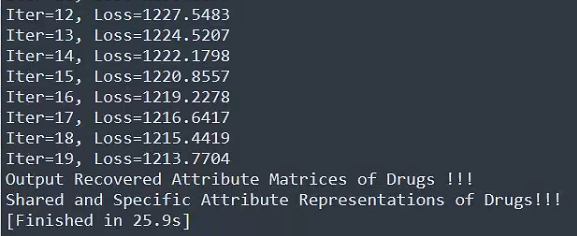
This folder contains the source codes of multi-attribute representation learning in MAMK model, which aims to explore the shared and specific attribute representations of drugs with the formal formulations in terms of the consistency and uniqueness of multi-attribute for recovering the absent information of multi-attribute. Since the adverse interactions among drugs generally originate from adverse drug reports associated with approved drugs with known molecular structures, the source codes of multi-attribute representation learning are to recover the absent information of drugs in terms of target, pathway, side effect, phenotype, gene, and disease and take no consideration of inferring molecular structures of drugs.

Except for two subfolders .idea and \_\_pycache\_\_ automatically generated by source codes, this folder consists of nine subfolders and a source code file named ‘MultiAttributeRepresentationLearning.py’. The detailed information of the files in this folder are described in the following.

Subfolders **1. MolecularStructure, 2. Target, 3. Pathway, 4. SideEffect, 5. Phenotype, 6. Gene, and 7. Disease** contain the seven attribute information of 752 drugs used for experiments. The feature dimensionalities for attributes molecular structure, target, pathway, side effect, phenotype, gene, and disease are 881, 1431, 427, 3750, 2656, 6267, and 613, respectively. Except for molecular structure, some drugs suffer from the absence of the remaining six attributes. The number of drugs with absent attribute information of target, pathway, side effect, phenotype, gene, and disease are 38, 47, 148, 154, 40, and 117. **The details descriptions for the files included in these seven subfolders can be referred to ./ Experimental Dataset\Drug Attribute Information\ReadMe.docx.**

**MultiAttributeRepresentationLearning.py** is the main source codes for multi-attribute representation learning. This file can be executed by ‘Python MultiAttributeRepresentationLearning.py’ in the software of Sublime Text 3 by Ctrl+B. For each iteration, the console outputs the number of iteration and the loss of the objective function. We conduct 20 iterations to output the shared attribute representation, the specific attribute representations of the seven attributes, and the recovered attribute information of the seven attributes. In particular, the shared and specific attribute representations are stored in subfolder **SharedSpecificAttributeRepresentations,** and the recovered attribute information of the seven attributes are stored in subfolder **RecoveredMultiAttributeRepresentations**. Some print results are shown as follows:



Subfolder **SharedSpecificAttributeRepresentations** contains the shared attribute representation and specific attribute representations of the seven attributes output by running file ‘**MultiAttributeRepresentationLearning.py**’. In this subfolder, the files with suffix ‘.txt’ are the shared and specific attribute representations output by the codes that can be opened and viewed by notepad, and the files with suffix ‘.npy’ are also the shared and specific attribute representations by the codes that cannot be opened and viewed directly but only can be loaded by Python codes for the convenience of the implementation for the source codes. Therefore, files **P\_CommonAttributeMatrix.npy** and **P\_CommonAttributeMatrix.txt** are the shared attribute representations, files **Q\_MolecularStructure.npy** and **Q\_MolecularStructure.txt** are the specific attribute representations of molecular structure, **Q\_Target.npy** and **Q\_Target.txt** are the specific attribute representations of target, **Q\_Pathway.npy** and **Q\_Pathway.txt** are the specific attribute representations of pathway, **Q\_SideEffect.npy** and **Q\_SideEffect.txt** are the specific attribute representations of side effect, **Q\_Phenotype.npy** and **Q\_Phenotype.txt** are the specific attribute representations of phenotype, **Q\_Gene.npy** and **Q\_Gene.txt** are the specific attribute representations of gene, and **Q\_Disease.npy** and **Q\_Disease.txt** are the specific attribute representations of disease.

Subfolder **RecoveredMultiAttributeRepresentations** contains the recovered attribute representations of the seven attributes. To be specific, files **X\_MolecularStructureRecovered.npy** and **X\_MolecularStructureRecovered.txt** are the recovered molecular structures of drugs. However, due to that adverse drug reports are associated with approved drugs with known molecular structures, the drugs in the experiments are all with existing molecular structures. Thus, the contents in files **X\_MolecularStructureRecovered.npy** and **X\_MolecularStructureRecovered.txt** are the same to file **DrugMolecularStructureMatrix.txt** in .\Experimental Dataset\Drug Attribute Information\1. MolecularStructure. In addition, files **X\_TargetRecovered.npy** and **X\_TargetRecovered.txt** are the recovered target representations of drugs, files **X\_PathwayRecovered.npy** and **X\_PathwayRecovered.txt** are the recovered pathway representations of drugs, files **X\_SideEffectRecovered.npy** and **X\_SideEffectRecovered.txt** are the recovered side effect representations of drugs, files **X\_PhenotypeRecovered.npy** and **X\_PhenotypeRecovered.txt** are the recovered phenotype representations of drugs, files **X\_GeneRecovered.npy** and **X\_GeneRecovered.txt** are the recovered gene information of drugs, and files **X\_DiseaseRecovered.npy** and **X\_DiseaseRecovered.txt** are the recovered disease information of drugs.