https://doi.org/10.48009/4_iis_2019_123-129

Issues in Information Systems
Volume 20, Issue 4, pp. 123-129, 2019

The U.S. - MEXICO BORDER WALL - A SENTIMENT ANALYSIS

Jeffrey Easter, Missouri University of Science and Technology, jpex29@mst.edu
Qiuyue Yang, Missouri University of Science and Technology, qyhv7@mst.edu
Choji Daches, Missouri University of Science and Technology, cbdfrs@mst.edu
Wen-Bin Yu, Missouri University of Science and Technology, yuwen@mst.edu
Craig Claybaugh, Missouri University of Science and Technology, claybaughc@mst.edu

ABSTRACT

The central focus of our study is to apply a computer-base sentiment analysis on twitter data on the U.S.-Mexico border wall. we want to be able to validate the results from traditional polling method and also create a pointer for decision making. From our results, we were able to show that a computer-base sentiment analysis is as valid as a traditional base method on topics such as the considered for the study. Our results compared favorably with the results obtained by organizations who did a polling survey on the issue. Also, we were able to visualize our results to present a clear picture of the sentiments trend.

Keywords: Sentiment Analysis, Twitter, Text mining

INTRODUCTION

The border line between the U.S. and Mexico stretches roughly 2,000 miles long and traverses four states, from California to Texas. The U.S.-Mexico border wall has been a long-standing government agenda. The border wall was created in 1848 following a treaty of Guadalupe Hidalgo which was barbed wired to prevent Mexican cattle from grazing on American crops. Currently, there is nearly 653-mile of border wall and Presidents Clinton, Bush and Obama are responsible for 97% of the present wall. Clinton began building the wall raising about 34 miles of it. Bush beginning in 2006, built nearly 465 miles of the wall during his time as president. Obama continued the construction building 133 miles. 37 miles of secondary fencing was added by Clinton, Bush, and Obama together (Guerrero & Castaneda 2017).

But why is Trump wall trending creating a lot of sentiments? The Trump proposal comes with a different twist “and I will have Mexico pay for the wall.” But the question of who will pay for the wall, still remains a contention in parliament and the nation at large. A wall along the U.S.-Mexico border will surely cost an arm and a leg. Mitch McConnell said it will cost an estimated $15 billion to build the Wall. If Mexico is not paying for the Wall, as has been refuted by the Mexican president, then American taxpayers would fund the building of the Wall if Trump remains obstinate about the Wall. And this is generating a torrent of feud across board. Eight prototypes of what Trump’s wall might look like were completed in San Diego in October 2017. They tower about 30 feet high — three times taller than the old primary fencing at the prototype site (Guerrero & Castaneda 2017).

The perceived objective of building a wall along the border is to curb the prevalent menace of illegal immigration and drug trafficking. The Trump declaration to build the “great wall”, has generated an avalanche of sentiments on both news and social media. Many polls have been set up by different organizations to gather the opinions and thoughts of Americans on the U.S.-Mexico border wall. Polling by Quinnipiac conducted on April 5, 2017 showed 33% support for "building a wall on the border with Mexico." A CBS News poll in January 2017 found 37% support for "… building a wall along the U.S.-Mexico border to try to stop illegal immigration." And `Pew Research Center in February 2017 found 35% support for "building a wall along the entire border with Mexico."

The “great wall” even though has not been built on land it has divided the American politico. The democrats and the republicans have been at loggerheads over the building of the Wall. Debates over funding of the Wall has remained an adamant issue in parliament. The “great wall” has even threatened government shutdown if it is not funded.
The sentiments shared among Americans is what we will like to harness using sentiment analysis to tell our story of the ‘Great Wall’.

LITERATURE REVIEW

According to dictionary.com sentiment analysis also known as opinion mining, or emotion AI is the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc., is positive, negative, or neutral. Sentiment analysis employs natural language processing (NLP) and text analysis to extract subjective information. Sentiment analysis is applied to text mined from web pages in order to discover trends and new information based on thoughts and opinions. Acquiring text from web pages in order to extract new and useful information is known as text mining. Pang and Lee (2008) explained that the sudden eruption of activity in the area of opinion mining and sentiment analysis, has occurred at least in part as a response to the surge of interest in new systems that deal directly with opinions as a first-class object. The authors mentioned that Opinion mining and sentiment analysis covers techniques and approaches that promise to directly enable opinion-oriented information-seeking systems.

The first approach to sentiment analysis is data gathering or information retrieval. Different methods have been applied depending on the type of study to collect data over different sources. For our project, we will be utilizing the web scraping technique to collect data on Twitter web site using the Python programming language and using SAS text miner to perform sentiment analysis on the data. Twitter which allows people to express their opinions in succinct sentences has become one of the most popular social networks for microblogging because of its wide public patronage (Guevara et al 2018). According to Hridoy et al (2015), social network data is one of the most effective and accurate indicators of public sentiment. In their work, the authors analyzed a methodology that allows for the utilization and interpretation of twitter data to ascertain public opinions.

Web scraping is a process of extracting data from websites. A web scraper (computer program) accesses web pages, finds specific data elements on the page, extracts the data, transforms and finally saves the data as a structured data (Boeing and Waddell 2016). Boeing and Waddell (2016) applied the technique of web scraping in extracting rental housing data on Craigslist. The authors built a web scraper to collect 11 million rental listings from Craigslist website, using the Python programming language and the web scraping framework to investigate spatial and temporal patterns in the rental housing market.

The importance of sentiment analysis is underscored by its application to different subjects whether it is product based, service based, political, government, personality, etc., to make deterministic and futuristic decisions. Tumasjan et al (2010) applied sentiment analysis on twitter data to predict election outcome. In their study, the authors used the context of the 2009 German federal election to investigate whether Twitter can be used as a forum for political deliberation and whether online messages on Twitter validly mirror offline political sentiment. They were able to show from analyzing over 100,000 messages containing a reference to either a political party or a politician that Twitter perhaps can be used profoundly for political deliberations based on the sentiments in the messages.

As at the time of this study, and to the best of our knowledge, there is no literature on a computer base analysis of people's sentiments on the U.S.-Mexico border wall. This makes our work a novel piece which can further drive research into mundane domains that may otherwise be impactful in decision making if the sentiments are well analyzed.

RESEARCH METHODOLOGY

For our project we intended to measure public sentiment of president Trump's proposal to build a border wall between the United States and Mexico. For a sense of timeliness, our dataset was finalized over the period November 19 to 23, 2018.

Software Tools Used

This research project was conducted utilizing Python3, Microsoft Excel, SAS Enterprise Miner Workstation 12.1, and SAS 9.3.

Purpose of Research

The intent for the project was: (1) To collect relevant data from a topical source which would provide a dataset with a wide variety of opinions. (2) To apply classification rules to the dataset for analyzation. (3) Based on this classification scheme, be able to define a general sentiment on the topic based on the collected dataset.

Data Collection

To begin the project, a suitable datasource was necessary. Since our project deals with sentiment analysis, social media was chosen as the platform to use because it provides the broadest array of public opinion on any topic. Twitter.com was finally selected due to its ubiquitous nature, sense of immediacy and supply of most current public opinion.

The next step was to compile a list of hashtags and search terms. Initially the list provided an overabundance of information. We would need to pare down our list in order to make forward movement on this project. Making use of a site called Ritetag.com, allowed us to find relevant hashtags and other search terms and decide what was trending with our subject matter and gave us the terms which would return the greatest amount of tweets.
Our final list was composed of the search terms: borderwall, buildthewall, immigrantcaravan, mexicanimmigration, mexicancaravan, mexicanwall, trumpcaravan, and trumpwall.

Data Preparation

After collecting our dataset, several processes were ran against our corpus in order to prepare for analysis. The first process was to remove any non alphanumeric characters from each tweet. This was followed by removing any tweets marked as retweets or duplicates. The next process was to apply a sentiment to each tweet. Overall, due to the collection limits set by Twitter, our total corpus resulted in 222 tweets. In applying sentiment, three groups were created: P – positive, N – negative, and D – do not care (neutral). Both programs for data cleansing and application of sentiment were custom written for this project and written in Python3. The program which applied sentiment utilized a NLP algorithm called TextBlob found here: https://textblob.readthedocs.io/en/dev/index.html.

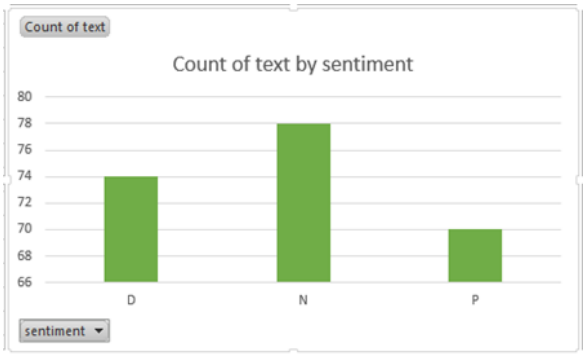


Figure 1. Comparison of sentiments (P: Positive, N: Negative, D: Neutral or "Don't care")

After running the sentiment program over the dataset to see how well the algorithm worked, we recorded the totals for each group as can be seen in figure 1.

After looking at some of the tweets to verify classification, we found examples such as in figure 2. In this case we agreed that it was, indeed, correctly marked as a negative against the wall.

131	N	The MexicanCaravan is humanitarian crisis its being treated like an invasion

Figure 2. Example of tweet with category applied

Having a smaller dataset to work with allowed us to define our training set with 50 tweets and the remaining 172 tweets became the testing set. Prior to running the supervised tests, the sentiments were removed from the testing set.

Model Preparation

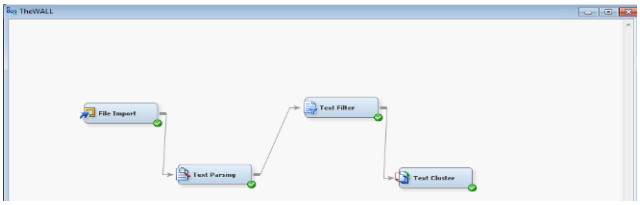


Figure 3. The SAS Diagram for Unsupervised processing over the full dataset.

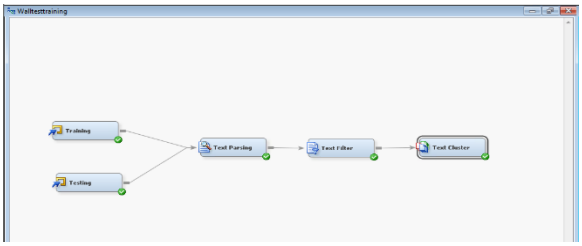


Figure 4. The SAS diagram for Supervised processing.

Training node holds the subset of data with sentiment applied and the Testing node holds the remainder of the dataset with no sentiment applied. The Training dataset is used as a guide for SAS when processing the Testing dataset which allows SAS to "learn" how to process the Testing data and group each tweet in the correct cluster.

Cluster Analysis

Neither processing model provided a clear and distinct result. However, the Supervised model did provide a better view of the dataset.

Beginning with the Unsupervised model results below, our default cluster size was Exactly 3.

Cluster	
Exact or Maximum Number	Exact
Number of Clusters	3
Cluster Algorithm	Expectation-Maximization
Descriptive Terms	5

Figure 5. Defined Cluster size of exactly 3.

Frequency	Percentage
69	31%
49	22%
104	47%

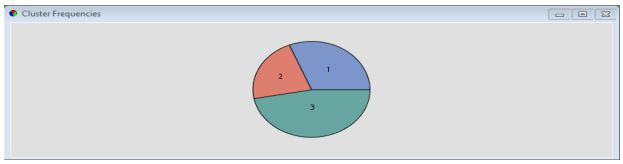


Figure 6. The percentage breakdown of each of the initial clusters.

Cluster	
Exact or Maximum Number	Exact
Number of Clusters	10
Cluster Algorithm	Expectation-Maximization
Descriptive Terms	5

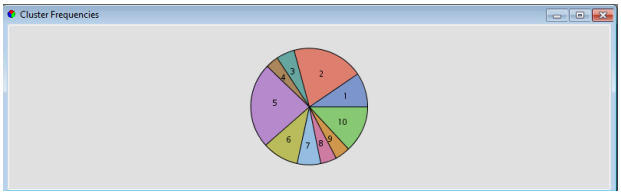


Figure 7. Exact clusters of10.

But again the percentages for each cluster did not reveal any leading distinctions.

Frequency	Percentage
21	9%
44	20%
11	5%
8	4%
53	24%
22	10%
15	7%
10	5%
9	4%
29	13%

Figure 8. 10 Cluster percentages.

Below are the terms per cluster for 3 vs. 10 clusters

Cluster ID	Descriptive Terms	
1	borderwall mexicanwall buildthewall america realdonaldtrump	...
2	+'migrant caravan' caravan migrant +trump +migrant	...
3	'border wall' +wall border trump +border	...

Figure 9. Descriptive terms for original 3 clusters.

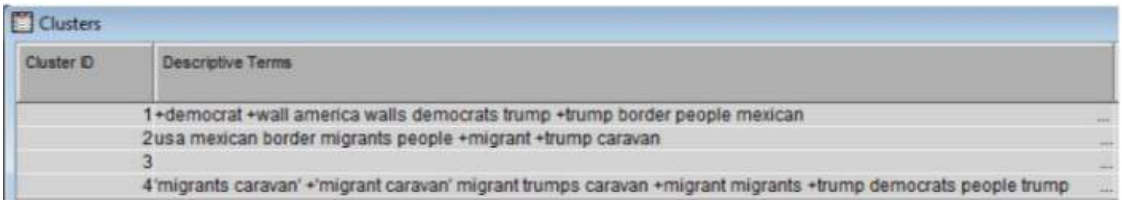
Cluster ID	Descriptive Terms	
1	+people +country illegally buildthewall day	...
2	+'migrant caravan' caravan migrant +trump entry	...
3	borderwall democrats +democrat +fund funding	...
4	+dollar +mexican migration illegal +election	...
5	'border wall' +keep +wall walls border	...
6	+border borders 'mexican border' +want +happen	...
7	+worry lies mexicancaravan invasion mexicanwall	...
8	mexico asylum women thousands +want	...
9	money usa fox asylum wall	...
10	today wall building mexican government	...

Figure 10. Descriptive terms for 10 clusters.

Concluding the Unsupervised processing, no changes, in addition to Term and Frequency weights or SVD Resolution and Max Dimensions had any effect on the results.

Using the Supervised model, again, did not provide a clear distinction, however, it did produce interesting results.

Changing the cluster size to 40, did produce a cluster with no descriptive terms as shown below.


Clusters

Cluster ID	Descriptive Terms
1	+democrat +wall america walls democrats trump +trump border people mexican
2	usa mexican border migrants people +migrant +trump caravan
3	
4	'migrants caravan' +'migrant caravan' migrant trumps caravan +migrant migrants +trump democrats people trump

Figure 11. Cluster showing anomalies.

Upon review of the cluster contents utilizing SAS 9.3, we obtained the following sample tweets which, in fact, did not include any of the original search terms.

31	worry A lot rhetoric that politicians indulge in prior elections is well mere rhetoric get elected
32	Asylum is legal for them in Mexico when they are actually in need asylum

Examples of cluster anomalies.

CONCLUSIONS AND FUTURE WORK

Conclusion one is that the number of Neutral (D) tweets were just enough to prevent the Positive or Negative from standing out over the other.

Two is that, based on the topic, although a tweet might have been categorized as Positive or Negative, there was not enough sentiment in the overall cluster to provide a distinctive result.

Going forward, additional research which could impact results would be:
(1) Collect small groupings of tweets over a longer time period. This could allow sentiment to be developed based on timely events related to the topic in the time period or allow for trending analysis of new keywords within the dataset to provide more search terms or generate new clusters.
(2) Expand the dataset to include more input from other social media outlets or news and journal articles..

REFERENCES

Albright, R. Taming Text with the SVD. SAS Institute Inc., Cary, NC. January, 2004.

Boeing, G. and Waddell, P. New insights into rental housing markets across the United States: Web scraping and analyzing craigslist rental listings. *Journal of Planning Education and Research, 37*(4), 457-476, 2017.

Guerrero, J., & Castaneda, L. Decades-long struggle to secure U.S.-Mexico border, 2017. http://border.inewsource.org

Guevara, J., Costa, J., Arroba, J., & Silva, C. Harvesting opinions in Twitter for sentiment analysis. In 2018 13th Iberian Conference on Information Systems and Technologies (CISTI) (pp. 1-7). IEEE, 2018.

Howland, P., & Park, H. Generalizing discriminant analysis using the generalized singular value decomposition. *IEEE transactions on pattern analysis and machine intelligence, 26*(8), 995-1006, 2004.

Hridoy, S. A. A., Ekram, M. T., Islam, M. S., Ahmed, F., & Rahman, R. M. (2015). Localized twitter opinion mining using sentiment analysis. *Decision Analytics, 2*(1), 8.

Kadhim, A. I., Cheah, Y. N., & Ahamed, N. H. Text document preprocessing and dimension reduction techniques for text document clustering. In Artificial Intelligence with Applications in Engineering and Technology (ICAIET), 2014 4th International Conference on (pp. 69-73). IEEE, 2014.

Kang, M., Ahn, J., & Lee, K. Opinion mining using ensemble text hidden Markov models for text classification. *Expert Systems with Applications, 94,* 218-227, 2018.

Lee, J. H., Jung, S. H., & Park, J. The role of entropy of review text sentiments on online WOM and movie box office sales. *Electronic Commerce Research and Applications, 22*, 42-52, 2017.

Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., & Ngo, D. C. L. Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment. *Expert Systems with Applications, 42*(1), 306-324, 2015.

Newport, F. Building a Wall Out of Sync with American Public Opinion, 2017. http://news.gullup.com

Pal, J. K., & Saha, A. (2010). Identifying themes in social media and detecting sentiments. In Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on (pp. 452-457). IEEE.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval, 2*(1–2), 1-135.

Roy, K., Kohli, D., Kumar, R. K. S., Sahgal, R., & Yu, W. B. Sentiment Analysis of Twitter Data for Demonetization In India–A Text Mining Approach. *Issues in Information Systems, 18*(4), 2017.

Suls, R. Most Americans continue to oppose U.S. border wall, doubt Mexico would pay for it, 2017. http://pewresearch.org

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. Predicting elections with twitter: What 140 characters reveal about political sentiment. *Icwsm, 10*(1), 178-185, 2010.