

Homework #9 ISYE6501

Retailer Case Study

Zach Olivier

7/16/2018

Summary:

My analytics solution for the Retailer follows this general format:

Given: market basket data - sales, output from collaborative filtering model, and margins

Use: Use clustering algorithm

To: determine a ranked set of high value pairs of store items (high revenue, high sales, high correlation)

Given: ranks of high value pairs, and subset of within store and across store variable self positions

Use: Greedy Multi-Armed Bandit experiment design (Bayesian Hypothesis Testing)

To: Figure out which pairs of items benefit the most from more space and in which stores

Recommendation:

My recommendation unfolds in two steps:

1. Determine the pairs of items that will benefit the retailer the most if space is optimized
2. Run a greedy experiment within stores and across stores to provide the most value while gathering information to recommend the best space requirements

My hypothesis is that there are too many pairs of products to control for to run tests and optimize the storage space. Further, I imagine that some products will benefit more for more space and placement closer to related items than others. Depending on the location of the store, different products might have benefit more from more space in one store than another. To tackle these problems I would set of a greedy experiment to provide the retailer the most value while we continue to learn about which products in which stores benefit the most from increased self space. If we work with a subset of certain stores and certain high value products I believe we have a better chance at quickly getting results. For example - investment and resources can be controlled easier, more stores would be able to participate, and we would be able to provide store specific recommendations. In a perfect scenario - we could find a set of products that benefitted from increased space and complementary placement in one store, and shift to exploitation of this relationship for that store only.

Below are the components of my analytics solution - with notes aiming to explain the thought process of my choices.

Data / Collaborative Filtering Model:

To determine the highest value set of pairs of products, we need sales data, data on which products are normally purchased together, and the margins of each product. Based on the market basket data - or customer data of products they purchased in their basket - we can run a quick collaborative filtering algorithm to determine which pairs of products are truly complementary. Products with the highest cosine distance in terms of market basket will be products that are purchased together the most.

Running a simple clustering algorithm on this data should give us clusters of products that sell well, have complimentary products, and provide revenue to the retailer. Having this data will allow us to make recommendations on the most relevant items, and avoid optimizing marginal products that do not improve the bottom line. Further this will allow us to narrow our experiment down to a smaller set of products which will reduce the resources and cost of having to move the products around to support the experiment. To me this is the most reasonable attempt at getting insights quickly without having to wait for more data to come in (retailer should have this data) or get extremely complex in our experiment structure.

Clustering:

A simple unsupervised clustering algorithm should give us a set of products to include in our experiment. We would select the "highest value" cluster of products and recommend we try to answer the space and placement questions with these products - finding positive results will allow us to switch to exploitation on products that will have a large effect on the business. The result of this step is a subset of products and their compliments that we can feasibly run a within store and across stores experiment with. To further ground this project in reality, we can manually inspect the top pairs to ensure space and placement adjustments can be done quickly or without high cost.

Experiment Design:

Designing a robust experiment within each store, across stores, and within the pairs of highest value products will be the engine of this analytical solution. The first step would be to work with the retailer to develop a list of possible variables that can be altered. What is the measurement for increasing space? Where should complimentary products be placed? These questions require domain knowledge to answer, but the end result should be a set of possible combinations to try. By setting up a factorial

design we can test the best subset of combinations of space and placement of our high value products in a select subset of participating stores.

Based on the timing constraints, we can leave certain interactions of the experiment up for days, weeks or months at a time. Based on which placements in which stores perform well, we can adjust our spacing for exploitation or hold to keep learning on certain combinations. This is all well if we have a good measure of performance. A retailer store is not a webpage where many experiments can be run quickly and easily. Based on our timing, we may only be able to test certain combinations of space and placement for a week. These small increments of time may not be enough to judge if a product placement iteration is truly working or if randomness has boosted sales.

Depending on timing - with a longer test period - a change detection model could be wrapped around each product and each store. This will show us if Product A at Store A with a unique iteration of our placement and space experiment design change significantly over Product A's baseline at Store A. If we had less time to compare experiments, I might recommend a non-parametric or Bayesian hypothesis test (does well with small data - we would know the posterior distribution of each product) to drive our greedy decisions. The test would measure: did the change of placement and / or self space change this product's sales versus the baseline - or if the change more likely due to chance (based on the past distribution of that product). These tests would have to be performed within the store's unique placement and across the all store's unique placement.

Finally we need a mechanism or a team of onboard store managers to move the products uniformly across the stores, based on our experiment's results. I believe with limited products to optimize will help minimize the work this experiment will be on the store managers. Especially if we can provide specific instructions on the space and placement quickly.

Results:

Ultimately the results of my analytic solutions would aim to drive more revenue while "working" to answer the space and placement questions the retailer has. If particular iteration of space and placement may work well in Store A, then keep using that placement for those products. That same placement might not work for Store B, then keep using Store B to test our other iterations of placement and space.

We should be able to answer the question about more space and complimentary placement generally being better for all stores while taking advantage of the the stores where there is a significant effect. At

the end of our timeline, we should have data to support an answer to the question, and a measurement of how much revenue we generated using the greedy experiment. The infrastructure and process should be set up at the end of the timeline for the retailer to continue to run these studies for any number or iteration of products that they have the resources for. I believe this analytical solution would be easy to sell to the team - let's gather our data to answer these overarching business questions which driving the most value possible.