清華大學

Tsinghua University

Department of Computer Science and Technology

Machine Learning

Homework 2

Sahand Sabour

2020280401

1 DeepWalk

1.1 Motivation

1.2 Methodology

1.3 Experiments

1.4 Personal Opinion

2 (GCN)

2.1 Motivation

2.2 Methodology

2.3 Experiments

2.4 Personal Opinion

2.5 Comparison with DeepWalk

3 CogDL

3.1 Testing CogDL Models

In this assignment, we are asked to run two models on two different datasets by running the provided scripts. The first script is as follows: python scripts/train.py –task unsupervised_node_classification – dataset wikipedia –model deepwalk; by running this script, we would be training and testing the DeepWalk model on the Wikipedia dataset for the task of node classification. The results of this process is provided in the figure below (Figure 1).

```
(cogdl) ➜ cogdl git:(master) python scripts/train.py --task unsupervised_node_classification --dataset wikipedia --model deepwalk
Failed to import Deep Graph Library (DGL)
Namespace(cpu=False, dataset=['wikipedia'], device_id=[0], enhance=None, hidden_size=128, iteration=10, lr=0.01, max_epoch=500, model=
['deepwalk'], num_shuffle=5, patience=100, save_dir='.', seed=[1], task='unsupervised_node_classification', walk_length=80, walk_num=4
0, weight_decay=0.0005, window_size=5, worker=10)
Downloading http://snap.stanford.edu/node2vec/POS.mat
Processing...
Done!
node number: 4777
generating random walks...
100%|                                        | 40/40 [00:59<00:00,  1.48s/it]
training word2vec...
| Variant                | Micro-F1 0.1   | Micro-F1 0.3   | Micro-F1 0.5   | Micro-F1 0.7   | Micro-F1 0.9   |
|------------------------|----------------|----------------|----------------|----------------|----------------|
| ('wikipedia', 'deepwalk') | 0.4253±0.0000  | 0.4720±0.0000  | 0.4951±0.0000  | 0.5056±0.0000  | 0.5214±0.0000  |
```

Figure 1: Result of running the first script

Accordingly, the second script (python scripts/train.py –task node_classification –dataset citeseer – model gcn) analyzes the same task but for training and testing the GCN model on the citeseer dataset. The obtained results are demonstrated in the below figure (Figure 2).

```
(cogdl) ➜ cogdl git:(master) python scripts/train.py --task node_classification --dataset citeseer --model gcn
Failed to import Deep Graph Library (DGL)
Namespace(cpu=False, dataset=['citeseer'], device_id=[0], dropout=0.5, enhance=None, hidden_size=64, lr=0.01, max_epoch=500, missing_r
ate=-1, model=['gcn'], num_classes=None, num_features=None, patience=100, save_dir='.', seed=[1], task='node_classification', weight_d
ecay=0.0005)
Downloading https://github.com/kimiyoung/planetoid/raw/master/data/ind.citeseer.x
Downloading https://github.com/kimiyoung/planetoid/raw/master/data/ind.citeseer.tx
Downloading https://github.com/kimiyoung/planetoid/raw/master/data/ind.citeseer.allx
Downloading https://github.com/kimiyoung/planetoid/raw/master/data/ind.citeseer.y
Downloading https://github.com/kimiyoung/planetoid/raw/master/data/ind.citeseer.ty
Downloading https://github.com/kimiyoung/planetoid/raw/master/data/ind.citeseer.ally
Downloading https://github.com/kimiyoung/planetoid/raw/master/data/ind.citeseer.graph
Downloading https://github.com/kimiyoung/planetoid/raw/master/data/ind.citeseer.test.index
Processing...
Done!
Epoch: 440, Train: 1.0000, Val: 0.7000:  88%|                                    | 440/500 [00:05<00:00, 84.81it/s]
Valid accurracy = 0.73
Test accuracy = 0.72
| Variant            | Acc          |
|--------------------|--------------|
| ('citeseer', 'gcn') | 0.7200±0.0000 |
```

Figure 2: Result of running the second script

3.2 API Design

Similar to what is demonstrated in the two scripts of the previous section, the main entry point of the model training is train.py located in the scripts folder.

3.3 Implementation

In this assignment, I implemented the Hierarchical Graph Pooling with Structure Learning model [1]. The pull request id for this implementation is .

3.4 Suggestions

There are a number of suggestions that I believe would improve the CogDL experience, both as a user and as a contributor:

1. hi

3.5 Contributions

References

[1] Zhen Zhang et al. "Hierarchical graph pooling with structure learning". In:arXiv(2019) arXiv:1911.05954