IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 26, 2019, accepted June 29, 2019, date of publication July 31, 2019, date of current version October 2, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2932136

Combining the Attention Network and Semantic Representation for Chinese Verb Metaphor Identification

DONGYU ZHANG[1,2], HONGFEI LIN[2], XIKAI LIU[2], HETING ZHANG[1], AND SHAOWU ZHANG[2]
[1]School of Software, Dalian University of Technology, Dalian 116620, China
[2]School of Computer Science and Technology, Dalian University of Technology, Dalian 116620, China

Corresponding author: Hongfei Lin (hflin@dlut.edu.cn)

This work was supported in part by the NSFC Programs under Grant 61602079 and Grant 61632011, and in part by the Ministry of Education under Grant 16YJCZH141.

ABSTRACT Metaphor is the central issue of language and thinking. Metaphor identification plays a significant preliminary role in the field of machine translation, reading comprehension, and automatic summarization, making it a focus of natural language processing. Recently, research into Chinese verb metaphor identification has become a widespread concern. The main problem is that the usage of semantic resources is relatively simple, and there is a lack of deep semantic support. Therefore, this paper proposes a word representation method suitable for metaphor classification tasks, which combines the traditional word vector with structural information from the Synonym Thesaurus, so that the word vector can contain the abstraction degree of the word in the metaphor. On this basis, we propose a verb metaphor attention network based on subject–predicate and verb–object relationships, which gives full consideration to the global syntactic information when we perform LSTM encoding and calculate the weight of each word. At the same time, it can facilitate the understanding of literalness and non-literalness. The experimental results show that the identification effect improves on the existing results, indicating that word representation combining semantic resources and attention network can improve the verb metaphor identification performance.

INDEX TERMS Attention network, metaphor identification, metaphor dataset, semantic representation.

I. INTRODUCTION

Metaphor is the central problem of language and thinking, and it is a common phenomenon in human language, in that that there is one metaphor in every three sentences in daily life [1]. In the Internet age, lots of electronic texts have emerged, which has made metaphor detection the focus of natural language processing. The effect of metaphor detection directly affects related tasks of natural language processing, such as machine translation, reading comprehension, and automatic summaries. Therefore, metaphor detection is gaining more and more attention from scholars as the basic job in metaphor processing. Moreover, verb metaphors appear very frequently in texts [2], so they deserve more attention.

Most scholars who study Chinese verb metaphors are from the linguistics study field, and they consider the problem mainly from a qualitative view, rather than using quantitative analysis [3]. The text processing should be automatic, as there are extensive Internet resources and limited human resources. For machine translation, reading comprehension, Q&A systems, and automatic summary systems, the understanding of verb metaphor is a vital technical link.

This task can refer to verb metaphor detection in English, but there are important differences in semantic expression, history, and culture between the Chinese and the English. On the basis of English metaphor detection, Chinese metaphor detection needs to combine more Chinese language characteristics and to use the abundant online resources and advanced and effective machine learning methods fully, especially the deep learning method, to produce better performance in Chinese verb metaphor detection.

Focusing on Chinese verb metaphor detection, this paper utilizes the current Chinese semantic resource – the Synonym Thesaurus [4] and the verb metaphor dataset. It constructs a verb metaphor attention network based on subject-verb-object information, starting with word representation, which

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

makes the system consider global syntax information when giving sentences LSTM encodes and computing the weight of each word. On a given Chinese verb metaphor dataset, the performance of this model is better than the current classification model, which is based on hand-crafted features and the mainstream deep learning model. Our contributions are as follows.

(1) We propose a word representation method suitable for Chinese metaphor classification tasks. On the basis of the original word vector, we fused the structural information of a Synonym Thesaurus so that the word vector could contain information such as the abstractness of the words concerned in metaphors.

(2) We constructed a verb metaphor attention network based on subject-verb-object information. Therefore, when we used LSTM to encode sentences and we calculated the weight of each word, we could fully consider the global subject-verb-object information.

In this paper, the second section introduces related research work, the third section explains Chinese verb metaphor detection, and the fourth section describes an experiment. The final section concludes this paper.

II. RELATED WORK

The study of English verb metaphors started some time ago, and it has had some good results. There are two main ideas: one is based on English semantic resources and the abstractness of words; the other is based on machine learning.

Many researchers identify metaphors on the basis of an abstract degree computed from semantic resources. In reference [5], an unsupervised method of metaphor detection was presented. They cluster metaphors that have similar syntactic structures based on a small, manually annotated dataset from the corpus. It extracts metaphoricity from verb and noun information. Reference [6] shows that the metaphoricity of words is related to the level of abstractness, and they design an algorithm according to this idea to classify words as literal or metaphorical. They use their algorithm to calculate the level of abstractness of words as a feature, and they apply supervised learning with logistic regression to classify and achieve state-of-the-art performance on related datasets. Reference [7] shows the usage of WordNet resources, compute affinity and similarity between words and clustered results. They parse sentences to get subject-predicate pairs and predicate-object pairs, combining existing metaphor resources and conceptual distance within pairs to complete metaphor identifications.

The method of verb metaphor detection based on machine learning mainly applies in LDA topic modeling and classifier and neuron network methods. Reference [8] shows a system using LDA topic modeling that supports low-resource languages and no labeled data to identify metaphors in unlabeled text. Her approach looks for overlapping semantic concepts by LDA topics, which can find cross-language metaphors. Reference [9] focuses on the regularities of behavior of verbs, and she attempts various way of defining semantic generalization for verbs to detect metaphors. She evaluates her ideas with logistic regression and random forest classifiers, and she has discovered the value of the lemma unigram. Reference [10] proposes using a neural network combined with word embedding trained on a large corpus to detect metaphor, inputting the target word and the surrounding words to multilayer perceptrons to obtain semantic and structure information, which is important, and to push the development of research into metaphor detection. Reference [11] presents a better method to get the level of abstractness of words automatically by using vector representations and to apply them to multi-words and words with multiple senses. At the same time, they published their resources and the metaphor detection performance of their work about abstractness.

For Chinese verb metaphor detection, scholars have made some progress by applying natural language understanding and machine learning methods to the research. On basis of studies of English verb metaphors, the current methods of studying can be divided into two classes: one is based on semantic resources and abstractness; the other is based on machine learning. The first is based on Chinese semantic resources such as the Synonym Thesaurus [4] and HowNet [12] to compute the level of abstractness of words. Reference [13] proposes a method of Chinese metaphor detection based on HowNet and Verbnet, using verb information such as the semantic and subject role from Verbnet to identify mismatching between the actor and the subject of the verb. She extracted 20 sentences from the reader corpus to experiment, with favorable results.

The second approach uses machine-learning methods, divided into unsupervised learning and supervised learning. The unsupervised learning mainly includes a clustering algorithm and a topic model. In [14], Jia uses an unsupervised method to recognize metaphors with selection preferences and source domain knowledge. Jia extracts selectional preferences from the corpus by clustering and choosing source domain candidates based on source domain knowledge. Then Jia considers whether the headword of the verb is a candidate for a metaphor. Reference [15] focuses on Chinese verb metaphors at the sentence level, and they propose a detection algorithm based on a topic model. They apply LDA, and they use the distribution of topics in the sentence as a feature, referring to machine-learning methods, to identify verb metaphors. The performance develops after considering topic annotation. In addition, supervised learning needs a certain number of annotated examples as a training dataset when applied to metaphor detection; then using machine learning and methods of natural language processing becomes possible. Reference [16] shows the process of utilizing conditional random fields and the maximum entropy model to identify verb metaphor matches. They added two semantic resources – the Synonym Thesaurus and HowNet – to their algorithm, choosing 10 Chinese verbs from the People's Daily corpus with some success.

If the data size of Chinese verb metaphors for study is small, depending on the available semantic resources and the lack of word representations and deep learning method applications, we use a deep learning method based on the attention mechanism for a large corpus. In the presentation layer we propose a way of expressing words as vectors for metaphor-classification tasks, on the basis of original vector fusing structure information from the Synonym Thesaurus, which contains abstractness information for words that is related to metaphor detection. Then we build an attention network of verb metaphors based on subject-verb-object information, and we let the system consider global information when transferring sentences to LSTM codes and computing the weights of each word. In a given Chinese verb metaphor dataset, the performance of this model is better than the current classification model, which is based on manual features and the mainstream deep learning model. It improves the precision of Chinese verb metaphor detection, and it lays a good foundation for other types of research such as machine translation, reading comprehension, Q&A systems, and automatic summary.

III. VERB METAPHOR DETECTION MODEL

A. METAPHOR DETECTION BASED ON THE ABSTRACT DEGREE IN THE SYNONYM THESAURUS

The original intention of the Synonym Thesaurus [4] is to provide more synonym words, which is helpful for creation and translation. This dictionary contains not only synonyms of a word, but also a certain number of similar words-namely, related words in a broad sense. The class code of each word in the Synonym Thesaurus contains all its information. For example, the first letter of the code represents the large class to which the word belongs, the second letter represents the middle class, and the last letters represent the small class.

Some research has shown that information on word abstractness plays an important role in metaphor detection. For this paper, the abstractness information came from the Synonym Thesaurus [4]. It contains 12 large categories of vocabularies, of which Class C and Class D represent abstract vocabularies. If the subject-verb-object contains abstract words, the abstractness of the sentence can be determined. For words in the subject-verb-object that do not appear in the Synonym Thesaurus, the pre-trained FastText model can calculate the similarity between the words and similar words in the Synonym Thesaurus. Once the first K closed words are obtained, then the most frequently occurring major category is classified as Synonym Thesaurus of the word, determining whether the word is abstract. Sentences that contain abstract words in the subject-verb-object will be identified as metaphorical sentences.

In fact, there are lots of synonyms in every language: this association is important, and it is usually sorted and processed. We use the Synonym Thesaurus, and we build a Chinese thesaurus based on it, which contains not only synonyms, but also a certain number of similar words that will help in creation and translation. The code for each word contains all its information. The thesaurus's information is structured into a tree and saved with lots of nodes, including multiple words. Therefore, the node vectors can help in finding abstraction information on words for a neural network model. Specifically, if the word appears in the Synonym Thesaurus, we can extract its abstract information directly to identify metaphors; if the word does not appear in the Synonym Thesaurus, we apply a pre-trained Fast Text model to compute the similarity between it and each word in the Synonym Thesaurus. We can get the K most similar words and choose the most frequent category of the K words as the target word's category to get abstraction information and to detect metaphors in sentences.

B. VERBAL METAPHORICAL ATTENTION NETWORK BASED ON THE REPRESENTATION OF NODES IN THE SYNONYM THESAURUS

We adopt the learning method of word representation to learn the structural information in the Synonym Thesaurus. In contrast to the direct application of abstract categories in the Synonym Thesaurus to metaphor detection, this method not only allows the model to learn information about abstract categories, but it also combines all kinds of information in the Synonym Thesaurus into the semantic representation process of words. In the process of encoding sentence semantics, the neural network makes more use of information in the Synonym Thesaurus to improve the effect of metaphor detection.

C. NODE REPRESENTATION LEARNING BASED ON THE SYNONYM THESAURUS

In this paper, the method of node representation learning based on the Synonym Thesaurus (Synonyms2Vec) combines the Neural Probabilistic Language Model with the Synonym Thesaurus, as shown in Figure 1. For the central word Wt, all the parent nodes of the word are $\{ S_{t1},S_{t2},S_{t3},S_{t4},S_{t5}\}$ in the Synonym Thesaurus. Then, under the condition of a given central word and its parent nodes, using a skip-gram structure predicts the context words: $(W_{t-2},W_{t-1},W_{t+1},W_{t+2})$. During the process of model training, the sliding window size is preset, and the maximum likelihood estimation is taken as the optimization goal. When the sentence $D = (W_1,W_2,W_3,\cdots,W_M)$ is given, the target function is shown below:

$$L = \frac{1}{M}\sum_{i=1}^{M}\sum_{-k\leq c\leq k, c\neq 0}(\log p(W_{i+c}|W_i)) + \sum_{j=1}^{N_i}\log p(W_{i+c}|S_{ij}) \tag{1}$$

M is the number of words in the training corpus, N_i is the number of parent nodes in the Synonym Thesaurus for the i word, and the sliding window size is 2k.

Synonyms2Vec is designed to learn the tree structure information from the Synonym Thesaurus and to provide a vector representation of the position information about words in the Synonym Thesaurus. To train faster, we adopt the optimization method in Word2Vec [17], and we used the negative

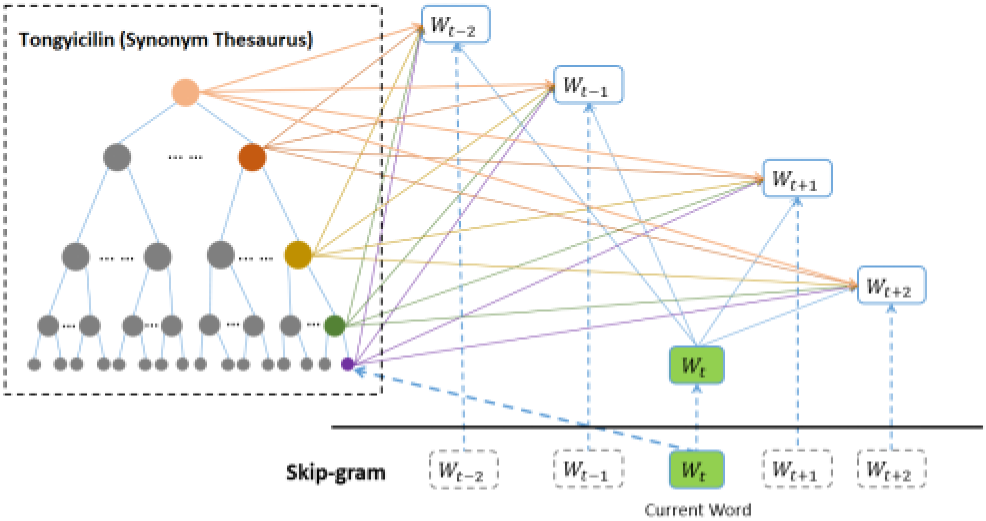

FIGURE 1. W_t **is the central word and** (W_{t-2},W_{t-1},W_{t+1},W_{t+2}) **is the model structure (Synonyms2Vec) of context words.**

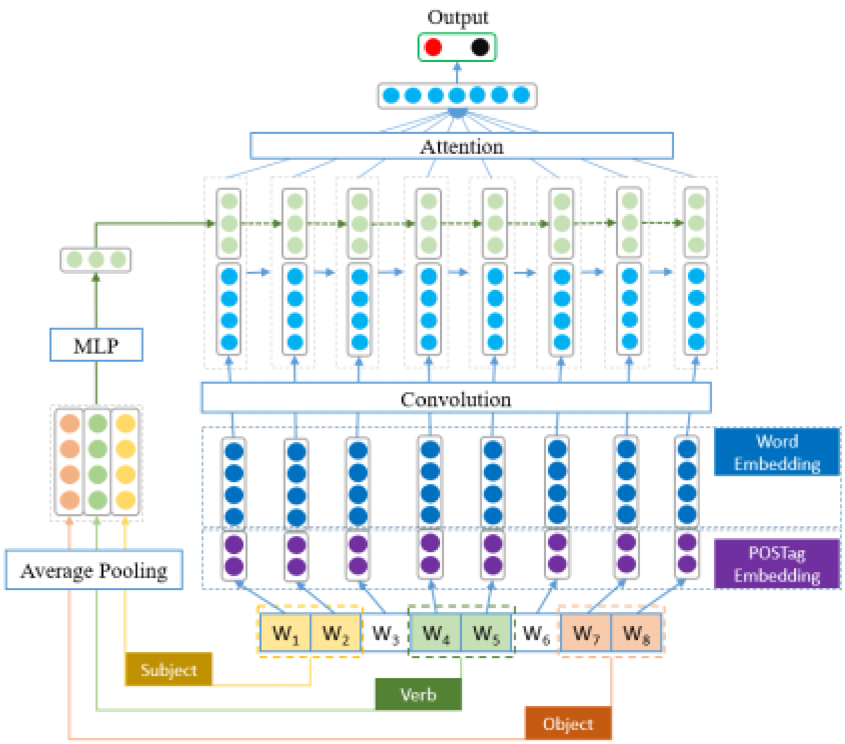

FIGURE 2. **Verb metaphor attention network.**

sampling method to approximate the Softmax result in the maximum likelihood estimation. The parameters in the model were updated using SGD and a backpropagation algorithm. Because the tree branch of words in the Synonym Thesaurus is of great help for metaphor detection, the node vector obtained by using Synonyms2Vec method is conductive to learning the abstract degree of words in the neural network model to help to complete the task of metaphor detection.

D. VERBAL METAPHOR ATTENTION NETWORK

On the basis of the node vector obtained from the pre-training of Synonyms2Vec, we proposed an attention network structure for verb metaphors. As shown in Figure 2, this involves taking sentences. For example, $\{Tag_1, Tag_2, \cdots, Tag_8\}$ is obtained by analyzing the part of speech of each word in the sentence, and Tag_1 means the POS tagging of W_i. In the neural network model, part-of-speech vector mapping is carried out. First, a possible part-of-speech number L is calculated, the dimension K of part-of-speech vector is set manually as a hyperparameter, and the matrix Q of L*K is initialized in the network. According to the part-of-speech number, the corresponding row vectors are taken from the matrix Q as the part-of-speech vectors written as V_{tag}. In the result of the pre-training using the Synonyms2Vec model, the node vector, written as V_{word}, corresponding to each word in the sentence is queried. The vector spliced by V_{word} and V_{word} passes the convolution kernel, and its length is 1. The vector of each word written as V_{conv} is obtained by a convolution operation with Relu as its activation function.

To make full use of the structure information of subject-verb-object in the sentence, we conducted a separate vectorization of subject-verb-object in the model. The subject of the sentence given is $\{W_1,W_2\}$, the verb is $\{W_4,W_5\}$, and the object is $\{W_7,W_8\}$. First, the node vector corresponding to each word is obtained. Then, the three phrases of subject-verb-object are average poding. Thus, the vector representation of subject-verb-object is obtained. The subject vector is written as V_{sub}, the verb vector is written as V_{verb}, the object vector is written as V_{obj}. Finally, after splicing the three vectors, the concatenated vector representation of subject-verb-object, written as V_{svo}, is obtained by passing the full connection layer. The formula is as follows:

$$V_{svo} = relu(W_{svo}^T[V_{sub}; V_{verb}; V_{obj}] + b_{svo}) \tag{2}$$

To obtain the global subject-verb-object information for every word in the sentence, the subject-verb-object vector is spliced onto the convolution word vector of each word. In this paper, the LSTM layer and self-attention structure are used to encode the word vector of the sentence and the subject-verb-object information. Finally, we get the metaphorical vector of the sentence. The calculation process of the LSTM layer is as follows:

$$x_t = V_{convt} \tag{3}$$

$$\begin{bmatrix} i_t \\ f_t \\ o_t \\ l_t \end{bmatrix} = \begin{bmatrix} sigmoid \\ sigmoid \\ sigmoid \\ tanh \end{bmatrix} W^T \cdot \begin{bmatrix} x_t \\ h_{t-1} \\ V_{svo} \end{bmatrix} \tag{4}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot l_t \tag{5}$$

$$h_t = o_t \cdot tanh(c_t) \tag{6}$$

where x_t represents the convolution vector of the t word in the sentence, W is the weight vector, which is randomly initialized and updated with the network backpropagation, i_t, f_t, and o_t are the input gate, forgetting gate, and output gate in the LSTM network structure respectively, c_t is memory unit, and h_t is the hidden layer output of the t word in the sentence. The structure allows the network to remember the information overlonger periods and to alleviate the gradient vanishing problem caused by excessive time steps. Each hidden layer added V_{svo} to the calculation process. When the model is used to calculate the output of each hidden layer and to calculate the weight of self-attention, the global subject-verb-object can be used. The calculation formula of self-attention is as follows:

$$e_t = v^T tanh(W_e^T[h_t; V_{svo}] + b_e) \tag{7}$$

$$\alpha_t = \frac{exp(e_t)}{\sum_{k=1}^{n} exp(e_k)} \tag{8}$$

$$V_{meta} = \sum_{t=1}^{n} \alpha_t h_t \tag{9}$$

where W_e,b_e, and v are initialized randomly, α_t represents the weight of the t hidden layer, and V_{meta} represents the final metaphorical vector of the sentence. V_{meta} is mapped to the output node via the full connection layer. Then, the cross-entropy between the ground truth labels and predicted labels is used as the loss function in the training process, and the optimization objective is to minimize the loss function. The calculation of the loss function is as follows:

$$L = -\sum_{i}^{N} y_i^t \log(y_i^p) + (1 - y_i^t)\log(1 - y_i^p) \tag{10}$$

where N is the number of all the training examples, y_i^t is the truth label of the *ith* training example, and y_i^p is the probability value predicted by the model.

TABLE 1. The size of the dataset.

Dataset	Size	Verb Metaphor	Noun Metaphor	Non Metaphor
Training set	4,394	2,040	2,035	319
Test set	1,100	517	507	76

TABLE 2. Distribution statistics of the dataset.

	Maximum sentence length	Average sentence length	Minimum sentence length	Vocabulary length	Verb-noun ratio
Training set	246	18.54	4	12,456	88.48% (9,052/ 10,230)
Test set	144	19.06	4	4,898	91.00% (2,366/ 2,600)
Training set+Test set	246	18.65	4	14,501	88.99% (11,418/ 12,830)

IV. EXPERIMENT

A. DATA

We used the data released publicly in the shared task of metaphor identification at the Seventeenth China National Conference on Computational Linguistics in 2018.[1] Table 1 gives the size of the dataset. Table 2 gives distribution statistics.

To verify the validity of our model, we performed a comparative experiment using our model and the state-of-art model on the constructed data set.

[1] http://ir.dlut.edu.cn/news/detail/496

TABLE 3. Model parameters.

Parameter	Value
The dimension of word vector	620
The dimension of POS	10
The number of convolution kernels	128
The dimension of the LSTM	128
The dimension of attention	128
The dimension of the full connection layer	100
Dropout	0.5
Optimization	RMSProp
The size of the word list	15,662

We completed the pre-training of word vector using the 5GB microblog corpus and the Synonyms2Vec model. The choice of the word vector dimension refers to the parameters in Table 3. The part-of-speech vector only has 10 dimensions. Because the number of parts of speech is small, the 10-dimensional vector can fully express the part-of-speech information. If the dimension is set too high, it will affect the operating speed of the model, and the training of each dimension of the vector will be inefficient. The other parameters of the model were adjusted according to the experimental results of the model on the validation set. The optimal combination of parameters in the validation set was selected. The specific parameter values are shown in Table 3.

B. COMPARATIVE EXPERIMENT

This section compares our model with the commonly used deep learning text classification model. The comparison model includes:

1) TEXTCNN MODEL [18]

CNN is the convolutional neural network model. In the text field, the model input is a word vector pre-trained by Word2Vec. The model input usually performs a one-dimensional convolution, that is, the convolution kernel width is the length of the word vector. TextCNN uses convolution kernels of various sizes, and it has better local feature extraction capabilities. In this paper, the TextCNN model is used to encode the user problem and the standard problem respectively, and then the encoded vector is spliced, and the output of the two classification is completed through the fully connected layer and the Sigmoid layer.

2) LSTM MODEL [19]

The LSTM model is Long Short-Term Memory model. It takes the pre-trained word vector as an input, and it can extract long-distance dependence semantic information for comparison with the original recurrent neural network. In this task, the LSTM model is used to participate in the encoding process of sentences, and the rest of the structure except the encoding structure is consistent with the TextCNN model.

3) LSTMATT MODEL [20]

The LSTMATT model adds a self-attention structure to the LSTM model, and it obtains the weights of the hidden layers of the LSTM. The final output vector is obtained by the weighted average of the weights of the hidden layers of the LSTM, replacing the way in which only the last hidden layer vector is outputted in the original LSTM model.

4) TAGLSTM MODEL

The TagLSTM model is a comparative model constructed by adding the information of the part-of-speech vector based on the original LSTM model.

5) METAATT-PARSER MODEL

The MetaATA-Parser model is the verb metaphor attention networks model proposed in this paper. The Harbin Institute of Technology Syntax Analyzer is used to extract the subject-predicate structure in sentences automatically.

6) METAATT-HUMAN MODEL

The MetaATT-Human model uses the subject-object structure manually labeled in the dataset instead of the syntax analyzer, and the rest is the same as the MetaATT-Parser.

We divide the data set into 80% train set, 10% validation set, and 10% test set. The train set is used to train and update the model. The optimal model parameters are selected according to the results of the verification set. The final effect of the model is evaluated on the test set. We make the model test results more stable through 10-fold cross-validation. The evaluation metrics in this paper include cross entropy loss as Loss, accuracy as Acc, Precision as P, recall as R, F-Measure, and area under curve as AUC [21]. The calculation formulas for each evaluation metrics are as follows:

$$Acc = \frac{T_p + T_n}{T_p + F_p + T_n + F_n} \tag{11}$$

$$P = \frac{T_p}{T_p + F_p} \tag{12}$$

$$R = \frac{T_p}{T_p + F_n} \tag{13}$$

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \tag{14}$$

Among these equations, T_p is the number of positive samples correctly predicted by the model, T_n is the number of negative samples correctly predicted by the model, F_p is the number of false predictions of positive samples by the model, and F_n is the number of false predictions as negative samples by the model. Because the predicted output of the model is a probability value, when calculating Acc, P, R, and F1, we need to specify a classification threshold manually. If the output probability is greater than the threshold, it is judged as a positive example, otherwise it is a negative example. The selection of this threshold will affect the final evaluation index, so this paper also uses the AUC index. The value is based on the true positive rate (TPR) as the vertical axis and the false positive rate (FPR) as the horizontal axis. The program continuously changes the threshold, draws multiple points, and forms a curve. The area under the curve is the AUC value.

TABLE 4. Evaluation results of verb grouping experiment validation set.

	Loss	AUC	ACC	P	R	F1
TextCNN	51.74	84.08	76.73	75.06	80.59	77.61
LSTM	45.82	87.19	78.08	78.95	76.70	77.76
LSTMATT	47.74	85.76	77.42	77.00	78.32	77.60
TagLSTM	47.35	85.77	77.22	75.92	80.15	77.87
MetaATT-Parser	46.28	86.96	79.09	77.95	81.14	79.51
MetaATT-Human	43.01	88.89	80.46	80.00	81.50	80.61

TABLE 5. Evaluation results of verb grouping experiment test set.

	Loss	AUC	ACC	P	R	F1
TextCNN	52.38	84.20	76.46	75.89	78.06	76.81
LSTM	47.02	87.17	77.83	79.52	75.04	77.07
LSTMATT	48.16	86.83	78.02	77.73	79.06	78.17
TagLSTM	47.36	86.19	77.92	76.51	80.85	78.52
MetaATT-Parser	46.61	86.87	78.89	78.76	79.93	79.12
MetaATT- Human	**44.45**	**88.91**	**80.18**	**79.66**	**81.57**	**80.37**

TABLE 6. Evaluation results of the randomized group experiment validation set.

	Loss	AUC	ACC	P	R	F1
TextCNN	47.95	85.90	77.79	78.36	78.95	78.57
LSTM	43.44	88.62	80.05	80.42	81.24	80.69
LSTMATT	43.63	88.05	79.61	80.65	79.43	79.94
TagLSTM	44.83	87.59	78.48	77.59	81.90	79.64
MetaATT-Human	42.11	89.02	81.08	80.33	84.19	82.12

TABLE 7. Evaluation results of the randomized group experiment test set.

	Loss	AUC	ACC	P	R	F1
TextCNN	54.33	84.23	76.18	77.09	76.95	76.84
LSTM	46.94	87.61	78.82	78.72	81.24	79.74
LSTMATT	47.58	86.63	78.38	79.81	77.81	78.66
TagLSTM	48.23	86.31	78.48	78.02	81.43	79.54
MetaATT-Parser	47.22	86.61	79.22	79.09	81.52	80.24
MetaATT-Human	**43.97**	**89.26**	**82.41**	**83.29**	**81.95**	**82.34**
MetaATT-Word2Vec	**44.90**	**89.07**	**81.96**	**82.01**	**81.24**	**81.62**

C. EXPERIMENTAL RESULTS AND ANALYSIS

The experimental results of the model in the test set are given in this section. The results of the validation set of the group experiments in units of verbs are shown in Table 4. The results of the test set are shown in Table 5.

As can be seen from Table 4 and Table 5, our model offers significant improvements on existing models in AUC, ACC, P, R, and F1 values, and in other indexes. Therefore, we can prove that when all our data are in the training set, our model can better learn the metaphor judgment of the verb in different contexts. The difference between the P value and the R value is not large, which proves that our model has a better predictive structure for data and better correlation with the data. As can be seen from Figure 2, the use of a parser to extract the subject-predicate of a sentence will result in a loss of precision, because the extracted subject-predicate object may not be completely correct. However, the result is still higher than the result of the state-of-the-art model. The reason is that the subject-predicate information is the key in the verb metaphor identification, which plays an important role in the process of semantic understanding of the whole sentence. We spliced the information of the subject-predicate into the vector representation of each word in the sentence, so that the words can fully refer to the global subject-predicate information in the process of LSTM hidden layer transmission. This proves that our model can still detect the metaphors of verbs without labeling information, and it can achieve good results. Through the comparison of MetaATT-Parser and MetaATT-Human results, it can also be verified that the method of using the component information of the subject-predicate can play a positive role in the recognition of verb metaphor.

The results of the validation set for the grouping experiments in a completely random manner are shown in Table 6, and the results of the test set are shown in Table 7.

It can be seen from the experimental results that the model effect still offers a significant improvement over randomly segmented data. When the model of this paper does not appear in the prediction training set, the detection of the verb is still very accurate, and it can learn the metaphorical representation of verbs that have not appeared in the training set. It can be seen from the experiment that randomized experiments are better than experiments with verbs grouped. The reason is that the number of positive and negative samples of each verb in the training corpus is limited, and after it is dispersed into the training set and the validation set, it is difficult for the model fully to understand the representation mode of the verb metaphor, and the network parameter update is not enough. In a randomized group experiment, the model can better compare the metaphorical representation of the same verb in different sentences. Also, the method of this paper can make up for the shortcomings of some verbs that have not appeared in the training set.

In the random group experiment, the model MetaATT-Parser, which uses the automated parser to extract the subject-predicate component, is still improved in each evaluation index compared with other mainstream deep learning models. This further proves our point of view: the subject-predicate information is key in the verb meta-identification, and it plays an important role in the process of semantic understanding of the whole sentence. Compared with Tag LSTM and LSTMATT, which do not use the subject-predicate information, the model can increase the F1score by more than 2.3%. After optimizing the results of the automated parser extraction using the subject-predicate structure labeled in the dataset, the model has been further improved.

V. CONCLUSION

Utilizing syntactic information from the dataset, we developed the labeling form of subject-verb-object, and we illustrated the usages of metaphor and non-metaphor. Aiming at the problem that the existing models are too simple to use semantic information, we propose a word vector representation method suitable for metaphor classification tasks. On the basis of the original word vector, we fused the structural information of a Synonym Thesaurus so that the word vector could contain information such as the abstractness of the words concerned in metaphors. This model not only utilizes the structural information of the Synonym Thesaurus, but it also digs into the semantic information at a deeper level to enhance word representations for metaphor detection. On this basis, we constructed a verb metaphor attention network based on subject-verb-object information. Therefore, when we used LSTM to encode sentences and we calculated the weight of each word, we could fully consider the global subject-verb-object information. In this paper, the model classification effect outperformed the existing manual features based on classification model and the mainstream deep learning model with the dataset of metaphors.

REFERENCES

[1] G. Lakoff and M. Johnson, *Metaphors we Live by*. Chicago, IL, USA: Univ. of Chicago Press, 1980, pp. 67–72.

[2] E. Shutova and S. Teufel, ''Metaphor corpus annotated for source-target domain mappings,'' in *Proc. LREC*, 2010, vol. 2, no. 2, p. 2.

[3] Y. L. Su, ''The frontier of multidisciplinary research on verb metaphor,'' *Mod. Foreign Lang.*, vol. 40, no. 5, pp. 705–714 and 731, 2017.

[4] J. J. Mei, Y. M. Lan, Y. Q. Gao, and H. X. Yin, *Chinese Thesaurus Tongyici Cilin*, 2nd ed. Shanghai, China: Shanghai Thesaurus Press, 1996.

[5] E. Shutova, L. Sun, and A. Korhonen, ''Metaphor identification using verb and noun clustering,'' in *Proc. 23rd Int. Conf. Comput. Linguistics*, 2010, pp. 1002–1010.

[6] P. D. Turney, Y. Neuman, D. Assaf, and Y. Cohen, ''Literal and metaphorical sense identification through concrete and abstract context,'' in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2011, pp. 680–690.

[7] J. Xu and J. S. Yu, ''Identifying English verb metaphors using statistical and clustering approaches,'' *Int. J. Knowl. Lang. Process.*, vol. 3, no. 1, pp. 15–27, 2012.

[8] I. Heintz, R. Gabbard, M. Srinivasan, D. Barner, D. S. Black, M. Freedman, and R. Weischedel, ''Automatic extraction of linguistic metaphors with lda topic modeling,'' in *Proc. 1st Workshop Metaphor NLP*, Atlanta, GA, USA, 2013, pp. 58–66.

[9] B. B. Klebanov, C. W. Leong, E. D. Gutierrez, E. Shutova, and M. Flor, ''Semantic classifications for detection of verb metaphors,'' in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, vol. 2, 2016, pp. 101–106.

[10] E.-L. Do Dinh and I. Gurevych, ''Token-level metaphor detection using neural networks,'' in *Proc. 4th Workshop Metaphor NLP*, 2016, pp. 28–33.

[11] M. Köper and S. S. Walde, ''Improving verb metaphor detection by propagating abstractness to words, phrases and individual senses,'' in *Proc. 1st Workshop Sense, Concept Entity Represent. Appl.*, 2017, pp. 24–30.

[12] Z. Z. Dong, ''HowNet and Chinese studies,'' *Contemp. Linguistics*, vol. 3, no. 1, pp. 33–44, 2001.

[13] H. Liu, ''Chinese verb metaphor detection based on HowNet and VerbNet,'' M.S. thesis, Dept. Comput. Sci., Xiamen Univ., Xiamen, China, 2014.

[14] Y. Jia and S. Yu, ''Unsupervised Chinese verb metaphor recognition based on selectional preferences,'' in *Proc. 22nd Pacific Asia Conf. Lang., Inf. Comput.*, 2008, pp. 207–214.

[15] Z. K. Bai, X. X. Huang, R. B. Wang, Z. Q. Chen, and X. H. Wang, ''Chinese verb metaphor recognition based on topic model,'' *Comput. Technol. Develop.*, vol. 26, no. 11, pp. 67–71, Nov. 2016.

[16] H. Y. Zhao, W. G. Qu, F. Zhang, and J. S. Zhou, ''Chinese verb metaphor recognition based on machine learning and semantic knowledge,'' *J. Nanjing Normal Univ., Eng. Technol. Ed.*, vol. 11, no. 3, pp. 59–64, 2011.

[17] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, ''Distributed representations of words and phrases and their compositionality,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 26, 2013, pp. 3111–3119.

[18] Y. Kim, ''Convolutional neural networks for sentence classification,'' 2014, *arXiv:1408.5882*. [Online]. Available: https://arxiv.org/abs/1408.5882

[19] S. Hochreiter and J. Schmidhuber, ''Long short-term memory,'' *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[20] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, ''Hierarchical attention networks for document classification,'' in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2016, pp. 1480–1489.

[21] A. P. Bradley, ''The use of the area under the ROC curve in the evaluation of machine learning algorithms,'' *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159, 1997.