

An Ontology-Based Deep Learning Approach for Knowledge Graph Completion with Fresh Entities

Elvira Amador-Domínguez¹ †, Patrick Hohenecker², Thomas Lukasiewicz²,
Daniel Manrique¹, and Emilio Serrano¹

¹ Department of Artificial Intelligence, Universidad Politécnica de Madrid, Madrid, Spain
`{eamador, dmanrique, emiloserra}@fi.upm.es`

² Department of Computer Science, University of Oxford, Oxford, United Kingdom
`{patrick.hohenecker, thomas.Lukasiewicz}@cs.ox.ac.uk`

Abstract. This paper introduces a new initialization method for knowledge graph (KG) embedding that can leverage ontological information in knowledge graph completion problems, such as link classification and link prediction. Although the initialization method is general and applicable to different KG embedding approaches in the literature, such as TransE or RESCAL, this paper experiments with deep learning and specifically with the neural tensor network (NTN) model. The experimental results show that the proposed method can improve link classification for a given relation by up to 15%. In a second contribution, the proposed method allows for addressing a problem not studied in the literature and introduced here as “KG completion with fresh entities”. This is the use of KG embeddings for KG completion when one or several of the entities in a triple (head, relation, tail) has not been observed in the training phase.

Keywords: Statistical Relational Learning, Ontological Knowledge Base, Knowledge Graph Embedding, Latent Feature Model.

1 Introduction

Knowledge representation has always been one of the main challenges of Artificial Intelligence. Throughout time, several approaches have been proposed to model knowledge in a structured and comprehensive way. *Ontologies* [7] are an important approach to knowledge representation. Ontologies present a formal definition of types, properties and relationships between entities applied to a concrete domain. These representations are fairly intuitive for humans, as well as being easily translated into machine languages to allow computers to share concepts. One of the main characteristics of this representation is that concepts or classes are organized in a hierarchical way. These classes tend to remain static, whereas individuals or instances are more dynamic.

Knowledge graphs (KGs) are usually supported on this structured form of representation. The approach considered in KGs is, however, different from the one existing in ontologies. Unlike ontologies, KGs aim to gather all the concrete facts existing in the domain instead of extracting general patterns. A fact follows usually the schema (*head*,

† Authors listed in alphabetical order.

relation, tail), representing the link between two certain entities by means of a specific relation. These KGs have an increasing popularity in the literature, because they support several tasks, such as question answering or recommendation systems [5].

KGs usually present a high degree of incompleteness, hindering them to be used in a number of domains. Even when completing KGs is mostly a handcrafted task, a number of approaches allow this *Knowledge Graph Completion* (KGC) to be automatic or at least supported automatically [5]. Popular KGC approaches essentially calculate a *Knowledge Graph Embedding* (KGE) that is used for tasks such as: (1) *link or triple prediction*, given two elements of a triple in the KG, ranking the most plausible entities to fulfill the missing element; and (2) *link or triple classification*, predicting if a triple belongs to a KG [10].

Figure 1 illustrates a simplistic KG about American politicians and their relatives. In this graph, Michelle Obama’s nationality is unknown. A KGC approach would allow: triple prediction, i.e., ranking the nationalities more likely to fulfill the triple (*Michelle Obama, nationality, ?*); or, triple classification, i.e., calculating the confidence of the triple (*Michelle Obama, nationality, United States*). KGE and KGC approaches could learn that American presidents’ wives are likely to be American, although this is not necessarily true as is the case with Donald Trump’s wife.

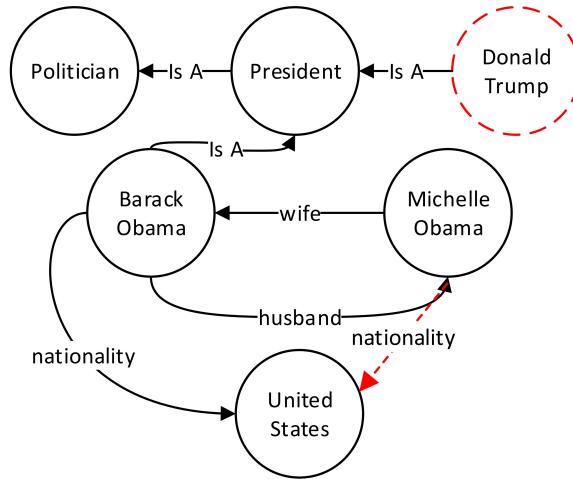


Fig. 1: An example of Knowledge Graph Completion.

Two major shortcomings found in current KGC and KGE proposals are that: (1) they do not leverage ontological information; (2) they are unable to predict facts not seen in training time or *fresh entities*. In Figure 1, there is not a special treatment for the relation “is A”, considered at the same level as relations such as wife or nationality. Furthermore, if the entity Donald Trump is unknown during training, no prediction containing it can be conducted. This paper contributes with a new initialization method for KGEs that addresses these two issues.

The paper outline is as follows. Section 2 presents the related works. Section 3 explains the approach presented. Section 4 details the methodology to evaluate this approach, and Section 5 offers the results and discussion. Finally, Section 6 concludes and discusses future works.

2 Related Works

2.1 Knowledge Graph Completion

Considering the vastness of the domain they cover, KGs usually present a considerable degree of incompleteness. Therefore, an important challenge in KGs is to complete them automatically. KGC tasks include triple prediction and triple classification. In a more general context, these tasks are in the wider domain of *statistical relational learning*. KGC methods can be categorized into three different groups: (1) graph-feature-based methods, (2) Markov random fields, and (3) latent-feature-based models [5]. In this work, we focus on improving the last group, which includes some of the most popular KGC and KGE models, such as RESCAL and TransE [10].

RESCAL [6] represents a KG as a tensor, where each slice of the tensor represents an existing relation. This method is easy to understand and to implement. On the other hand, the tensor size is increased with the number of entities and relations. Therefore, this approach has scalability problems. TransE, or several of its variants, such as TransD or TransR [10], are more scalable. TransE is based on exploiting distance-based scoring functions. Another of the earliest works in KGC [9] introduces *Neural Tensor Networks* (NTN), an approach based on the combination of the tensor representation with artificial neural networks.

Comparing these approaches is out of the scope of this paper, and there are extensive reviews [5,10]. The approach presented is an initialization method that can be combined with these classical works in KGC and KGEs. In this paper, we have experimented with the NTN model, since its implementation and results are available online, facilitating the evaluation and comparison of our proposal.

2.2 KGC with Ontologies

As noted by Paulheim [8], ontologies are mainly involved in the creation and data introduction phase of KGs. They are useful for this phase, as they provide restrictions for instances and relations.

Ontological information is not leveraged in the mainstream of KGE and KGC approaches [10]. Some works address its use for the *type inference* problem, i.e., predicting the type of a given entity. This can be considered as a KGC task, although less versatile than triple prediction and triple classification. Only Hohenecker and Lukasiewicz [2] recently propose to distinguish ontologies and facts for a number of KGC tasks. Unlike in this previous work, instead of presenting a new KGE approach, a simple initialization method is proposed to be combined with classical KGEs. This method, as shown in Section 5, has potential of enhancing KGC via the use of ontologies.

2.3 KGC with Fresh Entities

As explained in the introduction, fresh entities refer to those entities that have not been seen in training time when building a KGE model. Most KGE approaches cannot deal with these entities, because they do not receive an initial encoding in the one-hot vector initialization [5]. Some proposals technically can use fresh entities as input in the prediction phase. In this vein, Socher et al. [9] use Word2Vec [3] embeddings as input in the KGE model. However, beyond having a compatible interface, these works do not explore or evaluate the possibility of reasoning over fresh entities as in the presented contribution.

3 An Ontology-Based Deep Learning Approach for Knowledge Graph Completion with Fresh Entities

The hypothesis presented is that KGC can be enhanced if the initial representation of entities is complemented with a vector that encodes ontological information. In the example given in Figure 1, predicting links for the entity “Barack Obama”, let us call it e_1 , should be easier if the initial representation of e_1 is enriched with an embedding that tries to reflect that e_1 is a president and, therefore, a politician. Moreover, “fresh entities” could also receive this extra information about the ontology if their initial representation can deal with entities not seen in training time. In the example, predicting links for “Donald Trump”, denoted e_2 , would be feasible if, at least, e_2 ’s links with the ontology were known. That is, even when e_2 could be new in prediction time, a KGC approach could reason about it if e_2 ’s initial representation encodes that e_2 is a president and a politician. This is because the information about similar entities such as e_1 might be leveraged.

Figure 2 summarizes the method proposed and its application to general KGs. Firstly, the ontological information or *ontological knowledge base* (OKB), such as concepts and classes, has to be separated from the *general knowledge base* (KB), such as individuals or instances. In the example, only the relation *type_of*, which is present in a number of general KGs such as Freebase, is taken as the hierarchy of concepts in the ontology. Secondly, an initial representation is chosen for the entities in both the OKB and the KB. Following Socher et al.’s [9] initialization method, the mean of vectors using Word2Vec is used for each word contained in an entity. For example, the initial value “Bengal tiger” is the mean of vectors for the words “Bengal” and “tiger”. In the case of the OKB, this initial mean of vectors is enriched with the concepts higher up in the hierarchy. For instance, the initial value for the concept “tiger” is the mean of vectors for the words “tiger”, “cat”, and “mammal”. Thirdly, each entity vector in the KB is extended with its corresponding OKB vector codifying ontological information. Fourthly, a KGE is learned using known approaches such as NTN or RESCAL. Finally, this KGE can be used to solve KGC tasks such as link classification.

The main idea behind this approach is that relations tend to have a type restriction for the entities that can be linked. These restrictions are explicitly represented in ontologies. The proposed initialization method includes this type information explicitly in the initial representations for entities, releasing KGE approaches from the need to

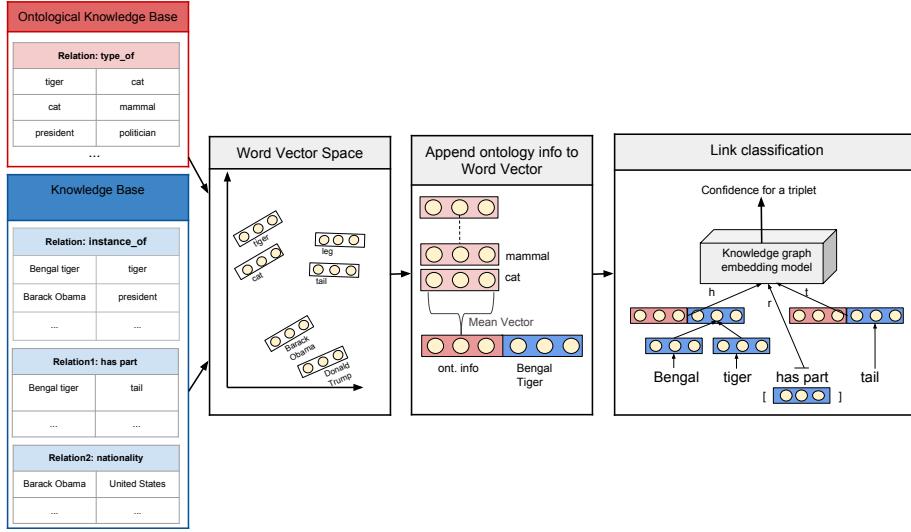


Fig. 2: An ontology-based deep learning approach for KGC with fresh entities

learn it. Moreover, considering an embedding for the hierarchy of types or concepts in the input can allow KGEs to reason about “fresh entities” in certain relations, even if nothing else is known about the new entity. Furthermore, pretrained Word2Vec models can leverage information from unsupervised texts and then transfer this information to the KGC process.

4 Methodology

4.1 Baseline Model

To evaluate the proposed method, a Neural Tensor Network (NTN)[9] is used as KGE approach³. One of the main advantages of this model is that both the datasets and the original code are publicly available, allowing an easy comparison of the use of NTN alone or with the proposed initialization method containing ontological information. In this model, the embedding matrix is incorporated as a parameter of the network, allowing it to be initialized externally.

4.2 Datasets

Two different datasets are considered to evaluate our approach. As mentioned, Socher et al. [9] provide not only the model but also the data used for training and evaluation. Moreover, these provided datasets are hegemonic in the literature and widely used to

³The implementation used in this work, as well as the employed datasets are available at <https://github.com/Elviish/ntn-pytorch-ontological-info>.

evaluate different KGC approaches. These datasets represent a reduced subset of the data existing in WordNet [4] and Freebase [1], providing facts about 11 and 13 relations, respectively.

4.3 Training Protocol

Before the training phase, a type-hierarchy retrieval is performed for each entity on the dataset as explained in Section 3. The DBpedia ontology is the selected source for type retrieval and ontological information. A SPARQL query is run to obtain each entity type or class and its upper classes. Then this set is transformed into a vector and appended to the initial embedding of each entity. The ontological part of the initial entity embedding is frozen throughout the training phase.

The model is trained for the triple classification, a binary classification problem. Therefore, the model has to be fed with both positive and negative examples in order to learn to discriminate. Since the training datasets are only composed by correct triples or positive samples, ten negative samples are generated for each positive example by exchanging the tail entity of the given fact with a randomly sampled one from the total entity pool. The NTN is trained with 500 epochs and a regularization value of 1e-4.

4.4 Evaluation Protocol

Regarding evaluation, two different scenarios are considered. Firstly, the impact of the ontological information on the model is quantified by comparing the results obtained with or without appending the ontological information vector to the entity. In these experiments, the provided test sets composed by the same amount of correct and incorrect facts are used. All facts are formed by entities that have been previously seen during training. However, not all the relations in the datasets are evaluated cause some of them are considered unsuitable for triple prediction even for humans, such as *place of death* in Freebase.

In the second scenario, a new dataset partition is considered to evaluate if the ontology-based initialization method allows reasoning with “fresh entities”. The first set is composed by facts where both head and tail entities are known, whereas the second set is composed by facts where either the head or the tail are unknown in training time. In this case, ontological information is always considered. This fresh entities set is generated by randomly selecting 1500 entities from the total set and extracting its associated facts in the training data.

5 Results and Discussion

Table 1 shows the results in terms of accuracy for the triple classification problem. As shown, there is an improvement with the proposed method for all relations studied in WordNet except *member_holonym* and *has_part*. For the relation *sysnet_domain_topic* the method produces a 9% enhancement. For Freebase, there is an improvement with the proposed method for all relations studied except in the relation *institution*. For the

Relation	WordNet				WordNet with FEs				Freebase				Freebase with FEs						
	KB	OKB	No FEs	FEs	Relation	KB	OKB	No FEs	FEs	Relation	KB	OKB	No FEs	FEs	Relation	KB	OKB		
member_meronym	0.51	0.53	0.54	0.52	nationality	0.37	0.5	0.53	0.32	profession	0.47	0.6	0.5	0.5	cause_of_death	0.52	0.53	0.57	0.7
member_holonym	0.52	0.5	0.53	0.53	institution	0.52	0.5	0.56	0.58	religion	0.39	0.54	0.55	0.55	ethnicity	0.5	0.53	0.36	0.32
part_of	0.52	0.54	0.55	0.53															
has_part	0.52	0.5	0.5	0.48															
domain_region	0.49	0.56	0.48	0.49															
sysnet_domain_topic	0.44	0.53	0.47	0.47															
domain_topic	0.51	0.57	0.49	0.44															
MEAN	0.51	0.53	0.52	0.51												0.45	0.53	0.53	0.41

Table 1: Results over WordNet and FreeBase, both considering only the knowledge based without and with ontological information, and results over WordNet and Freebase, both considering all entities in training time or with fresh entities in testing.

relation *religion*, there is a 15% increment. The relation *profession* also presents a significant increase, 13%.

The difference between WordNet and Freebase can be motivated cause, while WordNet relations are not restricted by the entity class but by its syntactical category, the entity types in Freebase relations are more homogeneous.

Regarding the fresh entities problem, results show that the approach presented can address it. There is a slight worsening, always less than 5%, in the prediction of three of seven relations evaluated for WordNet. For Freebase, only in two of six relations the accuracy is lower when fresh entities are considered. However, for the specific case of *nationality*, the loss of accuracy is considerably high, 21%.

These results support the hypothesis that the explicit ontological information used in the proposed initialization method allows KGEs to reason about entities not seen in training time by leveraging type restrictions.

6 Conclusion and Future Works

A novel initialization approach for *Knowledge Graph* (KG) completion models is presented. In this approach, explicit ontological information is introduced in the form of a vector that gathers the class hierarchy of a given entity. This information is fixed and shared across similar entities, which leads the KG completion model to infer type restrictions on the relations. This approach not only increases the predictive capability of the model in the canonical *triple classification* problem (up to 15%), but also enables reasoning over entities unknown in training time or “*fresh entities*”. To the best of our knowledge, this is the first research work introducing the fresh entities problem, proposing an approach to address it, and evaluating it.

Since KGs are very dynamic, and obtaining *KG Embeddings* (KGEs) is computationally expensive, addressing the fresh entities problem allow trained KGEs to be used when new entities and links are added in the KG without retraining the model. More importantly, this allow transfer learning with KGEs, i.e., using pretrained KGEs for a

number of different KGs as long as they share the same ontology. This is important because, unlike the instances than can be private and very dynamic, ontologies tend to be public, general, and a more static source of information. The presented experimental results show that dealing with this interesting and challenging problem is feasible, although with a cost in the predictive power, as expected.

Our future works include: applying this initialization approach to new knowledge bases; evaluating its use in other KGE methods, such as TransE, RESCAL, or ComplEx [10]; and evaluating transfer learning among KGs sharing the same ontological information.

Acknowledgments

This research work is supported by the Spanish Ministry of Science, Innovation and Universities under the program “Estancias de movilidad en el extranjero José Castillejo para jóvenes doctores” (CAS18/00229) and by the “Universidad Politécnica de Madrid” under the programs: “Ayudas al Personal Docente e Investigador para Estancias Breves en el Extranjero”, “Ayudas dirigidas a Jóvenes Investigadores para Fortalecer sus Planes de Investigación”, and “Ayudas para Contratos Predoctorales para la Realización del Doctorado”.

References

1. K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor. Freebase: A collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, SIGMOD ’08, pages 1247–1250, New York, NY, USA, 2008. ACM.
2. P. Hohenecker and T. Lukasiewicz. Ontology reasoning with deep neural networks. *CoRR*, abs/1808.07980, 2018.
3. T. Mikolov, K. Chen, G. S. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781, 2013.
4. G. A. Miller. Wordnet: A lexical database for english. *Communications of the ACM*, 38:39–41, 1995.
5. M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich. A review of relational machine learning for knowledge graphs: From multi-relational link prediction to automated knowledge graph construction. *CoRR*, abs/1503.00759, 2015.
6. M. Nickel, V. Tresp, and H.-P. Kriegel. A three-way model for collective learning on multi-relational data. 2011.
7. N. F. Noy and D. L. McGuinness. Ontology development 101: A guide to creating your first ontology, 2001.
8. H. Paulheim. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web*, 8(3):489–508, 2017.
9. R. Socher, D. Chen, C. D. Manning, and A. Ng. Reasoning with neural tensor networks for knowledge base completion. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 926–934. Curran Associates, Inc., 2013.
10. Q. Wang, Z. Mao, B. Wang, and L. Guo. Knowledge graph embedding: A survey of approaches and applications. *IEEE Trans. Knowl. Data Eng.*, 29(12):2724–2743, 2017.