

7.3 运行分布式训练/推断所需的工具

(English Version)

DGL提供了两个脚本来帮助用户进行分布式训练：

- `tools/copy_files.py` 用于将图分区复制到集群，
- `tools/launch.py` 用于在机器集群中启动分布式训练任务。

`copy_files.py` 将计算机(对图进行分区的计算机)中的分区数据和相关文件(例如，训练脚本)复制到(负责分布式训练的)机器集群上。在这些机器上，分布式训练将需要用到这些分区。该脚本包含四个参数：

- `--part_config` 指定分区配置文件，该文件包含本地计算机中分区数据的信息。
- `--ip_config` 指定集群的IP配置文件。
- `--workspace` 指定训练机器中存储与分布式训练有关的所有数据的目录。
- `--rel_data_path` 指定工作空间目录下存储分区数据的相对路径。
- `--script_folder` 指定工作空间目录下存储用户的训练脚本的相对路径。

Note: `copy_files.py` 会根据IP配置文件找到对应的计算机来存储图分区。因此，`copy_files.py`和`launch.py`应该使用相同的IP配置文件。

DGL提供了用于启动集群中的分布式训练任务的`tools/launch.py`。该脚本有以下假设：

- 分区数据和训练脚本都已被复制到集群或存在集群中所有计算机均可访问的全局存储空间(例如NFS)。
- 主计算机(执行启动脚本的计算机)具有对集群内所有其他计算机的无密码ssh访问权限。

Note: 必须在集群中的一台计算机上调用启动脚本。

下面展示了在集群中启动分布式训练任务的示例。

```
python3 tools/launch.py \
--workspace ~/graphsage/ \
--num_trainers 2 \
--num_samplers 4 \
--num_servers 1 \
--part_config data/ogb-product.json \
--ip_config ip_config.txt \
"python3 code/train_dist.py --graph-name ogb-product --ip_config ip_config.txt --num-epochs 5 -
--batch-size 1000 --lr 0.1 --num_workers 4"
```

配置文件 `ip_config.txt` 包含了集群中计算机的IP地址。 `ip_config.txt` 的典型示例如下：

```
172.31.19.1
172.31.23.205
172.31.29.175
172.31.16.98
```



每行是一个计算机的IP地址。IP地址后面还可以有一个端口，用来指定不同训练器之间的网络通信所使用的端口。如果未提供具体端口，则默认值为 `30050`。

启动脚本中指定的工作空间(`-workspace`)是计算机中的工作目录，里面保存了训练脚本、IP配置文件、分区配置文件以及图分区。文件的所有路径都应指定为工作空间的相对路径。

启动脚本会在每台计算机上创建指定数量的训练任务(`--num_trainers`)。另外，用户需要为每个训练器指定采样器进程的数量(`--num_samplers`)。采样器进程的数量必须匹配 `initialize()` 中指定的工作进程的数量。