
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

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 4 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语言模型设计的。然而，现有面向垂直领域的检索增强对话生成任务存在以下难点：1）内外部知识不一致问题，即现有方法多关注如何提升外部知识召回准确度，忽略了模型内部知识与之存在的分布差异，从而影响外部知识发挥作用；2）人类偏好对齐问题，即人类用户意图与模型理解之间存在偏差，造成模型生成不符合用户预期的回答。针对以上存在的两个问题，本文采取的研究方案如下：

1) 针对内外部知识不一致的问题，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，并将提取出来的信息分别用于构建外部知识库和监督训练数据集，最后使用知识对齐后的数据集微调对话模型，有效缓解了模型在预训练阶段注入的内部知识与推理阶段的外部知识不一致的问题。本文将该方法应用于金融领域、云计算领域和法律领域，在知识问答任务和金融领域股票价格预测任务上验证了该方法的有效性。实验结果表明，本文方法有助于避免语言模型在垂直领域上生成不符合事实的回复，在自动和人工评估上都优于现有方法。

2) 针对人类偏好对齐的问题，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信息，并利用大型语言模型的推理与分析能力进行用户问题优化，构成优化问题三元组数据集，用以训练单独的用户问题优化器，而无需训练对话模型，实现了与模型无关的、可解释的、效果稳定的人类偏好对齐。本文分别在两个不同的垂直领域基准测试集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，本文方法能够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，在用户偏好一致性和事实性方面都优于现有的方法。

**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 dialogue generation is an important basic technology in building intelligent dialogue system, which aims to solve problems in specific fields and provide professional answers for users. At present, the technology has been widely used in service assistants, question&answering robots and other applications for the fields of health care, finance, law, science, which has a high research value. At present, the mainstream vertical domain oriented dialogue generation methods are all designed based on retrieval augmented language models. However, the existing vertical domain oriented retrieval augmented dialogue generation has the following difficulties: 1) internal and external knowledge inconsistency problem, that is, the existing methods mainly focus on how to improve the accuracy of external knowledge recall, ignoring the distribution difference between the model's internal knowledge and it, thus affecting the role of external knowledge; 2) Human preference alignment problem, that is, there is a deviation between human user intent and model understanding, causing the model to generate answers that do not meet user expectations. In view of the above two problems, the research plan adopted in this paper is as follows:

1) Aiming at the problem of internal and external knowledge inconsistency, a retrieval augmented 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 information is used to build external knowledge base and supervised training dataset respectively. Finally, the knowledge aligned dataset is used to fine-tune the dialogue model, which effectively avoids the inconsistency between the internal knowledge injected into the model in the pre-training stage and the external knowledge in the inference stage. In this paper, the method is applied to financial analysis domain, and the validity of the method is verified on real-world stock trend prediction task and financial question&answering task. The experimental results show that the proposed method can avoid the language model to generate factually incorrect responses on vertical domain, and is superior to the existing methods in both automatic and manual evaluation.

2) Aiming at the problem of human preference alignment, a retrieval augmented dialogue

generation method based on human preference alignment is proposed. By collecting human preference information for real-world dialogue corpus, the method utilizes the reasoning and analysis capabilities of large language models to optimize user query, and constructs a triplet dataset of optimized queries to train a single user query optimizer instead of training the dialogue model, which achieves model-agnostic, interpretable and stable human preference alignment. In this paper, we compare with the current mainstream language model alignment methods on two different financial benchmarks, and prove that the method can improve the accuracy of knowledge document recall and the consistency of model understanding and user intent, and it is superior to the existing methods in terms of user preference consistency and fact.

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

# 目 录

# 插图目录

# 表格目录

# 第一章　绪论

## 1.1　研究背景和意义

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[1]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，垂直领域对话背景知识更丰富，实现稳定、可控、准确的对话生成的技术挑战性更高，使垂直领域检索增强对话生成存在以下几个难点：

- 难点 1：内外部知识不一致问题。模型在预训练期间学习到的内部知识为常识知识，因此当模型迁移到垂直领域上时的泛化能力有限。现有方法大多关注如何提升外部知识召回准确度，而忽略了模型内部知识与之存在的分布差异，从而影响外部知识发挥作用，导致模型难以准确回答垂直领域专业问题。
- 难点 2：人类偏好对齐问题。在垂直领域对话场景下，用户问题往往多样而复杂，模型难以从输入的用户问题中准确理解用户的真实意图，在语义检索和回答生成过程中存在数据分布偏差，导致模型生成不符合用户预期的回答。

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[2-4]。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[5-6]。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[7-8]。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

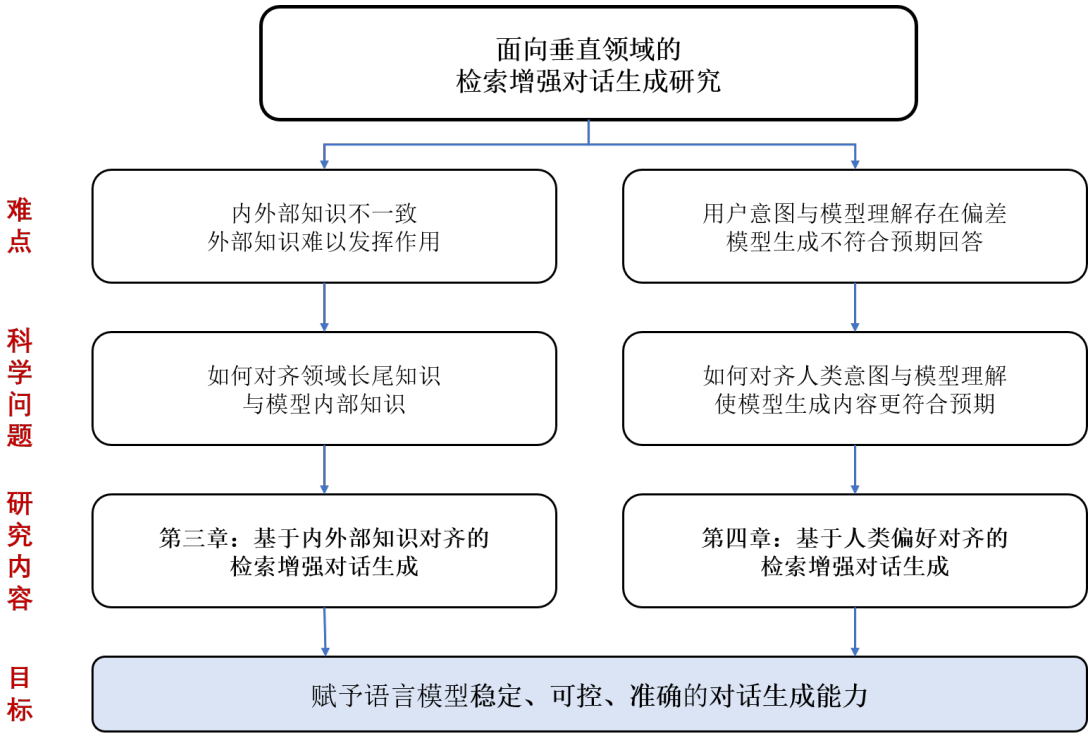


图 1-1 本文的研究思路与研究内容。

为解决上述挑战，本文主要研究在垂直领域场景下，如何利用对齐技术赋予语言模型稳定、可控、准确的对话生成能力，本文的研究思路与研究内容如图1-1所示。包括基于内外部知识对齐的检索增强对话生成和基于人类偏好对齐的检索增强对话生成。本文的具体研究内容如下：

1) 基于内外部知识对齐的检索增强对话生成

垂直领域知识背景复杂，语言模型在预训练中学习到的内部知识主要是常识知识，难以直接应用到垂直领域上做长尾问答，现有方法主要通过检索与用户问题相关的外部知识作为额外信息补充到模型输入中，然而，模型内部知识与外部知识之间存在较大分布差异，因此垂直领域外部知识难以发挥作用。针对上述问题，本文提出基于内外部知识对齐的检索增强对话生成方法，首先利用一个多粒度语义切分模块，从垂直领域知识文档中提取出文档级信息和实体级信息，并将提取出来的信息分别用于构建外部知识库和监督训练数据集，最后使用知识对齐后的数据集微调对话模型，有效避免了模型在预

训练阶段注入的内部知识与推理阶段获得的外部知识不一致的问题。另外，本文结合知识库文档数量大、专有名词多的特点，提出关键词检索与向量检索相结合的多级检索模块，提升知识文档的召回准确率，进而提升对话生成的准确性。本文将该方法应用于金融分析领域，在真实股票趋势预测任务和金融问答任务上验证了该方法的有效性。

2) 基于人类偏好对齐的检索增强对话生成

垂直领域对话场景下的用户问题往往多样而复杂，模型难以从输入的用户问题中准确理解用户的真实意图，在语义检索和回答生成过程中存在数据分布偏差，导致模型生成不符合用户预期的回答。另一方面，基于强化学习的人类偏好对齐算法训练成本高、稳定性差，使得对齐学习难度加深。针对上述问题，本文提出基于人类偏好对齐的检索增强对话生成方法，首先通过采集人类对真实场景对话样本的偏好信息，并利用大型语言模型的推理与分析能力进行用户问题优化，构成优化问题三元组数据集，用以训练单独的用户问题优化器，而无需训练对话模型，实现了与模型无关的、可解释、效果稳定的人类偏好对齐。本文分别在两个公开的金融基准测试集上与目前主流的语言模型对齐方法进行实验比较，证明了该方法的有效性。

## 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。

第二章是对话生成研究现状，概述了与对话生成相关的国内外工作研究现状，包括对话系统的发展及相关工作，对面向垂直领域的对话生成主流方法进行梳理，以及对基于大型语言模型的对话生成方法进行总结和概述。

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[9]。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（Pipeline-based Dialogue System）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。

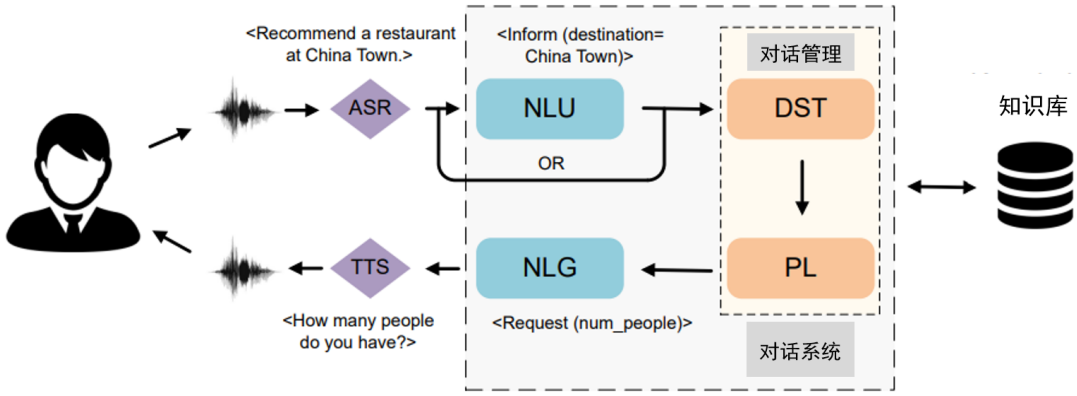


图 2-1 管道式对话系统结构示意图[10]。

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[11]和 Tur 等人[12]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。Sarikaya 等人[13]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[14]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[15]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[16]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[17]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[18]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[19]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[20]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[21]提出信念片段，即一种与槽名相对应的对话上下文片段，通过构建一个两阶段 CopyNet 网络来拷贝和存储对话历史中的槽值，并用以生成回答，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[22]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[23]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[24]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。Huang 等人[25]提出基于槽特征有限状态自动机的对话管理方法，通过树形的意图分层结构，实现多主题对话系统的主题检测与主题切换功能，并具有一定扩展性。

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[26]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[27]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[28]提出一种迭代矫正网络以迭代式修正所生成的标识符，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[29]，BERT[30]，T5[31]等序列到序列（Seq2Seq）[32-33]预训练模型的兴起，基于 Seq2Seq 模型的端到端对话系统（End-to-End Dialogue System）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[34]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[35]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[36]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[37]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[38]提出 Pointer Net，每个时间步的输出都来自输入序列，将标识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 仅从输入中选择标识符的局限性较大，Gu 等人[39]提出 CopyNet，在解码阶段的每个时间步中，决定复制输入中的标识符还是生成一个新的标识符，将复制概率和生成概率加和，得到最终的预测概率。Balakrishnan 等人[40]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[41]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标识符的生成，有效提升对话相关知识的检索精确度。Gao 等人[42]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[43]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[44]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[45]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[46]使用 GPT-2[47]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[48]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[49]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[50]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 面向垂直领域的对话生成

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。面向垂直领域的对话生成方法主要分为两类：（1）基于外部知识的方法；（2）基于知识注入的方法。

### 2.2.1 基于外部知识的垂直领域对话生成

Lin 等人[51]提出基于拷贝机制的知识对话方法 KIC，将循环知识交互解码器与知识感知的 Pointer Net 相结合，实现知识生成和知识拷贝，其算法框架如图2-2所示。Wu 等人[52]采用一个多类别分类器对生成的单词、生成的知识实体和拷贝的查询单词进行融合，提升了模型生成回答的准确性、一致性和知识性。Majumder 等人[53]同时进行角色信息选择和基于角色的响应生成，并使用强化学习训练对话智能体。Moon 等人[54]结合知识图谱与对话系统，将知识图谱信息作为外部知识，在强化学习框架中，智能体基于当前节点和状态选择相应的边，然后将知识组合到回答生成过程中。Xu 等人[55]将知识

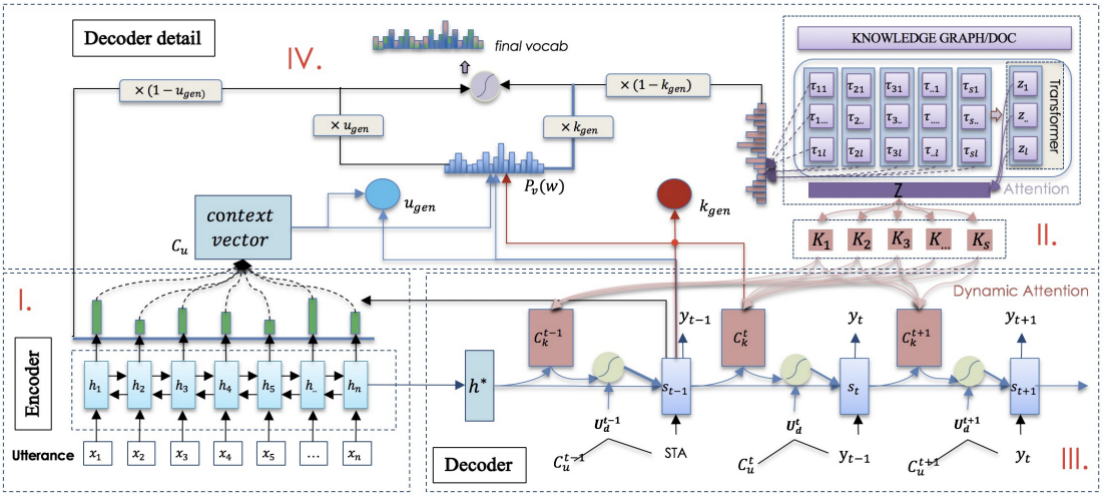


图 2-2 KIC 算法框架示意图[51]。

图谱作为外部知识，控制粗粒度的对话生成，对话具有常识性知识支撑，使得智能体能更好地引导对话话题。Jung 等人[8]提出双向图探索模型 AttnIO，用于在结构化知识图谱中检索外部知识，通过在遍历的每一步中计算注意力权重，使得模型能够选择更大范围的知识路径。Yang 等人[56]提出基于知识图谱的对话方法 GraphDialog，首先将对话历史解析为依赖关系树，并编码为嵌入向量，然后使用注意力机制对知识图谱进行多跳推理，最后由解码器从图谱中拷贝实体或预测单词，实现将垂直领域知识图谱中的外部知识整合到模型回复中。Lv 等人[57]提出嵌入知识语义的医疗领域对话系统，使用 BERT 融合 BiLSTM 的对话训练方法，将医疗知识图谱的语义信息融入对话系统。

总结：基于外部知识的垂直领域对话生成方法主要基于符号逻辑或知识图谱来实现外部知识的补充。然而，这类方法无法有效建模未见过的实体，且没有考虑互联网上大量的文本信息，导致泛化性和实时性不足。

### 2.2.2 基于知识注入的垂直领域对话生成

Wu 等人[59]提出 BloombergGPT 金融大模型，使用通用语料与金融领域语料混合的数据集，从头训练领域大模型，在模型学习基本语言语法和世界常识的同时，注入垂直领域知识。然而，从头开始预训练大模型需要收集大规模的通用语料，同时提高模型后续在语言层面的迁移难度。因此，较为主流的知识注入方式是在预训练通用模型的基础上进行继续预训练或监督微调。Zhang 等人[60]提出 XuanYuan 金融领域大模型在预训练语言模型 Bloom 的基础上使用混合数据继续预训练，混合数据包含通用领域数据和金融领域内的专业知识数据，并采用 Hybrid-Tuning 训练策略，在提升模型领域内能力的基础上，保证模型通用能力不会退化。Meng 等人[61]提出 DukeNet，构建了一个双向知

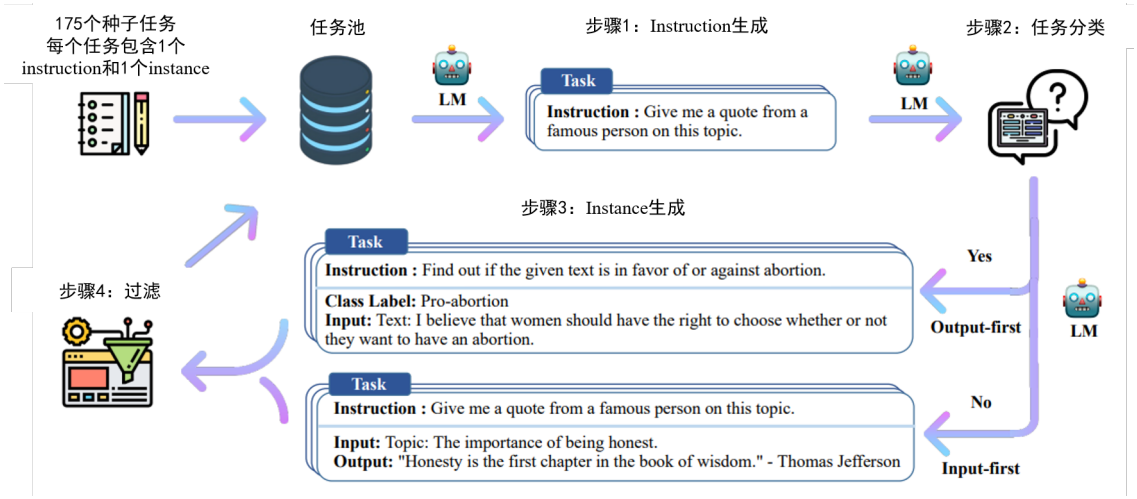


图 2-3 Self-Instruct 算法框架示意图[58]。

识交互网络，实现面向垂直领域知识的对话生成。Chen 等人[62]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。Zhou 等人[7]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。上述方法都需要构建大规模高质量领域内训练语料，对于垂直领域对话生成任务，往往还需要耗费人力，将其处理为“指令-回答”格式的对话数据，因此高质量的领域对话数据集十分稀缺。为解决数据稀缺的问题，Wang 等人[58]提出 Self-Instruct 方法，利用大型语言模型对小规模的种子样本进行数据扩充，以生成更多符合要求的微调数据，其算法框架如图2-3所示。Zhang 等人[63]提出 Self-QA 方法，直接基于非结构化文档数据生成指令数据，无需初始种子数据，进一步降低了数据扩充的人力成本。Wang 等人[64]提出 Self-KG 方法，基于中文医药知识图谱 CMeKG 生成指令数据，利用知识图谱中的节点关联信息，生成更高质量的指令数据。

总结：基于知识注入的垂直领域对话生成方法需要收集大规模高质量对话语料用于训练，同时知识被隐式存储在模型参数中，难以及时更新。

## 2.3 基于大型语言模型的对话生成

近年来，随着 ChatGPT[65]的发布，国内外先后出现许多预训练大型语言模型，如 ChatGLM[66]、文心一言[67]、通义千问[68]、LLaMA[69]、百川[70]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三类：（1）基于提示词工程的方法；（2）基于检索增强的方法；（3）基于模型微调的方法。

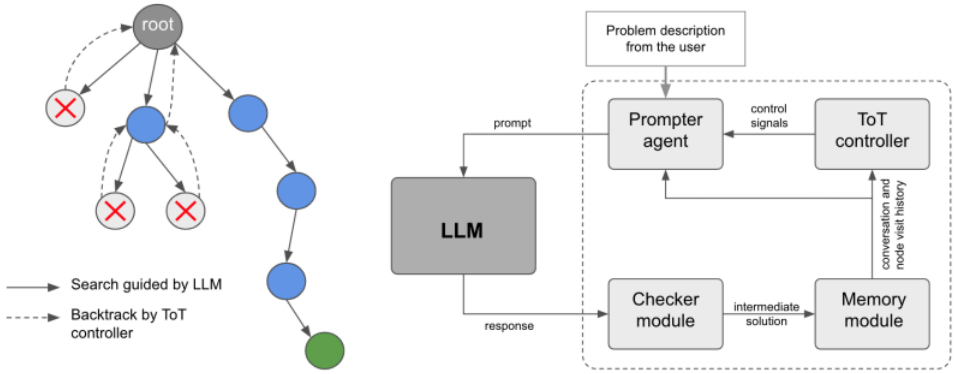


图 2-4 ToT 算法搜索策略与系统架构示意图[71]。

### 2.3.1 基于提示工程的对话生成

Radford 等人[47]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[29]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[72]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[73]引入自洽性（Self-Consistency）解码策略，通过采样生成多条推理链，然后采用投票方法从这些推理链中选出一致性最高的答案，在 GSM8K 复杂推理数据集上，Self-Consistency 相比于 CoT 方法有 17.9% 的提升，证明该方法能很好的提升 LLM 的复杂任务推理能力。Long 等人[71]提出思维树（Tree-of-Thought，ToT）提示方法，如图2-4所示，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[74]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，难以应对垂直领域对话。

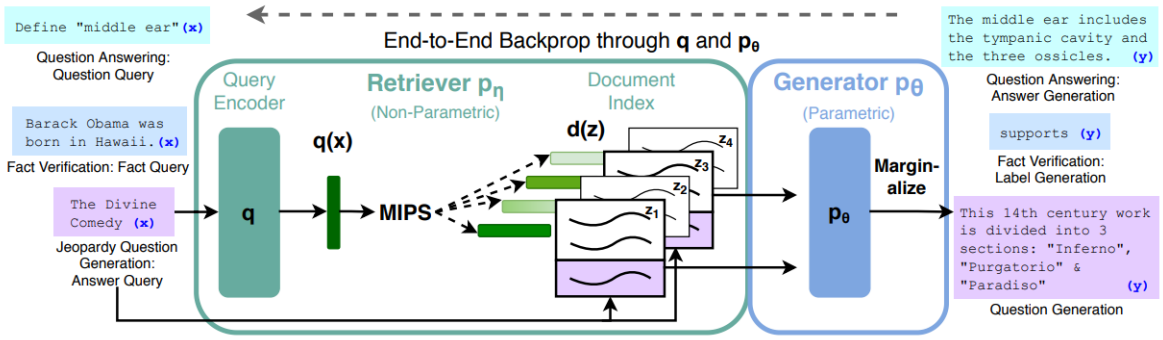


图 2-5 检索增强生成算法框架示意图[75]。

### 2.3.2 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[75]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，其算法框架如图2-5所示。Wang 等人[5]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[3]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[76]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[4]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[6]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。

### 2.3.3 基于模型微调的对话生成

根据研究目标的不同，基于模型微调的方法主要可分为两类：（1）参数高效的微调方法；（2）人类偏好对齐的微调方法。

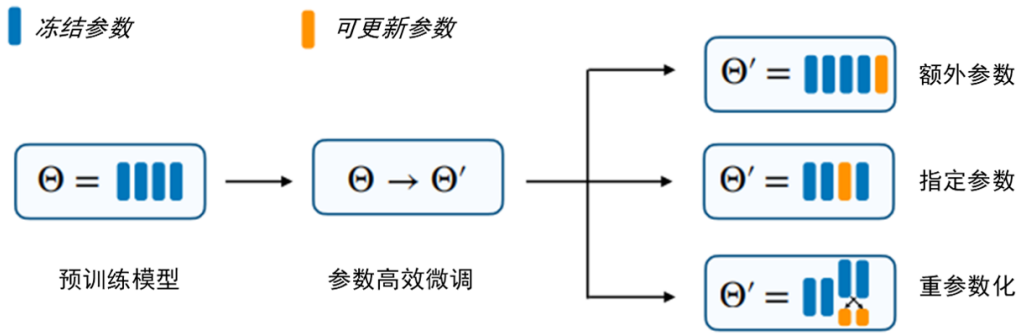


图 2-6 参数高效微调方法分类示意图[77]。

如图2-6所示，参数高效微调方法主要分为三类：额外参数型、指定参数型和重参数化型。Houlsby 等人[78]提出适配器微调（Adapter-Tuning）方法，在 Tranformer 层中插入小规模的神经网络模块，并在训练过程中冻结模型原有参数，仅更新这些适配器模块，实现与全量微调相近的性能。Mahabadi 等人[79]提出 Compacter 方法，通过结合超复数乘法和参数共享，将模型原有的线性层视为两个小矩阵的 Kronecker 积，实现在不损害模型性能的情况下显著减少适配器参数量。Pfeiffer 等人[80]提出 AdapterFusion 方法，先分别训练任务特定的适配器，再将不同任务的适配器融合起来，以利用跨任务知识来提升迁移学习的性能。除了在 Transformer 模型内部引入额外的神经网络模块外，也有研究人员在模型输入中引入额外的可训练上下文，实现参数高效微调。Li 等人[81]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。Liu 等人[82]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。Gu 等人[83]提出 PPT 方法，在模型与训练阶段插入软提示，以寻找更合适的参数初始化起点。Lee 等人[84]仅微调 BERT 和 RoBERTa 模型最后四分之一的中间层，冻结模型其他参数，达到了全量微调方法 90% 的性能。Zaken 等人[85]提出 BitFit 方法，仅对模型中的 bias 参数进行优化更新，并通过实验证明该方法在多个评测集上能够达到超过 95% 的性能。除了通过人工指定的方式确定需要更新的模型参数子集外，Zhao 等人[86]提出基于二维矩阵的掩码方法，通过学习掩码，自动选择需要更新的关键参数。Aghajanyan 等人[87]通过实验

发现，模型全参数微调的过程可以被重新参数化为一个低维子空间的优化过程。Li 等人[88]提出内在维度（intrinsic dimension）概念，用于衡量使模型达到预期性能所需要更新的最小参数量，并通过实验证明较低的内在维度重新参数化即可达到全量微调超过 85% 的性能，同时发现内在维度随模型尺寸增大而减小，且模型预训练过程能够隐式减小其内在维度。Hu 等人[89]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。Qiu 等人[90]进一步提出假设，认为多任务自适应存在一个通用的内在空间，因此能够将多任务学习过程重新参数化为一个低维内在子空间的优化过程。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。

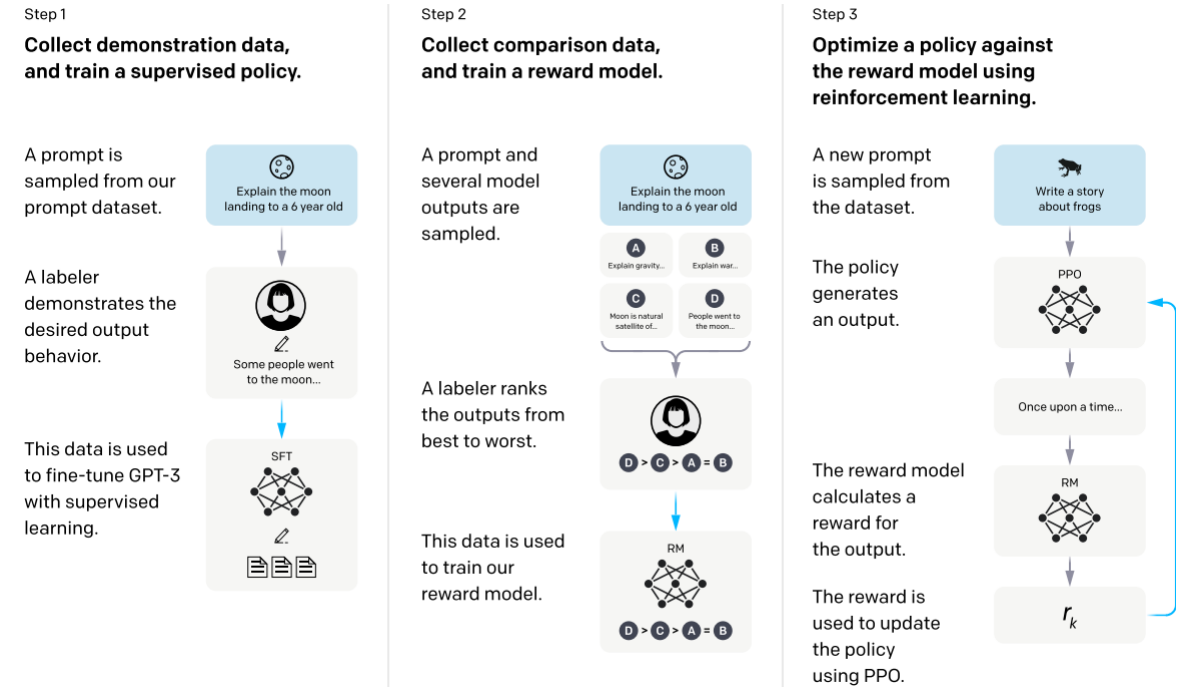


图 2-7 RLHF 算法框架示意图[65]。

Long 等人[65]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，其算法框架如图2-7所示。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[91]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[92]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[93]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[94]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[95]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[96]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[97]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[98]提出 RLAIF 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 RLAIF 能够达到与 RLHF 相当的性能。此外，Li 等人[99]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。Wei 等人[100]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。

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，难以及时跟进垂直领域的新知识。

## 2.4 本章小结

本章主要对研究相关的对话系统、面向垂直领域的对话生成以及基于大型语言模型的对话生成进行了介绍。总体而言，本章先介绍了对话系统在管道式对话系统和端到端式对话系统等方向的相关进展。然后，本章介绍了基于外部知识和基于知识注入的垂直领域对话生成方法的研究进展。最后，本章对基于大型语言模型的对话生成任务在基于提示工程、基于检索增强和基于模型微调方法的研究现状作了相关概述。

# 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

垂直领域对话生成的一个重要目标就是针对特定领域或行业为用户提供准确、专业、实用的问答服务，基于大型语言模型（LLM）的端到端式对话生成方法是目前实现这一目标的主流方法。LLM 在大规模语料上预训练时，广泛的常识知识被内化到模型内部的参数中，成为内部知识，从而使得 LLM 具有强大的文本理解和生成能力。然而，LLM 在生成训练数据之外的垂直领域对话内容时，会出现编造事实的现象，即“幻觉”现象。因此，使用检索增强生成（Retrieval Augmented Generation，RAG）[75]技术来提升垂直领域对话生成准确度的做法越来越受到学术界和工业界的关注。

近年来，大量研究[75,101]表明，与直接使用 LLM 生成对话相比，基于检索到的外部知识文档生成的对话内容具备更高的准确性和实时性。然而，基于 RAG 的垂直领域对话生成方法仍然存在一些局限性。具体而言，LLM 在预训练阶段没见过垂直领域的长尾知识，因此检索所补充的外部知识未能与模型内部知识完全对齐，而垂直领域对话涉及到广泛的领域背景知识，知识库检索得到的知识文档不足以覆盖回答用户问题所需的所有前置知识，导致模型无法给出正确的分析和解答。因此，通过对齐 LLM 内部和外部的知识以提升其生成对话的准确性和相关性是重要且必要的。

采用 RAG 方法构建外部知识库时，需要从互联网、领域信息平台等来源搜集大量垂直领域知识文档，这些知识文档包含丰富的垂直领域背景知识，可以为 LLM 内外部知识对齐提供长尾知识训练语料。本章受此启发，提出让 LLM 从知识库文档中学习垂直领域对话能力，从而解决现有方法的局限。然而，使用外部知识库中的知识文档构建训练数据集用以对齐模型内部知识，具有以下挑战：1）原始知识文档语料质量较低，不适用于对话生成任务；2）不同知识文档之间的知识复杂性不一样，学习难度不同，混合训练导致性能受限；3）推理过程中的检索结果准确率对模型生成效果存在较大影响。

为解决上述挑战，本章提出三点方法。为应对挑战 1），本章提出了一种基于多粒度语义切分的指令对生成方法。具体而言，利用基于提示工程的方法或人工标注的方式，从知识文档中分别提取文档级和实体级信息，用以构建数据同分布的外部知识库和监督训练数据集。为克服挑战 2），本章提出了一种基于两阶段微调的知识对齐方法。具体而言，将经过多粒度语义切分得到的文档块，按照其任务类别划分为两个数据集，首先在低难度的通用任务数据集上训练，然后在高难度的精标指令数据集上训练，以获得更好

的垂直领域对话性能。为解决挑战 3），本章提出了一种基于多级混合检索的文档块召回方法。具体而言，首先使用两类不同的检索方法分别对知识库文档块进行排序，然后对二者结果进行融合，最后再对结果进行重排序，以最大程度提升文档块召回准确率。

## 3.2 问题定义

面向垂直领域的检索增强对话生成任务要求对话系统根据给定的垂直领域用户问题，结合相关外部知识进行专业性的回复生成。将一个多轮对话视为两个对话者之间的诸多“问题-回复”对，给定在第 $t$ 轮对话时的用户问题 $Q_t$ 和对话历史 $H_t = [Q_0, R_0, \ldots, Q_{t-1}, R_{t-1}]$，在外部知识库中检索与 $Q_t$ 相关的知识文档 $d_k$，对话系统利用 $Q_t$、$H_t$ 和 $d_k$ 输出准确的分析与解答回复 $R_t$。

## 3.3 本章方法

### 3.3.1 方法总体框架

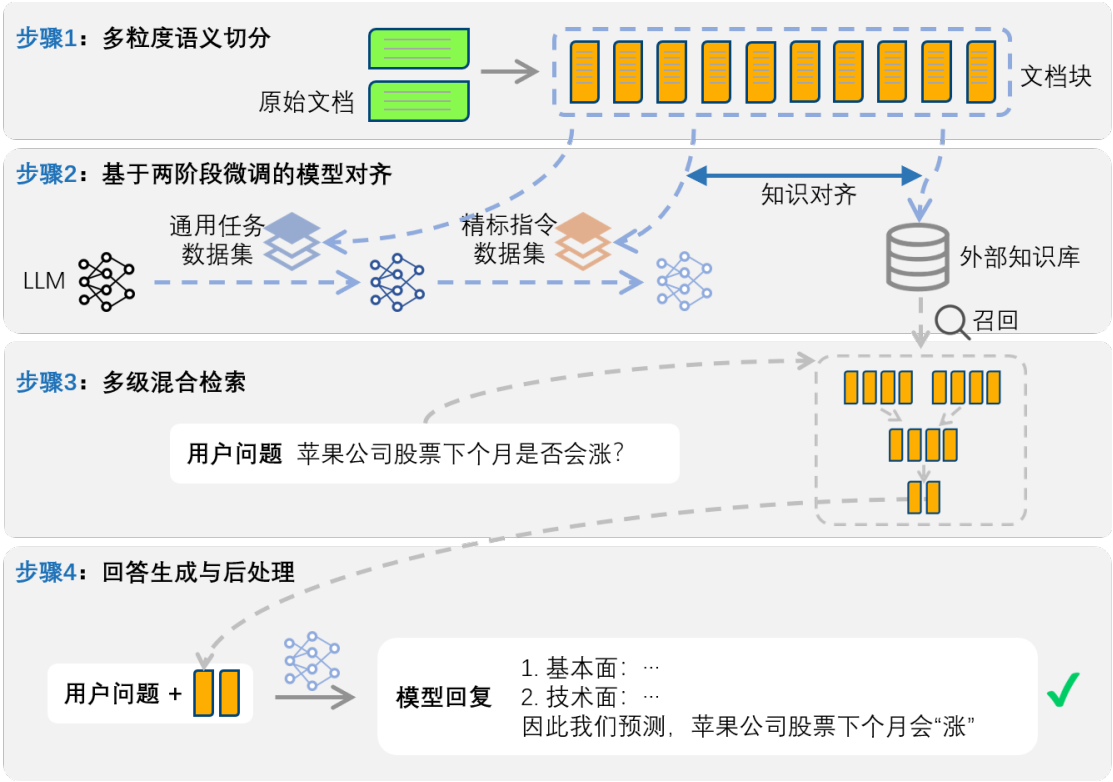


图 3-1 本章所提出的检索增强对话生成框架示意图。

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。整体方法框架如图3-1所示，整个框架主要包括三个部分：（1）对垂直领域知识文档进行多粒度语义切分，得到一系列垂直领域知识文档块，并构建外部知识库用于存储和检索这些知

识文档块；（2）构建垂直领域数据集，为模型注入垂直领域的内部知识，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，对用户问题和知识库中的文档块计算语义相似度，得到与用户问题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。

### 3.3.2 基于多粒度语义切分的指令对生成

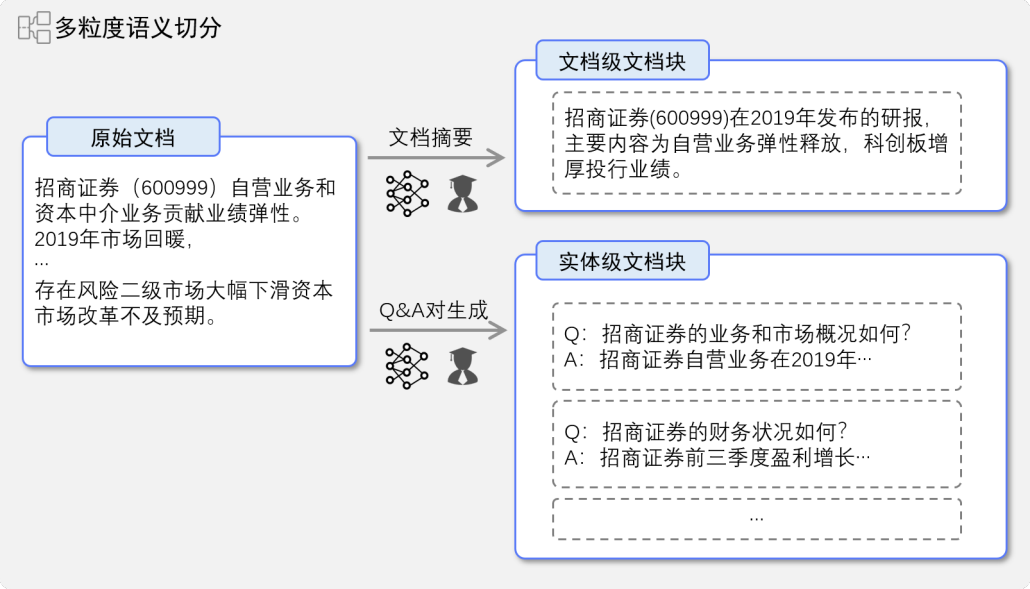


图 3-2 对话框架中的多粒度语义切分模块示意图。

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。外部知识是指显式存储在外部知识库中的知识，内部知识是指通过模型训练隐式存储在神经网络权重中的知识，二者的数据来源和数据分布存在差异，导致检索增强对话生成存在性能瓶颈。为了实现外部知识与模型内部知识的对齐，本节提出一种基于多粒度语义切分的指令对生成方法，从文档中提取出关键信息，构成指令对数据，用于外部知识库和监督微调数据集的构建。如图3-2所示，本章采用两种切分策略：粗粒度文档级总结和细粒度实体级 Q&A 对生成，以得到包含不同层级信息的文档块。文档级文档块是对原始文档的高度压缩，通过丢弃一部分细粒度信息，来对齐知识文档和用户问题之间的语义空间，从而提升外部知识库在信息密集型知识文档上的检索准确度。实体级文档块捕获原始文档中单个实体的特征和上下文信息，提升外部知识召回准确度，避免噪音信息对模型生成回答产生干扰。粗粒度文档级文档块和细粒度实体级文档块有效地提高了知识文档信息的一致性和互补性，并提升单个文档块内的信噪比，从而优化模型生成的回答质量。其中，语义切分过程由人类垂直领域专家编写，或使用 LLM（如

ChatGPT 模型）通过设计相应的提示词完成，本章所使用的语义切分提示词如表3-1所示。

表 3-1 知识文档语义切分提示词。

| 提示词 |
| --- |
| 基于 <content>，请提出多个 <domain> 领域的专业问题，并给出准确的回答。<br>输出格式如下：<br>问题：<question><br>回答：<answer> |

对于文档 $d_k$，其语义切分过程如下：

$$s_k = LLM_{sum}(d_k) \tag{3-1}$$

$$(q_{k0}, a_{k0}), (q_{k1}, a_{k1}), \cdots = LLM_{qa}(d_k) \tag{3-2}$$

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，以医疗领域为例，假设 $d_k$ 是与“核磁共振”相关的文档，$q_{k\_}$ 则可能是“如何根据核磁共振报告了解具体病情?”。$LLM_{sum}$ 表示用于文本摘要的 LLM，$LLM_{qa}$ 表示用于 Q&A 对生成的 LLM。

### 3.3.3 基于两阶段微调的知识对齐

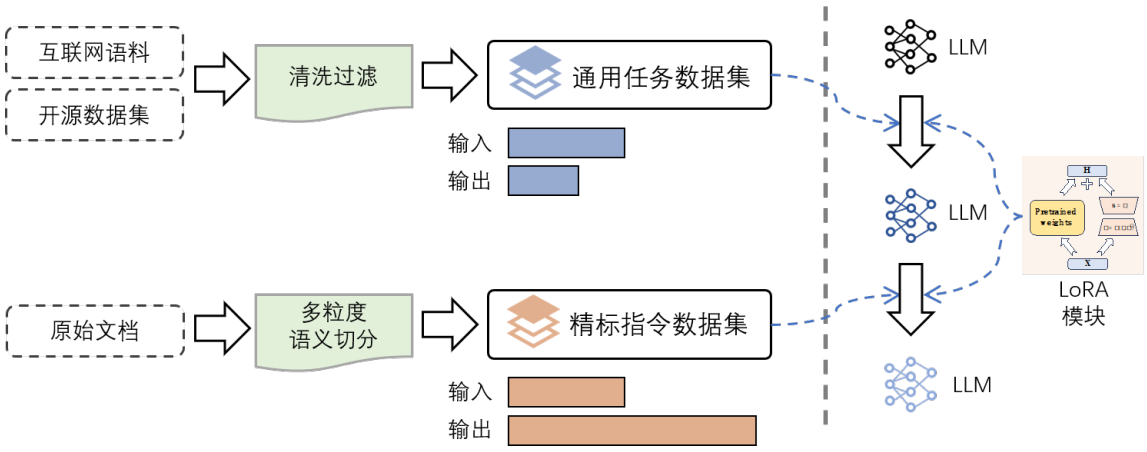


图 3-3 基于两阶段微调的知识对齐过程示意图。

垂直领域知识与模型预训练阶段的通用领域知识存在较大的数据分布差距，其本质是垂直领域 n-gram 词汇的差异和语料上下文的差异。由于领域数据分布差距的存在，

使用预训练 LLM 直接在垂直领域语料上微调不能达到最优的效果。因此本章使用两阶段微调策略，首先在较大数据量的垂直领域通用任务数据集上微调，引入垂直领域特定知识，适应垂直领域下游任务。其次，在较小数据量的高质量精标指令对数据集上微调，对垂直领域长尾数据的特征进行精准建模，提升模型对垂直领域知识的深度理解和推理能力。

本节构建了垂直领域监督微调数据集用于训练对话模型，以弥合内外部知识之间的分布差异，从而提升语言模型在垂直领域知识密集型对话任务上的性能。本节所构建的监督微调数据集包含两个部分：1）通用任务数据集，由较低难度的通用任务指令对数据构成，通过在互联网上搜集的垂直领域相关语料或开源数据集，经过清洗过滤后得到，包含情感分析、命名实体识别、文本分类、文本摘要等传统 NLP 任务，样本输出长度在 100 个标识符以内，用于增强模型在垂直领域的理解能力，缩小通用领域模型内部知识与垂直领域的差距，使得模型能更好地对齐垂直领域外部知识；2）精标指令数据集，由较高难度的逻辑推理、知识问答指令对数据构成，数据来源于多粒度语义切分后的垂直领域知识文档块，样本输出长度在 100 个标识符以上，用于对齐模型内部知识与外部知识的数据分布，使模型在垂直领域对话生成上达到最优效果。

模型微调过程分为两个阶段：1）首先在通用任务数据集上进行指令微调，目的是将专业领域的基本知识注入模型内部；2）然后，在上一步微调后的模型基础上进行继续训练，以进一步对齐知识文档中的长尾知识与模型内部的知识，同时使模型学会复杂任务的输出格式。

本章所使用的模型参数量为 6B，该数量级大小的预训练语言模型在下游任务上具有较小的内在秩，即能够用低维向量表示涵盖其解空间。同时，Prefix-Tuning 等基于额外参数的模型微调方法的性能受限于更新参数量，存在性能瓶颈。因此，本章所有的微调过程采用 LoRA-Tuning 方法，通过对模型 Attention 模块权重低秩分解，在低维子空间中进行优化，以缓解训练过拟合问题，在适应垂直领域任务的同时保持通用领域上的泛化能力，同时减少显存占用，降低训练成本。微调后的模型将作为对话框架中的对话生成模型，与用户交互。模型微调所使用的损失函数如下：

$$\mathcal{L} = -\frac{1}{N}\sum_{t=1}^{N}\log \pi_\theta\left(y_t \mid x, y_{<t}\right) \tag{3-3}$$

其中，$\pi_\theta$ 表示被训练的对话模型，$x$ 表示样本输入，$y_t$ 表示样本输出中的第 $t$ 个标识符。

### 3.3.4 基于多级混合检索的文档块召回

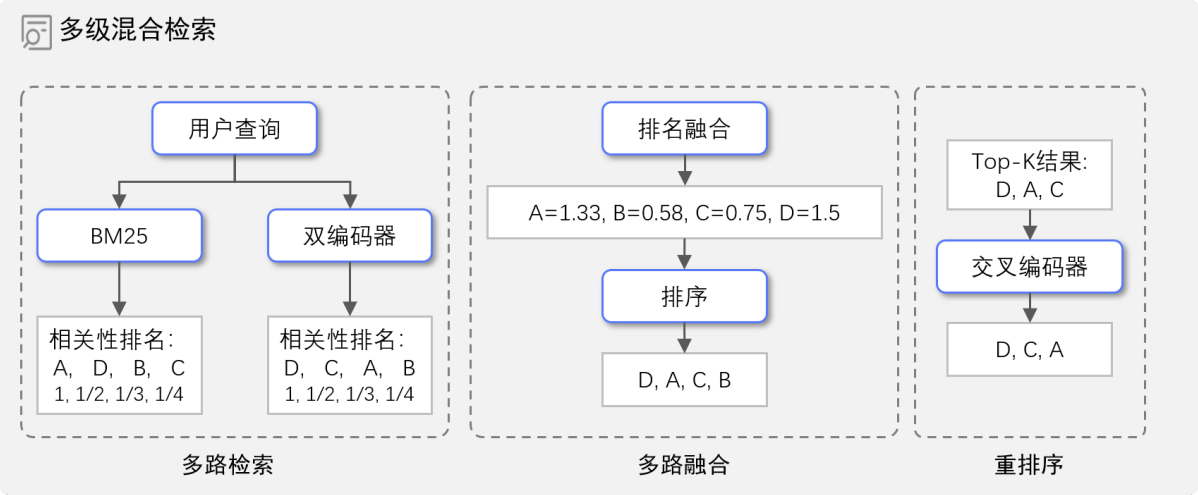


图 3-4 对话框架中的多级混合检索模块示意图。

如图3-4所示，多级混合检索模块主要包含三个步骤：（1）多路检索；（2）多路融合；（3）重排序。本节以文档级摘要切分策略为例，说明多级混合检索的具体计算过程。其中，对于细粒度实体级对话生成的切分策略，下式中的 $s_k$ 和 $d_k$ 可分别被替换为 $q_{k\_}$ 和 $a_{k\_}$。

**多路检索**本节使用稀疏向量检索和稠密向量检索两种方式。BM25 算法是常用的稀疏向量检索算法，通过词频逆文档频率（TF-IDF）计算用户问题和知识文档的词汇重叠率，从而得到二者的相似度得分。相较于稠密向量检索，BM25 算法能够较好地从通用实体泛化到垂直领域罕见实体，缓解长尾效应。同时，BM25 算法能够考虑到词在文档中的分布情况，更好的捕捉到词的上下文信息。给定用户问题 $Q$ 和文档摘要 $s_k$，其 BM25 相似度评分计算过程如下：

$$Score_{bm25}(Q,s_k)=\sum_{i}^{N} IDF(q_i,s_k)\frac{f_i(k_1+1)}{f_i+k_1(1-b+b\frac{sl}{avg(sl)})} \tag{3-4}$$

$$IDF(q_i,s_k)=log\frac{sl-n_i+0.5}{n_i+0.5} \tag{3-5}$$

其中，$N$ 表示 $Q$ 中的单词数量，$q_i$ 表示 $Q$ 中的第 $i$ 个单词，$k_1$、$k_2$ 和 $b$ 是调节因子，本节分别设置为 2、0、0.75。$f_i$ 表示单词 $q_i$ 在文档 $s_k$ 中出现的频率，$n_i$ 表示 $q_i$ 在 $s_k$ 出现的次数，$sl$ 为文档 $s_k$ 的长度，$avg(sl)$ 为知识库中所有文档的平均长度。基于评分 $Score_{bm25}$ 对所有文档进行排序，可得到文档排列 $R_{bm25}$。

然而，BM25 存在数据稀疏问题，导致文档中的重要信息被忽略。基于双编码器

（Bi-Encoder）的稠密向量检索方法能将文本映射为词嵌入向量，对目标词及其上下文进行建模，更好的捕捉到文本的语义信息，使得语义向量能够更好地表示用户问题的意图，获得更准确和全面的检索结果。因此本章同时还使用了基于双编码器的稠密检索方法，与稀疏检索互为补充。

给定用户问题 $Q$ 和文档摘要 $s_k$，分别通过双编码器模型得到对应的嵌入向量 $e_Q$ 和 $e_{sk}$。所有文档的嵌入向量 $e_{sk}$ 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。

$$e_{sk} = BE(s_k) \tag{3-6}$$

$$e_Q = BE(Q) \tag{3-7}$$

其中，$BE$ 是双编码器模型，如 BGE[102]、SGPT[103]等。基于评分 $Score_{emb}$ 对所有文档进行排序，可得到文档排列 $R_{emb}$。

**多路融合**由于不同检索方式使用的相关度评分存在量纲差异，因此不宜直接对多路检索评分结果求和。本节使用排名融合方法，对文档排列 $R_{bm25}$ 和 $R_{emb}$ 进行合并。文档摘要 $s_k$ 的排名融合评分计算过程如下：

$$Score_{RRF} = \sum_{r \in \{R_{bm25}, R_{emb}\}} \frac{1}{k + r(s_k)} \tag{3-8}$$

其中，$k$ 为超参数，本节设置为 60。基于排名融合评分重新排序，得到文档排列 $R_{RRF}$。

**重排序**交叉编码器计算速度慢，且无法预先计算文本嵌入向量，只能实时计算文本对的语义相似度分数，但是其准确率优于双编码器，因此适用于对双编码器结果的检索结果进行二次精排。本节选取多路融合结果的 Top-K 文档排列，分别与用户问题一起输入交叉编码器，得到语义相似度分数，选择相似度分数最高的文档作为最终的结果 $d^*$。

$$d^* = \arg\max_{d_k} CE(s_k), \quad s_k \in topk(R_{RRF}) \tag{3-9}$$

其中，$CE$ 表示交叉编码器模型，$topk(\cdot)$ 表示取排列中的前 k 个元素。

### 3.3.5 回答生成

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，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 $R_t$。

$$I_t = concat(Prompt, d^*, H_t, Q_t) \tag{3-10}$$

$$R_t = LLM(I_t) \tag{3-11}$$

## 3.4 实验结果

本节分别以金融领域、云计算领域、法律领域等垂直领域为应用场景对方法进行实验验证。本实验主要设置两类任务：1）知识问答任务；2）金融领域股票趋势预测任务。对于第一类任务，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。对于第二类任务，本节在金融领域进行模拟投资，以进一步验证本章方法在真实场景中的实用价值，该任务主要考察模型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。

### 3.4.1 数据集介绍

为说明本章方法在不同垂直领域下的有效性，本节在三个基准测试集上进行实验，即 AlphaFin-test、Aliyun-Docs 和 CAIL2021[104]。三个数据集都包含垂直领域对话数据。对于知识问答任务，由于本实验使用人工和 GPT-4 模型进行评估，评估成本较高，因此本实验对所有基准测试集均随机抽取 1000 个样本进行实验。对于金融领域股票趋势预测任务，本实验使用 AlphaFin-test 中的金融研报数据集进行实验。

- AlphaFin-test：AlphaFin-test 来自从开源金融数据接口库（如 Tushare[105]、AKshare[106]等）爬取的财经新闻、股票价格数据。AlphaFin-test 为中文数据集，包含金融新闻 & 问答数据集和金融研报数据集两个部分。金融新闻 & 问答数据集有 10000 个样本，每个样本包含系统指令、输入和输出字段，包含针对金融新闻、金融数据的单轮问答对话。金融研报数据集有 1000 个样本，涵盖 544 家上市公司在 2020 年 1 月到 2023 年 7 月的真实市场数据，每个样本包含系统指令、输入和输出三个字段，在输入字段中给出指定公司在指定日期的研报和当月股价数据，在系统指令中要求给出该公司股票次月涨跌分析及预测，输出字段为该股票的实际涨跌结果。
- Aliyun-Docs：Aliyun-Docs 来自从阿里云官网爬取的 3206 个云计算产品文档，经过 HTMl 源码清洗后得到 3005 个可用知识文档。同时，Aliyun-Docs 还包含阿里

云线上业务场景中采集到的 13995 个客户问题，所有数据均经过脱敏处理。

- CAIL2021：CAIL2021 阅读理解数据集包含 1500 个法律问答对，每个样本包括案例片段、案例名称、问答列表三个字段。案例片段为裁判文书网公开的裁判文书，问答列表为多个针对案例片段内容的“问题-回答”对。

### 3.4.2 基准方法

对于知识问答任务，本节使用语言模型作为基准方法，以对比本章方法相较于其他模型的性能提升。基准方法具体介绍如下。

- LLM：包括 ChatGLM2-6B[66]和 ChatGPT[65]，在通用任务上进行监督微调，具有对话能力的 LLM。
- 垂直领域 LLM：包括 FinMA[107]、FinGPT[108]和通义金融[68]，在金融领域数据集上微调后的 LLM，能够处理广泛的金融任务。

对于金融领域股票趋势预测任务，除了上述语言模型外，本节还使用股票指数和机器学习模型作为基准方法，具体介绍如下。

- 股票指数：本节选取了中国股票市场的重要指数，包括上证 50、沪深 300、上证指数和创业板指数。这类指数反映了某个交易板块中所有上市公司股票的整体涨跌情况，即市场综合水平。
- 机器学习模型：包括随机森林[109]、RNN[110]、BERT[30]、GRU[33]、LSTM[111]、逻辑回归[112]、XGBoost[113]、决策树[114]等适用于时序数据预测的机器学习模型。

### 3.4.3 评价指标

对于知识问答任务，本节使用 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，本节还使用人工与 GPT-4 作为评判员，对模型回复进行两两配对评估。同时，本节还使用 Ragas[115]检索增强评估框架中的 Context Precision、Context Recall、Faithfulness 三项指标对本章方法进行评估，以定量分析所召回知识文档的相关性和模型回复的幻觉程度。指标具体介绍如下。

- ROUGE：机器翻译、文本摘要等自然语言处理任务中的常用评估指标，基于模型生成的候选值和参考答案计算其 N-gram 召回率。其中，ROUGE-L 计算的是二者的最长公共子序列长度。
- 人类偏好评价：由人类担任评判员，从两个不同模型的回复中选择更优的回答，考察角度包括回答的有帮助性、相关性、准确性、深度、创造性和详细程度。

- GPT-4 偏好评价：由 GPT-4[116]模型担任评判员，从两个不同模型的回复中选择更优的回答，考察角度与人类偏好评价保持一致，并在系统指令中对 GPT-4 模型进行约束，指令提示词格式参考 MT-Bench[117]评估框架。
- 上下文准确率（Context Precision）：利用 LLM（如 GPT-4 模型）评估知识文档块与用户问题之间的相关性及文档块排名顺序。
- 上下文召回率（Context Recall）：利用 LLM 估计模型回答和文档块的 TP 和 FN，计算文档块的召回率，即衡量模型回答对知识文档块的引用比例。
- 可信度（Faithfulness）：利用 LLM 计算 (用户问题, 模型回答, 文档块) 三元组的自然语言推断（NLI）分数，即对模型回答的事实性进行量化评估。

对于金融领域股票趋势预测任务，本节使用两类指标。第一类是核心指标，包括衡量收益能力的年化收益率（ARR）和准确率（ACC）。第二类是辅助分析的观察指标，如最大回撤（MD），卡玛比率（CR），夏普比率（SR），用于评估投资组合风险情况。具体介绍如下。

- 年化收益率（ARR）：将当前收益率按照复利计算（即每年收益计入本金），折算成平均年度收益率。其计算公式为：$ARR = (Return/Principle)^{1/n} - 1$，其中 $n$ 为投资年数，$Return$ 为总收益，$Principle$ 为本金。
- 准确率（ACC）：当模型对股票涨跌预测结果与次月实际涨跌趋势一致时，可认为模型预测正确（ACC=1），否则预测错误（ACC=0）。
- 年化超额收益率（AERR）：投资组合的 ARR 超过基准 ARR 的部分，即 $AERR = ARR - ARR_b$，本实验中 $ARR_b$ 取沪深 300 的 ARR。
- 年化波动（ANVOL）：用于衡量投资组合的波动风险，其计算公式为：$ANVOL = \sigma_P * \sqrt{n}$，其中 $\sigma_P$ 为收益率标准差，$n$ 为投资年数。
- 最大回撤（MD）：在选定周期内任一时间点开始，产品净值下降到最低点时的收益率最大回撤幅度，用于表现投资组合可能出现的最糟糕情况。其计算公式为：$MD = max((D_i - D_j)/D_i)$，$D_i$ 表示第 $i$ 天的资产净值，且 $i < j$。
- 卡玛比率（CR）：表示投资组合收益和最大回撤之间的关系，代表每单位回撤能获得的收益率。其计算公式为：$CR = ARR/MD$。
- 夏普比率（SR）：投资组合每承受一单位总风险，所产生的超额收益。其计算公式为：$SharpRatio = (E(ARR) - R_f)/\sigma_P$，其中 $E(ARR)$ 表示投资组合的 ARR 期望，$R_f$ 为年化无风险利率，$\sigma_P$ 为投资组合收益率的标准差。

- 最大下潜期（MDD）：描述持有价值从回撤开始到再创新高所经历的时间，该指标可以反映资产创新高的频率。

### 3.4.4 实验细节

表 3-2 实验环境配置参数。

| 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<br>transformers 4.33.0 |

本节实验中所有语言模型的推理解码策略均为贪心搜索，以达到最稳定的性能。另外，在所有模型训练过程中，所使用的超参数如下：Batch Size=16，学习率调度器为CosineLRScheduler，Learning Rate=5e-5，Precision=bf16，其余硬件和软件环境如表3-2所示。模型首先在通用任务数据集上训练 20 个 epoch，然后再继续在精标指令数据集上进行 2 个 epoch 的增量微调。本实验中所有微调均使用 LoRA 方法，对 Transformer 模型的 Attention 模块中的 query_key_value 线性层增加 LoRA 模块，在训练过程中冻结模型原有的参数，仅更新 LoRA 模块中的参数。其中，LoRA Rank=8，LoRA Alpha=16。

同时，对于金融领域股票趋势预测任务，为观察模型在真实市场中的模拟投资表现，本节采用如下处理方法进行投资策略生成：

给定输入 $I_i$，通过 LLM 得到关于 $c_i$ 的回复文本 $Res_i$。然后，使用基于规则的方法从 $Res_i$ 中提取出趋势预测结果 $Pred_i$，选择所有被预测为“上涨”的股票，得到股票集合 $C_{chosen}$。最后，本节按月滚动执行该投资策略。即，在每个月月初买入或继续持有 $C_{chosen}$ 中的所有股票 $c_i$，持有时间为一个月。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。

$$AR_m = AR_{m-1} + \sum_{c_i \in C_{chosen}} \omega_{c_i} R_{c_i} \tag{3-12}$$

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。

$$\omega_{c_i} = \frac{v_i}{\sum_{c_n \in C_{chosen}} v_n} \tag{3-13}$$

### 3.4.5 与现有方法的性能比较

表 3-3 人工对模型回复的偏好评价结果。

| 数据集 | 模型 | Win | Tie | Lose | ΔWR |
|---|---|---|---|---|---|
| AlphaFin-test | 本章方法 v.s. FinMA | 85% | 14% | 1% | +84% |
| | 本章方法 v.s. ChatGLM | 60% | 25% | 15% | +45% |
| | 本章方法 v.s. FinGPT | 57% | 24% | 19% | +38% |
| | 本章方法 v.s. ChatGPT | 53% | 25% | 22% | +31% |
| Aliyun-Docs | 本章方法 v.s. ChatGLM | 78% | 13% | 9% | +69% |
| | 本章方法 v.s. ChatGPT | 61% | 18% | 21% | +40% |
| CAIL2021 | 本章方法 v.s. ChatGLM | 41% | 33% | 26% | +15% |
| | 本章方法 v.s. ChatGPT | 35% | 34% | 31% | +4% |

表 3-4 GPT-4 模型对模型回复的偏好评价结果。

| 数据集 | 模型 | Win | Tie | Lose | ΔWR |
|---|---|---|---|---|---|
| AlphaFin-test | 本章方法 v.s. FinMA | 95% | 4% | 1% | +94% |
| | 本章方法 v.s. ChatGLM | 73% | 3% | 24% | +49% |
| | 本章方法 v.s. FinGPT | 72% | 2% | 26% | +46% |
| | 本章方法 v.s. ChatGPT | 58% | 6% | 36% | +22% |
| Aliyun-Docs | 本章方法 v.s. ChatGLM | 85% | 2% | 13% | +72% |
| | 本章方法 v.s. ChatGPT | 66% | 5% | 29% | +37% |
| CAIL2021 | 本章方法 v.s. ChatGLM | 51% | 6% | 43% | +8% |
| | 本章方法 v.s. ChatGPT | 48% | 9% | 43% | +5% |

**知识问答任务**本节使用人类和 GPT-4 作为评判员，对每个 LLM 在测试数据集上的回复效果进行评分。人工评判结果如表3-3所示，在三个不同垂直领域的数据集上，本

章所提出的方法在内容有效性方面均优于其他 LLM。其中，本章方法相比于未经知识对齐的 ChatGLM 模型在三个数据集上分别取得 45%、69% 和 15% 的胜率，说明本章方法在不同领域上均能提升模型知识问答的回复质量。同时，FinMA 和 FinGPT 是在其他金融领域数据集上微调的 LLM，本章方法在 AlphaFin-test 上也分别取得 84% 和 38% 的胜率，这表明基于内外部知识对齐的微调方法能够提升模型对于垂直领域外部知识的理解能力，从而生成更准确、全面的回答。

表 3-5 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。 |

GPT-4 模型的偏好评价结果如表3-4所示，在所有数据集上，本章方法均取得优于其他基线方法的性能，GPT-4 模型对不同方法的偏好评价与人工评判结果具有较高的一致性。GPT-4 给倾向于给出明确的胜负结果，而“Tie”结果的比例显著低于人工评判结果，因此本章方法得到更高的胜率提升 ΔWR。值得注意的是，本章方法在 CAIL2021 数据集上相比于 ChatGLM 和 ChatGPT 的胜率提升均在 10% 以下，这是因为 CAIL2021 数据集的问题答案均可在上下文片段中找到，对垂直领域内其他背景知识的依赖性不强，因此性能提升不显著。

另外，本节以金融领域为例，展示基线方法与本章方法的模型回复结果。从表3-5中可以看出，即使通过 RAG 方法提供了与用户问题中的“万科 A”股票相关的研报和市场数据作为外部知识文档，FinMA、ChatGLM、FinGPT 和 ChatGPT 模型仍然无法很好理解外部知识，并将其用于股票未来涨跌预测分析，从而做出了拒绝回答的决策，而通义金融模型则给出了不具有有效性、事实性的回复。从表3-6中可以看出，本章方法通过内外部知识对齐，使得模型具有良好的金融领域理解和分析能力，并学习到专业的输出范式，因此能够输出准确、全面、详实的分析过程，并最终给出明确的涨跌预测结果。

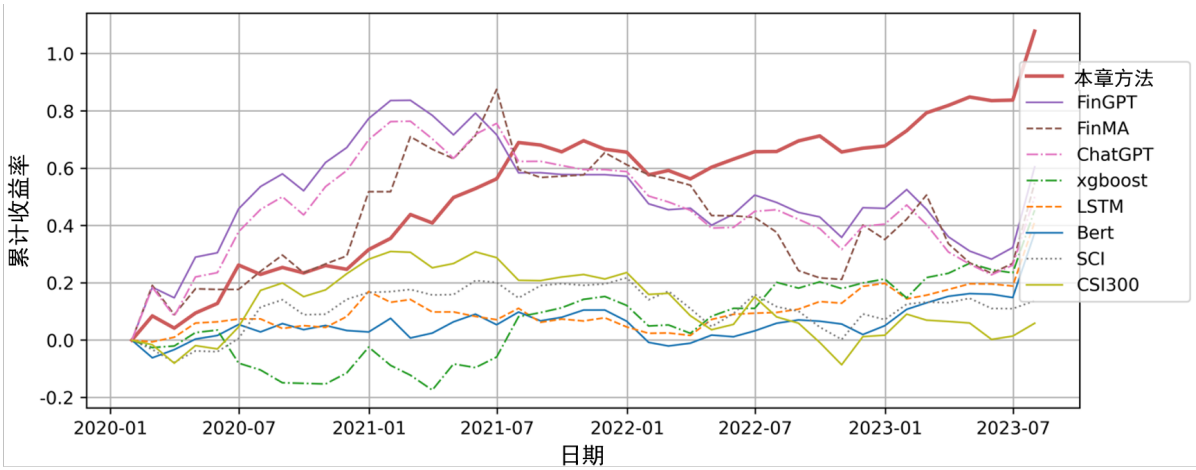


图 3-5 不同方法在股票趋势预测任务上的累计收益情况。

**金融领域股票趋势预测任务**如图3-5所示，曲线表示每种方法的年化收益。值得注意的是，从 2023 年开始，本章方法的收益达到最高并保持上升趋势。这表明了本章方法在投资中的有效性和可用性。同时，由表3-7可以得出以下结论：

首先，RNN、LSTM 等机器学习、时序预测模型在股票趋势预测方面具有一定的预测能力，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，总体上超过了机器学习、时序预测模

表 3-6 本章方法回复示例。

| | |
|---|---|
| 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。 |
| 本章方法回复 | 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”。 |

表 3-7 不同方法在股票趋势预测任务上的具体指标。

| 模型 | ARR ↑ | AERR ↑ | ANVOL ↓ | SR ↑ | MD ↓ | CR ↑ | MDD ↓ | ACC ↑ |
|---|---|---|---|---|---|---|---|---|
| 上证 50 | -1.0% | -2.7% | 19.3% | -0.054 | 45.9% | -0.023 | 29 | - |
| 沪深 300 | 1.7% | 0% | 18.2% | 0.092 | 39.5% | 0.043 | 30 | - |
| 上证指数 | 3.9% | 2.2% | 14.8% | 0.266 | 21.5% | 0.183 | 19 | - |
| 创业板指数 | 7.6% | 5.9% | 26.5% | 0.287 | 41.3% | 0.185 | 20 | - |
| 随机森林 | 9.8% | 8.1% | 19.5% | 0.501 | 16% | 0.608 | 22 | 55.5% |
| RNN | 8.1% | 6.4% | 10.9% | 0.742 | 15.7% | 0.515 | 12 | 54.1% |
| BERT | 10.7% | 9.0% | 16.1% | 0.664 | 13.5% | 0.852 | 14 | 51.4% |
| GRU | 11.2% | 9.5% | 13.7% | 0.814 | 14.6% | 0.765 | 21 | 54.7% |
| LSTM | 11.8% | 10.1% | 15.4% | 0.767 | 15.3% | 0.768 | 19 | 55.2% |
| 逻辑回归 | 12.5% | 10.8% | 27.1% | 0.463 | 32.5% | 0.385 | 18 | 54.8% |
| XGBoost | 13.1% | 11.4% | 20.5% | 0.633 | 20.9% | 0.619 | 17 | **55.9%** |
| 决策树 | 13.4% | 11.7% | 19.6% | 0.683 | **11.9%** | 1.126 | 20 | 55.1% |
| ChatGLM | 8.1% | 6.4% | 24.9% | 0.324 | 62.6% | 0.126 | 26 | 49.5% |
| ChatGPT | 14.3% | 12.6% | 27.7% | 0.516 | 53.6% | 0.267 | 23 | 51.4% |
| FinMA | 15.7% | 14.0% | 37.1% | 0.422 | 66.3% | 0.236 | 25 | 49.1% |
| FinGPT | 17.5% | 15.8% | 28.9% | 0.605 | 55.5% | 0.312 | 24 | 50.5% |
| 本章方法 | **30.8%** | **29.1%** | **19.6%** | **1.573** | 13.3% | **2.314** | **10** | 55.7% |

型等方法，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，通过在金融领域数据集上的微调，FinMA、FinGPT 等金融领域大模型具有更强的股票趋势预测能力。FinMA 和 FinGPT 模型的 ARR 分别达到 15.7% 和 17.5%。

最后，基于内外部知识对齐对模型进行微调后，本章方法实现了 30.8% 的 ARR 和 55.7% 的 ACC，说明内外部知识对齐对 LLM 能力提升起着至关重要的作用。

### 3.4.6 多级混合检索的有效性

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、增加多路融合方法、增加交叉编码器重排序后，模型在 AlphaFin-test 数据集上的回复效果。从表3-8可以看出，增加多路融合方法后，各项指标均有小幅度的提升，增加重

表 3-8 探究多级检索模块在 Ragas 评估指标上对性能的影响。

| 方法 | Precision ↑ | Recall ↑ | Faithfulness ↑ |
|---|---|---|---|
| 向量检索 | 0.6028 | 0.8195 | 0.7412 |
| + 多路融合 | 0.6189 | 0.8324 | 0.7691 |
| + 重排序 | **0.6717** | **0.8430** | **0.8005** |

排序后，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3-9所示。

表 3-9 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；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:<br>研报中提到, 青松建化在 2022 年有望迎来 4 重弹性，...<br>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。 |

### 3.4.7 两阶段微调对性能的影响

首先，本实验研究了在精标指令数据集上微调 LLM 后，LLM 回复内容与检索到的知识文档之间的 ROUGE 值。根据表3-10可以观察到，在通用任务数据集上训练后，模型在 ROUGE-1 和 ROUGE-L 上均有提升，在 ROUGE-2 上略有下降。在精标指令数据集上微调后，模型在三类 ROUGE 指标上均有所提升，且模型输出长度有所提升，本章

表 3-10 金融问答 ROUGE 指标下不同数据集对性能的影响。

| 数据集 | 模型 | ROUGE-1↑ | ROUGE-2↑ | ROUGE-L↑ |
|---|---|---|---|---|
| AlphaFin-test | ChatGLM | 0.2578 | 0.1960 | 0.2542 |
| | + 通用任务数据集 | 0.2604 | 0.1895 | 0.2688 |
| | + 精标指令数据集 | 0.4052 | 0.2956 | 0.4031 |
| CAIL2021 | ChatGLM | 0.6854 | 0.5108 | 0.6003 |
| | + 通用任务数据集 | 0.7015 | 0.5544 | 0.6159 |
| | + 精标指令数据集 | 0.7609 | 0.6031 | 0.7456 |

方法在三类 ROUGE 指标上分别达到了 0.4052、0.2956 和 0.4031，达到最优性能，这主要是因为精标指令数据集提升了模型对知识文档的理解能力和摘要能力，因此输出内容与知识文档的一致性更高。

表 3-11 股票涨跌预测指标下不同数据集对性能的影响。

| 数据集 | 模型 | ARR↑ | SR↑ | 输出长度↑ | 无效答案率↓ |
|---|---|---|---|---|---|
| AlphaFin-test | ChatGLM | 8.1% | 0.324 | 228.1 | 52.3% |
| | + 通用任务数据集 | 15.8% | 0.636 | 17.2 | 0% |
| | + 精标指令数据集 | 30.8% | 1.573 | 254.8 | 25.9% |

其次，本实验探究了两阶段数据集微调对模型在股票涨跌预测上的性能的影响；由表3-11可知，相对于未经过训练的基座模型 ChatGLM，LLM 对股票价格的预测能力在使用通用任务数据集进行微调后有所提高，实现了 15.8% 的收益提升。此外，在通用任务数据集上进行微调后，该输出范式仅包含“涨”或“跌”，较为简单，因此 LLM 通过微调即可遵循该范式，无效回答率为 0%。经过精标指令数据集的微调，本章方法以 30.8% 的 ARR 达到最优性能，无效答案比例相较于原始 ChatGLM 模型的结果也有所下降，达到 25.9%。

### 3.4.8 不同 LoRA 秩对性能的影响

本节通过设置消融实验探究不同 LoRA 秩对模型性能的影响，测试数据集为 AlphaFin-test，实验结果如表3-12所示。随着 LoRA 秩的增大，模型训练所消耗的显存和训练时长有轻微的增加。同时，LoRA 秩在非常小的取值，如 rank=1 时，模型就已经

表 3-12 不同 LoRA 秩对性能的影响。

| LoRA 秩 | Recall | Faithfulness | 显存占用（MB） | 训练时长（小时） |
|---|---|---|---|---|
| 1 | 0.8397 | 0.7986 | 30748 | 56 |
| 2 | 0.8411 | 0.7893 | 30748 | 57 |
| 4 | 0.8508 | 0.8122 | 30790 | 57.5 |
| 8 | 0.8430 | 0.8005 | 30804 | 58 |
| 64 | 0.8584 | 0.8097 | 31290 | 60 |
| 128 | 0.8462 | 0.8144 | 32114 | 63 |

表现出良好的 Recall 和 Faithfulness 性能。然而，模型性能并未与 LoRA 秩表现出正相关关系，本文认为这是因为 ChatGLM2-6B 模型在该项任务上的内在秩较小，因此选择较小的 LoRA 秩即可覆盖关键子空间。本章基于经验主义选择 rank=8 作为超参数 LoRA 秩的取值。

## 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。本章方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，本章还提出了基于多粒度语义切分的指令对生成方法、基于两阶段微调的知识对齐方法和基于多级混合检索的文档块召回方法，能有效提升知识文档召回的准确率。本章提出的方法在金融领域、云计算领域、法律领域三个垂直领域上获得了超越其他现有方法的性能，同时在金融领域进一步进行股票趋势预测任务，同样展现出优于基线方法的性能。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

垂直领域对话场景下的用户问题往往多样而复杂，模型难以从输入的用户问题中准确理解用户的真实意图，即在语义检索和回答生成过程中存在数据分布偏差，导致模型生成不符合用户预期的回答。如图4-1所示，对于描述不够清晰明确的垂直领域用户问题，知识库检索到的知识文档内容与用户意图存在偏差，这导致 LLM 生成的红色回答较为简略、不符合用户预期。若能对齐人类意图与模型理解，即可得到图中更详细和信息丰富的绿色回答。通过向人类偏好对齐，能够进一步提升垂直领域检索增强对话生成任务的用户交互体验，因此 LLM 的人类偏好对齐研究受到越来越多研究者的关注。

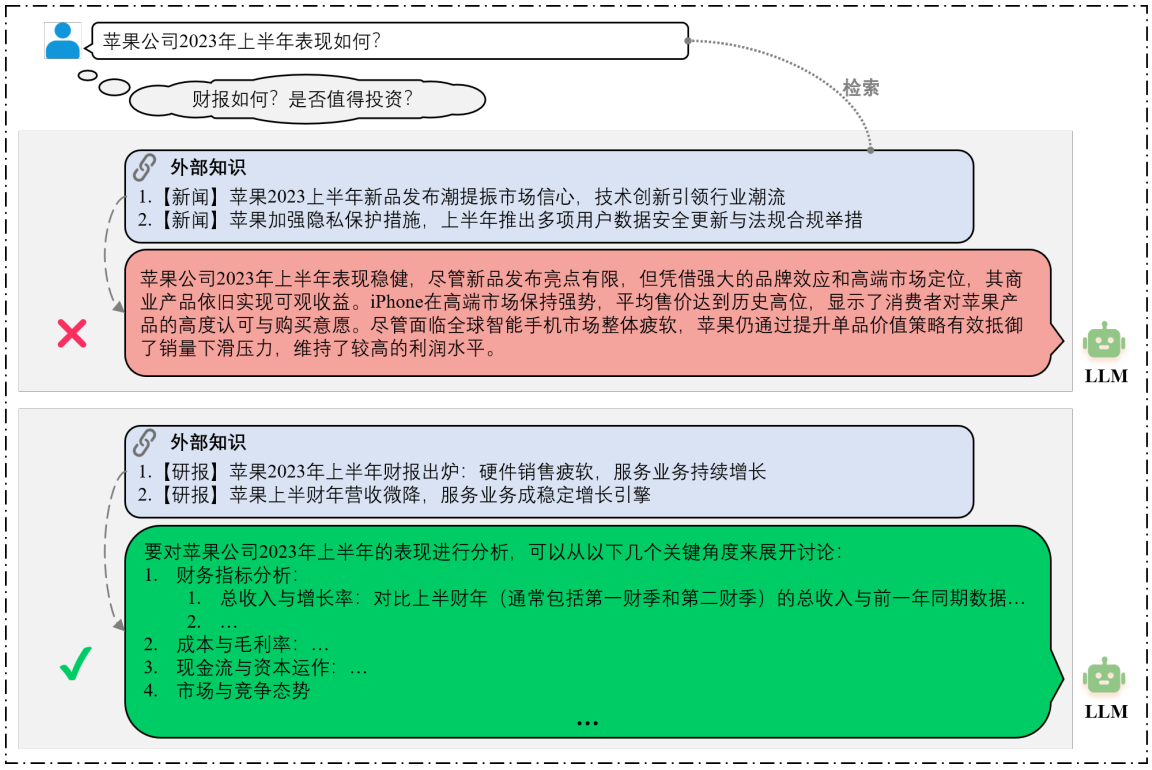


图 4-1 垂直领域对话场景下的模型理解偏差问题示例。

为了实现 LLM 的人类偏好对齐，Long 等人[65]提出了 RLHF 方法，首先采集高质量数据集对语言模型进行监督微调（SFT），再基于 SFT 模型的生成内容收集人类偏好排名数据集，并训练奖励模型（RM），最后执行 PPO 强化学习进一步微调 SFT 模型，显式地引入了人类偏好和主观意见，使得模型生成的内容更符合人类的偏好。然而，RLHF 方法存在一定的局限性。一方面，RLHF 需要大规模的人类偏好标注数据，垂直领域的训练数据则更加难以获取。另一方面，RLHF 使用 PPO 强化学习来训练 LLM，训练过程

复杂且效果稳定性差。为了应对 RLHF 方法存在的问题，Rafailov 等人[96]提出 DPO 算法，通过引入 SFT 约束，在不降低性能的情况下提升训练的稳定性。Dong 等人[93]提出 RAFT 方法，无需使用强化学习，进一步提升稳定性和鲁棒性。Bai 等人[98]提出 RLAIF 方法，使用 LLM 代替人类完成偏好标注工作，一定程度上缓解了数据稀缺问题。然而，上述方法需要训练对话模型本身，随着大型语言模型的发展，对话模型尺寸也不断增大，且部分闭源模型只能通过 API 访问，因此基于训练的方法受到对话模型的尺寸和访问限制难以应用和迁移。

与训练模型使之对齐人类意图不同，Cheng 等人[95]提出 BPO 算法，使用户问题中的用户意图对齐模型理解。然而，BPO 算法存在一些局限性。一方面，BPO 利用 LLM 直接对用户问题进行优化，在检索增强应用场景下，没有考虑检索到的知识文档信息，存在性能瓶颈。另一方面，BPO 使用 LLM 进行用户问题优化，所生成的内容可能受到 LLM 的幻觉问题影响，因此数据有效性不足。

针对上述问题，本文提出一种基于人类偏好对齐的检索增强对话生成方法，通过采集人类对检索增强对话的偏好反馈数据，利用 LLM 分别针对数据库检索和对话生成进行优化，各自得到一个新的用户问题，基于优化后的问题，再次进行相关文档块检索和对话生成，根据生成的回答内容进行样本有效性验证，并用有效样本构建三元组训练数据集，训练一个单独 Seq2Seq 模型作为问题优化器，在推理时自动将用户意图对齐模型理解，实现与模型无关的、稳定的人类偏好对齐。

## 4.2 问题定义

给定垂直领域用户问题 $Q$，外部知识库检索与此相关的文档序列 $D = [d_1, \ldots, d_k]$，对话系统需要基于 $Q$ 和 $D$ 输出符合用户偏好的问题回复 $R$。此任务的主要挑战是需要对齐用户意图和模型理解，使得所生成的回复符合人类偏好。

## 4.3 本章方法

### 4.3.1 方法总体方案

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

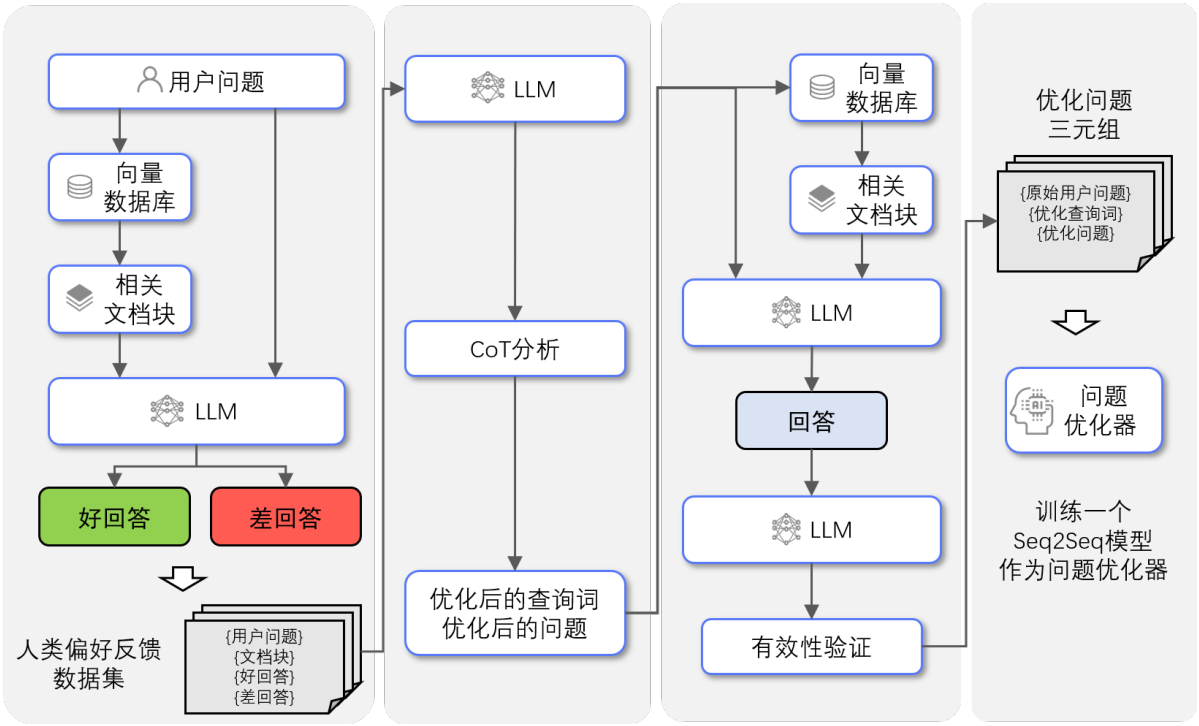


图 4-2 基于人类偏好对齐的检索增强对话生成框架图。

（Query Alignment with Human Feedback, QAHF），基于人类偏好反馈数据训练一个单独的问题优化器，避免直接训练对话模型，从而实现与模型无关的、非侵入的人类偏好对齐。如图4-2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）优化问题构建阶段；3）问题有效性验证阶段；4）问题优化器训练阶段。

### 4.3.2 人类偏好数据采集阶段

为了将模型理解与人类偏好对齐，本节需要对垂直领域对话中的人类偏好进行建模，以提取人类偏好特征，用于后续的优化问题生成。语言模型的训练方式采用 Next Token Prediction（NTP）任务，偏好对齐的目标是找到原始用户问题与真实用户意图的内在联系，从而对原始用户问题和反映真实意图的优化用户问题的条件概率分布进行建模，具体的数据采集步骤如下。

本节首先通过收集垂直领域开源数据集和爬取互联网媒体语料，构建了一个垂直领域问答数据集，该数据集样本为“问题-回答”对，数据样本的问答主题为垂直领域相关专业问题。然后，本章使用基于规则的方法，过滤掉数据集中的低质量样本，以保证样本的用户问题存在优化空间，具体做法是：1）先过滤筛除输入、输出长度低于 10 个标识符的过短样本；2）然后基于百科词条与开源数据集构建垂直领域专有名词词汇表，筛除样本中不包含专有名词的领域无关样本。本章主要关注单轮对话的对话生成。

**算法 1:** 基于人类偏好对齐的检索增强对话生成方法

**输入：**用户问题集合 $Q$，知识文档块集合 $D$，对话模型 $M$，问题优化器 $\pi_\theta$.
**for** q **in** Q **do**
    从 $D$ 中检索与 $q$ 相关性 Top-3 的知识文档 $\{d_1, d_2, d_3\}$
    将 $q$ 与 $\{d_1, d_2, d_3\}$ 输入 $M$，基于随机采样收集回答组 $\{r_1, r_2, ...\}$
    通过回答质量评测标准对 $r_j$ 计算偏好一致性 $c_j$
    选择偏好一致性分数最高和最低的回答 $r_{good}$ 和 $r_{bad}$
**end for**
**for** 偏序数据对 i = 1, ..., N **do**
    通过模型 $M$ 获取偏好特征思维链分析 $A$
    结合思维链分析 $A$，获取优化后的检索问题 $q^i_{search}$ 和优化后的用户问题 $q^i_{qa}$
    从 $D$ 中检索与 $q^i_{search}$ 相关性 Top-3 的知识文档 $\{d^i_1, d^i_2, d^i_3\}$
    将 $q^i_{qa}$ 与 $\{d^i_1, d^i_2, d^i_3\}$ 输入 $M$，收集回答 $r_i$
    通过模型 $M$ 计算优化问题有效性 $c_i$
    **if** $c_i > threshold$ **do**
        将 $c_i$ 加入优化问题数据集 $O$
**end for**
使用预训练模型参数初始化问题优化器的网络参数 $\theta$
**while** 不收敛 **do**
    从优化问题数据集 $O$ 中均匀采样三元组样本
    通过公式（4-1）得到样本输出
    通过 AdamW 优化算法更新 $\theta$
**end while**

随后，在向量数据库中检索与用户问题相关的文档块，并将结果输入 LLM，通过随机采样的生成策略，从 LLM 的语言空间中采样得到多组不同的回答。最后，通过人工标注的方式，根据回答质量评判标准对 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”，“好回答”与“坏回答”的 Pairwise 样本构成偏序回答对，其与原始用户问题和相关文档块共同组成人类偏好反馈数据集。该标注步骤引入人类的偏好信息，即回答评分和排名。其中，

回答质量评判标准包含 5 个维度，即准确性、完整性、逻辑性、表达能力和有效建议，以确保标注标准的客观性和一致性，每个维度的评分取值范围为 0、1、2 分，分别表示“不符合预期”、“不太符合预期”和“较为符合预期”，最终通过加权平均计算得到样本的偏好一致性评分 $C$。

$$C = \frac{1}{5}\sum_{i=1}^{5} c_i \tag{4-1}$$

其中，$c_i$ 分别表示不同维度的评分。

### 4.3.3 优化问题构建阶段

随后，本节通过利用 LLM 的基础推理能力，基于人类偏好反馈数据集中的偏序回答对，对原始用户问题进行优化。问题优化任务包含多个推理步骤，即首先从偏序回答对中提取出人类偏好特征，分析原始用户问题中包含的真实意图，并最终对原始用户问题进行重写，得到优化后的用户问题。问题优化任务属于复杂逻辑推理任务，且由于 LLM 采用自回归的文本生成方式，模型输出内容会受到前面单词的影响，因此直接基于偏序回答对进行问题优化，结果的不稳定性较大，导致进一步加剧 LLM 幻觉问题，无法有效完成人类偏好对齐。同时，由于检索增强对话输入包含多个外部知识文档块，问题优化任务的输入普遍具有较长的上下文，若使用 Few-shot 提示推理策略，模型输入将超过 2048 个标识符，导致模型推理出现外推问题，即推理长度超过模型训练长度后，模型推理的性能显著下降，具体表现为困惑度（Perplexity，PPL）等指标显著上升。为解决上述问题，本节使用基于 Zero-shot-CoT 的提示工程方法进行用户问题优化，通过扩展 LLM 推理步骤，在改写用户问题前补充流畅而符合逻辑的上下文，利用 LLM 的内部参数知识引导模型自身完成人类偏好特征识别，从而改变其推理结构，最大化发挥 LLM 在问题优化任务上的性能。

在检索增强对话生成任务中，用户问题会分别被用于本地知识库的知识文档召回和 LLM 回答生成。知识文档召回计算用户问题和文档块之间的本文相似度与语义相似度，而 LLM 回答生成需要以用户问题为指令进行精准解答，这两个场景的任务目标存在差异，使用相同的用户问题分别进行文档召回和回答生成存在性能瓶颈。因此，本节分别针对本地知识库的知识文档召回和 LLM 回答生成场景各自生成一个优化问题，使得在新的用户问题下，检索增强对话系统生成的回答分布更趋近于“好回答”的分布，区别于“差回答”的分布。

本节所使用的问题优化提示词如表4-1所示。当模型输出结果不符合提示词中的输

出格式要求时，后处理过程中无法提取出有效的优化问题。对此，本节的解决方法是，在后处理抛出匹配异常时，调整模型推理超参数，即提高 temperature 和 top_p，使预测词概率被拉平，并再次进行模型推理，以此循环直到模型输出符合预期格式要求。

表 4-1 问题优化提示词格式。

| 问题优化提示词 |
| --- |
| # SYSTEM<br>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。<br># QUERY<br>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。<br>问题是：<original query><br># DOCS<br><docs><br># GOOD RESPONSE<br><good response><br># BAD RESPONSE<br><bad response><br># TASK<br>你的任务是：<br>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；<br>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；<br>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；<br>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。 |

### 4.3.4 问题有效性验证阶段

LLM 生成的优化问题可能仍然存在理解偏差问题，这类无效样本将成为数据集噪声，若直接使用带噪声的数据进行后续模型训练，可能会进一步加重模型理解偏差，反而损害模型性能。因此，有必要对上一步骤所生成的优化问题数据进行筛选过滤，剔除对人类偏好对齐产生负向优化的低质量样本。相较于直接对数据集样本中的优化问题进行有效性评估，引入问题优化后的模型回复能够更好地观察优化问题对文档召回和回答生成的影响，从而提升有效性评估的准确率。因此，本节将优化后的用户问题重新输入系统，进行知识文档召回后得到新的模型回复，并利用大型语言模型的 Self-Criticism 能力[118]，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。与上一节相同，本节同样使用 Zero-shot-CoT 方法进行问题有效性验证，以提升 LLM 的准确性。

其中，本节所使用的问题有效性验证提示词如表4-2所示，有效性验证过程包含三个步骤：1）计算知识文档与用户问题的相关性，即验证优化后的问题是否与知识文档的语义空间完成对齐；2）计算模型回答的正确性，即验证优化后的问题与对话模型是否完成对齐；3）根据上述相关性和正确性结果，给出最终的问题优化有效性评分，取值范围为 [1, 10]。提示词中对模型输出格式进行了约束，故本节采用正则表达式匹配的方式从模型输出中提取评分结果。

### 4.3.5 问题优化器训练阶段

基于前述步骤所构建的人类偏好数据集，训练一个 Seq2Seq[68]模型作为问题优化器，实现模型无关的、可插拔的用户问题自动优化。形式上，给定原始用户问题 $Q_{ori}$，优化器生成优化后的用户问题 $Q_{opt}$：

$$Q_{opt} = [Q_{search}, [SEP], Q_{qa}] \tag{4-2}$$

其中，$Q_{search}$ 表示用于数据库检索的问题，$Q_{qa}$ 表示用于回答生成的问题，$[SEP]$ 为特殊标识符，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:

$$\mathcal{L} = -\frac{1}{N}\sum_{t=1}^{N} \log P\left(x_t \mid Q_{\text{ori}}, x_{<t}\right) \tag{4-3}$$

其中，$N$ 是 $Q_{opt}$ 的长度，$x_t$ 表示 $Q_{opt}$ 中的第 $t$ 个标识符。本章选择 Qwen-7B 作为基座

表 4-2 问题有效性验证提示词格式。

| 有效性验证提示词 |
| --- |
| #SYSTEM<br>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。<br># QUERY<br>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。<br>问题是：<original query><br># DOCS<br><docs><br># RESPONSE<br>< response><br># TASK<br>你的任务是：<br>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；<br>2. 判断 RESPONSE 是否能够准确、可靠地回答 QUERY；<br>3. 对 RESPONSE 进行评分，分数取值范围为 [1, 10]。 |

模型，以更好地学习 $Q_{ori}$ 和 $Q_{opt}$ 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，作为本方法中的问题优化器具有一定优势。

## 4.4 实验结果

### 4.4.1 数据集介绍

- FinGPT-FiQA[119]是 Wang 等人创建的人工评估数据集。它包含了 17.1k 个金融领域的用户问题与回答样本，这些样本均采样于实际应用场景。本节从中抽取了 1000 个样本作为实验的测试数据。
- AlphaFin-test 是本文第三章中构建的 AlphaFin 数据集的测试集，它包含 1000 条

金融与数据问答对，AlphaFin-test 数据集能够评估模型在金融领域对话上的性能表现。更详细的数据集介绍参考第三章内容。

### 4.4.2 评价指标

本节使用 GPT-4 偏好评价和 Ragas[115]评估框架作为实验评估指标。其中，本节使用 Ragas 中的 Context Precision、Context Recall、Faithfulness 三项指标。指标具体介绍如下。

- GPT-4 偏好评价：由 GPT-4[116]模型担任评判员，从两个不同模型的回复中选择更优的回答，考察角度与人类偏好评价保持一致，并在系统指令中对 GPT-4 模型进行约束，指令提示词格式参考 MT-Bench[117]评估框架。
- 上下文准确率（Context Precision）：利用 LLM（如 GPT-4 模型）评估知识文档块与用户问题之间的相关性及文档块排名顺序。
- 上下文召回率（Context Recall）：利用 LLM 估计模型回答和文档块的 TP 和 FN，计算文档块的召回率。
- 可信度（Faithfulness）：利用 LLM 计算 (用户问题, 模型回答, 文档块) 三元组的自然语言推断（NLI）分数，即对模型回答的事实性进行量化评估。

### 4.4.3 基座模型

- Qwen-7B-Chat[68]：通义千问-7B（Qwen-7B）是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[120]：FinGPT 是面向金融领域的大模型系列。它使用自建金融数据集在 LLaMA2-13B[121]、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[66]：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[65]：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，基于 GPT[122]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。

### 4.4.4 基准方法

- PPO[123]：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[95]：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。

### 4.4.5 实现细节

对于 QAHF，本节使用 Qwen-7B[68]作为优化器的基座模型，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。在训练阶段，本节使用 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 LinearLRScheduler 学习率调度方法。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 的推理解码策略，另外，所有测试的基座模型都使用默认的解码策略。

### 4.4.6 与现有方法的性能比较

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，其性能均优于原始问题，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 相比于原始问题的胜率分别提高了 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 ChatGPT 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 4-3 QAHF 在 FinGPT-FiQA Eval 和 AlphaFin-test 上的有效性实验。

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

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 | | | 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% |
| | QAHF | BPO | **36.8%** | 39.6% | 23.6% | **52.5%** | 12.7% | 34.8% | +15.4% |
| | QAHF | ori. | **58.0%** | 21.0% | 21.0% | **61.0%** | 5.2% | 33.8% | +32.1% |
| ChatGPT | BPO | ori. | **39.4%** | 12.3% | 48.3% | **43.6%** | 25.5% | 30.9% | +1.9% |
| | QAHF | BPO | **31.1%** | 38.5% | 30.4% | **40.9%** | 28.2% | 30.9% | +5.4% |
| | QAHF | 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 ChatGPT 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，进一步证明 QAHF 方法相较于 BPO 方法的优越性。

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。

本节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”。

表 4-5 QAHF 与 PPO 在 AlphaFin-test 上的性能对比。

| 模型 | 方法 | | AlphaFin-test | | | Δ WR |
|---|---|---|---|---|---|---|
| | A | B | A win | Tie | B win | |
| ChatGLM | PPO | ori. | **13.7%** | 75.4% | 10.9% | +2.8% |
| | QAHF | PPO | **49.3%** | 20.6% | 30.1% | +19.2% |
| | QAHF | ori. | **61.0%** | 5.2% | 33.8% | +27.2% |

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 (本章方法) | 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。 |

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。

### 4.4.7 问题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 ChatGPT 进行快速优化。如表4-8和表4-9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 4-7 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 (本章方法) | 作为知识答疑助手，我认为招商证券 2020 年下半年市场份额如下：<br>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。<br>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%。<br>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。<br>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，公司经纪业务、投行和资管业务仍保持了头部地位，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%。 |

表 4-8 在 FinGPT-FiQA 数据集上探究问题有效性验证模块对性能的影响。

| 模型 | 方法 | | FinGPT-FiQA Eval | | | Δ WR |
|---|---|---|---|---|---|---|
| | A | B | A win | Tie | B win | |
| ChatGLM | QAHF | ori. | **58.0%** | 21.0% | 21.0% | +32.1% |
| | w/o Eval | ori. | **49.8%** | 25.5% | 24.7% | +24.8% |
| | QAHF | w/o Eval | **8.6%** | 86.3% | 5.1% | +4.5% |

表 4-9 在 AlphaFin-test 数据集上探究问题有效性验证模块对性能的影响。

| 模型 | 方法 | | AlphaFin-test | | | Δ WR |
|---|---|---|---|---|---|---|
| | A | B | A win | Tie | B win | |
| ChatGLM | QAHF | ori. | **61.0%** | 5.2% | 33.8% | +32.1% |
| | w/o Eval | ori. | **46.9%** | 30.6% | 22.5% | +24.8% |
| | QAHF | w/o Eval | **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4-10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，可能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4-11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

表 4-10 在 Ragas 指标上问题有效性验证模块对性能的影响。

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

表 4-11 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 | 国庆黄金周对旅游股的影响主要体现在以下几个方面：<br>1. 客流量增加：…<br>2. 板块内部分红机会：…<br>3. 旅游股的业绩弹性：…<br>4. 旅游股面临了一定的市场风险：…<br>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。 |
| 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 | 杭州亚运会将为杭州的公司带来积极影响，具体来说：<br>1. 亚运会将带来大量游客：…<br>2. 亚运会将带动周边房价上涨：…<br>3. 亚运会将推动杭州文化产业发展：…<br>4. 亚运会将带动杭州基础设施建设：…<br>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。 |

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。

### 4.4.8 迭代优化对方法性能的影响

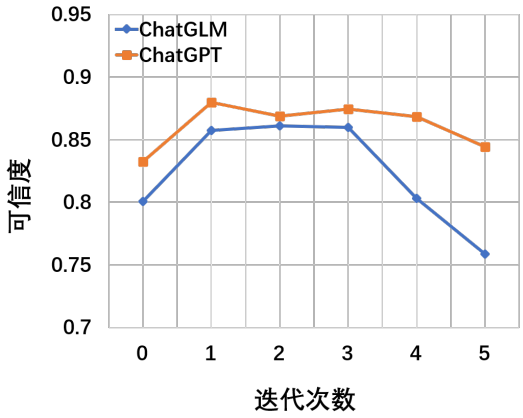


图 4-3 不同迭代优化次数对可信度的影响。

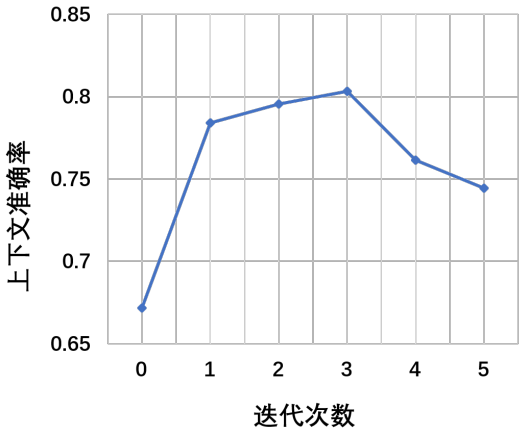


图 4-4 不同迭代优化次数对上下文准确率的影响。

由于 QAHF 可以优化用户问题以获得更好的模型回复，自然的想法是，是否可以多次迭代优化用户问题，逐步增强 LLM 的输出。因此，本节在 AlphaFin-test 数据集上进行了实验，探究迭代优化对方法性能的影响。具体来说，本节对原始用户问题进行了 5 次迭代优化，并与原始问题进行了 Ragas 可信度和上下文准确率指标的比较。如图4-3所示，在第 1 次迭代中，ChatGLM 和 ChatGPT 模型的回复可信度均有明显的改善。在前 3 次迭代中，可信度基本持平。而在第 4 次迭代时，ChatGPT 性能略有下降，ChatGLM 则显著降低。此外，本节还发现 QAHF 表现出良好的保留性，当输入的原始问题已经足够好时，QAHF 有很高的概率保留它。本节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。

同时，本节探究迭代优化对上下文准确率指标的影响情况，即探究其对数据库知识文档检索精度的影响，结果如图4-4所示。在经过第 1 次迭代优化后，上下文准确率由 0.6717 显著提升至 0.7839。在第 2、3 次优化后，准确率增速减缓，提升至 0.8032。在第 4、5 次优化后，准确率下降到 0.7444。可以看出，多次迭代优化对知识文档检索准确度的收益不高，在第 4 次优化后出现负面影响。

表 4-12 在偏好评价指标上人类反馈对性能的影响。

| 模型 | 方法 | | AlphaFin-test | | | Δ WR |
|---|---|---|---|---|---|---|
| | A | B | A win | Tie | B win | |
| ChatGLM | QAHF | ori. | **61.0%** | 5.2% | 33.8% | +27.2% |
| | w/o Human | ori. | **39.4%** | 35.8% | 24.8% | +14.6% |
| | QAHF | w/o Human | **45.7%** | 20.3% | 34.0% | +11.7% |

### 4.4.9 人类反馈对方法性能的影响

QAHF 的一个重要组成部分是利用人类反馈来优化用户问题。为了研究人类反馈对 QAHF 的优化效果有多大影响，本节进行了消融实验，以比较人类反馈学习优化(QAHF) 和直接使用 LLM（ChatGPT）进行优化的性能差异。如表4-12所示，使用 LLM 直接优化用户问题也能够提高模型回复性能，这表明 LLM 具有一定的用户问题优化的潜力。本章所提出的 QAHF 方法相较于没有人类反馈的直接优化方法，取得了 11.7% 的胜率增量，这表明显式引入人类反馈能够进一步激发 LLM 优化检索增强对话场景下用户问题的潜力，从而实现更有效的人类偏好对齐。

## 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

## 一、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方法，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：

- 针对模型内外部知识不一致，影响模型有效利用外部知识的问题，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外部知识库构建和内部知识注入，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。
- 针对人类用户意图与模型理解之间存在偏差，造成模型生成不符合用户预期回答的问题，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优越性。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领域、云计算领域和法律领域，进行垂直领域知识问答任务实验，并在金融领域真实股票市场价格趋势预测任务上，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。

## 二、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方法性能。
- 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] 杨州, 陈志豪, 蔡铁城, 等. 基于深度学习的情感对话响应综述[J]. 计算机学报, 2023, 46(12): 2489-2519.

[2] 徐凡, 徐健明, 马勇, 等. 基于知识增强的开放域多轮对话模型[J]. 软件学报, 2024, 35(02): 758-772.

[3] Jagerman R, Zhuang H, Qin Z, et al. Query Expansion by Prompting Large Language Models[J]. CoRR, 2023, abs/2305.03653. arXiv: 2305.03653.

[4] Asai A, Wu Z, Wang Y, et al. Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection[J]. CoRR, 2023, abs/2310.11511. arXiv: 2310.11511.

[5] Wang L, Yang N, Wei F. Query2doc: Query Expansion with Large Language Models [C].Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). 2023: 9414-9423.

[6] Cui J, Li Z, Yan Y, et al. ChatLaw: Open-Source Legal Large Language Model with Integrated External Knowledge Bases[J]. CoRR, 2023, abs/2306.16092. arXiv: 2306.16092.

[7] Zhou H, Zheng C, Huang K, et al. KdConv: A Chinese Multi-domain Dialogue Dataset Towards Multi-turn Knowledge-driven Conversation[C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2020: 7098-7108.

[8] Jung J, Son B, Lyu S. AttnIO: Knowledge Graph Exploration with In-and-Out Attention Flow for Knowledge-Grounded Dialogue[C].Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020: 3484-3497.

[9] 陈晨, 朱晴晴, 严睿, 等. 基于深度学习的开放领域对话系统研究综述[J]. 计算机学报, 2019, 42(07): 1439-1466.

[10] Ni J, Young T, Pandelea V, et al. Recent Advances in Deep Learning Based Dialogue Systems: a Systematic Survey[J]. Artificial intelligence review (AI), 2023, 56(4): 3055-3155.

[11] Deng L, Tür G, He X, et al. Use of Kernel Deep Convex Networks and End-to-End Learning for Spoken Language Understanding[C].Proceedings of Workshop for IEEE Spoken Language Technology (SLT). 2012: 210-215.

[12] Tür G, Deng L, Hakkani-Tür D, et al. Towards Deeper Understanding: Deep Convex

Networks for Semantic Utterance Classification[C].Proceedings of IEEE International Conference on Acoustics, Speech and SP (ICASSP). 2012: 5045-5048.

[13] Sarikaya R, Hinton G E, Deoras A. Application of Deep Belief Networks for Natural Language Understanding[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing (TASLP), 2014, 22(4): 778-784.

[14] Ravuri S V, Stolcke A. Recurrent Neural Network and LSTM Models for Lexical Utterance Classification[C].Conference of the International Speech Communication Association (INTERSPEECH). 2015: 135-139.

[15] Hashemi H B, Asiaee A, Kraft R. Query intent detection using convolutional neural networks[C].Proceedings of Workshop for ACM International Conference on Web Search and Data Mining (WSDM): vol. 23. 2016.

[16] Lee J Y, Dernoncourt F. Sequential Short-Text Classification with Recurrent and Convolutional Neural Networks[C].Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human LanguageTechnologies (NAACL-HLT). 2016: 515-520.

[17] Wu C, Hoi S C H, Socher R, et al. ToD-BERT: Pre-trained Natural Language Understanding for Task-Oriented Dialogues[J]. CoRR, 2020, abs/2004.06871. arXiv: 2004.06871.

[18] Henderson M, Thomson B, Young S J. Deep Neural Network Approach for the Dialog State Tracking Challenge[C].Proceedings of the Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL). 2013: 467-471.

[19] Mrksic N, Séaghdha D Ó, Thomson B, et al. Multi-domain Dialog State Tracking using Recurrent Neural Networks[C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2015: 794-799.

[20] Mrksic N, Séaghdha D Ó, Wen T, et al. Neural Belief Tracker: Data-Driven Dialogue State Tracking[C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2017: 1777-1788.

[21] Lei W, Jin X, Kan M, et al. Sequicity: Simplifying Task-oriented Dialogue Systems with Single Sequence-to-Sequence Architectures[C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2018: 1437-1447.

[22] Wang Y, Guo Y, Zhu S. Slot Attention with Value Normalization for Multi-Domain Dia-

logue State Tracking[C].Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020: 3019-3028.

[23] Zhang Z, Li X, Gao J, et al. Budgeted Policy Learning for Task-Oriented Dialogue Systems[C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2019: 3742-3751.

[24] Takanobu R, Liang R, Huang M. Multi-Agent Task-Oriented Dialog Policy Learning with Role-Aware Reward Decomposition[C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2020: 625-638.

[25] 黄民烈, 朱小燕. 对话管理中基于槽特征有限状态自动机的方法研究[J]. 计算机学报, 2004, 27(08): 1092-1101.

[26] Wen T, Gasic M, Kim D, et al. Stochastic Language Generation in Dialogue using Recurrent Neural Networks with Convolutional Sentence Reranking[C].Proceedings of the Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL). 2015: 275-284.

[27] Wen T, Gasic M, Mrksic N, et al. Multi-domain Neural Network Language Generation for Spoken Dialogue Systems[C].Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human LanguageTechnologies (NAACL-HLT). 2016: 120-129.

[28] Li Y, Yao K, Qin L, et al. Slot-consistent NLG for Task-oriented Dialogue Systems with Iterative Rectification Network[C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2020: 97-106.

[29] Brown T B, Mann B, Ryder N, et al. Language Models are Few-Shot Learners [C].Advances in Neural Information Processing Systems (NeurIPS). 2020.

[30] Devlin J, Chang M, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[C].Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human LanguageTechnologies (NAACL-HLT). 2019: 4171-4186.

[31] Raffel C, Shazeer N, Roberts A, et al. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer[J]. Journal of Machine Learning Research (JMLR), 2020, 21: 140:1-140:67.

[32] Sutskever I, Vinyals O, Le Q V. Sequence to Sequence Learning with Neural Networks [C].Advances in Neural Information Processing Systems (NeurIPS). 2014: 3104-3112.

[33] Cho K, van Merrienboer B, Gülçehre Ç, et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation[C].Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014: 1724-1734.

[34] Sordoni A, Bengio Y, Vahabi H, et al. A Hierarchical Recurrent Encoder-Decoder for Generative Context-Aware Query Suggestion[C].Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM). 2015: 553-562.

[35] Serban I V, Sordoni A, Lowe R, et al. A Hierarchical Latent Variable Encoder-Decoder Model for Generating Dialogues[C].Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). 2017: 3295-3301.

[36] Weston J, Chopra S, Bordes A. Memory Networks[C].International Conference on Learning Representations (ICLR). 2015.

[37] Sukhbaatar S, Szlam A, Weston J, et al. End-To-End Memory Networks[C].Advances in Neural Information Processing Systems (NeurIPS). 2015: 2440-2448.

[38] Vinyals O, Fortunato M, Jaitly N. Pointer Networks[C].Advances in Neural Information Processing Systems (NeurIPS). 2015: 2692-2700.

[39] Gu J, Lu Z, Li H, et al. Incorporating Copying Mechanism in Sequence-to-Sequence Learning[C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2016.

[40] Balakrishnan A, Rao J, Upasani K, et al. Constrained Decoding for Neural NLG from Compositional Representations in Task-Oriented Dialogue[C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2019: 831-844.

[41] Chen X, Xu J, Xu B. A Working Memory Model for Task-oriented Dialog Response Generation[C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2019: 2687-2693.

[42] Gao S, Zhang Y, Ou Z, et al. Paraphrase Augmented Task-Oriented Dialog Generation[C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2020: 639-649.

[43] Wang W, Zhang J, Li Q, et al. Incremental Learning from Scratch for Task-Oriented Dialogue Systems[C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2019: 3710-3720.

[44] Dai Y, Li H, Tang C, et al. Learning Low-Resource End-To-End Goal-Oriented Dialog for Fast and Reliable System Deployment[C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2020: 609-618.

[45] He W, Yang M, Yan R, et al. Amalgamating Knowledge from Two Teachers for Task-oriented Dialogue System with Adversarial Training[C].Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020: 3498-3507.

[46] Zhang Y, Sun S, Galley M, et al. DIALOGPT : Large-Scale Generative Pre-training for Conversational Response Generation[C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2020: 270-278.

[47] Radford A, Wu J, Child R, et al. Language models are unsupervised multitask learners [J]. OpenAI blog, 2019, 1(8): 9.

[48] Adiwardana D, Luong M, So D R, et al. Towards a Human-like Open-Domain Chatbot [J]. CoRR, 2020, abs/2001.09977. arXiv: 2001.09977.

[49] Roller S, Dinan E, Goyal N, et al. Recipes for Building an Open-Domain Chatbot [C].Proceedings of Conference of the European Chapter of the Association for Computational Linguistics (EACL). 2021: 300-325.

[50] Bao S, He H, Wang F, et al. PLATO: Pre-trained Dialogue Generation Model with Discrete Latent Variable[C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2020: 85-96.

[51] Lin X, Jian W, He J, et al. Generating Informative Conversational Response using Recurrent Knowledge-Interaction and Knowledge-Copy[C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2020: 41-52.

[52] Wu S, Li Y, Zhang D, et al. Diverse and Informative Dialogue Generation with Context-Specific Commonsense Knowledge Awareness[C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2020: 5811-5820.

[53] Majumder B P, Jhamtani H, Berg-Kirkpatrick T, et al. Like hiking? You probably enjoy nature: Persona-grounded Dialog with Commonsense Expansions[C].Proceedings of

the Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020: 9194-9206.

[54] Moon S, Shah P, Kumar A, et al. OpenDialKG: Explainable Conversational Reasoning with Attention-based Walks over Knowledge Graphs[C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2019: 845-854.

[55] Xu J, Wang H, Niu Z, et al. Conversational Graph Grounded Policy Learning for Open-Domain Conversation Generation[C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2020: 1835-1845.

[56] Yang S, Zhang R, Erfani S M. GraphDialog: Integrating Graph Knowledge into End-to-End Task-Oriented Dialogue Systems[C].Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020: 1878-1888.

[57] 吕学强, 张剑, 穆天杨, 等. 嵌入知识语义的医疗领域对话系统[J]. 计算机工程与设计, 2023, 44(12): 3794-3799.

[58] Wang Y, Kordi Y, Mishra S, et al. Self-Instruct: Aligning Language Models with Self-Generated Instructions[C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2023: 13484-13508.

[59] Wu S, Irsoy O, Lu S, et al. BloombergGPT: A Large Language Model for Finance[J]. CoRR, 2023, abs/2303.17564. arXiv: 2303.17564.

[60] Zhang X, Yang Q. XuanYuan 2.0: A Large Chinese Financial Chat Model with Hundreds of Billions Parameters[C].Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM). 2023: 4435-4439.

[61] Meng C, Ren P, Chen Z, et al. DukeNet: A Dual Knowledge Interaction Network for Knowledge-Grounded Conversation[C].Proceedings of the International Conference on Research on Development in Information Retrieval (SIGIR). 2020: 1151-1160.

[62] Chen Y, Wang Z, Xing X, et al. BianQue: Balancing the Questioning and Suggestion Ability of Health LLMs with Multi-turn Health Conversations Polished by ChatGPT[J]. CoRR, 2023, abs/2310.15896. arXiv: 2310.15896.

[63] Zhang X, Yang Q. Self-QA: Unsupervised Knowledge Guided Language Model Alignment[J]. CoRR, 2023, abs/2305.11952. arXiv: 2305.11952.

[64] Wang H, Liu C, Xi N, et al. HuaTuo: Tuning LLaMA Model with Chinese Medical

Knowledge[J]. CoRR, 2023, abs/2304.06975. arXiv: 2304.06975.

[65] Ouyang L, Wu J, Jiang X, et al. Training Language Models to Follow Instructions With Human Feedback[C].Advances in Neural Information Processing Systems (NeurIPS). 2022.

[66] Zeng A, Liu X, Du Z, et al. GLM-130B: An Open Bilingual Pre-trained Model [C].International Conference on Learning Representations (ICLR). 2023.

[67] Sun Y, Wang S, Feng S, et al. ERNIE 3.0: Large-scale Knowledge Enhanced Pre-training for Language Understanding and Generation[J]. CoRR, 2021, abs/2107.02137. arXiv: 2107.02137.

[68] Bai J, Bai S, Chu Y, et al. Qwen Technical Report[J]. CoRR, 2023, abs/2309.16609. arXiv: 2309.16609.

[69] Touvron H, Lavril T, Izacard G, et al. LLaMA: Open and Efficient Foundation Language Models[J]. CoRR, 2023, abs/2302.13971. arXiv: 2302.13971.

[70] Yang A, Xiao B, Wang B, et al. Baichuan 2: Open Large-scale Language Models[J]. CoRR, 2023, abs/2309.10305. arXiv: 2309.10305.

[71] Long J. Large Language Model Guided Tree-of-Thought[J]. CoRR, 2023, abs/2305.08291. arXiv: 2305.08291.

[72] Wei J, Wang X, Schuurmans D, et al. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models[C].Advances in Neural Information Processing Systems (NeurIPS). 2022.

[73] Wang X, Wei J, Schuurmans D, et al. Self-Consistency Improves Chain of Thought Reasoning in Language Models[C].International Conference on Learning Representations (ICLR). 2023.

[74] Yao Y, Li Z, Zhao H. Beyond Chain-of-Thought, Effective Graph-of-Thought Reasoning in Large Language Models[J]. CoRR, 2023, abs/2305.16582. arXiv: 2305.16582.

[75] Lewis P S H, Perez E, Piktus A, et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks[C].Advances in Neural Information Processing Systems (NeurIPS). 2020.

[76] Liu N F, Lin K, Hewitt J, et al. Lost in the Middle: How Language Models Use Long Contexts[J]. CoRR, 2023, abs/2307.03172. arXiv: 2307.03172.

[77] Ding N, Qin Y, Yang G, et al. Delta Tuning: A Comprehensive Study of Parameter Efficient Methods for Pre-trained Language Models[J]. CoRR, 2022, abs/2203.06904. arXiv: 2203.06904.

[78] Houlsby N, Giurgiu A, Jastrzebski S, et al. Parameter-Efficient Transfer Learning for NLP[C].Proceedings of Machine Learning Research: Proceedings of the International Conference on Machine Learning (ICML): vol. 97. 2019: 2790-2799.

[79] Mahabadi R K, Henderson J, Ruder S. Compacter: Efficient Low-Rank Hypercomplex Adapter Layers[C].Advances in Neural Information Processing Systems (NeurIPS). 2021: 1022-1035.

[80] Pfeiffer J, Kamath A, Rücklé A, et al. AdapterFusion: Non-Destructive Task Composition for Transfer Learning[C].Proceedings of Conference of the European Chapter of the Association for Computational Linguistics (EACL). 2021: 487-503.

[81] Li X L, Liang P. Prefix-Tuning: Optimizing Continuous Prompts for Generation [C].Proceedings of the Conference of the Association for Computational Linguistics and International Joint Conference on Natural Language Processing (ACL/IJCNLP). 2021: 4582-4597.

[82] 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]. CoRR, 2021, abs/2110.07602. arXiv: 2110.07602.

[83] Gu Y, Han X, Liu Z, et al. PPT: Pre-trained Prompt Tuning for Few-shot Learning [C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2022: 8410-8423.

[84] Lee J, Tang R, Lin J. What Would Elsa Do? Freezing Layers During Transformer Fine-Tuning[J]. CoRR, 2019, abs/1911.03090. arXiv: 1911.03090.

[85] Zaken E B, Goldberg Y, Ravfogel S. BitFit: Simple Parameter-efficient Fine-tuning for Transformer-based Masked Language-models[C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2022: 1-9.

[86] Zhao M, Lin T, Mi F, et al. Masking as an Efficient Alternative to Finetuning for Pre-trained Language Models[C].Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020: 2226-2241.

[87] Aghajanyan A, Gupta S, Zettlemoyer L. Intrinsic Dimensionality Explains the Effectiveness of Language Model Fine-Tuning[C].Proceedings of the Conference of the Association for Computational Linguistics (ACL). 2021: 7319-7328.

[88] Li C, Farkhoor H, Liu R, et al. Measuring the Intrinsic Dimension of Objective Landscapes[C].International Conference on Learning Representations (ICLR). 2018.

[89] Hu E J, Shen Y, Wallis P, et al. LoRA: Low-Rank Adaptation of Large Language Models [C].International Conference on Learning Representations (ICLR). 2022.

[90] Qin Y, Wang X, Su Y, et al. Exploring Low-dimensional Intrinsic Task Subspace via Prompt Tuning[J]. CoRR, 2021, abs/2110.07867. arXiv: 2110.07867.

[91] Li Y, Wei F, Zhao J, et al. RAIN: Your Language Models Can Align Themselves without Finetuning[J]. CoRR, 2023, abs/2309.07124. arXiv: 2309.07124.

[92] Zheng R, Dou S, Gao S, et al. Secrets of RLHF in Large Language Models Part I: PPO [J]. CoRR, 2023, abs/2307.04964. arXiv: 2307.04964.

[93] Dong H, Xiong W, Goyal D, et al. RAFT: Reward rAnked FineTuning for Generative Foundation Model Alignment[J]. CoRR, 2023, abs/2304.06767. arXiv: 2304.06767.

[94] Yuan Z, Yuan H, Tan C, et al. RRHF: Rank Responses to Align Language Models with Human Feedback without tears[J]. CoRR, 2023, abs/2304.05302. arXiv: 2304.05302.

[95] Cheng J, Liu X, Zheng K, et al. Black-Box Prompt Optimization: Aligning Large Language Models without Model Training[J]. CoRR, 2023, abs/2311.04155. arXiv: 2311.04155.

[96] Rafailov R, Sharma A, Mitchell E, et al. Direct Preference Optimization: Your Language Model is Secretly a Reward Model[C].Advances in Neural Information Processing Systems (NeurIPS). 2023.

[97] Liu T, Zhao Y, Joshi R, et al. Statistical Rejection Sampling Improves Preference Optimization[J]. CoRR, 2023, abs/2309.06657. arXiv: 2309.06657.

[98] Bai Y, Kadavath S, Kundu S, et al. Constitutional AI: Harmlessness from AI Feedback [J]. CoRR, 2022, abs/2212.08073. arXiv: 2212.08073.

[99] Li Z, Xu T, Zhang Y, et al. ReMax: A Simple, Effective, and Efficient Reinforcement Learning Method for Aligning Large Language Models[J]. CoRR, 2023, abs/2310.10505. arXiv: 2310.10505.

[100] Wei J, Bosma M, Zhao V Y, et al. Finetuned Language Models are Zero-Shot Learners [C].International Conference on Learning Representations (ICLR). 2022.

[101] Gao Y, Xiong Y, Gao X, et al. Retrieval-Augmented Generation for Large Language Models: A Survey[J]. CoRR, 2023, abs/2312.10997. arXiv: 2312.10997.

[102] Xiao S, Liu Z, Zhang P, et al. C-Pack: Packaged Resources To Advance General Chinese Embedding[J]. CoRR, 2023, abs/2309.07597. arXiv: 2309.07597.

[103] Muennighoff N. SGPT: GPT Sentence Embeddings for Semantic Search[J]. CoRR, 2022, abs/2202.08904. arXiv: 2202.08904.

[104] Zhong H, Xiao C, Tu C, et al. JEC-QA: A Legal-Domain Question Answering Dataset [C].Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). 2020: 9701-9708.

[105] TuShare[Z]. https://github.com/waditu/tushare. 2019.

[106] King A. AKShare[Z]. https://github.com/akfamily/akshare. 2019.

[107] Xie Q, Han W, Zhang X, et al. PIXIU: A Large Language Model, Instruction Data and Evaluation Benchmark for Finance[J]. CoRR, 2023, abs/2306.05443. arXiv: 2306.05443.

[108] Yang H, Liu X, Wang C D. FinGPT: Open-Source Financial Large Language Models [J]. CoRR, 2023, abs/2306.06031. arXiv: 2306.06031.

[109] Ho T K. The Random Subspace Method for Constructing Decision Forests[J]. IEEE Trans on Pattern Analysis and Machine Intelligence (TPAMI), 1998, 20(8): 832-844.

[110] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088): 533-536.

[111] Hochreiter S, Schmidhuber J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[112] Cox D R. The Regression Analysis of Binary Sequences[J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 1958, 20(2): 215-232.

[113] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System[C].Proceedings of the ACM Knowledge Discovery and Data Mining (SIGKDD). 2016: 785-794.

[114] Quinlan J R. C4.5: Programs for Machine Learning[M]. 1993.

[115] ES S, James J, Anke L E, et al. RAGAs: Automated Evaluation of Retrieval Augmented

Generation[C].Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics (EACL). 2024: 150-158.

[116] Achiam J, Adler S, Agarwal S, et al. GPT-4 Technical Report[J]. CoRR, 2023, abs/2303.08774. arXiv: 2303.08774.

[117] Zheng L, Chiang W, Sheng Y, et al. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena[C].Oh A, Naumann T, Globerson A, et al. Advances in Neural Information Processing Systems (NeurIPS). 2023.

[118] Tan X, Shi S, Qiu X, et al. Self-Criticism: Aligning Large Language Models with their Understanding of Helpfulness, Honesty, and Harmlessness[C].Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). 2023: 650-662.

[119] Wang N, Yang H, Wang C D. FinGPT: Instruction Tuning Benchmark for Open-Source Large Language Models in Financial Datasets[J]. CoRR, 2023, abs/2310.04793. arXiv: 2310.04793.

[120] Yang H, Liu X, Wang C D. FinGPT: Open-Source Financial Large Language Models [J]. CoRR, 2023, abs/2306.06031. arXiv: 2306.06031.

[121] Touvron H, Martin L, Stone K, et al. Llama 2: Open Foundation and Fine-Tuned Chat Models[J]. CoRR, 2023, abs/2307.09288. arXiv: 2307.09288.

[122] Radford A, Narasimhan K, Salimans T, et al. Improving Language Understanding by Generative Pre-training[J]. 2018.

[123] Schulman J, Wolski F, Dhariwal P, et al. Proximal Policy Optimization Algorithms[J]. CoRR, 2017, abs/1707.06347. arXiv: 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 (已投稿) | 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