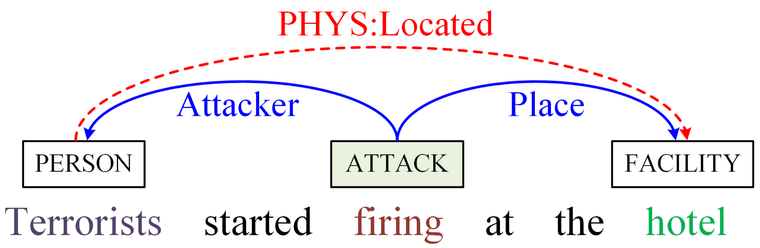
GATE: Graph Attention Transformer Encoder for Cross-Lingual Relation and Event Extraction

Wasi Uddin Ahmad, Nanyun Peng, and Kai-Wei Chang

Department of Computer Science, University of California, Los Angeles

Relation and Event Extraction

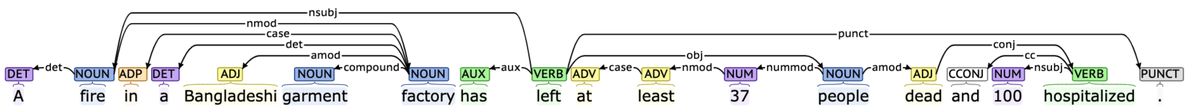


A relation (red dashed) between two entities and an event of type Attack (triggered by "firing") including two arguments and their role labels (blue) are highlighted.

Challenge: capturing long-range dependencies

A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized .

Distance (fire, hospitalized) = Sequential: 15, Structural: 4

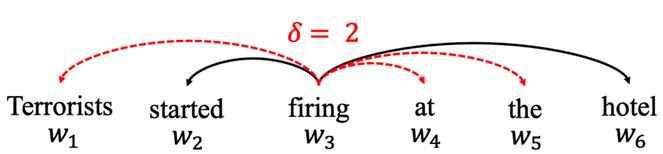


Dependency-guided self-attention

- Allow tokens to attend tokens that are within syntactic distance δ.
- Pay more attention to tokens that are closer and less attention to tokens that are faraway in the dependency tree.

$$A_l = F\left(\text{softmax}\left(\frac{QK^T}{\sqrt{d_k}} + M\right)\right) V_l.$$

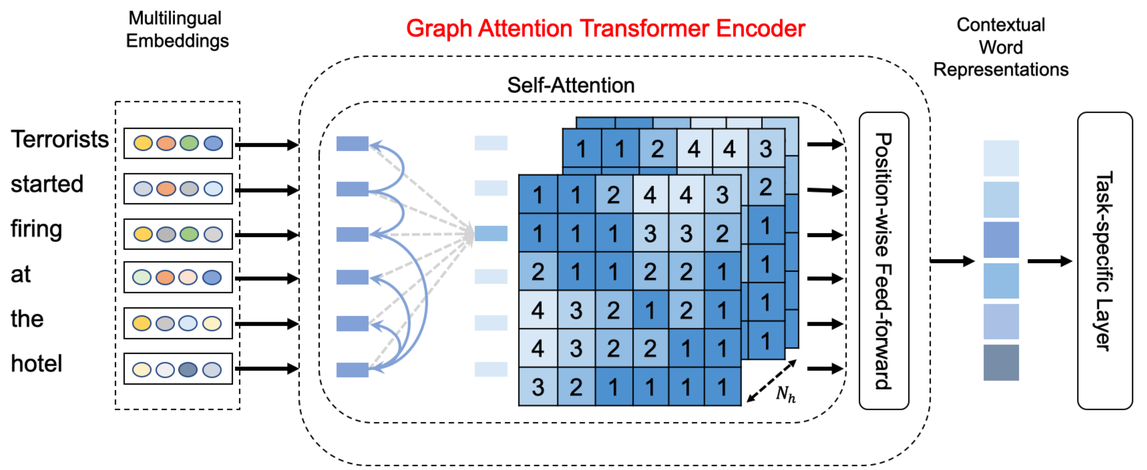
$$F(P)_{ij} = \frac{P_{ij}}{Z_i D_{ij}},\ M_{ij} = \begin{cases} 0, & D_{ij} \le \delta \\ -\infty, & \text{otherwise} \end{cases}$$



Syntactic Distance Matrix, D

	w_1	w_2	w_3	w_4	w_5	w_6
w_1	1	1	2	4	4	3
w_2	1	1	1	3	3	2
w_3	2	1	1	2	2	1
w_4	4	3	2	1	2	1
w_5	4	3	2	2	1	1
w_6	3	2	1	1	1	1

Graph Attention Transformer Encoder

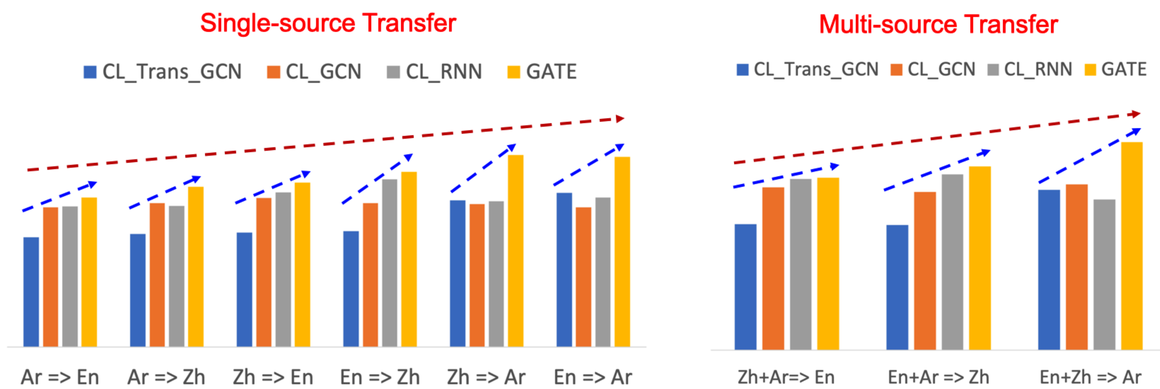


Evaluation Setup

Dataset: ACE 2005; Languages: English (En), Chinese (Zh), and Arabic (Ar).

Source code: https://github.com/wasiahmad/GATE

Zero-shot Cross-lingual Transfer



Acknowledgements

This work was supported in part by National Science Foundation (NSF) Grant OAC 1920462 and the Intelligence Advanced Research Projects Activity (IARPA) via Contract No. 2019-19051600007.