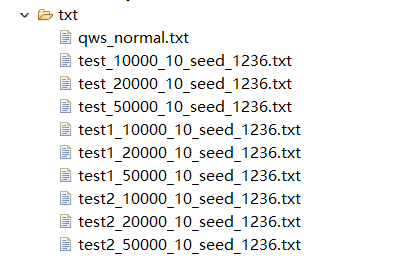
服务计算前沿实验报告

1173710107 任永鹏

1. 数据集解释



格式为testx\_y\_z\_seed\_1236

这里test代表独立 test1代表正相关 test2代表负相关

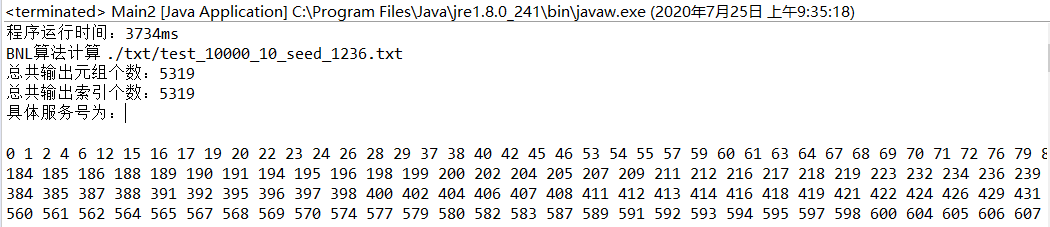
Y代表数据集行数

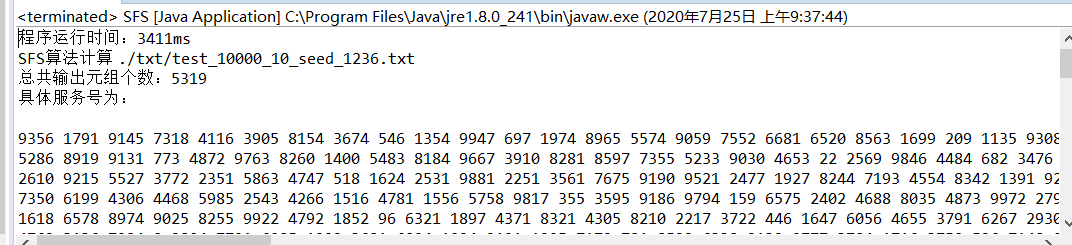
Z代表数据集属性个数

Seed\_1236代表使用1236作为随机数种子

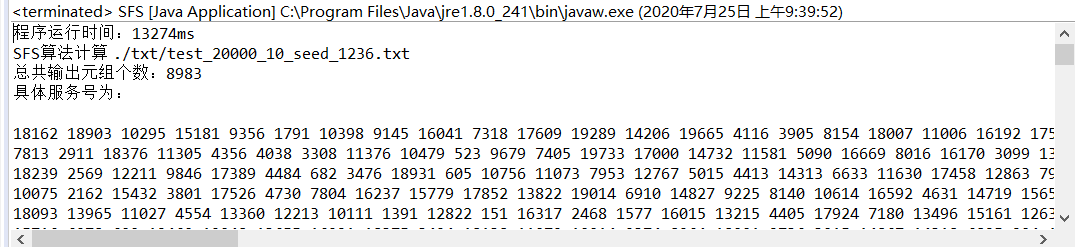
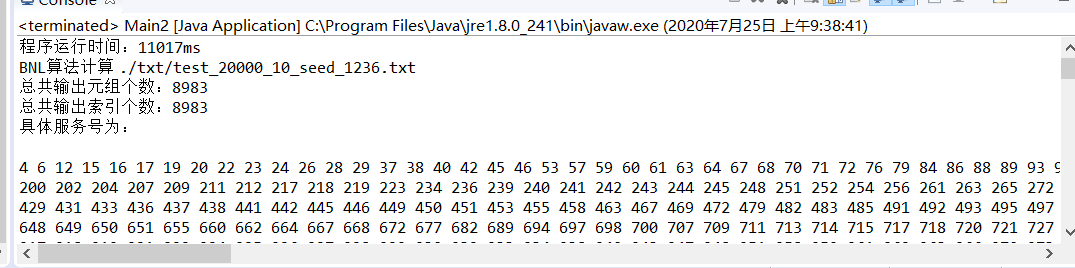
1. 运算结果正确性分析

Test\_10000:

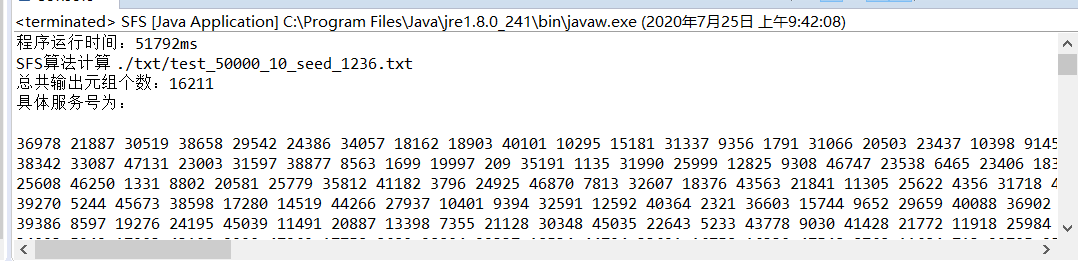
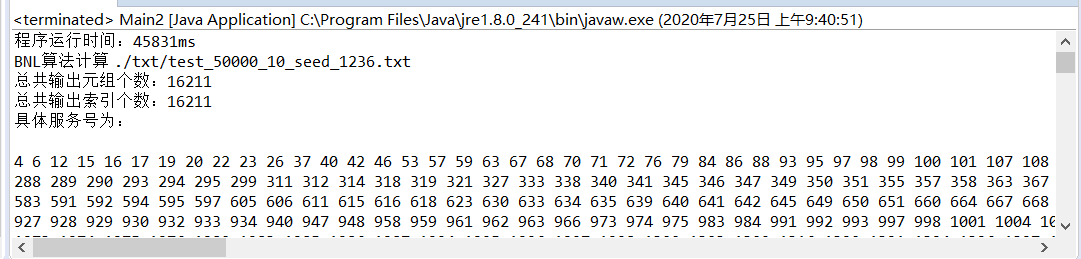




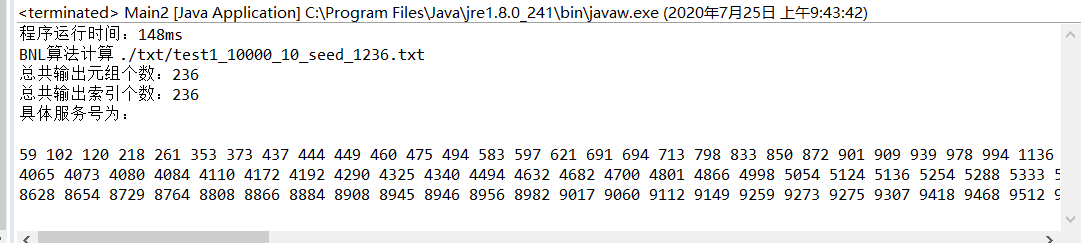
Test\_20000:

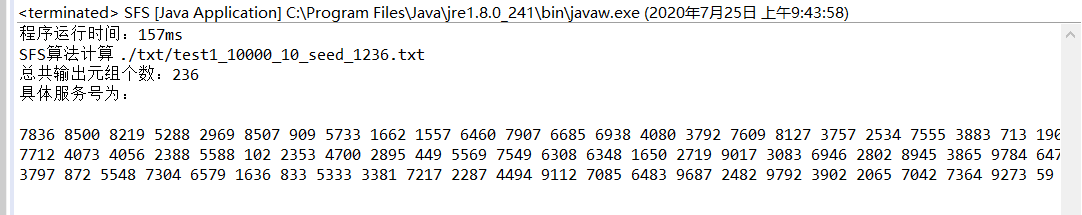


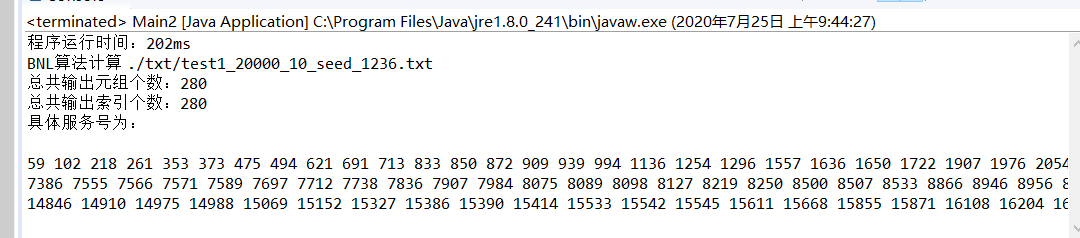
Test\_50000:

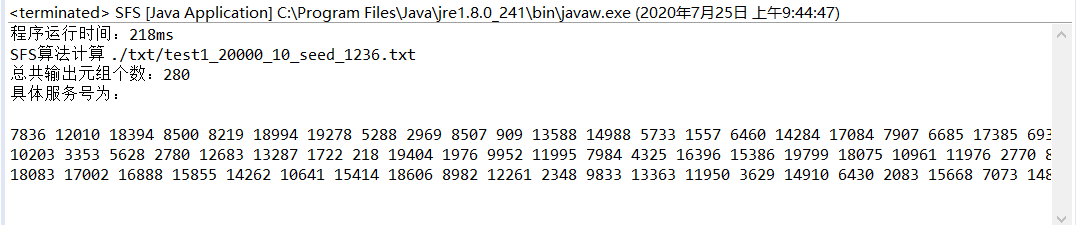


Test1\_10000:

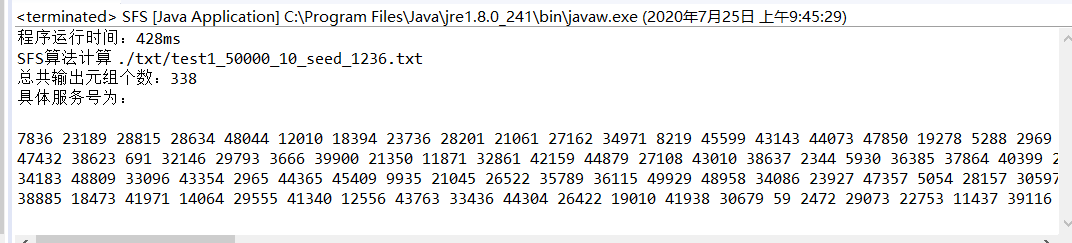
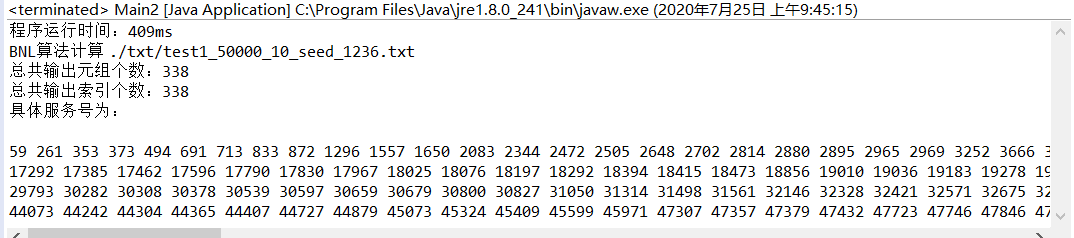


test1\_20000:

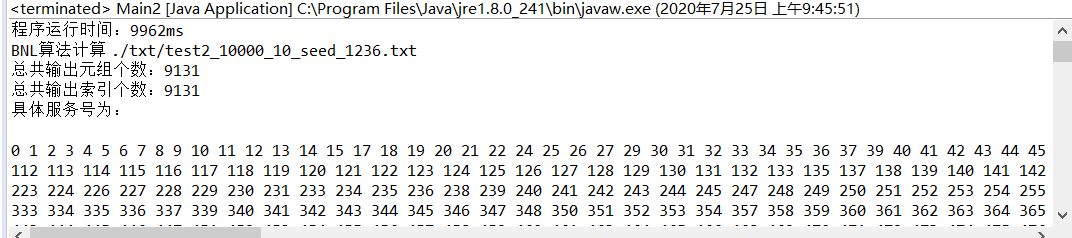


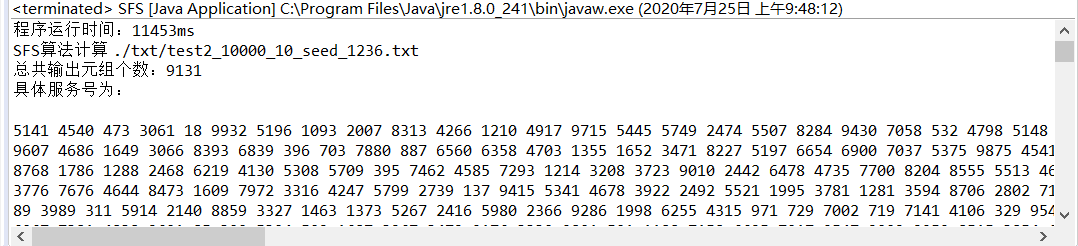


Test1\_50000:

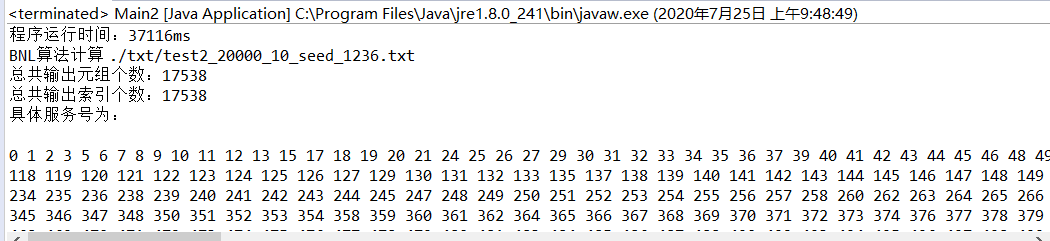


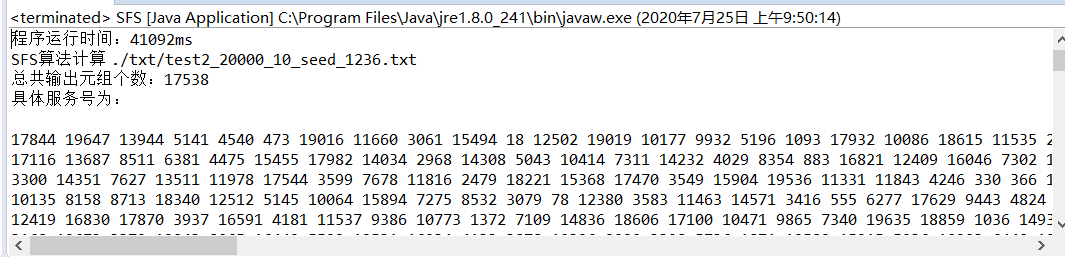
Test2\_10000:



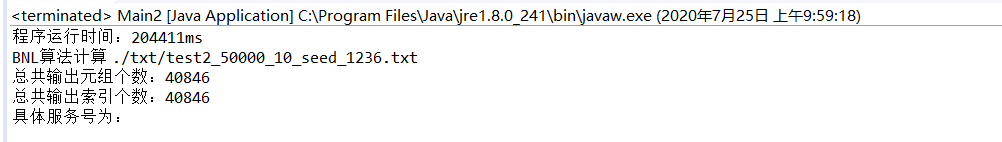


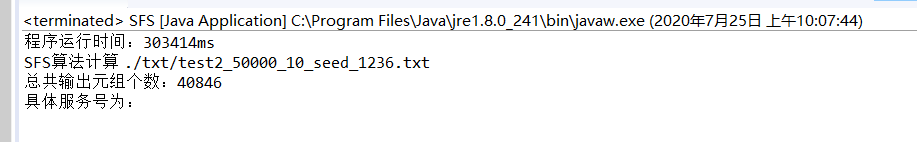
Test2\_20000:





Test2\_50000:





看结果可以看到，同样的数据集，结果都是一样的

1. 运算速度分析比较

BNL和SFS速度差别不是很大，应该是和SFS没有设置窗口大小有关系，但是在不同数据集上，算法时间大致趋势是一样的，正相关的数据集运算速度最快，时间最短，结果也最少，独立的适中，反相关的数据运算时间最长。同样类型的数据集，数据多大，运算时间越长，速度越慢。

1. 实验结果的启示

如果在SFS算法中设置窗口大小，根据SFS算法的原理，添加到窗口中的点，必然不能被后读入的点支配，那么窗口中的点就可以直接放到结果集合里面，这样迭代，运算时间应该是会比BNL算法小很多的。