



---

*Research article*

## **MRE: A translational knowledge graph completion model based on multiple relation embedding**

**Xinyu Lu, Lifang Wang, Zejun Jiang\*, Shizhong Liu and Jiashi Lin**

School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China

\* **Correspondence:** Email: [claud@nwpu.edu.cn](mailto:claud@nwpu.edu.cn).

**Abstract:** Knowledge graph completion (KGC) has attracted significant research interest in applying knowledge graphs (KGs). Previously, many works have been proposed to solve the KGC problem, such as a series of translational and semantic matching models. However, most previous methods suffer from two limitations. First, current models only consider the single form of relations, thus failing to simultaneously capture the semantics of multiple relations (direct, multi-hop and rule-based). Second, the data-sparse problem of knowledge graphs would make part of relations challenging to embed. This paper proposes a novel translational knowledge graph completion model named multiple relation embedding (MRE) to address the above limitations. We attempt to embed multiple relations to provide more semantic information for representing KGs. To be more specific, we first leverage PTransE and AMIE+ to extract multi-hop and rule-based relations. Then, we propose two specific encoders to encode extracted relations and capture semantic information of multiple relations. We note that our proposed encoders can achieve interactions between relations and connected entities in relation encoding, which is rarely considered in existing methods. Next, we define three energy functions to model KGs based on the translational assumption. At last, a joint training method is adopted to perform KGC. Experimental results illustrate that MRE outperforms other baselines on KGC, demonstrating the effectiveness of embedding multiple relations for advancing knowledge graph completion.

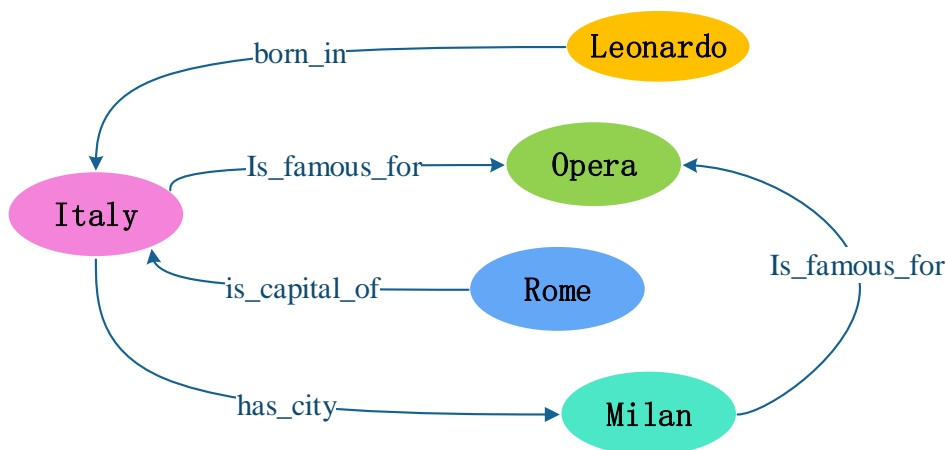
**Keywords:** knowledge graph completion; knowledge graphs; multiple relations; entities; semantic information

---

### **1. Introduction**

A knowledge graph (KG) [1,2] is a semantic network designed to describe entities of the real world and the relations between various entities, sometimes referred to as the Knowledge Base (KB) [3,4]. A KG also can be denoted as a triple [5] indicating that the head entity and the tail entity are connected

by a relation, e.g., (head entity, relation, tail entity). However, even some large-volume KGs, such as Freebase [6] and NELL [7], are still incomplete, i.e., missing a lot of correct KGs. Thus, many researchers have paid massive attention to knowledge graph completion (KGC) to validate whether a KG is correct or not. A KGC task is one of the research directions of knowledge representation learning (KRL) [8–11] that aims at validating whether a triple is correct or not while preserving the inherent structure of the KGs. Examples of knowledge graphs are illustrated in Figure 1. The colored ellipses represent entities, and arrows connect the relations from the head entities to the tail entities.



**Figure 1.** An example of knowledge graphs. Nodes (ellipses) represent entities, and edges (arrows) represent relations.

Most of the available methods perform the KGC tasks by embedding components of KGs into continuous vector spaces through two steps: embedding KG components (entities and relations) and then defining a scoring function to measure the plausibility of each triple. The current KGC models can be categorized as translational and semantic matching models. Translational models take relations to translate head entities to tail entities, such as TransE [12], TransH [13] and TransR [14]. Translational models can take advantage of the transitional character of KG components, but they only use addition operations, which limits the expressive power of KGC models. Semantic matching models match latent semantics of KG components in the embedding space, such as RESCAL [15] and DistMult [16]. Semantic matching models can capture the semantic similarities among different triples, but some models with fully connected layers usually cause overfitting problems. Consequently, some convolution neural network (CNN)-based methods [17, 18] have been proposed to perform KGC tasks. They can capture deep expressive features and alleviate overfitting problems.

Recently, some researchers have attempted to use path information to improve the performance of knowledge graph completion models. A path starts along the head entity, passes through intermediate entities, and reaches the tail entity. Taking a 2-hop path as an example, the path can be composed of two multi-hop relations and an intermediate entity, i.e.,  $r_1 \rightarrow e \rightarrow r_2$ , where  $r_1$  and  $r_2$  are 2-hop relations, and  $e$  denotes a intermediate entity. PTransE [19] is a typical translational model that performs simple addition operations on multi-hop paths to complete knowledge graphs. Other methods like RUGE [20] can inject logic rules [21] into the model and use the rules to guide the representation learning of

entities and relations. The rules are interpretable and contain rich semantic information [22], showing the power of knowledge reasoning. The process of reasoning can be defined as  $r \Leftarrow (r_3, r_4)$ , where  $r_3$  and  $r_4$  denote rule-based relations, and  $r$  can be inferred by these two rule-based relations.

Even though the above methods can achieve good performance, it remains challenging to conduct KGC tasks. There are two limitations: First, most knowledge graph completion models only consider the single form of relations since they only embed direct relations or multi-hop paths to capture semantic information of relations. From the perspective of natural language processing, different forms of the same relation are semantically related and complementary. Therefore, multiple (direct, multi-hop, and rule-based) relations can potentially be exploited to capture adequate semantic information and advance the performance of KGC. Second, the data-sparse problem of knowledge graphs would make part of relations challenging to embed since very little data is involved. To address these issues, we propose a novel knowledge graph completion model called multiple relation embedding (MRE), which can simultaneously embed multiple relations and jointly learn embeddings of entities and relations.

In this paper, we reconfirm that translational KGC methods are effective and apply the main idea of TransE [12] in our proposed work. MRE aims to take advantage of multiple relations to perform knowledge graph completion. Specifically, MRE mainly consists of four steps. MRE first extracts multi-hop and rule-based relations through corresponding tools. Then, MRE proposes different encoders to encode multi-hop and rule-based relations. Note that our proposed encoders of multiple relations do not learn each relation in isolation but continuously interact with entities during the learning process and use pre-trained embeddings as supervision to maximize the restoration of the semantic information of multiple relations. Next, MRE defines new energy functions to model knowledge graphs. Finally, a joint training method is adopted to train our proposed MRE.

In summary, the proposed KGC method can better use multiple relations and fully capture semantic information of knowledge graphs. Our contributions can be summarized as follows:

- We propose a novel method for knowledge graph completion that simultaneously embeds multiple relations (direct, multi-hop and rule-based) in a unified framework and defines new energy functions to learn representations of entities and relations.
- Our proposed multiple relation encoders can continuously interact with connected entities in the process of relation encoding, which is rarely considered in most knowledge graph completion methods.
- This paper evaluates the MRE on two benchmark datasets of FB15K-237 and NELL-995 with knowledge graph completion. Experimental results show that MRE has achieved the best performance on all evaluation metrics compared to several baselines.

The rest of this paper is organized as follows: Section 2 summarizes the related work of knowledge graph completion. Section 3 introduces some notations and definitions used in this paper. Section 4 introduces the details of our proposed MRE. Section 5 shows and analyzes our experimental results, including comparisons with current methods, further evaluations and visualization analysis. Section 6 concludes our work and points out future research directions.

## 2. Related work

This section introduces KGC models that embed KGs into continuous low-dimensional spaces to capture latent semantic representations, including translational and semantic matching methods.

**Translational methods:** Translational models leverage distance-based scoring functions and measure the plausibility of a triple as the distance between head entities and tail entities, specifically after the relations through a translation. TransE [12] is inspired by word2vec, where relations can be regarded as translations between head entities and tail entities in implicit semantic embedding space. However, TransE has its drawbacks in handling complex relations, such as one head entity and one relation corresponding to multiple tail entities. To tackle this problem, TransH [13] is proposed, in which each relation has a specific hyperplane while projecting the head entity and the tail entity onto the hyperplane. Since there may be infinite hyperplanes in each relation, TransH uses approximate orthogonality to select a hyperplane which may prevent the model from dealing with entities and relations properly. TransR [14] extends TransH [13] by using relation-specific spaces to complete a KGC task. PTransE [19] is an extension of TransE which proposes a path constraint resource allocation method to extract multi-hop paths and compose all the relations in each multi-hop path.

**Semantic matching methods:** Semantic matching models use similarity-based scoring functions that measure the plausibility of triples through matching implicit semantics of KGs embedded in continuous low-dimensional vector space. RESCAL [15] can capture KGs' implicit semantic information by associating each entity with a vector and each relation with a matrix to realize bilinear interactions. DistMult [16] is based on RESCAL [15], and each relation is restricted by a diagonal rather than a full matrix. DistMult can capture bilinear interactions between head and tail entities through the same embedding space and reduce training parameters. HOLE [23] represents both entities and relations as vectors by utilizing the circular correlation operation based on RESCAL and DistMult. This allows the model to deal with irreflexive or similar relations in KGs.

In addition to the above methods, there are still many ways to conduct knowledge graph completion tasks, such as RUGE [20], which iteratively learns entity and relation embeddings from labeled triples, unlabeled triples and soft rules, or MADLINK [24], which considers contextual information as well as the textual descriptions of the entities to perform KGC.

Although the experimental results of the above models are all impressive, they only use direct relations or multi-hop paths, limiting the models' representational power. Translational methods are one of the most valuable methods for KGC. Thus, our work, while building upon translational methods, is distinguished based on the following properties:

- Unlike most KGC methods that only embed single-form relations, MRE simultaneously embeds multiple relations and obtains multiple semantics of relations.
- Different from current KGC methods, MRE considers an interactive process in the relation encoding phase to obtain accurate relation representations.

## 3. Background

**Notations:** Table 1 shows the important symbols of this article. This paper uses lower-case boldface letters to denote vectors (e.g.,  $\mathbf{h}$ ) and boldface upper-case letters to represent matrices (e.g.,  $\mathbf{W}$ ). We use  $\|\mathbf{x}\|$  to denote the  $\ell_2$  norm.

**Knowledge graph:** For a given knowledge graph  $\mathcal{G} = \{\mathbf{E}, \mathbf{R}\}$ , we represent entities by  $\mathbf{E}$  and relations by  $\mathbf{R}$ . A KG  $\mathcal{G}$  includes many factual triples, and each triple is in the form of  $(\mathbf{h}, \mathbf{r}, \mathbf{t})$ , with  $\mathbf{h}, \mathbf{t} \in \mathbf{E}$  and  $\mathbf{r} \in \mathbf{R}$ .

**Knowledge graph completion:** Knowledge graph completion tasks can be carried out simply by predicting if a triple  $(h, r, t)$  is valid or not. In our work, valid triples obtain lower energy (scores) than invalid triples.

**Multiple relations:** Multiple relations consist of direct relations, multi-hop relations and rule-based relations. The direct relations are derived from the given knowledge graphs. The multi-hop and rule-based relations are extracted from the given knowledge graphs.

**Table 1.** A summary of symbols used in this paper.

Symbols	Notations
$\mathbf{E}$	a set of entities that include head and tail entities
$\mathbf{R}$	a set of relations
$\mathcal{G}$	a knowledge graph which can be denoted as $\{\mathbf{E}, \mathbf{R}\}$
$(h, r, t)$	a fact which contains a subject entity, a relation, and an object entity
KGC	knowledge graph completion
KRL	knowledge representation learning
CNN	convolution neural network
MLP	multilayer perceptions
MRE	multiple relation embedding

## 4. Proposed work

Knowledge graph completion aims to predict whether a triple is correct or not. Although much research has been devoted to KGC, most models cannot take advantage of the multiple relations and fail to capture multiple semantic information. This paper proposes a novel knowledge graph completion method called MRE to compensate for existing methods' deficiencies. We attempt to simultaneously embed multiple relations (direct, multi-hop and rule-based) in our model. The overall architecture of MRE is shown in Figure 2. First, we extract different kinds of relations from given triples (4.1). After vector initialization, we propose different encoders to capture multi-hop and rule-based semantics of relations (4.2). Furthermore, we embed multiple relations into the same semantic space based on the translational assumption (4.3). Finally, a joint training method is implemented for optimizing the objective of MRE (4.4). The following section shows the details of our work.

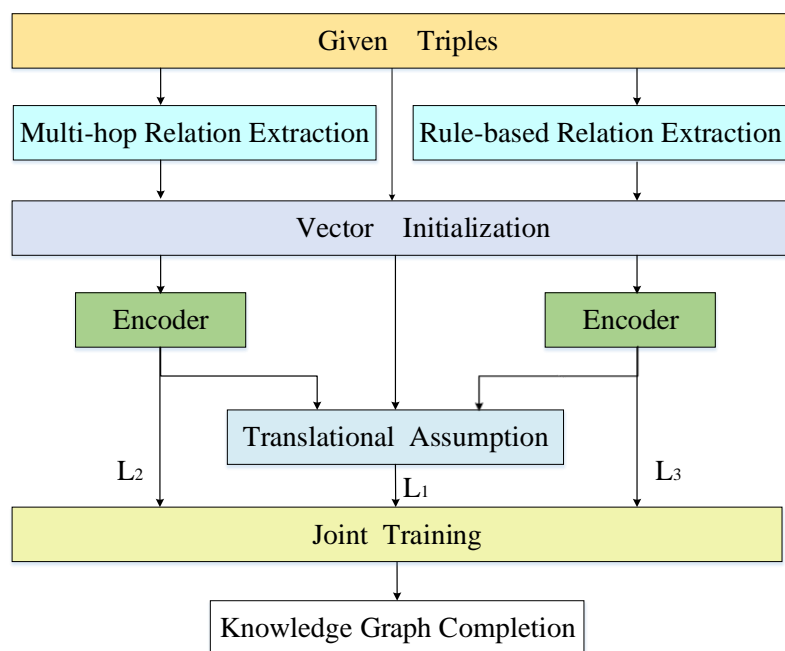
### 4.1. Compositions of multiple relations

In this paper, multiple relations include direct, multi-hop and rule-based relations. The direct relations come from given triples. The multi-hop and rule-based relations are derived from the following procedures.

**Multi-hop relations:** We follow the paths extraction procedure provided by PTransE [19] to extract multi-hop paths. In PTransE, each multi-hop path is extracted together with its reliability, which is achieved by the path-constraint resource allocation mechanism. The path-constraint resource allocation mechanism can extract different paths and compute a reliability metric for each path that flows from

the head entity to the tail entity. A multi-hop path can be linked by a head entity, multi-hop relations, intermediate entities and a tail entity. For a given entity pair, we manually select the path with the largest reliability metric as the optimal path and limit the path length to 2. In this paper, we regard path-determined relations as multi-hop relations. Given a triple  $(h, r, t)$ , a 2-hop path  $r_1 \rightarrow e \rightarrow r_2$  can be generated from  $(h, r_1, e)$  and  $(e, r_2, t)$ , where  $r_1$  and  $r_2$  are 2-hop relations, and  $e$  denotes an intermediate entity.

**Rule-based relations:** Rule-based relations can be generated using the rule extraction tool AMIE+ [25] which extracts relations based on horn rules. A horn rule can be defined as  $r \Leftarrow (r_3, r_4)$ , where  $r_3$  and  $r_4$  denote rule-based relations. Each rule has a confidence level to measure the matching degree of the rule. The higher the confidence level is, the higher the matching degree of the rule. In this paper, we limit the length of rules to 2 and select the highest confidence level to extract rule-based relations. For the relation ‘/film/film/country’, there are two horn rules:  $/film/film/country \Leftarrow (/film/film/executive\_produced\_by, /film/film/film\_format)$  and  $/film/film/country \Leftarrow (/film/film/music, /people/person/nationality)$ . Since the confidence level of the former is greater than that of the latter, we select the former as the rule-based relations, i.e.,  $/film/film/executive\_produced\_by$  and  $/film/film/film\_format$  are rule-based relations.



**Figure 2.** The overall architecture of MRE.  $L_1$  denotes pairwise ranking loss,  $L_2$  denotes error loss of multi-hop encoder, and  $L_3$  denotes error loss of rule-based encoder.

#### 4.2. Encoders of multiple relations

Through corresponding relation extraction procedures, we can obtain multi-hop and rule-based relations. To make better use of multiple relations, we propose two encoders that can encode different kinds of relations and achieve interactions between entities and relations.

#### 4.2.1. Multi-hop relation encoding

A multilayer perceptron (MLP) [26] is a simple feedforward neural network that can map a set of input vectors to output vectors and learn the best of the parameters of neural networks. To encode multi-hop relations, we use two MLP structures that load pre-trained embeddings as inputs and achieve multiplicative interactions between MLP entities and relations. The encoder of multi-hop relations is shown in Figure 3. The specific process is as follows:

$$\mathbf{r}'_1 = \sigma(\mathbf{W}_{11}\mathbf{h}\mathbf{r}_1 + \mathbf{b}_{11}) \quad (4.1)$$

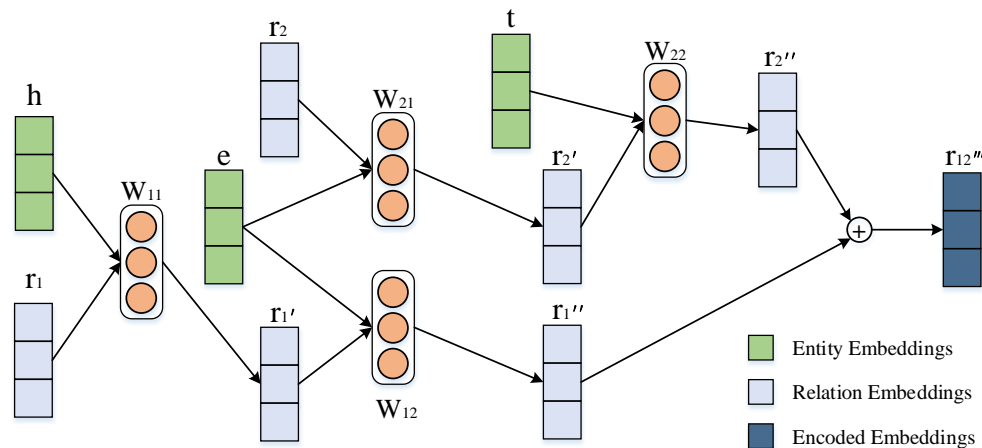
$$\mathbf{r}''_1 = \sigma(\mathbf{W}_{12}\mathbf{r}'_1\mathbf{e} + \mathbf{b}_{12}) \quad (4.2)$$

$$\mathbf{r}'_2 = \sigma(\mathbf{W}_{21}\mathbf{e}\mathbf{r}_2 + \mathbf{b}_{21}) \quad (4.3)$$

$$\mathbf{r}''_2 = \sigma(\mathbf{W}_{22}\mathbf{r}'_2\mathbf{t} + \mathbf{b}_{22}) \quad (4.4)$$

$$\mathbf{r}''_{12} = \mathbf{r}''_1 + \mathbf{r}''_2 \quad (4.5)$$

where  $\mathbf{r}_1$  represents the first multi-hop relation embedding, and  $\mathbf{h}$  represents a head entity embedding.  $\mathbf{W}_{11}$ ,  $\mathbf{W}_{12}$ ,  $\mathbf{W}_{21}$  and  $\mathbf{W}_{22}$  are the weights of MLP structures, and  $\mathbf{b}_{11}$ ,  $\mathbf{b}_{12}$ ,  $\mathbf{b}_{21}$  and  $\mathbf{b}_{22}$  are the biases of MLP structures.  $\sigma$  denotes a non-linear activation (Tanh),  $\mathbf{r}_2$  represents the second multi-hop relation embedding,  $\mathbf{e}$  represents an intermediate entity embedding,  $\mathbf{t}$  represents a tail entity embedding, and  $\mathbf{r}''_{12}$  denotes the encoded multi-hop relation embedding. It can be seen from the above formula that the multi-hop relation embeddings are obtained by encoding complete link information, since we attempt to capture the semantic information before and after each multi-hop relation.



**Figure 3.** The architecture of the encoder for multi-hop relations. Squares represent embedding vectors, and circles represent neural networks. The encoding process of multi-hop relations is realized by continuously interacting entities and relations through MLP.

#### 4.2.2. Rule-based relation encoding

Similarly, we use two MLP structures to obtain rule-based relation embeddings. The encoder of rule-based relations is shown in Figure 4. The specific process is as follows:

$$\mathbf{r}'_3 = \sigma(\mathbf{W}_{31}\mathbf{r}_3 + \mathbf{b}_{31}) \quad (4.6)$$

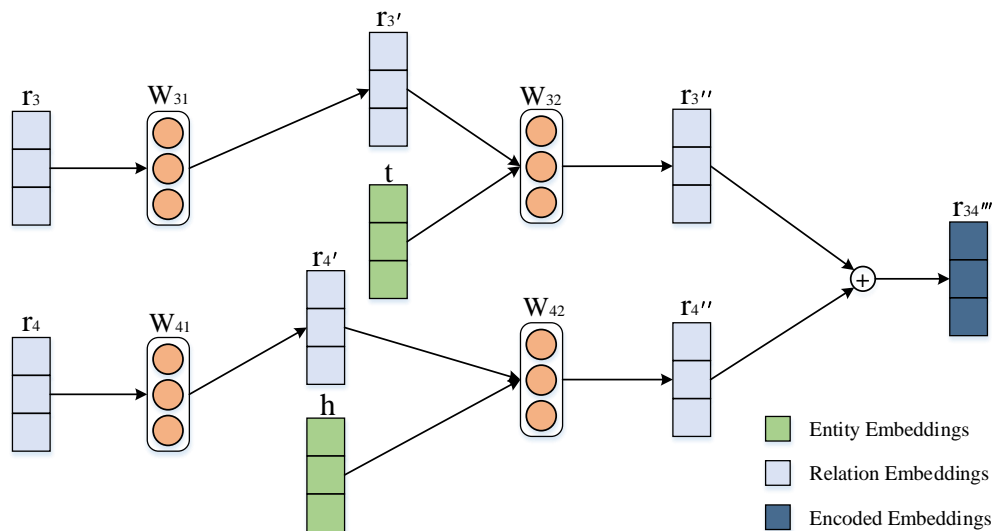
$$\mathbf{r}''_3 = \sigma(\mathbf{W}_{32}\mathbf{h}\mathbf{r}'_3 + \mathbf{b}_{32}) \quad (4.7)$$

$$\mathbf{r}'_4 = \sigma(\mathbf{W}_{41}\mathbf{r}_4 + \mathbf{b}_{41}) \quad (4.8)$$

$$\mathbf{r}''_4 = \sigma(\mathbf{W}_{42}\mathbf{r}'_4\mathbf{t} + \mathbf{b}_{42}) \quad (4.9)$$

$$\mathbf{r}'''_{34} = \mathbf{r}''_3 + \mathbf{r}''_4 \quad (4.10)$$

where  $\mathbf{r}_3$  represents the first rule-based relation embedding.  $\mathbf{W}_{31}$ ,  $\mathbf{W}_{32}$ ,  $\mathbf{W}_{41}$  and  $\mathbf{W}_{42}$  are the weights of MLP structures, and  $\mathbf{b}_{31}$ ,  $\mathbf{b}_{32}$ ,  $\mathbf{b}_{41}$  and  $\mathbf{b}_{42}$  are the biases of MLP structures.  $\mathbf{r}_4$  represents the second rule-based relation embedding, and  $\mathbf{r}'''_{34}$  denotes the encoded rule-based relation embedding. It can be seen from the above formula that the process of rule-based relations encoding is realized by the interactions between entities and relations in essence.



**Figure 4.** The architecture of the encoder for rule-based relations. Squares represent embedding vectors, and circles represent neural networks. The encoding process of rule-based relations is not affected by intermediate entities due to the composition of the rules.

### 4.3. Energy functions

TransE [12] is the most representative model in knowledge graph completion. Previous work shows that TransE and its extensions [13, 14] can obtain very competitive results. We reconfirm that TransE is a robust model and apply this structure to our method. In TransE, relations can be regarded as translations between head entities and tail entities in implicit semantic embedding space. The energy function of TransE is defined as

$$E_1 = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\| \quad (4.11)$$

where  $\mathbf{h}$  represents a head entity embedding,  $\mathbf{r}$  represents a relation embedding,  $\mathbf{t}$  represents a tail entity embedding, and  $E_1$  is an energy function of given triples.



Next, we transfer the energy function of given triples to the multi-hop information, incorporating the multi-hop relations under the translational assumption. The energy function can be defined as

$$E_2 = \|\mathbf{h} + \mathbf{r}_{12}'' - \mathbf{t}\| \quad (4.12)$$

where  $\mathbf{r}_{12}''$  represents an encoded multi-hop relation embedding. The energy function  $E_2$  can be regarded as an additional constraint for the given triples.

Following the above energy function  $E_2$ , we transfer the energy function of given triples to the rule-based information, incorporating the rule-based relations under the translational assumption. The energy function can be defined as

$$E_3 = \|\mathbf{h} + \mathbf{r}_{34}'' - \mathbf{t}\| \quad (4.13)$$

where  $\mathbf{r}_{34}''$  represents an encoded rule-based relation embedding. The energy function  $E_3$  can be regarded as another additional constraint for the given triples.

Therefore, the overall energy function of a triple can be defined as

$$E = E_1 + E_2 + E_3 \quad (4.14)$$

#### 4.4. Objective formalization

Based on the above energy functions, we propose a joint training method to take advantage of multiple relations and train our model. The overall loss function is defined as

$$\mathcal{L} = L_1 + L_2 + L_3 \quad (4.15)$$

where  $L_1$  denotes pairwise ranking loss,  $L_2$  denotes error loss of multi-hop encoder, and  $L_3$  denotes error loss of rule-based encoder. Given the positive triples  $\mathcal{G}$  and negative triples  $\mathcal{G}'$  constructed accordingly, we define the pairwise ranking loss as

$$L_1 = \sum_{(h,r,t) \in \mathcal{G}} \sum_{(h',r',t') \in \mathcal{G}'} \max(\gamma + E(h,r,t) - E(h',r',t'), 0) \quad (4.16)$$

where  $\gamma$  is a margin parameter that separates positive and negative triples. Following TransE [12], negative triples can be generated by changing the head entity or the tail entity at random, i.e.,

$$\mathcal{G}' = \{(h', r, t) \mid h' \in E\} \cup \{(h, r, t') \mid t' \in E\}, (h, r, t) \in \mathcal{G}. \quad (4.17)$$

Our method uses TransE to obtain pre-trained embeddings of entities and relations. In order to reduce the semantic errors of the encoders, we use pre-trained relation embeddings to supervise the encoded embeddings. The error loss function is defined as follows:

$$L_2 = \|\mathbf{r}_{12}'' - \mathbf{r}_{\text{pre}}\| \quad (4.18)$$

$$L_3 = \|\mathbf{r}_{34}'' - \mathbf{r}_{\text{pre}}\| \quad (4.19)$$

where  $\mathbf{r}_{\text{pre}}$  denotes pre-trained relation embeddings.

The central insight in developing MRE is as follows: Given three input embeddings, we view the KGC task as a predicting problem. To predict whether a triple is valid, MRE first selects multi-hop and rule-based relations through different extraction tools. Then, MRE encodes extracted relations by leveraging MLP structures. Next, MRE models the whole knowledge graphs based on the TransE. At last, the optimization objective is proposed to train MRE jointly.

## 5. Experiments

### 5.1. Datasets

The statistics of the two KGC datasets evaluated in this paper are given in Table 2. FB15K-237 and NELL-995 are created from FB15K [12] and NELL [7]. FB15K comes from the large real-world knowledge base Freebase [6]. FB15K-237 contains 237 relations, 14,541 entities and 310,116 triples; and the approximate ratio of the train set, valid set and test set is 14:1:1. NELL-995 is a subset of NELL created from the 995th iteration of the construction. NELL-995 includes 75,492 entities, 200 relations and 154,208 triples.

**Table 2.** The statistics of the KGC datasets.

Dataset	E	R	Train	Valid	Test
FB15K-237	14,541	237	272,115	17,535	20,466
NELL-995	75,492	200	123,370	15,000	15,838

### 5.2. Evaluation protocol

Following the convention, we employ mean reciprocal rank (MRR) and Hits@k as evaluation metrics. MRR is computed by the average reciprocal rank of correct entities. Higher MRR indicates better performance.

$$\text{MRR} = \frac{1}{|\mathcal{G}|} \sum_{i \in \mathcal{G}} \frac{1}{\text{Rank}(i)} \quad (5.1)$$

where  $|\mathcal{G}|$  represents the total number of triples, and  $\text{Rank}(i)$  represents the ranking of the correct label of the  $i$ -th test triple. Hits@k computes the proportion of correct entities that appear within the top-k predictions. Higher Hits@k indicates better performance.

$$\text{Hits @ k} = \frac{|\text{Rank-k}|}{|\mathcal{G}|} \quad (5.2)$$

where  $|\text{Rank-k}|$  represents the number of correct labels ranking in the top-k. MRR and Hits@k scores always range from 0 to 1.

### 5.3. Experimental settings

We use pre-trained embeddings to initialize our model. To obtain entity and relation embeddings, we follow the traditional settings initially provided in TransE [12]. After training, we can obtain entity and relation embeddings with embedding size  $k = 100$ . For our proposed MRE, we employed the path extracted method provided in PTransE to extract multi-hop relations. We employed AMIE+ to extract rule-based relations. The length of multi-hop and rule are limited to 2. For our proposed MRE, we use Adam [29] and apply an  $L_2$  norm for all the equations. The highest Hits@10 scores are obtained when using  $lr = 1e^{-4}$ , batch size  $b = 128$  and  $\gamma = 8$  on FB15K-237, and when using  $lr = 2e^{-3}$ , batch size  $b = 128$  and  $\gamma = 8$  on NELL-995.

#### 5.4. Results and discussion

To show the effectiveness and superiority of our proposed method, we list ten representative KGC algorithms as baselines. All the baselines have taken advantage of knowledge representation learning methods to advance the KGC tasks. These baselines contain translational methods TransE [12], TransH [13], TransR [14] and PTransE [19], rule-guided method RUGE [20], CNN-based method DMACM [18], hierarchical reasoning method ConE [27], path-based method MADLINK [24], Transformer-based method Hitter [28] and graph attention-based method MRGAT [30].

- TransE [12] is the most representative translational model, which embeds components of KGs in vector space and makes learned entity and relation embeddings follow the translational principle  $h + r = t$ .
- TransH [13] assigns each relation with a specific hyperplane and projects the head and tail entity onto the hyperplane.
- TransR [14] is an extension of TransH which introduces relation-specific spaces to complete a KGC task.
- PTransE [19] leverages multi-hop paths to complete knowledge graphs.
- RUGE [20] learns entity and relation representations through iterative guidance of soft rules.
- DMACM [18] captures directional information and the triple's inherent deep expressive characteristic using a CNN-based method.
- ConE [27] is a hierarchical reasoning method that embeds entities into hyperbolic cones and then models relations as conversions between the cones.
- MADLINK [24] incorporates path and contextual information in given knowledge graphs to learn embeddings.
- Hitter [28] can jointly learn entity and relation embeddings based on Transformer structure.
- MRGAT [30] proposes a multi-relational graph attention model to complete knowledge graphs.

The experimental results of different baselines on FB15K-237 and NELL-995 are listed in Table 3. From the results in Table 3, we have some observations. First, MRE consistently outperforms other baselines in MRR and Hits@10 on FB15K-237 and NELL-995. For the FB15K-237 dataset, MRE has achieved an improvement of  $0.399 - 0.373 = 0.026$  (+7%) in MRR and  $0.612 - 0.558 = 0.054$  (+9.7%) in Hits@10 when compared to the second-best results. For the NELL-995 dataset, MRE has achieved an improvement of  $0.327 - 0.318 = 0.009$  (+2.8%) in MRR and  $0.572 - 0.437 = 0.135$  (+31%) in Hits@10 when compared to the second-best results. This demonstrates the effectiveness of our proposed method and supports that embedding multiple relations in a unified semantic space is beneficial for knowledge graph completion. Second, we find that transitional methods [12–14, 19] can achieve competitive results on both datasets. This indicates the usefulness of considering transitional properties in KGC methods. Our proposed MRE differs from these transitional methods in that MRE can capture semantic information of multiple relations while maintaining translation characteristics between entities and relations. Third, we observe that Hitter [28] can obtain second-best MRR and Hits@10 on FB15K-237. However, Hitter leverages a complex structure to embed knowledge graphs. Compared with Hitter, MRE simply uses MLP structures to encode knowledge graphs. This shows the superiority of our method. In addition, MRE outperforms path-based models [19, 24] and the rule-guided method [20] mainly because our model can take advantage of multiple relations and provide more semantic information to complete knowledge graphs.

**Table 3.** Experimental results on FB15K-237 and NELL-995. MRR represents the mean reciprocal rank and Hits@10 represents the correct predictions appear within the top-10. The best result is in **bold**, while the second-best result is in underline.

Model	FB15K-237		NELL-995	
	MRR	Hits@10	MRR	Hits@10
TransE (2013) [12]	0.294	0.465	0.219	0.352
TransH (2014) [13]	-	-	0.223	0.358
TransR (2015) [14]	0.199	0.382	0.232	0.382
PTransE (2015) [19]	0.314	0.501	-	-
PTransE-RNN (2015) [19]	-	-	0.286	0.423
PTransE-ADD (2015) [19]	-	-	0.304	<u>0.437</u>
RUGE (2018) [20]	0.164	0.349	<u>0.318</u>	0.433
DMACM (2021) [18]	0.27	0.489	-	-
ConE (2021) [27]	0.345	0.540	-	-
MADLINK (2021) [24]	0.347	0.529	-	-
HittER (2021) [28]	<u>0.373</u>	<u>0.558</u>	-	-
MRGAT(2022) [30]	0.358	0.542	-	-
MRE(our)	<b>0.399</b>	<b>0.612</b>	<b>0.327</b>	<b>0.572</b>

### 5.5. Further evaluations

This paper proposes a multiple relation embedding method to make better use of different kinds of relations. In this section, we explore the impacts of margin, different numbers of relations and extra semantic information on our proposed MRE.

**Effect of margin** Evaluation results are shown in Table 4. From the table, we have two observations. First, we observe that, like most translational models [12–14, 19], our method is also affected by the margin, i.e., as the margin value changes, the experimental results of MRE will also fluctuate. Second, MRE performs well when the margin  $\gamma = 6/8/10$ . This indicates that the setting of a reasonable margin value is helpful for the model to achieve good performance.

**Table 4.** Effect of margin  $\gamma$  on FB15K-237 and NELL-995. The best result is in **bold**.

$\gamma$	FB15K-237					NELL-995				
	MRR	Hits@k				MRR	Hits@k			
		1	3	5	10		1	3	5	10
2	0.319	0.212	0.362	0.433	0.528	0.289	0.184	0.316	0.394	0.502
4	0.348	0.244	0.391	0.46	0.551	0.298	0.186	0.333	0.414	0.534
6	0.355	0.245	0.402	<b>0.474</b>	0.564	0.317	0.199	0.357	0.443	0.556
8	<b>0.399</b>	<b>0.257</b>	<b>0.406</b>	<b>0.474</b>	<b>0.612</b>	<b>0.327</b>	<b>0.208</b>	0.368	0.453	<b>0.572</b>
10	0.36	0.255	0.403	0.471	0.565	0.325	0.206	<b>0.369</b>	<b>0.46</b>	0.569

**Effect of different numbers of relations** To explore the effect of different numbers of relations, we randomly select 9 relations in FB15K-237 and NELL-995 to perform further evaluations. These relations are listed in descending order of quantity as shown in Table 5. Here, our model is trained

with all the triples and tested on every relation respectively. The performance of different numbers of relations on FB15K-237 and NELL-995 are shown in Tables 6 and 7. As can be seen from Table 6, for MRR, the relation ‘/people/person/profession’ with the largest number of relations did not achieve the best results, and the relation with the least number of relations ‘/business/business\_operation/industry’ did not obtain the worst result. As can be seen from Table 7, the relation ‘concept:coachesinleague’ which has the second smallest number of relations, achieves most of the best results in the NELL-995 dataset. From the above observations, we can conclude that the performance of MER has no obvious influence by the amount of training data. This suggests that MRE has the potential to achieve better experimental results even with a small number of relations, which benefits from the fact that MRE can simultaneously embed multiple relations. The multiple relation embedding method proposed in this paper can alleviate the sparsity of relations and provide extra semantic information for a smaller number of relations.

**Table 5.** The numbers of different kinds of relations on FB15K-237 and NELL-995.

Relation Name	Number		Relation Name	Number	
	Train	Test		Train	Test
/people/person/profession	7807	829	concept: mutualproxyfor	2631	119
/location/location/contains	4828	305	concept: agentcontrols	1378	30
/music/genre/artists	4443	513	concept: animalpreyson	1112	240
/film/film/genre	1598	151	concept: academicprogramatuniversity	1033	78
/people/ethnicity/people	1147	137	concept: musicartistgenre	619	52
/people/person/languages	708	86	concept: persongraduatedschool	345	11
/tv/tv_program/genre	158	13	concept: ismultipleof	101	14
/location/location/partially_contains	126	12	concept: coachesinleague	81	12
/business/business_operation/industry	60	10	concept: plantgrowinginplant	73	10

**Table 6.** The performance of different numbers of relations on FB15K-237. The best result is in **bold**.

Relation Name	MRR	Hits@k			
		1	3	5	10
/people/person/profession	0.646	0.479	<b>0.771</b>	<b>0.864</b>	<b>0.943</b>
/location/location/contains	0.13	0.066	0.144	0.174	0.249
/music/genre/artists	0.203	0.107	0.228	0.288	0.392
/film/film/genre	0.358	0.199	0.404	0.556	0.728
/people/ethnicity/people	0.096	0.036	0.095	0.117	0.182
/people/person/languages	<b>0.661</b>	<b>0.547</b>	0.709	0.849	0.919
/tv/tv_program/genre	0.534	0.385	0.615	0.692	0.846
/location/location/partially_contains	0.286	0.167	0.25	0.417	0.583
/business/business_operation/industry	0.615	0.5	0.7	0.7	0.9

**Table 7.** The performance of different numbers of relations on NELL-995. The best result is in **bold**.

Relation Name	MRR	Hits@k			
		1	3	5	10
concept:mutualproxyfor	0.1	0.042	0.084	0.151	0.210
concept:agentcontrols	0.143	0.067	0.167	0.2	0.267
concept:animalpreyson	0.247	0.138	0.258	0.35	0.463
concept:academicprogramatuniversity	0.511	0.372	0.603	0.641	0.769
concept:musicartistgenre	0.187	0.058	0.231	0.288	0.462
concept:persongraduatedschool	0.648	0.455	0.818	<b>0.909</b>	0.909
concept:ismultipleof	0.217	0.143	0.143	0.286	0.429
concept:coachesinleague	<b>0.766</b>	<b>0.667</b>	<b>0.833</b>	0.833	<b>0.917</b>
concept:plantgrowinginplant	0.222	0.2	0.2	0.2	0.2

**Effect of extra semantic information** To analyze the parameter sensitivity of extra semantic information, we set a threshold  $\Omega$  to change the overall energy. The changed energy function can be defined as  $E = E_1 + \Omega (E_2 + E_3)$ , where the larger  $\Omega$  is, the richer the extra semantic information. The results are shown in Table 8. We can see that the MRR and Hits@k of MRE will increase with extra semantic information. This once again confirms that sufficient semantic information helps to improve the performance of the knowledge graph completion.

**Table 8.** Parameter sensitivity of extra semantic information. The best result is in **bold**.

$\Omega$	FB15K-237					NELL-995				
	MRR	Hits@k				MRR	Hits@k			
		1	3	5	10		1	3	5	10
0.2	0.355	0.215	0.388	0.447	0.548	0.279	0.173	0.311	0.381	0.492
0.4	0.369	0.238	0.393	0.459	0.57	0.292	0.185	0.319	0.4	0.507
0.6	0.374	0.245	0.402	0.461	0.573	0.298	0.189	0.33	0.41	0.525
0.8	0.381	0.253	<b>0.409</b>	0.468	0.582	0.31	0.196	0.346	0.427	0.542
1	<b>0.399</b>	<b>0.257</b>	0.406	<b>0.474</b>	<b>0.612</b>	<b>0.327</b>	<b>0.208</b>	<b>0.368</b>	<b>0.453</b>	<b>0.572</b>

### 5.6. Visualization analysis

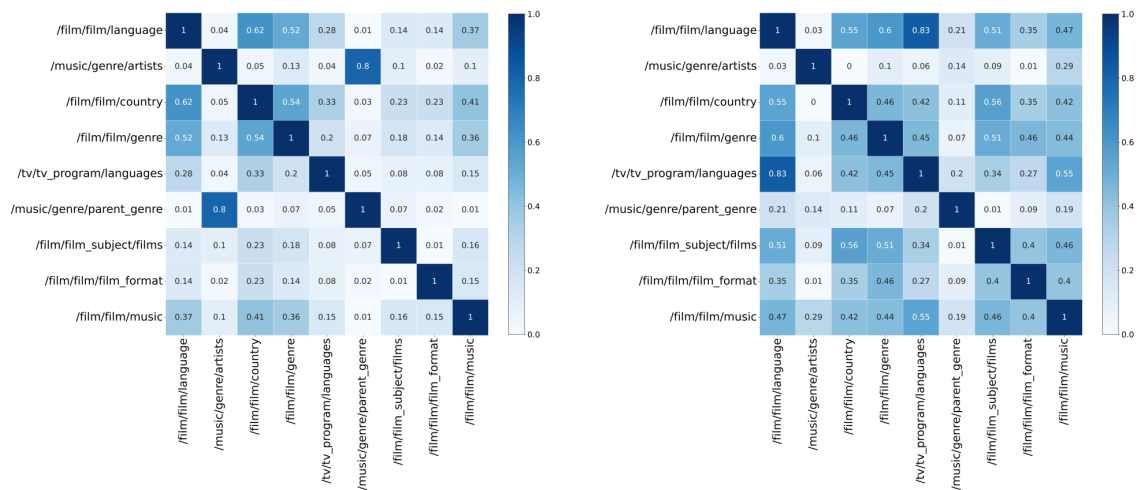
To understand and analyze semantic similarities among different relations, we visualize knowledge graph completion results of TransE [12] and MRE. We compute the similarities matrices of each two learned relation embeddings. Figures 5 and 6 show the semantic similarities of different relations on FB15K-237 and NELL-995, respectively. From the figures, we have three significant observations. First, heat maps from Figures 5 and 6 show an evident regularity for various relations in FB15K-237 and NELL-995, i.e., semantic similarities exist among various relations. For each two learned relation embeddings, the darker the heat map color is, the higher the similarity and the tighter the semantic associations. Second, not all relations have a high degree of semantic similarities. This indicates that TransE and MRE can obtain semantic differences among different relations. Third, compared

with TransE, the discrimination degree of heat maps in MRE is more obvious, i.e., the similarity of some relations in the heat map will be higher. For example, the similarity between the relation ‘/film/film/genre’ and the relation ‘/film/film/genre’ in Figure 5 has increased from 0.37 to 0.47. This shows that our proposed model can learn more accurate semantic information than TransE.

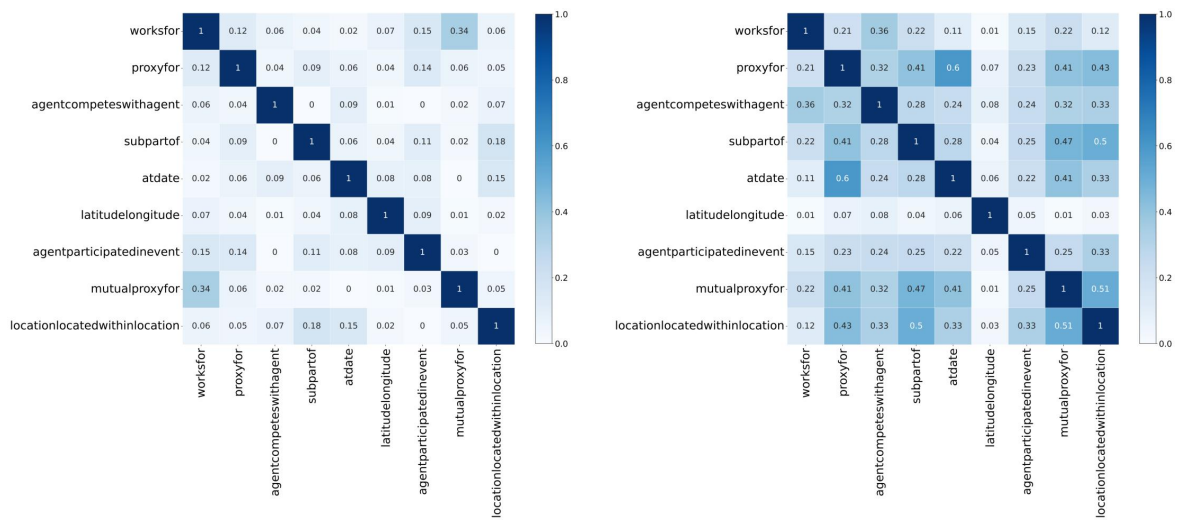
**Table 9.** Examples of tail entity prediction. Correctly predicted answers are marked in **bold**.

Dataset	Query Head Entity    Relation	Top-5 Prediction	Golden Answer	Rank
FB15K-237	/m/0418wg    /film/film/language	/m/06nm1 /m/02h40lc <b>/m/02bjrlw</b> /m/064_8sq /m/04306rv	/m/02bjrlw	3
	/m/063g71    /people/person/nationality	<b>/m/09c7w0</b> /m/0d060g /m/0rh6k /m/05fkf /m/05kkh	/m/09c7w0	1
	/m/02hnl    /music/instrument/instrumentalists	/m/01wl38s /m/05qhnq /m/050z2 <b>/m/0137g1</b> /m/01vsyg9	/m/0137g1	4
	/m/06pcz0    /people/person/profession	/m/018gz8 <b>/m/0dxtg</b> /m/03gjzk /m/02krf9 /m/0cbd2_r	/m/0dxtg	2
	/m/0190yn    /music/genre/artists	/m/01wgf6 <b>/m/01x1cn2</b> /m/04bbv7 /m/011qf49 /m/01jfr3y	/m/01x1cn2	2
	/m/07rd7    /film/director/film	/m/0g56t9t /m/050xxm /m/04pk1f <b>/m/09g7vfw</b> /m/01jfr3y	/m/09g7vfw	4

To showcase the process of completing knowledge graphs, we regard tail entity prediction as a simple question-answer system. Different tail entity prediction results are shown in Table 9. Given a query consisting of a head entity and a relation, the objective is to predict the golden answer (tail entity). Top-5 predictions are listed in Table 9, and we can compute the rank of the golden answer in candidate answers. It can be seen from Table 9 that our proposed MRE can generate a list of candidate answers and predict the golden answer.



**Figure 5.** Semantic similarities of different relations on FB15K-237 in TransE (left) and MRE (right).



**Figure 6.** Semantic similarities of different relations on NELL-995 in TransE (left) and MRE (right).



## 6. Conclusions

In this paper, we proposed a new kind of knowledge graph completion approach called MRE to predict if a triple in a knowledge graph is valid or not. Unlike most methods that only consider the single form of relations in the embedding phase, we embedded multiple relations in a semantic space. Specifically, we introduced two relation encoders to capture semantic information of multi-hop and rule-based relations. These two encoders proposed in this paper can realize the interactions between connected entities and relations in the encoding phase. In addition, we defined corresponding energy functions for multi-hop and rule-based relations to obtain new representations. Compared with current KGC methods, MRE can better use multiple relations properties and provide additional semantic information for single-form relations.

In order to verify the effectiveness and superiority of our work, we conducted massive experiments on two widely used benchmarks. The experimental results showed that our method could effectively capture multiple semantics. Further evaluations demonstrated that our work could keep stable when the margin is within a reasonable range and alleviate the sparse existing in knowledge graphs. Visualization analysis showed the semantic similarities of different relations and the working principle of MRE.

As for future work, we plan to study follow-up open problems. (i) MRE performs the knowledge graph completion based on MLP. How can we advance it using sophisticated neural architectures such as capsule or graph neural networks? (ii) Existing studies have shown that there is still much additional information in the knowledge graphs that is not used, such as entity types and spatial information. How can we integrate useful information with MRE to advance knowledge graph completion?

## Acknowledgments

We thank the editors and anonymous reviewers for their helpful comments for improving our work.

## Conflict of interest

All authors declare no conflicts of interest in this paper.

## References

1. L. F. Wang, X. Lu, Z. Jiang, Z. Zhang, R. Li, M. Zhao, et al., Frs: A simple knowledge graph embedding model for entity prediction, *Math. Biosci. Eng.*, **16** (2019), 7789–7807. <https://doi.org/10.3934/mbe.2019391>
2. K. Zhang, B. Hu, F. Zhou, Y. Song, X. Zhao, X. Huang, Graph-based structural knowledge-aware network for diagnosis assistant, *Math. Biosci. Eng.*, **19** (2022), 10533–10549. <https://doi.org/10.3934/mbe.2022492>
3. S. Dost, L. Serafini, M. Rospocher, L. Ballan, A. Sperduti, Aligning and linking entity mentions in image, text, and knowledge base, *Data Knowl. Eng.*, **138** (2022), 101975. <https://doi.org/10.1016/j.datak.2021.101975>

4. Z. Gomolka, B. Twarog, E. Zeslawska, E. Dudek-Dyduch, Knowledge base component of intelligent ALMM system based on the ontology approach, *Expert Syst. Appl.*, **199** (2022), 116975. <https://doi.org/10.1016/j.eswa.2022.116975>
5. P. Do, T. H. V. Phan, Developing a BERT based triple classification model using knowledge graph embedding for question answering system, *Appl. Intell.*, **52** (2022), 636–651. <https://doi.org/10.1007/s10489-021-02460-w>
6. K. D. Bollacker, C. Evans, P. K. Paritosh, T. Sturge, J. Taylor, Freebase: a collaboratively created graph database for structuring human knowledge, in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, ACM, Vancouver, Canada, (2008), 1247–1250. <https://doi.org/10.1145/1376616.1376746>
7. T. Mitchell, W. Cohen, E. Hruschka, P. Talukdar, B. Yang, J. Betteridge, et al., Never-ending learning, *Commun. ACM*, **61** (2018), 103–115. <https://doi.org/10.1145/3191513>
8. L. Hou, M. Wu, H. Y. Kang, S. Zheng, L. Shen, Q. Qian, et al., Pmo: A knowledge representation model towards precision medicine, *Math. Biosci. Eng.*, **17** (2020), 4098–4114. <https://doi.org/10.3934/mbe.2020227>
9. X. Lu, L. Wang, Z. Jiang, S. He, S. Liu, MMKRL: A robust embedding approach for multi-modal knowledge graph representation learning, *Appl. Intell.*, **52** (2022), 7480–7497. <https://doi.org/10.1007/s10489-021-02693-9>
10. N. D. Rodríguez, A. Lamas, J. Sanchez, G. Franchi, I. Donadello, S. Tabik, et al., Explainable neural-symbolic learning (*X-NeSyL*) methodology to fuse deep learning representations with expert knowledge graphs: The monumai cultural heritage use case, *Inf. Fusion*, **79** (2022), 58–83. <https://doi.org/10.1016/j.inffus.2021.09.022>
11. S. Chakrabarti, Deep knowledge graph representation learning for completion, alignment, and question answering, in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, Madrid, Spain, (2022), 3451–3454. <https://doi.org/10.1145/3477495.3532679>
12. A. Bordes, N. Usunier, A. García-Durán, J. Weston, O. Yakhnenko, Translating embeddings for modeling multi-relational data, in *Advances in Neural Information Processing Systems 26*, Curran Associates Inc., Lake Tahoe, United States, (2013), 2787–2795.
13. Z. Wang, J. Zhang, J. Feng, Z. Chen, Knowledge graph embedding by translating on hyperplanes, in *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, AAAI, Québec City, Canada, (2014), 1112–1119. <https://doi.org/10.1609/aaai.v28i1.8870>
14. Y. Lin, Z. Liu, M. Sun, Y. Liu, X. Zhu, Learning entity and relation embeddings for knowledge graph completion, in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI, Austin, USA, (2015), 2181–2187. <https://doi.org/10.1609/aaai.v29i1.9491>
15. M. Nickel, V. Tresp, H. Kriegel, A three-way model for collective learning on multi-relational data, in *Proceedings of the 28th International Conference on Machine Learning*, Omnipress, Bellevue, USA, (2011), 809–816.
16. B. Yang, W. Yih, X. He, J. Gao, L. Deng, Embedding entities and relations for learning and inference in knowledge bases, in *3rd International Conference on Learning Representations*, San Diego, USA, 2015.

17. T. Dettmers, P. Minervini, P. Stenetorp, S. Riedel, Convolutional 2d knowledge graph embeddings, in *Proceedings of the AAAI conference on artificial intelligence*, AAAI, New Orleans, Louisiana, USA, (2018), 1811–1818. <https://doi.org/10.1609/aaai.v32i1.11573>
18. J. Huang, T. Zhang, J. Zhu, W. Yu, Y. Tang, Y. He, A deep embedding model for knowledge graph completion based on attention mechanism, *Neural Comput. Appl.*, **33** (2021), 9751–9760. <https://doi.org/10.1007/s00521-021-05742-z>
19. Y. Lin, Z. Liu, H. Luan, M. Sun, S. Rao, S. Liu, Modeling relation paths for representation learning of knowledge bases, in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, The Association for Computational Linguistics, Lisbon, Portugal, (2015), 705–714. <https://doi.org/10.18653/v1/d15-1082>
20. S. Guo, Q. Wang, L. Wang, B. Wang, L. Guo, Knowledge graph embedding with iterative guidance from soft rules, in *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI, New Orleans, USA, (2018), 4816–4823. <https://doi.org/10.1609/aaai.v32i1.11918>
21. M. Pitsikalis, T. Do, A. Lisitsa, S. Luo, Logic rules meet deep learning: A novel approach for ship type classification (extended abstract), in *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, IJCAI, Vienna, Austria, (2022), 5324–5328. <https://doi.org/10.24963/ijcai.2022/744>
22. S. Matsuoka, T. Sawaragi, Recovery planning of industrial robots based on semantic information of failures and time-dependent utility, *Adv. Eng. Inf.*, **51** (2022), 101507. <https://doi.org/10.1016/j.aei.2021.101507>
23. M. Nickel, L. Rosasco, T. A. Poggio, Holographic embeddings of knowledge graphs, in *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI, Phoenix, USA, (2016), 1955–1961. <https://doi.org/10.1609/aaai.v30i1.10314>
24. R. Biswas, M. Alam, H. Sack, MADLINK: Attentive multihop and entity descriptions for link prediction in knowledge graphs, *Semant. Web*, (2021), 1–24. <https://doi.org/10.3233/SW-222960>
25. L. Galárraga, C. Teflioudi, K. Hose, F. M. Suchanek, Fast rule mining in ontological knowledge bases with AMIE+, *VLDB J.*, **24** (2015), 707–730. <https://doi.org/10.1007/s00778-015-0394-1>
26. J. Kalina, J. Tumpach, M. Holena, On combining robustness and regularization in training multilayer perceptrons over small data, in *2022 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Padua, Italy, (2022), 1–8. <https://doi.org/10.1109/IJCNN55064.2022.9892510>
27. Y. Bai, Z. Ying, H. Ren, J. Leskovec, Modeling heterogeneous hierarchies with relation-specific hyperbolic cones, in *Advances in Neural Information Processing Systems 34*, (2021), 12316–12327.
28. S. Chen, X. Liu, J. Gao, J. Jiao, R. Zhang, Y. Ji, Hitter: Hierarchical transformers for knowledge graph embeddings, in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, (2021), 10395–10407. <https://doi.org/10.18653/v1/2021.emnlp-main.812>
29. D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, in *3rd International Conference on Learning Representations*, Amsterdam Machine Learning lab, 2015.

- 
30. G. Dai, X. Wang, X. Zou, C. Liu, S. Cen, MRGAT: multi-relational graph attention network for knowledge graph completion, *Neural Networks*, **154** (2022), 234–245. <https://doi.org/10.1016/j.neunet.2022.07.014>



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)