EAI Endorsed Transactions
on Scalable Information Systems
Research Article EAI.EU

Hybrid neural network model based on multi-head attention for English text emotion analysis

Ping Li[1,*]

[1]Department of Public Instruction, Nanyang Medical College, Nanyang City 473000, China

Abstract

Traditional Convolutional Neural Network (CNN) ignores the contextual semantics information when performing emotion analysis tasks. And CNN will lose a lot of feature information during the maximum pooling operation, which will limit the text classification performance. CNN cannot extract the emotion features of English text more comprehensively, and relies heavily on a large number of language knowledge and emotion resources. In this paper, we propose a hybrid neural network model based on multi-head attention for English text emotion analysis. Firstly, the new model uses multi-head attention to learn the dependence between words and capture the emotion words in the English text. Secondly, the improved bidirectional gated recurrent unit is used to extract different granularity emotion features of English text. According to each emotion category and attention mechanism, feature vectors are generated to construct the emotion feature vector set. Finally, the text emotion categories are judged according to the model attributes. The model is tested on MR, IMDB and SST-5 data sets, the results show that the proposed method has better classification effect compared with other models.

Keywords: hybrid neural network model, multi-head attention, English text emotion analysis.

Received on 29 October 2021, accepted on 12 November 2021, published on 12 November 2021

doi: 10.4108/eai.12-11-2021.172103

*Corresponding author. Email: snowycry@qq.com

1. Introduction

In recent years, the Internet has evolved from static one-way information carrier to dynamic interactive media. More and more users are posting news or product reviews to express their opinions. Using emotion analysis technology to analyze these massive interactive information, users' emotional and psychological trajectories can be found to help research institutions grasp the dynamics of social emotions [1]. Text emotion analysis is to analyze, process, summarize and judge the emotion tendency of subjective and sexual text information with emotional color [2]. Efficient and rapid analysis of these thoughts and opinions with subjective emotions is the current hot research direction.

Traditional text emotion analysis methods mainly include emotion dictionary-based methods and machine learning-based methods. Although these methods perform well in terms of classification accuracy, they still face many difficulties. The emotion dictionary-based approach, which takes the emotion dictionary as the main basis for judging and commenting on emotion polarity [3], relies on a large amount of manual intervention, such as dictionary construction and judgment rule formulation, etc. It is difficult to cope with the emergence of new and unknown words and has the domain dependence problem [4]. Machine learning method neglects the order of words in the sentence and fails to distinguish the semantics of the

sentence, leading to the error of emotion classification [5]. Taking the bag of words model (BoW) [6] in machine learning as an example, the BoW model represents text as a collection of words, but this collection ignores the syntax and the order of words in the statement, resulting in the failure of the model to capture inter-word and contextual information.

In recent years, the application of deep learning technology in the field of natural language processing (NLP) has become the mainstream of the industry. Compared with traditional methods, convolutional neural network (CNN) and recurrent neural network (RNN) both show advantages in emotion classification tasks. In view of the problem that a large amount of existing emotion information is not fully utilized, more and more researchers [7-9] integrate linguistic knowledge and emotion information into the model. Lin et al., [10] combined word emotion sequence features with CNN to improve classification accuracy. Sun et al., [11] proposed a convolutional neural network model combining word level and word vector. Although these neural network models have achieved great success, they are difficult to extract multi-level and more comprehensive emotion features of text, and rely heavily on text information and emotion resources. Language knowledge [12] (emotion dictionary, negative words, adverbs of degree) needs to be integrated into the model to achieve the best potential in terms of prediction accuracy [13]. With the advent of capsules [14], Wang [15] first attempted to conduct emotion analysis through capsules, which did not require any assistance of language knowledge and had a higher classification accuracy compared to the baseline model. Capsule is rich neural unit. As a vector neuron, it replaces the scalar neuron node in the traditional neural network, changes the structure of connecting scalar in the traditional neural network, and reduces the loss of information.

The main contributions of this paper are as follows:

1) A capsule model combining CNN and Bi-GRU network is proposed for English text emotion analysis task. The model combines the attention mechanism to construct emotion capsules for each emotion category. Vector neurons (capsules) are used for feature representation of text emotion information to enhance model generalization ability and robustness. Compared with models that need to incorporate language knowledge and emotion information, the proposed model is more concise and has higher classification accuracy.
2) The model integrates the advantages of local feature extraction of CNN and the feature of Bi-GRU considering context semantics, which effectively improves the classification performance of the model.
3) Multi-head attention is introduced into the model to capture the emotion words in the text and to encode the dependence between words, so as to improve the feature expression ability of the model.

The structure of this paper is as follows. Section 2 detailed introduces the model. The experiments are conducted in section 3. There is a conclusion in section 4.

2. Proposed emotion analysis model

The deep learning-based approach learns the emotional English semantic features of the commentary English text through self-learning. The construction of the deep learning model in this paper is considered from the following three aspects: 1) Integrating the global semantic structure features of English sentences learned by BiGRU network into the output end of convolutional neural network; 2) An attention mechanism is introduced between the BiGRU output terminal and the convolutional layer and pooling layer of the convolutional neural network model, so as to reduce the interference of noise data; 3) Convolutional neural network and BiGRU learn local and global features of sentences by jointly training.

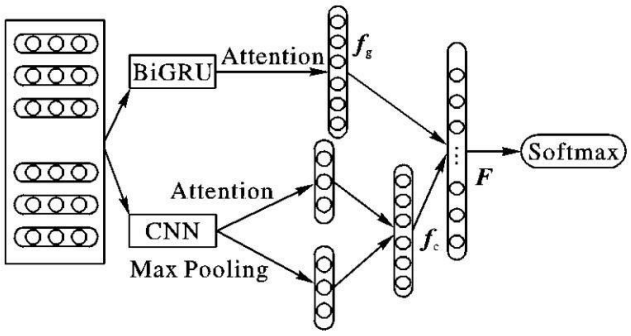

Figure 1. Structure of proposed model

2.1. BiGRU-Attention layer for global feature extraction

The attention mechanism can selectively focus on the important information of the text. In this paper, multi-head attention is used to capture the key information of the text sequence from multiple sub-spaces, as shown in figure 2.

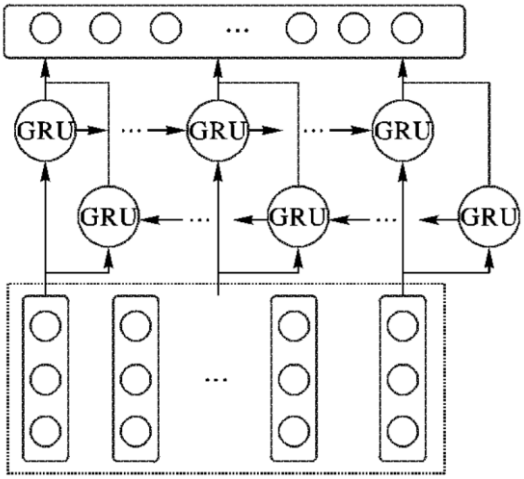

Figure 2. BiGRU-Attention model

For the given English text $S=\{w_1,w_2,\cdots,w_L\}$ with length L, w_i is the i-th word in sentence S, each word is mapped as a D-dimensional vector, i.e. $S\in R^{L\times D}$.

Firstly, the word vector matrix S is linearly transformed and cut into three matrices with the same dimensions, $Q\in R^{L\times D}$, $K\in R^{L\times D}$, $V\in R^{L\times D}$. And they are mapped to different sub-spaces as shown in equation (1).

$$\begin{aligned}[Q_1,\cdots,Q_h]&=[QW^{Q_1},\cdots,QW^{Q_k}]\\ [K_1,\cdots,K_h]&=[KW^{K_1},\cdots,KW^{K_k}]\\ [V_1,\cdots,V_h]&=[VW^{Q_1},\cdots,VW^{V_k}]\end{aligned}\quad(1)$$

Where Q_i, K_i, V_i are the query, key, and value matrices of each subspace. W^{Q_i}, W^{K_i}, W^{V_i} are transformation matrices. h is the number of heads. Then, the attention values of each subspace are calculated in parallel as shown in equation (2).

$$head_i=soft\max(\frac{Q_i,K_i^T}{\sqrt{d}})V_i\quad(2)$$

Where $head_i$ is the attention value of the i-th subspace. $\sqrt{d}$ changes the attention matrix into a standard normal distribution to prevent the gradient from disappearing in the process of back propagation.

Then the attention values of each subspace are spliced and linearly transformed, as shown in equation (3).

$$Multi-head=concat(head_1,\cdots,head_h)W^M\quad(3)$$

Wherein, W^M is the transformation matrix. Multi-head is the attention value of the whole sentence, and concat is the concatenation operation. Finally, the residuals of Multi-head and S are connected to get the sentence matrix, as shown in equation (4).

$$X=residual\_Connect(S,Multi-head)\quad(4)$$

Where $X\in R^{L\times D}$ is the output of multiple attention and $residual\_Connect$ is the residual operation.

2.2. Text local feature extraction combining CNN and Bi-GRU

In order to extract more comprehensive text emotional features, this paper integrates the text features extracted by CNN and Bi-GRU. The emotion features of text are modeled from local to global levels.

A. Text local feature extraction based on CNN

The inspiration of Convolutional Neural Network comes from the research on biological vision mechanism in the field of biology. Its powerful feature learning and feature representation ability has been widely used in natural language processing fields such as text classification and emotion classification. As shown in figure 3, in the text task, traditional CNN takes the word vector formed by the sentence as the input [16-21], and then uses multiple convolution kernels consistent with the dimension of the word vector to carry out the convolution operation and capture the features among multiple consecutive words.

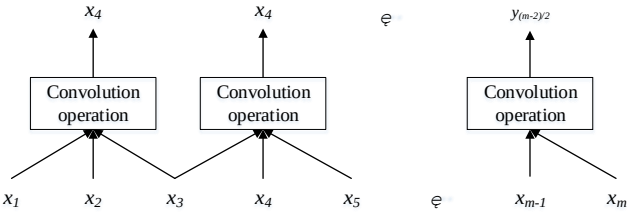

Figure 3. Schematic diagram of convolution operation

In this paper, B convolution filters are selected for local feature extraction of output matrix X in multi-head attention, and the feature matrix $C_i=[C_{i,1},C_{i,2},\cdots,C_{i,B}]\in R^{(L-k+1)\times B}$ is obtained. Where $C_{i,B}=[c_1,c_2,\cdots,c_{L-k+1}]\in R^{L-k+1}$ is the B-th column vector in C_i. The element c_j in this vector can be obtained by equation (5):

$$c_j=f(W\cdot x_{j:j+k-1}+b)\quad(5)$$

Wherein f is the activation function ReLU. $W\in R^{k\times D}$ is the convolution kernel. k is the width of the window, $x_{j:j+k-1}\in R^{k\times D}$ represents the head and tail concatenation of k word vectors. b is the offset term.

In order to extract the local text features of N-gram in the text, the feature vectors extracted from the different window convolution kernel sizes are spliced to form the fusion

feature sequence $C=[C_1,C_2,\cdots,C_n], C\in R^{l\times B}$. Where, $C_n\in R^{(L-k_n+1)\times B}$ is the feature sequence extracted by the convolution kernel with window size of k_n.

B. Text local feature extraction based on Bi-GRU

The traditional machine learning method only considers the limited prefix word information as the condition term of the semantic model, and the recurrent neural network (RNN) has the ability to take all the pre-order words in the language knowledge set into the model. However, the standard RNN has the problem of gradient disappearance or explosion. LSTM network and GRU network can overcome this problem by utilizing the structure of some "gates" to allow information to selectively influence the state of each moment in the model. As a variant of LSTM, GRU replaces forgetting gate and input gate with update gate in LSTM. The structure of GRU is shown in figure 4, and the relevant calculation is shown in equations (6) to (9).

$$z_t=\sigma(W^z x_t+U^z h_{t-1}) \quad (6)$$

$$r_t=\sigma(W^r x_t+U^r h_{t-1}) \quad (7)$$

$$\tilde{h}_t=\tanh(W^h x_t+U^h(h_{t-1}\Theta r_t)) \quad (8)$$

$$h_t=(1-z_t)\Theta\tilde{h}_t+z_t\Theta h_{t-1} \quad (9)$$

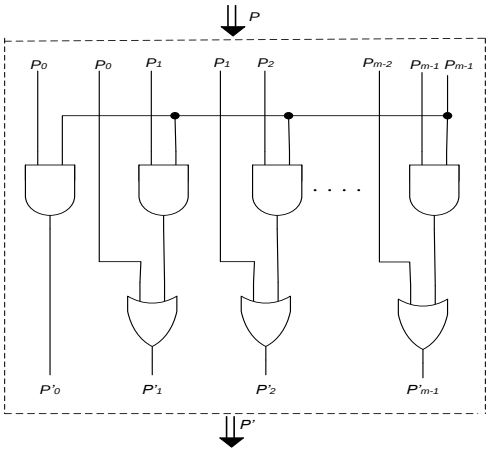

Figure 4. Structure of GRU

Where W^z, W^r, W^h, U^z, U^r, U^h are the weight matrices of the GRU. C stands for sigmoid function. Θ means elements multiplication. z_t is the update gate, which controls the update degree of GRU, and it is jointly determined by the current input state and the state of the previous hidden layer. r_t is the reset door, which integrates the new input information with the original information. h_t is the hidden layer. $\tilde{h}_t$ is the candidate hidden layer. In conclusion, compared with LSTM network, GRU network reduces model parameters and complexity, and it also reduces a large number of experimental costs.

In a classical RNN, the transmission of state is one-way from front to back. However, in some problems, the output at the current moment is related not only to the before state, but also to the after state. For example, the prediction of missing words in a sentence requires not only pre-context judgment, but also subsequent context content, which is solved by the bidirectional RNN as shown in figure 5.

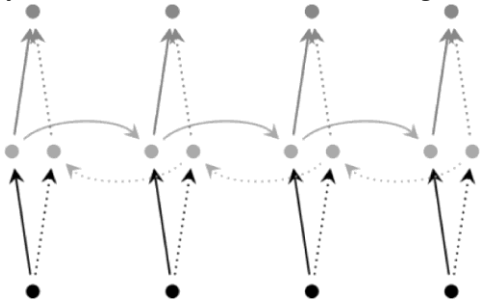

Figure 5. Structure of bidirectional RNN

Bidirectional RNN is combined two single RNNs. At each time, two RNNs in opposite directions are input at the same time. They determine the output together to make the result more accurate. The RNN in the bidirectional RNN is replaced with GRU structure to form Bi-GRU.

The model in this paper uses a Bi-GRU network to learn global semantic information from the multi-head attention output matrix X. In the training process, the network simultaneously uses two GRUs to conduct emotion modeling along the forward and backward direction of the text sequence. Finally, it outputs the hidden layer H_t. The specific calculation process is shown in equations (10)~(12).

$$\vec{h}_t=GRU(X,\vec{h}_{t-1}), t\in[1,L] \quad (10)$$

$$\overleftarrow{h}_t=GRU(X,\overleftarrow{h}_{t+1}), t\in[L,1] \quad (11)$$

$$H_t=[\vec{h}_t,\overleftarrow{h}_t] \quad (12)$$

Where $\vec{h}_0$ and $\overleftarrow{h}_{L+1}$ are initialized to the zero vector. $\vec{h}_t\in R^{L\times d}$ is the expression of emotion features of the word vector matrix X that integrates the information mentioned above. $\overleftarrow{h}_t\in R^{L\times d}$ is the emotion feature expression with fusing the later content. d is the vector dimension output by GRU unit. $H_t\in R^{L\times 2d}$ is connected in series by the two GRU units, and it integrates the contextual emotion information as the emotion representation of input text.

2.3. Feature fusion

Convolutional neural network can extract local features of text and reduce information loss. The bidirectional GRU network traverses the entire text sequence and extracts global semantic features. In this paper, the advantages of CNN and Bi-GRU network are integrated. The global average pooling method is used to integrate the local features and global semantic features of text to obtain the text instance feature representation V_s, which enhances the feature expression ability of the model.

During the experiment, the number of convolution kernel B in CNN is set to the same value as the output vector dimension 2d of the Bi-GRU network. The generated feature vectors by the two networks are combined and spliced as shown in equation (13).

$$H = concat(C, H_t) \tag{13}$$

Where, $H \in R^{(l+L)\times 2d}$ is the spliced vector. $C = [C_1, C_2, \cdots, C_n], C \in R^{l\times B}$ is the output vector of the convolutional neural network. $H_t = [h_1, h_2, \cdots, h_L], H_t \in R^{L\times 2d}$ is the output vector of the bidirectional GRU. $concat$ is a concatenation operation.

The global average pooling layer is used to average the vector H to form the feature points. These feature points are formed into the final feature vector V_s as the feature representation of the text emotion instance to avoid over-fitting and enhance the robustness of the model. The calculation is shown in equation (14).

$$V_s = gap(H) \tag{14}$$

In here, gap is the global average pooling operation.

2.4. Model training

The updated parameters of the hybrid neural network model in this paper include the parameters of convolutional neural network, BiGRU and attention mechanism. The Dropout layer is added before the fused features are fed into the Softmax classifier. Part of the trained parameters are abandoned in each iteration, so that the weight update is no longer dependent on part inherent features, and over-fitting is prevented. The probability that x is divided into category j with Sofmax regression in this paper is:

$$p(y^i = j \mid x^i, \theta) = \frac{\exp(\theta_j^T x^i)}{\sum_{i=1}^{k} \exp(\theta_j^T x^i)} \tag{15}$$

Where, k is the category number of labels. This paper focuses on positive, negative and neutral emotional polarities, so k=3. θ is the model parameter.

The training model parameter θ adopts categorical cross entropy as the loss function, and introduces L2 regularization to control the complexity of parameter value and avoid over-fitting. The specific calculation is shown in equation (16).

$$J(\theta) = -\frac{1}{N}[\sum_{i=1}^{N}\sum_{j=1}^{k} y_i \cdot \ln(p_j(\theta))] + \frac{\lambda}{2}\sum_{i=1}^{k}\sum_{j=1}^{n}\theta^2 \tag{16}$$

Where y_i is the true emotion value of the sentence. $p_j(\theta)$ is the predicted emotion value. N is the total number of samples. k indicates the number of label categories. n represents the number of parameter θ. λ represents the regularization coefficient of L2.

3. Experiments and analysis

This paper conducts experiments on three English datasets, including MR (Movie Review), IMDB, and SST-5 (Stanford Emotion Tree Library). All the above data sets have been widely used in emotion classification tasks, which makes the experimental results have a good evaluation effect. MR dataset is a collection of English film reviews. Each sentence is labeled as positive and negative according to emotion category with 5331 positive statements and 5331 negative statements. IMDB data set contains 50000 data sets from American film review sites, which are divided into positive and negative emotion categories for emotion orientation analysis. The SST-5 dataset is an extension of the MR dataset and provides divided training set, verification set and testing set with 11855 sentences in total. The data is labeled as "very positive," "positive," "neutral," "negative," and "very negative." In this paper, the training is carried out on the sentence-level of SST. The overview of each dataset is shown in Table 1.

Table 1. Experimental data set statistics

Dataset	training set	verification set	testing set	Number of categories
MR	8.6k	1.5k	1.1k	2
IMDB	25k	50k	25k	2

SST-5	8.5k	1.0k	2.2k	5

3.1. Experiment setting

This experiment is based on PyTorch. The English dataset uses the 300d Glove word vector to initialize the word embedding vector. For words that do not exist in the dictionary, it uses uniform distribution $U(-\varepsilon,\varepsilon)$ for random initialization. ε is set to 0.05. In order to train the English word vector in advance, the tool Fasthan is used to segment the text. Then the large-scale Wikipedia data is used to train the skip-gram model. The attention module adopts 8-head attention (h=8), and the Adam optimizer is used in the model training process. The learning rate is 0.001. The accuracy index is used to evaluate the model, and the specific super-parameters of the model are set in Table 2.

Table 2. Super-parameters of the model

Parameter	MR	IMDB	SST-5
convolution kernel width	2,3,4	3,4,5	3,4
convolution kernel number	512	256	256
Number of hidden units	256	128	128
Batch size	64	40	64
Dropout	0.5	0.4	0.5

3.2. Experiments comparison

In this paper, the Bi-GRU-MHA model in this paper is compared with the following four different methods: traditional machine learning method, deep learning method (CNN model and RNN model), language knowledge and model combination method, and capsule method, which are introduced as follows:

1) NBSVM [22]: Variants of Native Bayes (NB) and Support Vector Machines (SVM), it is often used as a baseline method for text classification.
2) CNN: The convolutional neural network uses filters with different sizes to convolve text word vectors, and then classifies them by connecting to the full connection layer after maximum pooling.
3) Bi-LSTM: A variant of LSTM network, which combines two-way text information to improve classification accuracy.
4) MC-CNN-LSTM: The model proposed in reference [23] utilizes multi-channel CNN to extract Ngram features of text as the input of LSTM, which effectively captures the key information in text.
5) LR-LSTM/LR-Bi-LSTM: The LSTM model based on language rules proposed in reference [12] integrates language knowledge in the model.
6) NCSL: The method learns the emotion value of text by using recurrent neural network is based on a simple weight sum model, but it requires complex language knowledge.
7) Multi-Bi-LSTM: An emotion model based on multi-channel bi-directional long and short-term memory network proposed in reference [24] needs to make the model fully learn the emotional information in sentences to achieve the best performance of the model.
8) Capsule-A/Capsule-B: Capsule network proposed in reference [25] is applied to text classification task.
9) RNN-capsule: the emotion classification Capsule model proposed in reference [26]. Compared with the model in this paper, this model only uses RNN to capture text sequence features.
10) Bi-GRU-MHA: A proposed text emotion analysis Capsule model combining multi-head attention and Bi-GRU network in this paper.

3.3. Results analysis

In this paper, experimental comparisons are made between three common datasets and the above models, the results are shown in Table 3. It can be seen from Table 3 that the proposed Bi-GRU-MHA model in this paper achieves better classification effect than other models on the three data sets. The classification accuracy values of the proposed model on the MR dataset, SST-5 dataset and IMDB dataset are 85.9% 51.6% and 92.7%, which improves by 2.1%, 2.2% and 3.5% than other models with the optimal classification.

Table 3. Experiments results of accuracy/%

Model	MR	SST-5	IMDB
NBSVM	75.5	46.8	83.5
CNN	76.1	46.9	85.6
Bi-LSTM	79.3	46.5	86.6
MC-CNN-LSTM	80.3	47.2	88.7
LR-LSTM	81.5	48.2	88.6
LR-Bi-LSTM	82.1	48.6	88.8
NCSL	82.9	47.1	88.9
Multi-Bi-LSTM	81.9	49.4	88.6
Capsule-A	81.3	46.2	89.2
Capsule-B	82.3	49.4	88.9
RNN-Capsule	83.8	49.3	88.9
Bi-GRU-MHA	85.9	51.6	92.7

First, CNN, Bi-LSTM and MC-CNN-LSTM models have better effect on the emotion classification task compared with the NBSVM method, but they are lower than Capsule-based methods. It is shown that using capsules to represent text emotional features can retain more emotion information and improve the classification performance. Moreover, the fusion of capsule method and linguistic knowledge model also shows the competition.

Second, the experiment performance of MC-CNN-LSTM in all data sets is superior to that of CNN and Bi-LSTM, which verifies the necessity of integrated convolutional neural network for local feature extraction and Bi-GRU for global text information extraction. On three publicly available English datasets, the accuracy of our model improves by 5.6%, 4.4% and 4.0% than that of MC-CNN-LSTM. It is shown that capsule model with vector neurons has stronger ability of emotion modeling. In terms of MR and SST-5 datasets, although the deep learning method integrating language knowledge and emotional resources shows a good classification performance compared with other baseline models, the accuracy of the proposed Bi-GRU-MHA model on MR dataset is higher than that of LR-Bi-LSTM, NSCL, multi-Bi-LSTM model, which improves by 3.8%, 3.0% and 4.0%, respectively. It also shows a better classification effect on multi-classification datasets. In addition, LR-Bi-LSTM, NSCL models rely excessively on linguistic knowledge, such as sentiment dictionaries and intensity regularizers. It is important to note that building such linguistic knowledge requires a great deal of human intervention. Compared with the above two modeling methods, the multi-Bi-LSTM model is more concise, but it is still a deep learning model based on language knowledge and emotional resources, which requires a lot of manpower and time costs. However, the model in this paper does not need to model any language knowledge and emotional resources. Using capsules to model text emotional features achieves better classification effect than the deep learning model that integrates language knowledge and emotional information, indicating that the model in this paper is more efficient and simpler.

Third, compared with the Capsule method, the classification accuracy of RNN-Capsule on MR dataset is higher than Capsule A (2.5%) and Capsule B (1.5%), but the classification performance of RNN-Capsule on IMDB dataset is slightly worse than Capsule B(0.1%). This is because the IMDB dataset is a long text dataset (average sentence length is 294), while the MR dataset is a short text dataset (average sentence length is 20). RNN-Capsule is used to extract text sequences and averages the hidden features according to the sentence length to obtain the final example feature representation. The longer sentence length denotes the worse instantiation representation of the vector. It cannot well represent the emotion category of the text, which affects the final performance of the model. So the performance of RNN-capsule on IMDB data set is poor. Capsule-A and Capsule-B are classified by dynamic routing

mechanism instead of the pooling layer to generate capsules. And it connects them to the fully connection Capsule layer, and the text length has little effect on them. The classification accuracy of the proposed model on the three datasets is higher than RNN-Capsule, and its classification performance on the IMDB data set is also higher than Capsule-A and Capsule-B, which effectively verifies the dependency relationship between words encoded with multi-head attention and the integrated convolutional neural network. It overcomes the limitation of RNN-Capsule long text vector representation, and the efficient performance of global average pooling layer to generate text instance features on English dataset, which shows the robustness and generalization capability of Bi-GRU-MHA.

In this paper, the concept of capsule is introduced into the model, and vector neurons are used to replace scalar neurons, which not only reduces the loss of information, but also enhances the emotion modeling ability. Moreover, this learning approach based on vector units is different from the general neural network model.

We conduct an experiment on the performance of the model based on vector learning on the MR dataset, and the results are shown in table 4. By changing the vector dimension of capsule model and the dimension of reconstructed vector, the change of accuracy on the testsets is obtained. The experiment results show that the classification accuracy of the model will be higher when the vector with larger dimension is used to represent the emotion features of text. Therefore, when the training object is a vector, the ability to express the emotion features of the text will be enhanced, and it may represent various attributes of the text.

Table 4. Average accuracy

MR			SST-5			IMDB		
512d	256d	128d	512d	256d	128d	512d	256d	128d
89.6%	87.3%	85.7%	91.2%	89.3%	87.8%	93.4%	91.2%	87.1%

In order to show more intuitively that multi-head attention can capture the dependency of emotion words and coded words in the text, this paper visualizes the distribution of word attention weight in the sentence to show the important emotion features in the text [27-29]. As shown in table 5, the positive and negative samples in the IMDB dataset are taken as examples to label the emotion features of the text, in which the weight of the darker part is larger, while the weight of the blue part is smaller.

Table 5. Visualization of Attention Weight

IMDB positive sample	IMDB negative sample
This was one of the **best** war movies. I've seen because it focuses on the characters more then the actual war. All of the cast do an **excellent** job and because most of them are relative unknowns it makes everything seem more believable. The camera footage is **great** is so was the pacing and editing. This movie will actually get to you and causes the audience to care for the characters.	I've written at least a half dozen **scathing** reviews of this abysmal little flick and none get published, so I must opine that someone at imdb.com really likes this **awful** movie. The idea that a bunch of oilmen can resurrect a military tank that has set in the desert for over a decade, and make a fighting machine of it again is **ludicrous**. So is the acting and direction. Pass on it.

4. Conclusions

In this paper, a capsule model combining convolutional neural network and Bi-GRU is proposed for English text emotion classification task. The model uses multi-head attention to capture emotion words to solve the problem that capsule network can not selectively focus on important words in text classification task. To extract multi-level and more comprehensive text emotional features, CNN is used to collect local features and Bi-GRU network is used to extract global semantic features. Using capsules to model text emotion, the method achieves better classification performance than other methods, which proves the feature expression ability of capsule model. The effectiveness of the proposed model is verified by conducting experiments on different datasets.

In the next step, the improvement of the internal mechanism of emotion capsule, such as the optimization of attention mechanism, can be considered. At the same

time, the ability of feature fusion is enhanced, so that the vector can better represent the emotional features, and the stability and efficiency of the model are improved.

Acknowledgements.

The author is grateful to the reviewers for their anonymous reviews.

References

[1] P. Ajitha, A. Sivasangari, R.-I. Rajkumar, et al., Design of text sentiment analysis tool using feature extraction based on fusing machine learning algorithms, Journal of Intelligent and Fuzzy Systems 40(1) (2020) 1-9.

[2] M. Giatsoglou, M.-G.Vozalis, K. Diamantaras, A. Vakali, G. Sarigiannidis, Konstantinos Ch.Chatzisavvas, Sentiment analysis leveraging emotions and word embeddings, Expert Systems with Application 69 (2017) 214-224.

[3] H.-A. Bouarara, Sentiment Analysis Using Machine Learning Algorithms and Text Mining to Detect Symptoms of Mental Difficulties Over Social Media, International Journal of Information Systems and Social Change, 2021.

[4] L. Teng, H. Li and S. Yin, Modified Pyramid Dual Tree Direction Filter-based Image De-noising via Curvature Scale and Non-local mean multi-Grade remnant multi-Grade Remnant Filter, International Journal of Communication Systems 31(16) (2018) e.3486.1-e.3486.12.

[5] M. A. El-Affendi, K. Alrajhi and A. Hussain, A Novel Deep Learning-Based Multilevel Parallel Attention Neural (MPAN) Model for Multidomain Arabic Sentiment Analysis, IEEE Access 9(2021) 7508-7518.

[6] Y. Zhao, H. Li, S. Yin, Y. Sun, A new Chinese word segmentation method based on maximum matching, Journal of Information Hiding and Multimedia Signal Processing 9(6) (2018) 1528-1535.

[7] J. Lytmki, P. Ohtonen, L Aa Kso M, et al., The role of linguistic and cognitive factors in emotion recognition difficulties in children with ASD, ADHD or DLD, International Journal of Language & Communication Disorders 55(2) (2020).

[8] J. Yu, H. Li, S. Yin, New intelligent interface study based on K-means gaze tracking, International Journal of Computational Science and Engineering 18(1) (2019) 12-20.

[9] L. Unsworth, K.-A. Mills, English language teaching of attitude and emotion in digital multimodal composition, Journal of Second Language Writing (2020) 47:100712.

[10] J. Lin, Y. Gu, Y. Zhou, A. Yang, J. Chen and X. Li, Combining Convolutional Neural Networks and Word Topic Features for Chinese Short Text Sentiment Analysis, 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), 2018, pp. 422-425.

[11] X. Sun, C. Li, F. Ren, Sentiment analysis for Chinese microblog based on deep neural networks with convolutional extension features, Neurocomputing 210 (2016) 227-236.

[12] Q. Qian, M. Huang, J. Lei, et al., Linguistically regularized LSTMs for sentiment classification, Computational Linguistics 14(4) (2016) 34-37.

[13] U.-M. Osmanoglu, O.-N. Atak, K. Alar, et al., Sentiment Analysis for Distance Education Course Materials: A Machine Learning Approach, Journal of Educational Technology and Online Learning 3 (2020).

[14] Y. Dong, Y. Fu, L. Wang, Y. Chen, Y. Dong and J. Li, A Sentiment Analysis Method of Capsule Network Based on BiLSTM, IEEE Access 8 (2020) 37014-37020.

[15] N. Antypa, A. Smelt, A. Strengholt, et al., Effects of omega-3 fatty acid supplementation on mood and emotional information processing in recovered depressed individuals, Journal of Psychopharmacology 26(5) (2012) 738.

[16] S. Yin, H. Li, L. Teng, M. Jiang & S. Karim, An optimised multi-scale fusion method for airport detection in large-scale optical remote sensing images, International Journal of Image and Data Fusion 11(2) (2020) 201-214.

[17] M.-Z. Boito, A. Villavicencio, L. Besacier, Investigating alignment interpretability for low-resource NMT, Machine Translation 34(1) (2020).

[18] Shoulin Yin, Hang Li, Asif Ali Laghari, et al. A Bagging Strategy-Based Kernel Extreme Learning Machine for Complex Network Intrusion Detection[J]. EAI Endorsed Transactions on Scalable Information Systems. http://dx.doi.org/10.4108/eai.6-10-2021.171247

[19] Qingwu Shi, Shoulin Yin, Kun Wang, et al. Multichannel convolutional neural network-based fuzzy active contour model for medical image segmentation. Evolving Systems (2021). https://doi.org/10.1007/s12530-021-09392-3

[20] S. Yin and H. Li. Hot Region Selection Based on Selective Search and Modified Fuzzy C-Means in Remote Sensing Images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 5862-5871, 2020, doi: 10.1109/JSTARS.2020.3025582.

[21] Shoulin Yin, Hang Li, Desheng Liu and Shahid Karim. Active Contour Modal Based on Density-oriented BIRCH Clustering Method for Medical Image Segmentation [J]. Multimedia Tools and Applications. Vol. 79, pp. 31049-31068, 2020.

[22] S. Wang, C.-D. Manning, Baselines and bigrams: Simple, good sentiment and topic classification, Proceedings of the 50th Annual Meeting of the Association for Computational

Linguistics: Short Papers Vol 2. Association for Computational Linguistics (2012) 90-94.

[23] H. Zhang, W. Jin, J. Zhang, et al., YNU-HPCC at SemEval 2017 Task 4: Using A Multi-Channel CNN-LSTM Model for Sentiment Classification, Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017). 2017.

[24] Z. F. Qian, L. Y. Li, Z. Q. Tao and L. L. Kun, Research on Sentiment Analysis of Two-way Long and Short Memory Network Based on Multi-Channel Data, 2020 IEEE 6th International Conference on Computer and Communications (ICCC), 2020, pp. 1728-1732.

[25] W. Zhao, J. Ye, M. Yang, et al., Investigating Capsule Networks with Dynamic Routing for Text Classification, Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018.

[26] C. Lin and H. Chang, Dimentional sentiment analysis by synsets and sense definitions, 2016 International Conference on Asian Language Processing (IALP), 2016, pp. 332-335.

[27] Shoulin Yin, Jie Liu, and Lin Teng. A Sequential Cipher Algorithm Based on Feedback Discrete Hopfield Neural Network and Logistic Chaotic Sequence [J]. International Journal of Network Security. Vol. 22, No. 5, pp. 869-873, 2020.

[28] Yin, S., Li, H. & Teng, L. Airport Detection Based on Improved Faster RCNN in Large Scale Remote Sensing Images [J]. Sensing and Imaging, vol. 21, 2020. https://doi.org/10.1007/s11220-020-00314-2

[29] Xiaowei Wang, Shoulin Yin, Ke Sun, et al. GKFC-CNN: Modified Gaussian Kernel Fuzzy C-means and Convolutional Neural Network for Apple Segmentation and Recognition [J]. Journal of Applied Science and Engineering, vol. 23, no. 3, pp. 555-561, 2020.