Learning Diverse Rules for Knowledge Graph Completion using Generative Models

1 Context

Knowledge graphs (KGs) are sets of triples of the form (subject, predicate, object), such as e.g. (*Paris*, *capital of*, *France*). KGs have emerged as the *de facto* standard for representing knowledge across both industry and academia. As KGs are notoriously incomplete, there is an active line of research that is dedicated to the problem of automatically finding plausible missing triples in KGs, a problem that is usually referred to as *link prediction*. The standard approach to solving this problem consists in learning a so-called KG embedding. Specifically, for each entity e, we learn a corresponding vector $\mathbf{e} \in \mathbf{R}^n$, called an entity embedding. Furthermore, for each relation r we learn a corresponding scoring function $s_r : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ such that $s_r(\mathbf{e}, \mathbf{f})$ reflects our confidence that the triple (e, r, f) is valid. For instance, in the seminal TransE model [1], the scoring function takes the following form: $s_r(\mathbf{e}, \mathbf{f}) = \|\mathbf{e} - \mathbf{f} - \mathbf{r}\|_2$, where $\mathbf{r} \in \mathbb{R}^n$ is a vector which intuitively represents the relation r. The entity embeddings $\mathbf{e}$ and the parameters of the scoring function s_r are learned by maximising the log-likelihood of a given knowledge graph. The underlying intuition is that these representations will then capture the regularities that exist in the KG. The criterion $s_r(\mathbf{e}, \mathbf{f})$ will thus allow us to uncover triples which are plausible but not actually present in the KG.

A very different approach consists in learning rules that can be used for finding plausible missing triples. For instance, we could have the following rule:

$$(X, \textit{lives in}, Y) \wedge (Y, \textit{has national language}, Z) \to (X, \textit{speaks}, Z)$$

Here X, Y, Z are variables, which can be instantiated with entities from the knowledge graph. The rule intuitively states that if someone lives in a country, then they are typically able to speak the official language from that country. Note that this is a default rule, which is often, but not necessarily, satisfied. The well-known AnyBURL method [2] learns a large set of such default rules from a given KG, where for each rule a corresponding confidence degree is also learned. This method has important advantages over KG embedding methods. For instance, it is completely **transparent**, as we know exactly why each predicted triple has been inferred. This means that we can provide explanations, or allow users to tweak how the method operates (e.g. by manually adding or removing rules). Furthermore, rule-based methods are also significantly more **efficient** than other methods. Unfortunately, however, their performance, while generally quite good, does not match what is possible with state-of-the-art methods.

The main limitation of rule based methods is that they can only learn rules that are witnessed sufficiently frequently in the given KG. In some domains, however, to perform well, we would have to learn many different variations of the same rule, including variations that are not actually occurring in the training data. This is illustrated, for a toy example, in Table 1. The rules in this table are concerned with the problem of predicting crime hotspots [3]. Suppose that the first three rules have been learned from the training data. Intuitively, these rules partially characterise the spatio-temporal diffusion pattern of crime hotspots. Now consider the last rule. Even if the situation that is described in the antecedent of this rule never occurs in the training data, the rule nonetheless seems plausible because the situation it covers is similar to the situations that are covered by the first three rules. The aim of this internship will be to address the fact that such rules cannot be learned using standard methods.

Specifically, the approach that will be taken is to use the given KG to learn a generative model, from which we can sample small sub-graphs. These samples may correspond to sub-graphs that occur in the given KG, but some samples will also be novel sub-graphs. Our hypothesis is that these samples will capture meaningful variations of the situations that arise in the KG itself. Therefore, by using a large number of sampled sub-graphs for learning rules, we should be able to learn a broader variety of rules, and thus outperform existing rule based methods.

if	*burglary*$(L, T-2)$, *burglary*$(L, T-1)$	**then** *burglary*(L, T)
if	*burglary*$(L, T-1)$, *burglary*$(L_1, T-1)$, *burglary*$(L_2, T-1)$, $n(L, L_1)$, $n(L, L_2)$, $L_1 \neq L_2$	**then** *burglary*(L, T)
if	*burglary*$(L, T-2)$, *burglary*$(L_1, T-1)$, *burglary*$(L_2, T-1)$, $n(L, L_1)$, $n(L, L_2)$, $L_1 \neq L_2$	**then** *burglary*(L, T)
if	*burglary*$(L_1, T-2)$, *burglary*$(L_2, T-2)$, *burglary*$(L, T-1)$, $n(L, L_1)$, $n(L, L_2)$, $L_1 \neq L_2$	**then** *burglary*(L, T)

Table 1: *Example rules about crime prediction, where burglary*(L, T) *means that a burglary took place in location* L *at time* T *and* $n(L, L_1)$ *means that* L *and* L_1 *are neighbouring locations.*

2 Objectives

The specific objectives of this internship will be two-fold:

1. Implement and compare different generative models of KGs. The considered models may involve generative graph neural networks, as well as path based methods (e.g. based on LSTMs).

2. Assess the effectiveness of the proposed strategy for learning rule-based KG completion methods.

3 Internship Details

The internship is proposed for a duration of 6 to 8 weeks at Cardiff University's School of Computer Science & Informatics. It will be supervised by Prof Steven Schockaert (`schockaerts1@cardiff.ac.uk`) and Dr Akash Anil (`AnilA@cardiff.ac.uk`).

References

[1] A. Bordes, N. Usunier, A. García-Durán, J. Weston, and O. Yakhnenko. Translating embeddings for modeling multi-relational data. In C. J. C. Burges, L. Bottou, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 2787–2795, 2013.

[2] C. Meilicke, M. W. Chekol, D. Ruffinelli, and H. Stuckenschmidt. Anytime bottom-up rule learning for knowledge graph completion. In S. Kraus, editor, *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 3137–3143. ijcai.org, 2019.

[3] Y. Zhang and T. Cheng. Graph deep learning model for network-based predictive hotspot mapping of sparse spatio-temporal events. *Comput. Environ. Urban Syst.*, 79, 2020.