2021

An LSTM Based Approach for the Classification of Customer Reviews: An Exploratory Study

Jun Liu
Dakota State University

Akhilesh Chauhan
Dakota State University

Piyush Vyas
Dakota State University

An LSTM Based Approach for the Classification of Customer Reviews: An Exploratory Study

Akhilesh Chauhan
Dakota State University

Jun Liu
Dakota State University

AMCIS 2021 Proceedings

Global, International, and Cross Cultural Research in Info Systems (SIG CCRIS)

Aug 9th, 12:00 AM

An LSTM Based Approach for the Classification of Customer Reviews: An Exploratory Study

Piyush Vyas
Dakota State University, piyush.vyas@trojans.dsu.edu

Jun Liu
Dakota State University, jun.liu@dsu.edu

Akhilesh Chauhan
Dakota State University, akhilesh.chauhan@trojans.dsu.edu

An LSTM Based Approach for the Classification of Customer Reviews: An Exploratory Study

Emergent Research Forum (ERF)

Piyush Vyas
Dakota State University
piyush.vyas@trojans.dsu.edu

Jun Liu
Dakota State University
jun.liu@dsu.edu

Akhilesh Chauhan
Dakota State University
akhilesh.chauhan@trojans.dsu.edu

Abstract

Significant research has been conducted to address the problem of identification and elimination of malicious content. The credibility of such information is always in question, especially in the E-commerce domain. This research proposes a classification model that automatically classifies customer reviews as credible or non-credible. This model encompasses a Long Short-Term Memory (LSTM) as a classification technique. The preliminary results have shown the potential of our model to classify customer reviews as credible / non-credible based on textual features.

Keywords

Customer reviews, e-commerce, text classification, long short-term memory.

Introduction

With the prevalence of E-commerce websites such as, Amazon.com and Ebay.com, online customer reviews are increasingly considered crowd-sourced consumer opinions that significantly influence online purchasing decisions and reputation of businesses. In a survey, 62% of respondents from the US showed their inclination towards customer reviews (Clement 2019), and mostly under 50 years of age group adults incorporated the reviews in their decision making (Smith and Abderson 2016). Therefore, due to increase in people's reliance on reviews, it is important to assess the credibility of customer reviews posted on online crowd-sourced review platforms (such as Yelp.com) and E-commerce websites.

However, fake reviews are so prevalent and are of such sophistication that they are rendering the use of consumer reviews largely ineffective and 48 % of customers are not certain whether reviews are real or fake (Smith and Abderson 2016). The evidence that fake reviews are undermining market effectiveness is compelling. The problem of fake customer reviews has been extensively studied in literature. However, most of the approaches such as (Delgado et al. 2021; Munzel 2015; Shukla et al. 2019) to detect fake reviews rely on either manual assessment of reviews or the use of the Amazon Mechanical Turks (AMT) service, which also involves human interaction to assess the reviews. Such assessment of customer reviews is not scalable in practice and raises questions on the quality of current approaches to detect fake reviews.

The objective of our research is to develop an automated classification model to distinguish credible reviews from non-credible ones. We first identified textual features of online customer reviews such as word count, sentence length, sentiments, and N-grams that can be used to effectively separate credible reviews from non-credible ones. We then developed a deep learning-based model that uses a Long Short-Term Memory (LSTM) for credibility analysis of customer reviews on E-commerce websites.

To fulfill the overall research goal, we investigated following research questions: 1) What features of the reviews can be used to separate credible reviews from non-credible ones? and 2) Can deep learning technologies such as Recurrent Neural Networks (RNN) with LSTM help enhance the effectiveness of fake review detection? Much of the current research on fake review focus on the content of reviews (Hu et al.

2012; Munzel 2015). We, however, believe that text is only part of an effective method for detecting fake reviews. Extraction of textual features fosters the process of model training to feed the information related to review's syntactic and semantics meanings. Therefore, our proposed model considered the aforementioned textual features of reviews. To address the second research question, we have compared our deep learning model that uses LSTM with other algorithms adopted by (Deng et al. 2017; Felipe Gutiérrez et al. 2020; Wang et al. 2018).

Furthermore, from the theoretical perspective, identification of appropriateness of textual features with LSTM will foster the theory related to credibility assessment. Practically, proposed model will help to mitigate the effects of non-credible customer reviews.

Literature Review

In the existing literature, various studies have been conducted to explore different features and machine learning (ML) techniques associated to solve the fake review detection problem.

Two major factors affect customer decisions. First, the description of the product. Second, product reviews by other customers (Wulff et al. 2015). According to Kolhar (2018), online exchanging opinions or experiences with others in the form of reviews affects business reputation and product sales. The textual information must be considered as an essential aspect because mainly, customer reviews are in textual form with pieces of evidence such as product images and videos. Customer review's textual features such as writing style also play an essential role in finding out the manipulated content, but people have different writing styles. Hence, it is challenging to find out misleading content if there isn't enough data available (Hu et al. 2012). User-centric features consist of information related to how users behave while posting reviews on platforms like Yelp. The information provided by users are divided into four types: Personal Features, Social Features, Reviewing Activity Features and Trusting Features. Furthermore, three N-gram feature sets, namely Unigrams, Bigrams, and Trigrams, showed their potential for classifying fictitious or deceptive text (Felipe Gutiérrez et al. 2020).

Popular product categories have a large number of reviews and it is hard for customers to read all reviews to identify the informative one, therefore, Hu and Liu (2004) stated that Opinion mining (i.e., extraction of customer sentiments) is a technique that has the potential to classify the customer reviews as fake or real. Customer reviews may change over time (Safi and Yu 2017). E-commerce companies are trying to build their models for the analysis of customer review's nature (real/fake). Therefore, classification is a way to classify customer reviews in various classes such as fake, real and neutral. Training of classifiers such as Naive Bayes (i.e., a probability-based classifier based on Bayes theorem) and Support Vector Machine (i.e., a supervised ML technique) with N-gram features will show promising results in classification (Ott et al. 2011). Felipe Gutiérrez et al. (2020) have utilized various ensemble-based ML techniques such as Random Forest(RF) to detect fake reviews. Ensemble techniques utilize the collective results of various similar ML techniques such as RF as an ensemble technique uses results of more than one decision tree technique. According to the Friedman test, the Ada Boost (i.e., an ensemble-based ML technique) classifier has been proven to be the best one by statistical means (Barbado et al. 2019).

As Kolhar (2018) mentioned that classification of customer reviews required labeled data and manual annotation creates difficulty to label correctly. Therefore, by complementing prior research, we have used the labeled dataset of Yelp Hotel customer reviews from (kaggle 2020; yelp 2020) and adopted textual features to classify the customer reviews as credible or non-credible. Furthermore, we have applied the RNN based LSTM deep learning technique that has potential to overcome the performance of traditional ML techniques adopted by (Barbado et al. 2019; Deng et al. 2017; Felipe Gutiérrez et al. 2020; Ott et al. 2011). Moreover, existing research such as (Delgado et al. 2021; Shukla et al. 2019) adopted AMT services during the process of fake review detection. This could lead to an expensive solution for a given large datasets. Therefore, we have incorporated the authentic labeled dataset provided by Yelp to train our model.

Research Methodology

Figure 1. shows the high-level view of our proposed approach to classify credible and non-credible customer reviews. Details are in following sections.

Data collection

We have utilized open-source Yelp Customer review data from (kaggle 2020; yelp 2020). There are many business categories in the Yelp dataset. We have used the Hotel category. This category has 359,052 reviews along with information regarding User ID, rating, and Label (1-Genuine / -1 Fake). We have considered Genuine as credible and Fake as non-credible reviews. We have used 70% of data for training, and 30% of data for testing the model.

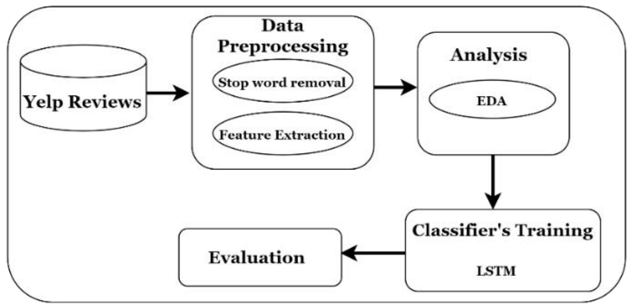


Figure 1. High-Level view of the proposed approach

Data Preprocessing and EDA

Prior to train a deep learning-based LSTM model it is essential to perform data preprocessing tasks to enhance the model performance. Hence, we cleaned the data by removing insignificant information such as User ID, which will not contribute well to model training. We kept the text of customer reviews. Further, we removed stop words such as "he"," she"," it" because these would not add value to prediction. We also performed lemmatization to get the base words (i.e., the lemma of "meeting" is meet). Furthermore, we extracted textual features such as word count, sentence length (i.e., number of characters), and N-grams (sequence of frequently used words). As an exploratory data analysis, we have used word cloud to explore the most frequently used words that are present in credible and non-credible customer reviews. We also analyzed top five Unigrams, Bigrams, and Trigrams to see the most frequently used phrases.

LSTM Classifier

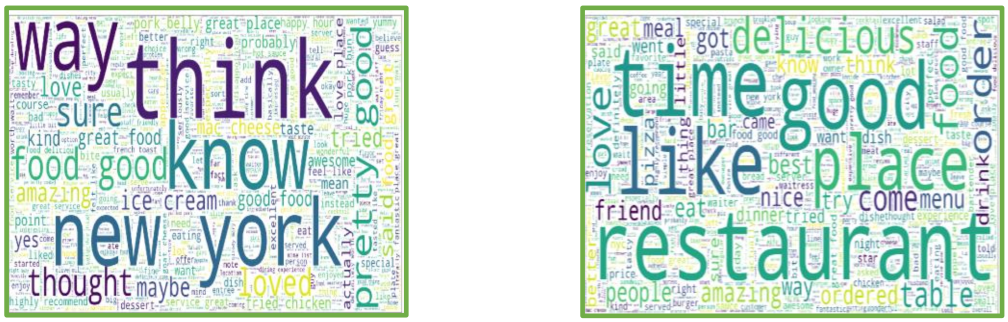
In this work, we classify customer reviews into credible vs. non-credible. The proposed model utilizes textual features from Yelp customer review data set. LSTM is RNN based deep learning neural network that comprises distinct layers with memory cells (to keep the information) and gate units (to regulate the information flow in memory cells). Basic RNN has the potential to simplify sequential data processing and majorly has applied to text or speech recognition (Heinrich et al. 2019). However, standard RNNs become unable to learn long-term dependencies.To overcome this discrepancy, LSTM was introduced to achieve significant performance (Zhou et al. 2015). Our purpose behind the usage of LSTM is, it facilitates the variable-length sequential data such as customer reviews, and integrated units of LSTM stores the learned knowledge for future utilization during classification task. Furthermore, LSTM showed noticeable results with short sentences and outperformed advanced models such as Bidirectional Encoder Representations from Transformers (BERT)(Ezen-Can 2020).

To train LSTM model, we have used embedding layer and dense layers provided by Keras, an open-source python-based library. Before LSTM layer, we have used an embedding layer that produced fixed-length vectors for each word. Therefore, the sequence of layers in our model is, 1) embedding layer, 2) LSTM layer, 3) dropout layer,4) dense layer. We have used dropout layer to restrain the entire model from overfitting (i.e., perfectly fit model). We have used Sigmoid activation function at the dense layer to produce binary output (i.e., credible or non-credible).

Preliminary Results and Discussion

As a result of EDA, Figure 2a shows the word cloud of frequently used word in credible customer reviews and Figure 2b shows the word cloud for non-credible reviews. The larger the word size in cloud, the more it is used by customers while writing the reviews. Table 1 shows top five Unigrams, Bigrams, and Trigrams that pinpoint the group of words frequently used together such as the sequence of words - "highly recommend place" indicates customer recommendations. These syntactic features help the model to learn about word's meaning.

Furthermore, as shown in Table 2 we compared our model by utilizing test accuracy as an evaluation measure with other ML-based models used by existing research. Test accuracy indicates that how proficiently our model performed the prediction of credible / non-credible reviews on an unknown dataset (i.e., test set). Our model achieved 89.55% of test accuracy and 90% of recall. Recall is another evaluation measure to judge the correctly classified reviews that often used for binary classification problem (Vyas and El-Gayar 2020).



(a) (b)

Figure 2. Word cloud of frequently used words: a) in credible reviews; b) in non-credible reviews

Unigram	Bigram	Trigram
good	Really good	seated right away
great	Food good	food great service
delicious	Great place	highly recommend place
got	Feel like	great service great
nice	Just right	staff super friendly

Table 1. Top Five N-grams

Models	Proposed LSTM based	Decision Tree (Felipe Gutiérrez et al. 2020)	(Deng et al. 2017)'s Semi-Supervised
Accuracy	89.7%	79.6%	87.3%

Table 2. Comparison of proposed approach with existing models

However, the preliminary results show that our model demonstrating a potential to extract textual features from reviews and thus outperforms the existing models. Although further hyperparameter tuning is required to select the optimum (i.e., best performing) parameters of LSTM, dropout, and dense layers to achieve higher performance. Additional model improvement plans will be to incorporate loss normalization (i.e., to minimize the loss) and batch normalization (i.e., to make the model work fast).

Conclusion

In this research, we proposed a deep learning-based customer reviews classification model that uses word count, sentence length, sentiments, and N-grams features to classify customer reviews into credible vs non-credible. Moreover, our preliminary results have shown that the proposed LSTM based approach has potential to outperform existing ML-based approaches. Thus, we successfully addressed our identified research questions and achieved our objective. In the future, successful implementation of the proposed model will automate the process of detecting fake online customer reviews and can be served as a unified model to detect fake content.

REFERENCES

Barbado, R., Araque, O., and Iglesias, C. A. 2019. "A Framework for Fake Review Detection in Online Consumer Electronics Retailers," *Information Processing and Management* (56:4), Elsevier Ltd, pp. 1234–1244.

Clement, J. 2019. "Online Reviews - Statistics & Facts | Statista," *Statista*. (https://www.statista.com/topics/4381/online-reviews/, accessed January 30, 2021).

Delgado, C., Alejandro, A., Glisson, W., Shashidhar, N., Mcdonald, J., Grispos, G., and Benton, R. 2021. "Deception Detection Using Machine Learning," in *Proceedings of the 54th Hawaii International Conference on System Sciences*, p. 7122. (https://hdl.handle.net/10125/71478).

Deng, H., Zhao, L., Luo, N., Liu, Y., ... G. G.-... on P. and, and 2017, U. 2017. "Semi-Supervised Learning Based Fake Review Detection," in *In 2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC)*, pp. 1278–1280.

Ezen-Can, A. 2020. "A Comparison of LSTM and BERT for Small Corpus," *ArXiv* (abs/2009.0).

Felipe Gutiérrez, L., Abri, F., Namin, A. S., Jones, K. S., Gutierrez-Espinoza, L., and Sears, D. R. W. 2020. "Ensemble Learning for Detecting Fake Reviews," *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, pp. 1320–1325.

Heinrich, K., Zschech, P., Skouti, T., Griebenow, J., and Riechert, S. 2019. "Demystifying the Black Box: A Classification Scheme for Interpretation and Visualization of Deep Intelligent Systems," in *AMCIS*, p. 8.

Hu, M., and Liu, B. 2004. "Mining Opinion Features in Customer Reviews," in *AAAI*, pp. 755–760.

Hu, N., Bose, I., Koh, N. S., and Liu, L. 2012. "Manipulation of Online Reviews: An Analysis of Ratings, Readability, and Sentiments," *Decision Support Systems* (52:3), North-Holland, pp. 674–684.

kaggle. 2020. "Kaggle." (https://www.kaggle.com/abidmeeraj/yelp-labelled-dataset, accessed February 28, 2021).

Kolhar, M. 2018. "E-Commerce Review System to Detect False Reviews.," *Science and Engineering Ethics* (24:5), pp. 1577–1588. (https://doi.org/10.1007/s11948-017-9959-2).

Munzel, A. 2015. "Malicious Practice of Fake Reviews: Experimental Insight into the Potential of Contextual Indicators in Assisting Consumers to Detect Deceptive Opinion Spam," *Recherche et Applications En Marketing (English Edition)* (30:4), SAGE Publications, pp. 24–50.

Ott, M., Choi, Y., Cardie, C., and Hancock, J. T. 2011. "Finding Deceptive Opinion Spam by Any Stretch of the Imagination," *ArXiv:1107.4557*, pp. 309–319.

Safi, R., and Yu, Y. 2017. "Online Product Review as an Indicator of Users' Degree of Innovativeness and Product Adoption Time: A Longitudinal Analysis of Text Reviews," *European Journal of Information Systems* (26:4), pp. 414–431. (https://doi.org/10.1057/s41303-017-0045-2).

Shukla, A., Wang, W., Gao, G. (Gordon), and Agarwal, R. 2019. "Catch Me If You Can — Detecting Fraudulent Online Reviews of Doctors Using Deep Learning," *SSRN Electronic Journal*, Elsevier BV.

Smith, A., and Abderson, M. 2016. "Online Reviews and Ratings | Pew Research Center," *PewResearchCenter*. (https://www.pewresearch.org/internet/2016/12/19/online-reviews/, accessed January 30, 2021).

Vyas, P., and El-Gayar, O. 2020. "Credibility Analysis of News on Twitter Using LSTM: An Exploratory Study," in *AMCIS 2020 Proceedings*, AMCIS 2020 Proceedings, p. 17.

Wang, X., Zhang, X., Jiang, C., and Liu, H. 2018. "Identification of Fake Reviews Using Semantic and Behavioral Features," in *2018 4th International Conference on Information Management, ICIM 2018*, Institute of Electrical and Electronics Engineers Inc., June 21, pp. 92–97.

Wulff, D. U., Hills, T. T., and Hertwig, R. 2015. "Online Product Reviews and the Description-Experience Gap," *Journal of Behavioral Decision Making* (28:3), Wiley-Blackwell, pp. 214–223.

yelp. 2020. "Yelp Dataset." (https://www.yelp.com/dataset, accessed February 28, 2021).

Zhou, C., Sun, C., Liu, Z., and Lau, F. C. M. 2015. "A C-LSTM Neural Network for Text Classification," *ArXiv Preprint ArXiv:1511.08630*. (http://arxiv.org/abs/1511.08630).