國立雲林科技大學  
資訊管理研究所

資料探勘專案作業4

指導教授：許中川教授  
學生：M11123039 廖顯億  
 M11123046 周子懷  
 M11123052 賴俊佑  
 M11123062 陳靖穎

Data Mining Homework: 交易資料關聯規則分析

摘要

Covid-19疫情的關係使得宅經濟變成普遍的消費習慣，在這百家爭鳴的電商平台市場，如何掌握消費者的商機成為非常重要的課題。本研究藉由Apriori演算法及FP-Growth演算法分析出適合推薦產品的方向。相比之下發現FP-Growth演算法具有較快的分析時間。

**關鍵字**：宅經濟、關聯規則分析、Apriori演算法、FP-Growth演算法

1. **緒論**

Covid-19疫情期間受政府防疫政策的影響，大多數的人們只能待在家裡，不管是手機遊戲、串流平台、網路購物平台甚至是外賣等等，宅經濟幾乎成為大多數人的消費行為（李沃牆，2020）。如何在這宅經濟的時代裡，透過精準的推薦相關產品使營收增加（黃琬琪，2014），已成為大多數網購平台發展的方向。

* 1. **動機**

在這電子商務平台飽和的時代，市場競爭亦非常激烈，尤其網路跨越地域，廣大的消費市場，如何快速地掌握商機，如何創造使消費者更願意花錢的動機已成為每個電商平台研究的課題。

* 1. **目的**

為了有更好的營收，找到潛在的商品組合來增加消費動機是很重要的，例如經典的案例，在禮拜五擺放尿布與啤酒的商品，創造了意外的收入。本研究期望利用消費紀錄來推薦相關可能購買的產品。

1. **方法**

本研究預計使用Apriori演算法及FP-Growth演算法來分析不同參數的情況下推薦產品數量多寡的關係。匯入資料後先進行資料預處理，將數量為零或負值的交易代表退貨或註銷的數值去除，去除後再將資料分組並模型資料，接著進行Apriori演算法及FP-Growth演算法兩個演算法分析，並將分析出來的規則進行存檔以便下次使用。

1. 實驗
2. 資料集

* 資料集名稱：交易資料集
* 資料筆數：157396
* 屬性數量：7

表1 交易資料集屬性簡介

|  |  |  |
| --- | --- | --- |
| 屬性名稱 | 型態 | 尺度 |
| ITEM\_ID | Int | 名目 |
| ITEM\_NO | String | 名目 |
| PRODUCT\_TYPE | String | 名目 |
| CUST\_ID | Int | 名目 |
| TRX\_DATE | Date | 名目 |
| INVOICE\_NO | String | 名目 |
| QUANTITY | Int | 名目 |

1. 前置處理

數量為零或負值的交易代表退貨或註銷，非本研究所需的資料，因此在前處理階段將其剔除。

1. 實驗設計

一開始先進行預處理，選擇需要的屬性ITEM\_ID及INVOICE\_NO，並利用INVOICE\_NO做資料分組。使用Apriori與FP-Growth演算法分析關聯規則，並比較兩種演算法所需花費的時間。

1. 實驗結果

表2 支持度0.0008所花費時間

|  |  |
| --- | --- |
| 演算法 | 花費時間（秒） |
| Apriori | 22.528 |
| FP-Growth | 1.096 |

表 3支持度0.001所花費時間

|  |  |
| --- | --- |
| 演算法 | 花費時間（秒） |
| Apriori | 10.534 |
| FP-Growth | 1.069 |

表 4支持度0.002所花費時間

|  |  |
| --- | --- |
| 演算法 | 花費時間（秒） |
| Apriori | 1.709 |
| FP-Growth | 1.052 |

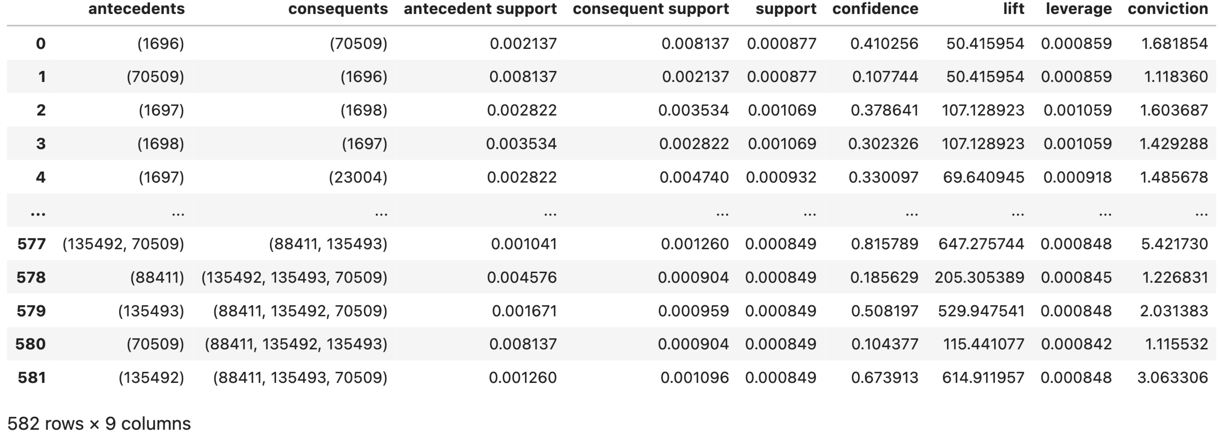


圖1 Apriori支持度0.0008之關聯結果

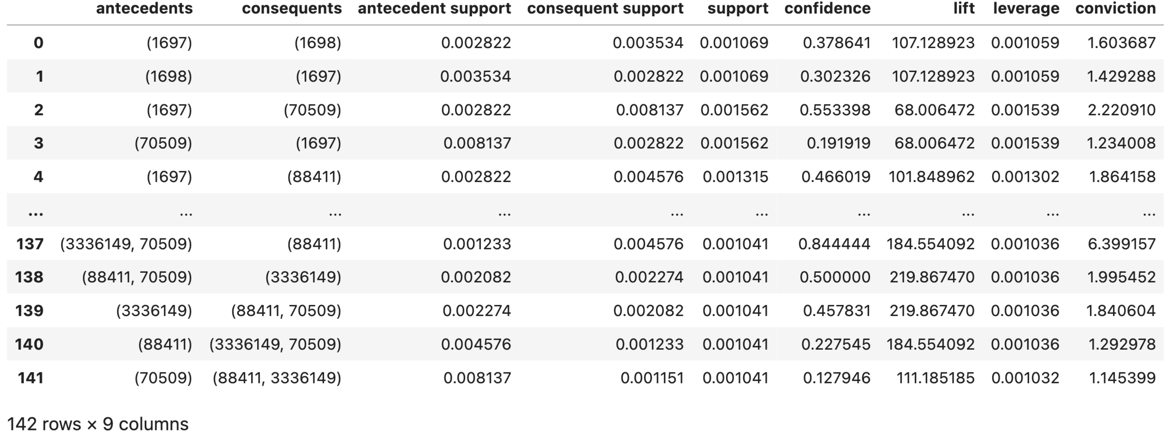


圖2 Apriori支持度0.001之關聯結果

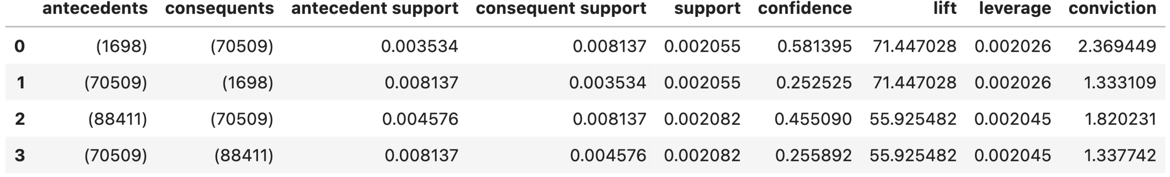


圖3 Apriori支持度0.002之關聯結果

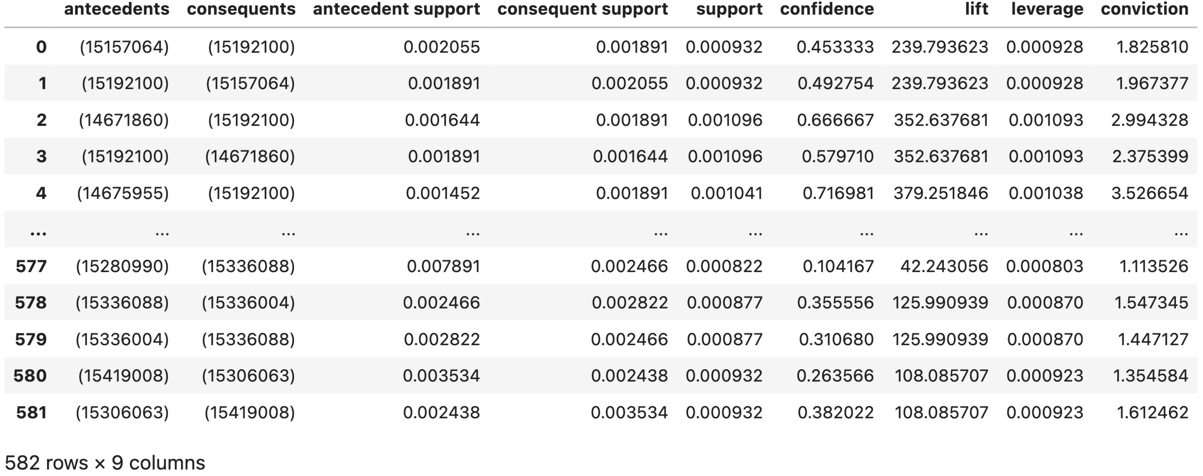


圖 4 FP-Growth支持度0.0008之關聯結果

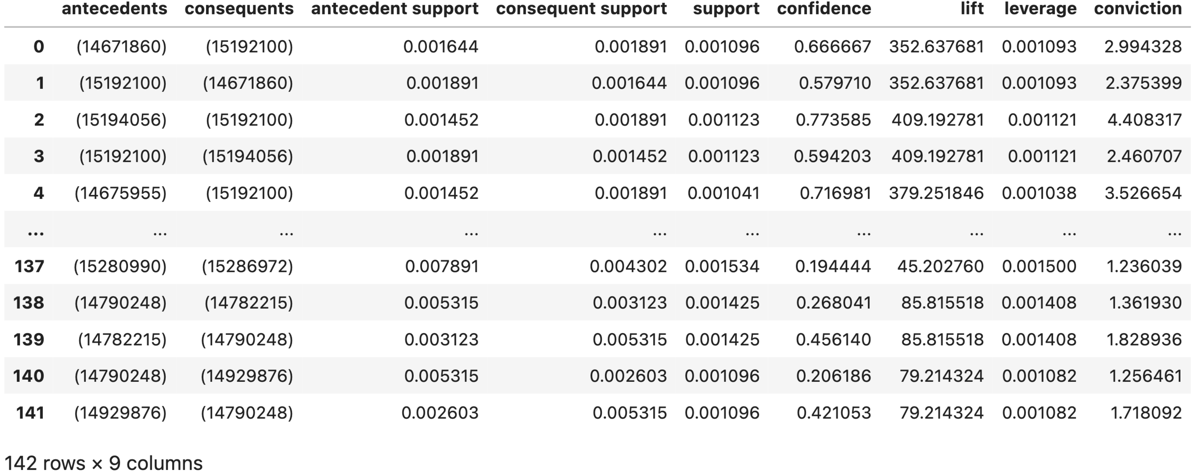


圖 5 FP-Growth支持度0.001之關聯結果

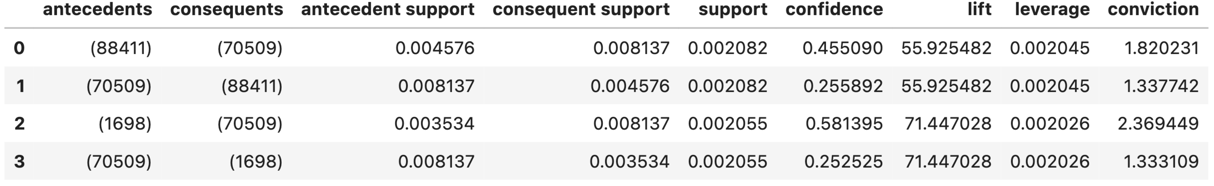


圖 6 FP-Growth支持度0.002之關聯結果

1. 結論

透過演算法的分析結果可以發現FP-Growth演算法所花費的分析時間較少。當支持度設定0.0008時，表示10,000次購買紀錄中只需出現過8次組合即可形成關聯規則，因此關聯規則的數量較多；支持度設定0.001時為10,000次購買紀錄出現過10次組合會形成關聯規則，而支持度0.002時為10,000次購買紀錄需要出現過20次組合才能形成關聯規則，因此支持度0.001與0.002相比，支持度0.001會出現較多的關聯規則。

參考文獻

Mike Ku (2020, November 1). [Pandas教學]資料分析必懂的Pandas DataFrame處理雙維度資料方法. LEARN CODE WITH MIKE.

<https://www.learncodewithmike.com/2020/11/python-pandas-dataframe-tutorial.html>

plusone (2017, December 29). [Day10]Pandas Groupby使用！ iT邦幫忙.

<https://ithelp.ithome.com.tw/articles/10194027>

Sebastian Raschka (2022, December 24). apriori: Frequent itemsets via the Apriori algorithm. mlxtend.

<http://rasbt.github.io/mlxtend/user_guide/frequent_patterns/apriori/#apriori-frequent-itemsets-via-the-apriori-algorithm>

Sebastian Raschka (2022, December 25). association\_rules: Association rules generation from frequent itemsets. mlxtend.

<http://rasbt.github.io/mlxtend/user_guide/frequent_patterns/association_rules/>

李沃牆（2020）。新冠肺炎引爆全球 [宅經濟] 商機。*會計研究月刊*，(413)，15-20。

黃琬琪（2014）。推薦產品之相關性對購物經驗之影響〔碩士論文，國立臺北科技大學〕。華藝線上圖書館。