Considering Sub-graph Structure in Representation Learning

Wentao Fu¹, Xiaoyu Yang² and Xiyao Sun¹*

¹ National & Local joint engineering research center of satellite navigation positioning and location service, Guilin University of Electronic Technology, Guilin, Guangxi, 541004, China
² Lierda Technology Group Co., Ltd., Hangzhou, Zhejiang, 311100, China
¹*Corresponding author’s e-mail: sunxiyan@163.com

Abstract. Representation learning aims to encode knowledge graphs into a low-dimensional vector space. However, since both the out-degree and in-degree of the entities in the knowledge graph follow the power-law distribution, only a small number of entities with higher frequency play a key role in training process, while the others have less effect. This leads to a more serious problem of data sparsity. In this paper, we propose two knowledge graph representation learning models for sub-graph structure fusion. Based on the original model, our models are trained by incorporating the sub-graph structure during training process. By incorporating the sub-graph information into the representation learning model, the structural associations between different relations can be incorporated while considering the semantic association between entities. So that the entities and relations can be modeled more accurately, and the data sparsity problem of the knowledge graph representation learning model can be alleviated effectively.

1. Introduction

Google formally proposed the concept of knowledge graph and introduced it into search engines [1] on May 17, 2012. Knowledge graphs is essentially a huge semantic network, expressed in the form of triples (head entity, relation, tail entity) (short for \( h, r, t \)). For instance, the triple (Hemingway, author, the old man and the sea) illustrates the fact that the author of the novel "the old man and the sea" is Hemingway. In recent years, a large number of knowledge graphs have been constructed, such as YAGO [2], Freebase [3], DBpedia [4], DeepDive [5], etc. Tasks like Question Answering System [6,7], Web Page Ranking [8] and other downstream natural language processing tasks can be solved by Large-scale knowledge graphs. Introducing knowledge graphs for large data analysis is a research hotspot in many vertical industries, such as finance and health care. However, like other types of large-scale dataset, large-scale knowledge graphs follow long-tailed distribution, and with serious data sparsity problem. Therefore, the research on knowledge graph completion technology is of great significance.

As shown in figure 1, it can be found that the out-degree and in-degree of entities in the knowledge graph follows power law distribution through the statistics of FB15k dataset. This shows that the amount of information contained in the triple in the knowledge graph inclines seriously. Only a few entities with high frequency play a key role in training process, while the others have less effect, which resulting in more serious data sparsity problem. In addition, statistics show that there is a shorter path between any two entities, and the average length of the shortest path is about three. If the above characteristics are utilized and the high-order structural information is incorporated into the training and learning process, data sparsely problems can be effectively alleviated. Furthermore, there are many sub-graph structures appearance in semantic networks frequently, in which two, three, four-point sub-graph structures are
common. In the knowledge graph, the high-order sub-graph structure contains not only the association information between entities, but also the association information between relations. It is possible to obtain a more accurate vector representation of entities and relations in the vector space via sub-graph structure.

The triples in the knowledge graph can be regarded as a two-point sub-graph structure, and only a single relation contained in the triple. If we want to learn the relation between different relations, a higher-order sub-graph structure for more information are needed. Three-point sub-graph is the most common high-order sub-graph with relatively simple structure. Since the average shortest path length between any two entities is about three, it is better to capture the relations among entities and relations by the four-point sub-graph structure in the knowledge graphs. It can be considered that the higher-order sub-graph structure has dense data volume and more abundant information. Although such a sub-graph structure does not contain all the information in the knowledge graph, but it expresses more abundant structural features to a certain extent.

Figure 1. Distribution of entities’ out-degree and in-degree in FB15k

2. Our Model
In addition to the existing two-point sub-graph structure, we take two other sub-graph structures into consideration: three-point sub-graph structure and four-point sub-graph structure. The knowledge graph representation learning based on sub-graph structure takes full account of the association information between entities and path information between relations, which provides more abundant constraint information for the model. For example, according to the path of a relation: "father + father" between "Li Dan" and "Li Zhi", extracted from (Li Dan, father, Li Xian) and (Li Xian, father, Li Zhi), combining with the triple group (Li Dan, grandfather, Li Zhi), the three-point sub-graph structure additionally provides the implied Relevant information, a new relation: "father + father = grandfather". In addition, supposing that there is a triple (Li Zhi, Father, Li Shimin) in the knowledge graph and another triple (Li Dan, Great Grandfather, Li Shimin), in fact, the four-point sub-graph structure provides more abundant information about the relation, that is, "Father + Father + Father = Great Grandfather". Obviously, sub-graph structure implies the transferability between relations. Therefore, if we can find an appropriate method to model sub-graph structure, it will greatly improve the accuracy of knowledge representation.

Generally speaking, sub-graph structure includes rich semantic association between nodes, interactions between nodes and relations, as well as semantic relations between relations. For simplicity, this paper chooses the simplest sub-graph structure, that is, the three-point sub-graph structure and the four-point sub-graph structure in which relations point to the same. Moreover, we choose the method which best conforms to the translation hypothesis between vectors in the triple to combine multiple relational vectors. As shown in figure 2, for sub-graph (a), the constraint is $e_i + r_y = e_f$. For sub-graph...
(b), in addition to constraint $e_i + r_y = e_j$, two additional constraints $e_i + r_x = e_s$ and $e_s + r_y = e_j$ are added. For sub-graph (c), in addition to constraint $e_i + r_y = e_j$, three additional constraints $e_i + r_y = e_s$, $e_s + r_y = e_i$, and $e_j + r_y = e_j$ are added.

$$\Delta = \{e_i, e_x, e_j, r_y, r_x, r_y\}$$

Figure 2. Examples of sub-graph structures in the knowledge graph

2.1. Knowledge Atlas Representation Learning Model Fusing Three-Point Sub-graph Structure

Based on the framework of the classical representation learning model TransE\[9\], we propose a knowledge atlas representation learning model (TransT) which integrates three-point sub-graph structure. In the TransE model and the TransT model, the entity set and the relation set are denoted as $E$ and $R$, respectively, and a knowledge graph is represented by a triple set. Then, starting from each triple, we can find three-point sub-graphs related to each triple and form a set of three-point sub-graphs. Among them, $e$ denotes entities, $r$ denotes relations, $i$, $x$ and $j$ denotes the numbers of entities, $ij$, $ix$ and $xj$ denotes the relations between entities numbered $i$ and $j$, $i$ and $x$, and $x$ and $j$, respectively. For each triple $(h, r, t)$, TransE regards relational vector $r$ as a translation operation between head and tail entity vectors. For each established triple, a smaller score means better model, while for the non-established triple, the score has to be as large as possible. In this section, we propose a knowledge graph representation learning method which can be used for integrating three-point sub-graph structure. We hope that by learning the three-point sub-graph structure, the energy score of the established three-point sub-graph is as small as possible, and the non-established three-point sub-graph is as large as possible.

Based on TransE, the proposed model will be integrated into the three-point sub-graph structure of knowledge graph to improve the accuracy of representation of learning entities and relational vectors. The score function of TransE is shown in formula 1.

$$f_e(h, t) = \|h + r - t\|^2_{\ell_2}$$

(1)

Obviously, TransE only considers a two-point structure which consisting of two entities and one relation. The proposed representation learning model incorporating three-point sub-graph structure which contains three entities and three relations, as shown in figure 2(b). The scoring function as shown in formula 2.

$$f(\Delta) = \|e_i + r_y - e_j\|^2_{\ell_2} + \|e_i + r_x - e_s\|^2_{\ell_2} + \|e_x + r_{ij} - e_j\|^2_{\ell_2}$$

(2)

Each item represents the scores of three triples in a three-point sub-graph, and the possibility of each triple established measured by learning the semantic similarity between entities.

The model optimizes the embedding vector representation of entities and relations through the objective function based on maximum interval.

$$L = \sum_{h \in E} \sum_{r \in R} \max(0, f(\Delta) + \gamma - f(\Delta'))$$

(3)

Among them, the negative case three-point sub-graph structure is represented. The three-point sub-graph contains three triples, so we need to construct negative examples for each triple, which is constructed in the same with TransE.

**Algorithm:** Learning TransT

**Input:** Training set $G = \{(h, r, t)\}$, entity and relation set $E$ and $R$, three-point sub-graph $T = \{\Delta\}$,

$\Delta = (e_i, e_x, e_j, r_y, r_x, r_y)$, margin $\gamma$, entity embedding dim $m$, relation embedding dim $n$, learning rate $\lambda$.
1 initialize The embedding vectors for each entity and relation as a result of TransE;
2 loop
3 \( S_{\text{batch}} = \text{sample} (S, b) \) : // sample a minibatch of size b
4 \( P_{\text{batch}} = \emptyset \) : // initialize the set of pairs of three-point sub-graph structure
5 for \( \Delta \in T_{\text{batch}} \) do
6 \( \Delta' := \text{sample}(T') \) : // sample an invalid three-point sub-graph structure
7 \( P_{\text{batch}} := P_{\text{batch}} \cup \{(\Delta), (\Delta')\} \):
8 end for
9 \( \sum_{\Delta, \Delta' \in P_{\text{batch}}} \nabla \left[ \gamma + f(\Delta) - f(\Delta') \right] \) : // update embedding \( h, r, t \).
10 end loop

2.2. Knowledge Atlas Representation Learning Model Fusing Four-Point Sub-graph Structure

Based on the framework of the classical representation learning model TransE, we propose a knowledge representation learning model (TransF) that integrates four-point sub-graph structure. In TransE and TransF model, the entity set and the relation set are denoted as E and R respectively, and a knowledge graph is represented by a triple set. Then, starting from each triple, we can find four-point sub-graphs related to each triple and form a set of four-point sub-graphs. Where \( e \) denotes entities, \( r \) denotes relations, \( i, x, y \) and \( j \) denotes the numbers of entities, \( ij, ix, xy \) and \( yj \) denotes the relations between entities numbered \( i \) and \( j \), \( i \) and \( x \), \( x \) and \( y \), and \( y \) and \( j \) respectively. In this section, we propose a knowledge graph representation learning method that integrates four-point sub-graph structure. We hope that by learning four-point sub-graph structure, the energy score of the established four-point sub-graph is as small as possible, and the non-established four-point sub-graph is as large as possible at the same time.

The proposed model in this subsection will be integrated into the four-point sub-graph structure of knowledge graph on the basis of TransE to improve the accuracy of the representation of entities and relational vectors. The scoring function of TransE is shown in formula 4.

\[
f_e(h, t) = \|h + r - t\|^2_{L_2} \quad (4)
\]

Obviously, TransE only considers a two-point structure consisting of two entities and one relation. The proposed representation learning model considers four-point structure which including four entities and four relations. As shown in figure 2(c), the scoring function as shown in formula 5.

\[
f(\Delta) = \|e_i + r_y - e_j\|^2_{L_2} + \|e_i + r_x - e_x\|^2_{L_2} + \|e_y + r_x - e_x\|^2_{L_2} + \|e_x + r_j - e_y\|^2_{L_2} \quad (5)
\]

Each item represents the scores of four triples in the four-point sub-graph. The possibility of each triple established is measured by leaning the semantic similarity between entities.

The model optimizes the embedding vector representation of entities and relations through the following objective function based on maximum interval as shown in formula 3.

\[
L = \sum_{\Delta \in T} \sum_{\Delta' \in T'} \max(0, f(\Delta) + \gamma - f(\Delta')) \quad (6)
\]

Where the negative examples four-point sub-graph structure is represented. The construction method is same with TransT.

**Algorithm:** Learning TransF

**Input:** Training set \( G = \{(h, r, t)\} \), entity and relation set \( E \) and \( R \), four-point sub-graph \( T = \{\Delta\} \), \( \Delta = (e_i, e_x, e_y, e_j, r_x, r_y, t_y) \) margin \( \gamma \), entity embedding dim \( m \), relation embedding dim \( n \), learning rate \( \lambda \).

1 initialize The embedding vectors for each entity and relation as a result of TransE;
2 loop
3 \( S_{\text{batch}} := \text{sample} (S, b) \); // sample a minibatch of size b
4 \( P_{\text{batch}} := \emptyset \); // initialize the set of pairs of four-point sub-graph structure
5 for \( \Delta \in T_{\text{batch}} \) do
6 \( \Delta' := \text{sample}(T') \); // sample an invalid four-point sub-graph structure
7 \( P_{\text{batch}} := P_{\text{batch}} \cup \{ (\Delta), (\Delta') \} \);
8 end for
9 \( \sum_{\Delta, \Delta' \in P_{\text{batch}}} \nabla [\gamma + f(\Delta) - f(\Delta')]. \) // update embedding h.r.t.
10 end loop

3. Evaluation
In order to verify the performance of the knowledge graph representation learning method proposed in this paper, two typical knowledge graph completion tasks link prediction and triples classification will be tested on FB15k.

First, some parameter settings are introduced: the range of learning rate in TransT and TransF is \{0.1, 0.01, 0.02, 0.001, 0.0001\}; the range of maximum interval distance between positive and negative samples are \{1,4\}; the range of dimension \( m \) of entity embedding and dimension \( n \) of relational embedding is \{20, 30, 50, 100\}; the range of batch size \( B \) is \{30, 120, 480, 1440,4800\}, and distance measure D.S is 1. The optimal parameter configuration of this experiment is determined according to the average ranking of the model on the verification set. The optimal parameter configuration of TransT is as follows: \( \lambda = 0.001 \), gamma = 1, \( M = D = 100 \), \( B = 4800 \). The best parameters of TransF are: \( \lambda = 0.001 \), gamma = 1.2, \( M = D = 100 \), \( B = 4800 \).

| Model | TransE (Raw) | TransT (Raw) | TransF (Raw) | TransE (Filt) | TransT (Filt) | TransF (Filt) |
|-------|--------------|--------------|--------------|--------------|--------------|--------------|
| Hits@10 | 34.9 | 42.8 | 43.7 | 47.1 | 68.0 | 68.6 |

The result of the top 10 percentage of entities in link prediction are shown in table 1. From the table 1, we can see that TransT (Raw) is 7.9% higher than TransE (Raw), and TransF (Raw) is 8.8% higher than TransE (Raw). TransT (Filt) is 20.9% higher than TransE (Filt), and TransF (Filt) is 21.5% higher than TransE (Filt). The results show that the representation learning models TransT and TransF proposed in this paper are significantly better than TransE in entity link prediction tasks, and the results of TransF are better than TransT.

| Model | TransE | TransT | TransF |
|-------|--------|--------|--------|
| Hits@10 | 34.9 | 42.8 | 43.7 |

The results of the classification accuracy of triples classification are shown in table 2. TransT is 7.4% higher than TransE, while TransF is 7.7% higher than TransE. The results indicate that the representation learning models TransT and TransF proposed in this paper are significantly better than TransE in terms of triples classification tasks, and TransF is better than TransT.

References
[1] Su Y, Yang S, Sun H, et al. Exploiting Relevance Feedback in Knowledge Graph Search. Proceedings of the 21th International Conference on Knowledge Discovery and Data Mining. ACM, 2015:1135-1144.
[2] Suchanek F M, Kasneci G, Weikum G. Yago: a core of semantic knowledge. International Conference on World Wide Web. OAI, 2007: 697-706.

[3] Bollacker K, Evans C, Paritosh P, et al. Freebase: a collaboratively created graph database for structuring human knowledge. Sigmod Conference. 2008: 1247-1250.

[4] AUER, Al S E. DBpedia: A Nucleus for a Web of Open Data. Semantic Web, International Semantic Web Conference, Asian Semantic Web Conference, Iswc + Aswc, Busan, Korea, November. DBLP, 2007: 722-735.

[5] Zhang C. DeepDive: A Data Management System for Automatic Knowledge Base Construction [J]. Dissertations & Theses - Gradworks, 2015.

[6] Jun-Wei B, Nan D, Zhao Y, et al. Constraint-Based Question Answering with Knowledge Graph [C]. COLING 2016: 2503-2514.

[7] Cui W, Xiao Y, Wang H, et al. KBQA: Learning Question Answering over QA Corpora and Knowledge Bases [J]. Proceedings of the VLDB Endowment, 2017, 10(5):565-576.

[8] Du Y J, Hu Q, Li X, et al. Ranking Web Page with Path Trust Knowledge Graph. Big Data and Machine Learning Techniques. Springer International Publishing, 2015: 66-75.

[9] Antoine B, Nicolas U, Alberto G, Jason W, Oksana Y. Translating Embeddings for Modeling Multi-relational Data. NIPS (2013)2787-2795.