3. Iterative Pooling

original image

first attention layer

second attention layer

Question Answering

Joe went to the kitchen.

Fred went to the kitchen.

Joe picked up the milk.

Joe travelled to the office.

Joe left the milk.

Joe went to the bathroom.

Where is the milk?

Question Answering

Joe went to the kitchen.

Fred went to the kitchen.

Joe picked up the milk.

Joe travelled to the office.

Joe left the milk.

Joe went to the bathroom.

Where is the milk?

Question Answering

Joe went to the kitchen.

Fred went to the kitchen.

Joe picked up the milk.

Joe travelled to the office.

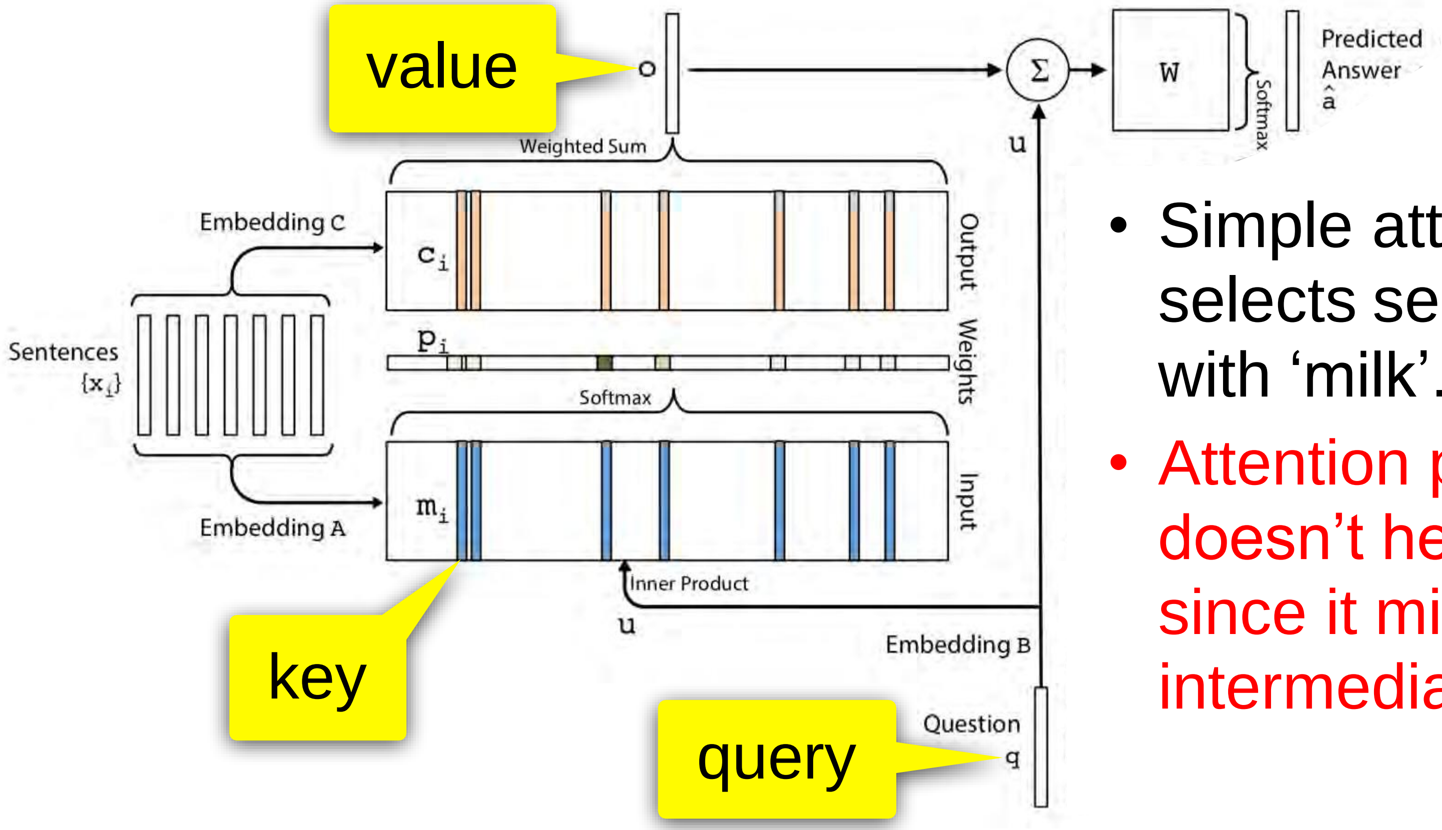
Joe left the milk.

Joe went to the bathroom.

Where is the milk?

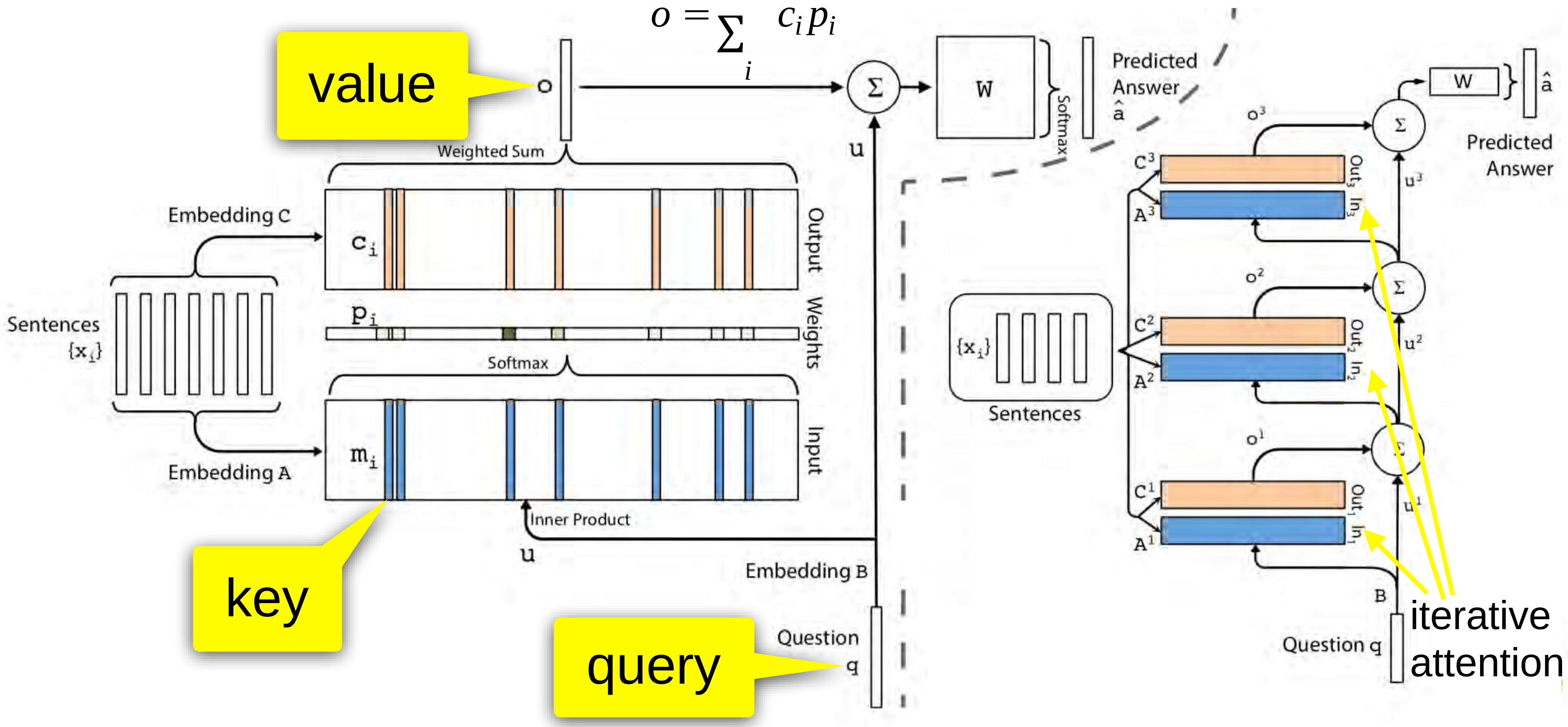
- Simple attention selects sentences with ‘milk’.
- Attention pooling doesn’t help much since it misses intermediate steps.

Question Answering with Pooling (Sukhbaatar et al., '15)



- Simple attention selects sentences with 'milk'.
- Attention pooling doesn't help much since it misses intermediate steps.

Question Answering with Pooling and Iteration (Sukhbaatar et al., ’15)



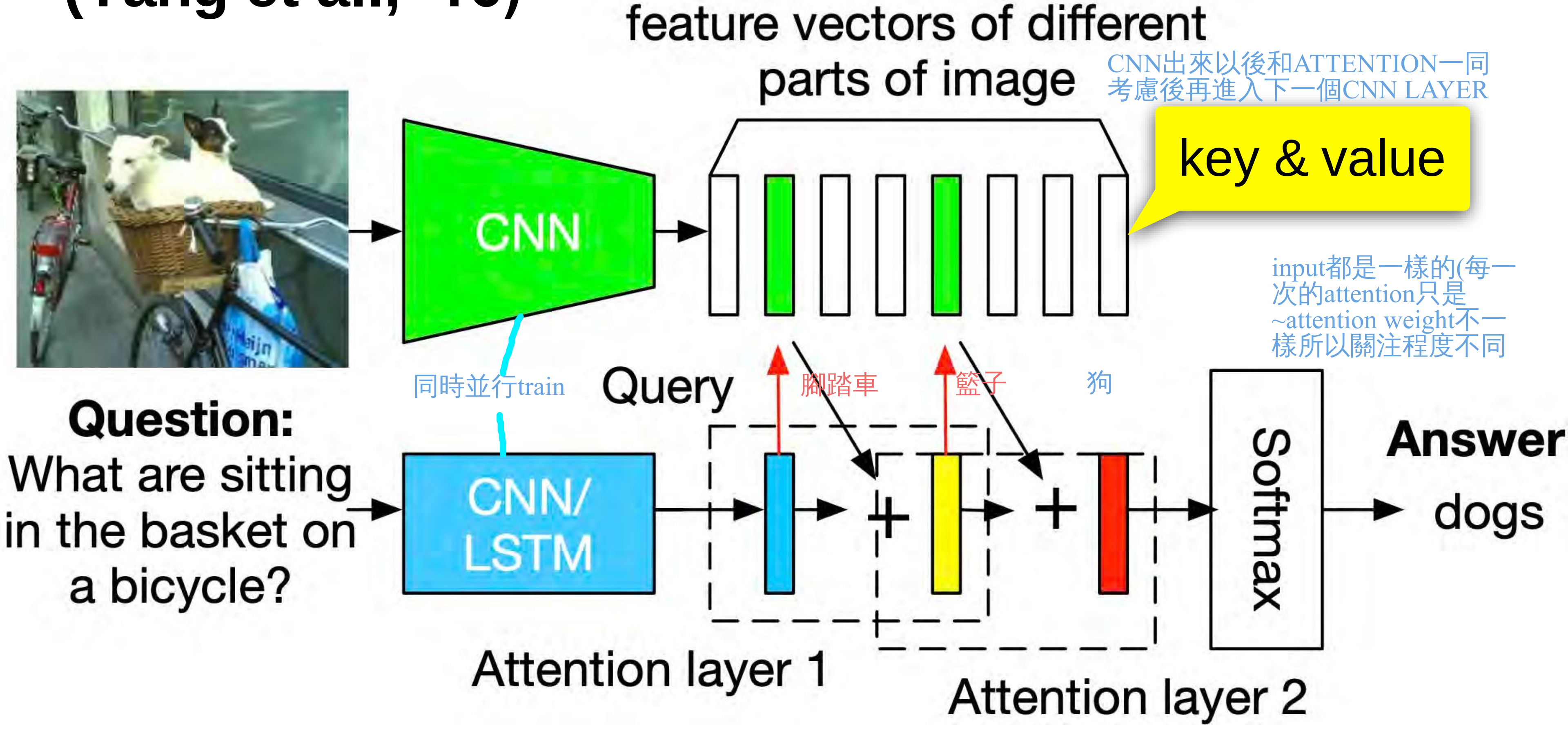
Question Answering with Pooling and Iteration (Sukhbaatar et al., ’15)

Sam walks into the kitchen.
Sam picks up an apple.
Sam walks into the bedroom.
Sam drops the apple.
Q: Where is the apple?
A. Bedroom

Brian is a lion.
Julius is a lion.
Julius is white.
Bernhard is green.
Q: What color is Brian?
A. White

Mary journeyed to the den.
Mary went back to the kitchen.
John journeyed to the bedroom.
Mary discarded the milk.
Q: Where was the milk before the den?
A. Hallway

Question Answering with Pooling and Iteration (Yang et al., '16)



Question Answering with Pooling and Iteration (Yang et al., ’16)

- Encode image via CNN
- Encode text query via LSTM
- Weigh patches according to attention and iterate

- Improving it (2019 tools)
  - Convolutionalize CNN (e.g. ResNet)
  - BERT for query encoding
  - Convolutional weighting (a la SE-Net)

(a) What are pulling a man on a wagon down on dirt road?
Answer: horses Prediction: horses

(b) What is the color of the box ?
Answer: red Prediction: red

(c) What next to the large umbrella attached to a table?
Answer: trees Prediction: tree

(d) How many people are going up the mountain with walking sticks?
Answer: four Prediction: four

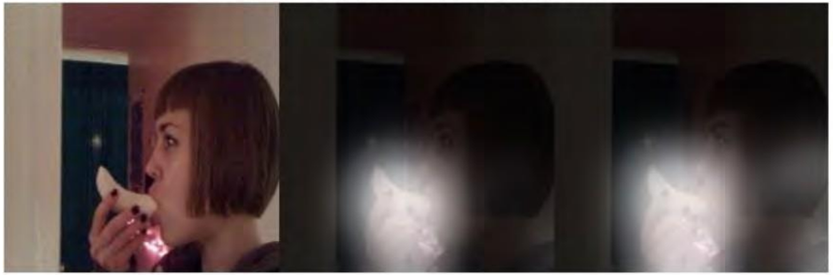
(e) What is sitting on the handle bar of a bicycle?
Answer: bird Prediction: bird

(f) What is the color of the horns?
Answer: red Prediction: red

(a) What swim in the ocean near two large ferries?
Answer: ducks Prediction: boats

(b) What is the color of the shirt?
Answer: purple Prediction: green

(c) What is the young woman eating?
Answer: banana Prediction: donut



(d) How many umbrellas with various patterns?
Answer: three Prediction: two

(e) The very old looking what is on display?
Answer: pot Prediction: vase

(f) What are passing underneath the walkway bridge?
Answer: cars Prediction: trains

Iterative Attention Summary

- Pooling

$$f(X) = \rho\left(\sum_{x \in X} \phi(x)\right)$$

- Attention pooling

$$f(X) = \rho\left(\sum_{x \in X} \alpha(x, w)\phi(x)\right)$$

- Iterative Attention pooling

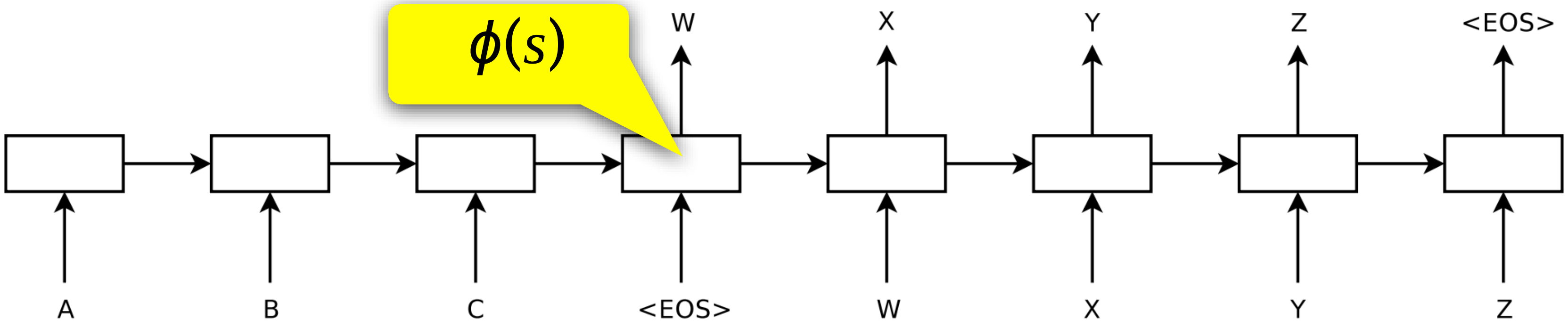
$$q_{t+1} = \rho\left(\sum_{x \in X} \alpha(x, q_t)\phi(x)\right)$$

Repeatedly update internal state


4. Iterative
Output

Seq2Seq for Machine Translation, Sutskever, Vinyals, Le ‘14

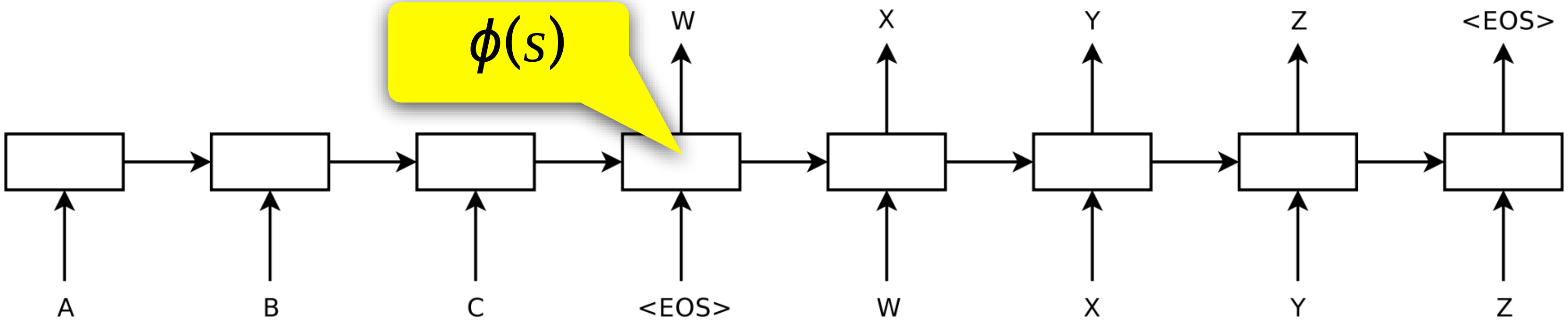
- Encode source sequence s via LSTM to representation $\phi(s)$
- Decode to target sequence one character at a time



- ‘The table is round.’ - ‘Der Tisch ist rund.’
- ‘The table is very beautiful with many inlaid patterns, blah blah blah blah’ - ‘Error …’

Seq2Seq for Machine Translation, Sutskever, Vinyals, Le ‘14

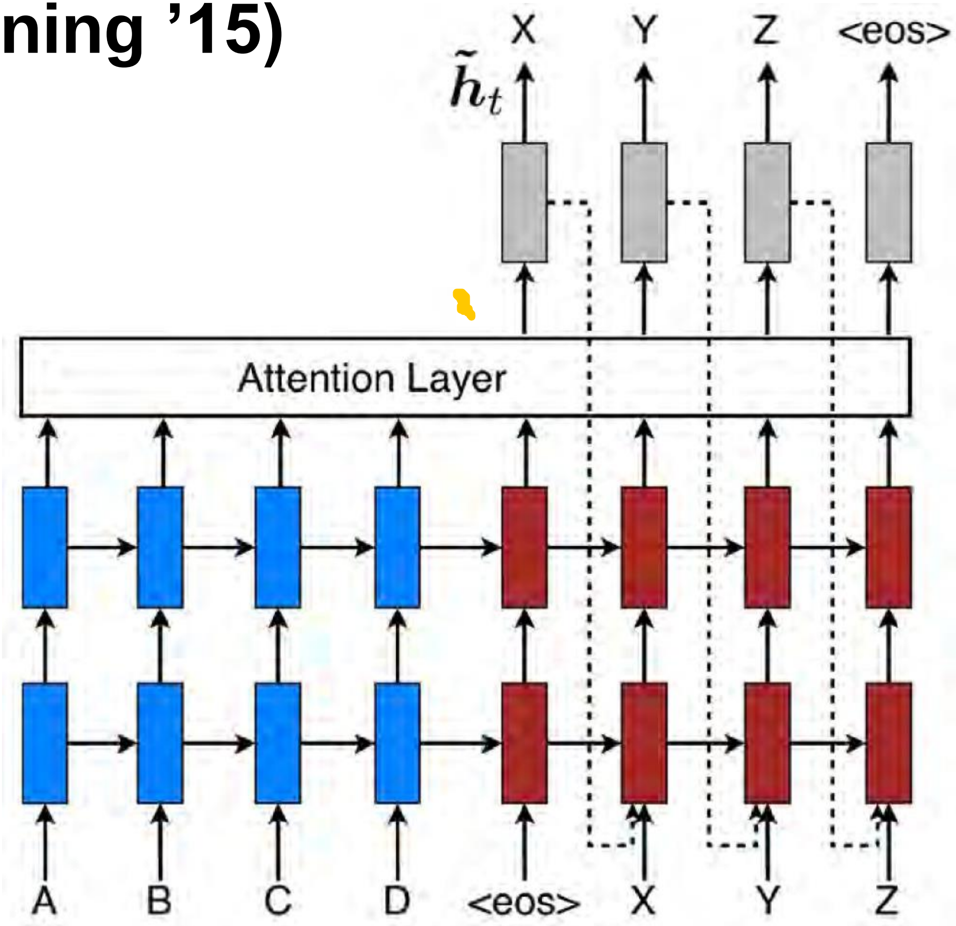
- Encode source sequence s via LSTM to representation $\phi(s)$
- Decode to target sequence one character at a time



- ‘The table is round.’ - ‘Der Tisch ist ru
- ‘The table is very beautiful with ma
  blah blah blah blah’ - ‘Error …’

Representation not rich enough

Seq2Seq with attention (Bahdanau, Cho, Bengio ’14) (Pham, Luong, Manning ’15)



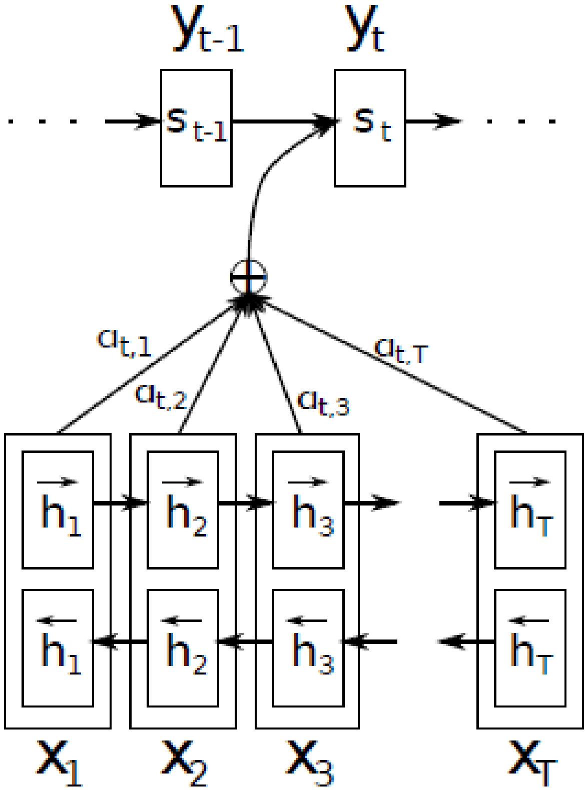
Seq2Seq with attention (Bahdanau, Cho, Bengio ’14) (Pham, Luong, Manning ’15)

$$e_{ij} = a(s_{i-1}, h_j)$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})},$$

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j.$$

$$s_i = f(s_{i-1}, y_{i-1}, c_i).$$



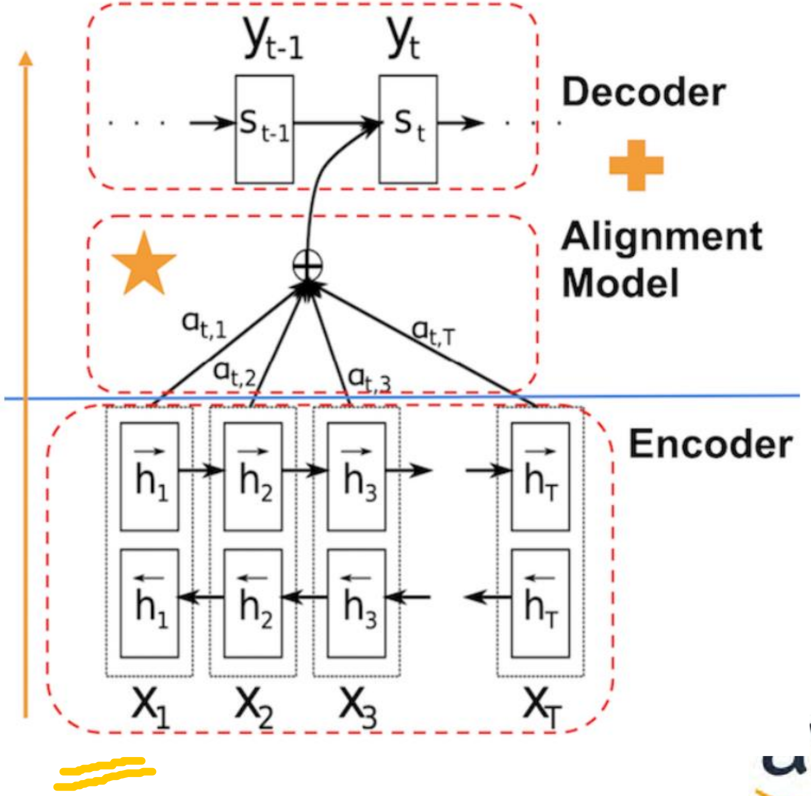
Seq2Seq with attention (Bahdanau, Cho, Bengio ’14) (Pham, Luong, Manning ’15)

$$e_{ij} = a(s_{i-1}, h_j)$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})},$$

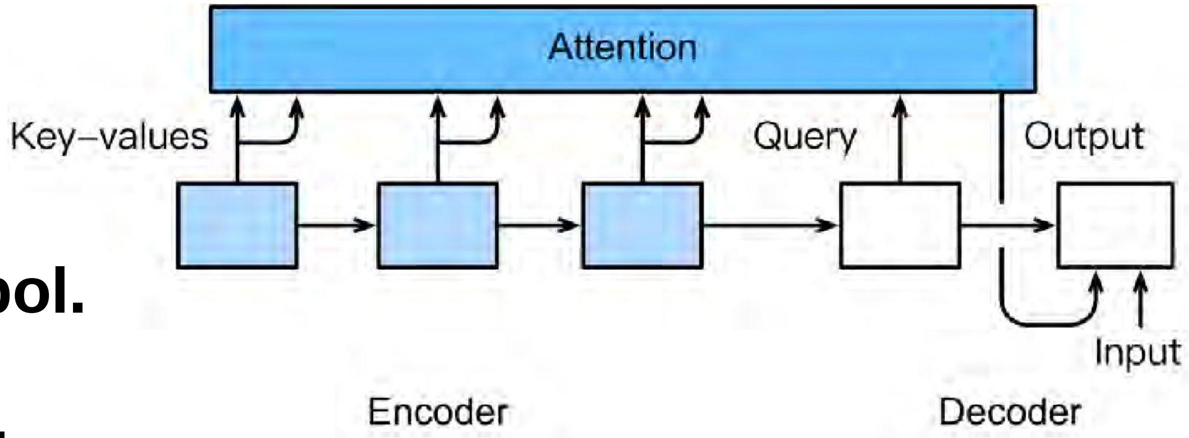
$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j.$$

$$s_i = f(s_{i-1}, y_{i-1}, c_i).$$

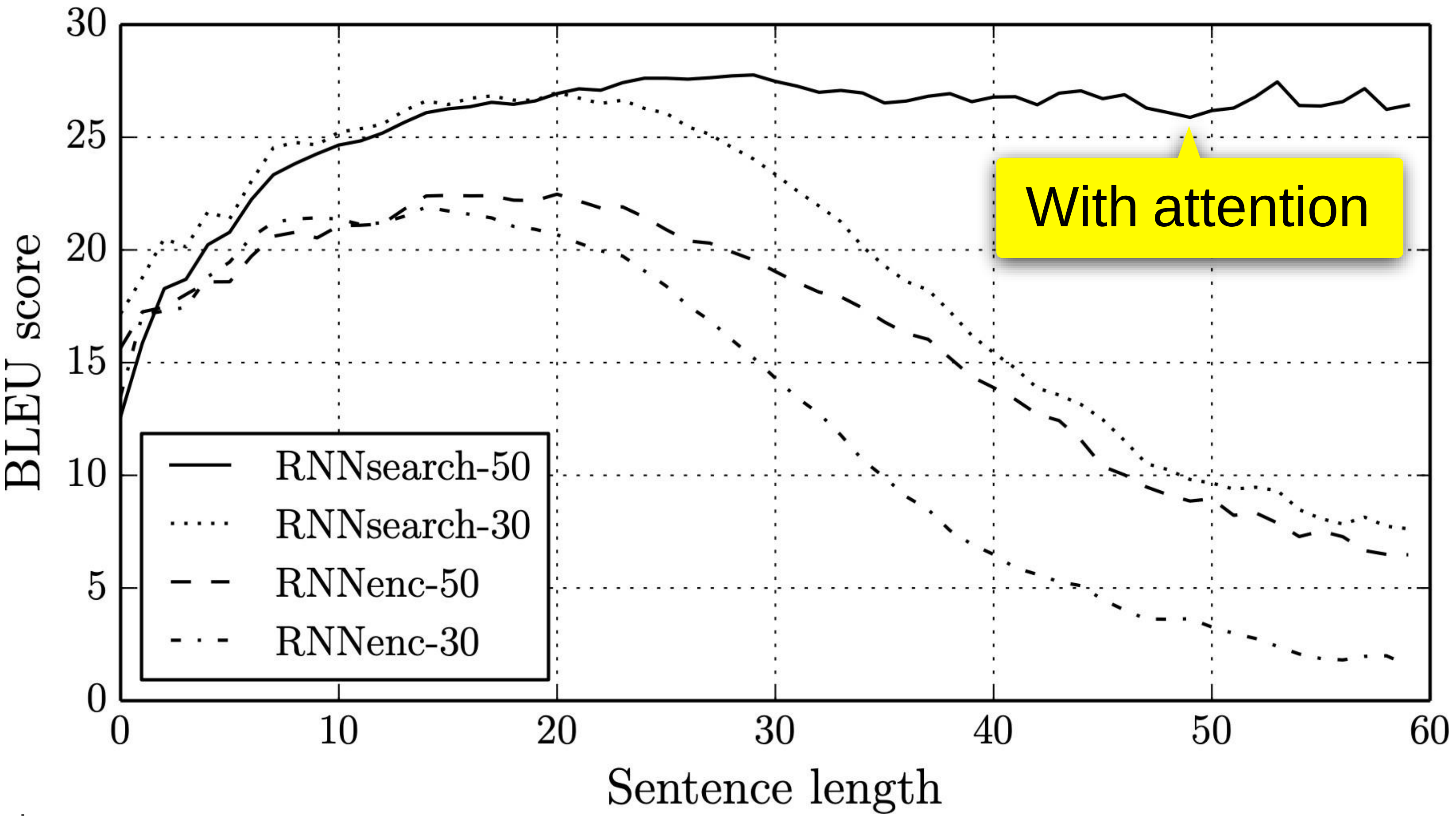


Seq2Seq with attention (Bahdanau, Cho, Bengio ’14) (Pham, Luong, Manning ’15)

- Iterative attention model
  - Compute (next) attention weights
  - Aggregate next state
  - Emit next symbol
- Repeat
- **Memory networks emit only one symbol.**
- **NMT with attention emits many symbols.**



Seq2Seq with attention (Bahdanau, Cho, Bengio ’14)



Variations on a Theme

BWV 988

(PART I)

J.S.Bach (1685-1750)

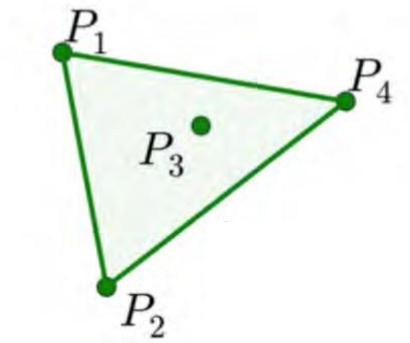
Aria

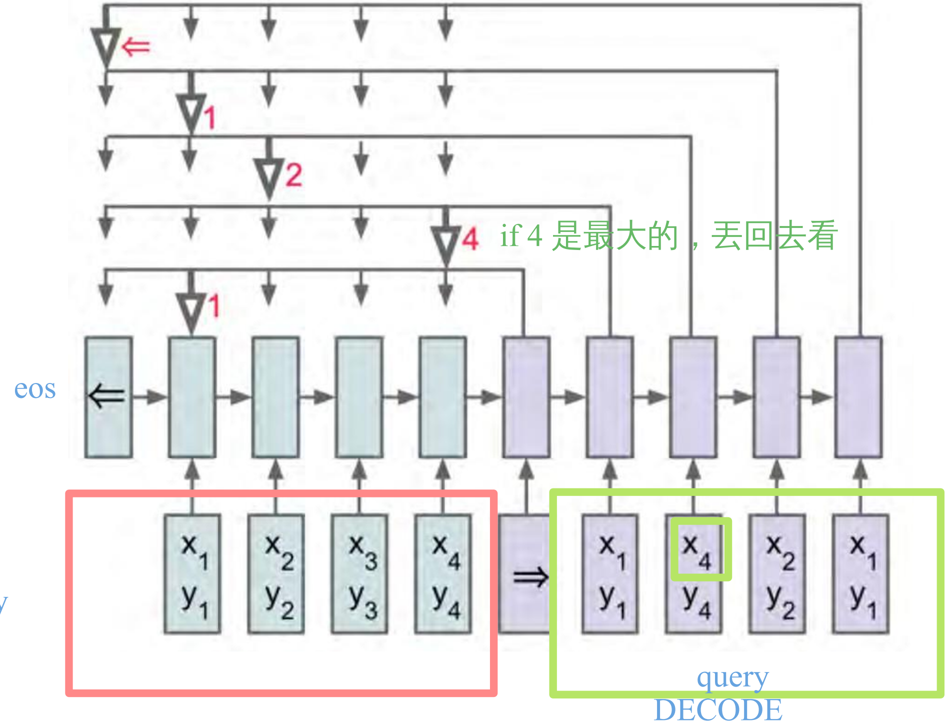
Pointer networks for finding convex hull (Vinyals et al., ‘15)

$$\text{Input } P = \{P_1, \ldots P_4\}$$

$$\text{Output } O = \{1,4,2,1\}$$

INPUT OUTPUT不固定 -> SEQ2SEQ





Pointer networks for finding convex hull (Vinyals et al., ‘15)

Input $P = \{P_1, \ldots P_4\}$

Output $O = \{1,4,2,1\}$

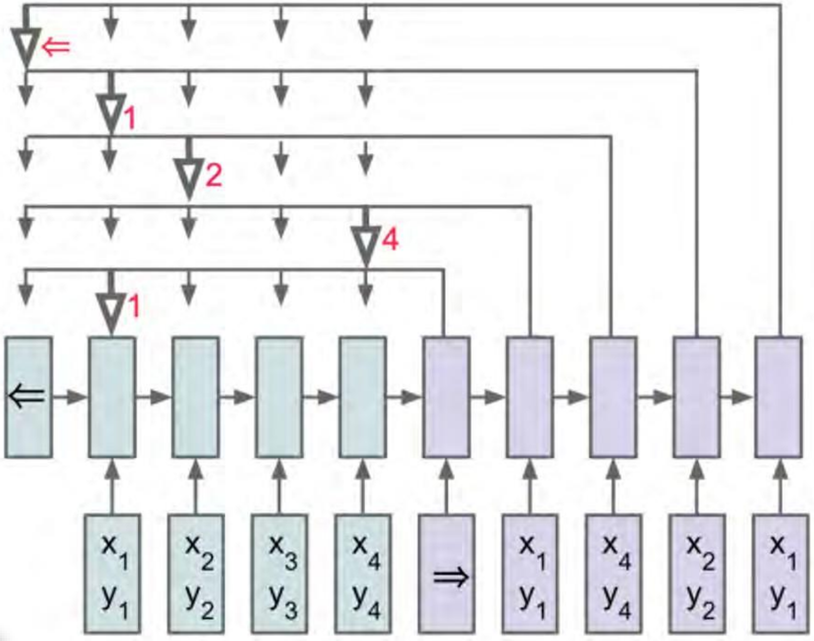
key

query

$$u_{ij} = v^\top \tanh(\overset{\text{attention}}{W}[e_j, d_i])$$

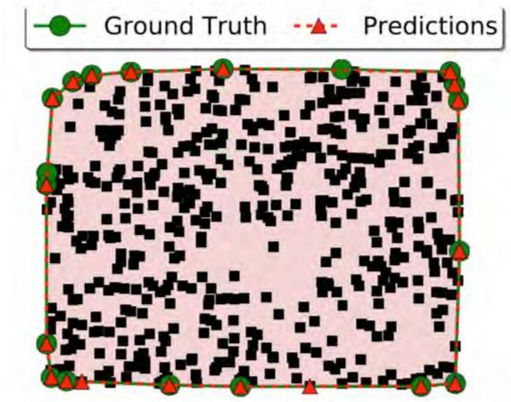
$$p(C_i \mid C_{[1:i-1]}, P) = \mathrm{softmax}(u_i)$$

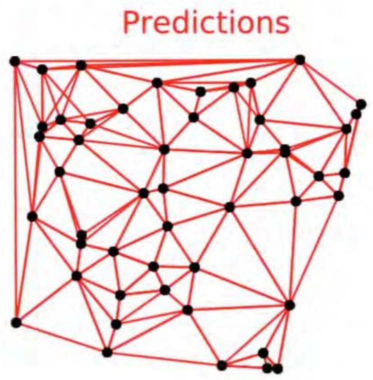
attention weight as prediction distribution

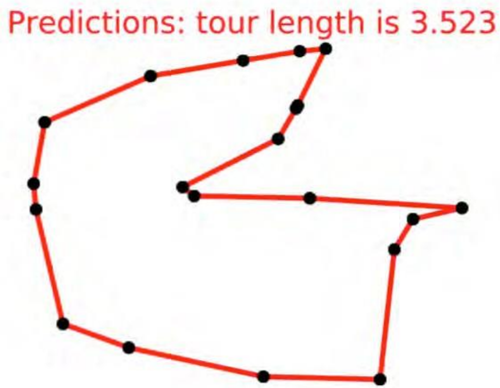


encoder state: e_j decoder state: d_i

Convex hulls, Delaunay triangulation, and TSP







2019 style improvements

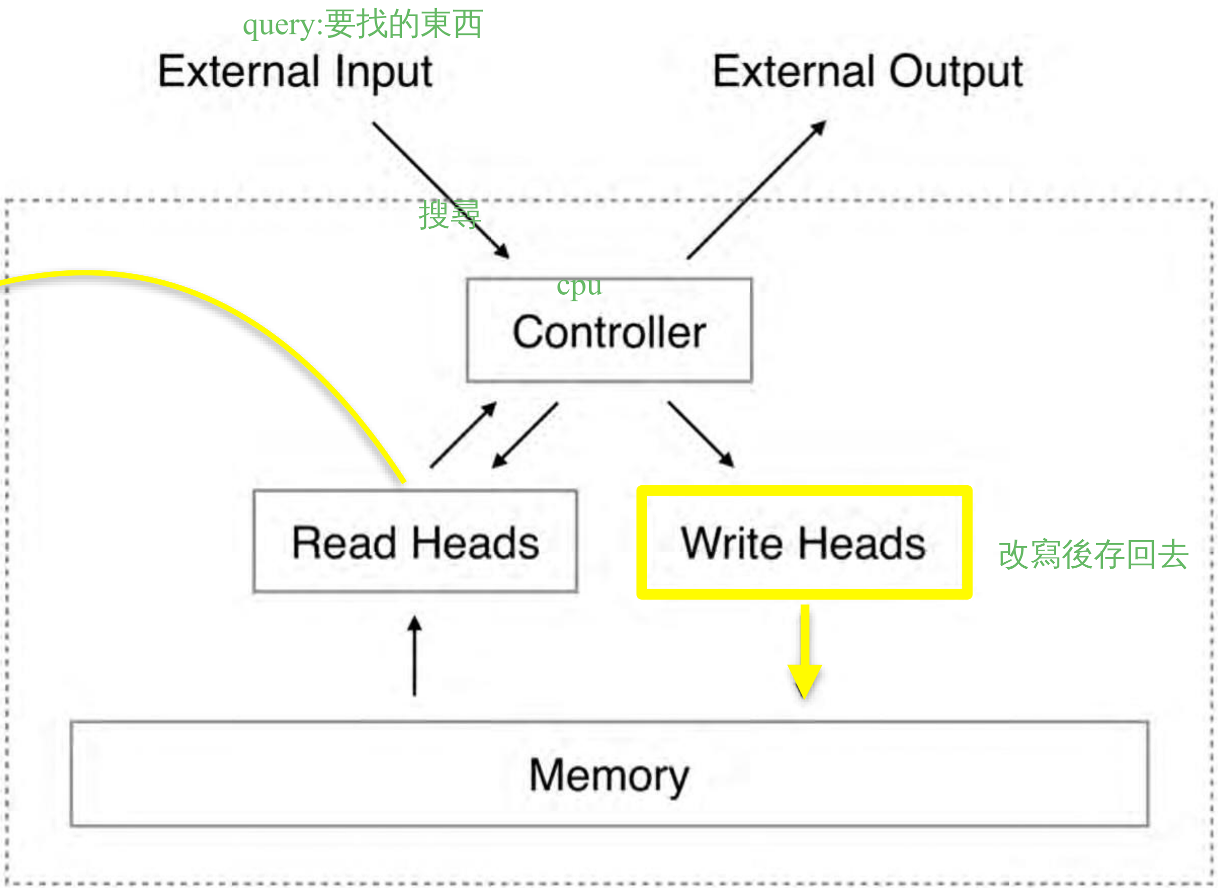
- Transformer to encode inputs (and outputs)
- Graph neural networks for local interactions

Neural Turing Machines (Graves et al., ‘14)

attention weight

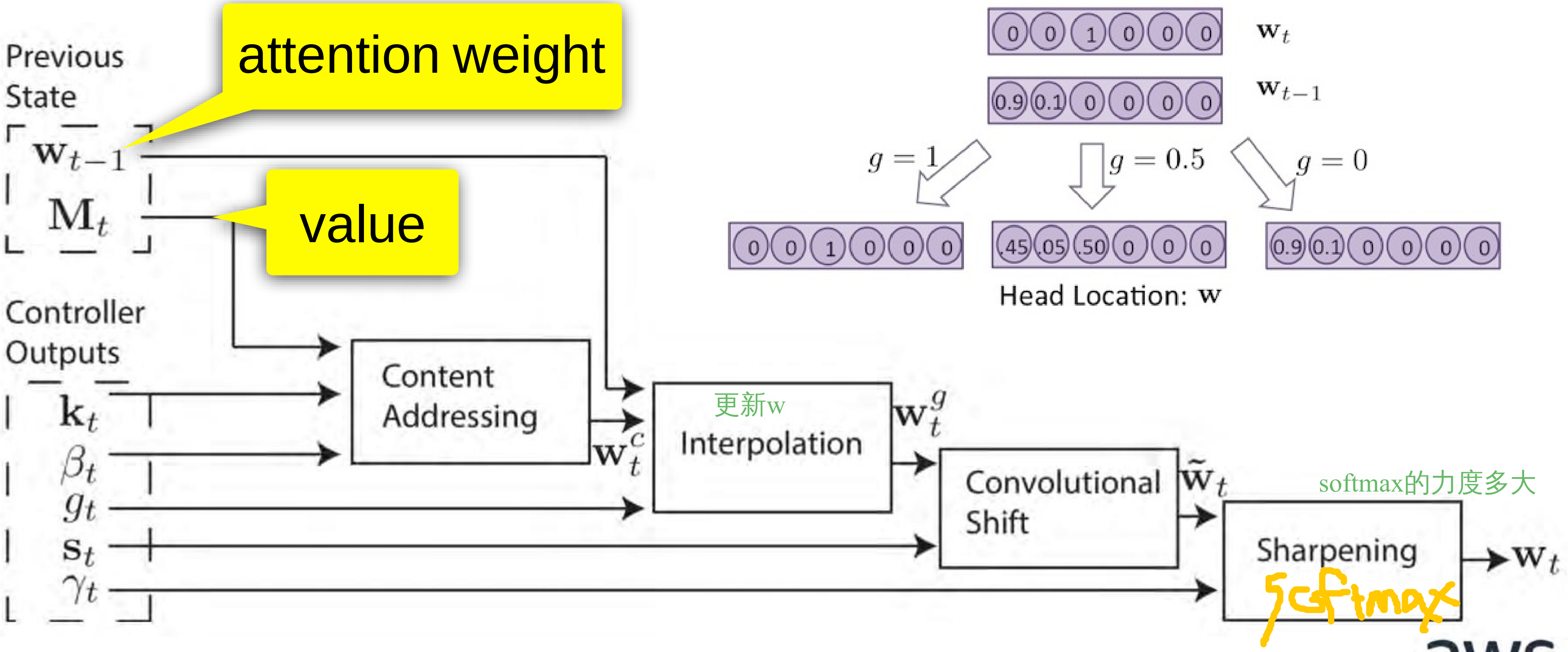
$$r_t \leftarrow \sum_i w_t(i) M_t(i)$$

value



https://d2l.ai

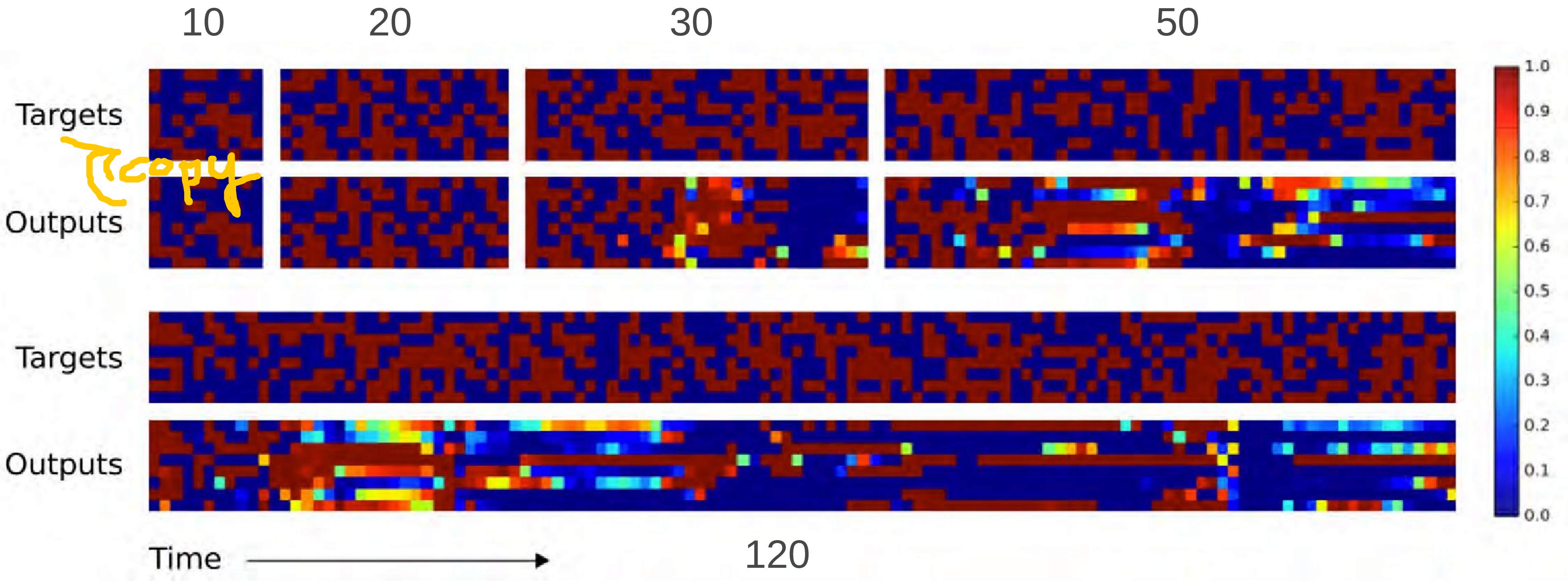
Attention weights can be stateful (values, too)



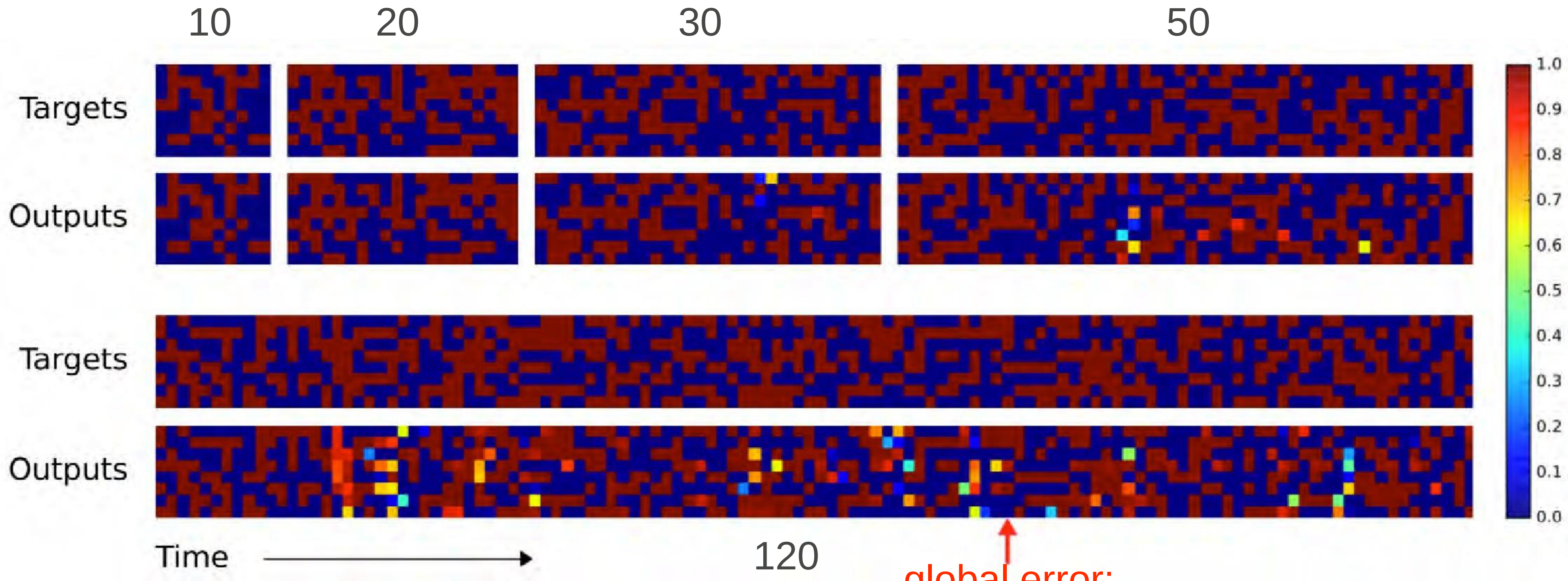
https://d2l.ai

Copying a sequence (with LSTM)

lstm在短的interval時候是表現不錯的，長的就不行



Copying a sequence (with NTM)



global error:
push all subsequent
vectors one step back

https://d2l.ai