News marketing recognition based on attention neural network model

Shuhua Cao[1], Pengxiang Gao[2]

[1]School of Data Science and Software Engineering, Qingdao University, Qingdao, Shan dong Province, China

[2]School of Data Science and Software Engineering, Qingdao University, Qingdao, Shan dong Province, China

[*]Corresponding author's e-mail: 2018025539@qdu.edu.cn

Abstract: Aiming at the problem of insufficient text feature extraction in the existing short news text classification methods for identifying marketing intentions and difficulty in capturing the semantic information in the text, a short text news marketing recognition method based on the attention mechanism neural network model is proposed. The model uses convolutional neural network to extract local features, and then uses bidirectional long and short-term memory network to extract contextual semantic information, strengthens the learning of features, and introduces a double-layer attention mechanism to calculate feature weights to further obtain influential features between sentences. Finally, perform text classification at the softmax function layer. Through experiments on two standard data sets, the experimental results show that this method has better classification performance, and has a significant improvement in accuracy compared with other models.

1. Introduction

With the rapid development of the information age, a large amount of text information data has been produced. Massive data information is faced with the problems of complexity and uneven quality. Obtaining effective information is of great significance to people. News reading is an important way for people to effectively obtain information. With the widespread dissemination of news, the number of news releases for marketing purposes is increasing. Negative marketing communication uses a large number of decontextualized, one-sided, dogmatic short paragraphs, popular articles, news and reviews, etc., which cause negative effects from multiple aspects and levels. Therefore, it is important to accurately identify malicious marketing news. News marketing recognition is a text classification task in the field of natural language processing. The key to recognizing text intent is to establish an effective classification algorithm to correctly determine the text category according to the text content [1]. Traditional text classification research focuses on rule-based classification and machine learning model classification. Rule-based methods rely on the definition of data content categories. The processing of regular text is fast and effective, but it is not effective for large-scale data processing. it is good. Short text classification methods based on machine learning have achieved good results in text classification. Common machine learning classification algorithms mainly include SVM[2], Naive Bayes[3], etc., through supervised learning of labeled samples, and the kernel function Target optimization reduces the loss, so as to establish a classifier on the feature set, which can effectively learn the characteristics of the training sample category, but its classification effect has a great dependence on feature selection and manual preprocessing, including manual feature engineering and over-learning process Fitting, training data size limitation, etc. Ensemble learning is also widely used. It uses the word vector model combined with the text classification framework to assign weights to different texts, which improves the accuracy of the classification model. However, the ensemble learning model has a complex structure, deep model hierarchy, and long training time, which is not conducive to large-scale Data processing [4].

In recent years, deep learning technology has achieved good results in natural language processing tasks. Deep learning based on neural networks has the advantages of manual feature engineering and semantic feature learning, which can effectively extract text features and improve classification effects. In the field of natural language processing, convolutional neural network (CNN) is applied to sentence classification tasks. Its basic structure is convolutional layer plus pooling layer[5]. When the number of convolutional layers of CNN increases, the model can extract Deeper features, but will encounter the problem of disappearing gradients[6].Kim[7] proposed to use pre-trained word vectors as the input of CNN to achieve sentence-level text classification. The Recurrent Neural Network (RNN) proposed by Zaremba [8] captures the semantic information of sentence context, but it has long-term dependence. The Long Short Term Memory (LSTM) is an improved model for RNN problems. Gers et al. [9] first proposed to use the LSTM model on text classification problems, which was significantly improved over machine learning algorithms, and then further introduced the two-way LSTM model. The Bi-LSTM (Bi-directional Long Short-Term Memory) model can perform text classification. Extract features in both directions. Sutskever et al.[10] proposed the Seq2Seq model incorporating the attention mechanism, which has been effectively applied in a variety of text tasks. In the current Chinese short text classification tasks, facing the problems of complex sentence semantics and large vocabulary, the research on processing such Chinese tasks mainly adopts feature expansion and deep learning methods. Zhou [11] proposed the C-LSTM model, which constructed a network model by linking CNN and LSTM, which enhanced the feature learning ability of the model and improved the model classification effect. Ganin et al. [12] proposed to use word vectors to input the CNN model to improve the feature expression ability of the model. Yin et al. [13] proposed the introduction of an attention mechanism model in a convolutional neural network to help the model improve its classification ability. Lu et al. [14] proposed to introduce an attention mechanism to automatically weight the output sequence information at each moment when RNN is performing sequence feature modeling, and reduce the loss of key features. Internet news is rich in information, subject information is diverse, and data is sparse. Although the above methods put forward many methods to solve the problem of short news classification, they cannot solve the marketing intention news well. The sentences of this type of news are colloquial. Features such as aggressiveness and suggestiveness, and insufficient feature information is not extracted. In response to the above problems, a neural network model based on the attention mechanism is proposed. The model integrates attention into CNN and LSTM networks.

The main contributions of our work are as follows:

1. Propose a CBiLSTM-ATT model, which combines the convolutional neural network and the bidirectional cyclic neural network, and improves the classification ability of the model through feature fusion from the local semantics of the text and the overall semantics.

2. The model adds a double-layer attention mechanism, and adds a layer of attention in the convolution process to help extract the local representative part of the text. In order to prevent the CNN network from ignoring the sentence context information, after the CNN and BiLSTM network features are fused, the model introduces an attention mechanism to improve the feature extraction ability of the model.

3. Experiments show that this method can extract features of marketing news data, thereby further improving the accuracy of recognition.

2.Model

In this part, we will introduce the deep learning model proposed in this article for short news marketing intent recognition. The model structure is as follows As shown in Figure 1. First, after the word embedding, it is used as a CNN network combined with the attention mechanism for feature extraction.

After the pooling layer, it is input into the BiLSTM network, and the attention layer is used for weight calculation to capture contextual semantic features. Finally, the classification result is obtained in the softmax layer.

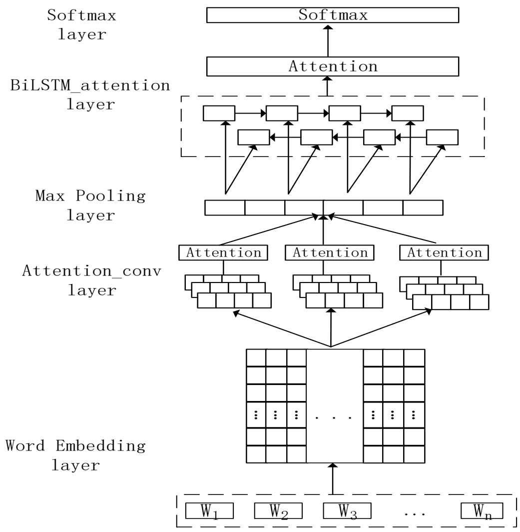

Figure.1 CBiLSTM-ATT network model

2.1 Text embedded representation

Define the maximum length of sentences in the data set as n, and each sentence $X\{w_1,w_2,w_3,\ldots w_n\}$ is a two-dimensional data matrix formed by vertical splicing of d-dimensional word vectors. If the sentence length is less than n, the missing vector is randomly initialized from the uniform Gaussian distribution U(-0.25,0.25). For each word $w_i \in R^d$ in the sentence, the vector can be expressed as $x_i \in R^d$ and d is expressed as the dimension of the vector.

2.2 Attention Convolutional Neural Network

The attention-fusion CNN network extracts important local features of sentences, and learns different types of features through multiple convolutional layer filters. During the convolution operation, a convolution kernel of size k is used in each of the feature map matrix H The window generates new features, the formula is as follows:

$$A_i = f(\omega \times x_{i:i+k-1} + b) \tag{1}$$

Among them, ω represents the convolution kernel, f represents the nonlinear function, b represents the bias, and obtains the feature matrix $C=[c_1,c_2,\ldots,c_{n-k+1}]$ after passing through the convolution layer. The attention mechanism is introduced to calculate the importance of different features in the text sequence to distinguish the key information and interference information that determine the text category, and the obtained convolution layer feature matrix is combined with the attention matrix. For the words t_i in a given sentence, the attention calculation method is:

$$a_i = \sigma(\sum(V[i-k:i+k] \circ H_a) + b_a) \tag{2}$$

V represents the context word vector, k is the window size, $H_a \in R^{d\times k}$ is the attention weight matrix, and b_a represents the attention bias. The attention mechanism here uses the sliding window principle, and by changing K, the content of the context can be reduced or increased. Then multiply the attention matrix $A[a_1,a_{2,\ldots},a_i]$ with the feature C of the convolutional layer to obtain a new feature combination, which is expressed as follows:

$$J[\mathrm{c}_1 \bullet a_1, \mathrm{c}_2 \bullet a_2, \ldots, \mathrm{c}_i \bullet a_i] \tag{3}$$

Journal of Physics: Conference Series **1792** (2021) 012048 doi:10.1088/1742-6596/1792/1/012048

The convolutional neural network further extracts features with relatively high attention weight through the maximum pooling layer, which is obtained after pooling, and serialized and connected to obtain a new feature matrix as the input of the BiLSTM layer.

2.3 Attention BiLSTM network

The BiLSTM network is composed of a forward LSTM and a reverse LSTM, which makes up for the lack of information in the LSTM network. The two-way structure provides complete past and future context information at each moment in the input sequence of the output layer.[15]. LSTM changes the hidden layer of the recurrent neural network to a memory unit, and at the same time uses a door switch to realize the memory function of the data, which effectively solves the problem of gradient disappearance. The LSTM memory unit can be used to calculate the vector of the hidden unit at the current moment. The hidden layer representation of the forward LSTM at time t is $\vec{h}_t$, and the reverse LSTM hidden layer representation is $\overleftarrow{h}_t$. The hidden layer representation of the BiLSTM network output is as follows:

$$h_t = \vec{h}_t \oplus \overleftarrow{h}_t \tag{4}$$

Add the attention layer again to assign different attention weights to the semantic coding of the hidden layer, representing the importance of words in the context, generate attention weights v_t , probabilize the weights, generate probability vectors p_t , and then assign the attention weights to the corresponding hidden Layer state codes h_t to obtain the weighted average a_t of attention, the specific calculation is as follows:

$$v_t = \tanh(h_t) \tag{5}$$

$$p_t = \frac{\exp(v_t)}{\sum_{t=1}^{m} \exp(v_t)} \tag{6}$$

$$a_t = \sum_{t=1}^{m} p_t \bullet h_t \tag{7}$$

3.Experiment

In order to verify the effectiveness of the model in this paper, experiments were carried out on the Sohu News Content Algorithm Competition dataset and Fudan University News Corpus dataset. Among them, the Sohu News data set is the largest Chinese data set used to identify marketing news. The training data includes labeled data and unlabeled data. 80% of the data is used as the training set and 20% is the test set. The Fudan University News Data Set contains 30,000 training data and 1,900 test data. The experiment is based on the Keras deep learning framework. The server configuration is as follows: Windows7, Python version: 3.6.8, Keras version: 2.3.1, CPU version: i5-4210U CPU @1.70GHz 2.40GHz.

3.1 Experimental parameter settings

The experiment uses the GLOVE model [16] to train word vectors, and sets the vector dimension to 200. The window size of the convolutional neural network filter is set to 4. The size of the hidden layer of the bidirectional cyclic neural network is set to 32, the number of iterations is set to 20, and the random inactivation rate is set to 0.5. The model evaluation indicators are Acc, F1, Loss, the accuracy rate acc is the ratio of the number of correctly classified samples to the total number of samples, and F1 is the weighted average of the accuracy rate (P) and the recall rate (R) [17]. The specific calculation formula is as follows:

$$F1 = \frac{2PR}{P+R} \tag{8}$$

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \tag{9}$$

3.2 Analysis of results

In order to verify the effectiveness of the method in this paper, experiments were conducted on two data sets to compare with traditional methods and deep learning neural network models. The comparative experimental results of different methods are shown in Table 1

Table 1 Comparison results of different model experiments

		SVM	XGboost	CNN	LSTM	BiLSTM	CNN-BiLSTM	CBiLSTM-ATT
SOUHU	F1	0.4510	0.4972	0.5249	0.6186	0.6838	0.7300	0.7458
	Acc	0.4537	0.5865	0.5629	0.6283	0.6945	0.7368	0.7637
FuDan	F1	0.4286	0.5795	0.5622	0.6303	0.6720	0.6935	0.7541
	Acc	0.4371	0.5900	0.5857	0.6339	0.6808	0.7201	0.7590

It can be seen from the results in Table 1 that when compared with traditional machine learning models and ensemble learning methods, the deep neural network model is significantly better than the traditional model. Experiments show that the neural network can obtain the semantic feature information of the text more effectively. The CNN model can effectively extract the local information of the sentence and improve the classification ability, but it cannot effectively obtain the sentence context information, resulting in loss of information. The LSTM network can obtain long-distance sequence information, fully extract the contextual semantic relationship, and further improve the classification ability of the model. Compared with LSTM, BiLSTM has the advantage of capturing semantic information in two directions. The combination of convolutional neural network and recurrent neural network model is better than a single CNN or LSTM or BiLSTM model. This type of model combines the advantages of neural network to extract local features and global features to obtain more text feature information. By comparing the above Acc and F1 indicators, the model in this paper has obvious advantages compared with a single neural network model and a combined neural network model. Based on the combined advantages of the model, the attention mechanism is introduced to focus on the important information that determines the news category. It is proved that the attention model assigns attention weights to input vectors, strengthens important semantic information during feature extraction, reduces irrelevant semantic information, and improves classification accuracy. In order to further verify the performance of the model, through experiments on data sets of different scales, the model in this paper outperforms other methods in terms of accuracy, and is relatively stable as the data scale increases. As shown below:

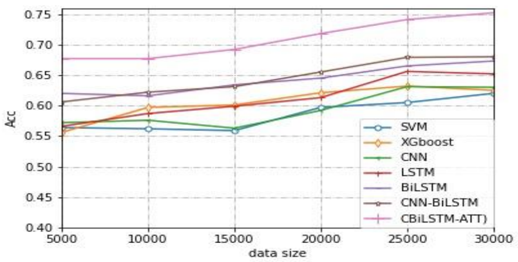

Figure.2 comparison of Acc values of models Based on the SOHU data

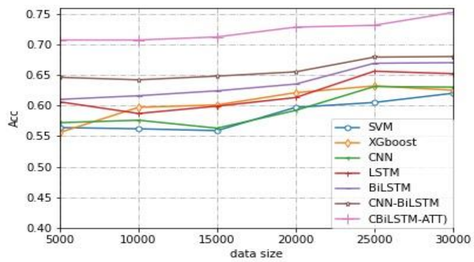

Figure.3comparison of Acc values of models Based on the FuDan data

4.Conclusion

This paper proposes a marketing news recognition model based on the attention neural network model. The model combines CNN and BiLSTM networks. The combined model can capture the advantages of local and global semantics separately. At the same time, a double-layer attention mechanism is introduced to further weight important features. Distribution, in addition to obtaining influential feature information, reduces the interference of unimportant information, and extracts text features more effectively, so as to achieve a good classification effect. The effect of this model is verified on the provided Chinese data set, and the model has good performance on different scales of data. Finally, in the research of news marketing identification, our model needs to further improve its classification ability. Incorporating Chinese dictionaries into the model is the research goal in future research tasks.

Acknowledgments
Shandong Province Key R&D Program (Major Innovation Project) Project Fund (2019JZZY020101), Qingdao University

References
[1] Wang Yufeng. Research and application of marketing intention recognition based on integrated learning and topic memory network [D]. University of Jinan, 2020.
[2] Liu Yongfen, Cheng Li, Chen Zhian. M-SVM Chinese text classification based on feature selection[J]. Software, 2019(9).
[3] Wenfeng G, Hong Z, Ruoyi C. A Study on Online Detection of micro-blog Rumors Based on Naive Bayes Algorithm[C]// 2020 Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC). 2020.
[4] Wang Y, Liu S, Li S, et al. Stacking-Based Ensemble Learning of Self-Media Data for Marketing Intention Detection[J]. Future Internet, 2019, 11(7):155.
[5] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J].Institute of Electrical and Electronics Engineers, IEEE, 1998, 86 (11): 2278 -2324.
[6] Computer ence, 2015.[7]KIM Y.Convolutional neural networks for sentence clas-sification[C]//Proceedings of the EMNLP, 1746-1751.
[7] Tao Yongcai, Yang Chaoyang, Shi Lei, et al. News text classification method combining pooling and attention[J]. Small Microcomputer System, 2019, 40(11).
[8] Zaremba, W., Sutskever, I., Vinyals, O., 2014. Recurrent neural network regularization.arXiv preprint arXiv:1409.2329.
[9] Gers F, J. Schmidhuber, and F. Cummins. Learning to Forget: Continual Prediction with LSTM. Istituto Dalle Molle Di Studi Sull Intelligenza Artificiale, 1999.
[10] Sutskever I, Vinyals O, Le Q V. Sequence to Sequence Learning with Neural Networks[J]. Advances in neural information processing systems, 2014.
[11] Zhou C, Sun C, Liu Z, et al. A C-LSTM Neural Network for Text Classification[J]. Computer ence, 2015, 1(4):39-44.
[12] GANIN Y, USTINOVA E， AJAKAN H, et al. Domain-adversarial training of neural networks[J]. Journal of Machine Learning Research, 2017, 17(1): 2096 - 2030.
[13] Yin W, Schütze, Hinrich, Xiang B, et al. ABCNN: Attention-Based Convolutional Neural Network for Modeling Sentence Pairs[J]. Computer ence, 2017.
[14] Lu Jian, Ma Chengxian, Yang Tengfei, et al. Multi-category text information classification under Text-CRNN+attention architecture[J]. Application Research of Computers, 2020, 037(006):1693-1696,1701.
[15] Li Wenhui, Zhang Yingjun, Pan Lihu. Improved short text classification method of biLSTM network [J]. Computer Engineering and Design, 2020, 041(003):880-886.
[16] Chen Zhenrui, Ding Zhiming. Word vector improvement method based on GloVe model[J]. Computer System Applications, 2019, 028(001):194-199.
[17] Wang Y, Ma K, Garcia-Hernandez L, et al. A CLSTM-TMN for marketing intention detection[J]. Engineering Applications of Artificial Intelligence, 2020, 91(May):103595.1-103595.9.