Monetization Case

HW 15 – Kelly “Scott” Sims

Internal Data

For the first part, we want to come up with a model/system that will help ACME company monetize their own data that they collect. That data, in no particular order, is the following:

- title
- first name
- middle initial
- last name
- credit card type
- credit card number
- list of products purchased in the past, with date of purchase and ship-to address
- which web pages the person looked at
- how long the person spent on each page
- what the person clicked on each page
- estimate of how long the user's eyes spent on each page viewed (for customers where the software was able to take over the device's camera)

There's a lot of different routes to go with all these potential features. But I think a recommender system based off of a Markov Matrix is most prudent here. Now, to calculate the state change probabilities across a whole customer based might be a little over zealous. So to better hone in on trends in our customer based, we can cluster them first based off of purchased items categories. Since we know what items were purchased, we can group these items into a higher order categories. E.g. computers, tvs, speakers would be grouped under the technology category. Couches, tables, rugs, etc. would fall under the furniture category. So on and so forth. Undoubtedly, customers could belong to more than one category. So we order those by number of items purchased within that category. For example:

Customer A Purchases: Laptop, batteries, monitor, xbox, shirts, shoes, pots & pans

Customer A Classification Rankings: Technology, Apparel, Home Goods

Customer B Purchases: Lawn mower, garden hose, rake, TV, Playstation, Earrings

Customer B Classification Rankings: Lawn & Garden, Technology, Jewelry

A customer could be a member of a lot of different classifications, therefore, we take the top 3 categories they belong to. We then cluster all customers together belonging to the same top 3 categories. We now have distinct clusters of customers and their purchasing trends. So if you are a categorized as person who loves tech, lawn & garden, and home goods in that order, we can analyze the items purchased by other customers categorized the same way and recommend something you might like based off of what they purchased. This is where the Markov Matrix comes in. It would be the probabilities of buying items B, C, D, etc. AFTER buying item A. So for a customer who has bought item A and hasn't bought item B,C,D, etc, we can recommend the one that has the highest state change probability.

GIVEN:

- List of products purchased in the past

USE:

- Clustering

RESULTS:

- In Unique clusters of customers with particular shopping habits

Then

GIVEN:

- Each particular cluster

USE:

- Markov Matrices in each cluster to build the state change probability matrices for each item in that cluster

RESULTS:

- A recommender system

Combined Datasets

There is room for improvement with our first model. We can recommend an item to a customer based off of purchasing habits of other "like" customers. But in order to maximize the probability of that customer buying ANYTHING, what if we could show them the recommended item and items just like it but at a different price point. For customers who earn more, or have great credit history, they might be more likely to buy the recommended item's more expensive counter part that has all the bells and whistles. This would increase revenue. Conversely, we could show a cheaper version of the recommended item for customers who don't earn as much or with bad credit. But we need to determine what salary ranges are those that represents a customers who would be more likely buy the more expensive version versus those who would buy the mid tier, versus those who would buy the cheapest option. For this, we could run various classification models. After determining the recommended item for a customer, before showing them that recommendation, a more expensive version and a less expensive version is selected. These 3 items are now the classes we are trying to predict on. From data set #1 and #2, we can use a customer's background as features to make this prediction. These are features will be fed into a "majority vote" ensemble model. In this model, the voters are

- Logistic Regression
- Random Forest
- Naive Bayes
- SVM
- k-Means Clustering

Each model makes a prediction of which item to show the customer. The item that gets the most votes from the voters, wins.

Therefore:

GIVEN:

- Marital status
- Number of Children
- Current city
- Financial net worth
- whether they've ever owned real estate
- list of monthly payment status'
- in default or not

USE:

- Majority Vote Ensemble method

RESULTS:

- Tier of recommended item (most expensive, mid tier, cheapest)