

國立雲林科技大學資訊管理系

資料探勘

專案作業四

組員

M11123037 許禾諭

M11123043 謝永盛

M11123050 李京樺

M11123053 葉家豪

摘要

近年來由於疫情導致供應鏈出現了一些問題，在疫情期間大量的生產線受到疫情的影響，而貨運的情況也出現了一些問題。終端產品的短缺使得生產方上游加大力度生產產品，但後疫情時代來臨，終端產品的需求卻沒有回溫，使得長鞭效應逐漸擴大造成廠商庫存成本創新高，因此本研究計畫利用 Apriori 以及 FP-Growth 進行關聯分析來研究終端產品的需求，已得出最佳的產品組合並盡可能的降低庫存。

關鍵字: Apriori、FP-Growth、關聯分析、降低庫存

壹、緒論

一、動機

於 2019 年在中國武漢市爆發的嚴重特殊傳染性肺炎(維基百科,嚴重特殊傳染性肺炎),迅速蔓延到全世界,而中國有世界工廠的美名(維基百科,中國製造),因此全世界許多工廠皆設於此,但中國不僅是疫情爆發的國家,同時疫情在近日正迅速升溫(王能斌, 2022),澳洲財政部長查默斯表示中國疫情的升溫可能影響到全世界的供應鏈(陳冠宇, 2023),此次疫情不僅影響了人們的健康,同時也影響了通運,間接影響了供應鏈,如何在現有資源下創造出最佳的效益,將是本研究探討的重點。

二、目的

本研究透過關聯分析,能夠使生產方依據此分析結果優化產品出貨組合,進一步降低庫存水平,以減少庫存損失。

貳、資料集

一、真實資料集

本研究使用的交易資料集共有六個欄位,資料有 157,396 筆,以下為欄位中英對照表(如表 1 所示)。

表 1
欄位中英對照表

中文	英文
產品編號	ITEM_ID
產品項目	ITEM_NO
產品名稱	PRODUCT_TYPE
客戶編號	CUST_ID
交易日期	TRX_DATE
發票號碼	INVOICE_NO
數量	QUANTITY

參、方法

一、實作說明

載入所需的函式庫後,在進行資料的前處理,處理需剔除的空值、負值和 Null 的資料,比較 Apriori 以及 FP-Growth 是如何影響分析結果的。

二、操作說明

下載 GitHub 上的資料集以及程式,使用 Jupyter 開發環境讀取檔案的路徑,確定檔案可成功讀取後,即可執程式。

肆、實驗

一、前置處理

- (一)、先將原始資料集轉為 csv2l 檔後使用 Dropna 方法將資料集匯入 python 程式,剔除為空值的資料
- (二)、因數量為零和負值的交易資料代表退貨或註銷,故予以剔除。

(三)、避免產品名稱 (PRODUCT_TYPE) 有 Null 值影響關聯規則分析結果，故予以剔除。

二、實驗設計

本研究將交易資料集讀入程式後進行資料前處理篩選出我們需要的資訊，完成資料前處理後我們利用 Apriori 以及 FP-Growth 來分析各項產品之間的關係，並比較之間的執行時間以及參數不同會如何影響結果，最後也利用統整出來的數據分析廠商在購買零組件時的情況。

三、實驗結果

(1) Apriori

首先我們將 Apriori 的 minsupport 設定為 0.002，並列出執行時間以及在此設定下關聯的項目，如圖 1 所示。

執行時間: 0.0320526999999949 秒

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(CHIPSET / ASP)	(CPU / MPU)	0.044112	0.110746	0.005343	0.121118	1.093658	0.000458	1.011802
1	(CPU / MPU)	(CHIPSET / ASP)	0.110746	0.044112	0.005343	0.048243	1.093658	0.000458	1.004341
2	(LINEAR IC)	(LOGIC IC)	0.260398	0.121541	0.043454	0.166877	1.373012	0.011805	1.054417
3	(LOGIC IC)	(LINEAR IC)	0.121541	0.260398	0.043454	0.357529	1.373012	0.011805	1.151184
4	(MEMORY_SYSTEM)	(MEMORY_EMBEDDED)	0.021234	0.066716	0.002055	0.096774	1.450540	0.000638	1.033279
...
91	(OTHERS, LOGIC IC)	(LINEAR IC, DISCRETE)	0.003754	0.034906	0.002219	0.591241	16.938077	0.002088	2.361033
92	(LINEAR IC)	(DISCRETE, OTHERS, LOGIC IC)	0.260398	0.002767	0.002219	0.008523	3.079827	0.001499	1.005805
93	(DISCRETE)	(LINEAR IC, OTHERS, LOGIC IC)	0.257905	0.002849	0.002219	0.008605	3.019901	0.001484	1.005806
94	(OTHERS)	(LINEAR IC, DISCRETE, LOGIC IC)	0.113403	0.014713	0.002219	0.019570	1.330100	0.000551	1.004954
95	(LOGIC IC)	(LINEAR IC, DISCRETE, OTHERS)	0.121541	0.003397	0.002219	0.018260	5.374535	0.001806	1.015139

96 rows × 9 columns

圖 1 Apriori 的 minsupport 設定為 0.002

其次我們將 Apriori 的 minsupport 設定為 0.005，並列出執行時間以及在此設定下關聯的項目，如圖 2 所示。

執行時間: 0.0200642000000002 秒

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(CHIPSET / ASP)	(CPU / MPU)	0.044112	0.110746	0.005343	0.121118	1.093658	0.000458	1.011802
1	(CPU / MPU)	(CHIPSET / ASP)	0.110746	0.044112	0.005343	0.048243	1.093658	0.000458	1.004341
2	(LINEAR IC)	(LOGIC IC)	0.260398	0.121541	0.043454	0.166877	1.373012	0.011805	1.054417
3	(LOGIC IC)	(LINEAR IC)	0.121541	0.260398	0.043454	0.357529	1.373012	0.011805	1.151184
4	(LINEAR IC, DISCRETE)	(LOGIC IC)	0.034906	0.121541	0.014713	0.421507	3.468026	0.010471	1.518530
5	(LINEAR IC, LOGIC IC)	(DISCRETE)	0.043454	0.257905	0.014713	0.338588	1.312841	0.003506	1.121986
6	(DISCRETE, LOGIC IC)	(LINEAR IC)	0.026988	0.260398	0.014713	0.545178	2.093634	0.007686	1.626134
7	(LINEAR IC)	(DISCRETE, LOGIC IC)	0.260398	0.026988	0.014713	0.056503	2.093634	0.007686	1.031282
8	(DISCRETE)	(LINEAR IC, LOGIC IC)	0.257905	0.043454	0.014713	0.057049	1.312841	0.003506	1.014417
9	(LOGIC IC)	(LINEAR IC, DISCRETE)	0.121541	0.034906	0.014713	0.121055	3.468026	0.010471	1.098014

圖 2 Apriori 的 minsupport 設定為 0.005

最後我們將 Apriori 的 minsupport 設定為 0.008，並列出執行時間以及在此設定下關聯的項目，如圖 3 所示。

執行時間：0.0133369000000054 秒

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(LINEAR IC)	(LOGIC IC)	0.260398	0.121541	0.043454	0.166877	1.373012	0.011805	1.054417
1	(LOGIC IC)	(LINEAR IC)	0.121541	0.260398	0.043454	0.357529	1.373012	0.011805	1.151184
2	(LINEAR IC, DISCRETE)	(LOGIC IC)	0.034906	0.121541	0.014713	0.421507	3.468026	0.010471	1.518530
3	(LINEAR IC, LOGIC IC)	(DISCRETE)	0.043454	0.257905	0.014713	0.338588	1.312841	0.003506	1.121986
4	(DISCRETE, LOGIC IC)	(LINEAR IC)	0.026988	0.260398	0.014713	0.545178	2.093634	0.007686	1.626134
5	(LINEAR IC)	(DISCRETE, LOGIC IC)	0.260398	0.026988	0.014713	0.056503	2.093634	0.007686	1.031282
6	(DISCRETE)	(LINEAR IC, LOGIC IC)	0.257905	0.043454	0.014713	0.057049	1.312841	0.003506	1.014417
7	(LOGIC IC)	(LINEAR IC, DISCRETE)	0.121541	0.034906	0.014713	0.121055	3.468026	0.010471	1.098014

圖 3 Apriori 的 minsupport 設定為 0.008

(2) FP-Growth

首先我們將 FP-Growth 的 minsupport 設定為 0.002，並列出執行時間以及在此設定下關聯的項目，如圖 4 所示

執行時間：0.2244238000000028 秒

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(LOGIC IC, OTHERS)	(LINEAR IC)	0.003754	0.260398	0.002849	0.759124	2.915247	0.001872	3.070470
1	(OTHERS, LINEAR IC)	(LOGIC IC)	0.005233	0.121541	0.002849	0.544503	4.479995	0.002213	1.928571
2	(LOGIC IC)	(OTHERS, LINEAR IC)	0.121541	0.005233	0.002849	0.023445	4.479995	0.002213	1.018649
3	(LINEAR IC)	(LOGIC IC, OTHERS)	0.260398	0.003754	0.002849	0.010943	2.915247	0.001872	1.007269
4	(LOGIC IC, OTHERS)	(DISCRETE)	0.003754	0.257905	0.002767	0.737226	2.858524	0.001799	2.824085
...
91	(LOGIC IC)	(OPTICAL AND SENSOR, DISCRETE)	0.121541	0.005562	0.002192	0.018034	3.242436	0.001516	1.012701
92	(OPTICAL AND SENSOR)	(LOGIC IC, DISCRETE)	0.039865	0.026988	0.002192	0.054983	2.037323	0.001116	1.029624
93	(DISCRETE)	(LOGIC IC, OPTICAL AND SENSOR)	0.257905	0.002438	0.002192	0.008499	3.485307	0.001563	1.006112
94	(CPU / MPU)	(CHIPSET / ASP)	0.110746	0.044112	0.005343	0.048243	1.093658	0.000458	1.004341
95	(CHIPSET / ASP)	(CPU / MPU)	0.044112	0.110746	0.005343	0.121118	1.093658	0.000458	1.011802

96 rows x 9 columns

圖 4 FP-Growth 的 minsupport 設定為 0.002

其次我們將 FP-Growth 的 minsupport 設定為 0.005，並列出執行時間以及在此設定下關聯的項目，如圖 5 所示

執行時間：0.1681397999999962 秒

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(LOGIC IC)	(LINEAR IC)	0.121541	0.260398	0.043454	0.357529	1.373012	0.011805	1.151184
1	(LINEAR IC)	(LOGIC IC)	0.260398	0.121541	0.043454	0.166877	1.373012	0.011805	1.054417
2	(LOGIC IC, LINEAR IC)	(DISCRETE)	0.043454	0.257905	0.014713	0.338588	1.312841	0.003506	1.121986
3	(LOGIC IC, DISCRETE)	(LINEAR IC)	0.026988	0.260398	0.014713	0.545178	2.093634	0.007686	1.626134
4	(LINEAR IC, DISCRETE)	(LOGIC IC)	0.034906	0.121541	0.014713	0.421507	3.468026	0.010471	1.518530
5	(LOGIC IC)	(LINEAR IC, DISCRETE)	0.121541	0.034906	0.014713	0.121055	3.468026	0.010471	1.098014
6	(LINEAR IC)	(LOGIC IC, DISCRETE)	0.260398	0.026988	0.014713	0.056503	2.093634	0.007686	1.031282
7	(DISCRETE)	(LOGIC IC, LINEAR IC)	0.257905	0.043454	0.014713	0.057049	1.312841	0.003506	1.014417
8	(CPU / MPU)	(CHIPSET / ASP)	0.110746	0.044112	0.005343	0.048243	1.093658	0.000458	1.004341
9	(CHIPSET / ASP)	(CPU / MPU)	0.044112	0.110746	0.005343	0.121118	1.093658	0.000458	1.011802

圖 5 FP-Growth 的 minsupport 設定為 0.005

最後我們將 FP-Growth 的 minsupport 設定為 0.008，並列出執行時間以及在此設定下關聯的項目，如圖 6 所示

執行時間: 0.1670167000000046 秒

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(LOGIC IC)	(LINEAR IC)	0.121541	0.260398	0.043454	0.357529	1.373012	0.011805	1.151184
1	(LINEAR IC)	(LOGIC IC)	0.260398	0.121541	0.043454	0.166877	1.373012	0.011805	1.054417
2	(LOGIC IC, LINEAR IC)	(DISCRETE)	0.043454	0.257905	0.014713	0.338588	1.312841	0.003506	1.121986
3	(LOGIC IC, DISCRETE)	(LINEAR IC)	0.026988	0.260398	0.014713	0.545178	2.093634	0.007686	1.626134
4	(LINEAR IC, DISCRETE)	(LOGIC IC)	0.034906	0.121541	0.014713	0.421507	3.468026	0.010471	1.518530
5	(LOGIC IC)	(LINEAR IC, DISCRETE)	0.121541	0.034906	0.014713	0.121055	3.468026	0.010471	1.098014
6	(LINEAR IC)	(LOGIC IC, DISCRETE)	0.260398	0.026988	0.014713	0.056503	2.093634	0.007686	1.031282
7	(DISCRETE)	(LOGIC IC, LINEAR IC)	0.257905	0.043454	0.014713	0.057049	1.312841	0.003506	1.014417

圖 6 FP-Growth 的 minsupport 設定為 0.008

(3) 推薦功能

除了根據不同的參數儲存關聯規則之外，我們同時也可以根據關聯規則進行查詢，如圖 7 所示。

以下為產品清單

```
=====
(CHIPSET / ASP) (CPU / MPU) (DISCRETE) (LINEAR IC) (LOGIC IC)
(MEMORY_EMBEDDED) (MEMORY_SYSTEM) (OPTICAL AND SENSOR) (OTHERS) (PEMCO)
=====
```

請輸入產品名稱:(如需查詢多項產品請以逗號隔開，例如:CPU / MPU,DISCRETE)
CPU / MPU,DISCRETE

minsupport共有三種可供選擇，分別為0.002,0.005,0.008

請輸入您要查詢的參數:

0.002

=====

依據關聯規則向您推薦 ['LINEAR IC', 'LOGIC IC']

圖 7 產品推薦功能

結論

根據本研究實驗結果可得知，雖然理論上 FP-Growth 會較 Apriori 快速，但在本實驗中兩者並無顯著差異，推估可能是跟選取的欄位有關。另外透過調整 minsupport 有助於所篩選資料，若是需要更緊密的關聯可將 minsupport 調高一點，就可得出更頻繁出現的關係。

參考文獻

- [1] fpgrowth: Frequent itemsets via the FP-growth 。
algorithmhttp://rasbt.github.io/mlxtend/user_guide/frequent_patterns/fpgrowth/
- [2] 王能斌.(2022 年 12 月 14 日).中國疫情升溫百度「發燒」搜索暴增 6 倍 官方今起不公布無症狀感染者數據.擷取自 上報:
https://www.upmedia.mg/news_info.php?Type=3&SerialNo=161565
- [3] 陳冠宇.(2023 年 1 月 4 日). 澳洲財長示警：應為供應鏈受大陸疫情衝擊做準備.擷取自 中時新聞網：
<https://www.chinatimes.com/realtimenews/20230104002026-260409?chdtv>
- [4] 維基百科.中國製造.擷取自 維基百科: <https://zh.wikipedia.org/zh-tw/%E4%B8%AD%E5%9C%8B%E8%A3%BD%E9%80%A0>
- [5] 維基百科.嚴重特殊傳染性肺炎.擷取自 維基百科:
<https://zh.m.wikipedia.org/zh-tw/2019%E5%86%A0%E7%8A%B6%E7%97%85%E6%AF%92%E7%97%85>