P77111037 樊紹萱 - hw1 answers

Testdata: ibm-2022.txt

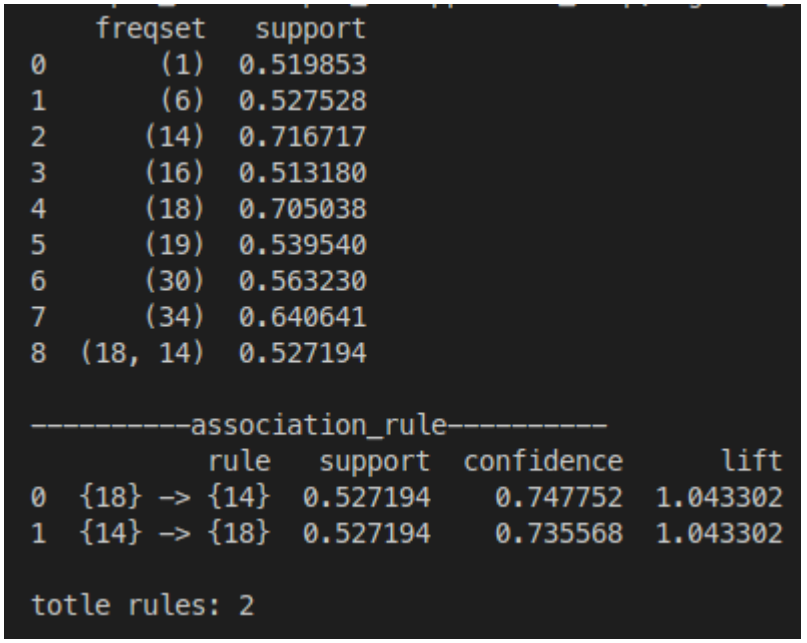
Q: What do you observe in the below 4 scenarios?

- **High support, high confidence**
  - 若 `min_sup = 0.5, min_conf = 0.5`，我的輸出結果如下圖，發現找到的 association rule 數量較少，因為minimum support & minimum confidence 的門檻都較高。
  -

```
    freqset   support
0       (1)  0.519853
1       (6)  0.527528
2      (14)  0.716717
3      (16)  0.513180
4      (18)  0.705038
5      (19)  0.539540
6      (30)  0.563230
7      (34)  0.640641
8  (18, 14)  0.527194

----------association_rule----------
           rule   support  confidence      lift
0  {18} -> {14}  0.527194    0.747752  1.043302
1  {14} -> {18}  0.527194    0.735568  1.043302

totle rules: 2
```

- **High support, low confidence**
  - 若 `min_sup = 0.5, min_conf = 0.1`，我的輸出結果如下圖，但我發現在這筆測試資料中，這次的輸出結果和上一題 (`min_sup = 0.5, min_conf = 0.5`) 一樣。
  - 

```
     freqset   support
0        (1)  0.519853
1        (6)  0.527528
2       (14)  0.716717
3       (16)  0.513180
4       (18)  0.705038
5       (19)  0.539540
6       (30)  0.563230
7       (34)  0.640641
8   (18, 14)  0.527194

----------association_rule----------
            rule   support  confidence      lift
0  {18} -> {14}  0.527194    0.747752  1.043302
1  {14} -> {18}  0.527194    0.735568  1.043302

totle rules: 2
```

- **Low support, low confidence**
  - 若 `min_sup = 0.1, min_conf = 0.1`，我的輸出結果如下圖，因為門檻都設的較低，所以找到`association rule` 數量非常多
  - ```
                        freqset   support
    0                       (0)  0.284284
    1                       (1)  0.519853
    2                       (2)  0.272606
    3                       (3)  0.238238
    4                       (4)  0.396730
    ...                     ...       ...
    1485  (14, 48, 18, 19, 30)  0.101768
    1486  (34, 14, 48, 18, 19)  0.112779
    1487  (34, 14, 18, 26, 30)  0.104771
    1488  (32, 34, 14, 18, 30)  0.102769
    1489  (34, 14, 48, 18, 30)  0.119453

    [1490 rows x 2 columns]

    ----------association_rule----------
                               rule   support  confidence      lift
    0                     {0} -> {1}  0.152486    0.536385  1.031801
    1                     {1} -> {0}  0.152486    0.293325  1.031801
    2                     {0} -> {4}  0.112446    0.395540  0.997000
    3                     {4} -> {0}  0.112446    0.283431  0.997000
    4                     {0} -> {6}  0.153153    0.538732  1.021240
    ...                          ...       ...         ...       ...
    10281  {34} -> {48, 18, 14, 30}  0.119453    0.186458  1.128920
    10282  {14} -> {48, 34, 18, 30}  0.119453    0.166667  1.153580
    10283  {48} -> {34, 18, 14, 30}  0.119453    0.249477  1.155616
    10284  {18} -> {48, 34, 14, 30}  0.119453    0.169427  1.123393
    10285  {30} -> {48, 34, 18, 14}  0.119453    0.212085  1.133012

    [10286 rows x 4 columns]

    totle rules: 10286
    ```
- **Low support, high confidence**
  - 若 `min_sup = 0.1, min_conf = 0.5`，我的輸出結果如下圖，發現與上一題(`min_sup = 0.1, min_conf = 0.1`)相比，這次找到`association rule` 少了許多，因為 `minimum support` 門檻提高，導致所符合資格的 `frequent itemsets` 也少了許多。
    ```

```
                    freqset   support
0                       (0)  0.284284
1                       (1)  0.519853
2                       (2)  0.272606
3                       (3)  0.238238
4                       (4)  0.396730
...                     ...       ...
1485  (14, 48, 18, 19, 30)  0.101768
1486  (34, 14, 48, 18, 19)  0.112779
1487  (34, 14, 18, 26, 30)  0.104771
1488  (32, 34, 14, 18, 30)  0.102769
1489  (34, 14, 48, 18, 30)  0.119453

[1490 rows x 2 columns]

----------association_rule----------
                            rule   support  confidence      lift
0                    {0} -> {1}  0.152486    0.536385  1.031801
1                    {0} -> {6}  0.153153    0.538732  1.021240
2                   {0} -> {14}  0.212546    0.747653  1.043163
3                   {0} -> {16}  0.154154    0.542254  1.056654
4                   {0} -> {18}  0.205539    0.723005  1.025483
...                          ...       ...         ...       ...
3289  {48, 18, 14, 30} -> {34}  0.119453    0.723232  1.128920
3290  {48, 34, 18} -> {14, 30}  0.119453    0.507801  1.205928
3291  {48, 34, 30} -> {18, 14}  0.119453    0.626970  1.189259
3292  {48, 14, 30} -> {34, 18}  0.119453    0.549080  1.177104
3293  {48, 18, 30} -> {34, 14}  0.119453    0.582114  1.225998

[3294 rows x 4 columns]

totle rules: 3294
```

- Any topics you are interested in:

  因為我在輸出的 association rule **中，也有計算出提升度**(lift)，**有發現**Lift(X -> Y) **與** Lift(Y -> X) **的**值是一樣的，例如：

  - min_sup = 0.3 & min_conf = 0.3
  - **擷取部份**內容：

```
50       {48} -> {34}  0.316316    0.660627  1.031198
51       {34} -> {48}  0.316316    0.493750  1.031198
52  {18, 19} -> {14}  0.305639    0.764608  1.066820
53  {18, 14} -> {19}  0.305639    0.579747  1.074522
54  {19, 14} -> {18}  0.305639    0.765246  1.085397
55  {18} -> {19, 14}  0.305639    0.433507  1.085397
56  {19} -> {18, 14}  0.305639    0.566481  1.074522
```