National University of Singapore
School of Computing

Hotel Review Spammer Detection and Behaviour Analysis

Ning Wei Wei (A0084106M), Prassanthi Muralidharan (A0123403Y), Raghavendhra Balaraman(A0123443R), Tran Minh Thy(A0074353H),

Zhang You (A0102049U)

Department of Information System, National University of Singapore,

IS4240 Project Assignment, Business Intelligence System

Submitted 15/04/2015

***Abstract* - Online reviews provide valuable information about products and services to consumers. Since they can promote or harm the brand of a product or service, buying or selling fake reviews would be a profitable business and a big threat. Previous attempts for spammer detection used reviewer's behaviour, text similarity, rating pattern or response time of reviewing. However, in reality, there are other kinds of spammers who can imitate behaviours of genuine reviewers, and thus, cannot be detected by the available techniques. In this paper, we analyse common behaviours of spammers to come out with unexpected rules that helps us build the opinion spam detection model.**

Keywords - Opinion spam detection, hotel opinion, spam model

---

1. Introduction

1.1 Problems

There are many websites for different industries which allow customers to write reviews or feedback based on their experience in utilising the product or service. There are industries in which products or services can only be evaluated by customer's experiences. Hotel industry is one such industry. Although the room size and equipment can be announced in form of figures or images, the hygiene, environment and services, etc. are very hard to represent. Customers tend to trust in other customers' past experiences. That is why many users take advantage of the existing reviews to make decisions on purchasing a product or availing some services.

With the increasing popularity of those review websites, many other businesses discovered the potentials in monetary gain through opinion spams. Those opinion spams are often inappropriate or fraudulent reviews, which are trying to attain self-promotion or bring other competitors down. According to current review website ranking system, hotels or restaurants with higher rating or more positive reviews often have higher ranking. Higher ranking attracts higher traffic and higher rating influences users in choosing them. Thus, it is intuitive for businesses to make the profile better and try to make other competitors' profiles look worse, so that the businesses themselves are able to occupy higher rankings. Opinion spams are then created, when the businesses try to make themselves rank better.

Spams do have very bad influence on customer experience or customer impressions. Imagine a low-quality hotel has a very high rating and successfully attracts many customers to stay in it whereas a competitive hotel receiving a low rating. The original purpose of the review websites are distorted, ending up in people choosing an unwanted product not as expected. The reputation of a review website itself will be affected. The review website industry is an industry heavily depending on strong customer network. Once the customer trust is lost, the website will also be very hard to sustain.

As a result, it is very important that the website have such spam detection tool to identify the spams, remove them and present purely actual customer reviews to other customers. In this paper, we will study a dataset of hotel reviews retrieved from Tripadvisor.com.sg and propose a method to identify opinion spams in hotel industries. (Trip Advisor Webpage , n.d.) Relevant spammer behaviours will also be discussed.

1.2 Challenges

1.2.1 Detecting fake reviews is hard

Unlike factual information, opinions and sentiments have an important characteristic, namely, they are subjective, which makes the detection of fake reviews difficult and uncertain.

First of all, there are several types of fake reviews. The most obvious type of fake reviews includes reviews containing link spam and commercial advertisement. Link spam is spam on hyperlinks, which hardly exist in reviews. Although advertising links are more common in other forms of social media, they are relatively easy to detect. However, dealing with fake reviews which are opinion spam, containing no link or commercial advertisement is the most difficult challenge for spam detection model. It is very hard, if not impossible, to recognize fake opinions even by manually reading them. In fact, in the extreme case, it is logically impossible to recognize spam by simply reading it. For example, one reviewer can write a truthful review for a good hotel and post it as a fake review for a bad hotel in order to promote it.

Secondly, there are two types of spammers which are individual spammers and group spammers. Individual spammers are persons who do not work with anyone, and simply write fake reviews her/himself using a single user name. Group spammers are a group of persons who work together to write reviews, and they usually vote higher rate for their own reviews. Groups' spammers can be one person who registers multiple user-names. Group spammers are really difficult to be detected.

Based on types of fake reviews and types of spammers, there are some requirements to find training data for the spam detection model:

- ✓ Star rating, user-id
- ✓ Time when a review was posted, and time taken to write/post the review
- ✓ Host IP address and MAC address
- ✓ Geo-location of the reviewer
- ✓ Sequence of clicks at the review site

Time taken to write a review is very important to detect spam reviews. Usually, spammers would copy and paste a fake review on a hotel's review websites, which logically takes them a shorter time to write than honest reviewers. Host IP address can help the model to detect whether a group of user-ids is registered by a single person. However, time and host IP address are difficult to use web crawler to get data from TripAdvisor website (Trip Advisor Webpage , n.d.).

1.2.2 Previous research and studies

A number of studies in the past have focused on traditional spam detection in e-mail and on the web. However, only recently there are studies examined the opinion spam. Jindal and Liu performed some of the first studies of this nature. (Jindal, 2008) They focused on three types of disruptive opinion spam, including spam containing advertisements and other non-related text. While these types of spam may be distracting where they are easily detectable by human readers.

On the other hand, the focus of our paper is detecting opinion spam, written with the specific intent of misleading customers which may be difficult for humans to detect. Opinion spam detection provides an unusual scenario in the assessment of human-created data, since machine-based methods have been shown to outperform human judges.

In this paper, we are interested in hotel reviews. Even if we can borrow some ideas from previous studies, their clues are not sufficient enough to define hotel review spammers. For example, although it is not usual that one reviewer can post several reviews on several hotels in Singapore or review several times about one particular hotel but it is quite normal because he or she can have experiences on different hotels or have several experiences on one hotel. Furthermore, one person has the same writing style on review writing, it may be quite usual to have similar style of reviews in words for one or more hotels. One hotel can be famous for its good services, but based on different perspectives of different customers, one review can give the opposite opinions to the rest of the crowd, it is not sufficient to prove that the customer is a spammer. Hence, there is a need to look for a more sophisticated and complementary framework.

2. Approaches

In order to build an effective spam detection model, we need to first find out the scenarios, where spammers try to create spams on the reviewing websites. Those spams or detective reviews are regarded as unexpected cases among all other true reviews. Next, we need to conclude the possible unexpectedness rules that may detect the spams. We apply all the unexpected rules into experiments and build a relevant model. After we have built the model, we collect spam data, which was identified manually, and mix them with true reviews. We use the model we build to detect the spams in the mixed data. The higher rate we can detect, the better spam detection tool it is.

2.1 Opinion Spam Incentives and Spam Scenarios

As mentioned above, users usually make decisions to purchase new products or services which have higher rate or more positive reviews. Based on this psychology, businesses might recruit spammers to write fake positive reviews to make themselves look better or write negative reviews for their competitors. The following are the detailed scenarios of opinion spams:

2.1.1 Self-promotion

Self-promotion aims to improve businesses' own profiles. For this incentive, businesses usually try to pump in as many positive review as possible on their profiles. The most efficient way for them is to use duplicate positive reviews. Simply uploading many same positive reviews on a brand's profile can easily promote its ranking in a very short period. More advanced, businesses can recruit different people pretending to be normal users to write positive reviews, which mimic the truthful reviews, called 'deceptive spams'.

2.1.2 Striking Other Competitors

Striking other competitors is actually a more insidious way to promote itself. By bombarding negative reviews on other competitors to drop their overall ratings and ranking, the hotel can improve its ranking in the review websites. Other than that, the business can also upload negative reviews with positive ratings for competitors. In this way, their competitors might not pay attention to these negative reviews, since the overall ratings will not be affected. However, after users view their profiles and go through the reviews, those negative reviews can destroy users' impressions on the competitors. Users will have higher chances to switch to the business, which creates the spams.

2.1.2 Irrelevant Product Promotion

Due to the high traffic in review websites, some advertisers try to upload advertisements on other products or services, directing users to clicking the link on it. Those kind of spammers usually spread identical advertisements as reviews across thousands of products or services being reviewed. They can also keep posting same advertisements under same products or services, trying to enhance their brand awareness and click-through rate.

2.2 Unexpectedness Rules and Relevant Detection Methodology

Usually, honest users have patterned reviewing behaviours on the review websites. Firstly, a user generally only gives one reviews to a hotel. Secondly, the review will only either be positive or negative. Thirdly, the ratings and the detailed reviews should be consistent, where positive reviews go with higher ratings and negative reviews go with lower ratings. Since spammers create fake spams, they often do things different from normal users. Thus, compared to normal users, their behaviours or reviews uploaded are deemed as unexpected cases. We have come out some unexpected rules to identify possible spams as following:

2.2.1 Contrasting Reviews and Ratings

Most of the spam detection mechanisms either take advantage of the ratings or the content of the reviews themselves. Usually, a review with 1 star out of 5, perceives a poor rating/satisfaction by the reviewer whereas a 5 star rating perceives high rating/satisfaction for a particular hotel. Nevertheless, spam detections mechanism that considers only ratings of the reviewer as a main indicator always suffers from the following shortcomings:

- Firstly, the rating used against each reviews does not completely reflect the sentiment of the reviewer who posted the review.
- Secondly, spammers nowadays make an educated guess and are quite aware of the fact that, most spam detection techniques consider either ratings alone or the review content alone as a main indicator.

This results in spammers posting negative reviews but with high rating (5 out of 5 stars) and positive reviews with low rating. This produces the noisy (spam) data.

No.	Review	Rating	Polarity
1	"**MBS is one of the worst hotels I have ever stayed at.** Not in my life have I been treated **so poorly** as a guest. The front desk was very **unaccommodating** when I asked for a smoke free room when they had made an error in my reservation. There was no bellhop available for some strange reason so I had to move all my luggage to the elevator and down a long hallway to my room by myself. If it wasn't already a bad stay, I ordered room service and it took over an hour and a half to be delivered. If they didn't have air conditioning in the room, I would say just about everything about this stay was **completely miserable**. If you are traveling to Singapore for any kind of business, I hope you decide not to choose this hotel. I was quite surprised, I like Singapore as a city but this stay definitely made my trip quite a negative experience."	4	Negative
2	"Pan Pacific in Singapore was **awesome**. The room was very **clean** and the hotel staff was very **professional**. One of the features I liked, was that in my room the internet access was wire and wireless, considering my laptop is not wireless, it help me out alot. Food was very good, quality was great. There was also a flat screen in my room...awesome. The hotel itself is locaated in the middle of alot of resturants with fin dinning. I also enjoyed the gym very much. Overall, I enjoyed myself, and I will stay again at the Pan Pacific when I return to Singapore"	1	Positive

Table 1. Examples of reviews in contrasting reviews and ratings case

From the above table, we can clearly see that the review contents and their corresponding rating are obviously contradicting. Thus, we understand that the sentiment of the review is equally important to the ratings of the review. In addition, we take advantage of these contradicting review contents and the rating score to classify those reviews as spam.

In this project, we try to implement the sentiment analysis techniques for each review through R-script. The relationship between the sentiment analysis score and the rating score is identified to discriminate Spam and No-Spam reviews. Our methodology in identifying spams from the given data set involves following three steps:

- ***Step 1:*** The rating scores of a particular hotel are collected and the average of the hotel is computed. Say, if the average rating for that hotel is 3.5, we consider all reviews whose rating score is 1, 2 or 5. These reviews are now fed into a sentiment analysis model performed in the next step.

- ***Step 2:*** The content of the reviews are now fed into R-script which performs sentiment analysis. In this step, the data set itself is divided into two sets namely,
    1. Training Data Set
    2. Test Data set

  With the training data set, feature-vector analysis is performed to build a term-document matrix. The accuracy of the training model are then identified using Weka. This training model is now used to predict the sentiments of the test data set. To ensure more correctness and accuracy, the prediction mechanism of the test data-set with the training data model is performed using two classifiers namely,
    1. Support Vector Machine Classifier
    2. Naive-Bayes Classifier

  The output from this step involves classifying whether a review is either 'Positive' or 'Negative'

Step 3: In this step, the rating score and the polarity of the review are compared against each other to arrive at a conclusion that whether a review is a 'spam' or 'not spam'. All those reviews that have a contradicting pair (polarity, rating) are classified to be spam and the rest are classified as 'not spam'.

Review - Polarity	Rating	Inference
Positive	1	SPAM
Negative	2	Not-Spam
Positive	5	Not-Spam
Negative	5	SPAM

Table 2. Examples of spam classification in Contrasting reviews and ratings case

2.2.2 Duplicate Reviews

According to many researches on human behaviour, the expectation of the number of reviews uploaded by one user for one hotel is only one and the reviews for different hotels should be different and relevant to the individual hotels. It is unexpected that one reviewer give multiple same reviews under one hotel or across hotels. Vice versa, multiple users upload identical reviews can also be deemed as unexpected. This kind of unexpectedness rule can be called 'Support Unexpectedness". The methodology in analysing such spammer behaviour is discussed below, followed by the descriptions on 4 cases of duplicate reviews.

2.1.2.1 Methodology in analysing duplicate spams behaviour

Due to data constraints, we are not able to build a spam detection model as what we did for the above case. Instead, we will carry out processes as following:

1. Use Sequel to filter out the unexpected cases as possible spams out of the dataset, based on each duplicate reviews rule we set. In this case, we will filter reviews with same review summary instead of same detailed review to detect more other duplicate patterns.
2. Manually go through the filtered possible spams on duplicate review summary and subjectively identify highly suspicious actual spams together with their patterns.
3. Figure out the proportion of suspicious actual spams out of the total possible spams, to conclude whether the unexpected duplicate reviews are highly featured spams.

2.1.2.2 Duplicates from different users on the same hotel

The first unexpected case we have identified is the duplicated reviews from different users within the same hotel. In this case, we first filter out the possible spam review to set it as a new dataset by comparing usernames, review summaries and the hotel names. What we will get from the query results will be a list of reviews by different users within one hotel as shown in Table3.

Summary Review	Detailed Review	Username	Hotel Name
SR1	DR1	User A	Marina Bay Sands
SR2	DR2	User B	Marina Bay Sands

Table 3. Examples of reviews in duplicates from different users on same hotels

After we get the possible spams, we manually look through the data to identify the spams. This method is used to identify the possible spam reviews written by same person who has multiple IDs.

2.1.2.3 Duplicates from same users on different hotels

It is also possible for some users to give same reviews on different hotels. In this case, we will compare the usernames, review summaries and the hotel names to filter out the same review summaries on different hotels given by same user. These reviews are possible to be spams because it is highly likely that the spammers will just copy and paste the reviews to different hotels in order to do advertising or promote certain hotel by posting negative reviews on the rest as shown in Table 4.

Summary Review	Username	Hotel Name
Same	Evangelos M	Marina Bay Sands
Same	Evangelos M	Fullerton Hotel
Same	Evangelos M	Raffles Hotel

Table 4. Examples of reviews in duplicates from same users on different hotels

2.1.2.4 Duplicates from same users on the same hotel

Same user can post same reviews on the same hotels. In this case, we will compare the usernames, review summaries and the hotel names, in order to figure out cases with all the 3 fields being the same. Examples are illustrated in the Table 5 below:

Review Summary	Username	Hotel Name
Same	Evangelos M	Marina Bay Sands
Same	Evangelos M	Marina Bay Sands
Same	Evangelos M	Marina Bay Sands

Table 5. Examples of spam reviews in duplicates from same users on same hotels

It is noticeable that there might be cases, where reviewers mistakenly submit the reviews twice or even more times, due to website design defects. That is why it is very important to manually look through the possible spams. If the submission time of the duplicate reviews is immediate with one followed by another, it is probably due to double submissions but not spams.

2.1.2.5 Duplicates from different users on different hotels

Duplicates can also be posted by different users on different hotels. We will just filter out same review summaries under distinct users and distinct hotels. Spammers can create different user accounts for posting reviews. They can post duplicate negative reviews to different competitors. This can be also seen for advertisements spread across hotels as shown in Table6.

Review Summary	Username	Hotel Name
Same	Evangelos M	Marina Bay Sands
Same	Nug_and_yub	Fullerton Hotel
Same	BaileyC_21	Raffles Hotel

Table 6. Examples of reviews in duplicates from different users on different hotels

2.2.3 Attribute Unexpectedness - Biased polarity of reviews by the same user

Another unexpected rule is biased polarity of reviews by a single user. According to a research “Finding unusual review patterns using unexpected rules” done by Jindal, Bing Liu and Lim, if a reviewer wrote all positive reviews on one hotel but all negative reviews on competing hotels, that reviewer may be a spammer (Nitin Jidal). A reviewer may find one hotel is better than the rest, but they will not rate too negatively for the rest.

Rule 1: Reviewer id 1, hotel 1 -> positive (Rate 5-star)

Rule 2: Reviewer id 2, hotel 2, 3, 4, 5 -> negative (rate 1 star)

Rule 2 is unexpected due to rule 1.

User id	Hotel Name	Review	Review Type
Reviewer 1	Hotel A	Good Hotel. Liked a lot !	Positive
Reviewer 1	Hotel B	Great hotel to spend your vacation	Positive
Reviewer 1	Hotel C	Fantastic Hotel. Loved the ambience!	Positive
Reviewer 1	Hotel D	Not good at all	Negative
Reviewer 1	Hotel D	Not even worth a penny. pls dont go to this hotel	Negative

Table 7. Examples of biased polarity of reviews by the same user.

From the above table, it is very obvious that the Reviewer 1 is completely biased to one particular hotel. We can also infer Reviewer 1 to be a potential spammer as his negativity towards Hotel D seems artificial with no supportive statements conveying as to why he/she didn't like Hotel D.

2.2.5 Other Unexpectedness

There are other unexpected cases, suggested by B. Liu (2012), which can also be possible spams:

- Confidential Unexpectedness

  It is unexpected that if a user give one or several positive reviews, while other users generally give negative reviews and low ratings.

- Attribute Distribution Unexpectedness

  If most of the positive or negative reviews under same hotel are contributed by one reviewer, although there are other reviewers giving both positive and negative reviews, the reviewer is likely to be a spammer.

- Submission Time

  If a reviewer use 5 seconds to write a review with 100 words, it is likely that the review is a spam. It is probable that the reviewer have already had the reviews and just copy and paste on it. Nevertheless, for normal users, people usually write the comments online after they log into the system.

However, due to data constraints, we are not able to conduct experimental studies for the three cases. For example, we cannot get the duration between the reviewer entering the web page and uploading the review.

3. Data collection and preparation

The main source of data for this project was from **Tripadvisor.com**, an American travel related company which provides user-generated travel related reviews. The entire process of data collection for this project underwent the following stages

Data Collection: During this stage, a python script using BeautifulSoup module was developed to parse the user reviews from **tripadvisor.com**. The parsed reviews are loaded into a .csv file

Data Cleaning: This stage involves removing unwanted information from the data set. The unwanted information includes all incomplete, inaccurate and irrelevant parsed reviews. This stage ensures that the dataset is consistent with all other data in the dataset.

Data Reduction: In this phase, the dataset undergoes feature selection, where only required fields/attributes that are helpful in our analysis alone are retained and the rest of the attributes are ignored

Data Loading: The consistent data is now loaded into a SQLite database for further query processing and analysis of the project.

After the data preparation, the below tables highlights the number of review count collected against each hotels

Hotel Name	**Review Count**
Marina Bay Sands	8528
Pan Pacific Singapore	3441
Village Hotel Changi by Far East Hospitality	1496
Oasia Hotel Singapore by Far East Hospitality	1324
PARKROYAL on Pickering	999
Moon 23 Hotel	331
Lloyd's Inn	203
Total Data Set	16322

Table 8. Dataset

In addition, the below table illustrates the attributes along with their description used in the data set.

Attribute Name	**Description**	**Data Type**
hotelname	The name of the hotel	STRING
reviewid	The url from where the hotel details are scrapped	INTEGER
summaryreview	Caption provided by the reviewers for each review(something like a summary)	STRING
detailedreview	The detailed review of his opinions about the hotel	STRING
reviewername	Name of individual who posted the review	STRING
totalreviewersreview	Total no. of reviews posted by the reviewer.	INTEGER
overallrating	Overall rating (out of 5)	INTEGER
sleeprating	Sleep Quality Rating (out of 5)	INTEGER

locationrating	Location Rating (out of 5)	INTEGER
roomrating	Room Quality Rating (out of 5)	INTEGER
servicerating	Service Rating (out of 5)	INTEGER
valuerating	Value Rating (out of 5)	INTEGER
cleanlinessrating	Cleanliness Rating (out of 5)	INTEGER

Table 9. Attribute Description of the dataset

3.1 Dataset - Contrasting Reviews and Ratings:

The data sets that are used for the identifying the contrasting review - rating spam are shown below. In the train dataset two additional attributes namely, polarity of review (Positive/Negative) and type of review (Spam/ Not Spam) are added. These two attributes are pre-classified to help build the training model

Model	Dataset Size
Training Dataset	400 reviews
Test Dataset	2142 reviews

Table 10. Training and Test Dataset

3.2 Dataset - Duplicate Reviews and Attribute Unexpectedness - Biased polarity of reviews by the same user

In order to identify the spam reviews in Duplicate Reviews Approach and Attribute Unexpectedness approach, the entire dataset of 16142 records has been used.

4. Experimental Results

We used the statistical package R (http://www.r-project.org/) to perform queries to do classification of our training data and the spam reviews are detected manually.

4.1 Contrasting Reviews and Ratings

In order to analyse the results, we built a model on training dataset using Support Vector Machine (SVM) and Naive bayes Classifier using R and Weka to perform sentiment analysis. This would help in identifying the sentiment (Positive/Negative) of a review. We then compare the output score of each review with its corresponding rating. The following table highlights the some important facts

	Dataset	Naïve Bayes Accuracy	SVM Classifier Accuracy
TRAINING	400	92.75	95.25

Table 11. Accuracy of training dataset

The test data set is now run against the training model to obtain the below results. The following are some of the assumptions made during the course of this project

- In order to reduce the run time of R-code, the test data set is reduced from 16322 to 2144.
- We consider reviews with rating 1,v2 as poor rating and 4,5 as high rated for the hotel MBS.

```
Time taken to build model: 1.59 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         371               92.75   %
Incorrectly Classified Instances        29                7.25   %
Kappa statistic                          0.8147
Mean absolute error                      0.0729
Root mean squared error                  0.2689
Relative absolute error                 19.4047 %
Root relative squared error             62.0953 %
Coverage of cases (0.95 level)          92.75   %
Mean rel. region size (0.95 level)      50.125  %
Total Number of Instances              400

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.930    0.080    0.972      0.930    0.951      0.818    0.941     0.972     Positive
                 0.920    0.070    0.814      0.920    0.864      0.818    0.940     0.779     Negative
Weighted Avg.    0.928    0.078    0.933      0.928    0.929      0.818    0.941     0.924

=== Confusion Matrix ===

   a   b   <-- classified as
 279  21 |   a = Positive
   8  92 |   b = Negative
```

Figure 1. Accuracy of training model using Naive Bayes Classifier

```
Time taken to build model: 1.31 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         381               95.25   %
Incorrectly Classified Instances        19                4.75   %
Kappa statistic                          0.8667
Mean absolute error                      0.0475
Root mean squared error                  0.2179
Relative absolute error                 12.6434 %
Root relative squared error             50.332  %
Coverage of cases (0.95 level)          95.25   %
Mean rel. region size (0.95 level)      50      %
Total Number of Instances              400

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.993    0.170    0.946      0.993    0.969      0.872    0.912     0.945     Positive
                 0.830    0.007    0.976      0.830    0.897      0.872    0.912     0.853     Negative
Weighted Avg.    0.953    0.129    0.954      0.953    0.951      0.872    0.912     0.922

=== Confusion Matrix ===

   a   b   <-- classified as
 298   2 |   a = Positive
  17  83 |   b = Negative
```

Figure 2. Accuracy of training model using SVM Classifier

Step 1: With this assumption we consider all the reviews with ratings 1,2,4,5 from the above dataset

Total Dataset : 2144	
Ratings	Count
1,2	204
4,5	1719

Table 12. Count of review whose rating is 1,2,4,5

Step 2: Now 1923 records are considered for the sentimental analysis performed against of training model and the following results are observed.

Total Dataset : 1923	
Positive	214
Negative	1709

Table 13. Count of Positive and Negative Reviews from Sentiment Analysis

Step 3: After classifying the reviews based on the polarity, the resultant data set is now compared against the rating score and the following behaviour was observed.

	1	2	4	5
Positive	8	8	73	126
Negative	693	827	75	113

Table 14. Ratings vs Polarity Count

From the above table, we conclude the following.

Total Dataset : 1923	
Spam	204
Not-Spam	1719

Table 15. Spam vs Not-Spam count

4.2 Duplicate review detection

In summary, we have identified 6896 reviews with duplicate summary out of 16142 reviews in total, which occupies 42.72%. Within all the duplicate reviews, only 0.74% can be concluded as actual spam. Duplicate reviews from different users under same hotels (Case 1) appeared most, followed by different users under different hotels (Case 4). Same users, different hotels (Case 2) and same users, same hotels (Case 3) are much fewer.

PROCESS	POSSIBLE SPAM	ACTUAL SPAM	ACTUAL SPAM %	TOTAL SPAM
CASE 1	4701	39	0.83	76.47
CASE 2	12	1	8.33	1.96
CASE 3	12	0	0	0
CASE 4	2171	11	0.51	21.57
TOTAL	6896	51	0.74	

Figure 3. Summary results of duplicate review

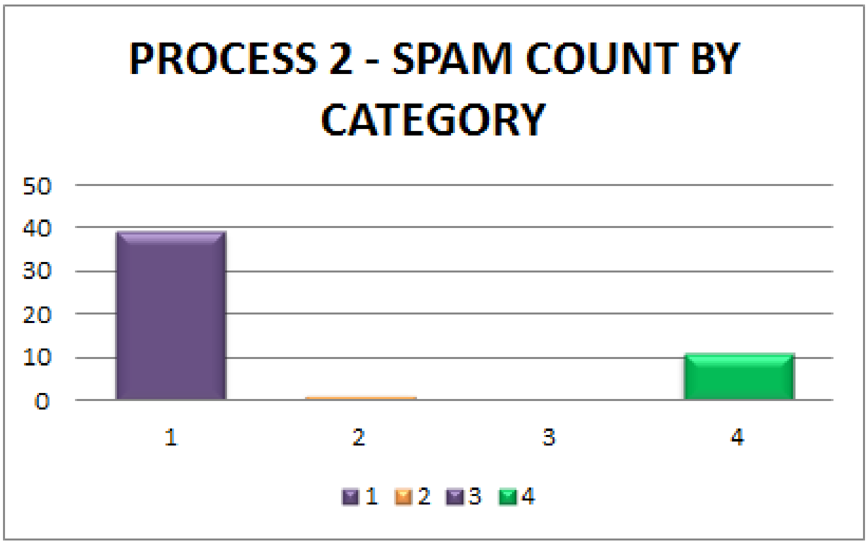


Figure 4. Total spam identified in each case

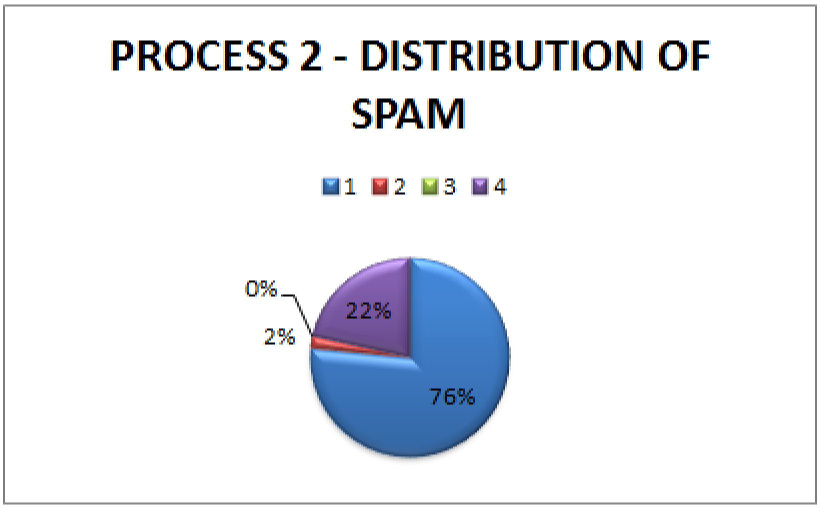


Figure 5. Spam distribution in each case

4.2.1 Different Users, Same Hotel

After we get the list of reviews by different users with same hotel. We manually look into the data to find the possible reviews that are actually written by same person. One possible method is to identify the reviews that contain the words with same spelling errors. These reviews are possible written by same person because it is highly likely that same person may spell a word wrongly in the same way. For example, if there are 3 reviews written by different users that contain the word "especially" spelt with "expecially", we may define them as spams as this particular mistake may probably made by one person who has created multiple accounts. In addition, the reviews contain same unusual expressions such as "very very very nice" by different user IDs are also be considered as spams.

SUMMARY REVIEW	DETAILED REVIEW	REVIEWER	RATING
Amazing pool and views	We chose to stay at the MBS because of the pool. And the pool was absolutely amazing and very worth the stay! The hotel as such was good but quite expensive, expecially the breakfast buff. The service we got by the pool staff was excellent and we really enjoyed the roof top bar (Ku De Ta) as well. I can't really say too much about the location because we chose to spend our full three days just hanging by the pool and enjoying the view! The gym was by far the best hotel gym I have ever been to - 5 stars!	Karolina F	4
Excellent	We only stayed 3 days at Marina Bay Sands but wished we had stayed for more. The hotel is just stunning. The club rooms and ammenities are outstanding and would recommend to anyone staying at Marina Bay Sands to stay in one of these rooms. The view and rooms are amazing. All the staff treated us with so much respect and nothing was a problem. The pool, what can we say it was amazing, expecially at night. Definitely do not miss this hotel in your trip to Singapore.	Moderne01	5
MARINA BAY SANDS	We stayed at the Marina Bay Sands and thoroughly enjoyed the entire experience. The view from the infiniti pool is one of the best in the city, and the service was excellent. We used the airport shuttle, which was convenient, and the check-in and check-out processes were efficient. One thing to be aware of is that this hotel accommodates thousands of guests nightly, and it's a tourist attraction in its own right. That means that both the public areas and the areas restricted to hotel guests are very crowded. Don't expect to have the pool to yourself or to find a quiet corner to read a book. Nonetheless, if you come with the right set of expectations, you will have a great time here.	cynthia3212	4
Stunning view	The city view room is simply incredible. Rooms are very spacious. We really like the separate bath and shower area. The Infiniti pool is amazing. My son really enjoyed it. We did not dine at any of the restaurants or shop as they were overpriced which was expected. The location was good as the Bayfront MRT line is directly underneath the hotel. Hotel service was excellent. We look forward to coming back sometime in the near future.	Alex T	5
very good	very nice this hotel. hotel infiniti pool has very very nice view!! very nice this hotel. hotel infiniti pool has very very nice view!!very nice this hotel. hotel infiniti pool has very very nice view!!	Youngah K	5
Fantastic	Perfect, fabulous hotel, it can be rated as [illegible] worth to stay some nights!!! Very if you wanna spend some money with anxiety- try casino))))"pleasant staff and everything is ok with the food. By the way	Viktor Z	5

Figure 6. Possible spams in Different Users, Same Hotel

In 16142 data we have, there are 4701 reviews on Marina Bay Sands from different users. And with these selected reviews, we are able to identify 39 reviews as spams based to the methods that mentioned above, which accounts for 0.83% of the 4701 reviews. Then we use this ratio to estimate the total spams in this case would be 76 reviews out of 16142.

4.2.2 Same user, different hotels

There are some users who are hired by a hotel to post reviews (most likely negative reviews) on its competitor hotels. These kind of spammers generally copy paste the reviews on different hotels that would give a negative impact about the hotels amongst the readers. In this case, we have filter out the 16142 data to get only 12 reviews on different hotels that written by same users, which we can see that it is quite a small portion for the duplicates cases, means that users seldom copy and paste their summary reviews to different hotels. In these 12 reviews, we identify only 1 spam manual reading, which consists of 8.3% of the selected data.

HOTELNAME	SUMMARY REVIEW	DETAILED REVIEW	REVIEWER	RATING
Marina Bay Sands	Another Great Stay	This is our 4th time to this property, and every stay we love it more, yes I have read the reviews about the crowded pool , lol	DeV67	5
Marina Bay Sands	Another Great Stay	Second visit to MBS after again attending the Singapore Grand Prix, this place never fails to impress check in was a little lo	DeV67	5
Marina Bay Sands	Unforgetable Experience	I booked MBS to experience the skypool, and the nearby Gardens by the Bay. I knew before from the reviews that this was not a hotel to look forwards to, but I thought may be those negative reviews were overstated. I decided to give it a try and booked the Club Room a garden view, because I like to stay with my two kids in the same room. Check in was prompt because one staff suddenly asked us to go over to counter and checked in for us but he didnt explain why we could jump	Gill W	1
Marina Bay Sands	Unforgetable Experience	I booked MBS to experience the skypool, and the nearby Gardens by the Bay. I knew before from the reviews that this was n	Gill W	1
Pan Pacific Singapor	Great value for your money	It's been renovated recently & this is my second stay at this hotel this year. Their staff are always friendly & welcoming. Loc	Hadi A	5
Pan Pacific Singapor	Great value for your money	It's been renovated recently & this is my second stay at this hotel this year. Their staff are always friendly & welcoming. Loc	Hadi A	5
Pan Pacific Singapor	Beautiful Hotel	Just home from our holiday in Singapore. Our stay at the hotel was really lovely. Unfortunately 2 of our first 3 nights there w	Jacqueline R	4
Marina Bay Sands	My experience at Marina Bay Sand:	Overall a splendid experience. Opening the Bayfront MTR was an added advantage. Travel was made so easy. Need not get wet in the rain. The restaurants and the casinos were awesome. Would like to come here again. I have already stayed here 4 times and would like to go there more often. If they have a better lobby would serve better.	sivacar	5
Marina Bay Sands	Lucky to have chosen this great hot	Excellent . Overall I enjoyed my stay. This is a beautiful hotel inside and outside.The breakfast spread was good. Even the vegetarians were taken good care.Casino lovers simply love it. The swimming pool on top made it very exciting.Now with the MTR station just close by makes travelling very easy instead of waiting for taxis. As you know getting taxis is not easy in Singapore especially during the change of duty hours.	sivacar	4
Lloyd's Inn	Lucky to have chosen this great hot	Fantastic stay! The hotel was beautifully decorated and immaculately clean.This is a beautiful hotel inside and outside. The rooms were very comfortable and clean. The hotel made my experience of visiting this city even more wonderful.	sivacar	4

Figure 7. Possible spams in Same User, Different Hotels

4.2.3 Same users, same hotel

We are able to get 12 reviews out of 16142 reviews in total as duplicate review summaries from same users and same hotels. It occupies 0.17% of all the duplicate reviews (6898 in total). According to the percentage, it is not a common case among all the duplicate cases.

After we manually look into the data, it is concluded that among the 12 possible spams from same users on same hotels, **none** of them can be concluded to be suspicious actual spams. We can see that spammers usually will not use same user account to post opinion spams on the same hotels. According to the cases illustrated in Figure 5 below, there are actually some reviews with exactly same details. However, the description in the reviews use a lot of subjectives to describe the feelings and experiences. The experiences are very detailed and unique. It should be a truthful review, which was accidentally posted twice. As for the other two reviews with same review summary, it is obvious that the reviewer gives reviews every time he check in. It should not be spam either.

Figure 7. Examples of reviews in dataset for duplicates from same users on same hotels

4.2.4 Different users, different hotels

Based on queries results, there are 2171 reviews with duplicate review summaries out of 16142 reviews, which is 31.48% of all the duplicate review summaries cases. It is surprising that many different users actually use the same review summary title in the reviews.

When we look into the details of the reviews, we find that 11 out of the 2171 reviews (0.51%) can be suspicious spams. We can see that among the 11 reviews with same review summary, many of them use the equipment in the hotels as the subjects in the sentences. They are mainly describing how good the services are, instead of describing users' own experience with a lot of 'I' and 'We'. It is also interesting that there are money appeared in the review, as indicated in the cases below in Figure 6.

Due to same writing styles with descriptions of the services and equipment in the hotel, the spammers can be actually one person creating several users to help different hotels promote themselves. It is a very interesting discovery that different hotels might actually recruit same person to assist on self-promotion. The scenarios of irrelevant product advertisements are not detected, which might be due to no market on the hotel review websites or the dataset is not big enough.

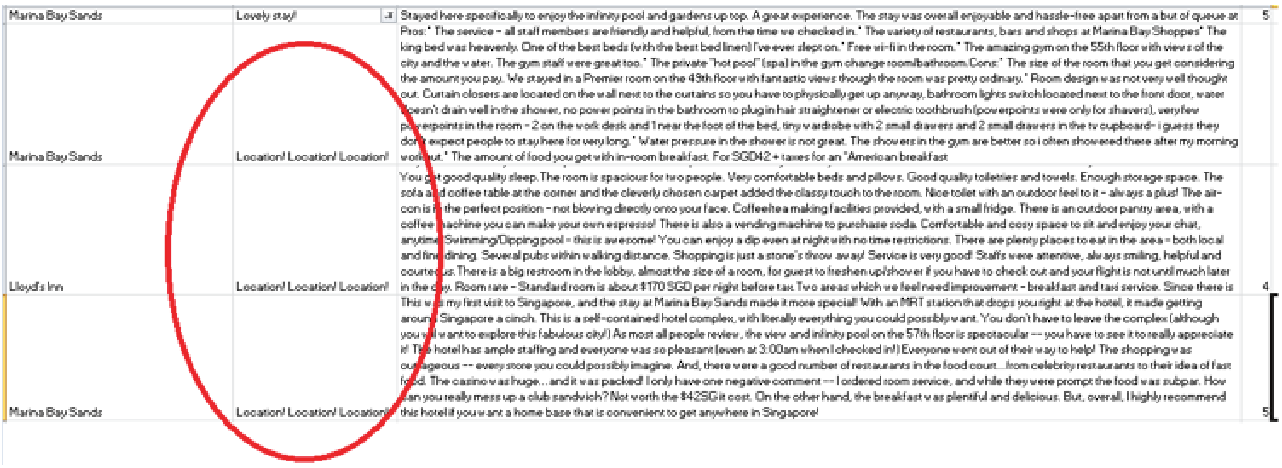


Figure 8. Examples of reviews in dataset for duplicates from different users on different hotels

4.3 Biased polarity of reviews by the same user

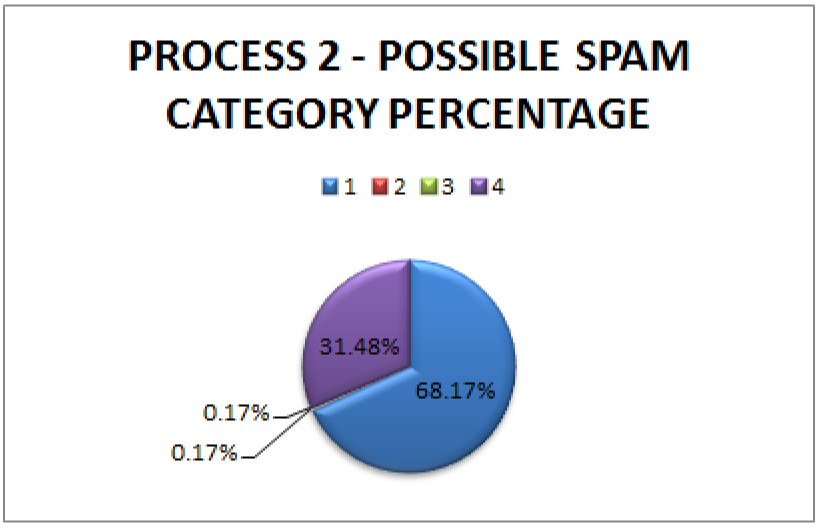
From the experiments performed on the dataset containing 16000 reviews, we try to find reviewers who write positive reviews for one hotel, while the rest (four other hotels) receive negative ratings.

Expected results: a data set contains reviews, written by some users, which are positive for one hotel (more than 2 ratings from one user) but too negative for the rest. *After we run this queries, from 16000 reviews, we are able to output 8 reviews written by 2 reviewers which are suspicious to be spams.*

5. Discussions and Conclusions

After setting up our expected rules, we have done several experimental studies and spam behaviours analysis. We have found out certain patterns in spammer behaviours which are discussed as below:

1. We are successful to build a spam detection model for detecting the reviews with contrasting contents and ratings. It is reported that 92.75% of spams with this pattern can be detected. It indicates that the expected rule of contrasting contents and ratings is significant. Many spammers love to use this method to hide their negative reviews covered with high ratings or bring down competitors' ratings covered with positive reviews. It seems a preferred method for spammers to strike on competitors in order to self-promote themselves.
2. It is surprising that 6896 out of 16142 reviews (42.72%) are detected as reviews with duplicate review summaries. However, only 51 out of 6896 (0.74%) are concluded as suspicious actual spams. Those spams mainly have a lot descriptive words to promote the hotels or same spelling mistakes or sentences appeared multiple times. The low ratio might be due to same normal users might prefer use same summary. It might also attribute to the limited dataset, which does not cover many featured spams.



3. There does exist a few spammers giving negative reviews to other hotels but positive reviews to only one hotel. 8 out of 16142 reviews (0.049%) belong to this attribute unexpectedness. Some might argue that there are users who actually only have positive experience with one hotel. When we look into the dataset, it is obvious that those reviewers with extreme bias have intentions to strike on other competitors in the contents.

5.1 Report's limitations

By reading this report so far, it is not hard to see that sentiment analysis for opinion is very challenging technically. Although the report set unexpected rules from natural behaviour spammers to do the spam opinion detection model, it is required to evaluate each review from the technical output to find which one is the actual review based on our several manual assessment methods. Hence, the manual process to detect actual spam reviews is subjective. The model we built is not expected to detect spammers who behaves naturally as an honest users and not follows our unexpected rules.

5.2 Future direction

In the later stage of our research on hotel review spam detection, we will find more relevant spam dataset that can be used to build spam detection model to test the rest of cases of spam reviews. Therefore, we will continue to work on the algorithms that we have proposed while not implemented due to time and technical restriction in this study, trying to build sentiment analysis and spam detection model for them.

In addition, we have discovered that other than the unexpected rules we have raised, there are many other suspicious actual spams with normal behaviour but suspicious contents. For example, not much individual experience is shared but a lot description of the equipment. Same spelling mistakes or even same structured sentences are shown in reviews under different hotels. All of those are 'deceptive opinion spam', according to (M.Ott, 2011). (2011). We would like to study on algorithms to find out all these opinion spams with dishonesty. In a nutshell, a well-defined spam detection system is what we are looking for in the future.

References

Jindal, L. (2008). Sentiment Analysis and Opinion Mining. 102 - 122.

M.Ott, Y. C. (June , 2011). *Finding Deceptive Opinion Spam by Any Stretch of the Imagination.* Retrieved from http://myleott.com/op_spamACL2011.pdf

Need someone to write positive reviews about our company. (n.d.). Retrieved from https://www.freelancer.com/jobs/Forum-Posting-Reviews/Need-someone-write-post-positive/

Nitin Jidal, B. L.-P. (n.d.). *Finding Unusual Review Patterns Using Unexpected Rules.* Retrieved from http://www.cs.uic.edu/~liub/publications/CIKM-final-unexpected.pdf

Trip Advisor Webpage . (n.d.). Retrieved from http://www.tripadvisor.com.sg/