Review Spam Detection

By: Ziyi Hu-100981750

Dongyang Li-7996582

**Abstract**: Review spam detection plays an important role in the e-commerce. In this work, we used four spam review features and three classifiers in weka. The best result is when we weight product feature more and use SVM in weka. The result of this approach shows that **80.7725%** accuracy was achieved in identifying the spam review from the amazon reviews.

**Keywords**: Spam Detection, SVM, Amazon reviews

**1 Introduction**

As the rapid development of Internet, people rely on web much more than before. However web spam is also huge that can be malicious manipulation of user’s generated data. There are several forms of web spam, such as email spam and search engine spam. In our project, we are more focusing on review spam from commercial websites like Amazon.com. On this kind of website, consumers shop and write down reviews for the product they bought which would be very useful for others to make decisions to purchase. The reviews are based on honesty. Unfair reviews can influence consumers’ perception of products.

Out aim is to detect spam reviews. It is hard to detect review spam one by one manually because spammers generate spam in so many different ways. However after we compare one with other reviews, we can find some patterns or rules to detect potential unfair reviews.

In this paper, we proposed a rule-based approach to figure out weather one review is spam or not. To find out how spammers spam is hard, but we can find out what are potential spam behavior for spam reviews. We can easily get five features for a review. Each feature represents a degree whether it is spam or not. Each has a weight and assigns a numeric score. Spam review can be detected by adding five features’ scores up and see whether it is larger than a threshold which is determined through experiments.

**2 Related Work**

We survey the related research on spam review detection. In this field, researchers present approaches for different applications and in different entry points. Then we compare our approach to others.

**Review sentiment analysis.** By analyzing the content, these works focused on reviews whether they are positive or negative[3]. In spite of the overall reviews, it does not consider whether it is spam or not.

**Review spam detection.** This part of work is essential important for commercial websites to improve the precision of average rating prediction and reviews validation. [4] shows investigation in this part. There are three types of spam in this work, untruthful reviews, reviews on band and non-reviews (advertisements). By a set of review-level, reviewer-level and product-level features to assign spam labels to reviews. Another aspect is to detect unfair ratings, [2] represents a way to cluster unfair high ratings and low ratings, then use third party ratings on producers of ratings and find ratings from less reputable producers as unfair ratings.

**Spam item detection.** This work is to identify items may be spam by figuring out singleton reviews [5]. Singleton review refers to a review is written by one reviewer, but this reviewer never write reviews on other items.

**Review spammers detection.** Rather than detecting spam review, this is aiming at finding out spammers [1]. It has great advantage to go for this way. It can simplify the process of detection. For example, if one review’s content is appeared in another product’s review. Then this review is spam with high possibility. Also the person who wrote this review should be spammer. So we can easily remove all reviews written by this person.

Compared with different strategy to detect spam review, our approach is a rule-based method that through several experiment we can achieve a high accuracy of detection. Also rule-based method is flexible and has a good extendability. Now we have five features such as content, rating. If we find another pattern that can help us to predict, we can add this feature easily.

**3 Dataset**

In this paper, we use the reviews from Amazon.com. Figure 1 shows

an example of a review from the website.

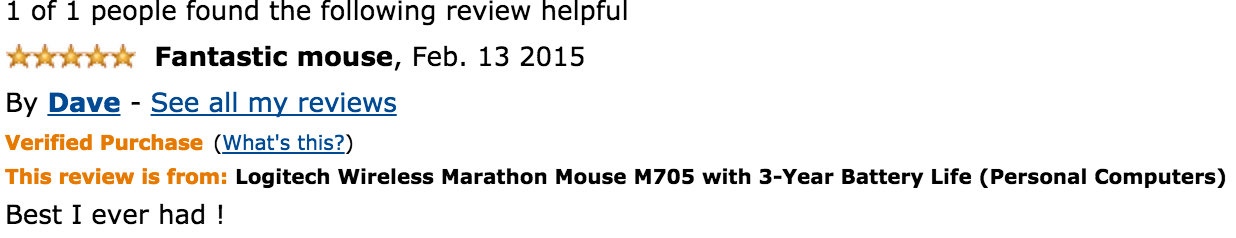


Fig 1. Example of review from Amazon.com

For each product, the dataset has a average rating. And for each review, it contains user name, rating, date, content, the number of people who think this review is useful or not useful.

The data we get is like below fig 2. It contains product ID, user ID, the number of people who think this review is useful, the number of people who rate this review, title and the content.

|  |
| --- |
| A1004AX2J2HXGL B0007RT9LC 3 4 5.0 The film speaks for itself The only thing missing is a presentation of the original 78 minute film "Some Folks Call it a Sling Blade" (1993) which later the feature length film was created from. Can't have it all, I guess. Perhaps when they release the extra-special edition? If you like Bill Bob in this film, you should also check out "Dead Man" (he's in exactly one hilarious scene) and "The Man Who Wasn't There". |

Fig 2. Example of data set

**Pre-processing.** In order to better use it in Weka, several steps need to be performed.

* + - * + Write Java to convert it into arff format. In the arff format, we have attributes content, category (rating user made) and the label (whether it is spam).
        + To improve the accuracy in Weka, we have to eliminate stop words, rare words.
        + Calculate average rating for each product.

Finally, we get a Weka format file. Then write Java program to label every review and put it into Weka to calculate the accuracy by using Machine Learning methods.

**4 Five Spam Detection Features**

The resource we get involves five features from product view, reviewer view and review view.

* + - **Product Feature.** Spam review’s rating usually is extremely high or low. The first step is to calculate the average rating for the product. If the rating is too high or too low, it’s very likely the reviews are faked.
    - **Reviewer Feature.** In the data set we have two numbers, one is the number of people who criticized this review noted as X and the other is the number of people who think this review is useful noted as Y. We consider a review is spam if Y/X < 10%.
    - **Review Feature.**  This feature corresponds to the relationship between the length of the review and the length of the title. If this review rating is around 4, and the review show the positive polarity, but people do not like it so much. If one person don’t like it very much and don’t hate it, they would not spend too much time on writing the review especially the title is very short. It’s very likely the review is faked.
    - **Textual Feature.** This feature measures the total number of numbers, positive words and negative words that occur in the sentence. In one review, if there are more than 5 numbers, we consider this review may be true because numbers can represent detailed judge to the product. If the positive words or negative words are more than 20. The review maybe faked, because people won’t just give too good words or bad words to a product. This review may not be true. We take this as the fourth feature.
    - **Overall Rating and Current Rating.** This feature measures average rating and the rating from this specific review. If the average rating is 5 and this rating is 1, it’s very likely this one is faked. If the average rating minus the current rating is more than 3.5 or less than -3.5, we consider it may be faked review.

We give each of them a weight. For each feature, we calculate a degree. Then multiply degree to corresponding weight and get a final result. Compare this result to a defined threshold, if bigger then this is spam. During the experiment, we will adjust the weight of each feature to improve the accuracy.

**5 Experiment and Results**

In this study we conducted several experiments to examine our proposed features for the spam review detection.

Product reviews on E-commerce website are the most popular test-bed for the spam review detection. We use the Amazon reviews dataset. We give each of the spam detection features a parameter. If all the true parameters add up over 0.6, we consider this review is spam, otherwise the review is non-spam. From the manually labelled dataset, and we have 434 spam reviews together with 744 non-spam reviews.

We build a system to detect the spam reviews and the system has five feature.java classes. Each time the system read one line from the dataset. It will use the five feature .java to determine whether the five feature is true or not for that line. The code is attached in the report. There are two java packages, import them into the eclipse and run from the spam1 package ReviewParser.java then get the arff file. The parameters can be changed in the spam package Product.java.

Name the ProdutFeature as the FPro ，the RatingFeature as the FRating , the ReviewFeature ad the FRe1 , the ReviewerFeature as the FRe2 ，the TextualFeature as the FTextual.

The dataset doesn’t have a lot of reviews for the same product and the number of the true RatingFeature is too small, so in the experiment we don’t consider the RatingFeature. However, the system has the RatingFeature.java and can determine the relationship between the average rating and current rating

In the experiment, we compare different parameters set and used three machine learning algorithms in the weka. They are Bayes, SVM and J48 to see the accuracy.

If we only consider one parameter, the FPro gives 608 spam reviews, the FRating gives 0 spam review, the FRe1 gives 39 spam reviews, the FRe2 gives 29 spam reviews, the FTextual gives 734 spam reviews. So the ProductFeature and TextualFeature should weight more than others in the review spam detection.

For the parameters set, we first consider two of them are more important than others, so we set two of them equals to 0.3 and two of them equals to 0.2, if the summary of the parameters is more than 0.6, then the review is spam, so only the two parameter set 0.3 is true or more than two of them is true, then the review is spam. Then, we set three of them equals to 0.3, only one feature is less important than others. For some parameters set, the accuracy of the result is too low and we delete them from the result table.

The training data is the manually labelled data.

The test data is after we changed the parameter in the system and the system will give the .arff. Put them in weka and get the result.

The method used in the weka is the 10 cross validation folds, that is 9 training data and one test data.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| FPro | FRe1 | FRe2 | FTextual | Bayes Accuracy | SVM  Accuracy | J48  Accuracy | Spam | Nonspam |
| 0.3 | 0.3 | 0.3 | 0.3 | 70.9871% | **78.2833%** | 76.1373% | 415 | 763 |
| 0.3 | 0.25 | 0.25 | 0.35 | 70.9013% | **77.3391%** | 75.7082% | 410 | 768 |
| 0.35 | 0.25 | 0.25 | 0.3 | 73.9056% | **80.7725%** | 78.6266% | 387 | 791 |
| 0.3 | 0.2 | 0.2 | 0.3 | 75.279% | **78.97%** | 76.9099% | 382 | 796 |
| 0.3 | 0.2 | 0.3 | 0.3 | 73.9056% | 79.2275% | **79.3991%** | 392 | 786 |
| 0.3 | 0.3 | 0.2 | 0.3 | 71.588% | **77.5966%** | 74.4206% | 405 | 773 |

The first one is every two of the features true, then the review is spam. The second one is that FTextual and any other feature is true, we set the review spam, while for the other three, only three of them is true, the review is considered as spam. For the third one, FPro add any other true feature is the spam review, for the other three, only all of them is true the review is labelled spam. After that, the only spam review for two true features is the FPro add the FTextual, the other spam review need three true feature. The next one is that all the other two true features make the review spam, while if the true features contain the FRe1 it need three features to make the review spam. The last one is the same as the fourth one, just change FRe1 to FRe2.

This experiment aims at examining these four features’ contributions to the spam review detection. From the results of the Amazon review dataset, we found that the parameters set up to 0.3, 0.3, 0.2, 0.3 achieved the lowest accuracy (77.3391%). The highest accuracy is the parameters set up to 0.35, 0.25, 0.25, 0.3. In particular,the weight of the feature FPro was shown to be the most effective indicator of spam review detection. The parameters set up to 0.3, 0.2, 0.3, 0.3 achieved the second highest spam reviews detection (79.3991%), only slightly lower than the best one. Generally, the rising weight of the FRe1 will increase the accuracy. On the other hand, the increasing weight of the FRe2 will decrease the accuracy. If increase the last feature FTextual and decrease the frist feature FPro , the accuracy will decrease.

The results show that, for each parameter set of the five features, SVM always outperformed the other two algorithms in terms of classification accuracy. Only for the parameters set up to 0.3, 0.2, 0.3, 0.3, the second highest accuracy the highest algorithm is the J48.

From the experiment results, we found that the five rule-based features to detect the spam review performed good on the Amazon dataset. The different features influenced the result differently. The first one is the product feature. Normally, the faked reviews are for two purposes, one is to attract others to buy this product and these reviews will give high rating for the product, the other one is to stop people buying this product and these reviews will give very low rating for the product. In the experiment, if the product’s rating has 5 star or 1 star feature trends to be the best base on the reason of most faked reviews have the rating of 1 or 5. The other feature which should weighted more is the reviewer feature, that is percentage of feedbacks think this review is useful out of all the feedback. This feature is based on other people’s opinion which makes this feature very useful. The other two features seems less important. The review feature is the relationship between the length of the review content and the length of the title. The textual feature is the number of numbers, positive words and negative words that occur in the content. This two feature do not give a good support for the review spam detection.

**6 Conclusion and Future Work**

In conclusion, we used five features to detect the spam review and build a system to identify the review from a txt file. We used the Amazon review dataset that contains more than 1000 reviews. The test data achieved high accuracy in spam review detection. However, there are some limitations for our project. First, the dataset does not contain a lot of reviews that average rating and current rating are very different. Secondly, the features we used is not so complicated, if anyone learned about the features, they may faked reviews that avoid those features. Finally, there is still a lot of room of performance improvement on our spam review detection. One promising line of future research is to refine rules in the spam reviews detection. Add more features to the spam review and do experiment based on that, find out the subset of the most useful features.

**References**

[1] LIM, Ee Peng; Liu, Bing; Jindal, Nitin; Nguyen, Viet-An; and Lauw, Hady Wirawan. Detecting Product Review Spammers using Rating Behaviors. (2010). ACM Conference on Information and Knowledge Management (CIKM2010). Research Collection School Of Information Systems.

[2]  C. Dellarocas. Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior. In ACM EC, 2000.

[3]  M. Hu and B. Liu. Mining and summarizing customer reviews. In KDD, 2004.

[4]  N. Jindal and B. Liu. Opinion spam and analysis. In WSDM, 2008.

[5]  G. Wu, D. Greene, B. Smyth, and P. Cunningham. Distortion as a validation criterion in the identification of suspicious reviews. Technical Report UCD-CSI-2010-04, University College Dublin, 2010.