

Received December 31, 2019, accepted January 23, 2020, date of publication February 24, 2020, date of current version March 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2973711

A Sentiment Analysis Method of Capsule Network Based on BiLSTM

YONGFENG DONG^{1,4,5}, YU FU^{1,4,5}, LIQIN WANG^{1,4,5}, YUNLIANG CHEN²,
YAO DONG^{1,4,5}, AND JIANXIN LI³

¹School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China

²School of Computer Science, China University of Geosciences, Hubei 430074, China

³Smart Networks Data Laboratory, Faculty of Science, Engineering and Built Environment, School of Information Technology, Deakin University, Burwood, VIC 3125, Australia

⁴Hebei Province Key Laboratory of Big Data Computing, Tianjin 300401, China

⁵Hebei Engineering Research Center of Data-Driven Industrial Intelligent, Tianjin 300401, China

Corresponding author: Liqin Wang (wangliqin@scse.hebut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61806072, and in part by the Natural Science Foundation of Tianjin City, China under Grant 19JCZDJC40000.


ABSTRACT Nowadays, capsule network model is widely used in image processing, whose feature engineering is not suitable for sentiment analysis based on texts obviously. In this paper, we propose a capsule network model with BiLSTM named caps-BiLSTM for sentiment analysis to solve the problem, and introduce the experimental results on different datasets. At the beginning of caps-BiLSTM, a convolution layer is used to transform the instance to hide vector. Then the capsule module constructs the capsule representation to the n-gram model. The state probability of a certain capsule is calculated by the capsule model. If the state probability of a given instance is the largest among all capsules, a higher coupling coefficient is assigned. Finally, in order to fusion the data features, the output of the capsule enters into a BiLSTM structure, which is used as a decoder to get the probability representation. Experimental results based on MR, IMDB and SST datasets show that the proposed method can achieve better performances than the traditional machine learning methods and the compared deeping learning models.

INDEX TERMS Sentiment analysis, neural network, capsule network, BiLSTM.

I. INTRODUCTION

Sentiment analysis is an important task in the field of natural language processing, which is to identify the emotional polarity in the original texts, such as positivity, negativity or neutrality. With the development of social networks, the role of Internet users is transforming. More and more people express their opinions on the Internet, and these opinions produce a complete text finally. These texts contain a large amount of available information, which are great of use to understand what the users think in evaluating social networks or products [1].

For a long time in the past, the methods of sentiment analysis focused on combining with sentiment dictionaries and machine learning. On the one hand, the lexicographical method analyzes the text through the position and grammar of words to determine the polarity of opinions [2], [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Shirui Pan .

On the other hand, the method of machine learning can learn sentiment classification through the texts manually annotated [4], [5]. However, in recent years, deep neural network structures based on convolution or recursion has become the main method of sentiment analysis, such as the convolutional neural network(CNN) used by Kim [6] and the long short time memory (LSTM) neural network used by Xu *et al.* [7].

As have been mentioned above, these methods, that provide different solutions to the problems of sentiment analysis, are used in various sentiment analysis tasks widely, and have achieved excellent results of performance evaluation. However, most of the traditional methods have the limitation of relying too much on language knowledge, or it is difficult to extract text features comprehensively. Moreover, the learning and training process after feature encoding is not effective enough in traditional models. The accuracy of evaluation results is less than satisfying. Therefore, the task of sentiment analysis still faces great challenges.

In order to solve the problems of the previous chapters, a BiLSTM-integrated capsule neural network model is considered, with the goal of extracting text features in a wider range, increasing the learning and generalization ability of the network to improve the accuracy of sentiment analysis. In this paper, we propose a model named caps-BiLSTM, which obtains word vectors used to represent semantic information first, and then inputs these vectors into the capsule network to learn the similarity between the inputs and the targets. Finally, the resulting vectors are used as the input of the BiLSTM network to train and predict the emotional category labels of the text. Experimental results show that the accuracy of our method is significantly improved compared with the baseline models. Our main work of this paper includes the following three parts:

- (i) using a convolutional layer based n-gram extracts the features of sentences and captures the dependencies on words with the capsule network, so as to present the rich structures of the feature sequence;
- (ii) the modified dynamic routing algorithm is adopted to update the weight parameters, so as to extract more accurate text information and improve the reliability of routing adaptive tuning;
- (iii) BiLSTM is used for a deeper learning in the outputs of the capsule network to increase the learning and generalization ability of the model to obtain a better performance.

II. RELATED WORK

It is obvious that the algorithms based on machine learning effectively show their superior performance in sentiment analysis task than other methods of previous work. Among the researches, Nakagawa *et al.* [8] proposed a Conditional Random Field(CRF) model based on the dependency tree, which achieved good results. Pang *et al.* [4] used Naive Bayes classification and maximum entropy to study the emotional polarity of texts. However, with the increasing size of datasets and more annotated datasets being exposed, the booming development of deep learning methods provides inspiration to address the challenge of sentiment analysis.

Benefiting from the large-scale corpus, several methods based on deep learning can learn the potential semantic and syntactic features of articles actively with better robustness and higher accuracy, which effectively make up for the shortage of artificial feature engineering. Convolutional Neural Network(CNN) [9] is able to extract the text features effectively, and has made a breakthrough in sentiment analysis. Liu *et al.* [10] conducted experiments on the task of sentiment analysis with CNN, obtaining satisfactory results. However, CNN is still unsatisfactory in capturing long-distance dependent information, since CNN only retains the size of feature but ignores the important information of feature direction and spatial position. The Recurrent Neural Network(RNN) [11] and capsule neural network [12] has addressed the above limitations.

In recent years, RNN has been applied in many fields, such as language model [13], speech recognition [14] and

so on. RNN can predict the outputs according to the long distance characteristics by learning the previous information. Since the original RNN still has some defects, such as gradient disappearance and gradient dispersion, researchers have proposed a series of variants of RNN, including Long Short-Term Memory [15] (LSTM), bidirectional LSTM(BiLSTM), Gated Recurrent Unit(GRU) [16]. These models overcome the gradient problem of RNN and perform better than the original RNN in many tasks. In the task of sentiment analysis, Tang *et al.* [17] combined LSTM and target-related information to process the texts, which improved the accuracy of the results. At present, the neural network model combining attention mechanism is also widely applied to sentiment analysis, and has achieved better results than the previous methods [18], [19].

With the vigorous development of CNN and RNN, Hinton *et al.* [20] put forward the concept of "capsule" for the first time. The core work of capsule network is to introduce spatial relationships and position directions on the basis of the traditional neural network, and identify objects by combining invariance and covariability. Referring to the work of Hinton, Sabour proposed the capsule network, which used the vector output capsule instead of the traditional convolution neural network scalar output, not using pooling but dynamic routing algorithm. The process of dynamic routing can adjust coupling coefficient of connection strength according to the similarity between the target and the input to get better results. Inspired by state-of-the-art performance of the capsule network on MNIST, Zhao *et al.* [21] have introduced the capsule network into text-related tasks and verified the superiority of its performance on multiple datasets, which is the first work done by the capsule network in the actual modeling process. Later, Wang *et al.* [22] combined capsule network and attention mechanism for sentiment analysis based on text and realized the complexity of routing calculation. Not only does this model pay special attention to the feature during training, but also can effectively adjust parameters of different features. The capsule network can extract richer text features while retaining the position information of words, grammar and syntax, so that effectively encode and improve the representation of text, and further obtain effective information. Some scholars [23] made a summary of the current research work on capsule networks, pointing out that there are not many applications of capsule networks so far, and had a broader development prospect.

III. OUR MODEL

In this paper, a capsule network model incorporating BiLSTM is proposed, which is called caps-BiLSTM, as shown in figure 1. This model consists of two parts, namely capsule combination module and BiLSTM learning module. First, a sentence represented by a word embedding is input into the model. Then, after feature extraction and dynamic routing iterations, the output vector is used to predict the polarity of sentiment.

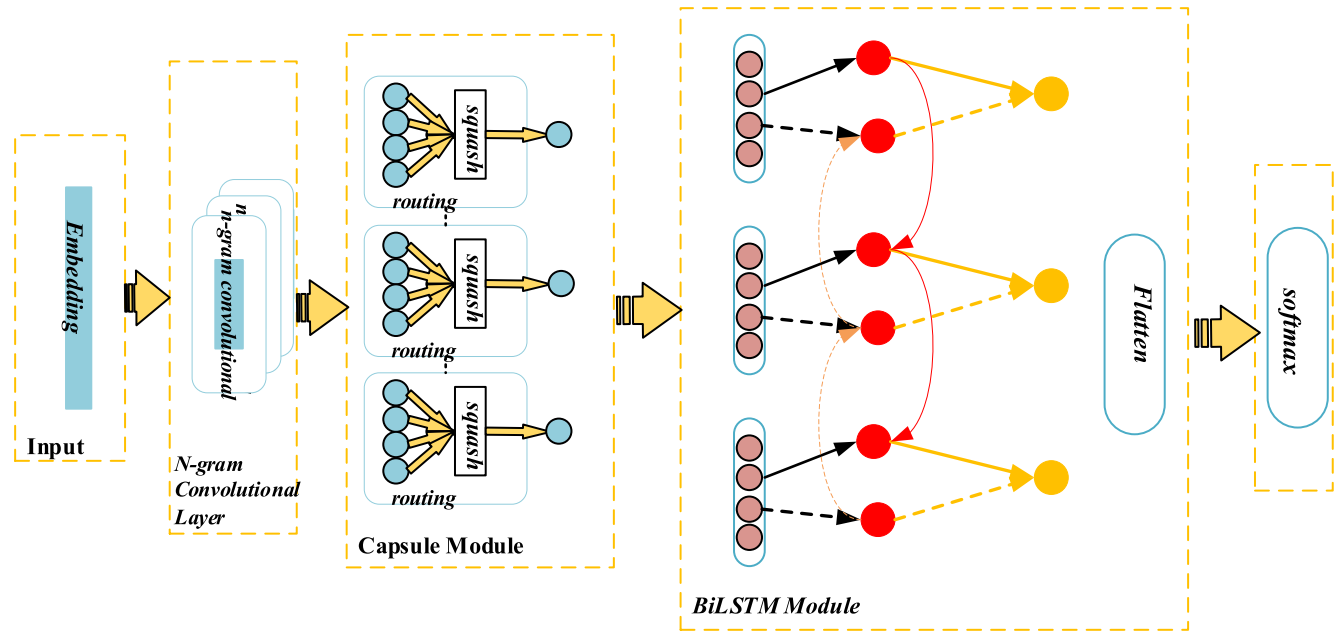


FIGURE 1. The structure of caps-BiLSTM.

A. CAPSULE COMBINATION MODULE

The capsule network was first proposed by Sabour [12] in 2017, and the first practical modeling application was in the work of Zhao et al. [21] in 2018. Pooling in traditional neural network means deleting several pieces of data that contain knowledge of features, so some features may be lost. The capsule network not only overcomes the point mentioned above, but also proposes a more intuitive way to group information on a hierarchical level to solve more complex problems as well. In the capsule network, the neuron vector replaces the single neuron node in the traditional neural network, and the training process uses dynamic routing algorithm. Dynamic routing algorithm can increase the weight of vital features and find more hidden features, which improves the performance of the network.

B. N-GRAM CONVOLUTIONAL LAYER

The convolutional layer is an intermediate process of text on capsule representation, whose input is word embedding representation of sentences. Word embedding is the distributed representation of a word, in which words of the vocabulary are mapped to vectors. The network performance can be improved by initializing the text with word embedding as the input of the convolutional layer, for they can capture the grammatical and semantic information of words from large scale text corpus. In this work, we used the vector representation of the text obtained by the GloVe [24] word embedding as the input of the model, which is used in natural language processing tasks commonly.

Different from image processing, considering the composition, hierarchy and structure of the text, n-gram features are extracted at different positions of the sentence in the

convolutional layer in the capsule network through the convolutional filter. $x \in R^{(L \times V)}$ represents the input of a sentence (L is the length of the sentence, V is the dimension of the word embedding), and $x_i \in R^V$ represents the word embedding corresponding to the i th word in the sentence. $W^\alpha \in R^{L-K_1+1}$ is the convolutional filter, where K_1 is the length of n-gram, that is, the length of the sliding window on the sentence, which is used to extract feature at different positions of the sentence.

Convolutional filter W^α generates a feature map m^α in each window (after sliding a step into the next window). Every element in this feature map is shown in Equation (1).

$$m_i^\alpha = f(x_{i:i+K_1-l} \circ W^\alpha + b_0) \tag{1}$$

where \circ is the unit multiplication, and b_0 is the bias, f is the nonlinear activation function, and l is the sliding step distance. The above is the process of extracting text features by filters. So for a total of B -filters with the same n-gram size, B feature maps can be generated and rearranged as shown in Equation (2).

$$M = [m_1, m_2, \dots, m_B] \tag{2}$$

C. CAPSULE MODULE

The input vectors of the original capsule layer come from the output neuron M of the convolutional layer, and the output of the capsule replaces the scalar of the traditional neural network with the vector so as to retain the instantiation parameters.

First, the activation function is used to transform the feature vector M_i of each n-gram sliding window into the corresponding feature capsule u_i , as shown in

Equation (3)-(4).

$$U_i = g(W_b M_i + b_1) \quad (3)$$

$$u_{ji} = W_{ij} U_i \quad (4)$$

where W_b is the filter shared by different sliding windows, and b_1 is the capsule bias. Meanwhile, g is the nonlinear activation function. W_{ij} represents the weight matrix of the correlation between the input layer and the output layer.

The capsule network updates the weight of the coupling coefficient through the dynamic routing process. The coupling coefficient of the adjacent capsule depends on the similarity of them, and the coupling coefficient of similar capsules is larger. The dynamic routing strategy is more efficient than the basic routing in CNN, ensuring that the output of each capsule is sent to the corresponding parent in the subsequent layer.

Given each prediction vector u_{ji} and its probability of existence a_{ji} , the purpose of the dynamic routing iteration process is to update the coupling coefficient c_{ij} , which symbolizes the similarity between the input vector and the target value, and to assign higher weight c_{ij} to the v_j and u_{ji} that are closer to each other. The equation for coupling coefficient updating is based on Zhao et al. [21]. The initial value of b_{ij} is 0. The dynamic routing process is shown in Equation (5)-(8). Equation (7) is the compression function, which can squeeze the modulus of the input vectors to [0,1).

$$c_{ij} = a_{ji} \cdot \exp(b_{ij}) \frac{\exp(b_{ij})}{\exp(b_{ik})} \quad (5)$$

$$S_j = \sum_j c_{ij} u_{ji} \quad (6)$$

$$v_j = \frac{\|S_j\|^2}{1 + \|S_j\|^2} \frac{S_j}{\|S_j\|}, \quad a_j = |v_j| \quad (7)$$

$$b_{ij} \leftarrow b_{ij} + u_{ji} \cdot v_j \quad (8)$$

D. BiLSTM MODULE

LSTM is a popular recursive neural network structure designed to possess better ability to capture long-term dependencies than standard RNN. Its emergence solves the long-term dependency problem of RNN and may make it better in discovering and exploiting dependencies of long distance data. LSTM is an improved model of RNN, which introduces some gates internally to try to solve the gradient problem. As a recursive neural network, LSTM computes an output vector on the basis of the current input and the output of the previous unit, which is then used as the input of the next unit. Finally, the output of the hidden layer is used for classification.

As mentioned above, in LSTM, each neuron introduces a new structure, which is mainly composed of four parts: input gate i_t , output gate o_t , forget gate f_t and memory unit c_t . The internal structure of individual components of LSTM module is shown in figure 2.

The calculation formula is shown in Equation (9)-(14).

$$f_t = \sigma(W_{(f)}x_t + U_{(f)}h_{t-1} + b_{(f)}) \quad (9)$$

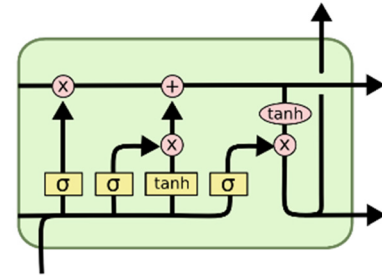


FIGURE 2. The internal structure of an individual component in LSTM.

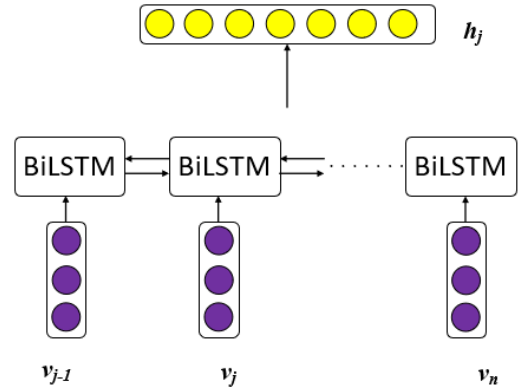


FIGURE 3. The structure of BiLSTM.

$$i_t = \sigma(W_{(i)}x_t + U_{(i)}h_{t-1} + b_{(i)}) \quad (10)$$

$$o_t = \sigma(W_{(o)}x_t + U_{(o)}h_{t-1} + b_{(o)}) \quad (11)$$

$$c_t = \tilde{c} + i_t \odot \tanh(W_{(c)}x_t + U_{(c)}h_{t-1} + b_{(c)}) \quad (12)$$

$$\tilde{c} = f_t \odot c_{t-1} \quad (13)$$

$$h_t = o_t \odot \tanh(c_t) \quad (14)$$

where x_t represents the input at time t , σ is the sigmoid activation function, and \odot is the corresponding multiplication of elements. c_t is used to reduce the problem of gradient disappearance or explosion, so LSTM can learn longer information dependence. The forget gate f_t is a reset memory unit, i_t and o_t represent the input and output gates, which are used to control the input and output of the memory unit.

But LSTM can not encode the information on back to front. In order to access the backward and forward features in a given time, the bidirectional LSTM network was proposed. The forward LSTM and backward LSTM are combined to BiLSTM, which can provide additional context information and has better and faster learning ability. The structure of BiLSTM is shown in Figure 3. In this module, BiLSTM is used to learn the output of the capsule layer, in order to enhance the fitting effect of the network features and the generalization ability on the new data set.

For the output of BiLSTM, the softmax activation function is used for classification, as shown in Equation (15).

$$p = \text{softmax}(W_c s + b_c) \quad (15)$$

TABLE 1. Experimental environment.

Experimental Environment	System Version
operating system	Windows10
Processor	Inter Core i5 7400
Graphics card	NVIDIA GeForce GTX 1060
Programing language	Python 3.6.0
Deep learning framework	Keras 2.2.2

IV. EXPERIMENTS AND RESULTS

A. DATASETS

To evaluate our proposed methods, we conducted experiments on three benchmark datasets: Movie Review [25](MR), Internet Movie Database [26] (IMDB) and Stanford Sentiment Treebank [27] (SST). These datasets are widely used in the sentiment analysis. The experimental results are compared with the published algorithm results to achieve the objective comparison effect.

MR is from Cornell University, a collection of English film reviews collected from www.rottentomatoes.com, including 5331 positive English film reviews and 5331 negative English film reviews. The average sentence length is 20.

IMDB is made up of 50000 Internet Movie Reviews with serious polarization. This dataset contains 25000 training data and 25000 test data, including 25000 positive and 25000 negative comments. The average sentence length is 294.

SST is an extension of MR data set, which provides the divided training set, verification set and test set, with 11855 sentences in total. According to the sentiment tendency of sentences, data labels are divided into five categories: very positive, positive, neutral, negative and very negative. The average sentence length is 19.

B. PARAMETER SETTINGS

In this experiment, in the text initialization stage, the pre trained GloVe word embedding is used to transform the text into a vector, each word vector dimension is 300. We got the best parameter values through repeated trainings. In the initial convolution layer, the size of the sliding window is set to 4; the 4-gram feature of the sentence is extracted; the number of filters is 32; and the activation function of the convolution layer is ReLU. In the capsule layer, the size of the capsule is set to 16; the number of capsules is 32; and the number of dynamic route iterations is 3. The number of neurons in BiLSTM layer is set to 128. In order to reduce the training loss, Adam optimizer is selected with learning rate of 0.001. The batch size is 128. In order to prevent from over fitting, a dropout layer is added with a parameter of 0.5. In order to reduce the training time and ensure that the training can be stopped in time when the over fitting phenomenon occurs, the early stopping mechanism is introduced. The initial number of epoch is set to 50. When the verification loss lasts for 10 epochs without improvement, the iteration is stopped. The experimental environment is shown in Table 1.

TABLE 2. Accuracy rate under different algorithm models (%). The bold represents the best result.

	MR	IMDB	SST
SVM	76.14	86.9	36.03
LSTM	75.90	87.64	45.60
BiLSTM	79.30	88.81	46.50
Tree-LSTM	80.70	--	48.10
LR-LSTM	81.50	--	48.20
CNN	76.10	86.84	46.90
ToWE-CBOW	65.10	90.80	--
CapsNet-dynamic-routing	81.00	89.80	--
caps-BiLSTM	81.47	91.96	48.34

C. COMPARISON OF CLASSIFICATION ACCURACY

The evaluation protocol of this experiment is accuracy, as shown in Equation (16).

$$\text{accuracy} = \frac{T}{N} \quad (16)$$

where T represents the number of samples with correct classification and N represents the total number of samples. We compare our proposed method caps-BiLSTM with the following method: SVM [4] (Pang 2002), LSTM [28] (CHO *et al.*, 2014), BiLSTM [28] (CHO *et al.*, 2014), tree structured LSTM [29] (Tree-LSTM) (Tai *et al.*, 2015), LSTM regulated by linguistic knowledge [30] (LR- LSTM) (Qian *et al.*, 2016), CNN [6] (Kim, 2014), ToWE-CBOW [31] (Liu *et al.*, 2018) and CapsNet-dynamic-routing [32] (Kim *et al.*, 2020). The sentiment analysis experiment was carried out on the above three datasets.

Table 2 presents the experimental results on all the datasets. The results clearly demonstrate that our proposed method significantly outperforms the compared models for IMDB and SST. And the deep learning methods are superior to the method of machine learning represented by the SVM. However, the experimental result of MR dataset is a bit poorer than LR-LSTM, but still better than other seven models. This is because n-gram model can extract the information on the text very well, and the dynamic routing algorithm of the capsule network can find the capsules with higher similarity in the target value. Hence, BiLSTM can learn the context characteristics of the sequence and classify accordingly. Compared with the experimental results, the accuracy rate of caps-BiLSTM model improved by 1.12% at least on IMDB than ToWE-CBOW, while the accuracy rates of MR and SST were not significantly different from that of the baseline models such as Tree-LSTM and CapsNet-dynamic-routing, indicating that the performance of our model in the long texts was better than that in the short texts.

D. VERIFICATION LOSS ANALYSIS

The loss of caps-BiLSTM on each validation data set (val_loss) changes with the epoch as shown in Figure 4. It can be seen that the loss of the model reached the minimum around the 5th to 7th epoch. After that, there is over fitting phenomenon in training, and the accuracy decreases

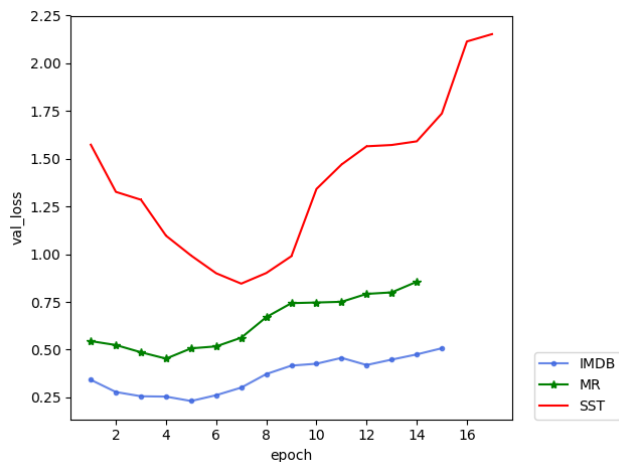


FIGURE 4. The chart of epoch and val_loss.

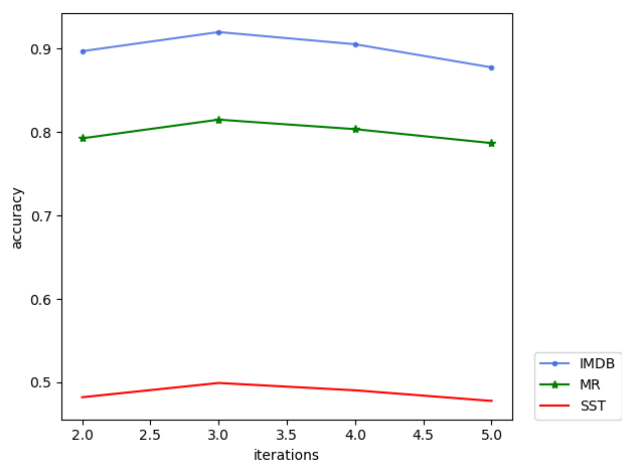


FIGURE 5. The chart of iterations and accuracy.

accordingly. Because the early stopping mechanism is added to the training, the training is stopped after gaining 10 epochs with the minimum loss, and the best classification result is obtained. In this way, the best model in training can be found automatically without setting the epoch number.

E. ANALYSIS OF DYNAMIC ROUTE ITERATIONS

The accuracy of the model on each data set changes in the number of dynamic routing iteration as shown in Figure 5. It can be seen that when the number of iterations is 3, all datasets can achieve the highest accuracy. This is because dynamic routing can't fully learn the characteristics of data when the number of iterations is less than 3, and it may cause over fitting phenomenon when the number of iterations is more than 3, both of which will reduce the accuracy.

V. CONCLUSION AND FUTURE WORK

In this paper, we propose a BiLSTM based capsule network model for sentiment analysis, which mainly improves the performance of the capsule network. By comparing the

machine learning method and some deep learning methods, the validity of caps-BiLSTM network in feature extraction and the accuracy in classification is proved. This model has a favorable performance in the sentiment analysis task. While the generalization ability of the model on the new data set is still weak. Consequently, to improve the generalization ability of the model and reduce the verification loss is the main goal in the future work.

REFERENCES

- [1] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lectures Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.
- [2] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2004, pp. 168–177.
- [3] S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," in *Proc. 20th Int. Conf. Comput. Linguistics (COLING)*, 2004, p. 1367.
- [4] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: Sentiment classification using machine learning techniques," in *Proc. Conf. EMNLP*, 2002, pp. 79–86.
- [5] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews," in *Proc. 12th Int. Conf. World Wide Web*, 2003, pp. 519–528.
- [6] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. Conf. EMNLP*, 2014, pp. 1746–1751.
- [7] J. Xu, D. Chen, X. Qiu, and X. Huang, "Cached long short-term memory neural networks for document-level sentiment classification," in *Proc. Conf. EMNLP*, 2016, pp. 1660–1669.
- [8] T. Nakagawa, K. Inui, and S. Kurohashi, "Dependency tree-based sentiment classification using CRFs with hidden variables," in *Proc. Int. Conf. Assoc. Comput. Linguistics*, 2010, pp. 786–794.
- [9] C. Huiping, W. Lidan, and D. Shukai, "Sentiment classification model based on word embedding and CNN," *Appl. Res. Comput.*, vol. 33, no. 10, pp. 2902–2905, 2016.
- [10] L. F. Liu, L. Yang, and S. W. Zhang, "Sentiment tendency analysis of weibo based on convolutional neural network," *J. Chin. Inf. Sci.*, vol. 29, no. 6, pp. 159–165, 2015.
- [11] R. Dey and Y. Hong, "CompNet: Complementary segmentation network for brain MRI extraction," 2018, *arXiv:1804.00521*. [Online]. Available: <http://arxiv.org/abs/1804.00521>
- [12] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3856–3866.
- [13] T. Mikolov, M. Karafiát, L. Burget, J. Černocký, and S. Khudanpur, "Recurrent neural network based language model," in *Proc. Conf. INTER-SPEECH*, 2010, pp. 1–24.
- [14] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Netw.*, vol. 18, nos. 5–6, pp. 602–610, Jul. 2005.
- [15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [16] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv:1412.3555*. [Online]. Available: <http://arxiv.org/abs/1412.3555>
- [17] D. Tang, B. Qin, and X. Feng, "Effective LSTMs for target-dependent sentiment classification," in *Proc. Int. Conf. Comput. Linguistics*, 2015, pp. 3298–3307.
- [18] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*. [Online]. Available: <http://arxiv.org/abs/1409.0473>
- [19] W. Wang, Y. X. Sun, and J. Q. Qin, "Text sentiment classification model based on BGRU-attention neural network," *Appl. Res. Comput.*, vol. 36, no. 12, pp. 126–137, 2018.
- [20] G. E. Hinton, A. Krizhevsky, and S. D. Wang, "Transforming auto-encoders," in *Proc. Int. Conf. Artif. Neural Netw.*, Espoo, Finland, 2011, pp. 44–51.
- [21] W. Zhao, J. Ye, M. Yang, Z. Lei, S. Zhang, and Z. Zhao, "Investigating capsule networks with dynamic routing for text classification," 2018, *arXiv:1804.00538*. [Online]. Available: <http://arxiv.org/abs/1804.00538>
- [22] Y. Wang, A. Sun, J. Han, Y. Liu, and X. Zhu, "Sentiment analysis by capsules," in *Proc. World Wide Web Conf. World Wide Web*, 2018, pp. 1165–1174.

- [23] R. Katarya and Y. Arora, "Study on Text Classification using Capsule Networks," in *Proc. 2019 5th Int. Conf. Adv. Computing Commun. Syst. (ICACCS)*, Mar. 2019, pp. 501–505.
- [24] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1532–1543.
- [25] B. Pang and L. Lee, "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales," in *Proc. 43rd Annu. Meeting Assoc. Comput. Linguistics*, 2005, pp. 115–124.
- [26] A. L. Maas, R. E. Daly, and P. T. Pham, "Learning word vectors for sentiment analysis," in *Proc. 49th Annu. Meeting Assoc. Comput. Linguistics: Hum. Lang. Technol.*, 2011, pp. 142–150.
- [27] R. Socher, A. Perelygin, and J. Wu, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proc. Conf. EMNLP*, 2013, pp. 1631–1642.
- [28] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*. [Online]. Available: <http://arxiv.org/abs/1406.1078>
- [29] K. S. Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term memory networks," in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Language Process.*, vol. 1, 2015, pp. 1556–1566.
- [30] Q. Qian, M. Huang, J. Lei, and X. Zhu, "Linguistically regularized LSTMs for sentiment classification," 2016, *arXiv:1611.03949*. [Online]. Available: <http://arxiv.org/abs/1611.03949>
- [31] Q. Liu, H. Y. Huang, Y. Gao, X. Wei, Y. Tian, and L. Liu, "Task-oriented word embedding for text classification," in *Proc. Int. Conf. Comput. Linguistics*, 2018, pp. 2023–2032.
- [32] J. Kim, S. Jang, E. Park, and S. Choi, "Text classification using capsules," *Neurocomputing*, vol. 376, pp. 214–221, Feb. 2020.

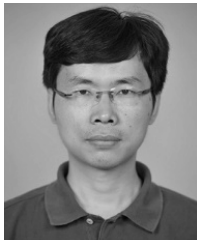


LIQIN WANG was born in Zhangbei, Hebei, China, in 1980. She received the B.S. and M.S. degrees in computer science and technology and the Ph.D. degree in electrical engineering from the Hebei University of Technology, Tianjin, in 2003, 2007, and 2015, respectively.

From 2007 to 2009, she held an assistant position at the Hebei University of Technology. Since 2010, she has been an Experimentalist with the School of Artificial Intelligence, Hebei University of Technology. She is the author of more than ten articles and over ten inventions. Her research interests include intelligent information processing-based big data, deep learning in image processing and applications, and representation learning in knowledge graph.



YUNLIANG CHEN received the B.Sc. and M.Eng. degrees from the China University of Geosciences and the Ph.D. degree from the Huazhong University of Science and Technology, China. He is currently an Associate Professor with the School of Computer Science, China University of Geosciences, Wuhan, China. His research interests include data mining, cloud computing, and the IoT.



YONGFENG DONG was born in Zhengding, Hebei, China, in 1977. He received the B.S. and M.S. degrees in computer science and technology and the Ph.D. degree in electrical theory and new technology from the Hebei University of Technology, Tianjin, in 2000, 2003, and 2008, respectively.

From 2010 to 2015, he was an Associate Professor with the Hebei University of Technology, where he has been a Professor with the School

of Artificial Intelligence, since 2016. He is the author of one book, over 20 articles, and more than 20 inventions. His research interests include intelligent information processing, big data technology, robot and intelligent control, and software engineering.

Dr. Dong is a Senior Member of the China Computer Federation, a member of the Education Committee of the China Computer Federation, the Standing Director of the Hebei Computer Federation, and the Director of the Hebei Institute of Electronics.



YAO DONG is currently a Senior Experimentalist with the School of Artificial Intelligence, Hebei University of Technology. She is a member of the China Computer Federation. Her interests mainly include artificial intelligence and robot path.



YU FU received the bachelor's degree from the School of Computer Science, Hebei University of Technology, Tianjin, China, in 2017, where she is currently pursuing the M.S. degree in computer technology. Her research interests include natural language processing, sentiment analysis, and deep learning.



JIANXIN LI is currently an Assistant Professor with the School of Information Technology, Deakin University, Australia. He has published 70 high-quality research articles in top international conferences and journals, including PVLDB, IEEE ICDE, ACM WWW, IEEE ICDM, EDBT, ACM CIKM, IEEE TKDE, and ACM WWW. His research interests include social computing, query processing and optimization, big graph data analytics, and educational data computation.

...