

## 机器之心报道

编辑：思

腾讯广告通过模型能力的提升助力广告主达成生意目标，同时兼顾用户体验，实现多赢局面。

模型是广告系统中的一个复杂且重要的组成部分。之所以说它很复杂，不仅在于庞大的用户量及广告场景，也在于广告这种天然的多模态数据，要求模型具有强大的拟合能力。

思考一下，如果你是一名工程师，怎样才能让你的模型理解广告的内容、主题、目标受众？对于手机、电脑等不同设备，食品、汽车等不同行业，销量、名声等不同目标，模型怎样才能理解不同的广告场景？又怎样才能从百万条广告中为受众选择合适的推送？面对流量竞争加剧的大环境，广告系统如何用技术破局，通过模型能力的提升助力广告主达成生意目标，同时兼顾用户体验，实现多赢局面？

不得不承认，抛开工程难度不说，想要拍脑袋得出一套好的解决方案，也基本是不可能的。

那么业界顶尖的广告系统，它到底是怎么做的呢？本文将

从腾讯广告混元AI大模型及广告大模型出发，介绍腾讯广告系统背后的技术，以及如何带来广告系统的硬指标——商品交易总额（GMV）的提升。

## 大模型：广告系统的灵魂

在了解腾讯广告系统具体细节之前，我们先要有一个概念：大规模预训练模型，或者说大模型，不仅是 NLP、CV 的未来，同时也是广告系统的灵魂。

归根究底，广告系统需要解决以下两个关键问题：

- 如何理解各种广告文案、图片与视频？
- 如何将合适的广告，在合适的时间推送给合适的人？

而这两点，恰恰是大模型能够处理的。一方面，多模态大模型能充分理解广告内容；另一方面，广告推荐大模型，能充分考虑庞大的用户与广告库，找到最合适的推荐。

拥有了大模型，手上也就有了武器。

混元AI大模型充分编码文本、图片、视频等各种类型的广告数据，并理解它们都是什么，以及想要表达什么；腾讯广告大模型则基于之前理解的广告内容，思考到底要推荐给哪些人群，才能既帮广告主达成营销目标，又能让信息满足受众的需要。

## 混元AI大模型：内容理解是一切的基础

在广告系统中，广告的视频、图片、文案是天然的多模态数据，能充分理解它们的也必然是最新的大规模多模态训练。BERT证明了Transformer无监督预训练能理解自然语言，MAE证明了Transformer无监督预训练能理解图像数据。这种情况下，统一用Transformer建模图像与文字等数据，构建一种统一的大规模多模态预训练模型就成为了炙手可热的领域。

为此，针对内容理解，混元AI大模型使用大量无监督广告内容，构建一个通用的多模态内容理解预训练模型，并预期它能为各种下游理解任务提供更强的基础能力。

该内容理解模型主要基于Transformer结构，并根据真实业务场景与广告数据特性，提出多尺度视觉特征融合、层次化注意力交互、大规模对抗训练等诸多改进和设计。

混元AI大模型的Transformer，通过预训练实现多模态内容理解。

上图所示为混元AI大模型针对多模态理解构建的Transformer架构。该模型会通过多种任务，例如预测被遮盖的语言（MLM）、图像文本是否匹配（ITM）和视频文本是否匹配（VTM）等等，让模型同时学会理解多种形式的广告。腾讯广告的混元AI大模型，已经霸榜了多个多模态内容理解榜单，包括VCR、CLUE、MSR-VTT等11个公开测评集，这也充分体现了其训练方式的有效性。

此外，腾讯广告还通过多种策略以加强模型的理解能力。例如在上图「Image」部分采用了多尺度视觉特征融合，即特征同时采用主体Region、Patch特征，以便获取更丰富的视觉语义信息。

其次在Transformer层级中增加了层次化注意力交互，通过采用全局+局部注意力的方式，在不损失重要信息交互学习下，节省计算开销，提升训练效率。最后就是每种数据都增加了对抗噪声（Adversarial Noise），进行大规模对抗训练，这主要针对大模型在下游任务容易出现过拟合现象，因此在预训练和微调两阶段都增加了对抗噪声，从而提升模型的稳健性。

当然，除了要理解多模态广告内容，混元AI大模型还需要实现生成文案、搜索 / 推荐相似广告等任务

，这些任务似乎并不能直接使用之前的多模态内容理解模型。腾讯的做法是采用更合适的架构，为它们打造独立的「内容理解」模型。例如跨模态检索，就更适合使用双塔Transformer结构，从而得到不同模态之间的对应关系。

混元AI大模型的双塔 Transformer，通过层级化对比学习，学习视频和文本两种模态的不同粒度的相似度。

## 广告大模型：合适的内容给合适的受众

对于广告系统来说，不仅需要准确理解广告所表达的信息，同时也要将下游任务做好

。在众多下游任务中，推荐系统又是最为突出的部分，毕竟它承担着将合适的内容推送给合适的受众，是广告投送方与广告接收方最能感知的部分。

将合适的内容推送给合适的受众，听上去很简单，但背后的广告大模型实际上涵盖了召回、粗排、精排等广告投放环节的一系列方法。简而言之，召回负责从百万量级广告中筛选出某类的广告；粗排在召回结果的基础上进一步筛选出数万量级的广告，选择某一典型的广告集合；最后，也是最重要的精排模型，则对粗排筛选的结果进一步排序，从而选择最终要向受众曝光的广告。因此，广告大模型需要在考虑哪个广告最能迎合用户诉求的同时，为广告主带来最大的效益。

总的来说，广告大模型可以实现千亿参数量的高效训练与推理，并提供可靠的广告匹配结果。得益于广告大模型的能力，腾讯广告整体的商品交易总额实现大幅提升。对于广告主来说，商品交易总额的提升，就是广告大模型实打实的能力。

## 打破天花板：广告大模型的创新解法

打破百亿参数量「小模型」的效果天花板，离不开高效的模型算法。

算法工程师

们需要考虑的是，千亿参数的广告大模型，怎样设计、优化才能解决广告业务中的个性化、场景化、行业化这三大难点。

1. 个性化说的是，广告场景需要更精准、更个性化的表征，从而突破小模型的表征能力天花板。通过改特征维度、Attention机制、让模型理解更长的时序跨度等方法，腾讯广告团队提升了模型对「个性」的建模能力。
2. 场景化说的是，对于腾讯上万的广告位，场景差异特别大。小模型只能引入场景特征，或者分场景建模，这样效果既达不到最好，维护起来也困难重重。而广告大模型，通过构建场景化模型，集成了多个算法特性组合，从而解决多场景联合建模问题。
3. 还有行业化，深入行业转化目标多达上百个且数据稀疏，大模型加入大量高维特征，造成样本相对特征参数空间不足。这里腾讯通过提出Multi-Embedding Net，采用多种不同尺寸的嵌入组合学习来减轻这个问题，同时也通过显示匹配 ( match ) 特征，降低特征稀疏度。

腾讯广告在召回、粗排

和精排算法上都有很多重要创新

，再借助千亿参数大模型的强大学习能力，它们共同为解决如上广告三大难点做出努力。

## 精排算法三大创新

第一点：从原始特征抽取信息。

精排模型是一种典型的稀疏神经网络架构，参数主要集中在Embedding层，或者说表示层。以前的小模型因为算力、表征能力等受限，原始行为特征会先通过降维算法，再输入到模型，这必然导致信息的损失。

来到大模型时代，腾讯广告团队直接把用户曝光、点击、转化、文章阅读等行为高维特征植入模型，减少了信息损失。但是研发团队也发现，简单增加原始行为序列特征并不能使模型完美学习，为此进一步提出了AutoAttention，即通过注意力机制，自动且高效地抽取用户行为序列中重要的部分，增加模型表

达能力。

研发团队也与阿里开源的DIN ( Deep Interest Network ) 点击率预估模型做了对比，AutoAttention有以下两项优势：

- DIN特征组合需要人工选择，AutoAttention能自动学习不同特征的权重；
- 相比DIN的外积+MLP方式，AutoAttention仅采用内积，更高效，且更节省计算量。

因为特征维度从亿级别提升到十亿级，线上效果也比较显著，其中朋友圈页面访问点击率模型 ( pCTR ) 商品交易总额也呈现提升。

第二点：让子网络解决广告场景多且差异大的问题。

腾讯的广告位多，且差异大，它们是一种高度不均衡的样本分布，甚至多样的场景在训练中会带偏少样本的场景。为此，腾讯广告团队叠加运用一系列子网络设计技巧解决该问题：

- 1) 表示层：学习共享Embedding在不同领域下的权重，并通过位置偏置网络和多维度场景交叉特征强化场景差异；
- 2) 隐藏层：采用Partitioned layer normalization，对不同领域的样本分别进行归一化，增强模型分领域的收敛速度；
- 3) 输出层：按照场景拆分公共塔和场景独立塔，强化场景个性表达。

通过叠加一系列算法优化，之前因为领域差异拆开的模型能重新合起来，这不仅会降低维护成本，如公众号页面点击预估现在只需要一个模型，同时也取得GMV的提升。

第三点：解决广告模型特有的高维与稀疏问题。

随着行业化深入，细粒度行业特征增多，与此同时，转化链路变长。虽然链路长的目标样本更稀疏，但其也更重要。腾讯广告团队提出的Multi-



## Embedding

Net构建了三种不同的Embedding子网络，并期待从Embedding Size大的子网络帮助浅层目标更准确，Embedding Size小的子网络帮助深层目标收敛更快。

腾讯广告团队表示，以朋友圈pCVR为例，采用Multi-Embedding模型后，不同目标AUC都有较显著的提升，线上的模型目前均已采用该结构。

## 粗排、召回算法也升级

为了配合精排模型的创新与提升，召回算法、粗排算法都选择了排序学习（Learn to Rank, LTR），它们以精排结果为学习目标，希望能更好地对接最终的精排模型。

召回候选广告队列达百万级，需在全库广告中择优，因此存在较大的选择性偏差问题。为此，腾讯广告团队引入用户和广告自监督对比学习，并与原有监督学习任务联合训练，提升了模型泛化能力。其次，团队还对负采样进行了优化，并在广告对比学习中引入 Momentum Contrast机制，提升了全库感知。

粗排在业界长期以来采用类似精排、但更简单的LiteCXR模型。然而，考虑到粗排的定位是典型的集合选择问题，其核心优化目标应该是排序一致性。所以，研发团队转向排序学习LTR模型。对比传统做法，LTR模型更加简洁高效，之前数据稀疏、样本选择偏差问题也得到大幅缓解。

## 物质基础：太极机器学习平台

与算法同等重要的是计算平台，尤其是对于广告系统，在线学习、高并发低延迟的推断，都离不开计算平台的支持，不然前文大模型的各种方案与创新，都只是空中阁楼。

从计算平台方面，首先需要它能支持千亿参数量模型的在线学习，也即广告大模型要能 7\*24小时实时迭代优化自己的效果；其次在发布模型时，需要快速

将300多GB的模型权重文件推送到全国上百个推理服务器，并启动、加载到计算设备中；最后，计算平台还要有能力支持快速推理，别看说的简单，但当模型达到千亿量级，优化推理速度可是难之又难。

混元AI大模型和广告大模型，其背后的物质基础，即腾讯「太极机器学习平台」。

太极机器学习平台由腾讯TEG云架构平台部的机智平台和数据平台部的Tesla平台协作共建而成，是该公司首个大规模云原生机器学习平台，更在中国信通院举办的2021云原生产业大会上获得了「云原生应用优秀案例」奖项。

先从技术上来看，通过整合与高效利用资源，太极机器学习平台旨在提供统一的机器学习框架和加速方案。该平台采用分布式参数服务器架构，具备业内领先的模型训练能力，提供GPU算力和训练加速框架，是业界第一梯队企业们公认的最佳选择。

太极机器学习平台在训练上的技术突破。除此之外，平台的大模型发布机制、推理机制都在技术上有本质的突破。

再从应用上来看，太极机器学习平台目前为多个公司级的业务提供训练平台与加速服务，其中针对广告应用场景及依附的大模型技术进行了性能优化，并得到了充分的工业应用检验。

太极机器学习平台在技术上具备独有优势，又能落地到实际应用场景中。那么，面对具有优化训练速度和推理需求的混元AI大模型和广告大模型技术，该平台是如何基于自身能力进行针对性优化和提升的呢？

首先是平台易用性提升大模型研发效率。

广告模型的开发通常涉及特征与样本生产、模型训练、模型服务等主要环节，在过去的系统中，这些环节的操作需要涉及多个子系统，算法开发者需要在多个系统之间切换，导致操作复杂，时间长成本高。为此，太极机器学习平台针

对广告场景打造了一站式广告平台，将主要的模型开发流程进行集成，过去多个子系统收拢到一处入口，大幅提升平台的易用性，提高模型研发的效率。

其次是针对不同类别的大模型「具体问题具体分析」。

混元AI大模型属于CV/NLP类预训练模型，基于太极机器学习平台进行研发。借助GPU算力，实现快速算法迭代和模型训练。该平台的预训练大模型训练框架支持数据并行和模型并行等分布式训练策略，其中模型并行包含张量并行、流水并行和专家并行，以及多种组合优化策略。

腾讯广告大模型属于广告推荐类稀疏大模型，太极机器学习平台对它采用了分布式参数服务器（Parameter Server, PS）架构。这种架构的优势在于存储模型参数和执行模型计算在各自的服务器上运行，如此一来，增加更多服务器意味着可以支持更大、计算需求更高的模型。这就为赋能腾讯亿级用户、海量广告内容的大模型提供了架构支撑。

此外，腾讯基于这种架构自研了参数服务器系统AngelPS，现在可以支持10TB级模型的训练。

Angle参数服务器架构设计。

最后是模型训练硬件加速能力和超大模型在线推理服务能力。

一方面，太极机器学习平台的模型训练硬件加速能力非常突出。我们就以混元AI大模型的训练为例，太极的加速方案高于业界其他方案3.7倍，能够有效提升模型的训练效率。

另一方面，该平台具备超大模型在线推理服务的能力。推理计算方面支持常规CPU计算和复杂模型的GPU计算加速；存储上，根据在线推理模型大小，自适应选择小模型本地内存加载和大模型AngelPS远程查询的部署策略。

就效果而言，目前在太极机器学习平台上，千亿参数、TB规模训练、百GB规模预测的腾讯广告大模型能够短时间内完成计算，并提供可靠的广告匹配结果。

可以这样说，太极机器学习平台实现了对混元AI大模型和广告大模型的助力，并最终提升了腾讯广告自身的推荐效率以及面向受众的推荐效果。



## 软硬联合：价值倍增

广告系统这种复杂的模型集合，也只有结合了硬件的算力与大模型的拟合力，才能尽可能高效。

很明显，在模型上，混元AI大模型对多模态的理解能力已经登顶各大榜单，广告大模型又直接展现出提升商品交易总额（GMV）的能力。算法上肯定属于领先水平了。

同时，计算平台能支持千亿参数模型的重度使用，训练、推理、部署发布一气呵成，可以说独此一份了。两大技术缺一不可，软硬联合起来，能力才可以得到最大程度的释放。

广告主能更精准、正确地将信息传递给目标受众，达成生意目标，用户也真正获取到自己想要的信息，形成双赢。

当然，GMV只是一个指标，只是大模型在当前阶段希望尽力提升的成果。我们还需要看到未来，腾讯广告大模型现在也一直在学习，它希望能更好地完成「广而告之」这一使命，将正确的信息传递给需要的人。也许当模型越来越强大，商品、服务信息能快速而精准地匹配给需要的人群，那整个社会的经济活力都将因为大模型而变得不同。

参考链接：

<https://toutiao.io/posts/0p8zys/preview>

<https://zhuanlan.zhihu.com/p/337444865>

<https://www.jiqizhixin.com/articles/2022-06-20-5>

<https://cloud.tencent.com/developer/article/1829554>