From One Point to A Manifold:
Knowledge Graph Embedding For Precise Link Prediction

Han Xiao[1], Minlie Huang[1], Hao Yu[1], Xiaoyan Zhu[1]

[1]Department of Computer Science and Technology, State Key Lab on Intelligent Technology and Systems,
National Lab for Information Science and Technology, Tsinghua University, Beijing, China

Abstract

Knowledge graph embedding aims at offering a numerical knowledge representation paradigm by transforming the entities and relations into continuous vector space. However, existing methods could not characterize the knowledge graph in a fine degree to make a precise prediction. There are two reasons: being an ill-posed algebraic system and applying an overstrict geometric form. As precise prediction is critical, we propose an manifold-based embedding principle (**ManifoldE**) which could be treated as a well-posed algebraic system that expands the position of golden triples from one point in current models to a manifold in ours. Extensive experiments show that the proposed models achieve substantial improvements against the state-of-the-art baselines especially for the precise prediction task, and yet maintain high efficiency.

1 Introduction

Knowledge is critical to artificial intelligence, and the embedded representation of knowledge offers an efficient basis of computing over symbolic knowledge facts. More specifically, knowledge graph embedding projects entities and relations into a continuous high-dimension vector space by optimizing well-defined objective functions. A variety of methods have been proposed for this task, including TransE [Bordes *et al.*, 2013], PTransE [Lin *et al.*, 2015a] and KG2E [He *et al.*, 2015].

A fact of knowledge graph is usually represented by a triple **(h,r,t)**, where **h,r,t** indicate the embedded vectors of the head entity, the relation and the tail entity, respectively. As a key branch of embedding methods, translation-based methods, such as TransE [Bordes *et al.*, 2013], PTransE [Lin *et al.*, 2015a] and KG2E [He *et al.*, 2015], treat the triple as a relation-specific translation from the head entity to the tail entity.

Despite the success of previous methods, none of previous studies has addressed the issue of ***precise link prediction, which finds the exact entity given another entity and the relation***. For a specific query fact, most existing methods would

extract a few candidate entities that may contain correct answers, but there is no mechanism to ensure that the correct answers rank ahead the candidate list.

Generally speaking, precise link prediction would improve the feasibility of knowledge completion, the effectiveness of knowledge reasoning, and the performance of many knowledge-related tasks. Taking knowledge completion as example, when we want to know about the birth place of Martin R.R., what we expect is the exact answer "U.S.", while a few other candidates do not make any sense.

The issue of precise link prediction is caused by two reasons: *the ill-posed algebraic system* and *the over-strict geometric form*. First, from the algebraic perspective, each fact can be treated as an equation of $\mathbf{h_r} + \mathbf{r} = \mathbf{t_r}$[1] if following the translation-based principle, embedding could be treated as a solution to the equation group. In current embedding methods, the number of equations is more than that of free variables, which is called an ill-posed algebraic problem as defined in [Tikhonov and Arsenin, 1978]. More specifically, $\mathbf{h_r} + \mathbf{r} = \mathbf{t_r}$ indicates d equations as $h_i + r_i = t_i$ where d is the dimension of embedding vector and i denotes each dimension. Therefore, there are Td equations where T is the number of facts, while the number of variables are $(E + R)d$, where E, R are the number of entities and relations, respectively. As is the case that triples are much more than the sum of entities and relations, the number of variables are much less than the number of equations, which is typically an ill-posed algebraic system. Mathematically, an ill-posed algebraic system would commonly make the solutions imprecise and unstable. In this paper, we propose to address this issue by replacing the translation-based principle $\mathbf{h_r} + \mathbf{r} = \mathbf{t_r}$ by a manifold-based principle $\mathcal{M}(\mathbf{h}, \mathbf{r}, \mathbf{t}) = D_r^2$ where $\mathcal{M}$ is the manifold function. With the manifold-based principle, our model can make a nearly well-posed algebraic system by taking $d \geq \frac{T}{E+R}$ so that the number of equations (T) is no more than that of the free parameters ($(E + R)d$).

Second, from the geometric perspective, the position of the golden facts in existing methods is almost one point, which is too strict for all relations and is more insufficient for complex relations such as many-to-many relations. For example, for entity *American Revolution*, there exist many triples such as (American Revolution, HasPart, Battle Bunker Hill),

---

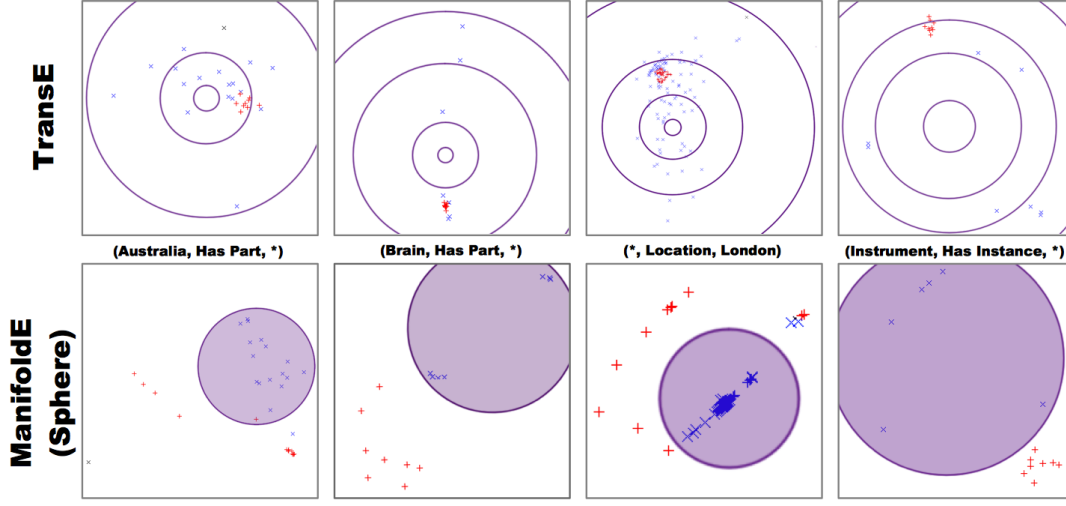[1]h_r and t_r are the relation-specific entity embeddings

Figure 1: The visualization comparison of TransE and ManifoldE(Sphere). For ManifoldE, the manifold collapsed to a solid circle by dimension reduction. The data are selected from Wordnet and Freebase. The blue crosses mean the correctly matched entities while the red crosses indicate the unmatched ones. The upper block corresponds to TransE where more near to the center , more plausible the triple is. It is clear that the true and false triples near the golden position is chaotically distributed. The lower block is ManifoldE (Sphere) where the triples inside the solid circle is matched and outside is unmatched. We can see that there are relatively less errors in ManifoldE than TransE.

(American Revolution, HasPart, Battle Cowpens). When many tail entities compete for only one point, there would be a major loss of objective function. Some previous work such as TransH [Wang *et al.*, 2014] and TransR [Lin *et al.*, 2015b] address this problem by projecting entities and relations into some relation-specific subspaces. However, in each subspace, the golden position is also one point and the over-strict geometric form is still existing. As can be seen from Fig.1, translation-based geometric principle involves too much noise. However, ManifoldE alleviates this issue by expanding the position of golden triples from one point to a manifold such as a high-dimensional sphere. By this mean, ManifoldE avoids much noise to distinguish the true facts from the most possible false ones, and improves the precision of knowledge embedding as Fig.1 shows.

To summarize, our contributions are two-fold: **(1.)**We have addressed the issue of ***precise link prediction*** and attributed this issue to two reasons: *the ill-posed algebraic system* and *the over-strict geometric form*. To our best knowledge, this is the first time to address this issue formally. **(2.)**We propose a manifold-based principle to alleviate this issue and design a new model, ManifoldE, which achieves remarkable improvements over the state-of-the-art baselines in experiments, especially for precise embedding tasks. Besides, our methods are also *very efficient*.

2 Related Work

2.1 Translation-based Methods

As a pioneering work of knowledge graph embedding, **TransE** [Bordes *et al.*, 2013] opens a line of translation-based methods. TransE treats a triple (h, r, t) as a relation-specific translation from a head entity to a tail entity, say $\mathbf{h} + \mathbf{r} = \mathbf{t}$

and the score function has a form of $||\mathbf{h} + \mathbf{r} - \mathbf{t}||_2^2$. Following this principle, a number of models have been proposed. For instance, **TransH** [Wang *et al.*, 2014] adopts a projection transformation, say $\mathbf{h_r} = \mathbf{h} - \mathbf{w_r}^\top \mathbf{h} \mathbf{w_r}$, $\mathbf{t_r} = \mathbf{t} - \mathbf{w_r}^\top \mathbf{t} \mathbf{w_r}$, while **TransR** [Lin *et al.*, 2015b] applies a rotation transformation, say $\mathbf{h_r} = \mathbf{M_r}\mathbf{h}, \mathbf{t_r} = \mathbf{M_r}\mathbf{t}$. Similar work also includes **TransD** [Ji *et al.*,] and **TransM** [Fan *et al.*, 2014]. Other approaches take into consideration extra information such as relation-type **[Wang *et al.*, 2015]**, paths with different confidence levels (**PTransE**) [Lin *et al.*, 2015a], and semantic smoothness of the embedding space **[Guo *et al.*, 2015]**. **KG2E** [He *et al.*, 2015] is a probabilistic embedding method for modeling the uncertainty in knowledge base. Notably, translation-based models demonstrate the state-of-the-art performance.

2.2 Other Methods

The **Unstructured Model (UM)** [Bordes *et al.*, 2012] is a simplified version of TransE, which ignores relation information and the score function is reduced to $f_r(h, t) = ||\mathbf{h} - \mathbf{t}||_2^2$. The **Structured Embedding (SE)** model [Bordes *et al.*, 2011] transforms the entity space with the head-specific and tail-specific matrices and the score function is defined as $f_r(h, t) = ||\mathbf{M_{h,r}}\mathbf{h} - \mathbf{M_{t,r}}\mathbf{t}||$. The **Semantic Matching Energy (SME)** model [Bordes *et al.*, 2012] [Bordes *et al.*, 2014] can enhance SE by considering the correlations between entities and relations with different matrix operators, as follows:

$$f_r(h, t) = (\mathbf{M_1}\mathbf{h} + \mathbf{M_2}\mathbf{r} + \mathbf{b_1})^\top (\mathbf{M_3}\mathbf{t} + \mathbf{M_4}\mathbf{r} + \mathbf{b_2})$$
$$f_r(h, t) = (\mathbf{M_1}\mathbf{h} \otimes \mathbf{M_2}\mathbf{r} + \mathbf{b_1})^\top (\mathbf{M_3}\mathbf{t} \otimes \mathbf{M_4}\mathbf{r} + \mathbf{b_2})$$

where $\mathbf{M_1}, \mathbf{M_2}, \mathbf{M_3}$ and $\mathbf{M_4}$ are weight matrices, $\otimes$ is the Hadamard product, $\mathbf{b_1}$ and $\mathbf{b_2}$ are bias vectors. The **Sin-**

gle Layer Model (SLM) applies neural network to knowledge graph embedding and the score function is defined as $f_r(h,t) = \mathbf{u_r}^\top g(\mathbf{M_{r,1}}\mathbf{h} + \mathbf{M_{r,2}}\mathbf{t})$ where $\mathbf{M_{r,1}}, \mathbf{M_{r,2}}$ are relation-specific weight matrices. The **Latent Factor Model (LFM)** [Jenatton *et al.*, 2012], [Sutskever *et al.*, 2009] makes use of the second-order correlations between entities by a quadratic form and the score function is as $f_r(h,t) = \mathbf{h}^\top\mathbf{W_r}\mathbf{t}$. The **Neural Tensor Network (NTN)** model [Socher *et al.*, 2013] defines a very expressive score function to combine the SLM and LFM: $f_r(h,t) = \mathbf{u_r}^\top g(\mathbf{h}^\top\mathbf{W}_{..r}\mathbf{t} + \mathbf{M_{r,1}}\mathbf{h} + \mathbf{M_{r,2}}\mathbf{t} + \mathbf{b_r})$, where $\mathbf{u_r}$ is a relation-specific linear layer, $g(\cdot)$ is the $tanh$ function, $\mathbf{W} \in \mathbb{R}^{d \times d \times k}$ is a 3-way tensor. Besides, **RESCAL** is a collective matrix factorization model which is also a common method in knowledge graph embedding [Nickel *et al.*, 2011], [Nickel *et al.*, 2012].

3 Methods

In this section, we introduce the novel manifold-based principle and then we analyze these methods from algebraic and geometric perspective.

3.1 ManifoldE : Manifold-Based Embedding Model

Instead of adopting the translation-based principle $\mathbf{h} + \mathbf{r} = \mathbf{t}$, we apply the manifold-based principle $\mathcal{M}(\mathbf{h}, \mathbf{r}, \mathbf{t}) = D_r^2$ for a specific triple (h, r, t). When a head entity and a relation are given, the tail entities lay in a high-dimensional manifold. Intuitively, our score function is designed by measuring the distance of the triple away from the manifold:

$$f_r(h,t) = ||\mathcal{M}(h,r,t) - D_r^2||^2 \tag{1}$$

where $\mathbf{h}, \mathbf{r}, \mathbf{t}$ are respectively the embeddings of *h, r, and t*. D_r is a relation-specific manifold parameter. $\mathcal{M} : \mathbb{E} \times \mathbb{L} \times \mathbb{E} \to \mathbb{R}$ is the manifold function, where $\mathbb{E}, \mathbb{L}$ are the entity set and relation set and $\mathbb{R}$ is the real number field.

Sphere. Sphere is a very typical manifold. In this setting, all the tail (or head) entities for a specific fact such as $(h, r, *)$ are supposed to lay in a high-dimensional sphere where $\mathbf{h} + \mathbf{r}$ is the center and D_r is the radius. Formally stated as:

$$\mathcal{M}(h,r,t) = ||\mathbf{h} + \mathbf{r} - \mathbf{t}||_2^2$$

Obviously, this is a straight-forward extension of translation-based model in which D_r is zero. From geometric perspective, the manifold collapses into a point when applying the translation-based principle.

RKHS (Reproducing Kernel Hilbert Space) usually provides a more expressive approach to represent the manifolds, which motivates us to apply the manifold-based principle with kernel. To this point, kernels are involved to lay the sphere in a Hilbert space (an implicit high-dimensional space) as below:

$$\begin{aligned}\mathcal{M}(h,r,t) &= ||\varphi(h) + \varphi(r) - \varphi(t)||^2 \\ &= \mathbf{K}(h,h) + \mathbf{K}(t,t) + \mathbf{K}(r,r) \\ &\quad -2\mathbf{K}(h,t) - 2\mathbf{K}(r,t) + 2\mathbf{K}(r,h)\end{aligned} \tag{2}$$

where φ is the mapping from the original space to Hilbert space, and $\mathbf{K}$ is the induced kernel by φ. Commonly, $\mathbf{K}$
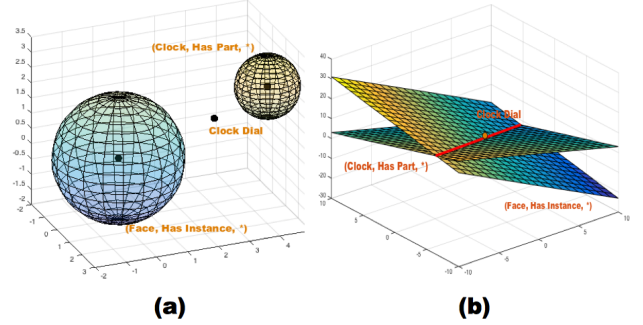
Figure 2: Visualization of Embedding for Manifold-based models. (a) corresponds to the Sphere setting where all the tail entities are supposed to lay in the sphere. As Clock Dial is matched by the two facts, it should lay in both spheres. (b) corresponds to the Hyperplane setting where Clock Dial should lay and does lay in both hyperplanes, making embedding more precise.

could take Linear kernel ($\mathbf{K}(a,b) = \mathbf{a}^\top\mathbf{b}$), Gaussian kernel ($\mathbf{K}(a,b) = e^{-\frac{||\mathbf{a}-\mathbf{b}||_2^2}{\sigma^2}}$), Polynomial Kernel ($\mathbf{K}(a,b) = (\mathbf{a}^\top\mathbf{b} + d)^p$), and so on. Obviously, if applying the linear kernel, the above function is reduced to the sphere manifold.

Hyperplane. As shown in Fig.2, we could see that when two manifolds are not intersected, there may be a loss in embedding. Two spheres would intersect only under some strict conditions, while two hyperplanes would intersect if their normal vectors are not in parallel. Motivated by this fact, we apply a hyperplane to enhance our model as below:

$$\mathcal{M}(h,r,t) = (\mathbf{h} + \mathbf{r_{head}})^\top(\mathbf{t} + \mathbf{r_{tail}})$$

where $\mathbf{r_{head}}$ and $\mathbf{r_{tail}}$ are two specific relation embeddings. From the geometric perspective, given the head entity and the relation, the tail entities lay in the hyperplane whose direction is $\mathbf{h} + \mathbf{r_{head}}$ and the bias is D_r^2. In practical cases, the two vectors $\mathbf{e_1} + \mathbf{r_{head}}$ and $\mathbf{e_2} + \mathbf{r_{tail}}$ for two entities e_1, e_2 have more chance to have intersection than two spheres. Therefore, there would be more solutions provided by the intersection of hyperplanes.

Motivated by enlarging the number of precisely predicted tail entities for the same head and relation, we apply the absolute operators as $\mathcal{M}(h,r,t) = |\mathbf{h} + \mathbf{r_{head}}|^\top|\mathbf{t} + \mathbf{r_{tail}}|$ where $|\mathbf{w}| \doteq (|w_1|, |w_2|, |w_3|, ...|w_n|)$. For an instance of one-dimensional case that $|h + r_{head}||t + r_{tail}| = D_r^2$, the absolute operator would double the solution number of t, meaning two tail entities rather than one could be matched precisely to this

Table 2: Evaluation results on link prediction

Datasets	WN18				FB15K			
Metric	HITS@10(%)		HITS@1	Time(s)	HITS@10(%)		HITS@1	Time(s)
	Raw	Filter	Filter	One Epos	Raw	Filter	Filter	One Epos
SE[Bordes *et al.*, 2011]	68.5	80.5	-	-	28.8	39.8	-	-
TransE [Bordes *et al.*, 2013]	75.4	89.2	29.5	**0.4**	34.9	47.1	24.4	**0.7**
TransH [Wang *et al.*, 2014]	73.0	82.3	31.3	1.4	48.2	64.4	24.8	4.8
TransR [Lin *et al.*, 2015b]	79.8	92.0	33.5	9.8	48.4	68.7	20.0	29.1
PTransE [Lin *et al.*, 2015a]	-	-	-	-	51.4	84.6	63.3	266.0
KG2E [He *et al.*, 2015]	80.2	92.8	54.1	10.7	48.9	74.0	40.4	44.2
ManifoldE Sphere	80.7	92.8	55.8	**0.4**	**55.7**	86.2	64.1	**0.7**
ManifoldE Hyperplane	**84.2**	**94.9**	**93.2**	0.5	55.2	**88.1**	**70.5**	0.8

head for this relation. For this reason, the absolute operator would promote the flexibility of embedding.

We also apply kernel trick to the Hyperplane setting, as below:

$$\mathcal{M}(h, r, t) = \mathbf{K}(\mathbf{h} + \mathbf{r_{head}}, \mathbf{t} + \mathbf{r_{tail}})$$

3.2 Algebraic Perspective

The ill-posed equation system that posses more equations than free variables always leads to some undesired properties such as instability, which may be the reason why the translation-based principle performs not so well in precise link prediction. To alleviate this issue, manifold-based methods model embedding within a nearly well-posed algebraic framework, since our principle indicates only one equation for one fact triple. Taking an example of sphere as $\sum_{i=1}^{d}(h_i + r_i - t_i)^2 = D_r^2$, we could conclude that if $d \geq \frac{T}{E+R}$, our embedding system would be more algebraically stable and this condition is easy to achieve by just enlarging the embedding dimension to a suitable degree. In theory, more embedding dimensions provide more solutions to embedding equations, which makes embedding more flexible. When the suitable condition is satisfied, the stable algebraic solution would lead the embedding to a fine characterization, therefore the precise link prediction would be promoted.

3.3 Geometric Perspective

The translation-based principle allocates just one position for a golden triple. We extend one point to a whole manifold such as a high dimensional sphere. For instance, all tail entities for a 1-N relation could lay on a sphere, which applies $\mathbf{h} + \mathbf{r}$ as the center and D_r as the radius. Obversely, there would be more suitable in a manifold setting than in a point setting.

3.4 Training

We train our model with the rank-based hinge loss, which means to maximize the discriminative margin between the golden triples and the false ones.

$$\mathcal{L} = \sum_{(h,r,t) \in \Delta} \sum_{(h',r',t') \in \Delta'} [f_r(h', t') - f_r(h, t) + \gamma]_+ \quad (3)$$

where $\mathcal{L}$ is the loss function which should be minimized, γ is the margin, and $[\cdot]_+ \doteq \max(0, \cdot)$ is the hinge loss. The false

triples are sampled with the Bernoulli Sampling Method as introduced in [Wang *et al.*, 2014]. We initialize the embedding vectors by the similar methods used in deep neural network [Glorot and Bengio, 2010]. SGD is applied to solve this problem.

In theory, our computation complexity relative to TransE is bounded by a very small constant, as $O(\lambda \times O(\text{TransE}))$ where $\lambda \geq 1$. This small constant λ is caused by manifold-based operations and kernelization. Commonly, TransE is most efficient among all the translation-based methods, while ManifoldE could be comparable to TransE in efficiency, hence better than other translation-based methods.

4 Experiments

Our experiments are conducted on four public benchmark datasets that are the subsets of Wordnet [Miller, 1995] and Freebase [Bollacker *et al.*, 2008]. The statistics of these datasets are listed in Tab.1. Experiments are conducted on two tasks : Link Prediction and Triple Classification. To further demonstrate how the proposed model performs the manifold-based principle, we present the visualization comparison between translation-based and manifold-based models in the section 4.3. Finally, we conduct case study and error analysis to further understand the benefit and limits of our models.

4.1 Link Prediction

Reasoning is the focus of knowledge computation. To verify the reasoning performance of embedding, link prediction task is conducted. This task aims to predict the missing entities. An alternative of the entities and a relation are given while the embedding methods infer the other missing entity. More specifically, in this task, we predict t given $(h, r, *)$, or predict h given $(*, r, t)$. The WN18 and FB15K are two benchmark datasets for this task. Notably, many AI tasks could be enhanced by "Link Prediction" such as relation extraction [Hoffmann *et al.*, 2011].

Evaluation Protocol. We adopt the same protocol used in previous studies. Firstly, for each testing triple (h, r, t), we corrupt it by replacing the tail t (or the head h) with every entity e in the knowledge graph. Secondly, we calculate a

Table 3: Evaluation results on FB15K by mapping properties of relations(%)

Tasks	Predicting Head(HITS@10)				Predicting Tail(HITS@10)			
Relation Category	1-1	1-N	N-1	N-N	1-1	1-N	N-1	N-N
TransE [Bordes *et al.*, 2013]	43.7	65.7	18.2	47.2	43.7	19.7	66.7	50.0
TransH [Wang *et al.*, 2014]	66.8	87.6	28.7	64.5	65.5	39.8	83.3	67.2
TransR [Lin *et al.*, 2015b]	78.8	89.2	34.1	69.2	79.2	37.4	90.4	72.1
PTransE [Lin *et al.*, 2015a]	90.1	92.0	58.7	86.1	90.1	70.7	87.5	88.7
KG2E [He *et al.*, 2015]	92.3	93.7	**66.0**	69.6	92.6	67.9	94.4	73.4
ManifoldE Sphere	92.4	95.5	53.2	84.4	92.8	68.7	94.8	86.8
ManifoldE Hyperplane	**92.5**	**97.0**	59.9	**88.2**	**93.4**	**70.8**	**94.9**	**90.0**

Table 4: Evaluation results for precise prediction on FB15K by mapping properties of relations(%)

Tasks	Predicting Head(HITS@1)				Predicting Tail(HITS@1)			
Relation Category	1-1	1-N	N-1	N-N	1-1	1-N	N-1	N-N
TransE [Bordes *et al.*, 2013]	35.4	50.7	8.6	18.1	34.5	10.6	56.1	20.3
TransH [Wang *et al.*, 2014]	35.3	48.7	8.4	16.9	35.5	10.4	57.5	19.3
TransR [Lin *et al.*, 2015b]	29.5	42.8	6.1	14.5	28.0	7.7	44.1	16.2
PTransE [Lin *et al.*, 2015a]	57.5	83.0	46.2	60.3	58.1	**58.4**	73.9	61.7
KG2E [He *et al.*, 2015]	62.3	73.9	39.4	30.4	62.3	33.9	76.2	33.8
ManifoldE Sphere	62.7	78.0	30.8	63.0	64.8	38.8	76.4	63.0
ManifoldE Hyperplane	**66.6**	**89.4**	**46.6**	**66.5**	**65.0**	54.2	**83.4**	**70.3**

probabilistic score of this corrupted triple with the score function $f_r(h, t)$. By ranking these scores in descending order, we then obtain the rank of the original triple. The evaluation metric is the proportion of testing triple whose rank is not larger than N (HITS@N). HITS@10 is applied for common reasoning ability and HITS@1 concerns the precise embedding performance. This is called "Raw" setting. When we filter out the corrupted triples that exist in the training, validation, or test datasets, this is the"Filter" setting. If a corrupted triple exists in the knowledge graph, ranking it ahead the original triple is also acceptable. To eliminate this case, the "Filter" setting is more preferred. In both settings, a higher HITS@N means better performance. Note that we do not report the results of "raw" setting for HITS@1, because they are too small to make a sense.

Implementation. As the datasets are the same, we directly reproduce the experimental results of several baselines from the literature for HITS@10. As to HITS@1, we request the results from the authors of PTransE and KG2E. We have attempted several settings on the validation dataset to get the best configuration. Under the "bern." sampling strategy, the optimal configurations of ManifoldE are as follows. For sphere, $\alpha = 0.001$, $k = 100$, $\gamma = 3.0$, Linear kernel, on WN18; $\alpha = 0.0005$, $k = 800$, $\gamma = 1.0$, Polynomial kernel($p = 2, d = 2$) on FB15K. For hyperplane, learning rate $\alpha = 0.01$, embedding dimension $k = 800$, margin $\gamma = 0.2$, Linear kernel, on WN18; $\alpha = 0.01$, $k = 1000$, $\gamma = 0.2$, Linear kernel, on FB15K. The experimental environment is a common PC with i7-4790 CPU, 16G Memory and Windows 10. Note that all the symbols are introduced in "Methods".

Results. Evaluation results on WN18 and FB15K are re-

ported in Tab.2 and Tab.3. We observe that:

1. ManifoldE beats all the baselines in all the sub-tasks, yielding the effectiveness and efficiency of the manifold-based principle.

2. From the algebraic perspective, it's reasonable to measure the algebraic ill-posed degree with the radio $T/(E + R)$ because with the translation-based principle, Td is the number of equations and $(E + R)d$ is the number of free variables, a larger radio means more ill-posed. Since the manifold-based principle alleviates this issue, ManifoldE(Sphere) would make a more promotion relatively to the comparable baselines (TransE) under a larger radio. As to the metric HITS@1, on WN18, the radio is 3.5 while ManifoldE(Sphere) achieves 55% and TransE achieves 29.5%, leading to a relative improvement of 85.1%. On FB15K the radio is 30.2 while ManifoldE(Sphere) achieves 64.1% and TransE achieves 24.4%, leading to a relative improvement of 162.7%. This comparison illustrates manifold-based methods could stabilize the algebraic property of embedding system, by which means, the precise embedding could be approached much better.

3. From the geometric perspective, traditional models attempt to express all the matched entities into one position, which leads to unsatisfactory performance on complex relations. Meanwhile, manifold-based model could perform much better for these complex relations as we discussed. As to the metric HITS@1, the simple relation 1-1 improves relatively by 87.8% by ManifoldE(Sphere) than TransE while the complex relations such as 1-N, N-1, N-N improve relatively by 266.5%, 36.2% and

Table 5: Triples classification: accuracy(%) for different embedding methods.

Methods	WN11	FB13	AVG.
SE	53.0	75.2	64.1
NTN	70.4	87.1	78.8
TransE	75.9	81.5	78.7
TransH	78.8	83.3	81.1
TransR	85.9	82.5	84.2
KG2E	85.4	85.3	85.4
ManifoldE Sphere	**87.5**	87.2	**87.4**
ManifoldE Hyperplane	86.9	**87.3**	87.1

215.7% respectively. This comparison demonstrates manifold-based method that extends the golden position from one point to a manifold could better characterize the true facts, especially for complex relations.

4.2 Triple Classification

In order to present the discriminative capability of our method between true and false facts, triple classification is conducted. This is a classical task in knowledge base embedding, which aims at predicting whether a given triple (h, r, t) is correct or not. WN11 and FB13 are the benchmark datasets for this task. Note that evaluation of classification needs negative samples, and the datasets have already been built with negative triples.

Evaluation Protocol. The decision process is very simple as follows: for a triple (h, r, t), if $f_r(h, t)$ is below a threshold σ_r, then positive; otherwise negative. The thresholds $\{\sigma_r\}$ are determined on the validation dataset. This task is somehow a triple binary classification.

Implementation. As all methods use the same datasets, we directly re-use the results of different methods from the literature. We have attempted several settings on the validation dataset to find the best configuration. The optimal configurations of ManifoldE are as follows with "bern" sampling. For sphere, $\alpha = 0.001$, $k = 100$, $\gamma = 10.0$, Linear kernel on WN18; $\alpha = 0.00005$, $k = 1000$, $\gamma = 0.3$, Gaussian kernel($\sigma = 1.0$) on FB13. For hyperplane, learning rate $\alpha = 0.01$, embedding dimension $k = 500$, margin $\gamma = 1.0$, Linear kernel, on WN18; $\alpha = 0.001$, $k = 1000$, $\gamma = 3.0$, Polynomial kernel ($p = 2, d = 2$), on FB13.

Results. Accuracies are reported in Tab.5. We observe that:

1. Overall, ManifoldE yields the best performance. This illustrates our method could improve the embedding.

2. More specifically, on WN11, the relation "Type Of" that is a complex one, improves from 71.4% of TransE to 86.3% of ManifoldE(Sphere) while on FB13, the relation "Gender" that is an extreme N-1 relation, improves from 95.1% to 99.5%. This comparison shows manifold-based method could handle complex relations better.

4.3 Visualization Comparison between Translation-Based and manifold-Based Principle

As the Fig.1 shows, the translation-based principle involves too much noise near the center where is supposed to lay the true facts. We attribute such issues to the precise link prediction issue as introduced previously, However, manifold-based principle alleviates this issue to enhance precise knowledge embedding, which could be seen from the visualization results.

4.4 Error Analysis

To analyze the errors in Link Prediction, we randomly sample 100 testing triples that could not rank at the top positions by ManifoldE (Hyperplane) and three categories of errors are summarized. Notably, we call the predicted top rank triple, which is not the golden one, as "top rank triple".

1. **True Facts (29%)**: The top rank triple is correct, but it is not contained in the knowledge graph, thus ranking it before the golden one is acceptable. This category is caused by incompleteness of the dataset. For example, reflexive semantics as $(\mathrm{Hot, Similar\ To, Hot})$, general expression as $(\mathrm{Animal\ Scientist, Type\ Of, Individual})$, professional knowledge as $(\mathrm{Crescentia, Type\ Of, Plant})$ and so on.

2. **Related Concepts (63%)**: The top rank triple is related concepts, but the fact is not exactly correct. This category is caused by the relatively simple manifolds applied by ManifoldE. For example, puzzled place membership as $(\mathrm{Afghanistan, Has\ Part, Hyderabad})$, similar mentions as $(\mathrm{New\ Haven, Part\ Of, One\ Star\ State})$, similar concepts as $(\mathrm{Sea, Has\ Instance, Aleutian\ Islands})$, possible knowledge as $(\mathrm{Accent, Type\ Of, Individual})$ and so on. We could further exploit complex manifolds to enhance the discriminative ability.

3. **Others (8%)**: There are always some top rank triples that are difficult to interpret. For example, $(\mathrm{Demonstration, Type\ Of, Body})$, $(\mathrm{Geographic\ Area, Has\ Instance, Argun})$ and so on.

5 Conclusions

In this paper, we study the precise link prediction problem and attribute two reasons to the problem: ill-posed algebraic system and over-restricted geometric form. To alleviate these issues, we propose a novel manifold-based principle and a few models ManifoldE (Sphere/Hyperplane) inspired by the principle. From algebraic perspective, ManifoldE is a nearly well-posed equation system and from a geometric perspective, it expands one point in translation-based principle to a manifold. Extensive experiments show our method achieves substantial improvements against the state-of-the-art baselines.

References

[Bollacker *et al.*, 2008] 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*, pages 1247–1250. ACM, 2008.

[Bordes *et al.*, 2011] Antoine Bordes, Jason Weston, Ronan Collobert, Yoshua Bengio, et al. Learning structured embeddings of knowledge bases. In *Proceedings of the Twenty-fifth AAAI Conference on Artificial Intelligence*, 2011.

[Bordes *et al.*, 2012] Antoine Bordes, Xavier Glorot, Jason Weston, and Yoshua Bengio. Joint learning of words and meaning representations for open-text semantic parsing. In *International Conference on Artificial Intelligence and Statistics*, pages 127–135, 2012.

[Bordes *et al.*, 2013] 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*, pages 2787–2795, 2013.

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

[Fan *et al.*, 2014] Miao Fan, Qiang Zhou, Emily Chang, and Thomas Fang Zheng. Transition-based knowledge graph embedding with relational mapping properties. In *Proceedings of the 28th Pacific Asia Conference on Language, Information, and Computation*, pages 328–337, 2014.

[Glorot and Bengio, 2010] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *International conference on artificial intelligence and statistics*, pages 249–256, 2010.

[Guo *et al.*, 2015] Shu Guo, Quan Wang, Bin Wang, Lihong Wang, and Li Guo. Semantically smooth knowledge graph embedding. In *Proceedings of ACL*, 2015.

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

[Hoffmann *et al.*, 2011] Raphael Hoffmann, Congle Zhang, Xiao Ling, Luke Zettlemoyer, and Daniel S Weld. Knowledge-based weak supervision for information extraction of overlapping relations. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 541–550. Association for Computational Linguistics, 2011.

[Jenatton *et al.*, 2012] Rodolphe Jenatton, Nicolas L Roux, Antoine Bordes, and Guillaume R Obozinski. A latent factor model for highly multi-relational data. In *Advances in Neural Information Processing Systems*, pages 3167–3175, 2012.

[Ji *et al.*,] Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. Knowledge graph embedding via dynamic mapping matrix.

[Lin *et al.*, 2015a] Yankai Lin, Zhiyuan Liu, and Maosong Sun. Modeling relation paths for representation learning of knowledge bases. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics*, 2015.

[Lin *et al.*, 2015b] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

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

[Nickel *et al.*, 2011] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. A three-way model for collective learning on multi-relational data. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 809–816, 2011.

[Nickel *et al.*, 2012] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. Factorizing yago: scalable machine learning for linked data. In *Proceedings of the 21st international conference on World Wide Web*, pages 271–280. ACM, 2012.

[Socher *et al.*, 2013] Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. Reasoning with neural tensor networks for knowledge base completion. In *Advances in Neural Information Processing Systems*, pages 926–934, 2013.

[Sutskever *et al.*, 2009] Ilya Sutskever, Joshua B Tenenbaum, and Ruslan Salakhutdinov. Modelling relational data using bayesian clustered tensor factorization. In *Advances in neural information processing systems*, pages 1821–1828, 2009.

[Tikhonov and Arsenin, 1978] A. N. Tikhonov and V. Y. Arsenin. Solutions of ill-posed problems. *Mathematics of Computation*, 32(144):491–491, 1978.

[Wang *et al.*, 2014] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, pages 1112–1119, 2014.

[Wang *et al.*, 2015] Quan Wang, Bin Wang, and Li Guo. Knowledge base completion using embeddings and rules. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, 2015.