

2024阿里：DUIN框架-心有灵犀，解锁推荐系统中的显性与隐性用户意图之谜



SmartMindAI 专注搜索、广告、推荐、大模型和人工智能最新技术，欢迎关注我

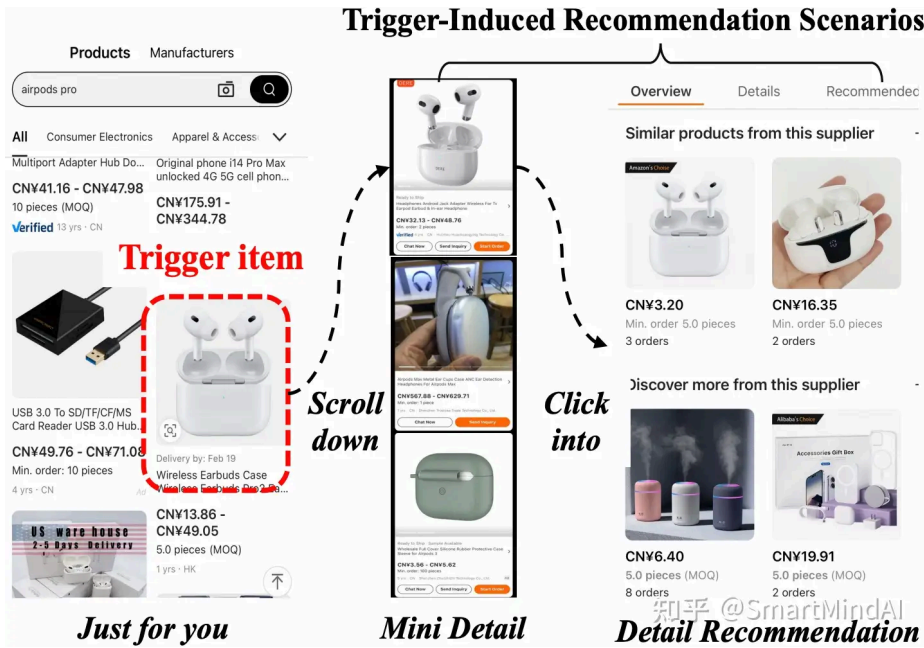
已关注

47 人赞同了该文章

收起

Introduction

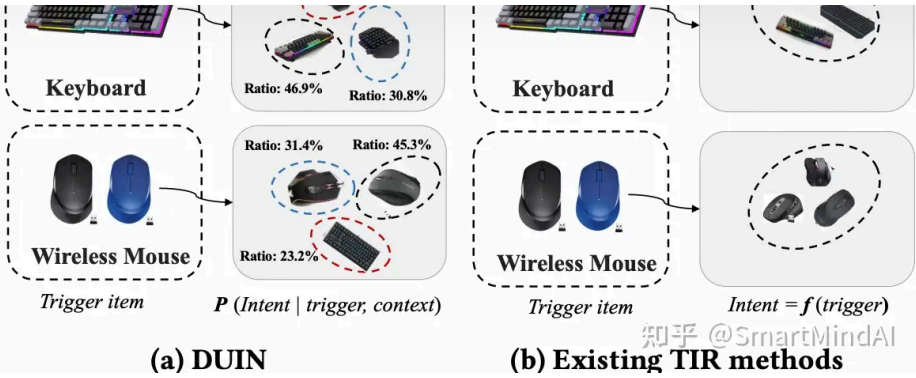
触发引荐 (Trigger-Induced Recommendation, TIR)，近期在行业引起了广泛关注，因为它能够通过建模用户在触发内容瞬间的点击历史，提供沉浸式的用户体验。例如，作为全球最大的电子商务平台之一，[阿里巴巴](#)⁺ 提供了许多TIR场景。如图所示，用户在"专属为您"场景中点击触发内容后，可以访问一些TIR场景，例如"迷你详情"或"详细推荐"。



TIR的主要目标是在点击触发内容的情况下，推荐满足用户需求的内容。如今，TIR在许多行业中扮演着越来越重要的角色。在阿里巴巴网站上，超过一半的活跃买家都是通过TIR在所有推荐场景中被贡献的。

传统的推荐建模方法往往侧重于使用各种[深度神经网络](#)⁺来建模用户的时序行为，忽视了触发内容的影响，因此无法直接应用于TIR场景。最近，已经提出了几种基于触发的方法，这些方法在TIR任务中展示了相对于传统方法的优越性。然而，现实世界电子商务平台上的用户行为更为复杂。如图展示了用户在点击触发内容后的后续行为的一个有洞察力的例子。

知乎



当用户点击了一个键盘，46.9%的用户随后购买了相同或相似的产品，比如另一个键盘，而30.8%的用户购买了热门商品（例如，无线键盘），其余的22.3%的用户购买了互补选择（例如，无线鼠标）。这一观察表明，用户在点击触发内容后的意图存在一定程度的不确定性，这也意味着平台需要提供更精确和全面的采购推荐。

用户点击触发内容后的后续行为展示了用户需求的多样性，平台应据此提供个性化推荐，以满足不同用户的具体需求。然而，如图1所示，现有的TIR方法过度依赖触发项，忽视了对用户意图的更深层次探索和利用，导致对用户意图的理解过于集中和收敛，用户只能从触发项中获得一小部分推荐内容。对于购买电脑配件的用户来说，键盘和无线鼠标等产品符合他们的购买意图，过度狭窄的推荐可能会阻碍他们长期购买体验的提升。此外，用户意图受到多种因素的影响，包括用户的历史交互和浏览上下文。这种复杂性在准确建模用户意图和捕捉相关不确定性方面提出了重大挑战。

为了解决这些挑战，我们在本文中提出了一种名为DUIN的新模型，简称DeepUncertainty Intent Network。它主要由三个部分组成：

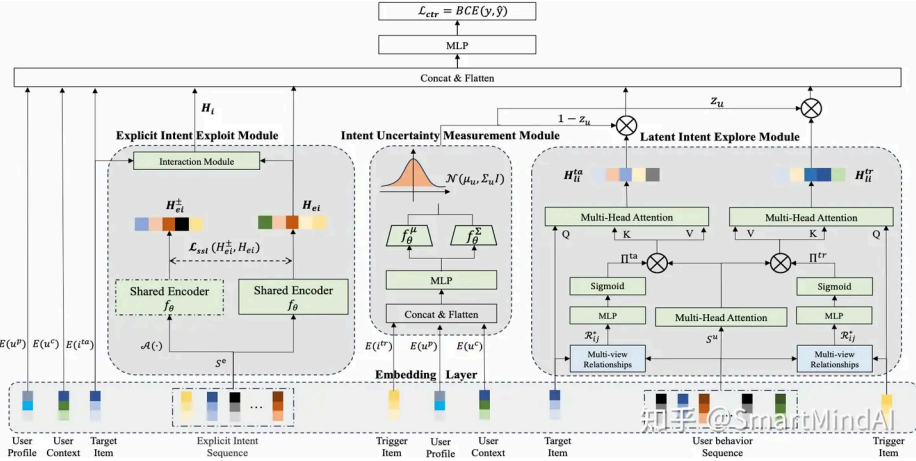
为了从触发项中提取明确的用户意图，我们设计了一个明确意图利用模块（EIEM），它采用对比学习范式来获取通用且可区分的意图表示。

我们还设计了一个潜在意图探索模块（LIEM），通过利用多视角关系有效地探索用户的潜在意图，适用于触发项和目标项。这使得我们能够利用触发项和目标项之间的关系来更好地理解用户的意图。

为了建模用户意图强度并捕捉与用户意图相关的不确定性，我们实现了一个意图不确定性度量模块（IUMM），它将用户意图强度建模为高斯分布⁺。

Method

本文提出的方法，由三个模块组成：1) 显式意图利用模块（EIEM），采用对比学习框架，根据触发项提取用户明确的意图；2) 潜在意图探索模块（LIEM），负责基于多个共现视图探索用户的潜在意图；3) 意图不确定性测量模块（IUMM），在潜在空间中给出分布估计，并将不确定性注入到意图建模中。



知乎

让 \mathcal{U} 代表用户集合 \mathcal{I} 代表内容集合。每个用户有一个行为序列 \mathcal{S}^u ，在给定触发内容 i^{tr} 的情况下，行为序列 \mathcal{S}^u 包含行为 $i_1^u, i_2^u, \dots, i_T^u$ ，其中 M 、 N 和 T 分别表示用户数量、内容数量和行为序列的长度。我们用 u^p 表示用户特征，用 u^c 表示环境特征或情境特征，TIR的主要目的是预测用户 u 与目标内容 i^{ta} 互动的可能性如下：

$$\hat{y} = \mathcal{F}_\theta(\mathbf{E}(u^p), \mathbf{E}(u^c), \mathbf{E}(i^{tr}), \mathbf{E}(i^{ta}), \mathbf{E}(\mathcal{S}^u))$$

其中 \mathcal{F}_θ 表示我们旨在学习的预测函数⁺， $\mathbf{E}(\cdot) \in \mathbb{R}^d$ 表示嵌入层，它将原始特征转化为维度为 d 的稠密向量。

Deep Uncertainty Intent Network

深度不确定性意图网络（DUIN）的整体架构由三个模块构成：明确意图的利用、潜在意图的探索以及意图不确定性测量。

Explicit Intent Exploit Module

用户按时间序列⁺排列的行为表现出因果关系，并传递明确的意图信息，包括用户倾向于互动的方面，如品牌、主题和类别。触发内容仅代表用户明确意图的部分表达。在我们的方法中，我们引入了明确意图利用模块，以学习用户明确的意图。

通常，具有相同特性的内容展现出高度的相似性。例如，在电子商务场景中，属于同一类别的产品往往具有相似的内在质量和互动用途。在用户行为序列中，我们考虑与触发内容具有相同特性的内容表达用户明确的意图。给定触发内容 i^{tr} ，我们从行为序列 \mathcal{S}^u 中检索与 i^{tr} 具有相同特性的内容。这些内容与触发内容一起构成了明确意图序列 \mathcal{S}^e ：给定触发内容 i^{tr} ，从行为序列 \mathcal{S}^u 中检索与 i^{tr} 具有相同特性的内容，这些内容与触发内容一起构成了明确意图序列 \mathcal{S}^e ：

$$\mathcal{S}^e = \{i^{tr}, i_1^e, \dots, i_l^e, \dots, i_L^e\}$$

其中 i_l^e 表示与触发内容 i^{tr} 具有相同属性的内容。 L 表示 i_l^e 的数量。

各种基于注意力的方法显著促进了用户行为建模⁺。然而，近期的研究表明，这些方法生成的用户表示倾向于退化为非对称形状，这可能导致表示之间的高语义相似性。为了解决这个问题，受到计算机视觉⁺领域对比学习范式近期进展的启发，基于明确意图序列，我们应用对比学习来获取具有辨别性和高质量的用户明确意图表示。正式地，明确意图对比学习任务的定义是：

$$f_\theta = \arg \min_{f_\theta} \mathcal{L}_{ssl}(f_\theta(\mathcal{S}^e), f_\theta(\tilde{\mathcal{S}}_+^e))$$

对比损失函数⁺ \mathcal{L}_{ssl} ，在实践中，我们以一定概率对显式意图序列 \mathcal{S}^e 进行掩码操作，以生成其增强视图 $\tilde{\mathcal{S}}_+^e \sim \mathcal{A}(\mathcal{S}^e, \gamma)$

其中 $\mathcal{A}(\cdot)$ 表示增广操作符 γ 是掩码⁺概率。背后的理念是，用户明确意图在一段时间内相对稳定。因此，尽管有些内容被掩码了，但明确意图的信息仍然保留，并且应该被视为正面视图的保留。因此，我们将 $(\mathcal{S}^e, \tilde{\mathcal{S}}_+^e)$

视为正面视图配对，而同一批中的其他增广示例则视为负面示例或负面视图。

此外，我们采用共享编码器 f_θ 来提取用户明确意图的表示： $\mathcal{H}_{ei} = f_\theta(\mathcal{S}^e)$

我们通过对比损失函数来优化 f_θ ，该函数可以表述为：

$$\mathcal{L}_{ssl} = -\log \frac{e^{\text{sim}(\mathcal{H}_{ei}, \mathcal{H}_{ei}^+)/\tau}}{e^{\text{sim}(\mathcal{H}_{ei}, \mathcal{H}_{ei}^+)/\tau} + \sum_{i=1}^{2B-1} e^{\text{sim}(\mathcal{H}_{ei}, \mathcal{H}_{ei}^-)/\tau}}$$

其中 \mathcal{H}_{ei}^+ 和 \mathcal{H}_{ei}^- 分别表示对 \mathcal{H}_{ei} 的增强正向视图和负向视图。 $\text{sim}(\cdot)$ 表示余弦相似度函数。 B 是批大小。 τ 是温度参数。值得一提的是，EIEM 还引入了一个特征交互模块来学习目标内容 i^{ta} 和用户明确意图 \mathcal{H}_{ei} 之间的明确交互关系，输出表示 \mathcal{H}_i 是由 \mathcal{H}_{ei}^+ 和 \mathcal{H}_{ei}^- 的余弦相似度⁺计算得出的：

$$\mathcal{H}_i = \text{sim}(\mathcal{H}_{ei}^+, \mathcal{H}_{ei}^-)$$

Latent Intent Explore Module

正如之前所述，用户在点击触发项后具有各种潜在意图，例如寻找相似内容、热门内容或互补内容。之前的潜在意图识别方法缺乏对用户潜在意图的探索，忽视了用户行为序列中与意图相关的潜在信息。在我们提出的方法中，我们引入了潜在意图探索模块来揭示用户的潜在意图。

我们考虑具有相似需求的用户通常表现出相似的行为，因此，我们将多个用户的时序定向行为聚合形成基本的概率图结构 \mathcal{G} 。图中的节点 \mathcal{V} 包含内容及其属性，我们将连接两个连续发生的节点以细化合作关系。

在获得了图 \mathcal{G} 中更细致的合作模式后，假设有一个具有属性 ϕ 的内容 i 和一个具有属性 ψ 的内容 j ，我们通过以下方式构建内容 i 和内容 j 之间的多视角关系：1. 构建内容 i 和内容 j 之间的多视角关系。因此，通过这些多视角关系的构建，我们能够更深入地理解用户的**行为模式**⁺和潜在需求，从而优化**推荐系统**⁺，提供更符合用户期待的内容推荐。

$$\begin{aligned}\mathcal{R}_{ij}^t &= \mathcal{C}_{i \rightarrow j}(\{(\mathcal{V}_i, \mathcal{V}_j) | (\mathcal{V}_i, \mathcal{V}_j) \in \mathcal{G}\}), \\ \mathcal{R}_{ij}^c &= \mathcal{C}_{\phi \rightarrow \psi}(\{(\mathcal{V}_\phi, \mathcal{V}_\psi) | (\mathcal{V}_\phi, \mathcal{V}_\psi) \in \mathcal{G}\}), \\ \mathcal{R}_{ij}^p &= \mathcal{C}_{\phi \rightarrow j}(\{(\mathcal{V}_\phi, \mathcal{V}_j) | (\mathcal{V}_\phi, \mathcal{V}_j) \in \mathcal{G}\}),\end{aligned}$$

其中 \mathcal{V}_\star 表示在图 \mathcal{G} 中的注释。 \mathcal{C}_\star 是聚合操作符，记录两个节点的共现频率。

上述多视角关联可以隐式捕获内容的高阶联系。具体地，我们使用 $\mathcal{C}_{i \rightarrow j}$ 来计算内容 i 和内容 j 之间的共现频率，使用 \mathcal{R}_{ij}^t 来代表它们的定向转换关系，这种关系是直观的。同样，属性之间的协作关系，由 \mathcal{R}_{ij}^c 表示，代表了内容 i 和内容 j 之间的互补关系。 \mathcal{R}_{ij}^p 用于描述在属性 ϕ 下，内容 j 的流行度。

对于内容 i 和内容 j ，我们通过多视角关联更好地编码它们之间的信息，以量化它们的潜在联系，即定义隐含意图相关分数 Π_{ij} ： Π_{ij} = 潜在联系的隐含意图相关分数

这有助于我们理解内容之间的复杂关联，包括但不限于它们的共现、定向转换、互补属性，以及在特定属性下的流行度。

$$\Pi_{ij} = \text{Sigmoid}(\text{MLP}(\mathbf{E}(\mathcal{R}_{ij}^t), \mathbf{E}(\mathcal{R}_{ij}^p), \mathbf{E}(\mathcal{R}_{ij}^c)))$$

受到先前工作的启发，我们利用多头自注意力（MHSA）来获取细化的行为表示 $\mathbf{E}(\mathcal{S}^u)$ 。接下来，我们使用两个MHSA模块，分别提取用户对目标内容和触发内容的隐式意图。更关键的是，我们提出了一种简单而有效的方法，以使**注意力机制**⁺意识到内容之间的先验多视图关系。具体来说，对于每个行为记录 $i_\star^u \in \mathcal{S}^u$ ，我们根据公式分别计算了与触发内容和目标内容之间的隐式意图相关性分数，以增强注意力机制对内容间多视图关系的理解。

$$\begin{aligned}\mathcal{H}_{ii}^{tr} &= \text{Attention}(\mathbf{E}(i^{tr}), \Pi^{tr} \mathbf{E}(\mathcal{S}^u), \Pi^{tr} \mathbf{E}(\mathcal{S}^u)), \\ \mathcal{H}_{ii}^{ta} &= \text{Attention}(\mathbf{E}(i^{ta}), \Pi^{ta} \mathbf{E}(\mathcal{S}^u), \Pi^{ta} \mathbf{E}(\mathcal{S}^u)),\end{aligned}$$

\mathcal{H}_{ii}^{tr} 和 \mathcal{H}_{ii}^{ta} 分别代表用户针对触发内容和目标内容的潜在意图表示。

其中， Π^{tr} 表示行为项与触发内容之间的隐含意图相关分数。

而 Π^{ta} 则表示行为项与目标内容之间的隐含意图相关分数。

Intent Uncertainty Measurement Module

在 TIR 任务中，假设有一个**连续映射**⁺空间： $\mathcal{X} \rightarrow \mathcal{Z}$

其中 $\mathbf{x}_u \in \mathcal{X}$ 表示用户个性化信息，包括触发物品 i^{tr} 、上下文特征 \mathbf{u}^c 和个人特征 \mathbf{u}^p ， $\mathbf{z}_u \in \mathcal{Z}$ 表示用户意图强度。随着用户意图的演变和浏览上下文的变化，点击触发物品后的用户意图强度会变得不确定。因此，在 DUIN 中，我们将用户交互行为视为受用户意图强度影响的不确定事件。我们使用意图不确定性测量模型将用户意图强度建模为一个分布，而不是静态值，这可以覆盖更广泛的用户意图空间，并为推荐结果注入新颖性和不确定性。具体来说，我们将每个用户的意图强度 \mathbf{z}_u 定义为高斯分布 \mathcal{N} ，其数学定义如下：

其中 μ_u 和 Σ_u 分别表示高斯分布的均值和方差 μ_u 表示预测的用户意图强度 Σ_u 可以视为 μ_u 的不确定性。为了简化，我们仅考虑**对角矩阵**⁺。实际上，我们使用两个独立的网络来获取 μ_u 和 Σ_u 。

$$\begin{aligned}\mu_u &= f_{\theta}^{\mu}(x_u), \\ \Sigma_u &= \text{softplus}(f_{\theta}^{\Sigma}(x_u)),\end{aligned}$$

其中 f_{θ}^{μ} 和 f_{θ}^{Σ} 被实现为两个**全连接层**⁺。softplus是一个**激活函数**⁺，并约束输出始终为正。我们将 z_u 视为用户对触发项意图强度的度量，它整合了由LIEM输出的用户潜在意图特征，可以表示为：

$$\mathcal{H}_{li} = (z_u \odot \mathcal{H}_{li}^{tr}; (1 - z_u) \odot \mathcal{H}_{li}^{ta})$$

Prediction and Optimization

从EIEM和LIEM中获取用户意图表示 \mathcal{H}_{ei} 、 \mathcal{H}_i 和 \mathcal{H}_{li} 后，类似于equation $eq:ctr$ (reference-type="ref" reference="eq:ctr")，我们把这些意图表示与上下文特征 u^c 、用户画像特征 u^p 以及其他特征进行拼接。接着，拼接后的特征被输入到**多层感知器**⁺（MLP）中，以估计用户交互的可能性 \hat{y}_i 在区间 $[0, 1]$ 内。我们采用二元**交叉熵**⁺损失函数用于模型优化，以提升模型的准确性：

$$\mathcal{L}_{ctr} = -\frac{1}{N} \sum_{i=1}^N (y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i))$$

$$\mathcal{L}_{final} = \mathcal{L}_{ctr} + \alpha \cdot \mathcal{L}_{ssl}$$

其中 α 是一个用于调整模型参数的超参数，它专门用于平衡对比度损失的权重。

Evaluation

Datasets.

在实验中使用了三个真实世界的数据集：阿里巴巴电子商务网站、**阿里妈妈**⁺（Alimama）、内容智慧的ContentWise。

Overall Comparison

Model	Alibaba.com		Alimama		ContentWise	
	AUC	RelaImpr	AUC	RelaImpr	AUC	RelaImpr
WDL	0.6096 ± 0.0019	−0.99%	0.6062 ± 0.0008	−7.97%	0.9469 ± 0.0003	−7.28%
DIN	0.6042 ± 0.0016	−5.87%	0.6154 ± 0.0007	0.00%	0.9774 ± 0.0002	−0.95%
DIEN	0.6047 ± 0.0025	−5.42%	0.6155 ± 0.0005	0.09%	0.9779 ± 0.0013	−0.85%
DMIN	0.6107 ± 0.0011	0.00%	0.6154 ± 0.0002	0.00%	0.9820 ± 0.0002	0.00%
WDL ^{TIR}	0.7412 ± 0.0014	111.89%	0.6075 ± 0.0018	−6.84%	0.9713 ± 0.0004	−2.22%
DIN ^{TIR}	0.7425 ± 0.0021	119.06%	0.6155 ± 0.0015	0.09%	0.9803 ± 0.0019	−0.35%
DIEN ^{TIR}	0.7419 ± 0.0019	118.52%	0.6157 ± 0.0004	0.26%	0.9796 ± 0.0015	−0.50%
DMIN ^{TIR}	0.7454 ± 0.0007	121.68%	0.6157 ± 0.0003	0.26%	0.9822 ± 0.0003	0.04%
DIHN	0.7462 ± 0.0006	122.40%	0.6166 ± 0.0008	1.04%	0.9786 ± 0.0012	−0.75%
DIAN	0.7480 ± 0.0016	124.03%	0.6168 ± 0.0002	1.21%	0.9764 ± 0.0003	−1.16%
DEI2N	0.7671 ± 0.0012	141.28%	0.6180 ± 0.0005	2.25%	0.9840 ± 0.0002	0.41%
DUIN	0.7782 ± 0.0014	151.31%	0.6194 ± 0.0003	3.47%	0.9881 ± 0.0004	1.27%

在三个数据集上的所有方法的总体性能总结在表中。从结果中，我们可以观察到，传统的CTR方法在与其他方法相比时，提供了最差的结果，这表明它们在TIR任务中的不足。通过引入触发内容，这些传统方法的升级版显示了显著改进的性能。例如，DMIN^{TIR}在阿里巴巴数据集上的AUC提升为22.1%，这表明在TIR场景中对触发内容进行详尽建模是必要的。

同时，现有的TIR方法，DIHAN，DIAN和DEI2N，通过精心设计的网络进一步超越了先前升级的传统方法的性能。值得注意的是，我们提出的"DUIN"方法在所有三个数据集上均实现了最佳性能。结果显示，我们的方法"DUIN"相对于当前最强的竞争者DEI2N的AUC增益分别为阿里巴巴数据集的1.45%，阿里妈妈数据集的0.22%，以及内容智慧数据集的0.41%。这充分证明了在TIR场景中考虑明确和潜在意图以及相应的不确定性对于提高性能的必要性。

Online A/B Testing Results

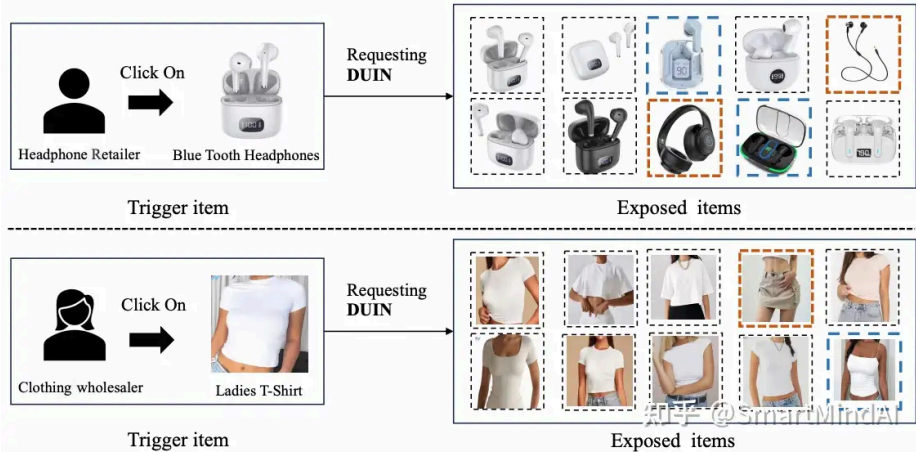
我们在线上A/B测试中，在两个14天周期的在线TIR场景中进行了实验，如表所示。我们选择DEI2N 作为基准，它在离线测试中的表现仅次于第一。结果表明，（CTR）和转化率（CVR）方面始终优于DEI2N。此外，由于暴露的商品可以分为两类-----与触发商品属于同一类的商品和与触发商品属于不同类的商品，CTR也可以分为同类别CTR（S-CTR）和跨类别CTR（C-CTR）。- CTR和C-CTR指标下都显示出显著的改进，这也表明。从跨类别曝光比率（CCER）的角度来看，为用户提供更多样化和新颖的推荐。因此，我们已经在阿里巴巴.com平台的主要TIR场景中部署了它。

Table 4: Online A/B testing results.

Scenario	Model	CTR	CVR	S-CTR	C-CTR	CCER
TIR 1	DEI2N	4.61%	3.96%	5.01%	2.70%	41.09%
	DUIN	4.75%	4.08%	5.25%	2.81%	46.81%
TIR 2	DEI2N	2.73%	1.59%	2.81%	2.92%	27.75%
	DUIN	2.82%	1.64%	2.88%	3.01%	28.50%

Case study

我们通过阿里巴巴平台上两个真实的案例展示了。如图所示，上部分说明了当用户，例如耳机零售商，参与我们的TIR方案时的情况。当他们点击"蓝色蓝牙耳机"时，"蓝色蓝牙耳机"相似的商品。同时，，比如"有线耳塞"和"耳机"。这说明"蓝色蓝牙耳机"相似的商品，还能准确预测用户的目的范围，提供更为精准和全面的采购建议。图的下部分展示了另一个真实世界的场景。在服装批发商点击"女士T恤"后，"女士T恤"相似的"女士T恤"，还推荐了与"女士T恤"搭配的"迷你裙"，以及平台上的热门商品"棉质背心"。



原文《Modeling User Intent Beyond Trigger: Incorporating Uncertainty for Trigger-Induced Recommendation》

编辑于 2024-09-04 11:10 · IP 属地北京

阿里巴巴集团 推荐系统 工业级推荐系统



理性发言，友善互动



发布