TranS: Transition-based Knowledge Graph Embedding with Synthetic Relation Representation

Xuanyu Zhang, Qing Yang and **Dongliang Xu**
Du Xiaoman Financial
{zhangxuanyu,yangqing,xudongliang}@duxiaoman.com

Abstract

Knowledge graph embedding (KGE) aims to learn continuous vector representations of relations and entities in knowledge graph (KG). Recently, transition-based KGE methods have become popular and achieved promising performance. However, scoring patterns like TransE are not suitable for complex scenarios where the same entity pair has different relations. Although some models attempt to employ entity-relation interaction or projection to improve entity representation for one-to-many/many-to-one/many-to-many complex relations, they still continue the traditional scoring pattern, where only a single relation vector in the relation part is used to translate the head entity to the tail entity or their variants. And recent research shows that entity representation only needs to consider entities and their interactions to achieve better performance. Thus, in this paper, we propose a novel transition-based method, TranS, for KGE. The single relation vector of the relation part in the traditional scoring pattern is replaced by the synthetic relation representation with entity-relation interactions to solve these issues. And the entity part still retains its independence through entity-entity interactions. Experiments on a large KG dataset, ogbl-wikikg2, show that our model achieves state-of-the-art results.

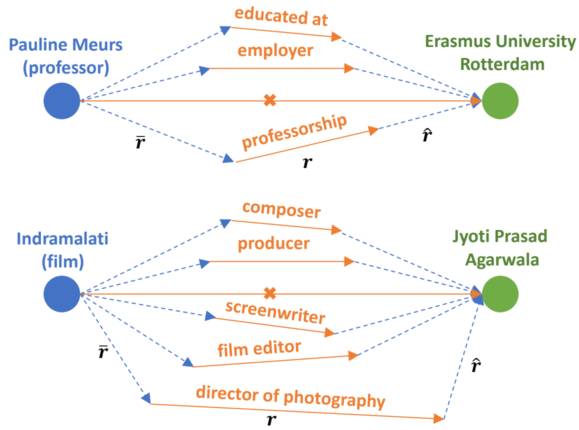

Figure 1: Examples from ogbl-wikikg2. It is difficult for a single relation vector to represent different relations between the same entity pairs.

1 Introduction

Knowledge graphs (KGs), such as Freebase (Bollacker et al., 2008), Wikidata (Vrandečić and Krötzsch, 2014), DBpedia (Lehmann et al., 2015) and Yago (Rebele et al., 2016), play a very important role in many fields, including question answering (Huang et al., 2019), semantic parsing (Yih et al., 2015), information retrieval (Xiong et al., 2017) and so on. KG, as a multi-relational graph, is composed of entities as nodes and relations as different types of edges. It is usually represented as the form of triplets (h, r, t), i.e., (*head entity*, *relation*, *tail entity*), where *relation* indicates the relationship between the two entities.

Knowledge graph embedding (KGE) is an important and fundamental research topic in KG. It aims to learn dense semantic representations of entities and relations for downstream tasks such as KG completion and link prediction. Generally speaking, KGE methods can be roughly divided into the following directions: translational distance (Bordes et al., 2013; Wang et al., 2014; Fan et al., 2014; Lin et al., 2015; Ji et al., 2015, 2016; Feng et al., 2016), semantic matching (Nickel et al., 2011; Bordes et al., 2011, 2014; García-Durán et al., 2014; Yang et al., 2015; Nickel et al., 2016; Balazevic et al., 2019) and neural networks (Socher et al., 2013; Dong et al., 2014; Liu et al., 2016; Dettmers et al., 2018; Nguyen et al., 2018). Because transition-based KGE method like TransE (Bordes et al., 2013) is simple and effective, this series of models are becoming more and more popular in both academia and industry. Specifically, TransE makes the difference between two entity vectors ($\mathbf{h}$ and $\mathbf{t}$) approximate to the relation vector ($\mathbf{r}$), i.e., $\mathbf{t} - \mathbf{h} \approx \mathbf{r}$. That is to say, the relation r is characterized by the translating vector $\mathbf{r}$.

Findings of the Association for Computational Linguistics: EMNLP 2022, pages 1202–1208
December 7-11, 2022 ©2022 Association for Computational Linguistics

However, TransE is not suitable to deal with complex relations like one-to-many/many-to-one/many-to-many. For example, in Figure 1, after graduating from Erasmus University Rotterdam, Pauline Meurs became a professor at the same university. And the composer, producer, screenwriter, editor and director of the film, Indramalati, can be the same person, Jyoti Prasad Agarwala. Although previous models (Wang et al., 2014; Lin et al., 2015; Qian et al., 2018; Chao et al., 2021; Yu et al., 2021) such as TransH/R/D have considered relevant issues, they still focus on the entity-relation projection or interaction in the entity part and continue the TransE pattern, $\mathbf{R_t} - \mathbf{R_h} \approx \mathbf{r}$, where $\mathbf{R_t}$ and $\mathbf{R_h}$ is the deformation of $\mathbf{t}$ and $\mathbf{h}$, $\mathbf{R_t} - \mathbf{R_h}$ is the entity part, and $\mathbf{r}$ is the relation part. Actually, recent research, InterHT (Wang et al., 2022), shows that the entity part only needs to consider the head and tail entities and their interaction information to achieve remarkable performance and outperform previous TransX series models. Unfortunately, it again ignores the problem of complex relation representation. Therefore, from the perspective of interaction, how to solve the problem in Figure 1 by introducing entity-relation interactions in the relation part under the condition that only entity-entity interactions are retained in the entity part needs to be further considered.

To this end, we propose a novel transition-based knowledge graph embedding model, TranS, which replaces traditional scoring pattern with synthetic relation pattern, i.e., $\mathbf{R_t} - \mathbf{R_h} \approx \bar{\mathbf{r}} + \mathbf{r} + \hat{\mathbf{r}}$. The final relation representation is the sum of multiple relation vectors. Two of them ($\bar{\mathbf{r}}$, $\hat{\mathbf{r}}$) are also related to the head entity *h* and the tail entity *t* in addition to the relation *r* (orange solid lines denote $\mathbf{r}$, and blue dotted lines denote $\bar{\mathbf{r}}, \hat{\mathbf{r}}$ in Figure 1). For one thing, in the entity part, instead of using entity-relation interaction and projection, it focuses only on entities and their interactions themselves to guarantee their independence and effectiveness. For another thing, different from other methods that utilize entity-relation interactions in the entity part, our method migrates their interactions to the relation part and forms synthetic relation representation, which can effectively solve the problem that a single relation vector cannot represent different relations when facing the same entity pair. Experiments on a large knowledge graph dataset, ogbl-wikikg2, show that our proposed model achieves the best results with fewer parameters.

2 Methodology

2.1 TranS

Our proposed TranS model first breaks the traditional scoring patterns $\mathbf{R_t} - \mathbf{R_h} \approx \mathbf{r}$ in previous models (Bordes et al., 2013; Wang et al., 2014; Fan et al., 2014; Lin et al., 2015; Chao et al., 2021; Yu et al., 2021; Wang et al., 2022). It replaces single relation vector $\mathbf{r}$ with synthetic relation vectors $\bar{\mathbf{r}} + \mathbf{r} + \hat{\mathbf{r}}$, i.e., $\mathbf{R_t} - \mathbf{R_h} \approx \bar{\mathbf{r}} + \mathbf{r} + \hat{\mathbf{r}}$, where $\bar{\mathbf{r}}$ is an adjoint relation vector related to the head entity and $\hat{\mathbf{r}}$ is another adjoint relation vector related to the tail entity. The illustration of TranS is shown in Figure 2 (f). Two entity and three relation representations together make up our proposed scoring function $f_r(h, t)$. That is to say, the synthetic relation representation in the right relation part consists of the sum of three different relation vectors. To make full use of context information, we use adjoint vectors and Hadamard product $\circ$ to interact with $\mathbf{h}$, $\mathbf{t}$, $\bar{\mathbf{r}}$ and $\hat{\mathbf{r}}$ separately:

$$
\begin{aligned}
f_r(h,t) &= -||\mathbf{R_h} - \mathbf{R_t} + \mathbf{R_r}||, \\
\mathbf{R_h} &= \mathbf{h} \circ \tilde{\mathbf{t}}, \\
\mathbf{R_t} &= \mathbf{t} \circ \tilde{\mathbf{h}}, \\
\mathbf{R_r} &= \bar{\mathbf{r}} \circ \mathbf{h} + \mathbf{r} + \hat{\mathbf{r}} \circ \mathbf{t},
\end{aligned}
\tag{1}
$$

where $\mathbf{h}$, $\mathbf{t}$ and $\mathbf{r}$ denote main vectors similar to those in traditional scoring patterns. $\tilde{\mathbf{h}}$ represents the adjoint head entity vector and $\tilde{\mathbf{t}}$ represents the adjoint tail entity vector. Accordingly, $\mathbf{R_h}$ is the representation of the head entity that combines information of the tail entity, and $\mathbf{R_t}$ is the representation of the tail entity integrating information of the head entity. $\bar{\mathbf{r}} \circ \mathbf{h}$ is the representation of the adjoint relation with the head entity information, and $\hat{\mathbf{r}} \circ \mathbf{t}$ is the representation of another adjoint relation with the tail entity information. Thus, the final equation can be represented as:

$$
f_r(h,t) = -||\mathbf{h} \circ \tilde{\mathbf{t}} - \mathbf{t} \circ \tilde{\mathbf{h}} + \underline{\bar{\mathbf{r}} \circ \mathbf{h} + \mathbf{r} + \hat{\mathbf{r}} \circ \mathbf{t}}||. \tag{2}
$$

Following previous works (Yu et al., 2021; Wang et al., 2022), we add an unit vector $\mathbf{e}$ to $\mathbf{R_h}$ and $\mathbf{R_t}$, i.e., $\mathbf{h} \circ \tilde{\mathbf{t}} \rightarrow \mathbf{h} \circ (\tilde{\mathbf{t}} + \mathbf{e})$, $\mathbf{t} \circ \tilde{\mathbf{h}} \rightarrow \mathbf{t} \circ (\tilde{\mathbf{h}} + \mathbf{e})$. And considering the out-of-vocabulary problem, we also use the NodePiece (Galkin et al., 2022) to learn a fixed-size entity vocabulary.

2.2 Training

Inspired by previous works (Chao et al., 2021; Zhang and Yang, 2021; Wang et al., 2022), we use

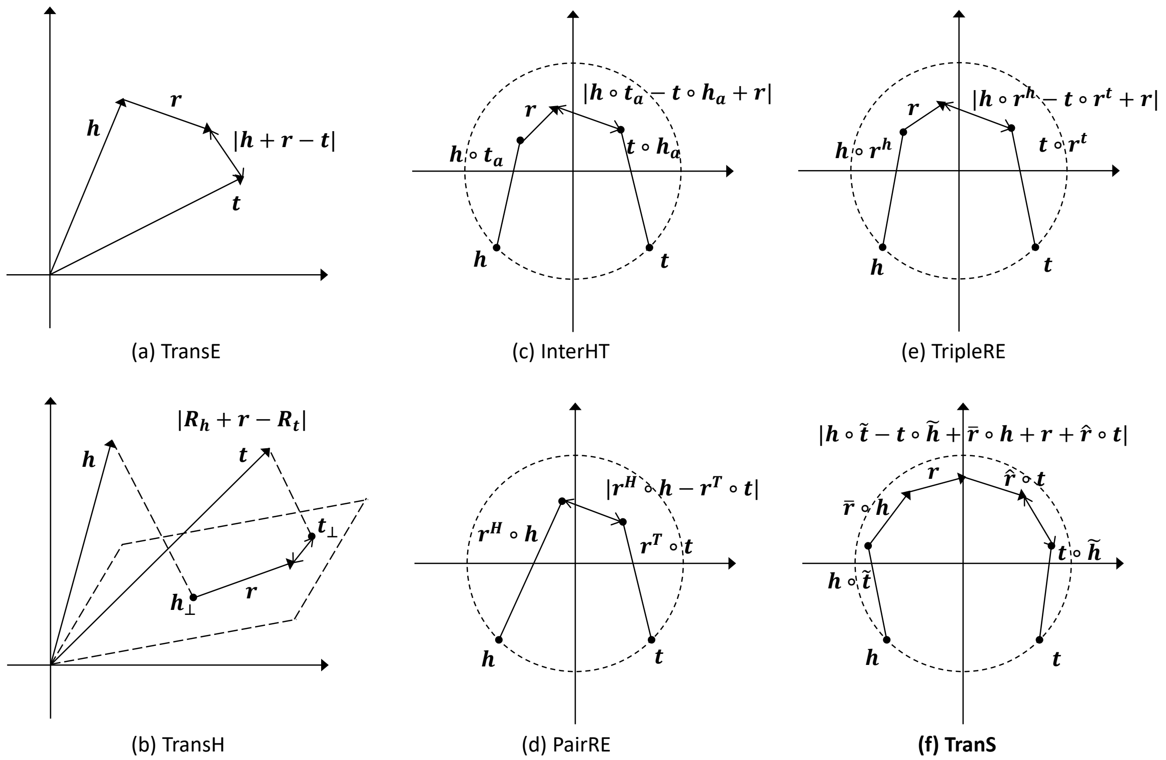

Figure 2: Comparison of different transition-based KGE models.

the self-adversarial negative sampling loss (Sun et al., 2019) as our loss function, which is defined as follows:

$$\mathcal{L} = -\, log\, \sigma(\gamma - f_r(h,t)) \\ - \sum_{i=1}^{n} p(h_i^{'}, r, t_i^{'}) log\, \sigma(f_r(h_i^{'}, t_i^{'}) - \gamma), \tag{3}$$

where γ is a fixed margin, σ is the sigmoid function, and $(h_i^{'}, r, t_i^{'})$ is the i-th of n randomly sampled negative triplets. And the weights of this negative sample $p(h_i^{'}, r, t_i^{'})$ can be calculated as follows:

$$p(h_i^{'}, r, t_i^{'}) = \frac{\exp f_r(h_i^{'}, t_i^{'})}{\sum_j \exp f_r(h_j^{'}, t_j^{'})}. \tag{4}$$

2.3 Comparison

As shown in Figure 2, the main difference between our model (f) and previous transition-based KGE methods (a,b,c,d,e) is the synthetic relation representation. That is to say, it changes single relation representation $\mathbf{r}$ in traditional scoring pattern $\mathbf{R_t} - \mathbf{R_h} \approx \mathbf{r}$ to synthetic relation representation $\bar{\mathbf{r}} + \mathbf{r} + \hat{\mathbf{r}}$ in our proposed new pattern $\mathbf{R_t} - \mathbf{R_h} \approx \bar{\mathbf{r}} + \mathbf{r} + \hat{\mathbf{r}}$. Specifically, different from InterHT (Wang et al., 2022), the relation part of our scoring function is the sum of multiple relation vectors $\mathbf{R_r} = \bar{\mathbf{r}} \circ \mathbf{h} + \mathbf{r} + \hat{\mathbf{r}} \circ \mathbf{t}$ rather than single vector $\mathbf{r}$. Comparing with TripleRE (Yu et al., 2021), where three relations are applied into three parts ($\mathbf{R_h} = \mathbf{h} \circ \mathbf{r^h}$, $\mathbf{R_t} = \mathbf{t} \circ \mathbf{r^t}$, $\mathbf{R_r} = \mathbf{r^m}$) of traditional scoring patterns with addition and subtraction operations, our proposed TranS only applies synthetic relation vectors into the relation part $\mathbf{R_r} = \bar{\mathbf{r}} \circ \mathbf{h} + \mathbf{r} + \hat{\mathbf{r}} \circ \mathbf{t}$ of scoring functions with addition operations.

3 Experiments

3.1 Dataset and Metric

Ogbl-wikikg2 (Hu et al., 2020) is a large KG dataset extracted from Wikidata (Vrandečić and Krötzsch, 2014). It contains a set of triplet edges, capturing the different types of relations between entities in the world. The statistics of the dataset are shown in Table 1. It contains 2,500,604 entities, 535 relation types and 17,137,181 edges. Following official guidelines, we evaluate the KGE performance by predicting new triplet edges according to the training edges. The evaluation metric follows the standard filtered metric widely used in KG. Specifically, each test triplet edge is corrupted by replacing its head or tail with randomly sampled negative entities, while ensuring the resulting

Type	Train	Validation	Test	Nodes	Relations	Edges
#Number	16,109,182	429,456	598,543	2,500,604	535	17,137,181

Table 1: Statistics of the ogbl-wikikg2 dataset.

Model	#Params	#Dims	Test MRR	Valid MRR
TransE (Bordes et al., 2013)	1251M	500	0.4256 ± 0.0030	0.4272 ± 0.0030
RotatE (Sun et al., 2019)	1250M	250	0.4332 ± 0.0025	0.4353 ± 0.0028
PairRE (Chao et al., 2021)	500M	200	0.5208 ± 0.0027	0.5423 ± 0.0020
AutoSF (Zhang et al., 2020)	500M	-	0.5458 ± 0.0052	0.5510 ± 0.0063
ComplEx (Trouillon et al., 2016)	1251M	250	0.5027 ± 0.0027	0.3759 ± 0.0016
TripleRE (Yu et al., 2021)	501M	200	0.5794 ± 0.0020	0.6045 ± 0.0024
ComplEx-RP (Chen et al., 2021)	250M	50	0.6392 ± 0.0045	0.6561 ± 0.0070
AutoSF + NodePiece	6.9M	-	0.5703 ± 0.0035	0.5806 ± 0.0047
TripleREv2 + NodePiece	7.3M	200	0.6582 ± 0.0020	0.6616 ± 0.0018
TripleREv3 + NodePiece	36.4M	200	0.6866 ± 0.0014	0.6955 ± 0.0008
InterHT + NodePiece	19.2M	200	0.6779 ± 0.0018	0.6893 ± 0.0015
TranS + NodePiece	**19.2M**	200	$\mathbf{0.6882 \pm 0.0019}$	$\mathbf{0.6988 \pm 0.0006}$

Table 2: Results on the ogbl-wikikg2 dataset.

triplets do not appear in KG. The goal is to rank the true head or tail entities higher than the negative entities, which is measured by Mean Reciprocal Rank (MRR).

We follow the original dataset partition. The triplets are split according to time to simulate a real KG completion scenario where missing triplets that are not present at a specific timestamp need to be filled. The training set contains 16,109,182 triplets, the validation set contains 429,456 triplets, and the test set contains 598,543 triplets.

3.2 Implementation Details

In our experiments, Adam (Kingma and Ba, 2014) is used as our optimizer with 0.0005 learning rate. The batch size of the model is set to 512. To prevent overfitting, we use the dropout technique and set it to 0.05. The negative sampling size is set to 128. And the dimension of each embedding vector in Eq. 2 is set to 200. The maximum number of training steps is 800 thousand. We validate the model every 20 thousand steps. The number of anchors for NodePiece is 20 thousand. And γ in the loss function is set to 6. The final model is evaluated with 10 different random seeds. Our code is publicly available at the link: `https://github.com/xyznlp/TranS`.

3.3 Results

The results are shown in Table 2. Our model achieves 0.6988 (validation set) and 0.6882 (test set) on MRR, which outperforms the previous best model, TripleREv3, on the ogbl-wikikg2 dataset. Especially, the parameters of our model (19.2M) are about half of TripleREv3 (36.4M). So the experimental results show that our proposed method can improve the model performance effectively with fewer parameters. Besides, we also construct a 38.4M TranS (large) model, the best score of which can reach 0.7101 (validation set) and 0.6992 (test set) on MRR. Comparing the two groups with similar numbers of parameters, i.e., TranS versus InterHT and TranS (large) versus TripleREv3, we can observe more significant improvements.

4 Related Work

Recently, graph structures are used widely in natural language processing, recommendation and other areas (Zhang, 2020; Zhang et al., 2021). KG, as one of the graph structures, uses triples consisting of head nodes, tail nodes and relation edges to represent structured knowledge. To further compare different transition-based knowledge graph embeddings, we summarize related methods in Table 3 with reference to recent research (Ji et al., 2021).

Model	Embedding	Scoring Function	Pattern
TransE	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d, \mathbf{r} \in \mathbb{R}^d$	$-\lVert\mathbf{h} + \mathbf{r} - \mathbf{t}\rVert_{1/2}$	$\mathcal{T}$
TransR	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d, \mathbf{r} \in \mathbb{R}^k, \mathbf{M}_r \in \mathbb{R}^{k\times d}$	$-\lVert\mathbf{M}_r\mathbf{h} + \mathbf{r} - \mathbf{M}_r\mathbf{t}\rVert_2^2$	$\mathcal{T}$
TransH	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d, \mathbf{r}, \mathbf{w}_r \in \mathbb{R}^d$	$-\left\lVert\left(\mathbf{h} - \mathbf{w}_r^\top\mathbf{h}\mathbf{w}_r\right) + \mathbf{r} - \left(\mathbf{t} - \mathbf{w}_r^\top\mathbf{t}\mathbf{w}_r\right)\right\rVert_2^2$	$\mathcal{T}$
ITransF	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d, \mathbf{r} \in \mathbb{R}^d$	$\left\lVert\boldsymbol{\alpha}_r^H \cdot \mathbf{D} \cdot \mathbf{h} + \mathbf{r} - \boldsymbol{\alpha}_r^T \cdot \mathbf{D} \cdot \mathbf{t}\right\rVert_\ell$	$\mathcal{T}$
TransAt	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d, \mathbf{r} \in \mathbb{R}^d$	$P_r\left(\sigma\left(\mathbf{r}_h\right)\mathbf{h}\right) + \mathbf{r} - P_r\left(\sigma\left(\mathbf{r}_t\right)\mathbf{t}\right)$	$\mathcal{T}$
TransD	$\mathbf{h}, \mathbf{t}, \mathbf{w}_h\mathbf{w}_t \in \mathbb{R}^d, \mathbf{r}, \mathbf{w}_r \in \mathbb{R}^k$	$-\left\lVert\left(\mathbf{w}_r\mathbf{w}_h^\top + \mathbf{I}\right)\mathbf{h} + \mathbf{r} - \left(\mathbf{w}_r\mathbf{w}_t^\top + \mathbf{I}\right)\mathbf{t}\right\rVert_2^2$	$\mathcal{T}$
TransM	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d, \mathbf{r} \in \mathbb{R}^d$	$-\theta_r\lVert\mathbf{h} + \mathbf{r} - \mathbf{t}\rVert_{1/2}$	$\mathcal{T}$
TranSparse	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d, \mathbf{r} \in \mathbb{R}^k, \mathbf{M}_r \in \mathbb{R}^{k\times d}$	$-\lVert\mathbf{M}_r\left(\theta_r\right)\mathbf{h} + \mathbf{r} - \mathbf{M}_r\left(\theta_r\right)\mathbf{t}\rVert_{1/2}^2$	$\mathcal{T}$
	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d, \mathbf{M}_r^1, \mathbf{M}_r^2 \in \mathbb{R}^{k\times d}$	$-\left\lVert\mathbf{M}_r^1\left(\theta_r^1\right)\mathbf{h} + \mathbf{r} - \mathbf{M}_r^2\left(\theta_r^2\right)\mathbf{t}\right\rVert_{1/2}^2$	$\mathcal{T}$
PairRE	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d, \mathbf{r^H}, \mathbf{r^T} \in \mathbb{R}^d$	$-\lvert\lvert\mathbf{h} \circ \mathbf{r^H} - \mathbf{t} \circ \mathbf{r^T}\rvert\rvert$	$\mathcal{T}$
TripleRE	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d, \mathbf{r^H}, \mathbf{r^T}, \mathbf{r^M} \in \mathbb{R}^d$	$-\lvert\lvert\mathbf{h} \circ \mathbf{r^H} - \mathbf{t} \circ \mathbf{r^T} + \mathbf{r^M}\rvert\rvert$	$\mathcal{T}$
InterHT	$\mathbf{h}, \mathbf{t}, \mathbf{h_a}, \mathbf{t_a} \in \mathbb{R}^d, \mathbf{r} \in \mathbb{R}^d$	$-\lvert\lvert\mathbf{h} \circ \mathbf{t_a} - \mathbf{t} \circ \mathbf{h_a} + \mathbf{r}\rvert\rvert$	$\mathcal{T}$
TransS	$\mathbf{h}, \mathbf{t}, \tilde{\mathbf{h}}, \tilde{\mathbf{t}} \in \mathbb{R}^d, \mathbf{r}, \bar{\mathbf{r}}, \hat{\mathbf{r}} \in \mathbb{R}^d$	$-\lvert\lvert\mathbf{h} \circ \tilde{\mathbf{t}} - \mathbf{t} \circ \tilde{\mathbf{h}} + \underline{\bar{\mathbf{r}} \circ \mathbf{h} + \mathbf{r} + \hat{\mathbf{r}} \circ \mathbf{t}}\rvert\rvert$	$\mathcal{S}$

Table 3: Summary of transition-based knowledge graph embedding models. $\mathcal{T}$ represents the traditional scoring pattern $-\lvert\lvert\mathbf{R_h} - \mathbf{R_t} + \mathbf{r}\rvert\rvert$. And $\mathcal{S}$ represents our proposed new scoring pattern $-\lvert\lvert\mathbf{R_h} - \mathbf{R_t} + \bar{\mathbf{r}} + \mathbf{r} + \hat{\mathbf{r}}\rvert\rvert$.

Transition-based methods measure the plausibility of fact triples (h,r,t) as the distance between entities. TransE (Bordes et al., 2013), as a representative method, models relationships by interpreting them as translations operating on the low-dimensional embeddings of the entities, i.e., $t-h \approx r$. Although it is simple and efficient, it cannot handle complex relations. Thus, several TransX models (TransH (Wang et al., 2014), TransR (Lin et al., 2015), TransD (Ji et al., 2015)) are proposed based on hyperplane or multiple embedding spaces for these issues. For example, TransR (Lin et al., 2015) projects entities from entity space to corresponding relation space and builds translations between projected entities. And recent works also begin to utilize multiple vectors to represent entities and relations and conduct their interactions. For example, PairRE (Chao et al., 2021) and TripleRE (Yu et al., 2021) employ two and three relation vectors to represent relation information, respectively. Especially, InterHT (Wang et al., 2022) outperforms previous models only with two head and tail vectors and their interactions in the entity part. But InterHT again ignores the problem of complex relation representation. Different from previous models, from the perspective of interaction (Zhang et al., 2022; Zhang and Wang, 2020; Zhang, 2019), our proposed TranS introduces entity-entity interaction in the entity part like InterHT and migrates entity-relation interaction from the entity part to the relation part. It can not only preserve the independence of entity representation, but also utilize entity-relation interaction in the relation part to solve the above problem.

5 Conclusion

In this paper, we propose a novel transition-based knowledge graph embedding model, TranS, to solve the representation problem of complex scenarios where the same entity pair has different relations. TranS replaces the single relation vector of the relation part in traditional scoring patterns with synthetic relation representation. It not only retains the independence of entity interaction in the entity part, but also introduces entity-relation interaction in the relation part. Experiments on a large KG dataset, ogbl-wikikg2, show that our model achieves the best results with fewer parameters.

Limitations

Although our model has achieved the best performance on relevant datasets, it still focuses on current or local KG triples to learn entity and relation representations. Actually, in large-scale knowledge graphs, neighborhoods can provide extra information for entity representation or initialization like NodePiece. Thus the performance of our model can be further improved by exploring additional neighbor information and encoding methods.

References

Ivana Balazevic, Carl Allen, and Timothy Hospedales. 2019. TuckER: Tensor factorization for knowledge graph completion. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5185–5194, Hong Kong, China. Association for Computational Linguistics.

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

Antoine Bordes, Xavier Glorot, Jason Weston, and Yoshua Bengio. 2014. A semantic matching energy function for learning with multi-relational data. *Mach. Learn.*, 94(2):233–259.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26.

Antoine Bordes, Jason Weston, Ronan Collobert, and Yoshua Bengio. 2011. Learning structured embeddings of knowledge bases. *Proceedings of the AAAI Conference on Artificial Intelligence*.

Linlin Chao, Jianshan He, Taifeng Wang, and Wei Chu. 2021. PairRE: Knowledge graph embeddings via paired relation vectors. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4360–4369, Online. Association for Computational Linguistics.

Yihong Chen, Pasquale Minervini, Sebastian Riedel, and Pontus Stenetorp. 2021. Relation prediction as an auxiliary training objective for improving multi-relational graph representations. In *3rd Conference on Automated Knowledge Base Construction*.

Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2d knowledge graph embeddings. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

Xin Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, N. Lao, Kevin P. Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. 2014. Knowledge vault: a web-scale approach to probabilistic knowledge fusion. *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*.

Miao Fan, Qiang Zhou, Emily Chang, and Fang Zheng. 2014. Transition-based knowledge graph embedding with relational mapping properties. In *Proceedings of the 28th Pacific Asia conference on language, information and computing*, pages 328–337.

Jun Feng, Minlie Huang, Mingdong Wang, Mantong Zhou, Yu Hao, and Xiaoyan Zhu. 2016. Knowledge graph embedding by flexible translation. In *Proceedings of the Fifteenth International Conference on Principles of Knowledge Representation and Reasoning*, KR'16, page 557–560. AAAI Press.

Mikhail Galkin, Etienne Denis, Jiapeng Wu, and William L. Hamilton. 2022. Nodepiece: Compositional and parameter-efficient representations of large knowledge graphs. In *International Conference on Learning Representations*.

Alberto García-Durán, Antoine Bordes, and Nicolas Usunier. 2014. Effective blending of two and three-way interactions for modeling multi-relational data. In *Proceedings of the 2014th European Conference on Machine Learning and Knowledge Discovery in Databases - Volume Part I*, ECMLPKDD'14, page 434–449, Berlin, Heidelberg. Springer-Verlag.

Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. 2020. Open graph benchmark: Datasets for machine learning on graphs. *arXiv preprint arXiv:2005.00687*.

Xiao Huang, Jingyuan Zhang, Dingcheng Li, and Ping Li. 2019. Knowledge graph embedding based question answering. *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*.

Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. 2015. 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)*, pages 687–696.

Guoliang Ji, Kang Liu, Shizhu He, and Jun Zhao. 2016. Knowledge graph completion with adaptive sparse transfer matrix. In *Thirtieth AAAI conference on artificial intelligence*.

Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and S Yu Philip. 2021. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N. Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, S. Auer, and Christian Bizer. 2015. Dbpedia - a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web*, 6:167–195.

Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *Twenty-ninth AAAI conference on artificial intelligence*.

Quan Liu, Hui Jiang, Zhen-Hua Ling, Si Wei, and Yu Hu. 2016. Probabilistic reasoning via deep learning: Neural association models. *CoRR*, abs/1603.07704.

Dai Quoc Nguyen, Tu Dinh Nguyen, Dat Quoc Nguyen, and Dinh Q. Phung. 2018. A novel embedding model for knowledge base completion based on convolutional neural network. In *NAACL*.

Maximilian Nickel, Lorenzo Rosasco, and Tomaso Poggio. 2016. Holographic embeddings of knowledge graphs. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, page 1955–1961. AAAI Press.

Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. 2011. A three-way model for collective learning on multi-relational data. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ICML'11, page 809–816, Madison, WI, USA. Omnipress.

Wei Qian, Cong Fu, Yu Zhu, Deng Cai, and Xiaofei He. 2018. Translating embeddings for knowledge graph completion with relation attention mechanism. In *IJCAI*, pages 4286–4292.

Thomas Rebele, Fabian Suchanek, Johannes Hoffart, Joanna Biega, Erdal Kuzey, and Gerhard Weikum. 2016. Yago: A multilingual knowledge base from wikipedia, wordnet, and geonames. In *International semantic web conference*, pages 177–185. Springer.

Richard Socher, Danqi Chen, Christopher D. Manning, and A. Ng. 2013. Reasoning with neural tensor networks for knowledge base completion. In *NIPS*.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. Rotate: Knowledge graph embedding by relational rotation in complex space. In *International Conference on Learning Representations*.

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *International conference on machine learning*, pages 2071–2080. PMLR.

Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85.

Baoxin Wang, Qingye Meng, Ziyue Wang, Dayong Wu, Wanxiang Che, Shijin Wang, Zhigang Chen, and Cong Liu. 2022. Interht: Knowledge graph embeddings by interaction between head and tail entities. *arXiv preprint arXiv:2202.04897*.

Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28.

Chenyan Xiong, Russell Power, and Jamie Callan. 2017. Explicit semantic ranking for academic search via knowledge graph embedding. *Proceedings of the 26th International Conference on World Wide Web*.

Bishan Yang, Scott Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *Proceedings of the International Conference on Learning Representations (ICLR) 2015*.

Wen-tau Yih, Ming-Wei Chang, Xiaodong He, and Jianfeng Gao. 2015. Semantic parsing via staged query graph generation: Question answering with knowledge base. In *Proceedings of the 53rd ACL*, pages 1321–1331, Beijing, China.

Long Yu, ZhiCong Luo, Deng Lin, HuanYong Liu, and YaFeng Deng. 2021. Triplere: Knowledge graph embeddings via triple relation vectors. *viXra preprint viXra:2112.0095*.

Xuanyu Zhang. 2019. MCˆ2: Multi-perspective convolutional cube for conversational machine reading comprehension. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 6185–6190, Florence, Italy. Association for Computational Linguistics.

Xuanyu Zhang. 2020. Cfgnn: Cross flow graph neural networks for question answering on complex tables. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):9596–9603.

Xuanyu Zhang and Zhichun Wang. 2020. Rception: Wide and deep interaction networks for machine reading comprehension (student abstract). *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(10):13987–13988.

Xuanyu Zhang and Qing Yang. 2021. Dml: Dynamic multi-granularity learning for bert-based document reranking. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*, CIKM '21, page 3642–3646, New York, NY, USA.

Xuanyu Zhang, Qing Yang, and Dongliang Xu. 2021. Combining explicit entity graph with implicit text information for news recommendation. In *Companion Proceedings of the Web Conference 2021*, WWW '21, page 412–416, New York, NY, USA. Association for Computing Machinery.

Xuanyu Zhang, Qing Yang, and Dongliang Xu. 2022. Deepvt: Deep view-temporal interaction network for news recommendation. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*, CIKM '22, page 2640–2650, New York, NY, USA.

Yongqi Zhang, Quanming Yao, Wenyuan Dai, and Lei Chen. 2020. Autosf: Searching scoring functions for knowledge graph embedding. In *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, pages 433–444. IEEE.