Research shows how statistics can aid in the fight against misinformation

2 December 2021, by Rebecca Basu

Credit: CC0 Public Domain

An American University math professor and his team have created a statistical model that can be used to detect misinformation in social posts. The model also avoids the problem of black boxes that occur in machine learning.

With the use of algorithms and computer models, machine learning is increasingly playing a role in helping to stop the spread of misinformation, but a main challenge for scientists is the black box of unknowability, where researchers don't understand how the machine arrives at the same decision as human trainers.

Using a Twitter dataset with misinformation tweets about COVID-19, Zois Boukouvalas, assistant professor in AU's Department of Mathematics and Statistics, College of Arts and Sciences, shows how statistical models can detect misinformation in social media during events like a pandemic or a natural disaster. In newly published research, Boukouvalas and his colleagues, including AU student Caitlin Moroney and Computer Science Prof. Nathalie Japkowicz, also show how the model's decisions align with those made by humans.

"We would like to know what a machine is thinking when it makes decisions, and how and why it agrees with the humans that trained it," Boukouvalas said. "We don't want to block someone's social media account because the model makes a biased decision."

Boukouvalas's method is a type of machine learning using statistics. It's not as popular a field of study as deep learning, the complex, multi-layered type of machine learning and artificial intelligence. Statistical models are effective and provide another, somewhat untapped, way to fight misinformation, Boukouvalas said.

For a testing set of 112 real and misinformation tweets, the model achieved a high prediction performance and classified them correctly, with an accuracy of nearly 90 percent. (Using such a compact dataset was an efficient way for verifying how the method detected the misinformation tweets.)

"What's significant about this finding is that our model achieved accuracy while offering transparency about how it detected the tweets that were misinformation," Boukouvalas added. "Deep learning methods cannot achieve this kind of accuracy with transparency."

Before testing the model on the dataset, researchers first prepared to train the model. Models are only as good as the information humans provide. Human biases get introduced (one of the reasons behind bias in facial recognition technology) and black boxes get created.

Researchers carefully labeled the tweets as either misinformation or real, and they used a set of pre-defined rules about language used in misinformation to guide their choices. They also considered the nuances in human language and

linguistic features linked to misinformation, such as a post that has a greater use of proper nouns, punctuation and special characters. A socio-linguist, Prof. Christine Mallinson of the University of Maryland Baltimore County, identified the tweets for writing styles associated with misinformation, bias, and less reliable sources in news media. Then it was time to train the model.

"Once we add those inputs into the model, it is trying to understand the underlying factors that leads to the separation of good and bad information," Japkowicz said. "It's learning the context and how words interact."

For example, two of the tweets in the dataset contain "bat soup" and "COVID" together. The tweets were labeled misinformation by the researchers, and the model identified them as such. The model identified the tweets as having hate speech, hyperbolic language, and strongly emotional language, all of which are associated with misinformation. This suggests that the model distinguished in each of these tweets the human decision behind the labeling, and that it abided by the researchers' rules.

The next steps are to improve the user interface for the model, along with improving the model so that it can detect misinformation social posts that include images or other multimedia. The statistical model will have to learn how a variety of elements in social posts interact to create misinformation. In its current form, the model could best be used by social scientists or others who are researching ways to detect misinformation.

In spite of the advances in machine learning to help fight misinformation, Boukouvalas and Japkowicz agreed that human intelligence and news literacy remain the first line of defense in stopping the spread of misinformation.

"Through our work, we design tools based on machine learning to alert and educate the public in order to eliminate misinformation, but we strongly believe that humans need to play an active role in not spreading misinformation in the first place," Boukouvalas said.

More information: Caitlin Moroney et al, The Case for Latent Variable Vs Deep Learning Methods in Misinformation Detection: An Application to COVID-19, *Discovery Science* (2021). DOI: 10.1007/978-3-030-88942-5_33

Provided by American University

APA citation: Research shows how statistics can aid in the fight against misinformation (2021, December 2) retrieved 26 May 2022 from https://techxplore.com/news/2021-12-statistics-aid-misinformation.html