Neuro-Symbolic Integration for Reasoning and Learning on Knowledge Graphs

Luisa Werner

Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG F-38000 Grenoble, France
luisa.werner@inria.fr

Abstract

The goal of this thesis is to address knowledge graph completion tasks using neuro-symbolic methods. Neuro-symbolic methods allow the joint use of symbolic information defined as rules in ontologies and knowledge graph embedding methods that represent the entities and the relations of the graph in the vector space. This approach has the potential to improve the resolution of knowledge graph completion tasks in terms of reliability, interpretability, data efficiency and robustness.

Background

Neural-Symbolic Integration

Neuro-Symbolic AI (Sarker et al. 2021) is an increasingly active research field that aims to integrate the paradigms of symbolic AI with connectionist models. While the strength of symbolic AI lies in its general and interpretable knowledge representation, sub-symbolic approaches, namely neural networks, contribute data-driven pattern recognition capabilities. Neuro-symbolic integration addresses the mutual limitations of purely symbolic or neural approaches, thereby enabling more robust, interpretable and reliable models. Further, the fusion of these traditionally complementary methodologies aims to enhance the efficiency of knowledge representations, reasoning, and learning in complex domains.

While the range of neuro-symbolic methods is wide, I focus on methods that use prior knowledge in the form of first-order logic rules and integrate them in a learning task. One such approach is knowledge-enhanced neural networks (Daniele and Serafini 2019), which involves stacking differentiable knowledge enhancement layers on top of neural architectures. In these layers, the predictions of a neural component are interpreted as grounded predicates and are modified with the objective of increasing the satisfaction of a set of logic rules.

Knowledge Graphs

Knowledge graphs are widely used in industry and academia, providing a rich and versatile representation of diverse information. Knowledge graphs such as YAGO, Freebase or Wikidata store real-world facts in the form of triples. However, many existing knowledge graphs suffer from incompleteness and noise. Hence, knowledge graph completion tasks have gained importance. These tasks are addressed by knowledge graph embedding methods (Dai et al. 2020), which aim to find low-dimensional vector representations for entities and relations in the knowledge graph.

Given a knowledge graph, rules can be defined that the triples must respect. These rules are typically encoded in an ontology. Given the example triple `(BillGates, founderOf, Microsoft)` the subject linked with the relation `founderOf` has to be of type `Person` and the object of type `Organization`. Such rules can also describe multiple relations with transitivity, symmetry or antisymmetry for example, such as $\texttt{children(x,y)} \implies \texttt{parent(y,x)}$ or $\texttt{children(x,y)} \wedge \texttt{children(y,z)} \implies \texttt{grandchildren(x,z)}$.

Goal of the Thesis

Given the symbolic nature of ontologies and the vector representation in knowledge graph embeddings, neuro-symbolic methods are a promising tool for leveraging both sources of information for knowledge graph completion tasks. The goal of my thesis is twofold: (1) produce a learning architecture with explicit representation of entities, relations and their compositions by (2) using ontological language as mean to enhance the learning process and inject reasoning capabilities.

Contributions

In the following, my (planned and past) contributions are listed.

Reproducibility Study on Knowledge Enhanced Neural Networks

As a first step in my thesis, I reproduced and replicated the results in (Daniele and Serafini 2023) to firstly better understand the method and, secondly, to make sure that my own experiments are reliable. In this context, I wrote a paper that gives lessons learned on reproducibility and discusses how research can be made more reproducible, exemplifying the work in (Daniele and Serafini 2023).
Status: Completed, accepted paper at AAAI24.

Knowledge Enhanced Graph Neural Networks for Graph Completion

Further, I investigated the combination of knowledge enhancement layers (Daniele and Serafini 2023) with graph neural networks (Wu et al. 2022). In related works, knowledge enhancement layers have been used in conjunction with non-relational neural networks such as MLPs. However, no significant improvement of a knowledge enhanced graph neural network in comparison to a standalone graph neural network could be found. The experiments gave evidence that this may be due to the fact that the knowledge enhancement layers encode that knowledge of the form $\texttt{Class(x)} \wedge \texttt{Edge(x,y)} \implies \texttt{Class(y)}$ which is implicitly encoded as inductive bias in the message passing algorithm of a graph neural network. The conclusion of this work is that it is required to encode knowledge that is complementary to the information prevalent in the data in order to achieve a significant improvement.
Status: Completed, accepted paper at DSAA23 (Werner et al. 2023), accepted workshop paper at KBCG@IJCAI23

Knowledge Enhancement of Node Classification on Heterogeneous Graphs

To tackle the previous limitations and formulate more complex knowledge, the next goal is to apply knowledge enhancement techniques to node classification on multi-relational graphs, since it has only been used on graphs with a single edge type and node type (homogeneous graphs). As real-world knowledge graphs are multi-relational, this is the next natural step. As benchmark, I use the Wikialumni dataset (Abboud and İsmail İlkan Ceylan 2021) which is a knowledge graph extracted from YAGO augmented with Wikipedia text as node features that is adapted to node classification on multi-relational graphs. In this setting, the formulation of rules with multiple binary predicates that correspond to various relations in the graph can be investigated.
Status: In Progress. Planned to be finished by 03/2024.

Ontology-Based Link Prediction

The beforementioned works tackle the task of node classification and are focused on attributed, homogeneous graph. They are not directly applicable to the incomplete, noisy and heterogeneous nature of real-world knowledge graphs. As main work of my PhD, I am developing a model that conducts link prediction on knowledge graphs and is able to reason with logical rules. Knowledge graph embedding models produce scores that represent the plausibility of a candidate triple in the context of link prediction. Scores that can be interpreted as probabilities can serve as input to a reasoning model, such as scallop (Li, Huang, and Naik 2023), for example. The whole architecture, including a knowledge graph embedding component and a reasoning component has to be differentiable. This allows to learn indirectly supervised predicates that can be described through rules such as $\texttt{parent(x,y)}$ and $\texttt{parent(y,z)} \implies \texttt{grandparent(x,z)}$.

This task poses multiple challenges. Knowledge graphs are sparse so that the amount of nodes that are not connected is much larger than the amount of nodes that are connected. Predicting probabiltiy scores for all potential triples results in a 3D tensor of size $\mathbb{R}^{|\text{entities}| \times |\text{entities}| \times |\text{relations}|}$. Since multiple relations can exist between two nodes, this amounts to a multi-label link prediction task. Consequently, one major challenge of this approach is scalability. The 3D tensor of predictions may exceed GPU memory capacity even for medium size graphs. Graph-specific sampling methods can be potential solutions. Another challenge is that knowledge graph embedding methods are usually trained under the closed world assumption. All triples that are not explicitly observed are assumed to be false. This is inconsistent with the assumption that knowledge graphs are incomplete. The model to be developed should also have the property of modeling uncertainties.

Status: In Progress. Planned to be completed by July 2024.

Acknowledgements

This work has been partially supported by the MIAI Knowledge communication and evolution chair (ANR-19-P3IA-0003).

References

Abboud, R.; and İsmail İlkan Ceylan. 2021. Node Classification Meets Link Prediction on Knowledge Graphs. arXiv:2106.07297.

Dai, Y.; Wang, S.; Xiong, N. N.; and Guo, W. 2020. A Survey on Knowledge Graph Embedding: Approaches, Applications and Benchmarks. *Electronics*, 9(5).

Daniele, A.; and Serafini, L. 2019. Knowledge Enhanced Neural Networks. In Nayak, A. C.; and Sharma, A., eds., *PRICAI 2019: Trends in Artificial Intelligence*, 542–554. Cham: Springer International Publishing. ISBN 978-3-030-29908-8.

Daniele, A.; and Serafini, L. 2023. Knowledge Enhanced Neural Networks for Relational Domains. In Dovier, A.; Montanari, A.; and Orlandini, A., eds., *AIxIA 2022 – Advances in Artificial Intelligence*, 91–109. Cham: Springer International Publishing. ISBN 978-3-031-27181-6.

Li, Z.; Huang, J.; and Naik, M. 2023. Scallop: A Language for Neurosymbolic Programming. *Proc. ACM Program. Lang.*, 7(PLDI).

Sarker, M. K.; Zhou, L.; Eberhart, A.; and Hitzler, P. 2021. Neuro-symbolic artificial intelligence. *AI Communications*, 34(3): 197–209.

Werner, L.; Layaïda, N.; Genevès, P.; and Chlyah, S. 2023. Knowledge Enhanced Graph Neural Networks. *2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA)*, 1–10.

Wu, L.; Cui, P.; Pei, J.; and Zhao, L. 2022. *Graph Neural Networks: Foundations, Frontiers, and Applications*. Singapore: Springer Singapore.