Evidence-aware Fake News Detection with Graph Neural Networks

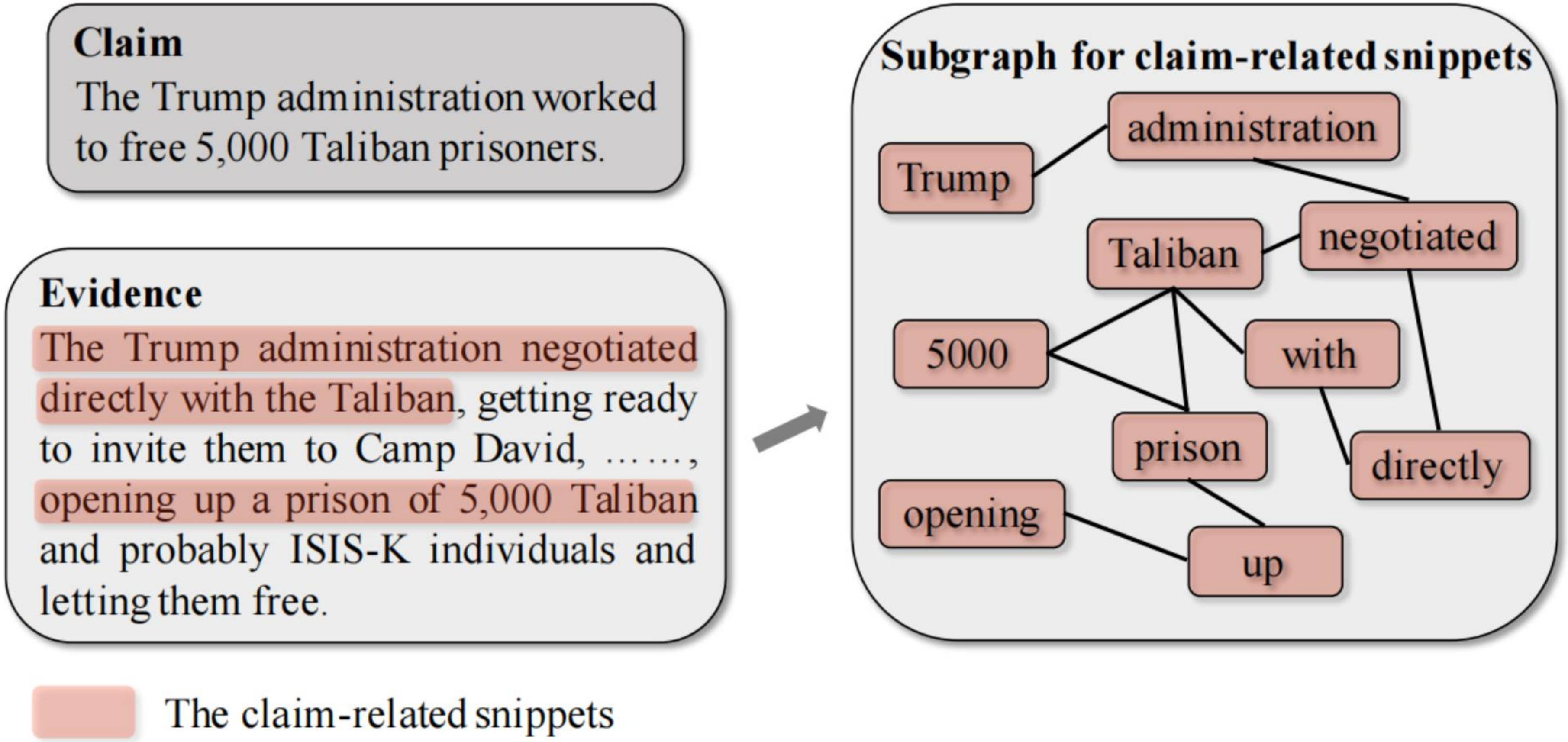
Weizhi Xu[1,2,*], Junfei Wu[3,*], Qiang Liu[1,2], Shu Wu[1,2,†], Liang Wang[1,2]
[1]CRIPAC, NLPR, Institute of Automation, Chinese Academy of Sciences
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences [3]Beijing Institute of Technology
weizhi.xu@cripac.ia.ac.cn,junfei.wu@bit.edu.cn,{qiang.liu,shu.wu,wangliang}@nlpr.ia.ac.cn

code： none

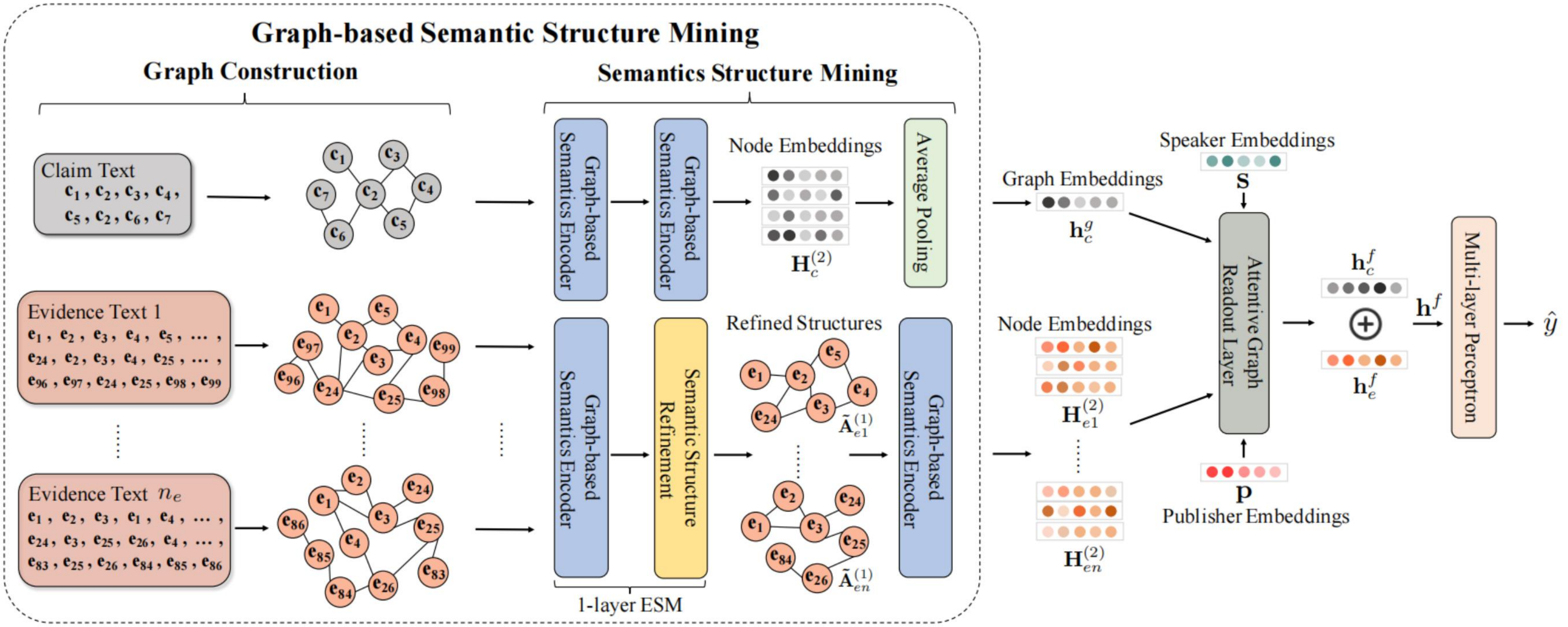
2022. 3. 13

Reported by Xiaoke Li

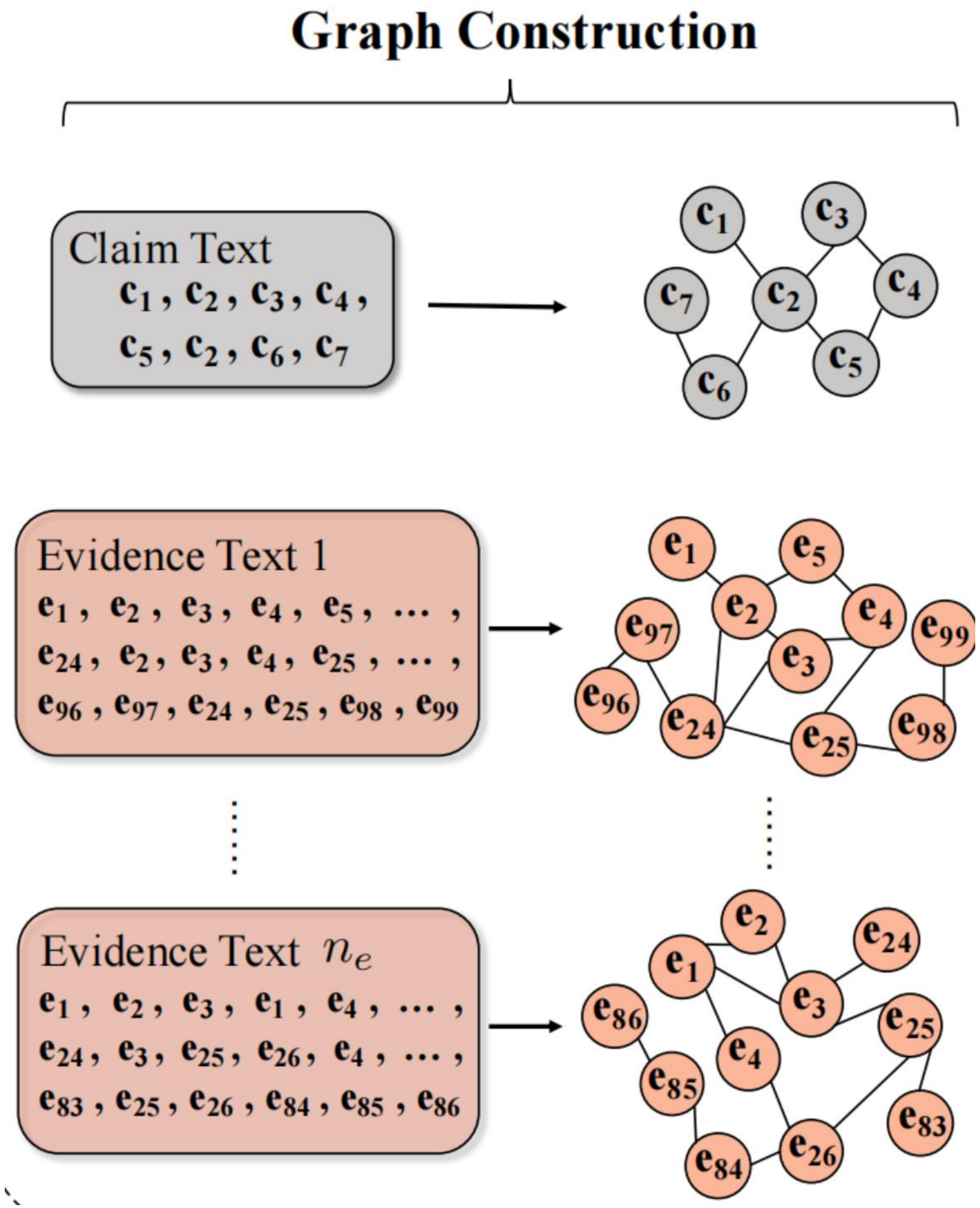
Introduction



Method

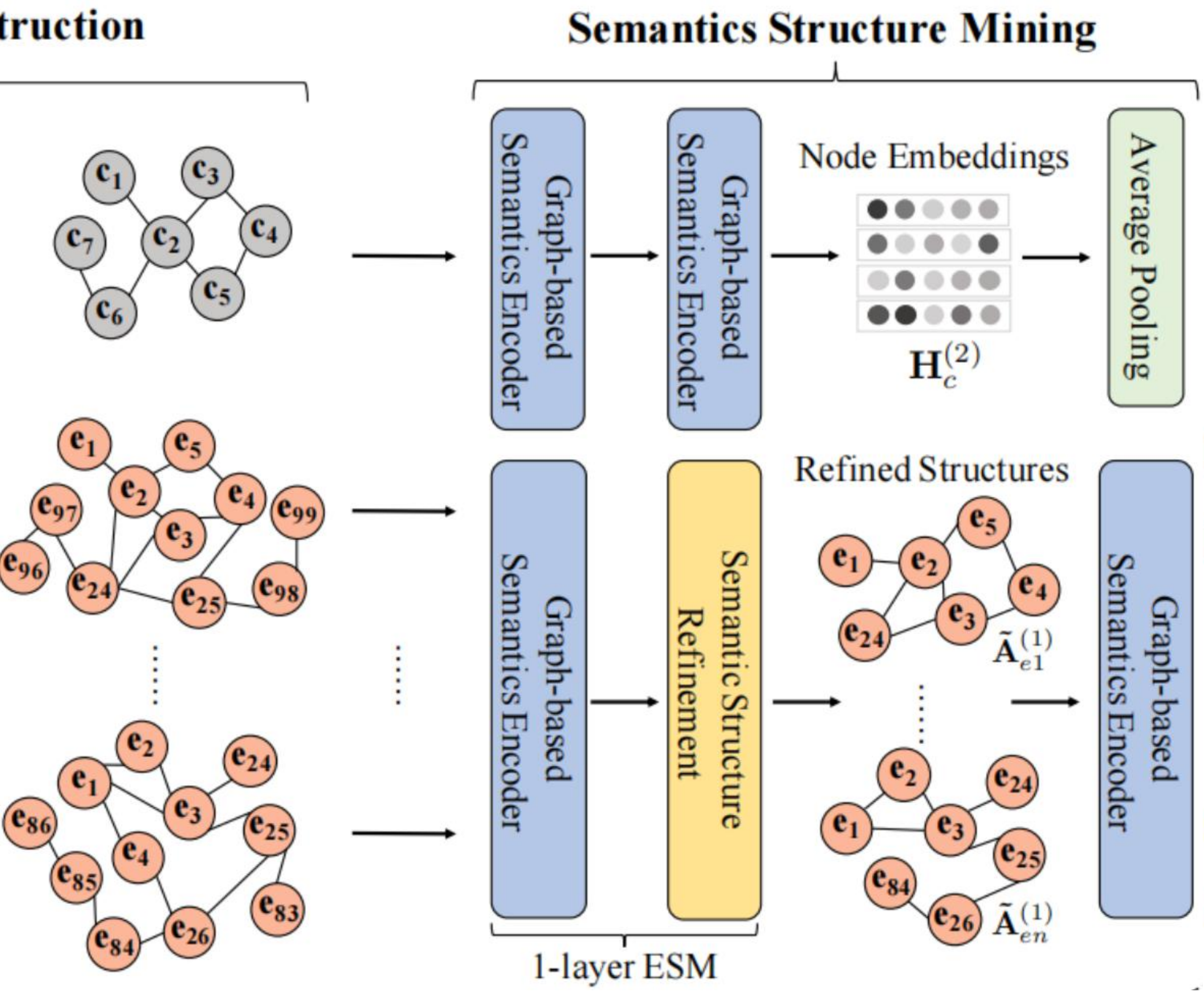

Graph-based Semantic Structure Mining
Graph Construction
Semantics Structure Mining
Claim Text
$c_1, c_2, c_3, c_4, c_5, c_2, c_6, c_7$
Evidence Text 1
$e_1, e_2, e_3, e_4, e_5, \dots, e_{24}, e_2, e_3, e_4, e_{25}, \dots, e_{96}, e_{97}, e_{24}, e_{25}, e_{98}, e_{99}$
Evidence Text n_e
$e_1, e_2, e_3, e_1, e_4, \dots, e_{24}, e_3, e_{25}, e_{26}, e_4, \dots, e_{83}, e_{25}, e_{26}, e_{84}, e_{85}, e_{86}$
Graph-based Semantics Encoder
Graph-based Semantics Encoder
Node Embeddings
$\mathbf{H}_c^{(2)}$
Average Pooling
Graph-based Semantics Encoder
Semantic Structure Refinement
Refined Structures
$\tilde{\mathbf{A}}_{e1}^{(1)}$
$\tilde{\mathbf{A}}_{en}^{(1)}$
Graph-based Semantics Encoder
1-layer ESM
Graph Embeddings
$\mathbf{h}_c^g$
Node Embeddings
$\mathbf{H}_{e1}^{(2)}$
$\mathbf{H}_{en}^{(2)}$
Speaker Embeddings
$\mathbf{s}$
Attentive Graph Readout Layer
Publisher Embeddings
$\mathbf{p}$
$\mathbf{h}_c^f$
$\mathbf{h}_e^f$
$\mathbf{h}^f$
Multi-layer Perceptron
$\hat{y}$

Method



$$\tilde{A} = D^{-\frac{1}{2}}(A + I)D^{-\frac{1}{2}}$$

$$\tilde{A}_c^{(0)} \in \mathbb{R}^{N_c \times N_c}, \tilde{A}_e^{(0)} \in \mathbb{R}^{N_e \times N_e} \text{ and } H_c^{(0)} \in \mathbb{R}^{N_c \times d}, H_e^{(0)} \in \mathbb{R}^{N_e \times d},$$



$$\mathbf{a}_i = \sum_{(w_i, w_j) \in C} \tilde{\mathbf{A}}_{ij} \mathbf{W}_a \mathbf{H}_j \quad (1)$$

$$\mathbf{z}_i = \sigma(\mathbf{W}_z \mathbf{a}_i + \mathbf{U}_z \mathbf{H}_i + \mathbf{b}_z) \quad (2)$$

$$\mathbf{r}_i = \sigma(\mathbf{W}_r \mathbf{a}_i + \mathbf{U}_r \mathbf{H}_i + \mathbf{b}_r) \quad (3)$$

$$\tilde{\mathbf{H}}_i = \tanh(\mathbf{W}_h \mathbf{a}_i + \mathbf{U}_h (\mathbf{r}_i \odot \mathbf{H}_i) + \mathbf{b}_h) \quad (4)$$

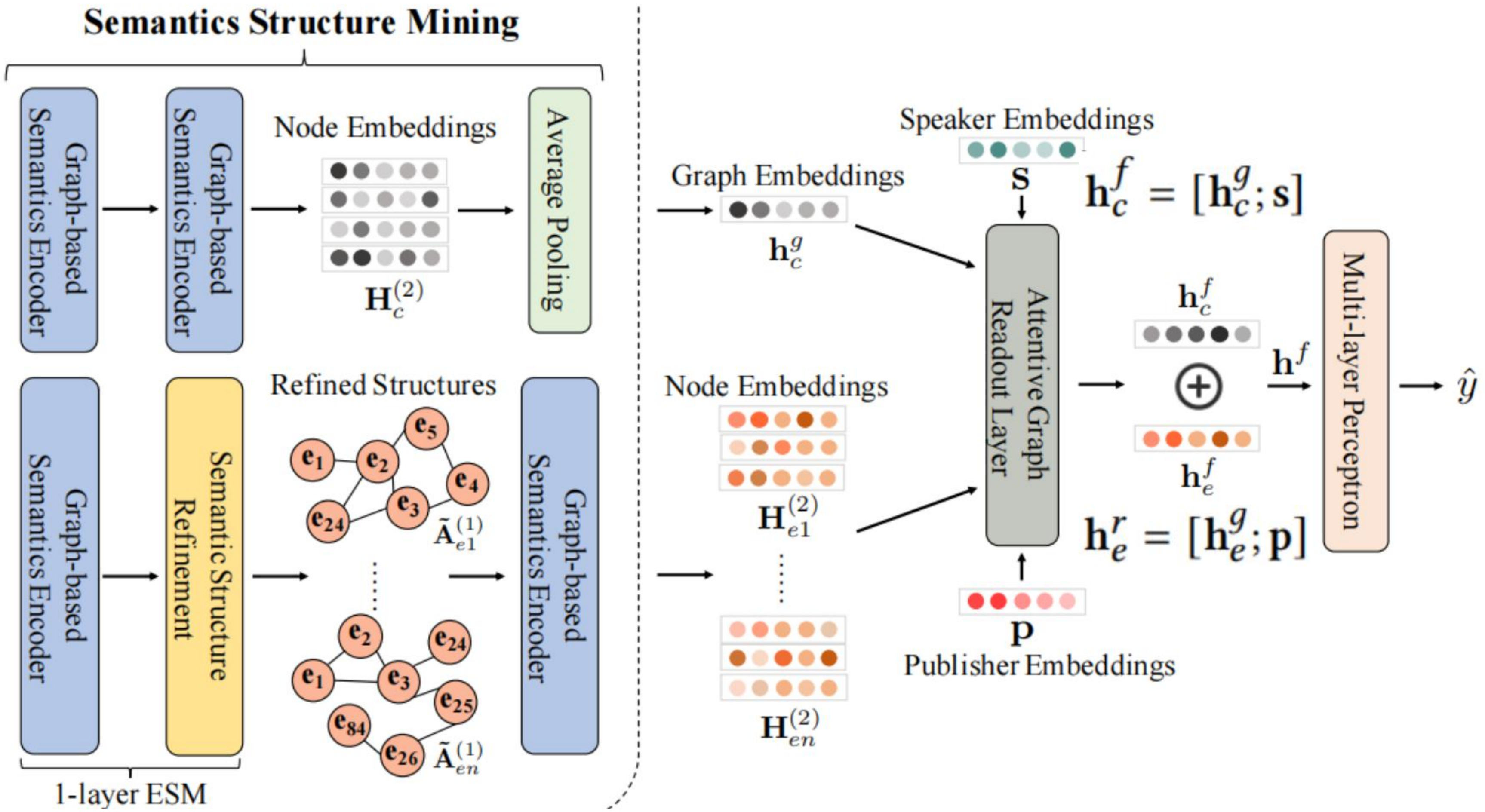
$$\hat{\mathbf{H}}_i = \tilde{\mathbf{H}}_i \odot \mathbf{z}_i + \mathbf{H}_i \odot (1 - \mathbf{z}_i) \quad (5)$$

$$\mathbf{s}_r = \mathbf{GGNN}(\tilde{\mathbf{A}}, \hat{\mathbf{H}}_\mathbf{e} \mathbf{W}_s) \quad (6)$$

$$idx = topk\_index(\mathbf{s}_r) \quad (7)$$

$$\tilde{\mathbf{A}}_{idx,:} = \tilde{\mathbf{A}}_{:,idx} = 0 \quad (8)$$

Method



$$\mathbf{h}_c^g = \frac{1}{l_c}\sum_{i=1}^{l_c}\mathbf{H}_{ci} \tag{9}$$

$$\mathbf{p}_j = \tanh\left(\left[\mathbf{H}_{ej}; \mathbf{h}_c^g\right]\mathbf{W}_c\right) \tag{10}$$

$$\alpha_j = \frac{\exp\left(\mathbf{p}_j\mathbf{W}_p\right)}{\sum_{i=1}^{l_e}\exp\left(\mathbf{p}_i\mathbf{W}_p\right)} \tag{11}$$

$$\mathbf{h}_e^g = \sum_{j=1}^{l_e}\alpha_j\mathbf{H}_{ej} \tag{12}$$

$$\mathbf{H}_e^r = [\mathbf{h}_{e1}^r; \mathbf{h}_{e2}^r; \ldots; \mathbf{h}_{en}^r] \tag{13}$$

$$\mathbf{h}_e^f = \mathbf{ATTN}(\mathbf{H}_e^r, \mathbf{h}_c^f) \tag{14}$$

$$\mathbf{h}^f = [\mathbf{h}_c^f; \mathbf{h}_e^f] \tag{15}$$

$$\hat{y} = \mathrm{Softmax}(\mathbf{W}_f\mathbf{h}^f + \mathbf{b}_f) \tag{16}$$

$$\mathcal{L}_\Theta(y, \hat{y}) = -(y\log\hat{y} + (1-y)\log(1-\hat{y})) \tag{17}$$

Experiments

Dataset	# True	# False	# Evi.	# Spe.	# Pub.
Snopes	1164	3177	29242	N/A	12236
PolitiFact	1867	1701	29556	664	4542

Table 1: The statistics of two datasets. The symbol "#" denotes "the number of". "True" and "False" stand for true claims and false claims, respectively. "Evi.', 'Spe.", and "Pub." denote evidences, speakers and publishers.

Experiments

Method	Snopes								PolitiFact							
	F1-Ma	F1-Mi	F1-T	P-T	R-T	F1-F	P-F	R-F	F1-Ma	F1-Mi	F1-T	P-T	R-T	F1-F	P-F	R-F
LSTM	0.621	0.719	0.430	0.484	0.397	0.812	0.791	0.837	0.606	0.609	0.618	0.632	0.613	0.593	0.590	0.604
TextCNN	0.631	0.720	0.450	0.482	0.430	0.812	0.799	0.826	0.604	0.607	0.615	0.630	0.610	0.592	0.591	0.604
BERT	0.621	0.716	0.431	0.477	0.407	0.810	0.793	0.830	0.597	0.598	0.608	0.619	0.599	0.586	0.577	0.597
DeClarE	0.725	0.786	0.594	0.610	0.579	0.857	0.852	0.863	0.653	0.652	0.675	0.667	0.683	0.631	0.637	0.625
HAN	0.752	0.802	0.636	0.625	0.647	0.868	0.876	0.861	0.661	0.660	0.679	0.676	0.682	0.643	0.650	**0.637**
EHIAN	0.784	0.828	0.684	0.617	0.768	0.885	0.882	0.890	0.676	0.679	0.689	0.686	0.693	0.655	0.675	0.636
MAC	0.786	0.833	0.687	0.700	0.686	0.886	0.886	0.887	0.672	0.673	0.718	0.675	0.735	0.643	0.676	0.617
CICD	0.789	0.837	0.691	0.632	**0.775**	0.893	0.890	0.895	0.682	0.685	0.702	**0.689**	0.714	0.657	0.691	0.629
GET	**0.800**‡	**0.846**‡	**0.705**‡	**0.721**‡	0.694	**0.895**‡	**0.890**	**0.902**‡	**0.691**‡	**0.694**‡	**0.723**‡	0.687	**0.764**‡	**0.660**‡	**0.708**‡	0.629

Table 2: The model comparison on two datasets Snopes and PolitiFact. "F1-Ma" and "Fi-Mi" denote the metrics F1-Macro and F1-Micro, respectively. "-T" represents "True News as Positive" and "-F" denotes "Fake news as Positive" in computing the precision and recall values. The best performance is highlighted in boldface. ‡ indicates that the performance improvement is significant with p-value ≤ 0.05.

Experiments

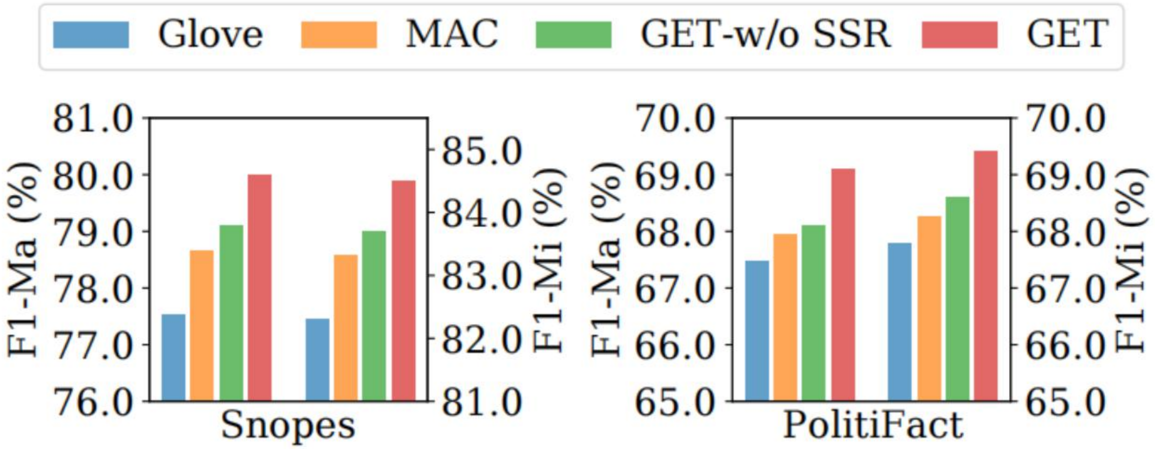


Figure 3: The performance comparison between GET and model variants with different semantic encoders (Glove and MAC) and without structure refinement (GET-w/o SSR).

Dataset	Metric	DeC	GET-DeC	EHI	GET-EHI
Snopes	F1-Ma	0.725	**0.761**	0.784	**0.795**
	F1-Mi	0.786	**0.813**	0.828	**0.841**
	F1-T	0.594	**0.649**	0.684	**0.693**
	F1-F	0.857	**0.873**	0.885	**0.897**
PolitiFact	F1-Ma	0.653	**0.681**	0.676	**0.688**
	F1-Mi	0.652	**0.685**	0.679	**0.690**
	F1-T	0.675	**0.714**	0.689	**0.713**
	F1-F	0.631	**0.647**	0.655	**0.663**

Table 3: The performance of GET with different claim-evidence interaction modules, compared to their corresponding baselines DeClarE (DeC) and EHIAN (EHI). The superior results are highlighted in boldface.

Experiments

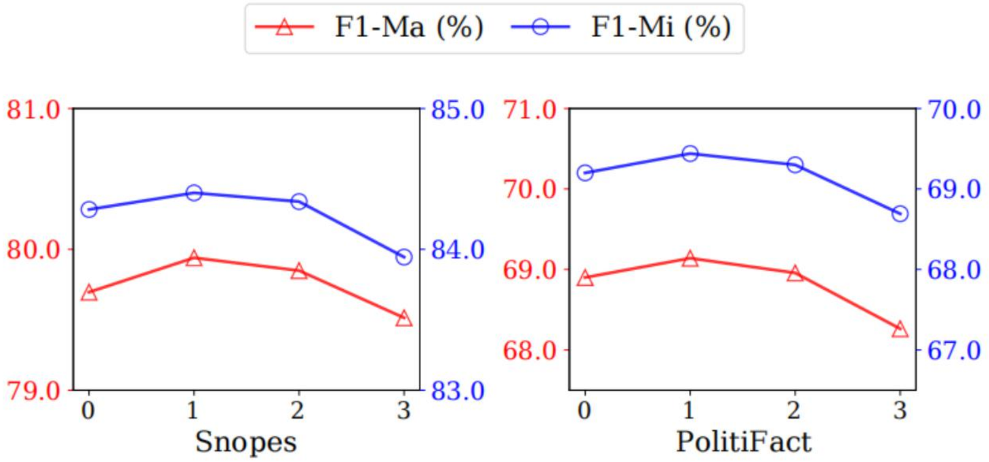


Figure 4: The influence of different semantics encoder layers T_E for claims on model performance.

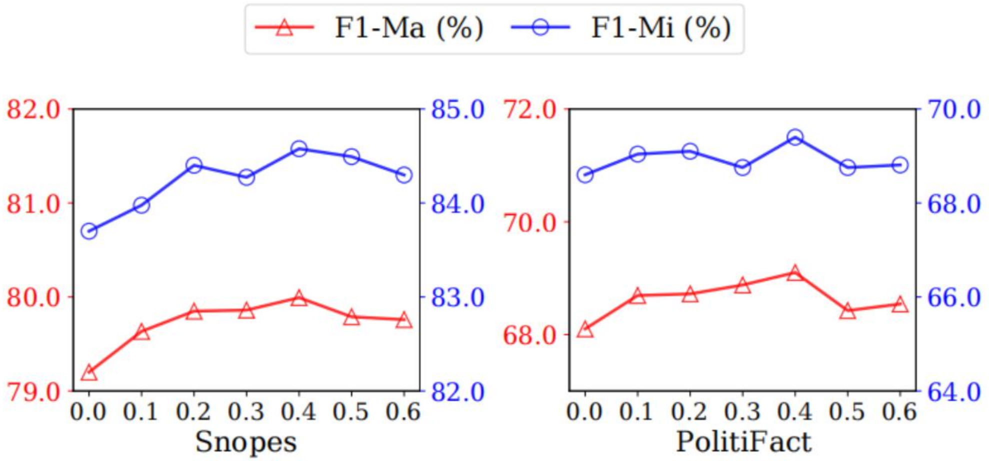


Figure 5: The influence of different discarding rates r on model performance.

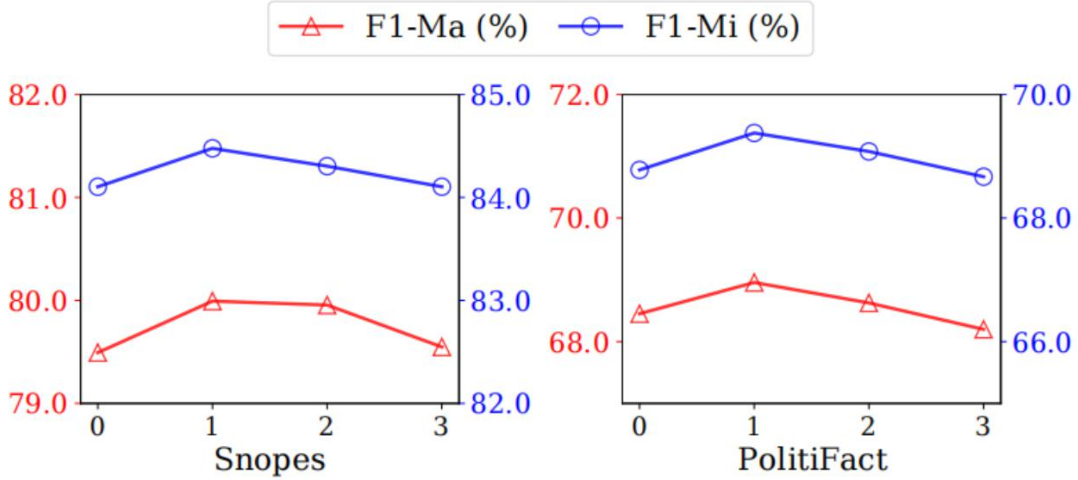


Figure 6: The influence of different evidence semantics miner layers T_R on model performance.

Experiments

Claim	*A new executive order being circulated will severely limit firearms ownership by the elderly **[False]***
Doc	…skip to content news local and beyond elderly elderly gun ban elderly gun ban claim a new executive order being circulated will severely limit firearms ownership by the elderly image image false image guns to be banned for elderly staff reports united press international washington obama deputy attorney general designate david ogden is circulating a draft of an executive order in which among other things firearms possession would be severely limited to people over 60 an assistant to ogden told us it appears that in these changing times it is no longer necessary to allow …
Claim	A chinese coal miner recently found alive in an abandoned mine 17 years *trapped inside by earthquake* ***[False]***
Doc	…images 4 a chinese coal miner was recently found alive in an abandoned mine 17 years after he had been trapped inside it by an earthquake okay i feel pretty silly because i genuinely thought this was real the first time i saw it it doesn't have the wackiness factor of the others that said it false origin …

Figure 7: Visualization of discarded words in the examples in Snopes Dataset. [True/False] indicates veracity of claims.

Thanks