ICIC Express Letters
Volume **16**, Number **11**, November **2022**

ICIC International ©2022 ISSN 1881-803X
pp. 1215–1222

A RELATIONAL REASONING METHOD BASED ON GRAPH ATTENTION NETWORK

Yongmei Zhang[1,*], Zhirong Du[1,2] and Ruiqi Li[1]

[1]School of Information Science and Technology
[2]School of Electrical and Control Engineering
North China University of Technology
No. 5, Jinyuanzhuang Road, Shijingshan District, Beijing 100144, P. R. China
2237zhirong@sina.com; 1393552139@qq.com
*Corresponding author: zhangym@ncut.edu.cn

Received January 2022; accepted April 2022

Abstract. *To solve the problem of real-time reasoning, quantitative thinking and the way of making full use of neighbors and relations in the knowledge graph are introduced to complete the rapid knowledge reasoning, and realize the relational reasoning based on the graph attention network. When constructing the knowledge graph, an entity alignment algorithm based on semantic aggregation and attribute attention is adopted to overcome the problem of different expressions for the same entity in different data sources. Combine the idea of quantization in knowledge reasoning, comprehensively consider neighbors and relation information, complete relation reasoning by changing the multi-head attention non-linear function and loss function, perform the non-linear transformation of all initial input values, map the results to a certain range, and reduce computing time. From the two perspectives of MR and Hits@k (k = 1, 3, 10) evaluation indicators and the relation between loss value and the number of iterations, the experiment results show the proposed algorithm has better performance than the other four methods including TransE, ComplEx, ConvE, and the relational graph convolutional network (R-GCN).*

Keywords: Knowledge graph, Graph attention network, Entity alignment, Quantitative method

1. **Introduction.** The continuous development of knowledge graphs (KG) has an increasing impact on applications such as information retrieval and recommendation systems [1]. At present, the general large-scale KG such as Freebase [2], NELL [3], and WordNet [4] have developed rapidly. Domain knowledge graphs are also rapidly emerging. However, whether it is a general KG or a domain KG, it is usually incomplete. There are many missing links between entities, which severely restricts the performance improvement of many downstream tasks.

The knowledge graph is a relational network obtained by connecting all different types of information [5], and it provides the ability to analyze problems from the perspective of "relation". There are many entity pairs and relations in the knowledge graph, but due to the update iteration and incompleteness of the data, the constructed knowledge graph is destined to be incomplete. Similarly, much hidden information cannot be easily found in the knowledge graph. Relational reasoning can better discover hidden information.

Currently, the reasoning based on distributed representation learning represented by the TransE and the reasoning on the basis of a neural network with R-GCN as a common method is the research focuses. The reasoning method based on distributed representation learning is mainly to find a mapping function to map the symbolic representation to the vector space for numerical representation, thereby reducing the dimensionality disaster while capturing the entity and implicit association between relations, and the calculation

DOI: 10.24507/icicel.16.11.1215

speed is fast [6]. The reasoning method on the basis of a neural network mainly utilizes the learning and generalization ability of the neural network to model the knowledge graph tuple.

Specifically, the TransE [7] has low accuracy in dealing with complex relations when the triple information is insufficient, but it cannot accurately infer entities with the same relation. The ComplEx [8] is effective for handling complex relations, but it does not consider neighbor information. The ConvE [9] is a link prediction model expressed by multi-layer nonlinear features, which utilizes few parameters and lacks interaction in reasoning. The R-GCN [10] adopts graph convolutional neural networks for reasoning. The experiment results of this model are unstable and have poor real-time performance.

The proposed method considers the neighbor information by introducing the idea of quantification. The convergence speed is accelerated, and the results are more stable. The main contributions of this paper are as follows.

1) Given the heterogeneity of crawled data, the complex calculation of entity alignment, and the lack of prior alignment data, this paper adopts an entity alignment algorithm based on semantic aggregation and attribute attention.

2) Aiming at the problem of slower convergence caused by insufficient considering of entity neighbors and relations in the process of knowledge graph reasoning, this paper proposes a relational reasoning algorithm based on a graph attention network.

The rest is arranged as follows. The entity alignment method of semantic aggregation and attribute attention is described in Section 2. The proposed relational reasoning algorithm based on the graph attention network is introduced in Section 3. Section 4 shows the experiment results and analysis. The conclusions are stated in Section 5.

2. Entity Alignment Method Based on Semantic Aggregation and Attribute Attention.

Entity alignment is a complex, time-consuming, error-prone task, and lacks prior alignment data. In response to the problems, this paper adopts an entity alignment algorithm based on semantic aggregation and attribute attention.

Entities and relations are embedded in encoding, that is, a range of numbers is randomly generated for vectorization, and function calculations are utilized in the algorithm. The entity alignment algorithm mainly introduces attention to aggregation, and adopts the attribute weight calculation method in [11]. The score value is calculated by Equation (1). The loss function is determined by Equation (2).

$$f(r+t,h)=\sum_{r_{ij}\in R}\sum_{t_{ij}\in E}(r_{ij}+t_{ij})-\sum_{r_{ij}\in R}\mu r_{ij}h_i \tag{1}$$

$$L=\sum_{(h,r,t)\in S}\sum_{(h',r,t')\in S'}\left[\gamma+f(r+t,h)-f\left(r+t',h'\right)\right]^+ \tag{2}$$

where $[x]^+ = \{0, x\}$ represents the maximum value between 0-x, $\gamma > 0$. The above entity alignment algorithm is used for experiment on the lung cancer dataset. The evaluation index adopts the accuracy rate and recall rate. The embedding dimension of the design entity and relation in the experiment is $d = 100$, the boundary $\gamma = \{0.5, 1.0, 1.5\}$, and the maximum batch size $n = \{500, 1000, 1500, 2000, 2500, 3000\}$, when the experiment parameters are selected as $\gamma = 1.0$ and $n = 3000$, the alignment algorithm has the best effect.

3. Relational Reasoning Algorithm Based on Graph Attention Network.

3.1. Problem definition.

Before defining the problem, the relevant mathematical symbols are given. $[X]$ represents the integer set $\{1, 2, \ldots, X\}$. The text gives the representation of the knowledge graph: $G = \{(h, r, t)\}$, where the head entities $h \in [V_e]$, relations $r \in [V_r]$, and tail entities $t \in [V_e]$, V_r is the number of relations in the knowledge graph, and V_e means the number of entities. If entities are uniformly represented by e, and relations are represented by r, the k-dimensional embedding vectors are represented by $\vec{e}$ and $\vec{r}$, respectively.

Definition 3.1. *Knowledge graph G. Knowledge graph is a directed graph, denoted as* $G = \{E, R, T\}$*, nodes are entities, and edges correspond to triple facts* $\langle h, r, t\rangle$*. Each edge* r *of* $\langle h, r, t\rangle$ *indicates there is a relation* r *from* h *to* t*.* E *and* R *respectively represent the entity collection and the relation collection, and* T *represents a set of fact triples. Therefore, the* G *can also be expressed as* $G = \{E, R, T\} = \{\langle h, r, t\rangle | h, t \in E, r \in R\}$.

Definition 3.2. *Relational reasoning. Given the* G *and support set* $S \in \{\langle h, r, t\rangle\}$ *of entities* h*,* t*, a relational reasoning task: given ternary* $\langle h, ?, t\rangle$*, reasoning "?".*

3.2. **Algorithm description.** A graph attention network is one graph neural network, which mainly assigns different weights to the neighbors of each node. Inspired by the information processing mechanism of humans in cognitive science, the attention mechanism is introduced into the graph neural network to improve the processing efficiency of the network. The core of the graph attention network is the attention mechanism, and the key of the attention mechanism is to assign different weights to different nodes.

The proposed method mainly considers the influence of neighbors in the knowledge graph on the embedding representation. Based on the graph convolutional neural network, an attention layer is added, and the neighbors of each node are assigned different weights according to the characteristics of the node neighbors. Calculate the attention value, in accordance with the idea of weighted summation for the attention mechanism, calculate the new embedding representation of the entity (node), and use the new embedding of the entity for reasoning. This reasoning method only relies on entities and neighbors in the scoring function and ignores another important information in the knowledge graph, namely relation, which is also a problem in many current reasoning models.

The presented reasoning algorithm combines the quantification idea, and still considers the two important structural information of the known relation between the neighbors of the triplet and the entity, which more accurately expresses the embedding of the triplet, and makes the inference more accurate. The quantization idea in the paper is mainly embodied in the original data encoding process. By mapping the vectorization tuples within a certain range, the running speed is improved. Through changing the nonlinear function of multi-head attention, the problem of non-zero average value is avoided, the loss function is changed, and the convergence time is shortened. The flowchart of the relational reasoning algorithm based on the graph attention network is shown in Figure 1.

Figure 1 shows the execution time of the relational reasoning algorithm is affected by the number of iterations and the loss value. The greater the number of iterations, the more unstable loss value and the longer the execution time, and vice versa. Among them, neighbors and relations are particularly important in the calculation of the loss value, so the time complexity of the proposed algorithm is $O(d*(Ne+Nr))$, where d is the number of iterations, Ne represents the number of input entities and neighbors, and Nr is the number of relations.

The real-time performance is reflected in the convergence time of the loss value. The quantization method is to map the vectorization value to a certain range through a nonlinear function. The algorithm appropriately adjusts the multi-head attention nonlinear function and changes the loss function while considering neighbors and relations.

The input layer includes three embedding matrices, which are the original triple entity embedding matrix E, the neighbor entity embedding matrix H, and the relation embedding matrix R of the connected entities. The output layer is the new embedding representation processed by the attention layer. The three embedding matrices adopt non-linear transform as shown in Equation (3). According to [12], the mapping method

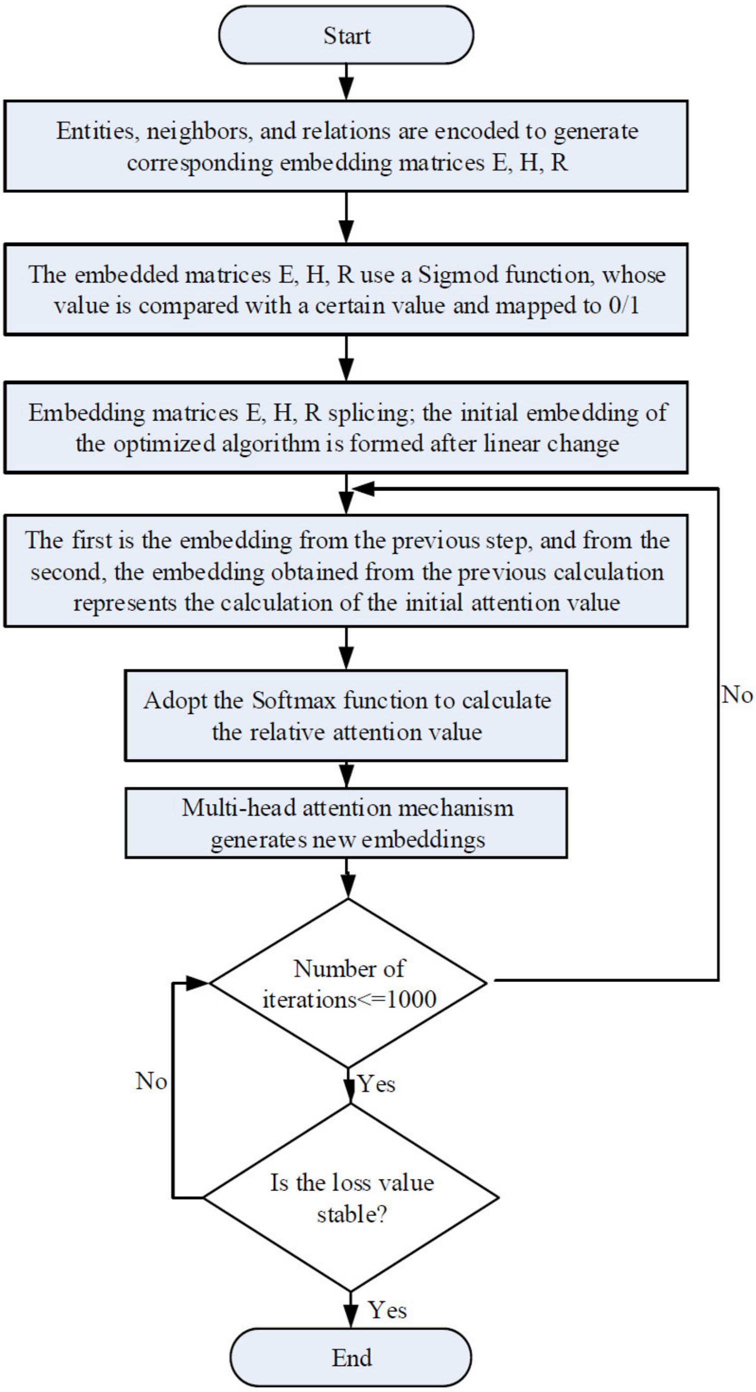

FIGURE 1. Flowchart of the relational reasoning algorithm based on graph attention network

is similar to the binarization method. This paper maps the range to 0/1.

$$p = 1/\left(1 + e^{h*w}\right) \tag{3}$$

where h represents the input vector, w is the weight value, and p represents the floating-point mapping value. The value obtained by Equation (3) is compared with a threshold value of 0.5, and the value is mapped to 1 if the value is greater than this value, and the value is mapped to 0 if the value is smaller than the value.

To obtain the new embedding of the entity e_i and learn the embedding representation of the neighbors, this paper concatenates the mapping values of the entities, neighbors, and relations, and obtains the initial embedding representation through linear changes, and the calculation is shown in Equation (4).

$$\vec{e}_{ijk} = W_1 \left(\vec{e}_i; \vec{h}_j; \vec{r}_k\right) \tag{4}$$

$\vec{e}_{ijk}$ is the vector representation of triples. The vector representations of head entity, tail entity and relation are $\vec{e}_i$, $\vec{h}_j$ and $\vec{r}_k$, respectively, and W_1 is linear transition matrices. According to [13], the calculation equation is shown in Equation (5).

$$b_{ijk} = \text{LeakyReLU}\left(W_2 \cdot c_{ijk}\right) \tag{5}$$

where '$\cdot$' represents a product operation, and b_{ijk} is the attention value from entity i to entity j and whose relation is k. The attention value is unified and normalized to achieve better weight distribution, and the relative attention value is calculated by Equation (6).

$$\alpha_{ijk} = \text{soft}\max x_{jk}\left(b_{ijk}\right) = \exp\left(b_{ijk}\right) \Bigg/ \sum_{n \in N_l} \sum_{r \in R_m} \exp\left(b_{inr}\right) \tag{6}$$

N_l represents the neighbors of the entity e_i, and R_m is the relation set connecting the entity e_i and entity h_j. The new embedding of the entity e_i is the sum of the embedding representations of each triplet, weighted according to its attention value, and the preliminary new embedding is calculated by the tanh function, as shown in Equation (7).

$$\vec{x}_i' = \tanh\left(\sum_{j \in N_l} \sum_{k \in R_{ij}} \alpha_{ijk} \cdot \vec{c}_{ijk}\right) \tag{7}$$

The algorithm adopts a multi-head attention mechanism, and each layer separately learns the vector representations of different entities and then connects them. The calculation is shown in Equation (8), and || indicates the connection operation with the same meaning as the symbol in the second subsection.

$$\vec{x}_i = \mathop{\Big\|}_{m=1}^{M} \tanh\left(\sum_{j \in N_l} \alpha_{ijk}^m c_{ijk}^m\right) \tag{8}$$

In this paper, the score function is shown in Equation (9), using different measures of L1-norm to embed the relation, where different measures of L1-norm can be calculated by $d_{t_{ij}} = \left\|\vec{e}_i + \vec{r}_k - \vec{h}_j\right\|_1$. Adopt Equation (10) as the loss function of the training model.

$$f\left(t_{ij}^k\right) = \left(\mathop{\Big\|}_{m=1}^{M} \text{ReLU}\left(\left[\vec{e}_i, \vec{h}_j, \vec{r}_k\right] * w^m\right)\right) \cdot W \tag{9}$$

$$L = \sum_{t_{ij}^k \in \{s\}} \log\left(1 + e^{f\left(t_{ij}^k\right)} + \sum_{t_{ij}^k \in \{s\}} \log\left(1 + e^{-f\left(t_{ij}^k\right)}\right)\right) + \frac{\lambda}{2}\|W\|_2^2 \tag{10}$$

In Equation (9), M represents the number of cores, w^m represents the weight of the m-th convolution kernel, $*$ is the convolution operation, and W is the linear transformation matrix. In Equation (10), λ is the learning rate, and s is the valid triplet.

4. **Experiment.**

4.1. **Experiment dataset.** The experiment data adopt the lung cancer dataset constructed in [14], NELL-995, WN18RR and FB15k-237, as shown in Table 1. The datasets have a sufficient number of triples and similar structures, which can be used for relational reasoning tasks. In Table 1, #Rel and #Ent represent the number of relationships and entities. #Train, #Valid and #Test are training, validation and test datasets, respectively.

For all datasets, the stochastic gradient descent (SGD) method is adopted in the experiment for small batches, and the maximum iterative training is 1000 times. Three common evaluation indicators are used including average ranking (MR) and Hits@k (the ratio of

Table 1. Experiment datasets

Datasets	#Rel	#Ent	#Train	#Valid	#Test
ClinicalM	188	14499	78657	5000	5000
NELL-995	200	75492	149678	543	3992
WN18RR	11	40943	86835	3034	3134
FB15k-237	237	14541	272115	17535	20466

valid test triples in the top k). Like the TransE, a self-confrontation sampling strategy [15] is used to replace its head entity or tail entity to generate all possible negative triples.

4.2. **Experiment and analysis.** Four classic methods such as TransE, ComplEx, ConvE and R-GCN are adopted as comparison methods. In the four methods, the input triples are embedded and coded, that is, converted into vectors, and then inferred. The first three scoring functions are closely related to the entities and relations of the input triples. R-GCN not only considers the input triples, but also adds its neighbor information. The proposed method is similar to the above four methods in this respect, but the initial reasoning in this paper is to quantify the input value that is mapped to a certain range and change the multi-head attention function to improve the expressive ability of the network.

The four classical reasoning methods are compared with the presented reasoning algorithm in relational reasoning tasks. If there are damaged triples in the triples of the knowledge graph, that is, the triples are correct, it should be reasonable for such triples to be ranked before the original triples. To eliminate the influence of this factor, in the experiment, before getting the ranking of each test triplet, these "disturbing" damaged triples were removed from the training set, validation set, and test set, thereby ensuring the damage triplet. Tuples do not belong to any datasets. This operation is called a Filt operation, and the one that has not been processed is called a Raw operation. In the two experimental settings, a lower MR and higher Hits@k mean better experiment results. Table 2 and Table 3 respectively show the results of relational inferences on the four datasets for the Filt operation and Raw operation. The bold scores in the table indicate the best inference result, and the slope scores show the second-best inference result.

Table 2 shows the results of TransE, ComplEx, ConvE, R-GCN and the proposed algorithm on the NELL-995 and WN18RR. The proposed algorithm is more accurate on the NELL-995. The result of removing the "disturbing" triplet is better than that of not removing it. After the Filt operation, the inferred result is the highest score. The value of MR is 2056, and the values of Hits@1, Hits@3, and Hits@10 are 45.5%, 54.3%, and 60.9%, respectively. ConvE ranked second in the evaluation index Hits@10, with the highest Hits@10 being 63.3%. On the WN18RR, the performance of the proposed algorithm is poor, and it is not even ranked in the top two on Hits@1 and Hits@3. The

Table 2. Relational reasoning results on the NELL-995 and WN18RR

Datasets	NELL-995								WN18RR							
Algorithms	MR		Hits@k (%)						MR		Hits@k (%)					
			@1		@3		@10				@1		@3		@10	
	Filt	Raw	Filt	Raw	Filt	Raw	Filt	Raw	Filt	Raw	Filt	Raw	Filt	Raw	Filt	Raw
TransE	*2100*	*2100*	34.4	34.9	47.2	47.4	50.1	50.6	*2300*	*2436*	4.27	4.92	44.1	42.67	**53.2**	51.97
ComplEx	4600	4730	39.9	40.4	52.8	**53.3**	60.6	60.8	7882	8018	*40.9*	**42.55**	*46.9*	*45.47*	53	**52.77**
ConvE	3560	**2098**	*40.3*	*40.8*	*53.1*	*53*	**63.3**	**61.8**	4464	4600	**43.9**	*41.55*	**47**	**45.57**	53.1	51.57
R-GCN	7600	7600	8.2	8.7	12.6	12.8	18.8	19.3	6700	6836	8	8.65	13.7	12.27	20.7	19.47
The proposed algorithm	**2056**	2210	**45.5**	**44**	**54.3**	50.5	*60.9*	*61.4*	**540**	**676**	15.6	16.25	31.2	29.77	*53.1*	*51.87*

TABLE 3. Relational reasoning results on the FB15k-237 and ClinicalM

Datasets	FB15k-237								ClinicalM							
Algorithms	MR		Hits@k (%)						MR		Hits@k (%)					
			@1		@3		@10				@1		@3		@10	
	Filt	Raw	Filt	Raw	Filt	Raw	Filt	Raw	Filt	Raw	Filt	Raw	Filt	Raw	Filt	Raw
TransE	323	501	19.8	**30.42**	**40.5**	38.83	44.1	42.71	6.76	22.06	0.4	2.5	0.67	0.55	0.88	0.65
ComplEx	546	724	19.4	19.22	29.7	**40.03**	45	43.61	*2.44*	*17.74*	72.8	74.9	80.5	80.38	*90.1*	*87.87*
ConvE	**230**	**408**	**30.6**	19.42	34.1	32.43	**52.96**	*51.57*	8.96	24.26	*73.5*	**79.6**	*82.16*	*82.04*	90	87.77
R-GCN	600	778	10	9.82	18.1	16.43	30	28.61	25.88	41.18	2.5	4.6	8.8	8.68	31.6	31.37
The proposed algorithm	*245*	*423*	*22.5*	*22.32*	*37.6*	*35.93*	*49.7*	**54.31**	**1.99**	**17.29**	**80.1**	*78.2*	**90.1**	**89.98**	**92.4**	**89.17**

main reason is that there are many entities in the dataset, but there are few relations, and the auxiliary information provided is also the corresponding reduction.

From Table 3, this reasoning ability of the proposed algorithm is better than the other four methods. The experiment results of the Filt operation are better than those of the Raw experiments. For the Filt operation, on the FB15k-237, the MR value and Hits@k value of the ConvE method are better than the proposed algorithm. The presented algorithm ranks second, but it is similar to the ConvE method. On the new dataset ClinicalM, the presented algorithm is significantly better than the other four methods. The values of MR, Hits@1, Hits@3, and Hits@10 are 1.99, 80.1%, 90.1%, and 92.4%, respectively. Observing the inference scores of the four datasets, it performs better on the new dataset ClinicalM, and shows the presented algorithm is more suitable for reasoning in the specific fields.

To further analyze the effect of the relational reasoning algorithm based on the graph attention network, the relation between the loss value and the number of iterations is given. This paper sets 1000 iterations, and averages the loss value every 100 iterations to obtain a graph of the loss value with the number of iterations as shown in Figure 2.

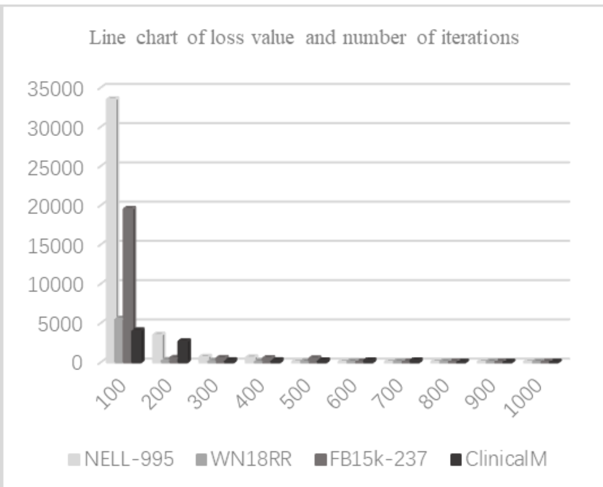

FIGURE 2. The relation between the loss value and the number of iterations

It can be seen that as the number of iterations increases, the loss value gradually decreases and tends to be stable. During the execution of 200 times, the loss value dropped sharply, and then stabilized after 300 times. The loss values on the four datasets are similar, and the loss value is the lowest on the ClinicalM, which is 79.53.

From the MR, Hits@k ($k = 1, 3, 10$) and the relation between loss value and the number of iterations, the proposed algorithm has better performance, especially suitable for knowledge graph reasoning in the specific fields such as other disease knowledge graph reasoning and financial domain knowledge graphs. These knowledge graphs have rich known relations and are more beneficial to accurately mining hidden relations.

5. **Conclusions.** This paper adopts an entity alignment algorithm based on semantic aggregation and attributes attention to solve heterogeneous problems. Aiming at the current reasoning algorithms without making full use of the knowledge graph entities and relations, and the real-time reasoning, a relational reasoning algorithm based on a graph attention network is proposed. Compared with other reasoning models, the proposed algorithm has advantages in reasoning in the specific fields. Future can consider the influence of paths on real-time knowledge reasoning, and realize the reasoning of complex relations.

Acknowledgment. This work is partially supported by the National Natural Science Fund of China (No. 61371143), the National Key Research and Development Program of China (No. 2020YFC0811004), the Basic Research Project of Shanxi Province (No. 202103 02123216), Science and technology project of Beijing Municipal Education Commission (No. KM202110009002). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] M. Hildebrandt, J. Serna, Y. Ma et al., Reasoning on knowledge graphs with debate dynamics, *AAAI 2020*, pp.4123-4131, 2020.

[2] K. Bollacker, C. Evans, P. Paritosh et al., Freebase: A collaboratively created graph database for structuring human knowledge, *Proc. of the 2008 ACM SIGMOD International Conference on Management of Data*, pp.1247-1250, 2008.

[3] A. Carlson, J. Betteridge, B. Kisiel et al., Toward an architecture for never-ending language learning, *AAAI 2010*, pp.1306-1313, 2010.

[4] G. A. Miller, WordNet: A lexical database for English, *Communications of the ACM*, vol.38, no.11, pp.39-41, 1995.

[5] Y. Wang, Z. Sun and Y. Han, Network attack path prediction based on vulnerability data and knowledge graph, *International Journal of Innovative Computing, Information and Control*, vol.17, no.5, pp.1717-1730, 2021.

[6] T. Li, *Research on Knowledge Representation and Reasoning Algorithm Based on Deduction Lattice in Clinical Decision Support System*, Master Thesis, Wuhan University, 2019.

[7] N. Lao and W. W. Cohen, Relational retrieval using a combination of path-constrained random walks, *Machine Learning*, vol.81, no.2, pp.53-67, 2010.

[8] Y. Feng, X. Chen, B. Y. Lin et al., Scalable multi-hop relational reasoning for knowledge-aware question answering, *2020 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference (EMNLP2020)*, pp.1295-1309, 2020.

[9] M. Patacchiola and A. Storkey, Self-supervised relational reasoning for representation learning, *The 34th Conference on Neural Information Processing Systems (NeurIPS2020)*, pp.126-132, 2020.

[10] P. Veličković, G. Cucurull, A. Casanova et al., Graph attention networks, *International Conference on Learning Representations (ICLR)*, pp.891-897, 2018.

[11] Y. Fan, T. Yang et al., Mining method of telecommunication fraud communication characteristics based on knowledge graph, *Computer Applications and Software*, vol.36, no.11, pp.182-187, 2019.

[12] I. Hubara, D. Courbariaux, R. Soudry et al., Binarized neural networks, *Advances in Neural Information Processing Systems*, pp.4107-4115, 2016.

[13] Y. Zou, *Research on Semi-Automatic Construction of Chinese Medical Knowledge Graph*, Master Thesis, Wuhan University of Science and Technology, 2016.

[14] Y. Zhang and Z. Du, A real-time inference method of graph attention network based on knowledge graph for lung cancer, *2021 5th International Conference on Digital Signal Processing (ICDSP2021)*, pp.326-331, 2021.

[15] Z. Sun, Z. H. Deng et al., RotatE: Knowledge graph embedding by relational rotation in complex space, *ICLR (Poster)*, 2019.