

Global News Sentiment Analysis



Eric Castellanos, Hang Xie, and Paul Brenner

Abstract The Journalist’s Creed, a declaration of the principles, values, and standards of a journalist, states that a journalist should “believe that clear thinking and clear statement, accuracy and fairness are fundamental to good journalism”. However, in recent years there has been concern that personal, corporate, and government biases and opinions have had an effect on the fairness and integrity of modern journalism. Many mainstream media outlets are well known to have certain political inclinations and occurrences of “fake news” have become common. Studying mainstream international news sources is important for numerous reasons, such as national security, cultural stability, and the development of social/economic policy. Through the use of modern natural language processing techniques, we performed sentiment analysis on 10,809 articles published by ten globally distributed sources covering eleven globally impactful events, spanning a period of eleven years. We tested three open-source sentiment analysis tools and one commercial tool from Microsoft. We also tested three text processing approaches, i.e., sentence, paragraph, and article-based approaches. We observed that sentiment polarity was generally consistent across news agencies, however, the relative percent differences in sentiment magnitude varied by multiple standard deviations. Further, we found general agreement in sentiment polarity across the four analysis tools strengthening our finding regarding the news source content sentiment trends. The work provides both a reproducible computer-based metric for quantifying differences in global journalism and insights into the effective use of modern sentiment analysis tools.

E. Castellanos · P. Brenner (✉)
Center for Research Computing, University of Notre Dame, Notre Dame, USA
e-mail: paul.r.brenner@nd.edu

H. Xie
Department of Computer Science, Wenzhou-Kean University, Wenzhou, China
e-mail: xieh@kean.edu

1 Introduction

Sentiment analysis is often used to measure and classify bias in social media posts and customer reviews [1, 2]. In comparison, however, there are a lack of sentiment analysis studies that focus on major/mainstream news outlets, and among those which do, most studies focus solely on U.S. news sources. Our study expands sentiment analysis to global news sources which can affect national security, social and economic policy, and cultural stability. While many believe that mass media does not have the ability to direct government [3, 4], several studies have provided evidence for the claim that news coverage does have the ability to shape public policy [5–7]. Most of these studies focus on how coverage of foreign events in U.S. media affects U.S. public policy. Few published studies focus on coverage of internationally significant events in foreign media and the coverage’s subsequent influence on public opinion. In the specific case of U.S. events with international impact, the coverage in foreign media shapes public policy and opinion in those foreign countries. For instance, if a foreign news outlet were to report negatively about the performance of automobiles manufactured in the U.S., this could lead to a decrease in purchases of the automobile in that country and other negative economic effects. As a second example, if a foreign news outlet were to report negatively about U.S. international intervention, it could lead to harboring hostility toward the U.S. and in the worst case, military intervention.

In addition, the loss of ground truth and the emergence of so-called “fake news”, also contributes to the shaping of public policy [8]. Negative news reporting of U.S. events and observations of “fake news” can lead to the loss of efficiency in a society and stifle democracy. This problem can be amplified by the development of country-level “echo chambers”. For example, Russia plans to introduce legislation that will isolate its internet servers from the rest of the world [9]. If Russia were to then publish untrue articles, Russian citizens would have few internal ways to verify the truth. Further, politicians could manipulate media and information to a point approaching a dictatorship. While it is important to recognize this rise of “fake news”, it is tangential to the focus of our study. The rise of “fake news” and political spin has more broadly brought into question the validity of reporting in today’s largest and most respected global newsrooms.

We must discuss the fine line between perspective and bias. Perspective can be defined as a viewpoint or frame from which a person sees an event. A journalist covering a particular event often has a deeper understanding of the supporting background than the average citizen. Providing this citizen reader with some perspective is part of the journalist’s task. However, it seems that sometimes reporting can go beyond perspective, and venture into bias. Bias implies that truth and results are unfairly prejudiced in favor of one person or group [10]. It is difficult to discern what is perspective and what is bias in journalism, and so our study seeks to enhance the discernment with reproducible algorithmic metrics.

We leverage mainstream sentiment analysis tools to better identify bias and differences in perspective across major global news sources in order to address these

concerns regarding public disinformation. The paper is organized as follows. The first section reviews relevant previous work that has been done for sentiment analysis of media. The second section will cover methodology/experimentation where issues with the collection and querying of data, details of the chosen sentiment analysis libraries, three approaches to process the news articles and design of experiment will be discussed. In the third section, we will present our data analysis results. Fourth, we discuss possible reasons for the output of our chosen sentiment analysis libraries. Finally, we will conclude with summary observations and acknowledgements.

2 Related Work

Multiple works have explored the relationship between computational sentiment analysis studies and bias. For example, Khan and Taimoor identify bias analysis as a subfield of sentiment analysis [11]. Additionally, Zhang, Kawai, Nakajima, and Matsumoto investigate sentiment bias when websites present information [12].

Multiple studies on media bias leverage sentiment analysis. Pang and Lee published a work describing the state of the field of sentiment analysis, which covers techniques and approaches that promise to directly enable opinion-oriented information-seeking systems [13]. Most of the works focus on social media rather than traditional news (newspapers, TV stations, etc.), and those that do focus on traditional news, generally restrict the study to U.S.-based sources. Abdul and Diab (a non-US media-focused work) describe their methodology for performing sentiment analysis on Arabic language social media [14]. Balahur describes the difficulty of analyzing sentiment of news articles, and notes the differences between analyzing sentiment of social media and sentiment of news articles [15]. In the following paragraphs, we discuss works similar to ours which focus on news articles.

Bautin et al. focus on machine translation and if sentiment will be affected when translating from English to a foreign language [16]. They explored an approach utilizing state-of-the-art machine translation technology and performs sentiment analysis on the English translation of a foreign language text. The study is relevant to ours because they used news articles from international sources as their data. One of our news sources is based in Argentina and the Spanish articles from this source required automatic translation from Spanish to English prior to the sentiment analysis phase.

Budak et al. describe a machine learning approach to classify news articles and quantify bias [2]. Through a combination of machine learning and crowdsourcing techniques, it investigates the selection and framing of political issues in fifteen major US news outlets. The study utilizes a database of over 800,000 articles. Like ours, the study aims to quantify bias, but does not use sentiment analysis to do so. In addition, the data used in the study is restricted to articles from only US news sources. However, the study serves as a good reference as their methodology/experimentation is similar to ours. Kaya et al. describe the difficulties of performing sentiment analysis on political news articles [17]. They state that sentiment analysis struggles on news

texts due to the lack of large-scale golden standard datasets and the high context-dependency of sentiment-inducing phrases.

3 Methodology and Experimentation

In this section, we break down our methodology into three sections. We first discuss how we gathered our data, initial challenges we faced and how we eventually utilized the LexisNexis data repository. We then discuss how we extracted and cleaned articles, followed by selection of the sentiment analysis algorithms (with basic details on each tool). Then, our reasoning for selection of topics will be discussed. Finally, we introduce three approaches to process news articles and explain why we chose a paragraph-based approach to process articles.

As pointed out by Hamborg et al., finding a data repository of news articles for a computational study is not an easy task [18]. Two datasets available to the public are Integrated Crisis Early Warning System (ICEWS) and Global Database of Events Languages and Tone (GDELT). ICEWS consists of coded interactions between socio-political actors such as cooperative or hostile actions between individuals, groups, sectors, and nation states [19]. GDELT is a large, comprehensive, open database of human society. It is a platform that monitors the world's news media in print, broadcast, and web formats, in over 100 languages, and stretches back to January 1, 1979 through present day [20]. The datasets are vast, date back many years, and contain useful information about events and the articles written about them. However, they do not provide access to full article text, which is essential to our study. We then discovered NewsAPI, a simple and easy-to-use API that returns JSON metadata for headlines and articles [21]. Further, NewsAPI gave us access to title and short, two-sentence summaries of news articles. Ultimately we decided that in order to perform a more complete study, we needed access to full article text. NexisUni, a data repository provided by LexisNexis, features more than 15,000 news, business and legal sources, including U.S. Supreme Court decisions dating back to 1790 [22]. The biggest advantage that NexisUni had to offer was the availability of an API that allows developers to query the NexisUni database on things like article topics and dates, and return the full article content in XML format.

Python was chosen as the language for our analysis. We utilized the lxml library for extraction of the article text and organization of files by article topic and news source. Next, we had to decide which sentiment analysis algorithm and library we were going to use. We researched many different popular natural language processing libraries available in the Python language [23]. While there are many natural language processing libraries and sentiment analysis algorithms available, many of those tools are ill-equipped for dealing with texts like news articles. Most sentiment analysis libraries are intended to be used to analyze small bodies of text, no more than a few sentences. In addition, many of the tools are not very good at dealing with highly contextualized texts, which news articles generally are [24]. As a result, we focused on selecting a sentiment analysis tool that attempted to tackle these chal-

lenges. We initially selected the VADER Sentiment Analysis tool available in the Natural Language Tool Kit 3.4.1 (NLTK) library in Python [25]. VADER's documentation states that the algorithm is equipped to deal with somewhat long bodies of text and can manage "tricky" sentences, and was evaluated by the data from Twitter, movie reviews, technical product reviews, and 500 New York Times opinion editorials [25]. In addition, VADER has been recognized as one of the best tools available for sentiment analysis. VADER combines the use of a lexicon and processing characteristics of sentences to determine the polarity of a sentence. For words in the lexicon, sentiment scores were assigned by human raters in the range of -4 (extremely negative) to 4 (extremely positive), and the overall sentiment score of a sentence is obtained by adding all sentiment scores of words that appear in both VADER's lexicon and sentence. Then, VADER will normalize this final score to a value between -1 and 1 (the normalize process will be discussed further in a later section). VADER also takes into account punctuation, capitalization, and n-gram examination for things like negation and context [26].

To improve our tool evaluation, we also leveraged three other sentiment analysis tools: TextBlob 0.15.3, Afinn 0.1, and Microsoft Azure Text Analytics 0.2.0 (MATA). TextBlob is considered as an easily accessible NLP package and it stands on the shoulder of NLTK and Pattern [27]. It is an easy-to-use library, equipped with impressive functionality, which attracts a wide range of people for their projects' initial prototyping purpose. TextBlob provides two different algorithms to perform sentiment analysis: one is the implementation of the "pattern-en" module [28], which is similar to the lexicon-based approach that VADER adopts, and another one is the Naive Bayes classifier which is trained on movie reviews. In our research, we choose to use the default approach, i.e., lexicon-based approach, as the way to calculate the sentiment. Afinn is also a popular lexical-based sentiment analysis toolkit whose lexicon mainly comes from Twitter [29]. According to Ribeiro, Araujo, Goncalves, Goncalves, and Benevenuto, for the 3-Classes sentiment analyzer, VADER ranked first and Afinn ranked third after comparing 24 sentiment analysis methods on a benchmark of eighteen labeled datasets [26]. Thus, we decided to use this package due to Afinn's excellent performance. Different from VADER, TextBlob, and Afinn which are free to use and can be implemented locally, MATA API is a cloud-based service for natural language processing and has been commercially available for several years [30]. One of the reasons we choose to use this tool was to determine if there is a significant difference between commercial and open-source natural language processing libraries.

For our study, we chose ten different news sources that cover the six populous continents. Choices for selection of news sources were made based on: rough global coverage, popularity of the news sources in their respective regions, availability in English, availability of articles mapping to target events, modestly opposing political spectrum (CNN vs New York Post) and their availability in the NexisUni data repository. The news sources are La Nacion (Argentina), All Africa, Xinhua, The Times of India, Al Jazeera, BBC, RIA Novosti, The Australian, CNN, and The New York Post as shown in Fig. 1. It is important to note that La Nacion is an Argentinian newspaper written in Spanish. Since Matheus et al. [31] suggested that applying machine



Fig. 1 Distribution map for selected news sources

translation to Spanish input text prior to using an English sentiment tool can be a competitive strategy, we utilized Google's translate API in order to prepare articles from La Nacion for sentiment analysis. We created an article count matrix to visualize how many articles were available for each topic, as Fig. 2 shows. As expected, there were significantly more articles returned for the mainstream popular news outlets like CNN, BBC, and Xinhua, while more regional media sources or those with a smaller number of English readers like All Africa and Al Jazeera, respectively, often did not have many articles available for the target events. Therefore, the results for sources such as BBC will have higher statistical relevance than those of All Africa.

The selected events are the execution of Saddam Hussein, the Saffron Protests of Burma/Myanmar, the Election of Barack Obama, the indictment of Omar Al-Bashir (Sudan), Kim Jong Un's succession, the Fukushima Nuclear Disaster, the Benghazi Attacks, the Snowden Incident, Russia's annexation of Crimea, the Legalization of Same-Sex Marriage in the US, and the Election of Donald Trump, as shown in Table 1. Articles were selected over an eleven year time period (2006–2016). One major event was selected for each year, and articles were gathered from the ten sources for each event. In attempt to observe differences in reporting, five selected events were about topics that primarily concerned the U.S., and the other six were about international events.

After querying the NexisUni database, extracting article text from the XML documents, and data cleaning, the articles were ready to be analyzed by the sentiment

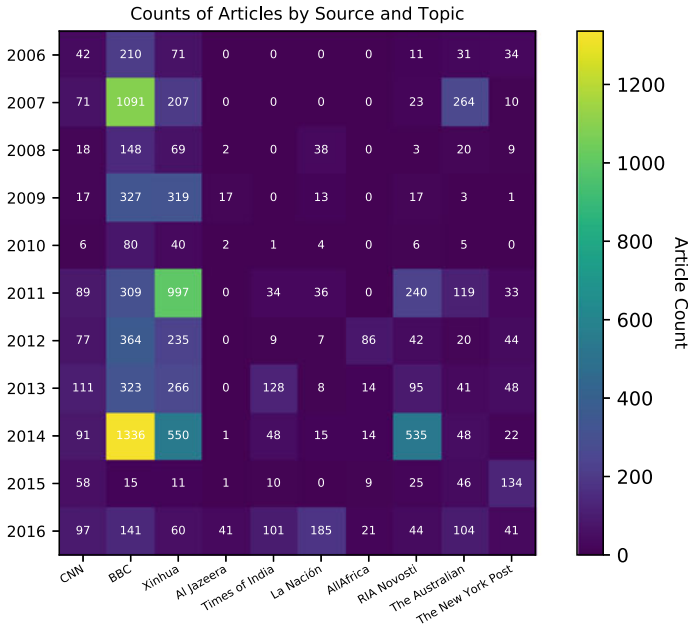


Fig. 2 Article count matrix

Table 1 Selected global events

Year	Event
2006	Saddam Hussein sentenced to death
2007	Saffron protests in Burma/Myanmar
2008	Barack Obama elected as US President
2009	Omar Al-Bashir convicted on humanity crimes
2010	Kim Jong Un announced as Kim Jong II sucesor
2011	Fukushima-Daichi nuclear disaster
2012	US facilities attacked in Benghazi
2013	Edward Snowden leaks sensitive information
2014	Russia annexes Crimea
2015	Same sex marriage legalized in the US
2016	Donald Trump elected as US President

analysis libraries. While the algorithms were reasonably equipped to handle large bodies of text, our initial study utilized paragraph-based approach to process the news articles, where each article was broken down by paragraph, a sentiment score was assigned to each paragraph, and finally each article was assigned the average of all its paragraphs’ sentiment scores.

4 Results

4.1 Initial Observation

We first calculated the mean sentiment values of all articles by source and then by event. Not all events were covered by all the news outlets, and thus we see some years missing in the charts. The majority of the sources had some polarity alterations for mean sentiment value for the various events, with some events having positive mean sentiment values and other events having negative or neutral mean sentiment values. Due to the publication space limits, the compound sentiment scores of all articles are available by contacting the authors.

Some events were chosen for their potential to be controversial, while other events had the potential to be either unanimously positive or unanimously negative. For instance, we chose the Fukushima Nuclear Disaster because intuitively, we expected the sentiment of news on this topic to be negative, while we chose Donald Trump’s election since we surmised it might be a controversial topic around the world. We analyze the article bias for these two events here in the paper and provide full data (online) for readers to evaluate bias in the other events and sources.

In Figs. 3 and 4, we show the sentiment distribution histogram for Fukushima Nuclear Disaster and Donald Trump’s Election from different sources. The X-axis

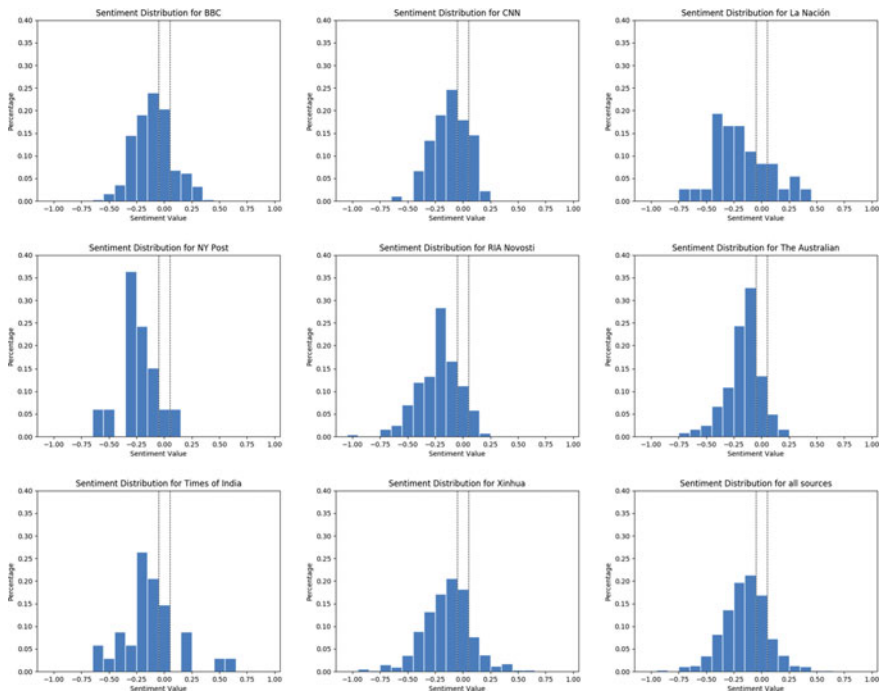


Fig. 3 Sentiment distribution for Fukushima nuclear disaster

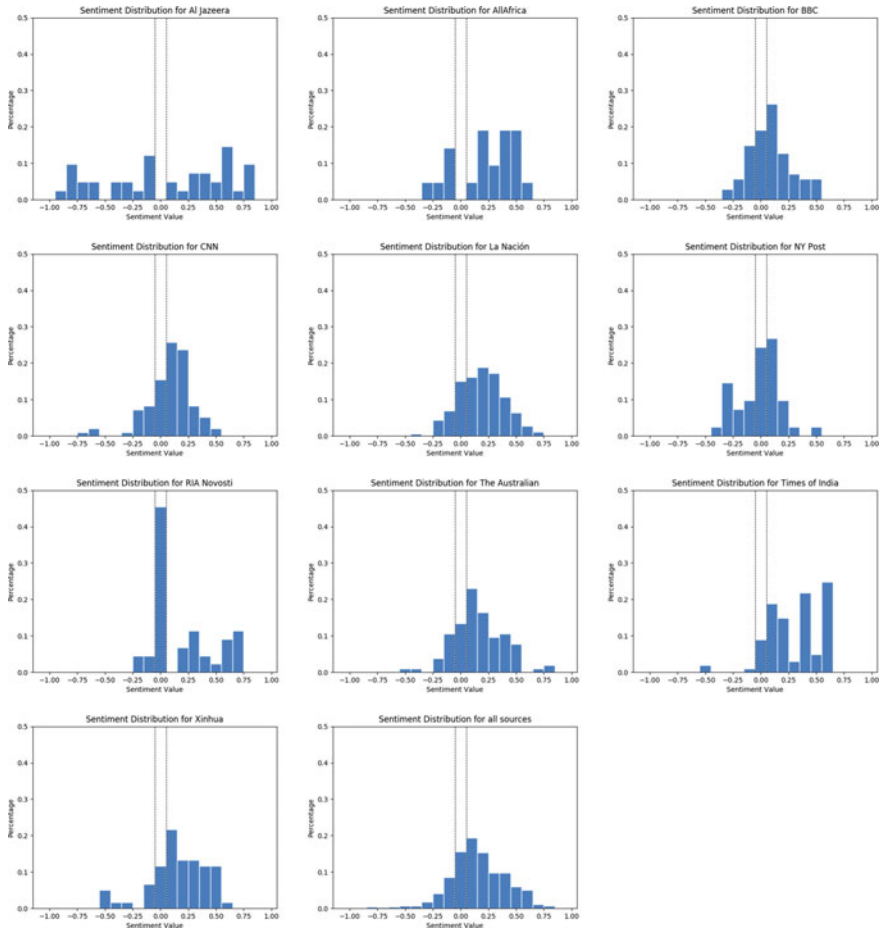


Fig. 4 Sentiment distribution for Trump’s election

represents the sentiment compound value of the news while the Y-axis denotes the percentage of news with a corresponding sentiment compound value. We calculate this probability by dividing the number of news articles in the range of emotion value by the total number of news under this topic. For sentiment compound value, VADER has defined that if the compound score is less or equal to -0.05 , then the text tested shows a negative sentiment and if it is higher or equal to 0.05 , then it shows a positive sentiment. Any compound score which lies in between -0.05 and 0.05 will be regarded as neutral sentiment.

Figure 4 demonstrates the sentiment distribution for Fukushima Nuclear Disaster from different sources and in the last subgraph we also present the sentiment distribution from all sources. For news from all sources, the probability of negative sentiment in the news outweighs the probability of positive sentiment. Specifically,

69.47% of news showed a negative feeling, 13.68% revealed a positive feeling, and 16.86% of news conveyed a neutral sentiment.

The sentiment distribution for Donald Trump's Election, as shown in Fig. 4, yields greater positive sentiment occurring in a news articles than the probability for negative sentiment. In detail, 67.11% of news demonstrated a positive sentiment while 17.34% demonstrated a negative sentiment.

We provide box plots to demonstrate the spread of our data points and check if mean values were being affected by outliers. Information on how whiskers were calculated can be found in the documentation of Matplotlib's Boxplot function, default values were used for all arguments. Where IQR is the interquartile range ($Q3 - Q1$), the upper whisker will extend to last data point less than $Q3 + 1.5 * IQR$. Similarly, the lower whisker will extend to the first data point greater than $Q1 - 1.5 * IQR$. Beyond the whiskers, data are considered outliers and are plotted as individual points.

The boxplots of sentiment by source demonstrate news publishers possess different viewpoints for each event. In addition, from the boxplot in Fig. 5, we observe that for publishers like BBC, Xinhua, and New York Post, many outliers are present, indicating that these publishers' viewpoints tend to vary widely. For example, for 2014 [Russia Annexes Crimea], news articles released by BBC are either quite positive or quite negative. Note, some news publishers, like Al Jazeera and The Times of India only report a few articles for some events, which provides less statistical mass.

From Fig. 6 we can see for most of the events, publishers tend to have similar sentiment polarity toward them. However, for events like 2009 [President of Sudan, Omar Al-Bashir, Convicted for Crimes Against Humanity], sentiment diverges for different news publishers, where NY Post holds a positive view toward it the rest of the publishers generally hold a negative view. We also find that for events like 2014 [Russia Annexes Crimea], there are many outliers in the graph, showing that the overall sentiment for this event is polarized and was quite controversial during that period.

4.2 Comparison of Four Sentiment Analysis Tools

We tested the other three tools for the same dataset, trying to compare VADER with those tools. Since the sentiment scores calculated by VADER and TextBlob are between -1 and 1 while those calculated by Afinn and MATA are not, we normalized the sentiment scores from the latter libraries to -1 to 1 for comparison purposes.

For MATA, the return sentiment score is ranged between 0 and 1 , making normalization straightforward. However, for Afinn, since it doesn't have the upper and lower bound of its return sentiment score, we normalized it in the following way: First, we try to calculate the average and standard deviation of the sentiment scores we have for Afinn. Then, we remove the data which is 15 standard deviations away from the average Afinn score. Finally, we divide all Afinn scores by the maximum absolute Afinn score we obtained from all news sources. In this phase, we removed

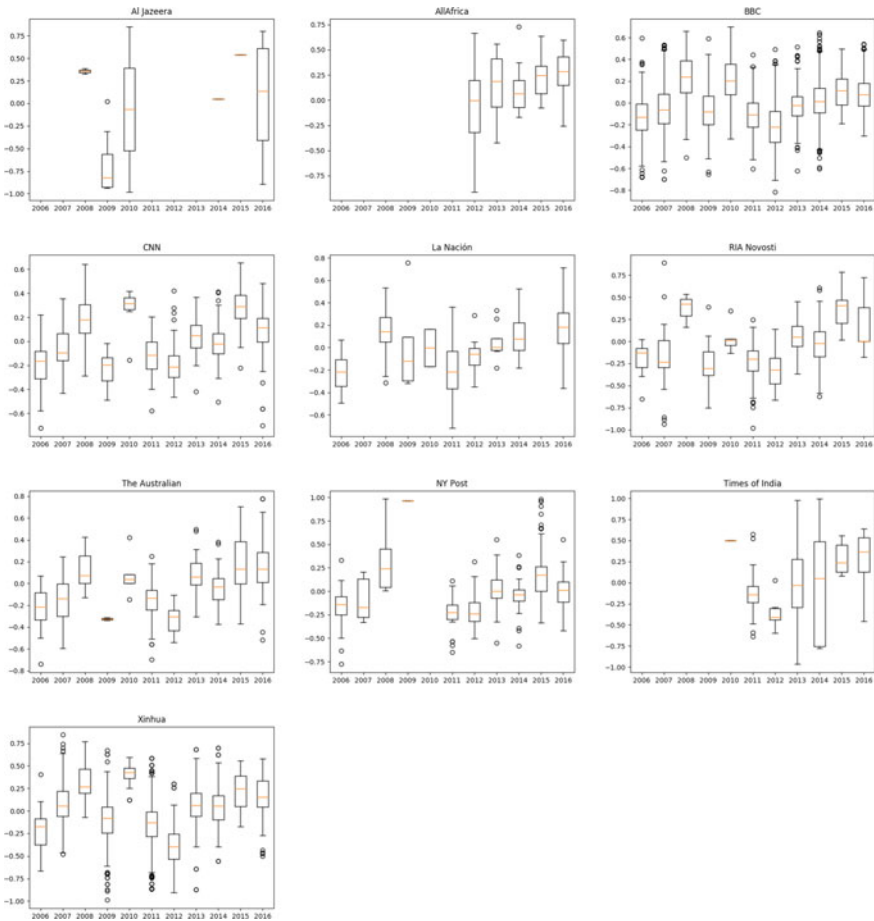


Fig. 5 The boxplots of sentiment based on source by VADER, paragraph-based

two news articles published by Al Jazeera: one for 2010 Kim Jong Un Succeeds, and another one for 2016 Trump elected.

Figures 7 and 8 show the calculated sentiment scores for different sources and events based on our chosen packages. From a practical perspective, there should be a center area that lies in between the positive and negative sentiment, which denotes neutral sentiment. For instance, although sentiment with score 0.01 may mean slightly positive while sentiment with score -0.02 may represent slightly negative, it is difficult to distinguish a difference. Thus, in our paper, we define sentiment score between -0.05 and 0.05 as neutral.

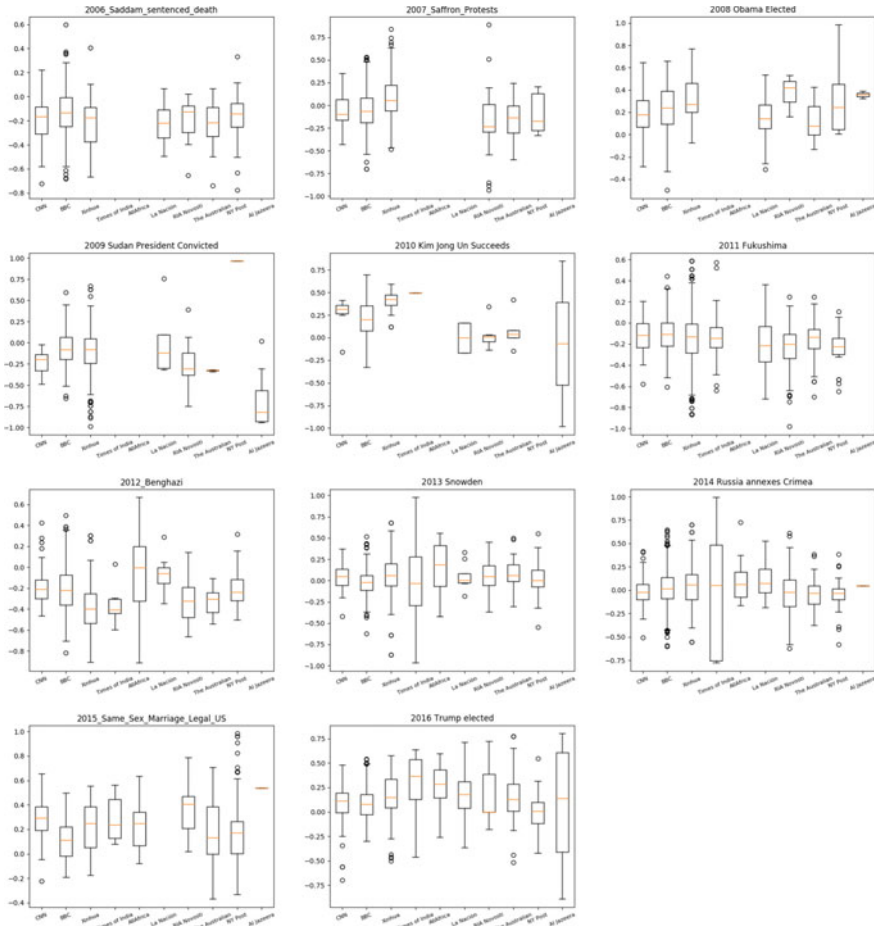


Fig. 6 The boxplots of sentiment based on topic by VADER, paragraph-based

From Fig. 8, we find that the sentiment for news published by a particular source for different events varies, indicating that publishers show different perspectives (sentiment) while reporting news versus consistently treating all news as negative, positive, or neutral. When looking into the detail, we find that for each event published by a specific publisher, the four tools tend to have an agreement concerning the sentiment polarity (when outside the neutral zone). For example, for news published by Al Jazeera concerning 2009 President of Sudan, Omar Al-Bashir, Convicted for Crimes Against Humanity, sentiment calculated by VADER, Afinn, and MATA shows a negative sentiment, while the sentiment calculated by TextBlob shows a neutral sentiment. This occurs throughout the events as demonstrated by both Figs. 8 and 9. We perceive this to be reasonable considering the difference of training data for each tool and sometimes it's difficult, even for humans, to discern the difference between



Fig. 7 Collective sentiment compound score from each source, paragraph-based

neutral and slightly positive or negative, since people tend to have different views on single news articles.

In Fig. 9, we group the sentiment compound score based on the event. For most events we choose, different publishers tend to have similar sentiment polarity. Surprisingly, for events like 2009 President of Sudan Indicted for Crimes Against Humanity, New York Post tends to have extremely positive sentiment while the rest of publishers hold negative or neutral-to-negative sentiment.

Overall, VADER is quite sensitive toward the sentiment underlying the news articles, as we can tell from Figs. 8 and 9, in general the compound score calculated by VADER is higher than the rest of the three and it can clearly detect the polarity of the news. However, TextBlob finds minimal sentiment magnitude variance in our study since most of the values it calculated lie in the range between -0.05 and 0.05 .

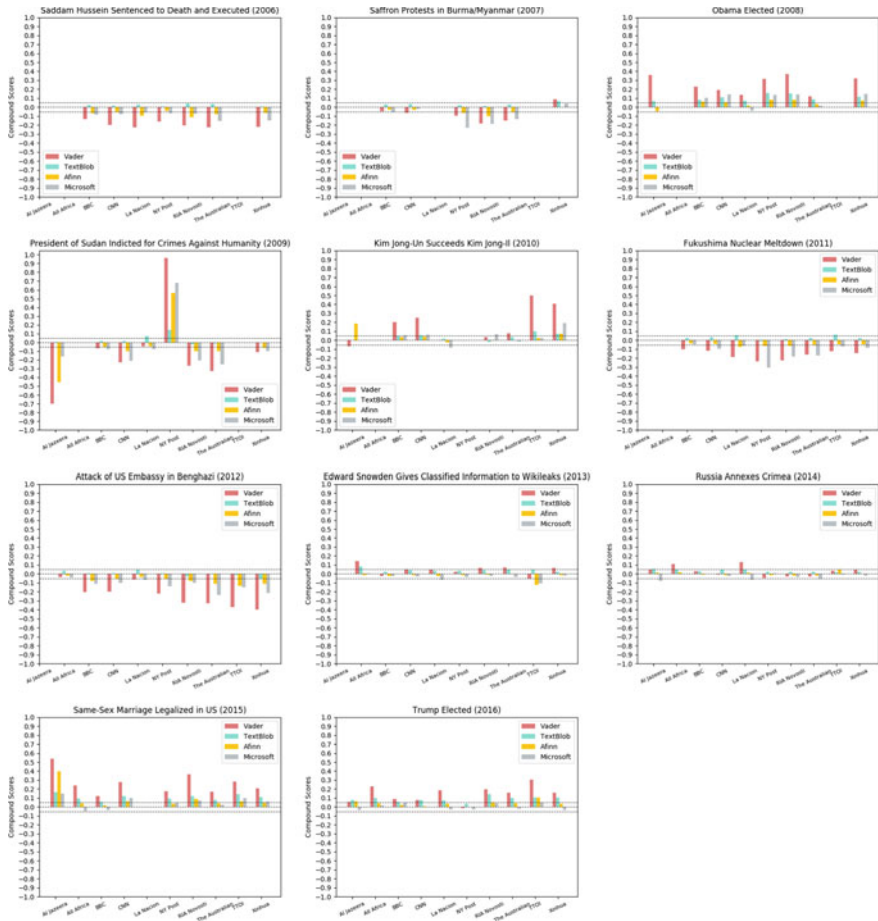


Fig. 8 Collective sentiment compound score for each event, paragraph-based

The performance of Afinn and MATA are fairly good regarding their ability to detect the polarity, but they are not as sensitive as VADER.

4.3 Three Article Processing Approaches

There are three common approaches to parsing text articles: sentence-based, paragraph-based, and article-based approaches. We tested all three approaches to process the news articles, and it turned out that the paragraph-based approach provided us with a relatively objective result which is neither too sensitive nor insensitive.

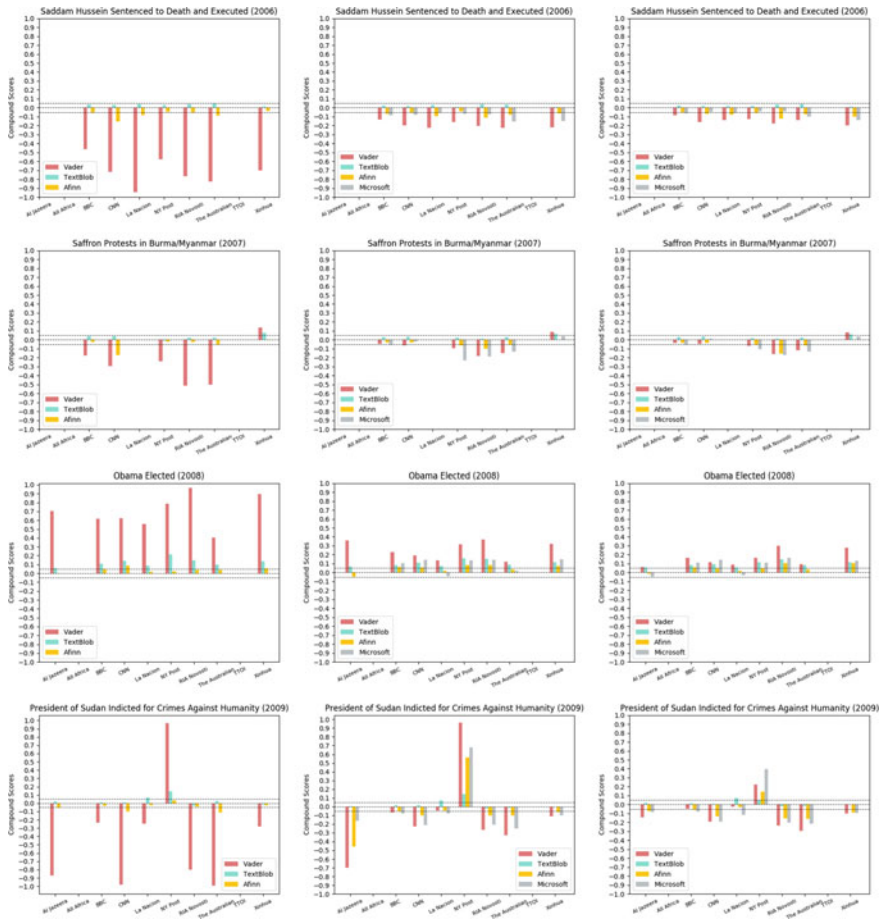


Fig. 9 Three article processing approaches. Column 1 = Article, Column 2 = Paragraph, Column 3 = Sentence

The article-based approach was easy to implement since we can directly put the text into the library to obtain its sentiment score. Unfortunately, since MATA has a limitation that the maximum size of a single document be less than 5120 characters and there were 1765 articles exceeding this limit, we didn't calculate the article-based approach for MATA.

It is also easy to calculate the sentiment score by paragraph-based approach, since the tag <p> in the LexisNexis XML files helps identify the paragraph. For sentence-based approach, however, there are many different cases that can be applied to determine a sentence. For example, if a text is ended by a period, exclamation mark, or question mark. Also, sometimes there are cases when a sentence ends with three continuous periods "...". Initially, we wanted to write a regular expression to extract the sentence out from a paragraph. However, given the linguistic difficulty

of handling all cases which can be used to form a sentence, we decided to use a well-trained tokenization package in the NLTK library to implement the sentence extraction task.

Our results in Fig. 9 show that VADER is more sensitive to the way we process the text. Nevertheless, for three different approaches calculated by VADER, although they have differences in the absolute value for the emotion score, they don't yield much difference in the polarity of the sentiment. Specifically, we find that VADER seems to be extremely sensitive to the emotion in the news articles processed by the article-based approach. For the result calculated by paragraph-based approach by VADER, it is less sensitive than the article-based approach, but is more sensitive than the sentence-based approach.

In a minority of cases, the sentiment polarity shifted as we alter the text processing approaches. For example, for news published by CNN for 2007 Saffron Protests in Burma/Myanmar, Afinn first possesses a negative view about it for article-based approach, but when we test the paragraph-based, the CNN then holds a neutral view toward it, though there is not much change in absolute sentiment score. Overall, the article-based approach is the most sensitive approach, followed by the paragraph-based approach and the sentence-based approach. Also for VADER, no matter which text processing approach we choose, it will rarely alter the sentiment polarity for our news. However, for the rest of tools, since it may alter the sentiment polarity when we use different approaches to process the text, we should think carefully concerning which approach works better before further application.

5 Discussion

The heightened sensitivity of VADER for different text processing approaches could be explained by its normalization approach. We know VADER uses a lexicon-based approach to obtain its sentiment score, where each sentiment-bearing word in a text will be mapped to a sentiment score which is given by human raters in a range of -4 to 4 , and then we sum it up as the overall sentiment for this text. However, we notice that the return value the VADER provides us is between -1 and 1 , indicating there has a normalization process before we get the sentiment score. Hutto and Gilbert do the normalization in the following ways, as it shows in VADER's source code:

$$\text{normalizedScore} = \frac{\text{score}}{\sqrt{\text{score}^2 + \alpha}}$$

where the score is the sum of the sentiment scores for each sentiment-bearing word, and the α , which is set to 15 as a default value, is used as a normalization parameter. For example, there is a sentence in one of the news articles concerning Trump's Election: "There are jurisdictions that fail to report hate crime statistics". In this sentence, three words get matched to the VADER's lexicon: fail, hate and crime, and their sentiment score are -2.5 , -2.7 , and -2.5 , respectively. Thus, the sentiment

score before we normalize it should be -7.7 , and when we replace the score in the above formula by -7.7 , and let α be the default value 15, we get our normalized sentiment score -0.8934 for this sentence.

As we can observe from this formula, if the score becomes larger and larger, then the effect of α toward the denominator will decrease, and the whole equation will approach to -1 or 1 . We know that the length of the text also matters for the performance of VADER. If the text is longer, then the drag effect of the α will be lesser, and the normalized score will be potentially higher, or more likely to get closer to -1 or 1 .

Similar to VADER, the default implementation of sentiment analyzer in TextBlob is based on the pattern.en module [28], which relies on the hand-coded sentiment lexicon to calculate sentiment. However, this lexicon only contains 1,528 adjectives that frequently appear in movie reviews. For example, the sentence “It is fun and unforgettable.” will yield a sentiment polarity 0.55 because the polarity of “fun” and “unforgettable” in our lexicon is 0.3 and 0.8, respectively, and after averaging these two we obtain 0.55.

One reason that yields the insensitivity of TextBlob is that the number of words in its lexicon is only 1,528, which is lesser than the VADER and AFINN, which contain 7,518 and 3,382 words, respectively. When we try to calculate the sentiment using TextBlob, especially in paragraph and sentence-based approaches, only a few adjectives in one sentence can match up with the adjectives in the lexicon, and in some paragraphs/sentences, there might exist some adjectives which cannot be mapped into the lexicon. After we average the polarity score of our matched adjectives, we still need to divide this average score by the number of paragraphs/sentences, and the final sentiment polarity we obtained becomes smaller.

AFINN also relies on the human-tagged sentiment lexicon and tries to map each sentiment-bearing word appearing in the text to the lexicon and then sum up all the sentiment scores to the final sentiment score. After that, we normalize the final result to the range of -1 to 1 , in accordance with VADER and TextBlob. Since the number of words in AFINN’s lexicon is as twice as the number of words in TextBlob, it performs better than TextBlob with regard to the sensitivity since it has higher possibility to match words in the text.

We do not know exactly how the commercial closed source MATA calculates sentiment.

6 Conclusion

We performed sentiment analysis on eleven different events by analyzing news articles from ten different international news sources. Our results demonstrated that most events produced similar sentiment polarities across all news sources, however, there was absolute variance among sentiment score magnitudes across news sources. For a few events, there was significant polarity shift for sentiment scores among news sources. Every news publisher expressed varied sentiment toward individual event,

versus a consistent trend of negative, positive, or neutral sentiment. We were pleased not to find major differences in sentiment polarity between major news sources for the same event which could have pointed to bias and drive tensions between the populations relying on these news outlets for objective reporting. We do acknowledge that this is a simple limited metric which fails to capture the specific differences in message of any two articles even if they share the same sentiment.

There are limitations present in our study. First, some events were not covered by all news sources mentioned in this study. Additionally, some events received little coverage in general (an event may have less than 100 articles in total dedicated to it for instance). In the future, we hope to expand this study to a larger dataset and possibly employ more advanced sentiment analysis techniques.

In addition, we hope to human validate the sentiment underlying the news articles. For instance, we can invite a group of people to read news articles and rate their sentiment toward them, to see the difference between the sentiment calculated by sentiment analysis tool and humans. Furthermore, we could label a part of news articles manually with their sentiment score, and try to apply machine learning techniques to train our own model and thus make predictions.

Acknowledgements The authors would like to thank the Notre Dame Center for Research Computing for financial support and Google who kindly provided a student research grant to utilize the Google Translation API cloud service for translating La Nacion articles into English. Finally, we would like to thank Mrs. Rachel O'Hare, Mr. Richard Jones, and Mrs. Victoria St. Martin for insightful discussions on perspective in journalism.

References

1. Mejova, Y., Srinivasan, P., Boynton, B.: *Gop primary season on twitter: popular political sentiment in social media*. In: *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, pp. 517–526. ACM (2013)
2. Budak, C., Goel, S., Rao, J.M.: *Fair and balanced? Quantifying media bias through crowd-sourced content analysis*. *Public Opin. Q.* **80**(S1), 250–271 (2016)
3. Stimson, J.A., MacKuen, M.B., Erikson, R.S.: *Dynamic representation*. *Am. Polit. Sci. Rev.* **89**(3), 543–565 (1995)
4. Page, B.I., Shapiro, R.Y.: *Effects of public opinion on policy*. *Am. Polit. Sci. Rev.* **77**(1), 175–190 (1983)
5. Seib, P.: *Effects of real-time news coverage on foreign policy*. *J. Confl. Stud.* **20**(2) (2000)
6. Zhang, C.: *International coverage, foreign policy, and national image: exploring the complexities of media coverage, public opinion, and presidential agenda*. *Int. J. Commun.* **6**, 20 (2012)
7. Robinson, P.: *Theorizing the influence of media on world politics: models of media influence on foreign policy*. *Eur. J. Commun.* **16**(4), 523–544 (2001)
8. Tambini, D.: *Fake news: public policy responses*. (2017)
9. Stratfor. *Russia: Duma vote paves way for Russian internet*. <https://worldview.stratfor.com/situation-report/russia-duma-vote-paves-way-russian-internet>. Accessed 26 February 2019
10. *Distinguishing biased and impartial perspectives*. <https://www.canadianmysteries.ca/en/4179.php>. Accessed 26 February 2019
11. Durrani, M., Ali, A., Khan, T., et al.: *Sentiment analysis and the complex natural language*. *Complex Adapt. Syst. Model.* **4**

12. Sentiment bias detection in support of news credibility judgment
13. Pang, B., Lee, L., et al.: Opinion mining and sentiment analysis. *Found. Trends® Inf. Retr.* **2**(1–2), 1–135 (2008)
14. Abdul-Mageed, M., Diab, M., Kübler, S.: Samar: subjectivity and sentiment analysis for Arabic social media. *Comput. Speech Lang.* **28**(1), 20–37 (2014)
15. Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Van Der Goot, E., Halkia, M., Pouliquen, B., Belyaeva, J.: Sentiment analysis in the news (2013). arXiv preprint [arXiv:1309.6202](https://arxiv.org/abs/1309.6202),
16. Bautin, M., Vijayarenu, L., Skiena, S.: International sentiment analysis for news and blogs. In: *ICWSM* (2008)
17. Kaya, M., Fidan, G., Toroslu, I.H.: Sentiment analysis of Turkish political news. In: *Proceedings of the 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology*, vol. 01, pp. 174–180. IEEE Computer Society (2012)
18. Hamborg, F., Donnay, K., Gipp, B.: Automated identification of media bias in news articles: an interdisciplinary literature review. *Int. J. Digit. Librar.* 1–25 (2018)
19. Boschee, E., Lautenschlager, J., O’Brien, S., Shellman, S., Starz, J.: *ICEWS Automated Daily Event Data* (2018)
20. Leetaru, K. Global database of events languages and tone. <https://www.gdeltproject.org>
21. News API. <https://newsapi.org>
22. LexisNexis. Nexis uni. <https://www.lexisnexis.com/en-us/support/nexis-uni/default.page>
23. Bobriakov, I.: Comparison of top 6 python nlp libraries. <https://medium.com/activewizards-machine-learning-company/comparison-of-top-6-python-nlp-libraries-c4ce160237eb>. Accessed 14 July 2018
24. Oelke, D., Geißelmann, B., Keim, D.A.: Visual analysis of explicit opinion and news bias in German soccer articles. In: *EuroVA@ EuroVis* (2012)
25. Hutto, C.J., Gilbert, E.: Vader: a parsimonious rule-based model for sentiment analysis of social media text. In: *Eighth International AAAI Conference on Weblogs and Social Media* (2014)
26. Ribeiro, F.N., Araújo, M., Gonçalves, P., Gonçalves, M.A., Benevenuto, F.: Sentibench—a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Sci.* **5**(1), 23 (2016)
27. Loria, S.: Textblob: simplified text. <https://textblob.readthedocs.io/en/dev/>
28. Computational Linguistics and Psycholinguistics Research Center. <https://www.clips.uantwerpen.be/pages/pattern-en>
29. Nielsen, F.: A new ANEW: evaluation of a word list for sentiment analysis in microblogs (2011). arXiv preprint [arXiv:1103.2903](https://arxiv.org/abs/1103.2903)
30. Microsoft Azure. Text analytics. <https://azure.microsoft.com/en-us/services/cognitive-services/text-analytics/>
31. Araujo, M., Reis, J., Pereira, A., Benevenuto, F.: An evaluation of machine translation for multi-lingual sentence-level sentiment analysis. In: *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, pp. 1140–1145. ACM (2016)