Learning Ontology Axioms over Knowledge Graphs via Representation Learning

Leyuan Zhao[1,3], Xiaowang Zhang[1,3,⋆], Kewen Wang[2], Zhiyong Feng[1,3], and Zhe Wang[2]

[1] College of Intelligence and Computing, Tianjin University, Tianjin 300350, China
[2] School of Information and Communication Technology, Griffith University, Brisbane, QLD 4111, Australia
[3] Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin, China
⋆ Corresponding author: xiaowangzhang@tju.edu.cn

Abstract. This presents a representation learning model called SetE by modeling a predicate into a subspace in a semantic space where entities are vectors. Within SetE, a type as unary predicate is encoded as a set of vectors and a relation as binary predicate is encoded as a set of pairs of vectors. A new approach is proposed to compute the subsumption of predicates in a semantic space by employing linear programming methods to determine whether entities of a type belong to a sup-type and thus an algorithm for learning OWL axioms is developed. Experiments on real datasets show that SetE can efficiently learn various forms of axioms with high quality.

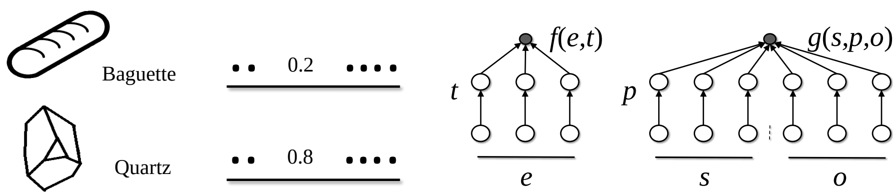
1 Introduction

Ontology construction is a core task of ontology engineering. It has been a research challenge in both knowledge representation and machine learning communities. This is because ontologies are often based on logical formalisms such as description logics (DLs), and contain more complex logical structures than graph databases or RDF triples. DL-Learner [1] is among the first practical systems to learn ontological expressions, including complex DL class descriptions. Many methods for learning new first order formulas and rules have been developed in Inductive Logic Programming (ILP) but they are often unable to handle very large ontologies. Recently, some attempts have been made to effectively learn rules, such as [4], over KG through techniques in knowledge representation learning, but the rules they learn are not typical ontological axioms. What is more, conventional embedding models (e.g.,TransE, TransR, DistMult and SimplE) mainly focus on KG completion, which only embed entities and relations without modeling unary predictes. TransC [3] firstly differentiate types (unary predictes) and entities, it encodes each type as a sphere and can learn the SubClassOf relationship between types. However, the encoding of relations and types in TransC is split, which prevents it from learning the relation SubPropertyOf and other complex axioms (e.g.SubClassOf(ObjectSomeValuesFrom(P, C), D)).

In this paper, we propose a novel unified embedding (called SetE) for KG unary predicates (types) and binary predicates (relations) treating types as sets of entities and relations as sets of entity pairs. On this basis, the subsumption is transformed to relative position of set boundaries which can be efficiently computed by linear programming (LP). We provide an algorithm for learning positive OWL axioms over large-scale knowledge graphs.

2 Our Approach

In this section, we will introduce SetE and the learning algorithm.



(a) A simple illustration of the embeddings of Baguette and Quartz.

(b) The structure of SetE

Fig. 1. An overview of SetE.

Embedding model Inspired by [2], we treat each dimension of embedding as a feature describing certain unique aspect of characteristics. For example, as shown in Figure 1(a), assume the third dimension representing the feature of "hardness", then the value of entity Quartz in this dimension is greater than the value of entity Baguette.

We use inner products to capture feature interactions between sets and their elements(in this case, types and entities). Because when they have common features, the inner product gets a larger value, as illustrated in Figure 1(b). An Thus the score function of the instance fact $< e, InstanceOf, t >$ is defined as follows, where $\boldsymbol{e}$ and $\boldsymbol{t}$ are embedding of e and t, resp.; n is the dimension of $\boldsymbol{e}$,

$$f(\boldsymbol{e}, \boldsymbol{t}) = \boldsymbol{e}^T \boldsymbol{t} = \sum_{i=1}^{n} [\boldsymbol{e}]_i * [\boldsymbol{t}]_i. \tag{1}$$

Following the same intuition, the entity pair $< s, o >$ can be considered as an instance of the relation p, so we model the fact $< s, p, o >$ as follows. Where s and o are head and tail entity of the relation p, $concate(\boldsymbol{s}, \boldsymbol{o})$ means concatenate the two vecotrs $\boldsymbol{s}$ and $\boldsymbol{o}$.

$$g(\boldsymbol{s}, \boldsymbol{p}, \boldsymbol{o}) = concate(\boldsymbol{s}, \boldsymbol{o})^T \boldsymbol{p} = \sum_{i=1}^{2n} [concate(\boldsymbol{s}, \boldsymbol{o})]_i * [\boldsymbol{p}]_i \tag{2}$$

To train the model, we introduce type boundary $B_t \in \mathbb{R}$. So that for all entity e of type t, there has $f(\boldsymbol{e}, \boldsymbol{t}) \geqslant B_t$; for $e \notin t$, there has $f(\boldsymbol{e}, \boldsymbol{t}) < B_t$. The relation boundary B_r is the same. Like previous models, we generate negative samples and use SGD to train SetE.

LP to Subsumption Subsumption in KG has SubClassOf and SubPropertyOf. We take SubClassOf as an example to show how this can be transformed into LP under our model. The axiom SubClassOf(C, D) means that all entities that are instances of C must be instances of D. i.e., $f(\boldsymbol{e}, \boldsymbol{t_C}) \geqslant B_t$ implies $f(\boldsymbol{e}, \boldsymbol{t_D}) \geqslant B_t$, where $\boldsymbol{t_C}$ and $\boldsymbol{t_D}$ are type embeddings of C and D. So we convert this to linear programming that computes the minimum value of $f(\boldsymbol{e}, \boldsymbol{t_D})$ subject to $e \in C$ $(f(\boldsymbol{e}, \boldsymbol{t_C}) \geqslant B_t)$. If the minimum value is greater than the boundary B_t, that is for all entity e in type C, e always satisfy D, so we get the axiom SubClassOf(C, D).

Learning Ontology Axioms Based on previous analysis, we use liner programming on embeddings to learn the following forms: A_1, SubClassOf(C, D); A_2, SubPropertyOf(P, Q); A_3,SubClassOf(ObjectSomeValuesFrom$(P, C), D)$;
A_4, SubClassOf(ObjectIntersectionOf($C, D)$, Range(F)). The algorithm learning Sub ClassOf(C, D) is as follows.Line 3 means that if values in $\boldsymbol{t_C}$ are smaller than or equal to $\boldsymbol{t_D}$ in every dimension, then we can directly get that for any $\boldsymbol{e}$, if $f(\boldsymbol{e}, \boldsymbol{t_C}) \geqslant B_t$ then $f(\boldsymbol{e}, \boldsymbol{t_D}) \geqslant B_t$. At last, Filter() returns axioms whose SC(standard confidence, defined in [4]) are greater than $MinSC$.

Algorithm 1 Learning SubClassOf Axioms from a KG

Input: a KG K, and two real numbers LB_t and $MinSC \in [0, 1]$
Output: a set O of SubClassOf axioms

1: $\mathcal{E}$:= SetE(K); $O := \emptyset$.
2: **for** type embeddings $\boldsymbol{t}_C$ and $\boldsymbol{t}_D$ in $\mathcal{E}$ **do**
3:     **if** $(\sum_{i=1}^{n}([\boldsymbol{t}_C]_i \leqslant [\boldsymbol{t}_D]_i)?1:0) == n$ **then**
4:         Add SubClassOf(C, D) to O
5:     **else if** LP$(\boldsymbol{t}_C, \boldsymbol{t}_D) \geqslant LB_t$ **then**
6:         Add SubClassOf(C, D) to O
7:     **end if**
8: **end for**
9: O := Filter$(O, MinSC)$
10: **return** O

3 Experiments and Evaluation

The experiment on YAGO39K aims to evaluate the effectiveness of SetE by comparing with the state-of-the-art model TransC[3] in SubClassOf classification. We retain four metrics: Accuracy, Precision, Recall and F1-score. TransC was trained with the configuration in their report. SubClassOfs were removed from the training set. To reflect real data that the negative samples far exceeds the positive(e.g.,#negative :#positive is 226:1 in DBpedia 2016 OWL), we add the proportion of negative samples during the experiment.

Result in Table 1 indicates that SetE outperforms TransC and is getting better when improving the proportion of negative samples. The Precision of SetE is much higher than TransC (up to 87.03% for rate 1:10). It shows that SetE is more cautious in making positive judgments, i.e., SetE distinguishes positive samples better.

Table 1. Classification results(%) based on the rate of #positive and #negative.

Model	rate	Accuracy	Precision	Recall	F1
TransC	1:1	57.95	56.61	**68.10**	**61.82**
	1:4	50.96	24.20	**68.10**	35.71
	1:10	48.17	11.23	**68.10**	19.28
SetE	1:1	65.20	**70.10**	53.10	60.41
	1:4	80.78	51.91	53.10	52.50
	1:10	**87.03**	35.66	53.10	42.67

4 Conclusion

In this paper, we present a new model SetE to specifically represent types and relations in a semantic space which can reduce subsumption into linear programming. Our proposal utilizes the logical relationship to characterize the semantic features of expressive types in learning shows certain interpretability. In the future, we will improve the quality of expressive axioms learned and considerate even negated axioms.

Acknowledgments

This work is supported by the National Key Research and Development Program of China (2017YFC0908401) and the National Natural Science Foundation of China (61976153,61972455). Xiaowang Zhang is supported by the Peiyang Young Scholars in Tianjin University (2019XRX-0032).

References

1. Bühmann, L., Lehmann, J., and Westphal, P.: DL-Learner - A framework for inductive learning on the semantic web. *J. Web Semant.*, (39),15–24 (2016).
2. Ding, B., Wang, Q., Wang, B., Guo, L.: Improving knowledge graph embedding using simple constraints. In: *Proc. of ACL 2018*, pp.110–121.
3. Lv, X., Hou, L., Li, J., Liu, Z.: Differentiating concepts and instances for knowledge graph embedding. In: *Proc. of EMNLP 2018*, pp.1971–1979.
4. Omran, P.G., Wang, K., Wang, Z.: Scalable rule learning via learning representation. In: *Proc. of IJCAI 2018*, pp.2149–2155.