Mixing symbolic and sub-symbolic representations in machine learning

Context - For several decades, ontologies, i.e. structured representation of domain concepts and their relations [3], have been promoted as the appropriate tool for making domain knowledge available for machine processing [4]. Injecting domain knowledge into a neural learner to alleviate reliance on high-quality data and improve explainability is a rapidly expanding research trend [6]. While most of the effort focused on regular topology formats such as sequences (e.g. text) and grids (e.g. image), our focus is graph data (e.g. knowledge graphs). We are investigating a novel approach to inject the knowledge from a domain ontology into a neural learning process [5]. It boils down to mining specific flavor of graph patterns and feed them as high-level features for the neural net. The expected impact of our ontologically-generalized graph-shaped features is multi-fold: improved learning accuracy, generalization capacity enhancement, interpretability, etc. The final approach is to be validated within a precision farming project aiming at optimizing milk production in Quebec dairy farms [2].

Research question - While designing, extracting and filtering such patterns is a challenge in itself, bridging the gap between patterns and a neural learner is not to be underestimated. Indeed, given the stark *impedance mismatch* between the symbolic representation mode of an ontology and the subsymbolic one of a neural network, the main research question becomes proper way of injecting the graph patterns into a neural learner so that their positive impact could be maximized.

Approach - A first, and rather naive, approach would be to encode the raw graph data with patterns as boolean features. However, a vast array of more subtle alternatives could be put to use. For example, recently Knowledge-Infused Learning [6] proposes to insert domain knowledge directly into low-level components of a neural learner (e.g. its loss function). Alternatively, patterns can be used to constrain entity embeddings as per [1]. Even more ambitious, and less explored, tracks could be investigated such as ontological pattern-based graph kernels or attention mechanisms, contextual ranking of patterns, etc.

Goal - Design one or more comprehensive methods for injecting ontological graph patterns into the neural learning process and practically asses it using the dairy production data.

Strategy - A sufficiently comprehensive study of the respective limitations of pattern-based representations and their sub-symbolic counterparts should be the starting point of the internship. The next step will be to implement and, at least partially, evaluate the above straightforward approach (patterns as features). The final step, based on the previous one, the design and implementation of one novel method for mixing original data and ontological graph patterns will be pursued, e.g. for direct or indirect embedding.

Environment, technical and non technical requirements - A theoretical understanding of deep neural networks (especially recurrent and convolutional) and mastering of deep learning platforms (TensorFlow, PyTorch) is key. Knowledge about graph neural networks [7] is also desirable, yet not a must. Basic knowledge of ontologies and the semantic web technology stack could help. Implementation will be Python and Java.

The intern will work in tight collaboration with the professor and a PhD candidate, Tomas Martin. Duration is 5 to 6 months.

References

[1] A. Bordes et al. A semantic matching energy function for learning with multi-relational data: Application to word-sense disambiguation. *Machine Learning*, 94(2):233–259, Feb. 2014.

[2] C. Frasco et al. Towards an Effective Decision-making System based on Cow Profitability using Deep Learning:. In *Proc. of the 12th ICAART*, pages 949–958, Valletta, Malta, 2020.

[3] T. Gruber et al. A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2):199–220, 1993.

[4] F. Kramer and T. Beißbarth. Working with ontologies. In *Bioinformatics*, pages 123–135. Springer, 2017.

[5] T. Martin et al. Leveraging a domain ontology in (neural) learning from heterogeneous datas. In *Proc. of CSSA@CIKM20*, 2020.

[6] A. Sheth et al. Shades of knowledge-infused learning for enhancing deep learning. *IEEE Internet Computing*, 23(6):54–63, Nov 2019.

[7] K. Xu et al. How powerful are graph neural networks? *arXiv:1810.00826 [cs, stat]*, Feb 2019.