**科⽬：1071\_資料探勘 DATA MINING**

**學⽣：楊沛霖**

**學號：Q36071156**

**日期：2018-10-23**

**Topic 1:**

先以老師上課資料確認程式運行是否正確。

Min\_sup : 2

執行時間 : 1.1854257

**使用資料(IBM Quest Data Generator):**

**FP-growth:**

老師講義上的圖(圖 二)與這次程式實作出來的圖(圖 一)

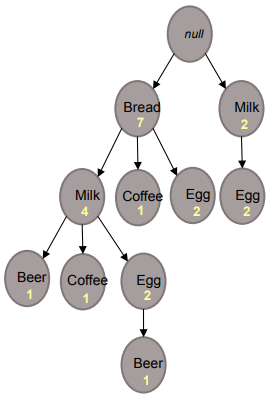
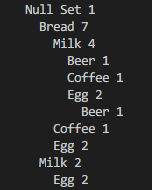


圖 一

圖 二

**Association analysis:**

分別為自己寫的程式(圖 四)與WEKA的FP-tree(圖 三)的關聯顯示,可以看到兩邊答案相同.

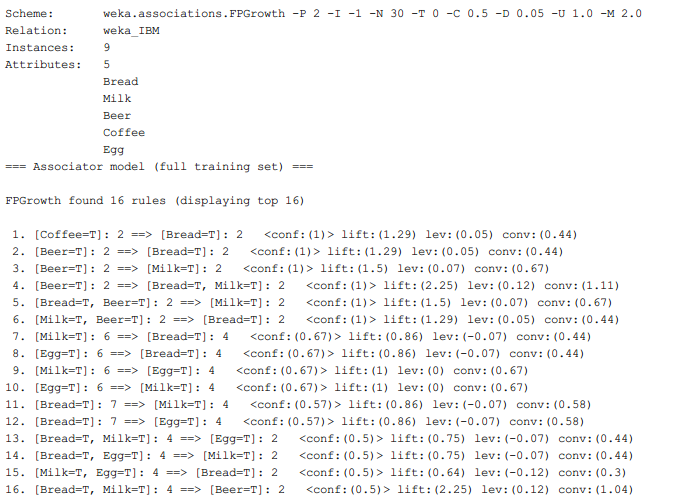
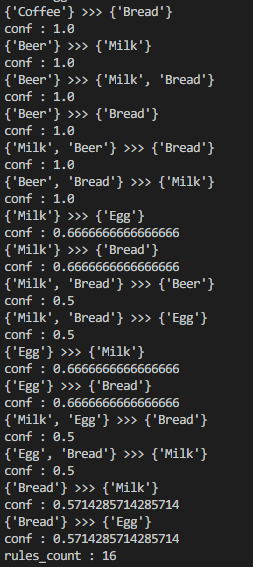
對應function為start.py中的associate(freqItems, simpDat)

圖 三

圖 四

比較結果:

**程式經過老師範例的洗禮過後**，**能正常運行並且和分析工具WEKA所顯示的資料相同**，**因此可以加入IBM資料去處理了**。

**Topic 2:**

實際運行IBM Quest Data Generator所生成的資料

Min\_sup : 2

執行時間 : 2.124941

**使用資料(IBM Quest Data Generator):**

|  |  |
| --- | --- |
| **TID** | **ItemSet** |
| 1 | {247,506,578} |
| 2 | {247,506,578} |
| 3 | {247,506,247} |
| 4 | {247,506,578} |
| 5 | {247,506,578} |
| 6 | {247,506,578} |

**使用參數 :**

lit -ntrans 1 -ascii -tlen 5 -nitems 10 -npats 3

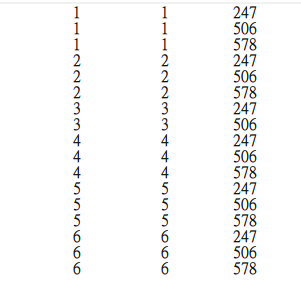
圖 五為IBM Quest Data Generator， 表一 為切換成老師投影片的形式。

圖 五

表 一

**FP-****growth:**

有程式生成的FP-growth(圖 六)

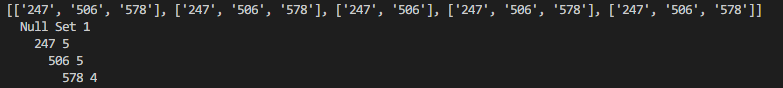


圖 六

**Association analysis:**

分別為自己寫的程式(圖 七)與WEKA的FP-tree(圖 八)的關聯顯示,可以看到兩邊答案相同.

對應function為start.py中的associate(freqItems, simpDat)

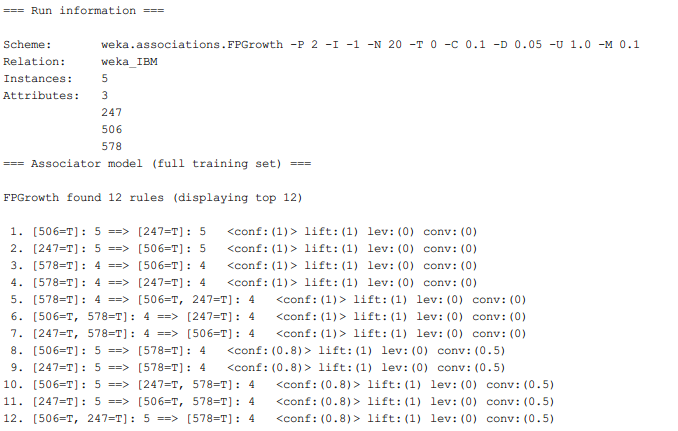
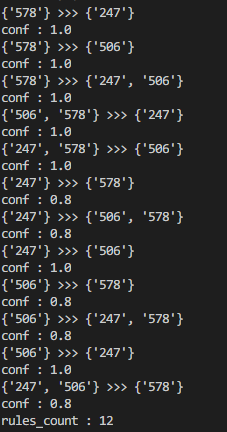


圖 七

圖 八

比較結果:

無論資料量大小皆與WEKA的分析結果一致，不過在自己associate的寫法中沒辦法像WEKA能夠把除了True或False的資料顯示出來，只能簡單的用存不存在去做對比。