Capturing Sentence Relation for Answer Sentence Selection with Multi-Perspective Graph Encoding

中科院自动化所 田志兴

Task: answer sentence selection

- **Given**: a passage, and a relevant question
- **Request**: select the sentence that can answer the question

Candidates:

…

S_1: The film series was rebooted in 2013 with Man of Steel, directed by Zack Snyder with **Henry Cavill** starring as Superman.

S_2: **Cavill** is the first British and non-American actor to play **the character**.

…

S_3: Man of Steel was released in theaters on June 14, 2013

…

Question:

Which British actor played Superman in Man of Steel?

S_2: Cavil is the first British and non-American actor to play the character.

Problem: relations among the sentences

- **Previous Works**: model each candidate sentences independently, and then calculate the relevance between the candidates and the question.

- **Consequences**: ignore the relations among the candidates. Thus could not model the semantic of them comprehensively, and tend to make a decision based on the similarity between the candidates and the question.

Problem: relations among the sentences

- **Consequences**:

ignore the relations among the candidates. Thus could not model the semantic of them comprehensively, and tend to make a decision based on the similarity between the candidates and the question.

Candidates:

...

S_1: The film series was rebooted in 2013 with **Man of Steel**, directed by Zack Snyder with **Henry Cavill** starring as **Superman**.

S_2: **Cavill** is the first **British** and non-American actor to play **the character**.

...

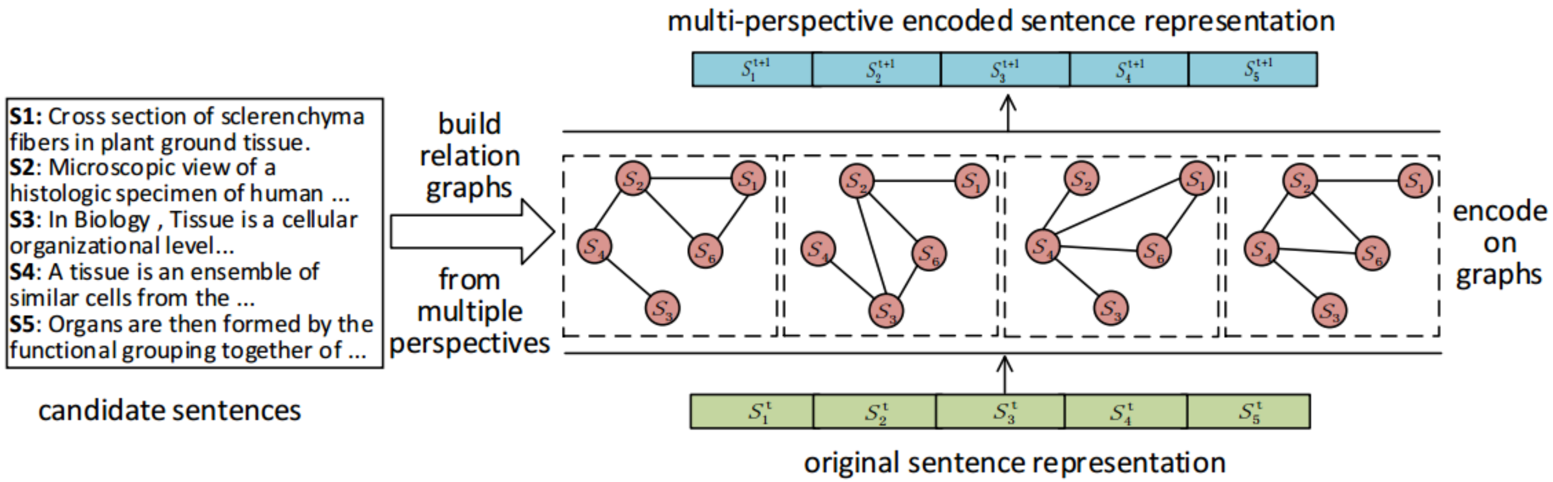
S_3: **Man of Steel** was released in theaters on June 14, 2013

...

Question:

Which **British** actor played **Superman** in **Man of Steel**?

Method: multi-perspective encoding



- **Based on** : graph, instead of sequence
- **Multi-perspective**: topic relevance, distance in passage, semantic similarity, and the others.

Method: multi-perspective graphs (static)

- **Topic relevance**: Two sentences are likely to describe the same entity and share a common topic, if there is entity co-occurrence between them.

$$A_{ij}^{ent} = \begin{cases} 1 & \text{if co-occurrence entity exists} \\ & \text{between } S_i \text{ and } S_j \\ 0 & \text{otherwise} \end{cases}$$

Method: multi-perspective graphs (static)

- **Distance in passage** : The sentences closer to each other tend to be more relevant in a passage.

$$A_{ij}^{dist} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(j-i)^2}{2\sigma^2}}$$

Method: multi-perspective graphs (static)

- **Semantic similarity** : The motivation is to build a connection between two candidates with similar semantics, and then conduct a rich information encoding for each of them.

$$A_{ij}^{simi} = \frac{r_i \cdot r_j}{\|r_i\|_2 \cdot \|r_j\|_2}$$

Method: multi-perspective graphs (dynamic)

- **Dynamic Graph**: No amount of pre-constructed graphs could coverage all of the relations between the candidate sentences. Thus the dynamic graph, as a complement, is designed to capture the instance-specific relations

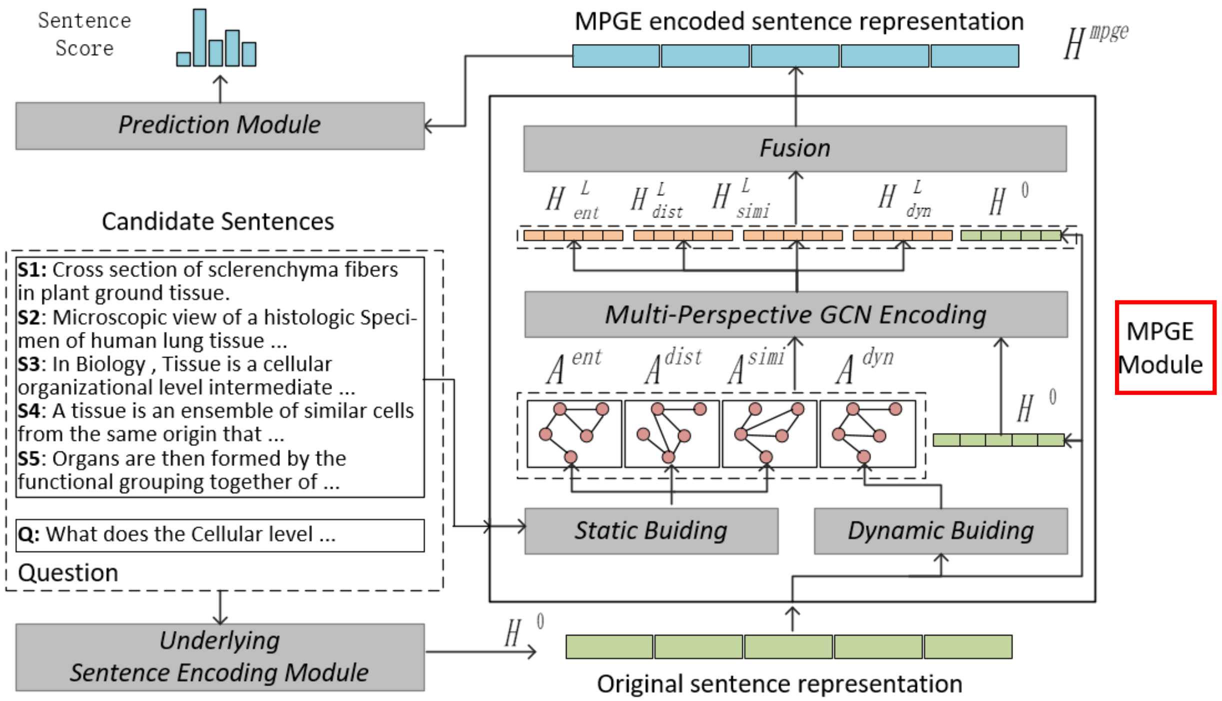
$$A_{ij}^{dyn} = \frac{exp(\alpha_{ij})}{\sum_{j'} exp(\alpha_{ij'})}$$

$$\alpha_{ij} = \sigma(w_s h_i^0)^T \sigma(w_s h_j^0)$$

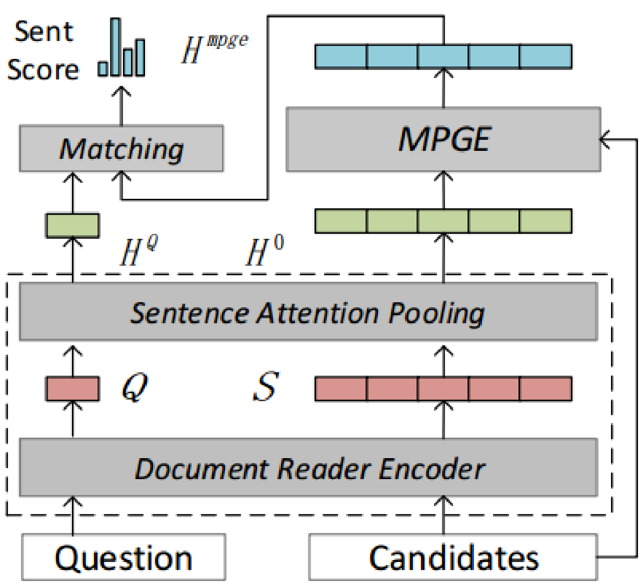
Method: detail of proposed module

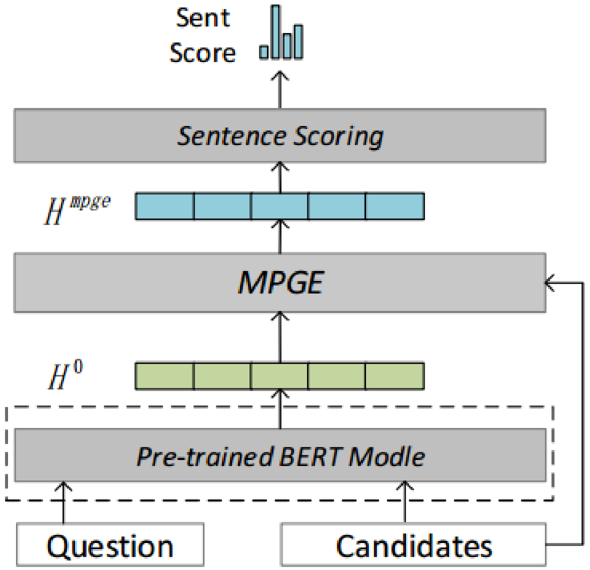
- **Static Building**:
  - topic similarity, A^{ent}
  - distance in passage, A^{dist}
  - semantic relevance, A^{simi}

- **Dynamic Building**:
  - Instance specific, A^{dyn}



Method: models based on different rep





- **Weak representation:**
  Glove+LSTM+Attention
- **Strong representation:**
  Pretrained BERT model

Experiments-effectiveness and universality

- **Achieve State of the art**
- **MPGE is effective for both kinds of representation**

Method	TOP 1	MAP
TF-IDF(Min et al. 2018)	81.2	89.0
CNN-MULT(Wang and Jiang 2017)	-	90.7
Selector(Min et al. 2018)	85.8	91.6[3]
wGRU-sGRU(Tan et al. 2018)	-	92.1
TR-AS	86.0	91.7
TR-MPGE-AS	89.0	93.2
BR-AS	89.5	93.3
BR-MPGE-AS	**92.1**	**95.0**

SQuAD

Method	MAP	MRR
AP-LSTM(Santos et al. 2016)	67.0	68.4
AP-CNN(Santos et al. 2016)	68.9	69.6
ABCNN(Yin et al. 2016)	69.2	71.0
KV-MemNN(Miller et al. 2016)	70.7	72.7
BiMPM(Wang, Hamza, and Florian 2017)	71.8	73.1
RNN-POA(Chen et al. 2017b)	72.1	73.1
Multihop(Tran and Niedereée 2018)	72.2	73.8
IARNN(Wang, Liu, and Zhao 2016)	73.4	74.1
CNN-CTK(Tymoshenko, Bonadiman, and Moschitti 2016)	74.1	75.8
CNN-MULT(Wang and Jiang 2017)	74.3	75.4
wGRU-sGRU(Tan et al. 2018)	76.3	78.2
TR-AS	72.1	73.6
TR-MPGE-AS	77.3	78.7
BR-AS	83.4	84.4
BR-MPGE-AS	**86.7**	**87.9**

WikiQA

Experiments-robust

- **Adversarial datasets**: AddSent and AddOneSent

	Original	AddOneSent	AddSent
TR-MPGE-AS	93.2	73.1	68.2
TR-AS	91.7	69.5	65.1
ΔMAP	1.5	3.6	3.1
BR-MPGE-AS	95.0	84.0	78.4
BR-AS	93.3	78.2	73.8
ΔMAP	1.7	5.8	4.6

- Also achieve better performance in adversarial datasets.

- Performance gap get wider. Thus on the adversarial examples which contain distracting sentences, the MPGE can better shows its strength.

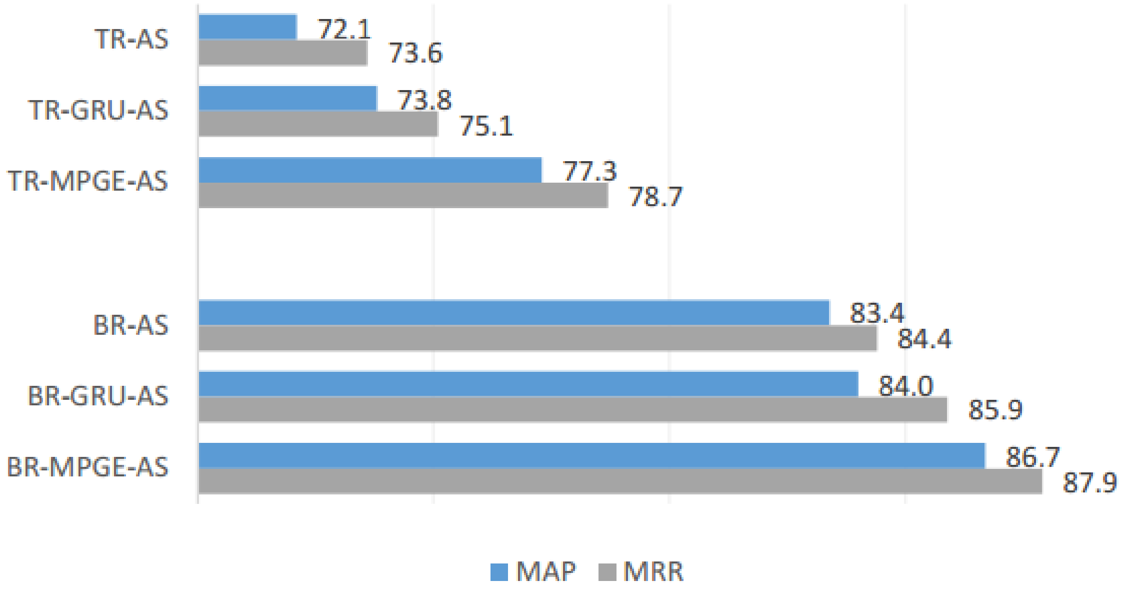
Experiments-ablation

TR Based	MAP	ΔMAP	MRR	ΔMRR	BR Based	MAP	ΔMAP	MRR	ΔMRR
TR-MPGE-AS	**77.3**	-	**78.7**	-	BR-MPGE-AS	**86.7**	-	**87.9**	-
- entity	76.3	-1.0	77.6	-1.1	- entity	86.1	-0.6	87.3	-0.6
- distance	76.6	-0.7	78.0	-0.7	- distance	85.8	-0.9	87.1	-0.8
- similarity	76.4	-0.9	77.8	-0.9	- similarity	86.2	-0.5	87.6	-0.3
- dynamic	75.4	-1.9	76.9	-1.8	- dynamic	85.7	-1.0	87.0	-0.9
TR-AS	72.1	-5.2	73.6	-5.1	BR-AS	83.4	-3.3	84.4	-3.5

- Every graph(perspective) contribute to the performance.
- Dynamic graph is important
- The effect of entity graph and similarity graph degrades when applied on BERT, while distance one keep its effect. BERT contain more semantic and topic knowledge.

Experiments-use GRU instead

- **GRU also works**: confirms the effectiveness of capturing modeling sentence relations.
- **Our MPGE is better**: multi-perspectives works better than the only perspective of sequence



Thanks