Sentiment Analysis of Movie Reviews using ELMO Word Representation on Hybrid Neural Network

MAIN PROJECT REPORT

Submitted by

SHAHANA MUSTHAFA

REG NO : TKM20CSCE14

In partial fulfillment for the award of the degree of

MASTER OF TECHNOLOGY
IN
COMPUTER SCIENCE AND ENGINEERING

Under the guidance of
Dr. Ansamma John
&
Prof. Anantha Padmanabhan N K

Thangal Kunju Musaliar College of Engineering
Kerala

SEPTEMBER 2022

Thangal Kunju Musaliar College of Engineering
Dept. of Computer Science & Engineering

C E R T I F I C A T E

This is to certify that this report titled ***Sentiment Analysis of Movie Reviews using ELMO Word Representation on Hybrid Neural Network*** is a bonafide record of the **Main project** presented by **SHAHANA MUSTHAFA (TKM20CSCE14)**, under our guidance and supervision, in partial fulfillment of the requirements for the award of the degree, **M.Tech in Computer Science & Engineering** in **APJ Abdul Kalam Technological University** .

Coordinators

Dr. Ansamma John
Professor
Computer Science & Engineering

Prof. Anantha Padmanabhan N K
Assistant Professor
Computer Science & Engineering

Head of the Department

Dr. Dimple A Shahjahan
Associate Professor
Computer Science & Engineering

ACKNOWLEDGEMENT

A successful main project is a fruitful culmination of efforts by many people, some directly involved and some others indirectly, by providing support and encouragement. Firstly I would like to thank the almighty for giving me the wisdom and grace for making my main project a memorable one. I thank him for steering me to the shore of fulfillment under his protective wings.

I express my sincere gratitude to **Dr. T A Shahul Hammed** , Principal, T.K.M College of Engineering for giving me an opportunity to present my mini project. I would like to thank **Dr. Dimple A Shahjahan**, Associate Professor and Head of the Department, CSE, TKMCE, for her constant support and encouragement throughout the project work.

With a profound sense of gratitude, I would like to express my heartfelt thanks to my guide **Dr. Ansamma John and Prof. Anantha Padmanabhan N K**, Professor, CSE, TKMCE, for their expert guidance, cooperation and immense encouragement. I also extend my thanks to the entire faculty members and staffs of the Department of Computer Science and Engineering, TKMCE, who has encouraged me throughout this work.

I also express my thanks to my loving parents, brothers and friends, for their support and encouragement in the successful completion of this project work.

Shahana Musthafa

Abstract

Sentiment analysis is an active research area in natural language processing where underlying emotions expressed in textual data are analysed and identified. It can either be categorized as positive and negative or multiclass such as happiness, sadness, anger, fear, etc that enables to understand the social sentiment of a particular brand, product or service while monitoring online conversations. In addition, it can be used for studying psychological problems, conducting emotional marketing and for improving customer experience. The overload of data in today's environment makes it impossible to analyze it manually and makes systematic sentiment analysis even more relevant.

Many approaches were used earlier for sentiment analysis like ruled - based analysis and machine learning based analysis. Recent research indicates that deep learning techniques shows better performance for sentiment analysis. Hence the hybrid neural network model consisting of CNN and BiLSTM is introduced. Initially, text are converted into vector format using embedding languages such as Word2Vec, GloVe and ELMo language model. Then CNN allows extracting local features of text vectors while global features are extracted by BiLSTM. The features extracted by the two models are then used together for performing the sentiment analysis. Experiments are conducted on IMDB movie reviews dataset. The trained hybrid neural network can then automatically classify the sentences achieving an accuracy rate of 91% for text categorization compared to other state-of-art approaches.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Sentiment analysis is an active research area in natural language processing where underlying emotions expressed in textual data are analysed and identified. Gaining in-depth knowledge of people's emotions is helpful to the business as it allows to gain an overview of the wider public opinion. The overload of data in today's environment makes it impossible to analyze it manually and makes systematic sentiment analysis even more relevant. Hence automatic analysis of customer opinion in the form of surveys, questionnaires, social media forums allow the brands to hear their customers about the product and services. This helps them to understand their customers better through sentiment analysis and opinion mining. For example, sentiment analysis can help to automatically read tons of product reviews and extract useful and meaningful information to discover if the customers are really satisfied with the product or not.

Sentiment analysis applies natural language processing and text analysis techniques to highlight the subjective information from the text. A person's feedback is more subjective rather than factual. It is an important factor when it comes to product and brand recognition, customer loyalty, customer satisfaction, advertising and promotion's success, and product acceptance. Understanding the psychology of consumers can help achieve success and to alter their product road map with greater precision. Some other notable applications of sentiment analysis are tracking sentiments in real time, studying psychological problems, conducting emotional marketing, boost employee morale and many more.

Generally sentiments can either be categorized as positive and negative or multi class such as happiness, sadness, anger, fear, etc that enables to understand the social sentiment of a particular brand, product or service while monitoring online conversations. In light with these categorization, it can be seen that sentiment analysis focuses on the polarity of a text (positive, negative, neutral) or to detect specific feelings and emotions (angry, happy, sad, etc), urgency (urgent, not urgent) and even intentions (interested, not interested). There are different algorithms that can be implemented for performing sentiment analysis, depending on how much data needs to analyzed and how accurate the model needs to be. Some of the approaches used earlier for sentiment analysis are ruled-based analysis and machine learning based analysis. Rule-based systems automatically perform sentiment analysis based on a set of manually crafted rules and automatic systems rely on machine learning techniques to learn from data. Recent research indicates that deep learning techniques shows better performance for sentiment analysis and among them accuracy of the hybrid model outperforms than those of single models such as CNN and LSTM.

A word embedding represents the word meaning as a numeric vector. It is thus very useful for finding relations between words. With vector operations, it is possible to discover words used with similar contexts, solve word analogies and create visualizations of similar words. Word embedding helps capture the semantic, syntactic context of a word or term and helps understand how similar or dissimilar it is to other terms in an article, blog, etc. It implements language modeling and feature extraction based techniques to map a word to vectors of real numbers. Some of the algorithms used to create word embeddings are:

- GloVe creates the embeddings by generating a matrix with the number of occurrences of the surrounding words and performing statistics on that matrix.
- Word2Vec will capture relationships of words with the help of window size by using skip-gram and CBOW methods.
- ELMo creates the word embeddings based on the context, which means that the same word can have different embeddings according to the words nearby.

ELMo(Embeddings from Language Models) is a state-of-the-art NLP framework developed by AllenNLP that resulted in one of the biggest breakthrough. ELMo is a novel way to represent words in vectors or embeddings. These word embeddings are helpful in achieving state-of-the-art results in several NLP tasks. ELMo word vectors are computed on top of a two-layer bidirectional language model (biLM). This biLM model has two layers stacked together. Each layer has 2 passes - forward pass and backward pass. The architecture uses a character-level convolutional neural network to represent words of a text string into raw word vectors. These raw word vectors act as inputs to the first layer of biLM. The forward pass contains information about a certain word and the context (other words) before that word. The backward pass contains information about the word and the context after it. This pair of information, from the forward and backward pass, forms the intermediate word vectors. These intermediate word vectors are fed into the next layer of biLM. The final representation (ELMo) is the weighted sum of the raw word vectors and the 2 intermediate word vectors.

A hybrid neural network model consisting of CNN and BiLSTM[1] is introduced in this project where CNN allows extracting local features of text vectors while global features of text vectors are extracted by BiLSTM model. The features extracted by the two models are then used together for performing the sentiment analysis on the data achieving an accuracy rate of 91% for text categorization compared to other state-of-art approaches. Experiments are conducted using the large data set of movie reviews collected from IMDB consisting of a training set and a test set. It contains 25000 positive emotional movie reviews and negative emotional movie reviews each. Each review comprises of positive or negative reviews along with its label. Wikipedia and GoogleNews datasets are selected as corpus. For comparison, single models Convolutional Neural Network (CNN)[2] and Bidirectional Long Short Term Memory (BiLSTM)[3] are studied.

The rest of the report is organized as follows: Literature Review is discussed in Chapter 2. Chapter 3 introduces the proposed system. The experiments are carried out and the evaluation results are discussed in Chapter 4. Conclusion and Future works are discussed in Chapter 5. The report also includes references which list the many works that were examined while working on the project.

Chapter 2

Related Works

Sentiment analysis is an important task due to its various applications in marketing research and various natural language processing tasks. An effective architecture for sentiment analysis with hybrid neural network CNN-BiLSTM is proposed in this project. This model combines convolutional neural network CNN and bidirectional LSTM (BiLSTM). Embedding of words are obtained using Word2Vec, GloVe and ELMo language model. The classifier is used so as to predict sentiment of movie reviews with the help of feature extraction from these embedding. Some of the related works that were done in this are:

2.1 Deep learning-based sentiment classification based on Multi-feature fusion

In this paper[4], a deep-learning-based method to classify a user's opinion expressed in reviews called RNSA is introduced. The RNSA employs the Recurrent Neural Network (RNN) which is composed by Long Short-Term Memory (LSTM) that takes advantage of sequential processing and overcome several flaws in traditional methods, where order and information about the word are vanished. This deep learning-based method takes a unified feature set which is representative of word embedding, sentiment knowledge, sentiment shifter rules, statistical and linguistic knowledge. Sentiment and linguistics knowledge features are the input of RNN-LSTM, then the encoded feature from RNN-LSTM is considered as the sentence level representation. Perform the augmentation of the extracted vector representation from last LSTM cell with the sentiment and linguistic knowledge features to form a final vector representation for each sentence. This vector is taken to the fully connected layer, then the sigmoid function reveals the sentiment label either as positive or negative.

2.2 Sentiment classification using attention mechanism and BiLSTM network

This paper[5] introduces a sentiment classification method for large scale microblog text based on the attention mechanism and the bidirectional long short-term memory network. Firstly, words in the microblog corpus is used as the basic unit to form a sequence of words. N-gram features are extracted from the microblog text based on word embedding methods. Each word is then mapped into a multidimensional continuous value vector based the trained word vector and a word vector matrix representation of the entire sentence sequence is produced as the input of BiLSTM. Secondly, BiLSTM processes the previous and subsequent contextual features together by pairwise combining each forward and backward hidden layer. Finally, the predecessor and successor contextual features are processed based on the attention mechanism layers respectively. The attention mechanism layers in the model can achieve better performance, as each representation of the text can better remain focused on the important information. Then the output of attention mechanism layers are connected together as the input for the softmax classifier, which result in two main sentiment classes which are the non-negative sentiment and negative sentiments.

2.3 Deep learning for sentiment analysis: successful approaches and future challenges

Sentiment analysis is an active research area in natural language processing that aims at identifying sentiments from user generated texts in social networks, blogs or product reviews. Earlier, a lot of studies used ruled based approaches and machine learning techniques to solve sentiment analysis tasks from different perspectives. Recently, deep learning approaches[6] emerge as powerful computational models that help discover semantic representations of texts automatically from data. One of the main focus is to learn continuous word representation, called word embedding as words are the basic computational units of natural language. Sentiment classification where sentiment polarity of text is discovered is also studied as another important aspect. Finally it discusses opinion extraction and sentiment lexicon building. In opinion extraction, elements of a sentiment quintuple including the opinion holder, the entity and the sentiment of an effective expression are obtained. A sentiment lexicon is a list of words and phrases each of which is assigned with a positive or negative score reflecting its sentiment polarity and strength. Sentiment lexicon is a fundamental component for sentiment analysis. This paper introduces to various successful approaches for embedding learning, sentiment classification, opinion extraction and sentiment lexicon learning . It also specifies various challenge on generating new sentiment words or phrases from new corpus, using coarse and fine-grained supervisions, interpreting meaning of word embeddings and so on.

2.4 ELMo - Deep contextualized word representations

ELMo[7] word representations are functions of the entire input sentence. They are computed on top of two-layer biLMs with character convolutions as a linear function of the internal network states. This allows to do semi-supervised learning, where the biLM is pre-trained at a large scale and easily incorporated into a wide range of existing neural NLP architectures. The pre-trained biLMs support joint training of both directions and add a residual connection between LSTM layers. Each token is assigned a representation that is a function of the entire input sentence. It uses vectors derived from a bidirectional LSTM that is trained with a coupled language model objective on a large text corpus. For this reason, they are called ELMo (Embeddings from Language Models) representations. ELMo representations are deep, in the sense that they are a function of all of the internal layers of the biLM as they learn a linear combination of the vectors stacked above each input word for each end task, which markedly improves performance over just using the top LSTM layer. Combining the internal states in this manner allows for very rich word representations. Using intrinsic evaluations, the higher-level LSTM states capture context-dependent aspects of word meaning while lower level states model aspects of syntax.

2.5 A ConvBiLSTM Deep Learning Model-Based Approach

An integrating structure of CNN and Bi-LSTM model is proposed in this paper[8]. CNN model help extracts local features from word embedding whereas Bi-LSTM captures long-distance dependencies. ConvBiLSTM is implemented with the help of a word embedding model which converts tweets into numerical values. CNN extracts text features and give its context information of text to Bi-LSTM. In other words, CNN layer receives feature embedding as input and produces smaller dimension of features, and the Bi-LSTM model takes the input from the CNN layer and produces classification result. Word2Vec and GloVe were distinctly applied to observe the impact of the word embedding result on the proposed model. These models are pre-train unsupervised word vectors that are trained with a large collection of words and can capture word semantics. The different word vector models helped to verify effectiveness of the model. ConvBiLSTM model with Word2Vec on retrieved Tweets dataset outperformed other models. A crucial step was deciding on the hyper-parameter and studies were conducted to tune hyper-parameter so as to optimise the model. Finally, the model could classify text sentiment effectively on both datasets.

2.6 Sentiment Analysis using Word2vec-CNN-BiLSTM Classification

Traditional neural network based short text classification algorithms for sentiment classification has many weakness and is easier to find errors. In order to solve this problem, the Word Vector Model (Word2vec), Bidirectional Long Short-term Memory networks (BiLSTM) and Convolutional Neural Network (CNN) are combined for addressing the individual weaknesses and leverage the distinct advantages of LSTM and CNN[9]. In the initial stage, data cleansing and pre-processing are performed. Then, distributed document representation using Doc2Vec embedding is applied to prepare data for convolution. The resulting vector is passed as an input to the next stage. In the convolution stage, convolution and max pooling layers are applied for feature extraction to extract high level features. In the final stage, BiLSTM and fully connected layers are applied for document sentiment classification. The output of this stage is the final classification of the document. The experiment shows that the accuracy of CNN-BiLSTM model associated with Word2vec word embedding is higher compared to other state-of-the-art approaches. This proves that the hybrid network model performs better than the single structure neural network in short text classification.

2.7 Hybrid Neural Network for Sina Weibo Sentiment Analysis

In this paper[10], a novel structure of a hybrid neural network model is proposed to deal with the polysemy phenomena of words and topic confusion with Sina Weibo. First, the embeddings from language models (ELMo) and some statistical methods based on the corpus and sentiment lexicon are employed to extract the features. This method uses latent semantic relationships in different linguistic contexts and co-occurrence statistical features between words in Weibo. Second, for the classification model, employ several filters with variable window sizes to extract a sequence of high-level word representation in different granularity distributions of text data in multichannel CNN. At the same time, obtain the sentence representation in Bi-LSTM. Then, concatenate the outputs of multichannel CNN and Bi-LSTM. In the output layer, the concatenated output is mapped to the full connection layer and then inputs to the softmax classifier giving the final classification. The experimental results indicate that the proposed model performs better on the precision, recall, and F1-score for Weibo sentiment analysis.

As a summary of these works, it can be concluded that sentiment analysis is an active research area. Also, hybrid architecture and embeddings can help improve performance while carrying out sentiment analysis. With pre-trained models, more semantic information could be included which other models could not incorporate. Various problems like improper context capturing were seen in the previous works which was properly addressed in this project. The developments in various analysis techniques with time could also be noted.

Chapter 3

Proposed System

Sentiment analysis using hybrid neural network CNN-BiLSTM was proposed in this project. It combines convolutional neural network and bidirectional long short-term memory. Embedding of words are obtained using language models like Word2Vec, GloVe and ELMo. A softmax classifier is used at the end to classify sentiments. In this section, various phases of proposed system are discussed. An architecture of hybrid neural network is shown in Figure 3.1

3.1 Overview of hybrid neural network

- The dataset is given to a data preprocessing module that remove useless words, invalid symbols, punctuation symbols and lowercase the characters.
- Then it is converted into numeric value called embeddings by the embedding layer since machines do not understand text.

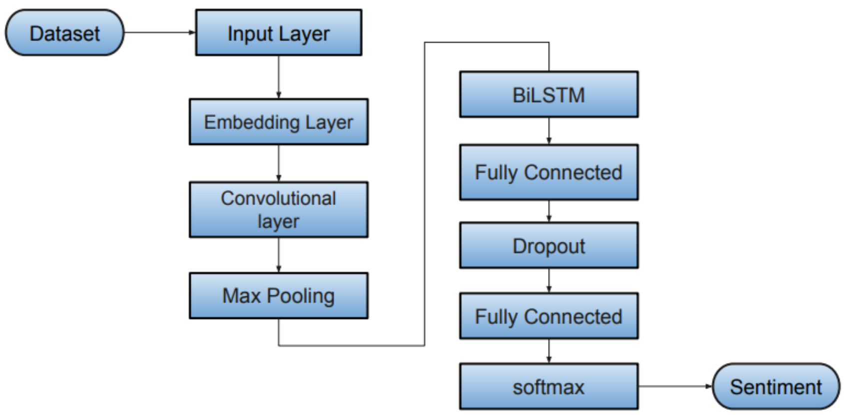

Figure 3.1: Framework of hybrid neural network model

- The embeddings are given to the convolutional layer that extracts the local features and pooling operation is applied continuously to reduce features.
- The BiLSTM module then captures the sequence correlation effectively by using its forward and backward LSTM.

- The output of BiLSTM are aggregated by the full connection layer to form features for classification.
- Softmax classifier is the last layer that takes features as input and find probability vectors to classify the sentiment either as positive or negative.

3.2 Convolutional neural network

Convolutional neural network(CNN)[11] is used for feature extraction in the model architecture introduced. Initially the words are transformed into vectors by word vectorization. CNN starts with an input matrix, followed by a shaping operation to convert data into a specific format[12]. Then, convolution blocks[13] and the largest pool block are applied continuously. Each block convolutes its input signal with an estimated kernels of specific step size, and step 1 is designed to extract a specific number of feature maps from the data. Then, the maximum pool layer is used to reduce output. Then, the output of the pooling layers is put into a full connection

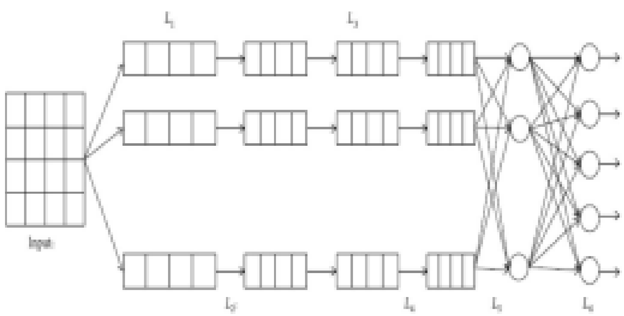

Figure 3.2: Convolutional neural network structure

layer, and the output features of the last layer are aggregated by the full connection layer to form a global feature for text emotional classification in an architecture only involving CNN. Finally, these features will be input into the last layer with five neurons, and the probability vectors will be obtained by using soft max. The structure of convolution neural network is shown in Fig. 3.2. In the proposed architecture, convolution and largest pool operations are applied to the word vectors continuously and the result is given to the BiLSTM module following it which captures sequence correlation.

3.3 Bidirectional long short-term memory

Long Short Term Memory(LSTM)[14] introduces three gates to maintain state. There are three gates, including input gate (i_t), forget gate (f_t) and output gate (o_t). The problem with LSTM is that it assumes the current text is only affected by the previous text frame, however the following frame data is also related to the current state. BiLSTM is very suitable for achieving this property as it consists of 2 layers which are the forward and backward LSTM. Hence BiLSTM[15] can solve the relationship between the two text frames and strengthen the bidirectional relationship between the current text frame and the next text frame.

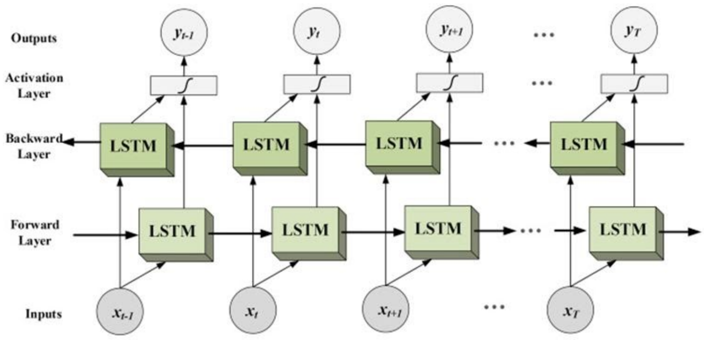

Figure 3.3: BiLSTM architecture

3.4 Data preprocessing

The data used in this project is the IMDB movie reviews dataset consisting of various positive and negative reviews. Since these reviews are very informal and unstructured, there can be many errors, useless words, invalid symbols and punctuation symbols in the reviews that needs to be removed. Fig 3.4 shows a sample of movie review before and after processing. Lowercasing is also performed as the methods in word vectorization module are case sensitive. The operations thus performed in data preprocessing are as follows:

- Removal of invalid symbols
- Removal of punctuation symbols

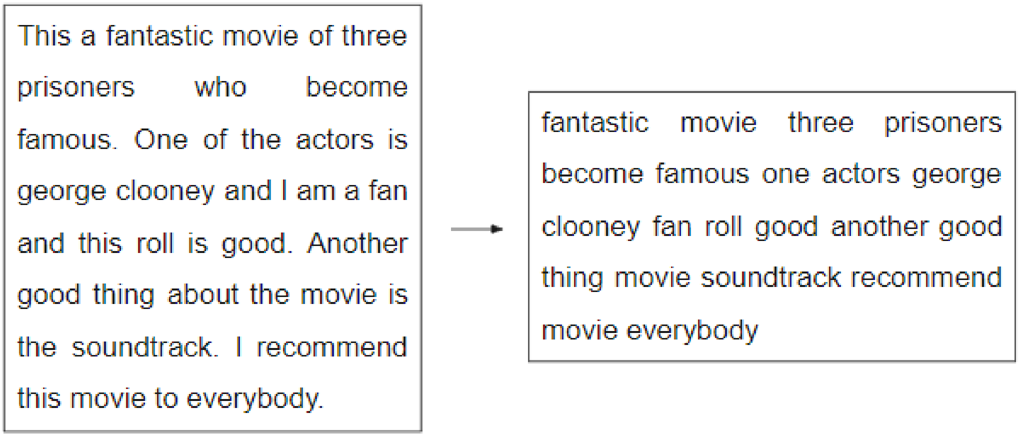

Figure 3.4: Data preprocessing

- Removal of stopwords
- Lowercasing

3.5 Word Vectorization

Word embedding language model[16] is used to obtain the input vector of text in K(100, 300 or 1024) dimension. This is then given as the input of CNN-BiLSTM feature extraction model that prevents difficulty of model training and also helps in obtaining more semantic information[17]. Word2Vec, GloVe, and ELMo embedding have dimensions of 100, 300, and 1024, respectively. The Word2Vec embeddings are based on the word's context, GloVe embeddings are based on the word's co-occurrence in the corpus, and ELMo uses a bidirectional LSTM language model to learn the word and its context.

3.6 Feature selection and extraction

The word embedding layer gives the matrix of the word embedding as input to the CNN layer. The next layer is the convolution layer, which carries out convolution operation and extracts local features. This is followed by a maximum pooling operation that extracts key features, discards redundant features and generates feature vectors with fixed dimensions. It is then followed by a layer of BiLSTM whose both layers are of size 128 hidden layers. This layers ensures the effective capturing of sequence correlation. Finally, the output parts of the two hidden layers are joined together to get the final output of BiLSTM.

The final feature is saved in output, as the input of the first full connection layer. Also a dropout mechanism is introduced between the first full connection layer and the second full connection layer, and some trained parameters are discarded in each iteration, so that weight updating does not depend on some inherent features and avoids over-fitting.

3.7 Classifier

At the final stage, input is given to the softmax classifier and it output the classification results. The main purpose of this training model is to correctly predict whether the sentiment of the input sentences are positive or negative. The algorithm design of sentiment analysis based on CNN–BiLSTM model is presented in Algorithm2.

Algorithm 1 Algorithm of CNN-BiLSTM

Loading corpus
Data preprocessing of corpus T_{train}, T_{test}
Get the set W of word vectors of text
Initialize CNN–BiLSTM model parameters
For each sentence $t \in T_{train}$
The set of word vectors of all the words found forms $t = [w_1, w_2, ..., w_{n-1}, w_n]$
Obtaining eigenvalue $c_j = f(w_j * m + b)$ by convolution operation
The most important features $c_j^k = max(c_{i,j})$ are obtained by maximum pooling method
Generate expression sequence and output eigenvector $h = [h_1, h_2, ..., h_{n-1}, h_n]$ through BiLSTM memory storage unit
Back propagation algorithm is used to adjust model parameters and text word vectors
Calculating the probability value $P(y^{(i)} = j|x^{(i)} : \theta) = \frac{exp(\theta_j^T x^{(i)})}{\sum_{l=1}^{k} exp(\theta_j^T x^{(i)})}$ of text sentiment label of input samples by using Softmax operation
End for
For each sentence $t \in T_{test}$
Classification of samples using trained CNN–BiLSTM model
Output emotional category label
End for

Chapter 4

Experimental Results and Discussion

Experiments are conducted on IMDB dataset and Tensorflow is used as the experimental tool. Embedding techinques Word2Vec, GloVe and ELMo are used to generate word vectors and implement the training of CNN-BiLSTM model[18]. For comparison, single models Bidirectional LSTM (BiLSTM)[19] and Convolutional Neural Network (CNN)[20] are studied and its performance is compared with hybrid model with each of the embedding technique.

4.1 Experimental data

Experiments are conducted using the large data set of movie reviews collected from IMDB consisting of a training set and a test set. It contains 25000 positive emotional movie reviews and 25000 negative emotional movie reviews and also each review comprises of label specifying its nature.A sample of dataset is shown in Fig. 4.1. Wikipedia and GoogleNews vectors are selected as corpus. Firstly, various preprocessing steps such as removal of useless words, invalid symbols and punctuation symbols are done on the data. A basic description of the dataset specifying count and unique number of reviews and sentiment along with highest frequency review is shown in Fig. 4.2.

	Review	Sentiment
0	One of the other reviewers has mentioned that...	Positive
1	A wonderful little production. The ...	Positive
2	Basically there's a family where a little boy ...	Negative

Figure 4.1: Dataset

4.2 Data preprocessing

Reviews are written by humans in an informal and unstructured manner so there are many possibilities of encountering misspellings,errors and symbols. For accurate results, the data needs to be processed by removing

	Review	Sentiment
count	50000	50000
unique	49582	2
top	Loved today's show!!! It was a variety and not...	Positive
freq	5	25000

Figure 4.2: Data description

useless words, invalid symbols, punctuation symbols and finally the lower-casing of characters due to case sensitivity. The output after performing preprocessing is shown in Fig. 4.3:

	Review	Sentiment
0	One reviewers mentioned watching 1 Oz episode..	1
1	wonderful little production filming technique ...	1
2	Basically family little boy Jake thinks zombie...	0

Figure 4.3: Processed data

4.3 Word Vectorization

The data obtained from input layer is in textual format that machine do not understand and process. So a transformation to numeric valued embedding is made using embedding language model. For experiment and analysis, Word2Vec, GloVe and ELMo of dimensions 300, 100 and 1024 are studied and implemented. The embedding are obtained in a matrix format as shown in Figure 4.4:

```
[ 0.17480469 -0.10986328 -0.20019531 ...
0.07958984  0.02807617  -0.02026367]
 [-0.0038147  -0.01916504 -0.13085938 ...
0.06347656  0.03417969  0.03344727]
 ...
 [ 0.28693558  0.5821401   0.17414214 ...
0.12220891  0.92546049 0.2759993 ]
```

Figure 4.4: Embedding

4.4 Feature selection and extraction

Feature selection and extraction section comprises of convolutional neural network and bidirectional long short-term memory. Once the embeddings are formed, it passes through convolution operation followed by pooling capturing local features and discarding some of them so as to eliminate redundant features. The output thus obtained is given to a BiLSTM consisting of 128 hidden layers that capture sequence correlation. It is then aggregated by full connection layer to form features for classification.

4.5 Classifier

This is the final section where sentiments are classified into either positive or negative by using the activation function softmax. It takes features from previous layer and find probability value for both negative and positive label thereby assigning the label with higher probability value. Some of the values obtained is shown in Fig. 4.5.

```
[[9.9247712e-01 5.2849147e-03]
 [9.7791630e-01 1.5029620e-02]
 [3.5588583e-04 9.9990344e-01]
 ...
 [4.5562745e-04 9.9987209e-01]
 [9.9995136e-01 4.6750749e-05]
 [7.9006764e-05 9.9999201e-01]]
```

Figure 4.5: Classifier output

4.6 Evaluation

Once the model has been built and trained, its goal is to correctly predict the sentiment categories of sentences. For this, a test set is used and the predictions are made using the trained model. Word2Vec, GloVe, and ELMo are the three embedding techniques used to build the model. These methods yielded results of 88.5, 86.5, and 90.58, respectively. The accuracy plot of CNN-BiLSTM with GLoVe and Word2Vec and CNN-BiLSTM with ELMo is shown in Fig. 4.6 and Fig. 4.7 respectively.

4.7 Comparison with single model

The hybrid neural network model CNN-BiLSTM is compared with single model convolutional neural network and bidirectional long short-term

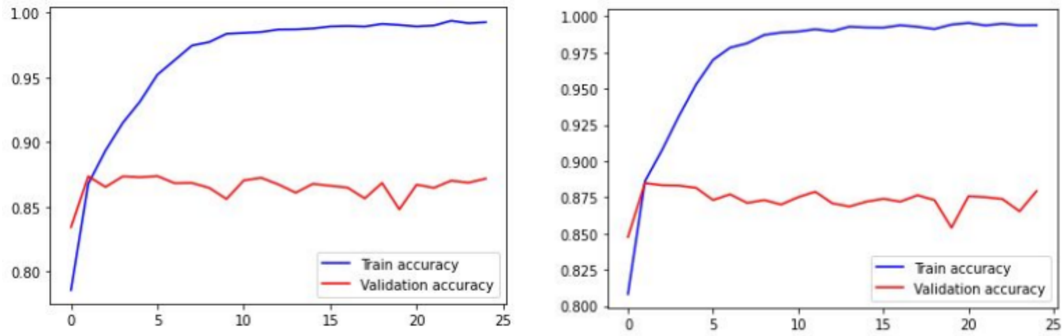

Figure 4.6: Accuracy plot of GloVe and Word2Vec with CNN-BiLSTM

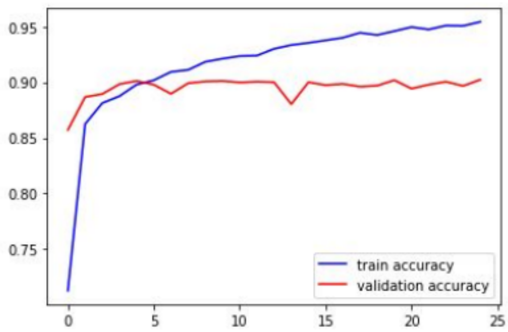

Figure 4.7: Accuracy plot of ELMo with CNN-BiLSTM

memory model. The accuracy plot with each of embedding techniques are obtained as shown and compared. CNN with embedding techniques GloVe, Word2Vec and ELMo are shown in Fig. 4.8 whereas single model BiLSTM with the three embedding techniques are shown in Fig. 4.9

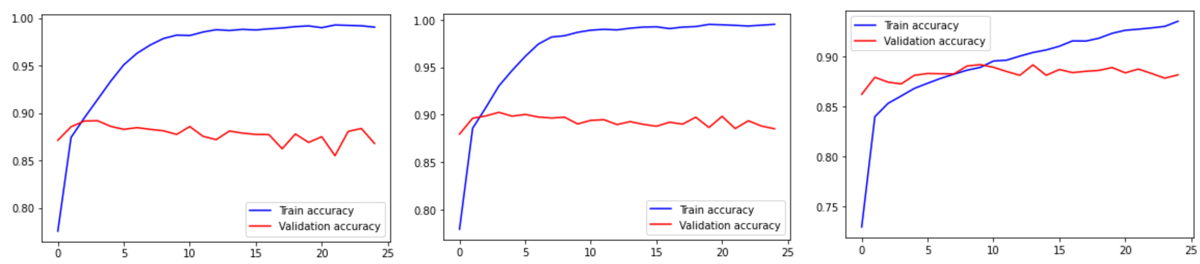

Figure 4.8: Accuracy plot of CNN

4.8 Comparison of embedding models

The values obtained on experimentation using the embedding techniques and single as well as hybrid models are summarised in Table 4.1.

On comparing with other baseline models, hybrid model CNN- BiLSTM is seen to achieve better accuracy of text classification compared to CNN

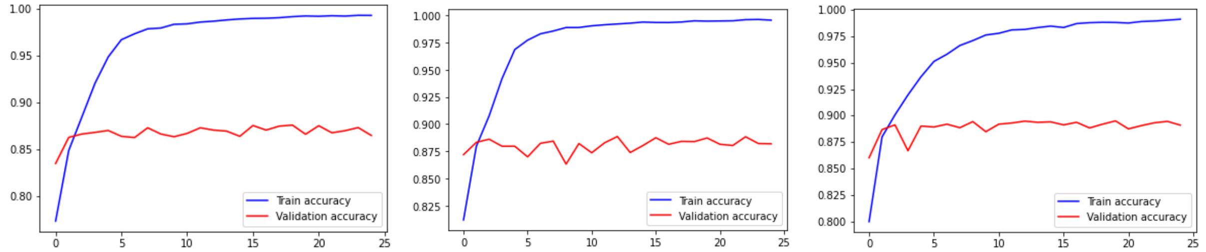

Figure 4.9: Accuracy plot of BiLSTM

Table 4.1: Evaluation of hybrid model

Embedding	**Model**	**Accuracy(%)**
Word2Vec	CNN	88
	BiLSTM	87
	CNN-BiLSTM	89
GloVe	CNN	86
	BiLSTM	86
	CNN-BiLSTM	87
ELMo	CNN	88
	BiLSTM	88
	CNN-BiLSTM	91

and BiLSTM. Each of these models with different embedding techniques was also analysed and among them ELMo achieved higher accuracy results compared to Word2Vec and GloVe models.

Chapter 5

Conclusion and Future works

Sentiment analysis is an active research area in natural language processing where underlying emotions expressed in textual data are analysed and identified. In this project, sentiment analysis using hybrid neural network CNN-BiLSTM was proposed. Here convolutional neural network is used to extract local features and BiLSTM model for capturing sequence correlation of text vectors effectively. It also incorporated word embedding models Word2Vec, GloVe and ELMo into hybrid model, which achieves state-of-the art result on IMDB movie reviews dataset. Experiments show that when word vectors constructed by ELMo model pass through CNN–BiLSTM model, it achieves higher accuracy compared to other embeddings like Word2Vec and GloVe. The hybrid model introduced is further compared with single CNN model and single BiLSTM model. The classification accuracy of the hybrid model is better than that of single CNN and single BiLSTM model. The results show that the feature fusion model is superior to the contrast model in classification accuracy and it effectively improves the accuracy of text classification

In the future, there are still some questions that need further study. Firstly, the model has a lot of hyper-parameters which can be further explored so as to improve accuracy of text classification. Secondly, emoticons, slang and jargon are said to express opinions more effectively but these elements were ignored in this work. The proper interpretability of embeddings is another issue to be resolved. As for applications, the hybrid model discussed could be used to perform sentiment analysis on twitter data, hate speech detection and text-based analysis of comments[21-23].

References

[1] Zi- Xian Liu, De- gan Zhang, Gu- Zhao Luo, Ming Lian, Bing Liu, "A new method of emotional analysis based on CNN–BiLSTM hybrid neural network", Springer, 2020.

[2] Y., Zhang, B., Xue, "Research on Chinese text classification method based on FOA-SVM", J. Sichuan University, 2016.

[3] Gao, J.P., Zhang, H., Zhao, X.J., et al., "Research on WEB domain knowledge classification based on feature words", 2016.

[4] Abdi, Asad, et al., "Deep learning-based sentiment classification of evaluative text based on Multi-feature fusion", Elsevier, 2019.

[5] Wu, Peng, et al., "Sentiment classification using attention mechanism and bidirectional long short-term memory network" , Elsevier, 2021.

[6] Tang, D., Qin, B., Liu, T., "Deep learning for sentiment analysis: successful approaches and future challenges", Wiley, 2015.

[7] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," in Proc. NAACL-HLT, 2018.

[8] Tam, Sakirin and Said, Rachid Ben and Tanri, "A ConvBiLSTM Deep Learning Model-Based Approach for Twitter Sentiment Classification", IEEE Access, 2021.

[9] Yue, Wang and Li, Lei, "Sentiment Analysis using Word2vec-CNN-BiLSTM Classification", IEEE, 2020.

[10] Mingjie Ling, Qiaohong Chen, Qi Sun, and Yubo Jia, "Hybrid Neural Network for Sina Weibo Sentiment Analysis", IEEE Transactions, 2020.

[11] Simonyan, K., Zisserman, A., "Two-Stream Convolutional Networks for Action Recognition in Videos", University of Oxford, 2014.

[12] Zhang, T., "Novel self-adaptive routing service algorithm for application of VANET", 2019.

[13] Zhang, D.G., Hui, G., "New multi-hop clustering algorithm for vehicular ad hoc networks", IEEE Transaction, 2019.

[14] Liu, S., "Dynamic analysis for the average shortest path length of mobile ad hoc networks under random failure scenarios", IEEE Access, 2019.

[15] Gao, J.X., "Novel approach of distributed adaptive trust metrics for MANET", 2019.

[16] Zhang, D.G., Chen, C., Yu, Y.C., et al., "New method of energy efficient subcarrier allocation based on evolutionary game theory", 2018.

[17] Liu, S., "Novel dynamic source routing protocol (DSR) based on genetic algorithm-bacterial foraging optimization", 2018.

[18] Salur, Mehmet Umut and Aydin, Ilhan, "A novel hybrid deep learning model for sentiment classification", IEEE Access, 2020.

[19] Xu, Guixian and Meng, Yueting and Qiu, Xiaoyu and Yu, Ziheng and Wu, Xu, "Sentiment analysis of comment texts based on BiLSTM", IEEE Access, 2019.

[20] Ouyang, Xi, et al., "Sentiment analysis using convolutional neural network", IEEE international conference on computer and information technology, IEEE, 2015.

[21] Mathur, Amrita and Kubde, Purnima and Vaidya, Sonali, "Emotional analysis using twitter data during pandemic situation: COVID-19", IEEE, 2020.

[22] Zhou, Yanling and Yang, Yanyan, Han and Liu, "Deep learning based fusion approach for hate speech detection", IEEE Access, 2020.

[23] Liu, Yi and Lu, Jiahuan and Yang, Jie and Mao, Feng, "Sentiment analysis for e-commerce product reviews by deep learning model of Bert-BiGRU-Softmax", 2020.