Fed-GAME：用于时间序列预测的图注意力专家混合个性化联邦学习

Yi Li[1], *Han Liu*[2], *Mingfeng Fan*[3],*Guo Chen*[4],*Chaojie Li*[5*], *Biplab Sikdar*[2]

[1] 中南大学自动化学院[2]新加坡国立大学电气与计算机工程系[3]新加坡国立大学机械工程系[4] 新南威尔士大学电气工程与电信学院[5] 香港城市大学电气工程系

arXiv:2603.01363v1 [cs.LG] 2 Mar 2026

摘要

图上的联邦学习（FL）在分布式时间序列预测方面展现出良好前景。然而，现有方法依赖静态拓扑，难以应对客户端异质性。我们提出 Fed-GAME，这是一种将个性化聚合建模为可学习动态隐式图上消息传递的框架。其核心是一种解耦的、基于参数差异的更新协议：客户端传输其微调后的私有模型与共享全局模型之间的参数差异。在服务器端，这些差异被分解为两个流：（1）用于更新全局模型以达成共识的平均差异；（2）输入新型图注意力专家混合（GAME）聚合器的选择性差异，以实现细粒度个性化。在该聚合器中，共享专家提供评分信号，而个性化门控则自适应地为选择性更新赋权，从而支持个性化聚合。在两个真实世界电动汽车充电数据集上的实验表明，Fed-GAME 的表现优于最先进的个性化联邦学习基线。

索引词— 个性化联邦学习，时间序列预测，隐式图注意力，专家混合

1. 引言

联邦学习（FL）支持在去中心化数据上进行协同训练 [1], ，但其性能往往会因统计异质性而下降。在分布式时间序列预测任务中，这一问题尤为严重，例如电动汽车充电需求预测，因为各个客户端呈现出不同的时间模式和季节性。个性化联邦学习（PFL）通过近端正则化、聚类或多任务学习等技术，根据局部特征定制模型，以应对这一挑战 [2, 3, 4, 5, 6]。然而，许多此类方法侧重于局部适配（例如延长本地训练或采用个性化正则化）来提升客户端性能，却忽视了联邦学习（FL）的一个核心方面：跨客户端的知识聚合。由于聚合会直接影响异质数据下个性化的质量，近期研究提出了联邦图学习（FGL），使客户端能够与相似的对等客户端进行更有选择性的协作，而非依赖统一聚合。 [7, 8, 9, 10]。例如， [7] 设计了双层优化方案，使用两类目标函数同时训练本地模型和图神经网络模型。 [8] 利用地理邻近性构建客户端图，以指导模型聚合过程。然而，现有的FGL及其他先进的个性化联邦学习方法面临两个根本性挑战。

(1) 不精确客户端图：现有 FGL 方法通常依赖预定义的静态图（例如，基于地理位置构建的图） [10, 11], ，这些图可能无法捕捉客户端之间真实的任务特定关系。一些研究尝试从完整模型更新中推断动态图 [12], ，但这需要客户端传输其完整模型，而且生成的图通常基于粗粒度的全局相似性度量，而非细粒度的个性化度量。

(2) 混合更新的次优聚合：在传统的个性化联邦学习框架中，每个客户端只交换一个更新向量，这会将全局共识与客户端特定个性化混为一谈 [13]。例如，FedAvg [1]只是对更新取平均，这会削弱通用知识并模糊个性化信号。由此得到的全局模型可能无法有效服务不同客户端。近期的参数分解技术 [14] 通过采用参数加性分解来解决这一问题：提取全局参数以实现知识共享，同时保留本地参数用于个性化。然而，这种分解忽视了客户端间丰富的关系，而这些关系对于实现真正有效的个性化至关重要。

为应对这些挑战，我们提出了用于时间序列预测的全新个性化联邦学习框架——图注意力专家混合（Fed-GAME）。我们的贡献如下：（1）我们通过仅传输每个客户端微调后的私有模型

*通讯作者：Chaojie Li（chaojili@cityu.edu.hk）

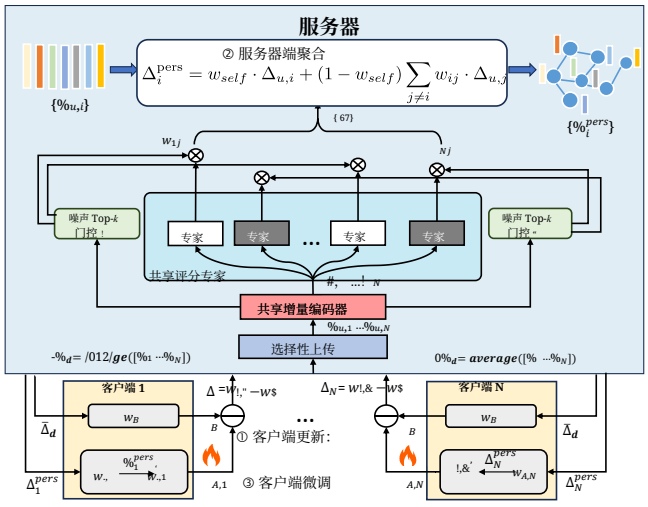


图 1。Fed-GAME 的框架

(M_A)与共享全局模型(M_B)之间的参数差异，将全局共识与个性化分离开来。这种设计能够实现细粒度的个性化更新，同时维持稳健的全局共识。（2）我们提出了一种服务器侧 GAME 聚合器，从传输的更新中学习动态的客户端特定聚合策略，使 GAME 能够以灵活、数据驱动的方式捕捉特定任务之间的关系。（3）我们设计了一种基于相似度的元损失，用于在服务器上训练聚合器，而无需直接访问客户端数据。

2. Fed-GAME 框架

2.1. 问题表述

本文旨在为联邦环境中的 N客户端实现个性化时间序列预测。每个客户端 $i \in \{1, \cdots, N\}$都拥有一个本地私有数据集 $D_i = \{(\mathbf{X}_i, \mathbf{Y}_i)\}$，其中 $\mathbf{X}_i = \{X_i^{t+1}, \cdots, X_i^{t+h}\}$ 表示长度为 h 的历史序列，$\mathbf{Y}_i = \{y_i^{t+h+1}, \cdots, y_i^{t+h+p}\}$ 表示待预测的对应未来 p 步序列。第 i 个客户端的 p 步预测序列记为 $\hat{\mathbf{Y}}_i^p = \{\hat{y}_i^{t+h+1}, \cdots, \hat{y}_i^{t+h+p}\}$。由于数据异质性，Fed-GAME 不再学习单一的全局模型，而是旨在为每个客户端学习一组独特的模型参数 $\boldsymbol{w}_i$。其目标是最小化所有客户端上的聚合损失：$\min_{\boldsymbol{w}} \sum_{i \in N} \lambda_i f_i(\boldsymbol{w}; \mathcal{D}_i)$，其中 N 表示客户端数量，$f_i(\cdot)$ 表示客户端 i的预测损失，λ_i 表示数据集权重，$\boldsymbol{w}$ 表示待优化的模型参数。

2.2. Fed-Game框架

为实现个性化时间序列预测，我们提出了 Fed-GAME。Fed-GAME 包含以下步骤，为清晰起见，各步骤均以图 1 中的阶段编号标注：

（1）客户端更新：每个客户端 i 都维护自己的持久化私有模型（M_A），其参数为 $w_{A,i}^{\tau}$，以及一个共享全局模型（M_B）的本地副本，其参数为w_B^{τ}，对应第 τ 轮次。客户端在第 τ 轮次使用本地数据对 M_A 进行微调后，i 会计算其参数相对于全局模型的差值：$\Delta_i^{\tau} = w_{A,i}^{\tau} - w_B^{\tau}$。随后，客户端将$\Delta_i^{\tau}$上传至服务器。

(2) 服务器端聚合：在服务器上，这些差异有两个用途：i. 全局共识更新：为维护共享知识库，服务器对客户端 Δ_i^{τ} 进行平均，并将结果应用于全局模型：$w_B^{\tau+1} = w_B^{\tau} + \eta \cdot \frac{1}{N} \sum_i \Delta_i^{\tau}$，。其中，$\eta$ 是控制全局更新步长的服务器学习率。ii. 个性化聚合：为提高通信效率，仅使用 Δ_i^{τ} 的选择性组件（记为 $\Delta_{u,i}^{\tau}$）进行个性化聚合。这里，我们限定只选择最后几层 MLP 层，因为普遍认为它们能够捕获高层次的可迁移特征，而较低层则仍具有客户端特定性。随后，每个客户端的选择性差异 $\Delta_{u,i}^{\tau}$ 由共享编码器（由线性层构建）编码为低维嵌入 $\mathbf{e}_i^{\tau} = \text{编码器}(\Delta_{u,i}^{\tau})$，以稳定训练并降低维度，从而高效计算注意力。接着，这些嵌入被输入 GAME聚合器，其中共享专家用于建模客户端间相关性，个性化门控则对其进行自适应加权。所得分数 w_{ij}^{τ} 作为 M_A 更新的聚合权重，从而得到个性化更新 $\Delta^{\text{pers},\tau}_i$：

$$\Delta_i^{\text{pers},\tau} = w_{self} \cdot \Delta_{u,i}^{\tau} + (1 - w_{self}) \sum_{j \neq i} w_{ij}^{\tau} \cdot \Delta_{u,j}^{\tau}, \quad (1)$$

其中，w_{self} 控制客户端本地更新与来自其同伴的聚合知识之间的平衡。这对应于客户端交互图上的基于图的消息传递，其中边编码最终层更新之间的相似度。图注意力权重 $\{w_{ij}^{\tau}\}$的计算详见第 2.3 节。

(3) 客户端微调：每个客户端应用个性化的 $\Delta^{pers,\tau}_i$ 并在本地数据上进行微调：

$$w_{A,i}^{\tau+1} = \text{FineTune}(w_{A,i}^{\tau'}) = \text{FineTune}(w_{A,i}^{\tau} + \gamma \cdot \Delta_i^{\text{pers},\tau}), \quad (2)$$

γ 表示控制个性化更新步长的学习率。微调后，计算下一轮的 $\Delta_i^{\tau+1}$，即 $\Delta_i^{\tau+1} = w_{A,i}^{\tau+1} - w_B^{\tau+1}$，以便在下一轮上传。

2.3. 用于聚合权重的 GAME 聚合器

我们服务器端个性化的核心是一个可学习的混合专家（MoE）聚合器。在每一轮（为清晰起见，省略上标 τ）中，我们构建一个全连接隐式图，其中每个上传的更新 $\Delta_{u,i}$ 都被视为一个节点。其低维嵌入 $\mathbf{e}_i$ 充当动态节点特征，而原始更新 $\Delta_{u,i}$ 则作为消息进行传播。不同于采用固定邻接矩阵的方法，我们的聚合器学习个性化图注意力系数 $\{w_{ij}\}_{j \neq i}$，从而自适应地对消息进行加权

（每个客户端的原始更新 $\Delta_{u,\ i}$）。这种个性化注意力机制通过两个组件实现：共享评分专家，用于捕获客户端间相关性；以及个性化门控模块，用于为每个客户端自适应地对专家输出进行加权。

(1) 共享评分专家：我们采用 M 共享评分专家 $\{E_k\}_{k=1}^{M}$（由线性网络构建）。这些专家的作用是评估邻居节点*j*向目标客户端 *i*发送的消息的相关性。标量评分 s_{ijk} 计算如下： $s_{ijk} = E_k([\boldsymbol{e}_j, \boldsymbol{e}_i])$。该评分表示特定专家对每一对客户端之间相关性的评估。

(2) 个性化噪声 Top-k 门控：每个客户端 i都配备一个个性化门控 G_i，用于分配各专家的组合权重。门控生成逻辑值$H_i(e_i) = (e_i \cdot w_{gi}) + \text{SN}(\text{softplus}((e_i \cdot w_{\text{noise}i})))$，其中$w_{gi}$ 和 $w_{\text{noise}i}$ 是可学习参数，SN 添加高斯噪声以进行探索，而 $\text{softplus}(z) = \log(1 + e^z)$。应用 Top-$k$ 选择后再进行 softmax，得到稀疏权重向量：$c_i = G_i(e_i) = \text{Softmax}(\text{Top-}k(H_i(e_i), k))$，将专家评分合成为单一相关性值：$v_{ij} = c_i^\top s_{ij}$，$s_{ij} = [s_{ij1}, \ldots, s_{ijM}]$。随后通过温度缩放 softmax 得到图注意力系数：$w_{ij} = \exp(v_{ij}/T) / \sum_{l \neq i} \exp(v_{il}/T)$，其中 T为温度。因此， w_{ij} 定义了每一轮的动态客户端中心图的边权重，并通过公式 (1) 中的加权消息传递计算； w_{ij} 也可解释为个性化权重。详情见 [15]。

备注1 与标准图注意力聚合器（例如 *GAT*）相比，*Fed-GAME* 摆脱了对静态拓扑的依赖，并采用混合专家范式，将单一注意力函数分解为多个共享专家和个性化路由器，从而实现多专家关系建模和客户端特定聚合策略。

2.4. 训练目标

Fed-GAME框架采用解耦的两级优化策略。**客户端侧**：每个客户端 i 通过最小化本地目标函数来优 M_A 化其私有模型，该目标函数结合了其本地数据上的预测损失与 FedProx [2]中的近端项，以规范本地训练并减轻相对于全局模型 M_B的漂移。客户端 i 的本地训练可表述为：$\min_{w_A} \left(\mathcal{L}^i_{\text{task}}(w_A) + \frac{\mu}{2} |w_A - w_B|_2^2 \right)$。**服务器侧**：服务器使用基于相似度的元损失 L_{meta}训练其可学习聚合模块（编码器和混合专家模型），该损失定义为均方误差（L2）损失与余弦不相似度损失的加权组合：

$$\mathcal{L}^i_{\text{meta}}(w_{\text{enc}}, w_{\text{moe}}) = \alpha |\Delta_i^{pers} - \Delta_{u,i}|_2^2 + \beta \left(1 - \frac{\Delta_i^{pers} \Delta_{u,i}}{|\Delta_i^{pers}||\Delta_{u,i}|}\right) \quad (3)$$

表 1. 评估指标分位数得分 $= \begin{cases} \frac{1}{n}\sum_{i=1}^{n}(1-q) & \\ \|y_i^t - \hat{y}_{i,q}^t\|, & y_i^t < \hat{y}_{i,} \\ \frac{1}{n}\sum_{i=1}^{n} q\|y_i^t - y_{i,q}^t\| & \begin{cases} 1, & y_i^t \geq y_{i,q}^t \\ \hat{y}_{i,q} \leq y_i^t \leq \hat{y}_{i,0} & \end{cases} \end{cases}$

ICP $= \frac{1}{n}\sum_{i=1}^{n}$ ，否则

MIL $= \frac{1}{n}\sum_{i=1}^{n} \|\hat{y}_{i,\ q}^t - \hat{y}_{i,\ q'}^t\|$

其中， α 和 β 是平衡超参数， w_{enc} 和 w_{moe}分别表示编码器和混合专家模型的参数。该元损失促使混合专家模型生成与客户端自身原始参数差异相一致的个性化更新，从而在保持全局一致性的同时提升个性化效果。

2.5. 通信成本分析

通信开销是联邦学习中的关键瓶颈。因此，我们分析Fed-GAME每轮次的通信开销，并将其与交换完整模型参数的标准联邦学习基线（如FedAvg和FedProx）进行比较。设 θ 为模型的参数总数。对于基线，上传和下载均涉及全部参数，总通信开销为 $C_{baseline} = 2N|\theta|$。对于Fed-GAME，设 r 为被选中用于通信的参数比例 $0 < r < 1$。总上行开销为 $N \cdot |\theta|$。对于下行，服务器发送全局参数差异（Δ_d）和个性化参数差异（Δ_i^{pers}）。总通信开销为$C_{\text{Fed-GAME}} = 2\overline{N}|\theta| + N \cdot r|\theta| = (2 + r) N|\theta|$。通过将其与基线进行比较，我们得出通信成本比：通信成本比 $= \frac{C_{\text{Fed-GAME}}}{C_{\text{Baseline}}} \approx \frac{(2+r)N|\theta|}{2N|\theta|} = 1 + \frac{r}{2}$。个性化头很小（$r < 1\%$），因此开销可忽略不计，同时能够实现更强的服务器端混合专家聚合。

3. 实验结果

我们在两个真实世界的数据集上开展案例研究：帕洛阿尔托 [16] 和深圳 [17]，，它们分别呈现小规模和大规模的非独立同分布时空需求模式。帕洛阿尔托数据集包含 8 个充电站，需求（kWh）按 5 分钟间隔记录；深圳数据集则包含 247 个站点，时间分辨率为 30 分钟。

我们将 Fed-GAME 与最先进方法进行比较，包括 FedAvg[4], FedProx [2], pFedMe [5], PAG-FedAVG [18] 和 GCRN- FedAvg [19]，，其中 FedProx 和 pFedMe 属于个性化联邦学习，而 PAG-FedAVG 和 GCRN-FedAVG 属于 FGL。我们还纳入无联邦学习，即在每个客户端本地训练预测模型，不进行任何通信。对于充电需求的不确定性，我们采用分位数回归进行时间序列预测 [20]。模型性能通过三项标准指标进行评估：分位数得分（QS）、平均区间长度（MIL）和区间覆盖率（ICP），结果汇总于表 1。对于帕洛阿尔托数据集，我们使用过去 3 小时的数据，预测未来 30 分钟（预测步长 6）和 60 分钟（预测步长 12）的充电需求。对于深圳，使用过去 6 小时的数据来预测 3 小时（预测步长 6 ）和 6

小时（预测步长 12）的充电需求。原则上，可以使用任何预测模型。然而，为体现实际设备限制，我们采用两层 LSTM+MLP 模型，LSTM 隐藏层维度分别为 256 和 128。对于 Fed-GAME，学习率设为 0.0005，$w_{self} = 0.6$,$\mu = 0.2$, $\alpha = \beta = 0.5$。采用四个共享评分专家，并使用 Top-$k = 2$。超参数通过初步实验确定。为了评估概率预测 $q = \{0.1, 0.5, 0.9\}$，ICP 应接近 0.8，而 QS 和 MIL 应尽可能小。所有实验均使用 Python 3.10 在 NVIDIA A100 GPU 上实现。

3.1. 对比分析

帕洛阿尔托和深圳的预测性能结果如表 2 所示。在帕洛阿尔托数据集上，Fed-GAME 的平均分位数得分和 ICP 得分均为最佳。由于我们上传的是参数差值，而非完整参数，因此本方法的 MIL 性能并非最佳；但在分位数得分和 ICP 得分方面仍优于基线方法。这证实了 Fed-GAME 的聚合策略能够在不直接共享数据的情况下有效利用客户端知识。在包含 200 个站点（其中 47 个在训练期间未见过）的深圳数据集上，与无联邦学习相比，Fed-GAME 将第6步的性能提升了 54.23%（分位数得分）和 43.87%（ICP）。对于第12步，性能分别提升了 45.14% 和 68.93%，表明其在非独立同分布和部分客户端参与的设置下具有很强的泛化能力。这些结果凸显了其在长期预测方面的优势，并证实 Fed-GAME 在不同部署场景中均具备出色的准确性和泛化能力。

3.2. GAME聚合器的性能

我们将 GAME 聚合器与两种强大的基于图的基线方法进行比较，以评估其有效性：GraphSAGE [21], 采用固定的聚合方式，而 GAT [22] 则学习节点间的注意力权重。在这项消融实验中，仅替换了服务器端的聚合模块。结果如表 3 所示。实验结果证明了 GAME 聚合器的优越性。在两个数据集上，Fed-GAME 在 QS 和 MIL 指标上均显著优于 GraphSAGE 和 GAT。这一性能展现了 GAME 的有效性。GAT 计算节点间的单一注意力分数，而我们的 GAME 聚合器则利用一组评分专家，从多个角度评估客户端之间的关系。个性化的 noisy Top-k 路由器会学习一种稀疏、自适应的策略，为每个客户端整合这些专家的判断。借助这种方法，GAME 能够学习有效的内容感知聚合函数，从而提升预测准确率。

3.3. 通信效率评估

我们将 Fed-GAME 的通信开销与基线方法（FedAvg、FedProx 和 pFedMe）进行比较。对于 LSTM 模型（$\approx$ $996K$ 参数），在 6步预测（帕洛阿尔托/深圳）任务中，线性头 $\approx$ $996K$ 为 322 个参数，选用于通信的参数比例为 $r = \frac{|\theta_{\text{linear}}|}{|\theta|} = \frac{996,013}{2322} \approx 0.0023$，因此开销仅为 0.115%。在 12步预测（帕洛阿尔托/深圳）任务中，线性头为= 4644 个参数，$r = \frac{4,644}{994,852} \approx 0.0047$，使开销增加了 0.235 %。因此，与全参数基线相比，Fed-GAME 仅以极低的成本实现了个性化。

表2. 帕洛阿尔托和深圳数据集上不同联邦学习预测模型的性能比较

	方法	聚合方法	帕洛阿尔托 QS	帕洛阿尔托 MIL	帕洛阿尔托 ICP	深圳 QS	深圳 MIL	深圳 ICP
步骤6	无联邦学习	—	1.5228	4.7110	0.7585	18.6698	39.8802	0.6988
	LSTM-FedAvg	平均值	1.4846	4.5517	0.7390	10.4788	10.1150	0.7381
	FedProx	平均值	1.5378	4.7564	0.6571	9.8572	**6.3500**	0.6587
	pFedMe	平均值	1.7047	5.0307	0.6677	10.5375	9.5557	0.7176
	PAG-FedAvg	平均值	1.5304	**4.2462**	0.7534	13.7907	13.0350	0.4305
	GCRN-FedAvg	平均值	1.4219	4.2518	0.7448	9.9119	9.7647	0.7002
	Fed-GAME	GAME	**1.4067**	4.4833	**0.7827**	**8.5461**	9.8635	**0.7432**
第12步	无联邦学习	—	2.4770	7.0310	0.6726	19.1557	35.0342	0.6310
	LSTM-FedAvg	FedAvg	1.8118	5.4041	0.6633	18.4644	14.2590	0.7351
	FedProx	FedAvg	2.1077	5.4319	0.6662	11.8104	12.4952	0.7260
	pFedMe	FedAvg	2.1397	5.0307	0.6963	12.7760	13.5238	0.7077
	PAG-FedAvg	FedAvg	1.8843	5.0488	0.6866	12.7842	**13.3762**	0.6907
	GCRN-FedAvg	FedAvg	1.6297	5.1615	0.7277	9.3349	17.9791	0.7431
	Fed-GAME	GAME	**1.6270**	**5.0027**	**0.7569**	**10.5097**	15.5639	**0.7475**

表 3. 不同数据集上采用不同聚合机制的 QS、MIL 和 ICP（预测步长=12）

	方法	聚合方法	QS	MIL	ICP
帕洛阿尔托	无联邦学习	—	2.477	7.031	0.6726
	Fed-GraphSAGE	GraphSAGE	1.8878	5.6907	0.756
	Fed-GAT	GAT	1.9081	5.6808	0.7462
	Fed-GAME	GAME	**1.627**	**5.0027**	**0.7569**
深圳	无联邦学习	—	19.1557	35.0342	0.631
	Fed-GraphSAGE	GraphSAGE	11.9333	19.5901	**0.756**
	Fed-GAT	GAT	12.0329	19.3822	0.7392
	Fed-GAME	GAME	**10.5097**	**15.5639**	0.7475

3.4. 学习到的聚合权重分析

我们进一步分析了学习到的聚合权重，以更充分地验证所提出的机制。结果表明，个性化权重 w_{ij} 并非仅由地理邻近性决定；相反，客户端在不同时间尺度上呈现出各自的专长，从而形成结构化的注意力模式。随着训练的推进，聚合权重从近均匀分布逐渐演变为差异更明显的分布，其方差更高、熵更低，这表明模型能够在非独立同分布数据分布下捕捉非平凡的客户端间相关性，并使深圳数据集上的 QS 提升了 54.23%（表 2）。

4. 结论

本文提出了一种新颖的个性化联邦学习框架Fed-GAME，用于时间序列预测，通过采用GAME聚合器学习动态的内容感知图注意力系数。实验结果表明，FED- GAME在异构时间序列数据上取得了最先进的预测性能。未来工作中，我们计划进一步研究选择性更新过程中参数选择的自适应策略。

5. 参考文献

[1]B. McMahan、E. Moore、D. Ramage等，“从去中心化数据中进行深度网络的通信高效学习”，载于人工智能与统计。PMLR，2017，第1273–1282页。

[2] T. Li、A. K. Sahu 等，“异构网络中的联邦优化”，机器学习与系统会议论文集，第 2 卷，页 429–450，2020。

[3] A. Z. Tan、H. Yu 等，“迈向个性化联邦学习”，*IEEE*神经网络与学习系统汇刊，第 34 卷，第 12 期，页 9587–9603，2022。

[4] Z. Y. Wang、H. L. Xu、J. C. Liu 等，“通过聚类构建和分层聚合加速联邦学习”，*IEEE*移动计算汇刊，第 22 卷，第 7 期，页 3805–3822，2022。

[5] C. T Dinh、N. Tran 和 J. Nguyen，“基于Moreau包络的个性化联邦学习”，神经信息处理系统进展，第 33 卷，页 21394–21405，2020。

[6]Y. J Cheng、Z. W Zhang 和 S. J. Wang，“Fed-sds：面向异构客户端联邦学习的自适应结构化动态稀疏”，载于 *2024 IEEE*国际声学、语音与信号处理会议（*ICASSP*），2024年，第9231–9235页。

[7]F. W. Chen、G. D. Long 等，“基于图的个性化联邦学习”，*arXiv*预印本 *arXiv:2203.00829*，2022年。

[8] C. Z. Meng、S. Rambhatla 和 Y. Liu，“用于时空数据建模的跨节点联邦图神经网络”，载于 第*27*届 *ACM SIGKDD* 知识发现与数据挖掘会议论文集，2021年，第1202–1211页。

[9] R. Liu、P. W. Xing、Z. C. Deng 等，“联邦图神经网络：概述、技术与挑战”，*IEEE*神经网络与学习系统汇刊，2024年。

[10] Z.R. Zhou、G. Y. Gao、X. H. Wu 等，“通过学习动态图实现个性化联邦学习”，*arXiv* 预印本 *arXiv:2503.05474*，2025。

[11] R. Ye、Z. Y. Ni、F. Z. Wu 等，“利用推断的协作图实现个性化联邦学习”，2023，机器学习研究论文集第 202 卷，第 39801–39817 页，PMLR。

[12]Y. Li、R. Y. Xie、C. J. Li等，“面向电动汽车充电需求预测的联邦图学习，结合个性化以抵御网络攻击，”*arXiv*预印本 *arXiv:2405.00742*，2024年。[13] X. H. Wu、X. F. Liu、J. W. Niu等，“大胆而谨慎：通过审慎激进的协作释放个性化联邦学习的潜力”，载于*IEEE/CVF*国际计算机视觉会议（*ICCV*）论文集，2023年，第19375–19384页。

[14] X. H. Wu、X. F. Liu 和 J. W. Niu，“通过加法与低秩分解解耦联邦学习中的通用知识与个性化知识”，2024，MM ’24，第 7172–7181 页。

[15] Y. Li、Fan. M. F、G. Chen、C. J Li 和 Biplab Sikdar，“利用带预训练大语言模型的多门专家混合实现零样本鹅类异常检测”，载于 *2025 IEEE Kiel PowerTech*，2025，第 1–6 页。

[16] 帕洛阿尔托市，“电动汽车充电站使用情况。opendata.city ofpalo alto，2021。”，https: // data.cityofpaloalto.org/dataviews/ 257812/electric-vehicle-charging-sta tion-usage-july-2011-dec-2020/。

[17] J. Zhou 和 L. Ma，“深圳居民出行结构演变特征分析及对公共交通发展政策的启示”，《城市轨道交通研究》，第 24 卷，第 7 期，第 63–68 页，2021。

[18] H. H. Qu、H. X. Kuang 等，“一种基于物理信息与注意力机制的图学习方法，用于区域电动汽车充电需求预测”，*IEEE Transactions on Intelligent Transportation Syste ms*，第 25 卷，第 10 期，第 14284–14297 页，2024。

[19] S. R. Fahim、R. Atat 等，“电动汽车充电需求预测：一种基于图卷积神经网络的方法”，载于*2024*年第*4*届智能电网与可再生能源国际会议（*SGRE*）。IEEE，2024，第1–6页。

[20] Y. Li、G. Chen 和 X. J. Zhou，“一种基于多尺度时间感知图 Mamba 神经网络的多能源系统消耗分位数预测方法”，*IEEE*系统期刊，第1–11页，2025年。

[21] W. Hamilton、Z. T. Ying 和 J. Leskovec，“大图上的归纳式表示学习”，载于神经信息处理系统进展。2017年，第30卷，Curran Associates, Inc.

[22] Petar V.、Guillem C.、Arantxa C. 等，“图注意力网络”，载于第六届国际学习表示会议，*2018*年，会议论文，2018年。