



# An Augmented Neural Network for Sentiment Analysis Using Grammar

Baohua Zhang, Huaping Zhang\*, Jianyun Shang and Jiahao Cai

School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China

Understanding human sentiment from their expressions is very important in human-robot interaction. But deep learning models are hard to represent grammatical changes for natural language processing (NLP), especially for sentimental analysis, which influence the robot's judgment of sentiment. This paper proposed a novel sentimental analysis model named MoLeSy, which is an augmentation of neural networks incorporating morphological, lexical, and syntactic knowledge. This model is constructed from three concurrently processed classical neural networks, in which output vectors are concatenated and reduced with a single dense neural network layer. The models used in the three grammatical channels are convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and fully connected dense neural networks. The corresponding output in the three channels is morphological, lexical, and syntactic results, respectively. Experiments are conducted on four different sentimental analysis corpuses, namely, hotel, NLPCC2014, Douban movie reviews dataset, and Weibo. MoLeSy can achieve the best performance over previous state-of-art models. It indicated that morphological, lexical, and syntactic grammar can augment the neural networks for sentimental analysis.

## OPEN ACCESS

### Edited by:

Kok-Lim Alvin Yau,  
Tunku Abdul Rahman  
University, Malaysia

### Reviewed by:

Shubhamoy Dey,  
Indian Institute of Management  
Indore, India  
Nouhaila Idrissi,  
Mohammed V University, Morocco

### \*Correspondence:

Huaping Zhang  
kevinzhang@bit.edu.cn

**Received:** 16 March 2022

**Accepted:** 25 May 2022

**Published:** 01 July 2022

### Citation:

Zhang B, Zhang H, Shang J and Cai J  
(2022) An Augmented Neural Network  
for Sentiment Analysis Using  
Grammar.  
*Front. Neurobot.* 16:897402.  
doi: 10.3389/fnbot.2022.897402

**Keywords:** sentiment analysis, multi-channel CNN, grammar, augmentation, morphological

## INTRODUCTION

In the human-computer dialog, it is very important to recognize and understand the human sentiment, which can make human-computer interaction more intuitive, genuine, and natural. In the same scenario, the robot needs to choose appropriate feedback methods according to the different emotions of users. Sentimental analysis algorithms can give robots the ability to understand human emotions, but real interaction scenarios have high requirements for speed and accuracy, and existing methods are not effective.

Sentimental analysis is one of the traditional tasks in the field of natural language processing (NLP). In recent years, in large part due to increased internet access, a massive corpus of user opinion data is now available. The amount of these data on social media sites increases almost exponentially every day (Cheng et al., 2017). Using sentiment analysis technology to mine these data quickly is a hotspot of modern computer science research and software development.

More researchers have begun to do sentiment analysis using deep learning methods, such as bi-directional long short-term memory (BiLSTM), TextCNN, recurrent convolutional neural networks (RCNNs) (Kim, 2014; Zhou et al., 2016; Yang et al., 2019; Yuan et al., 2020), etc. These neural network models can automatically learn the features of the text and have significantly increased classification accuracy over former models; however, by directly mapping the input text to a vector space first and then learning the characteristics of different types of text through the

optimization of a matrix used by the neural network model have adequately satisfactory results, but do not emphasize the influence of grammatical rules (Cepukenas et al., 2015) of the input text. Our research shows that using the sentimental feature of the lexical which is the sentiment word of the whole sentence, in addition to the sentence features, as the input of the model can improve the model's ability to acquire the sentiment features of the sentence. Our experiments on Chinese texts, sentence structures, sentence patterns, punctuation, and the variety of conjunctive words contained in the sentence (we called them collectively syntactic rules) can collectively have positive impacts on the process of calculating the sentimental value of the input sentences. For example, the following sentence can be considered: "Although the weather is bad today and I missed the car when I went out, I still felt very happy today (虽然天气不好且出门误了车, 但是我今天玩的很开心)." It can be seen that the sentiment category of the first-half sentence is negative and the second-half sentence is positive, but the total sentimental value of the sentence is positive. The sentiment of the first-half sentence does not affect the sentiment of the whole sentence. This part of the information will be lost if only the text features are used as input. Our experiments hypothesized that using grammatical features of the input sentence as another input of the neural network could obtain more comprehensive feature extraction and classification accuracy.

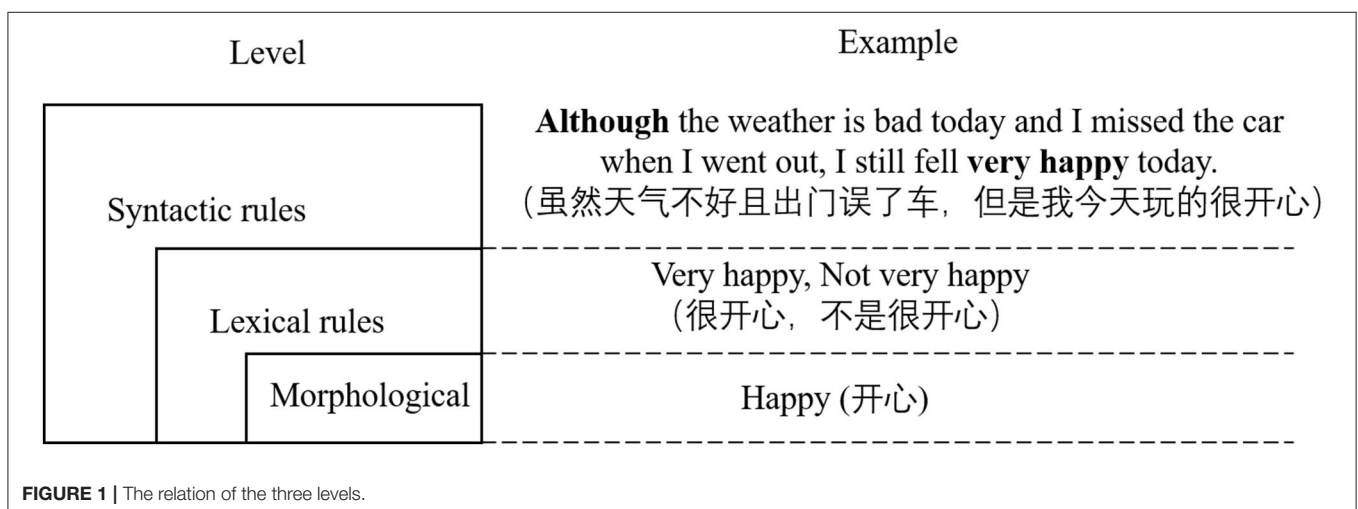
We divided the grammar rules into three levels, namely, morphological level, lexical rules (Zhang et al., 2015), and syntactic rules. **Figure 1** shows the relations of different levels. The morphological level is the bottom layer, which is the word input commonly used in sentimental analysis models. Complex sentiment word grammar rules are at the upper level of the morphological level, which are the rules about calculating the weight of complex sentimental words according to the different combinations of the simple sentiment words, negators, and degree adverbs. The syntactic rules are at the top level, which is explained in the "Syntactic rules level" section.

Based on the research of the neural network model and the three different grammar rule levels, this paper proposed MoLeSy, which takes the morphological level, lexical rule features, and syntactic rule features all as input in a parallel fashion. First, those inputs are simultaneously input to three independent neural networks and then merged into a single feature to get the input to the final dense neural network layer. The experiment results showed that MoLeSy can extract more comprehensive text features and have more accurate output in sentiment analysis in comparison to previously existing models. Thus, the three main contributions of this study are as follows:

- (1) The proposal of MoLeSy, an augmented neural network for sentiment analysis using grammar, which can obtain more comprehensive text features, and the accuracy of the inference of sentiment is higher than other models.
- (2) The novel proposal is to use morphological level, lexical rules, and syntactic rules as separate features and input them into different models for feature extraction to obtain other text features of sentences from the perspective of grammar rules.
- (3) The memorization of the grammar rules in Chinese, resulting in the research and development of a vector mapping method, and the construction of a feature extractor that directly extracts corresponding grammar features from sentences and converts them into feature vectors.

## RELATED WORK

Sentiment analysis is the main technical means for analyzing user attitudes and opinions. The current sentiment analysis techniques mainly include lexicon-based methods, machine learning algorithms, and deep learning models. The methods based on a sentiment lexicon need to construct or collect the sentiment lexicon first, then perform syntactic analysis on the sentence or directly match the sentiment word, and



finally, get the sentiment score of the sentence by accumulating the sentiment weight of the sentiment word. This method is simple but yields unsatisfactory results. The machine learning algorithms of sentiment analysis are often regarded as a classification of supervised machine learning, and the most commonly used methods are mainly maximum entropy, naive Bayes, and support vector machine (SVM). The machine learning method has a better classification result than the sentiment lexicon method, but also relies on manual feature selection and does not have satisfactory scalability.

Deep learning was first applied to the field of NLP by Collobert (Collobert et al., 2011) in 2011 to solve problems such as part-of-speech tagging. In 2014, Kim (2014) first used a convolutional neural network (CNN) (TextCNN) for text classification and achieved significantly improved classification results. Then, Kalchbrenner et al. (2014) proposed a wide convolution model and chose to use k-max pooling instead of the traditional CNN maximum pooling to retain more features. Zhang and Wallace (2017) compared the effects of different hyperparameters on the performance and stability of the CNN model structure through many repeated experiments. Gao et al. (2014) and Shen et al. (2014) introduced how to use CNN to learn to represent sentences into semantic structures. Zhang et al. (2016) proposed reliance on sensitive CNN and constructed a hierarchical text representation by processing pretrained word embeddings. CNNs are often used to capture local features, while recurrent neural networks (RNNs) can retain memory information due to their own feedback loop structure and have been applied in time series models (Cho et al., 2014). But, RNNs have shortcomings, as it has satisfactory results only on short input text, and when the length of the text increases, RNNs tend to exhibit gradient disappearance and gradient explosions (Li and Qian, 2016). LSTMs (Gers et al., 2000) and gated recurrent units (GRUs) (Ma et al., 2018) introduce the gate mechanism based on the traditional RNN, yielding improved even further improved results. Socher et al. (2013) proposed the tree-LSTM model, which can produce more text features. Tran et al. (2016) introduced an additional external memory unit based on LSTM to improve the model's ability to process historical information. However, due to the addition of a large number of parameters, the accuracy of this model was not improved much by this. Chen et al. (2017) proposed BiLSTMs, which both pre-order memory and post-order information, yielding better results. Wang et al. (2016a) established the AE-LSTM and ATAE-LSTM neural network models and finally obtained the sentimental classification of the aspect by modeling the LSTM, combining the hidden state of the text and the aspect information to generate the attention vector after the context modeling. But LSTM has shortcomings too. Although it can obtain the contextual semantic information of the text, it lacks the acquisition of the local information of the text. It can be seen from this that the simple model itself has certain limitations on text data processing due to its structure.

To use the advantages of RNN and CNN models at the same time, researchers began to merge the two types of models for text sentimental analysis. Wang et al. (2016b) constructed a fusion model by fusing a single-layer RNN and a single layer

CNN and conducted experiments on a short text data set, and the results proved that the fusion model was outperforming the single model effect. Xu et al. (2018) proposed a CNN with dual attention vectors. Zhou et al. (2000) proposed a model that combines CNNs and LSTMs serially. The text is first convolved through CNN and then input into LSTM to obtain features. Yoon and Kim (2017) proposed a multichannel CNN\_BiLSTM model combined with a dictionary. First, the text and dictionary are simultaneously input into the multichannel CNN to obtain the text features, and then the features are combined and then input into a BiLSTM for classification. Vo et al. (2017) suggested a parallel multichannel CNN and LSTM model. After embedding the text, it is input into the multichannel CNN and LSTM at the same time, and then the text features output by the two models are merged. After merging, they are input into the fully connected neural network. The experiment shows that the effect of this model on the separation of emotions in Vietnamese is better than that of a single model.

With the development of deep learning, an architecture combining multiple models has gradually become the mainstream of sentimental analysis methods. Wang et al. (2016c) proposed a regional CNN LSTM model. Yang et al. (2020) proposed a typical parallel two-layer network structure with multichannel CNN and BiGRU, which combines the features obtained by the two models when training. Li et al. (2020a) used a parallel structure to combine a BiLSTM and CNN. The difference is that they proposed a method of filling with emotional words, which can alleviate the problem of gradient disappearance. Li et al. (2020b) did not use a CNN model, but did use three BiLSTMs in parallel, and constructed the input of the three models by concatenating the part-of-speech vector, position vector, and dependent grammar vector with the text vector. Then, the attention mechanism is used to splice the three features. In this way, some grammar rules can be introduced to improve the accuracy of the model.

Yang et al. (2020) used the serial structure to combine CNNs and BiGRUs and added the attention mechanism to the last layer. Usama et al. (2020) used the serial method to combine an RNN and CNN, and the attention mechanism is added between the two models. Lin et al. (2020) proposed a comparison enhanced Bi-LSTM with multi-head attention (CE-B-MHA). Basiri et al. (2021) proposed a new model that not only has a parallel structure but also a serial structure. This model first combines a BiLSTM and BiGRU in parallel and then uses a CNN to perform convolution operations to extract local features and reduce feature dimensions. This architecture of combining multiple models has advantages over single independent models, but only Li et al. (2020a) incorporated the use of a few grammar rulers, while others ignore the influence of grammar entirely. Finally, Cheng et al. (2020) mentioned in their conclusion that it is necessary to consider integrating the syntactic structure features of traditional methods into the deep learning model, which is the main research area of this paper.

Although the existing research methods have begun to use the method of combining multiple models for sentimental analysis tasks, the current methods are still mainly based on input sentences, and a small number of models also use sentimental

dictionaries as input. However, few researchers pay attention to sentimental words and grammar features that play an important role in text sentiment. The research included in this paper indicates that when the sentimental analysis is performed, more text features can be obtained as text input at the same time, and the classification accuracy is improved.

## THE MOLESY

As shown in **Figure 2**, our model also includes three parts. The morphological level (refer to the “Morphological level” section for details), the lexical rules level (refer to the “Lexical rules level” section for details), and the syntactic rule level (refer to the “Syntactic rules level” section for details). In this model, the input of the multichannel CNN (MCNN) model is the word embedding vector matrix of the entire sentence. The text feature will be extracted from the multichannel CNN. The input of LSTM only includes the sentimental words, negators, and the degree adverbs, which have an important impact on the sentiment of the sentence. Since these words are more representative of the sentiment of the sentence, other factors are removed. We used LSTMs to extract the features of these words. The feature extractor mainly extracts the grammar features in sentences and converts them into feature vectors as the input of the fully connected layer. Then, we can get the feature extracted from the fully connected layer. Finally, the features obtained from the three models are merged and then used as input to the fully connected neural network to combine all the features of the text and classify them.

### Morphological Level

We used multichannel CNN to extract the morphological features from the text. **Figure 3** shows the details of our multichannel CNN. It is mainly composed of a convolutional layer, a pooling layer, and a fully connected layer. The convolutional layer is used to extract the features of the input data. It contains multiple convolution kernels. After the feature extraction of the convolutional layer, the output feature map will be passed to the pooling layer for feature selection and information filtering. The fully connected layer is equivalent to the hidden layer of the traditional feed-forward neural network and is generally connected to the output layer to achieve the final output.

In our model, we used  $N$  as the max-length of the sentences, and short sentences are padded using zeros. The embedding dim is  $d$ . We took  $S \in \mathbb{R}^{n \times d}$  as the input sentence matrix.  $W \in \mathbb{R}^{n \times d}$  is taken as the convolution kernel, and  $d$  is the length of the convolution kernel and is the same size as the dimension of the word embedding.  $h$  is the width of the convolution kernel. For the input matrix  $S \in \mathbb{R}^{n \times d}$ , the feature vector  $OC = (OC_0, OC_1, OC_2, \dots, OC_{n-h}) \in \mathbb{R}^{n-h+1}$  can be obtained by a convolution operation. The formula is as follows:

$$oc_i = W \cdot S_{i,i+h-1} \quad (1)$$

where  $i = 1, 2, 3, \dots, n - h$ ,  $(\cdot)$  represents the dot product operation of a matrix.  $S_{i,j}$  is the sub-matrix from row  $i$  to row

$j$  of  $S$ . In CNNs, pooling operations are generally performed after convolution is completed. In the pooling layer, using the maximum pooling operation, we can obtain the maximum value in each filter and extract the most significant features. Zhang et al. (2015) has proved that in various sentence classification tasks, the maximum pooling operation is always better than other pooling strategies in performance. In our model, we also used the 1-max pooling. The main idea is to capture the most important feature  $V$  corresponding to the specific feature map by selecting the maximum value of the specific feature map. The formula is as follows:

$$V = \max_{0 \leq s-h} \{oc_i\} \quad (2)$$

As shown in **Figure 3**, we used different sizes of convolution kernels to extract the features, and then we merged all of the three features into the next layer's input.

### Lexical Rule Level

At this level, the input of the model is the sentimental words of the text. In the work of sentimental lexicon build, we proposed that different combinations of negators, degree words, and simple sentimental words will have different sentimental weights. We used LSTM to extract the lexical rule features.

As mentioned in the related work section, RNNs can handle a certain short-term dependency, but if the sequence is long, then the gradient at the back of the sequence is difficult to backpropagate to the previous sequence, so it cannot handle the problem of solving for long-term dependency. In LSTMs, the cell state is introduced and the input is utilized. Different types of gates can save and control information, which can be used to solve the shortcomings of RNNs. So, in this section, we choose LSTMs to replace RNNs. The details of our model are shown in **Figure 4**.

At each step  $t$ , the LSTM unit will perform the following actions:

Step1. When the input gate  $i_t$  receives the current input  $x_t$ .

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \quad (3)$$

where  $h_{t-1}$  denotes the final hidden state and  $\sigma$  denotes a logistic sigmoid function.

Step2. The cell state at time  $t$  is  $\tilde{c}_t$ .

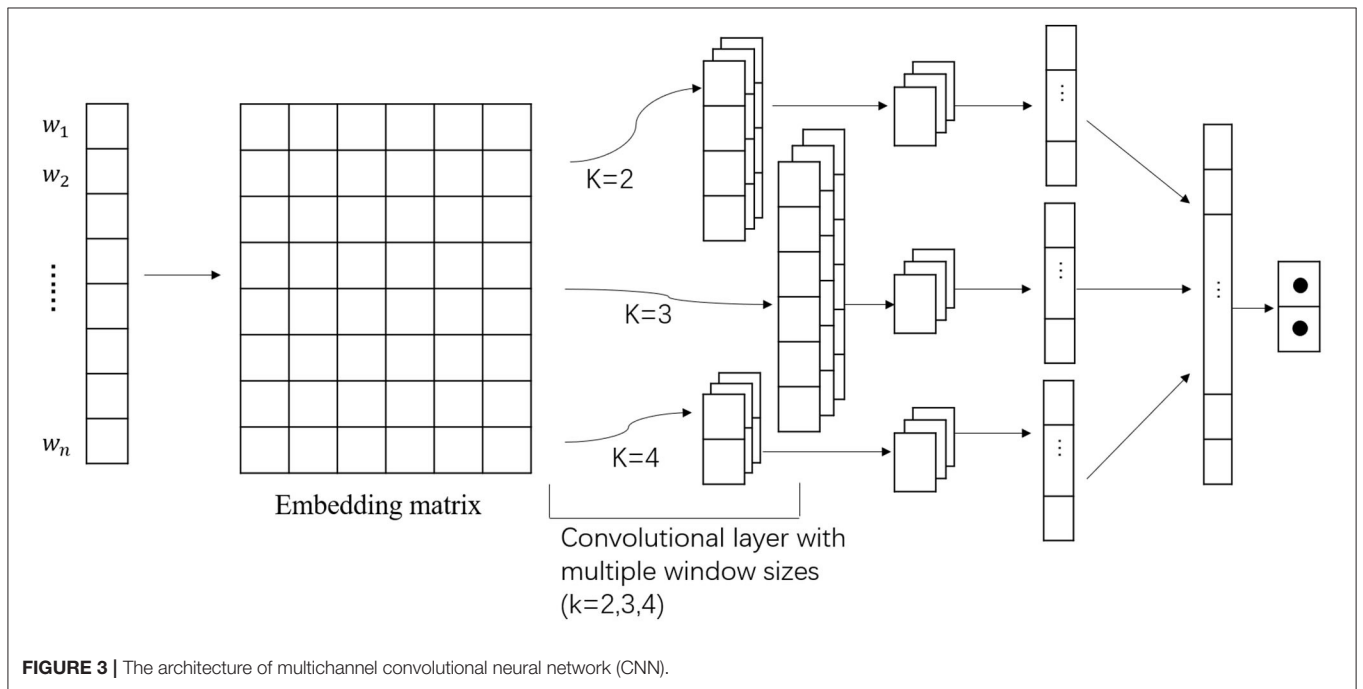
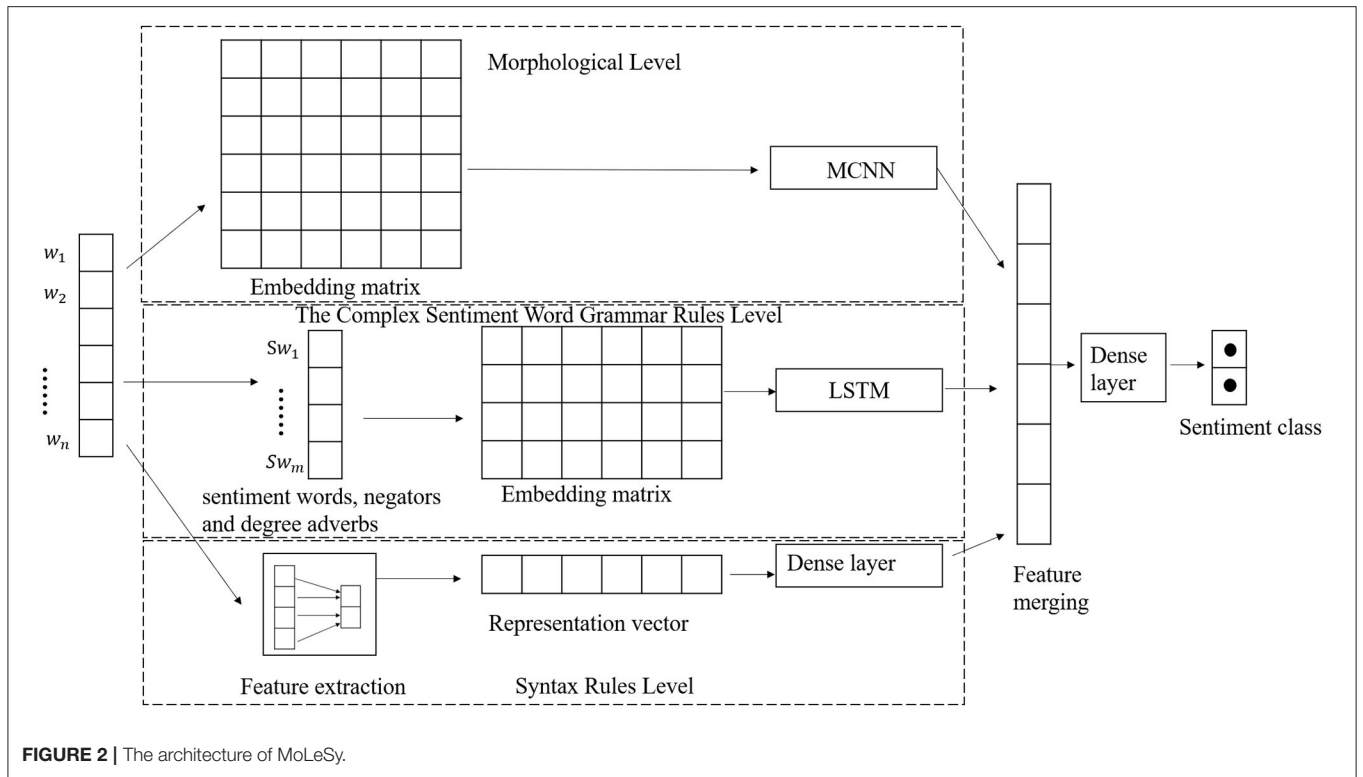
$$\tilde{c}_t = \tanh(W_{cx}x_t + W_{ch}h_{t-1} + b_c) \quad (4)$$

If  $\tilde{c}_t = 1$ , the current input information can enter the cell state, and if  $\tilde{c}_t = -1$ , the current input information cannot enter the cell state.

Step3. Then, the forget gate value at time  $t$  is calculated as  $f_t$ .

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \quad (5)$$

If  $f_t = 0$ , then the information does not pass, and if  $f_t = 1$ , then all information is passed to  $c_t$  by assigning  $c_{t-1}$  to  $c_t$ .



Step4. The cell state is then updated to be:

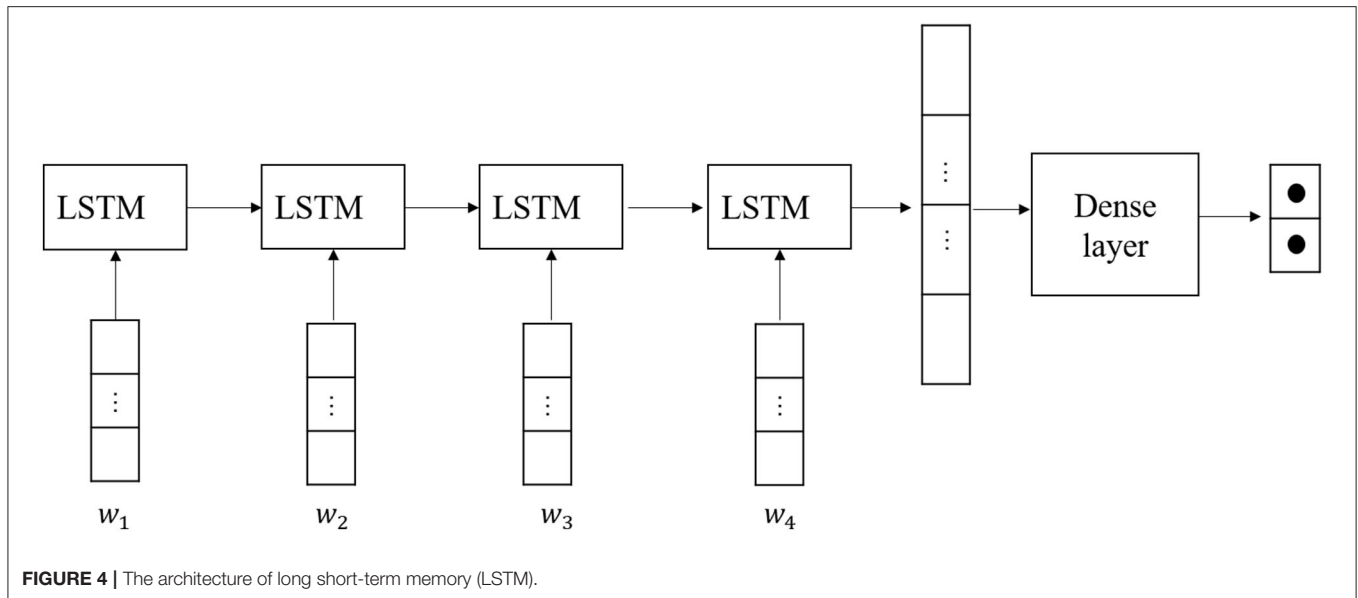
$$c_t = f_t c_{t-1} + i_t \tilde{c}_t \tag{6}$$

Step5. The LSTM is then assigned its final state value:

$$o_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \tag{7}$$

$$h_t = h_t \tanh(c_t) \tag{8}$$





## Syntactic Rule Level

According to our research, there are three main syntactic rules in a sentence: structural rules, sentence relation rules, and sentence pattern rules. We referred to these three rules collectively as grammar rules, and the following is the explanation for these three rules.

As for structural rules, we proposed the layer hierarchy for sentimental analysis before (Zhang et al., 2022), including chapters, complex sentences, single sentences, complex sentimental words, simple sentimental words, and characters. Since the sentimental analysis task is mainly in the sentence dimension, the structural rules in this study are only referred to complex sentences and single sentences. A complex sentence is composed of multiple single sentences. The sentimental weight of the complex sentence is accumulated by the sentimental weight of the single sentences according to some rules. The sentimental weight of a single sentence is its weight. Therefore, first, a sentence is marked as a single sentence or a complex sentence according to the punctuation in the sentence. If it is a complex sentence, it is mapped to the vector space according to the number of single sentences.

Sentence relation rules: there are mainly four types of inter-sentence relationship rules, namely, transition relationship, progressive relationship, causal relationship, and hypothetical relationship. In the transition relationship, according to the different transition words, a transition sentence can be divided into one sentence before the transition and another sentence after the transition. The sentimental polarity of the sentences before and after the transition is opposite, and the emotion of the latter sentence is emphasized; in the progressive relationship, the emotion of the preceding and following progressive sentences is gradually strengthened, and more emphasis is placed on the emotion of the latter progressive sentence; in

the causal relationship, the emotion of the cause is more emphasized; in the hypothetical relationship, more emphasis is placed on the condition, which weakens the emotion of the latter hypothetical sentence. Based on this, we mapped the turning words in every single sentence to the vector as input. Sentence pattern rules: according to the difference in ending punctuation, it is mainly divided into a declarative sentence, exclamation sentence, and interrogative sentence; among them, the sentence ending with the full stop is the declarative sentence, and the sentimental value of the sentence remains unchanged; the sentence ending with the exclamation mark is the exclamation sentence, and the emotional value of the sentence is enhanced; question sentences can be divided into antonym question sentences and question sentences according to whether they are antonym question words. Question sentences express no emotion, and antonym question sentences emphasize the reverse emotion of the sentence. Therefore, according to the ending symbol, the sentence pattern can be mapped into the vector space.

We can see that the three grammar rules all come from the sentimental analysis method based on the sentimental dictionary, and all influence the sentimental weight of the sentence. But, the existing deep learning models only focus on the text features, and some grammar ruler information will be ignored or changed when using a word embedding to map the text to the vector space and use a neural network model to extract features. This will lead to the lack of processing for some features, thereby reducing the accuracy of the model. We created the grammar vector (GV) feature with the **Pseudo-code 1**.

This model is a combination of three models, and the input of the LSTM and fully connected dense layer need to do some extra processing. Before we used a dataset to do training, we built the sentimental word lexicon for the datasets so that it could be used with this algorithm. Then, we used word2vec to do word

**Pseudo-code 1** | Framework of ensemble learning for our system.

---

**Input:** Sentence,  $S_i$ ; Maximum number of single sentences,  $MnS$ ;  
**Output:** Grammar rule vector,  $GV$ ;  
 Initialize  $GV = []$ ,  $GVsS = []$   $GVsS$  is the sentence struct feature vector;  
 $Sentence_{list} = Sentence\ segmentation(S_i)$ ;  
**if**  $length(Sentence_{list}) \leq MnS$  **then**  
    $GVsS = [1] * MnS$ ;  
**else**  $GVsS = [1] * length(Sentence_{list}) + [0] * [MnS - length(Sentence_{list})]$ ;  
**end if**  
**for** each  $simple_{sentence} \in Sentence_{list}$  **do**  
   Initialize  $GV_R = [0] * 8$   
   **if** the relation word in  $simple_{sentence}$  **then**  
      $Rwp = \text{find relation word position (relation word)}$   
      $GV_R[Rwp] = 1$   
   **end if**  
   Initialize  $GV_P = [0] * 4$   
   according to the type of punction in the simple sentence, set the  
   corresponding position of  $GV_P$  to 1.  
    $GV = GV + GV_R + GV_P$   
**end for**  
 $GV = GVsS + GV$

---

embedding. The multi-CNN layer uses the vector-matrix of the sentence as direct input. The LSTM just uses the vector-matrix of the sentimental words contained in the sentence as input. For the fully connected dense layer, we used a grammar vector as the input.

## EXPERIMENT

### Dataset and Parameter Setting

We evaluated our model on four datasets: the Tan Songbo hotel review dataset, the NLPCC2014 sentimental analysis task data set, the Douban review dataset (which was crawled by ourselves), and the Weibo comments data set. There are 5,000 positive comments and 5,000 negative comments in each hotel dataset and NLPCC2014 dataset. In addition, there are 50,000 positive comments and 50,000 negative comments in Douban and Weibo data sets, respectively (Fan and Li, 2021). **Table 1** shows the statistics of the datasets. The parameters are settled as detailed in **Table 2**.

### Ablation Experiments

To evaluate the contribution of different parts of this model, we removed various models and performed experiments. We choose accuracy (Acc) as the evaluation indicator.

As shown in **Table 3**, the models that include S\_LSTM in their names indicate that the input of that LSTM only includes sentimental words, and models that include NN in their names mean that the grammar ruler features were used as input. We can see that, after adding a dense NN layer, all models have higher accuracy than without a dense NN layer. This suggests that the syntactic features extracted from the dense NN model can help obtain more emotional information, thereby

**TABLE 1** | The statics of the datasets.

Dataset	Total number of comments	Maximum length of comments	Minimum length of sentences	Average number of simple sentences contained in each complex sentence	Total number of complex sentences
Douban	100,000	99	11	4.32	90,334
Hotel	10,000	1,985	4	8.74	9,861
nlpcc2014	10,000	1,004	3	5.48	9,258
weibo	119,988	260	1	4.23	100,146

**TABLE 2** | The parameter setting.

Maxlength of CNN and NN	300
Maxlength of LSTM	200
Embedding dim	300
Embedding method	Word2vec
CNN kernel size	2,3,4
The number of convolutional filters	150
CNN activation	Relu
LSTM units	128
Optimizer	Adadelta

improving the accuracy of the model. The results of the MoLeSy and MCNN\_S\_LSTM demonstrate that using sentimental words input plays a very important role in the inference of the sentiment of sentences.

### Comparative Experiment

We used Keras and Python version 2.7 to finish all models, and we trained those models on four GTX 1080 Ti with CUDA. We choose Acc, precision (P), recall (R), and F1 as the evaluation indicators, and we used all four datasets. We choose CNN, LSTM, multichannel CNN and LSTM (MCNNALSTM) (Wang et al., 2016c), SLCABG (Yang et al., 2020), ATTCNN RNN-rand (Usama et al., 2020), ABCDM (Basiri et al., 2021), SAMF-BILSTM (Li et al., 2020a), CNN-BiLSTM (sentimental word padding) (Li et al., 2020b), and MC-AttCNN-AttBiGRU (Cheng et al., 2020) as baseline.

The results are shown in **Tables 4, 5**. It can be seen from **Table 4** that the accuracy and the F1 value of our model are highest in all four datasets. This indicates that the use of grammar rules can help get more sentimental features and improve the result of the sentimental analysis. In addition, we found that many models take attention as an important part to extract the feature, which may be better when applied to the extraction of syntactic rules so that no feature extractor is needed. This will possibly be the focus of future research.

### The Influence of Different Kernel Sizes

To prove the influence of different sizes of convolution kernels on the experimental results, we conducted experiments on the

**TABLE 3** | The results of ablation experiments.

	Hotel	Douban	NLPCC
MoLeSy	<b>91.70</b>	<b>85.19</b>	<b>76.70</b>
MCNN_LSTM_NN	91.60	84.55	75.49
MCNN_S_LSTM	91.45	84.61	75.20
MCNN_LSTM	91.00	84.30	74.20
MCNN_NN	90.50	84.50	72.50
S_LSTM_NN	90.50	83.94	72.90
MCNN	90.35	83.55	71.25
S_LSTM	89.90	83.20	73.50
LSTM	88.04	83.02	71.55

*Bold values are the maximum number in each column.*

**TABLE 4** | The accuracy of different models on the Hotel and Douban.

	Hotel				Douban			
	P	R	F1	ACC	P	R	F1	ACC
CNN	88.18	89.40	88.78	88.61	82.61	86.17	84.36	83.94
LSTM	87.79	89.70	88.74	88.40	86.01	80.54	83.23	83.64
MCNNALSTM	90.61	90.79	90.70	90.52	85.38	84.16	84.77	84.16
SLCABG	91.35	88.91	90.12	90.15	86.13	83.89	85.00	85.11
ATTConv-RNN-rand	92.45	88.42	90.43	90.55	86.19	84.85	85.51	85.55
ABCDM	89.48	92.67	91.05	90.80	84.20	<b>86.91</b>	85.53	85.22
SAMF_BiLSTM	<b>93.25</b>	87.52	90.30	90.50	84.72	85.49	85.10	84.96
CNN-BiLSTM(SP)	92.32	88.12	90.17	90.30	86.13	83.89	84.77	84.79
MC-AttCNN-ArrBiGRU	91.35	90.99	91.17	91.10	85.53	86.80	86.16	85.99
MoLeSy	90.74	<b>93.08</b>	<b>91.87</b>	<b>91.72</b>	<b>88.27</b>	84.17	<b>86.17</b>	<b>86.42</b>

*Bold values are the maximum number in each column.*

Hotel data set. We controlled the size of the three convolution kernels to change, and other parameters remained unchanged, as shown in **Figure 5**, where the y-axis is the accuracy rate, and the x-axis is the size of the three convolution kernels. It can be seen that when the convolution kernel size is 2, 3, and 4, the accuracy is the highest. This is because, in Chinese, the two-character, three-character, and four-character words are the most used words, so the convolution kernel can extract the most morphological information.

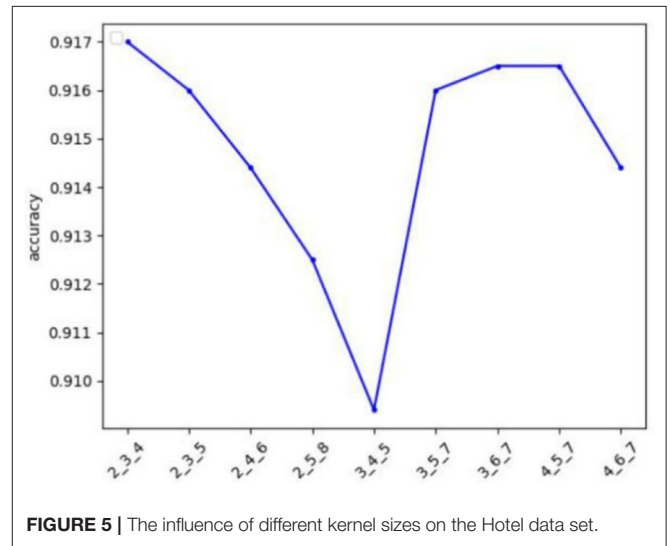
### Accuracy of LSTM-Based Models at Different Epochs

To study the accuracy of LSTM-based models in different epochs of training, we conducted experiments on the Hotel dataset, and the experimental results are shown in **Figure 6**, where epochs = 150, and it can be seen that the accuracy of LSTM\_CNN and MoLeSy starts to stabilize at 40 epochs. The accuracy of LSTM fluctuated greatly before 40 epochs and began to stabilize in the later epochs.

**TABLE 5** | The accuracy of different models on the Hotel and Douban.

	NLPCC				Weibo			
	P	R	F1	ACC	P	R	F1	ACC
CNN	78.32	67.67	72.57	73.56	91.26	96.85	93.97	93.77
LSTM	75.92	68.45	72.00	72.65	92.90	95.99	94.42	94.31
MCNNALSTM	74.36	75.50	74.93	73.80	94.03	95.66	94.84	94.78
SLCABG	79.08	71.46	75.08	75.40	96.19	96.96	96.57	96.57
ATTConv-RNN-rand	<b>81.70</b>	67.60	73.98	75.35	95.10	97.11	96.10	96.06
ABCDM	78.59	73.96	76.20	76.05	97.42	96.82	97.12	97.13
SAMF_BiLSTM	74.82	<b>78.20</b>	76.47	75.05	94.84	97.15	95.98	95.94
CNN-BiLSTM(SP)	78.32	71.07	74.52	74.80	94.04	<b>97.31</b>	95.65	95.58
MC-AttCNN-ArrBiGRU	75.99	78.11	77.03	75.85	98.30	96.79	97.54	97.56
MoLeSy	77.56	77.34	<b>77.45</b>	<b>76.65</b>	<b>99.72</b>	96.73	<b>98.20</b>	<b>98.23</b>

*Bold values are the maximum number in each column.*



**FIGURE 5** | The influence of different kernel sizes on the Hotel data set.

### Time Consumption

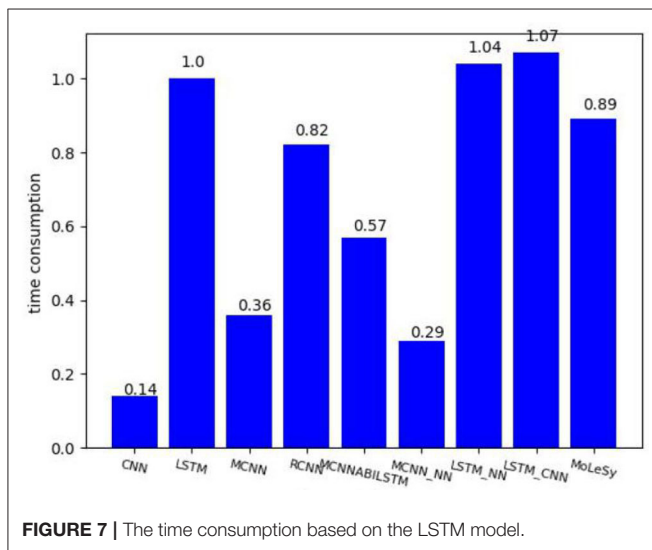
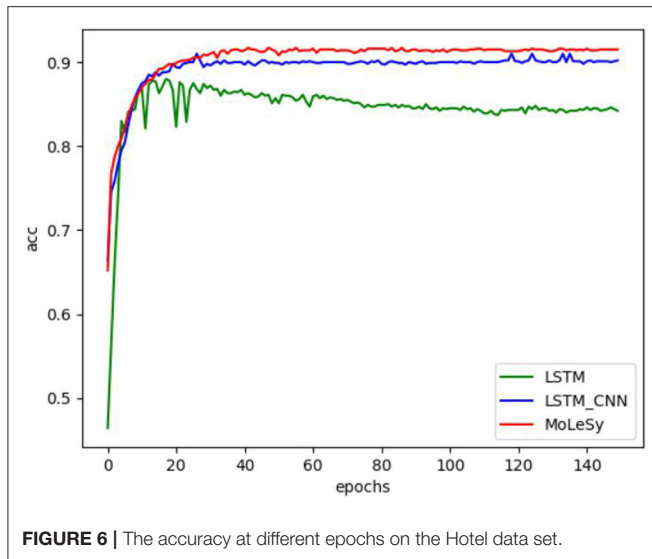
We also compared the time consumption of different models, took the LSTM model as the baseline, and the result is shown in **Figure 7**.

As we can see from **Figure 7**, the LSTM-based models need more time-consuming than the CNN-based model. MoLesy is a little faster than the LSTM model, and the reason is that the input of the LSTM layer in the MoLeSy model only contains sentimental words, which is shorter than the length of the whole sentences.

### CONCLUSION

In this paper, we proposed MoLeSy, an augmented neural network for sentimental analysis, based on the study of Chinese grammar rules and neural network models commonly used in sentimental analysis tasks. The multichannel CNN, LSTM,





fully connected dense neural networks, and input of different data for each model together obtain more comprehensive text features. At the same time, this paper proposes and constructs a syntactic rule extractor, which can extract syntactic rules from text and map them to vector space. Comparative experiments with other models on the Tan Songbo hotel data set, NLPCC2014 sentimental analysis task data set, Douban review data set, and

## REFERENCES

Basiri, M. E., Nemati, S., Abdar, M., Cambria, E., and Acharya, U. R. (2021). Abcdm: an attention-based bidirectional cnn-rnn deep model for sentiment analysis. *Fut. Generat. Comput. Syst.* 115, 279–294. doi: 10.1016/j.future.2020.08.005

Weibo data set showed that the model proposed in this paper can achieve the best performance over previous state-of-art models. Additionally, the model used in the experiments presented in this paper demonstrates that using the input text features as input while adding sentimental word features and grammatical features can achieve more accurate sentimental features. These results suggest a new methodology for Chinese sentimental analysis. This is the first time that all the rules in the method based on the sentimental lexicon for sentimental analysis have been used in the neural network model. The result of ablation experiments showed that the use of the grammar rule feature extraction model can help ascertain more sentimental information and increase sentimental features and final accuracy.

Our model can still maintain the highest accuracy with a faster processor speed than LSTM-based models, which proved that MoLeSy is suitable for the robot to understand human sentiment. We think this study will be helpful in human-robot interaction.

Our model is not without limitation, though: the method of extracting grammar rules is not spectacularly efficient, as only a small set of rules are encoded into the grammar vector, and may ignore some grammar rules as a result. Thus, in our future study, we intended on improving the extraction of grammar features, as we suspected that emphasizing this will reduce the proportion of text features lost. This may require many experiments to prove.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

BZ: conceptualization, methodology, data curation, software, and writing—original draft. HZ: funding acquisition and project administration. JS: writing—review and editing. JC: formal analysis. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by Fundamental Strengthening Program Technology Field Fund (2021-JCJQ-JJ-0059) and Natural Science Fund of Beijing (4212026).

Cepukenas, J., Lin, C., and Sleeman, D. (2015). “Applying rule extraction and rule refinement techniques to (blackbox) classifiers,” in *Proceedings of the 8th International Conference on Knowledge Capture*, pp. 1–5.

Chen, P., Sun, Z., Bing, L., and Yang, W. (2017). “Recurrent attention network on memory for aspect sentiment analysis,” in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 452–61.

- Cheng, K., Li, J., Tang, J., and Liu, H. (2017). "Unsupervised sentiment analysis with signed social networks," in *Thirty-First AAAI Conference on Artificial Intelligence*.
- Cheng, Y., Yao, L., Xiang, G., Zhang, G., Tang, T., and Zhong, L. (2020). Text sentiment orientation analysis based on multi-channel cnn and bidirectional gru with attention mechanism. *IEEE Access* 8, 134964–134975. doi: 10.1109/ACCESS.2020.3005823
- Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., et al. (2014). "Learning phrase representations using rnn encoder-decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734.
- Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa, P. (2011). Natural language processing (almost) from scratch. *J Mach Learn Res. (ARTICLE)* 12, 2493–2537. doi: 10.1016/j.chemolab.2011.03.009
- Fan, H., and Li, P. (2021). Short text sentiment analysis research based on fasttext word vector and two-way gru cyclic neural network—take weibo comment text as an example. *Inform. Sci.* 39, 15–22. doi: 10.13833/j.issn.1007-7634.2021.04.003
- Gao, J., Pantel, P., Gamon, M., He, X., and Deng, L. (2014). "Modeling interestingness with deep neural networks," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 2–13.
- Gers, F. A., Schmidhuber, J., and Cummins, F. (2000). Learning to forget: Continual prediction with lstm. *Neural Comput.* 12, 2451–2471. doi: 10.1162/089976600300015015
- Kalchbrenner, N., Grefenstette, E., and Blunsom, P. (2014). "A convolutional neural network for modelling sentences," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pp. 655–65.
- Kim, Y. (2014). "Convolutional neural networks for sentence classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1746–1751.
- Li, D., and Qian, J. (2016). "Text sentiment analysis based on long short-term memory," in *2016 First IEEE International Conference on Computer Communication and the Internet (ICCCI)* (New York, NY: IEEE), pp. 471–475.
- Li, W., Qi, F., Tang, M., and Yu, Z. (2020b). Bidirectional lstm with self-attention mechanism and multi-channel features for sentiment classification. *Neurocomputing* 387, 63–77. doi: 10.1016/j.neucom.2020.01.006
- Li, W., Zhu, L., Shi, Y., Guo, K., and Cambria, E. (2020a). User reviews: sentiment analysis using lexicon integrated two-channel cnn-lstm family models. *Appl. Soft Comput.* 94, 106435. doi: 10.1016/j.asoc.2020.106435
- Lin, Y., Li, J., Yang, L., Xu, K., and Lin, H. (2020). Sentiment analysis with comparison enhanced deep neural network. *IEEE Access* 8, 78378–78384. doi: 10.1109/ACCESS.2020.2989424
- Ma, D., Li, S., and Wang, H. (2018). "Joint learning for targeted sentiment analysis," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 4737–42.
- Shen, Y., He, X., Gao, J., Deng, L., and Mesnil, G. (2014). "A latent semantic model with convolutional-pooling structure for information retrieval," in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management—CIKM '14*, p. 101–110.
- Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., and Ng, A.Y., Potts, C. (2013). "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proceedings of the 2013 conference on empirical methods in natural language processing*, pp. 1631–42.
- Tran, K. M., Bisazza, A., and Monz, C. (2016). "Recurrent memory networks for language modeling," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 321–331.
- Usama, M., Ahmad, B., Song, E., Hossain, M. S., Alrashoud, M., and Muhammad, G. (2020). Attention-based sentiment analysis using convolutional and recurrent neural network. *Fut. Generat. Comput. Syst.* 113, 571–578. doi: 10.1016/j.future.2020.07.022
- Vo, Q.-H., Nguyen, H.-T., and Le B, Nguyen, M.-L. (2017). "Multi-channel lstm-cnn model for vietnamese sentiment analysis," in *2017 9th international conference on knowledge and systems engineering (KSE)* (New York, NY: IEEE), pp. 24–29.
- Wang, J., Yu, L. C., Lai, K. R., and Zhang, X. (2016c). "Dimensional sentiment analysis using a regional cnn-lstm model," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 225–230.
- Wang, X., Jiang, W., and Luo, Z. (2016b). "Combination of convolutional and recurrent neural network for sentiment analysis of short texts," in *Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical papers*, pp. 2428–2437.
- Wang, Y., Huang, M., Zhu, X., and Zhao, L. (2016a). "Attention-based lstm for aspect level sentiment classification," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 606–15.
- Xu, Y., Yang, Y., Han, J., Wang, E., Zhuang, F., and Xiong, H. (2018). "Exploiting the sentimental bias between ratings and reviews for enhancing recommendation," in *2018 IEEE International Conference on Data Mining (ICDM)*, (New York, NY: IEEE), pp. 1356–61.
- Yang, F., Du, C., and Huang, L. (2019). Ensemble sentiment analysis method based on r-cnn and c-rnn with fusion gate. *Int. J. Comput. Commun. Control* 14, 272–285. doi: 10.15837/ijccc.2019.2.3375
- Yang, L., Li, Y., Wang, J., and Sherratt, R. S. (2020). Sentiment analysis for e-commerce product reviews in chinese based on sentiment lexicon and deep learning. *IEEE Access* 8, 23522–23530. doi: 10.1109/ACCESS.2020.2969854
- Yoon, J., and Kim, H. (2017). "Multi-channel lexicon integrated cnn-bilstm models for sentiment analysis," in *Proceedings of the 29th Conference on Computational Linguistics and Speech Processing (Rocling 2017)*, pp. 244–415.
- Yuan, J., Wu, Y., Lu, X., Zhao, Y., Qin, B., and Liu, T. (2020). Recent advances in deep learning based sentiment analysis. *Sci. China Technol. Sci.* 20, 1–24. doi: 10.1007/s11431-020-1634-3
- Zhang, B., Li, E., Zhang, H., and Shang, J. (2022). The representation method of affective unit based on hierarchical system. *Comp. Eng. Sci.* 1, 149–158.
- Zhang, B. Y., and Wallace, A. (2017). "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification," in *Proceedings of the Eighth International Joint Conference on Natural Language Processing*, p. 253–63.
- Zhang, R., Lee, H., and Radev, D. (2016). "Dependency sensitive convolutional neural networks for modeling sentences and documents," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1512–21.
- Zhang, Z., Nie, J.-Y., and Wang, H. (2015). "Tjudem: a combination classifier for aspect category detection and sentiment polarity classification," in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pp. 772–777.
- Zhou, C., Sun, C., Liu, Z., and Lau, F. (2000). *A c-lstm neural network for text classification*, arXiv preprint arXiv:1511.08630.
- Zhou, P., Qi, Z., Zheng, S., Xu, J., Bao, H., and Xu, B. (2016). "Text classification improved by integrating bidirectional lstm with two-dimensional max pooling," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pp. 3485–3495.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Zhang, Shang and Cai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.