Matthew Shipley

IEMS 308 – Professor Klabjan

11/2/2016

Assignment 2: Affinity Analysis on Dillard's POS Data

Executive Summary:

By performing an affinity analysis algorithm on the Dillard's POS database, we discovered the 20 SKUs that could benefit most from relocating around the store. Using sales data from Arkansas only (Dillard's headquarters), we plan to make floor plan changes using these 20 SKUs, and will test the effectiveness of any changes before rolling changes out to national locations. The 20 SKUs are:

6062521	6072521	6032521	3988011	2716578	3968011
3898011	3908011	3988011	3998011	2726578	8798636
5528349	4992993	3161221	2783996	3978011	3968011
3690654	4108011				

Problem Statement:

Dillard's is looking to use information about sales of items in their stores to optimize their floor layout. By placing items that are frequently purchased in the same transaction close together, the store can ensure that customers don't miss the second item that they may be looking for. On the contrary, placing these two items far apart in the store can ensure that the customer spends a maximum amount of time in the store and is more likely to make an impulse buy.

Dillard's wants to test the effectiveness of using affinity analysis on floor layout optimization by test the effectiveness of using affinity analysis on floor layout optimization by running through a test trial. Since Dillard's is headquartered in Little Rock, Arkansas and Arkansas is thus one of Dillard's premiere markets, we are to examine purchase data from Dillard's Arkansas stores to find the strongest association rules, and uncover 100 SKUs which are the best candidates for planogram modification.

Assumptions:

- Disregarding any seasonal changes in potential association rules between items
- Ignoring sizes/different variations on the same products
- The current floorplan of Dillard's stores is random; no prior arrangement was done

Methodology:

To extract only sales data from Arkansas stores only, the *STORE* column from *trnsact* was imported into a Pandas dataframe. Using a cross reference to the *strinfo* table (which contains info about each Dillard's store), a new array *stateSelect* was created which contains *TRUE* values only for transactions that occurred in Dillard's Arkansas stores. Using this *stateSelect* array, another array named *stateKeepRows* was created to store an index of the row numbers containing Arkansas stores. Using an inverse of this array named *stateSkipRows*, the transactions were read into trnsactAR. Next, all instances of transactions that were of *STYPE* returns were removed from the dataframe.

To construct the market baskets, the *trnsactAR* array was grouped by the *STORE*, *REGISTER*, *TRANNUM*, and *SALEDATE* columns, as these four keys constitute a whole basket of a transaction. These groupings were looped through, and each SKU that was part of a single basket was added to a list of lists called *baskets*, where each list is a basket. Orange3-Associate was used to run an affinity analysis on this list of lists *baskets*. Using the set of frequent items and item pairings generated, an association rule algorithm using a minimum support of 250 and minimum confidence of .077 was run to find the top 50 association rules. From there, the best 20 association rules were selected as candidates for changing the Dillard's floorplan in Arkansas.

Analysis:

	Left Hand	Right Hand	Support	Confidence
1	6062521	6072521	625	0.7070
2	6072521	6062521	625	0.6779
3	6032521	6072521	549	0.6169
4	6072521	6032521	549	0.5954
5	6062521	6032521	516	0.5837
6	6032521	6062521	516	0.5798
7	3988011	2716578	680	0.2951
8	3968011	3898011	846	0.2902
9	3908011	3988011	528	0.2763
10	3908011	3998011	512	0.2679
11	3908011	2726578	506	0.2648
12	8798636	5528349	645	0.2585
13	2716578	3988011	680	0.2559
14	4992993	3161221	553	0.2509
15	8798636	2783996	610	0.2445
16	3898011	3978011	930	0.2311
17	3988011	3908011	528	0.2292
18	3898011	3968011	846	0.2102
19	3968011	3690654	553	0.1897
20	3690654	3968011	553	0.1864
21	4108011	4628597	2066	0.1862
22	3690654	3898011	550	0.1854
23	3898011	3524026	742	0.1843
24	3998011	2726578	689	0.1785
25	2726578	3998011	689	0.1676

	Left Hand	Right Hand	Support	Confidence
26	6318344	4628597	599	0.1590
27	3524026	4628597	1608	0.1584
28	3978011	3898011	930	0.1497
29	3161221	4628597	945	0.1494
30	3978011	3524026	924	0.1487
31	3978011	4628597	863	0.1389
32	3559555	4628597	671	0.1385
33	3898011	3690654	550	0.1366
34	9277426	4628597	946	0.1334
35	3998011	3908011	512	0.1326
36	803921	4628597	865	0.1311
37	3898011	803921	500	0.1242
38	2726578	3908011	506	0.1231
39	4208011	4628597	755	0.1188
40	3868338	4628597	552	0.1153
41	3559555	803921	542	0.1118
42	2698353	4628597	646	0.1006
43	2783996	8798636	610	0.0998
44	3978011	4108011	598	0.0962
45	3524026	3978011	924	0.0910
46	3161221	4992993	553	0.0874
47	5528349	8798636	645	0.0865
48	3978011	803921	514	0.0827
49	803921	3559555	542	0.0822
50	803921	3978011	514	0.0779

The table above shows the top 50 association rules sorted by confidence. The confidence for each rule was used to select the top 20 SKU's that are candidates for moving in the store. Any duplicate association rules that scored high on confidence both ways (left hand -> right hand and right hand -> left hand) were ignored.

Next Steps:

First, to make these floor arrangements on a national scale, the results of the changes made in Arkansas would need to be measured and analyzed for their effectiveness. Furthermore, there are several ways that we could make this affinity analysis model more accurate. First, putting SKUs into "buckets" could eliminate singling out items that go together and are already placed next to each other on retail floors. One example of this is bed sheets, where the pillow cases will already be next to the matching sheets on the floor. Bucketing could also eliminate problems with the size and colors of clothing. Rather than finding an association rules with a specific size and color of a certain item, it could be more beneficial to group these into one "item" and find larger trends in association rules. Finally, considering seasonality and time purchased could help create a more accurate model, as some association rules may be particularly strong during certain times of the month or year (i.e. the holiday season.