ONLINE REVIEW DETECTION SYSTEM – A SURVEY

Tejas Kale, Nikhil Khobragade, Aditya Gupta, Mahesh Biradar, Mr. Lalit Borase

Nutan Maharashtra Institute of Engineering and Technology

ABSTRACT

A lot of work is done on this particular topic which majorly deals with Sentimental Analysis part. Internet reviews are becoming a medium to judge a products quality, durability, life and many more. Right from the object like a pen to your pair of glasses or the electronic gadgets. To invest in something or not is the primary question one needs to deal with. The current systems are lacking with the accuracy, which needs to be appropriate as it matters. In this paper we are trying to categorize the users review based on different machine learning algorithms, where we apply them on the data set which is collected from Internet sites and also proposing a model to improve the accuracy compared to the existing system.

Keywords: *Sentimental Analysis, Supervised Classification, Review Detection*

INTRODUCTION

When customers want to make a decision about a service or a product, reviews have become the primary source of information. When customers take the initiative to book for a hotel, they read reviews on other people's opinions before booking a hotel. Customers, who have already used the hotel's services, judge the service they experienced. Depending on the previous responses, they decide whether or not to book a room based on the reviews and rating. If they do, they most likely received Real feedback from the reviews and continue to book the room. As a result, the hotel gets a business and one gets a shelter with good services. Since reviews are regarded as form of sharing genuine feedback on either positive or negative services. Today people have a big number of choices, even varieties have sub varieties. On what basis one should decide. As today E-Commerce is becoming people’s favorite choice, without knowing physically how something exists, people invest, just by having a look at the reviews, so by this we can get the depth of how today business can get shape. In this paper we are showing the amount of work done till now and proposed work we are going to do. In our proposed system we are using SVM(support vector machine) and NB (Naive Bayes) algorithm to improve the accuracy.

FIGURES AND TABLES

Detecting Fake Reviews Utilizing Semantic and Emotion Model.	*2016*	*Yuejun Li, Xiao Feng, Shuwu Zhang*	*NLP Algorithm*	*Spammer Detection; Semantic Model; Emotion Model*
A Survey on Fake Review Detection using Machine Learning Techniques	*2018*	*Nidhi A. Patel, Prof. Rakesh Pate*		*Sentiment Analysis, Opinion Spam, Machine learning*
Feature Analysis for Fake Review Detection through Supervised Classification	*2019*	Pradeep Kumar Tiwari, Rishi Gupta, Rohit Kumar Gupta	K Nearest Neighbour Logistic Regression	Review Spam Detection; Feature Learning; Document and Node Embeddings
Fake Reviews Detection: A Survey	*2021*	RAMI MOHAWESH SHUXIANG XU , SON N. TRAN, ROBERT OLLINGTON , MATTHEW SPRINGER , YASER JARARWEH , SUMBAL MAQSOOD1	C-LSTM, HIERARCHAL ATTENTION NETWORK (HAN), CONVOLUTIONAL HAN	Fake review, fake review detection, feature engineering,

DETECTING FAKE REVIEWS UTILIZING SEMANTIC AND EMOTION MODEL:

In this paper, Yuejun Li, Xiao Feng, Shuwu Zhang(2) have proposed the semantic analysis on the basis of review density, semantic and emotion related model. The review density includes category density, store density and time density. Maximum Review Density (MRD, F5): the maximum of number of reviews the reviewer posted in a day divides number of days that has at least one review. Active Time Frame (ATF, F7): Number of reviews reviewer posted divides number of Active Time Frame since reviewer registered in the website. The concept of emotion modeling is done by using the following formulae's:

$$\text{Index}_{\text{neutral}}(r) = \frac{\text{Num(neutralWord)}_r}{\text{wordNum}(r)} \quad \text{Index}_{\text{negative}}(r) = \frac{\text{Num(negativeWord)}_r}{\text{wordNum}(r)}$$

$$\text{Index}_{\text{positive}}(r) = \frac{\text{Num(positiveWord)}_r}{\text{wordNum}(r)}$$

The algorithms used by Yuejun Li, Xiao Feng, Shuwu Zhang are Naïve Bayes, SVM, and Decision Tree. Where they found that Naïve Bayes performs weak with the precision of 0.9, where as SVM and Decision Tree performs with the precision of 0.92, 0.93 respectively.

A SURVEY ON FAKE REVIEW DETECTION USING MACHINE LEARNING TECHNIQUES:

To detect the fake and real nature of reviews various kind of methods can be used, which are review centric approach, Reviewer centric approach and product centric approach. Various kind of features can be extracted from review data, reviewers' data and product data such as linguistic features, behavioral features, relational features, textual features. These kinds of features are used in many machine learning algorithms. In supervised learning techniques linguistic features containing parts of speech (POS) are used to classify data using classifiers like SVM (support vector machine), naïve bayes, decision tree, random forest, logistic regression etc. In semi-supervised learning techniques Positive Unlabeled (PU) learning technique is used to determine the real or fake nature of review using algorithms like k nearest neighbor, logistic regression, random forest, naïve bayes, SVM (support vector machine) etc. And in Unsupervised methods discussed in paper use Relational features to classify data into real or fake.

FEATURE ANALYSIS FOR FAKE REVIEW DETECTION THROUGH SUPERVISED CLASSIFICATION:

In this paper the model is based on the count vectorizer or a tf_idf matrix (i.e.) word tallies relatives to how often they are used in other articles in your dataset) can help. Since, this problem is a kind of text classification, implementing a naive Bayes classifier will be best as this is standard for textbased processing. However, classification algorithms such as Random Forest, SVM, and Logistic regression have also been used. Now the next step is to extract the most optimal features for count vectorizer or tf_idf vectorizer. After which feature extraction is done by encoding the words as integers or floating point values for use as input to a machine learning algorithm. The classifier is trained with a training data set. With the help of the trained data set, the testing and validation of review are performed. The output is shown as either the given review is fake or genuine or also a probabilistic approach is shown.

REFERENCES:

1. *Detecting Fake Reviews Utilizing Semantic and Emotion Model(2016 3rd International Conference on Information Science and Control Engineering) by Yuejun Li, Xiao Feng, Shuwu Zhang.*
2. *A Survey on Fake Review Detection using Machine Learning Techniques (2018 4th International Conference on Computing Communication and Automation (ICCCA)) by Nidhi A. Patel, Prof. Rakesh Pate*
3. *Feature Analysis for Fake Review detection using supervised learning(2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT) Manipal University Jaipur, Sep 28-29, 2019) by Pradeep Kumar Tiwari, Rishi Gupta, Rohit Kumar Gupta*
4. *Fake Reviews Detection: A Survey (IEEE volume 9 2021 Digital Object Identifier 10.1109/ACCESS.2021.3075573) by RAMI MOHAWESH 1 , SHUXIANG XU 1 , SON N. TRAN 1 , ROBERT OLLINGTON1 , MATTHEW SPRINGER 1 , YASER JARARWEH 2 , AND SUMBAL MAQSOOD1*