ISSN [Online]: 2583-2654

Improved Comment Sentiment Analysis Method using Deep Learning

Revathi[1] , Anushree R [1], Sindhu S[1], Suprith B[1], and Vanishree S[1]*
[1]Department Of Computer Science and Engineering, Dr. T. Thimmaiah Institute of Technology, Karnataka, India

Abstract-- Sentiment Analysis of the comment text from the social media is helpful for understanding the public opinion on the product review. The core of sentiment analysis is the text classification task, and different words have different contributions to classification. The classification provides that the product is positive or negative based on the comment text provided by the users of the product.

Our proposed system uses the traditional TF-IDF algorithm and generates weighted word vectors. The weighted term vectors are given as input to the bidirectional long short-term memory (BiLSTM) to capture the context information effectively. The sentiment is positive or negative of the comment is obtained by feed forward neural network classifier.

The system will be tested with the comment text collected by the user product review from social media, e-commerce website, and the result shows that the product has the positive or negative reviews

***Keywords*:** Sentiment Analysis; BiLSTM; Machine Learning Algorithms; Product Reviews.

I. INTRODUCTION

In recent years, with the rapid development of the Internet and social networks, more and more users begin to freely express their opinions on web pages. Therefore, the big data of user comments is generated on the Internet. For example, the product comments are generated on E-commerce websites such as Jingdong and Taobao, and hotel comments are created on travel websites such as Ctrip and ELong.

With the explosive increasing of comments, it is difficult to analyze them manually. In the era of big data, mining the emotional tendencies of comment texts through artificial intelligence technology is helpful for timely understanding of network public opinion. The research of sentiment analysis is very meaningful for obtaining the sentiment trend of the comments.

With the advent of Web 2.0 various platforms like Facebook, Twitter, LinkedIn, Instagram permits citizens to share their comments, views, feelings, judgements on the myriad of topics ranging from education to entertainment. These platforms contain the massive amount of the data in the form of tweets, blogs, and updates on the status, posts, etc.

Sentiment Analysis aims to determine the polarity of feelings like happiness, sadness, grief, hatred, anger and affection and opinions from the text, reviews, posts which are available online on these platforms. Opinion Mining finds the sentiment of the text with reverence to a given source of content. Sentiment analysis is complex because of the slang words, misspellings, short forms, repeated characters, use of regional language and new upcoming emoticons. So it is a significant mission to identify appropriate sentiment of each word.

Sentiment analysis is a kind of text classification, involving natural language processing, machine learning, data mining, information retrieval and other research fields. Sentiment analysis of comments mainly concentrates on the sentiment orientation analysis of comment corpus, which indicates that users express positive, negative or neutral sentiments towards products or events.
The essence of sentiment analysis is the text classification task, and the contribution of different

ISSN [Online]: 2583-2654

words is different to classification. For sentiment classification tasks, learning a low-dimensional, non-sparse word vector representation for a word is a key step. The widely used word re-presentation is the distributed word vector obtained by Word2vec technology. The word vector has a low dimension and contains the semantic information of the word. However, distributed word vectors do not contain sentiment information about words.

In this paper, the contribution of the word's sentiment information to text sentiment classification is embedded into the traditional TF-IDF algorithm, and the weighted word vector is generated.

II. LITERATURE REVIEW

A literature review is an objective, survey of the research work relevant to a topic that are under consideration. Here, is the replication of the literatures presented. Its purpose is to create familiarity with current thinking and research on a particular topic. **Andreea Salinca et.al [1]** presented Business reviews Classification Using Sentiment Analysis in which the author uses the two feature extraction methods namely bag of words(BOW) and POS(part-of-speech tag) and four machine learning algorithm for the text classification of sentiment analysis as positive or negative, namely Multinomial Naïve Bayes, Support Vector Machines, Logistic Regression and Stochastic Gradient Descent classifier. They compared all the four approach for business reviews dataset by yelp and got the best classifiers as SVM and SGD have obtained an accuracy of 94.4%. In terms of performance the Naïve Bayes and Logistic Regression got slightly worst result. **Yonas Woldemariam et.al [2]** presented Sentiment Analysis in A Cross-Media Analysis Framework in which author proposed the Sentiment analysis method with Lexicon-Based Sentiment Prediction Algorithm and Recursive Neural Tensor Network (RNTN) model. The Lexicon-Based uses sentiment dictionary containing words annotated with sentiment labels and other basic lexical features, and the RNTN is trained on Sentiment Treebank with 215,154 phrases, labelled using Amazon Turk. They have used the following components chat room cleaner, NLP and sentiment analyzer. The global performance evaluation shows that RNTN outperforms the Lexicon-based by 9.88% accuracy on variable length positive, negative and neutral comments. However, the lexicon-based shows improved performance on classifying positive comments. We also notice that the F1-score values of the Lexicon-based id greater by 0.16 from the RNTN. **M.Trupthi et.al [3]** presented Sentiment analysis on Twitter using streaming API in which the author proposed to get the people thinking and feel about their products and services in Twitter platform. This work is of tremendous use to the people and industries which are based on sentiment analysis. For example, Sales Marketing, Product Marketing, etc. This paper deals with the tasks that appear in the process of Sentiment Analysis, real time tweets are considered as they are rich sources of data for opinion mining and sentiment analysis. They have used the NLTK for text processing such as tokenization, stemming and Parsing. The Classifier Naïve Bayes used for classifying the sentiment and labelling them as positive, negative or neutral. For testing they get the twitter API in the web posted tweets by the users of the product and classifies them as number of positive, negative or neutral reviews about the product. This work is of tremendous use to the people and industried which are based on sentiment analysis. For example, Sales Marketing, product Marketing etc. **Rachana Bandana et.al [4]** presented Sentiment Analysis of Movie Reviews using Heterogeneous features in which heterogeneous features such as machine learning based and lexicon-based features and machine learning algorithms like Naive Bayes and Linear Support Vector Machine (LSVM). The Heterogeneous features created using a combination of lexicon like SentiWordNet, WordNet and machine learning like Bag of Words, TF-IDF, etc. They have collected the dataset from different sources like Bookmyshow, IMDB, Rotten Tomatoes and Netflix it contains text movie reviews. The text movie reviews are given to the classifies and can predict the sentiment class label as positive and negative. By using the heterogeneous features, we get the better result rather than using only machine learning or lexicon-based features and also the naïve Bayes algorithm achieved tremendous accuracy compared to Linear SVM for these heterogeneous features.**P. Karthika et.al [5]** presented Sentiment Analysis of Social Media Network Using Random Forest Algorithm in which the rating from the online

ISSN [Online]: 2583-2654

shopping website known as Fipkart.com is analyzed collected from the Kaggle resource, based on the aspects of the product the rating is classified as positive, neutral and negative. The proposed work is evaluated by using machine learning algorithm called Random Forest and simulated by using SPYDER. By using the product review start rating as splinted into 0,1 Class denotes the negative review and Class 2,3 denotes the neutral review and finally Class 4,5 denotes the positive reviews by extracting the features like Product Id, Product Name, Brand Name and Rating. The accuracy comparison is made for the product between the Random Forest and Support Vector Machine Algorithm and the Random Forest shows the best accuracy of 97% then the Support Vector Machine. Masses of customer share their feedback (or) reviews on social media, this helps the provider to enrich their brand and as well as the customer to gain information about the product.

Poornima.A et.al [6] presented A Comparative Sentiment Analysis of Sentence Embedding Using Machine Learning Techniques in which the aim in this paper is to comparing the performance of different Machine Learning Algorithms in finishing Sentiment Analysis of Twitter data. The proposed method uses term frequency to find the polarity of the sentence. They have used three machine learning Algorithms such as Multinomial Naïve Bayes, SVM(Support Vector Machine) and Logistic Regression. The Logistic regression performs better when compared to SVM and Multinomial Naïve Bayes. Logistic regression has achieved accuracy of approximately 86% when used bigram model.

III. PROPOSED MODEL

A system architecture is the conceptual mode; that defines structure, behavior and more interpretations of a system. A system architecture can comprises of system components and the sub systems developed, that will work together to implement the overall system. The figure.1 shows the system architecture, it shows a simple yet effective algorithm to identify the polarity and categorize the reviews.

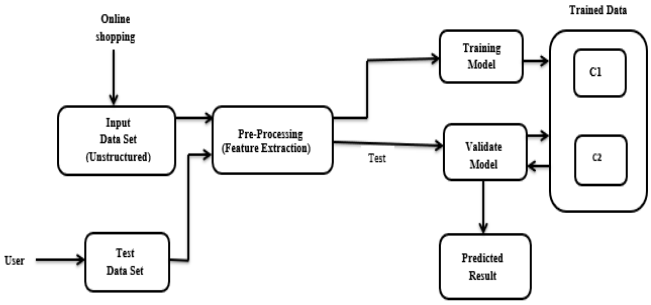


Fig 1: Architecture Diagram

STAGES IN THE PROCESS ARE

1. Data Collection

In this project methodology, the data is obtained from the E-commerce website (Amazon, Flipkart) by the product reviews given by the users. The collected user comments text may consist of unnecessary and redundant information, we have to remove this information for improving accuracy and efficiency.

2. Pre-Processing (Feature Extraction)

In this stage, the proposed system reduces number of words in each individual text command in the input text command set through the process of identifying the relevant (Subjective and Objective) and irrelevant (articles, proposition, connection) words and remove the irrelevant words based on predetermined pattern model and the respective Adjectives will be produced.In this phase the Tokenization and all the spaces from the review text are removed, and convert all capital letters to lowercase in order to reduce redundancy in the feature selection task.

The use of Bag of Words (BOW) technique as the feature extraction process. Better features can produce simpler and flexible models, and they often produce better results.

3. Training Data Set And Testing Data Set Split

In this stage, Once data are pre-processed this system has to split data into division such as training data and testing data. Usually training data should be large for accurate result. The Training

ISSN [Online]: 2583-2654

Data in Deep Learning is the actual dataset used to train the model for performing various actions. This is the actual data the ongoing development method models learn with various API and algorithm to train the machine to work automatically. The Test Data is the data typically used to provide an unbiased evaluation of the final system that are completed and fit on the training dataset. Actually, such data is used for testing the model whether it is working appropriately or not.

4. Training Stage

This stage, the training data set with label has given to one of the Deep Learning technique like BiLSTM (Bidirectional Long Short-Term Memory), this module will extract the feature from the label data and keep it ready for prediction process.

The process of training Deep Learning model involves providing an Deep Learning algorithm (that is, learning algorithms) with training data to learn from. The term Deep Learning (BiLSTM) Model refers to the model artifact that is created by the training process. The training data must contain the accurate answer, which is known as a target or target attribute. The learning algorithm discovers patterns in the training data that map the input data attributes to the target (the answer that you want to predict), and it outputs an DL model that captures these patterns.

5. Classifier Stage

In this stage, the test data without label has to be given to prediction model which is generated using training method, this prediction model accepts the test data and process. Classification is an important supervised learning method in which the computer program learns from the input given to it and then uses this learning to classify new observation. This data set may simply be bi-class (like identifying whether the comment text is Positive or Negative).

IV. DATAFLOW DIAGRAM OF PROPOSED MODEL

A data-flow diagram (DFD) is a graphical illustration of the "flow" of data through an information system. On a DFD, data items move from an external data source or an internal data store to an internal data store or an external data sink, via an internal process.

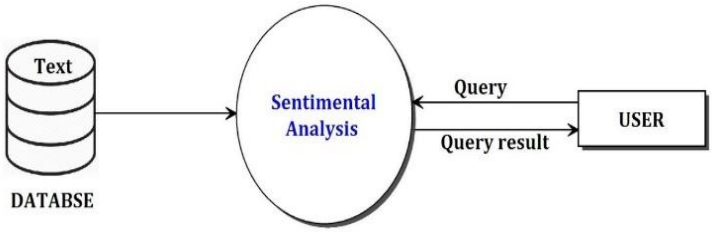


Fig 2. Level 0 Data Flow Diagram

A level 0 data flow diagram (DFD), also known as a context diagram, shows a data system as a whole and emphasizes the way it interacts with external entities. The top-level process and input, output for the top-level process is given in Fig 2.

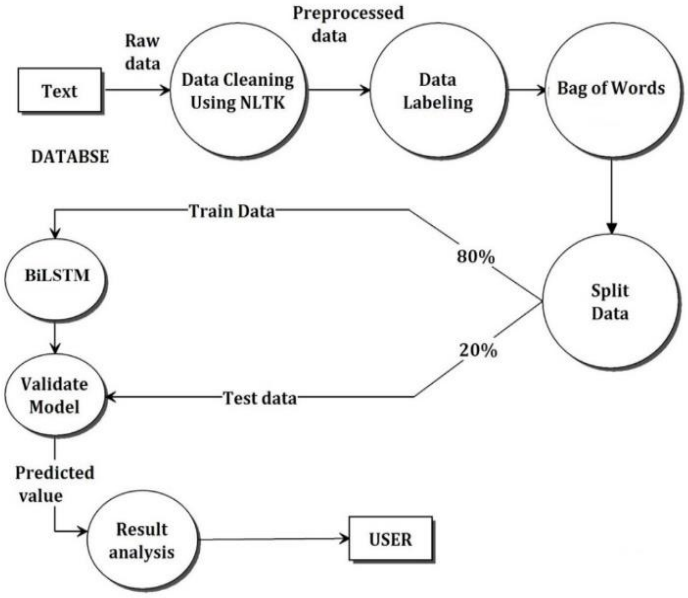
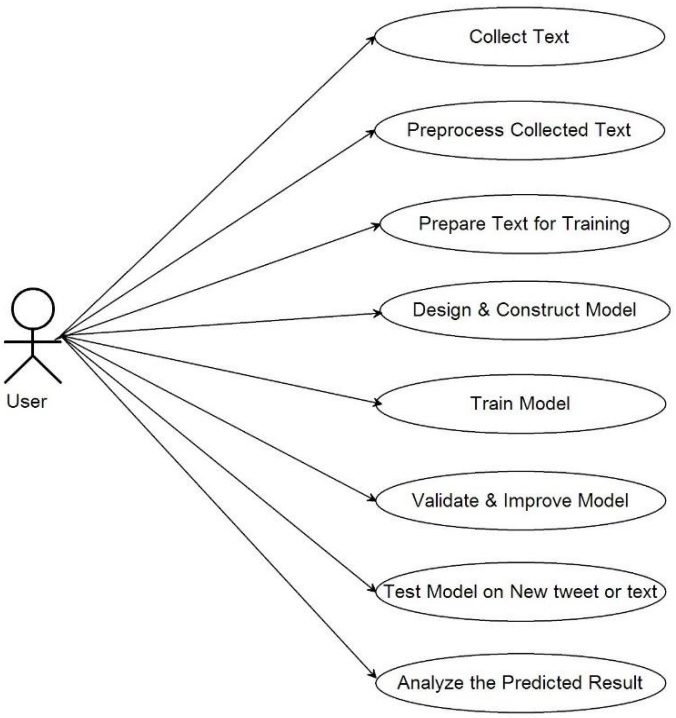


Fig 3. Level 1 Data Flow Diagram

A level 1 DFD notates each of the main sub-processes that together from the whole system. We can think of level 1 DFD as an "exploded view" of the context diagram. The process is split to sub process as given in Fig 3.

ISSN [Online]: 2583-2654

V. USE CASE DIAGRAM OF THE SYSTEM



VI RESULTS AND DISCUSSIONS

This section describes the detail of experimental result of proposed PBTC system that tested over the sample existing online user review command set. For the experimentation purpose, more than 100 unstructured consumer review text commands with different size taken from online related to product review. The best results were given by algorithm. The Bidirectional long short memory(BILSTM) achieved 85% accuracy.

VII CONCLUSION

This paper uses techniques to identify the polarity of the reviews. The algorithms performed was bidirectional long short term memory(BILSTM).Finding the polarity of the reviews can help in various domain. Intelligent systems can be developed which can provide the users with comprehensive reviews of movies, products, services etc. without requiring the user to go through individual reviews, he can directly take decisions based on the results provided by the intelligent system.

i).Sentiment analysis process is carried out to classify the highly unstructured data of product reviews into positive, negative or neutral.

ii).Online product datasets are selected as data used for this study which ultimately represents a comparatively model of discriminative classifier.

REFERENCES

[1] Andreea Salinca, "Business Reviews Classification Using Sentiment Analysis", 17th International Symposium on Symbolic and Numeric Algorithms for Scientific computing (SYNASC), pp. 247-250, Sept. 2015.

[2] Yonas Woldemariam, "Sentiment Analysis in A Cross-Media Analysis Framework", IEEE International Conference on Big Data Analysis (ICBDA), pp. 1-5, March 2016.

[3] M. Trupthi, Suresh Pabboju, G. Narasimha, "Sentiment Analysis on Twitter Using Streaming API", IEEE 7th International Advance Computing Conference (IACC), pp. 915-919, Jan. 2017.

[4] Rachana Bandana, "Sentiment Analysis of Movie Reviews Using Heterogeneous Features", 2nd International Conference on Electronics, Materials Engineering and Nano-Technology (IEMENTech), pp. 1-4, May 2018.

[5] P. Karthika, R. Murugeswari, R. Manoranjithem, "Sentiment Analysis of Social Media Network Using Random Forest Algorithm", IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), pp. 1-5, April 2019.

[6] Poornima. A, K. Sathiya Priya, "A Comparative Sentiment Analysis Of Sentence Embedding Using Machine Learning Techniques", 6th International Conference on Advanced Computing and Communication Systems (ICACCS), pp. 493-496, March 2020.

[7] Guixian Xu, Yueting Meng, Xiaoyu Qiu, Ziheng Yu, Xu Wu, "Sentiment Analysis of Comment Texts Based on BiLSTM", IEEE Access, vol. 7, pp. 51522-51532, April 2019.

[8] Suchita V. Wawrw, Sachin N. Deshmukh, "Sentimental Analysis of Movie Review using Machine Learning Algorithm with Tuned Hyperparameter", International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), Vol.4, Issue 6, pp. 12395-12402, June 2016.

[9] Yogesh Chandra, Antoreep Jana, "Sentiment Analysis using Machine Learning and Deep Learning", 7th International Conference on Computing for Sustainable Global Development (INDIACom), pp. 1-4, March 2020.

[10] Vishesh Kasturia, Shanu Sharma, Sachin Sharma, "Automatic Product Salability Prediction Using Sentiment Analysis on User Reviews", 10th International Conference on Cloud Computing, Data Science and Engineering (Confluence), pp. 102-106, Jan. 2020.

[11] Adith Shetty, Dhruv Makati, Monil Shah, Swati Nadkarni, "Online Product Grading Using Sentimental Analysis with SVM", 4th International Conference on Intelligent Computing and Control System (ICICCS), pp. 1079-1084, May 2020.

[12] Chaya Chauhan, Smriti Sehgal, "Sentiment Analysis on Product Reviews", International Conference on Computing, Communication and Automation (ICCCA), pp. 26-31, 2017.

[13] Kudakwashe Zvarevashe, Oludayo O. Olugbara, "A Framework for Sentiment Analysis with Opinion Mining of Hotel Reviews", Conference on Information Communication Technology and Society (ICTAS), pp. 1-4, March 2018.

[14] Ahlam Alrehili, Kholood Albalawi, "Sentiment Analysis of Customer Reviews Using Ensemble Method", International Conference on Computer and Information Sciences (ICCIS), pp. 1-6, April 2019.

[15] Pankaj, Prashant Pandey, Muskan, Nitasha Soni, "Sentiment Analysis on Customer Feedback Data: Amazon Product Reviews", International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), pp. 320-322, Feb. 2019.