超大规模多模态预训练模型M6的关键技术及产业应用

Key Technologies and Applications of Extremely Large-Scale Multimodal Pre-Trained Model M6

林俊旸/LIN Junyang，周畅/ZHOU Chang，
杨红霞/YANG Hongxia

（阿里巴巴达摩院，中国 杭州 311100）
(Alibaba DAMO Academy, Hangzhou 311100, China)

DOI：10.12142/ZTETJ.202202007
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20220410.2248.004.html
网络出版日期：2022-04-11
收稿日期：2022-02-25

摘要：阿里巴巴达摩院研发了超大规模中文多模态预训练模型M6，并陆续推出了百亿、千亿、万亿和十万亿参数规模的预训练模型，实现了高效低碳的预训练，推动超大规模预训练模型的产业化应用。同时，推出了M6服务化平台，帮助广大用户快速使用大模型。未来，大模型在产业领域的应用将更加丰富。

关键词：多模态预训练；大规模预训练；图像生成；文本生成

Abstract: The extremely large-scale Chinese multimodal pre-trained model M6 is proposed by Alibaba DAMO Academy, and the 10 B, 100 B, 1 T, and 10 T versions of M6 are released. M6 has been trained efficiently with low carbon emission, and it has been deployed in multiple scenarios, which leads to the creation of new products as well as performance improvement. Also, to provide better services, the easy-to-use M6 platform for users to leverage large-scale pre-trained models is released by DAMO Academy.

Keywords: multimodal pre-training; large-scale pre-training; image generation; text generation

近年来，预训练技术的诞生与迅速崛起成为人工智能（AI）发展史的一大标志。基于无监督学习和弱监督学习的预训练具备强大的迁移能力，可以充分利用海量无标注数据，因此能够应用于多种不同类型的下游任务中。此外，研究人员通过扩大模型容量、扩大训练数据、降低人工标注的依赖等方式让模型取得更好的效果及通用性。随着模型规模和数据规模的不断扩大，模型效果也会显著提升。预训练大模型的研究具有深远的学术意义，并有着广泛的应用前景。

传统的预训练多集中于单模态数据，且多数的预训练工作均在英文数据上实现。在很长一段时间里，中文数据都缺乏大规模预训练模型和多模态预训练模型。自2020年以来，阿里巴巴达摩院认识到这一问题的重要性，提出了超大规模中文多模态预训练的课题。基于多模态表示学习以及超大规模预训练模型的研究，达摩院掌握了基于超大规模多模态预训练的核心技术，于2021年提出了超大规模中文多模态预训练模型M6。在之后的一年内，达摩院陆续发布了百亿、千亿、万亿和十万亿参数规模的超大模型。这些工作推动了低碳AI的发展，同规模的M6耗电量不到GPT-3的1/100。达摩院还积极推进M6大模型的产业化落地，这包括手机淘宝推荐、支付宝搜索推荐等100余种算法场景。同时，M6利用其能力支持多个行业实现创新产品的孵化，如AI服饰设计能力可以支持服饰制造行业。同时，M6还推出了大规模预训练平台，使得大模型的应用以服务化的形式对外提供服务。该平台也是当前下游任务覆盖最广的预训练平台。平台化使得大模型同时服务于学界和产业界，大幅降低了大模型的门槛，并让AI大模型简单易用。

1 M6技术进展与突破

此前的研究[1]指出，随着数据规模、模型规模和计算资源的不断增长，模型能力也会不断提高。在过去的几年里，学习了海量无监督数据的预训练模型的规模实现了指数级增长。2018年，最大规模的BERT[2]和GPT[3]参数规模仅约为3亿。2019—2020年，具有15亿参数规模的GPT-2[4]、83亿参数规模的Megatron[5]、110亿参数规模的T5[6]、170亿参数规模的图灵自然语言生成（Turing-NLG）[7],以及史无前例

的1 750亿参数规模的GPT-3[8]陆续出现。随着模型规模的增大，模型学习大规模数据的能力逐渐增强，并展现出强大的小样本和零样本学习的能力。

以上工作大多集中于纯文本的预训练，多模态预训练的工作规模较小，且主要针对理解类的任务。另外，还缺少成熟的中文多模态预训练模型。针对以上问题，达摩院联合阿里云机器学习平台及清华大学在2021年1月提出了首个中文领域的超大规模多模态预训练模型M6[9]，并在KDD2021（2021年数据挖掘顶会）发表了相应的论文。M6在超大规模的中文多模态数据上做预训练，兼容多模态及单模态的理解与生成能力。M6的下游任务包括视觉问答、视觉描述、跨模态检索、基于文本的图像生成和图像编辑、文本摘要、诗歌生成等，覆盖领域广。同时，针对超大模型训练效率低的问题，达摩院联合阿里云机器学习平台实现了基于图形处理器（GPU）的混合专家（MoE）机制的开发，这不仅是中国首个基于MoE的超大规模预训练的实践，也是全球首个基于MoE的、最大规模的多模态预训练模型。随后，达摩院针对MoE机制做了细致的分析和优化，提出专家分组机制[10]，并通过一系列的优化，提升了训练效率，降低了资源消耗，在480个GPU上实现了万亿参数规模的M6预训练。该MoE机制是多模态领域的首个万亿参数规模的预训练模型。相比于同为万亿参数规模的MoE模型Switch Transformer[11]（使用了2 048个TPU），M6-T更为低碳和高效（仅使用480个GPU）。考虑到不断增长的参数规模，为了实现更为绿色环保的模型训练，达摩院开始研究极限参数规模即十万亿参数的M6预训练[12]，提出共享解除的训练机制，实现了512个GPU的十万亿参数规模M6模型的预训练，助力绿色环保的人工智能的发展。

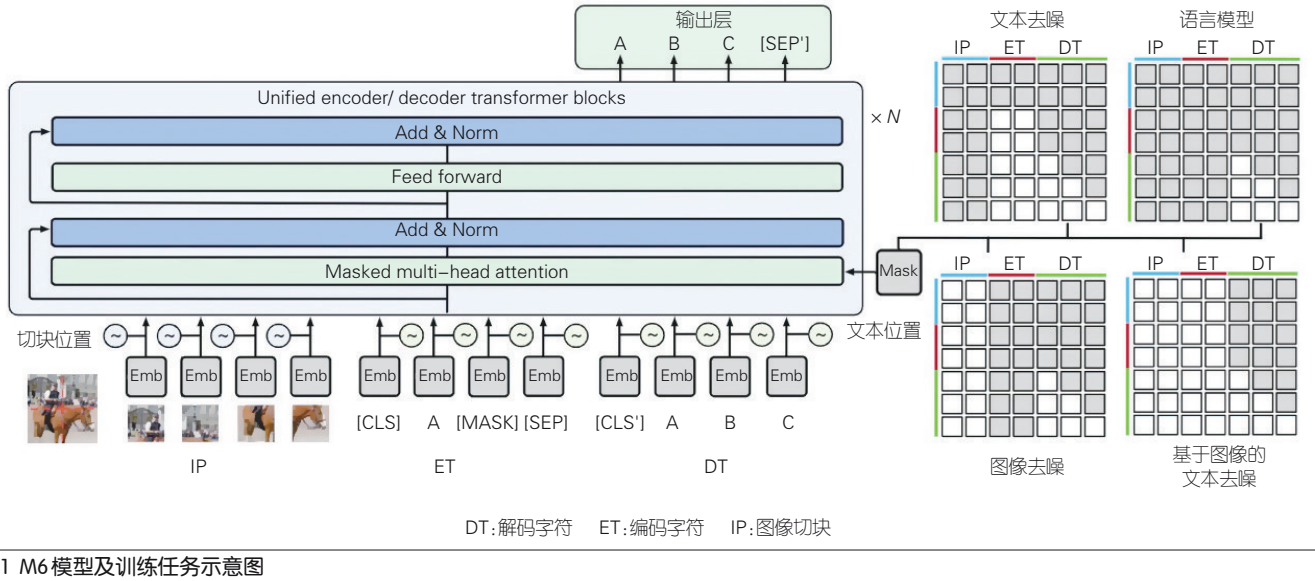
1.1 百亿参数规模M6

M6系列模型以自然语言预处理（NLP）和多模态领域中最为主流的Transformer架构为主体，针对不同模态的数据设计了特定的模块及预训练任务（如图1所示）。在多模态预训练模型中，研究人员针对图像数据和文本数据的差异，对图像数据进行特征提取的预处理。不同于物体检测提取特征的方式，M6采用了图像切块、backbone模型（如ResNet）等提取特征的方式，并根据块的位置提供位置表示。在主体架构上，M6依然采用Transformer block堆叠的方式，具体架构如图1所示。

在预训练任务上，为了让模型兼具多模态及单模态的理解和生成能力，研究人员设计了有无图像信息条件的文本去噪和语言模型的任务。这样一来，模型通过学习便可根据上下文还原和续写文本，从而掌握跨模态的理解和生成能力，并可以便捷地迁移到多种类型的下游任务中。

传统的分布式训练因其有限的显存，无法支持百亿参数规模的模型训练。为了打破显存的限制，研究人员在数据并行的基础上，增加了重计算的机制，并采用优化器状态分片和梯度分片的策略，在单台机器8个GPU的条件下即可训练一个百亿参数规模的M6模型。

结合上述模型架构、训练任务及大模型训练机制，达摩院采集并处理了超过2 TB的中文图像及接近300 GB的中文文本数据，然后使用这份大规模数据对M6进行预训练，并



▲图1 M6模型及训练任务示意图

将这些数据迁移到多种类型的下游任务中（包括视觉问答、视觉描述、跨模态检索、基于文本的图像生成和图像编辑、文本摘要、诗歌生成、故事生成、自然语言理解等）。实验结果证明，M6大模型在中文视觉问答和视觉描述中均能取得最优的效果。同时，M6大模型具备极强的小样本学习能力，在多个小样本任务评测上超出同期的中文预训练大模型（CPM）[13]。

1.2 千亿参数规模M6-MoE

在攻克了百亿参数规模M6的难关后，研究人员使用阿里巴巴自研框架Whale,实现了专家并行机制，并将其和M6模型相结合，在GPU集群上训练出首个千亿参数规模的多模态预训练模型。具体而言，研究人员使用Whale框架的算子拆分功能，将多个专家网络分配到多个GPU上，并使用all-to-all通信机制实现输入信息的分配和聚合。在此基础上，研究人员添加了峰值显存优化、通信优化和混合精度优化的一系列策略，在128个A100上达到1 440个样本/s的训练效率。实验结果表明，相较于百亿参数规模M6，M6-MoE模型的参数规模虽然增长了10倍，但训练效率依然高于M6。M6-MoE的语言模型困惑度评测具有较大优势。

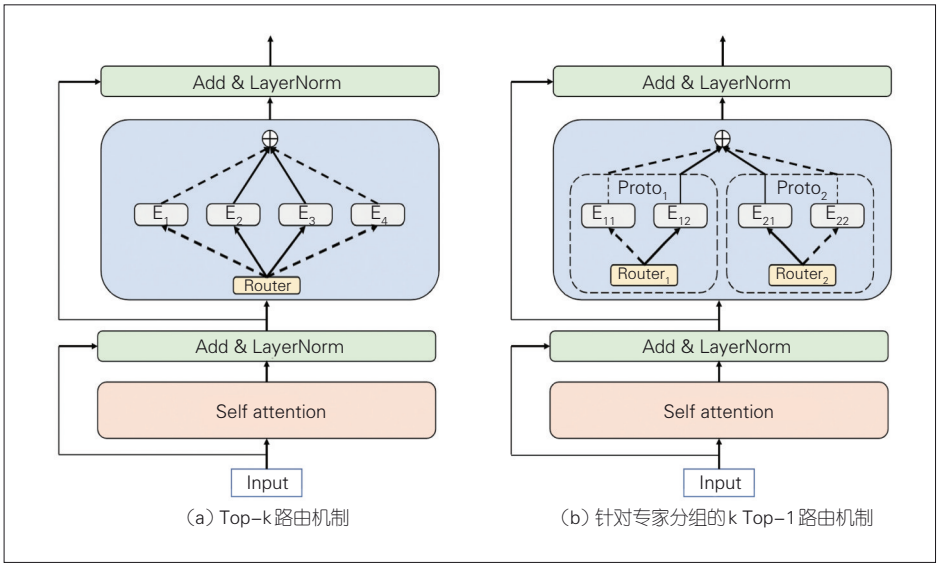
1.3 万亿参数规模M6-T

研究人员在研究混合专家机制与大模型训练的结合并分析其中的问题后发现，辅助损失的必要性比较低，并且其中的top-k路由机制中k的大小对模型效果具有决定性的影响。随着k的增大，模型性能逐渐提升，但同时也会出现边际效应递减及模型的训练效率显著下降的情况。针对以上问题，研究人员设计了专家分组机制[10]（如图2所示），将专家网络分成多组后再以并行的方式对每组进行top-k路由。实验表明，分组机制能够显著提升模型的训练效率；当k值较大时，分组机制的模型表现也优于传统方法；在上下游的语言模型困惑度评测上，分组后的模型均显著优于传统方法。

研究人员将此方法应用于万亿参数规模的M6-T模型的训练，并优化了大模型中的显存占用，即在480个V100-32G上仅用约3天的时间便实现了万亿参数规模M6-T模型的预训练。实验表明，相较于基线模型，结合了专家分组机制的M6-T收敛速度显著加快，同时损失也更少。

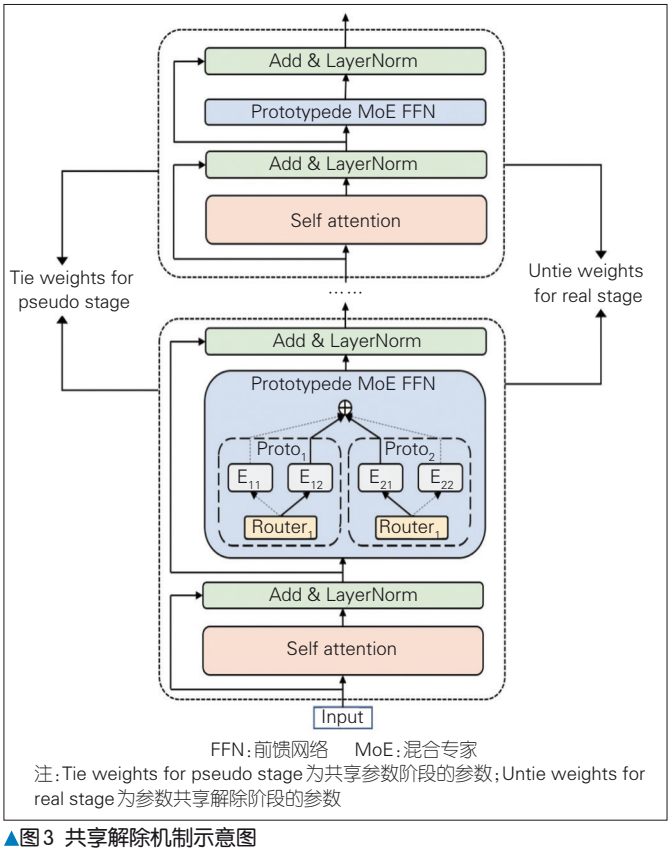
1.4 十万亿参数规模M6-10T

经过研究，达摩院提出了更加低碳的共享解除的训练机制（具体如图3），并且设计了粒度可控的CPU offload（中央处理器负载迁移）[12]，成功地用10天左右的时间在512个



（a）Top-k路由机制　（b）针对专家分组的k Top-1路由机制

▲图2 专家分组机制示意图



FFN：前馈网络　MoE：混合专家

注：Tie weights for pseudo stage为共享参数阶段的参数；Untie weights for real stage为参数共享解除阶段的参数

▲图3 共享解除机制示意图

GPU上完成十万亿参数规模的M6–10T的预训练。

实验证明，M6–10T方案在收敛和下游迁移的过程中具有有效性。同时，它在十万亿参数规模的M6–10T模型上做出了成功实践，仅用10天左右的时间便取得了非常突出的收敛效果。十万亿模型达到相同预训练损失所需的样本量仅为万亿模型的40%，这充分显示出该机制的优势。

2 M6的产业化应用

超大规模多模态预训练取得的成功也意味着它将成为社会发展的重要基建设施，为各类下游任务提供支持。目前，针对互联网生态中各种复杂的业务场景，M6做出了相应的优化，在服装设计、自动文案、金融服务、搜索推荐等业务场景中实现商业落地，产生了巨大的商业价值以及社会价值。

此外，千亿参数和万亿参数M6大模型的研发，大力推动了低碳大模型的发展，并助力绿色环保AI的发展，响应了中国的碳中和战略部署。相比于传统方法，M6元生款数智制造结合犀牛环保面料的研发应用，能在全链路中减少30%以上的碳排放。每卖出一件元生款链路生产的服装，就能减排0.35 kg二氧化碳。也就是说，卖出50件就相当于种下一棵树。

2.1 AI图像生成在服饰制造等行业的应用

M6具有的一项重要能力是基于文本的图像生成和图像编辑。为了充分利用Transformer架构对大数据的高效处理和泛化能力，研究人员在文到图的生成架构上，选用了两阶段模型：第一阶段，需要将图像进行离散编码；而在第二阶段，则利用预训练模型M6来建模文本和图像离散编码的关系。为了进一步提升图像生成的清晰度和细节丰富度，研究人员将序列长度从1 024延长至4 096，并加入了稀疏注意力，成功将生成图像分辨率提升至1 024×1 024。在服饰制造行业，M6模型能够生成具有高清晰度和丰富细节的图片（如图4），并利用AI服饰设计的能力在服装制造行业实现落地。

为了进一步提升M6在图像生成过程中的可控性和效率，研究人员提出了基于M6的非自回归的图像生成模型[14]，实现在不同控制条件下（包括文本、图像、风格等）的图像生成，使得模型具备了图像编辑的能力。

基于以上能力，M6以AI设计师的身份参与到服饰制造等行业中。对该行业传统企划链路来说，从样式规划到最终的生产上架，往往需要耗时半年。例如，一件冬季的羽绒服，需要在初夏时就决定其款式并开始漫长的人工设计和反复的打样修改。2021年，达摩院智能计算实验室与阿里巴巴犀牛智造深度合作，借助数字化的能力，自动化地捕捉流行机会，并结合M6的生成能力，为商家提供敏捷高效的设计和丰富的体验。

目前，M6生成的服装图片已通过质检并达到商家的质量标准。首期文到图生成的人工质检优质率约为10%，比人工设计师的效率高10～20倍。研究人员会根据人工反馈的结果不断优化模型的生成质量，提升优质率。研究人员不断优化模特试衣算法，协助犀牛智造为商家提供更多的模特试穿效果图。

为了进一步验证并应用M6的超强图像生成以及创新能力，达摩院和某车厂合作产出概念型车型以及未来型车型，以辅助汽车设计师进行车型设计，并和阿里云LOGO服务团队合作产出种类丰富的LOGO配图供客户挑选（如图5所示）；另外，还和蚂蚁花呗团队合作生成宠物头像图作为宠物唯一身份认证，通过少样本或者文本描述即可生成符合要求的图像。

2.2 工业级文案生成

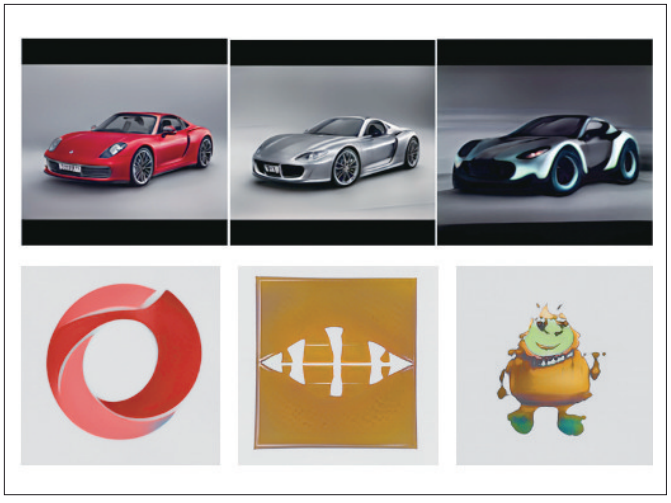
M6的另一个重要能力便是文本生成。该能力能够运用于视觉描述、视觉问答、文本摘要、问答、对话、文案创作等。目前M6的文本生成已经达到工业级标准。在训练语料较少的情况下，M6的优势更加明显，仅以此前5%的数据再

▲图4 羊羔绒毛衣生成示例

通过微调便可达到更优的效果，人工审核通过率最高可达85%。目前M6的文本生成已经成功应用于UC畅聊中台、手机淘宝和支付宝Push文案、阿里小蜜智能客服、阿里小蜜虚拟直播、支付宝腰封文案中。在UC畅聊平台中，算法团队利用畅聊评论中台沉淀的历史用户回复数据对M6模型进行微调，批量构造机器人自动回复，并通过M6模型每天的自动总结和生成推荐理由，覆盖社会、娱乐、体育、汽车等10个内容类目，引发用户回复率显著提升（约10%）。

虚拟直播中，在不足1 000条标注数据的条件下，M6结合非自回归生成方法（SANA）[15]实现了在双语评估替换（BLEU）评测上超过10%的大幅提升，人工审核通过率超70%，大大提升了智能剧本的生成效率（如图6所示）。

在支付宝的腰封文案场景中，M6利用其生成能力为该业务产出主副标题。相较于基线模型，M6产出的文案在线上实现了超5%的点击率。



▲图5 概念车LOGO配图示例

ASR原始文本	M6输出
那宝宝你用我们这款燕麦爽肤水真的是特别的合适啊。在我们7号链接这款燕麦爽肤水呢，它里面是顾名思义就是蕴含了一个燕麦成分，对不对？ 我们这个燕麦成分呢，它不是一般的燕麦，它它是北美进口的微米级别的燕麦添加在里面，然后形成了一个燕麦黄铜的成分。燕麦黄铜呢是可以有效的去强韧到我们肌肤的角质层， 然后呢也可以给我们肌肤建立。一个隐形的保护膜，这个保护膜呢是可以有效的隔绝住外界环境啊，天气对我们肌肤带来的一个损伤。然后也可以把补充到基底的一个水分养分牢牢的去锁起来啊，宝宝们，对， 而且我们这个基底呢也是。 可也是可以也是可以给我们基底带来一个密集性的保湿补水锁水的一个功效，让我们肌肤的每一个角落喝饱水。 因为它里面呢是蕴含了一个透明质酸的成分，再加脚上。 角鲨烷的一个成分是双重深水炸弹，一起打入到基底扩散出来， 给肌肤做到一个密集性的保湿补水啊。	宝宝我们这款燕麦爽肤水真的是特别的合适。这款燕麦爽肤水呢它里面是顾名思义就是蕴含了一个燕麦成分，我们这个燕麦成分呢，它不是一般的燕麦，它是北美进口的微米级别的燕麦添加在里面，然后形成了一个燕麦黄铜的成分。燕麦黄铜呢是可以有效的去强韧到我们肌肤的角质层，然后呢也可以给我们肌肤建立一个隐形的保护膜，这个保护膜呢是可以有效的隔绝住外界环境天气对我们肌肤带来的一个损伤。然后也可以把补充到基底的一个水分养分牢牢的去锁起来啊，而且我们这个基底也是可以给我们基底带来一个密集性的保湿补水锁水的一个功效，让我们肌肤的每一个角落喝饱水。因为它里面呢是蕴含了一个透明质酸的成分，再加角鲨烷的一个成分是双重深水炸弹，一打入到基底扩散出来，给肌肤做到一个密集性的保湿补水啊。

▲图6 生成剧本示例

2.3 表示提取功能助力搜索推荐等应用

在支付宝搜索场景中，结合经典的双塔结构，M6在场景语料上进行微调，然后应用于向量召回中。更加关注语义和知识信息的预训练模型能够有效提升检索的准确率。超大规模模型强大的小样本学习和理解能力则助力冷启场景和长尾查询词的召回，从而实现了搜索页点击率（PVCTR）的显著提升。

在更加依赖内容表示的内容导购轻应用中，M6的作用更加明显。在家居场景中，相比于传统的图像单模态表示，M6的多模态表示包含更加丰富的信息，在召回和排序场景均能实现效果的提升，在风格一致性上的表现尤其突出。这体现了多模态表示学习的特点。

在搜索推荐以外的场景中，达摩院联合斑马汽车，使用M6来提升车辆检索准确率。在该场景中，系统需要根据用户指令检索出最相关的车辆，而M6提供的用户指令表示能够更加准确地反映用户意图，帮助下游的车辆检索模型指代最相关车辆。这能够将准确率提升5%。

3 M6平台化服务

超大规模多模态预训练的应用意味着大模型开始实现对各行各业的支持。为了让大模型更加简便易用，达摩院推出了M6预训练平台，以服务化的形式将大模型的能力运用至各行各业中。凭借平台的易用性、训练高效低碳、下游任务覆盖广泛等特点，M6服务化平台服务了阿里巴巴集团内部各类业务。同时，M6预训练平台也已经通过阿里云对外发布。

M6多模态预训练服务化平台具有3个核心能力点：

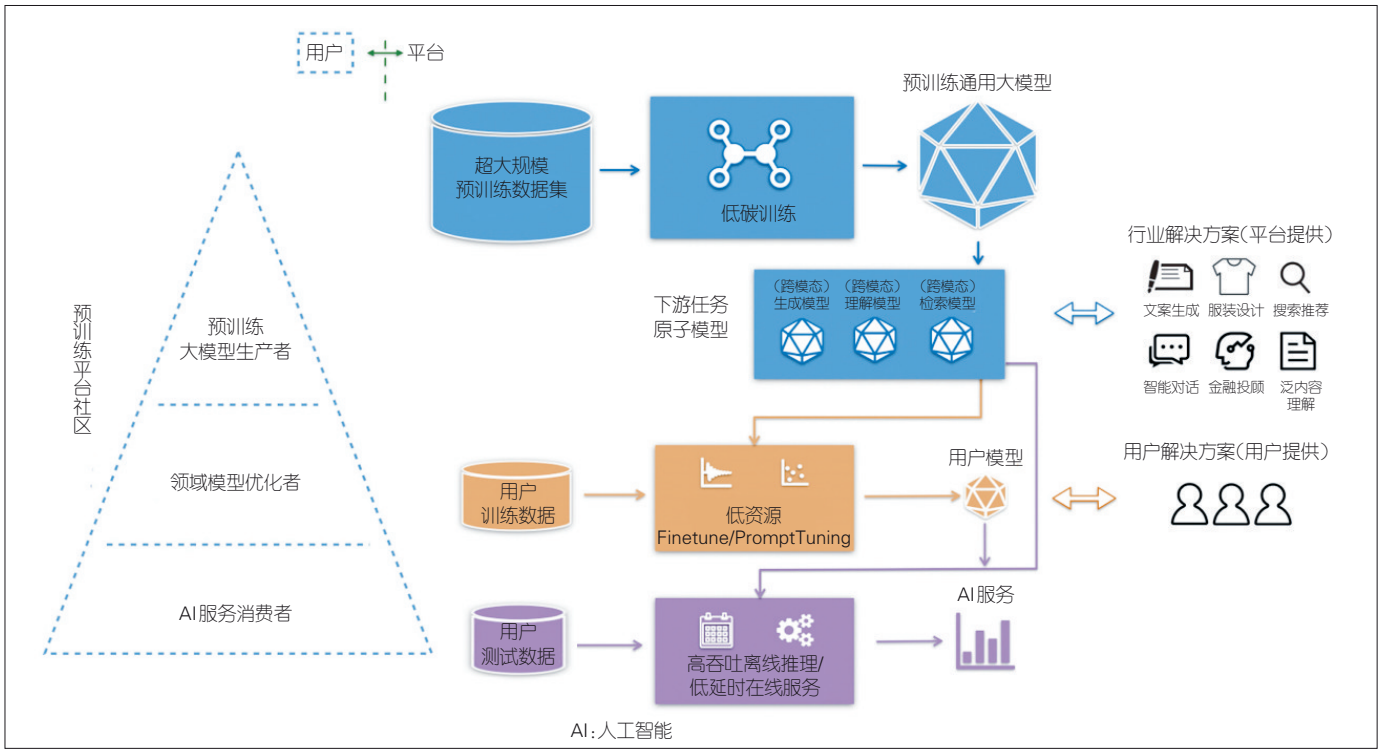
• 任务形式覆盖广。该平台是目前下游任务覆盖最广泛的预训练平台，覆盖多模态输入输出的常见任务超20个。

• 高性能&简单易用。无须关注分布式训练、数据输入输出、数据并行、底层实现以及训练评估流程等细节，准备好输入数据，再对参数进行简单修改，即可实现多机多卡的训练或推理任务。

• 下游任务内源+支持自定义模型改造。用户可以在M6的上层实现自定义模型，无须关注过多大模型细节。

M6平台集成了多种微调形式，并提供了高效&低碳的分布式训练、低延迟的模型服务、统一的数据&模型管理、一键式模型部署方案。用户可以根据自身数据情况，选择使用软件开发工具包（SDK）调用、微调、自定义模型等方式来灵活支持自身的下游任务。M6平台的整体框架如图7所示。

相较于传统小作坊式的AI服务，集中式的数据、算力



▲图7 平台整体架构示意图

开发模式有着更好的平台粘性，因此我们希望能将内部的成功经验以服务化平台的方式让更多外部用户获益。同时，基于阿里云强大的基础设施，M6预训练平台能提供更完善的全链路服务，实现AI普惠化目标。

4 结束语

自2020年以来，达摩院深入研究超大规模多模态预训练关键技术，在2021年陆续发布百亿、千亿、万亿、十万亿参数规模的M6模型。并且，达摩院持续发力，解决大模型应用落地难的问题，并专注于大模型的产业化落地，在服饰制造、工业级文案生产、搜索推荐等场景实现了应用。在此基础上，达摩院将M6的能力以服务化的形式集成到M6预训练平台中并对外发布，帮助行业、企业和个体快速使用大模型，推动普惠AI发展。

未来，大模型在产业领域的应用将更加丰富。前景主要包括：推动传统产业智能化转型，催生基于智能模型的新产业，以及改变人类社会的生产和管理模式等。预训练大模型还有亟待突破的几个难题：

（1）目前的主流实践是先通过训练大模型得到参数规模大、精度高的模型，再基于下游任务数据，通过剪枝、微调的方法将模型的体积压缩，在基本不损失精度的情况下减轻部署压力。目前，业界还没找到通用的、直接训练小型模型就能得到较满意精度的办法。

（2）训练千亿、万亿模型动辄需要上千个GPU卡，这对大模型的推广和普惠带来了很大的挑战。

（3）因为参数量大，目前预训练模型主要针对大量非结构化数据。如何与知识等结构化数据进行结合，更加有效地进行认知推理，也是一个非常大的挑战。

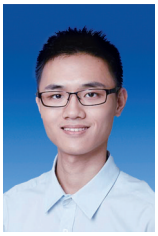
以上难题使得大模型参数竞赛进入“冷静期”，而大小模型在云边端协同进化则带来了新的突破可能性。云边端协同使小模型更容易获取通用的知识与能力。小模型专注于在特定场景做极致优化，从而获得性能与效率的提升。大小模型的协同进化可以更好地服务于更加复杂的新场景，例如元宇宙、数字人等。同时，该体系更有利于保护用户数据隐私。

参考文献

[1] KAPLAN J, MCCANDLISH S, HENIGHAN T, et al. Scaling laws for neural language models [EB/OL]. (2019-05-10) [2022-01-12]. https://paperswithcode.com/paper/mutual-information-scaling-and-expressive
[2] DEVLIN J, CHANG M, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [EB/OL]. [2022-01-10]. https://

paperswithcode.com/paper/bert-pre-training-of-deep-bidirectional
[3] RADFORD A, NARASIMHAN K. Improving language understanding by generative pre-training [EB/OL]. [2022-01-12]. https://paperswithcode.com/paper/improving-language-understanding-by
[4] RADFORD A, WU J, CHILD R, et al. Language models are unsupervised multitask learners [EB/OL]. [2022-01-10]. https://paperswithcode. com/paper/language-models-are-unsupervised-multitask
[5] SHOEYBI M, PATWARY M, PURI R, et al. Megatron-LM: training multi-billion parameter language models using GPU model parallelism [EB/OL]. [2022-01-10]. https://www. researchgate. net/publication/335908286_Megatron-LM_Training_Multi-Billion_Parameter_Language_Models_Using_GPU_Model_Parallelism
[6] RAFFEL C, SHAZEER N, ROBERTS A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer [EB/OL]. [2022-01-10]. https://paperswithcode.com/paper/exploring-the-limits-of-transfer-learning
[7] ROSSET C. Turing-NLG: a 17-billion-parameter language model by Microsoft [EB/OL]. (2020-02-13) [2022-01-10]. https://www. microsoft.com/en-us/research/blog/turing-nlg-a-17-billion-parameter-language-model-by-microsoft/
[8] BROWN T B, MANN B, RYDER N, et al. Language models are few-shot learners [EB/OL]. [2022-01-10]. https://paperswithcode. com/paper/language-models-are-few-shot-learners
[9] LIN J Y, MEN R, YANG A, et al. M6: a Chinese multimodal pretrainer [EB/OL]. (2021-03-01) [2022-01-12]. https://paperswithcode. com/paper/m6-a-chinese-multimodal-pretrainer
[10] YANG A, LIN J, MEN R, et al. Exploring sparse expert models and beyond [EB/OL]. (2021-03-31)[2022-01-12]. https://arxiv.org/abs/2105.15082v2
[11] FEDUS W, ZOPH B, SHAZEER N, et al. Switch transformers: scaling to trillion parameter models with simple and efficient sparsity [EB/OL]. (2021-11-11)[2022-01-12]. https://arxiv.org/abs/2101.03961
[12] LIN J, YANG A, BAI J, et al. M6-10T: a sharing-delinking paradigm for efficient multi-trillion parameter pretraining [EB/OL]. (2021-10-08)[2022-01-12]. https://paperswithcode. com/paper/m6-10t-a-sharing-delinking-paradigm-for
[13] ZHANG Z, HAN X, ZHOU H, et al. CPM: a large-scale generative Chinese pre-trained language model [EB/OL]. (2021-10-08)[2022-01-12]. https://paperswithcode.com/paper/cpm-a-large-scale-generative-chinese-pre
[14] ZHANG Z, MA J, ZHOU C, et al. M6-UFC: unifying multi-modal controls for conditional image synthesis [EB/OL]. (2021-10-08)[2022-01-12].https://paperswithcode.com/paper/ufc-bert-unifying-multi-modal-controls-for
[15] WANG P, LIN J, YANG A, et al. Sketch and refine: towards faithful and informative table-to-text generation [EB/OL]. (2021-10-08) [2022-01-12]. https://paperswithcode.com/paper/sketch-and-refine-towards-faithful-and

作 者 简 介



林俊旸，阿里巴巴达摩院智能计算实验室算法专家；主要研究领域包括自然语言处理及多模态表征学习（侧重于多模态预训练及其应用），并参与研究大模型的低碳训练、提示学习、轻量化应用等问题；参与研究的全球最大规模的十万亿参数多模态预训练模型M6广泛应用于下游应用（如图像生成、自然语言生成、跨模态检索等）；发表论文30余篇。

周畅，阿里巴巴达摩院高级算法专家，并担任NeurIPS/ICML/KDD/WWW等会议PC member；主要研究领域包括表征学习、推荐系统、多模态预训练；带领团队研发了全球最大参数规模的多模态预训练模型M6，大幅优化了大规模预训练模型的计算效率，取得了中国首个大模型商业化落地成果，并作为算法负责人研发了大规模GNN训练平台AliGraph，获得中国电子学会科学技术进步奖一等奖；发表文章30余篇。

杨红霞，阿里巴巴达摩院人工智能科学家，曾担任IBM全球研发中心Watson研究员、雅虎首席数据科学家等；主导阿里巴巴下一代人工智能突破性技术——认知智能技术的发展与场景应用落地；曾获2019世界人工智能大会最高奖卓越人工智能引领者（Super AI Leader，简称SAIL奖）、2020年国家科学技术进步奖二等奖、杭州市领军型创新团队、2021年电子学会科学技术进步奖一等奖；发表文章80余篇，获得中国和美国专利近20项。