Regarding some personal viewpoints on knowledge graph

A cute kid.
A digital version of the manuscript.

I study research on the representations of entities and relations in knowledge graphs for predicting links between triplets of a head entity, a tail entity, and their corresponding relation. Previous models were mostly based on negative sampling and were overly academic. In this paper, I present a simple model for knowledge graph embedding named TransQ, which is an abbreviation of "transfer quantity". Specifically, the TransQ model first performs a Hadamard product between the representation of the head entity and the corresponding representation of the relationship, and then maps it to the representation of the tail entity through a linear transformation. Furthermore, I analyzed from a semantic perspective to determine whether the entities and relationships in the dataset satisfy the logical concepts in the field of discrete mathematics. The experimental results show that, on the benchmark dataset, the proposed TransQ model can perform link prediction on knowledge graphs.

Introduction

Knowledge graphs are usually built on multiple triples, where each triplet includes the head entity, tail entity, and the relationship. In recent years, more and more scholars have entered the field of knowledge graph-related tasks. Link prediction refers to the task of predicting the entities and relationships in knowledge graphs, which is an indispensable part of knowledge graph construction and completion.

In this paper, I proposed the model named TransQ for knowledge graph link prediction task. My motivation comes from the Gate Recurrent unit, which consists of Hadamard product to generate or for discrimination task. Given a triplet $(\mathbf{h}, \mathbf{r}, \mathbf{t})$, let the $\mathbf{h}$, $\mathbf{r}$, and $\mathbf{t}$ be represented by a string of real numbers, like $\{x_1, x_2, ..., x_n\}, x_i \in \mathbb{R}$; define the logits of the $\mathbf{t}$ is $P(x)$, expect that $P(x) \sim \mathbf{h} \odot \mathbf{r} \cdot \mathbf{W}$; the final score function to be minimized is equation 1:

$$\mathbb{E}_{\mathbf{t} \sim \mathrm{P}}[-\log(\mathrm{Q}(\mathbf{h} \odot \mathbf{r}))] \tag{1}$$

where the P represent the distribution of the correspondding true label and the Q is the operation that makes the $\mathbf{h} \odot \mathbf{r}$ output its logical value. The $\mathbb{E}$ is the meaning of CrossEntropy.

The main contributions of this paper can be summarized in three points:

- Proposed the TransQ model.
- Perform dimensionality reduction visualization of entities and relationships in the model.
- Perform a visualization of the discrepancy analysis on the link prediction task.

Related work

In this section, I use examples to prove whether the entities and relationships in the WN18RR dataset satisfy the properties of binary relations in discrete mathematics.

Defination. Given a knoeledge graph dataset, let $\mathcal{E}$ denote the set of entities and $\mathcal{R}$ denote the set of relations, the number of the relations is 11, then a knowledge graph is a collection of factual triplets $(\mathbf{h}, \mathbf{r}, \mathbf{t})$, where $\mathbf{h}$, $\mathbf{t} \in \mathcal{E}$, and $\mathbf{r} \in \mathcal{R}$.

Property 1: symmetric. $\forall x, y \in \mathcal{E}, \forall r \in \mathcal{R}; r(x, y) \Rightarrow r(y, x)$.

Derivation process: Let the "activist" is x, "active" is y, and the r is "derivationally_related_from". The meaning of the $r(x, y)$ is the word "activist" is derivatoinally related from "active". And the word "active" is not derivatoinally related from the "activist". Further, the present x,y and r also satisfy the equation 2.

$$\exists x, y \in \mathcal{E}, \exists r \in \mathcal{R}; r(x, y) \nRightarrow r(y, x) \tag{2}$$

Thus, the WN18RR dataset is not satisfy the property of the symmetric.

Property 2: antisymmetry. $\forall x, y \in \mathcal{E}, \forall r \in \mathcal{R}; r(x, y) \Rightarrow \neg r(y, x)$.

Derivation process: Let the "daily" is x, "everyday" is y, and the r is "_similar_to". The meaning of the $r(x, y)$ is the word "daily" is similar to "everyday". And the word "daily" is also similar to the word "everyday". Further, the present x,y and r also satisfy the equation 3.

$$\exists x, y \in \mathcal{E}, \exists r \in \mathcal{R}; r(x, y) \nRightarrow \neg r(y, x) \tag{3}$$

Thus, the WN18RR dataset is not satisfy the property of the antisymmetry.

Property 3: inverse. $\forall x, y \in \mathcal{E}, \exists r_1, r_2 \in \mathcal{R}; r_2(x, y) \Rightarrow r_1(y, x)$.

Derivation process: The relations in wordnet like hypernym and hyponym are inverse. But the relations in the version of the dataset which I used is not exist the inverse relations. Thus, the WN18RR dataset is not satisfy the property of the inverse.

Property 4: composed. $\forall x, y, z \in \mathcal{E}, \exists r_1, r_2, r_3 \in \mathcal{R}; r_2(x, y) \wedge r_3(y, z) \Rightarrow r_1(x, z)$.

Derivation process: The relations in the version of the dataset which I used is not exist the composed relations. Thus, the WN18RR dataset is not satisfy the property of the inverse.

Methodology

The structure of the model involves performing a Hadamard product operation on the representation of the head entity and the representation of the relationship, followed by outputting the corresponding label distribution through a linear layer. The model is then trained by minimizing the cross-entropy loss with the true label.

Experiments

I evaluate my proposed model on WN18RR knowledge graph. The statistics of these knowledge graph is summarized into Table 1.

Dataset	**#entity**	**#relation**	**#training**	**#validation**	**#test**
WN18RR	40,943	11	86,835	3,034	3,134

Table 1: Number of entities, relations, and observed triples in each split for the benchmark.

Hyperparameter Settings. Table 2 demonstrates the value of the hperparaameters of the model.

Hyperparameters	**value**
batch size	1
epoch	100
learning rate	0.001
Embedding dimension	300

Table 2: Hyperparameters of the model.

Result and Analysis

Main Results. Table 3 demonstrates the model's accuracy on the WN18RR test dataset. Figure 1 illustrates the changes in loss value between training epochs. It can be seen that the loss value of the model is decreasing.

TransQ on the Dataset	Accuaracy
WN18RR	$\frac{19}{3134}$

Table 3: Hyperparameters of the model.

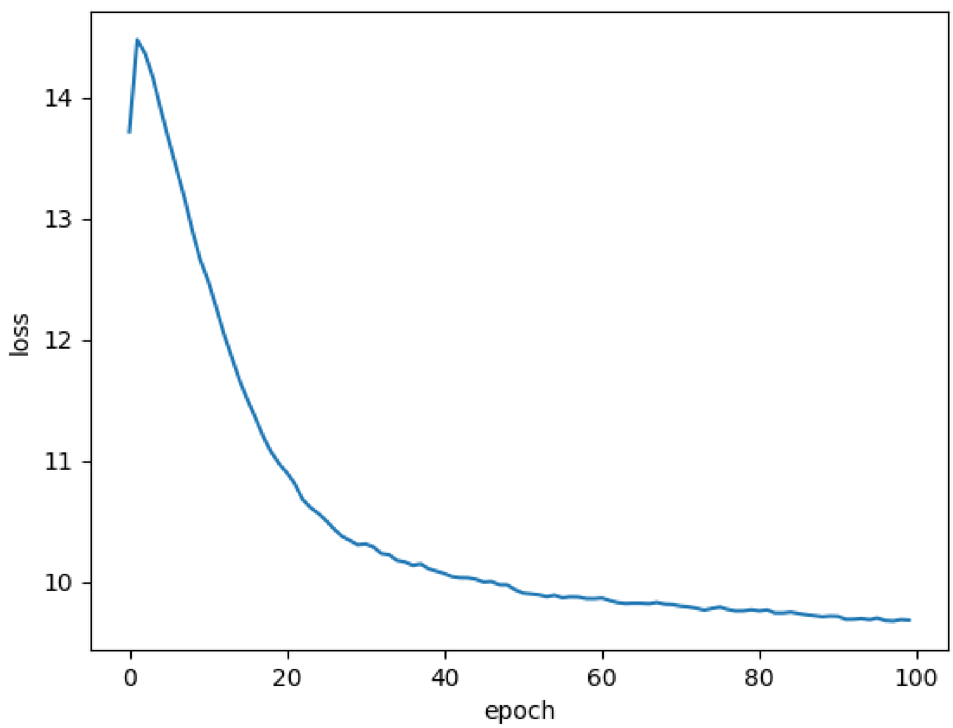

Figure 1: The model loss plot.

Dimension reduction analysis

Figure 2 demonstrates the results of the model dimensionality reduction analysis. The first two figures in Figure 2 show the dimensionality reduction and visualization of entity representations and relationship representations using SVD, respectively. The third figure in Figure 2 shows the dimensionality reduction analysis results after merging entities and relationship representations. Before merging, I used softmax to scale the representations of entities and relationships to the closed interval $[0, 1]$. The fourth figure in Figure 2 shows the results of the dimensionality reduction analysis of the logical values between the head entity and the relationship.

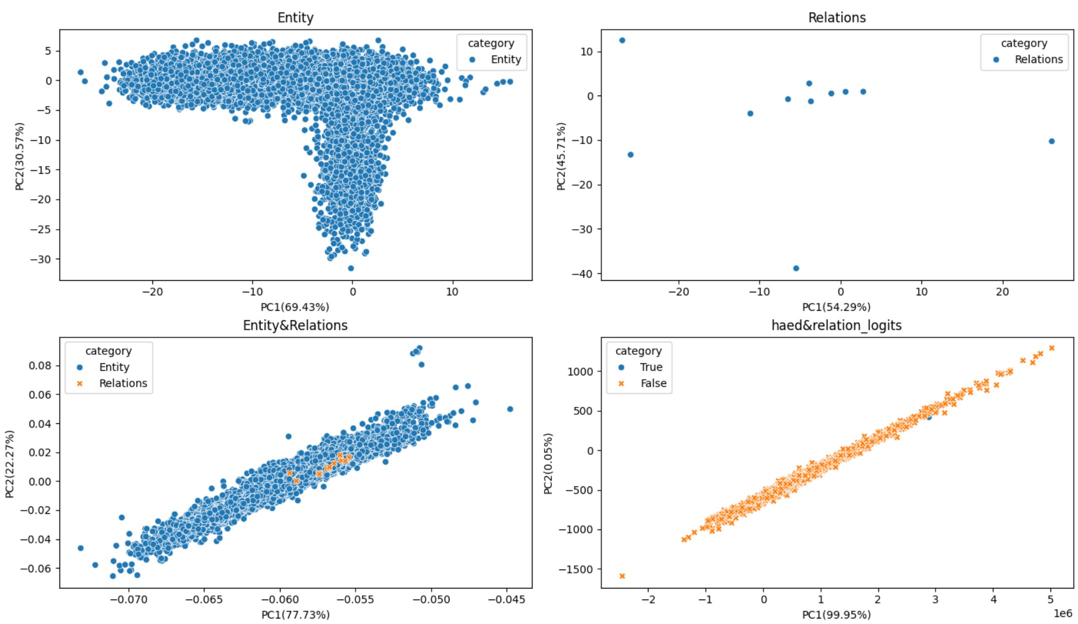

Figure 2: The dimension reduction analysis.

Discrepancy analysis

Figure 3 demonstrates the results of the model difference analysis. Towards displaying normal difference analysis results, I modified the calculation of the P_value. Specifically, I divided the predicted value by the sum of the logical values of the current sample for the final P_value. The "log2 fold change" refers to dividing the predicted value of a sample by the corresponding logical value of the true label and then performing a logarithmic operation with base 2. I also used shift operations to shift the "Up" samples to the right and the "Down" samples to the left, ultimately making the results clearer than the original results. Furthermore, the “Down” samples are more concentrated than the “Up” samples, while the “Up” samples are more dispersed.

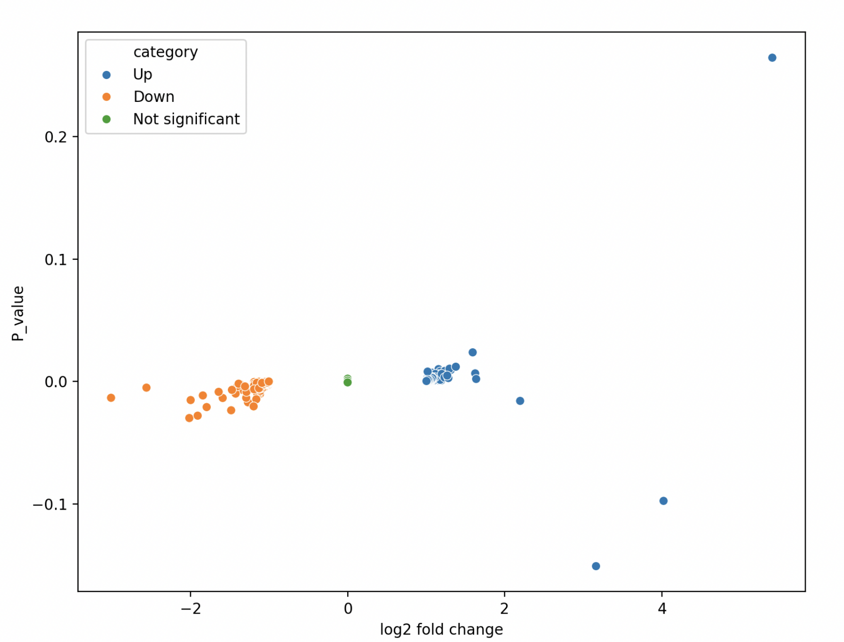

Figure 3: The discrepancy analysis.

Conclution

This paper shares some personal views on link prediction tasks in knowledge graphs.