Cosine-based Embedding for Completing Schematic Knowledge

Huan Gao[1,2], Xianda Zheng[3], Weizhuo Li[1,2(✉)], Guilin Qi[1,2,3], and Meng Wang[1,2]

[1] School of Computer Science and Engineering, Southeast University, Nanjing, China.
liweizhuo@amss.ac.cn {gh,zhengxianda,gqi,meng.wang}@seu.edu.cn
[2] Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, China.
[3] School of Cyber Science and Engineering, Southeast University, Nanjing, China.

Abstract. Schematic knowledge, as a critical ingredient of knowledge graphs, defines logical axioms based on concepts to support for eliminating heterogeneity, integration, and reasoning over knowledge graphs(KGs). Although some well-known KGs contain large scale schematic knowledge, they are far from complete, especially schematic knowledge stating that two concepts have subclassOf relations (also called subclassOf axioms) and schematic knowledge stating that two concepts are logically disjoint (also called disjointWith axioms). One of the most important characters of these axioms is their logical properties such as transitivity and symmetry. Current KG embedding models focus on encoding factual knowledge (i.e., triples) in a KG and cannot directly be applied to further schematic knowledge (i.e., axioms) completion. The main reason is that they ignore these logical properties. To solve this issue, we propose a novel model named CosE for schematic knowledge. More precisely, CosE projects each concept into two semantic spaces. One is an angle-based semantic space that is utilized to preserve transitivity or symmetry of an axiom. The other is a translation-based semantic space utilized to measure the confidence score of an axiom. Moreover, two score functions tailored for $subclassOf$ and $disjointWith$ are designed to learn the representation of concepts with these two relations sufficiently. We conduct extensive experiments on link prediction on benchmark datasets like YAGO and FMA ontologies. The results indicate that CosE outperforms state-of-the-art methods and successfully preserve the transitivity and symmetry of axioms.

Key words: Schematic knowledge, Embedding, Logical properties

1 Introduction

Schematic knowledge, as a critical ingredient of knowledge graphs (KGs), defines logical axioms based on concepts to support for eliminating heterogeneity, integration, and reasoning over knowledge graphs. Although some well-known knowledge graphs, e.g., WordNet [1], DBpedia [2], YAGO [3], contain lots of triples and axioms but they are far from complete [4]. Right now, DBpedia contains more than 800 concepts, but there are only 20 disjointWith axioms in it. In another way, traditional reason methods cannot obtain all missing axioms. For example, given two axioms (Boy, $subclassOf$, Children) and (Male_Person, $subclassOf$, Person), If the relationship $subclassOf$ between *Children* and *Male_Person* is missing, then it is hard to get the axiom (Boy, $subclassOf$, Person) by traditional rule-based reasoning methods. Moreover, the sparsity of schematic knowledge limits the applications of knowledge graphs such as question-answer and data integration. Therefore, it is of importance to improve the completion of schematic knowledge.

Knowledge graph embedding, which aims to encode a knowledge graph into a low dimension continuous vector space, has shown to be benefited knowledge graph completion by link prediction [5]. Each entity or relation is represented as a vector or a matrix which contains rich semantic information and can be applied to link prediction [6]. The typical KG embedding models, such as TransE [7], TransH [8] and TransR [9], treat a relation as a translation from head entity to tail entity. Some other models, like RESCAL [10], DistMult [11], HolE [12] and ComplEx [13], adopt different compositional operations to capture rich interactions of embeddings. Recently, several studies, such as EmbedS [14] and TransC [15], pay an attention to the completion of axioms for given schematic knowledge. They encoded instances as vectors and concepts as spheres so that they could preserve the transitivity of some relations. Although existing embedding methods have achieved certain success in KG completion, most of them only focus on entity-level triples but ignore the logical properties of axioms asserted in schematic knowledge. Therefore, It is hard to directly employ these methods on the tasks related to schematic knowledge such as completion, reasoning, repairing. For instance, two concepts C_i and C_j with symmetry relation denoted by (C_i, r, C_j) , in a typical translation-based method, e.g., TransE [7], all the concepts and relations are projected into a translation-based semantic space. In this space, the score of this axiom $||\mathbf{C_i} + \mathbf{r} - \mathbf{C_j}||_2$ is not equal to $||\mathbf{C_j} + \mathbf{r} - \mathbf{C_i}||_2$. Thus, the symmetry of this axiom is lost. To solve this problem, we need to explore a new embedding method for axioms that can simultaneously preserve the transitivity of $subclassOf$ and symmetry of $disjointWith$ well.

In this paper, we propose a novel embedding model, namely CosE (**Cos**ine-based **E**mbedding), for learning concepts and relationships in schematic knowledge. In previous studies [16], the authors showed that all the axioms could be reduced to subclassOf axioms and disjointWith axioms. Hence, our model is mainly tailored for learning the representation of axioms asserted with these two relations. To preserve logical properties and measure the confidence of a potential missing axiom, CosE is implemented by projecting concepts based on relations to an angle-based semantic space and a translation-based semantic space. In CosE, for a concept, its vector and valid length in the angle-based semantic space are learned and utilized to preserve logical properties. Another vector for this concept in the translation-based semantic space is used to measure the confidence score of an axiom. The effectiveness of CosE is verified by link prediction experiments on standard benchmark datasets. The experimental results indicate that CosE can achieve state-of-the-art performance comparing existing methods in most cases.

The main contributions of our work are summarized as follows:

1. To the best of our knowledge, we are the first to propose one kind of embedding method for schematic knowledge, which can simultaneously preserve the transitivity of $subclassOf$ and symmetry of $disjointWith$ well.
2. We define two score functions based on angle-based semantic space and translation-based semantic space which are tailored for subclassOf axioms and disjointWith axioms in order to sufficiently learn the representation of concepts.
3. We conduct extensive experiments on benchmark datasets for evaluating effectiveness of our methods. Experimental results on link prediction demonstrate that our method can outperform state-of-the-art methods in most cases and successfully preserve the transitivity and symmetry of axioms.

2 Related work

In this section, we divide the research efforts into knowledge graph embedding and schema embedding, and review them as follows.

2.1 Knowledge graph embedding

There are two mainstream methods for knowledge embedding: translational distance models and semantic matching models [5]. The former uses distance-based scoring functions, and the latter employs similarity-based ones.

In recent years, *knowledge graph embedding* has been widely studied, see [7–10, 17]. It aims to effectively encode a relational knowledge base into a low dimensional continuous vector space and achieves success on relational learning tasks like link prediction [18] and triple classification [17]. Various techniques have been devised for this task, several improved models are proposed. In order to find the most related relation between two entities, TransA [19] sets different weights according to different axis directions. In TransH [8], each entity is projected into the relation-specific hyperplane that is perpendicular to the relationship embedding. TransR [9] and TransD [20] still follow the principle of TransH. These two models project entities into relation-specific spaces in order to process complex relations. Moreover, the translation assumption only focuses on the local information in triples such as a single triple, which may fail to make full use of global graph information in KGs.

Another type of embedding methods conducted semantic matching using neural network architectures and obtained the encourage results on link prediction of KG completion such as MLP [21], NAM [22], R-GCN [23]. Moreover, ProjE [24] and ConvE [25] optimized the complex feature space and changed in the architecture of underlying models. Both of them achieved better performances compared with the models without complex feature engineering.

KG embedding methods mainly focus on instance-level triples of knowledge graphs, which utilize triples of KGs to obtain the representations of entities and relations, but they are not suitable for encoding the schematic knowledge of ontologies because they can not persevere transitivity and symmetry of axioms in their models well, which are essential characters applied in enriching incomplete data.

2.2 Schema embedding

Some methods have been proposed for embedding of schematic knowledge in a simple ontology language called RDF Schema (or RDFS). On2Vec [26] employed translation-based embedding method for ontology population, which integrated matrices that transformed the head and tail entities in order to characterize the transitivity of some relations. To represent concepts, instances, and relations differently in the same space, EmbedS [14] and TransC [15] encoded instances as vectors and concepts as spheres so that they can deal with the transitivity of isA relations (i.e., instanceOf, subclassOf). In addition, [27] proposed a joint model, called KALE, which embeds factual knowledge and logical rules simultaneously. [28] improved this model, referred to as RUGE, which could learn simultaneously from labeled triples, unlabeled triples, and soft rules in an iterative manner. Both of them treated RDFS as rules to improve the performances of embedding models [4].

Although above embedding models tailored for RDFS enable to preserve the transitivity of some relations (e.g., $subclassOf$) in their semantic spaces, it is not enough for them to express other kinds of schematic knowledge, especially those expressed with the $disjointWith$ relation. Moreover, it is hard for their score functions to simultaneously describe $disjointWith$ among concepts and preserve its symmetry well.

To the best of our knowledge, our method is the first embedding method for schematic knowledge of ontologies based on cosine measure that can simultaneously preserve the transitivity of $subclassOf$ and symmetry of $disjointWith$ well.

3 Cosine-based embedding for schematic knowledge

In this section, we first present the framework of CosE and introduce two semantic spaces. Then, we define two score functions for these two spaces, one is defined for angle-based semantic space to preserve logic properties of axioms. The other is defined for translation-based semantic space to predict the confidence score of a missing axiom. Finally, we present the training model of CosE.

3.1 CosE

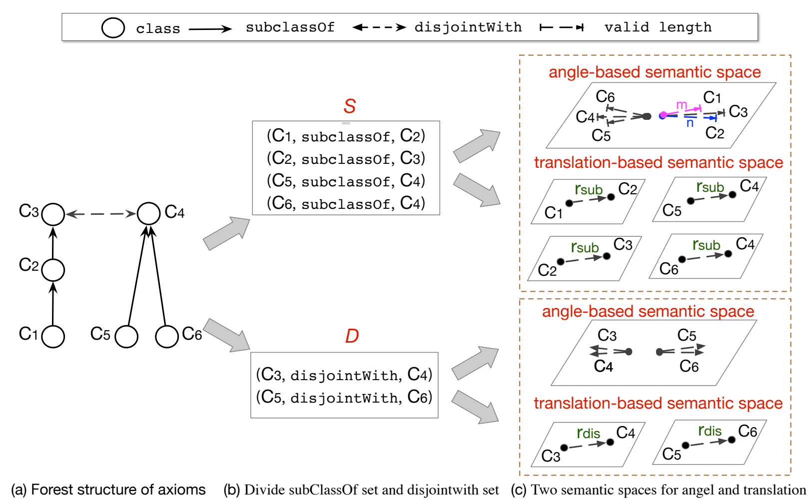

Fig. 1: An overview of cosine-based embedding for Schematic knowledge

In most cases, for an axiom (C_i, r, C_j) with transitive or symmetry asserted in schematic knowledge, existing KG embedding models only treat r as a symbol and ignore its logical property, so these models cannot represent transitivity and symmetry precisely. Therefore, in order to get a better embedding of schematic knowledge, the logical properties of relations should be considered.

Figure 1 shows a framework of CosE that describes how concepts are represented based on the logical properties of their relations. We use solid lines and dotted lines denoted as $subclassOf$ and $disjointWith$ relations, respectively.

Given a set of axioms, CosE first separates all axioms into two sets where S contains all subclassOf axioms and D contains all disjointWith axioms. For example, three axioms $(C_1, subclass\text{-}Of, C_2)$, $(C_2, subclassOf, C_3)$ and $(C_1, disjointWith, C_4)$ are asserted, we obtain S=$\{(C_2, subclassOf, C_3), (C_1, subclassOf, C_2)\}$ and D=$\{(C_1, disjointWith, C_4)\}$. Then, for each set, all concepts are projected into two semantic spaces, one of which is an angle-based semantic space for modeling the logical properties of relations and a translation-based semantic space for measuring confidence score of given axioms. As most of subclassOf axioms and disjointWith axioms are 1-to-n and n-to-n relations, existing score functions of translation-based methods are still not good at dealing with these complex relations. To measure the confidence score more precisely, concept vectors are projected into a translation-based semantic space by a mapping matrix $\mathbf{M_{C_iC_j}}$ where C_i and C_j are concepts in given schematic knowledge. For an axiom $(C_1, subclassOf, C_2)$, head concept C_1 and tail concept C_2 are projected by $\mathbf{M_{C_iC_j}}$ that means each axiom is projected into its own translation-based semantic space. In the above example, let $\mathbf{C^{12}_{1\perp}}$ and $\mathbf{C^{12}_{2\perp}}$ be the projected vectors of C_1 and C_2 by $\mathbf{M_{C_1C_2}}$, and let $\mathbf{C^{23}_{2\perp}}$ and $\mathbf{C^{23}_{3\perp}}$ be the projected vectors of C_2 and C_3 by $\mathbf{M_{C_2C_3}}$. $\mathbf{C^{12}_{2\perp}}$ and $\mathbf{C^{23}_{2\perp}}$ are the vectors of C_2, but they locate in different translation-based semantic spaces. It is helpful that all the axioms can be expressed well in their own translation-based semantic space. Given an axiom (C_i, r, C_j), its mapping matrix is defined as follows:

$$\mathbf{M_{C_iC_j}} = \mathbf{C_{ip}}\mathbf{C_{jp}^{\top}} + \mathbf{I^{n\times n}}, \tag{1}$$

where $\mathbf{C_{ip}} \in \mathbb{R}^n$ and $\mathbf{C_{jp}} \in \mathbb{R}^n$ are projection vectors for head concept C_i and the tail concept C_j. With the mapping matrix, the projected vectors about C_i and C_j are defined as follows:

$$\mathbf{C_{i\perp}} = \mathbf{M_{C_iC_j}}\mathbf{C_i}, \qquad \mathbf{C_{j\perp}} = \mathbf{M_{C_iC_j}}\mathbf{C_j}. \tag{2}$$

Notice that the logical properties are expressed precisely by the angle-based semantic space. In the above examples, to deal with the transitivity, if we have two correct axioms $(C_1, subclassof, C_2)$ and $(C_2, subclassOf, C_3)$, vectors of C_1 and C_2 should be similar (i.e., $cos(\mathbf{C_i}, \mathbf{C_j}) \approx 0$) in subclassOf angle-based semantic space and the length of C_1 is smaller than C_2. Similarly, the angle between vectors of C_2 and C_3 should be approximated $0°$ and the length of C_2 is also smaller than C_3. For dealing with the symmetry, CosE only removes the length constraints because cosine function has symmetry property. For example, if $(C_1, disjointWith, C_4)$ is a correct axiom, then the vectors of C_1 and C_4 are similar in the disjointWith angle-based semantic space. Therefore, CosE can simultaneously preserve transitivity and symmetry well.

3.2 Score function

In this section, we introduce score functions of CosE in detail. As CosE projects each concept into two semantic spaces, we design two kinds of score functions to measure the score of each axiom. One is utilized to preserve logical properties in angle-based semantic space. The other is served for measuring the confidence of axioms in translation-based semantic space. Given an axiom (C_i, r, C_j), the score function of this axiom is defined as:

$$f(C_i, r, C_j) = f_a(C_i, r, C_j) + f_t(C_i, r, C_j), \tag{3}$$

where $f_a(C_i, r, C_j)$ and $f_t(C_i, r, C_j)$ are score functions defined in the angle-based semantic space and translation-based semantic space, respectively.

The angle-based semantic space aims to preserve logical properties. We assume that relations with different properties should be measured by different score functions. For a subclassOf axiom (C_i, r_s, C_j), concepts C_i and C_j are encoded as $(\mathbf{C_i}, \mathbf{m})$ and $(\mathbf{C_j}, \mathbf{n})$, where $\mathbf{C_i}$ and $\mathbf{C_j}$ are the vector representations of concepts C_i and C_j, $\mathbf{m}$ and $\mathbf{n}$ are two vectors that defined to obtain their valid lengths for persevering the transitivity of concepts. The score function of the subclassOf axiom is defined as follows.

$$f_a(C_i, r_s, C_j) = 1 - cos(\mathbf{C_i}, \mathbf{C_j}) + ||\mathbf{m}||_2 - ||\mathbf{n}||_2, \tag{4}$$

where $\mathbf{C_i} \in \mathbb{R}^n$ and $\mathbf{C_j} \in \mathbb{R}^n$ are their vectors in the angle-based semantic space. $||\mathbf{m}||_2$ and $||\mathbf{n}||_2$ are valid length corresponding to C_i and C_j. Notice that these four vectors are parameters that could be obtained when training procedure is accomplished.

For one disjointWith axiom (C_i, r_d, C_j), CosE removes the length constraints of vectors. Its score function is defined as:

$$f_a(C_i, r_d, C_j) = 1 - cos(\mathbf{C_i}, \mathbf{C_j}), \tag{5}$$

where $\mathbf{C_i} \in \mathbb{R}^n$ and $\mathbf{C_j} \in \mathbb{R}^n$. For the axiom $(C_i, disjointWith, C_j)$, the score of $f_a(C_i, r_d, C_j)$ and $f_a(C_j, r_d, C_i)$ are the same because of the symmetry of cosine measure. It means CosE can preserve the symmetry of disjointWith axioms.

Although the angle-based semantic space can keep the logical properties of axioms, it is hard to measure the confidence score of each axiom. Particularly, the subclassOf axioms and disjointWith axioms are the typical multivariate relations so that the score function designed for angle-based semantic space is not enough to measure the confidence of axioms with multivariate relations. To solve this issue, we introduce a new score function of an axiom w.r.t the translation-based semantic space to measure the confidence of each axiom as follows.

$$f_t(C_i, r, C_j) = ||\mathbf{C_{i\perp}}' + \mathbf{r} - \mathbf{C_{j\perp}}'||_2 \tag{6}$$

where $\mathbf{C_{i\perp}}' \in \mathbb{R}^n$, $\mathbf{r} \in \mathbb{R}^n$ and $\mathbf{C_{j\perp}}' \in \mathbb{R}^n$ are the projection vectors in translation-based semantic space. In our experiments, we enforce constrains as $||\mathbf{C_i}||_2 \leq 1$, $||\mathbf{C_j}||_2 \leq 1$, $||\mathbf{C_{i\perp}}'||_2 \leq 1$ and $||\mathbf{C_{j\perp}}'||_2 \leq 1$.

3.3 Training Model

To train CosE, every axiom in our training set has been labeled to indicate whether the axiom is positive or negative. However, most of existing ontologies only contain positive axioms. Thus we need to generate negative axioms by corrupting positive axioms. For an axiom (C_i, r, C_j), we replace C_i or C_j to generate a negative triple (C_i', r, C_j) or (C_i, r, C_j') by a uniform probability distribution.

For each axiom, we adopt the margin rank loss to train the representation of concepts and relations, where ξ and ξ' denote a positive axiom and a negative one w.r.t the type of relation, respectively. $\mathcal{T}$ and $\mathcal{T}'$ are used to denote the sets of positive axioms and negative ones, respectively. For an axiom with $subclassOf$ relation, the margin-based ranking loss is defined as:

$$\mathcal{L}_{sub} = \sum_{\xi \in \mathcal{T}_{sub}} \sum_{\xi' \in \mathcal{T}'_{sub}} [\gamma_{sub} + f(\xi) - f(\xi')]_+, \tag{7}$$

where $[x]_+ \triangleq max(x, 0)$ and γ_{sub} is the margin separating the positive axiom and the negative one. Similarly, for the axioms with $disjointWith$ relation, the margin-based ranking loss is defined as:

$$\mathcal{L}_{dis} = \sum_{\xi \in \mathcal{T}_{dis}} \sum_{\xi' \in \mathcal{T}'_{dis}} [\gamma_{dis} + f(\xi) - f(\xi')]_+. \tag{8}$$

Finally, the overall loss function is defined as linear combinations of these two functions:

$$\mathcal{L} = \mathcal{L}_{sub} + \mathcal{L}_{dis} \tag{9}$$

The goal of training CosE is to minimize the above functions and iteratively update embeddings of concepts.

4 Experiments

To verify the effectiveness of our model, we compare CosE with some well-known KG embedding methods on the task of link prediction, a typical task commonly adopted in knowledge graph embedding. We also design other tasks which are variants of link prediction for transitivity and symmetry of relations in schematic knowledge.

4.1 Datasets

FB15K and WN18 are two benchmark datasets in most previous works, but they are not suitable to evaluate the embedding models for schematic knowledge. Both of them consists of many instances but contain few concepts and axioms. To evaluate CosE, we build a knowledge graph in named YAGO-On from a popular knowledge YAGO which contains a lot of concepts from WordNet and instances from Wikipedia. In our experiments, every subclassOf axiom in YAGO is saved in YAGO-On. Another benchmark dataset is Foundational Model of Anatomy (FMA) which is a real evolving ontology that has been under development at the University of Washington since 1994 [29]. Its objective is to conceptualize the phenotypic structure of the human body in a machine-readable form. It is a real-world, biomedical schematic knowledge and the version used is the OWL files provided by OAEI[4]. As these two datasets only contain subclassOf axioms, so we add disjointWith axioms into them by the simple heuristic rules in [30].

We also evaluate CosE on two new benchmark datasets, named YAGO-on-t and YAGO-on-s, which are two subsets of YAGO-On to test the effects of the link prediction for transitivity and symmetry inference. For any two axioms in YAGO-On, if $(C_i, subclassOf, C_j)$ and $(C_j, subclassOf, C_m)$ exist in YAGO-On, we save an axiom $(C_i, subclassOf, C_m)$ in YAGO-On-t. Similarly, if an axiom $(C_i, disjointWith, C_j)$ exists in YAGO-On, we save the axiom $(C_j, disjointWith, C_i)$ in YAGO-On-s. The statistics of YAGO-On, FMA, YAGO-On-t and YAGO-On-s are listed in Table 1.

4.2 Implementation details.

We employ several state-of-art KG embedding models as baselines, including TransE, TransH, TransR, TransD, ComplEx, Analogy, Rescal and TransC, which are implemented by OpenKE platform [31] and the source codes of methods.

[4] http://oaei.ontologymatching.org/

Table 1: Statistics of original datasets and generated ones

Dataset		YAGO-On	FMA	YAGO-On-t	YAGO-On-s
$\sharp$ Concept		46109	78988	46109	46109
Train	$\sharp$ $subclassOf$	29181	29181	11898	0
	$\sharp$ $disjointWith$	32673	32673	0	10000
Valid	$\sharp$ $subclassOf$	1000	2000	1000	1000
	$\sharp$ $disjointWith$	1000	2000	1000	1000
Test	$\sharp$ $subclassOf$	1000	2000	5949	0
	$\sharp$ $disjointWith$	1000	1000	0	10000

CosE is implemented in Python with the aid of Pytorch and OpenKE. The source code and data are available at https://github.com/zhengxianda/CosE. Mini-batch SGD is utilized on two datasets for training CosE model. For parameters, we use SGD 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 is chosen from the scope of $\{125, 250, 500, 1000\}$, batch size B range of $\{200, 512, 1024, 2048\}$, and fixed margin γ range of $\{3, 6, 9, 12, 18, 24, 30\}$. Both the real and imaginary parts of the concept embeddings are uniformly initialized, and the phases of the relation embeddings are uniformly initialized between 0 and 1. No regularization is used since we find that the fixed margin of γ could prevent our model from over-fitting. The best configuration is determined according to the mean rank in the validation set. The optimal parameters are $\alpha = 0.001$, $k = 200$, $\gamma = 3$ and $B = 200$.

4.3 Linked prediction

Link prediction is a task to complete the axiom (C_i, r, C_j) when C_i, r or C_j is missing. Following the same protocol used in [7], we take *MRR* and *Hits@N* as evaluation protocols. For each test axiom (C_i, r, C_j), we replace the concept C_i or C_j with C_n in concept set $\mathcal{C}$ to generate *corrupted triples* and calculate the score of each triple using the score function. Afterward, by ranking the scores in descending order, the rank of the correct concepts is then derived. *MRR* is the mean reciprocal rank of all correct concepts, and *Hits@N* denotes the proportion of correct concepts or relations ranked in the *top N*. Note that a corrupted triple ranking above a test triple could be valid, which should not be counted as an error. Hence, corrupted triples that already exist in schematic knowledge are filtered before ranking. The filtered version is denoted as "Filter," and the unfiltered version is represented as "Raw." The "Filter" setting is usually preferred. In both settings, a higher *Hits@N* and *MRR* implies the better performance of a model.

For link prediction, all models aim to infer the possible C_i or C_j concept in a testing axiom (C_i, r, C_j) when one of them is missing. The results of concept prediction on YAGO-On and FMA are shown in Table 2. From the table, we can conclude that:

- CosE significantly outperforms the models in term of *Hits@N* and *MRR*. It illustrates that CosE can simultaneously preserve the logical properties by means of two semantic spaces, which are helpful to learn better embeddings for completing schematic knowledge.
- Compared with the project matrices of TransH, TransR and TransD, the projection matrix $\mathbf{M_{C_iC_j}}$ in CosE can measure the confidence of axioms more precisely.

The reason may be that CosE projects axioms with the same relation into several translation-based semantic spaces. As most schematic knowledge only has few relations, so the projection strategy of CosE is more suitable.

Table 2: Experimental results on link prediction

Experiment	YAGO-On					FMA				
Metric	MRR		Hits@N(%)			MRR		Hits@N(%)		
	Raw	Filter	10	3	1	Raw	Filter	10	3	1
TransE	0.241	0.501	0.784	0.582	0.343	**0.066**	0.325	0.474	0.371	0.247
TransR	0.090	0.428	0.588	0.433	0.355	0.060	0.411	0.490	0.440	0.370
TransH	0.195	0.196	0.472	0.252	0.091	0.008	0.009	0.018	0.005	0.003
TransD	0.038	0.176	0.462	0.305	0.000	0.034	0.149	0.430	0.250	0.000
Analogy	0.037	0.301	0.496	0.429	0.160	0.037	0.277	0.487	0.415	0.130
ComplEx	0.034	0.237	0.491	0.403	0.058	0.033	0.201	0.484	0.372	0.011
Rescal	0.080	0.339	0.525	0.392	0.244	0.047	0.317	0.469	0.377	0.236
TransC[5]	0.112	0.420	0.698	0.502	0.298	–	–	–	–	–
CosE	**0.256**	**0.638**	**0.863**	**0.731**	**0.502**	0.053	**0.444**	**0.510**	**0.487**	**0.397**

Table 3 and Table 4 list the results of link prediction on subclassOf axioms and disjointWith axioms, respectively. In most cases, CosE has outperformed all models in terms of *Hits@N* and *MRR* that means these two semantic spaces work well in CosE. For link prediction results on disjointWith axioms, CosE performs a little bit worse than TransR and TransE in *MRR* raw. From further analysis, we find CosE prefer to give a higher score for a correct corrupted triple, so CosE is performing well. Particularly, in disjointWith axioms prediction, *Hits@1* of CosE is increased by 15% and 30% on the two benchmark datasets. It indicates that the angle-based semantic space can preserve symmetry property precisely.

Table 3: Experimental results of link prediction on subclassOf axioms

Experiment	YAGO-On					FMA				
Metric	MRR		Hits@N(%)			MRR		Hits@N(%)		
	Raw	Filter	10	3	1	Raw	Filter	10	3	1
TransE	0.375	0.116	0.722	0.472	0.179	0.113	0.113	0.260	0.110	0.035
TransR	0.063	0.063	0.216	0.020	0.000	0.010	0.010	0.050	0.050	0.050
TransH	0.377	0.724	0.494	0.179	0.179	0.110	0.110	**0.295**	0.080	0.040
TransD	0.011	0.011	0.018	0.008	0.000	0.050	0.050	0.050	0.000	0.000
Analogy	0.003	0.003	0.035	0.003	0.003	0.050	0.050	0.050	0.050	0.050
ComplEx	0.001	0.003	0.002	0.001	0.001	0.003	0.003	0.010	0.000	0.000
Rescal	0.069	0.069	0.143	0.073	0.035	0.009	0.009	0.010	0.005	0.005
CosE	**0.428**	**0.428**	**0.726**	**0.509**	**0.267**	**0.176**	**0.176**	0.290	**0.190**	**0.090**

[5] As experimental results of TransC are much worse than the ones mentioned in the paper [15], so we adopt its original results for comparison.

Table 4: Experimental results of link prediction on disjointWith axioms

Experiment	YAGO-On					FMA				
Metric	MRR		Hits@N(%)			MRR		Hits@N(%)		
	Raw	Filter	10	3	1	Raw	Filter	10	3	1
TransE	0.120	0.627	0.846	0.693	0.507	**0.122**	0.639	0.927	0.741	0.491
TransR	**0.132**	0.792	0.974	0.848	0.710	0.010	0.010	0.050	0.050	0.050
TransH	0.010	0.014	0.220	0.010	0.003	0.005	0.006	0.002	0.001	0.001
TransD	0.066	0.774	0.906	0.621	0.000	0.066	0.292	0.873	0.488	0.000
Analogy	0.074	0.598	0.988	0.854	0.317	0.069	0.557	0.979	0.823	0.264
ComplEx	0.066	0.470	0.970	0.820	0.110	0.003	0.003	0.010	0.000	0.000
Rescal	0.100	0.640	0.920	0.720	0.500	0.094	0.640	0.940	0.750	0.480
CosE	0.097	**0.917**	**0.990**	**0.970**	**0.860**	0.090	**0.870**	**0.990**	**0.950**	**0.780**

4.4 Transitivity and symmetry

In this section, we verify whether the logical properties are implicitly representation by CosE embeddings. To illustrate what kind of information is contained in concept vectors, we design two link prediction experiments on two special datasets. In YAGO-On-t, axioms of the training set are subjected to the rule $(C_i, subclassOf, C_j)$ and $(C_j, subclassOf, C_m)$ and the testing set contains the inferred axioms $(C_i, subclassOf, C_m)$ by applying the transitivity property of $subclassOf$. Thus, we train CosE by training set and use link prediction on the testing set to verify the performance on transitivity. Similarly, we verify the symmetry by YAGO-On-s. If the training set contains $(C_i, disjointWith, C_j)$, the test axiom $(C_j, disjointWith, C_i)$ is saved in the testing set.

As listed in Table 5, CosE is the only model which can achieve good performances on both two datasets. On YAGO-On-t, the *MRR* and *Hits@N* of CosE exceed the ones of other models. On YAGO-On-s, only *MRR* and *Hits@1* of CosE worse than TransE, but their results are very similar. These two experiments indicate that CosE is better than other models for reasoning axioms with transitivity or symmetry.

Table 5: Experimental results on link prediction for transitivity and symmetry

Experiment	YAGO-On-t					YAGO-On-s				
Metric	MRR		Hits@N(%)			MRR		Hits@N(%)		
	Raw	Filter	10	3	1	Raw	Filter	10	3	1
TransE	0.064	0.077	0.142	0.070	0.001	**0.043**	**0.369**	0.971	0.514	**0.080**
TransR	0.012	0.013	0.003	0.002	0.001	0.010	0.010	0.000	0.000	0.000
TransH	0.200	0.238	0.309	0.274	0.149	0.001	0.002	0.000	0.000	0.000
TransD	0.008	0.009	0.020	0.001	0.000	0.001	0.181	0.512	0.302	0.000
Analogy	0.001	0.001	0.001	0.001	0.000	0.043	0.315	0.932	0.538	0.000
ComplEx	0.001	0.001	0.001	0.000	0.000	0.036	0.253	0.743	0.439	0.000
Rescal	0.016	0.020	0.055	0.015	0.004	0.032	0.166	0.449	0.226	0.039
CosE	**0.203**	**0.334**	**0.429**	**0.280**	**0.270**	0.038	0.324	**0.990**	**0.558**	0.000

5 Conclusion and future work

In this paper, we presented a cosine-based embedding method for schematic knowledge called CosE, which could simultaneously preserve the transitivity of $subclassOf$ and the symmetry of $disjointWith$ very well. In order to sufficiently learn the representation of concepts, we defined two score functions based on angle-based semantic space and translation-based semantic space which are tailored for $subclassOf$ axioms and $disjointWith$ axioms. We conducted extensive experiments on link prediction on benchmark datasets. Experimental results indicated that CosE could outperform state-of-the-art methods and successfully preserve the transitivity and symmetry of relations.

As future work, we will explore the following research directions: (1) CosE is a simple model tailored for learning the representation of axioms, but it still has some limits. We will try to find a more expressive model instead of cosine measure to learn the representation of concepts. (2) The embedding of axioms can be applied in various tasks of knowledge graphs. We will merge CosE into these tasks for improving their performances such as noise detection [32] and approximating querying [33].

Acknowledgements

This work was partially supported by the National Key Research and Development Program of China under grant (2017YFB1002801,2018YFC0830200), the Natural Science Foundation of China grant (U1736204), the Fundamental Research Funds for the Central Universities (3209009601).

References

1. George Miller. *WordNet: An electronic lexical database*. MIT press, 1998.
2. Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick Van Kleef, Sören Auer, et al. DBpedia - A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web*, 6(2):167–195, 2015.
3. Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. YAGO: A Large Ontology from Wikipedia and WordNet. *J. Web Sem.*, 6(3):203–217, 2008.
4. Vıctor Gutiérrez-Basulto and Steven Schockaert. From Knowledge Graph Embedding to Ontology Embedding? An Analysis of the Compatibility between Vector Space Representations and Rules. In *KR*, pages 379–388, 2018.
5. Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. Knowledge Graph Embedding: A Survey of Approaches and Applications. *IEEE Trans. Knowl. Data Eng.*, 29(12):2724–2743, 2017.
6. Jason Weston, Antoine Bordes, Oksana Yakhnenko, and Nicolas Usunier. Connecting Language and Knowledge Bases with Embedding Models for Relation Extraction. In *EMNLP*, pages 1366–1371, 2013.
7. Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating Embeddings for Modeling Multi-relational Data. In *NIPS*, pages 2787–2795, 2013.
8. Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge Graph Embedding by Translating on Hyperplanes. In *AAAI*, pages 1112–1119, 2014.
9. Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. Learning Entity and Relation Embeddings for Knowledge Graph Completion. In *AAAI*, pages 2181–2187, 2015.

10. Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. A Three-Way Model for Collective Learning on Multi-Relational Data. In *ICML*, pages 809–816, 2011.
11. Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding Entities and Relations for Learning and Inference in Knowledge Bases. *CoRR*, abs/1412.6575, 2014.
12. Maximilian Nickel, Lorenzo Rosasco, Tomaso A Poggio, et al. Holographic Embeddings of Knowledge Graphs. In *AAAI*, pages 1955–1961, 2016.
13. Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex Embeddings for Simple Link Prediction. In *ICML*, pages 2071–2080, 2016.
14. Gonzalo I Diaz, Achille Fokoue, and Mohammad Sadoghi. EmbedS: Scalable, Ontology-aware Graph Embeddings. In *EBDT*, pages 433–436, 2018.
15. Xin Lv, Lei Hou, Juanzi Li, and Zhiyuan Liu. Differentiating Concepts and Instances for Knowledge Graph Embedding. In *EMNLP*, pages 1971–1979, 2018.
16. Xuefeng Fu, Guilin Qi, Yong Zhang, and Zhangquan Zhou. Graph-based approaches to debugging and revision of terminologies in DL-Lite. *Knowl.-Based Syst.*, 100:1–12, 2016.
17. Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. Reasoning With Neural Tensor Networks for Knowledge Base Completion. In *NIPS*, pages 926–934, 2013.
18. Antoine Bordes, Jason Weston, Ronan Collobert, and Yoshua Bengio. Learning Structured Embeddings of Knowledge Bases. In *AAAI*, pages 301–306, 2011.
19. Han Xiao, Minlie Huang, Yu Hao, and Xiaoyan Zhu. TransA: An Adaptive Approach for Knowledge Graph Embedding. *CoRR*, abs/1509.05490, 2015.
20. Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. Knowledge Graph Embedding via Dynamic Mapping Matrix. In *ACL*, pages 687–696, 2015.
21. Xin Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *SIGKDD*, pages 601–610, 2014.
22. Quan Liu, Hui Jiang, Andrew Evdokimov, Zhen-Hua Ling, Xiaodan Zhu, Si Wei, and Yu Hu. Probabilistic Reasoning via Deep Learning: Neural Association Models. *CoRR*, abs/1603.07704, 2016.
23. Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. Modeling Relational Data with Graph Convolutional Networks. In *ESWC*, pages 593–607, 2018.
24. Baoxu Shi and Tim Weninger. ProjE: Embedding Projection for Knowledge Graph Completion. In *AAAI*, pages 1236–1242, 2017.
25. Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2D Knowledge Graph Embeddings. In *AAAI*, pages 1811–1818, 2018.
26. Muhao Chen, Yingtao Tian, Xuelu Chen, Zijun Xue, and Carlo Zaniolo. On2Vec: Embedding-based Relation Prediction for Ontology Population. In *SIAM*, pages 315–323, 2018.
27. Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. Jointly embedding knowledge graphs and logical rules. In *EMNLP*, pages 192–202, 2016.
28. Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. Knowledge graph embedding with iterative guidance from soft rules. In *AAAI*, pages 4816–4823, 2018.
29. Natalya F. Noy, Mark A. Musen, José L. V. Mejino Jr, and Cornelius Rosse. Pushing the envelope: challenges in a frame-based representation of human anatomy. *Data Knowl. Eng.*, 48(3):335–359, 2004.
30. Huan Gao, Guilin Qi, and Qiu Ji. Schema induction from incomplete semantic data. *Intell. Data Anal.*, 22(6):1337–1353, 2018.
31. Xu Han, Shulin Cao, Xin Lv, Yankai Lin, Zhiyuan Liu, Maosong Sun, and Juanzi Li. OpenKE: An Open Toolkit for Knowledge Embedding. In *EMNLP*, pages 139–144, 2018.
32. Ruobing Xie, Zhiyuan Liu, Fen Lin, and Leyu Lin. Does William Shakespeare really write Hamlet? knowledge representation learning with confidence. In *AAAI*, pages 4954–4961, 2018.
33. Meng Wang, Ruijie Wang, Jun Liu, Yihe Chen, Lei Zhang, and Guilin Qi. Towards Empty Answers in SPARQL: Approximating Querying with RDF Embedding. In *ISWC*, pages 513–529, 2018.