



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 dialog generation technology is an important basic technology in building intelligent dialog system, which aims to solve problems in specific fields and provide professional answers for users. At present, the technology has been widely used in the fields of health care, finance, law, science and technology and other service assistants, answering robots and other applications, has high research value.

At present, the mainstream vertical-domain oriented dialogue generation methods are all designed based on retrieve-enhanced autoregressive models. However, there are the following difficulties in the current vertic-domain-oriented retrieval enhanced dialogue generation: 1) the background knowledge of the proprietary domain is rich, and the logic is relatively complex; 2) User questions come in many forms and their intentions are hard to understand. In view of the above two problems, the research plan adopted in this paper is as follows:

1) Aiming at the problem of rich background knowledge and relatively complex logic in proprietary domain, a search enhanced dialogue generation method based on internal and external knowledge alignment is proposed. In this method, a semantic segmentation module is used to extract document-level information and entity-level information from knowledge documents, and the extracted knowledge is used to construct external knowledge base and internal knowledge injection respectively, so as to realize internal and external knowledge alignment in vertical domain dialogue model. In this paper, the method is applied to financial analysis, and the validity of the method is verified on real stock price prediction task and financial question answering task.

2) Aiming at the problem of various forms of user questions and difficult to understand the intention, a search enhancement dialog generation method based on human preference alignment is proposed. By collecting human preferences for real scene dialogue samples, the method uses the understanding and analysis ability of large-scale language models to optimize the problem, and trains a single problem optimization language model, and realizes the model-independent, interpretable and stable human preference alignment. In this paper, two different benchmarking machines are used to compare with the current mainstream language model alignment methods, and it is proved that this method can improve the accuracy of knowledge document recall and

the consistency of model understanding and user intention.

**Keywords:** Dialogue System; Human Preference Alignment; Retrieval Augmented Generation; Regressive Language Model

# 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 本文主要研究内容 .....	2
1.3 本文组织结构 .....	3
第二章 对话生成研究现状 .....	4
2.1 对话系统 .....	4
2.1.1 管道式对话系统 .....	4
2.1.2 端到端式对话系统 .....	6
2.2 基于大型语言模型的对话生成 .....	7
2.2.1 基于提示工程的对话生成 .....	7
2.2.2 基于检索增强的对话生成 .....	8
2.2.3 基于模型微调的对话生成 .....	9
2.3 本章小结 .....	10
第三章 基于内外部知识对齐的检索增强对话生成 .....	11
3.1 问题描述 .....	11
3.2 基于内外部知识对齐的检索增强对话生成 .....	12
3.2.1 算法总体框架 .....	12
3.2.2 多粒度语义切分模块 .....	12
3.2.3 AlphaFin 数据集构建 .....	14
3.2.4 多级混合检索模块 .....	16
3.2.5 回答生成与后处理 .....	17
3.3 实验 .....	18
3.3.1 数据集介绍 .....	18
3.3.2 基准方法 .....	18
3.3.3 评价指标 .....	20
3.3.4 实验细节 .....	21

3.3.5	与现有方法的性能比较	21
3.3.6	多级检索模块的有效性	22
3.3.7	AlphaFin 数据集对性能的影响	22
3.4	本章小结	28
第四章	基于人类偏好对齐的检索增强对话生成	29
4.1	问题描述	29
4.2	基于人类偏好对齐的检索增强对话生成	29
4.3	总体方案	29
4.3.1	人类偏好数据采集阶段	30
4.3.2	优化查询构建阶段	30
4.3.3	查询有效性验证阶段	31
4.3.4	查询优化器训练阶段	32
4.4	实验	33
4.4.1	数据集介绍	33
4.4.2	评价指标	33
4.4.3	基础模型与基准方法	33
4.4.4	实现细节	34
4.4.5	与现有方法的性能比较	35
4.4.6	查询有效性验证模块的有效性	38
4.4.7	迭代优化对算法性能的影响	39
4.4.8	人类反馈对算法性能的影响	39
4.5	本章小结	41
总结与展望		42
4.6	全文总结	42
4.7	未来展望	42
参考文献		44
攻读博士/硕士学位期间取得的研究成果		56
致 谢		57





# 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 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。

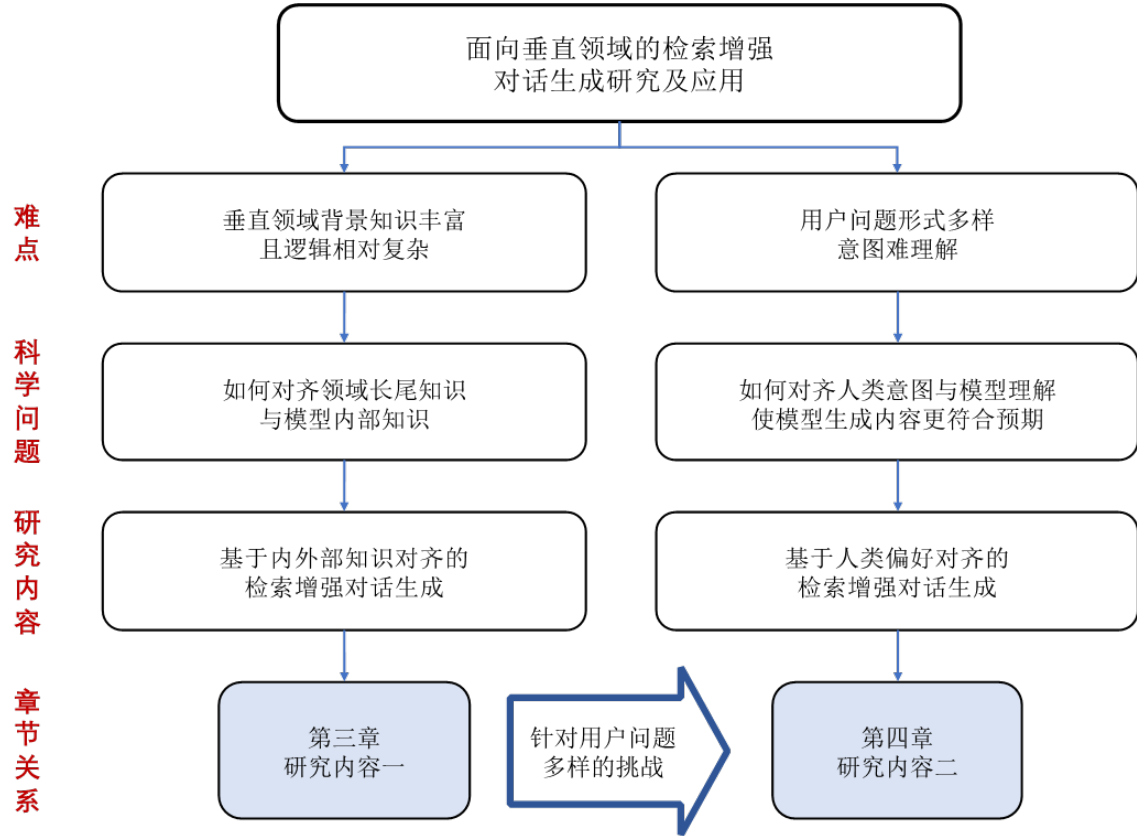


图 1-1 本文的研究思路与研究内容。

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：

### 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。

### 1.3 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 对话生成研究现状

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。本文所研究的面向垂直领域的对话生成是对话系统领域的其中一个下游任务，因此本章先对管道式对话系统和端到端式对话系统的研究进展和现状进行总结。然后，分别介绍面向垂直领域的对话生成和基于大型语言模型的对话生成的发展现状，并分析和总结其优点与不足。

### 2.1 对话系统

近年来，深度学习技术快速发展，促进了对话系统研究的发展。目前主流的对话系统实现方法从架构模式来看，可以分为管道式对话系统和端到端式对话系统。

#### 2.1.1 管道式对话系统

图2-1展示了管道式对话系统结构的示意图，管道式对话系统主要包含四个模块：（1）自然语言理解（NLU）模块；（2）对话状态跟踪（DST）模块；（3）对话策略学习（PL）模块；（4）自然语言生成（NLG）模块。

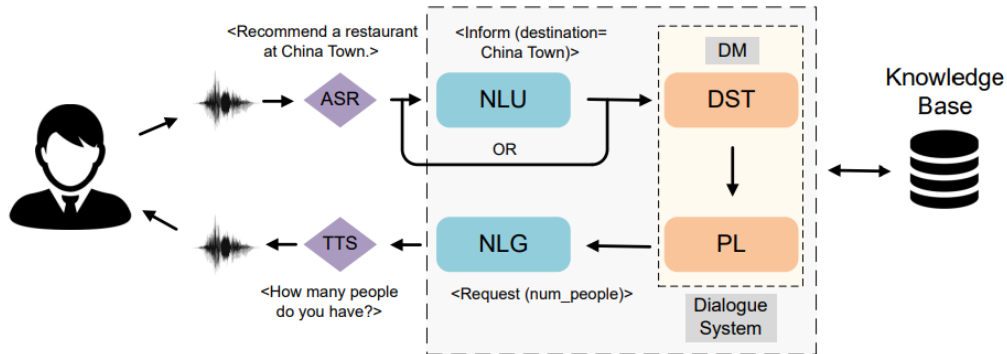


图 2-1 管道式对话系统结构示意图<sup>[1]</sup>。

自然语言理解模块对原始的用户输入进行领域分类和意图识别，并提取其关键信息填入槽位中。Deng 等人<sup>[2]</sup>和 Tur 等人<sup>[3]</sup>使用深凸网络将过去时间步的预测和当前消息结合起来，成功提高对话意图识别的准确率。Sarıkaya 等人<sup>[4]</sup>使用受限玻尔兹曼机和深度置信网络初始化深度神经网络参数，解决了深度网络难以在领域分类和意图识别任务上训练的问题。Ravuri 等人<sup>[5]</sup>利用循环神经网络（RNN）在序列预测上的优势，使用 RNN 对用户消息进行编码，以实现意图识别和领域分类。Hashemi 等人<sup>[6]</sup>使用卷积神经网络（CNN）提取对话的多层级文本特征，以实现意图识别，同时该工作说明了 CNN 在序列

预测任务上的可用性。Lee 等人<sup>[7]</sup>使用 RNN 和 CNN 整合对话历史，获得上下文信息作为额外输入补充，解决了短语句信息不足的问题。Wu 等人<sup>[8]</sup>提出预训练的面向任务对话 BERT (TOD-BERT) 模型提升意图识别任务上的准确率，同时所提出模型展现出小样本学习能力，缓解特定领域数据匮乏的问题。

对话状态跟踪模块根据当前用户消息和对话历史控制对话状态，以生成信息用于决定下一步动作。Henderson 等人<sup>[9]</sup>将深度学习模型用于对话状态跟踪，他们集成多种特征工程方法来预测每个槽-值对的概率。Mrksic 等人<sup>[10]</sup>将 RNN 应用于对话状态跟踪以获得对话上下文感知。随后，Mrksic 等人<sup>[11]</sup>提出一种多跳的神经网络对话状态跟踪器，它使用系统输出、用户消息、候选槽值对作为输入，并基于对话历史对当前槽值对进行二分类预测。但是上述方法基于预定义的槽名与槽值，每一轮对话都需要对所有槽分别进行一次二分类，系统响应慢。为了解决填槽类别多导致的计算复杂度高的问题，Lei 等人<sup>[12]</sup>提出信念片段，即一种与槽名相对应的对话上下文片段，他们构建了一个两阶段 CopyNet 网络来拷贝和存储对话历史中的槽值，用以生成回答，促进端到端训练，有效提升了系统在未登录词 (OOV) 下的准确度。Wang 等人<sup>[13]</sup>提出了一种基于 BERT 模型的槽值预测方法，使用槽位注意力方法对相关片段进行检索，并使用槽值归一化方法将片段转化为最终的槽值。

对话策略学习模块根据对话状态决定下一步行为，主流方法是监督学习和强化学习。Zhang 等人<sup>[14]</sup>提出一种高效资源调度方法，充分利用用户交互，在训练时使用概率调度器来分配对话样本，同时利用一个控制器决定使用真实样本还是模拟样本，有效提升强化学习训练的效果。Takanobu 等人<sup>[15]</sup>提出一种多机对话策略学习方法，分别使用两个机器人扮演用户和系统，同时学习对话策略，同时，他们还加入特定角色的强化学习反馈，促进角色回答的生成。

自然语言生成模块将系统的行为转化为自然语言，即最终输出，并返回给用户。Wen 等人<sup>[16]</sup>提出一种基于 RNN 的统计语言模型，通过语义约束和语法树学习回答生成，同时还使用 CNN 对候选回答进行排序，选出最佳回答。Wen 等人<sup>[17]</sup>对循环语言模型采取先在大规模通用语料上预训练，然后在小规模特定领域语料上微调的训练方法，有效提升了循环语言模型的领域适应能力。Li 等人<sup>[18]</sup>提出一种迭代矫正网络以迭代式修正所生成的字符，首先使用监督学习训练网络，然后通过强化学习进行微调，并将槽位不一致的惩罚项添加到训练奖励中。

总结：管道式对话系统利用神经网络模型进行信息抽取和文本分类，而对话流程基

于人工预定义的模板，因此特征相对固定，难以实现领域迁移。

### 2.1.2 端到端式对话系统

随着 GPT<sup>[19]</sup>，BERT<sup>[20]</sup>，T5<sup>[21]</sup>等序列到序列（Seq2Seq）预训练模型的兴起，基于 Seq2Seq 模型的端到端对话系统逐渐成为主流。这类方法将对话系统的核心功能隐式地集成到一个复杂的神经网络模型当中，降低了系统模块复杂度。从研究侧重点上来看，端到端式对话系统的研究可分为两类：（1）基于模型架构的研究；（2）基于训练方法的研究。

Sordoni 等人<sup>[22]</sup>提出一种上下文感知的 Seq2Seq 模型 HRED，该模型同时学习字符级和对话级的文本表示，有效解决上下文感知的在线查询建议问题，提供少见且高质量的结果。Serban 等人<sup>[23]</sup>进一步提出 VHRED 模型，对序列间的复杂依赖进行建模，通过在解码器中增加一个隐变量，在回答的多样性、长度和质量上都有一定的提升。Weston 等人<sup>[24]</sup>提出一种记忆网络，使用 RNN 网络在每个时间步上向后传递历史信息。然而，这种模型包含五个模块，每个模块都需要单独进行监督训练，因此不适用于端到端式对话系统。Sukhbaatar 等人<sup>[25]</sup>在他们的工作基础上，提出了一种端到端的记忆网络，引入注意力方法，依次进行权重计算、记忆选择、最终预测。在某些对话场景下，对话系统需要直接从用户输入中引用或摘要内容，而传统 Seq2Seq 模型的输出维度受限于输入的维度。为解决该问题，Oriol 等人<sup>[26]</sup>提出 Pointer Net，每个时间步的输出都来自输入序列，将字符预测问题转化为位置预测问题，以适应输入长度的变化。然而，Pointer Net 仅从输入中选择字符的局限性较大，Gu 等人<sup>[27]</sup>提出 CopyNet，在解码阶段的每个时间步中，决定复制输入中的字符还是生成一个新的字符，将复制概率和生成概率加和，得到最终的预测概率。Balakrishnan 等人<sup>[28]</sup>提出一种约束解码模块，以提升对话系统所生成回答的语义准确性。Chen 等人<sup>[29]</sup>使用两个长期记忆模块分别存储知识元组和对话历史，然后用一个工作记忆模块来控制字符的生成，有效提升对话相关知识的检索精确度。Gao 等人<sup>[30]</sup>使用一个释义模型对回答生成模型进行增强，释义模型与整个对话系统联合训练，用于增强训练样本。

Wang 等人<sup>[31]</sup>提出一种基于增量学习的训练方法，通过构建一个不确定性估计模块，在模型所生成回复的自信度低于阈值时使用人类回答作为结果，同时通过在线学习拟合人类回答，有效提升了模型回复质量。Dai 等人<sup>[32]</sup>使用模型无关的元学习方法，仅依靠少量训练样本，在真实线上服务场景下有效提升模型的迁移能力和可靠性。He 等人<sup>[33]</sup>提出一种“双教师单学生”知识蒸馏训练框架，首先两个教师模型分别以知识检索



和回答生成作为训练目标，进行强化学习训练，然后让学生模型模仿教师模型的输出，实现专业知识的迁移。Zhang 等人<sup>[34]</sup>使用 GPT-2<sup>[35]</sup>作为基座模型，在 Reddit 社交媒体语料上训练，并引入最大互信息（MMI）评分函数，使对话系统能生成更相关、内容丰富、上下文一致的回复。Daniel 等人<sup>[36]</sup>提出基于 Evolved Transformer 的 Meena 架构模型，将模型参数量扩充至 2.6B，同时提出对话评估指标 SSA，包含逻辑性和特异性两个维度，适用于对话模型的大规模人工评估场景，具有易于理解、一致性高等优势。Roller 等人<sup>[37]</sup>使用具有特定对话能力的 BST 数据集对预训练模型进行微调，得到 BlenderBot 模型，BST 数据集涵盖引人入胜、善于倾听、博学、同理心、有个性等对话能力，同时对比了检索式、生成式、混合式三种基于 Transformer 的模型架构，最终生成式模型的表现超过当时的最先进方法。Bao 等人<sup>[38]</sup>提出对话行为隐变量，用以表征不同的说话风格，进而生成多样的回答，使用该方法在对话语料上训练 BERT 得到 PLATO 模型，该模型具备多轮流畅对话能力。

总结：端到端式对话系统的架构相对简单，且生成的回复更多样。然而，这类对话系统通常需要大量高质量对话语料，且模型尺寸一般比较大，训练资源开销大。

## 2.2 基于大型语言模型的对话生成

近年来，随着 ChatGPT<sup>[39]</sup>的发布，国内外先后出现许多预训练大型语言模型，如 ChatGLM<sup>[40]</sup>、文心一言<sup>[41]</sup>、通义千问<sup>[42]</sup>、LLaMA<sup>[43]</sup>、百川<sup>[44]</sup>等。因此，越来越多的研究者将大型语言模型作为对话生成研究的基础。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。

### 2.2.1 基于提示工程的对话生成

Radford 等人<sup>[35]</sup>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，Brown 等人<sup>[19]</sup>提出 Few-shot 提示方法，在模型输入中提供高质量的示例，提升模型在复杂任务上的性能。Wei 等人<sup>[45]</sup>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。Wang 等人<sup>[46]</sup>引入自洽性（Self-Consistency）解码策略，Long 等人<sup>[47]</sup>提出思维树（Tree-of-Thought, ToT）提示方法，ToT 是对 CoT 方法的扩展，通过管理树结构的中间推理步骤，同时使用深度优先或广度优先等搜索算法，对推理链路进行系统性扩展，使模型在得到错误答案时能够进行回溯，ToT

在 24 类游戏任务上取得了 74% 的成功率，而 CoT 仅有 4%。Yao 等人<sup>[48]</sup>提出思维图（Graph-of-Thought, GoT）提示方法，使用基于图的架构，更好地适应人类非线性思考地特性。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。

总结：提示工程方法的复杂度和成本相对较低，具有较高的灵活性，但生成的结果准确性受限于基座模型自身的能力，通常会出现不符合事实的回答。

### 2.2.2 基于检索增强的对话生成

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，Lewis 等人<sup>[49]</sup>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。Wang 等人<sup>[50]</sup>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。Jagerman 等人<sup>[51]</sup>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。Liu 等人<sup>[52]</sup>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。Asai 等人<sup>[53]</sup>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。Cui 等人<sup>[54]</sup>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。

总结：检索增强方法能很好地改善大型语言模型的幻觉问题，得到高准确性的回复。然而，现有的检索增强方法更多的关注如何提升外部知识的召回准确性，没有考虑基座模型内部知识与外部知识之间的偏差，存在性能瓶颈。



### 2.2.3 基于模型微调的对话生成

Wei 等人<sup>[55]</sup>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。Li 等人<sup>[56]</sup>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，Brian 等人<sup>[57]</sup>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。Liu 等人<sup>[58]</sup>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。Hu 等人<sup>[59]</sup>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。

Long 等人<sup>[39]</sup>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数据集对语言模型监督微调（SFT）、收集人类偏好排名数据集并训练奖励模型（RM）、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。Li 等人<sup>[60]</sup>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。Zheng 等人<sup>[61]</sup>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。Dong 等人<sup>[62]</sup>提出 RAFT 方法，在采集的人类偏好数据中选取多个高质量样本，继续对 SFT 模型进行监督微调，利用更多次采样和更少的梯度计算，让模型更稳定和鲁棒。Yuan 等人<sup>[63]</sup>在 RLHF 的基础上提出 RRHF 算法，直接使用偏好数据训练语言模型，结合排名损失和 SFT 损失，在性能上与 PPO 方法相近，但实现相对简单且训练稳定。Cheng 等人<sup>[64]</sup>提出 BPO 黑盒提示优化算法，通过采集人类偏好数据，训练提示优化器，对用户指令进行优化，使模型生成的内容更符合用户期望。Rafailov 等人<sup>[65]</sup>提出 DPO 算法，其思想与 RRHF 相似，但

同时引入 SFT 模型的约束，保证在不使用 SFT 损失的情况下训练依然稳定，该方法在多个开源 RM 数据集上获得优于 RLHF 的奖励得分，且对超参数的敏感度更低，效果更稳定。Liu 等人<sup>[66]</sup>提出 RSO 算法，使用拒绝采样得到高质量回答的分布。上述方法依赖于人类反馈数据，标注成本高昂。为解决该问题，Bai 等人<sup>[67]</sup>提出 RLAIIF 方法，用宪法人工智能（CAI）代替人类进行偏好的标记工作，实验表明 RLAIIF 能够达到与 RLHF 相当的性能。此外，Li 等人<sup>[68]</sup>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。

总结：模型微调方法通过调整模型内部参数，能够控制模型内在的行为模式，对齐人类偏好，但是对动态数据缺乏实时性，且训练成本高昂。

## 2.3 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 问题描述

基于大型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。语言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，以金融分析领域为应用场景对算法进行验证。金融分析领域主要涵盖两个关键任务：（1）股票趋势预测任务；（2）金融问答任务。

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， $\pi$  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：

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

目前机器学习和深度学习的方法已经被广泛应用在该任务上，并取得了一定的进展。然而，这些方法通常只能基于过去的市场价格数据预测未来股票的涨跌，而无法给出具体的分析过程和原因，其预测结果对投资者用户来说缺乏可解释性。同时，这类方法难以将新闻、研报等非结构化的文本信息用于股票趋势预测中。

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  $\pi$  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。目前大部分的金融问答方法主要基于大

型语言模型，这类模型在大规模语料上经过预训练，具有强大的文本理解和文本生成能力。但由于金融领域数据集的稀缺性和模型微调的知识滞后性，语言模型存在不可控的“幻觉”问题，即编造没有现实基础的事实或细节。因此，一些基于检索增强生成的方法提出利用知识库检索出与用户查询相关的外部知识辅助模型生成回答，一定程度上缓解了大型语言模型的幻觉问题和实时性问题。然而，对于某些金融领域问题，涉及到广泛的金融背景知识，仅依靠知识库检索得到的知识文档并不足以让模型给出正确的分析和解答。

## 3.2 基于内外部知识对齐的检索增强对话生成

### 3.2.1 算法总体框架

针对以上问题，本章提出基于内外部知识对齐的检索增强对话生成方法。整体算法框架如图3-1所示，整个框架主要包括三个部分：（1）对金融知识文档进行多粒度语义切分，得到一系列金融知识文档块，并构建外部知识库用于存储和检索这些知识文档块；（2）构建 AlphaFin 金融数据集，为模型注入金融领域的内部知识，同时与知识库中的外部知识进行对齐；（3）利用多级混合检索，对用户查询和知识库中的文档块计算语义相似度，得到与用户查询相关性最高的一系列文档块，最后与指令提示词拼接输入模型，得到回答。

### 3.2.2 多粒度语义切分模块

如图3-2所示，外部知识库的构建是检索增强生成中的重要组成部分，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，本文所使用的提示词如表3-1所示。

表 3-1 实体级对话生成所使用的提示词。

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

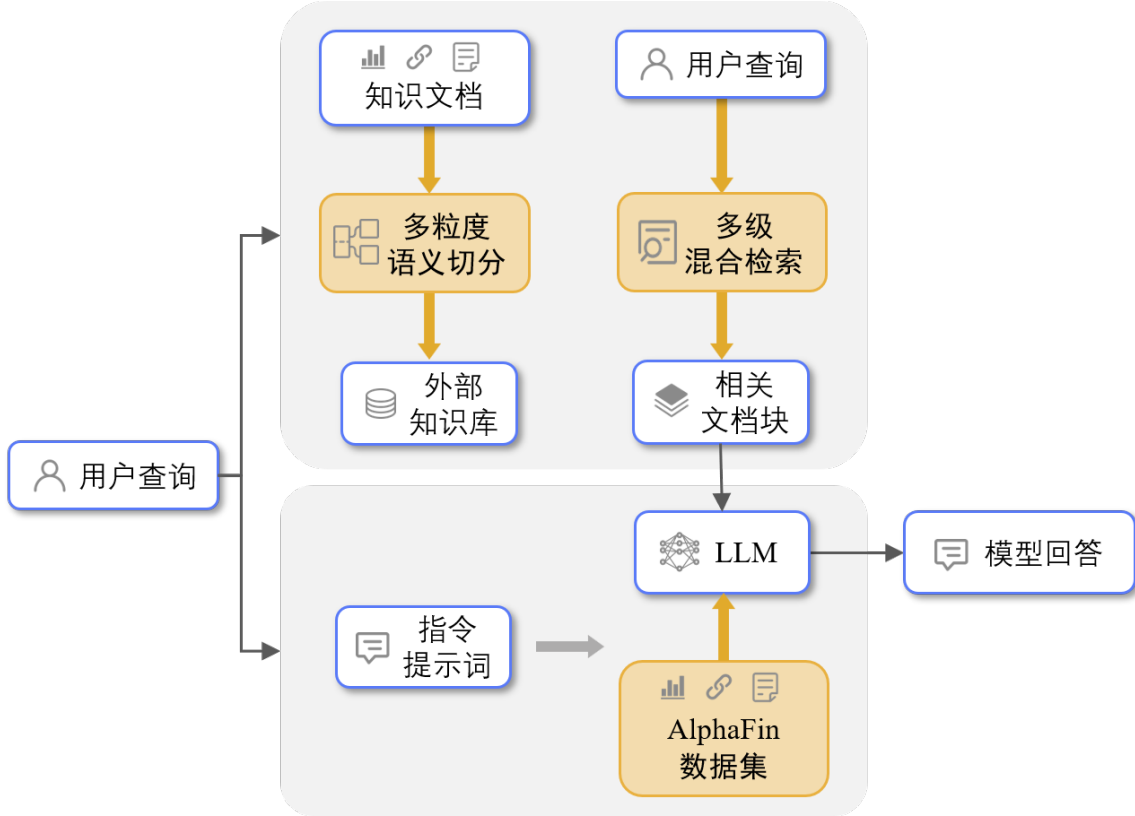


图 3-1 本章所提出的检索增强对话生成框架示意图。

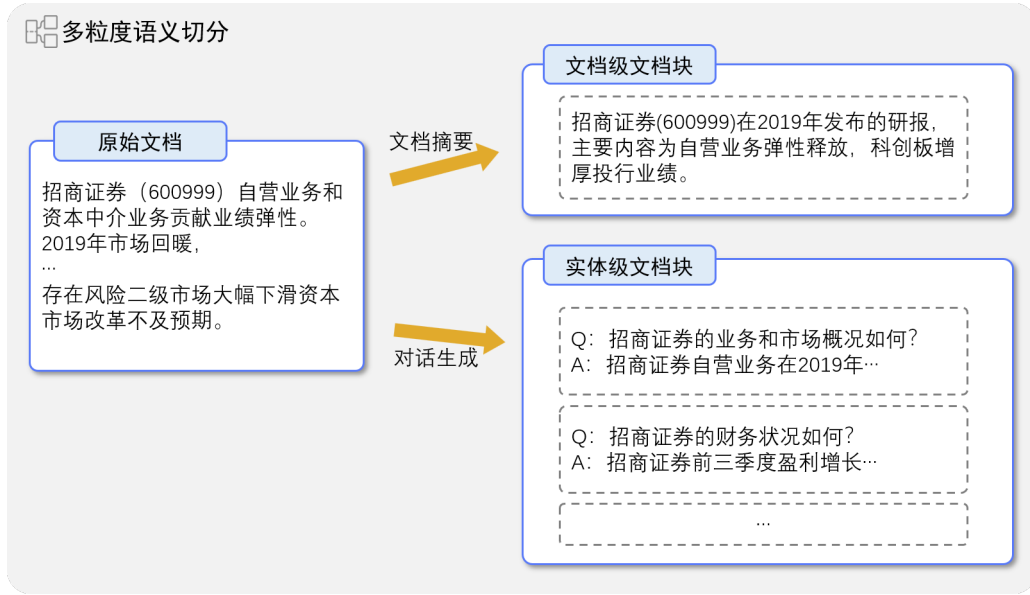


图 3-2 对话框架中的多粒度语义切分模块示意图。

对于文档  $d_k$ ，其语义切分过程如下：

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

$$(q_{k0}, a_{k0}), (q_{k1}, a_{k1}), \dots = LLM_{qa}(d_k) \quad (3-5)$$

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,  $\text{SentEmbed}$  是句子嵌入模型, 如 BGE、SGPT 等。

### 3.2.3 AlphaFin 数据集构建

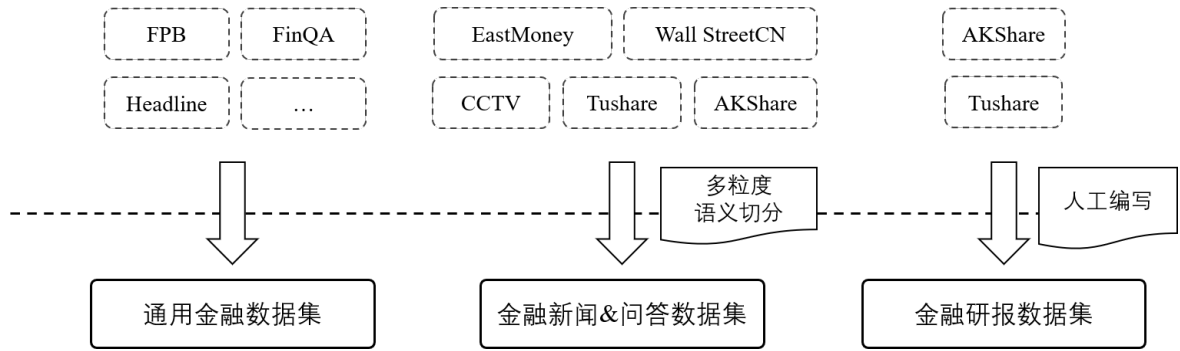


图 3-3 数据集预处理过程示意图。

如图3-3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<sup>[69]</sup>、FinQA<sup>[70]</sup>、ConvFinQA<sup>[71]</sup>、Headline<sup>[72]</sup>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, 用于增强模型的多语言能力、信息抽取能力和摘要能力。

**金融新闻数据集:** 为模型提供真实世界的金融知识, 我们整合了在线新闻资源, 如 CCTV 的金融板块和华尔街见闻。利用 ChatGPT<sup>[39]</sup>提取每条新闻的摘要, 构建金融新闻数据集。这个过程提高了模型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<sup>[73]</sup>和 AKshare<sup>[74]</sup>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。

**金融研报数据集:** 我们通过 DataYes 数据平台 API 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，并使用如表3-2格式来生成最终数据。其中，金融研报数据集包含两种类型的数据：1) Raw 类型，样本输出中仅包含股票涨跌结果和涨跌概率；2) CoT 类型，在给出股票涨跌预测结果前，先对输入中的公司研报和市场数据信息进行基本面和技术面的详细分析和逻辑推导，这部分 CoT 分析由金融领域专家人工编写，以保证其正确性和质量。

表 3-2 金融研报数据集样本格式。

数据集	Raw	CoT
系统指令	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。	
输入	<div>&lt;  公司研报  &gt;</div> <div>发布日期: &lt;date&gt;</div> <div>研报题目: &lt;title&gt;</div> <div>目标价格: &lt;target price&gt;</div> <div>研报摘要: &lt;abstract&gt;</div> <div>&lt;  市场数据  &gt;</div> <div>1. 股票价格（元）:&lt;price&gt;;</div> <div>2. 日涨跌幅（%）:&lt;change&gt;;</div> <div>3. 日成交量（亿）:&lt;volumn&gt;</div>	
输出	这个股票的下月最终收益结果是:’跌’, 下跌概率: 中上	<div>通过研报和市场走势数据可以得出以下结论:</div> <div>1、基本面: &lt;basic analysis&gt;。</div> <div>2、技术面: &lt;tech analysis&gt;。</div> <div>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</div>

最终得到的 AlphaFin 数据集各组成部分的统计信息如表3-3所示，任务类型涵盖 NLP 基础任务、金融新闻、金融问答、研报分析，不同任务类型、不同语言的数据量相当，以保证模型学习到的各项能力趋于平衡。

表 3-3 AlphaFin 数据集统计信息。

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

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。

### 3.2.4 多级混合检索模块

知识检索过程如图3-4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。

$$d^* = \arg \max_{d_k} \frac{e_Q^\top \cdot e_{sk}}{|e_Q| |e_{sk}|} \quad (3-8)$$

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  $a_{k\_}$ 。



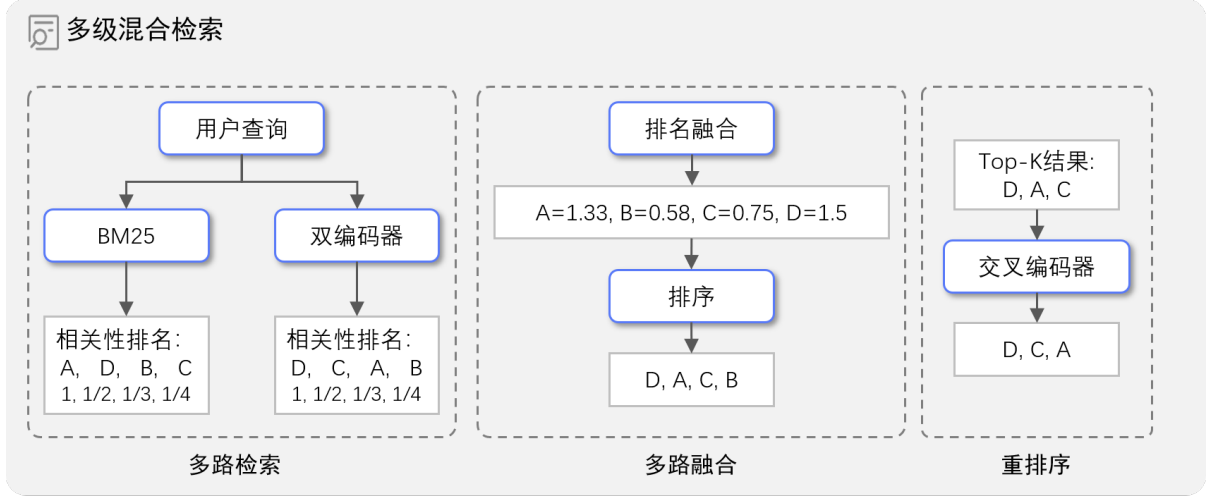


图 3-4 对话框中的多级混合检索模块示意图。

### 3.2.5 回答生成与后处理

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  $R_t$ 。

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

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

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  $\text{Res}_i$ 。

$$\text{Res}_i = \text{LLM}(I_i) \quad (3-11)$$

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

$$\text{Pred}_i = \begin{cases} \text{up}, & \text{if } \text{“up”} \in \text{Res}_i \\ \text{down}, & \text{else} \end{cases} \quad (3-12)$$

$$C_{\text{chosen}} = c_i | \text{Pred}_i = \text{up} \quad (3-13)$$

另外，我们按月滚动执行该投资策略。每个月，对于  $C_{\text{chosen}}$  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。

$$AR_m = AR_{m-1} + \sum_{c_i \in C_{chosen}} \omega_{c_i} R_{c_i} \quad (3-14)$$

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。  $\omega_{c_i}$  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。

$$\omega_{c_i} = \frac{v_i}{\sum_{c_n \in C_{chosen}} v_n} \quad (3-15)$$

### 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 数据集介绍

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。

#### 3.3.2 基准方法

**股票指数:** 我们选取了中国股票市场的指数,包括上证 50 (SSE50) 上证指数 (SCI)、沪深 300 (CSI300) 和创业板指数 (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 (RNN) 的一种。它的主要特点是具有两个门: 更新门 (update gate) 和重置门 (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 (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。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。

### 3.3.3 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 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<sup>[75]</sup>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。本节使用 ragas 框架中的 context\_precision、context\_recall、faithfulness 三项指标对本章方法进行评价：

- 1) Context\_Precision: 用于评估 contexts 的相关性及其 ranking
- 2) Context\_Recall: 通过估计 answer 和 context 的 TP 和 FN，计算 context 的召回率
- 3) 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。

3.3.4 实验细节

对于股票涨跌预测任务，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。对于金融问答任务，我们对比了模型生成能力，并分别使用人工和 GPT4 模型作为评判员。所有模型的生成策略都是贪心搜索，以达到最优和稳定的性能。其中，在所有模型训练过程中，所使用的超参数如下：batch size 16，LoRA rank 8，cosine lr scheduler，学习率 5e-5，参数精度 bf16，其余硬件和软件环境如表3-4所示。具体来说，在股票涨跌预测任务中，我们为第一步训练了 4 个 epoch，为第二步训练了 20 个 epoch。在金融问答任务中，我们使用股票涨跌预测中的模型作为基础模型，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

3.3.5 与现有方法的性能比较

如图3-5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3-5和表3-6可知，Stock-Chain 的 ARR 和 ACC 分别达到了最高的 30.8% 和 55.7%，体现了其有效性。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

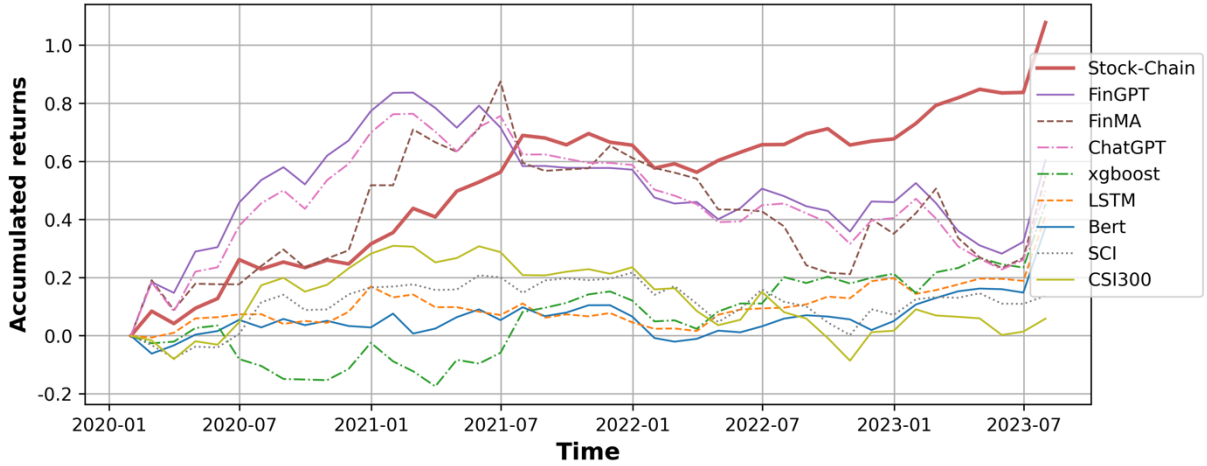


图 3-5 不同方法在 2020 年 1 月到 2023 年 7 月期间进行股票趋势预测的累计收益情况。

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，本章实现了 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。

本节使用人类和 GPT-4 作为评判员，对每个 LLM 在测试数据集上的输出性能进行评分。在人工评判结果如表 3-7 所示，本文所提出方法在内容有效性方面优于其他 LLM。从图 5 可以看出，本文方法相比 ChatGLM2 的胜率超过 60%，相比 FinLLM 如 FinGPT 的胜率达到 62%。GPT-4 评判结果如表 3-8，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%。具体回复示例如表 3-9 和表 3-10 所示。

### 3.3.6 多级检索模块的有效性

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 指标对效果进行量化。从表 3-11 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。同时，增加多级检索模块前后的回复结果如表 3-12 所示。

### 3.3.7 AlphaFin 数据集对性能的影响

本节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模型的股票趋势预测能力；由表 3-13 可知，相对于 ChatGLM2, llm 对股票价格的预测能力在使

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%	<b>55.9%</b>
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	<b>30.8%</b>	55.7%

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 的输出仅包含涨或跌，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%。

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3-14，可以观察到本章方法在 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，本章方法在经过新闻和研报数据集的微调后均取得了最优的性能。

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	<b>11.9%</b>	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	<b>29.1%</b>	<b>19.6%</b>	<b>1.573</b>	13.3%	<b>2.314</b>	<b>10</b>

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	$\Delta$ WR
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 基线方法回复示例。

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 本章方法回复示例。

问题	<p>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。</p>
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，大市指数在 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	<b>0.6717</b>	<b>0.8430</b>	<b>0.8005</b>

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“鹤鹑蛋”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。

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 $\uparrow$	SR $\uparrow$	输出长度 $\uparrow$	无效答案率 $\downarrow$
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 $\uparrow$	ROUGE-2 $\uparrow$	ROUGE-L $\uparrow$
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

### 3.4 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 问题描述

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<sup>[19]</sup>时代，提示工程<sup>[76]</sup>，以及自动提示搜索<sup>[77]</sup>和 Prefix-Tuning<sup>[56-57,78]</sup>，已经发展成为特定任务的校准者。然而，它们后来被指令微调<sup>[55]</sup>所取代和目前的 RLHF 方法<sup>[39,67]</sup>，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

### 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) 人类偏好数据采集阶段；2) 查询提示词构建阶段；3) 查询有效性验证阶段；4) 查询优化器训练阶段。

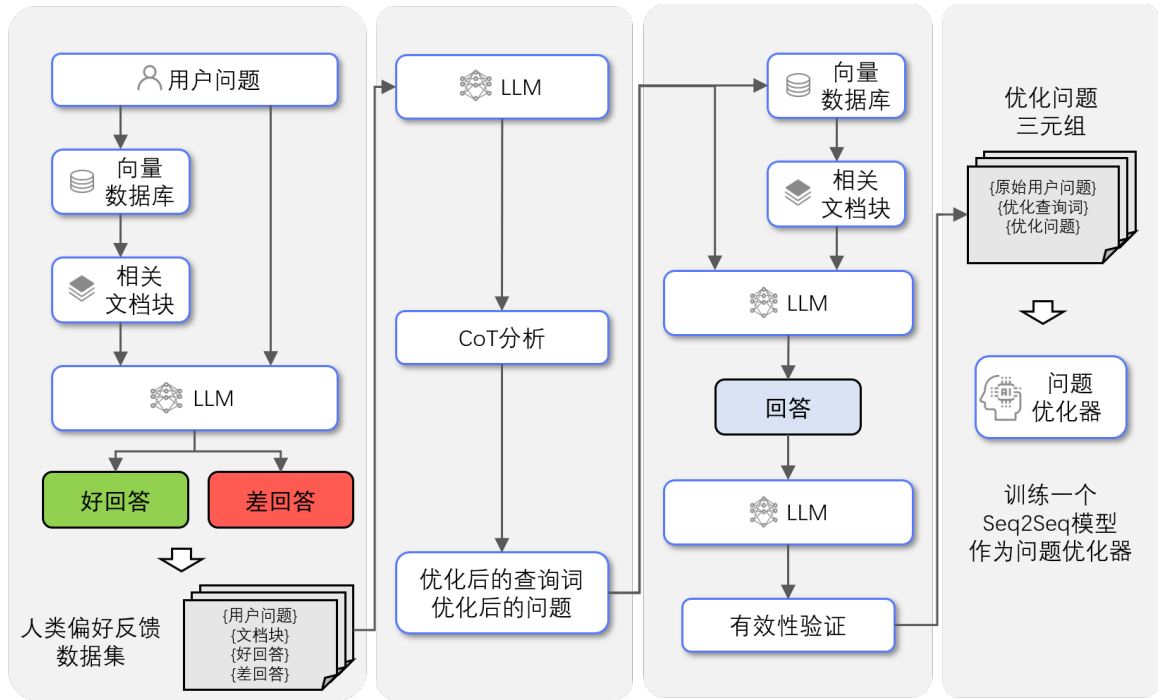


图 4-1 基于人类偏好对齐的检索增强对话生成框架图。

#### 4.3.1 人类偏好数据采集阶段

为了建模人类偏好，本节首先搜集了一系列金融问答数据集，该数据集样本为“查询-回答”对，查询主题为金融相关专业问题。此外，本章使用基于规则的方法，过滤掉数据集中与金融领域无关、样本长度过短、查询范围过窄的低质量样本，以保证样本查询存在优化空间。然后，在向量数据库中检索与用户查询相关的文档块，并将结果输入 LLM，通过随机采样的生成策略，得到 LLM 的多组回答。最后，通过人工标注的方式，从准确性、有帮助性、全面性等方面对 LLM 回答进行偏序关系标注，从每组回答中选出综合评分更高的回答作为“好回答”，评分最低的回答作为“差回答”，得到人类偏好反馈数据集。本章主要关注单轮对话的对话生成。

#### 4.3.2 优化查询构建阶段

随后，本节使用 LLM 来对数据集中的原始用户查询进行优化。本节使用基于 CoT 的提示工程方法，引导 LLM 通过人类偏好反馈数据识别出用户偏好特征，分析造成 LLM 生成差回答的原因以及好回答中蕴含的人类偏好特征。接着，分别针对数据库检索和 LLM 对话场景各自生成一个优化后的查询语句，使得在新的查询语句下，检索增强对话系统生成的回答分布更趋近于好回答分布，区别于差回答的分布。本节所使用的查询优化提示词如表 4-1 所示。

表 4-1 查询优化提示词格式。

查询优化提示词
<pre># SYSTEM  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。  # QUERY  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。 问题是: &lt;original query&gt;  # DOCS  &lt;docs&gt;  # GOOD RESPONSE  &lt;good response&gt;  # BAD RESPONSE  &lt;bad response&gt;  # TASK  你的任务是：  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。</pre>

### 4.3.3 查询有效性验证阶段

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。其中，本阶段所使用的查询有效性验证提示词如表4-2所示。

表 4-2 查询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：&lt;original query&gt;</p> <p># DOCS</p> <p>&lt;docs&gt;</p> <p># RESPONSE</p> <p>&lt; response&gt;</p> <p># TASK</p> <p>你的任务是：</p> <ol style="list-style-type: none"> <li>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；</li> <li>2. 判断 RESPONSE 是否能够准确、可靠地回答 QUERY；</li> <li>3. 对 RESPONSE 进行评分，分数取值范围为 [1, 10]。</li> </ol>

#### 4.3.4 查询优化器训练阶段

基于前述步骤所构建的人类偏好数据集，训练一个 Seq2Seq 模型作为查询优化器，实现模型无关的、可插拔的用户查询自动优化。形式上，给定原始用户查询  $Q_{ori}$ ，优化器生成优化后的查询  $Q_{opt}$ ：

$$Q_{opt} = [Q_{search}, [SEP], Q_{qa}] \quad (4-1)$$

其中， $Q_{search}$  表示用于数据库检索的查询， $Q_{qa}$  表示用于回答生成的查询， $[SEP]$  为特殊字符，用于分隔  $Q_{search}$  和  $Q_{qa}$ 。



本节使用交叉熵损失函数作为训练目标，损失函数定义为：

$$L = -\frac{1}{N} \sum_{t=1}^N \log P(x_t | Q_{\text{ori}}, x_{<t}) \quad (4-2)$$

其中， $N$  是  $Q_{\text{opt}}$  的长度， $x_t$  表示  $Q_{\text{opt}}$  中的第  $t$  个字符。本章选择 Qwen-7B 作为基座模型，以更好地学习  $Q_{\text{ori}}$  和  $Q_{\text{opt}}$  之间的隐式偏好映射。同时，在目前的主流 LLM 中，7B 模型的参数量相对较小，能够以更低的成本进行训练和推理，具有更低的推理延迟，作为本方法中的查询优化器具有一定优势。

## 4.4 实验

### 4.4.1 数据集介绍

FinGPT-FiQA<sup>[79]</sup> 是 Wang 等人创建的人工评估数据集。它包含了 17.1k 个金融领域的用户查询与回答样本，这些样本均采样于真实应用场景。我们从中抽取了 1000 条样本作为本节实验的测试数据。

AlphaFin-test 是本文第三章中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。

### 4.4.2 评价指标

本节使用人类和 GPT4 模型作为评判员，对两种不同模型或不同方法的生成结果进行两两对比，以更好地评估不同方法之间性能差异。另外，本节还采用 Ragas<sup>[75]</sup> 指标自动化评估模型回复的质量。本节所使用的 ragas 评价指标包括：（1）Context Precision；（2）Context Recall；（3）Faithfulness。

### 4.4.3 基础模型与基准方法

基座模型：

Qwen-7B-Chat<sup>[80]</sup>：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<sup>[81]</sup>：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<sup>[40]</sup>: 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。

ChatGP<sup>[39]</sup>T: 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<sup>[82]</sup>: 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<sup>[64]</sup>: 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。

#### 4.4.4 实现细节

对于 QAHF, 我们使用 Qwen-7B 作为优化器基座模型, 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 优化器,  $\beta_1=0.9$  和  $\beta_2=0.999$ 。本实验将学习率设置为  $2e-5$ , 热启动步长为 0.1%, 使用线性衰减学习率。每个 GPU 的 batch size 大小为 4。对于 RLHF 训练, RM 模型训练和 PPO 优化只训练 1 个 epoch。其中, 本实验 RM 模型的训练数据来自于本章方法所构建的人类偏好数据集, RM 模型在同分布测试集上达到了 78% 的准确率。。所有实验均在 8×80GB NVIDIA A800 gpu 上进行。QAHF 采用 Top-p=0.9 和

Temperature=0.6 进行推理解码，而所有测试的基座模型都使用默认的解码策略。

4.4.5 与现有方法的性能比较

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

模型	方法		FinGPT-FiQA Eval			AlphaFin-test			$\Delta$ 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%

详细的实验结果如表4-3所示。基于本章所提出的 QAHF 方法，在所有基座模型和所有数据集上，其性能均优于原始查询，取得了更高的胜率，证明了本章方法的有效性和广泛适用性。值得注意的是，在 ChatGLM、FinGPT、Qwen 等相对小尺寸的开源模型上，QAHF 相比于原始查询的胜率分别提高了 32.1%、30.5% 和 20.1%，FinGPT 在数据集 FinGPT-FiQA Eval 上甚至达到了 37.4% 的提升，而在 gpt-3.5-turbo 这类模型上，胜率提升在 10% 以内，说明本章方法对于尺寸更小、基础能力相对更弱的基座模型能够收获更大的对齐收益。

表 4-4 QAHF 与 BPO 在 FinGPT-FiQA Eval 和 AlphaFin-test 上的性能对比

模型	方法		FinGPT-FiQA Eval			AlphaFin-test			$\Delta$ 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%

如表4-4所示，BPO 和 QAHF 均成功提升了 ChatGLM 和 gpt-3.5-turbo 模型的性能。此外，在 QAHF 与 BPO 的所有对比实验中，QAHF 均取得了正向的胜率提升，近一步

证明 QAHF 方法相较于 BPO 方法的优越性。

表 4-5 QAHF 与 PPO 在 FinGPT-FiQA Eval 和 AlphaFin-test 上的性能对比

模型	方法		AlphaFin-test			$\Delta$ WR
	A	B	A win	Tie	B win	
ChatGLM	PPO	ori.	13.7%	75.4%	10.9%	+2.8%
	Ours	PPO	49.3%	20.6%	30.1%	+19.2%
	Ours	ori.	61.0%	5.2%	33.8%	+27.2%

同时，本实验还对比了 QAHF 和 PPO 强化学习对齐方法的性能差异，结果如表4-5所示。从表中可以看出，由于 PPO 训练稳定性较差，虽然训练实现了收敛，但 PPO 方法对模型性能的提升较为有限，75.4% 的样本没有明显提升，胜率仅有 2.8%。而 QAHF 方法依然能获得 10% 以上的稳定提升。

表 4-6 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。

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4-6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4-7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

表 4-7 使用不同对齐方法优化用户查询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>

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。

#### 4.4.6 查询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4-8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 4-8 在偏好评价指标上查询有效性验证模块对性能的影响

模型	方法		FinGPT-FiQA Eval			AlphaFin-test			$\Delta$ 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%

本节基于 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4-9所示。从表中结果可以看出，去除提示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。

同时，表4-10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



表 4-9 在 Ragas 指标上查询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)

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。

#### 4.4.7 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，本节用 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 4-2 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。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。

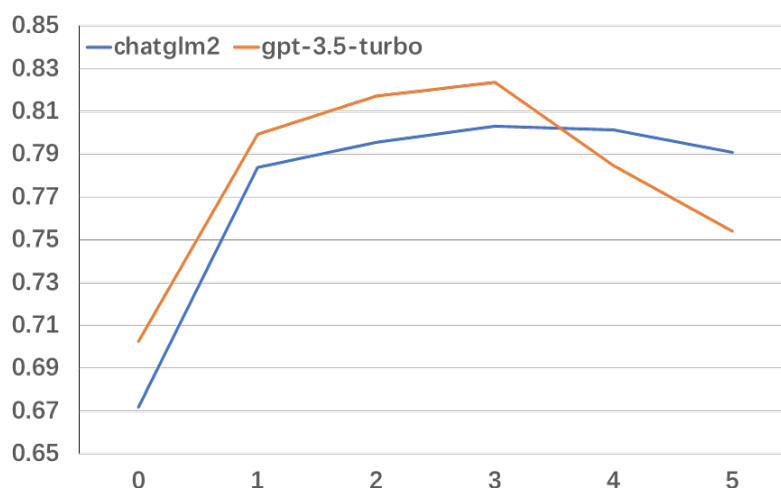


图 4-2 经过不同迭代优化次数的 ragas precision 变化示意图。

#### 4.4.8 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 4-11 所示，直接优化可以提高模型性能，这验证了 llm

表 4-10 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"> <li>1. 客流量增加： …</li> <li>2. 板块内部分红机会： …</li> <li>3. 旅游股的业绩弹性： …</li> <li>4. 旅游股面临了一定的市场风险： …</li> </ol> <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"> <li>1. 亚运会将带来大量游客： …</li> <li>2. 亚运会将带动周边房价上涨： …</li> <li>3. 亚运会将推动杭州文化产业发展： …</li> <li>4. 亚运会将带动杭州基础设施建设： …</li> </ol> <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>



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 4-11 在偏好评价指标上人类反馈对性能的影响

模型	方法		AlphaFin-test			$\Delta$ WR
	A	B	A win	Tie	B win	
ChatGLM	Ours	ori.	61.0%	5.2%	33.8%	+27.2%
	w/o Human	ori.	39.4%	35.8%	24.8%	+14.6%
	Ours	w/o Human	45.7%	20.3%	34.0%	+11.7%

## 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检索增强对话生成方法，本文仅考虑了文本模态的外部

知识，而其他模态，如图像、语音等模态数据可能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，未来工作将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。

- 本文所提出的两种方法仅在金融分析这一领域中验证了其有效性，未来工作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。
- 本文所提出的两种方法都基于大型语言模型的检索增强生成技术，方法的有效性在一定程度上对知识文档的质量和所使用的语言模型的指令遵循能力存在依赖性。未来工作将针对低信噪比外部知识和小尺寸基座模型的检索增强对话生成方式进行改进，提出一种新的方法，降低对话生成对文档和基座模型的敏感度。
- 对于基于人类偏好对齐的检索增强对话生成方法，该方法显式地引入了人类标注者的主观偏好，因此最终生成的对话回复可能会受到标注者的价值观影响。未来工作将探究如何减弱标注者个人偏好对方法性能的影响，实现更广泛的人类偏好对齐。

## 参考文献

- [1] Ni J, Young T, Pandelea V, et al. Recent advances in deep learning based dialogue systems: a systematic survey[J/OL]. Artif. Intell. Rev., 2023, 56(4): 3055-3155. <https://doi.org/10.1007/s10462-022-10248-8>. DOI: 10.1007/S10462-022-10248-8.
- [2] Deng L, Tür G, He X, et al. Use of kernel deep convex networks and end-to-end learning for spoken language understanding[C/OL]//2012 IEEE Spoken Language Technology Workshop (SLT), Miami, FL, USA, December 2-5, 2012. IEEE, 2012: 210-215. <https://doi.org/10.1109/SLT.2012.6424224>. DOI: 10.1109/SLT.2012.6424224.
- [3] Tür G, Deng L, Hakkani-Tür D, et al. Towards deeper understanding: Deep convex networks for semantic utterance classification[C/OL]//2012 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2012, Kyoto, Japan, March 25-30, 2012. IEEE, 2012: 5045-5048. <https://doi.org/10.1109/ICASSP.2012.6289054>. DOI: 10.1109/ICASSP.2012.6289054.
- [4] Sarikaya R, Hinton G E, Deoras A. Application of Deep Belief Networks for Natural Language Understanding[J/OL]. IEEE ACM Trans. Audio Speech Lang. Process., 2014, 22(4): 778-784. <https://doi.org/10.1109/TASLP.2014.2303296>. DOI: 10.1109/TASLP.2014.2303296.
- [5] Ravuri S V, Stolcke A. Recurrent neural network and LSTM models for lexical utterance classification[C/OL]//INTERSPEECH 2015, 16th Annual Conference of the International Speech Communication Association, Dresden, Germany, September 6-10, 2015. ISCA, 2015: 135-139. <https://doi.org/10.21437/Interspeech.2015-42>. DOI: 10.21437/INTERSPEECH.2015-42.
- [6] Hashemi H B, Asiaee A, Kraft R. Query intent detection using convolutional neural networks[C]//International conference on web search and data mining, workshop on query understanding: vol. 23. 2016.
- [7] Lee J Y, Deroncourt F. Sequential Short-Text Classification with Recurrent and Convolutional Neural Networks[C/OL]//Knight K, Nenkova A, Rambow O. NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June

- 12-17, 2016. The Association for Computational Linguistics, 2016: 515-520. <https://doi.org/10.18653/v1/n16-1062>. DOI: 10.18653/V1/N16-1062.
- [8] Wu C, Hoi S C H, Socher R, et al. ToD-BERT: Pre-trained Natural Language Understanding for Task-Oriented Dialogues[J/OL]. CoRR, 2020, abs/2004.06871. arXiv: 2004.06871. <https://arxiv.org/abs/2004.06871>.
- [9] Henderson M, Thomson B, Young S J. Deep Neural Network Approach for the Dialog State Tracking Challenge[C/OL]//Proceedings of the SIGDIAL 2013 Conference, The 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue, 22-24 August 2013, SUPELEC, Metz, France. The Association for Computer Linguistics, 2013: 467-471. <https://aclanthology.org/W13-4073/>.
- [10] Mrksic N, Séaghdha D Ó, Thomson B, et al. Multi-domain Dialog State Tracking using Recurrent Neural Networks[C/OL]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 2: Short Papers. The Association for Computer Linguistics, 2015: 794-799. <https://doi.org/10.3115/v1/p15-2130>. DOI: 10.3115/V1/P15-2130.
- [11] Mrksic N, Séaghdha D Ó, Wen T, et al. Neural Belief Tracker: Data-Driven Dialogue State Tracking[C/OL]//Barzilay R, Kan M. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers. Association for Computational Linguistics, 2017: 1777-1788. <https://doi.org/10.18653/v1/P17-1163>. DOI: 10.18653/V1/P17-1163.
- [12] Lei W, Jin X, Kan M, et al. Sequicity: Simplifying Task-oriented Dialogue Systems with Single Sequence-to-Sequence Architectures[C/OL]//Gurevych I, Miyao Y. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers. Association for Computational Linguistics, 2018: 1437-1447. <https://aclanthology.org/P18-1133/>. DOI: 10.18653/V1/P18-1133.
- [13] Wang Y, Guo Y, Zhu S. Slot Attention with Value Normalization for Multi-Domain Dialogue State Tracking[C/OL]//Webber B, Cohn T, He Y, et al. Proceedings of the

- 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020. Association for Computational Linguistics, 2020: 3019-3028. <https://doi.org/10.18653/v1/2020.emnlp-main.243>. DOI: 10.18653/V1/2020.EMNLP-MAIN.243.
- [14] Zhang Z, Li X, Gao J, et al. Budgeted Policy Learning for Task-Oriented Dialogue Systems[C/OL]//Korhonen A, Traum D R, Màrquez L. Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers. Association for Computational Linguistics, 2019: 3742-3751. <https://doi.org/10.18653/v1/p19-1364>. DOI: 10.18653/V1/P19-1364.
- [15] Takanobu R, Liang R, Huang M. Multi-Agent Task-Oriented Dialog Policy Learning with Role-Aware Reward Decomposition[C/OL]//Jurafsky D, Chai J, Schluter N, et al. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020. Association for Computational Linguistics, 2020: 625-638. <https://doi.org/10.18653/v1/2020.acl-main.59>. DOI: 10.18653/V1/2020.ACL-MAIN.59.
- [16] Wen T, Gasic M, Kim D, et al. Stochastic Language Generation in Dialogue using Recurrent Neural Networks with Convolutional Sentence Reranking[C/OL]//Proceedings of the SIGDIAL 2015 Conference, The 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue, 2-4 September 2015, Prague, Czech Republic. The Association for Computer Linguistics, 2015: 275-284. <https://doi.org/10.18653/v1/w15-4639>. DOI: 10.18653/V1/W15-4639.
- [17] Wen T, Gasic M, Mrksic N, et al. Multi-domain Neural Network Language Generation for Spoken Dialogue Systems[C/OL]//Knight K, Nenkova A, Rambow O. NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016. The Association for Computational Linguistics, 2016: 120-129. <https://doi.org/10.18653/v1/n16-1015>. DOI: 10.18653/V1/N16-1015.
- [18] Li Y, Yao K, Qin L, et al. Slot-consistent NLG for Task-oriented Dialogue Systems with Iterative Rectification Network[C/OL]//Jurafsky D, Chai J, Schluter N, et al. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL

- 2020, Online, July 5-10, 2020. Association for Computational Linguistics, 2020: 97-106. <https://doi.org/10.18653/v1/2020.acl-main.10>. DOI: 10.18653/V1/2020.ACL-MAIN.10.
- [19] Brown T B, Mann B, Ryder N, et al. Language Models are Few-Shot Learners[C/OL]//Larochelle H, Ranzato M, Hadsell R, et al. Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual. 2020. <https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfc4967418bfb8ac142f64a-Abstract.html>.
- [20] Devlin J, Chang M, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[C/OL]//Burstein J, Doran C, Solorio T. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers). Association for Computational Linguistics, 2019: 4171-4186. <https://doi.org/10.18653/v1/n19-1423>. DOI: 10.18653/V1/N19-1423.
- [21] Raffel C, Shazeer N, Roberts A, et al. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer[J/OL]. J. Mach. Learn. Res., 2020, 21: 140:1-140:67. <http://jmlr.org/papers/v21/20-074.html>.
- [22] Sordoni A, Bengio Y, Vahabi H, et al. A Hierarchical Recurrent Encoder-Decoder for Generative Context-Aware Query Suggestion[C/OL]//Bailey J, Moffat A, Aggarwal C C, et al. Proceedings of the 24th ACM International Conference on Information and Knowledge Management, CIKM 2015, Melbourne, VIC, Australia, October 19 - 23, 2015. ACM, 2015: 553-562. <https://doi.org/10.1145/2806416.2806493>. DOI: 10.1145/2806416.2806493.
- [23] Serban I V, Sordoni A, Lowe R, et al. A Hierarchical Latent Variable Encoder-Decoder Model for Generating Dialogues[C/OL]//Singh S, Markovitch S. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA. AAAI Press, 2017: 3295-3301. <https://doi.org/10.1609/aaai.v31i1.10983>. DOI: 10.1609/AAAI.V31I1.10983.
- [24] Weston J, Chopra S, Bordes A. Memory Networks[C/OL]//Bengio Y, LeCun Y. 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA,

- May 7-9, 2015, Conference Track Proceedings. 2015. <http://arxiv.org/abs/1410.3916>.
- [25] Sukhbaatar S, Szlam A, Weston J, et al. End-To-End Memory Networks[C/OL]//Cortes C, Lawrence N D, Lee D D, et al. Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada. 2015: 2440-2448. <https://proceedings.neurips.cc/paper/2015/hash/8fb21ee7a2207526da55a679f0332de2-Abstract.html>.
- [26] Vinyals O, Fortunato M, Jaitly N. Pointer Networks[C/OL]//Cortes C, Lawrence N D, Lee D D, et al. Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada. 2015: 2692-2700. <https://proceedings.neurips.cc/paper/2015/hash/29921001f2f04bd3baee84a12e98098f-Abstract.html>.
- [27] Gu J, Lu Z, Li H, et al. Incorporating Copying Mechanism in Sequence-to-Sequence Learning[C/OL]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers. The Association for Computer Linguistics, 2016. <https://doi.org/10.18653/v1/p16-1154>. DOI: 10.18653/V1/P16-1154.
- [28] Balakrishnan A, Rao J, Upasani K, et al. Constrained Decoding for Neural NLG from Compositional Representations in Task-Oriented Dialogue[C/OL]//Korhonen A, Traum D R, Màrquez L. Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers. Association for Computational Linguistics, 2019: 831-844. <https://doi.org/10.18653/v1/p19-1080>. DOI: 10.18653/V1/P19-1080.
- [29] Chen X, Xu J, Xu B. A Working Memory Model for Task-oriented Dialog Response Generation[C/OL]//Korhonen A, Traum D R, Màrquez L. Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers. Association for Computational Linguistics, 2019: 2687-2693. <https://doi.org/10.18653/v1/p19-1258>. DOI: 10.18653/V1/P19-1258.
- [30] Gao S, Zhang Y, Ou Z, et al. Paraphrase Augmented Task-Oriented Dialog Generation [C/OL]//Jurafsky D, Chai J, Schluter N, et al. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020.



- Association for Computational Linguistics, 2020: 639-649. <https://doi.org/10.18653/v1/2020.acl-main.60>. DOI: 10.18653/V1/2020.ACL-MAIN.60.
- [31] Wang W, Zhang J, Li Q, et al. Incremental Learning from Scratch for Task-Oriented Dialogue Systems[C/OL]//Korhonen A, Traum D R, Màrquez L. Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers. Association for Computational Linguistics, 2019: 3710-3720. <https://doi.org/10.18653/v1/p19-1361>. DOI: 10.18653/V1/P19-1361.
- [32] Dai Y, Li H, Tang C, et al. Learning Low-Resource End-To-End Goal-Oriented Dialog for Fast and Reliable System Deployment[C/OL]//Jurafsky D, Chai J, Schluter N, et al. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020. Association for Computational Linguistics, 2020: 609-618. <https://doi.org/10.18653/v1/2020.acl-main.57>. DOI: 10.18653/V1/2020.ACL-MAIN.57.
- [33] He W, Yang M, Yan R, et al. Amalgamating Knowledge from Two Teachers for Task-oriented Dialogue System with Adversarial Training[C/OL]//Webber B, Cohn T, He Y, et al. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020. Association for Computational Linguistics, 2020: 3498-3507. <https://doi.org/10.18653/v1/2020.emnlp-main.281>. DOI: 10.18653/V1/2020.EMNLP-MAIN.281.
- [34] Zhang Y, Sun S, Galley M, et al. DIALOGPT : Large-Scale Generative Pre-training for Conversational Response Generation[C/OL]//Celikyilmaz A, Wen T. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, ACL 2020, Online, July 5-10, 2020. Association for Computational Linguistics, 2020: 270-278. <https://doi.org/10.18653/v1/2020.acl-demos.30>. DOI: 10.18653/V1/2020.ACL-DEMOS.30.
- [35] Radford A, Wu J, Child R, et al. Language models are unsupervised multitask learners [J]. OpenAI blog, 2019, 1(8): 9.
- [36] Adiwardana D, Luong M, So D R, et al. Towards a Human-like Open-Domain Chatbot [J/OL]. CoRR, 2020, abs/2001.09977. arXiv: 2001.09977. <https://arxiv.org/abs/2001.09977>.

9977.

- [37] Roller S, Dinan E, Goyal N, et al. Recipes for Building an Open-Domain Chatbot[C/OL]//Merlo P, Tiedemann J, Tsarfaty R. Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, EACL 2021, Online, April 19 - 23, 2021. Association for Computational Linguistics, 2021: 300-325. <https://doi.org/10.18653/v1/2021.eacl-main.24>. DOI: 10.18653/V1/2021.EACL-MAIN.24.
- [38] Bao S, He H, Wang F, et al. PLATO: Pre-trained Dialogue Generation Model with Discrete Latent Variable[C/OL]//Jurafsky D, Chai J, Schluter N, et al. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020. Association for Computational Linguistics, 2020: 85-96. <https://doi.org/10.18653/v1/2020.acl-main.9>. DOI: 10.18653/V1/2020.ACL-MAIN.9.
- [39] Ouyang L, Wu J, Jiang X, et al. Training language models to follow instructions with human feedback[C/OL]//Koyejo S, Mohamed S, Agarwal A, et al. Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022. 2022. [http://papers.nips.cc/paper%5C\\_files/paper/2022/hash/b1efde53be364a73914f58805a001731-Abstract-Conference.html](http://papers.nips.cc/paper%5C_files/paper/2022/hash/b1efde53be364a73914f58805a001731-Abstract-Conference.html).
- [40] Zeng A, Liu X, Du Z, et al. GLM-130B: An Open Bilingual Pre-trained Model[C/OL]//The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023. OpenReview.net, 2023. <https://openreview.net/pdf?id=-Aw0rrPUF>.
- [41] Sun Y, Wang S, Feng S, et al. ERNIE 3.0: Large-scale Knowledge Enhanced Pre-training for Language Understanding and Generation[J/OL]. CoRR, 2021, abs/2107.02137. arXiv: 2107.02137. <https://arxiv.org/abs/2107.02137>.
- [42] Bai J, Bai S, Chu Y, et al. Qwen Technical Report[J/OL]. CoRR, 2023, abs/2309.16609. arXiv: 2309.16609. <https://doi.org/10.48550/arXiv.2309.16609>. DOI: 10.48550/ARXIV.2309.16609.
- [43] Touvron H, Lavril T, Izacard G, et al. LLaMA: Open and Efficient Foundation Language Models[J/OL]. CoRR, 2023, abs/2302.13971. arXiv: 2302.13971. <https://doi.org/10.48>

- 550/arXiv.2302.13971. DOI: 10.48550/ARXIV.2302.13971.
- [44] Yang A, Xiao B, Wang B, et al. Baichuan 2: Open Large-scale Language Models[J/OL]. CoRR, 2023, abs/2309.10305. arXiv: 2309.10305. <https://doi.org/10.48550/arXiv.2309.10305>. DOI: 10.48550/ARXIV.2309.10305.
- [45] Wei J, Wang X, Schuurmans D, et al. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models[C/OL]//Koyejo S, Mohamed S, Agarwal A, et al. Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022. 2022. [http://papers.nips.cc/paper%5C\\_files/paper/2022/hash/9d5609613524ecf4f15af0f7b31abca4-Abstract-Conference.html](http://papers.nips.cc/paper%5C_files/paper/2022/hash/9d5609613524ecf4f15af0f7b31abca4-Abstract-Conference.html).
- [46] Wang X, Wei J, Schuurmans D, et al. Self-Consistency Improves Chain of Thought Reasoning in Language Models[C/OL]//The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023. OpenReview.net, 2023. <https://openreview.net/pdf?id=1PL1NIMMrw>.
- [47] Long J. Large Language Model Guided Tree-of-Thought[J/OL]. CoRR, 2023, abs/2305.08291. arXiv: 2305.08291. <https://doi.org/10.48550/arXiv.2305.08291>. DOI: 10.48550/ARXIV.2305.08291.
- [48] Yao Y, Li Z, Zhao H. Beyond Chain-of-Thought, Effective Graph-of-Thought Reasoning in Large Language Models[J/OL]. CoRR, 2023, abs/2305.16582. arXiv: 2305.16582. <https://doi.org/10.48550/arXiv.2305.16582>. DOI: 10.48550/ARXIV.2305.16582.
- [49] Lewis P S H, Perez E, Piktus A, et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks[C/OL]//Larochelle H, Ranzato M, Hadsell R, et al. Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual. 2020. <https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html>.
- [50] Wang L, Yang N, Wei F. Query2doc: Query Expansion with Large Language Models [C/OL]//Bouamor H, Pino J, Bali K. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, EMNLP 2023, Singapore, December 6-10, 2023. Association for Computational Linguistics, 2023: 9414-9423. <https://aclanthology>

- .org/2023.emnlp-main.585.
- [51] Jagerman R, Zhuang H, Qin Z, et al. Query Expansion by Prompting Large Language Models[J/OL]. CoRR, 2023, abs/2305.03653. arXiv: [2305.03653](https://arxiv.org/abs/2305.03653). <https://doi.org/10.48550/ARXIV.2305.03653>. DOI: [10.48550/ARXIV.2305.03653](https://doi.org/10.48550/ARXIV.2305.03653).
  - [52] Liu N F, Lin K, Hewitt J, et al. Lost in the Middle: How Language Models Use Long Contexts[J/OL]. CoRR, 2023, abs/2307.03172. arXiv: [2307.03172](https://arxiv.org/abs/2307.03172). <https://doi.org/10.48550/ARXIV.2307.03172>. DOI: [10.48550/ARXIV.2307.03172](https://doi.org/10.48550/ARXIV.2307.03172).
  - [53] Asai A, Wu Z, Wang Y, et al. Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection[J/OL]. CoRR, 2023, abs/2310.11511. arXiv: [2310.11511](https://arxiv.org/abs/2310.11511). <https://doi.org/10.48550/ARXIV.2310.11511>. DOI: [10.48550/ARXIV.2310.11511](https://doi.org/10.48550/ARXIV.2310.11511).
  - [54] Cui J, Li Z, Yan Y, et al. ChatLaw: Open-Source Legal Large Language Model with Integrated External Knowledge Bases[J/OL]. CoRR, 2023, abs/2306.16092. arXiv: [2306.16092](https://arxiv.org/abs/2306.16092). <https://doi.org/10.48550/ARXIV.2306.16092>. DOI: [10.48550/ARXIV.2306.16092](https://doi.org/10.48550/ARXIV.2306.16092).
  - [55] Wei J, Bosma M, Zhao V Y, et al. Finetuned Language Models are Zero-Shot Learners [C/OL]//The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022. OpenReview.net, 2022. <https://openreview.net/forum?id=gEZrGCozdqR>.
  - [56] Li X L, Liang P. Prefix-Tuning: Optimizing Continuous Prompts for Generation[C/OL] //Zong C, Xia F, Li W, et al. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021. Association for Computational Linguistics, 2021: 4582-4597. <https://doi.org/10.18653/v1/2021.acl-long.353>. DOI: [10.18653/v1/2021.acl-long.353](https://doi.org/10.18653/v1/2021.acl-long.353).
  - [57] Lester B, Al-Rfou R, Constant N. The Power of Scale for Parameter-Efficient Prompt Tuning[C/OL]//Moens M F, Huang X, Specia L, et al. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Online: Association for Computational Linguistics, 2021: 3045-3059. <https://aclanthology.org/2021.emnlp-main.243>. DOI: [10.18653/v1/2021.emnlp-main.243](https://doi.org/10.18653/v1/2021.emnlp-main.243).
  - [58] Liu X, Ji K, Fu Y, et al. P-Tuning v2: Prompt Tuning Can Be Comparable to Fine-tuning

- Universally Across Scales and Tasks[J/OL]. CoRR, 2021, abs/2110.07602. arXiv: [2110.07602](https://arxiv.org/abs/2110.07602). <https://arxiv.org/abs/2110.07602>.
- [59] Hu E J, Shen Y, Wallis P, et al. LoRA: Low-Rank Adaptation of Large Language Models [C/OL]//The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022. OpenReview.net, 2022. <https://openreview.net/forum?id=nZeVKeeFYf9>.
- [60] Li Y, Wei F, Zhao J, et al. RAIN: Your Language Models Can Align Themselves without Finetuning[J/OL]. CoRR, 2023, abs/2309.07124. arXiv: [2309.07124](https://arxiv.org/abs/2309.07124). <https://doi.org/10.48550/ARXIV.2309.07124>. DOI: [10.48550/ARXIV.2309.07124](https://doi.org/10.48550/ARXIV.2309.07124).
- [61] Zheng R, Dou S, Gao S, et al. Secrets of RLHF in Large Language Models Part I: PPO [J/OL]. CoRR, 2023, abs/2307.04964. arXiv: [2307.04964](https://arxiv.org/abs/2307.04964). <https://doi.org/10.48550/arXiv.2307.04964>. DOI: [10.48550/ARXIV.2307.04964](https://doi.org/10.48550/ARXIV.2307.04964).
- [62] Dong H, Xiong W, Goyal D, et al. RAFT: Reward rAnked FineTuning for Generative Foundation Model Alignment[J/OL]. CoRR, 2023, abs/2304.06767. arXiv: [2304.06767](https://arxiv.org/abs/2304.06767). <https://doi.org/10.48550/arXiv.2304.06767>. DOI: [10.48550/ARXIV.2304.06767](https://doi.org/10.48550/ARXIV.2304.06767).
- [63] Yuan Z, Yuan H, Tan C, et al. RRHF: Rank Responses to Align Language Models with Human Feedback without tears[J/OL]. CoRR, 2023, abs/2304.05302. arXiv: [2304.05302](https://arxiv.org/abs/2304.05302). <https://doi.org/10.48550/arXiv.2304.05302>. DOI: [10.48550/ARXIV.2304.05302](https://doi.org/10.48550/ARXIV.2304.05302).
- [64] Cheng J, Liu X, Zheng K, et al. Black-Box Prompt Optimization: Aligning Large Language Models without Model Training[J/OL]. CoRR, 2023, abs/2311.04155. arXiv: [2311.04155](https://arxiv.org/abs/2311.04155). <https://doi.org/10.48550/arXiv.2311.04155>. DOI: [10.48550/ARXIV.2311.04155](https://doi.org/10.48550/ARXIV.2311.04155).
- [65] Rafailov R, Sharma A, Mitchell E, et al. Direct Preference Optimization: Your Language Model is Secretly a Reward Model[C/OL]//Oh A, Naumann T, Globerson A, et al. Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023. 2023. [http://papers.nips.cc/paper%5C\\_files/paper/2023/hash/a85b405ed65c6477a4fe8302b5e06ce7-Abstract-Conference.html](http://papers.nips.cc/paper%5C_files/paper/2023/hash/a85b405ed65c6477a4fe8302b5e06ce7-Abstract-Conference.html).
- [66] Liu T, Zhao Y, Joshi R, et al. Statistical Rejection Sampling Improves Preference Optimization[J/OL]. CoRR, 2023, abs/2309.06657. arXiv: [2309.06657](https://arxiv.org/abs/2309.06657). <https://doi.org/10.48550/ARXIV.2309.06657>.

- 8550/arXiv.2309.06657. DOI: 10.48550/ARXIV.2309.06657.
- [67] Bai Y, Kadavath S, Kundu S, et al. Constitutional AI: Harmlessness from AI Feedback [J/OL]. CoRR, 2022, abs/2212.08073. arXiv: 2212.08073. <https://doi.org/10.48550/arXiv.2212.08073>. DOI: 10.48550/ARXIV.2212.08073.
- [68] Li Z, Xu T, Zhang Y, et al. ReMax: A Simple, Effective, and Efficient Reinforcement Learning Method for Aligning Large Language Models[J/OL]. CoRR, 2023, abs/2310.10505. arXiv: 2310.10505. <https://doi.org/10.48550/arXiv.2310.10505>. DOI: 10.48550/ARXIV.2310.10505.
- [69] Malo P, Sinha A, Korhonen P J, et al. Good debt or bad debt: Detecting semantic orientations in economic texts[J/OL]. J. Assoc. Inf. Sci. Technol., 2014, 65(4): 782-796. <https://doi.org/10.1002/asi.23062>. DOI: 10.1002/ASI.23062.
- [70] Zhu F, Lei W, Huang Y, et al. TAT-QA: A Question Answering Benchmark on a Hybrid of Tabular and Textual Content in Finance[C/OL]//Zong C, Xia F, Li W, et al. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021. Association for Computational Linguistics, 2021: 3277-3287. <https://doi.org/10.18653/v1/2021.acl-long.254>. DOI: 10.18653/V1/2021.ACL-LONG.254.
- [71] Feng D, Dai Y, Huang J, et al. Empowering Many, Biasing a Few: Generalist Credit Scoring through Large Language Models[J/OL]. CoRR, 2023, abs/2310.00566. arXiv: 2310.00566. <https://doi.org/10.48550/arXiv.2310.00566>. DOI: 10.48550/ARXIV.2310.00566.
- [72] Sinha A, Khandait T. Impact of News on the Commodity Market: Dataset and Results [J/OL]. CoRR, 2020, abs/2009.04202. arXiv: 2009.04202. <https://arxiv.org/abs/2009.04202>.
- [73] TuShare[Z]. <https://github.com/waditu/tushare>. 2019.
- [74] King A. AKShare[Z]. <https://github.com/akfamily/akshare>. 2019.
- [75] ES S, James J, Anke L E, et al. RAGAs: Automated Evaluation of Retrieval Augmented Generation[C/OL]//Aletas N, Clercq O D. Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2024 - Sys-

- tem Demonstrations, St. Julians, Malta, March 17-22, 2024. Association for Computational Linguistics, 2024: 150-158. <https://aclanthology.org/2024.eacl-demo.16>.
- [76] Reynolds L, McDonell K. Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm[C/OL]//Kitamura Y, Quigley A, Isbister K, et al. CHI '21: CHI Conference on Human Factors in Computing Systems, Virtual Event / Yokohama Japan, May 8-13, 2021, Extended Abstracts. ACM, 2021: 314:1-314:7. <https://doi.org/10.1145/3411763.3451760>. DOI: 10.1145/3411763.3451760.
- [77] Shin T, Razeghi Y, IV R L L, et al. AutoPrompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts[C/OL]//Webber B, Cohn T, He Y, et al. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020. Association for Computational Linguistics, 2020: 4222-4235. <https://doi.org/10.18653/v1/2020.emnlp-main.346>. DOI: 10.18653/V1/2020.EMNLP-MAIN.346.
- [78] Liu X, Zheng Y, Du Z, et al. GPT Understands, Too[J/OL]. CoRR, 2021, abs/2103.10385. arXiv: 2103.10385. <https://arxiv.org/abs/2103.10385>.
- [79] Wang N, Yang H, Wang C D. FinGPT: Instruction Tuning Benchmark for Open-Source Large Language Models in Financial Datasets[J]. NeurIPS Workshop on Instruction Tuning and Instruction Following, 2023.
- [80] Bai J, Bai S, Chu Y, et al. Qwen Technical Report[J]. arXiv preprint arXiv:2309.16609, 2023.
- [81] Yang H, Liu X Y, Wang C D. FinGPT: Open-Source Financial Large Language Models [J]. FinLLM Symposium at IJCAI 2023, 2023.
- [82] Schulman J, Wolski F, Dhariwal P, et al. Proximal Policy Optimization Algorithms [J/OL]. CoRR, 2017, abs/1707.06347. arXiv: 1707.06347. <http://arxiv.org/abs/1707.06347>.

## 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	发表或投稿刊物/会议名称	作者（仅注明第几作者）	发表年份	与学位论文哪一部分（章、节）相关	被索引收录情况
1	International Conference on Computational Linguistics (COLING), CCF-B 类会议	共同第一作者	2024	第三章	
2	Annual Meeting of the Association for Computational Linguistics (ACL), CCF-A 类会议	共同第一作者	2024	第三章	

注：1. 请在“作者”一栏填写本人是第几作者，例：“第一作者”或“导师第一，本人第二”等；  
2. 若文章未发表或未被接受，请在“发表年份”一栏据实填写“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）

1. 发明专利：已受理一项发明专利，导师第一发明人，本人第二发明人，2024



## 致 谢

感谢世界

李振宇

2024 年 3 月 30 日

于华南理工大学