



Big Data, Small Talk

By: Rahul Narula, Guoqiang Chen, Ben Ku
Faculty Advisors: Janet Duncan, Xiyue Liao

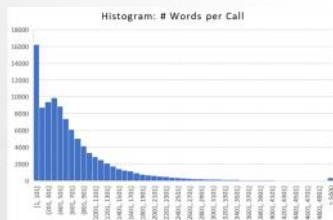
University of California, Santa Barbara – Department of Statistics

Abstract

The purpose of this project is to help an auto and homeowners' insurance company improve their customer service, make call center operations more efficient, and reduce costs by analyzing their customer calls. Text mining methods such as topic modeling and sentiment analysis are used to study roughly 100,000 non-claim customer calls from the last year. These models provide insight into what customers call about and how they feel during their calls, allowing the company to adjust their processes and make improvements accordingly. Results show what the most frequent topics of call are and how customer sentiment differs across topics.

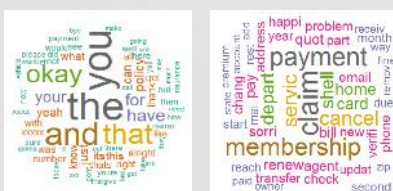
Dataset

- Random sample of 10,000 calls of non-claim calls from the last year
- Only contains text data
- All numbers removed to protect customer confidentiality
- Calls transcribed using NICE software – No human involvement
- Words per Call: 1-7,442
Mean: 679 Std. Dev: 750

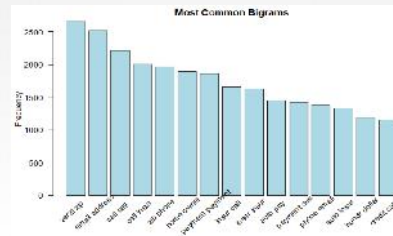


Pre-Processing

- Changing text case
- Removing punctuation and numbers
- Stemming words
- Removing stop words
 - Standard stop words
 - Low frequency stop words
 - Manually selected

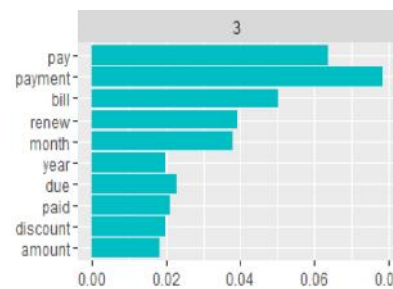


N-Grams



N-Grams are contiguous strings of n tokens in text or speech. They reveal underlying patterns and associations in language and are instrumental in text mining analysis. We analyzed consecutive words to discover and model recurrent phrases.

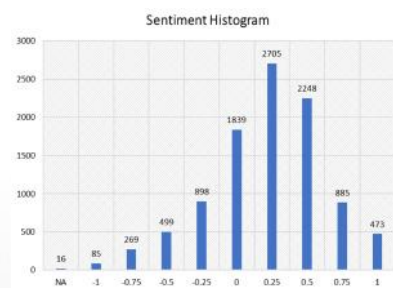
Topic Modeling



Topic	Proportion
[Uncategorized]	30%
Add/Remove Vehicle	10%
Admin	8%
Auto Policy	5%
Billing	20%
Cancel	9%
Change/Renew	7%
Coverage	5%
Premium/Rates	6%

Topic modeling is a text mining technique used to identify the various topics in a set of documents. We used three methods, Latent Dirichlet Allocation, Correlated Topic Models, and Non-Negative Matrix Factorization to discover the most common topics in customer calls.

Sentiment Analysis



Topic	Avg.	Std. Dev
[Uncategorized]	0.218	0.404
Add/Remove Vehicle	0.223	0.405
Administrative	0.233	0.412
Auto Policy	0.240	0.411
Billing	0.220	0.413
Cancellation	0.203	0.397
Change/Renew	0.221	0.404
Coverage	0.216	0.397
Premium/Rates	0.220	0.421

Sentiment analysis is the use of analytical methods to study emotions in text. Using sentiment analysis allowed us to see how customers felt in their calls and if their experiences varied between subject matter. Understanding this information helps the company improve their customer experience.

Conclusion & Recommendations

Using N-Grams, topic modeling, and sentiment analysis allowed us to understand what customers are calling about and how they feel while talking. The company can use this information to improve customer experience and save money. We also recommend they continue this analysis in further depth and supplement these methods with machine learning and time series.

References

- r-bloggers.com/free-guide-to-text-mining-with-r
- tidytextmining.com/sentiment.html
- cran.r-project.org
- ncbi.nlm.nih.gov/pmc/articles/PMC2841631
- ncbi.nlm.nih.gov/pmc/articles/PMC4597325

Acknowledgments

- Professor Janet Duncan, UCSB
- Professor Xiyue Liao, UCSB
- James Xia, Insurance Company
- Department of Statistics, UCSB