ARTICLE IN PRESS
JID: PATREC [m5G;October 3, 2016;9:20]

Pattern Recognition Letters 000 (2016) 1–7

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Distributed representation learning for knowledge graphs with entity descriptions☆

Miao Fan[a,*], Qiang Zhou[a], Thomas Fang Zheng[a], Ralph Grishman[b]

[a] CSLT, Division of Technical Innovation and Development, Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China
[b] Department of Computer Science, Courant Institute of Mathematical Sciences, New York University, NY 10003, USA

A R T I C L E I N F O

Article history:
Available online xxx

Keywords:
Knowledge graph
Representation learning
Entity description
Knowledge graph completion
Entity type classification

A B S T R A C T

Recent studies of knowledge representation attempt to project both entities and relations, which originally compose a high-dimensional and sparse knowledge graph, into a continuous low-dimensional space. One canonical approach *TransE* [2] which represents entities and relations with vectors (embeddings), achieves leading performances solely with triplets, i.e. (*head_entity, relation, tail_entity*), in a knowledge base. The cutting-edge method *DKRL* [23] extends *TransE* via enhancing the embeddings with entity descriptions by means of deep neural network models. However, *DKRL* requires extra space to store parameters of inner layers, and relies on more hyperparameters to be tuned. Therefore, we create a single-layer model which requests much fewer parameters. The model measures the probability of each triplet along with corresponding entity descriptions, and learns contextual embeddings of entities, relations and words in descriptions simultaneously, via maximizing the loglikelihood of the observed knowledge. We evaluate our model in the tasks of knowledge graph completion and entity type classification with two benchmark datasets: *FB500K* and *EN15K*, respectively. Experimental results demonstrate that the proposed model outperforms both *TransE* and *DKRL*, indicating that it is both efficient and effective in learning better distributed representations for knowledge bases.

1. Introduction

A typical large-scale knowledge base, such as Freebase [1], mainly contains billions of triplets (*head_entity, relation, tail_entity*), abbreviated as (h, r, t), and each represents a fact that there is a relation r between the two entities (h and t). These triplets conventionally compose a knowledge graph in which each entity is a node, and relations between two entities are regarded as directed edges. This symbolic representation facilitates storing and displaying knowledge, but makes the inference of knowledge infeasible, especially when the volume of knowledge base grows and data becomes sparse.

Therefore, recent research on knowledge representations attempts to address the issue via projecting both entities and relations into a continuous low-dimensional space [10,17]. One canonical approach is *TransE* [2], which solely uses triplets in a knowledge base without requiring extra text to make the inference of knowledge computable. It learns low-dimensional vector representations (embeddings) of both entities and relations by minimizing a margin-based loss function. *TransE* attracts many successive studies [6–8,12,21], not only because of its leading performances, but also owing to fewer parameters required.

However, knowledge is expected to reinforce other intelligent applications, such as question–answering (QA) systems [19] in which unstructured text is also involved. On the other hand, mainstream knowledge repositories, such as Freebase [1] and NELL [3], contain concise entity descriptions and relation mentions in addition. The extra text provides contextual evidence to help learn better embeddings. Therefore, the study of context-enhanced representation learning for knowledge bases becomes prosperous [4,5,20,22]. The cutting-edge method *DKRL* [23] extends *TransE* via enhancing the embeddings with entity descriptions by means of deep neural network models. However, *DKRL* demands extra space to store parameters of inner layers, and relies on more hyperparameters to be tuned.

In this paper, we create a single-layer model which requires much fewer parameters for representation learning of knowledge bases (*RLKB*). *RLKB* measures the probability of each triplet along with corresponding entity descriptions. During the phase of

☆ This paper has been recommended for acceptance by "Eva Armengol".

* Corresponding author.
E-mail address: fanmiao.cslt.thu@gmail.com (M. Fan).

http://dx.doi.org/10.1016/j.patrec.2016.09.005

training, the model learns contextual embeddings of entities, relations and words in descriptions simultaneously via maximizing the loglikelihood of the whole observed knowledge base, so as to encoding those embeddings into the same low-dimensional vector space. We evaluate our model with two benchmark datasets: *FB500K* and *EN15K*. *FB500K* contains nearly 500,000 triplets from Freebase, and it is a wide-spread dataset adopted by many recent studies [2,6–8,12,21] to test the performance of knowledge graph completion. *EN15K* is composed by almost 15,000 entities occurring in *FB500K* for the task of entity type classification [23], and each entity usually belongs to multiple principal entity types in Freebase. Experimental results demonstrate that the embeddings acquired by *RLKB* outperforms both *TransE* and *DKRL* in the tasks of knowledge graph completion and entity type classification, indicating that the proposed model is both efficient and effective in learning better knowledge representations. We also explore the reason why *RLKB* achieves such leap forwards on the two experimental tasks, and find out that the acquired embeddings can obtain semantic relatedness among entities, relations, and even words in descriptions. Intuitively, the semantic relatedness revealed by *RLKB* within knowledge repositories, not only helps narrow down the scope of searching missing entities/relations, but also provides similar representations between entities that share the same types.

Overall, we contribute a single layer neural network model which requires fewer parameters to learn knowledge embeddings. In order to acquire contextual embeddings, both structured knowledge graphs and unstructured text are used, so that the semantic relatedness among entities, relations and even words is captured. These embeddings learnt by our approach not only assist the task of knowledge graph completion, but also produce better features for the task of entity type classification.

2. Related work

The trend of studying distributed representations for knowledge bases has emerged in recent years. A group of papers [2,6–8,12,21] engage in exploring representation models based on structured knowledge graphs without requiring extra text, and we will talk about them in Section 2.1. Section 2.2 is going to give a review on context-enhanced approaches, in which entity descriptions [20,23] and relation mentions [4,5,22] are additionally considered while learning knowledge embeddings.

2.1. Structure-based representation learning

TransE [2] is a classical approach on learning vector representations of both entities and relations solely with knowledge graphs. The approach regards relations between two entities as translating operations in vector spaces, and uses the scoring function $||\mathbf{h}+\mathbf{r}-\mathbf{t}||$ to measure the plausibility of each triplet (h, r, t). Its strength lies in requiring fewer parameters to represent triplets. However, *TransE* cannot cope well with multi-relations between two entities, since these relations tend to gain the same embeddings. To address the issue, Wang et al. propose *TransH* [21] in which two entities are projected into different relation-dependent hyperplanes, so that each relation can distinguish from the others. Fan et al. [6] simply adapt the learning rates along with the number of multi-relations, and achieve great improvements. Several state-of-the-art models, such as *IIKE* [7], *LMNNE* [8] and *TransR* [12], are created to learn better embeddings of entities and relations within knowledge graphs, but none of them considers extra information from text, such as entity descriptions and relation mentions which are included in most knowledge repositories as well.

2.2. Context-enhanced representation learning

Weston et al. [22] firstly concern about encoding words in relation mentions together with entities and relations. They match the mentions with corresponding relations, based on the assumption of distant supervision [15]. In addition, Fan et al. [4,5] leverage the relation mentions already aligned by NELL, and propose several jointly embedding models.

On the other hand, Wang et al. [20] start to align entities with anchors in Wikipedia to obtain contextual descriptions. This framework has constraints of application scenario, because linking by entity names severely pollutes the embeddings of words, and using Wikipedia anchors completely relies on the special data source. Therefore, Xie et al. [23] propose *DKRL* which directly uses the concise descriptions of entities in Freebase, and achieves state-of-the-art performances on the tasks of knowledge graph completion and entity type classification. However, *DKRL* adopts deep neural network models to refine embeddings of entities and descriptions, so that it needs more space to store parameters of inner layers and more hyperparameters to be tuned. Therefore, we design a single-layer model which will be described in the subsequent section to address those issues.

3. Model

Given a knowledge repository Δ which contains enormous number of items (h, r, t, d_h, d_t), where each item is composed by a head entity h with its descriptions d_h, a tail entity t with its descriptions d_t, and a relation r between the two entities, our model aims to maximize the loglikelihood of the observed Δ, expressed by Eq. (1), to obtain the embeddings of entities, relations and words in entity descriptions:

$$\arg\max_{h,r,t,d_h,d_t} \sum_{(h,r,t,d_h,d_t)\in\Delta} \log Pr(h,r,t,d_h,d_t), \tag{1}$$

in which the probability of each item $(h, r, t, d_h, d_t) \in \Delta$ is influenced by two factors:

$$\log Pr(h,r,t,d_h,d_t) = \log Pr(h,r,t) + \log Pr(d_h,d_t|h,r,t). \tag{2}$$

$Pr(h, r, t)$ represents the probability of the observed triplet (h, r, t), and $Pr(d_h, d_t|h, r, t)$ is the conditional probability of observing entity descriptions given the triplet.

If a triplet (h, r, t) is a positive example, we need to make sure that any one of the three objects (h, r and t) is plausible given the other two. In other words, $Pr(h, r, t)$ is the trade-off among the three conditional probabilities:

$$\log Pr(h,r,t) = \frac{\log Pr(h|r,t) + \log Pr(r|h,t) + \log Pr(t|h,r)}{3}. \tag{3}$$

To represent $Pr(h|r, t)$, the conditional probability of observing h given r and t, we firstly adopt the geometric modeling of a triplet proposed by *TransE* [2] which is illustrated by Fig. 1: the relation ($\mathbf{r}$) between two entities is considered as a translation in the vector space from the head entity ($\mathbf{h}$) to the tail entity ($\mathbf{t}$). Furthermore, we define a scoring function Θ as follow:

$$\Theta(h,r,t) = \alpha - \frac{1}{2}||\mathbf{h}+\mathbf{r}-\mathbf{t}||_2^2 \tag{4}$$

to quantify the plausibility of a triplet, and α is a positive bias for structured knowledge. The larger Θ scores, the higher possibility that the evaluated triplet is positive. It is obvious that many incorrect/negative head entities, denoted by h', are likely to replace h and to collapse the triplet. Suppose that the set of these negative head entities is E'_h. We define $Pr(h|r, t)$ as

$$Pr(h|r,t) = \frac{\exp^{\Theta(h,r,t)}}{\exp^{\Theta(h,r,t)} + \sum_{h'\in E'_h} \exp^{\Theta(h',r,t)}}. \tag{5}$$