Predicting relations of embedded RDF entities by Deep Neural Network

Yohei Onuki[1], Tsuyoshi Murata[1], Shun Nukui[1], Seiya Inagi[2], Xule Qiu[2], Masao Watanabe[2], and Hiroshi Okamoto[2]

[1] Department of Computer Science, School of Computing
Tokyo Institute of Technology, W8-59 2-12-1 Ookayama, Meguro, Tokyo, 152-8552 Japan
[2] Research & Technology Group, Fuji Xerox Co., Ltd.
6-1 Minatomirai, Nishi-ku, Yokohama, Kanagawa, 220-8668 Japan

Abstract. The goal of our research is to predict a relation (predicate) of two given RDF entities (subject and object). Link prediction between entities is important for developing large-scale ontologies and for knowledge graph completion. TransE and TransR have been proposed as the methods for such a prediction. However, TransE and TransR embed both entities and relations in the same (or different) semantic space(s). Since entity embedding is enough for predicting relations, we propose a method for predicting a predicate from a subject and an object by using a Deep Neural Network (DNN), and developed RDFDNN. RDFDNN embeds entities only; given subject and object are embedded and concatenated to predict probability distribution of predicates. Experimental results showed that predictions by RDFDNN are more accurate than those by TransE and TransR. Although RDFDNN learns from RDF triples only, its accuracy is comparable to that of DKRL which uses both RDF triples and entity descriptions for learning.

1 Introduction

Ontology learning is one of the important topics for developing the Semantic Web. In general, there are many entity pairs where the relations between them are unknown [2][5]. If we can predict such relations accurately, we can augment a given ontology. Since many Semantic Web data (such as Google's Knowledge Graph) are already available, techniques for predicting relations between entities are important for developing large-scale ontologies.

The goal of our research is to predict relations between two given entities in Resource Description Framework (RDF) accurately. RDF is the framework for representing Web resources, and each triple in RDF is composed of three entities (subject, predicate, and object). Subject and object are entities, and predicate is the relation between the entities. Suppose (Tokyo, is-capital-of, Japan) is an example of such a triple. We would like to predict "is-capital-of" when "Tokyo" and "Japan" are given. For this purpose, we propose a method for predicting a predicate from a subject and an object by using a Deep Neural Network (DNN), and developed RDFDNN. The code is available at https://github.com/yo0826jp/RDFDNN.

2 RDFDNN

RDFDNN predicts the relation between two entities represented as a RDF triple. When h and t of a RDF triple (h, l, t) are given as inputs, RDFDNN will output l. Relational prediction is similar to classification, and DNN is good at classification. Therefore we predict the relation between two entities by DNN.

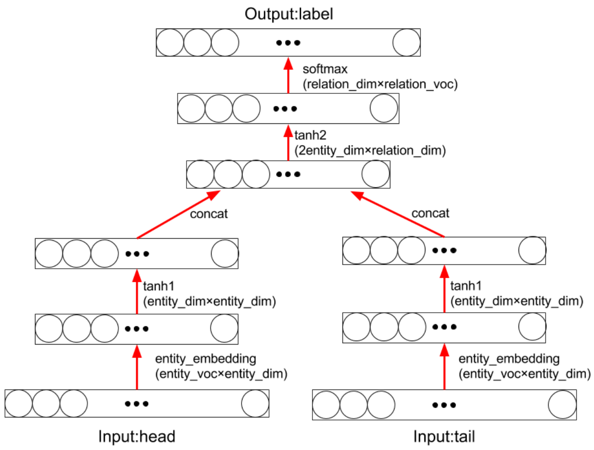

Fig. 1. The structure of RDFDNN

entity_voc and *relation_voc* are the numbers of entities and relations, respectively. *entity_dim* and *relation_dim* are the dimensions of weight matrices. Embedding is the transformation from RDF entities to their vector representations. The length of the transformed vector is called embedding dimension. The vector representation of entities are learned by RDFDNN. We employ simple concatenation of two embedding vectors for the concat. It is better than element wise multiplication and element wise addition in RDFDNN.

Since the output of RDFDNN is the probability distribution of one-hot representation of relation l, the following cross entropy is used as the objective function for training RDFDNN:

$$E = -\sum_{(h,t,l)\in S} \sum_{k \in relation\_voc} l_k log P(h, t)_k, \tag{1}$$

where S is the set of triples in training data, $P(h, t)$ the output of RDFDNN when h and t are given as its inputs, k is the integer index that satisfies $0 \leq k < relation\_voc$. As the optimizer of the above objective function, Adam is used.

Since entity embedding is enough for predicting relations, RDFDNN focuses on entity embedding. Although RDFDNN cannot predict an object from a predicate and a subject, it has abilities of predicting a predicate accurately from a subject and an object.

3 Evaluation

In our experiments, we have used the FB15k and WN18. The datasets are the same as the ones used in the experiments of previous research [1][3]. We have compared RDFDNN with TransE[1], TransR[3] and DKRL[4]. TransE, TransR and DKRL's parameters are same as their original papers. For the comparison of accuracy with previous methods, we set the RDFDNN's parameters as $(entity\_dim, relation\_dim) = (30, 30)$ for FB15k and WN18. RDFDNN learns 10 epochs.

3.1 Comparison with previous methods

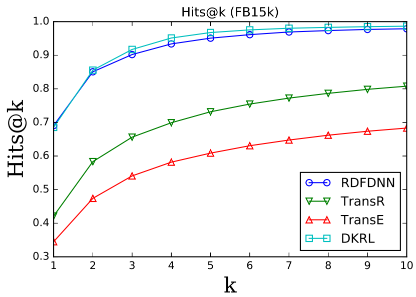

Fig. 2. Hits@k (FB15k)

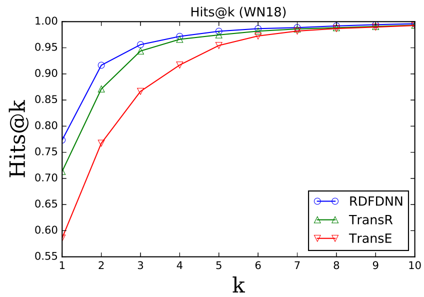

Fig. 3. Hits@k (WN18)

Fig.2 and Fig.3 are the results of comparison with the FB15k and WN18 datasets, respectively. The X-axis is k, and the Y-axis is $Hits@k$. As shown in both figures, RDFDNN is more accurate than TransE and TransR in both datasets. The results are a clear victory for RDFDNN. Although DKRL uses both RDF triples and entity descriptions for learning, its accuracy is comparable or slightly better than RDFDNN. We can claim that RDFDNN can be widely applicable for accurate prediction of entity relations even when entity descriptions are not available.

3.2 Failure analysis

For this failure analysis, 100 triples of RDFDNN failures are randomly sampled. Then the triples are manually evaluated and classified into the above four categories.

The most frequent failure is deceived by majority cases. As an example, RDFDNN's prediction of relation between "Leslie Dilley" and "Raiders of the Lost Ark" is "performer", while its correct answer is "art director". This is because the relation "performer" is the most frequent one for the relation between people and movies. The second most frequent failure is complete failure. The third most frequent failure is too abstract or too concrete compared with correct answers. As an example of this type, RDFDNN predicts the relation between "Park Chu-yong" and "South Korea" as "citizenship", while its correct answer is "Olympic representative". The least frequent failure is structurally similar, but this failure means that RDFDNN recognizes structural similarity between relations. As an example of this type, RDFDNN predicts the relation between "Washington Wizards" and "Michael Jordan" as "belonging states", while its correct answer is "team member". From the above failure analysis, we can say that even when RDFDNN failed, more than half of its failed prediction are valid in some sense.

4 Conclusion

In this paper, we propose RDFDNN for predicting relations of RDF from two given entities. RDFDNN is more accurate compared with TransE and TransR. RDFDNN is comparable with DKRL which uses both RDF triples and entity descriptions for learning.

Acknowledgement

This work was supported by Tokyo Tech - Fuji Xerox Cooperative Research (Project Code KY260195), JSPS Grant-in-Aid for Scientific Research(B) (Grant Number 17H01785) and JST CREST (Grant Number JPMJCR1687).

References

1. Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, Oksana Yakhnenko, "Translating Embeddings for Modeling Multi-relational Data", Part of Advances in Neural Information Processing Systems 26 (NIPS'13), pp.2787-2795. (2013)
2. Martin Kavalec, Vojtech Svatek, "A Study on Automated Relation Labelling in Ontology Learning", Ontology Learning from Text: Methods, Evaluation and Applications, IOS Press. (2003)
3. Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, Xuan Zhu, "Learning Entity and Relation Embeddings for Knowledge Graph Completion", in Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI'15), pp.2181-2187. (2015)
4. Xie Ruobing, Liu Zhiyuan, Jia Jia, Luan Huanbo, Sun Maosong, "Representation Learning of Knowledge Graphs with Entity Descriptions", in Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI'16), pp.2659-2665. (2016)
5. Albert Weichselbraun, Gerhard Wohlgenannt, Arno Scharl, "Refining non-taxonomic relation labels with external structured data to support ontology learning", Data and Knowledge Engineering, Volume 69, Issue 8, pp.763-778. (2010)