

Article

An Approach to Knowledge Base Completion by a Committee-Based Knowledge Graph Embedding

Su Jeong Choi ¹, Hyun-Je Song ² and Seong-Bae Park ^{3,*}

- ¹ Institute of Convergence Technology, KT, 151 Taebong-ro, Seocho-gu, Seoul 06763, Korea; sujeong.choi@kt.com
- ² Department of Information Technology, Jeonbuk National University, 567 Baekje-daero, Deokjin-gu, Jeonju-si 54896, Jeollabuk-do, Korea; hyunje.song@jbnu.ac.kr
- ³ Department of Computer Science and Engineering, Kyung Hee University, 1732 Deogyeong-daero, Yongin-si 17104, Gyeonggi-do, Korea
- * Correspondence: sbpark71@khu.ac.kr

Received: 15 March 2020; Accepted: 8 April 2020; Published: 11 April 2020



Abstract: Knowledge bases such as Freebase, YAGO, DBPedia, and Nell contain a number of facts with various entities and relations. Since they store many facts, they are regarded as core resources for many natural language processing tasks. Nevertheless, they are not normally complete and have many missing facts. Such missing facts keep them from being used in diverse applications in spite of their usefulness. Therefore, it is significant to complete knowledge bases. Knowledge graph embedding is one of the promising approaches to completing a knowledge base and thus many variants of knowledge graph embedding have been proposed. It maps all entities and relations in knowledge base onto a low dimensional vector space. Then, candidate facts that are plausible in the space are determined as missing facts. However, any single knowledge graph embedding is insufficient to complete a knowledge base. As a solution to this problem, this paper defines knowledge base completion as a ranking task and proposes a committee-based knowledge graph embedding model for improving the performance of knowledge base completion. Since each knowledge graph embedding has its own idiosyncrasy, we make up a committee of various knowledge graph embeddings to reflect various perspectives. After ranking all candidate facts according to their plausibility computed by the committee, the top-k facts are chosen as missing facts. Our experimental results on two data sets show that the proposed model achieves higher performance than any single knowledge graph embedding and shows robust performances regardless of k. These results prove that the proposed model considers various perspectives in measuring the plausibility of candidate facts.

Keywords: knowledge base completion; knowledge graph construction; knowledge graph embedding; committee machine

1. Introduction

Large-scale knowledge bases such as Freebase [1], DBPdia [2], and NELL [3] are now publicly available, and they contain massive volumes of facts involving diverse entities and relations. Due to their huge size, they are used as an essential resource in many language-related tasks such as information retrieval, question-answering, and text mining. However, no matter how large the knowledge bases are, they are not yet complete since they are constructed manually. For instance, Freebase has three million entities for the 'Person' concept, but only 25% of them have nationality information [4]. In addition, the entities with parent information occupy just 6% of the 'Person' entities, while it is natural that every person has a nationality or parent(s) in the real world. Such missing information appears for nearly every relation, and then is piled up. As a result, these cumulative missing facts limit the effective use of knowledge bases. Therefore, it is significant to fill in missing facts of a knowledge base automatically.



There have been many previous studies that attempted to fill in missing facts. One promising approach to this task is to use knowledge graph embedding [5–9], which represents all entities and relations of a knowledge base as vectors in a low-dimensional space. The candidates for missing facts are also represented as vectors in the low-dimensional space and their plausibility is measured by vector similarity to the existing facts in the space. Then, the candidates with high plausibility are discovered as possible facts for the knowledge base. According to the experimental results of previous studies [10], this approach achieves over 70% in Hits@10 (Hits@k is the rate of correct entities appearing at top-k ranked entities. That is, Hits@10 is the proportion of correct entities ranked in the top 10) in knowledge base completion. However, its Hits@1 is lower than 30%. This implies that it is still difficult to find correct missing facts using knowledge graph embedding alone.

A couple of methods have been proposed to improve Hits@1 performance in discovering missing facts. Wang et al. conducted knowledge base completion using both knowledge graph embedding and rules derived from knowledge base schema [11]. Their experimental results show that adopting the rules helps improve Hits@1 performance. However, their method still does not cover *n*-to-*n* relations, since it is based on knowledge graph embedding which utilizes the characteristics of 1-to-1 relations. On the other hand, Choi et al. proposed a re-ranking model that uses both internal and external information of a knowledge graph for more accurate knowledge base completion [12]. Their model first extracts top-*k* candidates according to the plausibility computed by knowledge graph embedding. Then, the candidates are re-ranked by considering two kinds of additional information: knowledge base schema information and web search results. This is a good approach in that it is the first attempt to exploit such additional information, but it has still some problems. The main problem of the model is that it depends severely on the first top-*k* results. If a correct fact is excluded in the first step, there is no chance to include it in the second step. Note that, in this model, the top-*k* candidates are determined by a single knowledge graph embedding while no knowledge graph embedding is yet perfect. As a result, many correct facts are missed by this model.

This paper proposes a sole ranking model that adopts a committee of knowledge graph embeddings for accurate knowledge base completion. Unlike previous work that represents knowledge base completion as a re-ranking task [12], we formulate it as a ranking task. Given a knowledge base, the proposed model generates candidate facts, and then the plausibility of each candidate is determined by a committee of knowledge graph embeddings, not by a single knowledge graph embedding. After that, the candidates are sorted according to their plausibility. Since the proposed model evaluates all candidates, the probability of missing correct facts gets reduced. Another advantage of the proposed model is that it can reflect various perspectives of a knowledge graph for measuring candidate plausibility, since it is a kind of committee machine. That is, the diversity of the committee members allows the model to have less variance error. According to our experiments on two standard data sets, the proposed committee-based model outperforms every single knowledge graph embedding. In addition, it shows higher Hits@k performance as k decreases, which implies it predicts missing facts more accurately.

The rest of this paper is organized as follows. Section 2 describes related work on knowledge graph embedding for knowledge base completion. Section 3 proposes the overall idea of the proposed model and Section 4 explains how the proposed model works as a committee machine. Section 5 describes how to measure the plausibility of each candidate using the proposed model and Section 6 shows the experimental results. Finally, Section 7 draws our conclusions.

2. Related Work

There have been a number of previous studies on knowledge base completion. One promising approach in the studies is to use knowledge graph embedding [13–16]. Knowledge graph embedding represents all entities and relations of knowledge facts as low-dimensional vectors. These vectors of entities and relations are trained by preserving the inherent structure of a knowledge base. Thus, the plausibility of a knowledge fact can be measured using the vectors, and then knowledge base completion is done by filling up the facts with a high plausibility.

Knowledge graph embeddings are clustered into two kinds of approach [17,18]. One is a translation-based model and the other is a latent semantics-based model. In the translation-based model, the entities are translated according to their relation and the plausibility of facts is measured by a distance-based plausibility function. The best-known representative translation-based model is TransE [6]. TransE represents entities and relations as vectors in the same space assuming that the sum of a head entity vector and a relation vector is equal to a tail entity vector. Therefore, the plausibility function of TransE considers the distance between the sum vector and tail entity vector. It is simple and effective, especially for one-to-one relations.

There have been many extensions of TransE since its first appearance [19–22]. TransH [23], TransR [24], and TransD [25] are the extensions, but they all adopt relation-specific entity embedding. Since TransH projects entities onto relation-specific hyperplanes, it can represent an entity as a different vector according to its relation. Similarly, TransR adopts a relation-specific space rather than hyperplanes. It embeds entities into relation-specific spaces with a projection matrix. TransD simplifies TransR by decomposing the projection matrix into a product of two vectors, where the two vectors are mapping matrices of a head and a tail. Since entities are projected according to their role, it can represent one-to-*n*, *n*-to-*n*, and *m*-to-*n* relations.

Latent semantics-based models capture the latent semantics of entities and relations using a similarity-based plausibility function [5,26,27]. In RESCAL [7], entities are represented as vectors, and relations are matrices derived from pairwise interactions between entities. It works well for all relation types from one-to-one to *n*-to-*m*. However, it suffers from high complexity [28]. DistMult solves this problem by simplifying the relation matrices of RESCAL [29]. It represents a relation as diagonal matrices instead of matrices of relations. HolE combines RESCAL and DistMult effectively [30]. The entities and relations are all represented as vectors, and the plausibility function adopts a circular correlation operation [31] to compress the pairwise interactions. By this operation, HolE is able to model asymmetric relations that DistMult cannot. ComplEX is another method that extends DistMult to model asymmetric relations [32]. Since it embeds entities and relations into a complex vector space, its plausibility function is based on a Hermitian dot product. As a result, it is scalable to manage a large data set.

Table 1 summarizes the knowledge graph embeddings explained above. For a given triple $\langle h, r, t \rangle$, all methods but ComplEx embeds h, r, and t into a space \mathbb{R}^d , where d is a space dimension. ComplEx embeds them into a complex space \mathbb{C}^d . Then, h and t are represented as vectors \mathbf{h} and \mathbf{t} , and r is as a vector \mathbf{r} or as projection matrices \mathbf{w} and \mathbf{M} . In order to train these embedding vectors, each embedding has its own plausibility function $f_r(\mathbf{h}, \mathbf{t})$. In Table 1, $diag(\cdot)$ implies a diagonal matrix, \otimes is the circular correlation operation [31], and $Re(\cdot)$ is the real part of a complex value.

	Entity Embedding	Relation Embedding	Plausibility Function
TransE	h , t $\in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$\ \mathbf{h} + \mathbf{r} - \mathbf{t}\ _{2}^{2}$
TransH	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r}, \mathbf{w}_r \in \mathbb{R}^d$	$\ (\mathbf{h} - \mathbf{w}_r^{T}\mathbf{h}\mathbf{w}_r) + r - (\mathbf{t} - \mathbf{w}_r^{T}\mathbf{t}\mathbf{w}_r)\ _2^2$
TransR	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^k$, $\mathbf{M}_r \in \mathbb{R}^{k imes d}$	$\ \mathbf{M}_r\mathbf{h}+\mathbf{r}-\mathbf{M}_r\mathbf{t}\ _2^2$
TransD	$\mathbf{h}, \mathbf{w}_h, \mathbf{t}, \mathbf{w}_t \in \mathbb{R}^d$	$\mathbf{r}, \mathbf{w}_r \in \mathbb{R}^k$	$\ (\mathbf{w}_r\mathbf{w}_h^{T}+\mathbf{I})\mathbf{h}+\mathbf{r}-(\mathbf{w}_r\mathbf{w}_t^{T}+\mathbf{I})\mathbf{t}\ _2^2$
RESCAL	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{M}_r \in \mathbb{R}^{d imes d}$	$\mathbf{h}^{\top}\mathbf{M}_{r}\mathbf{t}$
DistMult	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$\mathbf{h}^{T} diag(\mathbf{r})\mathbf{t}$
HolE	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$\mathbf{r}^{ op}(\mathbf{h}\otimes\mathbf{t})$
ComplEx	h , t $\in \mathbb{C}^d$	$\mathbf{r} \in \mathbb{C}^d$	$Re(\mathbf{h}^{\top}diag(\mathbf{r})\mathbf{\bar{t}})$

Table 1. Entity and relation embeddings of knowledge graph embeddings.

There have been a few studies in which the advantages of various embeddings are combined. Krompaß et al. proposed an ensemble of knowledge graph embeddings for knowledge base completion [33]. Their method combines three knowledge graph embeddings of TransE, RESCAL, and the embedding proposed by Xin et al. [34]. They applied a logistic regression to each embedding to normalize the plausibility by each embedding. As a result, as the number of knowledge graph embeddings increases, the number of logistic regressions to train also increases. Another problem of their method is that it ignores the relative importance of each embedding.

3. Knowledge Base Completion

In general, a knowledge base stores a number of entities and relations, but it is usually incomplete in that it has many missing facts that exist in the real world. The applications of knowledge bases are limited due to this incompleteness. Therefore, it is important to solve knowledge base incompleteness.

The knowledge base completion is to find a set of facts missing from a knowledge base. Assume that a knowledge base *S* with a great number of facts is given. Each fact in *S* is represented as a triple $\langle h, r, t \rangle$, where *h* is a head entity, *t* is a tail entity, and *r* is their relation, respectively. The head entity *h* and the tail entity *t* belong to *E*, a set of entities, while the relation *r* is a member of *R*, a set of relations. Then, the knowledge base completion completes *S* by finding a set of missing triples $T = \{\langle h_c, r_c, t_c \rangle | h_c \in E, r_c \in R, t_c \in E, \langle h_c, r_c, t_c \rangle \notin S \}.$

On difficult problem in finding *T* is that not all possible candidates for the missing triples generated from *S* belong to *T*. For instance, a triple (Donald Trump, *nationality*, China) should not be a member of *T*, even if a knowledge base *S* has 'Donald Trump' and 'China' as its entities and *nationality* as its relation. Therefore, it is critical to measure the plausibility of every candidate missing triple. If a candidate triple is plausible enough with respect to a knowledge base *S*, it should be a member of *T*. Otherwise, it should be discarded.

4. Committee-Based Knowledge Base Completion

This paper proposes committee-based knowledge base completion that adopts a committee of knowledge graph embeddings to measure the plausibility of candidate missing triples. Figure 1 shows how the set of missing triples, *T*, is found systematically by the proposed committee-based knowledge base completion method. The proposed method first generates C_{cand} , a set of candidate missing triples from *S* as in the work of Bordes et al. [35]. Let C_h be a concept of *h*, C_t be that of *t*, H_h be a set of all entities belonging to C_h , and T_t be a set of all entities belonging to C_t . Then, from every triple $\langle h, r, t \rangle \in S$, the candidate triples are generated by replacing *h* with one of the elements in H_h except *h* or replacing *t* with one of the elements in T_t but *t*. Thus, $(|H_h| - 1) \cdot (|T_t| - 1)$ candidates are prepared from each triple in *S*.

Once C_{cand} is prepared, all its members are sorted according to their plausibility. There could be a number of ways to compute the plausibility of candidate triples, but knowledge graph embeddings are adopted to compute the plausibility in this paper. A knowledge graph embedding represents all entities and relations in *S* as low-dimensional vectors thereby preserving the inherent structure of *S*. Since the embedding is trained to preserve the structure of its knowledge graph, it can be used to measure how plausible the embedded vectors are in the space spanned by the knowledge graph embedding.

Even if any knowledge graph embedding can be used to compute the plausibility of triples, every embedding has its own idiosyncrasy. Table 2 proves this. This table shows the top-10 candidate tail entities suggested by four famous knowledge graph embeddings when a head entity *yogurt* and a relation *hypernym* are known. The knowledge graph embeddings used in this table are TransE, TransR, DistMult, and ComplEx, and the entities are the synsets in the WordNet knowledge base. The bold entities are correct tail entities for *yogurt* and *hypernym* (Multiple entities can be correct since the relation *hypernym* is transitive), and the shaded entities are those shared at least by two embeddings. As shown in this table, different knowledge graph embeddings suggest different entities. TransE shares three entities *dairy product, solid food*, and *foodstuff* out of ten with TransR, and two entities of

dairy product and *solid food* with DistMult and ComplEx. Moreover, even the shared entities are ranked differently. For instance, *dairy product* is ranked first by TransE, fifth by TransR, ninth by DistMult, and second by ComplEx, while *solid food* is ranked second, third, sixth, and first by them.

The ranking differences according to different embedding can be shown numerically by *Spearman's rank correlation coefficient*, which is a widely-used metric to evaluate rank correlation. Table 3 shows the Spearman's rank correlation coefficient between different knowledge graph embeddings. These coefficients are obtained using 500 triples sampled from the development set of WN18 dataset [6]. The average coefficient among the embeddings is 0.5158. The coefficient between TransE and TranR is highest as 0.5904, and that between TransR and DistMult is lowest as 0.4916. These values prove that the embeddings are positively related, but not that strongly related. DistMult, in particular, shows lower coefficients against other embeddings, which implies that it ranks entities differently from the others.

In order to reflect the idiosyncrasies of knowledge graph embeddings in knowledge base completion while maximizing their effectiveness, the proposed method computes the plausibility of triples by a committee of knowledge graph embeddings rather than by a single knowledge graph embedding. That is, the plausibility of a candidate triple in C_{cand} is determined by the committee of knowledge graph embeddings. Then, all the members of C_{cand} are sorted by their plausibility, and the sorted set is C_{rank} . Finally, the top-*k* candidate triples from C_{rank} are selected as new facts to form *T*, a final set of missing triples.



Figure 1. Overall process of knowledge base completion by knowledge graph embedding committee.

Table 2. Top-10 candidate triples for a head entity *yogurt* and a relation *hypernym*. Different embeddings suggest different entities as a tail for *yogurt* and *hypernym*.

Rank	TransE	TransR	DistMult	ComplEx
1	dairy product	leave	cheddar cheese	solid food
2	solid food	convertor	chalcidfly	dairy product
3	lover	solid food	tear down	degrade
4	free-reed instrument	internal secretion	immovableness	financial aid
5	fish	dairy product	fire	mixture
6	travel	wrong	solid food	tug
7	meat	foodstuff	pelargonium peltatum	copper-base alloy
8	preserves	baked goods	tyrant flycatcher	swamp plant
9	foodstuff	hair	dairy product	terabit
10	pull together	disposition	feather palm	previous question

	TransE	TransR	DistMult	ComplEx
TransE	-	0.5904	0.4940	0.4946
TransR	0.5904	-	0.4916	0.4926
DistMulti	0.4940	0.4916	-	0.5317
ComplEx	0.4946	0.4926	0.5317	-

Table 3. Spearman's rank correlation coefficients among the knowledge graph embeddings.

5. Measuring Plausibility by Embedding Committee

The product of experts (PoE) proposed by Garcia-Durán et al. [36] is adopted as a committee machine to combine knowledge graph embeddings. When all experts are probabilistic models, PoE models their overall probability distribution by combining their outputs. While each expert considers a particular aspect of a target task, PoE manages the task comprehensively. As a result, it produces a better distribution than individual experts.

In this paper, the plausibility of a triple x is represented as its probability, where P(x), the probability of x, is estimated by

$$\mathbf{P}(\mathbf{x}|\theta_1,...,\theta_n) = \frac{\prod_m f_m(\mathbf{x}|\theta_m)}{\sum_{\mathbf{a}\in\mathbf{A}}\prod_m f_m(\mathbf{a}|\theta_m)}.$$
(1)

Here, θ_m is the parameter of the *m*-th individual embedding, $f_m(\mathbf{x}|\theta_m)$ is the score of \mathbf{x} by the *m*-th embedding, and \mathbf{A} is all possible candidate triples. $\mathbf{P}(\mathbf{x})$ is basically a product of all outputs by individual embeddings. Thus, the more plausible \mathbf{x} is, the higher $\mathbf{P}(\mathbf{x})$ it has.

Four knowledge graph embeddings of TransE [6], TransR [24], DistMult [29], and ComplEx [32] are adopted as members of a committee to determine the final plausibility of candidate triples. As shown in Table 3, these four embeddings have diverse tendencies for ranking triples. These embeddings output a real value as a score for a triple, but each expert should be a probabilistic model in PoE. Thus, the sigmoid function σ is used to covert the score to a probability. That is, the score functions of the embeddings are

$$f_{transE} = \sigma(-\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_2^2), \qquad (2)$$

$$f_{transR} = \sigma(-\|\mathbf{M}_r\mathbf{h} + \mathbf{r} - \mathbf{M}_r\mathbf{t}\|_2^2), \qquad (3)$$

$$f_{distMult} = \sigma(\mathbf{h}^{\mathsf{T}} diag(\mathbf{r})\mathbf{t}), \tag{4}$$

$$f_{complEx} = \sigma(Re(\mathbf{h}^{\mathsf{T}} diag(\mathbf{r})\bar{\mathbf{t}})).$$
(5)

TransE and TransR have a distance-based score function, which implies that the smaller the score of a triple is, the more plausible the triple is. Thus, in Equation (2) and (3), a negative score is used in the sigmoid.

Note that the parameter θ_m of an embedding in Equation (1) is determined to maximize the performance the embedding without considering other committee members. Therefore, in order to rank the candidate triples optimally by the committee, all θ_m 's should be fine-tuned by considering neighboring embeddings. To fine-tune the parameters, the negative log-likelihood loss is used, which is defined as

$$L(\mathbf{x}) = -\log \mathbf{P}(\mathbf{x}|\theta_1, ..., \theta_n).$$
(6)

The derivative of $L(\mathbf{x})$ with respect to θ_m is

$$\frac{\partial L}{\partial \theta_m} = \frac{\partial \log \mathbf{P}(\mathbf{x}|\theta_1, ..., \theta_n)}{\partial \theta_m}$$
$$= \frac{\partial \log f_m(\mathbf{x}|\theta_m)}{\partial \theta_m} - \frac{\partial \log \sum_{\mathbf{c} \in \mathbf{C}} \prod_m f_m(\mathbf{c}|\theta_m)}{\partial \theta_m}.$$
(7)

Since the second term of Equation (7) is intractable, it is approximated through a negative sampling [37]. In this paper, negative samples are generated by swapping the head or tail entity of training triples. When a training triple $\mathbf{x} = \langle h, r, t \rangle$ is given, the negative triples $\langle h, r, t' \rangle$ are made by random-sampling t' from E, and replacing it with t in x. Another negative triples $\langle h', r, t \rangle$ are generated in the same way. Then, Equation (7) becomes

$$\frac{\partial L}{\partial \theta_m} = \frac{\partial \log \mathbf{P}(\mathbf{x}|\theta_1, ..., \theta_n)}{\partial \theta_m}$$
$$= \frac{\partial \log f_m(\mathbf{x}|\theta_m)}{\partial \theta_m} - \frac{\partial \log \sum_{\mathbf{g} \in \mathbf{G}} \prod_m f_m(\mathbf{g}|\theta_m)}{\partial \theta_m}, \tag{8}$$

where **G** is a set of negative samples.

6. Experiments

6.1. Experimental Settings

Two benchmark data sets of FB15K and WN18 are used to evaluate the proposed model, where FB15K is a subset of Freebase [1] that contains general facts about the world and WN18 is originated from WordNet [38] which provides the semantic relations among words. Table 4 gives simple statistics on these data sets. In this table, '# of Entitties' and '# of Relations' denote the number of entities and relation, respectively. '# of Training Triples', '# of Validation Triples', and '# of Test Triples' are the number of triples in the training, validation, and test set. FB15K has 592,213 triples while WN18 has 151,441 triples. The total number of triples in FB15K is larger than that in WN18, but the number of entities in FB15K is smaller than that in WN18. In addition, the number of relations in FB15K is about ten times larger than that of WN18.

Data Set	FB15K	WN18
# of Entities	14,951	40,943
# of Relations	1345	18
# of Training Triples	483,142	141,441
# of Validation Triples	50,000	5000
# of Test Triples	59,071	5000
# of Total Triples	592,213	151,441

Table 4. Simple statistics on data sets.

Four different knowledge graph embeddings: TransE [6], TransR [24], DistMult [29] and ComplEx [32] are adopted as members of the proposed embedding committee. Each member embedding is pre-trained with a training set from the data sets to initialize the committee. In TransE and TransR, entities and relation are embedded into a 50-dimensional space with 512 batch size. On the other hand, DistMult and ComplEx represent them as 100-dimensional vectors with 32 batch size. Margin and learning rate are set to 1 and 0.1 for all member embeddings.

In training the committee, 512 and 0.001 are used for batch size and learning rate, respectively, for WN18. The size of negative samples is double of the batch size. For FB15K, 1024 batch size is used, and the learning rate is equal to that for WN18. Adam [39] is adopted as an optimizer.

Hits@*k*, a commonly-used evaluation metric for knowledge base completion task is used to measure the performance of the models. This metric returns the proportion of correct triples in the top-*k* ranked triples. In all experiments, we report Hits@*k*'s following *raw* settings [6].

6.2. Experimental Results

Table 5 shows the results of the proposed model for WN18 data set. 'Head' implies candidate triples are generated by swapping head entities. Thus, it is the problem of finding correct head entities.

Similarly, 'Tail' is the problem of finding correct tail entities. As noted above, all values in this table are measured by Hits@k. Thus, 'Average' implies the average of Hits@k of 'Head' and that of 'Tail'.

Overall, the performance of all models gets worse as k gets smaller. That is, the values of Hits@10's are higher than those of Hits@3's and Hits@1's, and Hits@3's are higher than Hits@1's. This is natural because the smaller k gets, the less probable it is that a candidate triple becomes an actual fact. In addition, ComplEx always achieves the highest performance among the baseline models, and TransR achieves the lowest, on average. When k = 10, ComplEx achieves the highest Hits@10 of 0.8, while TransR does the lowest Hits@10 of 0.579. The Hits@3 of ComplEx is 0.647 which is nearly double of that of TransR. Specifically, it achieves 0.455 in Hits@1, and is around 10 times higher than that of TransE. The proposed model noted as 'committee-based model' outperforms all these baseline models. Its Hits@10 is 0.824, Hits@3 is 0.691, and Hits@1 is 0.495.

The experimental results for FB15K data set are shown in Table 6. The performance for FB15K data set is lower than that for WN18 data set. This is because that the size of FB15K is larger than that of WN18 despite the smaller number of entities. The performances of TransE and TransR in Hits@10 are 0.467 and 0.377, respectively. Their Hits@3's are 0.228 and 0.187, while Hits@1's are 0.030 and 0.052. Although both DistMult and ComplEx achieve higher results for all Hits@10, Hits@3, and Hits@1, DistMult outperforms ComplEx for this data set. That is, DistMult is the best baseline model for this data set. However, the proposed committee-based model is superior to DistMult even in FB15K. Hits@10 of the proposed model is 0.532, its Hits@3 is 0.305, and its Hits@1 is 0.162. All these values are much higher than those of the baseline models regardless of k.

		TransE	TransR	DistMult	ComplEx	Committee-Based Model
Hits@10	Head	0.721	0.575	0.787	0.794	0.813
	Tail	0.733	0.584	0.805	0.807	0.835
	Average	0.727	0.579	0.796	0.800	0.824
Hits@3	Head Tail Average	$\begin{array}{c} 0.474 \\ 0.474 \\ 0.474 \end{array}$	0.330 0.314 0.322	0.600 0.626 0.613	0.630 0.664 0.647	0.673 0.709 0.691
Hits@1	Head	0.057	0.080	0.340	0.441	0.477
	Tail	0.034	0.054	0.368	0.469	0.514
	Average	0.045	0.067	0.354	0.455	0.495

Table 5. Experimental results for WN18 data set.

Table 6. Experimental results for FB15K data set.

		TransE	TransR	DistMult	ComplEx	Committee-Based Model
10	Head	0.434	0.344	0.483	0.465	0.501
8	Tail	0.501	0.410	0.551	0.526	0.564
Hil	Average	0.467	0.377	0.517	0.495	0.532
3 3	Head	0.213	0.167	0.272	0.260	0.281
ts	Tail	0.244	0.207	0.327	0.305	0.330
Ηi	Average	0.228	0.187	0.299	0.282	0.305
1 1	Head	0.054	0.059	0.129	0.128	0.145
ts	Tail	0.007	0.046	0.170	0.160	0.180
Η	Average	0.030	0.052	0.149	0.144	0.162

These experimental results prove the effectiveness of the proposed committee model. Since the proposed model can reflect various aspects of a knowledge graph by combining various knowledge graph embeddings, it outperforms the baseline models. Figure 2 proves this fact. This figure depicts how Hits@k changes as k decreases from ten to one. Figure 2a shows the change of Hits@k for WN18 data set, and Figure 2b is about FB15K data set. The performance of all models naturally decreases as k

decreases, since the task gets more difficult with smaller k. However, the proposed model outperforms all its competitors consistently for all k's in both figures. One thing to note here is that the performances of the translation-based models (TransE and TransR) decrease rapidly as k gets smaller. Especifically, the performance of TransE becomes worst when k = 1, even though its performance is much higher than that of TransR for k > 1 and similar to those of DistMult and ComplEx for large k's. However, the proposed model consistently shows the best performance even for small k's, which implies that the weaknesses of the translation-based models are compensated by DistMult or ComplEx. In addition, the fact that the proposed model outperforms both DistMult and ComplEx for all k's proves that the shortcomings of DistMult or ComplEx vanish in the proposed model due to the influence of the translation-based models. Therefore, the proposed model reflects the synergistic advantages of the various knowledge graph embeddings, and achieves the best performance.



Figure 2. Change of *Hits@k as k decreases*.

7. Conclusions

This paper has proposed a committee model of knowledge graph embeddings for knowledge base completion, where knowledge base completion is a problem of discovering missing triples in a knowledge base. The previous research on this task has used a single knowledge graph embedding. Since every embedding has its own idiosyncrasy, no single knowledge graph embedding is fully effective in solving the task. Thus, we address the knowledge base completion task by organizing a committee that reflects the synergistic advantages of various knowledge graph embeddings. When a knowledge base is given, candidate triples are first generated from the knowledge base to discover missing triples and then their plausibilities are computed by the committee. The candidates with a high plausibility are selected as missing facts of the knowledge base. This paper incorporates TransE, TransR, DistMult, and ComplEx into the embedding committee. According to our experimental results, the proposed committee-based model shows higher performance than any single knowledge graph embedding. In addition, it is robust even when the model accepts only a small partition of candidate triples. This is because the proposed model can combine these knowledge graph embeddings effectively. For future work, we are going to combine another knowledge graph embeddings into the committee. In addition, the performance of knowledge base completion depends much on the negative sampling method. Thus, we are going to study better negative sampling methods to improve the committee machine.

Author Contributions: Conceptualization, S.J.C. and H.-J.S.; Funding acquisition, S.-B.P.; Investigation, S.J.C.; Methodology, S.J.C. and H.-J.S.; Project administration, S.-B.P.; Supervision, S.-B.P.; Validation, H.-J.S.; Visualization, S.J.C.; Writing—original draft, S.J.C.; Writing—review & editing, H.-J.S. and S.-B.P. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: This work was supported by Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korean government(MSIT)(No.2013-2-00109, WiseKB: Big data based self-evolving knowledge base and reasoning platform).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Bollacker, K.; Evans, C.; Paritosh, P.; Sturge, T.; Taylor, J. Freebase: a collaboratively created graph database for structuring human knowledge. In Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, Vancouver, BC, Canada, 10–12 June 2008; pp. 1274–1250.
- Lehmann, J.; Isele, R.; Jakob, M.; Jentzsch, A.; Kontokostas, D.; Mendes, P.N.; Hellmann, S.; Morsey, M.; van Kleef, P.; Auer, S.; et al. DBpedia—A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia. *Semant. Web J.* 2015, *6*, 167–195.
- Carlson, A.; Betteridge, J.; Kisiel, B.; Settles, B.; Hruschka, E.R., Jr.; Mitchell, T.M. Toward an architecture for never-ending language learning. In Proceedings of the 24th AAAI Conference on Artificial Intelligence, Atlanta, Georgia, USA, 11–15 July 2010; pp. 1306–1313.
- 4. West, R.; Gabrilovich, E.; Murphy, K.; Sun, S.; Gupta, R.; Lin, D. Knowledge Base Completion via Search-Based Question Answering. In Proceedings of the 23th International Conference on World Wide Web, Seoul, Korea, 7–11 April 2014; pp. 515–526.
- 5. Bordes, A.; Glorot, X.; Weston, J.; Bengio, Y. A semantic matching energy function for learning with multi-relational data. *Mach. Learn.* **2014**, *94*, 233–259.
- Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; Yakhnenko, O. Translating Embeddings for Modeling Multi-relational Data. In Proceedings of the 26th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–8 December 2013; pp. 2787–2795.
- Nickel, M.; Tresp, V.; Kriegel, H.P. A three-way model for collective learning on multi-relational data. In Proceedings of the 28th International Conference on International Conference on Machine Learning, Bellevue, WA, USA, 28 June–2 July 2011; pp. 809–816.
- Socher, R.; Chen, D.; Manning, C.D.; Ng, A. Reasoning with Neural Tensor Networks for Knowledge Base Completion. In Proceedings of the 26th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–8 December 2013; pp. 926–934.
- Xiao, H.; Huang, M.; Zhu, X. From one point to a manifold: Knowledge graph embedding for precise link prediction. In Proceedings of the 25th International Joint Conference on Artificial Intelligence, New York, NY, USA, 9–15 July 2016.; pp. 1315–1321.
- 10. Kadlec, R.; Bajgar, O.; Kleindienst, J. Knowledge Base Completion: Baselines Strike Back. In Proceedings of the 2nd Workshop on Representation Learning for NLP, Vancouver, BC, Canada, 3 August 2017; pp. 69–74.
- Wang, Q.; Wang, B.; Guo, L. Knowledge Base Completion Using Embeddings and Rules. In Proceedings of the 24th International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015; pp. 1859–1865.
- 12. Choi, S.J.; Song, H.J.; Yoon, H.G.; Park, S.B.; Park, S.Y. A Re-ranking Model for Accurate Knowledge Base Completion with Knowledge Base Schema and Web Statistic. In Proceedings of the IEEE Congress on Evolutionary Computation, San Sebastián, Spain, 6–11 July 2016; pp. 4958–4964.
- 13. Fan, M.; Zhou, Q.; Zheng, T.F.; Grishman, R. Distributed representation learning for knowledge graphs with entity descriptions. *Pattern Recognit. Lett.* **2017**, *93*, 31–37.
- He, S.; Liu, K.; Ji, G.; Zhao, J. Learning to Representation Knowledge Graphs with Gaussian Embedding. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, Melbourne, Australia, 19–23 October 2015; pp. 623–632.
- 15. Jia, Y.; Wang, Y.; Jin, X.; Lin, H.; Cheng, X. Knowledge Graph Embedding: A Locally and Temporally Adaptive Translation-Based Approach. *ACM Trans. Web* **2017**, *12*, 1–33.

- Xie, R.; Liu, Z.; Sun, M. Representation Learning of Knowledge Graphs with Hierarchical Types. In Proceedings of the 25th International Joint Conference on Artificial Intelligence, New York, NY, USA, 9–15 July 2016; pp. 2955–2971.
- 17. Nickel, M.; Murphy, K.; Tresp, V.; Gabrilovich, E. A review of relational machine learning for knowledge graphs. *Proc. IEEE* **2016**, *104*, 11–33.
- 18. Wang, Q.; Mao, Z.; Wang, B.; Guo, L. Knowledge graph embedding: A survey of approaches and applications. *IEEE Trans. Knowl. Data Eng.* **2017**, *29*, 2724–2743.
- Fan, M.; Zhou, Q.; Chang, E.; Zheng, T.F. Transition-based knowledge graph embedding with relational mapping properties. In Proceedings of the 28th Pacific Asia Conference on Language, Information and Computing, Phuket, Thailand, 12–14 December 2014; pp. 328–337.
- 20. Feng, J.; Huang, M.; Wang, M.; Zhou, M.; Hao, Y.; Zhu, X. Knowledge Graph Embedding by Flexible Translation. In Proceedings of the 15th International Conference on Principles of Knowledge Representation and Reasoning, Cape Town, South Africa, 25–29 April 2016; pp. 557–560.
- Ji, G.; Liu, K.; He, S.; Zhao, J. Knowledge graph completion with adaptive sparse transfer matrix. In Proceedings of the 30th AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; pp. 985–991.
- 22. Zhu, J.Z.; Jia, Y.T.; Xu, J.; Qiao, J.Z.; Cheng, X.Q. Modeling the Correlations of Relations for Knowledge Graph Embedding. *J. Comput. Sci. Technol.* **2018**, *33*, 323–334.
- Wang, Z.; Zhang, J.; Feng, J.; Chen, Z. Knowledge graph embedding by translating on hyperplanes. In Proceedings of the 28th AAAI Conference on Artificial Intelligence, Quebec City, QC, Canada, 27–31 July 2014; pp. 1112–1119.
- Lin, Y.; Liu, Z.; Sun, M.; Liu, Y.; Zhu, X. Learning Entity and Relation Embeddings for Knowledge Graph Completion. In Proceedings of the 29th AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015; pp. 2181–2187.
- 25. Ji, G.; He, S.; Xu, L.; Liu, K.; Zhao, J. 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, Beijing, China, 26–31 July 2015; pp. 687–696.
- 26. García-Durán, A.; Bordes, A.; Usunier, N.; Grandvalet, Y. Combining two and three-way embedding models for link prediction in knowledge bases. *J. Artif. Intell. Res.* **2016**, *55*, 715–742.
- Jenatton, R.; Roux, N.L.; Bordes, A.; Obozinski, G.R. A latent factor model for highly multi-relational data. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 3167–3175.
- 28. Liu, H.; Wu, Y.; Yang, Y. Analogical inference for multi-relational embeddings. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2018; pp. 2168–2178.
- Yang, B.; Yih, W.t.; He, X.; Gao, J.; Deng, L. Embedding Entities and Relations for Learning and Inference in Knowledge Bases. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.
- 30. Nickel, M.; Rosasco, L.; Poggio, T. Holographic Embeddings of Knowledge Graphs. In Proceedings of the 30th AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; pp. 1955–1961.
- 31. Plate, T.A. Holographic Reduced Representations. IEEE Trans. Neural Netw. 1995, 6, 623–641.
- 32. Trouillon, T.; Dance, C.R.; Gaussier, É.; Welbl, J.; Riedel, S.; Bouchard, G. Knowledge graph completion via complex tensor factorization. *J. Mach. Learn. Res.* **2017**, *18*, 4735–4772.
- Krompaß, D.; Tresp, V. Ensemble Solutions for Link-Prediction in Knowledge Graphs. In Proceedings of the 2nd Workshop on Linked Data for Knowledge Discovery, Porto, Portugal, 7–11th September 2015; pp. 1–10.
- 34. Dong, X.; Gabrilovich, E.; Heitz, G.; Horn, W.; Lao, N.; Murphy, K.; Strohmann, T.; Sun, S.; Zhang, W. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; pp. 601–610.
- Bordes, A.; Weston, J.; Collobert, R.; Bengio, Y. Learning Structured Embeddings of Knowledge Bases. In Proceedings of the 25th AAAI Conference on Artificial Intelligence, Francisco, CA, USA, 7–11 August 2011; pp. 301–306.

- García-Durán, A.; Niepert, M. KBLRN: End-to-End Learning of Knowledge Base Representations with Latent, Relational, and Numerical Features. In Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence, Monterey, CA, USA, 6–10 August 2018.
- 37. Mikolov, T.; Yih, W.; Zweig, G. Linguistic Regularities in Continuous Space Word Representations. In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Atlanta, GA, USA, 9–14 June 2013; pp. 746–751.
- 38. Fellbaum, C. WordNet: An Electronic Lexical Database; Bradford Books: Cambridge, MA, USA, 1998.
- 39. Kinga, D.; Adam, J.B. A method for stochastic optimization. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).