# Towards Learning Better Representations for Completion of Real-World Knowledge Bases

**Anonymous Author(s)**
Affiliation
Address
`email`

## 1 Introduction

Knowledge bases (KBs) play a fundamental role in compiling human knowledge. They structure open information available at the web over topological and non-topological relations. Despite the success of KBs as core resources on multiple tasks, they usually experience incompleteness and noisy information. To address this issue, several previous works have proposed techniques to predict and infer new relationships from the existing information aiming at improving the completeness of the KB [Socher et al., 2013, Bordes et al., 2013, Trouillon et al., 2016].

In this last decade, methods that learn embeddings representations of multi-relational data (*e.g.*, KBs, ontologies, and knowledge graphs (KGs)) have emerged with good performance on tasks like link predicton [Socher et al., 2013, Bordes et al., 2013, Trouillon et al., 2016]. Such methods have been driven by the recent advances in learning representations of symbolic data methods (*e.g.*, words). In the same context, techniques of graph embeddings such as DeepWalk [Perozzi et al., 2014] and Graph Neural Networks models [Wang et al., 2017, Hamilton et al., 2017a, Xu et al., 2019] have improved the capability of structure-related tasks, for example, node classification.

However, we observe that most of the techniques aforementioned suffer at least two issues. On the first issue, most of them rely on transductive inference, and, as such, the embeddings are only learned to represent the KB components already known in training time. There are only a few recent approaches accounting for the Out-of-Knowledge-Base (OOKB) problem [Hamaguchi et al., 2017, Hamilton et al., 2017b, Shi and Weninger, 2018]. The second issue is related to the fact that most of the mentioned approaches explicitly include only entities and relations, leaving aside more general elements such as concepts and definitions. As a consequence, they ignore the topological structure underlying the concepts definitions and semantics of the KBs. Recently, some proposals have emerged to mitigate part of this issue by building the embeddings within more complex spaces rather than Euclidean to accommodate concepts and hierarchies of concepts. For instance, Holographic Embeddings [Nickel et al., 2016] and Poincaré Embeddings [Nickel and Kiela, 2017] seek to improve the representations by explicitly relating the latent hierarchies characterized on the KBs.

As a consequence, we argue that two challenges arise to enhance the existing approaches. The first challenge is the development of an approach that enhances the link prediction task while providing inductive representations of the knowledge bases by considering: (i) the prediction of relations between existing entities ( *i.e.*, traditional link prediction with closed-world KBC [Shi and Weninger, 2018]), (ii) the prediction of relations between existing entities and new incoming entities ( *i.e.*, OOKB [Hamaguchi et al., 2017] and open-world KB [Shi and Weninger, 2018]), (iii) the introduction of new incoming relations between entities, (iv) the introduction of new concepts and (v) the prediction of new relations between entities and concepts. The second challenge relies on an approach for automatically learning topological relations between concepts and non-topological relations between entities. In this work, we stand for the discussion of the prior-mentioned challenges and how tackling each one could lead to the development of new techniques for embedding knowledge bases in a relational-oriented space and that consider predicting relations between seen and unseen entities.

Submitted to 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Do not distribute.

## 2 Challenges

The knowledge base completion (KBC) problem, is defined by assuming that a KB is *incomplete* and the objective is to identify possible triples that complete it. KBC is formulated as follows:

**Definition 2.1.** Let $\mathcal{KB} = (E, R, T)$ be an incomplete knowledge base, in which $E$ is the set of entities, $R$ is the set of relations, $T$ is the set of triples $T = \{t | t = (h, r, t)\}$, where $r \in R$ and $h, t \in E$. The objective is to find a new triple $\mathsf{t} = (x, k, y)$, where $\mathsf{t} \notin T$ and $x, y \in E$ to improve the completeness of $\mathcal{KB}$.

In this work, we acknowledge the importance of also including concepts into the KBC task. Therefore, we extend the above definition by the addition of one more element, namely $\mathcal{KB}_e = (E, R, C, T)$, where $C$ is the set of concepts and $T$ also includes triples of the form $t = (c_i, r, c_j)$ and $t = (e, r, c_j)$, where $c_i, c_j \in C$ and $e \in E$.

Notwithstanding, for real cases, most KBs evolve over time, *i.e.*, they incorporate, remove, and update concepts, entities, and relations. The classical definition of KBC does not explicitly include new unseen concepts, entities, and relations, which drives the first challenge. To address this KBs evolving nature, [Hamaguchi et al., 2017] defines the *OOKB entity* task, which consists of adding new unseen entities on the KB based on its relations. Along similar lines, [Shi and Weninger, 2018] define the *closed-world knowledge graph completion (KGC)* task and the *open-world KGC* task. Both works first introduce the need for updating (on inference time, *i.e.*, without retraining the model) the embeddings space model with the newly learned representation of the unseen *entities* and *relations*. We advocate that they still miss a fundamental component of real-world knowledge bases: the introduction of *new concepts* and predictions of *new relations* between concepts and entities.

On the second challenge, as surveyed by [Trouillon et al., 2019], the state-of-the-art on latent representations of knowledge bases, such as [Bordes et al., 2013, Trouillon et al., 2016], lacks the abilities to learn topological semantics and relations between concepts and between concepts and entities. Therefore, the challenge here consists of providing a technique able to generate embedding representations that incorporate relation semantics, structural properties of the KB and its concepts, entities, and relations.

## 3 Our Vision

In the portrayed perspective of this paper, a technique that produces adequate embedding representations of KBs' entities and their relations should consider not only those components but also the inner structure between all the elements whose semantics must be captured by a KB. Besides that, we claim that it is necessary for the emergence of a proper task to evaluate such technique in terms of the provided embedding and its induction capability for new coming concepts, entities, and relations.

Like so, first, we extend the Definition 2.1 to augment the real-world KBC task in the context of KB embeddings. Then, we define a new task, named *Knowledge Base Progression* (KBP) as follows:

**Definition 3.1.** Let $\mathcal{KB} = (E, C, R, T)$ be an incomplete knowledge base, where $E$ is the set of entities, $C$ is the set of concepts, $R$ is the set of relations and $T$ is the set of triples, *i.e.*, the set of the relationship between relations, concepts and entities. A *knowledge base progression* task defines, in inference time: *(i)* incorporate to the learned space-model of $\mathcal{KB}$, *(i).a*, new unseen entities and concepts of a $\mathcal{KB}$, $\mathcal{KB}' = (E', C', R, T')$, where $E' - E$ and $C' - C$ is the set of unseen entities and concepts, respectively, or, *(i).b*, new unseen relations of a $\mathcal{KB}'$, $\mathcal{KB}'' = (E, C, R'', T'')$, where $R'' \cap R = \varnothing$; *(ii)* predict new relations between elements from sets $E$ and $C$; and *(iii) consider the KB's structural properties,* i.e.*, the topology of its concepts to provide relations between concepts and entities*.

Second, we point out some of the requirements that would make such a technique capable of solving the challenge defined in Definition 3.1. The technique should be able to learn an embedding space that considers the relational semantics and structural properties of the KB and its elements. Also, the embeddings of each KB's element, *i.e.*, concept, entity or relation should be provided by a learning function capable of generalizing the element regarding its set of associated triples, to provide support for new unseen elements (*e.g.*, [Hamaguchi et al., 2017, Hamilton et al., 2017b]). Further, be able to predict new relations between the learned embeddings.

## 4 Conclusion

Our perception is towards a new technique able to learn embeddings representations of symbolic data in the form of multi-relational data. Also, we pointed out an extension of a traditional task in KBs representation learning. We argue in favor of a technique that takes into account the topological structure of the data, the semantics of relations and its elements, *i.e.*, concepts, entities, and relations. Likewise, we define the pointed challenge as the *Knowledge Base Progression* problem. Potential future paths include the formal definition of this technique and extend existing benchmark on KBC problem for the new KBP challenge.

## References

A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. Translating Embeddings for Modeling Multi-relational Data. In *Advances in Neural Information Processing Systems 26, NIPS 2013*, pages 2787–2795. Curran Associates, Inc., 2013. URL https://papers.nips.cc/paper/5071-translating-embeddings-for-modeling-multi-relational-data.

T. Hamaguchi, H. Oiwa, M. Shimbo, and Y. Matsumoto. Knowledge Transfer for Out-of-Knowledge-Base Entities: A Graph Neural Network Approach. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 1802–1808. International Joint Conferences on Artificial Intelligence, 2017.

W. L. Hamilton, R. Ying, and J. Leskovec. Representation Learning on Graphs: Methods and Applications. *IEEE Data Engineering Bulletin*, sep 2017a. URL http://arxiv.org/abs/1709.05584.

W. L. Hamilton, R. Ying, and J. Leskovec. Inductive Representation Learning on Large Graphs. In *Advances in Neural Information Processing Systems 30, NIPS 2017*, pages 1024–1034. Curran Associates, Inc., 2017b. URL http://papers.nips.cc/paper/6703-inductive-representation-learning-on-large-graphs.

M. Nickel and D. Kiela. Poincaré Embeddings for Learning Hierarchical Representations. In *Advances in Neural Information Processing Systems 30, NIPS 2017*, pages 6338–6347. Curran Associates, Inc., 2017.

M. Nickel, L. Rosasco, and T. Poggio. Holographic Embeddings of Knowledge Graphs. In *Proceedings of the 30th Conference on Artificial Intelligence, AAAI 2016*, pages 1955–1961. AAAI Press, 2016. URL https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/download/12484/11828.

B. Perozzi, R. Al-Rfou, and S. Skiena. DeepWalk: Online Learning of Social Representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2014*, pages 701–710. ACM Press, 2014. doi: 10.1145/2623330.2623732. URL http://dx.doi.org/10.1145/2623330.2623732.

B. Shi and T. Weninger. Open-World Knowledge Graph Completion. In *Proceedings of the 32th Conference on Artificial Intelligence, AAAI 2018*, pages 1957–1964. AAAI Press, 2018.

R. Socher, D. Chen, C. D. Manning, and A. Ng. Reasoning With Neural Tensor Networks for Knowledge Base Completion. In *Advances in Neural Information Processing Systems 26, NIPS 2013*, pages 926–934. Curran Associates, Inc., 2013. URL http://papers.nips.cc/paper/5028-reasoning-with-neural-tensor-networks-for-knowledge-base-completion.

T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, and G. Bouchard. Complex Embeddings for Simple Link Prediction. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning, ICML 2016*, pages 2071–2080. JMLR.org, 2016. URL https://dl.acm.org/citation.cfm?id=3045609.

T. Trouillon, E. Gaussier, C. R. Dance, and G. Bouchard. On Inductive Abilities of Latent Factor Models for Relational Learning. *Journal of Artificial Intelligence Research*, 64:21–53, 2019. ISSN 1076-9757. doi: 10.1613/jair.1.11305. URL https://www.jair.org/index.php/jair/article/view/11305.

Q. Wang, Z. Mao, B. Wang, and L. Guo. Knowledge Graph Embedding: A Survey of Approaches and Applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12):2724–2743, 2017. doi: 10.1109/TKDE.2017.2754499. URL http://ieeexplore.ieee.org/document/8047276/.

K. Xu, W. Hu, J. Leskovec, and S. Jegelka. How Powerful are Graph Neural Networks? In *International Conference on Learning Representations, ICLR 2019*, 2019. URL https://openreview.net/forum?id=ryGs6iA5Km.