

Researchers collaborate on method to explain 'fake news' to users

July 18 2019, by Jessica Hallman



While numerous deep learning methods currently exist to detect fake news, they are unable to explain why it is recognized as such. New research from Penn State and Arizona State could help to explain why a piece of news is detected as being false. Credit: Adobe Stock/georgejmlittle

Social media can expose users to misinformation, including fake

news—news stories with intentionally false information. In fact, during the 2016 U.S. presidential election, fake news engaged more people than real news, according to a [BuzzFeed News analysis](#).

Numerous deep learning methods currently exist to detect fake [news](#), but these methods are unable to explain why it is recognized as such. Now, a team of researchers from Penn State and Arizona State is working to help explain why any piece of fake news is detected as being false.

The team's recent findings are to be presented at the Association for Computing Machinery (ACM)'s Knowledge Discovery and Data Mining (KDD), a flagship [data mining](#) conference, held Aug. 4-8 in Anchorage, Alaska.

"Detection is one thing, but how to present it to the user to explain why it's fake is more challenging," said Dongwon Lee, associate professor in the Penn State College of Information Sciences and Technology and researcher on the project. "If we don't provide a good explanation, it has a limited impact to curtail the distribution of misinformation because people won't accept it."

In their study, the researchers built an explainable fake-news detection framework, which they call DEFEND (Explainable FakeE News Detection). The framework consists of three components: (1) a news content encoder, to detect opinionated and sensational language styles commonly found in fake news; (2) a user comment encoder, to detect activities such as skeptical opinions and sensational reactions in comments on news stories; and (3) a sentence-comment, co-attention component, which detects sentences in news stories and user comments that can explain why a piece of news is fake.

The new detection algorithm designed and developed in this novel approach has outperformed seven other state-of-the-art methods in

detecting fake news, according to the researchers.

"Among the users' comments, we can pinpoint the most effective explanation as to why this [piece of news they are reading] is fake news," explained Lee. "Some users expressed discontent but others provide particular evidence, such as linking to a fact-checking website or to an authentic news article. These techniques can concurrently find such evidence and present it to the user as potential explanation."

He added, "The democracy [in the United States] as we know it is based on the premise of sharing one's ideas and opinions freely. If we cannot trust what has been said in the media, and start suspecting it may be false, it could be undermining an entire ecosystem of democracy. As such, this research makes an important and huge societal impact."

The researchers are working on a prototype of the system, which they hope to share in late 2019, so that others can use the tool to detect fake news and better understand it.

"Early fake-news detection is another important issue," said Suhang Wang, assistant professor in the College of IST and collaborator on the project. "When [fake] news comes out, within a few hours, we want to detect it. Once [fake news](#) spreads, the damage has already been done. It's important to detect and curtail it as soon as possible."

More information: dEFEND: Explainable Fake News Detection.
doi.org/10.1145/3292500.3330935

Provided by Pennsylvania State University

Citation: Researchers collaborate on method to explain 'fake news' to users (2019, July 18)

retrieved 26 September 2023 from

<https://techxplore.com/news/2019-07-collaborate-method-fake-news-users.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.