先运行formatDatasetsGO.R，以下全部去除根节点

再运行exceptroottable.R，产生except.root.table2是全部数据集（4085行）与除root外所属GO标签类别（1048列）的dataframe。20201216，阈值设为100的节点数为264，20201231

至此完成全部原始训练集生成4031\*1048节点\*50属性

再运行nodescreation.R，产生go.for.level2是除root外全部GO标签所属的13个层级，产生nodes.to.index2是全部GO标签的标号索引，产生nodes.to.children，是每一个GO的子节点索引，产生go.leaf.nodes2，检索出全部叶子节点。产生祖先节点，子孙节点索引（20210131改）

再运行nodesclean.R，减去8个不属于BP功能，减去这8个的子孙节点，剩下223个节点.20210202加入recall值不好节点的去除，总共剩下除根结点的204个节点

至此完成全部节点操作，共204节点

此后为训练集拆分以fit in SVM

*再运行testdataprocessing，traindataprocessing，traindataprocessing，产生train.data2.total，valid.data2.total，test.data2.total等，包含xxx.select.data2是数据集与特征值构成的dataframe，xxx.select.table2是数据集中每一个样本与所含GO标签构成的二值dataframe，以及数据集每一个样本所含的全部GO标签。（不用了，直接在nodesalldatasetcreation里generate了）*

*运行nodesalldatasetcreation.R，针对每一个节点的训练集*

*Plotgraph.R是针对属于BP功能的204个GO绘制成的结构图*

*recremove是筛选recall值较低的节点并进行去除，正常流程中不用运行，已将需要去除的节点加在nodesclean.R中了*

*20210221完成dataset\_select2.R封装成DatasetSelect函数，根据dataset.index拆分数据集，4031拆成4\*806+807，每一个806（807）用dataset\_divide封装成的DatasetDivide拆成训练集和测试集2:1*

运行partdatasetprocessing.R，选择datasetindex=1/2/3/4/5,生成5个数据集，以及各自的traindataset，trainclass，validdataaset，validclass，testdataset，testclass， xxx.select.data2是数据集与特征值构成的dataframe，xxx.select.table2是数据集中每一个样本与所含GO标签构成的二值dataframe，以及数据集每一个样本所含的全部GO标签。另外根据每个节点拆分出了训练集。所有文件存储在data下的datasetx文件夹里。

运行evaluate.R对某数据集进行结果评价，运行evaluateall.R对全部数据集进行结果评价。

关于DAGlabel的实现，先运行forDAGlabel.R，产生DAGlabel需要的节点索引，然后运行chloss2.R，产生节点的σ值。最后运行DAGlabel.R，

以上全部都有hr版本，对应于HR-SVM，evaluate.hr,chloss2.hr,DAGlabel.hr，并产生结果HRSVM（HR-SVM），DAGlabelhr。

运行resultcomparison.R，产生TPR，TOPDOWN，DOWNTOP，CLUS-HMC，HR-SVM，HMC-DAG-SVM对比结果，图+表

运行proveeffectnew.R,产生关于DAGlabel和TPR的对比，tochangelabels是SVM出来违反层级约束的个数，DAGchange是DAGlabel改变的标签对数，TPRchange是TPR改变的标签对数，DAGwork是经DAGlabel后从不满足层级约束到满足层级约束的改变的对数，TPRwork是经TPR后从不满足层级约束到满足层级约束的改变的对数，TPRright是TPR在改变了的对数中分对的标签对，DAGright是DAGlabel在改变了的对数中分对的标签对

关于对比实验clus，首先在partdatasetprocessing.R中设置index，产生训练集验证集测试集数据之后，运行generatearff.R生成arff格式的数据，存在了arff//arff1/2/3/4/5的文件夹中，然后打开命令提示符，先进入newtrain/valid/testarff所在的文件夹中,

cd C:\Users\1231\Desktop\dataprocessing\data\223arff\arff1(,2,3,4,5)

再在其中根据.s的设置运行clus

java -jar F:\R\Clus-2.12\Clus\Clus.jar new.s

输出的文件有output,有test.predict.arff，回到R中运行readarfflabels.R产生包含层级多标签分类的评价指标。把clus12345.txt复制到data//result文件夹