

# Ask Me Anything : Dynamic Memory Networks for Natural Language Processing

Ankit Kumar, Peter Ondruska, Mohit Iyyer, James Bradbury, Ishaan Gulrajani, Victor Zhong, Romain Paulus, Richard Socher

모두의연구소 김승일

# 대부분의 NLP 문제는 QA Task로 볼 수 있음

QA Task Sentiment

Classification

POS-Tagging Task

I: Jane went to the hallway.  
I: Mary walked to the bathroom.  
I: Sandra went to the garden.  
I: Daniel went back to the garden.  
I: Sandra took the milk there.  
Q: Where is the milk?  
A: garden  
I: It started boring, but then it got interesting.  
Q: What's the sentiment?  
A: positive  
Q: POS tags?  
A: PRP VBD JJ , CC RB PRP VBD JJ .

# QA Task Example

hallway

W,  
Mary got the football there.

John moved to the bedroom.

Sandra went back to the kitchen.

Mary travelled to the hallway.

John got the football there.

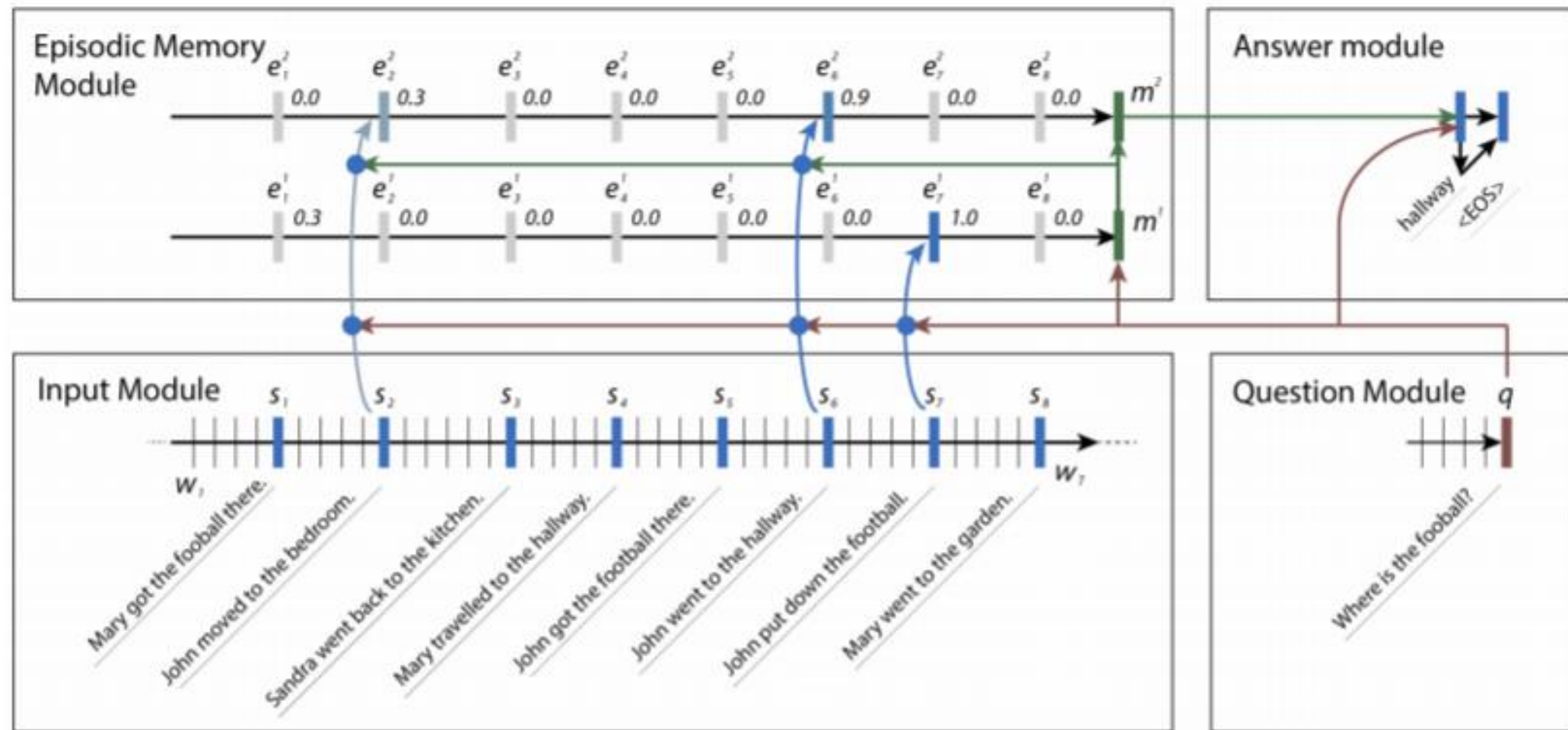
John went to the hallway.

John put down the football.

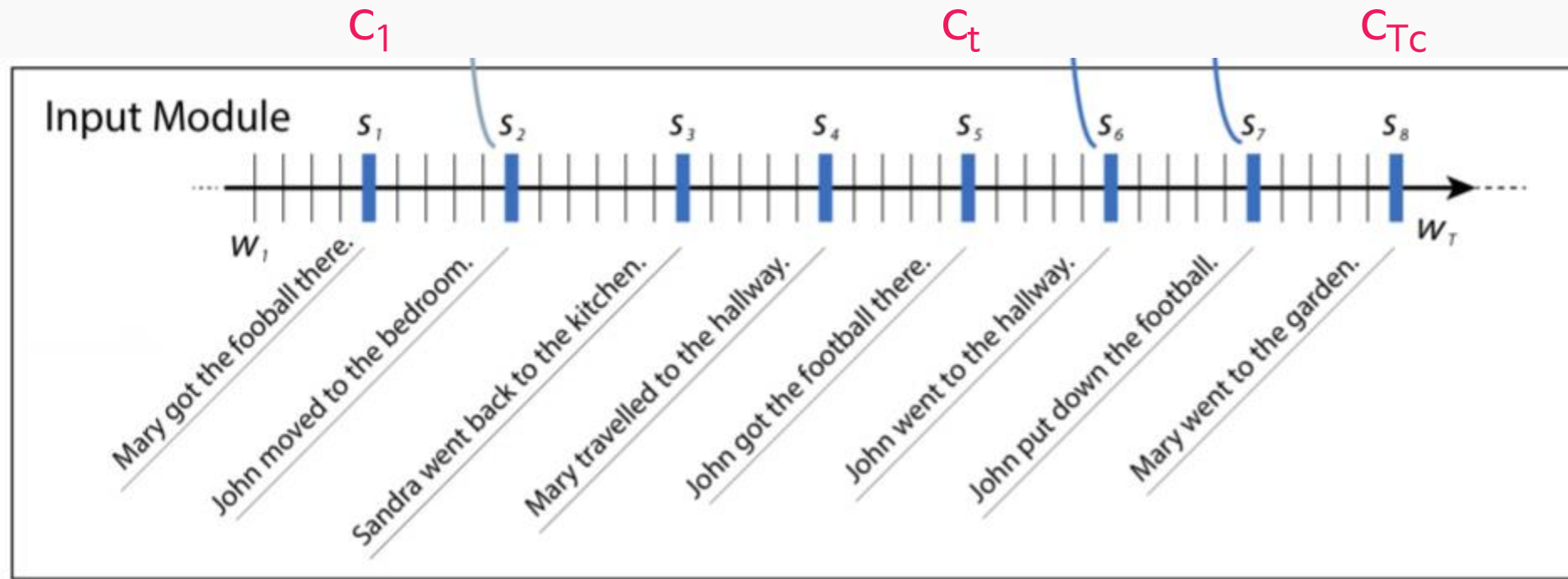
Mary went to the garden.

Where is the football?

# Dynamic Memory Networks



# Dynamic Memory Networks : Input Module



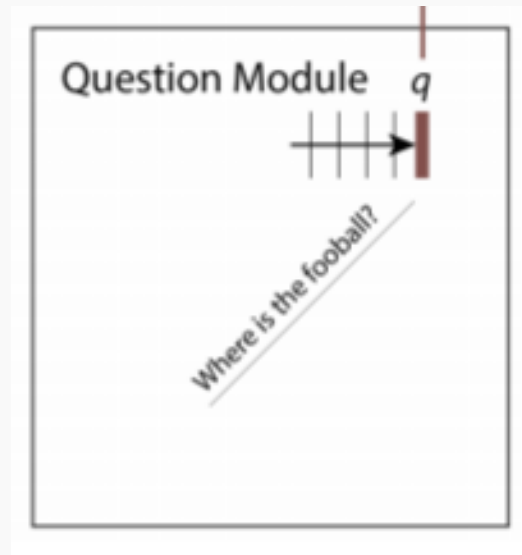
word vector

End-of-sentences tokens

$$h_t = GRU(x_t, h_{t-1}).$$

Output of input module = T sequence of hidden state

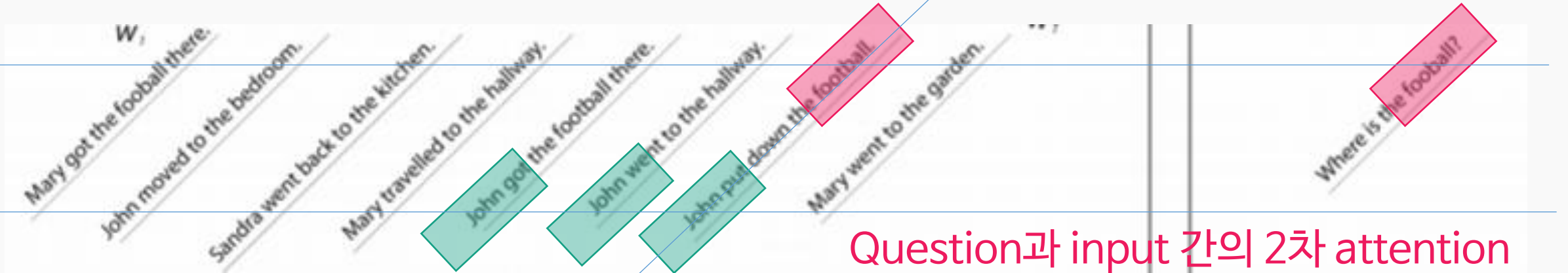
# Dynamic Memory Networks : Question Module



Output of question module = final hidden state

# Dynamic Memory Networks : Episodic Memory Module

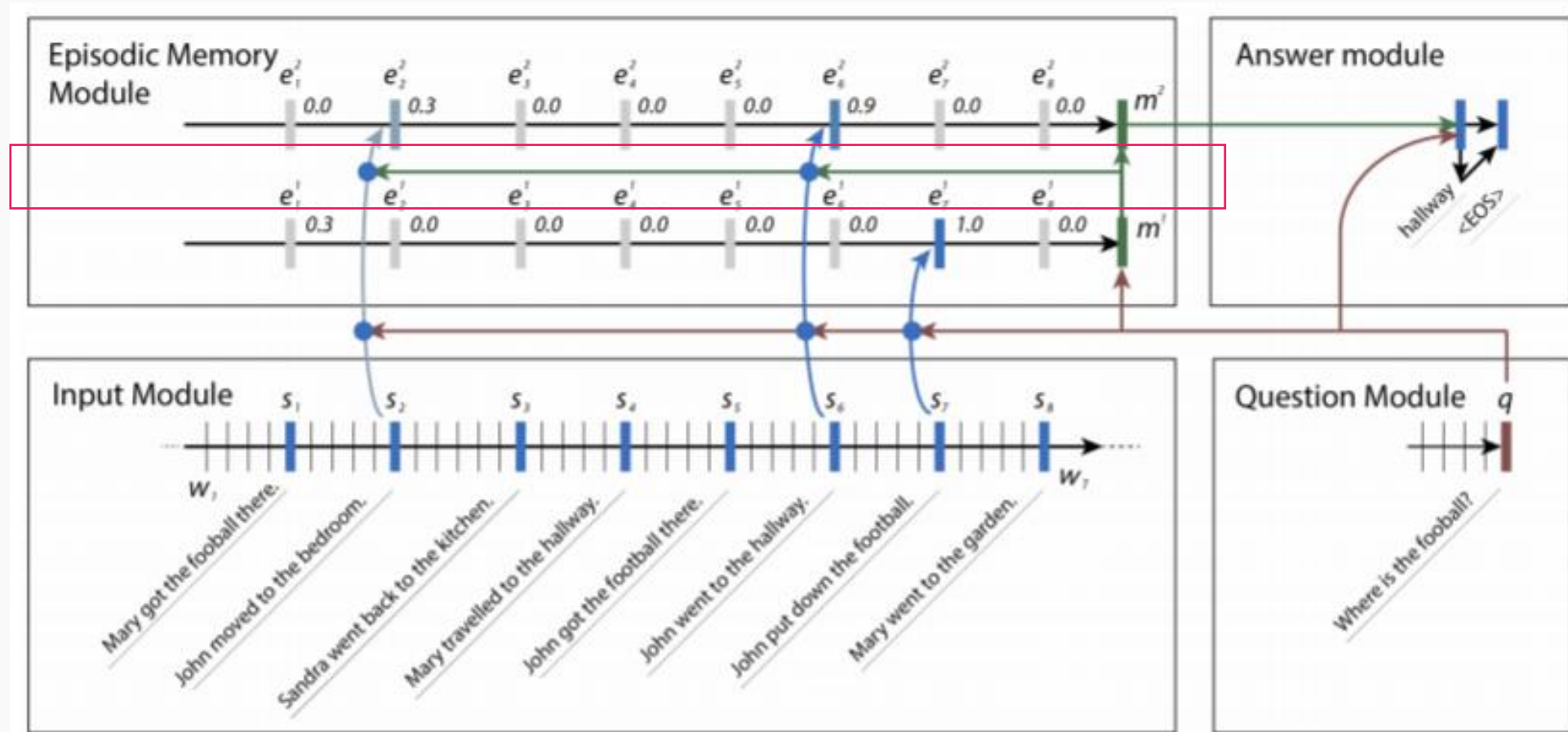
Question과 input 간의 1차 attention



+ 기존의 기억과의 attention

- \* 결국, previous memory, input, question 을 입력으로하는 attention mechanism + multiple hierarchy memory구조가 필요.

# Dynamic Memory Networks : Episodic Memory Module



Input / question / previous memory 간의 attention



# Dynamic Memory Networks : Episodic Memory Module

## attention mechanism

: prev memory, input, question 간의 relation을 다 봐야함.

뭐가 좋은지 모르니 입력으로 그냥 다 때려넣음.

(5) 식을 입력으로 하는 2 Layer NN를 통해 attention weight (여기서는 gating)를 구함

$$\left[ c, m, q, c \circ q, c \circ m, |c - q|, |c - m|, c^T W^{(b)} q, c^T W^{(b)} m \right], \quad (5)$$

$$h_t^i = g_t^i GRU(c_t, h_{t-1}^i) + (1 - g_t^i) h_{t-1}^i \quad (7)$$

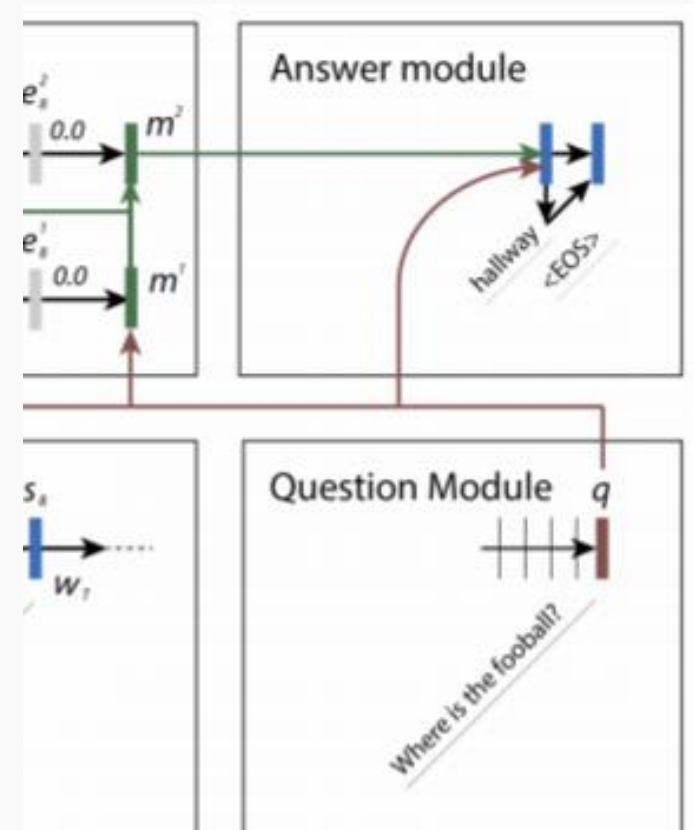
$$e^i = h_{T_C}^i \quad (8)$$

# Dynamic Memory Networks : Answer Module

Initial hidden m1  
Input q (+prev.output)

$$y_t = \text{softmax}(W^{(a)} a_t) \quad (9)$$

$$a_t = \text{GRU}([y_{t-1}, q], a_{t-1}), \quad (10)$$



# Experiments : QA Task

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	<b>93.6</b>

# Experiments : Sentiment Analysis

Task	Binary	Fine-grained
MV-RNN	82.9	44.4
RNTN	85.4	45.7
DCNN	86.8	48.5
PVec	87.8	48.7
CNN-MC	88.1	47.4
DRNN	86.6	49.8
CT-LSTM	88.0	51.0
DMN	<b>88.6</b>	<b>52.1</b>

# Experiments : POS tagging

Model	Acc (%)
SVMTool	97.15
Sogaard	97.27
Suzuki et al.	97.40
Spoustova et al.	97.44
SCNN	97.50
DMN	<b>97.56</b>

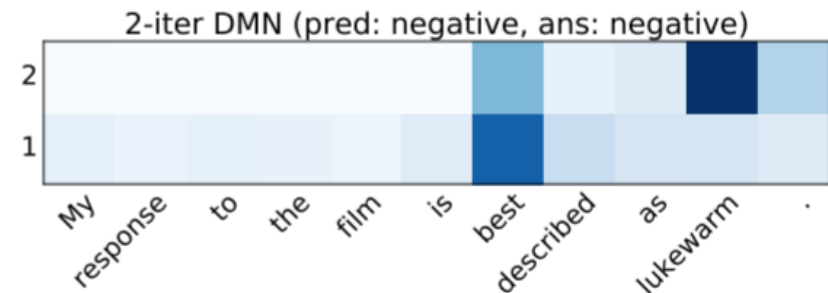
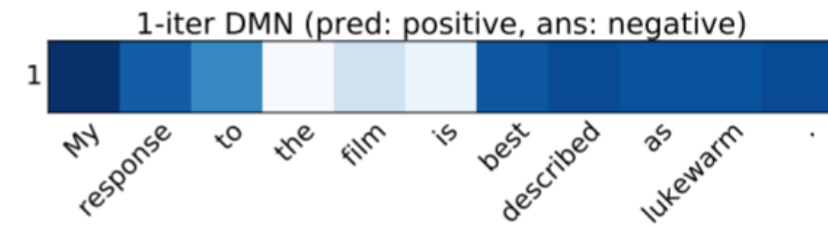
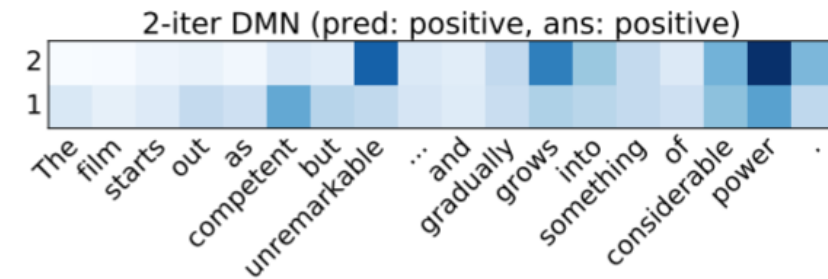
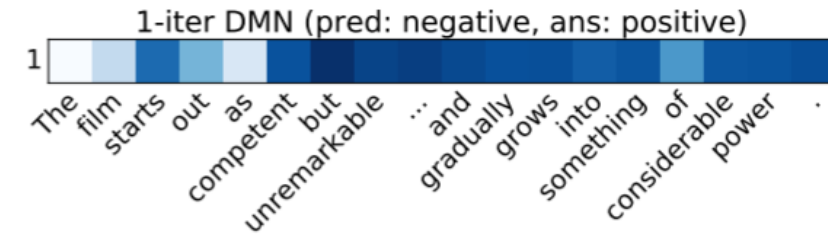
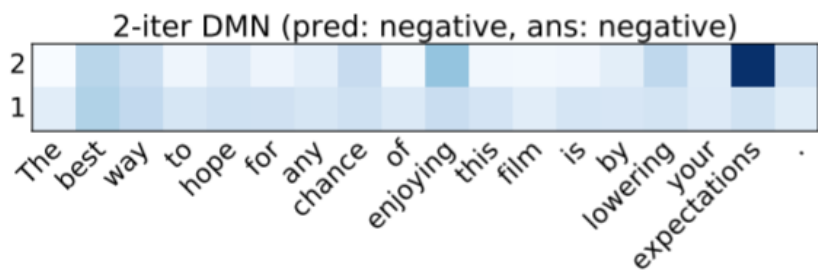
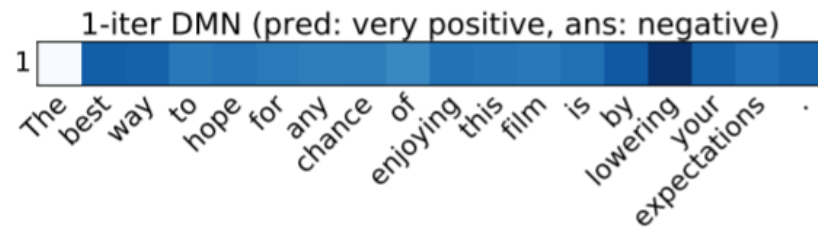
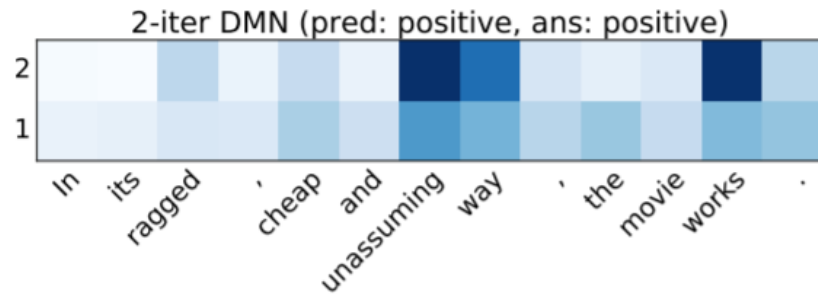
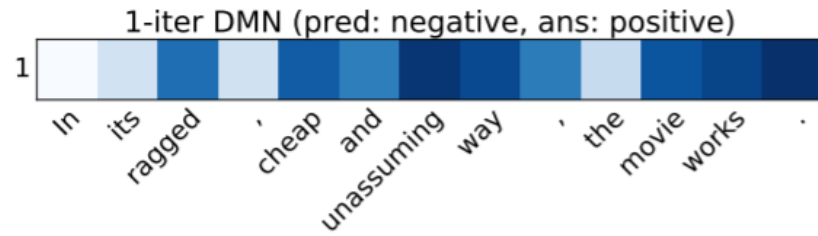
Table 3. Test accuracies on WSJ-PTB

# Experiments

## : Effectiveness of Episodic Memory Module

Max passes	task 3 three-facts	task 7 count	task 8 lists/sets	sentiment (fine grain)
0 pass	0	48.8	33.6	50.0
1 pass	0	48.8	54.0	51.5
2 pass	16.7	49.1	55.6	<b>52.1</b>
3 pass	64.7	83.4	83.4	50.1
5 pass	<b>95.2</b>	<b>96.9</b>	<b>96.5</b>	N/A

# Experiments : Attention Weight





김승일 연구소장

E-mail : [si.kim@modulabs.co.kr](mailto:si.kim@modulabs.co.kr)

FB: [www.facebook.com/lab4all](http://www.facebook.com/lab4all)  
[www.facebook.com/groups/modulabs](http://www.facebook.com/groups/modulabs)

Blog : [www.whyDSP.org](http://www.whyDSP.org)