

# 作業(一)

學號: P77091106

姓名: 侯則瑜

Github: <https://github.com/houzeyu2683/Data-Mining.git>

使用 2022 released testdata 測試演算法			
Apriori			
min_sup	min_conf	run_time (sec)	item_set
0.1	0.1	194	1495
0.1	0.3	202	1495
0.3	0.1	2.52	50
0.3	0.3	2.55	50

在尋找不同組合的itemset frequency的時候, min\_conf不會影響速度, 只有min\_sup的限制會影響, min\_sup小, 會考慮出現頻率較少的itemset, 所以在掃描table的時候會花比較多的時間, 反之, min\_sup稍微調高點, 則會在header table建立的時候過濾掉不滿足條件的itemset, 在搜索上時間比較快。

至於min\_conf在哪裡會影像? 這邊有兩個想法。

1. 在搜索到所有的itemset frequency後, 計算兩個item set之間的confidence的時候會影響, 在計算兩個item set的confidence之前, 需要將聯集的support算出來, 如果聯集的support不在我們的itemset frequency中, 那就找不到。
2. 在搜索到所有的itemset frequency後, 基於前者, 對所有兩個item set的聯集, 如果聯集後的itemset在itemset frequency中, 就可計算彼此的confidence。

使用 2022 released testdata 測試演算法			
fptree			
min_sup	min_conf	run_time (sec)	item_set
0.1	0.1	194	465(X)
0.1	0.3	202	465(X)
0.3	0.1	2.52	50
0.3	0.3	2.55	50