Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 1167462, 8 pages
https://doi.org/10.1155/2022/1167462

*Research Article*

# Attention-Based Bi-DLSTM for Sentiment Analysis of Beijing Opera Lyrics

**Cong Jin,[1] Zhen Song,[2] Jiaqi Xu,[2] and Huiyue Gao[2]**

[1] *School of Information and Communication Engineering, Communication University of China, Beijing 100024, China*
[2] *Advanced Research Center for Digitalization of Traditional Drama, The Central Academy of Drama, Beijing 100006, China*

Correspondence should be addressed to Zhen Song; songzhen@zhongxi.cn

Received 28 August 2021; Revised 15 October 2021; Accepted 19 April 2022; Published 6 May 2022

Academic Editor: Carles Gomez

We propose a sentiment analysis model based on Bi-DLSTM to solve the problem of sentiment analysis of Beijing Opera lyrics. A Bi-LSTM network with dilated recurrent skip connections (Bi-DLSTM) is introduced in this model, which can improve the ability to exact long-sequence information. The proposed model can learn the dependence of long sequences in different time dimensions and effectively improve the semantic extraction performance of lyrics. The attention mechanism is introduced to ensure the recognition of the more important words in the text sequence, which further improves the performance of the network. In order to solve the problem of lack of data on lyric sentiment analysis on the Internet, we build a dataset that can be used for lyric sentiment analysis. This paper completes multiple experiments on four datasets and verifies the effectiveness of the proposed model.

## 1. Introduction

Beijing Opera has a profound cultural heritage and literary value. Using artificial intelligence to analyze the emotions of Beijing Opera is helpful for its inheritance and development. With the development and improvement of Internet music software, various diversified services have been introduced, and intelligent song search and recommendation systems have gradually emerged. The emotional analysis of lyrics is the key technology to realize this function. Studies have shown that [1–3] lyrics are an important feature in the emotional classification of songs. When analyzing the emotional content of Beijing Opera, it is extremely valuable to consider the emotional attributes of the lyrics. At the same time, the field of natural language processing has developed rapidly, and neural networks have been widely used. Text sentiment analysis has been an important field in natural language processing, and the technology of text sentiment analysis has been very mature.

Lyric as a special form of text representation has more lyrical and artistic expressions and more complex semantics than ordinary written texts, such as lyrics: "叹只叹东风起火烧战船" and "何处悲声破寂寥." Therefore, sentiment analysis of the lyrics is more difficult. At present, the current work on sentiment analysis of lyrics is relatively scarce. When the ordinary text sentiment analysis method is used to deal with the lyrics with artistic expression, the extraction of sentiment semantics and deep features in the lyrics will be insufficient, which will affect the accuracy of lyric sentiment classification. Therefore, it is the key to excavate the emotional semantic information in lyrics accurately.

In this paper, we propose a sentiment analysis model based on Bi-DLSTM network, aiming at the problem that the traditional LSTM architecture will have a certain deviation when processing long-sequence information, because it can only capture semantics in one direction. We designed a Bi-DLSTM network composed of two layers of Dilated LSTM [4] network stack. The two layers have the same structure and opposite directions, which can capture bidirectional semantics. This variant LSTM network structure can not only realize parallel computing when the information transmission span becomes larger, but also can interpret and extract sequence information from a context perspective, which is more conducive to the retention of long-sequence information. In the model, the Bi-LSTM network and the Bi-DLSTM network are combined to extract information in different time dimensions to achieve deep mining of textual semantic information and emotional representation. The Bi-DLSTM network alleviates the problem of long-sequence learning, but it cannot identify the more important words in the text sequence. Therefore, the model introduces an attention mechanism to further improve the performance of the network.

## 2. Related Work

Sentiment analysis based on texts in specific fields such as lyrics belongs to the category of text sentiment analysis and is common research in natural language processing. At present, deep learning methods can simulate human brain thinking analysis methods obtaining data feature representations through automatic learning and encode the semantic and grammatical attributes of text to provide relatively accurate text representation information.

In the early research on natural language processing, the bag-of-words representation model (BOW) is one of the most common text representation models. Later, with the rise of neural networks, the concept of word embedding was proposed to solve the problem of dimensional disaster. The initial word embedding method was not mature and was not applied on a large scale. Bengio et al. [5] first proposed the NNLM language model using neural networks, which can be trained to obtain a low-dimensional dense word vector representation. Mikolov et al. [6] proposed a new word embedding method Word2Vec, which simplifies the structure of the NNLM model, and the pretrained word vector can effectively represent the text. Later, in order to solve the shortcoming that the model usually ignores the order of words in the text, researchers introduced sequence representation models, such as recurrent neural network (RNN) [7–9] and convolutional neural network (CNN) [10, 11]. With the development of the field of natural language processing and the change of various new technologies, many novel sentiment analysis models have emerged, such as structure-enhanced LSTM [12, 13] and the combination of RNN and CNN models [14]. Attention mechanism is a novel and effective technique, which has been widely used in natural language processing in recent years. It has shown advantages in many fields such as document classification [15], sentiment classification [16], and sentence representation [17].

At present, a large number of music-based researches have shown that it is necessary to conduct in-depth research on lyrics when analyzing and categorizing music. In most song classification tasks, the performance of music classifiers incorporating text features is better than pure audio classifiers [18, 19]. Lyrics are usually easier to obtain and process than audio data. For nonprofessional musicians, when interacting with the music system, they tend to pay more attention to the lyrics than audio information [20]. Kim and Kwon [21] proposed a feature selection method based on syntactic analysis, using four syntactic analysis rules to extract emotional features from lyrics for emotional classification of lyrics. Mihalcea and Strapparava [22] used music and lyrics to characterize songs and completed the classification of song emotions. Fell and Sporleder [23] proposed a lyrics analysis method that combines *n*-gram model and complex features to model lyrics text from different dimensions and achieved performance improvements in three different classification tasks. Rachman et al. [24] combined lyrics and audio features, using two emotional models, classification model, and dimensional model, to complete the emotional classification of songs.

## 3. Learning Model

*3.1. Embedding Layer.* In the first stage of the model, an embedding layer is applied to encode the lyrics text and emotional tags to obtain their unified representation. The word vector tool selected is Word2Vec. Before inputting the word vector, it is necessary to train the model based on the corpus. The lyrics of 1500 Beijing Opera songs collected on the Internet are selected as the corpus for training for the model. Previous studies [25, 26] showed that the highest performance of the model can be achieved when the dimensionality of the word vector is set in 100 dimensions. In the 100 dimension, it has good convergence, and the loss value is the smallest, so 100-dimensional word vector is selected for training of Beijing Opera lyrics. The trained lyrics vector is input into the Bi-LSTM network in the form of sequence $w = [x_1, x_2, \cdots, x_n]$ to learn the semantic and syntactic information of the lyrics.

After the training is completed, the word vector library can realize the functions of printing the corresponding word vector, comparing the semantic similarity, and adding and subtracting the word vector. For example, if the input word is "离开," the system can output its corresponding sequence of synonyms. The top three are "离去," "别离," and "消失," and their similarity scores are "0.846201," "0.759595," and "0.697776." The structure of the sentiment analysis model based on the Bi-DLSTM network is shown in Figure 1.

*3.2. Semantic Extraction Layer.* The semantic extraction layer uses the Bi-LSTM network because it is suitable for language and time series data [27]. The LSTM network maps the input $x$ to the output $y$ by learning the hidden representation $h_t$; the formula is as follows:

$$y = f(h_{t-1}, x_t). \tag{1}$$

$t$ represents different moments. The loss function needs to be minimized during training. The formula is as follows:

$$L(x, y) = -\frac{1}{N}\sum_{n\in N} x_n \log y_n. \tag{2}$$

The LSTM network manages its weight updates through gate structures, which determine the amount of information that should be retained and forgotten at each moment. The gate structure contains three kinds of "gates": "input gate" $i_t$, "forgotten gate" $f_t$, and "output gate" $o_t$. The "input gate" determines how much new information is added at each moment to update the current cell state; the "forgotten gate" determines not to keep irrelevant information, and the "output gate" determines the output and the value of the next hidden state. The calculation process is as follows:

$$\begin{aligned} i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i), \\ f_t &= \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f\right), \\ c_t &= f_t c_{t-1} + i_t \tan h(W_{xc}x_t + W_{hc}h_{t-1} + b_c), \\ o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o), \\ h_t &= o_t \tan h(c_t). \end{aligned} \tag{3}$$

$\sigma$ is the sigmoid function, $c_t$ is the unit state, $W$ is the weight matrix, $b$ is the bias term, and $h_t$ is the hidden state. Then Bi-LSTM is used to obtain information from both directions of each word to capture contextual information. It combines forward and backward hiding states. Forward hiding states $\overleftarrow{h_t}$ reading every word in the lyrics from $w_i1$ to $w_iT$, and backward hiding states $\overrightarrow{h_t}$ reading words in the lyrics from $w_iT$ to $w_i1$. The formula is as follows:

$$\begin{aligned} x_i t &= W_e w_i t, t \in [1, T], \\ \overleftarrow{h_t} &= \overleftarrow{LSTM}(x_i t), t \in [1, T], \\ \overrightarrow{h_t} &= \overrightarrow{LSTM}(x_i t), t \in [1, T]. \end{aligned} \tag{4}$$

Each lyric $t$ contains $w_i$ words, where $w_i, i \in [0, T]$ represents the $i$th word in the librettos and $x_i t$ represents the word vector matrix in the embedding.

Finally, all the outputs of the forward hidden state and the backward hidden state are connected to obtain the final output $o$, which contains the contextual semantic information in the librettos, then input $o$ into the Bi-DLSTM network.

*3.3. Bi-DLSTM Layer.* This layer of network consists of two layers of Dilated LSTM network stacks with opposite information transfer directions, and the dilation of each layer of network is set to 2, in order to achieve the effect of focusing on the extraction of sequence information in different time dimensions with the previous Bi-LSTM network layer. The most important part of the architecture is the dilated recurrent skip connection in the LSTM cell. The formula is as follows:

$$c_t^{(l)} = LSTM\left(o_t, c_{t-2}^{(l)}\right). \tag{5}$$

$c_t^{(l)}$ represents the cell state of the $l$th layer at time $t$, and $o_t$ is the input at time $t$ in the LSTM layer. The calculation process of the Bi-DLSTM network layer is as follows:

$$\begin{aligned} \overleftarrow{h_t} &= DLSTM\left(\overleftarrow{h_{t-1}}, s_t\right), \\ \overrightarrow{h_t} &= DLSTM\left(\overrightarrow{h_{t+1}}, s_t\right), \\ h_t &= \left[\overleftarrow{h_t}, \overrightarrow{h_t}\right]. \end{aligned} \tag{6}$$

$s_t$ represents the state of the input at time $t$, $\overleftarrow{h_t}$ represents the forward hidden state, and $\overrightarrow{h_t}$ represents the backward hidden state. We obtain the semantic information of a given word by connecting the forward and backward hidden state. Finally, the final semantic layer hidden representation $h_t$ is fed into the attention layer.

Compared with the traditional Bi-LSTM network, the Bi-DLSTM network layer proposed in this section has two advantages over the traditional Bi-LSTM network on the problem of long sequences: (1) it can focus on the extraction of sequence information in different dilated dimensions; (2) it reduces the path length of information transmission in the network and enables the network model to learn complex long-term dependencies. By combining the Bi-LSTM layer and the Bi-DLSTM layer, the model can learn the text sequence in two dilated dimensions, and extract the deep semantic information and emotional representation of the librettos text.

*3.4. Attention Layer.* The Bi-DLSTM layer solves the problem of long-sequence learning and performs more in-depth semantic feature extraction, but not every word or word in the lyrics sequence has the same importance and weight. Therefore, the model introduces an attention mechanism to expand this network.

The attention layer of this chapter is inspired by [14], through the attention network to find the most important words for the meaning of the lyrics. Therefore, an attention mechanism is introduced to extract words that are important to the meaning of a sentence, and the representations of these informative words are aggregated to form a sentence vector. Using the output of the Bi-DLSTM network as the input of the attention layer, the calculation formula is as follows:

$$\begin{aligned} u_{tw} &= \tan h(W_w h_t + b_w), \\ \alpha_{tw} &= \frac{\exp\left(u_{tw}^T u_w\right)}{\sum_t \exp\left(u_{tw}^T u_w\right)}, \\ L_t &= \sum_i (\alpha_{tw} h_t). \end{aligned} \tag{7}$$

$h_t$ represents the corresponding word annotation representation output by the Bi-DLSTM layer, $W_w$ and $b_w$ represent paranoid parameters, $u_{tw}$ is the corresponding hidden representation obtained through word annotations $h_t$ using a layer of MLP network and then calculates an importance weight $\alpha_{tw}$ by calculating the similarity between the hidden representation $u_{tw}$ and the word-level context vector $u_w$, which can measure the importance of the word in the document, and $L_t$ represents the corresponding lyrics vector.

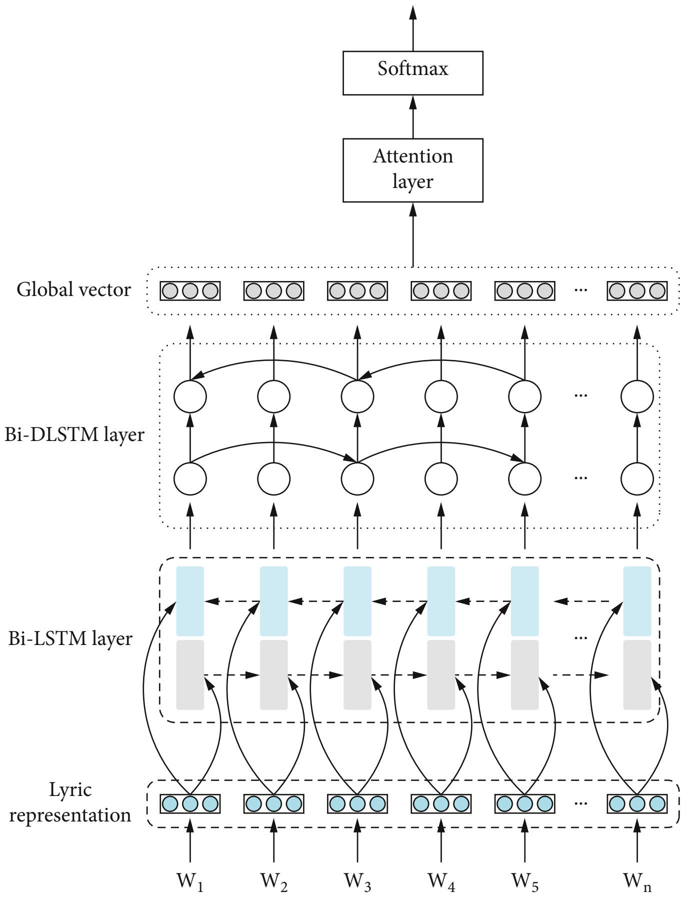

Figure 1: Model structure.

Then, the output of the attention layer obtained is calculated by the softmax function, and the classification result is obtained, and the formula is as follows:

$$P = softmax(W_t L_t + b_t). \tag{8}$$

$W_t$ and $b_t$ represent $softmax$ layer parameters. Finally, use the negative log likelihood of the correct label as the training loss:

$$Loss = -\sum_{L} \log p_{Ly}. \tag{9}$$

$y$ is the emotion label of the corresponding lyric document $L$.

## 4. Self-Built Data

At present, the open-source datasets for lyrics sentiment analysis on the Internet are relatively scarce and of low quality. We combined the working mechanism of the sentiment analysis model based on deep learning and the hardware experimental environment to construct a database that can be used for sentiment analysis of Beijing Opera lyrics.

In text sentiment analysis, the distinction between sentiment categories is usually more obvious. For example, based on the six basic emotions of Ekman [59], each sentiment category contains independent emotional attributes and does not contain dependencies. In this paper, it is not feasible to divide the lyrics into six basic emotions, because although the lyrics contain rich emotions, the types of emotions expressed are more concentrated, such as happiness, praise, love, sadness, and melancholy, while the emotions in lyrics such as fear, disgust, and anger are rare and not easy to define the classification. At the same time, each emotion has a different degree of expression in different lyrics, and there are different emotions. If the data is classified according to the six basic emotions, there will be great differences in the quantity and quality of the data between each category, and the difficulty of data sorting is also a big challenge. Therefore, for the creation of the dataset for lyrics sentiment analysis, we chose to build on the basis of two major emotional categories, positive and negative.

The original data of the lyrics comes from the lyrics of more than 1500 Beijing Opera songs of different styles on the Internet. Based on these lyrics, through the steps of sentence segmentation, emotional subjective analysis, and tagging, a dataset that can be used for lyrics emotional polarity analysis is created. Lyrics are divided into two major

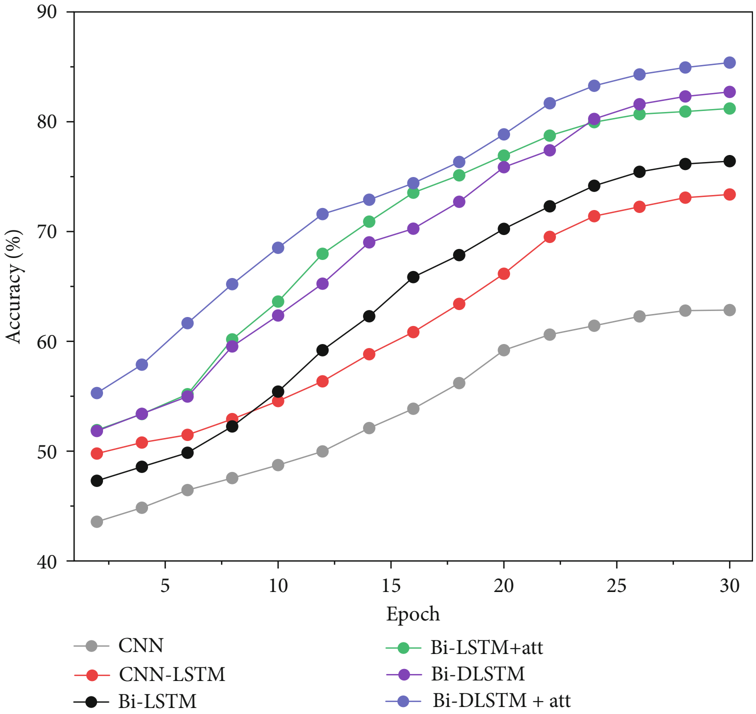

Figure 2: Classification accuracy of six models in self-built lyrics dataset.

emotional categories, positive and negative, and each degree of emotional category includes about 3500 lyrics; the number of words in each lyrics ranges from 7 to 20, and the average number of words is about 10.

## 5. Experiments

*5.1. Comparative Data.* We will use the self-built lyrics dataset and three Chinese pop songs based on datasets for comparative experiments, and the three Chinese pop songs based on datasets are divided into seven categories of emotions, including happiness, like, anger, sadness, fear, disgust, and surprise. There are some obvious expression differences between the lyrics of Beijing Opera and Chinese pop songs. For example, in Beijing Opera, "Happy" is expressed more frequently with the word "欢喜," but in Chinese pop songs, it is expressed more frequently with the word "快乐." We have already paid attention to that when collecting the data. We will use 80% of the data for training and the remaining 20% for testing.

*5.1.1. The Chinese Pop Song Lyrics Dataset One.* The Chinese pop song lyrics dataset one sentiment dataset contains 14000 Chinese pop songs lyrics with sentiment annotations and is based on seven emotions.

*5.1.2. The Chinese Pop Song Lyrics Dataset Two.* The Chinese pop song lyrics dataset two sentiment dataset contains close to 20,000 Chinese pop songs lyrics with sentiment annotations and is based on seven emotions.

*5.1.3. The Chinese Pop Song Lyrics Dataset Three.* The Chinese pop song lyrics dataset three contains close to 40,000 Chinese pop songs lyrics with sentiment annotations and is based on seven emotions.

Table 1: Model experiment comparison.

| Learning model | Precision | Recall | F1-score |
|---|---|---|---|
| CNN | 0.625 | 0.619 | 0.628 |
| CNN-LSTM | 0.731 | 0.743 | 0.741 |
| Bi-LSTM | 0.763 | 0.757 | 0.759 |
| Bi-LSTM+attention | 0.804 | 0.798 | 0.810 |
| Bi-DLSTM | 0.823 | 0.833 | 0.824 |
| Bi-DLSTM+attention | *0.854* | *0.846* | *0.849* |

The reason for collecting three Chinese pop songs lyrics dataset instead of only one is because Chinese pop songs have different genres; for instance, some of them are rock style with a fast rhythm, and some of them are romantic with a slow rhythm. So the lyrics may vary a lot according to the difference in the rhythm. Therefore, we built three datasets to reduce the influence from the different genres.

*5.2. Baselines.* We use five baseline models and the proposed model for comparative experiments.

*5.2.1. CNN.* The semantic extraction layer uses a CNN 2-D convolutional network with two fully connected layers. The size of the convolution kernel is 3, 4, and 5; the number of convolution kernels is 300; the learning rate is 0.001; batch size is 128; Adam optimizer, dropout is 0.5; and epoch is 30.

*5.2.2. CNN-LSTM.* The semantic extraction layer uses a network architecture that combines CNN and LSTM. The size

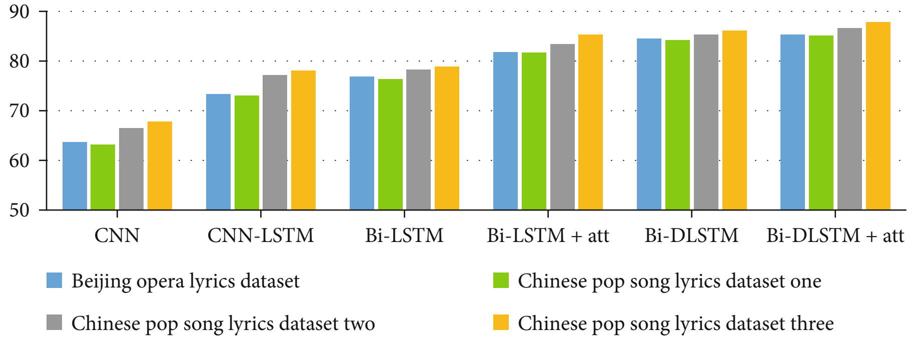

Figure 3: Comparison of classification accuracy in self-built lyrics dataset and Chinese pop song lyrics dataset.

of the convolution kernel is 3, 4, and 5; the number of convolution kernels is 300; the learning rate is 0.001; batch size is 128; Adam optimizer, dropout is 0.5; epoch is 30; and the number of LSTM hidden units is 128.

*5.2.3. Bi-LSTM.* The semantic extraction layer uses a Bi-LSTM network. The learning rate is 0.001, the batch size is 128, Adam optimizer, dropout is 0.5, epoch is 30, and the number of LSTM hidden units is 128.

*5.2.4. Bi-LSTM+Attention.* The semantic extraction layer uses a Bi-LSTM network with attention mechanism. The learning rate is 0.001, batch size is 128, Adam optimizer, dropout is 0.5, epoch is 30, and the number of LSTM hidden units is 128.

*5.2.5. Bi-DLSTM.* The semantic extraction layer uses a Bi-DLSTM network. The learning rate is 0.001, batch size is 128, Adam optimizer, dropout is 0.5, epoch is 30, the number of LSTM hidden units is 128, and the dilation is 2.

*5.2.6. Bi-DLSTM+Attention.* The semantic extraction layer uses a Bi-DLSTM network with attention mechanism. The learning rate is 0.001, batch size is 128, Adam optimizer, dropout is 0.5, epoch is 30, the number of LSTM hidden units is 128, and the dilation is 2.

*5.3. Sentiment Analysis Experiment.* In order to conduct a better comparison experiment, the texts marked as happy and favorite emotions in the three Chinese pop songs sentiment datasets were selected as positive sentiment categories, and other sentiments were used as negative sentiment categories and compared with the self-built lyrics dataset to evaluate the performance of the proposed sentiment analysis model.

*5.4. Evaluation Indicator.* The experiment uses four basic sentiment analysis evaluation indicators for evaluation: accuracy, precision, recall, and F1-score.

## 6. Results

As shown in Figure 2 and Table 1, the classification result of models without skip connections and attention mechanism is low. Among them, the CNN network has the lowest classification performance of all models, and the network model combining CNN and LSTM has achieved better classification results. It is experimental indicators that are equivalent to the performance of the two-way LSTM network. It is not difficult to see from the experimental data that the LSTM network with expanded jump connections shows better classification performance in sentiment analysis. After introducing a bidirectional Dilated LSTM network layer, the Bi-LSTM and Bi-LSTM+attention models have achieved about 5% improvement in classification accuracy. Similarly, the Bi-LSTM network has an accuracy increase of about 5% after the introduction of the attention mechanism, and the Bi-DLSTM network has an accuracy increase of about 3% after the introduction of the attention mechanism.

Among the accuracy, recall, and F1-score trained in the self-built lyrics dataset, the model proposed in this paper is also better than the other five comparison models. After the Bi-LSTM network introduces the two-way Dilated LSTM network layer, the three evaluation indicators are increased by 6%, 7.7%, and 6.5%, respectively. After the Bi-LSTM +attention network introduces the bidirectional Dilated LSTM network layer, the three evaluation indicators are increased by 5%, 4.8%, and 3.9%, respectively. At the same time, the Bi-LSTM network and the Bi-DLSTM network with the introduction of the attention mechanism have improved in all evaluation indicators. It fully shows that the Bi-DLSTM+attention network can improve the performance of long-sequence information extraction and sentiment analysis.

From Figure 3, we can tell that Bi-DLSTM+att has a better performance on all four datasets compared with all other models. On average, the results on Beijing Opera lyrics dataset are not as good as the outcome on the other three Chinese pop song lyrics dataset. One reason we suggest is that the context for Beijing Opera lyrics is different from Chinese pop song lyrics, and the other reason is that the sentence of Beijing Opera is shorter which may lead to less accuracy.

## 7. Conclusion

In this paper, we proposed a sentiment analysis model based on Bi-DLSTM network to solve the problem of semantic extraction and sentiment classification of long-sequence Beijing Opera lyrics, and construct a lyrics dataset suitable for lyrics sentiment analysis tasks. In order to explore the sentiment analysis performance of the model in this chapter, a coarse-grained sentiment analysis experiment was carried out on four datasets including self-built lyrics datasets, and the evaluation indicators obtained by training are better than the other five baseline methods. It reflects the excellent performance of this model in semantic extraction and sentiment analysis of long-sequence texts.

## Data Availability

The self-built dataset is collected from the website below: https://www.xikao.com/. The three comparative dataset (NLPCC2013/NLPCC2014/Chinese Pop Songs Emotion Corpus) is collected from the website below: https://www.mulanci.org/.

## Conflicts of Interest

There is no conflict of interest regarding the publication of this paper.

## Acknowledgments

This work is supported by the Beijing Outstanding Young Scientist Program (BJJWZYJH01201910048035), National Natural Science Foundation of China (61631016), Key Projects of National Key R D Program of China (2018YFB14039002), and Fundamental Research Funds for the Central Universities (CUC210B011).

## References

[1] C. Strapparava and A. Valitutti, "Word net-affect: an affective extension of word net," *LREC*, vol. 4, pp. 1083–1086, 2004.

[2] H. Corona and M. P. O. Mahony, "An exploration of mood classification in the million songs dataset," in *12th Sound and Music Computing Conference*, Dublin, Ireland, 2015.

[3] B. Rocha, R. Panda, and R. P. Paiva, "Music Emotion Recognition: The Importance of Melodic Features," in *International Workshop on Machine Learning and Music (MML)*, 2013, http://hdl.handle.net/10316/95166.

[4] A. M. Schoene, "Dilated LSTM with attention for classification of suicide notes," in *Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019)*, pp. 136–145, Hong Kong, 2019.

[5] Y. Bengio, R. Ducharme, and P. Vincent, "A neural probabilistic language model," *The Journal of Machine Learning Research*, vol. 3, pp. 1137–1155, 2003.

[6] T. Mikolov, W. Yih, and G. Zweig, "Linguistic regularities in continuous space word representations," in *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies*, pp. 746–751, Atlanta, Georgia, 2013.

[7] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[8] A. Graves, A. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 6645–6649, Vancouver, BC, Canada, 2013.

[9] J. Chung, C. Gulcehre, K. H. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, https://arxiv.org/abs/1412.3555.

[10] A. Rakhlin, *Convolutional Neural Networks for Sentence Classification*, GitHub, 2016.

[11] Q. Liu, H. Wu, Y. Ye, H. Zhao, C. Liu, and D. Du, "Patent litigation prediction: a convolutional tensor factorization approach," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)*, pp. 5052–5059, 2018.

[12] W. Li, F. Qi, M. Tang, and Z. Yu, "Bidirectional LSTM with self-attention mechanism and multi-channel features for sentiment classification," *Neurocomputing*, vol. 387, pp. 63–77, 2020.

[13] Y. Dong, Y. Fu, L. Wang, Y. Chen, Y. Dong, and J. Li, "A sentiment analysis method of capsule network based on BiLSTM," *IEEE Access*, vol. 8, pp. 37014–37020, 2020.

[14] M. E. Basiri, S. Nemati, M. Abdar, E. Cambria, and U. R. Acharya, "ABCDM: an attention-based bidirectional CNN-RNN deep model for sentiment analysis," *Future Generation Computer Systems*, vol. 115, pp. 279–294, 2021.

[15] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "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, San Diego, California, 2016.

[16] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based LSTM for aspect-level sentiment classification," in *Proceedings of the 2016 conference on empirical methods in natural language processing*, pp. 606–615, 2016.

[17] Z. Lin, M. Feng, C. N. Santos et al., "A structured self-attentive sentence embedding," 2017, https://arxiv.org/abs/1703.03130.

[18] R. Mayer, R. Neumayer, and A. Rauber, "Combination of audio and lyrics features for genre classification in digital audio collections," in *Proceedings of the 16th ACM international conference on Multimedia*, pp. 159–168, Vancouver, BC, Canada, 2008a.

[19] R. Mayer and A. Rauber, "Music genre classification by ensembles of audio and lyrics features," in *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, pp. 675–680, Miami, Florida, USA, 2011.

[20] D. Bainbridge, S. J. Cunningham, and J. S. Downie, "How people describe their music information needs: a grounded theory analysis of music queries," in *Proceedings of the International Symposium on Music Information Retrieval (ISMIR 2003)*, pp. 221-222, Baltimore, Maryland, USA, 2003.

[21] M. Kim and H. C. Kwon, "Lyrics-based emotion classification using feature selection by partial syntactic analysis," in *2011 IEEE 23rd International Conference on Tools with Artificial Intelligence*, pp. 960–964, Boca Raton, FL, USA, 2011.

[22] R. Mihalcea and C. Strapparava, "Lyrics, music, and emotions," in *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 590–599, Jeju Island, Korea, 2012.

[23] M. Fell and C. Sporleder, "Lyrics-based analysis and classification of music," in *Proceedings of COLING 2014, the 25th international conference on computational linguistics: Technical papers*, pp. 620–631, Dublin, Ireland, 2014.

[24] F. H. Rachman, R. Sarno, and C. Fatichah, “Music emotion classification based on lyrics-audio using corpus based emotion,” *International Journal of Electrical and Computer Engineering*, vol. 8, no. 3, 2018.

[25] J. Turian, L. Ratinov, and Y. Bengio, “Word representations: a simple and general method for semi-supervised learning,” in *Proceedings of the 48th annual meeting of the association for computational linguistics*, pp. 384–394, Uppsala, Sweden, 2010.

[26] A. M. Dai, C. Olah, and Q. V. Le, “Document embedding with paragraph vectors,” 2015, https://arxiv.org/abs/1507.07998.

[27] S. Hochreiter and J. Schmidhuber, “LSTM can solve hard long time lag problems,” in *Advances in Neural Information Processing Systems 9*, pp. 473–479, MIT Press, Cambridge MA, 1997.