Energy Based Ontology Alignment

The problem of ontology alignment is to map two discrete entities or relation from different sources over each other two get a unified representation. Each such mapping has to be probabilistic in nature as it is impossible to map entities or relations with centainity. So the problem of ontology alignment can be described as

$$M(x_{kb:1}, x_{kb:2}) = \begin{cases} 1 & \underset{x_{kb:2}}{arg\ max}\ P(x_{kb:1} \equiv x_{kb:2}) \\ 0 & otherwise \end{cases} \tag{1}$$

where $x_{kb:1}$ and $x_{kb:2}$ are either entities or relations from two KBs $KB1$ and $KB2$ and $M(x_1, x_2)$ is matching matrix where entities are 1 is there is a match or 0 otherwise. Ontology can be treated as a triple of facts of the form (e_l, r, e_r) where e_l and e_r are entities and r is a relation that join them. In our case we don't discriminate between entities and literals so we can use same notation for both.

Our task now is to model the probability measure or function in 1. We will use gibbs measure to model this probability in following way. Lets say the $E(x_{kb:1}, x_{kb:2})$ be the energy of the assignment $x_{kb:1} = x_{kb:2}$, we can model the probability of this alignment as

$$P(x_{kb:1} = x_{kb:2}) = \frac{exp(-E(x_{kb:1}, x_{kb:2}))}{\sum_{x \in X_{kb:2}} exp(-E(x_{kb:1}, x))}$$

For estimation task we don't need the normalizing term as it will be same for all, so the we can rewrite 1 as

$$\underset{x_{kb:2}}{argmax}\ exp(-E(x_{kb:1}, x_{kb:2})) \tag{2}$$

This can be restated as energy minimization problem instead of probability maximization. so rewriting 2, we get

$$\underset{x_{kb:2}}{arg\ min}\ E(x_{kb:1}, x_{kb:2}) \tag{3}$$

Uptil now we have reduced the problem of alignment to minimzation of pairwise energy of two matching entities or relations. Now we need to define an seperate

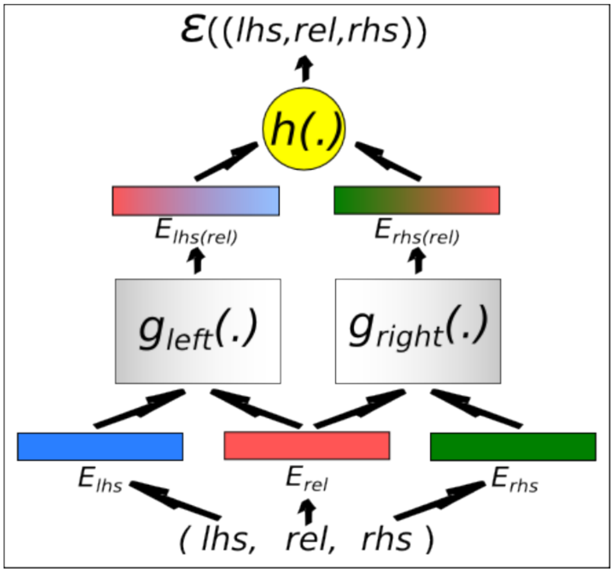

Figure 1: Architecture for Semantic Matching Energy

energy functions E_{rel}, E_{ent} for relations and entities such that the energy is minimized for similar relations and entities from different KBs respectively.

The characterstics of E_{rel} and E_{ent} should be as follows:

1. E_{ent} should be symmetic where as E_{rel} should not be. The reason for this is that relation $r_{kb:1}$from $KB1$ might be a sub-relation of of relation $r_{kb:2}$ from $KB2$.

2. E_{ent} and E_{rel} should be minimum for matching pair.

First challenge we face in defining this energy function is that entities and relations being discrete, there is no direct way to measure to the similarities. To overcome this problem we need to project both the knowledge bases in the same latent space. In this latent space we define an energy function on a triple such that energy is minimum for valid triple. In similar setting to [2], we try to emebed relations and entities in space using architecture defined in the paper in such a way that mimizes the energy of a valid triple. We now discuss this procedure in brief.

The process is two steps. In first step we embed the entities from both the KBs into latent space satisfying the constraint of minimum energy for valid triple. This is followed by alignment by searching for entities or relations that when replaced minimizes the energy.

Let E be a embedding matrix and (e_{lhs}, r, e_{rhs}) be a valid triple from a knowledge base. The embedding can be calculated as

$$(E_{lhs}, E_{rel}, E_{rhs}) = (Ee_{lhs}, Er, Ee_{rhs})$$

Now we calculate relation dependent embedding of left and right embedding $E_{lhs(rel)}$ and $E_{rhs(rel)}$ as

$$E_{lhs(rel)} = g_{left}(E_{lhs}, E_{rel})$$
$$E_{rhs(rel)} = g_{right}(E_{rhs}, E_{rel})$$

The energy function ξ is calculated as

$$\xi(e_{lhs}, r, e_{rhs}) = h(E_{lhs(rel)}, E_{rhs(rel)})$$

We can start by using the bilinear function for g_{left},g_{right} and dot product for h.

The training objective is as follows. Let (e_i^l, r_i, e_i^r) be a valid triplet. Let D be a set of all triplets from both the KBs.

$$\xi(e_i^l, r_i, e_i^r) < \xi(e_j^l, r_i, e_i^r), \forall j : (e_j^l, r_i, e_i^r) \notin D \tag{4}$$
$$\xi(e_i^l, r_i, e_i^r) < \xi(e_i^l, r_j, e_i^r), \forall j : (e_i^l, r_j, e_i^r) \notin D \tag{5}$$
$$\xi(e_i^l, r_i, e_i^r) < \xi(e_i^l, r_i, e_j^r), \forall j : (e_i^l, r_i, e_j^r) \notin D \tag{6}$$

This is achieved by max-margin training between valid and pseudo negative triple. The loss function is given by

$$\sum_{x_i \in D} \sum_{\hat{x_i} \in Q(x_i|\hat{x_i})} max(\xi(x) - \xi(\hat{x_i}) + 1, 0) \tag{7}$$

where $Q(\hat{x_i}|x_i)$ is the corruption process that transforms a positive triplet to a pseudo negative example.

The pseudo-negative triple is created by randomly selecting an entity or relation in a valid triplet and replace it from a random relation or entity respectively. Training algorithm can be defined as follows:

1. Select a positive training triplet x_i at random from set D.
2. Randomly select a constraint from 45 or 6.
3. Create a negative example by randomly selecting an left entity, relation or right entity e_{lhs}^{neg},r^{neg} or e_{rhs}^{neg} depending upon the constraint and replacing it with valid entity or relation in x_i. Lets call this negative sample as $\hat{x_i}$.
4. If $\xi(x_i) < \xi(\hat{x_i}) - 1$, make a stochastic gradient step to minimize the criterion 7

5. Enforce the constraint that each embedding vector is normalized, $||E_i|| = 1 \forall i$.

Now once we have the energy functions on the triplets we need to define E_{ent} and E_{rel}. They can be defined as follows:

$$E_{ent}(e_{kb:1}, e_{kb:2}) = \sum_{(e_{kb:1}, r, e_{rhs}) \in D} \xi(e_{kb:2}, r, e_{rhs}) + \sum_{(e_{lhs}, r, e_{kb:1}) \in D} (e_{lhs}, r, e_{kb:2})$$
$$E_{rel}(r_{kb:1}, r_{kb:2}) = \sum_{(e_{lhs}, r_{kb:1}, e_{rhs}) \in D} \xi(e_{lhs}, r_{kb:2}, e_{rhs})$$

The process here is very similar to one followed in [1], only difference being the we define entity and relation matching energy. The matches can be found by using equation 3 and 1.

References

[1] Antoine Bordes, Xavier Glorot, Jason Weston, and Yoshua Bengio. Joint learning of words and meaning representations for open-text semantic parsing. In *International Conference on Artificial Intelligence and Statistics*, pages 127–135, 2012.

[2] Antoine Bordes, Xavier Glorot, Jason Weston, and Yoshua Bengio. A semantic matching energy function for learning with multi-relational data. *Machine Learning*, 94(2):233–259, 2014.