Schema-Guided Event Graph Completion

Hongwei Wang[1], Zixuan Zhang[1], Sha Li[1], Jiawei Han[1], Yizhou Sun[2], Hanghang Tong[1], Joseph P. Olive[3], Heng Ji[1]

[1] University of Illinois Urbana-Champaign
[2] University of California, Los Angeles
[3] JP Olive S&T Management LLC

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

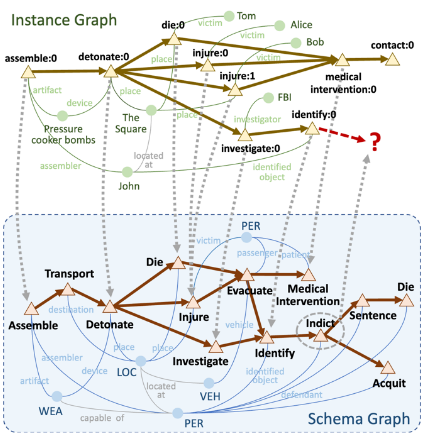
BLENDER

AUTOMATED KNOWLEDGE BASE CONSTRUCTION 2022 | London

INTRODUCTION

- We tackle a new task, event graph completion, which aims to predict missing event nodes for event graph
- We utilize event schemas, a type of generalized representation that describes the stereotypical structure of event graphs, to address drawbacks of existing methods
- Our approach (1) first maps an instance event graph to a schema subgraph, then (2) it predicts whether a candidate event node in the schema graph should be instantiated by characterizing two aspects of local topology: neighbors and paths. (3) The neighbor module and the path module are later combined for final prediction

SCHEMA-GUIDED EVENT GRAPH PREDICTION



- **Task**: given an incomplete instance event graph (above), predict its missing events (the red question mark)
- **Insight**: Map events in the instance graph to nodes in the schema graph. The event "Indict" in the grey circle is very likely to be the missing event

THE PROPOSED METHOD

1. Subgraph Matching

- The exact graph matching between instance graph I and schema graph S is a NP-hard problem. Therefore, we propose a two-step heuristic approach:
- For a node $e_i \in I$ to be matched, map e_i to a node in S that has the same event type with e_i
- If more than one node in S can be matched, select the node in S that has the most similar neighborhood structure as the matched node

2. Neighborhood Module

- Using graph neural networks to learn the representation of nodes and subgraphs in the event schema graph:

$$\mathbf{h}_i^k = \sigma\Big(\mathbf{W}^k \sum\nolimits_{j \in \mathcal{N}(i)\cup\{i\}} \frac{1}{\sqrt{|\mathcal{N}(i)|\cdot|\mathcal{N}(j)|}} \mathbf{h}_j^{k-1} + \mathbf{b}^k\Big) \qquad \mathbf{h}_{I'}^K = \text{READOUT}\left(\{\mathbf{h}_{e_i}^K\}_{e_i \in I'}\right)$$

- The probability that e is missing for I': $p_{neighbor}(e, I') = \text{MLP}_{neighbor}\left([\mathbf{h}_e^K, \mathbf{h}_{I'}^K]\right)$

3. Path Module

- Collect all paths (length $\le L$) connecting node e and subgraph I', and represent it as a multi-hot bag-of-path vector $\mathbf{p}_{e \to I'}^{\le L}$
- The probability that e is missing for I': $p_{path}(e, I') = \text{MLP}_{path}\left(\mathbf{p}_{e \to I'}^{\le L}\right)$

Final prediction: $f(e, I') = \left(p_{neighbor}(e, I') + p_{path}(e, I')\right) / 2$

EXPERIMENTS

- **Statistics of the three datasets**:

Dataset	Car-Bombings	IED-Bombings	Suicide-IED	Pandemic
# train/val/test instance graphs	75 / 9 / 10	88 / 11 / 12	176 / 22 / 22	40 / 5 / 6
# train/val/test samples	2,368 / 288 / 320	2,904 / 363 / 396	5,808 / 726 / 726	3,200 / 400 / 480
Corresponding schema name	Car-IED	General-IED		Disease-Outbreak
# event/entity nodes	32 / 134	33 / 140		102 / 17
# ev-ev/ev-en/en-en links	41 / 138 / 261	42 / 143 / 530		200 / 75 / 1

- **Binary classification task**: predict if a given event is missing for a given event subgraph:

Dataset	Car-Bombings		IED-Bombings		Suicide-IED		Pandemic	
Metrics	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
AddAll	55.3	50.0	35.3	50.0	40.4	50.0	19.0	50.0
AddNeighbor	59.4	56.9	58.6	64.1	60.9	64.5	64.8	56.8
ID-MLP	78.4 ± 3.2	88.9 ± 1.8	72.1 ± 0.8	85.7 ± 0.5	79.2 ± 0.7	88.3 ± 0.6	91.2 ± 0.6	96.3 ± 0.6
Type-MLP	80.2 ± 1.2	89.8 ± 0.8	72.4 ± 2.1	85.8 ± 0.9	79.9 ± 0.7	89.0 ± 0.4	91.4 ± 0.7	96.4 ± 0.4
TransE	79.5 ± 2.0	89.0 ± 1.9	73.3 ± 3.8	84.5 ± 3.1	78.4 ± 1.4	87.6 ± 0.9	90.5 ± 1.2	95.2 ± 1.0
RotatE	74.3 ± 4.0	86.0 ± 4.5	70.8 ± 2.6	86.5 ± 0.6	73.4 ± 1.3	85.1 ± 1.1	88.2 ± 1.5	94.0 ± 1.1
SEGC	82.8 ± 1.7	**92.3** ± 0.3	**81.5** ± 1.2	**88.9** ± 0.2	**82.4** ± 0.3	**90.0** ± 0.5	**92.8** ± 0.5	**97.6** ± 0.1
SEGC-neighbor	**83.6** ± 1.4	92.1 ± 0.4	80.0 ± 2.1	88.8 ± 0.6	81.8 ± 0.6	89.8 ± 0.3	91.0 ± 0.3	95.6 ± 0.5
SEGC-path	80.9 ± 0.8	89.3 ± 0.7	79.5 ± 0.9	86.1 ± 0.4	81.8 ± 1.1	88.3 ± 0.8	91.9 ± 0.6	96.6 ± 0.2

- **Graph completion task**: predict all missing events for a given event subgraph:

Dataset	Car-Bombings		IED-Bombings		Suicide-IED		Pandemic	
Metrics	Jaccard	F1	Jaccard	F1	Jaccard	F1	Jaccard	F1
AddAll	17.3	27.8	8.3	15.0	9.9	17.3	3.2	6.1
AddNeighbor	16.5 ± 0.9	25.5 ± 1.9	10.9 ± 1.0	18.7 ± 1.6	12.4 ± 0.6	21.0 ± 1.1	4.7 ± 1.1	8.7 ± 2.0
ID-MLP	33.2 ± 1.9	46.6 ± 2.2	17.8 ± 3.3	28.2 ± 3.6	28.2 ± 2.4	39.1 ± 2.8	29.2 ± 5.0	38.0 ± 5.0
Type-MLP	38.7 ± 1.8	52.2 ± 2.2	18.7 ± 6.0	28.9 ± 7.5	33.6 ± 5.9	42.8 ± 5.3	30.3 ± 7.0	41.2 ± 6.4
TransE	34.9 ± 2.2	46.7 ± 2.9	19.6 ± 5.6	28.6 ± 6.3	30.2 ± 6.8	40.1 ± 6.4	27.3 ± 6.6	33.7 ± 8.2
RotatE	29.1 ± 6.5	41.8 ± 6.6	14.6 ± 3.7	23.5 ± 3.3	20.8 ± 2.4	30.1 ± 3.1	26.6 ± 5.8	31.9 ± 6.2
SEGC	**45.8** ± 5.0	**59.2** ± 5.4	**34.8** ± 3.1	**48.3** ± 3.2	**44.9** ± 3.8	54.3 ± 5.0	**33.1** ± 4.3	**45.5** ± 7.3
SEGC-neighbor	42.0 ± 5.2	56.4 ± 5.4	32.8 ± 9.1	43.4 ± 10.3	44.4 ± 7.4	**55.5** ± 6.1	25.6 ± 0.9	38.9 ± 0.7
SEGC-path	41.6 ± 7.2	53.8 ± 6.7	28.8 ± 5.5	42.4 ± 7.1	37.8 ± 5.6	49.8 ± 5.5	31.1 ± 7.5	43.0 ± 7.8

Conclusion

- We propose a new schema-guided method for event graph completion, which enables the model to predict missing events in instance event graphs
- We consider neighbors and paths when modeling the event schema graph to fully capture its high-order topological and semantic information
- Experimental results on four datasets and three schemas demonstrate that our method achieves state-of-the-art performance on event graph completion task