

Project1(補交): Association Analysis

E54036219 張富嘉

I. Apriori Software usage

此次補交加了 Apriori 的部分，程式執行方式為：

```
./Apriori.py -f <datafile> -s minSupport -c minConfidence
```

預設值為 filename="data1.data"， minSupport=0.2， minConfidence=0.3。

II. Data Description

在 Data 資料夾中共有一大一小兩個資料檔，分別是 data1.data 和 data3.data，兩者皆為 IBM Data Generator 產生的資料集，data1 總共有 1000 個 items，1000 筆 transactions，而 data3 總共有 3000 個 items，10000 筆 transactions，其餘的參數設定皆為 IBM Data Generator 的預設值。

III. Result Comparison

在時間的比較方面，使用同樣一筆資料(data1.data)，單純計算演算法的執行時間(不包括印出 rules)，FP-Growth 所花費的時間為 9.3782 sec，而 Apriori algorithm 所花費的時間為 3.74344 sec，可以發現在較少的資料量之下 Apriori algorithm 所用的時間筆 FP-Growth 還要來得少。接下來比較較大的資料集(data3.data)，FP-Growth 總共花了 45.9102 sec，Apriori algorithm 總共花了 18.77387 sec，依然較為快速。

IV. Discussion

原本預期 Apriori algorithm 的速度應該為筆 FP-Growth 慢上許多，但是以此次測試的兩筆資料來看卻是相反的結果，推測的原因有可能是資料量不夠大，Apriori algorithm 在資料量更多的情況下執行速度應該就會比 FP-Growth 還要慢了。