

# 亚马逊COSMO：LLM构建高质量电商知识图谱

原创 方方 方方的算法花园 2024年11月12日 08:46 北京

## 0 ▶ 论文概况

### 1. 论文名称：

COSMO: A Large-Scale E-commerce Common Sense Knowledge Generation and Serving System at Amazon

《COSMO：亚马逊大规模电子商务常识知识生成与服务系统》

2. 论文链接：<https://dl.acm.org/doi/10.1145/3626246.3653398>

3. 论文作者所在机构：亚马逊公司、香港科技大学

4. 一句话概括：该论文提出了 COSMO 系统，它通过从大规模用户行为中挖掘用户中心的常识知识来构建知识图谱，利用LLM和指令微调生成高质量知识，应用于多种电子商务任务，有效提升了购物体验和业务指标。

## 1 ▶ 挑战



Figure 1: An example of mining implicit commonsense knowledge from e-commerce user behaviors.

### 1. 意图理解与提取

(1) **用户意图不明确**：在电子商务场景中，用户意图通常不直接表达，需要借助常识来理解，但机器难以从海量且嘈杂的用户行为数据中提取这些隐含意图。

(2) **现有方法的局限性**：以往方法如 FolkScope 在提取常识知识时，虽能通过提示LLM从用户行为中发现知识，但存在诸多局限。例如，仅探索了少量类别中的共同购买意图，未充分考虑搜索购买等多样化行为产生的复杂意图；且在实际生产环境中，面对每天数百万用户产生的复杂、多样且嘈杂的行为数据，难以有效选择代表性行为来生成多样化意图。

### 2. 知识生成与优化

(1) **知识质量与偏好对齐**：从语言模型生成的知识候选可能不符合人类偏好，存在通用性强但不具体、不准确或无用的情况，如生成“顾客一起购买是因为喜欢”等缺乏针对性的意图解释，而理想的生成应能典型地解释电子商务行为。

**(2) 数据驱动的关系发现：**由于计算资源限制，无法简单采用以往方法对齐常识关系，需要从大规模生成中进行数据驱动的关系发现，以满足电子商务场景需求并使生成的知识结构化。

### 3.数据标注与效率

**(1) 标注成本与数据量平衡：**随着电子商务场景的拓展，准确标注数据以训练高质量模型变得愈发困难。如 FolkScope 中对生成知识进行精细标注（分别标记合理性和典型性分数）的成本随着类别和用户行为类型的增加而显著上升，同时大规模数据的处理效率也面临挑战。

**(2) 推理效率瓶颈：**在应用 FolkScope 等方法于下游任务时，推理开销可能成为瓶颈。例如，对于新用户行为的知识生成，需经过大语言模型生成和分类器评分的流程，像 OPT - 30b 这样的大语言模型计算成本高，不适用于在线服务。

### 4.系统部署与实时性

**(1) 满足搜索延迟要求：**在将系统部署到实际电子商务应用中时，需要满足严格的搜索延迟要求，确保用户体验不受影响，同时控制存储成本。

**(2) 处理实时信息的挑战：**尽管系统每日刷新模型，但在处理如限时抢购等实时信息时仍存在局限，难以快速响应和处理短时间内快速变化的信息，影响系统在快节奏电子商务活动中的敏捷性。

## 2 ▶ 论文贡献点

### 1.构建大规模行业知识系统

率先采用LLM构建高质量知识图谱，并将其应用于在线电子商务服务，填补了行业空白，为电子商务领域的知识图谱构建和应用提供了新的范例。

Table 1: Comparison among existing commonsense knowledge graphs. ‘Rel’ represents relation types. Our new KG covers more nodes and edges in more domains compared to existing e-commerce related KGs for intention understanding.

KG	# Nodes	# Edges	# Rels	Source	Node Type	E-commerce	Intention	User Behavior
ConceptNet [30]	8M	21M	36	Crowdsourc <sup>1</sup>	concept	✗	✓	✗
ATOMIC [25]	300K	870K	9	Crowdsourc	daily situation, event	✗	✓	✗
AliCoCo [13, 14]	163K	813K	91	Extraction	concept	✓	✗	search logs
AliCG [47]	5M	13.5M	1	Extraction	concept, entity	✗	✗	search logs
FolkScope [45]	1.2M	12M	19	LLM Generation	product, intention	2 domains	✓	co-buy
COSMO (Ours)	6.3M	29M	15	LLM Generation	product, query, intention	18 domains	✓	co-buy&search-buy

### 2.创新指令调优方法

通过指令调优生成电子商务常识知识，使语言模型能更好地与人类偏好对齐，有效解决了从语言模型生成的知识可能不符合人类实际需求的问题，提高了生成知识的准确性和实用性。

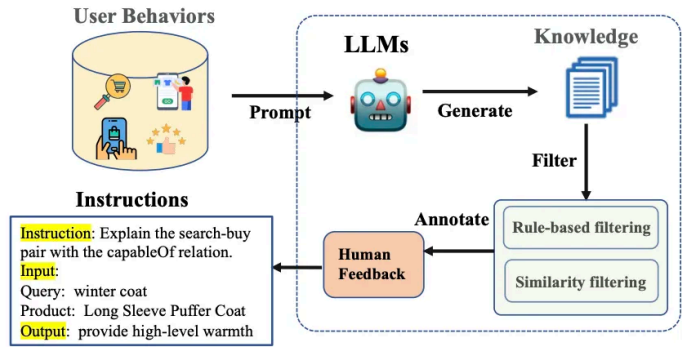
### 3.优化数据收集与处理

设计了基于大规模用户行为的自动指令生成pipelines，通过扩展产品领域、关系类型和微调任务，实现了从数百万用户行为中高效提取电子商务意图知识，并以较少的标注工作量生成高质量的指令数据，显著提升了数据处理效率和知识提取的可扩展性。

### 4.成功应用于多种任务

将生成的意图知识广泛应用于搜索相关性、基于会话的推荐和搜索导航等多个实际电子商务任务中，实验结果表明，这些应用取得了显著的改进，充分展示了该方法在电子商务领域的巨大潜力和广泛适用性，为提升电子商务服务质量和用户体验提供了有力支持。

## 3 ▶ COSMO 框架



**Figure 2: Overall framework of generating high-quality instruction data from massive user behaviors and large language models.**

**初步定义**

- 1.用户行为 (User Behaviors)：** 电子商务平台每天产生海量用户行为日志，选择搜索 - 购买 (search - buy) 和共同购买 (co - buy) 这两种典型行为进行研究。搜索 - 购买行为定义为用户在短时间内点击查询并购买产品的查询 - 产品对；共同购买行为则用共同购买的产品对表示，且每个产品可归为一个主要领域。
- 2.常识知识 (Commonsense Knowledge)：** 利用关系感知提示 (relation - aware prompts) 让大语言模型将用户行为解释为知识候选，以三元组 (h, r, t) 表示，其中 r 和 t 分别表示关系和尾实体。通过数据驱动的关系发现方法，从大规模生成中挖掘知识关系类型和尾类型，以满足电子商务场景需求并结构化生成的知识。

**Table 2: Mined e-commerce commonsense relations for the COSMO KG.**

Relation Type	Tail Type	Example
USED_FOR_FUNC	Function / Usage	dry face
USED_FOR_EVE	Event / Activity	walk the dog
USED_FOR_AUD	Audience	daycare worker
CAPABLE_OF	Function / Usage	hold snacks
USED_TO	Function / Usage	build a fence
USED_AS	Concept / Product Type	smart watch
IS_A	Concept / Product Type	normal suit
USED_ON	Time / Season / Event	late winter
USED_IN_LOC	Location / Facility	bedroom
USED_IN_BODY	Body Part	sensitive skin
USED_WITH	Complementary	surface cover
USED_BY	Audience	cat owner
xINTERSTED_IN	Interest	herbal medicine
xIs_A	Audience	pregnant women
xWANT	Activity	play tennis

**3.指令数据 (Instruction Data)：** 指令数据是一组用自然语言定义任务的集合，每个任务包含输入 - 输出对实例。对于常识生成任务，输入为用户行为对，输出为典型知识尾实体，其质量由人类标注的合理性和典型性分数衡量，选择高典型性分数的知识作为期望的模型输出，并添加辅助任务训练语言模型以实现知识泛化和在线服务。

**知识生成**

**1.用户行为采样 (User Behavior Sampling)：**为生成多样化和高质量知识，从产品采样和行为对采样两方面进行精细采样。涵盖亚马逊常见热门类别选择产品，采用产品类型标签采样，并通过多种策略（如阈值设置、查询特异性考虑、随机采样排除等）对共同购买和搜索 - 购买行为对进行采样，最终获得数百万行为对，且采样具有多样性。

Table 3: Statistics of COSMO knowledge graph including the sampled user behavior pairs, annotated knowledge candidates, and remaining edges after knowledge refinement.

Category	Co-buy			Search-buy		
	# Behavior Pairs	# Annotations	# Edges	# Behavior Pairs	# Annotations	# Edges
Clothing, Shoes & Jewelry	233,989	1303	2,147,605	176,018	597	887,130
Sports & Outdoors	251,713	1302	2,140,491	126,130	970	556,233
Home & Kitchen	426,070	1991	3,380,502	225,377	3798	1,054,764
Patio, Lawn & Garden	117,871	542	908,158	56,754	263	280,932
Tools & Home Improvement	258,480	1184	1,988,346	122,613	585	629,004
Musical Instruments	24,206	84	174,238	9,385	24	33,786
Industrial & Scientific	385,990	1820	3,002,352	177,400	1317	814,266
Automotive	166,234	782	1,330,580	55,201	456	258,340
Electronics	178,938	777	1,316,937	119,764	768	549,716
Baby Products	111,204	430	721,727	30,156	38	135,702
Arts, Crafts & Sewing	13,1131	616	1,095,531	62,135	232	274,015
Health & Household	233,945	1198	1,906,447	215,349	67	930,307
Toys & Games	148,455	646	1,165,692	73,512	536	291,107
Video Games	16,436	60	106,449	10,306	30	29,681
Grocery & Gourmet Food	99,660	504	775,016	116,765	2123	577,986
Office Products	136,519	650	1,086,735	79,470	2063	364,767
Pet Supplies	43,541	206	302,839	51,807	1122	219,143
Others	182,738	905	1,351,257	160,189	11	648,765
Total	3,147,120	15,000	24,900,902	1,868,331	15,000	5,093,795

**2.问答式生成 (QA - Prompted Generation)：**利用大语言模型在回答情境化问题方面的能力，通过提供问答上下文（如特定的任务描述、查询和产品信息）来生成知识候选。使用 OPT175b 和 OPT30b 模型在 16 个 A100 GPU 上进行生成推理，并通过在问题后添加数字字符“1”来生成知识候选列表。

Task: Please provide typical explanation for the following search-purchase behavior and complete the answer.

Search Query: {Query}

Product: {Product Title}

Question: what is the product capable of, which exactly match the intention of the search query?

Answer: the query means customers want the product that is capable of

1.

Figure 3: Prompts used for generating knowledge candidates.

知识精炼

1.粗粒度过滤 (Coarse - grained Filtering)

(1) **规则过滤 (Rule - based Filtering)：**借助自然语言处理工具和语言模型 (GPT - 2) 计算困惑度得分来过滤不完整句子，同时直接排除与查询、产品类型或标题相同（或编辑距离小于阈值）的生成内容，以及通过频率和熵识别通用知识情况。

(2) **相似性过滤 (Similarity Filtering)：**使用预训练的内部电子商务语料库语言模型获取生成知识尾实体、查询和产品的嵌入，通过计算余弦相似度来处理语义相似情况，过滤掉本质上是原始用户行为上下文的释义且具有句法转换的生成内容，保留典型知识。

2.人工标注 (Human - in - the - loop Annotation)

- (1) 为平衡知识候选数量和成本，结合知识频率和产品或查询的流行度进行加权采样，对两种用户行为各采样 15000 个知识候选进行标注。
- (2) 采用专业数据标注公司和严格内部审计，通过分解合理性和典型性判断为五个明确问题（完整性、相关性、信息性、合理性、典型性）来降低标注者认知负担和分歧率，对标注数据进行随机抽样审计确保准确性，构建分类模型对粗粒度过滤后的知识候选评分，保留合理性分数高于 0.5 的知识，以相对低成本获得高质量电子商务知识。

指令调优的 COSMO 语言模型

- 1.收集 30000 个多样知识样本的人类判断后，基于标注数据创建大规模指令数据，涵盖 18 个产品领域、15 种关系类型和 5 种不同任务类型。
- 2. 发现搜索 - 购买生成的典型知识比例较高可直接用作指令数据，而共同购买的典型比例低。期望微调后的语言模型具备生成典型知识和预测知识质量的能力，为此将所有标注转换为指令数据用于相关任务。考虑到用户行为数据中的噪声，将共购买预测和搜索相关性预测纳入微调任务，设计不同模板使模型对不同格式具有鲁棒性，使用收集的指令数据微调 LLaMA 7b 和 13b 模型。

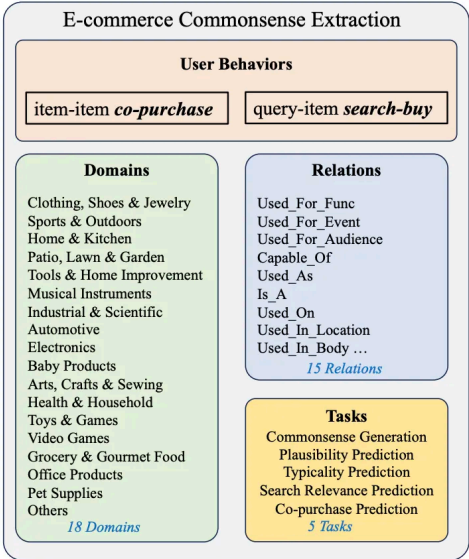
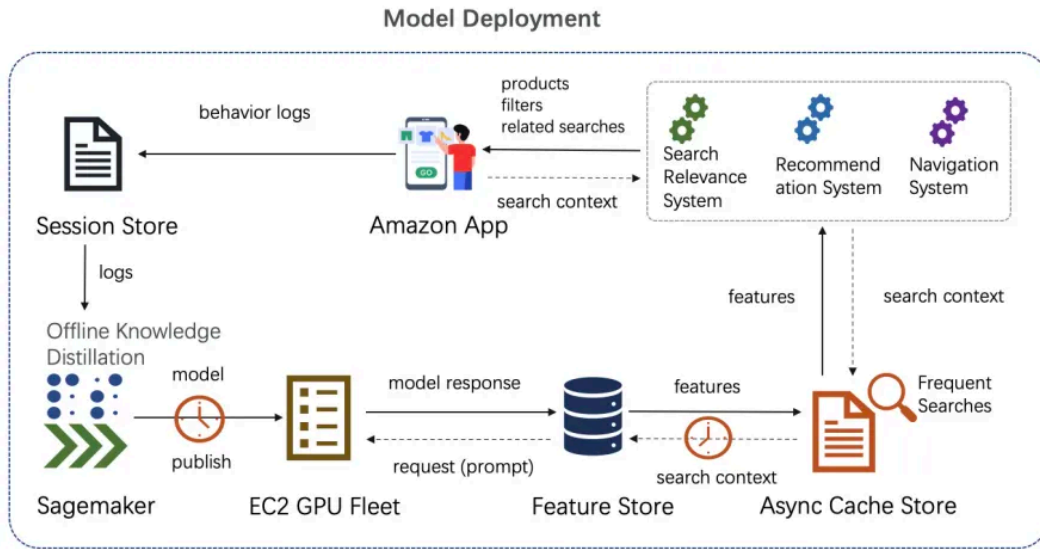


Figure 4: Illustration of finetuning COSMO-LM to generate e-commonsense knowledge from two typical user behaviors. We scale up product domains, relation types and tasks.

在线部署





**Figure 5: Illustration of COSMO-LM deployment, featuring the Asynchronous Cache Store and Feature Store as central components. It depicts the efficient processing of user queries and dynamic daily updates, crucial for meeting Amazon’s search latency requirements.**

**1.部署策略 (Deployment Strategy)**

(1) **部署管理**：使用 SageMaker 刷新 COSMO - LM 模型，实现客户行为会话日志的动态摄取和模型的高效更新。

(2) **特征存储集成**：将模型响应转换为结构化特征，供下游应用使用，处理产品键值对、语义子类别表示和强意图检测等功能。

(3) **异步缓存存储**：采用两层缓存策略管理频繁搜索，结合预加载的年度频繁搜索和批量处理的每日请求，适应每日流量模式。

**2. 操作流程 (Operational Flow)**

(1) **模型部署**：将 COSMO - LM 部署在 SageMaker 上处理用户行为会话日志并更新模型。

(2) **请求处理**：初始查询先在异步缓存存储中检查，频繁查询直接获取响应，其他查询则转发进行批量处理。

(3) **批量处理和缓存更新**：特征存储将语言模型响应格式化为结构化洞察，更新缓存以供未来查询使用。

(4) **与下游应用通信**：缓存中的结构化数据增强下游应用，提供丰富特征改善用户交互。

(5) **反馈循环**：将用户交互反馈给 COSMO - LM 以持续改进模型，确保对不断变化的用户行为做出及时响应。

**3. 影响与限制**：该部署策略有效满足亚马逊的搜索延迟要求，同时保持与实时服务相当的存储成本，显著提升了在线请求管理能力。然而，尽管每日刷新模型，但在处理实时信息（如限时抢购）方面仍存在局限性，此类时间敏感事件的快速变化对系统的敏捷性构成挑战，凸显了进一步发展以增强系统应对电子商务活动动态变化能力的必要性。

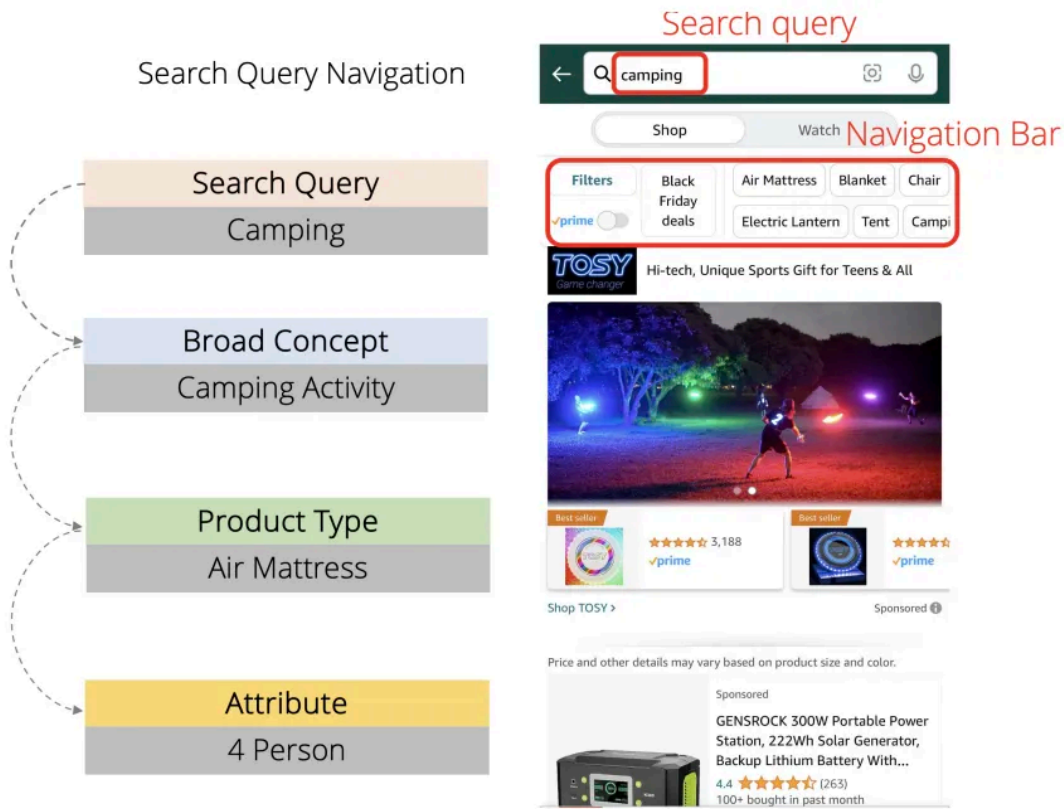


Figure 9: Search Navigation Experience using COSMO

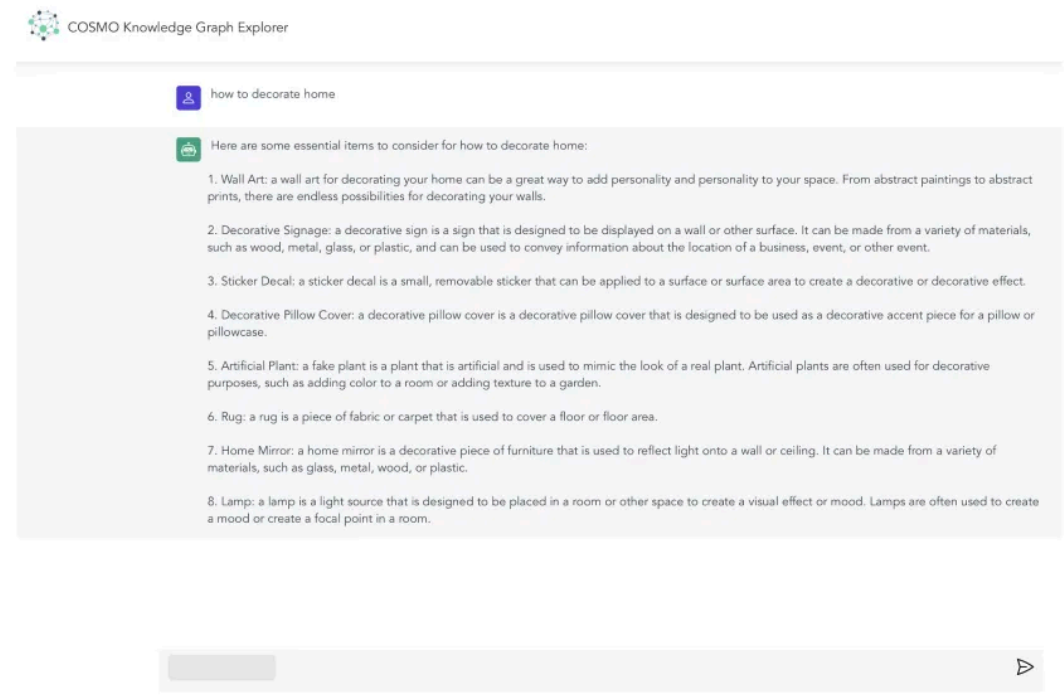


Figure 10: An example of generation from COSMO-LM

END