Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/38/20230524

Movie sentiment analysis based on Long Short-Term Memory Network

Siyao Li[1], Rui Qin[2,4] and Zijian Zhou[3]

[1]College of Computer Science and Technology, Harbin Engineering University, Harbin, 150000, China
[2]Department of Automation, Guangdong University of Technology, Guangzhou, 510006, China
[3]Dundee International Institute, Central South University, Changsha, Hunan, 410083, China

[4]3220000984@mail2.gdut.edu.cn

Abstract. An important task in the study of Natural Language Processing (NLP) is the analysis of movie reviews. It finishes the task of classifying movie review texts into sentiment, such as positive, negative or neutral sentiment. Previous works mainly follow the pipeline of LSTM (Long Short-Term Memory Network). The network model is a variant of Recurrent Neural Network (RNN) and particularly suitable for processing natural language texts. Though existing LSTM-based works have improved the performance significantly, we argue that most of them deal with the problem of analyzing the sentiment of movie reviews while ignore analyze the model performance in different application scenarios, such as different lengths of the reviews and the frequency of sentiment adverbs in the reviews. To alleviate the above issue, in this paper, we constructed a simple LSTM model containing an embedding layer, a batch normalization layer, a dropout layer, a one-dimensional convolutional layer, a maximal pooling layer, a bi-directional LSTM layer and a fully connected layer. We used the existing IMDB movie review dataset to train the model, and selected two research scenarios of movie review length and frequency of occurrence of sentiment adverbs to test the model, respectively. From the experimental results, we proposed a model for the scenarios in which the LSTM model handles the problem of sentiment analysis with respect to the dataset construction, model stability and generalization ability, text fragment processing, data preprocessing and feature extraction, model optimization and improvement.

Keywords: Sentiment Analysis, LSTM, Movie Review, Different Application Scenarios.

1. Introduction

In today's society, the information explosion brings a huge amount of data. How to extract valuable information from the data has become one of the hotspots of research, which has attracted increasing attention from both industry and academia. Among them, movie review analysis, as an important task in the realm of natural language processing, seeks to gather sentiment data from textual reviews of films from viewers and it can be used in a variety of contexts, including marketing, product improvement, and public opinion monitoring[1-3].

A kind of recurrent neural network is called LSTM (Long Short-Term Memory Network), has become a common tool in movie review analysis research due to its excellent performance in processing sequential data, especially in natural language text processing. The LSTM and the adaptive boosting technique (Adaboost) were integrated by Zhang et al., who then trained the model using the IMDB movie review dataset. By comparing CNN and LSTM results on performance metrics like accuracy and precision, conclusions can finally be reached [4]. After deciding on the most straightforward LSTM-CNN architecture to extract the features for sentiment categorization, Mahesh et al. employed a multi-layer perceptron network to identify positive and negative attitudes [5]. The feelings of the IMDb movie reviews were analyzed by Qaisar et al. using the LSTM classifier. They preprocessed the data effectively and enhanced the post classification performance. Their study supports the feasibility of incorporating the developed approach in contemporary text-based sentiment analyzers [6]. However, in practice, the problem of unstable model performance in different application scenarios is often faced. In particular, when dealing with movie reviews, factors such as the length of the review and the frequency of emotional adverbs will have an impact on the model's performance, but these factors are often ignored.

To alleviate this issue, in this paper, we provide an in-depth investigation of LSTM models in the problem of movie review sentiment analysis. Specifically, a simple LSTM model including Embedding Layer, Batch Normalization Layer, Dropout Layer, 1-D Convolutional Layer, Maximum Pooling Layer, Bidirectional LSTM Layer and Fully Connected Layer is constructed to be able to better adapt to the characteristics of variant movie reviews. Meanwhile, this paper chooses the specific research scenario of movie review length and uses the existing IMDB movie review dataset to train the model adequately. On the trained model, this paper conducts a series of rigorous tests, evaluates the performance of the model under different movie review lengths for the delineated test sets, and analyzes the experimental results in detail. These experimental results not only provide this paper with information about the model's effectiveness in various settings, but also provide a strong basis for further optimization and improvement of the model. Based on the experimental results, this paper proposes outlooks and concepts for the scenarios in which the LSTM model handles sentiment analysis problems in terms of dataset construction, model stability and generalization ability, text fragment processing, data preprocessing and feature extraction, model optimization and improvement, and other aspects. Ultimately, this paper argues that in practical applications, a more reasonable network structure should be designed according to the actual length of the movie reviews, the weight of sentiment adverbs and the context in which they appear should be reasonably considered, and data preprocessing as well as feature extraction should be strengthened to further improve the stability and generalization ability of the model.

2. Method

2.1. Revisiting LSTM

The Forget Gate, Input Gate, and Output Gate are the three basic gating units that are used in LSTM to govern information input, output, and forgetting. The Output Gate controls the output of the current moment, the Input Gate controls the input of recent information, and the Forget Gate determines whether the memory of the previous instant is kept. The gating mechanism allows the LSTM to maintain and utilize long-term contextual information in long sequences, thus effectively capturing dependencies in the sequence.

(1) Forget Gate. The forgetting gate decides which information should be erased from the preceding moment C_t memory , it is calculated using the following formula:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t]) + b_f \quad (1)$$

In the function, W_f and b_f are the weight matrix and bias vector of the forgetting gate, respectively. h_{t-1} is the hidden state of the previous moment, x_t is the input at the current moment, σ denotes Sigmoid function.

(2) Input Gate. The new information is decided by the input gate at the current moment g_t which determines what parts of the memory state C_t will be added to and it is calculated using the following formula:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t]) + b_i \tag{2}$$

$$\widetilde{C_t} = \sigma(W_C \cdot [h_{t-1}, x_t]) + b_C \tag{3}$$

where W_i and b_i are the input gate's weight matrix and bias vector. W_C and b_C are the weight matrix and bias vector to generate candidate memory states, $\widetilde{C_t}$ denotes the hyperbolic tangent function.

(3) Updating Memory Status. Utilizing forgetting gates, input gates, and the resulting candidate memory states $\widetilde{C_t}$, memory status C_t can be updated, it is calculated using the following formula:

$$C_t = f_t * C_{t-1} + i_i * C_t \tag{4}$$

(4) Output Gate. The concealed state is determined by the output gate. at the current moment h_t, it is calculated using the following formula:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t]) + b_o \tag{5}$$

$$h_t = o_t * tanh(C_t) \tag{6}$$

$\mathrm{W_o}$ and $\mathrm{b_o}$are the weight matrix and bias vector of the output gate, $\mathrm{h_t}$ is the hidden state at the current moment, tanh denotes the hyperbolic tangent function.

As shown in Figure 1, the basic structure of an LSTM network includes components such as an embedding layer, multiple LSTM layers, and a fully connected layer, which are briefly described as follows:

(1) Embedding Layer: first, the input discretized data (e.g., words in text) is transformed into continuous vectors, a process done by the embedding layer. The embedding layer helps to extract features of the input data by learning to map different symbols into a continuous vector space with semantic associations.

(2) LSTM Layer: the LSTM layer, which stacks numerous LSTM units to increase the model's learning capacity, is the fundamental part of the LSTM model. There are forgetting gates, input gates, output gates, and memory units in every LSTM unit. These gating mechanisms control the information flow based on the current input, the output from the previous moment, and the memory state from the previous moment.

(3) Fully Connected Layer: a Fully Connected Layer is often put after the LSTM layer., which is used to further process the features output from the LSTM layer for final tasks such as classification, regression or generation. The fully connected layer can be designed according to the requirements of specific tasks, such as adding a Dropout layer to prevent overfitting.

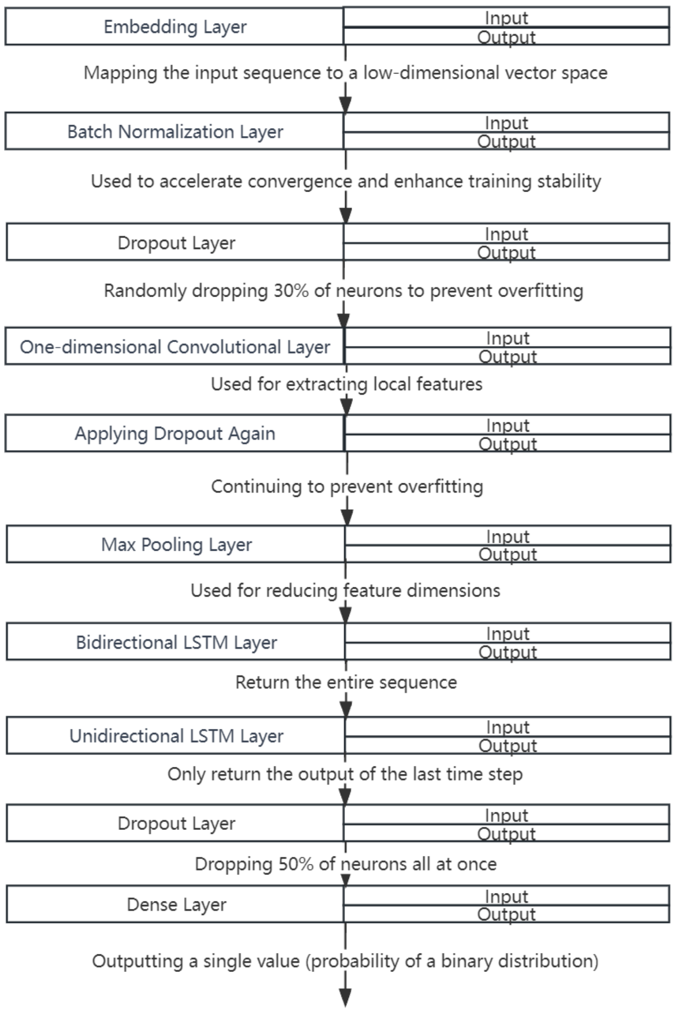

Figure 1. The Network structure of LSTM.

2.2. *Original dataset and pre-processing*

Utilizing the LSTM model enables sentiment analysis on movie reviews, but it necessitates a substantial dataset for training. Therefore, having a high-quality dataset is crucial. Currently, there are several large-scale datasets available, including the IMDb dataset, Amazon movie review dataset, Rotten Tomatoes dataset, Yelp dataset, MovieLens dataset, and Cornell movie review dataset. Among these, the IMDb dataset contains a substantial amount of data related to movies. In this study, due to the necessity of training the LSTM model for various scenarios, we require a more detailed preprocessing of these datasets.

In the process of dataset acquisition, our team first looked for existing large-scale datasets of movie reviews. Subsequently, our team performed various preprocessing steps on the dataset for different scenarios. One involved categorizing reviews based on their length, and the other involved categorizing the frequency of adverbs of degree appearing in the reviews. These diverse datasets for various scenarios will aid us in assessing the performance of the LSTM model across different application contexts.

In the first application scenario, we categorized movie reviews of different lengths into five classes within the dataset. These classes were: reviews containing 100 words or fewer, reviews containing 100 to 150 words, reviews containing 150 to 200 words, reviews containing 200 to 250 words, and reviews

DOI: 10.54254/2755-2721/38/20230524

containing over 250 words. In this dataset, the first column represents the specific content of the movie review, while the second column indicates the sentiment of the review (whether the text expresses a positive or negative sentiment). We tested the accuracy of an LSTM model for sentiment analysis on movie reviews of varying sentence lengths.

In the second application scenario, we categorized the dataset into four groups based on the number of adverbs of degree present in the movie reviews: 1 to 5 adverbs, 5 to 10 adverbs, 10 to 15 adverbs, and over 15 adverbs. In this dataset, the first column represents the specific content of the movie review, while the second column indicates the sentiment of the review (whether the text expresses a positive or negative sentiment). We evaluated the accuracy of the LSTM model in performing sentiment analysis on movie reviews with varying numbers of adverbs of degree.

3. Experiment

The following methods are adopted to validate the impact of the LSTM model on sentiment classification for datasets with varying lengths and different degrees of sentiment. First, the count of words and the frequency of occurrences of sentiment adverbs in each sample of the original dataset are counted to derive the distribution of the number of words and the degree of sentiment of each comment in the original dataset. Then, the original dataset is further divided into five sub-datasets containing different comment word counts respectively according to the derived word count distribution statistics. Based on the number of occurrences of sentiment adverbs the original dataset is divided into four sub-datasets containing comments with different levels of sentiment. Ultimately, the sentiment of the two sub-datasets' categories is categorized using the trained LSTM model, and the impact of the LSTM model on sentiment classification of the datasets with different lengths and datasets with different degrees of sentiment is discussed.

3.1. Experiment setting

The evaluation index of the experiment is classification accuracy. First, the features of the input training set are fed into the LSTM for training and learning, and then it is the sub-test set generated for sentence length and sentiment level that is fed into the LSTM for classification testing. The parameters of the LSTM network are configured according to the details provided in Table 1.

Table 1. Hyperparameters of the LSTM network.

HYPERPARAMETERS	VALUE
EPOCH	50
BATCH_SIZE	32
SEQUENCE_MAX_LEN	293
EMBEDDING_DIM	293

3.2. Convergence analysis

The original dataset is utilized to train the LSTM model, and the training outcomes are depicted in Figure 2 below, which shows the variation of accuracy versus loss when the model is trained. It can be observed that in the final training results, the accuracy of the validation set reaches about 0.87, and the loss function curve and accuracy function curve of the model have converged, indicating that the model has been trained to be more mature, with a certain degree of generalization ability, and can be used for the next prediction and comparison of the test set divided under different application scenarios.

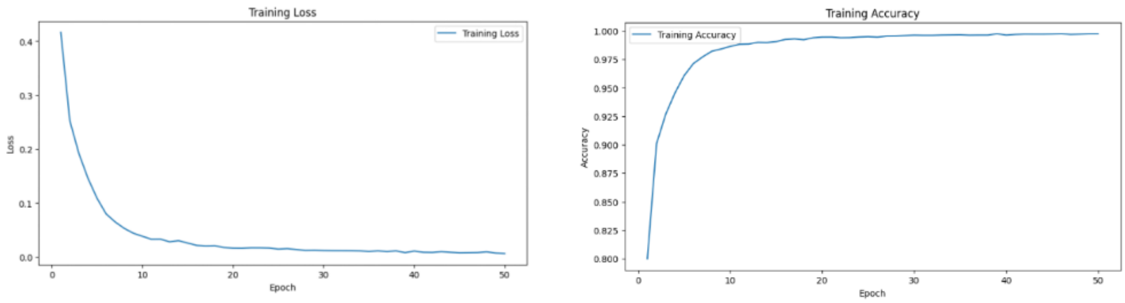

Figure 2. Training accuracy and loss of model.

3.3. Results analysis

3.3.1. Sentiment classification results based on sentence length. We obtained the probability density distribution curves for the number of comment entries regarding the number of words contained in different comments by counting the number of words contained in each comment in the test set, as shown in Figure 3.

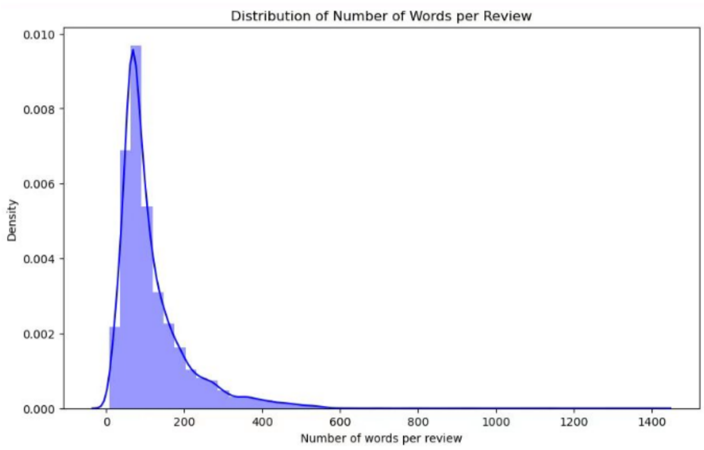

Figure 3. Distribution of Number of Words per Review.

Observe the study of this probability density distribution graph, and make a reasonable estimation of it, to ensure that the test set can be divided into two cases, namely, the extreme number of comment words and the concentrated number of comment words, and can also show its certain reasonableness for other cases. In the end, we divide the test sets into five categories: less than 100 comments, less than 150 comments but more than 100 comments, less than 200 comments but more than 150 comments, less than 250 comments but more than 250 comments, and more than 250 comments. Name them as set_1~set_5 respectively. The accuracies obtained using the trained LSTM model for the sentiment categorization task for set_1~set_5 are shown in Table 2 and Figure 4 as follows.

Table 2. Accuracy and sample size of the set_1~set_5.

SUBTEST SET	SAMPLE SIZE	ACCURACY
SET_1	5764	87.27%
SET_2	1922	86.63%
SET_3	904	87.06%
SET_4	489	85.07%
SET_5	838	86.87%

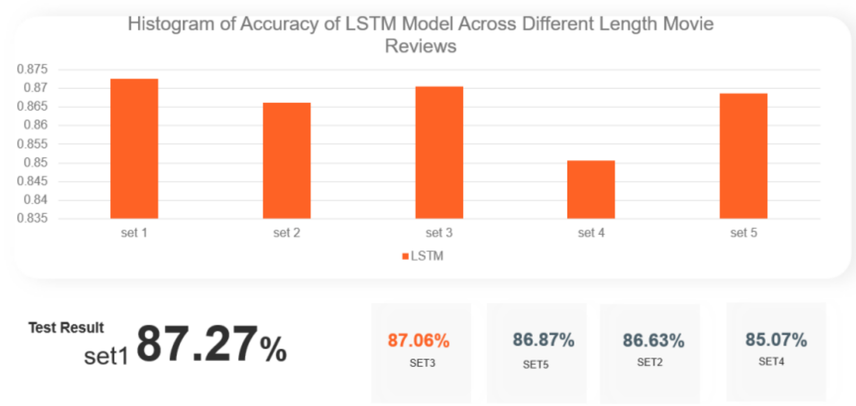

Figure 4. Accuracy of LSTM accross different review lengths.

This means that the LSTM model is less effective in classifying the test set with longer sentence length. This may be due to the fact that the dataset of 200-250 words in the English movie review text dataset is relatively small. This will lead to insufficient training samples in this category during model training, which makes it difficult for the model to learn and generalize sufficiently, and results in a reduction in the precision of the LSTM model for this text type when contrasted with other categories. Although short movie reviews contain fewer words, they may also contain less invalid and interfering information compared to longer reviews, so the model's accuracy on these reviews is also better. However, if the length of a movie review exceeds a certain threshold such as the 250 words counted in the figure above, the model accuracy rises back to a certain high level, which most likely indicates that the amount of invalid and valid information contained in the movie review has reached some kind of balance.

3.3.2. Sentiment classification results based on experience level. In order to study the LSTM model's ability to analyze the sentiment of movie reviews with different degrees of sentiment, this paper counts the number of times the sentiment adverbs (such as Absolutely, Feelingly, etc.) appear in each sample of movie reviews. The number of occurrences of emotion adverbs in each sample is used as the basis for judging its emotion degree, and the sub-test set for the emotion degree is divided. Eventually, the original test set was divided into four sub-test sets, namely, four sub-test sets with the number of occurrences of affective adverbs as 1-5, 6-10, 10-15, and 15 or more. These four test sets are named set_6~set_9 respectively. The accuracy obtained using the trained LSTM model for the sentiment categorization task for set_6~set_9 are shown is shown in the Table 3 and Figure 5.

Table 3. Accuracy and the sample size of the set_6~set_9.

Subtest set	Sample size	Accuracy
SET_6	664	85.84%
SET_7	1368	87.35%
SET_8	551	88.93%
SET_9	289	85.47%

DOI: 10.54254/2755-2721/38/20230524

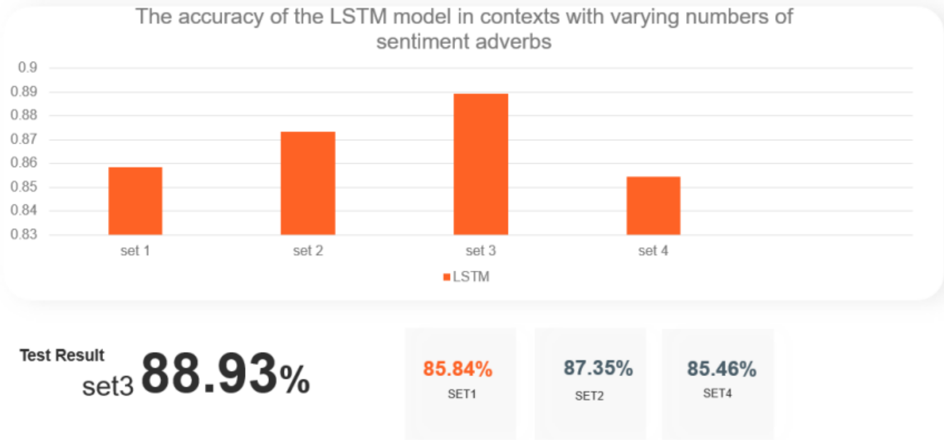

Figure 5. Accuracy of LSTM is measured by different experience levels.

The analysis reveals that the LSTM model achieves its peak accuracy in cases where the count of sentiment words falls within the range of 10 to 15. Thus, it is possible to know the classification ability of the LSTM model for text datasets with different experience levels. In response to the results of the LSTM model's poor classification of the set9 test set, the reason may be that too many sentiment adverbs lead to overemphasis of the sentiment information in the text, which may mask other more subtle sentiment signals. This makes the LSTM model pay more attention to these adverbs when analyzing sentiment and ignore other more recognizable sentiment features. The poorer results of LSTM on the set9 test set can also be explained by the fact that too many sentiment adverbs may introduce noise, especially if there is an imbalance in the sentiment distribution in the training data. The model may over-adapt to the frequent occurrence of sentiment adverbs in the training data, leading to a decrease in predictive power on unseen data, as the occurrence of these frequent adverbs may not be a universal sentiment signal.

4. Discussion

Through this experience, the result shows that the accuracy of LSTM model sentiment analysis can be influenced by sentiment adverbs and sentence length. Therefore, these are the two factors influencing LSTM in the sentiment analysis of movie reviews. In order to inspire subsequent research, this paper further discusses the development of subsequent research on this topic in terms of technology and application.

(1) Building upon the foundation that CNN models effectively extract precise characteristics through the application of convolutional layers and max-pooling layers, and that LSTM models capture extended relationships among sequences of words., in the work of Raza et al. (2020) [7], they introduced an enhanced Hybrid CNN-LSTM Model that effectively merges LSTM with CNN. This novel approach outperforms preceding neural network models. Therefore, our team can use Hybrid CNN-LSTM Model to experiment so that it can help us get the newest statics. (2) Technology in experiment, in our experiment, we have obtained the accuracy of the LSTM model for movie reviews with varying numbers of sentiment adverbs and the accuracy for movie reviews of different lengths. In our subsequent work, we can consider these two metrics as influencing factors at the sentence level. We can combine these two metrics to define specific sentence patterns, such as 'sentence length between 100 and 150, with 1 to 5 sentiment adverbs.' Once we have established these combinations, we can further refine our experiment by selecting different sentence patterns based on movie genres, such as 'Suspense Film,' 'Science Fiction Film,' and others, and then reevaluate the accuracy of the LSTM model for sentiment analysis of movie reviews. This approach allows us to delve deeper into our experimentation, yielding more specific and detailed results.

Sentiment analysis stands out as one of the primary applications of natural language processing and text classification, capturing significant interest from researchers and professionals across diverse domains for a multitude of use cases. Sentiment analysis can assist businesses in understanding the social sentiment surrounding their brand, products, or services across social media and online discussions. Sanjan (2023) [8] highlights that analysis of virtual entertainment streams often remains confined to rudimentary sentiment analysis and quantitative measurements.

(1) In cinema website terms, by incorporating sentiment analysis to account for user preferences, it becomes possible to provide users with more accurate movie recommendations, thereby increasing their engagement on the website and extending their time spent (Barick, 2023) [9]. Thus, utilizing the LSTM model for sentiment analysis on user reviews can help identify their preferences for specific movie genres. This information can be leveraged in subsequent recommendations to suggest movies of that genre. (2) In health care terms, neural network model models have been employed to predict significant healthcare facts. Furthermore, exploring the emotions associated with healthcare information can play a pivotal role in generating crucial insights. Several global and domestic institutions, including the United States' Centers for Disease Control and Prevention (CDC), the World Health Organization (WHO), among others, disseminate vital healthcare data. (Singh, 2021)[10]. Indeed, we can also utilize LSTM for sentiment analysis to predict people's evaluations and sentiments towards healthcare-related matters. (3) In commodity analysis terms, Ge and colleagues (2021) [11] introduced an approach to evaluate the sentiment orientation of product reviews through the utilization of the BERT-BiLSTM-ATT model. Therefore, sentiment analysis tasks can also be applied in product analysis. E-commerce websites can leverage user reviews of previously purchased items to infer user preferences, enabling personalized product and brand recommendations.

Our research findings indicate that the LSTM model exhibits preferences within specific contextual environments. Given that our project's research indicates varying accuracy of LSTM across different textual characteristics, we can select datasets that align with LSTM's preferences for the aforementioned application scenarios. For instance, training LSTM using sentences under 100 words in length and containing 10 to 15 sentiment adverbs could enhance its accuracy in movie analysis.

5. Conclusion

Sentiment analysis can aid businesses in analyzing user habits to refine their innovation direction and assist leaders in understanding employee motivation for improved company management strategies. Governments, organizations, and individuals can utilize sentiment analysis to monitor public opinion and societal emotions. This holds significant value in predicting social trends, crisis management, and informing public policy decisions. Through this project, we conducted sentiment analysis by partitioning and examining the LSTM sentence-level factors. Firstly, we found that accuracy is highest for sentences under 100 words in length. Secondly, the accuracy is optimized when sentiment adverbs occur between 10 and 15 times within a sentence. These discoveries contribute to the nuanced development of the sentiment analysis field.

Author contribution

All the authors contributed equally, and their names were listed in alphabetical order.

References

[1] Ullah K, Rashad A, Khan M, Ghadi Y, Aljuaid H, Nawaz Z. A Deep Neural Network-Based Approach for Sentiment Analysis of Movie Reviews. Complexity. 2022;2022. doi:https://doi.org/10.1155/2022/5217491

[2] Chen, C., Xu, B., Yang, J. H., & Liu, M. (2022). Sentiment analysis of animated film reviews using intelligent machine learning. Computational Intelligence and Neuroscience, 2022.

[3] Ghosh, S., & Ahammed, M. T. (2022). Effects of sentiment analysis on feedback loops between different types of movies. Journal of Media, Culture and Communication (JMCC) ISSN: 2799-1245, 2(02), 14-20.

[4] L. Zhang, M. Wang, M. Liu and H. Li, "Sentiment Analysis of Movie Reviews Based on LSTM-Adaboost," 2022 IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China, 2022, pp. 186-190, doi: 10.1109/IMCEC55388.2022.10019969.

[5] M. Mishra and A. Patil, "Sentiment Prediction of IMDb Movie Reviews Using CNN-LSTM Approach," 2023 International Conference on Control, Communication and Computing (ICCC), Thiruvananthapuram, India, 2023, pp. 1-6, doi: 10.1109/ICCC57789.2023.10165155.

[6] S. M. Qaisar, "Sentiment Analysis of IMDb Movie Reviews Using Long Short-Term Memory," 2020 2nd International Conference on Computer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, 2020, pp. 1-4, doi: 10.1109/ICCIS49240.2020.9257657.

[7] Rehman, A.U., Malik, A.K., Raza, B. et al. A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis. Multimed Tools Appl 78, 26597–26613 (2019). https://doi.org/10.1007/s11042-019-07788-7 .

[8] Ranjith, V., Barick, R., Pallavi, C.V., Sandesh, S., Raksha, R. (2023). Sentiment Enhanced Smart Movie Recommendation System. In: Shakya, S., Balas, V.E., Haoxiang, W. (eds) Proceedings of Third International Conference on Sustainable Expert Systems . Lecture Notes in Networks and Systems, vol 587. Springer, Singapore. https://doi.org/10.1007/978-981-19-7874-6_4.

[9] S. Namitha, P. Sanjan, N. C. Reddy, Y. Srikar, H. Shanmugasundaram and B. P. Andraju, "Sentiment Analysis: Current State and Future Research Perspectives," 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2023, pp. 1115-1119, doi: 10.1109/ICICCS56967.2023.10142318.

[10] Rameshwer Singh, Rajeshwar Singh, Applications of sentiment analysis and machine learning techniques in disease outbreak prediction – A review, Materials Today: Proceedings,Volume 81, Part 2, 2023, Pages 1006-1011, ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2021.04.356. (https://www.sciencedirect.com/science/article/pii/S2214785321032764)

[11] H. Ge, S. Zheng and Q. Wang, "Based BERT-BiLSTM-ATT Model of Commodity Commentary on The Emotional Tendency Analysis," 2021 IEEE 4th International Conference on Big Data and Artificial Intelligence (BDAI), Qingdao, China, 2021, pp. 130-133, doi: 10.1109/BDAI52447.2021.9515273.