Enhancing Emotion Recognition in Text Data Based on Bi-LSTM and Attention Approach

Zhuojun Lyu

School of Computer Science, Central South University, Changsha, 410083, China
8209210519@csu.edu.cn

Abstract. Emotion recognition stands as a cornerstone across various domains, propelling the evolution of artificial intelligence. This paper introduces a pioneering approach to emotion recognition, employing a Bi-directional Long Short-Term Memory (Bi-LSTM) neural network fused with an attention mechanism (Att). The primary aim is to enrich the representation of emotional features within text data, particularly for sentiment analysis endeavors. The Bi-LSTM network effectively captures bidirectional dependencies within text sequences, while the Att meticulously focuses on pivotal segments of the input text, thereby enhancing performance and accuracy. Through experimentation on the Sentiment140 dataset, the model's efficacy is demonstrated, showcasing heightened accuracy and adaptability in contrast to conventional methods. The fusion of Bi-LSTM with Att presents a promising pathway for advancing sentiment analysis tasks, offering valuable insights into the intricacies of emotion recognition within text data. The outcomes of this research not only hold significance for social media analysis and intelligent customer service but also pave the way for potential applications in medical domains such as emotional health monitoring and mental illness diagnosis. Thus, fostering the application of artificial intelligence in diverse fields.

Keywords: Emotion Recognition, Bi-LSTM, Attention Mechanism, Sentiment Analysis.

1 Introduction

In recent years, with the rapid development of artificial intelligence and big data and the massive spread of the Internet, more and more researchers have begun to combine artificial intelligence (AI) with various fields, achieving varying degrees of innovative development. At the same time, a large number of researchers have begun to pay attention to the field of emotion recognition. Emotion analysis has always been one of the important branches in the field of artificial intelligence development. With the continuous development of the field of natural language processing, emotion recognition has also attracted much attention because of its unique ability to discover potential information from many social media. In recent years, the emergence of application products such as smart medical care and AI companions has been inseparable from the development of emotion recognition technology. More accurate and effec-

Y. Wang (ed.), *Proceedings of the 2024 2nd International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2024)*, Advances in Computer Science Research 115,
https://doi.org/10.2991/978-94-6463-540-9_84

tive recognition and judgment of emotions will greatly promote the development of the field of artificial intelligence and provide people with more convenience in their daily lives.

From the current research, emotion recognition is still a classification task. Koolagudi et al. [1] clearly proposed the definition of conversational emotion recognition for the first time and studied the context dependencies in conversations based on the Long Short-Term Memory (LSTM) model. In the initial stage, relevant researchers mainly used support vector machines [2] to complete emotion recognition work. With the rapid development of deep learning, the excellent performance of neural networks in emotion recognition has attracted more and more researchers. In 2015, Chen [3] proposed to build a deep Convolutional Neural Networks (CNN) model through two serial convolution layers. Realize sentence meaning capture [4]. Hazarika et al. [5] have successively used the Conversational Memory Network (CMN) to simulate human interaction and perform emotion recognition through the memory of the conversational person. Majumder et al. [6] proposed using three Gated Recurrent Units (GRU) based on Recurrent Neural Network (RNN) to implement emotion detection by tracking the status and context of the conversational person. In recent years, emotion recognition work has made much progress. Han et al. [7] combined the eXtreme Language Network (XLNet) model [8] with two other models (Bidirectional Gated Recurrent Units (BiGRU) and Attention Mechanism (Att)) to further develop the XLNet-BiGRU-Att model. User texts in social networks are highly subjective, so users will perceive different emotions in the same or similar texts [9]. In order to solve this problem, Ghafoor et al. [10] developed a model for emotion recognition from the perspective of context and subjectivity.

To enhance emotional feature representation across diverse datasets and optimize sentiment analysis model performance, this study proposes a bi-directional long short-term memory (Bi-LSTM) neural network integrated with Atts. Leveraging the Sentiment140 dataset, which comprises 1.6 million sentiment-labeled tweets, serves as the foundation for model training and evaluation. By refining and introducing novel techniques to the LSTM model, a refined emotion recognition model is developed, enhancing accuracy by capturing long-range data dependencies. The implementation procedure involves initial dataset loading, and preprocessing for data quality assurance, followed by the construction of bidirectional LSTM neural network models with integrated Atts to capture contextual dependencies and salient features. Subsequently, the optimized models are centrally trained and evaluated on the Sentiment140 dataset. Comparative analysis against traditional models demonstrates significant improvements in training efficiency and overall accuracy. Furthermore, this model exhibits enhanced generalization capabilities across multiple datasets, offering valuable insights and research avenues for future investigations in sentiment analysis and related domains, thereby guiding researchers and facilitating advancements in diverse fields.

2 Methodology

2.1 Dataset Description and Preprocessing

The dataset used in this study is named Sentiment140 dataset with 1.6 million tweets, which is obtained from Kaggle [11]. The Sentiment140 dataset consists of 1.6 million tweets labeled with a sentiment polarity (positive or negative). It has six fields: target, ids, date, flag, user, and text. In the preprocessing stage, in addition to cleaning and normalizing the data, in order to analyze the effective feature columns in the extracted data set, a feature selection part is added, which only needs the target and text for emotion recognition, so only these two fields are retained. The training efficiency of the model is effectively improved.

2.2 Proposed Approach

The primary objective of this study is to devise a robust text data emotion recognition model capable of accurately identifying and classifying textual data. To fulfill this objective, the article introduces a novel neural network model termed Bi-LSTM-Attention-Sentiment (BLASTS). The selection of this specific architecture offers several advantages. Firstly, the Bi-LSTM network efficiently captures bidirectional dependencies within text sequences, enabling a more comprehensive understanding of textual content. Compared to conventional unidirectional LSTMs, Bi-LSTM demonstrates superior memory capabilities, crucial for tasks like sentiment analysis that rely on discerning long-term dependencies in text. Secondly, Att empower the model to focus on salient sections of input text during predictions, thereby enhancing overall performance and accuracy. By assigning varying attention weights to individual words, the model can flexibly process input texts of varying lengths and complexities. The integration of the BLASTS model leverages both Bi-LSTM networks and Atts, facilitating deeper comprehension and analysis of textual data, thereby enhancing the efficacy of sentiment analysis tasks. This amalgamation enables the model to effectively capture semantic information and emotional nuances within text sequences, thereby delivering more precise and reliable results in sentiment analysis tasks. The schematic representation of the overall process is depicted in Fig. 1.

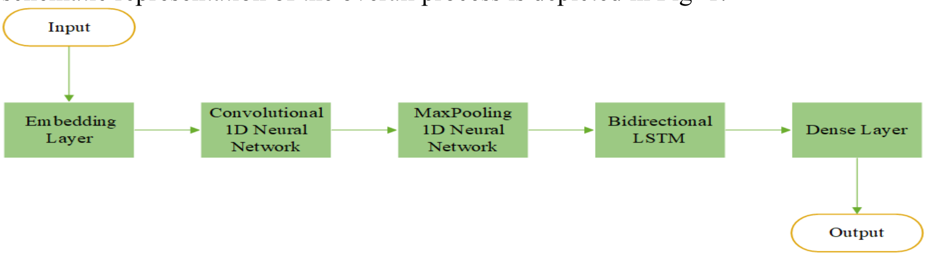

Fig. 1. The flowchart of the model.

As shown in Fig. 1, the input text data is preprocessed to prepare it for analysis. The preprocessed data is then fed into a bidirectional LSTM neural network, where

Att highlights important features. Finally, the model produces an emotion recognition output based on the learned input text representation.

Bi-LSTM. The Bi-LSTM model is a variant of the RNN architecture specifically designed to capture bidirectional dependencies in sequence data. Unlike traditional LSTM networks that process input sequences in one direction, Bi-LSTM networks process forward and backward sequences simultaneously. This enables effectively capture contextual information from past and future inputs, making particularly suitable for tasks involving natural language processing, such as sentiment analysis. The key feature of the Bi-LSTM network is its ability to capture long-term dependencies and contextual information from the input sequence. By processing input data in both forward and backward directions, the Bi-LSTM model can effectively capture the semantic meaning and sequence structure of text data, making it suitable for tasks that require understanding and analyzing the context of words in sentences or documents.

In the context of sentiment analysis, the Bi-LSTM model has several advantages. That can effectively capture the nuances and context of language, allowing to make more accurate predictions of the emotions expressed in textual data. In addition, the Bi-LSTM network can handle variable-length input sequences, making it flexible to adapt to different types of text data. During the research process, the Bi-LSTM model is implemented as a key component of the BLASTS (Bi-LSTM-attention-sentiment) model for sentiment analysis. The structure of the Bi-LSTM layer consists of two LSTM sub-layers, one processing the forward input sequence and the other processing the backward input sequence. Each sub-layer contains multiple LSTM units, which are responsible for processing and capturing information in the input data.

The process of the Bi-LSTM layer in this paper involves three steps. The input text data is first converted into a dense vector representation. Separate LSTM sub-layers are then used to process the input sequence in the forward and backward directions. The outputs of the forward and backward LSTM sub-layers are finally concatenated to obtain a unified representation of the input sequence.

Attention Mechanism (Att). Attention is a technique used in neural network architectures to dynamically focus on specific parts of the input data when making predictions. It gives different weights to different elements of the input sequence based on their relevance to the current task, allowing the model to selectively focus on important information. Att enhances the performance of neural network models by enabling focus on the most relevant parts of the input data. In the context of natural language processing tasks such as sentiment analysis, Att can help models prioritize important words or phrases in input text, leading to more accurate sentiment predictions.

Att dynamically adjusts its attention based on the context of the input sequence, allowing the model to adaptively allocate resources to different parts of the input data. This improves the model's ability to capture the semantics and nuances present in text data, allowing for more efficient emotion recognition. In this paper, Att and the Bi-LSTM layer are simultaneously integrated into the BLASTS model to further improve its performance in sentiment analysis. Att works by assigning an attention weight to each word in the input sequence, which is then used to compute a weighted sum of word embeddings. This weighted sum represents the context vector, which captures the most relevant information in the input sequence for sentiment analysis.

The implementation steps of Att are roughly as follows: first, calculate the attention weight of each word in the input sequence based on its correlation with the sentiment analysis task. The weighted sum of word embeddings is then calculated using attention weights to obtain context vectors. Then integrated with the Bi-LSTM output, the context vector is integrated with the output of the Bi-LSTM layer to generate the final emotion prediction. In summary, Att allows it to dynamically focus on important parts of the input text data while performing emotional prediction, improves the overall training and prediction efficiency of the model, and alleviates the gradient in long sequence processing when processing long sequences. problem, improving the performance of the model.

Loss Function. Loss functions play a crucial role in training deep learning models, including the improved sentiment analysis model discussed earlier. In this study, the author adopts the binary cross-entropy loss function, which is well-suited for binary classification tasks such as sentiment analysis. The binary cross-entropy loss measures the difference between the predicted probability and the true label, and the penalty versus the true situation. deviation. Mathematically, it is expressed as the negative logarithm of the log-likelihood function, providing a principle-based method for optimizing model parameters. The formula is as follows.

$$\mathscr{L}(\hat{y}, y) = -\frac{1}{N}\sum_{i=1}^{N}[y_i \, log(\hat{y}_i) + (1 - y_i) \, log(1 - y_i)] \quad (1)$$

where $\hat{y}$ is the prediction output of the model, indicating the probability that the sample belongs to the positive category, y is the true label of the sample, the value is 0 or 1, and N is the number of samples.

2.3 Implementation Details

This study used Python 3.11 and the Scikit-learn library to build the model and used PyCharm to complete the project under the Windows 10 system. During the training process, 10% of the data is used as a verification set for model verification, the number of convolution kernels is set to 64, the convolution kernel size is set to 5, the number of samples in each batch (batch size) is set to 80, and 16 Rounds of training to get final results. The implementation of the system also involves components such as data preprocessing, data enhancement, and hyperparameter tuning. First, the raw tweet data is preprocessed to remove noise and irrelevant information such as special characters and URLs. Then, data enhancement techniques, such as oversampling and data synthesis, are used to solve the sample imbalance problem and enhance the generalization ability of the model. Finally, hyperparameters (including learning rate, batch size, optimizer settings, etc.) are fine-tuned through cross-validation to optimize the performance of the model. The above implementation details ensure the robustness and effectiveness of the sentiment analysis system in real-world applications.

3 Results and Discussion

In this section, the author conducts an in-depth performance analysis of the sentiment analysis model trained on the Sentiment140 dataset. By analyzing the accuracy/loss, confusion matrix, Area Under the Curve (ROC) curve, and other indicators of training and verification, the accuracy, efficiency, and stability of the model in emotion classification are comprehensively and in-depth explored.

3.1 Training and Verification Performance Analysis

The training and verification performance of the sentiment analysis model is shown in Fig. 2 and Fig. 3 below, where its accuracy and loss curve are evaluated. Fig. 2 shows the accuracy of the model on the training set and validation set as the training period changes. The author observes that during the training process, the accuracy of the model increased from about 50% to 91%, and the loss value dropped from about 0.69 to about 0.22. This shows that the model gradually converges on the training set and shows a similar trend on the validation set, which means that the model has better generalization ability and is less likely to overfit. It shows that the model effectively learns sentiment classification.

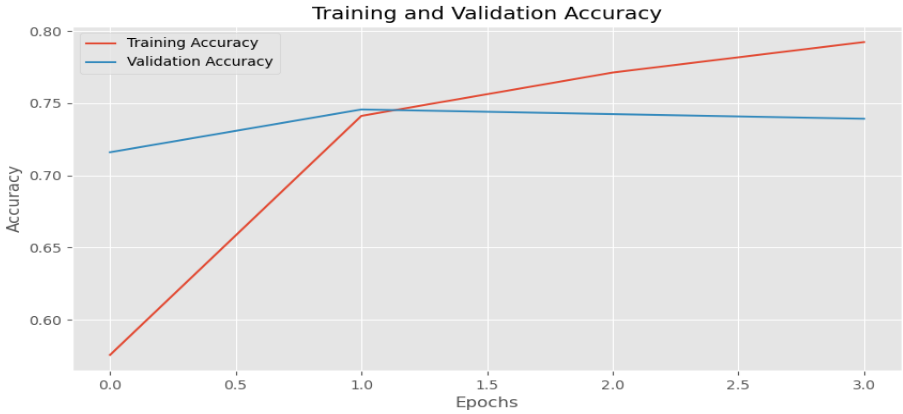

Fig. 2. Training and validation accuracy.

Fig. 3. Training and validation loss.

Fig. 3 depicts the corresponding changes in training and validation losses. As expected, the training loss decreases with increasing epochs, indicating that the model optimizes its parameters to minimize classification errors. In contrast, the validation loss initially decreases but then levels off, suggesting that the model's performance may not improve significantly after a certain point.

These results highlight the importance of monitoring accuracy and loss during model training. While high training accuracy is desirable, it must be accompanied by consistent validation accuracy and loss trends to ensure the model's ability to generalize.

3.2 Model Evaluation and Performance Metrics

To further evaluate the performance of the model, the author analyzes the ROC curve and confusion matrix. The ROC curve shown in Fig. 4 below provides the trade-off between the true positive rate (sensitivity) and the false positive rate (1-specificity) at different classification thresholds. The larger the area under the curve (AUC), the better the model performance. The model achieves an AUC of 0.74, indicating that the model has high accuracy and reliability in classifying positive and negative emotions.

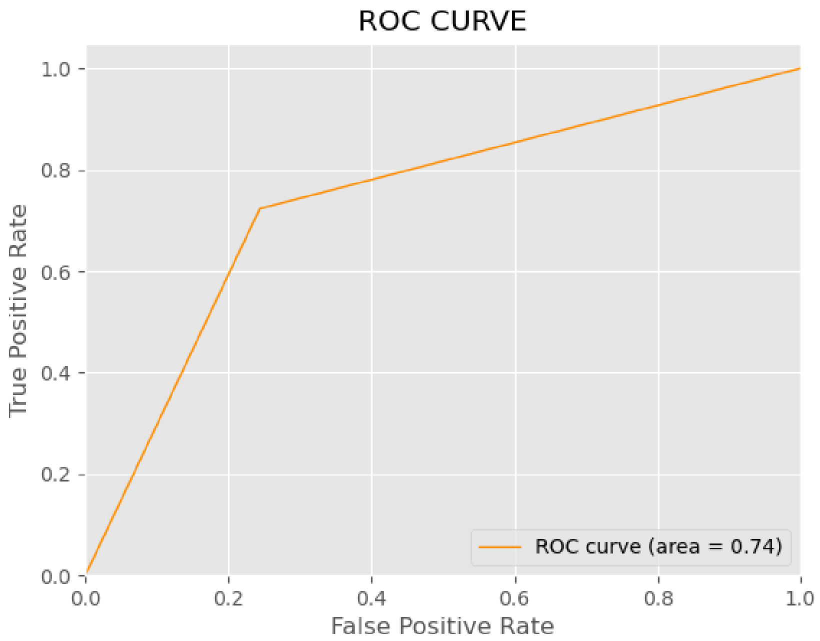

Fig. 4. ROC curve.

Table 1. Confusion matrix.

	Predicted Negative	Predicted Positive
Actual Negative	4733	1276
Actual Positive	1349	4642

At the same time, the author drew a confusion matrix, as shown in Table 1, which summarizes the classification results of the model on the test set and provides details of the prediction of true examples, true negative examples, false positive examples, and false negative examples. As can be seen from the confusion matrix, the model shows a good balance in predicting negative emotions and positive emotions, but there is a certain number of misclassifications. Overall, the model's performance in handling sentiment analysis tasks is still considerable, and it achieved high accuracy and low loss on the test set.

3.3 Discussion

Experimental results demonstrate the effectiveness of the proposed sentiment analysis model. Despite high training accuracy, the model's performance on the validation set plateaued, indicating a potential risk of overfitting. To solve this problem, the author uses techniques such as dropout regularization and learning rate decay to improve the generalization ability of the model. In addition, ROC curve and confusion matrix analysis demonstrated issues regarding model classification performance and error patterns, which are resolved by adjusting the model structure, optimizing algorithm parameters, and adopting appropriate loss functions. These methods enable the model to show good performance on the training set, validation set, and test set, with high accuracy and low loss. Future research can further improve the structure and algorithm of the model to improve the performance and generalization ability of the model in handling sentiment analysis tasks.

4 Conclusion

In this investigation, a novel sentiment analysis approach merges Bi-LSTM with the Att to accurately scrutinize sentiment expressions within textual data. The hybrid Bi-LSTM-Att model adeptly discerns sentiment expressions, achieving an accuracy rate of approximately 80%. Compared to utilizing a standalone Bi-LSTM model, this integration yields enhanced model training, recognition, and processing speeds without compromising accuracy. The superior performance of the Bi-LSTM-Att model stems from its concurrent capture of contextual dependencies and pivotal features in the input text. By amalgamating Bi-LSTM with Att, the model leverages bidirectional information flow and selective attention, facilitating a comprehensive understanding of text sentiment. Experimentation on the Sentiment140 dataset demonstrates the combined approach's superiority over traditional methods, showcasing heightened accuracy and adaptability across diverse text inputs.

Future investigations will confront potential limitations and delve into the model's performance in varied text domains and emotional contexts. Advanced deep learning architectures, such as transformer-based models, will be explored to augment model interpretability and generalization capabilities. Additionally, the study will pivot towards "word-sentiment" analysis, delving deeper into the semantic nuances of sentiment conveyed through text. This endeavor is anticipated to enrich the comprehension of emotion expression recognition and propel advancements in sentiment analysis research.

References

1. Koolagudi, S. G., & Rao, K. S.: Emotion recognition from speech: a review. International journal of speech technology, 15, 99-117 (2012).
2. Altman, D., Feigenblat, G., Herzig, J., Konopnicki, D., Rafaeli, A., Shmueli-Scheuer, M., & Spivak, D.: Classifying emotions in customer support dialogues in social media. In Proceedings of the 17th annual meeting of the special interest group on discourse and dialogue pp. 64-73. September (2016).
3. Chen, Y.: Convolutional neural network for sentence classification. Master's thesis, University of Waterloo (2015).
4. Guo, Y., Wang, Y., Yuan, Y., & Zhang, J. Sentiment boosting model for emotion recognition in conversation text. Journal of Computer Applications, 43(3), 706 (2023).
5. Cambria, E., Hazarika, D., Morency, L. P., Poria, S., Zadeh, A., & Zimmermann, R.: Conversational memory network for emotion recognition in dyadic dialogue videos. In Pro-

ceedings of the conference. Association for Computational Linguistics. North American Chapter, p. 2122. NIH Public Access (2018).
6. Cambria, E., Gelbukh, A., Hazarika, D., Majumder, N., Mihalcea, R., & Poria, S.: Dialoguernn: An attentive rnn for emotion detection in conversations. In Proceedings of the AAAI conference on artificial intelligence, vol. 33(1), pp. 6818-6825 (2019).
7. Dong, C., Han, T., Jiang, X., Ren, M., Zhang, Z., & Zhuang, Q.: Text emotion recognition based on xlnet-bigru-att. Electronics, 12(12), 2704 (2023).
8. Carbonell, J., Dai, Z., Le, Q. V., Salakhutdinov, R. R., Yang, Y., & Yang, Z.: Xlnet: Generalized autoregressive pretraining for language understanding. Advances in neural information processing systems, 32 (2019).
9. Mulcrone, K.: Detecting emotion in text. University of Minnesota–Morris CS Senior Semminar Paper, 2012.
10. Calderon, F. H., Chen, K. T., Chen, Y. S., Ghafoor, Y., Huang, Y. H., & Jinping, S.: TERMS: textual emotion recognition in multidimensional space. Applied Intelligence, 53(3), 2673-2693 (2023).
11. Sentiment dataset, https://www.kaggle.com/datasets/kazanova/sentiment140, last accessed 2023/10/7.