July 08, 2019 | By Ioana Patringenaru

Hate Spoilers? This AI Tool Spots Them for You

Neural networks-based SpoilerNet is designed to catch online spoilers before they catch you off guard

Did social media spoil the Avengers' Endgame movie for you? Or maybe one of the Game of Thrones books? A team of researchers from the University of California San Diego is working to make sure that does not happen again. They have developed an AI-based system that can flag spoilers in online reviews of books and TV shows.

"Spoilers are everywhere on the internet, and are very common on social media. As internet users, we understand the pain of spoilers, and how they can ruin one's experience," said Ndapa Nakashole, a professor of computer science at UC San Diego and one of the paper's senior authors.

Some websites allow people to manually flag their posts with tags that serve as 'spoiler ahead' warning signs. But this doesn't always happen. So researchers wanted to develop an artificial intelligence tool powered by neural networks to automatically detect spoilers. They named the tool SpoilerNet.

On a theoretical level, researchers want to better understand how people write spoilers and what kind of linguistic patterns and common knowledge mark a sentence as a spoiler.

Researchers will present their findings at the 2019 annual meeting of the Association for Computational Linguistics in Florence, Italy, July 28 to Aug. 2. The tool the researchers developed could be used to build a browser extension to shield people from spoilers.

To train and test SpoilerNet, the UC San Diego team went looking for large datasets of sentences containing spoilers. Spoiler alert! They found none. So they created their own by collecting more than 1.3 million book reviews annotated with spoiler tags by book reviewers. The tags encompass sentences that include spoilers and hide them behind a "view spoiler" link in the text. The reviews were collected from Goodreads, a social networking site that allows people to track what they read, and share thoughts and reviews with other readers.

"To our knowledge, this is the first dataset with spoiler annotations at this scale and at such a fine-grained granularity," said Mengting Wan, a Ph.D. student in computer science at UC San Diego and the paper's first author.

Researchers found that spoiler sentences tend to clump together in the latter part of reviews. But they also found that different users had different standards to tag spoilers, and neural networks needed to be carefully calibrated to take this into account.

In addition, the same word may have different semantic meanings in different contexts. For example, 'green' is just a color in one book review, but it can be the name of an important character and a signal for spoilers in another book. Identifying and understanding these differences is challenging, Wan said.

Researchers trained SpoilerNet on 80 percent of the reviews on Goodreads, running the text through several layers of neural networks. The system could detect spoilers with 89 to 92 percent accuracy.

They also ran SpoilerNet on a dataset of more than 16,000 single-sentence reviews of about 880 TV shows. The accuracy of the tool to detect spoilers was 74 to 80 percent.

Most of the errors came from the system getting distracted by words that are usually loaded and revelatory—for example murder or killed.

Looking forward, the Goodreads dataset can be used as a powerful tool to train algorithms to detect spoilers in different types of content—say, tweets containing spoilers.

MEDIA CONTACT

Ioana Patringenaru, 858-822-0899, ipatrin@ucsd.edu

UC San Diego's <u>Studio Ten 300</u> offers radio and television connections for media interviews with our faculty, which can be coordinated via <u>studio@ucsd.edu</u>. To connect with a UC San Diego faculty expert on relevant issues and trending news stories, visit <u>https://ucsdnews.ucsd.edu/media-resources/faculty-experts</u>.