Dynamically Fused Graph Network for Multi-hop Reasoning

Yunxuan Xiao Yanru Qu   Lin Qiu

Hao Zhou   Lei Li   Weinan Zhang   Yong Yu

Shanghai Jiao Tong University   ByteDance AI Lab, China

ACL19

Repoter : Xiachong Feng

Outline

- Author
- Background
- Task and Challenge
- Motivation
- Model
- Experiments
- Conclusion

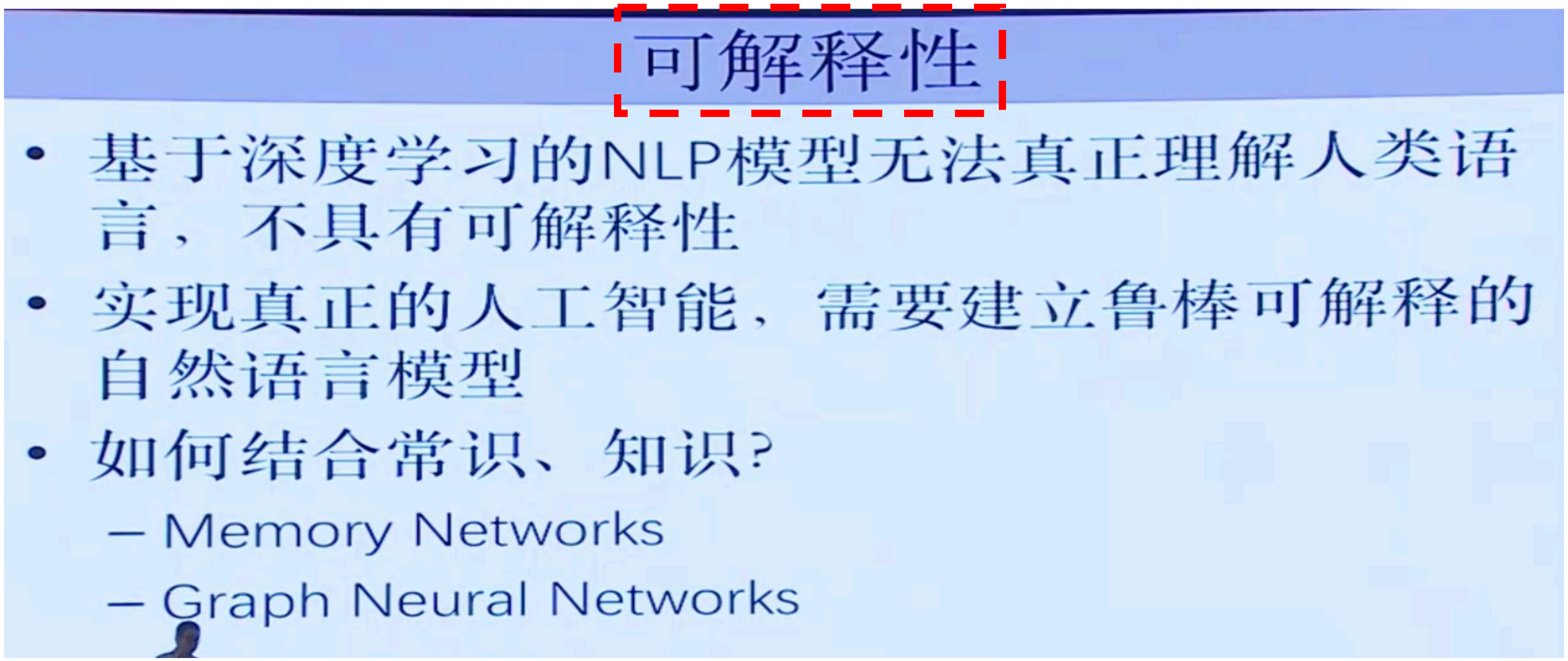
Author

Yunxuan Xiao(肖云轩)
Junior undergraduate

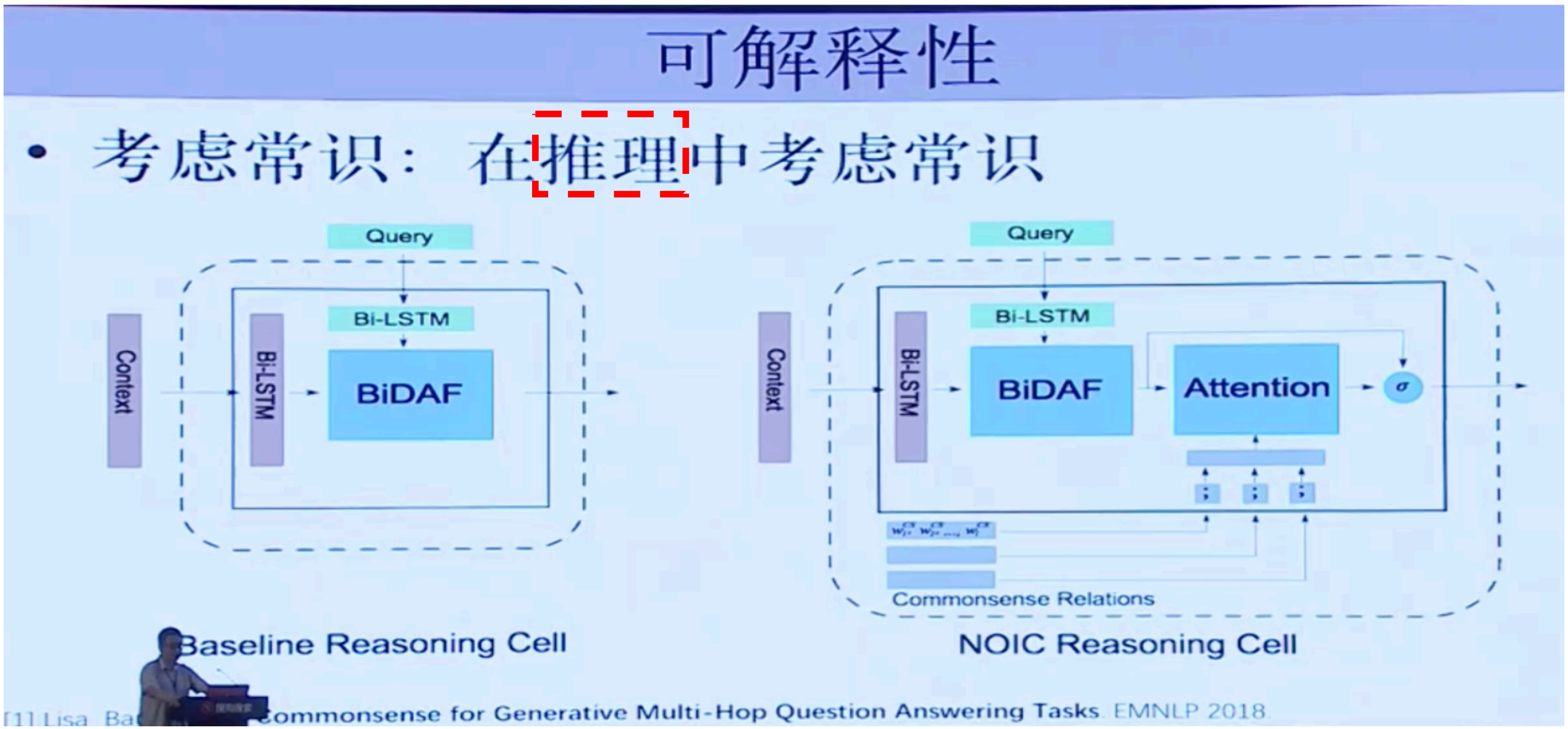
Yanru Qu
University of Illinois,
Urbana-Champaign
fall 2019

Lin Qiu

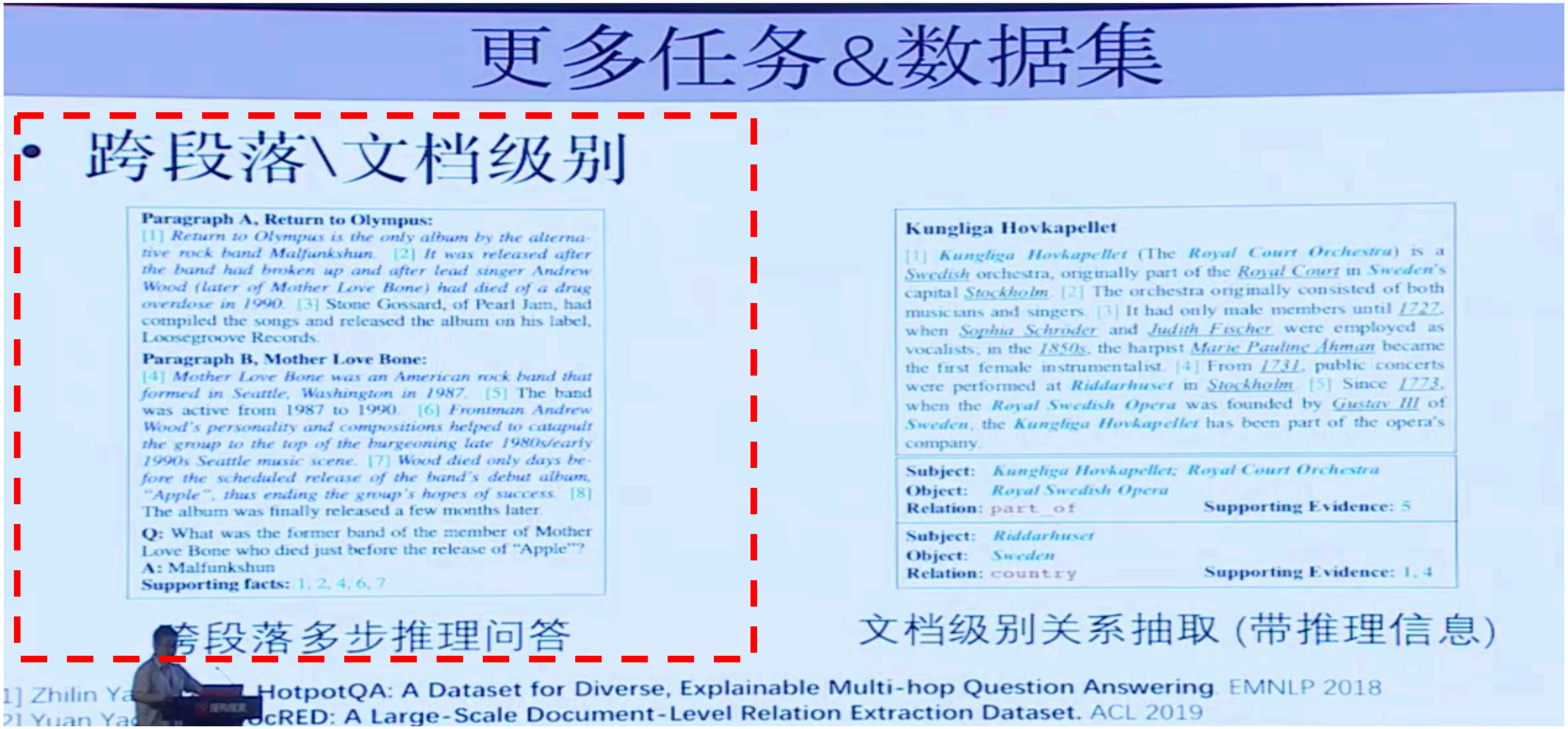
Background


可解释性
• 基于深度学习的NLP模型无法真正理解人类语言，不具有可解释性
• 实现真正的人工智能，需要建立鲁棒可解释的自然语言模型
• 如何结合常识、知识?
– Memory Networks
– Graph Neural Networks

Background


可解释性
• 考虑常识：在推理中考虑常识
Query
Bi-LSTM
Context
Bi-LSTM
BiDAF
Query
Bi-LSTM
Context
Bi-LSTM
BiDAF
Attention
σ
Commonsense Relations
Baseline Reasoning Cell
NOIC Reasoning Cell
[1] Lisa Ba
ommonsense for Generative Multi-Hop Question Answering Tasks. EMNLP 2018

Background


更多任务&数据集
• 跨段落\文档级别
Paragraph A, Return to Olympus:
[1] Return to Olympus is the only album by the alternative rock band Malfunkshun. [2] It was released after the band had broken up and after lead singer Andrew Wood (later of Mother Love Bone) had died of a drug overdose in 1990. [3] Stone Gossard, of Pearl Jam, had compiled the songs and released the album on his label, Loosegroove Records.
Paragraph B, Mother Love Bone:
[4] Mother Love Bone was an American rock band that formed in Seattle, Washington in 1987. [5] The band was active from 1987 to 1990. [6] Frontman Andrew Wood's personality and compositions helped to catapult the group to the top of the burgeoning late 1980s/early 1990s Seattle music scene. [7] Wood died only days before the scheduled release of the band's debut album, "Apple", thus ending the group's hopes of success. [8] The album was finally released a few months later.
Q: What was the former band of the member of Mother Love Bone who died just before the release of "Apple"?
A: Malfunkshun
Supporting facts: 1, 2, 4, 6, 7
Kungliga Hovkapellet
[1] Kungliga Hovkapellet (The Royal Court Orchestra) is a Swedish orchestra, originally part of the Royal Court in Sweden's capital Stockholm. [2] The orchestra originally consisted of both musicians and singers. [3] It had only male members until 1727, when Sophia Schröder and Judith Fischer were employed as vocalists; in the 1850s, the harpist Marie Pauline Åhman became the first female instrumentalist. [4] From 1731, public concerts were performed at Riddarhuset in Stockholm. [5] Since 1773, when the Royal Swedish Opera was founded by Gustav III of Sweden, the Kungliga Hovkapellet has been part of the opera's company.
Subject: Kungliga Hovkapellet; Royal Court Orchestra
Object: Royal Swedish Opera
Relation: part_of
Supporting Evidence: 5
Subject: Riddarhuset
Object: Sweden
Relation: country
Supporting Evidence: 1, 4
跨段落多步推理问答
文档级别关系抽取（带推理信息）
HotpotQA: A Dataset for Diverse, Explainable Multi-hop Question Answering. EMNLP 2018
cRED: A Large-Scale Document-Level Relation Extraction Dataset. ACL 2019

Question Answering

Sentence having the right answer

```
'context': 'Beyoncé Giselle Knowles-Carter (/biːˈjɒnseɪ/ bee-YON-say) (bor
n September 4, 1981) is an American singer, songwriter, record producer an
d actress. Born and raised in Houston, Texas, she performed in various sin
ging and dancing competitions as a child,and rose to fame in the late 1990
s as lead singer of R&B girl-group Destiny\'s Child. Managed by her father
, Mathew Knowles, the group became one of the world\'s best-selling girl g
roups of all time. Their hiatus saw the release of Beyoncé\'s debut album,
Dangerously in Love (2003), which established her as a solo artist worldwi
de, earned five Grammy Awards and featured the Billboard Hot 100 number-on
e singles "Crazy in Love" and "Baby Boy".',
'text': 'in the late 1990s’
'question': 'When did Beyonce start becoming popular?'
```

Exact Answer

Question Answering

- **Question Answering**
  - Knowledge-based (KBQA)
  - **Text-based (TBQA)**
  - Mixed
  - Others

- **KBQA** : supporting information is from **structured knowledge bases (KBs)**
- **TBQA** : supporting information is **raw text**
  - SQuAD
  - **HotpotQA**

Multi-Hop QA

Input Paragraphs:

The Sum of All Fears is a best-selling thriller novel by Tom Clancy ... It was the fourth of Clancy's Jack Ryan books to be turned into a film ...

Dr. John Patrick Jack Ryan Sr., KCVO (Hon.), Ph.D. is a fictional character created by Tom Clancy who appears in many of his novels and their respective film adaptations ...

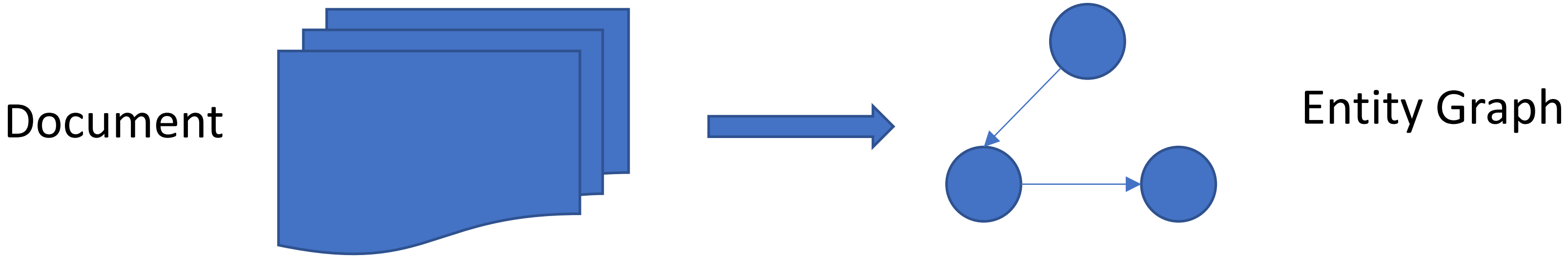
Net Force Explorers is a series of young adult novels created by Tom Clancy and Steve Pieczenik as a spin-off of the military fiction series ...

Question: What fictional character created by Tom Clancy was turned into a film in 2002?

Answer: Jack Ryan

Challenge

1. Filtering out noises from multiple paragraphs and extracting useful information.
2. Previous work on multi-hop QA aggregates document information to an **entity graph**, and **answers** are then directly **selected on entities of the entity graph**. However, in a more realistic setting, **the answers may even not reside in entities of the extracted entity graph.**



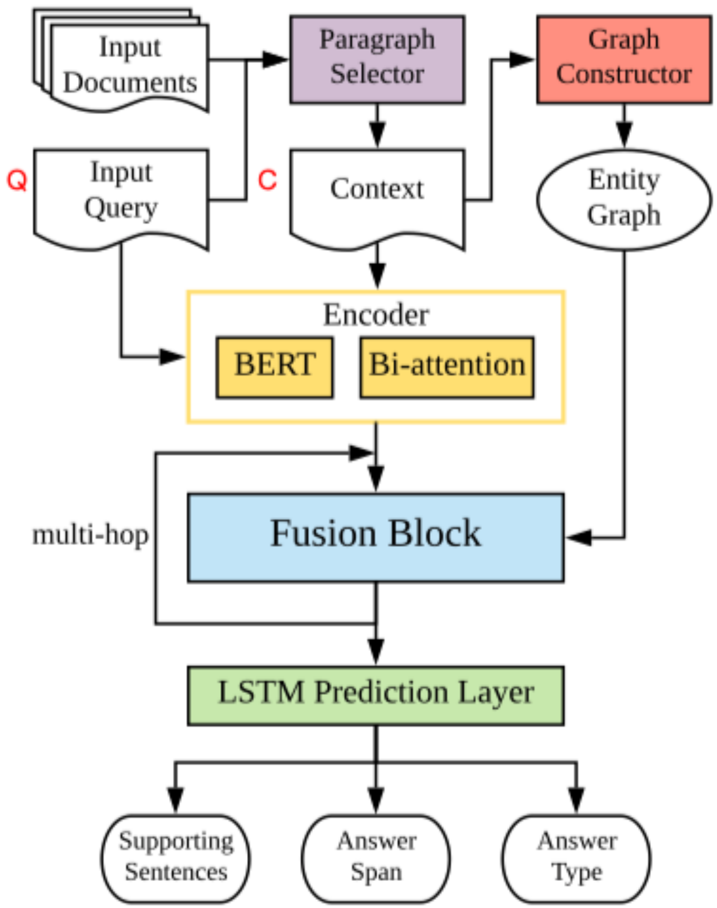
Motivation

Human's step-by-step reasoning behavior

1. One starts from an entity of interest in the query
2. Focuses on the words surrounding the start entities.
3. Connects to some related entity either found in the neighborhood or linked by the same surface mention.
4. Repeats the step to form a reasoning chain.
5. Lands on some entity or snippets likely to be the answer.

Model

- Dynamically Fused Graph Network



- Paragraph selection sub-network
- Module for entity graph construction
- Encoding layer
- Fusion block for multi-hop reasoning
- Final prediction layer

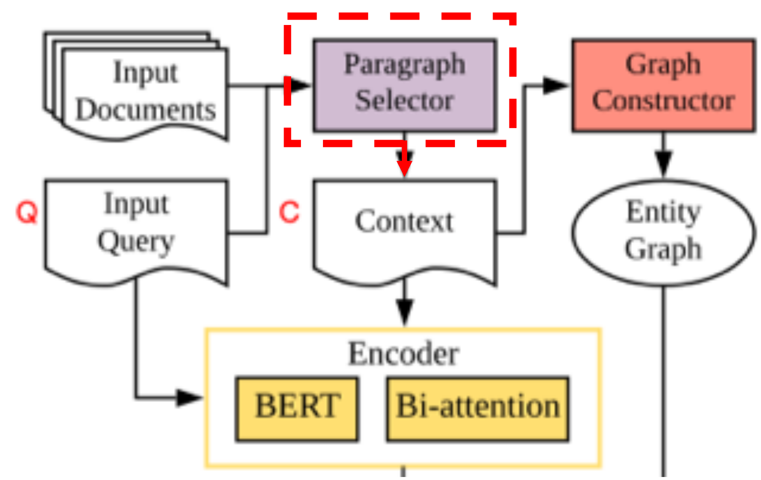
Paragraph Selection

- 1 question → N_p paragraphs
- **Model:** Pre-trained BERT followed by a sentence classification layer with sigmoid prediction (> 0.1)
- **Label**: least one supporting sentence

Paragraph A, Return to Olympus:
[1] *Return to Olympus is the only album by the alternative rock band Malfunkshun.* [2] *It was released after the band had broken up and after lead singer Andrew Wood (later of Mother Love Bone) had died of a drug overdose in 1990.* [3] Stone Gossard, of Pearl Jam, had compiled the songs and released the album on his label, Loosegroove Records.

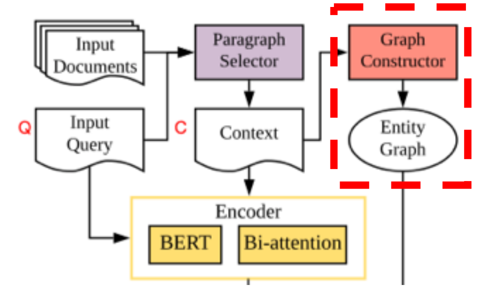
Paragraph B, Mother Love Bone:
[4] *Mother Love Bone was an American rock band that formed in Seattle, Washington in 1987.* [5] The band was active from 1987 to 1990. [6] *Frontman Andrew Wood's personality and compositions helped to catapult the group to the top of the burgeoning late 1980s/early 1990s Seattle music scene.* [7] *Wood died only days before the scheduled release of the band's debut album, "Apple", thus ending the group's hopes of success.* [8] The album was finally released a few months later.

Q: What was the former band of the member of Mother Love Bone who died just before the release of "Apple"?
A: Malfunkshun
Supporting facts: 1, 2, 4, 6, 7

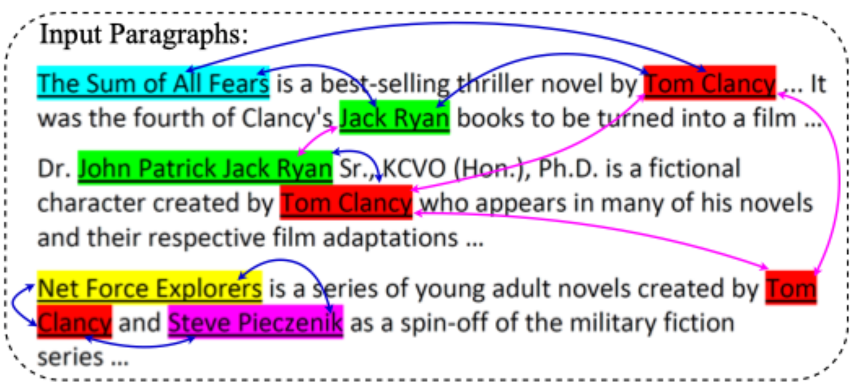


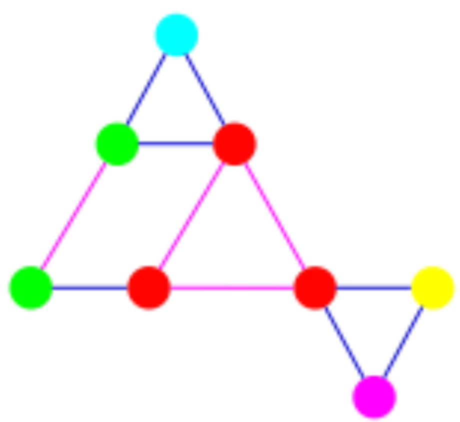
concatenated together as the context ***C***

Constructing Entity Graph



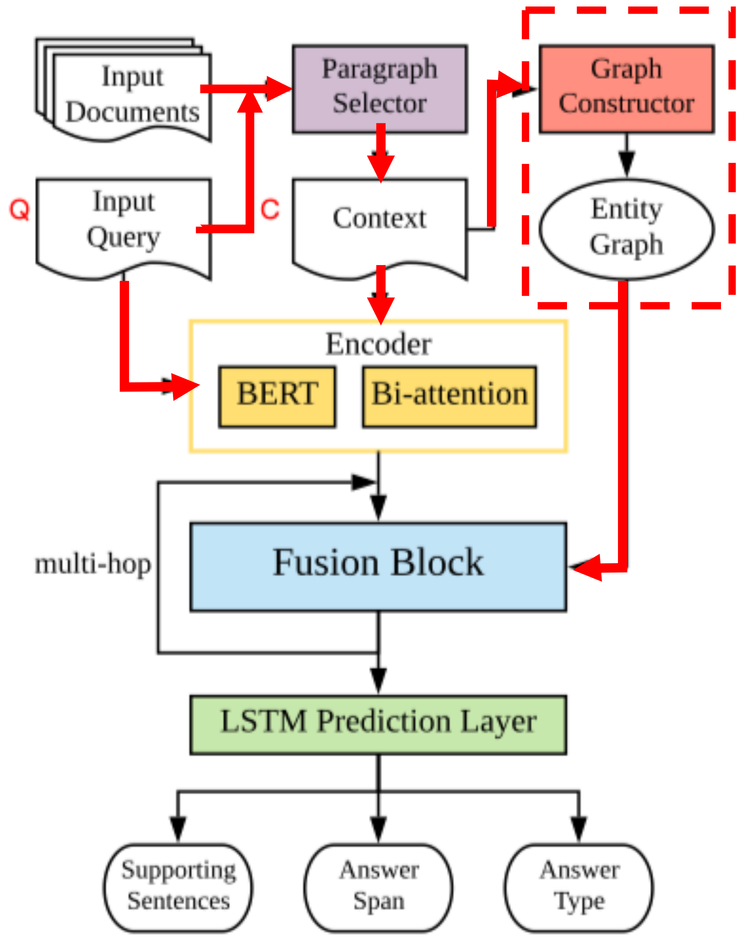
- **Nodes: NER(**Person, Organization, and Location)
- **Edge**
  1. For every pair of entities appear in the same sentence in C
  2. For every pair of entities with the same mention text in C
  3. Between a **central entity** node and other entities within the same paragraph





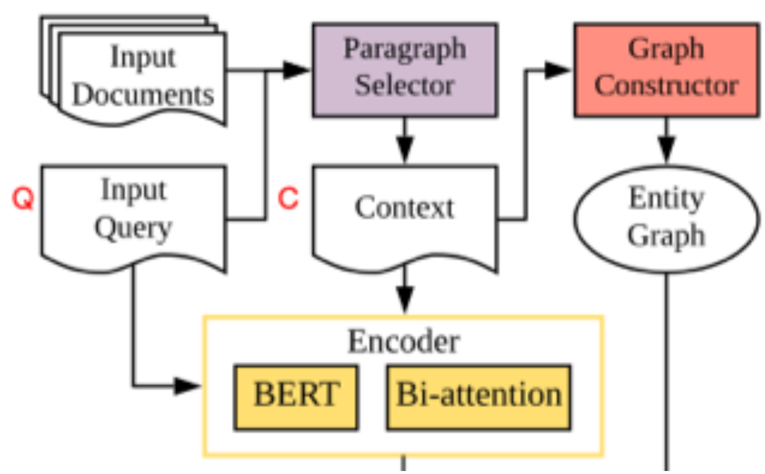
Model

- Dynamically Fused Graph Network



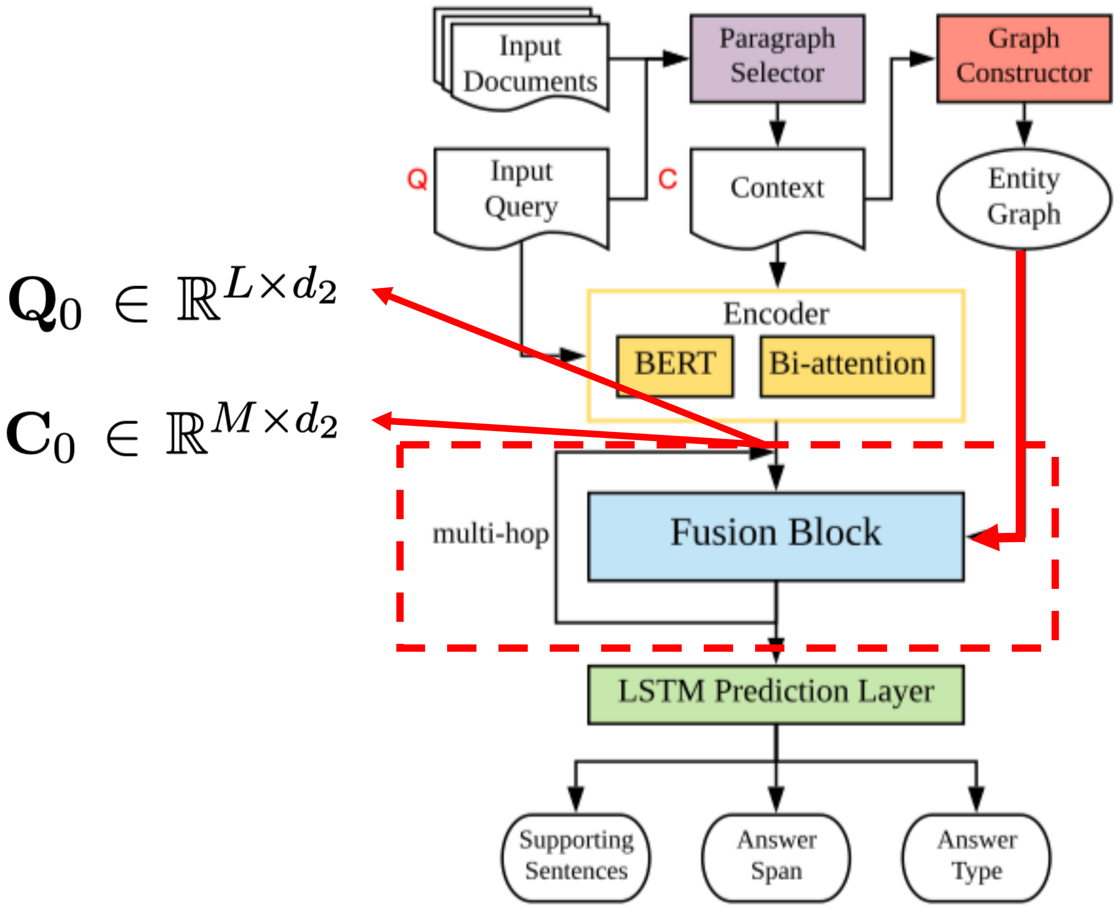
Encoding Query and Context

- Concatenate the **query Q** with the **context C**
- Pass the resulting sequence to a **pre-trained BERT model**
- The representations are further passed through a **bi-attention layer**

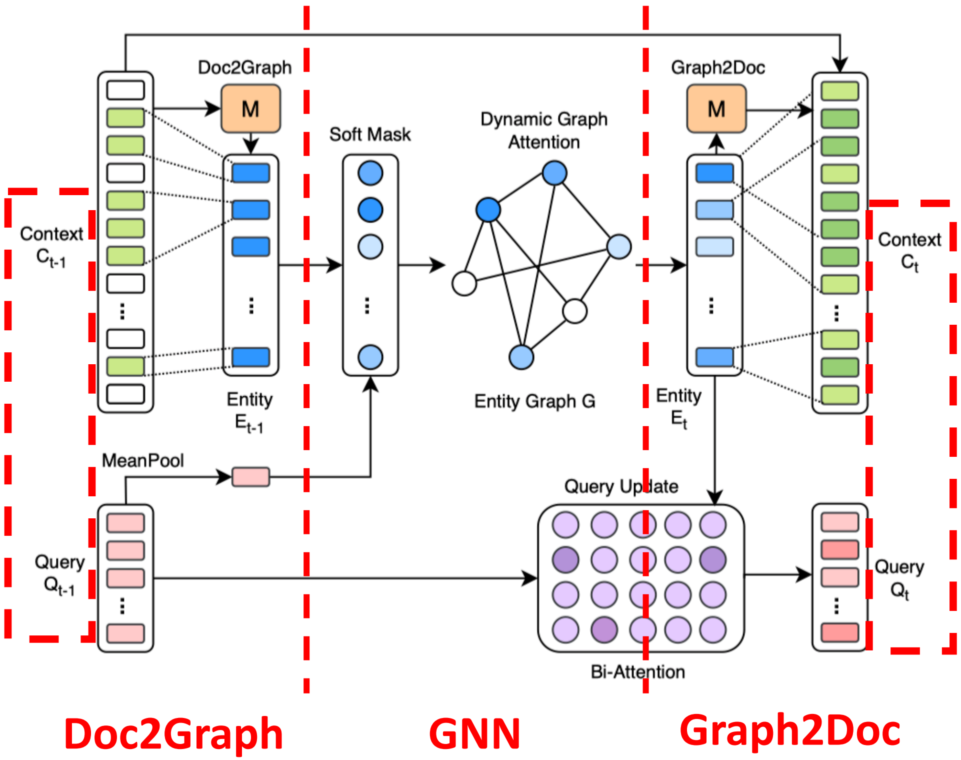


$$\mathbf{Q}_0 \in \mathbb{R}^{L \times d_2} \qquad \mathbf{C}_0 \in \mathbb{R}^{M \times d_2}$$

Reasoning with the Fusion Block

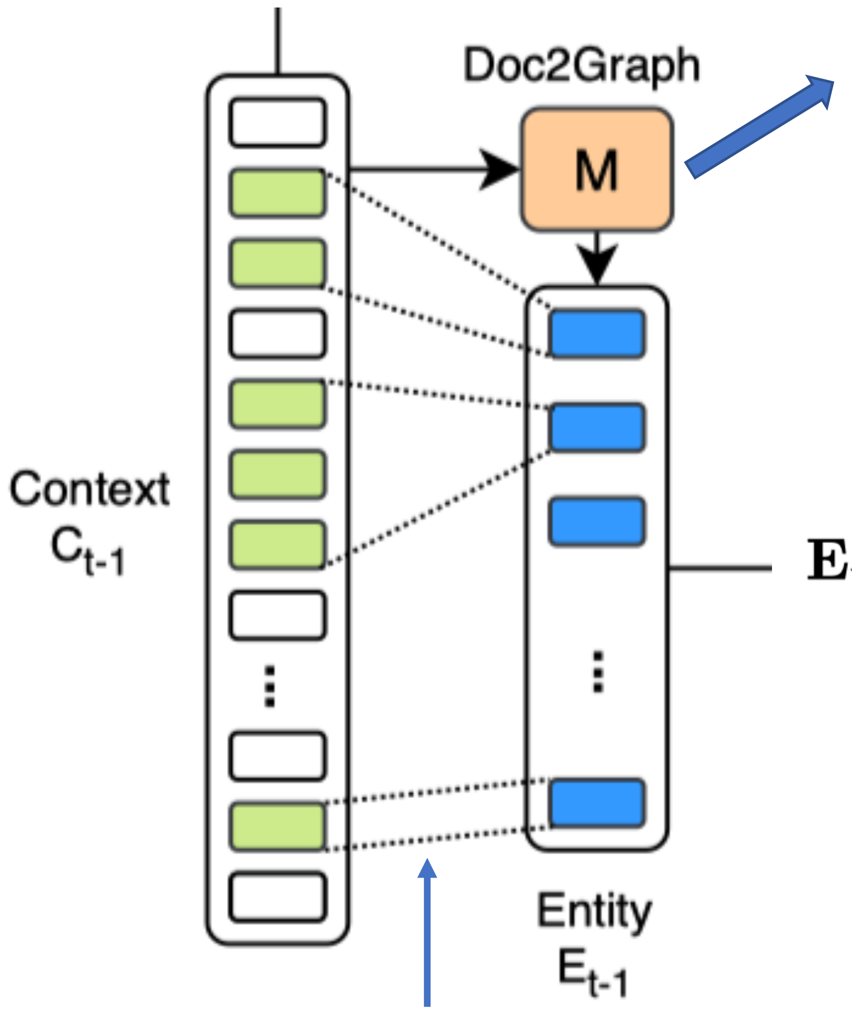


Reasoning with the Fusion Block



1. Passing information from tokens to entities by computing entity embeddings from tokens **(Doc2Graph flow)**;
2. Propagating information on entity graph; **(GNN)**
3. Passing information from entity graph to document tokens since the final prediction is on tokens **(Graph2Doc flow).**

Document to Graph Flow

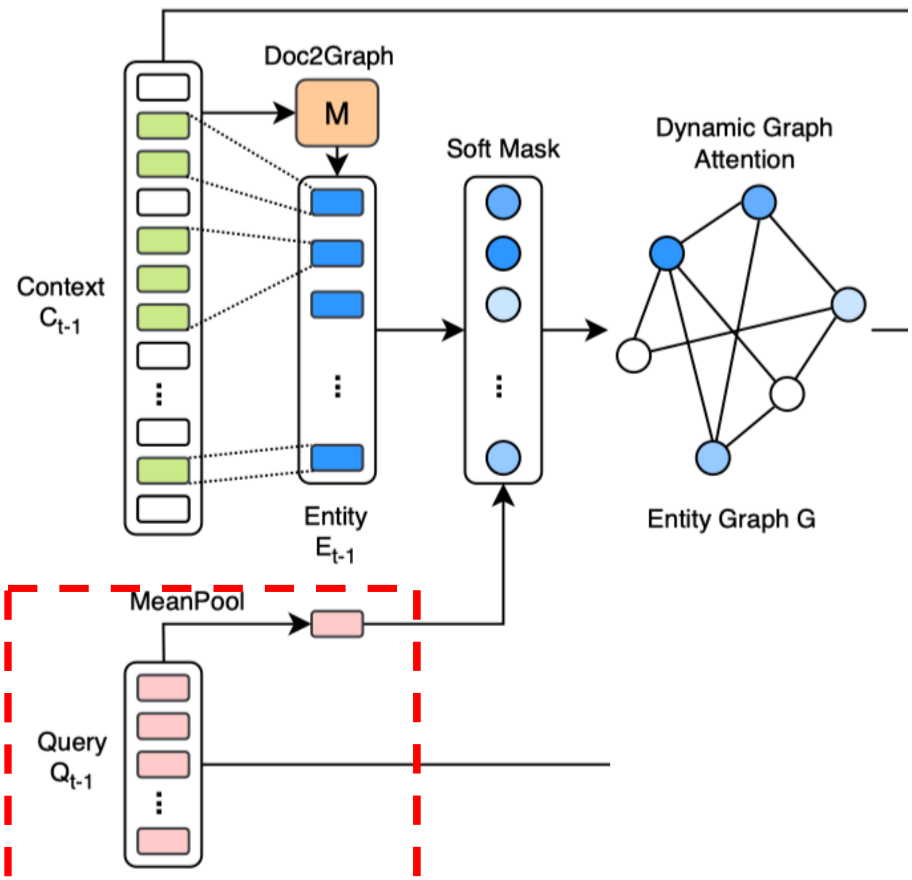


	E1	E2	E3
w1			
w2	1		
w3	1		
w4			
w5			

$$\mathbf{E}_{t-1} = [\mathbf{e}_{t-1,1}, \ldots, \mathbf{e}_{t-1,N}]$$

Mean-max pooling

Dynamic Graph Attention



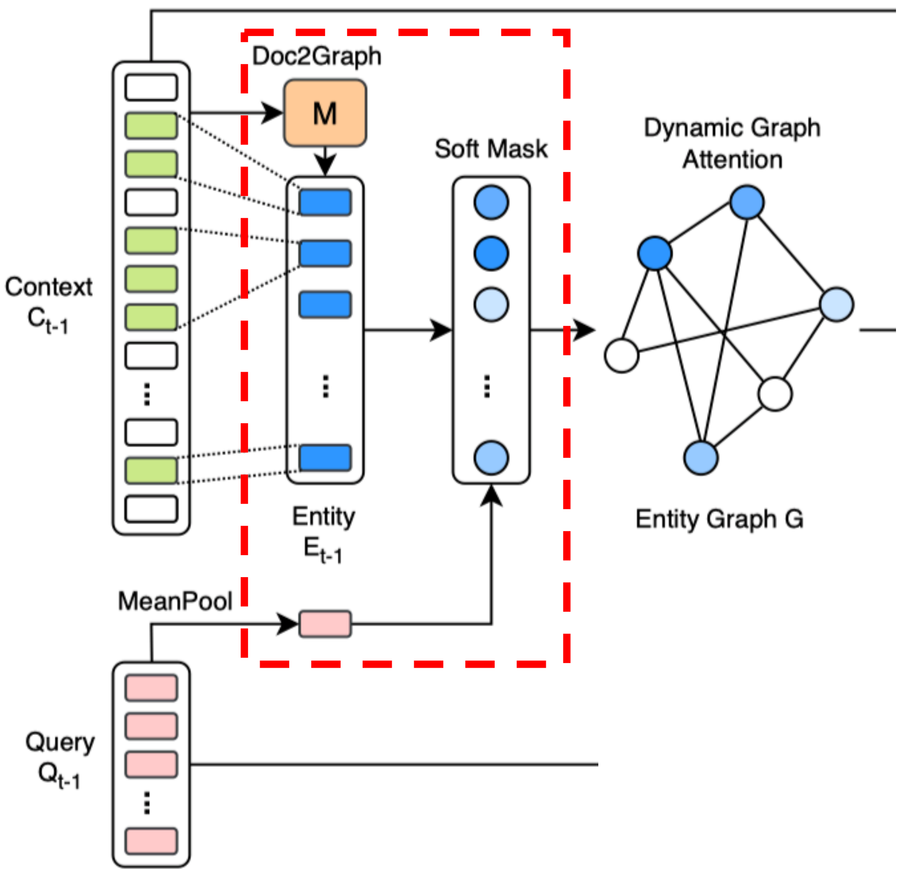
$$\boxed{\tilde{\mathbf{q}}^{(t-1)} = \mathrm{MeanPooling}(\mathbf{Q}^{(t-1)})}$$

$$\gamma_i^{(t)} = \tilde{\mathbf{q}}^{(t-1)}\mathbf{V}^{(t)}\mathbf{e}_i^{(t-1)}/\sqrt{d_2}$$

$$\mathbf{m}^{(t)} = \sigma([\gamma_1^{(t)}, \cdots, \gamma_N^{(t)}])$$

$$\tilde{\mathbf{E}}^{(t-1)} = [m_1^{(t)}\mathbf{e}_1^{(t-1)}, \ldots, m_N^{(t)}\mathbf{e}_N^{(t-1)}]$$

Dynamic Graph Attention

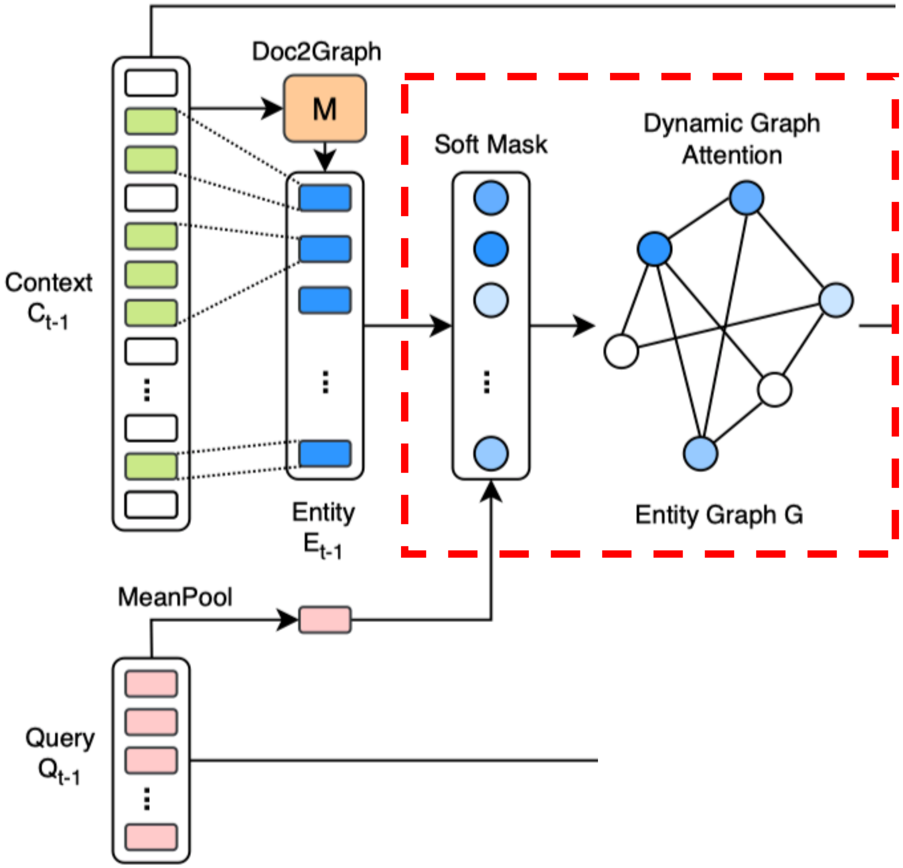


$$\tilde{\mathbf{q}}^{(t-1)} = \text{MeanPooling}(\mathbf{Q}^{(t-1)})$$

$$\boxed{\begin{aligned}\gamma_i^{(t)} &= \tilde{\mathbf{q}}^{(t-1)}\mathbf{V}^{(t)}\mathbf{e}_i^{(t-1)}/\sqrt{d_2} \\ \mathbf{m}^{(t)} &= \sigma([\gamma_1^{(t)}, \cdots, \gamma_N^{(t)}])\end{aligned}}$$

$$\tilde{\mathbf{E}}^{(t-1)} = [m_1^{(t)}\mathbf{e}_1^{(t-1)}, \ldots, m_N^{(t)}\mathbf{e}_N^{(t-1)}]$$

Dynamic Graph Attention



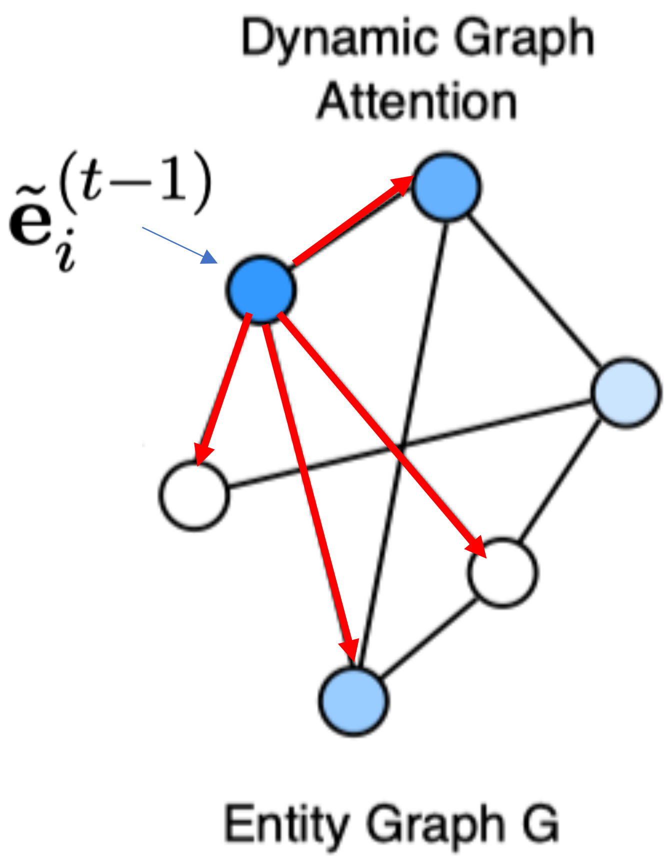
$$\tilde{\mathbf{q}}^{(t-1)} = \mathrm{MeanPooling}(\mathbf{Q}^{(t-1)})$$

$$\gamma_i^{(t)} = \tilde{\mathbf{q}}^{(t-1)}\mathbf{V}^{(t)}\mathbf{e}_i^{(t-1)}/\sqrt{d_2}$$

$$\mathbf{m}^{(t)} = \sigma([\gamma_1^{(t)}, \cdots, \gamma_N^{(t)}])$$

$$\boxed{\tilde{\mathbf{E}}^{(t-1)} = [m_1^{(t)}\mathbf{e}_1^{(t-1)}, \ldots, m_N^{(t)}\mathbf{e}_N^{(t-1)}]}$$

Dynamic Graph Attention



GAT

$$\mathbf{h}_i^{(t)} = \mathbf{U}_t \tilde{\mathbf{e}}_i^{(t-1)} + \mathbf{b}_t$$

$$\beta_{i,j}^{(t)} = \mathrm{LeakyReLU}(\mathbf{W}_t^\top [\mathbf{h}_i^{(t)}, \mathbf{h}_j^{(t)}])$$

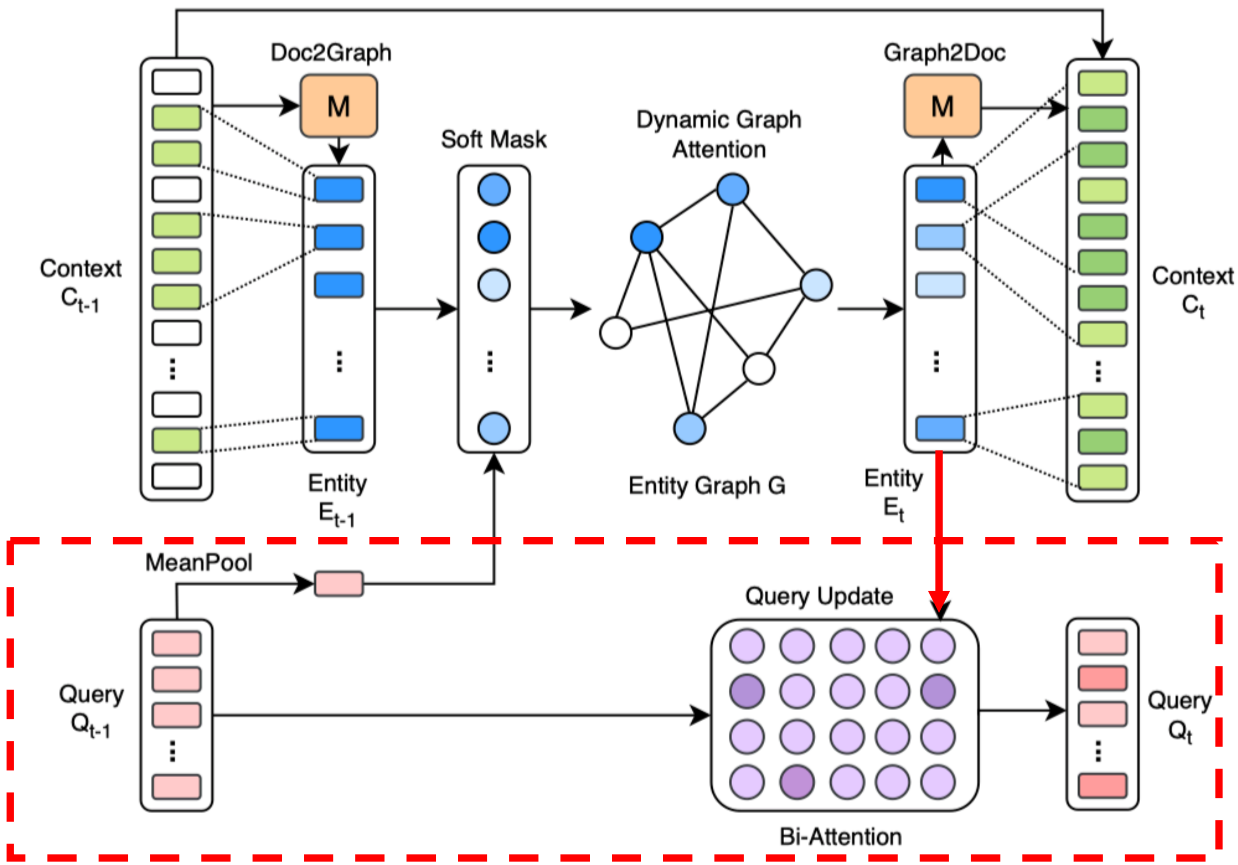
$$\alpha_{i,j}^{(t)} = \frac{\exp(\beta_{i,j}^{(t)})}{\sum_k \exp(\beta_{i,k}^{(t)})}$$

$$\boxed{\mathbf{e}_i^{(t)} = \mathrm{ReLU}(\sum_{j \in B_i} \alpha_{j,i}^{(t)} \mathbf{h}_j^{(t)})}$$

set of neighbors of entity i

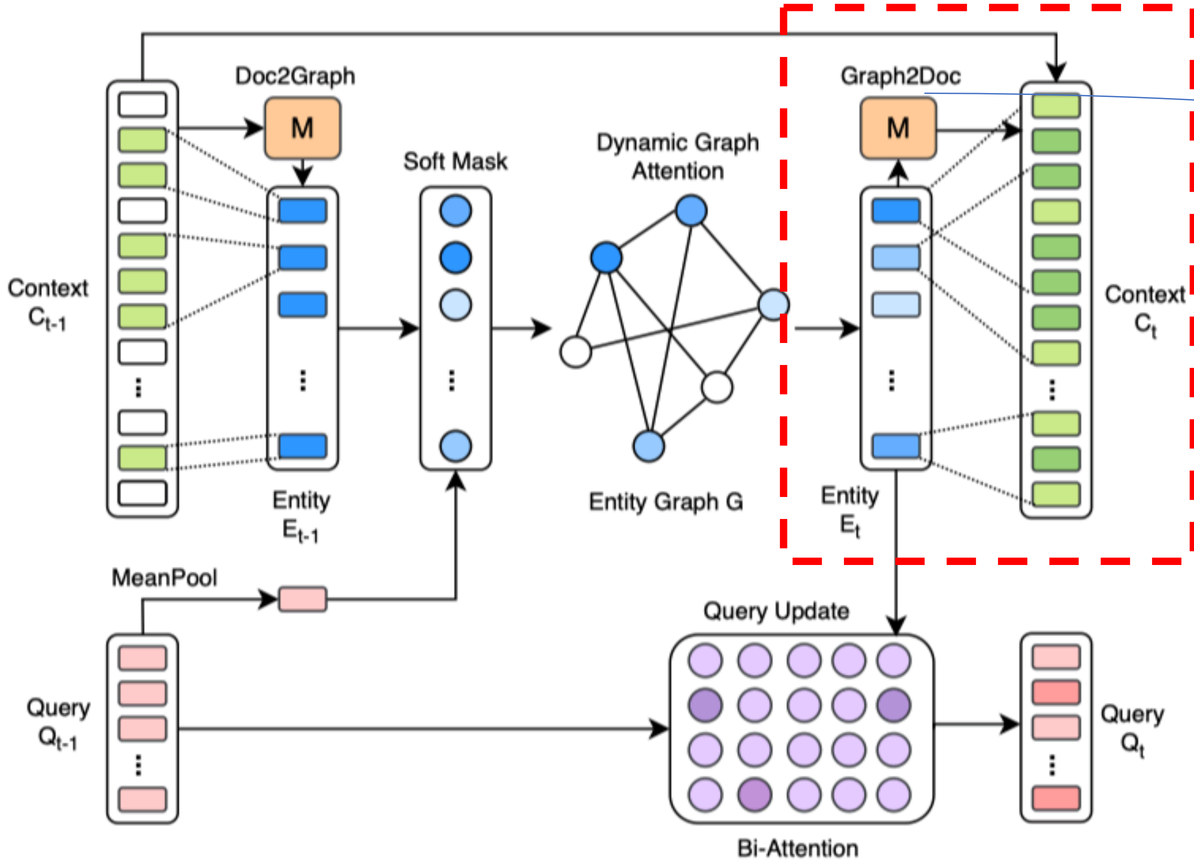
Updating Query

- In order to predict the expected start entities for the next step



$$\mathbf{Q}^{(t)} = \text{Bi-Attention}(\mathbf{Q}^{(t-1)}, \mathbf{E}^{(t)})$$

Graph to Document Flow



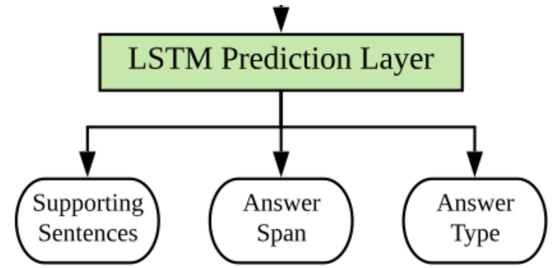
	E1	E2	E3
w1	0	0	0
w2	1	0	0

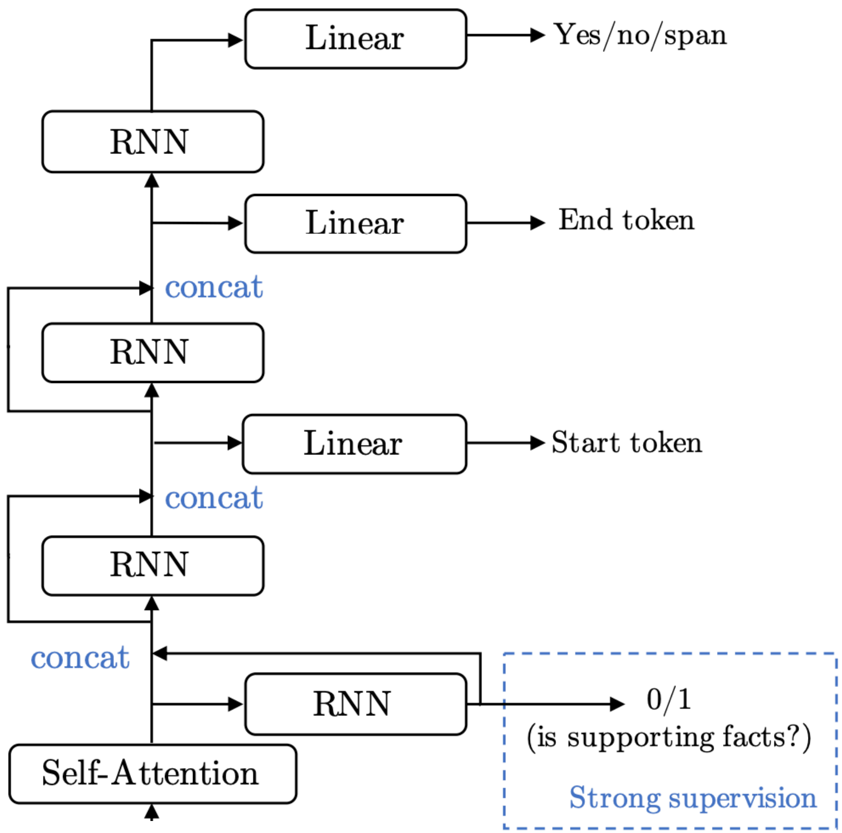
$$\mathbf{C}^{(t)} = \mathrm{LSTM}([\mathbf{C}^{(t-1)}, \mathbf{M}\mathbf{E}^{(t)\top}])$$

?

- The previous token embeddings in Ct−1 are **concatenated** with the associated entity embedding corresponding to the token.
- **;** refers to concatenation

Prediction





$$\mathbf{O}_{sup} = \mathcal{F}_0(\mathbf{C}^{(t)})$$
$$\mathbf{O}_{start} = \mathcal{F}_1([\mathbf{C}^{(t)}, \mathbf{O}_{sup}])$$
$$\mathbf{O}_{end} = \mathcal{F}_2([\mathbf{C}^{(t)}, \mathbf{O}_{sup}, \mathbf{O}_{start}])$$
$$\mathbf{O}_{type} = \mathcal{F}_3([\mathbf{C}^{(t)}, \mathbf{O}_{sup}, \mathbf{O}_{end}])$$

$$\mathcal{L} = \mathcal{L}_{start} + \mathcal{L}_{end} + \lambda_s \mathcal{L}_{sup} + \lambda_t \mathcal{L}_{type}$$

HOTPOTQA: A Dataset for Diverse, Explainable Multi-hop Question Answering

Weak Supervision

- Soft masks at each fusion block to match the **heuristic masks**.
- **Heuristic masks**
  - Start mask detected from the query
  - Additional BFS masks obtained by applying breadth first search (BFS) on the adjacent matrices give the start mask
- A **binary cross entropy loss** between the predicted soft masks and the heuristics is then added to the objective.

Experiments

- Distractor setting
  - a question-answering system reads 10 paragraphs to provide an answer (Ans) to a question.
- Fullwiki Setting
  - a question-answering system must find the answer to a question in the scope of the entire Wikipedia.

https://hotpotqa.github.io/

Main Results

Model	Answer		Sup Fact		Joint	
	EM	F1	EM	F1	EM	F1
Baseline Model	45.60	59.02	20.32	64.49	10.83	40.16
ASNet*	**56.01**	**69.90**	42.30	80.56	26.86	58.13
DFGN (Ours)	55.17	68.49	49.85	81.06	31.87	58.23
QFE*	53.86	68.06	**57.75**	**84.49**	**34.63**	**59.61**

Table 1: Performance comparison on the private test set of HotpotQA in the distractor setting. Our DFGN is the second best result on the leaderboard now (on March 1st). The baseline model is from Yang et al. (2018) and the results with * is unpublished.

Leaderboard (Distractor Setting)

In the *distractor* setting, a question-answering system reads 10 paragraphs to provide an answer (Ans) to a question. They must also justify these answers with supporting facts (Sup).

	Model	Code	Ans		Sup		Joint	
			EM	F_1	EM	F_1	EM	F_1
1 Apr 11, 2019	HiPaR-net + BERT (single model) *Anonymous*		**60.13**	**73.31**	52.55	83.20	35.40	**63.41**
2 May 16, 2019	SAE (single model) *Anonymous*		59.77	72.77	52.53	82.82	**35.54**	62.92
3 Apr 19, 2019	GRN + BERT (single model) *Anonymous*		55.12	68.98	52.55	84.06	32.88	60.31
4 Apr 22, 2019	DFGN (single model) *Shanghai Jiao Tong University & ByteDance AI Lab* (Xiao, Qu, Qiu et al. ACL19)		56.31	69.69	51.50	81.62	33.62	59.82

Ablation study

Setting	EM	F1
DFGN (2-layer)	55.42	69.23
- BFS Supervision	54.48	68.15
- Entity Mask	54.64	68.25
- Query Update	54.44	67.98
- E2T Process	53.91	67.45
- 1 Fusion Block	54.14	67.70
- 2 Fusion Blocks	53.44	67.11
- 2 Fusion Blocks & Bi-attn	50.03	62.83
gold paragraphs only	55.67	69.15
supporting facts only	57.57	71.67

Model ablation

Dataset ablation

Table 2: Ablation study of question answering performances on the develop set of HotpotQA in the distractor setting. We use a DFGN with 2-layer fusion blocks as the origin model. The upper part is the model ablation results and the lower part is the dataset ablation results.

- Using 1-layer fusion block leads to an obvious performance loss, which implies the significance of performing multi-hop reasoning in HotpotQA.
- Model is not very sensitive to the noise paragraphs

Evaluation on Graph Construction and Reasoning Chain

- Missing supporting entity
    - **Limited accuracy of NER model** and **the incompleteness of our graph construction**, 31.3% of the cases in the develop set are unable to perform a complete reasoning process
- Focus on the rest 68.7% good cases in the following analysis.

ESP (Entity-level Support) scores

- **Path**
  - sequence of entities visited by the fusion blocks

$[e_{p_1}, \ldots, e_{p_{t+1}}]$ (suppose t-layer fusion blocks)

- **Path Score**
  - multiplying corresponding soft masks and attention scores along the path

$$score(P) = \prod_{i=1}^{t} m_{i,p_i} \alpha_{t,p_i,p_{i+1}}$$

- **Hit**
  - Given a path and a supporting sentence, if at least one entity of the supporting sentence is visited by the path, we call this supporting sentence is **hit.**

ESP (Entity-level Support) scores

- **ESP EM (Exact Match)**
  - For a case with m supporting sentences, if all the m sentences are hit, we call this case is **exactly match**
  - ESP EM score is the ratio of exactly matched cases.
- **ESP Recall**
  - For a case with m supporting sentences and h of them are hit, this case has a recall score of h/m.

k	1	2	5	10
ESP EM	7.4%	15.5%	29.8%	41%
ESP Recall	37.3%	46.1%	58.4%	66.4%

top-k paths

Table 3: Evaluation of reasoning chains by ESP scores.

Case study-Good

0.67	0.01	0.01	Barrack
0	0.8	0.67	Provisional Irish Republican Army
0.01	0.69	1.09	IRA
0.02	0	0	British Royal Navy
0.01	0	0	British Army Gazelle
0	0	0	Falkland Islands
0.74	0	0.01	British Army Lynx
0	0.82	0.41	Provisional Irish Republican Army
0	0.81	0.73	IRA
0	0.73	0.13	Northern Ireland
0.01	0.33	0.61	IRA
0	0.47	0.05	South Armagh Brigade

Q: Who used a Barrack buster to shoot down a British Army Lynx helicopter?
Answer: IRA **Prediction:** IRA
Top 1 Reasoning Chain: British Army Lynx, Provisional Irish Republican Army, **IRA**

Supporting Fact 1:
"Barrack buster is the colloquial name given to several improvised mortars, developed in the 1990s by the engineering group of the Provisional Irish Republican Army (**IRA**)."
Supporting Fact 2:
" On 20 March 1994, a British Army Lynx helicopter was shot down by the Provisional Irish Republican Army (**IRA**) in Northern Ireland."

- **Mask1** : as the start entity mask of reasoning, where "Barrack" and "British Army Lynx"
- **Mask2** : mentions of the same entity "IRA"

Case study-Bad

Mask1	Mask2	End	
0.01	0.07	0	Sasanid
0	0.07	0	Iran
0	0.02	0	Islam
0	0.02	0	House of Sasan
0	0.02	0	Roman-Byzantine Empire
0	0.11	0	Samo
0.04	0.03	0	King
0	0.03	0.01	Samo
0	0	0	Moravia

Q: From March 631 to April 631, Farrukhzad Khosrau V was the king of an empire that succeeded which empire?
Answer: the Parthian Empire **Prediction:** Parthian Empire
Top 1 Reasoning Chain: n/a

Supporting Fact 1:
"Farrukhzad Khosrau V was briefly king of the Sasanian Empire from March 631 to ..."
Supporting Fact 2:
"The Sasanian Empire, which succeeded the Parthian Empire, was recognised as ... the Roman-Byzantine Empire, for a period of more than 400 years."

- Due to the malfunction of the NER module, the only start entity, "Farrukhzad Khosrau V", was not successfully detected.

Conclusion

- DFGN, a novel method for the multi-hop text-based QA problem
- Provide a way to explain and evaluate the reasoning chains via interpreting the entity graph masks predicted by DFGN. The mask prediction module is additionally weakly trained.
- Provide an experimental study on a public dataset (HotpotQA) to demonstrate that our proposed DFGN is competitive against state-of-the-art unpublished works.

Thanks!