Paper 223-2011

Improving Customer Loyalty Program through Text Mining of Customers' Comments

Maheshwar Nareddy and Goutam Chakraborty
Oklahoma State University, Stillwater, OK, USA

ABSTRACT

Typical surveys contain closed-end questions that generate structured numerical data and open-ended questions that generate unstructured textual data. Given the perceived difficulties in analyzing texts, most companies often ignore textual data or simply look at summaries of comments. We illustrate how textual data can be grouped together to generate insights into customers' expectations and how such groupings can be used as input variables to build better predictive models than models based on numerical data alone. Data was collected at a national conference via a survey that had numerical and four open-ended questions. Ten clusters were found in grouping of textual comments. The target is a binary (Yes/No) variable about best loyalty program in the industry. Data was split into training and validation before building predictive models. The best predictive model using numerical data has a misclassification rate of only 26.5% and a sensitivity of 60% in the validation sample. The addition of the cluster memberships as input variables substantially increased the performance of the predictive model (misclassification reduced to 18.7% and sensitivity increased to 82.8%).

INTRODUCTION

Customer Relationship Management (CRM) is considered to be one of the important responsibilities of marketing. In order to retain existing customers and attract new ones, many companies use customer loyalty programs. This practice is very common among retail businesses where it has been popular for quite some time. Periodic evaluation of these loyalty programs through customer surveys helps in identifying the important updates that need to be done to the programs to keep them effective.

Customer surveys typically contain closed-end questions that generate structured numerical data and open-ended questions or comments that generate unstructured textual data. Survey analysis so far has been confined mostly to structured quantitative data analysis. According to Dobson (2010), many companies fail to analyze the patterns in their large sets of data. However, textual comments and responses provided in these surveys contain a wealth of information (Boire, 2009; Dobson, 2010). When these textual comments are further explored using text mining techniques, they have the potential to increase the predictive and/or explanatory ability of models (Sullivan and Ellingsworth, 2003). Kleij and Musters (2003) demonstrate how text analysis of open-ended survey responses can complement preference mapping.

We illustrate the importance of analyzing textual data and using it to predict whether a customer feels that a company's loyalty program is the best compared to those offered by its competitors. The textual responses from the customers are first clustered using the text mining node in SAS® Enterprise Miner™ 6.1, and later these clusters are used as inputs to predict the target variable.

COMPANY BACKGROUND

The data used for this paper was collected via a survey administered at a national conference by a client company (that wishes to remain anonymous) to determine whether the company has the best loyalty program in the industry. This company provides both B2B and B2C products and services in the retail sector. The company has a loyalty program in place through which their customers can redeem their loyalty points for free items such as general merchandise, coffee, drinks, etc. The company sells products and services through a network of wholly owned and franchised stores all over the USA.

SURVEY DETAILS

The original survey had 20 questions (with numerical responses) and four open-ended questions (textual responses). A total of 315 completed surveys were obtained by a market research company (working on behalf of the client company) at a national industry conference. We use eight numerical questions and four open-ended textual questions to predict the target variable. The target variable captures customers' response to the question of whether the client company has the best loyalty program. Of the 315 respondents, 55.56% felt that the client company has the best loyalty program in the industry.

The structured quantitative variables comprised responses to questions similar to the following.

1. What percent of the company's products is bought by the respondent? (1% -19%, 20%-39% and so on)
2. Does the respondent have complete freedom to buy the client company's products/services or does he/she work for a company that has an existing contract to buy the products/services of the client company? (Yes/No)

3. What method does the respondent prefer to redeem the loyalty points? (redemption coupons, loyalty card or either method)
4. Is the respondent aware of other services provided by the loyalty program beyond reward points?(Yes/No)

The textual comments comprised responses to open-ended questions similar to the following.

1. Why does the respondent feel that a company has the best or worst loyalty program?
2. Why does the respondent feel that a store is his/her favorite or least favorite stop-to-shop?

Based on the type of questions, the survey data was divided into two data sets – textual data set and numeric data set. Text mining techniques were applied on the textual data set and later this textual data set with clusters was merged with the numeric data set to build predictive models.

METHODOLOGY

TEXT MINING

The textual data set has four variables capturing textual responses. Each of these variables was used to cluster the respondents using the text mining node (Refer to Figure 1) in SAS® Enterprise Miner™ 6.1. As shown in the partial view of the properties used for text mining (Refer to Figure 2), a 10-cluster solution was used for clustering textual comments. Based on the discriminative power of the weights calculated in the Term-Document Matrix (Refer to Figure 3), a customized stop-list specific to the domain was developed by the authors, and this list was used during the clustering of textual comments. Figure 4 shows that there is one large cluster that also includes all observations that had no textual comments. Descriptive terms of each cluster indicate the common words used for clustering the textual comments. For example, cluster six has descriptive terms like "bad service," "bad experience," etc. Similarly, cluster nine has descriptive terms like "slow," "customer service," "rude people," etc. All observations with these kinds of descriptive terms in their textual comments fall into respective clusters. We can identify the important factors that drive customers' perceptions of best or worst loyalty programs and favorite or least favorite stop-to-shop by analyzing these clusters and their descriptive terms. Such important factors that drive customers' perception are summarized in Table 1.

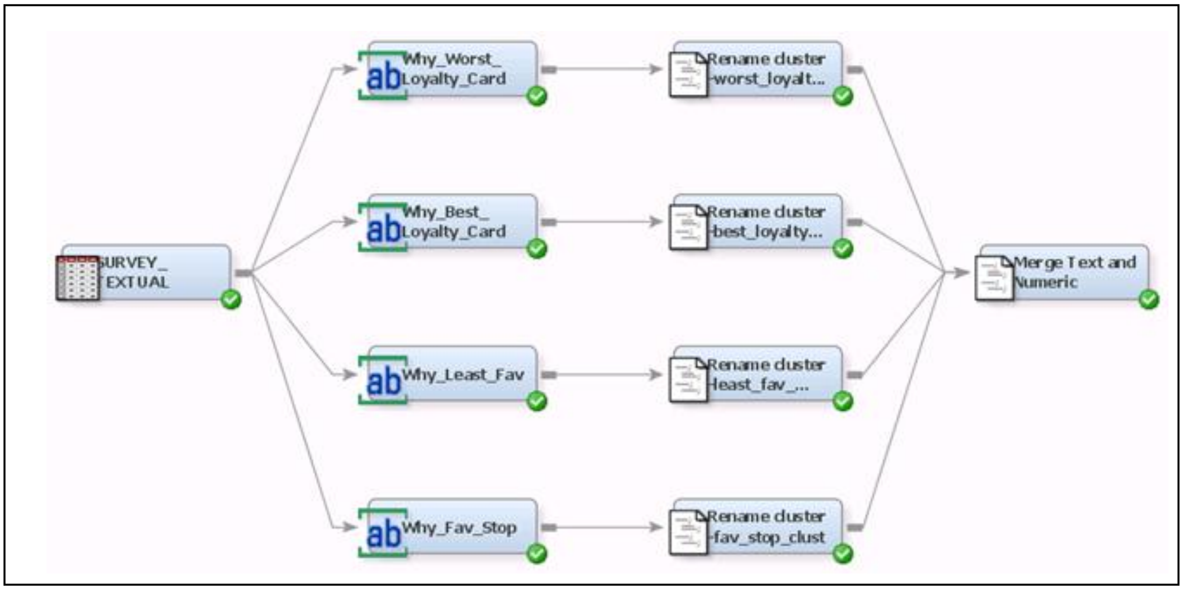

Figure 1. Application of Text Mining on Textual Data

Property	Value
Different Parts of Speech	Yes
Ignore Parts of Speech	...
Noun Groups	Yes
Synonyms	Sashelp.engsynms ...
Find Entities	No
Types of Entities	...
Transform	
Compute SVD	Yes
SVD Resolution	Low
Max SVD Dimensions	100
Scale SVD Dimensions	No
Frequency weighting	Log
Term Weight	Entropy
Roll up Terms	No
No. of Rolled-up Terms	100
Drop Other Terms	No
Cluster	
Automatically Cluster	Yes
Exact or Maximum Number	Exact
Number of Clusters	10
Cluster Algorithm	EXPECTATION-MAXIMIZATIOI
Ignore Outliers	No
Hierarchy Levels	.
Descriptive Terms	15

Figure 2. Properties of Text Mining Node

Terms

TERM	FREQ	# DOCS	KEEP	WEIGHT ▼	ROLE	ATTRIBUTE
bad experience	2	2	☑	0.88	NOUN_GROUP	Alpha
cost	2	2	☑	0.88	Noun	Alpha
crowded	2	2	☑	0.88	Adj	Alpha
desk	2	2	☑	0.88	Noun	Alpha
discount	2	2	☑	0.88	Noun	Alpha
experience	2	2	☑	0.88	Noun	Alpha
inconvenience	2	2	☑	0.88	Noun	Alpha
lack	2	2	☑	0.88	Noun	Alpha
much	2	2	☑	0.88	Adj	Alpha
nasty	2	2	☑	0.88	Adj	Alpha
productx desk	2	2	☑	0.88	NOUN_GROUP	Alpha
purchase	2	2	☑	0.88	Noun	Alpha
quality	2	2	☑	0.88	Noun	Alpha
restraunts	2	2	☑	0.88	Noun	Alpha
room	2	2	☑	0.88	Noun	Alpha
rude people	2	2	☑	0.88	NOUN_GROUP	Alpha
slow service	2	2	☑	0.88	NOUN_GROUP	Alpha
smoke	2	2	☑	0.88	Noun	Alpha
stop	2	2	☑	0.88	Verb	Alpha
suck	2	2	☑	0.88	Verb	Alpha
system	2	2	☑	0.88	Noun	Alpha

Figure 3. Partial View of Customized Stop-List Based on Term-Document Matrix

#	DESCRIPTIVE TERMS	FREQ	PERCENTAGE	RMS STD.
1	+ little, suck, smoke, inconvenience, di scount	109	0.3460317460...	0.0
2	clean, + have, coffee, clean, outdated , + facility, + high, productx	19	0.0603174603...	0.1714528...
3	+ park lot, + lot, restraunts, + park	3	0.0095238095...	0.0389897...
4	slow, + crowd, nasty, nasty, ta, dirty	19	0.0603174603...	0.0525598...
5	many, enough, cleanliness, + restaura nt, + long, + line, out, don, lack, room , + bad food, + long line, + pilot, + pr oductx price, + reason	55	0.1746031746...	0.1779368...
6	bad service, bad experience, experien ce, + bad shower, + hole, bad, + servi ce, + shower, + park, productx	26	0.0825396825...	0.1498996...
7	don, + driver, like	13	0.0412698412...	0.1244634...
8	expensive, + employee, crowded, muc h, purchase, system, + discount, + sto re, + dirty, + attitude, friendly, + truc k, poor, bad, productx	31	0.0984126984...	0.1751188...
9	slow, customer service, rude people, sl ow service, visa, + productx card, + vi sa card, people, + card, + customer, r ude, productx, + service, bad, poor	30	0.0952380952...	0.1775152...
10	+ high price, + price, + high	10	0.0317460317...	0.0612858...

Figure 4. Descriptive Terms of Each Cluster

PREDICTIVE MODELING

Before applying the models, a tree imputation method was used to impute the missing values of all numerical variables. After imputation, a stratified partitioning method was applied to the data set to split it into training (80%) to build predictive models and validation (20%) to fine-tune and test the models. Two different types of models were built. In the first model type, only numerical (quantitative) responses were considered as inputs for prediction purposes. In the second model type, the clusters representing the textual comments were added along with the numerical responses as inputs to predict the target.

Various models, including Logistic Regression (both stepwise and forward variable selection methods); Decision Trees with different criteria such as ProbChisq, Gini, and Entropy; and Artificial Neural Networks (ANN) with variable selection were used to predict the target as shown in Figure 5.

A model comparison node was used to evaluate the performance of each model against other models using the validation misclassification rate criteria. In the final stage of the model, one best model from each of these two types (numerical inputs only and numerical and textual inputs) was selected to evaluate the effect of textual input data.

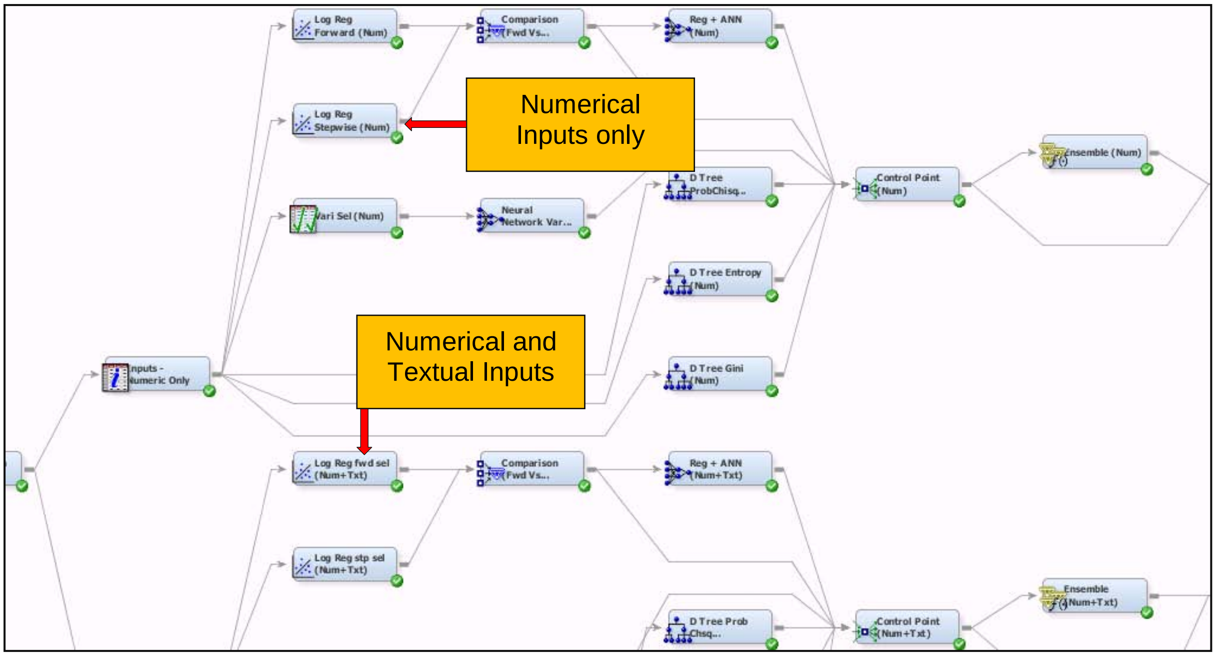

Figure 5. Model Diagram Showing Different Predictive Models and Types of Inputs Used

RESULTS

The grouping of the textual data revealed several important factors for each of the four textual questions relating to best/worst loyalty programs and most/least favorite stop-to-shop. These factors are inferred based on our subjective interpretation of terms in each of the clusters for each question. These factors are summarized in Table 1.

Customers' perception of favorite stop to shop	Customers' perception of least favorite stop to shop	Customers' perception of best loyalty program	Customers' perception of worst loyalty program
Faster customer service	Slow customer service & bad service experience	Free drink refill	Mailing the redemption coupons
Large & convenient parking	Dirty restrooms & facilities	Free showers & Unlimited showers	
Free coffee & drinks	Long queues/lines	Double reward points per dollar spent	
	Rude employees	Redemption points expire late	
		Instant redemption of points	

Table 1. Summary of Different Factors Based on Clustering of Textual Responses

Comparison of the best model using numerical data only (referred to as Num in the screenshots) with the best model using numerical plus textual data (referred to as Num +Txt in the screenshots) clearly shows the utility of using textual data in addition to the numerical data. Table 2 shows that the Num + Txt model substantially outperforms the Num model. The validation misclassification rate was reduced from 26.56% to 18.75%.

Fit Statistics

Selected Model	Predecessor Node	Model Node	Model Description	Train: Misclassification Rate	Valid: Misclassification Rate	Train: Number of Wrong Classifications	Valid: Number of Wrong Classifications
Y	MdlComp4	Neural4	Reg + ANN (Num+Txt)	0.195219	0.1875	49	12
	MdlComp2	Neural2	Reg + ANN (Num)	0.258964	0.265625	65	17

Table 2. Partial View of the Results of Model Comparison Node

The ROC charts in Figure 6 and the summary statistics reported in Table 3 provide more diagnostics that show the superiority of the Num + Txt model over the Num model. The area under the ROC curve of the model with numerical and textual inputs (green curve) is more compared to that of the model with numerical inputs alone (red curve).

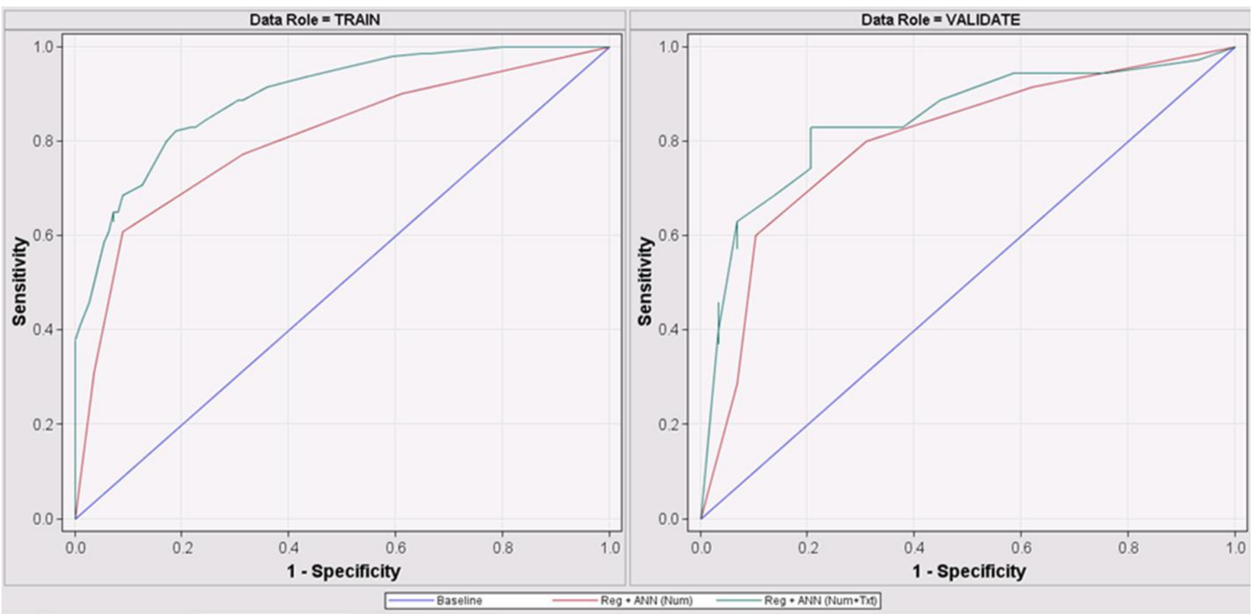

Figure 6. ROC Chart Showing the Differences in the Predictive Ability of Models

The sensitivities of different models are calculated based on the results of the model comparison node. Sensitivities of the best models for both train and validation data sets are summarized in Table 3.

The formula used for calculating sensitivity is as follows:

Sensitivity = Number of True Positives***/*** (Number of True Positives + Number of False Negatives)

Model	Role	Misclassification	Sensitivity
Reg + ANN (Num+Txt)	Train	19.52%	82.86%
Reg + ANN (Num+Txt)	Validate	**18.75%**	82.86%
Reg + ANN (Num)	Train	25.90%	60.71%
Reg + ANN (Num)	Validate	**26.56%**	60.00%

Table 3. Misclassification Rates and Sensitivities of the Two Best Models

DISCUSSION

As shown in Table 1, customers' perceptions of least favorite stop-to-shop depends on the factors similar to: rude employee behavior, long lines, bad/slow service experience, high prices, and unclean facilities. On the other hand, a customer's perception of his/her favorite stop-to-shop is influenced by the factors similar to: faster customer service, free coffee/drinks, and large and convenient parking. These findings provide actionable diagnostics for the client company to work on improving the image of their shops.

As shown in Table 1, customers' perception of best or worst loyalty program depends on the factors similar to: late expiration of loyalty points, free/unlimited refills, double reward points per dollar spent, and instant redemption of points rather than mailing back redemption coupons. These findings provide actionable diagnostics for the client company to work on improving its loyalty program.

As shown in Table 2, the best model selected using the validation misclassification criteria is Reg + ANN (Num + Txt). This model is an ANN predictive model using a regression node for input variable selection with both numerical and textual inputs. Reg + ANN (Num) is another ANN predictive model using a regression node for input variable selection with numerical inputs only.

A comparison of different statistics (refer to Tables 2 and 3) of these two models clearly shows that beyond just reduction of the validation misclassification rate, there is a substantial difference in the sensitivity of both the models. The best model with numerical and textual inputs has a sensitivity of 82.86% in the validation data, whereas the other model has a sensitivity of merely 60%. Another important finding from Table 3 is that the misclassification rates and sensitivities of both training and validation data of each of these models are closer. This indicates that the predictive models developed in this demonstration are stable.

From the ROC chart (refer to Figure 6), we can conclude that the winning ANN model (with numerical and textual inputs) definitely fares well over the ANN model (with numerical inputs only) in terms of predicting whether a customer feels a company's loyalty program is the best compared to its competitors. In the validation data, the area under the green curve (best model with numerical and textual inputs) is much greater than the red curve (best model with numerical inputs only). More importantly, the green curve is above the red curve for the entire range of horizontal axis.

CONCLUSION

Periodic evaluation of loyalty programs through survey analysis not only helps companies understand customers' ever-changing needs, but also provides insightful information about where they stand in providing service relative to their competitors. When analyzed together, unstructured textual comments along with the quantitative data from the customer surveys can create a wealth of information to better understand customers' expectations and predict customers' satisfaction levels. This in turn will assist companies in designing and revamping their loyalty programs, thus allowing them to position themselves well ahead of the competition.

REFERENCES

Boire, Richard. 2009. "Data Mining For Customer Loyalty." *Direct Marketing*.

Dobson, David. 2010. "Segmenting Textual Data for Automobile Insurance Claims." *Proceedings of the SAS® Global 2010 Conference.* Available at http://support.sas.com/resources/papers/proceedings10/125-2010.pdf

Kleij, Frederieketen and Musters, Pieter A.D. 2003. "Text Analysis of Open-Ended Survey Responses: A Complementary Method to Preference Mapping." *Food Quality and Preference*, Vol.14 (1). pp. 43–52.

Sullivan, Dan and Ellingsworth, Marty. 2003. "Text Mining Improves Business Intelligence and Predictive Modeling in Insurance." *DM Review Magazine*. Available at http://www.information-management.com/issues/20030701/6995-1.html

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Name: Maheshwar Nareddy
Enterprise: Oklahoma State University
Address: Department of MSIS, Oklahoma State University
City, State ZIP: Stillwater, OK - 74074
Work Phone: (405)564-3719
E-mail: maheshwar.nareddy@okstate.edu

Name: Dr. Goutam Chakraborty
Enterprise: Oklahoma State University
Address: Department of Marketing, Oklahoma State University
City, State ZIP: Stillwater, OK - 74074
Work Phone: (405)744-7644
E-mail: goutam.chakraborty@okstate.edu