

FAKE REVIEW PREVENTION FROM A PRODUCT REVIEW USING QUICK RESPONSE CODE

P. JEYA PRISKILA¹

²*Professor: Prof. Dr. A. MALARVIZHI, M.C.A., M.Phil., Ph.D., Assistant Professor, P.G. and Research Department of Computer Science, Pudukkottai- 622 001, Tamil Nadu, India.*

Abstract - The rapid growth of the Internet influenced many of our daily activities. One of the very rapid growth area is e-commerce. Generally, e-commerce provide facility for customers to write reviews related with its service. The existence of these reviews can be used as a source of information. For example, companies can use it to make design decisions of their products or services, while potential customers can use it to decide either to buy or to use a product.

Unfortunately, the importance of the review is misused by certain parties who tried to create fake reviews, both aimed at raising the popularity or to discredit the product. This research aims to detect fake reviews for a product by using the text and rating property from a review. In short, the proposed QR Code system will not save the honesty value of a review, the trustiness value of the reviewers and the reliability value of a product.

The honesty value of a review will be measured by scanning the QR code image printed in the product delivery package. Since each product are embedded with individual QR code image, only the real purchased customer can give the feedback will prevent the fake reviews

Key Words: QR, Fake Review.

1. INTRODUCTION

With in the display situation, clients are more subordinate on making choices to purchase items either one commerce destinations or off l in retail stores. Since these surveys are diversion changers for victory or disappointment in deals of a item, surveys are being controlled for positive or negative suppositions. Controlled surveys can moreover be alluded to as fake fraudulent surveys or conclusion spam or untruthful surveys.

In today's computerized world tricky conclusion spam has end end up a danger to both clients and companies. Recognizing these fake surveys is an critical and troublesome errand. These tricky commentators are regularly paid to type in these surveys. As a result, it

could be a herculean assignment for an standard client to distinguish false surveys from veritable ones, by looking at each audit. There have been genuine charges almost multi-national companies that are reveling in slandering competitor's items within the same division.

Quick Response Code

QR code (abbreviated from **Quick Response Code**) is the trademark for a type of matrix barcode (or two-dimensional barcode) first designed for the automotive industry in Japan. A barcode is a machine-readable optical label that contains information about the item to which it is attached. A QR code uses four standardized encoding modes (numeric, alphanumeric, byte / binary, and kanji) to efficiently store data; extensions may also be used.

The QR Code system has become popular outside the automotive industry due to its fast readability and greater storage capacity compared to standard UPC barcodes. Applications include product tracking, item identification, time tracking, document management, general marketing, and much more.

A QR code consists of black modules (square dots) arranged in a square grid on a white background, which can be read by an imaging device (such as a camera) and processed using Reed-Solomon error correction until the image can be appropriately interpreted. The required data are then extracted from patterns present in both horizontal and vertical components of the image.

RUN LENGTH CODING

Run-length encoding (RLE) is a very simple form of lossless data compression in which runs of data (that is, sequences in which the same data value occurs in many consecutive data elements) are stored as a single data value and count, rather than as the original run. This is most useful on data that contains many such runs. Consider, for example, simple graphic images such as icons, line drawings, Conway's Game of Life, and animations. It is not useful with files that don't have many runs as it could greatly increase the file size.

RLE may also be used to refer to an early graphics file format supported by CompuServe for compressing black and white images, but was widely supplanted by their later Graphics Interchange Format. RLE also refers to a little-used image format in Windows 3.x, with the extension rle, which is a Run Length Encoded Bitmap, used to compress the Windows 3.x startup screen.

Run-length encoding schemes were employed in the transmission of television signals as far back as 1967. It is particularly well suited to palette-based bitmapped images such as computer icons, and was a popular image compression method on early online services such as CompuServe before the advent of more sophisticated formats such as GIF. It does not work well at all on continuous-tone images such as photographs, although JPEG uses it quite effectively on the coefficients that remain after transforming and quantizing image blocks.

Common formats for run-length encoded data include True vision TGA, Pack Bits, PCX and ILBM. The ITU also describes a standard to encode run-length- colour for fax machines, known as T.45. The standard, which is combined with other techniques into Modified Huffman coding, [citation needed] is relatively efficient because most faxed documents are generally white space, with occasional interruptions of black.

ONLINE REVIEWS

As of late e-commerce mammoth amazon.com had conceded that it had fake audits on its location and sued three websites denouncing them of giving fake audits, stipulating that they halt the practice.

Fakespot.com has taken a lead in recognizing fake audits of items recorded on amazon.com and its auxiliary ecommerce locales by giving rate of fake surveys and grade.

Reviews and evaluations can straightforwardly impact client buy choices. They are considerable to the victory of businesses. Whereas positive audits with great evaluations can give monetary advancements, negative surveys can hurt the notoriety and cause financial loss. Fake surveys and evaluations can contaminate a trade. It can influence how others see or buy a item or benefit. So it is basic to decide fake/ false surveys.

DETECTING FAKE REVIEWS

Conventional strategies of information investigation have long been utilized to distinguish fake/fraudulent surveys. Early information examination strategies were arranged toward

extricating quantitative and factual information characteristics. A few of these procedures encourage valuable information elucidations and can offer assistance to urge superior bits of knowledge into the method behind information. To go beyond a conventional framework, a information investigation framework has got to be prepared with impressive sum of foundation information, and be able to perform thinking errands including that information. In exertion to meet this objective analysts have turned to the areas of machine learning and counterfeit intelligence.

A survey can be classified as either fake or veritable either by utilizing administered and/or unsupervised learning methods. These strategies look for reviewer's profile, audit information and action of the commentator on the Web for the most part utilizing treats by creating client profiles. Utilizing either administered or unsupervised strategy gives us as it were an sign of extortion probability.

SPAM DETECTION METHODS

Spamming Discovery In later a long time, Web or e-mail spam have been significantly examined. For case, a study is given on Web spam location. Mail spam discovery is additionally considered. Web journal spam or arrange spam are moreover expectation examined. For the audit spams, Fei et al. considered the behavior of fake audits and given conceivable spam design.

SUPERVISED SPAM DETECTION

Supervised Spam Detection were among the primary who proposed arrangements on the discovery of fake or beguiling surveys. In common, fake audit location can be classified into two categories, directed and unsupervised. Liu given a nitty gritty study on spam detection. The method proposed by Liu et al. may be a administered survey discovery learning strategy in which fake surveys are respected as a set of copy or close rehash surveys and other surveys are non-fake reviews.

A FRAMEWORK HAS BEEN PROPOSED TO DETECT FAKE PRODUCT REVIEWS OR SPAM REVIEWS:

In this work 1. Opinion Mining (Anusha Sinha et. al, 2018) has been proposed to leverage the peeling properties from Autoencoder and random forest, a stochastic decision tree model is implemented to guide the global parameter learning process (Manqing Dong et.al, 2018).

In this work Salma Farooq and Hilal Ahmad Khanday, 2016 has been proposed to solve the problem of review spam detection and the performance of different approaches for classification and detection of review spam. To detect spam and fake reviews, and filter out reviews with expletives, vulgar and curse words, by incorporating sentiment analysis has proposed.

2. Existing System

Initially there was an issue of alluding valuable survey among numerous. This issue was unraveled in existing framework by implies of conclusion mining. Client should studied all conceivable surveys for selecting that item.

Conclusion mining strategies recognizes the extremity of each sentence in all audits given to item, and after that computes the whole of all comparative items utilizing the standard Capacities.

Existing framework has given one highlight that anybody can donate input almost any item. This causes impediment for this system. The individual from challenging e- shopping site can deliver fake criticism to the first site.

This can be done for accomplishing ubiquity in web showcasing. There's no extraordinary usefulness portray in existing framework for maintaining a strategic distance from this fake looking into. So existing framework is able to mine negative and positive criticism but it falls flat to distinguish genuine one and fake one survey in list of item survey.

3. Proposed System

In this proposed system, QR code scanning is used to avoid fake review.

Each and every product will be packed by tagging QR code embedded with secret pin . Review page will be accessed only when the scanning QR code matches.

Then only customer can write their reviews on particular products.

TO CLASSIFY MOVIE REVIEWS INTO GROUPS OF POSITIVE OR NEGATIVE POLARITY BY USING MACHINE LEARNING ALGORITHMS:

Elshrif Elmurngi 2018 has proposed a method to recognizing the untruthful reviews that are given by the users which is having distinct semantic content based on sentiment analysis as the reviews of movies

has proposed (Rashmi Gomatesh Adike and Vivekanand Reddy, 2016).

- A system used to classify tweets into different groups as spam and non spam tweets has proposed (Abha Tewari and Smita Jangale ,2016).
- To detect fake reviews for a product by using the text and rating property from a review has proposed (Eka Dyar Wahyuni and Arif Djunaidy, 2016).

In this work (Gurneet Kaur and Abhinash Singla, 2016) has proposed an empirical study of efficacy of classifying product review by semantic meaning. Bhanu Prakash Battula et.al, 2015 a studies of different approaches for identifying manipulated reviews and proposes a new approach.

TO IDENTIFY THOSE MANIPULATED REVIEWS USING DECISION TREE:

In this work (Rajashree S. Jadhav and Deipali V. Gore , 2014) has proposed to identifying whether a review is fake or truthful one by Naïve Bayes Classifier, Logistic regression and Support Vector Machines has been proposed.

Qing-yun dai et.al, 2018 in this work has proposed a decision tree model is used to classify a record is to find a path that from root to leaf by measuring the attributes test, and the attribute on the leaf is classification result has been proposed. A new decision tree algorithm IQ Tree for classification problem has been proposed to extract a kind of "structure" from a sample of objects.

Bhaskar N. Patel et.al, 2012 an implementation of decision tree algorithm and for comparative study and to analysis the performance has been proposed. There is a lot of work that has been done on the fake reviews, fake news and fake social edia Ids. Spotting groups of fake reviewers finds the fake reviews and an individual fake reviewers group who are working on writing the fake reviews on the e-commerce Websites for promoting or demote the seller's product.

DEVELOPERS USE THE "FREQUENT ITEMSET MINING (FIM)" METHOD TO SEARCH THE FAKE REVIEWERS GROUPS:

In this work Pooja Sharma and Rupali Bhartiya, 2012 has proposed the system uses the behavioral model and relational models to find the relationship among the fake reviewers groups which also called "spammer

groups". Before this work, no labeled dataset is available so to check their method they produce a labeled dataset using expert human judges. This system uses novel relation based model called "GSRank" which finds the fake reviewers and relationship between the spammer groups.

1) In this technique, the set of item I is the set of reviewer Id and each transaction is the set of Id who gives the review to a particular product and then the system uses the FIM method to find the groups who gives review of the multiple different products together.

2) The system of detecting fakes reviews of Yelp analyzes that what Yelp Is doing by filtering and analyzing these filtered reviews.

3) There are two main approaches to filter reviews: supervised and unsupervised and in terms of features there are also two main types: behavioral features and linguistic features.

4) Behavioral eatures perform well for their system rather than the linguistics features. There starting point is the work of the Ott which uses the Amazon Mechanical Turk (AMT) to find the Turkers who are writing the fake reviews for the hotels.

FAKE REVIEWS DETECTION ON MOVIE REVIEWS THROUGH SENTIMENTAL ANALYSIS USING SUPERVISED LEARNING TECHNIQUES:

In this work has proposed with the aim to classify the movie reviews into positive and negative polarity by using the machine learning algorithm (Kolli Shivagangadhar et.al, 2019). The proposed system applies the sentimental analysis and text classification methods to detect the fake reviews related to the movies. The proposed work find out that Yelp may be using sentimental analysis but there is not hundred percent surety that they are using this technique to filter their reviews.

Rayana and Akoglu 2020 Fake reviews detection for the Yelp is worked with the intention to filter the fake reviews from the original reviews as this is becoming the need of the hour.

The proposed system classifier takes the reviews text and other information and produces the output whether the reviews are reliable or not.

The data set which is used in this project is taken for the Yelp.com which is firstly used by the Rayana and Akoglu.

They use 16282 reviews and split these into 0.7 training set, 0.2 dev set and 0.1 test set. Extracting predictive features from the reviews is the most challenging part of the project. Basically they extract two types of features: review-centric feature and reviewer-centric features. Firstly they count the percentages of each unigram and bigram tokens for fake and non-fake reviews. They then take out the top 100 unigrams and bigrams that have the most different percentages in fake and non-fake reviews. The second approach leads to the better performance because it processed all the unigrams and bigrams. They tested multiple algorithms of machine learning but by using the Neural Networks they achieved the highest accuracy of about 81.92%. This system is good in finding the fake reviews but still there is a need to improve the accuracy in filtering the reviews.

SPAM REVIEWS DETECTION BY USING TEMPORAL PATTERN DISCOVERY:

In this work has been proposed to observe the reviews related to the normal reviewers arrival pattern and fake reviewers arrival pattern and they observe that the normal reviewer arrival pattern is stable and uncorrelated to their rating pattern temporally.

FAKE PRODUCT REVIEW MONITORING AND REMOVAL METHODOLOGY

In this work Detecting fake reviews from product is very much important in this era. As there are two types of purchases in every e-commerce Website which are verified purchase and Non-verified purchase. Verified purchase means that the customer who was writing the review purchased the product from the online store and who did not receive the product at a great discount. To detect fake reviews, data is gathered from trip advisor, MTurk and Yelp. So that model can be trained in a best possible way. For the verified purchase, it is an obvious thing that the customer has bought that product for which he/she is giving the review. Now, in this situation, if the customer leaves a positive review for the product and gives the rating of 1 or 2 than it is cleared that it is a fake review. That is why system uses sentimental analysis for the verified purchase. For the word like 'good, excellent' etc. the sentiment is to be positive and for the 'bad, poor quality' etc. type words the sentiment is to be negative.

For the Non-verified purchase anyone can give the review of the product without having any type of purchase history on e-commerce Website. Now in this situation, it is possible that the person who is giving the positive review related to the product also gives the good rating to that product. But, here this possibility also falls that the person may be giving the positive review and rating to increase the rating of that product and he is a spammer. In this type of reviews the technique of sentimental analysis is not worked properly because by using sentimental analysis it is considered as a genuine review. So that is the reason the proposed system is using another technique in which Support Vector Machine (SVM) is a classifier.

4. Conclusion

In this ponder Web Analytics and Client Mapping conveys lightning quick web location log investigation creating valuable insights and reports for any measure of organization. Logs examination is the finest tool available for understanding our clients. We have to be anticipate how our clients discover our Web location and why they are searching for it. This data will deliver an thought approximately the attitude of our arriving guests. The venture strategy decides reliance on items and shopping propensities. Besides, estimate deals characterize the advancements of items and client profiles. Affiliation run the show mining was utilized as an approach for recognizing clients buying designs and as a result client profiles were decided.

Future Enhancements:

Segmentation (clustering) utilizing more point by point behavioral information and opportunity acknowledgment utilizing affiliation calculations inside the sections discovered.

Association of items and client cluster for cross-selling (offering modern items) and up-selling (offering more of what clients as of now buy).

We ought to discover out for modern, advanced ways of exploring in information created amid broad information profiling.

The dependable investigation of metadata is most critical for the choice prepare, we are working on approaches to produce, store, and investigate metadata of tall quality.

REFERENCES

BOOKS:

- [1]Jindal, N and Liu, B. (2008) Opinion spam and analysis[C]. In: Proceedings of the 1st International ACM Conference on Web Search and Data Mining, ACM: 219-230.
- [2]DANDIBHOTLA, T.S. (2018) Finding and classifying the deceptive spam reviews using LIWC dictionary variables and decision tree classifier[J]. International Journal of Engineering & Technology, 10(1): 308-315.
- [3]Gilbert, E., Karahalios, K. (2010) Understanding deja reviewers[C]. Proceedings of the 2010 ACM Conference on Computer Supported Cooperative work. ACM:225-228.
- [4]Heterogeneous Collective Classification (MHCC) and then extend it to Collective Positive and Unlabeled learning (CPU) has been proposed]. scoring methods to measure the degree of spam for each reviewer and apply them on an Amazon review dataset has been proposed .
- [5](Elshrif Elmurngi 2018) There are many research work published in the area of fake review identification. rominent machine learning techniques that have been proposed to solve the problem of review spam detection and the performance of different approaches for classification and detection of review spam (Michael Crawford et. al, 2015).