Research article

Scalable deep learning framework for sentiment analysis prediction for online movie reviews

Peter Atandoh [a,*], Fengli Zhang [a], Mugahed A. Al-antari [b,*], Daniel Addo [a], Yeong Hyeon Gu [b,*]

[a] School of Information and Software Engineering, University of Electronic Science and Technology of China, North Jianshe Road, Chengdu, 610054, Sichuan, China
[b] Department of Artificial Intelligence and Data Science, College of Software and Convergence Technology, Daeyang AI Center, Sejong University, Seoul, 05006, Republic of Korea

ARTICLE INFO

ABSTRACT

Sentiment analysis has broad use in diverse real-world contexts, particularly in the online movie industry and other e-commerce platforms. The main objective of our work is to examine the word information order and analyze the content of texts by exploring the hidden meanings of words in online movie text reviews. This study presents an enhanced method of representing text and computationally feasible deep learning models, namely the PEW-MCAB model. The methodology categorizes sentiments by considering the full written text as a unified piece. The feature vector representation is processed using an enhanced text representation called Positional embedding and pretrained Glove Embedding Vector (PEW). The learning of these features is achieved by inculcating a multichannel convolutional neural network (MCNN), which is subsequently integrated into an Attention-based Bidirectional Long Short-Term Memory (AB) model. This experiment examines the positive and negative of online movie textual reviews. Four datasets were used to evaluate the model. When tested on the IMDB, MR (2002), MRC (2004), and MR (2005) datasets, the (PEW-MCAB) algorithm attained accuracy rates of 90.3%, 84.1%, 85.9%, and 87.1%, respectively, in the experimental setting. When implemented in practical settings, the proposed structure shows a great deal of promise for efficacy and competitiveness.

1. Introduction

More and more people are expressing themselves online as a result of The fast advancement of the web and the proliferation of mobile Internet. In the present day, this occurrence is common. For example, many people want to express themselves on online movie sites like Rotten Tomatoes, IMDB, Letterboxd, Reddit, Metacritic, etc. After that, it might be categorized as subjective or emotional data. On top of that, using sentiment analysis facilitates the process of differentiating the good from the negative parts of content related to a specific disagreement. Therefore, a marketing plan that uses positive product reviews and user emotion can encourage the promotion of specific goods and services by endorsing them to consumers.

* Corresponding authors.
E-mail addresses: peteratandoh@std.uestc.edu.cn (P. Atandoh), fzhang@uestc.edu.cn (F. Zhang), en.mualshz@sejong.ac.kr (M.A. Al-antari), daddo@std.uestc.edu.cn (D. Addo), yhgu@sejong.ac.kr (Y. Hyeon Gu).

Analysis of topics discussed in online discourse can help predict rapid changes in customer preferences and popular trends in destinations using big data from online movie sites and social media [1].

It is possible to achieve this by promoting particular goods and services. To analyze user sentiment, the majority of current systems use ML, deep learning, and lexicons. At now, techniques such as lexicon, machine learning, and deep learning are being prioritized in sentiment analysis [2]. To determine the sentiments in documents or sentences, lexicon-based procedures use sentiment terms. Sentiment words are important for recognizing and understanding the emotional qualities expressed in writing. The lexicon-based method determines the document's orientation by looking at the semantic orientation of its words and phrases. Building classifiers from labeled text or words is the lexical approach to text categorization. Manual and automated dictionary creation are both possible using lexicon-based techniques. Research-based on dictionaries has made heavy use of adjectives to demonstrate the text's semantic direction The conventional methods of text sentiment analysis have been successful up until recently. These methods often include shallow learning techniques like Support Vector Machines (SVM) [3,4], Decision trees [5], and Bayes classifiers [6] for sentiment classification, or intensive manual feature extraction [7]. Although traditional approaches are usually easy to understand and work well with little data. These systems have significant drawbacks despite their minimal hardware requirements. Inadequate or incorrect expression could emerge from manual extraction of emotions. Furthermore, they have limited capabilities when it comes to representing complex functions and performing intricate categorization tasks. There has been a lot of research on ways to increase the capacity, capable of handling high-dimensional problems. A major step forward has been using deep learning technologies, ensemble technologies, and Fuzzy systems [8–10]. Several elements are included by the neural network model, used to reflect this viewpoint in their research [11], an individual's statement-making process and the associated emotions are the focus of this discussion. Belief in an emotionally charged remark shapes an individual's outlook on life, making them more optimistic or pessimistic. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are two widely used deep learning architectures. Convolutional Neural Network (CNN) based text sentiment analysis methods [12] can eliminate the need for custom-built features by extracting them through a weight-sharing network. Crucially, the network is straightforward and capable of running concurrently. To further improve textual sentiment analysis' efficacy, we use the attention-based bidirectional long short term memory [13], unique enhanced text representation [14], and multichannel CNN [15]. However, CNN requires the inclusion of numerous layers in order to adequately understand contextual information, and it is not very good at learning knowledge from successive dimensions. Similar to sequential modeling, textual sentiment analysis uses recurrent neural networks (RNN) to learn sequence-specific contextual information. However, due to gradient vanishing, traditional RNNs can only record information from short-term memory. The need to analyze text sentiment led to the development of specialized RNN iterations such as the Gated Recurrent Unit (GRU) and the Long Short-Term Memory (LSTM). To achieve this, models like BiGRU and Bidirectional LSTM (BiLSTM) (e.g., [16,17] are employed to collect forward and backward contextual information. Unfortunately, due to their high degree of complexity, these techniques cannot resolve the long-dependence problem for very long sequences. There have been attempts to solve these issues by combining CNN with RNN [18]. Most of the time, they do a good job, but occasionally they fail to recognize the vital importance of the initial traits.

It is possible to obtain word vector space representations with intricate syntactic patterns using glove embeddings. To convert words into vector representations, the Glove technique employs an unsupervised learning mechanism [19]. In order to measure how well a word is represented, the majority of word-vector algorithms look at the distances or angles between nearby word vectors. Instead of just computing the scalar distance between words, a new evaluation method has been proposed that investigates the intricate structure of the word vector space using word analogies. Using latent factorizations of matrices is an application of latent semantic analysis (LSA), as initially proposed by [20]. Mikolov [21] proposed the skip-gram model, which shows a contextualized window at the local level. There are currently major challenges that necessitate a resolution from both families. Although it makes effective use of statistical data, the Latent Semantic Analysis (LSA) technique performs poorly when the word analogy problem is presented. A less-than-ideal arrangement of the vector space is implied by this restriction. Although skip-gram algorithms are quite good at similar tasks, they hinder data efficiency since they only use small context windows instead of full co-occurrence counts.

Moreover, document-level sentiment analysis encounters numerous obstacles. Firstly, the absence of word order information in word embedding for sentiment analysis. For instance, the sentences "Author is great, bad book" and "Book is great, bad author" illustrate a matrix that solely captures information about word order. While the language remains unchanged, the sentences have significantly different meanings. Therefore, it is crucial to provide contextual signals in order to comprehend their individual interpretations.

Secondly, in a textual sentiment analysis, long sentences in a single review that have strong polarity may be assigned a sentiment label. However, these sentences express sentiments through words with weaker sentiments, making it challenging to accurately identify the sentiments. For instance, "The movie was certainly different." The sentence portrays the reviewer's assessment of the movie as unclear, allowing for both positive and negative meanings.

This study proposes an improved method of text representation utilizing scalable deep learning models, expanding upon the previously discussed methodologies and problems. With the help of Positional embedding and Glove, the text representation is enhanced. To improve upon Glove's pre-trained embedding, the model focuses on using positional embedding word order information. A multichannel Convolutional Neural Network (CNN) is employed by the model to automatically obtain hierarchical features, quickly parse input text, and produce a precise representation of the text for sentiment classification. According to several tests, increasing the number of channels can improve performance. The approach also makes use of global attention and local attention to identify the differences in traits. This makes it easier to distinguish between distinct emotions and improves the model's accuracy in identifying them. At the end of the day, the results show that our model is better than the state-of-the-art methods. The following is a brief description of this study's main contributions.

- First, we propose a PEW-MCAB model for sentiment classification in online reviews.
- Our proposed version of text representation incorporates both positional and Glove techniques. The model's main focus is on improving Glove's pre-trained embedding by using positional embedding to comprehend word-order information in texts.
- Second, to extract high-level properties of different sizes, a scalable multi-channel Convolutional Neural Network (MCNN) and attention is proposed. Its goal is to capture context features at several scales in addition to the original context information and also understand the implicit meaning in long text.
- Furthermore, the model incorporates an attention-based bidirectional LSTM to simulate the interactions between each pair of words in the input. Subsequently, the Softmax function utilizes the outputs from the attention layer to generate a prediction regarding the sentiment label of the document.

The rest of the sections of this manuscript are organized in the following manner: Section 2 offers an extensive summary of the most recent advancements in sentiment classification research. The proposed model, which was influenced by previous research, was introduced in sections 3 and 4. In the 5th section, a number of experimental processes were conducted to assess the efficacy of our framework concept. Section 5 concludes with a comprehensive overview of our empirical discoveries and offers suggestions for the future trajectory of our research work.

2. Related work

It is of the utmost importance to investigate the feelings expressed in written language because words are a straightforward and effective means of expressing them. Modern approaches to textual sentiment analysis are mostly covered in this part, with an emphasis on both traditional and techniques Utilizing deep learning.

A great deal of conventional techniques have been used for text sentiment analysis recently. This study illustrates the persistent high performance of Semantic Orientation Calculator (SO-CAL) in various areas, even when faced with unfamiliar data [22]. The authors provide a refinedsentiment analysis technique that effectively detects and interprets sentiment expressed in concise and informal text messages such as tweets and SMS. This system is specifically built to do sentiment analysis at two distinct levels: the message-level, which entails identifying the overall sentiment of the communication, and the term-level, which requires assessing the sentiment of individual words or phrases inside the message. To illustrate the point, Seref and Bostanci [23] used a supplemental Naive Bayes classifier to analyze Amazon movie reviews for sentiment, and they achieved an optimum accuracy rate higher than 65%. An approach for evaluating twitter sentiment using a Support Vector Machine (SVM) classifier employing -gram features was introduced by Naz et al. [3] in the same year. An accuracy rate of about 75% was attained using this method. Nonetheless, Rathi et al. [24] contended that present-day ML approaches fail to produce adequate results when it comes to sentiment classification. So, to analyze the sentiment of tweets, they came up with a blend machine learning approach that integrated Support Vector Machines (SVM) with Decision Trees. This led to an impressive 2% improvement in total categorization performance, which was a major accomplishment. Furthermore, to ascertain the sentiment weight for classifying the sentiments, the research in [7] used a sentiment lexicon in conjunction with other dictionaries. Using this method yielded more than a 5% improvement. A contextual lexicon for feature words and an emotion lexicon were created by Huang and Cao [25]. With these dictionaries, sentiment analysis in text reviews is more precise. A new study [5] created a sentiment analysis algorithm tailored to tweets about news. Part of speech (POS) tags, negation presence, and sentiment words were just a few of the thoughtfully crafted elements that the system made use of, along with decision trees. A remarkable result exceeding 85% sentiment classification accuracy was thus attained by the algorithm. This study introduces a novel machine learning (ML) method, namely the Local Search Improvised Bat algorithm-based Elman Neural Network, for the sentiment analysis of online product evaluations [26]. A fuzzy cognitive map is employed for the purpose of categorizing sentiments into either positive or negative categories. The suggested methodology comprises several sequential stages, including pre-processing, feature extraction, feature selection, and sentiment classification (SC) [27].

As deep learning technology advances, more and more researchers are utilizing it for textual sentiment analysis. This research presents a hybrid approach for categorizing the sentiment of online reviews from different fields. The system combines a convolutional neural network with long-short-term-memory. This method is based on Weibo's statistical data, which analyzes the co-occurrence of words and latent semantic links in various settings. Unlike the conventional C-LSTM approach, where the output of CNN is sent into LSTM, our method involves employing a set of filters with variable window widths. This allows us to retrieve a series of words at a higher degree of representations with changing levels of detail from the input text in a multichannel CNN architecture [28]. The authors [29] implemented a Convolutional Neural Network (CNN) based system to analyze the sentiment and identify emotions in a collection of tweets. The model outperforms standard methods by more than 7% in terms of performance. In addition, Yang et al. [30] developed a convolutional neural network (CNN) that incorporates emotion and semantic information. This model employs emoticon embedding and word-level text vectors to extract sentiment features, thereby enhancing the performance of CNN in sentiment classification of text reviews. The model achieved an improvement of 1% compared to eight frequently utilized approaches in NLP. In a similar vein, Song et al. [31] conducted research on a novel position-sensitive CNN. This CNN enhances the capability of extracting features by acquiring positional attributes across three different levels of language: single words, phrases, and sentences. Nevertheless, the aforementioned study [32] discovered that existing research primarily concentrates on identifying dominant sentiments, disregarding the concealed sentiments. To address this gap, the study proposed an approach that combines multiple levels of semantic features and a model based on Convolutional Neural Networks (CNN). The aforementioned methods were utilized to detect concealed attitudes in Weibo and car reviews by utilizing three levels of features. The findings indicated that the proposed methods achieved an identification accuracy of more than 70%, and a polarity classification accuracy of 75%.

Table 1
Dataset.

Name of dataset	Labels	Positive	negative	Total
IMDB	2	25,000	25,000	50,000
Movie Review (2002)	2	700	700	1400
Movie Review (2004)	2	1000	1000	2000
Movie Review (2005)	2	5331	5331	10662

The Recurrent Neural Network (RNN) is an alternative kind of neural network. RNN can read a text word by word and update its hidden state to reflect the meaning and sentiment it contains, and can dynamically represent contextual information, making them extensively employed in textual sentiment analysis. In a study conducted by Abdi, the researchers [33] conducted a study on a model based on Long Short-Term Memory (LSTM), which resulted in an enhancement of more than 4% in sentiment categorization of reviews. This improvement was obtained through the utilization of a multi-feature fusion method. In a similar vein, this study [34] examined Chinese aspect-level sentiment analysis using a multi-grained aspect target sequence and two fusion techniques. Their findings indicate that The model based on RNNs can accomplish a comprehensive enhancement of over 1% on Chinese review datasets. In addition, this study [35] utilized a context attention mechanism to enhance the performance of LSTM, resulting in a huge increase in the accuracy of tweet classification. In a similar vein, the study [36] introduced a multi-resource attention mechanism for sentiments that enables the model to gather a more extensive range of sentiment-related information from three different sentiment resources. This approach resulted in a performance enhancement of 2% on text reviews. Similarly, this study [37] conducted optimization on a dynamic memory architecture. They encoded word embeddings and position embeddings using the GRU paradigm. After that, they used a multiple-attention approach to identify significant sentiment-related aspects in the recollected data As a result, they achieved an accuracy of slightly more than 70% in sentiment analysis. Over time, researchers have integrated each of the variants of neural networks to improve the outcomes of sentiment analysis. The authors employ a combination of Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (LSTM) model to improve the outcomes of sentiment analysis [38]. The study introduced an alternating framework for textual classification tasks, specifically for sentiment classification of reviews [39]. The framework utilized BiLSTM to capture overall long-term relationships and CNN to extract local semantic characteristics. The purpose of local-guided global attention and global-guided local attention is to learn from each other and improve the way information is weighted. In a carried out in 2019, the authors utilized a sequential framework for sentiment and question categorization [40]. The convolutional layer, attention mechanism, and bidirectional LSTM (BiLSTM) make up the architecture, all integrated into a cohesive framework. Furthermore, two attention layers were incorporated to assign weights to future and historical high-level context features.

This study aims to gain insight into online movie reviews through the consideration of the polarity of sentiments (document level) and the lack of word order information in textual representations. In this regard, the study used an enhanced method of representing text and computationally feasible deep learning models, namely the PEW-MCAB model for classifying sentiments.

2.1. Research objectives

- To examine and evaluate the challenges associated with sentiment analysis based on the document level of online movie texts.
- To develop a PEW-MCAB framework for sentiment analysis at the document level.
- To study how different Glove representation dimensions work to ascertain which one works best with the proposed framework.
- To compare the findings of the PEW-MCAB model in comparison to the current most current models, and thereafter conduct an empirical analysis to find any distinct variations.

3. Method

3.1. The dataset

On four separate internet movie review datasets, we ran sentiment classification tests: The versions of Movie Review from 2002 and 2004, as well as IMDB [41] [42] [43], and MR [44]. The movie review datasets are depicted in Table 1.

- **IMDB**: Movie reviews pulled straight from the IMDB website make up the IMDB dataset. An example of a lengthy piece that uses a star rating system.
- **MR 2002**: Positive and negative statements are what make up the movie review dataset.
- **MR 2004**: Reviews in the 2004 movie review dataset are rated positively or negatively.
- **MR 2005**: Dataset containing movie review sentences and their corresponding polarity. The text comprises both positive and negative excerpts gathered from web pages of Rotten Tomatoes. Reviews classified as "fresh" are considered positive, while those labeled as "rotten" are considered negative.

3.2. Data pre-processing

The datasets were cleaned up at this part of the process. At first, the text reviews were edited to remove any HTML tags that were present. Following the removal of stop words in the manner stated, brief reviews were also removed from consideration. Additionally, single characters and numerous spaces were removed from the list of characters. Tokenization is performed on the white space that exists between each sentence, and the punctuation and numerals that are present in the sentences in the text are removed. All of the phrases that had terms that were made up of impure alphabet and numerical digits (alphanumeric characters) were removed. Every single stop-word and every single word that contained a special character has been eliminated from the sentences.

3.3. Baselines approaches

- **RAE** [45]: Introduced a Recursive Autoencoder model that utilizes pre-trained word vectors sourced from Wikipedia on MR (2005) dataset.
- **MNB** [46]: They considered multinomial Naive Bayes on IMDB dataset.
- **SVM** [47]: Multi-class sentiment categorization using an improved one-vs-one technique and the support vector machine (SVM) algorithm on the MR dataset.
- **NBSVM** [46]: They considered basic versions of NB (Naive Bayes) and SVM (Support Vector Machine) algorithms on the IMDB dataset.
- **MTL-ML4** [51]: An emotion classifier that learns many tasks using the IMDB dataset.
- **LSTM** [40]: They employed LSTM on IMDB AND MR (2005) datasets.
- **BiLSTM** [40]: They employed BiLSTM on IMDB AND MR (2005) datasets
- **AC-BiLSTM** [40]: They employed Attention and BiLSTM on IMDB AND MR (2005) datasets.
- **Static CNN** [48]: They employed CNN without fine-tuning pre-trained vectors on MR (2005) dataset.
- **Non-Static CNN** [48]: They employed CNN and fine-tuned pretrained vectors on the MR (2005) dataset.
- **Multichannel CNN** [48]: A CNN model consists of two collections of vector words, where each vector is considered as a separate channel on MR (2005) dataset.
- **BERT** [49]: A Bidirectional Encoder Representation for Transformer on IMDB MR (2002 and 2005) dataset.
- **CNN** [49]: A CNN IMDB MR (2002 and 2005) dataset.
- **B-cNN** [49]: A Combination of BERT and CNN model IMDB MR (2002 and 2005) dataset.
- **B-MLCNN** [49]: A Combination of BERT and Multi channel CNN model IMDB MR (2002 and 2005) dataset.
- **Glove CNN-BiLSTM** [38]: A Combination of Glove, CNN model BiLSTM on IMDB and MR (2005) datasets.
- **LSTM** [50]: It is employed to circumvent the issue of recurrent neural networks (RNNs), that impede the capacity to retain information over a long duration on IMDB and MR (2005) datasets.
- **BiLSTM** [50]: The BiLSTM model employs two LSTM components, one for processing the input in a left-to-right manner and the other for processing it in a right-to-left manner on IMDB and MR (2005) datasets.

3.4. Evaluation metrics

We utilized evaluation measures and benchmarks such as Accuracy (A), Precision (P), Recall (R), and F1 score (F). The parameters used in this calculation are explained in full below and are in line with those used in previous studies. Accuracy refers to the measure of accurately predicted reviews concerning all reviews. A True Positive (TP) refers to a result that correctly predicts positive feedback. A false positive (FP) occurs when a result incorrectly indicates a positive outcome that was not intended. A True Negative (TN) occurs when the results accurately predict favorable ratings. False negative (FN): This is noteworthy considering the outcome, which indicates that incorrect predictions were made for unfavorable assessments.

-
$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

-
$$P = \frac{TP}{TP + FP} \tag{2}$$

-
$$R = \frac{TP}{TP + FN} \tag{3}$$

-
$$F1 = \frac{2 * precision * recall}{P + R} \tag{4}$$

3.5. Training details

The model utilized a 300-dimensional Glove pre-trained embedding and Absolute Positional embedding to explicitly incorporate word order information into the representation of text. The input maximum sentence length is set to 400 for IMDB, and 100 for

Algorithm 1 Pseudo-code for PEW-MCAB.

1. Load Data (d)
2. Data is Preprocessed and Tokenized
3. Create an input word embedding matrix by using Eq. (1);
4. Input embedding is enhanced by integrating Absolute Positional Embedding (APE) using Eq. (2) and (3);
5. The vector c_i is concatenated with the GloVe vector w_i and the positional vector p_i for a word at location i using Eq. (4);
6. The concatenated vector is fed into a multi-channel convolutional neural network (CNN) using Eqs. (6)-(8);
7. Calculate the hidden states in both directions using Bidirectional Long Short-Term Memory (BiLSTM) using Eqs. (9) and (10)
8. The model utilizes attention layers to focus on the most relevant textual portions.
9. Compute the attention score e_i and The context vector c_i with score function and weighted attention respectively using Eqs. (11)-(13);
10. Softmax classifies the sentiment label based on the context vector using Eq. (14);
10. Select Initial Parameters;
11. Data is split into train and test;
12. Apply Adam optimizer to minimize the loss function of the model.
End Process

MR (2002, 2004, and 2005). The Conv1D1, Conv1D2, and Conv1D3 layers utilize 128 CNN filters with kernel sizes of 4, 6, and 8 correspondingly. The ReLU activation function is incorporated into the Convolutional Neural Network (CNN) to enhance the speed of the training process. It is advantageous in facilitating the convergence of gradients and addressing the problem of disappearing gradients. Adam optimizer is used as the network trainer, and the total number of training epochs is 100. Following every training session, the network undergoes testing using validation data. Every dataset uses the same fixed-size training batch size of 100. To prevent overfitting of the neural networks the dropout rate was set to 0.3. The BiLSTM layer utilizes an LSTM with 128 hidden units. Furthermore, the loss function utilized during training is the cross-entropy loss.

3.6. Proposed PEW-MCAB model

The Glove embedding technique is a well-acknowledged method for natural language processing (NLP) embedding, as stated in reference [19]. To accurately represent a word, the system must gather information on both the local and global statistical characteristics associated with that specific term. The methodology is commonly known as a hybrid strategy for continuous word representation, as it successfully combines the advantages of both the co-occurrence matrix and Word2Vec approaches. Nevertheless, it faces difficulty in accurately interpreting the sequential arrangement of words in their respective contexts. Additionally, the challenge of comprehending the implicit significance of phrases in lengthy review text is also observed. These issues impact the accuracy of sentiment analysis. This research aims to enhance the performance of text representation and sentiment classifier by designing an architecture that integrates positional embedding with Glove. This integration serves to incorporate word order information into the text representation, hence addressing the mentioned limitations. In addition, we incorporated the enhanced text representation into a scalable multi-convolutional neural network, and an attention-based bidirectional long short-term memory to improve the accuracy of sentiment classification. The proposed architecture is referred to as PEW-MCAB, which stands for Positional Embedding (PE) with Glove Word Vectorization (G) with Multi Convolutional Neural Network (MC) and attention-based BiLSTM (AB). The Positional Embedding Word Vector (PEG) is a method for learning vector representations that incorporate word order information. The convolutional layers extract n-gram characteristics from the text to model sentences. The BiLSTM (B) model incorporates a forward hidden layer and a backward hidden layer to make use of both preceding and following contextual factors. Sentence semantics understanding is aided by the Attention mechanism (A) for single-word representation, which gives priority to words that are pertinent to the text's sentiment. The PEW-MCAB model processes the contextual features via two layers of the attention mechanism that come before and after a given point. The characteristics processed by the Attention layers are combined and subsequently inputted into the softmax classifier. Fig. 1 displays the architectural structure of the PEW-MCAB model.

3.7. Embedding

The model first utilized GloVe embedding to represent our words in the model. The Glove embedding technique is a widely recognized method for embedding natural language processing (NLP). To represent a word, it acquires knowledge of both the local and global statistics associated with that word. It is commonly known as a hybrid strategy for continuous word representation since it incorporates the strengths of both the co-occurrence matrix and Word2Vec methods. The embedding is derived from the statistical analysis of the frequency of words together in a text. The objective of the GloVe model is to acquire word embeddings that effectively represent the semantic associations of words by considering their probability of co-occurrence. The objective is to get word vectors w_i and w_j from a word-word co-occurrence matrix X. In this matrix, $X_{i,j}$ denotes the frequency of word j appearing in the context of the word i. The aim is to find w_i and w_i in such a way that their dot product closely approximates the logarithm of the probability of co-occurrence between words i and j.

The GloVe model trains the word vectors and bias terms by minimizing this loss function, which leads to word embeddings that effectively represent the semantic connections between words, derived from their co-occurrence statistics. The resultant equation obtained by optimizing the loss function is depicted in Eq. (5).
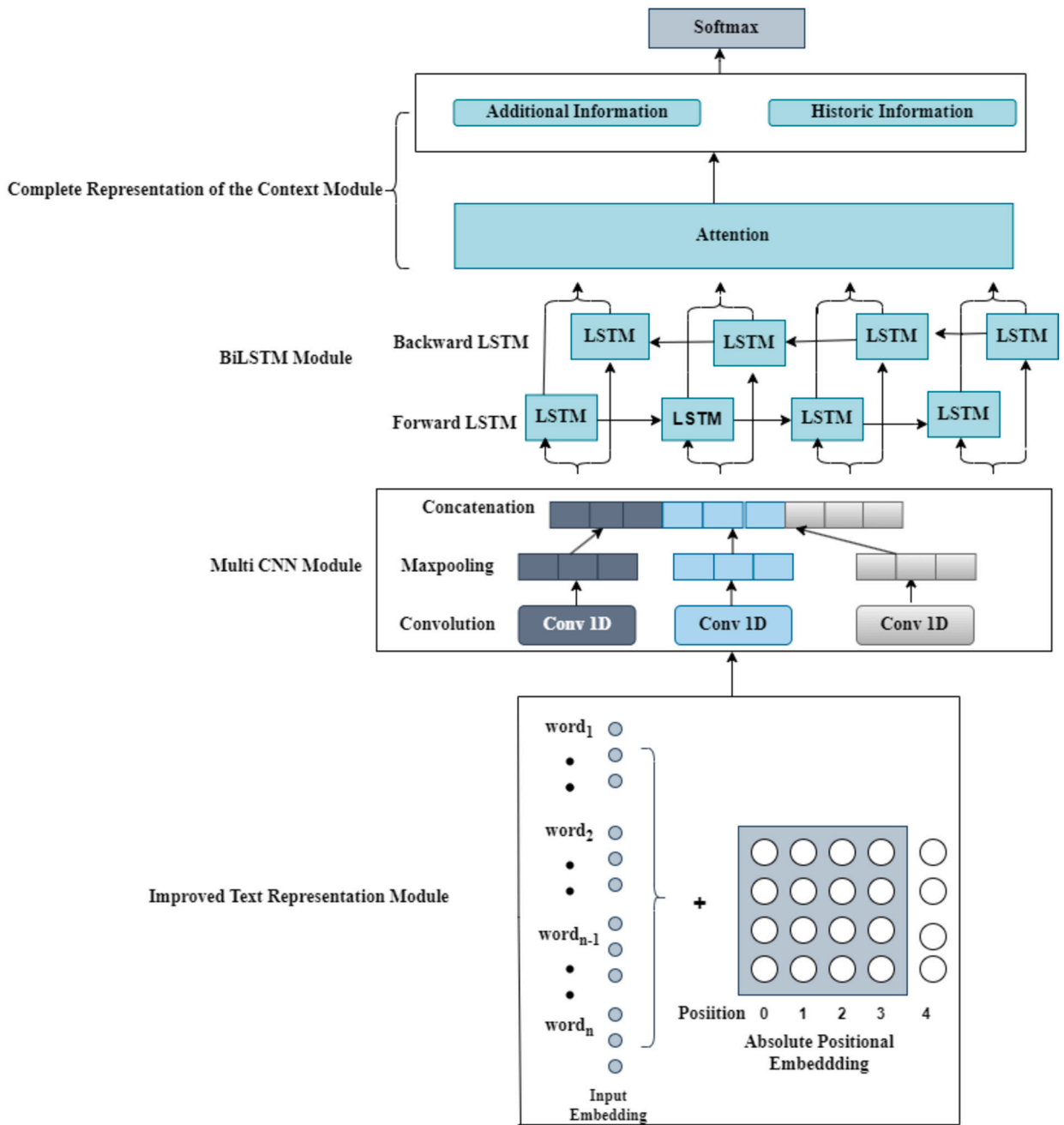
Fig. 1. Proposed Model of PEW-MCAB.

Optimization aims to reduce the loss function given below:

$$J = \sum_{i,j=1}^{|w|} f(X_{ij})(w_i^T w_j' + b_i + b_j' - log(X_{ij}))^2 \tag{5}$$

- w and w' depict the word vectors for i and j respectively.
- b_i and b' depict the bias term.
- f depicts the Weight function that reduces the impact of highly frequent co-occurrences.

The input embedding is further incorporated with Absolute Positional embedding (APE) to activate positional information. Positional encoding assigns a distinct representation to the position of each entity inside a sequence. Thus, positional embedding was

utilized to incorporate word order information. Furthermore, we improve the Glove embedding by integrating it with the positional embedding to better the precision of sentiment categorization. An effective method for obtaining shift invariance when utilizing absolute locations involves introducing random position shifts during the training process. An analogous concept can be observed in question-answering in natural language processing [56]. APE assigns a distinct embedding to each position. APE calculates the input representation by adding together the word embedding and the position embedding for each token. For each location in the input sequence of the model, we create a positional vector by applying the sine and cosine functions as outlined in the Absolute Positional Embedding equations. This vector represents the semantic order of a word within a sentence.

To calculate the absolute positional embedding function $PE(i, 2k)$ and $PE(i, 2k + 1)$, a series of input tokens is needed $x_1, x_2, x_3, x_4, \ldots x_n$ and the position i.

$$(PE(i, 2k) = \sin(\frac{i}{10000^{2k/model}}) \tag{6}$$

$$(PE(i, 2k + 1) = \cos(\frac{i}{10000^{2k/model}}) \tag{7}$$

- i depicts the location in the sequence.
- k depicts the position of the positional encoding.
- d_{model} depicts the dimensions used to represent the positional information.

For every word in the input sequence, we combine the GloVe vector with its corresponding positional vector. As a consequence, a new vector is generated that encompasses both the semantic significance of the word and its positional value inside the sentence. The concatenated vector c_i for a word at position i, with a GloVe vector w_i, and a positional vector p_i, can be calculated as follows:

$$c_i = w_i + p_i \tag{8}$$

3.8. Scalable MultiChannel CNN and attention based BiLSTM

Concatenated vector is inputted into a multi-channel convolutional neural network (CNN). The proposed approach utilizes a scalable multichannel convolutional layer to reduce the number of dimensions in the input data and obtain sequence information. The convolution procedure in the convolutional layer is performed unidirectionally. The convolutional layers utilize kernel sizes of (128 x 4), (128 x 6), and (128 x 8) to extract features from the textual representation. During this step, several sequences that capture semantic properties are constructed. The convolutional layer is a crucial element in any Convolutional Neural Network (CNN) as it is responsible for interpreting the CNN's parameters and constructing channels from various trainable neurons. The channel takes into account the dimensions of the input and calculates the output between the filter and the input pixels using the dot product technique. In the CNN architecture, the convolutional layer performs convolutional operations on the input matrix, allowing semantic fragments to be synthesized [54,55]. By identifying a distinct feature category at a certain location on the feature map, the framework can gain an understanding of the subsequent channels that are generated.

The input tensor x is defined in relation to the filter tensor w and the output tensor C.

$$C = f((X * W) + b) \tag{9}$$

- f depicts the activation function
- X depicts the input tensor
- W depicts filter with dimensions
- b depicts the bias tensor

Let C_1, C_2, and C_3 denote the output tensors of the three Convolutional Neural Network (CNN) layers. These multi-channel convolutional operations are depicted in Eqs. (10)-(12). The filter tensors for each layer are denoted as W1, W2, and W3, while the bias of the tensors is represented by b_1, b_2, and b_3. The convolution action is represented by the symbol $*$, and the activation function (ReLU) is applied to each element of the output obtained from the convolution process.

$$C_1 = ReLU((X * W_1) + b_1) \tag{10}$$

$$C_2 = ReLU((C_1 * W_2) + b_2) \tag{11}$$

$$C_3 = ReLU((C_2 * W_3) + b_3) \tag{12}$$

The BiLSTM model is a sophisticated adaptation of the conventional Recurrent Neural Network (RNN). Its design is tailored to handle sequential data, making it well-suited for jobs involving natural language processing. LSTMs, unlike regular RNNs, possess a gated architecture that enables them to retain information about long-term dependencies. This is of utmost importance in sentiment analysis, as the sentiment frequently depends on the context that may extend across multiple sentences. The bidirectional nature of BiLSTMs enhances the analysis of the text by considering information from both ways, hence assuring comprehensive capture of the contextual information. Nevertheless, BiLSTMs are not without their constraints. In sentiment analysis, all segments of the input sequence are given equal importance, but this is not always the norm. In this vein, we utilized the attention-based BiLSTM.

The attention-based Bidirectional Long Short-Term Memory (BiLSTM) model computes the attention weights $\alpha = (\alpha_1, \alpha_2, \alpha_3....., \alpha_n)$ and the context vector c by applying the following equations to the input sequence $X = (x_1, x_2, x_3...., x_n)$. The Bidirectional LSTM (BiLSTM) calculates the hidden states in both the forward and backward directions, these operations are depicted in Eqs. (13) and (14). Thus, the integration of Attention-based BiLSTM can effectively handle sequential tasks and semantics in text and yield excellent results [53].

$$\overrightarrow{h_1} = LSTM(x_i, \overrightarrow{h}_{i-1}) \tag{13}$$

$$\overleftarrow{h_1} = LSTM(x_i, \overleftarrow{h}_{i+1}) \tag{14}$$

The attention employed enables the model to zero in on the most pertinent aspects of the input data, similar to how people prioritize specific words or phrases when reading a text. The attention mechanism allocates weights to various segments of the input sequence, emphasizing those that are more reflective of the overall sentiment. The incorporation of the attention mechanism into a BiLSTM model yields a potent tool for our sentiment analysis task. The model possesses the capability to give priority to important information while retaining the capacity to store and analyze input from both directions across extended sequences. This leads to a heightened level of precision and subtlety in comprehending sentiment. The mathematical representation for obtaining the attention scores is depicted in Eq. (15).

The attention scores e_i are computed with a score function, represented in Eq. (16):

$$e_i = score(\overrightarrow{h_i}, \overleftarrow{h_i}) \tag{15}$$

The context vector c is computed by summing the obscured states, with each state being multiplied by its corresponding attention weight:

$$c = \sum_{i=1}^{n} \alpha_i \cdot [\overrightarrow{h_i}, \overleftarrow{h_i}] \tag{16}$$

The last element of this architectural design is the softmax classifier. Once the attention-based BiLSTM has completed processing the input, the softmax classifier assumes control to classify the sentiment. The BiLSTM output is converted into probabilities that collectively add up to one, thereby identifying the most probable sentiment category for the provided text. Eqs. (17) and (18) depict the softmax function and the operation softmax uses for predicting sentiment labels. The attention weights, denoted as α_i, are derived by applying the softmax function to the scores:

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{n} \exp(e_j)} \tag{17}$$

The softmax classifier uses the context vector to make predictions about the sentiment label.

$$\hat{y} = softmax(W_c \cdot c + b_c) \tag{18}$$

- x_i depicts the input at time step i
- α_i depict the attention weights
- c depicts the context vector
- $\hat{y}$ depicts sentiment label
- W_c depicts the weight of softmax classifier
- b_c depicts the bias parameter of the softmax classifier

4. Experimental setup

In this section, we will detail the specific experimental setups that will be used for the rest of the study. Following this, we give the datasets that were utilized to conduct the experimental tests of the aforementioned method. In what follows, we take a close look at the dataset filtering pre-processing options. After that, the baselines used for comparison using the previous method are presented. The accuracy, precision, recall, and F1-measure are the measures used in this study.

4.1. Experimental results

This research presents a sentiment classification of online movie reviews using the PEW-MCAB method. The PEWoVe model is an enhanced and resilient text representation that can effectively comprehend ambiguities and information sequencing in textual data by establishing contextual understanding through referencing the textual context. The PEW-MCAB model incorporated a convolutional neural network with multiple channels that utilized distinct kernel dimensions to extract features. The model's multichannel convolutional neural network improved feature extraction by utilizing max pooling and convolutional layers, and successfully identified minor grammatical patterns within the training set. Consequently, these components collaborate to improve the performance of the suggested methodology in the experiment. The attention-based Bidirectional Long Short-Term Memory (BiLSTM) model is employed to analyze the connections between emotional words and their surrounding words to comprehensively comprehend the

context. This improves the precision of our model. Due to the cumulative influence of multiple input parameters, this strategy is the most dependable and comprehensive. The efficiency of the PEW-MCAB is assessed in this study using four widely used movie datasets: IMDB movie review, and Movie Review (2002, 2004, and 2005) versions datasets. To the best of our knowledge and skill, we assert that our method is among the pioneering approaches to extract and classify sentiment data at the document level utilizing the PEW-MCAB. Cross-validation is a notable approach for assessing the effectiveness of a model in predicting unseen data. To optimize the model, we partitioned the training data for each dataset through the use of cross-validation (ten folds) [52]. Eight of the subsets are designated for training and two are designated for validation. They are equally distributed and mutually exclusive. Every baseline study's experimental designs were closely examined by the PEW-MCAB. The proposed methodology aims to ascertain the impact of different parameter configurations on performance outcomes during the analysis. The PEW-MCAB algorithm obtains optimal parameter settings by doing a comprehensive examination of both the baseline approaches and the method of parametric optimization. To do this, it uses approaches for parameter optimization and similar investigations. Our models were trained and evaluated sequentially, starting with Glove, followed by PEW, PEW-C, PEW-MC, and finally PEW-MCAB. This document highlights the performance of four datasets, namely Glove, PEW, PEW-C, PEW-MC, and PEW-MCAB. The tabular notations demonstrate the excellent achievements of the B-MLCNN when compared to the acquisition of baseline methods utilizing standardized datasets. This demonstrates its empirical advancement in all of the applicable domains.

4.2. Experiment configurations

A graphics processing unit (GPU) and 16 GB of random access memory (RAM) from NVIDIA were used in the experimental arrangement. A 2.30 GHz dual-core Intel Core i7 serves as the central processing unit. All of the environment configurations were made in the Jupyter Notebook IDE with Python 3.6.9 and Keras 2.4.3.

4.3. Comparative study models

- **PEW**: This model utilizes a glove pretrain model with positional embedding to create a global matrix cooccurrence matrix representation of text. This approach incorporates word order information into the representation.
- **PEW-C**: The model describes the use of a glove pretrain model and positional embedding to build a global matrix cooccurrence matrix. This approach incorporates word order information into the representation. The enhanced text representation is fused with a Convolutional Neural Network to augment feature extraction.
- **PEW-MC**: The model refers to the utilization of a glove pretrain model and positional embedding technique to enhance the global matrix cooccurrence matrix text representation. This approach aims to incorporate word order information into the existing representation. An improved feature extraction method is implemented in the text representation through the use of a multichannel Convolutional Neural Network (CNN) with distinct kernel sizes.
- **PEW-MCAB**: This model involves the use of a glove pretrain model and positional embedding technique to improve the representation of the global matrix cooccurrence matrix in text. The objective of this strategy is to integrate word order information into the current representation. To increase the extraction of features, the text representation is augmented by including a multichannel Convolutional Neural Network (CNN) using a range of kernel sizes. To improve the understanding of the relationship between words and prioritize key parts of a sentence in a text review, the model utilized bidirectional long short-term memory that is attention-based.

4.4. Comparative study models

The PEW, PEW-C, and PEW-MC models are being compared with the PEW-MCAB model. The four models are assessed using identical training and test sets to see which model achieved the highest degree of accuracy. Additionally, Table 2 displays the performance metrics of the IMDb dataset for PEW, PEW-C, PEW-MC, and PEW-MCAB. The PEW model exhibits an accuracy of 86.4%, precision of 87.9%, recall of 85.3%, and an F-1 score of 86.5%. The results for PEW-C exhibit a high level of consistency, with values of 86.9%, 88.0%, 86.1%, and 86.9%. PEW-MC recorded consistent percentages of 87.7%, 88.2%, 87.2%, and 87.7%. The PEW-MCAB framework we proposed demonstrated an accuracy of 90.3%, precision of 91.5%, recall of 89.3.3%, and an F1 score of 90.3%. Table 3 contains the performance metrics of PEW, PEW-C, PEW-MC, and PEW-MCAB on the MR (2002) dataset. The PEW model attains an F-1 score of 81.8%, an accuracy of 81.2%, a precision of 85.0%, and a recall of 78.8%. The PEW-C consistently produces results of 83.7%, 83.1%, 86.4%, and 81.2%. PEW-MC achieved a continuous accuracy rate of 83.7%, 87.1%, 87.1%, and 84.4% on, precision, recall, and f-1 score respectively. The proposed framework (PEW-MCAB) achieved an accuracy of 84.1%, precision of 87.8%, recall of 82.5%, and an F1 score of 85.1% in this case. Table 4 displays the results for the MR (2004) dataset for PEW, PEW-C, PEW-MC, and PEW-MCAB. The PEW model achieves an accuracy of 83.4%, precision of 87.5%, recall of 81.0%, and an F-1 score of 84.4%. The PEW-C model has an accuracy rate of 84.1%, a precision rate of 88.0%, a recall rate of 81.9%, and an F-1 score rate of 84.8%. PEW-MC likewise yielded similar outcomes, with accuracies of 84.9%, 88.5%, 83.1%, and 85.7%. The PEW-MCAB framework demonstrated an accuracy of 85.9%, precision of 89.5%, recall of 83.6%, and an F1 score of 86.4%. The performance indicators of PEW, PEW-C, PEW-MC, and PEW-MCAB on the MR (2005) dataset are presented in Table 5. The PEW performance exhibits an accuracy rate of 85.9%, a precision rate of 87.1%, a recall rate of 84.9%, and an F-1 score of 85.9%. The findings achieved utilizing PEW-C are similarly comparable, with percentages of 84.1%, 86.8%, 84.8%, and 85.6%. The scores recorded by PEW-MC were 86.2%, 87.4%, 85.3%, and 86.3%. The PEW-MCAB framework demonstrated an accuracy of 87.1%, precision of 88.2%, recall of 86.2%, and an F1 score of 87.2%.

Table 2
Empirical Findings on IMDB dataset.

Model	Accuracy	Precision	Recall	F-1 score
PEW	86.4	87.9	85.3	86.5
PEW-C	86.9	88.0	86.1	86.9
PEW-MC	87.7	88.2	87.2	87.7
PEW-MCAB	90.3	91.5	89.3	90.3

Table 3
Empirical Findings on MR (2002) dataset.

Model	Accuracy	Precision	Recall	F-1 score
PEW	81.2	85.0	78.8	81.8
PEW + C	83.1	86.4	81.2	83.7
PEW + MC	83.7	87.1	81.8	84.4
PEW +MCAB	84.1	87.8	82.5	85.1

Table 4
Empirical Findings on MR(2004) dataset.

Model	Accuracy	Precision	Recall	F-1 score
PEW	83.4	87.5	81.0	84.4
PEW + C	84.1	88.0	81.9	84.8
PEW + MC	84.9	88.5	83.1	85.7
PEW +MCAB	85.9	89.5	83.6	86.4

Table 5
Empirical Findings on MR (2005) dataset.

Model	Accuracy	Precision	Recall	F-1 score
PEW	85.9	87.1	84.9	85.9
PEW + C	84.1	86.8	84.8	85.6
PEW + MC	86.2	87.4	85.3	86.3
PEW +MCAB	87.1	88.2	86.2	87.2

4.5. Model comparison

From the accuracy data presented in Table 6, we draw several conclusions regarding the performance of the different models on the IMDB, MR (2002), MR (2004), and MR (2005) datasets. Firstly, on the IMDB dataset, the B-MLCNN, B-CNN, and BERT models exhibit superior performance compared to all other models in terms of overall performance. The BERT approach may outperform other neural networks because of its pre-trained contextual understanding of language, combined with the multi-channel CNN that improves feature extraction in text through several convolutional neural network (CNN) layers. The proposed technique demonstrated remarkable accuracy in measurements on the IMDB dataset. It surpasses most of the basic methods with a substantial margin of achievement. The accuracy rates for the different models on the IMDB dataset are as follows: 3.7% for MNB, 1.1% for SVM, 3.3% for LSTM, 2.4% for Bi-LSTM, 2.2% for Multichannel CNN, 0.8% for Glove CNN-BiLSTM, 8.6% for LSTM, and 5.3% for BiLSTM. Furthermore, among the several models evaluated on the MR (2002) and MR (2004) datasets, the B-MLCNN, B-CNN, and BERT models demonstrate superior performance in terms of overall effectiveness. Among all the models evaluated on the MR (2005) dataset, the B-MLCNN, B-CNN, and BERT models demonstrate superior performance in terms of overall effectiveness. The proposed technique demonstrated remarkable accuracy performance on the MR (2005) dataset. It surpasses most of the traditional techniques with an impressive degree of achievement. The percentages of contributions from different models are as follows: 9.4% from RAE, 8.1% from MNB, 7.7% from NBSVM, 7% from LSTM, 6.8% from Bi-LSTM, 3.9% from AC-BiLSTM, 6.1% from Static CNN, 5.6% from Non-Static CNN, 6% from Multichannel CNN, 2.7% from Glove CNN-BiLSTM, 14.4% from LSTM, and 12.4% from BiLSTM. It is worth noting that LSTM achieved the lowest accuracy on the MR (2002) and MR (2005) datasets. The LSTM model is probably unable to generate a more significant output by integrating layers from both forward and backward directions compared to other neural networks that employ BiLSTTM and attention.

4.6. Discussion

This research paper introduces the PEW-MCAB approach, an enhanced deep-learning algorithm designed specifically for performing sentiment analysis on online movie reviews. Subsequently, we seek to analyze the possible factors contributing to the higher performance of the proposed PEW-MCAB model in comparison to existing baseline models. The PEW-MCAB method integrates an

Table 6

Performance comparison in term of overall accuracy (%) of various Glove embedding dimensions on all datasets.

Model	IMDB	MR (2002)	MR (2004)	(2005)
RAE [45]	-	-	-	77.7
MNB [46]	86.6	-	-	79.0
SVM [47]	89.2	-	-	-
NBSVM [46]	91.2	-	-	79.4
MTL-ML4 [51]				
LSTM [40]	87.0			80.1
BiLSTM [40]	87.9	-	-	80.3
AC-BiLSTM [40]	91.8	-	-	83.2
Static CNN [48]	-	-	-	81.0
Non-Static CNN [48]	-	-	-	81.5
Multichannel CNN [48]	88.1	88.4	81.7	81.1
BERT [49]	94.0	88.0	89.0	-
CNN [49]	91.1	84.2	89.1	-
B-CNN [49]	93.0	87.1	90.3	-
B-MLCNN [49]	95.1	88.2	95.0	-
Glove CNN-BiLSTM [38]	89.5	-	-	84.4
LSTM [50]	81.7	-	-	72.7
BiLSTM [50]	85.0	-	-	74.7
PEW (Proposed)	86.4	81.2	83.4	85.9
PEW + C (Proposed)	86.9	83.1	84.1	84.1
PEW + MC (Proposed)	87.7	83.7	84.9	86.2
PEW +MCAB (Proposed)	90.3	84.1	85.9	87.1

enhanced feature extraction technique that combines positional embedding and glove-pre-trained embedding. Additionally, it adds a Multichannel CNN and an attention-based Bidirectional LSTM. In terms of accuracy, precision, recall, F-1 score, and confusion matrix analysis, the proposed sentiment analysis model performed well. Figs. 2–5 display the confusion matrix for each of the baseline techniques tested, across all four datasets. The confusion matrix displays the predicted sentiment categorization results for PEW, PEW-C, PEWW-MC, and PEW-MCAB summaries in matrix format. The display indicates the number of accurate and inaccurate predictions for each class. It aids in comprehending the sentiment labels or classes the model mistakenly identifies as a different class.

Fig. 2–5 shows that adding enhanced word embeddings to PEW, PEW-C, PEW-MC, and PEW-MCAB resulted in good performance by the baselines. The confusion matrix of comparative models for the IMDB dataset is depicted in Fig. 2 (a, b, c, and d). Fig. 3 (a, b, c and d) depicts the confusion matrix for the MR (2002) dataset. Fig. 4(a, b, c and d) depicts the confusion matrix for the MR (2004) dataset, and Fig. 5 (a, b, c and d) depicts the confusion matrix of comparative study models for MR (2005) dataset. The PEW model displayed some weaknesses in classification due to the absence of the local and global semantic feature extraction layers. Similarly, the PEW-C model struggled a bit after including the local semantic feature extraction layer. Compared with the PEW and PEW-C, the PEW-MC and PEW-MCAB displayed superior performance in classification. An important reason for this can be attributed to the dynamic integration of enhanced word embedding, a multi-layer CNN, and an attention-based BiLSTM.

Tables 2–5 display the evaluation criteria for the datasets acquired from the PEW, PEW-C, PEW-MC, and PEW-MCAB research. The evaluation procedure gauges the model's performance by utilizing parameters like accuracy, precision, recall, and f-1 score. These metrics are calculated using standardized datasets. The experimental results demonstrate that the PEW-MCAB methodology displays the utmost precision, with a precision rate of 91.5%. Similarly, the PEW-MC method exhibits a similar degree of accuracy, measuring at 88.2%. Conversely, the PEW-MC approach attains a little lower precision rate of 88%. Finally, the PEW strategy demonstrates the lowest level of precision compared to the other tested methods, with a minimal precision rate of 87.9% when utilized on the IMDB dataset. The recall values for PEW, PEW-C, PEW-MC, and PEW-MCAB are 89.3%, 87.2%, 86.1%, and 85.3% correspondingly. The F-1 scores obtained successively on the IMDB dataset were 90.3%, 87.7%, 86.9%, and 86.5%. The proposed model achieves a precision rate of 87.8% on the MR (2002) dataset, which is the highest among all models. The PEW-MC and PEW-C models attain precision rates of 87.1% and 86.4% respectively, while the PEW model achieves a slightly lower precision rate of 85.0%. The recall values for PEW, PEW-C, PEW-MC, and PEW-MCAB are 82.5%, 81.8%, 81.2%, and 78.8% correspondingly. The F-1 scores obtained were 85.1%, 84.4%, 83.7%, and 81.8% for the MR (2002) dataset. The PEW-MCAB has the highest precision rate of 89.5% on the MR (2004) dataset, followed by PEW-MC with 88.5% precision, PEW-C with 88.0% precision, and PEW with a somewhat lower precision rate of 87.5%. The recall values for PEW, PEW-C, PEW-MC, and PEW-MCAB are 83.6%, 83.1%, 81.9%, and 81.0% respectively. The F-1 score had sequential values of 86.4%, 85.7%, 84.8%, and 84.4% when assessed on the MR (2004) dataset. The PEW-MCAB model achieves the highest precision rate of 88.2% on the MR (2005) dataset. It is followed by PEW-MC with a precision rate of 88.2%, PEW with a precision rate of 87.1%, and PEW-CC with the lowest precision rate of 86.8%. The recall values for PEW, PEW-C, PEW-MC, and PEW-MCAB are 86.2%, 85.3%, 84.9%, and 84.8% respectively. The F-1 score achieved sequential values of 87.2%, 86.3%, 85.9%, and 85.6% when evaluated on the MR (2005) dataset. The precision indicates that the PEW-MCAB model predicts sentiment labels more
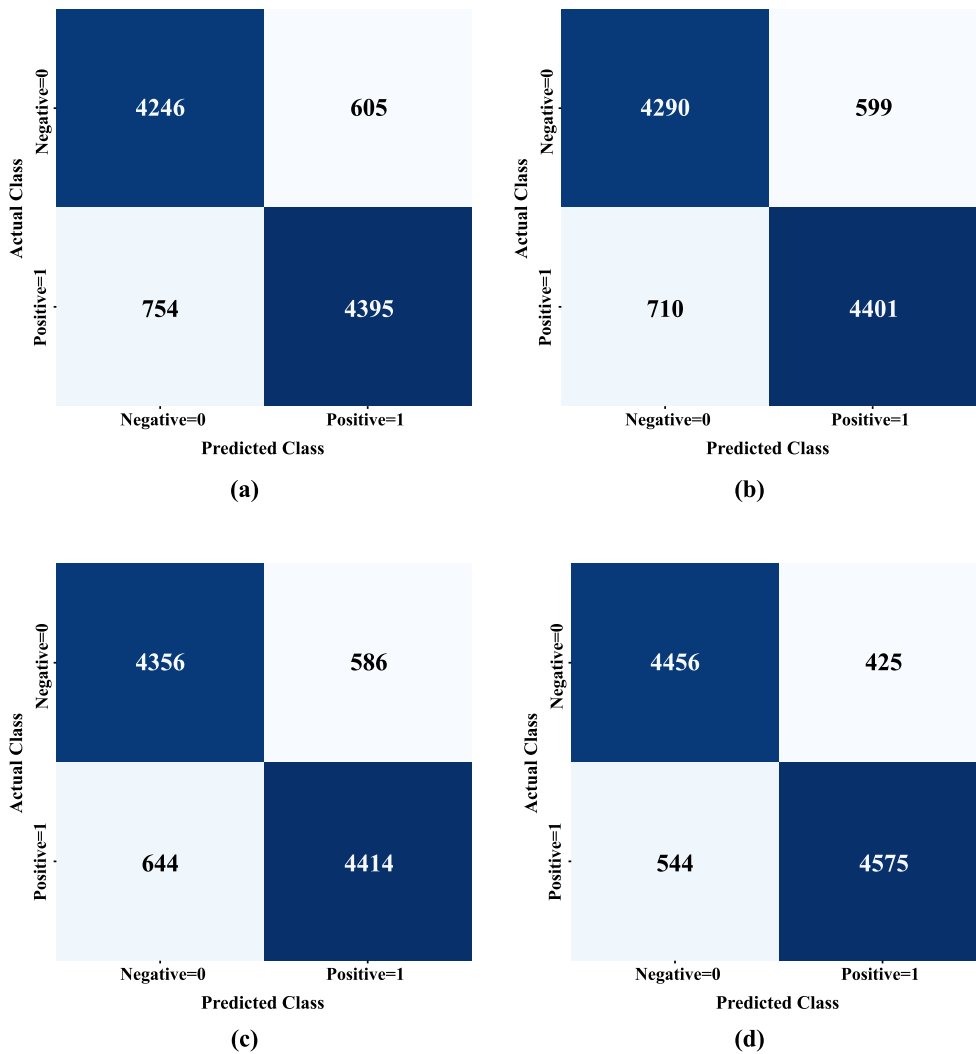
Fig. 2. Confusion Matrix of Models for IMDB Dataset: (a) PEW (b)PEW-C (c)PEW-MC (d)PEW-MCAB.

accurately than the other baseline models. Consequently, when the PEW-MCAB predicts a positive sentiment, it is likely to be accurate.

The recall values suggest that the PEW-MCAB is more effective at capturing all positive sentiments compared to other baseline models. It follows that the recall values of all baseline models showed fewer missing positive sentiments that should have been detected. A comparison of the F-1 scores obtained from all baseline models demonstrated that the proposed sentiment analysis model is competitive. On average, all baseline models recorded an F-1 score of over 86%, demonstrating a balance between precision and recall as well as a robust sentiment analysis model. Fig. 6 (a, b, c, and d) depicts the performance of all movie review datasets employed in the experiment with respect to the evaluation metrics: precision, recall, and F-1 score.

Fig. 7 illustrates the Receiver Operating Characteristic (ROC) curve, which visually displays the performance of the classification model in comparison to the baseline methods across the four datasets. Based on a classification threshold, the ROC curve illustrates the true positive rate in relation to the false positive rate.

The AUC values for the PEW, PEW-C, PEW-MC, and PEW-MCAB models applied to the IMDB dataset, as shown in Fig. 7a, are 87.5%, 88.2%, 88.7%, and 91.4%, respectively. The MR (2002) dataset includes the following percentages: 83.8% for PEW, 86.5% for PEW-C, 87.1% for PEW-MC, and 88.1% for PEW-MCAB, as displayed in Fig. 7b. The MR (2004) dataset in Fig. 7c produced receiver operating characteristic (ROC) scores of 85.2%, 86.2%, 86.9%, and 87.8%. The ROC curve values for the MR (2005) dataset are as follows: PEW - 87.4%, PEW-C - 88.1%, PEW-MC - 88.9%, and PEW-MCAB - 89.5%. These results are displayed in Fig. 7d.
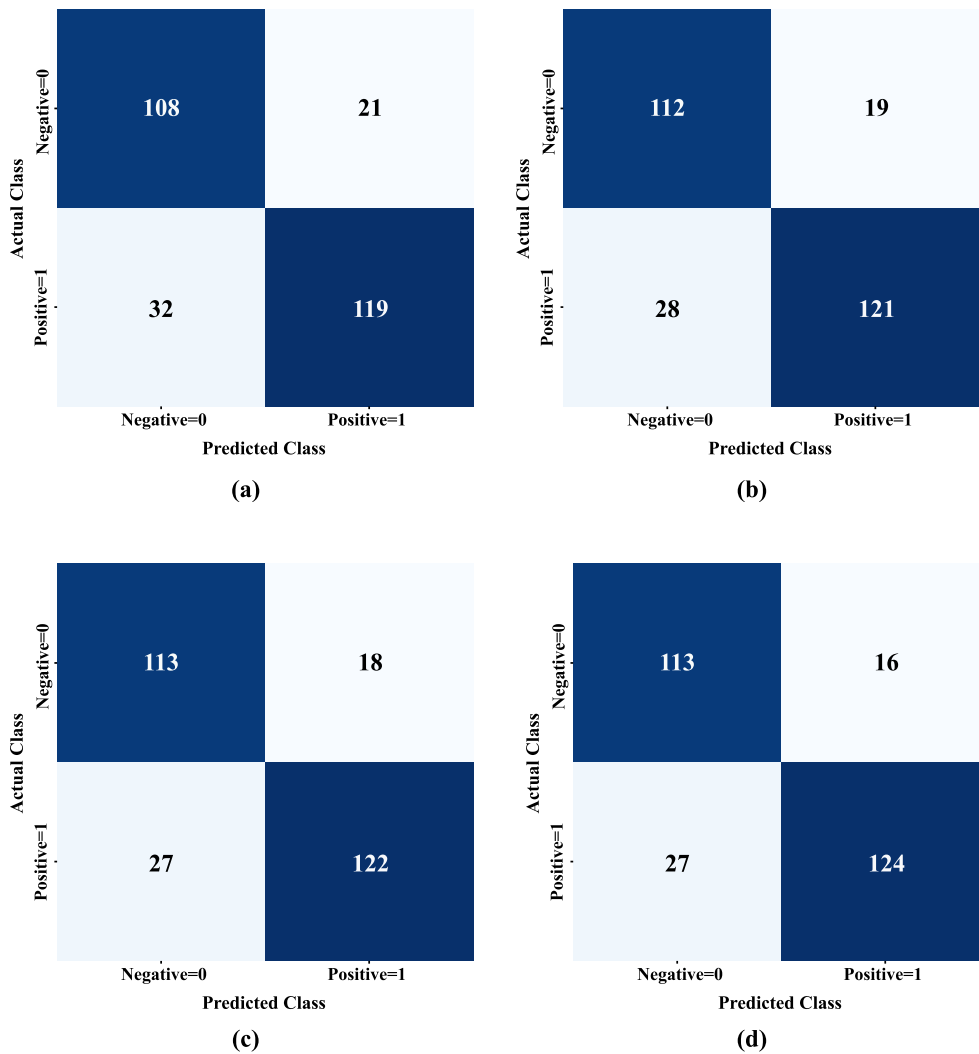
Fig. 3. Confusion Matrix of Models for MR (2002) Dataset:(a) PEW (b)PEW-C (c)PEW-MC (d)PEW-MCAB.

From the ROC curves, it is apparent that the proposed model across all four datasets is closer to the top-left corner or closer to 1, indicating the model performed better in classifying sentiments and capable of distinguishing between positive and negative sentiments.

The aforementioned findings indicate that the proposed methodology is dependable and has the capacity to accurately categorize sentiments. The PEW-MCAB method proposes an enhanced approach to sentiment classification. This is accomplished by extracting high-level features at many scales and integrating both word order information (Positional Embedding) and global cooccurrence matrix (Glove pre-trained embedding), together with Multichannel CNN and attention-based BiLSTM. Thus, this approach aims to acquire fundamental contextual information and the key contextual characteristics at different levels of magnitude. We conduct a comprehensive evaluation of the performance of PEW-MCAB in sentiment categorization using multiple measures across four online movie review datasets. The framework demonstrates a substantial enhancement in performance across these datasets, in contrast to the bulk of the fundamental approaches. The above illustration showcases the model's ability to improve precision in conducting sentiment classification at a document level. This scientific framework offers a robust basis and showcases the potential for successful application in practical situations. The empirical analysis conducted on various online movie datasets shows the importance of incorporating broad feature connectivity and utilizing Positional embedding with Glove, multichannel convolutional neural networks (CNN) along with attention-based bidirectional long-short-term memory (BiLSTM) models. Our research achieved superior results by modifying the algorithm to address the challenges of sentiment analysis conducted at the document level for complicated online movie review datasets, compared to other methods.
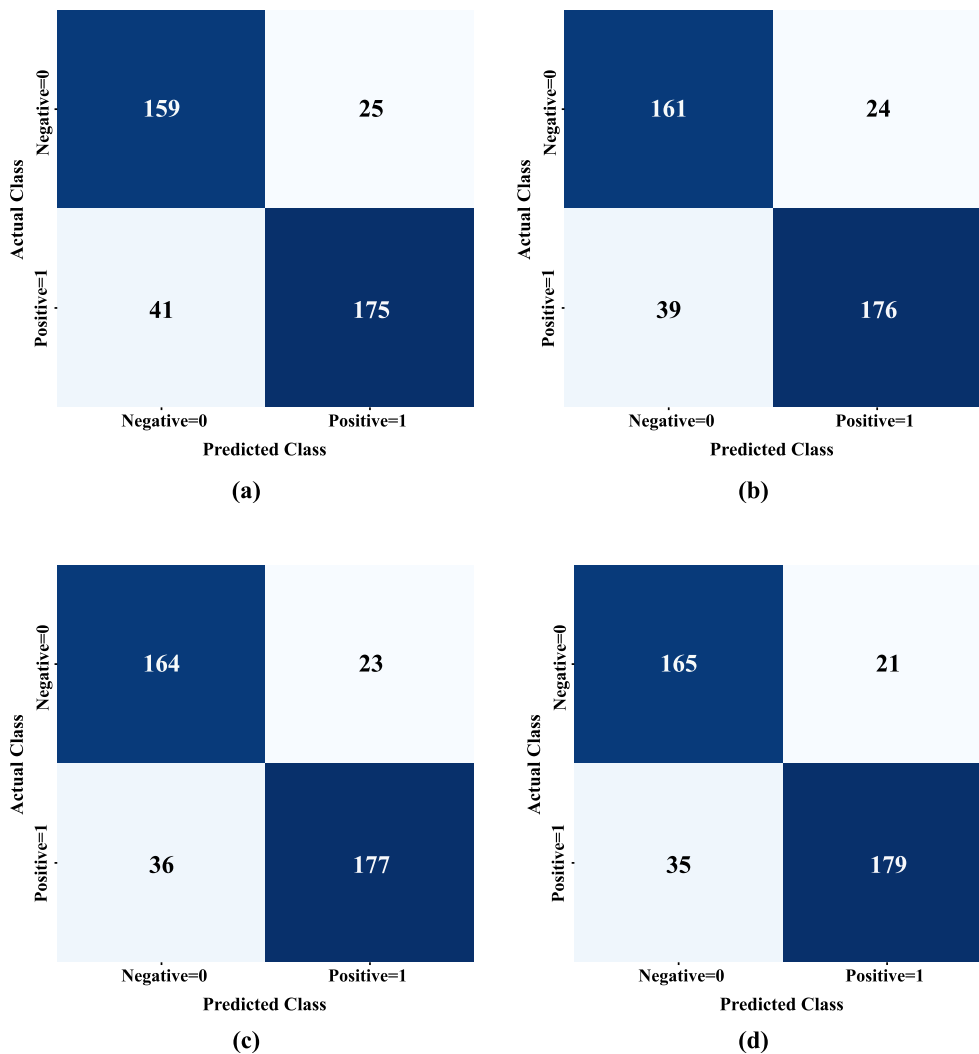
Fig. 4. Confusion Matrix of Models for MR (2004) Dataset: (a)PEW (b)PEW-C (c)PEW-MC (d)PEW-MCAB.

5. Conclusion

Several factors make it difficult to accurately identify sentiment, including the absence of word order information in word embeddings and the nuances associated with sentences, which can lead to weaker sentiments being expressed. The study developed a sentiment classification model to address these challenges by studying the word information order and analyzing the content of online movie text reviews in order to interpret the hidden meanings of words. This work introduces a model called PEW-CAB that combines word order information from textual reviews with Glove pretrain word representation. The effectiveness of the model is enhanced by combining multichannel CNN and attention-based BiLSTM. Four movie datasets were subjected to experiments, wherein four distinct assessment criteria were employed, namely accuracy, precision, recall, and F1-measure. The experimental results demonstrated that incorporating multichannel convolution and attention-based bidirectional long-short-term memory significantly enhanced the model's performance in sentiment analysis. The classification performance of the PEW-MCAB model surpassed the majority of the baselines. The present research focuses on enhancing global characteristics for our text representation problem by utilizing word order information. To enhance the integration of text data with positional information, future research will investigate how to integrate positional embeddings with other encodings. Additionally, future research will examine how to effectively apply regulation techniques to optimize positional embeddings in achieving better accuracy in sentiment classification.
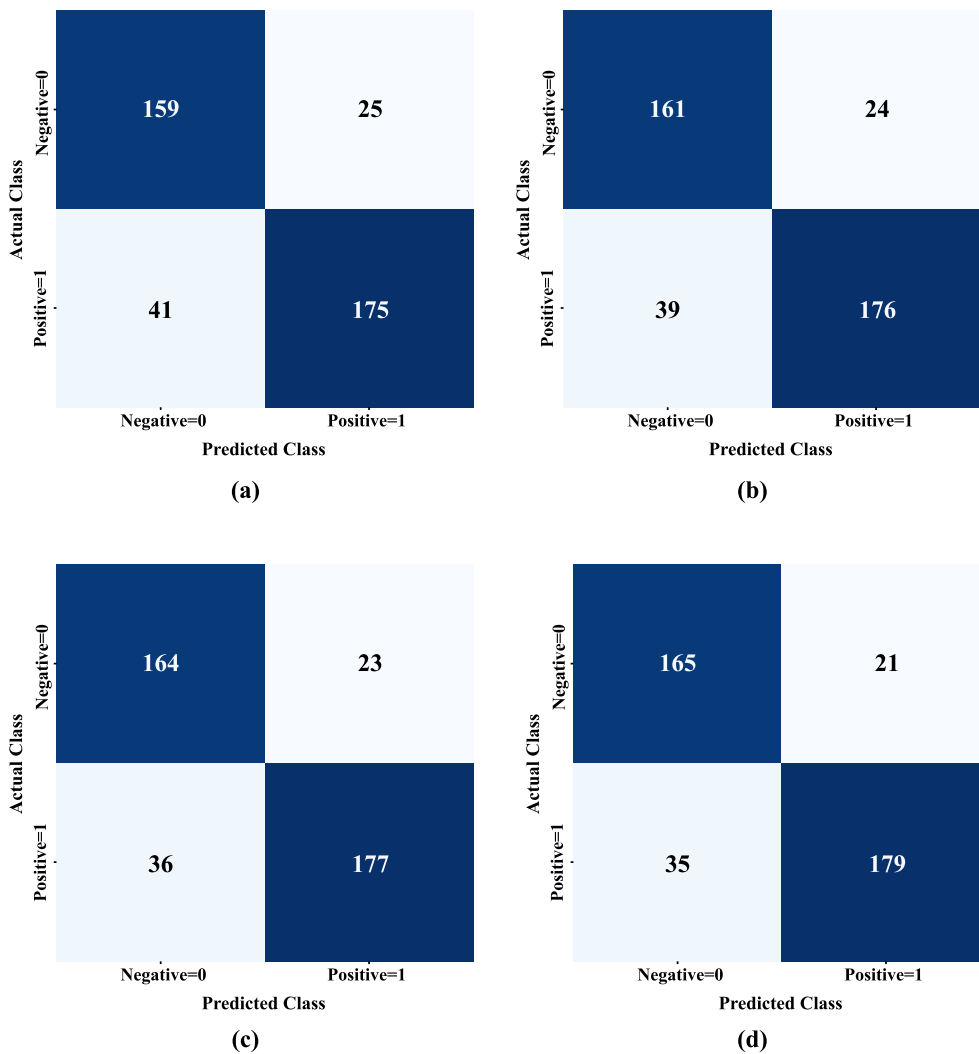
Fig. 5. Confusion Matrix of Models for MR (2005) Dataset: (a)PEW (b)PEW-C (c)PEW-MC (d)PEW-MCAB.

CRediT authorship contribution statement

Peter Atandoh: Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Fengli Zhang:** Writing – review & editing, Validation, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Mugahed A. Al-antari:** Writing – review & editing, Visualization, Project administration, Methodology, Investigation, Funding acquisition. **Daniel Addo:** Visualization, Formal analysis. **Yeong Hyeon Gu:** Methodology, Funding acquisition, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The datasets used in our experiments are available online at the following links: IMDB Reviews at https://ai.stanford.edu/~amaas/data/sentiment/ (accessed in February 2024) and the three movie reviews at https://www.cs.cornell.edu/people/pabo/movie-review-data/ (accessed in February 2024).
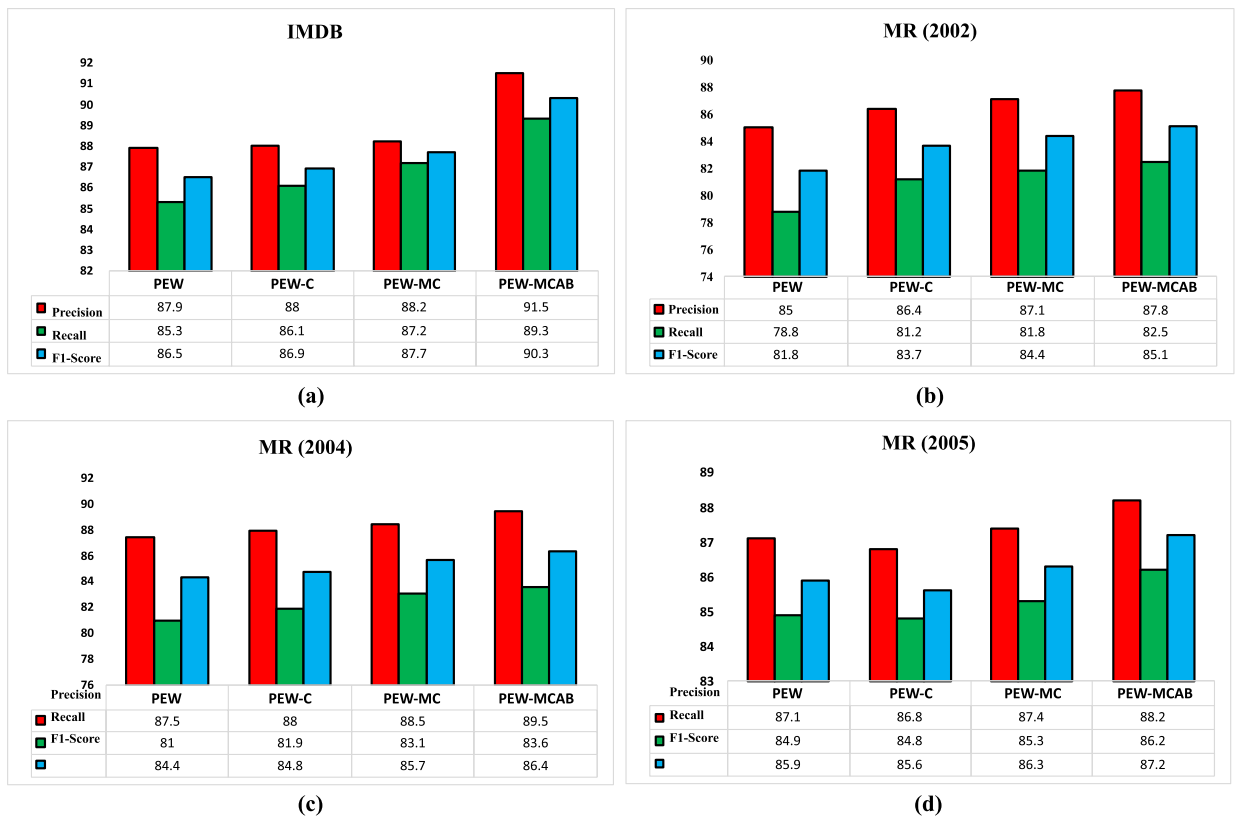
Fig. 6. Comparative Analysis of Models for all Datasets.
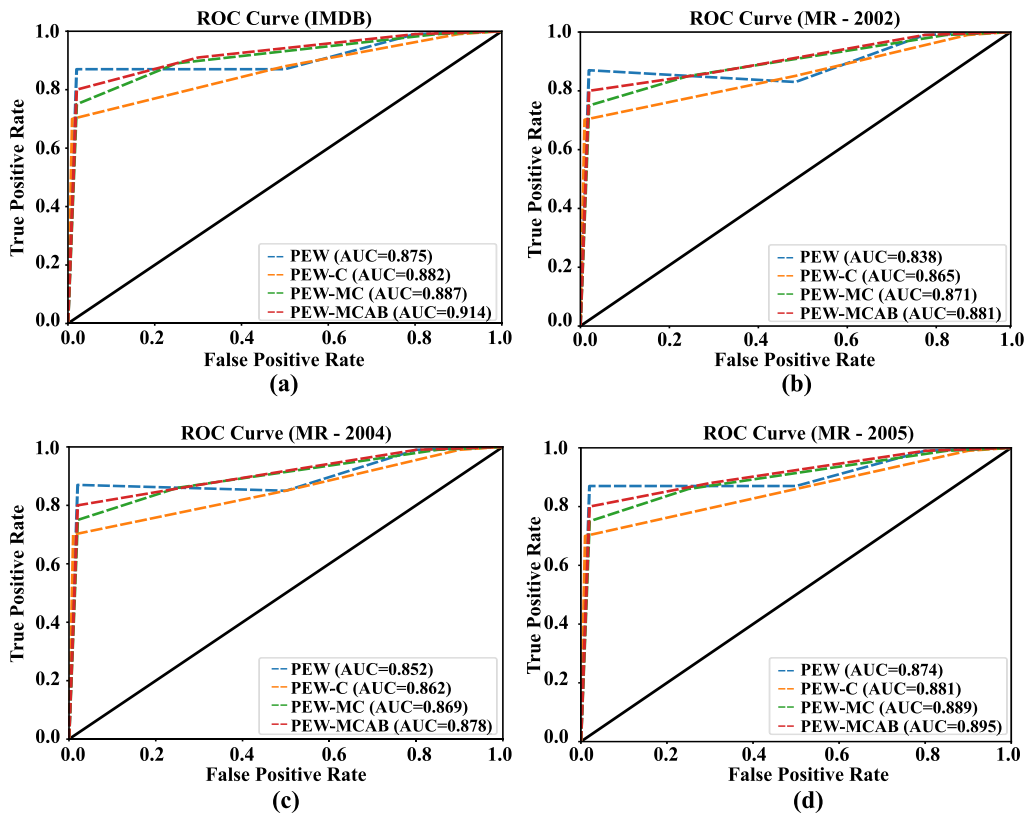
Fig. 7. ROC curves of Baselines Techniques on all Datasets.

Acknowledgements

References

[1] B. Wu, L. Wang, Y. Zeng, Interpretable tourism demand forecasting with temporal fusion transformers amid COVID-19, Appl. Intell. 53 (2023) 14493–14514, https://doi.org/10.1007/s10489-022-04254-0.

[2] J. Khan, A. Alam, Y. Lee, Intelligent hybrid feature selection for textual sentiment classification, IEEE Access 9 (2021) 140590–140608.

[3] S. Naz, A. Sharan, N. Malik, Sentiment classification on Twitter data using support vector machine, in: 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), 2018, pp. 676–679, https://api.semanticscholar.org/CorpusID:58005797.

[4] Y. Zhang, G. Wang, F. Chung, S. Wang, Support vector machines with the known feature-evolution priors, Knowl.-Based Syst. 223 (2021) 107048, https://doi.org/10.1016/j.knosys.2021.107048.

[5] R. Bibi, U. Qamar, M. Ansar, A. Shaheen, Sentiment analysis for urdu news tweets using decision tree, in: 2019 IEEE 17th International Conference on Software Engineering Research, Management and Applications (SERA), 2019, pp. 66–70.

[6] B. Seref, E. Bostanci, Sentiment analysis using naive Bayes and complement naive Bayes classifier algorithms on Hadoop framework, in: 2018 2nd International Symposium On Multidisciplinary Studies And Innovative Technologies (ISMSIT), 2018, pp. 1–7.

[7] S. Zhang, Z. Wei, Y. Wang, T. Liao, Sentiment analysis of Chinese micro-blog text based on extended sentiment dictionary, Future Gener. Comput. Syst. 81 (2018) 395–403.

[8] A. Jentzen, B. Kuckuck, P. Wurstemberger, Mathematical introduction to deep learning: methods, implementations, and theory, CoRR, arXiv:2310.20360 [abs], 2023, https://doi.org/10.48550/arXiv.2310.20360.

[9] J. Mendes-Moreira, T. Mendes-Neves, Towards a Systematic Approach to Design New Ensemble Learning Algorithms, 2024.

[10] Y. Zhang, G. Wang, T. Zhou, X. Huang, S. Lam, J. Sheng, K. Choi, J. Cai, W. Ding, Takagi-Sugeno-Kang fuzzy system fusion: a survey at hierarchical, wide and stacked levels, Inf. Fusion 101 (2024) 101977, https://doi.org/10.1016/j.inffus.2023.101977.

[11] H. Chen, M. Sun, C. Tu, Y. Lin, Z. Liu, Neural sentiment classification with user and product attention, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016, pp. 1650–1659, https://aclanthology.org/D16-1171.

[12] D. Stojanovski, G. Strezoski, G. Madjarov, I. Dimitrovski, I. Chorbev, Deep neural network architecture for sentiment analysis and emotion identification of Twitter messages, Multimed. Tools Appl. 77 (2018) 32213–32242.

[13] Z. Zhang, Y. Zou, C. Gan, Textual sentiment analysis via three different attention convolutional neural networks and cross-modality consistent regression, Neurocomputing 275 (2018) 1407–1415, https://api.semanticscholar.org/CorpusID:20819925.

[14] H. Tao, S. Tong, H. Zhao, T. Xu, B. Jin, Q. Liu, A radical-aware attention-based model for Chinese text classification, in: Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, 2019.

[15] C. Gan, L. Wang, Z. Zhang, Multi-entity sentiment analysis using self-attention based hierarchical dilated convolutional neural network, Future Gener. Comput. Syst. 112 (2020) 116–125, https://www.sciencedirect.com/science/article/pii/S0167739X19329383.

[16] Z. Zhang, L. Wang, Y. Zou, C. Gan, The optimally designed dynamic memory networks for targeted sentiment classification, Neurocomputing 309 (2018) 36–45, https://www.sciencedirect.com/science/article/pii/S0925231218305241.

[17] A. Abdi, S. Shamsuddin, S. Hasan, J. Piran, Deep learning-based sentiment classification of evaluative text based on multi-feature fusion, Inf. Process. Manag. 56 (2019) 1245–1259, https://www.sciencedirect.com/science/article/pii/S0306457318307416.

[18] M. Basiri, S. Nemati, M. Abdar, E. Cambria, U. Acharrya, ABCDM: an attention-based bidirectional CNN-RNN deep model for sentiment analysis, Future Gener. Comput. Syst. 115 (2021) 279–294, https://api.semanticscholar.org/CorpusID:224858759.

[19] J. Pennington, R. Socher, C. Manning, GloVe: global vectors for word representation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 1532–1543, https://aclanthology.org/D14-1162.

[20] S. Deerwester, S. Dumais, G. Furnas, T. Landauer, R. Harshman, Indexing by latent semantic analysis, J. Am. Soc. Inf. Sci. 41 (1990) 391–407, https://EconPapers.repec.org/RePEc:bla:jamest:v:41:y:1990:i:6:p:391-407.

[21] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: Proceedings of the 26th International Conference on Neural Information Processing Systems, vol. 2, 2013, pp. 3111–3119.

[22] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, M. Stede, Lexicon-based methods for sentiment analysis, Comput. Linguist. 37 (2011) 267–307, https://doi.org/10.1162/COLI.

[23] B. Seref, E. Bostanci, Sentiment analysis using naive Bayes and complement naive Bayes classifier algorithms on Hadoop framework, in: 2018 2nd International Symposium On Multidisciplinary Studies And Innovative Technologies (ISMSIT), 2018, pp. 1–7.

[24] M. Rathi, A. Malik, D. Varshney, R. Sharma, S. Mendiratta, Sentiment analysis of tweets using machine learning approach, in: 2018 Eleventh International Conference on Contemporary Computing (IC3), 2018, pp. 1–3, https://api.semanticscholar.org/CorpusID:53288405.

[25] C. Hung, Y. Cao, Sentiment classification of Chinese cosmetic reviews based on integration of collocations and concepts, Electron. Libr. 38 (2019) 155–169, https://api.semanticscholar.org/CorpusID:212972315.

[26] H. Zhao, Z. Liu, X. Yao, Q. Yang, A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach, Inf. Process. Manag. 58 (2021) 102656, https://www.sciencedirect.com/science/article/pii/S0306457321001448.

[27] D. Jain, P. Boyapati, J. Venkatesh, M. Prakash, An intelligent cognitive-inspired computing with big data analytics framework for sentiment analysis and classification, Inf. Process. Manag. 59 (2022) 102758, https://www.sciencedirect.com/science/article/pii/S0306457321002399.

[28] M. Ling, Q. Chen, Q. Sun, Y. Jia, Hybrid neural network for sina Weibo sentiment analysis, IEEE Trans. Comput. Soc. Syst. 7 (2020) 983–990.

[29] D. Stojanovski, G. Strezoski, G. Madjarov, I. Dimitrovski, I. Chorbev, Deep neural network architecture for sentiment analysis and emotion identification of Twitter messages, Multimed. Tools Appl. 77 (2018) 32213–32242, https://doi.org/10.1007/s11042-018-6168-1.

[30] G. Yang, H. He, Q. Chen, Emotion-semantic-enhanced neural network, IEEE/ACM Trans. Audio Speech Lang. Process. 27 (2019) 531–543, https://api.semanticscholar.org/CorpusID:57379095.

[31] Y. Song, Q. Hu, L. He, P-CNN: enhancing text matching with positional convolutional neural network, Knowl.-Based Syst. 169 (2019) 67–79, https://www.sciencedirect.com/science/article/pii/S0950705119300371.

[32] J. Liao, S. Wang, D. Li, Identification of fact-implied implicit sentiment based on multi-level semantic fused representation, Knowl.-Based Syst. 165 (2019) 197–207, https://www.sciencedirect.com/science/article/pii/S0950705118305756.

[33] A. Abdi, S. Shamsuddin, S. Hasan, J. Piran, Deep learning-based sentiment classification of evaluative text based on multi-feature fusion, Inf. Process. Manag. 56 (2019) 1245–1259.

[34] H. Peng, Y. Ma, Y. Li, E. Cambria, Learning multi-grained aspect target sequence for Chinese sentiment analysis, Knowl.-Based Syst. 148 (2018) 167–176, https://doi.org/10.1016/j.knosys.2018.02.034.

[35] F. Huang, X. Li, C. Yuan, S. Zhang, J. Zhang, S. Qiao, Attention-emotion-enhanced convolutional LSTM for sentiment analysis, IEEE Trans. Neural Netw. Learn. Syst. 33 (2022) 4332–4345.

[36] Z. Lei, Y. Yang, M. Yang, Y. Liu, A multi-sentiment-resource enhanced attention network for sentiment classification, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2018, pp. 758–763, https://aclanthology.org/P18-2120.

[37] M. Zulqarnain, R. Ghazali, M. Ghouse, M. Mushtaq, Efficient processing of GRU based on word embedding for text classification, Int. J. Inf. Vis. 3 (2019) 377–383.

[38] P. Atandoh, Z. Fengli, D. Adu-Gyamfi, H. Leka, P. Atandoh, A glove CNN-Bilstm sentiment classification, in: 2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2021, pp. 245–249.

[39] Q. Ma, L. Yu, S. Tian, E. Chen, W. Ng, Global-local mutual attention model for text classification, IEEE/ACM Trans. Audio Speech Lang. Process. 27 (2019) 2127–2139.

[40] G. Liu, J. Guo, Bidirectional LSTM with attention mechanism and convolutional layer for text classification, Neurocomputing 337 (2019) 325–338, https://www.sciencedirect.com/science/article/pii/S0925231219301067.

[41] A. Yenter, A. Verma, Deep CNN-LSTM with combined kernels from multiple branches for IMDb review sentiment analysis, in: 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), 2017, pp. 540–546.

[42] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? Sentiment classification using machine learning techniques, in: Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing, EMNLP 2002, Philadelphia, PA, USA, July 6-7, 2002, 2002, pp. 79–86, https://aclanthology.org/W02-1011/, 2002.

[43] B. Pang, L. Lee, A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts, in: Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, 21-26 July, 2004, Barcelona, Spain, 2004, pp. 271–278, https://aclanthology.org/P04-1035/.

[44] B. Pang, L. Lee, Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales, 2005, ArXiv Preprint Cs/0506075.

[45] R. Socher, J. Pennington, E. Huang, A. Ng, C. Manning, Semi-supervised recursive autoencoders for predicting sentiment distributions, in: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, 2011, pp. 151–161.

[46] S. Wang, C. Manning, Baselines and bigrams: simple, good sentiment and topic classification, in: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2012, pp. 90–94, https://aclanthology.org/P12-2018.

[47] Y. Liu, J. Bi, Z. Fan, A method for multi-class sentiment classification based on an improved one-vs-one (OVO) strategy and the support vector machine (SVM) algorithm, Inf. Sci. 394–395 (2017) 38–52, https://www.sciencedirect.com/science/article/pii/S002002551631043X.

[48] Y. Kim, Convolutional neural networks for sentence classification, ArXiv Preprint, arXiv:1408.5882, 2014.

[49] P. Atandoh, F. Zhang, D. Adu-Gyamfi, P. Atandoh, R. Nuhoho, Integrated deep learning paradigm for document-based sentiment analysis, J. King Saud Univ, Comput. Inf. Sci. 35 (2023) 101578, https://www.sciencedirect.com/science/article/pii/S1319157823001325.

[50] P. Liu, X. Qiu, X. Huang, Adversarial multi-task learning for text classification, ArXiv Preprint, arXiv:1704.05742, 2017.

[51] L. Gui, J. Leng, J. Zhou, R. Xu, Y. He, Multi task mutual learning for joint sentiment classification and topic detection, IEEE Trans. Knowl. Data Eng. 34 (2022) 1915–1927.

[52] S. Bates, T. Hastie, R. Tibshirani, Cross-validation: what does it estimate and how well does it do it?, J. Am. Stat. Assoc. (2023) 1–12, https://doi.org/10.1080/01621459.2023.2197686.

[53] G. Liu, J. Guo, Bidirectional LSTM with attention mechanism and convolutional layer for text classification, Neurocomputing 337 (2019) 325–338.

[54] B. Wu, Z. Wang, L. Wang, Interpretable corn future price forecasting with multivariate time series, J. Forecast. (2024), https://api.semanticscholar.org/CorpusID:268169971.

[55] Y. Xie, R. Raga, Convolutional Neural Networks for Sentiment Analysis on Weibo Data: a Natural Language Processing Approach, 2023.

[56] M. Geva, A. Gupta, J. Berant, Injecting numerical reasoning skills into language models, ArXiv Preprint, arXiv:2004.04487, 2020.