

A Semantic Sentiment Analysis System on Product Reviews -A Neural Network-Based Approach

Dr.Eldho K J

*Asst. Professor, Department of Computer Science,
Mary Matha Arts and Science College, Mananthavady, Kerala*

eldhorvs@gmail.com

Abstract: Objectives: Sentiment analysis is a method used to extract emotions, thoughts and opinions from text documents. In the aspects of online shopping, sentiment analysis is used to identify the opinions and feedback of customers about a product, allows brands to identify customers are happy or not. So that they can tailor the products to meet customers satisfaction.

Findings: Sentiment analysis is a part of Natural Language Processing and Artificial Intelligence which is used to extract the opinions, thoughts or emotions from text documents. With the rapid development of the internet, E-commerce websites and social media, a huge volume of unstructured and unspecified data is available on the web. The identification of sentiment on public views will help individuals and decision-makers. In this proposed study.

Novelty: Evaluation The standard performance metrics for classification problems were used to assess all the models, namely accuracy, Precision, Recall, F-measure and Area under the ROC (Receiver Operating Characteristic) curve (AUC-ROC). This work we proposed a sentiment analysis on product reviews obtained from E-commerce websites which can measure the sentiments on public views. Here the methodology we used neural network methods (NN Method) – Word embedding and Bidirectional Long Short Term Memory (Bi-LSTM). As per the proposed work the precision and recall parameters of the proposed sentiment analysis system are evaluated as 84.2% and 85.1% respectively. The efficiency of the system is greater than that of RNN and LSTM methods.

Key Words- *Sentiment Analysis, Word Embedding, CBOW, SkipGram, iLSTM.*

1. INTRODUCTION

Online shopping and the effect of the internet and social media on human life is growing rapidly day by day. Online shopping or E-commerce means buying and selling products and services over the internet. In recent times especially in the Covid lock down period, people prefer online shopping as it has many advantages. The customers can able to search and compare different products available in different outlets. Comparison will give an outcome of quality purchase. People can compare the products based on their price, quality and user reviews. The most important factor in E-commerce is user review. Different users have different opinions about the particular products. So analyzing and identifying the sentiment or emotion of user reviews is the most important factor. Artificial Intelligence plays an important role in analysing the emotions based on different data science methods.

There are different works are conducted by different eminent persons. *Xing Fang* proposed a system for sentiment analysis of online product reviews. *Aytug Onan* proposed a weighted word embedding and deep neural network-based sentiment analysis system for product

reviews. The other works are compared and discussed in Review of literature.

The organization of this paper is as follows. The following section discussed the review of related works. Section 3 contained the methodologies and system design. Section 4 contained the experimental results and analysis. Finally, section 5 discussed the conclusion and future works.

2. OBJECTIVES OF THE STUDY

The main objective of this work is to find out the sentiment analysis by using different methods in an experimental way.

Sentiment analysis is a method used to extract emotions, thoughts and opinions from text documents. In the aspects of online shopping, sentiment analysis is used to identify the opinions and feedback of customers about a product, allows brands to identify customers are happy or not. So that they can tailor the products to meet customers satisfaction.

There are different types of sentiment analysis- 1) Polarity based:- In this, the sentiments are classified as positive, negative and neutral. 2) Feelings or Emotions based - Here the sentiments are classified as happy, sad, angry etc.. 3) Intention based - The sentiment is classified as interested v/s not interested. Etc.[1].

Here we proposed a polarity based sentiment analysis on product reviews, i.e. the reviews from various E-commerce websites like Amazon, Flipkart etc, are classified as positive, negative or neutral by using neural network methods. Word embedding algorithm word2vec was used for semantic modelling of comment texts and BiLSTM [4] was used for extract contextual information and finally, a sigmoid activation function was used for sentiment classification.

Evaluation The standard performance metrics for classification problems were used to assess all the models, namely accuracy, Precision, Recall, F-measure and Area under the ROC (Receiver Operating Characteristic) curve (AUC-ROC).

2. REVIEW OF RELATED WORKS

Xing Fang et al[2] proposed a system for sentiment analysis of online product reviews. They proposed a sentiment polarity categorization based on sentence-level and review level. They compare the performance of three different classification algorithms- Naive Bayes, Random Forest and Support Vector Machine.

Aytug Onan [3] proposed a weighted word embedding and deep neural network-based sentiment analysis system for

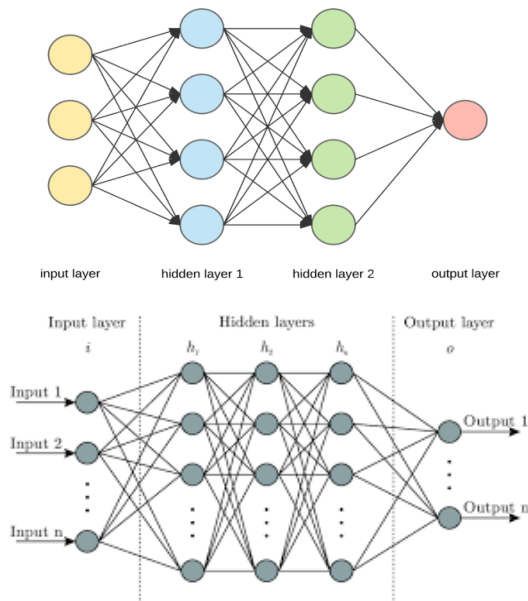
product reviews. The empirical analysis of different word embedding schemes with several weighting functions is evaluated in conjunction with LSTM.

Huyen Tran et al [4] proposed a method to improve the performance of sentiment analysis of tweets using the feature ensemble model, related to tweets containing fuzzy sentiment by taking elements such as lexical, word-type, semantic, position and sentiment polarity of words. The CNN model is used to analyze the sentiment of tweets. The F1 score of the result shows the effective improvement of the performance of the sentiment analysis process.

Giuxian et al [5] developed a sentiment analysis system with an improved word representation, which combines the contribution of sentiment information into the TF-IDF algorithm. The weighted word vectors are input to a BiLSTM to capture the context information. The sentiment of the comment is found by feed-forward neural classifier. Then the proposed algorithm is compared with other methods. The experimental result shows that the proposed system has high accuracy than other methods.

Rajkumar et al [6] proposed a sentiment analysis system on product reviews using machine learning techniques. They used machine learning algorithms – Naive Bayes and Support Vector Machine for sentiment classification. From the literature, we identified many sentiment analysis works are happening. Most of the systems are on social media sentiment analysis. The latest systems are based on neural network methods.

3. METHODOLOGY AND SYSTEM DESIGN



The methodology used and detailed implementation details of the proposed work sentiment analysis on product review of different users is explained in this section with detailed discussion.

a) Data Set

The data set used for this study is various product reviews collected from Amazon and Flipkart covering the details from 2015 to 2020 randomly. The data set contained a total of 2 lakh product reviews. The reviews contained product ID, rating, time and review comment. Here we focussed only on review comments.

b) Data Preprocessing

Preprocessing is the very first step and plays much important for every natural language processing application. Especially the task such as sentiment analysis and question answering deals with human's texts. It may contain wrong spelling, short words, meaningless texts, special symbols and emojis. Need to clean this kind of noisy data before processing. Preprocessing is used to convert the document texts into more meaningful forms. Here we used the pre-processing techniques – punctuation removal, text normalization, tokenization and part of speech tagging.

First of all, removed the punctuations and special symbols from the review texts. Then converted the texts into lower case letters. After that tokenized the comment texts. During this phase, the comment texts are separated into more meaningful tokens. Finally performed Part Of Speech tagging. In sentiment analysis, a Part of Speech tagger has an important role. Because the identification of the different parts of speech such as noun, verb, pronoun and adverb has an important role in sentiment analysis. In the process of sentiment analysis, the pronouns and adverbs have more roles than nouns and verbs. All the preprocessing techniques are implemented by using python with NLTK libraries.

c) Word Embedding and Vector Space Modelling

After preprocessing, the comment texts are converted as vectors for processing. Here we used the word embedding algorithm – word2vec [7] for document modelling. It is a neural network-based method used to map words from a higher dimension to vector space with a lower dimension. After processing the word2vec algorithm, the documents are converted as weighted word vectors.

d) Sentiment Analysis using BiLSTM

The weighted word vectors from the word2vec model are input to BiLSTM to identify the context information of the comment texts. BiLSTM is a combination of a Bidirectional RNN and LSTM. In Bidirectional RNN, the contexts are identified from both left to right and right to left directions. LSTM is a type of RNN, which learns the contextual meaning of words in the document. Finally, a sigmoid activation function [8] is used to find the sentiment polarity of the comments. The block diagram of proposed sentiment analysis system is shown in Figure.1

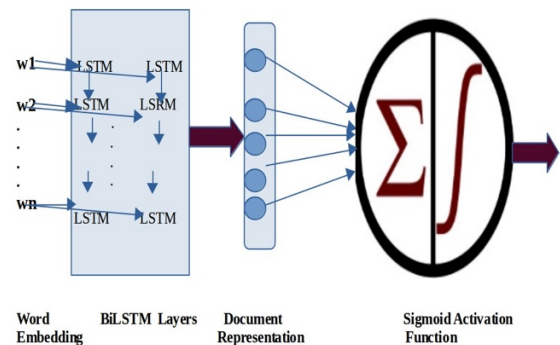


Figure 1 : Block diagram of Sentiment Analysis System

4. RESULTS AND DISCUSSION

Evaluation The standard performance metrics for classification problems were used to assess all the

models. The experiment is conducted with the review comments collected from Amazon and Flipkart. The data set is divided into the training set and test set in the ratio of 4:1.

The sample data from our data set is shown in Table 1. The sample results taken from our experiment is shown in Table 2.

Table1: Sample data taken from review data set.

No	Review
1	In this budget, it has exceeded all expectations .
2	I will never purchase Samsung phones. Phone is lagging too much and slow.
3	Excellent product.For gaming I am not sure but otherwise best laptop in this price range.
4	Not worth buying as the battery drains out within 3.30 hours of usage despite the fact that they claim 9 hours of battery usage on the official HP website.
5	Not proper working and touch pad is reduculas. very bad performance
6	The product is good, but the price is not affordable.
7	Build & Looks are awesome. Very thin & sleek. Silver color adds to the classy looks.
8	This product is best.battery life 7-8days, heart rate sensor is best quality,blood o2 sensor is best quality,16+mod,still body,gorila glass,full water resistant.
9	laptop is good and Fast but the GHZ mentioned is not correct they say 4.2 ghz but what we receive is only 2.4 Ghz. overall laptop is good in this price range.
10	poor sound quality , battery back up not good

Table 2: Sample results taken from results.

No	Review	Sentiment
1	In this budget, it has exceeded all expectations .	Positive
2	I will never purchase Samsung phones. Phone is lagging too much and slow.	Negative
3	Excellent product.For gaming I am not sure but otherwise best laptop in this price range.	Positive
4	Not worth buying as the battery drains out within 3.30 hours of usage despite the fact that they claim 9 hours of battery usage on the official HP website.	Negative
5	Not proper working and touch pad is reduculas. very bad performance	Negative
6	The product is good, but the price is not affordable.	Neutral
7	Build & Looks are awesome. Very thin & sleek. Silver color adds to the classy looks.	Positive
8	This product is best.battery life 7-8days, heart rate sensor is best quality,blood o2 sensor is best quality,16+mod,still body,gorila glass,full water resistant.	Positive
9	Laptop is good and Fast but the GHZ mentioned is not correct they say 4.2 ghz but what we receive is only 2.4 Ghz. overall laptop is good in this price range.	Positive
10	Poor sound quality , battery back up not good	Neutral

The precision and recall parameters of the proposed sentiment analysis system are evaluated as 84.2% and 85.1% respectively. It is greater than the accuracy compared with RNN and LSTM methods

5. CONCLUSION AND FUTURE WORKS

This study proposed a semantic, sentiment analysis system on product reviews. Here we identified the polarity of review texts. The neural network methods - weighted word embedding and BiLSTM are used for the implementation. The system has experimented with reviews collected from Amazon and Flipkart websites.

As per the detailed study with the new data set **the precision and recall parameters of the proposed sentiment analysis system are evaluated as 84.2% and 85.1% respectively. It is greater than the accuracy compared with RNN and LSTM methods.** The proposed method has high precision and recall values such as sentiment analysis system are evaluated as 84.2% and 85.1% respectively as per the experimental results.

In Future As now only a single method is tested, It is required to test with more algorithms or combination of more than one approaches will increase the accuracy of the system. In future, we planned to extend our work to other native languages too as well as the results can be compared with the existing studies.

REFERENCES

- [1]. Nhan Cach Dang, María N. Moreno-García, and Fernando De la Prieta. "Sentiment Analysis Based on Deep Learning:A Comparative Study". *Electronics* 2020, 9, 483; doi:10.3390/electronics9030483.
- [2]. Xing Fang and Justin Zhan."Sentiment analysis using product review data". Fang and Zhan *Journal of Big Data* (2015) 2:5.D OI 10.1186/s40537-015-0015-2. Springer.
- [3]. Aytu g Onan. "Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks". *Concurrency Computat Pract Exper.* 2020. <https://doi.org/10.1002/cpe.5909>.
- [4]. Huyen Trang Phan, Van Cuong Tran, Ngoc Thanh Nguyen And Dosam Hwang. "Improving the Performance of Sentiment Analysis of Tweets Containing Fuzzy Sentiment Using the Feature Ensemble Model". *Digital Object Identifier 10.1109/ACCESS.2019.2963702*. IEEE Access.
- [5]. Guixian Xu,Yueting Meng, Xiaoyu Qiu, Ziheng Yu, Xu Wu. " Sentiment Analysis of Comment Texts Based on BiLSTM". *Special Section On Artificial Intelligence And Cognitive Computing For Communication And Network*. *Digital Object Identifier IEEE Access 10.1109/ACCESS.2019.2909919*..
- [6]. Rajkumar S. Jagdale, Vishal S. Shirsat and Sachin N. Deshmukh. " Sentiment Analysis on Product Reviews Using Machine Learning Techniques". p. 1301-3781. *Advances in Intelligent Systems and Computing* 768, https://doi.org/10.1007/978-981-13-0617-4_61. @2019, Springer Nature Singapore Pte Ltd.
- [7]. Mikolov T, Chen K, Corrado G, Dean J. Efficient Estimation of Word Representations in Vector Space. In *Proceedings of Workshop at ICLR*, arXiv; 2013.
- [8]. Kulkarni A., Shivananda A. " Deep Learning for NLP. In: *Natural Language Processing Recipes*". Apress, Berkeley, CA. <https://doi.org/10.1007/978-1-4842-4267-4-6>
- [9]. Jain, A.P.; Dandannavar, P. "Application of machine learning techniques to sentiment analysis". *Proceedings of the 2nd International Conference on Applied and Theoretical Computing and Communication Technology (ICATecT)*, Karnataka, India, 21–23 July 2016.
- [10]. Z. Hao, R. Cai, Y. Yang, W. Wen, and L. Liang, "A dynamic conditional random field based framework for sentence-level sentiment analysis of chinese microblog," in *Proceedings of Internatioanal Conference of Computer Science and Engineering* , Guangzhou, China, Jul. 2017.
- [11]. A deep learning approach in predicting products' sentiment ratings: a comparative analysis Vimala Balakrishnan, · Zhongliang Shi, Chuan Liang Law, · Regine Lim1 , Lee Leng The, · Yue Fan.