计算机科学
COMPUTER SCIENCE

图联邦学习:问题、方法与挑战

王鑫, 熊书博, 孙凌云

引用本文

王鑫, 熊书博, 孙凌云. 图联邦学习:问题、方法与挑战[J]. 计算机科学, 2025, 52(1): 362-373.

WANG Xin, XIONG Shubo, SUN Lingyun. Federated Graph Learning:Problems,Methods and Challenges [J]. Computer Science, 2025, 52(1): 362-373.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.240500118

图联邦学习:问题、方法与挑战

王　鑫[1,2]　熊书博[1]　孙凌云[2]
1 浙江工业大学计算机科学与技术学院　杭州 310023
2 浙江大学计算机科学与技术学院　杭州 310058

摘　要　图作为一种高效、灵活、通用的数据结构,在多个学科领域得到了广泛应用。近年来,基于图的深度学习算法不断涌现,并在社交网络、生物信息学、推荐系统等领域取得显著成效。尽管公开的图数据量在增加,但高质量的数据往往分散在不同的数据所有者手中。随着社会对数据隐私保护要求的提高,现有的图学习算法面临着许多挑战。图联邦学习作为一种有效的解决方案应运而生。文中系统回顾了图联邦学习领域近五年的研究进展,将该领域的核心问题划分为3个部分,并在结构上进行了垂直整合,在关系上进行了递进阐述,包括:1)原始图数据差异导致的结构异构性;2)图联邦特性导致的模型聚合问题;3)模型整体调优方面的挑战。针对每个问题,详细分析了代表性工作及其优缺点,并总结了图联邦学习领域的典型应用和未来挑战。

关键词:联邦学习;图神经网络;图联邦学习;隐私计算
中图分类号　TP181;TP309

Federated Graph Learning:Problems,Methods and Challenges

WANG Xin[1,2],XIONG Shubo[1] and SUN Lingyun[2]
1 College of Computer Science and Technology,Zhejiang University of Technology,Hangzhou 310023,China
2 College of Computer Science and Technology,Zhejiang University,Hangzhou 310058,China

Abstract　Graph has been widely used in various fields for many years as an efficient,flexible,and versatile data structure. In recent years,graph-based deep learning algorithms have emerged,achieving significant success in areas like social network,bioinformatics,and recommendation systems. Although publicly graph data online is increasing,high-quality data remains scattered among different owners. With society's growing demand for data privacy protection,existing graph learning algorithms require enhancement. Graph federated learning is a novel approach to addresses this issue. This paper systematically reviews the research progress in the field of federated graph learning over the past five years. The core problems in the field are divided into three parts,and the structure is vertically integrated and the relationships are progressively explained:1)structural heterogeneity from differences in raw graph data;2)model aggregation issues due to federated graph learning characteristics;3)overall model tuning. For each section,it provides a detailed analysis of representative works and their advantages and disadvantages,summarizes the typical applications and future challenges in the field of federated graph learning.

Keywords　Federated learning,Graph neural network,Federated graph learning,Privacy computing

1　引言

随着计算机技术的不断进步,图这一能够有效表达现实世界中物品之间连接关系的数据结构,在医疗[1]、交通[2]以及推荐系统等各领域,节点分类[3]和链接预测[4]等各任务中得到广泛运用。这一趋势促进了大量图学习技术的发展,为挖掘图中的隐含信息提供了新的手段。

训练优秀的网络模型在实际应用中有巨大潜力,然而其需要大量训练数据,且数据通常分散在不同的持有者手中。由于隐私问题的存在,数据持有者倾向于保护原始数据,导致出现数据孤岛问题。此外,不同数据持有者存储的图数据在一定程度上具有非独立同分布(Non-Independent and Identically Distributed,Non-IID)特性,这种Non-IID性质可能体现在图结构或节点特征的差异上,从而进一步加剧数据孤岛现象。因此,如何在分布式、隐私保护的环境下有效训练模型,成为当前图机器学习领域亟待解决的重要挑战之一。

针对上述挑战,近年来涌现出许多研究工作,其中联邦学习作为一种较为有效的解决方案备受关注。联邦学习作为

到稿日期:2024-05-26　返修日期:2024-10-08
基金项目:浙江工业大学科技项目(KYY-HX-20220288,KYY-HX-20180649)
This work was supported by the Zhejiang University of Technology Science and Technology Project(KYY-HX-20220288,KYY-HX-20180649).
通信作者:王鑫(xinw@zjut.edu.cn)

一种机器学习技术，允许多个参与者在不共享原始数据的情况下共同训练机器学习模型。在联邦学习中，模型训练并不直接将显式数据发送到服务器，而是每个本地设备都拥有自己的数据，并在本地执行模型的训练，模型的更新参数被发送到中央服务器进行聚合。通过这种方式，联邦学习可以有效地保护用户的隐私。同时，联邦学习还能够降低通信成本，因为只有模型参数被传输，而不是整个数据集。这使得联邦学习在具有大规模分布式数据的情况下尤为有用，例如在医疗保健[5]、金融和物联网领域。

联邦学习可以通过显著提高图模型训练数据量来大幅提升模型精度。然而，由于图结构的复杂性和联邦学习框架的多样性，现有的综述难以全面概括图联邦学习技术的发展现状。例如，FedGraphNN[6]根据图数据的分布将联邦图神经网络(FedGNN)分为3类：图级联邦学习、子图级联邦学习以及节点级联邦学习。由于当时对该技术的研究尚不深入，该分类方法缺乏针对不同类别问题的具体分析。此外，联邦学习在应用于图数据时存在不同的处理方法，也未在该文中得到体现。文献[7]则尝试从两个角度对图联邦学习进行总结：1)主要分类法分析了图神经网络如何增强联邦学习训练以及联邦学习如何辅助图神经网络训练，为两者结合提供了清晰的视角；2)辅助分类法探讨了图联邦学习如何处理联邦学习客户端之间的异质性。该文从多个角度对图联邦学习进行了较为全面的介绍，但仍存在一些不足，例如两个分类方法的层次不够清晰，难以指导实际应用。在对垂直和水平分类中，某些共性问题被错误地局限在了特定的联邦学习类型上，如数据安全性和隐私保护，以及模型的收敛速度、通信开销和跨参与方协同训练的复杂性，这些问题是所有图联邦学习面临的共同挑战，而不仅仅局限于垂直或水平图联邦学习。

本文在此基础上进行改进，首先将两个水平的分类方法调整为3个垂直递进的部分，更符合实际应用场景。其次，本文将各个子领域的问题一一对应以便于理解。例如，在安全增强方面，对水平和垂直两个方面分别进行了探讨。这种全面的方法不仅有助于弥补现有综述文献[7]的不足，还为未来的研究提供了有价值的方向和见解。

本文的主要贡献如下：

1)提出了基于不同情景的分类，并总结了每种情景中的关键挑战。对图联邦中现有的技术进行了全面的综述。与现有的综述相比，不仅研究了更广泛的相关工作，而且提供了更精细的技术分析，而不是简单地列出每种方法的步骤。

2)对每个分支下的真实世界应用例子技术进行举例。

3)对图联邦技术进行了系统的展望分析，指出了现有方法的局限性，并提出了图联邦可能的未来发展方向。

本文主体分类方法的组织架构如图1所示。第2章介绍了对公式符号的界定；第3章由图结构出发，对图联邦学习进行了分类，并列举了某些分支可能出现的问题以及已有的解决方法；第4章则对图联邦的聚合问题进行了讨论；第5章讨论了图联邦的一些共性问题，以及可以增强模型收敛效果的一些方法；第6章列举了现实生活中的应用；最后总结全文并对未来方向进行了展望。

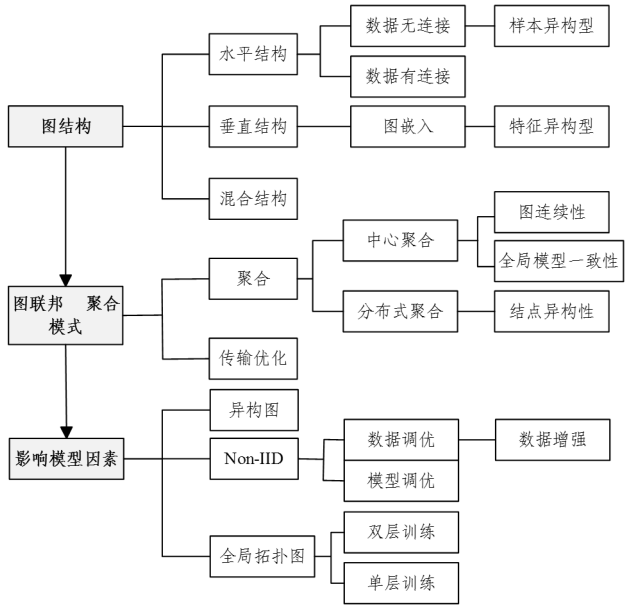


图1　文章主体分类方法架构图

Fig. 1　Architecture diagram of main classification methods

2　相关符号界定

2.1　图神经网络

GNN是一类深度学习模型，旨在对图数据进行特征嵌入和推理。GNN需要两个输入：1)图，它由节点和边组成；2)节点特征，一张图[8]可以被定义为 $G=(V,E)$，V 表示顶点或节点的集合，E 为边的集合。

图神经网络为了根据输入节点邻居信息更新节点状态，将局部转移函数 f 定义为循环递归函数的形式，每个节点以周围邻居节点和相连的边作为信息来源来更新自身的表达 h。为了得到节点的输出 o，引入局部输出函数 g。因此，有以下定义：

$$h_n=f(X_n,X_{co[n]},X_{ne[n]},h_{ne[n]}) \tag{1}$$

$$o_n=g(h_n,X_n) \tag{2}$$

其中 X_n 表示节点 n 的特征，$X_{co[n]}$ 指与节点 n 相连的边的特征，$X_{ne[n]}$ 表示与节点 n 相邻的节点的特征，$h_{ne[n]}$ 是与节点 n 相邻的节点的隐藏状态。

2.2　联邦学习

联邦学习(Federated Learning，FL)是一种协作机器学习范式，它在不交换原始数据的情况下，跨多个数据所有者训练模型[9]。在FL中，如果具有敏感本地数据的数据所有者由称为服务器的中央实体进行统一协调，则可以将其称为客户端。定义 N 个数据拥有者$\{F_1,\sim F_N\}$，他们希望通过合并各自的数据$\{D_1,\sim D_N\}$来训练机器学习模型。传统的方法是将所有数据放在一起并使用 $D=D_1\cup D_2\sim D_N$ 训练模型 M_{sum}。联邦学习系统是数据拥有者协同训练模型M_{fed}的一个学习过程，该过程中任何数据拥有者F_i都不会将其数据D_i暴露给其他客户。

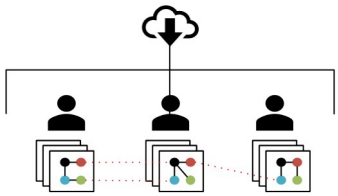
2.3　图联邦学习

在图数据中，客户端拥有私有的结构化数据集，并联合训练由中心服务器编排的图模型。形式上，每个客户 F 拥有其

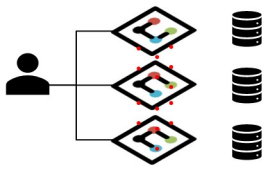
私有局部数据$\{G_1, \sim G_N\}$，其中每个$G=(V,E)$是一个图。图联邦学习的目标是每个客户端基于其局部图数据集与其他客户端协同训练一个学习模型。数据可能包含一个单一的(子)图或多个图。通常情况下，当每个客户端F拥有多个图，客户端会为图级任务训练图机器学习模型；相反，当每个客户端F拥有一个单一的图G或整个图的一个子图G时，图机器学习模型用于节点级任务。

3 数据结构异质分类

图联邦技术的目标是在一个数据分散的情况下，联合训练一个图模型，同时保证数据的隐私性。本章主要介绍图联邦技术分类中的数据部分。根据客户端之间图节点的拓扑关系与节点ID的重叠程度，图联邦数据可以被分为三大类：水平结构、垂直结构、混合结构。



(a)水平数据示意图



(b)垂直数据示意图

图2 图结构示意图

Fig. 2 Graph structure

3.1 水平结构

水平图联邦学习为联邦学习较为直观的扩展，其中各个客户端的样本为图数据，全局模型执行图级或子图级任务。水平分布的数据最主要的特点是客户端之间的数据ID和特征各不相同，各个节点的重叠程度较低，客户端拥有众多小图或者子图(区分点为是否有隐式连接)。

3.1.1 无隐式图

无隐式连接图是一种较为简单的图联邦学习场景，在此场景中，每个客户端样本都是一个图数据，全局模型则执行图级任务[10]。每一个数据库客户端内的数据相互独立，而端与端之间数据也相互独立，没有隐式边相连接。FedGSL[11]可以对FedGNNs的局部子图进行扩充，在隐藏特征和真实图的同时，为每个客户端获取多个加密视图和伪视图，并将其匿名传播集成到学习到的图结构中。此算法的泛化能力强，在不暴露个体隐私的情况下，一定程度地提高了模型精度，但该算法的前提条件较为苛刻。

该方向的典型应用之一在生物化学领域，特别是在药物研究中。在此领域，研究人员使用图神经网络来研究分子的图结构。分子可以被表示为一个图，其中原子是节点，化学键是边。通过这种方式，研究人员可以利用图神经网络来分析分子的结构和性质，从而帮助设计新的药物。研究人员可以联合多家药企的隐私数据来共同训练图神经网络模型，以此来预测新的药物结构[12]。

3.1.2 有隐式图

图数据与CV以及NLP等其他领域的数据属性差异较大，后者通常是孤立和独立的，而图数据则相互连接且相关。此情境下，客户端之间的图数据通常存在链接关系，完整的图网络被分割为多个子图，子图之间节点互为邻居或重叠。通过重新建立子图之间的关系或部分扩展子图，可以在一定程度上提高模型精度，但是现实数据中子图之间的关联程度各异，这会导致对最后精度的提升程度不同。传统方法中进行多轮训练会占用较多的计算资源，现有研究中还未有对连接最后提高的精度以及计算所需资源的平衡算法。

在传统的计算中每个节点只能对其邻居节点，即同个子图内的一跳信息进行GNN聚合，忽略了其他客户端潜在的链接节点或者相同ID节点。子图间的链接关系可以减少跨客户端信息点嵌入产生的误差，从而提高模型的性能。由于知识图谱的不完整性和隐私敏感性，简单地将不同知识图谱的数据集合在一起来共同补全知识图谱完全是不可行的。Chen等[13]引入了联邦学习的概念，使得不同知识图谱在不共享原始三元组的情况下进行合作。FKGE[14]提出了一种新颖的去中心化的可扩展学习框架来帮助联邦知识图谱嵌入，其中来自不同知识图谱的嵌入 以异步和对等的方式进行学习，同时具有隐私保护功能，利用知识图谱对之间的对抗生成，将不同领域的相同实体和关系转化到近似嵌入空间中。

尽管图中节点的相互关联可以提高模型性能，但过多的关联关系会显著增加计算开销，并且隐式关系通常是未知的，需要在联邦学习时同步计算以获取这些关系。

在全局图已知的情况下，可以基于整体的网络结构，生成全图信息的节点嵌入。该方法首先在客户端训练本地图神经网络(GNN)模型，得到对应的图表示以及节点嵌入。然后，采用联邦学习算法进行模型参数或梯度的聚合，以提升最终的节点嵌入质量。具体来说，FL服务器从客户端收集模型参数或梯度，进行聚合更新，并将更新后的参数发送回客户端以进行下一轮训练。该方法能够有效利用全局图信息，并解决数据隐私和安全问题，适用于大规模图数据的处理。Wu等[15]提出了一种针对推荐问题的FL方法，其考虑的场景假设子图具有重叠的项目即隐式连接，并且用户与物品的交互是分布式的，但完全存储在系统中，其中每个数据所有者在整个推荐用户-物品图的子图上学习。

在全局图未知的情况下，即不同客户端之间数据节点连接关系未知，现有方法[16]根据来自其他客户端F的子图的节点信息来估计局部子图的节点，然后用估计的节点扩展现有节点。然而，这种增强方案会产生高昂的通信成本，因为它需要在客户端之间共享节点信息，这也可能违反数据隐私约束。

最近也有提出使用生成缺失邻居的方法，来估计缺失邻居节点的概率分布，并根据这一分布生成边缘节点缺失多跳信息。FedSage+[17]是一种用于子图联邦学习的改进方法，旨在解决数据异构性和缺失链接问题。该方法在每个数据所有者的本地子图上生成缺失邻居节点，然后利用这些生成的节点修补原始子图，采用随机保留节点和链接的策略提高缺失邻居生成器的性能。但是该方法未考虑客户端之间存在

重叠节点的情况,使得对节点生成的解决方案不完善。FedNI[18]首先使用图生成对抗网络(GAN)联合训练缺失节点和边预测器,以补全本地网络的缺失信息,使用图联邦学习平台训练跨机构的全局 GCN 节点分类器。FedLit[19]提出了一个新颖的图结构 FL 框架,该框架通过基于期望最大化的聚类算法发现潜在的链接类型,并通过分布式本地客户端的协作来建模与这些识别出的链接类型相关的消息传递。

当存在一个由所有子图组成的全局图,其中客户端内的子图的彼此连接往往比外界子图更紧密。基于网络同质性,同一 FL 客户端内的密集连接的子图具有相似的属性,而两个不同 FL 客户端中的子图则不然。这种差异会导致朴素的联邦学习算法无法有效融合来自不同客户端的知识,甚至导致模型性能下降。为了解决这个问题,研究人员提出了 FED-PUB[16],引入了一种新的子图 FL 方法,通过对局部模型参数进行个性化权重平均和屏蔽来个性化子图 FL 算法,从而捕获相互关联的子图之间的客户端结构,其重点是相互关联的局部 GNN 的联合改进,而不是学习单个全局模型。该方法可以进一步扩展到客户端图级计算,即在无隐式图状态下或不考虑边连接的状态下进行计算,通过相似度计算为选定图选择相关图网。

该方向现有的解决方案效果大多一般,或者在解决方案上并不完善。对如何寻找相邻节点、重叠节点的研究较少。该问题的难点在于如何在不破坏联邦隐私保护的基础上,从不同的客户端提取出节点进行对比。并且上述算法大多忽略了由全局图的不同社区组成的子图之间不可避免的异质性,从而使局部 GNN 模型的不相容知识崩溃。

3.1.3　水平数据安全增强

由于需要通过共享模型参数和图拓扑结构来实现客户端之间的交流,FedGNN 很容易被攻击,因此需要更多地关注 FL 系统的安全增强。水平数据安全增强主要分为两个方面:1)基于 Secure Aggregation[20]的水平联邦学习;2)基于同态加密[21]的水平联邦学习。Secure Aggregation 设想每个参与者在上传模型数据前,在自身的模型上加入大量噪声保护传输安全。同态加密技术对参与者的梯度信息进行保护,客户端加密本地训练,服务器使用同态加密保护全局模型聚合。Geisler 等提出了一种新颖的鲁棒聚合函数 Soft Medoid[22],通过适当的预算,攻击者只能扰乱聚合输入的子集,以达到跨越决策边界的目的。只要攻击者仅控制少数输入,无论攻击特征如何,鲁棒估计器都会出现有限误差(即没有攻击可以任意扭曲聚合结果)。

3.2　垂直结构

当图数据以垂直方式排列时,FL 客户端持有节点 ID 相同但图内数据的特征和连接方式可能不同。在垂直联邦(VFL)环境中,不同客户端虽拥有重叠的节点 ID,但这些节点所包含的特征空间各异。全局模型在训练时综合考虑各个客户端数据的特征,以及各个客户端数据的重叠关系,从而提高模型精度。

在现实应用场景中,一个实例或者一个实例的一部分可以属于多个 FL 客户端。分布在不同客户端的图数据由于节点重叠,通常包含互补信息。由于这些重叠的节点无法共享和聚合数据,因此每个客户端上的图结构并不全面。如图2(b)所示,不同客户端中节点 ID 相同,但是连接关系、节点数量各不相同,在一定程度上互补了信息。在这种情景下,FL 训练期间来自不同客户端的重叠实例的嵌入可能来自不同的嵌入空间,解决该问题的关键在于找出重叠实例,根据客户端的本地实例嵌入来学习全局实例嵌入。

根据图数据的完整性,垂直结构可以进一步细分为两个子场景。

1)具有部分数据的客户端。在这种情况下,图拓扑、节点特征和节点标签由不同客户端拥有。例如,在一个有 3 个客户端的 FL 系统中,一个客户端可能仅拥有节点特征,另一个拥有图拓扑,第三个拥有节点标签。如何在保护隐私的同时使这些客户端协同工作是一个关键挑战。

2)具有全部数据的客户端。在这种设置中,假设客户端包含完整的图数据,包括图拓扑和节点特征。然而,它们的节点特征类型可能各不相同。安全集成这些不同的节点特征,是使用这类分布式图数据训练 GNN 模型的关键策略。

这样的细分有助于更深入理解垂直 FedGNN 的应用场景和面临的挑战。由于篇幅限制,本节不会对每个子场景进行详尽展开,但通过上述说明,读者可以更明确地区分垂直结构的不同应用环境,与水平结构形成鲜明对比。本节将主要阐述垂直结构中主要面临的问题以及可能的解决方法。

3.2.1　图嵌入

经典的 VFGNN[23]将图嵌入主要分为 3 个部分:1)通过协助加密进行初始化节点嵌入;2)基于初始节点嵌入,通过在图上使用多跳邻域聚合来生成本地节点嵌入;3)结合来自数据持有者的本地节点嵌入并获取全局节点嵌入。

在某些情景中,节点中可能包含时间序列数据且局部图拓扑可能随时间变化,因此需要动态地嵌入分布式图数据时空信息。Feddy[24]通过应用动态 GNN 来考虑图嵌入中的时间信息。4D-FED-GNN+[25]专注于缺失时间点的进化图学习任务。每个客户端为每个时间步训练一个 GNN 模型。根据当前以及下一个时间步数据是否可用来决定算法作为生成器或自编码器。它提高了本地模型的预测性能,同时受益于具有与缺失时间点相对应的数据的其他客户端。

FedGL[26]上传每个客户端的预测结果和节点嵌入到服务器,发现全局自监督信息,包括全局伪标签和全局伪图,从而减轻异构性并利用互补性。通过融合节点嵌入并重构全局伪图,实现了高质量的全局模型。全局伪图的发现过程同样保护了隐私,并充分利用了互补性。

3.2.2　重叠实例对齐

以知识图谱为例,每个客户端都拥有一个知识图谱,并且每个知识图谱可能具有存在于其他客户端上的重叠实体。当存在重叠节点时,由于每个客户端的知识图谱是独立嵌入的,这使得不同客户端的重叠节点可能会被映射到不同的位置,从而造成最后结果的偏差。而对齐模块旨在将不同客户数据的嵌入统一到同样的向量空间中,这样就能够识别出对齐的实体了,而这个统一操作也是知识图谱对齐最大的挑战。这里需要注意的是,有些基于 GNN 的方法可能在嵌入阶段中进行共享模型权重之类的操作,从而使二者的嵌入尽可能

接近，因此嵌入模块事实上也可能完成了一些对齐功能。

在每轮 t 之后，服务器从每个客户端收集每个局部嵌入矩阵以更新全局嵌入矩阵。然后服务器将全局嵌入分发给相应的客户端以进行后续的本地训练。FedE[13]技术的主要思想是在保护隐私的前提下，仅在每个客户端上训练实体嵌入，并在服务器上聚合这些嵌入，而不将客户端的三元组数据发送到中央服务器。

3.2.3 垂直数据安全增强

从分布式计算的角度考虑，水平联邦是一种数据并行的思路，而垂直联邦是模型并行的思路。它们的不同点在于水平图联邦与模型耦合度较弱，目前主流模型都基于水平联邦的框架进行实现；而垂直图联邦中，每个隐私保护方案都和模型紧密绑定，一般需要对单机版本的模型进行拆解，规定每个参与者本地需要计算的内容以及多方之间需要交换的信息。在设计垂直分布的图数据模型时，需要对每种算法进行安全性分析，不同的算法实现方式可能会有不同的安全性考虑。

在 VerFedGNN[27]中通过随机投影传递了邻域嵌入聚合，以及受三元量子化机制扰动的公共参数梯度，显著增强了跨图交互的隐私保护，同时保持了竞争预测的准确性。

一种较为通用但效果较差的方式为在传输过程中统一加密。目前有研究人员致力于在此情景下提高安全保护效果，如 FedVGCN[28]，这是一种图神经网络垂直联合学习算法，用于数据垂直分区设置下的隐私保护任务，可以推广到现有的 GCN 模型。该算法将计算数据分为两部分，对于训练过程的每次迭代，双方都在同态加密下互相传输中间结果。

3.3 混合结构

目前大多数工作都集中在水平或垂直数据分布上，其中每个客户端分别拥有具有共享特征的不同样本，或者每个客户端完全共享样本索引。但是现实生活中数据大多是混合的，同时包含水平和垂直数据。混合图集成了水平图和垂直图的各个方面。在此配置中，不同客户端的数据可能不具有相同的特征空间或样本 ID。混合框架中的一个常见挑战是客户端仅持有部分样本 ID 和特征集的场景。这些数据集通常重叠，但与其他客户端持有的数据并不完全相同。此外，与集中式数据设置相比，混合联邦学习中的数据集通常不完整，这使得学习过程进一步复杂化。

传统的混合联邦学习方法通常采用聚合客户端模型参数的策略来构建服务器模型。尽管这种方式简便，但其无法充分利用多个客户端的表征来进行深入的协同合并，因为它仅仅是对客户端的知识（即模型参数）进行简单的组合。Zhang 等[29]为混合 FL 建立了一种新的基于模型匹配的问题表述，然后提出了一种有效的算法，可以协同训练全局和局部模型来处理完整和部分特征数据。FedHD[30]引入了跟踪变量，使客户端能够跟踪全局梯度信息并基于本地数据更新模型，允许客户端执行局部随机梯度下降的多个步骤，从而提高通信效率。

FedGraph[31]提出了一种新颖的混合联邦学习策略，其不仅聚合训练好的本地模型，还通过图卷积网络（GCN）有效地聚合多个客户端的表示，以实现更优的预测结果。此方法不单单强调模型参数的合并，而是着重于客户端的特征表示的整合，使得在确保数据隐私的前提下，能够充分利用跨客户端的特征共享信息，从而显著提升模型的预测性能。

本节主要介绍了图联邦学习在不同数据分布情况下的分类和关键技术。首先讨论了水平分布数据，其中每个客户端的数据样本都是图数据，但各个客户端之间的数据 ID 和特征存在差异，并可能存在不同程度的重叠。针对水平分布数据，主要介绍了包括无隐式连接图和有隐式连接图两种主要情况下的应用及其关键技术，例如图结构信息获取、图融合和生成缺失邻居节点等。其次，探讨了垂直分布数据，即客户端持有相同节点 ID 但图内数据的特征和连接方式可能不同的情况。在垂直分布数据方面，详细介绍了图嵌入方面的关键技术，包括动态图神经网络利用时间信息和时空信息嵌入分布式图数据的方法。最后，提到了混合结构的数据，即同时包含水平和垂直数据的情况。针对混合数据，讨论了局部优化和模型聚合等关键技术。通过对这些不同数据结构下的图联邦学习技术进行分类和总结，有助于深入理解联邦学习并将其应用到图数据领域。表 1 总结了基于数据分类的代表性文献中的图数据类型、关键技术以及应用任务的情况。

表 1 基于数据分类的代表性工作总结

Table 1 Summary of representative works based on data classification

	代表性文献	关键技术	应用任务
水平数据	FedGSL[11]	图结构增强	优化图结构学习
	FedE[13]	图融合	预测缺失的三元组来进行知识图谱补全
	FKGE[14]	图融合	联邦知识图谱嵌入
	FedGNN[15]	图融合	针对推荐算法的联邦学习
	FEDPUB[16]	图融合	局部 GNN 的联合改进
	FedSage+[17]	生成缺失邻居节点	处理跨局部子图的缺失链接
	FedNI[18]	生成缺失邻居节点	利用网络修复和机构间数据
	FedLit[19]	生成缺失邻居节点	发现潜在的链接类型
垂直数据	Feddy[24]	图嵌入	多用户图序列中学习对象表示
	4D-FED-GNN+[25]	图嵌入	联合脑图演化轨迹预测框架
	FedGL[26]	图嵌入	在保护数据隐私的同时获得高质量的全局图模型
	VerFedGNN[27]	安全增强	基于联邦 GNN 的推荐系统
	VFGNN[23]	安全增强	节点分类
混合数据	FedHD[30]	局部优化	混合图模型整合
	FedGraph[31]	模型聚合	混合图模型整合

3.4 图联邦分类与经典分类的异同

联邦学习（FL）作为一种分布式学习范式，旨在解决数据隔离问题。在 FL 中，多个客户端协作进行训练，无需集中式的本地数据。尽管 FL 在计算机视觉领域和自然语言处理

领域取得了成功应用，但在图数据的机器学习领域应用仍然有限。主要挑战包括：1）缺乏对不同图 FL 设置和任务的统一形式化描述；2）现有 FL 框架对多样化数据集和学习任务的支持不足；3）基于模拟的联邦训练系统对大规模和私密图数据的研究存在效率和安全性问题。

本文受经典联邦学习启发，根据图结构可将联邦学习进一步细分为水平、垂直、混合 3 种结构。与经典联邦学习中的分类方法不同，经典联邦学习分类方法是根据样本重合程度来划分的。横向联邦学习指两个样本集合中样本的特征重合较高，但是样本来源不一样。纵向联邦学习指两个数据集 ID 重合较高，但是特征不一样。联邦迁移学习则侧重于利用源领域学习过的模型，基于数据、任务或模型之间的相似性，将其应用于目标领域的学习过程。

而图联邦学习分类方法根据客户端之间图节点的拓扑关系与节点 ID 的重叠程度来划分。水平图联邦指两个图样本集合中样本的节点 ID 重合度较低（图之间拓扑差异性较大）。垂直图联邦学习与经典联邦学习中的纵向联邦学习类似，样本间的 ID 重合度都较高，但特征不同。而混合图联邦中主要考虑的是现实应用场景下的图数据往往错综复杂，其中垂直与水平数据交织。

4　图联邦的聚合模式

聚合问题指如何有效地将各个设备或节点上训练得到的模型参数进行集成，以更新全局模型。由于图数据的特殊性，在图结构上实现有效的模型更新和参数聚合成为一个主要挑战。聚合过程需要考虑图结构中节点之间的连接关系，并确保在模型参数聚合后保持图的拓扑结构和节点属性的一致性。

在解决聚合问题时，需要应对节点的异构性，这可以通过引入节点嵌入技术来实现。同时，保持局部更新与全局一致性之间的平衡至关重要，常见的方法包括加权平均和模型压缩等技术。此外，为了保持图结构的连续性和一致性，通常利用邻居节点的信息来实现。解决聚合问题需要综合考虑图结构的特性、节点的异构性以及隐私保护等因素，并设计相应的聚合算法来确保模型参数的有效集成和全局模型的收敛性。在解决聚合问题的过程中，研究者们提出了多种方法来应对不同方面的挑战。表 2 总结了现有研究工作在聚合问题上的优缺点。

表 2　水平与垂直结构优缺点对比

Table 2　Comparison of advantages and disadvantages of horizontal and vertical structures

分类	子分类	优点	缺点
水平联邦	无隐式图	计算方便 泛化性强 安全性强	各个客户端之间节点关联度低，Non-IId 问题严重
	有隐式图	综合考虑节点联系，减少 Non-iid 问题	节点信息难以获取，处理不当可能会暴露数据
垂直联邦	无完整数据	垂直对比数据，可发现更多隐藏信息	安全防护较为繁琐，需针对任务个性化设计； 目前垂直对比的层数有限
	有完整数据		

4.1　中心服务器聚合

在中央服务器聚合和客户端聚合两种情况下，聚合的方式各有不同。对于中心服务器聚合而言，关键在于如何在保证数据隐私的前提下，有效地集成来自不同客户端的模型参数[32]。这要求设计出能够处理异构节点属性和保护隐私的聚合算法，并在聚合过程中维护全局模型的一致性和图结构的连续性。

最经典的 FedAvg[33] 在联邦学习的过程中，每个本地机构训练本地模型，然后将本地模型参数进行聚合。通过平均本地模型的参数来更新全局模型，从而实现在多个机构之间共同训练全局模型的目的。FedGCN[34] 提出了一种基于在线可调注意力机制的联邦聚合算法。该算法将可训练参数输入注意力机制中，聚合方法为每个局部模型分配相应的注意力系数，降低了低效的局部模型参数对全局模型造成的损害，提高了 FL 算法的容错性和准确性。FedU[35] 用于解决通信集中式方案下的公式化 FL 问题，服务器对客户端子集 $S(t)$ 进行统一采样，并将本地模型的最新更新发送到每个客户端，在执行 R 个本地更新步骤后，服务器从采样的客户端接收最新的本地更新，以对每个本地模型进行模型正则化。

4.2　客户端聚合

对于客户端聚合，重点在于如何在设备端进行局部模型更新后，实现全局模型的有效更新[36]。聚合过程中不仅要考虑到图结构的特点和节点的异构性，还需要兼顾隐私保护和全局模型的优化。

FedGTA[37] 提出了一种通过拓扑感知局部平滑置信度和混合邻居特征进行优化的个性化优化策略，其结合了 Non-param LP，允许在每个客户端中显式地考虑模型预测和拓扑结构。Deng 等提出了可以在整个训练过程中根据局部模型的训练情况自适应调整聚合权重的 FedGraph[38] 模型，从粗到细考虑了 3 个因素：每个本地数据集大小的比例、模型图的拓扑因子和模型权重。dFedU[35] 假设每个客户端都有一个任务，并且预先给出了一个完全连接的客户端间图。一旦每个客户端从其邻居获取本地更新的模型，它就会通过图正则化执行模型更新。Lalitha 等[39] 提出了一种分布式学习算法，其中节点通过聚合本地观测数据和其一跳邻居模型的信息来更新自身，以共同学习一个最适合整个网络观测的模型。SpreadGNN[40] 假设每个客户端解决多个任务，客户端间任务关系图是根据任务分类器模型参数初始化的。客户端应用分散周期平均随机梯度下降来优化目标函数，并在收敛保证下迭代更新模型权重及其相应的任务关系图。

4.3　客户端通信问题

不论是中心服务器聚合还是客户端聚合，通信开销和传输效率一直是主要的问题。数据传输时的开销问题主要是由客户端和中央服务器之间的网络连接和传输数据（例如模型、参数）所引起的。为解决通信开销问题，最简单直接的方法是在联邦学习框架中牺牲模型准确度，仅训练占用通信空间较小的低容量模型。CCESA[41] 提出了一种低复杂度方案，将秘密分享节点的拓扑设计为稀疏的随机图，而不是与现有解决方案对应的完全图，大幅减少了通信/计算资源开销。

为了提高 O-GFML 的通信效率，Gogineni 等开发了一种

基于部分共享的 O-GFML(PSO-GFML)[42]。PSO-GFML 允许参与的客户端在全局迭代期间仅与各自的服务器交换一部分模型参数,而非参与的客户端如果可以访问新数据,则更新其本地模型。图注意网络(GAT)与之类似,可以在训练过程中学习图,而不是提前知道邻接矩阵。此外,Malinovsky等[43]提出了一种高效通信的分布式定点优化方法,从解决优化问题和寻找凸凹函数的鞍点的角度出发限制客户端本地计算,从而解决了联邦学习通信开销瓶颈问题。FedGraph[44]通过调整采样策略和模型参数来降低通信和计算成本,使用了一种基于强化学习的采样策略来优化服务器端的采样过程,同时也对客户端的 GNN 模型参数进行了优化。在 FedGCN[45]算法中,模型是在客户端上进行训练的,每个客户端在本地计算对模型的局部更新,然后将这些更新发送到中心服务器进行聚合。在训练开始之前,客户端之间只进行一轮通信,将跨客户端边缘的信息发送到中心服务器,以便模型可以利用这些信息进行训练。这种方式可以大大减少通信开销,并提高隐私保护性能。

表 3 聚合问题代表性工作总结

Table 3 Summary of representative works based on aggregation issues

	代表性文献	关键技术
聚合算法	FedAvg[33]	平均模型参数
	FedGTA[37]	个性化优化
	FedGraph[38]	自适应调整聚合权重
	dFedU[35]	图正则化执行模型更新
	Peer-to-Peer Federated[39]	结合邻居模型更新
	SpreadGNN[40]	稀疏随机图代替完整图
通信策略	CCESA[41]	限制客户端之间交换的参数量
	PSOGFML[42]	限制客户端本地计算
	Local Fixed-Point Methods[43]	调整采样策略和模型参数
	FedGraph[44]	限制客户端通信轮次
	FedGCN[34]	稀疏随机图代替完整图

5 模型调优

在图联邦学习的实践中,为了解决数据分布不均、模型参数更新效率低下等问题,不断有优化策略被提出,其中全局拓扑图作为一种较为新颖的优化策略脱颖而出。全局拓扑图的建立有助于维护节点之间的连接关系,促进信息传播和模型聚合的高效性。此外,针对非独立同分布问题,既可从数据方面优化数据采样和分布策略,确保节点数据的多样性和代表性,也可从算法方面调整学习算法以适应不同的数据分布情况。异构图处理则是针对图数据的多样性和复杂性,开发适用于不同节点和边属性的处理方法,以更好地捕捉图数据的特征。这些辅助手段的综合应用将有助于提高图联邦学习的性能和适用性。

5.1 全局拓扑图

联邦学习中,客户端分散计算数据,通过本地或中心服务器进行计算聚合。但是在计算中客户端与客户端之间的拓扑关系[46]常常被忽视。考虑到这些连接,客户端可以形成全局拓扑图,如图 3 所示。全局拓扑图使单独的客户端可以在某种程度上获取全局信息,优化预测能力。但要注意的是,客户端之间的拓扑关系往往是时空关系,各个客户端存储数据之间时空的差距进一步生成了图拓扑关系,一般数据往往不具备构建拓扑图的条件。例如,CNFGNN[47]是一种联邦时空模型,其同时考虑了节点的时间和空间特性,在设备上的时间动态建模和服务器上的空间动态建模,利用交替优化来降低通信成本,促进边缘设备上的计算。SFL[48]提出了一种新颖的结构化联邦学习框架,旨在利用客户之间基于图的结构信息来加强 PFL 中的知识共享,利用客户关系图和客户的私有数据同时学习全局模型和个性化模型。

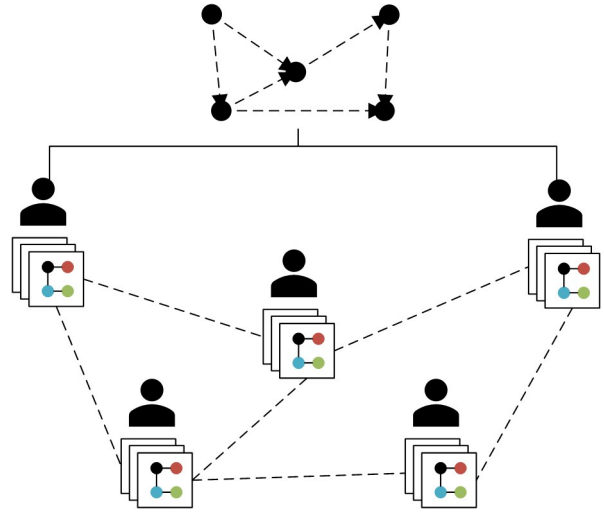


图 3 全局拓扑图示意

Fig. 3 Schematic diagram of global topology

在已知拓扑图的情况下,存在两种不同的训练方式:双层训练和单层训练。

双层训练设计了内部和外部两个目标函数、多个权重。内部权重使本地任务能够向个性化发展,外部共享权重针对非独立同分布问题,促使各个任务能够向全局约束空间发展。BiG-Fed[49]分别设置了内外部函数,利用联邦学习客户端的连接性(即全局图结构中的拓扑信息)作为外层函数的指导,并将其映射为权重,衡量相邻客户端模型的相似性。内部函数将外部层权重引入内部任务中,以保持新权重的聚合和局部权重的更新之间的有界距离。

与双层优化训练不同,单层训练采用单独的目标函数依次训练局部模型和 GNN 模型。Lee 等[50]受到图正则化的启发,提出了一种新的融合框架,只需要对本地估计进行一次传输即可。具体地,该方法将局部估计结果线性组合,为每个任务生成改进的估计,最终通过综合考虑任务相似度与任务难度来确定理想权重。

5.2 Non-IID 问题的缓解

导致 Non-IID 问题的原因多种多样,包括每个客户端对应于特定用户、特定地理位置或特定时间窗口,以及客户端上数据分布的差异。图数据由数据和图组成。Non-IID 问题主要体现在两个方面:数据分布和图拓扑。数据分布中的非独立同分布问题,包括属性分布和标签分布,与普通 FL 中的非独立同分布问题相同。图拓扑中的 Non-IID 问题则是指客户端之间的节点度、边权重、边类型或图结构的非独立同分布。不同的 Non-IID 问题,有不同的解决方法。目前的解决方法主要针对数据分布,分别集中在数据层面和算法优化层面[51]两个

角度，而对图拓扑的 Non-IID 特性的研究较少。Xie 等[52]分析了跨客户端的 Non-IID 图结构，但没有完全解决此问题。

5.2.1　*数据方面*

图联邦中不同图数据集具有异质性，这种图形异质性既可以在统计意义上存在，即尽管节点和边的类型相似，但图数据集的节点和边数量也相差较大；也可以在结构方面存在，即来自不同图数据集的实体可能相同，但邻域不同。

数据增强最初是一种通过随机变换或知识迁移来增加训练数据多样性的技术，可以用于缓解 FL 中的局部数据不平衡问题。经典的处理方式为采样、调整权重等对数据进行改进的方法。例如日本会津大学的相关研究人员提出的 FedGraph[44]算法，采用了两种不同的图网络采样方式，一种是不直接依据节点而是通过计算后的隐藏层采样；另外一种是利用 layer 节点去采样。

通过调整权重学习，可以参考 ASFGNN[53]，通过数据分布之间的差异计算 JS 散度调整权重，通过不同节点计算 JS 散度，然后调整学习权重。ADASYN[54]根据学习难度对不同的少数类示例使用加权分布。与那些更容易学习的少数类示例相比，ADASYN 为更难学习的少数类示例生成更多的综合数据，通过减少类不平衡带来的偏差，以及自动地将分类决策边界转向困难的例子，来对抗样本不平衡的问题。FLIT+[5]通过实例重新加权和调整跨客户的训练来缓解在跨客户处理异构分子时出现的问题。

当前的 FL 算法通常假设独立同分布（IID）的训练数据才能获得良好表现，但图联邦特有的图数据的不可分离性和复杂的图模型设计问题也可能会严重降低联合训练模型的性能。FedAlign[55]将损失函数限制为 L-Lipschitz 平滑并测量隐藏层的最佳传输（OT）距离。

5.2.2　*模型方面*

模型方面的改进主要包括提高局部模型的适应能力以及学习一个强大的全局联邦学习模型两个方面。从这两个方面出发，对每个客户端进行个性化改进，主要包括模型插值、正则化的本地损失、元学习、知识蒸馏等几种方法。

在 GraphFL[56]中将模型无关元学习（MAML）合并到 FL 中，遵循 MAML 的训练方案在服务器上学习一个全局模型，从而缓解 Non- IID 图数据引起的问题。此外，利用现有的 FL 方法进一步更新全局模型，使其能够在测试节点上实现良好的泛化。FedStar[57]提出了通过结构知识共享进行图联邦学习，充分利用图中自然结构信息来处理图间 FL，捕捉并分享多个客户的通用结构知识，以提高 FGL 客户的本地表现。Hanzely 等[58]通过增加一个正则化项，设计了一个新形式的代价函数来研究局部模型和全局模型之间的权衡。Dinh 等[59]在提出的 pFedMe 算法中引入 Moreau Envelopes，以克服客户端之间的统计多样性。然而，当本地数据差异巨大时，为每个客户协作训练个性化模型比学习单个全局模型更合适。FedPer[60]作为一种个性化学习模型，共享基础层，同时为每个客户端提供本地个性化层，以保留本地知识。

集群 FL 框架[61]在群组级别上融合了个性化，同时有效地减少了通信成本。该方法旨在通过将客户端组织成集群，在每个集群内进行个性化 FL，从而在个性化和通信效率之间取得平衡，进而解决客户端异质性问题。

5.3　异构图处理

在商业化领域，如保险中的车险反欺诈场景，异构图的应用相当普遍。在这种场景中，人、车、物分别作为不同的节点，节点之间的类型关系至关重要，并非所有节点之间都有直接连接，这导致了节点和边之间的特殊性。在处理异构图时，有两种主要方法：1）首先对图进行嵌入，在嵌入图上做节点分类；2）引入伪标签修正损失函数，将标签和节点分离，这种方法适用于简单节点和边组合的场景。目前普遍采用的异构图处理流程如图 4 所示。

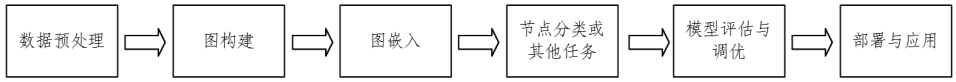


图 4　异构图处理流程

Fig. 4　Processing workflow of heterogeneous graph

FedGL[26]引入伪标签做损失的计算，把伪标签和全局隐含梯度结合起来形成好的循环，以实现不同节点标签之间的互补。通过全局学习对伪标签进行修正，使伪标签越来越准确，从而与本地标签形成互补。GCFL[52]框架根据 GNN 的梯度动态查找本地系统的集群，并从理论上证明此类集群可以减少本地系统所拥有的图之间的结构和特征异质性。FedHG+[62]联合训练了一个类型感知的缺失邻居生成器和一个类型感知的 GCN，以处理不完整的子异构邻域。

6　典型应用

图联邦学习因其能够在保护数据隐私的前提下进行跨机构、多方协作的高效图分析，在金融、医疗、交通等领域的应用正在显著增加。其独特的优势在于，不仅能从分布式的图数据中挖掘出更深层次的知识，还能有效避免数据孤岛问题。表 4 总结了图联邦学习在不同领域的代表性应用及其对应的案例。

表 4　代表性应用总结

Table 4　Summary of representative applications

应用场景	代表性文献	具体解决问题
金融	Anti-Money Laundering[63]	反洗钱监测
	GraphSniffer[64]	虚拟货币保护
医疗	FedGP[65]	医学影像噪声问题
	Hyper-Graph[66]	心理健康症状检测模型
	FedNI[18]	疾病预测 FL 框架
交通	FedAGCN[2]	交通流预测（基于异步图卷积网络）
	MFVSTGNN[67]	交通流预测（基于时空图）

6.1　金融

跨域的图神经网络技术已经在金融犯罪监控（如诈骗、偷盗、洗钱等）和药物发现等领域得到应用。在处理跨国金融犯罪行为（如跨国诈骗和洗钱）时，单一银行或国家/地区的数据

通常难以发现这些犯罪行为。

部分企业和研究机构在图联邦领域已经取得了显著的进展。例如 IBM[63] 针对反洗钱监测问题，联合多家银行进行图神经网络建模，利用各家银行的转账交易数据、不同银行之间的跨行交易数据进行全局图的构建，相比单个银行建模反洗钱可疑账户的识别准确率提高了 20%。

虚拟货币的蓬勃发展为犯罪分子提供了天然的庇护所，导致各类恶意交易层出不穷，严重危害了数字货币的金融秩序。许多研究人员开始关注这一领域，如 GraphSniffer[64] 可以在保护交易特征数据安全和模型隐私的同时，实现对恶意交易的联合识别和分析。但现有的基于联邦图的交易欺诈检测方法侧重于垂直框架，该框架仅支持每个字段一个参与者。然而，仅仅利用垂直框架不足以在金融网络中构建有效的交易欺诈检测模型。首先，同一领域的不同参与者（如不同的银行）无法自由地与他人共享数据。此外，由于私有集交叉过程，大量的本地数据会被丢弃。现有研究在这些方面都未提出一个较为完善的解决方案。

6.2 医疗

大多数医疗机构面临的一个关键挑战是在数据不共享的情况下解决疾病预测问题。FedGP[65] 提出了一个联合图纯化的联合框架，通过服务器和客户端协作来解决 FL 中的标签噪声问题。Ahmed 等[66] 提出了一个结构超图以及一个用于单词表示的情感词典，开发了基于联邦学习的嵌入模型用于心理健康症状检测，该模型将文本数据视为连续单词的集合。FedNI[18] 提出了一种新的疾病预测 FL 框架，使用从全球人口图划分的不相交的小型局部图来制定用于疾病分类的跨孤岛联合图学习。该方法通过新颖的网络修复模块提高了 GCN 在自我网络上的有效性，并通过在 FL 框架中训练节点生成器和节点分类器来提高性能。

6.3 交通

准确、实时的交通流量预测是智能交通系统的重要组成部分，平衡预测模型的预测精度和时间成本是一个具有挑战性的问题。FedAGCN[2] 提出了一种基于联邦学习和异步图卷积网络的深度学习框架来实时准确地预测交通流，应用异步时空图卷积来模拟交通数据中的时空依赖性。为了降低深度学习模型的时间成本，提出了一种图联邦学习策略 GraphFed 来训练模型。

MFVSTGNN[67] 是一种基于联邦学习的智能交通流预测模型，该模型将基于时空图的深度学习模型集成到多级联邦学习框架（MFL）中。该 MFL 用于允许不同数据所有者之间的数据协作来训练有效的模型，而无需共享其私有数据，同时实现通信开销和计算性能之间的权衡。

7 未来挑战与展望

7.1 图联邦学习中的公平性问题

公平性问题主要包括两方面：1）数据方面的公平性；2）设备之间的公平性。在数据方面，不同的客户可能拥有具有不同属性的图。例如，一些客户端具有高节点度的图，而其他客户端具有低节点度的图。这种有偏图会影响最后的聚合。设备之间的公平性问题则是由不同设备的处理能力、可用性等造成的，较为简单的缓解方式为仅根据设备过去的贡献概况来确定贡献的权重，但是这并不能从根源上解决问题。

未来的研究可以探索开发更加公平和健壮的图联邦学习算法，考虑如何有效处理不同客户端间图数据的属性差异，以及如何在分配贡献权重时考虑到设备间的性能差异，从而提高整体系统的公平性和稳定性。

7.2 安全问题

图联邦学习由于其天然属性，存在许多易受攻击的方面，如共享节点嵌入、图拓扑和模型参数。Chen 等[68] 提出了多个纵向图联邦可能被攻击的角度，并针对纵向图联邦提出了一种新型对抗性攻击方法 Graph-Fraudster。该方法通过 GVFL 的隐私泄露和成对节点梯度，混淆 GVFL 中的服务器模型，基于噪声添加的全局节点嵌入生成对抗性扰动。Li 等[69] 提出了一种由诚实但好奇的参与方基于生成式网络发动的嵌入表示重构攻击方法。该攻击在多个数据集上成功且完整地重构了参与方的嵌入表示，进一步凸显了在图联邦学习中参与方嵌入表示时存在的隐私泄露风险。

未来的研究可以集中于开发更加安全的图联邦学习方法，包括改进隐私保护技术、设计抗攻击的模型参数共享机制，并探索新型的对抗性攻击检测和防御方法，以应对日益复杂和普遍存在的安全威胁。

7.3 图联邦的设计基准以及标准数据集问题

现有文献对各种图联邦设置和任务缺乏统一的表述，使得关注基于 SGD 的联邦优化算法的研究者难以理解图联邦中的本质挑战。现有的 FL 库不支持多样化的数据集和学习任务，难以满足不同模型和训练算法的需求。考虑到图数据的复杂性，在 FL 环境下训练，GNN 的动态性可能不同于训练视觉或语言模型。

未来的研究可以致力于建立更加完善和多样化的图联邦学习基准测试平台，包括标准化开放数据集的建立和相关实现工具的开发，以促进不同模型和算法在真实环境下的比较评估，并推动图联邦学习领域的标准化和进步。

7.4 混合图联邦问题

现有的图联邦研究主要聚焦于水平和垂直两种数据分布，但现实应用场景中的图数据往往不是纯粹的水平或垂直类型，而是这两种类型的交织且通常不完整，这一特性导致现有的图联邦算法难以在这些场景中表现良好。目前针对这一研究方向的成果尚少，但其对图联邦实际应用的影响极为显著。因此，当前研究应重点关注开发适用于混合图数据的图联邦算法，包括探索如何有效处理水平与垂直数据混合及不完整性问题，以增强图联邦算法在复杂应用环境中的适用性和性能。通过这种方式，可以更好地适应实际应用中的图数据特性，从而提高算法的整体效果和应用价值。

结束语 本文系统全面地概述了图联邦学习的关键技术和方法。从图数据的异构性处理到图模型的构建和优化，详细讨论了如何通过建立全局拓扑图来维护节点间的连接关系，以及如何通过改进的聚合算法和通信策略提高模型的效率和效果。此外，本文还深入探讨了模型调优的各种策略，包括应对非独立同分布问题的算法调整和异构图数据的处理

方法，这些策略极大地提升了模型在实际应用中的适用性和性能。

图联邦学习作为一种新兴的学习范式，正逐步克服初期技术困难，朝着成熟和广泛应用方向迈进。随着技术的进步和更多实验的验证，图联邦学习有潜力解决更多现实世界的复杂问题，特别是那些要求高度数据隐私和安全性的应用场景，如跨领域数据融合。未来研究需要集中于优化算法的设计，提高处理效率，以及探索更多能够支持异构和动态数据源的图模型结构。

参 考 文 献

[1] MITTONE G, SVOBODA F, ALDINUCCI M, et al. A Federated Learning Benchmark for Drug-Target Interaction[C]// Companion Proceedings of the ACM Web Conference 2023. New York, NY, USA: Association for Computing Machinery, 2023: 1177-1181.

[2] QI T, CHEN L, LI G, et al. FedAGCN: A traffic flow prediction framework based on federated learning and Asynchronous Graph Convolutional Network [J]. Applied Soft Computing, 2023, 138: 110175.

[3] MEI G, GUO Z, LIU S, et al. SGNN: A Graph Neural Network Based Federated Learning Approach by Hiding Structure[C]// 2019 IEEE International Conference on Big Data (Big Data). 2019: 2560-2568.

[4] GUAN Z L, DU J P, XUE Z, et al. Personalized Public Safety Emergency Detection Method Based on Enhanced Federated Graph Neural Network [J]. Journal of Software, 2024, 35(4): 1774-1789.

[5] ZHU W, LUO J, WHITE A D. Federated learning of molecular properties with graph neural networks in a heterogeneous setting[J]. Patterns, 2022, 3(6): 100521.

[6] HE C, BALASUBRAMANIAN K, CEYANI E, et al. FedGraphNN: A Federated Learning System and Benchmark for Graph Neural Networks[J]. arXiv: 2104.07145, 2021.

[7] LIU R, XING P, DENG Z, et al. Federated Graph Neural Networks: Overview, Techniques and Challenges[J]. arXiv: 2202.07256, 2024.

[8] WU Z, PAN S, CHEN F, et al. A Comprehensive Survey on Graph Neural Networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1): 4-24.

[9] KAIROUZ P, MCMAHAN H B, AVENT B, et al. Advances and Open Problems in Federated Learning[M]// Now Foundations and Trends. 2021: 1-210.

[10] ZHANG H, SHEN T, WU F, et al. Federated Graph Learning -- A Position Paper[J]. arXiv: 2105.11099 2021.

[11] ZHAO G, HUANG Y, TSAI C H. FedGSL: Federated Graph Structure Learning for Local Subgraph Augmentation [C]// 2022 IEEE International Conference on Big Data (Big Data). 2022: 818-824.

[12] CHEN S, XUE D, CHUAI G, et al. FL-QSAR: a federated learning-based QSAR prototype for collaborative drug discovery [J]. Bioinformatics (Oxford, England), 2021, 36(22-23): 5492-5498.

[13] CHEN M, ZHANG W, YUAN Z, et al. FedE: Embedding Knowledge Graphs in Federated Setting[C]// Proceedings of the 10th International Joint Conference on Knowledge Graphs. New York, NY, USA: Association for Computing Machinery, 2022: 80-88.

[14] PENG H, LI H, SONG Y, et al. Differentially Private Federated Knowledge Graphs Embedding[C]// Proceedings of the 30th ACM International Conference on Information & Knowledge Management. New York, NY, USA: Association for Computing Machinery, 2021: 1416-1425.

[15] WU C, WU F, LYU L, et al. FedGNN: A federated graph neural network framework for privacy-preserving personalization[J]. Nature Communications, 2022, 13(1): 3091.

[16] BAEK J, JEONG W, JIN J, et al. Personalized Subgraph Federated Learning[J]. arXiv: 2206.10206, 2023.

[17] ZHANG K, YANG C, LI X, et al. Subgraph Federated Learning with Missing Neighbor Generation [J]. arXiv: 2106.13430, 2021.

[18] PENG L, WANG N, DVORNEK N, et al. FedNI: Federated Graph Learning With Network Inpainting for Population-Based Disease Prediction[J]. IEEE Transactions on Medical Imaging, 2023, 42(7): 2032-2043.

[19] XIE H, XIONG L, YANG C. Federated Node Classification over Graphs with Latent Link-type Heterogeneity[C]// Proceedings of the ACM Web Conference 2023. Austin TX USA: ACM, 2023: 556-566.

[20] BONAWITZ K, IVANOV V, KREUTER B, et al. Practical Secure Aggregation for Privacy-Preserving Machine Learning [C]// Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. New York, NY, USA: Association for Computing Machinery, 2017: 1175-1191.

[21] ACAR A, AKSU H, ULUAGAC A S, et al. A Survey on Homomorphic Encryption Schemes: Theory and Implementation[J]. ACM Computing Surveys, 2018, 51(4): 79: 1-79: 35.

[22] GEISLER S, ZÜGNER D, GÜNNEMANN S. Reliable Graph Neural Networks via Robust Aggregation [C]// Advances in Neural Information Processing Systems. 2020: 13272-13284.

[23] CHEN C, ZHOU J, ZHENG L, et al. Vertically Federated Graph Neural Network for Privacy-Preserving Node Classification[J]. arXiv: 2005.11903, 2022.

[24] JIANG M, JUNG T, KARL R, et al. Federated Dynamic GNN with Secure Aggregation[J]. arXiv: 2009.07351, 2022.

[25] GÜRLER Z, REKIK I. Federated Brain Graph Evolution Prediction Using Decentralized Connectivity Datasets With Temporally-Varying Acquisitions[J]. IEEE Transactions on Medical Imaging, 2023, 42(7): 2022-2031.

[26] CHEN C, HU W, XU Z, et al. FedGL: Federated Graph Learning Framework with Global Self-Supervision[J]. arXiv: 2105.03170, 2021.

[27] MAI P, PANG Y. Vertical Federated Graph Neural Network for

Recommender System[J]. arXiv:2303.05786,2021.
[28] NI X,XU X,LYU L,et al. A Vertical Federated Learning Framework for Graph Convolutional Network[J]. arXiv:2106.11593,2021.
[29] ZHANG X,YIN W,HONG M,et al. Hybrid federated learning: Algorithms and implementation[J]. arXiv:2012.12420,2020.
[30] GAO H,GE S,CHANG T H. FedHD:Communication-efficient federated learning from hybrid data[J]. Journal of the Franklin Institute,2023,360(12):8416-8454.
[31] JANG J,KLABJAN D,MENDIRATTA V,et al. Hybrid FedGraph: An efficient hybrid federated learning algorithm using graph convolutional neural network[J]. arXiv: 2404.09443,2024.
[32] RIZK E,SAYED A H. A Graph Federated Architecture with Privacy Preserving Learning[C]// 2021 IEEE 22nd International Workshop on Signal Processing Advances in Wireless Communications(SPAWC). 2021:131-135.
[33] MCMAHAN B,MOORE E,RAMAGE D,et al. Communication-efficient learning of deep networks from decentralized data [C]// Artificial Intelligence and Statistics. PMLR,2017:1273-1282.
[34] HU K,WU J,LI Y,et al. FedGCN:Federated Learning-Based Graph Convolutional Networks for Non-Euclidean Spatial Data [J]. Mathematics,2022,10(6):1000.
[35] DINH C T,VU T T,TRAN N H,et al. A New Look and Convergence Rate of Federated Multi-Task Learning with Laplacian Regularization[J]. arXiv:2102.07148,2022.
[36] PEI Y,MAO R,LIU Y,et al. Decentralized federated graph neural networks[C]// International Workshop on Federated and Transfer Learning for Data Sparsity and Confidentiality in Conjunction with IJCAI. 2021.
[37] LI X,WU Z,ZHANG W,et al. FedGTA:Topology-aware Averaging for Federated Graph Learning[J]. arXiv: 2401.11755, 2024.
[38] DENG Z,HUANG X,LI D,et al. FedGraph:an Aggregation Method from Graph Perspective[J]. arXiv:2210.02733,2022.
[39] LALITHA A,KILINC O C,JAVIDI T,et al. Peer-to-peer Federated Learning on Graphs[J]. arXiv:2210.02733,2022.
[40] HE C,CEYANI E,BALASUBRAMANIAN K,et al. SpreadGNN:Decentralized Multi-Task Federated Learning for Graph Neural Networks on Molecular Data[J]. Proceedings of the AAAI Conference on Artificial Intelligence,2022,36(6):6865-6873.
[41] CHOI B,YONG S J,HAN D J,et al. Communication-Computation Efficient Secure Aggregation for Federated Learning[J]. arXiv:2012.05433,2021.
[42] GOGINENI V C,WERNER S,HUANG Y F,et al. Decentralized Graph Federated Multitask Learning for Streaming Data [C]// 2022 56th Annual Conference on Information Sciences and Systems(CISS). 2022:101-106.
[43] MALINOVSKIY G,KOVALEV D,GASANOV E,et al. From local SGD to local fixed-point methods for federated learning [C]// International Conference on Machine Learning. PMLR, 2020:6692-6701.
[44] CHEN F,LI P,MIYAZAKI T,et al. FedGraph:Federated Graph Learning with Intelligent Sampling[J]. arXiv: 2111.01370,2021.
[45] YAO Y,JIN W,RAVI S,et al. FedGCN:Convergence-Communication Tradeoffs in Federated Training of Graph Convolutional Networks[J]. arXiv:22041.12433,2022.
[46] ZHANG C,ZHANG S,YU J J Q,et al. FASTGNN:A Topological Information Protected Federated Learning Approach for Traffic Speed Forecasting[J]. IEEE Transactions on Industrial Informatics,2021,17(12):8464-8474.
[47] MENG C,RAMBHATLA S,LIU Y. Cross-Node Federated Graph Neural Network for Spatio-Temporal Data Modeling[J]. arXiv:2160.05223,2021.
[48] CHEN F,LONG G,WU Z,et al. Personalized Federated Learning With Graph[J]. arXiv:2203.00829,2022.
[49] XING P,LU S,WU L,et al. BiG-Fed:Bilevel Optimization Enhanced Graph-Aided Federated Learning[EB/OL]. https://flicml.github.io/2021/papers/FL-ICML21_paper_74.pdf.
[50] LEE H,BERTOZZI A L,KOVAČEVIĆ J,et al. Privacy-Preserving Federated Multi-Task Linear Regression: A One-Shot Linear Mixing Approach Inspired By Graph Regularization [C]// ICASSP 2022 — 2022 IEEE International Conference on Acoustics,Speech and Signal Processing(ICASSP). 2022:5947-5951.
[51] ZHU H,XU J,LIU S,et al. Federated Learning on Non-IID Data:A Survey[J]. arXiv:2106.06843,2022.
[52] XIE H,MA J,XIONG L,et al. Federated Graph Classification over Non-IID Graphs[C]// Advances in Neural Information Processing System. 2021:18839-18852.
[53] ZHENG L,ZHOU J,CHEN C,et al. ASFGNN:Automated Separated-Federated Graph Neural Network[J]. arXiv: 2011.03248,2020.
[54] HE H,BAI Y,GARCIA E A,et al. ADASYN:Adaptive synthetic sampling approach for imbalanced learning[C]// 2008 IEEE International Joint Conference on Neural Networks(IEEE World Congress on Computational Intelligence). 2008: 1322-1328.
[55] LIN Y,CHEN C,CHEN C,et al. Improving Federated Relational Data Modeling via Basis Alignment and Weight Penalty [J]. arXiv:2011.11369,2020.
[56] WANG B,LI A,LI H,et al. GraphFL:A Federated Learning Framework for Semi-Supervised Node Classification on Graphs [J]. arXiv:2012.04187,2020.
[57] TAN Y,LIU Y,LONG G,et al. Federated Learning on Non-IID Graphs via Structural Knowledge Sharing[J]. Proceedings of the AAAI Conference on Artificial Intelligence,2023,37(8):9953-9961.
[58] HANZELY F,RICHTÁRIK P. Federated Learning of a Mixture of Global and Local Models[J]. arXiv:2002.05516,2021.
[59] DINH C,TRAN N,NGUYEN J. Personalized Federated Lear-

ning with Moreau Envelopes[C]// Advances in Neural Information Processing Systems. 2020：21394-21405.

[60] ARIVAZHAGAN M G，AGGARWAL V，SINGH A K，et al. Federated Learning with Personalization Layers[J]. arXiv：1912.00818，2019.

[61] SATTLER F，MÜLLER K R，SAMEK W. Clustered Federated Learning：Model-Agnostic Distributed Multi-Task Optimization under Privacy Constraints[J]. arXiv：1910.01991，2019.

[62] ZHANG K，XIE H，GU Z，et al. Subgraph federated learning over heterogeneous graphs[J]. arXiv：2106.13430，2022.

[63] WEBER M，CHEN J，SUZUMURA T，et al. Scalable Graph Learning for Anti-Money Laundering：A First Look[J]. arXiv：1812.00076，2018.

[64] DU H，SHEN M，SUN R，et al. Malicious Transaction Identification in Digital Currency via Federated Graph Deep Learning[C]// IEEE Conference on Computer Communications Workshops. 2022：1-6.

[65] CHEN Z，LI W，XING X，et al. Medical federated learning with joint graph purification for noisy label learning[J]. Medical Image Analysis，2023，90：102976.

[66] AHMED U，LIN J C W，SRIVASTAVA G. Hyper-Graph Attention Based Federated Learning Methods for Use in Mental Health Detection[J]. IEEE Journal of Biomedical and Health Informatics，2023，27(2)：768-777.

[67] LIU L，TIAN Y，CHAKRABORTY C，et al. Multilevel Federated Learning-Based Intelligent Traffic Flow Forecasting for Transportation Network Management[J]. IEEE Transactions on Network and Service Management，2023，20(2)：1446-1458.

[68] CHEN J，HUANG G，ZHENG H，et al. Graph-Fraudster：Adversarial Attacks on Graph Neural Network-Based Vertical Federated Learning[J]. IEEE Transactions on Computational Social Systems，2023，10(2)：492-506.

[69] LI R C，ZHENG H B，ZHAO W H，et al. Data Reconstruction Attack for Vertical Graph Federated Learning[J]. Computer Science，2023，50(7)：332-338.

WANG Xin，born in 1984，Ph.D，associate professor，master supervisor，is a member of CCF（No. 11687M）. His main research interests include machine learning，big data analysis and federated learning.

（责任编辑：何杨）

CCF NOI-Pre 公益项目举办走进学校公益讲座系列活动

2024 年 12 月 14 日，CCF NOI-Pre 公益项目走进云南红河哈尼族彝族自治州，在云南省蒙自市第一高级中学凤凰校区成功举办了青少年信息学奥赛公益课的专题讲座。CCF 常务理事、NOI 主席助理、NOI-Pre 主席李轩涯博士，蒙自市第一高级中学党委书记、校长吴仕君，蒙自市第一高级中学科创竞赛中心主任邓凯维，清华大学江禹墨等嘉宾，以及近 500 位中小学家长及学生参与了此次活动。

CCF NOI-Pre 主席李轩涯博士在专题讲座上以《青少年信息学学习的一些想法及思考》为题，从为什么要学信息学，信息学竞赛介绍及特点，学习信息学的考虑因素三方面入手，为学生的培养和生涯规划做了介绍。最后，重点详细介绍了 CCF NOI-Pre 公益项目的宗旨和内容，指出虽然参与信息学学习的中小学生日益增多，但教育资源的地域差异却愈发显著。为此，CCF 特别推出了 NOI-Pre 网上公益课，旨在为广大对信息学充满热情的 18 岁及以下中小学生提供高质量的教学资源和学习机会。

蒙自市一中党委书记、校长吴仕君，科创竞赛中心主任邓凯维向现场家长及学生也做了精彩的报告，并对学校情况进行了详实的介绍。蒙自市第一高级中学作为红河州信息学教育的领军者，一直以来都高度重视学生创新思维的培养。通过周末公益课，假期公益冬令营、夏令营的形式，让更多学生了解计算机，了解编程。在保证普及面的同时，选拔优秀学生组建了从小学到初中的梯队。后续，学校也将号召更多的孩子参与 CCF NOI-Pre 公益课的学习，让优质的教学资源普惠更多的学生。

CCF NOI-Pre 公益课项目于 2024 年 5 月 12 日正式发布，已运行半年，目前在全国有近 3000 名学员。项目以线上网络平台为依托，通过“培训课程＋上机练习”的方式，为中小学生打造了一个全方位、多层次的学习平台。所有课程及上机练习均免费开放，并支持回看功能，让学员们可以根据自己的需求随时随地进行学习。课程内容涵盖了直播、录播课程等多种形式，为学员们提供了丰富多样的学习体验。

CCF NOI-Pre 项目希望通过此次专题讲座的举办，能够激发更多中小学生对信息学的兴趣和热情，为他们提供优质的教学资源和学习机会。同时，项目组也将继续加强与各中小学的合作与交流，尤其是中西部学校，举办系列线下讲座，共同推动信息学的普及和发展，为培养更多具有创新精神和实践能力的科技人才贡献力量。

据 CCF 微信公众号