Enhancing Knowledge Graph Embedding with Relational Constraints

1st Mingda Li[1,2] limingda2018@ia.ac.cn

2nd Zhengya Sun[1,2] zhengya.sun@ia.ac.cn

3rd Siheng Zhang[1,2] zhangsiheng2015@ia.ac.cn

4th Wensheng Zhang[1,2] zhangwenshengia@hotmail.com

[1]*Research Center of Precision Sensing and Control, Institute of Automation, Chinese Academy of Sciences, Beijing, China*
[2]*University of Chinese Academy of Sciences, Beijing, China*

***Abstract*—Knowledge graph embedding is studied to embed entities and relations of a knowledge graph into continuous vector spaces, which benefits a variety of real-world applications. Among existing solutions, translation-based models, which employ geometric translation to design score function, have drawn much attention. However, these models primarily concentrate on evidence from observing whether the triplets are plausible, and ignore the fact that the relation also implies certain semantic constraints on its subject or object entity. In this paper, we present a general framework for enhancing knowledge graph embedding with relational constraints (KRC). Specifically, we elaborately design the score function by encoding regularities between a relation and its arguments into the translation-based embedding space. Additionally, we propose a soft margin-based ranking loss for effectively training the KRC model, which characterizes different semantic distances between negative and positive triplets. Furthermore, we combine regularities with distributional representations to predict the missing triplets, which possesses certain robust guarantee. We evaluate our method on the task of knowledge graph completion. Extensive experiments show that KRC achieves substantial improvements against baselines.**

***Keywords*-knowledge graph embedding, relational constraints, knowledge graph completion**

I. Introduction

As a collection of human knowledge, knowledge graph (KG) can represent information about real-world entities and their relations. A large number of knowledge graphs, such as Freebase [1], DBpedia [2] and YAGO [3], have been created and recently available. These knowledge graphs usually contain billions of vertices, multi-typed edges and triplets. However, traditional knowledge representation methods are logic and symbolic [4] and not easy to model such a large-scale graph. Thus, knowledge graph embedding (KGE) is studied to embed the entities and relations of a knowledge graph into low-dimensional vector spaces, which benefits various real-world applications such as machine translation [5], question answering [6] and recommendation [7].

Specifically, KGE represents a symbolic triplet (h, r, t) as continuous vectors $(\mathbf{h}, \mathbf{r}, \mathbf{t})$, each of which corresponds to head entity, relation and tail entity, respectively. The framework of KGE can be abstracted as follows. First, each embedding method proposes a score function of triplet (h, r, t), such as $f_r(h,t) = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_p$ for TransE [8], where $p \geq 1$ denotes l_p norm. Then, based on the score function, they formulate and minimize a loss function. One of the mainly used loss functions is margin-based ranking loss [9] on each triplet (h, r, t) as $\max(0, \gamma + f_r(h,t) - f'_r(h', t'))$, where (h', r', t') is a negative triplet corrupting the subject or object of positive triplet (h, r, t), and γ is the margin separating them. Last, the representations are obtained and used in the downstream tasks. To summarize, the main differences between KGE methods are the principles of the design of score functions.

Among these methods, the translation-based methods, which are simple and efficient, have drawn much attention. They obtain entity and relation embeddings by regarding a relation as a translation from a head entity to a tail one in the same embedding space (formally as $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$). Then the score function is designed by measuring the error of geometric translation. Recently, there exist many variants in this branch. Different methods employ different embedding spaces to project the entities, such as hyperplane [10], static mapping matrix [11] and dynamic mapping matrix [12].

Though achieving outstanding performance, translation-based methods primarily concentrate on entity-relation-entity triplets [13], and ignore the fact that the relation also implies certain semantic constraints on its subject or object entity. For example, given a golden triplet (e_i, r_k, e_j), if the relation r_k is *president_of*, we can infer that the entity e_i is probably a politician, and e_j is a country. This motivates us to propose a method for improving translation-based models by capturing the regularities between the relation and its arguments.

Additionally, based on the score function, we propose a novel soft margin-based ranking loss via the local closed-world assumption (LCWA) [14], which can be applied to approximate domain and range constraints for a certain relation solely based on observed triplets in the graph. With the soft margin scheme, we can characterize different semantic distances between negative and positive triplets. Referring to Fig.1, the negative triplet *(Beckham, place_of_birth, Honolulu)* is closer to the particular positive one *(Obama, place_of_birth, Honolulu)* in semantic space than *(Britain, place_of_birth, Honolulu)*. Because *Beckham* agrees with the domain constraints of *place_of_birth*, while *Britain* dosen't. When predicting the missing triplets, we combine the regularities with distributional representations

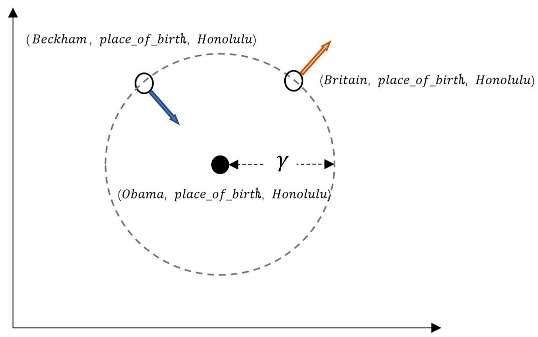

Figure 1. An example for the soft margin that separates positive and negative triplets. *(Obama, place_of_birth, Honolulu)* is a positive triplet, while *(Beckham, place_of_birth, Honolulu)* and *(Britain, place_of_birth, Honolulu)* are both negative triplets

to make KRC have more predictive power.

To summarize, KRC can be seen as an extension module for translation-based models. And our contributions are three folds:

- We propose KRC, a universal approach to enhance knowledge graph embedding with relational constraints. Our method can improve the translation-based models by capturing the regularities between the relation and its arguments.
- We introduce a novel soft margin-based ranking loss to reflect different semantic distances between negative triplets and positive ones; When predicting the missing triplets, we combine regularities with distributional representations, which possesses certain robust guarantee.
- We evaluate our method on the task of knowledge graph completion on two benchmark datasets. Experimental results illustrate that our method can significantly improve the performance of translation-based models.

This paper is organized as follows. In Section II, we survey and categorize the related researches. In Section III, we present our method for enhancing knowledge graph embedding. In Section IV, we conduct experiments for testing our method. Finally, we conclude our paper in Section V.

II. Related work

A. Translation-based Model

The first translation-based model is TransE [8], which translates the head entity to the tail one by the relation vector, formally as $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$. TransE gains attention because of its effectiveness and simplicity, while it has some issues for modelling 1-N, N-1, and N-N relations. Thus, many variants of TransE are proposed, and they transform entities into different subspaces. TransH [10] is proposed to project the relation-specific head and tail entities into the same hyperplane and TransR [11] rotates the entity space with a relation-specific matrix. TransD [12] follows the work TransR and it uses an additional vector of entity/relation to construct the mapping matrix dynamically.

B. Bilinear Model

RESCAL [15] is proposed as a pioneering work of the bilinear model, where each entity is associated with a vector to capture its latent features. And each relation is represented as matrix $\mathbf{M_r}$, which models pairwise interactions between latent features. The score of a triplet (h, r, t) is defined by a bilinear function, formally as $\mathbf{h^T M_r t}$. Extensions of RESCAL have been proposed by restricting the bilinear functions. For example, DisMult [16] and ComplEx [17] restrict the relational matrix to diagonal matrix. HolE [18] introduces circular correlation to combine the expressive power of RESCAL with the efficiency and simplicity of DistMult.

C. Neural Network-based Model

Neural network-based models have layers and an activation function like a neural network. The Neural Tensor Network (NTN) [19] has a standard linear neural network structure and a bilinear tensor structure, and can be considered as a generalization of RESCAL, where the weight of the network is trained for each relation. MLP [20] and NAM [21] can be seen as variants of NTN. ConvE [22] exploits the convolution operator to capture the interactive information between latent features.

D. KGE Model with Type Information

Though above KGE models have yielded great results, they suffer from the fact that entity-relation-entity triplets are less informative, which limits their performance. Recently, some KGE methods are studied to exploit the semantic information of entity type. Different methods interpret the entity type as different embedding spaces which are appiled to modify the latent representations of entities, such as mapping matrix [23], n-sphere [24], hyperplane [13]. However, type information can also suffer from incompleteness and inconsistency present in the data [14]. Therefore, we study the fact that the relation can imply certain semantic information of its argument entities and propose a KRC model to improve the performance of translation-based models.

III. Methodology

In this section, we first introduce the local closed-world assumption (LCWA)[14]. According to the LCWA, we extract four triple features for each triplet. Afterwards, we present the KRC model for enhancing knowledge graph embedding with relational constraints, and a novel soft margin-based ranking loss to effectively train the model. We also provide a scheme to combine regularities with distributional representations. Fig.2 shows the whole framework of our method.

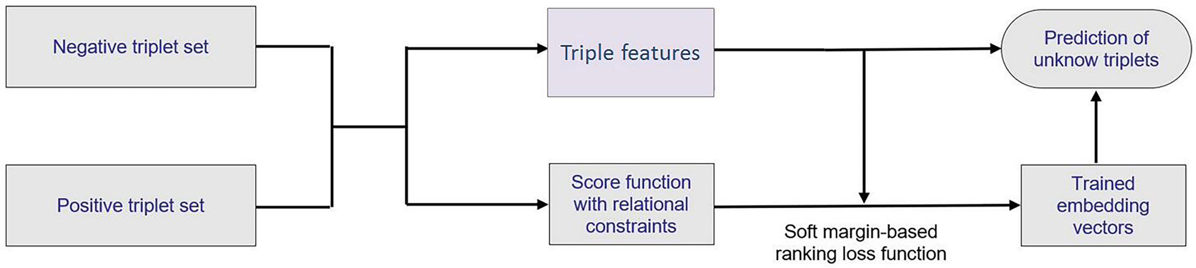

Figure 2. The overall architecture of our proposed framework. Firstly, we obtain the triple features and design the score function via negative and positive triplets. Then the triple features will be applied to train the model and filter out obviously false triplets.

A. Triple Features under the LCWA

Entity and relation types might miss complete typing, leading to fuzzy type-constraints in the knowledge graph. In these cases a local closed-world assumption (LCWA) can be applied, which approximates the domain and range constraints of the target relation not on class level, but on instance level based solely on observed triplets [14]. The idea is shown in Fig.3. Given all observed triplets, under this LCWA, the domain of a relation r consists of all entities that are related by the relation r as subject. The range is accordingly defined, but contains all the entities related as object by relation r. According to LCWA, we extract the following two features 1 and 2 to indicate whether an entity e agrees with the domain and range constraints of the target relation r.

$$\mathbf{I_1}(e,r) = \begin{cases} 1, & (e,r,e') \in TrainingSet \\ 0, & otherwise \end{cases} \tag{1}$$

$$\mathbf{I_2}(e,r) = \begin{cases} 1, & (e',r,e) \in TrainingSet \\ 0, & otherwise \end{cases} \tag{2}$$

In the above equations, TrainingSet is the set of golden triplets. We use e' to denote any possible entity occurring in the knowledge graph. However, this LCWA is so restrictive that it may exclude some entities from the domain or range constraints that agree with the type-constraints. Therefore, we relax the LCWA, and the idea is shown in Fig.4. According to relaxed LCWA, if there is a connection between relation r_1 and r_2, the domain or range constraints of these two relations are the same. Thus, we extract another two features 3 and 4 to indicate whether an entity e agrees with the domain and range constraints of the target relation r under the relaxed LCWA.

$$\mathbf{I_3}(e,r) = (\mathbf{I_1}(e,r')\wedge\mathbf{S_1}(r',r))\vee(\mathbf{I_2}(e,r')\wedge\mathbf{S_2}(r',r)) \tag{3}$$

$$\mathbf{I_4}(e,r) = (\mathbf{I_1}(e,r')\wedge\mathbf{S_3}(r',r))\vee(\mathbf{I_2}(e,r')\wedge\mathbf{S_4}(r',r)) \tag{4}$$

where

$$\mathbf{S_1}(r_1,r_2) = \mathbf{I_1}(e',r_1) \wedge \mathbf{I_1}(e',r_2) \tag{5}$$

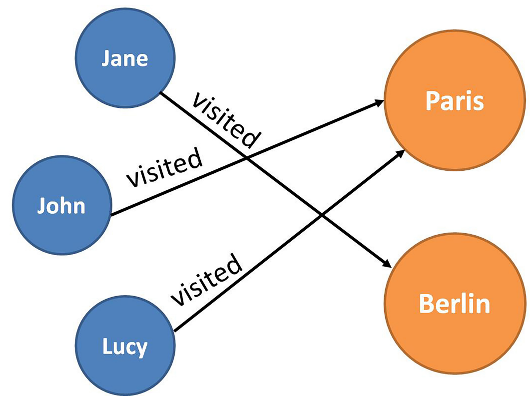

Figure 3. llustration of how the local closed-world assumption is applied. All left entities *(John, Jane and Lucy)* are assigned to the domain of relation $visited$, because the graph contains triplets where these entities occur as subject. For the range only the cities Paris and Berlin are added and the other entities are excluded from the range, because no relation $visited$ has been observed in the graph where they occur as object.

$$\mathbf{S_2}(r_1,r_2) = \mathbf{I_2}(e',r_1) \wedge \mathbf{I_1}(e',r_2) \tag{6}$$

$$\mathbf{S_3}(r_1,r_2) = \mathbf{I_1}(e',r_1) \wedge \mathbf{I_2}(e',r_2) \tag{7}$$

$$\mathbf{S_4}(r_1,r_2) = \mathbf{I_2}(e',r_1) \wedge \mathbf{I_2}(e',r_2) \tag{8}$$

In the above equations, S_1-S_4 are designed to describe the head-to-head(5), tail-to-head(6), head-to-tail(7), tail-to-tail(8) connection between relation r_1 and r_2. We use e' and r' to denote any possible entity and relation occurring in the knowledge graph. Thus, for each triplet (h,r,t), we can extract four triple features $\mathbf{I_1}(h,r)$, $\mathbf{I_2}(t,r)$, $\mathbf{I_3}(h,r)$, $\mathbf{I_4}(t,r)$. Intuitively, the more triple features a candidate triplet matches with, the more likely it is a positive one.

B. The KRC Model

Given a triplet (h,r,t), the score function of translation-based model can be described as follows:

$$f_r(h,t) = \|\mathbf{h}^* + \mathbf{r} - \mathbf{t}^*\|_p \tag{9}$$

where $\mathbf{h}^*, \mathbf{t}^* \in \mathbb{R}^d$ are obtained by employing different embedding spaces to project $\mathbf{h}$ and $\mathbf{t}$, which are the representations of head and tail entity. The score function favours

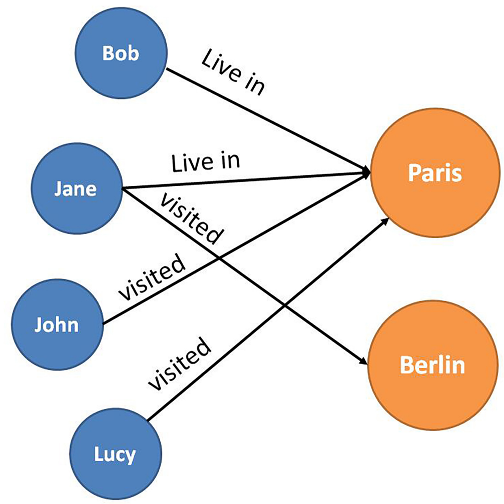

Figure 4. Illustration of how the relaxed local closed-world assumption is applied. Because there is a head-to-head connection between relation *visited* and *live in*, the domain of these two relations are the same. Then the Bob is assigned to the domain of relation *visited*, even though no relation *visited* has been observed in the graph where Bob occurs as subject.

a lower value of the energy for a positive triplet than for a negative one.

As described in introduction, a relation can imply certain semantic constraints on its subject or object entity. That is, the relation r can influence the latent representations of its subject entity $\mathbf{h}^*$, and object entity $\mathbf{t}^*$. To capture the influence, we elaborately design the relational constraints, which are imposed on the score function of translation-based model. In this paper, we use standardized Euclidean distance (SED) to define relational constraints. Because SED has a low time complexity and can be applied in large-scale knowledge representation, which is defined as:

$$\mathbf{X}^* = \frac{\mathbf{X} - \mu}{\sigma} \tag{10}$$

Here $\mathbf{X}$ and $\mathbf{X}^*$ represent the feature vector before and after standardization, respectively. And μ and σ are expectation and standard deviation vectors. As we know, different relations can imply different semantic constraints on its subject and object entity. Thus, μ and σ are variables relevant with r. We introduce additional mapping vectors $\mathbf{w}_{rs}, \mathbf{w}_{ro} \in \mathbb{R}^d$ and bias vectors $\mathbf{b}_{rs}, \mathbf{b}_{ro} \in \mathbb{R}^d$ along with the relation representation $\mathbf{r} \in \mathbb{R}^d$. Then the constraints of $\mathbf{h}^*$ and $\mathbf{t}^*$ are:

$$C_s(\mathbf{h}^*) = \left\| \frac{\mathbf{h}^* - \mu_s(r)}{\sigma_s(r)} \right\|_2 = \left\| \mathbf{w}_{rs}^\top (\mathbf{h}^* - \mathbf{b}_{rs}) \right\|_2 \tag{11}$$

$$C_o(\mathbf{t}^*) = \left\| \frac{\mathbf{t}^* - \mu_o(r)}{\sigma_o(r)} \right\|_2 = \left\| \mathbf{w}_{ro}^\top (\mathbf{t}^* - \mathbf{b}_{ro}) \right\|_2 \tag{12}$$

With the relational constraints of subject entity $C_s(\mathbf{h}^*)$ and object entity $C_o(\mathbf{t}^*)$, we modify the traditional score function (Eq.(9)) as follows:

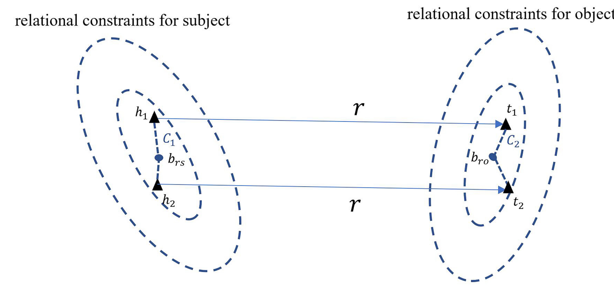

Figure 5. Entity and relation space of KRC, where (h_1, r, t_1) and (h_2, r, t_2) are candidate triplets

$$f_{krc}(h, r, t) = \left\| \mathbf{h}^* + \mathbf{r} - \mathbf{t}^* \right\|_p + \lambda_1 C_s(\mathbf{h}^*) + \lambda_2 C_o(\mathbf{t}^*) \tag{13}$$

where hyper-parameters $\lambda_1, \lambda_2 > 0$ are used to trade off the weight of relational constraints in the subject and object position. The score function indicates that the embeddings $\mathbf{h}^*, \mathbf{t}^*$ should not only be compatible with translation constraints (formally as $\mathbf{h}^* + \mathbf{r} \approx \mathbf{t}^*$), but also get close to $\mathbf{b}_{rs}, \mathbf{b}_{ro}$ respectively based on SED metric, as shown in Fig.5. And in experiments, we can extend relational constraints to $p-$norm to keep in the same scale with $f_r(h, t)$, and impose regularization as $\left\| \mathbf{w}_{rs}^\top \mathbf{h}^* \right\|_2 \leq 1$, $\left\| \mathbf{w}_{rs}^\top \mathbf{b}_{rs} \right\|_2 \leq 1$, $\left\| \mathbf{w}_{ro}^\top \mathbf{t}^* \right\|_2 \leq 1$ and $\left\| \mathbf{w}_{ro}^\top \mathbf{b}_{ro} \right\|_2 \leq 1$.

C. Soft Margin-based Ranking Loss

To effectively train translation-based model, margin-based ranking loss is usually employed. Here the margin preserves certain preferences between positive and negative triplets. As the negative triplets don't have equal semantic distance to the particular positive one. We introduce semantic weight into the margin and design a novel soft margin-based loss. Thus, we reform the embedding learning problem as that of minimizing the following loss function:

$$L = \sum_{\xi \in \Delta} \sum_{\xi' \in \Delta'} [f_{krc}(\xi) - f_{krc}(\xi') + \gamma(1 + \nu(\xi, \xi'))]_+ \tag{14}$$

where Δ, Δ' denote the positive and negative triplet set, $[x]_+ = \mathbf{max}(0, x)$, γ is the margin value, and the weight function $\nu(\xi, \xi')$ quantifies the semantic distance between any triplet pairs. Intuitively, the large the semantic distance between positive and negative triplets, the large the margin is. Formally, we define the weight function according to the predefined triple features which have different influence on the semantic distance. To distinguish their contributions to the weight function, we let:

$$\nu(\xi_1, \xi_2) = \delta_1 \cdot [\mathbf{m}_{z_1}(\xi_1, \xi_2) - \frac{|Z_1|}{2}] + \delta_2 \cdot [\mathbf{m}_{z_2}(\xi_1, \xi_2) - \frac{|Z_2|}{2}] \tag{15}$$

where $\delta_1, \delta_2 > 0$ are weight parameters chosen through validation set, $\mathbf{m}_z(\xi_1, \xi_2)$ denotes the manhattan distance between ξ_1 and ξ_2 in the triple feature space z, and $|Z|$ is the cardinality of set Z. Z_1 is a feature space comprising feature 1 and 2, while Z_2 is a space comprising feature 3 and 4.

Following the soft margin-based ranking loss function, the more similar triplet pairs are in the feature space, the closer they are in the semantic space. After the training process, we can obtain the embedding of entities and relations in the knowledge graph, which have stronger semantics and can benefit the downstream tasks.

D. Prediction with Triple Features

When predicting the missing triplets, previous studies compute scores of the candidate triplets only by their score function, which limits the performance. As described above, triple features 3 and 4 indicate whether the triplet agrees with the domain and range constraints of the target relation, which can be used to filter out some obviously false candidates. Thus, we apply these features to amend the output of our score function, and the score of a candidate triplet (h, r, t) can be calculated as follows:

$$Score_{(h,r,t)} = f_{krc}(h, r, t) + \gamma[2 - \mathbf{I}_3(h, r) - \mathbf{I}_4(t, r)] \quad (16)$$

where γ is the margin value used in the loss function. According to the equation, if the triple feature 3 or 4 of a candidate triplet is zero, we will add the margin value to its score computed by score function, which indicates it is perhaps a false triplet.

E. Model Complexity

Table I lists the complexity of the baselines and our proposed framework, including the number of parameters and the number of multiplication operations in an epoch. N_e and N_r represent the number of entities and relations, respectively. N_t represents the number of triplets in a knowledge graph. m is the dimension of entity embedding space, and n is that of relation embedding space. In some models, $m = n$. As described above, we should preserve the triple features 1-4 before the process of model training, and hence cost $2N_eN_r$ parameters. Additionally, the relational constraints of score function need $4N_rn$ parameters. Thus, KRC will totally cost $2N_eN_r + 4N_rn$ extra parameters against its base model. Given that $N_r \ll N_e$ and $n \ll N_e$, our method will not significantly increase the model complexity.

Table I
COMPARISON OF MODEL AND TIME COMPLEXITY

Model	#Parameters	#Operations
TransE	$O(N_em + N_rn) \quad (m = n)$	$O(nN_t)$
KRC(TransE)	$O(N_em + 5N_rn + 2N_eN_r)$	$O(2nN_t)$
TransH	$O(N_em + 2N_rn) \quad (m = n)$	$O(2nN_t)$
KRC(TransH)	$O(N_em + 6N_rn + 2N_eN_r)$	$O(4nN_t)$
TransD	$O(2N_em + 2N_rn) \quad (m = n)$	$O(2nN_t)$
KRC(TransD)	$O(2N_em + 6N_rn + 2N_eN_r)$	$O(4nN_t)$

IV. EXPERIMENTS

In this section, we evaluate our method on the task of knowledge graph completion. Afterwards, we carry out semantic analysis and a case study to demonstrate the effectiveness of relational constraints.

A. Experimental Settings

Datasets. Our experiments are conducted on two public benchmark datasets FB15K [8], and FB15K-237 [25]. These datasets have been widely used in previous studies for evaluating model performance. FB15K is extracted from Freebase. As discussed by [25], the dataset has redundancy in the form of reverse relations. For this reason, FB15K-237 is created by removing inverse relations from FB15K. The statistics of these datasets are list in Table II.

Table II
STATISTICS OF DATASETS USED IN THE EXPERIMENTS

Dataset	#Rel	#Ent	#Train	#Valid	#Test
FB15K	1345	14951	483142	50000	59071
FB15K-237	237	14541	272115	17535	20466

Implementation. We implement all baseline methods for comparison, and directly reproduce the results with the reported optimal parameters. KRC can be seen as an extension of the translation-based model. Thus, to perform a thorough comparison with the baselines, we use and implement TransE, TransH, TransD (TransR can be seen as a variant of TransD) as our base models. In this paper, we take l_1 norm and fine-tune the hyper-parameters on the validation dataset. The ranges of the hyper-parameters for the grid search are set as follows: embedding dimension $d \in$ {100, 200, 300}, margin value $\gamma \in$ {1.0, 2.0, 3.0, 4.0}, the additional hyper-parameters of relational constraints $\lambda_1, \lambda_2 \in$ {0.1, 0.2, 0.3}, $\delta_1 \in$ {0.05, 0.1, 0.15}, and $\delta_2 \in$ {0.01, 0.02, 0.03}. The optimal settings of KRC are the following: (1) On FB15K, embedding dimension $d = 200$, initial learning rate $\alpha = 0.0005$, margin $\gamma = 2.0$, which are the same as the base translation-based models, and the additional hyper-parameters of relational constraints $\lambda_1, \lambda_2 = 0.1$, soft margin $\delta_1 = 0.05, \delta_2 = 0.01$. (2) On FB15K-237, embedding dimension $d = 100$, initial learning rate $\alpha = 0.0005$, margin $\gamma = 3.0$, hyper-parameters of relational constraints $\lambda_1, \lambda_2 = 0.2$, soft margin $\delta_1 = 0.1, \delta_2 = 0.02$. We train our model until convergence but stop at most 3000 rounds. Each experiment of KRC is repeated for 3 times. We report the average performance in the following results.

B. Knowledge Graph Completion

Knowledge graph completion aims to predict the missing triplets in KGs under the supervision of existing knowledge graph, which is the most important benchmark task for knowledge graph embedding.

Evaluation protocol. For each test triplet (h, r, t), the head entity h (or the tail t) is removed and replaced by each of all other entities to create a set of corrupted triplets. We use "Filtered" setting protocol, which will not take any corrupted triplets that appear in the KG into account. Then, the scores of those corrupted triplets are computed. We rank the valid test triplet and corrupted triplets in ascending order of their scores. The following evaluation metrics are employed: mean rank(MR), mean reciprocal rank(MRR) and the proportion of the valid test triples ranking in top 10 predictions (Hits@10). Lower MR, higher MRR, higher Hits@10 indicate better performance.

Comparison With Translation-based Models: Evaluation results are reported in Tables III and IV, noting that +R, +S, +T refers to utilizing the relational constraints, soft margin-ranking loss and triple features for prediction. We could observe that:

Table III
COMPARISON WITH TRANSLATION-BASED MODELS ON FB15K

FB15K	**MRR**	**MR**	**HITS@10**
TransE	0.622	92	83.8
TransE+R	0.665	96	85.0
TransE+R+S	0.675	91	84.8
TransE+R+S+T	0.682	82	84.9
TransH	0.629	85	83.8
TransH+R	0.668	86	84.7
TransH+R+S	0.676	82	84.6
TransH+R+S+T	0.682	77	84.8
TransD	0.627	90	84.1
TransD+R	0.672	92	84.9
TransD+R+S	0.675	90	84.7
TransD+R+S+T	0.683	83	84.9

Table IV
COMPARISON WITH TRANSLATION-BASED MODELS ON FB15K237

FB15K237	**MRR**	**MR**	**HITS@10**
TransE	0.282	255	46.9
TransE+R	0.304	298	48.6
TransE+R+S	0.307	293	49.0
TransE+R+S+T	0.309	263	49.1
TransH	0.293	254	48.0
TransH+R	0.306	280	49.2
TransH+R+S	0.312	268	49.4
TransH+R+S+T	0.313	239	49.5
TransD	0.291	254	47.6
TransD+R	0.309	270	49.0
TransD+R+S	0.311	265	49.4
TransD+R+S+T	0.313	245	49.6

- From the results of knowledge graph completion, we justify that each module of our method is efficient. Compared with the base model TransD, KRC achieves 10% relative improvement on MRR on FB15k and 8% on FB15K-237.
- Utilizing the relational constraints (+R) works generally better than the soft margin-based loss (+S). Possibly because most of negative triplets' triple features defined in Section III are zero. And in that case, soft margin-based ranking loss degrades into normal margin-based ranking loss.
- Additionally, predicting the missing triplets with triple features (+T) can achieve a consistent improvement.

Comparison With other KGE Models: We compare KRC(TransD) with other KGE models described in Section II, which always leads to the best results on MRR. Evaluation results are reported in Tables V and VI, and we could observe that:

Table V
COMPARISON WITH OTHER KGE MODELS ON FB15K

FB15K	**MRR**	**MR**	**HITS@10**
ComplEx	0.634	89	83.5
HolE	0.468	160	70.6
ConvE	0.679	**61**	84.1
KRC(TransD)	**0.683**	83	**84.9**

Table VI
COMPARISON WITH OTHER KGE MODELS ON FB15K237

FB15K237	**MRR**	**MR**	**HITS@10**
ComplEx	0.276	401	43.3
HolE	0.264	342	43.5
ConvE	0.310	277	49.5
KRC(TransD)	**0.313**	**245**	**49.6**

- KRC outperforms the baselines substantially on almost all metrics, which demonstrates the effectiveness of our model and the correctness of our intuition analysis.
- The comparison illustrates our method improves translation-based models to be competitive in the field of knowledge graph completion.

C. Semantic Analysis: Statistic Justification

As discussed above, the relation can imply certain semantic information of its argument entities, and KRC model can obtain better embedding results with the relational constraints. To justify this statement, we conduct the entity classification experiment that aims to predict the types of entities on FB15K-237. For a fair comparison, our front-end classifier is identically the Logistic Regression in a one-versus-rest setting for multi-label classification. The evaluation applies Macro-F1 that is commonly used in multi-label classification. Type@N means the task is involved with N types to be predicted. As Table VII shows, KRC outperforms its base models consistently, demonstrating the effectiveness of our model. Additionally, Entity types represent some level of semantics, thus the better results illustrate that our method is indeed more semantics specific.

Table VII
EVALUATION RESULTS OF ENTITY CLASSIFICATION ON FB15K237

FB15K237	T@25	T@50	T@75
TransE	79.3	71.1	66.7
KRC(TransE)	**80.6**	**72.4**	**67.7**
TransH	80.2	71.7	66.7
KRC(TransH)	**80.6**	**72.7**	**68.3**
TransD	79.2	70.8	66.8
KRC(TransD)	**80.3**	**72.6**	**67.9**

D. Relational Constraints Analysis: Case Study

Utilizing relational constraints is beneficial for correctly classifying the triplets that could not be discriminated by traditional triple-level learning methods. To prove this, we conduct a case study about link prediction on FB15K-237. The results are listed in Table VIII. Specifically, *(Hungarian kingdom, capital, Budapest)* is a golden triplet in the test set. We replace the object *Budapest* with other entities and obtain the remaining negative triplets in the table. From this table, we can observe that the relational constraints C_s, C_o can capture a bias of the entity occurring in subject or object position for the specific relation *capital*. For example, according to the score of C_o, *Washington, D.C.* matches with the relation in the object position, while *Maleness* dosen't. Additionally, by virtue of relational constraints, we can effectively classify confusing triplets such as *(Hungarian kingdom, capital, Moldavia)*, where *Moldavia* is a long-tail entity and has only one instance in the training set. Thus, its embedding obtained by triple-level learning model, such as TransE, has low expressiveness, and can't be discriminated correctly. The experiment also demonstrates the effectiveness of our model and indicates the potential usage of the relational constraints.

V. CONCLUSION

In this paper, we propose a method KRC that improves translation-based models with relational constraints. Then triple features are applied to effectively train KRC model and obtain a more predictive power. We also evaluate our method with extensive studies. Experimental results show that our method achieves consistent and significant improvements compared with the state-of-the-art baselines.

Currently, our method is based on the translation-based models. As described in related work, bilinear and neural network-based models also concentrate on entity-relation-entity triplets and ignore the reltional constraints. In the future, we will extend KRC methods to improve the performance of these models.

ACKNOWLEDGMENT

The authors are thankful for the financial support from the National Natural Science Foundation of China (U1636220 and 61876183).

REFERENCES

[1] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, J. Taylor, "Freebase: a collaboratively created graph database for structuring human knowledge," in: Proceedings of the 2008 ACM SIGMOD international conference on Management of data, ACM, 2008, pp. 1247–1250.

[2] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. Van Kleef, S. Auer, et al., "Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia," Semantic Web 6 (2015) 167–195.

[3] F. M. Suchanek, G. Kasneci, G. Weikum, "Yago: a core of semantic knowledge," in: Proceedings of the 16th international conference on World Wide Web, ACM, 2007, pp. 697–706.

[4] J. F. Sowa, "Knowledge representation: logical, philosophical, and computational foundations," PWS, 2000.

[5] A. Kazemi, A. Toral, A. Way, A. Monadjemi, M. Nematbakhsh, "Syntax-and semantic-based reordering in hierarchical phrase-based statistical machine translation," Expert systems with applications 84 (2017) 186–199.

[6] Y. Zhang, H. Dai, Z. Kozareva, A. J. Smola, L. Song, "Variational reasoning for question answering with knowledge graph," in: Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[7] H. Wang, Z. Wang, S. Hu, X. Xu, S. Chen, Z. Tu, "Duskg: A finegrained knowledge graph for effective personalized service recommendation," Future Generation Computer Systems 100 (2019) 600–617.

[8] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in: Advances in neural information processing systems, 2013, pp. 2787–2795.

[9] M. Nickel, K. Murphy, V. Tresp, E. Gabrilovich, "A review of relational machine learning for knowledge graphs," Proceedings of the IEEE 104 (2015) 11–33.

[10] Z. Wang, J. Zhang, J. Feng, Z. Chen, "Knowledge graph embedding by translating on hyperplanes," in: Twenty-Eighth AAAI conference on artificial intelligence, 2014.

[11] Y. Lin, Z. Liu, M. Sun, Y. Liu, X. Zhu, "Learning entity and relation embeddings for knowledge graph completion," in: Twenty-ninth AAAI conference on artificial intelligence, 2015.

Table VIII
THIS IS A CASE STUDY. *(Hungarian kingdom, capital, Budapest)* IS A GOLDEN TRIPLET, AND THE REMAINING ARE NEGATIVE TRIPLETS

Triplet	KRC(TransE)			TransE
	$\|\mathbf{h}+\mathbf{r}-\mathbf{t}\|_1$	C_s	C_o	$\|\mathbf{h}+\mathbf{r}-\mathbf{t}\|_1$
(Hungarian kingdom, capital, Budapest)	0.106	0.074	0.059	0.107
(Hungarian kingdom, capital, Washington, D.C.)	0.120	0.074	0.050	0.123
(Hungarian kingdom, capital, Maleness)	0.142	0.074	0.095	0.130
(Hungarian kingdom, capital, Moldavia)	0.100	0.074	0.092	0.099

[12] G. Ji, S. He, L. Xu, K. Liu, J. Zhao, "Knowledge graph embedding via dynamic mapping matrix," in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2015, pp. 687–696.

[13] N. Guan, D. Song, L. Liao, "Knowledge graph embedding with concepts," Knowledge-Based Systems 164 (2019) 38–44.

[14] Krompaβ, Denis, Stephan Baier, and Volker Tresp. "Type-Constrained Representation Learning in Knowledge Graphs," international semantic web conference (2015): 640-655.

[15] M. Nickel, V. Tresp, H.-P. Kriegel, "A three-way model for collective learning on multi-relational data," in: ICML, volume 11, 2011, pp.809–816.

[16] B. Yang, W.-t. Yih, X. He, J. Gao, L. Deng, "Embedding entities and relations for learning and inference in knowledge bases," unpublished.

[17] T. Trouillon, J.Welbl, S. Riedel, É. Gaussier, G. Bouchard, "Complex embeddings for simple link prediction," in: Proceedings of the 33th International Conference on Machine Learning, 2016, pp. 2071–2080.

[18] M. Nickel, L. Rosasco, T. Poggio, "Holographic embeddings of knowledge graphs," in: Thirtieth AAAI conference on artificial intelligence, 2016.

[19] R. Socher, D. Chen, C. D. Manning, A. Ng, "Reasoning with neural tensor networks for knowledge base completion," in: Advances in neural information processing systems, 2013, pp. 926–934.

[20] X. Dong, E. Gabrilovich, G. Heitz, W. Horn, N. Lao, K. Murphy, T. Strohmann, S. Sun, W. Zhang, "Knowledge vault: A web-scale approach to probabilistic knowledge fusion," in: Proceedings of the 20th ACMSIGKDD international conference on Knowledge discovery and data mining, ACM, 2014, pp. 601–610.

[21] Q. Liu, H. Jiang, A. Evdokimov, Z.-H. Ling, X. Zhu, S. Wei, Y. Hu, "Probabilistic reasoning via deep learning: Neural association models," unpublished.

[22] T. Dettmers, P. Minervini, P. Stenetorp, S. Riedel, "Convolutional 2d knowledge graph embeddings," in: Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[23] Xie, Ruobing, Zhiyuan Liu, and Maosong Sun. "Representation learning of knowledge graphs with hierarchical types," international joint conference on artificial intelligence (2016): 2965-2971.

[24] Diaz, Gonzalo, Achille Fokoue, and Mohammad Sadoghi. "EmbedS: Scalable, Ontology-aware Graph Embeddings," extending database technology (2018): 433-436.

[25] K. Toutanova, D. Chen, "Observed versus latent features for knowledge base and text inference," in: Proceedings of the 3rdWorkshop on Continuous Vector Space Models and their Compositionality, 2015, pp. 57–66.