Lies and Language: Understanding how context can improve the automated early detection of fake news.

Noah Rose Ledesma, Final Draft, 1602 Words

Abstract

Due to the explosive growth and influential power of fake news, the demand for methods that accurately detect fake news, and intervene before it propagates, has dramatically increased. Human fact-checking cannot keep up with the quantity of disinformation entering the web. Hence automated systems have been developed. Preliminary evidence suggests that, for such systems, analyzing the text of an article alone may not be sufficient for distinguishing real from fake news. Other research has established a set of style features that are effective indicators of fake news on sample datasets. In this study, I bring these two findings together to investigate the relative effectiveness of different style features at distinguishing real from fake news across multiple areas of news. My research proposes questions such as "Can the number of words in a sentence be used to distinguish real from fake news?" and "Is the number of words in a sentence a better indicator of real vs. fake news in politics or sports articles?" Experiments conducted using language processing software on a real-world data set of N=400 articles indicate that the effectiveness of these language indicators does vary by news type. These variances could prove particularly useful for the development of novel early detection systems focused on textual content.

1.1. INTRODUCTION

The world is facing an age of disinformation. With Russian troll armies, vast networks of social media bots, and fake news stories running rampant on the web, new methods are being used to undermine democracies and disrupt political systems (Pomerantsev, 2019). In a survey conducted by Mitchell et al. (2019) fake news and disinformation were found to have had a great effect on American's confidence in each other and in government. To combat the presence of disinformation online, fact-checking websites such as PolitiFact, Snopes, Fiskkit, and TextThresher have emerged. While the accuracy and utility of these services are profound in establishing ground truth, manual fact-checking cannot scale with the volume of disinformation entering social media. Additionally, when overloaded with deceptive information, a human's ability to detect deception is only slightly better than chance (Rubin, 2010). Hence, there has been a great deal of research in developing automated systems for detecting disinformation. Such systems can be classified as using *propagation-based* and *content-based* methods.

Propagation-based methods for detecting disinformation online leverage how fake news propagates farther, faster, and deeper on social networks than real news (Vosoughi et al., 2018). While proposed *propagation-based* models exhibit good performance on real-world articles, they perform poorly on articles that have yet to become widespread (Jin et al., 2014). For the purposes of detecting fake news, and intervening before it goes viral, one must turn to *content-based* methods.

Content-based methods for detecting fake news analyze the text of an article under the assumption that fake statements differ from real ones in writing style. This assumption is known as the Undeutsch hypothesis and has been the subject of much research in forensic psychology. Analysis of textual content can take place on many levels such as lexicon, syntax, semantic, and discourse. A survey of detection methods conducted by Zhou & Zafarani (2018) provides a detailed description of these levels.

However, content alone may not be sufficient for distinguishing fake from real news. Rather, *content-based* methods may be improved by factoring in the context of an article (Pierri, 2018). In this study, I analyze how the effectiveness of language features used by *content-based* methods vary across different areas of news. If the effectiveness of these language features differs among news types, I propose that *content-based* detection methods should include the area of news as a feature within a machine learning framework. Such methods would be considering both content and context.

2. METHODS

As fake news potentially differs from real news in writing style, A dataset compiled by Pérez-Rosas et al. (2018), containing 400 real-world news articles labeled as *Real* or *Fake*, have been analyzed for use of language features. In addition to being labeled as *Real* or *Fake*, each of these articles has been labeled as belonging to one of the following areas: business, education, entertainment, politics, or sports. As suggested by Zhou et al., I investigate the difference in writing style between real and fake news by counting the frequencies of language features relative to the total number of words in an article. Features used in this study are defined by and measured using *Linguistic Inquiry and Word Count 2015* (Pennebaker et al., 2015). A detailed description of the language features used in this study is provided in Pennebaker et al. (2015). Using the Complementary Cumulative Distribution Function (CCDF) of attribute distributions, I will identify common patterns of language features across different areas of news. Then, I will establish statistical significance to the differences between the presence of these language features using a two-sample Kolmogorov-Smirnov goodness-of-fit hypothesis test.

2.2. TERMINOLOGY

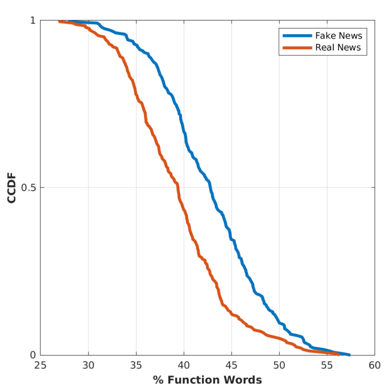
Before delving into my experimental design, some terminology should be clarified. By “language feature” I mean any output variable produced by LIWC2015. This includes *Summary Language Variables*, *Linguistic Dimensions*, *Psychological Processes*, and *Grammar* as they are defined by Pennebaker et al. (2015). When I refer to language features that are representative of writing style, I am referring to all of the aforementioned categories excluding *Grammar*.

3. RESULTS

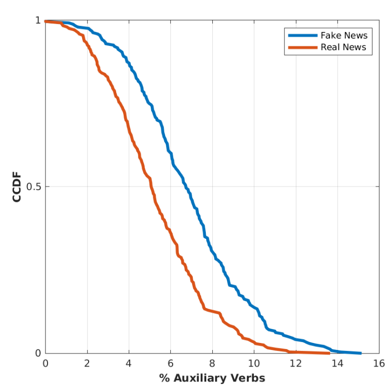
The results of my experiment are summarized in *Appendix A*. The effectiveness of each lanugage feature in different areas of news was determined by the p-value of the aforementioned *Kolmogorov-Smirnov* test. A p-value of 0.05 or less was considered strongly significant and a p-value of 0.1 or less was considered significant. My findings indicate that i) some language features may be effective in distinguishing real from fake news and ii) the effectiveness of these language features varies by news area. In addition to language features that relate to writing style, I also found that the frequency of punctuation may be an effective indicator.

4. DISCUSSION

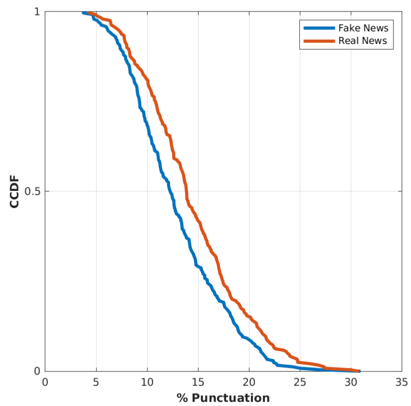
Our findings indicate that while the *polarity* (positive vs. negative) of an indicator tends to remain constant across multiple areas of news, it's *strength* varies. To illustrate, while the percentage of regular verbs in an article seems to be a strong indicator for news in general, it is a better indicator in sports and politics than it is in business news. Additionally, the frequencies of regular verbs in real education and entertainment news were not found to be significantly different than their fake counterparts. Noteably, I found that i) of the language features which were strong indicators fake news tends to have a higher frequency of these features than real news and ii) real news tends to have a higher frequency of punctuation and numbers.



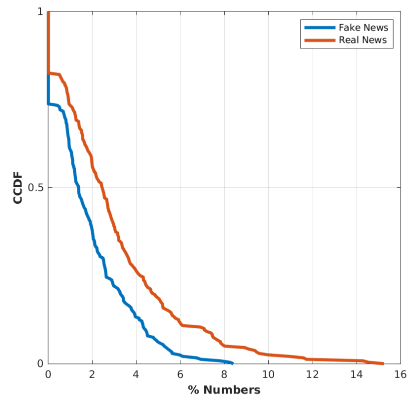
CCDF of the frequency of function words across all news types



CCDF of the frequency of auxillary verbs across all news types

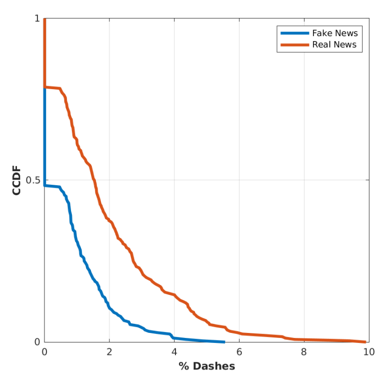


CCDF of the frequency of punctionation across all news types



CCDF of the frequency of numbers across all news types

Two such language features which were identified as being strong indicators that an article is fake were the frequency of function words and the frequency of auxillary verbs. The figures above plot the probability that the frequency of a feature was at least a certain percentage. For example, for the frequency of function word plots, the probablity that a real news article was composed of at least 40% function words was less than 50%. However, the probability that a fake news article had at least 40% function words was greater than 50%.



CCDF of the frequency of dashes across all news types

The most effective indicator of fake news across all areas of news was not a language feature relating to writing style. Rather, it was the frequency of dashes. The frequencies of other types of punctuation were also found to be effective indicators as well. The findings that I have noted here are consistent with that of Zhou & Jain et al. (2019) Matsumoto & Hwang (2015) and several other studies in a survey by Zhou & Zafarani (2018). This suggests that content-based detection systems may be improved by considering the quantity of punctuation in addition to traditional content features.

Additionally, it may be that the use of punctuation, regardless of quantity, is an effective indicator of fake news.

While this study only considers five different areas of news, my findings suggest that style assessed at the lexical-level varies among types of news. As such, existing methods for the content-based detection of fake news may be improved by including the area of news as a feature in machine learning models. It is important to note that this study only examines style at the lexicon-level. There are several more layers on which style may be analyzed, including but not limited to: syntax-level, semantic-level, and discourse-level as suggested by Zhou & Jain et al. (2019). Based on my findings at the lexicon-level, I hypothesize that style at higher levels varies across different areas of news.

5. CONCLUSION

In this study, I analyze how the presence of certain language features varies between real and fake news across multiple areas of news. Experimental results on a real-world dataset suggest that while language features may be positive or negative indicators of fake news, the effectiveness of these indicators does vary based on the area of news. Additionally, fake news tends to have a higher presence of language features that relate to writing style than real news. I acknowledge that effective use of this my findings remains an open issue and that I have not provided evidence that a detection method that considers this information will outperform existing methods. Further research is needed to explore the role that context plays in the content-based detection of fake news.

Attribute	All News	Business	Education	Entertainment	Politics	Sports
Categorical-Dynamic Index	--	--			--	--
# Words Per Sentence	--				--	--
# Words Longer Than 6 Letters	--		-		--	
% Function Words	++	++	++		++	++
% Pronouns	++				++	++
% Personal Pronouns	++	++			+	+
% First Person Plural			++			
% Third Person Plural	++	++				
% Impersonal Pronouns	+				++	+
% Auxiliary Verbs	++	++	++		++	++
% Common Adverbs					++	+
% Negations	++			++		++
% Regular Verbs	++	+			++	++
% Numbers	--		--	--	--	--
% Social	++	+				++
% Cognitive Processes	++					++
% Discrepancies	++					++
% Certainty	++		+			
% Hearing	--	--				
% Reward Focus	++	++				
% Present Focus	++	++			+	+
% Future Focus	+		+			
% Work			-		--	-
% Punctuation	--		--			
% Commas	++	++	--	++		++
% Colons	--					
% Dashes	--	--	--	--	--	--
% Apostrophes	--	--	--			--
% Parentheses (pairs)	--	--				

Appendix A

+: The feature is a positive indicator of fake news

++: The feature is a strong positive indicator of fake news

-: The feature is a negative indicator of fake news

--: The feature is a strong negative indicator of fake news

Empty cells indicate that the feature was not an effective indicator

References

Jin, Z., Cao, J., Jiang, Y.-G., & Zhang, Y. (2014). News Credibility Evaluation on Microblog with a Hierarchical Propagation Model. ICDM ’14: Proceedings of the 2014 IEEE International Conference on Data Mining. https://doi.org/10.1109/icdm.2014.91

Matsumoto, D., & Hwang, H. C. (2015). Differences in word usage by truth tellers and liars in written statements and an investigative interview after a mock crime. Journal of Investigative Psychology and Offender Profiling, 12(2), 199-216.

Mitchell, A., Gottfried, J., Fedeli, S., Stocking, G., & Walker, M. (2019). Many Americans say made-up news is a critical problem that needs to be fixed. Pew Research Center. June, 5, 2019.

Pennebaker, J.W., Boyd, R.L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. Austin, TX: University of Texas at Austin

Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. (2018). Automatic Detection of Fake News. International Conference on Computational Linguistics (COLING).

Pierri, F., & Ceri, S. (2019). False news on social media: A data-driven survey. ACM SIGMOD Record, 48(2), 18-27.

Pomerantsev, P. (2019, July 27). The disinformation age: a revolution in propaganda. *The Guardian*. https://www.theguardian.com/books/2019/jul/27/the-disinformation-age-a-revolution-in-propaganda

Rubin, V. L. (2010). On deception and deception detection: Content analysis of computer-mediated stated beliefs. Proceedings of the American Society for Information Science and Technology, 47(1), 1–10. https://doi.org/10.1002/meet.14504701124

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Zhou, X., Jain, A., Phoha, V. V., & Zafarani, R. (2019). Fake news early detection: A theory-driven model. arXiv preprint arXiv:1904.11679.

Zhou, X., & Zafarani, R. (2018). Fake news: A survey of research, detection methods, and opportunities. arXiv preprint arXiv:1812.00315.