Analysis of change in sentiments towards Chick-fil-A after Dan Cathy’s statement about same sex marriage using SAS® Text Miner and SAS® Sentiment Analysis Studio

Authors: Swati Grover, Jeffin V Jacob, Goutam Chakraborty
Oklahoma State University, Stillwater, USA

ABSTRACT
Social media analysis along with text analytics is playing a very important role in keeping a tab on consumer sentiments. Tweets posted on twitter are one of the best ways to analyze customer’s sentiments following any post-corporate event. Although there are a lot of tweets, only a fraction of them are relevant to a specific business event. This paper demonstrates application of SAS Text Miner and SAS Sentiment Analysis Studio to perform text mining and sentiment analysis on tweets written about Chick-fil-A before and after the company’s president’s statement supporting traditional marriage. We find there is a huge increase in negative sentiments immediately following the company’s president’s statement. We also track and show that the change in sentiment persists through an extended period of time.

INTRODUCTION
President of Chick-fil-A, Dan Cathy, made public comments supporting traditional marriage over same-sex marriage. His remarks went viral on social media giving rise to a big public battle generating strong negative and positive sentiments. By analyzing the data (tweets) we could explore the sentiments of the public towards Chick-fil-A before and after the statement was publicized. We extracted about 10,000 tweets before Dan Cathy’s statement (1-15 July) and about 15,000 after he made the statement (16-31 July). Text mining and sentiment analysis were performed on these pre- and post-tweets. By generating topics from tweets for each time period along with generating concept link diagrams we could spot the differences in sentiments expressed by Twitter users.

To carry the research further and see whether the public sentiments neutralize over time we extract about 5,000 tweets from 1-15 August, 3,000 tweets from 16-31 August and 1,000 tweets from 1-15 September. Again, text mining and sentiment analysis were performed on these post-tweets. The results helped us broaden our understanding about the impact of time on changes in public sentiments following a specific incident.

DATA EXTRACTION AND PROCESS
We extracted tweets from twitter via web scraping using third party resources. Upon close inspection, we found that many of those were re-tweets and we had to remove the duplicates to get a wide variety of topics for better analysis.

We divided the data collection into 5 groups,
Group 1 consisted of about 10,000 Tweets collected before Dan Cathy’s statement on same sex marriage i.e. between July 1-15, 2012.
Group 2 consisted of about 15,000 Tweets collected immediately after Dan Cathy’s statement on same sex marriage i.e. between July 16-31, 2012.
Group 3 consisted of about 5,000 Tweets collected between August 1-15, 2012.
Group 4 consisted of about 3,000 Tweets collected between August 16-31, 2012.
Group 5 consisted of about 1,000 Tweets between September 1-15, 2012.

TEXT MINING
Text mining was initiated by parsing the data to find tokens (terms), parts of speech tags, entities, etc. Special synonym lists, stop lists and dictionary were created and incorporated while performing text parsing. Then, we filtered the data by using the filter node by using Entropy as the term weight and log as the frequency weighting. Finally, we attached a topic node to the filtered data to find out the various topics and the association of terms relating to the topics.

SENTIMENT ANALYSIS
From all the five group data sets used for text mining, we extracted five random samples as modeling data sets with 500 tweets each. Also, five random test data sets were extracted consisting of 200 tweets each for every group. Then, we performed a feature-level sentiment analysis using SAS Sentiment Analysis Studio on those 500 tweets in the modeling data set. A rule based model was built from each data set to find out the sentiment distribution before and after the statement. Different rules were written to specify positive, negative and product descriptors in accordance with the study objectives. We started by taking Chick-fil-A as a product and then Dan Cathy as a feature. Finally, all the models were tested against the test data to see to what extent Dan Cathy’s statement played a role in overall sentiment change towards Chick-fil-A and how these sentiments were carried over the course of time.

TEXT MINING RESULTS
The topics formed in each group of data were used for comparing the results. A synonym list was initially drafted based on the first group. This was further modified for the other four groups of data. A single synonym list was eventually used for all the groups.

Custom stop lists were used for each group because each group had their own set of terms that had to be removed. Terms that were less frequent were removed to prevent them from dominating topic formation. The descriptive terms identified in each topic were used for understanding the topics. The topics formed in all the groups along with their descriptive terms are shown below.

SENTIMENT ANALYSIS RESULTS
Impact of Dan Cathy’s Statement
Before Dan Cathy’s statement, the overall distribution of sentiments towards Chick-fil-A was mostly neutral and positive. 35% of the people were writing positive comments about Chick-fil-A while only 8% were negative and the rest being neutral. People were either neutral about Dan Cathy or discussing about him in a positive way.

There was a huge change in sentiments towards Chick-fil-A, as soon as Dan Cathy made his comments about same-sex marriage. After the incidence, only 16.5% of people were talking positively about Chick-fil-A whereas 20% were negative and the rest being neutral. 62.8% of people talked negatively about Dan Cathy and that in turn likely impacted the overall negative sentiment for Chick-fil-A.
IMPACT OF TIME ON PUBLIC SENTIMENTS

Our further analysis to understand the effect of time on change in public sentiments showed that people tend to forget things and their strong negative emotions can neutralize over time though not completely turning to positive again. Our analysis showed that distribution of sentiments observed for Group 5 was more neutral and less positive or negative than what was observed for Group 1 i.e., before Dan Cathy made the statement. For Group 5, only 8.5% were talking positively about Chick-fil-A, while 6% negatively and rest were neutral. For Group 3, 15.5% of people wrote positive comments, 13% negative and rest neutral. Similarly for Group 4, 12% of people were writing positively, 11% negatively and rest were neutral.

Group 3: Sentiment Distribution for tweets from August 1-15 in modeling and test data

Group 4: Sentiment Distribution for tweets from August 15-31 in modeling and test data

Group 5: Sentiment Distribution for tweets from September 1-15 in modeling and test data

OVERALL DISTRIBUTION OF SENTIMENTS:

During the course of study, the increased percentage of negative sentiments caused by Dan Cathy’s statement gradually decreased. On the other hand, positive sentiment towards Chick-fil-A also decreased significantly over the time. There was a continuous increase in neutral sentiments which may be the result of people with positive opinions about Chick-fil-A turning to neutral. So, even after two months of Dan Cathy’s statement, Chick-fil-A suffered from its impact and sentiments it generated in public and as the trend showed it might take people a couple of more months to turn to positive from neutral.

CONCLUSION

In our analysis, we find an immediate change in sentiments towards Chick-fil-A after Dan Cathy’s statement and how those sentiments stabilized over the course of time. Apart from change in the topics, the general tone of tweets first shifted from positive to negative and then changed to neutral with gradual passage of time. Text Miner helped us identifying the key topics associated with text. This was taken to next level by performing sentiment analysis for a deeper analysis. We were able to quantify the sentiments in terms of change in percentage of sentiment distribution associated with Dan Cathy impacting Chick-fil-A’s image. Our study has many limitations including convenient sampling of tweets. Notwithstanding those limitations, our study results clearly show how text mining and sentiment mining can be used to measure and track the impact of any post-corporate event by analyzing tweets.

Grover, Jacob, Chakraborty