Proceedings of the 2020 Annual General Donald R. Keith Memorial Capstone Conference
West Point, New York, USA
April 30, 2020
A Regional Conference of the Society for Industrial and Systems Engineering

Using Media Benchmarking to Measure Social Media Activity at the United States Military Academy

Sara Klena

Department of Systems Engineering
United States Military Academy, West Point, NY

Corresponding Author: sara.klena@westpoint.edu

Author Note: Sara Klena is a senior at the United States Military Academy at West Point and conducted this project as part of an independent research study with the Department of Systems Engineering's Honors Program. She worked under her advisor, MAJ John Caddell, who provided tremendous insight into the current relationship between social media benchmarking and the United States Military Academy. He also coordinated meetings with the USMA G5 Office, a stakeholder (along with the Public Affairs Office) for this research project. The author would like to thank MAJ Caddell, the Department of Systems Engineering, and MAJ Andrew Bond for their support and assistance.

Abstract: As social media and digital content continue to grow, the United States Military Academy (USMA) finds itself looking for a method to measure and assess its actions and presence in these spheres. This work attempts to quantify and assess USMA's digital presence by examining the relative quantity of online queries and sentiment of relevant news articles. By benchmarking key incidents with this method, the Academy can better assess the relative scale and tone of significant events in the future. This work benchmarked significant events involving the Academy within the current academic year to demonstrate the differing scale and sentiment of each event and relating these results to one another. We conclude that if USMA considers benchmarked data and a range of search terms relating to the Academy it can be better equipped to handle future significant events in the form of an emergency operations center and the daily analysis of social media interactions.

Keywords: benchmarking, Google Trends, natural language processing, Emergency Operating Center

1. Background

In the recent past, the United States Military Academy has experienced a range of both positive and negative attention for various events. Most negative "hot topics" received national news coverage, like the 2017 publication of LTC(R) Heffington entitled "Open Letter to Grads" and training accident resulting in the death of a Cadet during 2019 Cadet Summer Training. Positive "hot topics" also gain legitimate national coverage, like the 2018 visit between celebrity Reese Witherspoon and USMA First Captain 1LT Simone Askew. Additionally, Army football games gain significant media attention, particularly when they involve opponents with large internet followings like the University of Michigan, Oklahoma University, and Ohio State University. Sandhurst competitions at USMA also receive attention as their competitors travel internationally and represent their institutions and countries while at the Academy.

With each of the negative "hot topic" events, the G5 Office and Public Affairs Office (PAO) need a mechanism to describe the relative size and tone of a situation. This mechanism will help with triage, providing context for a response from key leadership, and to track reactions to any USMA messaging. The G5 Office supplies concise summaries on relevant incidents so key leadership may respond appropriately. Scale and sentiment are important metrics to allow the G5 to see how much a topic is being discussed and the tone in which it is discussed. Both scale and sentiment must be analyzed over a period of time to monitor continual public reaction both before and after official USMA response. The PAO manages USMA's social media, issuing official statements and monitoring the frequency of views, comments, and overall tone of responses. Although this overview of interaction is readily available to USMA, analytical capabilities are not extensive. Improving analysis will serve a two-fold utility to USMA personnel. First, it will allow for daily response monitoring even in the absence of a significant "hot topic" event. Second, this improvement will provide USMA with vital information for the Emergency Operations Center (EOC). In the event of a serious problem, G5 personnel can turn to their analytic tools to promptly determine the best way to respond to a situation. With events like training accidents, missing Cadets or serious injuries, time is of utmost importance in allowing key leadership to deal with a crisis in the appropriate way. With accurate and timely information gathered by data analysis, the G5 Office and PAO can identify the general sentiment of the public regarding a situation and the frequency with which it is being discussed. Thus, if an issue requires widespread and detailed information, key leadership knows to disseminate the widest form of communication possible and can provide the answers that the public demands.

2. Literature Review

This section describes the relevant work in natural language processing and digital benchmarking. Understanding these topics will allow for an understanding of the scope of previous study involving Twitter and Google Trends, as well as other findings in the importance of benchmarking and sentiment analysis.

Natural language processing is a subfield of linguistics focused on how computers process and analyze large amounts of natural language data. Sentiment analysis is a component of natural language processing involving the interpretation and classification of emotions within text data. Top techniques in natural language processing show the utility of social media, Google Trends and news data sets. Ma et al. demonstrated a method of natural language processing involving the separation of Twitter queries by user to visualize the site's "sharing" capabilities, like the quantity and spread of retweets, tagged tweets and comments (Ma et al, 2013). Time and location are important metrics to visualize what population is concerned with a specific topic. Colleoni et al. discussed the use of the Twitter sentiment algorithm by large corporations to monitor the real-time evolution of their company's reputation (Colleoni et al, 2011). This data helps the corporations to not only visualize the changes in online sentiment regarding their company but allows them to respond to specific concerns if sentiment is relatively negative. Gunter et al. examined the explosive nature of news sites and the corresponding response of social media. They focused on the use of natural language processing by political campaigns to gather the sentiment of news articles and any corresponding comments (Gunter et al, 2014). With this data, political campaigns can analyze coverage after significant events like debates to help the team frame an appropriate response and any follow-on messaging within its own social media presence. D'avanzo et al. supplemented Twitter sentiment analysis with analysis of Google Trends data, commenting on the easy access to these interfaces and the comprehensive nature of both search data and sentiment data (D'avanzo et al, 2017). Abbasi et al. introduced a caveat that D'avanzo's work did not mention: they identified several roadblocks in analyzing sentiment with Twitter. Twitter text data often includes non-English languages and false positives, which are sentiment attributions that do not match the intended sentiment of words. False positives typically derive from sarcasm and slang, where a standalone word may seem positive but was intended by the user to carry negative sentiment (or vice versa) (Abbasi et al, 2014).

Benchmarking provides a means to evaluate by comparison to a standard or similar occurrence. This technique becomes extremely important when institutions and corporations are dealing with public appearance or progress over time. Top techniques include benchmarking in both natural language processing and sentiment analysis to understand the significance of an analysis. Homburg et al. discussed the utility of social media in marketing firms, analyzing how consumers react to active participation with consumers on various forums. Their study found that companies' interactions with consumers were perceived positively when addressing consumers' functional needs but were not received positively when addressing their social needs (Homburg et al, 2015). This is an important conclusion: it allowed corporations to recognize that benchmarking functional needs instead of social needs would save them time and resources, since customers were most concerned with interactions of this nature. It also allowed corporations to compare previous functionally related responses to determine the best method of response to maintain a positive sentiment from the consumer. Oliveira et al. described benchmarking by another group: the Higher Education Sector. They introduced a regulatory communication framework, or weekly monitoring of a university's social media sites, and a corresponding editorial model (Oliveira et al, 2015). This framework allows universities to make press releases when communication reaches a threshold of sentiment in social media, addressing both the magnitude of sentiment and any keywords common to this discussion. By benchmarking individual weeks, universities can study sentiment and quantity of social media interactions over time to notice specific jumps in either category and respond accordingly. Oliveira et al. also identifies key social media sites in regard to higher education: Facebook and Twitter are the most commonly used by the American public and should have correspondingly large focus from higher education institutions (2015). Finally, Mellon et al. asked if internet search data can provide measures of salience. They measured search data with Google Trends using key benchmarked events in the United States between 2004 and 2010. Mellon et al. chose to benchmark these events based on Gallup polling on the nation's "most important problem," where three topics were considered most important: the economy, immigration and terrorism (Mellon et al, 2014). By benchmarking these significant events, perceived as important by the American public, Mellon could visualize the corresponding response of media coverage to analyze the relationship between media coverage and salience (Mellon et al, 2014). They concluded that the two interact reciprocally: higher media coverage involves public salience and vice versa. This is an important conclusion because it asserts the utility of analyzing both social media and news coverage, as well as the utility of benchmarking to focus on high-impact events.

Despite the wealth of previous relevant work, there are still several gaps in research. Abbasi et al. identified the roadblock of false positives in sentiment analysis, but they offer no remedy to this problem and do not explain the impact this has on data. False positive detection and mitigation is an active area of research that leverages advanced techniques, but they are currently unrealistic to implement due to expertise and complexity. This indicates that Twitter sentiment analysis may not be statistically relevant to an institution like USMA because its results cannot be guaranteed with confounding variables like sarcasm. Additionally, Mellon et al. commented on the tendency for emotional responses in social media. This presents a response bias in gathered data, since users typically feel inclined to comment on topics when they feel either strong negative or positive emotions. This could lead an organization like USMA to feel an unnecessarily large desire to respond to these responses although they do not necessarily convey an overall public opinion. In an effort to provide a quick and easily understandable solution, we chose to analyze news articles which contain more formal language that yields itself to more convention analysis techniques; this analysis will be further addressed below.

3. Methodology

With this previous research in mind, and an understanding of Twitter, Google Trends, and Google News features, it is possible to develop a complete methodology for how an ideal system will work. It consists of four steps that are outlined in Figure 1 below and explained in the subsequent sections. The ultimate goal of this system is to ingest data, conduct analyses and utilize natural language processing to develop an app dashboard for use by EOC and PAO capabilities.

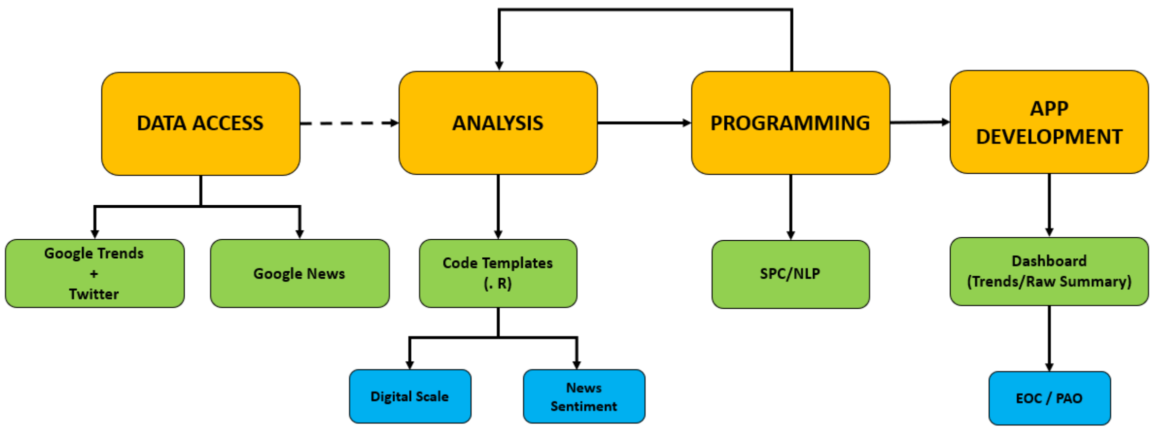

Figure 1. Methodology Overview

3.1 Data Access

Data acquisition in this methodology comes from two primary sources. For digital scale data, this work will analyze Google Trends. By leveraging the pre-existing API for Google Trends, *gtrendsR*, we can extract data on specific search queries and the quantity, location and timestamp of these searches (Massicotte & Eddelbuettel, 2019). This data will help us analyze the overall digital scale of internet searches involving specific USMA events, as well as overall searches for the Academy itself. Similarly, from the Twitter API *twitteR* we can pull data on specific content (both tweets and hashtags) and the quantity and timestamp of these searches (Gentry, 2015). This data will help us to further analyze the digital scale of certain events beyond Google searches to determine how much the Academy or an event is discussed in social media.

3.2 Benchmarking

When trying to understand the digital environment, relative scales provide a quick method to provide perspective to an analyst or decision maker. Working with the G5 Office, we developed a list of recent and institutionally relevant events to provide this benchmarking function. This work focuses on four main events that occurred between June and December 2019. Two of these events brought positive attention to the Academy: the close result of Army football's overtime loss against Michigan in September and the Army-Navy football game in December. Two events brought negative attention: the training accident in June involving the death of CDT Morgan and the disappearance and death of CDT Kurita in October. These are good benchmarks because they pertain to events that illicit both a large response and strong emotions from the public. As such, they should provide a large scale of responses and significant sentiment values in data.

3.3 Digital Scale

Digital scale data predominantly takes the form of plots showing interest over time: the output of Google Trends analysis. Figure 2 shows the output of a Google Trends search for four USMA-related search terms: "United States Military Academy," "USMA," "West Point" and "Army football." The time window of this output goes from April 4, 2019 to April 4, 2020: this includes all four benchmarked events. The plot identifies relative interest on the y-axis: this metric is the interest relative to the highest point on the chart for the entire time period (a value of 100 being peak popularity). As the plot shows, there are four large spikes in interest for "West Point:" each occurs in coordination with one of the events identified by the G5. Smaller peaks coordinate with other events in sports or press releases from USMA but focusing on the benchmarked events allows for analysis of more significant events (and corresponding emotional responses).

USMA can also augment digital scale data by searching Twitter for traffic. As mentioned before, sentiment data from social media (including Twitter) is difficult to use due to the presence of slang and sarcasm; however, Twitter can be used to measure the amount of posts with reference to the Academy in a given time period. Twitter data collection is limited to only the past week, so for the scope of this work it is impossible to sift through tweets pertaining to previous events. However, USMA can still use this method for future "hot topic" events and stay within the seven-day bound of the collection parameter.

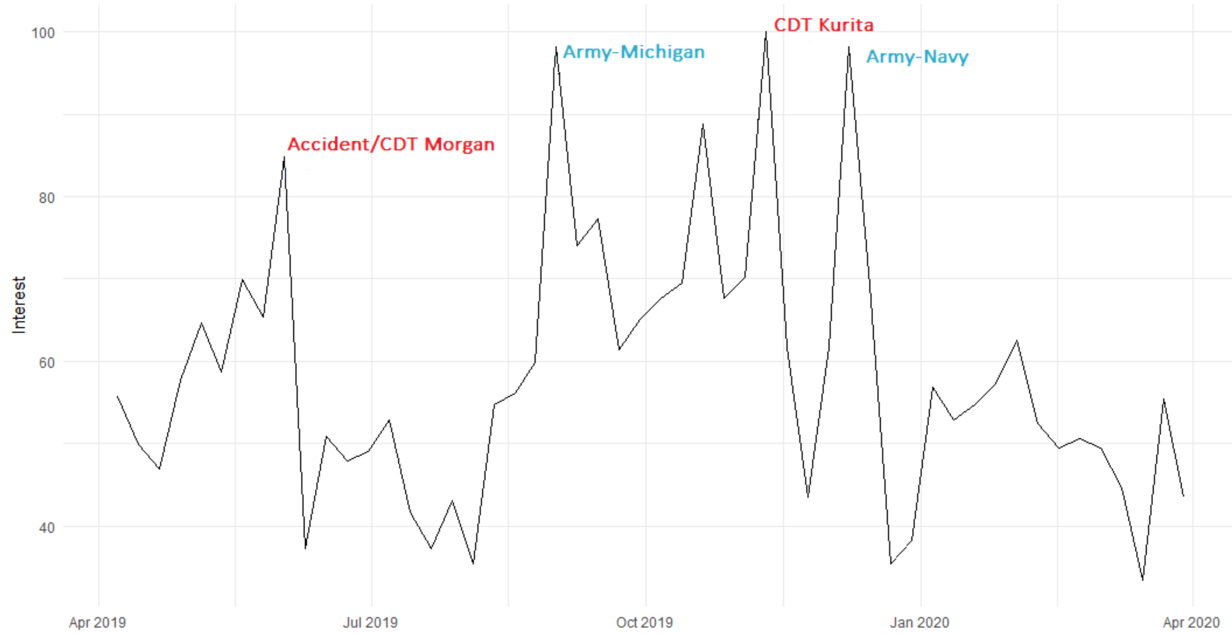

Figure 2. Rescaled Google Trends Interest Data for the United States Military Academy

3.4 Sentiment

For sentiment data, this work will analyze Google News. Google News, similar to Google Trends, uses web scraping via *httR* to provide the total quantity of articles pertaining to a search topic over a period of time (Wickham, 2019). We can then provide sentiment analysis on specific news articles. Analyzing sentiment from news articles provides a wealth of results while eliminating the confusing variables of social media. Social media posts face the difficulty of slang and sarcasm, which can confound data during analysis. News articles are generally written without these components, instead using formal language that is more valuable in an analytical capacity.

Sentiment data is best visualized in a histogram showing the overall news articles pertaining to a topic and the average sentiment of those articles. After gathering news articles, the analysis tool *sentimentR* gives each word in the article a value between -1 and 1: a more negative value is a more negative word, and a more positive value is a more positive word (Rinker, 2019). These values come from predefined libraries of English words, where each word is assigned an associated sentiment value based on common understanding. Each article then gets an overall sentiment value based on the average of all words' individual values, and these sentiment values are plotted in a final histogram to show the topic's sentiment distribution. Figure 3 depicts a sentiment analysis of news concerning the benchmarked event of the Army-Navy football game.

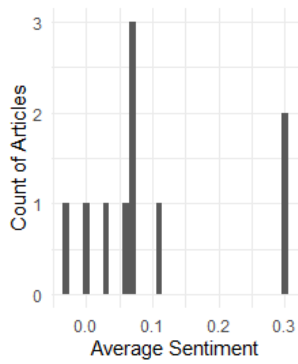

Figure 3. Army-Navy Football Game Sentiment

4. Results

By combining the digital scale and sentiment capabilities, we can conduct analysis on the benchmarked events. We will focus first on the negative events and then the positive events so they can be comparatively analyzed.

4.1 Negative Events

First, the negative press on 6 June 2019 regarding the vehicle accident and death of Cadet CJ Morgan. This event gathered a digital scale of 85% relative interest. Figure 4 shows the sentiment of news articles gathered surrounding this event in red. These articles have an average sentiment value of -0.035. This is the most negative average sentiment value of the benchmarked events, which corresponds with the gravity of the situation compared to football games and other common events. However, we also notice that there is a high concentration of articles with positive sentiment. This is likely due to the objective coverage of the accident, explaining the situation without editorializing and adding emotional words. From this analysis, we conclude that tragic events like accidents gather both large-scale and negative sentiment in public response but are also reasonably objective and aimed at describing the facts of a situation.

The second negative event involves the negative press on the disappearance and death of Cadet Kade Kurita on 23 October 2019. This event gathered a digital scale of 100% relative interest, meaning that this event brought the largest amount of searches for the Academy in the observed time period. Figure 4 above shows the sentiment of news articles pertaining to the event in blue, with an average sentiment value of 0.050. This is an abnormal sentiment value considering the severity and negativity of the situation and is likely due to the extremely low quantity of news articles surrounding the event (only three articles were available for analysis). The most probable reason for the low coverage is the nature of the topic: the circumstances around CDT Kurita's death were associated with suicide from an early point, and news outlets tend to stay away from discussions of suicide. This is just one example of an event that the news will likely not cover in depth; this bolsters the reasoning that news data should not be used as a single point of sampling. It is imperative to pull data from multiple sources when assessing a situation to understand the full scope of public opinion and discussion.

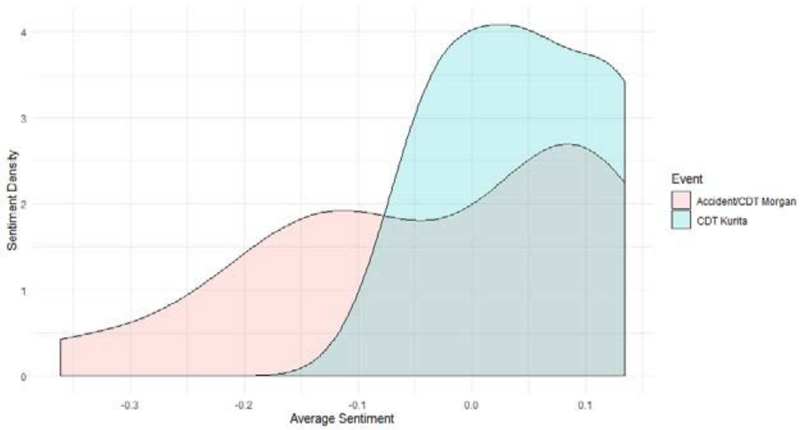

Figure 4. Negative Event Sentiment

4.2 Positive Events

When analyzing the positive benchmarked events, we begin with the close loss in the Army-Michigan football game on 7 September 2019. This event gathered a digital scale of 97% relative interest: this means it was the second-most significant event in terms of public interest in the Academy over the observed time period. Figure 5 shows the sentiment of news articles surrounding this event in red, with a slightly negative sentiment value of 0.025. The positive sentiment value corresponds with the perceived positivity of the event. However, it is important to observe that sentiment analysis regarding sports articles cannot differentiate between which team receives the positive sentiment. As such, it is likely that these articles *involve* Army football to some degree but focus on the fact that *Michigan* won the game. Although many Army football fans considered it a “win” that the team stayed in play with Michigan into overtime, the sentiment analysis does not take any perspectives into account. From these findings, we conclude that sports analysis for USMA will require a more in-depth analysis than simple sentiment of news articles, once again bolstering the importance of analyzing multiple sources.

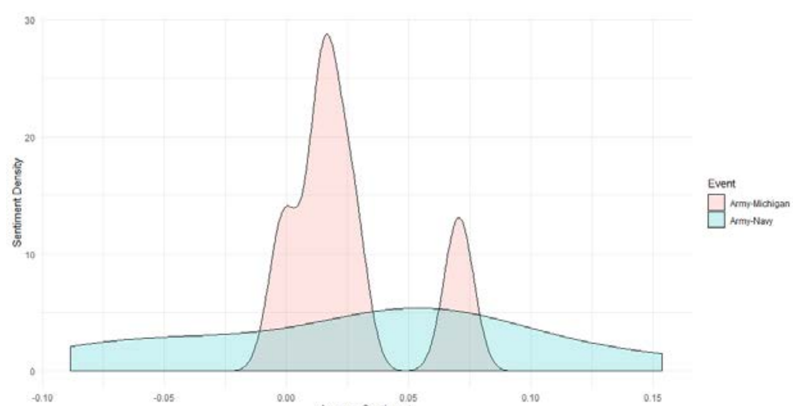

Figure 5. Positive Event Sentiment

The final benchmarked event involves the positive press on the Army-Navy football game on 14 December 2019. This event gathered a 97% relative interest. Figure 5 above shows the sentiment of news articles involving Army football in December in blue, with an average sentiment value of 0.031. We notice that this is the highest average sentiment value of all benchmarked events. This average value shows the difficulty of sentiment analysis: although the sentiment here is slightly *positive*, a basic scrape of the news articles shows that the coverage included *negative* press surrounding accusations of Cadets demonstrating “White Power” symbols during the game. Sentiment values are based on the words themselves, but the word “power” holds a positive connotation in normal speech. The algorithm has no way of defining the phrase “White Power” as

ISBN: 97819384961-8-9

clearly derogatory and negative; as such, the values will seem slightly positive. However, we also notice that the sentiment data is normally distributed: there is an equal spread of both positive and negative articles, indicating that there *are* some articles that focus on the negative perceived actions of Cadets with more critical verbiage. Additionally, similar to the Army-Michigan analysis, sentiment values like this reinforce why benchmarking is so important: this value is far more meaningful when comparing it to the average sentiment values of other large-scale events to visualize its relative positivity with a level of confidence. USMA can conclude that the overall positivity of the Army-Navy football game coverage superseded that of the accusations, and that relative to the Academy's recent tragedies this event is not significantly negative.

With the sentiment analysis of each benchmarked event, we found it necessary to manipulate the specific search query to gather the most results pertaining to the exact event. For the negative events of CDT Morgan and CDT Kurita's deaths, the Academy was most often referenced as "West Point" in news coverage: as such, searching for this query instead of "United States Military Academy" gathered the most accurate results. Conversely, with respect to the Army-Michigan and Army-Navy football games, the phrase "Army football" was the most common query related to the events. Thus, we can conclude that West Point will gather the most accurate sentiment data by searching through a series of possible queries involving the school instead of simply searching the same one with each event.

5. Discussion

The features developed in this project are generalizable across corporations and Army units alike. As previous research shows, large corporations and higher education entities are wary of the increasing quantity of available information online and its impact on the internet perception of their companies. As long as an organization has access to Google Trends and Google News- both public and free entities- it can quickly analyze search and post queries over time and location for both frequency and sentiment. Groups with similar goals to the PAO and G5 Office, whether in emergency situations or not, could benefit from the comprehensive nature of scale and sentiment output.

The successes of this research include demonstrating the utility of benchmarking and sentiment and the combination of Google Trends with news data. Benchmarking significant events allows for a simpler interpretation of digital scale data by focusing on the relative peaks of interest in USMA. The institution can prevent unnecessary panic by comparing each event's sentiment to other benchmarked events and determine the relative impact on public perception. Our analyses of benchmarked events confirm the scale and sentiment that USMA has already experienced, but also introduce potential failures in current analysis procedure.

The failures of this research include the inability to differentiate sentiment of slang and sports as well as dealing with small data sets. As previous research indicated, the presence of slang and sarcasm produces false positives that confound data. We experienced this firsthand when analyzing the tweets surrounding the Army-Navy football game "White Power" controversy in December. Without a way to eliminate the positive connotation with "power," the sentiment appears positive while in reality skewing slightly negative. Additionally, we discuss the failure of sentiment analysis to differentiate between the context of the winning and losing team in the Army-Michigan football game. We can conclude that sentiment analysis is difficult in sports, but this does not solve the identified problem of eliminating false positives. Further, some of the benchmarked events showed that news data may involve a limited number of samples. With a small dataset, it is difficult to draw statistically significance results appropriate for analysis. We must work to increase the scope of data for sentiment analysis by including multiple sources in order to gather significant data on sentiment.

6. Limitations & Future Work

Although this tool could be very useful to USMA, its data collection could take on several forms. Google Trends data is free and available to the public and provides a wealth of data on specific internet searches over time. Google News data, while also free, can produce a limited data set depending on the specific event of concern. Additionally, the collection of news data in general can be time-consuming. Custom scrapers must be written for each benchmark using specific key words involving that event. Additionally, searching for "West Point" in news can lead to articles that involve the Academy but do not pertain to the benchmarking analysis: for example, smaller-scale news outlets often write pieces on individual Cadets or servicemembers that the Academy is not particularly concerned with analyzing. More specific keyword searches provide better data but are often low in quantity. This balance is difficult to remedy but is vital to understanding the details of a situation and gathering relevant data.

This project demonstrates the versatility of media benchmarking and the clients it can cater to when necessary. The PAO benefits from this technology, visualizing the frequency and sentiment of posts related to its events. Google Trends data gives quick and easy access to typical public response monitoring, including web traffic and common searches related to the institution. The G5 Office benefits more significantly from these capabilities in an emergency capacity, using the speed and breadth of this process to quickly gather large amounts of data on public sentiment in the wake of a significant event. Using a concise methodology and comprehensive programming, the United States Military Academy can quickly gather data on all situations, whether "hot topic" or simple daily life. This knowledge empowers key leaders to provide quick and concise summaries on all relevant information.

7. References

Abbasi, A., Dhar, M. & Hassan, A. (2014, May). Benchmarking Twitter Sentiment Analysis Tools. Research Gate.

Colleoni, E., Arvidsson, A., Hansen, L. & Marchesini, A. (2011, May). Monitoring Corporate Reputation in Social Media using Real-time Sentiment Analysis. *The 15th International Conference on Corporate Reputation: Navigating the Reputation Economy, New Orleans.*

D'Avanzo, E., Pilato, G. & Lytras, M. (2017, September). Using Twitter sentiment and emotions analysis of Google Trends for decision making. Emerald Insight, Electronic Library and Information Systems.

Gentry, J. (2019). *Package twitteR.* CRAN.

Gunter, B., Koteyko, N. & Atanasova, D. (2014). Sentiment analysis: A market-relevant and reliable measure of public feeling? *University of Leicester, International Journal of Market Research Vol. 56 (2)*. DOI: 10.2501/IJMR-2014-014.

Homburg, C., Ehm, L & Artz, M. (2015, October). Measuring and Managing Consumer Sentiment in an Online Community Environment. University of Mannheim. DOI: 10.1509/jmr.11.0448.

Jun, S.P., Yoo, H.S. & Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Elsevier, Technological Forecasting and Social Change Vol. 130*, 69-87.

Lloyd, L., Kechagias, D. & Skiena, S. (2010, April). *Lydia: A System for Large-Scale News Analysis.* Stony Brook Department of Computer Science.

Ma, H., Qian, W., Wei, J., Xia, F., Yu, C. & Zhou, A. (2013, June). On benchmarking online social media analytical queries. *First International Workshop on Graph Data Management Experiences and Systems Vol. 10.* DOI: 10.1145/2484425.2484435.

Massicotte, P. & Eddelbuettel, D. (2019). *Package gtrendsR.* CRAN.

Mellon, J. (2014). Internet Search Data and Issue Salience: The Properties of Google Trends as a Measure of Issue Salience. *Journal of Elections, Public Opinion & Parties, Vol. 24 (1)*, 45-72. DOI: 10.1080/17457289.2013.846346.

Oliveira, L. & Figueira, A. (2015, October). Benchmarking analysis of social media strategies in the Higher Education Sector. *Procedia Computer Science, Conference on ENTERprise Information Systems.*

Rinker, T. (2019). *Package sentimentr.* CRAN.

Wickham, H. (2019). *Package httR.*" CRAN.