。该向量将被存储在知识库中, 作为数据库索引被用于后续的检索步骤。

$$e_{sk} = \text{SentEmbed}(s_k) \quad (3-6)$$

其中, SentEmbed 是句子嵌入模型, 如 BGE、SGPT 等。

3.2.3 知识检索与回答生成

知识检索过程如图 3-3 所示。从知识库中检索相关知识文档, 首先将用户问题 Q 输入与 3.2.2 节相同的句子嵌入模型中, 获得其嵌入向量 e_Q 。

$$e_Q = \text{SentEmbed}(Q) \quad (3-7)$$

依次计算用户文档嵌入向量 e_Q 与各文档嵌入向量 e_{sk} 的余弦相似度, 选择相似度

最高的文档作为外部知识，辅助 LLM 生成回复。

$$d^* = \argmax_{\mathbf{T}}(d_k) \quad ||e_Q||e_{sk}||/||e_Q||e_{sk}|| \quad (3-8)$$

其中，对于细粒度实体级对话生成的切分策略，上式中的 s_k 和 d_k 可分别被替换为 $q(k_-)$ 和 $a(k_-)$ 。



图 3-3 对话框架中的多级混合检索模块示意图。

给定对话历史 H_t ，用户问题 Q_t ，以及检索到的与用户问题 Q_t 相关的文档 d^* ，目标是获得第 t 轮对话的回复 R_t 。然后，我们拼接提示词模板、知识文档、对话历史和用户问题，以得到 LLM 的输入 I_t 。将 I_t 传入 LLM，即可得到回复 R_t 。

$$I_t = \text{concat}(\text{Prompt}, d^*, H_t, Q_t) \quad (3-9)$$

$$R_t = LLM(I_t) \quad (3-10)$$

/subsection 趋势预测与后处理

如图所示，给定输入 I_i ，我们利用 LLM 来预测股票的涨跌趋势，这可以看作是一个二分类任务。通过将 I_i 输入 LLM，得到关于 c_i 的回复文本 Res_i 。

$$Res_i = LLM(I_i) \quad (3-11)$$

然后，我们使用基于规则的方法从 Res_i 中提取出趋势预测结果 $Pred_i$ 。最后，我们选择所有被预测为“上涨”的股票，得到股票集合 C_{chosen} 。

$$Pred_i = \begin{cases} up, & \text{if "up" } \in Res_i \\ down, & \text{else} \end{cases} \quad (3-12)$$

$$C_{chosen} = c_i \mid |Pred_i = up \quad (3-13)$$

另外，我们按月滚动执行该投资策略。每个月，对于 C_{chosen} 中的所有股票 c_i ，我们会持有一整个月。投资组合中的每种股票的比例是通过市值加权计算得到的。

$$AR_m = AR(m-1) + \sum_{(c_i \in C_{chosen})} v_i * R(c_i) \quad (3-14)$$

其中， AR_m 表示第 m 个月的累计收益， $R(c_i)$ 表示股票 c_i 的收益。 v_i 代表股票 c_i 在投资组合中所占的比例。 v_i 是公司 c_i 的市值。

$$= v_i / (\sum_{(n=1)}^N v_n) \quad (3-15)$$

3.2.4 数据集预处理

如图所示，我们基于金融领域知识文档构建 AlphaFin 数据集，该数据集包含三部分：通用金融数据集、金融新闻与问答数据集、金融研报数据集。各数据集具体来源和

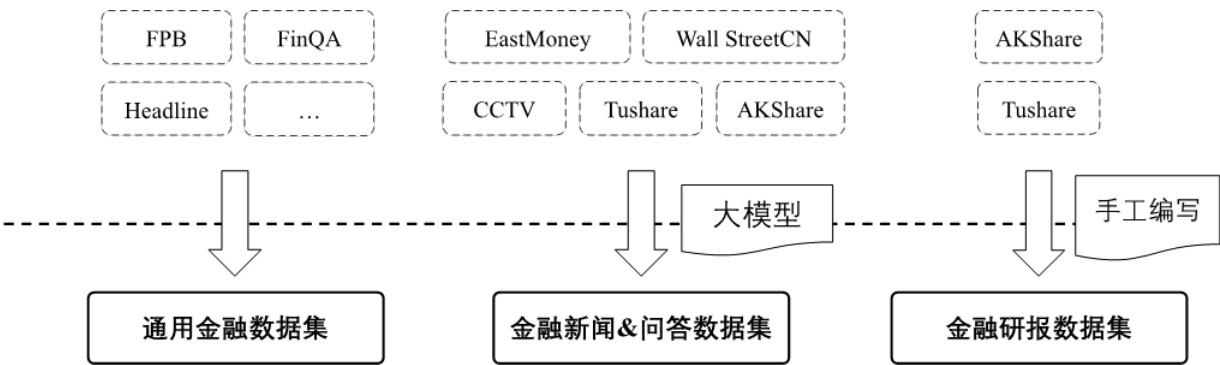


图 3-4 数据集预处理过程示意图。

预处理方式如下：

通用金融数据集：由传统金融开源数据集构成，如 FPB、FinQA、ConvFinQA、Headline 等。这些开源数据集主要是英文数据集，因此我们从中采样一部分，用于增强 LLM 的多语言能力、信息抽取能力和摘要能力。

金融新闻数据集：为 llm 提供真实世界的金融知识，我们整合了在线新闻资源，如中央电视台的金融板块和华尔街 CN。利用 ChatGPT 提取每条新闻的摘要，构建金融新闻数据集。这个过程提高了 llm 为金融新闻生成汇总的能力。

金融问答数据集：这部分包含了 Tushare (Tushare, 2021) 和 AKshare (AK- share, 2021) 的股票价格和其他财务数据。它利用序列数据，例如真实世界的股票价格趋势 (例如…，170, 173, 171, 175, 173, 170, …)。给定源数据以顺序格式呈现，我们利用 ChatGPT 和以下提示，在其上生成财务问题。

表 3-1 金融问答数据集预处理提示词

提示词
基于 <content>，请提出一个金融问题。
输入：<sequential data>
输出：<question>

金融研报数据集：我们通过 DataYes (DataYes, 2021) 构建了财务报告数据集，包括机构对公司进行的专业分析和知识。我们手动对齐公司的财务报告及其在报告发布当天的股票价格，并使用以下格式来生成最终数据。

表 3-2 金融问答数据集预处理提示词

数据集	Raw	CoT
系统指令	<p>请根据下方提供的该股票相关研报与数据，对该股票的下个月的涨跌，进行预测，请给出明确的答案，“涨”或者“跌”。同时给出这个股票下月的涨跌概率，分别是：极大，较大，中上，一般。</p>	
输入	<p>< 公司研报 ></p> <p>发布日期: <date></p> <p>研报题目: <title></p> <p>目标价格: <target price></p> <p>研报摘要: <abstract></p> <p>< 市场数据 ></p> <p>1. 股票价格（元）:<price>;</p> <p>2. 日涨跌幅（%）:<change>;</p> <p>3. 日成交量（亿）:<volumn></p>	
输出	<p>这个股票的下月最终收益结果是:’跌’, 下跌概率: 中上</p>	<p>通过研报和市场走势数据可以得出以下结论:</p> <p>1、基本面: <basic analysis>。</p> <p>2、技术面: <tech analysis>。</p> <p>因此，我们预测，这个股票的下月最终收益结果是:’跌’, 下跌概率: 极大</p>

表 3-3 金融问答数据集预处理提示词

数据集	类别	大小	输入长度	输出长度	语言
通用金融数据集		42,373	712.8	5.6	英文
金融新闻 & 问答数据集	新闻	21,000	1313.6	40.8	中文
	问答	79,000	497.8	64.2	中文
金融研报数据集	Raw	120,000	2203.0	17.2	中文
	CoT	200	2184.8	407.8	中文

3.2.5 两阶段模型微调

为了实现外部知识与模型内部知识的对齐，我们对 LLM 进行微调。首先，我们使用专业领域的知识文档构建两类数据集：通用 QA 数据集和 CoT 分析数据集。通用 QA 数据集的构建过程与 3.2.2 节中的细粒度实体级对话生成过程相同，其作用是让模型具备基本的专业领域理解能力；CoT 分析数据集的构建需要先采集针对知识文档的专业领域问题，然后人工撰写 CoT 详细分析的专业解答，其作用是让模型具备专业领域的详细分析能力和文本生成能力。

LLM 微调过程分为两个阶段：1) 首先在通用 QA 数据集上进行微调，将专业领域的基本知识注入模型内部；2) 然后在 CoT 分析数据集上进行微调，进一步对齐知识文档中的长尾知识与模型内部的知识。所有的微调过程均采用 LoRA 方法，以缓解训练过拟合问题，同时降低训练成本。微调后的 LLM 将作为对话框架中的对话生成模型，与用户交互。

3.3 实验

在本节中，我们以金融领域为应用场景对算法进行实验验证。金融领域分析任务主要包括两个部分：1) 股票趋势预测；2) 金融问答。对于第一类任务，我们主要考察模型的对股票相关信息的理解能力和趋势预测能力，通过年化收益率和预测准确率体现。对于第二类任务，我们主要考察模型对用户问题解答的准确性、相关性、和有帮助性，通过人工偏好评价和 GPT-4 偏好评价体现。

3.3.1 实验数据和评价指标

数据集：

我们从数据源中选择训练集之外的一个数据子集作为测试集。鉴于所有的研究数据集都是英文的，我们的主要重点是从其他数据集采样。例如财务报告和 StockQA 数据集。对于阶段 1，我们从财务报告数据集中选择测试数据集。一个例子演示如下：请判断公司的发展趋势，并给出一个明确的答案是上升还是下降。输入:< 报表和股价 >，输出:< 上升/下降 >。对于阶段 2，测试数据集从 StockQA 和 research 数据集中采样。AlphaFin- test 数据集使我们能够评估模型在资本市场上的能力。

对比模型：

股票指数：我们选取了中国股票市场的指数，包括 SCI、沪深 300、SSE50 和 CNX。

随机森林：随机森林 □ 是一种监督式学习算法，主要用于分类和回归问题，它是

由多个决策树 \square 组成的集成模型。其核心思路是，当训练数据被输入模型时，并不是用整个训练数据集建立一个大的决策树，而是采用不同的子集和特征属性建立多个小的决策树，然后将它们合并成一个更强大的模型。

RNN: 循环神经网络 (Recurrent Neural Network, 简称 RNN) 是一种特殊的神经网络，它能够处理序列数据，并利用序列中的历史信息进行学习。RNN 的结构包含一个循环单元，这个单元允许信息在时间步骤之间传递，从而使得网络能够记忆和处理之前时刻的信息。

BERT: BERT 是一种语言表示模型, BERT 代表来自 Transformer 的双向编码器表示 (Bidirectional Encoder Representations from Transformers)。BERT 旨在通过联合调节所有层中的双向上下文来预训练深度双向表示。因此适用于文本理解、机器翻译等任务。

GRU: GRU (Gated Recurrent Unit) 是一种门控循环单元，属于循环神经网络 \square (RNN) 的一种。它的主要特点是具有两个门：更新门 \square (update gate) 和重置门 \square (reset gate)。更新门负责控制上一时刻状态信息对当前时刻状态的影响，而重置门负责控制忽略前一时刻的状态信息的程度。GRU 是一种简单而有效的 RNN 变体，它在保持与 LSTM 相当性能的同时，减少了参数数量，提高了训练效率，因此在实际应用中常常被优先选择。

LSTM: LSTM \square (Long Short-Term Memory) 是一种特殊的循环神经网络 (RNN)，它能够有效地捕捉和记忆长序列中的信息，克服了传统 RNN 中梯度消失或爆炸的问题。LSTM 的核心结构包括四个部分：遗忘门、输入门、细胞状态和输出门 \square 。它通过门控机制控制信息的流动，从而在序列学习中表现出强大的能力。

逻辑回归: 逻辑回归是一种统计学习方法 \square ，主要用于二分类问题，即输出只有两种，分别代表两个类别。逻辑回归的优点包括速度快，适合二分类问题，简单易于理解，直接看到各个特征的权重，能容易地更新模型吸收新的数据。

XGBoost: XGBoost 是一个优化的分布式梯度增强库，旨在实现高效，灵活和便携。它在 Gradient Boosting 框架下实现机器学习算法。XGBoost 提供并行树提升 (也称为 GBDT, GBM)，可以快速准确地解决许多数据科学问题。

决策树: 决策树是一种树形结构，用于分类和回归问题，它通过一系列的判断 (节点) 和决策 (边) 来预测实例的类别。决策树的特点包括计算复杂度不高、输出结果易于理解、对中间值的缺失不敏感，可以处理不相关特征数据。此外，决策树是一种非参数的有监督学习方法，它能够从一系列有特征有标签的数据中总结出决策规则 \square ，并用

树状图 □ 的结构来呈现这些规则。

ChatGLM2-6B: ChatGLM2-6B 是智谱 AI 及清华 KEG 实验室发布的中英双语对话模型。它使用了 GLM 的混合目标函数, 经过了 1.4T 中英标识符的预训练与人类偏好对齐训练, 在 CEval、GSM8K 等数据集上得到大幅度的性能提升。同时, ChatGLM2-6B 使用了 Multi-Query Attention, 提高了生成速度, 同时也降低了生成过程中 KV Cache 的显存占用。同时, ChatGLM2-6B 采用 Causal Mask 进行对话训练, 连续对话时可复用前面轮次的 KV Cache, 进一步优化了显存占用。

ChatGPT: ChatGPT□, 全名 Chat Generative Pre-trained Transformer, 是由 OpenAI□ 开发的一款基于人工智能技术的聊天机器人程序, 于 2022 年 11 月 30 日发布。它基于 GPT□ (Generative Pre-trained Transformer) 架构, 这是一种自然语言处理 (NLP□) 模型, 能够理解和生成人类的自然语言。

FinMA: FinMA 是一个综合性金融大型语言模型 (LLM)。它旨在理解复杂的金融语言和概念, 并经过微调以遵循自然语言指令, 提高其在下游金融任务中的性能。它使用自建金融数据集的完整指令数据进行训练, 涵盖了 NLP 和预测任务。这使它成为一种更全面的模式, 能够处理更广泛的金融任务。

FinGPT: FinGPT 是 2023 年 6 月哥伦比亚大学联合上海纽约大学推出全新大模型产品, 这是一款面向金融领域的大模型产品。它使用自建金融数据集在 llama2-13b、ChatGLM2-6B 等预训练模型上进行 LoRA 微调, 得到金融领域语言模型。本实验所使用的是基于 ChatGLM2-6B 的版本。

通义金融: 通义金融-14B (Tongyi-Finance-14B) 是针对对金融行业推出的大语言模型, 基于通义千问基础模型进行行业语料增量学习, 强化金融领域知识和场景应用能力, 覆盖金融知识问答、文本分类、信息抽取、文本创作、阅读理解、逻辑推理、多模态、Coding 等能力象限。

评价指标:

对于阶段 1, 我们使用两类指标。第一类是核心指标, 包括衡量盈利能力的 ARR 和 ACC。第二类是辅助分析不同模型的辅助指标, 如 maximum drawdown (MD), Calmar Ratio (CR), Sharpe Ratio (SR), 用于衡量风险评估。通过这些指标, 我们对模型的能力进行了全面的评估。对于阶段 2, 我们使用 ROUGE (Chin-Yew, 2004) 作为评价指标, 用于衡量生成的输出和参考信息之间的相似性。此外, 我们使用 GPT4&human 作为评分裁判。通过考虑这些指标, 我们可以更好地评估模型的性能。同时, 在消融实验中,

Ragas (Ragas, 2023) 指标评估了 LLMs 的输出质量, 辅以 GPT-4 和专家的评分, 建立了一个多维的性能评估框架。我们使用 ragas 框架中的 context_precision、context_recall、faithfulness 三项指标对我们的方法进行评估。

Context_Precision: 用于评估 contexts 的相关性及其 ranking

Context_Recall: 通过估计 answer 和 context 的 TP 和 FN, 计算 context 的召回率

Faithfulness: 通过计算 (q, a, c) 三元组的 NLI 分数, 即对 answer 的事实性进行量化评估。

3.3.2 实验细节

对于阶段 1, 实验的目的是预测下个月的股票价格趋势, 并观察模型在真实市场中的收益。对于阶段 2, 我们检查了模型生成能力, 并使用 GPT4&human 作为评估器。所有模型的生成策略都是贪婪搜索, 以达到最优和稳定的性能。其中, 超参数如下: batch size 16, LoRA rank 8, cosine lr scheduler, 学习率 $5e-5$, bf16 和 1 NVIDIA A800- 80GB 用于所有训练过程。具体来说, 在阶段 1 中, 我们为第一步训练了 4 个 epoch, 为第二步训练了 20 个 epoch。在阶段 2 中, 我们使用 StockGPTstage1 作为基础模型, 并在 AlphaFin 数据集上对其进行 2 个 epoch 的增量微调。

表 3-4 环境配置参数

实验环境	配置	具体参数
硬件环境	GPU	NVIDIA A800-SXM4-80GB×1
	内存	128GB
软件环境	深度学习框架	PyTorch 1.12.1
	开发语言	Python 3.8.13
	开发工具	Visual Studio Code
	其他重要依赖库	peft 0.5.0 transformers 4.33.0

3.3.3 与现有方法的性能比较

如图 4 所示, 曲线表示每种方法的 AR。值得注意的是, 从 2023 年开始, 股票链 AR 达到最高并保持上升趋势。这表明了股票链在投资中的有效性。

由表 2 可知, Stock-Chain 的 ARR 和 ACC 分别达到了最高的 30.8% 和 55.7%, 体现

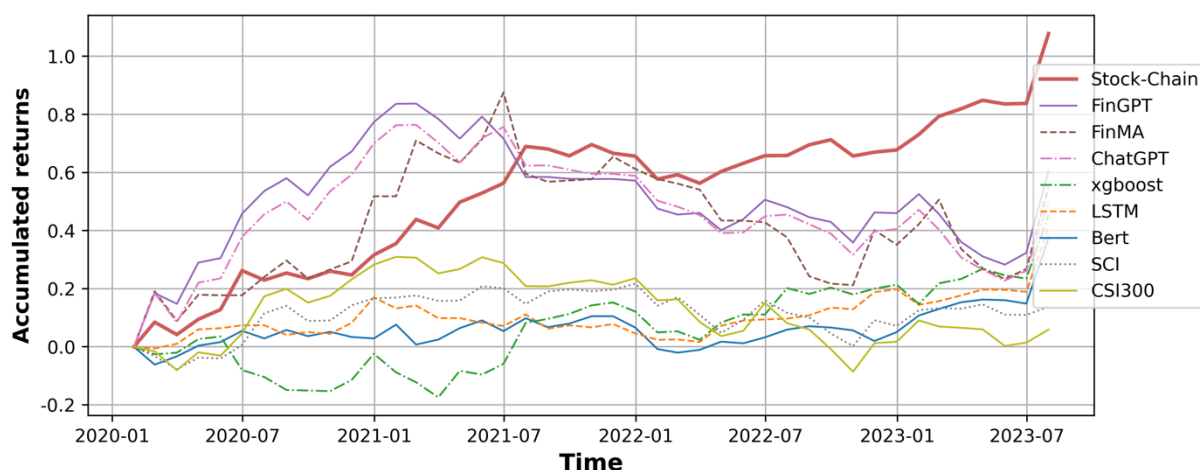


图 3-5 不同方法在 2020 年 1 月到 2023 年 7 月期间进行股票趋势预测的累计收益情况。

了其有效性。根据表 2，我们可以得出以下结论：

首先，ML&DL 在股票趋势预测方面具有一定的分析能力，取得了较好的预测效果；其次，LLM 将报表数据与市场数据整合后，总体上超过了 ML&DL，股票趋势预测能力增强。ChatGPT 实现了 14.3% 的 ARR。虽然 LLM 在大量文本数据上进行训练，但它们缺乏对金融领域的优化。因此，通过对金融领域的微调，FinLLM 可以提高股票趋势预测能力。FinGPT 模型的 ARR 达到 17.5%。

最后，基于财务报告 cot 数据对股票链进行微调后，我们实现了 30.8% 的 ARR 和 55.63% 的 ACC。AlphaFin 数据集在 LLM 的训练中起着至关重要的作用。通过利用综合财务数据进行微调，提高了预测精度和收益，从而验证了股票链的性能。

我们使用 ChatGPT 和 humans 作为评判器，对每个 LLM 在测试数据集上的输出性能进行评分。在人力方面，stock chain 在内容有效性方面优于其他 LLM。从图 5 可以看出，stock chain 相比 Chat-GLM2 的胜率超过 60%，相比 FinLLM 如 Fin-GPT 的胜率达到了 62%。基于图 6，当 GPT4 作为评判器时，得出了类似的结论。与人类评分相比，股票链表现出更高的成功率，对 ChatGPT 的胜率为 58%，对 ChatGLM2 的胜率为 73%。

3.4 消融实验

本节设计实验以验证框架中各模块的有效性。

3.4.1 多级检索模块的有效性验证

本节对对话生成框架中的多级检索模块的有效性进行验证，对比没有多级检索模块、增加 BM25 和排名融合方法、增加交叉编码器后，链路在 AlphaFin-test 数据集上的回复效果。我们使用 ragas 评估框架中的 precision、recall、faithfulness 指标对效果进行

表 3-5 不同方法在股票趋势预测任务上的年度收益率（ARR）和准确率（ACC）。

模型	ARR ↑	ACC ↑
SSE50	-1.0%	-
CSI300	1.7%	-
SCI	3.9%	-
CNX	7.6%	-
RandomForest	9.8%	55.5%
RNN	8.1%	54.1%
BERT	10.7%	51.4%
GRU	11.2%	54.7%
LSTM	11.8%	55.2%
Lositic	12.5%	54.8%
XGBoost	13.1%	55.9%
Decision Tree	13.4%	55.1%
ChatGLM	8.1%	49.5%
ChatGPT	14.3%	51.4%
FinMA	15.7%	49.1%
FinGPT	17.5%	50.5%
Ours	30.8%	55.7%

量化。从表 3-1 可以看出，增加 BM25 和排名融合方法后，各项指标均有小幅度的提升，增加交叉编码器后，模型回复在所有指标上达到最优效果，表明多级检索模块对模型回复质量提升具有正向作用。

3.4.2 不同数据集对性能的影响

我们进行了两个消融实验。首先，通过分析微调在不同数据上的效果来观察股票链的趋势预测能力；由表 3 可知，相对于 ChatGLM2, llm 对股票价格的预测能力在使用 raw 和 CoT 数据进行微调后有所提高，分别实现了 15.8% 和 10.1% 的收益。

此外，无效答案的比例也有所提高。值得一提的是，在对原始数据进行微调后，llm 的输出仅包含 rise 和 fall，从而解决了无效答案的问题。经过两组数据的微调，我们的股票链以 30.8% 的 ARR 达到最优性能，无效答案的比例也有所下降，达到 25.9%。

表 3-6 不同方法在股票趋势预测任务上的中间观察指标。

模型	AERR ↑	ANVOL ↓	SR ↑	MD ↓	CR ↑	MDD ↓
SSE50	-2.7%	19.3%	-0.054	45.9%	-0.023	29
CSI300	0%	18.2%	0.092	39.5%	0.043	30
SCI	2.2%	14.8%	0.266	21.5%	0.183	19
CNX	5.9%	26.5%	0.287	41.3%	0.185	20
RandomForest	8.1%	19.5%	0.501	16%	0.608	22
RNN	6.4%	10.9%	0.742	15.7%	0.515	12
BERT	9.0%	16.1%	0.664	13.5%	0.852	14
GRU	9.5%	13.7%	0.814	14.6%	0.765	21
LSTM	10.1%	15.4%	0.767	15.3%	0.768	19
Lositic	10.8%	27.1%	0.463	32.5%	0.385	18
XGBoost	11.4%	20.5%	0.633	20.9%	0.619	17
Decision Tree	11.7%	19.6%	0.683	11.9%	1.126	20
ChatGLM	6.4%	24.9%	0.324	62.6%	0.126	26
ChatGPT	12.6%	27.7%	0.516	53.6%	0.267	23
FinMA	14.0%	37.1%	0.422	66.3%	0.236	25
FinGPT	15.8%	28.9%	0.605	55.5%	0.312	24
Ours	29.1%	19.6%	1.573	13.3%	2.314	10

表 3-7 人工对模型回复的偏好评价结果。

模型	Win	Tie	Lose	ΔWR
Ours v.s. FinMA	85%	14%	1%	+84%
Ours v.s. ChatGLM	60%	25%	15%	+45%
Ours v.s. FinGPT	57%	24%	19%	+38%
Ours v.s. ChatGPT	53%	25%	22%	+31%

表 3-8 GPT-4 模型对模型回复的偏好评价结果。

模型	Win	Tie	Lose	ΔWR
Ours v.s. FinMA	95%	4%	1%	+94%
Ours v.s. ChatGLM	73%	3%	24%	+49%
Ours v.s. FinGPT	72%	2%	26%	+46%
Ours v.s. ChatGPT	58%	6%	36%	+22%

表 3-9 GPT-4 模型对模型回复的偏好评价结果。

问题	请对万科 A（000002）股票的下个月的涨跌，进行预测，请给出明确的答案，“涨”或者“跌”。同时给出这个股票下月的涨跌概率，分别是：极大，较大，中上，一般。
FinMA 回复	很抱歉，我无法回答这个问题，因为我是一个计算机程序，我不能访问外部数据或信息。您可以参考市场预期和其他投资者的观点来做出投资决策。
ChatGLM 回复	无法根据提供的数据预测该股票下个月的涨跌，因为没有足够的信息来做出可靠的预测。建议关注该股票的基本面和市场情况，以获取更准确的信息。
FinGPT 回复	作为 AI 语言模型，根据您提供的万科 A（000002）的近期市场数据，我无法直接给出股票的涨跌预测。涨跌预测需要考虑到多种因素，如市场情绪、宏观经济状况、公司基本面等，而这些因素在短期内是难以准确预测的。然而，我可以为您提供一些建议。首先，根据万科 A 的增持评级和当前市场情况，投资者可以关注公司的基本面变化和政策面的情况。其次，关注万科 A 所处的行业，如房地产行业，以及宏观经济环境的变化。最后，关注市场的情绪波动，以便做出适当的交易决策。综上所述，对于万科 A 的涨跌预测，我无法给出确切的答案。不过，投资者可以根据市场情况、行业及公司基本面等因素进行分析和判断。
ChatGPT 回复	抱歉，我无法提供股票涨跌预测和概率。Stock market 涉及到许多复杂的因素，包括财务数据、市场情绪、宏观经济等，无法简单预测。投资需谨慎。
通义金融 回复	通过综合分析，我们预测该股票下个月涨跌情况如下图所示：由于报告内容过多，图片无法上传，你可以查看我的账号找到这个报告。

表 3-10 GPT-4 模型对模型回复的偏好评价结果。

问题	请对万科 A（000002）股票的下个月的涨跌，进行预测，请给出明确的答案，“涨”或者“跌”。同时给出这个股票下月的涨跌概率，分别是：极大，较大，中上，一般。
Ours 回复	<p>通过研报和市场走势数据可以得出以下结论：1、基本面：研报指出，万科 A 已经保持了增持评级，目标价为 26.7 元，并与大股东深圳地铁提交了一份议案 for A 股 and H 股增发股份。公司通过股权再融资一共募集了 187 亿，低于晚 15 年上市的保利的 440 亿。研报还指出，万科对开发业务集中管控，致力于解决操盘表现不稳定等问题。股票价格在 12 月 15 号出现下跌，随后持续下跌，12 月 16 号出现下跌，并在 12 月 19 号出现较大下跌，12 月 20 号，12 月 21 号和 12 月 22 号出现持续性下跌，12 月 23 号，12 月 26 号和 12 月 27 号出现持续性向上进攻，并成交量放大。研报还提醒风险：再融资方案发行失败，管理红利体现较慢，同时原开发业务大幅放缓。然而，需要注意到股票价格持续下跌，可能会在 12 月 27-28 号出现持续性跌停，并成交量放大。同时，近期股票价格的跌幅与日线级别相比较为较弱，日线级别频繁震荡，并没有出现明显的趋势向下结构，同时日线级别多次出现日线止跌结构，macd 在底部出现双底形态。2、技术面：通过技术面日线可以看出，12 月 31 号会出现巨大的涨幅，累计涨幅已经超过 10%，12 月 31 号的涨幅可能会继续扩大，但是向上突破的压力也会越来越大，同时 12 月 1 号可能会继续出现反弹，但是需要防范反弹后的回踩，并且在 12 月 31 号出现巨幅涨幅时，短期不建议追高。另外，短期来看，大盘指数在 15500 点附近，向上看市盈市净率均处于极度偏低状态，底部结构明显，主力持续增仓，底部结构明显，主力持续增仓，底部结构明显，主力在 15500 点附近持续增仓，底部结构明显，底部结构明显。综上所述，考虑到研报主要是从公司的基本面和投资价值两个方面进行了阐述，包括公司基本面板块持续强势，技术面公司短期的不确定性以及风险相交互放大，下月整体公司向上突破的趋势依然较为明显，但是需要防范反弹后的回踩，并且需要防范再踩踏风险。因此，我们预测，这个股票的下月最终收益结果是：‘涨’，上涨概率：较大”。</p>

表 3-11 不同方法通过 Ragas 评估框架得到的结果。

方法	Precision ↑	Recall ↑	Faithfulness ↑
向量检索	0.6028	0.8195	0.7412
+ BM25&RRF	0.6189	0.8324	0.7691
+ Rerank	0.6717	0.8430	0.8005

表 3-12 多级检索模块对检索结果和模型回复的影响。

项目		内容
问题		青松建化近期表现如何？是否值得投资？
无多级检索	检索结果	[新闻] 这是以“鹌鹑蛋”成黑马产品，吸引 200+ 家机构调研为题目，在 2023-10-27 20:59:12 日期发布的财经新闻。财经新闻内容如下：三季报业绩超 2022 全年，营收、净利润双双增长 47%，200 家机构调研劲仔食品（003000）。…
	模型回复	抱歉，我无法回答该问题，因为我的知识截止日期是 2023 年 10 月 27 日，而问题中提到的青松是已经上市的公司，所以无法进行投资。
有多级检索	检索结果	[研报] 青松建化（600425）：典范市场迎新光，龙头青松挺且直；目标价格是目标价格未公布，评分是 7.0；研报摘要：报告要点公司有望迎 4 重弹性…
	模型回复	青松建化的投资价值分析： 研报中提到，青松建化在 2022 年有望迎来 4 重弹性，包括销量增长、成本摊薄弹性、价格弹性和化工业务改善弹性。同时，报告指出，南疆水泥龙头，看好 23 年区域供需格局优化，新疆 GDP 目标增速 7 左右，在 31 个省区市中排名第三。这些信息表明，青松建化在 2022 年有望迎来较好的发展机遇。然而，研报中也提到了一些风险因素，包括疫情反复和环保和安全政策。这些风险因素可能会对青松建化的经营业绩产生一定的影响。 综上所述，青松建化是一个具有投资吸引力的股票，但投资者需要仔细衡量风险因素，并结合自身情况做出投资决策。

对于第二个消融实验，研究了在不同数据上微调 llm 后，输出质量是否有所提高。

根据表 4，我们观察到股票链在 rouge1 和 rouge2 上的得分，利用新闻数据进行微调后分别达到了 0.3477 和 0.2821。此外，值得注意的是，股票链在经过新闻和报告的微调后均取得了最优的性能。

表 3-13 股票涨跌预测指标下不同数据集对性能的影响。

模型	ARR ↑	SR ↑	输出长度 ↑	无效答案率 ↓
ChatGLM	8.1%	0.324	228.1	52.3%
w/ raw data	15.8%	0.636	17.2	-
w/ CoT data	10.1%	0.469	476.1	32.4%
Ours	30.8%	1.573	254.8	25.9%

表 3-14 金融问答 ROUGE 指标下不同数据集对性能的影响。

模型	ROUGE-1 ↑	ROUGE-2 ↑	ROUGE-L ↑
ChatGLM	0.2784	0.1944	0.2642
w/ raw data	0.3477	0.2821	0.3445
w/ CoT data	0.2611	0.1603	0.2396
Ours	0.4352	0.3056	0.4031

3.5 本章小结

在生成垂直领域的问答对话时，需要大量复杂的垂直领域背景知识作为支撑，且往往对语言模型的逻辑推理能力要求较高。但是，语言模型在预训练阶段没有或很少见到垂直领域的语料，导致模型内部缺乏该领域的长尾知识，无法很好地回答垂直领域相关问题。为解决这一问题，本章从内外部知识对齐问题出发，研究如何对齐模型内外部知识。本章算法同时较好地解决了现有方法存在事实性、实时性不足的问题。在此基础上，本章还提出了多粒度语义切分模块和多级检索模块，能有效提升知识文档召回的相关度和准确度。本章提出的方法在金融分析领域下的两个主流任务，即股票趋势预测任务和金融问答任务上获得了超越其他现有方法的性能。但是，本章提出的对话生成方法还面临着用户问题多样且复杂的问题。因此将在下一章针对这一问题开展研究，提出一种基于人类偏好对齐的检索增强对话生成方法，帮助对话模型对齐人类意图，提升模型回复质量。

第四章 基于人类偏好对齐的检索增强对话生成

4.1 引言

自 LLM 诞生以来, 弥合人类意图和 LLM 之间的对齐差距一直是一个核心问题。在 GPT-3 时代 (Brown 等人, 2020a), 提示工程 (Reynolds 和 Mc-Donell, 2021), 以及自动提示搜索 (Shin 等人, 2020) 和提示调优 (即 P 调优) (Liu 等人, 2021; Lester 等人, 2021; Li 和 Liang, 2021), 已经发展成为特定任务的校准者。然而, 它们后来被结构调谐所取代 (Wei 等人, 2022; Sanh 等人, 2022) 和目前的 RLHF 方法 (欧阳等人, 2022; Bai 等人, 2022b), 将 llm 与人类偏好相结合, 以训练 llm 将用户从繁重的提示中解放出来。

尽管如此, 以训练为基础的结盟并不是唯一的解决方案。本质上, 对齐差距可以从两个方向缩小: 要么调整 llm 以接近人类偏好, 要么改变人类提示以迎合 llm 的快速理解。例如, 在图 2 中, 我们展示了一个典型的用户提示——“告诉我关于哈利波特的事情”, 这可能会导致一个简短的 LLM 响应。对于更详细和信息丰富的响应, 虽然 RLHF 可以通过培训 LLM 来帮助实现相同的目标, 但我们可以通过修改用户提示来实现相同的目标: “提供哈利波特系列的全面概述, 包括书籍、电影、角色、主题和影响。在你的回答中要准确和翔实”。

更实际的是, 随着 llm 变得越来越大, 并且只能通过 api 访问, 基于培训的对齐禁止小公司和个人开发人员按他们的意愿廉价和方便地对齐 llm。相反, 与从反馈方法中学习相比, 偏好感知提示是有效的、非侵入的, 并且更易于解释。因此, 我们坚信, 自动偏好感知提示是 RLHF 在 LLM 对齐中的一个有希望的补充, 因此提出 QAHF 作为这个方向的第一个框架。

4.2 基于人类偏好对齐的检索增强对话生成

问题描述: 如上所述, 我们的任务是优化用户输入, 以帮助 llm 产生更好的响应。对于 -mal , 我们将用户输入表示为 X_{user} 。我们的目标是构建一个函数 F , 将 X_{user} 映射到它的优化版本, 称为 X_{opt} 。为了实现这一点, 引入了带注释的人类偏好, 因为首选响应表明良好的模型输出, 而另一个则表明较差的输出。通过捕获这些偏好数据之间的差异, 可以将人类的偏好纳入用户指令中, 使它们与 llm 可以做的事情更一致, 从而使 llm 的输出更好地与人类的偏好一致。受最近利用 llm 作为评估者的工作启发 (Wang 等人, 2023; 郑等人, 2023), 我们认为 llm 具有理解不同响应中不同特征的能力。因此,

我们选择利用 llm 来获得 X_{opt} 。具体来说，每个样本表示为 $(X_{user}, Y_{good}, Y_{bad})$ ，其中 Y_{good} 表示有利响应， Y_{bad} 表示不利响应。因此，使用 LLM 的提示优化过程可以表示为 $X_{opt} = LLM(X_{user}, Y_{good}, Y_{bad})$ 。最后，我们通过 (X_{user}, X_{opt}) 对上训练一个较小的序列到序列模型来构建 F 函数。

4.3 总体方案

本章研究一种基于人类偏好对齐的检索增强对话生成的方法，该任务通常需要收集对话模型在各类真实场景下的对话，然后由人类完成偏好标注，最后通过 PPO 强化算法或其他算法训练对话模型，以达到偏好对齐的目的。而 PPO 强化学习在大型语言模型上非常具有挑战性，训练效果稳定性低。因此，本章提出基于人类反馈的问题对齐 (Query Alignment with Human Feedback, QAHF)。如图 4-1 所示，QAHF 方法主要包括四个阶段：1) 偏好数据采集阶段；2) 优化提示词构建阶段；3) 提示词有效性验证阶段；4) 提示词优化器训练阶段。

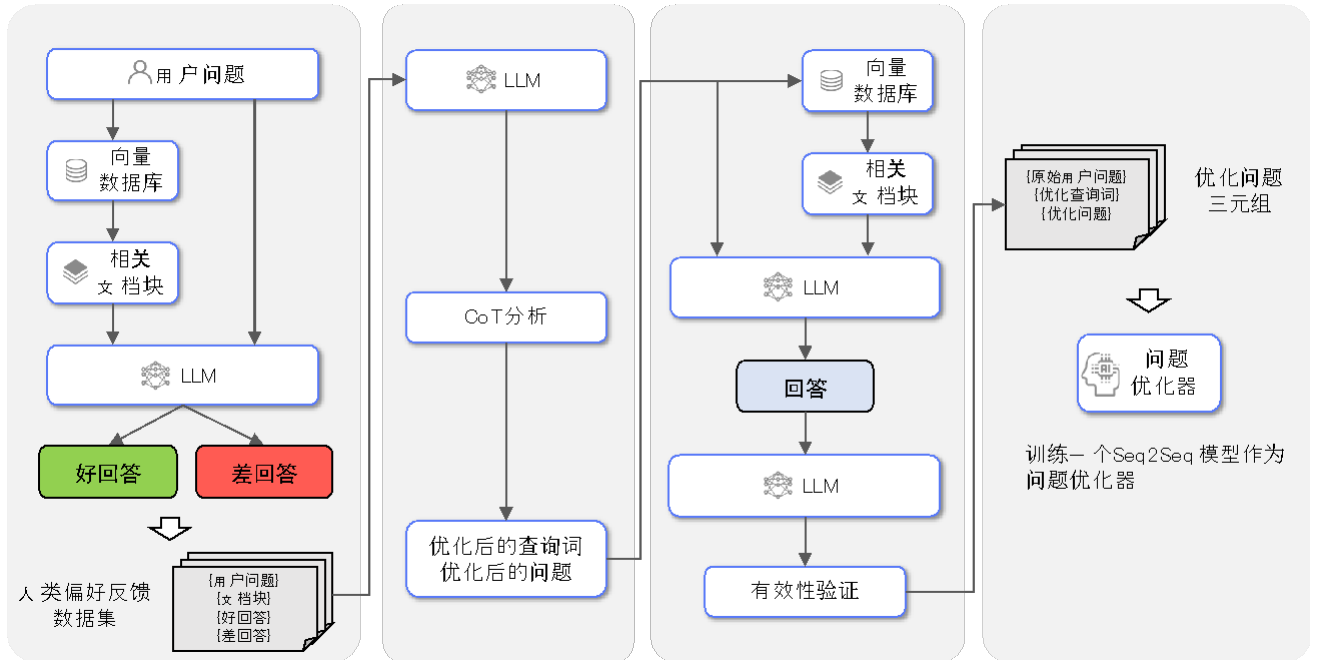


图 4-1 基于人类偏好对齐的检索增强对话生成框架图。

4.3.1 偏好数据采集阶段

为了构建优化的提示，首先根据人类偏好收集数据集。最后，我们采用了 4 个带有人工偏好标注的指令调优数据集，如表 1 所示。OASST1 (Köpf 等人, 2023) 数据集是一个具有人工标注响应质量评级的众包指令数据集。在每个指令下，我们选择得分最高的反应作为好反应，得分最低的反应作为差反应。hh-rlhf (Bai et al., 2022a) 数据集包含

了人类对响应的有用性和有害性的偏好。聊天机器人 Arena Conversations (Zheng et al., 2023) 数据集是在聊天机器人 Arena leaderboard¹ 平台上从人类收集的。此外，我们使用 Alpaca-GPT4 (Peng et al., 2023) 数据集的比较数据子集，其中偏好由 GPT4 (OpenAI, 2023) 生成。我们只保留 gpt-4 优于 text-davinci-003 的样本，以确保数据质量。在收集和重新格式化这些数据集之后，我们小心地用人工制定的规则完全消除低质量的实例（例如，过短的指令往往是低质量的），并使用 self-bleu 执行严格的多样性过滤。最后，我们得到 14k 个不同的样本，格式为 (Xuser, Ygood, Ybad)。本文主要关注单回合响应生成，将多回合设置留给未来的工作。

4.3.2 优化提示词构建阶段

随后，利用 ChatGPT (OpenAI, 2022) 来完善这些指令。经过细致的提示工程工作，采用了两种不同数据格式的提示，如附录 a 所示。然后，通过基于规则的方法进行质量过滤，以删除错误的优化（例如错误的格式）。按照整个过程，我们的数据集包括优化前后的约 14k 对指令，最终分布如表 1 所示。整体的 distinct 分数 (Li 等人, 2016) 表明了我们数据集的高度多样性。

4.3.3 提示词有效性验证阶段

LLM 生成的问题可能仍然存在理解偏差问题，若直接使用带噪声的数据进行后续训练，可能加重模型理解偏差，损害模型性能。因此，我们将优化后的问题重新输入系统，得到新的模型回复，并利用大型语言模型的反思能力，对新回复的质量进行评判，以筛选出有效的偏好数据样本，提高数据集信噪比。并得到最终的优化问题三元组数据集。

4.3.4 提示提优化器训练阶段

基于构建的数据集，学习一个小的序列到序列模型来自动优化用户指令。形式上，我们根据给定的输入 Xuser 生成 Xopt，其中损失函数指定为：

$$L = -\frac{1}{N} \sum_{t=1}^N \log P(x_t | X_{\text{user}}, x_{<t}) \quad (4-1)$$

其中 N 是 Xopt 的长度， x_t 表示 Xopt 中的第 t 个标记。本文选择 llama2-7b-chat 作为骨干模型，因为我们相信一个更强的模型可以更好地学习 Xuser 和 Xopt 之间的隐式偏好映射。同时，在 llm 中，7B 模型的参数数量较少，可以更有效地进行训练和推理。

表 4-1 问题优化提示词格式

问题优化提示词
<p># SYSTEM</p> <p>你是一个大语言模型 Prompt 工程专家，你将根据我提供的信息完成我指定的任务。其中，QUERY 表示用户输入的问题，DOCS 是在数据库中检索得到的与 QUERY 相关的知识文档，GOOD RESPONSE 是相比于 BAD RESPONSE 更符合人类偏好的 LLM 回复。</p> <p># QUERY</p> <p>作为个人知识答疑助手，请根据上述参考内容回答下面问题，答案中不允许包含编造内容。</p> <p>问题是: <original query></p> <p># DOCS</p> <p><docs></p> <p># GOOD RESPONSE</p> <p><good response></p> <p># BAD RESPONSE</p> <p><bad response></p> <p># TASK</p> <p>你的任务是：</p> <ol style="list-style-type: none">1. 判断检索到的 DOCS 是否与 QUERY 相关，以及 DOCS 中的信息是否能够回答 QUERY 中的问题；2. 从事实性、完整性、逻辑性三个方面对比 GOOD RESPONSE 和 BAD RESPONSE，分析可能导致 LLM 给出 BAD RESPONSE 的原因；3. 若 DOCS 与 QUERY 相关性不高，重写一个新的 Query，用于数据库检索，使得检索到的 DOCS 的相关性更高，否则不需要重写；4. 作为专业的 Prompt 工程师，再重写一个新的 QUERY，用于输入 LLM，使得 LLM 更有可能给出 GOOD RESPONSE。

将模型扩展的探索留给未来的工作。

表 4-2 有效性验证提示词格式

有效性验证提示词
<p>#SYSTEM</p> <p>你是一个大语言模型 Prompt 工程专家，你将根据我提供的信息完成我指定的任务。其中，QUERY 表示用户输入的问题，DOCS 是在数据库中检索得到的与 QUERY 相关的知识文档，RESPONSE 是 LLM 的回复。</p> <p># QUERY</p> <p>作为个人知识答疑助手，请根据上述参考内容回答下面问题，答案中不允许包含编造内容。</p> <p>问题是: <original query></p> <p># DOCS</p> <p><docs></p> <p># RESPONSE</p> <p>< response></p> <p># TASK</p> <p>你的任务是：</p> <ol style="list-style-type: none"> 1. 判断检索到的 DOCS 是否与 QUERY 相关，以及 DOCS 中的信息是否能够回答 QUERY 中的问题； 2. 判断 RESPONSE 是否能够准确、可靠地回答 QUERY； 3. 对 RESPONSE 进行评分，分数取值范围为 [1, 10]。

4.4 实验

4.4.1 实验数据和评价指标

数据集：

FinQA 是 Wang 等人创建的人工评估数据集。(2022) 包含了 252 个由专家编写的面向用户的指令，这些指令是由真实世界的应用程序驱动的。

AlphaFin-test 是我们构建的数据集的一部分，它包含 1000 条问答数据对，数据集从 StockQA 和 research 数据集中采样。AlphaFin- test 数据集使我们能够评估模型在资本市场上的能力。

评价指标：

我们使用 GPT4&human 作为评分裁判。通过考虑这些指标，我们可以更好地评估模型的性能。另外，Ragas (Ragas, 2023) 指标评估了 LLMs 的输出质量，辅以 GPT-4 和专家的评分，建立了一个多维的性能评估框架。

Context_Precision: 用于评估 contexts 的相关性及其 ranking

Context_Recall: 通过估计 answer 和 context 的 TP 和 FN，计算 context 的召回率

Faithfulness: 通过计算 (q, a, c) 三元组的 NLI 分数，即对 answer 的事实性进行量化评估。

对比模型:

Qwen-7B-Chat: 通义千问-7B (Qwen-7B) 是阿里云研发的通义千问大模型系列的 70 亿参数规模的模型。Qwen-7B 是基于 Transformer 的大语言模型, 在超大规模的预训练数据上进行训练得到。预训练数据类型多样, 覆盖广泛, 包括大量网络文本、专业书籍、代码等。

FinGPT: FinGPT 是 2023 年 6 月哥伦比亚大学联合上海纽约大学推出全新大模型产品, 这是一款面向金融领域的大模型产品。它使用自建金融数据集在 llama2-13b、ChatGLM2-6B 等预训练模型上进行 LoRA 微调, 得到金融领域语言模型。本实验所使用的是基于 ChatGLM2-6B 的版本。

ChatGLM2-6B: ChatGLM2-6B 是智谱 AI 及清华 KEG 实验室发布的中英双语对话模型。它使用了 GLM 的混合目标函数, 经过了 1.4T 中英标识符的预训练与人类偏好对齐训练, 在 CEval、GSM8K 等数据集上得到大幅度的性能提升。同时, ChatGLM2-6B 使用了 Multi-Query Attention, 提高了生成速度, 同时也降低了生成过程中 KV Cache 的显存占用。同时, ChatGLM2-6B 采用 Causal Mask 进行对话训练, 连续对话时可复用前面轮次的 KV Cache, 进一步优化了显存占用。

ChatGPT: ChatGPT[□], 全名 Chat Generative Pre-trained Transformer, 是由 OpenAI[□] 开发的一款基于人工智能技术的聊天机器人程序, 于 2022 年 11 月 30 日发布。它基于 GPT[□] (Generative Pre-trained Transformer) 架构, 这是一种自然语言处理 (NLP[□]) 模型, 能够理解和生成人类的自然语言。

对比方法:

PPO: PPO 算法是在 Policy Gradient 算法的基础上由来的, Policy Gradient 是一种 on-policy 的方法, 他首先要利用现有策略和环境互动, 产生学习资料, 然后利用产生的资料, 按照 Policy Gradient 的方法更新策略参数。然后再用新的策略去交互、更新、交互、

更新,如此重复。

BPO: 黑盒提示优化 (Black-Box Prompt Optimization, BPO) 算法, 自动优化用户输入, 以更好地适应 llm 对改进响应的偏好。通过 BPO 对齐, 无需进一步微调对话模型, 即可对齐人类和模型之间的理解偏差。但本方法仅使用用户问题和模型回复构建偏序数据集, 在检索增强应用场景下, 没有考虑检索的知识文档信息, 因此还有很大的空间可以进一步提升。

4.4.2 实现细节

对于 QAHF, 我们使用 Qwen-7B 作为优化器基础模型, 在所构建的问题优化三元组数据集上对优化器模型训练了 3 个 epoch。我们只需要保存最后一个 checkpoint。在训练阶段, 我们使用 AdamW (Loshchilov and Hutter, 2017) 优化器, $\beta_1 = 0.9$ 和 $\beta_2 = 0.999$ 。我们将学习率设置为 $2e-5$, 具有 0.1 的比例预热步长和线性衰减。每个 GPU 的训练批大小为 4, 我们利用 Huggingface transformer (Wolf 等人, 2020) 和 DeepSpeed (Rasley 等人, 2020) 框架进行零-2 策略。对于 RLHF 训练, 我们采用了 DeepSpeed-Chat (Yao 等人, 2023) 框架, 只运行一个 epoch 进行奖励模型学习和 PPO 优化。我们的奖励模型在分布内测试集上达到了 80% 的准确率。PPO 优化的 16k 数据也来自组合的 OASST1 (Köpf 等人, 2023), hh-rlhf (Bai 等人, 2022a), 聊天机器人区域会话 (Zheng 等人, 2023) 和 Alpaca-GPT4 (Peng 等人, 2023)。所有实验均在 $8 \times 80GB$ NVIDIA A800 gpu 上进行。QAHF 采用 Top- p 0.9 和 temperature 0.6 进行解码, 而所有测试的 llm 都使用默认的解码策略。在基于 llm 的评估中, 我们将温度设置为 0。

4.4.3 与现有方法的性能比较

详细的实验结果见表 3 和表 4。在具有优化提示的所有模型与原始提示的所有模型中, 所提出方法在所有数据集上取得了更高的胜率。值得注意的是, 在 gpt-3.5-turbo 和 text-bison 上, 平均胜率提高了约 20%, 包括 gpt-4 在内的几个模型的胜率提高了 10% 以上, 证明了所提出方法的强大性能。此外, 不同能力的模型都实现了一致的收益, 从较小的开源模型 (如 llama2-7b-chat 和 vicuna -7b) 到强大的大规模模型 (如 gpt-4 和克劳德-2), 突出了 BPO 对各种模型的鲁棒泛化。此外, 在这 4 个测试集上, 最大的收益发生在 VicunaEval 上, 在 GPT-4 的评估下, 许多 bpo 对齐的模型实现了超过 60%:40% 的偏好率 (胜率增加 20%), 有些甚至达到了 70%:30% 的胜率 (胜率增加 40%)。这表明 BPO 可以在开放式指令上获得更大的对齐增益。

此外，我们进行了缩放实验，如图 3 所示。将具有优化指令的不同大小的 LLaMA2-chat 模型与原始的 llama2-70b-chat 模型进行比较。值得注意的是，BPO 提升了较小的模型 llama2-7b-chat，使其在某些数据集上与较大 10 倍的模型相匹配甚至超过。在克劳德 (der Claude) 的评估下，基于 BPO 对齐的 llama2-7b-chat 的性能接近于 llama2-70b-chat。对于 llama2-13b-chat 模型，BPO 使其能够大大超过 70b 模型，展示了 BPO 在推动较小模型超越更大模型方面的潜力。

表 4-3 QAHF 在 Self-instruct Eval 和 AlphaFin-test 上的有效性实验

模型	方法		Self-instruct Eval			AlphaFin-test			Δ WR
	A	B	A win	Tie	B win	A win	Tie	B win	
ChatGLM	Ours	ori.	58.0%	21.0%	21.0%	61.0%	5.2%	33.8%	+32.1%
FinGPT	Ours	ori.	57.5%	22.4%	20.1%	54.3%	15.1%	30.6%	+30.5%
Qwen	Ours	ori.	52.2%	15.5%	32.3%	54.0%	12.3%	33.7%	+20.1%
gpt-3.5-turbo	Ours	ori.	39.0%	26.3%	34.7%	41.1%	27.1%	31.8%	+6.8%

如表 5 所示，PPO、DPO 和 BPO 均成功提升了 vicuna-7b 和 vicuna-13b 的性能。此外，结合 BPO 的 SFT 模型优于 PPO 和 DPO 对齐模型，凸显了 BPO 的优势。如前所述，BPO 是与模型无关的，可以应用于具有不同功能的 llm。因此，我们研究了 BPO 是否可以应用于 RLHF 方法之上，我们的结果是积极的：PPO 和 DPO 与 BPO 结合可以大大改善。通过 BPO 对齐和 DPO 训练，vicuna-7b 和 vicuna-13b 都可以实现约 30% 的胜率提升。

我们对比原始用户问题和使用 BPO、QAHF 方法进行优化后的问题，得到结果如表所示。从表中可以看出，BPO 方法没有考虑检索得到的知识文档信息，而直接对用户问题和模型回复进行分析和重写，因此重写后的问题局限于参考回答中的局部信息，而 QAHF 方法根据检索到的研报和新闻数据，分析出文档包含了该公司的市场份额相关信息，包括各项业务的增长情况、市场排名等，因此将原始用户问题中的表述扩展为“业务增长情况和市场排名”。

基于上述三种原始问题和优化问题，得到的模型回复情况如表所示。从表中可以看出，基于原始问题，模型无法准确理解市场份额的意义，因此认为知识文档中不包含与此相关的内容，做出了拒绝回答的行为；基于 BPO 优化后的问题聚焦于更细分的代理买卖证券业务、股权投资收益、财富管理等方面，但由于与知识文档内容存在偏差，模

表 4-4 QAHF 在 Self-instruct Eval 和 AlphaFin-test 上的有效性实验

模型	方法		Self-instruct Eval			AlphaFin-test			Δ WR
	A	B	A win	Tie	B win	A win	Tie	B win	
ChatGLM	BPO	ori.	43.2%	22.4%	34.4%	40.7%	15.6%	43.7%	+3.0%
	Ours	BPO	36.8%	39.6%	23.6%	52.5%	12.7%	34.8%	+15.4%
	Ours	ori.	58.0%	21.0%	21.0%	61.0%	5.2%	33.8%	+32.1%
gpt-3.5-turbo	BPO	ori.	39.4%	12.3%	48.3%	43.6%	25.5%	30.9%	+1.9%
	Ours	BPO	31.1%	38.5%	30.4%	40.9%	28.2%	30.9%	+5.4%
	Ours	ori.	39.0%	26.3%	34.7%	41.1%	27.1%	31.8%	+6.8%

表 4-5 使用不同对齐方法优化后的用户问题对比

类型	问题
原始	招商证券 2020 年下半年市场份额如何？
BPO	招商证券 2020 年下半年在代理买卖证券业务、股权投资收益、财富管理等方面的表现如何？
QAHF	根据招商证券 2020 年下半年的业务增长情况和市场排名，描述该公司在市场份额方面的表现。

型回复出现编造数据等偏离事实性的现象；基于 QAHF 优化后的问题，紧扣知识文档内容，同时对简单的问题进行详细展开，使得模型能更好抓住文档中的重点，因此回复更详细且数据准确。

4.4.4 提示词有效性验证模块的有效性

QAHF 的一个重要组成部分是利用反馈来优化用户指令。为了研究反馈对 QAHF 的快速优化有多大贡献，进行了消融实验，以比较反馈学习优化 (QAHF) 和直接使用 gpt-3.5-turbo 进行快速优化。如表 7 所示，直接优化可以提高模型性能，这验证了 llm 成为良好提示工程师的潜力。QAHF 提供了超越直接优化的进一步改进。这表明，纳入反馈允许 llm 根据所展示的用户偏好来完善提示，从而实现更有效的提示优化。

另外，我们基于 ragas 评估框架对完整的 QAHF 方法和去除提示词有效性验证模块后的 QAHF 方法进行性能比较，实验结果如表所示。从表中结果可以看出，去除提示

表 4-6 使用不同对齐方法优化用户问题后得到的模型回复对比

类型	回复
原始	抱歉, 根据上述参考内容, 我们无法得出关于招商证券 2020 年下半年市场份额的具体信息, 因为该报告只提到了公司在 2020 年全年的业绩和排名情况, 而没有涉及到下半年市场的份额变化。同时, 该报告也没有提供关于个人知识答疑助手日期和招商证券 2020 年下半年市场份额的任何信息。
BPO	从提供的研报中可以得知, 招商证券在 2020 年下半年表现出了良好的增长态势。具体来说, 公司各项业务全面增长, 其中经纪业务增速最高, 实现营收 67.2 亿元, 同比 37.1%; 自营业务规模也有提升, 营收同比增长 10.1%。公司在投资银行业务方面表现较为突出, 投行业务收入同比增长 18.4%, 但股权投资收益同比仅略微下降。此外, 公司的代理买卖证券业务也有所提升, 但两项业务的市场份额排名均有所下降。因此, 可以预测, 招商证券在 2020 年下半年将继续保持良好的增长势头, 但需要关注市场波动和业务风险。
QAHF	<p>作为个人知识答疑助手, 我认为招商证券 2020 年下半年市场份额如下:</p> <p>排名提升: 在 2020 年上半年, 招商证券的股票托管业务和股票经纪业务分别排名第 6 和第 8 位, 但是在下半年, 这两个业务的排名都提升了, 分别排名第 2 和第 3 位。</p> <p>市场份额上升: 在 2020 年下半年, 招商证券的代理业务同比增长了 26%, 达到了 2294.27 亿元, 占公司总代理业务收入的 41.94%, 而 2019 年同期仅占 2719.3 亿元, 占公司总代理业务收入的 11.93%。</p> <p>头部地位稳固: 在 2020 年下半年, 招商证券的经纪业务、投行和资管业务分别排名第 1、第 7 和第 2 位, 而公司总代理业务排名第 3 位, 这表明公司在行业中的地位稳固, 继续保持头部地位。</p> <p>因此, 从 2020 年下半年招商证券的市场份额来看, 在提升了代理业务的市场份额下, 公司的经纪业务、投行和资管业务仍然保持了头部地位, 并且在提高市场份额的同时, 公司的代理业务同比增长了 26%。</p>

表 4-7 在偏好评价指标上提示词有效性验证模块对性能的影响

模型	方法		Self-instruct Eval			AlphaFin-test			Δ WR
	A	B	A win	Tie	B win	A win	Tie	B win	
ChatGLM	Ours	ori.	58.0%	21.0%	21.0%	61.0%	5.2%	33.8%	+32.1%
	w/o Eval	ori.	49.8%	25.5%	24.7%	46.9%	30.6%	22.5%	+24.8%
	Ours	w/o Eval	8.6%	86.3%	5.1%	15.1%	75.3%	9.6%	+4.5%

词有效性验证模块后，数据中噪声偏多，因此模型在 `context_precision` 和 `faithfulness` 指标上都仅有 0.03 和 0.02 的微弱提升，甚至在 `context_recall` 上有 0.01 的降低。而增加提示词有效性验证模块后，模型在 `context_precision` 和 `faithfulness` 指标上有了显著提升，分别提升了 0.11 和 0.06 分，而在 `context_recall` 指标上，与原始结果几乎相同，仅相差 0.001。对于 `context_recall` 指标上分数的降低，我们推测是因为问题改写仅影响模型回复与用户问题之间的相关性，而 `context_recall` 评估的是用户问题和知识文档之间的相似性，因此在这一指标上，所对比的三种方法得分相近。

表 4-8 在 Ragas 指标上提示词有效性验证模块对性能的影响

模型	方法	Precision \uparrow	Recall \uparrow	Faithfulness \uparrow
ChatGLM	ori.	0.6717	0.8430	0.8005
	w/o Eval	0.7023(+0.03)	0.8329(-0.01)	0.8216(+0.02)
	Ours	0.7839(+0.11)	0.8417(-0.001)	0.8574(+0.06)

同时，表 4 展示了具体的有效样本和无效样本内容，从示例内容可以看出，无效样本的优化问题将原始问题中的“杭州公司股票”改为了“杭州的公司”，导致问题对象发生变化，导致最终模型回复不符合用户初始意图。有效性验证模块能够准确地从初始数据集中过滤无效样本，提高训练数据集的信噪比，避免进一步加重模型的幻觉现象。

4.4.5 迭代优化对算法性能的影响

由于 QAHF 可以优化用户提示以获得更好的响应，自然的想法是我们是否可以迭代优化问题，逐步增强 LLM 的输出。因此，用 `gpt-3.5-turbo` 在 AlphaFin-test 数据集上进行了实验。具体来说，我们对原始指令进行了 5 次迭代优化，并与原始指令进行了 ragas 指标比较。如图 4 所示，通过 3 次迭代，有明显的改善，在第 4 次迭代时略有下降。附

表 4-9 有效样本与无效样本对比

类型		内容
有效 样本	原始问题	国庆黄金周对旅游股有什么影响
	优化问题	国庆黄金周对旅游股有什么影响，列点回答，从宏观到微观依次分析对旅游行业、旅游公司、旅游股的影响，最后给出综合性结论。
	模型回复	<p>国庆黄金周对旅游股的影响主要体现在以下几个方面：</p> <ol style="list-style-type: none">1. 客流量增加： …2. 板块内部分红机会： …3. 旅游股的业绩弹性： …4. 旅游股面临了一定的市场风险： … <p>综上所述，国庆黄金周对旅游股的影响主要体现在客流量增加、板块内部分红机会、旅游股的业绩弹性和面临的的市场风险等方面。投资者在关注国庆黄金周带来的投资机会的同时，还需注意国庆期间的市场风险，合理制定投资策略。</p>
无效 样本	原始问题	杭州亚运会对杭州公司股票有什么积极作用
	优化问题	杭州亚运会对杭州的公司有什么积极作用
	模型回复	<p>杭州亚运会将为杭州的公司带来积极影响，具体来说：</p> <ol style="list-style-type: none">1. 亚运会将带来大量游客： …2. 亚运会将带动周边房价上涨： …3. 亚运会将推动杭州文化产业发展： …4. 亚运会将带动杭州基础设施建设： … <p>总之，杭州亚运会将为杭州的公司带来诸多积极影响，包括游客、房价、文化产业和基础设施等方面。这些都将为杭州的公司提供更多的商业机会和发展动力。</p>

录 D 给出了一个在每次优化迭代后提示的案例。此外，我们还发现 BPO 表现出良好的保留性，当输入提示已经足够好时，它有很高的概率保留它。我们认为这是实现迭代增强的关键因素，因为它避免了对用户的原始意图进行不合理的更改。多次执行 CoT 分析与优化问题生成。

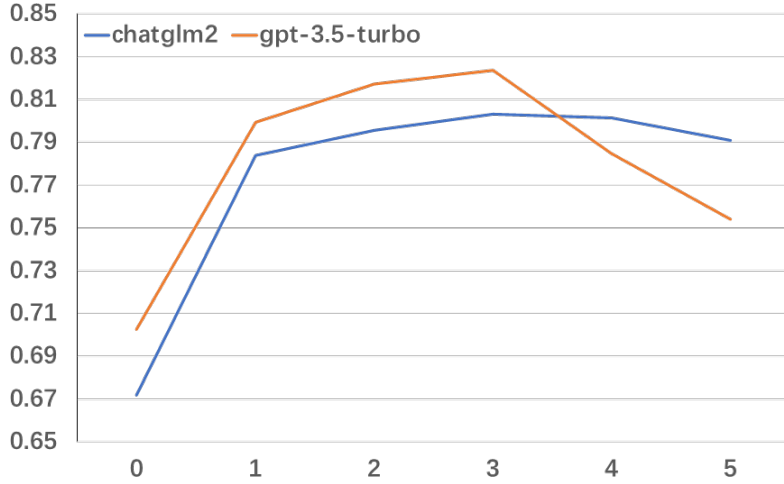


图 4-2 经过不同迭代优化次数的 ragas precision 变化示意图。

4.4.6 人类反馈对算法性能的影响

QAHF 的一个重要组成部分是利用反馈来优化用户指令。为了研究反馈对 QAHF 的快速优化有多大贡献，进行了消融实验，以比较反馈学习优化 (QAHF) 和直接使用 gpt-3.5-turbo 进行快速优化。如表 7 所示，直接优化可以提高模型性能，这验证了 llm 成为良好提示工程师的潜力。QAHF 提供了超越直接优化的进一步改进。这表明，纳入反馈允许 llm 根据所展示的用户偏好来完善提示，从而实现更有效的提示优化。

表 4-10 在偏好评价指标上提示词有效性验证模块对性能的影响

模型	方法		Self-instruct Eval			AlphaFin-test			Δ WR
	A	B	A win	Tie	B win	A win	Tie	B win	
ChatGLM	Ours	ori.	58.0%	21.0%	21.0%	61.0%	5.2%	33.8%	+32.1%
	w/o Eval	ori.	49.8%	25.5%	24.7%	46.9%	30.6%	22.5%	+24.8%
	Ours	w/o Eval	8.6%	86.3%	5.1%	15.1%	75.3%	9.6%	+4.5%

4.5 本章小结

本章提出基于人类偏好对齐的检索增强对话生成方法，能够实现与模型无关的、可解释、效果稳定的人类偏好对齐。此外，本章提出的人类偏好对齐方法还面临着受限于标注者的主观偏好的问题，因此在未来的工作中，这也是需要进一步研究和完善的方向。

总结与展望

4.6 全文总结

针对现有垂直领域检索增强对话生成的缺点与不足，本文对两方面内容展开了研究：一是基于内外部知识对齐的检索增强对话生成；二是基于人类偏好对齐的检索增强对话生成。通过所提出的两个新算法，本文能够较好的解决现有大部分垂直领域检索增强对话生成算法存在的问题。本论文的主要工作总结如下：

- 针对专有领域背景知识丰富、逻辑相对复杂的问题，本文提出了基于内外部知识对齐的检索增强对话生成方法。该方法利用一个语义切分模块提取知识文档的文档级信息和实体级信息，并将提取出来的知识分别用于构建外部知识库和内部知识注入，实现垂直领域对话模型内外部知识对齐，更好地帮助提高模型回复的事实性和可靠性。
- 针对用户问题形式多样、意图难理解的问题，本文提出了基于人类偏好对齐的检索增强对话生成方法。该方法通过采集人类对真实场景对话样本的偏好，利用大型语言模型的理解与分析能力进行问题优化，并训练单独的问题优化语言模型，实现了与模型无关的、可解释、效果稳定的人类偏好对齐，使得对话模型生成的回复更准确有效。
- 本文通过大量的实验证明了所提出的两个方法的有效性和。对于基于内外部知识对齐的垂直领域对话生成方法，本文将该方法应用于金融分析领域，并在真实股票市场价格趋势预测任务和金融问答任务上，通过多个评价维度验证其优于所有比较的方法。实验结果表明：内外部知识对齐有助于提升垂直领域对话生成的质量。对于基于人类偏好的对话生成对齐方法，本文分别在两个不同的基准测试机上与目前主流的语言模型对齐方法进行实验比较。实验结果表明：对用户问题进行优化，能同时提升知识文档召回准确率和模型理解与用户意图的一致性。

4.7 未来展望

本文围绕面向专有领域的检索增强对话生成对齐课题开展研究并取得了一定的成果。然而，作为深度学习和自然语言处理的前沿研究方向之一，检索增强对话生成对齐仍然面临诸多困难与挑战。未来的工作主要总结为如下几个方面：

- 对于基于内外部知识对齐的垂直领域对话生成方法，本文仅考虑了文本模态的外部

知识，而其他模态，如图像、语音等模态数据包含更多有助于模型进行金融分析的信息，进一步提升对话模型性能。为了解决这一问题，我们计划在接下来的工作中进行多模态文档增强的探索，并探究如何利用多模态大模型的图文理解与生成能力，从而实现利用多模态信息的知识提升算法性能。

- 本文所提出的两个方法方法仅在金融分析这一领域中验证了其有效性，我们未来将在更多的专有领域和数据集上验证这两个方法，探究其广泛适用性。
- 依赖于文档质量和语言模型的指令遵循能力
- 进仅虑文档充足的情况，然而，不同领域场景中
- 受限于标注者的主观偏好

攻读博士/硕士学位期间取得的研究成果

一、已发表（包括已接受待发表）的论文，以及已投稿、或已成文打算投稿、或拟成文投稿的论文情况(只填写与学位论文内容相关的部分):

序号	作者（全体作者，按顺序排列）	题目	发表或投稿刊物名称、级别	发表的卷期、年月、页码	与学位论文哪一部分（章、节）相关	被索引收录情况
1						
2						

注：在“发表的卷期、年月、页码”栏：

1. 如果论文已发表，请填写发表的卷期、年月、页码；
2. 如果论文已被接受，填写将要发表的卷期、年月；
3. 以上都不是，请据实填写“已投稿”，“拟投稿”。

不够请另加页。

二、与学位内容相关的其它成果（包括专利、著作、获奖项目等）

致 谢

这次你离开了没有像以前那样说再见, 再见也他妈的只是再见
我们之间从来没有想象的那么接近, 只是两棵树的距离
你是否还记得山阴路我八楼的房间, 房间里唱歌的日日夜夜
那么热的夏天你看着外面, 看着你在消逝的容颜
我多么想念你走在我身边的样子, 想起来我的爱就不能停止
南京的雨不停地下不停地下, 就像你沉默的委屈
一转眼, 我们的城市又到了夏天, 对面走来的人都眯着眼
人们不敢说话不敢停下脚步, 因为心动常常带来危险
我多么想念你走在我身边的样子, 想起来我的爱就不能停止
南京的雨不停地下不停地下, 有些人却注定要相遇
你是一片光荣的叶子, 落在我卑贱的心
像往常一样我为自己生气并且歌唱
那么乏力, 爱也吹不动的叶子

作者姓名

2020 年 7 月 10 日

于华南理工大学