ROTATE: KNOWLEDGE GRAPH EMBEDDING BY RELA-TIONAL ROTATION IN COMPLEX SPACE

Zhiqing Sun,Zhi-Hong Deng,Jian-Yun Nie,Jian Tang

Abstract

We study the problem of learning representations of entities and relations in knowledge graphs for predicting missing links. The success of such a task heavily relies on the ability of modeling and inferring the patterns of (or between) the relations. In this paper, we present a new approach for knowledge graph embedding called RotatE, which is able to model and infer various relation patterns including: symmetry/antisymmetry, inversion, and composition. Specifically, the RotatE model defines each relation as a rotation from the source entity to the target entity in the complex vector space. In addition, we propose a novel self-adversarial negative sampling technique for efficiently and effectively training the RotatE model. Experimental results on multiple benchmark knowledge graphs show that the proposed RotatE model is not only scalable, but also able to infer and model various relation patterns and significantly outperform existing state-of-the-art models for link prediction.

Keywords：KG, embedding.

1 Introduction

Knowledge graphs are collections of factual triplets, where each triplet $(\mathtt{h}, \mathtt{r}, \mathtt{t})$ represents a relation $\mathtt{r}$ between a head entity $\mathtt{h}$ and a tail entity $\mathtt{t}$. Examples of real-world knowledge graphs include Freebase (Bollacker et al., 2008), Yago (Suchanek et al., 2007), and WordNet (Miller, 1995). Knowledge graphs are potentially useful to a variety of applications such as question-answering (Hao et al., 2017), information retrieval (Xiong et al., 2017), recommender systems (Zhang et al., 2016), and natural language processing (Yang , Mitchell, 2017). Research on knowledge graphs is attracting growing interests in both academia and industry communities.

Since knowledge graphs are usually incomplete, a fundamental problem for knowledge graph is predicting the missing links. Recently, extensive studies have been done on learning low-dimensional representations of entities and relations for missing link prediction (a.k.a., knowledge graph embedding) (Bordes et al., 2013; Trouillon et al., 2016; Dettmers et al., 2017) These methods have been shown to be scalable and effective. The general intuition of these methods is to model and infer the connectivity patterns in knowledge graphs according to the observed knowledge facts. For example, some relations are symmetric (e.g., marriage) while others are antisymmetric (e.g., filiation);

some relations are the inverse of other relations (e.g., hypernym and hyponym);

and some relations may be composed by others (e.g., my mother's husband is my father). It is critical to find ways to model and infer these patterns, i.e., **symmetry/antisymmetry**, **inversion**, and **composition**, from the observed facts in order to predict missing links.

Model	Score Function	
SE	$-\|\boldsymbol{W}_{r,1}\mathbf{h} - \boldsymbol{W}_{r,2}\mathbf{t}\|$	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^k, \boldsymbol{W}_{r,\cdot} \in \mathbb{R}^{k\times k}$
TransE	$-\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^k$
TransX	$-\|g_{r,1}(\mathbf{h}) + \mathbf{r} - g_{r,2}(\mathbf{t})\|$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^k$
DistMult	$\langle \mathbf{r}, \mathbf{h}, \mathbf{t} \rangle$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^k$
ComplEx	$\mathrm{Re}(\langle \mathbf{r}, \mathbf{h}, \bar{\mathbf{t}} \rangle)$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{C}^k$
HolE	$\langle \mathbf{r}, \mathbf{h} \otimes \mathbf{t} \rangle$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^k$
ConvE	$\langle \sigma(\mathrm{vec}(\sigma([\bar{\mathbf{r}}, \bar{\mathbf{h}}] * \boldsymbol{\Omega}))\boldsymbol{W}), \mathbf{t} \rangle$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^k$
RotatE	$-\|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\|$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{C}^k, \|r_i\| = 1$

Table 1: The score functions $f_r(\mathbf{h}, \mathbf{t})$ of several knowledge graph embedding models, where $\langle \cdot \rangle$ denotes the generalized dot product, $\circ$ denotes the Hadamard product, $\otimes$ denotes circular correlation, σ denotes activation function and $*$ denotes 2D convolution. $\bar{\cdot}$ denotes conjugate for complex vectors, and 2D reshaping for real vectors in ConvE model. TransX represents a wide range of TransE's variants, such as TransH (Wang et al., 2014), TransR (Lin et al., 2015b), and STransE (Nguyen et al., 2016), where $g_{r,i}(\cdot)$ denotes a matrix multiplication with respect to relation $\mathtt{r}$.

Indeed, many existing approaches have been trying to either implicitly or explicitly model one or a few of the above relation patterns (Bordes et al., 2013; Wang et al., 2014; Lin et al., 2015b;Yang et al., 2014; Trouillon et al., 2016).For example, the TransE model (Bordes et al., 2011), which represents relations as translations, aims to model the inversion and composition patterns; the DisMult model (Yang et al., 2014), which models the three-way interactions between head entities, relations, and tail entities, aims to model the symmetry pattern. However, none of existing models is capable of modeling and inferring all the above patterns. Therefore, we are looking for an approach that is able to model and infer all the three types of relation patterns.

In this paper, we propose such an approach called RotatE for knowledge graph embedding. Our motivation is from Euler's identity $e^{i\theta} = \cos\theta + i\sin\theta$, which indicates that a unitary complex number can be regarded as a rotation in the complex plane. Specifically, the RotatE model maps the entities and relations to the complex vector space and defines each relation as a rotation from the source entity to the target entity. Given a triplet $(\mathtt{h}, \mathtt{r}, \mathtt{t})$, we expect that $\mathbf{t} = \mathbf{h} \circ \mathbf{r}$, where $\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{C}^k$ are the embeddings, the modulus $|r_i| = 1$ and $\circ$ denotes the Hadamard (element-wise) product. Specifically, for each dimension in the complex space, we expect that:

$$t_i = h_i r_i, \text{ where } h_i, r_i, t_i \in \mathbb{C} \text{ and } |r_i| = 1. \tag{1}$$

It turns out that such a simple operation can effectively model all the three relation patterns: symmetric/antisymmetric, inversion, and composition. For example, a relation $\mathtt{r}$ is symmetric if and only if each element of its embedding $\mathbf{r}$, i.e. r_i, satisfies $r_i = e^{0/i\pi} = \pm 1$; two relations $\mathtt{r}_1$ and $\mathtt{r}_2$ are inverse if and only if their embeddings are conjugates: $\mathbf{r}_2 = \bar{\mathbf{r}}_1$; a relation $\mathbf{r}_3 = e^{i\boldsymbol{\theta}_3}$ is a combination of other two relations $\mathbf{r}_1 = e^{i\boldsymbol{\theta}_1}$ and $\mathbf{r}_2 = e^{i\boldsymbol{\theta}_2}$ if and only if $\mathbf{r}_3 = \mathbf{r}_1 \circ \mathbf{r}_2$ (i.e. $\boldsymbol{\theta}_3 = \boldsymbol{\theta}_1 + \boldsymbol{\theta}_2$). Moreover, the RotatE model is scalable to large knowledge graphs as it remains linear in both time and memory.

To effectively optimizing the RotatE, we further propose a novel self-adversarial negative sampling technique, which generates negative samples according to the current entity and relation embeddings. The proposed technique is very general and can be applied to many existing knowledge graph embedding models. We evaluate the RotatE on four large knowledge graph benchmark datasets including FB15k (Bordes et al., 2013), WN18 (Bordes et al., 2013), FB15k-237 (Toutanova , Chen, 2015) and WN18RR (Dettmers et al., 2017). Experimental results show that the RotatE model significantly outperforms existing state-of-the-art approaches. In addition, RotatE also outperforms state-of-the-art models on Countries (Bouchard et al., 2015), a benchmark explicitly designed for composition pattern inference and modeling. To the best of our knowledge, RotatE is the first model that achieves state-of-the-art performance on all the benchmarks.

2 Related works

Predicting missing links with **knowledge graph embedding** (KGE) methods has been extensively investigated in recent years. The general methodology is to define a score function for the triplets. Formally, let $\mathcal{E}$ denote the set of entities and $\mathcal{R}$ denote the set of relations, then a knowledge graph is a collection of factual triplets $(\mathtt{h}, \mathtt{r}, \mathtt{t})$, where $\mathtt{h}, \mathtt{t} \in \mathcal{E}$ and $\mathtt{r} \in \mathcal{R}$. Since entity embeddings are usually represented as vectors, the score function usually takes the form $f_r(\mathbf{h}, \mathbf{t})$, where $\mathbf{h}$ and $\mathbf{t}$ are head and tail entity embeddings. The score function $f_r(\mathbf{h}, \mathbf{t})$ measures the salience of a candidate triplet $(\mathtt{h}, \mathtt{r}, \mathtt{t})$. The goal of the optimization is usually to score true triplet $(\mathtt{h}, \mathtt{r}, \mathtt{t})$ higher than the corrupted false triplets $(\mathtt{h}', \mathtt{r}, \mathtt{t})$ or $(\mathtt{h}, \mathtt{r}, \mathtt{t}')$. Table1 summarizes different score functions $f_r(\mathbf{h}, \mathbf{t})$ in previous state-of-the-art methods as well as the model proposed in this paper. These models generally capture only a portion of the relation patterns. For example, TransE represents each relation as a bijection between source entities and target entities, and thus implicitly models inversion and composition of relations, but it cannot model symmetric relations; ComplEx extends DistMult by introducing complex embeddings so as to better model asymmetric relations, but it cannot infer the composition pattern. The proposed RotatE model leverages the advantages of both.

A relevant and concurrent work to our work is the TorusE (Ebisu , Ichise, 2018) model, which defines knowledge graph embedding as translations on a compact Lie group. The TorusE model can be regarded as a special case of RotatE, where the modulus of embeddings are set fixed; our RotatE is defined on the entire complex space, which has much more representation capacity. Our experiments show that this is very critical for modeling and inferring the composition patterns. Moreover, TorusE focuses on the problem of regularization in TransE while this paper focuses on modeling and inferring multiple types of relation patterns.

There are also a large body of relational approaches for modeling the relational patterns on knowledge graphs (Lao et al., 2011; Neelakantan et al., 2015; Das et al., 2016; Rockt □aschel , Riedel, 2017; Yang et al., 2017).However, these approaches mainly focus on explicitly modeling the relational paths while our proposed RotatE model implicitly learns the relation patterns, which is not only much more scalable but also provides meaningful embeddings for both entities and relations.

Another related problem is how to effectively draw negative samples for training knowledge graph embeddings. This problem has been explicitly studied by Cai , Wang (2017), which proposed a generative adversarial learning framework to draw negative samples. However, such a framework requires simultaneously training the embedding model and a discrete negative sample generator, which are difficult to optimize and also computationally expensive. We propose a self-adversarial sampling scheme which only relies on the current model. It does require any additional optimization component, which make it much more efficient.

Model	Score Function	Symmetry	Antisymmetry	Inversion	Composition
SE	$-\|\boldsymbol{W}_{r,1}\mathbf{h} - \boldsymbol{W}_{r,2}\mathbf{t}\|$	✗	✗	✗	✗
TransE	$-\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|$	✗	✓	✓	✓
TransX	$-\|g_{r,1}(\mathbf{h}) + \mathbf{r} - g_{r,2}(\mathbf{t})\|$	✓	✓	✗	✗
DistMult	$\langle \mathbf{h}, \mathbf{r}, \mathbf{t} \rangle$	✓	✗	✗	✗
ComplEx	$\mathrm{Re}(\langle \mathbf{h}, \mathbf{r}, \bar{\mathbf{t}} \rangle)$	✓	✓	✓	✗
RotatE	$-\|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\|$	✓	✓	✓	✓

Table 2: The pattern modeling and inference abilities of several models.

3 Method

In this section, we introduce our proposed RotatE model. We first introduce three important relation patterns that are widely studied in the literature of link prediction on knowledge graphs. Afterwards, we introduce our proposed RotatE model, which defines relations as rotations in complex vector space. We also show that the RotatE model is able to model and infer all three relation patterns.

3.1 MODELING AND INFERRING RELATION PATTERNS

The key of link prediction in knowledge graph is to infer the connection patterns, e.g., relation patterns, with observed facts. According to the existing literature e (Trouillon et al., 2016; Toutanova , Chen, 2015; Guu et al., 2015; Lin et al., 2015a), three types of relation patterns are very important and widely spread in knowledge graphs: symmetry, inversion and composition. We give their formal definition here:

Definition 1. *A relation* $\mathtt{r}$ *is* ***symmetric (antisymmetric)*** *if* $\forall \mathtt{x}, \mathtt{y}$

$$\mathtt{r}(\mathtt{x}, \mathtt{y}) \Rightarrow \mathtt{r}(\mathtt{y}, \mathtt{x}) \ \ (\ \mathtt{r}(\mathtt{x}, \mathtt{y}) \Rightarrow \neg \mathtt{r}(\mathtt{y}, \mathtt{x}) \ )$$

A clause with such form is a ***symmetry (antisymmetry)*** *pattern.*

Definition 2. *Relation* $\mathtt{r}_1$ *is* ***inverse*** *to relation* $\mathtt{r}_2$ *if* $\forall \mathtt{x}, \mathtt{y}$

$$\mathtt{r}_2(\mathtt{x}, \mathtt{y}) \Rightarrow \mathtt{r}_1(\mathtt{y}, \mathtt{x})$$

A clause with such form is a ***inversion*** *pattern.*

Definition 3. *Relation* $\mathtt{r}_1$ *is* ***composed*** *of relation* $\mathtt{r}_2$ *and relation* $\mathtt{r}_3$ *if* $\forall \mathtt{x}, \mathtt{y}, \mathtt{z}$

$$\mathtt{r}_2(\mathtt{x}, \mathtt{y}) \wedge \mathtt{r}_3(\mathtt{y}, \mathtt{z}) \Rightarrow \mathtt{r}_1(\mathtt{x}, \mathtt{z})$$

A clause with such form is a ***composition*** *pattern.*

According to the definition of the above three types of relation patterns, we provide an analysis of existing models on their abilities in inferring and modeling these patterns. Specifically, we provide an analysis on TransE, TransX, DistMult, and ComplEx. We did not include the analysis on HolE and ConvE since HolE is equivalent to ComplEx (Hayashi , Shimbo, 2017), and ConvE is a black box that involves two-layer neural networks and convolution operations, which are hard to analyze. The results are summarized into Table 2. We can see that no existing approaches are capable of modeling all the three relation patterns.

3.2 MODELING RELATIONS AS ROTATIONS IN COMPLEX VECTOR SPACE

In this part, we introduce our proposed model that is able to model and infer all the three types of relation patterns. Inspired by Euler's identity, we map the head and tail entities $\mathtt{h}, \mathtt{t}$ to the complex embeddings, i.e.,

$\mathbf{h}, \mathbf{t} \in \mathbb{C}^k$; then we define the functional mapping induced by each relation $\mathtt{r}$ as an element-wise rotation from the head entity $\mathbf{h}$ to the tail entity $\mathbf{t}$. In other words, given a triple $(\mathtt{h}, \mathtt{r}, \mathtt{t})$, we expect that:

$$\mathbf{t} = \mathbf{h} \circ \mathbf{r}, \quad \text{where } |r_i| = 1, \tag{2}$$

and $\circ$ is the Hadmard (or element-wise) product. Specifically, for each element in the embeddings, we have $t_i = h_i r_i$. Here, we constrain the modulus of each element of $\mathbf{r} \in \mathbb{C}^k$, i.e., $r_i \in \mathbb{C}$, to be $|r_i| = 1$. By doing this, r_i is of the form $e^{i\theta_{r,i}}$, which corresponds to a counterclockwise rotation by $\theta_{r,i}$ radians about the origin of the complex plane, and only affects the phases of the entity embeddings in the complex vector space. We refer to the proposed model as RotatE due to its rotational nature. According to the above definition, for each triple $(\mathtt{h}, \mathtt{r}, \mathtt{t})$, we define the distance function of RotatE as:

$$d_r(\mathbf{h}, \mathbf{t}) = \|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\| \tag{3}$$

By defining each relation as a rotation in the complex vector spaces, RotatE can model and infer all the three types of relation patterns introduced above. Formally, we have following results[1]:

Lemma 1. *RotatE can infer the symmetry/antisymmetry pattern.*

Lemma 2. *RotatE can infer the inversion pattern.*

Lemma 3. *RotatE can infer the composition pattern.*

These results are also summarized into Table 2. We can see that the RotatE model is the only model that can model and infer all the three types of relation patterns.

Connection to TransE. From Table 2, we can see that TransE is able to infer and model all the other relation patterns except the symmetry pattern. The reason is that in TransE, any symmetric relation will be represented by a $\mathbf{0}$ translation vector. As a result, this will push the entities with symmetric relations to be close to each other in the embedding space. RotatE solves this problem and is a able to model and infer the symmetry pattern. An arbitrary vector $\mathbf{r}$ that satisfies $r_i = \pm 1$ can be used for representing a symmetric relation in RotatE, and thus the entities having symmetric relations can be distinguished. Different symmetric relations can be also represented with different embedding vectors. Figure 1,Figure 2,Figure3 provides illustrations of TransE and RotatE with only 1-dimensional embedding and shows how RotatE models a symmetric relation.

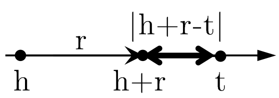

Figure 1: TransE models r as translation in real line.

[1]We relegate all proofs to the appendix.

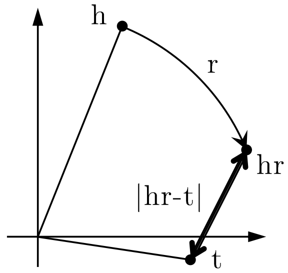

Figure 2: RotatE models r as rotation in complex plane

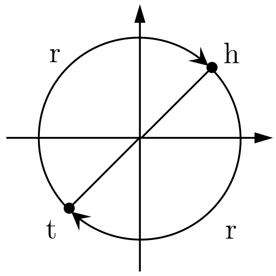

Figure 3: RotatE: an example of modeling symmetric relations r with ri = −1

3.3 OPTIMIZATION

Negative sampling has been proved quite effective for both learning knowledge graph embedding (Trouillon et al., 2016) and word embedding (Mikolov et al., 2013). Here we use a loss function similar to the negative sampling loss (Mikolov et al., 2013) for effectively optimizing distance-based models:

$$L = -\log \sigma(\gamma - d_r(\mathbf{h}, \mathbf{t})) - \sum_{i=1}^{n} \frac{1}{k} \log \sigma(d_r(\mathbf{h}'_i, \mathbf{t}'_i) - \gamma), \tag{4}$$

where γ is a fixed margin, σ is the sigmoid function, and $(\mathrm{h}'_i, \mathrm{r}, \mathrm{t}'_i)$ is the i-th negative triplet.

We also propose a new approach for drawing negative samples. The negative sampling loss samples the negative triplets in a uniform way. Such a uniform negative sampling suffers the problem of inefficiency since many samples are obviously false as training goes on, which does not provide any meaningful information. Therefore, we propose an approach called self-adversarial negative sampling, which samples negative triples according to the current embedding model. Specifically, we sample negative triples from the following distribution:

$$p(h'_j, r, t'_j | \{(h_i, r_i, t_i)\}) = \frac{\exp \alpha f_r(\mathbf{h}'_j, \mathbf{t}'_j)}{\sum_i \exp \alpha f_r(\mathbf{h}'_i, \mathbf{t}'_i)} \tag{5}$$

where α is the temperature of sampling. Moreover, since the sampling procedure may be costly, we treat the above probability as the weight of the negative sample. Therefore, the final negative sampling loss with self-adversarial training takes the following form:

$$L = -\log \sigma(\gamma - d_r(\mathbf{h}, \mathbf{t})) - \sum_{i=1}^{n} p(h'_i, r, t'_i) \log \sigma(d_r(\mathbf{h}'_i, \mathbf{t}'_i) - \gamma) \tag{6}$$

In the experiments, we will compare different approaches for negative sampling.

4 Implementation details

4.1 Experimental Setting

We evaluate our proposed model on four widely used knowledge graphs. The statistics of these knowledge graphs are summarized into Table 3.

Dataset	#entity	#relation	#training	#validation	#test
FB15k	14,951	1,345	483,142	50,000	59,071
WN18	40,943	18	141,442	5,000	5,000
FB15k-237	14,541	237	272,115	17,535	20,466
WN18RR	40,943	11	86,835	3,034	3,134

Table 3: Number of entities, relations, and observed triples in each split for four benchmarks.

- FB15k (Bordes et al., 2013) is a subset of Freebase (Bollacker et al., 2008), a large-scale knowledge graph containing general knowledge facts. Toutanova , Chen (2015) showed that almost 81% of the test triplets $(\mathtt{x}, \mathtt{r}, \mathtt{y})$ can be inferred via a directly linked triplet $(\mathtt{x}, \mathtt{r}', \mathtt{y})$ or $(\mathtt{y}, \mathtt{r}', \mathtt{x})$. Therefore, the key of link prediction on FB15k is to model and infer the **symmetry/antisymmetry** and **inversion** patterns.
- WN18 (Bordes et al., 2013) is a subset of WordNet (Miller, 1995), a database featuring lexical relations between words. This dataset also has many inverse relations. So the main relation patterns in WN18 are also **symmetry/antisymmetry** and **inversion**.
- FB15k-237 (Toutanova , Chen, 2015) is a subset of FB15k, where inverse relations are deleted. Therefore, the key of link prediction on FB15k-237 boils down to model and infer the **symmetry/antisymmetry** and **composition** patterns.
- WN18RR (Dettmers et al., 2017)is a subset of WN18. The inverse relations are deleted, and the main relation patterns are **symmetry/antisymmetry** and **composition**.

Hyperparameter Settings. We use Adam (Kingma , Ba, 2014) as the optimizer and fine-tune the hyperparameters on the validation dataset. The ranges of the hyperparameters for the grid search are set as follows: embedding dimension $k \in \{125, 250, 500, 1000\}$, batch size $b \in \{512, 1024, 2048\}$, self-adversarial sampling temperature $\alpha \in \{0.5, 1.0\}$, and fixed margin $\gamma \in \{3, 6, 9, 12, 18, 24, 30\}$. Both the real and imaginary parts of the entity embeddings are uniformly initialized, and the phases of the relation embeddings are uniformly initialized between 0 and 2π. No regularization is used since we find that the fixed margin γ could prevent our model from over-fitting.

Evaluation Settings. We evaluate the performance of link prediction in the filtered setting: we rank test triples against all other candidate triples not appearing in the training, validation, or test set, where candidates are generated by corrupting subjects or objects: $(\mathtt{h}', \mathtt{r}, \mathtt{t})$ or $(\mathtt{h}, \mathtt{r}, \mathtt{t}')$. Mean Rank (MR), Mean Reciprocal Rank (MRR) and Hits at N (H@N) are standard evaluation measures for these datasets and are evaluated in our experiments.

Baseline. Apart from RotatE, we propose a variant of RotatE as baseline, where the modulus of the entity embeddings are also constrained: $|h_i| = |t_i| = C$, and the distance function is thus $2C \left\| \sin \frac{\boldsymbol{\theta}_h + \boldsymbol{\theta}_r - \boldsymbol{\theta}_t}{2} \right\|$. In this way, we can investigate how RotatE works without modulus information and with only phase information. We refer to the baseline as pRotatE. It is obvious to see that pRotatE can also model and infer all the three relation patterns.

	FB15k					WN18				
	MR	MRR	H@1	H@3	H@10	MR	MRR	H@1	H@3	H@10
TransE [♥]	-	.463	.297	.578	.749	-	.495	.113	.888	.943
DistMult [♦]	42	.798	-	-	**.893**	655	.797	-	-	.946
HolE	-	.524	.402	.613	.739	-	.938	.930	.945	.949
ComplEx	-	.692	.599	.759	.840	-	.941	.936	.945	.947
ConvE	51	.657	.558	.723	.831	374	.943	.935	.946	.956
pRotatE	43	**.799**	**.750**	.829	.884	**254**	.947	.942	.950	.957
RotatE	**40**	.797	.746	**.830**	.884	309	**.949**	**.944**	**.952**	**.959**

Table 4: Results of several models evaluated on the FB15K and WN18 datasets. Results of [♥] are taken from (Nickel et al., 2016) and results of [♦] are taken from (Kadlec et al., 2017). Other results are taken from the corresponding original papers.

	FB15k-237					WN18RR				
	MR	MRR	H@1	H@3	H@10	MR	MRR	H@1	H@3	H@10
TransE [♥]	357	.294	-	-	.465	3384	.226	-	-	.501
DistMult	254	.241	.155	.263	.419	5110	.43	.39	.44	.49
ComplEx	339	.247	.158	.275	.428	5261	.44	.41	.46	.51
ConvE	244	.325	.237	.356	.501	4187	.43	.40	.44	.52
pRotatE	178	.328	.230	.365	.524	**2923**	.462	.417	.479	.552
RotatE	**177**	**.338**	**.241**	**.375**	**.533**	3340	**.476**	**.428**	**.492**	**.571**

Table 5: Results of several models evaluated on the FB15k-237 and WN18RR datasets. Results of [♥] are taken from (Nguyen et al., 2017). Other results are taken from (Dettmers et al., 2017).

4.2 Main Results

We compare RotatE to several state-of-the-art models, including TransE (Bordes et al., 2013), DistMult (Yang et al., 2014), ComplEx (Trouillon et al., 2016), HolE (Nickel et al., 2016), and ConvE (Dettmers et al., 2017), as well as our baseline model pRotatE, to empirically show the importance of modeling and inferring the relation patterns for the task of predicting missing links.

Table 4 summarizes our results on FB15k and WN18. We can see that RotatE outperforms all the state-of-the-art models. The performance of pRotatE and RotatE are similar on these two datasets. Table 5 summarizes our results on FB15k-237 and WN18RR, where the improvement is much more significant. The difference between RotatE and pRotatE is much larger on FB15k-237 and WN18RR, where there are a lot of composition patterns. This indicates that modulus is very important for modeling and inferring the composition pattern.

Moreover, the performance of these models on different datasets is consistent with our analysis on the three relation patterns (Table 2):

5 Results and analysis

- On FB15K, the main relation patterns are symmetry/antisymmetry and inversion. We can see that ComplEx performs well while TransE does not perform well since ComplEx can infer both symmetry/antisymmetry and inversion patterns while TransE cannot infer symmetry pattern. Surprisingly, DistMult achieves good performance on this dataset although it cannot model the antisymmetry and inversion patterns. The reason is that for most of the relations in FB15K, the types of head entities and tail entities are different. Although DistMult gives the same score to a true triplet $(\mathtt{h}, \mathtt{r}, \mathtt{t})$ and its opposition triplet $(\mathtt{t}, \mathtt{r}, \mathtt{h})$, $(\mathtt{t}, \mathtt{r}, \mathtt{h})$ is usually impossible to be valid since the entity type of t does not match the head entity type of h. For example, DistMult assigns the same score to (Obama, nationality, USA) and (USA, nationality, Obama). But (USA, nationality, Obama) can be simply predicted as false since USA cannot be the head entity of the relation nationality.

- On WN18, the main relation patterns are also symmetry/antisymmetry and inversion. As expected, ComplEx still performs very well on this dataset. However, different from the results on FB15K, the performance of DistMult significantly decreases on WN18. The reason is that DistMult cannot model antisymmetry and inversion patterns, and almost all the entities in WN18 are words and belong to the same entity type, which do not have the same problem as FB15K.
- On FB15k-237, the main relation pattern is composition. We can see that TransE performs really well while ComplEx does not perform well. The reason is that, as discussed before, TransE is able to infer the composition pattern while ComplEx cannot infer the composition pattern.
- On WN18RR, one of the main relation patterns is the symmetry pattern since almost each word has a symmetric relation in WN18RR, e.g., $also\_see$ and $similar\_to$. TransE does not well on this dataset since it is not able to model the symmetric relations.

Besides, we changed the fomula and get worse outcomes,compared to the former one.These experiments justify the validity of the original one,using Hadmard (or element-wise) product.

6 Conclusion and future work

We have recurrented the classical knowledge graph embedding method called RotatE, which represents entities as complex vectors and relations as rotations in complex vector space. In addition, we reused the proposed self-adversarial negative sampling technique for efficiently and effectively training the RotatE model. Our experimental results show get close outcomes of the original paper's. Moreover, we changed the fomula and get worse outcomes,which justifys the validity of the original one. In the future, we plan to evaluate the RotatE model on more datasets and leverage a probabilistic framework to model the uncertainties of entities and relations.

7 References

Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In Proceedings of the 2008 ACM SIGMOD international conference on Management of data, pp. 1247–1250. AcM, 2008.

Antoine Bordes, Jason Weston, Ronan Collobert, Yoshua Bengio, et al. Learning structured embeddings of knowledge bases. In AAAI, volume 6, pp. 6, 2011.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In Advances in neural information processing systems, pp. 2787–2795, 2013.

Guillaume Bouchard, Sameer Singh, and Theo Trouillon. On approximate reasoning capabilities of lowrank vector spaces. AAAI Spring Syposium on Knowledge Representation and Reasoning (KRR): Integrating Symbolic and Neural Approaches, 2015.

Liwei Cai and William Yang Wang. Kbgan: Adversarial learning for knowledge graph embeddings.arXiv preprint arXiv:1711.04071, 2017.

Rajarshi Das, Arvind Neelakantan, David Belanger, and Andrew McCallum. Chains of reasoning over entities, relations, and text using recurrent neural networks. arXiv preprint arXiv:1607.01426, 2016.

Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2d knowledge graph embeddings. arXiv preprint arXiv:1707.01476, 2017.

Takuma Ebisu and Ryutaro Ichise. Toruse: Knowledge graph embedding on a lie group. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, pp. 1819–1826. AAAI Press, 2018.

Kelvin Guu, John Miller, and Percy Liang. Traversing knowledge graphs in vector space. arXiv preprint arXiv:1506.01094, 2015.

Yanchao Hao, Yuanzhe Zhang, Kang Liu, Shizhu He, Zhanyi Liu, Hua Wu, and Jun Zhao. An endto-end model for question answering over knowledge base with cross-attention combining global knowledge. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), volume 1, pp. 221–231, 2017.

Katsuhiko Hayashi and Masashi Shimbo. On the equivalence of holographic and complex embeddings for link prediction. arXiv preprint arXiv:1702.05563, 2017.

Shizhu He, Kang Liu, Guoliang Ji, and Jun Zhao. Learning to represent knowledge graphs with gaussian embedding. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, pp. 623–632. ACM, 2015.

Rudolf Kadlec, Ondrej Bajgar, and Jan Kleindienst. Knowledge base completion: Baselines strike back. arXiv preprint arXiv:1705.10744, 2017.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.698 2014.

Ni Lao, Tom Mitchell, and William W Cohen. Random walk inference and learning in a large scale knowledge base. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 529–539. Association for Computational Linguistics, 2011.

Yankai Lin, Zhiyuan Liu, Huanbo Luan, Maosong Sun, Siwei Rao, and Song Liu. Modeling relation paths for representation learning of knowledge bases. arXiv preprint arXiv:1506.00379, 2015a.

Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. Learning entity and relation embeddings for knowledge graph completion. In AAAI, volume 15, pp. 2181–2187, 2015b.

Farzaneh Mahdisoltani, Joanna Biega, and Fabian M Suchanek. Yago3: A knowledge base from multilingual wikipedias. In CIDR, 2013.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems, pp.

3111–3119, 2013.

George A Miller. Wordnet: a lexical database for english. Communications of the ACM, 38(11): 39–41, 1995.

Arvind Neelakantan, Benjamin Roth, and Andrew McCallum. Compositional vector space models for knowledge base completion. arXiv preprint arXiv:1504.06662, 2015.

Dai Quoc Nguyen, Tu Dinh Nguyen, Dat Quoc Nguyen, and Dinh Phung. A novel embedding model for knowledge base completion based on convolutional neural network. arXiv preprint arXiv:1712.02121, 2017.

Dat Quoc Nguyen, Kairit Sirts, Lizhen Qu, and Mark Johnson. Stranse: a novel embedding model of entities and relationships in knowledge bases. arXiv preprint arXiv:1606.08140, 2016.

Maximilian Nickel, Lorenzo Rosasco, Tomaso A Poggio, et al. Holographic embeddings of knowledge graphs. In AAAI, volume 2, pp. 3–2, 2016.

Tim Rocktäschel and Sebastian Riedel. End-to-end differentiable proving. In Advances in Neural Information Processing Systems, pp. 3788–3800, 2017.

Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. Yago: a core of semantic knowledge. In Proceedings of the 16th international conference on World Wide Web, pp. 697–706. ACM, 2007.

Kristina Toutanova and Danqi Chen. Observed versus latent features for knowledge base and text inference. In Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality, pp. 57–66, 2015.

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In International Conference on Machine Learning, pp. 2071–2080, 2016.

Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph embedding by translating on hyperplanes. In AAAI, volume 14, pp. 1112–1119, 2014.

Chenyan Xiong, Russell Power, and Jamie Callan. Explicit semantic ranking for academic search via knowledge graph embedding. In Proceedings of the 26th international conference on world wide web, pp. 1271–1279. International World Wide Web Conferences Steering Committee, 2017.

Bishan Yang and Tom Mitchell. Leveraging knowledge bases in lstms for improving machine reading. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), volume 1, pp. 1436–1446, 2017.

Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. arXiv preprint arXiv:1412.6575, 2014.

Fan Yang, Zhilin Yang, and William W Cohen. Differentiable learning of logical rules for knowledge base completion. CoRR, abs/1702.08367, 2017.

Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. Collaborative knowledge base embedding for recommender systems. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 353–362. ACM, 2016.