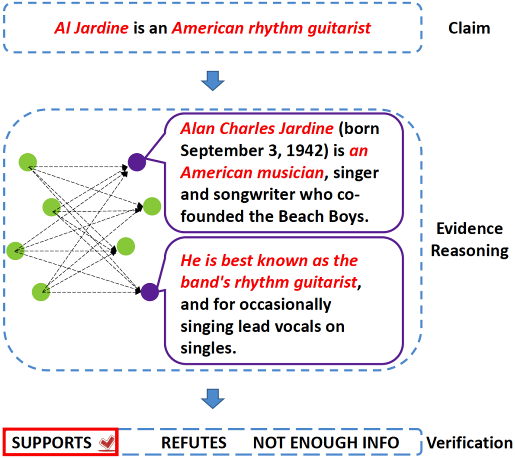
Fine-grained Fact Verification with Kernel Graph Attention Network

Zhenghao Liu, Chenyan Xiong, Maosong Sun, Zhiyuan Liu

ACL 2020

Fact Verification Task

Verify the integrity of statements using trustworthy corpora, e.g., Wikipedia



Notation

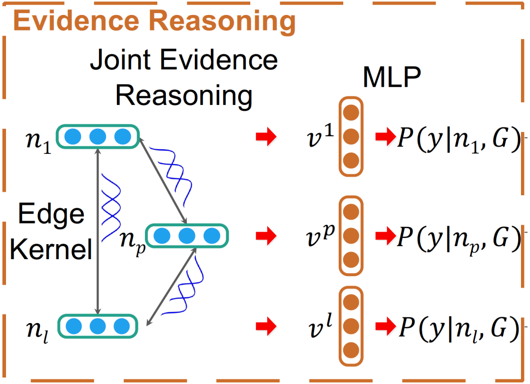
- Claim: c and claim label y
- Evidence sentences for claim c is $D = \{e^1, \cdots, e^p, \cdots, e^l\}$
- Evidence graph G where each node $n^p(e^p, c)$
- Modeling: $P(y|c, D) = \sum_{p=1}^{l} P(y|c, e^p, D)P(e^p|c, D)$
  - Label prediction in each node conditioned on the whole graph
  - Evidence selection

$$P(y|G) = \sum_{p=1}^{l} P(y|n^p, G)P(n^p|G)$$

Evidence Propagation with Edge Kernel

$$v^p = Edge - Kernel(n^p, G)$$

$$P(y|n^p, G) = softmax_y(Linear(v^p))$$



Initial Node Representation

Node tokens: [CLS] + claim + [SEP] + evidence + [SEP]

Initial node representation:

$$z_0^p = H_0^p$$

where $H^p = BERT(n^p)$

Claim representation: $H_{1:m}^p$

Claim representation: $H_{m+1:m+n}^p$

Token level attention for Node Representation

Translation matrix $M_{ij}^{q \to p} = cos(H_i^q, H_j^p)$

Kernal-based feature extraction: $K(M_i^{q \to p}) = \{K_1(M_i^{q \to p}), \cdots, K_k(M_i^{q \to p})\}$

where the Gaussian kernel (μ_k, σ_k^2)

$$K_k(M_i^{q \to p}) = \log \sum_j \exp\left(-\frac{M_{ij}^{q \to p} - \mu_k}{2\sigma_k^2}\right)$$

Attention weight $\alpha_i^{q \to p} = softmax_i(Linear(K))$

Token representation

$$\hat{z}_i^{q \to p} = \sum_{i=1}^{m+n} \alpha_i^{q \to p} H_i^q$$

Sentence Level Attention for Node Representation

Attention weight

$$\beta^{q \rightarrow p} = softmax_q(MLP(z^p \circ \hat{z}^{q \rightarrow p}))$$

Node representation

$$v^p = \left(\sum_{q=1}^{l} \beta^{q \rightarrow p} \cdot \hat{z}^{q \rightarrow p} \right) \circ z^p$$

Evidence Aggregation with Node Kernel

Calculate the evidence representation

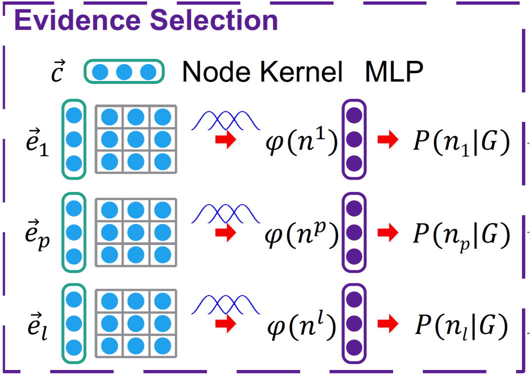
$$\phi(n^p) = Node - Kernel(n^p)$$
$$= \frac{1}{m}\sum_{i=1}^{m} K(M_i^{c \rightarrow e^p})$$

Evidence prediction

$P(n^P|G) = softmax_p(Linear(\phi(n^p)))$

Prediction:

$P(y|G) = \sum_{p=1}^{l} P(y|n^p, G)P(n^p|G)$



Result

Model	Dev		Test	
	LA	FEVER	LA	FEVER
Athene (Hanselowski et al., 2018)	68.49	64.74	65.46	61.58
UCL MRG (Yoneda et al., 2018)	69.66	65.41	67.62	62.52
UNC NLP (Nie et al., 2019a)	69.72	66.49	68.21	64.21
BERT Concat (Zhou et al., 2019)	73.67	68.89	71.01	65.64
BERT Pair (Zhou et al., 2019)	73.30	68.90	69.75	65.18
GEAR (Zhou et al., 2019)	74.84	70.69	71.60	67.10
GAT (BERT Base) w. ESIM Retrieval	75.13	71.04	72.03	67.56
KGAT (BERT Base) w. ESIM Retrieval	**75.51**	**71.61**	**72.48**	**68.16**
SR-MRS (Nie et al., 2019b)	75.12	70.18	72.56	67.26
BERT (Base) (Soleimani et al., 2019)	73.51	71.38	70.67	68.50
KGAT (BERT Base)	**78.02**	**75.88**	**72.81**	**69.40**
BERT (Large) (Soleimani et al., 2019)	74.59	72.42	71.86	69.66
KGAT (BERT Large)	**77.91**	**75.86**	**73.61**	**70.24**
KGAT (RoBERTa Large)	**78.29**	**76.11**	**74.07**	**70.38**