

# 【LLM论文阅读】ProductAgent:电商场景中产品检索的多轮澄清实现策略

原创 方方 方方的算法花园 2024年11月06日 08:54 北京

## 0 ▶ 论文概况

### 1. 论文名称:

ProductAgent: Benchmarking Conversational Product Search Agent with Asking Clarification Questions 《ProductAgent: 通过询问澄清问题对对话式产品检索agent进行基准测试》

2. 论文链接: <https://arxiv.org/pdf/2407.00942>

3. 论文作者所在机构: 清华大学, 阿里达摩院, 鹏城实验室

4. 一句话概括: 该论文提出电商场景中产品需求澄清任务, 设计了具备战略澄清问题生成和动态产品检索能力的 ProductAgent, 介绍了 PROCLARE 基准以自动评估其性能, 经实验验证了 ProductAgent 的有效性, 同时分析了其局限性。

## 1 ▶ 挑战

在电商场景中, 用户通常难以迅速找到符合自身需求的产品, 因为他们可能不熟悉哪些产品特性适合自己的具体使用情况, 且在表达购买需求时可能存在模糊性。

## 2 ▶ 论文贡献点

### 1. 提出产品需求澄清任务及解决方案:

(1) **明确了产品需求澄清任务** (如下图), 即在电商场景中, 针对用户初始query模糊的问题, 通过agent与用户对话, 以澄清问题的方式精准定位用户需求, 从而提高产品搜索的准确性。

(2) **设计了ProductAgent作为解决方案**, 它能在对话中整合产品数据库、记忆模块和工具, 执行产品特征总结、query生成和产品检索等操作, 逐步明确用户需求, 提升产品搜索效果。

### 2. 引入可靠的自动评估基准:

(1) **构建了PROCLARE基准**, 用于对产品需求澄清任务进行全面评估。该基准关注自动对话检索评估, 涵盖传统设置和对话设置两种场景, 能有效测试agent在不同情况下的检索性能。

(2) **借助LLM驱动的用户模拟器实现自动评估**, 通过模拟用户与agent的对话, 生成模拟用户响应, 避免了人工标注的高成本, 提高了评估效率, 为产品需求澄清任务提供了可靠且自动化的评估pipelines。

### 3. 实验验证及分析改进方向

(1) **开展广泛实验**, 验证了ProductAgent在对话过程中随着交互轮次增加, 检索性能显著提升, 证明了其有效性。

(2) 深入分析了ProductAgent的性能，包括交互轮次对检索性能的影响、澄清操作对性能的提升作用、不同策略获取产品统计数据的效果以及agent失败的原因等。通过这些分析，为进一步改进agent性能提供了有价值的参考，指出了未来探索更有效规划策略、优化SQL生成、改进数据集和提升提示设计等研究方向。

3 ▶ 产品需求澄清任务

A conversation example
A: I am a product sales assistant. Please just type in the targeted product category first, and I would search for you.
U: Sports shoes.
A: I have search several relevant items for you: <Item <sub>1</sub> >, <Item <sub>2</sub> >, ... Also, you can answer these questions so that I could recommend more suitable items for you!
Q1: What is the suitable scenario for the sports shoes you want to purchase?
C1: ["Outdoor", "Brisk walking", "Basketball", "Dance", "Travel", "Other"]
Q2: What style of sports shoes do you like?
C2: ["Versatile", "Internet celebrity", "Foreign style", "Basketball", "Breathable style", "Other"]
Q3: Please choose your preferred color for sports shoes.
C3: ["Light gray", "White gray", "Moonlight", "Yellow", "Rose Red", "Other"]
U: Answers for Q1: "Outdoor", "Basketball".
Answers for Q2: "Basketball", "Breathable style".
Answers for Q3: I like green.
A: I have search several relevant items for you: <Item <sub>1</sub> >, <Item <sub>2</sub> >, ... Also, you can answer these questions...

Table 1: A conversation example between ProductAgent (A) and the user (U). Since the clarification questions are formulated as multi-choice questions, the user can conveniently respond by selecting the preferred options or typing in the answers when needed.

产品需求澄清任务的目标是**通过人与agent的协作来改善电子商务搜索过程**。该任务基于一种典型的在线购物场景，即用户已经决定购买某一类别产品，但他们尚未确定所感兴趣产品的详细方面。因此，他们经常难以清晰地表达自己的购买需求，特别是对于带有陌生术语的产品类别。

为了应对这一挑战，该任务旨在**通过明确提出澄清问题来识别用户需求，这些问题有助于消除产品搜索方面的任何不确定性**。该任务以对话的形式进行，智能体主动与用户进行交互。具体而言，给定产品类别  $U_1$ ，他们之间的对话  $D$  表述如下： $D = \{U_1, A_1, P_1, U_2, A_2, P_2, \dots\}$ ，其中  $U_t$  ( $t = 1, 2, \dots$ ) 是用户表述， $A_t$  是智能体表述， $P_t$  是检索到的产品项目。智能体表述主要包括多个多项选择澄清问题  $A_t = \{Q_{t1}, Q_{t2}, \dots, Q_{tn}\}^2$ 。每个问题包含问题本身的内容以及一组答案候选。

需要注意的是，在每一轮中，agent都会输出问题和检索到的项目作为及时反馈，以方便提供友好的客户应用界面。**该任务面临几个关键挑战：**

- 1) 生成澄清问题需要大量的产品知识。
- 2) agent应具备准确理解用户意图的能力，以便检索到最相关的产品项目，并生成动态且有价值的澄清问题。
- 3) 有必要建立定量和自动化的评估协议，以促进针对该任务开发有效的agent。

4 ▶ ProductAgent

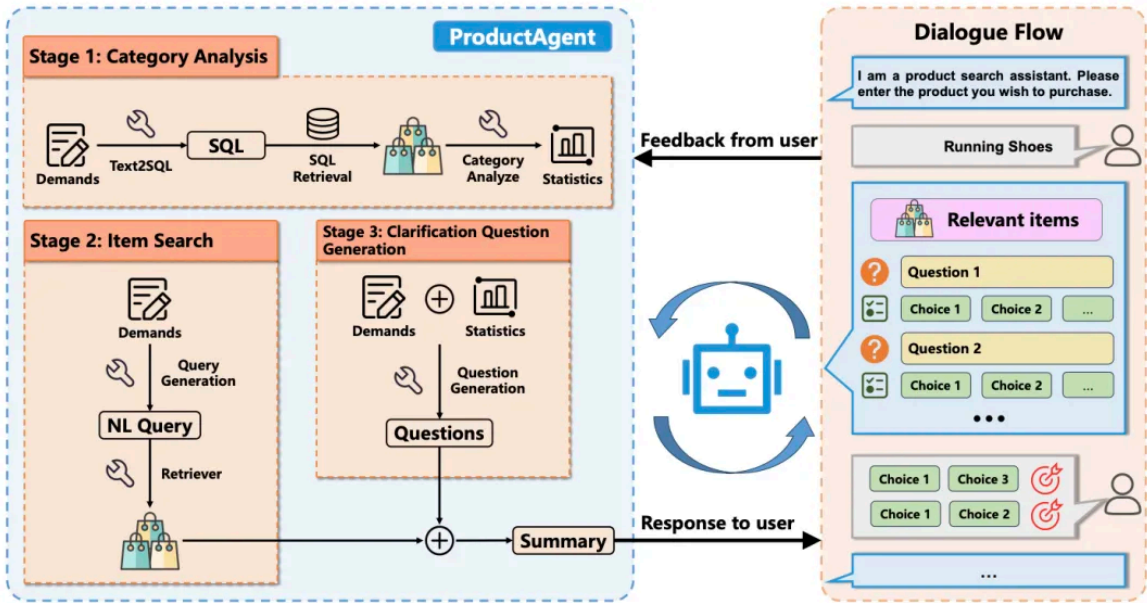


Figure 2: Overview of our proposed ProductAgent.

ProductAgent执行对话循环，每个循环可以分解为三个主要步骤。在初始化循环之前，用户通过输入产品类别作为原始query来开始对话。

**(1) 类别分析：**根据每一轮已知的用户需求收集和总结相关且详细阐述的产品知识。agent首先综合一个反映所有已知用户需求的query，然后使用该query从数据库中检索相关产品项目。检索到的项目（可能涵盖感兴趣的理想项目）随后被总结为统计数据。这些统计数据涵盖了各种产品方面，并且作为一个动态知识库，根据正在进行的用户需求实时构建，为后续步骤奠定基础。

**(2) 项目搜索：**根据所有已知需求动态地检索不同的相关产品项目。ProductAgent首先利用query生成工具，根据当前需求生成自然语言query。然后，使用生成的query来检索最符合用户需求的项目。

**(3) 澄清问题生成：**agent精心策划提出一些新的澄清问题及其答案候选，这些问题是结合知识统计数据和已知用户需求而生成的。在交互界面中，用户通过选择选项或提供能反映其兴趣的额外个人信息，来高效地澄清自身需求。这种用户反馈具有重要价值，能助力提高产品搜索的准确性。

上述步骤在每一轮对话中不断重复，旨在让后续轮次的用户需求比前一轮更加精准。也正因为如此，随着时间的推移，产品搜索的准确性得以稳步提升。

## 5 PROCLARE Benchmark

PROCLARE Benchmark是用于评估产品需求澄清任务中对话式电商搜索agent性能的基准，具体内容如下：

**(1) 文档集：**创建了一个包含100万份文档（产品项）的集合，涵盖20个类别，作为基准的知识库。利用AliMe KG构建数据集，从每个类别中随机抽取50,000份文档，并对所有数据进行命名实体识别（NER）预处理，将文档存储在关系型SQL数据集中，以确保高效的数据管理。

**(2) 传统信息检索：**在PROCLARE的传统设置中，使用与先前工作相同的方法评估传统产品搜索任务的检索性能（此设置不涉及agent）。从整个文档集中为每个类别随机抽取100份文档，共2000份，使用Doc2Query技术将其转换为合成query，然后测量不同检索方法基于这些合成query的检索性能。

**(3) 检索策略：**实验采用了三种广泛使用的检索器：BM25、General Text Embedding (GTE) 和CoROM，并结合了重新排序 (bge-reranker-base) 和融合技术。

**(4) 对话式信息检索：**在对话式设置中，通过用户模拟器自动评估ProductAgent的性能，用户模拟器在每个回合中回答澄清问题以模拟真实对话。对于每个query文档，模拟10个回合（5个用户回合和5个代理回合），用户模拟器从输入所需产品类别开始对话。

**(5) 自动评估指标：**采用信息检索基准中常用的评估方法，使用平均倒数排名（MRR@10）和前10命中率（HIT@10）来评估两种设置下的检索性能。对于重新排序结果，仅报告MRR@10指标。

6 ▶ Prompts && Examples

1. Prompts for ProductAgent

(1) Text2SQL（翻译版）

```
1  你是一个 SQL 生成助手。给定一些约束条件，你需要生成一个语法正确的{dialect}SQL 查询语
2  为了生成合理的查询，你必须遵循以下规则：
3  1. 直接生成只有 SQL 查询语句，不输出任何解释或推理信息。
4  2. 直接使用“*”来检索所有列。
5  3. 使用“LIMIT”关键字将检索到的记录的最大数量限制为{{max_number}}。
6  4. 仔细构建查询的 where 条件，并尽可能多地使用“LIKE”关键字。
7
8  给定以下表结构描述，只允许检索该表：
9  Table schema: {schema}
10
11 Example input:
12 Product category: 休闲裤
13 Question: 你想购买的休闲裤的主要材料是什么？
14 Answer: 聚酯纤维
15
16 Example output:
17 SQL Query: SELECT * From item WHERE category='休闲裤'
18 AND material LIKE '%聚酯纤维%' LIMIT {max_number};
19
20 Input:
21 Product category: {category}
22 Question: {question_1}
23 Answer: {answer_1}
24 ...
25 SQL Query:
```

(2) Query 生成 (翻译版)

- 1 您是一个query生成助手。鉴于用户的购买需求，您需要生成一个用于检索目标产品的简短自然语言
- 2 为了生成一个合理的query，你必须遵循以下规则：
- 3
- 4 1.生成的query要简洁，由关键字组成，空格分隔。
- 5 2.生成的query应涵盖用户的所有购买需求。
- 6 3.不要输出任何解释或推断信息，不要使用引号等不必要的标点符号。
- 7
- 8 Product category: {category}
- 9 Question: {question\_1}
- 10 Answer: {answer\_1}
- 11 ...
- 12 Query:

(3) 澄清问题生成 (翻译版)

- 1 你是一个能够准确识别用户需求的产品购物助手，你能够为需求澄清生成三个多项选择问题。
- 2 为了帮助你提出有价值的问题，这里有关于 {category} 的统计摘要。
- 3 Statistics: {statistics}
- 4
- 5 1.生成的内容必须聚焦于产品类别（ {category} ），有助于准确识别用户需求。
- 6 2.禁止生成与之前重复的新问题。
- 7 3.在构建选项时，尽量直接从统计数据中选择。
- 8
- 9 你需要根据历史问答为需求澄清生成新的多项选择题。以下是历史问答：
- 10 Question: {question\_1}
- 11 Answer: {answer\_1}
- 12 ...
- 13
- 14 生成的结果应严格遵循 JSON 格式，不应包含其他额外信息。
- 15 JSON format description: json\_description
- 16 ```json

2. 用户模拟器的prompt

- 1 你正在与一位购物助手交谈，希望他们能帮助你搜索合适的产品。
- 2 为了给出合理的答案，你必须遵循以下规则：
- 3 1. 你必须根据目标产品准确地回答问题。
- 4 2. 尽量使用所提供的选项来回答，如果你的答案不在选项范围内，只需简单地回答“Other”。
- 5 3. 直接输出答案，不要提供任何解释或推理信息，并且避免不必要的标点符号，如引号。
- 6 4. 每个问题的答案应单独占一行。
- 7 这里是目标产品信息: {item}
- 8 这里是助手提出的正式问题: {questions}



9 直接回答上述问题，不要重复问题，\*\*每个答案应单独占一行\*\*。

3. 对话示例

基于GPT-4的ProductAgent和用户模拟器的对话示例。

1 A: 我是产品销售助理。请先输入目标产品类别，我会为您搜索。

2 U: 帆布鞋

3 A: 我为您搜索了几个相关项目: <Item1>、<Item2>, . . . . 此外, 您可以回答这些问题, 以

4 Q1: 当你需要帆布鞋时, 你主要关注哪个颜色?

5 C1: ["浅灰色", "天蓝色", "黄色", "军绿色", "玫瑰红", "其他"]

6 Q2: 你更偏向在什么场合穿帆布鞋?

7 C2: ["旅行", "舞蹈", "健身", "建筑工地", "官方旗舰店", "其他"]

8 Q3: 在选择帆布鞋时, 你重视哪些功能?

9 C3: ["防滑", "增高 10 厘米", "防水鞋底", "防穿刺", "抗菌", "其他"]

10

11 U: Answers for Q1: "军绿色".

12 Answers for Q2: "旅行".

13 Answers for Q3: "其他".

14

15 A: 我已经为你搜索了几个相关的物品: <物品 1>, <物品 2>, . . . . . 此外, 你可以回答这些问

16 Q1: 在选择帆布鞋时, 你更喜欢哪种风格?

17 C1: ["休闲", "内增高", "系带", "平跟", "低帮", "其他"]

18 Q2: 你通常关注鞋面的哪些装饰属性?

19 C2: ["粘贴", "透气", "内里", "网面", "鞋底", "其他"]

20 Q3: 在购买帆布鞋时, 你更喜欢用哪种材质制作它们?

21 C3: ["网布", "橡胶", "帆布", "其他"]

22

23 U: Answers for Q1: "休闲".

24 Answers for Q2: 咖啡豹纹, 豹纹, 圆头。

25 Answers for Q3: "帆布".

26

27 A: 我已经为你搜索了几个相关的物品: <物品 1>, <物品 2>, . . . . . 此外, 你可以回答这些问

28 ...

29

30

31 Ground truth product item:

32 标题:2023 年春季女款飞跃厚底帆布休闲百搭馒头鞋 8925 号

33 类别: 帆布鞋

34 品牌: ["飞跃"]

35 系列: []

36 目标客户: ["女性", "18-40 岁", "青年"]

37 适用场景: ["春季"]

38 装饰属性: ["鞋底", "厚底", "横带", "咖啡色豹纹", "浅口", "豹纹", "图案", "圆头", "卡

39 材料: ["橡胶", "帆布"]

40 风格: ["系带", "低帮", "豹纹", "休闲", "面包", "馒头", "百搭", "硫化工艺"]

41 规格: ["EUR34", "EUR35", "EUR36", "EUR37", "EUR38", "....."]

42 颜色: [“紫色”, “湖蓝和橙色”, “军绿和白色”, “粉色”, “混色”]

43 功能: [ ]

END