Grenze International Journal of Engineering and Technology, Jan Issue

Sentiment Analysis for Movie Reviews on Social Media based on Sentiment Lexicon and Deep Learning

Dr. Yogesh Kumar Jakhar[1] and Dr. Nidhi Mishra[2]
[1]Assistant Professor, Department of Computer Application, University of Engineering & Management, Jaipur
Email: yogeshjakhar@gmail.com
[2]Assistant Professor, Amity University, Noida
Email: nmishra1@amity.edu

***Abstract*— Now days, with the fast improvement in internet technology and social media network, user give his/her feedback on many things and companies also asked to give feedback about their services as product or service delivered to the customers in the form of review. User's review about product or service playing major role in improving services as user pleasure. In this paper we present a sentiment analysis model, which is constructed with the help of sentiment lexicon with association of attention-based Bidirectional Gated Recurrent Unit (BiGRU) and CNN (Convolutional Neural Network). For experimental work we used movie review dataset and matched the sentiment analysis results of the proposed method with the mutual sentiment analysis models (SVM, Naïve Bayes, Decision Tree, Random Forest, CNN, and BiGRU) on the dataset. The Experiment result shows that proposed model perform better than other classifiers.**

***Index Terms*— Sentiment Analysis, Movie Review, Attention Mechanism, CNN, BiGRU.**

I. INTRODUCTION

The advent of social media platforms has facilitated the spreading of knowledge and opinions on a variety of global issues. People use the internet to exchange information, opinions, and feelings about products, events, services, and political issues. Social media communication platforms such as Twitter, Instagram, YouTube, and Facebook enable people to discuss a wide range of subjects, issues, and challenges and make them express themselves in a variety of ways, such as through text, images, and videos [1]. It is challenging to design a sentiment analysis algorithm with high accuracy due to the following issues. A key factor that seriously affects the accuracy is feature extraction. The traditional feature extraction approaches are commonly based on knowledge-based or statistical techniques [3].

In the past forty years, a multitude of methods have been developed to accomplish the Sentiment Analysis assignment, which looks for subjectivity and polarity in texts at the sentence, document, and aspect levels. Significant research on sentiment analysis, including the analysis of subjective and objective texts, the cognitive aspects of sentiments, and the development of affective lexicons, took place in the 1980s. In the 1990's, WordNet, Part of Speech (POS) Tagging, Parsing Trees based on Statistical Methods, directional interpretation (positive/negative/neutral) of a given text, predicting the semantic orientation of adjectives, and Fuzzy Model were introduced for data mining and used for sentiment analysis [5]. Sentiment categorization is used to identify the positive or negative undertones in a given content. Sentiment classification is widely utilized in a wide range of corporate fields to understand consumer attitudes and improve products or services. Deep learning produces

Grenze ID: 01.GIJET.10.1.20_2

cutting-edge outcomes in a number of difficult disciplines. With the success of deep learning, many studies have proposed deep-learning-based sentiment classification models and achieved better performances compared with conventional machine learning models [8].

The reminder of the paper is structured as follows. Section II gives a momentary summary of associated research work. Methodology is given in Section III. Section IV represent experimental results and section V give conclusion and some future directions of the research.

II. Related Work

In this section addresses recent researches in the sentimental analysis domain that have been conducted by a variety of researchers in recent years. Urooba Sehar, Summrina Kanwal, Kia Dashtipur and et al. outlined a brand-new multimodal sentiment analysis (MSA) framework that uses text, visual, and audio responses to identify sentiments that are aware of context. The experimental results verified that integration of multimodal features increased the polarity detection capability of the proposed algorithm from 84.32% to 95.35% [1]. Zenun Kastrati, Ali Shariq Imran and Arianit Kurti proposed a structure that takes the benefit of weakly supervised annotation of MOOC-related features and promulgates the weak supervision signal to successfully identify the aspect categories discussed in the unlabeled students' evaluations. The experimental results indicated that proposed framework attained inspiring performance with respect to both the aspect category identification and the aspect sentiment classification [2].

Tong GU, Guoliang XU and D Jiangtao LUO designed a novel sentiment analysis model named MBGCV. The proposed model MBGCV employs a multichannel paradigm and integrates Bidirectional Gated Recurrent Unit (BiGRU), CNN and VIB (Variational Information Bottleneck). The experimental results demonstrated that proposed model can achieve superior accuracy and improve the performance of text sentiment analysis [3]. ZHI LI, RUI LI and Guanghao JIN constructed a danmaku sentiment dictionary and presented a novel technique using sentiment dictionary and Naïve Bayes for the sentiment analysis of danmaku reviews. Experimental results demonstrated that the proposed method has a significant effect on sentiment score and polarity detection [4].

Fuad Alattar and Khaled Shaalan presented a Filtered-LDA framework that significantly overtook existing methods of interpreting sentiment variations on Twitter. The proposed framework implements cascaded LDA Models with multiple settings of hyper-parameters to capture candidate reasons that cause sentiment variations. Then it applies a filter to remove tweets that discuss old topics, followed by a Topic Model with a high Coherence Score to extract Emerging Topics that are interpretable by a human [5]. Ryosuke Harakawa, Takahiro Ogawa and Miki Haseyama proposed a new technique based on sentiment-aware signed network analysis that extracts the hierarchical structure of Web video groups to enable Web video retrieval. Initially, the suggested approach uses sentiment features gleaned from texts attached to Web videos and multimodal features of contents to estimate latent links between Web videos. The Results of experiment show that using a new benchmark dataset, YouTube-8M, have proven the hypothesis that sentiment in multimedia contents has an influence on their topics and have confirmed the improvement in performance by extracting the hierarchical structure for Web video retrieval [6]. Mohd Usama, Wenjing Xiao, Belal Ahmad and et al., proposed an approach to taking multiple neural networks' multilevel and multiple features and combining them. While multiple features come from various network architectures, multilevel features come from various layers within the same network. Experiment results demonstrate that the proposed model based on multilevel and multitype weighted features fusion outperforms than many exiting works with an accuracy of 80.2%, 48.2%, and 87.0% on MR, SST1, and SST2 datasets respectively [7]. Seungwan Seo, Czangyeob Kim, Haedong Kim and et al., presented a proportional study of different deep-learning-based sentiment classification model structures to derive meaningful suggestions for constructing sentiment classification models. Specifically, eight deep learning models, three based on convolutional neural networks and five based on recurrent neural networks [8].

Avinash Kumar, Vishnu Teja Narapareddy and et al., proposed a novel interactive gated convolutional network (IGCN) that uses a bidirectional gating mechanism to learn mutual relation between the target and corresponding review context. The experimental results on SemEval 2014 datasets show the effectiveness of the proposed IGCN model [9]. Waweru Mwangi, James Mutinda and George Okeyo presented a model using sentiment lexicon, N-grams, and BERT vectorise words selected from a section of the input text and presented model was evaluated on Amazon products' reviews, Imbd movies' reviews, and Yelp restaurants' reviews datasets. Proposed model was performed well on these dataset with F-measure score of 88.73% in binary sentiment classification [11]. Kifayat Ullah, Anwar Rashad and et al. proposed a deep neural network which consist of seven layers for movie reviews' sentiment analysis. Proposed model was applied of two different datasets with sample records of 25,000 and 50,000. Presented model was achieved 92% accuracy in binary classification [12].

III. METHODOLOGY

To boost the accurateness of sentiment analysis on movie reviews, we joined the benefits of sentiment lexicon, CNN model, GRU model and attention mechanism to get more accurate result with the help of new model. The model consist of six layers. First layer is embedded layer and it is followed by convolutional layer, the third layer is pooled layer and it is followed by a BiGRU layer, fifth layer is the attention layer, and sixth and last layer is fully connected layer. The configuration of proposed model is shown in following Figure-1.

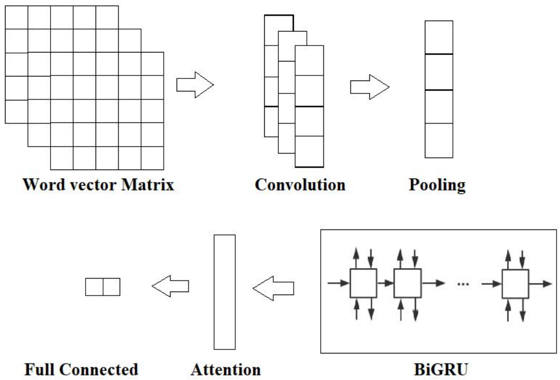

Figure 1. Model structure

Here the input text (Movie review) statement is R = $\{w_1, w_2, w_3, w_4, \ldots w_j \ldots w_n\}$, where w_n represent a word in R, and the main task of proposed model is to forecast the sentiment polarity P of the statement R. The task of the sentiment lexicon is to provide each word *w* in R a conforming sentimental weight Rw_n.

Embedded layer is used to display the text statement R as a weighted word vector matrix. In this model used the weighted word vector matrix as the result of the embedded layer or output of this layer. The second layer or convolution layer is used to extract the most significant local features of the input matrix. The primary features are extracted by the pooling layer by wrapping the text features that the convolutional layer acquired. The input matrix's context features are extracted using the BiGRU layer. Sequence information is often processed using the GRU model, which is a variation of the recurrent neural network model. It retrieves the context features from the sequence data and combines it with the historical data from the previous instant to affect the current output. Different words in a sentence are given different weights using the attention layer. Utilizing a fully connected layer, the input feature matrix is classified.

A. Dataset and Data Preprocessing

For experimental work we used movie review dataset. The dataset is downloaded from https://www.kaggle.com. The dataset in .CSV file with many parameters. The movie reviews in the original dataset are divided into five levels. Audience have given one to five stars in the rating column. To perform experiments, we divide the five levels into two main categories. One (1) and two (2) stars are treated as negative review and three(3) to five(5) stars are treated as positive review. The dataset includes 40,000 reviews, of which 19,981 are positive and 20,019 are negative.

For data preprocessing process many packages are available in python, in this experiment we used python word segmentation tool jieba package to perform word segmentation on the comment data in the dataset. We made Make text in lowercase, Remove punctuation, Remove emoji's and replace with suitable word, Remove stop words and Lemmatization. After perform all the steps of data preprocessing counted the length of the longest word included in each review, the number of distinct words in the dataset, the number of occurrences of each word, and the length of the word contained in each review in the computed dataset.

B. Performance Measurement

To compute the effectiveness of algorithms including proposed method, we are going to use various catalogs, like accuracy, specificity, sensitivity, and precision to measure the models' strength. These indices (equation 1-4) are deliberate by the confusion matrix. The calculation parameters are defined as follows [10]:

(1) *True Positive (TP):* the number of comments categorizing positive movie reviews comments as positive.
(2) *False Positive (FP):* the total number of comments that classify negative movie reviews comments as positive.

(3) *True Negative (TN):* the total number of negative comments classified as negative comments.
(4) *False Negative (FN):* the total number of comments categorizing positive movie reviews as negative.
(5) *Accuracy*: the ratio of correctly predicted comments to the total comments.

$$accuracy = \frac{\#\, of\ correctly\ classified\ documents}{total\ number\ of\ documents} \quad (1)$$

(6) *Recall*: the ratio of properly predicted positive comments to the all comments in actual class.

$$recall = \frac{true\ positive}{true\ positive + false\ negative} \quad (2)$$

(7) *Precision*: the ratio of properly predicted positive comments to the total predicted positive comments.

$$precision = \frac{true\ positive}{true\ positive + false\ positive} \quad (3)$$

(8) *F-measure*: the weighted average of precision and recall.

$$f - measure = \frac{2 \times recall \times precision}{recall + precision} \quad (4)$$

IV. EXPERIMENTAL RESULTS

Here we evaluate the performance of proposed model for sentimental analysis on movie review. In Table-1 we have shown the parameters list, which is going to use for experiment work. We split the dataset using 5*2 cross-validation and 10-fold cross-validation to more accurately assess the performance of the suggested model. Using a random division of the data set into ten parts, in which nine are use as the training set and one as the verification set in 10-fold cross-validation method. The evaluation result of our model is determined by taking the mean of these ten results. Using a random division of the dataset, we used one half as the training set and the other as the test set in the 5*2 cross-validation method. The evaluation result of our model was determined by taking the average of the five results. The experimental results of the suggested model under 10--fold cross-validation and 5*2 fold cross-validation are shown in Table-II. We will input the length of the text statement to a specific value when we input the model because the length of the text statement varies in the dataset. For our experiments, we take the longest sentence possible and the longest sentence on average from the dataset. Table-III displays the results of the experiment. We discover that for sentences longer than the average sentence length, using the average sentence length as the fixed length of the input sentence causes the loss of a portion of the context feature, which in turn impacts the model's performance.
In the the experiment, we exposed that the model's performance is somewhat influenced by the thesaurus's word count. We begin with 20,000 words in the thesaurus, and we reduce the frequency of occurrence of the words from the lowest frequency words in the thesaurus. This process is repeated every 2000 words. The experimental outcomes are displayed in Figure 2 and Table-IV. The table indicates that the model operates at its best when there are between 12,000 and 20,000 words in the lexicon. The performance of the model declines as the thesaurus's word count grows or shrinks.

TABLE I. PARAMETER DESCRIPTION AND CORRESPONDING VALUES OF THE PROPOSED MODEL

Parameters	Value
The Length of the movie review	324
The Dimension of the word vector	364
The Thesaurus Size	18000
The size of the convolution kernel	3x3
Hidden neurons in convolution layer	128
Hidden neurons in BiGRU layer	128
Dropout	0.4

TABLE II. 10-FOLD & 5*2 CROSS VALIDATION RESULTS ON THE PROPOSED MODEL

Method	Accuracy	Precision	Recall	F1
10-fold cross Validation	91.7%	90.6%	92.4%	92.0%
5*2 cross Validation	92.1%	91.3%	91.8%	92.4%

TABLE III. THE INFLUENCE OF THE THESAURUS SIZE ON THE PROPOSED MODEL

The number of words	Accuracy	Precision	Recall	F1
20000	92.7%	90.6%	92.4%	93.2%
18000	92.3%	91.3%	93.8%	93.0%
16000	92.5%	91.1%	94.1%	93.1%
14000	92.6%	91.2%	93.3%	92.7%
12000	92.7%	90.9%	93.7%	92.9%
10000	92.2%	90.8%	93.8%	93.1%

TABLE IV. THE INFLUENCE OF THE DROPOUT ON THE PROPOSED MODEL

Dropout	Accuracy	Precision	Recall	F1
0.2	93.7%	91.6%	93.4%	93.2%
0.3	93.3%	92.3%	93.3%	93.3%
0.4	93.2%	92.1%	93.1%	93.1%
0.5	92.9%	91.9%	93.0%	92.9%

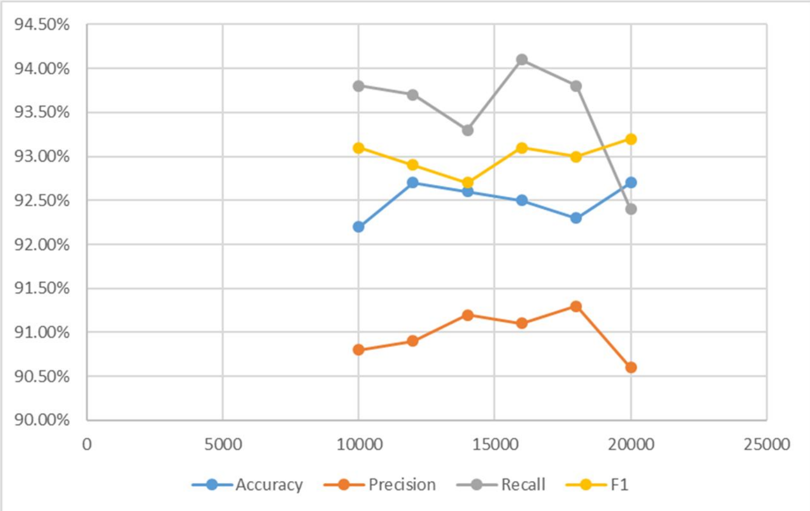

Figure 2. The influence of the thesaurus size on the proposed model.

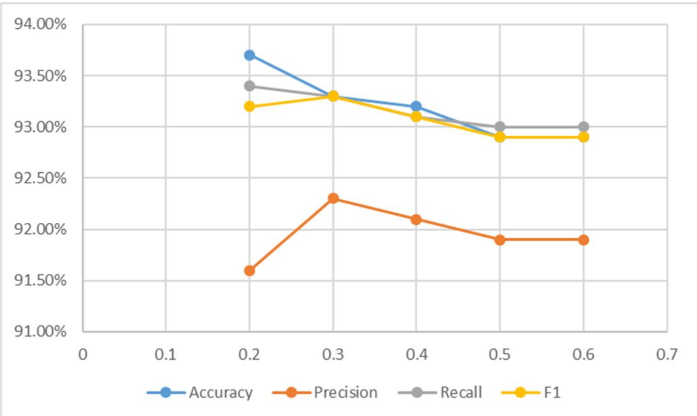

Figure 3. The impact of the dropout value on the proposed model.

We compared the sentiment analysis effects of the proposed model with the common sentiment analysis models (SVM, Naïve Bayes, Decision Tree, Random Forest, CNN, and BiGRU) on the dataset. The comparison results are shown in Table-5 and Figur-4. The experimental outcomes represents that the classification performance of the deep learning model (CNN and BiGRU) is expressively improved wit compared with the machine learning model (SVM, Naïve Bayes, Decision Tree and Random Forest). The outcomes also represents that the attention mechanism based on the deep learning model can increase the classification performance of the proposed idea.

TABLE-V: COMPARISON OF PERFORMANCE BETWEEN PROPOSED MODEL AND OTHER SENTIMENTAL ANALYSIS MODELS

Model	Accuracy	Precision	Recall	F1
SVM	83.7%	90.6%	92.4%	91.2%
Naïve Bayes	65.3%	88.3%	60.7%	58.3%
Decision Tree	84.2%	86.3%	85.1%	82.1%
Random Forest	89.9%	87.7%	90.0%	90.9%
CNN	91.2%	89.7%	90.4%	91.2%
BiGRU	92.2%	91.7%	90.8%	91.8%
Proposed model	93.9%	92.7%	91.0%	92.9%

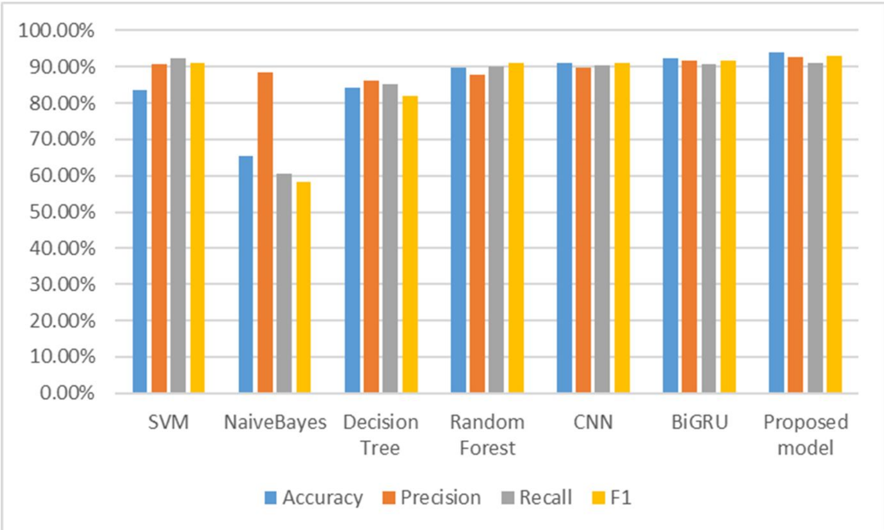

Figure 4. Model performance comparison

V. CONCLUSION

Through the fast improvement of internet technology and social media network, the sentiment analysis knowledge of product reviews has expanded more and more consideration. User's review about product or service playing major role in improving services as user fulfillment. In this paper we present a sentiment analysis model using movie review dataset, which is constructed with the help of sentiment lexicon and combines CNN and attention-based Bidirectional Gated Recurrent Unit (BiGRU).

The suggested model's performance is compared with other model like SVM, Random Forest, Naïve Bayes, Decision Tree, CNN, and BiGRU. The proposed model execute superior in comparison of the other models with highest accuracy of 93.9%. The method anticipated in this paper can only applied on movie dataset to categorized positive and negative review of the peoples. Proposed approach can be apply with other type of sentimental analysis with combining other model as an ensemble to get better performance.

REFERENCES

[1] Urooba Sehar, Summrina Kanwal, Kia Dashtipur and et al., "Urdu Sentiment Analysis via Multimodal Data Mining Based on Deep Learning Algorithms", IEEE Access, 2021, VOLUME 9, pp. 153072- 153082.

[2] Zenun Kastrati, Ali Shariq Imran and Arianit Kurti, "Weakly Supervised Framework for Aspect-Based Sentiment Analysis on Students' Reviews of MOOCs", IEEE Access, 2020, VOLUME 8, pp. 106799- 106810.

[3] TONG GU, GUOLIANG XU AND JIANGTAO LUO, "Sentiment Analysis via Deep Multichannel Neural Networks With variational Information Bottleneck", IEEE Access, 2020, VOLUME 8, pp. 121014- 121021.

[4] ZHI LI, RUI LI and GUANGHAO JIN, “Sentiment Analysis of Danmaku Videos Based on Naïve Bayes and Sentiment Dictionary”, IEEE Access, 2020, VOLUME 8, pp. 75073- 75074.
[5] Fuad Alattar and Khaled Shaalan, “Using Artificial Intelligence to Understand What Causes Sentiment Changes on Social Media”, IEEE Access, 2020, VOLUME 9, pp. 61756- 61767.
[6] Ryosuke Harakawa, Takahiro Ogawa and Miki Haseyama, “Extracting Hierarchical Structure of Web Video Groups Based on Sentiment-Aware Signed Network Analysis”, IEEE Access, 2017, VOLUME 5, pp. 16963- 16973.
[7] Mohd Usama, Wenjing Xiao, Belal Ahmad and et al., “Deep Learning Based Weighted Feature Fusion Approach for Sentiment Analysis”, IEEE Access, 2019, VOLUME 7, pp. 140252- 140260.
[8] Seungwan Seo, Czangyeob Kim, Haedong Kim and et al., “Comparative Study of Deep Learning-Based Sentiment Classification”, IEEE Access, 2020, VOLUME 8, pp. 6861- 6875.
[9] Avinash Kumar, Vishnu Teja Narapareddy and et al., “Aspect-Based Sentiment Classification Using Interactive Gated Convolutional Network”, IEEE Access, 2020, VOLUME 8, pp. 22445- 22453.
[10] Yogesh Kumar Jakhar, Nidhi Mishra and Rakesh Poonia, “Performance Analysis of Data Mining Algorithms in the Prediction of Rain Fall”, International Journal of Recent Technology and Engineering (IJRTE), 2019, ISSN: 2277-3878, Volume-8 Issue-2, pp. 558-563
[11] Waweru Mwangi, James Mutinda and George Okeyo, “Sentiment Analysis of Text Reviews Using Lexicon-Enhanced Bert Embedding (LeBERT) Model with Convolutional Neural Network”, Applied Sciences MDPI, 2023, 13, pp. 1-14.
[12] Kifayat Ullah, Anwar Rashad and et al., “ADeep Neural Network-Based Approach for Sentiment Analysis of Movie Reviews”, Hindawi Complexity, Volume 2022, Article ID 5217491.