Market Basket Analysis – Performance Assessment

Steven Schindler

Data Mining II – D212

Kamara, Kesselly; PhD

College of I.T., Western Governors University

Table of Contents

A1.

Which items are the most frequently purchased together?

A2.

The goal of this analysis is to use Market Basket Analysis(MBA) to explore the purchasing habits of the customers from the telecommunications company. Also to discover which items are purchased together for the executive leaders.

B1.

Market basket analysis is used to uncover associations between items, it looks for combinations of items that occur frequently together in transactions.(Sharma, A. (2019, May 15).) It analyses the data set by uncovering links between items that are frequently purchased together. It uses association rules such as support, confidence and lift that are used to measure how strong of a combination a pair of items in a transaction will be. The expected outcome is to use MBA on the transaction dataset and get a list of transactions that have high numbers in regard to the association rules. Such as

(HP 61 ink)	(Dust-Off Compressed Gas 2 pack)

Which has a lift of 1.34, a support of .052 and a confidence of .321.

B2.

	A	B
1	Item01	Item02
5957	Apple Pencil	Dust-Off Compressed Gas

One example of a transaction is shown in the picture above an Apple Pencil was purchased with Dust-Off Compressed Gas.

B3.

Market Basket Analysis assumes that customers who buy one item are more likely to buy another similar item or group of items if they are associated together.(Tianhua, S. (n.d.).)

C1.

Using TransactionEncoder to fit and transform the data the new data file is called clean_d212_MBA.csv.

C2.

Using this code to generate the apriori rules with a minimum support of .05:

a_rules = apriori(df_cleaned, min_support = 0.05, use_colnames = True)
a_rules.head()
Then using the head method we get the first five rows:

```
: a_rules = apriori(df_cleaned, min_support = 0.05, use_colnames = True)
  a_rules.head()
```

:

	support	itemsets
0	0.050527	(10ft iPHone Charger Cable 2 Pack)
1	0.068391	(Anker USB C to HDMI Adapter)
2	0.087188	(Apple Lightning to Digital AV Adapter)
3	0.179709	(Apple Pencil)
4	0.132116	(Apple USB-C Charger cable)

C3.

assoc_rules = association_rules(a_rules, metric = 'lift', min_threshold = 1)

Running the above code, we get association rules using lift as a filter with a minimum threshold as 1.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(Apple Pencil)	(Dust-Off Compressed Gas 2 pack)	0.179709	0.238368	0.050927	0.283383	1.188845	0.008090	1.062815	0.193648
1	(Dust-Off Compressed Gas 2 pack)	(Apple Pencil)	0.238368	0.179709	0.050927	0.213647	1.188845	0.008090	1.043158	0.208562
2	(HP 61 ink)	(Dust-Off Compressed Gas 2 pack)	0.163845	0.238368	0.052660	0.321400	1.348332	0.013604	1.122357	0.308965
3	(Dust-Off Compressed Gas 2 pack)	(HP 61 ink)	0.238368	0.163845	0.052660	0.220917	1.348332	0.013604	1.073256	0.339197
4	(Dust-Off Compressed Gas 2 pack)	(VIVO Dual LCD Monitor Desk mount)	0.238368	0.174110	0.059725	0.250559	1.439085	0.018223	1.102008	0.400606
5	(VIVO Dual LCD Monitor Desk mount)	(Dust-Off Compressed Gas 2 pack)	0.174110	0.238368	0.059725	0.343032	1.439085	0.018223	1.159314	0.369437

antecedents	consequents	support	confidence	lift
(Apple Pencil)	(Dust-Off Compressed Gas 2 pack)	0.050927	0.283383	1.188845
(Dust-Off Compressed Gas 2 pack)	(Apple Pencil)	0.050927	0.213647	1.188845
(HP 61 ink)	(Dust-Off Compressed Gas 2 pack)	0.05266	0.3214	1.348332
(Dust-Off Compressed Gas 2 pack)	(HP 61 ink)	0.05266	0.220917	1.348332
(Dust-Off Compressed Gas 2 pack)	(VIVO Dual LCD Monitor Desk mount)	0.059725	0.250559	1.439085
(VIVO Dual LCD Monitor Desk mount)	(Dust-Off Compressed Gas 2 pack)	0.059725	0.343032	1.439085

C4.

1]:
```
a_rules = apriori(df_cleaned, min_support = 0.05, use_colnames = True)
a_rules.nlargest(3, ['support'])
```

1]:

	support	itemsets
6	0.238368	(Dust-Off Compressed Gas 2 pack)
3	0.179709	(Apple Pencil)
24	0.174110	(VIVO Dual LCD Monitor Desk mount)

The top three item sets generated by the apriori algorithm are shown above in the screenshots. Using the item sets generated by the apriori algorithm I get these association rules:

```
In [40]: assoc_rules = association_rules(a_rules, metric = 'lift', min_threshold = 1)
         assoc_rules
```

Out[40]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(Apple Pencil)	(Dust-Off Compressed Gas 2 pack)	0.179709	0.238368	0.050927	0.283383	1.188845	0.008090	1.062815	0.193648
1	(Dust-Off Compressed Gas 2 pack)	(Apple Pencil)	0.238368	0.179709	0.050927	0.213647	1.188845	0.008090	1.043158	0.208562
2	(Dust-Off Compressed Gas 2 pack)	(HP 61 ink)	0.238368	0.163845	0.052660	0.220917	1.348332	0.013604	1.073256	0.339197
3	(HP 61 ink)	(Dust-Off Compressed Gas 2 pack)	0.163845	0.238368	0.052660	0.321400	1.348332	0.013604	1.122357	0.308965
4	(Dust-Off Compressed Gas 2 pack)	(VIVO Dual LCD Monitor Desk mount)	0.238368	0.174110	0.059725	0.250559	1.439085	0.018223	1.102008	0.400606
5	(VIVO Dual LCD Monitor Desk mount)	(Dust-Off Compressed Gas 2 pack)	0.174110	0.238368	0.059725	0.343032	1.439085	0.018223	1.159314	0.369437

These are the top 6 rules with a minimum lift of 1. Narrowing further we get the top three rules with a lift of at least 1.30, confidence at least .25 and support at least .05.

```
In [13]: assoc_rules[ (assoc_rules['lift'] >= 1.30) &
                     (assoc_rules['confidence'] >= .25) & (assoc_rules['support'] >= .05)  ]
```

Out[13]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
3	(HP 61 ink)	(Dust-Off Compressed Gas 2 pack)	0.163845	0.238368	0.052660	0.321400	1.348332	0.013604	1.122357	0.308965
4	(Dust-Off Compressed Gas 2 pack)	(VIVO Dual LCD Monitor Desk mount)	0.238368	0.174110	0.059725	0.250559	1.439085	0.018223	1.102008	0.400606
5	(VIVO Dual LCD Monitor Desk mount)	(Dust-Off Compressed Gas 2 pack)	0.174110	0.238368	0.059725	0.343032	1.439085	0.018223	1.159314	0.369437

This shows that HP 61 ink and Dust-Off Compressed Gas 2 pack are frequently purchased together with a confidence of 32%. Dust-Off Compressed Gas 2 pack and VIVO Dual LCD Monitor Desk mount are purchased together with a confidence of 25%. Finally, when VIVO Dual LCD Monitor Desk mount is purchased Dust-Off Compressed Gas 2 pack will also be purchased together with a confidence of 34%.

D1.

Support is the frequency an itemset occurs, Dust-Off Compressed Gas 2 pack has the highest support of .23 which means of all the transactions it is purchased in 23% of the transactions. Confidence is the probability that items are purchased together. VIVO Dual LCD Monitor Desk mount and Dust-Off Compressed Gas 2 pack have the highest confidence of .34. This means that when the monitor desk is purchased the compressed gas will also be purchased with 34% confidence.

Lift summarizes the strength of association between two items, the larger the lift the greater the link is. Dust-Off Compressed Gas 2 pack and VIVO Dual LCD Monitor Desk mount have a lift of 1.43 which is the highest lift. This means that the presence of the compressed gas increases the probability that the monitor desk will also be purchased. (Select Statistics. (n.d.).)

D2.

The practical significance is that the company now knows which items are frequently purchased together. Such as the compressed gas will often be purchased with the monitor desk mount. Which have the highest metrics of .059 support, confidence of .34, and a lift of 1.43. These two items are the most frequently purchased together.

D3.

Since we know that Dust-Off Compressed Gas 2 pack and the VIVO Dual LCD Monitor Desk mount are the most frequently purchased together, they should be near each other in a physical store. On an online store they should be recommended to the customers, if a customer buys one the other should be recommended. Discounts can also be offered on these two products when purchased together.

References:

Select Statistics. (n.d.). Market Basket Analysis: Understanding Customer Behaviour. Retrieved from https://select-statistics.co.uk/blog/market-basket-analysis-understanding-customer-behaviour/

Sharma, A. (2019, May 15). A Gentle Introduction on Market Basket Analysis & Association Rules. Towards Data Science. Retrieved from https://towardsdatascience.com/a-gentle-introduction-on-market-basket-analysis-association-rules-fa4b986a40ce

Tianhua, S. (n.d.). Market Basket Analysis. Retrieved from https://sarahtianhua.wordpress.com/portfolio/market-basket-analysis/