Pertanika J. Sci. & Technol. 32 (1): 79 - 98 (2024)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Attention-based Spatialized Word Embedding Bi-LSTM Model for Sentiment Analysis

Kun Zhu and Nur Hana Samsudin*

School of Computer Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia

ABSTRACT

Movie reviews provide a medium of communication for the movie fans community. Movie reviews not only help viewers and potential viewers to obtain a general opinion about a movie but also allow the fans to construct an opinion of the movie. In this work, an analysis of over 60,000 movie reviews has been implemented to find meaningful text representation via text embedding. We improved the text embedding by proposing an attention-based Bidirectional Long-Short Term Memory (Bi-LSTM) network by using over 60,000 movie review text data as the training set and over 20,000 movie review text data as the testing set. Based on the data features, we performed a probabilistic analysis of the information related to words and phrases, combined the analysis results with text embedding, spatialized the text embedding, and compared the performance of the proposed attention-based spatialized word embedding Bi-LSTM model with several traditional machine learning models. The attention-based spatialized word embedding Bi-LSTM model proposed in this paper achieves an F1 score of 0.91 on the movie review sentiment classification dataset, with a prediction accuracy of 91%, outperforming the results of the current state-of-the-art research. The model can effectively identify the sentimental tendencies of movie reviews and use the analyzed sentimental tendencies to guide consumers in their consumption and obtain feedback on movie content.

Keywords: Attention-based deep neural network, data mining, deep learning, natural language processing, sentiment analysis

ARTICLE INFO

Article history:
Received: 11 December 2022
Accepted: 14 June 2023
Published: 06 November 2023

DOI: https://doi.org/10.47836/pjst.32.1.05

E-mail addresses:
zhu.kun@student.usm.my (Kun Zhu)
nurhana.samsudin@usm.my (Nur Hana Samsudin)
* Corresponding author

INTRODUCTION

Nowadays, people are no longer limited to asking friends and family for recommendations when they want to watch a movie or buy an item, as there are many user reviews and discussions about the product online. Movie reviews are an important

ISSN: 0128-7680
e-ISSN: 2231-8526

piece of data on the social services platform, directly reflecting the audience's evaluation of the movie's plot, scenes, director, actors, shooting techniques, script, setting, colors, and sound effects (Zhen et al., 2018). Sentiment analysis of movie reviews in different periods has different reference and application values. Through analyzing the movie reviews, audiences who want to watch the movie can quickly obtain public opinion about a movie. From the review, film investors may also adjust their marketing strategies in time to maximize their investment return. Therefore, film review analysis has become one of a popular tool of sentiment analysis research in recent years (Wang et al., 2020).

Sentiment analysis is an important research direction in natural language processing (NLP)(Abdullah & Rusli, 2021; Fernandes & Mannepalli, 2021a). Manual analysis is unrealistic because of the massive amount of movie review data. It is necessary to use technical tools such as machine learning to perform sentiment analysis of movie reviews (Nayak et al., 2018). As one of the current popular recurrent neural networks (RNN), LSTM can mitigate the gradient disappearance and gradient explosion problems of RNN through the gating mechanism (Cheng et al., 2020; Munshi et al., 2022), so it is often used for the task of text analysis. Sentiment analysis models of LSTM tend to better capture long-term memory dependencies in textual data sequences. Meanwhile, adding an attention mechanism to LSTM networks can reduce the dimensionality of the data, reducing the computational effort of the data and giving the model the ability to capture more important inter-word information (Fernandes & Mannepalli, 2021b; Lai et al., 2015).

This paper proposes an attention-based approach to spatialize word embedding in a bidirectional LSTM model based on the above overview. Spatialization sees words as points in space and word-to-word connections as a vector. Commonly used word combinations are close, and less common word combinations are farther apart. The same vector distance is the same part of the sentence with the same combination of words. It is used to increase the efficiency and accuracy of the model. The bidirectional attentional structure can effectively capture the connections between sentences in context. The attention mechanism is used to capture the relationships between sentences. It allows sentiment analysis tasks to be performed without losing information. Based on Gu et al. (2021), the relevant data are calculated to build a matrix and combined with word embeddings to form spatially structured word embeddings. It allows for a richer input characterization and, to a certain extent, improves information extraction and avoids information loss. Then, add spatialized word embeddings in both directions of the Bi-LSTM and use the Bi-LSTM to encode and decode the text information so that sentence features and information at different locations can be concatenated. In the final part of the model, important movie review text features are re-extracted and weighted using the attention mechanism, thus reducing the difficulty of processing the data and improving the accuracy of the output results. The final result is to improve the performance of the text sentiment classification tasks.

This paper presents an experimental study of attention-based spatialized word embedding using a bidirectional LSTM model for sentiment analysis to understand the applicability and advantages of the method over baseline machine learning methods by evaluating accuracy, precision, recall, and F1 score.

RELATED WORK

Opinions mined from user-generated content are invaluable assets (Liu, 2012). Since Nasukawa and Yi (2003) introduced the concept of sentiment analysis at the beginning of the 21st century, it has become of increasing interest (Liu, 2012). In this context, social text analytics are considered the original application of big data analytics today, as they can derive underlying opinions and sentiments from the vast amount of user-generated data shared online. Traditional sentiment analysis methods are divided into lexicon-based and machine-learning-based approaches. Machine learning methods for sentiment analysis are considered superior to lexicon-based methods because they are unsupervised techniques that rely on external resources to associate polarity scores with each feature used for opinion mining (Behera et al., 2021). Several studies on text classification as well as clustering proposed various methods that aim to enhance the performance of text classification and clustering using ensemble learning methods, clustering, and optimization algorithms (Onan, 2018a; Onan, 2018b; Onan, 2019a; Onan, Korukoğlu et al., 2017). Onan, Bulut et al. (2017), and Onan and Korukoğlu (2017) also evaluated the proposed methods on benchmark datasets and compared their performance with traditional methods. The experimental results showed that the proposed methods outperformed traditional methods in most cases. However, some proposed methods may require more computing resources or manual tuning. Overall, the studies suggest that using machine learning methods such as integrated learning, clustering, and optimization algorithms can improve the performance of text classification and clustering.

Deep learning-based approaches to emotional text analysis are considered superior to traditional sentiment analysis methods, as deep learning can extract features with very high accuracy (Li et al., 2019; Lim et al., 2021). In previous research on sentiment analysis, the most commonly used deep learning models are RNN and LSTM (Hochreiter & Schmidhuber, 1997). Word embeddings have been widely used for text-based representation and understanding since the late 1990s, while LSTM only started to flourish in 2013 when data science was developing (Jain et al., 2021). Onan (2019b) discusses the challenge of detecting irony in sentiment analysis and proposes a deep learning approach that combines word embedding techniques and feature sets. The LDA2vec algorithm outperforms other word embedding techniques, and combining word embedding-based features with traditional feature sets yields good results. In another paper, Onan (2019c) presents a method for topic extraction in scientific literature using word embedding models. The method improves the performance of clustering algorithms, but the study may be limited to certain subject areas. In 2020, Onan proposed a text mining approach for sentiment analysis on instructor evaluation reviews, which achieved a high classification accuracy using deep learning methods with the GloVe word embedding scheme. However, the limitations of this approach are not discussed (Onan, 2020).

Despite their inconsistent emergence, LSTM and word embeddings are now well used in text-based analysis, such as sentiment analysis and opinion mining (AlKhwiter & Al-Twairesh, 2021). In 2012, Socher et al. began an approach to using semantic data to improve the performance of sentiment analysis. However, traditional approaches do not take into account the semantic associations between sentences or document content. In 2015, Chen proposed the Text-CNN model, which achieved good results in sentiment analysis, and was one of the first researchers to apply CNN to sentiment analysis. Later, researchers found that deep learning-related models worked well for emotional text sequence problems (Jianqiang et al., 2018) and began to apply deep learning on a gradually larger scale in NLP (Kumar et al., 2021). In 2019, Rani and Kumar proposed a deep learning approach based on CNN.

This CNN-based deep learning model can effectively perform sentiment analysis on Hindi movie reviews. The CNN-based deep learning approach uses a hand-labeled dataset by three native Hindi experts, and the model achieves classification accuracy of up to 95%. In the same year, Shrivastava et al. (2019) proposed a method combining CNN and Attention to perform sentiment analysis on textual datasets, yielding good results. The research shows that CNN-based deep learning sentiment analysis models have the potential to perform better than machine learning solutions. On novel symbolic and picture emojis, Solanki et al. (2019) propose an opinion-mining method based on sentiment markers to identify positive, negative, and neutral opinions on sentiment topics.

Although this is a traditional method, it can be useful for the sentiment analysis of new emoticons. In 2021, Sarzynska et al. proposed a new ELMo model based on the LSTM structure, and the results were better than the original LSTM model. Onan (2021a) proposed methods for sentiment analysis on MOOC comments using machine learning, ensemble learning, and deep learning techniques. Their research showed that ensemble learning methods outperformed supervised learning methods and that deep learning architectures, particularly the LSTM network, achieved the highest predictive performance. Onan (2021b) also proposed a deep learning architecture for sentiment analysis that combined TF-IDF weighted glove word embedding with a CNN-LSTM architecture. Onan and Tocoglu (2021) developed a deep learning framework for identifying sarcasm in social media using term weighting and LSTM network structure, which outperformed traditional deep neural networks. These studies demonstrate the effectiveness of deep learning techniques and the importance of selecting appropriate text representation schemes and classification algorithms for sentiment analysis in different contexts.

However, further research is needed to explore the potential of other methods and address the limitations of current approaches, such as model ambiguity and fuzzy interference. Tan et al. (2022) built on this method by applying sliding windows to low-resource languages with very good results. Onan (2022) proposed a deep-learning framework for detecting sarcasm in text. The framework includes a three-layer bidirectional LSTM network with a weighted trigram model and a weighted word embedding model. The proposed approach outperformed traditional deep neural network architectures in prediction performance and word embedding. However, efficiency issues when processing large-scale data may be a limitation. Some of the work related to sentiment text analysis is summarized in Table 1.

Our work compares and improves the above research results and proposes an attention-based approach to spatializing word embedding in a Bi-LSTM model. The spatialized input provides more effective information to the model, such as the distance of the same word vectors in a sentence, but increases the model's running time and computational space. Based on the model's accuracy, we use the attention mechanism to assign weights to the individual features within the model, thus reducing the computational burden of processing high-dimensional input data, reducing the dimensionality of the data, and producing high-quality results. The addition of the attention mechanism increases the complexity of the model but reduces the computational effort and increases the accuracy of the model, so we believe that the complexity of the model is worth the sacrifice. This paper focuses on improving the accuracy of deep learning for sentiment analysis of movie reviews using Bi-LSTM models with spatialized inputs and attention mechanisms.

METHODOLOGY

Data Description and Pre-processing

The primary valid data used in this model was obtained from IMDB. The IMDB dataset was originally used by Pang et al. (2002), and then IMDB was widely used and extended as a benchmark sentiment dataset. This dataset was extracted from a large dataset of movie reviews used by the Computer Science Department of Stanford University. The IMDB dataset is a collection of 50,000 sentiment texts labeled with positive and negative polarity reviews. All these English movie review texts are balanced and contain only two sentiment categories (i.e., positive and negative sentiment). The dataset consists of a training set, a test set, and a set of unlabeled data. The total training set has 41758 data, of which 20,652 are positive and 21,106 are negative. The test set is 25,000, with 12,500 positive and 12,500 negative data. Based on the IMDB data, Easy Data Augmentation (EDA) was performed on the data related to the IMDB training set using a cloud server in this paper (Wei & Zou, 2019). We back-translated the sentiment text data from the training set while the test set was left unchanged, thus ensuring the fairness of the evaluation. The use of EDA techniques can effectively increase the amount of data for training and improve the

Table 1
The list of some related tasks for sentiment text analysis

Authors	Method	Task	Precision	Recall	F1	Data set
Nasukawa and Yi (2003)	Sentiment keyword extraction	Sentiment Analysis	0.943	0.286	-	Benchmark Corpus
Shen and Huang (2016)	Attention-Based CNN	Semantic Relation Extraction	-	-	0.859	SemEval-2010 Task 8
Mondal et al. (2018)	Probabilistic model	Rumor identification	0.623	0.708	0.6672	Twitter
Dai et al. (2018)	LSTM	Relation Classification	-	-	0.857	Wikidata knowledge graph
Chen et al. (2018)	RNNs Autoencoders	Rumor identification	0.9249	0.8799	0.8916	Sina Weibo
Lee et al. (2019)	Bi-LSTM	Semantic Relation Classification	-	-	0.852	SemEval-2010 Task 8
Rani and Kumar (2019)	CNN	Sentiment Analysis	0.935	0.934	0.934	Hindi Movie reviews
Liu et al. (2019)	LSTM	Rumor identification	0.9475	0.9485	0.9475	Sina Weibo
Li et al. (2020)	SAMF-Bi-LSTM	Sentiment Classification	-	-	0.638	YELP3
Demotte et al. (2020)	Sentence-State LSTM	Sentiment Analysis	0.8705	0.888	0.8791	Sinhala News articles
Ranathunga and Liyanage (2021)	LSTM fastText	Sentiment Analysis	0.67	0.66	0.67	Sinhala News articles
Muhammad et al. (2021)	LSTM	Sentiment Analysis	0.825	-	-	Indonesian Hotel Reviews
Kumar et al. (2021)	Bi-LSTM-CRF	Sentiment Analysis	0.781	0.821	0.8	Reviews benchmark datasets
Rasool et al. (2021)	LSTM-CNN	Sentiment Analysis	0.87	0.88	0.881	Twitter
Asghar et al. (2021)	Bi-LSTM-CNN	Rumor identification	0.86	0.86	0.86	Twitter
Islam et al. (2021)	LSTM	Fake News Detection	0.96	0.96	0.96	COVID dataset

Note. A dash (–) represents that the item is not presented in that instance generalization ability of the model, increase the noise data to improve the robustness of the model, and, to some extent, address the problem of insufficient data. The following multiple operations on the review texts are required to facilitate the construction of a lexicon based on a corpus of movie reviews:

(1) Noise data removal: Remove useless symbols such as “,” “@””,” “#”” //” “</br>” from the text, which are useless for building sentiment dictionaries.
(2) Unification of text: Unified uppercase and lowercase, unified text length. Uniform capitalization will help feature extraction during word embedding. The unified case is because the length of text characters needs to be the same during the LSTM model training process. In this case, for texts less than the maximum characters, this paper uses ‘PAD’ characters for padding to ensure that the maximum length of the text is consistent.
(3) Tokenization: The text is broken down into tokens, symbols, words, or unique elements with clear meaning to reduce the burden on the model.

Word Embedding and Spatialization

This step first performs tokenization on the paragraphs that need to be analyzed. Usually, there are two ways of word segmentation: First, sentences can be dissected into individual words and expressions, allowing for a more granular analysis of language. Second, an alternative approach involves treating the entire sentence as a single entity, which can be particularly useful in certain contexts. In this step, we combine the above two methods, use regular expressions to segment words and retain certain special emotional symbols. Moreover, we use the N-gram algorithm to optimize word segmentation. The N-Gram formula is as stated in Equation 1:

$$\mathrm{p(Q)=p(w_1w_2w_3w_4\cdots w_n)=p(w_1^n)} \tag{1}$$

p(Q) is the probability of sentence Q appearing in the corpus, n is the length of sentence Q, and $\mathrm{w_i}(1 \le i \le n)$ is a word transformed by word segmentation in sentence Q. According to the chain rule and conditional probability formula, the probability of p(Q) is equal to the probability multiplication of each word $\mathrm{w_i(1 \le i \le n)}$ in Q, and the formula is as stated in Equation 2:

$$p(Q) = p(w_1^n)$$

$$= p(w_1)p(w_2|w_1)p(w_3|w_2w_1)\cdots p(w_n|w_{n-1}\cdots w_2w_1) \tag{2}$$

When training an N-gram model, we also need to know the parameter estimation conditional probability of the model. Suppose $\mathrm{f(w_{i-1})}$ is the number of times the word

$\mathrm{w_{i-1}}$ appears in the corpus. In that case, $f(\mathrm{w_{i-1}}, \mathrm{w_i})$ is the number of times the two-tuple $(\mathrm{w_{i-1}}, \mathrm{w_i})$ appears in the training corpus, and $\mathrm{p}(\mathrm{w_i}|\mathrm{w_{i-1}})$ does the model require the estimated conditional probability, then the formula is as stated in Equation 3:

$$p(w_i|\mathrm{w}_{i-1}) = \frac{p(\mathrm{w}_{i-1}, w_i)}{p(w_i)} = \frac{f(\mathrm{w}_{i-1}, w_i)}{f(w_i)} \tag{3}$$

This research uses word embedding because if one-hot encoding is adopted, the dimensionality will be too high, and the amount of calculation will increase. Unlike one-hot encoding, word embedding uses a floating-point dense matrix to represent word segmentation. Based on the size of the dictionary, vectors are used to represent the words' dimensions. A vector represents words and sentences. Before that, we will use numbers to represent word segmentation and vectors to represent numbers: word segmentation → number → vector. The vectors consisting of words, sentences, and the information between them (e.g., word-to-word, sentence-to-sentence, word-to-sentence similarity, and distance) then spatialize the word embeddings to improve the model's accuracy. This paper uses a dictionary to save each word and the corresponding number in the text and implements a method to map the sentence into a list containing numbers through the dictionary. After completing this serialization, we also need to record the number of times and filter the words. For the problem of different sentence lengths, each batch of sentences cannot be constructed with the same length. For those not in the dictionary during training, This article also uses special characters to replace the words.

Model Structure

In this paper, we first extract the text for sentence feature information, then perform word embedding and combine the results with the extracted sentence information features as the input to the neural network. Features contain information about spatialized features, i.e., the distance between the same words in a sentence. LSTM is mainly designed to extract high-level features from the text, and Bi-LSTM can get a deeper temporal relationship, enhancing the accuracy of high-level feature extraction. After completing the forward LSTM and reverse LSTM, the results are concatenated, and the results are then handed over to the Attention mechanism for processing. This layer of Attention processing gets the global information after the Bi-LSTM processing and performs weighting based on the relevant feature information. The Attention mechanism here will also get the fault information related to the model, which will also greatly improve the accuracy of the prediction. After completing these parts, it will be handed over to the fully connected layer for softmax processing and classification, with the final output being the sentiment classification of the movie text. The model structure is shown in Figure 1.

The process described above is a process of encoding. After the encoding has been completed, attention and decoder work simultaneously. The model needs to calculate the correlation, with the input Bi-LSTM result l_i and the un-updated state Y_0. The first step is to transform linearly using the two-parameter matrices W_k and W_q to get the two vectors k_i and q_0. The third step is to do a softmax transform on these values. The correlation is expressed as a, a is also the weight for the attention mechanism, and each a corresponds to a state l_i. We need a weighted average of the states in a and get the content vector C. Each C has a corresponding decoding state, Y. The content vector C_i formula is as stated in Equation 4:

$$C_i = a_1 l_1 + a_2 l_2 + \cdots + a_i l_i \tag{4}$$

The new state Y_i is a concat of the old state S_{i-1}, the new input function X with the content vector C_i, and then multiplied by the weight matrix plus the bias b. The formula is as stated in Equation 5:

$$Y_i = tanh\left(S_i \cdot \begin{bmatrix} X \\ Y_{i-1} \\ C_i \end{bmatrix}\right) \tag{5}$$

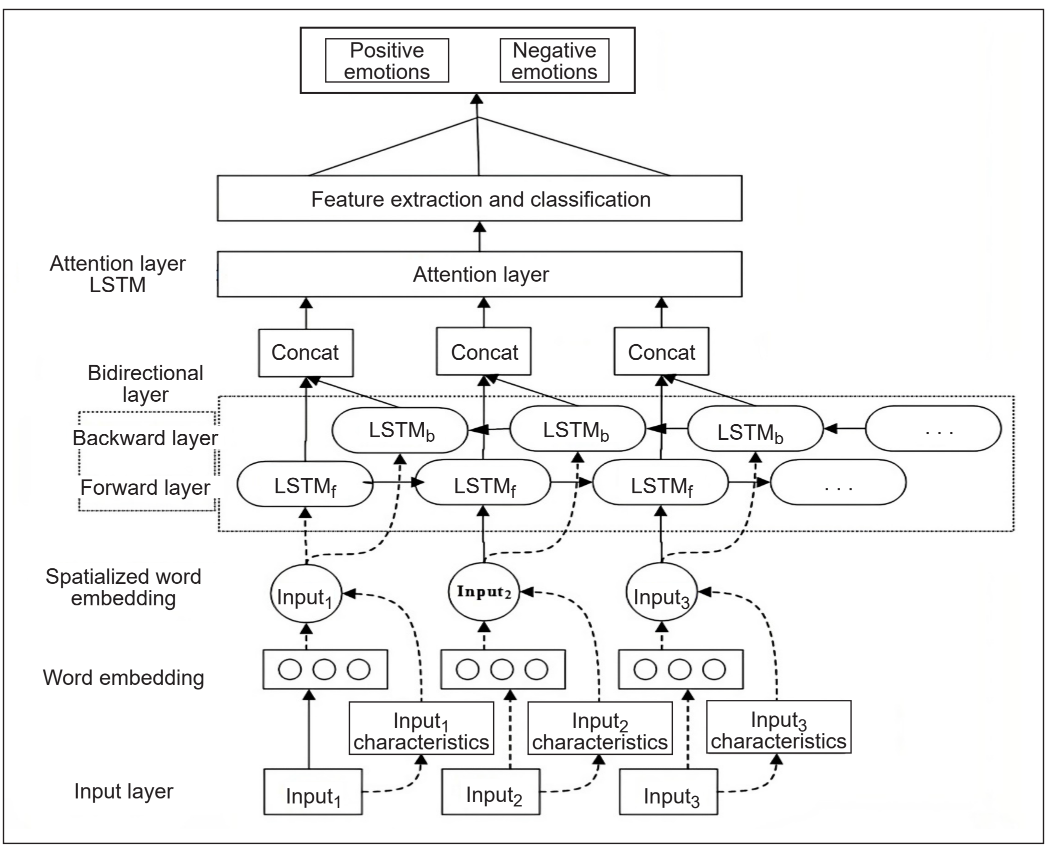


Figure 1. Model structure

Model Training and Analysis

This paper uses torch.nn from the Pytorch library to build a Bi-LSTM model. When building the model, we need to input the necessary parameters. Table 2 shows the parameter settings for the model in this paper and the comparison model.

Table 2
Model parameters

Parameters	**Spatialized word embedding**	**Spatialized word embedding + Bi-LSTM**	**Spatialized word embedding + Bi-LSTM + attention**
Number of layers	2-Dense	1-Bi-LSTM, 1-Dense	1-Bi-LSTM, 1-Dense
Dimension of hidden state vector	-	200, 128	200, 128
Number of neurons (Dense)	256, 2	2	2
Activation	ReLU, Softmax	Softmax	Softmax
Learning rate	0.001	0.001	0.001
Optimizer	Adam	Adam	Adam
Dropout	0.1	0.05	0.05
Sentence Max	200	200	200
Batch size	8	128	128
Epochs	30	30	30

RESULTS AND COMPARISON

This section discusses the results of different models of traditional machine learning and deep learning for sentiment analysis attention-based spatialized word embedding LSTM. The attention-based spatialized word embedding Bi-LSTM model was constructed based on the Pytorch deep learning framework, using a Bi-LSTM +Attention deep learning network structure. In the case of deep learning using the attention-based spatialized word embedding Bi-LSTM model, the total sentiment text dataset collected was 91,758, split into a training and test set in a 3:2 ratio. In this section, we perform comparative experiments on our model and provide a detailed comparative analysis of the model's results after spatialized word embedding.

Results

The performance of the proposed model was evaluated using the evaluation metrics of Accuracy, Precision, Recall, F1-score, and AUC.

As shown in Table 3, Spatialized word embedding methods have significant advantages over regular word embedding methods for sentiment analysis tasks, provided that the dataset remains unchanged. As shown in Table 4, among the models using only word embedding, the models using the spatialized word embedding method had a 10.12% higher prediction rate and an 11.9% higher F1 score than those using the regular word embedding method. The attention-based Bi-LSTM model with spatialized word embedding had a 5.34% higher prediction rate and a 6.27% higher F1 score than the attention-based Bi-LSTM model with regular word embedding. The model of the spatialized word embedding method is able to capture more word and semantic information before the word is embedded, and the model can encode better based on these features.

It is one of the important reasons why the spatialized word embedding approach models work better. The Attention mechanism helps the model to reduce the computational burden of processing high-dimensional input data by structurally selecting a subset of the inputs, reducing the dimensionality of the data, and allowing the model to focus more on finding useful information in the input data that is significantly relevant to the current output, thus improving the quality of the output. As shown in Table 5, the performance of the Bi-LSTM model with the Attention mechanism was higher than that of the Bi-LSTM model without the Attention mechanism. The F1-score of the model can be improved by 7.02% after using the Attention mechanism.

Precision represents the degree of prediction precision in the sample results. Figure 2 compares the precision and recall of different models for positive and negative categories after spatialized word embedding: (a) the precision of the positive category and (b) the precision of the negative category; (c) the recall of the positive category and (d) the recall of the negative category. The combination of spatialized word embeddings, attention mechanisms, and a Bi-LSTM model has significant results for the sentiment text

Table 3
Comparison of spatialized word embedding

Method	Precision	AUC	Recall	F1
Word embedding	64.63%	70.43%	62.31%	62.56%
Word embedding (spatialized)	74.75%	82.96%	74.29%	74.46%

Table 4
Comparison of the attention-based Bi-LSTM model with spatialized word embedding

Method	Precision	AUC	Recall	F1
Word embedding + Bi-LSTM + attention	85.26%	88.74%	84.15%	84.23%
Word embedding (spatialized) + Bi-LSTM + attention	90.60%	95.69%	90.41%	90.50%

Table 5
Model comparison with an attention mechanism

Method	Precision	AUC	Recall	F1
Word embedding (spatialized) + Bi-LSTM	84.14%	90.78%	83.65%	83.48%
Word embedding (spatialized) + Bi-LSTM +attention	90.60%	95.69%	90.41%	90.50%

classification task (Figure 2). The stability of the model and the recognition of positive and negative categories are significantly improved. It is demonstrated that the combination of spatialized word embeddings and Bi-LSTM can better capture sequence information, and the classification precision is higher with the addition of the Attention mechanism.

Figures 3 and 4 show images of the loss and accuracy processes for the three model runs after spatialization. In Figure 3, the X-axis is the epoch of the model's run, and the Y-axis is the loss of the model. In Figure 4, the X-axis is also the epoch of the models run, and the Y-axis is the accuracy of the models. The loss of the three models decreases and gradually stabilizes as the number of training steps increases in Figure 3. In Figure 4, it can be seen from the learning curve that the models are learning from the data. As the epoch of training rounds increases, the accuracy of the models increases and stabilizes as it reaches its peak. The comparison of the different models in Figures 3 and 4 shows that the attention-based spatialized word embedding Bi-LSTM model is more stable has a faster training process, and is more accurate than the other models. The comparison of the different models in Figures 3 and 4 shows that the attention-based spatialized word embedding Bi-

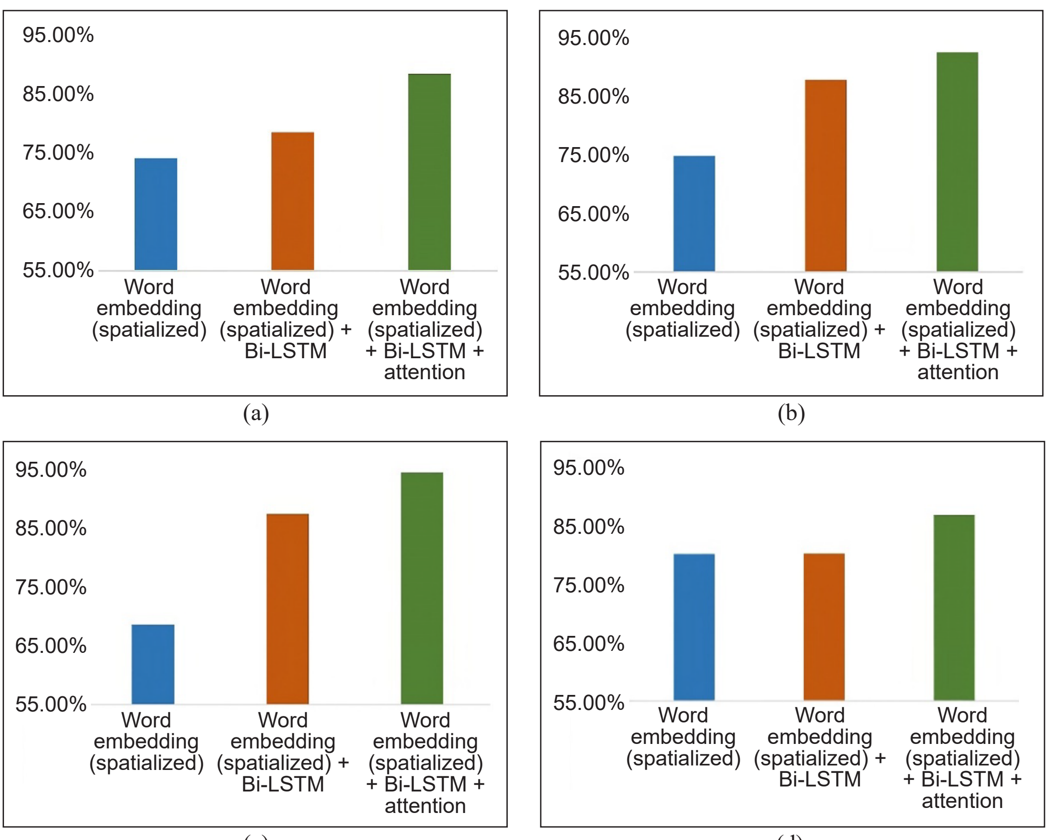
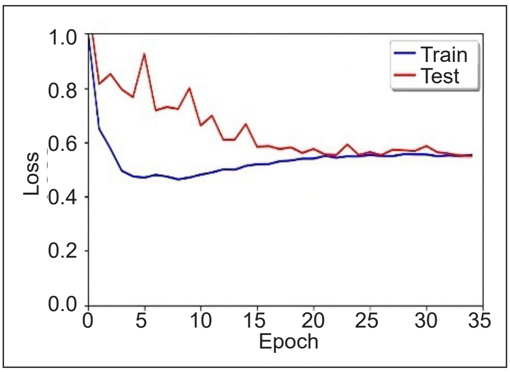
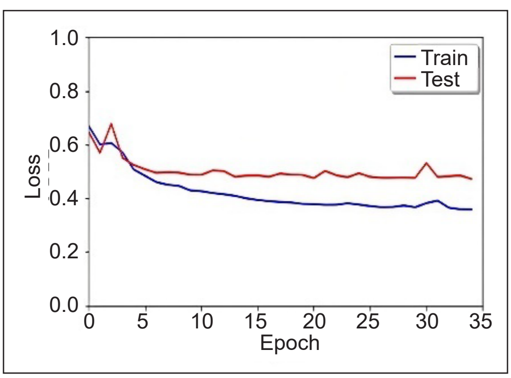


Figure 2. Precision: (a) positive category, (b) negative category; and recall: (c) positive category, (d) negative category

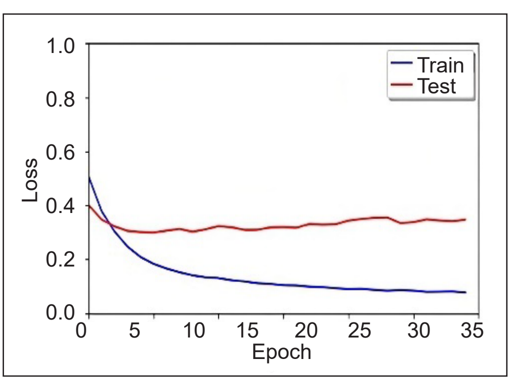
LSTM model is more stable has a faster training process, and is more accurate than the other models. It also proved that there is a gap between the traditional word embedding and deep learning models and that our Bi-LSTM model with the Attention mechanism has better results than the traditional Bi-LSTM model.



(a)

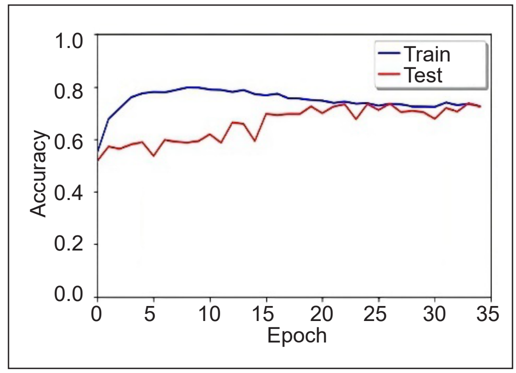


(b)

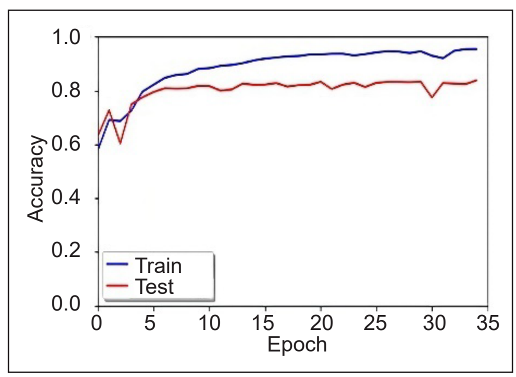


(c)

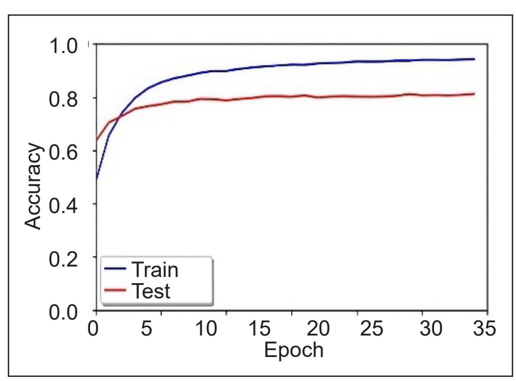
Figure 3. Loss of models: (a) word embedding (spatialized); (b) word embedding (spatialized) + Bi-LSTM; and (c) word embedding (spatialized) + Bi-LSTM + attention



(a)



(b)



(c)

Figure 4. Accuracy of models: (a) word embedding (spatialized); (b) word embedding (spatialized) + Bi-LSTM; and (c) word embedding (spatialized) + Bi-LSTM + attention

Comparison of Related Models

In this subsection, we evaluate the performance of our attention-based spatialized word embedding Bi-LSTM model on the same test set using the work in Table 6. Traditionally, feature extraction methods were the most commonly used in the early period of targeting sentiment analysis tasks. The first three models in Table 6 mainly use traditional feature extraction methods (Yang et al., 2016), all of which have achieved good effectiveness on sentiment analysis tasks, in particular the emotions and word embeddings method proposed by Giatsoglou et al. (2017), which achieved a precision of 87.8%, which is a very good result. The spatialized word embedding method in this paper also achieved a precision of only 74.75% in a stand-alone test, but the combination of traditional feature extraction methods with deep learning in this paper achieved a precision of 90.60%. We might achieve better results if we learn from Giatsoglou et al. (2017) and add the recognition of sentiment word categories to the spatialized word embeddings. It is one of the next steps in our future work.

In the last three years, the methods used for sentiment analysis have mainly been neural networks or deep learning. Chiny et al. (2021) achieved 82.9% precision, 82.7% recall, and 83.5% F1-score using LSTM on the same test set as our paper, but were respectively 7.7%, 7.7%, and 7% lower than the model proposed in our paper. We believe that the main difference between the two models is that the LSTM is unable to obtain more complete features before the input and cannot perform importance analysis on the features after the output, which is one of the areas where we differ from most models. Based on previous work, Leng et al. (2021) and Gupta et al. (2021) added an Attention mechanism to the LSTM, which improved the precision of the model by about 5% compared to the ordinary LSTM model. A unidirectional LSTM has the limitation that it only retains information from the past and not from the future or the next state to understand the context of the sentence better.

In our paper, the Bi-LSTM model captures information from both current and previous inputs and avoids information loss, thus enabling effective prediction by retaining contextual information (current and previous) over a long period. Briskilal and Subalalitha (2022) used the BERT model to solve the sentiment analysis task with very good results, and it improved the precision value of the movie sentiment classification task to over 90%. To a certain extent, the Attention mechanism in this paper is similar to the Attention mechanism of BERT. Anisotropy in the vectors in which BERT encodes sentences can lead to the semantics of even a high-frequency word and a low-frequency word being equivalent. This paper uses a method of spatialized word embeddings that reduces this by altering the vector somewhat before input. Compared to the BERT model, the precision of the proposed model in this paper is improved by 0.2%.

Through comparative experiments, the model proposed in this paper is the best compared to other classical models on the same test set of IMDB. This paper applies the Bi-LSTM and attention mechanism methods to the automatic codec model. It allows the automatic codec to have bi-directional computational capabilities. In addition, using spatialized word embedding, the text information can be well reconstructed using this traditional method to improve the capture of text features. It is also shown from the comparative tests that the deep learning methods are, to some extent, superior to the traditional methods and that the models with attention mechanisms outperform those without.

Table 6
Comparison of related models

Authors	Method	Precision	Recall	F1
Yang et al. (2016)	Hierarchical attention networks	-	-	49.4%
Giatsoglou et al. (2017)	Emotions and word embeddings	87.8%	-	-
Kumar et al. (2019)	Hybrid feature extraction	78.3%	-	78.0%
Chiny et al. (2021)	LSTM	82.9%	82.7%	83.5%
Leng et al. (2021)	HRNaEMSA	87.5%	87.4%	87.4%
Gupta et al. (2021)	Senti_ALSTM	87.4%	-	-
Briskilal et al. (2022)	BERT	90.4%	-	89.0%
Our proposed model	Word embedding (spatialized) + Bi-LSTM + attention	90.6%	90.4%	90.5%

CONCLUSION

In this paper, we propose an approach based on attention spatialized word embedding Bi-LSTM model to improve the accuracy of movie review sentiment analysis effectively and enrich the movie review training dataset using Easy Data Augmentation Techniques.

Firstly, Bi-LSTM can effectively capture the connections between sentences in the context. Secondly, spatialized word embeddings provide the model with more usable text data features while retaining the basic textual information features. Finally, the attention mechanism weighs the processed data to perform sentiment analysis tasks faster and better without losing information. We experimented with classifiers from a variety of traditional and deep-learning methods. They were tested on the same IMDB test set, and our results were compared with previous work. The experimental results show that classifiers from the attention-based spatialized word embedding Bi-LSTM model approach outperformed those from other approaches, achieving the best results in terms of precision (90.6%), recall (90.4%), and F1-score (90.5%).

In the future, we will extend the model to more scenarios of sentiment text analysis. In addition, we will also combine the features learned from unstructured data, such as sound and video, using multimodal joint representation to do sentiment text information retrieval and apply sentiment text information retrieval to sentiment text classification tasks, which is one of our important future works.

ACKNOWLEDGMENTS

This manuscript publication fee was partially funded by the Universiti Sains Malaysia's Short-term Grant No.: 304/PKOMP/6315273.

REFERENCES

Abdullah, N. A. S., & Rusli, N. I. A. (2021). Multilingual sentiment analysis: A systematic literature review. *Pertanika Journal of Science and Technology*, *29*(1), 445-470. https://doi.org/10.47836/pjst.29.1.25

AlKhwiter, W., & Al-Twairesh, N. (2021). Part-of-speech tagging for Arabic tweets using CRF and Bi-LSTM. *Computer Speech and Language*, *65*, Article 101138. https://doi.org/10.1016/j.csl.2020.101138

Asghar, M. Z., Habib, A., Habib, A., Khan, A., Ali, R., & Khattak, A. (2021). Exploring deep neural networks for rumor detection. *Journal of Ambient Intelligence and Humanized Computing*, *12*(4), 4315-4333. https://doi.org/10.1007/s12652-019-01527-4

Behera, R. K., Jena, M., Rath, S. K., & Misra, S. (2021). Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data. *Information Processing and Management*, *58*(1), Article 102435. https://doi.org/10.1016/j.ipm.2020.102435

Briskilal, J., & Subalalitha, C. N. (2022). An ensemble model for classifying idioms and literal texts using BERT and RoBERTa. *Information Processing and Management*, *59*(1), Article 102756. https://doi.org/10.1016/j.ipm.2021.102756

Chen, W., Zhang, Y., Yeo, C. K., Lau, C. T., & Lee, B. S. (2018). Unsupervised rumor detection based on users' behaviors using neural networks. *Pattern Recognition Letters*, *105*, 226-233. https://doi.org/10.1016/j.patrec.2017.10.014

Chen, Y. (2015). *Convolutional Neural Network for Sentence Classification* (Unpublished Master's thesis). University of Waterloo, Canada. https://uwspace.uwaterloo.ca/handle/10012/9592

Cheng, Y., Yao, L., Xiang, G., Zhang, G., Tang, T., & Zhong, L. (2020). Text sentiment orientation analysis based on multi-channel CNN and bidirectional GRU with attention mechanism. *IEEE Access, 8*, 134964-134975. https://doi.org/10.1109/ACCESS.2020.3005823

Chiny, M., Chihab, M., Chihab, Y., & Bencharef, O. (2021). LSTM, VADER and TF-IDF based hybrid sentiment analysis model. *International Journal of Advanced Computer Science and Applications*, *12*(7), 265-275. https://doi.org/10.14569/IJACSA.2021.0120730

Dai, Y., Guo, W., Chen, X., & Zhang, Z. (2018). Relation classification via LSTMs based on sequence and tree structure. *IEEE Access, 6*, 64927-64937. https://doi.org/10.1109/ACCESS.2018.2877934

Demotte, P., Senevirathne, L., Karunanayake, B., Munasinghe, U., & Ranathunga, S. (2020). Sentiment analysis of Sinhala news comments using sentence-state LSTM networks. In *MERCon 2020 - 6th International Multidisciplinary Moratuwa Engineering Research Conference* (pp. 283-288). IEEE Publishing. https://doi.org/10.1109/MERCon50084.2020.9185327

Fernandes, B., & Mannepalli, K. (2021a). An analysis of emotional speech recognition for tamil language using deep learning gate recurrent unit. *Pertanika Journal of Science and Technology, 29*(3), 1937-1961. https://doi.org/10.47836/pjst.29.3.37

Fernandes, B., & Mannepalli, K. (2021b). Speech emotion recognition using deep learning LSTM for tamil language. *Pertanika Journal of Science and Technology, 29*(3), 1915-1936. https://doi.org/10.47836/pjst.29.3.33

Giatsoglou, M., Vozalis, M. G., Diamantaras, K., Vakali, A., Sarigiannidis, G., & Chatzisavvas, K. C. (2017). Sentiment analysis leveraging emotions and word embeddings. *Expert Systems with Applications, 69*, 214-224. https://doi.org/10.1016/j.eswa.2016.10.043

Gu, W., Tandon, A., Ahn, Y. Y., & Radicchi, F. (2021). Principled approach to the selection of the embedding dimension of networks. *Nature Communications, 12*(1), 1-10. https://doi.org/10.1038/s41467-021-23795-5

Gupta, C., Chawla, G., Rawlley, K., Bisht, K., & Sharma, M. (2021). Senti_ALSTM: Sentiment analysis of movie reviews using attention-based-LSTM. In *Proceedings of 3rd International Conference on Computing Informatics and Networks: ICCIN 2020* (pp. 211-219). Springer. https://doi.org/10.1007/978-981-15-9712-1_18

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735-1780. https://doi.org/10.1007/978-1-4757-5388-2_2

Islam, M. U., Hossain, M. M., & Kashem, M. A. (2021). COVFake: A word embedding coupled with LSTM approach for COVID related fake news detection. *International Journal of Computer Applications, 174*(10), 1-5. https://doi.org/10.5120/ijca2021920977

Jain, P. K., Saravanan, V., & Pamula, R. (2021). A hybrid CNN-LSTM: A deep learning approach for consumer sentiment analysis using qualitative user-generated contents. *ACM Transactions on Asian and Low-Resource Language Information Processing, 20*(5), Article 84. https://doi.org/10.1145/3457206

Jianqiang, Z., Xiaolin, G., & Xuejun, Z. (2018). Deep convolution neural networks for twitter sentiment analysis. *IEEE Access, 6*, 23253-23260. https://doi.org/10.1109/ACCESS.2017.2776930

Kumar, A., Verma, S., & Sharan, A. (2021). ATE-SPD: Simultaneous extraction of aspect-term and aspect sentiment polarity using Bi-LSTM-CRF neural network. *Journal of Experimental and Theoretical Artificial Intelligence, 33*(3), 487-508. https://doi.org/10.1080/0952813X.2020.1764632

Kumar, K., Harish, B. S., & Darshan, H. K. (2019). Sentiment analysis on IMDb movie reviews using hybrid feature extraction method. *International Journal of Interactive Multimedia and Artificial Intelligence, 5*(5), Article 109. https://doi.org/10.9781/ijimai.2018.12.005

Lai, S., Xu, L., Liu, K., & Zhao, J. (2015). Recurrent convolutional neural networks for text classification. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 29, No. 1). AAAI Press. https://doi.org/10.1609/aaai.v29i1.9513

Lee, J., Seo, S., & Choi, Y. S. (2019). Semantic relation classification via bidirectional LSTM networks with entity-aware attention using latent entity typing. *Symmetry, 11*(6), Article 785. https://doi.org/10.3390/sym11060785

Leng, X. L., Miao, X. A., & Liu, T. (2021). Using recurrent neural network structure with enhanced multi-head self-attention for sentiment analysis. *Multimedia Tools and Applications, 80*(8), 12581-12600. https://doi.org/10.1007/s11042-020-10336-3

Li, W., Liu, P., Zhang, Q., & Liu, W. (2019). An improved approach for text sentiment classification based on a deep neural network via a sentiment attention mechanism. *Future Internet, 11*(4), Article 96. https://doi.org/10.3390/FI11040096

Li, W., Qi, F., Tang, M., & Yu, Z. (2020). Bidirectional LSTM with self-attention mechanism and multi-channel features for sentiment classification. *Neurocomputing, 387*, 63-77. https://doi.org/10.1016/j.neucom.2020.01.006

Lim, C. T., Bong, C. H., Wong, W. S., & Lee, N. K. (2021). A comprehensive review of automated essay scoring (AES) research and development. *Pertanika Journal of Science and Technology, 29*(3), 1875-1899. https://doi.org/10.47836/pjst.29.3.27

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies, 5*(1), 1-167. https://doi.org/10.1142/9789813100459_0007

Liu, Y., Jin, X., & Shen, H. (2019). Towards early identification of online rumors based on long short-term memory networks. *Information Processing and Management, 56*(4), 1457-1467. https://doi.org/10.1016/j.ipm.2018.11.003

Mondal, T., Pramanik, P., Bhattacharya, I., Boral, N., & Ghosh, S. (2018). Analysis and early detection of rumors in a post disaster scenario. *Information Systems Frontiers, 20*(5), 961-979. https://doi.org/10.1007/s10796-018-9837-8

Muhammad, P. F., Kusumaningrum, R., & Wibowo, A. (2021). Sentiment analysis using Word2vec and long short-term memory (LSTM) for Indonesian hotel reviews. *Procedia Computer Science, 179*, 728-735. https://doi.org/10.1016/j.procs.2021.01.061

Munshi, A. A., AlSabban, W. H., Farag, A. T., Rakha, O. E., Al Sallab, A., & Alotaibi, M. (2022). Automated Islamic jurisprudential legal opinions generation using artificial intelligence. *Pertanika Journal of Science and Technology, 30*(2), 1135-1156. https://doi.org/10.47836/pjst.30.2.16

Nasukawa, T., & Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd International Conference on Knowledge Capture* (pp. 70-77). John Wiley & Sons. https://doi.org/10.1111/j.1469-185X.1956.tb01550.x

Nayak, S. K., Rout, P. K., Jagadev, A. K., & Swarnkar, T. (2018). Elitism-based multi-objective differential evolution with extreme learning machine for feature selection: A novel searching technique. *Connection Science, 30*(4), 362-387. https://doi.org/10.1080/09540091.2018.1487384

Onan, A. (2018a). An ensemble scheme based on language function analysis and feature engineering for text genre classification. *Journal of Information Science, 44*(1), 28-47. https://doi.org/10.1177/0165551516677911

Onan, A. (2018b). Biomedical text categorization based on ensemble pruning and optimized topic modelling. *Computational and Mathematical Methods in Medicine*, *2018*, Article 2497471. https://doi.org/10.1155/2018/2497471

Onan, A. (2019a). Consensus clustering-based undersampling approach to imbalanced learning. *Scientific Programming*, *2019*, Article 5901087. https://doi.org/10.1155/2019/5901087

Onan, A. (2019b). Topic-enriched word embeddings for sarcasm identification. In *Software Engineering Methods in Intelligent Algorithms: Proceedings of 8th Computer Science Online Conference* (pp. 293-304). Springer. https://doi.org/10.1007/978-3-030-19807-7_29

Onan, A. (2019c). Two-stage topic extraction model for bibliometric data analysis based on word embeddings and clustering. *IEEE Access, 7*, 145614-145633. https://doi.org/10.1109/ACCESS.2019.2945911

Onan, A. (2020). Mining opinions from instructor evaluation reviews: A deep learning approach. *Computer Applications in Engineering Education, 28*(1), 117-138. https://doi.org/10.1002/cae.22179

Onan, A. (2021a). Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach. *Computer Applications in Engineering Education, 29*(3), 572-589. https://doi.org/10.1002/cae.22253

Onan, A. (2021b). Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks. *Concurrency and Computation: Practice and Experience, 33*(23), Article e5909. https://doi.org/10.1002/cpe.5909

Onan, A. (2022). Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism for text sentiment classification. *Journal of King Saud University - Computer and Information Sciences, 34*(5), 2098-2117. https://doi.org/10.1016/j.jksuci.2022.02.025

Onan, A., Bulut, H., & Korukoğlu, S. (2017). An improved ant algorithm with LDA-based representation for text document clustering. *Journal of Information Science, 43*(2), 275-292. https://doi.org/10.1177/0165551516638784

Onan, A., & Korukoğlu, S. (2017). A feature selection model based on genetic rank aggregation for text sentiment classification. *Journal of Information Science, 43*(1), 25-38. https://doi.org/10.1177/0165551515613226

Onan, A., Korukoğlu, S., & Bulut, H. (2017). A hybrid ensemble pruning approach based on consensus clustering and multi-objective evolutionary algorithm for sentiment classification. *Information Processing and Management, 53*(4), 814-833. https://doi.org/10.1016/j.ipm.2017.02.008

Onan, A., & Tocoglu, M. A. (2021). A term weighted neural language model and stacked bidirectional LSTM based framework for sarcasm identification. *IEEE Access, 9*, 7701-7722. https://doi.org/10.1109/ACCESS.2021.3049734

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In *EMNLP '02: Proceedings of the ACL-02 conference on Empirical methods in natural language processing* (pp. 79-86). ACM Publishing. https://doi.org/10.3115/1118693.1118704

Ranathunga, S., & Liyanage, I. U. (2021). Sentiment analysis of Sinhala news comments. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *20*(4), Article 59. https://doi.org/10.1145/3445035

Rani, S., & Kumar, P. (2019). Deep learning based sentiment analysis using convolution neural network. *Arabian Journal for Science and Engineering, 44*(4), 3305-3314. https://doi.org/10.1007/s13369-018-3500-z

Rasool, A., Jiang, Q., Qu, Q., & Ji, C. (2021). WRS: A novel word-embedding method for real-time sentiment with integrated LSTM-CNN model. In *2021 IEEE International Conference on Real-Time Computing and Robotics (RCAR)* (pp. 590-595). IEEE Publishing. https://doi.org/10.1109/RCAR52367.2021.9517671

Sarzynska-Wawer, J., Wawer, A., Pawlak, A., Szymanowska, J., Stefaniak, I., Jarkiewicz, M., & Okruszek, L. (2021). Detecting formal thought disorder by deep contextualized word representations. *Psychiatry Research, 304*, Article 114135. https://doi.org/10.1016/j.psychres.2021.114135

Shen, Y., & Huang, X. J. (2016). Attention-based convolutional neural network for semantic relation extraction. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers* (pp. 2526-2536). The COLING 2016 Organizing Committee.

Shrivastava, K., Kumar, S., & Jain, D. K. (2019). An effective approach for emotion detection in multimedia text data using sequence based convolutional neural network. *Multimedia Tools and Applications, 78*(20), 29607-29639. https://doi.org/10.1007/s11042-019-07813-9

Socher, R., Huval, B., Manning, C. D., & Ng, A. Y. (2012). Semantic compositionality through recursive matrix-vector spaces. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (pp. 1201-1211). Association for Computational Linguistics.

Solanki, V. K., Cuong, N. H. H., & Lu, Z. J. (2019). Opinion mining: using machine learning techniques. In *Extracting Knowledge from Opinion Mining* (pp. 66-82). IGI Global. https://doi.org/10.4018/978-1-5225-6117-0.ch004

Tan, T. P., Lim, C. K., & Rahman, W. R. E. A. (2022). Sliding window and parallel LSTM with attention and CNN for sentence alignment on low-resource languages. *Pertanika Journal of Science and Technology, 30*(1), 97-121. https://doi.org/10.47836/pjst.30.1.06

Wang, Q., Zhu, G., Zhang, S., Li, K., Chen, X., & Xu, H. (2020). Extending emotional lexicon for improving the classification accuracy of Chinese film reviews. *Connection Science, 33*(2), 153-172. https://doi.org/10.1080/09540091.2020.1782839

Wei, J., & Zou, K. (2019). *EDA: Easy data augmentation techniques for boosting performance on text classification tasks*. ArXiv Preprint. https://doi.org/10.18653/v1/d19-1670

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016). Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies* (pp. 1480-1489). Association for Computational Linguistics.

Zhen, F., Yi, G., Zhenhao, Z., & Meiqi, H. (2018). Sentiment analysis of movie reviews based on dictionary and weak tagging information. *Journal of Computer Applications, 38*(11), 3084-3088.