Opinion Spam And Analysis

Nitin Jindal And Bing Liu

WSDM, 2008

Present By Lingjie Zhang

Motivation

- Used by both customer and manufacturers
- Significant impact on product sales
- High review => more sales?
- Different from Web Spam and Email Spam

Defined Three Types Of Spam Reviews

- **Type 1:** Untruthful opinions
  - Also known as *fake reviews* or *bogus reviews*.
  - Deliberately mislead(promote/damage the reputation) readers or opinion mining systems.

Type 2: Reviews on brands only

- Do not contain specific product reviews but only brands / manufacturers / sellers
- May be useful; treated as spam in present study

Type 3 : Non-reviews

- Non-reviews, such as ads, or other irrelevant text without opinions

Amazon Dataset

- June 2006
- 5.8mil reviews, 1.2mil products and 2.1mil reviewers.

Amazon review components:
<Product Configuration>
<Reviewer ID>
<Rating>
<Date>
<Review Title>
<Review Body>
<Number of Helpful Feedbacks>
<Number of Feedbacks>

New feature:
Verified Purchase

Did you write reviews?
How many stars you would give?

Reviews, Reviewer, and Products

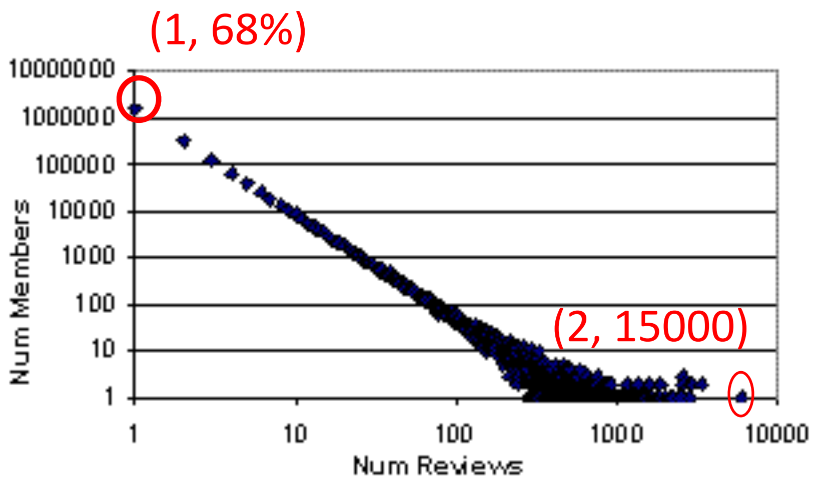


Figure 1. reviews and reviewers

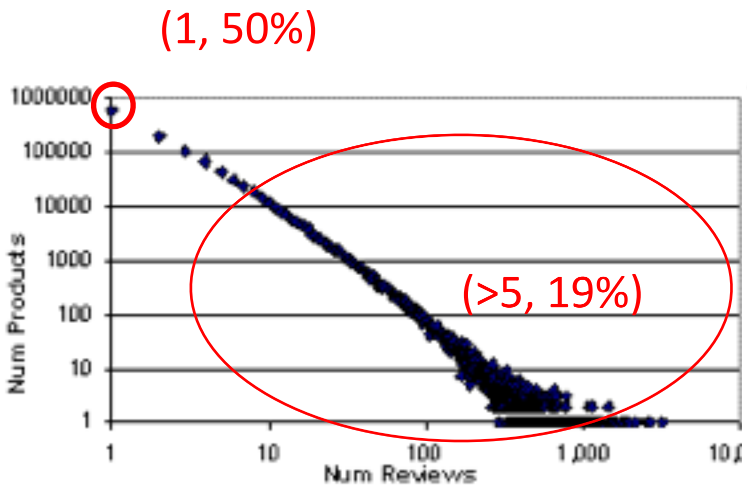


Figure 2. reviews and products

On average, a review gets 7 feedbacks.

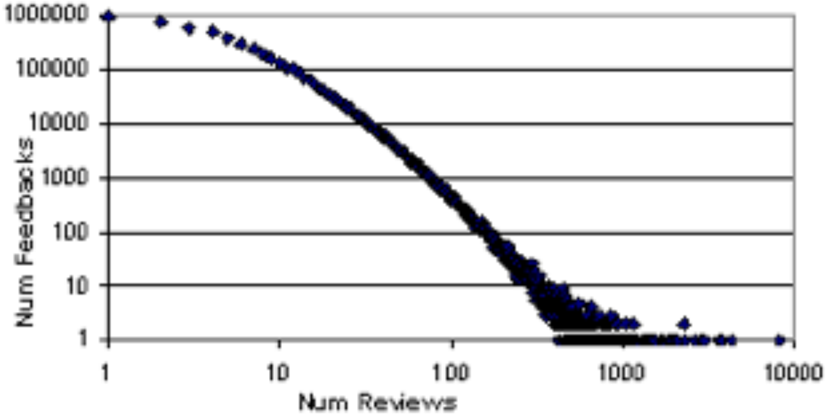


Fig. 3 reviews and feedbacks

Review Ratings and Feedbacks

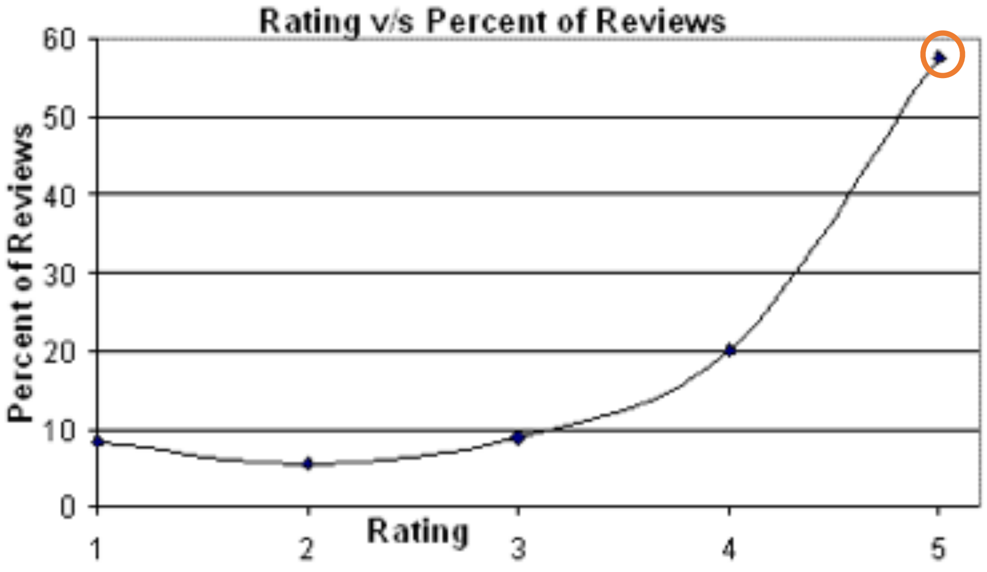


Figure 4. Rating vs. percent of reviews

- Rating of 5
  60% reviews
  45% of products
  59% of members

- Reviews and Feedbacks
  1st review – 80% positive feedbacks
  10th review – 70% positive feedbacks

Duplicates Duplicates Everywhere Everywhere!

- Three kinds of duplicates
  - Different user-ids on same product
  - Same user-id on different product
  - Different user-id on the different products

Detection of Duplicate Reviews

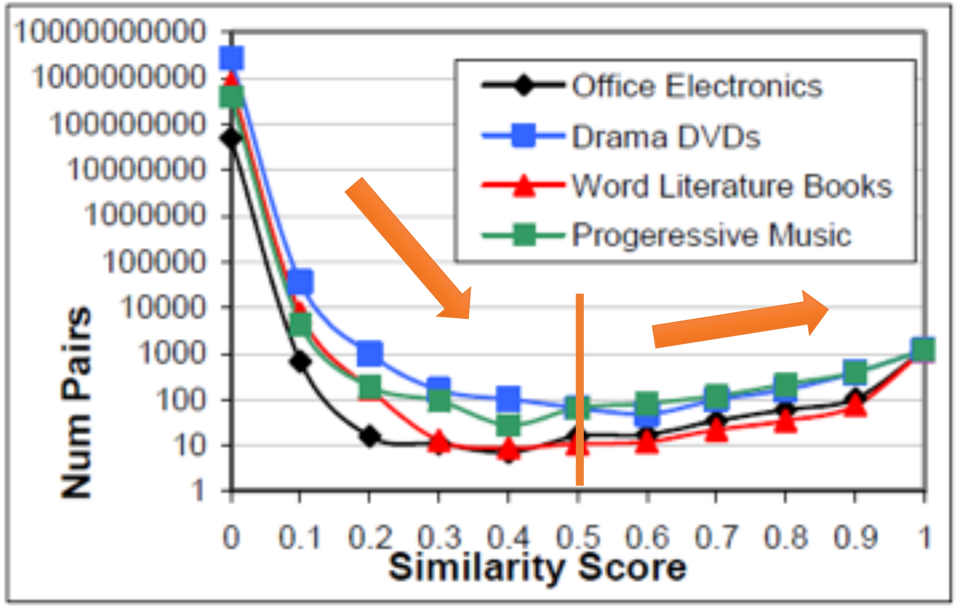


Figure 5. Similarity score and number of pairs of reviews for different sub-categories. Points on X axis are intervals. For example, 0.5 means between interval [0.5, 0.6).

(90%, <0.1), review diff products

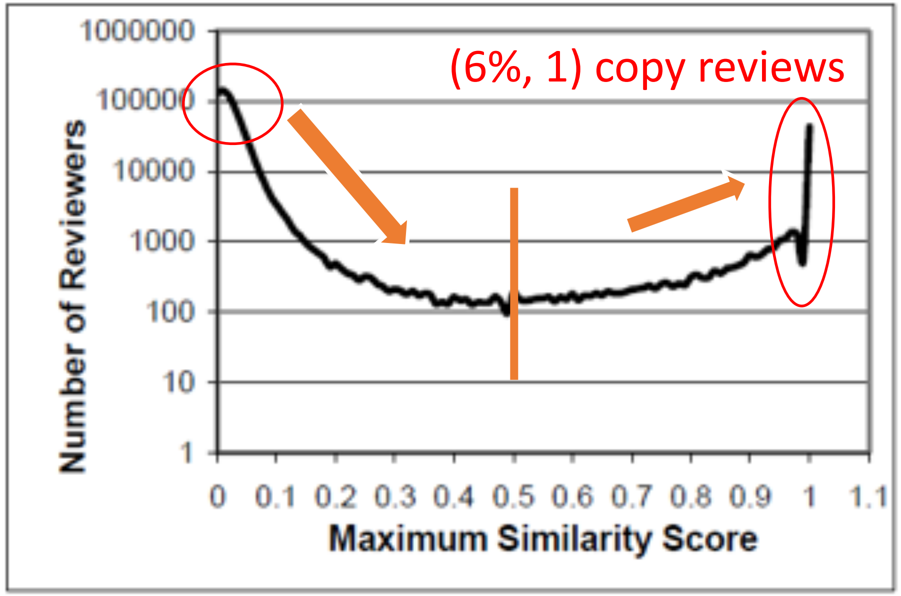


Figure 6. Maximum similarity score and number of members.

- Shingle method (2-grams)
- Jaccard distance (Similarity score, $J(S,T) = \frac{|S \cap T|}{|S \cup T|}$) > 90% ➔ duplicates.
- *The maximum similarity score* : the maximum of similarity scores between different reviews of a reviewer.

Detecting Type 2 and Type 3 Spam Reviews

- Spam types 2 and 3 are easy to classify manually, labeled 470 spam
- Use logistic regression
- 10-fold cross validation
- 38 features

Table 3. AUC values for different types of spam

Spam Type	Num reviews	AUC	AUC – text features only	AUC – w/o feedbacks
Types 2 & 3	470	98.7%	90%	98%
Type 2 only	221	98.5%	88%	98%
Type 3 only	249	99.0%	92%	98%

*High AUC(Area under ROC Curve) -> Easy to detect

Issues of Spam Detection

- Logistic regression works well for Type 2 (non-specific reviews) and Type 3 (non-reviews) spam
- Manual labeling of Type 1 Spam extremely difficult
- WHY?

*Paper published before Amazon review added a new feature

Making Use of Duplicates

- treat all duplicate spam reviews as positive examples, and the rest of the reviews as negative examples.

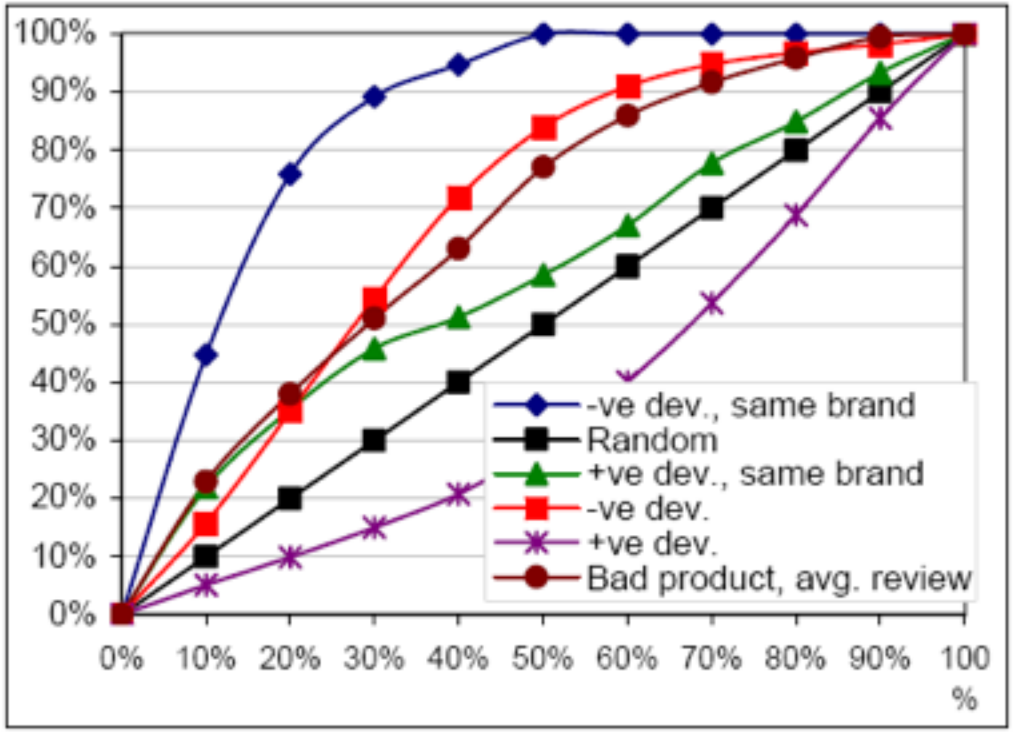
Table 5. AUC values on duplicate spam reviews.

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

  - good predictive power
  - How to check if it can detect type 1 reviews? (outlier reviews)

Lift Curve for outlier reviews

cumulated percentage of reviews of the current bin



X% of reviews (the test data)

The model built using duplicated spam as positive data is also predictive of non-duplicate spam reviews to a good extent.

Conclusions

- Review Spam and Detection
- Categorization into three types
- Type 2 and 3 easy to detect
- Type 1 difficult to label manually
  - Proposed to use duplicate reviews for detecting type 1 spam
  - Predictive power on outlier reviews

Questions and Thanks