

Roadmap

- Opinion Mining Problem
- Document sentiment classification
- Sentence subjectivity & sentiment classification
- Aspect-based sentiment analysis
- Aspect-based opinion summarization
- Opinion lexicon generation
- Mining comparative opinions
- Some other problems
- ➔ ■ **Opinion spam detection**
- Utility or helpfulness of reviews
- Summary

Opinion spam detection

(Jindal and Liu 2007, 2008)

- Opinion spamming refers to people giving fake or untruthful opinions, e.g.,
 - Write undeserving positive reviews for some target entities in order to promote them.
 - Write unfair or malicious negative reviews for some target entities in order to damage their reputations.
- Opinion spamming has become a business in recent years.
- Increasing number of customers are wary of fake reviews (biased reviews, paid reviews)

Problem is wide-spread

Professional Fake Review Writing Services

- [Post positive reviews](#)
- [Fake review writer](#)
- [Product review writer for hire](#)
- [Hire a content writer](#)

Manipulating Social Media (sock puppets - fake identities - fake personas)

- [Revealed: US spy operation that manipulates social media](#), Guardian.co.uk, Thursday 17 March 2011.
- [America's absurd stab at systematising sock puppetry](#), Guardian.co.uk, Thursday 17 March 2011.

China's Internet "Water Army" (Shuijun) - Opinion Spammers

- You can hire people to write and post fake reviews or comments, and even bribe staff at review, forum
- ['Water Army' Whistleblower Threatened](#), January 7, 2011, People's Daily.
- [The Chinese Online "Water Army"](#), June 25, 2010, Wired.com.
- If you read Chinese, see [this description](#) from Baidu Baike at baidu.com.

An example practice of review spam

Belkin International, Inc

- Top networking and peripherals manufacturer | Sales ~ \$500 million in 2008
- Posted an ad for writing fake reviews on amazon.com (65 cents per review)

Timer: 00:00:00 of 60 minutes

Want to work on this HIT? Want to see other HITs?

Write Product Reviews 25-50 Words
Requester: Mike Bayard
Qualifications Required: HIT approval rate (%) is not less than 95

Write a Positive 5/5 Review for Product on Website

Positive review writing.

- Use your best possible grammar and write in US English only
- Always give a 100% rating (as high as possible)
- Keep your entry between 25 and 50 words
- Write as if you own the product and are using it
- Tell a story of why you bought it and how you are using it
- Thank the website for making you such a great deal
- Mark any other negative reviews as "not helpful" once you post yours

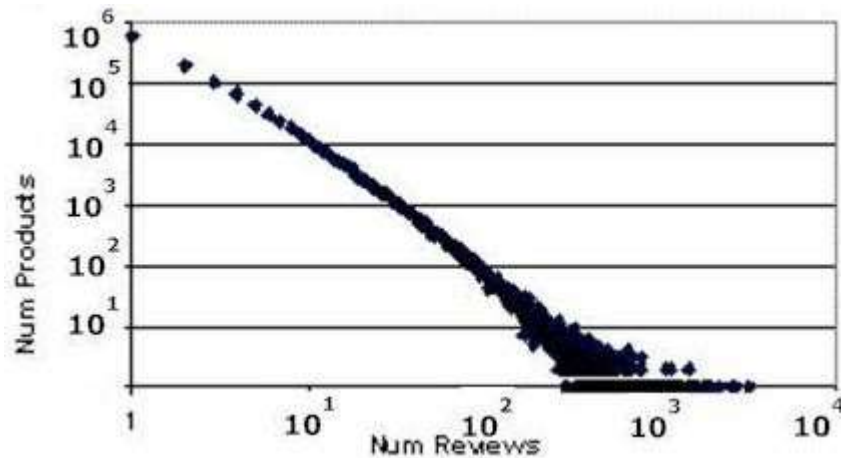
Instructions:

The link below leads to a product on a website. Read-through the product's features and write a positive review for it using the guidelines above to the best of your ability. I have also provided the part number for this product and you can click on the links below to see it on several alternative websites. In order to post some reviews you will need to create an account on the site. You can use your own email address or open a new free webmail account (gmail, yahoo...) and use it to post with.

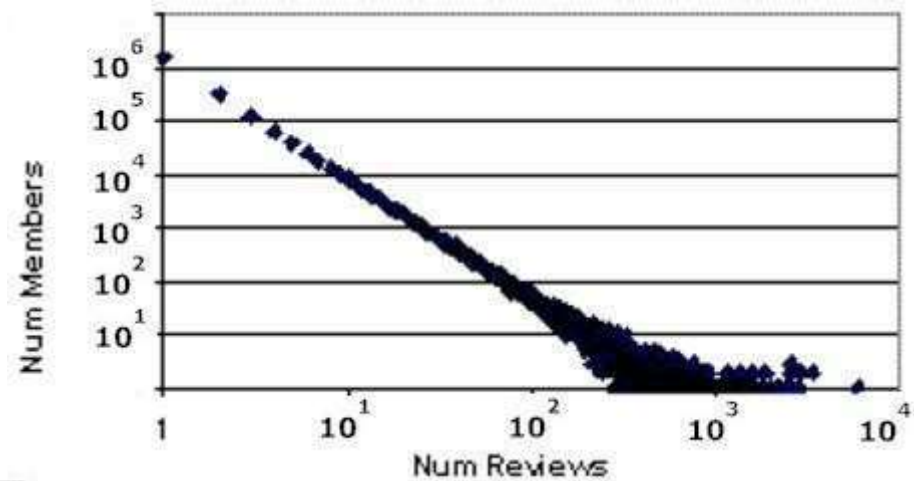
Jan 2009

Log-log plot

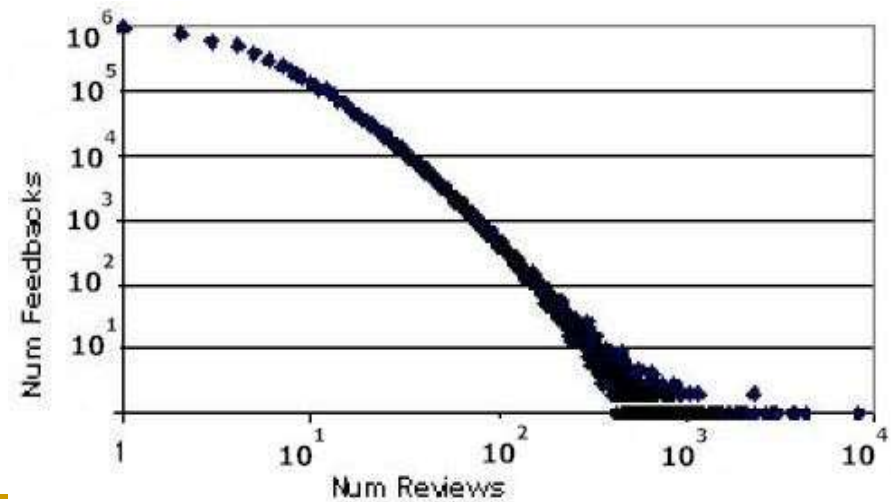
Amazon reviews, reviewers and products



■ Fig. 2 reviews and products



■ Fig. 1 reviews and reviewers



■ Fig. 3 reviews and feedbacks

Categorization of opinion spam

(Jindal and Liu 2008)

- Type 1 (fake reviews)

Ex:

- Type 2 (Reviews on Brands Only) (?)

Ex: *"I don't trust HP and never bought anything from them"*

- Type 3 (Non-reviews)

- Advertisements

Ex: *"Detailed product specs: 802.11g, IMR compliant, ..."*
"...buy this product at: compuplus.com"

- Other non-reviews

Ex: *"What port is it for"*
"The other review is too funny"
"Go Eagles go"

Type 1 Spam Reviews

- Hype spam – promote one's own products
- Defaming spam – defame one's competitors' products

Table 4. Spam reviews vs. product quality

	Positive spam review	Negative spam review
Good quality product	1	2
Bad quality product	3	4
Average quality product	5	6

Harmful Regions

- Very hard to detect manually

Harmful spam are outlier reviews?

- **Assumption:** Most reviewers and reviews are honest,
 - Not true when a group of people spam on a product (called group spam, discussed later).
- **Outliers reviews:** Reviews which deviate a great deal from the average product rating
- **Harmful spam reviews:**
 - Outliers are necessary but not sufficient condition for harmful spam reviews.
 - This idea helps us identify learning features.

Types of spam reviews

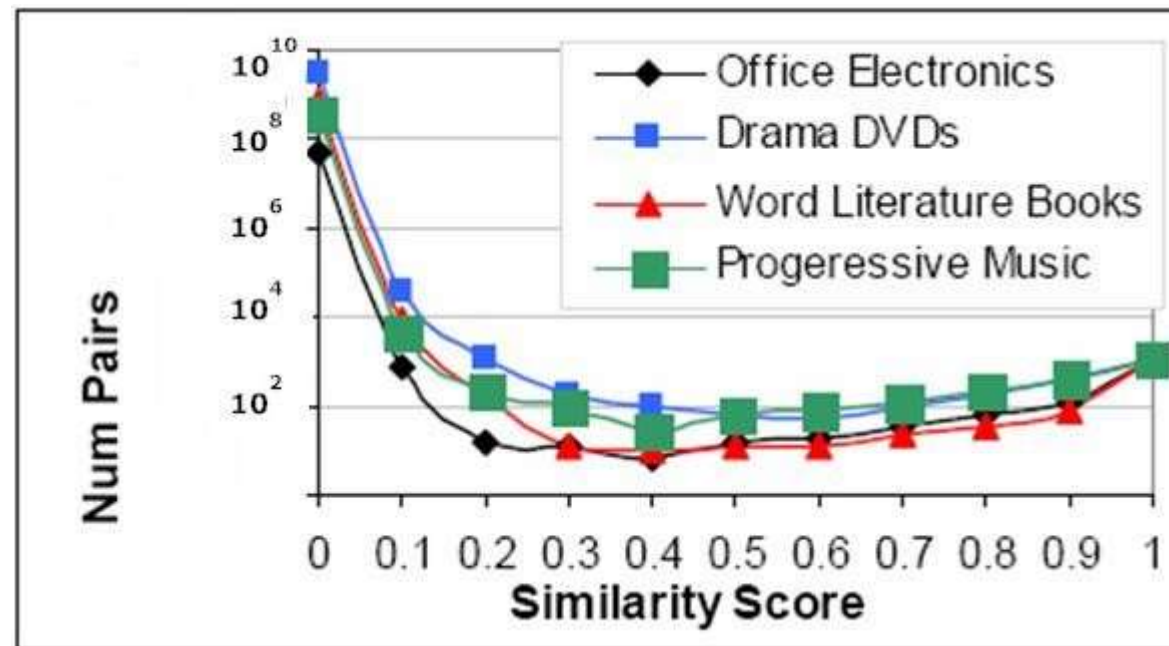
- Type 1 (fake review): These are reviews that give undeserving positive or negative opinions to some target entities.
- Type 2 (review on brand only): These reviews do not comment on the specific products that they are supposed to review, but only comment on the brands, the manufacturers, or the sellers of the products. Example: “I hate HP. I never buy any of their products”.
- Type 3 (non-review): These are not reviews or opinionated although they appear as reviews. There are two main sub-types:
 - Advertisements
 - Other irrelevant texts containing no opinions (e.g., questions, answers, and random texts).

Spam detection

- Type 2 and Type 3 spam reviews are relatively easy to detect
 - Supervised learning, e.g., logistic regression
 - It performs quite well, and not discuss it further.
- Type 1 spam (fake) reviews
 - **Manual labeling is extremely hard**
 - Propose to use duplicate and near-duplicate reviews as positive training data

Duplicate reviews

Two reviews which have similar contents are called duplicates



Four types of duplicates

1. Same userid, same product
2. Different userid, same product
3. Same userid, different products
4. Different userid, different products

■ The last three types are very likely to be fake!

Supervised model building

- Logistic regression

- Training: duplicates as spam reviews (positive) and the rest as non-spam reviews (negative)

- Use the follow data attributes

- Review centric features (content)
 - About reviews (contents (n-gram), ratings, etc)
 - Reviewer centric features
 - About reviewers (different unusual behaviors, etc)
 - Product centric features
 - Features about products reviewed (sale rank, etc)

Predictive power of duplicates

- Representative of all kinds of spam
- Only 3% duplicates accidental
- Duplicates as positive examples, rest of the reviews as negative examples

Table 5. AUC values on duplicate spam reviews.

Features used	AUC
All features	78%
Only review features	75%
Only reviewer features	72.5%
Without feedback features	77%
Only text features	63%

- reasonable predictive power
- Maybe we can use duplicates as type 1 spam reviews(?)

Tentative classification results

- Negative outlier reviews tend to be heavily spammed
- Those reviews that are the only reviews of products are likely to be spammed
- Top-ranked reviewers are more likely to be spammers
- Spam reviews can get good helpful feedbacks and non-spam reviews can get bad feedbacks

Roadmap

- Opinion Mining Problem
- Document sentiment classification
- Sentence subjectivity & sentiment classification
- Aspect-based sentiment analysis
- Aspect-based opinion summarization
- Opinion lexicon generation
- Mining comparative opinions
- Some other problems
- Opinion spam detection
- ➡ ■ **Utility or helpfulness of reviews**
- Summary

Utility or quality of reviews

- **Goal:** Determining the usefulness, helpfulness, or utility of each review.
 - It is desirable to rank reviews based on utilities or qualities when showing them to users, with the highest quality review first.
- Many review aggregation sites have been practicing this, e.g., amazon.com.
 - “*x of y people found the following review helpful.*”
 - Voted by user - “*Was the review helpful to you?*”

Application motivations

- Although review sites use helpfulness feedback to rank,
 - A review takes a long time to gather enough feedback.
 - New reviews will not be read.
 - Some sites do not provide feedback information.
- It is thus beneficial to score each review once it is submitted to a site.

Regression formulation

(Zhang and Varadarajan, 2006; Kim et al. 2006)

- **Formulation:** Determining the utility of reviews is usually treated as a **regression** problem.
 - A set of features is engineered for model building
 - The learned model assigns an utility score to each review, which can be used in review ranking.
- Unlike fake reviews, the ground truth data used for both training and testing are available
 - Usually the user-helpfulness feedback given to each review.

Features for regression learning

- Example features include
 - review length, review rating, counts of some POS tags, opinion words, tf-idf scores, wh-words, product aspect mentions, comparison with product specifications, timeliness, etc (Zhang and Varadarajan, 2006; Kim et al. 2006; Ghose and Ipeirotis 2007; Liu et al 2007)
- Subjectivity classification was applied in (Ghose and Ipeirotis 2007).
- Social context was used in (O'Mahony and Smyth 2009; Lu et al. 2010).