

SOCIAL MEDIA SENTIMENT ANALYSIS OF THE RUSSIA-UKRAINIAN CONFLICT THROUGH NATURAL LANGUAGE PROCESSING: TOWARDS A NEW STRATEGIC STUDIES METHODOLOGY

ANDREW BROSNAN, SHAURYA MALHOTRA, FERGUS JAYES, ANOOP SABU, PETER SYLVESTER, MATHEW MURPHY

ABSTRACT

This paper presents a new methodology in the repertoire of strategic studies practitioners. Social media users generate large amounts of raw data that could provide actionable information in advising decision-making within strategic studies, as evidenced by the high volumes of posts related to the Russia-Ukrainian conflict. However, this potential resource has yet to be leveraged within strategic studies and international relations. This paper outlines how natural language processing can be leveraged to enable social media sentiment analysis, a methodology already in use in a variety of fields to inform decision-making based on large amounts of social media data. As a proof of concept and to outline how the methodology has a low skill and financial barrier of entry, this paper presents five key performance indicators that have been analysed and visualised using natural language processing of over four thousand tweets, collected in the first four weeks of the Russian-Ukrainian conflict. As such, it presents the utility of natural language processing as a methodology for strategic studies practitioners.

Keywords: Natural Language Processing, Social Media Sentiment Analysis, Strategic Studies, Russian-Ukrainian Conflict

INTRODUCTION

The Russian invasion of Ukraine on February 24th 2022, generated a swift yet disjointed global response with significant deviation between the reactions of geopolitical world leaders, military forces on either side and the global community at large. As has been the case throughout most of

recorded history, the views of influential leaders, celebrities, politicians and those with widely accessible platforms of communication have been circulated and recorded through the press and other forms of media (Jiang et al., 2017: 2-3; Zeitzoff, 2017: 2). However, this paper presents a new methodology to augment the toolsets of international relations and geopolitical strategists through the leveraging of natural language processing (NLP) and social media sentiment analysis. In this context, natural language processing refers to the use of computer science, programming and machine learning to analyse larger amounts of natural, ordinary or written language than humans could be expected to analyse manually (Chowdhary, 2020: 604; Raina & Krishnamurthy, 2022: 63-64). Additionally, social media sentiment analysis will be taken to mean the insight that can be derived from using NLP to mine opinions, reactions and feelings from a written text on social media platforms like Facebook, Twitter and LinkedIn among others (Khader et al., 2018: 3; Solangi et al., 2018: 1). This can present a new dimension to strategic studies in which the views and opinions of a silent majority can be collected and analysed into actionable information to inform strategies and decision-making. As a proof of concept, this paper has used NLP to collect, pre-process and analyse over four thousand tweets during the first four weeks of the Russian-Ukrainian conflict. As such, five key performance indicators (KPIs) were created during analysis of the Twitter data as quantifiable measurements to gauge the views of the social media-using public in relation to the conflict (Păvăloaia et al., 2019: 7). These KPIs were visualised into a dashboard to demonstrate how the decision-making process can be optimised through the

aggregation and visualisation of large amounts of data (Appendix 1; Alattar et al., 2021: 61763). This paper does not present a definitive strategy to address the conflict in Ukraine but rather a guideline for strategists to incorporate a wider perspective into their research and strategies. Nevertheless, to outline the capabilities of NLP in strategic studies, this paper presents a sentiment analysis of geopolitical actors during the first month of the conflict, the social media reaction to NATO's response, a heatmap of the perceived influence and sentiment of various stakeholders in the conflict, a visualisation of the countries most implicated in the conflict on social media and a word cloud of the most popular words and phrases of the conflict. Considering the short period of time between the beginning of the conflict, with the extensive time commitments required for the peer review process, this research has had limited success in comparing and contrasting the social media data with analogous academic research. This has limited the quality of discussion with respect to the geopolitical or international relations theoretical implications of the study but still presents a valuable methodology within the repertoire of strategic studies, geopolitics or international relations practitioners.

BACKGROUND LITERATURE: PRELUDE TO THE CONFLICT

The decision for Russia to invade Ukraine in late February 2022 cannot be traced to a singular cause but rather the crescendo of several interlacing geopolitical events. Ukraine had been part of the Soviet Union from 1922 to August 1991 following the collapse of the USSR (Bowker, 2013: 238). This was a major geopolitical blow to the Russian Federation, which had extensively used Ukraine's natural resources like iron ore, natural gas, oil and coal to fuel its economy and extensive military (Stulberg, 2017: 73, 80). Additionally, Ukraine had been a major recruitment source for the Soviet military and its large workforce contributed to a significant portion of Russian economic and political influence (Colás & Pozo, 2011: 213). Ukraine had also provided a crucial

buffer zone between geopolitical rivals in Western Europe and the Russian mainland in addition to enabling access to uncontested key naval territories like the Black Sea and Azov Sea (Treisman, 2016: 47-48). Since 1991, Russia's respective governments have attempted to reintegrate Ukraine within its sphere of influence through political, economic and most notably, military means, as evidenced by the annexation of the Crimean Peninsula in February 2014 (Kuzio, 2018: 5). Treisman presents that in addition to Russia being a threat to Ukrainian security, the Ukrainian state also threatened the Russian Federation from a geopolitical perspective. This stems from the risk of Ukraine joining NATO, potentially enabling the positioning of near-peer rivals close to the Russian border (2016: 48). President Vladimir Putin had several rounds of troop and military equipment migrations to half a dozen zones along the Russia-Ukrainian border from March 2021 until February 2022 (Aris, 2022: 3). This culminated in Russian forces invading Ukraine in late February, generating significant international uproar and entering the global zeitgeist, notably generating high volumes of social media discourse.

METHODOLOGY

In order to highlight the utility and ease of integrating NLP and social media sentiment analysis within strategic studies research, this research pursued a methodology with low barriers of entry in terms of prerequisite programming abilities, IT hardware and software requirements and financial cost. The proposed methodology requires no additional expenditure beyond the cost of a computer with internet access, although further optimisation such as automation, increased storage and machine learning will have commensurate financial expenses. Twitter was the social media platform of choice as it has a large, active user base with access to key data points to inform analysis like geographical information, timestamps and dates among others (Brosnan et al., 2021a: 23). RStudio was the integrated development environment used to collect, process and analyse the tweets using the

TwitterR package as an Application Programming Interface (API) into Twitter (Brosnan et al., 2021b: 2). These APIs were leveraged using libraries like 'twitterR', 'dplyr', 'tm', and 'purrr' as the integrated

development environment supports user viewership and interaction with objects in the environment indefinitely (fig 1).

```

1  ## Packages used for the Twitter Analysis
2  install.packages("purrr")
3  install.packages("twitterR")
4  install.packages("ROAuth")
5  install.packages("stringr")
6  install.packages("RCurl")
7  install.packages("plyr")
8  install.packages("tm")
9  install.packages("SnowballC")
10 install.packages("wordcloud")
11 library(wordcloud)
12 library(purrr)
13 library(twitterR)
14 library(dplyr)
15 library(ROAuth)
16 library(stringr)
17 library(RCurl)
18 library(plyr)
19 library(tm)
20 library(SnowballC)
21 library(ggplot2)

```

Fig 1. These libraries are free to access and significantly decrease the complexity and time requirements of NLP through allowing the leveraging of generic configuration data, documentation, pre-written code and classes among others. This lowers the barrier of entry associated with social media data analysis.

This can be a major benefit when collecting social media data as the same scripts can be used each week to reduce the number of variables that may influence the data collected. This enables a more accurate view of social media sentiment without any influence from external variables (Sarlan et al., 2014: 215). With respect to collecting tweets, it was decided to disallow retweets as they had the potential to skew the results by outliers in addition to the research finding value in collecting only original tweets to gain as rounded a perspective as possible (Giachano & Crestani, 2016: 5, 13). There was no geocode applied to the script meaning that data could be collected from any country with access to Twitter (Zhang & Gelernter, 2014: 39). However,

the code excluded tweets in languages other than English as the lexicon used to analyse the tweets only included English words. The majority of analysis was conducted by keyword searches like '#Ukraine' or '#NATO' and tweets were collected at a rate 1,200 a week over four weeks between February 28 to March 21, 2022. Data was collected at 6pm GMT each Monday to ensure consistency. Following collection, a function was created to clean and mine the data in order to remove any inaccurate, incomplete or otherwise unusable data resulting in a sample of 4,157 tweets. This function also assisted with the sentiment analysis of the tweets through the use of a lexicon of 4,783 negative words with 2,006 positive words to give a sentiment score to each tweet (fig 2).

```

46 # List of Positive and NEgative words
47 pos.words <- scan('/Users/shauryamalhotra/Desktop/positive.words.txt', what = 'character', comment.char = ';')
48 neg.words <- scan('/Users/shauryamalhotra/Desktop/neg.words.txt', what = 'character', comment.char = ';')

```

neg.words	chr [1:4783] "2-faced" "2-faces" "abnorm...
pos.words	chr [1:2006] "a+" "abound" "abounds" "ab...

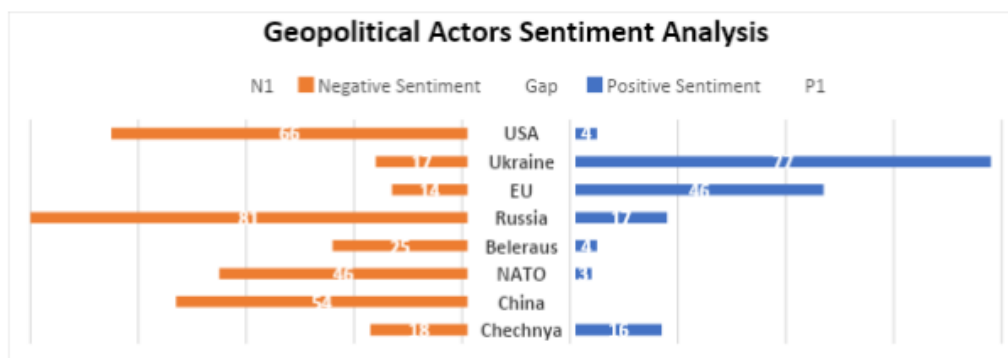
Fig 2. This lexicon is free to integrate into RStudio and was significantly more accurate and time efficient than manual analysis.

The function provided a score of +1 point for every positive word and -1 point for every negative word in a tweet. With regards to visualising the data after analysis, Google Sheets was chosen as it is free to access with a standard Google account although the data could have also been visualised within RStudio or exported to a more specialised platform

like Microsoft Power BI or Tableau. This enabled the formation and visualisation of the KPIs and dashboard discussed herein.

SOCIAL MEDIA SENTIMENT KPIS

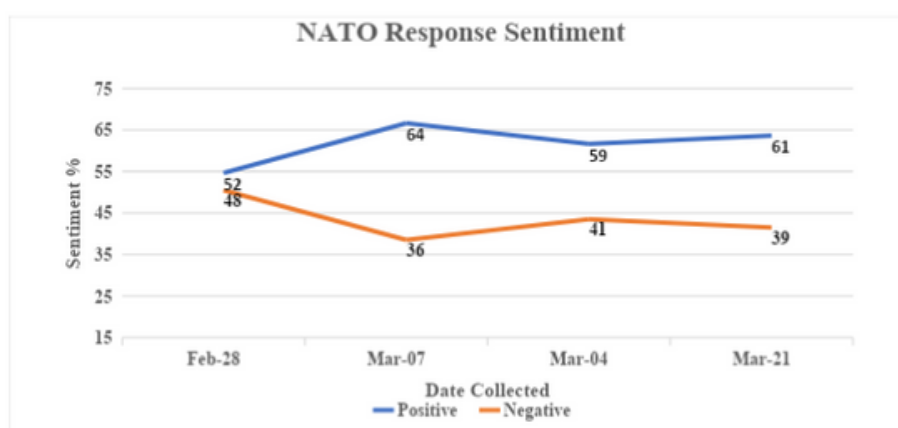
KPI 1: Geopolitical Actors Sentiment Analysis



The analysed tweets highlight that the Russian-Ukrainian conflict is not isolated to two countries but envelops several global powers, nation-states and institutions. NLP analysis indicates that different geopolitical actors are perceived in different favour by the global community; Ukraine is viewed in a resoundingly positive light, while Russia in an almost entirely negative one. Interestingly, sanctions were well received in relation to tweets directed toward the EU but were viewed less favourably when mentioning the US. The US received a significant portion of the negative sentiment with both the entertaining of US military intervention and neutrality in the conflict being treated with almost the same levels of criticism by the social media community. This can also be seen in relation to NATO, with tweets discussing the mobilisation of NATO forces being treated harshly as an escalation to world war, while perceived inaction in not

supporting the Ukrainians has also garnered severe derision. The concept of nations taking in refugees within their borders received overall very positive sentiment with EU nations, such as Poland and Ireland, gaining support for their efforts so far. On the whole however, positive sentiment appears to generally be directed towards geopolitical actors that support Ukraine while negative sentiment is directed toward pro-Russian actors. However, the impacts of only analysing tweets written in English should not be forsaken considering that the majority of people involved in the conflict do not speak English as a first language and so the majority of people directly impacted by the conflict could not be reflected in the data. The data also indicated fluctuations in sentiment over the four week period as seen with NATO.

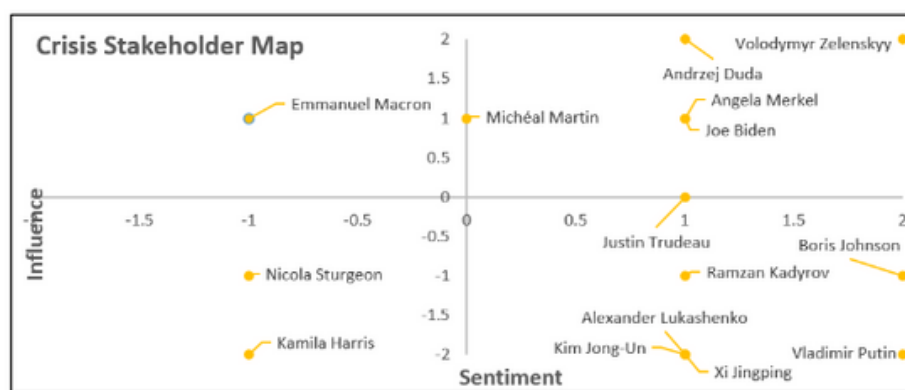
KPI 2: NATO Response Sentiment



Sentiment during the Russia-Ukrainian conflict was not fixed with significant increases and decreases in positive and negative opinions of nations, stakeholders and tactics on a weekly if not daily basis. As an example, in the first days of the conflict, NATO received almost as many negative tweets as positive ones with some social commentators critical of NATO's perceived inaction and inability to predict such a conflict from occurring. Additionally, other tweets speculated that had NATO included Ukraine within their alliance earlier, Russia would not have invaded. However, NATO received marginally more positive sentiment with many tweets presenting that Russia would not risk expanding beyond Ukraine as it would demand military action from neighbouring NATO powers. Some Twitter users were adamant that the economic and military capabilities of NATO would be enough to deter Russian aggression beyond the Ukrainian border. By the second week, NATO sentiment increased with political sanctions and strong denunciation of Russia by NATO leaders being received favourably by the global community.

However, the mobilisation of US troops in Germany was not received well. This indicates that the wider community was not in favour of NATO military intervention in the conflict. This sentiment largely extended to week three with tweets indicating that the social media community were assured that the conflict was being contained within Ukraine, minimising the need for NATO intervention. Additionally, Ukrainian resistance and publicised victories using NATO equipment was received favourably although the perceived inaction in not abating the column of Russian supply and military vehicles en route to Kyiv was not received well. In the final week of analysis, the news of NATO sending anti-air munitions and other equipment to support the Ukrainian resistance was mostly met with positive sentiment indicating that as a whole, NATO is viewed most positively when it supports Ukraine through political action and supplying the resistance as opposed to overt military intervention.

KPI 3: Crisis Stakeholder Map



The stakeholder map evaluates the perceived influence of individuals involved in the conflict with the sentiment that could be recorded from their tweets. As such, a score of (2, 2) would denote an individual with a high level of perceived influence over the conflict who was viewed favourably by Twitter users. In this case, one individual received such a score, Volodymyr Zelensky was the President of Ukraine during the first month of the conflict and he has been credited with rallying the international community while enabling the resistance in Ukraine. Other individuals with high sentiment scores but slightly fewer degrees of influence over the conflict were Poland's President Andrej Duda, German

Chancellor Angela Merkel and the US President Joe Biden who all denounced Russia and have been in support of their countries taking in Ukrainian refugees. Individuals with a strong degree of influence but low support online were French President Emmanuel Macron and Irish Taoiseach Michéal Martin, indicating that both may not be doing enough to support the war effort. France is a member of NATO which indicates that Twitter users may have wanted more action from the nation whereas Ireland has traditionally been neutral throughout its history which has been met with some negative sentiment from social media users, although the Irish

Taoiseach is viewed more favourably than Macron. With respect to individuals with high degrees of influence over the conflict but low support online, Vladimir Putin received the strongest score at (2, -2) for his role as President of Russia in the conflict. Individuals aligned with Russia like Alexander Lukashenko of Belarus and Ramzan Kadyrov of Chechnya also received comparable scores. This can also be seen with the leaders of North Korea and the People’s Republic of China, although this largely stems from fears of involvement in a hypothetical World War Three as opposed to connections to Russia. Prime Minister of the UK, Boris Johnson also received a negative score with tweets indicating that

the UK should be doing more to support the Ukrainian military and its people. The final category was individuals with low degrees of influence and low levels of support, namely Nicola Sturgeon of Scotland and Kamala Harris, the Vice President of the US. This indicates that both leaders are viewed as ineffective online which in the case of Harris, in particular, presents a major vulnerability as the US is one of Ukraine’s largest allies in the conflict. The stakeholder map is a valuable resource in visualising the impacts of individuals associated with the conflict. The data also indicates the country’s involvement at a global scale.



KPI 4: Crisis Stakeholder Map

The Russia-Ukrainian conflict encapsulated many states beyond the two involved in the fighting. The heatmap visualises the number of times a country or their leader was named in relation to the conflict with darker shades of orange denoting the most active nations. Unsurprisingly, Russia and Ukraine were the most active with NATO states like the US, Canada, UK, Australia, France and Germany also implicated within the conflict. Generally, countries aligned with Russia were not as often mentioned in relation to the conflict, indicating that Russia may be viewed as a sole entity in the conflict as opposed to the global coalition siding with Ukraine. Countries like China, India, Iran and North Korea were mentioned in relation to the conflict although this

generally was as a result of theorising how the conflict would develop into World War Three as opposed to denoting the involvement of these states in the Russia-Ukraine conflict. Areas without collected Twitter data were localised in large regions like South America, Africa and South-East Asia which highlights how the conflict had not yet become a completely global issue as of the four week period in which the social media data was collected.

- Brosnan, A. Malhotra, S. Murphy, M. Sabu, A. and Sylvester, P. (2021b) 'A Social Media Sentiment Analysis on President Joseph R. Biden's Presidency: Issues To Be Addressed and Recommendations For The Future', University College Cork White Paper, pp. 1-17.
- Brosnan, A. (2021c) 'Digitalising Defense: The Potential Role of Digital Transformation in Augmenting European Military Doctrine', *European Studies Review*, 1(6), pp. 52-61.
- Chowdhary, K. (2020) 'Natural Language Processing,' in Chowdhary, K. (ed), *Fundamentals of Artificial Intelligence*, New Delhi: Springer.
- Colás, A. and Pozo, G. (2011) 'The Value of Territory: Towards a Marxist Geopolitics', *Geopolitics*, 16(1), pp. 211-220.
- Giachanou, A. and Crestani, F. (2016) 'Like It or Not: A Survey of Twitter Sentiment Analysis Methods', *ACM Computing Surveys*, 49(2), pp. 1-41.
- Khader, M. Awajan, A. and Al-Naymat, G. (2018) 'The Effects of Natural Language Processing on Big Data Analysis: Sentiment Analysis Case Study', *In: 2018 International Arab Conference on Information Technology (ACIT)*, Werdanye, Lebanon, November 28-30, pp. 1-7.
- Kuzio, T. (2018) 'Russia-Ukraine Crisis: The Blame Game, Geopolitics and National Identity', *Europe-Asia Studies*, 70(3), pp. 462-473.
- Jiang, H. Luo, Y. and Kulemeka, O. (2017) 'Strategic Social Media Use in Public Relations: Professionals' Perceived Social Media Impact, Leadership Behaviors, and Work-Life Conflict', *International Journal of Strategic Communication*, 11(1), pp. 18-41.
- Păvăloaia, V. Teodor, E. Fotache, D. and Danileț, M. (2019) 'Opinion Mining on Social Media Data: Sentiment Analysis of User Preferences', *Sustainability*, 11(16), pp. 1-21.
- Raina, V. and Krishnamurthy, S. (2022) *Building an Effective Data Science Practice: A Framework to Bootstrap and Manage a Successful Data Science Practice*. Berkley, California: Apress.
- Sarlan, A. Nadam, C. and Basri, S. (2014) 'Twitter Sentiment Analysis', *In: 2014 International Conference on Information Technology and Multimedia*, Putrajaya, Malaysia, November 18-20, pp. 212-216.
- Solangi, Y. Solangi, Z. Aarain, S. Abro, A. Mallah, G. and Shah, A. (2018) 'Review on Natural Language Processing (NLP) and Its Toolkits for Opinion Mining and Sentiment Analysis', *In: 2018 IEEE 5th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, Bangkok, Thailand, November 22-23 2018, pp. 1-4.
- Stulberg, A. (2017) 'Natural gas and the Russia-Ukraine crisis: Strategic restraint and the emerging Europe-Eurasia gas network', *Energy Research & Social Science*, 24, pp. 71-85.
- Treisman, D. (2016) 'Why Putin Took Crimea: The Gambler in the Kremlin', *Foreign Affairs*, 95(3), pp. 47-54.
- Zhang, W. and Gelernter, J. (2014), *Geocoding Location Expressions in Twitter Messages: A Preference Learning Method*, *Journal of Spatial Information Science*, 9, pp. 37-70.
- Zeitzoff, T. (2017) 'How Social Media Is Changing Conflict', *Journal of Conflict Resolution*, 61(9), pp. 1970-1991.

