

What's in your Wallet? *Lots of receipts.*

Marcos Carzolio, Andy Hoegh, Xinran Hu, Lucas Robers, Yuhyun Song

Department of Statistics, Virginia Tech

November 3, 2013

Abstract

To predict consumers' future behavior, we extract patterns in historical transaction data. Shopper specific variables including total number of transactions, number of merchants visited, and the number of zip codes visited are used to segment customers into classes. Merchant specific variables are also created, such as whether a customer regularly shops in a zip code where the merchant is located and frequency of shopping at the target merchant. The random forest algorithm is used to estimate the probability of a customer visiting any of the target merchants. Invoking statistical decision theory, an optimal decision is made regarding whether coupons should be issued and if so to which merchants.

1 Overview of Solution

Predicting human behavior is notoriously challenging. However, in this “big data” era massive amounts of information available through data sources such as social media and transaction histories make this problem more manageable. Furthermore, intelligent and responsible use of big data can benefit all involved parties. In particular within the context of predicting consumer transactions, customers can be provided with coupons for discounts which also provides additional business for merchants and instills customer loyalty for the credit card company. This process supports Capital One’s mission of *Making Banking Better for Good*.

Our predictive modeling scheme can be described in a series of four steps: feature extraction, customer segmentation, model building, and model use. An overview of each is provided directly below.

Feature Extraction

Raw transaction data does not lend itself directly to predictive modeling. This is a common trait of massive, and often unstructured data which require preprocessing and feature extraction to retain signals and filters noise present in the information. So, in addition to past history at a particular merchant, we also look at customer specific information such as transaction count by industry type, transaction frequency, number of zip codes shopped in and customer-merchant specific variables such as whether the customer shops in a zip code where the merchant is present. Features are calculated over a twelve month period from July 2010 - June 2011. The response is denoted as a positive outcome if a shopper visits a merchant over the three month period July 2011 - September 2011. The complete list of features is more than 270.

Customer Segmentation:

In an ideal scenario the target customers would be drawn the same population; hence, allowing a single modeling framework to be applied to the entire set of customers. This is clearly not the case with the data set presented in this competition. The transaction records indicated heterogenous segments of customers. Hence, a model based clustering approach was used to segment the customers, then a separate model was applied to each segment. This makes intuitive sense and results in more accurate predictions.

Model Building:

For each segmentation of customers, a collection of well established classification tools were used to evaluate performance. Given that our feature set contained so many variables we favored methods that encourage some sparsity in the covariate space, with the random forest algorithm ultimately being selected. It should be noted that these models are computing the probability of shopping at a merchant in a univariate manner. This has the built in assumption that the probabilities of visiting separate merchants are independent. To

lessen the correlation structure, indicator variables for shopping at other target merchants are introduced.

Model Use:

Given that the goal of the project is to make a decision on issuing coupons for each customer, statistical decision theory proves useful when determining the optimal decision for each customer. The crux of the question is the expected return for issuing coupons to a customer. The mathematical details are provided later on, but five coupons are issued if the expected ‘gain’ is greater than the \$1 penalty for not issuing a coupon.

The following sections will provide additional details on each of the four components that go into our modeling framework. Then a discussion section will follow identifying some closing thoughts and touching on some ideas for future work in this area.

2 Data Preparation

There are two main components in the data preparation stage, feature extraction and customer segmentation.

2.1 Feature Extraction

Meaningful features are extracted from the transaction level data. The key is reducing the data while obtaining information relevant to where a customer may shop in the future. All variables are constructed on the twelve month period, with the exception of small frequency customers. These customers have variables constructed on only nine months as three are held out of the validation set for model fitting. A list of customer specific variables are shown below.

- proportion of purchases online
- transaction count by industry (for all purchases and online purchases)
- total transaction count and transaction count for last 3 months
- number of zip codes for which purchases are made
- number of merchants customer shopped at

The transaction counts by industry encompass nearly one hundred variables corresponding to the range of industries. The customer specific variables are also used in the cluster segmentation scheme. The customer– merchant specific variables are as follows:

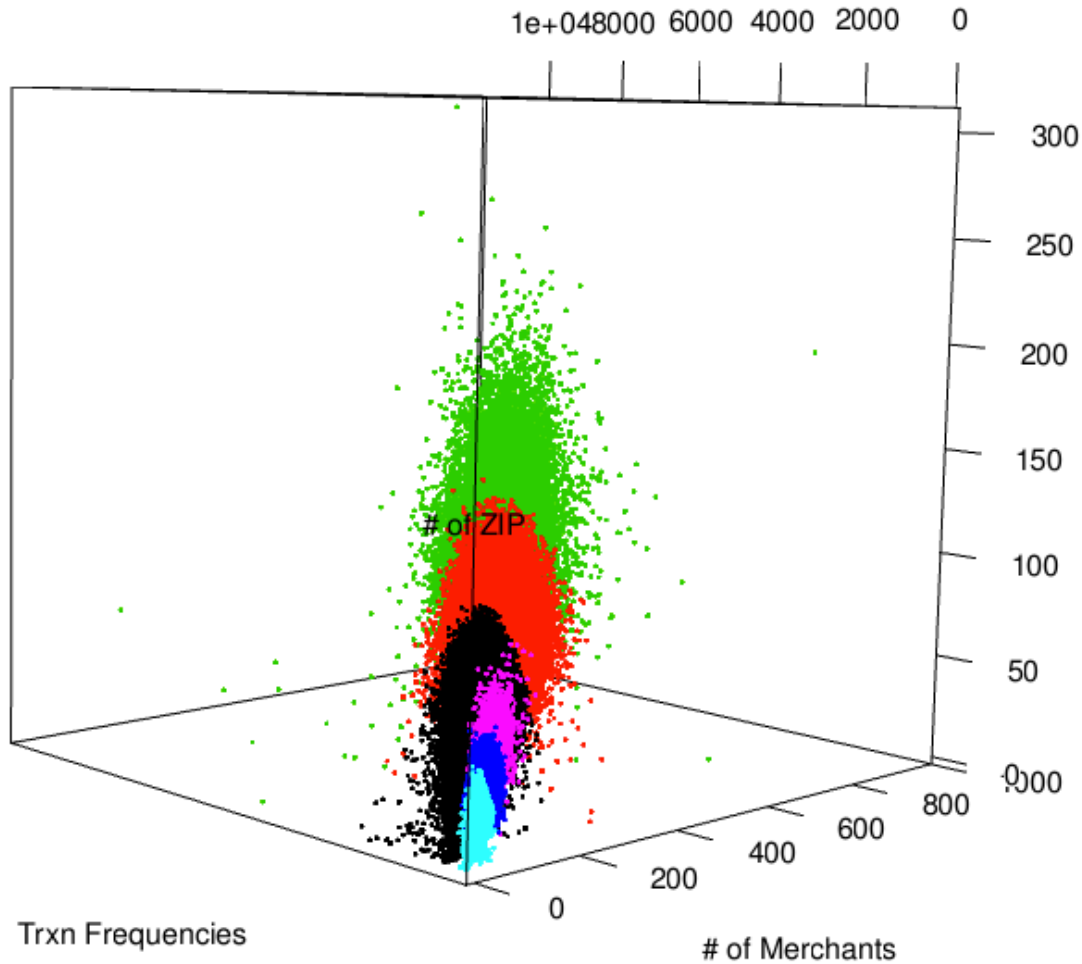
- frequency of shopping at merchant
- merchant located in customer’s home zipcode
- merchant located in zip code customer regularly shops at

- customer shop at competitor (same industry)
- customer shop at competitor in same zip code

2.2 Customer Segmentation

There are two reasons clustering is both useful and necessary in this scenario. First, there is heterogeneity in the customers - suggesting separate models for intuitive and predictive advantages. The second is the inclusion of low frequency customers in the validation set. Fitting a model on customers with high credit card use and then applying it to these low use customers would be a classic interpolation error.

To differentiate heterogenous groups, K-means clustering is applied to three constructed variables: total transactions, number of merchants, number of zip codes shopped in. The goal is to identify homogenous groups of shoppers so that their behavior can be modeled in the same framework. Figure 2.2 provides a visual of the constructed clusters. These



clusters break apart customers not only by total transactions, but also take into account

the diversity of stores and locales a credit card is used in. Summaries of the cluster characteristics can be seen in Table 1.

Table 1: Shopper Segmentations from K means clustering

Cluster	Cluster Color	# Customers	Avg. Transaction Freq	Avg. # of ZIP	Avg. # of Merchant
1	green	152	1337	161	131
2	red	845	824	115	93
3	black	2264	510	67	58
4	pink	1894	282	60	49
5	blue	3907	175	39	34
6	sky blue	4538	121	23	21

The second reason for clustering is the inclusion of small frequency customers with less than 360 annual transactions in the validation set. These customers clearly exhibit different spending habits than the high frequency customers that are provided in the sample test sets. This necessitated a bit of creative thinking to develop a different type of model for these customers. Three options were considered: use the models constructed on high frequency customers, do not issue coupons to these customers, or use the 12 months of data to learn about these customers. Given that options one and two were unsatisfactory, we choose option three. While not a perfect scenario, it was the best option. The idea is to use the twelve months of data to learn how these customers behave. The choices were either use data from July - March to predict April - June *or* October - June to predict back in time for July - September. With additional time, we would like to combine these two models in an intelligent way, but given the time constraints we choose to predict for the same time frame as the required output (July - September). Our thought was that the seasonal patterns would be strong enough that it would be beneficial to fit the model during the same months.

3 Model Building

The random forest algorithm is an ensemble classifier that builds a series of decision trees which are combined to create a stronger classifier. Each tree is constructed on a randomly selected subset of data and predictors. In many problems of this type, the goal is accuracy in the classification itself, but rather the importance lies in accurately estimating the class probabilities. Class probabilities of a random forest algorithm are obtained as the proportion of trees that make a specification classification. Part of the value in this modeling approach is random forest does variable selection itself - less influential features are not selected to construct trees. Additionally, the additive nature of the trees doesn't impose structural assumptions on the relationships between variables.

For the low frequency customers, a model is fit using the the nine months of data immediately following September 30th, to predict use during July - September. This is backwards, in some sense, as we use transactions from after our target period; however, this was favorable to using a model based strictly on high use customers. Then the predictions are made for July-September of the following year.

As mentioned earlier, these models assume independence in the probabilities of shopping at merchants. In reality, there is likely some correlation. For computational and actual time constraints we estimated the probabilities separately, but included variables denoting whether a customer shopped at another target merchant in our model. These will adjust the probabilities accordingly. A more complete solution could be achieved through a multivariate probit model or using a copula.

Our evaluations (on high frequency customers) showed good model results. There would be several valid criteria to look at, but we considered the expected return and its association with the realized return. This is conducted by removing sets of customers the model building step. Then predictions can be made and compared to actual results. For several runs, the expected and observed returns were very close to each - suggesting the probabilities computed with the random forest were valid.

4 Model Use

Using statistical decision theory with the specified loss function an optimal strategy can be obtained for each customer. In particular, given the probabilities of a customer shopping at a merchant and the specified loss function:

$$L(\delta, C_i) = \begin{cases} -3 & \text{offer suite issued with no redemption} \\ 5n & \text{offer suite issued with } n \text{ coupons redeemed} \\ -1 & \text{no offer suite extended} \end{cases}$$

an optimal decision can be made to maximize the expected return. That is, for each customer we seek to maximize the expected return:

$$E(\text{return}_{\text{issue}}) = \$25P_5 + \$20P_4 + \$15P_3 + \$10P_2 + \$5P_1 - \$3P_0, \quad (1)$$

where P_i is the probability of i coupons being redeemed. If $E(\text{return}_{\text{issue}}) > E(\text{return}_{\text{notissue}}) = -\1 then coupons are issued. In this case, the decision ultimately comes down to whether five coupons or zero coupons should be issued.

5 Discussion

One subtle, but important aspect about this competition is that the models are designed to predict where a customer will shop in upcoming months – without any influence from an issued coupon. As the problem statement outlines, merchants are not as interested in providing discounts to frequent customers, but would like to identify new sets of customers. Even though the target is customer behavior without a coupon, our framework includes features that to identify new customers. One in particular is a variable identifying a merchant’s competitor’s customers - that is customers shopping at the same industry in the same zip code. The predictive power of this variable is less than past history at the merchant; however, in reality issuing a coupon would cause a behavioral adjustment such that a customer may visit a merchant they ordinarily would not. An interesting

evolution that would be particularly interesting for Capital One's business partners would be to conduct an experiment by issuing coupons to customers in order to model behavioral change. Our model could easily be extended to estimate the difference in probability of a customer (or more likely a segmentation of customers) visiting a merchant after being issued a coupon. This would be more valuable to a merchant, as issuing coupons to regular customers might actually be a net loss for the company - assuming discount on goods that would have been purchased anyway.