

Computers learn to spot 'opinion spam' in online reviews

July 26 2011, By Bill Steele

(PhysOrg.com) -- If you read online reviews before purchasing a product or service, you may not always be reading the truth. Review sites are becoming targets for "opinion spam" -- phony positive reviews created by sellers to help sell their products, or negative reviews meant to downgrade competitors.

The bad news: Human beings are lousy at identifying deceptive [reviews](#). The good news: Cornell researchers are developing [computer software](#) that's pretty good at it. In a test on 800 reviews of Chicago hotels, a computer was able to pick out deceptive reviews with almost 90 percent accuracy. In the process, the researchers discovered an intriguing correspondence between the [linguistic structure](#) of deceptive reviews and fiction writing.

The work was reported at the 49th annual meeting of the Association for Computational Linguistics in Portland, Ore., June 24, by Claire Cardie, professor of [computer science](#); Jeff Hancock, associate professor of communication; and graduate students Myle Ott and Yejin Choi.

"This is the first look at this, and there's a lot more to be done, but I think there is a potential that [review sites] could apply it," Ott said.

The researchers created what they believe to be the first "gold standard" collection of opinion spam by asking a group of people to deliberately write false positive reviews of 20 Chicago hotels. These were compared with an equal number of carefully verified truthful reviews.

As a first step, the researchers submitted a set of reviews to three human judges -- volunteer Cornell undergraduates -- who scored no better than chance in identifying deception. The three did not even agree on which reviews they thought were deceptive, reinforcing the conclusion that they were doing no better than chance. Historically, Ott noted, humans suffer from a "truth [bias](#)," assuming that what they are reading is true until they find evidence to the contrary. When people are trained at detecting deception they may become overly skeptical and report deception too often, still scoring at chance levels.

The researchers then applied computer analysis based on subtle features of text. Truthful hotel reviews, for example, are more likely to use concrete words relating to the hotel, like "bathroom," "check-in" or "price." Deceivers write more about things that set the scene, like "vacation," "business trip" or "my husband." Truth-tellers and deceivers also differ in the use of keywords referring to human behavior and personal life, and sometimes in features like the amount of punctuation or frequency of "large words." In parallel with previous analysis of imaginative vs. informative writing, deceivers use more verbs and truth-tellers use more nouns.

Using these approaches, the researchers trained a computer on a subset of true and false reviews, then tested it against the rest of the database. The best results, they found, came from combining keyword analysis with the ways certain words are combined in pairs. Adding these two scores identified deceptive reviews with 89.8 percent accuracy.

Ott cautions that the work so far is only validated for hotel reviews, and for that matter, only reviews of hotels in Chicago. The next step, he said, is to see if the techniques can be extended to other categories, starting perhaps with restaurants and eventually moving to consumer products. He also wants to look at negative reviews.

This sort of software might be used by review sites as a "first-round filter," Ott suggested. If, say, one particular hotel gets a lot of reviews that score as deceptive, the site should investigate further.

"I think cutting down on deception would help everyone," he said. "Customers would not be fooled, and it would help [sellers] and review sites because people would trust their reviews."

Provided by Cornell University

Citation: Computers learn to spot 'opinion spam' in online reviews (2011, July 26) retrieved 26 April 2024 from <https://phys.org/news/2011-07-opinion-spam-online.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.