Motivation: 在前两问中，可以得出基本结论，Jaccard Distance不适合这类问题，因为矩阵太稀疏，那么距离矩阵也同样十分稀疏，最终的效果必然不佳。但是我们可以想办法把feature做一个agglomeration，合并后的特征能够使矩阵变得稠密，必然能够减少聚类的难度。

Process:

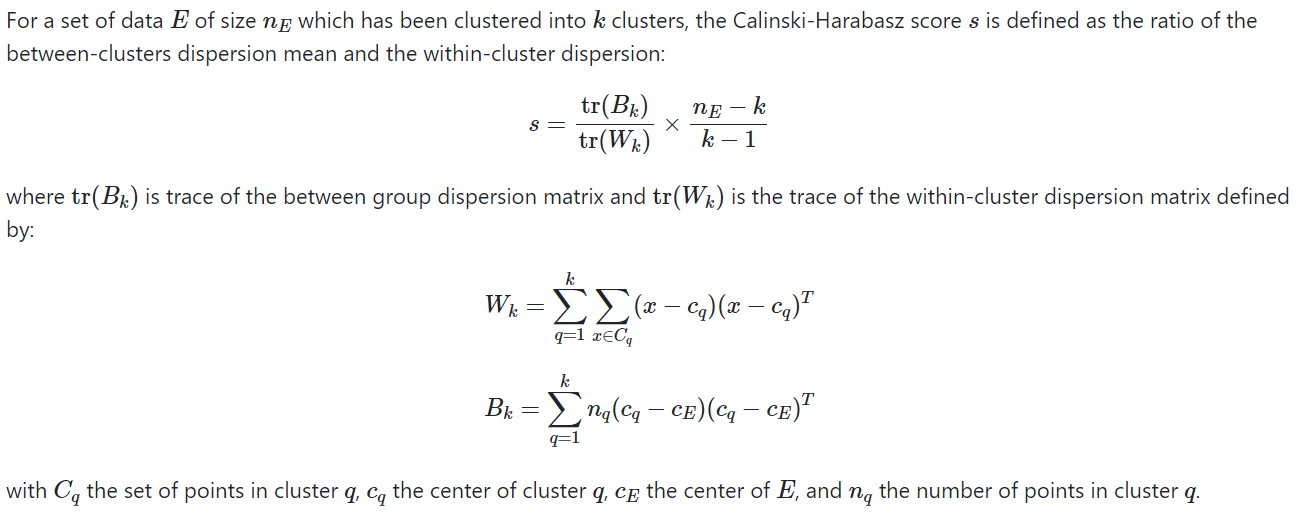
* 使用层次聚类进行 feature agglomeration
* 使用 bi-clustering 同时对 users 和 item-sets 进行谱聚类
* 计算 metrics

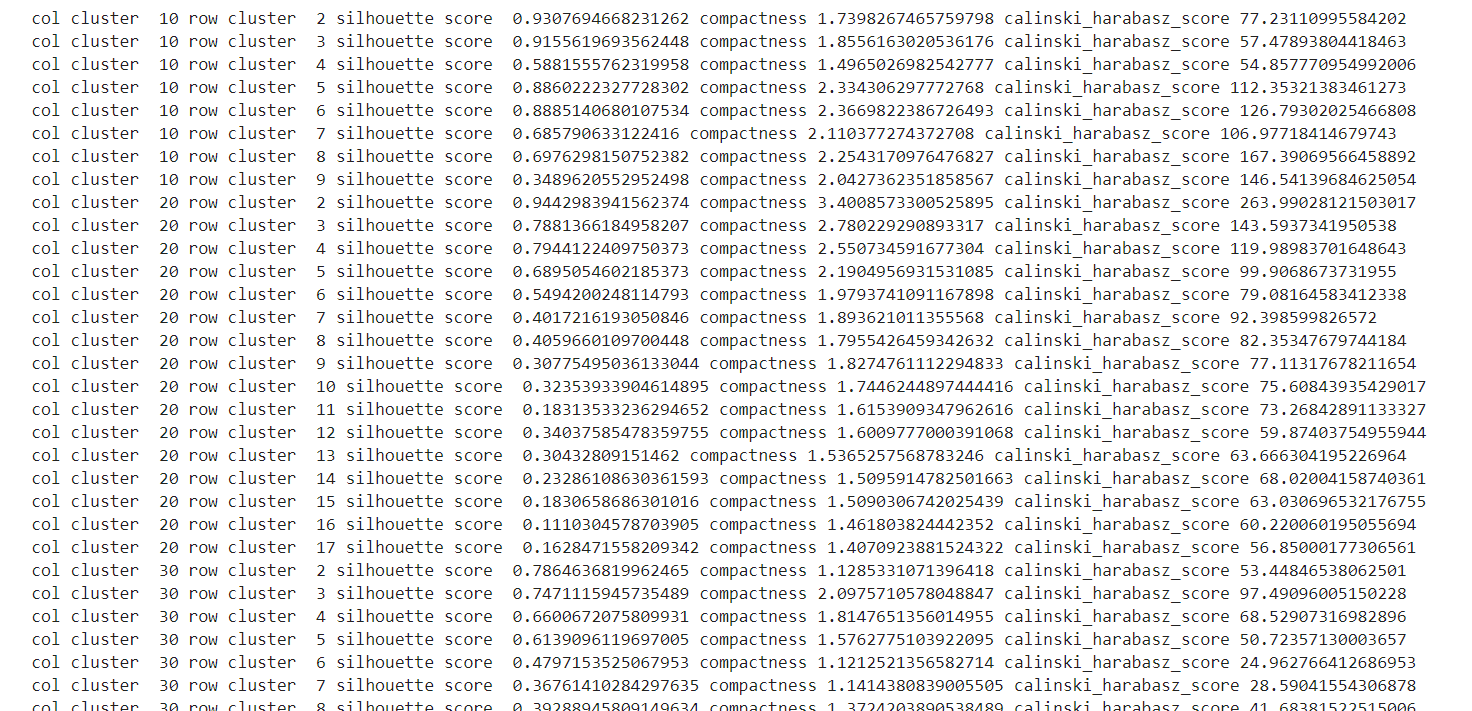
不同的参数对应的结果如下：

这是进行parameter tuning的输出

解释：

* Col cluster 表示 item cluster number
* Row cluster 表示 user cluster number
* Sc和cp 不能完全体现聚类结果，因此使用了 calinski\_harabasz\_score，该评分越高越好，这个评分同时体现了类内-类间的聚类标准





该算法的缺点：

* Cluster number 需要人为指定

该算法的优点：

* 运行速度快，能在1min内跑完