Fake News?

Fact-Checking Individual Statements With Web Scraping and Sentiment Analysis

Sofia Escoto

July 25, 2022

1 Introduction

The spread of misinformation is a large societal problem, exacerbated by the widespread use of social media. Today, anyone can log into their computer and say something that the entire world can see. While the malicious spread of misinformation is often seen in regards to politics, it's not exclusive to it, and though anyone with a computer can also access a search engine and tell if a statement is false based on the results, it's rare that a person will take the time out of their day to do so. Because of this, many people spread false, and potentially harmful, statements online due to ignorance.

With a machine learning algorithm that takes a statement (ex. a blue whale can weigh up to 200 tons) and classifies the statement as either true or false, a person doesn't have to take any time to Google it on their own. With this model used in conjunction with or integrated into websites like Twitter, Facebook, and Reddit, misinformation could be easily recognized as false and therefore would be unlikely to "go viral" and spread to more users.

Though machine learning has been used before for fact-checking and determining the credibility of a claim, most of its application has been in analyzing either full-length articles using context clues and sources or on subjective statements that were trained to use 'common sense'. Obtaining truth values of objective statements has not been previously attempted in the public sphere, (i.e. outside of scientific claims [1]) or without significant amounts of pre-validated data available (i.e. as a part of a broader algorithm with a question-and-answer input-output model that relies on fact databases [2]).

With this proposed model, the truth values of objective statements will be available with a single click to ordinary Internet users, who may have otherwise just taken whatever they read online at face value.

2 Literature Review

Most often when a person wants to know whether something online is true or not, they're concerned with political articles. Using machine learning to analyze fake news (which is defined as any article that is "intentionally and verifiably false") has been studied widely in recent years, using context, word usage, and sources determine credibility [3]. This is very useful when dealing with long-form journalism, but isn't applicable for finding truth values of individual statements. Because a large part of this method involves analyzing the source the article came from, any lone statement from social media sites like Twitter or Facebook would be at a disadvantage. Alternatively, if attempting to fact-check a single sentence in an article, the truth value of this statement may not match up with that of the article as a whole, and could lead to inaccurate results.

Other research being done in this sphere, especially considering individual sentences, are essentially trying to get machines to use 'common sense.' Some of these models aim for the computer to 'understand' a given sentence, taking meaning from it and allowing a user to get the gist of a claim without needing to know exactly what was said. This can include determining the emotional connotation of a sentence, classifying opinions and reviews as negative or positive, or categorizing sentences based on broad subject matter. These models can start without prior knowledge of words and their meanings, only given an alphabet and large training set [4], or build off of individual words and their definitions in addition to grammatical rules [5], because they rely on previous experience. This means that simple, common-sense statements can be analyzed with relative ease, but statements that require more than just reasoning can not.

The most relevant research to the proposed model is about fact-checking and trying to determine the truth value of individual claims as described in *Using NLP for fact checking: a survey*. In this procedure, though, the algorithm must first determine whether the statement given is a claim at all. Because this seemed outside the scope of 5-week course and, "though research has gone into claim detection, there is no formal definition of what a claim is yet [2]," this step will not be included in my proposed model. Once a statement is confirmed to be a claim and is therefore able to be fact-checked, the algorithm searches through a fact database that stores prior claims that have already been manually fact-checked by experts. If the database does not have the information needed, the Internet can be utilized, in some situations just returning a snippet of the first search result to the user, and in other cases using sentiment analysis on the results to determine the credibility of the claim. Search results that include unassertive words can suggest a false claim, as well as

the use of first and second person pronouns as opposed to third person. When trying to fact-check a claim that is not included in a given fact database, "methods that utilize powerful search engines on the Internet perform best [2]."

3 Data

Since there isn't an open-source dataset of statements with truth values, I'm going to collect this data by scraping trivia websites or articles that tell the user whether the sentence is true or not. When I have enough statements, ideally at least 500 instances, I'll use an API to enter the statement in a Google search and scrape the results for 'snippets' to input into the dataset as well [7]. Once added to the dataset, cleaning will include removing markdown commands.

References

[1] David Wadden, Shanchuan Lin, Kyle Lo, Lucy Lu Wang, Madeleine van Zuylen, Arman Cohan, and Hannaneh Hajishirzi. Fact or fiction: Verifying scientific claims, 2020. URL `https://arxiv.org/abs/2004.14974`.

[2] Eric Lazarski, Mahmood Al-Khassaweneh, and Cynthia Howard. Using NLP for fact checking: A survey. *Designs*, 5 (3), 2021. ISSN 2411-9660. doi: 10.3390/designs5030042. URL `https://www.mdpi.com/2411-9660/5/3/42`.

[3] Kai Shu and Huan Liu. *Detecting Fake News on Social Media*. Synthesis Lectures on Data Mining and Knowledge Discovery. Morgan & Claypool Publishers, 2019. doi: 10.2200/S00926ED1V01Y201906DMK018. URL `https://doi.org/10.2200/S00926ED1V01Y201906DMK018`.

[4] Xiang Zhang and Yann LeCun. Text understanding from scratch. *CoRR*, abs/1502.01710, 2015. URL `http://arxiv.org/abs/1502.01710`.

[5] Pei Wang. Experience-grounded semantics: a theory for intelligent systems. *Cognitive Systems Research*, 6(4):282–302, 2005. ISSN 1389-0417. doi: https://doi.org/10.1016/j.cogsys.2004.08.003. URL `https://www.sciencedirect.com/science/article/pii/S1389041704000373`.

[6] Yoon Kim. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar, oct 2014. Association for Computational Linguistics. doi: 10.3115/v1/D14-1181. URL `https://aclanthology.org/D14-1181`.

[7] Roi Krakovski. Web search API. `https://rapidapi.com/user/contextualwebsearch`. Accessed: 2022-07-24.