



International Journal of Intellectual Advancements and Research in Engineering Computations

Intelligent interface for fake review monitoring system

**Jeladaranaa. KC [17101031], Kamilah .A [17101035], Meera .S [17101051]
Mr.R.T. Dinesh Kumar, M.Tech, Assistant Professor**

Computer Science And Engineering, Vivekanandha College Of Engineering For Women
(Autonomous), Tiruchengode, Tamilnadu.

kcjela@gmail.com, kamalahamerjan19@gmail.com, meerasaravanandpi@gmail.com

ABSTRACT

As the style to save on-line is growing day through day and extra human beings are involved in shopping for the merchandise of their need from the on-line stores. This kind of buying does not take a lot of time of a customer. Customer goes to on line store, search the object of his/her want and region the order. But, the thing by means of which humans face subject in shopping for the products from on-line shop is the terrible first-rate of the product. Customer place the order solely by using searching at the ranking and by means of reading the opinions associated to the specific product. Such comments of different humans are the supply of pride for the new product buyer. Here, it can also be viable that the single negative review adjustments the perspective of the client now not to purchase that product. In this situation, it would possibly feasible that this one review is fake. So, in order to eliminate this kind of pretend evaluations and provide the customers with the unique evaluations and ranking related to the products, we proposed a Fake Product Review Monitoring and Removal System (FaRMS) which is an Intelligent Interface and takes the Uniform Resource Locator (URL) associated to merchandise of Amazon, Flipkart and Dares and analyzes the reviews, and affords the patron with the original rating. It is a special nice of the proposed system that it works with the three e-commerce Websites and no longer only analyze the evaluations in English however additionally the critiques written in Urdu and Roman Urdu. Previous work on faux critiques does not guide function to analyze the evaluations written in languages like Urdu and Roman Urdu and can't deal with the opinions of multiple e-commerce Websites. The proposed work achieved the accuracy of 87% in detecting pretend evaluations of written in English by using the usage of shrewd getting to know methods which is greater than the accuracy of the preceding systems.

INTRODUCTION

There are one-of-a-kind approaches to keep like you can purchase a particular aspect of your want via going to a shop or mall. In this fashion of purchasing the vendor offers you the comments of the product, you do no longer understand whether or not he/she is giving a pretend remark or original. Because, it is upon vendor honesty, how tons the vendor is proper in his/her phrases and you have to cautiously take a look at the product due to the fact you do now not have any different alternative in inspecting the product. If you do not pay interest in shopping for that product then it might also be

proved a waste for you. On the different hand, in modern times source of buying has been changed. You can purchase the merchandise from the online shops of one of a kind brands. Here, you have to area the order besides seeing and analyzing the unique product. You examine the critiques and purchase the product. Therefore, you are structured on the critiques about the product. These critiques may additionally be the authentic or fake. The patron needs to purchase a unique and dependable product, which is viable solely when you get the unique comments associated to that product. Research indicates that U.S. buyers spend \$6 billion in Black Friday sale 2018. Americans spend 36% of

Author for correspondence:

Computer Science And Engineering, Vivekanandha College Of Engineering For Women (Autonomous), Tiruchengode, Tamil Nadu.

the buying price range online. In 2017, E-commerce shops earned \$2.3 trillion in income and predicted to attain \$4.5 trillion with the aid of 2021. Today, nearly 12-24 million ecommerce shops are working round the world. Study determined that 61% of Amazon opinions that belongs to Electronics Category are fake. There are some web sites which are working to realize the pretend reviews. Fake spot is an on-line Website that detects faux evaluations the usage of suspicious patterns and reviewers activity. As in the system of shopping for the product from the online shops you have to examine all the evaluations one with the aid of one to test for the exceptional of that product and to get a properly satisfactory product. It is a very time eating process. Here this opportunity additionally falls that the critiques might also be pretend or original.

LITERATURE REVIEW

Fake Product Review Monitoring Using Opinion Mining

Product reviews play an important role in deciding the sale of a particular product on the e-commerce websites or applications like Flipkart, Amazon, Snap deal, etc. In this paper, we propose a framework to detect fake product review or spam reviews by using Opinion Mining. The Opinion mining is also known as Sentiment Analysis. In sentiment analysis, we try to figure out the opinion of a customer through a piece of text. We first take the review and check if the review is related to the specific product with the help of Decision tree. We use Spam dictionary to identify the spam words in the reviews. In Text Mining we apply several algorithms and on the basis of these algorithms we get the specific results.

Black Friday Purchases

The top selling products on Black Friday were laptops, with lots of people taking advantage of the savings to pick up a new machine. Other top sellers were games like God of War and Let's Go Pikachu, and children's toys like Fingerlings. This is a reminder that although Black Friday is increasingly thought of as a day for adults to pick up big ticket items like electronics and appliances, the market for kid's toys is still a significant part of sales.

Study Finds 61 Percent of Electronics Reviews on Amazon Are Fake

As the trend to shop online is increasing day by day and more people are interested in buying the products of their need from the online stores. This

type of shopping does not take a lot of time of a customer. Customer goes to online store, search the item of his/her need and place the order. But, the thing by which people face difficulty in buying the products from online store is the bad quality of the product. Customer place the order only by looking at the rating and by reading the reviews related to the particular product. Such comments of other people are the source of satisfaction for the new product buyer. Here, it may be possible that the single negative review changes the angle of the customer not to buy that product. In this situation, it might possible that this one review is fake. So, in order to remove this type of fake reviews and provide the users with the original reviews and rating related to the products, we proposed a Fake Product Review Monitoring and Removal System (FaRMS) which is an Interface and takes the Uniform Resource Locator (URL) related to products of Amazon, Flipkart and analyzes the reviews, and provides the customer with the original rating. The proposed work achieved the accuracy of 87% in detecting fake reviews of written in English by using intelligent learning techniques which is greater than the accuracy of the previous systems.

Urdu Sentiment Analysis

The entire world is transforming quickly under the present innovations. The Internet has become a basic requirement for everybody with the Web being utilized in every field. With the rapid increase in social network applications, people are using these platforms to voice their opinions with regard to daily issues. Gathering and analyzing peoples' reactions toward buying a product, public services, and so on are vital. Sentiment analysis (or opinion mining) is a common dialogue preparing task that aims to discover the sentiments behind opinions in texts on varying subjects. In recent years, researchers in the field of sentiment analysis have been concerned with analyzing opinions on different topics such as movies, commercial products, and daily societal issues. Twitter is an enormously popular microblog on which clients may voice their opinions. Opinion investigation of Twitter data is a field that has been given much attention over the last decade and involves dissecting "tweets" (comments) and the content of these expressions. As such, this paper explores the various sentiments analysis applied to Twitter data and their outcomes.

Spotting Fake Reviewer Groups in Consumer Reviews

Opinionated social media such as product reviews are now widely used by individuals and

organizations for their decision making. However, due to the reason of profit or fame, people try to game the system by opinion spamming (e.g., writing fake reviews) to promote or demote some target products. For reviews to reflect genuine user experiences and opinions, such spam reviews should be detected. Prior works on opinion spam focused on detecting fake reviews and individual fake reviewers. However, a fake reviewer group (a group of reviewers who work collaboratively to write fake reviews) is even more damaging as they can take total control of the sentiment on the target product due to its size. This paper studies spam detection in the collaborative setting, i.e., to discover fake reviewer groups. The proposed method first uses a frequent item set mining method to find a set of candidate groups. It then uses several behavioral models derived from the collusion phenomenon among fake reviewers and relation models based on the relationships among groups, individual reviewers, and products they reviewed to detect fake reviewer groups. Additionally, we also built a labelled dataset of fake reviewer groups. Although labelling individual fake reviews and reviewers is very hard, to our surprise labelling fake reviewer groups is much easier. We also note that the proposed technique departs from the traditional supervised learning approach for spam detection because of the inherent nature of our problem which makes the classic supervised learning approach less effective. Experimental results show that the proposed method outperforms multiple strong baselines including the state-of-the-art supervised classification, regression, and learning to rank algorithms.

Existing System

The device in which you can discover the authentic remarks and ranking associated to a precise product. Then, it is the supply of delight and reliability for you. In the proposed technique, the opinions associated to a product for which the URL is given are extracted. After it, the gadget finds the faux evaluations and in the end by means of examining these critiques device finds the authentic opinions of the product. Previous researches notice pretend opinions the use of one of a kind processes which includes identification address, opinion mining and sentiment analysis, machine-learning approach. There are many researches handy that notice the faux evaluations associated to English however no work is carried out so a long way that realize pretend opinions for Urdu (highly spoken language in Asia) and Roman Urdu. Therefore, we have proposed Fake Product Review Monitoring and Removal System (FaRMS) in which a purchaser

can get the fantastic viable object from the online shop in a quick time and with the authentic critiques related with that product. This gadget offers you the authentic phrases of human beings associated to the product with proper reviews. Some famous merchandise can get lots of evaluations at some massive service provider websites and FaRMS offers you the promising opinions through filtering pretend critiques and then you can determine whether or not you prefer to purchase or not.

Disadvantages

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.

Proposed System

A technique to ranks the product is worked to present a product ranking model that applies weights to product review factors to calculate a product ranking score. In this proposed system, the sentences that are not related with the quality of a product such as customer service or sentence related to the. In this paper the pre processing is done by Support Vector Machine (SVM). First of all, it removes the comments which neither is nor related with the quality of the product. Second stage describes the weights of the reviews based on the votes. The final stage calculates the overall ranking of the product. The ranking score is calculated by the relevance of the review with quality of the product, review content, and posting date of the review. They use 10-fold cross validation on the

training set. In the evaluation process they use two measures to quantify effectiveness of the ranking model which are as following: correlation between the ranking method and the Amazon' s rank and second is the Mean Average Precision (MAP), which is a very commonly used technique for evaluating ranking accuracy. As this system is finding the fake reviews by using the only two properties of the reviews but as per the future work describes in the paper more properties can be used to find out the fake reviews more accurately.

Spam reviews detection by using Temporal Pattern Discovery is 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. On the opposite side the spam attacks are usually bursty and either positively or negatively correlated with the rating pattern. The data set which they have taken is snapshot of a review Website on October 6, 2010. It includes 408469 reviews which are written by 343629 reviewers, which are written for 25034 stores of a Website. For each review they collect the following information like rating, postdate and whether it is a Spammer Review (SR) or not. In the evaluation process they select 53 stores each of which has more than 1000 reviews. Human evaluators make decision about the stores to be SR spam attack or not if two or more evaluators declared a store as SR spam attack than system considers the store a dishonest in its selling. Out of 53 stores 34 are suspicious one and the remaining are normal ones. Out of 34 stores 22 stores have at least two votes for being suspicious. The recall related to the system is 75.86% which shows that the system detects most of the stores having SR spam attack. The precision related to the proposed approach is 61.11%. This proposed system is good in terms of the training of

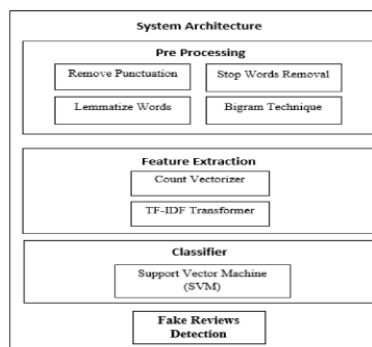
their model for finding the relation as the model is trained by using the large number of reviews contains in their dataset.

Advantages

The previous works detect fake reviews by using IP address, opinion mining, sentiment analysis, and some of them uses machine learning approaches. In some approaches dataset is very small and other uses few properties related to the reviews to find the fake reviews. In the proposed system the large dataset of English reviews is used to train the model. In this way, the system can find the hidden patterns in the reviews more efficiently. The accuracy which is achieved by using the proposed technique is greater than the accuracy of the previous systems in terms of the English reviews. This system is also worked to detect the fake reviews of Urdu and Roman Urdu reviews.

Fake review monitoring system focuses on detecting spam and fake reviews by using sentimental analysis removes the reviews which have curse and vulgar words. In the proposed system web crawler is used to scrapped the data on the Website. In the pre processing, the data is converted into the required format and then the fake reviews are removed from the mixture of original and spam reviews. Fake reviews are detected by the Fake Review Detector. Each review has to be passed from the classifier which calculates the review's sentiments score. Cosine similarity is used to measure this type of similarity. If the cosine value to be calculated is greater than 0.5 than the review is considered as the fake review. The developed system detected the 111 reviews to be faked out of 300 reviews. However, the data set which is used for training the model is very small which cannot find the suspicious patterns more accurately.

Farms Architecture



MODULE & MODULE DESCRIPTION

Module:

- Data set module
- Feature extraction module
- Training module
- Classification module
- Output module

Module Description

Data set module

A data set (or dataset) is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question. The data set lists values for each of the variables, such as height and weight of an object, for each member of the data set. Each value is known as a datum. Data sets can also consist of a collection of documents or files.

In the open data discipline, data set is the unit to measure the information released in a public open data repository. The European Open Data portal aggregates more than half a million data sets. In this field other definitions have been proposed, but currently there is not an official one. Some other issues (real-time data sources, non-relational data sets, etc.) increases the difficulty to reach a consensus about it.

Feature extraction module

Feature extraction is a process of dimensionality reduction by which an initial set of raw data is reduced to more manageable groups for processing. A characteristic of these large data sets is a large number of variables that require a lot of computing resources to process. Feature extraction is the name for methods that select and /or combine variables into features, effectively reducing the amount of data that must be processed, while still accurately and completely describing the original data set.

The process of feature extraction is useful when you need to reduce the number of resources needed for processing without losing important or relevant information. Feature extraction can also reduce the amount of redundant data for a given analysis. Also, the reduction of the data and the machine' s efforts in building variable combinations (features) facilitate the speed of learning and generalization steps in the machine learning process.

Training module

A training module is one structured section of a course. The content within a training module should be designed, and created, to support the learner's intake and retention of the information it contains. Grouping training modules together is used to create step-by-step learning.

Grouping training modules together is used to create step-by-step learning. Each module forms one part of an overall topic, enabling learners to gradually progress through a course, module by module, to reach their training goals. It' s a tactic that makes training delivered through a learning management system more digestible.

Classification module

Classification is a machine learning method that uses data to determine the category, type, or class of an item or row of data. For example, you can use classification to:

- Classify email filters as spam, junk, or good.
- Determine whether a patient's lab sample is cancerous.
- Categorize customers by their propensity to respond to a sales campaign.
- Identify sentiment as positive or negative.
- Classification tasks are frequently organized by whether a classification is binary (either A or B) or multiclass (multiple categories that can be predicted by using a single model).

RESULT AND SCREEN SHOTS

In signal communication within the output module, the external signal line must be isolated from the body. Optical-coupled photo couplers provide simple and reliable isolation. A transistor output photo coupler or an IC output photo coupler which has high-speed communication capability is recommended. In addition, it is essential to observe the safety standards established in each country. Toshiba Photo couplers have been approved by UL1577, VDE: EN60747-5-5, EN62368-1, and other organizations. The output stage also contributes to lower power dissipation through MOSFET of low on-resistance and higher output through the use of transistor arrays with high withstand voltage and large current.

Yelp Dataset: The proposed architecture uses the SVM classifier to train the model that consists of 147,440 rows.

	Predicted False	Predicted True
Actual False	13,542	1185
Actual True	2700	12,061

Table. Confusion Matrix of Yelp Reviews Dataset

Urdu Dataset: System uses the SVM classifier to train the model that consists of 400 rows.

	Predicted False	Predicted True
Actual False	25	8
Actual True	16	31

Table. Confusion Matrix of Urdu Reviews Dataset

Roman Urdu Dataset: The proposed architecture uses SVM classifier to train the model that consists of 400 rows.

	Predicted False	Predicted True
Actual False	29	15
Actual True	10	26

CONCLUSION

In this project, dataset is developed that contains Urdu and Roman Urdu reviews. It is difficult to detect fake reviews by yourself. So, n-gram approach is used to detect fake reviews for multiple languages. It is observed that the text categorization with SVM classifier is best approach for the detection of fake reviews. Now a days, as the technology is growing day by day and there are so many Websites and applications available in the online market by which seller can sell their products and, on that products, there are millions of reviews

available. There are some organizations posting fake reviews for the products of the seller in order to increase or decrease the rating of the products. Therefore, the system is proposed that detects the fake reviews in multiple languages including English, Urdu, and Roman Urdu, classify the reviews in genuine. It helps the user to get the products from Daraz, Flipkart and Amazon with the satisfaction of their mind and pay for the good quality product. As, there are a lot of e-commerce stores like AliExpress and Alibaba which have reviews of multiple languages. It would be great if the proposed system finds the way, to process and filter the reviews for other multiple languages.

REFERENCES

- [1] Sinha, N. Arora, S. Singh, M. Cheema, and A. Nazir, " Fake Product Review Monitoring Using Opinion Mining, 119(12), 13203-13209, 2018.
- [2] Torbet, Georgina. " U.S. Customers Spent over \$6 Billion on Black Friday Purchases." Digital Trends, Digital Trends, 25 Nov. 2018, www.digitaltrends.com/web/shopping-totals-black-friday/
- [3] Sterling, Greg. " Study Finds 61 Percent of Electronics Reviews on Amazon Are 'Fake'." Marketing Land, 19 Dec. 2018, marketingland.com/study-finds-61-percent-of-electronics-reviews-on-amazon-are-fake-254055.
- [4] K. Khan, W. Khan, A. Rehman, A. Khan, Asfandyar. Khan, A. Ullah Khan, B. Saqia, "Urdu Sentiment Analysis," (IJACSA) International Journal of Advanced Computer Science and Applications, 9(9), 2018.
- [5] Mukherjee, B. Liu, and N. Glance, "Spotting Fake Reviewer Groups in Consumer Reviews," 2012.
- [6] Mukherjee, V. Venkataraman, B. Liu, and N. Glance, " What Yelp Fake Review Filter Might Be Doing?," Aaai, 409-418, 2013
- [7] Z. Wang, Y. Zhang, and T. Qian, " Fake Review Detection on Yelp Dataset and features," pp.
- [8] S. Xie, G. Wang, S. Lin, and P. S. Yu, " Review spam detection via temporal pattern discovery," p. 823, 2012.

- [9] Paper, “ Mining millions of reviews: A technique to rank products based on importance of reviews Mining Millions of Reviews: A Technique to Rank Products Based on Importance of Reviews,” no. November, 2015.
- [10] V. K. Madhura N Hegde, Sanjeetha K Shetty, Sheikh Mohammed Anas, “ Fake product review monitoring,” Int. Res. J. Eng. Technol., 05(06), p. 4, 2018.
- [11] S. Rayana and L. Akoglu, “ Collective Opinion Spam Detection: Bridging Review Networks and Metadata,” Sigkdd, 985-994, 2015.