Homework 9

Questions 19.1

Describe analytics models and data that could be used to make good recommendations to the retailer. How much shelf space should the company have, to maximize their sales or their profit? Of course, there are some restrictions – for each product type, the retailer imposed a minimum amount of shelf space required, and a maximum amount that can be devoted; and of course, the physical size of each store means there's a total amount of shelf space that has to be used. But the key is the division of that shelf space among the product types.

For the purposes of this case, I want you to ignore other factors – for example, don't worry about promotions for certain products, and don't consider the fact that some companies pay stores to get more shelf space. Just think about the basic question asked by the retailer, and how you could use analytics to address it.

As part of your answer, I'd like you to think about how to measure the effects. How will you estimate the extra sales the company might get with different amounts of shelf space – and, for that matter, how will you determine whether the effect really exists at all? Maybe the retailer's hypotheses are not all true – can you use analytics to check?

Think about the problem and your approach. Then talk about it with other learners, and share and combine your ideas. And then, put your approaches up on the discussion forum, and give feedback and suggestions to each other.

You can use the {given, use, to} format to guide the discussions: Given {data}, use {model} to {result}. One of the key issues in this case will be data – in this case, thinking about the data might be harder than thinking about the models.

Student Answer:

1. Remove trend and seasonality from data
2. Identify Complimentary Products
3. Optimize shelf space
4. Validate experiment results

Given: Past sales data

Use: Exponential Smoothing

To: Remove trend and seasonality from the sales data.

All our further analysis will be based on this sales data and it's important that our proposed solution is applicable all year round. By removing the effects of trend and seasonality on sales data, we remove products that do well seasonally but do not require shelf space throughout the year (for example, halloween or christmas decorations). These seasonal items that sell high volumes during certain times of the year can skew the data in terms of actual shelf space required for non-seasonal items. Thus with exponential smoothing, we will be able to get data that's more representative of the sales on the average day of the year and not place greater weights on seasonal items which will lead to a more efficient optimization model later.

Given:

- Smoothed sales data from exponential smoothing

Use: Association rule mining algorithm, specifically the FP Growth algorithm

To: Identify complimentary products that retail better together than alone.

This analysis is known as market basket analysis. By identifying products that sell well together, we can place them close together on the shelf to influence customers to purchase more products.

We will use the FP Growth algorithm which is one of the algorithms used for market basket analysis. It uses a tree structure to find frequent sets of items (a set of items is termed an "itemset") that appear in transactions and identifies association rules between these items.

Every node in the tree is an item in the dataset, with the most frequently appearing items appearing first. Once the FP tree is constructed, frequent itemsets can be generated by recursively mining the tree.

The FP growth algorithm is better for large datasets compared to the more common association mining algorithm, the apriori algorithm. Depending on the size of the store this may or may not be beneficial, but I prefer to err on the side of caution and choose the algorithm that works on both large and small datasets, instead of the algorithm that only works on small datasets.

Given:

- Complimentary products
- Shelf space occupied by each product
- Shelf space available
- Inventory of all products available

Use: Optimization model

To: Optimize shelf space, determine effects of optimization

We can use an optimization model to determine how much shelf space the retailer should provide to each product to maximize profit.

Some of the contraints we would be working with are

- the minimum shelf space for a product, and
- maximum shelf space available

The results should give us the optimal combinations of products to put on shelves based on shelf size and sales volume to maximize profit.

A possible drawback to this optimization model is that it may not be very likely for the company to keep records on the dimensions of a product. We cannot optimize shelf space if we don't know the dimensions of products. To combat this, we could just measure the dimensions of a fraction of the complimentary products. We can set a threshold for profit and prioritize the products that are predicted to bring a profit above this threshold. We can then run the optimization model on a smaller set of products but with a more limited shelf space to match the smaller set of products. The other non-prioritized products can then be placed where space is available. I believe that we do not need to spend as much resources optimizing lower profit pairs of complimentary products when they can bring in the same amount of profit on their own.

Validation of experiment

Given:

- Data from our optimization model

Use: A/B Testing

To: Determine if the effects we have predicted really exist in reality

We should be able to verify if our solution works with a simple A/B test. Following the analysis, we can collect sales data for a set period of maybe 1 month with no changes, and then another month using the changes suggested from the optimization model.

Ideally we would also want to avoid festive months like december so that seasonality doesn't skew our collected data.

The benefit of A/B testing is that it is simple and doesn't require a lot of logistic on the part of the retailer. For example, having to re-organize the shelves frequently, especially in a bigger store, this is a huge time and labour commitment. Less frequent changes also means less disorientation for the customer (eg, if they walk in next week and whole store has been re-organized).

The downside is that A/B testing is slow compared to something like a multi-armed bandit experiment which is better for maximizing profit and quickly moving to the best solution for the most profit.

We may consider using a simulation for each store, that way we can condense multiple weeks of A/B testing into a few hours of running simulations.