

# AutoETER: Automated Entity Type Representation for Knowledge Graph Embedding

Guanglin Niu<sup>1</sup>, Bo Li<sup>1,2</sup>, Yongfei Zhang<sup>1,2,3\*</sup>, Shiliang Pu<sup>4</sup>, Jingyang Li<sup>1</sup>

<sup>1</sup>Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing 100191, China

<sup>2</sup>State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China <sup>3</sup>Pengcheng Laboratory, Shenzhen 518055, China

<sup>4</sup>Hikvision Research Institute, Hangzhou 311500, China

{beihangnli, boli, yfzhang, lijingyang}@buaa.edu.cn, pushiliang.hri@hikvision.com

## Abstract

Recent advances in Knowledge Graph Embedding (KGE) allow for representing entities and relations in continuous vector spaces. Some traditional KGE models leveraging additional type information can improve the representation of entities which however totally rely on the explicit types or neglect the diverse type representations specific to various relations. Besides, none of the existing methods is capable of inferring all the relation patterns of symmetry, inversion and composition as well as the complex properties of 1-N, N-1 and N-N relations, simultaneously. To explore the type information for any KG, we develop a novel KGE framework with **Automated Entity Type Representation** (AutoETER), which learns the latent type embedding of each entity by regarding each relation as a translation operation between the types of two entities with a relation-aware projection mechanism. Particularly, our designed automated type representation learning mechanism is a pluggable module which can be easily incorporated with any KGE model. Besides, our approach could model and infer all the relation patterns and complex relations. Experiments on four datasets demonstrate the superior performance of our model compared to state-of-the-art baselines on link prediction tasks, and the visualization of type clustering provides clearly the explanation of type embeddings and verifies the effectiveness of our model.

## 1 Introduction

In recent years, knowledge graph (KG) has been viewed as a powerful technique for recognition systems and prevalent in many fields such as E-commerce, intelligent healthcare, and public security. Knowledge graphs collect and store a great deal of commonsense or domain knowledge in factual triples composed of entity pairs with their relations. The existing large scale KGs such as

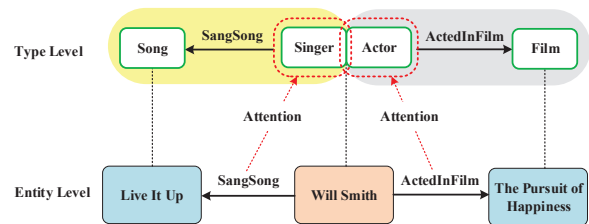


Figure 1: An actual example of the entity-specific triples and the type-specific triples with relation-aware projection mechanism. *Will Smith* has multiple types such as *Singer* and *Actor*, but only the type *Singer* should be focused on for the relation *SangSong*.

Freebase (Bollacker et al., 2008), WordNet (Miller, 1995), YAGO (Suchanek et al., 2007) have shown their validity in various applications, including question answering (Diefenbach et al., 2018), dialogue generation (He et al., 2017) and recommender systems (Wang et al., 2019).

However, the existing KGs are inevitably incomplete whether they are constructed manually or automatically, limiting the effectiveness when exploited for downstream applications. Some existing KG inference approaches such as inductive logic programming algorithm (Ray, 2009), Markov logic networks based method (Qu and Tang, 2019) and reinforcement learning-based approach (Lin et al., 2018) try to predict entities or relations in KGs but face the limited performance and suffer from the low efficiency. Compared to the above approaches, knowledge graph embedding models could learn the latent representations of the entities and relations and show the best performance on the KG completion task. However, most of the KG embedding models such as TransE (Bordes et al., 2013) and its variants TransH (Wang et al., 2014), TransR (Lin et al., 2015b) learn KG embeddings relying on single triples, which simply exploit the structure information implied in KGs.

Entity types define categories of entities that are valid to enhance the representation of entities. In

\*Corresponding Author

many type-embodied models such as TKRL (Xie et al., 2016) and TransT (Ma et al., 2017), the explicit types are necessary while some KGs (i.e., WordNet) lack them, which limits the versatility of these approaches. JOIE (Hao et al., 2019) jointly encodes both the ontology and instance views of KGs. Nevertheless, ontologies’ concepts always represent the general categories of entities but cannot reflect the specific types, primarily associated with different relations. Jain (Jain et al., 2018) learned the type embeddings by defining the compatibility between an entity type and a relation. Still, it ignores the semantics implied in a whole triple consisting of a relation jointly with its linked two entity types. Moreover, all the previous type-based approaches neglect the diversity of entity type representations specific to various relations. As Figure 1 shows, contrary to the previous researches considering entity types, the triples in the entity level could be extended to triples in the type level. Each entity has multiple types, and diverse types should be focused on for different specific relations.

Additionally, some models embed the entities and relations into the complex vector space instead of the frequently-used real space to improve the capability of representation learning, including ComplEx (Trouillon et al., 2016) and RotatE (Sun et al., 2019). Nevertheless, none of the existing embedding models could model and infer all the relation patterns and the complex 1-N, N-1 and N-N relations, simultaneously.

To conduct the KG inference from the perspectives of both entity-specific triples and type-specific triples on any KG, whether the explicit types exist, we propose AutoETER to automatically learn the diverse type representations of each entity when focusing on the various associated relations. Intuitively, the high-dimensional entity embeddings imply the individual features to distinguish the diverse entities. In contrast, the low-dimensional type embeddings capture the general features to discover the similarity of entities according to their categories. Inspired by the translational-based principle in TransE, we expect that given a head entity and its associated relation, the tail entity’s type representation can be obtained by  $\mathbf{type}_{head} + \mathbf{relation} = \mathbf{type}_{tail}$ . Particularly, the latent type embeddings of two head or two tail entities focused on the same relation should be close to each other since they imply the same type. Furthermore, the embeddings

of the entity-specific triples and the type-specific triples are capable of modeling and inferring symmetry, inversion, composition, and complex 1-N, N-1, N-N relations.

The contributions of this work are summarized as follows:

- We model type representations to enrich the general features of entities. A novel model AutoETER is proposed to learn the embeddings of entities, relations and entity types from entity-specific triples and type-specific triples without explicit types in KGs. Furthermore, the type embeddings can be incorporated with the entity embeddings for inference.
- To the best of our knowledge, we are the first to model and infer all the relation patterns, including symmetry, inversion and composition, as well as complex relations of 1-N, N-1 and N-N for the KG inference.
- We conduct extensive experiments on link prediction on four real-world benchmark datasets. The evaluation results demonstrate the superiority of our proposed model over other state-of-the-art algorithms. The visualization of clustering type embeddings validates the effectiveness of automatically representing entity types with relation-aware projection.

## 2 Related Works

### 2.1 Knowledge Graph Inference

To address the inherent incompleteness of KGs, multiple KG inference methods are investigated and have made significant progress. Traditional researches devote to generate logic rules based on inductive logic programming such as HAIL (Ray, 2009) to predict the missing entities in KGs. However, employing logic rules in KG inference limits generalization performance. Path ranking algorithm (PRA) (Lao et al., 2011) extracts the relational path features based on random-walk to infer the relationships between entity pairs. DeepPath (Lin et al., 2018) is a foundational approach that formulates the multi-hop reasoning as a Markov decision process and leverages reinforcement learning (RL) to find paths in KGs. However, the RL-based multi-hop KG reasoning approaches consume much time in searching paths.

## 2.2 KG Embedding Models

Various KG embedding models have been extensively developed for KG inference in recent years (Wang et al., 2017). KGE models are capable of capturing latent representations of entities and relations in KGs independently from hand-crafted rules, and they have shown a strong capacity of efficient computation in many knowledge-aware applications (Ji et al., 2020). TransE (Bordes et al., 2013) is the foundational translation-based method, which regards a relation as a translation operation from the head entity to the tail entity. Along with TransE, multiple variants are proposed to improve the embedding performance of KGs (Niu et al., 2020; Yuan et al., 2019; Xiao et al., 2016). ConvE (Dettmers et al., 2018) is a typical method representing entities and relations based on convolutional neural networks (CNN). Another category of KG embedding contains many tensor decomposition models, including DisMult (Yang et al., 2015). Particularly, ComplEx (Trouillon et al., 2016) extends DisMult to learn the KG embeddings in the complex space. RotatE (Sun et al., 2019) defines a relation as a rotation from source to target entities in a complex space but cannot infer the complex relations 1-N, N-1 and N-N. What’s more, all the approaches above purely depend on the triples directly observed in KGs.

## 2.3 Models Incorporating Entity Types

To further improve the performance of KG embedding, various auxiliary information is introduced, such as paths (Lin et al., 2015a; Niu et al., 2020), graph structure (Michael et al., 2018) and entity types (Xie et al., 2016; Krompaß et al., 2015; Ma et al., 2017). Among such information, entity types contain less noise and are appropriate for providing more general semantics for each entity. TKRL (Xie et al., 2016) projects each entity with the type-specific projection matrices. TransT (Ma et al., 2017) measures the semantic similarity of entities and relations utilizing types. However, all the above type-based KG embedding models require the supervision of explicit types and cannot work on KGs without explicit types. JOIE (Hao et al., 2019) links entities to their concepts in the ontology for jointly embed the instance-view graph and the ontology-view graph, but the concepts in ontologies provide too broad or even noisy information to represent the specific and precise types of each entity. (Jain et al., 2018) introduces the compatibil-

ity between the embeddings of an entity type and a relation for link prediction. Still, all the existing type-enhanced models neglect that an entity’s diverse types should be focused on when this entity is associated with various relations. Meanwhile, the association property implied in the embeddings of the type-specific triples has not been well modeled.

## 3 AutoETER: KGE with Automated Entity Type Representation

To cope with the above limitations, we describe the proposed model AutoETER, which aims to automatically learn a variant of type representations semantically compatible with various relations and infer all the relation patterns and complex relations. As figure 2 shows, we first embed the entities and relations into complex space via the entity-specific triple encoder with a hyper-plane projection strategy (§3.1). Additionally, the type-specific triple encoder is developed to learn type embeddings incorporated with a relation-aware projection mechanism (§3.2). Meanwhile, the type embeddings are constrained by their similarity derived from the associated relations (§3.3). Afterward, we propose the overall optimization objective with both entity-specific triple and type-specific triple representations and the similarity constraint of the type embeddings (§3.4).

### 3.1 Entity-specific Triple Encoder

We embed the entities and relations into the complex space and regard a relation as the rotation operation from the head entity to the tail entity as in RotatE (Sun et al., 2019). To further model and infer the complex relations such as 1-N, N-1 and N-N, we project entities into their associated relation hyper-planes to ensure each entity has various representations concerning the specific relations. In terms of an entity-specific triple  $(h, r, t)$ , the energy function  $E_1(h, r, t)$  is defined as

$$\mathbf{e}_{h,r} = \mathbf{h} - \mathbf{h}^\top \mathbf{w}_r \mathbf{h}, \quad \mathbf{e}_{t,r} = \mathbf{t} - \mathbf{t}^\top \mathbf{w}_r \mathbf{t} \quad (1)$$

$$E_1(h, r, t) = \|\mathbf{e}_{h,r} \circ \mathbf{r} - \mathbf{e}_{t,r}\| \quad (2)$$

where  $\mathbf{h} \in \mathbb{C}^k$ ,  $\mathbf{t} \in \mathbb{C}^k$ ,  $\mathbf{r} \in \mathbb{C}^k$  are the embeddings of head entity  $h$ , tail entity  $t$  and relation  $r$  in the complex space with dimension  $k$ .  $\mathbf{w}_r \in \mathbb{R}^k$  denotes the normal vector of the hyper-plane involved in the relation  $r$ .  $\mathbf{e}_{h,r} \in \mathbb{C}^k$  and  $\mathbf{e}_{t,r} \in \mathbb{C}^k$  represent the entity embeddings of  $h$  and  $t$  projected in the hyper-plane  $\mathbf{w}_r$ .  $\circ$  is the Hadamard product.

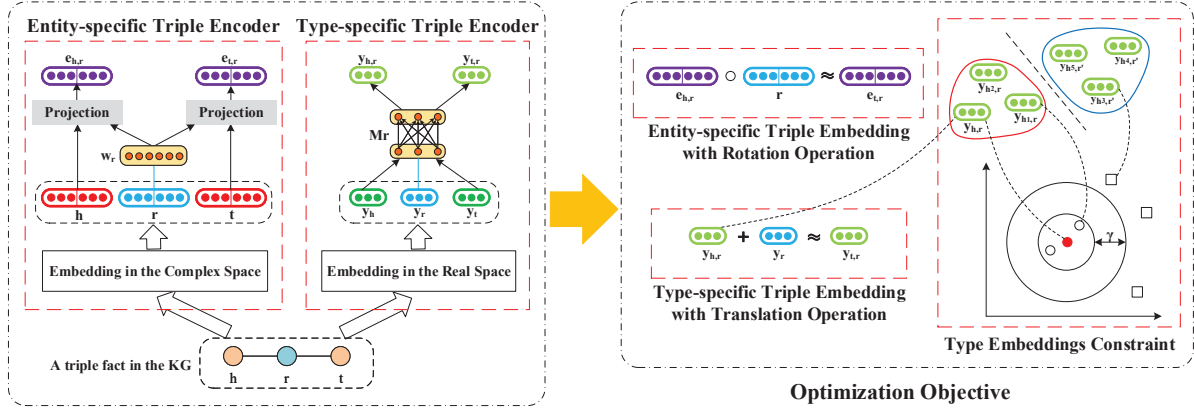


Figure 2: The architecture of AutoETER. Given a triple fact  $(h, r, t)$ ,  $\mathbf{e}_{h,r}$  and  $\mathbf{e}_{t,r}$  are the projected entity embeddings in the hyper-plane of relation  $r$ ,  $\mathbf{y}_{h,r}$  and  $\mathbf{y}_{t,r}$  are type embeddings focusing on relation  $r$ . Furthermore, we expect the embeddings of entity-specific triple satisfies rotation operation and type-specific triple satisfies translation operation from head to tail entities. Type embeddings associated with the same relation  $r$  are constrained to be closer, where  $\gamma$  is the margin enforced between two clusters of type embeddings related to different relations.

On account of the embeddings of entity-specific triples, our model can infer all the relation patterns via the rotation operation from head to tail entities as in RotatE. Particularly,  $\mathbf{r}$  is constrained to be  $|r_i| = 1, i = 1, 2, \dots, k$  for inferring the symmetric relation pattern and at least one element of  $\mathbf{r}$  is -1 to ensure the diverse representations of head and tail entities. Moreover, the projection operation shown in Eq. 1 enables our model to infer the complex relations via various representations of entities regarding different relations.

### 3.2 Type-specific Triple Encoder

Given an entity  $e$  and its associated relation  $r$  in a triple, we aim to learn the type and relation embeddings with a relation-aware projection mechanism to output the most important information of the type representations:

$$f_{att}(e, r) = \mathbf{M}_r \mathbf{y}_e \quad (3)$$

where  $\mathbf{y}_e \in \mathbb{R}^d$  denotes the type embedding of entity  $e$  in the real space with dimension  $d$ .  $\mathbf{M}_r \in \mathbb{R}^{d \times d}$  is defined as the projection weight matrix associated with the relation  $r$ , which could automatically select the latent information of each type embedding most relevant to the relation  $r$ .

With the relation-aware projection defined in Eq. 3, the energy function involved in type-specific triples is defined as

$$\begin{aligned} \mathbf{y}_{h,r} &= f_{att}(h, r), \quad \mathbf{y}_{t,r} = f_{att}(t, r) \\ E_2(h, r, t) &= \|\mathbf{y}_{h,r} + \mathbf{y}_r - \mathbf{y}_{t,r}\| \end{aligned} \quad (4)$$

where  $\mathbf{y}_{h,r} \in \mathbb{R}^d$ ,  $\mathbf{y}_{t,r} \in \mathbb{R}^d$  are the type embeddings of entities  $h$  and  $t$  both focusing on the relation  $r$  and  $\mathbf{y}_r \in \mathbb{R}^d$  denotes the embedding of the relation  $r$  in the type-specific triple. In terms of the energy function in Eq. 4, we expect that

$$\mathbf{y}_{h,r} + \mathbf{y}_r = \mathbf{y}_{t,r} \quad (5)$$

Furthermore, with the type and relation embeddings learned in the real spaces, our model cost fewer parameters and could model and infer all the relation patterns including symmetry (Lemma 1), inversion (Lemma 2) and composition (Lemma 3) as well as the complex properties of relations:

**Lemma 1.** *Our model could infer relation pattern of symmetry by type-specific triple embeddings.*

*Proof.* If a relation  $r$  is symmetric, two triples  $(h, r, t)$  and  $(t, r, h)$  will hold. From Eq. 5, the correlations among the embeddings of types and relations can be obtained as:

$$\mathbf{y}_{h,r} + \mathbf{y}_r = \mathbf{y}_{t,r}, \quad \mathbf{y}_{t,r} + \mathbf{y}_r = \mathbf{y}_{h,r} \quad (6)$$

From Eq. 6, we can further derive that

$$\mathbf{y}_{h,r} = \mathbf{y}_{t,r}, \quad \mathbf{y}_r = \mathbf{0} \quad (7)$$

We prove that the embedding of a symmetric relation should be zero vector, and the type embeddings of head and tail entities should be equal. The above results are reasonable owing to the focused types of two entities linked by the symmetric relation are supposed to be same.  $\square$



**Lemma 2.** *Our model is able to infer relation pattern of inversion by type-specific triple embeddings.*

*Proof.* With the inverse relations  $r_1$  and  $r_2$ , two triples  $(h, r_1, t)$  and  $(t, r_2, h)$  hold. From Eqs. 3, 4 and 5, it can be retrieved that

$$\mathbf{M}_{r_1}\mathbf{y}_h + \mathbf{y}_{r_1} = \mathbf{M}_{r_1}\mathbf{y}_t \quad (8)$$

$$\mathbf{M}_{r_2}\mathbf{y}_t + \mathbf{y}_{r_2} = \mathbf{M}_{r_2}\mathbf{y}_h \quad (9)$$

We can define a transform matrix  $\mathbf{P} \in \mathbb{R}^{d \times d}$  that satisfies

$$\mathbf{M}_{r_1} = \mathbf{P}\mathbf{M}_{r_2} \quad (10)$$

Substituting Eq. 10 into Eq. 9, the latter can be modified as

$$\mathbf{M}_{r_1}\mathbf{y}_t + \mathbf{P}\mathbf{y}_{r_2} = \mathbf{M}_{r_1}\mathbf{y}_h \quad (11)$$

Then, substituting Eq. 11 into Eq. 8, it yields that

$$\mathbf{y}_{r_1} = -\mathbf{P}\mathbf{y}_{r_2} \quad (12)$$

We can model and infer the inverse relations with the relation embeddings satisfying the relationship as in Eq. 12.  $\square$

**Lemma 3.** *Our model is capable of inferring the relations of composition by type-specific triple embeddings.*

*Proof.* On account of the relations of composition pattern  $r_3(a, c) \Leftarrow r_1(a, b) \wedge r_2(b, c)$ , the corresponding triples  $(a, r_1, b)$ ,  $(b, r_2, c)$  and  $(a, r_3, c)$  hold. Meanwhile, considering Eqs. 3, 4 and 5, it can be obtained that

$$\mathbf{M}_{r_1}\mathbf{y}_a + \mathbf{y}_{r_1} = \mathbf{M}_{r_1}\mathbf{y}_b \quad (13)$$

$$\mathbf{M}_{r_2}\mathbf{y}_b + \mathbf{y}_{r_2} = \mathbf{M}_{r_2}\mathbf{y}_c \quad (14)$$

$$\mathbf{M}_{r_3}\mathbf{y}_a + \mathbf{y}_{r_3} = \mathbf{M}_{r_3}\mathbf{y}_c \quad (15)$$

We can define two transform matrices  $\mathbf{P} \in \mathbb{R}^{d \times d}$  and  $\mathbf{Q} \in \mathbb{R}^{d \times d}$  to satisfy

$$\mathbf{P}\mathbf{M}_{r_1} = \mathbf{M}_{r_3} \quad (16)$$

$$\mathbf{Q}\mathbf{M}_{r_2} = \mathbf{M}_{r_3} \quad (17)$$

Substituting Eq. 16 into Eq. 13 and Eq. 17 into Eq. 14, respectively, we can derive that

$$\mathbf{M}_{r_3}\mathbf{y}_a + \mathbf{P}\mathbf{y}_{r_1} = \mathbf{M}_{r_3}\mathbf{y}_b \quad (18)$$

$$\mathbf{M}_{r_3}\mathbf{y}_b + \mathbf{Q}\mathbf{y}_{r_2} = \mathbf{M}_{r_3}\mathbf{y}_c \quad (19)$$

Substituting Eq. 18 into Eq. 19, it can be retrieved that

$$\mathbf{M}_{r_3}\mathbf{y}_a + \mathbf{P}\mathbf{y}_{r_1} + \mathbf{Q}\mathbf{y}_{r_2} = \mathbf{M}_{r_3}\mathbf{y}_c \quad (20)$$

Combining Eqs. 15 and 20, we can model the correlation among the relation embeddings of composition pattern as

$$\mathbf{y}_{r_3} = \mathbf{P}\mathbf{y}_{r_1} + \mathbf{Q}\mathbf{y}_{r_2} \quad (21)$$

We prove that we can model and infer the relations of composition pattern for type-specific triples with the relation embeddings as shown in Eq. 21.  $\square$

Specific to the inference on type-specific triples with the relations of complex properties 1-N, N-1 and N-N, we could exploit the various representations of an entity type associated with different relations via the relation-aware projection mechanism defined in Eq. 3 to infer on these relations.

### 3.3 Type Embeddings Similarity Constraint

In addition to learning type embeddings by the type-specific triple encoder (§3.2), the type representations should be constrained by the similarity between the entity types. The type embeddings of head entities involved in the triples with the same relation are closer to each other (the same as type embeddings of tail entities). Thus, as for two triples with the same relation, we expect that

$$\mathbf{y}_{h_1, r} = \mathbf{y}_{h_2, r}, \quad \mathbf{y}_{t_1, r} = \mathbf{y}_{t_2, r} \quad (22)$$

where  $\mathbf{y}_{h_1, r}$  and  $\mathbf{y}_{h_2, r}$  are type embeddings of head entities while  $\mathbf{y}_{t_1, r}$  and  $\mathbf{y}_{t_2, r}$  are type embeddings of tail entities. Particularly, they all focus on the same relation  $r$  by the relation-aware projection mechanism of Eq. 3.

Now, considering any two triples  $(h_1, r_1, t_1)$  and  $(h_2, r_2, t_2)$ , we design the energy function for evaluating the dissimilarity of the type embeddings as

$$E_3((h_1, r_1, t_1), (h_2, r_2, t_2)) = \frac{1}{2} (\|\mathbf{y}_{h_1, r_1} - \mathbf{y}_{h_2, r_2}\| + \|\mathbf{y}_{t_1, r_1} - \mathbf{y}_{t_2, r_2}\|) \quad (23)$$

where  $\mathbf{y}_{h_1, r_1}$  and  $\mathbf{y}_{h_2, r_2}$  are two head entity type embeddings,  $\mathbf{y}_{t_1, r_1}$  and  $\mathbf{y}_{t_2, r_2}$  are two tail entity type embeddings, and they are all associated with the relation  $r_1$  or  $r_2$ . Therefore, we expect the value derived from Eq. 23 tends to be smaller if  $r_1$  and  $r_2$  are the same relation.

### 3.4 Optimization Objective

The designed entity-specific triples encoder, type-specific triples encoder and type representations similarity constraint could be trained as a unified

end-to-end model. We optimize our model according to a three-component objective function:

$$L = \sum_{(h,r,t) \in S} \left\{ \sum_{(h',r,t') \in S'} \left\{ L_1 + \alpha_1 L_2 \right\} + \alpha_2 L_3 \right\} \quad (24)$$

in which the overall training objective consists of three components:  $L_1$  and  $L_2$  are two pair-wise loss functions that correspond to the entity-specific triple encoder and the type-specific triple encoder, respectively, and  $L_3$  is a triple loss function for constraining the type embeddings.  $\alpha_1$  and  $\alpha_2$  denote the weights of  $L_2$  and  $L_3$  for the tradeoff between the entity-specific triple, the type-specific triple and the type similarity constraint.  $S$  contains all the triples in the train set, and  $S'$  is the negative sample set generated by replacing the entities in  $S$ . Specifically,  $L_1$ ,  $L_2$  and  $L_3$  are defined as

$$L_1 = -\log \sigma(\gamma_1 - E_1(h, r, t)) - \log \sigma(E_1(h', r, t') - \gamma_1) \quad (25)$$

$$L_2 = \max [0, E_2(h, r, t) + \gamma_2 - E_2(h', r, t')] \quad (26)$$

$$L_3 = \sum_{(hp,r,tp) \in Y} \sum_{(hn,r',tn) \in Y'} \max [0, E_3((h, r, t), (hp, r, tp)) + \gamma_3 - E_3((h, r, t), (hn, r', tn))] \quad (27)$$

where  $\gamma_1$ ,  $\gamma_2$  and  $\gamma_3$  denote the fixed margins in  $L_1$ ,  $L_2$  and  $L_3$ , respectively. In specific,  $L_3$  can be viewed as the regularization in optimization for restraining the entity type embeddings.  $\sigma$  denotes the sigmoid function.  $\max[0,x]$  is the function to select the larger value between 0 and  $x$ . Particularly, in Eq. 27, the triple  $(h, r, t)$  is regarded as the anchor instance and  $(h_p, r, t_p)$  is a positive instance in the set  $Y$  containing other triples correlated to the same relation  $r$ , while  $(h_n, r', t_n)$  is any negative instance in the set  $Y'$  containing the other triples without the relation  $r$ . Besides, we employ self-adversarial sampling as in (Sun et al., 2019).

## 4 Experiment Results

In this section, we evaluate our model AutoETER for KG completion on four real-world benchmark datasets. Additionally, we visualize the clustering results of type embeddings for demonstrating the effectiveness of representing types automatically.

Dataset	WN18	YAGO3-10	FB15K	FB15K-237
#Entity	40,943	123,182	14,951	14,505
#Relation	18	37	1,345	237
#Train	141,442	1,079,040	483,142	272,115
#Valid	5,000	5,000	50,000	17,535
#Test	5,000	5,000	59,071	20,466

Table 1: Statistics of datasets used in the experiments.

## 4.1 Experimental Setup

### 4.1.1 Datasets

We utilize four standard datasets<sup>1</sup> for link prediction tasks: FB15K (Bordes et al., 2013) is a widely used dataset that is a subgraph of the commonsense knowledge graph Freebase. WN18 (Bordes et al., 2013) is a subset of the lexical knowledge graph WordNet. YAGO3-10 (Dettmers et al., 2018) is a subset of YAGO. Each of the three datasets consists of all the relation patterns, including symmetry, inversion, composition and complex 1-N, N-1 and N-N of relations. FB15K-237 (Toutanova and Chen, 2015) is a subset of FB15K and removes all the inverse relations. Table 1 exhibits the statistics of all the datasets exploited.

### 4.1.2 Evaluation Protocol

The link prediction task aims to predict when the head or tail entity of a triple in the test set is missing. For link prediction, all the entities in the KG are respectively replaced with the missing entity to generate the candidate triples. Then, on account of each candidate triple  $(h, r, t)$ , we combine the two perspectives of the entity-specific triple jointly with the type-specific triple to evaluate the plausibility of this candidate triple, and the energy function for evaluation is designed as follows:

$$E_{pred}(h, r, t) = E_1(h, r, t) + \alpha_1 E_2(h, r, t) \quad (28)$$

The above energy function  $E_{pred}(h, r, t)$  is composed of the energy functions  $E_1(h, r, t)$  (with regard to the entity-specific triple) and  $E_2(h, r, t)$  (with respect to the type-specific triple) defined in Eqs. 2 and 4, respectively.  $\alpha_1$  is the weight which is the same as in Eq. 24 for a trade-off. Then, the scores with respect to all the candidate triples are calculated by Eq. 28. Subsequently, these scores are sorted in ascending order, and further, the correct triple rank can be obtained.

Three standard metrics are employed to evaluate the performance of link prediction:

<sup>1</sup>Datasets could be found at onedrive: [https://1drv.ms/u/s!Ajh\\_jEjaTE0SbbceogcmdwSu9ME?e=zfw6sN](https://1drv.ms/u/s!Ajh_jEjaTE0SbbceogcmdwSu9ME?e=zfw6sN)

Model	FB15K					WN18				
	MR	MRR	Hits@1	Hits@3	Hits@10	MR	MRR	Hits@1	Hits@3	Hits@10
TransE (Bordes et al., 2013)	-	0.463	0.297	0.578	0.749	-	0.495	0.113	0.888	0.943
DistMult (Yang et al., 2015)	42	0.798	-	-	0.893	655	0.797	-	-	0.946
HolE (Nickel et al., 2016)	-	0.524	0.402	0.613	0.739	-	0.938	0.930	0.945	0.947
ComplEx (Trouillon et al., 2016)	-	0.692	0.599	0.759	0.840	-	0.941	0.936	0.945	0.947
ConvE (Dettmers et al., 2018)	51	0.657	0.558	0.723	0.831	374	0.943	0.935	0.946	0.956
RotatE (Sun et al., 2019)	<u>40</u>	<u>0.797</u>	<u>0.746</u>	<u>0.830</u>	<u>0.884</u>	<u>309</u>	<u>0.949</u>	<u>0.944</u>	<u>0.952</u>	<u>0.959</u>
QuatE (Zhang et al., 2019)	<u>40</u>	<u>0.765</u>	<u>0.692</u>	<u>0.819</u>	<u>0.878</u>	<u>393</u>	<u>0.950</u>	<u>0.942</u>	<b>0.954</b>	0.959
R-GCN (Michael et al., 2018)	-	0.696	0.601	0.760	0.852	-	0.819	0.697	0.929	<b>0.964</b>
PTransE (Lin et al., 2015a)	54	0.679	0.565	0.768	0.855	472	0.890	0.931	0.942	0.945
TKRL (Xie et al., 2016)	68	-	-	-	0.694	-	-	-	-	-
TypeComplex (Jain et al., 2018)	-	0.753	0.677	-	0.869	-	0.939	0.932	-	0.951
AutoETER	<b>33</b>	<b>0.799</b>	<b>0.750</b>	<b>0.833</b>	<b>0.896</b>	<b>174</b>	<b>0.951</b>	<b>0.946</b>	<b>0.954</b>	<u>0.961</u>

Table 2: Evaluation Results on FB15K and WN18. Best results are in **bold** and second best results are underlined.

Model	FB15K-237					YAGO3-10				
	MR	MRR	Hits@1	Hits@3	Hits@10	MR	MRR	Hits@1	Hits@3	Hits@10
TransE (Bordes et al., 2013)	357	0.294	-	-	0.465	-	-	-	-	-
DistMult (Yang et al., 2015)	254	0.241	0.155	0.263	0.419	5926	0.34	0.24	0.38	0.54
ComplEx (Trouillon et al., 2016)	339	0.247	0.158	0.275	0.428	6531	0.36	0.26	0.40	0.55
ConvE (Dettmers et al., 2018)	244	0.325	0.237	0.356	0.501	<u>1671</u>	0.44	0.35	0.49	0.62
RotatE (Sun et al., 2019)	177	0.338	<u>0.241</u>	<u>0.375</u>	<u>0.533</u>	1767	<u>0.495</u>	<u>0.402</u>	<u>0.550</u>	<u>0.670</u>
QuatE (Zhang et al., 2019)	<u>172</u>	0.311	0.220	0.344	0.495	-	-	-	-	-
R-GCN (Michael et al., 2018)	-	0.249	0.151	0.264	0.417	-	-	-	-	-
PTransE (Lin et al., 2015a)	302	<b>0.363</b>	0.234	0.374	0.526	-	-	-	-	-
TypeComplex (Jain et al., 2018)	-	0.259	0.186	-	0.411	-	0.411	0.319	-	0.609
AutoETER	<b>170</b>	<u>0.344</u>	<b>0.250</b>	<b>0.382</b>	<b>0.538</b>	<b>1179</b>	<b>0.550</b>	<b>0.465</b>	<b>0.605</b>	<b>0.699</b>

Table 3: Evaluation Results on FB15K-237 and YAGO3-10 datasets.

- 1) Mean Rank (MR) of the correct triples.
- 2) Mean Reciprocal Rank (MRR) of the correct triples.
- 3) Hits@n measures the proportion of the correct triples in top-n candidate triples.

We also follow the filtered setting as the previous study (Dettmers et al., 2018) that evaluates the performance by filtering out the corrupt triples already exist in the KG.

#### 4.1.3 Baselines and Hyper-parameters

We compare the developed model AutoETER with two categories of the state-of-the-art baselines: (1) Models only considering entity-specific triples including TransE, DisMult, HolE, ComplEx, ConvE, RotatE and QuatE; (2) Models introducing additional information such as TKRL with explicit types and the type-sensitive model TypeComplex, R-GCN with graph structure and PTransE with paths. All the baselines are selected because they achieve good performance and provide source codes for ensuring the reliability and reproducibility of the results. The results of R-GCN are from (Zhang et al., 2019). The results of TKRL are from (Xie et al., 2016). The results of PTransE<sup>2</sup>,

<sup>2</sup><https://github.com/thunlp/KB2E/tree/master/PTransE>

TypeComplex<sup>3</sup> and QuatE<sup>4</sup> are obtained by using their source codes. The other results of the baselines are from (Sun et al., 2019).

We tune our model utilizing a grid search to select the optimal hyper-parameters. The optimal configurations are provided as: the batch size is set as 1024, the learning rate is  $lr = 0.0001$ , and the weights in optimization are  $\alpha_1 = 0.1, \alpha_2 = 0.5$ . The dimension of the entity and relation embeddings in entity-specific triples is  $k = 1000$ , the dimension of the type and relation embeddings in type-specific triples is  $d = 200$ . For datasets FB15K and YAGO3-10, the three fixed margins are set as  $\gamma_1 = 22, \gamma_2 = 8, \gamma_3 = 6$ . For datasets WN18 and FB15K-237,  $\gamma_1 = 10, \gamma_2 = 6, \gamma_3 = 3$ .

#### 4.2 Evaluation Results and Analyses

Table 2 and Table 3 report the evaluation results of link prediction on the four datasets. We can observe that our model AutoETER outperforms all the baselines, including the state-of-the-art models RotatE and QuatE. These results demonstrate the superiority of modeling and inferring all the relation patterns and the complex relations by our model. Specifically, AutoETER performs better than the

<sup>3</sup><https://github.com/dair-iitd/KBI/tree/master/kbi-pytorch>

<sup>4</sup><https://github.com/cheungdaven/QuatE>

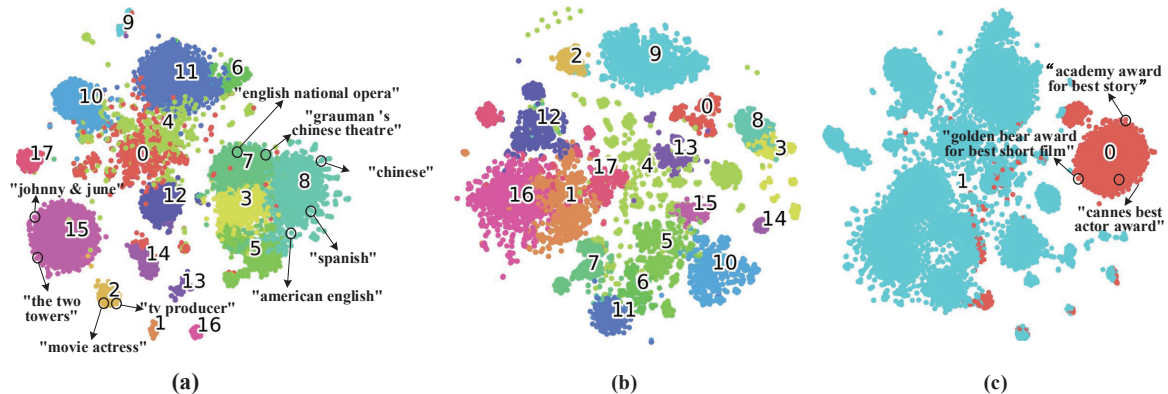


Figure 3: The visualization of type embeddings clustering on FB15K-237. (a) The clustering of the original type embeddings. (b) The clustering of the entity embeddings. (c) The clustering of the type embeddings all focusing on the relation */award/award\_category/nominated\_for*.

Model	Head Entity Prediction (Hits@10)				Tail Entity Prediction (Hits@10)			
	1-1	1-N	N-1	N-N	1-1	1-N	N-1	N-N
TransE (Bordes et al., 2013)	0.437	0.657	0.182	0.472	0.437	0.197	0.667	0.500
TransH (Wang et al., 2014)	0.668	0.876	0.287	0.645	0.655	0.398	0.833	0.672
RotatR (Lin et al., 2015b)	0.788	0.892	0.341	0.692	0.792	0.374	0.904	0.721
RotatE (Sun et al., 2019)	<u>0.922</u>	<u>0.967</u>	0.602	<u>0.893</u>	<u>0.923</u>	0.713	<u>0.961</u>	<u>0.922</u>
PTransE (Lin et al., 2015a)	0.910	0.928	<u>0.609</u>	0.838	0.912	<b>0.740</b>	0.889	0.864
AutoETER	<b>0.933</b>	<b>0.979</b>	<b>0.618</b>	<b>0.903</b>	<b>0.931</b>	<u>0.717</u>	<b>0.968</b>	<b>0.927</b>

Table 4: Evaluation results on FB15K by mapping properties of relations.

type-embodied models TKRL and TypeComplex, emphasizing the type representations learned automatically with relation-aware projection by AutoETER are more effective for inference than totally leveraging the explicit types or ignoring the diversity of type embeddings focusing on various relations. Furthermore, AutoETER outperforms RotatE because AutoETER could infer the complex relations of 1-N, N-1 as well as N-N and takes advantage of type representations. These results all illustrate the type representations learned from KGs are available to predict entities more accurately by restricting the candidate entities with type embeddings.

In view of more diverse relations existed in FB15K compared with the other three datasets, we select FB15K to evaluate link prediction performance by mapping 1-1, 1-N, N-1 and N-N relations. The results are shown in Table 4. Our model achieves better performance on both head entity prediction and tail entity prediction than other baselines particularly RotatE, which illustrates the superiority of capturing various representations of entities specific to different relations with the relation-aware projection mechanism to represent entity

Model	MR	MRR	H@1	H@3	H@10
AutoETER	<b>170</b>	<b>0.344</b>	<b>0.250</b>	<b>0.382</b>	<b>0.538</b>
-TSC	175	0.342	0.246	0.379	0.536
-TR	177	0.340	0.244	0.377	0.534

Table 5: Ablation study on FB15K-237. ‘‘H@’’ is the abbreviation of ‘‘Hits@’’.

types.

### 4.3 Ablation Study

We conduct the ablation study of our model on dataset FB15K-237 when we only omit the type similarity constraint (-TSC) and omit the type representation (-TR) from our model. Table 5 demonstrates that our model performs better than the two ablated models. It illustrates the type representation and the type similarity constraint both significantly impact the performance of link prediction and suggests that our automatically learned type representations play a pivotal role in our approach.

### 4.4 Visualization of Clustering Entity Type Representations

We utilize Kmeans to cluster the type embeddings and further employ t-SNE to implement dimen-



sional reduction for 2d visualization. As Figure 3(a) shows, some type embeddings are clustered into independent categories, while some clusters stay close to each other because these entities share many common types. For instance, *johnny&june* and *the two towers* are clustered into the same category which actually represents the type *movie* as we know. Figure 3(b) shows the clustering of the entity embeddings. It can be clearly observed that entity type clustering has better compactness than entity clustering, which demonstrates that entity type embeddings could reflect the characteristics of types. The type embeddings focusing on relation */award/award\_category/nominated\_for* are visualized in Figure 3(c). It is evident that some type embeddings representing the type *award* such as *academy award for best story* and *cannes best actor award* are clustered into the same category while others stay far away. These visualization results explain the effectiveness of our type embeddings learned automatically with relation-aware projection from the KG.

## 5 Conclusion and Future Work

In this paper, we propose an AutoETER framework to learn type representations for enriching KG embedding automatically. We introduce two classes of encoders to learn the entity-specific triple and type-specific triple embeddings, which could model and infer all the relation patterns of symmetry, inversion and composition as well as the complex 1-N, N-1 and N-N relations. We also constrain the type embeddings by the type similarity. Our experiments on four real-world datasets for link prediction illustrate the superiority of our model and the visualization of the type embeddings clustering verifies the availability of representing types automatically. In future work, we intend to extend our approach to obtain the better type representations incorporating the supervision of ontologies.

## Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (No. 61772054, 62072022), and the NSFC Key Project (No. 61632001) and the Fundamental Research Funds for the Central Universities.

## References

- Kurt Bollacker, Colin Evans, Praveen Paritosh, and Tim Sturge. 2008. Freebase: A collaboratively created graph database for structuring human knowledge. In *SIGMOD*, page 1247–1250.
- Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *NIPS*, page 2787–2795.
- Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2D knowledge graph embeddings. In *AAAI*, page 1811–1818.
- Dennis Diefenbach, Kamal Singh, and Pierre Maret. 2018. Wdaqua-core1: a question answering service for rdf knowledge bases. In *WWW*, pages 1087–1091.
- Junheng Hao, Muhao Chen, Wenchao Yu, Yizhou Sun, and Wei Wang. 2019. Universal representation learning of knowledge bases by jointly embedding instances and ontological concepts. In *KDD*, pages 1709–1719.
- He He, Anusha Balakrishnan, Mihail Eric, and Percy Liang. 2017. Learning symmetric collaborative dialogue agents with dynamic knowledge graph embeddings. In *ACL*, pages 1766–1776.
- Prachi Jain, Pankaj Kumar, Mausam1, and Soumen Chakrabarti. 2018. Type-sensitive knowledge base inference without explicit type supervision. In *ACL*, pages 75–80.
- Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S. Yu. 2020. A survey on knowledge graphs: Representation, acquisition and applications. *arXiv preprint arXiv:2002.00388*.
- Denis Krompaß, Stephan Baier, and Volker Tresp. 2015. Type-constrained representation learning in knowledge graphs. In *ISWC*, page 640–655.
- Ni Lao, Tom Mitchell, and William W. Cohen. 2011. Random walk inference and learning in a large scale knowledge base. In *EMNLP*, pages 529–539.
- Xi Victoria Lin, Richard Socher, and Caiming Xiong. 2018. Multi-hop knowledge graph reasoning with reward shaping. In *EMNLP*, page 3243–3253.
- Yankai Lin, Zhiyuan Liu, Huanbo Luan, Maosong Sun, Siwei Rao, and Song Liu. 2015a. Modeling relation paths for representation learning of knowledge bases. In *EMNLP*, pages 705–714.
- Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015b. Learning entity and relation embeddings for knowledge graph completion. In *AAAI 2015*, page 2181–2187.

- Shiheng Ma, Jianhui Ding, Weijia Jia, Kun Wang, and Minyi Guo. 2017. TransT: Type-based multiple embedding representations for knowledge graph completion. In *ECML PKDD*, pages 717–733.
- Schlichtkrull Michael, Kipf Thomas N., Bloem Peter, van den Berg Rianne, Titov Ivan, and Welling Max. 2018. Modeling relational data with graph convolutional networks. In *ESWC*.
- George A. Miller. 1995. Wordnet: A lexical database for english. *Communications of the ACM*, 38(11):39–41.
- Maximilian Nickel, Lorenzo Rosasco, and Tomaso A Poggio. 2016. Holographic embeddings of knowledge graphs. In *AAAI*, pages 1955–1961.
- Guanglin Niu, Yongfei Zhang, Bo Li, Peng Cui, Si Liu, Jingyang Li, and Xiaowei Zhang. 2020. Rule-guided compositional representation learning on knowledge graphs. In *AAAI*, pages 2950–2958.
- Meng Qu and Jian Tang. 2019. Probabilistic logic neural networks for reasoning. In *NeurIPS*.
- Oliver Ray. 2009. Nonmonotonic abductive inductive learning. *Journal of Applied Logic*, 7:329–340.
- Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. YAGO: A core of semantic knowledge. In *WWW*, pages 697–706.
- Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. RotatE: Knowledge graph embedding by relational rotation in complex space. In *ICLR*.
- Kristina Toutanova and Danqi Chen. 2015. Observed versus latent features for knowledge base and text inference. In *CVSC*, pages 57–66.
- Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *ICML*, page 2071–2080.
- Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. 2017. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12):2724–2743.
- Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. 2019. KGAT: knowledge graph attention network for recommendation. In *KDD*, pages 950–958.
- Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In *AAAI 2014*, page 1112–1119.
- Han Xiao, Minlie Huang, and Xiaoyan Zhu. 2016. From one point to a manifold: Knowledge graph embedding for precise link prediction. In *IJCAI*, pages 1315–1321.
- Ruobing Xie, Zhiyuan Liu, and Maosong Sun. 2016. Representation learning of knowledge graphs with hierarchical types. In *IJCAI*, page 2965–2971.
- Bishan Yang, Wen tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *ICLR*.
- Jun Yuan, Neng Gao, , and Ji Xiang. 2019. TransGate: Knowledge graph embedding with shared gate structure. In *AAAI*.
- Shuai Zhang, Yi Tay, Lina Yao, and Qi Liu. 2019. Quaternion knowledge graph embeddings. In *NeurIPS*, pages 2731–2741.