FAKE REVIEW DETECTION THROUGH DEEP LEARNING ENSEMBLE MODELS

Dr. C. Geetha[1], Lakkamaneni Chandana Manaswini[2], Mavillapalli Amrutha[2], Bapathi Aparna[2]

[1]Associate Professor, Department of Computer Science and Engineering

[2]Final Year Students, Department of Computer Science and Engineering

R.M.K. Engineering College, Kavaraipettai - 601206

Abstract

Online reviews about the acquisition of products or services provided became the most source of users' opinions. To acquire benefit or acclaim, as a rule spam audits are composed to advance or downgrade a couple of target items or administrations. This practice is known as review spamming. In the past few years, a spread of methods are suggested so as to unravel the difficulty of spam reviews. The effect of online audits on organizations has developed altogether during the most recent years, being critical to decide business achievement in a wide exhibit of areas, going from restaurants , hotels to e-commerce. Unfortunately, a few clients utilising dishonest intentions to upgrade their online standing by composing counterfeit surveys of their organisations or rivals. Previous research has addressed fake review detection during a number of domains, like product or business reviews in restaurants and hotels. The proposed work is detecting fake reviews based on the ensemble model of Convolutional Neural Network (CNN) models that has been evaluated in the Yelp restaurant domain.

Keywords

Term Frequency, Inverse Document Frequency, Natural language processing , Natural Language Toolkit, Convolutional Neural Network.

I. Introduction

Online consumer product reviews are playing an increasingly important role for patrons , constituting a new type of word of-mouth (WOM) information. Recent research shows that [3]52% of online consumers use the Internet to search for product information, while 24% of them use the Internet to browse products before making purchases. Online reviews have a robust impact on consumers' decision purchase in e-commerce, affecting the foremost relevant areas, like travel and accommodations, online retailers, and entertainment.

There are thousands of reviews online, [4]which makes it convenient for people to make decisions, but the amount of data makes it difficult to sort through. The real value of online reviews is in its content and the certainty that the reviewer indeed received products or services prior to writing the review. Promotion or demotion of the products and services is one among the most reasons for deceptive reviews. At times, to make better ratings for the venue, [5]hotel owners pay employees to fabricate false reviews. Alternatively, some reviewers write negative reviews for malicious reasons, wishing to distort the reputation of the business reviewed.

Yelp.com is one among the most important online review sites. [2]It uses a filtering algorithm to detect fake reviews. However, the algorithm is a trade secret. In this work, we collected reviews from yelp.com for 100 random hotels in the Charleston area. We labeled filtered reviews as real and unfiltered reviews as fake. We extracted part-of-speech features, trained and tested the data set, built a model and compared results to related work.

Besides, online surveys of a similar item can be found in products wellsprings of data, which can be grouped by the gatherings that have WOM data into internal WOMs , hosted by retailers (for example Amazon, Walmart, BestBuy, and so on) and outer ones, facilitated by autonomous item review suppliers (for example CNET, Yelp, TripAdvisor, Epinions, etc.). Nevertheless, only credible reviews have a significant impact on consumers' purchase decisions. Moreover, the product category significantly affects the credibility of WOMs. Consumer electronics product category is the most online reviewed, supported variety of things[6]. On the one hand, consumer electronics usually require a big investment, and therefore the more valuable and expensive an item is, the more it's researched. According to a study, consumer electronics are the merchandise most affected by online reviews, influencing the 24% of products acquired during this category, and being WOMs the second most influential source after search engines during this product category. On the other hand, consumers tend to research consumer electronics products because these products change very frequently, with new products and updates of existing ones. Thus, consumers frequently trust reviews to avoid making a wrong purchase decision. As a result, Horrigan et al. report that more than 50% of consumer electronics buyers tend to consult several WOMs before making a purchase decision.

Some studies show that retailer hosted online WOM influences enormously sales in low involvement products, [7] like books or CDs. However, consumers usually conduct a pre-sales research in high-involvement products, like consumer electronics. Thus, in consumer electronics, a retailer's internal WOM features a limited influence, while external WOM sources have a big impact on the retailer’s reputation and sales. Hence, consumer electronics are more sensible to the consequences of external WOMs, since they can't easily act on them.

Since both consumers and retailers become overwhelmed by the large number of obtainable opinions in WOM internal and external sources[8], automatic tongue processing and sentiment analysis techniques are frequently applied. a number of the foremost frequent application domains are review polarity classification, review summarization, competitive intelligence acquisition and reputation monitoring.

Given the importance of reviews for businesses and therefore the difficulty of obtaining an honest reputation on the web , several techniques are wont to improve online presence, including unethical ones. Fake reviews are one of the foremost popular unethical methods which are present on sites like Yelp or TripAdvisor. However, consistent with Jindal and Liu, not all fake reviews are equally harmful. Fake negative reviews on good quality products are really harmful for enterprises, and in conjunction with fake positive reviews on poor quality products, result also harmful for consumers. Counterfeit positive surveys on low quality items are likewise unsafe to contenders who offer normal or great quality items yet don't have various audits on them.

The goal of this text is analyzing the fake review problem within the buyer electronics field, more precisely studying [9]Yelp businesses from four of the foremost important cities of the USA. No prior research has been administered during this concrete field, being restaurants and hotels the foremost previously studied cases. We would like to prove that fake review detection problems in online consumer electronics retailers are often solved by machine learning means and to point out if the problem of achieving it depends on geographic location.

In order to realize this goal, we've followed a principled approach. Based on literature review and experimentation, a feature framework for fake review detection is proposed[10], which incorporates some contributions like the exploitation of the social perspective. This structure, the purported Fake Feature Framework (F3), assists with reworking and describes highlights for counterfeit survey determination. F3 considers data coming from both the client (individual profile, inspecting action, confiding in data, and social communications) and survey components (audit text), setting up a system with which to sort existing research.

In order to gauge the effectiveness of the features defined in F3, a dataset from the social Yelp in four different cities has been collected and a classification model has been developed and evaluated.

The other sections of this paper is organized as follows. Section II describes the detailed literature survey. Section III explains the methodology and section IV discussed the results. The section V gives the conclusion.

II. Literature Survey

The system used Feature extraction and sentiment analysis for the processing of knowledge so as to detect fake reviews[1]. The text representation model and sentiment analysis are accustomed to enrich the text features, and therefore the user’s abnormal comments are analyzed to represent the user’s behavior. The extracted data is quantized, and therefore the extracted features are divided into two feature sets consistent with attributes. The disadvantage is While sentiment analysis is useful , it isn't a whole replacement for reading survey responses. Often, there are useful nuances within the comments themselves.

Online surveys give extra item data to decrease vulnerability. Hence, consumers often believe online reviews to form purchase decisions[11]. However, an explosion of online reviews brings the matter of knowledge overload to individuals. Identifying reviews containing valuable information from large numbers of reviews becomes increasingly important to both consumers and corporations , especially for experience products, like attractions. A few online survey stages give a capacity to perusers to rate an audit as "accommodating" when it contains important data. Different from consumers, companies want to detect potential valuable reviews before they're rated [12]to avoid or promote their negative or positive influence, respectively.

People utilize online surveys to settle on choices about accessible items and administrations. In recent years, businesses and thus the research community have shown a superb amount of interest within the identification of faux online reviews[14].Applying exact calculations to recognize counterfeit online surveys can shield people from spam and falsehood. We accumulated separated and unfiltered online surveys for a few hotels from yelp.com. We extracted part-of-speech features from the info set, applied three classification models, and compared accuracy results to related works

[l]One of the most problems about opinion-sharing websites is that spammers can easily create hype[13] for a few particular products by writing spam reviews. These spam reviews may play a key role in increasing the price of the merchandise or service. As an example , if a customer wants to urge any product online, they typically attend the review section to know about other buyers' feedback. If the reviews are mostly positive, the user may buy, otherwise, they could not buy that specific product[15].This all shows that spam reviews became the foremost problem in online shopping which can cause a loss for both the customer and thus the manufacturer. The aim of the project is to acknowledge spamming or fake reviews.

III. Methodology

The proposed application should be able to identify fake or real reviews. Feature extraction model used is Glove Vectorizer. We used classification models ensemble type as combine CNN models

- ❖ Extract the feature using Glove Vectorizer
- ❖ Split train and test set
- ❖ Create CNN model 1 and Model 2
- ❖ Create trained model and given input for ensemble CNN
- ❖ Predict the review types
- ❖ Evaluate the performance of the algorithm

a. Algorithm

The proposed work is implemented in Python 3.6.4 with libraries tensor flow, keras, pandas, matplotlib and other mandatory libraries. We downloaded the dataset from yelp.com. The data downloaded contains train set and test set separately with four two classes of label namely fake and real. The train dataset is considered as a trained set and test dataset considered as test set. Deep learning algorithms are applied to the Convolutional Neural Network and Ensemble model.

The dataset is downloaded real time from yelp.com, the authentication keys are generated and keyword given "Restaurant" and location zip code is given to retrieve data through API from yelp.com

The dataset contains the following attributes

Table 1: Attributes and details of Dataset

Attribute	Details
ID	Restaurant Id
Name	Restaurant name
Address	Restaurant address
Phone	Phone number of the restaurant
Zip Code	Restaurant zip code
State	Restaurant State
Review count	Number of review counts
Rating	Rating for the restaurant
Author	Author Name
Date	Date of review
Review	Review content
Label	Labeled as REAL or FAKE

Fake review detection is done the taken dataset by applying feature extraction techniques

- ❖ Glove vectorizer model

Deep learning algorithm applied on the above extracted features

- ❖ CNN Static
- ❖ CNN Dynamic
- ❖ Ensemble model

The glove is an unsupervised learning for getting vector portrayals for words. Training is on accumulated worldwide word-word co-event measurements from a corpus, and the subsequent portrayals showcase fascinating direct foundations of the word vector space.

To quantitatively catch the subtlety important to recognize man from lady, a model should relate in excess of a solitary number to the word pair. A characteristic and straightforward contender for an amplified set of discriminative numbers is the vector distinction between the two-word vectors. The glove is planned with the goal that such vector contrasts catch, however much as could be expected the significance determined by the juxtaposition of two words.

In Fig 1 ,The above Glove features selection output as trained and test input given to the deep learning classification algorithm as input and arrives at the results. We used a convolutional 2F neural network available in keras for training and testing our model.

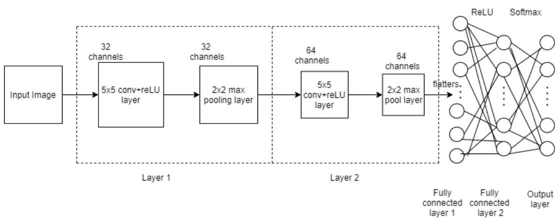

Fig 1: **Architecture of Conv2F**

Models in Keras are accessible in two structures – Sequential and by means of the Functional API. For most profound learning organizations, the Sequential model is likely. It permits to just stack consecutive layers (and even repetitive layers) of the organization so as to contribute to yield.

Add a 2D convolutional layer to handle the 2D info pictures. The principal contention passed to the Conv2D() layer work is that the quantity of yield channels – during this case we've 32 yield channels. The following information is the kernel_size, which for this situation we have decided to be a 5×5 moving window, trailed by the steps in the x and y bearings (1, 1). Then, the enactment work is an amended straight unit lastly we need to supply the model with the size of the contribution to the layer. Proclaiming the info shape is only expected of the essential layer – Keras is sufficiently sweet to sort out the components of the tensors moving through the model from that point. Add a 2D max pooling layer. We essentially determine the components of the pooling inside the x and y bearings – (2, 2) during this case, and thus the steps.

Next is to straighten the yield from these to enter our completely associated layers. The following two lines proclaim our completely associated layers – utilizing the Dense() layer in Keras, we determine the size – in accordance with our design, we indicate 1000 hubs, each actuated by a ReLU work. The second is our delicate max grouping, or yield layer, which is the size of the quantity of our classes.

In the preparation model, we need to indicate the misfortune capacity, or mention to the structure what kind of optimiser to utilize (for example angle plummet, Adam optimiser and so on)

Lass function of ordinary cross entropy for categorical class classification (keras.losses.categorical_crossentropy). We use the Adam optimizer (keras.optimizers.Adam). Finally, we will specify a metric which will be calculated once we run evaluate() on the model.

We first pass altogether our training data – during this case x_train and y_train. The subsequent argument is the batch size. During this case we are employing a batch size of 32. Next we pass the measure of training ages (2 during this case). The verbose banner, set to 1 here, determines on the off chance that you might want definite data being printed inside to reassure you about the advancement of the preparation.

In Fig2, a troupe model is utilized , we made two CNN models to blend and get a group one. Both CNN calculations are given the contribution of vectorized information inside the info layer and a learned model is made . The outfit method of Convolutional Neural Network is implied by giving contribution of the over two prepared models to encourage the enhanced outcomes

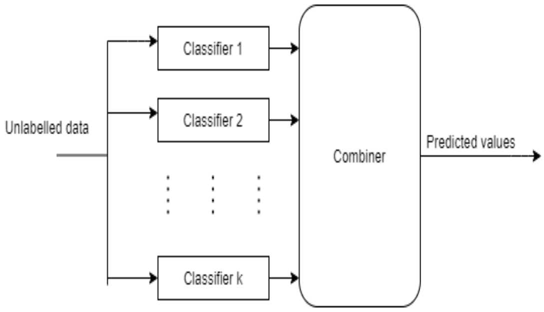

Fig2: CNN Ensemble model

b. Architecture

The Fig 3 represents the system architecture of the proposed system in which we represented all modules including data collection, pre-processing and applying algorithm and prediction modules.

- ❖ Data Collection: The data collection process involves the selection of quality data for analysis. Here we used yelp API and downloaded the dataset real time for our fake review research using machine learning implementation. The job of a knowledge analyst is to seek out ways and sources of collecting relevant and comprehensive data, interpreting it, and analyzing results with the assistance of statistical techniques.
- ❖ Information Pre-preparing: The point of preprocessing is to change over information into a structure that matches AI. Organized and clean information permits an information researcher to ask more exact outcomes from an applied AI model. The procedure incorporates information designing, cleaning, and testing.
- ❖ Dataset parting: A dataset utilized for AI ought to be parceled into three subsets — preparing, test, and approval sets. Preparing set. An information researcher utilizes a preparation set to prepare a model and characterize the ideal boundaries it needs to gain from information. A test set is needed for an assessment of the prepared model and its capacity for speculation
- ❖ After an information researcher has preprocessed the gathered information and part it into train and test can continue with a model preparing. That is the improvement of model boundaries to accomplish a calculator's best exhibition. At long last, dissect the outcomes.

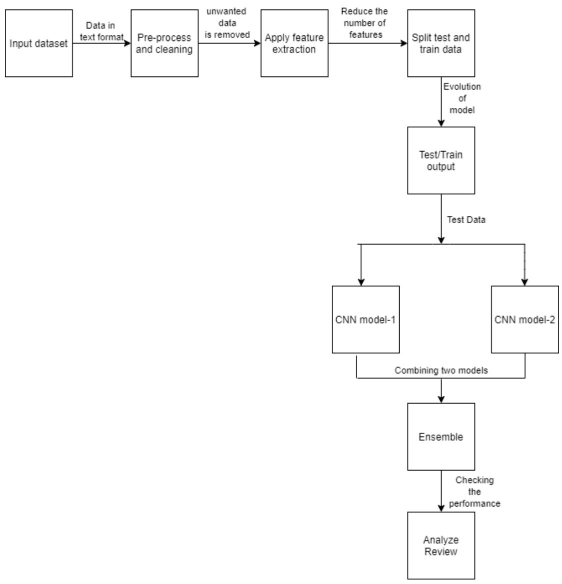

Fig 3: System Architecture

IV. Results

We have executed Fake audit recognition taken dataset by applying Glove Vectorizer highlight extraction procedures. The separated highlights are prepared and anticipated utilizing CNN model 1 and CNN model 2. The prepared models are given contributions for the group model. This outfit orders the surveys as genuine or phony.

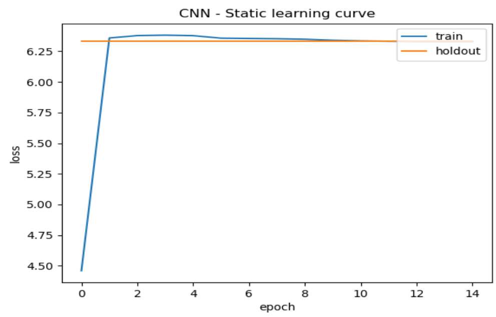

Fig 4: CNN model 1 learning curve

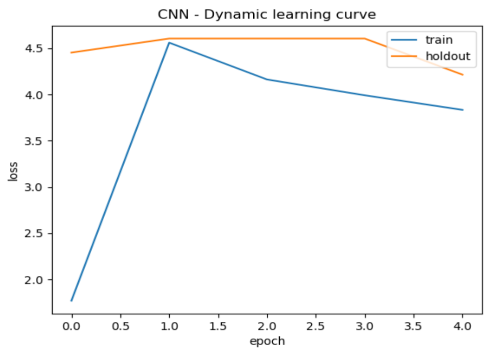

Fig 5: CNN model 2 learning curve

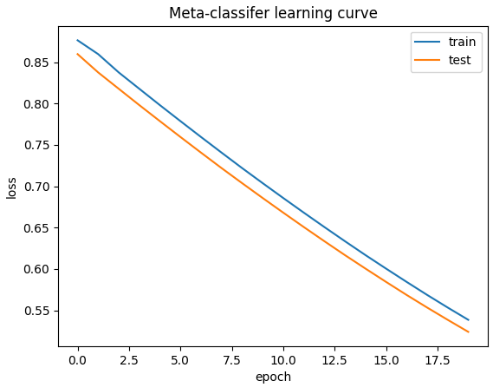

Fig 6 : CNN Meta Classifier Curve

V. Conclusion

In this paper, the review is successfully determined as fake or genuine by using convolutional neural networks. The feature extraction model used is glove vector and the ensemble algorithms considered are in order to produce the efficient product. The data considered here is from yelp.com and in further the data is integrated with some more datasets and can be used to retrieve fake reviews from the respective website.

References

[1]. Fake Review Detection Based onMultiple Feature Fusion and Rolling Collaborative Training by Jingdong Wang, Haitao Kan, Fanqi Meng IEE access Digital Object Identifier 10.1109/ACCESS.2020.3028588

[2]. Xue, Seo, H.; Pluretti, R. Trust-aware review spam detection. In Proceedings of the 2015 IEEE Trustcom/BigDataSE/ISPA, Helsinki, Finland, 20–22 August 2015; Volume 1, pp. 726–733

[3]. Liao, S.Y.; Kwok. Text mining and Probabilistic language modelling for online spam review identification. ACM Trans. Manag. 2011, 2, 25.

[4]. Olivas, J.A.; Romero, Serrano-Guerrero, Conclusion investigation: An audit and near examination of web administrations. Inf. Sci. 2015, 311, 18–38.

[5]. Venkataraman, V.; Mukherjee, A.; Glance, N.S. What Yelp fake review filter might be doing? In the Conference on Social Media, Cambridge, MA, USA, 8–11 July 2013; pp. 409–418.

[6]. Rashid, A.; Anwer, N.; Iqbal, M.; Sher, M. Areas, Techniques, Challenges of Opinion Mining. Int. J. Comput. Sci. 2013, 10, 18–31.

[7]. Venkataraman, V.; Mukherjee, A.; Glance, N.S. Survey on Detection of Shill Reviews by calculating its Linguistic characteristics. Int. J. Emerg. Trends Technol. Comput. Sci. (Ijettcs) 2014, 3, 269–272.

[8].Spirin, N.; Han, J. Survey on web spam detection: Principles and algorithms. ACM Sigkdd Explor. Newsl. 2011, 13, 50–64.

[9]. Chakraborty, M.;Chowdary, C.R. Pal, S.; Pramanik, R. Ongoing advancements in social spam detection and battling procedures: A study.Inf. Process. Manag. 2016, 52, 1053–1073.

[10]. Peng, J.; Choo, K.K.; Ashman, H. User profiling in intrusion detection: A review. J. Netw. Comput. Appl. 2016, 72, 14–27.

[11]. Keele, S. Rules for Performing Systematic Literature Reviews in Software Engineering; Ver. 2.3 EBSE Technical Report: Software Engineering Group, School of Computer Science and Mathematics Keele University, UK and Department of Computer Science University of Durham, UK: 2007; pp. 1–57.

[12].Fang, Bin & Ye, Qiang & Kucukusta, Deniz & Law, Rob, 2016. "Investigation of the apparent value of online tourism reviews: Influence of coherence and reviewer qualities," Tourism Management, Elsevier, vol. 52(C), pages 498-506.

[13]. S. Banerjee and A. Y. K. Chua, "Understanding the way toward composing fake online reviews," Ninth International Conference on Digital Information Management (ICDIM 2014), Phitsanulok, 2014, pp. 68-78.

[14]. Sandifer, Anna & Wilson, Casey & Olmsted, Aspen. (2017). Detection of fake online hotel reviews. 501-502. 10.23919/ICITST.2017.8356460].

[15]. El Murni, Elshrif & Gherbi, Abdelouahed. (2018).Identifying Fake Reviews through Sentiment Analysis Using Machine Learning Techniques.

[16] W. Zhang,R. Y. K. Lau and Li. Chunping, “Adaptive Big Data Analytics for Deceptive Review Detection in Online Social Media”, Thirty Fifth International Conference on Information Systems, Auckland 2014,pp.1-19,2014.

[17] C. Lai, K. Xu, R. Y. Lau, Y. Li, and L. Jing, “Toward a Language Modeling Approach for Consumer Review Spam Detection,” 2010 IEEE 7th International Conference on E-Business Engineering, pp. 1–8, 2010.

[18] M. I. Ahsan, T. Nahian, A. A. Kafi, M. I. Hossain, and F. M. Shah, “Review spam detection using active learning,” 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2016.

[19] M. Ott, Y. Choi, C. Cardie and J.T. Hancock, “Finding deceptive opinion spam by any stretch of the imagination”, ACM, pp.309- 319,2011.

[20] N. Jindal and B. Liu., “Opinion spam and analysis”, Proceedings of the international conference on Web

search and web data mining - WSDM 08 (2008), ACM, pp. 219 –230,2008.

[21] N. Jindal and B. Liu, "Review spam detection", Proceedings of the 16th international conference on World Wide Web - WWW 07 (2007), ACM, pp. 1189–1190, 2007

[22] S. Shojaee, A. Azman, M. Murad, N. Sharef and N. Sulaiman, "A Framework for Fake Review Annotation", 2015 17th UKSIM-AMSS International Conference on Modelling and Simulation, IEEE, pp. 153-158,2015.

[23] J. Koven, H. Siadati, and C. Y. Lin, "Finding Valuable Yelp Comments by Personality, Content, Geo, and Anomaly Analysis," 2014 IEEE International Conference on Data Mining Workshop, pp. 1215–1218, 2014.

[24] S. Banerjee and A.Y.K. Chua. 2014. "Applauses in hotel reviews: Genuine or deceptive ?", 2014 Science and Information Conference (2014), pp. 938–942,2014.

[25] P.Rosso, D.Cabrera, M. Gomez, "Detecting positive and negative deceptive opinions using PU-learning", Elsevier, pp.1-11, 2014.

[26] R. Dewang, A.Singh," Identify of Fake review using new set of lexical and syntactic feature", Proceedings of the Sixth International Conference on Computer and Communication Technology 2015 , ACM, pp. 115-119 ,2015.

[27] J. Fontanarava, G. Pasi, and M. Viviani, "Feature Analysis for Fake Review Detection through Supervised Classification," 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 658–666, 2017.

[28] X. Wang, X. Zhang, C. Jiang, and H. Liu, "Identification of fake reviews using semantic and behavioral features," 2018 4th International Conference on Information Management (ICIM), pp. 92–97, 2018.

[29] S. Feng, L. Xing, A Gogar and Y. Choi,"Distributional footprints of deceptive product reviews". In Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM), pp. 98- 105, 2012.

[30] Akoglu, L, R Chandy and C Faloutsos , "Opinion fraud detection in online reviews by network effects". In Proceedings of the 7th AAAI International Conference on Weblogs and Social Media (ICWSM'13), pp. 2-11, 2013.