TOWARD UNDERSTANDING THE EFFECT OF LOSS FUNCTION ON THE PERFORMANCE OF KNOWLEDGE GRAPH EMBEDDING

Anonymous authors
Paper under double-blind review

ABSTRACT

Knowledge graphs (KGs) represent world's facts in structured forms. KG completion exploits the existing facts in a KG to discover new ones. Translation-based embedding model (TransE) is a prominent formulation to do KG completion. Despite the efficiency of TransE in memory and time, it suffers from several *limitations* in encoding relation patterns such as symmetric, reflexive etc. To resolve this problem, most of the attempts have circled around the revision of the *score* function of TransE i.e., proposing a more complicated score function such as Trans(A, D, G, H, R, etc) to mitigate the limitations. In this paper, we tackle this problem from a different perspective. We show that existing theories corresponding to the limitations of TransE are inaccurate because they ignore the effect of *loss* function. Accordingly, we pose theoretical investigations of the main *limitations* of TransE in the light of *loss* function. To the best of our knowledge, this has not been investigated so far comprehensively. We show that by a proper selection of the loss function for training the TransE model, the main limitations of the model are mitigated. This is explained by setting upper-bound for the scores of positive samples, showing the region of truth (i.e., the region that a triple is considered positive by the model). Our theoretical proofs with experimental results fill the gap between the capability of translation-based class of embedding models and the loss function. The theories emphasise the importance of the selection of the loss functions for training the models. Our experimental evaluations on different loss functions used for training the models justify our theoretical proofs and confirm the importance of the loss functions on the performance.

1 INTRODUCTION

Knowledge is considered as commonsense facts and other information accumulated from different sources. Throughout history, civilizations have evolved due to increase in the knowledge. With the passage of time, humans obtain many relations among different entities. Therefore, development of proper knowledge representation (KR) and management systems is essential.

The aim of KR is to study how the beliefs can be represented in an explicit, symbolic notation proper to automated reasoning. Knowledge Graph (KG) is a new direction for KR. KGs are usually represented as a set of triples (h, r, t) where h, t are entities and r is a relation, e.g. $(iphone, hyponym, smartphone)$. Entities and relations are nodes and edges in the graph, respectively.

KGs are inherently incomplete, making prediction of missing links always relevant. Among different approaches used for KG completion, KG Embedding (KGE) has recently received growing attentions. KGE embeds entities and relations as low dimensional vectors. To measure the degree of plausibility of a triple, a scoring function is defined over the embeddings.

TransE, Translation-based Embedding model, Bordes et al. (2013) is one of the most widely used KGEs. The original assumption of TransE is to hold: $\mathbf{h} + \mathbf{r} = \mathbf{t}$, for every positive triple (h, r, t) where $\mathbf{h}, \mathbf{r}, \mathbf{t} \in R^d$ are embedding vectors of head (h), relation (r) and tail (t) respectively.

TransE and its many variants like TransH Wang et al. (2014) and TransR Lin et al. (2015b), underperform greatly compared to the current state-of-the-art embedding models due to the inherent limitations of their scoring functions.

Recent work has the main limitations of Translation-based models. Wang et al. (2018) reveals that TransE cannot encode a relation pattern which is neither reflexive nor irreflexive. Sun et al. (2019) prove that TransE is incapable of encoding symmetric relation. Wang et al. (2014) adds that TransE cannot properly encode reflexive, one-to-many, many-to-one and many-to-many relations.

TransH, TransR and TransD (Wang et al., 2014; Lin et al., 2015b; Ji et al., 2015) can handle the mentioned problems of TransE (i.e. one-to-many, many-to-one, many-to-many and reflexive) by projecting entities to relation space before applying translation. However, Kazemi & Poole (2018) investigate three additional limitations of TransE, FTransE (Feng et al., 2016), STransE (Nguyen et al., 2016), TransH and TransR models: (i) if the models encode a reflexive relation r, they automatically encode symmetric, (ii) if the models encode a reflexive relation r, they automatically encode transitive and, (iii) if entity e_1 has relation r with every entity in $\Delta \in \mathcal{E}$ and entity e_2 has relation r with one of entities in Δ, then e_2 must have the relation r with every entity in Δ.

The mentioned works have investigated these limitations by focusing on the capability of *scoring* functions in encoding relation patterns. However, we prove that the selection of *loss* function affects the boundary of score functions; consequently, the selection of loss functions significantly affects the limitations. Therefore, the above mentioned theories corresponding to the limitations of translation-based embedding models in encoding relation patterns are inaccurate. We pose new theories about the limitations of TransX(X=H,D,R, etc) models considering the *loss* functions. To the best of our knowledge, it is the first time that the effect of loss function is investigated to prove theories corresponding to the limitations of translation-based models.

In a nutshell, the key contributions of this paper is as follows: (i) We show that different loss functions enforce different upper-bounds and lower-bounds for the scores of positive and negative samples respectively. This implies that existing theories corresponding the limitation of TransX models are inaccurate because the effect of loss function is ignored. We introduce new theories accordingly and prove that the proper selection of loss functions mitigates the main limitations. (ii) We reformulate the existing loss functions and their optimization problems as an standard constrained optimization problem. This makes perfectly clear that how each of the loss functions affect on the boundary of triples scores and consequently ability of relation pattern encoding. (iii) using symmetric relation patterns, we obtain the optimal upper-bound of positive triples score to enable encoding of symmetric patterns. (iv) We additionally investigate the theoretical capability of translation-based embedding model when translation is applied in complex space (TransComplEx). We show that TransComplEx is a more powerful embedding model with fewer theoretical limitations in encoding different relation patterns such as symmetric while it is efficient in memory and time.

2 Related Works

Most of the previous work have investigated the capability of translation-based class of embedding models considering solely the formulation of the score function. Accordingly, in this section, we review the score functions of TransE and some of its variants together with their capabilities. Then, in the next section the existing limitations of Translation-based embedding models emphasized in recent works are reviewed. These limitations will be reinvestigated in the light of *score* and *loss* functions in the section 4.

TransE (Bordes et al., 2013) is one of the earlier KGE models which is efficient in both time and space. The score function of TransE is defined as: $f_r(h,t) = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|$.

TransH (Wang et al., 2014) projects each entity ($\mathbf{e}$) to the relation space ($\mathbf{e}_\perp = \mathbf{e} - \mathbf{w}_r\mathbf{e}\mathbf{w}_r^T$). The score function is defined as $f_r(h,t) = \|\mathbf{h}_\perp + \mathbf{r} - \mathbf{t}_\perp\|$. TransH can encode reflexive, one-to-many, many-to-one and many-to-many relations. However, recent theories (Kazemi & Poole, 2018) prove that encoding reflexive results in encoding the both symmetric and transitive which is undesired.

TransR (Lin et al., 2015b) projects each entity ($\mathbf{e}$) to the relation space by using a matrix provided for each relation ($\mathbf{e}_\perp = \mathbf{e}\mathbf{M}_r$, $\mathbf{M}_r \in R^{d_e \times d_r}$). TransR uses the same scoring function as TransH.

TransD (Ji et al., 2015) provides two vectors for each individual entities and relations $(\mathbf{h}, \mathbf{h}_p, \mathbf{r}, \mathbf{r}_p, \mathbf{t}, \mathbf{t}_p)$. Head and tail entities are projected by using the following matrices:

$\mathbf{M}_{rh} = \mathbf{r}_p^T \mathbf{h}_p + \mathbf{I}^{m \times n}, \mathbf{M}_{rt} = \mathbf{r}_p^T \mathbf{t}_p + \mathbf{I}^{m \times n}$

The score function of TransD is similar to the score function of TransH.

RotatE (Sun et al., 2019) rotates the head to the tail entity by using relation. RotatE embeds entities and relations in Complex space. By inclusion of constraints on the norm of entity vectors, the model would be degenerated to TransE. The scoring function of RotatE is $f_r(h, t) = \|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\|$, where $\mathbf{h}, \mathbf{r}, \mathbf{t} \in C^d$, and $\circ$ is element-wise product. RotatE obtains the state-of-the-art results using very big embedding dimension (1000) and a lot of negative samples (1000).

TorusE (Ebisu & Ichise, 2018) fixes the problem of regularization in TransE by applying translation on a compact Lie group. The model has several variants including mapping from torus to Complex space. In this case, the model is regarded as a very special case of RotatE Sun et al. (2019) that applies rotation instead of translation in the target Complex space. According to Sun et al. (2019), TorusE is not defined on the entire Complex space. Therefore, it has less representation capacity. TorusE needs a very big embedding dimension (10000 as reported in Ebisu & Ichise (2018)) which is a limitation.

3 THE MAIN LIMITATIONS OF TRANSLATION-BASED EMBEDDING MODELS

Here we review the six limitations of translation-based embedding models in encoding relation patterns (e.g., reflexive, symmetric) mentioned in the literature: Wang et al. (2014); Kazemi & Poole (2018); Wang et al. (2018); Sun et al. (2019).

Limitation L1: Wang et al. (2014): TransE cannot encode reflexive relations when relation vector is non-zero.

Limitation L2 Wang et al. (2018): if TransE encodes a relation r, which is neither reflexive nor irreflexive the following equations should be held simultaneously: $\mathbf{h}_1 + \mathbf{r} = \mathbf{h}_1, \mathbf{h}_2 + \mathbf{r} \neq \mathbf{h}_2$. Therefore, both $\mathbf{r} = \mathbf{0}, \mathbf{r} \neq \mathbf{0}$ should be held, which result in contradiction. In this regard, TransE cannot encode a relation which is neither reflexive nor irreflexive.

Limitation L3 Sun et al. (2019): If relation r is symmetric, the following equations should be held: $\mathbf{h} + \mathbf{r} = \mathbf{t}$ and $\mathbf{t} + \mathbf{r} = \mathbf{h}$. Therefore, $\mathbf{r} = \mathbf{0}$ and so all entities appeared in head or tail parts of training triples will have the same embedding vectors which is undesired. Therefore, TransE cannot properly encode symmetric relation when $\mathbf{r} \neq \mathbf{0}$.

The following limitations are held for TransE, FTransE Feng et al. (2016), STransE Nguyen et al. (2016), TransH and TransR.

Limitation L4 Kazemi & Poole (2018): if a relation r is reflexive on $\Delta \in \mathcal{E}$, where $\mathcal{E}$ is the set of all entities in the KG, r must also be symmetric.

Limitation L5 Kazemi & Poole (2018): if r is reflexive on $\Delta \in \mathcal{E}$, r must also be transitive.

Limitation L6 Kazemi & Poole (2018): if entity e_1 has relation r with every entity in $\Delta \in \mathcal{E}$ and entity e_2 has relation r with one of entities in Δ, then e_2 must have the relation r with every entity in Δ.

4 OUR MODEL

TransE and its variants underperform compared to other embedding models due to their limitations we iterated in Section 3. In this section, we reinvestigate the limitations. We show that the corresponding theoretical proofs are inaccurate because the effect of loss function is ignored. So we propose new theories and prove that each of the limitations of TransE are resolved by revising either the *scoring* function or the *loss*. In this regard, we consider several loss functions and their effects on the boundary of the TransE scoring function. For each of the loss functions, we pose theories corresponding to the limitations. we additionally investigate the limitations of TransE using each of the

loss functions while translation is performed in Complex space (TransComplEx). TransComplEx with a proper selection of loss function further mitigates the limitations as we discuss as follows.

4.1 TRANSCOMPLEX: TRANSLATIONAL EMBEDDING MODEL IN COMPLEX SPACE

Inspired by Trouillon et al. (2016), in this section we propose TransComplEx that translates head entity vector to the conjugate of tail entity vector using relation vector in Complex space. The score function is defined as follows:

$$f_r(h,t) = \|\mathbf{h} + \mathbf{r} - \bar{\mathbf{t}}\| \tag{1}$$

where $\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathcal{C}^d$ are complex vectors i.e., each elements of the vectors is a complex number. For example, the i-th element of the vector $\mathbf{h}$ is denoted by $h_i = Re(h_i) + Im(h_i)$. Respectively, $Re(.), Im(.)$ denote real and imaginary parts of a complex number. The complex vector $\mathbf{h}$ contains real and imaginary vectors parts i.e. $\mathbf{h} = Re(\mathbf{h}) + Im(\mathbf{h})$. $\bar{\mathbf{t}} = Re(\mathbf{t}) - Im(\mathbf{t})$ is conjugate of the complex vector $\mathbf{t}$.

Advantages of TransComplEx:

i) Comparing to TransE and its variants, TransComplEx has less limitations in encoding different relation patterns. The theories and proofs are provided in the next part.

ii) Using conjugate of tail vector in the formulation enables the model to make difference between the role of an entity as subject or object. This cannot be properly captured by TransE and its variants.

iii) Given the example $(A, Like, Juventus)$, $(Juventus, hasPlayer, C.Ronaldo)$, that $C.Ronaldo$ plays for $Juventus$ may affect the person A to like the team. This type of information cannot be properly captured by models such as CP decomposition Hitchcock (1927) where two independent vectors are provided Kazemi & Poole (2018) for $Juventus$ (for subject and object). In contrast, our model uses same real and imaginary vectors for $Juventus$ when it is used as subject or object. Therefore, TransComplEx can properly capture dependency between the two triples with the same entity used as subject and object.

iiii) ComplEx Trouillon et al. (2016) has much more computational complexity comparing to TransComplEx because it needs to compute eight vector multiplications to obtain score of a triple while our model only needs to do four vector summation/subtractions. In the experiment section, we show that TransComplEx outperforms ComplEx on various dataset.

4.2 REINVESTIGATION OF THE LIMITATIONS OF TRANSLATION-BASED EMBEDDING MODELS

The aim of this part is to analyze the limitations of Translation-based embedding models (including TransE and TransComplEx) by considering the effect of both *score* and *loss* functions. Different loss functions provide different upper-bound and lower-bound for positive and negative triples scores, respectively. Therefore, the loss functions affect the limitations of the models to encode relation patterns. To investigate the limitations, we redefine the conditions that a triple is considered as positive or negative by defining upper-bound and lower-bound for the scores.

Lets $f_r(h,t)$, $f_r(h^{'}, t^{'})$ be the scores of a positive (h,r,t) and negative $(h^{'}, r, t^{'})$ triples respectively. The negative triple $(h^{'}, r, t^{'})$ is generated by corruption of either head or tail of the triple (h,r,t) as mentioned in Bordes et al. (2013). Four conditions are defined as follows:

Table 1: Region of truth and falsity

Condition	Positive	Negative	γ_1, γ_2
(a)	$\mathrm{f}_r(h,t) = \gamma_1,$	$\mathrm{f}_r(h^{'}, t^{'}) \geq \gamma_2$	$\gamma_1 = 0,\ \gamma_2 > 0$
(b)	$\mathrm{f}_r(h,t) = \gamma_1$	$\mathrm{f}_r(h^{'}, t^{'}) \geq \gamma_2$	$\gamma_2 > \gamma_1 > 0$
(c)	$\mathrm{f}_r(h,t) \leq \gamma_1$	$\mathrm{f}_r(h^{'}, t^{'}) \geq \gamma_2$	$\gamma_2 > \gamma_1 > 0$
(d)	$\mathrm{f}_r(h,t) \leq \gamma_{1(h,r,t)}$	$\mathrm{f}_r(h^{'}, t^{'}) \geq \gamma_{2(h,r,t)}$	$\gamma_{2(h,r,t)} > \gamma_{1(h,r,t)} > 0$

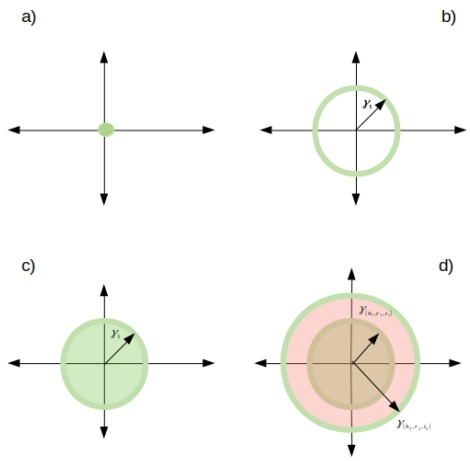

Figure 1: The region of truth for a triple: A triple is positive if (a) its residual vector (i.e., $\epsilon = \mathbf{h}+\mathbf{r}-\mathbf{t}$) becomes $\mathbf{0}$ (b) its residual vector (i.e., ϵ) lies on the border of a sphere with radius γ_1, (c) its residual vector (i.e., ϵ) lies inside of a sphere with radius γ_1, (d) its residual vector (i.e., $\epsilon_{(h_1,r_1,t_1)}$) lies inside of a sphere with radius $\gamma_{(h_1,r_1,t_1)}$.

.

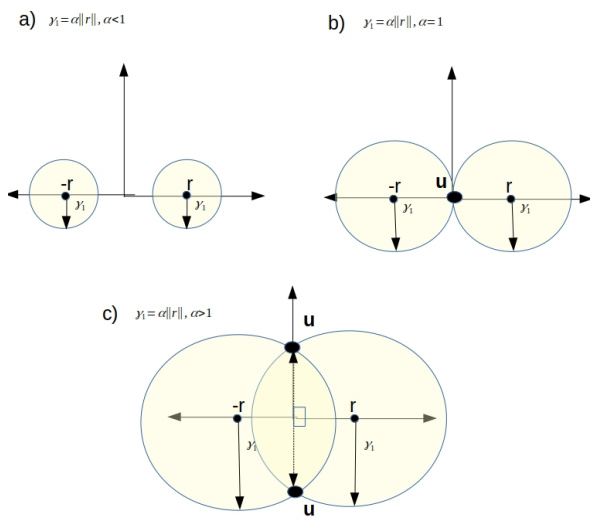

Figure 2: Necessity condition for encoding symmetric relation: (a) when $\alpha < 1$, the model cannot encode symmetric relation.There is not any common points between two hyperspheres). (b) when $\alpha = 1$, the intersection of two hyperspheres is a point. $\mathbf{u} = \mathbf{0}$ means embedding vectors of all entities should be same. Therefore, symmetric relation cannot be encoded. (c) if $\alpha > 1$, symmetric relation can be encoded because there are several points which are intersection of two hyperspheres.

Figure 1 visualizes different conditions mentioned above. The condition (a) indicates a triple is positive if $\mathbf{h} + \mathbf{r} = \mathbf{t}$ holds. It means that the length of *residual vector* i.e., $\epsilon = \mathbf{h} + \mathbf{r} - \mathbf{t}$, is zero. It is the most strict condition that expresses being positive. Authors in Sun et al. (2019); Kazemi & Poole (2018) consider this condition to prove their theories.

Condition (b) considers a triple to be positive if its residual vector lies on a hyper-sphere with radius γ_1. It is less restrictive than the condition (a) which considers a point to express being positive. The optimization problem that satisfies the conditions (a) ($\gamma_1 = 0$) and (b) ($\gamma_1 > 0$) is as follows:

$$\begin{cases} \min_{\xi_{h,t}} \sum_{(h,r,t)\in S^+} {\xi_{h,t}}^2 \\ \text{s.t.} \\ f_r(h,t) = \gamma_1,\ (h,r,t) \in S^+ \\ f_r(h^{'},t^{'}) \geq \gamma_2 - \xi_{h,t},\ (h^{'},r,t^{'}) \in S^- \\ \xi_{h,t} \geq 0 \end{cases} \tag{2}$$

where S^+, S^- are the set of positive and negative samples. The loss function that satisfies the conditions (a) ($\gamma_1 = 0$) and (b) ($\gamma_1 > 0$) is:

$$\begin{aligned} \mathcal{L}_{a|b} = \sum_{(h,r,t)\in S^+} & \lambda_1 \|f_r(h,t) - \gamma_1\| + \\ & \lambda_2\ \max(\gamma_2 - f_r(h^{'},t^{'}), 0). \end{aligned} \tag{3}$$

Condition (c) considers a triple to be positive if its residual vector lies inside a hyper-sphere with radius γ_1. The optimization problem that satisfies the condition (c) is as follows (Nayyeri et al., 2019):

$$\begin{cases} \min_{\xi_{h,t}} \sum_{(h,r,t)\in S^+} {\xi_{h,t}}^2 \\ \text{s.t.} \\ f_r(h,t) \leq \gamma_1,\ (h,r,t) \in S^+ \\ f_r(h^{'},t^{'}) \geq \gamma_2 - \xi_{h,t},\ (h^{'},r,t^{'}) \in S^- \\ \xi_{h,t} \geq 0 \end{cases} \tag{4}$$

The loss function that satisfies the condition (c) is as follows Nayyeri et al. (2019):

$$\begin{aligned} \mathcal{L}_c = \sum_{(h,r,t)\in S^+} & \lambda_1 \max(f_r(h,t) - \gamma_1, 0) + \\ & \lambda_2\ \max(\gamma_2 - f_r(h^{'},t^{'}), 0) \end{aligned} \tag{5}$$

Remark: The loss function which is defined in Zhou et al. (2017b) is slightly different from the loss 3. The former slides the margin while the later fixes the margin by inclusion of a lower-bound for the score of negative triples. The both losses put an upper-bound for scores of positive triples.

Condition (d) is similar to (c). But it provides different γ_1, γ_2 for each triples. Using the condition (d), there is not a unique region of truth for all positive triples, rather for each triple (h,r,t) and its corresponding negative samples $(h^{'},r,t^{'})$ there are triple specific region of truth and falsity. Margin ranking loss (Bordes et al., 2013) satisfies the condition (d). The loss is defined as:

$$\mathcal{L}_d = \sum \sum [f_r(h,t) + \gamma - f_r(h^{'},t^{'})]_+ \tag{6}$$

where $[x]_+ = \max(0,x)$. Considering the conditions (a), (b), (c) and (d), we investigate the limitations L1 ,..., L6. We prove that existing theories are invalid under some conditions. During the following investigations of the limitations, we assume that the relation vectors shouldn't be null because the null vector for relation results same embedding vectors for entities appeared in head and tail parts when conditions (a) is used.

Limitation L1: *Lemma 1:* Let assumption (a) holds, then TransE and TransComplEx cannot infer a reflexive relation pattern with non-zero relation vector. With assumptions (b), (c) and (d), however, this is not true anymore and the models can infer reflexive relation patterns

Proof: the proofs are provided in the supplementary material file.

Limitation L2: *Lemma 2:* 1) TransComplEx can infer a relation pattern which is neither reflexive nor irreflexive with condition (b), (c) and (d). 2) TransE cannot infer the relation pattern which is neither reflexive nor irreflexive.

Limitation L3: *Lemma 3*: 1) TransComplEx can infer symmetric patterns with condition (a), (b), (c) and (d). 2) TransE cannot infer symmetric patterns with condition (a) with non-zero vector for relation. 3) TransE can infer a relation pattern which is symmetric with conditions (b).

Proof: proof of 1), 2) and 3) are included in the supplementary material.

3) For TransE with condition (b), there is

$$\|\mathbf{h}+\mathbf{r}-\mathbf{t}\| = \gamma_1, \tag{7}$$

$$\|\mathbf{t}+\mathbf{r}-\mathbf{h}\| = \gamma_1. \tag{8}$$

The necessity condition for encoding symmetric relation is $\|\mathbf{h}+\mathbf{r}-\mathbf{t}\| = \|\mathbf{t}+\mathbf{r}-\mathbf{h}\|$. This implies $\|h\|cos(\theta_{h,r}) = \|t\|cos(\theta_{t,r})$. Let $h-t=u$, by definition we have $\|\mathbf{u}+\mathbf{r}\| = \gamma_1, \|\mathbf{u}-\mathbf{r}\| = \gamma_1$.

Let $\gamma_1 = \alpha\|r\|$. We have

$$\begin{cases} \|\mathbf{u}\|^2 + (1-\alpha^2)\|\mathbf{r}\|^2 = -2\langle \mathbf{u}, \mathbf{r}\rangle \\ \|\mathbf{u}\|^2 + (1-\alpha^2)\|\mathbf{r}\|^2 = 2\langle \mathbf{u}, \mathbf{r}\rangle \end{cases} \tag{9}$$

Regarding 9, there is

$\|\mathbf{u}\|^2 + (1-\alpha^2)\|\mathbf{r}\|^2 = -(|\mathbf{u}\|^2 + (1-\alpha^2)\|\mathbf{r}\|^2). \rightarrow \|\mathbf{u}\|^2 = (\alpha^2-1)\|\mathbf{r}\|^2.$

To avoid contradiction, $\alpha > 1$. If $\alpha > 1$ we have $\cos(\theta_{u,r}) = \pi/2$. Therefore, TransE can encode symmetric pattern with condition (b), if $\gamma_1 = \alpha\|r\|$ and $\alpha > 1$. Figure 2 shows different conditions for encoding symmetric relation.

Limitation L4: *Lemma 4:* 1) Let (a) holds. Limitation L4 holds for both TransE and TransComplEx. 2) Limitation L4 is not valid when assumptions (b), (c) and (d) hold.

Limitation L5: *Lemma 5:* 1) Under condition (a), the limitation L5 holds for both TransE and TransComplEx. 2) Under conditions (b), (c) and (d), L5 is not valid for both TransE and TransComplEx.

Limitation L6: *Lemma 6:* 1) With condition (a), the limitation L6 holds for both TransE and TransComplEx. 2) With conditions (b), (c) and (d), the limitation L6 doesn't hold for the models.

4.3 Encoding Relation Patterns in TransComplEx

Most of KGE models learn from triples. Recent work incorporates relation patterns such as transitive, symmetric on the top of triples to further improve performance of models. For example, ComplEx-NNE+AER Ding et al. (2018) encodes implication pattern in the ComplEx model. RUGE Guo et al. (2018) injects First Order Horn Clause rules in an embedding model. SimplE Kazemi & Poole (2018) captures symmetric, antisymmetric and inverse patterns by weight tying in the model. Inspired by Minervini et al. (2017) and considering the score function of TransComplEx, in this part, we derive formulae for equivalence, symmetric, inverse and implication to be used as regularization terms in the optimization problem. Therefore, the model incorporates different relation patterns to optimize the embeddings.

Symmetric: In order to encode symmetric relation r, the following should be held:

$$f_r(h,t) \Longleftrightarrow f_r(t,h),$$

Therefore the following algebraic formulae is proposed to encode the relation: $\|f_r(h,t) - f_r(t,h)\| = 0$. According to the definition of score function of TransComplEx, we have the following algebraic formulae: $\mathcal{R}_{\mathcal{S}} = \|Re(\mathbf{h}) - Re(\mathbf{t})\| = 0$. Using similar argument for symmetric, the following formulae are derived for transitive, composition, inverse and implication:

Equivalence: Let p, q be equivalence relations i.e., $f_p(h,t) \Longleftrightarrow f_q(h,t)$. we obtain $\mathcal{R}_{\mathcal{E}} = \|\mathbf{p} - \mathbf{q}\| = 0$.

Implication: Let $p \rightarrow q$, we obtain $\mathcal{R}_{\mathcal{I}} = \max(f_p(h,t) - f_q(h,t), 0) = 0$.

Inverse: Let $r \longleftrightarrow r^{-1}$, we obtain $\mathcal{R}_{In} = \|\mathbf{r} - \mathbf{r}^{-1}\|$.

Finally, the following optimization problem should be solved:

$$\min_{\theta} \ \mathcal{L} + \sum \eta_i \mathcal{R}_i \tag{10}$$

where θ is embedding parameters, $\mathcal{L}$ is one of the losses 3, 5 or 6 and $\mathcal{R}$ is one of the derived formulae mentioned above.

5 EXPERIMENTS AND EVALUATIONS

In this section, we evaluate performance of our model, TransComplEx, with different loss functions on link prediction task. The aim of the task is to complete the triple $(h, r, ?)$ $((?, r, t))$ by prediction of the missed entity h or t. Filtered Mean Rank (MR), Mean Reciprocal Rank (MRR) and Hit@10 are used for evaluations (Wang et al., 2017; Lin et al., 2015b).

Dataset. We use two dataset extracted from Freebase (Bollacker et al., 2008) (i.e., FB15K (Bordes et al., 2013) and FB15K-237 (Toutanova & Chen, 2015)) and two others extracted from WordNet (Miller, 1995) (i.e. WN18 (Bordes et al., 2013) and WN18RR (Dettmers et al., 2018)). FB15K and WN18 are earlier dataset which have been extensively used to compare performance of KGEs. FB15K-237 and WN18RR are two dataset which are supposed to be more challenging after removing inverse patterns from FB15K and WN18. Guo et al. (2018) and Ding et al. (2018) extracted different relation patterns from FB15K and WN18 respectively. The relation patterns are provided by their confidence level, e.g. $(a, BornIn, b) \xrightarrow{0.9} (a, Nationality, b)$. We drop the relation patterns with confidence level less than 0.8. Generally, we use 454 and 14 relation patterns for FB15K and WN18 respectively. We do grounding for symmetric and transitive relation patterns. Thanks to the formulation of score function, grounding is not needed for inverse, implication and equivalence.

Experimental Setup. We implement TransComplEx with the losses 3, 5 and 6 and TransE with the loss 5 in Pytorch. Adagrad is used as an optimizer. We generate 100 mini-batches in each iteration. The hyperparameter corresponding to the score function is embedding dimension d. We add slack variables to the losses 3 and 5 to have soft margin as in (Nayyeri et al., 2019). The loss 5 is rewritten as follows Nayyeri et al. (2019):

$$\min_{\xi^r_{h,t}} \sum_{(h,r,t)\in S^+} \lambda_0 {\xi^r_{h,t}}^2 + \lambda_1 \max(f_r(h,t) - \gamma_1, 0) + \lambda_2 \max(\gamma_2 - f_r(h^{'}, t^{'}) - \xi^r_{h,t}, 0) \tag{11}$$

We set λ_1 and λ_2 to one and search for the hyperparameters $\gamma_1(\gamma_2 > \gamma_1)$ and λ_0 in the sets $\{0.1, 0.2, \ldots, 2\}$ and $\{0.01, 0.1, 1, 10, 100\}$ respectively. Moreover, we generate $\alpha \in \{1, 2, 5, 10\}$ negative samples per each positive. The embedding dimension and learning rate are tuned from the sets $\{100, 200\}, \{0.0001, 0.0005, 0.001, 0.005, 0.01\}$ respectively. All hyperparameters are adjusted by early stopping on validation set according to MRR. RPTransComplEx# denotes the TransComplEx model which is trained by the loss function # (3, 5, 6). RP indicates that relation patterns are injected during learning by regularizing the derived formulae (see 10). TransComplEx# refers to our model trained with the loss # without regularizing relation patterns formulae. The same notation is used for TransE#. The optimal configurations for RPTransComplEx3 are $d = 200, \lambda_0 = 100, \gamma_1 = 0.4, \gamma_2 = 0.5, \alpha = 10$ for FB15K, $d = 200, \lambda_0 = 100, \gamma_1 = 1.5, \gamma_2 = 2, \alpha = 10$ for FB15K-237, $d = 200, \lambda_0 = 100, \gamma_1 = 1, \gamma_2 = 2, \alpha = 10$ for WN18; for RPTransComplEx5 are $d = 200, \lambda_0 = 10, \gamma_1 = 0.4, \gamma_2 = 0.5, \alpha = 10$ for FB15K, $d = 200, \lambda_0 = 100, \gamma_1 = 1.5, \gamma_2 = 2, \alpha = 10$ for FB15K-237, $d = 200, \lambda_0 = 100, \gamma_1 = 0.6, \gamma_2 = 1.7, \alpha = 2$ for WN18; for RPTransComplEx6 are $d = 200, \gamma = 5, \alpha = 10$ for FB15K, $d = 200, \gamma = 10, \alpha = 10$ for FB15K-237, $d = 200, \gamma = 10, \alpha = 10$ for WN18; for TransComplEx5 are $d = 200, \lambda_0 = 10, \gamma_1 = 0.4, \gamma_2 = 0.5, \alpha = 10$ for FB15K, $d = 200, \lambda_0 = 100, \gamma_1 = 1.5, \gamma_2 = 2, \alpha = 10$ for FB15K-237, $d = 200, \lambda_0 = 100, \gamma_1 = 0.6, \gamma_2 = 1.7, \alpha = 2$ for WN18, $d = 200, \lambda_0 = 1, \gamma_1 = 1.6, \gamma_2 = 2.7, \alpha = 2$ for WN18RR, for TransE5 are $d = 200, \lambda_0 = 10, \gamma_1 = 0.4, \gamma_2 = 0.5, \alpha = 10$ for FB15K, $d = 200, \lambda_0 = 100, \gamma_1 = 0.4, \gamma_2 = 0.5, \alpha = 10$ for FB15K-237, $d = 200, \lambda_0 = 1, \gamma_1 = 1, \gamma_2 = 2, \alpha = 10$ for WN18, $d = 200, \lambda_0 = 1, \gamma_1 = 0.6, \gamma_2 = 1.7, \alpha = 2$ for WN18RR.

Results. Table 2 presents comparison of TransComplEx and its relation pattern encoded variants (RPTransComplEx) with three classes of embedding models including Translation-based models (e.g. TransX, TorusE), relation pattern encoded models (e.g. RUGE, ComplEx-NNE+AER, SimplE, SimplE+), and other state-of-the-art embedding models (e.g. ConvE, ComplEx, ANALOGY). To investigate our theoretical proofs corresponding to the effect of *loss* function, we train TransComplEx with different loss functions. As previously discussed, FB15K-237 and WN18RR are two

	FB15k			WN18		
	MR	MRR	Hits @10	MR	MRR	Hits @10
TransE (Bordes et al., 2013)	125	-	47.1	251	-	89.2
TransH (bern) (Wang et al., 2014)*	87	-	64.4	388	-	82.3
TransR (bern) (Lin et al., 2015b)*	77	-	68.7	225	-	92.0
TransD (bern) (Ji et al., 2015)*	91	-	77.3	**212**	-	92.2
TransE-RS (bern) (Zhou et al., 2017a)*	63	-	72.1	371	-	93.7
TransH-RS (bern) (Zhou et al., 2017a)*	77	-	75.0	357	-	94.5
TorusE (Ebisu & Ichise, 2019)	-	73.3	83.2	-	94.7	95.4
TorusE(with WNP) (Ebisu & Ichise, 2019)	-	**75.1**	83.5	-	**94.7**	95.4
R-GCN (Schlichtkrull et al., 2018)+	-	65.1	82.5	-	81.4	**95.5**
ConvE (Dettmers et al., 2018)++	51	68.9	85.1	504	94.2	**95.5**
ComplEx (Trouillon et al., 2016)++	106	67.5	82.6	543	94.1	94.7
ANALOGY (Liu et al., 2017)++	121	72.2	84.3	-	94.2	94.7
RotatE (Sun et al., 2019)	48	69.0	86.1	433	**94.8**	**95.5**
SimplE (Kazemi & Poole, 2018)	-	72.7	83.8	-	94.2	94.7
SimplE+ (Fatemi et al., 2018)	-	72.5	84.1	-	93.7	93.9
PTransE (Lin et al., 2015a)	58	-	84.6	-	-	-
KALE (Guo et al., 2016)	73	52.3	76.2	241	53.2	94.4
RUGE (Guo et al., 2018)	97	76.8	86.5	-	-	-
ComplEx-NNE+AER (Ding et al., 2018)	116	80.3	87.4	450	94.3	94.8
RPTransComplEx3	**38**	70.5	88.3	451	92.7	94.8
RPTransComplEx5	**38**	72.4	**88.8**	275	92.4	95.4
RPTransComplEx6	59	61.7	82.2	547	94.0	94.7
TransComplEx5	**38**	68.2	87.5	284	92.2	**95.5**
TransE5	46	64.8	87.2	703	68.7	94.5

Table 2: Link prediction results. Rows 1-8: Translation-based models with no injected relation patterns. Rows 9-13: basic models with no injected relation patterns. Rows 14-18: models which encode relation patterns. Results labeled with *, + and ++ are taken from Zhou et al. (2017a), Ebisu & Ichise (2019) and Akrami et al. (2018) while the rest are taken from original papers/code. Dashes: results could not be obtained.

	FB15k-237			WN18RR		
	MR	MRR	Hits @10	MR	MRR	Hits @10
TransE (Bordes et al., 2013)+	-	25.7	42.0	-	18.2	44.4
DistMult (Bordes et al., 2013)+	-	24.1	41.9	-	43.0	49.0
ComplEx (Trouillon et al., 2016)+	-	24.0	41.9	-	44.0	51.0
R-GCN (Schlichtkrull et al., 2018)+	-	24.8	41.7	-	-	-
ConvE (Dettmers et al., 2018)+	-	31.6	49.1	-	46.0	48.0
TorusE (Ebisu & Ichise, 2019)	-	30.5	48.4	-	45.2	51.2
TorusE (with WNP) (Ebisu & Ichise, 2019)	-	30.7	48.5	-	46.0	53.4
RotatE (Sun et al., 2019)	211	31.1	49.4	4789	**47.3**	**54.9**
RPTransComplEx3	210	27.7	46.4	-	-	-
RPTransComplEx5	226	**31.9**	**49.5**	-	-	-
RPTransComplEx6	216	25.3	43.8	-	-	-
TransComplEx5	223	31.7	49.3	4081	38.9	49.8
TransE5	**205**	27.2	45.3	3850	20.0	47.5

Table 3: Link prediction results. Rows 1-8: basic models with no injected relation patterns. Results labeled with + are taken from Ebisu & Ichise (2019) while the rest are taken from original papers/code. Dashes: results could not be obtained.

more challenging dataset provided recently. Therefore, in order to have a better evaluation, Table 3 presents comparison of our models with state-of-the-art embedding methods on these two dataset. For WN18RR, we do not use any relation patterns to be encoded. The results labeled with "*", "+" and "++" are taken from Zhou et al. (2017a), Ebisu & Ichise (2019) and Akrami et al. (2018) respectively. To have a fair comparison, we ran the code of RotatE (Sun et al., 2019) in our setting e.g., embedding dimension 200 and 10 negative samples while the original paper reported the results of RotatE using a very big embedding dimension and a lot of negative samples (embedding dimension 1000 and 1000 negative samples).

Boosting techniques: There are several ways to improve the performance of embedding models: 1) designing a more sophisticated scoring function, 2) proper selection of loss function, 3) using more negative samples 4) using negative sampling techniques, 5) enriching dataset (e.g., adding reverse triples). Among the mentioned techniques, we focus on the first and second ones and avoid using other techniques. We keep the setting used in Trouillon et al. (2016) to have a fair comparison. Using other techniques can further improve the performance of every models including ours. For example, TransComplEx with embedding dimension 200 and 50 negative samples gets 52.2 for Hits@10.

Dissuasion of Results. According to the Table 2, FB15K dataset part, PRTransComplEx trained by the loss 5 significantly outperforms all Translation-based embedding models including the recent work TorusE. Note that TorusE is trained by embedding dimension 10000 while our model uses embedding dimension at most 200. Comparing to relation pattern encoded embedding models including recent works ComplEx-NNE+AER, RUGE, SimplE and SimplE+, our model outperforms them in the terms of MR and Hit@10. Moreover, the model significantly outperforms popular embedding models including ConvE and ComplEx. Regarding our theories, the loss 5 has less limitations comparing to the loss 3. This is consistent with our theories where RPTransComplEx5 outperforms RPTransComplEx3. TransComplEx without encoding relation patterns still obtains accuracy as good as state-of-the-art models. TransComplEx outperforms TransE while both are trained by the loss 5 in the terms of MR, MRR and Hit@10 which is consistent with our theories (TransComplEx score function has less limitations than TransE). Regarding the results on WN18, the accuracy of TransComplEx is very close to the state-of-the-art models. Encoding relation patterns cannot improve the performance on WN18 because the models learn relation patterns from data well. The loss 6 provides different upper-bounds and lower-bounds for the score of positive and negative triples respectively and also the margin can slide. Therefore, the accuracy would be degraded Zhou et al. (2017a). Generally, the loss 5 gets better performance which is consistent to our theoretical results. As shown in the Table 3, FB15K-237 part, with and without encoding relation patterns, TransComplEx trained by the loss 5 outperforms all the baselines in terms of MRR and Hit@10. TransComplEx5 outperforms TransE5 showing the effectiveness of our proposed score function. Regarding WN18RR, TorusE has better performance comparing to our model. However, the results are obtained with a very big embedding dimension ($d = 10000$).

6 Conclusion

In this paper, we reinvestigated the main limitations of Translation-based embedding models from two aspects: *score* and *loss*. We showed that existing theories corresponding to the limitations of the models are inaccurate because the effect of loss functions has been ignored. Accordingly, we presented new theories about the limitations by consideration of the effect of score and loss functions. We proposed TransComplEx, a new variant of TransE which is proven to be less limited comparing to the TransE. The model is trained by using various loss functions on standard dataset including FB15K, FB15K-237, WN18 and WN18RR. According to the experiments, TransComplEx with proper loss function significantly outperformed translation-based embedding models. Moreover, TransComplEx got competitive performance comparing to the state-of-the-art embedding models while it is more efficient in time and memory. The experimental results conformed the presented theories corresponding to the limitations.

REFERENCES

Farahnaz Akrami, Lingbing Guo, Wei Hu, and Chengkai Li. Re-evaluating embedding-based knowledge graph completion methods. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pp. 1779–1782. ACM, 2018.

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

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

Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2d knowledge graph embeddings. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

Boyang Ding, Quan Wang, Bin Wang, and Li Guo. Improving knowledge graph embedding using simple constraints. *arXiv preprint arXiv:1805.02408*, 2018.

Takuma Ebisu and Ryutaro Ichise. Toruse: Knowledge graph embedding on a lie group. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

Takuma Ebisu and Ryutaro Ichise. Generalized translation-based embedding of knowledge graph. *IEEE Transactions on Knowledge and Data Engineering*, 2019.

Bahare Fatemi, Siamak Ravanbakhsh, and David Poole. Improved knowledge graph embedding using background taxonomic information. *arXiv preprint arXiv:1812.03235*, 2018.

Jun Feng, Minlie Huang, Mingdong Wang, Mantong Zhou, Yu Hao, and Xiaoyan Zhu. Knowledge graph embedding by flexible translation. In *Fifteenth International Conference on the Principles of Knowledge Representation and Reasoning*, 2016.

Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. Jointly embedding knowledge graphs and logical rules. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 192–202, 2016.

Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. Knowledge graph embedding with iterative guidance from soft rules. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

Frank L Hitchcock. The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, 6(1-4):164–189, 1927.

Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun 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)*, volume 1, pp. 687–696, 2015.

Seyed Mehran Kazemi and David Poole. Simple embedding for link prediction in knowledge graphs. In *Advances in Neural Information Processing Systems*, pp. 4284–4295, 2018.

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

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

Hanxiao Liu, Yuexin Wu, and Yiming Yang. Analogical inference for multi-relational embeddings. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 2168–2178. JMLR. org, 2017.

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

Pasquale Minervini, Luca Costabello, Emir Munoz, Novacek, and Pierre-Yves Vandenbussche. Regularizing knowledge graph embeddings via equivalence and inversion axioms. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 668–683. Springer, 2017.

Mojtaba Nayyeri, Sahar Vahdati, Jens Lehmann, and Hamed Shariat Yazdi. Soft marginal transe for scholarly knowledge graph completion. *arXiv preprint arXiv:1904.12211*, 2019.

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

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *European Semantic Web Conference*, pp. 593–607. Springer, 2018.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. *arXiv preprint arXiv:1902.10197*, 2019.

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

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

Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12): 2724–2743, 2017.

Yanjie Wang, Rainer Gemulla, and Hui Li. On multi-relational link prediction with bilinear models. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph embedding by translating on hyperplanes. In *Twenty-Eighth AAAI conference on artificial intelligence*, 2014.

Xiaofei Zhou, Qiannan Zhu, Ping Liu, and Li Guo. Learning knowledge embeddings by combining limit-based scoring loss. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 1009–1018. ACM, 2017a.

Xiaofei Zhou, Qiannan Zhu, Ping Liu, and Li Guo. Learning knowledge embeddings by combining limit-based scoring loss. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 1009–1018. ACM, 2017b.

A SUPPLEMENTARY MATERIAL

The proof of lemmas are provided as follows:

Lemma 1: Let assumption (a) holds, then TransE and TransComplEx cannot infer a reflexive relation pattern with non-zero relation vector. With assumptions (b), (c) and (d), however, this is not true anymore and the models can infer reflexive relation patterns

Proof 1) Let r be a reflexive relation and condition a) holds. For TransE, we have

$$\mathbf{h} + \mathbf{r} - \mathbf{h} = \mathbf{0}. \tag{12}$$

Therefore, the relation vector collapses to a null vector ($\mathbf{r} = \mathbf{0}$). As a consequence of $\mathbf{r} = \mathbf{0}$, embedding vectors of head and tail entities will be same which is undesired. Therefore, TransE cannot infer reflexive relation with $\mathbf{r} \neq \mathbf{0}$.

For TransComplEx, we have

$$\mathbf{h} + \mathbf{r} - \bar{\mathbf{h}} = \mathbf{0}. \tag{13}$$

We have

$$\begin{aligned} Re(\mathbf{r}) &= \mathbf{0}, \\ Im(\mathbf{r}) &= -2Im(\mathbf{h}). \end{aligned} \tag{14}$$

Therefore, all entities will have same embedding vectors which is undesired.

2) Using condition (b), we have

$$\|\mathbf{h} + \mathbf{r} - \mathbf{t}\| = \gamma_1.$$

It gives $\|\mathbf{r}\| = \gamma_1$. Therefore, in order to infer reflexive relation, the length of the relation vector should be γ_1. Consequently, TransE and TransComplEx can infer reflexive relation. The same procedure can be used for the conditions (c) and (d).

Lemma 2: 1) Let the assumption b) or c) or d) holds. TransComplEx can infer a relation pattern which is neither reflexive nor irreflexive. 2) TransE cannot infer the relation pattern.

proof: 1) Let the relation r be neither reflexive nor irreflexive and two triples $(e_1, r, e_1), (e_2, r, e_2)$ be positive and negative respectively. Therefore the following inequalities hold:

$$\begin{cases} \|\mathbf{e}_1 + \mathbf{r} - \bar{\mathbf{e}_1}\| \leq \lambda_1, \\ \|\mathbf{e}_2 + \mathbf{r} - \bar{\mathbf{e}_2}\| \geq \lambda_2. \end{cases} \tag{15}$$

Equation 15 is rewritten as follows:

$$\begin{aligned} \|Re(\mathbf{r}) + i(Im(\mathbf{r}) + 2Im(\mathbf{e}_1))\| &\leq \gamma_1, \\ \|Re(\mathbf{r}) + i(Im(\mathbf{r}) + 2Im(\mathbf{e}_2))\| &\geq \gamma_2, \end{aligned} \tag{16}$$

For TransE in real space, $\|Re(\mathbf{r})\| \leq \gamma_1$ and $\|Re(\mathbf{r})\| \geq \gamma_2$ cannot be held simultaneously when $\gamma_2 > \gamma_1$. Therefore, TransE in real space cannot encode a relation which is neither reflexive nor irreflexive. In contrast, TransE in complex space can encode the relation by proper assignment of imaginary parts of entities. Therefore, theoretically TransComplEx can infer a relation which is neither reflexive nor irreflexive.

Lemma 3: 1) TransComplEx can infer symmetric patterns with condition a), b), c) and d). 2) TransE cannot infer symmetric patterns with condition a) with non-zero vector for relation. 3) TransE can infer a relation pattern which is symmetric and reflexive with conditions b), c) and d).

Proof: 1), 2) Let r be a symmetric relation and a) holds. We have

$$\begin{aligned} \mathbf{h} + \mathbf{r} &= \bar{\mathbf{t}}, \\ \mathbf{t} + \mathbf{r} &= \bar{\mathbf{h}}. \end{aligned} \tag{17}$$

Trivially, we have

$$\begin{aligned} Re(\mathbf{h}) + Re(\mathbf{r}) &= Re(\mathbf{t}), \\ Re(\mathbf{t}) + Re(\mathbf{r}) &= Re(\mathbf{h}), \\ Im(\mathbf{h}) + Im(\mathbf{r}) &= -Im(\mathbf{t}), \\ Im(\mathbf{t}) + Im(\mathbf{r}) &= -Im(\mathbf{h}), \end{aligned} \tag{18}$$

For TransE in real space, there is

$$\begin{aligned} Re(\mathbf{h}) + Re(\mathbf{r}) &= Re(\mathbf{t}), \\ Re(\mathbf{t}) + Re(\mathbf{r}) &= Re(\mathbf{h}), \end{aligned}$$

Therefore, $Re(\mathbf{r}) = \mathbf{0}$. It means that TransE cannot infer symmetric relations with condition a). For TransComplEx, additionally we have

$$\begin{aligned} Im(\mathbf{h}) + Im(\mathbf{r}) &= -Im(\mathbf{t}), \\ Im(\mathbf{t}) + Im(\mathbf{r}) &= -Im(\mathbf{h}), \end{aligned}$$

It concludes $Im(\mathbf{h}) + Im(\mathbf{r}) + Im(\mathbf{t}) = \mathbf{0}$. Therefore, TransE in complex space with condition a) can infer symmetric relation. Because a) is an special case of b) and c), TransComplEx can infer symmetric relations in all conditions.

3) For TransE with condition b), there is

$$\|\mathbf{h} + \mathbf{r} - \mathbf{t}\| = \gamma_1, \tag{19}$$

$$\|\mathbf{t} + \mathbf{r} - \mathbf{h}\| = \gamma_1. \tag{20}$$

The necessity condition for encoding symmetric relation is $\|\mathbf{h} + \mathbf{r} - \mathbf{t}\| = \|\mathbf{t} + \mathbf{r} - \mathbf{h}\|$. This implies $\|h\| cos(\theta_{h,r}) = \|t\| cos(\theta_{t,r})$. Let $h - t = u$, by 20 we have $\|\mathbf{u} + \mathbf{r}\| = \gamma_1$, $\|\mathbf{u} - \mathbf{r}\| = \gamma_1$.

Let $\gamma_1 = \alpha \|r\|$. We have

$$\begin{cases} \|\mathbf{u}\|^2 + (1 - \alpha^2)\|\mathbf{r}\|^2 = -2\langle \mathbf{u}, \mathbf{r} \rangle \\ \|\mathbf{u}\|^2 + (1 - \alpha^2)\|\mathbf{r}\|^2 = 2\langle \mathbf{u}, \mathbf{r} \rangle \end{cases} \tag{21}$$

Regarding 21, we have

$\|\mathbf{u}\|^2 + (1 - \alpha^2)\|\mathbf{r}\|^2 = -(|\mathbf{u}\|^2 + (1 - \alpha^2)\|\mathbf{r}\|^2)$.

$\rightarrow \|\mathbf{u}\|^2 = (\alpha^2 - 1)\|\mathbf{r}\|^2$.

To avoid contradiction, $\alpha \geq 1$. If $\alpha \geq 1$ we have $\cos(\theta_{u,r}) = \pi/2$. Therefore, TransE can encode symmetric pattern with condition b), if $\gamma_1 = \alpha \|r\|$ and $\alpha \geq 1$. From the proof of condition b), we conclude that TransE can encode symmetric patterns under conditions c) and d).

Lemma 4: 1) Let a) holds. Limitation L4 holds for both TransE and TransComplEx. 2) Limitation L4 is not valid when assumptions b), c) and d) hold.

Proof: 1) The proof of the lemma with condition a) for TransE is mentioned in the paper Kazemi & Poole (2018). For TransComplEx, the proof is trivial. 2) Now, we prove that the limitation L4 is not valid when b) holds.

Let condition b) holds and relation r be reflexive, we have $\|\mathbf{e}_1 + \mathbf{r} - \mathbf{e}_1\| = \gamma_1$, $\|\mathbf{e}_2 + \mathbf{r} - \mathbf{e}_2\| = \gamma_1$.

Let $\|\mathbf{e}_1 + \mathbf{r} - \mathbf{e}_2\| = \gamma_1$. To violate the limitation L4, the triple (e_2, r, e_1) should be negative i.e.,

$\|\mathbf{e}_2 + \mathbf{r} - \mathbf{e}_1\| > \gamma_1$,

$\rightarrow \|\mathbf{e}_2 + \mathbf{r} - \mathbf{e}_1\|^2 > \gamma_1^2$,

$\rightarrow \|\mathbf{e}_2\|^2 + \|\mathbf{e}_1\|^2 + \|\mathbf{r}\|^2 + 2 < \mathbf{e}_2, \mathbf{r} > -2 < \mathbf{e}_2, \mathbf{e}_1 > -2 < \mathbf{e}_1, \mathbf{r} > > \gamma_1^2$.

Considering $\|\mathbf{e}_1 + \mathbf{r} - \mathbf{e}_2\| = \gamma_1$, we have

$< \mathbf{e}_2, \mathbf{r} > - < \mathbf{e}_1, \mathbf{r} > > 0,$
$\rightarrow < \mathbf{e}_2 - \mathbf{e}_1, \mathbf{r} > > 0,$
$\rightarrow \ cos(\theta_{(\mathbf{e}_2 - \mathbf{e}_1), \mathbf{r}}) \ > 0,$

Therefore, the limitation L4 is not valid i.e., if a relation r is reflexive, it may not be symmetric. TransE is special case of TransComplEx and also condition b) is special case of condition c). Therefore using conditions b), c) and d), the limitation L4 is not valid for TransE and TransComplEx.

Lemma 5: 1) Under condition a), the limitation L5 holds for both TransE and TransComplEx. 2) Under conditions b), c) and d), L5 is not valid for both TransE and TransComplEx.

proof

1) Under condition a), equation $\mathbf{h} + \mathbf{r} - \mathbf{t} = \mathbf{0}$ holds. Therefore, according to the paper Kazemi & Poole (2018), the model has the limitation L5.

2) If a relation is reflexive, with condition b), we have $\|\mathbf{e}_1 + \mathbf{r} - \mathbf{e}_1\| = \gamma_1, \|\mathbf{e}_2 + \mathbf{r} - \mathbf{e}_2\| = \gamma_1$. Therefore, $\|r\| = \lambda_1$. Let

$$\begin{cases} \|\mathbf{e}_1 + \mathbf{r} - \mathbf{e}_2\| = \gamma_1, \\ \|\mathbf{e}_2 + \mathbf{r} - \mathbf{e}_3\| = \gamma_1. \end{cases} \tag{22}$$

we need to show the following inequality wouldn't give contradiction: $\|\mathbf{e}_2 + \mathbf{r} - \mathbf{e}_3\| > \gamma_1$.

From 22 we have $< \mathbf{e}_2, (\mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3) >< 0$, which is not contradiction.

Therefore, with conditions b) and c), the limitation L5 is not valid for both TransE and TransComplEx.

Limitation L6: *Lemma 6:* 1) With condition (a), the limitation L6 holds for both TransE and TransComplEx. 2) With conditions (b), (c) and (d), the limitation L6 doesn't hold for the models.

Proof: 1) With condition (a), the limitation L6 is proved in Kazemi & Poole (2018). 2) Considering the assumption of L6 and the condition (b), we have

$$\begin{cases} \|\mathbf{e}_1 + \mathbf{r} - \mathbf{s}_1\| = \gamma_1, \\ \|\mathbf{e}_1 + \mathbf{r} - \mathbf{s}_2\| = \gamma_1. \\ \|\mathbf{e}_2 + \mathbf{r} - \mathbf{s}_1\| = \gamma_1. \end{cases} \tag{23}$$

We show the condition that $\|\mathbf{e}_2 + \mathbf{r} - \mathbf{s}_2\| > \gamma_1$ holds.

Substituting 23 in $\|\mathbf{e}_2 + \mathbf{r} - \mathbf{s}_2\| > \gamma_1$, we have

$cos(\theta_{(s_1 - s_2), (e_1 - e_2)}) < 0$. Therefore, there are assignments to embeddings of entities that the limitation L6 is not valid with condition (b), (c) and (d).

Figure 3 shows that the limitation L6 is invalid by proper selection of loss function.

A.1 Further limitations and future work

In the paper, we have investigated the six limitations of TransE which are resolved by revision of loss function. However, revision of loss functions can resolve further limitations including 1-N, N-1 and M-N relations. More concretely, setting upper-bound for the scores of positive samples can mitigate the M-N problem. We will leave it as future work.

Our theories can be extended to every distance-based embedding models including RotatE etc.

Moreover, the negative likelihood loss has been shown to be effective for training different embedding models including RotatE and TransE. This can also be explained by reformulation of negative likelihood loss as standard optimization problem, showing the the loss put a boundary for the score functions.

We will consider the mentioned points as future work.

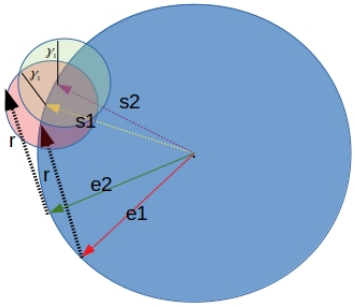

Figure 3: Investigation of L6 with condition (c): The limitation is not valid, because the triple (e_2, r, s_2) can get an score to be considered as negative while triples $((e_1, r, s_1), (e_1, r, s_2), (e_2, r, s_1))$ are positive.