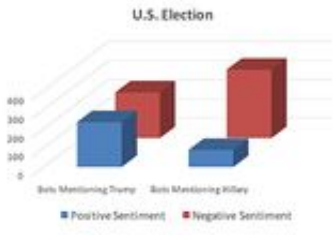
Researchers create technology to detect bad bots in social media

November 1 2016



Credit: University of New Mexico

When you check your Twitter feed, do you assume there is a real person behind each Tweet that is posted or shared? After all, there is a name and a photo, so it must be a real person behind the words, right?

Not so fast, say UNM researchers. Although Twitter has rules against "botting," impersonation and similar activity, the bots have multiplied so fast that the rule is nearly unenforceable, said Abdullah Mueen, assistant professor of computer science at The University of New Mexico.

"Twitter rules allow some bots for informational purposes but doesn't allow them to be created with the intention of swaying public behavior, but they are being created so quickly and are so hard to detect that they are going undetected," he said.

That's where the technology created in the Department of Computer

Science can help. The technology that Mueen and his group created called DeBot led to the formation of a startup company called BotAlert Inc. Mueen said the hope with that company is that it can assist the government and Twitter in detecting bad bots and preventing new ones from forming.

Since his research group (which includes students Nikan Chavoshi, Noor Abu-El-Rub, Amanda Minnich and Hossein Hamooni) has been tracking them in October 2015, their group has detected about 700,000 bots. Mueen said that about 1,500 bots a day are created, some of which are legal and some of which are not.

An example of a "good" use of a bot would be how feeds like CNN sports or CNN politics continuously update news around the clock. Bots in those cases are permitted because they are simply sharing news and not trying to impersonate other people with the intent to sway public opinion. Bot usage is efficient because it doesn't require someone to have to continuously and manually update feeds.

But bots are considered "bad" when they are created to impersonate someone under a fake identity and if those bots are created to influence public opinion. One example is in music contests, where the public is invited to vote on a winner. Mueen said in a recent iHeart Media popular choice contest, they observed "bot armies" engaged in massive campaigns in support of competing bands.

And the world of politics is using bots, especially in the presidential election. Mueen and his team have been tracking bots between Hillary Clinton and Donald Trump since the debates began. Through the use of IBM's Watson open-source analytics, Mueen's team has been able to monitor bots mentioning the two candidates and rate whether the content is "positive" or "negative." Their analysis has shown that there are twice as many bots mentioning Clinton in a negative light than there

are mentioning Trump in a negative light.

Mueen said it will take significant research to determine where these bots are originating from and what the possible effects may be on public opinion. Complicating matters is the fact that hybrid bots have been created to essentially "clean up" bad content to delete it after it's posted to erase the evidence.

To the user, it may be nearly impossible to detect a bad bot, but their technology has found a way to detect in real time what are likely fake bots through activity correlation. On the group's DeBot website, it shows examples of identical Tweets by two completely different users.

"To the user seeing one of these posts, this may not seem suspicious, but our technology has found many examples of highly-correlated Twitter accounts that are retweeting identical content within 10 seconds of each other," Mueen said. "Of course, two users could not be simultaneously posting the exact same content for hours, so this is naturally suspicious and identifies a bot."

UNM's DeBot technology works by listening to keywords, indexing and picking up on suspicious words, monitoring suspicious users, and clustering suspicious users to find highly-correlated user accounts. The technology is identifying bots at a higher rate than Twitter is suspending accounts, Mueen said.

While Mueen said the constantly-evolving technology and the increasing amount of information being disseminated is a challenge, he's hopeful this technology could have a positive impact.

"There are 313 million active Twitter users, and Twitter only provides us 1 percent of the data, but of that data, we've been able to detect thousands of bad bots," he said. "This is a huge computational task, but

our eventual goal is to be able to understand who is behind bots in order to stop their creation and spread."

More information: Identifying Correlated Bots in Twitter.
www.researchgate.net/publicati … ated_Bots_in_Twitter

DeBot: Twitter Bot Detection via Warped Correlation.
www.researchgate.net/publicati … a_Warped_Correlation

Provided by University of New Mexico