Question Answering

Sourangshu Bhattacharya

Story Comprehension

Joe went to the kitchen. Fred went to the kitchen. Joe picked up the milk. Joe travelled to his office. Joe left the milk. Joe went to the bathroom.

Questions from Joe's angry mother:

Q1 : Where is Joe?

Q2 : Where is the milk now?

Q3 : Where was Joe before the office?

Dialogue System

Machine

Hello! What can I do for you today?

I'd like to reserve a table for 6.

Sure! When would you like that reservation?

At 7 PM, please.

Okay. What cuisine would you like?

Actually make that 7:30 PM

Updated! What cuisine?

Is there anything better than a medium rare steak?

Nothing at all! Blackdog has a 4.7 on Yelp.

Sounds perfect! Also, add one more person.

Reservation done for 7, 7:30pm at Blackdog. Enjoy!

Human

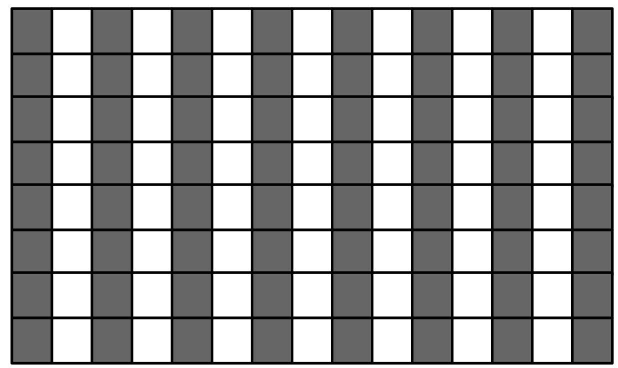
ML models need memory!

Deeper AI tasks require explicit memory and multi-hop reasoning over it

- RNNs have short memory
- Cannot increase memory without increasing number of parameters
- Need for compartmentalized memory
- Read/Write should be asynchronous

Memory Networks (MemNN)

- Class of Models with memory $\boldsymbol{m}$ - Array of objects $\boldsymbol{m_i}$



m_i

Each memory here is a dense vector

Four Components :

I - Input Feature Map : Input manipulation

G - Generalization : Memory Manipulation

O - Output Feature Map : Output representation generator

R - Response : Response Generator

Memory Networks, Weston et. al., ICLR 2015

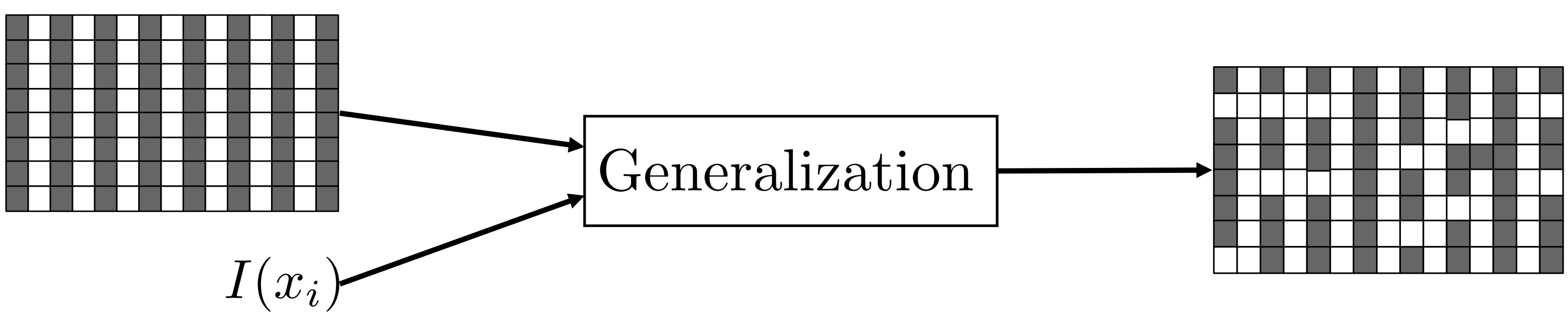
MemNN

1. Input Feature Map

- Imagine input as a sequence of sentences x_i

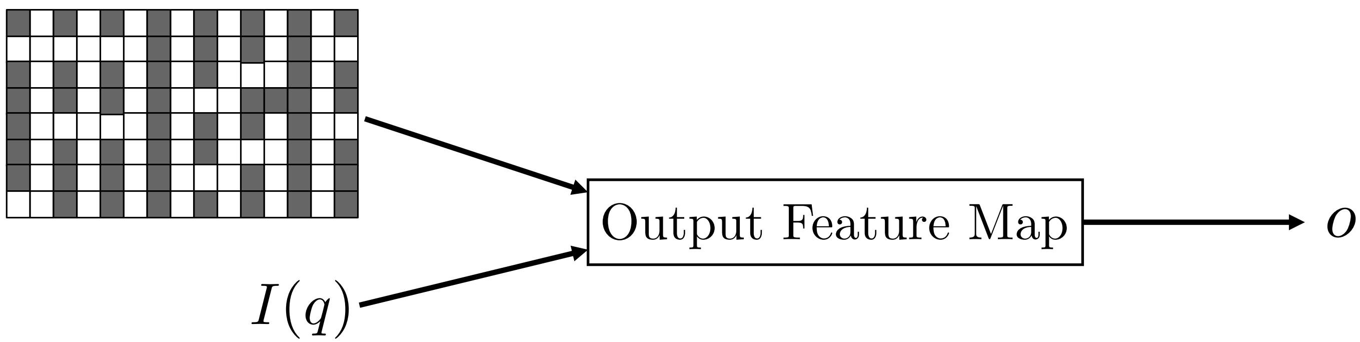
$$x_i \rightarrow \boxed{\text{Input Feature Map}} \rightarrow I(x_i)$$

2. Update Memories

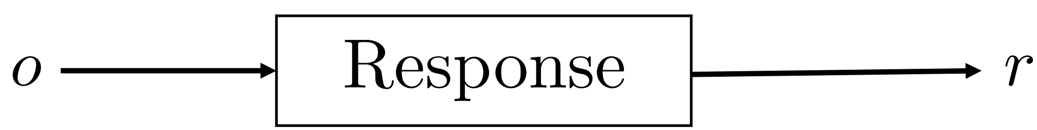


MemNN

3. Output Representation
   - Say if q is a question, compute output representation



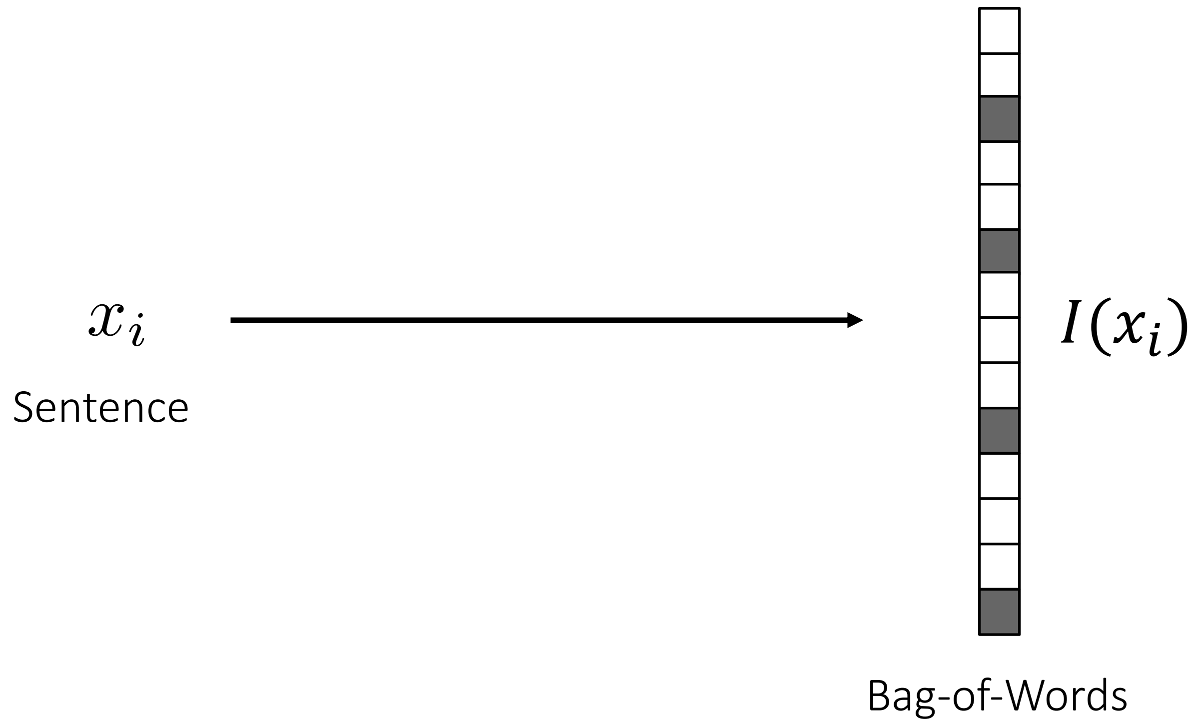
4. Generate Answer Response



Memory Networks, Weston et. al., ICLR 2015

Simple MemNN for Text

1. **Input Feature Map** - Bag-of-Words representation

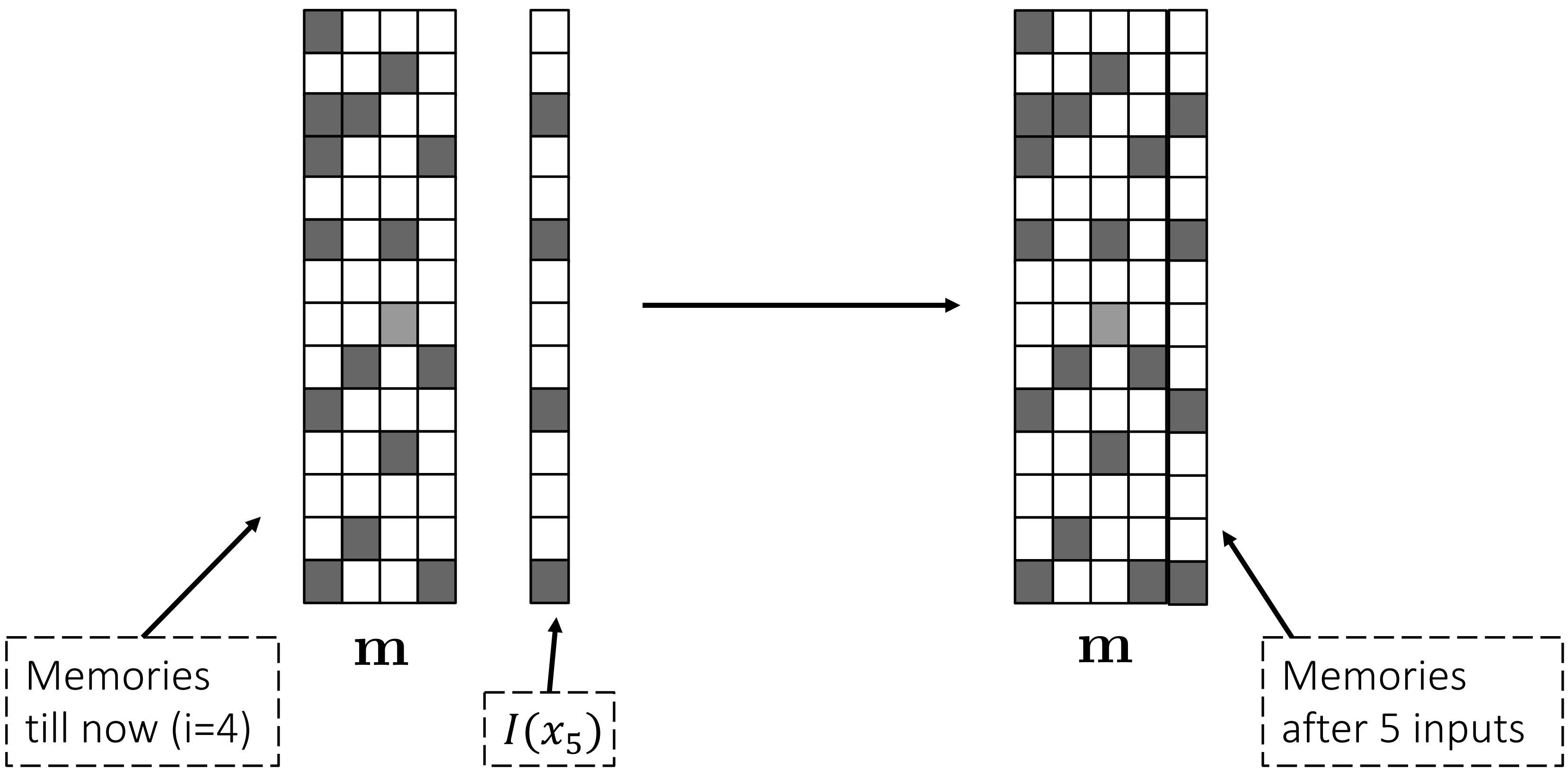


Memory Networks, Weston et. al., ICLR 2015

Simple MemNN for Text

2. **Generalization** : Store input in new memory

$$m_i = I(x_i)$$



Memory Networks, Weston et. al., ICLR 2015

Simple MemNN for Text

3. **Output:** Using $k = 2$ memory hops with query x

1[st] Max scoring memory index

$$o_1 = O_1(x, \mathbf{m}) = \underset{i=1,\dots,N}{argmax}\ s_O(I(x), m_i)$$

Score all memories against input

2[nd] Max scoring memory index

$$o_2 = O_2(x, \mathbf{m}) = \underset{i=1,\dots,N}{argmax}\ s_O([I(x), m_{o_1}], m_i)$$

Score all memories against input & o_1

4. **Response** - Single Word Answer

Max scoring word

$$r = argmax_{w \in W}\ s_R([I(x), m_{o_1}, m_{o_2}], w)$$

Score all words against query and 2 supporting memories

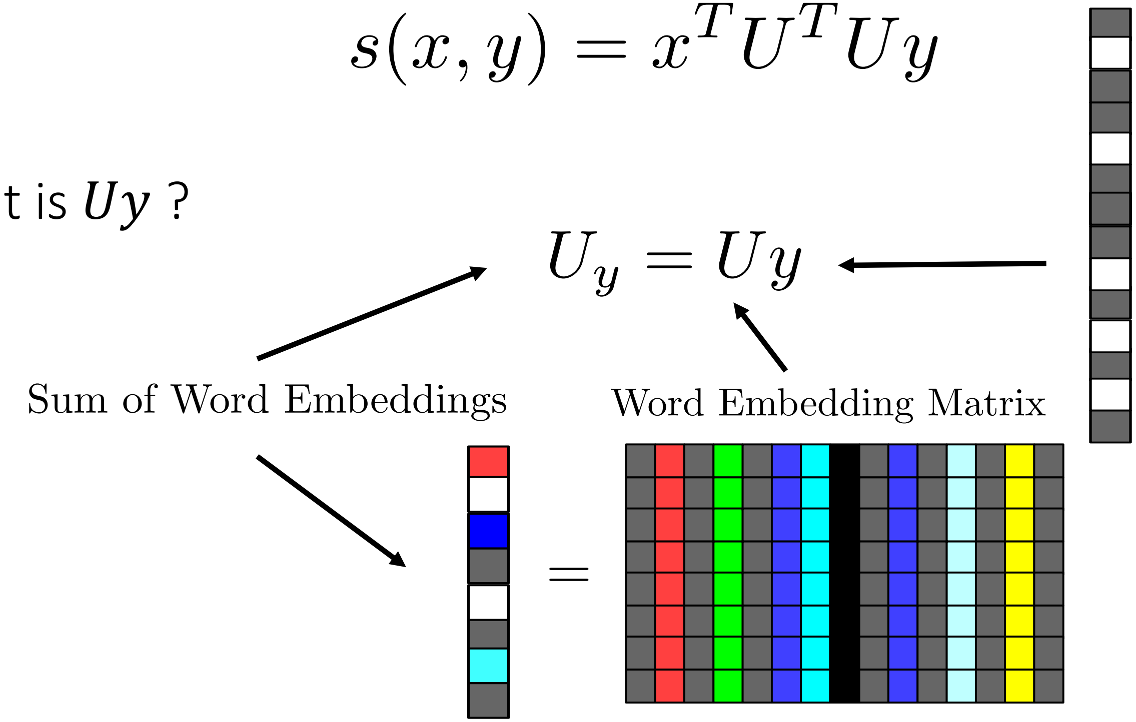
Memory Networks, Weston et. al., ICLR 2015

Scoring Function

- Scoring Function is an embedding model

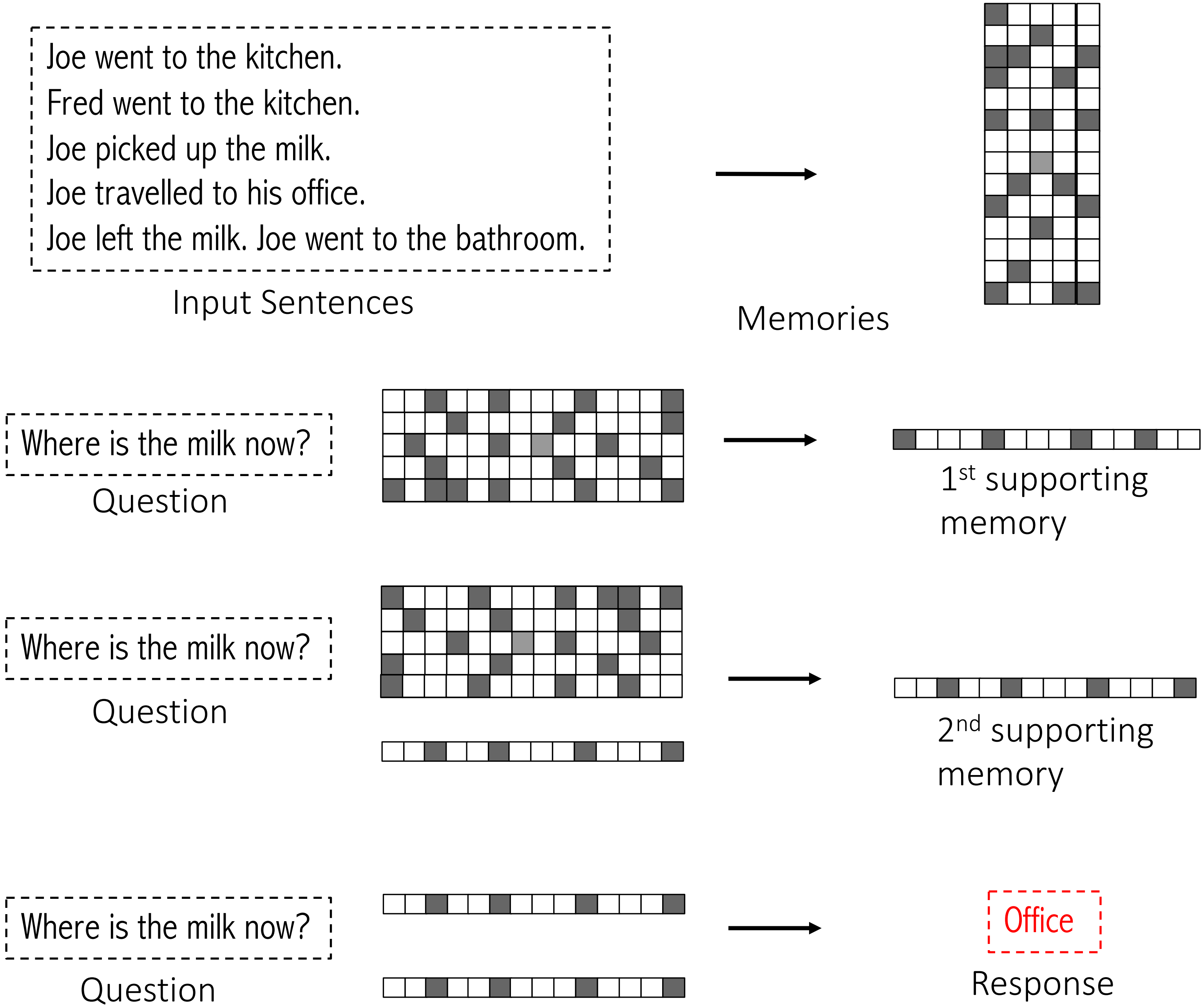
$$s(x, y) = x^T U^T U y$$

- What is $\boldsymbol{Uy}$?



Scoring Function is just dot-product between sum of word embeddings!!!

Memory Networks, Weston et. al., ICLR 2015



Memory Networks, Weston et. al., ICLR 2015

Training Objective

Score for true 1st memory

Score for a negative memory

$$\sum_{\bar{f} \neq \mathbf{m}_{o_1}} max(0, \gamma - \boxed{s_O(x, \mathbf{m}_{o_1})} + \boxed{s_O(x, \bar{f})}) \ +$$

Memory Networks, Weston et. al., ICLR 2015

Training Objective

Score for true 2nd memory

Score for a negative memory

$$\sum_{\bar{f} \neq \mathbf{m}_{o_1}} max(0, \gamma - s_O(x, \mathbf{m}_{o_1}) + s_O(x, \bar{f})) \;+$$

$$\sum_{\bar{f}' \neq \mathbf{m}_{o_2}} max(0, \gamma - \boxed{s_O([x, \mathbf{m}_{o_1}], \mathbf{m}_{o_2})} + \boxed{s_O([x, \mathbf{m}_{o_1}], \bar{f}')}) \;+$$

Memory Networks, Weston et. al., ICLR 2015

Training Objective

Score for true response

Score for a negative response

$$\sum_{\bar{f} \neq \mathbf{m}_{o_1}} max(0, \gamma - s_O(x, \mathbf{m}_{o_1}) + s_O(x, \bar{f})) \ +$$

$$\sum_{\bar{f}' \neq \mathbf{m}_{o_2}} max(0, \gamma - s_O([x, \mathbf{m}_{o_1}], \mathbf{m}_{o_2}) + s_O([x, \mathbf{m}_{o_1}], \bar{f}')) \ +$$

$$\sum_{\bar{r} \neq r} max(0, \gamma - \boxed{s_R([x, \mathbf{m}_{o_1}, \mathbf{m}_{o_2}], r)} + \boxed{s_R([x, \mathbf{m}_{o_1}, \mathbf{m}_{o_2}], \bar{r})})$$

Memory Networks, Weston et. al., ICLR 2015

Experiment

- Large – Scale QA
  - 14M Statements – *(subject, relation, object)*
  - Memory Hops; $k = 1$
  - Only re-ranked candidates from other system

Stored as memories

Output is highest scoring memory

Method	F1
Fader et. al. 2013	0.54
Bordes et. al. 2014b	0.73
Memory Networks (This work)	0.72

Why does Memory Network perform exactly as previous model?

Memory Networks, Weston et. al., ICLR 2015

Experiment

- Large – Scale QA
  - 14M Statements – *(subject, relation, object)*
  - Memory Hops; $k = 1$
  - Only re-ranked candidates from other system

Method	F1
Fader et. al. 2013	0.54
Bordes et. al 2014b	0.73
Memory Networks (This work)	0.72

USELESS EXPERIMENT

Why does Memory Networks not perform as well?

Useful Experiment

- Simulated World QA
  - 4 characters, 3 objects, 5 rooms
  - 7k statements, 3k questions for training and same for testing
  - Difficulty 1 (5) – Entity in question is mentioned in last 1 (5) sentences
  - For $k = 2$, annotation has intermediate best *memories* as well

Joe went to the garden then Fred picked up the milk; Joe moved to the bathroom and Fred dropped the milk, and then Dan moved to the living_room.
Where is Dan? A: living room I believe
Where is Joe? A: the bathroom
Fred moved to the bedroom and Joe went to the kitchen then Joe took the milk there and Dan journeyed to the bedroom; Joe discarded the milk.
Where is the milk now ? A: the milk is in the kitchen
Where is Dan now? A: I think he is in the bedroom
Joe took the milk there, after that Mike travelled to the office, then Joe went to the living_room, next Dan went back to the kitchen and Joe travelled to the office.
Where is Joe now? A: I think Joe is in the office

Memory Networks, Weston et. al., ICLR 2015

Limitations

- Simple BOW representation
- Simulated Question Answering dataset is too trivial
- Strong supervision i.e. for intermediate memories is needed

Memory Networks, Weston et. al., ICLR 2015

End-to-End Memory Networks (MemN2N)

- What if the annotation is:
  - Input sentences $x_1, x_2, \ldots, x_n$
  - Query q
  - Answer a

Joe went to the kitchen. Fred went to the kitchen.
Joe picked up the milk. Joe travelled to his office.
Joe left the milk. Joe went to the bathroom.

Where is the milk now?

Office

- Model performs by:
  - Generating memories from inputs
  - Transforming query into suitable representation
  - Process query and memories jointly using multiple hops to produce the answer
  - Backpropagate through the whole procedure

End-To-End Memory Networks, Sukhbaatar et. al., NIPS 2015

MemN2N

1. Convert input to memories $x_i \rightarrow m_i$

$$m_i = Ax_i$$

Sum of word-embeddings

Word-Embedding Matrix

BOW input

2. Transform query q into same representation space

$$u = Bq$$

3. Output Vectors $x_i \rightarrow c_i$

$$c_i = Cx_i$$

End-To-End Memory Networks, Sukhbaatar et. al., NIPS 2015

MemN2N

3. Scoring memories against query

$$p_i = \text{Softmax}(u^T m_i)$$

Score for input/memory

Memories

Query (transformed)

4. Generate output

$$o = \sum_i p_i c_i$$

Weighted average of all inputs (transformed)

End-To-End Memory Networks, Sukhbaatar et. al., NIPS 2015

MemN2N

5. Generating Response

$$\hat{a} = \text{Softmax}(W(u + o))$$

Distribution over response words

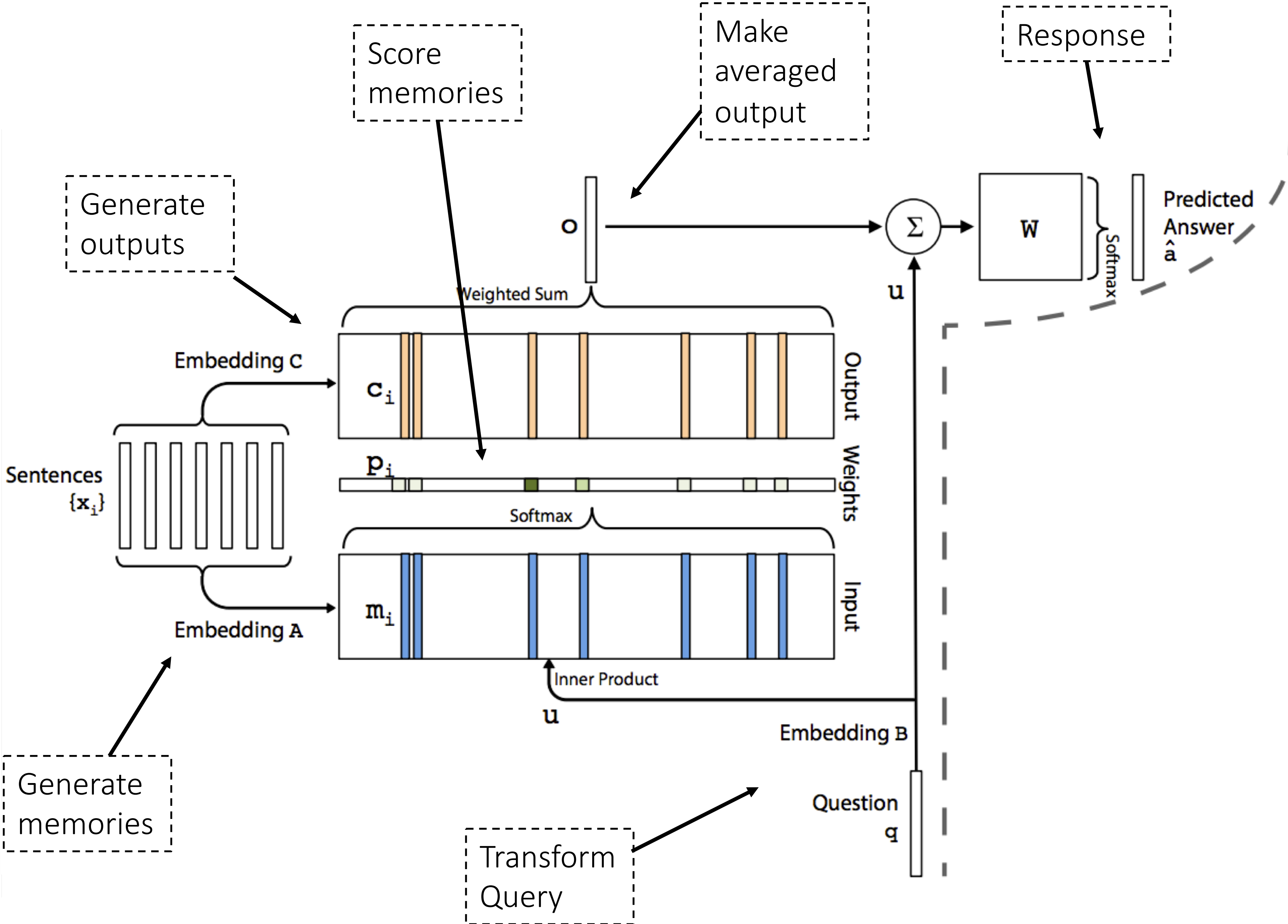
Query

Averaged-output

Training Objective – Maximum Likelihood / Cross Entropy

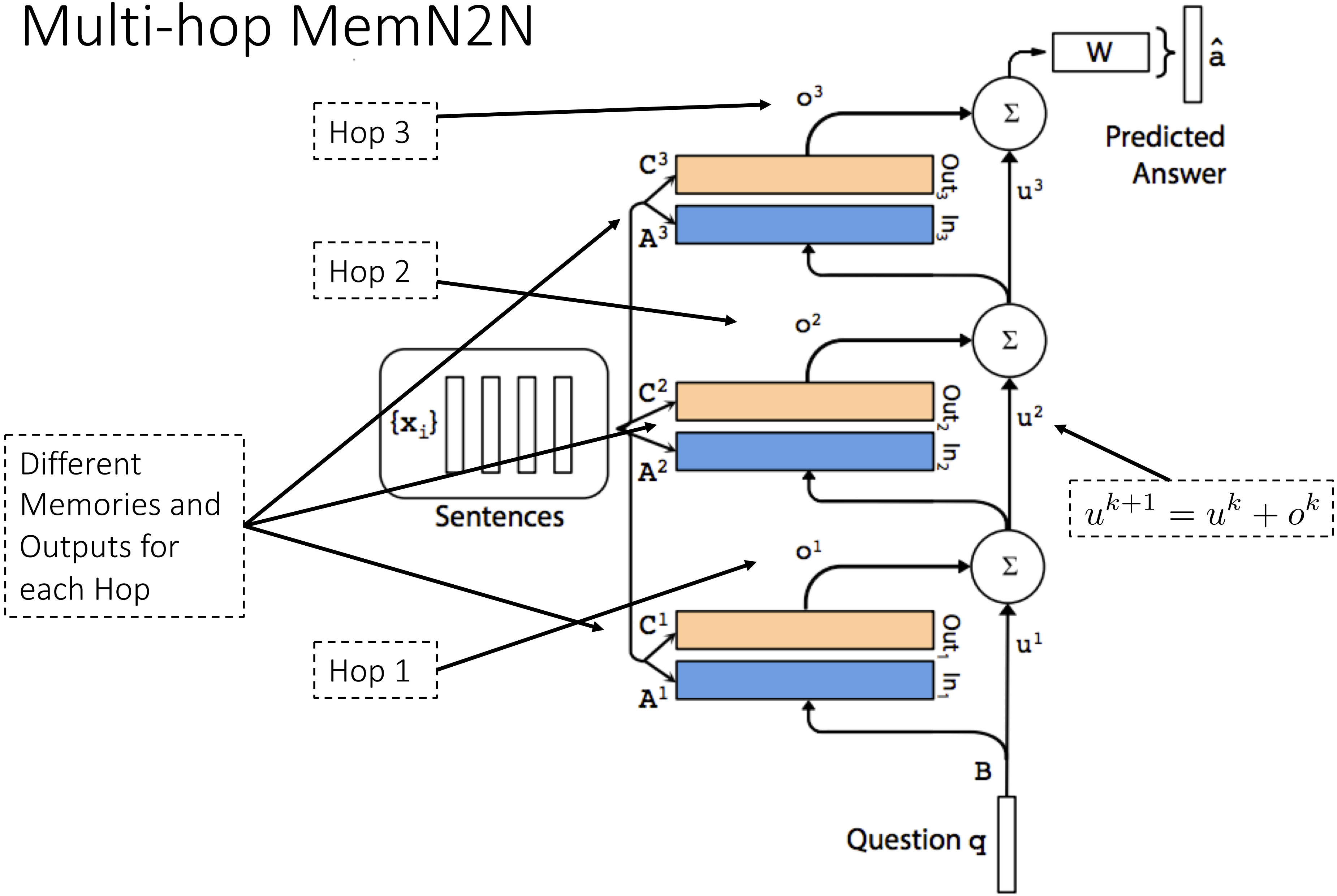
$$\hat{\Theta} = \text{argmax} \sum_{s=1}^{N} \log P(\hat{a}_s)$$

End-To-End Memory Networks, Sukhbaatar et. al., NIPS 2015



End-To-End Memory Networks, Sukhbaatar et. al., NIPS 2015

Multi-hop MemN2N



End-To-End Memory Networks, Sukhbaatar et. al., NIPS 2015

Experiments

- Simulated World QA
  - 20 Tasks from bAbI dataset - 1K and 10K instances per task
  - Vocabulary = 177 words only!!!!!
  - 60 epochs
  - Learning Rate annealing
  - Linear Start with different learning rate
  - *"Model diverged very often, hence trained multiple models"*

End-To-End Memory Networks, Sukhbaatar et. al., NIPS 2015

Story (1: 1 supporting fact)	Support	Hop 1	Hop 2	Hop 3
Daniel went to the bathroom.		0.00	0.00	0.03
Mary travelled to the hallway.		0.00	0.00	0.00
John went to the bedroom.		0.37	0.02	0.00
John travelled to the bathroom.	yes	0.60	0.98	0.96
Mary went to the office.		0.01	0.00	0.00
Where is John? Answer: bathroom Prediction: bathroom				

Story (16: basic induction)	Support	Hop 1	Hop 2	Hop 3
Brian is a frog.	yes	0.00	0.98	0.00
Lily is gray.		0.07	0.00	0.00
Brian is yellow.	yes	0.07	0.00	1.00
Julius is green.		0.06	0.00	0.00
Greg is a frog.	yes	0.76	0.02	0.00
What color is Greg? Answer: yellow Prediction: yellow				

	MemNN	MemN2N
Error % (1k)	6.7	12.4
Error % (10k)	3.2	7.5

End-To-End Memory Networks, Sukhbaatar et. al., NIPS 2015

Movie Trivia Time!

- Which was *Stanley Kubricks*'s first movie?

  Fear and Desire

- When did *2001:A Space Odyssey* release?

  1968

- After *The Shining*, which movie did its director direct?

  Full Metal Jacket

Subject | Relation | Object

(2001:a_space_odyssey, *directed_by*, stanley_kubrick)

(fear_and_dark, *directed_by*, stanley_kubrick)

…

(fear_and_dark, *released_in*, 1953)

(full_metal_jacket, *released_in*, 1987)

…

(2001:a_space_odyssey, *released_in*, 1968)

…

(the_shining, *directed_by*, stanley_kubrick)

…

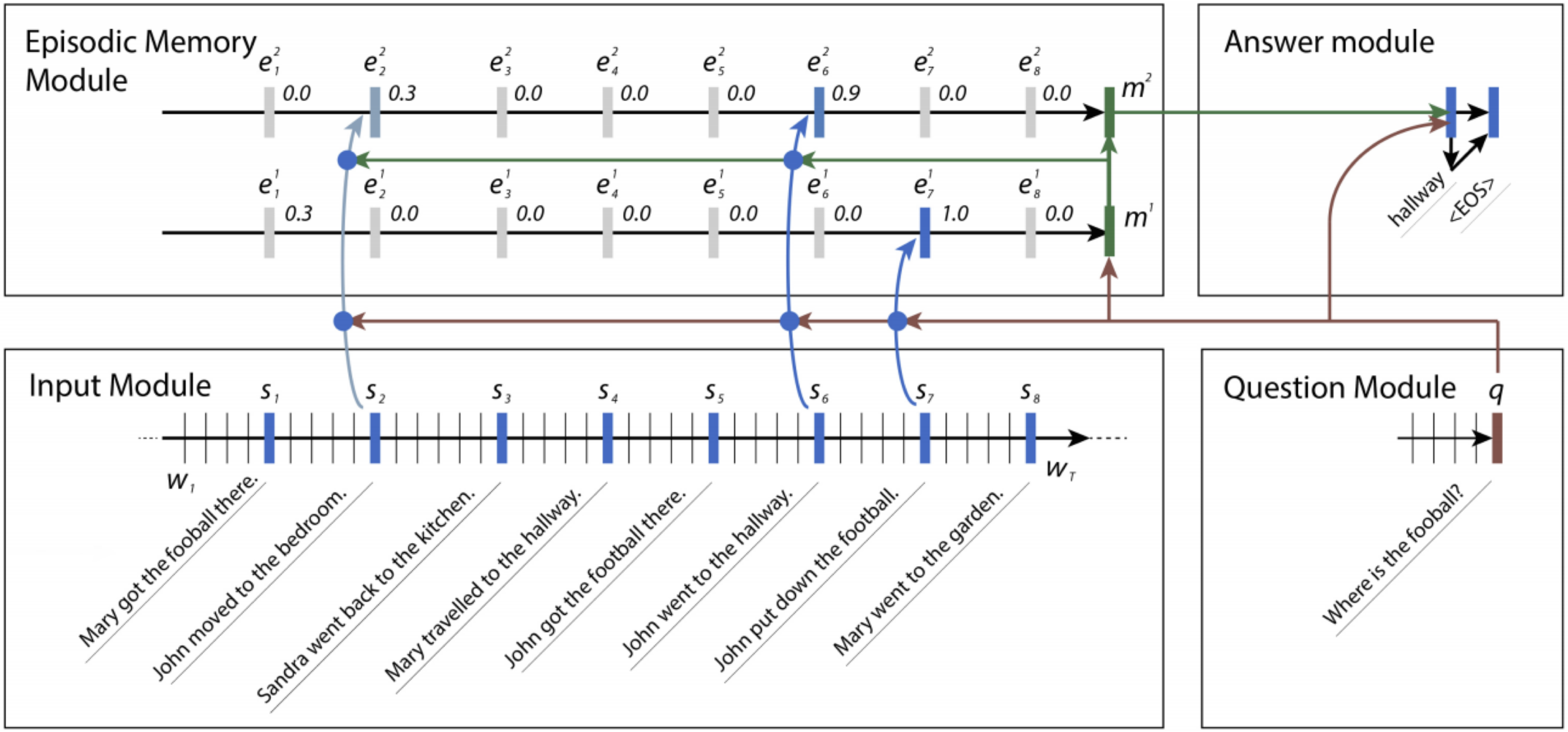
(AI:artificial_intelligence, *written_by*, stanley_kubrick)

Knowledge Base

CNN : Computer Vision :: ___RNN___ : NLP

Key-Value Memory Networks for Directly Reading Documents, Miller et. al., EMNLP 2016

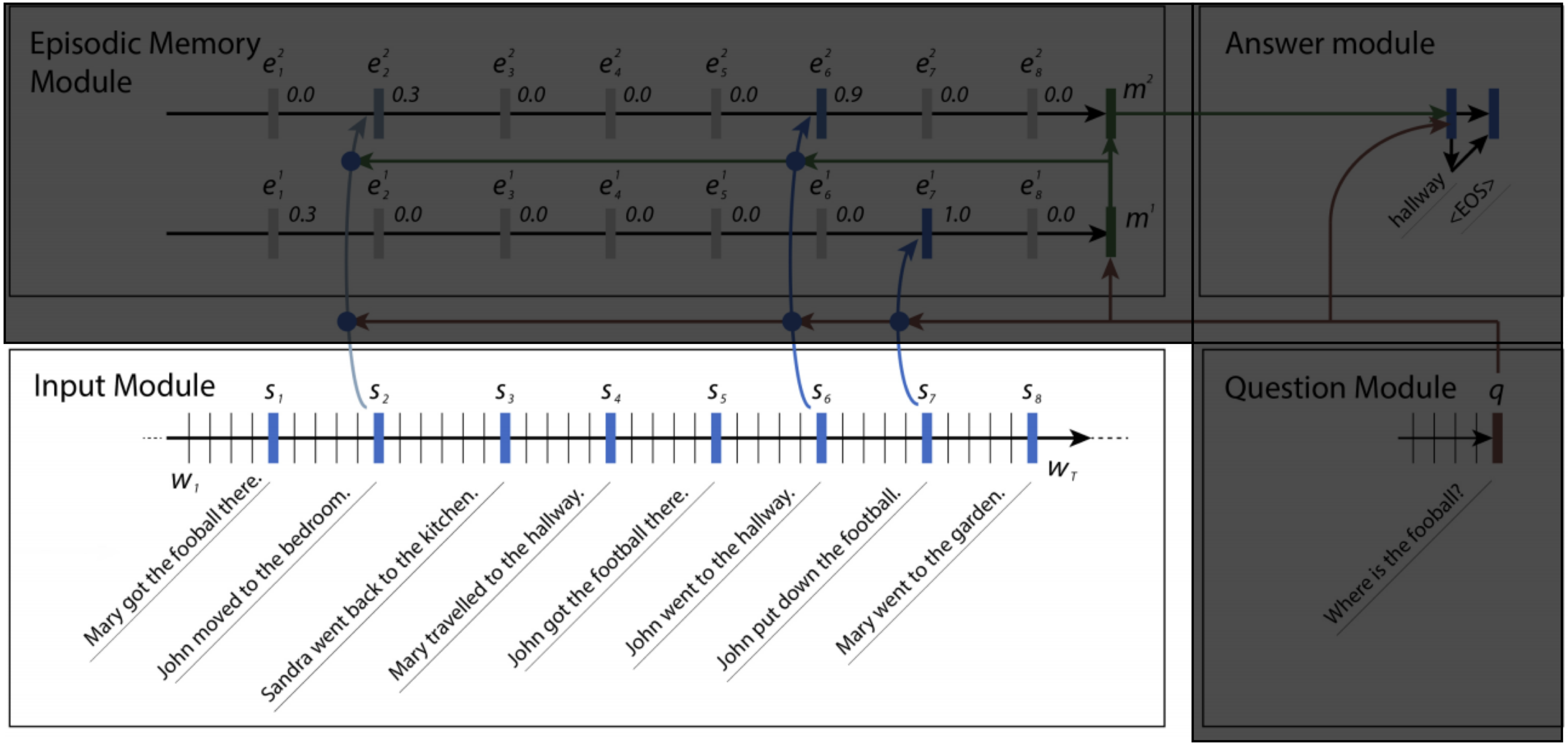
Dynamic Memory Networks – **The Beast**



Use RNNs, specifically GRUs for every module

Ask Me Anything: Dynamic Memory Networks for Natural Language Processing, Kumar et. al. ICML 2016

DMN

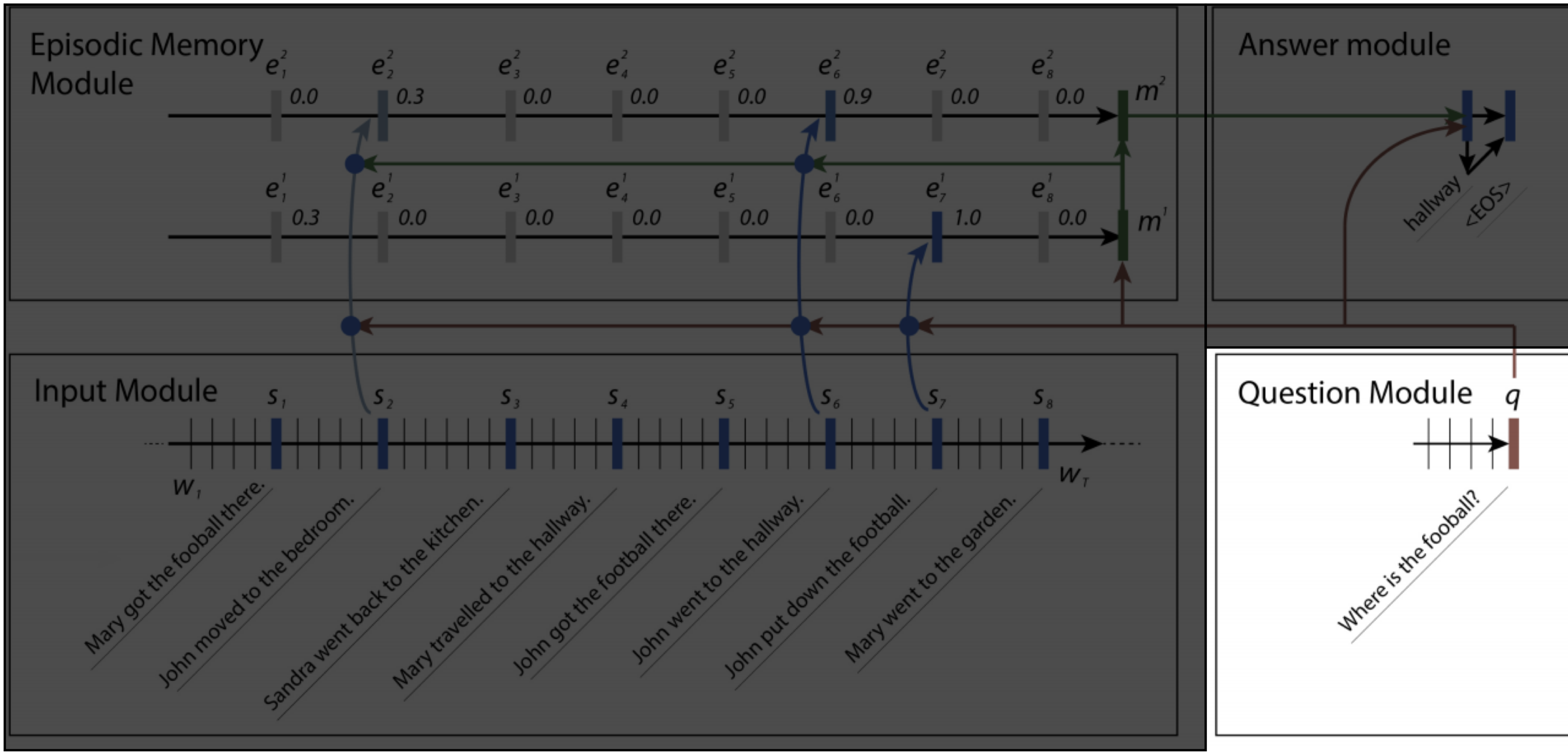


Final GRU Output for t^{th} sentence

$$c_t = \mathrm{GRU}(w_t^i, c_t^{i-1})$$

Ask Me Anything: Dynamic Memory Networks for Natural Language Processing, Kumar et. al. ICML 2016

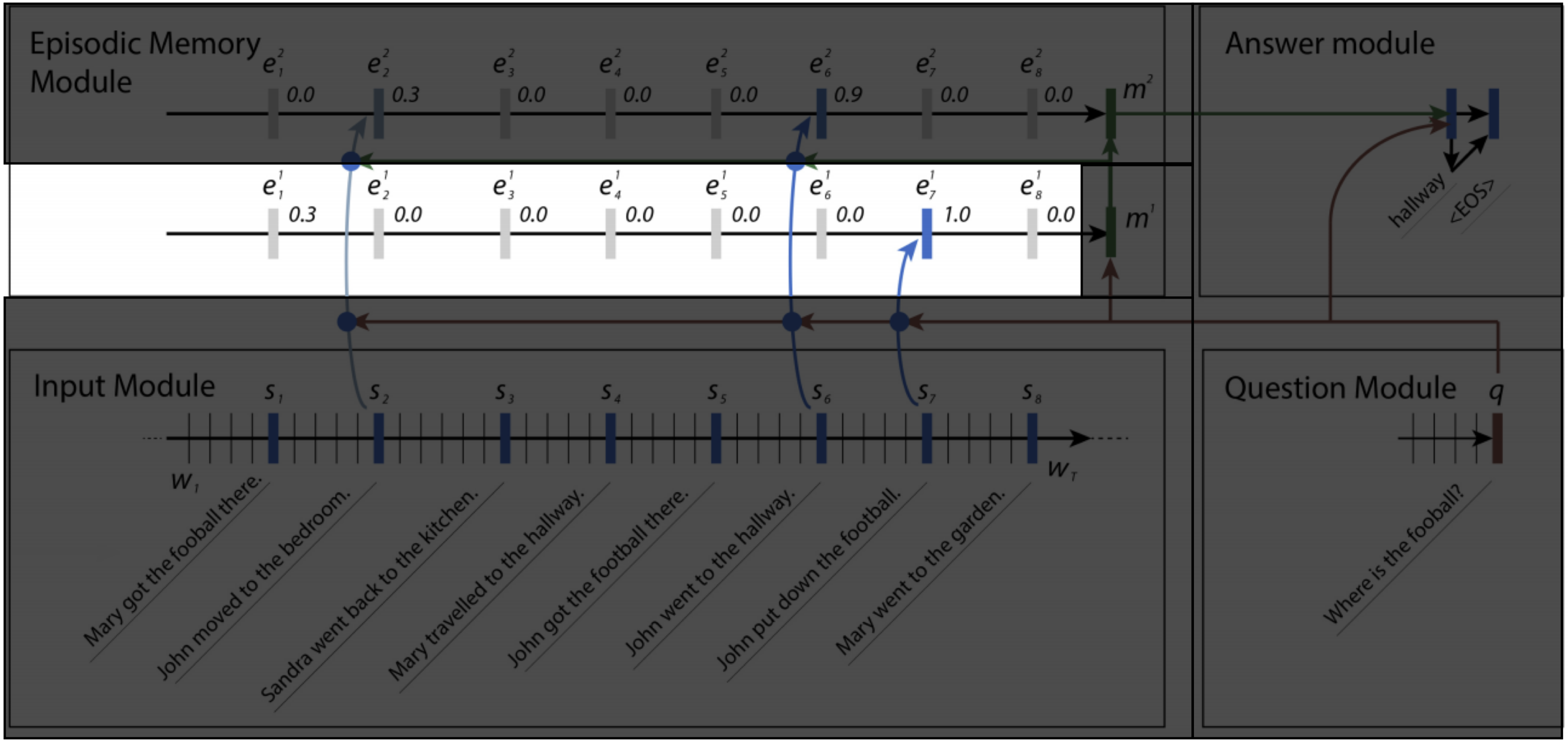
DMN



$$q = \mathrm{GRU}(q_w^i, q^{i-1})$$

Ask Me Anything: Dynamic Memory Networks for Natural Language Processing, Kumar et. al. ICML 2016

DMN



$$\boxed{Hop = i}$$

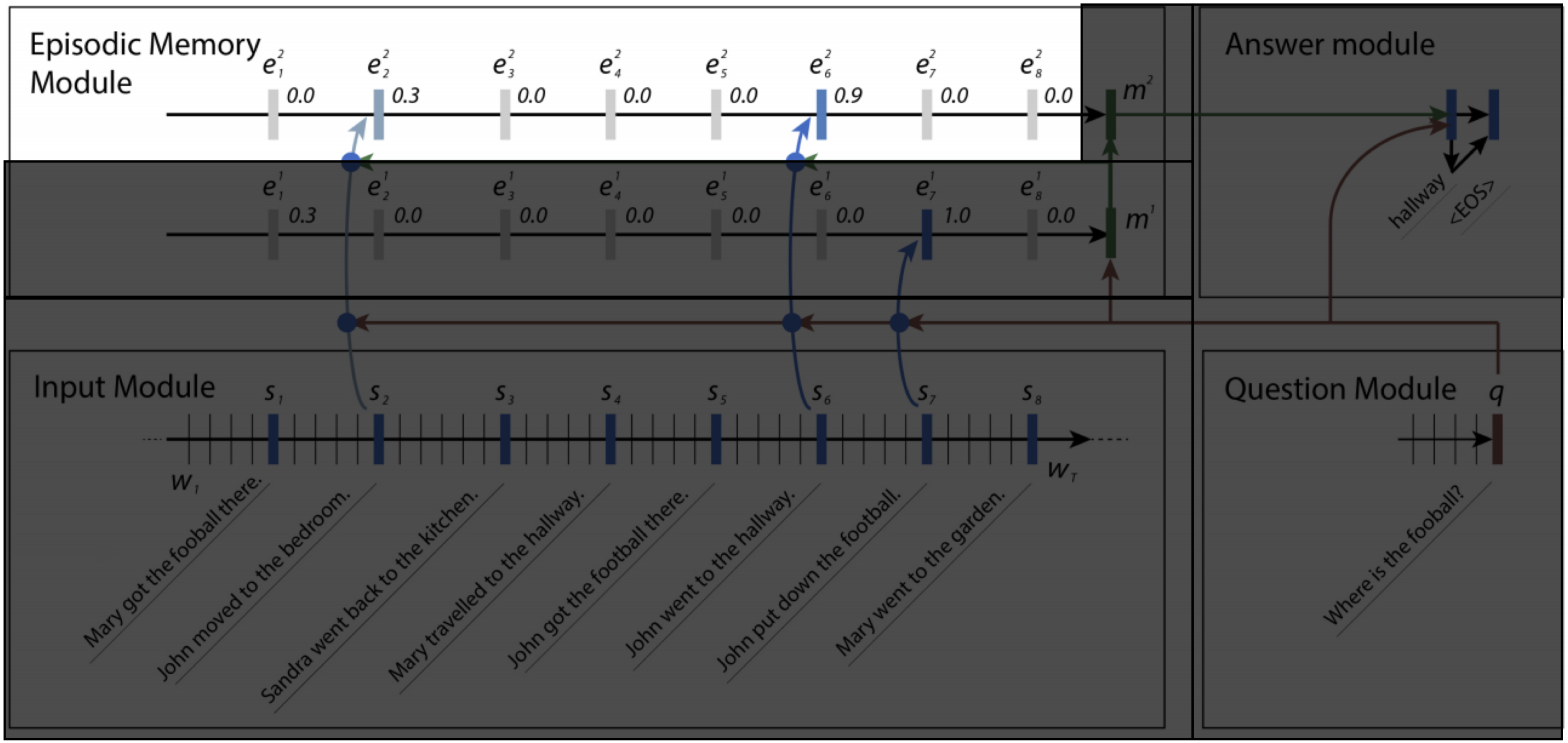
$$\boxed{i = 1}$$

$$h_t^i = g_t^i \mathrm{GRU}(c_t, h_{t-1}^i) + (1 - g_t^i) h_{t-1}^i$$

$$e^i = h_{T_C}^i$$

Ask Me Anything: Dynamic Memory Networks for Natural Language Processing, Kumar et. al. ICML 2016

DMN



$Hop = i$

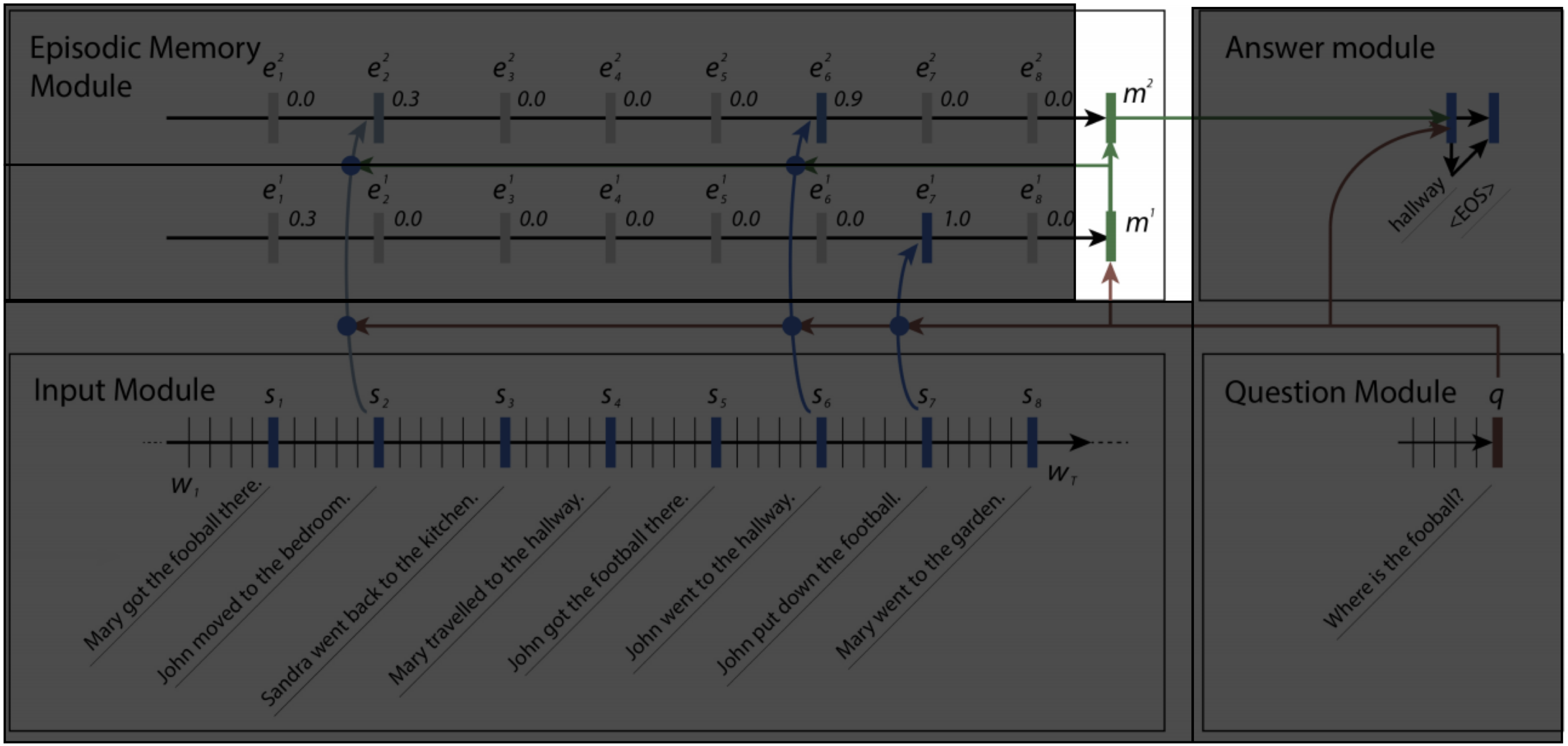
$i = 2$

$$h_t^i = g_t^i \mathrm{GRU}(c_t, h_{t-1}^i) + (1 - g_t^i) h_{t-1}^i$$

$$e^i = h_{T_C}^i$$

Ask Me Anything: Dynamic Memory Networks for Natural Language Processing, Kumar et. al. ICML 2016

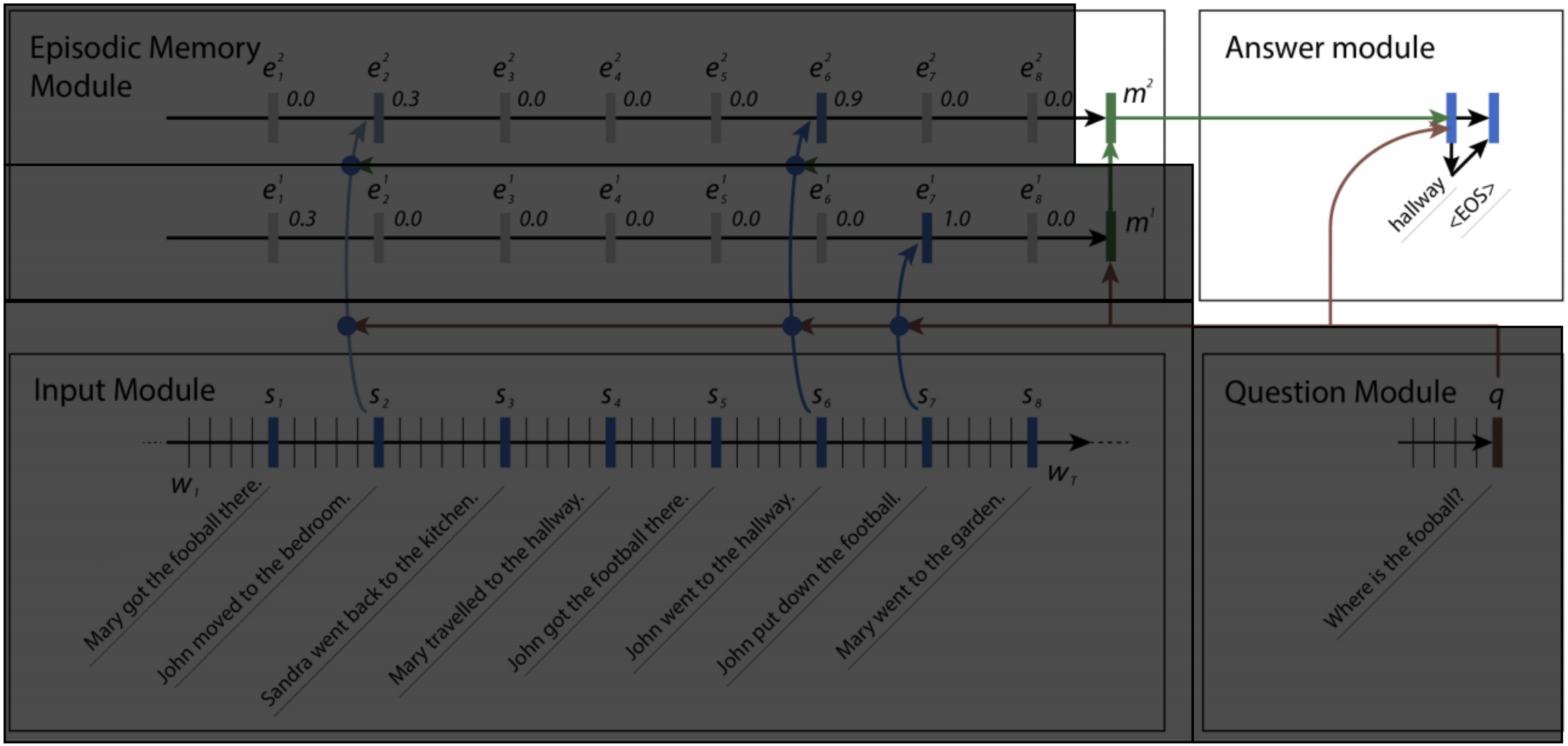
DMN



$$m^i = \mathrm{GRU}(e^i, m^{i-1})$$

Ask Me Anything: Dynamic Memory Networks for Natural Language Processing, Kumar et. al. ICML 2016

DMN

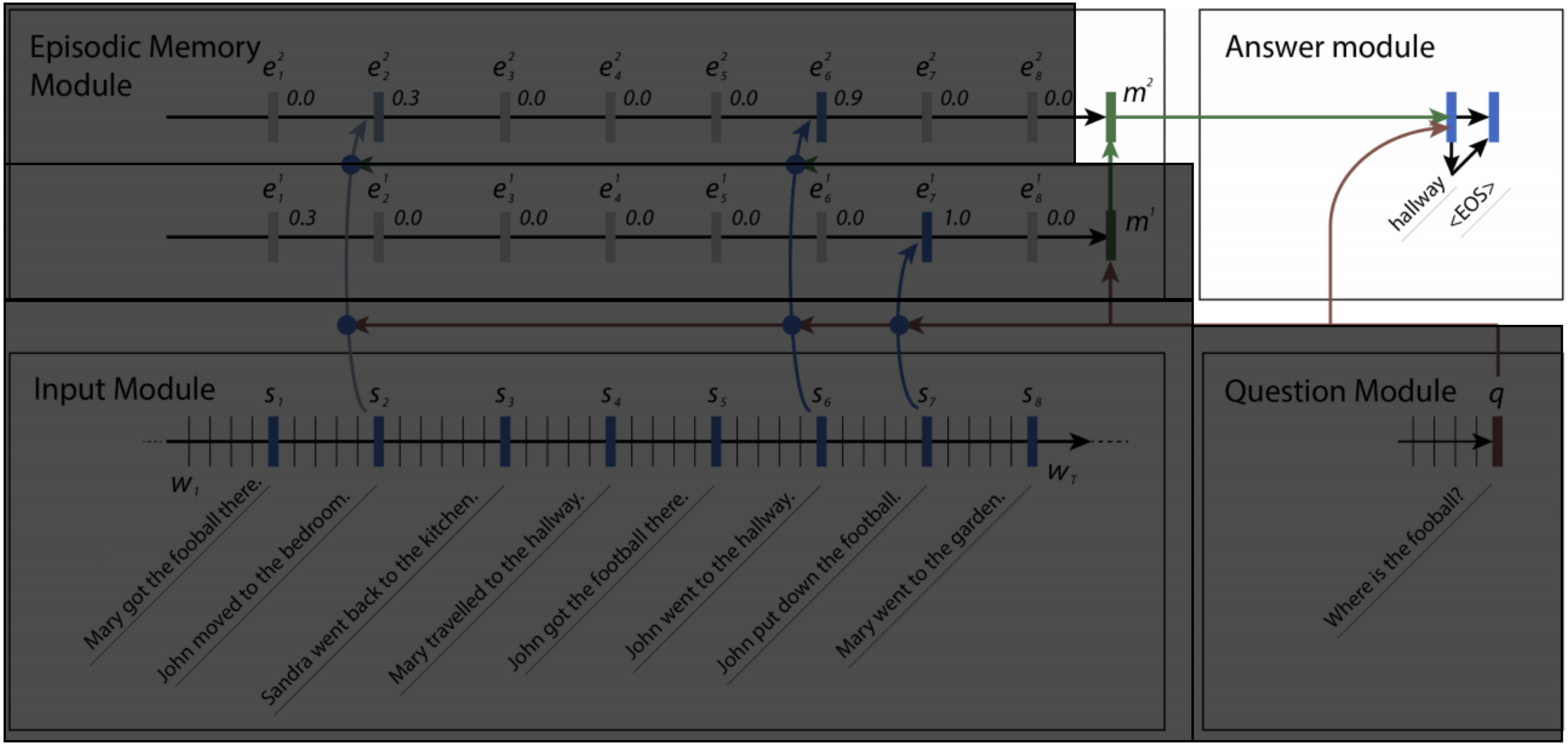


$$y_t = \mathrm{Softmax}(W^{(a)} \alpha_t) \qquad \alpha_0 = m^{T_m}$$

$$\alpha_t = \mathrm{GRU}([y_{t-1}, q], \alpha_{t-1})$$

Ask Me Anything: Dynamic Memory Networks for Natural Language Processing, Kumar et. al. ICML 2016

DMN



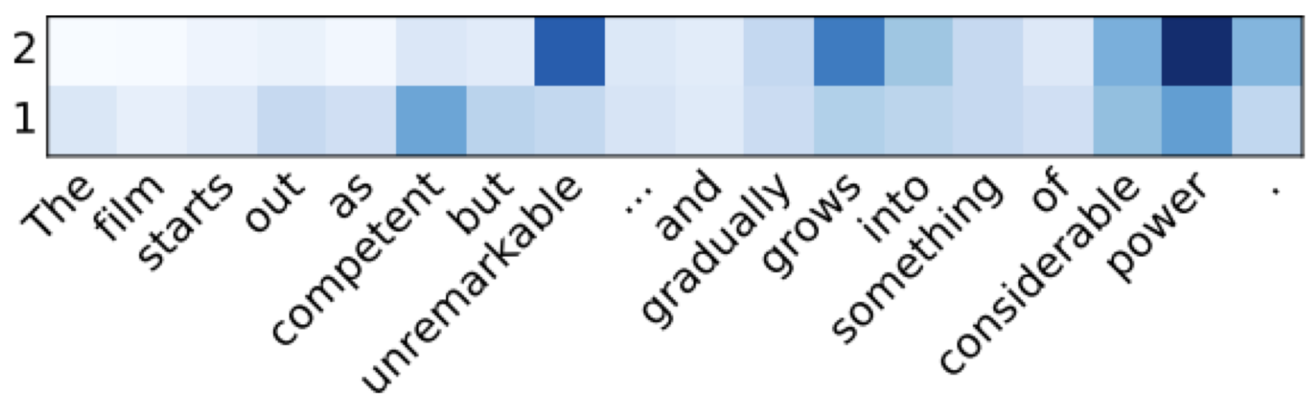
How many GRUs were used with 2 hops?

Ask Me Anything: Dynamic Memory Networks for Natural Language Processing, Kumar et. al. ICML 2016

DMN – Qualitative Results

Question: Where was Mary before the Bedroom?
Answer: Cinema.

Facts	Episode 1	Episode 2	Episode 3
Yesterday Julie traveled to the school.			
Yesterday Marie went to the cinema.			
This morning Julie traveled to the kitchen.			
Bill went back to the cinema yesterday.			
Mary went to the bedroom this morning.			
Julie went back to the bedroom this afternoon.			
[done reading]			



Ask Me Anything: Dynamic Memory Networks for Natural Language Processing, Kumar et. al. ICML 2016