Memory and Neural Networks

Slides were drawn in part from Hugo Larochelle and in part from the EMNLP 2017 tutorial of Caglar Gulcehre and Sarath Chandar

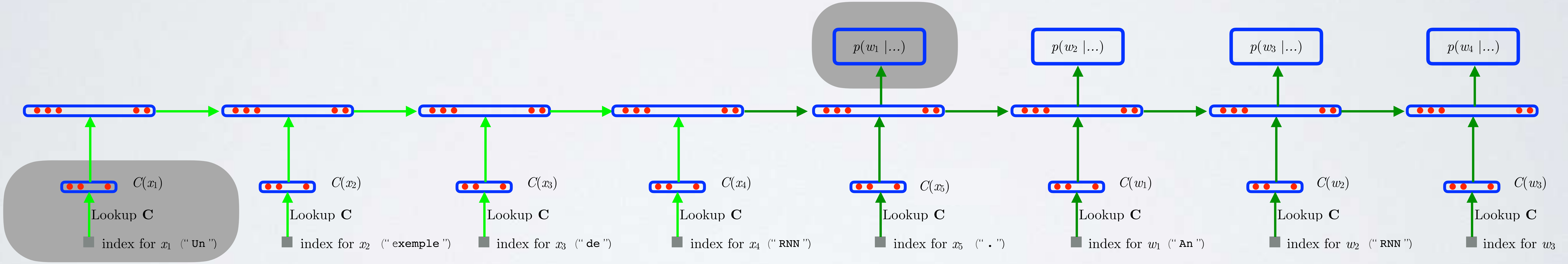
SEQUENCE TO SEQUENCE LEARNING

REMINDER

Topics: sequence to sequence (Seq2Seq) learning

- View of RNN unrolled through time
  - example: $\mathbf{w} = [\text{``}\,\texttt{An}\,\text{''},\ \text{``}\,\texttt{RNN}\,\text{''},\ \text{``}\,\texttt{example}\,\text{''},\ \text{``}\,\texttt{.}\,\text{''}]\ (T = 4)$

    $\mathbf{x} = [\text{``}\,\texttt{Un}\,\text{''},\ \text{``}\,\texttt{exemple}\,\text{''},\ \text{``}\,\texttt{de}\,\text{''},\ \text{``}\,\texttt{RNN}\,\text{''},\ \text{``}\,\texttt{.}\,\text{''}]\ (T_{\mathbf{x}} = 5)$



Capturing long-term dependencies is crucial

MEMORY AND ATTENTION

Topics: memory and attention

- Departs from the RNN representation of $\mathbf{x}$ using 2 concepts

1. Represent input $\mathbf{x}$ as a structured **memory** matrix instead of a single vector

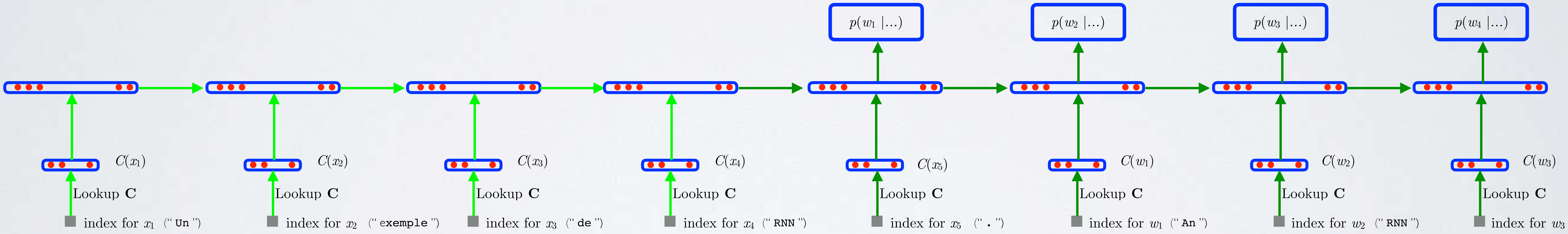
2. When predicting each output w_t, **attend** to most relevant location in memory

SEQUENCE TO SEQUENCE LEARNING

Topics: Seq2Seq with memory and attention

- View of RNN with memory/attention unrolled through time
  - example: $\mathbf{w} = [\text{``}\,\texttt{An}\,\text{''}, \text{``}\,\texttt{RNN}\,\text{''}, \text{``}\,\texttt{example}\,\text{''}, \text{``}\,\texttt{.}\,\text{''}]\ (T = 4)$
    $\mathbf{x} = [\text{``}\,\texttt{Un}\,\text{''}, \text{``}\,\texttt{exemple}\,\text{''}, \text{``}\,\texttt{de}\,\text{''}, \text{``}\,\texttt{RNN}\,\text{''}, \text{``}\,\texttt{.}\,\text{''}]\ (T_{\mathbf{x}} = 5)$

Neural Machine Translation by Jointly Learning to Align and Translate
Bahdanau, Cho and Bengio, ICLR 2015



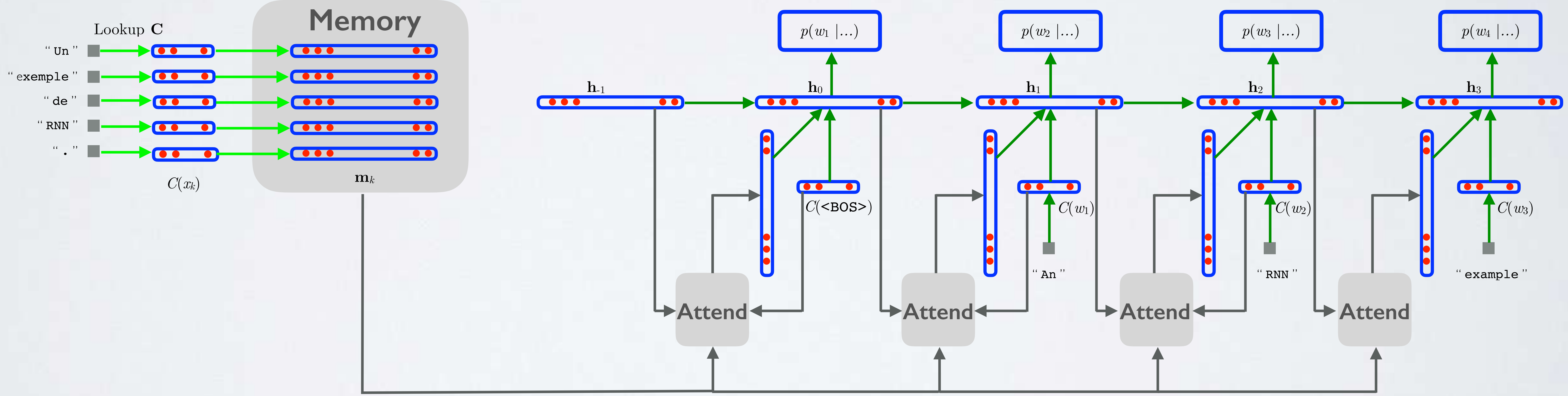
SEQUENCE TO SEQUENCE LEARNING

Topics: Seq2Seq with memory and attention

- View of RNN with memory/attention unrolled through time
  - example: $\mathbf{w} = [\text{``An''}, \text{``RNN''}, \text{``example''}, \text{``.''}]\ (T = 4)$

    $\mathbf{x} = [\text{``Un''}, \text{``exemple''}, \text{``de''}, \text{``RNN''}, \text{``.''}]\ (T_{\mathbf{x}} = 5)$

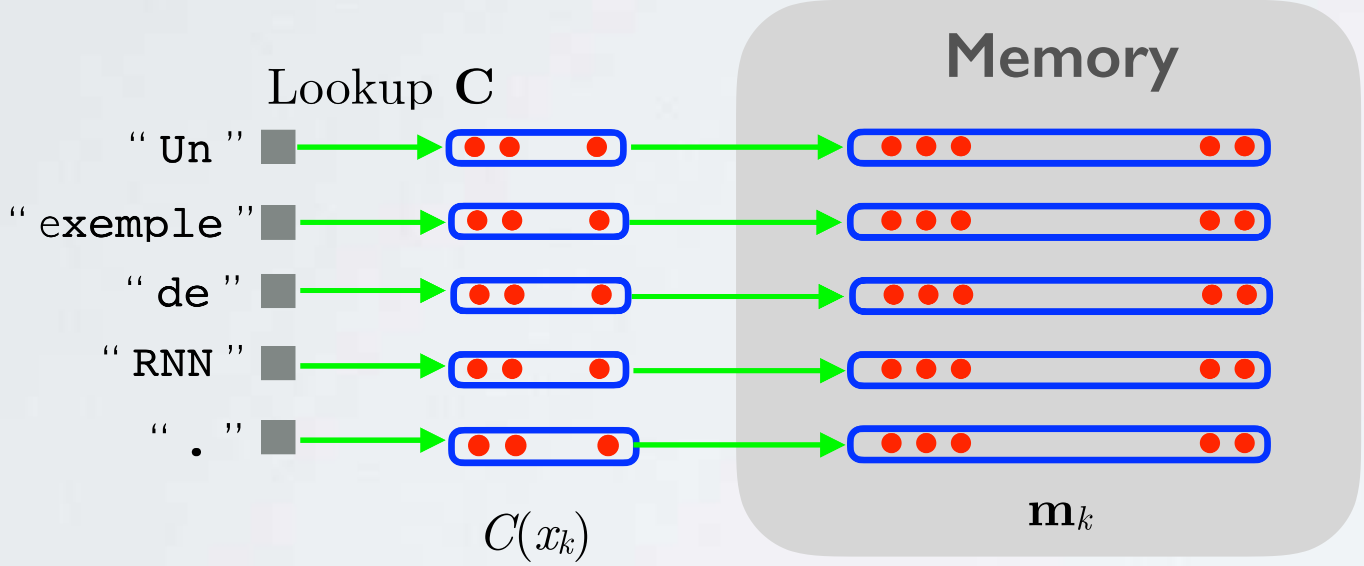
Neural Machine Translation by Jointly Learning to Align and Translate
Bahdanau, Cho and Bengio, ICLR 2015



MEMORY

Topics: memory

- Memory is just a list of vectors $\{\mathbf{m}_1, \ldots, \mathbf{m}_{T\mathbf{x}}\}$
  - by stacking each as a row vector, can be thought of as a matrix

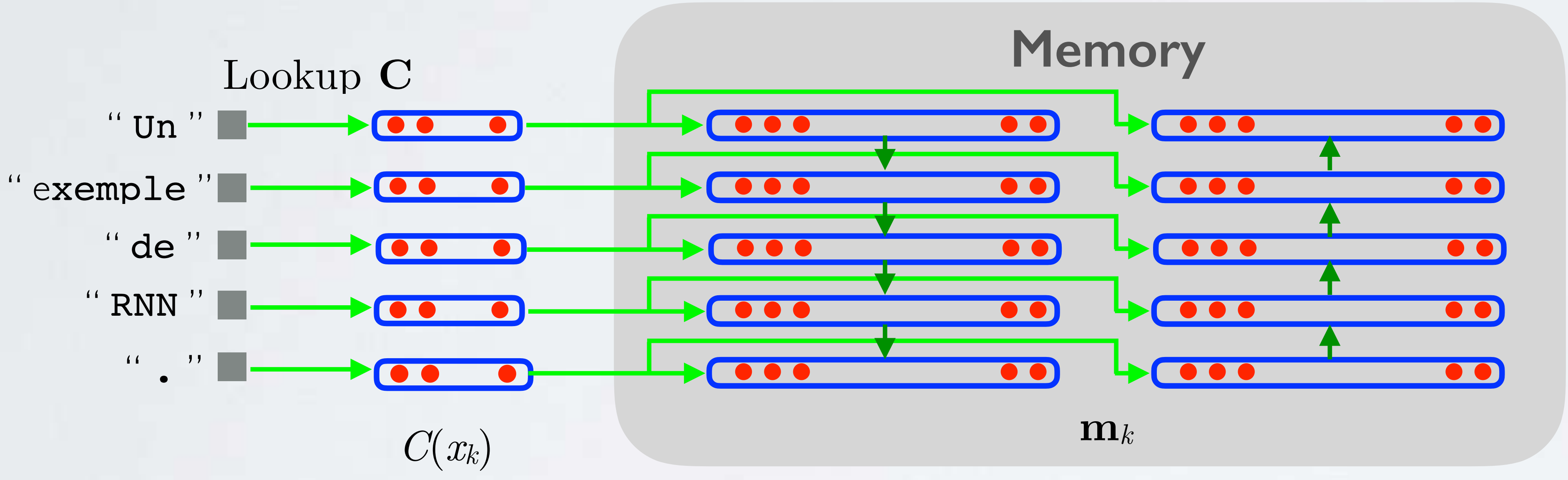


- Vectors $\mathbf{m}_k$ can be arbitrarily complex
  - can be the word representations ($\mathbf{m}_k = C(x_k)$)
  - can be the concatenated hidden layers of a bidirectional LSTM

MEMORY

Topics: memory

- Memory is just a list of vectors $\{\mathbf{m}_1, \ldots, \mathbf{m}_{T\mathbf{x}}\}$
    - by stacking each as a row vector, can be thought of as a matrix

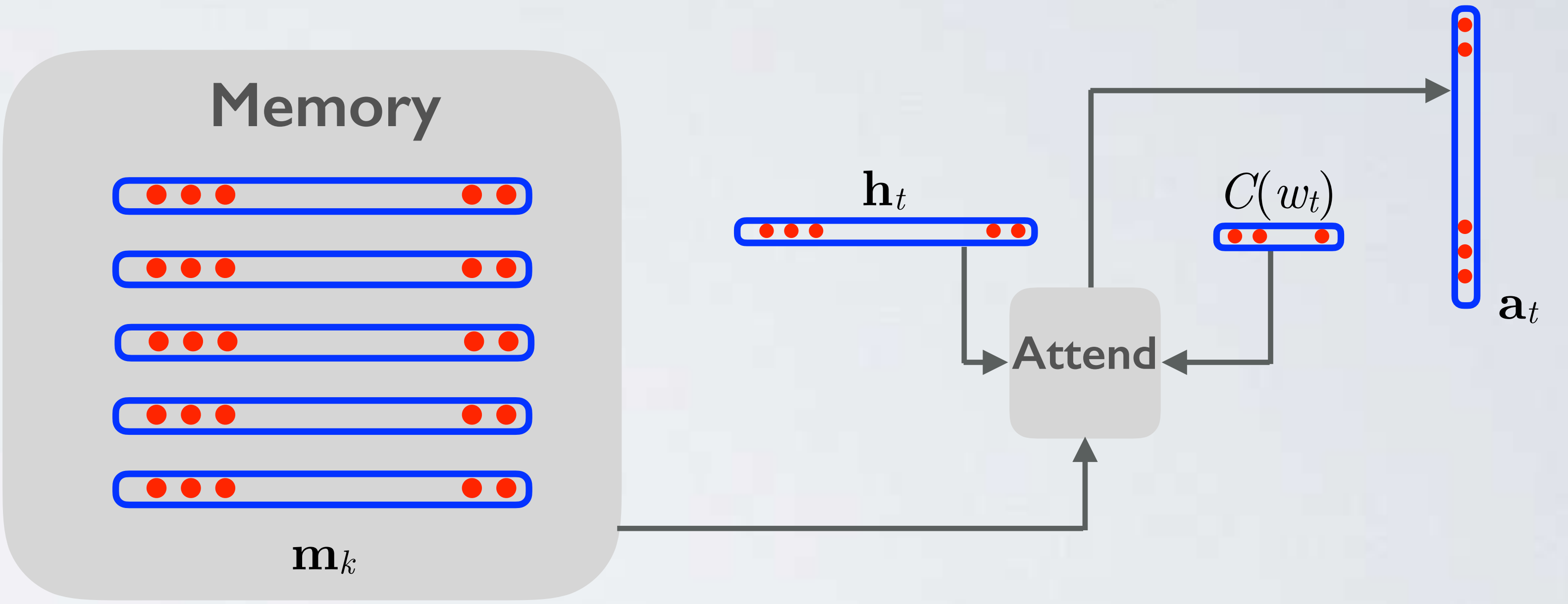


- Vectors $\mathbf{m}_k$ can be arbitrarily complex
    - can be the word representations ($\mathbf{m}_k = C(x_k)$)
    - can be the concatenated hidden layers of a bidirectional LSTM

ATTENTION

Topics: attention

- The attention mechanism:
  - uses the concatenation $C(w_t)$ of $\mathbf{h}_t$ and as a **query**
  - **compares** the query with all memory vectors $\mathbf{m}_k$
  - returns a read vector $\mathbf{a}_t$ from the memory



- A popular form for the attention module is one that
  - computes **attention (*addressing*) weights** $\alpha_{t,k}$:

$$\alpha_{t,k} = \frac{\exp(s(\mathbf{m}_k, [C(w_t), \mathbf{h}_t])}{\sum_{k'=1}^{T_{\mathbf{x}}} \exp(s(\mathbf{m}_{k'}, [C(w_t), \mathbf{h}_t])}$$

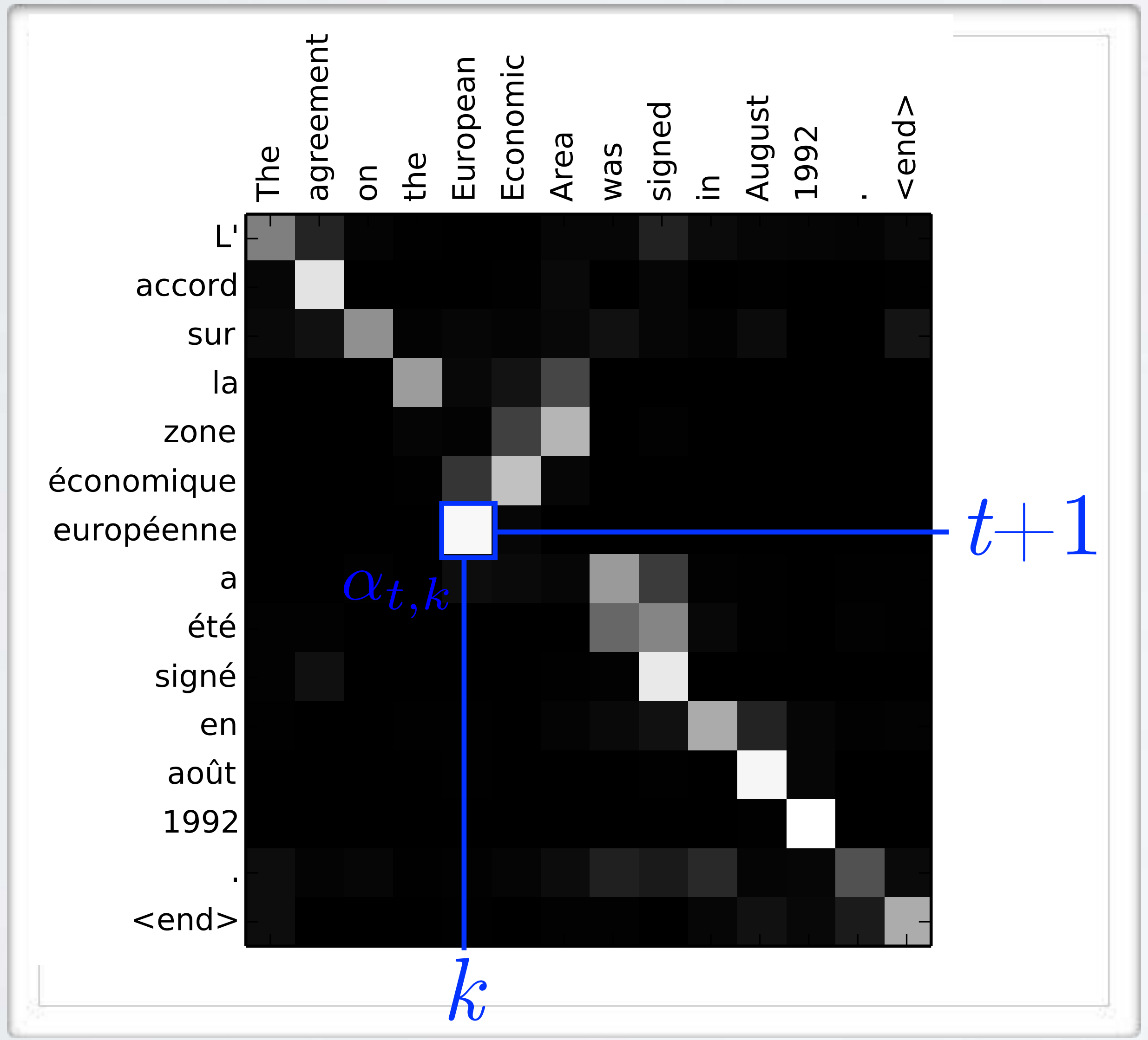
  where s could be a small neural network or $s(\mathbf{m},\mathbf{q}) = \mathbf{m}^\top \mathbf{U}\ \mathbf{q}$

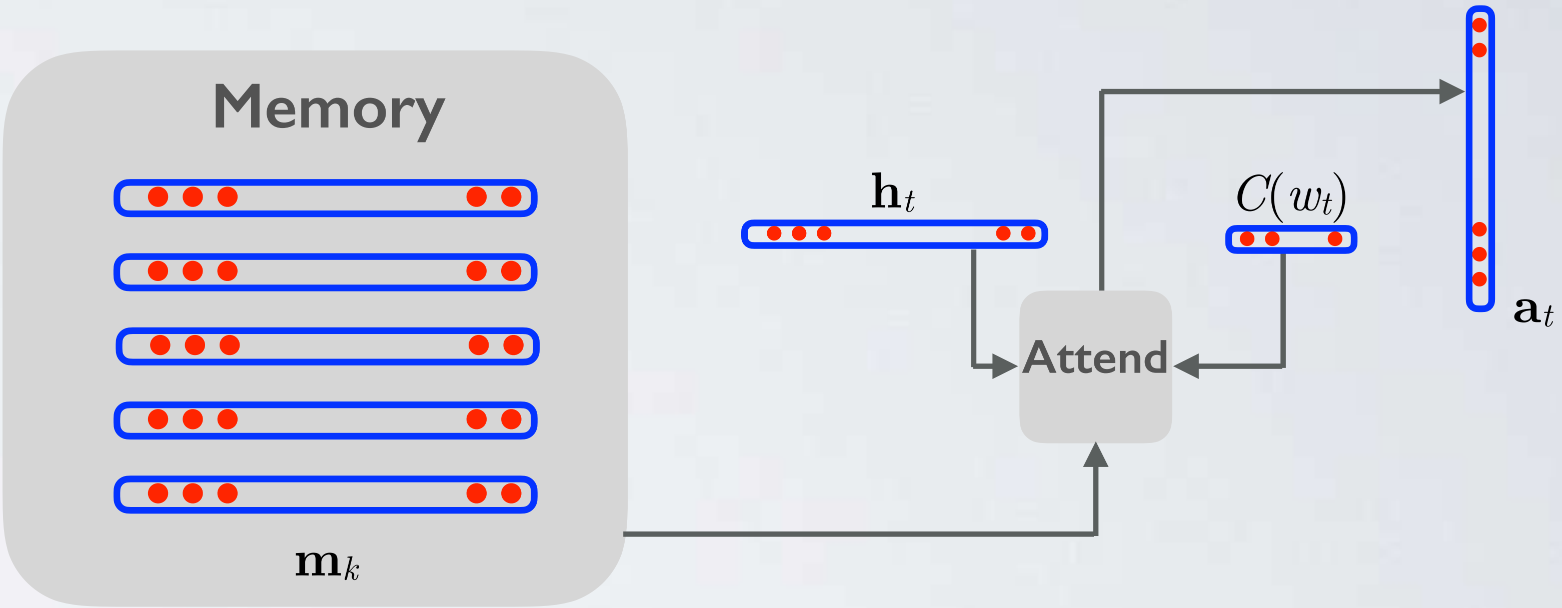
  - returns the **weighted average** of memory vectors $\mathbf{a}_t = \sum_{k=1}^{T_{\mathbf{x}}} \alpha_{t,k} \mathbf{m}_k$

ATTENTION

Topics: attention

- In (neural) machine translation, the attention learns an alignment model between the source sentence $\mathbf{x}$ and the target sentence $\mathbf{w}$





For more details, see:

Neural Machine Translation by Jointly Learning to Align and Translate
Bahdanau, Cho and Bengio, ICLR 2015

… and many more papers!

BABI TASKS

TOWARDS AI-COMPLETE QUESTION ANSWERING A SET OF PREREQUISITE TOY TASKS. **Weston et al. (2015)**

bAbi Tasks: Inspired by desire to go beyond classification, probe deeper into language understanding and reasoning.

- Each task checks one skill that a reasoning system should have.
- Systems should be able to solve all tasks: no task specific engineering.
- **Small training set:** 1000 questions / task.

BABI TASKS

TOWARDS AI-COMPLETE QUESTION ANSWERING A SET OF PREREQUISITE TOY TASKS. **Weston et al. (2015)**

20 Tasks: Examples from each of the bAbI tasks

Task 1: Single Supporting Fact
Mary went to the bathroom.
John moved to the hallway.
Mary travelled to the office.
Where is Mary? A:office

Task 2: Two Supporting Facts
John is in the playground.
John picked up the football.
Bob went to the kitchen.
Where is the football? A:playground

Task 3: Three Supporting Facts
John picked up the apple.
John went to the office.
John went to the kitchen.
John dropped the apple.
Where was the apple before the kitchen? A:office

Task 4: Two Argument Relations
The office is north of the bedroom.
The bedroom is north of the bathroom.
The kitchen is west of the garden.
What is north of the bedroom? A: office
What is the bedroom north of? A: bathroom

Task 5: Three Argument Relations
Mary gave the cake to Fred.
Fred gave the cake to Bill.
Jeff was given the milk by Bill.
Who gave the cake to Fred? A: Mary
Who did Fred give the cake to? A: Bill

Task 6: Yes/No Questions
John moved to the playground.
Daniel went to the bathroom.
John went back to the hallway.
Is John in the playground? A:no
Is Daniel in the bathroom? A:yes

Task 7: Counting
Daniel picked up the football.
Daniel dropped the football.
Daniel got the milk.
Daniel took the apple.
How many objects is Daniel holding? A: two

Task 8: Lists/Sets
Daniel picks up the football.
Daniel drops the newspaper.
Daniel picks up the milk.
John took the apple.
What is Daniel holding? milk, football

Task 9: Simple Negation
Sandra travelled to the office.
Fred is no longer in the office.
Is Fred in the office? A:no
Is Sandra in the office? A:yes

Task 10: Indefinite Knowledge
John is either in the classroom or the playground.
Sandra is in the garden.
Is John in the classroom? A:maybe
Is John in the office? A:no

Task 11: Basic Coreference
Daniel was in the kitchen.
Then he went to the studio.
Sandra was in the office.
Where is Daniel? A:studio

Task 12: Conjunction
Mary and Jeff went to the kitchen.
Then Jeff went to the park.
Where is Mary? A: kitchen
Where is Jeff? A: park

Task 13: Compound Coreference
Daniel and Sandra journeyed to the office.
Then they went to the garden.
Sandra and John travelled to the kitchen.
After that they moved to the hallway.
Where is Daniel? A: garden

Task 14: Time Reasoning
In the afternoon Julie went to the park.
Yesterday Julie was at school.
Julie went to the cinema this evening.
Where did Julie go after the park? A:cinema
Where was Julie before the park? A:school

Task 15: Basic Deduction
Sheep are afraid of wolves.
Cats are afraid of dogs.
Mice are afraid of cats.
Gertrude is a sheep.
What is Gertrude afraid of? A:wolves

Task 16: Basic Induction
Lily is a swan.
Lily is white.
Bernhard is green.
Greg is a swan.
What color is Greg? A:white

Task 17: Positional Reasoning
The triangle is to the right of the blue square.
The red square is on top of the blue square.
The red sphere is to the right of the blue square.
Is the red sphere to the right of the blue square? A:yes
Is the red square to the left of the triangle? A:yes

Task 18: Size Reasoning
The football fits in the suitcase.
The suitcase fits in the cupboard.
The box is smaller than the football.
Will the box fit in the suitcase? A:yes
Will the cupboard fit in the box? A:no

Task 19: Path Finding
The kitchen is north of the hallway.
The bathroom is west of the bedroom.
The den is east of the hallway.
The office is south of the bedroom.
How do you go from den to kitchen? A: west, north
How do you go from office to bathroom? A: north, west

Task 20: Agent's Motivations
John is hungry.
John goes to the kitchen.
John grabbed the apple there.
Daniel is hungry.
Where does Daniel go? A:kitchen
Why did John go to the kitchen? A:hungry

BABI TASKS

TOWARDS AI-COMPLETE QUESTION ANSWERING A SET OF PREREQUISITE TOY TASKS. Weston et al. (2015)

Task 1: Single Supporting Fact

Mary went to the bathroom.
John moved to the hallway.
Mary travelled to the office.
Where is Mary? A:office

BABI TASKS

TOWARDS AI-COMPLETE QUESTION ANSWERING A SET OF PREREQUISITE TOY TASKS. Weston et al. (2015)

Task 2: Two Supporting Facts

John is in the playground.
John picked up the football.
Bob went to the kitchen.
Where is the football? A:playground

BABI TASKS

TOWARDS AI-COMPLETE QUESTION ANSWERING A SET OF PREREQUISITE TOY TASKS. **Weston et al. (2015)**

Task 3: Three Supporting Facts

John picked up the apple.
John went to the office.
John went to the kitchen.
John dropped the apple.
Where was the apple before the kitchen? A:office

BABI TASKS

TOWARDS AI-COMPLETE QUESTION ANSWERING A SET OF PREREQUISITE TOY TASKS. **Weston et al. (2015)**

Task 4: Two Argument Relations

The office is north of the bedroom.
The bedroom is north of the bathroom.
The kitchen is west of the garden.
What is north of the bedroom? A: office
What is the bedroom north of? A: bathroom

BABI TASKS

TOWARDS AI-COMPLETE QUESTION ANSWERING A SET OF PREREQUISITE TOY TASKS. **Weston et al. (2015)**

Task 5: Three Argument Relations

Mary gave the cake to Fred.
Fred gave the cake to Bill.
Jeff was given the milk by Bill.
Who gave the cake to Fred? A: Mary
Who did Fred give the cake to? A: Bill

BABI TASKS

TOWARDS AI-COMPLETE QUESTION ANSWERING A SET OF PREREQUISITE TOY TASKS. **Weston et al. (2015)**

20 Tasks: Examples from each of the bAbI tasks

Task 1: Single Supporting Fact
Mary went to the bathroom.
John moved to the hallway.
Mary travelled to the office.
Where is Mary? A:office

Task 2: Two Supporting Facts
John is in the playground.
John picked up the football.
Bob went to the kitchen.
Where is the football? A:playground

Task 3: Three Supporting Facts
John picked up the apple.
John went to the office.
John went to the kitchen.
John dropped the apple.
Where was the apple before the kitchen? A:office

Task 4: Two Argument Relations
The office is north of the bedroom.
The bedroom is north of the bathroom.
The kitchen is west of the garden.
What is north of the bedroom? A: office
What is the bedroom north of? A: bathroom

Task 5: Three Argument Relations
Mary gave the cake to Fred.
Fred gave the cake to Bill.
Jeff was given the milk by Bill.
Who gave the cake to Fred? A: Mary
Who did Fred give the cake to? A: Bill

Task 6: Yes/No Questions
John moved to the playground.
Daniel went to the bathroom.
John went back to the hallway.
Is John in the playground? A:no
Is Daniel in the bathroom? A:yes

Task 7: Counting
Daniel picked up the football.
Daniel dropped the football.
Daniel got the milk.
Daniel took the apple.
How many objects is Daniel holding? A: two

Task 8: Lists/Sets
Daniel picks up the football.
Daniel drops the newspaper.
Daniel picks up the milk.
John took the apple.
What is Daniel holding? milk, football

Task 9: Simple Negation
Sandra travelled to the office.
Fred is no longer in the office.
Is Fred in the office? A:no
Is Sandra in the office? A:yes

Task 10: Indefinite Knowledge
John is either in the classroom or the playground.
Sandra is in the garden.
Is John in the classroom? A:maybe
Is John in the office? A:no

Task 11: Basic Coreference
Daniel was in the kitchen.
Then he went to the studio.
Sandra was in the office.
Where is Daniel? A:studio

Task 12: Conjunction
Mary and Jeff went to the kitchen.
Then Jeff went to the park.
Where is Mary? A: kitchen
Where is Jeff? A: park

Task 13: Compound Coreference
Daniel and Sandra journeyed to the office.
Then they went to the garden.
Sandra and John travelled to the kitchen.
After that they moved to the hallway.
Where is Daniel? A: garden

Task 14: Time Reasoning
In the afternoon Julie went to the park.
Yesterday Julie was at school.
Julie went to the cinema this evening.
Where did Julie go after the park? A:cinema
Where was Julie before the park? A:school

Task 15: Basic Deduction
Sheep are afraid of wolves.
Cats are afraid of dogs.
Mice are afraid of cats.
Gertrude is a sheep.
What is Gertrude afraid of? A:wolves

Task 16: Basic Induction
Lily is a swan.
Lily is white.
Bernhard is green.
Greg is a swan.
What color is Greg? A:white

Task 17: Positional Reasoning
The triangle is to the right of the blue square.
The red square is on top of the blue square.
The red sphere is to the right of the blue square.
Is the red sphere to the right of the blue square? A:yes
Is the red square to the left of the triangle? A:yes

Task 18: Size Reasoning
The football fits in the suitcase.
The suitcase fits in the cupboard.
The box is smaller than the football.
Will the box fit in the suitcase? A:yes
Will the cupboard fit in the box? A:no

Task 19: Path Finding
The kitchen is north of the hallway.
The bathroom is west of the bedroom.
The den is east of the hallway.
The office is south of the bedroom.
How do you go from den to kitchen? A: west, north
How do you go from office to bathroom? A: north, west

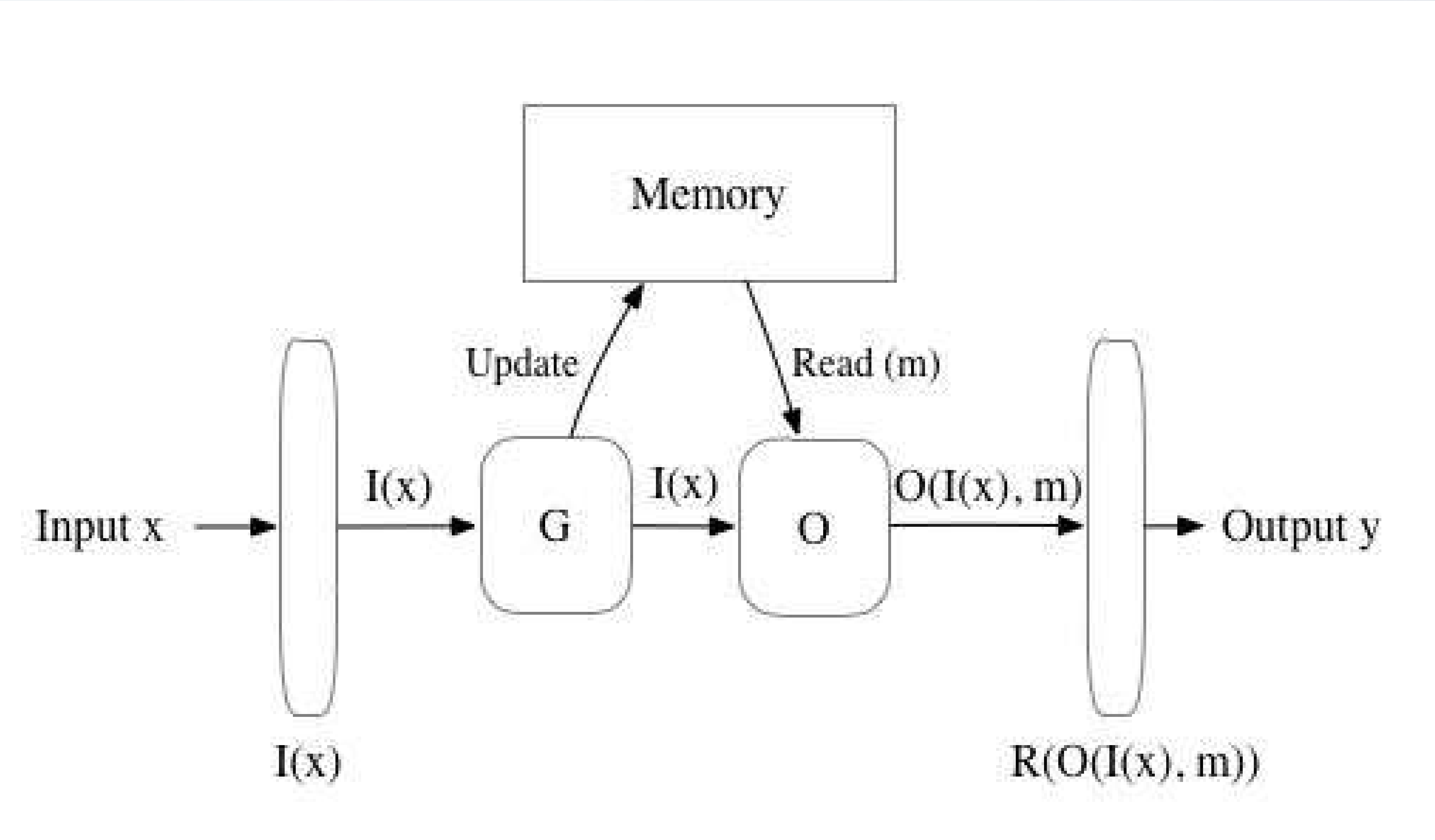
Task 20: Agent's Motivations
John is hungry.
John goes to the kitchen.
John grabbed the apple there.
Daniel is hungry.
Where does Daniel go? A:kitchen
Why did John go to the kitchen? A:hungry

MEMORY NETWORKS

Memory networks
Weston, Chopra and Bordes, ICLR 2015

Modular Structure

- **Significance**: One of the first application of memory augmented networks on an NLP task.
- Each module is a simple linear model.
- **Fully supervised**: The model is not trained on end to end.
- Memory representation is not learned (i.e. sentences or relations)
- Experiments focused both on bAbI (Episodic QA), factoid & knowledge based QA tasks.



MEMORY NETWORKS

Memory networks
Weston, Chopra and Bordes, ICLR 2015

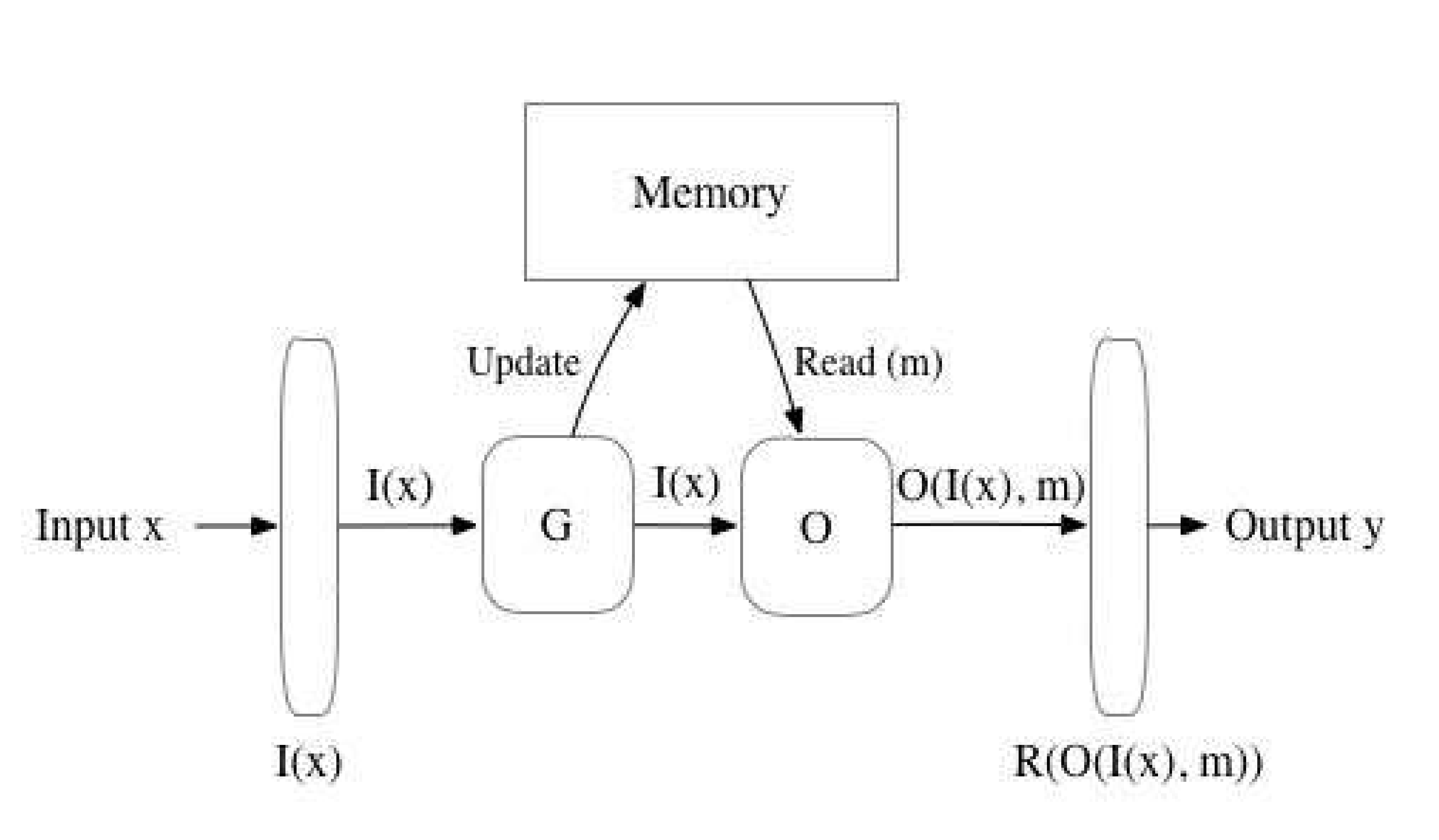
I: (input feature map)
- converts incoming input to internal feature representation.

G: (generalization)
- updates old memories given the new input.

O: (output feature map)
- produces a new output (in feature representation space), given the new input and the current memory state.

R: (response)
- converts the output into the response format desired. For example, a textual response or an action.

Modular Structure



END-TO-END MEMORY NETWORKS

Sukhbaatar, Szlam, Weston and Fergus.
NIPS 2015

Learned with much less supervision than the original MemNet

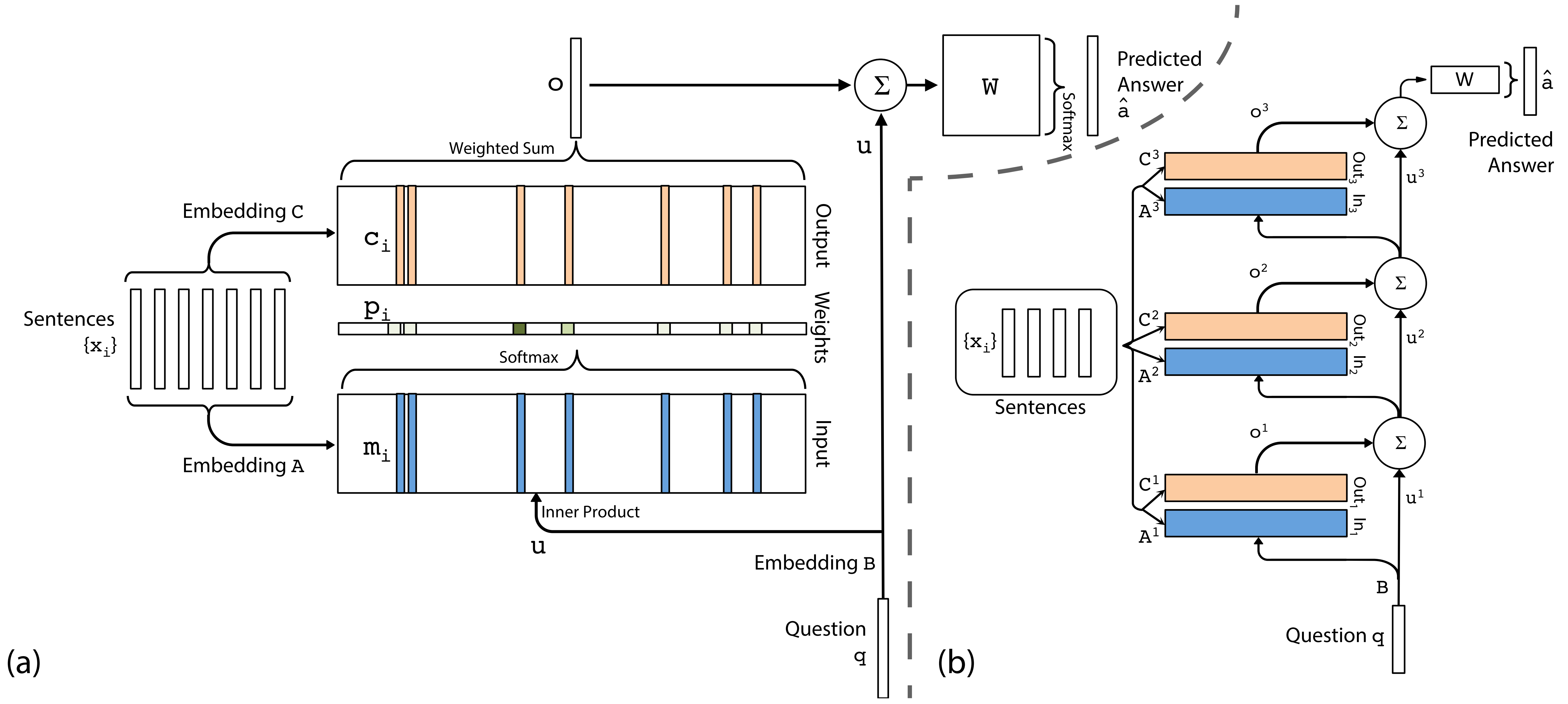


Figure 1: (a): A single layer version of our model. (b): A three layer version of our model. In practice, we can constrain several of the embedding matrices to be the same (see Section 2.2).

MemN2N: (on bAbi)

Task	Baseline			MemN2N									
	Strongly Supervised MemNN [22]	LSTM [22]	MemNN WSH	BoW	PE	PE LS	PE LS RN	1 hop PE LS joint	2 hops PE LS joint	3 hops PE LS joint	PE LS RN joint	PE LS LW joint
1: 1 supporting fact	0.0	50.0	0.1	0.6	0.1	0.2	0.0	0.8	0.0	0.1	0.0	0.1
2: 2 supporting facts	0.0	80.0	42.8	17.6	21.6	12.8	8.3	62.0	15.6	14.0	11.4	18.8
3: 3 supporting facts	0.0	80.0	76.4	71.0	64.2	58.8	40.3	76.9	31.6	33.1	21.9	31.7
4: 2 argument relations	0.0	39.0	40.3	32.0	3.8	11.6	2.8	22.8	2.2	5.7	13.4	17.5
5: 3 argument relations	2.0	30.0	16.3	18.3	14.1	15.7	13.1	11.0	13.4	14.8	14.4	12.9
6: yes/no questions	0.0	52.0	51.0	8.7	7.9	8.7	7.6	7.2	2.3	3.3	2.8	2.0
7: counting	15.0	51.0	36.1	23.5	21.6	20.3	17.3	15.9	25.4	17.9	18.3	10.1
8: lists/sets	9.0	55.0	37.8	11.4	12.6	12.7	10.0	13.2	11.7	10.1	9.3	6.1
9: simple negation	0.0	36.0	35.9	21.1	23.3	17.0	13.2	5.1	2.0	3.1	1.9	1.5
10: indefinite knowledge	2.0	56.0	68.7	22.8	17.4	18.6	15.1	10.6	5.0	6.6	6.5	2.6
11: basic coreference	0.0	38.0	30.0	4.1	4.3	0.0	0.9	8.4	1.2	0.9	0.3	3.3
12: conjunction	0.0	26.0	10.1	0.3	0.3	0.1	0.2	0.4	0.0	0.3	0.1	0.0
13: compound coreference	0.0	6.0	19.7	10.5	9.9	0.3	0.4	6.3	0.2	1.4	0.2	0.5
14: time reasoning	1.0	73.0	18.3	1.3	1.8	2.0	1.7	36.9	8.1	8.2	6.9	2.0
15: basic deduction	0.0	79.0	64.8	24.3	0.0	0.0	0.0	46.4	0.5	0.0	0.0	1.8
16: basic induction	0.0	77.0	50.5	52.0	52.1	1.6	1.3	47.4	51.3	3.5	2.7	51.0
17: positional reasoning	35.0	49.0	50.9	45.4	50.1	49.0	51.0	44.4	41.2	44.5	40.4	42.6
18: size reasoning	5.0	48.0	51.3	48.1	13.6	10.1	11.1	9.6	10.3	9.2	9.4	9.2
19: path finding	64.0	92.0	100.0	89.7	87.4	85.6	82.8	90.7	89.9	90.2	88.0	90.6
20: agent's motivation	0.0	9.0	3.6	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.2
Mean error (%)	6.7	51.3	40.2	25.1	20.3	16.3	13.9	25.8	15.6	13.3	12.4	15.2
Failed tasks (err. $> 5\%$)	4	20	18	15	13	12	11	17	11	11	11	10
On 10k training data												
Mean error (%)	3.2	36.4	39.2	15.4	9.4	7.2	6.6	24.5	10.9	7.9	7.5	11.0
Failed tasks (err. $> 5\%$)	2	16	17	9	6	4	4	16	7	6	6	6

Table 1: Test error rates (%) on the 20 QA tasks for models using 1k training examples (mean test errors for 10k training examples are shown at the bottom). Key: BoW = bag-of-words representation; PE = position encoding representation; LS = linear start training; RN = random injection of time index noise; LW = RNN-style layer-wise weight tying (if not stated, adjacent weight tying is used); joint = joint training on all tasks (as opposed to per-task training).

KEY-VALUE MEMORY NETWORKS

Miller et al. (2016): Key-Value Memory Networks for Directly Reading Documents.

- Introduces key-value memory as a generalization of MemN2N
- Key-value memory allows for better query inference and scalability.

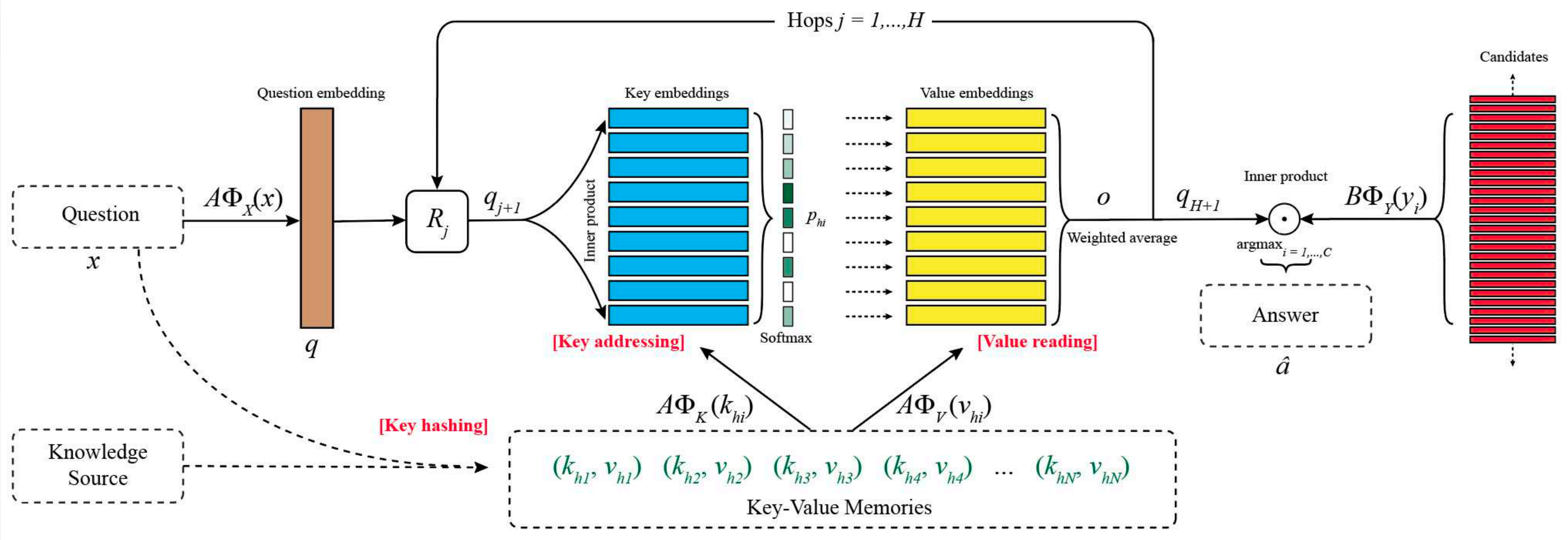


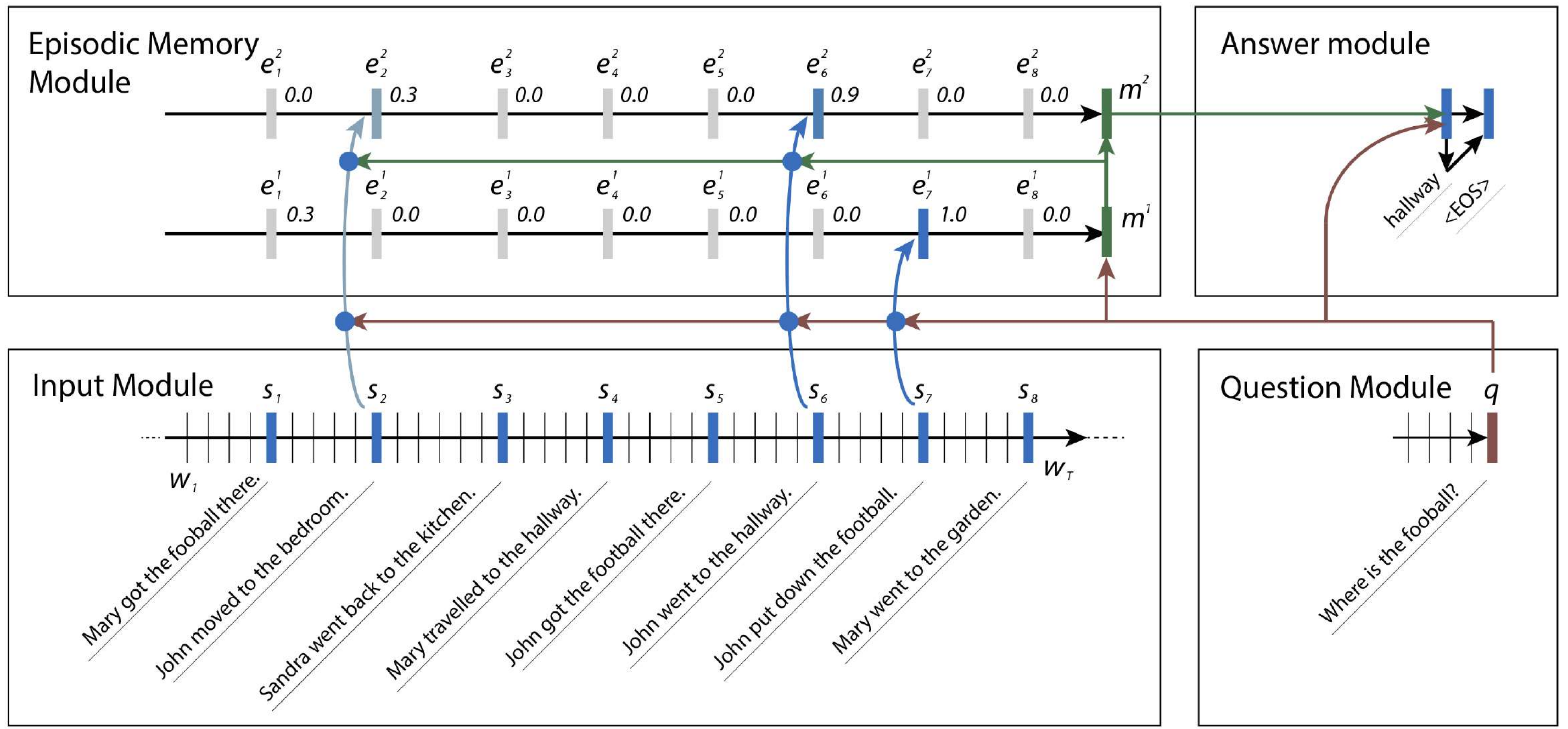
Figure 1: The Key-Value Memory Network model for question answering. See Section 3 for details.

DYNAMIC MEMORY NETWORKS

Ask Me Anything: Dynamic Memory Networks for Natural Language Processing
Kumar, Ondruska, Iyyer, Bradbury, Gulrajani, Zhong,Paulus, Socher, ICML 2016

- Read vector obtained from hidden layer of special **gated** **RNN**
- Multiple hops process represented by a **GRU**

- Figure: Note that the second iteration has wrongly placed some weight in sentence 2, which makes some intuitive sense, as sentence 2 is another place John had been.
- Figure: Connections for gates that are close to zero are not shown.

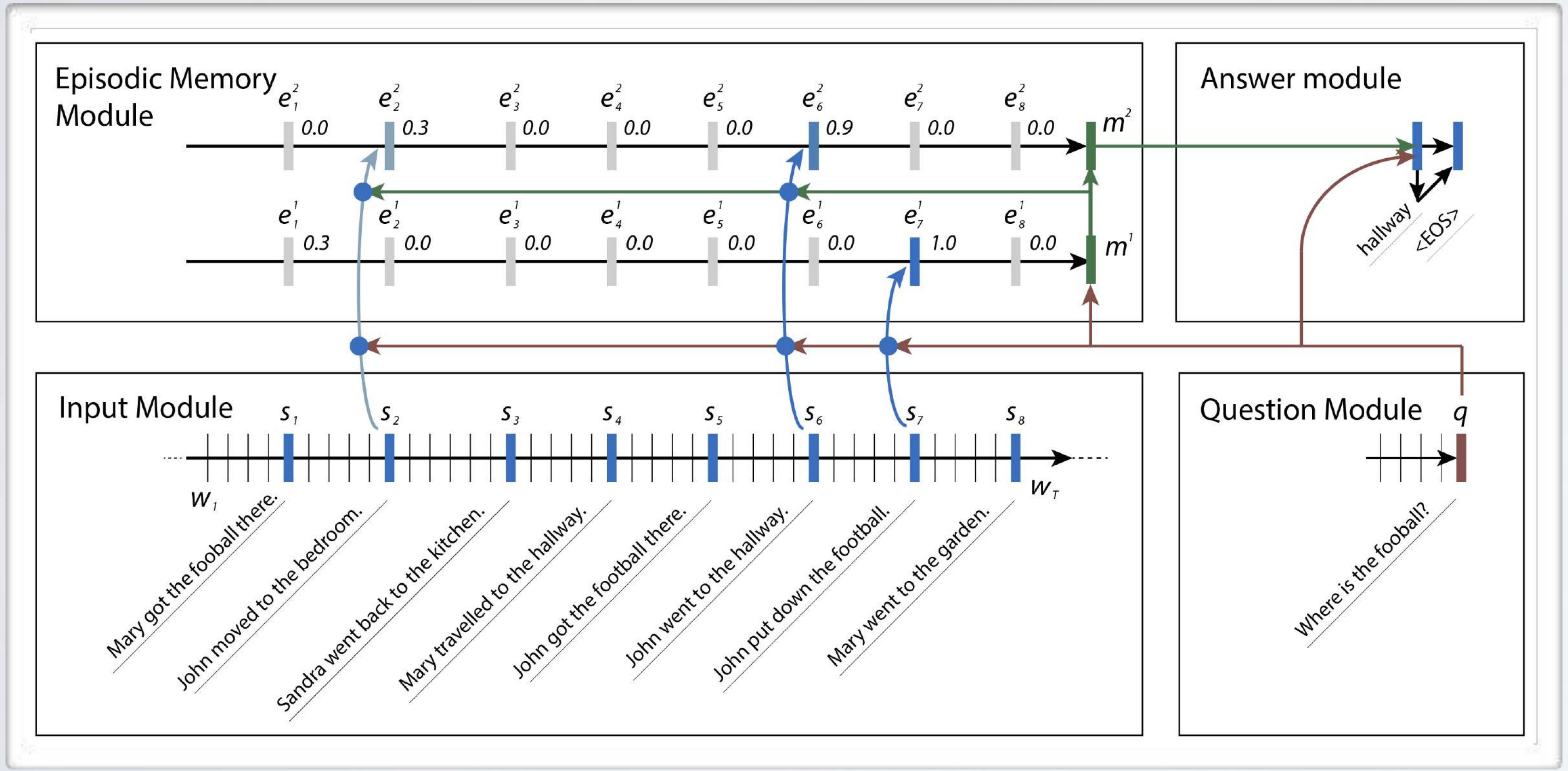


DYNAMIC MEMORY NETWORKS

- Results of DMN on bAbi
- DMN was also evaluated on other NLP tasks:
  - sentiment analysis
  - part-of-speech tagging

Task	MemNN	DMN
1: Single Supporting Fact	100	100
2: Two Supporting Facts	100	98.2
3: Three Supporting Facts	100	95.2
4: Two Argument Relations	100	100
5: Three Argument Relations	98	99.3
6: Yes/No Questions	100	100
7: Counting	85	96.9
8: Lists/Sets	91	96.5
9: Simple Negation	100	100
10: Indefinite Knowledge	98	97.5
11: Basic Coreference	100	99.9
12: Conjunction	100	100
13: Compound Coreference	100	99.8
14: Time Reasoning	99	100
15: Basic Deduction	100	100
16: Basic Induction	100	99.4
17: Positional Reasoning	65	59.6
18: Size Reasoning	95	95.3
19: Path Finding	36	34.5
20: Agent's Motivations	100	100
Mean Accuracy (%)	93.3	**93.6**

Table 1. Test accuracies on the bAbI dataset. MemNN numbers taken from Weston et al. (Weston et al., 2015a). The DMN passes (accuracy > 95%) 18 tasks, whereas the MemNN passes 16.

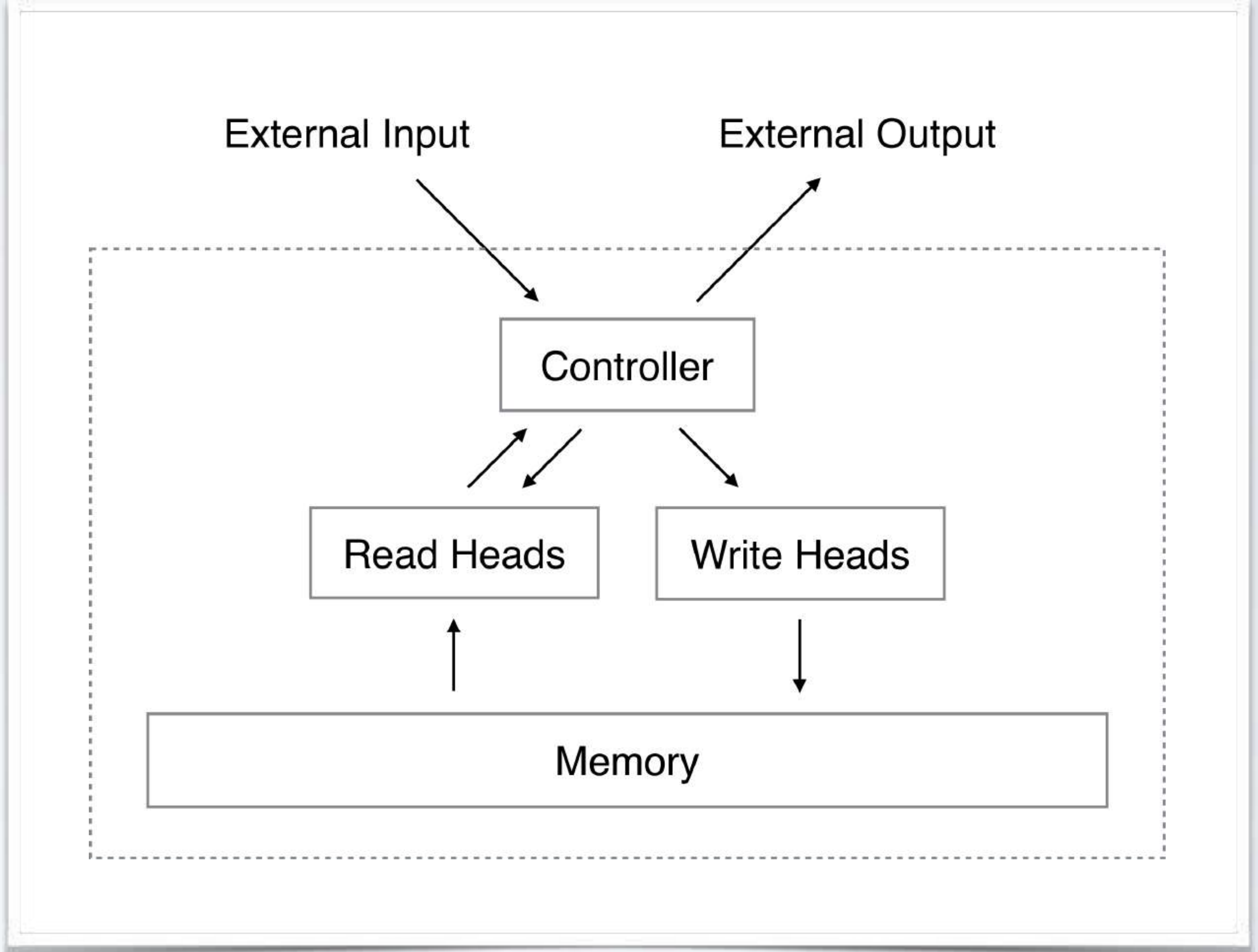


NEURAL TURING MACHINES

Neural Turing Machines
Graves, Wayne, Danihelka, arXiv 2014

Topics: neural turing machines

- In **neural turing machines**
  - introduces a mechanism to **erase and write** to the memory
  - distinguishes between **content-based** and **location-based** addressing



NEURAL TURING MACHINES

Neural Turing Machines
Graves, Wayne, Danihelka, arXiv 2014

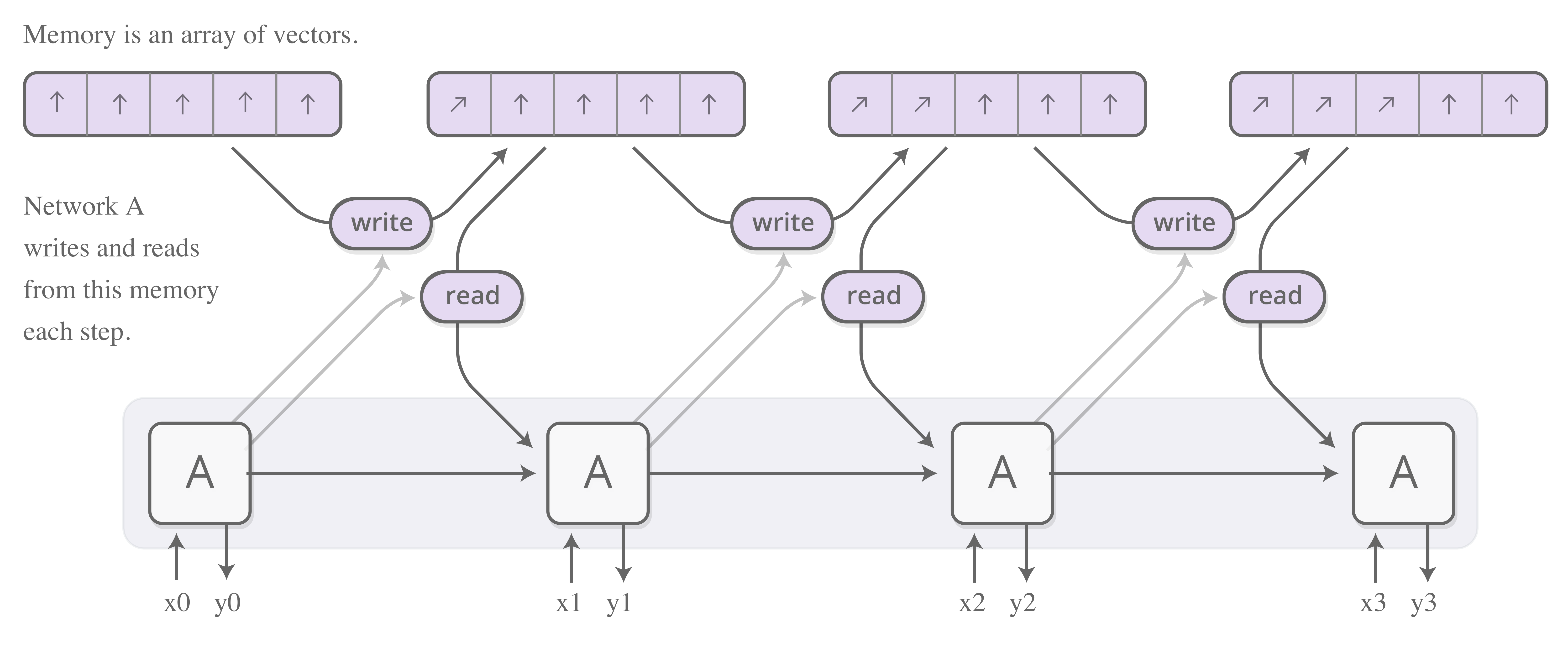
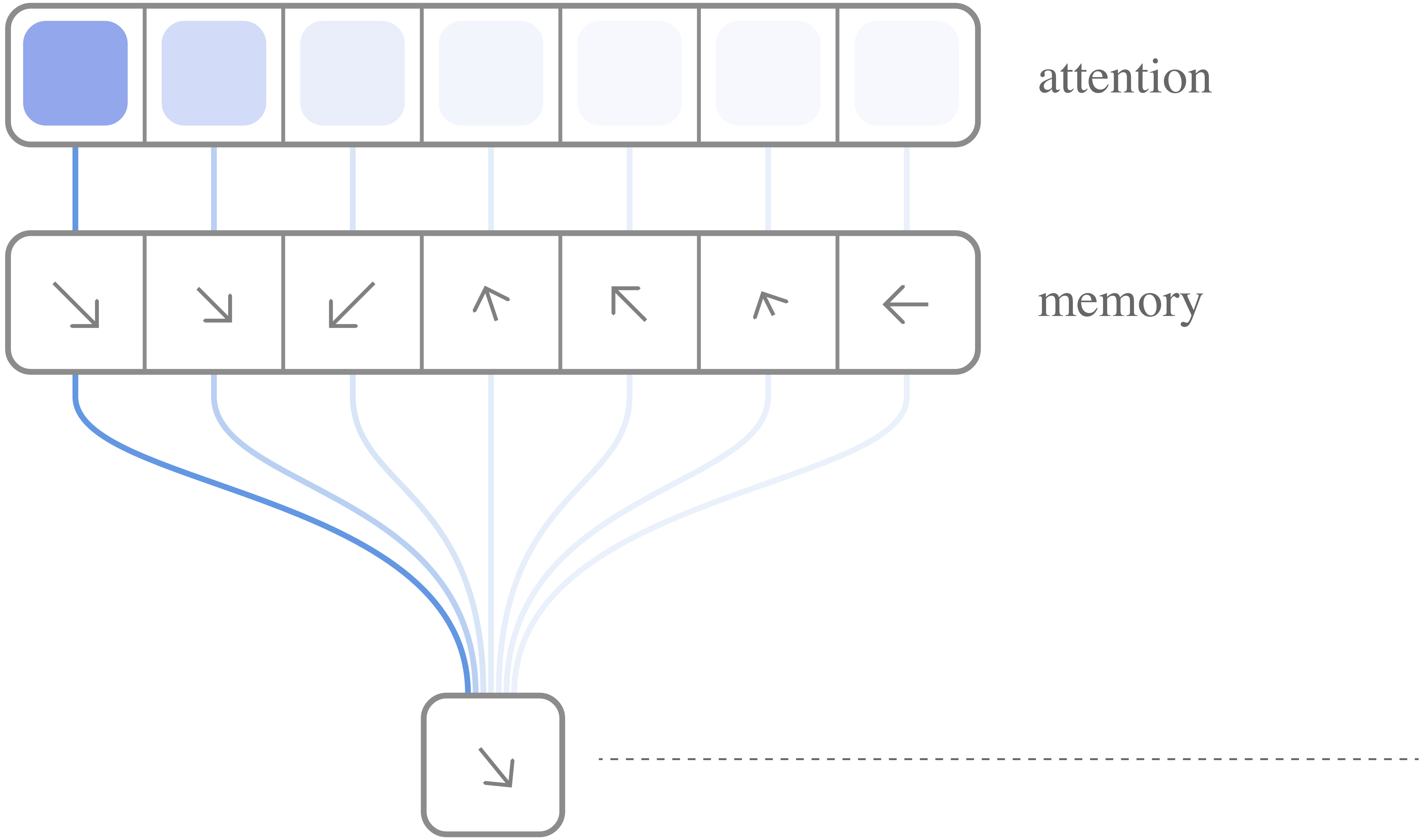


Image Credit: Olah & Carter (2016) *Attention and Augmented Recurrent Neural Networks*, Distill

NEURAL TURING MACHINES

Neural Turing Machines
Graves, Wayne, Danihelka, arXiv 2014

Reading: NTMs use RNN controlled soft-attention mechanism to read from memory



The RNN gives an attention distribution which describe how we spread out the amount we care about different memory positions.

The read result is a weighted sum.

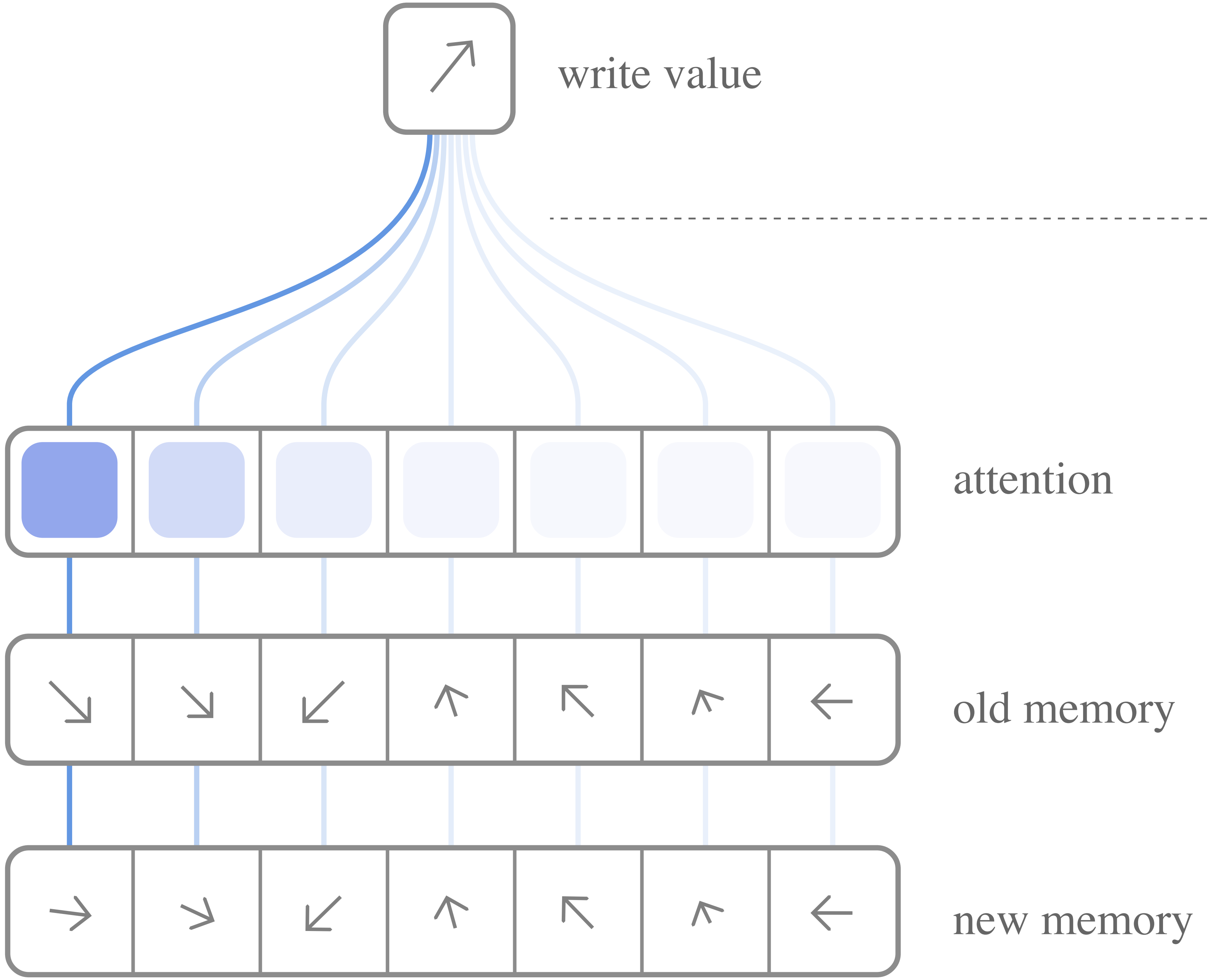
$$r \leftarrow \sum_i a_i M_i$$

Image Credit: **Olah & Carter (2016)** *Attention and Augmented Recurrent Neural Networks*, Distill

NEURAL TURING MACHINE

Neural Turing Machines
Graves, Wayne, Danihelka, arXiv 2014

Writing: NTMs use RNN controlled soft-attention mechanism to write to memory.



Instead of writing to one location, we write everywhere, just to different extents.

The RNN gives an attention distribution, describing how much we should change each memory position towards the write value.

$$M_i \leftarrow a_i w + (1 - a_i) M_i$$

Image Credit: **Olah & Carter (2016)** *Attention and Augmented Recurrent Neural Networks*, Distill

NTMS

Attention: Combination of content-based and location-based addressing.

- First interpolate between the previous weight and the current content based weight.

$$w_t^g = g_t w_t^c + (1 - g_t) w_{t-1} \qquad g_t \in (0,1)$$

- Generate shift weighting s_t which is a softmax distribution over -1,0,1 and use circular convolution to rotate w_t^g by s_t

$$\tilde{w}_t(i) = \sum_{j=0}^{N-1} w_t^g(j) s_t(i-j)$$

- Convolution can cause leakage of weights over time. Sharpen the final weighting using a scalar $\gamma_t \geq 1$

$$w_t(i) = \frac{\tilde{w}_t(i)^{\gamma_t}}{\sum_j \tilde{w}_t(j)^{\gamma_t}}$$

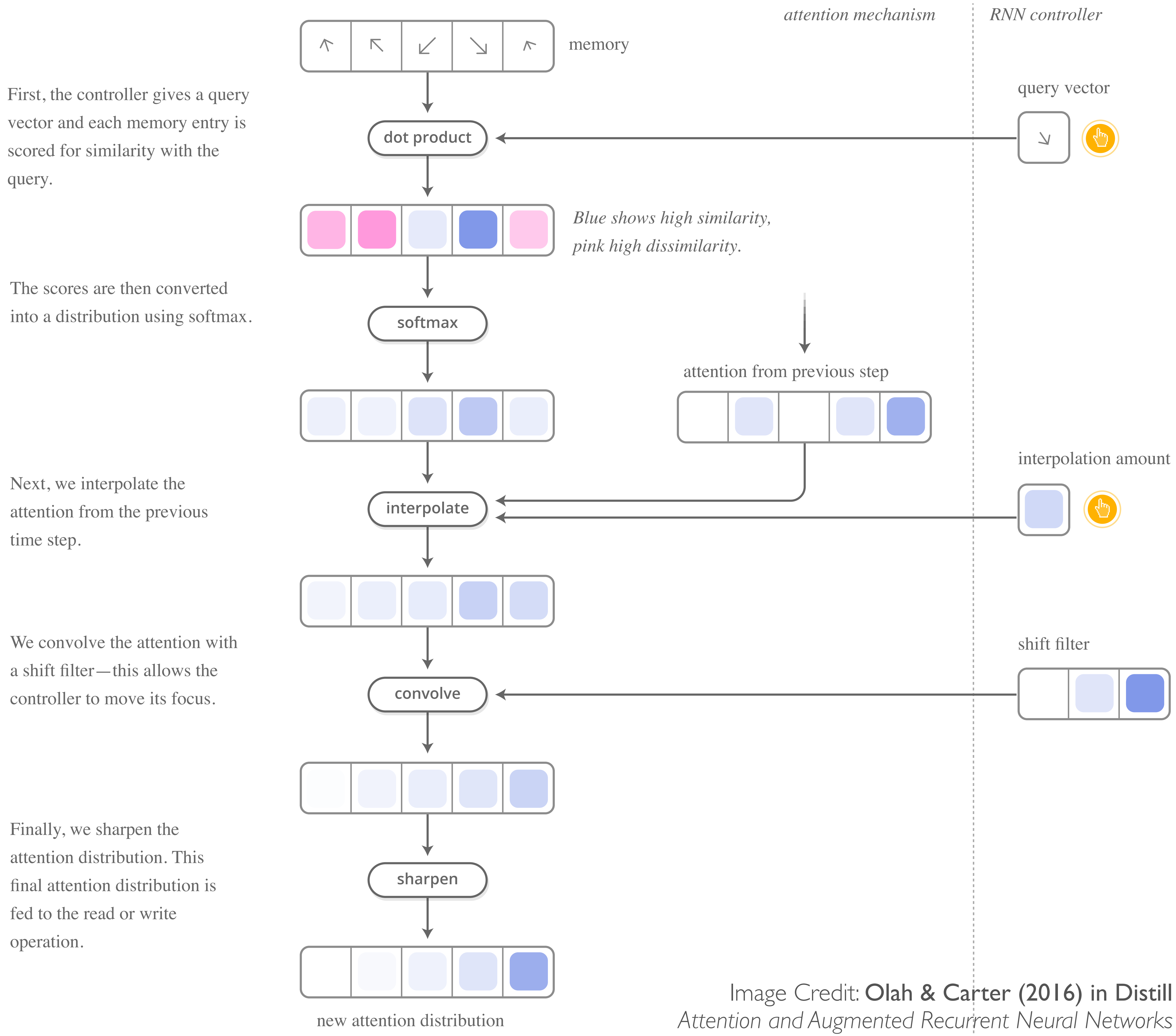


Image Credit: **Olah & Carter (2016) in Distill**
Attention and Augmented Recurrent Neural Networks

VARIATIONS ON MEMORY AND ATTENTION

Neural Turing Machines
Graves, Wayne, Danihelka, arXiv 2014

Topics: neural turing machines

- In **neural turing machines**
  - distinguishes between **content-based** and **location-based** addressing
  - introduces a mechanism to **erase and write** to the memory

- For more details, see the paper or this talk:

  https://www.youtube.com/watch?v=_H0i0IhEO2g