



6G网络面向大模型的分布式学习白皮书

Foundation Model-Oriented

6G Network Distributed Learning Whitepaper

目录

摘要	3
1 关于 6G 智能普惠的断言	4
1.1 AGI 服务将在 6G 时代普及	4
1.2 大模型提供智能服务的成本接近于零	4
1.3 每个人都将拥有自己的大模型	5
1.4 网络传输内容将从数据转向模型	5
2 面向大模型的 MFD 分布式学习模式	7
2.1 联邦学习模式的挑战	7
2.2 MFD 分布式学习模式	8
2.3 模型在网络内像 IP 包一样传输	11
2.4 模型可以自由缩放	13
2.5 支持异构模型间的知识分享	15
2.6 特轻量化自组织的学习流程	17
3 MFD 应用示例	19
3.1 Net4LM 的智能协作机器人	19
3.2 LM4Net 站间协同	21
4 总结和展望	22
5 参考文献	23

摘要

6G愿景是实现智能普惠，支持将AI服务随时、随地的提供给每个人，这将驱动6G网络成为一个分布式AI计算平台。而到6G时代，端边云都将部署不同尺寸的大模型，大模型无处不在，端边云节点之间交互的主要内容将从数据转为模型，由此，本文提出一种新的面向大模型的、去中心化的、原生支持无线网络边缘异构动态环境的MFD (Model Follow Data)分布式学习模式，旨在让模型在靠近数据的地方进行训练和推理，避免大量数据传输带来的传输能耗和隐私问题，同时可以充分利用6G网络的分布式AI算力。针对联邦学习存在依赖可信中心节点、通信开销大和异构支持性差等问题，MFD将神经网络模型构建为通信数据包，使得模型在网络中能够像数据包一样传输，并且可以自由缩放，实现异构模型间高效的知识分享和轻量化自组织的学习流程。可以预测，MFD将成为6G网络的原生分布式学习模式，支撑AGI服务的普及。

1 关于 6G 智能普惠的断言

1.1 AGI 服务将在 6G 时代普及

大模型 (Foundation model) 技术的快速崛起并已取得了非凡的成功。而大模型技术还将会持续快速演进，以GPT为例，从2020年的GPT3到2023年的GPT4，再到未来的GPT8乃至更高版本，大模型的尺寸不断增加，并且可以达到令人印象深刻的智力水平；而StarGate项目旨在推动人工智能研究和开发的界限，并提出到2028年实现人工通用智能 (AGI) 的目标。基于这样的趋势，我们预测基于AGI的智能服务将在6G时代普及。

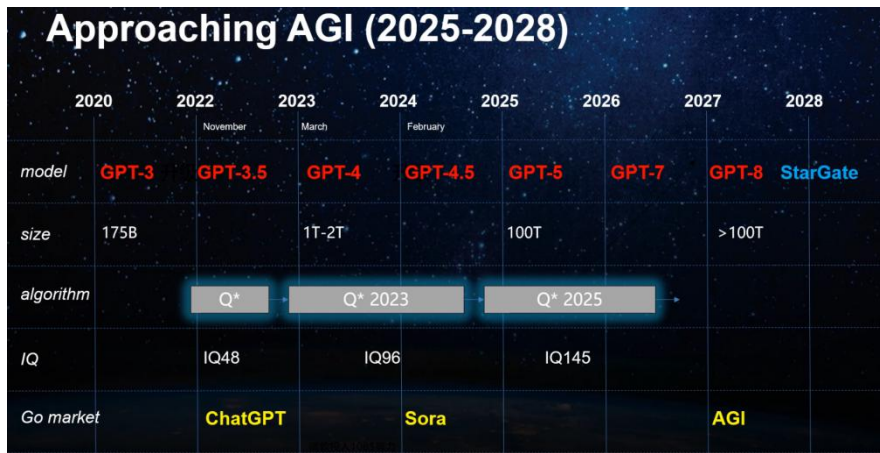


图 1 通用人工智能 (AGI) 发展进程

1.2 大模型提供智能服务的成本接近于零

规模理论 (Scaling Law) 是大模型背后的核心逻辑，从目前看也是通往AGI最为可行的路径。而从经济价值上看，大模型可以通过模型缩放来生成或产生几乎免费的满足用户需求的各类智能服务，而无需手工制作。以OpenAI提供的GPT系列服务为例，其收费持续下降，近期推出的GPT3.5 Turbo替代产品GPT-4o mini，收费大幅下降到百万 token输入/输出15美分/60美分，比GPT3.5 Turbo便宜超过60%以上，而对比GPT-4o

的收费更是便宜了近97%，甚至一些公司已开始推出了免费的大模型服务。而随着大模型参数共享、权重剪枝、压缩与蒸馏、硬件加速、分布式训练/推理等相关技术的发展和突破，基础大模型在推理速度和效率将不断提升，并且随着大模型的普及和规模效应，也使得单次服务的边际成本下降，到6G时代，大模型的智能服务成本将接近于零。

1.3 每个人都将拥有自己的大模型

目前大模型通常部署在云端，用户通过互联网访问各种智能服务。这种完全在数据中心集中部署的方式往往意味着巨量的算力，及随之带来惊人的能耗，如Stargate项目预期需要百万级AI芯片，数十亿瓦的功率来运行。而随着移动终端和在网算力能力的极大提升，已可以将大模型分布式部署在终端和网络中，一方面是数据安全和用户隐私保护的提升；另一方面，分布式部署也有助于解决大模型的所有权问题，如部署在终端的大模型个性化增量更新过程中，不可避免的涉及用户拥有的个性化数据，由此，用户将不可避免的倾向于获得更新后的大模型所有权。

1.4 网络传输内容将从数据转向模型

从大模型的规模理论来看，模型性能与计算量、模型参数量和训练数据量之间存在正相关的关系。这也导致模型的参数量在不断增加，从百亿到千亿甚至万亿。相应的，用来训练大模型的数据量，也需要以指数级暴增。据网络公开的数据，以 OpenAI 为例，从 GPT-1 到 GPT-3，其训练数据集就从 4.5GB 指数级增长到了 570GB，以此类推，那 GPT-5、GPT-6 需要的训练数据更是天文数字。



图 2 大模型性能与参数量和数据量之间的关系

构建 AGI 大模型，数据将成为首要障碍。二十几年来互联网积累的海量数据可以说为此次 AI 浪潮兴起奠定了坚实基础，而人工智能研究和预测组织 Epoch 在其发表的论文里预测，现有的高质量文本数据会在 2027 年消耗殆尽，未来的大模型进一步演进甚至实现 AGI，并应用到社会生活及各行业中，势必需要挖掘更多的高价值数据。值得注意的是，传统的模型训练方式是将大量数据收集到数据中心进行集中训练，即数据跟随模型（Data Follow Model, DFM），将带来巨大的通信开销和安全隐私风险，这种数据跟随模型的模式需要改变。而随着大模型被直接部署在终端和网络，将为进一步挖掘本地数据价值提供了可能性，如通过模型迁移等方式进行个性化的智能定制，相比与原始大模型可以更加高效准确的提供推理决策。而这些分布式大模型在各自学到新的知识后，将很自然的通过相互学习分享来进一步提升模型能力，随着大模型无处不在，可以预见，未来网络中传输的主要内容将从数据转变为模型。

基于以上的判断，到 6G 时代要实现大模型智能服务随时、随地的提供给每个人，这其中 6G 网络将扮演重要角色。一方面，结合 6G 的通感能力，大模型可以更容易与

物理世界同步，成为世界模型；另一方面，完全基于集中的数据中心来提供智能服务的模式，存在海量通信开销、安全隐私及集中电力供给等挑战，而 6G 网络可以作为一个分布式 AI 计算平台，来更高效实现智能普惠。

2 面向大模型的 MFD 分布式学习模式

2.1 联邦学习模式的挑战

联邦学习模式做到了传输的是模型而不是数据，并是一个解决数据所有权和隐私问题的价值路径，已在一些领域取得了成功的应用。联邦学习的目标是将数据放在用户本地进行模型更新，然后将模型参数发送给数据中心。数据中心收集参与者的模型更新参数，并进行聚合操作然后再下发给愿意参与联合训练的个体。联邦学习虽然传递的也是模型，但是在面向未来分布式大模型的学习和 6G 网络场景仍存在一些不足。

首先，为了进行多节点的模型聚合，需要有一个可信的中心服务器，所有的节点都是与中心服务器进行模型交互，这不仅带来了通信和计算瓶颈，而且也存在较大的安全风险[5]。虽然去中心的联邦学习（DFL）也有大量的研究工作[6][7][8]，却也还面临失去中心管理导致效率降低、模型版本混乱、和性能下降等问题[5]。

其次，联邦学习频繁的模型梯度交互带来了巨大的通信开销[9]，特别是面向大模型的联邦学习。以 Llama2-7B 模型为例，10 个节点进行 100 轮次的全局训练需要的传输量达到了 28TB。即使通过模型压缩和设备抽样等优化技术，通信依然会成为最大的瓶颈。此外，无线网络传输速率的不对称性（上行速率远低于下行速率）会进一步加剧移动终端的传输差异，从而引发通信瓶颈[10]，影响联邦学习的性能

再次，联邦学习的异构支持性差，其分布式训练通常要涉及多样的终端设备，以覆盖大量的数据样本去提升深度神经网络模型分布式训练的性能。在现实中，参与训练的终端设备大多是异构的，即具有不同的存储容量、计算能力和网络条件，由于联邦学习分布式训练的同步性原则[11]，整个联邦学习系统的训练时延可能会因为一些过长训练时延的终端设备而急剧增加，产生“落后者效应”[12]。另外，不同设备上运行的模型也存在异构性，在结构，尺寸和超参配置上不尽相同，为联邦学习进行模型的聚合带来了挑战[13]。为此，6G 网络需要一种非联邦学习的分布式学习模式，需要结合无线网络和无处不在的大模型特点进行原生设计。

2.2 MFD 分布式学习模式

在 6G 时代，模型将成为重心，数据在哪里，模型就去哪里，让模型跟随数据 (Model Follow Data, MFD)。6G 网络在连通海量终端的同时，原生支持模型在靠近数据的地方进行训练和推理，并在无线网络内实现模型传输和分享，这是模式的转变 (shift of paradigm)，称之为 MFD 分布式学习模式。

MFD 分布式学习模式避免了大量数据传输带来的传输能耗和隐私问题，同时可以充分利用分布式的算力，减少对数据中心的述求。MFD 的理念是将神经网络模型构建为通信数据包，使得模型在网络中能够像数据包一样被高效传输。同时，MFD 不需要像联邦学习一样，有一个控制锚点对整个系统的模型交换进行控制，提高了部署的灵活性。

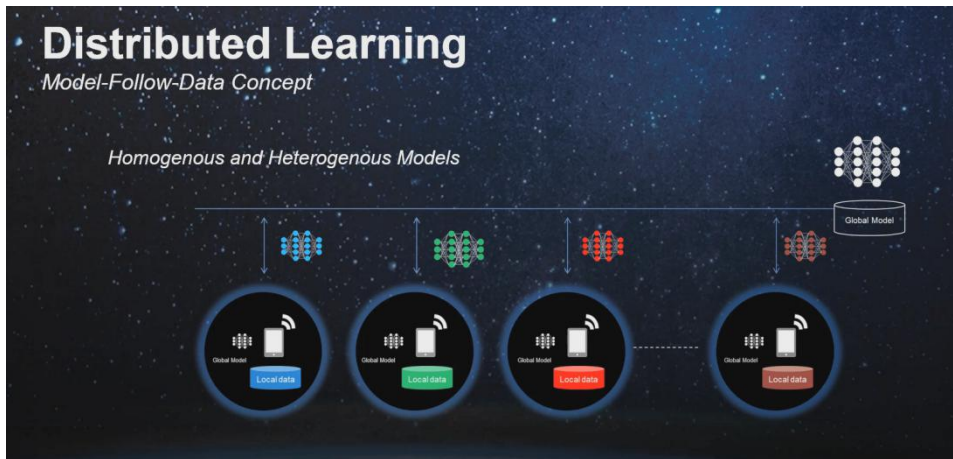


图 3 MFD分布式学习框架

MFD 分布式学习一个重要的出发点是原生支持无线网络边缘场景，充分考虑无线网络的具有特点，包括：

超动态性：终端与基站之间的无线连接受到的干扰和损耗会受到周边街道环境、其他基站的业务状态变化的影响，从而影响终端与基站的上下行连接速率；终端具有很强的移动性，可能从小区中心移动到边缘，切换到其他小区；基站会在忙时和闲时之间变换，终端在没有上下行业务时也会进入空闲状态；

超异构性：网络中的物理设备终端来自不同的供应商，导致了复杂的网络异构性。例如旗舰手机的计算能力比低端手机强；受观察空间的限制，各网络节点采集的数据在类型和分布上具有明显的异构性；另外近中远点的终端连接速率也有很大差别，还有不同的模型结构和任务类型等；

超分布性：无线网络的整体设计趋向分布式和扁平化，网络中的网元、基站和终端的在大范围内分布式部署，使得网络设备在各个区域间需要建立复杂的连接。特别是UE和网络通过空口传输，随着越来越多的业务，特别是大模型业务被迁移到移动终端上，频谱资源的稀缺性将进一步凸显。在此背景下，分布式学习算法应用到网络内，需要进

一步减少通信开销，提升传输效率；

超规模性：无线网络中同时参与分布式学习的节点规模可以非常大。这种分布式协作可以打破地域限制，实现跨区域协作。例如，部署分布式学习在不同省市的设备之间。组织网络中的大规模分布式学习需要网元、基站和终端间进行系统性管控，这对协作效率和性能是一个非常大的挑战。

MFD 分布式学习需要从以上无线网络的固有特点出发进行设计和优化，应具备以下关键特征：

模型在网络内像 IP 包一样传输：模型作为知识载体在网络设备间高效传输，而不是传输原始数据，或者频繁交互中间梯度。同时，模型在网络中传输的同时应该能感知到数据和计算资源的；

模型可以自由的缩放：模型可以自由地缩小和扩展，通过对模型进行压缩，抽象出要传递出去的核心知识；接收到模型的设备按照需要进行模型的扩展，得到可正常使用的模型；

支持异构模型间的知识分享：不同厂家设备上的算力和数据类型不同，支持的模型结构、尺寸、任务类型不一，导致在这些设备间进行智能协作比较困难，需要 MFD 分布式学习具备良好的异构支持性；

轻量化自组织的学习流程：一次分布式学习的协作范围可能非常大，甚至跨越不同覆盖地域。在传统网络中只需要进行邻区间的协作，如果要组织协调如此大范围的分布

式协作，对于网络的管控能力是相当大的挑战。因此 6G 网络内的 MFD 分布式学习更适合轻量化自行组织的方式，降低管控复杂度。

MFD 分布式学习的关键技术特征与无线网络特点之间的关系如图 4 所示，将在后续章节进行详细解释和阐述。

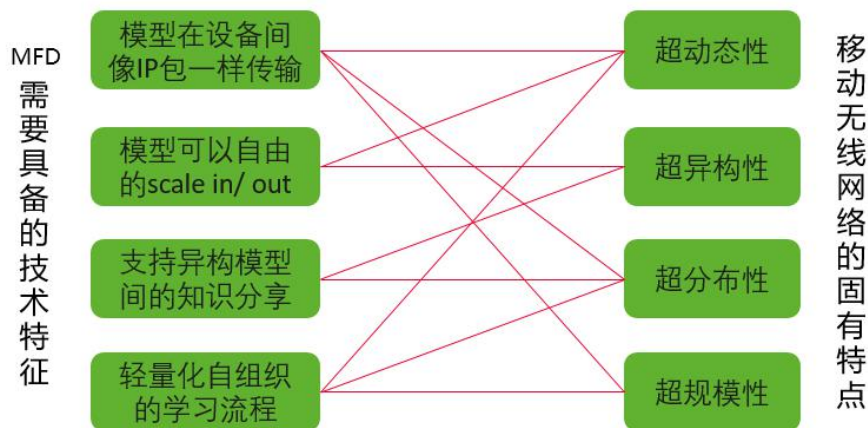


图 4 MFD 分布式学习关键技术特征与无线网络特点之间的关系

2.3 模型在网络内像 IP 包一样传输

模型经过大量数据的训练，在参数的不断调整中，总结并记忆数据中呈现出的特征和规律。因此，模型是对大量数据中携带信息的提取和压缩，相比于传输原始数据，传输模型是更加高效的信息交互方式。然而，MFD 并不希望像联邦学习一样，在网络中传输大量的临时性的模型梯度，这些梯度是损失函数相对于模型参数的导数，它们反映了模型在训练过程中对数据变化的敏感程度。利用梯度可以很容易地推导出数据的某些隐含信息，造成严重的隐私风险，即基于梯度的深度隐私泄露，知道了梯度信息基本等于知道了数据[1]。同时，这些频繁交互的模型梯度也会造成大量的通信开销。

在 MFD 分布式学习中，我们更倾向于在网络中传递模型本身，即实时的模型参数，可以通过参数共享、模型导出、模型集成和模型复制等多种形式进行传递。网络边缘设备基于本地的数据去训练模型，当模型训练达到一定程度，如满足基础的推理精度，认为该模型中已经具备了本地的数据知识。将这样的模型传递到其他边缘设备，进行知识的传递。模型在不断传递间，吸纳不同边缘设备上的数据知识，从而不断成长。

联邦学习的模型成长方式可以用如下公式描述：

$$\sum (\dots (\sum \tilde{W}_{n,i}) \sum \Delta \tilde{W}_{n,i})$$

其中， $\Delta \tilde{W}_{n,i}$ 是边缘设备 i 在第 n 轮本地迭代的模型参数梯度， $\tilde{W}_{n,i}$ 是边缘设备 i 在第 n 轮迭代的模型更新权重。

则 MFD 分布式学习的模型成长方式则是可以描述为基础模型在网络节点间传递不断学习的过程，以 LoRA (Low Rank Adaptation) 调参的方式为例：

$$(((W_0 + A_1 B_1) + A_2 B_2) + A_3 B_3) + \dots$$

其中， W_0 是初始模型参数， $A_i B_i$ 是经过某边缘设备 i 上的数据训练后更新的模型参数值。

无线网络作为连接的基础设施，可以原生支持模型以 IP 包的形式在网络中进行传输，称为 AI 数据包 (AI Packet)，AI 数据包的报头包含源地址和目的地址，网络将基于这些地址进行路由。图 5 是一个具体基于 AI 数据包传递模型的例子，其中终端 D1 首先作为目的节点，在接收到神经网络模型后基于本地数据进行模型更新；之后 D1 又作为源节点，利用更新后的模型重建数据包并将新模型传递给目的节点 D2。

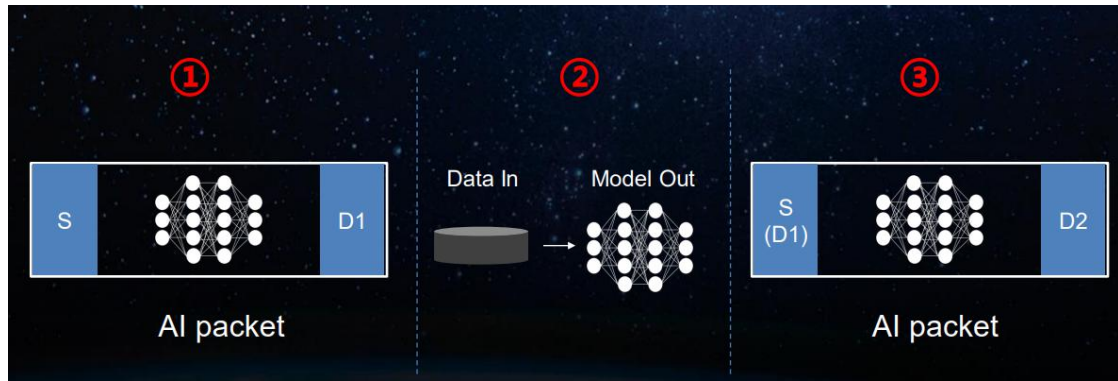


图 5 AI数据包传递流程

在以上过程中，需要设计新的AI数据包路由机制，包括AI数据包转发协议，数据兴趣值计算和路由更新算法等，支持将AI数据包路由到合适的节点以进行更新。AI数据包的路由机制类似IP，或者基于SRV6等协议拓展。

2.4 模型可以自由缩放

在 MFD 分布式学习中，模型既需要在边缘设备间传递，也需要在边缘设备上进行训练甚至推理。在模型传递时，我们希望在损失或者少损失信息量的情况下模型的尺寸尽量小，而在边缘设备上训练或者推理时，模型尺寸就需要适配本地算力、数据量和任务类型。为此，我们就需要模型可以自由的缩放。

如图 7 所示，假设在边缘设备间传递的全局模型设定为 W_0 ，边缘设备 1 上的算力和本地数据集都比较大，适配模型尺寸 W_{+1} ；而边缘设备 2 上的算力或者本地数据集比较小，适配模型尺寸 W_{-1} 。当设备 1 接收到传递过来的模型时，首先需要对 W_0 进行放大至 W_{+1} 的尺寸，然后再利用本地数据进行训练，得到更新后的 W_{+1} 模型。这个模型可以在设备 1 本地进行部署使用，其中既包含了 W_0 中传递过来的历史训练数据的知识，也针对本地数据进行了个性化。随后，设备 1 将 W_{+1} 重新压缩到 W_0 尺寸并传递到

下一个边缘设备。而对于设备 2 来说，过程正好相反，首先需要对 W_0 进行压缩至 W_{-1} 的尺寸，本地训练后再放大至 W_0 的尺寸来进行传递。

同理，假设边缘设备 3 和 4 上的模型尺寸为 W_{+2} 和 W_{-2} ，也需要与传递的全局模型间进行缩放，只是缩放的幅度比设备 1 和 2 要更大。模型压缩可能会带来信息量的损失，如何在尽量保证不损失信息量的情况下提升压缩效率也一直是业界的课题。目前比较流行的模型压缩方法包括剪枝、量化和蒸馏等。而对于模型放大，业界的研究还不是很多。其实对于模型放大，可以避免从零开始训练一个大模型，例如在 MFD 中，如果某个设备想要训练一个大一些的模型，从接收到全局模型 W_0 进行扩展然后继续训练显然比重头训练要高效。从理论上来说，大模型的拟合能力肯定大于小模型，但是具体如何扩展还是有很多问题需要研究的。蒸馏是一种直接可以想到的方法，但是需要经过一定的样本训练。更加高效的方法是仅仅对权重做一些确定性的变换，而不用通过梯度下降来训练；并能保证针对同一个输入，小模型和大模型给出的预测结果是完全一致的，或者说它们表面上看起来不一样，但数学上它们是完全一致的函数。在[2]中，作者通过“重复”、“补零”等操作实现了模型权重的自然放大。Zhang 在[3]中也用类似的方式实现了 Transformer 模型的放大。

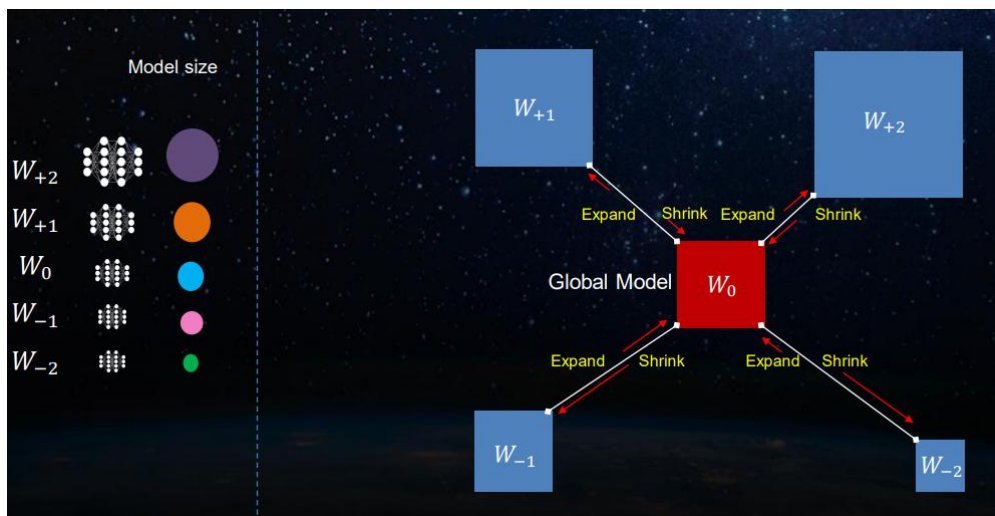


图 6 MFD分布式学习模型缩放机制

2.5 支持异构模型间的知识分享

网络中各厂家设备支持的模型多种多样，支持跨设备间的异构模型协同是 6G 原生 MFD 分布式学习必须要具备的特征。通过协同不同设备的异构模型，可以充分利用多样化的数据和计算资源，提高模型的泛化能力和鲁棒性。

具体来说，异构模型是指不同边缘设备使用的模型结构或训练方法不同，包括不同的模型结构，例如某些设备使用卷积神经网络，而另一些使用递归神经网络；不同的超参数设置，例如某些设备使用较大的尺寸，而另一些使用较小的尺寸；不同的数据处理方法，例如某些设备对数据进行了标准化处理，而另一些设备没有进行任何预处理等。模型异构性导致模型参数的集成和融合难度较大，也会出现数据偏差和计算负担和效率不均衡等问题。

实际上，在联邦学习中已经有大量工作进行这方面的研究，例如联邦蒸馏[4]，通过传递模型在公共数据集上的 logits 来避免直接聚合模型。在 MFD 分布式学习模式下，

同样可以参考知识蒸馏的方式来进行异构模型间的聚合。如下图所示，设备 1 在进行模型训练时希望集成设备 2 上的本地知识，首先将本地蓝色模型发送给设备 2，设备 2 使用知识蒸馏技术，以本地红色模型为 teacher，蓝色模型为 student，对本地数据进行蒸馏训练，并生成紫色模型。这个紫色模型即是设备 1 的蓝色模型融合设备 2 上的本地知识后的更新模型，此后再返回给设备 1，可以基于新的数据集进行持续更新。至此，在不传递原始数据的情况下完成了设备 2 向设备 1 的异构模型知识分享。同理，设备 2 也可以通过蒸馏的方法获得设备 1 的模型知识分享。

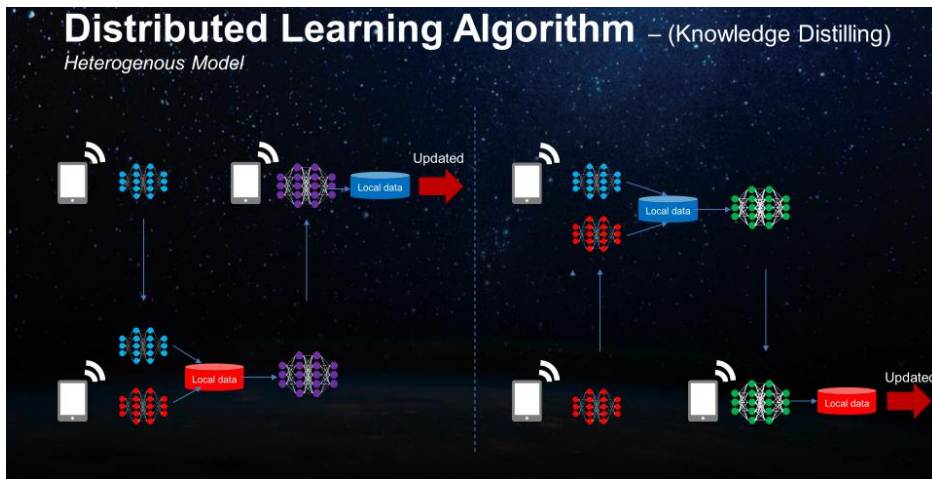


图 7 MFD分布式学习算法——知识蒸馏

另一种可行的方式是通过构建生成模型来进行知识分享，如下图所示。设备 1 基于本地模型和数据训练出一个生成模型(黄色)，该生成模型能够生成 AI 合成数据。将该生成模型发送给设备 2，设备 2 利用生成模型生成的数据集和本地数据集一起训练本地蓝色模型。至此，在不传递原始数据的情况下完成了设备 1 向设备 2 异构模型的知识分享。同理，设备 2 也可以构建自己的生成模型然后传递给设备 1 来完成知识的共享。

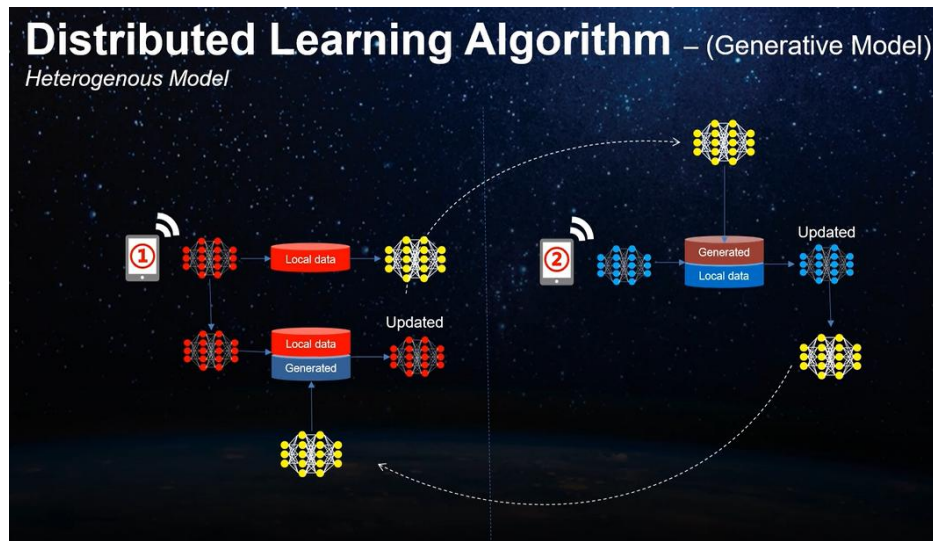


图 8 MFD分布式学习算法——知识蒸馏

上述方法都有很好的可伸缩性，但是对于模型参数特别大的基础模型，需要进一步优化，例如通过融合LoRA微调机制从而只交换少量参数。设备各自生成本地的LoRA参数，然后彼此分享，每个设备聚合本地和其他设备的LoRA参数生成新的LoRA模型。

2.6 特轻量化自组织的学习流程

网络中的分布式学习的协作范围可以非常大，甚至跨越不同覆盖地域。在传统网络中只需要进行邻区间的协作，如果要组织协调如此大范围的分布式协作，对于网络的管控能力是相当大的挑战。另外，在传统分布式学习中，往往考虑静态的协作集，协作对象间是相对稳定的关系，从而获得稳定的协作增益。而网络环境动态变化性强，可能出现终端切换到其他站点甚至跨出覆盖区域。因此 MFD 需要具备轻量化自组织特性，降低管控复杂度。

首先是模型在数据节点间的传输路径选择上，如果需要提前对模型完整的训练和交互流程进行管控，需要网络提前收集所有涉及的数据节点的本地信息进行决策，这样的多目标优化问题求解的复杂度随着数据节点数量的增加急剧上升。另外，随着网络状态

的变化如终端的移动，决策时优化效果往往很难达成。在 MFD 分布式学习模式下，模型在网络设备间进行传输时，应该能够自行的根据实时信息决定下一跳。例如计算周边节点与当前节点的数据分布间的 KL 散度来评估数据特征的相似性，选取 KL 距离短的节点作为下一跳；也可以评估周边节点的算力和传输时延，选取算力强时延短的节点作为下一跳；

其次是各网络设备能够根据自身连接、数据、算力等状态自适应更新模型参数。随着环境干扰等信道环境的变化，或者终端本身的移动，网络与终端的无线传输速率也在不停变化，影响模型的传输时延；同时网络设备的当前可用算力以及收集到的数据也是动态变化的，为此 MFD 允许模型在每个数据节点自行决策更新模型的全部或部分参数。其中的关键技术挑战是待训练的模型参数如何选择。例如可以通过计算信息熵来优先选择与本地数据最相关的参数，或者优先选择未被更新的参数，既可以让模型尽快得到完整遍历，又可以减少对之前训练结果的影响。

最后，MFD 分布式学习在具体部署应用过程中，还需要考虑模型管理上的安全可靠。MFD 分布式学习会形成大量模型版本，包括长期共享的基础模型，各节点根据自身需要更新后的模型等，需要进行完备的模型版本管理，并建立模型分布式账本系统，确保在任何一轮训练中都进行模型版本的验证和审计。

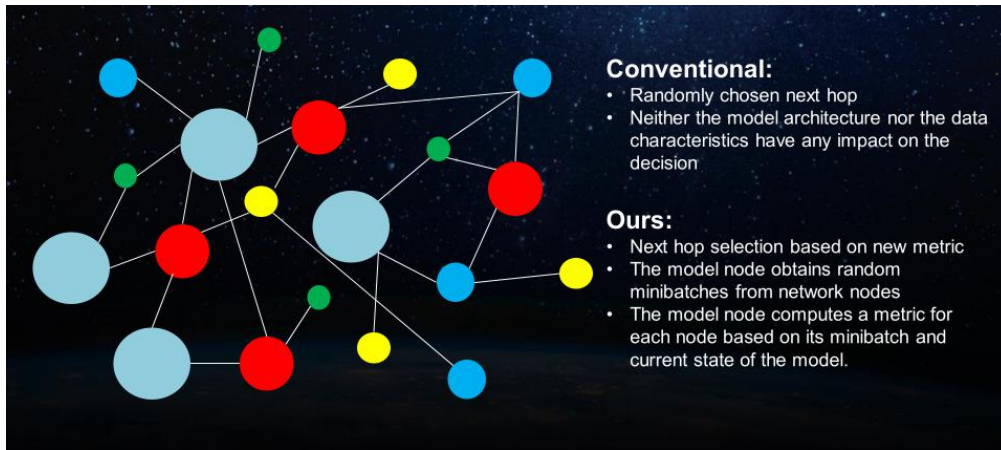


图 9 数据和计算感知模型路由

3 MFD 应用示例

3.1 Net4LM 的智能协作机器人

随着 AI 大模型和机器人关键技术的持续突破和飞速发展，基于大模型控制的智能协作机器人 (Cobots) 广受关注，预期在工业、医疗、服务等领域得到广泛应用。虽然大模型本身具备很强的复杂任务处理能力，可以显著提升机器人的感知、决策和执行能力；然而，由于未来机器人可能被用于处理各类动态、未知的任务，并且在物理世界发生的事件很难完全被大模型预先学习，因此，每个机器人就需要基于自身完成的任务，以及在完成任务过程中收集的本地数据进行持续学习，优化自身模型，而多个机器人之间采取分布式学习，共享知识将被认为是一种高效的提升机器人智能水平的手段。

通过 MFD 分布式学习机制，将可以有效优化机器人之间的通信开销，并解决数据安全隐私、模型所有权等挑战问题，支持跨不同品牌机器人之间的自组织学习，实现更广泛的机器人之间的知识共享。

假设一个社区有多种类型/品牌的服务机器人,共同自主完成社区所需的各种各样的工作,每个机器人基于各自工作学习新的知识,并通过 MFD 共享,具体包括:

- 1) **MFD 分布式学习任务的建立:** 机器人间的知识分享需要在一定范围的协作集内进行,且各厂家机器人之间的角色是平等的,不存在 host。通过 6G 网络为 MFD 分布式学习任务建立协作集,例如某个基站覆盖下的社区范围组成一个协作集,各机器人可以通过 6G 网络向协作范围的其他机器人发送和接收模型,模型在协作集内自由流动,不断更新。
- 2) **AI 模型路由:** 模型在社区机器人之间路由,选择合适的机器人来进行模型的更新。各机器人根据本地任务模型和数据计算并发出自身的数据兴趣值,在 6G 网络中建立 AI 数据包的路由表。模型根据路由表中的兴趣值路由到对应的机器人处进行更新。
- 3) **本地模型更新:** 机器人收取到 AI 数据包并解析出外来模型后,根据需要进行模型的缩放处理,适配本地的任务、算力和数据规模。然后该机器人将结合外来模型和本地数据,进行本地模型的更新,为模型融入新的知识。更新后的模型可以再作为 AI 数据包发送出去,继续路由到其他机器人。
- 4) **模型安全可靠:** 为了保障模型学习过程中的安全可靠,每个机器人分享的新模型都需要进行标记,通过 6G 网络建立对应的分布式账本,支持可验证和追溯,防止出现恶意的信息投毒;

3.2 LM4Net 站间协同

AI 技术正在被不断的引入网络中用于优化网络自身的性能。传统 5G 基站的很多 AI 算法还是基于逐个模型设计的方式来进行部署、训练和应用，例如专门为信道预测、编码调制、干扰消除、资源调度等细分任务设计的专家模型。随着大模型技术的发展，越来越多的研究集中在通过构建无线大模型来进行基站上多任务的推理和决策。基于无线大模型底座，可以通过 few-shot 或者 fine-tuning 的方式来进行下游任务的快速适配。

无线大模型在部署后也需要随着基站部署情况、业务流量等的变化进行更新。然而需要注意的是，基站上的大量数据，例如 SRS 和 CSI 的信道测量数据、干扰信号强度、SINR 下的解调数据以及空口资源的调度数据等，都无法传输到云上进行集中处理。通过 MFD 模式进行站间协同，实现更大范围的知识分享，是更适合的方式。具体包括：

- 1) **MFD 分布式学习任务的建立**：基站之间由 X2 口互相连接，模型可以通过承载网进行连通范围内的传输和共享。且各基站之间的角色是平等的，不存在 host。通过 6G 网络为 MFD 分布式学习任务建立协作集，例如某个核心网区域下的基站组成一个协作集，各基站可以通过 X2 口向协作范围的其他基站发送和接收模型，实现站间协同。
- 2) **AI 模型路由**：模型在基站之间路由，选择合适的基站来进行模型的更新。各基站根据本地任务模型和数据计算并发出自身的数据兴趣值，在无线网络相关节点上建立 AI 数据包的路由表。模型根据路由表中的兴趣值路由到对应的基站处进行更新。

- 3) **本地模型更新**：基站收取到 AI 数据包并解析出外来模型后，根据需要进行模型的缩放处理，适配本地的任务、算力和数据规模。然后该基站将结合外来模型和本地数据，进行本地模型的更新，为模型融入新的知识。更新后的模型可以再作为 AI 数据包发送出去，继续路由到基站。

4 总结和展望

当前 3GPP 已经在进行无线网络分布式学习的标准化研究工作，包括 R18 的横向联邦以及 R19 的纵向联邦等，这些标准化研究目前局限在 NWDAF 之间，并较少考虑大模型的影响。而随着大模型技术的快速发展，引起了业界的广泛关注，认为将为 6G 网络设计带来全新的路线和驱动力。面对未来大模型在端边云无处不在的部署，6G 网络需要原生的大模型分布式学习模式，MFD 正是在这个背景下提出。MFD 的设计理念是模型跟随数据，并让模型可以如 IP 数据包一样传输，使得知识可以在端边云高效的分享。MFD 包含算法设计和网络架构设计两个方面，一是需要原生设计符合 MFD 特征的高效可用的分布式算法，二是需要原生设计对应的网络架构，包括功能、流程和接口，支持各类 MFD 算法的部署运行。

本文重点描述了未来 6G 时代的智能普惠的相关断言，分析了联邦学习模式的挑战，定义了 MFD 分布式学习模式、其主要关键特征和应用示例。对于具体符合 MFD 模式的分布式算法和网络架构设计，将在后续系列文章中进行具体阐述，同时，也期待产业和学术界的专家共同研究探索，形成更多共识和研究成果。

5 参考文献

- [1]. Zhu, L., Liu, Z., Han, S.: Deep leakage from gradients. In: Annual Conference on Neural Information Processing Systems (NeurIPS) (2019)
- [2]. Jianlin Su. (Jun. 02, 2021). Can we magnify a Transformer model without loss? [Blog post]. Retrieved from <https://spaces.ac.cn/archives/8444>
- [3]. Zhang H. Transfer training from smaller language model[J]. arXiv preprint arXiv:2104.11390, 2021.
- [4]. Li L, Gou J, Yu B, et al. Federated Distillation: A Survey[J]. arXiv preprint arXiv:2404.08564, 2024.
- [5]. Yuan L, Wang Z, Sun L, et al. Decentralized federated learning: A survey and perspective[J]. IEEE Internet of Things Journal, 2024.
- [6]. A. Lalitha, S. Shekhar, T. Javidi, and F. Koushanfar, “Fully decentralized federated learning,” in Third workshop on Bayesian Deep Learning (NeurIPS), 2018.
- [7]. D. Monschein, J. A. P. Perez, T. Piotrowski, Z. Nocht, O. P. Waldhorst, ´ and C. Zircins, “Towards a peer-to-peer federated machine learning environment for continuous authentication,” in 2021 IEEE Symposium on Computers and Communications (ISCC). IEEE, 2021, pp. 1 – 6.
- [8]. V. P. Chellapandi, A. Upadhyay, A. Hashemi, and S. H. Zak, “On the convergence of decentralized federated learning under imperfect information sharing,” IEEE Control Systems Letters, 2023.
- [9]. J. Konecny, H. B. McMahan, F. X. Yu, P. Richtarik, A. T. Suresh, and D. Bacon, “Federated learning: Strategies for improving communication efficiency,” 2016. [Online]. Available: arXiv:1610.05492.
- [10]. Wang L, Wang W, Li B. CMFL: Mitigating communication overhead for federated learning[C]. In Proceedings of the IEEE International Conference on Distributed Computing Systems, 2019: 954–964.
- [11]. McMahan B, Moore E, Ramage D, Hampson S, Arcas B A Y. Communication-efficient learning of deep networks from decentralized data[C]. In

- Proceedings of the International Conference on Artificial Intelligence and Statistics, 2017: 1273–1282.
- [12]. Chen M, Poor H V, Saad W, Cui S. Convergence time minimization of federated learning over wireless networks[C]. In Proceedings of the IEEE International Conference on Communications, 2020: 1–6.
- [13]. Lin T, Kong L, Stich S U, et al. Ensemble distillation for robust model fusion in federated learning[J]. Advances in neural information processing systems, 2020, 33: 2351–2363.