

Online Review Spam Detection using Language Model and Feature Selection

Manali S.Patil
PG Student

Department Of Information Technology,
Pune Institute of Computer Technology,
S. No. 27, Dhankawadi, Pune-Satara Road,
Pune – 411043.

A.M.Bagade

Associate Professor
Department Of Information Technology,
Pune Institute of Computer Technology,
S. No. 27, Dhankawadi, Pune-Satara Road,
Pune – 411043.

ABSTRACT

In today's age of web 2.0, large numbers of product reviews posted on the Internet. Such reviews are important to customers or users and to companies. Customers use the reviews for deciding quality of product to buy. Companies or vendors use opinions to take a decision to improve their sales according to intelligent things done by other competitors. However, all reviews are given by customers or users are not true reviews. These reviews are given to promote or to demote the product. Some reviews are given on brand of product, and others are related to advertising of another product. There is need to find how many reviews are spam or non spam. In this paper, the system is proposed for detecting untruthful spam reviews using n-gram language model and reviews on brand spam detection using Feature Selection. Given system separately identifies spam and joined the result showing spam and non spam reviews.

General Terms

Data Mining, Natural Language Processing, Machine learning, Text Analytics.

Keywords

Review spam detection, Data Mining, Language Model, Features Selection.

1. INTRODUCTION

Web is widely used to interact with each other and to express different things as posting reviews of product, movies, etc., that means web users giving their own opinions. These opinions or reviews have importance in both users and vendors application. Consider the product review example, then on the user side, many opinions or reviews posted by different user or customers on the Internet for different products. During purchasing any product, public give importance to opinion of other publics about the same product. However, single query on the Internet gives many results that are giving different reviews or opinion. There is a need to develop a user application that gives exact result, i.e. opinion about a product. It becomes helpful to the customer or user to take a decision to buy the product. Furthermore, on vendor side, opinions are important for management of reputation and brand perception of product. Reputation management means finding out overall view about existing brand. Brand perception means how brand perceived by the customer. Opinions are beneficial to both user and vendors if the opinion posted are appropriately without any wrong intention.

However, today these opinions are fake for promoting or demoting product, giving a comment on another review or opinion, giving advertising links, giving a thumb up or down opinion, giving exclamatory marks, etc. Public intention to give an opinion is not true. Public giving spam or fake opinions.

In [1] N. Jindal et al. identifies three types of spam as follows:

Type 1 (Untruthful opinion / review): These reviews are given for misleading readers or opinion mining by giving more positive reviews to increase profit of the product called hyper spam or giving more negative reviews to damage the reputation of any product called defaming spam.

Type 2 (Review on band only): These reviews are not posted on product features, but on brand, manufacturer or sellers of product.

Type 3 (Non-Reviews): These reviews include advertising, other irrelevant things as questions, answers or smiley, etc.

This paper arranged in following section. Section 1 gives an introduction to review spam detection. Section 2 is a literature survey of opinion spam detection and spammer detection. Section 3 gives proposed work of opinion spam detection. Section 4 conclude with opinion spam detection.

2. LITERATURE SURVEY

This paper gives study of review spam detection in two different section first with different methods of detecting spam in review, and second is identifying the spammer groups.

2.1 Spam Detection

In [1], is the first attempt to study of spam detection that gives two methods for spam detection as duplicate detection and spam classification. They consider duplicate review is positive reviews, i.e. spam and others are negative reviews, and they use it for training a model to find out non-duplicate review. But text content is not enough for identification so that they use Naïve Bayes classification to classify spam and non spam review. They find out three types of duplicate positive reviews that used as a spam: (1) duplicates from different user id on the same product, (2) duplicates from the same user id on different products; and (3) duplicates from different user id on different products.

In [2], they identify three types of spam reviews as untruthful reviews, review on brand only and non review, then they gave following strategy for spam detection as: First detection of duplicate and near-duplicates using shingle method. The detection of review on brand and non review is based on machine learning and manual labeled example. Finally, detect

untruthful opinion spam that finds out three types of duplicates.

In[3], identify eight criteria as Proportion of Positive Singletons (PPS), Concentration of Positive Singletons (CPS), Reactive Positive Singletons (RPS), Review Weighted Rating (RWR), Contribution Weighted Rating (CWR), Truncated Rating (TR), Sentiment Shift (SS), Positive Review Length Difference (PRLD), then find the score matrix with these criteria for all hotels. The aggregation methods are used as Singular value decomposition (SVD) and Unsupervised Hedge algorithm to obtain suspicious review.

In[4], another work related to spam detection is finding unusual review pattern using Class Association Rules (CAR) that satisfy user given minimum support and minimum confidence constraints.

In[5], propose three approaches for finding deceptive opinion. First, Genre Identification test has carried out for each review to find out relation frequency distribution of part of speech tags in a text and is depended upon the genre of text. Second, Psycholinguistics Deception detection uses a tool as Linguistics Enquiry and word Count (LIWC) to detect four categories: Linguistics processes to find functional aspects, Psychological Processes to find all social, emotional, cognitive, perceptual and biological processes and anything related to timing and space, Personal Concern considers any references to work, leisure, money, religion, etc. Spoken categories have primarily filler and agreement words. Third, Text categorization approach allows us to model to both content and context with a n-gram features.

In[6], this is recent work in review spam detection is concerned with a problem of singleton review, i.e. the reviewer written only one review using time series pattern discovery in that they find the correlation between rating and volume of singleton reviews because as the review increases, rating is increases or decreases dramatically. They give a hierarchical framework for robust Singleton review spam detection.

In[7], proposes the method for untruthful review spam detection that done using text mining model and integrated into semantic language model and Non review spam detection that done by identifying different stylistic, syntactical and lexical features and SVM classifier applied to them.

2.2 Spammer Group

A spammer group is referred to group of reviewers who work together who work together writing fake reviews to promote or demote a set of target product.

In [8], proposes different behavior models based on review pattern. They focus on pattern of review content and ratings to define models, i.e. Targeting Product (TP), Targeting Group (TG), General Rating Deviation (GRD), Early Rating Deviation (ERD) and finally shows that spammer has more impact on rating.

In[9], proposes the review graph to show the relationship among reviewers, reviews and store that reviewer has reviewed. It also proposes computation method to calculate *trustiness* of reviewer, *honesty* of reviews and *reliability* of stores.

In[10], concentrate on Detecting spammer groups who write the reviews on the different product because it is easier to detect than groups who write on the single product. To meet this, propose a technique consist of three steps as: First

step is to find out candidate Spammer Group using frequent pattern mining by extracting review facts to find transaction with Reviewer Id and Product Id. Second step is to compute Spam Indicator values has been following eight criteria: Time Window (TW), Group Deviation (GD), Group Content Similarity (GCS), Member Content Similarity (MCS), Early Time Frame (ETF), Ratio of Group Size (RGS), Group Size (GS), Support Count (SC). Third step is ranking spammer group using SVM Rank.

In[11], find out two behavior indicators first are Group Spam Behaviour indicators that consist of above eight criteria[10] and second, Individual Spam Behaviour Indicators has four criteria as Individual Rating Deviation (IRD), Individual Content Similarity (ICS), Individual Early Time Frame (IETF), Individual Member Coupling in a group (IMC). Again, this study uses frequent item set mining to find spammer group and the use a behavioral model that derived from the relationship between product, individual reviewer and group reviewer and lastly uses GsRank to rank the spammer group.

3. PROPOSED WORK

The proposed architecture separately finds out three types of spam detection for each review and joined the result and give a last opinion to both vendors and users. Fig.1 is proposed architecture of review spam detection. First, there is need to collect the review from the Internet. Next, pre-processing, POS tagging, stemming and text mining of reviews is done. Next, spam detection for three types is done. Next, joined the result and do the analysis for identification of whether each review is spam or not spam.

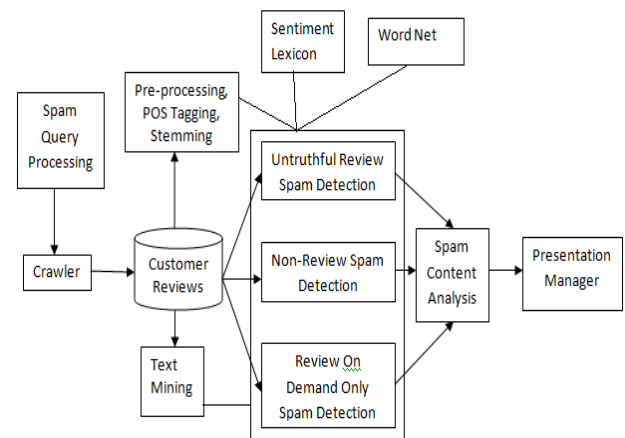


Fig. 1 Proposed architecture of review spam detection.

3.1 Spam Query Processing

While collecting reviews from the Internet search query is given, i.e. detection scope is decided for collecting review. Then, to find a similarity among search query and collected reviews cosine similarity is used. To find cosine similarity there is need to find out term frequency –inverse document frequency. To find tf-idf, each review and query is represented in vector form.

Let 'R' is collected reviews and $r \in R$. Review Vector is given as,

$$r_j = (w_{1,j}, w_{2,j}, w_{3,j}, \dots, w_{t,j}).$$

Query Vector, i.e detection scope for search is given as,

$$q = (w_{1,q}, w_{2,q}, w_{3,q}, \dots, w_{t,q})$$

Then Term frequency given by,

$$tf(t, r) = \frac{tc(t, r)}{\max\{tc(w, r) : w \in r\}}$$

where, $tc(t, r)$ = term count of 't' in review 'r'

$\max\{tc(w, r) : w\}$ = maximum term count of any word in review.

Inverse Document frequency calculated by,

$$idf(t, R) = \log \frac{|R|}{|r \in R : t \in r|}$$

where, $|R|$ = Total number of reviews and

$|r \in R : t \in r|$ = Number of reviews where term 't' appears.

Weight vector of review r is given as,

$$V_j = [w_{1,j}, w_{2,j}, w_{3,j}, \dots, w_{N,j}]^T$$

where, $w_{t,j} = tf(t, j).idf(t, R)$

Next, to measure similarity between two vectors cosine angle between review vector and query vector is measured and is given as,

$$\begin{aligned} sim(r_j, q) &= \frac{r_j \cdot q}{\|r_j\| \|q\|} \\ &= \frac{\sum_{i=1}^N w_{i,j} \cdot w_{i,q}}{\sqrt{\sum_{i=1}^N w_{i,j}^2} \sqrt{\sum_{i=1}^N w_{i,q}^2}} \end{aligned}$$

The resulting similarity -1 meaning exactly opposite, 1 meaning exactly the same, with 0 usually indicating independence, and in among -1 to 1 indicating intermediate similarity or dissimilarity among query and reviews [12].

3.2 Untruthful Review Spam Detection

Untruthful review contains either more positive or more negative review. Again, there is a possibility of three duplicate review [1]. The work only consider duplicate review of type different user id on same product using a n-gram language model. Language model is nothing but assign probability to sequence of 'm' words $p(w_1, w_2, \dots, w_m)$ by mean of probabilistic distribution.

N-gram language model used to predict next words in word sequence as (n-1)-order Markov model. The general formula for predicting next words is given by,

$$p(w_i | w_{i-(n-1)}, \dots, w_{i-1})$$

Relative frequency is given by,

$$p(w_n | w_{n-1}) = \frac{c(w_{n-1}, w_n)}{c(w_{n-1})} \dots (1)$$

An algorithm given is to find duplicate review in that each review is compare with all other reviews. The algorithm shows how two reviews are compared for duplicate identification. Suppose 'R1' and 'R2' are two reviews. Threshold is a value lies between 0 to 1 decided by taking different iteration of an algorithm with different values in 0 to 1 and value that give the more possible set of user id or spammer considered the threshold. Threshold can change according to product.

Algorithm: To find Untruthful Review Spam Detection.

Input : Set of all reviews on the same product.

Output : Possible user id set giving to duplicate review.

1. Find out probability of words sequence of review 'R2', i.e. P (R2) about words sequence of review 'R1' using n-gram model equation in (1).
2. If P (R2) >= Threshold
 - Review 'R2' is duplicated review.
 - Find out adjectives and frequency of adjectives.
 - Decide the review is positive or negative using SentiWordNet.
3. If P (R2) < Threshold
 - Check for words semantics among two reviews using Word Net.
 - If some words are semantic
 - Reviews are duplicated with modifying some words.
 - If not
 - Reviews are not duplicate.

3.3 Review on Brand Spam Detection

These reviews are not posted on product, but on brand, manufacturer or seller of product. To find this spam, there is a need to find features in reviews using feature selection algorithms. Feature selection algorithms are of two types: feature ranking and subset selection. Subset selection finds the set of all possible features for given data. Here decision tree is used to make a decision of reviews brand.

An algorithm considers the level of tree. Level-0 is root giving the product name. Level 1 giving general details and the level 2 onwards giving more details. Review giving above level 2 features are review on brand spam.

Algorithm: To Find Review on brand Spam Detection.

Input: Tree of set of possible features of the product in that leaf node giving more details features of its predecessor.

Output: A feature of reviews.

1. Find out the set of features from reviews that do not give wrong interpretation of phrases in the reviews.
2. For all possible combinations of features from reviews by checking from the root to leaf node of tree.
 - i. Find out features co-references and remove antecessor.
 - ii. Find out frequency of low referenced features and remove it.
3. If step two give features that match to the features in the tree, then according to level of a tree we say that review is on brand or not.

3.4 Non-Review Spam Detection

These reviews include advertising, other irrelevant things as questions, answers or similes, etc. To find an advertising links different link spam detection techniques [13] can be used. Again, work given in [6] by identifying lexical, syntactic and stylistic feature and use of SVM classifier is useful for such a detection. Furthermore, reviews compare to search query using cosine similarity to take decisions of non-review spam. Furthermore, rating, feedback given to review are helpful to know importance of reviews[8].

3.5 Spam Content Analysis

At this stage, result of three types of spam detection is analysed and decide each review is spam or not spam. This result is helpful to both users and vendor application during making their respective decision.

4. CONCLUSION

As individual users and companies use reviews and opinion for decision making, it is important to detect opinion spam and opinion spammer. This paper mainly concentrates on review spam detection. This paper proposes work for identifying untruthful review spam detection using n-gram model and review on brand spam detection using features selection techniques. In this paper, discussion regarding finding duplicate review, i.e. duplicate from different user id on the same product. Similarly, algorithm for duplicates from the same user id on different products and duplicates from different user id on different products can be written. The current study represents the initial investigation. Much work remains to be done. There is possibility for improving spam detection algorithm and finding spam in other domains.

The recent work related to spam detection is done with semantic language model. Our algorithm is using more sophisticated n-gram model give more effectiveness and efficiency. The result is useful to both customers while purchasing any product. And to company to improve their sales performance using true reviews.

5. REFERENCES

- [1] Nitin Jindal, Bing Liu, "Review Spam Detection", *ACM Proceedings of the 16th international conference on World Wide Web*, pp-1189-1190,2007.
- [2] Nitin Jindal, Bing Liu, "Opinion Spam and Analysis", *ACM Proceedings of the international conference on Web search and web data mining*, pp.219-229, 2008.
- [3] Guangyu Wu, Derek Greene, Pádraig Cunningham, "Merging multiple criteria to identify suspicious reviews", *Proceedings of the fourth ACM conference on Recommender systems*, pp.241-244,2010.
- [4] Nitin Jindal, Bing Liu, Ee-Peng Lim "Finding unusual review pattern using unexpected rules", *Proceedings of the 19th ACM international conference on Information and knowledge management*, pp.1549-1552,2010.
- [5] Myle Ott, Yejin Choi, Claire Cardie, Jeffrey T. Hancock, "Finding deceptive opinion spam by any stretch of imagination", *ACM Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, pp.309-319,2011.
- [6] Sihong Xie, Guan Wang, Shuyang Lin, Philip S. Yu "Review spam detection via time series pattern discovery", *ACM Proceedings of the 21st international conference companion on World Wide Web*, pp.635-636,2012.
- [7] Raymond Y. K. Lau, S. Y. Liao, Ron Chi-Wai Kwok, Kaiquan Xu, Yunqing Xia, Yuefeng Li, "Text mining and probabilistic modeling for online review spam detection" *ACM Transactions on Management Information Systems (TMIS)*, Volume 2 Issue 4, Article 25,2011.
- [8] Ee-Peng Lim, Viet-An Nguyen, Nitin Jindal, Bing Liu, Hady Wirawan Lauw, "Detecting product review spammer using rating behaviors", *Proceedings of the 19th ACM international conference on Information and knowledge management*, pp-939-948,2010.
- [9] Guan Wang, Sihong Xie, Bing Liu, Philip S. Yu "Review graph based online store review spammer detection", *Proceedings of the 2011 IEEE 11th International Conference on Data Mining*, pp.1242-1247,2011.
- [10] Arjun Mukherjee, Bing Liu, Junhui Wang, Natalie Glance, Nitin Jindal, "Detecting group review spam", *ACM Proceedings of the 20th international conference companion on World wide web*, pp.93-94,2011.
- [11] Arjun Mukherjee, Bing Liu, Natalie Glance, "Spotting fake reviewer group in consumer reviews", *ACM Proceedings of the 21st international conference on World Wide Web*, pp.191-200,2012.
- [12] en.wikipedia.org.
- [13] Dr.S.K.Jayanthi, Ms.S.Sasikala, "link spam detection based on DBSpamClust with Fuzzy C-Mean Clustering", *International Journal of Next-Generation Networks (IJNGN) Vol.2, No.4*, pp.1-10, 2010.