A First Attempt to Detect Misinformation in Russia-Ukraine War News through Text Similarity

Nina Khairova[1,2✉] and **Bogdan Ivasiuk**[3] and **Fabrizio Lo Scudo**[4]
Carmela Comito[5] and **Andrea Galassi**[3]
[1] Umeå University, Sweden
[2] National Technical University "Kharkiv Polytechnic Institute" , Ukraine
[3] DISI University of Bologna, Bologna, Italy
[4] DEMACS, University of Calabria, Rende, Italy
[5] ICAR-CNR, Rende, Italy
ninakh@cs.umu.se

Abstract

The paper focuses on misinformation detection in established global news outlets' texts covering significant and well-known events of the Russian-Ukraine war. We created the RUWA dataset and applied unsupervised ML approaches as the first dimension of misinformation detection. We consider several different aspects of semantic similarity identification of the articles from various regions in order to confirm the hypothesis that if the news covering the same event from the outlets of various regions over the world are similar enough it means they reflect each other or, instead, if they are completely divergent it means some of them are likely not trustworthy.

1 Introduction

Since the 2016 U.S. presidential election, passing through the U.K. Brexit referendum and the COVID-19 pandemic, misinformation is becoming one of the more significant problems of Modern Society (Zhou and Zafarani, 2020). Two major reasons for this relate to the huge amount of people relying mainly on online sources to get their information and news and the high speed of information spreading via the Internet. Large-scale misinformation campaigns carried out by a big corporation, a political party, or even a government of a certain country can affect various social, economic, and political events. Usually, these kinds of campaigns involve socially sensitive domains such as elections, coronavirus, or military operations and can not only threaten public security and social stability but even affect the results of elections and wars. This has been especially evident in the coverage of the current Russia-Ukraine war when misinformation has become a part of an information war and propaganda activities. The information warfare strategy has a twofold goal: the first is to manipulate the attitudes of people directly involved in the war, and the second is to modify societies' opinions of other countries (Thomas, 2014; Theohary, 2018).

To this effect, since the beginning of the Russian invasion of Ukraine, misleading information has been spreading online on social media and by many media outlets. Wide dissemination of misinformation was made possible by two main factors: assessing the truthfulness of facts is highly complex to war events, and news outlets are often inclined to lower the bar of the fact-checking process to provide information as quickly as possible. (Claudia et al., 2021). In this context, careful human-made fact-checking is thus not always possible. However automatic misinformation detection can not always help as well due to lacking labeled benchmark datasets of the particular domain, which relates to the war or military conflicts. While previous works regarding the automatic detection of misinformation do exist, they typically address specific domains, and to the best of our knowledge, little progress has been made regarding the domain of armed military conflicts.

The aim of our work is to analyze and compare news from several established outlets in an unsupervised fashion. Drawing similarities and differences between sources could facilitate future work on fact-checking aimed at establishing finding patterns of reliability of sources and information truthfulness. Specifically, we compare full texts, titles, meaningful sentences, and perform a sentiment analysis.

For this purpose, we create a novel dataset of news in English related to the Russian-Ukrainian war and release it publicly.[1] While we observed some relevant patterns and similarities, our results are not conclusive. Moreover, we find that the number of articles available for each source and

[1] https://github.com/ninakhairova/dataset_RUWA

the length of such articles strongly influence the outcome. Nonetheless, we deem our results useful for future works on this topic.

2 Related work

Machine Learning and Deep Learning methods require a large amount of labelled data to effectively train. Applying automatic misinformation detection approaches based on supervised machine learning methods is reasonably common (Capuano et al., 2023; Agrawal et al., 2021). However, in order to use the methods that provide good results it is necessary to train them on specific domain data, which are not available in this context.

We can distinguish several major approaches to misinformation labeling. Most of existed labeled datasets containing political news and some other kinds of news are manually labeled (Silverman et al., 2016; Wang, 2017) or utilize fact-checking websites such as PolitiFact or GossipCop (Shu et al., 2020). For instance, a corpus that is described in Choudhary and Arora (2021) comprises 1,627 articles that were manually fact-checked by professional journalists from BuzzFeed. In some cases, the real news was extracted from a special group of trustworthy sources, while the fake news was extracted from sources of the fake news list like "Business Insider’s Zimdars Fake news list" (Janicka et al., 2019). One more approach to annotating the fake news dataset was applied to the AMT dataset (Potthast et al., 2018), which contains 480 articles annotated as fake and true. While fake news articles were imitated by journalists intentionally, the real news was obtained from outlets of several domains.

In general, there are only a few labeled misinformation detection datasets that cover war and military topics (Salem et al., 2019). Furthermore, designing such kind of dataset becomes a much more challenging task due to the fact that the dataset must be created during the ongoing war when actual fact-checking is impossible, there is a good chance of the existence of a bias in various information sources, and so-called “fog of war” effect always can be inherent.

3 Data

Following the requirements of fake news corpus information balance (Rubin et al., 2016; Golbeck et al., 2018), We create a novel dataset called "RUWA" (Russian-Ukraine WAr), composed of several media outlets from Ukraine, Russia, European, Asia, and the USA. We selected nine of the most information-significant events of Russia’s Invasion of Ukraine and aligned articles in English language from all the outlets according to these events. The list of events includes widely-known events such as "The Bucha massacre" and "Sinking of the warship Moskva". To collect articles from the selected news outlets we applied a keyword-based research strategy, conditioned by specific time intervals and topic classification of the sites rubrics. We identified about 100 keywords, which range from geographical names (e.g., Bucha or Olenivka), specific buildings names (e.g., Kramatorsk train station or Mariupol theatre), organizations names (e.g., Red Cross), prominent individual names (e.g., Zelenskyi, Putin), to proper nouns and phrases (e.g., Nuclear Power Plant).

Currently, the RUWA dataset includes more than 16,500 news articles covering the Russian-Ukraine war events that occurred from February 2022 to September 2022. Table 1 shows the article distributions by selected news outlets and events.

4 Methodology of Analysis

Being aware of the complexity of assessing the truthfulness of facts for war events in the absence of the necessary resources to carry out a *journalistic*-oriented process of fact-checking, we decide to relax the problem to assess the veracity of reported facts. We assume that the news reported by news outlets located in the two countries that are directly involved in the conflict can be expected to be highly different. Discrepancies can be substantial up to the point of denying events such as a bombing of residential areas or civilian killings. Additionally, we assume that even though events reported by selected trustworthy independent news agencies and media should be accurate, however, their narrative perspective can remain not neutral.

Thus, as the first dimension of analysis, we focus on textual similarity, comparing the news and assessing if they have a similar meaning. We want to establish whether the news covering the same event from the outlets of various regions over the world are similar enough to indicate they reflect each other or, instead, they are completely divergent and consequently some of them are likely, not trustworthy. We will consider and aggregate several similarity measures that represent many different aspects (Hövelmeyer et al., 2022).

Source	Azovstal	Beginning	Bucha	Nuclear Plant	Prisoners	Railway	Moskva Sinking	Supermarket	Mariupol Theatre	Total
Al Jazeera	23	143	79	186	31	16	34	32	56	600
BBC	22	137	34	236	22	16	17	41	25	550
Censor.Net	826	1730	397	747	117	749	31	173	324	5094
News Front	10	28	18	16	7	7	5	2	1	94
NBC News	8	155	86	129	13	29	36	13	37	506
Reuters	68	924	143	649	23	16	38	15	133	1993
Russia Today	32	14	102	485	236	12	15	22	22	940
Ukrinform	827	3359	570	925	129	601	22	153	163	6749
Total	1816	6490	1429	3373	578	1468	175	436	761	16526

Table 1: RUWA Dataset Selected News outlets and War events

4.1 The similarity between articles based on pre-trained vectors

As the first dimension of analysis, we focus on pairwise evaluating the semantic similarity of all outlets' articles, aggregating all the articles from the same source as a single textual document. As textual encoder, we use FastText (Mikolov et al., 2018).

4.2 Similarity between the title of articles

Authors and correspondents of news agencies and media try to aggregate a major idea of an article, its narrative, or its specific message in the title. Therefore, we analyze similarities between articles over the same topic and use a hierarchical method to aggregate them into similarities between sources. We match each title of every article covering the particular event of the one source with comparable articles titles of the other source. Then we average the similarity scores of titles of two sources that cover the same event and thus we obtain a score similarity for the higher level of the hierarchy, namely for two sources. More formally, our purpose is to obtain a measure of similarity between two sources based on sets of articles titles covering the same event

4.3 Similarity between semantically meaningful sentences

Even if news articles carry different narratives, and contain different informational messages, their semantic similarity score based on the semantics of words or even semantics sentences, can be close enough. Obviously, this is due to the fact that all news articles include a lot of close-meaning sentences or phrases like "correspondent claimed" or 'it seems not obvious' and so on. In order to compare more semantically concentrated texts that only focus on the information of a particular event we extract sets of sentences from all articles of a source that describe only military and close-to-military actions regarding this particular event.

We utilize two approaches to compare the semantic similarity of such kinds of sentences. In the first one, we process only the sentences that contain keywords related to the considered event. For the second, we add additional knowledge via the lists of verbs that represent the actions involved in certain events. In order to generate such lists, we primarily based on the open list of words associated with the Russian-Ukrainian war from Solopova et al. (2023) and supplemented it with the verbs obtained from the articles. We selected only verbs that relate to a military domain and a given event from all the verbs extracted from the texts. For instance, for the "Moskva sinking" event the list of verbs related to the event includes more than 120 verbs. We also experiment with pre-processing, namely stemming and stop word removal.

4.4 Sentiment Analysis

Given an event for each media outlet, we compute the sentiment analysis for each article concerning that event. We performed sentence-level sentiment analysis and computed the article's overall sentiment by averaging the sentiment of every single sentence. Sentiment analysis has been performed using a statistical approach based on a Convolutional Neural Network for Sentence Classification (Kim, 2014) provided within the NLP toolkit STANZA (Qi et al., 2020).

Due to the linguistic journalist style and jargon, most sentences used within the articles do not provide valuable insights. Hence, we perform a preliminary step and restrict our analysis to a subset of all sentences we consider more informative. To assess the informativeness of a sentence, we employ a keyword-based approach. For each event, we collect all the articles related to that event and rely on TF-IDF to identify the most "significant" words. Then, we maintain only the sentences containing the extracted keywords for each article.

5 Results and discussion

5.1 Leveraging the pre-trained vectors

The experiment confirms our hypothesis. It shows that the semantic similarities between the outlets' texts of countries involved in the conflict (e.g., Censor.net and RT) and websites articles texts of other countries (e.g., Reuters and The Guardian) are less than the similarity of all other considered sites among themselves for almost all events. Also, the semantic similarity coefficients do not have a significant difference, ranging from 91% to 99%.

This can be explained primarily by the special military topic of the news, which is not stipulated by the lexis of the linguistic models. In addition, articles covering the same events may produce different narratives or real and fake facts, but their semantics remain the same.

Table 2 shows the pairwise cosine semantic similarity coefficients for articles of all outlets for the "Sinking of the Moskva" topic based on fastText's subword pre-trained vector from Facebook AI.

5.2 The articles headlines comparison

Leveraging the pre-trained FastText model for headlines' semantic similarity score calculation produces more distributive semantic similarity scores than for full-text articles. However, we observe that the headlines of articles on the same topic and belonging to the same outlet also produce relatively low similarity values, so we can not regard this approach as accurate.

Table 3 shows the example of the distribution of the pairwise cosine semantic similarity coefficients for articles headlines of all outlets for the "Sinking of the Moskva" topic.

We assume that there are a few reasons for this. First of all, the result of handling the titles of the articles depends on the size of the dataset even more than the processing of the articles' full texts. However, in the case of some websites for some events, we do not have a large number of articles (Table 1). Secondly, the effectiveness of the approach based on the semantic similarity of titles may depend on the quality and informativeness of the headlines themselves and their compliance with a particular event. But based on the considered domain we can assume that titles often not only call or describe an event but also reflect the ongoing tensions that can include the authors' biased opinions and feelings.

5.3 Use of extra knowledge for semantic similarity detection

As we mentioned in Section 4.3, we utilize keywords and military action verbs to supplement semantic similarity calculation with additional knowledge about an event. Leveraging sentences that contain keywords related to the considered event enables producing more specific and directly related to the subject of the event texts. However, this inevitably entails losing a large amount of information. Using extra knowledge via the lists of verbs that represent the actions involved in certain events allows us to determine the semantic similarity of news articles, focusing more on the semantic content of articles regarding a particular event. Table 4 shows the example of the semantic similarity for selected sentences that include action verbs for the "Sinking of the Moskva" topic.

The last experiment most explicitly confirms our hypothesis that the semantic similarity coefficient between established outlets of countries involved in the war from two different sides is the smallest. Consequently, we can assume that the value of the semantic similarity coefficient can correlate with producing some other information about the same event that can be identified as misinformation

5.4 Sentiment Analysis

As described in Section 4.4, we perform the sentiment analysis of each document at the sentence level. This is due to the issues Sentiment analysis tools have when working at the document level (Behdenna et al., 2018). In an attempt to mitigate such issues, we decided to perform our analysis at the sentence level and collect the result by simply counting the occurrences for the three classes: *Negative*, *Neutral*, and *Positive*. For each source, we thus aggregate the sentiment counting over all the sentences of the collected articles that focus on a specific event. In Table 5, we report the sentiment analysis made with the NLP toolkit STANZA for the event "Sinking of the Moskva".

Table 5 shows that most Neutral sentences are a common trait among all the sources. That is an expected result due to the journalistic nature of the analyzed documents, which might also be considered a potential noise source for any downstream task. We thus attempted to mitigate that by restricting our analysis to only the sentences that report event-specific keywords, assuming that such sentences would be more suitable to contain potential

	The Guardian	Reuters	Al Jazeera	Censor	CNN	Ukrinform	Russia Today
The Guardian	100%	99.7%	99.9%	94.7%	99.9%	99.8%	99.5%
Reuters	99.7%	100%	99.7%	94.8%	99.6%	99.6%	99.6%
Al Jazeera	99.9%	99.7%	100%	94.5%	99.9%	99.7%	99.4%
Censor	94.7%	94.8%	94.5%	100%	94.8%	95.1%	93.5%
CNN	99.9%	99.6%	99.9%	94.8%	100%	99.8%	99.3%
Ukrinform	99.8%	99.6%	99.7%	95.1%	99.8%	100%	99.3%
Russia Today	99.5%	99.6%	99.4%	93.5%	99.3%	99.3%	100%

Table 2: The semantic similarity for articles of all outlets for the “Sinking of the Moskva” topic based on fastText’s pre-trained vectors

	The Guardian	Reuters	Al Jazeera	Censor	CNN	Ukrinform	Russia Today
The Guardian	75.6%	72.5%	72.5%	74.3%	74.5%	70.7%	69.7%
Reuters	72.5%	77.5%	69.5%	77.2%	72.2%	71.5%	75.5%
Al Jazeera	72.5%	69.5%	70.7%	75.0%	68.9%	64.6%	70.3%
Censor	74.3%	77.2%	75.0%	92.5%	75.2%	66.7%	80.9%
CNN	74.5%	72.2%	68.9%	75.2%	76.6%	70.5%	70.5%
Ukrinform	70.7%	71.5%	64.6%	66.7%	70.5%	81.7%	68.5%
Russia Today	69.7%	75.5%	70.3%	80.9%	70.5%	68.5%	97.5%

Table 3: The semantic similarity for articles headlines of all outlets for the “Sinking of the Moskva” topic

misinformation. We report the results in Table 6. We hypothesize that such a sentence subset could provide more representative information to assess potential source polarization.

6 Conclusion

Creating high-quality resources about a controversial topic such as the Russian-Ukrainian war is a challenging task. In this work, we presented a novel dataset about the conflict, by identifying specific events and imposing a set of constraints on the selection of articles. In our view, such constraints should guarantee a better semantic alignment among articles from news sources, which in turn should facilitate subsequent tasks, such as media bias and misinformation detection. Such a dataset can provide a rich perspective of the different journalistic narrations of the Russian-Ukrainian war and support future research.

Additionally, as a first attempt to detect misinformation in Russia-Ukraine war news, we applied the text similarity approach and Sentiment Analysis. We analyzed the advantages and disadvantages of several approaches to comparing the semantic similarity of news covering the same event in various established outlet news sources.

We also we demonstrated that even though sentiment analysis alone may not be sufficient for misinformation detection, it can provide useful insights that can be combined with other techniques to improve detection accuracy.

We hope that our study contributes to the further development of unsupervised ML approaches to misinformation detection in established outlets news articles.

Acknowledgements

This work was supported by the EU H2020 ICT48 project “Humane AI Net" under contract #952026.

References

Chetan Agrawal, Anjana Pandey, and Sachin Goyal. 2021. A survey on role of machine learning and nlp in fake news detection on social media. In *GUCON*, pages 1–7. IEEE.

S. Behdenna, F. Barigou, and G. Belalem. 2018. Document level sentiment analysis: A survey. *EAI Endorsed Transactions on Context-aware Systems and Applications*, 4(13):e2.

Nicola Capuano, Giuseppe Fenza, Vicenzo Loia, and Francesco David Nota. 2023. Content based fake news detection with machine and deep learning: a systematic review. *Neurocomputing*.

Anshika Choudhary and Anuja Arora. 2021. Linguistic feature based learning model for fake news detection and classification. *Expert Systems with Applications*, 169.

Padovani Claudia, Giuliano Bobba, Baroni Alice, Marinella Belluati, Cecilia Biancalana, Bomba Mauro, Fubini Alice, Marrazzo Francesco, Rega Rossella, Ruggiero Christian, et al. 2021. Italy: A highly regulated system in search of equality. *The Media for Democracy Monitor 2021*, pages 315–386.

	The Guardian	Reuters	Al Jazeera	Censor	CNN	Ukrinform	Russia Today
The Guardian	-	16.8%	40.9%	18.6%	24.7%	25.2%	17.6%
Reuters	16.8%	-	14.7%	17.6%	10.1%	7.0%	19.4%
Al Jazeera	40.9%	14.7%	-	15.5%	24.6%	21.5%	16.4%
Censor	18.6%	17.6%	15.5%	-	7.2%	9.3%	8.1%
CNN	24.7%	10.1%	24.6%	7.2%	-	42.8%	9.7%
ukrinform	25.2%	7.0%	21.5%	9.3%	42.8%	-	11.5%
Russia Today	17.6%	19.4%	16.4%	8.1%	9.7%	11.5%	-

Table 4: The semantic similarity for selected sentences including action verbs for the "Sinking of the Moskva" topic

Source	Articles	Sentences	Negative (%)	Neutral (%)	Positive (%)
Al Jazeera	34	2501	25.83	72.53	1.64
BBC	17	545	30.64	65.32	4.04
Censor	31	314	26.43	69.43	4.14
News Front	5	264	28.03	68.56	3.41
Reuters	15	253	33.99	64.03	1.98
Russia Today	15	463	27.43	68.25	4.32
Ukrinform	22	300	25.0	71.33	3.67

Table 5: Sentiment analysis results for the event "Sinking of the Moskva".

Source	Articles	Sentences	Negative (%)	Neutral (%)	Positive (%)
Al Jazeera	34	328	31.1	67.99	0.91
BBC	17	96	36.46	57.29	6.25
Censor	31	26	34.62	65.38	0.0
News Front	5	12	33.33	66.67	0.0
Reuters	15	14	71.43	28.57	0.0
Russia Today	15	19	63.16	36.84	0.0
Ukrinform	22	27	40.74	59.26	0.0

Table 6: Sentiment analysis on the TF-IDF filtered sentences for the event "Sinking of the Moskva".

Jennifer A. Golbeck, Matthew Louis Mauriello, Matthew Louis Mauriello, Keval H Bhanushali, Christopher Bonk, Mohamed Amine Bouzaghrane, and Cody Buntain. 2018. Fake news vs satire: A dataset and analysis. In *WebSci*, volume 5, pages 17–21.

Alica Hövelmeyer, Katarina Boland, and Stefan Dietze. 2022. Simba at checkthat!-2022: Lexical and semantic similarity based detection of verified claims in an unsupervised and supervised way. In *Working Notes of CLEF*, volume 3180, pages 511–531. CEUR-WS.org.

Maria Janicka, Maria Pszona, and Aleksander Wawer. 2019. Cross-domain failures of fake news detection. *Computación y Sistemas*, 23:1089–1097.

Yoon Kim. 2014. Convolutional neural networks for sentence classification.

Tomas Mikolov, Edouard Grave, Piotr Bojanowski, Christian Puhrsch, and Armand Joulin. 2018. Advances in pre-training distributed word representations. In *LREC*.

Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2018. A stylometric inquiry into hyperpartisan and fake news. In *ACL*, pages 231–240. Association for Computational Linguistics.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A python natural language processing toolkit for many human languages.

Victoria Rubin, Niall Conroy, Yimin Chen, and Sarah Cornwell. 2016. Fake news or truth? using satirical cues to detect potentially misleading news. In *Proceedings of the Second Workshop on Computational Approaches to Deception Detection*, pages 7–17. Association for Computational Linguistics.

Fatima K. Abu Salem, Al Feel Roaa, Shady Elbassuoni, Jaber Mohamad, and Farah May. 2019. Fa-kes: A fake news dataset around the syrian war. In *ICWSM*, volume 13, pages 573–582.

Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2020. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data*, 8(3):171–188.

Craig Silverman, Lauren Strapagiel, Hamza Shaban, Ellie Singer-Vine, and Rong Jeremy. 2016. Hyperpartisan facebook pages are publishing false and misleading information at an alarming rate. In *Buzzfeed News*.

Veronika Solopova, Oana-Iuliana Popescu, Christoph Benzmüller, and Tim Landgraf. 2023. Automated multilingual detection of pro-kremlin propaganda in newspapers and telegram posts. *Datenbank-Spektrum*, pages 1–10.

Catherine A Theohary. 2018. Information warfare: Issues for congress. *Congressional Research Service*, pages 7–5700.

Timothy Thomas. 2014. Russia's information warfare strategy: Can the nation cope in future conflicts? *The Journal of Slavic Military Studies*, 27(1):101–130.

William Yang Wang. 2017. "liar, liar pants on fire": A new benchmark dataset for fake news detection. In *ACL*, pages 422–426, Vancouver, Canada. Association for Computational Linguistics.

Xinyi Zhou and Reza Zafarani. 2020. A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, 53(5):1–40.