Named Entity Recognition in Chinese E-commerce Domain Based on Multi-Head Attention

Hongliang Mao[1], Azragul Yusup[1,2*], Yifei Ge[1] and Degang Chen[1]
[1]College of Computer Science and Technology, Xinjiang Normal University Urumqi, Xinjiang 830054, China
[2] National Language Resource Monitoring & Research Center of Minority Languages, Beijing 100081, China
Wrinkle_625@163.com, Azragul2010@126.com,1453259830@qq.com, iloveqq001@vip.qq.com
*corresponding author

***Abstract*—At present, the demand for entity recognition for product query and product recommendation systems of e-commerce domain is growing. Traditional methods that rely on experts to define artificial features and domain knowledge can no longer meet the needs of the domain due to their low recognition accuracy. According to the above background, a Chinese named entity recognition method in the e-commerce domain that incorporates a multi-headed attention mechanism and bidirectional long-short term memory network is proposed. A word vector represents the product title statements of e-commerce platforms. The generated word vector sequences are fed into a bi-directional long-short term memory network that mines their contextual semantic features. A multi-headed attention mechanism is introduced to focus on the entity words in the text to unearth their hidden features. In contrast, the entity recognition results are labeled by calculating the joint probability of sequence labels through conditional random fields. The experimental data show that the method can reach 86.16% accuracy and 86.57% F1 value for entity recognition in the e-commerce domain, which is practical and feasible.**

Keywords- e-commerce; named entity recognition; bilstm; multi-headed attention; recommend system

I. INTRODUCTION

E-commerce is a new generation of business forms formed by computer technology to electronic and information various aspects of traditional business activities. With the continuous progress and development of Internet technology, the scale of Chinese online shopping users has gradually expanded, and many e-commerce platforms have been created. In e-commerce platforms, product information usually consists of text descriptions and pictures, while users' needs are typically expressed in text. How to extract keywords such as entities from the text search information entered by users and match them with the entity keywords in the product text description information becomes the key for each e-commerce platform to provide sellers with consumers' previous shopping information and buyers with accurate product search information. Named entity recognition tasks for the e-commerce domain have also become the focus of commercial platforms and academia.

NER was first proposed in MUC-6 (Message Understanding Conference), which is an essential task for text information extraction. Its purpose is to identify entities with strong denotations in text data, including names of people, places, organizations, dates and times, etc. Yan et al. [1]proposed improving the accuracy of Chinese named entity recognition by embedding BERT word vectors, fusing bidirectional threshold cyclic units and attention mechanism, and its F1 value reached 94.31%, which has a significant significance impact on the field of named entity recognition. Huang et al. [2]proposed to achieve named entities recognition in Chinese judicial domain by fusing self-attention mechanism and dilated convolution recognition. Their results show that the iterative dilated convolutional neural network makes full use of the parallelism of GPUs to reduce significantly the time cost of using long-short term memory networks and achieves better results for entity recognition of legal documents. Luo et al. [3]proposed a named entity recognition approach for chemical domains based on a bidirectional long-short term memory network and an attention mechanism, and its recognition rate reached 92.57%. Yang et al[4]proposed a two-layer attention model based on the principle of attention mechanism, which is word level and sentence level, and this structure can bring out the critical information in the document and get greater attention, proving that the attention mechanism can assign higher weights to the critical information in the text. Qiu et al. [5]proposed a deep learning-based approach for named entity recognition of online shopping reviews based on the problem that keyword information is easily ignored in entity recognition in online shopping reviews, which improves the accuracy of named entity recognition of online shopping reviews by adding a bidirectional long-short term memory network and a self-attention mechanism to the traditional conditional random field. The above analysis demonstrates that named entity recognition for e-commerce domain has not only practical commercial value and research value but also reliable theoretical support and methodological guidance. Previously, although linguistics-based and statistics-based methods can achieve the goal of domain named entity recognition, they are primarily based on artificial features labeled by the logical intuition of professionals, which cannot tap the hidden information features in the text, and their recognition methods are too weak in generalization.

II. RELATED WORK

At present, the following difficulties mainly exist for named entity recognition in the e-commerce domain: (1) The semantic entities in the e-commerce domain are rapidly

2767-6684/22/$31.00 ©2022 IEEE
DOI 10.1109/DSA56465.2022.00083

updated and diversified, and it is difficult to classify the entity categories.(2) E-commerce domain entities lack natural separators on Chinese, with boundary ambiguity, semantic ambiguity, nested ambiguity and other recognition difficulties.(3) The length of entity words in the text of commodity titles in the e-commerce domain varies, and there are specific difficulties in recognition. (4) Although the traditional linguistics-based and statistics-based methods can achieve the goal of identifying domain entities, they are primarily based on the artificial features labeled by the logical intuition of professionals, which cannot tap the hidden information features in the text, and the generalization ability of their identification methods is too weak.

To address the above problems, based on multi-headed attention mechanism and bi-directional long and short term memory network, we propose a Chinese named entity recognition approach in the e-commerce domain by integrating multi-headed attention mechanism and bi-directional long and short term memory network, namely BiLSTM-MAM-CRF (bi-directional long and short term memory network-conditional random fields) model. and short term memory network-multi headed attention - conditional random fields) model, which mainly accomplishes the following aspects: (1) The two-way long-short term memory network is used to extract the semantic features of sentence context, calculate the interdependencies between words of commodity title text, highlight the relationships between semantic entities in the text such as modification and collocation, and improve the accuracy of entity recognition.(2) Fine-grained delineation of entity categories in e-commerce domain alleviates the problem of unclear semantic entity categories and difficult identification. (3) The word weight assignment mechanism in multi-headed attention is introduced to add weights to entity categories by the different contribution of each entity category to entity recognition in the product title text in e-commerce domain to alleviate the problem of inaccurate recall caused by the same attention of semantic entities in different entity categories under the limited arithmetic power.(4) The problem of information loss due to equal contribution of semantic entity recognition is mitigated by dividing attention weights at the entity and non-entity levels.

III. MODEL

The general framework of the proposed Chinese named entity recognition model in the e-commerce domain incorporating multi-headed attention mechanism and bi-directional long and short-term memory network is shown in Fig. 1. The model mainly consists of input layer, word vector layer, BiLSTM layer, multi-headed attention mechanism layer, and CRF layer. The input layer obtains each character vector of the item title text statement as the input of the whole model, the BiLSTM layer is used to capture the contextual semantic features, the multi-headed attention mechanism layer is used to assign attention weights between entities and non-entities and between different classes of entities, which makes the model recognition more accurate and efficient, and the CRF layer is used to calculate the joint probability between the current character and the forward and backward characters and output the label optimal results.

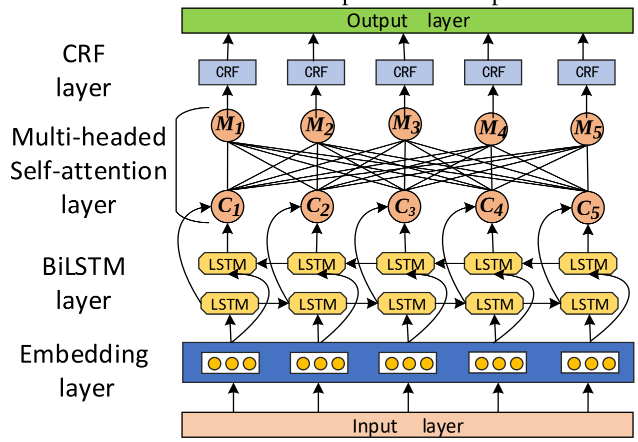

Fig. 1. Structure of BiLSTM-MAM-CRF

A. *Input Layer*

The primary role of the input layer is to pre-process and vectorize the text of product headings in the e-commerce domain. The text preprocessing process mainly involves processing or eliminating sentences that are too short, too long, and have many special symbols in the product title text.

In the model training process, the text cannot be directly fed into the neural network for training, and it needs to be vectorized to represent it. In this paper, we use the Word2vec model for text vectorization; firstly, we encode the text with one-hot to form a high-dimensional sparse vector and then input it to the Word2vec model for training to convert it into a low-dimensional continuous dense vector. The specific procedure is as follows: let the sample sentence X consist of n words, denoted as $X=\{t_1,t_2,...,t_n\}$, where t_t is the one-hot representation of the t word, and x_t is the word embedding[5]

$$x_t = W^{emb} t_t \tag{1}$$

where: $W^{emb} \in R^{d\times|v|}$ is the vector lookup table, $t_t \in R^{|v|}, x_t \in R_d$ is the vector dimension, and $|v|$ is the size of the dictionary.

B. *BiLSTM Layer*

LSTM has the same framework as a typical recurrent neural network but uses a different way to compute the hidden state, especially suitable for temporal problems. It can well solve the gradient disappearance and gradient explosion problems generated by RNN in extracting long-range sentence features.

The LSTM cell state diagram is shown in Fig. 2. At a specific moment t, each LSTM cell has three inputs, which are the cell state C_{t-1} of the last moment, the output value h_{t-1} of the LSTM cell of the last moment, and the input value X_t of the cell of the current moment, and two outputs, which are the cell state C_t and the output value h_t of the LSTM cell of the present moment. The LSTM cell is controlled by forgetting gates input, In essence, the LSTM gate mechanism is equivalent to a fully connected layer, where the input is a multidimensional vector, and the output is a number between 0 and 1, with 0 indicating that the information of the previous unit cannot be entirely inherited and 1 indicating that the knowledge of the last unit can be entirely inherited. The current moment unit can get the valid information of all previous units. The disadvantage is that it cannot get the unit's information after the current unit. To improve this drawback, the concept of bi-directional long–short term memory network, namely BiLSTM, is proposed.

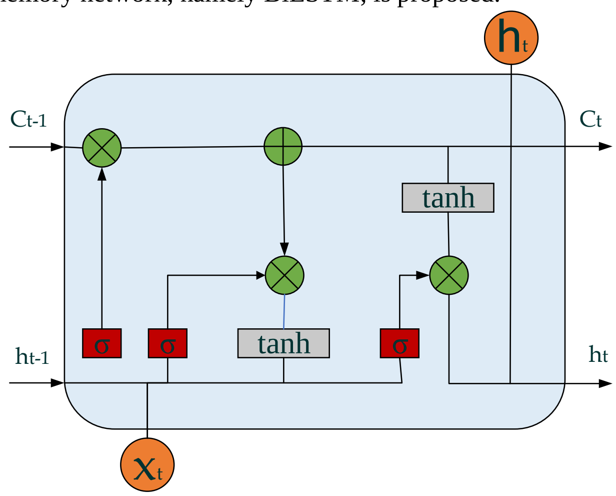

Fig. 2. Structure of LSTM cell

BiLSTM is combined with bi-directional LSTM, and its model structure is shown in Fig. 3. BiLSTM[6] obtains the information above the sentence by forward LSTM and the info below the sentence by backward LSTM, it splices the results of forward LSTM and backward LSTM and outputs them to the CRF layer to get the sequence label of the utterance.

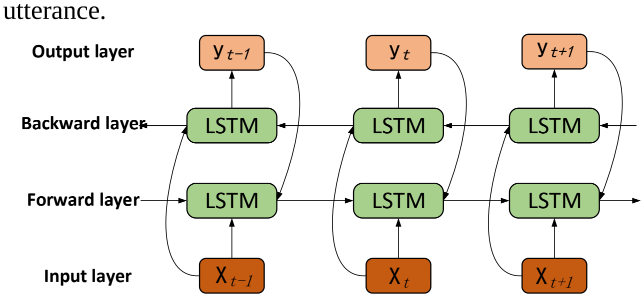

Fig. 3. Schematic diagram of BiLSTM model

The specific process is as follows: X_t is the word embedding vector obtained in the previous layer, fed into the model and computed.

Step 1 Compute the forward LSTM.

$$h_t^{forward} = LSTM^{forward}(h_{t-1}, x_t, c_{t-1}) \tag{2}$$

Step 2 Compute the backword LSTM.

$$h_t^{backward} = LSTM^{backward}(h_{t-1}, x_t, c_{t-1}) \tag{3}$$

Step 3 Concat the $h_t^{forward}$ and reverse $h_t^{backward}$.

$$h_t = [h_t^{forword}, h_t^{backward}] \tag{4}$$

where: h_{t-1} denotes the output value of the network at the moment $t-1$, C_{t-1} denotes the cell state at the moment $t-1$, h_t is the output vector at moment t, and the matrix formed by the sequence of output vectors of this layer is denoted as $H=(h_1,...h_i,...h_t)$.

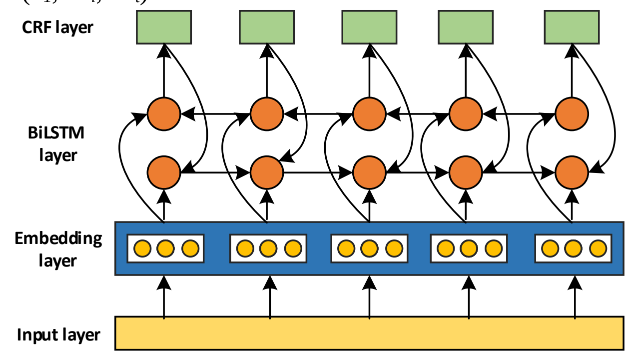

Fig. 4. Schematic of BiLSTM-CRF model

BiLSTM combined with conditional random field mainly solves the validity problem of sequence labeling. Although BiLSTM can also obtain the labels of input utterances, it cannot get the dependency relationship between the labels and, therefore cannot accurately obtain whether the labels are valid. The conditional random field calculates and learns the constraint relationship between the sequence labels from the training data, and then it can ensure the accurate validity of the predicted labels. As shown in Fig. 4, the BILSTM-CRF neural network model is composed of four parts together: the input layer, the vector layer, the BiLSTM layer, and the CRF layer. First, a sentence is input into the input layer. Then the vector layer converts each character into a low-dimensional continuous character vector and inputs it to the BiLSTM network, which obtains the sentence context semantic information vector through the unique gating mechanism of BiLSTM, and finally inputs the vector containing the sentence context semantic information to the CRF layer for decoding, calculates the label probability of each character through the conditional random field, and uses the calculated optimal label sequence as the result of model recognition.

C. *Multi-Headed Attention Layer*

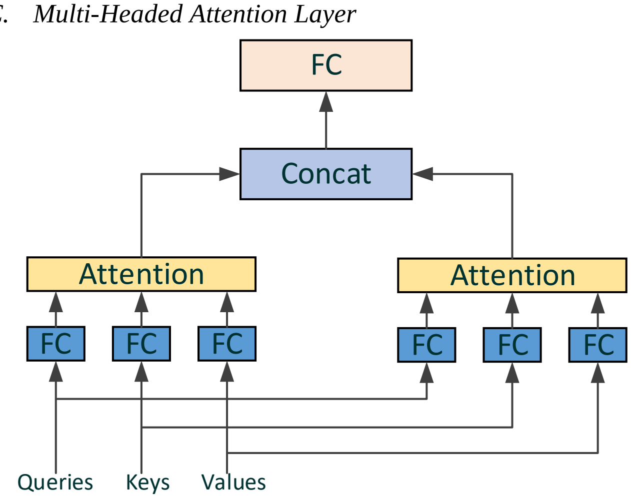

Fig. 5. Structure of multi-head attention

The multi-head attention mechanism is a variation from the self-attention mechanism, whose main feature is the ability to run multiple attention mechanisms simultaneously and compute the scaled dot product attention for each head separately, then stitch and linearly transform the output into the standard dimensions[7].Intuitively, multiple attention heads allow to focus on parts of the sequence in different ways; given a query *Q*, key *K*, and value *V*, they are then transformed into subqueries, subkeys and subvalues by computing the scaled dot product attention independently and finally, splicing each attention head and combining it with the final weight matrix to form a multi-headed attention mechanism, which is schematically shown in Fig. 5.

The multi-headed attention mechanism has been widely used in several fields. Its computational procedure is as follows:

Step 1 A linear mapping is performed for *Q*, *K*, and *V*

$$Q' = Q * W_j^Q \tag{5}$$

$$K' = K * W_j^K \tag{6}$$

$$V' = V * W_j^V \tag{7}$$

where: $W_j^{\{Q,K,V\}}$ is the weight matrix obtained by training, and $Q=K=V=H$.

Step 2 Compute the scaled dot product attention

$$M_j = Softmax\left(\frac{Q'K'^T}{\sqrt{d_k}}\right)V' \tag{8}$$

Where: Q' is the vector after linear mapping of the query matrix Q, K' is the vector after linear mapping of the query matrix K, $1 / \sqrt{d_k}$ is the scaling factor introduced to prevent the vector dot product from being too large, and d_k represents the vector dimension of V, K'.

Finally, the V' after linear mapping of the value matrix is weighted and summed according to the obtained weight coefficients.

Step 3 Calculate the multi-headed attention

$$M = Concat(M_1, \ldots, M_j, \ldots, M_h) \tag{9}$$

Where: *M* is the result of linear mapping after cycling one and two steps many times.

D. CRF Layer

The CRF layer obtains the dependencies between entity labels by introducing a state transfer matrix to improve named entity recognition, and the calculation process is shown below.

Step 1 Given a product title text statement sentence $(s_1, s_2, \ldots, s_n)$, use CRF to calculate the probability of all its entity label sequences.

$$p_{(y|x)} = \frac{1}{Z(X)} exp(\sum_{i=1}^{n} T_{y_{i-1},y_i} + \sum_{i=1}^{n} M_{i,y_i}) \tag{10}$$

where T_{y_{i-1},y_i} denotes the score from entity label y_i; M_{i,y_i} is the matrix element of the output of the multi-headed attention mechanism layer, which denotes the probability that the *i* word in the product title text statement is entity label y_i; and $Z(X)$ is the normalization factor.

$$\mathrm{Z(X)} = \sum_{y \in Y} \exp(\sum_{i=1}^{n} T_{y_{i-1},y_i} + \sum_{i=1}^{n} M_{i,y_i}) \tag{11}$$

Where: y is the real entity label; *Y* is the sequence of all possible entity labels.

Step 2 The Viterbi algorithm[8]is used to determine the most likely label sequences and use them as the final annotation results for named entity recognition.

IV. EXPERIMENTAL

To verify the effectiveness of the proposed method on the task of Chinese named entity recognition in e-commerce, evaluation metrics such as accuracy (P), recall (R), and F1 value are introduced. The F1 value is representative and is a comprehensive index considering various situations. These three evaluation metrics have great authority in named entity recognition and domain term extraction.

$$P = \frac{n}{M} \times 100\% \tag{12}$$

$$R = \frac{n}{N} \times 100\% \tag{13}$$

$$F1 = \frac{2PR}{P + R} \times 100\% \tag{14}$$

Where: *n* is the number of entities whose predictions are true and correct, *M* is the number of entities whose predictions are accurate, and *N* is the number of entities in the expected results.

A. Construction of data set.

To verify the effectiveness of the proposed model in Chinese named entity recognition in the e-commerce domain, the E-Commercial NER Dataset of the Tianchi competition platform is selected as the data for this experiment. The original corpus of the dataset is derived from the titles of Taobao product texts, and a total of 4 major categories and 9 minor entity categories are labeled, as shown in Table 1.

TABLE 1 Label type and number

Grouped Type	**Entity Type**	**Entity numbers**
PATTERN	Model Type	2173
PRODUCT	Product Description	5506
	Core Product	21958
BRAND	Brand Description	331
	Core Brand	3430
MISC	Location	1893
	Person	367
	Literature	814
	Product Specification	2732

B. Experimental settings

The paper continuously adjusts the parameter settings in E-Commercial NER Dataset through the method of control parameters through several comparison experiments, and summarizes the experimental parameter scheme with the best performance of the model, whose settings are shown in Table 2 .

.

TABLE 2 Super parameter setting

Parameter	Value
Vector dimension	100
Learning rate	0.001
Iteration	100
LSTM dimension	128
LSTM hidden dimension	128
Attention weight dimension	128
Dropout	0.5
BatchSize	20
Head number	8
Optimizer	ADAM

C. *Experimental Results*

a) effect of different learning rates on the performance of the BiLSTM-MAM-CRF model

The ADAM optimizer is a comprehensive consideration of the gradient mean and uncentered variance for step update, which has mathematical solid properties. To prevent the learning rate from being too high or too low to affect the loss function convergence, a set of different learning rates 0.0002,0.0005,...,0.001,...,0.005 is set to conduct experiments, and the variation of F1 values of the BiLSTM-MAM-CRF model is shown in Fig. 6.

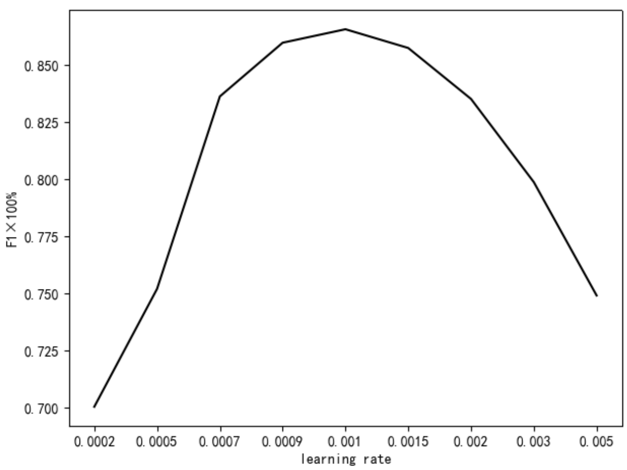

Fig. 6. F1 values with different learning rates

It can be seen that when the learning rate is in the interval [0.0002,0.0007], the F1 value becomes an increasing trend, i.e., the learning rate is positively correlated with the F1 value, and when the learning rate is in the interval [0.0007,0.0015], the F1 value first rises and then falls, and at Ir=0.001, the F1 achieves the maximum value of 86.57%. The best effect of training the model proposed in this paper at this learning rate, and when When the learning rate is in the interval [0.0015,0.005], the F1 value becomes a decreasing trend, i.e., the learning rate is negatively correlated with the F1 value, and the training effect gradually becomes worse.

b) effect of different learning rates on the performance of BiLSTM-MAM-CRF models.

The control parameter method was used to set the optimizer as ADAM and Ir=0.001 to observe the effect of different attention mechanism heads on the F1 value of BiLSTM-MAM-CRF model. The results are shown in Fig. 7:

It can be seen that when the number of multi-headed attention heads is from 0 to 8, the F1 value is on an increasing trend, and when the number of heads is from 8 to 12 the F1 value tends to be smooth with slight fluctuation in change, so the choice of The turning point 8 which tends to be soft is the most suitable.

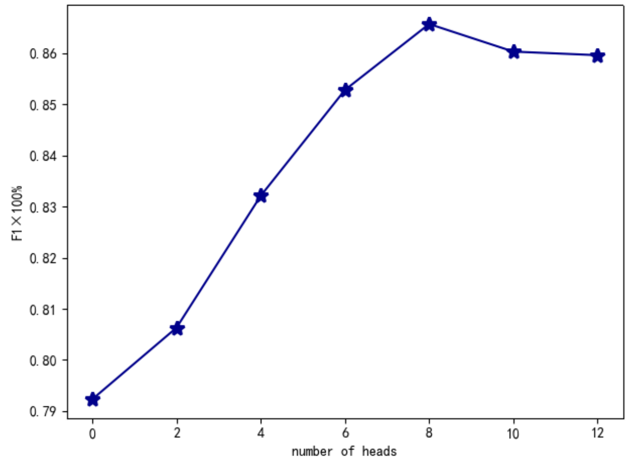

Fig. 7. F1 values for different multi-headed attention mechanism heads

c) effect of different learning rates on the performance of BiLSTM-MAM-CRF models.

To verify the superiority of the BiLSTM-MAM-CRF model for Chinese named entity recognition task in the e-commerce domain, six sets of comparison experiments were done for analysis, namely BiLSTM-CRF, BiLSTM-IDCNN-CRF, Albert-BiLSTM-IDCNN-CRF, BiLSTM-MAM-IDCNN-CRF, Albert-BiLSTM-MAM-CRF,BiLSTM-MAM-CRF experimental results are shown in Table 3.

TABLE 3 Comparison of experimental results(%)

Model	P	R	F1
BiLSTM-CRF	85.06	85.18	85.12
BiLSTM-IDCNN-CRF	85.56	85.78	85.66
Albert-BiLSTM-IDCNN-CRF	77.70	76.70	77.20
BiLSTM-MAM-IDCNN-CRF	84.83	85.59	85.21
Albert-BiLSTM-MAM-CRF	82.77	79.39	81.04
BiLSTM-MAM-CRF	86.16	86.98	86.57

From the experimental results, by comparing the BiLSTM-IDCNN-CRF with BiLSTM-CRF with the addition of the inflated convolutional neural network, the accuracy rate is improved by 0.5%, the recall rate is improved by 0.6%, and the F1 value is increased by 0.54%. Thus it can be seen that the addition of the inflated convolutional neural network to the BiLSTM-CRF model can effectively improve the product title text. The accuracy rate and F1 value of entity recognition can be effectively improved. By comparing the experimental results of BiLSTM-IDCNN-CRF with Albert-BiLSTM-IDCNN-CRF and BiLSTM-MAM-CRF with Albert-BiLSTM-MAM-CRF, we can see that the normal word vectors perform better than dynamic word vectors in the online shopping title text named entity recognition task, with F1 values improved by 8.46% and 4.17%. By comparing the experimental results of BiLSTM-MAM-CRF and BiLSTM-IDCNN-CRF, it can be seen that the model with a multi-headed attention mechanism and inflated convolutional neural network is added to BiLSTM-CRF, respectively. The model with multi-headed attention mechanism has better performance with 1.1% increase in accuracy, 1.8% increase in recall, and 1.45% increase in F1 value. By comparing the experimental results of BiLSTM-MAM-IDCNN-CRF and BiLSTM-MAM-CRF, it can be seen that adding both the multi-headed attention mechanism and the expanded convolutional neural network to BiLSTM-CRF, the recognition effect is not as good as the BiLSTM-MAM-CRF model, with a 1.33% decrease in accuracy, 1.39% decrease in recall, and a decrease in F1 value of 1.36%.

In summary, the method proposed in this paper, which integrates multi-headed attention mechanism and bidirectional LSTM, can not only effectively obtain the local features and global features of the sequence but also can mine the hidden information contained in the keywords in the sequence, so that the performance of the model can be significantly improved. Although the computational volume and complexity of the method in this paper are increased compared with the traditional methods, it effectively solves the problems of information loss and incomplete acquisition of sequence feature information caused by the insufficient dependency between different word vectors in the commodity title text. It shows better results on the experimental data set.

V. CONCLUSION

To summarize, this paper proposes a named entity recognition model for the e-commerce domain, namely BiLSTM-MAM-CRF model, which combines multi-headed attention mechanism to address the problems of low entity recognition rate and accuracy due to insufficient information acquisition of sequence features and weak dependency relationship between word vectors in e-commerce domain. The model achieves the goal of entity recognition rate improvement by applying the multi-headed attention mechanism based on bi-directional LSTM, and achieves an accuracy rate of 86.16%, a recall rate of 86.98%, and an F1 value of 86.57% on accurate experimental data, with solid robustness. The next step will be to optimize the practical method based on increasing the empirical corpus to make the entity recognition results more accurate and effective and make the model have better generalization performance.

ACKNOWLEDGMENT

This work was supported in part by Autonomous Region Innovation a (Talents, Bases) Construction Special-Natural Science Plan (Special Training of Ethnic Scientific and Technological Talents) Project (2022D03001), in part by National Natural Science Foundation of China (No. 61662081), and in part by National Social Science Foundation of China (No. 14AZD11).

REFERENCES

[1] Yan S, Chai J, Wu L. Bidirectional GRU with Multi-Head Attention for Chinese NER[C]//2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC). IEEE, 2020: 1160-1164.

[2] Huang W, Zhang J, Xiao Y, et al. Named Entity Recognition in Chinese Judicial Domain Based on Self-attention mechanism and IDCNN[C]//2020 8th International Conference on Digital Home (ICDH). IEEE, 2020: 51-56.

[3] Luo L, Yang Z, Yang P, et al. An attention-based BiLSTM-CRF approach to document-level chemical named entity recognition[J]. Bioinformatics, 2018, 34(8): 1381-1388.

[4] Yang Z, Yang D, Dyer C, et al. Hierarchical attention networks for document classification[C]//Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies. 2016: 1480-1489.

[5] Qiu Zenghui,He Mingjie,Lin Zhengkui. Deep learning-based named entity recognition method for online shopping reviews[J]. Computer Engineering and Science,2020,42(12):2287-2294.

[6] Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging[J]. arXiv preprint arXiv:1508.01991, 2015.

[7] An Y, Xia X, Chen X, et al. Chinese clinical named entity recognition via multi-head self-attention based BiLSTM-CRF[J]. Artificial Intelligence in Medicine, 2022, 127: 102282.

[8] Viterbi A J. A personal history of the Viterbi algorithm[J]. IEEE Signal Processing Magazine, 2006, 23(4): 120-142.

[9] Yu B, Wei J. IDCNN-CRF-based domain named entity recognition method[C]//2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT. IEEE, 2020: 542-546.

[10] Zhang Xiang-wei.，Li Zhi. Named entity recognition of Chinese electronic medical records based on multir feature fusion[J]. Software Guide, 2017. 16(2):128-131. (in Chinese)

[11] Li P H, Fu T J, Ma W Y. Why attention? Analyze BiLSTM deficiency and its remedies in the case of NER[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(05): 8236-8244.

[12] Goutte C, Gaussier E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation[C]//European conference on information retrieval. Springer, Berlin, Heidelberg, 2005: 345-359.

[13] Wen M, Vasthimal D K, Lu A, et al. Building large-scale deep learning system for entity recognition in e-commerce search[C]//Proceedings of the 6th IEEE/ACM International Conference on Big Data Computing, Applications and Technologies. 2019: 149-154.

[14] Joshi M, Hart E, Vogel M, et al. Distributed word representations improve ner for e-commerce[C]//Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing. 2015: 160-167.

[15] Wang M, Hu X, Xie P, et al. Automatic Construction of a Domain-specific Knowledge Graph for Chinese Patent Based on Information Extraction[C]//2021 International Conference on Management Science and Software Engineering (ICMSSE). IEEE, 2021: 1-8.

[16] Cheng X, Bowden M, Bhange B R, et al. An end-to-end solution for named entity recognition in ecommerce search[J]. arXiv preprint arXiv:2012.07553, 2020.