A given multi-task combination

Spliting training and test set for DDI or DTI prediction

python create_sample.py:

- --offset: the choice of downstream tasks (i.e., DDI or DTI)
- --warm_ratio: the ratio of test dataset in warm start prediction
- --cold_ratio: the ratio of test dataset in cold start prediction

Constructing initial features for each node in the BioHN

python degree_code.py:

- --downstream: the name of downstream tasks
- --scenario: the prediction scenario (i.e., warm or cold start)
- --dataclean: whether to remove the test data from SSL dataset

Generating samples for the given multi-task combination

Run some of programs according to the given combination:

python ClusterPre.py or python pathClass.py or python PairDistance.py or python EdgeMask.py or python SimReg.py or python SimCon.py

Split training and test set for multiple tasks by looping the code: python create_dataset.py

• --input_file: the path of input file

Executing multi-task self-supervised representation learning

Iterate some of programs according to the given combination:

python train_class.py or python train_Reg.py or python train_sim.py

Extracting representations of nodes in the BioHN

python get_feature.py

- --model: the path of multi-task SSL model
- --length: the number of tasks in the combinations

ClusterPre and SimReg are regression models.

python train_reg.py

- --save: the path for saving multi-task SSL model
- --train_file: the path of SSL training sample
- --nclass: the number of hidden units in output layer
- --length: the length of each sample
- --ntask: the number of tasks in combinations
- --task: the identifier of current task
- --share: the path of share model
- --refine: the path of private model
- --time: the identifier of current epoch

pathClass, PairDistance, and EdgeMask are classification models.

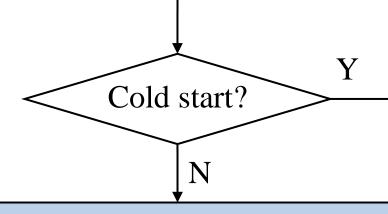
python train_class.py

- --save: the path for saving multi-task SSL model
- --train_file: the path of SSL training sample
- --nclass: the number of hidden units in output layer
- --length: the length of each sample
- --sub: the parameter is set as 1 in only PairDistance task
- --ntask: the number of tasks in combinations
- --task: the identifier of current task
- --share: the path of share model
- --refine: the path of private model
- --time: the identifier of current epoch

SimCon is a similarity-based contrast model.

python train_sim.py

- --save: the path for saving multi-task SSL model
- --train_file: the path of SSL training sample
- --nclass: the number of hidden units in output layer
- --length: the length of each sample
- --ntask: the number of tasks in combinations
- --task: the identifier of current task
- --share: the path of share model
- --refine: the path of private model
- --time: the identifier of current epoch



Executing drug warm start prediction

python warm_start.py

- --input_file: the path of input data
- --feature: the path of node representations

Executing drug cold start prediction

python cold_start.py

- --input_file: the path of input Data
- --feature: the path of node representations

DDI or **DTI** prediction results