Author = Song Jian

#### 阅读论文： Benchmarking Graph Neural Networks

|  |  |  |
| --- | --- | --- |
| **@article{Dwivedi2020,**  **author = {Dwivedi, Vijay Prakash and Joshi, Chaitanya K. and Laurent, Thomas and Bengio, Yoshua and Bresson, Xavier},**  **title = {{Benchmarking Graph Neural Networks}},**  **year = {2020}**  **}** | **[1] V. P. Dwivedi, C. K. Joshi, T. Laurent, Y. Bengio, and X. Bresson, “Benchmarking Graph Neural Networks,” 2020.**  文章链接：<https://arxiv.org/pdf/2003.00982.pdf>  源代码：<https://github.com/graphdeeplearning/benchmarking-gnns> | **南洋理工大学**  洛约拉马利蒙特大学  蒙特利尔大学 |

#### 文档由以下几个部分组成：

* 文章简述
* 作者贡献
* 模型介绍
* 实验部分
* 可以学到什么

#### 文章简述

摘要

图神经网络（GNN）已成为分析和学习图数据的标准工具包。它们已成功应用于很多领域，包括化学，物理，社会科学，知识图谱，推荐系统和神经科学。随着领域的发展，识别跨图大小通用的体系结构和机制变得至关重要，这使我们能够处理更大，更复杂的数据集和领域。不幸的是，在缺乏统一的实验设置和大型数据集的情况下，衡量新GNN的有效性和比较模型的难度越来越大。在本文中，作者提出了一个可复现的GNN基准框架，为研究人员提供了添加新数据集和模型的便利。作者将此基准框架应用于数学建模，计算机视觉，化学和组合问题等新颖的中型图形数据集，以在设计有效的GNN时建立关键操作。精确地，图卷积，各向异性扩散，残差连接和规范化层是用于开发健壮且可扩展的GNN的通用构件。

简介

随着领域的发展，如何构建强大的GNN已经成为研究的中心问题。哪些框架、基本原理和体系结构是可以通用的？另一个重要问题是如何量化理论发展对GNN的影响？基准测试为回答这些基本问题提供了范例，事实证明，基准测试在推动科学进步，确定基本思想和解决特定领域的问题方面是有益的，例如ImageNet。但是设计成功的基准非常具有挑战性：

成功的基准需要定义适当的数据集，强大的编码接口和通用的实验设置，以便进行公平的比较，同时可复现。

#### 作者贡献

1. 用PyTorch和DGL在GitHub上发布了一个基线框架，简单容易上手，链接：<https://github.com/graphdeeplearning/benchmarking-gnns>
2. 提出了一系列中等规模的数据集，包括数学建模，计算机视觉，组合优化，化学等领域；

GNN的领域主要使用CORA和TU数据集。这些数据集是真实的，但也很小。尽管小型数据集对于快速发展新想法很有用，但从长远来看，它们可能成为负担，因为新的GNN模型将被设计为适合小型测试集，而不是寻找更通用的体系结构。另一个主要问题是缺乏实验结果的可重复性。大多数发表的论文没有使用相同的trainvalidation-test拆分。因此作者提出以下中等规模数据集。

数据集分别是MNIST, CIFAR10，ZINC，PATTERN 和 CLUSTER ， TSP，进行的任务是依次分类（acc），回归（溶解度性能指标预测, MAE），节点分类（acc），边分类（acc）

本文进行实验的模型有MLP, GCN, GAT, GaphSAGE, DiffPool, GIN, MoNet-Gaussian Mixture Model， GatedGCN等。验证了残差连接，Batch Normalization, Graph Size Normalization等模块的作用。

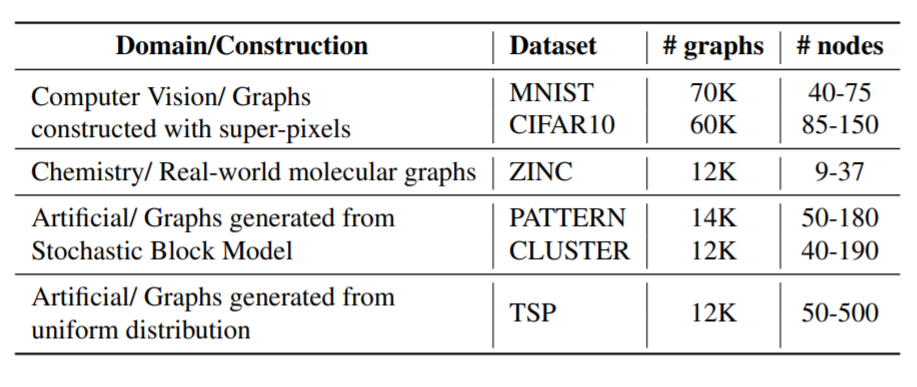


Table 1. Summary statistics of proposed

1. 确定了GNN关键部件的有效性，如异性扩散，残差连接，正则化等；
2. 论文没有对已有的模型进行ranking，而是固定参数来确定GNN重要的机制；
3. 好安装，易上手，可复现。

#### 模型介绍

图神经网络节点的表示是根据下面的公式由一层更新到另一层。 为节点i的第层的d维嵌入表示， 为非线性的， 是一个可学习的参数。

在mean版本的GraphSage中使用下式替换上式中的内容。

在GIN中则使用下式替换，是可学习的参数，BN是Batch Normalization layer，更重要的是GIN将所有的中间层都用于预测。

，

上面的模型对中央节点的更新做出同等的贡献，称为isotropic，它们同等的对待每个“边缘方向”。

另一方面MoNet–Gaussian Mixture Model Networks，GatedGC以及GAT则是anisotropic，根据各自的机制更新权重

#### 实验部分

作者首先用以上模型在之前的图分类数据集上进行了实验。红蓝黑加粗的颜色分别代表第一，第二，第三的表现。

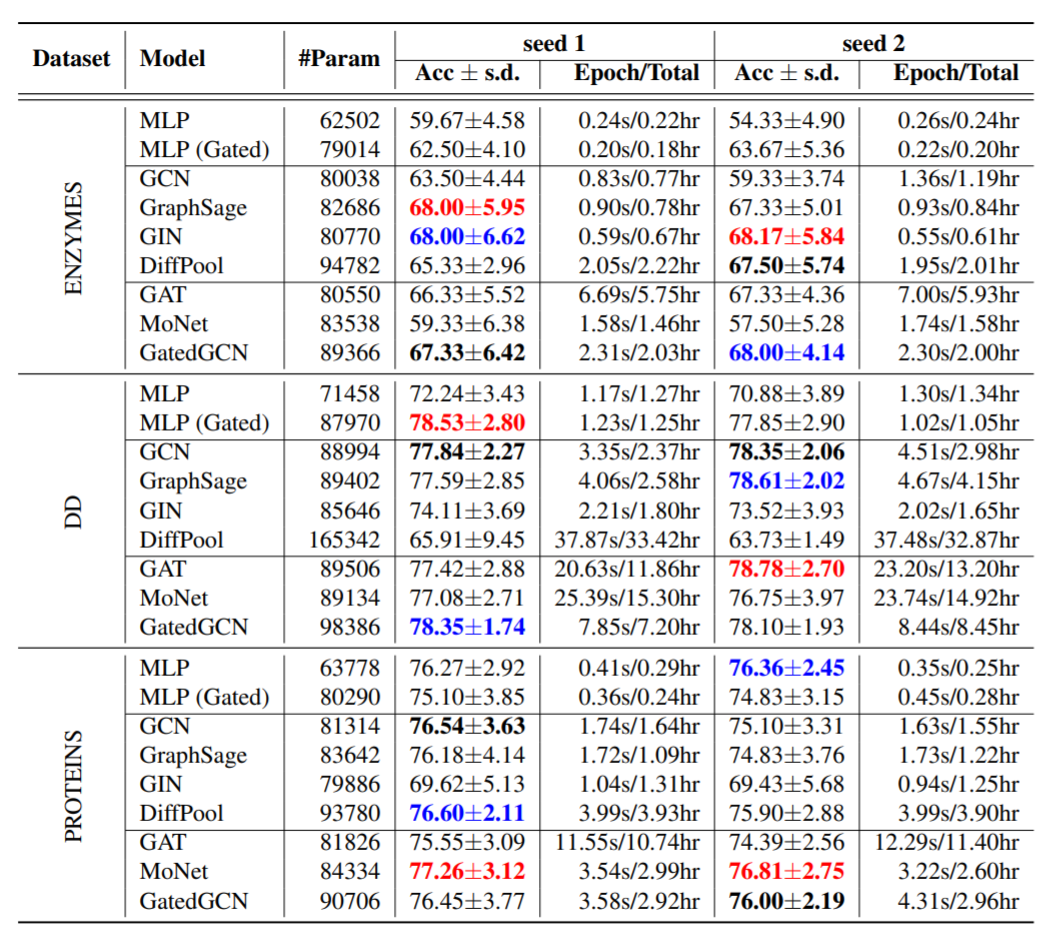


Table 2 Performance on the standard TU test sets.

可以看出来，结果标准偏差相当大，因为数据量小，按照交叉验证的思路，不同的分割方式会导致实验结果有很大的不同，这侧面反应了所有GNN的统计性能相似。另外，作者还报告了这些实验的第二次运行结果，采用相同的10折，但是不同的初始化方式，结果有较大的变化。这都可以归因于数据集的尺寸小和梯度下降优化器的不确定。还可以观察到，对于DD和Proteins数据集，MLP基线有时甚至比GNN还要好

**作者开始在自己提出的数据集上进行实验**。

作者提出以下基准测试设置。

分割：使用标准的MNIST和CIFAR10拆分。.

训练.：使用Adam优化器. For all GNNs, an initial learning rate is set to 1e−3, the reduce factor is 0.5, the patience value is 5, and the stopping learning rate is 1e−5.

准确性：性能指标是预测标签和真实标签之间的分类准确性.

重现性：报告了4次运行的平均值，有4种不同的随机种子。 在每次运行中，相同的种子

用于所有神经网络层。

图分类层：使用与之前的TU数据集中相同的分类器.

##### Graph Classification with SuperPixel Datasets

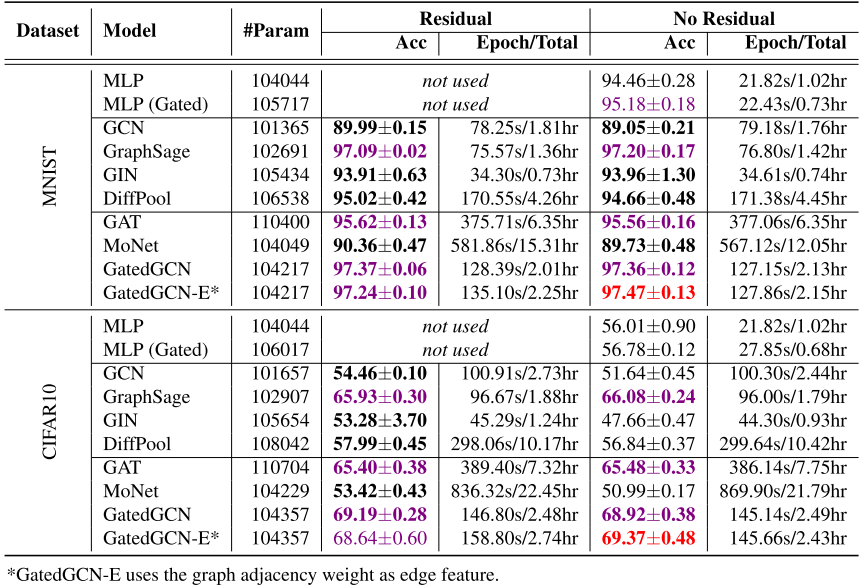


Table 3. Performance on the standard test sets of MNIST and CIFAR10 (higher is better).

##### Graph Regression with Molecular Dataset

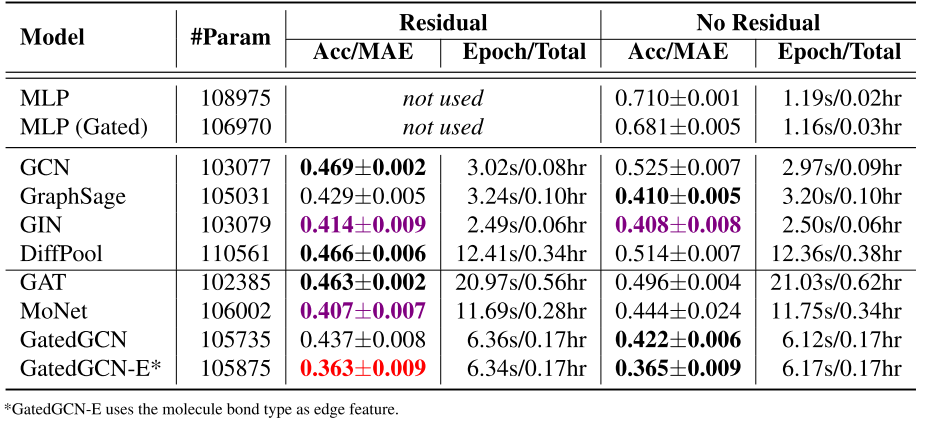


Table 4. Performance on the standard test sets of ZINC (lower is better).

##### Node Classification with SBM Datasets

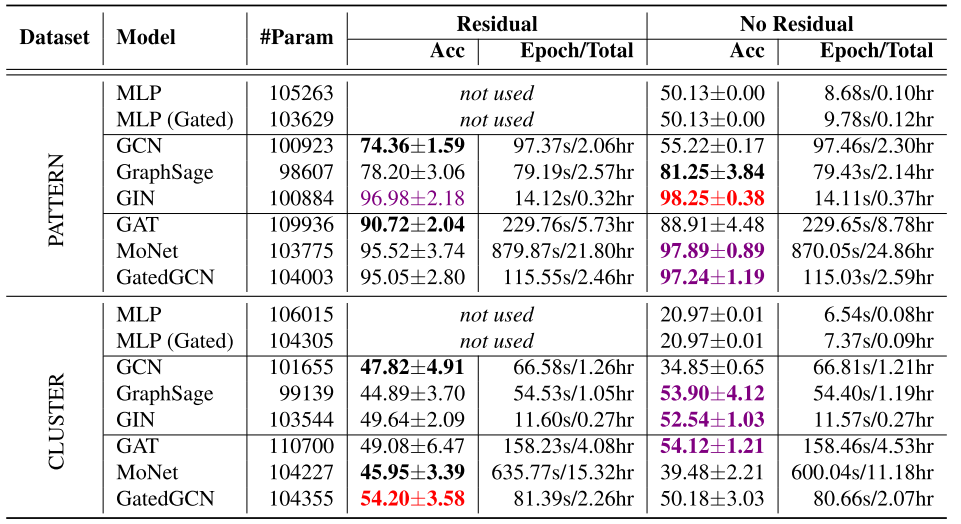


Table 5. Performance on the standard test sets of PATTERN and CLUSTER SBM graphs (higher is better).

##### Edge Classification with TSP Dataset

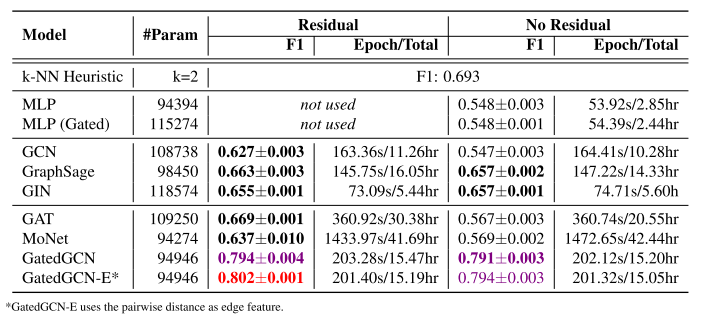


Table 6. Performance on TSP test set graphs with and without residual connections (higher is better).

#### 可以学到什么

**与图形无关的NN（MLP）在小型数据集上的表现与GNN相同**

**对于较大的数据集，GNN改进了与图无关的NN**

**最简单形式的GNN表现较差**

**各向同性（isotropic）GNN架构在原始GCN上有所改进**。GraphSage证明了在图卷积层中使用中心节点信息的重要性。GIN采用了中心节点特征以及一个新的分类器层，该分类器层在所有中间层均与卷积特征相连。DiffPool考虑了一种可学习的图形池化操作，其中在每个分辨率级别使用GraphSage。除CLUSTER外，这三个各向同性的GNN可以显着提高所有数据集的GCN性能。

**各向异性（Anisotropic）的GNN是有效的**。除了PATTERN以外，各向异性模型，例如GAT，MoNet和GatedGCN均能获得最佳结果。另外，注意到，GatedGCN在所有数据集上的性能始终都很好。注：各向同性的GNN大多依赖于相邻特征的简单总和，各向异性的GNN采用复杂的机制（GAT的稀疏关注机制，GatedGCN的边缘门）

**残差连接(Residual)能够提升模型的性能**

**正则化能够提升模型的性能**