

实验要求：

使用 pytorch 或者 tensorflow 的相关神经网络库，编写图卷积神经网络模型(GCN)，并在相应的图结构数据集上完成节点分类和链路预测任务，最后分析自环、层数、DropEdge、PairNorm、激活函数等因素对模型的分类和预测性能的影响。

实验步骤

1. **网络框架：**要求选择 pytorch 或 tensorflow 其中之一，依据官方网站的指引安装包。本次实验继续推荐大家安装 GPU 版本。如果你没有 GPU 机器，可考虑使用一些云资源，例如 Google Colab。
2. **数据准备：**本次实验使用的数据包含三个常用的图结构数据集：Cora、Citeseer、PPI。下面分别进行介绍。
 - (1) **Cora：**该数据集是由 2708 篇机器学习论文作为节点、论文间引用关系作为有向边构成的图数据。具体的数据描述见 <https://relational.fit.cvut.cz/dataset/CORA>。数据集下载链接 <https://lings-data.soe.ucsc.edu/public/lbc/cora.tgz>。另外，提供一个数据处理范例链接 <https://graphsandnetworks.com/the-cora-dataset/>。请同学们仔细阅读相关材料，了解文件的具体结构和数据格式。
 - (2) **Citeseer：**该数据集是由 3312 篇论文及相互引用构成的图数据集。数据集下载链接 <https://lings-data.soe.ucsc.edu/public/lbc/citeseer.tgz>。文件的结构和数据格式与 Cora 类似。
 - (3) **PPI：**PPI 网络是蛋白质相互作用 (Protein-Protein Interaction,PPI) 网络的简称，数描述可参考链接 <https://blog.csdn.net/ziqingnian/article/details/112979175>。数据集下载链接 <http://snap.stanford.edu/graphsage/ppi.zip>。
3. **数据预处理：**你需要通过 pytorch 或 tensorflow 所提供的标准数据接口，将原始数据处理为方便模型训练脚本所使用的数据结构，如 torch.utils.data.Dataset 等。由于这三个数据集是非常常见的公开数据集，你可以参考一些公开代码片段，尤其是 github 上典型的 GCN 教程级实现或相关论文的源码。
4. **图网络模型：**搭建 GCN 模型，这一步可以参考网络上公开的源码，但不能直接使用封装过的库。
5. **节点分类：**在三个数据集上按照节点分类任务的需求自行划分训练集、验证集、测试集，并用搭建好的 GCN 模型进行节点分类。
6. **链路预测：**在三个数据集上按照链路预测任务的需求自行划分训练集、验证集、测试集，并用搭建好的 GCN 模型进行链路预测。
7. **测试性能：**选择你认为最合适的（例如，在验证集上表现最好的）一组超参数，重新训练模型，并在测试集上测试（注意，这理应是你的实验中**唯一**一次在测试集上的测试），并记录测试的结果。

实验提交

本次实验截止日期为 **2023 年 2 月 19 日 23:59:59**，需在 bb 系统提交代码源文件及实验报告，具体要求如下：

1. 全部文件打包在一个压缩包内，压缩包命名为【学号】-姓名-作业 4.zip
2. 代码仅包含.py 文件，请勿包含数据集和实验中间结果（例如中间保存的模型等），如果有多个文件，放在 src/文件夹内
3. 实验报告要求 pdf 格式，要求包含姓名、学号，内容包括简要的**实验过程**和关键代码展示，对数据集的**处理设置超参数的实验分析**以及测试集上的**实验结果**。