

MAFiD: Moving Average Equipped Fusion-in-Decoder for Question Answering over Tabular and Textual Data

Sung-Min Lee, Eunhwan Park, Daeryong Seo, Donghyeon Jeon,
Inho Kang, Seung-Hoon Na

Jeonbuk National University

NAVER Corporation



NAVER

Introduction

Multi-Hop Question Answering over Tabular and Textual Data

The 2016 Summer Olympics officially known as the Games of the XXXI Olympiad (Portuguese : Jogos da XXXI Olimpíada) and commonly known as **Rio 2016** , was an international multi-sport event

Name	Year	Season	Flag bearer
XXXI	2016	Summer	Yan Naing Soe
XXX	2012	Summer	Zaw Win Thet
XXIX	2008	Summer	Phone Myint Tayzar
XXVIII	2004	Summer	Hla Win U
XXVII	2000	Summer	Maung Maung Nge
XX	1972	Summer	Win Maung

Yan Naing Soe (born **31 January 1979**) is a Burmese judoka . He competed at the 2016 Summer Olympics in the **men 's 100 kg event** , He was the flag bearer for Myanmar at the **Parade of Nations** .

Zaw Win Thet (born **1 March 1991** in Kyonpyaw , Pathein District , Ayeyarwady Division , Myanmar) is a Burmese runner who

Myint Tayzar Phone (Burmese : မြင့်တေဇာဖုန်း) born **July 2 , 1978**) is a sprint canoer from Myanmar who competed in the late 2000s .

.....

Win Maung (born **12 May 1949**) is a Burmese footballer . He competed in the men 's tournament at the 1972 Summer Olympics ...

Hardness
↓

Q: In which year did the judoka bearer participate in the Olympic opening ceremony?

A: 2016

Q: Which event does the does the XXXI Olympic flag bearer participate in?

A: men's 100 kg event

Q: Where does the Burmesse jodoka participate in the Olympic opening ceremony as a flag bearer?

A: Rio

Q: For the Olympic event happening after 2014, what session does the Flag bearer participate?

A: Parade of Nations

Q: For the XXXI and XXX Olympic event, which has an older flag bearer?

A: XXXI

Q: When does the oldest flag Burmese bearer participate in the Olympic ceremony?

A: 1972

HybridQA examples of annotated question answering pairs from Wikipedia page

Prior work: Fusion-in-Decoder

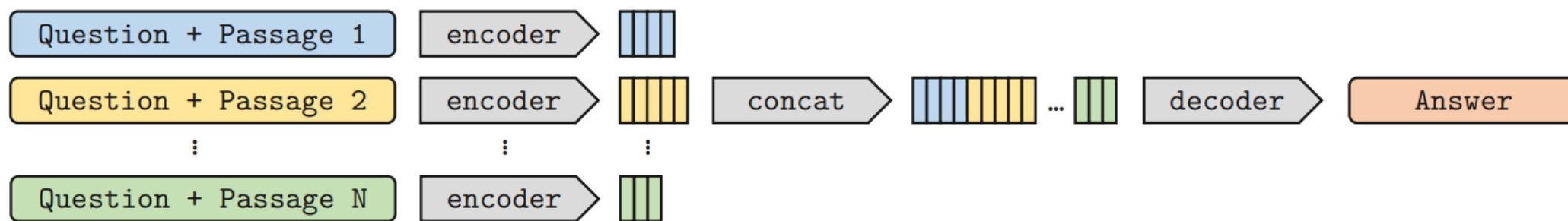


Figure 2: Architecture of the Fusion-in-Decoder method.

In FiD, All question augmented retrieved passages are encoded by LM encoder and decoder performs attention over the concatenation of the resulting representations of all the retrieved passages.

Prior work: MEGA

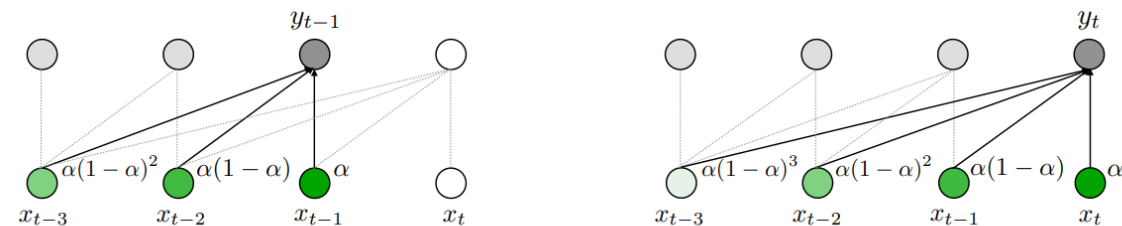
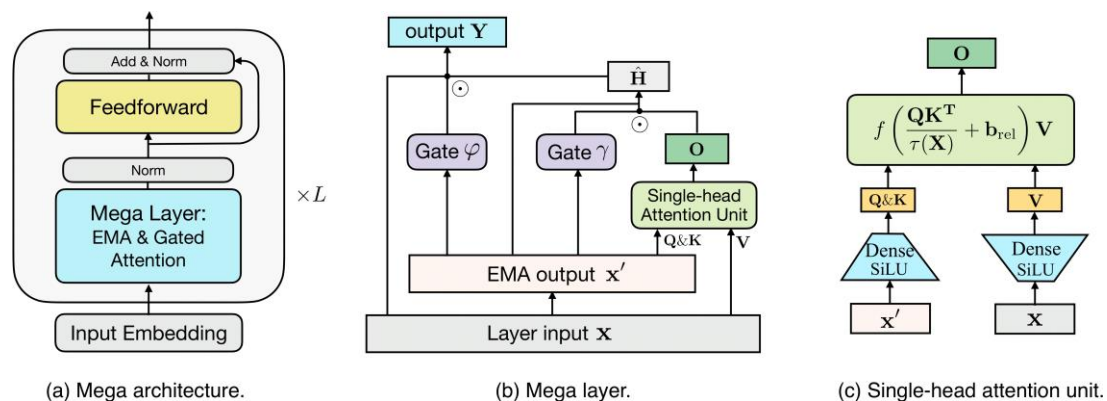
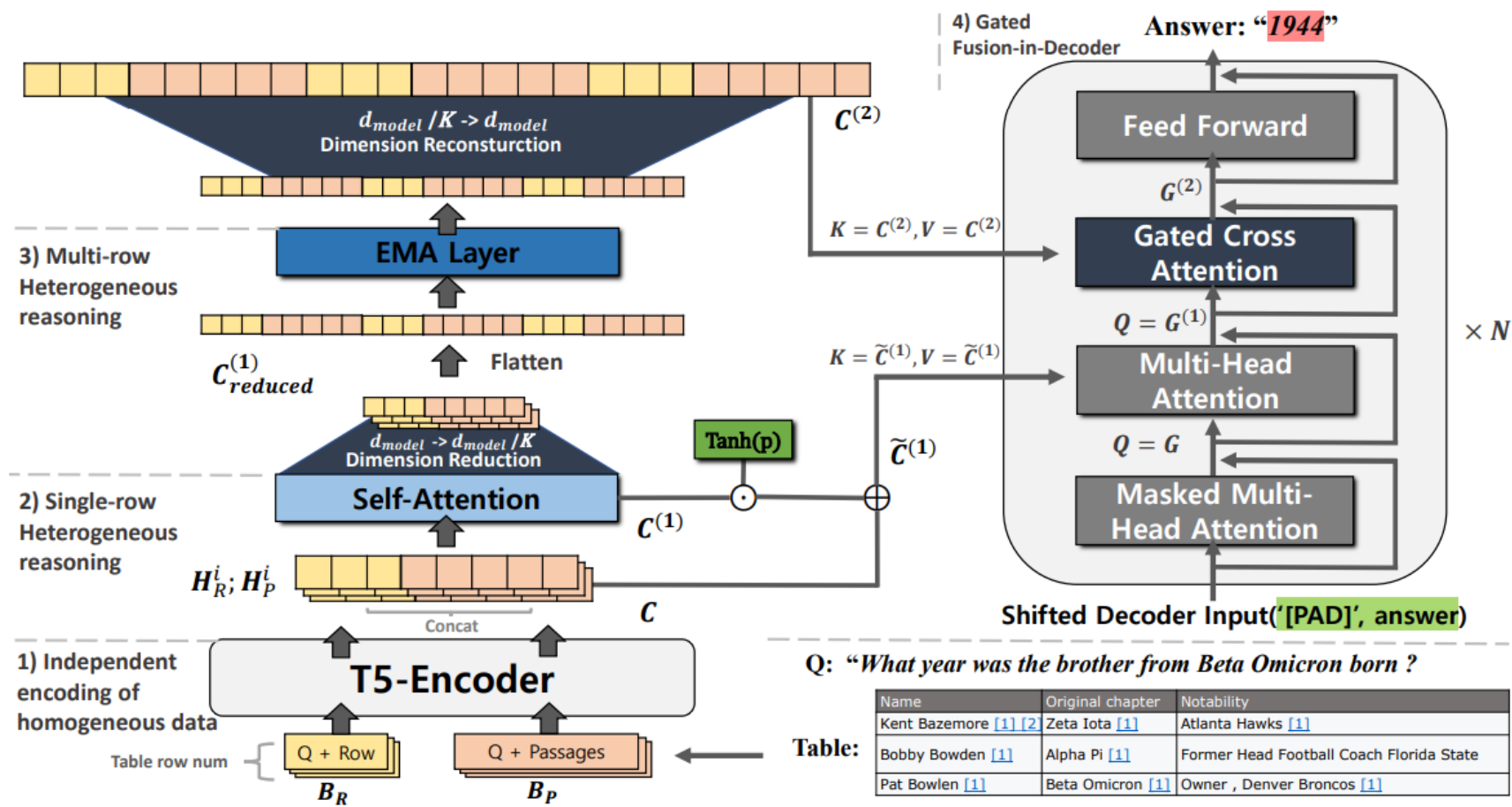


Figure 1: Illustration of the exponential moving average (EMA) approach, which averages the input values X with weights decaying exponentially over timesteps.

- MEGA incorporates an exponential moving average to provide position-aware local dependencies, adding a strong inductive bias to the position-agnostic attention mechanism.
- EMA: more efficient than self-attention (quadratic complexity) and complements the limitations of self-attention that does not effectively perform long sequence reasoning.
- **Question: Can we enhance long sequence reasoning for Table-Text QA by incorporating the advantages of EMA and FiD?**

Our approach: MAFiD

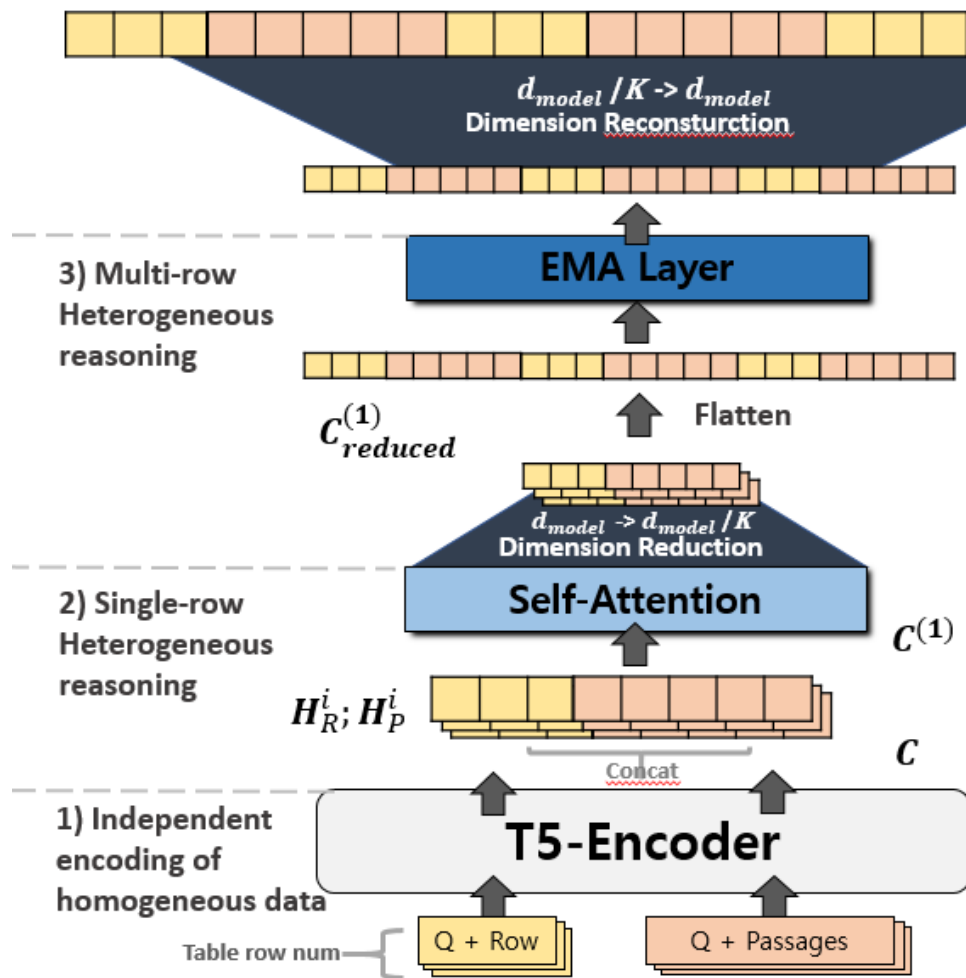
MAFiD: Moving Average Equipped Fusion-in-Decoder for Question Answering over Tabular and Textual Data



- The overall neural architecture of the proposed MAFiD -

MAFiD: Moving Average Equipped Fusion-in-Decoder for Question Answering over Tabular and Textual Data

1) Independent Encoding of Homogeneous Data: the Basic Encoder for FiD



$$\text{row}^i = [q; [\text{SEP}]; b_R^i], \text{psg}^i = [q; [\text{SEP}]; b_P^i]$$

$$H_R^i = \text{T5-enc}(\text{row}^i) \in \mathbb{R}^{|\text{row}^i| \times d_{model}}$$

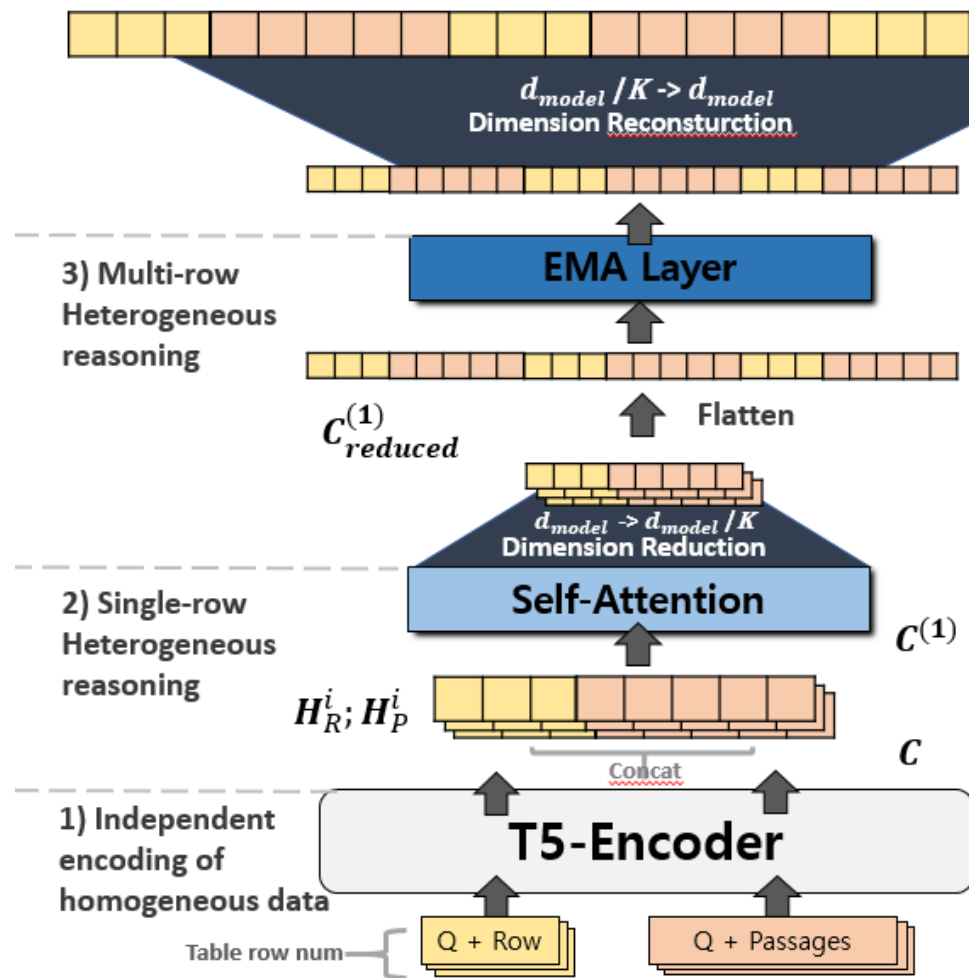
$$H_P^i = \text{T5-enc}(\text{psg}^i) \in \mathbb{R}^{|\text{psg}^i| \times d_{model}}$$

$$C_i = [H_R^i; H_P^i] \in \mathbb{R}^{(|\text{row}^i| + |\text{psg}^i|) \times d_{model}}$$

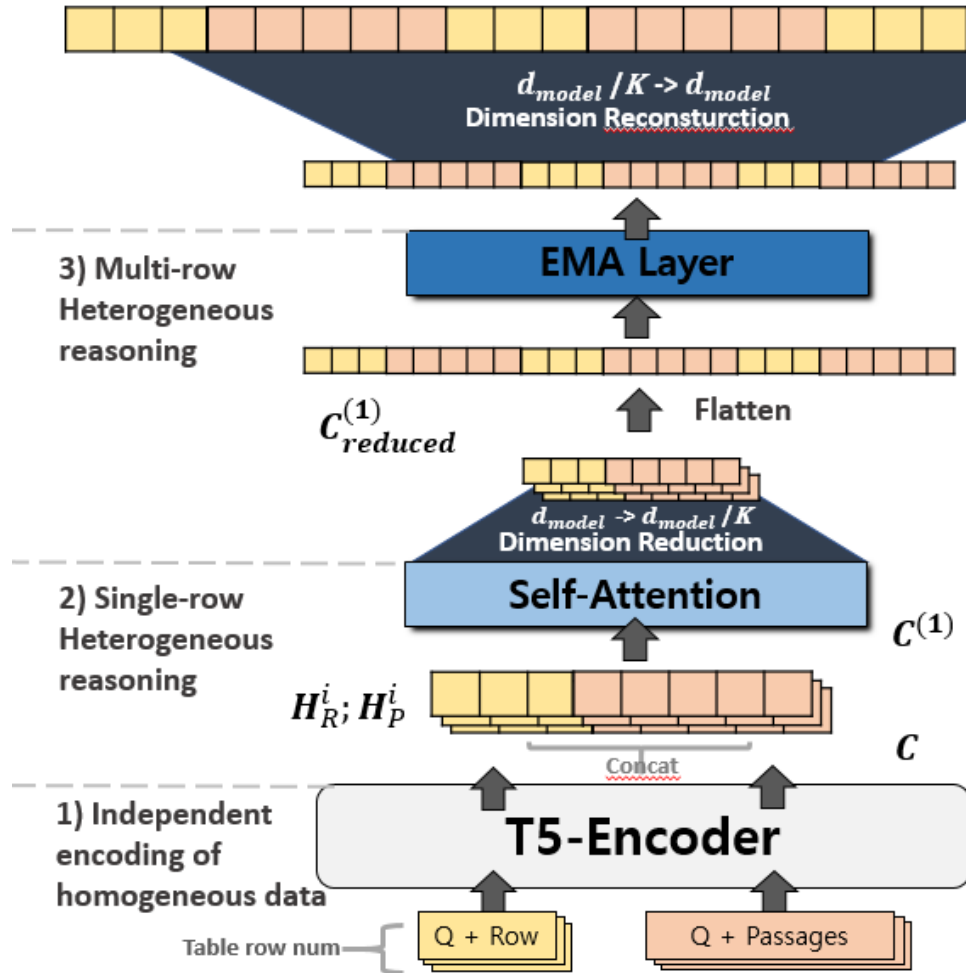
MAFiD: Moving Average Equipped Fusion-in-Decoder for Question Answering over Tabular and Textual Data

2) Single-row Heterogeneous Reasoning

$$C_i^{(1)} = \text{SHA}(C_i, C_i, C_i)$$



MAFiD: Moving Average Equipped Fusion-in-Decoder for Question Answering over Tabular and Textual Data



3) Multi-row Heterogeneous Reasoning by the Low-dimensional EMA

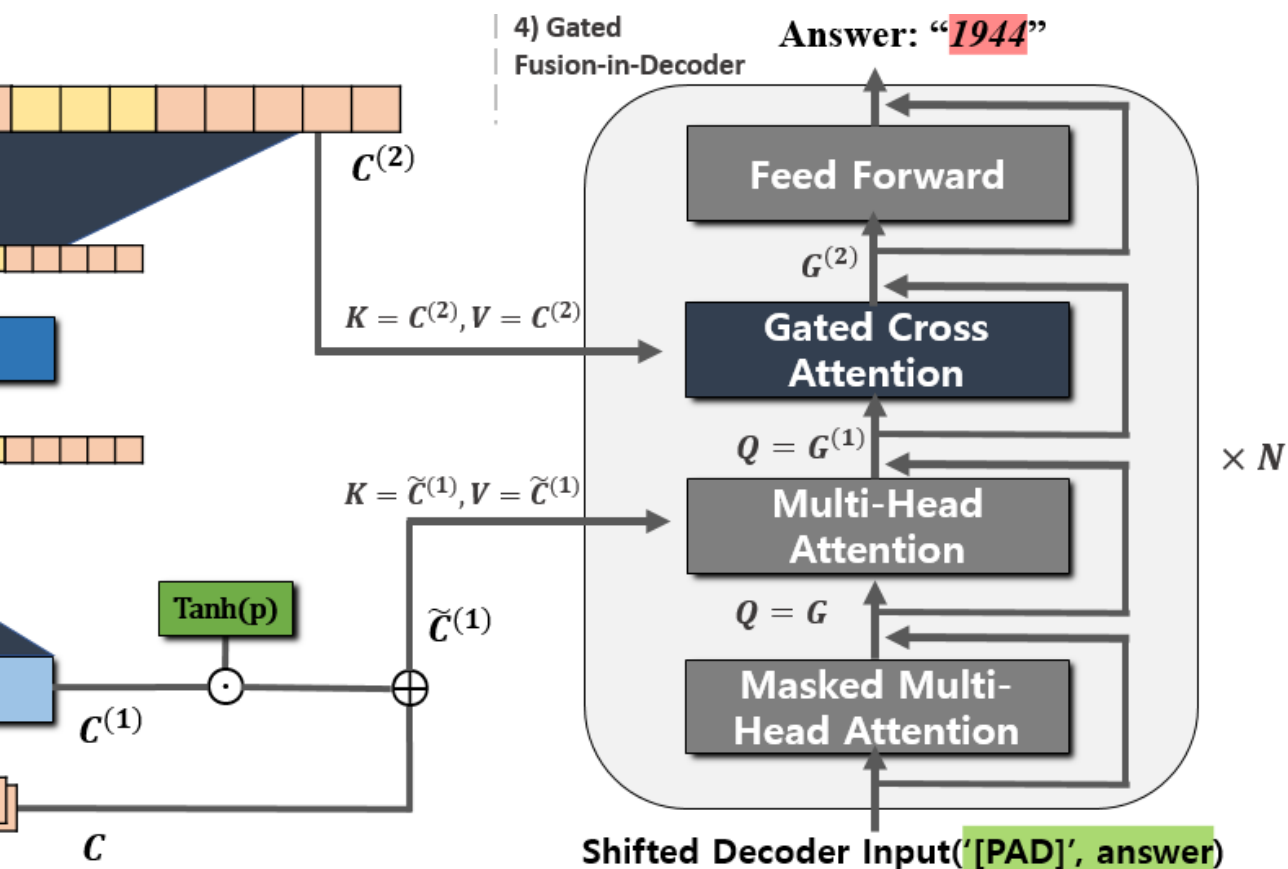
$$C^{(1)} = [C_1^{(1)}; \dots; C_L^{(1)}]$$

$$C_{reduced}^{(1)} = \text{Linear} \left(C^{(1)} \right)$$

$$C_{reduced}^{(2)} = \text{EMA} \left(C_{reduced}^{(1)} \right)$$

$$C^{(2)} = \text{Linear} \left(C_{reduced}^{(2)} \right)$$

MAFiD: Moving Average Equipped Fusion-in-Decoder for Question Answering over Tabular and Textual Data



4) Gated Fusion-in-Decoder

$$C = [C_1; \cdots; C_L]$$

$$\tilde{C}^{(1)} = C + \tanh(p) \odot C^{(1)}$$

$$G^{(1)} = \text{MHA}(G, \tilde{C}^{(1)}, \tilde{C}^{(1)})$$

$$G^{(2)} = G^{(1)} + \tanh(q) \odot \text{MHA}(G^{(1)}, C^{(2)}, C^{(2)})$$

Experiments

Main results

MAFiD shows the state-of-the-art performance by increasing EM and F1 by 1.1 and 1.7 over MITQA (Kumar et al., 2021) on the blind test set.

	Table				Passage				Total			
	Dev		Test		Dev		Test		Dev		Test	
	EM	F1	EM	F1	EM	F1	EM	F1	EM	F1	EM	F1
HYBRIDER	51.5	58.6	52.1	59.3	40.5	47.9	38.1	46.3	43.7	50.9	42.5	50.2
HYBRIDER-Large	54.3	61.4	56.2	63.3	39.1	45.7	37.5	44.4	44.0	50.7	43.8	50.6
DocHopper	-	-	-	-	-	-	-	-	47.7	55.0	46.3	53.3
POINTR + TAPAS	68.1	73.9	67.8	73.2	62.9	72.0	62.0	70.9	63.3	70.8	62.7	70.0
POINTR + MATE	68.6	74.2	66.9	72.3	62.8	71.9	62.8	71.9	63.4	71.0	62.8	70.2
MITQA	68.1	73.3	68.5	74.4	66.7	75.6	64.3	73.3	65.5	72.7	64.3	71.9
Ours	69.4	75.2	68.5	74.9	66.5	75.5	65.7	75.3	66.2	74.1	65.4	73.6
Human	-	-	-	-	-	-	-	-	-	-	88.2	93.5

Comparison results on the dev and blind test dataset in HybridQA.

- **HYBRIDER** (Chen et al., 2020) employs a sparse passage retriever to find relevant cells and performs the reasoning step consisting of the ranking, the hop, and the reading comprehension models to extract an answer.

- **DocHopper** (Sun et al., 2021) uses the “iterative hierarchical attention” to retrieve short or long contents in a multi-step navigational manner.

- **POINTR + (TAPAS or MATE)** (Herzig et al., 2020; Eisenschlos et al., 2021a). POINTR extends the cell with its entity description and performs a two-stage method that consists of “cell selection” and “passage reading” steps. Either TAPAS (Herzig et al., 2020) or MATE (Kumar et al., 2021) is considered as a transformer encoder.

- **MITQA** (Kumar et al., 2021) uses the pipelined module including a retriever, a reader, and a joint row+span reranker, etc., being trained using the multi-instance distant supervision approach.

Ablation Studies

	Table		Passage		Total	
	EM	F1	EM	F1	EM	F1
Ours	68.48	74.92	65.75	75.34	65.38	73.56
w/o Multi-row reasoning	67.44	73.74	65.50	75.23	64.86	73.08
w/o Multi-row, Single-row reasoning	41.97	49.46	60.20	69.42	51.46	59.86
w/o Single-row tanh gate	67.21	73.44	64.86	74.82	64.45	72.75
w/o Multi-row tanh gate	67.58	73.96	66.43	75.47	65.46	73.29
w/o Single-row, Multi-row tanh gate	66.09	72.51	64.81	75.22	64.01	72.65

Ablation study on blind test dataset in HybridQA.

- MAFiD without both reasonings significantly deteriorates the performance of EM and F1 by 13.92 and 13.7, respectively indicating the cross-modal interaction should be performed at least within a specific row, whereas the between-row interaction is somehow effectively proceeded by the proposed EMA module.
- MAFiD without the single-row tanh gate ($\tanh(p) = 1$) slightly decreases EM and F1 by approximately 11.5, indicating that the gated FiD is helpful for further improvements.

Impact of EMA

	Total			
	Dev		Test	
	EM	F1	EM	F1
EMA	66.2	74.1	65.4	73.6
sliding window attention	65.7	73.3	65.3	73.1
Human	-	-	88.2	93.5

Comparison results on the dev and blind test sets in HybridQA between EMA and the sliding window attention of (Beltagy et al., 2020) for long-range reasoning.

- The use of EMA increases F1 and EM by 0.1 and 0.5, respectively, suggesting that EMA is more helpful for promoting the enhanced local sequence representation.

Impact of Sequential Order

	Total			
	Dev		Test	
	EM	F1	EM	F1
original rows	66.2	74.1	65.4	73.6
permuted rows	51.5	59.4	51.1	59.2
Human	-	-	88.2	93.5

Comparison results of MAFiD on HybridQA between the case using original rows and that with permuted rows for tabular contents.

- A variant of MAFiD by randomly permuting rows in tabular contents both for training and inference, referred to as “permuted row”, comparing to the original case; the results strongly indicate that keeping original row orders is important for MAFiD.

Error Analysis

Q: “How many annual visitors are pulled in by the marathon that Tegna Loroupe won in 2002 ?

Table:

Year	Competition	Venue	Position	Notes
2000	New York City Marathon [1]	New York City , United States [1]	6th	Marathon
2002	Nagoya Women 's Marathon [1]	Nagoya , Japan [1]	7th	Marathon
2002	Lausanne Marathon [1]	Lausanne , Switzerland [1]	1st	Marathon

The New York City Marathon (currently branded TCS New York City Marathon for sponsorship reasons) is an annual marathon (42.195 km or 26.219 mi) that courses through the five boroughs of New York City . It is the largest marathon in the world , with 52,812 finishers in 2018 and 98,247 applicants for the 2017 race .

The Lausanne Marathon is one of the rare Swiss races that organises a half-marathon for wheelchairs and handcycles . The Lausanne Marathon is one of the largest annual sporting events of the Canton de Vaud , [citation needed] and attracts up to 2,500 tourists each year . In 2009 , a record 10,658 runners participated .

A: “2,500” MAFiD: “2,500” HYBRIDER-Large: “52,812”

Q: “What is the month of birth of the player with the fourth most National Football League career rushing yards ?

Table:

Rank	Player	Team (s) by season	Yards
3	Frank Gore [1]	San Francisco 49ers ... [1], ... , [10]	15,347
4	Barry Sanders [1]	Detroit Lions [1],[2],[3]	15,269
5	Adrian Peterson [1]	Minnesota Vikings [1],[2],[3], ... [8],[9]	14,216

Franklin Delano Frank Gore (born May 14 , 1983) is an American professional football player who is a running back for the Buffalo Bills of the National Football League (NFL) . He played college football for the University of Miami , and was drafted by the San Francisco 49ers ...

Barry Sanders (born July 16 , 1968) is an American former professional football player who was a running back for the Detroit Lions of the National Football League (NFL) . A Pro Bowl invitee in each of his ten NFL seasons and two-time NFL Offensive Player of the Year , Sanders led the league in rushing yards four times ...

A: “July” MAFiD: “July” HYBRIDER-Large : “May”

Q: “Which defender is the youngest ?”

Table:

Pick #	MLS team	Player	Position
22	Portland Timbers [1]	Chris Taylor	Defender [1]
25	New York Red Bulls [1]	John Rooney [1]	Midfielder [1]
26	Toronto FC [1]	Demitrius Omphroy [1]	Defender [1]

John Richard Rooney (born 17 December 1990) is an English professional footballer who plays for Barrow as an attacking midfielder . He is the younger brother of Derby County and former England forward Wayne Rooney . Although born in England , Rooney has expressed a desire to represent the Republic of Ireland at international level .

A: “Demitrius Omphroy” MAFiD: “Chris Taylor” HYBRIDER-Large : “John Rooney”

Illustrating examples of HYBRIDER-Large (Chen et al., 2020) and MAFiD in HybridQA.

Conclusion

- We proposed MAFiD, which extends FiD by equipping EMA and the gated cross-attention layer to design an effective way of combining various types of encoded representations.
- The experimental results on HybridQA showed that the proposed MAFiD achieved state-of-the-art performances in both the development and blind test sets.
- **Future works:** we will extend MAFiD to open-domain table-and-text QA and explore a unified approach that integrates single-row and multi-row reasoning.

Thanks!

- To my professor(Seung-Hoon Na) and NAVER Corporation!
- To all anonymous reviewers for their valuable comments and suggestions
- Code will be available at <https://github.com/ZIZUN/MAFiD>