Using SAS® Enterprise Miner for Categorization of Customer Comments to Improve Services at the U.S. Postal Service

Olayemi Olatunji, United States Postal Service Office of Inspector General, Arlington, VA

ABSTRACT

Delivering high-quality service and providing excellent customer experiences are performance outcome goals the U.S. Postal Service has established to measure corporate strategy success and continuous improvement efforts.

Social media has opened the door for customer engagement and decision making. With the help of Twitter, Facebook, and Yelp, government agencies are more informed about how customers feel about their service and experience. Using Yelp data we will text mine comments about U.S. Postal Service customer service, retail service, mail delivery, and facility services using SAS® Text Miner of SAS® Enterprise Miner 7.1. The aim of this paper is to provide ways to categorize consumer comments regarding U.S. Postal Service services to improve the customer experiences at stations.

INTRODUCTION

In the mail delivery industry, customers are able to choose among multiple service providers. In this competitive industry, customers are influenced by the service and shipping option. With the rise of social media, customer perspectives are being heard and influencing change for businesses.

DATA PREPARATION

The data was collected using Python from a Yelp API. The customer comments, ratings, and locations were extracted for analysis.

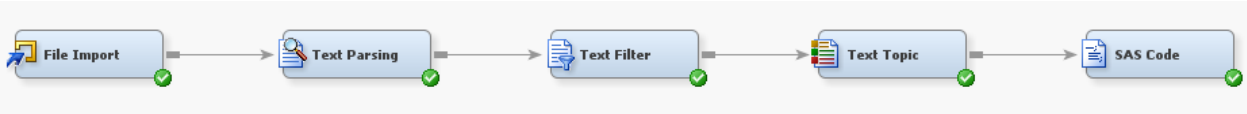

Figure 1: Process Flow

FILE IMPORT

The Yelp data was saved as a CSV file. File Import node was used to import the data. The import file was directed to the location of the csv file and the delimiter was set to comma. This process gathered the data into the SAS® Environment to begin the analysis.

Train	
Variables	
Import File	R:\IRAT_Misc\oolatunji\Social Media
Maximum rows to import	1000000
Maximum columns to import	10000
Delimiter	,
Name Row	Yes
Number of rows to skip	0
Guessing Rows	500
File Location	Local
File Type	csv

Figure 2: File Import Node

TEXT PARSING

The Text Parsing node is the first step to begin text mining. Text parsing deconstructs the textual data, calculates the frequency of occurrence of each term in the document and determines the part of speech.

Train	
Variables	
Parse	
Parse Variable	REVIEW
Language	English
Detect	
Different Parts of Speec	No
Noun Groups	No
Multi-word Terms	SASHELP.ENG_MULTI
Find Entities	Standard
Custom Entities	
Ignore	
Ignore Parts of Speech	'Aux' 'Conj' 'Det' 'Inte
Ignore Types of Entities	
Ignore Types of Attribut	'Num' 'Punct'
Synonyms	
Stem Terms	Yes
Synonyms	SASHELP.ENGSYNMS
Filter	
Start List	
Stop List	SASHELP.ENGSTOP
Select Languages	

Figure 3: Text Parsing Node

The following changes were made to the node:

- Detect different parts of speech was set to ‘No’ to eliminate repetitive terms
- Detect Noun Groups was set to ‘No’ to eliminate repetitive terms
- Find Entities was set to ‘Standard’ so that the node would be able to identify organizations and time period
- Synonyms for post office, passport, deliver, and parking lot were created

The results of the text parsing node showed the most common words were terms such as ‘post office’, ‘customer service’, ‘line’, and ‘time’.

TEXT FILTER

After the Text Parser node, the Text Filter node is connected to reduce the number of terms for building the topics and clusters. The terms that were kept were terms identified as most significant and useful.

Train	
Variables	
Spelling	
Check Spelling	Yes
Dictionary	
Weightings	
Frequency Weighting	Default
Term Weight	Default
Term Filters	
Minimum Number of Documents	6
Maximum Number of Terms	.
Import Synonyms	
Document Filters	
Search Expression	
Subset Documents	
Results	
Filter Viewer	
Spell-Checking Results	EMWS5.TextF
Exported Synonyms	

Figure 4: Text Filter Node

The following changes were made to text filter:

- Check Spelling was set to 'Yes' to correct misspelled terms
- Term Filters Minimum Number of Document was set to '6' to reduce the number of insignificant terms

The results of the text filter show low weighted terms were dropped.

Terms

Term	Role	Attribute	Status	Weight	Imported Frequency
+ be ...		Alpha	Drop	0.000	345
+ post office ...	Organization	Entity	Keep	0.068	215
+ have ...		Alpha	Drop	0.000	145
not ...		Alpha	Drop	0.000	140
+ do ...		Alpha	Drop	0.000	103
+ customer service ...		Alpha	Keep	0.161	67
+ get ...		Alpha	Drop	0.000	66
+ go ...		Alpha	Drop	0.000	68
s ...		Alpha	Drop	0.000	70
+ line ...		Alpha	Keep	0.182	54
here ...		Alpha	Drop	0.000	54
+ time ...		Alpha	Keep	0.180	51
+ package ...		Alpha	Keep	0.210	62
+ mail ...		Alpha	Keep	0.213	49

Figure 5: Text Filter Dropped and Kept Terms

Analyzing the results in Filter Viewer allowed us to drop additional terms considered irrelevant in building the topics for the next phase. We also see the importance of terms and how they correlate with other terms.

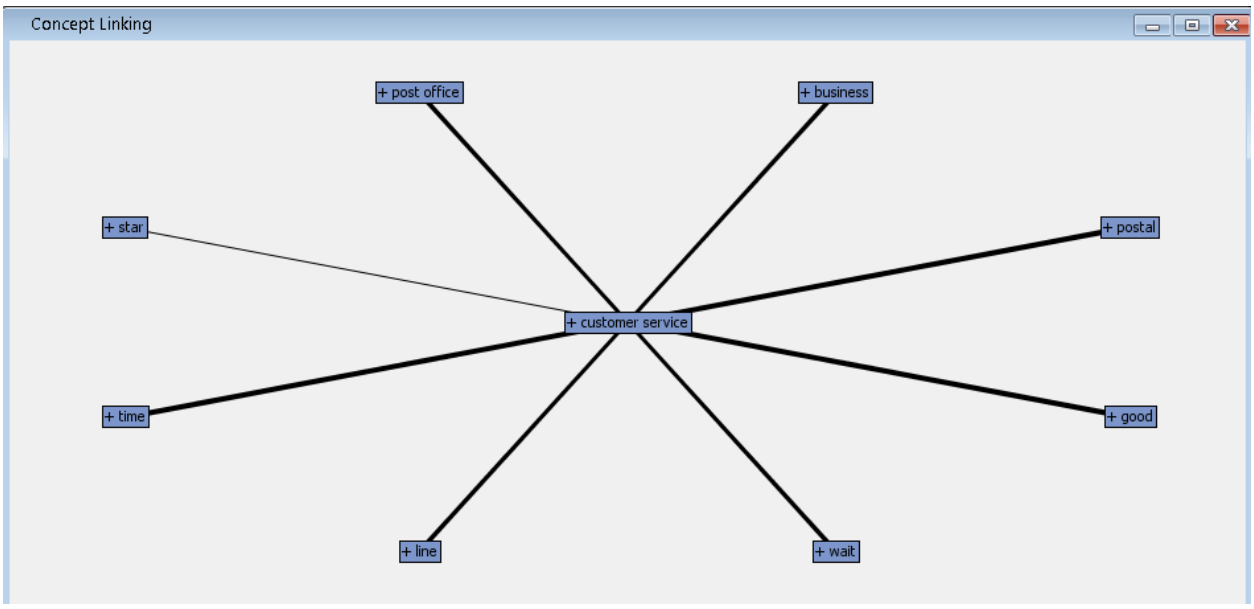

Figure 6: Concept Link diagram of 'customer service'

The concept link of 'customer service' show strong association with terms 'line', 'time', 'good', and 'wait'. Customers generally felt the service at the post office was good. Their complaints were about the wait time in line to be seen by an employee.

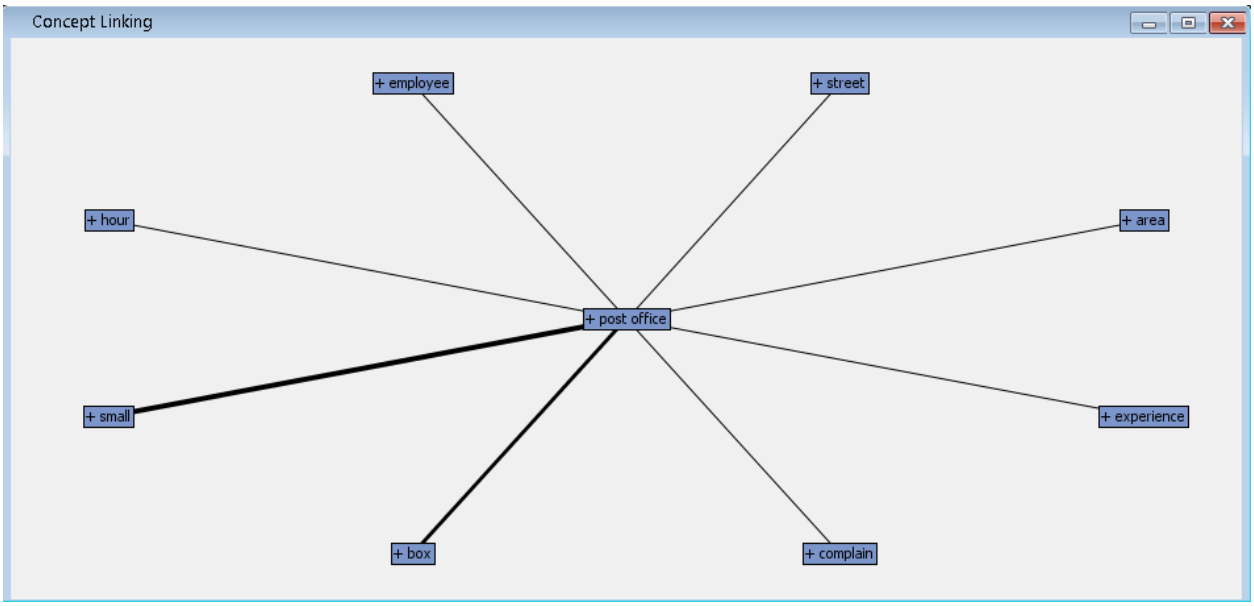

Figure 7: Concept Link diagram of 'post office'

The concept link of 'post office' show strong association with terms 'small', 'box', and 'hour'. Customers' comments about the post office were about the hours of operation, experience with the employee shipping or picking up a package, or lack of parking.

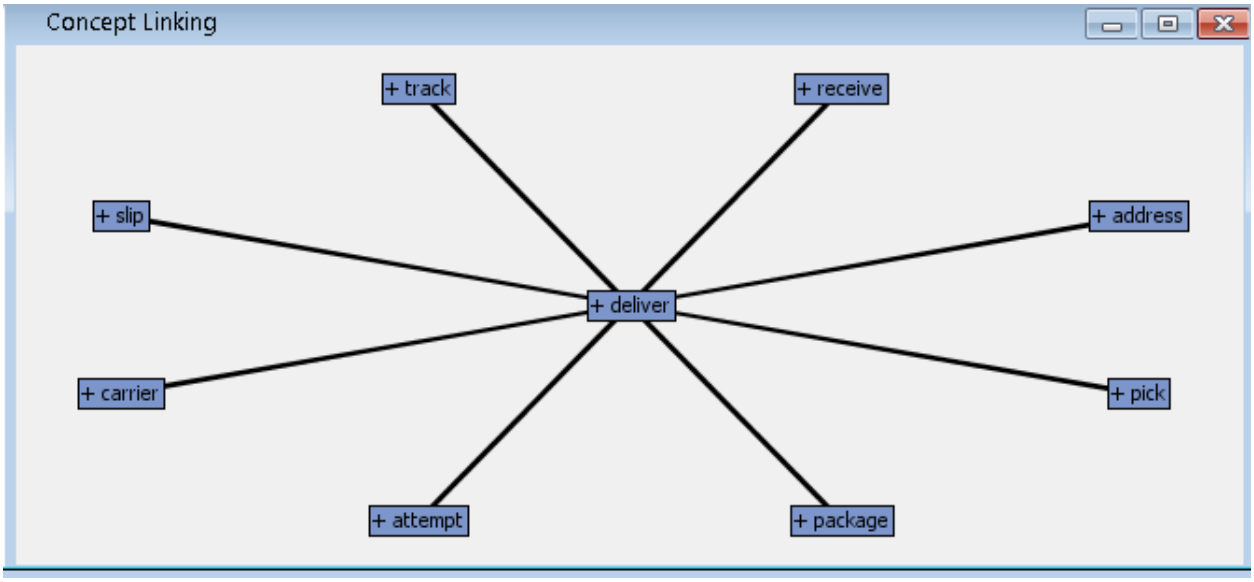

Figure 8: Concept Link diagram of 'deliver'

The concept link of 'deliver' show strong association with terms 'attempt', 'package', 'receive', 'pick'. Comments about delivery were a mix of positive and negative. For example, complaints of failed attempts by the postal carrier delivering mail or unable to track their package to its final destination. Customers also had a great experience picking up their package from the post office because the employee was helpful and friendly.

TEXT TOPIC

After filtering out low weight terms, the Text Topic node identifies terms that frequently associate together. 7 topics were created using the node after adjusting and analyzing the number of multi-terms topics. Yelp users' comments contain reviews of several topics pertaining to the appearance and location of the post office, the customer service, and experience of passport services. Because of these mixed focus in the reviews, text topic was used to allow terms replication in the topic.

Topic ID	Topic	Topic Category
1	+staff,+always,+friendly,+great,helpful	Customer Service
2	+deliver,+package,+carrier,+mail,+attempt	Mail Delivery
3	+passport,+application,+process,+photo,+renew	Retail Service
4	+park,+lot,+town,+small,+street	Facility Service
5	+box,+ship,+stamp,+tape,+buy	Retail Service
6	+wait,+line,+long,+minute,+open	Customer Service
7	+rude,+work,+know,+lady,+woman	Customer Service

Table 1: Text Topic Categories

SAS® CODE

SAS® Code node was connected to create unique categories of the topics. Topics were grouped into buckets for 'Customer Service' if it references the user experience with a clerk at the post office. Topics that contain terms about package services were categorized as 'Mail Delivery'. Topics that referenced passport or photo services were categorized as 'Retail Services'. Topics referencing the location and access to the post office in their town were categorized as 'Facility Service'.

The following code was executed to categorize the topics:

```
data EMWS.Yelp_Data;
set EMWS.Texttopic_train;
length Category $ 16;
if max(of TextTopic2_raw1-TextTopic2_raw7) = TextTopic2_raw1 then Category='Customer
Service';
else if max(of TextTopic2_raw1-TextTopic2_raw7) = TextTopic2_raw2 then Category='Mail
Delivery';
else if max(of TextTopic2_raw1-TextTopic2_raw7) = TextTopic2_raw3 then
Category='Retail Service';
else if max(of TextTopic2_raw1-TextTopic2_raw7) = TextTopic2_raw4 then
Category='Facility Service';
else if max(of TextTopic2_raw1-TextTopic2_raw7) = TextTopic2_raw5 then
Category='Retail Service';
else if max(of TextTopic2_raw1-TextTopic2_raw7) = TextTopic2_raw6 then
Category='Customer Service';
else if max(of TextTopic2_raw1-TextTopic2_raw7) = TextTopic2_raw7 then
Category='Customer Service';
else Category='Other';
run;
```

TEXT RULE BUILDER

The Text Rule Builder node generated rules based on the terms to predict each target category. The 'Generalization Error', 'Purity Rules', and 'Exhaustiveness' properties are set to 'medium' to ensure optimum rule generation without overfitting the model. The 'Change Target Values' in the property was used to reassign target values that were incorrectly classified.

Rules Obtained

Target Value	Rule #	Rule	Precision
FACILITY SERVICE	1	park & lot	92.45%
FACILITY SERVICE	2	park & ~post office	83.93%
FACILITY SERVICE	3	town & small	83.46%
CUSTOMER SERVICE	4	staff & ~park & ~address & ~passport & ~box	96.04%
CUSTOMER SERVICE	5	postal employee	96.30%
MAIL DELIVERY	6	deliver	83.87%
MAIL DELIVERY	7	delivery & package	85.71%
MAIL DELIVERY	8	carrier	84.52%
MAIL DELIVERY	9	claim	83.15%
RETAIL SERVICE	10	passport	96.43%
RETAIL SERVICE	11	box & ~park & ship	96.15%
RETAIL SERVICE	12	box & automate	96.43%
RETAIL SERVICE	13	pay & regular	96.63%
RETAIL SERVICE	14	tape	92.93%
RETAIL SERVICE	15	credit card	92.23%

Figure 9: Classification Rules Obtained

The text rule builder generated a set of 15 rules. The results can be interpreted as follows:

- Rule 1 specifies the presence of the term ‘park’ and ‘lot’, with a precision of 92.45% — a review that references facility services
- Rule 11 specifies the presence of the term ‘box’ and ‘ship’ and the absence of the term ‘park’, with a precision of 96.15% — a review that references retail services
- The validation data misclassification rate for the model is 15%

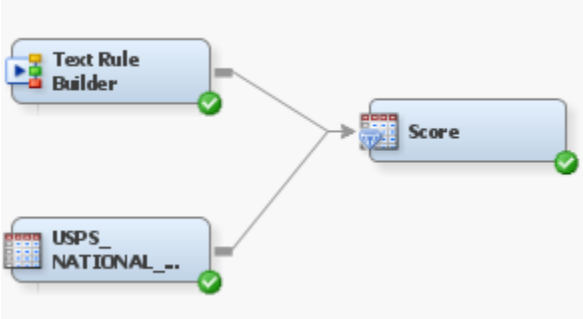

Figure 10: Diagram to Score New Data

We scored data collected about post offices across the nation to classify comments using the score node in SAS® Enterprise Miner.

CONCLUSION

The results of this analysis show that text mining can help the Postal Service understand common concerns of customers, such as long lines and rude and unhelpful employees; as well as customer concerns that are not always intuitive, such as parking limitations and need in variety of post office hours of operation. By taking action about these concerns, the postal service will be able to compete with other mail delivery services.

REFERENCES

Text Analytics Using SAS®Text Miner. Course Notes. 2014. SAS Institute Inc., Cary, NC

Chakraborty, Goutam, Murali Pagolu, and Satish Garla. 2013. *Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using* SAS®,. Cary, NC: SAS Institute Inc. 2013

CONTACT INFORMATION

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

Olayemi Olatunji
United States Postal Service Office of Inspector General
1735 North Lynn Street
Arlington, VA 22209
oolatunji@uspsoig.gov

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

The authors of the paper/presentation have prepared these works in the scope of their employment with U.S. Postal Service Office of Inspector General and the copyrights to these works are held by Olayemi Olatunji.

Therefore, Olayemi Olatunji hereby grants to SCSUG Inc a non-exclusive right in the copyright of the work to the SCSUG Inc to publish the work in the Publication, in all media, effective if and when the work is accepted for publication by SCSUG Inc.

This the 30th day of August, 2017.