

2024华为实践：利用召回数据提升精排CTR预估AUC



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

已关注

31 人赞同了该文章

Introduction

推荐系统⁺已普遍应用于解决用户信息过载问题。点击率⁺（CTR）预测作为其核心功能，旨在预测用户在特定环境下对某一内容点击的可能性，以优化用户体验和增加平台收益。近年，多种模型依据用户历史行为序列建模用户偏好，但面对海量内容时，这些序列表现出明显的同质性和稀缺性。

目前，预训练的模型倾向于使用目标注意力机制⁺，过度重视重复或类似内容，从而形成强化的同质循环，如用户购买口红后，系统推荐与其品牌相似的彩妆产品。然而，用户可能更倾向于探索如香水或耳环等与初始购买相关的但未被互动过的商品，以寻求多样化的需求满足。

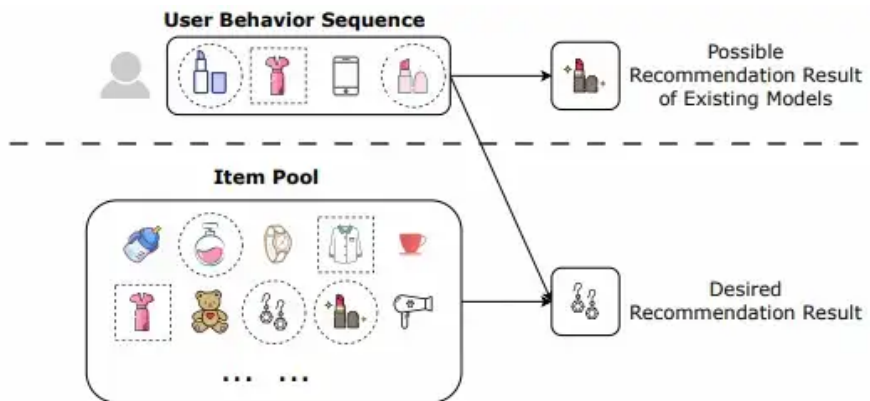


Figure 1: An illustrated example for motivations of RAR.

本研究将从数据驱动的方法出发，通过集成不同信息源来丰富用户表示，以提高CTR预测的准确性，从而促进用户探索与信号发现之间的平衡，推动推荐系统向更个性化、多样化的方向发展。

此外，传统的CTR预测主要关注单一的用户-内容交互，往往忽视了用户和内容间各种相互关系的存在，从而导致长尾内容建模不足。与此形成对比，召回阶段自然生成了相似的用户-内容列表，提供了一种跨实例的建模能力。

我们从交叉阶段框架的角度来看，识别用户画像⁺和召回内容作为两大理想数据源，分别涵盖了用户到用户（u2u）和内容到内容（i2i）的方面。在本文中，我们旨在利用这些交叉阶段数据来增强CTR预测的准确性，而非如何构造这两个集合。为此，我们提出了一种名为召回增强排序（RAR）的新型架构，旨在基于交叉阶段数据提升模型准确性。

RAR架构包含两个关键组成部分：交叉阶段用户与内容选择模块和协同交互模块。这两个部分通过整合广泛的类似用户和召回内容集合的信息，构建了更丰富且多元化的用户表示。



本论文的主要贡献如下：

- 针对只依据用户行为序列建模用户偏好的局限，我们提出了一种革新方法-----召回增强排序（RAR）架构。RAR架构巧妙地利用跨阶段数据，旨在深入挖掘并丰富用户表示，优化预测模型⁺的性能。
- RAR架构特别引入了两个数据补充来源:相似用户集合与召回物品集合。这一整合创新，在点击率预测领域的集对集建模方面尚属首次尝试，旨在全面提升模型性能。

Approach

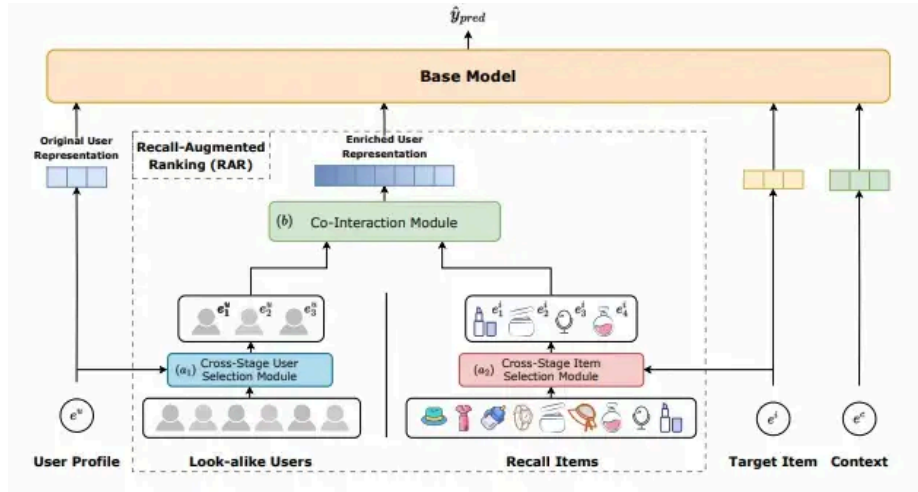


Figure 3: RAR applied in existing CTR prediction models.

Cross-Stage User/Item Selection Module

跨阶段用户/项目选择模块选择最相似的用户和相关的项目。选择过程可以抽象为两步，以召回项目的选取为例：首先，通过相似性函数 $f(\cdot)$ 测量目标项目与每个召回项目的相似性。接着，根据相似性评分筛选出最相关的前 k 个召回项目，公式如下：

$$S_{\mathcal{L}} = f(E_{\mathcal{L}}, e^u), S_{\mathcal{R}} = f(E_{\mathcal{R}}, e^i)$$

$$E_{\mathcal{L}'} = [e_{k_1}^u, e_{k_2}^u, \dots, e_{k_l}^u]^T, E_{\mathcal{R}'} = [e_{k_1}^i, e_{k_2}^i, \dots, e_{k_r}^i]^T$$

直观的想法是使用嵌入并基于点积⁺搜索 k 个最近邻。然而，大规模的乘法会使现实世界的应用变得不切实际。考虑到选择的复杂性，我们在实验中采用了SimHash函数。SimHash利用局部敏感性特性，通过随机投影和带符号轴，使相似输入得到相似输出，从而简化嵌入为二进制指纹。这一过程详细描述于公式⁺中，其中 e_k^i 表示第 k 个召回项目的嵌入，而 m 则是哈希函数集合中的第 m 个哈希函数⁺。通过使用汉明距离⁺进行高效的比较，这一过程实现了存储效率和选择速度快的目的。

$$\text{sig}_k^i[m] = \sum_{n=1}^{d_2} \text{sgn}(e_k^i[n] \cdot P[n][m])$$

$$\text{sig}_k^i[m] \leftarrow \mathbf{1}_{\text{sig}_k^i[m] > 0} (\text{sig}_k^i[m])$$

Co-Interaction Module

协同交互模块提供了一组精细的对对之间的建模。它改进了单纯将所有选择的召回项等权重对待的做法，这种方法忽略了层级信息。我们引入了一个匹配矩阵来评估用户与项目之间兴趣相容性的程度。匹配分数被表示为高层次潜在向量的点积，如示于公式⁺中所示。然后，我们在公式⁺中计算匹配矩阵，其中使用Sigmoid函数将匹配分数映射到 $(0, 1)$ 区间内。

$$E'_{\mathcal{LH}} = l$$

$$\mathcal{M}_M = \text{Sigmoid}(E'_{\mathcal{L}_H} \cdot E'^T_{\mathcal{R}_H})$$

为了使模型能够更清楚地了解哪些召回项目更重要，利用信号 y_{ui}^{ep} 监督匹配矩阵的训练。由于曝光信号非常稀疏，我们定义了 y_{ui}^{ep} 如公式所示。

$$\hat{y}_{ui}^{ep} = \mathcal{M}_M$$

$$y_{ui}^{ep} = \begin{cases} 0, & i \text{ has never been exposed to } u' \text{ for } \forall u' \text{ in } \mathcal{L}'_u \\ 1, & \text{otherwise} \end{cases}$$

最后，通过对匹配矩阵进行行平均和列平均，得到项目和用户加权向量。通过将加权向量与相应的嵌入相乘，我们得到用户共同兴趣 v_{ui}^c 和用户多样性兴趣 v_{ui}^d 。然后，通过连接 v_{ui}^c 和 v_{ui}^d ，得到用户增强表示 v_u^{enr} 。

$$w_i = \text{Mean}(\mathcal{M}_M, axis = 0), w_u = \text{Mean}(\mathcal{M}_M, axis = 1)$$

$$v_{ui}^c = w_u \cdot E_{\mathcal{L}'}^c, v_{ui}^d = w_i \cdot E_{\mathcal{R}'}^d$$

$$v_u^{enr} = \text{Concat}(v_{ui}^c, v_{ui}^d)$$

Objective function

RAR的**损失函数**⁺可以用公式表示，其中 \mathcal{L}_{clk} 旨在准确预测点击率 \mathcal{L}_{ep} 旨在为模型提供有关哪些召回项目最重要的更明确的指示。 $\alpha \in [0, 1]$ 是一个可以调整的参数，用于平衡两个损失。两个损失均采用**交叉熵损失**⁺，以点对点的方式监督训练过程。RAR的所有模块通过最小化训练数据集上的联合损失函数联合训练。

$$\mathcal{L} = \alpha \cdot \mathcal{L}_{clk} + (1 - \alpha) \cdot \mathcal{L}_{ep}$$

Experiments

Experimental Setup

数据集。实验在三个公开数据集上进行:KKBox、Movielens和工业数据集CandiCTR-Pub。KKBox用于音乐推荐，而Movielens包含用户对电影的标记数据。CandiCTR-Pub是一个包括召回项集的实用、大规模工业数据集。**基本模型。**探索了六种基本模型，包括IPNN、WDL、DeepFM、DCN、xDeepFM和AutoInt⁺，以及新的模型DeepIM和DCN-V2。这些模型在BARS**基准测试**⁺中已有评估记录。**基础模型比较。**FRNet是通过学习上下文感知的特征表示以捕获特征间的交叉关系而成为最新主导（SOTA）模型的推荐系统。CIM则通过变换器将所有潜在内容编码为上下文向量，以建模用户的隐式偏好。**评估指标。**使用了AUC（Area Under Curve）和gAUC（按用户分组的加权平均AUC）作为性能评估的指标。AUC提供了一个全面的评价指标，而gAUC更进一步考虑了用户之间的推荐差异性。

Performance Evaluation with SOTA Models

对现有最先进的模型（SOTA方法）进行评估时，引入RAR（Sequential Attention and Reward）后，观察到显著的性能提升。

Datasets		KKBox							
Modules	Raw		+FRNet		+CIM		+RAR		
Models	gAUC(%)	AUC(%)	gAUC(%)	AUC(%)	gAUC(%)	AUC(%)	gAUC(%)	AUC(%)	
IPNN	78.75	85.25	78.27	84.94	78.31	85.25	80.15	86.45	
WDL	78.44	85.02	78.26	84.85	78.67	85.36	79.74	86.23	
DeepFM	78.76	<u>85.32</u>	<u>78.70</u>	<u>85.26</u>	78.90	85.68	80.14	86.51	
DCN	78.66	85.25	78.69	<u>85.26</u>	<u>79.16</u>	<u>85.74</u>	<u>80.22</u>	<u>86.58</u>	
xDeepFM	78.60	85.25	78.65	85.22	78.72	85.56	80.12	86.50	
AutoInt+	<u>78.78</u>	85.34	78.71	85.28	78.94	85.64	80.20	86.55	
DeepIM	78.79	85.29	78.56	85.16	78.92	85.63	80.26	86.59	
DCN-V2	78.64	85.17	78.62	85.22	79.45	85.77	80.12	86.49	
Best RelImp	0.0%	0.0%	0.1%	0.1%	1.0%	0.7%	2.0%	1.6%	

Datasets		Movielens							
Modules	Raw		+FRNet		+CIM		+RAR		
Models	gAUC(%)	AUC(%)	gAUC(%)	AUC(%)	gAUC(%)	AUC(%)	gAUC(%)	AUC(%)	
IPNN	95.53	96.53	95.14	96.15	95.28	96.38	95.92	97.02	
WDL	95.29	96.23	95.17	96.19	95.36	96.44	95.73	96.67	
DeepFM	94.84	95.90	94.65	96.11	95.06	96.23	95.40	96.40	
DCN	<u>95.32</u>	<u>96.35</u>	95.21	96.33	95.30	96.37	95.51	96.54	
xDeepFM	95.27	96.20	95.21	96.26	95.16	96.29	<u>95.82</u>	<u>96.81</u>	
AutoInt+	95.22	96.24	95.26	<u>96.28</u>	95.31	96.38	95.61	96.58	
DeepIM	95.29	96.29	95.21	<u>96.28</u>	95.30	<u>96.39</u>	95.51	96.61	
DCN-V2	94.95	96.00	<u>95.23</u>	96.25	<u>95.35</u>	<u>96.39</u>	95.66	96.63	
Best RelImp	0.0%	0.0%	0.3%	0.3%	0.4%	0.4%	0.7%	0.7%	

Datasets		CandiCTR-Pub							
Modules	Raw		+FRNet		+CIM		+RAR		
Models	gAUC(%)	AUC(%)	gAUC(%)	AUC(%)	gAUC(%)	AUC(%)	gAUC(%)	AUC(%)	
IPNN	52.87	60.92	52.35	61.08	53.76	61.86	54.35	62.53	
WDL	<u>52.82</u>	60.90	52.48	60.82	53.92	62.73	54.43	<u>63.92</u>	
DeepFM	<u>52.82</u>	60.99	52.48	60.89	53.94	<u>62.80</u>	54.50	<u>63.92</u>	
DCN	52.78	60.95	52.55	60.61	53.75	61.75	54.50	62.53	
xDeepFM	52.87	<u>61.19</u>	52.48	60.78	<u>53.93</u>	62.79	54.40	64.06	
AutoInt+	52.61	61.10	52.73	<u>60.95</u>	53.82	62.85	<u>54.47</u>	63.86	
DeepIM	52.72	61.23	52.42	60.94	53.48	61.65	54.44	62.72	
DCN-V2	52.65	61.06	<u>52.64</u>	60.79	53.32	61.69	54.30	62.70	
Best RelImp	0.0%	0.0%	0.2%	0.3%	2.6%	<u>3.9%</u>	<u>3.5%</u>	<u>5.0%</u>	

以xDeepFM+RAR为例，在不同数据集上的AUC值⁺最高提升达4.7%。这一结果有力地表明，通过在不同阶段整合数据，可以有效地丰富用户表示，进而提升推荐系统的性能。

原文《Recall-Augmented Ranking: Enhancing Click-Through Rate Prediction Accuracy with Cross-Stage Data》

编辑于 2024-06-25 11:38 · IP 属地北京

ctr 推荐系统 协同过滤



理性发言，友善互动

2 条评论

默认 最新



落魄的阿黄

有线上的实验效果吗

06-25 · 北京

● 回复 ● 喜欢

