

Graph Attention Network Towards Prediction and Classification of Fake Reviews in E-Commerce Application

Dr. C. S. Saratha¹, C. Harikrishnan², N. Kapildesh³, J. Leenu⁴, M. Meena⁵

Associate Professor¹, PG student^{2,3, 4,5}

Department of Information Technology^{1,2,3,4,5}

CMS College of Science and Commerce (Autonomous), Coimbatore, India^{1,2,3,4,5}

ARTICLE INFO

Received: 30 Oct 2024

Revised: 05 Nov 2024

Accepted: 25 Dec 2024

ABSTRACT

Nowadays digital marketplace is growing tremendously due to changes in lifestyle of peoples and rapid increase of the technologies. Especially E commerce is becoming driving growth of many businesses through its cost effective strategies. In particular, E commerce gathers user reviews to their businesses as a user interaction and user intention to enhance purchase decision, improve user experiences and to build trust on their product and services. Thus, it becomes mandatory to analyze and evaluate the user reviews on their products and services. Traditionally, opinion mining and sentiment analysis were used to analyze the sentiment associated along the reviews in form of positive and negative aspects. Despite of several advantages of those models fails to capture the fake reviews which lead to several consequences in form of consumer deception, loss of trust and unfair competitions. In order to mitigate those challenges, a new deep learning architecture is designed to detect and classify the fake reviews propagating in the e commerce platform such as amazon, bigbasket and flipcart. In this work, Convolution Neural Network along Long Short Term Memory Mechanism has been designed as deep learning model to detect the fake reviews on product reviews. Initially data preprocessing is carried out using stop word removal, stemming and weighted tokenization technique. Weighted Token in form of word vector is projected to convolution layers of the Convolution Neural Network to extract the user specific features and product specific features separately. Further those extracted user specific and product specific feature is projected to pooling layer. Pooling layer incorporates Long Short Term Memory Mechanism to identify the long dependency map among the user specific features and product specific features. Finally long dependency map is projected to fully connected layer containing softmax function to predict and classify original review and fake reviews. Experimental analysis and performance analysis of the proposed model is performed using amazon dataset. On experimental and performance analysis, it is proved that proposed model outperforms state of art approaches in classifying the original and fake reviews with accuracy of 98.7%.

Keywords: Product Review, E Commerce Platform, Convolution Neural Network, Long Short Term Memory Model, Sentiment Analysis, Opinion Mining, Fake Review Prediction

1. INTRODUCTION

Shopping Behavior of the customer is changing tremendously after huge advancement and increment of the technologies utilization among peoples. Especially online shopping has increased from last decade due to lifestyle changes of the people. In particular, E commerce platform has attracted peoples with its cost effective strategies. Meanwhile, user reviews to the online products is changing purchase characteristics among the customer as it becoming significant strategies among many businesses in the e-commerce platform as it enhances purchase decision, improve user experiences and to build trust on their product and services. Traditionally, opinion mining and sentiment analysis were used to analyze evaluate and classify the sentiment associated with reviews in form of positive and negative opinions. Despite of several advantages of those models fails to capture the fake reviews which lead to several consequences in form of consumer deception, loss of trust and unfair competitions among products and services.

In order to mitigate those challenges, a new deep learning architecture is designed to detect and classify the reviews propagating in the e commerce platform such as amazon, bigbasket and flipcart with respect to user intention and its evolutions on user opinions along accessing their long-term reliance of users over numerous time frames. In this work, Convolution Neural Network along Long Short Term Memory Mechanism has been designed as deep learning model to detect the fake reviews on product reviews as it is efficient in handling temporal and time changing contents. Initially data preprocessing is carried out using stop word removal, stemming and weighted tokenization technique. Weighted Token in form of word vector is projected to convolution layers of the Convolution Neural Network to extract the user specific features and product specific features separately. Further those extracted user specific and product specific feature is projected to pooling layer. Pooling layer incorporates Long Short Term Memory Mechanism to identify the long dependency map among the user specific features and product specific features. Finally long dependency map is projected to fully connected layer containing softmax function to predict and classify original review and fake reviews

The remaining part of the article is organized as follows, section 2 mentions the review of related work on basis of sentiment analysis and opinion mining, section 3 defines a new design as long short term memory integrated convolution neural network to predict and classify the user reviews. Experimental analysis of the proposed model is carried out using benchmark amazon dataset and performance analysis of model is evaluated using accuracy metrics in section 4. Finally article is summarized with major findings in the section 5.

2. RELATED WORK

In this section, review of related work on basis of sentiment analysis and opinion mining to the user review were analyzed in terms of its design and experimental aspects of machine learning and deep learning model through benchmark amazon dataset.

2.1. Recurrent Neural Network towards Opinion Classification

In this literature, Recurrent Neural Network is employed to predict and classify the user opinion of the various product and services of the amazon. Initially model uses preprocessing technique to obtain the token vector. Obtained token vector is processed in the recurrent layer to identify the complex relationships among opinion of the users to the specified products and related products. On more similarities, it is considered as original reviews and huge dissimilarities among reviews is considered as fake review.

2.2. Gated Recurrent Network for Opinion Evolution Prediction

In this literature, Gated Recurrent Network is employed to predict evolution of the user opinion on the product and services to the products in the amazon platform. Initially model uses preprocessing technique to obtain the token vector. Obtained token vector is processed in the hidden layer to identify the complex relationships and those relationships are stored as hidden state information. Further gating mechanism uses different cell states to identify the evolution of the user opinion on the products and its services to classify the product review of the specified user as original or fake review.

3. PROPOSED MODEL

In this section, design of fake review prediction and classification deep learning framework is performed using long short term memory integrated convolution neural network along preprocessing processes such as stop word removal, stemming, tokenization and token weighting and token embedding.

2.1.Data pre-processing

Data Preprocessing is initial phase of the model which transforms textual dataset to numerical dataset with several processing methods is as follow

3.1.1. Stop Word Removal

Stop word removal is a pre-processing step performed to eliminate the stop words in form of noun, pronouns articles, preposition and conjunctions in the user reviews. Specified process reduces size of the review data. Further

it eliminates punctuation, whitespaces, and delimiters in the review data for post processing. In particular, Stop word removal process never impacts prediction outcomes.

3.1.2. Stemming process

Stemming is another pre-processing step performed to reduce character length of the review phrases or words as it mostly affixed to suffixes and prefixes.

3.1.3. Tokenization

Tokenization is primary step in pre-processing as it splits the user reviews words as separate tokens which is referred as word based tokenization. Figure 1 represents architecture of the proposed user fake review prediction model.

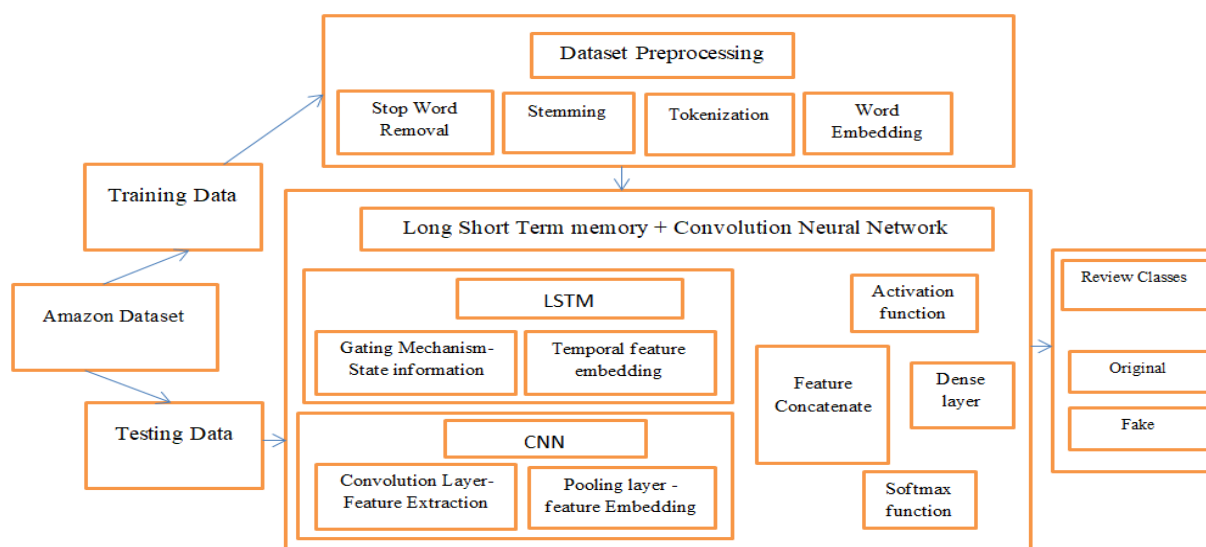


Figure 1: Fake Review Prediction and Classification Framework

3.1.4. Word Embedding or Weighted Tokenization

Word Embedding or weighted tokenization transforms the text into numerical format. It assigns numerical vector to each token with weight reflecting word significance. In this work, term frequency is employed for word embedding which is referred as ratio of number of occurrence of the term or word in the specified review to total number of words in the reviews. Term frequency computation for word embedding is as follows

$$TF(\text{Token}) = \frac{\text{No of occurrence of the Specified word in the review}}{\text{Total number of words in the review}}$$

3.2. Long Short Term Memory Integrated Convolution Neural Network

Token Vector from the pre-processing mechanism is applied to Long Short Term Memory Integrated Convolution Neural Network. In specific, token vector composed of user information and product information. Thus convolution layer is used to extract the user specific and product specific separately.

3.2.1. Convolution Layer

In this layer, Two Kernel Filters with its function were employed to the token vector. Filter function extracts the user information and product information separately.

User specific feature is represented as

$$U_f = \{U_1, U_2, \dots, U_n\}$$

Product specific features is represented as

$$P_f = \{P_1, P_2, \dots, P_n\}$$

3.2.2. Pooling layer- Long Short Term Memory

In this layer, long short term memory is interfaced to process user specific features and product specific features to establish long dependency map. Long dependency map combines the user similar characteristics to product features. Table 1 illustrates the hyper parameter of long short term memory mechanism integrated convolution neural network.

Table 1: Hyper Parameter Tuning of Convolution Neural Network

Network Parameter	Values
Batch size	108
Learning Rate	0.06
Epoch	45
Activation Function	ReLu
Error function	Cross entropy
Pooling function	LSTM mechanism
Softmax function	Support Vector Machine

3.2.3. Fully connected layer

In this layer, softmax function, loss function is employed along activation function to process long dependency map to predict and classify it. Detailed working of those functions is as follows

- **Softmax Function**

Softmax function uses support vector machine to process the long dependency map. Decision boundary function is applied to long dependency map to classify the user review.

- **Loss function**

Loss function uses cross entropy function as it capable of eliminating the underfitting and overfitting issues of the model.

Algorithm 1: Prediction of fake review on user opinion evolution to product review

Input: Amazon Dataset

Output: fake review prediction

Process

 Data Pre- Process ()

 Preprocess_Dataset ()

S_t = Stop word Removal (Dataset)

S_s = Stemming (S_t)

T = Word based tokenization(S_s)

T_f = Term Frequency(T)

CNN()

 Convolution Layer()

 Kernel (Tokenized Vector)

 Computes user specific features and Product Specific features

 User feature map and product feature map

Pooling Layer_LSTM ()

Generates the long dependency map

Fully Connected Layer ()

Sofmax function _SVM(Dependency map)

Opinion Class: { Normal , Fake}

4. EXPERIMENTAL RESULTS

Experimental analysis of Proposed Long short term memory mechanism integrated Convolution Neural Network approach is carried out using amazon dataset in python environment with tensor flow functionalities and multiple libraries such as NumPy and Pandas for data manipulation, scikit learn for baseline models and Matplotlib for data visualization. Panda's libraries for data processing using amazon dataset extracted from kaggle repository [15]. Dataset is portioned as 60 percent employed for model training and 40 percent for model testing. Grid search was conducted to identify the optimal hyperparameter for model training to LSTM and CNN.

4.1. Performance analysis

Performance analysis of the model is performed using test data through confusion matrix. Confusion matrix generates the values to the true positive, false negative, true negative and false positive parameters. Those parameter values of the matrix demonstrates strong performance of the LSTM +CNN model in classifying the user reviews on integrating the time series data(processed by LSTM network) and relational data(processed RNN network). Model achieved enhanced prediction accuracy of 98.4 percent. Further precision analysis, recall analysis and accuracy analysis were performed as follows

4.1.1. Precision Analysis

Precision analysis is performed to determine no of the aggregated feature correctly classified into user review classes among the total aggregated features. Precision analysis represented on parameter of confusion matrix is as follows

$$\text{Precision} = \frac{TP}{TP+FP}$$

Figure 2 represents precision analysis of the model against conventional approaches. It represents model ability towards identifying user review to products of the e-commerce application with respect to continuously updating the user behavior.

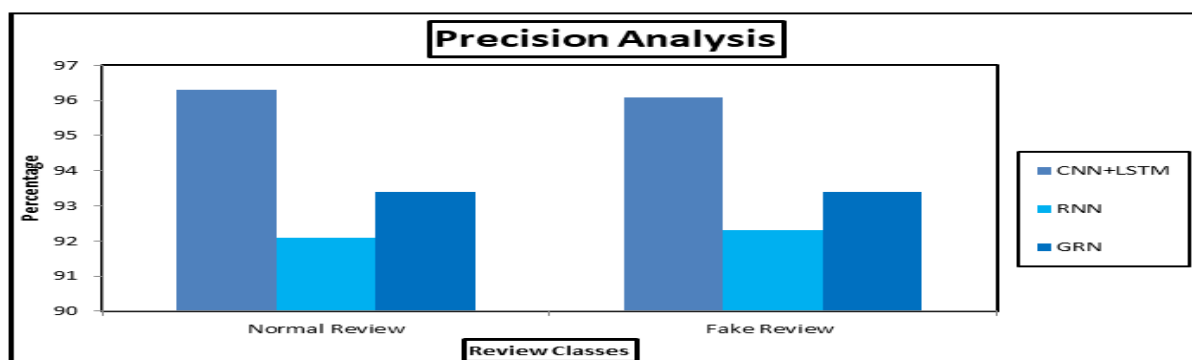


Figure 2: Precision Analysis

4.1.2. Recall Analysis

Recall analysis is performed to determine no of the aggregated feature incorrectly classified into user review classes among the total aggregated features. Recall analysis represented on parameter of confusion matrix is as follows

$$\text{Recall} = \frac{TN}{TP+FP}$$

Figure 3 represents recall analysis of the model against conventional approaches. It represents model ability towards identifying user reviews to products of ecommerce applications on continuously updating the user behavior and perceptions.

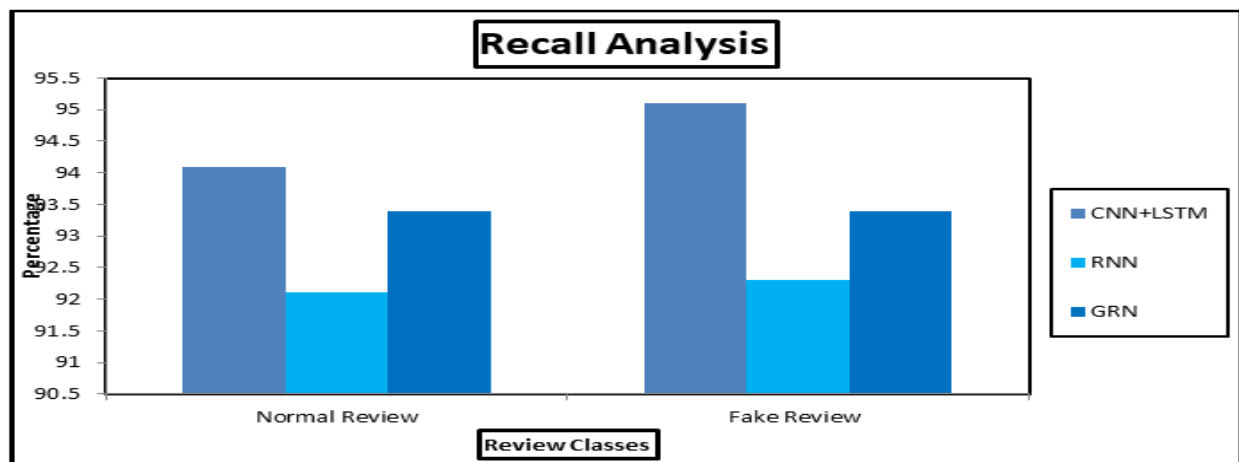


Figure 3: Recall Analysis

4.1.3. Accuracy Analysis

Accuracy analysis is performed as aggregation of recall and precision towards detecting aggregated feature among total features into user review classes. Accuracy analysis represented on parameter of confusion matrix is as follows

$$\text{Accuracy} = \frac{TP+TN}{TN+FN+TP+FP}$$

Figure 4 represents accuracy analysis of the model against conventional approaches. It represents model ability towards identifying user review to product of the e commerce application on continuously updating the user behavior and perceptions.

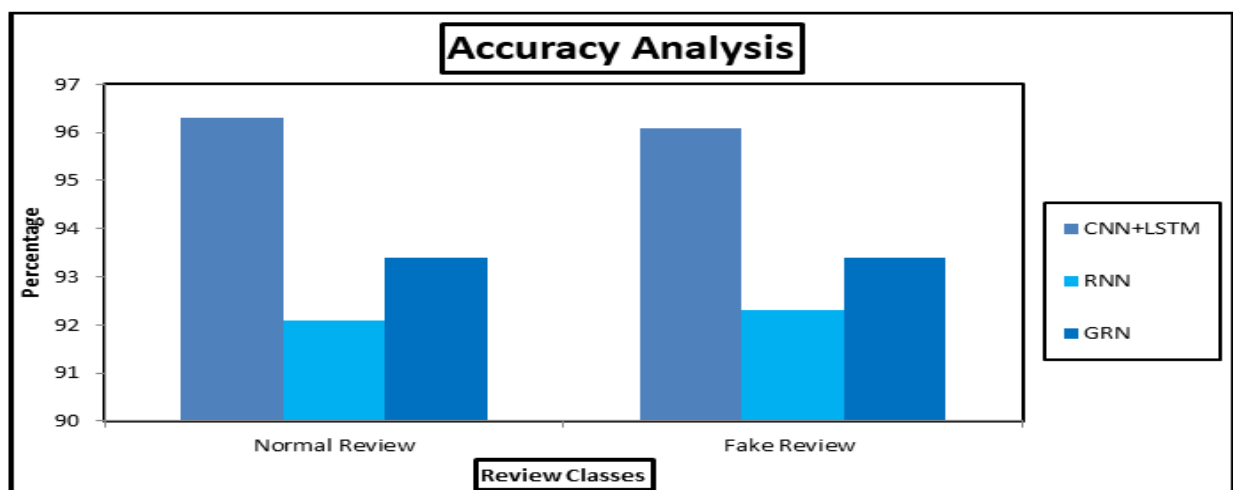


Figure 4: Accuracy analysis

Finally performance of CNN+LSTM architecture performs better with detection accuracy 98.4% on compared to conventional approaches. Table 2 mentions the performance evaluation of the deep learning architectures in user review classification

Table 2: Performance Evaluation of Deep learning architecture in Fake Review Prediction

Technique	Classes	Precision	Recall	Accuracy
CNN+LSTM – Deep learning	Normal Review	98.1	94.2	98.4
	Fake Review	98.4	94.5	98.2
RNN- Deep learning	Normal Review	96.4	91.7	96.7
	Fake Review	96.6	91.8	96.3
GRN – Deep learning	Normal Review	94.6	90.1	94.6
	Fake Review	94.3	90.5	94.3

CONCLUSION

We designed and implemented Convolution Neural Network along Long Short Term Memory Mechanism towards sentiment analysis of the user review to product review from amazon dataset to detect the fake reviews on product reviews. Initially data preprocessing is carried out using stop word removal, stemming and weighted tokenization technique. Weighted Token in form of word vector is projected to convolution layers of the Convolution Neural Network to extract the user specific features and product specific features separately. Further those extracted user specific and product specific feature is projected to pooling layer. Pooling layer incorporates Long Short Term Memory Mechanism to identify the long dependency map among the user specific features and product specific features. Finally long dependency map is projected to fully connected layer containing softmax function to predict and classify original review and fake reviews. Experimental analysis and performance analysis of the proposed model using amazon dataset proved that proposed model outperforms state of art approaches in classifying the original and fake reviews with accuracy of 98.7%

REFERENCES

- [1] Min E, Guo X, Qiang “A survey of clustering with deep learning: from the perspective of network architecture” in IEEE Access, Vol. 6, issue.39, pp: 501–14, 2018.
- [2] Chowdary NS, Prasanna DS, Sudhakar P. “Evaluating and analyzing clusters in data mining using different algorithms”. International Journal of Computer Science and Mobile Computing, Vol.3, PP: 86–99, 2014.
- [3] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 8, pp. 1798–1828, 2013
- [4] Dizaji KG, Herandi A, Cheng,” Deep clustering via joint convolutional autoencoder embedding and relative entropy minimization. In: 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017, 5747–56.
- [5] Xie J, Girshick R, Farhadi A ”Unsupervised deep embedding for clustering analysis. In: International Conference on Machine Learning. New York City, NY, USA: ICMLR, 2016, 478–87.
- [6] S.Praveen & R.Priya ” A Deep Conceptual Incremental learning Based High Dimensional Data Clustering model- A Deep Learning Approach” in Turkish Journal of Computer and Mathematics, 2021
- [7] H. Liu, M. Shao, S. Li, and Y. Fu, “Infinite ensemble for image clustering,” in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, , pp. 1745–1754, 2016
- [8] J. L. Elman, “Distributed representations, simple recurrent networks, and grammatical structure,” Machine learning, vol. 7, no. 2-3, pp. 195–225,1991.122, 2015.
- [9] K. G. Dizaji, A. Herandi, C. Deng, W. Cai, and H. Huang, “Deep clustering via joint convolutional autoencoder embedding and relative entropy minimization,” in 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, , pp. 5747–5756, 2017.
- [10] P. Xiao, G. K. Venayagamoorthy, and K. A. Corzine, “Combined training of recurrent neural networks with particle swarm optimization and backpropagation algorithms for impedance identification,” in Proceedings of the IEEE Swarm Intelligence Symposium (SIS ’07), pp. 9–15, April 2007.
- [11] P. Huang,Y. Huang,W.Wang, and L.Wang, ``Deep embedding network for clustering," in Proc. 22nd International. Conference. Pattern Recognition. (ICPR), Aug. 2014, pp. 1532-1537.
- [12] S. Zhang, Y. Zhan, and D. N. Metaxas, “Deformable segmentation via sparse representation and dictionary learning,” Med. Image Anal., vol. 16, no. 7, pp. 1385–1396, Oct. 2012.

- [13] W. Harchaoui, P. A. Mattei, and C. Bouveyron, "Deep adversarial Gaussian mixture auto-encoder for clustering," in Proc. ICLR, 2017, pp. 1-5.
- [14] N. Dilokthanakul et al. (2016). "Deep unsupervised clustering with Gaussian mixture variational autoencoders." [Online]. Available: <https://arxiv.org/abs/1611.02648>
- [15] G. Chen. (2015). "Deep learning with nonparametric clustering." [Online]. Available: <https://arxiv.org/abs/1501.03084>