Mining Product Reviews for Spam Detection Using Supervised Technique

M. Daiyan¹, Dr. S. K. Tiwari², M. A. Alam³

¹Research Scholar, Institute of Computer Science & IT, Magadh University, Bodh Gaya.
²Associate Professor, PG Department of Mathematics, Magadh University, Bodh Gaya.
³Research Scholar, School of Computer Science, Singhania University, Rajasthan

Abstract— By the introduction of web 2.0 and cheap accessibility of web, many merchant sites are operating and providing space for their users to share their experiences in form of customer reviews. Such reviews contain precious knowledge useful for both customers as well as manufacturers. E-customer accesses these reviews to know opinion expressed by existing users on a product before making purchase decision. Further, such reviews are used by manufacturers to know shortcoming in their existing products as well as to know strength of a competitor products for making business plans. Since Internet has no quality control, anyone can write anything which results in low quality reviews that contain biased information known as spam, and may mislead the customer affecting his buying decisions. Thus, it is very essential to have a mechanism which is capable of assessing the trustworthiness of reviews for proper decision making or for marketing intelligence. In this paper we propose a supervised method for spam detection. Dataset are taken from merchant sites like amazon.com. However our experimental results show our proposed method is very effective over the existing method.

Keywords— Text Mining, Feature Extraction, opinion Mining

I. INTRODUCTION

The Internet is public cooperated & self sustaining facility accessible to billions of people worldwide. Internet which literary means ‘network of networks’ includes the use of information entertainment, socialization, education and advertisement. Internet is growing fields now days. It has got impetus in the last decade of the century. The Internet has a lot of potential as commercial medium.

Introduction of Web 2.0, use of internet becomes more flexible as far as the possibility of choosing technologies is concerned. Easier and faster access to information, when and where it is needed is possible. Also, it integrates verity of technologies in the teaching-learning activities. It promotes amateurishness by invaluable contents generated by users. Due to rapid expansion of e-commerce, web documents are receiving great attention as a new medium that describes individual experiences.

Due to the existence of numerous merchant sites, forums, discussion groups and blogs, individual users are participating more actively and generating vast amount of new data which is symbolized by new terms such as “Blog journalism”, “Consumer generated mediums” and “User Generated Contents”. As customer feedbacks on the Web influence other customer’s decision, these feedbacks have become an important source of information for business to take into account while developing marketing and product development plans. It has been observed that user generated contents includes junk and unnecessary information known as “spam”, whose identification is important to safeguard the interest of large web community.

Spam is basically an electronic message system (including most broadcast media, digital delivery systems) to send Spontaneous bulk messages aimlessly same message large number of recipients on the Internet¹. Most spam on the Internet today is advertisements from individuals and the occasional small business looking for a way to make a fast buck. Spam messages are usually sent using sophisticated techniques designed to mask the messages true senders and points of origin. There exist difference between electronic spam and conventional marketing techniques, thus identification and development of preventive methods is required to stop spam, otherwise it could ultimately destroy the usefulness of the Internet. Following sub-section describes types of spam.

1.1 Types of Spam

1.1.1 Email Spam: Email spam targets individual users with direct mail messages. Email spam lists are often created by scanning Usenet postings, stealing Internet mailing lists, or searching the Web for addresses¹.

1.1.2 Comment Spam: Another category includes, comment spam which is widely used by spammer by posting comments for their nefarious purpose¹.
1.1.3 Instant Messaging spam: This type of spam makes use of instant messaging systems. Instant messaging (IM) is a form of real-time direct text-based chatting communication in push mode between two or more people using personal computers or other devices, along with shared clients. Text is conveyed over a network. It is very common on many instant messaging systems such as Skype¹.

1.1.4 Junk Fax: It is form of telemarketing where unsolicited advertisements are sent via fax transmission. Junk faxes are the faxed equivalent of spam or junk mail. It is an advertising medium¹.

1.1.5 Unsolicited Text Messages Spam or SMS Spam: This type of spam (SMS) is hard to filter. SMS spam is going to be the next big challenge as the fast development of technology allows SMS spam to be sent at relatively low cost using Internet SMS portals¹.

1.1.6 Social Networking Spam: It is spam directed at users of internet social networking services such as MySpace, Facebook or LinkedIn. The Users of social networking services can send note that include embedded links to other social network locations or even outside sites, to one another. This is where the social network spammer comes in. Utilizing the social network’s search tools, target

-----------------------------------------------------------------------


a certain demographic segment of the users, or use common fan pages or groups to send notes to them from an account disguised as that of a real person. Such notes may include embedded links to pornographic or other product sites designed to sell something¹.

1.1.7 Review Spam: In merchant sites such as www.amazon.com, www.cnet.com etc. review spam is designed to give unfair view of some products so as to influence the consumer perception of the products by directly or indirectly inflating or damaging the product’s reputation. It was found that 10 to 15% of reviews essentially echo the earlier reviews and may potentially be influenced by reviews spam [1]. There exist three types of review spam:

Type 1 (Untruthful opinions): It purposely mislead readers or opinion mining system by giving undeserving positive reviews to some target objects in order to promote the objects and or by giving malicious negative reviews to some other objects in order to damage the reputation.

Type 2 (reviews on brands only): Spam that comments only the brands, the manufacturers or the sellers of the products. Although they may be useful, it is consider as spam because they are not targeted at the specific products and are often biased.

Type 3 (non-reviews): It is non-reviews, which have two main sub-types: (a) advertisements and (b) other irrelevant reviews containing no opinions (e.g., questions, answers, and random texts).

Spam detection can be regarded as a classification problem with two classes, spam and non-spam. However, due to the specific nature of different types of spam, it requires dealing in different ways. For spam reviews of type 2 and type 3, detection can be based on traditional classification learning using manually labeled spam and non-spam reviews because these two types of spam reviews are recognizable manually[2]. It is important to identify and filter out the review spam. There is no reported study on the trustworthiness of opinions in reviews. As internet has no quality control, anyone can write anything on the web which results in many low quality reviews, and worse still review spam which is often biased and may mislead the customer affecting his buying decisions. Thus, it is very essential to have a mechanism which is capable of assessing the trustworthiness of reviews for proper decision making or for marketing intelligence. Trusted customer reviews are useful for both potential buyers and product manufacturers. It is more convenient and less time consuming for buyer to see at a glance feature by feature comparison of reviews written by most of the customers in taking buying decisions without getting biased and product manufacturer gets to know strengths and weaknesses of his/her own products and also that of the competitors, consumer preferences and interests by which profits could be maximized [3].

II. RELATED WORK

A Spam is many copies of the same message, in an attempt to force the message on people who would not otherwise choose to receive it. Most spam is commercial advertising, often for dubious products, get-rich-quick schemes, or quasi-legal services. Spam costs the sender very little to send most of the costs are paid for by the recipient or the carriers rather than by the sender. In [1], the author proposes scoring methods to measure the degree of spam for each reviewer and apply them on an Amazon review dataset. They mainly focus on Spam found in online product review sites commonly known as, review spam or opinion spam.
Study in [2], attempts to detect a review is a spam or a non spam, in order to provide a trusted review to help the customer in making the proper buying decision. The trustworthiness of the reviews is assessed as spam or a non spam review which includes both duplicate and near duplicate reviews classified as spam reviews, and partially related and unique reviews classified as non spam reviews. The author have proposed a novel and effective technique, namely Shingling technique, for detecting spam reviews based on the product features that have been commented in the reviews. In [4], a preliminary study was performed and it has been stated that, review spam detection is a relatively new research problem which has not yet been well studied. A more in-depth investigation has been done in [5] in which, three types of review spam were identified, namely untruthful reviews (reviews that promote or defame products), reviews on brands but not products, and non-reviews (e.g., advertisements). By representing a review using a set of review, reviewer and product-level features, classification techniques are used to assign spam labels to reviews. In particular, untruthful review detection is performed by using duplicate reviews as labeled data. Further researchers attempted in [6], identify items that are targets of spamming by identifying singleton reviews on the reviewed items. Singleton reviews are the reviews written by users who contribute only one review each. These users subsequently contribute no other reviews. Proportion of positive singleton reviews, concentration of positive singleton reviews, and rating distortion caused by singleton reviews are thus used to analyses possibly spammed hotels in Trip Advisor [4]. Study in [7], performs on usability of linguistic features in the Web spam classification task. The features were computed on two Web spam corpora i.e. Webspam-Uk2006 and Webspam-Uk2007. Further in [8], Miklos Spam Filtering in Internet Archives illustrate spam filtering needs, opportunities and blockers for Internet archives via analyzing several crawl snapshots and the difficulty of grating filter models across different crawls via the example of the 13 .uk snapshots performed by UbiCrawler that include WEBSPAM-UK2006 and WEBSHAMUK2007. Keeping in mind the above, our assumption is that effective spam detection and filtering system is required to save the interest of e-community from spam.

III. SPAM DETECTION TECHNIQUE

In general, spam detection techniques can be broadly classified in to two types namely Supervised Techniques and Unsupervised Techniques.

A further detail of these techniques is given in following sub section.

3.1 Supervised Technique: Supervised spam detection techniques require labeled review spam data set to identify review spam. Its uses several supervised methods, including SVM, logistic regression, Naive Bayes etc.

3.2 Unsupervised Techniques: In unsupervised methods refers to problem of trying to find hidden structure in unlabeled data.

Unsupervised methods consist of:
- clustering (e.g., k-means, mixture models, hierarchical clustering),
- Blind signal separation using feature extraction technique for dimensionality reduction.

IV. MOTIVATION AND PURPOSED WORK

Spam is a big problem facing by Internet community in today’s world. As internet has no quality control, anyone can write anything on the web which results in many low quality reviews, and worse still review spam which is often biased and may mislead the customer affecting his buying decisions. Thus, it is very essential to have a mechanism which is capable of assessing the trustworthiness of reviews for proper decision making or for marketing intelligence.

The information contained in a review message is divided into two fields first the header (details containing general information on the message, such as the subject, sender and recipient) and body (the actual contents of the message). Before the available information can be used by a classifier in a filter, appropriate pre-processing steps are required.

We have purpose supervised classification approach for spam filtering. The main steps involved in our work consist of following components.

Web Crawler: used to download target documents from merchant / web sites & store locally in database for further processing.

Document Pre-processor: converts the set of words present in the message to a specific format required by the machine learning algorithm. In documents Pre-processor main steps are followings:

Tokenization: used to extracts the word from messages.

Lemmatization: it reduces words to their root forms (e.g., getting to get).

Word Remover: remove words that frequently happen in many messages.
Spam classifier: classifies the message or documents as spam or legitimate using supervised classification techniques.

V. DATASET

For an experiment we use product reviews which are obtained from various merchant sites like www.amazon.com. It is very large and covers all range of products. In amazon site, there is a facility for users can evaluate the posted review after the review is posted. The user can provide achieve to symbolize if this review is helpful, or write comments for the reviews. To review spam, we manually build a review spam corpus. In this work we used different products of mobile and camera. The Dataset consist of 600 product reviews which are crawled from amazon.com.

<table>
<thead>
<tr>
<th>Product Name</th>
<th>Review</th>
<th>Spam</th>
<th>Legitimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sony</td>
<td>115</td>
<td>55</td>
<td>60</td>
</tr>
<tr>
<td>Nikon</td>
<td>100</td>
<td>39</td>
<td>61</td>
</tr>
<tr>
<td>Acer</td>
<td>120</td>
<td>50</td>
<td>70</td>
</tr>
<tr>
<td>Samsung</td>
<td>150</td>
<td>80</td>
<td>70</td>
</tr>
<tr>
<td>Konica</td>
<td>115</td>
<td>75</td>
<td>40</td>
</tr>
</tbody>
</table>

VI. EXPERIMENT

We used supervised learning model to build review spam or legitimate detectors based on the product features. These models or classifier forecast definite class labels. In a product review, to find whether a review is spam or legitimate, a model or classifier is constructed to envisage class labels, such as “Spam” or “Legitimate”. Firstly supervised learning model builds the classifier by analyzing or “learning from” a training set made up of database tuples and their associated class labels. In the second step, the model is used for classification [9]. We used 10-fold cross validation to calculate performance of our system. We use support for value of the patterns and build a feature vector for each sample. Subsequently we applied support vector machine as classifier. We conduct the experiments using Weka [10] and SVM light tools to train Naïve Bayes and Support Vector Machine classifiers respectively. The data has been described in Section 5. We divided the data set into training set and test set and conducts 10-fold cross-validation: the data set is randomly split into ten folds, where nine folds are selected for training and the tenth fold is selected for test. We calculate experimental results using standard Information Retrieval (IR) metrics Precision, Recall and F-score. Where TP indicates true positive FP indicates false positive and FN indicates false negatives.
Table 2
Result of different camera model dataset

<table>
<thead>
<tr>
<th>Product Name</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-score</th>
<th>Accuracy NB-FM</th>
<th>Accuracy NB-Accuracy</th>
<th>Accuracy SVM-FM</th>
<th>Accuracy SVM-Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sony</td>
<td>41</td>
<td>26</td>
<td>31</td>
<td>81.50</td>
<td>57.81</td>
<td>71.15</td>
<td>91.46</td>
<td>89.32</td>
<td>93.76</td>
<td>94.78</td>
</tr>
<tr>
<td>Nikon</td>
<td>60</td>
<td>28</td>
<td>76</td>
<td>74.83</td>
<td>42.97</td>
<td>73.16</td>
<td>85.02</td>
<td>86.79</td>
<td>82.34</td>
<td>86.71</td>
</tr>
<tr>
<td>Acer</td>
<td>48</td>
<td>29</td>
<td>67</td>
<td>66.67</td>
<td>41.12</td>
<td>64.71</td>
<td>89.85</td>
<td>89.91</td>
<td>91.78</td>
<td>94.71</td>
</tr>
<tr>
<td>Samsung</td>
<td>36</td>
<td>31</td>
<td>22</td>
<td>71.43</td>
<td>64.00</td>
<td>74.29</td>
<td>89.90</td>
<td>91.71</td>
<td>91.79</td>
<td>91.89</td>
</tr>
<tr>
<td>Konica</td>
<td>26</td>
<td>16</td>
<td>17</td>
<td>78.20</td>
<td>48.94</td>
<td>69.16</td>
<td>88.14</td>
<td>89.76</td>
<td>88.78</td>
<td>94.89</td>
</tr>
</tbody>
</table>

Figure II: A comparisons of Precision, Recall, F-score and Accuracy values of different camera

VII. RESULTS & DISCUSSION

Figure III: A comparison of Naive Bayes and Support vector Machine.

VIII. CONCLUSION AND FUTURE WORK

In this paper, we represent a method to identify product review Spam or Legitimate. We applied SVM and Naive Bayes classifier on different dataset of camera and its performance was evaluated manually.

We use SVM\textsuperscript{light} and Weka tools for conducting the experiments. Our results shows that our proposed method is effective in compared to similar method. In future, we try improving our accuracy with same corpus using another method.

REFERENCES


