AI can predict Twitter users likely to spread disinformation before they do it

December 14 2020

Credit: CC0 Public Domain

A new artificial intelligence-based algorithm that can accurately predict which Twitter users will spread disinformation before they actually do it has been developed by researchers from the University of Sheffield.

A team of researchers, led by Yida Mu and Dr. Nikos Aletras from the University's Department of Computer Science, has developed a method

for predicting whether a social media user is likely to share content from unreliable news sources. Their findings have been published in the journal *PeerJ*.

The researchers analyzed over 1 million tweets from approximately 6,200 Twitter users by developing new natural language processing methods—ways to help computers process and understand huge amounts of language data. The tweets they studied were all tweets that were publicly available for anyone to see on the social media platform.

Twitter users were grouped into two categories as part of the study—those who have shared unreliable news sources and those who only share stories from reliable news sources. The data was used to train a machine-learning algorithm that can accurately predict (79.7 percent) whether a user will repost content from unreliable sources sometime in the future.

Results from the study found that the Twitter users who shared stories from unreliable sources are more likely to [tweet](#) about either politics or religion and use impolite language. They often posted tweets with words such as 'liberal," 'government," 'media," and their tweets often related to politics in the Middle East and Islam, with their tweets often mentioning "Islam' or "Israel."

In contrast, the study found that Twitter users who shared stories from reliable news sources often tweeted about their personal life, such as their emotions and interactions with friends. This group of users often posted tweets with words such as
"mood." "wanna," "gonna," "I'll," "excited," and "birthday."

Findings from the study could help social media companies such as Twitter and Facebook develop ways to tackle the spread of disinformation online. They could also help [social scientists](#) and

psychologists improve their understanding of such user behavior on a large scale.

Dr. Nikos Aletras, Lecturer in Natural Language Processing at the University of Sheffield, said: "Social media has become one of the most popular ways that people access the news, with millions of users turning to platforms such as Twitter and Facebook every day to find out about key events that are happening both at home and around the world. However, social media has become the primary platform for spreading disinformation, which is having a huge impact on society and can influence people's judgment of what is happening in the world around them.

"As part of our study, we identified certain trends in user behavior that could help with those efforts—for example, we found that users who are most likely to share news stories from unreliable sources often tweet about politics or religion, whereas those who share stories from reliable news sources often tweeted about their personal lives.

"We also found that the correlation between the use of impolite language and the spread of unreliable content can be attributed to high online political hostility."

Yida Mu, a Ph.D. student at the University of Sheffield, said: "Studying and analyzing the behavior of users sharing content from unreliable news sources can help social media platforms to prevent the spread of fake news at the user level, complementing existing fact-checking methods that work on the post or the news source level."

  More information: Identifying Twitter users who repost unreliable news sources with linguistic information, Yida Mu, Nikolaos Aletras, *PeerJ*, doi.org/10.7717/peerj-cs.325