

说明书

发明名称: 一种大模型交互方法和系统

技术领域

[0001] 本发明涉及大模型技术领域，特别涉及一种大模型交互和系统。

背景技术

[0002] 大模型技术发展，用户可以通过与大模型的交互，快速获取到答案。用户侧，应用提示词工程、检索增强开发的聊天机器人或对话平台，用户在有限的聊天记忆窗口、上下文信息对话中，只能获得直白返回的请求结果；企业侧，在看到应用数量飞速增长、流量陡增的对话平台时，无法有效地对大模型用户分发营销内容。

[0003] 当前绝大多数的大模型交互系统为检索增强类型，如图1所示，其主要通过本地专有数据库或第三方api检索来实现专有内容、实时内容的交互对话。但具有如下缺点：答案结果为与会话请求直接关联的纯文本；以及未得到更准确的答案，用户需尽可能准确详细地描述需求。

[0004] 以perplexity平台为例，作为体验较好的对话平台之一，该perplexity平台在传统检索增强直接回答的基础上，补充引用源信息和相关问题反馈，用户可以进一步了解答案详情、简化用户提问交互。其页面截图与逻辑分析结构分别如图2和图3所示。其缺点是：当前平台返回结果为与会话请求直接关联的内容；没有进一步推理用户需求意图，从而提供服务推荐。营销能力的缺失，也导致平台商业模式只能局限在用户订阅付费的方式。

发明概述

技术问题

[0005] 本发明要解决的技术问题，在于提供一种大模型交互方法和系统，在检索增强、有源检索的基础之上提出营销类信息检索增强，为用户提供更好的体验且能承接企业侧营销需求。

技术解决方案

- [0006] 第一方面，本发明提供了一种大模型交互方法，包括下述步骤：
- [0007] S1、获取用户会话输入，并对用户会话输入进行解析得到用户请求；
- [0008] S2、针对所述用户请求进行检索并汇总，得到答案检索结果，所述答案检索结果包含引用源链接和检索内容；
- [0009] 同时综合分析用户个人设置信息、第三方用户画像信息、历史对话信息与意图、当前对话上下文，总结推理得到当前用户意图；基于所述用户会话输入和所述当前用户意图，在营销数据库中检索营销投放信息并进行汇总，得到营销投放检索结果，所述营销投放检索结果包含投放链接、投放内容、投放类型以及投放策略；
- [0010] S3、综合所述答案检索结果与所述营销投放检索结果的链接权重，以及所述营销投放检索结果的竞价价格，对所述答案检索结果的引用来源与所述营销投放检索结果的引用来源进行综合排序后，作为引用来源结果返回给用户；
- [0011] S4、对所述引用来源结果和所述营销投放检索结果进行总结和整合后，得到答案结果并返回给用户；
- [0012] S5、对所述用户请求、所述答案结果、所述营销检索结果进行整合，得到推荐关联问题后返回给用户，引导用户进一步多轮对话。
- [0013] 第二方面，本发明提供了一种大模型交互系统，包括：
- [0014] 会话解析模块，用于获取用户会话输入，并对用户会话输入进行解析得到用户请求；
- [0015] 检索模块，用于针对所述用户请求进行检索并汇总，得到答案检索结果，所述答案检索结果包含引用源链接和检索内容；
- [0016] 营销投放模块，用于在检索模块进行检索的同时，综合分析用户个人设置信息、第三方用户画像信息、历史对话信息与意图、当前对话上下文，总结推理得到当前用户意图；基于所述用户会话输入和所述当前用户意图，在营销数据库中检索营销投放信息并进行汇总，得到营销投放检索结果，所述营销投放检索结果包含投放链接、投放内容、投放类型以及投放策略；
- [0017] 引用来源生成模块，用于综合所述答案检索结果与所述营销投放检索结果的链接权重，以及所述营销投放检索结果的竞价价格，对所述答案检索结果的引用

来源与所述营销投放检索结果的引用来源进行综合排序后，作为引用来源结果返回给用户；

[0018] 答案结果生成模块，用于对所述引用来源结果和所述营销投放检索结果进行总结和整合后，得到答案结果并返回给用户；

[0019] 推荐关联问题生成模块，用于对所述用户请求、所述答案结果、所述营销检索结果进行整合，得到推荐关联问题后返回给用户，引导用户进一步多轮对话。

有益效果

[0020] 本发明实施例中提供的一个或多个技术方案，至少具有如下技术效果或优点：本发明融合大模型推理检索能力、个性化内容生成，提升用户体验；同时满足企业类用户营销需求，融合营销信息流策略，将营销内容融入问答，实现更广泛、高效的营销内容分发，拓展大模型应用方式，为用户提供更好的体验且能承接企业侧营销需求。

[0021] 上述说明仅是本发明技术方案的概述，为了能够更清楚了解本发明的技术手段，而可依照说明书的内容予以实施，并且为了让本发明的上述和其它目的、特征和优点能够更明显易懂，以下特举本发明的具体实施方式。

附图说明

[0022] 下面参照附图结合实施例对本发明作进一步的说明。

[0023] 图1为现有大模型交互方法的流程图；

[0024] 图2为perplexity平台的返回结果时的界面示意图；

[0025] 图3为perplexity平台返回结果的流程逻辑示意图；

[0026] 图4为本发明方法的流程图；

[0027] 图5为本发明系统的框架示意图。

本发明的实施方式

[0028] 本申请实施例通过提供一种大模型交互方法和系统，在检索增强、有源检索的基础上提出营销类信息检索增强，为用户提供更好的体验且能承接企业侧营销需求。

- [0029] 本申请实施例中的技术方案，总体思路如下：本发明融合大模型推理检索能力、个性化内容生成，提升用户体验；同时满足企业类用户营销需求，融合营销信息流策略，将营销内容融入问答，实现更广泛、高效的营销内容分发，拓展大模型应用方式，为用户提供更好的体验且能承接企业侧营销需求。
- [0030] 实施例一
- [0031] 如图4所示，本实施例提供一种大模型交互方法，包括下述步骤：
- [0032] S1、获取用户会话输入，并对用户会话输入进行解析得到用户请求；
- [0033] S2、针对所述用户请求进行检索并汇总，得到答案检索结果，所述答案检索结果包含引用源链接和检索内容；
- [0034] 同时综合分析用户个人设置信息、第三方用户画像信息、历史对话信息与意图、当前对话上下文，总结推理得到当前用户意图；基于所述用户会话输入和所述当前用户意图，在营销数据库中检索营销投放信息并进行汇总，得到营销投放检索结果，所述营销投放检索结果包含投放链接、投放内容、投放类型以及投放策略；
- [0035] S3、综合所述答案检索结果与所述营销投放检索结果的链接权重，以及所述营销投放检索结果的竞价价格，对所述答案检索结果的引用来源与所述营销投放检索结果的引用来源进行综合排序后，作为引用来源结果返回给用户；所述综合排序包括：
- [0036] （a1）适用于资讯站点流量引导，即用户在搜索某条资讯时，提升相似资讯中链接权重排名；
- [0037] （a2）适用于官方链接保护与引流，基于官方信息保护、最新营销方案分发、方，确保用户收到结果为官方最新信息；
- [0038] （a3）适用于同品类对象、榜单或渠道竞争引流，在回答用户问题的同时，推介同类产品或服务。
- [0039] S4、对所述引用来源结果和所述营销投放检索结果进行总结和整合后，得到答案结果并返回给用户；所述总结和整合是基于所述营销投放检索结果的投放策略来实现，并包括：

- [0040] (b1) 基于意图推测的补充额外信息：即基于当前用户意图推测，基于可能的用户服务类请求意图，补充提供服务功能对接；
- [0041] (b2) 官方最新营销方案的拓展补充；
- [0042] (b3) 同类榜单、竞品、渠道信息、推介信息拓展与替换。
- [0043] S5、对所述用户请求、所述答案结果、所述营销检索结果进行整合，还整合以下内容：
- [0044] (c1) 系统基于意图推测的服务功能引导；
- [0045] (c2) 官方营销方案的引导类问题；
- [0046] (c3) 同类榜单、竞品、渠道信息、推介信息引导类问题；
- [0047] 从而得到推荐关联问题后返回给用户，引导用户进一步多轮对话。
- [0048] 基于同一发明构思，本申请还提供了与实施例一中的方法对应的装置，详见实施例二。
- [0049] 实施例二
- [0050] 如图5所示，在本实施例中提供了一种大模型交互系统，包括：
- [0051] 会话解析模块，用于获取用户会话输入，并对用户会话输入进行解析得到用户请求；
- [0052] 检索模块，用于针对所述用户请求进行检索并汇总，得到答案检索结果，所述答案检索结果包含引用源链接和检索内容；
- [0053] 营销投放模块，用于在检索模块进行检索的同时，综合分析用户个人设置信息、第三方用户画像信息、历史对话信息与意图、当前对话上下文，总结推理得到当前用户意图；基于所述用户会话输入和所述当前用户意图，在营销数据库中检索营销投放信息并进行汇总，得到营销投放检索结果，所述营销投放检索结果包含投放链接、投放内容、投放类型以及投放策略；
- [0054] 引用来源生成模块，用于综合所述答案检索结果与所述营销投放检索结果的链接权重，以及所述营销投放检索结果的竞价价格，对所述答案检索结果的引用来源与所述营销投放检索结果的引用来源进行综合排序后，作为引用来源结果返回给用户；所述综合排序包括：

- [0055] (a1) 适用于资讯站点流量引导, 即用户在搜索某条资讯时, 提升相似资讯中链接权重排名;
- [0056] (a2) 适用于官方链接保护与引流, 基于官方信息保护、最新营销方案分发、方, 确保用户收到结果为官方最新信息;
- [0057] (a3) 适用于同品类对象、榜单或渠道竞争引流, 在回答用户问题的同时, 推介同类产品或服务。
- [0058] 答案结果生成模块, 用于对所述引用来源结果和所述营销投放检索结果进行总结和整合后, 得到答案结果并返回给用户; 所述总结和整合是基于所述营销投放检索结果的投放策略来实现, 并包括:
- [0059] (b1) 基于意图推测的补充额外信息: 即基于当前用户意图推测, 基于可能的用户服务类请求意图, 补充提供服务功能对接;
- [0060] (b2) 官方最新营销方案的拓展补充;
- [0061] (b3) 同类榜单、竞品、渠道信息、推介信息拓展与替换。
- [0062] 推荐关联问题生成模块, 用于对所述用户请求、所述答案结果、所述营销检索结果进行整合, 还整合以下内容:
- [0063] (c1) 系统基于意图推测的服务功能引导;
- [0064] (c2) 官方营销方案的引导类问题;
- [0065] (c3) 同类榜单、竞品、渠道信息、推介信息引导类问题;
- [0066] 从而得到推荐关联问题后返回给用户, 引导用户进一步多轮对话。
- [0067] 由于本发明实施例二所介绍的系统, 为实施本发明实施例一的方法所采用的系统, 故而基于本发明实施例一所介绍的方法, 本领域所属人员能够了解该装置的具体结构及变形, 故而在此不再赘述。凡是本发明实施例一的方法所采用的装置都属于本发明所欲保护的范围。
- [0068] 以下是本发明的应用场景举例:
- [0069] 用户画像包括:
- [0070] 用户个人信息设置: 两个孩子的母亲;
- [0071] 用户标签: 热爱生活、热爱游戏;
- [0072] 交互过程包括:

[0073] 一、药物与检测结果询问交互

[0074] 第一轮交互：

[0075] 用户会话输入：“阿莫西林”；

[0076] 大模型平台（下称平台）解析用户请求，对答案进行检索；推理交互意图，无关联历史记录，默认科学场景；

[0077] 平台返回答案与信息源“阿莫西林(Amoxicillin)是一种半合成广谱 β -内酰胺类抗生素，主要用于治疗由敏感细菌引起的各种感染。以下是关于阿莫西林的一些主要信息。作用机理：xxx，适应症：xxx，不良反应：xxx，注意事项与建议：xxx”；

[0078] 第二轮交互：

[0079] 用户会话输入：“阿莫西林克拉维酸钾”；

[0080] 平台解析用户请求，对答案进行检索；推理交互意图，历史查询阿莫西林、科学场景；

[0081] 平台返回答案与信息源“阿莫西林克拉维酸钾是一种复方制剂，由阿莫西林和克拉维酸钾两种成分组成。它结合了两药物的优势，扩大了抗菌谱xxx。作用机理：xxx，适应症：xxx，不良反应：xxx，注意事项与建议：xxx”；

[0082] 第三轮交互：

[0083] 用户会话输入：“儿童试纸检测，鼻病毒阳性，处方是阿莫西林克拉维酸钾”

[0084] 平台解析用户请求，对答案进行检索；

[0085] 推理用户交互意图，先获取下述信息：

[0086] “

[0087] 会话历史：询问主要围绕阿莫西林及其复方制剂的药理特性，以及在儿童鼻病毒感染中应用该药物的原因和注意事项展开；

[0088] 可能意图：想全面了解这两种抗生素药物的情况，以及在儿童鼻病毒感染中应用它们的合理性和注意事项，从而对医生的用药方式有进一步的认识和把握；

[0089] 情感反应：反映出作为家长或患者对药物使用的关注和疑虑，希望能够获得专业的解答和指导；

[0090] 平台策略：

- [0091] 表示理解和关注，表达出对用户疑虑的理解和重视，用友好、亲和的语气，让用户感受到诚挚关注。这有助于建立良好的沟通氛围。
- [0092] 提供一般性信息，可以提供一些该药物的一般性信息，如适应症、作用机制、常见不良反应等。
- [0093] 强调遵从医嘱，客观地指出，在未经专业人士确认的情况下，最好还是遵从医生的专业判断和医嘱，避免自行中断或调整用药方式。
- [0094] 提供权威医疗资源链接，搜索引擎可以在结果中优先展示来自权威医疗机构的官方网站或文章，如食品药品监督管理局、疾病预防控制中心、医院等。这些资源可以为用户提供准确、可信的药物信息。
- [0095] 专业医生的在线解答，一些医疗咨询网站有医生在线值班服务，搜索引擎可以汇总这些网站上医生对该药物使用疑问的回答，为用户提供专业意见。
- [0096] 提供正规医院的导诊挂号入口，如果用户表达希望复诊或二诊的需求，提供当地大型综合医院官方的导诊挂号入口；
- [0097] 在线购药服务，如果是需要凭处方购买的药品，不进行推荐任何在线药品购买渠道，因为存在一定的安全性和合法性风险，该类药品建议用户到当地正规的线下药房或医院药房购买。
- [0098] ”
- [0099] 平台返回答案结果与引用来源结果“对于儿童鼻病毒检测呈阳性，医生开具阿莫西林克拉维酸钾的处方，可能是出于以下几个考虑：
- [0100] 预防继发细菌感染：鼻病毒感染后，机体防御力下降，细菌更容易合并感染，如中耳炎、肺炎等。阿莫西林克拉维酸钾能有效预防和治疗这些细菌感染。
- [0101] 缓解症状：虽然病毒性感冒本身不需要服用抗生素，但一些细菌引起的并发症(如鼻窦炎)可能会加重症状。阿莫西林可以治疗这些细菌感染，从而缓解症状。
- [0102] 病情评估：有时医生难以确定是纯粹的病毒感染还是细菌感染，会根据症状预防性用药，待病情观察后适当调整。
- [0103] 患儿情况：如患儿免疫力较低或存在其他基础疾病，医生会考虑预防性使用抗生素，以防细菌并发症。

- [0104] 然而，也有医生认为对于普通儿童病毒感冒，不必过度使用抗生素。因此，家长可以向医生了解具体用药原因，并严格按医嘱用药，同时也要重视对症支持治疗，如常规饮食、休息等。如症状未缓解或加重，须及时复诊。”
- [0105] 同时推荐关联问题，即补充返回在线导诊咨询、本地就诊挂号或本地非处方药购买服务入口和问题引导。
- [0106] 二、热门游戏查询交互
- [0107] 用户会话输入：“近期热门游戏《游戏名1》”
- [0108] 系统查找知识库，汇总答案检索结果R1如下：
- [0109] 链接1，“近期非常火爆的新游戏有《游戏名1》，.简单介绍...从用户口碑和下载量来看反响非常不错”；
- [0110] 链接2，“《游戏名1》官方网站内容xxx”；
- [0111] 链接3，“《游戏名1》玩家社区，评价与攻略xxx”；
- [0112] 用户意图识别，根据用户画像和用户请求，检索营销投放信息，汇总营销检索结果R2；
- [0113] 系统综合R1与R2链接权重、竞价价格对答案检索结果R1、营销投放检索结果R2的引用来源进行排序后，作为引用来源结果A1返回给用户，原R1结果可能受R2结果影响，出现下述情况：
- [0114] 如果R2存在资讯站点流量引导类营销投放（相似内容），链接1-1替换链接1，“近期非常火爆的新游戏有《游戏名1》，.简单介绍..。”；
- [0115] 如果有官方营销投放，链接2为官方最新投放链接；
- [0116] 如果有同品类对象、榜单或渠道竞争引流，R1的第三条记录可能受影响，排名靠后，返回用户A1可能变成链接3-1，“《游戏名1》同类产品介绍xxx 榜单xxx”；
- [0117] 系统对引用来源结果A1和营销投放检索结果R2进行总结和整合后，得到答案结果A2并返回给用户；总结和整合是基于R2投放策略来实现，包括下述三种类型：

- [0118] 系统基于意图推测的补充额外信息：系统基于意图推测，用户画像“热爱游戏”，可能意图“了解与体验游戏”，基于可能的用户服务类请求意图，答案内容补充“可以通过下述正版渠道下载或购买体验：xxxx”；
- [0119] 官方最新营销方案的拓展补充，在原始游戏基本介绍之外，补充最新游戏资料片或运营活动简介；
- [0120] 同类榜单、竞品、渠道信息、推介信息拓展与替换，返回结果中的体验下载链接替换为“xx宝”应用下载链接；
- [0121] 系统对用户请求Q、答案结果A2和营销检索结果R2进行整合，在返回推理直接关联可能问题的同时，还可整合以下内容，作为推荐关联问题A3返回，引导用户进一步多轮对话：
- [0122] 系统基于意图推测的服务功能引导：系统基于意图推测，基于“了解与体验游戏”请求意图，补充提供服务功能对接类交互引导问题“《游戏名1》的下载链接”，“《游戏名1》的攻略”，“《游戏名1》的玩家社区是什么？”；
- [0123] 官方营销方案的引导类问题：“《游戏名1》最新资料片活动体验下载”；
- [0124] 同类榜单、竞品、渠道信息、推介信息引导类问题：“《游戏名1》类型与同品类MMO榜单”，“《游戏名1》竞品《游戏名2》口碑与体验”。
- [0125] 虽然以上描述了本发明的具体实施方式，但是熟悉本技术领域的技术人员应当理解，我们所描述的具体的实施例只是说明性的，而不是用于对本发明的范围的限定，熟悉本领域的技术人员在依照本发明的精神所作的等效的修饰以及变化，都应当涵盖在本发明的权利要求所保护的范围内。

权利要求书

[权利要求 1]

一种大模型交互方法，其特征在于：包括下述步骤：

S1、获取用户会话输入，并对用户会话输入进行解析得到用户请求；

S2、针对所述用户请求进行检索并汇总，得到答案检索结果，所述答案检索结果包含引用源链接和检索内容；

同时综合分析用户个人设置信息、第三方用户画像信息、历史对话信息与意图、当前对话上下文，总结推理得到当前用户意图；基于所述用户会话输入和所述当前用户意图，在营销数据库中检索营销投放信息并进行汇总，得到营销投放检索结果，所述营销投放检索结果包含投放链接、投放内容、投放类型以及投放策略；

S3、综合所述答案检索结果与所述营销投放检索结果的链接权重，以及所述营销投放检索结果的竞价价格，对所述答案检索结果的引用来源与所述营销投放检索结果的引用来源进行综合排序后，作为引用来源结果返回给用户；

S4、对所述引用来源结果和所述营销投放检索结果进行总结和整合后，得到答案结果并返回给用户；

S5、对所述用户请求、所述答案结果、所述营销检索结果进行整合，得到推荐关联问题后返回给用户，引导用户进一步多轮对话。

[权利要求 2]

根据权利要求1所述的一种大模型交互的方法，其特征在于：所述S3中的综合排序包括：

（a1）适用于资讯站点流量引导，即用户在搜索某条资讯时，提升相似资讯中链接权重排名；

（a2）适用于官方链接保护与引流，基于官方信息保护、最新营销方案分发、方，确保用户收到结果为官方最新信息；

（a3）适用于同品类对象、榜单或渠道竞争引流，在回答用户问题的同时，推介同类产品或服务。

- [权利要求 3] 根据权利要求1所述的一种大模型交互的方法，其特征在于：所述S4的总结和整合是基于所述营销投放检索结果的投放策略来实现，并包括：
- （b1）基于意图推测的补充额外信息：即基于当前用户意图推测，基于可能的用户服务类请求意图，补充提供服务功能对接；
 - （b2）官方最新营销方案的拓展补充；
 - （b3）同类榜单、竞品、渠道信息、推介信息拓展与替换。
- [权利要求 4] 根据权利要求1所述的一种大模型交互的方法，其特征在于：所述S5中，所述推荐关联问题还整合以下内容：
- （c1）系统基于意图推测的服务功能引导；
 - （c2）官方营销方案的引导类问题；
 - （c3）同类榜单、竞品、渠道信息、推介信息引导类问题。
- [权利要求 5] 一种大模型交互系统，其特征在于：包括：
- 会话解析模块，用于获取用户会话输入，并对用户会话输入进行解析得到用户请求；
- 检索模块，用于针对所述用户请求进行检索并汇总，得到答案检索结果，所述答案检索结果包含引用源链接和检索内容；
- 营销投放模块，用于在检索模块进行检索的同时，综合分析用户个人设置信息、第三方用户画像信息、历史对话信息与意图、当前对话上下文，总结推理得到当前用户意图；基于所述用户会话输入和所述当前用户意图，在营销数据库中检索营销投放信息并进行汇总，得到营销投放检索结果，所述营销投放检索结果包含投放链接、投放内容、投放类型以及投放策略；
- 引用来源生成模块，用于综合所述答案检索结果与所述营销投放检索结果的链接权重，以及所述营销投放检索结果的竞价价格，对所述答案检索结果的引用来源与所述营销投放检索结果的引用来源进行综合排序后，作为引用来源结果返回给用户；

答案结果生成模块，用于对所述引用来源结果和所述营销投放检索结果进行总结和整合后，得到答案结果并返回给用户；

推荐关联问题生成模块，用于对所述用户请求、所述答案结果、所述营销检索结果进行整合，得到推荐关联问题后返回给用户，引导用户进一步多轮对话。

[权利要求 6]

根据权利要求5所述的一种大模型交互系统，其特征在于：所述引用来源生成模块中的综合排序包括：

（a1）适用于资讯站点流量引导，即用户在搜索某条资讯时，提升相似资讯中链接权重排名；

（a2）适用于官方链接保护与引流，基于官方信息保护、最新营销方案分发、方，确保用户收到结果为官方最新信息；

（a3）适用于同品类对象、榜单或渠道竞争引流，在回答用户问题的同时，推介同类产品或服务。

[权利要求 7]

根据权利要求5所述的一种大模型交互系统，其特征在于：所述答案结果生成模块的总结和整合是基于所述营销投放检索结果的投放策略来实现，并包括：

（b1）基于意图推测的补充额外信息：即基于当前用户意图推测，基于可能的用户服务类请求意图，补充提供服务功能对接；

（b2）官方最新营销方案的拓展补充；

（b3）同类榜单、竞品、渠道信息、推介信息拓展与替换。

[权利要求 8]

根据权利要求5所述的一种大模型交互系统，其特征在于：所述推荐关联问题生成模块在对所述用户请求、所述答案结果、所述营销检索结果进行整合时，还整合以下内容：

（c1）系统基于意图推测的服务功能引导；

（c2）官方营销方案的引导类问题；

（c3）同类榜单、竞品、渠道信息、推介信息引导类问题。