Research on Text Sentiment Analysis Based on Attention C_MGU

Diangang Wang[1], Lin Huang[1], Xiaopeng Lu[2](✉), Yan Gong[1], and Linfeng Chen[3]

[1] State Grid Sichuan Information and Communication Company, Chengdu 610041, Sichuan, China

[2] School of Computer Science and Technology, Shanghai University of Electric Power, Shanghai 200090, China
2217297003@qq.com

[3] School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, Zhejiang, China

Abstract. Combining the advantages of the convolutional neural network CNN and the minimum gated unit MGU, the attention mechanism is merged to propose an attention C_MGU neural network model. The preliminary feature representation of the extracted text is captured by the CNN's convolution layer module. The Attention mechanism and the MGU module are used to enhance and optimize the key information of the preliminary feature representation of the text. The deep feature representation of the generated text is input to the Softmax layer for regression processing. The sentiment classification experiments on the public data sets IMBD and Sentiment140 show that the new model strengthens the understanding of the sentence meaning of the text, can further learn the sequence-related features, and effectively improve the accuracy of sentiment classification.

Keywords: Sentiment analysis · C_MGU · Attention mechanism

1 Introduction

With the widespread application of the Internet, the Internet (such as blogs, forums, and social service networks such as Douban, Public Comment) has generated a large number of user-engaged comment information with emotional factors such as people, events, and products. Recording various life states in the form of text to express emotions and attitudes has occupied an increasingly important position in people's daily communication. As a carrier of emotion, through the study of network text information, it is possible to analyze the emotional changes of network users, understand the selection preferences of network users from multiple aspects, and better understand the behaviors of network users. In addition, the government can supervise and manage the network,

Science and technology projects funded by State Grid Sichuan Electric Power Company (NO.: 521947140005).

J. Liu et al. (Eds.): MobiCASE 2020, LNICST 341, pp. 163–173, 2020.
https://doi.org/10.1007/978-3-030-64214-3_11

maintain Internet order, and use the power of public opinion to guide network users to form correct values and prevent network deviations from extending into real life.

Emotion is the advanced behavior of human intelligence, and people express emotions in a variety of ways. In order to understand the emotions in the text, the emotions need to be classified. The purpose is to classify the emotions into positive, negative emotions or more detailed emotion categories [1]. Text sentiment analysis is also known as opinion mining, which uses natural language processing, text analysis, and other methods to analyze, process, reason, and summarize texts with emotions [2, 3]. Existing sentiment analysis methods are mainly divided into two categories based on sentiment dictionary matching methods and machine learning-based methods.

As deep learning has gradually become a research hotspot in the field of natural language processing, the technology of using deep learning methods based on sentiment dictionary matching to solve sentiment analysis problems has also developed rapidly. Many scholars have optimized the sentiment characteristics of text sentiment analysis [4]. Zhang et al. [5] proposed a strategy based on sentiment dictionary, which successfully classified sentiment of online user-generated text. Zhang et al. [6] proposed a sentiment dictionary-based sentiment analysis method for Chinese Weibo texts. The sentiment values of Weibo texts were obtained by calculating weights, and then sentiment classification was performed. Wu et al. [7] help to perform sentiment analysis of social media content by using Web resources to build an easily maintained and expandable sentiment dictionary.

In the study of sentiment analysis based on machine learning, because deep learning has the advantages of local feature abstraction and memory function, it can avoid a large number of artificial feature extraction and other advantages. Research has applied deep neural network-based text classification methods to sentiment analysis, the most popular of which are CNN and RNN models. Hu et al. [1] used a recurrent neural network to perform multimodal fusion sentiment analysis on English text. Kim et al. [8] proposed an improved CNN for text classification. Zhao et al. [9] improved the proposed cyclic random walk network model by using back propagation method, and successfully classified sentiment on TWitter's blog posts, showing good performance. Bai et al. [10] used the Bilstm recursive neural network to extract features and combined context semantics to perform position detection on Sina Weibo Chinese text sentiment. Hu et al. [11] combined a long-term and short-term memory network in a recurrent neural network with a feedforward attention model and proposed a text sentiment classification scheme. Chen et al. [12] used the proposed multi-channel convolutional neural network model to classify sentiment in Chinese microblog text.

CNN, RNN and other neural networks have achieved good results in the field of sentiment analysis in natural language processing, but there are still some problems. For example, CNN was originally widely used in the field of image processing, and it is not very suitable for processing text sequence problems. Although deep neural networks RNNs can achieve good results in processing text classification tasks, RNNs have long-term dependencies and may face gradient explosion or gradient disappearance problems [13]. In order to solve this series of problems, many neural network variants such as LSTM and GRU have been proposed successively, and have been successfully applied in the field of sentiment analysis. LSTM is able to memorize the context of a sequence, which has obvious advantages for the emotional feature extraction of text sequences, and can solve the problem of vanishing gradients. As a variant of LSTM, GRU has a simpler structure than LSTM, with fewer parameters and faster convergence, and can solve the problem of long-distance dependence. In the latest text classification research, it is found that compared with GRU, the smallest gated unit MGU has the advantages of simpler structure, fewer parameters, and less training time. It is very suitable for tasks with strong time dependence, while using fewer parameters. The MGU model can reduce the workload of selecting parameters and improve the generalization ability of the model [14].

The MGU model does not fully learn the sequence-related features such as the sequence in which the text is generated in time. The convolutional neural network CNN can extract the features of the data through convolution operations, and can enhance certain features of the original data and reduce the noise. Impact. When a similar human brain recognizes a picture, it does not recognize the whole picture at once, but it first perceives each feature in the picture first, and then integrates the parts at a higher level. Operation to get global information.

The Attention mechanism is a resource allocation mechanism, that is, at a certain time, your attention is always focused on a fixed position on the screen without paying attention to other parts. The Attention mechanism was originally only applied to the task of image recognition in computer vision, and then applied to image-text conversion. In natural language processing, Attention is often used in combination with various neural networks to help neural networks focus on some important information in text sequences.

Therefore, in this paper, a new attention C_MGU model is proposed by combining the network structure of CNN and the minimum gating unit MGU and introducing the attention mechanism. This model combines the convolutional layer and the smallest gated unit MGU in CNN with a unified architecture, and adds the attention mechanism between the convolutional layer and the MGU layer. The new model not only reflects the advantages of the CNN model and the smallest gated unit MGU, but also uses the attention mechanism to further optimize the semantic feature representation. The experimental results show that the new model can not only highlight the key information of the text, but also mine richer semantics, and has a better performance in the sentiment classification of the text.

2 Recurrent Neural Network with Gate Structure

2.1 LSTM and GRU

RNN has been proven to be very successful in the field of natural language processing in practice, but when faced with long sequences of text, the gradient of the hidden layer variables of RNN may attenuate or explode. Although gradient clipping can cope with gradient explosion, it cannot solve gradient attenuation. Therefore, given a text sequence, it is actually more difficult for a recurrent neural network RNN to capture text elements (words or words) with a large distance between two moments. LSTM (Long Short Term Memory) memory unit is based on the RNN memory unit with a gate control mechanism. Its structure is shown in Fig. 1. It implements three gate calculations, namely the forget gate, input gate and output gate [15]. The Forget Gate is responsible for deciding how many unit states from the previous moment to the unit state for the current moment; the Input Gate is responsible for deciding how many unit states from the current moment to the unit state at the current moment; Output Gate Responsible for determining how many outputs the unit status has at the current moment. Each LSTM contains three inputs, namely the unit status at the last moment, the output of the LSTM at the moment in time, and the input at the current moment. The historical information can be filtered to solve the problem of gradient disappearance.

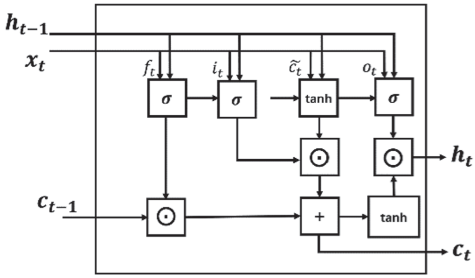

Fig. 1. LSTM

Because there are more learning parameter settings in LSTM and longer training time, GRU is proposed as an improved version of LSTM [16]. Compared with LSTM, the GRU has one less gated unit and fewer parameters, so the computing time and convergence speed are greatly improved. Its structure is shown in Fig. 2.

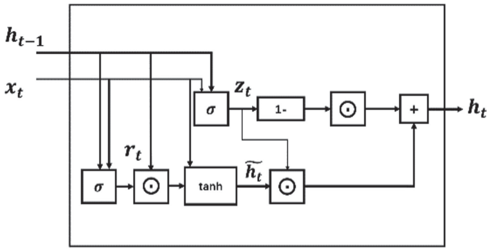

Fig. 2. GRU

The GRU model is shown in the figure. It has only two gates, namely the update gate and the reset gate, which are z and r in the figure. The update gate is used to control the degree to which the state information of the previous moment is brought into the current state. The larger the value of the update gate is, the more state information is brought into the previous moment. The reset gate is used to control the degree of ignoring the state information of the previous moment. The smaller the value of the reset gate is, the more it is ignored. At time t, for a given input x_t, the hidden output h_t of the GRU is calculated as follows:

$$r_t = \sigma\left(W_r \cdot \left[h_{t-1}, x_t\right]\right) \tag{1}$$

$$z_t = \sigma\left(W_z \cdot \left[h_{t-1}, x_t\right]\right) \tag{2}$$

$$\tilde{h}_t = \tanh\left(W_{\tilde{h}} \cdot \left[r_t * h_{t-1}, x_t\right]\right) \tag{3}$$

$$h_t = (1 - Z_t) * h_{t-1} + Z_t * \tilde{h}_t \tag{4}$$

Where [] indicates that the two vectors are connected, W is the weight matrix of the connection layer, and tanh is the activation function.

2.2 Minimum Gated Unit MGU

The minimum gated unit MGU is a simplified recurrent neural network structural unit. The gate number is the smallest in any gated unit, and there is only one forget gate. The input gate is merged into the forget gate. The structure is shown in Fig. 3.

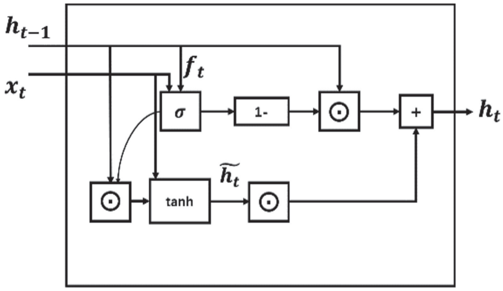

Fig. 3. MGU

As can be seen from the structure diagram of the MGU, the structure of the MGU is obviously simpler than that of the LSTM and GRU. The LSTM has four sets of parameters to be determined, and the GRU needs three sets of parameters, but the MGU only needs two sets of parameters. In the gating unit, the recurrent neural network has a simpler structure, fewer parameters, and faster training.

3 Sentiment Analysis Model Based on Attention C_MGU

3.1 Model Design

Combining the respective advantages of CNN and MGU, in order to achieve the goal of text sentiment analysis, this paper proposes a network structure that combines CNN with the smallest gating unit MGU and introduces the attention mechanism. The model structure is shown in Fig. 4. First, the model captures the preliminary feature representation of the extracted text through the CNN's convolutional layer module, and then uses the Attention mechanism and the minimum gating unit MGU module to strengthen the key information of the preliminary feature representation of the text. With further optimization, the final deep feature representation of the text is generated in the hidden layer of the MGU, and it is input to the Softmax layer for regression processing, and the classification of the text is finally completed.

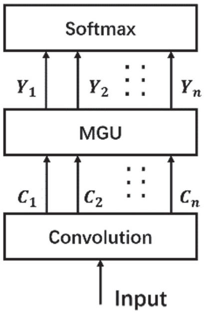

Fig. 4. C_MGU attention

3.2 Algorithm Design

The specific process of the model structure proposed in this paper is: firstly capture the preliminary feature representation of the text through the CNN's convolution layer, and then use the attention mechanism and the MGU model to obtain the deep features of the text with keyword discrimination in the hidden layer of the MGU Indicates that the state of the hidden layer is finally input to the Softmax layer for regression classification processing to complete the classification processing of the text.

Convolution Layer Feature Extraction. The first module of this model is to extract the preliminary feature representation of the text through the convolutional layer in CNN. By defining the region vector RSV to maintain the original sequence corresponding to the input sentence, the MGU model of the specified sequence input provides a reasonable input. The purpose of using the convolution layer in this module is to ensure a reasonable sequence vector as the input vector for subsequent MGUs. The pooling layer will destroy the original word order of the original sentence and cannot be used as a reasonable input for the MGU, so only the convolution layer is used. Define filters in a one-dimensional convolutional layer to extract local feature representations of different text positions, $\boldsymbol{w_j} \in \boldsymbol{R^D}$ means D-dimensional word embedding representing the jth word in a sentence. $\boldsymbol{x} \in \boldsymbol{R^{I \times D}}$ represents an input sentence of length L, $\boldsymbol{f} \in \boldsymbol{R^{K \times D}}$ is the filter with length $\boldsymbol{K}$ in the convolutional layer. The region sequence vector is marked with $\boldsymbol{S_i}$ and is composed of $\boldsymbol{K}$ word embeddings starting from the position $\boldsymbol{i}$ of the input sentence.

$$s_i = \left[w_i, w_{i+1} \cdots, w_{i+k-1}\right] \tag{5}$$

The input sentence is processed by the filter of the convolution layer to generate a new feature map $c \in R^{L-K+1}$. The conversion formula is as follows:

$$c_i = ReLU\left(s_i^{\circ} f + \theta\right) \tag{6}$$

Where θ represents the offset and the ReLU function is a non-linear activation function. Use N filters of the same length to generate the feature matrix:

$$F = \begin{bmatrix} c_{11} \cdots c_{1i} \\ \cdots\cdots\cdots \\ c_{n1} \cdots c_{ni} \end{bmatrix} \tag{7}$$

Attention Layer. With the development of deep learning in recent years, the attention mechanism has been widely used in image caption generation, machine translation, speech recognition and other fields, and has achieved outstanding achievements [17]. The Attention mechanism simulates the attention model of the human brain. For example, when we admire the painting, we can see the whole picture, but when we look closely, the glasses focus on only a small part. The main concern lies on this small pattern, that is, when there is a certain weight differentiation, the human brain's attention to the entire painting is not balanced. The Attention mechanism is a mechanism that highlights local important information by assigning sufficient attention to key information [18]. In the model proposed in this paper, each output element is obtained by clicking the formula:

$$y_i = F(C_i, y_1, y_2 \cdots y_{i-1}) \tag{8}$$

$$C_i = \sum\nolimits_{j-1}^{T} a_{ij} S\left(x_j\right) \tag{9}$$

$$a_{ij} = \frac{\exp\left(e_{ij}\right)}{\sum_{k-1}^{T} \exp(e_{ik})} \tag{10}$$

$$e_{ij} = score\left(s_{i-1}, h_j\right) = vtanh\left(Wh_j + Us_{i-1} + b\right) \tag{11}$$

Among them, C_i is an input sentence $x_1, x_2 \ldots$ which is obtained by non-linear transformation after the convolution layer operation, represents the corresponding hidden state of the j word embedding, T represents the number of input sequence elements, and a_{ij} represents the input j corresponding to the output Y_i attention distribution probability. Among them, e_{ij} is a verification model, which aims to reflect the influence evaluation score of the j input on the i output, h_j is the hidden state of the j input in the convolution layer, W and U are weight matrices, and b is Offset. The semantic coding formed by the Attention mechanism will be used as the input of the MGU module.

MGU Layer. After the Attention mechanism, this paper uses a simplified recurrent neural network structural unit, namely the smallest gate unit MGU. Compared with LSTM and GRU, MGU has fewer parameters [19]. At time t, the MGU model calculates the current state as:

$$h_t = (1 - f_t) \odot c_{t-1} + f_t \odot \tilde{h}_t \tag{12}$$

The forget gate controls the degree of memory forgetting at the last moment and how much new information is added. The forget gate is expressed as:

$$f_t = \sigma\left(W_f + U_f h_{t-1} + b_f\right) \tag{13}$$

$$\tilde{h}_t = tanh(W_h x_t + f_t \odot (U_t h_{t-1}) + b_h) \tag{14}$$

Compared with LSTM and GRU, MGU has fewer parameters, simpler structure, and faster convergence. Through the above model, a deep-level feature representation of the text can be obtained. Finally, Softmax is used to perform regression to obtain the final text classification result.

4 Experimental Analysis

4.1 Data Set

For the sentiment classification task, the publicly used IMDB and Sentiment140 datasets are selected in this paper. The IMDB film review dataset is a binary sentiment classification dataset, with positive and negative reviews each accounting for 50%. The Sentiment140 dataset is also a dataset that can be used for sentiment analysis. It is composed of 160,000 tweets, and emoticons have been removed. In this paper, 20,000 pieces of data are selected according to the principle of positive and negative balance, and divided into a training set and a test set according to a ratio of 8:2.

4.2 Experimental Settings

This article uses the above two data sets for pre-training. After word segmentation and processing of the above corpus, the word vector is trained using Word2vec. The dimension of the word vector is 100, and the vector representation of the text word can be obtained. If a word that is not in the word vector vocabulary appears when the model is running, random initialization is used. Since the convolutional layer module requires fixed-length input, this article defines a Maxsize to indicate the maximum length of the allowed text sentence. Through the threshold Maxsize to limit the text length, a fixed-length input text sentence can be obtained. In addition, this article Set the region sequence length to 5 and the number of filters to 2.

For the C_MGU model based on the attention mechanism, this paper defines the context vector in the attention mechanism as 100 dimensions, the number of MGU hidden units is 100, the learning rate is 0.01, and the batch-size is 20. The hidden layer state of the MGU is used as the input of the Softmax. The interval steps of the classifier are set to 100 and the batch-size is set to 20. The stochastic gradient descent method is used as the optimization method.

This experiment uses the Keras deep learning framework. The bottom layer is TensorFlow. The TensorFlow platform integrates CNN, RNN and LSTM, GRU, MGU and other deep learning models. It is implemented using Python programming.

4.3 Analysis of Experimental Results

Parameter Setting Comparison Experiment. The word vector is a more important text processing task. Its dimension directly affects the classification accuracy of the model. In this paper, Word2Vec is used to train the word vector. The text is given an initial value so that the network can better learn and adjust parameters during the training process. To make the model have better classification performance. Therefore, the choice of the word vector dimension is particularly important. The Att-C_MGU and CNN, LSTM, GRU and other models proposed in this paper perform on the data set IMBD when the word vector dimensions are different, thereby determining the choice of the word vector dimension in this paper problem. The specific impact of different word vector dimensions on the model's classification effect on the IMBD dataset is shown in Fig. 5.

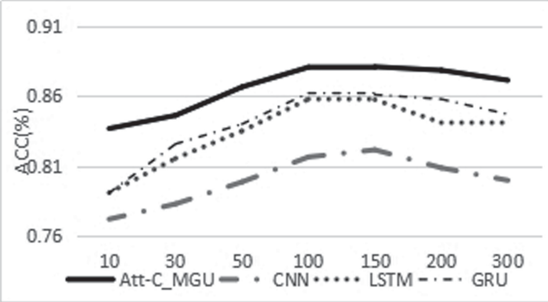

Fig. 5. Influence of word vector dimensions

It can be seen from Fig. 5 that when Word2Vec is used to train word vectors, when the word vector dimension is less than 100, the accuracy of the selected model has a significant upward trend. When the word vector dimension is greater than 100, the selected model is in the data The classification performance of the set shows fluctuations, and some rises more gently, but as the dimension of the word vector increases, the model does not learn the feature information of the word vector well, but it decreases, so this article The dimension of the word vector is selected as 100.

After determining the size of the word vector, the choice of model iterations in this article is also explained through experiments. The number of iterations is the number of times the entire training set is trained. As the number of iterations increases, the results of the network model gradually approach the optimal, When the number of iterations exceeds a certain number of times, it will lead to overfitting and reduce the generalization ability of the model.

It can be seen from Fig. 6 that with the increase of the number of iterations, the accuracy of the model will gradually increase. When the number of iterations is about 30, the accuracy will stabilize, the change will be relatively gentle, and the performance effect will be better.

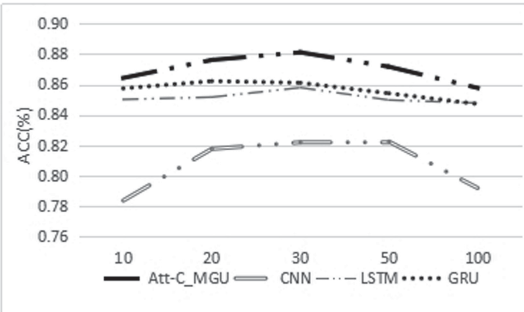

Fig. 6. Iteration times

Attention Mechanism Affects Experiments. In the experiments in this paper, after pre-processing the text data, jieba word segmentation is used to process the text sentence, then Word2vec is used to train the word vector, and the trained word vector is used as the model input. After the attention C_MGU model proposed in this paper, Finally, the output results are compared with the labeled data to show the effectiveness of the model. In order to illustrate the necessity of the attention mechanism and the role of the attention mechanism, the performance of the attention C_MGU model and the C_MGU without the attention mechanism were compared on the IMBD and Sentiment140 datasets. The experimental results are shown in Fig. 7. As shown. As can be seen from Fig. 7, due to the introduction of the attention mechanism in the sentiment analysis algorithm, the performance of the algorithm has been improved to a certain extent, and the accuracy of the IMBD, IMBD2, and Sentiment140 dataset classification has been improved by 2.8%, 2.3%, 1.7%. This illustrates the necessity of the Attention mechanism for the new model.

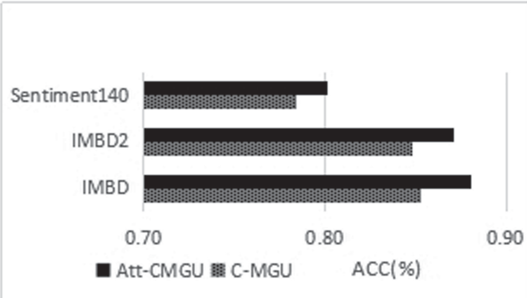

Fig. 7. Impact of attention mechanism

5 Conclusions

This paper proposes a C_MGU-based hybrid neural network model. By performing sentiment analysis experiments on IMBD film review data and Sentiment140 data, it compares accuracy and F1 values with other mainstream text classification methods. Effective. Compared with GRU and LSTM, MGU has certain advantages in the field of text sentiment analysis due to its advantages such as small calculation amount and fast convergence speed.

In the next research work, we will carry out research in the following aspects: In the research of this paper, there are only two types of emotional polarity: positive and negative, the classification of emotion is not refined, and the issue of emotional intensity is not considered; During data processing, some data with only emojis or only pictures and no text descriptions were discarded, which could not be comprehensively analyzed from multiple angles, which caused some limitations to the model. The next step will be in-depth research in sentiment analysis of multimodal text.

References

1. Hu, A., Flaxman, S.: Multimodal sentiment analysis to explore the structure of emotions. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, United Kingdom, 19–23 August 2018, pp. 350-358. ACM Press (2018)
2. Chen, L., Guan, Z.Y., He, J.H., et al.: Research progress of sentiment classification. Comput. Res. Dev. **54**(6), 1150–1170 (2017)
3. Chen, L., Guan, Z.Y., He, J.H., et al.: A survey on sentiment classification. J. Comput. Res. Dev. **54**(6), 1150–1170 (2017)
4. Li, S., Huang, C.R., Zhou, G.: Employing personal impersonal views in supervised and semisupervised sentiment classification. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, 11–16 July 2010, pp. 414–423. ACL, Stroudsburg (2010)
5. Zhang, Y., Fu, J., She, D., Zhang, Y., Wang, S., Yang, J.: Text emotion distribution learning via multi-task convolutional neural network. In: IJCAI (2018)
6. Zhang, S., Wei, Z., Wang, Y., et al.: Sentiment analysis of Chinese micro-blog text based on extended sentiment dictionary. Future Gener. Comput. Syst. **81**, 395–403 (2017). S0167739X17307835
7. Wu, L., Morstatter, F., Liu, H.: SlangSD: building and using a sentiment dictionary of slang words for short-text sentiment classification. Lang. Resour. Eval. **52**(6), 1–14 (2016)
8. Kim, Y.: Convolutional neural networks for sentence classification. Eprint Arxiv arxiv.org/abs/1408.5882 (2014)
9. Zhuang, Y., Zhao, Z., Lu, H., et al.: Microblog sentiment classification via recurrent random walk network learning. In: Twenty-Sixth International Joint Conference on Artificial Intelligence (2017)
10. Bai, J., Li, Y., Ji, D.H.: Attention-based BiLSTM-CNN Chinese Weibo position detection model. J. Comput. Appl. Softw. **35**(3), 266–274 (2018)
11. Hu, R.L., Rui, L., Qi, X., et al.: Text sentiment analysis based on recurrent neural network and attention model. Appl. Res. Comput. (11) (2019)
12. Chen, K., Liang, B., Ke, W.D.L., et al.: Sentiment analysis of Chinese Weibo based on multi-channel convolutional neural network. J. Comput. Res. Dev. **55**(05), 55–67 (2018)
13. Mikolov, T., Sutskever, I., Chen, K., et al.: Distributed representations of words and phrases and their compositionality. In: Proceedings of the 27th Annual Conference on Neural Information Processing Systems. Cambridge, Nevada, 5–10 December 2013, pp. 3111–3119. MIT Press, Cambridge (2013)
14. Cleeremans, A., Servan-Schreiber, D., Mcclelland, J.L.: Finite state automata and simple recurrent networks. Neural Comput. **1**(3), 372–381 (1989)
15. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
16. Cho, K., Merrienboer, B.V., Gulcehre, C., et al.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: Proceedings of the Empirical Methods in Natural Language Processing, Doha, 25–29 October 2014, pp. 1724–1735. ACL, Stroudsburg (2014)
17. Wang, C., Jiang, F., Yang, H.: A hybrid framework for text modeling with convolutional RNN. In: The 23rd ACM SIGKDD International Conference. ACM (2017)
18. Seo, S., Huang, J., Yang, H., et al.: Interpretable convolutional neural networks with dual local and global attention for review rating prediction. In: The Eleventh ACM Conference. ACM (2017)
19. Zhou, G.B., Wu, J., Zhang, C.L., et al.: Minimal gated unit for recurrent neural networks. Int. J. Autom. Comput. **13**(3), 226–234 (2016)