Misinformation Detection: Using Linguistic Cues

Astrid Krickl[1], Sabrina Kirrane[1]

[1]*Vienna University of Economics and Business, Institute for Information Systems & New Media, Vienna, Austria*

Abstract

Misinformation could potentially have severe consequences for society, ranging from healthcare to politics. In order to address the negative impact of misinformation, there is a need for tools and technologies that can automatically identify misinformation. Towards this end, we examine the effectiveness of seven different linguistic cues with respect to three datasets. Our results show that some linguistic cues proposed in the literature have a tenuous relationship to either true or false articles.

Keywords

Misinformation, Linguistics, Fake News, Identification

1. Introduction

Although misinformation is not a new phenomenon, with the prevalence of social media platforms false information can be spread further and faster. Consequently, web based misinformation can have a much broader impact, as evidenced by the US election, where people were injured during protests, or during the Covid-19 pandemic, where medical myths about potential cures for Covid-19 lead to poisonings and toxic exposures. Given that manual fact-checking is a challenging, time consuming task that depends on pro-activeness in terms of social media platform providers and/or users, there is a need for tools and technologies that can detect misinformation automatically. In this regard, researchers have demonstrated the potential of feature based learning approaches (c.f., [1, 2]). However, the proposed approaches are not yet accurate enough to serve as a reliable means for distinguishing between real and fake. Here the field of linguistics, which is tightly connected to the area of semantics, could potentially improve the status quo [3, 4, 5]. In order to better understand the potential of linguistic approaches, in this paper, we (i) propose a workflow for detecting misinformation based on seven different linguistic cues applied to three different datasets; (ii) assess the effectiveness of the different cues over the different datasets; and (iii) identify open research challenges and opportunities.

2. Related Work

Current misinformation researchers suggest using features that indicate text readability, such as the average use of long words and sentences [3], complex words [5], and repetition [4], to

SEMANTICS 2022 EU: 18th International Conference on Semantic Systems, September 13-15, 2022, Vienna, Austria

Acknowledgements. Sabrina Kirrane is funded by the FWF Austrian Science Fund and the Internet Foundation Austria under the FWF Elise Richter and netidee SCIENCE programmes as project number V 759-N.

astrid.krickl@wu.ac.at (A. Krickl); sabrina.kirrane@wu.ac.at (S. Kirrane)

distinguish between misinformation and credible information. According to Pérez-Rosas et al. [5] some lexical markers, such as swear words or sexual words, appear more frequently in fake news than in trusted news. While, Connell [6] argue that when authors use a lot of quotes, the credibility shifts to the person quoted, which can be an indicator of something being false.

From a practical perspective, Potthast et al. [2] use a variety of style features, readability scores, and information about the similarity between two texts as input for their machine learning classifier. However, with their style based approach (F1 score of 46%) they can not outperform current baselines by far. In turn, Rashkin et al. [1] use Linguistic Inquiriy Word Count (LIWC) features and enhance them with Term Frequency-Inverse Document Frequency (TF-IDF) vectors and different machine learning classification models. Interestingly, only a few of the machine learning models (Naive Bayes and Maximum Entropy) tested by the authors could be improved with additional knowledge about the text's features. Without adding linguistic knowledge, Long Short Term Memory (LSTM) performed as well as or better.

One of the biggest limitations of existing approaches is the fact that they only use some of the metrics applied to a single dataset. Although there are a variety of linguistic cues and combinations that could potentially be used to identify misinformation, it is currently not clear which cues work best in which context. Thus, in this paper, we perform a broad analysis of linguistic cues by examining a combination of metrics over different datasets.

3. Proposed Architecture

In this section, we outline our three-step workflow for detecting misinformation using linguistic cues (depicted in Figure 1).

Dataset Selection. For the task of identifying misinformation using linguistic cues it is necessary to use data that includes entire article bodies and annotations that indicate if articles are fake or not. Unfortunately many datasets are unsuitable as they only include the headline and a link to the original article, or to the article's fact check. Thus, we choose three datasets that meet our head, body, and annotation criteria: MisInfoText Snopes [7], MisInfoText Buzzfeed [7], and FakeNewsAMT & Celebrity [5]. Both of the MisInfoText datasets are included in the same repository, however their data comes from different sources and covers different topics. We only use the Celebrity part from the FakeNewsAMT & Celebrity dataset as it includes fake and real articles from authentic sources.

Metric Selection and Implementation. We combine seven linguistic metrics gleaned from the literature and apply them to the three datasets. Additionally, we use different Python libraries, such as NLTK (average word and sentence length), Fuzzywuzzy (repetition), Empath (swear and sexual words), ScispaCy (technical words), and Regular Expressions (quotes) in order to compute the various metrics appear. Metric thresholds were used to ensure there is at least a certain amount of occurrences of a linguistic cue in the text (see Figure 1). We first set the threshold through experimentation and spot checking, and then tune it based on prediction quality.

Analysis. In order to learn more about the various metrics, we graph the results and conduct a comparative analysis. Subsequently, we perform manual inspections and investigate in the data to see why certain metrics may appear in fake or real news comparing it to the findings

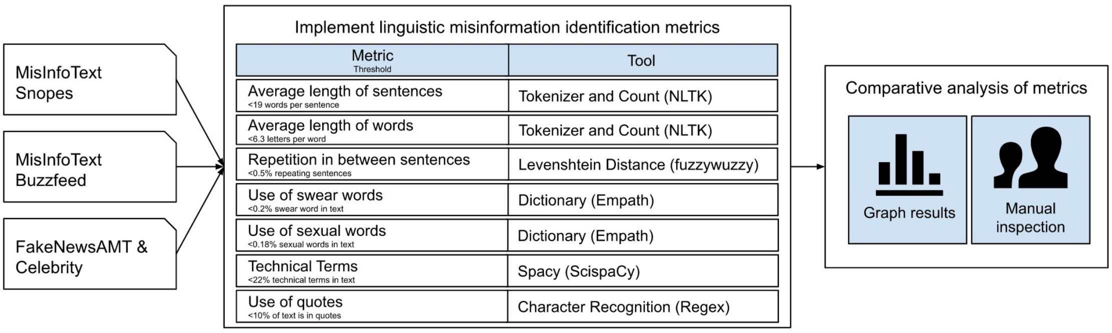

Figure 1: Workflow Diagram to Identify Misinformation with Linguistic Cues on Three Datasets

from the literature.

4. Experiment Results and Discussion

In this section, we describe the datasets used, the results of our experiment, and conduct a critical analysis on our results in the discussion.

4.1. Misinformation Datasets

For this study, we use datasets that include the entire text of an article because we want to learn more about writing styles and linguistic features.

MisInfoText Snopes Dataset [7]. This dataset consists of Snopes articles where the authors followed the links to the original articles and used Snope's veracity labels. They sampled 312 articles and manually assessed them due to quality issues with some of the articles. They employ a variety of Snopes' veracity labels, including true (65), mostly true (71), a mixture of true and false (72), and mostly false (53) and false (51). We focus on articles with the labels true and false without random sampling for our experiments because they are well balanced, yielding a total of 116 articles.

MisInfoText Buzzfeed Dataset [7]. The authors of this dataset used a list of links provided by Buzzfeed that included Facebook posts related to the 2016 US election. They gathered the titles, article bodies, authors, and dates and used Buzzfeed's four veracity labels. There are 1380 news articles in the dataset (1090 mostly true, 170 mixture of true and false, 64 mostly false, and 56 no factual content). Because this dataset is unbalanced, we conduct our experiment with a random sample of 64 mostly true articles and all 64 mostly false articles, which results in a total of 128 articles.

FakeNewsAMT & Celebrity [5]. This dataset is divided into two parts. The first is FakeNewsAMT, in which the authors used crowdsourcing to create fake versions of real articles they discovered. Second is Celebrity, where the authors gathered articles about celebrities, including both true and false stories. We only use the Celebrity part of this dataset because we want to investigate linguistic features that occur in news, and the fake and real articles are from authentic sources. They used GossipCop to double-check the information in articles and cross-referenced it with information from other sources. They used the labels fake and legit as

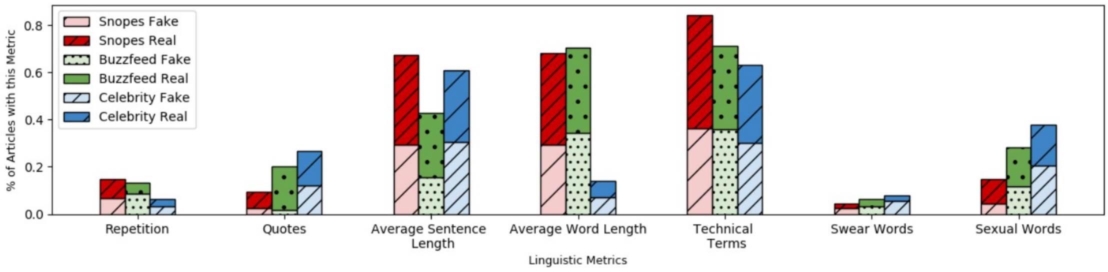

Figure 2: Comparison of Linguistic Cues used for Misinformation Detection between Three Datasets

they searched for articles in pairs (one true fitting to the false one and vice versa). They provide 500 articles in total, 250 of which are legit and 250 of which are fake.

4.2. The Experimental Results

We plot the percentage of articles that exceeded the metrics' threshold in Figure 2. On the x-Axis we display the overall results for each of the three datasets per metric.

Repetition. The repetition of information in an article may be a sign of a fake article [4]. We can confirm this finding for the MisInfoText Buzzfeed Dataset, where fake articles use more repetition than true articles. For the MisInfoText Snopes Dataset, as well as the Celebrity dataset, on the other hand, we see little difference between false and true articles.

Quotes. Using quotes, according to Connell [6], indicates that authors shift their credibility to the quoted person, which could be a sign of fake articles because authors are not holding their credibility. However, the MisInfoText Snopes and Buzzfeed datasets show that the majority of quotes are used in true articles, which contradicts the literature. There is little difference between true and false articles in the Celebrity dataset.

Sentence length. Longer sentences can be an indicator of high-credible writing, thus being indicative of true articles [3]. We can confirm this finding for the Celebrity dataset where more true articles have longer sentences, however in the MisInfoText Buzzfeed and Snopes datasets there is only a slight difference between true and false articles.

Word length. Using longer words on average can be an indicator of complicated words and high credibility writing, and thus true articles [3]. In our experiments, we do not see a significant difference in word length between true and false articles across all datasets.

Technical terms. According to Pérez-Rosas et al. [5], a high use of technical terms and complex words can indicate well-written articles for real information. In our experiment, we found no tendency for the use of technical terms that only refer to true or false articles.

Swear words. Swear words appear more frequently in false articles [5]. We can confirm this finding for the MisInfoText Snopes dataset and the Celebrity dataset. There was no clear indication of which type of article had more swear words in the MisInfoText Buzzfeed dataset.

Sexual words. Sexual words, like swear words, appear more frequently in false news [5]. In our experiments, however, we find more sexual words are used in real articles for the MisInfoText Snopes dataset. This may be related to the high use of quotes for real articles in this dataset, because swear words may be used in a large number of quotes that the author publishes without making any changes to underline statements for their true articles.

4.3. Discussion

Based on our experiments we see that some linguistic cues found in the literature have a tenuous relationship to either true or false articles. We also see that the datasets have a significant impact on the results. We can see clear relationships to either true or false articles if the dataset is not balanced, which is not represented after balancing the data. Some measurements are greatly influenced by the quality of the data. There were articles with metric repetition, for example, but upon closer inspection, the same text was duplicated in one article. This can occur during the automatic data curating process, as websites may include the article twice, for example, in a preview and on the page. We also discovered articles that were file dumps and thus unreadable, but still qualified for not fitting metrics. Working with measurements such as sentence length or word length reveals that some articles are empty or only have one sentence, which means that not only the average sentence length but also the repetition in between sentences cannot be performed correctly. Based on our analysis, there is a clear need for benchmarks to evaluate the quality of the datasets and to ensure that experiments do not suffer unduly from poor data quality.

5. Conclusion

In this paper, we examined the relationship between various linguistic cues that can be used to identify false and true articles. We discovered that some cues provide valuable insights into whether an article is true or false, but only for a few metrics and only for certain datasets. For future work, we plan to investigate other approaches that can identify misinformation across a variety of topics or domains and with a higher certainty.

References

[1] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, Y. Choi, Truth of varying shades: Analyzing language in fake news and political fact-checking, in: Proceedings of the 2017 conference on empirical methods in natural language processing, 2017.

[2] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, B. Stein, A stylometric inquiry into hyperpartisan and fake news, CoRR (2017).

[3] Q. Su, M. Wan, X. Liu, C.-R. Huang, Motivations, methods and metrics of misinformation detection: an nlp perspective, Natural Language Processing Research (2020).

[4] B. D. Horne, S. Adali, This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news, in: Eleventh international AAAI conference on web and social media, 2017.

[5] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, R. Mihalcea, Automatic detection of fake news, in: Proceedings of the 27th International Conference on Computational Linguistics, 2018.

[6] I. Connell, Mistaken identities: Tabloid and broadsheet news discourse, Javnost-the public (1998).

[7] F. T. Asr, M. Taboada, Misinfotext. a collection of news articles, with false and true labels. dataset., 2019. URL: https://github.com/sfu-discourse-lab/Misinformation_detection.