IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 15 February 2023, accepted 16 March 2023, date of publication 21 March 2023, date of current version 28 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3260139

RESEARCH ARTICLE

TransH-RA: A Learning Model of Knowledge Representation by Hyperplane Projection and Relational Attributes

YONGKANG WANG[1], AISHAN WUMAIER[2], (Member, IEEE), WEIWEI SUN[2], YANGXIN LIU[2], AND JIANGTAO HE[2]

[1]School of Software, Xinjiang University, Ürümqi 830091, China
[2]College of Information Science and Engineering, Xinjiang University, Ürümqi 830046, China

Corresponding author: Aishan Wumaier (Hasan1479@xju.edu.cn)

This work was supported by the Autonomous Region Natural Science Foundation Joint Fund Project, Research on Xinjiang Tourism Sentiment Analysis Technology Based on Deep Learning, under Grant 2021D01C081.

ABSTRACT The TransE model plays a key role in dealing with data sparsity and promotes the development of knowledge graphs completion. However, TransE has some difficulties in dealing with one-to-many, many-to-one, many-to-many and transmission relationships. In order to solve this problem, this paper proposes a knowledge representation learning model based on hyperplane projection and relational attributes, namely TransH-RA. First of all, we introduce the idea of hyperplane projection based on the TransE model, this idea is inspired by TransH, which makes different entities have different roles in a specific relationship, thus reducing the constraints of TransE translation rules, and map the head entity **h** and tail entity **t** to the plane of special relation **r**; Secondly, considering that it is not easy to identify different similar entities, the neighborhood information of entities is added to learn the neighborhood of entities around different entities; Then, in order to further strengthen the ability to deal with complex relationships, attribute features of relationships are added and attribute knowledge is embedded; Eventually, during the training of the model, the probability method is chosen to replace the head and tail entities. Link prediction experiments are conducted on the public datasets FB15K and WN18, and the triple classification experiments on the datasets WN11, FB13 and FB15K are carried out to analyze and verify the effectiveness of the proposed method. The evaluation results show that our method achieves state-of-the-art performance on MeanRank, Hits@10 and ACC indicators compared with TransE and TransH.

INDEX TERMS Knowledge graphs, knowledge representation, hyperplane projection, relational attributes, link prediction, triple classification.

I. INTRODUCTION

Knowledge graphs [1], [2] belong to a kind of semantic network. This semantic network [3] consists of entity nodes and relationships, which can be represented by the triples (denoted (h,r,t)). In the triple (h,r,t), h represents the head entity, r represents the relationship, and t represents the tail entity. Since the knowledge graphs were put forward, many large-scale knowledge bases have been built, such as WordNet, Knowledge Cube, Knowledge Center and so on. These large-scale knowledge bases greatly facilitate people's search [4] for knowledge. However, the large-scale knowledge bases adopt the way of triples (denoted (h,r,t)) to represent knowledge, it is more complex to calculate, and affects the efficiency of calculation. It does not make full use of the semantic information of entities in different spaces, which brings some difficulties to knowledge reasoning. Therefore, scholars gradually turn their attention to knowledge representation learning [5], hoping to solve the problems faced by the knowledge base [6] through knowledge representation learning. In recent years, because of the continuous break-through of machine learning technology, it promoted the development

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

TABLE 1. The number of parameters and time complexity of the knowledge representation learning model.

Model	#Parameters	# Operations (Time complexity)
Unstructured	$\mathrm{O}(E_n u)$	$\mathrm{O}(A_n)$
SE	$\mathrm{O}\left(E_n u+2R_n v^2\right)(u=v)$	$\mathrm{O}\left(2u^2 A_n\right)$
SME(linear)	$\mathrm{O}(E_n u+R_n v+4up+4p)(u=v)$	$\mathrm{O}(4upA_n)$
SME (bilinear)	$\mathrm{O}(E_n u+R_n v+4ups+4p)(u=v)$	$\mathrm{O}(4upsA_n)$
LFM	$\mathrm{O}\left(E_n u+R_n v^2\right)(u=v)$	$\mathrm{O}\left(\left(u^2+u\right)A_n\right)$
NTN	$\mathrm{O}\left(E_n u+R_n\left(v^2 s+2vs+2s\right)\right)(u=v)$	$\mathrm{O}\left(\left(\left(u^2+u\right)s+2up+p\right)A_n\right)$
TransE	$\mathrm{O}(E_n u+R_n v)(u=v)$	$\mathrm{O}(A_n)$
TransH	$\mathrm{O}(E_n u+2R_n v))(u=v)$	$\mathrm{O}(2uA_n)$
TransH-RA(this paper)	$\mathrm{O}(E_n u+2R_n v+u)(u=v)$	$\mathrm{O}(2uA_n+u)$

of knowledge representation learning to some extent. Knowledge representation [7] learning attempts to project entities and relations into low dimensional dense vector space [8]. The prediction of unknown entities and relations is translated into a calculation of the distance between two entities, the smaller the distance is, the more similar the entities are. This makes it possible to use a lot of deep learning [9], [10], machine learning [11], and mathematical knowledge in the computation of downstream tasks, furthermore, this promotes the development of knowledge reasoning [12], [13], personalized recommendation systems, and intelligent question answering [14]. For this reason, scholars have carried out a lot of research work and made remarkable achievements, such as TransE [15], [16], TransH [17], TransR [18], STransH [19], TransA [20] and so on. In these models, TransE is considered as the most potential model because of its simple operation, high computing efficiency [21], handling multiple relational data problems, and TransE is widely used. However, we note that TransE does not do well in dealing with one-to-many, many-to-one, many-to-many relationships.

Therefore, a knowledge representation learning model based on hyperplane projection and relational attributes [22], namely TransH-RA, is proposed in this paper. To be specific, firstly, we introduce the idea of hyperplane projection [23] based on the TransE model, we also use the TransH model for reference and introduce the mechanism of projecting to a particular relational hyperplane, and map the head entity **h** and tail entity **t** to the plane of special relation **r**, so that entities have different semantics on different relationships; Secondly, considering that it is not easy to identify different similar [24] entities, the neighborhood information of entities is added to learn the neighborhood of entities around different entities; Then, in order to further strengthen the ability to deal with complex relationships, attribute features of relationships [25] are added and the knowledge of attribute is embedded; Finally, in the training of the model, the probability method [26] is selected to replace the head and tail entities. Link prediction and triple classification tasks are evaluated on four publicly available datasets: WN18, FB15K, WN11, and FB13. The experimental results showed that the TransH-RA made better performance on Mean-Rank, Hits@10 and ACC, which verified the validity of the model.

The main contributions of this paper are as follows:

- A new model, TransH-RA, is proposed to embed entities and relations by introducing the idea of hyperplane projection;
- when embedding the relation, the attribute features of the relation are added to enhance the learning ability of the model.
- In the training of the model, the selection probability method is used to generate the negative triples.
- In order to enhance the model's ability of recognizing similar different entities, the neighborhood information of entities is added.

II. RELATED WORK

We collect the existing knowledge representation learning models and compare the parameters and time complexity [27] of each model, as shown in Table 1. Next, we will expand it from two aspects: TransE, TransH and other methods. In Table 1, We use E_n denotes the number of entities in a triple; R_n denotes the number of relationships in a triple; A_n represents the number of triples in the knowledge graph; The lowercase u denotes the dimension of the entity embedding. The lowercase letters v, p, and s represent the dimension of the relationship embedding, the number of nodes and the number of tensors in the network model, respectively.

A. TRANSE AND TRANSH

In the TransE model, in order to facilitate the calculation between entities and relationships, the method of mapping entities and relations to vector space is adopted, that is, the rules of translation are used in the computation between entities and relations, which transform them into the computation between corresponding vectors. The score function of TransE is:

$$f_r(h,t) = \|h+r-t\|_{l_1/l_2} \tag{1}$$

The key problem that TransH model solves is that TransE can not deal well with the one-to-many, many-to-one, many-to-many relationships. The core idea is that the hyperplane projection method is adopted, the head entity and the tail entity are projected to the relational plane respectively through the relational mapping matrix w_r, and then the legal vector is combined for transformation. Finally, the principle of translation is used again for calculation. The components after projection are:

$$h_{\perp} = h - W_r^T h W_r \tag{2}$$

$$t_{\perp} = t - W_r^T t W_r \tag{3}$$

By combining formula (2) and formula (3), the score function of the model can be obtained, namely:

$$f_r(h,t) = \left\| \left(h - W_r^T h W_r\right) + d_r - \left(t - W_r^T t W_r\right) \right\|_{l_1/l_2} \tag{4}$$

B. OTHER MODELS

In the Unstructured [28] model, the energy model captures semantic information between words, entities, and their combinations and scores the relationships between ambiguous lemmas and unambiguous entities, and learns on multiple resources. The score function of the Unstructured model is:

$$f_r(h,t) = -\|h - t\|_2^2 \tag{5}$$

The idea behind the SE [29] model is that embedding is built by a neural network with a special structure which allows the integration of the original data structure into the learning representation. M_{rh} and M_{rt} are represented as two independent matrices, The score function of the SE model is:

$$f_r(h,t) = -\|M_{rh}h - M_{rt}t\|_1 \tag{6}$$

The algorithm idea of the SME [30] model is to embed multiple graphs into a flexible continuous vector space where the original data is preserved and enhanced. Networks are trained to encode the semantics of these graphs in order to assign high probabilities to reasonable components. SME represents the tail entity and head entity with low dimensional vector, and defines several projection matrices, SME defines two energy functions, the single linear form is:

$$g_\eta = M_{\eta 1} e_\eta + M_{\eta 2} r + b_\eta \tag{7}$$

The bilinear form is:

$$g_\eta = \left(M_{\eta 1} e_\eta\right) \otimes \left(M_{\eta 2} r\right) + b_\eta \tag{8}$$

In formula (9), $\eta = \{left, right\}$, $e_{left} = h$, $e_{\mathrm{right}} = t$, $\otimes$ represents Hadamard product. The score function of the SME model is:

$$f_r(h,t) = g_{\mathrm{left}}^T g_{\mathrm{right}} \tag{9}$$

The idea behind the LFM [31] model is based on a bilinear structure that captures the order of various interactions between data, shares sparse potential factors in different relationships. The model extracts different semantics of different entities and relationships, and encodes each entity as a vector, and sets up a matrix for each relationship. The score function of the LFM model is:

$$f_r(h,t) = h^T M_r t \tag{10}$$

The idea behind the NTN [32] model is that when defining an embedded relationship, it is not defined randomly, but it is defined by means of a neural tensor network, which can be well related to two specific entities, and can well relate entities of different dimensions. The score function of the NTN model is:

$$f_r(h,t) = u_r^T f\left(h^T W_r t + W_{rh} h + W_{rt} t + b_r\right) \tag{11}$$

In formula (11), W_{rh}, W_{rt} represent the parameter of relation r, b_r denotes deviation, $f()$ represents the tanh operation.

The idea of TransR model is that for each type of relationship, there is not only a vector r to describe itself. In addition, there is a mapping matrix M_{r} to describe the relationship space where the relationship is located. At the same time, entities and relationships are mapped to entity vector space and relationship vector space respectively, and finally transformed in the corresponding relationship space.The score function of TransR is:

$$f_r(h,t) = \|h_r + r - t_r\|_2^2 \tag{12}$$

The idea of TransAH [33] model is to adopt an adaptive measurement method, add a diagonal weight matrix to convert Euclidean distance into weighted Euclidean distance, and introduce a hyperplane model oriented to a specific relationship. Finally, the head entity and the tail entity are mapped to a given relational hyperplane for differentiation.The score function of TransAH is:

$$\begin{aligned} f_r(h,t) = &\left(\left(h - n_r^T h n_r\right) + r - \left(t - n_r^T t n_r\right)\right)^T \bullet D_r \\ &\bullet \left(\left(h - n_r^T h n_r\right) + r - \left(t - n_r^T t n_r\right)\right) \end{aligned} \tag{13}$$

The idea of STransH model is to model in entity space and relationship space respectively, and use the nonlinear operation of single-layer neural network to strengthen the semantic relationship between entities and relationships. At the same time, the mechanism of projecting to specific relationship hyperplane is introduced, so that entities have different roles in different relationships.The score function of STransH is:

$$f_r(h,t) = \mathrm{g}\left(\left\| W_{r,1} h_{\perp} + r - W_{r,2} t_{\perp} \right\|_{L_1/L_2}\right) \tag{14}$$

In equation (14), g() represents nonlinear operation, L_1 and L_2 represents distance parameter.

III. OUR METHODOLOGY

At first, we introduce the motivation of the TransH-RA model, mainly from the model to deal with the problem of knowledge map completion; And then we analyze the algorithm idea of TransH-RA model; In the end, the training process of the model is introduced.

A. OUR MOTIVATION

Researchers have proposed several knowledge representation models, such as TransE, TransAH, TransENMM [34], TransF [35], TransD [36], etc. In these models, TransE is the most famous model inspired by the translation invariance of word vectors in the semantic space in Word2vec model, the score function of TransE is $f_r(h,t) = \|h + r - t\|_2^2$, which represents transformation of entities and relationships into vectors, and the calculation of Euclidean distance between the head entity and the tail entity after the relationship [37] transformation using the translation principle. TransE achieves good prediction results while maintaining the problem of few parameters and being able to handle multiple relationships. In addition, it can alleviate the problem of data sparsity, but its performance in dealing with complex attributes is not satisfactory. That is to say, the TransE method has limitations when dealing with complex relationships such as one-to-many, many-to-one, many-to-many, and transitivity. For example, when dealing with (US president, is, Biden) and (US president, is, Trump), it is easy for TransE to conclude that Biden equals Trump's result according to TransE's model idea, but actually these two people are different entities and should be represented by different vectors. For other many-to-one, many-to-many complex relationships, it is also easy to confuse different entities with the same relationship, In addition, TransE adopts the translation strategy of head entity (h) + relation (r) = tail entity (t), which is too strict. On the one hand, the proposed model can reduce the constraints of this translation, on the other hand, it can also enhance the learning and representation ability of the model. So TransE does not work well in dealing with complex relationships.

B. TRANSH-RA

The core idea of TransH-RA model is: Above all, based on the TransE model, the idea of hyperplane projection is introduced, we also use the TransH model for reference and introduce the mechanism of projecting to a particular relational hyperplane, the head entity h and tail entity t are mapped to the plane of special relation r, and the transformed head entity $h_\perp$ and tail entity $t_\perp$ are obtained, this is conducive to dealing with complex relationships; Secondly, considering that it is not easy to identify different similar entities, the neighborhood information of entities is added to learn the neighborhood of entities [38] around different entities. Specifically, when the number of entity neighbors(N) around the entity exceeds 4, the nearest 4 entities are selected, The reciprocal of the distance dis_i (i = 0,1,2,3,4) is the weight β of each neighbor node, and the reciprocal of the sum of the distances is used as the weight, and the weight is used to update the parameters of model training. This has the advantages of high training effect and low memory consumption, and the domain information of entities can be fully utilized; When the number of neighborhood of entities around the entity is less than 4, select all entities, and take the reciprocal of the sum of these entity distances as the weight;

$$\beta = \frac{1}{dis_i} \quad (N > 4) \tag{15}$$

$$\beta = \frac{1}{\sum_{0}^{4} dis_i} \quad (N <= 4) \tag{16}$$

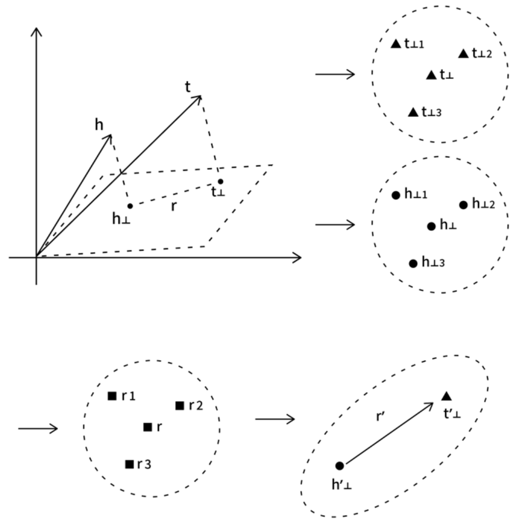


FIGURE 1. The core idea of TransH-RA model.

After such processing, although the model's ability to handle complex relationships has been enhanced, it is still not significant. For this reason, attribute features of the relationship are added, because the attribute features of the relationship are related to the head entity or tail entity itself, ignoring these factors will easily affect the learning ability of the model. Specifically, because each relationship has many attributes, in the selection of attribute relationships, if all are selected, it will waste resources and affect the efficiency of operation. Therefore, only the one with the smallest distance is selected as the attribute, and the attribute knowledge is embedded [39] r′. Among them, in the selection of relationship attributes, the selection is based on the type of entity, for example, if the entity belongs to the "sports type", Then choose the relationship of "sports type". Generally speaking, we choose "sports type". the "sports type" entity has a relatively large relationship with sports related attributes, so we choose sports type, and the minimum distance is taken as the attribute; Finally, in the training of the model, the probability method is chosen to replace the head and tail entities. The score function of the TransH-RA model is:

$$f_r(h,t) = \beta * \left\| h'_\perp + \left(r + r'\right) - t'_\perp \right\|_{l_1/l_2} \tag{17}$$

$$h'_\perp = \frac{1}{\sum_{i=1}^{3} h_i} \bullet h_\perp \tag{18}$$

C. MODEL TRAINING

In model training, negative example triples [40] need to be constructed from positive example triples. The method of TransE model is to randomly select positive triples from the knowledge map, and then scramble positive triples to construct negative triples. However, the TransH-RA model uses a probabilistic approach to replace the head-and-tail entities [41]. Considering that both head and tail entities have many attributes, different attributes have different semantic relationships. Therefore, when choosing the probability method to replace the head and tail entities, it is necessary to select the probability according to the type of relationship. Specifically, for many-to-one relationships, a higher probability is selected to replace the tail entity. The advantage is that multiple attributes of the tail entity can be fully trained, which can improve the discrimination of the tail entity; For one-to-many relationships, the advantage of choosing a higher probability to replace a head entity is that multiple attributes of the head entity can be adequately trained, so that the head entity can be better distinguished. In the process of model training, in order to distinguish positive triples from negative triples, the following loss functions are used in the model training:

$$L=\sum_{(h,r,t)\in S}\sum_{(h',r,t')\in S'}\max\left(f_r(h,t)+\gamma-f_r\left(h',t'\right),0\right) \quad (19)$$

In formula (19), the content of S is the set to which the regular triple belongs, the content of S' denotation is the set of negative triples, the content of max(y, x) denotation is the return of the larger value between y and x, and the content of γ refers to the European distance between the wrong triple loss function score and the correct triple loss function score. Therefore, the optimization objective of the objective function is to separate the wrong triplet from the correct triplet to the greatest extent, so as to distinguish the correct triplet from the wrong triplet.

Algorithm 1 The Algorithm of TransH-RA

Input: Training set $S=\{(h,r,t)\}$,entities and rel, sets entity set to E and margin γ,embeddings dim.d.

Initialize r $\leftarrow$ uniform $\left(-\frac{6}{\sqrt{d}},\frac{6}{\sqrt{d}},\right)$ for each r $\in$ R
r $\leftarrow$ r/$\|$r$\|$ for each r $\in$ R
e $\leftarrow$ uniform $\left(-\frac{6}{\sqrt{d}},\frac{6}{\sqrt{d}},\right)$ for each entity e $\in$ E
loop
  e $\leftarrow$ e/$\|$e$\|$ for each entity e $\in$ E
  $S_{batch} \leftarrow$ sample(S,b) //sample a minibatch of size b
  $T_{batch} \leftarrow \emptyset$ //initialize the set of pairs of triplets
  for (h, r, t) $\in S_{batch}$ **do**
    $\left(h',r,t'\right) \leftarrow$ sample $\left(S'_{(h,r,t)}\right)$ //add relationship attribute and sample a corrupted triple
    $T_{batch} \leftarrow T_{batch} \cup \{((h,r,t),t\quad ph/(tph+hpt)\left(uh',r,t\right),h\,pt/(tph+hpt)\left(h,r,u\,t'\right))\}$
  end for
  Update embeddings w.r.t.
  $$\sum_{((h,r,t),(h',r,t'))\in T_{batch}}\nabla\left[\gamma+d(h+r,t)-d\left(h'+r,t'\right)\right]_+$$
  //The embedding is updated after the head and tail entities are replaced by the probability method
end loop//Algorithm stop

IV. EXPERIMENT

This part mainly introduces the design of the experiment and the discussion of the results. In terms of configuration, pycharm needs to be used for code writing, Ubuntu needs to be used for the system, and GTX1080 needs to be used for model training. The model training platform is openke. In the process of the experiment, we mainly completed the experiments of link prediction and triple classification. Among them, link prediction needs to carry out the test of complex relationships.

TABLE 2. Data set statistics.

Data Set	WN18	FB15K	WN11	FB13
#Entities	40943	14951	38696	75043
#Relationships	18	1345	11	13
#Train	141442	483142	112581	316232
#Valid	5000	50000	2609	5908
#Test	5000	59071	10544	23733

A. DATA SET

This experiment completed two tasks, including link prediction and triple classification. For link prediction tasks, two indicators, MeanRank and Hits@10, are used mainly as the basis for comparison; For triple classification, ACC and training time are used mainly as the indicators for comparison. In order to facilitate data comparison with existing classical models, four data sets used by TransE are selected in the experiment: two subsets WN18 and WN11 in WordNet and two subsets FB15k and FB13 in Freebase, in which FB15K is considered to be a large data set due to the large number of relationship coefficients, as shown in Table 2:

B. LINK PREDICTION

The goal of link prediction is to predict the missing h or t in triples (h, r, t). We remove the head entity or tail entity from the triples (h, r, t), and then replace each triplet [42] in this article's test set with all the entities in the set in turn. We first calculate the score for these corrupted triples, then rank them in descending order, and finally record the ranking of the correct entities. The task emphasizes the ranking of the right entities, rather than just finding the best one.

1) EVALUATION INDEX

There are two evaluation indicators adopted by TransE and TransH: first, the average ranking of correct entities is recorded as MeanRank [43]; Second, the probability that the correct entity ranks in the top 10 is recorded as Hits@10 [44]. The lower the MeanRank is, the better the experimental results are. The higher Hits@10 is, the better the effect of experimental prediction is. In consideration of data comparison with TransE and TransH models, we also select MeanRank and Hits@10 as an evaluation index. The specific process of scoring is as follows: First, replace the tail entity or the head entity with each entity y in the data set; Then, score the replaced triples (h, r, y) with the score function; Finally, the entities are ranked from high to low according to the size of the score.

TABLE 3. The Link Prediction results. TransH-RP means that the model is not added to the entity domain. The bold number is the best performance of the experiment.

Method	WN18				FB15K			
	MeanRank		Hits@10/%		MeanRank		Hits@10/%	
	Raw	Filt	Raw	Filt	Raw	Filt	Raw	Filt
Unstructured	315	304	35.3	38.2	1074	979	4.5	6.3
RESCAL [46]	1180	1163	37.2	52.8	828	683	28.4	44.1
SE	1011	985	68.5	80.5	273	162	28.8	39.8
SME(Linear)	545	533	65.1	74.1	274	154	30.7	40.8
SME(Bilinear)	526	509	54.7	61.3	284	158	31.3	41.3
LFM	469	456	71.4	81.6	283	164	26.0	33.1
TransE	263	251	75.4	89.2	243	125	34.9	47.1
TransH	401	388	73.0	82.3	212	87	45.7	64.4
TransA	405	392	80.1	91.3	199	74	51.1	70.4
TransR	238	255	79.8	92.0	198	77	48.2	68.7
STransH	347	330	77.1	90.6	196	68	46.6	69.5
TransD	224	212	79.6	92.2	194	91	51.4	70.3
OPTransE [47]	211	199	79.2	91.7	141	53	51.0	69.9
RPJE [48]	205	183	79.1	91.1	186	50	51.5	70.3
TransR* [49]	206	195	**81.6**	94.6	190	62.6	51.5	69.3
ERDERP [50]	258	246	79.9	93.2	189	54	49.1	71.1
TransH-RP(unif)	321	306	77.2	90.6	187	52.0	48.6	70.8
TransH-RP(bern)	388	374	77.1	90.0	200	103	51.1	70.1
TransH-RA(unif)	120	**107**	80.0	92.9	**141**	**48**	**52.1**	**72.0**
TransH-RA(bern)	**119**	107.5	80.2	**95.0**	143	52	52.0	71.7

In the actual situation [45], a damaged triplet may be encountered in the knowledge map, that is, the triplet is actually correct, and it is also correct to place the triplet before the original triplet. In the experiment, this triple will interfere with the ranking score of the original triple. In order to eliminate this interference factor, it is necessary to filter out these interfering triples when generating negative example triples to ensure that the negative example triples do not belong to the training set, verification set and test set. We call the experimental settings that have been filtered "Filt", and the experimental settings that have not been filtered "Raw". In general, the experimental data after filtering is better than that without processing.

2) EXPERIMENTAL REALIZATION

The TransH-RA model in this paper is compared with several existing models, including SME, NTN, TransE, TransH, etc. Considering the problems of experimental realization and parameter adjustment, we do not obtain the best results in the corresponding literature when we reproduce. Because the experimental data sets of these models are the same, we directly use the optimal experimental results of each model in the corresponding literature as the comparison basis. In order to reduce the influence of random initialization of parameters on the results, 10 experiments are conducted on each set of parameters, and the average values are taken as the final results. When training TransH-RA, the learning rate α in {0.0001, 0.001, 0.005, 0.01} is used in the process of random gradient descent, Margin γ in {1, 1.5, 2, 3, 4, 4.5, 5, 10}, the embedded dimension k in {20, 50, 100, 200, 250}, and the size B of batch in {20, 50, 120, 1200, 4800, 9600}. The optimal parameters are determined by the validation set.

We use "unif" [51] to represent the traditional method of equal probability uniform distribution to replace the head entity or tail entity, and "bern" to represent the method of Bernoulli distribution sampling strategy. That is, according to different relationship types, we use different probabilities to replace the head entity and tail entity, which has the advantage of reducing the number of wrong triples. Under the "unif" setting, the best configuration is: On WN18, $\alpha = 0.0001, \gamma = 4.5, \text{k} = 100,\ \text{B} = 1200$; On FB15K, $\alpha = 0.0001, \gamma = 1.5, \text{k} = 100,\ \text{B} = 9600$. Under the "bern" setting, the best configuration is: On WN18, $\alpha = 0.0001, \gamma = 4.5, \text{k} = 100,\ \text{B} = 1200$; On FB15K, $\alpha = 0.0001, \gamma = 1.5, \text{k} = 100,\ \text{B} = 1200$. For both data sets, all training triples were iterated 500 times in this experiment.

3) EXPERIMENTAL RESULTS

From Table 3, we can see that TransH-RA (unif) and TransH-RA (bern) are better than other methods on the Maen-Rank index for WN18 and FB15K datasets. For Hits@10, Compared with TransE and TransH, TransH-RA increased 5.8% and 12.7% on WN18, and increased 24.9% and 7.6% on FB15K, respectively, with significant performance improvement.

To verify the ability of TransH-RA to deal with complex relationships, we select FB15K as the data set, change the file names of 1-1, 1-n, n-1, n-n to test2id, and then carry out an experiment of complex relational link prediction. The optimal combination of parameters on FB15K is still referenced in

TABLE 4. The value of Hits@10 for various relationships on FB15K. TransH-RP means that the model is not added to the entity domain.

Method	Predicting Left(Hits@10)				Predicting Right(Hits@10)			
	1-1	1-n	n-1	n-n	1-1	1-n	n-1	n-n
Unstructured	34.5	2.5	6.1	6.6	34.3	4.2	1.9	6.6
SE	35.6	62.6	17.2	37.5	34.9	14.6	68.3	41.3
SME	30.9	69.6	19.9	38.6	28.2	13.1	76.0	41.8
TransE	43.7	65.7	18.2	47.2	43.7	19.7	66.7	50.0
TransH	66.8	87.6	28.7	64.5	65.5	39.8	83.3	67.2
TransRD [52]	79.9	92.7	33.9	66.7	79.3	37.2	92.7	71.1
TransA	86.8	94.4	42.7	77.8	86.7	54.3	91.4	80.6
TransD	86.1	95.5	39.8	78.5	85.4	50.6	92.4	81.2
OPTransE	**93.1**	93.4	55.0	80.8	90.8	57.4	91.7	81.3
RPJE	92.2	96.0	54.4	81.6	91.1	73.9	91.3	83.3
TransR*	84.2	92.8	39.0	68.7	82.3	39.3	93.2	72.6
ERDERP	78.1	86.3	49.5	70.3	79.1	51.7	85.6	73.3
TransH-RP(unif)	85.1	94.9	56.6	80.1	84.0	60.3	92.5	82.7
TransH-RP(bern)	85.5	95.0	49.6	75.7	83.7	49.4	93.9	78.9
TransH-RA(unif)	88.5	94.7	**58.7**	76.5	88.1	59.5	94.0	80.1
TransH-RA(bern)	92.5	**96.5**	52.1	**81.9**	**91.4**	**75.4**	**95.1**	**84.6**

the configuration of parameters. From Table 4, we can see that TransH-RA achieves the best performance compared to TransE model and TransH, which 1-n of Hits@10 value reaches 96.5% on Predicting Left. On the Predicting Right, the Hits@10 value for n-1 is 95.1%. This suggests that TransH-RA does show a significant improvement in complex relationship types.

C. TRIPLE CLASSIFICATION

The purpose of triad classification is to give a triad (h, r, t) and judge whether (h, r, t) is correct. The main problem it solves is to classify a triad as "correct" or "wrong". For a triple (h, r, t), if its score is less than the given threshold σ_r, then the prediction is correct, and vice versa, it is wrong. Through experiments, we find that the threshold value for obtaining the maximum classification accuracy in the verification set determines the value size of σ_r.

In the triple classification experiments, we select a subset of WordNet, WN11, and a subset of FreeBase, FB13. Considering that WN11 and FB13 contain very few relationships, we choose FB15K, which contains more relationships. The statistical information of the experimental dataset is shown in Table 2.

1) EVALUATION INDICATORS

The accuracy rate (ACC) and training time (t) are used to evaluate the triad classification task. The higher the ACC is, the better the model will perform on the task of triple classification. The smaller the t is, the shorter the training time is, and the lower the time complexity of the model is. The formula for calculating the accuracy is as follows:

$$ACC = \frac{T_p + T_n}{N_{pos} + N_{neg}} \tag{20}$$

In formula (20), T_p refers to the correct number of positive triples to be predicted; T_n denotes the number of negative triples with correct predictions. N_{pos} represents the number of positive triples in the training set, while N_{neg} represent the number of negative triples in the training set.

2) EXPERIMENTAL REALIZATION

During the SDG process, the learning rate α in $\{0.0001, 0.001, 0.01, 0.1\}$ is selected, Margins γ in $\{1, 2, 4, 4.5, 5, 10\}$, Dimension K of entity and relationship vectors is selected from $\{20, 50, 100, 150, 200\}$, and batch size B is selected from $\{20, 120, 480, 1200, 4800, 9600\}$. We find that the validation set measures the precision of the optimal configuration. The best configuration on WN11 is: $\alpha = 0.1$, $\gamma = 4.5$, $\mathrm{k} = 20$, $\mathrm{B} = 4800$, and l_1 is used as the similarity measure; The best configuration on FB13 is: $\alpha = 0.001$, $\gamma = 5$, $\mathrm{k} = 50$, $\mathrm{B} = 4800$, and l_1 is used as the similarity measure; The best configuration on FB15K is: $\alpha = 0.0001$, $\gamma = 2$, $\mathrm{k} = 100$, $\mathrm{B} = 4800$, and l_1 is used as a measure of similarity.

3) EXPERIMENTAL RESULTS

Table 5 shows the evaluation results of the triple classification. It can be seen that the TransH-RA model works best on WN11, with 13.6% and 10.6% improvement over the TransE and TransH models, respectively. On FB13, TransH-RA model perform better than TransE and TransH, with 9.8% and 7.5% improvement, respectively. TransH-RA model perform slightly weaker than TransR* in FB15K data set, which improves by 16.0% and 8.0% compared with TransE and TransH, indicating that TransH-RA model is suitable for both sparse data set and large dense data set.

In order to select the optimal number of neighbors, we use FB15K to perform experiments on the number of neighbors and accuracy (ACC) on TransH-RA and TransE. The specific results are shown in Figure 2.

TABLE 5. The triplet classification accuracy of different models. 40h, 5m, 30m and 35m represent time loss. TransH-RP means that the model is not added to the entity domain. "—" indicates the unobtained experimental results.

Method	WN11	FB13	FB15K
Distant	53.0	75.2	—
SLM	69.0	85.3	—
SME	73.8	84.3	—
NTN	70.4	87.1	66.5(≈ 40 h)
TransE	75.8	81.5	79.7(≈ 5 m)
TransH	78.8	83.8	87.7(≈ 30 m)
TransR	75.5	84.7	88.0(≈ 30 m)
STransH	79.6	85.2	89.6(≈ 35 m)
TransA	83.2	87.3	—
TransD	86.4	89.1	88.0(≈ 30 m)
TransAH	85.2	88.1	92.0(≈ 30 m)
OPTransE	82.3	87.2	90.5(≈ 30 m)
RPJE	84.7	—	91.3(≈ 30 m)
TransR*	86.2	81.7	**97.1(≈ 30m)**
ERDERP	—	—	91.2(≈ 30 m)
TransH-RP(unif)	85.6	84.2	90.1(≈ 30 m)
TransH-RP(bern)	86.1	85.3	92.0(≈ 30 m)
TransH-RA(unif)	84.7	89.2	94.3(≈ 30 m)
TransH-RA(bern)	**89.4**	**91.3**	95.7(≈ 30 m)

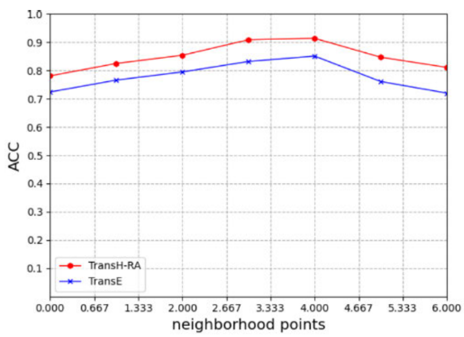


FIGURE 2. The number of adjacent nodes and accuracy experiment on FB15K.

From Figure 2, it can be seen that TransH-RA and TransE have an upward trend in the value of ACC in the range of [0, 4]. In the range of [4], [6], the value of ACC shows a downward trend, which indicates that the number of selected neighbor nodes will affect the result of ACC, and the optimal number of neighbor nodes is 4.

In order to further analyze the effect of learning rate and marginal value on the experimental results, we do experiments on learning rate(α) and accuracy(ACC), marginal value (λ) and accuracy(ACC) using the control variable method, respectively. The results are shown in Figure 3 and Figure 4.

It can be seen from Figure 3 that, on the whole, the accuracy of TransH-RA is higher than that of TransE. When the learning rate is within the range of [0, 0.1], the increase of the accuracy of TransH-RA is significantly higher than that of TransE. Within the range of [0, 0.1], the fluctuation of

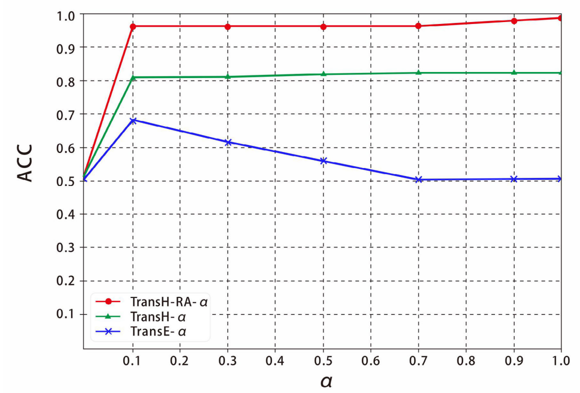


FIGURE 3. The experiment of learning rate and accuracy on FB15K.

the accuracy of TransH-RA is relatively small and gradually tends to be flat, while the fluctuation of TransE is large, which shows that the robustness of TransH-RA is stronger than that of TransE, and the impact of the learning rate on ACC is higher in TransH-RA than in TransE.

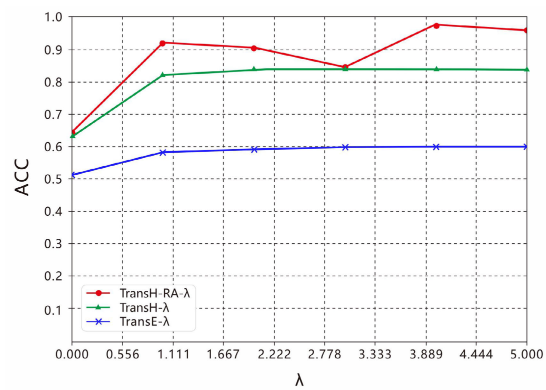


FIGURE 4. Marginal value and accuracy experiment on FB15K.

It can be seen from Figure 4 that, on the whole, the accuracy of TransH-RA is higher than that of TransE. When the boundary value is within the range of [0, 5], the accuracy rate of TransH-RA increases significantly more than that of TransE. However, the accuracy rate of TransH-RA fluctuates greatly while that of TransE fluctuates gently, which indicates that the robustness of TransH-RA is weaker than that of TransE, and the learning rate has a greater impact on ACC than that of TransE. According to the analysis of the reasons, TransH-RA is greatly affected by positive and negative samples due to the limitation of negative sampling strategy, and the sampling strategy of negative samples needs to be further improved in the future.

V. DISCUSSION

Considering the importance of AI interpretability [53], First of all, we introduce the idea of hyperplane projection, so that different entities have different roles in a specific relationship, which reflects the reliability of our model. Considering that it is not easy to identify different similar entities, we add the information of the entity field, and the fourth part of the experiment verifies that the selection of neighbor nodes will affect the training time consumption, memory consumption, and the impact on ACC, which reflects the innovation of our model. In addition, the relationship attribute feature itself has a certain correlation with the head and tail entities, so we add the attribute feature of the relationship to enhance the processing ability of TransH-RA to complex relationships, which reflects the flexibility of our model. Finally, the use of the probability method proves that our model TransH-RA has a certain effect on the triple classification experiment and link prediction experiment. Therefore, our model is interpretable.

Knowledge map is a very promising research direction. With the development of machine learning technology, the knowledge representation method in the knowledge map has ushered in a significant development. For the transportation field [54], [55], the knowledge map has been used for the search of intelligent transportation destinations with its excellent decision-making, analysis and recommendation capabilities, and can recommend a suitable travel route according to people's needs, which not only facilitates our travel, but also improves our work efficiency. For the medical field [56], [57], knowledge atlas can establish a more systematic and complete knowledge base and provide efficient retrieval. This application can facilitate doctors to accurately find the cause of cancer, and can help doctors diagnose cancer and save patients' lives. Based on the application of transportation and medicine, we can see that knowledge graph has become an indispensable part of our life, providing information retrieval, auxiliary diagnosis, automatic question answering and other help for our life.

VI. CONCLUSION AND FUTURE WORK

In this paper, a knowledge representation learning model based on hyperplane projection and relational attributes, namely TransH-RA, is proposed. This model mainly solves the problem of defect in dealing with the relations of one-to-many, many-to-one, many-to-many and reflexive in TransE. First of all, we introduce the idea of hyperplane projection based on the TransE model, which maps the head entity **h** and the tail entity **t** to the plane of a specific relationship **r**. This idea is inspired by TransH, which makes different entities have different roles in a specific relationship. Secondly, considering that it is not easy to identify different similar entities, the neighborhood information of entities is added to learn the neighborhood of entities around different entities; Then, in order to further strengthen the ability to deal with complex relationships, attribute features of relationships are added and attribute knowledge is embedded; Finally, in the model training, the probability method is chosen to replace the head and tail entities. Link prediction and triad classification tasks are evaluated on four publicly available datasets: WN18, FB15K, WN11, and FB13. The experimental results show that TransH-RA in MeanRank, Hits@10, as well as ACC three indicators have been improved, thus verifying the validity of the model.

In future work, further improvements to the proposed TransH-RA approach are planned. Because TransH-RA does not significantly improve performance on large datasets during the link prediction experiment, in the course of reading the literature, it is found that LSTM can enhance the learning and prediction ability of the model. Therefore, in future studies, we will try to combine LSTM with our proposed model to improve the overall prediction ability of the model. In addition, in the experiment of the influence of boundary values on accuracy, the model is greatly affected by positive and negative samples, and the sampling strategy of negative samples needs to be further improved in the future.

ACKNOWLEDGMENT

The authors thank all anonymous reviewers for their constructive comments.

REFERENCES

[1] P. Liu, X. Wang, Q. Fu, Y. Yang, Y.-F. Li, and Q. Zhang, "KGVQL: A knowledge graph visual query language with bidirectional transformations," *Knowl.-Based Syst.*, vol. 250, Aug. 2022, Art. no. 108870.
[2] S. Guo, Q. Wang, B. Wang, L. Wang, and L. Guo, "Semantically smooth knowledge graph embedding," in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process.*, 2015, pp. 84–94.
[3] Z. Zhang, F. Zhuang, H. Zhu, Z. Shi, H. Xiong, and Q. He, "Relational graph neural network with hierarchical attention for knowledge graph completion," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 5, pp. 9612–9619.
[4] R. Lijuan, L. Jun, and G. Wei, "Multi-source knowledge embedding research of knowledge graph," in *Proc. IEEE 3rd Int. Conf. Circuits, Syst. Devices (ICCSD)*, Aug. 2019, pp. 163–166.
[5] D. Santra, S. K. Basu, J. K. Mandal, and S. Goswami, "Rough set based lattice structure for knowledge representation in medical expert systems: Low back pain management case study," *Expert Syst. Appl.*, vol. 145, May 2020, Art. no. 113084.
[6] L. Guo, F. Yan, T. Li, T. Yang, and Y. Lu, "An automatic method for constructing machining process knowledge base from knowledge graph," *Robot. Comput.-Integr. Manuf.*, vol. 73, Feb. 2022, Art. no. 102222.
[7] S. Ji, S. Pan, E. Cambria, P. Marttinen, and P. S. Yu, "A survey on knowledge graphs: Representation, acquisition, and applications," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 2, pp. 494–514, Feb. 2022.
[8] P. Rosso, D. Yang, and P. Cudré-Mauroux, "Beyond triplets: Hyper-relational knowledge graph embedding for link prediction," in *Proc. Web Conf.*, Apr. 2020, pp. 1885–1896.
[9] K. Kumarasinghe, N. Kasabov, and D. Taylor, "Deep learning and deep knowledge representation in spiking neural networks for brain-computer interfaces," *Neural Netw.*, vol. 121, pp. 169–185, Jan. 2020.
[10] H. Balabin, C. T. Hoyt, C. Birkenbihl, B. M. Gyori, J. Bachman, A. T. Kodamullil, P. G. Plöger, M. Hofmann-Apitius, and D. Domingo-Fernández, "STonKGs: A sophisticated transformer trained on biomedical text and knowledge graphs," *Bioinformatics*, vol. 38, no. 6, pp. 1648–1656, Mar. 2022.
[11] Z. Lei, Y. Sun, Y. A. Nanehkaran, S. Yang, M. S. Islam, H. Lei, and D. Zhang, "A novel data-driven robust framework based on machine learning and knowledge graph for disease classification," *Future Gener. Comput. Syst.*, vol. 102, pp. 534–548, Jan. 2020.
[12] H.-C. Liu, X. Luan, Z. Li, and J. Wu, "Linguistic Petri nets based on cloud model theory for knowledge representation and reasoning," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 4, pp. 717–728, Apr. 2018.

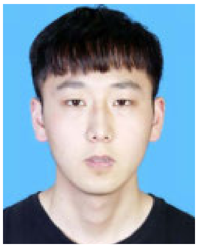
[13] N. Jain, T.-K. Tran, M. H. Gad-Elrab, and D. Stepanova, ''Improving knowledge graph embeddings with ontological reasoning,'' in *Proc. Int. Semantic Web Conf.* Springer, 2021, pp. 410–426.
[14] S. Katalnikova, L. Novickis, N. Prokofyeva, V. Uskov, and C. Heinemann, ''Intelligent collaborative educational systems and knowledge representation,'' *Proc. Comput. Sci.*, vol. 104, pp. 166–173, Jan. 2017.
[15] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenk, ''Translating embeddings for modeling multi-relational data,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 26, 2013, pp. 1–9.
[16] H. Yang and J. Liu, ''Knowledge graph representation learning as groupoid: Unifying TransE, RotatE, QuatE, ComplEx,'' in *Proc. 30th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2021, pp. 2311–2320.
[17] Z. Wang, J. Zhang, J. Feng, and Z. Chen, ''Knowledge graph embedding by translating on hyperplanes,'' in *Proc. AAAI Conf. Artif. Intell.*, vol. 28, 2014, pp. 1–8.
[18] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, ''Learning entity and relation embeddings for knowledge graph completion,'' in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 1–7.
[19] C. Xiaojun and X. Yang, ''STransH: An improved knowledge representation model based on translation model,'' *Comput. Sci.*, vol. 46, no. 9, pp. 184–189, 2019.
[20] H. Xiao, M. Huang, Y. Hao, and X. Zhu, ''TransA: An adaptive approach for knowledge graph embedding,'' 2015, *arXiv:1509.05490*.
[21] X. Benavent, A. Castellanos, E. de Ves, A. García-Serrano, and J. Cigarrán, ''FCA-based knowledge representation and local generalized linear models to address relevance and diversity in diverse social images,'' *Future Gener. Comput. Syst.*, vol. 100, pp. 250–265, Nov. 2019.
[22] D. W. Wardani and J. Kung, ''Semantic mapping relational to a directed property hypergraph model,'' in *Proc. IEEE Int. Conf. Comput. Inf. Technol., Ubiquitous Comput. Commun., Dependable, Autonomic Secure Comput., Pervasive Intell. Comput.*, Oct. 2015, pp. 152–159.
[23] I. Balažević, C. Allen, and T. M. Hospedales, ''Hypernetwork knowledge graph embeddings,'' in *Proc. Int. Conf. Artif. Neural Netw.* Springer, 2019, pp. 553–565.
[24] Y. Jia, Y. Wang, X. Jin, and X. Cheng, ''Path-specific knowledge graph embedding,'' *Knowl.-Based Syst.*, vol. 151, pp. 37–44, Jul. 2018.
[25] E. Bayram, A. García-Durán, and R. West, ''Node attribute completion in knowledge graphs with multi-relational propagation,'' in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 3590–3594.
[26] T. Wu, A. Khan, M. Yong, G. Qi, and M. Wang, ''Efficiently embedding dynamic knowledge graphs,'' *Knowl.-Based Syst.*, vol. 250, Aug. 2022, Art. no. 109124.
[27] K. AbuDahab, D.-L. Xu, and Y.-W. Chen, ''A new belief rule base knowledge representation scheme and inference methodology using the evidential reasoning rule for evidence combination,'' *Expert Syst. Appl.*, vol. 51, pp. 218–230, Jun. 2016.
[28] A. Bordes, X. Glorot, J. Weston, and Y. Bengio, ''Joint learning of words and meaning representations for open-text semantic parsing,'' in *Proc. Artif. Intell. Statist.*, 2012, pp. 127–135.
[29] A. Bordes, J. Weston, R. Collobert, and Y. Bengio, ''Learning structured embeddings of knowledge bases,'' in *Proc. 25th AAAI Conf. Artif. Intell.*, 2011, pp. 301–306.
[30] A. Bordes, X. Glorot, J. Weston, and Y. Bengio, ''A semantic matching energy function for learning with multi-relational data,'' *Mach. Learn.*, vol. 94, no. 2, pp. 233–259, Feb. 2014.
[31] R. Jenatton, N. Roux, A. Bordes, and G. R. Obozinski, ''A latent factor model for highly multi-relational data,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1–9.
[32] R. Socher, D. Chen, C. D. Manning, and A. Ng, ''Reasoning with neural tensor networks for knowledge base completion,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 26, 2013, pp. 1–10.
[33] F. Yang, Z. Xiang, T. Zhen, Y. Shiyu, and X. Weidong, ''A revised translation-based method for knowledge graph representation,'' *J. Comput. Res. Develop.*, vol. 55, no. 1, pp. 139–150, 2018.
[34] D. Q. Nguyen, K. Sirts, L. Qu, and M. Johnson, ''Neighborhood mixture model for knowledge base completion,'' 2016, *arXiv:1606.06461*.
[35] J. Feng, M. Huang, M. Wang, M. Zhou, Y. Hao, and X. Zhu, ''Knowledge graph embedding by flexible translation,'' in *Proc. 15th Int. Conf. Principles Knowl. Represent. Reasoning*, 2016, pp. 557–560.
[36] G. Ji, S. He, L. Xu, K. Liu, and J. Zhao, ''Knowledge graph embedding via dynamic mapping matrix,'' in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process.*, 2015, pp. 687–696.
[37] Y. Jia, Y. Wang, H. Lin, X. Jin, and X. Cheng, ''Locally adaptive translation for knowledge graph embedding,'' in *Proc. 13th AAAI Conf. Artif. Intell.*, 2016, pp. 1–7.
[38] X. Zheng, B. Wang, Y. Zhao, S. Mao, and Y. Tang, ''A knowledge graph method for hazardous chemical management: Ontology design and entity identification,'' *Neurocomputing*, vol. 430, pp. 104–111, Mar. 2021.
[39] S. Guo, Q. Wang, L. Wang, B. Wang, and L. Guo, ''Knowledge graph embedding with iterative guidance from soft rules,'' in *Proc. AAAI Conf. Artif. Intell.*, 2018, vol. 32, no. 1, pp. 1–8.
[40] L. Cai and W. Y. Wang, ''KBGAN: Adversarial learning for knowledge graph embeddings,'' 2017, *arXiv:1711.04071*.
[41] Z. Sun, J. Yang, J. Zhang, A. Bozzon, L.-K. Huang, and C. Xu, ''Recurrent knowledge graph embedding for effective recommendation,'' in *Proc. 12th ACM Conf. Recommender Syst.*, Sep. 2018, pp. 297–305.
[42] Z. Li, H. Liu, Z. Zhang, T. Liu, and N. N. Xiong, ''Learning knowledge graph embedding with heterogeneous relation attention networks,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 8, pp. 3961–3973, Aug. 2022.
[43] Y. Luan, L. He, M. Ostendorf, and H. Hajishirzi, ''Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction,'' 2018, *arXiv:1808.09602*.
[44] Y. Yang, X. Yin, H. Yang, X. Fei, H. Peng, K. Zhou, K. Lai, and J. Shen, ''KGSynNet: A novel entity synonyms discovery framework with knowledge graph,'' in *Proc. Int. Conf. Database Syst. Adv. Appl.* Springer, 2021, pp. 174–190.
[45] H. Wang, H. Ren, and J. Leskovec, ''Relational message passing for knowledge graph completion,'' 2020, *arXiv:2002.06757*.
[46] M. Nickel, V. Tresp, and H. P. Kriegel, ''A three-way model for collective learning on multi-relational data,'' in *Proc. ICML*, 2011, pp. 3104482–3104584.
[47] Y. Zhu, H. Liu, Z. Wu, Y. Song, and T. Zhang, ''Representation learning with ordered relation paths for knowledge graph completion,'' 2019, *arXiv:1909.11864*.
[48] G. Niu, Y. Zhang, B. Li, P. Cui, S. Liu, J. Li, and X. Zhang, ''Rule-guided compositional representation learning on knowledge graphs,'' in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 3, pp. 2950–2958.
[49] Z. Zhang, J. Jia, Y. Wan, Y. Zhou, Y. Kong, Y. Qian, and J. Long, ''TransR*: Representation learning model by flexible translation and relation matrix projection,'' *J. Intell. Fuzzy Syst.*, vol. 40, no. 5, pp. 10251–10259, Apr. 2021.
[50] L. Lin, J. Liu, F. Guo, C. Tong, L. Zu, and H. Guo, ''ERDERP: Entity and relation double embedding on relation hyperplanes and relation projection hyperplanes,'' *Mathematics*, vol. 10, no. 22, p. 4182, Nov. 2022.
[51] S. Wang, X. Wei, C. N. Nogueira dos Santos, Z. Wang, R. Nallapati, A. Arnold, B. Xiang, P. S. Yu, and I. F. Cruz, ''Mixed-curvature multi-relational graph neural network for knowledge graph completion,'' in *Proc. Web Conf.*, Apr. 2021, pp. 1761–1771.
[52] Z. Yanli, Y. Xiaoping, W. Liang, and Z. Zhiyu, ''TransRD: A knowledge graph embedding representation model with asymmetric features,'' *Chin. J. Inf.*, vol. 33, no. 11, pp. 73–82, 2019.
[53] J. Xi, D. Wang, X. Yang, W. Zhang, and Q. Huang, ''Cancer omic data based explainable AI drug recommendation inference: A traceability perspective for explainability,'' *Biomed. Signal Process. Control*, vol. 79, Jan. 2023, Art. no. 104144.
[54] S. Wang, Y. Lv, Y. Peng, X. Piao, and Y. Zhang, ''Metro traffic flow prediction via knowledge graph and spatiotemporal graph neural network,'' *J. Adv. Transp.*, vol. 2022, pp. 1–13, Sep. 2022.
[55] L. Zhang, M. Zhang, J. Tang, J. Ma, X. Duan, J. Sun, X. Hu, and S. Xu, ''Analysis of traffic accident based on knowledge graph,'' *J. Adv. Transp.*, vol. 2022, pp. 1–16, Aug. 2022.
[56] J. Xi, L. Ye, Q. Huang, and X. Li, ''Tolerating data missing in breast cancer diagnosis from clinical ultrasound reports via knowledge graph inference,'' in *Proc. 27th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, Aug. 2021, pp. 3756–3764.
[57] J. Xi, Z. Miao, L. Liu, X. Yang, W. Zhang, Q. Huang, and X. Li, ''Knowledge tensor embedding framework with association enhancement for breast ultrasound diagnosis of limited labeled samples,'' *Neurocomputing*, vol. 468, pp. 60–70, Jan. 2022.

YONGKANG WANG received the B.S. degree from Ludong University, Yantai, China, in 2019. He is currently pursuing the M.S. degree with the School of Software, Xinjiang University, Ürümqi, China. His research interest includes knowledge graph.

AISHAN WUMAIER (Member, IEEE) received the Ph.D. degree from Xinjiang University, in 2010. He is currently a Professor with Xinjiang University. His research interests include multimodal natural language processing, visual understanding, speech recognition, and machine translation.

WEIWEI SUN received the B.S. degree from the Xuhai College, China University of Mining and Technology, China, in 2019. He is currently pursuing the M.S. degree with the School of Computer Science and Engineering, Xinjiang University, Ürümqi, China. His research interest includes natural language processing.

YANGXIN LIU received the B.S. degree from the Guangdong University of Science and Technology, Dongguan, China, in 2021. He is currently pursuing the M.S. degree with the School of Information Science and Engineering, Xinjiang University, Ürümqi, China. His research interest includes computer vision.



JIANGTAO HE received the B.S. degree from the Nanyang Institute of Technology, Nanyang, China, in 2020. He is currently pursuing the M.S. degree with the School of Computer Science and Engineering, Xinjiang University, Ürümqi, China. His research interest includes aspect-based sentiment analysis.