ISSN 2616-5775 Vol. 4, Issue 4: 21-29, DOI: 10.25236/AJCIS.2021.040404

Sentiment Analysis of Film Reviews Based on BI-GRU+Attention+Capsule Fusion

Zhifei Hu[1,2,a]

[1]*SAIC Commercial Vehicle Technology Center, Shanghai, China*
[2]*Renmin University of China, Beijing, China*
[a]*huzhifei@saicmotor.com*

Abstract: *In this paper, we introduce a novel model called BGAC for text semantic analysis task. The proposed network is a new integration of two Bi-GRU, self-Attention mechanism and Capsule network architecture. In our experiment on the task of sentiment analysis in dataset IMDB (a public film review dataset), our model achieve the state-of-the-art results. We compare it with six other deep learning models, such as LSTM, CNN, GRU, BI-GRU, CNN+GRU and GRU+CNN. The results of the experiments show that the experimental effect of the bidirectional GRU fusion self-attention mechanism and the capsule network outperforms than the other six neural network models. In addition, the experiments show that combination of GRU with CNN is better than that combination of CNN and GRU, and the combination of CNN with GRU performs better than employ CNN model alone. The accuracy of using single CNN is successively higher than that of LSTM, BI-GRU and GRU model. Our model which the combination of the BI-GRU, Attention and Capsule network introduced in this paper achieves the highest accuracy, precision and F1 score. In conclusion, the bidirectional GRU with self-attention mechanism and capsule network model significantly improves the accuracy of text sentiment classification task.*

Keywords: *Sentiment analysis; Film review; Capsule network; Self-attentional mechanism; Neural network*

1. Introduction

With the rapid development of the Internet, the fast and convenient ways to purchase online products, and the diversification of types, have made people increasingly rely on online purchases. After consumers buy goods online, they will leave a comment on the goods in the network backend. Most of the comments in this part of the content indicate the consumer's emotional inclination towards the product. When consumers are satisfied with the purchased product, the product reviews are often inclined to be positive, and vice versa. Product reviews tend to be negative [4]. This part of the existing consumer reviews on the product will affect the willingness of other consumers to buy the product. Therefore, it is very important to study the emotional tendency of consumers to product reviews. According to different reviews, consumers' attitudes towards products can be obtained, and the advantages or disadvantages of the products can be judged in a short time. Consumers can quickly pick out their ideal products, and merchants can also use the advantages and disadvantages of the products in the reviews. Further improve the advantages and disadvantages of commodities, adjust sales strategies, and maximize the purchase rate of commodities [13].

Text sentiment analysis or opinion mining is the use of text mining related techniques to determine people's opinions, emotions, and assess their emotional preferences towards entities such as products, services, and organizations. The main research object of sentiment analysis is the mass text of commodity comments, blogs, microblogs and public opinions of various forum posts in the Internet [7]. The development and rapid start of this field is due to the rapid development of social media on the Internet, such as product reviews, forum discussions, Weibo, WeChat, because this is the first time in human history that such a large number of digital forms have been recorded. Since the beginning of 2000, text sentiment analysis has grown into one of the most active research fields in natural language processing (NLP) [16]. In fact, it has spread from computer science to management science and social science, such as marketing, finance, political science, communications, medical science, and even history, attracting the common attention of the whole society due to its important commercial nature. This proliferation is due to the fact that opinion is central to almost all human activity, to a considerable extent, very much concerned with what others think. For this reason, whenever we need to make a decision, we often seek

ISSN 2616-5775 Vol. 4, Issue 4: 21-29, DOI: 10.25236/AJCIS.2021.040404

other people's opinions. This is not only true for businesses but also for individuals. In recent years, with the wide application of deep learning technology in many fields, text mining using neural networks has become one of the research hotspots in the field of natural language processing [5,6]. Among them, various neural network algorithms are applied to text sentiment analysis [9].

The most basic task of text sentiment analysis is to extract valuable information units from the text of sentiment comments. Sentiment information extraction is to extract the words or phrase elements that contribute to sentiment analysis from the data of text sentiment comments. To some extent, information extraction plays an important role in feature dimension reduction and system performance improvement. Common statistical analysis methods include information gain based, mutual information, expected entropy, word frequency, document frequency, etc. The extraction of emotional information is mainly divided into the extraction and discrimination of evaluation words, the extraction of evaluation objects and the extraction of opinion holders. The extraction and judgment of the evaluation words mainly realize the recognition, polarity and metric judgment of the evaluation words. This part is usually a systematic work, which can be divided into corpus-based and dictionary-based methods. The extraction and discrimination of evaluative words based on corpus are mainly based on statistical characteristics. By comparing and observing large corpora, evaluative words in corpus are dug out and polarity is judged. The advantage is that it is simple and easy to operate, while the disadvantage is that the comment corpus is limited, and the distribution of the comment words in the large corpus is not easy to conclude. The lexicographical method of extracting and discriminating evaluative words mainly utilizes the semantic connections between the words in the dictionary to realize the mining of evaluative words. However, the biggest difficulty of this method is that the degree of dictionary update determines the semantic analysis. The extraction of evaluation object is mainly to extract the subject involved in the comment, which specifically refers to the object modified by the evaluation words in the comment text. Among them, the most traditional method to extract evaluation objects is the rule-based method, and the formulation of rules generally requires a series of language analysis and preprocessing, such as part of speech tagging, named entity recognition, and syntactic analysis. The extraction of opinion holders mainly completes the identification of who is the opinion subject of emotional text.

Affective computing is an emerging field of interdisciplinary research, including but not limited to computer science, psychology, cognitive and social sciences, natural language processing and other fields. In recent years, this field has gradually attracted more and more academic and industrial attention. Affective computing makes intelligent systems automatically identify emotions through the principles of recognition, perception, reasoning and interpretation. Bengio et al. [10] is one of the first groups to use neural networks to build language models. Lin Shiping [11] proposed a text representation method fused with knowledge graphs. This method uses a two-way long and short-term memory LSTM network to learn the contextual relationship between words. Anwar et al. [15] proposed a hybrid CNN-LSTM sentiment analysis model. A good combination of CNN can extract local features, and LSTM can capture the advantage of long-term dependencies between word sequences. Cheng Yan et al. [2] proposed a neural network model based on the attention mechanism of multi-channel CNN and bidirectional gated recurrent unit. This model can not only focus on the words that are important for emotional polarity classification in the sentence through the attention mechanism, but also combines the advantages of CNN to extract local features of text and Bi-GRU network to extract long text context semantic information, which improves the model's text feature extraction ability. Sun Min et al. [1] compared the sentiment classification effects of several models of CNN, LSTM, and CNN-LSTM based on the IMDB data set, and showed that the accuracy of the CNN-LSTM fusion model is significantly improved compared to the separate CNN and LSTM models. Tensor et al. [3] implemented the Bi-LSTM+Attention sentiment analysis model, analyzed its sentiment tendency on the IMDB movie review data set, and used it with three deep learning algorithms such as Bi-LSTM, Bi-GRU, and MLP. The accuracy and recall rate under the application scenario were compared. Experimental results show that the accuracy of the Bi-LSTM model is slightly higher than that of the base GRU model. After using the Attention mechanism at the same time, the accuracy of each model has been improved. Tao Chen et al. [17] proposed a sequence model based on neural network, which divided sentences into three types according to the number of targets appearing in the sentence. Then each set of sentences is input into a one-dimensional convolutional neural network for sentiment classification. This method shows that sentence type classification can improve the performance of sentence-level sentiment analysis. The proposed approach achieves state-of-the-art results on several benchmarking datasets. Pang et al. [18] tried different machine learning algorithms on a film review data set, and finally achieved an accuracy of 82.9% on a large number of texts through analysis and comparison of characters. Socher et al. [19] tried a novel method, namely recursive neural tensor network (RNTN), on the same data set, and the experimental results showed that RNTN achieved better accuracy (85%). Yu and Hatzivassiloglou[20] model sentence-level

ISSN 2616-5775 Vol. 4, Issue 4: 21-29, DOI: 10.25236/AJCIS.2021.040404

features from a semantic perspective to judge the emotional orientation of the text. Melville et al. [21] developed a framework to analyze emotions from the perspective of relational information of part of speech. Turney et al. [22] propose an unsupervised knowledge-based approach to sentiment analysis that uses seed words to calculate the polarity and semantic direction of phrases. Hu et al. [23] applied mathematical model to emotion recognition, and this data model was obtained by extracting emotional clues from blogs. Gangemi et al. [24] proposed a method for identifying opinion holders and subjects using an unsupervised framework.

2. Model Principle

In this chapter, we formulate sentiment analysis of film reviews as a text classification task which is defined as follows: given a film review set S ={S_1, S_2, S_3... , S_n}, for each S_i and its label y_i, y_i ∈Y, we need to automated learn function f that transform film review s_i to label y_i when given train dataset S and label Y.

In this paper, we propose a new deep neural network called BI-GRU Attention Capsule neural network (BGACN) which explicitly models the high-order text features in an end-to-end fashion. Experiments on public dataset show that BGAC can effectively capture text context information and extract effective semantic information, significantly outperforms state-of-the-art methods like LSTM, CNN. Figure 1 shows the model framework, which consists of five main components:

1) Word embedding layer: Word embedding layer is a word type representation, each word into a vector, the purpose is to facilitate the calculation. Word embedding layer can be divided into word vector and word part of speech vector. The part of speech information of words can provide important syntactic information for the model.

2) Bi-GRU layer: We employ bidirectional unit gates to capture text context dependency information. Because of its special gate structure, gate network has a certain memory ability, especially suitable for long text modeling. Because the effective information in text may be separated by a certain distance, general filter can only observe the local information, can not grasp the overall architecture characteristics.

3) Capsule layer: Hinton proposed that the capsule network was used in the image field at the beginning. It was mainly used to solve the problem that the general convolution network could not grasp the spatial location information of features. Similarly, in the text field, the location information was equally important. We used the dynamic routing algorithm of the capsule network to retain the location information and local spatial information of the text for the final prediction of the model.

4) Self-Attention layer: Bi-directional GRY focuses on capturing the dependency relationship between different features, which is not enough to highlight the capture of key information of the whole task. Therefore, we introduce attention mechanism, use hidden layer state to perform self attention operation, strengthen the extraction of key word features, and further extract features of great significance for classification.

5) Pooling layer: after self attention mechanism, we add pooling to reduce the number of invalid features.

6) output layer: the pooled attention mechanism layer and the output of the capsule network are spliced together as the fully connected input to get the final sentiment classification output. The whole model structure is as follows:

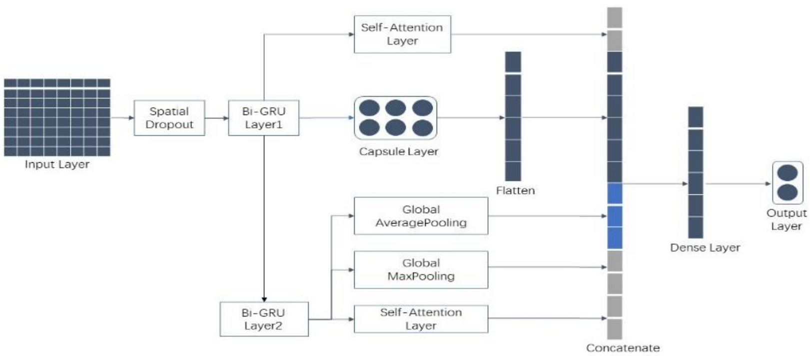

Figure 1 Bi-GRU+Attention+Capsule model diagram

ISSN 2616-5775 Vol. 4, Issue 4: 21-29, DOI: 10.25236/AJCIS.2021.040404

Figure 1 above is the model diagram of Bi-GRU+Attention+Capsule proposed in this paper.

2.1 Word Embedding Layer

First, do some preprocessing of the text. The length of all text data is processed to ensure that the length of all text is consistent. If the text is too long or too short, it will be clipped or filled. Secondly, a dictionary of text corresponding to the index is automatically mapped through the automatic function interface of the relevant module of NLP, and then the words are serialized and the text is converted into numbers. Finally, the trained Glove vector was represented by the global word vector, which was automatically initialized to form an embedding layer with rich semantics. Converts the text to a dense vector using an Embedding Lookup. In this article, the embedding dim size is set to 100.

2.2 Bi-directional GRU Layer

GRU is a new generation of recurrent neural network, very similar to LSTM, and a very effective variant of LSTM network [12]. Compared with LSTM, the structure of the LSTM network is simpler and easier to train, which can greatly improve training efficiency. And the effect is also very good, there are fewer tensor operations, so it is also a very manifold network at present, and it is more inclined to use GRU in many cases. GRU removes the cell shape State, use hidden state to transfer information. It contains only two doors: update door and reset door. Intuitively speaking, the function of the update gate is similar to the forget gate and input gate in LSTM, which determines which information to forget and which new information needs to be added. The reset gate determines how to combine the new input information with the previous memory, and is used to determine the degree of forgetting the previous information [14].

The structure of the GRU model is shown in Figure 2:

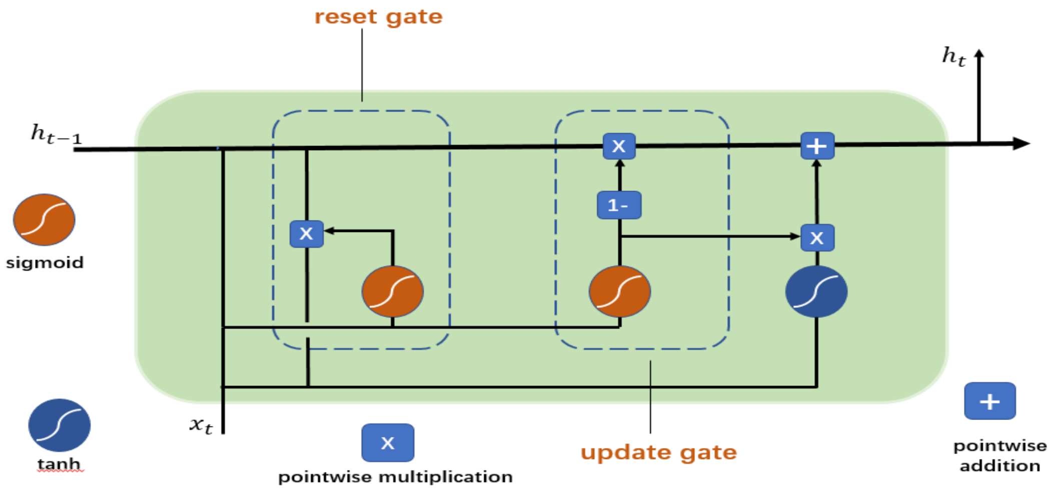

Figure 2 GRU structure diagram

The formula derivation of GRU is as follows:

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

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \tag{2}$$

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

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

If the reset gate is set to 1 and the update gate is set to 0, the standard RNN model will be obtained again. The input and forget gate in LSTM correspond to the update gate of GRU, and the reset gate directly acts on the previous hidden state. No second-order nonlinearity is applied when calculating the output. This article uses a bidirectional GPU as an overlay. The two-way GRU makes it possible to obtain both the previous information and the subsequent information at a certain point in the sequence.

The two-way GRU is based on the one-way GRU, which processes the input sequence forward and backward in turn, and splices the output into the final output layer. In this way, the output node of each time step contains the complete past and future context information of the current moment in the input

ISSN 2616-5775 Vol. 4, Issue 4: 21-29, DOI: 10.25236/AJCIS.2021.040404

sequence.

2.3 Self-attention Mechanism Layer

The attention mechanism is a kind of attention resource allocation mechanism similar to the human brain. By means of probability weight distribution, the probability weight of word vectors at different moments is calculated, so that some words can get more attention [11], thereby improving the hidden layer The quality of feature extraction. The traditional attention mechanism can only express the semantics of a sentence based on a certain aspect, which leads to the loss of part of the semantic information. The attention algorithm formula is as follows:

$$Attention(Q,K,V) = \mathrm{softmax}(\frac{QK^T}{\sqrt{d_k}})\mathrm{V} \tag{5}$$

Where $Attentio(Q, K, V)$ is the value of the obtained attention. Q, K, and V are the query vector matrix, key vector matrix, and value vector matrix. Each row of these three matrices represents a corresponding vector. And $Q \in Rn{\times}dk$, $K \in Rm{\times}dk$, $V \in Rm{\times}dv$. This is the traditional proportional dot product attention mechanism.

In actual applications, Q, K, and V are different in different scenarios. When K=V=Q, it is the self-attention mechanism. The self-attention mechanism is an improvement of the attention mechanism. By doing Attention in the sequence itself, looking for the internal connections of the sequence, it reduces the dependence on external information and is better at capturing the internal correlation of features. The self-attention model can establish the long-distance dependence relationship within the sequence, and can directly connect any two words in the sentence through a calculation. It is easier to capture the semantic features such as the word dependence in the sentence and the internal structure of the sentence, and it can also increase the calculation. Parallelism. But models based on self-attention mechanisms are not good at capturing local dependencies in text. The self-attention model can be used as a layer of the neural network, can also be used to replace the convolutional layer or the recurrent layer, and can also be used in cross-stacking with the convolutional layer or the recurrent layer. This article uses the self-attention model as a layer.

2.4 Capsule Network Layer

The capsule network is composed of individual capsules. Each capsule is composed of a group of neurons and expressed in the form of vectors, which can be called tensor neurons. The pooling operation of the traditional convolutional neural network in the forward propagation process will lose a lot of semantic information, it is impossible to understand the object from a new perspective, and it is difficult to identify the precise spatial relationship. Convolutional neural networks train neurons to detect different patterns, even from different angles of the same pattern, making the number of convolution kernels and layers more and more. The capsule network hopes to recognize the same type of pattern through a capsule. The length of the output vector through the capsule represents the probability estimate of the existence of the target, and the direction of the vector represents the attribute of the entity. The capsule network replaces the output of a single neuron in CNN with vector output, and replaces the pooling layer of CNN with dynamic routing. Dynamic routing refers to the realization of segmentation of highly overlapping objects. The author proves that the dynamic routing mechanism is an effective way. The capsule network can retain the location information and local spatial features of the feature data through the feature transfer method of dynamic routing [8].

The network structure of the capsule network is similar to that of the convolutional neural network. The author uses the length of the capsule output vector to represent the probability that the entity represented by a capsule appears in the input. Therefore, the author uses a nonlinear function to "compress" the vector. Short vectors are compressed to almost zero, and long vectors are compressed to a length less than 1. Make full use of this nonlinear function in discriminative learning.

$$V_j = \mathrm{Squash}(s_j) = \frac{\|s_j\|^2}{1+\|s\|^2}\frac{s_j}{\|s_j\|} \tag{6}$$

Among them, Vj is the output vector of capsule j, and sj all the inputs of capsule j. The Squash function is a new activation function specially designed by the z author for the capsule network. The vector is compressed to obtain the vector modulus length, and the modulus length of the vector represents the feature strength. Except for the first-layer capsule, all inputs of the capsule sj are the weighted summation of the prediction vector $uj|i$. These prediction vectors are all generated by the lower layer of

ISSN 2616-5775 Vol. 4, Issue 4: 21-29, DOI: 10.25236/AJCIS.2021.040404

the capsule, which is multiplied by the output of the capsule ui and a weight matrix Wij, and the impact of the perspective is modeled by matrix multiplication.

$$s_j = \sum_i c_{ij}\hat{u}_{j|i} \tag{7}$$

$$\hat{u}_{j|i} = W_{ij}u_i \tag{8}$$

Among them, cij is the weight determined by the iterative dynamic path process. The sum of the weight coefficients of capsule i and all capsules in the layer above it is 1.

The calculation formula of cij is as follows:

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \tag{9}$$

Among them, the initial logical value bij is the log prior probability that capsule i is coupled to capsule j. This log prior can be discriminatively learned along with other weights. They are determined by the position and type of the two capsules, not by the current input.

The forward propagation process of vector neurons:

1) Multiply the input vector u and the matrix W to obtain a new input vector U;

2) Multiply the input vector U by the weight C;

3) Sum the weighted input vector into a vector S;

4) Convert the vector S into a vector V using a non-linear function.

This paper mainly uses the strong text feature learning ability of the capsule network, the ability to retain text space information and the higher-speed training ability to perform text sentiment classification tasks. The results prove that the effect of the capsule network on multi-label text classification tasks has been significantly improved.

2.5 Pooling Layer & Fully Connected Layer

Finally, we put the pooled attention mechanism layer together with the output of the capsule network as the fully connected input to get the final emotion classification output.

$$\tilde{y} = concat(y_1||y_2||y_3||y_4||y_5) \tag{10}$$

$$y_o = \sigma(w_o\tilde{y} + b_o) \tag{11}$$

y_1,y_2, y_3,y_4, y_5 are the output of capsule network, global max pooling and global average pooling layer and two self-attention layers.$w_o$and b_o are the weight and bias of output layer and σ represents softmax function which can normalize the output probability.

We calculate the cross entropy loss between the final output of the model and the label. And we employ Adam optimizer to train the model. In order to prevent over fitting, L2 regularization of all training weights is added to the loss function to limit the weights.

$$\text{loss} = -\sum_{i=1}^{n} p(x_i)\, log\, q\,(x_i) + \lambda||f||_2^2 \tag{12}$$

where $p(x_i)$ is true label of dataset and $q(x_i)$ is the prediction of model. The second term is the L_2 regularization. λ is hyperparameters in our experiment.

3. Experimentations

3.1 Experimental Data and Experimental Environment

The experimental data in this article is the public movie review sentiment classification data set IMDB downloaded on kaggle. The total number of IMDB datasets is 50,000. Among them, the number of training sets of this model is 32,000, the number of verification sets is 8,000, and the number of test sets is 10,000. Each piece of data includes two columns of features: text and classification labels. The text data is movie reviews, and the classification label data is positive and negative emotion labels.

All experimental environments and specific configurations in this article are shown in Table 1 below.

ISSN 2616-5775 Vol. 4, Issue 4: 21-29, DOI: 10.25236/AJCIS.2021.040404

Use the deep learning framework keras, the backend uses Tensorflow, and the programming language uses python.

Table 1 Experimental environment configuration

Name of software	Software type	Software version
Window 10	Operating system	Windows 10 Home
Tensorflow	Deep learning framework	2.5.0
Keras	Deep learning framework	2.4.3
Python	Programming language	3.6
Numpy	Python scientific computing library	1.19.5
Pandas	Python data structure library	1.2.3
Sklearn	Python machine learning library	0.0
GPU	Processor	—
Jupyter Notebook	Tool	—
NLTK	Natural language processing tool	3.2.4
Seaborn	Python visualization library	0.8.1
Matplotlib	Python visualization library	2.1.0

3.2 Feature Extraction and Data Processing

Keras deep learning framework is used to implement the model experiment in this paper. In the part of word embedding, the Glove method is used for vectorization representation of words [20]. The word vector dimension is set to 100 dimensions. The number of capsules is set as 10, the dimension of capsules is set as 10, the number of dynamic routing iterations is set as 4, and the optimizer is set as Adam. In the process of model training, the learning rate, batch size and epochs were set as 0.2, 32 and 30 respectively. In order to prevent overfitting, the dropout mechanism parameter was set to 0.2 during the training process.

3.3 Model Structure and Parameter Settings

Table 2 below shows the parameter settings of all our model proposed in this paper.

Table 2 Parameter Settings

Parameter name	Parameter value
Word vector dimension	100
Learning rate	0.2
Batch size	128
Epochs	30
Maximum number of features in text	55395
Sentence length	55
Number of capsules	10
Capsule dimension number	10
Number of dynamic routes	4

3.4 Experimental Results and Analysis

In order to verify the effectiveness of the method proposed in this paper, multiple sets of comparative experiments are set up, including single network, hybrid network and network comparison with the introduction of attention mechanism. Specific experimental models include LSTM, CNN, GRU, Bi-GRU, CNN+GRU, GRU+CNN, and the newly proposed Bi-GRU+Attention+Capsule fusion model. The experimental comparison results of IMDB data set are shown in Table 3 below.

Table 3 Experimental results on IMDB dataset

Model	Accuracy	Precision	Recall	F1-score
GRU	0.8470	0. 8487	0. 8487	0. 8487
LSTM	0.8552	0.8563	0.8582	0.8572
Bi-GRU	0.8551	0.8577	0.8550	0.8563
CNN	0.9254	0.9255	0.9255	0.9255
CNN+GRU	0.9529	0.9514	0.9514	0.9514
GRU+CNN	0.9596	0.9598	0.9598	0.9598
Bi-GRU+Attention+Capsule	0.9687	0.9693	0.9693	0.9693

As shown in Table 3 above, the best experimental effect of this article on the IMDB data set is the Bi-GRU fusion capsule network And introduce the combined model of self-attention mechanism, followed by GRU+CNN, CNN+GRU, CNN, LSTM, Bi-GRU, GRU and other models. It can be seen from the experiments that the task of emotion recognition is often more concerned with specific words and phrases, so it is not particularly prominent for those networks that are more inclined to the global information modeling, like GRU, LSTM, Bi-GRU. Those networks that focus on local features will get excellent results. For example, only using convolution network, we still achieved 92% of F1! Compared with the mixture of recurrent neural network and convolutional neural network, our capsule network and self-attention mechanism can achieve better results. This is because the attention mechanism can better capture those local features which are particularly important for classification, and the capsule network can capture the location information, which provides a great help for the prediction of the model.

4. Conclusion

In this paper, a neural network model combining self-attention mechanism, capsule network and bidirectional gating loop unit is proposed. And it is applied to IMDB film review sentiment analysis task. The model uses the local feature extraction ability of capsule network, the ability of bidirectional gating loop unit network to capture bidirectional semantic dependence, and finally the self-attention mechanism to effectively identify the implicit emotional features in the text. Compared with the other six models, the experimental results show that the method proposed in this paper has achieved the best results in accuracy, accuracy, recall rate, F1 value and other indicators, which verifies that this model has better generalization ability, robustness and effectiveness. In the subsequent work, the performance of the model on other data sets will be explored to enhance the mobility of the model. In addition, it can be considered to try to integrate other neural network algorithms in different ways, as well as more modal feature extraction, such as the fusion of visual and speech emotion recognition, and the study of multi-modal emotion recognition. So as to further improve the accuracy of emotion classification task.

References

[1] Sun Min, Li Yang, et al. Sentiment analysis based on CNN-LSTM movie reviews[J]. Journal of Luoyang Institute of Technology (Natural Science Edition), 2019, 4(29): 71-77.

[2] Cheng Yan, Yao Leibo, et al. Multi-channel CNN and BiGRU based on attention mechanism text sentiment tendency analysis[J]. Computer Research and Development, 2020, 57(12): 2583-2592.

[3] Liang Zhang, Yuanfeng Yang, et al. Design and Implementation of Bi-LSTM+Attention Sentiment Analysis Model[J]. Database Technology, 2018: 177-179.

[4] MD Devika, etl. Sentiment Analysis: A Comparative Study On Different Approaches[J]. Procedia Computer Science, 2016, 87:44-49.

[5] Arnab Hara, etl. A Comparative Study of Different Classifcation Techniques for Sentiment Analysis. International Journal of Synthetic Emotions, 2020,11(1):49-57.

[6] Zeeshan Shaukat,etl. Sentiment analysis on IMDB using lexicon and neural networks. SN Applied Sciences,2020,2(148).

[7] Doaa Mohey,etl. A survey on sentiment analysis challenges[J]. Engineering Sciences,30:330-338.

[8] Chen Zhuang,etl. Transfer Capsule Network for Aspect Level Sentiment Classification. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistic,547-556.

[9] Fang Biyun. Network movie data analysis based on Keras framework [J].Computer Knowledge and Technology, 2019, 15(34):14-16.

[10] Wang Liya, et al. Text sentiment analysis based on CNN-BiLSTM network introducing attention model. Journal of Wuhan Institute of Technology, 2019, 41(4):386-391.

[11] Lin Shiping, et al. Text sentiment analysis combined with knowledge graphs. Journal of Fuzhou University (Natural Science Edition), 2020, 48(3): 270-275.

[12] Li Songru, et al. Sentiment analysis using recurrent neural network attention Force Model. Journal of Huaqiao University (Natural Science Edition), 2018, 39(2):252-255.

[13] Fan Bo, et al. Optimization and improvement of text sentiment analysis technology based on neural network. Electronic technology and software engineering, 180-182.

[14] Qu Zhaowei, etl. Hierarchical attention network sentiment analysis algorithm based on transfer learning. Computer Applications, 2018, 38(11):3053-3056.

[15] Anwar Ur Rehman, etl. A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis. Multimedia Tools and Applocations, 2019.

[16] H.M.Keerthi Kumar, etl. Sentiment Analysis on IMDb Movie Reviews Using Hybrid Feature

Extraction Method. Regular Issue, 2019, 109-114.
[17] Chen Tao, etl. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. Expert Systems with Applications, 2017, 221-230.
[18] B. Pang, etl. Thumbs up: sentiment classification using machine learning techniques, in: Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, Association for Computational Linguistics, 2002, 79–86.
[19] R. Socher, etl. Recursive deep models for semantic compositionality over a sentiment treebank, in: Proceedings of the conference on empirical methods in natural language processing (EMNLP), 2013, 1631-1642.
[20] H. Yu, etl. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences, in: Proceedings of the 2003 conference on Empirical methods in natural language processing, Association for Computational Linguistics, 2003, 129–136.
[21] P. Melville, etl. Sentiment analysis of blogs by combining lexical knowledge with text classification, in: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2009,1275–1284.
[22] P.D. Turney, Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews, in: Proceedings of the 40th annual meeting on association for computational linguistics, Association for Computational Linguistics, 2002, 417–424.
[23] X. Hu, etl. Unsupervised sentiment analysis with emotional signals, in: Proceedings of the 22nd international conference on World Wide Web, International World Wide Web Conferences Steering Committee, 2013, 607–618.
[24] A. Gangemi, etl. Frame-based detection of opinion holders and topics: a model and a tool, Comput. Intell. Mag. IEEE, 2014, 9 (1):20–30.