

## 4.5 使用ogb包导入OGB数据集

(English Version)

[Open Graph Benchmark \(OGB\)](#) 是一个图深度学习的基准数据集。官方的 `ogb` 包提供了用于下载和处理OGB数据集到 `dgl.data.DGLGraph` 对象的API。本节会介绍它们的基本用法。

首先使用pip安装ogb包：

```
pip install ogb
```

以下代码显示了如何为 *Graph Property Prediction* 任务加载数据集。

```
# 载入OGB的Graph Property Prediction数据集
import dgl
import torch
from ogb.graphproppred import DglGraphPropPredDataset
from dgl.data.loading import GraphDataLoader

def _collate_fn(batch):
    # 小批次是一个元组(graph, Label)列表
    graphs = [e[0] for e in batch]
    g = dgl.batch(graphs)
    labels = [e[1] for e in batch]
    labels = torch.stack(labels, 0)
    return g, labels

# 载入数据集
dataset = DglGraphPropPredDataset(name='ogbg-molhiv')
split_idx = dataset.get_idx_split()
# dataLoader
train_loader = GraphDataLoader(dataset[split_idx["train"]], batch_size=32, shuffle=True,
                                collate_fn=_collate_fn)
valid_loader = GraphDataLoader(dataset[split_idx["valid"]], batch_size=32, shuffle=False,
                                collate_fn=_collate_fn)
test_loader = GraphDataLoader(dataset[split_idx["test"]], batch_size=32, shuffle=False,
                                collate_fn=_collate_fn)
```

加载 *Node Property Prediction* 数据集类似，但要注意的是这种数据集只有一个图对象。

```
# 载入OGB的Node Property Prediction数据集
from ogb.nodeproppred import DglNodePropPredDataset

dataset = DglNodePropPredDataset(name='ogbn-proteins')
split_idx = dataset.get_idx_split()

# there is only one graph in Node Property Prediction datasets
# 在Node Property Prediction数据集里只有一个图
g, labels = dataset[0]
# 获取划分的标签
train_label = dataset.labels[split_idx['train']]
valid_label = dataset.labels[split_idx['valid']]
test_label = dataset.labels[split_idx['test']]
```

每个 Link Property Prediction 数据集也只包括一个图。

```
# 载入OGB的Link Property Prediction数据集
from ogb.linkproppred import DglLinkPropPredDataset

dataset = DglLinkPropPredDataset(name='ogbl-ppa')
split_edge = dataset.get_edge_split()

graph = dataset[0]
print(split_edge['train'].keys())
print(split_edge['valid'].keys())
print(split_edge['test'].keys())
```