Knowledge Integration

Task definition

In this part, we fuse the two set of triplets into one set, which means we should integrate the entities from structured data and unstructured data into one knowledge graph. It is not means that we can just add one set to the other. Since there are many entities may be represent the same one in nature in both sets. So the problem is how can we make the entities which are the same into one and simply add the last, i.e. give two set of entities S1={e11,e12,....,e1n} and S2={e21,e22,....,e2n}, e is an entity. Consider that there are some e in S1 and S2 have the same meaning, we should fuse the two set and integrate the entities which have the same meaning.

Approach

The theory we use to integrate the two knowledge graph is Entity Alignment. For Entity Alignment, we should find which pair of entities are representing the same thing in the nature.

To this end, we first embed the entities into low-dimension space to get the embedding of every entity. So we need to embed the whole knowledge graph and the approach we use is DeepWalk algorithm[1]. We use deepwalk to embed all entities in both set of triplets. And we set a threshold to judge whether two entities are representing the same thing, i.e., if the Euclidean distance between the embedding of two entities, then they are the same. If two entities are the same, we set the id of the entity in the second graph to be the same as the first entity. Otherwise, we just add the entity in the second set to the first set and set the id to be Len(S1)+1.

Experiment

In our experiment, we use the approach we introduced above the integrate the triplets extracted from the structured data and the unstructured data. The final result can be seen in Table 1. From the experiment result we can find that there are--- entities in the two set which are representing the same thing and finally we can --- entities and ---- triplets totally.

Table 1. The number of entities and triplets

|  |  |  |  |
| --- | --- | --- | --- |
| Set  Type | S1 | S2 | Final Set |
| Entity |  |  |  |
| Triples |  |  |  |

References

[1] Perozzi, B., Al-Rfou, R., & Skiena, S. (2014, August). Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 701-710).