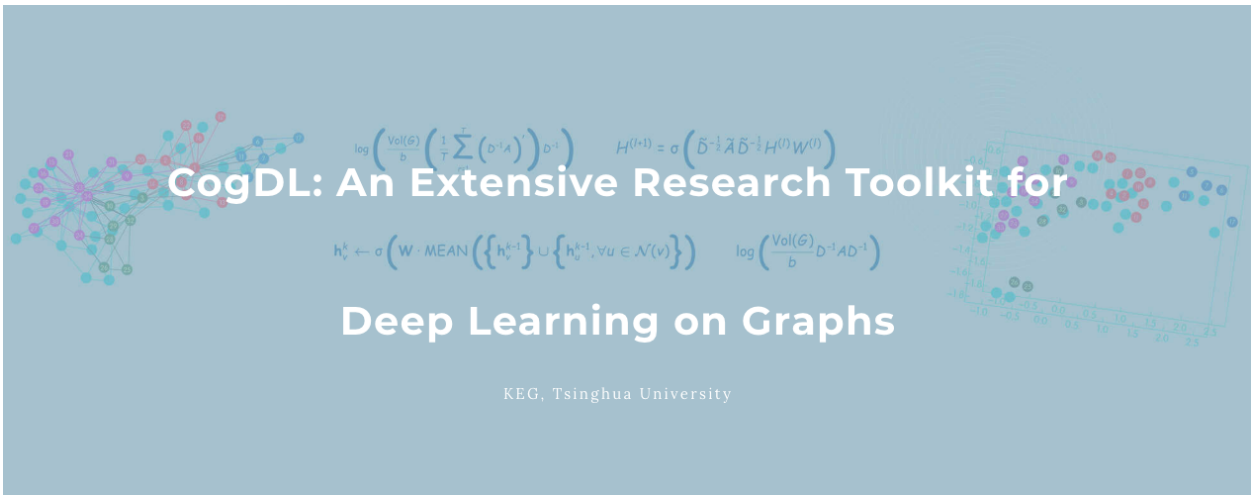
CogDL Documentation

Release 0.3.0

KEG

Mar 03, 2021

Get Started



CogDL is a graph representation learning toolkit that allows researchers and developers to easily train and compare baseline or customized models for node classification, graph classification, and other important tasks in the graph domain.

We summarize the contributions of CogDL as follows:

- **High Efficiency**: CogDL utilizes well-optimized operators to speed up training and save GPU memory of GNN models.
- **Easy-to-Use**: CogDL provides easy-to-use APIs for running experiments with the given models and datasets using hyper-parameter search.
- **Extensibility**: The design of CogDL makes it easy to apply GNN models to new scenarios based on our framework.
- **Reproducibility**: CogDL provides reproducible leaderboards for state-of-the-art models on most of important tasks in the graph domain.

CHAPTER 1

News

- The new **v0.3.0 release** provides a fast spmm operator to speed up GNN training. We also release the first version of CogDL paper in arXiv. You can join our slack for discussion.
- The new **v0.2.0 release** includes easy-to-use `experiment` and `pipeline` APIs for all experiments and applications. The `experiment` API supports automl features of searching hyper-parameters. This release also provides `OAGBert` API for model inference (`OAGBert` is trained on large-scale academic corpus by our lab). Some features and models are added by the open source community (thanks to all the contributors).
- The new **v0.1.2 release** includes a pre-training task, many examples, OGB datasets, some knowledge graph embedding methods, and some graph neural network models. The coverage of CogDL is increased to 80%. Some new APIs, such as `Trainer` and `Sampler`, are developed and being tested.
- The new **v0.1.1 release** includes the knowledge link prediction task, many state-of-the-art models, and `optuna` support. We also have a Chinese WeChat post about the CogDL release.

CHAPTER 2

Citing CogDL

Please cite our paper if you find our code or results useful for your research:

```
@article{cen2021cogdl,
   title={CogDL: An Extensive Toolkit for Deep Learning on Graphs},
   author={Yukuo Cen and Zhenyu Hou and Yan Wang and Qibin Chen and Yizhen Luo and
↪Xingcheng Yao and Aohan Zeng and Shiguang Guo and Peng Zhang and Guohao Dai and Yu
↪Wang and Chang Zhou and Hongxia Yang and Jie Tang},
   journal={arXiv preprint arXiv:2103.00959},
   year={2021}
}
```

2.1 Install

- Python version >= 3.6
- PyTorch version >= 1.6.0

Please follow the instructions here to install PyTorch (https://github.com/pytorch/pytorch#installation).

When PyTorch has been installed, cogdl can be installed using pip as follows:

```
pip install cogdl
```

Install from source via:

```
pip install git+https://github.com/thudm/cogdl.git
```

Or clone the repository and install with the following commands:

```
git clone git@github.com:THUDM/cogdl.git
cd cogdl
pip install -e .
```

If you want to use the modules from PyTorch Geometric (PyG), and Deep Graph Library (DGL), you can follow the instructions to install PyTorch Geometric (https://github.com/rusty1s/pytorch_geometric/#installation) and Deep Graph Library (https://docs.dgl.ai/install/index.html).

2.2 Quick Start

2.2.1 API Usage

You can run all kinds of experiments through CogDL APIs, especially `experiment()`. You can also use your own datasets and models for experiments. A quickstart example can be found in the quick_start.py. More examples are provided in the examples/.

```
from cogdl import experiment

# basic usage
experiment(task="node_classification", dataset="cora", model="gcn")

# set other hyper-parameters
experiment(task="node_classification", dataset="cora", model="gcn", hidden_size=32,␣
↪max_epoch=200)

# run over multiple models on different seeds
experiment(task="node_classification", dataset="cora", model=["gcn", "gat"], seed=[1,␣
↪2])

# automl usage
def func_search(trial):
    return {
        "lr": trial.suggest_categorical("lr", [1e-3, 5e-3, 1e-2]),
        "hidden_size": trial.suggest_categorical("hidden_size", [32, 64, 128]),
        "dropout": trial.suggest_uniform("dropout", 0.5, 0.8),
    }

experiment(task="node_classification", dataset="cora", model="gcn", seed=[1, 2], func_
↪search=func_search)
```

2.2.2 Command-Line Usage

You can also use `python scripts/train.py --task example_task --dataset example_dataset --model example_model` to run example_model on example_data and evaluate it via example_task.

- `--task`, downstream tasks to evaluate representation like `node_classification`, `unsupervised_node_classification`, `graph_classification`. More tasks can be found in the cogdl/tasks.
- `--dataset`, dataset name to run, can be a list of datasets with space like `cora citeseer ppi`. Supported datasets include 'cora', 'citeseer', 'pumbed', 'ppi', 'wikipedia', 'blogcatalog', 'flickr'. More datasets can be found in the cogdl/datasets.
- `--model`, model name to run, can be a list of models like `deepwalk line prone`. Supported models include 'gcn', 'gat', 'graphsage', 'deepwalk', 'node2vec', 'hope', 'grarep', 'netmf', 'netsmf', 'prone'. More models can be found in the cogdl/models.

For example, if you want to run LINE, NetMF on Wikipedia with unsupervised node classification task, with 5 different seeds:

```
python scripts/train.py --task unsupervised_node_classification --dataset wikipedia --
↪model line netmf --seed 0 1 2 3 4
```

Expected output:

Variant	Micro-F1 0.1	Micro-F1 0.3	Micro-F1 0.5	Micro-F1 0.7	Micro-F1 0.9
('wikipedia', 'line')	0.4069±0.0011	0.4071±0.0010	0.4055±0.0013	0.4054±0.0020	0.4080±0.0042
('wikipedia', 'netmf')	0.4551±0.0024	0.4932±0.0022	0.5046±0.0017	0.5084±0.0057	0.5125±0.0035

If you want to run parallel experiments on your server with multiple GPUs on multiple models, GCN and GAT, on the Cora dataset with node classification task:

```
python scripts/parallel_train.py --task node_classification --dataset cora --model␣
↪gcn gat --device-id 0 1 --seed 0 1 2 3 4
```

Expected output:

Variant	Acc
('cora', 'gcn')	0.8236±0.0033
('cora', 'gat')	0.8262±0.0032

2.2.3 Fast-Spmm Usage

CogDL provides a fast sparse matrix-matrix multiplication operator called GE-SpMM to speed up training of GNN models on the GPU. You can set `fast_spmm=True` in the API usage or `--fast-spmm` in the command-line usage to enable this feature. Note that this feature is still in testing and may not work under some versions of CUDA.

2.3 Tasks

2.3.1 Node Classification

In this tutorial, we will introduce a important task, node classification. In this task, we train a GNN model with partial node labels and use accuracy to measure the performance.

Semi-supervied Node Classification Methods

Method	Sampling	Inductive	Reproducibility
GCN			
GAT			
Chebyshev			
GraphSAGE			
GRAND			
GCNII			
DeeperGCN			
Dr-GAT			
U-net			
APPNP			
GraphMix			
DisenGCN			
SGC			
JKNet			
MixHop			
DropEdge			
SRGCN			

Tip: Reproducibility means whether the model is reproduced in our experimental setting currently.

First we define the *NodeClassification* class.

```
@register_task("node_classification")
class NodeClassification(BaseTask):
    """Node classification task."""

    @staticmethod
    def add_args(parser):
        """Add task-specific arguments to the parser."""

    def __init__(self, args):
        super(NodeClassification, self).__init__(args)
```

Then we can build dataset and model according to args. Generally the model and dataset should be placed in the same device using *.to(device)* instead of *.cuda()*. And then we set the optimizer.

```
self.device = torch.device('cpu' if args.cpu else 'cuda')
# build dataset with `build_dataset`
dataset = build_dataset(args)
self.data = dataset.data
self.data.apply(lambda x: x.to(self.device))
args.num_features = dataset.num_features
args.num_classes = dataset.num_classes

# build model with `build_model`
model = build_model(args)
self.model = model.to(self.device)
self.patience = args.patience
self.max_epoch = args.max_epoch

# set optimizer
self.optimizer = torch.optim.Adam(
```

(continues on next page)

(continued from previous page)

```
    self.model.parameters(), lr=args.lr, weight_decay=args.weight_decay
)
```

For the training process, *train* must be implemented as it will be called as the entrance of training. We provide a training loop for node classification task. For each epoch, we first call *_train_step* to optimize our model and then call *_test_step* for validation and test to compute the accuracy and loss.

```
def train(self):
    epoch_iter = tqdm(range(self.max_epoch))
    for epoch in epoch_iter:
        self._train_step()
        train_acc, _ = self._test_step(split="train")
        val_acc, val_loss = self._test_step(split="val")
        epoch_iter.set_description(
            f"Epoch: {epoch:03d}, Train: {train_acc:.4f}, Val: {val_acc:.4f}"
        )

def _train_step(self):
    """train step per epoch"""
    self.model.train()
    self.optimizer.zero_grad()
    # In node classification task, `node_classification_loss` must be defined in␣
↪model if you want to use this task directly.
    self.model.node_classification_loss(self.data).backward()
    self.optimizer.step()

def _test_step(self, split="val"):
    """test step"""
    self.model.eval()
    # `Predict` should be defined in model for inference.
    logits = self.model.predict(self.data)
    logits = F.log_softmax(logits, dim=-1)
    mask = self.data.test_mask
    loss = F.nll_loss(logits[mask], self.data.y[mask]).item()

    pred = logits[mask].max(1)[1]
    acc = pred.eq(self.data.y[mask]).sum().item() / mask.sum().item()
    return acc, loss
```

In supervied node classification tasks, we use early stopping to reduce over-fitting and save training time.

```
if val_loss <= min_loss or val_acc >= max_score:
    if val_loss <= best_loss:  # and val_acc >= best_score:
        best_loss = val_loss
        best_score = val_acc
        best_model = copy.deepcopy(self.model)
    min_loss = np.min((min_loss, val_loss))
    max_score = np.max((max_score, val_acc))
    patience = 0
else:
    patience += 1
    if patience == self.patience:
        self.model = best_model
        epoch_iter.close()
        break
```

Finally, we compute the accuracy scores of test set for the trained model.

```
test_acc, _ = self._test_step(split="test")
print(f"Test accuracy = {test_acc}")
return dict(Acc=test_acc)
```

The overall implementation of *NodeClassification* is at (https://github.com/THUDM/cogdl/blob/master/cogdl/tasks/node_classification.py).

To run NodeClassification, we can use the following command:

```
python scripts/train.py --task node_classification --dataset cora citeseer --model␣
↪gcn gat --seed 0 1 --max-epoch 500
```

Then We get experimental results like this:

Variant	Acc
('cora', 'gcn')	0.8220±0.0010
('cora', 'gat')	0.8275±0.0015
('citeseer', 'gcn')	0.7060±0.0050
('citeseer', 'gat')	0.7060±0.0020

2.3.2 Unsupervised Node Classification

In this tutorial, we will introduce a important task, unsupervised node classification. In this task, we usually apply L2 normalized logisitic regression to train a classifier and use *F1-score* or *Accuracy* to measure the performance.

Unsupervied node classificatioin includes *network embedding* methods(DeepWalk, LINE, ProNE adn etc.) and *GNN self-supervied* methods(DGI, GraphSAGE and etc.). In this section, we mainly introduce the part for *network embeddings* and the other will be presented in next section *trainer*.

Unsupervised Graph Embedding Methods

Method	Weighted	shallow network	Matrix Factorization	Reproducibility	GPU support
DeepWalk					
LINE					
Node2Vec					
NetMF					
NetSMF					
HOPE					
GraRep					
SDNE					
DNGR					
ProNE					

Unsupervised Graph Neural Network Representation Learning Methods

Method	Sampling	Inductive	Reproducibility
DGI			
MVGRL			
GRACE			
GraphSAGE			

First we define the *UnsupervisedNodeClassification* class, which has two parameters *hidden-size* and *num-shuffle* . *hidden-size* represents the dimension of node representation, while *num-shuffle* means the shuffle times in classifier.

```
@register_task("unsupervised_node_classification")
class UnsupervisedNodeClassification(BaseTask):
    """Node classification task."""

    @staticmethod
    def add_args(parser):
        """Add task-specific arguments to the parser."""
        # fmt: off
        parser.add_argument("--hidden-size", type=int, default=128)
        parser.add_argument("--num-shuffle", type=int, default=5)
        # fmt: on

    def __init__(self, args):
        super(UnsupervisedNodeClassification, self).__init__(args)
```

Then we can build dataset according to input graph's type, and get *self.label_matrix*.

```
dataset = build_dataset(args)
self.data = dataset[0]
if issubclass(dataset.__class__.__bases__[0], InMemoryDataset):
    self.num_nodes = self.data.y.shape[0]
    self.num_classes = dataset.num_classes
    self.label_matrix = np.zeros((self.num_nodes, self.num_classes), dtype=int)
    self.label_matrix[range(self.num_nodes), self.data.y] = 1
    self.data.edge_attr = self.data.edge_attr.t()
else:
    self.label_matrix = self.data.y
    self.num_nodes, self.num_classes = self.data.y.shape
```

After that, we can build model and run *model.train(G)* to obtain node representation.

```
self.model = build_model(args)
self.is_weighted = self.data.edge_attr is not None


def train(self):
    G = nx.Graph()
    if self.is_weighted:
        edges, weight = (
            self.data.edge_index.t().tolist(),
            self.data.edge_attr.tolist(),
        )
        G.add_weighted_edges_from(
            [(edges[i][0], edges[i][1], weight[0][i]) for i in range(len(edges))]
        )
    else:
        G.add_edges_from(self.data.edge_index.t().tolist())
    embeddings = self.model.train(G)
```

The spectral propagation in ProNE/ProNE++ can improve the quality of representation learned from other methods, so we can use *enhance_emb* to enhance performance. ProNE++ automatically searches for the best graph filter to help improve the embedding.

```
if self.enhance is True:
    embeddings = self.enhance_emb(G, embeddings)
```

When the embeddings are obtained, we can save them at *self.save_dir*.

At last, we evaluate embedding via run *num_shuffle* times classification under different training ratio with *features_matrix* and *label_matrix*.

```
def _evaluate(self, features_matrix, label_matrix, num_shuffle):
    # shuffle, to create train/test groups
    shuffles = []
    for _ in range(num_shuffle):
        shuffles.append(skshuffle(features_matrix, label_matrix))

    # score each train/test group
    all_results = defaultdict(list)
    training_percents = [0.1, 0.3, 0.5, 0.7, 0.9]
    for train_percent in training_percents:
        for shuf in shuffles:
```

In each shuffle, split data into two parts(training and testing) and use *LogisticRegression* to evaluate.

```
# ... shuffle to generate train/test set X_train/X_test, y_train/y_test

clf = TopKRanker(LogisticRegression())
clf.fit(X_train, y_train)

# find out how many labels should be predicted
top_k_list = list(map(int, y_test.sum(axis=1).T.tolist()[0]))
preds = clf.predict(X_test, top_k_list)
result = f1_score(y_test, preds, average="micro")
all_results[train_percent].append(result)
```

Node in graph may have multiple labels, so we conduct multilbel classification built from TopKRanker.

```
from sklearn.multiclass import OneVsRestClassifier

class TopKRanker(OneVsRestClassifier):
    def predict(self, X, top_k_list):
        assert X.shape[0] == len(top_k_list)
        probs = np.asarray(super(TopKRanker, self).predict_proba(X))
        all_labels = sp.lil_matrix(probs.shape)

        for i, k in enumerate(top_k_list):
            probs_ = probs[i, :]
            labels = self.classes_[probs_.argsort()[-k:]].tolist()
            for label in labels:
                all_labels[i, label] = 1
        return all_labels
```

Finally, we get the results of Micro-F1 score under different training ratio for different models on datasets.

`Cogdl` supports evaluating the trained embeddings ignoring the training process. With *–load-emb-path* set to the path of your result, `Cogdl` will skip the training and directly evalute the embeddings.

The overall implementation of *UnsupervisedNodeClassification* is at (https://github.com/THUDM/cogdl/blob/master/cogdl/tasks/unsupervised_node_classification.py).

To run UnsupervisedNodeClassification, we can use following instruction:

```
python scripts/train.py --task unsupervised_node_classification --dataset ppi␣
↪wikipedia --model deepwalk prone -seed 0 1
```

Then We get experimental results like this:

Variant	Micro-F1 0.1	Micro-F1 0.3	Micro-F1 0.5	Micro-F1 0.7	Micro-F1 0.9
('ppi', 'deepwalk')	0.1547±0.0002	0.1846±0.0002	0.2033±0.0015	0.2161±0.0009	0.2243±0.0018
('ppi', 'prone')	0.1777±0.0016	0.2214±0.0020	0.2397±0.0015	0.2486±0.0022	0.2607±0.0096
('wikipedia', 'deepwalk')	0.4255±0.0027	0.4712±0.0005	0.4916±0.0011	0.5011±0.0017	0.5166±0.0043
('wikipedia', 'prone')	0.4834±0.0009	0.5320±0.0020	0.5504±0.0045	0.5586±0.0022	0.5686±0.0072

2.3.3 Supervised Graph Classification

In this section, we will introduce the implementation "Graph classification task".

** Supervised Graph Classification Methods **

Method	Node Feature	Kernel	Reproducibility
GIN			
DiffPool			
SortPool			
PATCH_SAN			
DGCNN			
SAGPool			

Task Design

1. Set up "SupervisedGraphClassification" class, which has two specific parameters.
   - *degree-feature*: Use one-hot node degree as node feature, for datasets such as lmdb-binary and lmdb-multi, which don't have node features.
   - *gamma*: Multiplicative factor of learning rate decay.
   - *lr*: Learning rate.
2. Build dataset convert it to a list of *Data* defined in Cogdl. Specially, we reformat the data according to the input format of specific models. *generate_data* is implemented to convert dataset.

```
dataset = build_dataset(args)
self.data = self.generate_data(dataset, args)

def generate_data(self, dataset, args):
     if "ModelNet" in str(type(dataset).__name__):
         train_set, test_set = dataset.get_all()
         args.num_features = 3
         return {"train": train_set, "test": test_set}
    else:
        datalist = []
        if isinstance(dataset[0], Data):
            return dataset
        for idata in dataset:
            data = Data()
            for key in idata.keys:
                data[key] = idata[key]
                datalist.append(data)

        if args.degree_feature:
            datalist = node_degree_as_feature(datalist)
```

(continues on next page)

(continued from previous page)

```
            args.num_features = datalist[0].num_features
        return datalist
```

3. Then we build model and can run *train* to train the model.

```
def train(self):
    for epoch in epoch_iter:
         self._train_step()
         val_acc, val_loss = self._test_step(split="valid")
         # ...
         return dict(Acc=test_acc)

def _train_step(self):
    self.model.train()
    loss_n = 0
    for batch in self.train_loader:
        batch = batch.to(self.device)
        self.optimizer.zero_grad()
        output, loss = self.model(batch)
        loss_n += loss.item()
        loss.backward()
        self.optimizer.step()

def _test_step(self, split):
    """split in ['train', 'test', 'valid']"""
    # ...
    return acc, loss
```

The overall implementation of GraphClassification is at (https://github.com/THUDM/cogdl/blob/master/cogdl/tasks/graph_classification.py).

Create a model

To create a model for task graph classification, the following functions have to be implemented.

1. *add_args(parser)*: add necessary hyper-parameters used in model.

```
@staticmethod
def add_args(parser):
     parser.add_argument("--hidden-size", type=int, default=128)
     parser.add_argument("--num-layers", type=int, default=2)
     parser.add_argument("--lr", type=float, default=0.001)
     # ...
```

2. *build_model_from_args(cls, args)*: this function is called in 'task' to build model.
3. *split_dataset(cls, dataset, args)*: split train/validation/test data and return correspondent dataloader according to requirement of model.

```
def split_dataset(cls, dataset, args):
    random.shuffle(dataset)
    train_size = int(len(dataset) * args.train_ratio)
    test_size = int(len(dataset) * args.test_ratio)
    bs = args.batch_size
    train_loader = DataLoader(dataset[:train_size], batch_size=bs)
    test_loader = DataLoader(dataset[-test_size:], batch_size=bs)
    if args.train_ratio + args.test_ratio < 1:
```

(continues on next page)

(continued from previous page)

```
        valid_loader = DataLoader(dataset[train_size:-test_size], batch_size=bs)
    else:
        valid_loader = test_loader
    return train_loader, valid_loader, test_loader
```

4. *forward*: forward propagation, and the return should be (predication, loss) or (prediction, None), respectively for training and test. Input parameters of *forward* is class *Batch*, which

```
def forward(self, batch):
 h = batch.x
 layer_rep = [h]
 for i in range(self.num_layers-1):
     h = self.gin_layers[i](h, batch.edge_index)
     h = self.batch_norm[i](h)
     h = F.relu(h)
     layer_rep.append(h)

 final_score = 0
 for i in range(self.num_layers):
 pooled = scatter_add(layer_rep[i], batch.batch, dim=0)
 final_score += self.dropout(self.linear_prediction[i](pooled))
 final_score = F.softmax(final_score, dim=-1)
 if batch.y is not None:
     loss = self.loss(final_score, batch.y)
     return final_score, loss
 return final_score, None
```

Run

To run GraphClassification, we can use the following command:

```
python scripts/train.py --task graph_classification --dataset proteins --model gin␣
↪diffpool sortpool dgcnn --seed 0 1
```

Then We get experimental results like this:

Variants	Acc
('proteins', 'gin')	0.7286±0.0598
('proteins', 'diffpool')	0.7530±0.0589
('proteins', 'sortpool')	0.7411±0.0269
('proteins', 'dgcnn')	0.6677±0.0355
('proteins', 'patchy_san')	0.7550±0.0812

2.3.4 Unsupervised Graph Classification

In this section, we will introduce the implementation "Unsupervised graph classification task".

Unsupervised Graph Classificaton Methods

Method	Node Feature	Kernel	Reproducibility
InfoGraph			
DGK			
Graph2Vec			
HGP_SL			

Task Design

1. Set up "UnsupervisedGraphClassification" class, which has two specific parameters.
    - *num-shuffle* : Shuffle times in classifier
    - *degree-feature*: Use one-hot node degree as node feature, for datasets such as lmdb-binary and lmdb-multi, which don't have node features.
    - *lr*: learning

```
@register_task("unsupervised_graph_classification")
class UnsupervisedGraphClassification(BaseTask):
    r"""Unsupervised graph classification"""
    @staticmethod
    def add_args(parser):
        """Add task-specific arguments to the parser."""
        # fmt: off
        parser.add_argument("--num-shuffle", type=int, default=10)
        parser.add_argument("--degree-feature", dest="degree_feature", action="store_
↪true")
        parser.add_argument("--lr", type=float, default=0.001)
        # fmt: on
   def __init__(self, args):
     # ...
```

2. Build dataset and convert it to a list of *Data* defined in Cogdl.

```
dataset = build_dataset(args)
self.label = np.array([data.y for data in dataset])
self.data = [
     Data(x=data.x, y=data.y, edge_index=data.edge_index, edge_attr=data.edge_attr,
             pos=data.pos).apply(lambda x:x.to(self.device))
             for data in dataset
]
```

3. Then we build model and can run *train* to train the model and obtain graph representation. In this part, the training process of shallow models and deep models are implemented separately.

```
self.model = build_model(args)
self.model = self.model.to(self.device)

def train(self):
     if self.use_nn:
        # deep neural network models
             epoch_iter = tqdm(range(self.epoch))
        for epoch in epoch_iter:
            loss_n = 0
            for batch in self.data_loader:
                batch = batch.to(self.device)
                predict, loss = self.model(batch.x, batch.edge_index, batch.batch)
             self.optimizer.zero_grad()
             loss.backward()
             self.optimizer.step()
             loss_n += loss.item()
     # ...
    else:
       # shallow models
        prediction, loss = self.model(self.data)
        label = self.label
```

4. When graph representation is obtained, we evaluate the embedding with *SVM* via running *num_shuffle* times under different training ratio. You can also call *save_emb* to save the embedding.

```
return self._evaluate(prediction, label)
def _evaluate(self, embedding, labels):
    # ...
    for training_percent in training_percents:
         for shuf in shuffles:
            # ...
            clf = SVC()
            clf.fit(X_train, y_train)
            preds = clf.predict(X_test)
            # ...
```

The overall implementation of UnsupervisedGraphClassification is at (https://github.com/THUDM/cogdl/blob/master/cogdl/tasks/unsupervised_graph_classification.py).

Create a model

To create a model for task unsupervised graph classification, the following functions have to be implemented.

1. *add_args(parser)*: add necessary hyper-parameters used in model.

```
@staticmethod
def add_args(parser):
  parser.add_argument("--hidden-size", type=int, default=128)
  parser.add_argument("--nn", type=bool, default=False)
  parser.add_argument("--lr", type=float, default=0.001)
  # ...
```

2. *build_model_from_args(cls, args)*: this function is called in 'task' to build model.

3. *forward*: For shallow models, this function runs as training process of model and will be called only once; For deep neural network models, this function is actually the forward propagation process and will be called many times.

```
# shallow model
def forward(self, graphs):
     # ...
    self.model = Doc2Vec(
        self.doc_collections,
             ...
    )
    vectors = np.array([self.model["g_"+str(i)] for i in range(len(graphs))])
    return vectors, None
```

Run

To run UnsupervisedGraphClassification, we can use the following command:

```
python scripts/train.py --task unsupervised_graph_classification --dataset proteins --
↪model dgk graph2vec
```

Then we get experimental results like this:

Variant	Acc
('proteins', 'dgk')	0.7259±0.0118
('proteins', 'graph2vec')	0.7330±0.0043
('proteins', 'infograph')	0.7393±0.0070

2.3.5 Link Prediction

In this tutorial, we will introduce a important link prediction. Overall speaking, the link prediction in CogDL can be divided into 3 types.

1. Network embeddings based link prediction(*HomoLinkPrediction*). All unsupervised network embedding methods supports this task for homogenous graphs without node features.
2. Knowledge graph completion(*KGLinkPrediction* and *TripleLinkPrediction*), including knowledge embedding methods(TransE, DistMult) and GNN base methods(RGCN and CompGCN).
3. GNN base homogenous graph link prediction(*GNNHomoLinkPrediction*). Theoretically, all GNN models works.

	Models
Network embeddings methods	DeepWalk, LINE, Node2Vec, ProNE NetMF, NetSMF, SDNE, Hope
Knowledge graph completion	TransE, DistMult, RotatE, RGCN, CompGCN
GNN methods	GCN and all the other GNN methods

To implement a new GNN model for link prediction, just implement *link_prediction_loss* in the model which accepting thre parameters:

- Node features.
- Edge index.
- Labels. 0/1 for each item, indicating the edge exists in the graph or is a negative sample.

The overall implementation can be found at https://github.com/THUDM/cogdl/blob/master/cogdl/tasks/link_prediction.py

2.3.6 Other Tasks

Heterogeneous Graph Embedding Methods

Method	Multi-Node	Multi-Edge	Supervised	Attribute	MetaPath
GATNE					
Metapath2Vec					
PTE					
Hin2Vec					
GTN					
HAN					

Attributed Graph Clustering

Method	Content	Spectral
kmeans		
spectral		
PRONE		
NetMF		
deepwalk		
line		
AGC		
DAEGC		

Pretrained Graph Models

- STPGNN: Strategies for pretraining graph neunral networks
- GCC: GCC: Graph Contrastive Coding for Graph Neural Network Pre-Training

2.3.7 Create new tasks

You can build a new task in the CogDL. The BaseTask class are:

```
class BaseTask(object):
    @staticmethod
    def add_args(parser):
        """Add task-specific arguments to the parser."""
        pass

    def __init__(self, args):
        pass

    def train(self, num_epoch):
        raise NotImplementedError
```

You can create a subclass to implement 'train' method like CommunityDetection, which get representation of each node and apply clustering algorithm (K-means) to evaluate.

```
@register_task("community_detection")
class CommunityDetection(BaseTask):
    """Community Detection task."""

    @staticmethod
    def add_args(parser):
        """Add task-specific arguments to the parser."""
        parser.add_argument("--hidden-size", type=int, default=128)
        parser.add_argument("--num-shuffle", type=int, default=5)

    def __init__(self, args):
        super(CommunityDetection, self).__init__(args)
        dataset = build_dataset(args)
        self.data = dataset[0]

        self.num_nodes, self.num_classes = self.data.y.shape
        self.label = np.argmax(self.data.y, axis=1)
        self.model = build_model(args)
        self.hidden_size = args.hidden_size
        self.num_shuffle = args.num_shuffle

    def train(self):
        G = nx.Graph()
        G.add_edges_from(self.data.edge_index.t().tolist())
        embeddings = self.model.train(G)

        clusters = [30, 50, 70]
        all_results = defaultdict(list)
        for num_cluster in clusters:
            for _ in range(self.num_shuffle):
                model = KMeans(n_clusters=num_cluster).fit(embeddings)
                nmi_score = normalized_mutual_info_score(self.label, model.labels_)
```

(continues on next page)

(continued from previous page)

```
                all_results[num_cluster].append(nmi_score)

        return dict(
            (
                f"normalized_mutual_info_score {num_cluster}",
                sum(all_results[num_cluster]) / len(all_results[num_cluster]),
            )
            for num_cluster in sorted(all_results.keys())
        )
```

After creating your own task, you could run the task on different models and dataset. You can use 'build_model', 'build_dataset', 'build_task' method to build them with coresponding hyper-parameters.

```
from cogdl.tasks import build_task
from cogdl.datasets import build_dataset
from cogdl.models import build_model
from cogdl.utils import build_args_from_dict

def run_deepwalk_ppi():
    default_dict = {'hidden_size': 64, 'num_shuffle': 1, 'cpu': True}
    args = build_args_from_dict(default_dict)

    # model, dataset and task parameters
    args.model = 'spectral'
    args.dataset = 'ppi'
    args.task = 'community_detection'

    # build model, dataset and task
    dataset = build_dataset(args)
    model = build_model(args)
    task = build_task(args)

    # train model and get evaluate results
    ret = task.train()
    print(ret)
```

2.4 Trainer

In this section, we will introduce how to implement a specific *Trainer* for a model.

In previous section, we introduce the implementation of different *tasks*. But the training paradigm varies and is incompatible with the defined training process in some cases. Therefore, *CogDL* provides *Trainer* to customize the training and inference mode. Take *NeighborSamplingTrainer* as the example, this section will show how to define a trainer.

Design

1. A self-defined trainer should inherits *BaseTrainer* and must implement function *fit* to define the training and evaluating process. Necessary parameters for training need to be added to the *add_args* in models and can be obtained here in *__init__*.

```
class NeighborSamplingTrainer(BaseTrainer):
    def __init__(self, args):
        # ... get necessary parameters from args
```

(continues on next page)

(continued from previous page)

```
    def fit(self, model, dataset):
        # ... implement the training and evaluation

    @classmethod
    def build_trainer_from_args(cls, args):
        return cls(args)
```

2. All training and evaluating process, including data preprocessing and defining optimizer, should be implemented in *fit*. In other words, given the model and dataset, the rest is up to you. *fit* accepts two parameters: model and dataset, which usually are in cpu. You need to move them to cuda if you want to train on GPU.

```
def fit(self, model, dataset):
    self.data = dataset[0]

    # preprocess data
    self.train_loader = NeighborSampler(
        data=self.data,
        mask=self.data.train_mask,
        sizes=self.sample_size,
        batch_size=self.batch_size,
        num_workers=self.num_workers,
        shuffle=True,
    )
    self.test_loader = NeighborSampler(
        data=self.data, mask=None, sizes=[-1], batch_size=self.batch_size,␣
↪shuffle=False
    )
    # move model to GPU
    self.model = model.to(self.device)

    # define optimizer
    self.optimizer = torch.optim.Adam(self.model.parameters(), lr=self.lr, weight_
↪decay=self.weight_decay)
    # training
    best_model = self.train()
    self.model = best_model
    # evaluation
    acc, loss = self._test_step()
    return dict(Acc=acc["test"], ValAcc=acc["val"])
```

3. To make the training of a model use the trainer, we should assign the trainer to the model. In Cogdl, a model must implement *get_trainer* as static method if it has a customized training process. GraphSAGE depends on *NeighborSamplingTrainer*, so the following codes should exsits in the implementation.

```
@staticmethod
def get_trainer(taskType, args):
    return NeighborSamplingTrainer
```

The details of training and evaluating are similar to the implementation in *Tasks*. The overall implementation of trainers is at https://github.com/THUDM/cogdl/tree/master/cogdl/trainers

2.5 Model

In this section, we will create a spectral clustering model, which is a very simple graph embedding algorithm. We name it spectral.py and put it in cogdl/models/emb directory.

First we import necessary library like numpy, scipy, networkx, sklearn, we also import API like 'BaseModel' and 'register_model' from cogl/models/ to build our new model:

```
import numpy as np
import networkx as nx
import scipy.sparse as sp
from sklearn import preprocessing
from .. import BaseModel, register_model
```

Then we use function decorator to declare new model for CogDL

```
@register_model('spectral')
class Spectral(BaseModel):
    (...)
```

We have to implement method 'build_model_from_args' in spectral.py. If it need more parameters to train, we can use 'add_args' to add model-specific arguments.

```
@staticmethod
def add_args(parser):
    """Add model-specific arguments to the parser."""
    pass

@classmethod
def build_model_from_args(cls, args):
    return cls(args.hidden_size)

def __init__(self, dimension):
    super(Spectral, self).__init__()
    self.dimension = dimension
```

Each new model should provide a 'train' method to obtain representation.

```
def train(self, G):
    matrix = nx.normalized_laplacian_matrix(G).todense()
    matrix = np.eye(matrix.shape[0]) - np.asarray(matrix)
    ut, s, _ = sp.linalg.svds(matrix, self.dimension)
    emb_matrix = ut * np.sqrt(s)
    emb_matrix = preprocessing.normalize(emb_matrix, "l2")
    return emb_matrix
```

All implemented models are at https://github.com/THUDM/cogdl/tree/master/cogdl/models.

2.6 Dataset

In order to add a dataset into CogDL, you should know your dataset's format. We have provided several graph format like edgelist, matlab_matrix and pyg. If the format of your dataset is the same as the *ppi* dataset, which contains two matrices: *network* and *group*, you can register your dataset directly use the following code.

```
@register_dataset("ppi")
class PPIDataset(MatlabMatrix):
    def __init__(self):
        dataset, filename = "ppi", "Homo_sapiens"
        url = "http://snap.stanford.edu/node2vec/"
        path = osp.join("data", dataset)
        super(PPIDataset, self).__init__(path, filename, url)
```

You should declare the name of the dataset, the name of file and the url, where our script can download resource. More implemented datasets are at https://github.com/THUDM/cogdl/tree/master/cogdl/datasets.

2.7 data

class cogdl.data.**Data**(*x=None*, *edge_index=None*, *edge_attr=None*, *y=None*, *pos=None*, ***kwargs*)
Bases: object

A plain old python object modeling a single graph with various (optional) attributes:

Args:

x (Tensor, optional): Node feature matrix with shape :obj:'[num_nodes, num_node_features]'. (default: None)

edge_index (LongTensor, optional): Graph connectivity in COO format with shape [2, num_edges]. (default: None)

edge_attr (Tensor, optional): Edge feature matrix with shape [num_edges, num_edge_features]. (default: None)

y (Tensor, optional): Graph or node targets with arbitrary shape. (default: None)

pos (Tensor, optional): Node position matrix with shape [num_nodes, num_dimensions]. (default: None)

The data object is not restricted to these attributes and can be extented by any other additional data.

apply(*func*, **keys*)
Applies the function func to all attributes *keys. If *keys is not given, func is applied to all present attributes.

cat_dim(*key*, *value*)
Returns the dimension in which the attribute key with content value gets concatenated when creating batches.

Note: This method is for internal use only, and should only be overridden if the batch concatenation process is corrupted for a specific data attribute.

---

clone()

contiguous(**keys*)
Ensures a contiguous memory layout for all attributes *keys. If *keys is not given, all present attributes are ensured to have a contiguous memory layout.

cuda(**keys*)

edge_subgraph(*edge_idx*, *require_idx=False*)
Return the induced edge subgraph.

static from_dict(*dictionary*)
Creates a data object from a python dictionary.

static from_pyg_data(*data*)

is_coalesced()
Returns True, if edge indices are ordered and do not contain duplicate entries.

keys
Returns all names of graph attributes.

num_classes

num_edges
Returns the number of edges in the graph.

num_features
Returns the number of features per node in the graph.

num_nodes

sample_adj(*batch*, *size=-1*, *replace=True*)

subgraph(*node_idx*)
Return the induced node subgraph.

to(*device*, **keys*)
Performs tensor dtype and/or device conversion to all attributes `*keys`. If `*keys` is not given, the conversion is applied to all present attributes.

class cogdl.data.**Batch**(*batch=None*, ***kwargs*)
Bases: cogdl.data.data.Data

A plain old python object modeling a batch of graphs as one big (dicconnected) graph. With *cogdl.data.Data* being the base class, all its methods can also be used here. In addition, single graphs can be reconstructed via the assignment vector `batch`, which maps each node to its respective graph identifier.

cumsum(*key*, *item*)
If `True`, the attribute `key` with content `item` should be added up cumulatively before concatenated together.

Note: This method is for internal use only, and should only be overridden if the batch concatenation process is corrupted for a specific data attribute.

---

static from_data_list(*data_list*, *follow_batch=[]*)
Constructs a batch object from a python list holding `torch_geometric.data.Data` objects. The assignment vector `batch` is created on the fly. Additionally, creates assignment batch vectors for each key in `follow_batch`.

num_graphs
Returns the number of graphs in the batch.

class cogdl.data.**Dataset**(*root*, *transform=None*, *pre_transform=None*, *pre_filter=None*)
Bases: torch.utils.data.dataset.Dataset

Dataset base class for creating graph datasets. See here for the accompanying tutorial.

Args: root (string): Root directory where the dataset should be saved. transform (callable, optional): A function/transform that takes in an

cogdl.data.Data object and returns a transformed version. The data object will be transformed before every access. (default: `None`)

pre_transform (callable, optional): A function/transform that takes in an *cogdl.data.Data* object and returns a transformed version. The data object will be transformed before being saved to disk. (default: `None`)

pre_filter (callable, optional): A function that takes in an *cogdl.data.Data* object and returns a boolean value, indicating whether the data object should be included in the final dataset. (default: `None`)

static add_args(*parser*)
Add dataset-specific arguments to the parser.

download()
Downloads the dataset to the `self.raw_dir` folder.

get(*idx*)
Gets the data object at index `idx`.

get_evaluator()

get_loss_fn()

num_classes
The number of classes in the dataset.

num_features
Returns the number of features per node in the graph.

process()
Processes the dataset to the `self.processed_dir` folder.

processed_file_names
The name of the files to find in the `self.processed_dir` folder in order to skip the processing.

processed_paths
The filepaths to find in the `self.processed_dir` folder in order to skip the processing.

raw_file_names
The name of the files to find in the `self.raw_dir` folder in order to skip the download.

raw_paths
The filepaths to find in order to skip the download.

class cogdl.data.**DataLoader**(*dataset*, *batch_size=1*, *shuffle=True*, ***kwargs*)
Bases: `torch.utils.data.dataloader.DataLoader`

Data loader which merges data objects from a `cogdl.data.dataset` to a mini-batch.

Args: dataset (Dataset): The dataset from which to load the data. batch_size (int, optional): How may samples per batch to load.

(default: `1`)

shuffle (bool, optional): If set to `True`, the data will be reshuffled at every epoch (default: `True`)

static collate_fn(*batch*)

class cogdl.data.**MultiGraphDataset**(*root=None*, *transform=None*, *pre_transform=None*, *pre_filter=None*)
Bases: `cogdl.data.dataset.Dataset`

static from_data_list(*data_list*)
Borrowed from PyG

get(*idx*)
Gets the data object at index `idx`.

len()

num_classes
The number of classes in the dataset.

2.8 datasets

2.8.1 GATNE dataset

class cogdl.datasets.gatne.**AmazonDataset**
Bases: *cogdl.datasets.gatne.GatneDataset*

class cogdl.datasets.gatne.**GatneDataset**(*root*, *name*)
Bases: cogdl.data.dataset.Dataset

The network datasets "Amazon", "Twitter" and "YouTube" from the "Representation Learning for Attributed Multiplex Heterogeneous Network" paper.

Args: root (string): Root directory where the dataset should be saved. name (string): The name of the dataset ("Amazon",

"Twitter", "YouTube").

download()
Downloads the dataset to the self.raw_dir folder.

get(*idx*)
Gets the data object at index idx.

process()
Processes the dataset to the self.processed_dir folder.

processed_file_names
The name of the files to find in the self.processed_dir folder in order to skip the processing.

raw_file_names
The name of the files to find in the self.raw_dir folder in order to skip the download.

url = 'https://github.com/THUDM/GATNE/raw/master/data'

class cogdl.datasets.gatne.**TwitterDataset**
Bases: *cogdl.datasets.gatne.GatneDataset*

class cogdl.datasets.gatne.**YouTubeDataset**
Bases: *cogdl.datasets.gatne.GatneDataset*

cogdl.datasets.gatne.**read_gatne_data**(*folder*)

2.8.2 GCC dataset

class cogdl.datasets.gcc_data.**Edgelist**(*root*, *name*)
Bases: cogdl.data.dataset.Dataset

download()
Downloads the dataset to the self.raw_dir folder.

get(*idx*)
Gets the data object at index idx.

num_classes
The number of classes in the dataset.

process()
Processes the dataset to the self.processed_dir folder.

processed_file_names
The name of the files to find in the `self.processed_dir` folder in order to skip the processing.

raw_file_names
The name of the files to find in the `self.raw_dir` folder in order to skip the download.

url = 'https://github.com/cenyk1230/gcc-data/raw/master'

class cogdl.datasets.gcc_data.**GCCDataset**(*root*, *name*)
Bases: cogdl.data.dataset.Dataset

download()
Downloads the dataset to the `self.raw_dir` folder.

get(*idx*)
Gets the data object at index `idx`.

preprocess(*root*, *name*)

processed_file_names
The name of the files to find in the `self.processed_dir` folder in order to skip the processing.

raw_file_names
The name of the files to find in the `self.raw_dir` folder in order to skip the download.

url = 'https://github.com/cenyk1230/gcc-data/raw/master'

class cogdl.datasets.gcc_data.**KDD_ICDM_GCCDataset**
Bases: *cogdl.datasets.gcc_data.GCCDataset*

class cogdl.datasets.gcc_data.**SIGIR_CIKM_GCCDataset**
Bases: *cogdl.datasets.gcc_data.GCCDataset*

class cogdl.datasets.gcc_data.**SIGMOD_ICDE_GCCDataset**
Bases: *cogdl.datasets.gcc_data.GCCDataset*

class cogdl.datasets.gcc_data.**USAAirportDataset**
Bases: *cogdl.datasets.gcc_data.Edgelist*

2.8.3 GTN dataset

class cogdl.datasets.gtn_data.**ACM_GTNDataset**
Bases: *cogdl.datasets.gtn_data.GTNDataset*

class cogdl.datasets.gtn_data.**DBLP_GTNDataset**
Bases: *cogdl.datasets.gtn_data.GTNDataset*

class cogdl.datasets.gtn_data.**GTNDataset**(*root*, *name*)
Bases: cogdl.data.dataset.Dataset

The network datasets "ACM", "DBLP" and "IMDB" from the "Graph Transformer Networks" paper.

Args: root (string): Root directory where the dataset should be saved. name (string): The name of the dataset (`"gtn-acm"`,

`"gtn-dblp"`, `"gtn-imdb"`).

apply_to_device(*device*)

download()
Downloads the dataset to the `self.raw_dir` folder.

get(*idx*)
Gets the data object at index `idx`.

num_classes
The number of classes in the dataset.

process()
Processes the dataset to the `self.processed_dir` folder.

processed_file_names
The name of the files to find in the `self.processed_dir` folder in order to skip the processing.

raw_file_names
The name of the files to find in the `self.raw_dir` folder in order to skip the download.

read_gtn_data(*folder*)

class cogdl.datasets.gtn_data.**IMDB_GTNDataset**
Bases: *cogdl.datasets.gtn_data.GTNDataset*

2.8.4 HAN dataset

class cogdl.datasets.han_data.**ACM_HANDataset**
Bases: *cogdl.datasets.han_data.HANDataset*

class cogdl.datasets.han_data.**DBLP_HANDataset**
Bases: *cogdl.datasets.han_data.HANDataset*

class cogdl.datasets.han_data.**HANDataset**(*root*, *name*)
Bases: cogdl.data.dataset.Dataset

The network datasets "ACM", "DBLP" and "IMDB" from the "Heterogeneous Graph Attention Network" paper.

Args: root (string): Root directory where the dataset should be saved. name (string): The name of the dataset (`"han-acm"`,

`"han-dblp"`, `"han-imdb"`).

apply_to_device(*device*)

download()
Downloads the dataset to the `self.raw_dir` folder.

get(*idx*)
Gets the data object at index `idx`.

num_classes
The number of classes in the dataset.

process()
Processes the dataset to the `self.processed_dir` folder.

processed_file_names
The name of the files to find in the `self.processed_dir` folder in order to skip the processing.

raw_file_names
The name of the files to find in the `self.raw_dir` folder in order to skip the download.

read_gtn_data(*folder*)

class cogdl.datasets.han_data.**IMDB_HANDataset**
Bases: *cogdl.datasets.han_data.HANDataset*

cogdl.datasets.han_data.**sample_mask**(*idx*, *length*)
Create mask.

2.8.5 KG dataset

class cogdl.datasets.kg_data.**BidirectionalOneShotIterator**(*dataloader_head*, *dataloader_tail*)

Bases: object

static one_shot_iterator(*dataloader*)
Transform a PyTorch Dataloader into python iterator

class cogdl.datasets.kg_data.**FB13Datset**
Bases: *cogdl.datasets.kg_data.KnowledgeGraphDataset*

class cogdl.datasets.kg_data.**FB13SDatset**
Bases: *cogdl.datasets.kg_data.KnowledgeGraphDataset*

url = 'https://raw.githubusercontent.com/cenyk1230/test-data/main'

class cogdl.datasets.kg_data.**FB15k237Datset**
Bases: *cogdl.datasets.kg_data.KnowledgeGraphDataset*

class cogdl.datasets.kg_data.**FB15kDatset**
Bases: *cogdl.datasets.kg_data.KnowledgeGraphDataset*

class cogdl.datasets.kg_data.**KnowledgeGraphDataset**(*root*, *name*)
Bases: cogdl.data.dataset.Dataset

download()
Downloads the dataset to the self.raw_dir folder.

get(*idx*)
Gets the data object at index idx.

num_entities

num_relations

process()
Processes the dataset to the self.processed_dir folder.

processed_file_names
The name of the files to find in the self.processed_dir folder in order to skip the processing.

raw_file_names
The name of the files to find in the self.raw_dir folder in order to skip the download.

test_start_idx

train_start_idx

url = 'https://raw.githubusercontent.com/thunlp/OpenKE/OpenKE-PyTorch/benchmarks'

valid_start_idx

class cogdl.datasets.kg_data.**TestDataset**(*triples*, *all_true_triples*, *nentity*, *nrelation*, *mode*)
Bases: torch.utils.data.dataset.Dataset

static collate_fn(*data*)

class cogdl.datasets.kg_data.**TrainDataset**(*triples*, *nentity*, *nrelation*, *negative_sample_size*, *mode*)
Bases: torch.utils.data.dataset.Dataset

static collate_fn(*data*)

static count_frequency(*triples*, *start=4*)
Get frequency of a partial triple like (head, relation) or (relation, tail) The frequency will be used for subsampling like word2vec

static get_true_head_and_tail(*triples*)
Build a dictionary of true triples that will be used to filter these true triples for negative sampling

class cogdl.datasets.kg_data.**WN18Datset**
Bases: *cogdl.datasets.kg_data.KnowledgeGraphDataset*

class cogdl.datasets.kg_data.**WN18RRDataset**
Bases: *cogdl.datasets.kg_data.KnowledgeGraphDataset*

cogdl.datasets.kg_data.**read_triplet_data**(*folder*)

2.8.6 Matlab matrix dataset

class cogdl.datasets.matlab_matrix.**BlogcatalogDataset**
Bases: *cogdl.datasets.matlab_matrix.MatlabMatrix*

class cogdl.datasets.matlab_matrix.**DblpNEDataset**
Bases: *cogdl.datasets.matlab_matrix.NetworkEmbeddingCMTYDataset*

class cogdl.datasets.matlab_matrix.**FlickrDataset**
Bases: *cogdl.datasets.matlab_matrix.MatlabMatrix*

class cogdl.datasets.matlab_matrix.**MatlabMatrix**(*root*, *name*, *url*)
Bases: cogdl.data.dataset.Dataset

networks from the http://leitang.net/code/social-dimension/data/ or http://snap.stanford.edu/node2vec/

Args: root (string): Root directory where the dataset should be saved. name (string): The name of the dataset (`"Blogcatalog"`).

download()
Downloads the dataset to the `self.raw_dir` folder.

get(*idx*)
Gets the data object at index `idx`.

num_classes
The number of classes in the dataset.

num_nodes

process()
Processes the dataset to the `self.processed_dir` folder.

processed_file_names
The name of the files to find in the `self.processed_dir` folder in order to skip the processing.

raw_file_names
The name of the files to find in the `self.raw_dir` folder in order to skip the download.

class cogdl.datasets.matlab_matrix.**NetworkEmbeddingCMTYDataset**(*root*, *name*, *url*)
Bases: cogdl.data.dataset.Dataset

download()
Downloads the dataset to the `self.raw_dir` folder.

get(*idx*)
Gets the data object at index `idx`.

num_classes
The number of classes in the dataset.

num_nodes

process()
Processes the dataset to the `self.processed_dir` folder.

processed_file_names
The name of the files to find in the `self.processed_dir` folder in order to skip the processing.

raw_file_names
The name of the files to find in the `self.raw_dir` folder in order to skip the download.

class cogdl.datasets.matlab_matrix.**PPIDataset**
Bases: *cogdl.datasets.matlab_matrix.MatlabMatrix*

class cogdl.datasets.matlab_matrix.**WikipediaDataset**
Bases: *cogdl.datasets.matlab_matrix.MatlabMatrix*

class cogdl.datasets.matlab_matrix.**YoutubeNEDataset**
Bases: *cogdl.datasets.matlab_matrix.NetworkEmbeddingCMTYDataset*

2.8.7 PyG OGB dataset

class cogdl.datasets.ogb.**OGBArxivDataset**
Bases: *cogdl.datasets.ogb.OGBNDataset*

class cogdl.datasets.ogb.**OGBCodeDataset**
Bases: *cogdl.datasets.ogb.OGBGDataset*

class cogdl.datasets.ogb.**OGBGDataset**(*root*, *name*)
Bases: cogdl.data.dataset.Dataset

get(*idx*)
Gets the data object at index `idx`.

get_loader(*args*)

get_subset(*subset*)

num_classes
The number of classes in the dataset.

class cogdl.datasets.ogb.**OGBMAGDataset**
Bases: *cogdl.datasets.ogb.OGBNDataset*

class cogdl.datasets.ogb.**OGBMolbaceDataset**
Bases: *cogdl.datasets.ogb.OGBGDataset*

class cogdl.datasets.ogb.**OGBMolhivDataset**
Bases: *cogdl.datasets.ogb.OGBGDataset*

class cogdl.datasets.ogb.**OGBMolpcbaDataset**
Bases: *cogdl.datasets.ogb.OGBGDataset*

class cogdl.datasets.ogb.**OGBNDataset**(*root*, *name*)
Bases: cogdl.data.dataset.Dataset

get(*idx*)
Gets the data object at index `idx`.

get_evaluator()

get_loss_fn()

class cogdl.datasets.ogb.**OGBPapers100MDataset**
Bases: *cogdl.datasets.ogb.OGBNDataset*

class cogdl.datasets.ogb.**OGBPpaDataset**
Bases: *cogdl.datasets.ogb.OGBGDataset*

class cogdl.datasets.ogb.**OGBProductsDataset**
Bases: *cogdl.datasets.ogb.OGBNDataset*

class cogdl.datasets.ogb.**OGBProteinsDataset**
Bases: *cogdl.datasets.ogb.OGBNDataset*

cogdl.datasets.ogb.**coalesce**(*row*, *col*, *edge_attr=None*)

2.8.8 PyG strategies dataset

This file is borrowed from https://github.com/snap-stanford/pretrain-gnns/

class cogdl.datasets.strategies_data.**BACEDataset**(*transform=None*, *pre_transform=None*, *pre_filter=None*, *empty=False*)
Bases: cogdl.data.dataset.MultiGraphDataset

download()
Downloads the dataset to the self.raw_dir folder.

process()
Processes the dataset to the self.processed_dir folder.

processed_file_names
The name of the files to find in the self.processed_dir folder in order to skip the processing.

raw_file_names
The name of the files to find in the self.raw_dir folder in order to skip the download.

class cogdl.datasets.strategies_data.**BBBPDataset**(*transform=None*, *pre_transform=None*, *pre_filter=None*, *empty=False*)
Bases: cogdl.data.dataset.MultiGraphDataset

download()
Downloads the dataset to the self.raw_dir folder.

process()
Processes the dataset to the self.processed_dir folder.

processed_file_names
The name of the files to find in the self.processed_dir folder in order to skip the processing.

raw_file_names
The name of the files to find in the self.raw_dir folder in order to skip the download.

class cogdl.datasets.strategies_data.**BatchAE**(*batch=None*, ***kwargs*)
Bases: cogdl.data.data.Data

cat_dim(*key*)
Returns the dimension in which the attribute key with content value gets concatenated when creating batches.

Note: This method is for internal use only, and should only be overridden if the batch concatenation process is corrupted for a specific data attribute.

static from_data_list(*data_list*)
Constructs a batch object from a python list holding `torch_geometric.data.Data` objects. The assignment vector `batch` is created on the fly.

num_graphs
Returns the number of graphs in the batch.

class cogdl.datasets.strategies_data.**BatchFinetune**(*batch=None*, ***kwargs*)
Bases: `cogdl.data.data.Data`

static from_data_list(*data_list*)
Constructs a batch object from a python list holding `torch_geometric.data.Data` objects. The assignment vector `batch` is created on the fly.

num_graphs
Returns the number of graphs in the batch.

class cogdl.datasets.strategies_data.**BatchMasking**(*batch=None*, ***kwargs*)
Bases: `cogdl.data.data.Data`

cumsum(*key*, *item*)
If `True`, the attribute `key` with content `item` should be added up cumulatively before concatenated together. .. note:

```
This method is for internal use only, and should only be overridden
if the batch concatenation process is corrupted for a specific data
attribute.
```

static from_data_list(*data_list*)
Constructs a batch object from a python list holding `torch_geometric.data.Data` objects. The assignment vector `batch` is created on the fly.

num_graphs
Returns the number of graphs in the batch.

class cogdl.datasets.strategies_data.**BatchSubstructContext**(*batch=None*, ***kwargs*)
Bases: `cogdl.data.data.Data`

cat_dim(*key*)
Returns the dimension in which the attribute `key` with content `value` gets concatenated when creating batches.

Note: This method is for internal use only, and should only be overridden if the batch concatenation process is corrupted for a specific data attribute.

cumsum(*key*, *item*)
If `True`, the attribute `key` with content `item` should be added up cumulatively before concatenated together. .. note:

```
This method is for internal use only, and should only be overridden
if the batch concatenation process is corrupted for a specific data
attribute.
```

static from_data_list(*data_list*)
Constructs a batch object from a python list holding `torch_geometric.data.Data` objects. The assignment vector `batch` is created on the fly.

num_graphs
Returns the number of graphs in the batch.

class cogdl.datasets.strategies_data.**BioDataset**(*data_type='unsupervised'*, *empty=False*, *transform=None*, *pre_transform=None*, *pre_filter=None*)
Bases: `cogdl.data.dataset.MultiGraphDataset`

download()
Downloads the dataset to the `self.raw_dir` folder.

process()
Processes the dataset to the `self.processed_dir` folder.

processed_file_names
The name of the files to find in the `self.processed_dir` folder in order to skip the processing.

raw_file_names
The name of the files to find in the `self.raw_dir` folder in order to skip the download.

class cogdl.datasets.strategies_data.**ChemExtractSubstructureContextPair**(*k*, *l1*, *l2*)
Bases: `object`

class cogdl.datasets.strategies_data.**DataLoaderAE**(*dataset*, *batch_size=1*, *shuffle=True*, ***kwargs*)
Bases: `torch.utils.data.dataloader.DataLoader`

class cogdl.datasets.strategies_data.**DataLoaderFinetune**(*dataset*, *batch_size=1*, *shuffle=True*, ***kwargs*)
Bases: `torch.utils.data.dataloader.DataLoader`

class cogdl.datasets.strategies_data.**DataLoaderMasking**(*dataset*, *batch_size=1*, *shuffle=True*, ***kwargs*)
Bases: `torch.utils.data.dataloader.DataLoader`

class cogdl.datasets.strategies_data.**DataLoaderSubstructContext**(*dataset*, *batch_size=1*, *shuffle=True*, ***kwargs*)
Bases: `torch.utils.data.dataloader.DataLoader`

class cogdl.datasets.strategies_data.**ExtractSubstructureContextPair**(*l1*, *center=True*)
Bases: `object`

class cogdl.datasets.strategies_data.**MaskAtom**(*num_atom_type*, *num_edge_type*, *mask_rate*, *mask_edge=True*)
Bases: `object`

Borrowed from https://github.com/snap-stanford/pretrain-gnns/

class cogdl.datasets.strategies_data.**MaskEdge**(*mask_rate*)
Bases: `object`

Borrowed from https://github.com/snap-stanford/pretrain-gnns/

class cogdl.datasets.strategies_data.**MoleculeDataset**(*data_type='unsupervised'*, *transform=None*, *pre_transform=None*, *pre_filter=None*, *empty=False*)

Bases: cogdl.data.dataset.MultiGraphDataset

download()
Downloads the dataset to the self.raw_dir folder.

process()
Processes the dataset to the self.processed_dir folder.

processed_file_names
The name of the files to find in the self.processed_dir folder in order to skip the processing.

raw_file_names
The name of the files to find in the self.raw_dir folder in order to skip the download.

class cogdl.datasets.strategies_data.**NegativeEdge**

Bases: object

Borrowed from https://github.com/snap-stanford/pretrain-gnns/

class cogdl.datasets.strategies_data.**TestBioDataset**(*data_type='unsupervised'*, *root='testbio'*, *transform=None*, *pre_transform=None*, *pre_filter=None*)

Bases: cogdl.data.dataset.MultiGraphDataset

class cogdl.datasets.strategies_data.**TestChemDataset**(*data_type='unsupervised'*, *root='testchem'*, *transform=None*, *pre_transform=None*, *pre_filter=None*)

Bases: cogdl.data.dataset.MultiGraphDataset

cogdl.datasets.strategies_data.**graph_data_obj_to_nx**(*data*)

cogdl.datasets.strategies_data.**graph_data_obj_to_nx_simple**(*data*)
Converts graph Data object required by the pytorch geometric package to network x data object. NB: Uses simplified atom and bond features, and represent as indices. NB: possible issues with recapitulating relative stereochemistry since the edges in the nx object are unordered. :param data: pytorch geometric Data object :return: network x object

cogdl.datasets.strategies_data.**nx_to_graph_data_obj**(*g*, *center_id*, *allowable_features_downstream=None*, *allowable_features_pretrain=None*, *node_id_to_go_labels=None*)

cogdl.datasets.strategies_data.**nx_to_graph_data_obj_simple**(*G*)
Converts nx graph to pytorch geometric Data object. Assume node indices are numbered from 0 to num_nodes - 1. NB: Uses simplified atom and bond features, and represent as indices. NB: possible issues with recapitulating relative stereochemistry since the edges in the nx object are unordered. :param G: nx graph obj :return: pytorch geometric Data object

cogdl.datasets.strategies_data.**reset_idxes**(*G*)
Resets node indices such that they are numbered from 0 to num_nodes - 1 :param G: :return: copy of G with relabelled node indices, mapping

2.8.9 TU dataset

class cogdl.datasets.tu_data.**CollabDataset**
Bases: *cogdl.datasets.tu_data.TUDataset*

class cogdl.datasets.tu_data.**ENZYMES**
Bases: *cogdl.datasets.tu_data.TUDataset*

class cogdl.datasets.tu_data.**ImdbBinaryDataset**
Bases: *cogdl.datasets.tu_data.TUDataset*

class cogdl.datasets.tu_data.**ImdbMultiDataset**
Bases: *cogdl.datasets.tu_data.TUDataset*

class cogdl.datasets.tu_data.**MUTAGDataset**
Bases: *cogdl.datasets.tu_data.TUDataset*

class cogdl.datasets.tu_data.**NCT109Dataset**
Bases: *cogdl.datasets.tu_data.TUDataset*

class cogdl.datasets.tu_data.**NCT1Dataset**
Bases: *cogdl.datasets.tu_data.TUDataset*

class cogdl.datasets.tu_data.**PTCMRDataset**
Bases: *cogdl.datasets.tu_data.TUDataset*

class cogdl.datasets.tu_data.**ProtainsDataset**
Bases: *cogdl.datasets.tu_data.TUDataset*

class cogdl.datasets.tu_data.**RedditBinary**
Bases: *cogdl.datasets.tu_data.TUDataset*

class cogdl.datasets.tu_data.**RedditMulti12K**
Bases: *cogdl.datasets.tu_data.TUDataset*

class cogdl.datasets.tu_data.**RedditMulti5K**
Bases: *cogdl.datasets.tu_data.TUDataset*

class cogdl.datasets.tu_data.**TUDataset**(*root*, *name*)
Bases: cogdl.data.dataset.Dataset

download()
Downloads the dataset to the `self.raw_dir` folder.

get(*idx*)
Gets the data object at index `idx`.

num_classes
The number of classes in the dataset.

num_edge_attributes

num_edge_labels

num_node_attributes

num_node_labels

process()
Processes the dataset to the `self.processed_dir` folder.

processed_file_names
The name of the files to find in the `self.processed_dir` folder in order to skip the processing.

raw_file_names
The name of the files to find in the `self.raw_dir` folder in order to skip the download.

url = 'https://www.chrsmrrs.com/graphkerneldatasets'

cogdl.datasets.tu_data.**cat**(*seq*)

cogdl.datasets.tu_data.**coalesce**(*index*, *value*, *m*, *n*)

cogdl.datasets.tu_data.**normalize_feature**(*data*)

cogdl.datasets.tu_data.**parse_txt_array**(*src*, *sep=None*, *start=0*, *end=None*, *dtype=None*, *device=None*)

cogdl.datasets.tu_data.**read_file**(*folder*, *prefix*, *name*, *dtype=None*)

cogdl.datasets.tu_data.**read_tu_data**(*folder*, *prefix*)

cogdl.datasets.tu_data.**read_txt_array**(*path*, *sep=None*, *start=0*, *end=None*, *dtype=None*, *device=None*)

cogdl.datasets.tu_data.**segment**(*src*, *indptr*)

cogdl.datasets.tu_data.**split**(*data*, *batch*)

2.8.10 Module contents

cogdl.datasets.**build_dataset**(*args*)

cogdl.datasets.**build_dataset_from_name**(*dataset*)

cogdl.datasets.**build_dataset_from_path**(*data_path*, *task*)

cogdl.datasets.**register_dataset**(*name*)
New dataset types can be added to cogdl with the *register_dataset()* function decorator.

For example:

```
@register_dataset('my_dataset')
class MyDataset():
    (...)
```

Args: name (str): the name of the dataset

cogdl.datasets.**try_import_dataset**(*dataset*)

2.9 tasks

2.9.1 Base Task

class cogdl.tasks.base_task.**BaseTask**(*args*)
Bases: abc.ABC

static add_args(*parser: argparse.ArgumentParser*)
Add task-specific arguments to the parser.

get_trainer(*model*, *args*)

load_from_pretrained()

save_checkpoint()

set_evaluator(*dataset*)

set_loss_fn(*dataset*)

train()

class cogdl.tasks.base_task.**LoadFrom**
Bases: abc.ABCMeta

2.9.2 Node Classification

class cogdl.tasks.node_classification.**NodeClassification**(*args*, *dataset=None*, *model: Optional[cogdl.models.supervised_model.Supervised = None*)

Bases: *cogdl.tasks.base_task.BaseTask*

Node classification task.

static add_args(*parser: argparse.ArgumentParser*)
Add task-specific arguments to the parser.

train()

2.9.3 Unsupervised Node Classification

class cogdl.tasks.unsupervised_node_classification.**TopKRanker**(*estimator*, ***, *n_jobs=None*)

Bases: sklearn.multiclass.OneVsRestClassifier

predict(*X*, *top_k_list*)
Predict multi-class targets using underlying estimators.

X [(sparse) array-like of shape (n_samples, n_features)] Data.

y [(sparse) array-like of shape (n_samples,) or (n_samples, n_classes)] Predicted multi-class targets.

class cogdl.tasks.unsupervised_node_classification.**UnsupervisedNodeClassification**(*args*, *dataset=Nor*, *model=None*)

Bases: *cogdl.tasks.base_task.BaseTask*

Node classification task.

static add_args(*parser: argparse.ArgumentParser*)
Add task-specific arguments to the parser.

enhance_emb(*G*, *embs*)

save_emb(*embs*)

train()

2.9.4 Heterogeneous Node Classification

class cogdl.tasks.heterogeneous_node_classification.**HeterogeneousNodeClassification**(*args*, *dataset=*, *model=N*)

Bases: *cogdl.tasks.base_task.BaseTask*

Heterogeneous Node classification task.

static add_args(*_: argparse.ArgumentParser*)
Add task-specific arguments to the parser.

train()

2.9.5 Multiplex Node Classification

class cogdl.tasks.multiplex_node_classification.**MultiplexNodeClassification**(*args*, *dataset=None*, *model=None*)

Bases: *cogdl.tasks.base_task.BaseTask*

Node classification task.

static add_args(*parser: argparse.ArgumentParser*)
Add task-specific arguments to the parser.

train()

2.9.6 Link Prediction

class cogdl.tasks.link_prediction.**GNNHomoLinkPrediction**(*args*, *dataset=None*, *model=None*)

Bases: torch.nn.modules.module.Module

static get_link_labels(*num_pos*, *num_neg*, *device=None*)

train()
Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. Dropout, BatchNorm, etc.

Args:

mode (bool): whether to set training mode (True) or evaluation mode (False). Default: True.

Returns: Module: self

static train_test_edge_split(*edge_index*, *num_nodes*, *val_ratio=0.1*, *test_ratio=0.2*)

class cogdl.tasks.link_prediction.**HomoLinkPrediction**(*args*, *dataset=None*, *model=None*)

Bases: torch.nn.modules.module.Module

train()
Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. Dropout, BatchNorm, etc.

Args:

mode (bool): whether to set training mode (True) or evaluation mode (False). Default: True.

Returns: Module: self

class cogdl.tasks.link_prediction.**KGLinkPrediction**(*args*, *dataset=None*, *model=None*)

Bases: torch.nn.modules.module.Module

train()

Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. Dropout, BatchNorm, etc.

Args:

mode (bool): whether to set training mode (True) or evaluation mode (False). Default: True.

Returns: Module: self

class cogdl.tasks.link_prediction.**LinkPrediction**(*args*, *dataset=None*, *model=None*)

Bases: *cogdl.tasks.base_task.BaseTask*

static add_args(*parser*)

Add task-specific arguments to the parser.

load_from_pretrained()

save_checkpoint()

train()

class cogdl.tasks.link_prediction.**TripleLinkPrediction**(*args*, *dataset=None*, *model=None*)

Bases: torch.nn.modules.module.Module

Training process borrowed from *KnowledgeGraphEmbedding<https://github.com/DeepGraphLearning/KnowledgeGraphEmbeda*

train()

Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. Dropout, BatchNorm, etc.

Args:

mode (bool): whether to set training mode (True) or evaluation mode (False). Default: True.

Returns: Module: self

cogdl.tasks.link_prediction.**divide_data**(*input_list*, *division_rate*)

cogdl.tasks.link_prediction.**evaluate**(*embs*, *true_edges*, *false_edges*)

cogdl.tasks.link_prediction.**gen_node_pairs**(*train_data*, *test_data*, *negative_ratio=5*)

cogdl.tasks.link_prediction.**get_score**(*embs*, *node1*, *node2*)

cogdl.tasks.link_prediction.**log_metrics**(*mode*, *step*, *metrics*)

Print the evaluation logs

cogdl.tasks.link_prediction.**randomly_choose_false_edges**(*nodes*, *true_edges*, *num*)

cogdl.tasks.link_prediction.**save_model**(*model*, *optimizer*, *save_variable_list*, *args*)

Save the parameters of the model and the optimizer, as well as some other variables such as step and learning_rate

cogdl.tasks.link_prediction.**select_task**(*model_name=None*, *model=None*)

cogdl.tasks.link_prediction.**set_logger**(*args*)
Write logs to checkpoint and console

2.9.7 Multiplex Link Prediction

class cogdl.tasks.multiplex_link_prediction.**MultiplexLinkPrediction**(*args*, *dataset=None*, *model=None*)

Bases: *cogdl.tasks.base_task.BaseTask*

static add_args(*parser: argparse.ArgumentParser*)
Add task-specific arguments to the parser.

train()

cogdl.tasks.multiplex_link_prediction.**evaluate**(*embs*, *true_edges*, *false_edges*)

cogdl.tasks.multiplex_link_prediction.**get_score**(*embs*, *node1*, *node2*)

2.9.8 Graph Classification

class cogdl.tasks.graph_classification.**GraphClassification**(*args*, *dataset=None*, *model=None*)

Bases: *cogdl.tasks.base_task.BaseTask*

Superiviced graph classification task.

static add_args(*parser: argparse.ArgumentParser*)
Add task-specific arguments to the parser.

generate_data(*dataset*, *args*)

train()

cogdl.tasks.graph_classification.**node_degree_as_feature**(*data*)
Set each node feature as one-hot encoding of degree :param data: a list of class Data :return: a list of class Data

cogdl.tasks.graph_classification.**uniform_node_feature**(*data*)
Set each node feature to the same

2.9.9 Unsupervised Graph Classification

class cogdl.tasks.unsupervised_graph_classification.**UnsupervisedGraphClassification**(*args*, *dataset=N*
model=N

Bases: *cogdl.tasks.base_task.BaseTask*

Unsupervised graph classification

static add_args(*parser: argparse.ArgumentParser*)
Add task-specific arguments to the parser.

save_emb(*embs*)

train()

2.9.10 Attributed Graph Clustering

class cogdl.tasks.attributed_graph_clustering.**AttributedGraphClustering**(*args*, *dataset=None*, *_=None*)

Bases: *cogdl.tasks.base_task.BaseTask*

Attributed graph clustring task.

static add_args(*parser: argparse.ArgumentParser*)
Add task-specific arguments to the parser.

train() → Dict[str, float]

2.9.11 Similarity Search

class cogdl.tasks.similarity_search.**SimilaritySearch**(*args*, *dataset=None*, *model=None*)

Bases: *cogdl.tasks.base_task.BaseTask*

Similarity Search task.

static add_args(*_: argparse.ArgumentParser*)
Add task-specific arguments to the parser.

train()

2.9.12 Pretrain

class cogdl.tasks.pretrain.**PretrainTask**(*args*)

Bases: *cogdl.tasks.base_task.BaseTask*

static add_args(*_: argparse.ArgumentParser*)
Add task-specific arguments to the parser.

train()

2.9.13 Task Module

cogdl.tasks.**build_task**(*args*, *dataset=None*, *model=None*)

cogdl.tasks.**register_task**(*name*)
New task types can be added to cogdl with the *register_task()* function decorator.

For example:

```
@register_task('node_classification')
class NodeClassification(BaseTask):
    (...)
```

Args: name (str): the name of the task

2.10 models

2.10.1 BaseModel

class cogdl.models.base_model.**BaseModel**
Bases: torch.nn.modules.module.Module

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(**args*)

static get_trainer(*task: Any*, *args: Any*) → Optional[Type[cogdl.trainers.base_trainer.BaseTrainer]]

graph_classification_loss(*batch*)

node_classification_loss(*data*, *mask=None*)

predict(*data*)

set_device(*device*)

set_loss_fn(*loss_fn*)

2.10.2 Supervised Model

class cogdl.models.supervised_model.**SupervisedHeterogeneousNodeClassificationModel**
Bases: *cogdl.models.base_model.BaseModel*, abc.ABC

evaluate(*data: Any*, *nodes: Any*, *targets: Any*) → Any

static get_trainer(*taskType: Any*, *args: Any*) → Optional[Type[SupervisedHeterogeneousNodeClassificationTrainer]]

loss(*data: Any*) → Any

class cogdl.models.supervised_model.**SupervisedHomogeneousNodeClassificationModel**
Bases: *cogdl.models.base_model.BaseModel*, abc.ABC

static get_trainer(*taskType: Any*, *args: Any*) → Optional[Type[SupervisedHomogeneousNodeClassificationTrainer]]

loss(*data: Any*) → Any

predict(*data: Any*) → Any

class cogdl.models.supervised_model.**SupervisedModel**
Bases: *cogdl.models.base_model.BaseModel*, abc.ABC

loss(*data: Any*) → Any

2.10.3 Embedding Model

class cogdl.models.emb.hope.**HOPE**(*dimension*, *beta*)
Bases: *cogdl.models.base_model.BaseModel*

The HOPE model from the "Grarep: Asymmetric transitivity preserving graph embedding" paper.

Args: hidden_size (int) : The dimension of node representation. beta (float) : Parameter in katz decomposition.

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

model_name = 'hope'

train(*G*)
The author claim that Katz has superior performance in related tasks S_katz = (M_g)^-1 * M_l = (I - beta*A)^-1 * beta*A = (I - beta*A)^-1 * (I - (I -beta*A)) = (I - beta*A)^-1 - I

class cogdl.models.emb.spectral.**Spectral**(*dimension*)
Bases: *cogdl.models.base_model.BaseModel*

The Spectral clustering model from the "Leveraging social media networks for classication" paper

Args: hidden_size (int) : The dimension of node representation.

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

model_name = 'spectral'

train(*G*)
Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. `Dropout`, `BatchNorm`, etc.

Args:

mode (bool): whether to set training mode (`True`) or evaluation mode (`False`). Default: `True`.

Returns: Module: self

class cogdl.models.emb.hin2vec.**Hin2vec**(*hidden_dim*, *walk_length*, *walk_num*, *batch_size*, *hop*, *negative*, *epochs*, *lr*, *cpu=True*)
Bases: *cogdl.models.base_model.BaseModel*

The Hin2vec model from the "HIN2Vec: Explore Meta-paths in Heterogeneous Information Networks for Representation Learning" paper.

Args: hidden_size (int) : The dimension of node representation. walk_length (int) : The walk length. walk_num (int) : The number of walks to sample for each node. batch_size (int) : The batch size of training in Hin2vec. hop (int) : The number of hop to construct training samples in Hin2vec. negative (int) : The number of nagative samples for each meta2path pair. epochs (int) : The number of training iteration. lr (float) : The initial learning rate of SGD. cpu (bool) : Use CPU or GPU to train hin2vec.

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

model_name = 'hin2vec'

train(*G*, *node_type*)
Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. `Dropout`, `BatchNorm`, etc.

Args:

mode (bool): whether to set training mode (`True`) or evaluation mode (`False`). Default: `True`.

Returns: Module: self

class cogdl.models.emb.netmf.**NetMF**(*dimension*, *window_size*, *rank*, *negative*, *is_large=False*)

Bases: *cogdl.models.base_model.BaseModel*

The NetMF model from the "Network Embedding as Matrix Factorization: Unifying DeepWalk, LINE, PTE, and node2vec" paper.

Args: hidden_size (int) : The dimension of node representation. window_size (int) : The actual context size which is considered in language model. rank (int) : The rank in approximate normalized laplacian. negative (int) : The number of nagative samples in negative sampling. is-large (bool) : When window size is large, use approximated deepwalk matrix to decompose.

static add_args(*parser*)

Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)

Build a new model instance.

model_name = 'netmf'

train(*G*)

Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. `Dropout`, `BatchNorm`, etc.

Args:

mode (bool): whether to set training mode (`True`) or evaluation mode (`False`). Default: `True`.

Returns: Module: self

class cogdl.models.emb.distmult.**DistMult**(*nentity*, *nrelation*, *hidden_dim*, *gamma*, *double_entity_embedding=False*, *double_relation_embedding=False*)

Bases: cogdl.models.emb.knowledge_base.KGEModel

The DistMult model from the ICLR 2015 paper *"EMBEDDING ENTITIES AND RELATIONS FOR LEARNING AND INFERENCE IN KNOWLEDGE BASES" <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/ICLR2015_updated.pdf>* borrowed from *KnowledgeGraphEmbedding<https://github.com/DeepGraphLearning/KnowledgeGraphEmbedding>*

model_name = 'distmult'

score(*head*, *relation*, *tail*, *mode*)

class cogdl.models.emb.transe.**TransE**(*nentity*, *nrelation*, *hidden_dim*, *gamma*, *double_entity_embedding=False*, *double_relation_embedding=False*)

Bases: cogdl.models.emb.knowledge_base.KGEModel

The TransE model from paper *"Translating Embeddings for Modeling Multi-relational Data" <http://papers.nips.cc/paper/5071-translating-embeddings-for-modeling-multi-relational-data.pdf>* borrowed from *KnowledgeGraphEmbedding<https://github.com/DeepGraphLearning/KnowledgeGraphEmbedding>*

model_name = 'transe'

score(*head*, *relation*, *tail*, *mode*)

class cogdl.models.emb.deepwalk.**DeepWalk**(*dimension*, *walk_length*, *walk_num*, *window_size*, *worker*, *iteration*)

Bases: *cogdl.models.base_model.BaseModel*

The DeepWalk model from the "DeepWalk: Online Learning of Social Representations" paper

Args: hidden_size (int) : The dimension of node representation. walk_length (int) : The walk length. walk_num (int) : The number of walks to sample for each node. window_size (int) : The actual context size which is considered in language model. worker (int) : The number of workers for word2vec. iteration (int) : The number of training iteration in word2vec.

static add_args(*parser: argparse.ArgumentParser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*) → cogdl.models.emb.deepwalk.DeepWalk
Build a new model instance.

model_name = 'deepwalk'

train(*G: networkx.classes.graph.Graph*, *embedding_model_creator=<class 'gensim.models.word2vec.Word2Vec'>*)
Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. `Dropout`, `BatchNorm`, etc.

Args:

mode (bool): whether to set training mode (`True`) or evaluation mode (`False`). Default: `True`.

Returns: Module: self

class cogdl.models.emb.rotate.**RotatE**(*nentity*, *nrelation*, *hidden_dim*, *gamma*, *double_entity_embedding=False*, *double_relation_embedding=False*)

Bases: `cogdl.models.emb.knowledge_base.KGEModel`

Implementation of RotatE model from the paper *"RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space" <https://openreview.net/forum?id=HkgEQnRqYQ>*. borrowed from *KnowledgeGraphEmbedding<https://github.com/DeepGraphLearning/KnowledgeGraphEmbedding>*

model_name = 'rotate'

score(*head*, *relation*, *tail*, *mode*)

class cogdl.models.emb.gatne.**GATNE**(*dimension*, *walk_length*, *walk_num*, *window_size*, *worker*, *epoch*, *batch_size*, *edge_dim*, *att_dim*, *negative_samples*, *neighbor_samples*, *schema*)

Bases: *cogdl.models.base_model.BaseModel*

The GATNE model from the "Representation Learning for Attributed Multiplex Heterogeneous Network" paper

Args: walk_length (int) : The walk length. walk_num (int) : The number of walks to sample for each node. window_size (int) : The actual context size which is considered in language model. worker (int) : The number of workers for word2vec. epoch (int) : The number of training epochs. batch_size (int) : The size of each training batch. edge_dim (int) : Number of edge embedding dimensions. att_dim (int) : Number of attention dimensions. negative_samples (int) : Negative samples for optimization. neighbor_samples (int) : Neighbor samples for aggregation schema (str) : The metapath schema used in model. Metapaths

are splited with ",", while each node type are connected with "-" in each metapath. For example:"0-1-0,0-1-2-1-0"

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

model_name = 'gatne'

train(*network_data*)
Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. `Dropout`, `BatchNorm`, etc.

Args:

mode (bool): whether to set training mode (`True`) or evaluation mode (`False`). Default: `True`.

Returns: Module: self

class cogdl.models.emb.dgk.**DeepGraphKernel**(*hidden_dim*, *min_count*, *window_size*, *sampling_rate*, *rounds*, *epoch*, *alpha*, *n_workers=4*)
Bases: *cogdl.models.base_model.BaseModel*

The Hin2vec model from the "Deep Graph Kernels" paper.

Args: hidden_size (int) : The dimension of node representation. min_count (int) : Parameter in word2vec. window (int) : The actual context size which is considered in language model. sampling_rate (float) : Parameter in word2vec. iteration (int) : The number of iteration in WL method. epoch (int) : The number of training iteration. alpha (float) : The learning rate of word2vec.

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

static feature_extractor(*data*, *rounds*, *name*)

forward(*graphs*, ***kwargs*)

model_name = 'dgk'

save_embedding(*output_path*)

static wl_iterations(*graph*, *features*, *rounds*)

class cogdl.models.emb.grarep.**GraRep**(*dimension*, *step*)
Bases: *cogdl.models.base_model.BaseModel*

The GraRep model from the "Grarep: Learning graph representations with global structural information" paper.

Args: hidden_size (int) : The dimension of node representation. step (int) : The maximum order of transitition probability.

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

model_name = 'grarep'

train(*G*)
Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. `Dropout`, `BatchNorm`, etc.

Args:

mode (bool): whether to set training mode (`True`) or evaluation mode (`False`). Default: `True`.

Returns: Module: self

class cogdl.models.emb.dngr.**DNGR**(*hidden_size1*, *hidden_size2*, *noise*, *alpha*, *step*, *max_epoch*, *lr*, *cpu*)
Bases: *cogdl.models.base_model.BaseModel*

The DNGR model from the "Deep Neural Networks for Learning Graph Representations" paper

Args: hidden_size1 (int) : The size of the first hidden layer. hidden_size2 (int) : The size of the second hidden layer. noise (float) : Denoise rate of DAE. alpha (float) : Parameter in DNGR. step (int) : The max step in random surfing. max_epoch (int) : The max epoches in training step. lr (float) : Learning rate in DNGR.

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

get_denoised_matrix(*mat*)

get_emb(*matrix*)

get_ppmi_matrix(*mat*)

model_name = 'dngr'

random_surfing(*adj_matrix*)

scale_matrix(*mat*)

train(*G*)
Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. `Dropout`, `BatchNorm`, etc.

Args:

mode (bool): whether to set training mode (`True`) or evaluation mode (`False`). Default: `True`.

Returns: Module: self

class cogdl.models.emb.pronepp.**ProNEPP**(*filter_types*, *svd*, *search*, *max_evals=None*, *loss_type=None*, *n_workers=None*)
Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

model_name = 'prone++'

class cogdl.models.emb.graph2vec.**Graph2Vec**(*dimension*, *min_count*, *window_size*, *dm*, *sampling_rate*, *rounds*, *epoch*, *lr*, *worker=4*)

Bases: *cogdl.models.base_model.BaseModel*

The Graph2Vec model from the "graph2vec: Learning Distributed Representations of Graphs" paper

Args: hidden_size (int) : The dimension of node representation. min_count (int) : Parameter in doc2vec. window_size (int) : The actual context size which is considered in language model. sampling_rate (float) : Parameter in doc2vec. dm (int) : Parameter in doc2vec. iteration (int) : The number of iteration in WL method. epoch (int) : The max epoches in training step. lr (float) : Learning rate in doc2vec.

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

static feature_extractor(*data*, *rounds*, *name*)

forward(*graphs*, ***kwargs*)

model_name = 'graph2vec'

save_embedding(*output_path*)

static wl_iterations(*graph*, *features*, *rounds*)

class cogdl.models.emb.metapath2vec.**Metapath2vec**(*dimension*, *walk_length*, *walk_num*, *window_size*, *worker*, *iteration*, *schema*)

Bases: *cogdl.models.base_model.BaseModel*

The Metapath2vec model from the "metapath2vec: Scalable Representation Learning for Heterogeneous Networks" paper

Args: hidden_size (int) : The dimension of node representation. walk_length (int) : The walk length. walk_num (int) : The number of walks to sample for each node. window_size (int) : The actual context size which is considered in language model. worker (int) : The number of workers for word2vec. iteration (int) : The number of training iteration in word2vec. schema (str) : The metapath schema used in model. Metapaths are splited with ",", while each node type are connected with "-" in each metapath. For example:"0-1-0,0-2-0,1-0-2-0-1".

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

model_name = 'metapath2vec'

train(*G*, *node_type*)
Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. `Dropout`, `BatchNorm`, etc.

Args:

mode (bool): whether to set training mode (`True`) or evaluation mode (`False`). Default: `True`.

Returns: Module: self

class cogdl.models.emb.node2vec.**Node2vec**(*dimension*, *walk_length*, *walk_num*, *window_size*, *worker*, *iteration*, *p*, *q*)

Bases: *cogdl.models.base_model.BaseModel*

The node2vec model from the "node2vec: Scalable feature learning for networks" paper

Args: hidden_size (int) : The dimension of node representation. walk_length (int) : The walk length. walk_num (int) : The number of walks to sample for each node. window_size (int) : The actual context size which is considered in language model. worker (int) : The number of workers for word2vec. iteration (int) : The number of training iteration in word2vec. p (float) : Parameter in node2vec. q (float) : Parameter in node2vec.

static add_args(*parser*)

Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)

Build a new model instance.

model_name = 'node2vec'

train(*G*)

Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. `Dropout`, `BatchNorm`, etc.

Args:

mode (bool): whether to set training mode (`True`) or evaluation mode (`False`). Default: `True`.

Returns: Module: self

class cogdl.models.emb.complex.**ComplEx**(*nentity*, *nrelation*, *hidden_dim*, *gamma*, *double_entity_embedding=False*, *double_relation_embedding=False*)

Bases: `cogdl.models.emb.knowledge_base.KGEModel`

the implementation of ComplEx model from the paper *"Complex Embeddings for Simple Link Prediction"<http://proceedings.mlr.press/v48/trouillon16.pdf>* borrowed from *KnowledgeGraphEmbedding<https://github.com/DeepGraphLearning/KnowledgeGraphEmbedding>*

model_name = 'complex'

score(*head*, *relation*, *tail*, *mode*)

class cogdl.models.emb.pte.**PTE**(*dimension*, *walk_length*, *walk_num*, *negative*, *batch_size*, *alpha*)

Bases: *cogdl.models.base_model.BaseModel*

The PTE model from the "PTE: Predictive Text Embedding through Large-scale Heterogeneous Text Networks" paper.

Args: hidden_size (int) : The dimension of node representation. walk_length (int) : The walk length. walk_num (int) : The number of walks to sample for each node. negative (int) : The number of negative samples for each edge. batch_size (int) : The batch size of training in PTE. alpha (float) : The initial learning rate of SGD.

static add_args(*parser*)

Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)

Build a new model instance.

model_name = 'pte'

train(*G*, *node_type*)
Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. `Dropout`, `BatchNorm`, etc.

Args:

mode (bool): whether to set training mode (`True`) or evaluation mode (`False`). Default: `True`.

Returns: Module: self

class cogdl.models.emb.netsmf.**NetSMF**(*dimension*, *window_size*, *negative*, *num_round*, *worker*)
Bases: *cogdl.models.base_model.BaseModel*

The NetSMF model from the "NetSMF: Large-Scale Network Embedding as Sparse Matrix Factorization" paper.

Args: hidden_size (int) : The dimension of node representation. window_size (int) : The actual context size which is considered in language model. negative (int) : The number of nagative samples in negative sampling. num_round (int) : The number of round in NetSMF. worker (int) : The number of workers for NetSMF.

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

model_name = 'netsmf'

train(*G*)
Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. `Dropout`, `BatchNorm`, etc.

Args:

mode (bool): whether to set training mode (`True`) or evaluation mode (`False`). Default: `True`.

Returns: Module: self

class cogdl.models.emb.line.**LINE**(*dimension*, *walk_length*, *walk_num*, *negative*, *batch_size*, *alpha*, *order*)
Bases: *cogdl.models.base_model.BaseModel*

The LINE model from the "Line: Large-scale information network embedding" paper.

Args: hidden_size (int) : The dimension of node representation. walk_length (int) : The walk length. walk_num (int) : The number of walks to sample for each node. negative (int) : The number of nagative samples for each edge. batch_size (int) : The batch size of training in LINE. alpha (float) : The initial learning rate of SGD. order (int) : 1 represents perserving 1-st order proximity, 2 represents 2-nd, while 3 means both of them (each of them having dimension/2 node representation).

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

model_name = 'line'

train(*G*)
Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. `Dropout`, `BatchNorm`, etc.

Args:

mode (bool): whether to set training mode (`True`) or evaluation mode (`False`). Default: `True`.

Returns: Module: self

class cogdl.models.emb.sdne.**SDNE**(*hidden_size1*, *hidden_size2*, *droput*, *alpha*, *beta*, *nu1*, *nu2*, *max_epoch*, *lr*, *cpu*)
Bases: *cogdl.models.base_model.BaseModel*

The SDNE model from the "Structural Deep Network Embedding" paper

Args: hidden_size1 (int) : The size of the first hidden layer. hidden_size2 (int) : The size of the second hidden layer. droput (float) : Droput rate. alpha (float) : Trade-off parameter between 1-st and 2-nd order objective function in SDNE. beta (float) : Parameter of 2-nd order objective function in SDNE. nu1 (float) : Parameter of l1 normlization in SDNE. nu2 (float) : Parameter of l2 normlization in SDNE. max_epoch (int) : The max epoches in training step. lr (float) : Learning rate in SDNE. cpu (bool) : Use CPU or GPU to train hin2vec.

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

model_name = 'sdne'

train(*G*)
Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. `Dropout`, `BatchNorm`, etc.

Args:

mode (bool): whether to set training mode (`True`) or evaluation mode (`False`). Default: `True`.

Returns: Module: self

class cogdl.models.emb.prone.**ProNE**(*dimension*, *step*, *mu*, *theta*)
Bases: *cogdl.models.base_model.BaseModel*

The ProNE model from the "ProNE: Fast and Scalable Network Representation Learning" paper.

Args: hidden_size (int) : The dimension of node representation. step (int) : The number of items in the chebyshev expansion. mu (float) : Parameter in ProNE. theta (float) : Parameter in ProNE.

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

model_name = 'prone'

train(*G*)
Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. `Dropout`, `BatchNorm`, etc.

Args:

mode (bool): whether to set training mode (`True`) or evaluation mode (`False`). Default: `True`.

Returns: Module: self

2.10.4 GNN Model

class cogdl.models.nn.dgi.**DGIModel**(*in_feats*, *hidden_size*, *activation*)

Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

embed(*data*, *msk=None*)

forward(*x*, *edge_index*, *edge_attr=None*)

static get_trainer(*task*, *args*)

loss(*data*)

model_name = 'dgi'

node_classification_loss(*data*)

class cogdl.models.nn.mvgrl.**MVGRL**(*in_feats*, *hidden_size*, *sample_size=2000*, *batch_size=4*, *sparse=False*, *dataset='cora'*)

Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

embed(*data*, *msk=None*)

forward(*x*, *edge_index*, *edge_attr=None*)

static get_trainer(*taskType*, *args*)

loss(*data*)

model_name = 'mvgrl'

node_classification_loss(*data*)

preprocess(*x*, *edge_index*, *edge_attr=None*)

class cogdl.models.nn.patchy_san.**PatchySAN**(*batch_size*, *num_features*, *num_classes*, *num_sample*, *stride*, *num_neighbor*, *iteration*)

Bases: *cogdl.models.base_model.BaseModel*

The Patchy-SAN model from the "Learning Convolutional Neural Networks for Graphs" paper.

Args: batch_size (int) : The batch size of training. sample (int) : Number of chosen vertexes. stride (int) : Node selection stride. neighbor (int) : The number of neighbor for each node. iteration (int) : The number of training iteration.

static add_args(*parser*)
Add model-specific arguments to the parser.

build_model(*num_channel*, *num_sample*, *num_neighbor*, *num_class*)

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*batch*)

model_name = 'patchy_san'

classmethod split_dataset(*dataset*, *args*)

class cogdl.models.nn.pyg_cheb.**Chebyshev**(*in_feats*, *hidden_size*, *out_feats*, *num_layers*, *dropout*, *filter_size*)
Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*x*, *edge_index*)

model_name = 'chebyshev'

predict(*data*)

class cogdl.models.nn.gcn.**TKipfGCN**(*in_feats*, *hidden_size*, *out_feats*, *num_layers*, *dropout*)
Bases: *cogdl.models.base_model.BaseModel*

The GCN model from the "Semi-Supervised Classification with Graph Convolutional Networks" paper

Args: in_features (int) : Number of input features. out_features (int) : Number of classes. hidden_size (int) : The dimension of node representation. dropout (float) : Dropout rate for model training.

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*x*, *edge_index*, *edge_weight=None*)

get_embeddings(*x*, *edge_index*)

model_name = 'gcn'

predict(*data*)

class cogdl.models.nn.gdc_gcn.**GDC_GCN**(*nfeat*, *nhid*, *nclass*, *dropout*, *alpha*, *t*, *k*, *eps*, *gdctype*)
Bases: *cogdl.models.base_model.BaseModel*

The GDC model from the "Diffusion Improves Graph Learning" paper, with the PPR and heat matrix variants combined with GCN

Args: num_features (int) : Number of input features in ppr-preprocessed dataset. num_classes (int) : Number of classes. hidden_size (int) : The dimension of node representation. dropout (float) : Dropout rate for model training. alpha (float) : PPR polynomial filter param, 0 to 1. t (float) : Heat polynomial filter param

k (int) : Top k nodes retained during sparsification. eps (float) : Threshold for clipping. gdc_type (str) : "none", "ppr", "heat"

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*x*, *edge_index*)

model_name = 'gdc_gcn'

node_classification_loss(*data*)

predict(*data=None*)

preprocessing(*data*, *gdc_type='ppr'*)

reset_data(*data*)

class cogdl.models.nn.pyg_hgpsl.**HGPSL**(*num_features*, *num_classes*, *hidden_size*, *dropout*, *pooling*, *sample_neighbor*, *sparse_attention*, *structure_learning*, *lamb*)
Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*data*)

model_name = 'hgpsl'

classmethod split_dataset(*dataset*, *args*)

class cogdl.models.nn.graphsage.**Graphsage**(*num_features*, *num_classes*, *hidden_size*, *num_layers*, *sample_size*, *dropout*)
Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(**args*)

static get_trainer(*task: Any*, *args: Any*)

inference(*x_all*, *data_loader*)

mini_forward(*x*, *edge_index*)

mini_loss(*data*)

model_name = 'graphsage'

node_classification_loss(**args*)

predict(*data*)

sampling(*edge_index*, *num_sample*)

set_data_device(*device*)

class cogdl.models.nn.compgcn.**LinkPredictCompGCN**(*num_entities*, *num_rels*, *hidden_size*, *num_bases=0*, *layers=1*, *sampling_rate=0.01*, *score_func='conve'*, *penalty=0.001*, *dropout=0.0*, *lbl_smooth=0.1*)

Bases: *cogdl.layers.link_prediction_module.GNNLinkPredict*, *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

add_reverse_edges(*edge_index*, *edge_types*)

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*edge_index*, *edge_types*)

loss(*data*, *split='train'*)

model_name = 'compgcn'

predict(*edge_index*, *edge_types*)

class cogdl.models.nn.drgcn.**DrGCN**(*num_features*, *num_classes*, *hidden_size*, *num_layers*, *dropout*)

Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*x*, *edge_index*)

model_name = 'drgcn'

predict(*data*)

class cogdl.models.nn.pyg_gpt_gnn.**GPT_GNN**

Bases: *cogdl.models.supervised_model.SupervisedHomogeneousNodeClassificationModel*, *cogdl.models.supervised_model.SupervisedHeterogeneousNodeClassificationModel*

static add_args(*parser*)
Add task-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

evaluate(*data: Any*, *nodes: Any*, *targets: Any*) → Any

static get_trainer(*taskType: Any*, *args*) → Optional[Type[Union[cogdl.trainers.gpt_gnn_trainer.GPT_GNNHomogeneo
cogdl.trainers.gpt_gnn_trainer.GPT_GNNHeterogeneousTrainer]]]

loss(*data: Any*) → Any

model_name = 'gpt_gnn'

predict(*data: Any*) → Any

class cogdl.models.nn.pyg_graph_unet.**GraphUnet**(*in_feats: int, hidden_size: int, out_feats: int, pooling_layer: int, pooling_rates: List[float], n_dropout: float = 0.5, adj_dropout: float = 0.3, activation: str = 'elu', improved: bool = False, aug_adj: bool = False*)

Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*x: torch.Tensor*, *edge_index: torch.Tensor*, *edge_attr: Optional[torch.Tensor] = None*) → torch.Tensor

model_name = 'unet'

predict(*data*)

class cogdl.models.nn.gcnmix.**GCNMix**(*in_feat*, *hidden_size*, *num_classes*, *k*, *temperature*, *alpha*, *rampup_starts*, *rampup_ends*, *final_consistency_weight*, *ema_decay*, *dropout*)

Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*x*, *edge_index*)

forward_ema(*x*, *edge_index*)

model_name = 'gcnmix'

node_classification_loss(*data*)

predict(*data*)

class cogdl.models.nn.diffpool.**DiffPool**(*in_feats*, *hidden_dim*, *embed_dim*, *num_classes*, *num_layers*, *num_pool_layers*, *assign_dim*, *pooling_ratio*, *batch_size*, *dropout=0.5*, *no_link_pred=True*, *concat=False*, *use_bn=False*)

Bases: *cogdl.models.base_model.BaseModel*

DIFFPOOL from paper Hierarchical Graph Representation Learning with Differentiable Pooling.

in_feats [int] Size of each input sample.

hidden_dim [int] Size of hidden layer dimension of GNN.

embed_dim [int] Size of embeded node feature, output size of GNN.

num_classes [int] Number of target classes.

num_layers [int] Number of GNN layers.

num_pool_layers [int] Number of pooling.

assign_dim [int] Embedding size after the first pooling.

pooling_ratio [float] Size of each poolling ratio.

batch_size [int] Size of each mini-batch.

dropout [float, optional] Size of dropout, default: *0.5*.

no_link_pred [bool, optional] If True, use link prediction loss, default: *True*.

static add_args(*parser*)
Add model-specific arguments to the parser.

after_pooling_forward(*gnn_layers*, *adj*, *x*, *concat=False*)

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*batch*)

graph_classificatoin_loss(*batch*)

model_name = 'diffpool'

reset_parameters()

classmethod split_dataset(*dataset*, *args*)

class cogdl.models.nn.gcnii.**GCNII**(*in_feats*, *hidden_size*, *out_feats*, *num_layers*, *dropout=0.5*, *alpha=0.1*, *lmbda=1*, *wd1=0.0*, *wd2=0.0*, *residual=False*)
Bases: *cogdl.models.base_model.BaseModel*

Implementation of GCNII in paper *"Simple and Deep Graph Convolutional Networks"* <https://arxiv.org/abs/2007.02133>.

in_feats [int] Size of each input sample

hidden_size [int] Size of each hidden unit

out_feats [int] Size of each out sample

num_layers : int dropout : float alpha : float

Parameter of initial residual connection

lmbda [float] Parameter of identity mapping

wd1 [float] Weight-decay for Fully-connected layers

wd2 [float] Weight-decay for convolutional layers

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*x*, *edge_index*, *edge_attr=None*)

get_optimizer(*args*)

model_name = 'gcnii'

predict(*data*)

class cogdl.models.nn.sign.**MLP**(*num_features*, *hidden_size*, *num_classes*, *num_layers*, *dropout*, *dropedge_rate*, *undirected*, *num_propagations*, *asymm_norm*, *set_diag*, *remove_diag*)
Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*x*, *edge_index*)

model_name = 'sign'

node_classification_loss(*data*, *mask=None*)

predict(*data*)

reset_parameters()

class cogdl.models.nn.pyg_gcn.**GCN**(*num_features*, *num_classes*, *hidden_size*, *num_layers*, *dropout*)
Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*x*, *edge_index*, *weight=None*)

get_embeddings(*x*, *edge_index*, *weight=None*)

get_trainer(*task*, *args*)

model_name = 'pyg_gcn'

predict(*data*)

class cogdl.models.nn.mixhop.**MixHop**(*num_features*, *num_classes*, *dropout*, *layer1_pows*, *layer2_pows*)
Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*x*, *edge_index*)

model_name = 'mixhop'

predict(*data*)

class cogdl.models.nn.gat.**GAT**(*in_feats*, *hidden_size*, *out_features*, *num_layers*, *dropout*, *alpha*, *nhead*, *residual*, *last_nhead*, *fast_mode=False*)
Bases: *cogdl.models.base_model.BaseModel*

The GAT model from the "Graph Attention Networks" paper

Args: num_features (int) : Number of input features. num_classes (int) : Number of classes. hidden_size (int) : The dimension of node representation. dropout (float) : Dropout rate for model training. alpha (float) : Coefficient of leaky_relu. nheads (int) : Number of attention heads.

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*x*, *edge_index*)

model_name = 'gat'

predict(*data*)

class cogdl.models.nn.han.**HAN**(*num_edge*, *w_in*, *w_out*, *num_class*, *num_nodes*, *num_layers*)
Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

evaluate(*data*, *nodes*, *targets*)

forward(*A*, *X*, *target_x*, *target*)

loss(*data*)

model_name = 'han'

class cogdl.models.nn.ppnp.**PPNP**(*nfeat*, *nhid*, *nclass*, *num_layers*, *dropout*, *propagation*, *alpha*, *niter*, *cache=True*)
Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*x*, *adj*)

model_name = 'ppnp'

predict(*data*)

class cogdl.models.nn.grace.**GRACE**(*in_feats: int, hidden_size: int, proj_hidden_size: int, num_layers: int, drop_feature_rates: List[float], drop_edge_rates: List[float], tau: float = 0.5, activation: str = 'relu', batch_size: int = -1*)
Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

batched_loss(*z1: torch.Tensor*, *z2: torch.Tensor*, *batch_size: int*)

classmethod build_model_from_args(*args*)
Build a new model instance.

contrastive_loss(*z1: torch.Tensor*, *z2: torch.Tensor*)

drop_adj(*edge_index: torch.Tensor*, *edge_weight: Optional[torch.Tensor] = None*, *drop_rate: float = 0.5*)

drop_feature(*x: torch.Tensor*, *droprate: float*)

embed(*data*)

forward(*x: torch.Tensor*, *edge_index: torch.Tensor*, *edge_weight: Optional[torch.Tensor] = None*)

static get_trainer(*task*, *args*)

model_name = 'grace'

node_classification_loss(*data*)

prop(*x: torch.Tensor*, *edge_index: torch.Tensor*, *edge_weight: Optional[torch.Tensor] = None*, *drop_feature_rate: float = 0.0*, *drop_edge_rate: float = 0.0*)

class cogdl.models.nn.dgl_jknet.**JKNet**(*in_features*, *out_features*, *n_layers*, *n_units*, *node_aggregation*, *layer_aggregation*)

Bases: *cogdl.models.supervised_model.SupervisedHomogeneousNodeClassificationModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*graph*, *x*)

static get_trainer(*taskType*, *args*)

loss(*data*)

model_name = 'jknet'

predict(*data*)

set_graph(*graph*)

class cogdl.models.nn.pprgo.**PPRGo**(*in_feats*, *hidden_size*, *out_feats*, *num_layers*, *alpha*, *dropout*, *activation='relu'*, *nprop=2*)

Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*x*, *targets*, *ppr_scores*)

static get_trainer(*taskType: Any*, *args: Any*)

model_name = 'pprgo'

node_classification_loss(*x*, *targets*, *ppr_scores*, *y*)

predict(*x*, *edge_index*, *batch_size*, *norm_func*)

class cogdl.models.nn.gin.**GIN**(*num_layers*, *in_feats*, *out_feats*, *hidden_dim*, *num_mlp_layers*, *eps=0*, *pooling='sum'*, *train_eps=False*, *dropout=0.5*)

Bases: *cogdl.models.base_model.BaseModel*

Graph Isomorphism Network from paper "How Powerful are Graph Neural Networks?".

Args:

num_layers [int] Number of GIN layers

in_feats [int] Size of each input sample

out_feats [int] Size of each output sample

hidden_dim [int] Size of each hidden layer dimension

num_mlp_layers [int] Number of MLP layers

eps [float32, optional] Initial *epsilon* value, default: `0`

pooling [str, optional] Aggregator type to use, default:`sum`

train_eps [bool, optional] If True, *epsilon* will be a learnable parameter, default: `True`

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*batch*)

model_name = 'gin'

classmethod split_dataset(*dataset*, *args*)

class cogdl.models.nn.pyg_dgcnn.**DGCNN**(*in_feats*, *hidden_dim*, *out_feats*, *k=20*, *dropout=0.5*)
Bases: *cogdl.models.base_model.BaseModel*

EdgeConv and DynamicGraph in paper *"Dynamic Graph CNN for Learning on Point Clouds" <https://arxiv.org/pdf/1801.07829.pdf>__* .

in_feats [int] Size of each input sample.

out_feats [int] Size of each output sample.

hidden_dim [int] Dimension of hidden layer embedding.

k [int] Number of neareast neighbors.

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*batch*)

model_name = 'dgcnn'

classmethod split_dataset(*dataset*, *args*)

class cogdl.models.nn.grand.**Grand**(*nfeat*, *nhid*, *nclass*, *input_droprate*, *hidden_droprate*, *use_bn*, *dropnode_rate*, *tem*, *lam*, *order*, *sample*, *alpha*)
Bases: *cogdl.models.base_model.BaseModel*

Implementation of GRAND in paper *"Graph Random Neural Networks for Semi-Supervised Learning on Graphs"* <https://arxiv.org/abs/2005.11079>

nfeat [int] Size of each input features.

nhid [int] Size of hidden features.

nclass [int] Number of output classes.

input_droprate [float] Dropout rate of input features.

hidden_droprate [float] Dropout rate of hidden features.

use_bn [bool] Using batch normalization.

dropnode_rate [float] Rate of dropping elements of input features

tem [float] Temperature to sharpen predictions.

lam [float] Proportion of consistency loss of unlabelled data

order [int] Order of adjacency matrix

sample [int] Number of augmentations for consistency loss

alpha : float

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

consis_loss(*logps*, *train_mask*)

dropNode(*x*)

forward(*x*, *edge_index*, *edge_weight=None*)

model_name = 'grand'

node_classification_loss(*data*)

normalize_x(*x*)

predict(*data*)

rand_prop(*x*, *edge_index*, *edge_weight*)

class cogdl.models.nn.pyg_gtn.**GTN**(*num_edge*, *num_channels*, *w_in*, *w_out*, *num_class*, *num_nodes*, *num_layers*)
Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

evaluate(*data*, *nodes*, *targets*)

forward(*A*, *X*, *target_x*, *target*)

loss(*data*)

model_name = 'gtn'

norm(*edge_index*, *num_nodes*, *edge_weight*, *improved=False*, *dtype=None*)

normalization(*H*)

class cogdl.models.nn.rgcn.**LinkPredictRGCN**(*num_entities*, *num_rels*, *hidden_size*, *num_layers*, *regularizer='basis'*, *num_bases=None*, *self_loop=True*, *sampling_rate=0.01*, *penalty=0*, *dropout=0.0*, *self_dropout=0.0*)
Bases: *cogdl.layers.link_prediction_module.GNNLinkPredict*, *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*edge_index*, *edge_type*)

loss(*data*, *split='train'*)

model_name = 'rgcn'

predict(*edge_index*, *edge_type*)

class cogdl.models.nn.deepergcn.**DeeperGCN**(*in_feat*, *hidden_size*, *out_feat*, *num_layers*, *connection='res+'*, *activation='relu'*, *dropout=0.0*, *aggr='max'*, *beta=1.0*, *p=1.0*, *learn_beta=False*, *learn_p=False*, *learn_msg_scale=True*, *use_msg_norm=False*)

Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*x*, *edge_index*, *edge_attr=None*)

static get_trainer(*taskType: Any*, *args*)

model_name = 'deepergcn'

predict(*data*)

class cogdl.models.nn.drgat.**DrGAT**(*num_features*, *num_classes*, *hidden_size*, *num_heads*, *dropout*)

Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*x*, *edge_index*)

model_name = 'drgat'

predict(*data*)

class cogdl.models.nn.infograph.**InfoGraph**(*in_feats*, *hidden_dim*, *out_feats*, *num_layers=3*, *sup=False*)

Bases: *cogdl.models.base_model.BaseModel*

Implimentation of Infograph in paper '"InfoGraph: Unsupervised and Semi-supervised Graph-Level Representation
Learning via Mutual Information Maximization" <https://openreview.net/forum?id=r1lfF2NYvH>__. '

in_feats [int] Size of each input sample.

out_feats [int] Size of each output sample.

num_layers [int, optional] Number of MLP layers in encoder, default: `3`.

unsup [bool, optional] Use unsupervised model if True, default: `True`.

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*batch*)

graph_classification_loss(*batch*)

static mi_loss(*pos_mask*, *neg_mask*, *mi*, *pos_div*, *neg_div*)

model_name = 'infograph'

reset_parameters()

classmethod split_dataset(*dataset*, *args*)

sup_forward(*x*, *edge_index=None*, *batch=None*, *label=None*, *edge_attr=None*)

sup_loss(*pred*, *batch*)

unsup_forward(*x*, *edge_index=None*, *batch=None*)

unsup_loss(*graph_feat*, *node_feat*, *batch*)

unsup_sup_loss(*x*, *edge_index*, *batch*)

class cogdl.models.nn.dropedge_gcn.**DropEdge_GCN**(*nfeat*, *nhid*, *nclass*, *nhidlayer*, *dropout*, *baseblock*, *inputlayer*, *outputlayer*, *nbaselayer*, *activation*, *withbn*, *withloop*, *aggrmethod*)

Bases: *cogdl.models.base_model.BaseModel*

DropEdge: Towards Deep Graph Convolutional Networks on Node Classification Applying DropEdge to GCN @ https://arxiv.org/pdf/1907.10903.pdf

The model for the single kind of deepgcn blocks. The model architecture likes: inputlayer(nfeat)–block(nbaselayer, nhid)–. . . –outputlayer(nclass)–softmax(nclass)

|—— nhidlayer —-|

The total layer is nhidlayer*nbaselayer + 2. All options are configurable.

Args: Initial function. :param nfeat: the input feature dimension. :param nhid: the hidden feature dimension. :param nclass: the output feature dimension. :param nhidlayer: the number of hidden blocks. :param dropout: the dropout ratio. :param baseblock: the baseblock type, can be “mutigcn”, “resgcn”, “densegcn” and “inceptiongcn”. :param inputlayer: the input layer type, can be “gcn”, “dense”, “none”. :param outputlayer: the input layer type, can be “gcn”, “dense”. :param nbaselayer: the number of layers in one hidden block. :param activation: the activation function, default is ReLu. :param withbn: using batch normalization in graph convolution. :param withloop: using self feature modeling in graph convolution. :param aggrmethod: the aggregation function for baseblock, can be “concat” and “add”. For “resgcn”, the default

is “add”, for others the default is “concat”.

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*fea*, *adj*)

model_name = 'dropedge_gcn'

predict(*data*)

reset_parameters()

class cogdl.models.nn.disengcn.**DisenGCN**(*in_feats*, *hidden_size*, *num_classes*, *K*, *iterations*, *tau*, *dropout*, *activation*)

Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*x*, *edge_index*)

model_name = 'disengcn'

predict(*data*)

reset_parameters()

class cogdl.models.nn.mlp.**MLP**(*in_feats*, *out_feats*, *hidden_size*, *num_layers*, *dropout=0.0*, *activation='relu'*, *norm=None*)

Bases: *cogdl.models.base_model.BaseModel*

Multilayer perception with normalization

$$x^{(i+1)} = \sigma(W^i x^{(i)})$$

in_feats [int] Size of each input sample.

out_feats [int] Size of each output sample.

hidden_dim [int] Size of hidden layer dimension.

use_bn [bool, optional] Apply batch normalization if True, default: 'True).

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*x*, **args*, ***kwargs*)

model_name = 'mlp'

predict(*data*)

class cogdl.models.nn.sgc.**sgc**(*in_feats*, *out_feats*)

Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*x*, *edge_index*)

model_name = 'sgc'

predict(*data*)

class cogdl.models.nn.stpgnn.**stpgnn**(*args*)

Bases: *cogdl.models.base_model.BaseModel*

Implementation of models in paper *"Strategies for Pre-training Graph Neural Networks"*. <https://arxiv.org/abs/1905.12265>

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

model_name = 'stpgnn'

class cogdl.models.nn.sortpool.**SortPool**(*in_feats*, *hidden_dim*, *num_classes*, *num_layers*, *out_channel*, *kernel_size*, *k=30*, *dropout=0.5*)

Bases: *cogdl.models.base_model.BaseModel*

Implimentation of sortpooling in paper *"An End-to-End Deep Learning Architecture for Graph Classification" <https://www.cse.wustl.edu/~muhan/papers/AAAI_2018_DGCNN.pdf>__.*

in_feats [int] Size of each input sample.

out_feats [int] Size of each output sample.

hidden_dim [int] Dimension of hidden layer embedding.

num_classes [int] Number of target classes.

num_layers [int] Number of graph neural network layers before pooling.

k [int, optional] Number of selected features to sort, default: `30`.

out_channel [int] Number of the first convolution's output channels.

kernel_size [int] Size of the first convolution's kernel.

dropout [float, optional] Size of dropout, default: `0.5`.

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*batch*)

model_name = 'sortpool'

classmethod split_dataset(*dataset*, *args*)

class cogdl.models.nn.pyg_srgcn.**SRGCN**(*in_feats*, *hidden_size*, *out_feats*, *attention*, *activation*, *nhop*, *normalization*, *dropout*, *node_dropout*, *alpha*, *nhead*, *subheads*)
Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*x*, *edge_index*)

model_name = 'srgcn'

predict(*data*)

class cogdl.models.nn.dgl_gcc.**GCC**(*load_path*)
Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

model_name = 'gcc'

train(*data*)
Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. `Dropout`, `BatchNorm`, etc.

Args:

mode (bool): whether to set training mode (`True`) or evaluation mode (`False`). Default: `True`.

Returns: Module: self

class cogdl.models.nn.pairnorm.**PairNorm**(*pn_model*, *hidden_layers*, *nhead*, *dropout*, *nlayer*, *residual*, *norm_mode*, *norm_scale*, *num_features*, *num_classes*)

Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*x*, *edge_index*)

model_name = 'pairnorm'

predict(*data*)

class cogdl.models.nn.unsup_graphsage.**SAGE**(*num_features*, *hidden_size*, *num_layers*, *sample_size*, *dropout*, *walk_length*, *negative_samples*)

Bases: *cogdl.models.base_model.BaseModel*

Implementation of unsupervised GraphSAGE in paper *"Inductive Representation Learning on Large Graphs"* <https://cs.stanford.edu/people/jure/pubs/graphsage-nips17.pdf>

num_features [int] Size of each input sample

hidden_size : int num_layers : int

The number of GNN layers.

samples_size [list] The number sampled neighbors of different orders

dropout : float walk_length : int

The length of random walk

negative_samples : int

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

embed(*data*)

forward(*x*, *edge_index*)

static get_trainer(*taskType*, *args*)

loss(*data*)

model_name = 'unsup_graphsage'

node_classification_loss(*data*)

sampling(*edge_index*, *num_sample*)

class cogdl.models.nn.pyg_sagpool.**SAGPoolNetwork**(*nfeat*, *nhid*, *nclass*, *dropout*, *pooling_ratio*, *pooling_layer_type*)

Bases: *cogdl.models.base_model.BaseModel*

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*batch*)

model_name = 'sagpool'

classmethod split_dataset(*dataset*, *args*)

2.10.5 AGC Model

class cogdl.models.agc.daegc.**DAEGC**(*num_features*, *hidden_size*, *embedding_size*, *num_heads*, *dropout*, *num_clusters*)
Bases: *cogdl.models.base_model.BaseModel*

The DAEGC model from the "Attributed Graph Clustering: A Deep Attentional Embedding Approach" paper

Args: num_clusters (int) : Number of clusters. T (int) : Number of iterations to recalculate P and Q gamma (float) : Hyperparameter that controls two parts of the loss.

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

forward(*x*, *edge_index*)

get_2hop(*edge_index*)
add 2-hop neighbors as new edges

get_features(*data*)

get_trainer(*task*, *args*)

model_name = 'daegc'

recon_loss(*z*, *adj*)

class cogdl.models.agc.agc.**AGC**(*num_clusters*, *max_iter*)
Bases: *cogdl.models.base_model.BaseModel*

The AGC model from the "Attributed Graph Clustering via Adaptive Graph Convolution" paper

Args: num_clusters (int) : Number of clusters. max_iter (int) : Max iteration to increase k

static add_args(*parser*)
Add model-specific arguments to the parser.

classmethod build_model_from_args(*args*)
Build a new model instance.

get_features(*data*)

get_trainer(*task*, *args*)

model_name = 'agc'

2.10.6 Model Module

cogdl.models.**build_model**(*args*)

cogdl.models.**register_model**(*name*)

New model types can be added to cogdl with the *register_model()* function decorator.

For example:

```
@register_model('gat')
class GAT(BaseModel):
    (...)
```

Args: name (str): the name of the model

cogdl.models.**try_import_model**(*model*)

2.11 layers

2.11.1 GCC module

class cogdl.layers.gcc_module.**ApplyNodeFunc**(*mlp*, *use_selayer*)

Bases: torch.nn.modules.module.Module

Update the node feature hv with MLP, BN and ReLU.

forward(*h*)

class cogdl.layers.gcc_module.**GATLayer**(*g*, *in_dim*, *out_dim*)

Bases: torch.nn.modules.module.Module

edge_attention(*edges*)

forward(*h*)

message_func(*edges*)

reduce_func(*nodes*)

class cogdl.layers.gcc_module.**GraphEncoder**(*positional_embedding_size=32*, *max_node_freq=8*, *max_edge_freq=8*, *max_degree=128*, *freq_embedding_size=32*, *degree_embedding_size=32*, *output_dim=32*, *node_hidden_dim=32*, *edge_hidden_dim=32*, *num_layers=6*, *num_heads=4*, *num_step_set2set=6*, *num_layer_set2set=3*, *norm=False*, *gnn_model='mpnn'*, *degree_input=False*, *lstm_as_gate=False*)

Bases: torch.nn.modules.module.Module

MPNN from Neural Message Passing for Quantum Chemistry

node_input_dim [int] Dimension of input node feature, default to be 15.

edge_input_dim [int] Dimension of input edge feature, default to be 15.

output_dim [int] Dimension of prediction, default to be 12.

node_hidden_dim [int] Dimension of node feature in hidden layers, default to be 64.

edge_hidden_dim [int] Dimension of edge feature in hidden layers, default to be 128.

num_step_message_passing [int] Number of message passing steps, default to be 6.

num_step_set2set [int] Number of set2set steps

num_layer_set2set [int] Number of set2set layers

forward(*g*, *return_all_outputs=False*)
Predict molecule labels

g [DGLGraph] Input DGLGraph for molecule(s)

n_feat [tensor of dtype float32 and shape (B1, D1)] Node features. B1 for number of nodes and D1 for the node feature size.

e_feat [tensor of dtype float32 and shape (B2, D2)] Edge features. B2 for number of edges and D2 for the edge feature size.

res : Predicted labels

class cogdl.layers.gcc_module.**MLP**(*num_layers*, *input_dim*, *hidden_dim*, *output_dim*, *use_selayer*)
Bases: torch.nn.modules.module.Module

MLP with linear output

forward(*x*)

class cogdl.layers.gcc_module.**SELayer**(*in_channels*, *se_channels*)
Bases: torch.nn.modules.module.Module

Squeeze-and-excitation networks

forward(*x*)

class cogdl.layers.gcc_module.**UnsupervisedGAT**(*node_input_dim*, *node_hidden_dim*, *edge_input_dim*, *num_layers*, *num_heads*)
Bases: torch.nn.modules.module.Module

forward(*g*, *n_feat*, *e_feat*)

class cogdl.layers.gcc_module.**UnsupervisedGIN**(*num_layers*, *num_mlp_layers*, *input_dim*, *hidden_dim*, *output_dim*, *final_dropout*, *learn_eps*, *graph_pooling_type*, *neighbor_pooling_type*, *use_selayer*)
Bases: torch.nn.modules.module.Module

GIN model

forward(*g*, *h*, *efeat*)

class cogdl.layers.gcc_module.**UnsupervisedMPNN**(*output_dim=32*, *node_input_dim=32*, *node_hidden_dim=32*, *edge_input_dim=32*, *edge_hidden_dim=32*, *num_step_message_passing=6*, *lstm_as_gate=False*)
Bases: torch.nn.modules.module.Module

MPNN from Neural Message Passing for Quantum Chemistry

node_input_dim [int] Dimension of input node feature, default to be 15.

edge_input_dim [int] Dimension of input edge feature, default to be 15.

output_dim [int] Dimension of prediction, default to be 12.

node_hidden_dim [int] Dimension of node feature in hidden layers, default to be 64.

edge_hidden_dim [int] Dimension of edge feature in hidden layers, default to be 128.

num_step_message_passing [int] Number of message passing steps, default to be 6.

num_step_set2set [int] Number of set2set steps

num_layer_set2set [int] Number of set2set layers

forward(*g*, *n_feat*, *e_feat*)
Predict molecule labels

g [DGLGraph] Input DGLGraph for molecule(s)

n_feat [tensor of dtype float32 and shape (B1, D1)] Node features. B1 for number of nodes and D1 for the node feature size.

e_feat [tensor of dtype float32 and shape (B2, D2)] Edge features. B2 for number of edges and D2 for the edge feature size.

res : Predicted labels

2.11.2 GPT-GNN module

class cogdl.layers.gpt_gnn_module.**Classifier**(*n_hid*, *n_out*)
Bases: torch.nn.modules.module.Module

forward(*x*)

class cogdl.layers.gpt_gnn_module.**GNN**(*in_dim*, *n_hid*, *num_types*, *num_relations*, *n_heads*, *n_layers*, *dropout=0.2*, *conv_name='hgt'*, *prev_norm=False*, *last_norm=False*, *use_RTE=True*)
Bases: torch.nn.modules.module.Module

forward(*node_feature*, *node_type*, *edge_time*, *edge_index*, *edge_type*)

class cogdl.layers.gpt_gnn_module.**GPT_GNN**(*gnn*, *rem_edge_list*, *attr_decoder*, *types*, *neg_samp_num*, *device*, *neg_queue_size=0*)
Bases: torch.nn.modules.module.Module

feat_loss(*reps*, *out*)

forward(*node_feature*, *node_type*, *edge_time*, *edge_index*, *edge_type*)

link_loss(*node_emb*, *rem_edge_list*, *ori_edge_list*, *node_dict*, *target_type*, *use_queue=True*, *update_queue=False*)

neg_sample(*souce_node_list*, *pos_node_list*)

text_loss(*reps*, *texts*, *w2v_model*, *device*)

class cogdl.layers.gpt_gnn_module.**GeneralConv**(*conv_name*, *in_hid*, *out_hid*, *num_types*, *num_relations*, *n_heads*, *dropout*, *use_norm=True*, *use_RTE=True*)
Bases: torch.nn.modules.module.Module

forward(*meta_xs*, *node_type*, *edge_index*, *edge_type*, *edge_time*)

class cogdl.layers.gpt_gnn_module.**Graph**
Bases: object

add_edge(*source_node*, *target_node*, *time=None*, *relation_type=None*, *directed=True*)

add_node(*node*)

get_meta_graph()

get_types()

node_feature = None
edge_list: index the adjacancy matrix (time) by <target_type, source_type, relation_type, target_id, source_id>

update_node(*node*)

class cogdl.layers.gpt_gnn_module.**HGTConv**(*in_dim*, *out_dim*, *num_types*, *num_relations*, *n_heads*, *dropout=0.2*, *use_norm=True*, *use_RTE=True*, ***kwargs*)
Bases: torch_geometric.nn.conv.message_passing.MessagePassing

forward(*node_inp*, *node_type*, *edge_index*, *edge_type*, *edge_time*)

message(*edge_index_i*, *node_inp_i*, *node_inp_j*, *node_type_i*, *node_type_j*, *edge_type*, *edge_time*)
j: source, i: target; <j, i>

update(*aggr_out*, *node_inp*, *node_type*)
Step 3: Target-specific Aggregation x = W[node_type] * gelu(Agg(x)) + x

class cogdl.layers.gpt_gnn_module.**Matcher**(*n_hid*, *n_out*, *temperature=0.1*)
Bases: torch.nn.modules.module.Module

Matching between a pair of nodes to conduct link prediction. Use multi-head attention as matching model.

forward(*x*, *ty*, *use_norm=True*)

class cogdl.layers.gpt_gnn_module.**RNNModel**(*n_word*, *ninp*, *nhid*, *nlayers*, *dropout=0.2*)
Bases: torch.nn.modules.module.Module

Container module with an encoder, a recurrent module, and a decoder.

forward(*inp*, *hidden=None*)

from_w2v(*w2v*)

class cogdl.layers.gpt_gnn_module.**RelTemporalEncoding**(*n_hid*, *max_len=240*, *dropout=0.2*)
Bases: torch.nn.modules.module.Module

Implement the Temporal Encoding (Sinusoid) function.

forward(*x*, *t*)

cogdl.layers.gpt_gnn_module.**args_print**(*args*)

cogdl.layers.gpt_gnn_module.**dcg_at_k**(*r*, *k*)

cogdl.layers.gpt_gnn_module.**defaultDictDict**()

cogdl.layers.gpt_gnn_module.**defaultDictDictDictDictDictInt**()

cogdl.layers.gpt_gnn_module.**defaultDictDictDictDictInt**()

cogdl.layers.gpt_gnn_module.**defaultDictDictDictInt**()

cogdl.layers.gpt_gnn_module.**defaultDictDictInt**()

cogdl.layers.gpt_gnn_module.**defaultDictInt**()

cogdl.layers.gpt_gnn_module.**defaultDictList**()

cogdl.layers.gpt_gnn_module.**feature_OAG**(*layer_data*, *graph*)

cogdl.layers.gpt_gnn_module.**feature_reddit**(*layer_data*, *graph*)

cogdl.layers.gpt_gnn_module.**load_gnn**(*_dict*)

cogdl.layers.gpt_gnn_module.**mean_reciprocal_rank**(*rs*)

cogdl.layers.gpt_gnn_module.**ndcg_at_k**(*r*, *k*)

cogdl.layers.gpt_gnn_module.**normalize**(*mx*)
Row-normalize sparse matrix

cogdl.layers.gpt_gnn_module.**preprocess_dataset**(*dataset*) → cogdl.layers.gpt_gnn_module.Graph

cogdl.layers.gpt_gnn_module.**randint**()

cogdl.layers.gpt_gnn_module.**sample_subgraph**(*graph*, *time_range*, *sampled_depth=2*, *sampled_number=8*, *inp=None*, *feature_extractor=<function feature_OAG>*)
Sample Sub-Graph based on the connection of other nodes with currently sampled nodes We maintain budgets for each node type, indexed by <node_id, time>. Currently sampled nodes are stored in layer_data. After nodes are sampled, we construct the sampled adjacancy matrix.

cogdl.layers.gpt_gnn_module.**sparse_mx_to_torch_sparse_tensor**(*sparse_mx*)
Convert a scipy sparse matrix to a torch sparse tensor.

cogdl.layers.gpt_gnn_module.**to_torch**(*feature*, *time*, *edge_list*, *graph*)
Transform a sampled sub-graph into pytorch Tensor node_dict: {node_type: <node_number, node_type_ID>} node_number is used to trace back the nodes in original graph. edge_dict: {edge_type: edge_type_ID}

2.11.3 Link Prediction module

class cogdl.layers.link_prediction_module.**ConvELayer**(*dim*, *num_filter=20*, *kernel_size=7*, *k_w=10*, *dropout=0.3*)
Bases: torch.nn.modules.module.Module

concat(*ent*, *rel*)

forward(*sub_emb*, *obj_emb*, *rel_emb*)

predict(*sub_emb*, *obj_emb*, *rel_emb*)

class cogdl.layers.link_prediction_module.**DistMultLayer**
Bases: torch.nn.modules.module.Module

forward(*sub_emb*, *obj_emb*, *rel_emb*)

predict(*sub_emb*, *obj_emb*, *rel_emb*)

class cogdl.layers.link_prediction_module.**GNNLinkPredict**(*score_func*, *dim*)
Bases: torch.nn.modules.module.Module

forward(*edge_index*, *edge_type*)

get_edge_set(*edge_index*, *edge_types*)

get_score(*heads*, *tails*, *rels*)

cogdl.layers.link_prediction_module.**cal_mrr**(*embedding*, *rel_embedding*, *edge_index*, *edge_type*, *scoring*, *protocol='raw'*, *batch_size=1000*, *hits=[]*)

cogdl.layers.link_prediction_module.**get_filtered_rank**(*heads*, *tails*, *rels*, *embedding*, *rel_embedding*, *batch_size*, *seen_data*)

cogdl.layers.link_prediction_module.**get_rank**(*scores*, *target*)

cogdl.layers.link_prediction_module.**get_raw_rank**(*heads*, *tails*, *rels*, *embedding*, *rel_embedding*, *batch_size*, *scoring*)

cogdl.layers.link_prediction_module.**sampling_edge_uniform**(*edge_index*, *edge_types*, *edge_set*, *sampling_rate*, *num_rels*, *label_smoothing=0.0*, *num_entities=1*)

Args: edge_index: edge index of graph edge_types: edge_set: set of all edges of the graph, (h, t, r) sampling_rate: num_rels: label_smoothing(Optional): num_entities (Optional):

Returns: sampled_edges: sampled existing edges rels: types of smapled existing edges sampled_edges_all: existing edges with corrupted edges sampled_types_all: types of existing and corrupted edges labels: 0/1

2.11.4 Mean Aggregator module

class cogdl.layers.maggregator.**MeanAggregator**(*in_channels*, *out_channels*, *bias=True*)

Bases: torch.nn.modules.module.Module

forward(*x*, *adj_sp*)

static norm(*x*, *adj_sp*)

class cogdl.layers.maggregator.**SumAggregator**(*in_channels*, *out_channels*, *bias=True*)

Bases: torch.nn.modules.module.Module

static aggr(*x*, *adj*)

forward(*x*, *adj*)

2.11.5 MixHop module

class cogdl.layers.mixhop_layer.**MixHopLayer**(*num_features*, *adj_pows*, *dim_per_pow*)

Bases: torch.nn.modules.module.Module

adj_pow_x(*x*, *adj*, *p*)

forward(*x*, *edge_index*)

reset_parameters()

2.11.6 PPRGo module

class cogdl.layers.pprgo_modules.**PPRGoDataset**(*features: torch.Tensor*, *ppr_matrix: scipy.sparse.csr.csr_matrix*, *node_indices: torch.Tensor*, *labels_all: torch.Tensor = None*)

Bases: torch.utils.data.dataset.Dataset

cogdl.layers.pprgo_modules.**build_topk_ppr_matrix_from_data**(*edge_index*, **args*, ***kwargs*)

cogdl.layers.pprgo_modules.**calc_ppr_topk_parallel**

cogdl.layers.pprgo_modules.**construct_sparse**(*neighbors*, *weights*, *shape*)

cogdl.layers.pprgo_modules.**ppr_topk**(*adj_matrix*, *alpha*, *epsilon*, *nodes*, *topk*)
Calculate the PPR matrix approximately using Anderson.

cogdl.layers.pprgo_modules.**topk_ppr_matrix**(*adj_matrix*, *alpha*, *eps*, *idx*, *topk*, *normalization='row'*)
Create a sparse matrix where each node has up to the topk PPR neighbors and their weights.

2.11.7 ProNE module

class cogdl.layers.prone_module.**Gaussian**(*mu=0.5*, *theta=1*, *rescale=False*, *k=3*)
Bases: object

prop(*mx*, *emb*)

class cogdl.layers.prone_module.**HeatKernel**(*t=0.5*, *theta0=0.6*, *theta1=0.4*)
Bases: object

prop(*mx*, *emb*)

prop_adjacency(*mx*)

class cogdl.layers.prone_module.**HeatKernelApproximation**(*t=0.2*, *k=5*)
Bases: object

chebyshev(*mx*, *emb*)

prop(*mx*, *emb*)

taylor(*mx*, *emb*)

class cogdl.layers.prone_module.**NodeAdaptiveEncoder**
Bases: object

- shrink negative values in signal/feature matrix
- no learning

static prop(*signal*)

class cogdl.layers.prone_module.**PPR**(*alpha=0.5*, *k=10*)
Bases: object

applying sparsification to accelerate computation

prop(*mx*, *emb*)

class cogdl.layers.prone_module.**ProNE**
Bases: object

class cogdl.layers.prone_module.**SignalRescaling**
Bases: object

- **rescale signal of each node according to the degree of the node:**
  - sigmoid(degree)
  - sigmoid(1/degree)

prop(*mx*, *emb*)

cogdl.layers.prone_module.**get_embedding_dense**(*matrix*, *dimension*)

cogdl.layers.prone_module.**propagate**(*mx*, *emb*, *stype*, *space=None*)

2.11.8 SELayer module

class cogdl.layers.se_layer.**SELayer**(*in_channels*, *se_channels*)
Bases: torch.nn.modules.module.Module

Squeeze-and-excitation networks

forward(*x*)

2.11.9 SRGCN module

class cogdl.layers.srgcn_module.**ColumnUniform**
Bases: torch.nn.modules.module.Module

forward(*edge_index*, *edge_attr*, *N*)

class cogdl.layers.srgcn_module.**EdgeAttention**(*in_feat*)
Bases: torch.nn.modules.module.Module

forward(*x*, *edge_index*, *edge_attr*)

class cogdl.layers.srgcn_module.**HeatKernel**(*in_feat*)
Bases: torch.nn.modules.module.Module

forward(*x*, *edge_index*, *edge_attr*)

class cogdl.layers.srgcn_module.**Identity**(*in_feat*)
Bases: torch.nn.modules.module.Module

forward(*x*, *edge_index*, *edge_attr*)

class cogdl.layers.srgcn_module.**NodeAttention**(*in_feat*)
Bases: torch.nn.modules.module.Module

forward(*x*, *edge_index*, *edge_attr*)

class cogdl.layers.srgcn_module.**NormIdentity**
Bases: torch.nn.modules.module.Module

forward(*edge_index*, *edge_attr*, *N*)

class cogdl.layers.srgcn_module.**PPR**(*in_feat*)
Bases: torch.nn.modules.module.Module

forward(*x*, *edge_index*, *edge_attr*)

class cogdl.layers.srgcn_module.**RowSoftmax**
Bases: torch.nn.modules.module.Module

forward(*edge_index*, *edge_attr*, *N*)

class cogdl.layers.srgcn_module.**RowUniform**
Bases: torch.nn.modules.module.Module

forward(*edge_index*, *edge_attr*, *N*)

class cogdl.layers.srgcn_module.**SymmetryNorm**
Bases: torch.nn.modules.module.Module

forward(*edge_index*, *edge_attr*, *N*)

cogdl.layers.srgcn_module.**act_attention**(*attn_type*)

cogdl.layers.srgcn_module.**act_map**(*act*)

cogdl.layers.srgcn_module.**act_normalization**(*norm_type*)

2.11.10 Strategies module

class cogdl.layers.strategies_layers.**ContextPredictTrainer**(*args*)
Bases: *cogdl.layers.strategies_layers.Pretrainer*

static add_args(*parser*)

get_cbow_pred(*overlapped_rep*, *overlapped_context*, *neighbor_rep*)

get_skipgram_pred(*overlapped_rep*, *overlapped_context_size*, *neighbor_rep*)

class cogdl.layers.strategies_layers.**Discriminator**(*hidden_size*)
Bases: torch.nn.modules.module.Module

forward(*x*, *summary*)

reset_parameters()

class cogdl.layers.strategies_layers.**Finetuner**(*args*)
Bases: *cogdl.layers.strategies_layers.Pretrainer*

static add_args(*parser*)

build_model(*args*)

fit()

split_data()

class cogdl.layers.strategies_layers.**GINConv**(*hidden_size*, *input_layer=None*, *edge_emb=None*, *edge_encode=None*, *pooling='sum'*, *feature_concat=False*)
Bases: torch.nn.modules.module.Module

Implementation of Graph isomorphism network used in paper *"Strategies for Pre-training Graph Neural Networks"*. <https://arxiv.org/abs/1905.12265>

hidden_size [int] Size of each hidden unit

input_layer [int, optional] The size of input node features if not *None*.

edge_emb [list, optional] The number of edge types if not *None*

edge_encode [int, optional] Size of each edge feature if not *None*

pooling [str] Pooling method.

aggr(*x*, *edge_index*, *num_nodes*)

forward(*x*, *edge_index*, *edge_attr*, *self_loop_index=None*, *self_loop_type=None*)

class cogdl.layers.strategies_layers.**GNN**(*num_layers*, *hidden_size*, *JK='last'*, *dropout=0.5*, *input_layer=None*, *edge_encode=None*, *edge_emb=None*, *num_atom_type=None*, *num_chirality_tag=None*, *concat=False*)
Bases: torch.nn.modules.module.Module

forward(*x*, *edge_index*, *edge_attr*, *self_loop_index=None*, *self_loop_type=None*)

class cogdl.layers.strategies_layers.**GNNPred**(*num_layers*, *hidden_size*, *num_tasks*, *JK='last'*, *dropout=0*, *graph_pooling='mean'*, *input_layer=None*, *edge_encode=None*, *edge_emb=None*, *num_atom_type=None*, *num_chirality_tag=None*, *concat=True*)

Bases: torch.nn.modules.module.Module

forward(*data*, *self_loop_index*, *self_loop_type*)

load_from_pretrained(*path*)

pool(*x*, *batch*)

class cogdl.layers.strategies_layers.**InfoMaxTrainer**(*args*)

Bases: *cogdl.layers.strategies_layers.Pretrainer*

static add_args(*parser*)

class cogdl.layers.strategies_layers.**MaskTrainer**(*args*)

Bases: *cogdl.layers.strategies_layers.Pretrainer*

static add_args(*parser*)

class cogdl.layers.strategies_layers.**Pretrainer**(*args*, *transform=None*)

Bases: torch.nn.modules.module.Module

Base class for Pre-training Models of paper *"Strategies for Pre-training Graph Neural Networks"*. <https://arxiv.org/abs/1905.12265>

fit()

get_dataset(*dataset_name*, *transform=None*)

class cogdl.layers.strategies_layers.**SupervisedTrainer**(*args*)

Bases: *cogdl.layers.strategies_layers.Pretrainer*

static add_args(*parser*)

split_data()

2.12 options

cogdl.options.**add_dataset_args**(*parser*)

cogdl.options.**add_model_args**(*parser*)

cogdl.options.**add_task_args**(*parser*)

cogdl.options.**add_trainer_args**(*parser*)

cogdl.options.**get_default_args**(*task: str*, *dataset*, *model*, ***kwargs*)

cogdl.options.**get_display_data_parser**()

cogdl.options.**get_download_data_parser**()

cogdl.options.**get_parser**()

cogdl.options.**get_task_model_args**(*task*, *model=None*)

cogdl.options.**get_training_parser**()

cogdl.options.**parse_args_and_arch**(*parser*, *args*)

The parser doesn't know about model-specific args, so we parse twice.

2.13 utils

class cogdl.utils.utils.**ArgClass**
Bases: object

cogdl.utils.utils.**add_remaining_self_loops**(*edge_index*, *edge_weight=None*, *fill_value=1*, *num_nodes=None*)

cogdl.utils.utils.**add_self_loops**(*edge_index*, *edge_weight=None*, *fill_value=1*, *num_nodes=None*)

cogdl.utils.utils.**alias_draw**(*J*, *q*)
Draw sample from a non-uniform discrete distribution using alias sampling.

cogdl.utils.utils.**alias_setup**(*probs*)
Compute utility lists for non-uniform sampling from discrete distributions. Refer to https://hips.seas.harvard.edu/blog/2013/03/03/the-alias-method-efficient-sampling-with-many-discrete-outcomes/ for details

cogdl.utils.utils.**batch_mean_pooling**(*x*, *batch*)

cogdl.utils.utils.**batch_sum_pooling**(*x*, *batch*)

cogdl.utils.utils.**build_args_from_dict**(*dic*)

cogdl.utils.utils.**coalesce**(*row*, *col*, *value=None*)

cogdl.utils.utils.**coo2csc**(*edge_index*, *edge_attr*)

cogdl.utils.utils.**coo2csr**(*edge_index*, *edge_attr*, *num_nodes=None*)

cogdl.utils.utils.**cycle_index**(*num*, *shift*)

cogdl.utils.utils.**download_url**(*url*, *folder*, *name=None*, *log=True*)
Downloads the content of an URL to a specific folder.

Args: url (string): The url. folder (string): The folder. name (string): saved filename. log (bool, optional): If `False`, will not print anything to the

console. (default: `True`)

cogdl.utils.utils.**dropout_adj**(*edge_index: torch.Tensor*, *edge_weight: Optional[torch.Tensor] = None*, *drop_rate: float = 0.5*, *renorm: bool = True*)

cogdl.utils.utils.**edge_softmax**(*indices*, *values*, *shape*)

Args: indices: Tensor, shape=(2, E) values: Tensor, shape=(N,) shape: tuple(int, int)

Returns: Softmax values of edge values for nodes

cogdl.utils.utils.**filter_adj**(*row*, *col*, *edge_attr*, *mask*)

cogdl.utils.utils.**get_activation**(*act: str*)

cogdl.utils.utils.**get_csr_from_edge_index**(*edge_index*, *edge_attr*, *size*)

cogdl.utils.utils.**get_csr_ind**(*edge_index*, *edge_weight=None*, *num_nodes=None*)

cogdl.utils.utils.**get_degrees**(*indices*, *num_nodes=None*)

cogdl.utils.utils.**initialize_spmm**(*args*)

cogdl.utils.utils.**makedirs**(*path*)

cogdl.utils.utils.**mul_edge_softmax**(*indices*, *values*, *shape*)

Args: indices: Tensor, shape=(2, E) values: Tensor, shape=(E, d) shape: tuple(int, int)

Returns: Softmax values of multi-dimension edge values for nodes

cogdl.utils.utils.**negative_edge_sampling**(*edge_index: torch.Tensor*, *num_nodes: Optional[int] = None*, *num_neg_samples: Optional[int] = None*, *undirected: bool = False*)

cogdl.utils.utils.**print_result**(*results*, *datasets*, *model_name*)

cogdl.utils.utils.**remove_self_loops**(*indices*, *values=None*)

cogdl.utils.utils.**row_normalization**(*num_nodes*, *edge_index*, *edge_weight=None*)

cogdl.utils.utils.**set_random_seed**(*seed*)

cogdl.utils.utils.**spmm**(*edge_index*, *edge_weight*, *x*, *num_nodes=None*)

Args: edge_index : Tensor, shape=(2, E) edge_weight : Tensor, shape=(E,) x : Tensor, shape=(N,) num_nodes : Optional[int]

cogdl.utils.utils.**spmm_adj**(*indices*, *values*, *x*, *num_nodes=None*)

cogdl.utils.utils.**spmm_scatter**(*indices*, *values*, *b*)

Args: indices : Tensor, shape=(2, E) values : Tensor, shape=(E,) b : Tensor, shape=(N,)

cogdl.utils.utils.**symmetric_normalization**(*num_nodes*, *edge_index*, *edge_weight=None*)

cogdl.utils.utils.**tabulate_results**(*results_dict*)

cogdl.utils.utils.**to_undirected**(*edge_index*, *num_nodes=None*)

Converts the graph given by `edge_index` to an undirected graph, so that $(j, i) \in \mathcal{E}$ for every edge $(i, j) \in \mathcal{E}$.

Args: edge_index (LongTensor): The edge indices. num_nodes (int, optional): The number of nodes, *i.e.* `max_val + 1` of `edge_index`. (default: `None`)

Return type `LongTensor`

cogdl.utils.utils.**untar**(*path*, *fname*, *deleteTar=True*)

Unpacks the given archive file to the same directory, then (by default) deletes the archive file.

cogdl.utils.evaluator.**accuracy**(*y_pred*, *y_true*)

cogdl.utils.evaluator.**bce_with_logits_loss**(*y_pred*, *y_true*, *reduction='mean'*)

cogdl.utils.evaluator.**cross_entropy_loss**(*y_pred*, *y_true*)

cogdl.utils.evaluator.**multiclass_f1**(*y_pred*, *y_true*)

cogdl.utils.evaluator.**multilabel_f1**(*y_pred*, *y_true*, *sigmoid=False*)

class cogdl.utils.sampling.**RandomWalker**(*adj=None*, *num_nodes=None*)

Bases: `object`

build_up(*adj*, *num_nodes*)

walk(*start*, *walk_length*)

cogdl.utils.sampling.**random_walk**

Parameters: start : np.array(dtype=np.int32) length : int indptr : np.array(dtype=np.int32) indices : np.array(dtype=np.int32)

Return: list(np.array(dtype=np.int32))

2.14 experiments

class cogdl.experiments.**AutoML**(*task*, *dataset*, *model*, *n_trials=3*, ***kwargs*)
Bases: object

Args: func_search: function to obtain hyper-parameters to search

run()

cogdl.experiments.**auto_experiment**(*task: str*, *dataset*, *model*, ***kwargs*)

cogdl.experiments.**check_task_dataset_model_match**(*task*, *variants*)

cogdl.experiments.**experiment**(*task: str*, *dataset*, *model*, ***kwargs*)

cogdl.experiments.**gen_variants**(***items*)

cogdl.experiments.**output_results**(*results_dict*, *tablefmt='github'*)

cogdl.experiments.**raw_experiment**(*task: str*, *dataset*, *model*, ***kwargs*)

cogdl.experiments.**set_best_config**(*args*)

cogdl.experiments.**train**(*args*)

cogdl.experiments.**variant_args_generator**(*args*, *variants*)
Form variants as group with size of num_workers

2.15 pipelines

class cogdl.pipelines.**DatasetPipeline**(*app: str*, ***kwargs*)
Bases: *cogdl.pipelines.Pipeline*

class cogdl.pipelines.**DatasetStatsPipeline**(*app: str*, ***kwargs*)
Bases: *cogdl.pipelines.DatasetPipeline*

class cogdl.pipelines.**DatasetVisualPipeline**(*app: str*, ***kwargs*)
Bases: *cogdl.pipelines.DatasetPipeline*

class cogdl.pipelines.**OAGBertInferencePipepline**(*app: str*, *model: str*, ***kwargs*)
Bases: *cogdl.pipelines.Pipeline*

class cogdl.pipelines.**Pipeline**(*app: str*, ***kwargs*)
Bases: object

cogdl.pipelines.**check_app**(*app: str*)

cogdl.pipelines.**pipeline**(*app: str*, ***kwargs*) → cogdl.pipelines.Pipeline

CHAPTER 3

Indices and tables

- genindex
- modindex
- search

Python Module Index

C

Index

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

R

S

T

U

V

W

Y