New Recommendation Algorithm - Apriori

Nick Knauer[1]

Abstract

Determining which advertisers to place in front of placements has been a time consuming task for the Ad Operations team. A great amount of searching through the portal, comparison of performances among different partners, and checking CPM restrictions are tasks that are required currently for demand alignment. A new algorithm is introduced called Apriori which can help solve a majority of this issue.

Text based on Optimatic Team insights, see www.publishers.optimatic.com

Introduction

Objective. The Ad Operations team has spent a great amount of time re-ranking the AGORA waterfall and conducting demand alignment by hand. A solution has been created that will aid the Ad Operations team to better the current optimization daily tasks.

For demand alignment, if the Ad Operations team is seeing poor performance on some placements, then the search for appropriate demand takes place. This can be done either by: 1) Finding high filling tags 2) Finding tags with decent load times 3) Finding tags that meet the CPM requirement

One solution to this, excluding load time, is an algorithm called Apriori.

What is Apriori?. Wikipedia definition:

Apriori is an algorithm for frequent item set mining and association rule learning over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database: this has applications in domains such as market basket analysis.

In our case, the placements represent the items. We can link the association rules of the placements back to the advertiser to create a recommendation system per advertiser.

Methodology

The previous day's data for this analysis is from the following path: www.publishers.optimatic.com -> TOOLS -> MISC -> PLACEMENT BY AD SOURCE DATASET

Using this dataset, we can transform it so that it becomes a bucket of all the placements each advertiser bought on with a fill rate of 5% or higher.

Below is the original dataset in which the first column is the advertiser and the 2nd column in the placement:

Email address: `nknauer@optimatic.com` (Nick Knauer)

Advertiser Name	Ad Name
A-[redacted] ML ENG WL T14570 LKQD	[redacted]-009- Desktop - RS - WL
A-[redacted] PC-PR-DU-EXP-WW-RON-$6-LQ	[redacted]
A-[redacted] PC-PR-DU-WW-RON-$6-STREAM	[redacted]
A-[redacted] Media LP RON 60270	[redacted]
A-[redacted] WL US RTB	[redacted] -009- Desktop - RS - WL
A-[redacted] Collider Backfill	[redacted] -009- Desktop - RS - WL

Below is the transformed matrix in which each row represents an advertiser and every column after that shows the placements (items) they filled 5% or more on.

Advertiser Name	Col1	Col2
A-[redacted] Exponential (T)	[redacted]_$5_US_PC_VPAID_SP_InDirect_VP	OP-1-2918-[redacted]m–EN-$3
A-[redacted] CN 700 WL	[redacted] APAC HK	[redacted]_$3.5_US_PC_VPA
A-[redacted] ES GEOS WL 10	Sh[redacted] VM $5.2 ($4)	R[redacted] VM-$9 ($6.75)
A-[redacted] GCC Geos $10	[redacted]_Internal_7 USD	S[redacted] VM $8.71 ($6.7)
A-[redacted] WL AU	[redacted]_Desktop_WW_Mix_5.5	[redacted]_Desktop_WW_Mix_5
A-[redacted] WL SG MY	[redacted] APAC SG	

Summary of Dataset

Converting the dataframe to this format allows us to use the *arules* package in R and obtain interesting stats about the data.

We can see a few things by seeing the summary of the dataset:

Density Value of 0.03281992 (3.2%): refers to the proportion of nonzero matrix cells.

Since there are (55 x 839) = 46,145 positions in the matrix, we can calculate that a total of 46,145 x 0.03187778 = 1,471 placements were purchased with a 5% or more fill rate during the previous day for AGORA (ignoring the fact that duplicates of the same placements might have been purchased.)

With an additional step, we can determine that the average transaction contained 1,471 / 55 = 26.7 distinct placements.

You can also see the most frequent placements bought with a 5% or more fill rate. Mobiletech was the most frequent bought placement in AGORA at a fill rate of 5% or more. It had a fill rate greater than 5% for yesterday for 15 advertisers. Since 15/55 = 0.2727, we can also determine that [redacted] appeared in 27.3% of the transactions. [redacted] also contribute a good bit to the dataset.

Summary of dataset:

```
## transactions as itemMatrix in sparse format with
##  55 rows (elements/itemsets/transactions) and
##  839 columns (items) and a density of 0.03187778
##
## most frequent items:
##         [redacted]          [redacted]
##                15                  14                       14
##          [redacted]           [redacted]               [redacted]
##                14                  13                     1401
##
## element (itemset/transaction) length distribution:
```

```
## sizes
##   3   4   5   6   7   8   9  10  11  12  13  14  16  18  19  20  21  22
##   6   2   5   2   2   4   4   2   1   3   1   1   1   1   1   1   1   1
##  23  26  29  31  34  35  42  47  89  93 120 366
##   2   1   1   2   2   1   1   1   1   2   1   1
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    3.00    6.00   11.00   26.75   24.50  366.00
##
## includes extended item information - examples:
##   labels
## 1      1
## 2  1-1-A
## 3  1-2-A
```

Associations Among Placements

Going further, we can determine the associations with this. Looking at this table below, we can read in plain language as:

"If an advertiser fills 5% or more on [redacted]Optimatic-DT-WW-LP-$2.5, they will also fill 5% or more on [redacted]$2 Small 08-02-17".

With support of 0.0909 and a confidence of 1 we can determine that this rule covers 9% of the transactions and is correct in 100% percent of purchases involving [redacted]Optimatic-DT-WW-LP-$2.5.

On top of this, there is also a calculation called "lift". Lift implies, that advertisers who fill 5% or more on [redacted]Optimatic-DT-WW-LP-$2.5, are nearly **11x more likely** to buy [redacted]$2 Small 08-02-17 than the typical advertiser.

rules	support
{[redacted]_DT_RS_OPT_20Oct_S/M/L_RON_US} => {Adzos floor 2}	0.0909091
{[redacted] floor 2} => {[redacted]_DT_RS_OPT_20Oct_S/M/L_RON_US}	0.0909091
{[redacted]TV_$3(Feb)_US_PC_VPAID_LP_Direct _VP} => {[redacted]TV-Desktop-$2.5-US-L-InHouse-Optimatic}	0.0909091
{[redacted]TV-Desktop-$2.5-US-L-InHouse-Optimatic} => {[redacted].TV_$3(Feb)_US_PC_VPAID_LP_Direct _VP}	0.0909091
{T2288_Media Group- V_AP_RON_RS_} => {[redacted]_G1_RON_Player}	0.0909091
{[redacted]_G1_RON_Player} => {T2288_Media Group- V_AP_RON_RS_}	0.0909091
{vcientertainment} => {autoline.tv}	0.1090909
{autoline.tv} => {vcientertainment}	0.1090909
{vcientertainment} => {mobiletech}	0.1090909
{mobiletech} => {vcientertainment}	0.1090909
{vcientertainment} => {SDN_3$_dsk}	0.0909091
{SDN_3$_dsk} => {vcientertainment}	0.0909091
{vcientertainment} => {giletravel}	0.0909091
{[redacted]} => {vcientertainment}	0.0909091
{vcientertainment} => {vzcentral}	0.0909091

Filter out Specific Placements

Along with seeing the entire dataset, you can also filter out placements to see their associations. Below is a subset of 5 rules for the [redacted]G1 RON Player Placement:

```
##     lhs                                  rhs                                  support confidence     l
## [1] {T2288_Media Group- V_AP_RON_RS_} => {[redacted]_G1_RON_Player}           0.09090909  1.0000000 3.9285
## [2] {[redacted]1_RON_Player}           => {T2288_Media Group- V_AP_RON_RS_} 0.09090909  0.3571429 3.9285
## [3] {[redacted]floor 2}                => {[redacted]_G1_RON_Player}        0.09090909  0.7142857 2.8061
## [4] {[redacted]G1_RON_Player}          => {[redacted]floor 2}               0.09090909  0.3571429 2.8061
## [5] {SPR-ALL}                          => {[redacted]_G1_RON_Player}        0.09090909  1.0000000 3.9285
```

Recommendations

From here, we can examine the associations of the rules. Even better, we can link this back to the advertiser. Examining the association rules, we can see if all the placements from the lhs side were bought. If they were, we can then check if the rhs side was also bought. If the rhs side was not bought, then that would be considered a recommendation because they have not bought it yet and it could open up the door for more inventory to that advertiser.

Recommendations: (Excluding 406 other recommendations)

Advertiser.Name	Publisher	Recommendations	Target CPM	LIFT
A-ADO Exponential (T)	(SS) [redacted]	OP-1-3298-F[redacted]ML-EN-$2.1	[redacted]	7.638889
A-ADO Exponential (T)	(SS) [redacted]	G[redacted]- Small B - Feb -2	[redacted]	6.875000
A-ADO Exponential (T)	(SS) [redacted]	OP-1-3295-F[redacted]ML-EN-$1.5	[redacted]	6.547619
A-ADO Exponential (T)	(SS) [redacted]	1-4-A	[redacted]	5.729167
A-ADO Exponential (T)	(SS) [redacted]	[redacted]Desktop-$2.5-EN-S-Optimatic-	[redacted]	5.729167
A-ADO Exponential (T)	(SS) [redacted]	HEX - 008 - Desktop - RS - New2	[redacted]	5.729167
A-ADO Exponential (T)	(SS) [redacted]	OP-1-3299-F[redacted]-ML-EN-$2.3	[redacted]	5.729167
A-ADO Exponential (T)	(SS) [redacted]	OP-1-3300-F[redacted]ML-EN-$2.5	[redacted]	5.729167
A-ADO Exponential (T)	(SS) [redacted]	H[redacted]_Small_3 USD	[redacted]	5.092593
A-ADO Exponential (T)	(SS) [redacted]	GLO-Desktop-RS60-US-ML-Other	[redacted]	4.910714
A-ADO Exponential (T)	(SS) [redacted]	Media-5	[redacted]	4.365079
A-ADO Exponential (T)	(SS) [redacted]	[redacted]_G1_RON_Player	[redacted]	2.455357
A-ALTI RON 600	(SS) [redacted]	1-2-A	[redacted]	9.166667
A-ALTI RON 600	(SS) [redacted]	F[redacted]-Optimatic-DT-WW-LP-$2.5	[redacted]	9.166667
A-ALTI RON 600	(SS) [redacted]	F[redacted] $2 Small 08-02-17	[redacted]	9.166667
A-ALTI RON 600	(SS) [redacted]	G[redacted] - Small C - Feb -2	[redacted]	9.166667
A-ALTI RON 600	(SS) [redacted]	OP-1-2899-F[redacted]SP-EN-$2.25	[redacted]	9.166667
A-ALTI RON 600	(SS) [redacted]	T1349_[redacted]_LP_EN_3	[redacted]	9.166667
A-ALTI RON 600	(SS) [redacted]	MK_[redacted]_Q2_5_REV70	[redacted]	6.875000
A-ALTI RON 600	(SS) [redacted]	M[redacted]_Desktop_2.5_SP_WW_al	[redacted]	4.910714

Conclusion

The apriori algorithm has the capability of providing great insights into our marketplace. This example was simply just for AGORA demand.

Some applications this could be used for:

1) Provide as a service for SSP partners with their own demand.
2) Use in production as a recommendation system. It is easily scalable. This example took less than 30 seconds to run.
3) A hands-on tool for ad ops when it is not in production to see all possible advertisers not tested yet for their placements.

4) A clustering tool to see how placements are related to each other. For example, this algorithm helped Walmart realize that beer and diapers should be sold together especialy on Fridays. They put the items closer together in the isles and their revenue boosted.
5) Used to package up Deal Ids.

There are a number of uses for this and this is just the start of how machine learning can be helpful for Optimatic Media.