

Predicting Dissatisfied Credit Card Customers

Zachary Arens

4/25/2001

INFT 979
Dr. Wegman

Introduction

Customer satisfaction is essential to retaining customers and maintaining profitability of a firm. A 1999 J.D. Power and Associates study of credit cardholders indicated that customer service is the primary driver of satisfaction. Factors including payment processing, call center service, and issuer reputation were among the most important (Cards International). Although these factors do play a significant role in satisfaction, another 1998 cardholder study indicated “customer data is crucial in cardholder satisfaction strategy” (“Broad Approach”). Using a method that relies upon

the integration of internal and external customer data, one can create models to predict behavior. Ultimately, one could predict the likelihood of dissatisfaction for each customer and alter the company-customer relationship in an attempt to increase retention, usage, and therefore profitability.

The data for this analysis was collected from two sources. The first set was comprised of data from a survey of Capital One credit card customers. The variables from the survey included topics such as customer purchasing habits, employment status, and non-Capital One credit card usage, as well as others. The internal data was drawn from Capital One's consumer database and included data on each customer's balance, APR, credit limit, credit worthiness, cash advances, and others. Each customer record also included a binary rating indicating satisfaction or dissatisfaction with their Capital One credit card.

The total dataset included 22,242 records and 25 variables. A full list of all the variables can be found in appendix F. The variables were a mix of scale and ordinal data. Depending on the format, this dataset has roughly 5 MB of data, so according to the Huber taxonomy of datasets, this dataset is in the large category.

Cleaning

Before any model could be built, the data had to be cleaned. Throughout the data, there were missing values. The variables with the most occurrences of missing values were v11 (average daily balance), v4 (over limit in the past 30 days?), and v5 (balance on non-Capital One cards). These variables were missing 210, 198, and 171 values respectively. Out of 22,242 cases, 7.87% had one or more missing values.

Many statistical packages such as SPSS, which was used in this analysis, have built-in features, which allow missing values to be transformed into analyzable numbers. The methods often use an average or a linear trend. This process is appropriate for the scale data, and was utilized in such cases. For the ordinal variables, the cases with missing variables were deleted. This dataset is large enough to delete the cases with missing values without affecting the results. The only concern with deleting 1,477 records is that they may contain significant output. Perhaps a high proportion of respondents who were apathetic completing the survey were also dissatisfied with their

cards. Among the cases with missing variables, 15.1% were dissatisfied. The overall dataset included 14.8% dissatisfied customers. An independent samples t-test produces an F of 0.495. Therefore, no statistically significant difference exists between the two groups, and the cases with missing values are deleted.

Clustering and Neural Networks

A model based on the entire dataset will result in a suitable prediction of dissatisfaction. However, such a model also assumes that each customer is affected equally by the same factors. Perhaps, this assumption, so common in many models is false. To compensate for this assumption, we can cluster the data into groups of customers with similar characteristics. Similar customers are more likely to be dissatisfied by the same factors.

After clustering the data, we can build a model using a simple backpropagation neural network. Each cluster trains its own neural network separately. After adjusting the number of hidden nodes and other factors, the clustering and model can be applied to other datasets to predict dissatisfaction.

Two fundamental questions needed to be answered to perform the cluster. First, how many groups should be clustered, and second, which variables to use for the clustering. There are a number of methods available to determine the number of clusters. However, many of these methods are simply heuristic guidelines. To determine the appropriate number of groups for this analysis, we performed a hierarchical cluster on a subset of the data.

Hierarchical clusters require a large amount of computing resources and memory to perform. To economize, 1% of the cases were randomly selected and clustered. We used the between-groups linkage method, and squared Euclidean distance to determine the interval. Examination of the dendrogram indicates the existence of three clusters.

The variables used to cluster the data were not selected using any statistical method. Since clustering is not an empirical method, only a select number of variables were used rather than the entire set of variables. Statisticians recommend using an explicit theory to select the appropriate variables (Aldenderfer). Since no theory (that was available for this analysis) has been thoroughly tested, we simply use the best

judgement of which variables are most appropriate. The implicit hypothesis of this study indicates that dissatisfaction is a result of the consumer's spending habits and the features of their Capital One credit card. Therefore, we selected ten relevant variables. From the survey data, we selected the following:

- V2: How much money did you spend on purchases in the last 30 days?
- V3: How many times did you make purchases in the last 30 days?
- V10: How many years have you had any credit card?

From the internal data, we selected the following variables:

- V11: The average daily balance.
- V12: The current balance.
- V13: The current credit limit.
- V14: How many months the customer is past due.
- V15: The annual percentage rate.
- V16: Index of credit worthiness.
- V17: The number of months with a Capital One credit card.
- V18: Initial credit limit assigned when account was opened.

These variables will combine to indicate three areas groups. Variables v13, v15, v18 indicates the features of the credit card. Capital One offers cards using plans with the same feature, such as the Visa Gold and Visa Platinum plans. Therefore, it is likely that many cases will have identical or similar values for these variables.

Variables v2, v3, v10, v11, v12, and v17 will indicate the purchasing behavior of the customer. Notice it includes information on frequency of purchases, average balance, and length of customer status. Finally, variables v14 and v16 rate each customer's credit worthiness, and verifies whether the credit plan meets the needs of the customer.

Once we know the number of clusters, and the appropriate variables, we can cluster the dataset using a k-means cluster. Set the maximum iterations to ten and the convergence criterion to zero. The results produce three distinct groups of customers. The complete listing of cluster centers can be found in appendix A.

Cluster 1 members have the lowest average daily balance (center: \$1,485.66), make the fewest number of purchases. They have had a Capital One credit card longer than the other groups (center: 17.09 months).

Cluster 2 members lie between clusters 1 and 3. They make the most purchases (center: 6.02 in the last 30 days). Their average daily balance is \$2620.85, and their credit worthiness centers on 135.

Cluster 3 members are the biggest spenders of the clusters. The center of their current balance is \$4,097.69. They have a high credit limit and a high credit worthiness. Although they have had a credit card more years than the other clusters, they have had a Capital One credit card fewer months than the other groups (center: 16.84 months).

A view of the data confirms the existence of three clusters on the factors of current balance, average balance, and credit limit. These factors are directly related to each other. The credit limit especially indicates that the dataset contains three different types of credit cards, with centers around \$2,000, \$3,000 and \$5,000.

After the data has been clustered, each group can be trained separately using a neural network. To train each cluster's network, all the variables (v1-v24) were used as input nodes, with variable v25 as the only output node.

For each cluster, we used a backpropagation network. Fifty hidden nodes were used, which is slightly below the default number of nodes provided by the software package. The learn rate was kept at 0.6, and the momentum was set at 0.9. We also used a very simple complexity, and a rotational pattern selection. Clusters were allowed to learn from 4 minutes up to 2 hours. Approximately ten minutes of learning was all that was required to determine a decent test set.

The cluster 1 model had an R squared of 0.0379. The mean absolute error was 0.310. The output file produced predictions ranging from approximately 0.8 to 0. By changing the cutoff value between satisfaction and dissatisfaction, the model varies in its accuracy. At a cutoff value of 0.32, the model predicts 123 dissatisfied customers. Of those, 63 were incorrectly predicted and 60 were correct; or 48.7% correct.

The model from cluster 2 had an R squared of 0.032 and a mean absolute error of 0.217. Predictions in the output file ranged from 0.79 to 0. A cutoff of 0.25 can produce

a success rate of up to 80%. Yet at that value, the model only predicts six cases as dissatisfied. A more useful range of 0.2 predicts 268 dissatisfied customers, 33.2% of which are correct.

The model for cluster 3 had an R squared of 0.0204 and a mean absolute error of 0.180. However, even at a cutoff of 0.16, the model was only predicting dissatisfaction accurately 17.8% of the time. Only 10% of the customers in cluster 3 were dissatisfied, compared to 13.2% in cluster 2, and 20.1% in cluster 1

CART

It is good practice to experiment with a number of different methods when modeling or mining data. Different techniques may shed new light on a problem or confirm previous conclusions. We can use CART analysis in this respect.

A trial version of the CART software is available from the Salford Systems website. It provides a number of options to adjust and refine the tree model. It supports five different splitting methods including two Gini, two twoing, and a class probability method. In addition, it permits the user to incorporate linear combination splits.

CART produces trees that use an elegant color scheme to identify the best terminal nodes. Terminal nodes with high proportions of the selected target are a dark red, while nodes with low proportions of the target are dark blue. Nodes with moderate amounts of both targets are colored light blue and pink. Therefore, the user can immediately see if a tree is accurate, in which case the tree would appear with dark red and dark blue nodes.

CART is a very appropriate method for this dataset since the goal is to predict satisfaction, which is a binary variable. The Gini method of splitting works well with such binary variables. In addition, CART, unlike neural networks allows users to alter the costs associated with misclassifications. This feature proves useful for this analysis, since we are concerned with targeting only the dissatisfied not the satisfied customers. We can increase the cost of misclassifying a satisfied as a dissatisfied customer. A third advantage of CART is that it allows the user to set the prior probabilities (or just priors). We are trying to identify a relatively rare occurrence. In many regression and neural network methods, the large proportion of satisfied customers swamp the dissatisfied customers. Indeed, with this dataset, we can predict every customer as

satisfied, and have a very respectable error rating of 15%. Adjusting the priors helps to prevent this swamping effect.

For this analysis, we experimented with a number of different splitting methods, costs, and pruning methods. In the end two methods emerged as the best: a Gini method with the priors set to match the data, and a Gini method using linear combinations. For both trials one-third of the data was selected randomly for testing. The cost of misclassifying a satisfied (1) as dissatisfied (0) was set to 1.5 for the regular tree and 1.1 for the linear combination. The favor-even-splits parameter was set to 0.28 for the regular Gini and 0.26 for the linear combination method. Since so few cases are dissatisfied, it is unnecessary to attempt even splits. Doing so would only force satisfied customers into dissatisfied terminal nodes.

The regular tree began with 1150 terminal nodes, from which it was pruned down to an optimal tree of 20 nodes with a relative cost of 0.990. The root node, at the base of the tree split the dataset by the current balance. Those with a balance less than or equal to \$1,950 (approx. 30%) were split further, while the other 70% were placed in a terminal node largely composed of satisfied customers. The current balance was an important factor in all of the trees regardless of splitting method. Other important factors included the credit worthiness index, years with a credit card and credit on non-Capital One cards. For more details on the full tree see Appendix B.

In fact, other trees provided better relative costs. However, those trees assumed equal priors and costs, and few of the nodes accurately predicted dissatisfied customers. The greatest concern for this analysis isn't the overall relative cost. The concern is to create a tree, which has a few terminal nodes with a high purity of numerous dissatisfied customers.

Interestingly, linear combination methods will produce trees with only marginally improved performance. The relative costs of these trees are comparable to the regular trees. The relative cost of the optimal tree was 0.997, compared to 0.990 above. The optimal tree had only four terminal nodes, which were pruned down from 963 nodes. However, we allow the tree to grow four levels resulting in 18 terminal nodes. The root node splits the dataset on balance on non-Capital One cards (v5), credit on non-Capital

One cards (v6), the current balance (v12), and credit worthiness index (v16) with the following formula:

$$\text{Node 1} = 0.659(v5) - 0.751(v6) + 0.0014(v12) - 0.0253(v16).$$

The linear combination had purer learning nodes, but less pure test nodes. The test nodes are the better indicator of the overall performance of the model. The purest learning node had over 70% dissatisfied customers. However, on the test set, the best node had only 33% dissatisfied customers, with the worst node including only 11%. These results are similar, if not worse than the non linear combination tree. This is likely due to the lower cost of misclassifying a satisfied as a dissatisfied customer. For a detailed look at the structure of this tree, see Appendix C.

Method Comparison

Since this study is essentially a business problem, the most logical means of comparing CART to the clustering/neural network method is in terms of profitability. To translate the models into profit requires a number of assumptions about the costs and benefits of a retention program. Values will be assumed for the following factors:

- Revenue per customer per year
- Cost of the retention program per customer
- Rate of retention of dissatisfied customers exposed to retention program

Also assume that

- Satisfied customers do not defect
- Dissatisfied customers not exposed to the retention program defect

Fortunately, many of these values can be estimated using data in the dataset. It is known that the mean of the average daily balances per customer is \$2,532. Furthermore, the average annual percentage rate is 11%. Ignoring other revenues, each customer provides \$278.52 in revenue per year.

The cost-benefit calculation is as follows:

$$P = (Y * M * D) - C * (D + S)$$

Where:

C – Cost of the program per customer

Y – The benefit of retaining one customer for a year (\$278.52)

M – The rate of retention of dissatisfieds exposed to program

D – Number dissatisfieds exposed to program (correct classifications)

S – Number satisfieds exposed to program (misclassifications)

We can also assume that C is related to M, probably by a limited growth function. However, without further studies we are uncertain of the exact relationship between these factors. Therefore we will allow the decision makers to adjust them in a spreadsheet, see Appendix D. For this analysis, we will set the cost of the program per customer, $C = \$15.00$, and the rate of retention, $M = 25\%$.

The values for D and S are gathered from the results of the models above. To provide a fair comparison of CART with and without linear combinations and clustering/neural network we will choose customers with the top 300 to 400 predictions. However, we cannot get an exact amount since we have to use cutoff values for the neural network predictions and groups of predictions in the tree nodes. To further ensure fairness, the CART results are from the combined learning and training dataset. It was stated above that the training set is a more accurate prediction of the model. However, the neural network predictions were verified using the entire dataset, so we will use the entire dataset for the CART predictions as well.

Method	Num Pred.	Num Satisfied	Num Dissat.	Missclass Rate	Profit	Profit per cust
Cluster/Neural Net	391	242	149	62%	\$4509.87	\$11.53
Regular Gini	442	222	220	50%	\$7787.72	\$17.62
Linear Combo Gini	304	179	125	59%	\$4143.36	\$13.63

This table illustrates that both CART trees are superior to the clustering/neural network method. The regular Gini will have an average profit of \$17.62 per customer exposed to

the program. The linear combo Gini will realize an average profit of \$13.63 per customer, while the clustering/neural net will only average \$11.53.

Results and Conclusions

The underlying structure of this data becomes apparent with the application of the methods described above. Clustering indicates the existence of three separate groups, based mainly on the balance and credit limit criteria. Visualization of these factors through histograms confirms the presence of three distinct products with different credit limits. As expected, the credit worthiness indices have a positive relationship with the credit limit.

Although the underlying structure is easily identified, predicting the target variable, satisfaction remains a difficult task. The difficulty arises not in the creation of a general model, but in a model that can accurately pinpoint the minority of dissatisfied customers. It appears that CART is the best technique for this purpose, since it allows the user to adjust the priors and costs essential to isolating those few dissatisfied cases. Even with these capabilities, the terminal nodes in the tree produced by CART still have a high degree of impurity.

A perfect model would be a simple tree in which the terminal nodes are completely pure. Since a perfect model is infeasible, the CART model requires the user to tradeoff between two factors:

- A. Isolating nodes of purely dissatisfied customers
- B. Capturing all of the dissatisfied customers

Approaching this problem from a marketing objective, the most attractive alternative is A, which would maximize the return of the retention program. However, this does not account for lost revenue from customer defections, which would tend to make alternative B a priority.

The accuracy of the model is not limited by CART, but rather the methodology of this study, which assumes that customer satisfaction is related to the features of the product and the account. In reality, a number of other factors such as customer service, personal attitudes, and competition impact satisfaction.

Regardless, the model produced in this analysis along with an effective retention campaign can be used to increase profitability. Based on the assumptions of the cost-benefit analysis above, profits are estimated to range between \$10 and \$20 per customer included in the program. At that level, it would be foolhardy not to engage in such a campaign.

Programs used in this analysis

SPSS 10.0

Microsoft Excel 2000

NeuroShell 2

CART 4.0

Bibliography

Aldenderfer, Mark and Blashfield, Roger. Cluster Analysis. Sage Publications. London. 1985.

“Broad approach to cardholder satisfaction is best.” Card News
31 August, 1998.

“J.D. Power study gauges Canadian cardholder behavior, desires.” Card News
12 October, 1998.

“MBNA, Citi, and AmEx top in cardholder loyalty.” Cards International.
20 August, 1999: 4.