Notes on Holographic Embedding of Knowledge Graph

Jianing

April 2019

1 Concepts

1.1 Knowledge Graph

Knowledge graph (G, V, E, R) is a knowledge bases represented in the form of multigraph(i.e. allow multiple edges between two nodes) with entities as nodes and binary relations as typed edges. For example, WordNet is a KG with words(e.g. cat) as nodes, relations between two words as typed edges. One example of such relation is a super-subordinate between "bed" and "furniture" where "furniture" INCLUDEs "bed". Another example is a Meronymy relation between "leg" and "chair" where "leg" is PART of "chair".

In a Knowledge Graph, every edge is a triple (s, p, o) which can be read as SUBJECT s has RELATION R_p to OBJECT o. For each type of relations R_p in the KG, we can associate a characteristic function $\phi_p : V \times V \to \{\pm 1\}$ which indicates whether or not relation R_p is present between ordered pair $(s, o) \in V \times V$.

1.2 Compositional representations of Knowledge Graph

We want to represent a Knowledge Graph in a denser way, that is, to represent nodes and typed relations in terms of vectors. However this representation should have the advantage that allows us to predict existence of relations between nodes. We want the following mapping

$$s \in V \to e_s \in \mathbf{R}^d \tag{1}$$

$$p \in R \to r_p \in \mathbf{R}^d \tag{2}$$

such that a target function is minimized. We introduce the target function below.

As usual in machine learning, we want to model the following probability function

$$Pr(\phi_p(s, o) = 1|\theta) \tag{3}$$

which is the probability of a possible relation in KG. A usual way to do that is to use a compositional operator and sigmoid function.

$$Pr(\phi_p(s,o)=1|\theta) := \sigma(r_p^\intercal(e_s \circ e_o)) \tag{4}$$

where operator $\circ$ is called the compositional operator. The task is then to learn parameters θ, that is representation e_n's and r_p's.

With the expression of probability expressed in **??**, the target function is then the cross entropy.

$$E_{(s,p,o)\sim q}(-log(P)) = -\sum_{y=1} log(\sigma(r_p^\intercal(e_s \circ e_o))) - \sum_{y=-1} log(1-\sigma(r_p^\intercal(e_s \circ e_o))) \tag{5}$$

$$= -\sum_{i=1}^{m} log(1+exp(-y_i r_p^\intercal(e_s \circ e_o))) \tag{6}$$

2 Algorithms

Given a set of true or false relations (x_i, y_i)s in a KG where $x_i \in V \times V \times R$ is a relational triple and $y_i \in \{\pm 1\}$ is whether the relation exists or not, we want to learn compositional representations of the KG. This is a supervised binary classification task. Various compositional functions has been tried. For example, in RESCAL, tensor product $e_s \otimes e_o$ is used as composition function. Another way is to concatenate e_s and e_o and use a projection W to project onto a lower dimensional space. The composition function used in HoLE is a circular correlation:

$$a \circ b := a \star b \tag{7}$$

where $\star : R^d \times R^d \to R^d$ is defined as

$$[a \star b]_k = \sum_{i=0}^{d-1} a_i b_{(k+i)\, mod\, d} \tag{8}$$

3 Train

3.1 Dataset

WN18: subset of WordNet.

FB15K: FreeBase

3.2 Evaluation

To query for example masking one of the (s,p,o) in the true triple, we can associate a score to the ranking given by our model. For example, if o is masked and our model gives ranking $o_1, o_2, \cdots$, and the true o has rank $rank_{(s,p)}(o)$, then we associate the following score to this ranking

$$\frac{1}{rank_{(s,p,\_)}(o)}. \tag{9}$$

For $|Q|$ such query, the score is

$$MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank(Q)}. \tag{10}$$

Another evaluation metric proposed is hit at n which is the ratio of the true label within rank n in the ranking proposed by the model. For example, hit at 1 is the ratio of the true label proposed at the first place by the model.

References