

CIKM 2024 - 华为 QueryRec: 利用搜索域query emb提升推荐域CTR预测



Recommender
会有成为大佬的那一天

关注

51 人赞同了该文章

收起

Enhancing CTR prediction in Recommendation Domain with Search Query Representation

Yuening Wang*
Huawei Noah's Ark Lab
Markham, Canada
yuening.wang@huawei.com

Man Chen*
Huawei Noah's Ark Lab
Montreal, Canada
man.chen1@huawei.com

Yaochen Hu
Huawei Noah's Ark Lab
Montreal, Canada
yaochen.hu@huawei.com

Wei Guo
Huawei Noah's Ark Lab
Singapore, Singapore
guowei67@huawei.com

Yingxue Zhang
Huawei Noah's Ark Lab
Markham, Canada
yingxue.zhang@huawei.com

Huifeng Guo
Huawei Noah's Ark Lab
Shenzhen, China
huifeng.guo@huawei.com

Yong Liu
Huawei Noah's Ark Lab
Singapore, Singapore
liu.yong6@huawei.com

Mark Coates
McGill University
Montreal, Canada
mark.coates@mcgill.ca

@Recommender

出处: Enhancing CTR prediction in Recommendation Domain with Search Query Representation CIKM 2024

关键词: 搜索与推荐, Diffusion, query表征

论文: arxiv.org/pdf/2410.2148...

1 问题背景

许多在线平台同时提供推荐和搜索服务, 推荐系统在用户浏览时主动推荐内容, 搜索服务则允许用户主动查找特定商品。因此, 如何有效利用推荐和搜索领域的数据成为一个重要问题。该paper聚焦于搜索和推荐领域联动下的CTR预测问题。

现有研究已针对推荐和搜索领域的联合学习进行了多种探索, 一些研究通过引入搜索数据来更好地理解推荐领域中的用户行为, 另一些则将推荐行为视为具有零查询的搜索行为特例, 从而实现两者的整合。

然而, 现有方法仍存在一些问

- 跨域兴趣转移**: 现有方法通常隐含假设用户在各个领域中的兴趣具有统一性, 而忽略了用户在浏览和搜索之间的意图变化, 未考虑到这两个领域之间的兴趣转移。因此, 有效捕捉跨域之间用户偏好的潜在变化对于准确预测用户兴趣至关重要。
- query与item对齐问题**: 搜索领域中的query通常与对应item紧密相关, 或可以提供item描述信息。深入挖掘query与item关系可增强query表征, 从而提升推荐系统效果。然而, 许多query没有点击记录, 现有方法简单地忽略了这些无点击query, 无法充分捕捉查询与物品之间的关系, 如何处理这些无点击查询仍是一个挑战。

2 方法

▲ 赞同 51 ▼

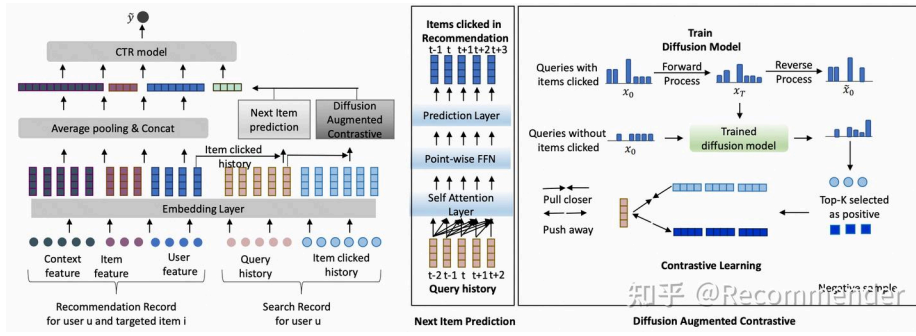
● 5 条评论

🔗 分享

♥ 喜欢

★ 收藏

📄



2.1 Next Item Prediction

针对跨域兴趣转移问题，作者设计了下一项预测模块。将 t 时间以前的query序列通过自注意力层和FFN⁺进行编码，然后预测用户在推荐领域 $t + 1$ 时会点击的项目。我们期望引入的query表征可以更好地匹配用户在推荐领域中的兴趣，理解用户在推荐领域中的偏好，以提供更多信息。为此，按照时间顺序映射推荐和搜索领域的交互记录，并通过直接预测用户在推荐领域下一个可能点击的item，来学习query的表征。

具体做法：在推荐领域中，从所有交互历史记录 (u, i, y, t) 中，构建两个item列表 $r_{u,t}^+$ 和 $r_{u,t}^-$ ，分别为用户 u 在时间 t 之后 $y = 1$ （有点击）和 $y = 0$ （无点击）的item列表。query序列 q_u^t 由自注意力序列(SAS)编码成emb $e_{q_u^t}$ 。使query emb与用户在推荐领域的点击物品（正样本）更接近，而与未点击物品（负样本）更远离：

$$\mathcal{L}'_2 = -\frac{1}{|r_{u,t}^+|} \sum_{j \in r_{u,t}^+} \log \langle e_{q_u^t}, e_j \rangle - \frac{1}{|r_{u,t}^-|} \sum_{k \in r_{u,t}^-} \log(1 - \langle e_{q_u^t}, e_k \rangle)$$

其中 $\langle \cdot, \cdot \rangle$ 表示点积⁺。为了更加高效，进行采样：

$$\mathcal{L}_2 = -\frac{1}{n_{\text{pos}}} \sum_{j \in r_{u,t}^+} \log \langle e_{q_u^t}, e_j \rangle - \frac{1}{n_{\text{neg}}} \sum_{k \in r_{u,t}^-} \log(1 - \langle e_{q_u^t}, e_k \rangle)$$

2.2 Diffusion-augmented Contrastive Learning

针对query与item表征对齐的问题，作者利用了对比loss，来学习query与item之间的相似性和差异性。将query和其相关的item对齐

1. query-item对比学习（Contrastive Learning）

从搜索域的交互历史中，收集query q 对应的正样本集，将其表示为 \mathcal{I}_q^+ 。其中，对于每个item $i \in \mathcal{I}_q^+$ ，至少有一位用户在通过 q 搜索后， i 曝光并且被点击。 q^t 表示用户在 t 时间之前的最后一次query。对比loss：

$$\mathcal{L}_3 = \frac{1}{|\mathcal{I}_q^+|} \sum_{i \in \mathcal{I}_q^+} \log \frac{\exp(s(e_{q^t}, e_i)/\beta)}{\sum_{j \in \mathcal{I} \setminus \mathcal{I}_q^+} \exp(s(e_{q^t}, e_j)/\beta)}$$

2. 基于Diffusion的数据增强（Diffusion Data Augmentation）

搜索数据集的搜索点击记录稀疏，同时大部分的query没有正向互动的item，即对于大量的查询 q ， \mathcal{I}_q^+ 为空，导致这些query在对比loss中没有标签。尽管如此，每个query都有一组未点击的item \mathcal{I}_q^- ，对于item $i \in \mathcal{I}_q^-$ ，用户在通过 q 搜索后， i 曝光但是未被点击。可以将 \mathcal{I}_q^- 用作对比信号，并使用Diffusion Model⁺来增强被搜索点击的item \mathcal{I}_q^+

3 实验结果

Dataset	KuaiSAR-small		KuaiSAR-large		Industrial	
Model	AUC	Imp %	AUC	Imp %	AUC	Imp %
DIN	0.6793	0	0.6892	0	0.8943	0
DCN	0.6729	-0.94	0.6749	-2.1	0.8948	0.02
PLE	0.6825	0.47	0.6879	-0.19	-	-
SESRec	0.6843	0.74	0.6859	-0.48	-	-
IV4REC+	0.6640	-2.25	0.6580	-4.53	-	-
QueryRec	0.6972	2.64	0.6996	1.5	0.8999	0.63

4 结论&想法

该paper的核心在于query表征的构建，提取推荐和搜索双域的信息，并将其融入到CTR模型*中，实现知识从搜索领域到推荐领域的转移。这篇文章提出了一个关键问题：推荐域和搜索域*之间由于用户消费意图和推荐曝光机制的差异，存在跨域兴趣转移。一方面，通过Next Item Prediction提取query特征，引入搜索后消费的正、负样本视频信息。在推荐域通过预测下一个，实现query表征与用户兴趣的对齐，使query表征具有推荐域的语义；另一方面，在搜索域通过对搜索行为历史emb结合对比学习、Diffusion进行优化，使其在搜索域的表达更加充分。基于此将推荐与搜索双域的信息都表征到query这一粒度。

跨域推荐系统往期文章

Recommender: WWW 2024 - 淘宝 CDAnet: 基于显式特征增强提升跨域CTR预测

Recommender: WWW 2024 - UDMCF: 跨域推荐中基于协同过滤的用户分布映射建模

Recommender: KDD 2024 - LCN: 用于CTR预测的跨域长序列建模

Recommender: RecSys2024 - MLoRA 用于CTR预估的多域低秩自适应网络

编辑于 2024-12-10 10:45 · IP 属地北京

内容所属专栏



跨域推荐系统

订阅专栏

ctr 搜索推荐系统 跨域



理性发言，友善互动

5 条评论

默认 最新



命运粘合剂

感觉就是搜索行为序列的事情。。。又水一篇

2024-12-11 · 浙江

回复

3



绒默

这领域学术文章不一直这样吗

01-13 · 内蒙古

回复

1



绒默

命运粘合剂

有些组这个就是KPI

01-13 · 北京