Bad news for fake news: New research helps combat social media misinformation

December 10 2020

Anshumali Shrivastava is an assistant professor of computer science at Rice University. (Photo by Jeff Fitlow/Rice University) Credit: Jeff Fitlow/Rice University

Rice University researchers have discovered a more efficient way for

social media companies to keep misinformation from spreading online using probabilistic filters trained with artificial intelligence.

The new approach to scanning social media is outlined in a study presented today at the online-only 2020 Conference on Neural Information Processing Systems (NeurIPS 2020) by Rice computer scientist Anshumali Shrivastava and statistics graduate student Zhenwei Dai. Their method applies machine learning in a smarter way to improve the performance of Bloom filters, a widely used technique devised a half-century ago.

Using test databases of fake news stories and computer viruses, Shrivastava and Dai showed their Adaptive Learned Bloom Filter (Ada-BF) required 50% less memory to achieve the same level of performance as learned Bloom filters.

To explain their filtering approach, Shrivastava and Dai cited some data from Twitter. The social media giant recently revealed that its users added about 500 million tweets a day, and tweets typically appeared online one second after a user hit send.

"Around the time of the election they were getting about 10,000 tweets a second, and with a one-second latency that's about six tweets per millisecond," Shrivastava said. "If you want to apply a filter that reads every tweet and flags the ones with information that's known to be fake, your flagging mechanism cannot be slower than six milliseconds or you will fall behind and never catch up."

If flagged tweets are sent for an additional, manual review, it's also vitally important to have a low false-positive rate. In other words, you need to minimize how many genuine tweets are flagged by mistake.

"If your false-positive rate is as low as 0.1%, even then you are

mistakenly flagging 10 tweets per second, or more than 800,000 per day, for manual review," he said. "This is precisely why most of the traditional AI-only approaches are prohibitive for controlling the misinformation."

Shrivastava said Twitter doesn't disclose its methods for filtering tweets, but they are believed to employ a Bloom filter, a low-memory technique invented in 1970 for checking to see if a specific data element, like a piece of computer code, is part of a known set of elements, like a database of known computer viruses. A Bloom filter is guaranteed to find all code that matches the database, but it records some false positives too.

"Let's say you've identified a piece of misinformation, and you want make sure it is not spread in tweets," Shrivastava said. "A Bloom filter allows to you check tweets very quickly, in a millionth of a second or less. If it says a tweet is clean, that it does not match anything in your database of misinformation, that's 100% guaranteed. So there is no chance of OK'ing a tweet with known misinformation. But the Bloom filter will flag harmless tweets a fraction of the time."

Within the past three years, researchers have offered various schemes for using machine learning to augment Bloom filters and improve their efficiency. Language recognition software can be trained to recognize and approve most tweets, reducing the volume that need to be processed with the Bloom filter. Use of machine learning classifiers can lower how much computational overhead is needed to filter data, allowing companies to process more information in less time with the same resources.

"When people use machine learning models today, they waste a lot of useful information that's coming from the machine learning model," Dai said.

The typical approach is to set a tolerance threshold and send everything that falls below that threshold to the Bloom filter. If the confidence threshold is 85%, that means information that the classifier deems safe with an 80% confidence level is receiving the same level of scrutiny as information it is only 10% sure about.

"Even though we cannot completely rely on the machine-learning classifier, it is still giving us valuable information that can reduce the amount of Bloom filter resources," Dai said. "What we've done is apply those resources probabilistically. We give more resources when the classifier is only 10% confident versus slightly less when it is 20% confident and so on. We take the whole spectrum of the classifier and resolve it with the whole spectrum of resources that can be allocated from the Bloom filter."

Shrivastava said Ada-BF's reduced need for memory translates directly to added capacity for real-time filtering systems.

"We need half of the space," he said. "So essentially, we can handle twice as much information with the same resource."

Provided by Rice University

Citation: Bad news for fake news: New research helps combat social media misinformation (2020, December 10) retrieved 27 April 2024 from https://techxplore.com/news/2020-12-bad-news-fake-combat-social.html