

非局部图神经网络

匿名作者
双盲评审论文

抽象的

现代图神经网络 (GNN) 通过多层局部聚合学习节点嵌入,并在分类图上的应用中取得了巨大成功。然而,非分类图上的任务通常需要非局部聚合。此外,我们发现局部聚合甚至对一些分类图有害。在这项工作中,我们提出了一个简单而有效的非本地聚合框架,该框架具有针对 GNN 的有效注意力引导排序。在此基础上,我们开发了各种非本地 GNN。我们进行了彻底的实验来分析异类图数据集并评估我们的非本地 GNN。实验结果表明,就模型性能和效率而言,我们的非局部 GNN 在六个非分类图基准数据集上明显优于以前的最先进方法。

1简介

图神经网络 (GNN) 处理图并将每个节点映射到嵌入向量 (Zhang 等人,2018b;Wu 等人,2019)。这些节点嵌入可以直接用于节点级应用,例如节点分类 (Kipf & Welling, 2017) 和链路预测 (Schutt et al., 2017)。

此外,它们可用于通过图池学习图表示向量 (Ying et al., 2018; Zhang et al., 2018a; Lee et al., 2019; Yuan & Ji, 2020),以拟合图级任务 (Yanardag & Vishwanathan,2015)。已经提出了 GNN 的许多变体,例如 ChebNets (Defferrard et al., 2016)、GCNs (Kipf & Welling, 2017)、GraphSAGE (Hamilton et al., 2017)、GATs (Velickovic et al., 2018)、LGCN (Gao et al., 2018) 和 GINs (Xu et al., 2019)。它们的优势已在各种图形数据集和任务中得到体现 (Errica 等人,2020)。然而,这些 GNN 共享一个多层局部聚合框架,类似于图像和文本等网格状数据上的卷积神经网络 (CNN) (LeCun et al., 1998)。

近年来,非局部聚合的重要性在计算机视觉 (Wang et al., 2018; 2020)和自然语言处理 (Vaswani et al., 2017)领域的许多应用中得到了证明。特别是,注意力机制已被广泛探索以实现非本地聚合并从远处捕获远程依赖。基本上,注意力机制测量每对位置之间的相似性,并使信息能够在相距遥远但相似的位置之间进行交流。在图方面,非局部聚合对于非分类图也至关重要,而之前对 GNN 的研究主要集中在分类图数据集上 (第 2.2 节)。此外,我们发现局部聚合甚至对某些异类图有害 (第 4.3 节)。最近提出的 Geom-GCN (Pei 等人,2020 年)探索了非分类图中捕获长程依赖性。它包含一个类似注意力的步骤,计算每对节点之间的欧几里得距离。然而,这一步对于大规模图来说在计算上是禁止的,因为计算复杂度是节点数量的二次方。此外,Geom-GCN 采用预训练节点嵌入 (Tenenbaum 等人,2000 年;Nickel & Kiela, 2017 年;Ribeiro 等人,2017 年),这些嵌入不是特定于任务的,限制了有效性和灵活性。

在这项工作中,我们为 GNN 提出了一个简单而有效的非局部聚合框架。该框架的核心是一个有效的注意力引导排序,它通过一般深度学习中的经典本地聚合运算符实现非本地聚合。所提出的框架可以灵活地用于以低成本增强通用 GNN。基于框架工作,我们构建了各种高效的非本地 GNN。此外,我们对现有的非分类图数据集进行了详细分析,并相应地应用了不同的非局部 GNN。实验结果表明,我们的非局部 GNN 在六个非分类图基准数据集上的节点分类任务上明显优于以前的最新方法。

2背景及相关工作

2.1图神经网络

我们专注于通过图神经网络 (GNN) 学习每个节点的嵌入向量。
大多数现有的 GNN 都受到卷积神经网络 (CNN) 的启发 (LeCun 等人,1998 年)并遵循局部聚合框架。一般来说,GNN 的每一层都会扫描图中的每个节点,并从直接连接的节点 (即 1 跳邻居)聚合本地信息。

具体来说,GNN 的一个公共层执行类似于深度可分离卷积 (Chollet, 2017) 的两步处理:空间聚合和特征变换。第一步使用空间相邻节点的嵌入向量更新每个节点嵌入。例如,GCNs (Kipf & Welling, 2017) 和 GATs (Velickovic et al., 2018) 计算 1-hop 邻域内节点嵌入的加权和,其中权重来自节点的程度和节点之间的交互,分别。GraphSAGE (Hamilton 等人,2017 年)应用最大池化,而 GINs (Xu 等人,2019 年)只是对节点嵌入求和。特征变换步骤类似于 1×1 卷积,其中每个节点嵌入向量通过共享线性变换映射到一个新的特征空间 (Kipf & Welling, 2017; Hamilton et al., 2017; Velickovic et al., 2018) 或多层感知器 (MLP) (Xu et al., 2019)。与这些研究不同,LGCN (Gao et al., 2018) 探索通过 top-k 排名直接应用常规卷积。

然而,这些 GNN 的每一层都只聚合 1-hop 邻域内的局部信息。虽然堆叠多层理论上可以实现跨多跳邻域的节点之间的通信,但聚合本质上是本地的。此外,深度 GNN 通常会遇到过度平滑问题 (Xu et al., 2018; Li et al., 2018; Chen et al., 2020) 。

2.2分类图和分类图

文献中的图有很多种,比如引文网络 (Kipf & Welling, 2017)、社区网络 (Chen et al., 2020)、共现网络 (Tang et al., 2009)、网页链接网络等 (Rozemberczki 等人,2019 年)。我们专注于与节点分类任务相对应的图数据集。特别是,我们根据标签方面的节点同质性,即具有相同标签的节点在图中彼此靠近的可能性,将图数据集分为分类和非分类数据集 (Newman, 2002; Ribeiro et al., 2017) 。

分类图是指那些具有高节点同质性的图。常见的分类图数据集是引文网络和社区网络。另一方面,非分类图数据集中的图包含更多具有相同标签但彼此远离的节点。示例异类图数据集是共现网络和网页链接网络。

如上所述,大多数现有 GNN 仅执行局部聚合并在分类图上取得良好性能 (Kipf & Welling,2017;Hamilton 等人,2017;Velickovic 等人,2018;Gao 等人,2018)。然而,它们可能在异类图上失败,其中同一类中的信息节点往往不在局部多跳邻域内,需要非局部聚合。

因此,在这项工作中,我们探索了非本地 GNN。

2.3注意力机制

注意机制 (Vaswani et al., 2017) 已广泛用于 GNNs (Velickovic et al., 2018; Gao & Ji, 2019; Knyazev et al., 2019) 以及其他深度学习模型 (Yang et al., 2016 年;Wang 等人,2018 年;2020 年)。典型的注意力机制以三组向量作为输入,即查询向量 q ,键向量 (k_1, k_2, \dots, k_n),值向量 (v_1, v_2, \dots, v_n)。请注意,键和值向量具有一一对应关系,有时可能相同。注意机制计算输出向量 o 为

$$a_i = \text{ATTEND}(q, k_i) \in \mathbb{R}, i = 1, 2, \dots, n; \text{ 名词}; o = \sum_{i=1}^n a_i v_i \quad \text{艾夫}, \tag{1}$$

其中ATTEND(\cdot)函数可以是 从 q 和 k_i 之间的交互输出标量注意力分数 a_i 的任何函数,例如点积 (Gao & Ji, 2019) 甚至是神经网络

工作 (Velicković 等人, 2018 年)。三组输入向量的定义取决于模型和应用程序。

值得注意的是, 现有的 GNN 通常使用注意力机制进行局部聚合 (Velicković et al., 2018; Gao & Ji, 2019)。具体来说, 当聚合节点 v 的信息时, 查询向量是 v 的嵌入向量, 而键和值向量来自 v 的直接连接节点的节点嵌入。并且针对每个 $v \in V$ 迭代该过程。

值得注意的是, 注意力机制可以很容易地扩展到非局部聚合 (Wang et al., 2018; 2020), 通过在为每个节点聚合信息时让键和值向量对应于图中的所有节点。然而, 在给定大规模图的情况下, 它在计算上是令人望而却步的, 因为对 n 个节点的图中的每个节点进行迭代需要 $O(n^2)$ 聚合方法, 而该聚合方法只需要 $O(n \log n)$ 时间。

$O(n)$ 时间。在这项工作中, 我们提出了一种新颖的非本地

3 提议的方法

3.1 具有注意力引导排序的非本地聚合

我们考虑一个图 $G = (V, E)$, 其中 V 是节点集, E 是边集。每条边 $e \in E$ 连接两个节点, 使得 $E \subseteq V \times V$ 。

每个节点 $v \in V$ 都有一个节点特征向量 $x_v \in \mathbb{R}^d$ 。 v 的 k 跳邻域是指在 k 条边内可以到达 v 的节点集合 $N_k(v)$ 。例如, v 的直接连接节点的集合是它的 1 跳邻域 $N_1(v)$ 。

我们提出的非局部聚合框架由三个步骤组成, 即局部嵌入、注意力引导排序和非局部聚合。下面, 我们一一描述。

局部嵌入: 我们提出的框架建立在局部嵌入步骤之上, 该步骤从节点特征向量中提取局部节点嵌入。局部嵌入步骤可以像 $z_v = \text{MLP}(x_v) \in \mathbb{R}^f$ 这样简单

$$, \forall v \in V.$$

$\text{MLP}(\cdot)$ 函数是一个多层感知器 (MLP), f 是局部节点嵌入 z_v 的维数。请注意, $\text{MLP}(\cdot)$ 函数在图中的所有节点之间共享。应用 MLP 仅考虑节点本身, 而无需聚合来自邻居的信息。如第 4.3 节所示, 此属性对于某些异配图非常重要。

另一方面, 图神经网络 (GNN) 也可以用作局部嵌入步骤, 因此我们提出的框架可以很容易地用于扩充现有的 GNN。如 2.1 节所述, 现代 GNN 执行多层局部聚合。通常, 对于每个节点, GNN 的一层会聚合来自其 1-hop 邻居的信息。堆叠 L 个这样的本地聚合层允许每个节点访问 L 跳远的信息。具体来说, L 层 GNN 的第 l 层 ($l = 1, 2, \dots, L$) 可以描述为

$$z_v^{(l)} = \text{TRANSFORM}() \text{ AGGREGATE}() \{z_u^{(l-1)} : u \in N_1(v) \cup \{v\}\} \in \mathbb{R}^f, \forall v \in V, \quad (3)$$

其中 $z_v^{(0)}$ 是第 2.1 节中介绍的空间聚合 x_v 表示局部节点嵌入。 $\text{AGGREGATE}()(\cdot) = x_v$, $z_v = z$ 和 $\text{TRANSFORM}()(\cdot)$ 函数分别表示和特征转换步骤。通过上述框架, GNN 可以从局部邻域内的节点捕获节点特征信息以及结构信息。

当使用 MLP 或 GNN 作为局部嵌入步骤时, 局部节点嵌入 z_v 仅包含节点 v 的局部信息。但是, z_v 可用于指导非局部聚合, 因为距离远但信息丰富的节点可能具有相似的节点特征和局部结构。基于这种直觉, 我们提出了注意力引导排序来实现非局部聚合。

注意力引导排序: 注意力引导排序的基本思想是学习节点的排序, 其中距离较远但信息丰富的节点彼此靠近。具体来说, 给定通过局部嵌入步骤获得的局部节点嵌入 z_v , 我们通过以下方式计算一组注意力分数

$$a_v = \text{ATTEND}(c, z_v) \in \mathbb{R}, \forall v \in V, \quad (4)$$

其中 c 是在训练期间随机初始化和联合学习的校准向量 (Yang et al., 2016)。在这个注意力运算符中, c 作为查询向量, z_v 是关键向量。在

此外,我们还将 z_v 视为值向量。然而,与 2.3 节中介绍的注意力机制不同,我们使用注意力分数对值向量进行排序,而不是计算加权和来聚合它们。请注意,最初在图中的节点之间没有排序。具体而言,由于 a_v 和 z_v 通过等式 (4) 具有一一对应关系,将注意力分数按非递减顺序排序为 (a_1, a_2, \dots, a_n) 提供了节点之间的排序,其中 $n = |V|$ 是图中的节点数。局部节点嵌入的结果序列可以表示为 (z_1, z_2, \dots, z_n) 。

式 (4) 中的注意力过程也可以理解为局部节点嵌入到一维空间的投影。投影取决于具体的 $\text{ATTEND}(\cdot)$ 函数和校准向量 c 。顾名思义,校准向量 c 用于校准一维空间,以便在该空间中将距离较远但信息丰富的节点推近。这个目标是通过以下非局部聚合步骤和校准向量 c 的训练来实现的,如下所示。

非局部聚合:我们指出,通过注意力引导排序,非局部聚合可以通过深度学习中最常见的局部聚合算子卷积来实现。具体来说,给定局部节点嵌入的排序序列 (z_1, z_2, \dots, z_n) ,我们计算

$$(\hat{z}_1, \hat{z}_2, \dots, \hat{z}_n) = \text{CONV}(z_1, z_2, \dots, z_n), \quad (5)$$

其中 $\text{CONV}(\cdot)$ 函数表示具有适当填充的一维卷积。请注意,只要输入和输出向量的数量保持不变, $\text{CONV}(\cdot)$ 函数就可以用一维卷积神经网络代替。

为了解 $\text{CONV}(\cdot)$ 函数如何通过注意力引导排序执行非局部聚合,我们举一个例子,其中 $\text{CONV}(\cdot)$ 函数是内核大小为 $2s + 1$ 的一维卷积。在这种情况下,计算 \hat{z}_i 来自 $(z_{i+s}, \dots, z_{i-s})$, 对应于 $\text{CONV}(\cdot)$ 函数的感受野。因此,如果注意力引导排序导致 $(z_{i+s}, \dots, z_{i-s})$ 包含距离 z_i 较远但信息丰富的节点,则输出 \hat{z}_i 会聚合非局部信息。另一种观点是,我们可以将注意力引导排序视为重新连接图中的节点,其中 $(z_{i+s}, \dots, z_{i-s})$ 可以视为 z_i 的 1-hop 邻域。

在 $\text{CONV}(\cdot)$ 函数之后,将 z_i 和 \hat{z}_i 连接起来作为分类器的输入,以预测相应节点的标签,其中可以捕获非本地和本地依赖项。为了能够对校准向量 c 进行端到端的训练,我们将等式 (5) 修改为

$$(\hat{z}_1, \hat{z}_2, \dots, \hat{z}_n) = \text{CONV}(a_1 z_1, a_2 z_2, \dots, a_n z_n), \quad (6)$$

我们将注意力分数乘以相应的局部节点嵌入。因此,校准向量 c 在训练期间通过注意力分数接收梯度。

剩下的问题是如何确保注意力引导的排序将距离较远但信息丰富的节点推到一起。简短的回答是没有必要保证这一点,因为非局部聚合的要求取决于具体的图。事实上,我们提出的框架赋予 GNN 非局部聚合的能力,但让端到端的训练过程决定是否使用非局部信息。来自监督损失的反向传播将调整校准向量 c 并鼓励 z_i 捕获未由 z_i 编码的有用信息。

在非分类图的情况下, z_i 通常需要聚合来自遥远但信息丰富的节点的信息。因此,校准向量 c 倾向于安排注意力引导的排序,以将距离较远但信息丰富的节点放在一起,如第 4.5 节中的实验所示。另一方面,本地邻域内的节点通常比分类图中的远距离节点提供更多信息。在这种情况下, z_i 可以简单地执行类似于 GNN 的局部聚合。

在第 4 节中,我们展示了我们提出的非局部聚合框架在六个异类图数据集上的有效性。特别是,我们在所有数据集上实现了最先进的性能,并且比以前有了显著改进。

3.2 时间复杂度分析

我们对我们提出的框架的时间复杂度进行了理论分析。如 2.3 节所述,使用注意力机制 (Vaswani 等人,2017 年; Wang 等人,2018 年;2020 年) 实现非局部聚合需要 $O(n)$ 个节点图的时间。本质上, $O(n)$

时间复杂度是由于需要在每对节点之间计算 $\text{ATTEND}(\cdot)$ 函数。特别是,最近提出的 Geom-GCN (Pei et al., 2020) 包含类似的非局部聚合步骤。对于每个 $v \in V$, Geom-GCN 找到与 v 的欧几里得距离小于预定义数字的节点集,其中需要计算每对节点之间的欧几里得距离。由于计算两个节点之间的欧氏距离可以理解为 $\text{ATTEND}(\cdot)$ 函数, Geom-GCN 至少有 $O(n^2)$ 的时间复杂度。

29

相比之下,我们提出的非本地聚合框架只需要 $O(n \log n)$ 时间。要看到这一点,请注意方程 (4) 中的 $\text{ATTEND}(\cdot)$ 函数只需要计算一次,而不是为每个节点迭代它。因此,计算注意力分数只需要 $O(n)$ 的时间。

因此,排序的时间复杂度,即 $O(n \log n)$,决定了我们提出的框架的总时间复杂度。在 4.6 节中,我们比较了普通 GNN、Geom-GCN 和下一节介绍的非局部 GNN 在不同数据集上的实际运行时间。

3.3 高效的非局部图神经网络

我们应用我们提出的非局部聚合框架来构建高效的非局部 GNN。回想一下,我们提出的框架从局部嵌入步骤开始,然后是注意力引导排序和非局部聚合步骤。

特别是,局部嵌入步骤可以通过 MLP 或常见的 GNN 实现,例如 GCN (Kipf & Welling, 2017) 或 GAT (Velickovic et al., 2018)。MLP 仅从节点特征向量中提取局部节点嵌入,并排除来自局部邻域内节点的信息。此属性对某些异类图很有帮助,其中局部邻域内的节点提供的噪声多于有用信息。在其他非分类图中,信息节点位于本地邻域和远距离位置。在这种情况下,GNN 更适合作为局部嵌入步骤。根据手头的分类图,我们使用 MLP 或 GNN 构建不同的非局部 GNN 作为局部嵌入步骤。在第 4.3 节中,我们表明可以通过简单的实验来区分这两类异类图,我们相应地应用不同的非局部 GNN。具体来说,MLP 和 GNN 的层数都设置为 2。

在注意力引导排序方面,我们只需要在等式 (4) 中指定 $\text{ATTEND}(\cdot)$ 函数。为了使其尽可能高效,我们选择 $\text{ATTEND}(\cdot)$ 函数作为

$$\text{ATTEND}(c, z_v) = c \cdot \frac{\exp(z_v^T z_v)}{\sum_{v \in V} \exp(z_v^T z_v)}, \quad z_v \in \mathbb{R}^d, \forall v \in V, \quad (7)$$

其中 c 是训练参数的一部分,如第 3.1 节所述。

借助注意力引导排序,我们可以通过卷积实现非局部聚合步骤,如第 3.1 节所述和等式 (6) 所示。具体来说, $\text{CONV}(\cdot)$ 函数被设置为由两个一维卷积组成的 2 层卷积神经网络。内核大小设置为 3 或 5,具体取决于数据集。激活函数是 ReLU (Krizhevsky et al., 2012)。

最后,我们使用一个线性分类器,它将 z_i 和 z_i 的串联作为输入,并对相应的节点进行预测。根据局部嵌入步骤,我们构建了三个高效的非局部 GNN,即非局部 MLP (NLMLP)、非局部 GCN (NLGCN) 和非局部 GAT (NLGAT)。可以使用分类损失对模型进行端到端训练。

4 个实验

4.1 数据集

我们对六个分类图数据集 (Rozemberczki 等人, 2019 年; Tang 等人, 2009 年; Pei 等人, 2020 年) (Chameleon, Squirrel, Actor, Cornell, Texas, Wisconsin) 和三个分类图数据集 (Sen 等人, 2008 年) (Cora, Citeseer, Pubmed)。这些数据集通常用于评估 GNN 在节点分类任务上的表现 (Kipf & Welling, 2017 年; Velickovic et al., 2018 年; Gao 等人, 2018 年; Pei 等人, 2020 年)。我们在附录 A.1 中提供了异类图数据集的详细描述。为了区分分类图数据集和分类图数据集, Pei 等人。 (2020)

提出一个度量来衡量图 G 的同质性,定义为

高 (克)=
$$\frac{1}{|V|} \sum_{v \in V} \frac{v \text{ 与 } v \text{ 具有相同标签的直连节点数}}{v \text{ 的直连节点数}} \quad (8)$$

直观上,大的 H(G) 表示分类图,反之亦然。表 1 总结了 H(G) 和其他统计数据。

表 1:我们实验中使用的九个数据集的统计数据。H(G) 的定义在第 4.1 节中提供。H(G) 可用于区分分类图数据集和非分类图数据集。

分类数据集				异配			
Cora	Citeseer	Pubmed	Chameleon	Squirrel	Aqtor	Cornell	Texas
Wisconsin	H(G)	0.83	0.71	#Nodes	2708	3327	#Edges
5429	4732	#Features	1433	0.79	0.25	0.22	0.24
3703	#Classes	19717	2277	5201	183	183	251
		44338	36101	217073	33544	295	309
		500	2325	2089	931	5	5
7		3	5	5	5	5	5

在我们的实验中,我们专注于比较模型在异类图数据集上的性能,以证明我们的非局部聚合框架的有效性。提供了分类图数据集上的性能以供参考,表明当不强烈需要非局部聚合时,所提出的框架不会损害性能。

4.2基线

我们将我们提出的非局部 MLP (NLMLP)、非局部 GCN (NLGCN) 和非局部 GAT (NLGAT) 与各种基线进行比较:(1) MLP 是最简单的深度学习模型。它仅基于节点特征向量进行预测,而不聚合任何本地或非本地信息。(2) GCN (Kipf & Welling, 2017) 和 GAT (Velickovi c et al., 2018) 是最常见的 GNN。如 2.1 节所述,它们仅执行本地聚合。(3) Geom-GCN (Pei et al., 2020) 是最近提出的一种可以捕获长程依赖的 GNN。它是几个非分类图数据集上的当前最先进模型。Geom-GCN 需要使用不同的节点嵌入方法,例如 Isomap (Tenenbaum et al., 2000)、Poincare (Nickel & Kiela, 2017) 和 struc2vec (Ribeiro et al., 2017)。我们只是报告 Pei 等人的最佳结果。(2020) 对于 Geom-GCN 和以下两个变体,没有指定节点嵌入方法。(4)

Geom-GCN-g (Pei et al., 2020) 是 Geom-GCN 的一个变体,它只执行局部聚合。它类似于常见的 GNN。(5) Geom-GCN-s (Pei et al., 2020) 是 Geom-GCN 的变体,不强制进行局部聚合。设计的功能类似于我们的 NLMLP。

我们使用 Pytorch (Adam et al., 2017) 和 Pytorch Geometric (Fey & Lenssen, 2019) 实现 MLP、GCN、GAT 和我们的方法。正如在公平环境中所讨论的那样,GCN 和 GAT 的结果与 Pei 等人的结果不同。(2020)。

在每个数据集上,我们遵循 Pei 等人。(2020) 并将每个类的节点随机分成 60%、20% 和 20% 用于训练、验证和测试。实验以不同的随机拆分重复运行 10 次,并报告了这 10 次运行的平均测试准确度。当每次运行的验证准确度达到最大值时,就会执行测试。除了第 3.3 节中指定的细节外,我们为我们提出的模型单独调整以下超参数:(1)隐藏单元的数量 ∈ {16,48,96},(2)辍学率 ∈ {0,0.5,0.8},(3)权重衰减 ∈ {0, 5e-4, 5e-5, 5e-6},(4)学习率 ∈ {0.01, 0.05}。

4.3解配图数据集分析

如第 3.3 节所述,异类图数据集可以分为两类。在属于第一类的非分类图中,局部邻域内的节点提供的噪声多于有用信息。因此,在此类异类图的模型中应避免局部聚合。对于第二类,信息节点位于本地邻域和远距离位置。直观上,具有较低 H(G) 的图更有可能属于第一类。但是,这并不是确定这两个类别的准确方法。

1<https://openreview.net/forum?id=S1e2agrFvS¬elId=8tGKV1oSzCr>

了解分类图的确切类别至关重要,因为我们需要相应地应用非局部 GNN。如上分析,关键区别在于本地聚合是否有效。
因此,我们可以通过比较 MLP 和普通 GNN (GCN,GAT)在六个异配图数据集的每一个上的性能来区分两类异配图。

表 2:MLP 与普通 GNN 的比较。这些分析实验用于确定两类非分类图数据集,如第 4.3 节所述。

数据集	分类			异配					
	科拉	Citeseer	Pubmed	Chameleon	Squirrel	演员	康乃尔	得克萨斯州	威斯康星州
多层板 GCN 洞	76.5±1.3	73.6 ± 1.9	87.5±0.4	48.5±3.0					
	88.2±1.2	75.7±1.3	88.4±0.6	35.1±0.8					
	88.4±1.4	76.1±1.0	87.0±0.3	81.6±6.3					
				48.5±3.0	31.5±1.4	35.1±0.8	81.6±6.3	81.3±7.1	84.9±5.3
				54.9 ± 1.9	30.3±1.6	54.2 ± 7.3	61.1 ± 7.0	59.6±4.5	65.0±3.7
				±2.5	29.4±1.2	56.3±4.3	57.9± 6.1	57.8±4.3	51.3

结果总结在表2中。我们可以看到Actor,Cornell,Texas和Wisconsin属于第一类,而Chameleon和Squirrel属于第二类。我们添加了分类图数据集的性能以供参考,其中局部聚合是有效的,因此 GNN 往往优于 MLP。

4.4与基线的比较

根据第 4.3 节的见解,我们根据非分类图数据集的类别应用不同的非局部 GNN,并与相应的基线进行比较。

具体来说,我们在 Actor,Cornell,Texas 和 Wisconsin 上采用了 NLMLP。
相应的基线是 MLP,Geom-GCN 和 Geom-GCN-s,如表 2 所示,GCN 和 GAT 在这些数据集上的表现比 MLP 差得多。而Geom-GCN-g与GCN类似,性能比Geom-GCN-s差,见附录A.2。比较结果报告在表 3 中。虽然 Geom-GCN-s 是这些数据集上以前最先进的 GNN (Pei 等人,2020),但我们发现 MLP 始终优于 Geom-GCN-s大利润。特别是,虽然 Geom-GCN-s 没有明确执行局部聚合,但它仍然优于 MLP。一个可能的解释是 Geom-GCN-s 使用预训练的节点嵌入,它隐式地聚合了来自局部邻域的信息。

表 3:我们的 NLMLP 与四个异类图数据集上的强基线之间的比较渴望第 4.3 节中定义的第一类。

数据集	演员	康乃尔	得克萨斯州	威斯康星州
MLP	35.1±0.8	81.6±6.3	81.3±7.1	84.9±5.3
Geom-GCN	31.6	Geom-GCN-	60.8	67.6
s 34.6		75.4	73.5	80.4
NLMLP	37.9±1.3	84.9±5.7	85.4±3.8	87.3±4.3

相比之下,我们的 NLMLP 是建立在 MLP 之上的,具有所提出的非局部聚合框架,它排除了局部噪声并从非局部信息节点收集有用信息。
NLMLP 在这些分类图数据集上设置了新的最先进的性能。

在属于第二类异类图数据集的 Chameleon 和 Squirrel 上,我们相应地应用了 NL GCN 和 NLGAT。基线是 GCN,GAT,Geom-GCN 和 Geom-GCN-g。在这些数据集上,这些显式执行局部聚合的基线显示出优于 MLP 和 Geom-GCN-s 的优势,如附录 A.2 所示。如表 4 所示,我们提出的 NLGCN 在两个数据集上都取得了最佳性能。

表 4:我们的 NL GCN,NLGAT 和强基线在属于第 4.3 节中定义的第二类的两个非分类图数据集上的比较。

数据集	变色龙松鼠	
GCN	67.6±2.4	54.9±1.9
洞	65.0±3.7	51.3±2.5
Geom-GCN	60.9	38.1
Geom-GCN-g	68.0	46.0
NLGCN	70.1±2.9	59.0±1.2
NLGAT	65.7±1.4	56.8±2.5

此外,值得注意的是,我们的 NLGCN 和 NL GAT 分别建立在 GCN 和 GAT 之上。他们展示了相对于同行的改进,这表明我们提出的非本地聚合框架的优势对于普通 GNN 是普遍的。

我们在附录 A.2 中提供了所有模型在所有数据集上的结果以供参考。

4.5注意力引导排序分析

我们分析了我们提出的框架中注意力引导排序的结果,以表明我们的非本地 GNN 确实执行了非本地聚合。

假设注意力引导排序导致排序序列 (z_1, z_2, \dots, z_n) ,它通过卷积或 CNN 变成 $(\hat{z}_1, \hat{z}_2, \dots, \hat{z}_n)$ 。正如 3.1 节中所讨论的,我们可以将序列 (z_1, z_2, \dots, z_n) 视为重新连接的图 G^\wedge ,其中我们将 z_i 感受野内的节点视为直接连接到 z_i ,即 z_i 的 1-跳邻居。这个新的 1 跳邻域内的信息将被聚合。如果我们的非局部 GNN 确实执行非局部聚合,则重新连接的图 G^\wedge 的同质性应该大于原始图。因此,我们为每个数据集计算 $H(G^\wedge)$ 以验证此陈述。在第 4.4 节之后,我们将 NLMLP 应用于 Actor、Cornell、Texas 和 Wisconsin,将 NLGCN 应用于 Chameleon 和 Squirrel。

图 1 比较了每个数据集的 $H(G^\wedge)$ 和 $H(G)$ 。我们可以观察到 $H(G^\wedge)$ 比 $H(G)$ 大得多,表明在重新连接的图 G^\wedge 中距离较远但信息丰富的节点彼此靠近。我们还为康奈尔大学和德克萨斯大学提供了排序序列的可视化。我们可以看到具有相同标签的节点倾向于聚集在一起。这些事实表明我们的非本地 GNN 通过注意力引导排序执行非本地聚合。

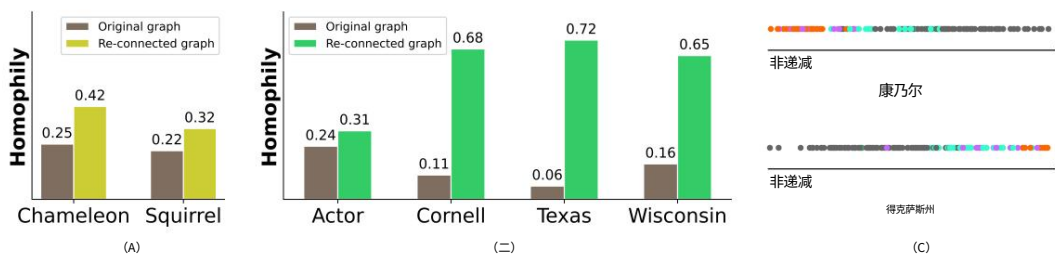


图 1: (a)我们的 NLGCN 在变色龙和松鼠上给出的原始图和重新连接图之间的同质性比较。(b) 我们的 NLMLP 在 Actor、Cornell、Texas 和 Wisconsin 上给出的原始图和重新连接图之间的同质性比较。(c) 康奈尔和得克萨斯州注意力引导排序后排序节点序列的可视化。颜色表示节点标签。详情在第 4.5 节中解释。

4.6 效率比较

正如 3.2 节中分析的那样,我们提出的非局部聚合框架比以前基于原始注意机制的方法更有效,例如 Geom-GCN (Pei 等人, 2020)。具体而言,与 $O(n)$ 相比,我们的方法仅需要 $O(n \log n)$ 计算时间。在本节中,我们通过比较真实的运行时间来验证我们的分析。具体来说,我们将 NLGCN 与 Geom-GCN 以及 GCN 和 GAT 进行了比较。对于 Geom-GCN,我们使用 Pei 等人提供的代码。(2020)。每个模型在每个数据集上训练 500 个时期,并报告每个时期的平均训练时间。

表 5:实际运行时间 (毫秒)方面的比较。

	变色龙	松鼠 22.2
GCN	(1.0×) 14.3 (1.0×)	
GAT	33.2 (1.5×)	163.3 (11.4×)
Geom-GCN	3615.0 (163.1×)	10430.0 (727.3×)
NLGCN	39.6 (2.8×)	26.3 (1.2×)

结果如表 5 所示。虽然我们的 NLGCN 建立在 GCN 之上,但它只比 GCN 慢一点,比 GAT 快一点,显示了我们的非本地聚合框架的效率。另一方面,由于 Geom-GCN 的时间复杂度为 $O(n)$,因此速度明显较慢。

5 结论

在这项工作中,我们为 GNN 提出了一个简单而有效的非局部聚合框架。该框架的核心是一种高效的注意力引导排序,它通过卷积实现非局部聚合。所提出的框架可以很容易地用于构建具有低计算成本的非本地 GNN。我们对节点分类任务进行了彻底的实验,以评估我们提出的方法。特别是,我们通过实验分析现有的异类图数据集,并相应地应用不同的非局部 GNN。结果表明,在准确性和速度方面,我们的非局部 GNN 在六个非分类图基准数据集上明显优于以前的最先进方法。

参考

Paszke Adam, Gross Sam, Chintala Soumith, Chanan Gregory, Yang Edward, D Zachary, Lin Zeming, Desmaison Alban, Antiga Luca 和 Lerer Adam. pytorch 中的自动微分。在神经信息处理系统 Autodiff 研讨会论文集中, 2017 年。

Deli Chen, Yankai Lin, Wei Li, Peng Li, Jie Zhou 和 Xu Sun. 从拓扑的角度衡量和缓解神经网络的过度平滑问题。在第三十四届 AAAI 人工智能会议上, 2020 年。

弗朗索瓦·肖莱。Xception: 具有深度可分离卷积的深度卷积。在 IEEE 计算机视觉和模式识别会议记录中, 第 1251–1258 页, 2017 年。

Michael Defferrard, Xavier Bresson 和 Pierre Vandergheynst. 具有快速局部光谱过滤的图上的卷积神经网络。在神经信息处理系统的进展中, 第 3844–3852 页, 2016 年。

Federico Errica, Marco Podda, Davide Bacciu 和 Alessio Micheli. 用于图分类的图神经网络的公平比较。在 2020 年学习表示国际会议上。

Matthias Fey 和 Jan E. Lenssen. 使用 PyTorch Geometric 快速图形表示学习。在 ICLR 图和流形表示学习研讨会, 2019 年。

红阳高、水王姬。通过硬和通道注意网络进行图形表示学习。在第 25 届 ACM SIGKDD 知识发现与数据挖掘国际会议论文集中, 第 741–749 页, 2019 年。

高红阳, 王正阳, 季水旺。大规模可学习图卷积网络。在第 24 届 ACM SIGKDD 知识发现与数据挖掘国际会议论文集中, 第 1416–1424 页, 2018 年。

Will Hamilton, Zhitao Ying 和 Jure Leskovec. 大图上的归纳表示学习。在神经信息处理系统的进展中, 第 1024–1034 页, 2017 年。

Thomas N. Kipf 和 Max Welling. 使用图卷积网络进行半监督分类。在国际学习表示会议上, 2017 年。

Boris Knyazev, Graham W Taylor 和 Mohamed Amer. 理解神经网络中的注意力和泛化。在神经信息处理系统的进展中, 第 4204–4214 页, 2019 年。

Alex Krizhevsky, Ilya Sutskever 和 Geoffrey E Hinton. 具有深度卷积神经网络的 Imagenet 分类。在神经信息处理系统的进展中, 第 1097–1105 页, 2012 年。

Yann LeCun, Leon Bottou, Yoshua Bengio 和 Patrick Haffner. 基于梯度的学习应用于文件识别。IEEE 会议记录, 86(11):2278–2324, 1998。

Junhyun Lee, Inyeop Lee 和 Jaewoo Kang. 自注意图池。在国际会议上关于机器学习的文章, 第 3734–3743 页, 2019 年。

李启迈、韩志超和吴晓明。更深入地了解用于半监督学习的图卷积网络。在 2018 年第 30 届 AAAI 人工智能会议上。

马克 EJ 纽曼。网络中的分类混合。物理评论快报, 89(20):208701, 2002。

Maximilian Nickel 和 Douwe Kiela. 用于学习层次表示的 Poincare 嵌入。在神经信息处理系统的进展中, 第 6338–6347 页, 2017 年。

Hongbin Pei, Bingzhe Wei, Kevin Chen-Chuan Chang, Yu Lei 和 Bo Yang. Geom-GCN: 几何图形卷积网络。在 2020 年学习表示国际会议上。

Leonardo FR Ribeiro, Pedro HP Saverese 和 Daniel R Figueiredo. struc2vec: 从结构标识中学习节点表示。在第 23 届 ACM SIGKDD 知识发现与数据挖掘国际会议论文集中, 第 385–394 页, 2017 年。

Benedek Rozemberczki, Carl Allen 和 Rik Sarkar. 多尺度属性节点嵌入。arXiv 预印本 arXiv:1909.13021, 2019。

Kristof Schutt, Pieter-Jan Kindermans, Huziel Enoc Saucedo Felix, Stefan Chmiela, Alexandre Tkatchenko 和 Klaus-Robert Muller. Schnet: 一种用于模拟量子相互作用的连续滤波器卷积神经网络。在神经信息处理系统的进展中, 第 991–1001 页, 2017 年。

Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher 和 Tina Eliassi-Rad. 网络数据中的集体分类。人工智能杂志, 29(3):93–93, 2008。

Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. Social influence analysis in large-scale networks. 第 15 届 ACM SIGKDD 知识发现和数据挖掘国际会议论文集, 第 807–816 页, 2009 年。

Joshua B Tenenbaum, Vin De Silva 和 John C Langford. 全球几何框架非线性降维。科学, 290(5500):2319–2323, 2000。

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser 和 Illia Polosukhin. 注意力就是你所需要的。在神经信息处理系统的进展中, 第 5998–6008 页, 2017 年。

Petar Velicković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio 和 Yoshua Bengio. 图注意力网络。在国际学习表示会议上, 2018 年。

Xiaolong Wang, Ross Girshick, Abhinav Gupta 和 Kaiming He. 非局部神经网络。在 IEEE 计算机视觉和模式识别会议论文集中, 第 7794–7803 页, 2018 年。

Zhengyang Wang, Na Zou, Dinggang Shen, and Shuiwang Ji. Non-local U-Nets for biomedical image segmentation. In Thirty-Fourth AAAI Conference on Artificial Intelligence, 2020.

Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S Yu. A comprehensive survey on graph neural networks. arXiv preprint arXiv:1901.00596, 2019.

Keyulu Xu, Chengtao Li, Yonglong Tian, Tomohiro Sonobe, Ken-ichi Kawarabayashi 和 Stefanie Jegelka. 具有跳跃知识网络的图形表示学习。在机器学习国际会议上, 第 5449–5458 页, 2018 年。

Keyulu Xu, Weihua Hu, Jure Leskovec 和 Stefanie Jegelka. 图神经网络有多强大? 在国际学习表示会议上, 2019 年。

Pinar Yanardag 和 SVN Vishwanathan. 深图内核。在第 21 届 ACM SIGKDD 知识发现和数据挖掘国际会议论文集中, 第 1365–1374 页。

美国计算机学会, 2015 年。

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola 和 Eduard Hovy. 用于文档分类的分层注意力网络。在计算语言学协会北美分会 2016 年会议记录中: 人类语言技术, 第 1480–1489 页, 2016 年。

Zhitao Ying, Jiaxuan You, Christopher Morris, Xiang Ren, Will Hamilton 和 Jure Leskovec. Hierarchical graph representation learning with differentiable pooling. 在神经信息处理系统的进展中, 第 4800–4810 页, 2018 年。

浩远和水王姬. StructPool: 通过条件随机场进行结构化图池化。在 2020 年学习表示国际会议上。

张慕涵、崔志成、马里恩·诺伊曼和陈一新。用于图分类的端到端深度学习架构。在第 30 届 AAAI 人工智能会议上,2018a。

Si Zhang、Hanghang Tong、Jiejun Xu 和 Ross Maciejewski。图卷积网络:算法、应用和开放挑战。在计算社交网络国际会议上,第 79-91 页。施普林格,2018b。

附录

A.1解配图数据集的详细信息

以下是我们实验中使用的异类图数据集的详细信息：

- Chameleon 和 Squirrel 是维基百科网络 (Rozemberczki 等人,2019 年) ,其中节点代表来自维基百科的网页,边表示页面之间的相互链接。节点特征向量是相应页面中信息名词的词袋表示。根据网页的平均每月流量数量,每个节点被标记为五个类别之一。
- Actor 是一个参与者共现网络,其中节点表示参与者,边表示来自维基百科的同一网页上的共同出现。它是从唐等人提出的电影导演演员作家网络中提取的。(唐等人,2009 年)。节点特征向量是演员维基百科页面中关键词的词袋表示。根据演员维基百科页面的主题,每个节点都标有五个类别之一。
- Cornell,Texas 和 Wisconsin 来自卡内基梅隆大学收集的 WebKB 数据集。节点代表网页,边代表它们之间的超链接。节点特征向量是相应网页的词袋表示。每个节点都标有学生、项目、课程、职员和教员之一。

A.2完整的实验结果

表 6:我们的 NLMLP,NLGCN,NLGAT 和所有九个数据集上的基线之间的比较。

数据集	分类				异配				
	科拉	Citeseer	Pubmed	Chameleon	Squirrel	演员	康乃尔	得克萨斯州	威斯康星州
多层板	76.5±1.3	73.6 ± 1.9	87.5±0.4	48.5±3.0	31.5±1.4	35.1±0.8	81.6±6.3	81.3±7.1	84.9±5.3
GCN	88.2±1.2	75.7±1.3	88.4±0.6	67.6±2.4	30.3±1.6	54.2± 7.3	61.1±7.0	59.6±4.5	51.3±2.5
洞	88.4±1.4	76.1±1.0	87.0±0.3	65.0±3.7	± 4.3	57.9 ± 6.1	57.8±4.3		
Geom-GCN	85.3	Geom-GCN-	78.0	90.1	60.9	38.1	31.6	60.8	67.6
g	87.0	Geom-GCN-s	73.3	80.6	68.0	46.0	32.0	67.0	73.1
				72.2	61.6	38.0	34.6	75.4	73.5
NLMLP	76.9±1.8	73.4 ± 1.9	88.2 ± 0.5	50.7±2.2	33.7±1.5	37.9±1.3	84.9±5.7	85.4±3.8	87.3±4.3
NLGCN	88.1±1.0	75.2±1.4	89.0±0.5	70.1±2.9	31.6±1.0	57.6±5.5	65.5± 6.6	60.2 ± 5.3	56.8±2.5
NLGAT	88.5±1.8	76.2±1.6	88.2±0.3	65.7±1.4	54.7 ± 7.6	62.6 ± 7.1	56.9±7.3		