2016

Sentiment Expression on Twitter Regarding the Middle East

Byron Boots
byron.boots@umontana.edu

Sentiment Expression on Twitter Regarding the Middle East

Byron Boots
University of Montana Senior
Graduating Spring 2016
Management Information Systems & International Business
Accounting Information Systems Certificate

Abstract

Social media has transformed the awareness of events around the world as it allows for instant, up-to-the-second data transmission and communication for a variety of interested parties. Due to the ongoing turmoil surrounding the Middle East and its heightened media attention, I chose to research what types of emotions and interactions are found on Twitter with regard to the region and related topics. I selected Twitter due to the relative accessibility, workability, and anticipated sufficient size of data samples available. Twitter reports 1 billion created accounts with 320 million active accounts as of December 31, 2015. These active accounts, defined as a ratio of followers to followed accounts, generate roughly 500 million tweets per day from around the world. In this research, I am looking at scholarly works, journalism sources, and other reports to learn more about some of the ways Twitter has been used as it relates to the Middle East and better establish context for my data analysis. This information helps guide me in performing real-time sentiment analysis – or opinion mining – on Twitter data using open-source sentiment dictionaries with machine learning algorithms to provide highly accurate analysis of emotional response as it relates to the Middle East. This sentiment analysis is performed by assigning numerical values to words to help quantify positive, neutral, or negative emotion associated with my topic. My findings will help to draw conclusions as to whether there are specific emotions correlated with the region and associated topics, the degrees of emotion felt when tweeting about specific subjects, and how spot-checked dates after different events influence the sentiment broadcast on Twitter. This unfiltered look at people’s emotions on Twitter serves to quantify how Twitter users perceive the Middle East and related topics.

I. Introduction

I.I What did I want to know?

The research that I completed in this project began with determining my specific interest in the Middle East and ultimately deciding that I was most interested in discovering if a difference existed between popular media's portrayal of the Middle East and related topics and whether society viewed the region positively or negatively. I decided to focus on answering the question: *Do the statements on Twitter discussing the Middle East and related topics have an emotional trend, and if so, is that trend positive or negative?*

I.II Why this topic?

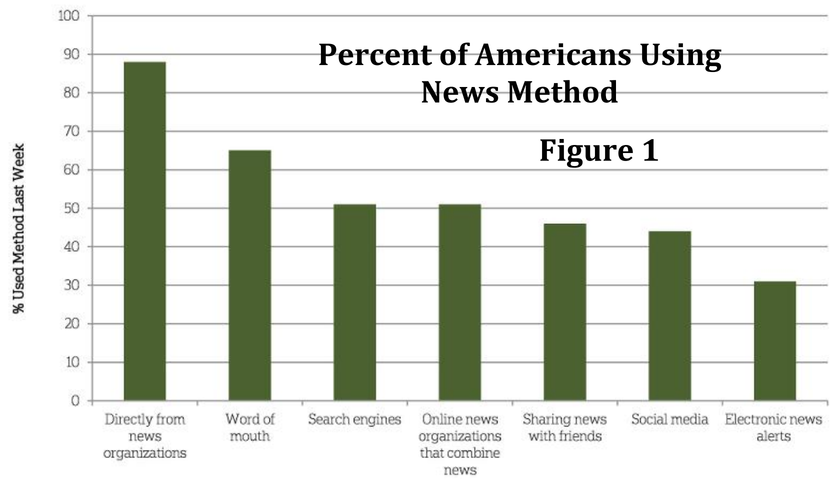
The Middle East has been involved in a number of hot button issues throughout history and therefore narrowing my scope of research to a certain time frame and specific geographic area was a difficult task. Countless stakeholders are involved in the region and issues such as the refugee migration have been influenced by global politics and affected global politics. For this reason, I chose only to look at the Middle East and its related topics beginning after December 2010, the beginning of the Arab Spring, and evaluate emotion on a real-time basis off of the social media platform, Twitter.

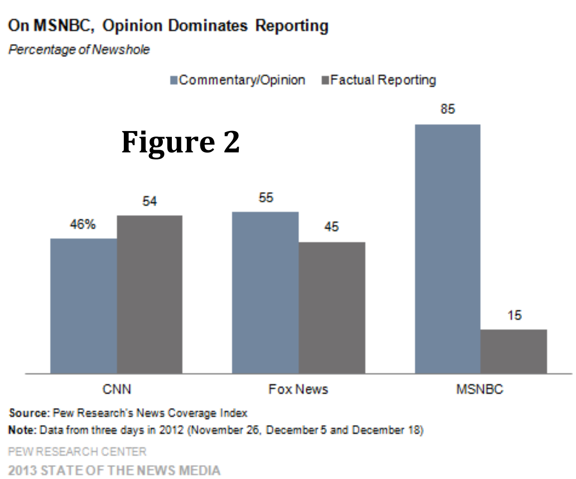
I.III Why Twitter?

Twitter was chosen due to its accessibility of information and the relative ease of analyzing this information using platforms I was already familiar with and had access to for reliable usage. On a daily basis, Twitter sees over 500 million tweets, or roughly 7,200 tweets per second, equating to a more-than-adequate data source in terms of size. Additionally, Twitter is a global application with 330 million active users globally, meaning that when I collected tweets I would be looking at not only at American, but also global viewpoints.

I.V Who cares?

According to the American Press Institute, 87 percent of Americans receive news from television and 88 percent say that they receive news directly from news organizations. These statistics are shown in Figure 1 with percent of Americans on the vertical axis and method of news reception on the horizontal axis. Further research shows that globally these statistics remain relatively stable, however the greatest research on the topic has been conducted on the American public. The prevalence of news organizations as Americans' primary news source translates to a huge amount of influence being given to established news organizations, especially those with television channels, over what the public receives regarding news. In my own experience and perception, the news broadcast by these organizations is almost entirely negative in the viewpoint and language chosen to discuss topics. This perception is confirmed by the Pew Research Center's Project for Excellence in Journalism where a study in 2012 concluded that almost all major cable news stations devote more air-time



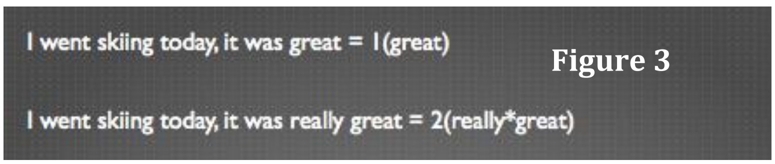


to opinions than factual reporting by a large margin and that those opinions are generally negative. I was interested in finding out whether or not the general public viewed and discussed these topics differently from how established news organizations broadcast and if they did, what that difference was. If there truly is a difference, then it shows not only a bias from a media perspective but also that the general public is not so negative as one might perceive from the media.

II. Methods

II.I Sentiment Analysis

Sentiment analysis, or opinion mining, is defined by Oxford Dictionaries as: *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.* Originally used by businesses, the foundations of sentiment analysis rest in product reviews for online retailers where reviewing customers could choose to rate a product on a set scale and write a review about the product. Analysts found that when products had a large number of reviews, it was inefficient and expensive to task an employee with reading through each review to determine whether any of them held any information that could be useful to the company. Worth noting, is that word-of-mouth review is consistently rated as the most influential component in purchasing decisions, so negative and positive reviews play a major role in purchase decisions. Therefore businesses created a “thumbs-up” capability where customers could rate reviews of a product based on how useful they found the review. The introduction of the ability to rate review usefulness allowed businesses to determine which reviews were worth reading and improved product development to satisfy customers better and fix faults in products. The downside to this method, however, was that a negative review stating a flaw in a product might not be flagged as important and useful until after more consumers had purchased the product and become dissatisfied, resulting in future lost business for the online retailers. Furthermore, many dissatisfied consumers never returned to the retailer to write a review but simply blogged about the item or ranted about its downsides on another platform so that the business could never hear about its product issues other than in the lost sales experienced. The need arose for a method to determine what consumer perception of a product was prior to fellow consumers purchasing the product, or not, based on the word-of-mouth from another consumer.



Sentiment analysis as an answer to this need was developed as computing technology and access to data improved. Opinion mining essentially involves defining a scale, say -5 (negative sentiment) to 5 (positive sentiment) with 0 being neutral sentiment, assigning numerical values to every word, and then having a computer "read" a sentence and average or sum the numerical values of the words to determine an overall sentiment for a statement. An example would be the statement: *I went skiing today; it was great.* For simplicity purposes, we assign a numerical value of 0 to all words in the statement except *great,* which receives a value of 1. When summed, this statement scores 1 and therefore receives a plot on a chart of 1, slightly positive. Now take the statement: *I went skiing today; it was really great.* If we again assign a numerical value to all terms *I went skiing today, it was*, but then assign a multiplier of 2 to *really* and our established value of 1 to *great*, then our score for the statement becomes 2, more positive than our prior statement. Roadblocks do exist for this work as human language and slang play a role in the effectiveness of the analysis. An example of a potentially problematic statement for sentiment analysis would be the statement: *I went skiing today; it was sick.* Once again, the words *I went skiing today, it was* all receive a value of 0, but the term *sick* could have a positive or negative meaning depending on context. Sarcasm, slang, and human spelling errors all can have a large impact on the effectiveness of sentiment analysis, something that many analysts deal with on a daily basis.

II.II Machine Learning

Due to the tendency of human beings not to operate in computer language, program developers began searching for a way to mitigate the risk of improper analysis, and as computing power improved, computer scientists were able to search for patterns in text with computers and develop machine learning. As defined by WhatIs: *Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can teach themselves to grow and change when exposed to new data.* With regards to sentiment analysis, machine learning allows the computer to "learn" what complex sentences and phrases hold for sentiment by recognizing text patterns and improve the accuracy of sentiment scores without human intervention.

II.III Visual Studio Streams and Sentiment140

To collect Twitter data, the first method used for pulling tweets was through the program Visual Studio that sent out a call for tweets every 5 seconds based on specific keywords and pulled up to 150 tweets at a time from the Twitter Application Program Interface (API) to be sentiment scored and cached into a data file. For this method of sentiment analysis, I used the sentiment dictionary Sentiment140 due to budget constraints, as it is a free, machine learning sentiment dictionary to score words and evaluate the average sentiment for a tweet on a scale of 0 through

4 with 0 being negative, 2 being neutral, and 4 being positive. After being scored, these tweets were sent to another application, Microsoft Azure Streaming Analytics to be further filtered and more easily manipulated into a variety of applications.

II.IV Microsoft Azure and Power Bi

After receiving the scored tweets from Visual Studio, Microsoft Azure would sort the tweets and cache them in a JSON file for further analysis and storage. The tweets would simultaneously be sent to Power Bi, a visualization application where I could watch tweets stream and their score being plotted over time.

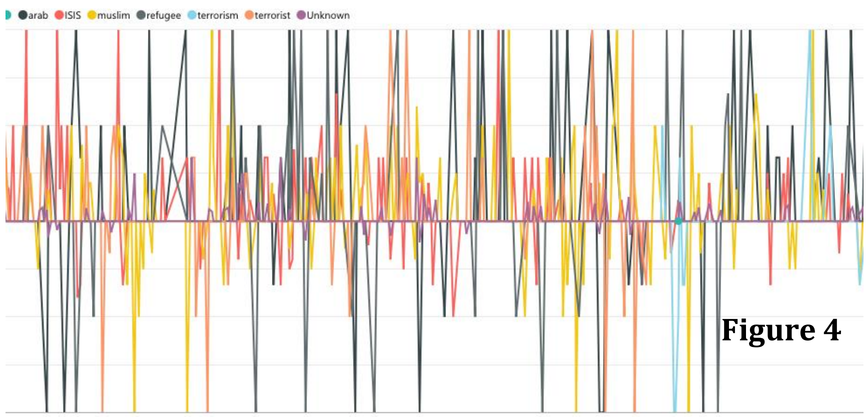


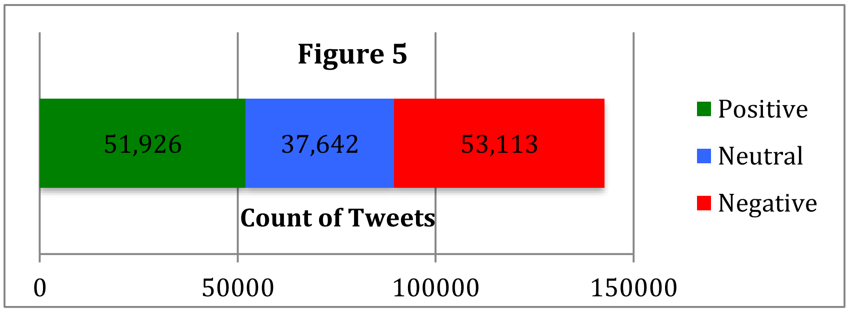
Figure 4

II.V Python (Twython Package)

As a check, I chose to run another analysis on tweets from another source to ensure that the sentiment scores I was receiving were valid and not being scored incorrectly. To perform this check I utilized the Twython package in Python to pull tweets and again attached the Sentiment140 dictionary as my sentiment-scoring dictionary. These tweets were also stored in JSON format on my own computer.

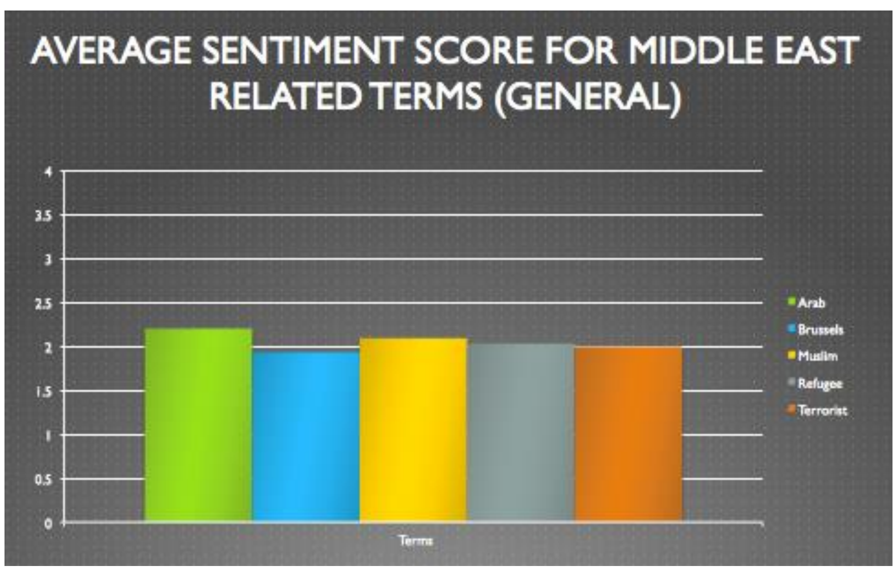
III. Findings

III.I Key Terms

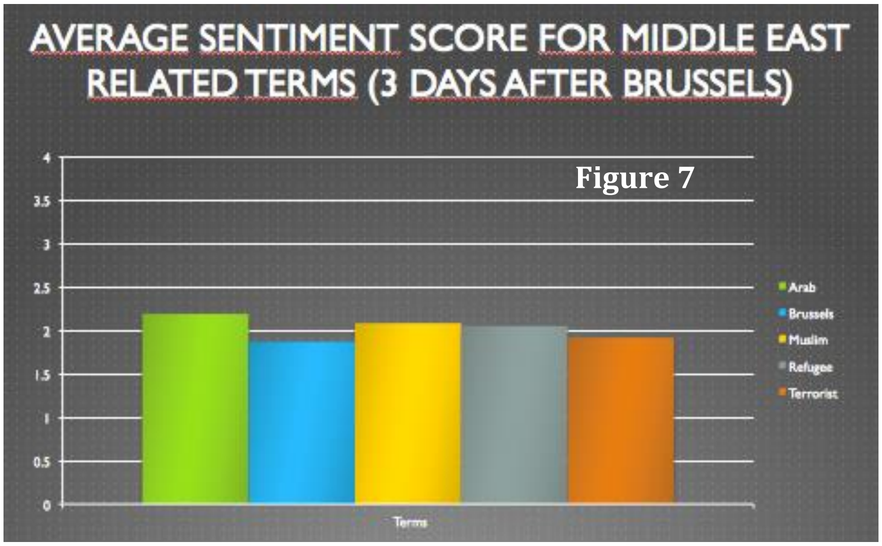


Though sentiment analysis was run on a number of keywords, I chose to focus on five based on the volume of tweets I received and relevance to my research question. The five terms shown in my findings are *Arab, Brussels, Muslim, Refugee, and Terrorist.* I felt that I collected enough Twitter data containing each of these key terms in the text of the tweet and that the terms strongly related to my research question and personal interest in this topic to draw conclusions based off of data (with these key terms) that is not random. Additionally, due to time constraints, I was forced to reduce my scope from my original intentions of closely analyzing sentiment for more key terms as I felt that fewer closely analyzed terms were more relevant than numerous minimally checked and analyzed terms.

III.II Polarized Sentiment



After collecting and analyzing the tweets I had received from both sources, it became clear that sentiment surrounding the Middle East and related topics was not so one-sided as popular media portrays. The Twitter sentiment scores showed that the Middle East is a polarizing topic with the majority of tweets holding either a strong positive or strong negative sentiment with neutral tweets existing but not in the same volume as tweets with strong sentiment in either direction. When doing a count of positive, negative and neutral tweets, it became clear that the Middle East and related topics receive almost the same number of tweets with positive feedback as negative feedback, and that the people on Twitter discussing those topics feel strongly in one direction or another rather than holding a neutral viewpoint about the issues. As is shown in Figure 5 with green being positive sentiment, blue being neutral sentiment, and red being negative sentiment, 51,926 tweets contained positive sentiment and 53,113 tweets contained negative sentiment. Comparatively, only 37,642 tweets contained neutral sentiment proving that these issues are polarizing, but not necessarily in one direction. Of the 142.681 total tweets, 36.4 percent of all tweets were positive and 37.2 percent were negative, a negligible difference.



III.II Polarization = Neutral Average

The implications of such polarized viewpoints make for a neutral average on the key terms that I chose to look at when running my analysis. Over an approximate one-month timespan, I ran my streams to collect Twitter data and process it, with the 2016 Brussels terrorist attacks taking place in the middle of my timeframe. For this reason I chose to split my data into a “General” file of almost all of the data except for a “Brussels” file composed of the 3 days after and including the Brussels attacks. As can be seen when comparing the two charts, there was almost no difference between the average sentiment in the “General” file and the timeframe just after the Brussels attacks, with all five key terms remaining relatively neutral.

III.III Interesting Finding

Something that I did notice but did not have the chance to look into further was the prevalence of the terms Hillary, Trump, Nazi, and Holocaust all appearing together frequently throughout the Twitter data. It was unclear whether or not these tweets contained positive or negative sentiment; however, the terms generally were found in reply or re-tweeted tweets. I did not have time in my research to conduct further analysis of these terms and their context.

IIII. Conclusion

IIII.I Takeaways

Through my research, I gained a surface-level understanding of Twitter and in-depth knowledge of sentiment analysis. With further resources I believe that I could draw firm conclusions as to the causes behind the polarized but neutral average sentiment that I found, as well as look deeper into the interesting text I found in tweets. Through the work that I did, however, I feel confident in concluding that popular media broadcasts a generally opinionated negative sentiment in their portrayal of issues that does not align with the public sentiment as found on Twitter. This is backed up by the Pew Research Center’s Project for Excellence in Journalism. I found that the views on Twitter represent the entire spectrum of emotions relating to the Middle East and that those emotions were equally distributed positively and negatively, contrary to media sentiment. Perception of the media skews largely negative as is verified by a Political Science study out of

UCLA, something that is not found on Twitter. Although negative viewpoints on all issues in the news certainly do exist, positive viewpoints are just as prevalent when taking in data from the public and not just major media sources.

IIII.II Errors and Limitations

The sentiment analysis process has countless limitations, something that makes drawing serious conclusions without domain expertise extremely dangerous as it could convey a false message accidentally. Potential areas for error and limitations in this project were:

- Human error by the researcher in the developing of the programs and interpreting of the data.
- Human errors by Twitter users in their writing of tweets (misspellings, strong sarcasm, lying).
- Size of data makes accuracy checks with regard to sentiment score very difficult.

IIII.III Moving Forward

To build upon the research completed already, I would recommend several next steps to understand better and answer this question. Further research I feel would be beneficial would be:

- Removing news media sites from the Twitter streams analyzed to determine what strictly public viewpoints were without media involvement.
- Analyzing of Twitter streams using expensive sentiment dictionaries and algorithms rather than the open source tools that performed this research.
- Separation of tweets by geographic location to determine if sentiment changes depending on location.
- Longer testing periods to collect greater time samples and overall a greater sample size.

- Increased budget for purchasing specialty analytics software, sentiment dictionaries, and conducting research using the entire Twitter API rather than a sample.