Supplementary file for Episodic Memory Reader: Learning What to Remember for Question Answering from Streaming Data

Moonsu Han[1*] **Minki Kang**[1*] **Hyunwoo Jung**[1] **Sung Ju Hwang**[1,2]

KAIST[1], Daejeon, South Korea
AITRICS[2], Seoul, South Korea
{mshan92, zzxc1133, hyunwooj, sjhwang82}@kaist.ac.kr

A bAbI dataset

We provide examples of the Original and Noisy datasets, as well as visualization of the memorized examples to show what our EMR models have remembered, for bAbI (Weston et al., 2015) dataset.

Dataset We visualize an example for **Original** and **Noisy** tasks in Figure 1.

Results and Analysis As shown in Figure 2, we further report the performance of the baseline models and our EMR variants, on how many supporting facts they retrain in the memory (denoted as solvable), by considering the QA performance with a perfect QA model. We observe that both EMR variants, EMR-Independent and EMR-Transformer, significantly outperform rule-based memory scheduling agents as well as EMR-Independent.

B TriviaQA

We provide more experiment details and additional examples for analyzing what our EMR models have remembered, for TriviaQA (Joshi et al., 2017).

Dataset The objective of our model is to learn general importance in situations where not knowing the question from streaming data. In terms of scalability, our model is able to access sequentially a large amount of streaming data by replacing the most uninformative memory entry in the external memory. When comparing TriviaQA with a common question-answering dataset (Rajpurkar et al., 2016; Weston et al., 2015), it is an appropriate dataset to prove the efficiency of our model since its average word number is approximately 3K which cannot be accessed using conventional models that predicts answer indices using a pointer network (Seo et al., 2016; Back et al., 2018; Yu et al., 2018).

To preprocess TriviaQA according to problem setting, we truncate all documents within 1200 words for a training set, in order to reduce the cost of training process. Unlike the training set, a test set takes all words in the documents. Although TriviaQA does not provide the answer indices in a document, we extract the documents that can be spanned to adopt Deep Bidirectional Transformers (BERT) (Devlin et al., 2018), which is state-of-the-art reading comprehension model using a pointer network. Additionally, we made all letters lowercase and removed all special characters.

Experiment Details As described in the main paper, we employed the pre-trained BERT to solve TriviaQA. A more specific implementation is described here. We encode the current input $x^{(t)}$ to $m_i^{(t)}$ using the BERT encoding layer and a bi-directional GRU whose output size is 768 and 128, respectively. The reason for using it is to convert the words into a sentence. By doing this, it can make accessing a possible chunk of words and computation cost is reduced. We utilize $m_i^{(t)}$ to output relative importance between the memory entries $\{m_1^{(t)}, ..., m_i^{(t)}\}$, where i indicates an address in the memory entry, as described in the main paper. In addition to using the pre-trained BERT, we finetune it with truncated documents (Up to 400 words) in the same way as LIFO (Last-In-First-Out) since hoping our model focuses on learning what to remember in the external memory and generalizes well even watching limited contents in the documents. We train our model and the baseline models using ADAM optimizer (Kingma and Ba, 2014), with the initial learning rate of 1e-5 and dropout probability of 0.1 for 1M steps. For A3C (Mnih et al., 2016), we set the discount factor to 0.1 and entropy regularization to 0.01 for all

* Equal contribution

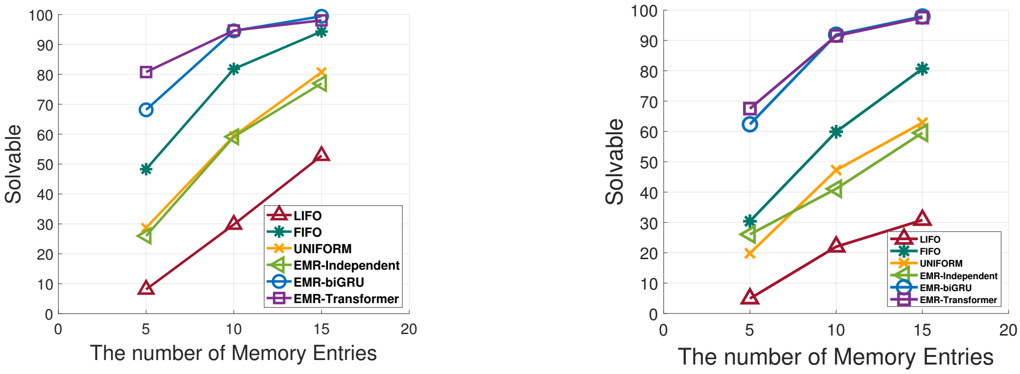
Index	Context
1	Mary journeyed to the bathroom
2	Sandra went to the garden
⋮	⋮
6	Sandra put down the milk there
Where is the milk?	Garden [2, 6]
8	Daniel went to the garden
⋮	⋮
17	Daniel dropped the football
Where is the football?	Bedroom [12, 17]
19	Sandra left the milk there
20	Daniel grabbed the football there
Where is the milk?	Bedroom [16, 19]
22	Sandra grabbed the milk there
23	Daniel went to the kitchen
Where is the football?	Kitchen [20, 23]
25	John travelled to the kitchen
26	Mary moved to the hallway
Where is the football?	Kitchen [7, 9]

(a) Original

Index	Context
1	Sandra moved to the kitchen
2	Wolves are afraid of cats
3	Sandra put down the apple
4	Sandra took the milk
5	Sandra is yellow
6	Mary is green
7	Sandra grabbed the apple
8	Mary went to the hallway
Where is the milk?	Kitchen [1, 4]
⋮	⋮
38	Mice are afraid of wolves
39	Wolves are afraid of wolves
40	Sandra let go of the football
41	Mary is green
42	Mary journeyed to the kitchen
43	Mary grabbed the milk
44	John is bored
Where is the apple?	Kitchen [34, 42]

(b) Noisy

Figure 1: Example of (a) Original task and (b) Noisy task. Sentences in green are noise sentences and ones in blue are supporting facts of each question.



(a) Original (Solvable)

(b) Noisy (Solvable)

Figure 2: The results for our model (EMR-biGRU and EMR-Transformer) and the baselines. The reported results are averages over 3 runs. The Solvable represents an accuracy that when the model encounters a question, it contains supporting facts in the memory to solve the question.

Index	Context
23	Mary moved to the bedroom
26	Sandra moved to the garden
28	Sandra left the milk
31	Sandra put down the football
32	John journeyed to the bedroom
Where is the milk?	Garden [26, 28]

(a) EMR-biGRU (Original)

Index	Context
28	John journeyed to the hallway
34	Sandra journeyed to the kitchen
39	John grabbed the football
40	John grabbed the milk
42	Mary went to the garden
Where is the milk?	Hallway [28, 40]

(b) EMR-biGRU (Noisy)

Index	Context
3	Mary went back to the office
12	John journeyed to the garden
14	Mary put down the apple
17	Daniel went to the kitchen
18	John discarded the milk
Where is the apple?	Apple [3, 14]

(c) EMR-Transformer (Original)

Index	Context
28	Mary went to the hallway
32	Mary journeyed to the garden
40	Mary grabbed the milk
42	John moved to the hallway
44	Mary got the milk
Where is the milk?	Garden [32, 44]

(d) EMR-Transformer (Noisy)

Figure 3: Example of Original and Noisy task for EMR-biGRU and EMR-Transformer. Sentences in blue are supporting facts of each question. The Index on the figure represents the order of sentences in the context.

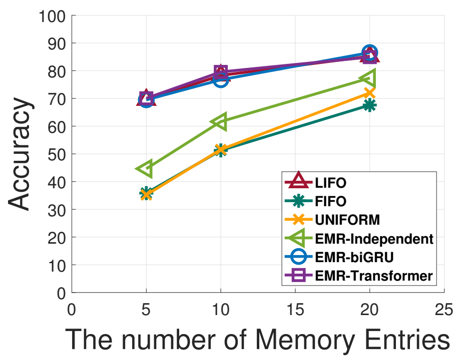


Figure 4: Oracle score

experiments.

Results and Analysis As shown in Figure 4, we report the score of each method using a perfect QA model, to see how many of the important facts are remembered by each method. We see that on TriviaQA dataset, LIFO contains similar amount of words as EMR-biGRU and EMR-Transformer. This is mostly due to the dataset bias, where most of the answers are found in earlier parts of the documents (Figure 6). However, our models outperform LIFO in QA task, since it observed more sentences during training which help the QA model to perform better, compared to LIFO that observed less number of training examples during training due to Last-In-First-Out policy that discards all words that come after the memory is filled.

C TVQA

We provide more experimental details and examples to show what our EMR models have remembered for the TVQA dataset. Each frame illustrated in the figure are the frames in the external memory at the last time step. The stars with different colors denote the supporting frames for different questions.

Experiment Details As described in the main paper, we use the Multi-stream model for Multi-Modal Video QA, which is suggested in Lei et al. (2018). We also pretrain the QA model for a delicate check of the performance of our EMR model. We use only the annotated frame when training the QA model. Since we use only the subtitle and frame image feature as input, we pretrain the QA model until reaching the reported performance of S+V model with the annotated time stamp in Lei et al. (2018).

Below is the detailed implementation of our model EMR. Since we have two kinds of input $x^{(t)}$ in TVQA, we need to blend them to one memory feature to be fitted to our model. In the case of subtitle input, we use GloVe (Pennington et al., 2014) to embed words to 300-size vectors. Then, we use bi-directional GRU to make the sentence 128-size vector from word vectors. In the case of video frame input, we use 2048-size feature vectors extracted from a ResNet-101 pretrained on the ImageNet dataset. Then, we compress video frame vectors to 128-size vector using a linear layer. Then, we add two 128-size feature vectors from the subtitle and the video frame to make 128-size of memory feature vector $m_i^{(t)}$. Other details including optimizer and reinforcement learning setting are described in the main paper.

References

Seohyun Back, Seunghak Yu, Sathish Reddy Indurthi, Jihie Kim, and Jaegul Choo. 2018. Memoreader: Large-scale reading comprehension through neural memory controller. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Mandar Joshi, Eunsol Choi, Daniel S. Weld, and Luke Zettlemoyer. 2017. Triviaqa: A large scale distantly supervised challenge dataset for reading comprehension. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017*.

Diederik P. Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980.

Jie Lei, Licheng Yu, Mohit Bansal, and Tamara L. Berg. 2018. TVQA: localized, compositional video question answering. *CoRR*, abs/1809.01696.

Volodymyr Mnih, Adrià Puigdomènech Badia, Mehdi Mirza, Alex Graves, Timothy P. Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. 2016. Asynchronous methods for deep reinforcement learning. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016*.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014*.

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. Squad: 100, 000+ questions for machine comprehension of text. In *Proceedings of* *the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016*.

Min Joon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hannaneh Hajishirzi. 2016. Bidirectional attention flow for machine comprehension. *CoRR*, abs/1611.01603.

Jason Weston, Antoine Bordes, Sumit Chopra, and Tomas Mikolov. 2015. Towards ai-complete question answering: A set of prerequisite toy tasks. *CoRR*, abs/1502.05698.

Adams Wei Yu, David Dohan, Minh-Thang Luong, Rui Zhao, Kai Chen, Mohammad Norouzi, and Quoc V. Le. 2018. Qanet: Combining local convolution with global self-attention for reading comprehension. *CoRR*, abs/1804.09541.

Index	(State 1) Memory
000	flanders dutch vlaanderen today normally refers to dutchspeaking northern portion of **belgium** it is
001	one of communities regions and language areas of **belgium** demonym associated with flanders …
	⋮
006	flanders to refer to entire dutchspeaking part of **belgium** stretching all way to river maas in
007	accordance with late 20th century belgian state reforms area was made into two political …
	⋮
012	ing those of northern italy **belgium** was one of centres of 19th century industrial revolution but …
	⋮
019	north consists of 22 exclaves surrounded by netherlands terminology in **belgium** term flanders has …
*020	community or flemish nation ie social cultural and linguistic scientific and educational economical …

Index	(State T) Memory
000	flanders dutch vlaanderen today normally refers to dutchspeaking northern portion of **belgium** it is
001	one of communities regions and language areas of **belgium** demonym associated with flanders …
	⋮
006	flanders to refer to entire dutchspeaking part of **belgium** stretching all way to river maas in
007	accordance with late 20th century belgian state reforms area was made into two political …
	⋮
012	ing those of northern italy **belgium** was one of centres of 19th century industrial revolution but …
	⋮
*245	flanders north brabant and limburg in north and east and with france french flanders and north sea in
416	longest time at 1 on chart

Question: Flanders is part of what country?
Prediction: **belgium**
Answer: **Belgium**

Figure 5: An example visualization of the memory. The answer word 'belgium' (Red / Thick) arrives at first timestep, and our model retains sentences at state T, which means after reading all the contexts. The star shape (*) indicates our model's selection which memory entry is deleted.

Index	(State 1) Memory
000	fens also known as is naturally marshy region in eastern england most of fens were drained several
001	centuries ago resulting in flat damp lowlying agricultural region fen is local name for individual …
	⋮
010	fens have been referred to as holy land of english because of churches and cathedrals of **ely** ramsey
	⋮
018	s around them were flooded largest of fenislands is isle of **ely** on which cathedral city of **ely**
*019	was built its highest point is 39 m above mean sea level without artificial drainage and flood …
020	would be liable to periodic flooding particularly in winter due to heavy load of water flowing down …

Index	(State T) Memory
000	fens also known as is naturally marshy region in eastern england most of fens were drained several
	⋮
010	fens have been referred to as holy land of english because of churches and cathedrals of **ely** ramsey
	⋮
018	s around them were flooded largest of fenislands is isle of **ely** on which cathedral city of **ely**
032	and internal drainage of land between rivers internal drainage was organised by levels or districts …
033	parts of one or several parishes details of organisation vary with history of their development but …
* 213	been set in fens bedford level appears in video game tom clancys endwar as possible battlefield

Question: The cathedral in which British city is known as 'The Ship of the Fens'?
Prediction: **ely**
Answer: **Ely**

Figure 6: An example visualization of the memory. The answer word 'ely' (Red / Thick) arrives at first timestep, and our model retains it after reading in all the context sentences. The star shape (*) indicates our model's selection which memory entry is deleted.

Index	(State 10) Memory
000	meringue is type of dessert often associated with french swiss and italian cuisine made from …
001	or aquafaba and sugar and occasionally acid such as lemon vinegar or cream of tartar binding agent
002	such as salt cornstarch or gelatin may also be added to eggs addition of powdered sugar
	⋮
017	c 1570 – c 1647 of gloucestershire and called pets in manuscript of collected recipes written by …
*018	fane 161213 – 1680 of knole kent slowly baked meringues are still referred to as
025	method best known to home cooks fine white sugar castor sugar is beaten into egg whites …
030	used for decoration on pie or spread on sheet or baked **alaska** base and baked swiss meringue is

Index	(State T) Memory
000	meringue is type of dessert often associated with french swiss and italian cuisine made from …
001	or aquafaba and sugar and occasionally acid such as lemon vinegar or cream of tartar binding agent
002	such as salt cornstarch or gelatin may also be added to eggs addition of powdered sugar
	⋮
*017	c 1570 – c 1647 of gloucestershire and called pets in manuscript of collected recipes written by …
025	method best known to home cooks fine white sugar castor sugar is beaten into egg whites …
030	used for decoration on pie or spread on sheet or baked **alaska** base and baked swiss meringue is
061	hydrates from refined sugar

Question: Which US state lends its name to a baked pudding, made with ice cream, sponge and meringue?
Prediction: **alaska**
Answer: **Alaska**

Figure 7: An example visualization of the memory. The answer word 'alaska' (Red / Thick) arrives at timestep 10, and our model retains it after reading in all the context sentences. The star shape (*) indicates our model's selection which memory entry is deleted.

[State T]

★Q1 : What did House say he had this morning when walking to Kutner and Taub?
★Q2 : What did House take out when he said oufr hours, not four months?
★Q3 : Where are Taub hands when he tells House that you couldn't share publicly?
★Q4 : What did Taub say he was going to run when talking to House in his office about last night?
★Q5 : How many hours of memory did House say he lost when he was talking to Taub in his office?
★Q6 : Who goes in House's office after House dismisses everyone to do what they told in the conference room?
★Q7 : Who does Taub think House took to a bar the night before when he is in House's office?

Figure 8: An example of clip from drama 'House'. Each frame with star is corresponding to question with the star of same color.

[State T]

★Q1 : Who enters the coffee shop after Ross shows everyone the paper?
★Q2 : Why is Monica excited after she enters the coffee shop?
★Q3 : What does Monica notice after she enters the shop?
★Q4 : What does Joey recall after he sees the newspaper?
★Q5 : What happens to Monica after she reads the paper?
★Q6 : What is Ross carrying when he walks into the coffee house?
★Q7 : What does Monica sit when she is holding a newspaper?

Figure 9: An example of clip from drama 'Friends'. Each frame with star is corresponding to question with the star of same color.

[State T]

★Q1 : Who was Valerie to calderon before she was killed and before they were intimate
★Q2 : Why did Calderon give Valerie the bracelette when he was questioned by beckett and castle?
★Q3 : What did the killer take from Valerie after killing her?
★Q4 : What did calderon say was significant about the necklace and the bracelet when castle and beckett said it looked familiar?
★Q5 : Who gave Valerie the bracelet before she was killed?
★Q6 : What did Calderon do after he take the picture from Castle?
★Q7 : What did Ryan took from Beckett hands when he was talking to Castle?

Figure 10: An example of clip from drama 'Castle'. Each frame with star is corresponding to question with the star of same color.

[State T]

★ Q1 : What drink bottle is at the table when Robin, Lily, Marshall, and Ted are talking to each other?
★ Q2 : What did Robin do after Simon called her and told her to wait?
★ Q3 : Who says that her folk's put in a pool when talking to Robin?
★ Q4 : What does Simon say Robin forgot to load when she is about to leave?
★ Q5 : What does Robin have in her hair when she tells Simon about a sprinkler she had?
★ Q6 : What was Ted doing when Lily explained why Robin wanted to meet the guy?
★ Q7 : What kind of jacket is Simon wearing when he talks to Robin about a pool?

Figure 11: An example of clip from drama 'When I met your mother'. Each frame with star is corresponding to question with the star of same color.