Sentiment Analysis on Movie Reviews

[1]Ms. Khanagavalle G R, [2]Ms. Kamalishree A, [3]Ms. Kavya Rajendran, [4]Ms. Priya Shrinithi R

[1]Assistant Professor, [2,3,4]Students
Department of Computer Science and Engineering,
Sri Venkateswara College of Engineering, Sriperumbudur Tk., Tamilnadu, India.

***Abstract*: Sentiment analysis is a kind of opinion mining technique that focuses on the extraction of the emotions of people towards a specific topic using structured or unstructured data. It is one of the most vital research areas in Natural Language Processing and Text Mining in recent times. It has a wide range of applications since most of the activities happening today revolve around the opinions and feedback of people. Handling such kind of immeasurable data manually is highly impossible and there everyone finds the necessity of sentiment analysis using a deep learning algorithm. In today's world, where numerous movies of different genres are released every day, people simply cannot afford to spend their time trying to figure out whether to go to a particular movie or not. An efficient deep learning model built using proper methodology would assist people in classifying a movie as good or bad based on the reviews it receives. To implement so, two deep learning algorithms named Recurrent Neural Networks and a variant of RNN named Long Short Term Memory are used, their performance is compared and the algorithm which gives better accuracy is determined.**

***Index Terms*: Natural Language Processing, Deep Learning, Recurrent Neural Network, Long Short Term Memory.**

I. Introduction

Natural Language Processing (NLP) is a branch of artificial intelligence where computers are able to understand, interpret and manipulate human language. NLP combines computational linguistics—rule-based modeling of human language—with statistical, machine learning, and deep learning models. Together, these technologies enable computers to process human language in the form of text or voice data and derive useful insights from those data. Several tasks like speech recognition, parts of speech tagging, name entity recognition, word sense disambiguation, sentiment analysis are being performed using NLP. It plays a key role in various real time applications like digital assistants, speech-to-text dictation software, customer service chatbots and so on.

Sentiment analysis can be referred to as a subfield of Natural Language Processing and a technique using which high-quality information can be acquired using people's opinion on a specific topic. This is performed on a dataset that contains data about a particular topic. After cleaning the dataset using various preprocessing methods, the training dataset is fed into the chosen algorithm for training. After training the model, it is tested using the test set to determine its accuracy. Sentiment Analysis has been widely used in a lot of real time applications like product analysis, social media monitoring, customer support,customer feedback, brand monitoring and so on. Sentiment Analysis can be performed using machine learning-based approach and lexicon-based approach in which the former approach trains a text classifier on a human labeled dataset and the latter involves calculating orientation for a document.

Recurrent neural networks (RNN) are a class of neural networks that are helpful in modeling sequence data. Derived from feed forward networks, RNNs exhibit similar behavior to how human brains function. Recurrent neural networks produce predictive results in sequential data that other algorithms cannot. In order to understand RNN, it is important to know about the working of normal feed-forward neural networks and sequential data. In a feed-forward neural network, the information only moves in one direction — from the input layer, through the hidden layers, to the output layer. The information moves straight through the network and never touches a node twice. Feed-forward neural networks have no memory of the input they receive and are bad at predicting what is coming next. Because a feed-forward network only considers the current input, it has no notion of order in time. It simply can't remember anything about what happened in the past except its training. In RNN, the information cycles through a loop. When it makes a decision, it considers the current input and also what it has learned from the inputs it received previously.

II. Literature Survey

1. Sentiment Analysis On Comment Texts Based On BiLSTM

Authors: Guixian Xu, Yueting Meng, Xiaoyu Qiu, Ziheng Yu, and Xu Wu

Findings

Sentiment Analysis was performed on comment text using an improved word representation method which integrated the contribution of sentiment information into the traditional TF-IDF algorithm and generated weighted word vectors. The reason for using an improved word representation method was that, the most widely used distributed word representation considers only the semantic information of the word and ignores its sentiment information. In order to mitigate this issue, the contribution of sentiment information was integrated with the TF-IDF algorithm. The weighted word vectors were input into BiLSTM (Bidirectional Long Short Term Memory) to capture the context information effectively, and the comment vectors were better represented. The experimental results showed that the proposed sentiment analysis method has higher precision, recall and F1 score.

2. Sentiment Classification Using A Single Layered BiLSTM Model

Authors: Zabit Hameed , (Member, IEEE), and Begonya Garcia-Zapirain, (Member,IEEE)

Findings

The ultimate aim of the project was to present a computationally perfect deep learning model for binary sentiment classification, which aimed to decide the sentiment polarity of people's opinions, attitudes, and emotions expressed in written text. In order to accomplish this, three widely used datasets about movie reviews were used. The project used merely one bidirectional long short-term memory (BiLSTM) layer along with a global pooling mechanism and achieved an accuracy of 80.500%, 85.780%, and 90.585% on MR, SST2 and IMDb datasets, respectively. The performance metrics of the proposed approach were competitive with the recently published models, having comparatively complex architectures. It was also inferred that the proposed single-layered BiLSTM based architecture is computationally efficient and can be recommended for real-time applications in the field of sentiment analysis.

The model consists of an input layer, an embedding layer, a BiLSTM layer followed by the ensemble of global average and global maximum pooling layers,and one sigmoid layer at the output. The input layer carries data samples as a sequence of unique indices of the same length. In the embedding layer, every index corresponding to a unique word is transformed into real-valued feature vectors. After this, the feature vectors are input into the BiLSTM layer where the input is propagated in forward and backward direction. The outputs of the BiLSTM layer are simultaneously passed to the global maximum and global average pooling layers and are merged into a single layer after passing through the concatenate layer. After all this process, the output is acquired in the output layer.

3. A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis

Authors: Anwar Ur Rehman & Ahmad Kamran Malik & Basit Raza &Waqar Ali

Findings

The project proposed a hybrid model using LSTM (Long Short Term Memory) and a very deep CNN (Convolutional Neural Networks) model named Hybrid CNN-LSTM Model to overcome the sentiment analysis problem. The reason for using a hybrid model using LSTM and CNN was that Long Short-Term Memory (LSTM) model and Convolutional Neural Network (CNN) model have been applied to different Natural Language Processing (NLP) tasks with remarkable and effective results. The CNN model efficiently extracts higher level features using convolutional layers and max-pooling layers. The LSTM model was capable of capturing long-term dependencies between word sequences. The proposed model also used dropout technology, normalization and a rectified linear unit for accuracy improvement. The proposed Hybrid CNN-LSTM Model outperformed traditional deep learning and machine learning techniques in terms of precision, recall, f-measure, and accuracy.

Firstly, the word to vector (Word2Vc) approach was used to train initial word embedding. The Word2Vc translates the text strings into a vector of numeric values, computes distance between words, and makes groups of similar words based on their meanings. After word embedding is performed in which the proposed model combines a set of features that are extracted by convolution and global max-pooling layers with long term dependencies.

.

4. Halal Products On Twitter: Data Extraction And Sentiment Analysis Using Stack Of Deep Learning Algorithms

Authors: Ali Feizollah, Sulaiman Ainin, Nor Badrul Anuar, Nor Aniza Binti Abdulla,and Mohamad Hazim

Findings

The main motive of the project was to collect and extract social media posts and perform sentiment analysis using two approaches. First approach was dictionary based, where each value has a numerical value as polarity. The second approach was machine learning based where word embedding is incorporated. The project focused on tweets of two halal products, i.e., halal tourism and halal cosmetics. Twitter data (over a 10-year span) were extracted using the Twitter search function and an algorithm was used to filter the data. Then, an experiment was conducted to calculate and analyze the tweets' sentiment using deep learning algorithms. In addition, convolutional neural networks (CNN), long short-term memory (LSTM), and recurrent neural networks (RNN) were utilized to improve the accuracy and construct prediction models. The various feature extraction techniques used in the project are Word2seq and Word2vec. The datasets used for model training are eRezeki, IMDB, Amazon Product Reviews and Yelp.

The data goes through preprocessing and vectorization in order to be prepared for the algorithms. The project used stacking of the deep learning algorithms in the form of layers. Different algorithms are stacked up, and results of one algorithm are passed to another one. In this way, the weakness of one algorithm is compensated by the strength of the other algorithm. Also, an extensive collection of data was used to train the algorithms.

5. Improving Sentiment Polarity Detection Through Target Identification

Authors: Mohammad Ehsan Basiri, Member, IEEE, Moloud Abdar, ArmanKabiri,Shahla Nemati, Member, IEEE, Xujuan Zhou, Forough Allahbakhshi, and Neil Y. Yen, Member, IEEE

Findings

The main focus of the project was to extract the expressions and terms carrying the sentiment of the reviews called potential terms. In addition to extracting potential terms from sentences, the main target of each review is determined, with the help of which the final polarity of the review is estimated. After detecting the main target, five policies, including most occurring first (MOF), most general first (MGF), most specific first (MSF), first occurring first (FOF), and last occurring first (LOF), are proposed to come up with the main target of the review. Finally, using the part-of-speech (POS) tags, potential terms in the sentences are specified and a comprehensive sentiment lexicon is employed to compute the polarity of the sentences. In order to evaluate the proposed method,three data sets of user reviews about different topics, including digital equipment,hotels, and movies, are created as no previous study addressed the problem of target identification in the Persian language.

The first step, preprocessing, contains tokenization, normalization, and POS tagging phases that are applied on sentences of reviews sequentially. The next step is named "potential terms extraction," which is responsible for extracting potential terms from the sentences. These terms usually contain adjectives, adverbs,and negations. The third step called "target identification" is to specify the main target of the whole review after determining the targets of each sentence. The final step is the classification phase that is in charge of calculating the final polarity of the review.

6. Tree-Structured Regional CNN-LSTM Model for Dimensional Sentiment Analysis

Authors: Jin Wang, Liang-Chih Yu, Member, IEEE, K. Robert Lai, and XuejieZhang

Findings

To extract task-relevant phrases for structured information and to predict the Valence-Arousal (VA) ratings of texts using the Tree-Structured Regional CNN-LSTM model. The project proposes a tree-structured regional CNN-LSTM model consisting of two parts: regional CNN and LSTM to predict the VA ratings of texts.The model can capture both local (regional) information within sentences and long-distance dependencies across regions. The proposed regional CNN uses a part of the text as a region, dividing an input text into several regions such that the useful affective information in each region can be extracted and weighted according to their contribution to the VA prediction. Such regional information is sequentially integrated across regions using LSTM for VA prediction.

III. METHODOLOGY

Sentiment analysis on movie reviews is done to classify the given movie review as positive or negative. Deep learning algorithms namely Recurrent Neural Networks that deals with sequential data and is a generalization of feed forward neural networks. The performance of the algorithms with different layers is evaluated and the algorithm which comes out better in terms of accuracy is determined. This could particularly be useful when a person has to determine whether to go to a particular movie or not without having to read all the reviews for that movie. The dataset used for the project is IMDB dataset that consists of 50000 movie reviews labeled as positive and negative.

The dataset is first subject to preprocessing where the non-word characters and stopwords are removed, all letters are changed to lowercase, and lemmatization is performed to reduce words to its root. After preprocessing,vectorization is done, where the document gets transformed into vector representations. This is done using Word2Vc. After this step, the preprocessed data is fed into three kinds of RNN and a single layer LSTM algorithm.

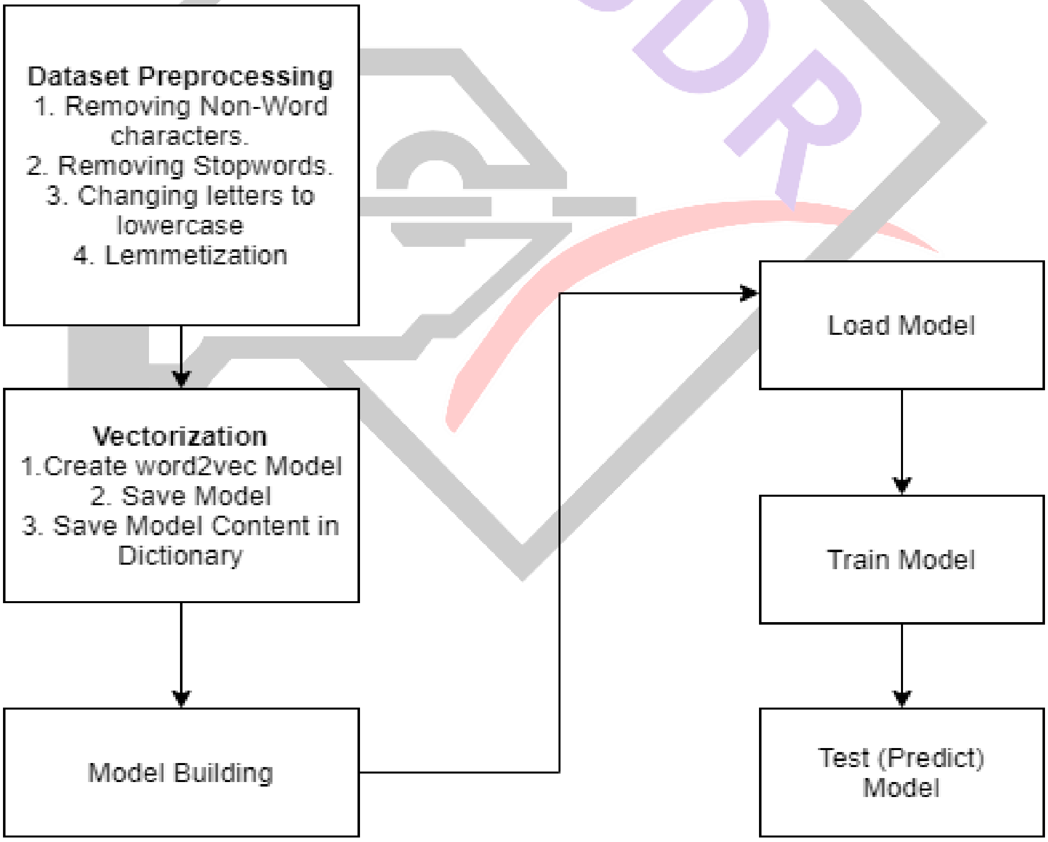


Fig.1 Proposed Work – Detailed Architecture Diagram

The dataset is preprocessed in the initial stage where the non-word characters, stopwords are removed and lemmatization is performed to reduce the words to its root. After this step, vectorization is carried out, where the text data is converted to vectors. Followed by that, the respective model is built,loaded and trained. After training the model, it is subject to testing to observe how accurately the model works..

IV. Results

Dataset

We have used IMDB Dataset for our project that consists of 50000 movie reviews in which 25000 movie reviews are classified as positive and the remaining 25000 movie reviews are classified as negative. During Exploratory Data Analysis of the dataset, it was found that, out of 50000 reviews 418 were duplicate rows. After dropping the duplicate rows, the number of positive reviews is 24884 and the number of negative reviews is 24698. In addition, the dataset was cleaned by removing the special characters,stopwords, punctuations and all the letters were made lowercase.

Experiments

1. Single Layered RNN

Batch Size = 64, Epoch = 50, Accuracy = 71.584

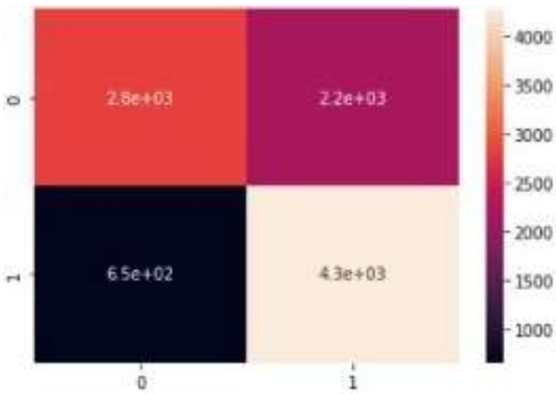


Fig.2 Confusion Matrix – Single Layered RNN

Report:

	precision	recall	f1-score	support
0	0.81	0.57	0.67	4987
1	0.66	0.87	0.75	4930
accuracy			0.72	9917
macro avg	0.74	0.72	0.71	9917
weighted avg	0.74	0.72	0.71	9917

Fig.3 Classification Report - Single Layered RNN

2. Two Layered RNN

Batch Size = 64, Epoch = 50, Accuracy = 74.246

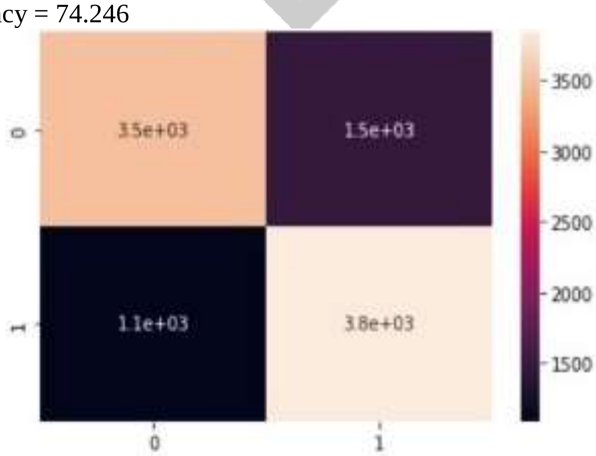
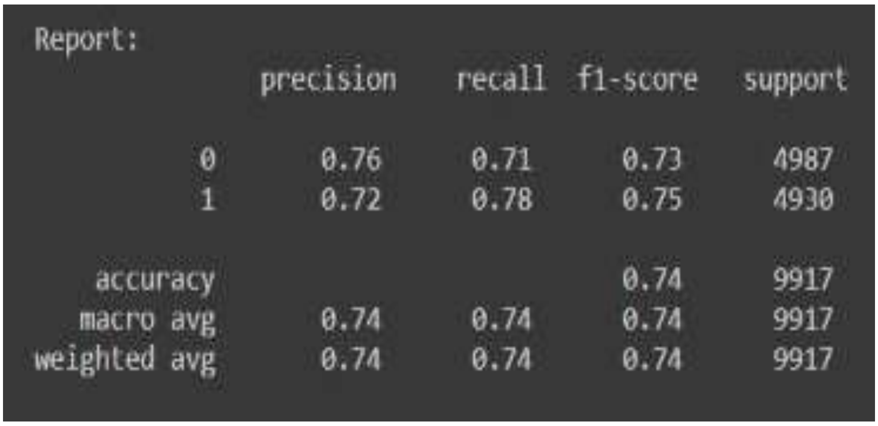


Fig.4 Confusion Matrix – Double Layered RNN



Report:

	precision	recall	f1-score	support
0	0.76	0.71	0.73	4987
1	0.72	0.78	0.75	4930
accuracy			0.74	9917
macro avg	0.74	0.74	0.74	9917
weighted avg	0.74	0.74	0.74	9917

Fig.5 Classification Report - Double Layered RNN

3. Two Layered RNN with different number of neurons

Batch Size = 64, Epoch = 50, Accuracy = 59.634

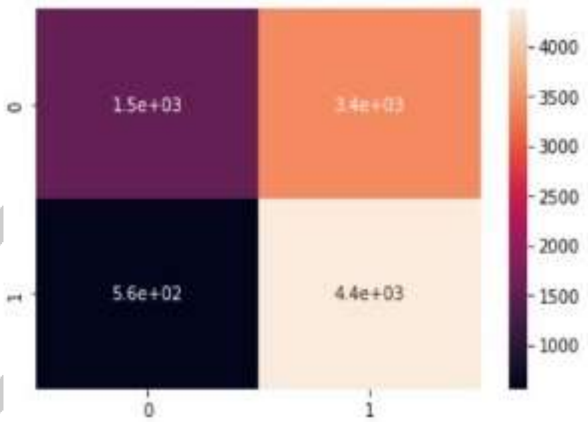
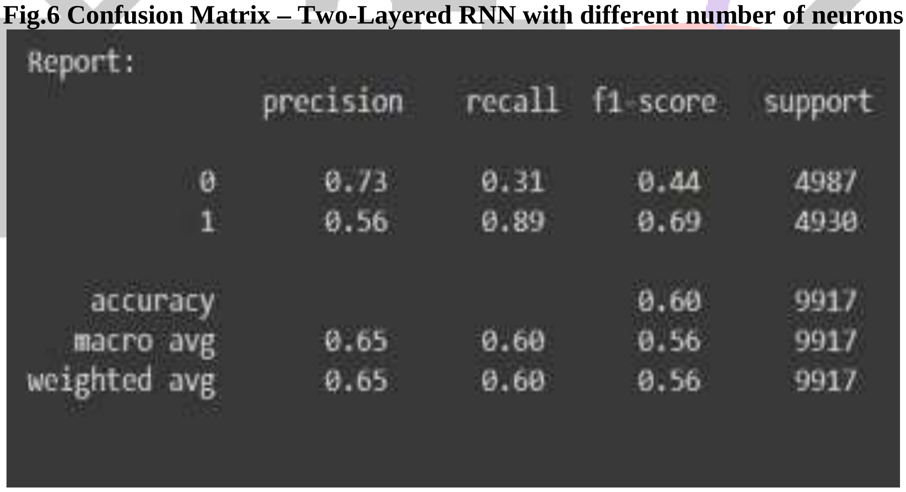


Fig.6 Confusion Matrix – Two-Layered RNN with different number of neurons



Report:

	precision	recall	f1-score	support
0	0.73	0.31	0.44	4987
1	0.56	0.89	0.69	4930
accuracy			0.60	9917
macro avg	0.65	0.60	0.56	9917
weighted avg	0.65	0.60	0.56	9917

Fig.7 Classification Report - Two-Layered RNN with different number of neurons

Accuracy Report of Models

Table 1 Accuracies of Models

S.No	*Model*	*Accuracy*
1	Single Layered RNN	71.584
2	Two Layered RNN	74.246
3	Two Layered RNN with Different number of Neurons	59.634

V. Conclusion

Thus, sentiment analysis is performed on the IMDB dataset using Recurrent Neural Networks of different layers and their performances are studied.

In this generation, sentiment analysis has become an important tool in analyzing and understanding people's emotion towards a particular product, movies and many other things. Predicting a customer's emotion accurately is vital for a brand to improve or shape its quality in order to fulfill the customer's requirements. Thus, there is a need for a better model that could handle voluminous data and still give results with better accuracy. The current project is aimed at bringing better results and in the future, can also be improved using a better hybrid approach that could further improve the performance of the model.

References

[1] Mohammad Ehsan Basiri, Member, IEEE, Moloud Abdar, ArmanKabiri,Shahla Nemati, Member, IEEE, Xujuan Zhou, Forough Allahbakhshi, and Neil Y. Yen, Member, IEEE (2019) "Improving Sentiment Polarity Detection Through Target Identification", IEEE Transactions on Computational SocialSystems, Vol. 7, No. 1, pp. 113-128.

[2] Ali Feizollah, Sulaiman Ainin, Nor Badrul Anuar, Nor Aniza Binti Abdulla, and Mohamad Hazim (2019) "Halal Products on Twitter: Data Extraction and Sentiment Analysis Using Stack of Deep Learning Algorithms", IEEEAccess,Vol. 7, pp. 83354-83362.

[3] Jin Wang, Liang-Chih Yu, Member, IEEE, K. Robert Lai, and Xuejie Zhang (2019) "Tree-Structured Regional CNN-LSTM Model for DimensionalSentiment Analysis", IEEE/ACM Transactions on Audio, Speech and Language Processing, Vo. 28, pp. 581-591.

[4] Guixian Xu, Yueting Meng, Xiaoyu Qiu, Ziheng Yu, and Xu Wu (2019) "Sentiment Analysis of Comment Texts based on BiLSTM", IEEE Accesson Artificial Intelligence and Cognitive Computing for Communication and Network, Vol.7, pp. 51522-51532.

[5] Anwar Ur Rehman & Ahmad Kamran Malik & Basit Raza &Waqar Ali (2019)"A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis", SpringerLink on Multimedia Tools and Applications, Vol.78, pp. 26597-26613.

[6] Zabit Hameed, (Member, IEEE), and Begonya Garcia-Zapirain, (Member,IEEE) (2020) "Sentiment Classification Using a Single-Layered BiLSTM Model", IEEE Access, Vol.8, pp. 73992 - 74001.