

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	conf_item
0.1	0.1	194	1495	6009
0.1	0.3	202	1495	5328
0.3	0.1	2.52	50	9
0.3	0.3	2.55	50	9

在尋找 itemset frequency 的時候, 只有 min_sup 的限制會影響, 從上表可以知道 min_sup 的設定稍微有點變化, 找出來的 itemset frequency 就會差很多, 只要稍微調小 min_sup, 時間就會大幅增加, 找出來的 itemset 數量也是。

根據 min_sup 找出所有的 itemset frequency, 接著從裡面去跟據 confidence 的定義去兩兩計算, 當 min_sup 設為 0.1 的時候, min_conf 從 0.1 調高到 0.3, 少了 681 項, 當 min_sup 設定 0.3 的時候, 不管 min_conf 是 0.1 還是 0.3, 數量減少到 9 項。從上面的幾次實驗, 在這筆資料的情況下, min_sup 的設定相對於 min_conf 還要來的敏感。

以下使用 Kaggle 網站的教學文件中的 ([連結](#)) 資料集來進行測試

Grocery Store				
Apriori				
min_sup	min_conf	run_time (sec)	item_set	conf_item
0.1	0.1	0.02	46	40
0.1	0.3	0.02	46	40
0.3	0.1	2.52	6	0
0.3	0.3	2.55	6	0