Relational Learning to Capture the Dynamics and Sparsity of Knowledge Graphs

Mehrnoosh Mirtaheri
USC Information Sciences Institute
mehrnoom@usc.edu

Abstract

The rapid growth of large scale event datasets with timestamps has given rise to the dynamically evolving multi-relational knowledge graphs. Temporal reasoning over such data brings on many challenges and is still not well understood. Most real-world knowledge graphs are characterized by a long-tail relation frequency distribution where a significant fraction of relations occurs only a handful of times. This observation has given rise to the recent interest in low-shot learning methods that are able to generalize from only a few examples. The existing approaches, however, are tailored to static knowledge graphs and not easily generalized to temporal settings, where data scarcity poses even bigger problems, due to the occurrence of new, previously unseen relations. The goal of my doctoral research is to introduce new approaches for learning meaningful representation that capture the dynamics of temporal knowledge graphs while tackling various existing challenges such as data scarcity.

Introduction

Large-scale knowledge graphs (KGs) have become a crucial component for performing various Natural Language Processing (NLP) tasks, including cross-lingual translation, Q&A, and relational learning. KGs typically suffer from incompleteness, making automatic KG completion, i.e. predicting missing links in KG, crucial to avoid the potential performance degradation in the downstream tasks.

Previous methods of KG completion have traditionally focused on learning representation over static knowledge graphs. Due to the rapid growth of temporal event datasets, there has also been significant recent interest in learning with Temporal Knowledge Graphs (TKG). The existing methods, however, suffer from major drawbacks. First, most of the existing KG completion methods rely on a sufficiently large number of training examples per relation. This could become an issue as most real-world KGs have a long-tail structure so that many relationships occur only a handful of times. The data scarcity issue is exacerbated for temporal graphs, where the distribution of occurrences of certain events over time can be highly inhomogeneous and bursty, and in fact, new types of relationships might emerge that have not been observed before. While recent research has addressed the data scarcity issue for static KGs, existing approaches cannot account for the dynamics in TKGs. Finally, most of the existing methods disregard the streaming nature of the incoming data in real-world KGs. Therefore, they have to be re-trained which could potentially be very time-consuming, or fine-tuned which could cause over-fitting.

The focus of my doctoral research is to introduce models that not only capture the relational dynamics of the TKGs but also tackle the data scarcity problem.

Completed Research

A TKG can be represented as a set of quadruples $G = \{(s, r, o, t)|s, o \in \mathcal{E}, r \in \mathcal{R}\}$, where $\mathcal{E}$ is the set of entities, $\mathcal{R}$ is the set of relations and t is the timestamp. KG completion for a static KG involves predicting new facts by either predicting an unseen object entity for a given subject and relation $(s, r, ?)$ or predicting an unseen link between the subject and object entity $(s, ?, o)$. This section includes a brief explanation of my work on building models for temporal knowledge graphs completion. Both proposed models are evaluated on two popular benchmarks for TKG completion tasks and outperform the state-of-the-art baselines.

Tensor-based Method for Temporal Geopolitical Event Forecasting (Mirtaheri et al. 2019)

In this work, we want to predict new events $(s, ?, o, t)$ at time t by predicting an unseen link between subject and object entity. We represent the interaction data as a 4-dimensional Tensor M of size $|\mathcal{E}| \times |\mathcal{E}| \times T \times |\mathcal{R}|$, where T is the number of time steps, and $|\mathcal{R}|$ is the number of relations and an entry m_{sotr} on matrix M corresponds to the number of interactions of type r from entity s to o at timestamp t. Given Tensor M, we want to extrapolate entries along the third (i.e. time) dimension. Specifically, we want to output a tensor with dimensions $|\mathcal{E}| \times |\mathcal{E}| \times C \times |\mathcal{R}|$, where C is the number of time steps we would like to predict in the future. This output tensor is an estimate of the number of different interactions, between every entity pair, that will happen in the upcoming C time steps. Our proposed algorithm includes the following steps:

Tensor Factorization. Tensor factorization methods identify the underlying hidden structure of the data. More specifically, the 4-dimensional matrix M can be factorized into four low-rank factor matrices $\theta^S \in \mathbb{R}^{|\mathcal{E}| \times N}, \theta^O \in \mathbb{R}^{|\mathcal{E}| \times N}, \theta^T \in \mathbb{R}^{T \times N}$, and $\theta^R \in \mathbb{R}^{|\mathcal{R}| \times N}$, and their outer tensor product should recover M. We use Bayesian Poisson Tensor Factorization (BPTF) proposed in (Schein et al. 2015), which is a probabilistic approach for identifying the latent structures. BPTF assumes that m_{sotr} is coming from a Poisson distribution (as it is suitable for count data).

Forecasting. We extrapolate θ^T producing θ^C through a simple autoregressive convolutional model. In particular, we train convolutional filter $W \in \mathbb{R}^{l \times N \times N}$, where filter height l allows us to process l timesteps in the past for predicting a single timestep. Given $\theta^S, \theta^O, \theta^R$ and extrapolated rows θ^C, we predict the future tensor using the PARAFAC method.

One-shot Learning for Temporal Knowledge Graphs (Mirtaheri et al. 2020)

As mentioned earlier, data scarcity is an even bigger problem in relational learning with TKGs. Few-shot episodic training has been proven to be effective to tackle this problem for static KGs (Xiong et al. 2018). In this work we want to predict new events $(s, r, ?, t)$ at time t by ranking the true object entities higher than others, under a hard condition where there is only one training example for each relation. The goal is to learn a metric space, that can be used during the inference to generate a similarity score between the one given example and a given query (a potential event). The similarity score is proportional to the likelihood of that event.

We extend the framework proposed by Xiong et al. 2018 for TKG completion. We divide the relations of a given TKG into two groups based on their frequency: frequent relations $\mathcal{F}$ and sparse relations $\mathcal{T}$. We assume to have a set of tasks where each task corresponds to a sparse relation $r \in \mathcal{T}$, and has its own training and test set denoted as support and query set respectively defined as follows:

$$\begin{aligned} S_r^t &= \{(s_0, r, o_0, t_0) | s_0, o_0 \in \mathcal{E}\} \\ Q_r^t &= \{(s_q, r, o_q, t_q) | s_q, o_q \in \mathcal{E}, t_q \in [t, t+w]\}, \end{aligned} \tag{1}$$

The loss function l_θ at each episode optimizes a score function $\mathcal{P}_\theta$ that ranks the true test queries in Q_r^t, higher than the others. The final optimization loss is:

$$\mathcal{L} = E_{r \sim \mathcal{T}} \left[E_{Q_r^t \sim G, S_r^t \sim G} \left[l_\theta(Q_r^t | S_r^t) \right] \right] \tag{2}$$

The relations in $\mathcal{T}$ are divided into *meta-train*, *meta-val* and *meta-test* relations, and any pair of these sets are mutually exclusive, meaning that the model can handle unseen relations during the test time. We assume there is also no time overlap between the quadruples in the *meta-train*, *meta-val* and *meta-test*. Finally, we assume that the model has access to a background knowledge graph defined as $G' = \{(s, r, o, t) | s, o \in \mathcal{E}, r \in \mathcal{F}\}$, and the entity set $\mathcal{E}$ is a closed set, i.e., there are no unseen entities during the inference time. Our model is comprised of two major components:

Neighborhood Encoder. The neighborhood encoder represents the neighborhood information of a given entity e as a d dimensional vector h_e and preserves the relational/sequential graph structure. It encodes the one-hop neighborhood during the past l timesteps as a sequence. The adjacent nodes at each timestamp are aggregated and given to a self-attention network to make a time-aware neighborhood representation.

Metric Learning. A similarity function parameterized by a neural network, $\mathcal{M}(q, S)$, that takes the representation of the support S and a potential event $q = (s_q, r, o_q, t)$ as input, and outputs a scalar similarity score between them proportional to the likelihood of the event.

Our model improves the link prediction performance over the sparse relations by 20%.

Current Research

Due to the dynamic nature of knowledge graphs, it might become necessary to do a frequent model re-training, which requires time and memory for keeping a large training dataset. Instead, we are working on developing efficient models for KG and TKG completion capable of handling new relations and entities and do not require re-training. Our goal is to employ the continual learning (Song and Park 2018; Zhou et al. 2020) framework to avoid the model over-fitting and catastrophic forgetting. We have designed the problem setup and are now conducting initial experiments.

Future Work

To make the technology useful for large-scale real-world applications, we would like to combine continual learning and meta learning approaches to engineer an end-to-end system that works for a broad variety of knowledge graphs. Specifically, we would like to exploit synergies between different problems so that learning a task or set of tasks for one KG can help to inform the learning process for another KG.

References

Mirtaheri, M.; Abu-El-Haija, S.; Hossain, T.; Morstatter, F.; and Galstyan, A. 2019. Tensor-based Method for Temporal Geopolitical Event Forecasting. *ICML Workshop on Learning and Reasoning with Graph-Structured Representations, Long Beach, CA* .

Mirtaheri, M.; Rostami, M.; Ren, X.; Morstatter, F.; and Galstyan, A. 2020. One-shot Learning for Temporal Knowledge Graphs. *arXiv preprint arXiv:2010.12144* .

Schein, A.; Paisley, J.; Blei, D. M.; and Wallach, H. 2015. Bayesian poisson tensor factorization for inferring multilateral relations from sparse dyadic event counts. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1045–1054. ACM.

Song, H.-J.; and Park, S.-B. 2018. Enriching Translation-Based Knowledge Graph Embeddings Through Continual Learning. *IEEE Access* 6: 60489–60497.

Xiong, W.; Yu, M.; Chang, S.; Guo, X.; and Wang, W. Y. 2018. One-Shot Relational Learning for Knowledge Graphs. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 1980–1990.

Zhou, F.; Cao, C.; Zhong, T.; Zhang, K.; Trajcevski, G.; and Geng, J. 2020. Continual Graph Learning. *arXiv preprint arXiv:2003.09908* .