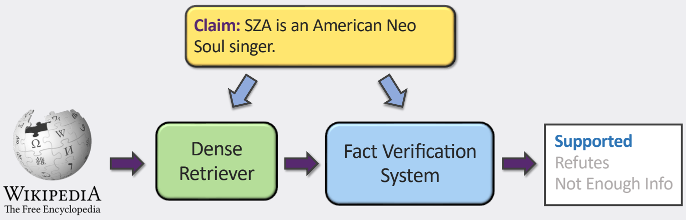
SISER: Semantic-Infused Selective Graph Reasoning for Fact Verification

Eunhwan Park♠, Jong-Hyeon Lee♠, Donghyeon Jeon, Seonhoon Kim, Inho Kang, Seung-Hoon Na

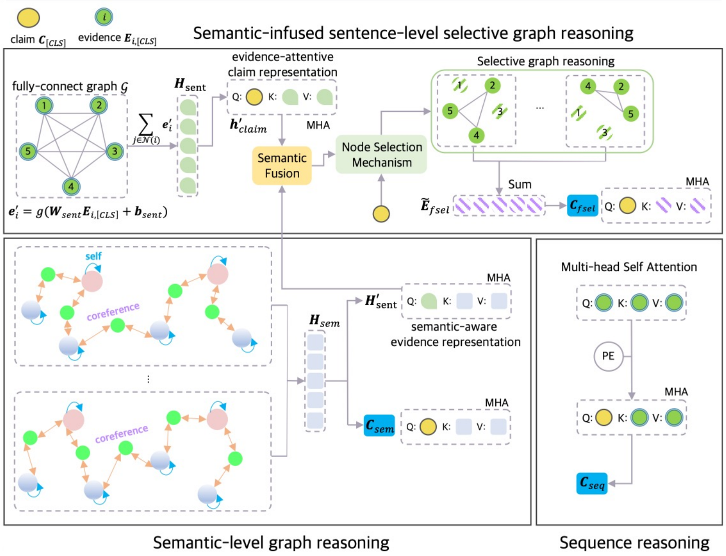
Introduction

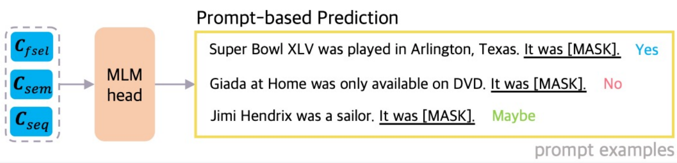
- Fact verification task aims to automatically classify a human-generated claim into "Supported", "Refuted", or "Not Enough Info" based on retrieved evidence sentences from Wikipedia.



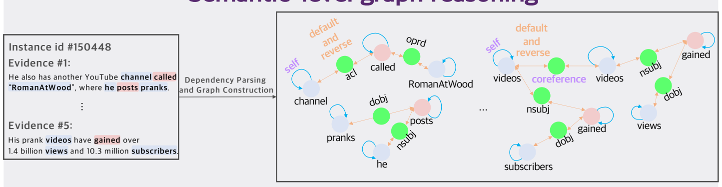
- Graph reasoning may suffer from:
  - **Unit-biased reasoning**; when relying on a single type of semantic unit for nodes of a graph, the semantic interaction between claim and evidence is restricted to a single graph type.
  - **Over-smoothing**; causing all node representations to converge to a stationary point at the extreme.

Our Approach





Semantic-level graph reasoning



Semantic-level graph reasoning employs a relational graph convolutional network defined as:

$$h_i^{(l+1)} = f\left(\sum_{r\in\mathcal{R}}\sum_{j\in\mathcal{N}_{sem}^r(i)}\frac{1}{|\mathcal{N}_{sem}^r(i)|}W_r^{(l)}h_j^{(l)} + W_0^{(l)}h_i^{(l)}\right)$$

$$H_{sem} = H^{(L)} = \left[h_1^{(L)},\cdots,h_{|\mathcal{V}_{sem}|}^{(L)}\right]$$

Semantic-infused Sentence-level Selective Graph Reasoning

- In selective graph reasoning, we prepare K different subgraphs by applying the selection mechanism K times, and combine the selective representations performed over K subgraphs.

Semantic fusion

$$\mathrm{sfu}(x,y) = g * x + (1-g) * y,\ g = \sigma(W_1 x + W_2 y)$$
$$H_{fused} = \mathrm{sfu}(H'_{claim}, H'_{sent})$$

Node Selection Mechanism

- Choosing K subsets of nodes to be selected because there is no ground-truth answer for the nodes to be selected. The node selection probabilities $p_{sent} \in \mathbb{R}^m$ described as:

$$p_{sent} = \sigma\left(g(H_{sent}W_3) + H_{fused}W_4C_{[CLS]}^T\right)$$

- The node selection mechanism creates a subsets of evidence nodes denoted $\mathcal{V}'$ by filtering out with low probabilities given the threshold τ as follows:

$$\mathcal{V}' = \{j \mid j \in \mathcal{V}\ and\ p_{sent,j} \geq \tau\}$$

- Then, we define $p'_{sent} \in \mathbb{R}^m$ by zeroing the probabilities of the filtered nodes as follows:

$$p'_{sent} = p_{sent} * i_{\mathcal{V}'}$$

where $i_{\mathcal{V}'} = [\mathcal{I}(k \in \mathcal{V}')]_{k=1}^m$ is the k-hot vector and $\mathcal{I}(e)$ is the indicator function.

Selective Graph Reasoning

- Given probabilities p'_{sent}, we perform graph reasoning using only the selected set of nodes, $\mathcal{V}'$.

$$h_i^{sel} = \sum_{j\in\mathcal{N}_{sent}(i)} p'_{sent,j}\cdot H_j^{fused}$$

Then, the reasoning-enhanced representation is obtained as follows:

$$h_i^{fsel} = \sum_{j\in\mathcal{N}_{sent}(i)} p'_{sent,j}\cdot v_j \cdot H_j^{fused}$$

where $v_j = \sigma\left(\langle w_{sel} \| [h_i^{sel}; e'_i]\rangle\right)$.

Sequence Reasoning

- Our sequence reasoning is based on MHA over only sentence-level evidence representation $E_{seq} \in \mathbb{R}^{m\times d_{model}}$, described as follows:

$$E_{seq} = PE\left(E_{1,[CLS]},\cdots,E_{m,[CLS]}\right)$$
$$H_{seq} = E_{seq} + MHA\left(E_{seq}, E_{seq}, E_{seq}\right)$$

where PE is the absolute positional encoding.

Main Results

Model	Dev LA	Dev F.S	Test LA	Test F.S
UNC NLP	69.72	66.49	68.21	64.21
GEAR (BERT-base)	74.84	70.69	71.60	67.10
DREAM (XLNet-large)	79.16	-	76.85	70.60
KGAT (BERT-large)	77.91	75.86	73.61	70.24
KGAT (RoBERTa-large)	78.29	76.11	74.07	70.38
LOREN (BERT-large)	78.44	76.21	74.43	70.71
LOREN (RoBERTa-large)	81.14	78.83	76.42	72.93
MLA (RoBERTa-large)	79.31	75.96	77.05	73.72
Ours (RoBERTa-large)	**83.13**	**79.87**	**77.50**	**73.90**