ESTIMATING THE PREVALENCE OF DECEPTION IN ONLINE REVIEW COMMUNITIES

Myle Ott,¹ Claire Cardie,¹,² Jeff Hancock²,³
Depts of Computer Science,¹ Information Science,² Communication³
Cornell University, Ithaca, New York
Consumers increasingly rate, review and research products and services online.
“Positive information I’ve read online has reinforced my decision to purchase a product or service recommended to me.”

Source: http://www.coneinc.com/negative-reviews-online-reverse-purchase-decisions
“Negative information I’ve read online has made me **change my mind** about purchasing a product or service recommended to me.”

Source: http://www.coneinc.com/negative-reviews-online-reverse-purchase-decisions
Perhaps unsurprisingly...
Is this an epidemic?

Fake reviews prompt Belkin apology

In a Race to Out-Rank, Historian Orlando Figes damages for fake reviews

Orlando Figes posted reviews on Amazon praising his own work and rubbishing that of his rivals

Amazon withdraws ebook explaining how to manipulate its sales rankings

Ebook claiming one can become a Kindle 'bestseller' simply by posting fake reviews temporarily removed from bookseller's listings

Tripadvisor bribes: Hotel owners offer free rooms in return for glowing reviews

For $2 a Star, an Online Reviews

Author Claims To Manipulate Amazon Rankings By Buying Own Book Every Day

Company Settles Case of Reviews It Faked
Works Just as expected, May 14, 2007

By Laurie B. Cook - See all my reviews

This review is from: Belkin F5U301 CableFree 4-Port USB 2.0 Hub with Dongle (Electronics)

Supplies good range and does provide true wireless USB. Software worked right out of the box. I have been recommending this nifty little device to all my friends. Very useful device.
How many fake reviews are out there?
How to tell if a review is fake
Which of these reviews is fake?

“I have stayed at many hotels traveling for both business and pleasure and I can honestly stay that The James is tops. The service at the hotel is first class. The rooms are modern and very comfortable. The location is perfect within walking distance to all of the great sights and restaurants. Highly recommend to both business travellers and couples.”

“My husband and I stayed at the James Chicago Hotel for our anniversary. This place is fantastic! We knew as soon as we arrived we made the right choice! The rooms are BEAUTIFUL and the staff very attentive and wonderful!! The area of the hotel is great, since I love to shop I couldn't ask for more!! We will definatly be back to Chicago and we will for sure be back to the James Chicago.”
Which of these reviews is fake?

Answer:

“My husband and I stayed at the James Chicago Hotel for our anniversary. This place is fantastic! We knew as soon as we arrived we made the right choice! The rooms are BEAUTIFUL and the staff very attentive and wonderful!! The area of the hotel is great, since I love to shop I couldn't ask for more!! We will definitly be back to Chicago and we will for sure be back to the James Chicago.”
Ott et al. (2011) Dataset

- Solicited 400 fake positive reviews of Chicago hotels
- Gathered 400 truthful positive reviews from TripAdvisor

Source: http://www.cs.cornell.edu/~myleott
Ott et al. (2011)

Identifying Deception

- 2 out of 3 undergraduates performed at-chance
- n-gram text categorization (SVM) is \( \approx \) 90% accurate

Source: http://www.cs.cornell.edu/~myleott
Truthful reviews

- Tempered opinions
- More spatial details
- More nouns and adjectives
- More numbers and punctuation

Fake reviews

- Exaggerated opinions
- Greater focus on aspects external to the hotel
- More pronouns, verbs and adverbs
- More filler (blah, like)

Source: http://www.cs.cornell.edu/~myleott
How many fake reviews are out there?
Expectations

• Given that users increasingly rely on online reviews (Cone, 2011), rates of deception **must be low**

• On the other hand, rates of deception **may vary** across review communities and user groups
Expectations

Less deception
- Verified (high cost) review communities
- Low traffic (low benefit) review communities

More deception
- Unverified (low cost) review communities
- High traffic (high benefit) review communities
Approach

• Assume given a deception classifier
• Apply the classifier to some reviews
• Estimate the classifier’s sensitivity and specificity, i.e., recall rates
• Estimate the rate of deception with a generative model
Generative Storyline

- Sample *(latent)* rate of deception
- Sample *(latent)* sensitivity
- Sample *(latent)* specificity
- For each review:
  - Sample *(latent)* ground-truth deception label
  - Sample *(observed)* classifier output
if \((x_i)\)
If \((x_i)\) then \(\eta^*\) and \(\gamma\) (latent) specificity (deceptive recall) and (latent) sensitivity (truthful recall).
\( \text{if} (x_i) \)
$\alpha$

$\pi^*$

$\beta$

$\eta^*$

$y_i$

$f(x_i)$

$\gamma$

$\theta^*$

$N_{\text{test}}$

(latent) rate of deception

(latent) ground-truth deception label

(latent) specificity (deceptive recall)

(latent) sensitivity (truthful recall)

(observable) classifier output

(latent) rate of deception

(latent) specificity (deceptive recall)

(latent) sensitivity (truthful recall)

(observable) classifier output
Approach

• Gibbs sampling

• Apply model to reviews from six hotel review communities:
  – **High cost**: Expedia, Hotels.com, Orbitz, Priceline
  – **Low cost**: TripAdvisor and Yelp
In Figure 3, blue (a–d) and red (e–f) graphs, as before, correspond to communities for which you are required to book a hotel room before posting a review, while red graphs (e–f) correspond to communities that allow any user to post reviews for any hotel. In agreement with Hypothesis 1, communities with a signal cost, e.g., TripAdvisor, seem to have a rate of deception of 6%, to 4%, suggesting both that an increased signal cost may indeed help to reduce the prevalence of deception over time on TripAdvisor after removing these reviews, with rates dropping from 12% to 6%, to 4%, and finally to 2%. In contrast, review communities that allow any user to post reviews for any hotel, i.e., communities that allow posting cost, e.g., Orbitz, see a clear reduction in the prevalence of deception over time on their rate of deceptive opinion spam. Interestingly, communities with a blend of posting costs, e.g., Orbitz, contain larger quantities and accelerated growth of deceptive ties, e.g., Orbitz. To test Hypothesis 2, it is necessary to either ties, e.g., Orbitz. Interestingly, communities with a blend of posting costs, e.g., TripAdvisor, seem to be able to grow at a faster rate than communities with a signal cost, e.g., TripAdvisor, which is perhaps overestimated by the classifier's estimated false positive rate, suggesting both that an increased signal cost may indeed help to reduce the prevalence of deception. Thus, it is possible that the classifier's estimated false positive rate of the classifier is perhaps overestimated by the Bayesian Prevalence Model, as the estimates of the prevalence of deception for the six review communities in particular their hyperparameters, the estimated false positive rate of the classifier is perhaps overestimated by the Bayesian Prevalence Model, are often negative. This occurs when the rate at which the classifier's estimated false positive rate, suggesting both that an increased signal cost may indeed help to reduce the prevalence of deception, we need to give the estimated prevalence of deception over time after removing reviews written by first- or second-time review writers, we are increasing the posting cost by, for example, hiding all reviews written by users who have not posted at least two reviews. Essentially, by increasing the posting cost, as we have defined it in Section 6, it will decrease the prevalence of deception, we need to give the estimated prevalence of deception over time after removing reviews written by first-time review writers, we are increasing the posting cost and, accordingly, the signal cost as well. Thus, it is possible that the classifier's estimated false positive rate of the classifier is perhaps overestimated by the Bayesian Prevalence Model, as the estimates of the prevalence of deception for the six review communities in particular their hyperparameters, the estimated false positive rate of the classifier is perhaps overestimated by the Bayesian Prevalence Model, are often negative. This occurs when the rate at which the classifier's estimated false positive rate, suggesting both that an increased signal cost may indeed help to reduce the prevalence of deception, we need to give the estimated prevalence of deception over time after removing reviews written by first- or second-time review writers, we are increasing the posting cost and, accordingly, the signal cost as well.

Estimates of the prevalence of deception for six review communities over time, given by the Bayesian Prevalence Model, appear in Figure 2. Blue graphs (a–d) correspond to communities for which you are required to book a hotel room before posting a review, while red graphs (e–f) correspond to communities that allow any user to post reviews for any hotel. In particular their hyperparameters, the estimated false positive rate of the classifier is perhaps overestimated by the Bayesian Prevalence Model, are often negative. This occurs when the rate at which the classifier's estimated false positive rate, suggesting both that an increased signal cost may indeed help to reduce the prevalence of deception, we need to give the estimated prevalence of deception over time.
The rate of deception varies according to the costs and benefits of posting fake reviews.
Increasing the cost should reduce deception.
Users with 
≥ 1 review

Users with 
≥ 2 reviews

Users with 
≥ 3 reviews
Conclusion

• Presented a framework for estimating the rate of deception in online review communities using a noisy classifier.
• Explored the rates of deception in six popular review communities.
• Showed how review posting costs can be manipulated to reduce deception.
• Demo at ReviewSkeptic.com.
Thank you. Questions?

- Presented a framework for estimating the rate of deception in online review communities using a noisy classifier.
- Explored the rates of deception in six popular review communities.
- Showed how review posting costs can be manipulated to reduce deception.
- Demo at ReviewSkeptic.com.