

赞同 22

分享

Facebook 2023：揭秘Que2Engage，创新向量召回技术，兼顾相关性与吸引力，引领新趋势



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

22 人赞同了该文章

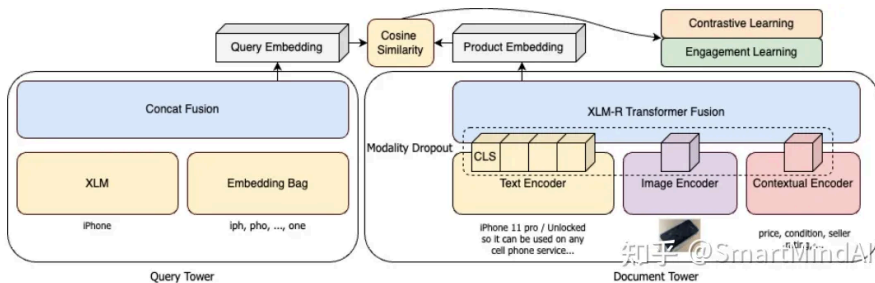
Introduction

向量检索⁺（EBR）在电商搜索引擎中发挥着关键作用，例如在Facebook Marketplace、Walmart和Instacart等平台上。EBR模型主要学习query和item的向量表示，通过近似最近邻搜索（ANN search）来检索语义上接近查询的文档。然而，搜索引擎作为复杂的多阶段系统，需要优化多个业务目标，因此仅优化语义相关性可能并不总能得到最佳结果。例如下游重新排序系统可能无法对EBR检索到的结果进行恰当排序。在**电子商务⁺**平台如Facebook Marketplace上，产品价格、生产条件、卖家评分等上下文信息也是确保检索产品吸引力的关键信号。然而，在EBR环境中利用这些上下文信息提高**搜索者⁺**参与度并非易事，因为

- （1）传统的基于对比学习的EBR技术主要强调语义相关性，简单应用上下文信息可能效果不佳；
- （2）产品语义相关并不保证其对搜索者的吸引力，因此同时保持相关性和吸引力具有挑战性。

本文提出Que2Engage，对Que2Search进行扩展以应对上述挑战。它采用**多模态⁺**方法，将上下文信号作为独立模态融入transformer融合主干网络。该模型通过**多任务⁺**学习进行训练，结合对比学习和排序训练，不仅检索语义相关产品，还提升有吸引力产品的排名。我们为多任务评估提出了框架，以了解模型在不同领域的性能。我们使用此框架进行详细基线比较和消融研究，验证了我们的多阶段一致性论点和方法在利用上下文信息上的有效性。Que2Engage 已集成在Facebook Marketplace**搜索引擎⁺**中，每天为数百万的搜索查询提供服务。通过为期两周的在线A/B测试，证明其在搜索者参与度上有显著改善。

Model Architecture



Query tower

我们采用了多粒度搜索查询表示，包含原始查询文本和字符**三元组⁺**。原始查询文本通过2层XLM编码器进行编码，字符三元组则使用EmbeddingBag编码器进行编码。在将这两种表示传入MLP层前，我们选择串联而非注意力融合来组合它们，实验发现前者的性能更佳。

Contextual information as a modality

BatchNorm层处理，再输入至MLP，以确保数值范围固定且输出长度一致。该编码器的输出称为“上下文令牌”，因为在多模态融合步骤中，其处理方式与文本和图像令牌类似，即上下文信息被视为一种独特的模态。

Multimodal fusion⁺

我们采用文本编码器将产品标题和描述转为单词标记，再输入到transformer获取文本嵌入。

CLS标记用于整句编码。对于图像，我们用预训练图像表示处理，并用共享MLP层和深度集合融合获得图像密集表示。参照中的transformer-fusion架构，我们将文本、图像和上下文标记连接后输入到**多模态融合⁺**编码器中。文本编码器和多模态融合编码器从6层的XLM-R初始化。所述，文本编码器继承前K层，多模态融合模型继承其余M层。最后，我们从多模态融合编码器最后一层提取**CLS**标记的隐藏输出，投影到所需维度作为文档嵌入。

Modality dropout

为了确保模型不过度依赖单一模态，并在推理过程中对缺失信息具有**鲁棒性⁺**，我们采取了模态dropout机制。具体来说，我们以 δ_c 、 δ_i 和 δ_t 的概率随机屏蔽上下文编码器、图像编码器和文本编码器的输出，被屏蔽部分用零替换。

Multitask Training

Contrastive learning

我们采用了基于batch负采样的对比学习作为训练目标的一部分。正样本是从搜索日志中抽取的用户参与的\<查询，产品>对，而负样本是通过**随机组合⁺**查询和产品生成的。形式上，我们定义了相关性损失 $L_{relevance}$ 如下：

$$L_{relevance} = \frac{1}{B} \sum_{i=1}^B -\log \left\{ \frac{\exp\{s \cdot \kappa(q_i, d_i)\}}{\sum_{j=1}^B \exp\{s \cdot \kappa(q_i, d_j)\}} \right\}$$

其中，B为batch大小， κ 是一个采用**余弦相似性⁺**的相似性核，s为固定缩放因子，值为20，与Que2Search模型中的设定相同。在这个版本中，去掉了原文中的一些不必要的词汇和表述，简化了语句结构，同时保留了所有的数学公式和符号，以保证语法清晰和整体可读性。

Learning contextual information

我们定义**损失函数⁺**L如下：

$$L_{engagement} = -(y_i \log(c_i) + (1 - y_i) \log(1 - c_i))$$

其中 $c_i = s \cdot \kappa(q_i, d_i)$ 。为了结合相关性和参与度损失，我们定义最终的多任务损失为：

$$L_{total} = \alpha L_{relevance} + (1 - \alpha) L_{engagement}$$

其中， α 是权衡两者重要性的**超参数⁺**。

$$L(\theta) = \lambda_1 \cdot L_{relevance} + \lambda_2 \cdot L_{engagement}$$

在论文中，我们已设定 θ 作为模型参数，而 λ_1 和 λ_2 则是基于经验的权重参数。

Offline Experiments

Dataset

我们从Facebook Marketplace的搜索日志中收集了1.5亿对\<查询，产品>展示，其中7500万作为正面样本，7500万作为负面样本。数据经过去标识和聚合处理，以保障隐私。评估使用了2.6万

Baselines and Ablation Studies

我们选取Que2Search作为基准模型，这是一个双塔模型⁺，基于注意力融合预训练的XLM编码器和图像表示。我们增强了模型，用基于上下文信息的编码器，并采用混合批次法纳入了展示给搜索者的产品的硬负样本。处理组使用具有预计算图像嵌入的Que2Engage，因其简单性更受青睐。我们另外分享了基于CommerceMM编码器微调的替代图像编码器的比较。

Experimental Setup

Parameters

我们在Nvidia A100 GPU上使用PyTorch Multimodal框架开发了所有模型。模型使用批量大小为512进行训练，并使用学习率为 $4e-4$ 的Adam优化器。权重参数 λ_1 和 λ_2 分别设为0.8和0.2。模态dropout参数为： $\delta_c=0.5$ ， $\delta_i=0$ 和 $\delta_t=0.5$ 。文本令牌和其他模态令牌直接输入到多模态变压器中，这是一个具有0层文本编码器和6层多模态融合编码器的早期融合模型。在比较预计算图像嵌入与微调CommerceMM编码器时，因GPU内存消耗增加，batch大小调整为64。为避免损失函数计算中出现NaN，CommerceMM的训练学习率调整为 $5e-5$ 。

Results

Analysis of baseline results

表上半部分展示了基准模型的性能。Que2Search在相关性评估上表现良好，但在参与度数据集上表现欠佳，这可能是因为它主要针对语义相关性进行了优化。添加上下文信息到Que2Search并不能直接提高参与度评估的性能。但当我们更改训练目标以引入混合负面因素后，预测性能得到显著提升，如表第4行所示，参与度评估的ROC_AUC达到最高。这与假设相符，普通的batch负采样可能因采样偏差不足以学习参与度预测的细节，而混合负采样可以有效地利用上下文特征，缓解这一问题。另外，请注意表的第4行和第5行显示的相关性评估回归以及参与度评估的改进，这是预期的，因为上下文信息（如产品价格和状况）与我们为评估者提供的查询-产品相关性指南无关。因此，我们不希望这些上下文信号对相关性评估有所改善。

Model	Engagement	Relevance
Que2Search[7]	55.88	67.14
Que2Search w/ contextual encoder	55.85	66.74
Que2Search w/ mixed batch loss	63.63	63.79
Que2Search w/ contextual encoder + mixed batch loss	64.45	61.17
Que2Engage w/ mixed batch loss	64.70	60.36
Que2Engage w/ multitask training	76.13	65.63
Que2Engage w/ multitask training + modality dropout	76.90	67.21

Table 1: Results for baseline comparison and ablation studies

Method	Engagement	Relevance
pre-computed image embedding	74.35	60.19
fine-tuned CommerceMM encoder	74.67	60.55

Table 2: Results for image encoder comparison

Que2Engage and ablation studies

表第二部分揭示，通过运用Que2Engage方法，参与度的评估得到了显著提升。在两种评估手段中，结合了多任务学习和模态丢弃的完整Que2Engage效果最佳。对损失函数的消融研究进一步证明，多任务训练对参与度评估的提升作用最为显著，这证实了对硬负样本的集中损失函数效果优于简单混合负样本。最后，模态丢弃对这两个指标也有进一步的优化，尤其在相关性评估上，这表明强制缺失模态能防止模型对单一任务过度拟合，有助于其对其他任务的回归。

Image encoders

端微调需更多GPU内存和训练时间，我们未在生产模型中使用此技术。消融研究在批次大小为64的情况下完成，这对绝对相关性评估有影响，因为更大的批次大小在对比学习中被证明有帮助。我们希望通过内存效率优化和更强大的硬件来实现该技术产品化。

Online Experiments

我们已将Que2Engage 部署在Facebook Marketplace搜索中，作为传统的基于词汇的搜索检索的并行检索源。在在线A/B测试中，对比了Que2Engage 与之前的生产模型Que2Search，后者仅针对语义相关性进行了优化。测试主要测量了NDCG和搜索者参与度。两周的测试结果显示，Que2Engage 提高了搜索者参与度4.5%，同时保持了NDCG的中立性，这与我们之前的离线多任务评估结果相符。

Related Work

近年来，向量检索（EBR）在电子商务搜索中得到了应用，作为词汇检索的补充手段，用于检索语义相关的产品。许多搜索和推荐系统中的EBR模型选择了使用对比学习损失训练的Siamese中性网络。然而，这种方法可能会错过有趣的负样本，且内存需求较大。为解决这个问题，一些对比学习的变体采用了更智能的负采样方法，并优化了内存使用。另外，教师-学生蒸馏模型也被用作辅助任务来改进相关性。预训练语言模型（PLM）在搜索检索中得到了广泛应用，因为检索的主要关注点是文本相关性。不过，上下文信息对于EBR系统端到端性能的影响也得到了重视。最近，有研究提出了一种统一的训练方案来平衡多个优化目标，但在多目标设置中上下文信息的作用尚未得到充分讨论。

Conclusion

我们提出了在Facebook Marketplace等应用中平衡相关性和参与度的EBR建模需求，并介绍了我们的最新搜索EBR系统。通过基准比较和消融研究，展示了结合上下文信号、多模态技术、表示学习和多任务学习的创新有效性。已在Facebook Marketplace搜索中部署了Que2Engage。两周的A/B测试证明，其性能优于现有最先进的搜索EBR系统，并显著提高了搜索者在Facebook Marketplace上列出产品的参与度。

论文原文《Que2Engage: Embedding-based Retrieval for Relevant and Engaging Products at Facebook Marketplace》

跟踪最新技术，关注我不迷路
www.zhihu.com/people/smartmindai



编辑于 2023-11-25 12:08 · IP 属地北京

相关性分析 Facebook 搜索

理性发言，友善互动

2 条评论

默认 最新

Photon

吸引力标签怎么获取的呢

2023-11-01 · 上海

回复

喜欢

马萨卡

成交标签

2023-11-09 · 上海

回复

喜欢