HomeWork-2 频繁模式挖掘

21302010042 侯斌洋

一、文件介绍

- `data/`：存放数据集
- `Apriori.py`：Apriori算法实现
- `ECLAT.py`：Eclat算法实现
- `FPGrowth.py`：FP-Growth算法实现
- `FPMining.py`：主程序
- `README.md`：说明文档

二、运行结果

运行`FPMining.py`文件，输出结果如下：

```
Dataset len: 10000
Min support: 5

Running Apriori ...
Apriori time: 2.1716320514678955
Apriori result len: 36
[['H.VincentPoor'], ['MohsenGuizani'], ['Sebasti'], ['ShigekiSugano'],
['LichengJiao'], ['VinceD.Calhoun'], ['ChinChenChang0001'], ['SajalK.Das0001'],
['lvarez'], ['C.L.PhilipChen'], ['DonaldF.Towsley'], ['JieLi0002'], ['PengWang'],
['BernardDeBaets'], ['FathiE.AbdElSamie'], ['JingChen'], ['LiuqingYang0001'],
['YanLiu'], ['YingLi'], ['PengShi0001'], ['XuelongLi0001'], ['FranciscoHerrera'],
['GeyongMin'], ['LaurenceT.Yang'], ['FlorentinSmarandache'], ['JeanFran'],
['QianWang'], ['NanCheng'], ['XueminShen'], ['RobertSchober'], ['HsiaoHwaChen'],
['VictorC.M.Leung'], ['JieZhang'], ['XiaojiangDu'], ['NanCheng', 'XueminShen'],
['MohsenGuizani', 'XiaojiangDu']]

Running FPGrowth ...
FPGrowth time: 0.01661825180053711
FPGrowth result len: 36
[['JieZhang'], ['HsiaoHwaChen'], ['RobertSchober'], ['NanCheng'], ['NanCheng',
'XueminShen'], ['QianWang'], ['FlorentinSmarandache'], ['LaurenceT.Yang'],
['GeyongMin'], ['FranciscoHerrera'], ['XuelongLi0001'], ['YingLi'], ['YanLiu'],
['LiuqingYang0001'], ['FathiE.AbdElSamie'], ['BernardDeBaets'], ['PengWang'],
['JieLi0002'], ['DonaldF.Towsley'], ['C.L.PhilipChen'], ['SajalK.Das0001'],
['VinceD.Calhoun'], ['LichengJiao'], ['ShigekiSugano'], ['XiaojiangDu'],
['MohsenGuizani', 'XiaojiangDu'], ['VictorC.M.Leung'], ['XueminShen'],
['PengShi0001'], ['JingChen'], ['H.VincentPoor'], ['JeanFran'], ['lvarez'],
['ChinChenChang0001'], ['Sebasti'], ['MohsenGuizani']]

Running ECLAT ...
ECLAT time: 0.04321432113647461
ECLAT result len: 36
```

```
[['JieZhang'], ['HsiaoHwaChen'], ['RobertSchober'], ['NanCheng'], ['NanCheng',
'XueminShen'], ['QianWang'], ['FlorentinSmarandache'], ['LaurenceT.Yang'],
['GeyongMin'], ['FranciscoHerrera'], ['XuelongLi0001'], ['YingLi'], ['YanLiu'],
['LiuqingYang0001'], ['FathiE.AbdElSamie'], ['BernardDeBaets'], ['PengWang'],
['JieLi0002'], ['DonaldF.Towsley'], ['C.L.PhilipChen'], ['SajalK.Das0001'],
['VinceD.Calhoun'], ['LichengJiao'], ['ShigekiSugano'], ['XiaojiangDu'],
['XiaojiangDu', 'MohsenGuizani'], ['VictorC.M.Leung'], ['XueminShen'],
['PengShi0001'], ['JingChen'], ['H.VincentPoor'], ['JeanFran'], ['lvarez'],
['ChinChenChang0001'], ['Sebasti'], ['MohsenGuizani']]
```

三、实验结果分析

1. 在数据集长度为10000，最小支持度为5的情况下，三种算法得到的频繁模式相同。相同的结果可以在一定程度上说明三种算法的正确性。
2. 从运行时间上看，FP-Growth算法的运行时间最短，Apriori算法的运行时间最长，ECLAT算法的运行时间居中。
3. FP-Growth算法与ECALT算法的运行时间都远小于Apriori算法，这是因为Apriori算法在每次迭代时都需要扫描整个数据集，而FP-Growth算法与ECLAT算法都通过不同的方式减少数据集的扫描次数。