Know-Evolve: Deep Temporal Reasoning for Dynamic Knowledge Graphs (Supplementary Material)

Rakshit Trivedi [1] Hanjun Dai [1] Yichen Wang [1] Le Song [1]

Appendix

A. Algorithm for Global BPTT Computation

As mentioned in Section 4 of main paper, the intricate relational and temporal dependencies between data points in our setting limits our ability to efficiently train by decomposing events into independent sequences. To address this challenge, we design an efficient Global BPTT algorithm presented below. During each step of training, we build computational graph using consecutive events in the sliding window of a fixed size. We then move sliding window further and train till the end of timeline in similar fashion which allows to capture dependencies across batches while retaining efficiency.

Algorithm 2 Global-BPTT

```
Input: Global Event Sequence O, Steps s, Stopping Condition max_iter
cur_index = 0, t_begin = 0
for iter = 0 to max_iter do
  if cur_index > 0 then
    t_begin = O[cur_index − 1] → t
  end if
  e_mini_batch = O[cur_index : cur_index + s]
  Build Training Network specific to e_mini_batch
  Feed Forward inputs over network of s time steps
  Compute Total Loss L over s steps:
  L = − Σ_{p=1}^{s} log(λ_r^{e^s,e^o}(t_p | t̄_p)) + Survival loss computed using Algorithm 1 in main paper
  Backpropagate error through s time steps and update all weights
  if cur_index + s > O.size then
    cur_index = 0
  else
    cur_index = cur_index + s
  end if
end for
```

[1] College of Computing, Georgia Institute of Technology. Correspondence to: Rakshit Trivedi <rstrivedi@gatech.edu>, Le Song <lsong@cc.gatech.edu>.

Proceedings of the 34^{th} International Conference on Machine Learning, Sydney, Australia, PMLR 70, 2017.

B. Data Statistics and Sparsity of Knowledge Tensor

Table 1. Statistics for each dataset.

Dataset Name	# Entities	# Relations	# Events
GDELT-full	14018	20	31.29M
GDELT-500	500	20	3.42M
ICEWS-full	12498	260	0.67M
ICEWS-500	500	256	0.45M

Table 2. Sparsity of Knowledge Tensor.

Dataset Name	# Possible Entries	# Available Entries	% Proportion
GDELT-full	3.93B	4.52M	0.12
GDELT-500	5M	0.76M	15.21
ICEWS-full	39.98B	0.31M	7e-3
ICEWS-500	64M	0.15M	0.24

C. Implementation Details

Know-Evolve. Both Algorithm 1 and Algorithm 2 demonstrate that the computational graph for each mini-batch will be significantly different due to high variations in the interactions happening in each window. To facilitate efficient training over dynamic computational graph setting, we leverage on graph embedding framework proposed in (Dai et al., 2016a) that allows to learn over graph structure where the objective function may potentially have different computational graph for each batch. We use Adam Optimizer with gradient clipping for making parameter updates. Using grid search method across hyper-parameters, we set mini-batch size = 200, weight scale = 0.1 and learning rate = 0.0005 for all datasets. We used zero initialization for our entity embeddings which is reasonable choice for dynamically evolving entities.

Competitors. We implemented all the reported baselines in Tensorflow and evaluated all methods uniformly. For each method, we use grid search on hyper-parameters and embedding size and chose the ones providing best performance in respective methods. All the baseline methods are trained using contrastive max-margin objective function described in (Socher et al., 2013). We use Adagrad optimization provided in Tensorflow for optimizing this objective function. We randomly initialize entity embeddings as typically done for these models.

D. Parameter Complexity Analysis

We report the dimensionality of embeddings and the resulting number of parameters of various models. Table 3 illustrates that Know-Evolve is significantly efficient in the number of parameters compared to Neural Tensor Network while being highly expressive as demonstrated by its prediction performance in Section 5 of main paper. The overall number of parameters for different dataset configurations are comparable to the simpler relational models in order of magnitude.

Method	Memory Complexity	GDELT		ICEWS	
		$H_e/H_r/H_a/H_b$	# Params	$H_e/H_r/H_a/H_b$	# Params
NTN	$N_e^2 H_b + N_r(H_b + H_a) + 2N_r N_e H_a + N_e H_e$	100/16/60/60	11.83B	60/32/60/60	9.76B
RESCAL	$N_r H_e^2 + N_e H_e$	100/-/-/-	1.60M	60/-/-/-	1.69M
TransE	$N_e H_e + N_r H_e$	100/-/-/-	1.40M	60/-/-/-	0.77M
TransR	$N_e H_e + N_r H_r + N_r H_r^2$	100/20/-/-	1.41M	60/32/-/-	1.02M
ER-MLP	$N_e H_e + N_r H_r + H_a + H_a(2H_e + H_r)$	100/20/100/-	1.42M	60/32/60/-	0.77M
Know-Evolve	$H_e(N_e + N_r H_e) + N_r H_r + H_a * (2H_e + H_r) + H_a * H_b + 2H_b$	100/20/100/100	1.63M	60/32/60/60	1.71M

Table 3. Comparison of our method with various relational methods for memory complexity. Last two columns provide example realizations of this complexity in full versions for GDELT and ICEWS datasets. H_a and H_b correspond to hidden layers used in respective methods.H_e and H_r correspond to entity and relation embedding dimensions respectively. N_e and N_r are number of entities and relations in each dataset. For GDELT, $N_e = 14018$ and $N_r = 20$. For ICEWS, $N_e = 12498$ and $N_r = 260$. We borrow the notations from (Nickel et al., 2016a) for simplicity.

E. Exploratory Analysis

E.1. Temporal Reasoning

We have shown that our model can achieve high accuracy when predicting a future event triplet or the time of event. Here, we present two case studies to demonstrate the ability of evolutionary knowledge network to perform superior reasoning across multiple relationships in the knowledge graphs.

CASE STUDY I: ENEMY'S FRIENDS IS AN ENEMY

We concentrate on the prediction of a quadruplet (Cairo,Assault,Croatia,July 5,2015) available in test set. This event relates to the news report of an assault on a Croation prisoner in Cairo on July 6 2015. Our model gives rank-1 to the object entity Croatia while the baselines did not predict them well ($rank > 250$).

We first consider relationship characteristics for Cairo and Croatia. In the current train span, there are 142 nodes with which Cairo was involved in a relationship as a subject (total of 1369 events) and Croatia was involved in a relationship as an object (total of 1037 events). As a subject, Cairo was involved in an assault relationship only 59 times while as an object, Croatia was involved in assault only 5 times. As mentioned earlier, there was no direct edge present between Cairo and Croatia with relationship type assault.

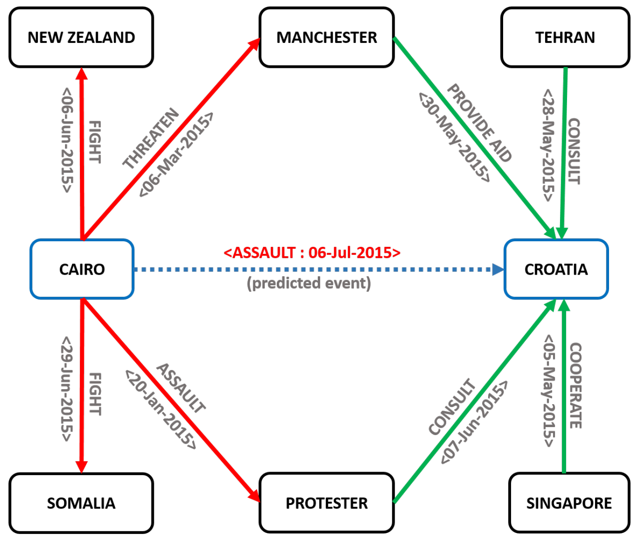


Figure 1. Relationship graph for Cairo and Croatia. Dotted arrow shows the predicted edge. Direction of the arrow is from subject to object entity.

While the conventional reasoning methods consider static interactions of entities in a specific relationship space, they fail to account for the temporal effect on certain relationships and dynamic evolution of entity embeddings. We believe that our method is able to capture this multi-faceted knowledge that helps to reason better than the competitors for the above case.

Temporal Effect. It is observed in the dataset that many entities were involved more in negative relationships in the last month of training data as compared to earlier months of the year. Further, a lot of assault activities on foreign prisoners were being reported in Cairo starting from May 2015. Our model successfully captures this increased intensity of such events in recent past. The interesting observation is that overall, Cairo has been involved in much higher number of positive relationships as compared to negative ones and that would lead conventional baselines to use that path to reason for new entity – instead our model tries to capture effect of most recent events.

Dynamic Knowledge Evolution. It can be seen from the dataset that Cairo got associated with more and more negative events towards the mid of year 2015 as compared to start of the year where it was mostly involved in positive and cooperation relationships. While this was not very prominent in case of Croatia, it still showed some change in the type of relationships over time. There were multiple instances where Cairo was involved in a negative relationship with a node which in turn had positive relationship with Croatia. This signifies that the features of the two entities were jointly and non-linearly evolving with the features of the third entity in different relationship spaces.

Below we provide reference links for the actual event news related to the edges in Figure 1.

Predicted Edge.

- (Cairo, Assault, Croatia, 06-Jul-2015): https://www.bloomberg.com/news/articles/2015-08-05/islamic-state-egypt-affiliate-threatens-to-kill-croatian-citizen

Other Edges.

- (Cairo, Assault, Protester, 20-Jan-2015):http://usa.chinadaily.com.cn/world/2015-04/22/content_20501452

- (Cairo, Threaten, Manchester, 06-Mar-2015): http://www.manchestereveningnews.co.uk/news/greater-manchester-news/anthony-filz-stashed-deadly-machine-8788541
- (Protester, Consult, Croatia, 07-Jun-2015): http://globalvoicesonline.org/2015/06/07/veterans-of-croatias-war-of-independence-are-still-knocking-on-the-governments-door/
- (Manchester, Provide Aid, Croatia, 30-May-2015): http://www.offthepost.info/blog/2015/05/liverpool-meet-inter-to-discuss-mateo-kovacic-deal/

CASE STUDY II: COMMON ENEMY FORGES FRIENDSHIP

We concentrate on the prediction of a quadruplet (Colombia,Engage in Material Cooperation,Ottawa,July 2 2015) available in test set. This event relates to the news report of concerns over a military deal between Colombia and Canada on July 2 2015 and reported in Ottawa Citizen. Our model gives rank-1 to the object entity Ottawa while the other baselines do not predict well ($rank > 250$). The above test event is a new relationship and was never seen in training.

As before, we consider relationship characteristics between Colombia and Ottawa. In the current train span, there are 165 nodes for which Colombia was involved in a relationship with that node as a subject (total of 1604 events) and on the other hand, Ottawa was involved in a relationship with those nodes as an object total of 733 events). As a subject, Colombia was involved in a cooperation relationship 71 times while as an object, Ottawa was involved in cooperation 24 times.

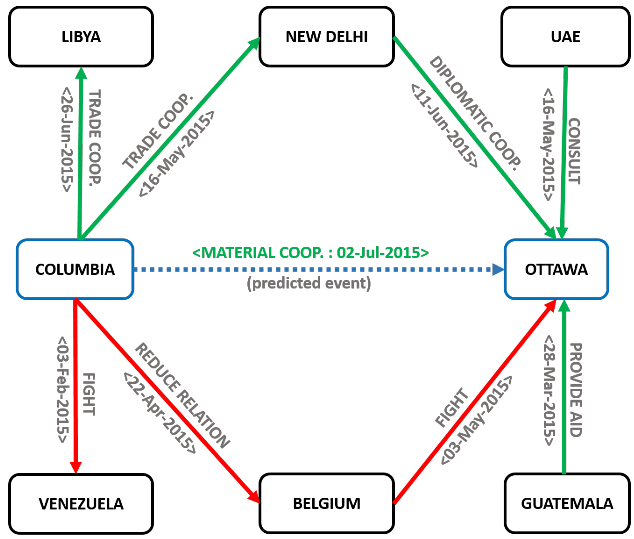


Figure 2. Relationship graph for Columbia and Ottawa. Dotted arrow shows the predicted edge. Direction of the arrow is from subject to object entity.

Temporal Effect. It is observed in the dataset that Colombia has been involved in hundreds of relationships with Venezuela (which is natural as they are neighbors). These relationships range across the spectrum from being as negative as “fight” to being as positive as “engagement in material cooperation”. But more recently in the training set (i.e after May 2015), the two countries have been mostly involved in positive relationships. Venezuela in turn has only been in cooperation relationship with Ottawa (Canada). Thus, it can be inferred that Colombia is affected by its more recent interaction with its neighbors while forming relationship with Canada.

Dynamic Knowledge Evolution. Overall it was observed that Colombia got involved in more positive relationships towards the end of training period as compared to the start. This can be attributed to events like economic growth, better living standards, better relations getting developed which has led to evolution of Colombia’s features in positive direction. The features for Ottawa (Canada) have continued to evolve in positive direction as it has been involved very less in negative relationships.

More interesting events exemplifying mutual evolution were also observed. In these cases, the relationship between Colombia and third entity were negative but following that relationship in time, the third entity forged a positive relationship with Ottawa (Canada). One can infer that it was in Colombia’s strategic interest to forge cooperation (positive relation) with Ottawa so as to counter its relationship with third entity. Below we provide reference links for the actual event news related to the edges in Figure 2.

Predicted Edge.

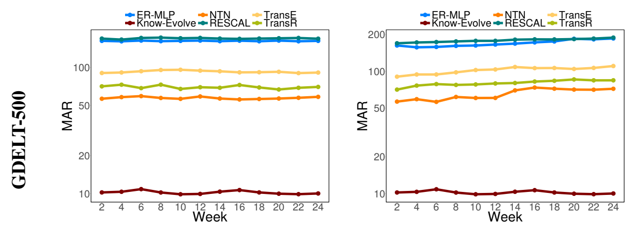
- (Columbia, Material Coop., Ottawa, 02-Jul-2015): http://ottawacitizen.com/news/politics/report-on-military-executions-casts-shadow-over-lav-deal-with-colombia

Other Edges.

- (Columbia, Trade Coop., New Delhi, 16-May-2015): http://www.newindianexpress.com/business/2015/may/16/Petroleum-Minister-Dharmendra-to-Lead-Business-Delegation-to-Mexico-Colombia-761494.html
- (Columbia, Fight, Venezuela, 03-Feb-2015):http://www.turkishpress.com/news/421947/
- (New Delhi, Diplomatic Coop., Ottawa, 28-May-2015):http://www.marketwatch.com/story/art-of-living-set-to-showcase-the-yoga-way-2015-06-11-61734555
- (Belgium, Fight, Ottawa, 05-May-2015): https://www.durhamregion.com/news-story/5597504-9-facts-about-in-flanders-fields-on-its-100th-anniversary/

E.2. Sliding Window Training Experiment

Unlike competitors, the entity embeddings in our model get updated after every event in the test, but the model parameters remain unchanged after training. To balance out the advantage that this may give to our method, we explore the use of sliding window training paradigm for baselines: We train on first six months of dataset and evaluate on the first test window. Next we throw away as many days (2 weeks) from start of train set as found in test set and incorporate the test data into training. We retrain the model using previously learned parameters as warm start. This can effectively aid the baselines to adapt to the evolving knowledge over time. Figure 3 shows that the sliding window training contributes to stable performance of baselines across the time window (i.e.the temporal deterioration is no longer observed significantly for baselines). But the overall performance of our method still surpasses all the competitors.

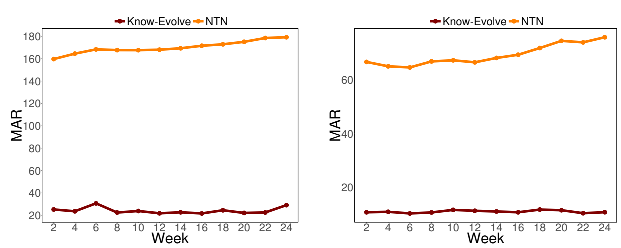


(c) Sliding Window Training (d) Non-sliding window Training

Figure 3. Performance comparison of sliding window vs. non-sliding window training (in terms of link prediction rank).

E.3. Recurrent Facts vs. New facts

One fundamental distinction in our multi-relational setting is the existence of recurrence relations which is not the case for traditional knowledge graphs. To that end, we compare our method with the best performing competitor - NTN on two different testing setups: 1.) Only Recurrent Facts in test set 2.) Only New facts in test set. We perform this experiment on GDELT-500 data. We call a test fact “new” if it was never seen in training. As one can expect, the proportion of new facts will increase as we move further in time. In our case, it ranges from 40%-60% of the total number of events in a specific test window. Figure 4 demonstrates that our method performs consistently and significantly better in both cases.



(a) New facts only (b) Recurrent Facts Only

Figure 4. Comparison with NTN over recurrent and non-recurrent test version.