

2016 한국컴퓨터종합학술대회

시간적인 동적 메모리 네트워크와 데이터 확장을 통한 질의응답 최적화

한동식

이충연

장병탁

서울대학교 컴퓨터공학부

2016년 7월 1일

차례

- 서론
- 동적 메모리 네트워크
- 질의 응답을 위한 DMN 성능 개선 방법
 - ✓ 학습 데이터 확장
 - ✓ 시간 순서 인식 개선
 - ✓ 성능 평가
- 실험 결과 및 논의
- 결론 및 향후 연구

질의 응답

1. Mary moved to the bathroom.
 2. Sandra journeyed to the bedroom.
 3. Mary got the football there.
 4. John went to the kitchen.
 5. Mary went back to the garden.
- Q. Where is the football?

질의 응답

1. Mary moved to the bathroom.
2. Sandra journeyed to the bedroom.
3. Mary got the **football** there.
4. John went to the kitchen.
5. Mary went back to the garden.

Q. Where is the football?

질의 응답

1. Mary moved to the bathroom.
2. Sandra journeyed to the bedroom.
3. Mary got the **football** there.
4. John went to the kitchen.
5. Mary **went back** to the **garden**.

Q. Where is the football?

질의 응답

1. Mary moved to the bathroom.
2. Sandra journeyed to the bedroom.
3. Mary got the **football** there.
4. John went to the kitchen.
5. Mary **went back** to the **garden**.

Q. Where is the football?



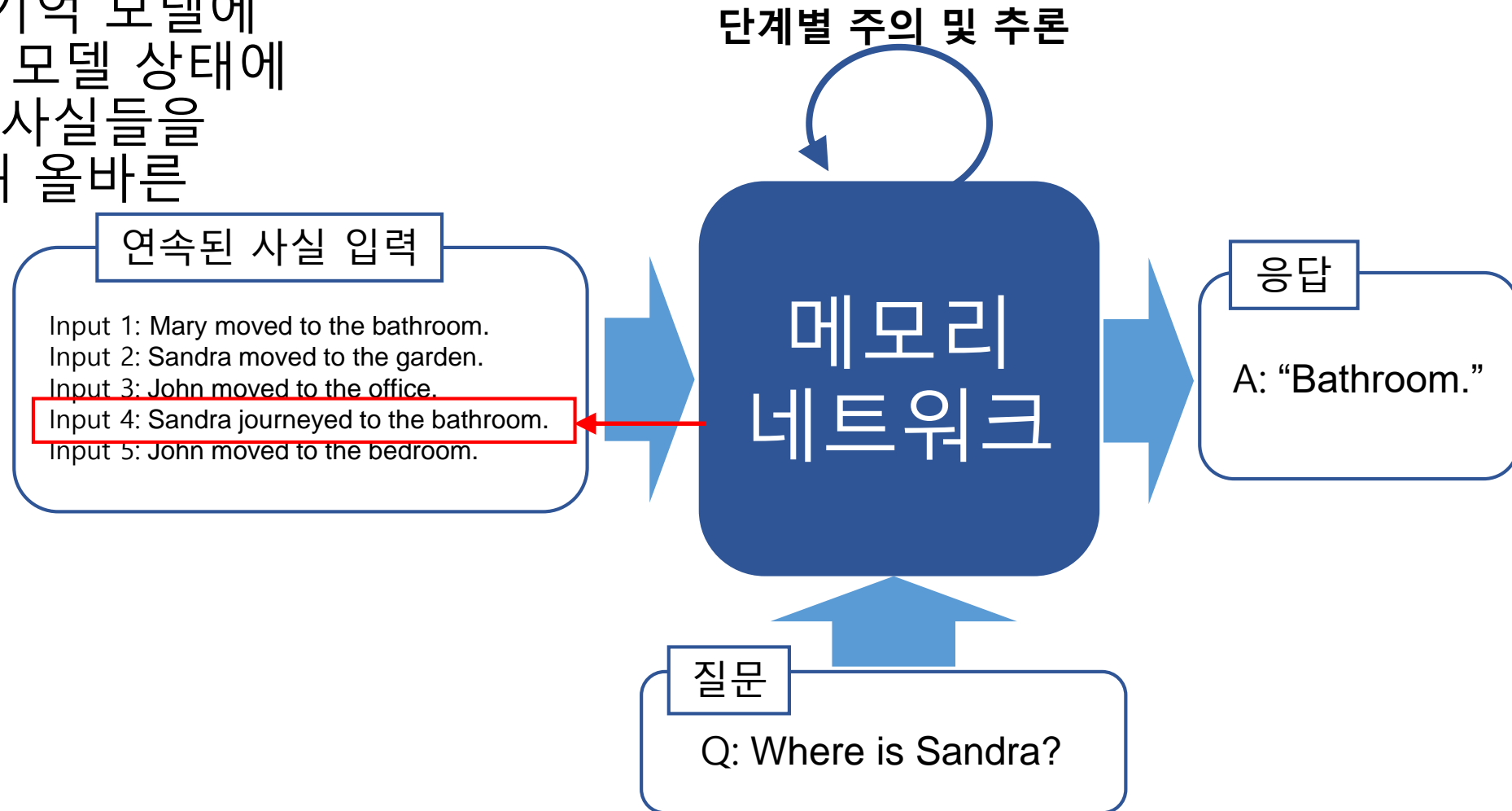
garden

질의 응답

- 새로운 지식을 학습하기 위한 효과적인 방법이다.
- 올바른 응답을 도출하기 위한 인지 체계를 내포하고 있다.

메모리 네트워크

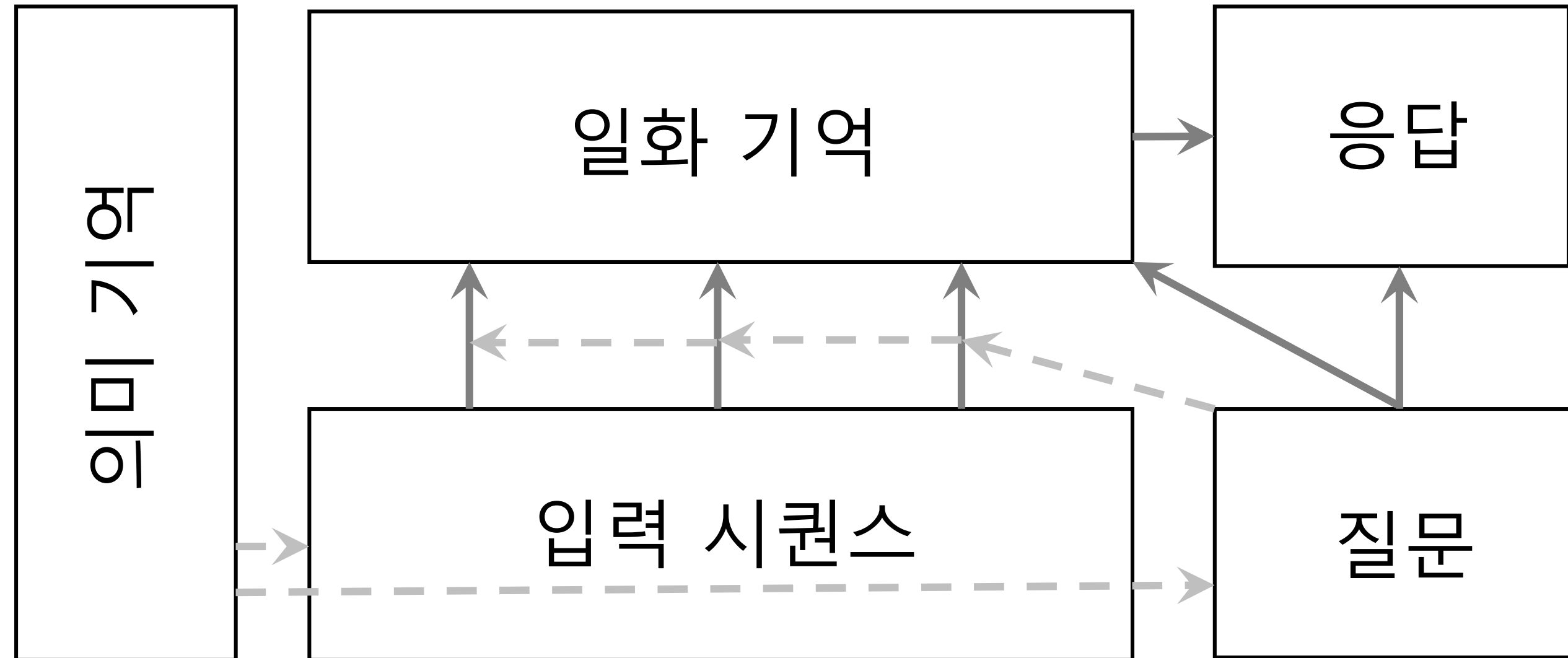
- 사실들의 나열을 기억 모델에 저장한 뒤, 질문과 모델 상태에 따라 가장 적합한 사실들을 순차적으로 선택해 올바른 응답을 나타낸다.



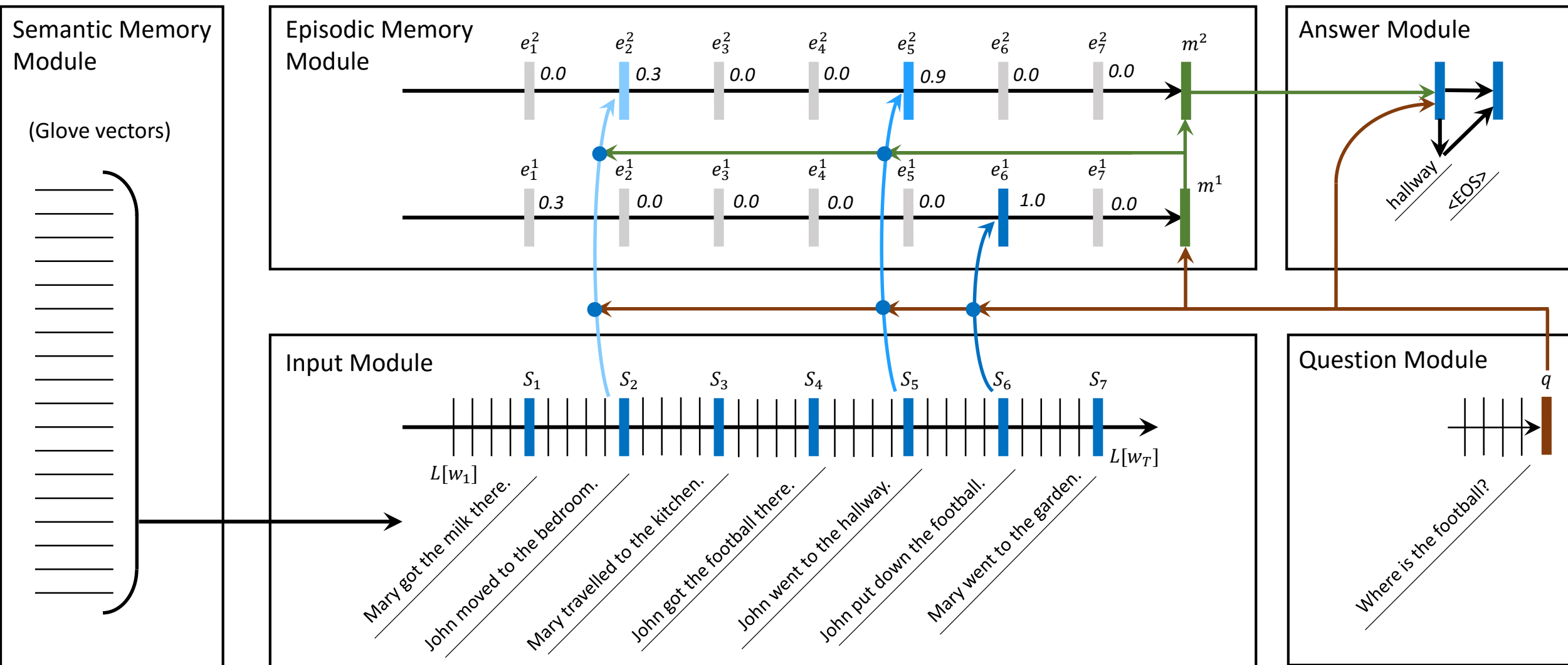
관련 연구

- 모델
 - 메모리 네트워크 – ICLR 2015
 - End-to-End 메모리 네트워크 – NIPS 2015
 - 동적 메모리 네트워크 – ICML 2016
- 질의 응답 데이터
 - bAbI 문제 – ICLR 2016

동적 메모리 네트워크



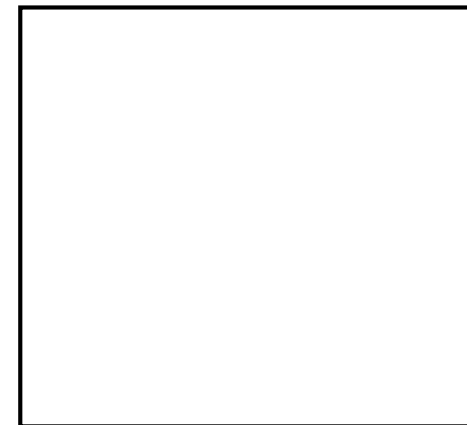
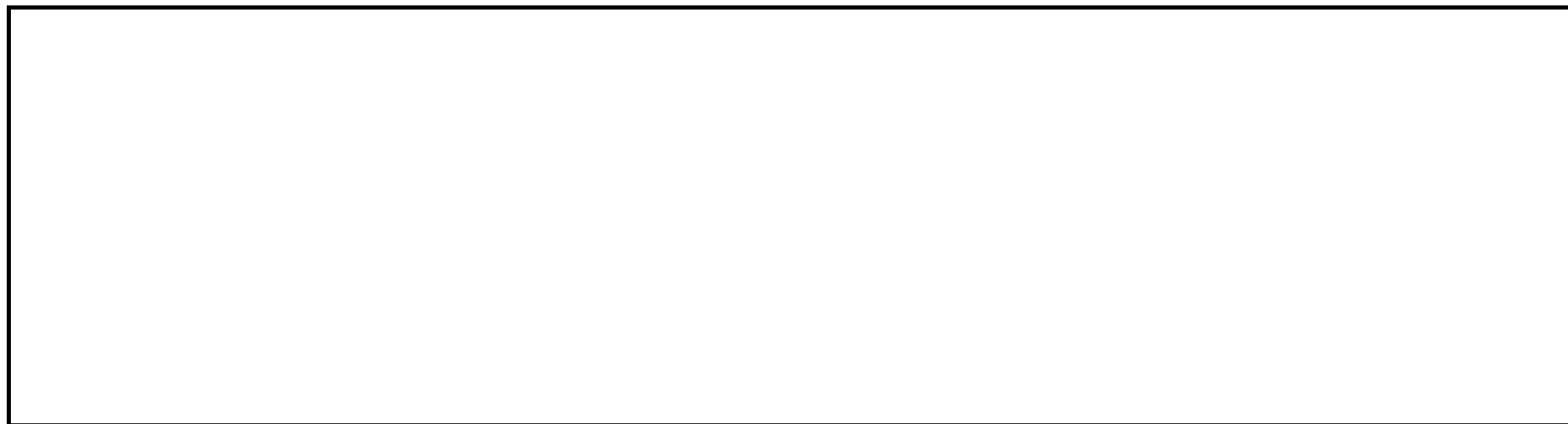
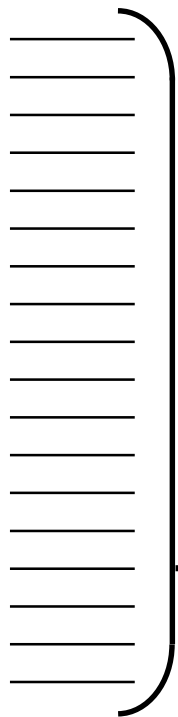
동적 메모리 네트워크



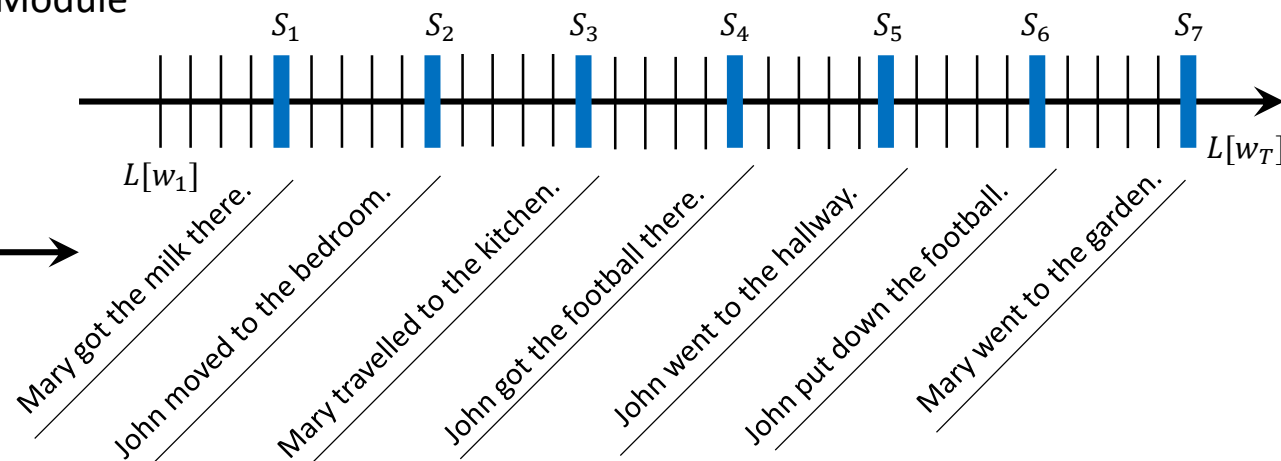
의미 기억 모듈, 입력 모듈, 질문 모듈

Semantic Memory
Module

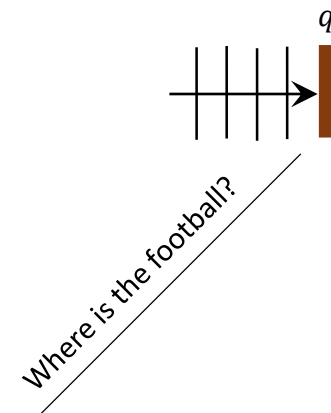
(Glove vectors)



Input Module



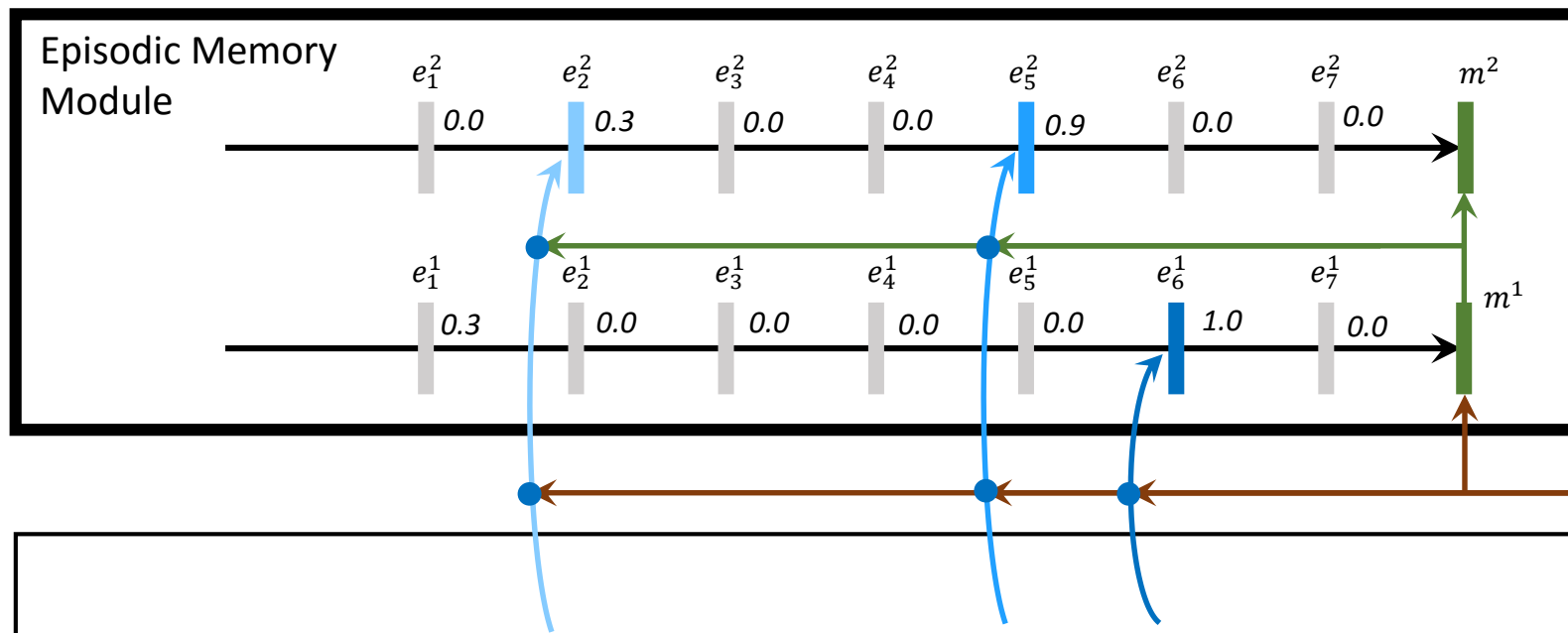
Question Module



의미 기억 모듈, 입력 모듈, 질문 모듈

- 의미 기억 모듈: 단어 임베딩 벡터. $L = \mathbb{R}^{W \times 300}$ 의 행렬을 통해 단어의 벡터 표현 생성
- 입력 모듈: 문장들의 입력을 벡터의 집합 $F = [f_1, \dots, f_N]$ 로 표현해 내는 것을 목적으로 함
 - 입력 문장의 단어들을 단어 임베딩을 통해 고정된 차원의 벡터들로 변환
 - 회귀 신경망을 통해 문장마다 최종 은닉 상태들을 추출
 - 한 문장 내의 i 번째 단어 w_i 와 전 은닉 상태 h_{i-1} 에 대해 Gated Recurrent Unit (GRU)에 의한 새로운 은닉 상태 $h_t = GRU(L[w_t], h_{t-1})$ 를 구하는 과정을 문장 내 마지막 단어까지 반복
- 질문 모듈
 - 본 모듈은 질문의 벡터 표현인 q 를 $q_t = GRU(L[w_t^q], q_{t-1})$ 로 계산
 - 실제 구현에서는 입력 모듈과 모든 변수를 공유

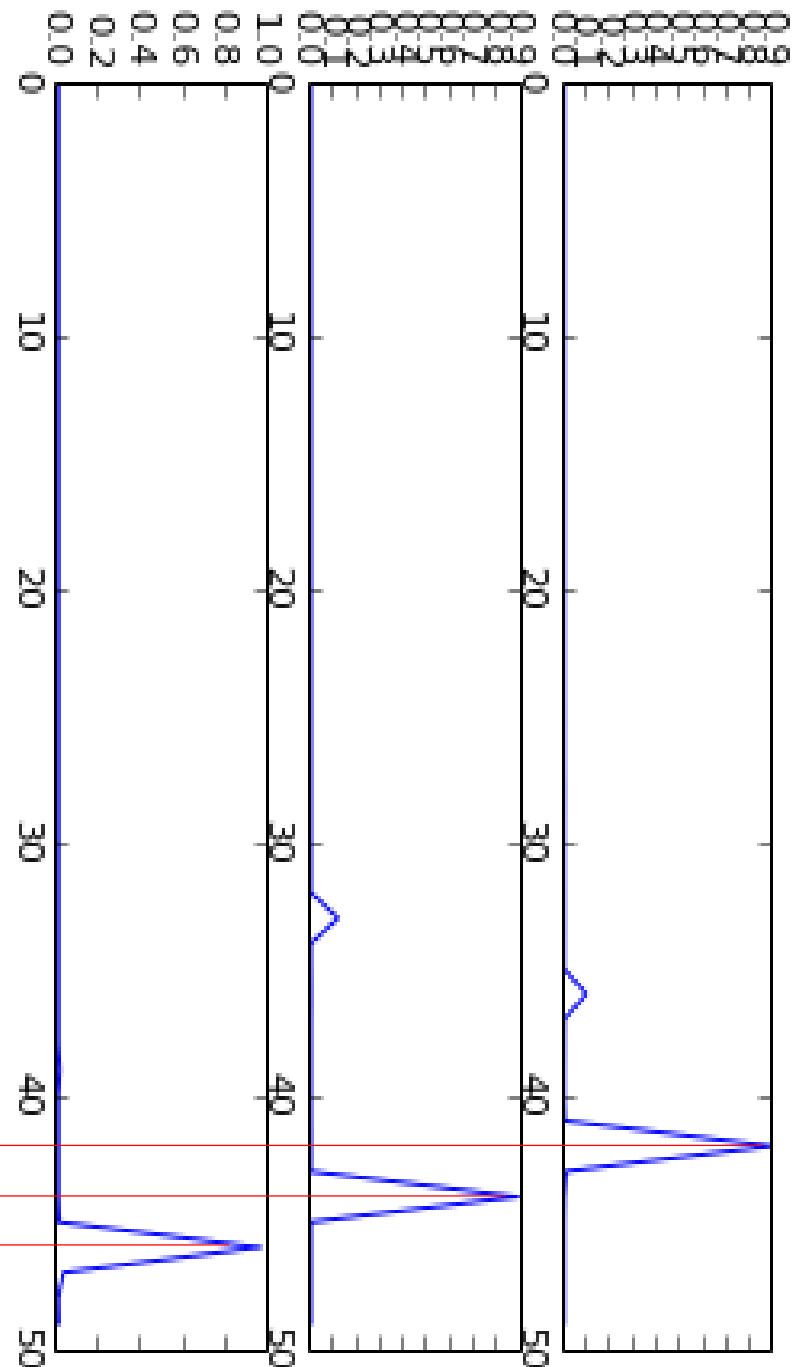
일화 기억 모듈



일화 기억 모듈의 주의 과정

0: mary picked up the apple .
 1: sandra went back to the kitchen .
 2: mary discarded the football there .
 3: mary went to the office .
 4: mary dropped the apple .
 5: mary picked up the apple .
 6: daniel moved to the garden .
 7: daniel moved to the bathroom .
 8: mary dropped the apple .
 9: mary took the apple .
 10: mary travelled to the hallway .
 11: mary journeyed to the office .
 12: sandra journeyed to the bathroom .
 13: mary dropped the apple .
 14: mary grabbed the apple .
 15: john moved to the kitchen .
 16: mary put down the apple .
 17: sandra moved to the hallway .
 18: john grabbed the milk there .
 19: mary got the apple .
 20: john left the milk .
 21: daniel went back to the bedroom .
 22: daniel went back to the garden .
 23: mary travelled to the hallway .
 24: john moved to the office .
 25: sandra went back to the bedroom .
 26: john moved to the kitchen .
 27: mary went to the bedroom .
 28: sandra moved to the office .
 29: mary went back to the kitchen .
 30: mary picked up the milk .
 31: mary went to the bedroom .
 32: mary moved to the kitchen .
 33: sandra moved to the garden .
 34: mary left the milk .
 35: mary got the milk .
 36: sandra journeyed to the bedroom .
 37: mary went back to the hallway .
 38: daniel journeyed to the office .
 39: sandra grabbed the football there .
 40: mary left the milk .
 41: daniel went back to the kitchen .
 42: sandra journeyed to the bathroom .
 43: daniel went to the hallway .
 44: sandra went to the garden .
 45: mary grabbed the milk there .
 46: sandra put down the football .
 47: mary put down the apple .
 48: daniel travelled to the bathroom .
 49: mary got the apple there .
 where was the football before the garden ?
 =====

The model predicted 'bathroom', where the target is 'bathroom'



일화 기억 모듈

- q 가 요구하는 응답을 하기 위해 필요한 사실들을 순차적으로 검색
- 질문과 입력에 대한 이해를 더 높이기 위한 이 과정은 일화적 통과(episodic pass)라 불리며 이를 통해 일화 메모리 벡터를 추출한다. 초기 메모리 벡터가 $m^0 = q$ 일 때, i 번째 일화 기억 모듈을 통과하여 생성되는 m^i 는 다음과 같다.

$$z_t^i = [f, m^{i-1}, q, f \circ q, f \circ m^{i-1}, |f - q|, |f - m^{i-1}|, f^T W^{(b)} q, f^T W^{(b)} m^{i-1}]$$

사실과 일화 메모리, 질문간의 관계 계산

$$g_t^i = W^{(2)} \tanh(W^{(1)} z_t^i + b^{(1)}) + b^{(2)}$$

관계를 통해 episodic gate (스칼라값) 계산

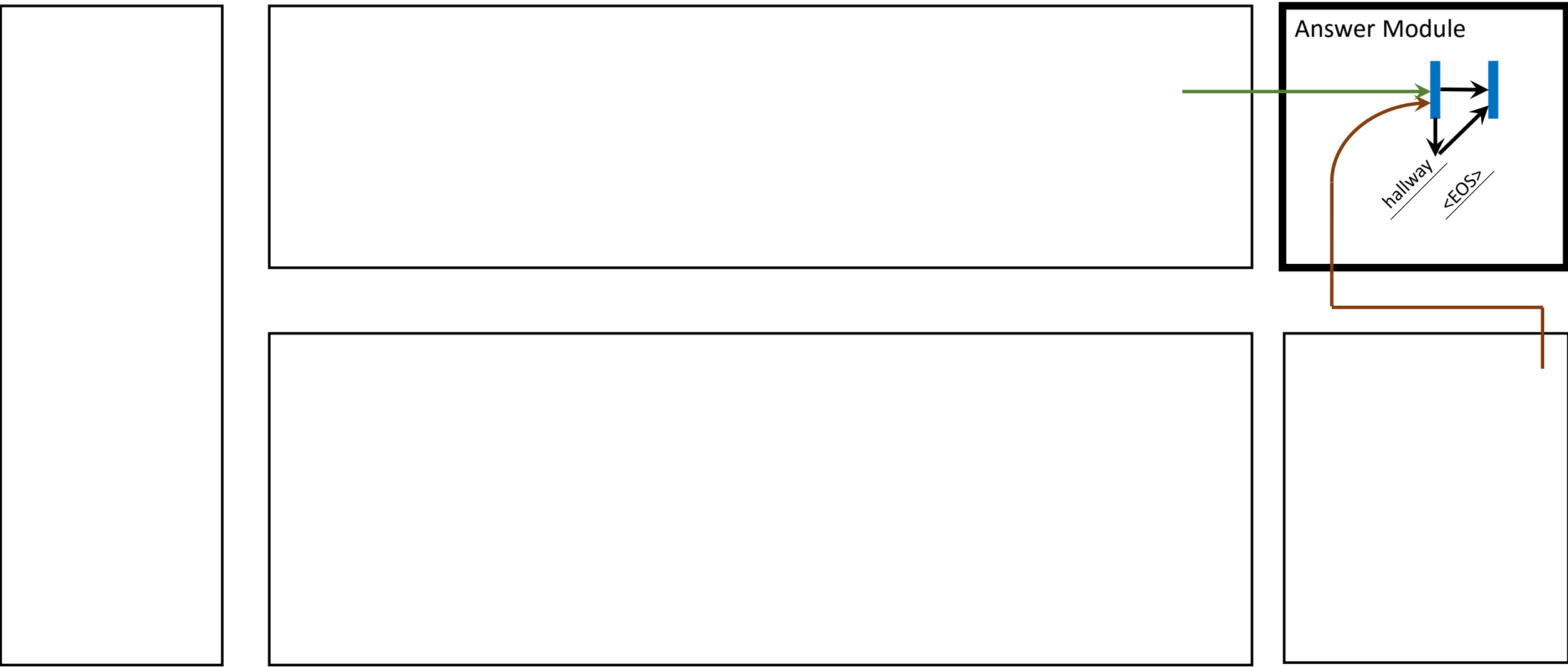
$$e^i = \sum_{t=1}^T \left(\frac{\exp(g_t^i)}{\sum_{j=1}^T \exp(g_j^i)} \right) f_t$$

소프트맥스 함수를 이용해 사실 검색, 요약

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

Episodic pass 수에 따라 GRU계산 반복

응답 모듈



응답 모듈

- 질문 모듈의 q 와 일화 기억 모듈에서 요약된 $m^{T_{pass}}$ 를 이용하여 올바른 정답을 예측한다.
- 소프트맥스 층을 통해 단어 수준의 답변 형식, 회귀 신경망을 사용하여 연속적 단어 답변을 얻을 수 있다.

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

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

질의 응답을 위한 DMN 성능 개선 방법

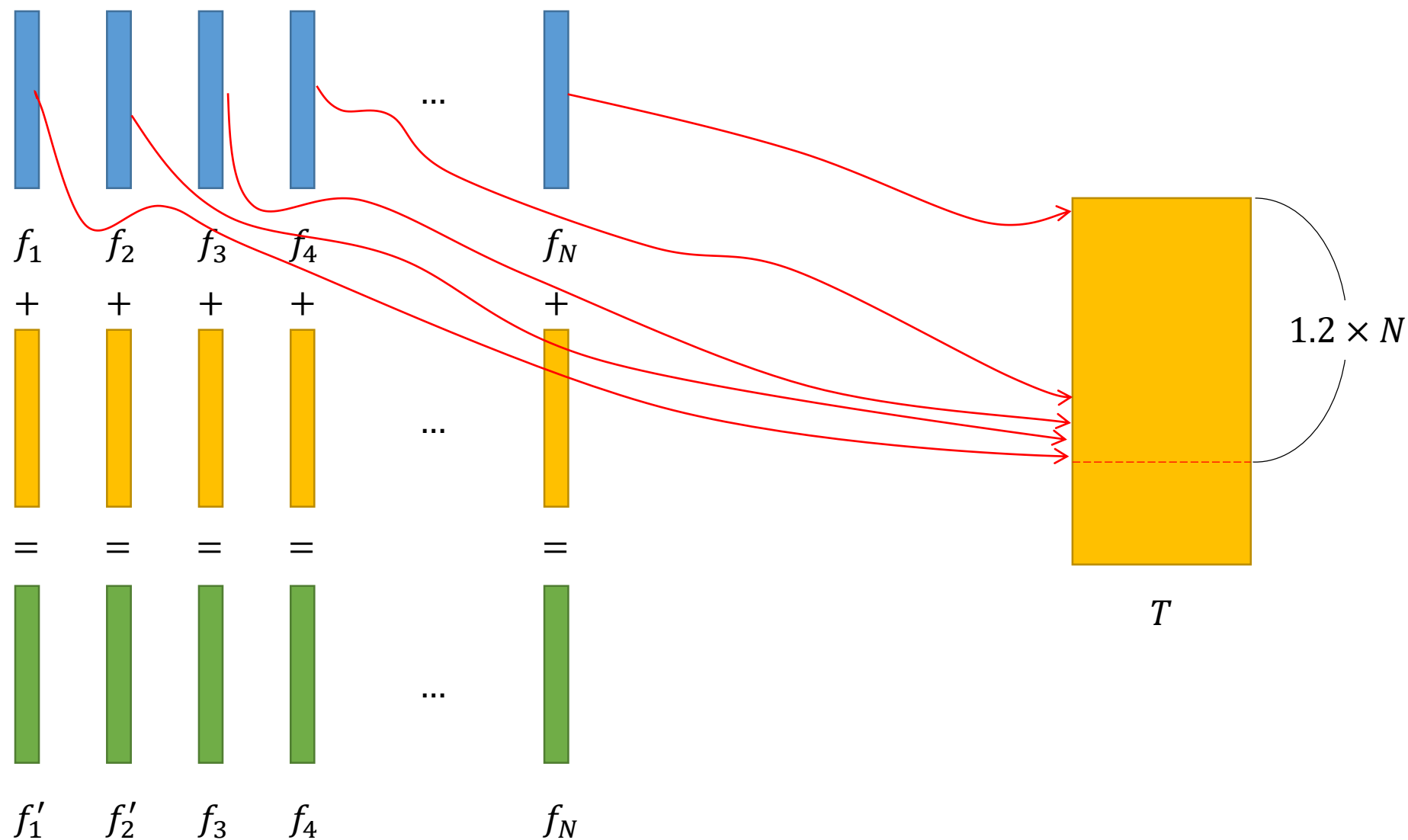
- 1K bAbI 문제의 20개 과제들 (학습 데이터 20,000 개, 테스트 데이터 20,000개)
- 오류 역전파 알고리즘
- Adam 최적화 방법

학습 데이터 개선

- 데이터가 충분치 않을 경우 학습 데이터에 과적합 문제가 발생
- 어휘 셔플링(vocabulary shuffling): 학습시 이름, 장소, 사물 등 같은 범주로 사용되는 어휘끼리 매 학습 횟수 마다 임베딩 벡터의 연결을 임의로 바꾸어 매번 새로운 일화를 학습하는 효과에 가깝도록 함

기술	원본 데이터	생성된 데이터
어휘 셔플링	1 <u>Mary</u> moved to the <u>bathroom</u> . 2 <u>John</u> went to the <u>hallway</u> . Q: Where is <u>Mary</u> ? A: <u>bathroom</u>	1 <u>Lily</u> moved to the <u>hallway</u> . 2 <u>Daniel</u> went to the <u>kitchen</u> . Q: Where is <u>Lily</u> ? A: <u>hallway</u>

시간 순서 개선



시간 순서 개선

- 기존 DMN이 가진 주의 기제의 취약점 중 하나는 명시적인 상대 시간 정보를 이용하지 않는다는 점
- 따라서 DMN의 주의 기제는 질문이 의도한 시점이 아닌 과거 혹은 미래 시점의 사실들을 찾을 가능성이 높을 수 있다.
- 본 연구에서는 시간 임베딩(temporal embedding)을 채용한다.
- T 라는 행렬에 대해 다음과 같이

$$f'_i = f_i + T(N - i + 1)$$

$N - i + 1$ 의 행 벡터를 합한 상대 시간에 의존적인 $F' = [f'_1, \dots, f'_N]$ 를 일화 기억 모듈에 사용

- 실제 학습에서는 시퀀스 여분의 20% 만큼 랜덤하게 상대적 순서를 유지하면서 index를 재배열한다.

성능 평가

- Strongly supervised DMN: 정답과 이 정답을 도출하기 위한 근거 문장 정보를 학습에 사용.

$$J = \alpha E_{CE}(GATES) + \beta E_{CE}(Answers)$$

where $\alpha + \beta = 1$

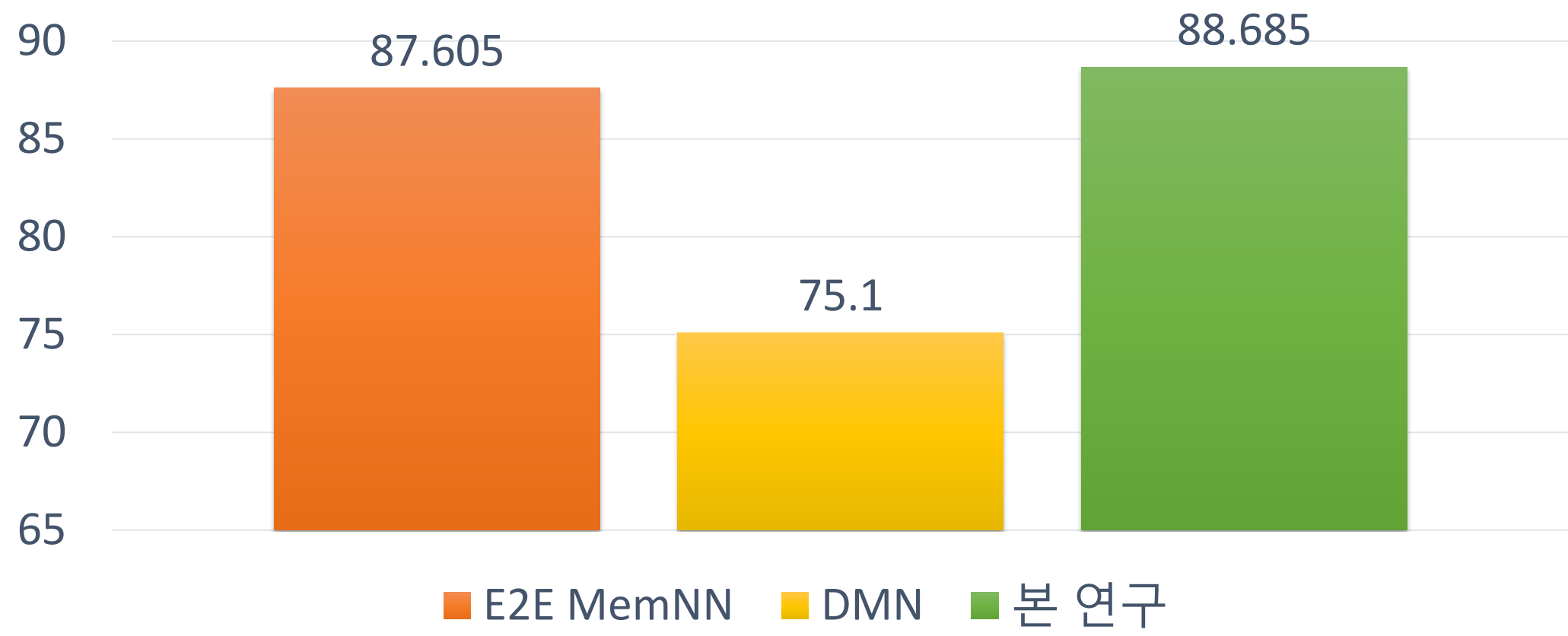
- Weakly supervised DMN: 정답만을 목적으로 신경망을 학습

$$J = E_{CE}(Answers)$$

질의 응답을 위한 DMN 성능 개선 방법

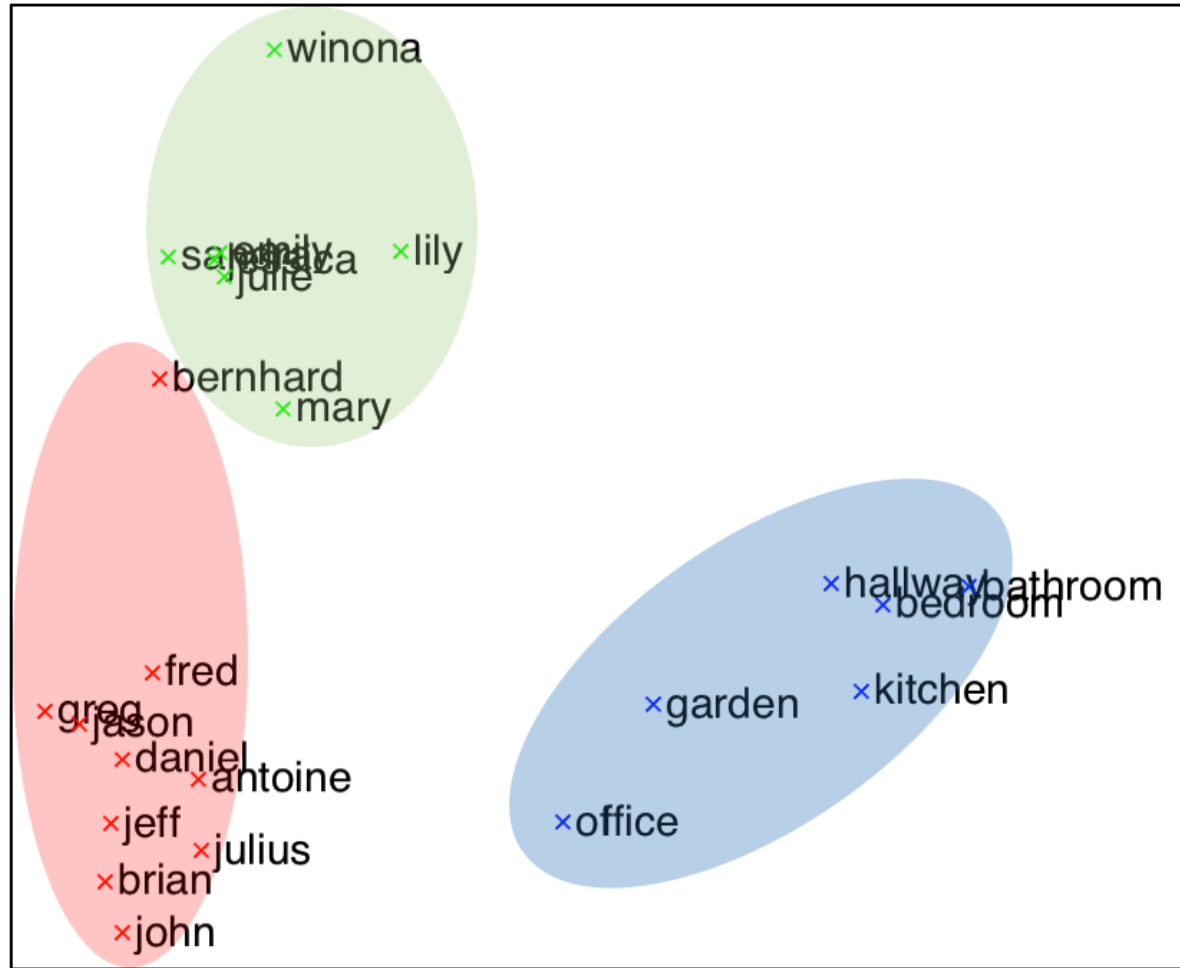
과제	비교 모델		개선된 DMN	
	End-to-End MemNN	DMN	Strongly Supervised DMN	Weakly Supervised DMN
1: 1 supporting fact	100	100	98.6	98.6
2: 2 supporting fact	88.6	39.7	72.3	78.6
3: 3 supporting fact	78.1	41.5	69.3	84.2
4: 2 argument relations	86.6	75.5	98	97.5
5: 3 argument relations	85.6	50.1	99	98.9
6: yes/no questions	97.2	97.7	96	99.2
7: counting	81.7	91.4	82.7	87.4
8: lists/sets	90.7	95.2	88.2	95.3
9: simple negations	98.1	99	96.4	98
10: indefinite knowledge	93.5	87.3	88.8	95.4
11: basic coreference	99.7	100	91.1	85.3
12: conjunction	99.9	87	98	96
13: compound coreference	99.8	97.4	93	92.4
14: time reasoning	93.1	73.1	97.8	98.3
15: basic deduction	100	53.9	87.6	98.2
16: basic induction	97.3	49.5	95.6	65.9
17: positional reasoning	59.6	59.3	59.6	63
18: size reasoning	90.6	98.3	97.1	96.6
19: path finding	12	9	29.1	45.6
20: agent's motivation	100	97.1	100	99.3
평균 정답률 (%)	87.605	75.1	86.91	88.685
통과 과제 (정답률 > 95%)	9	8	10	12

1K bAbI 문제에서의 평균 정답률

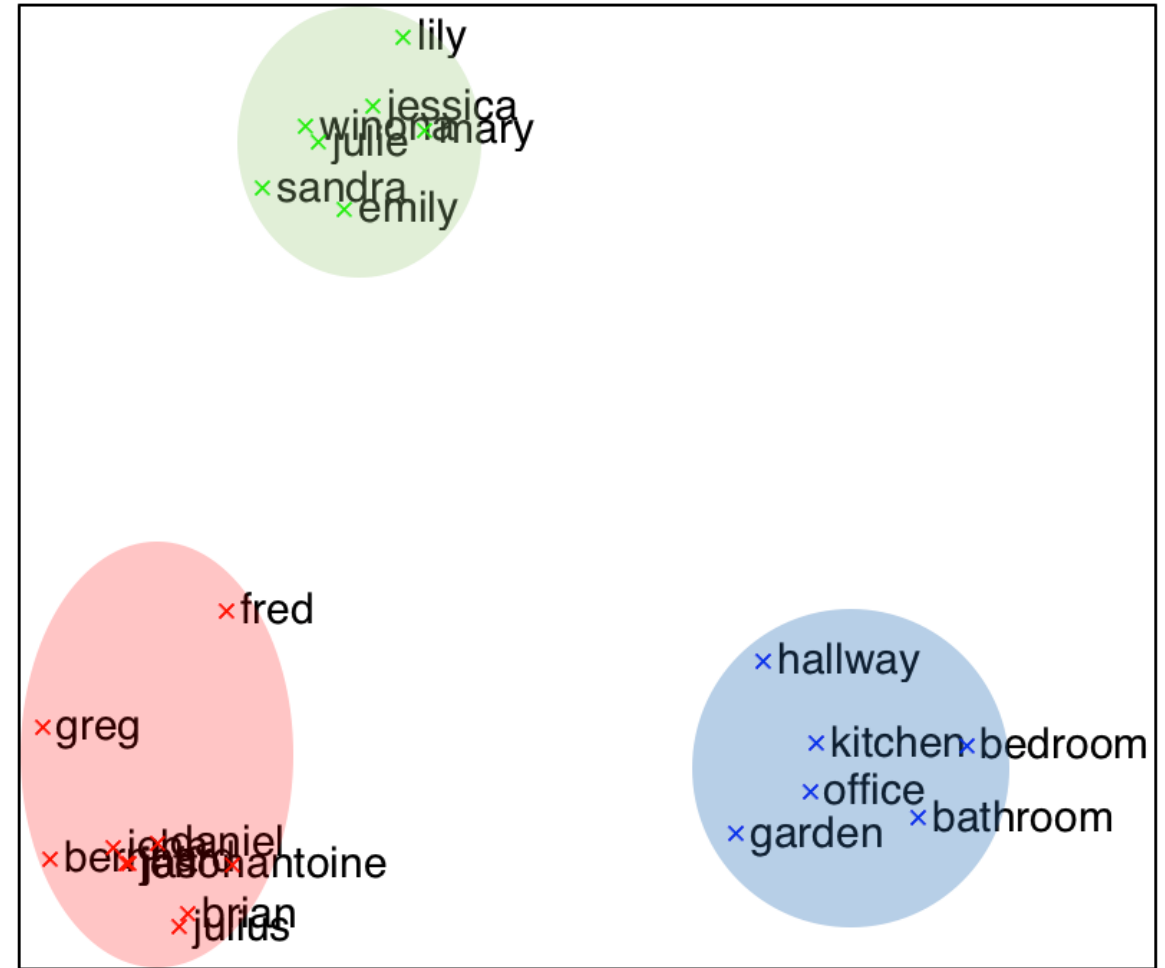


의미 기억 모듈의 명사 단어 지식 분석

GloVe

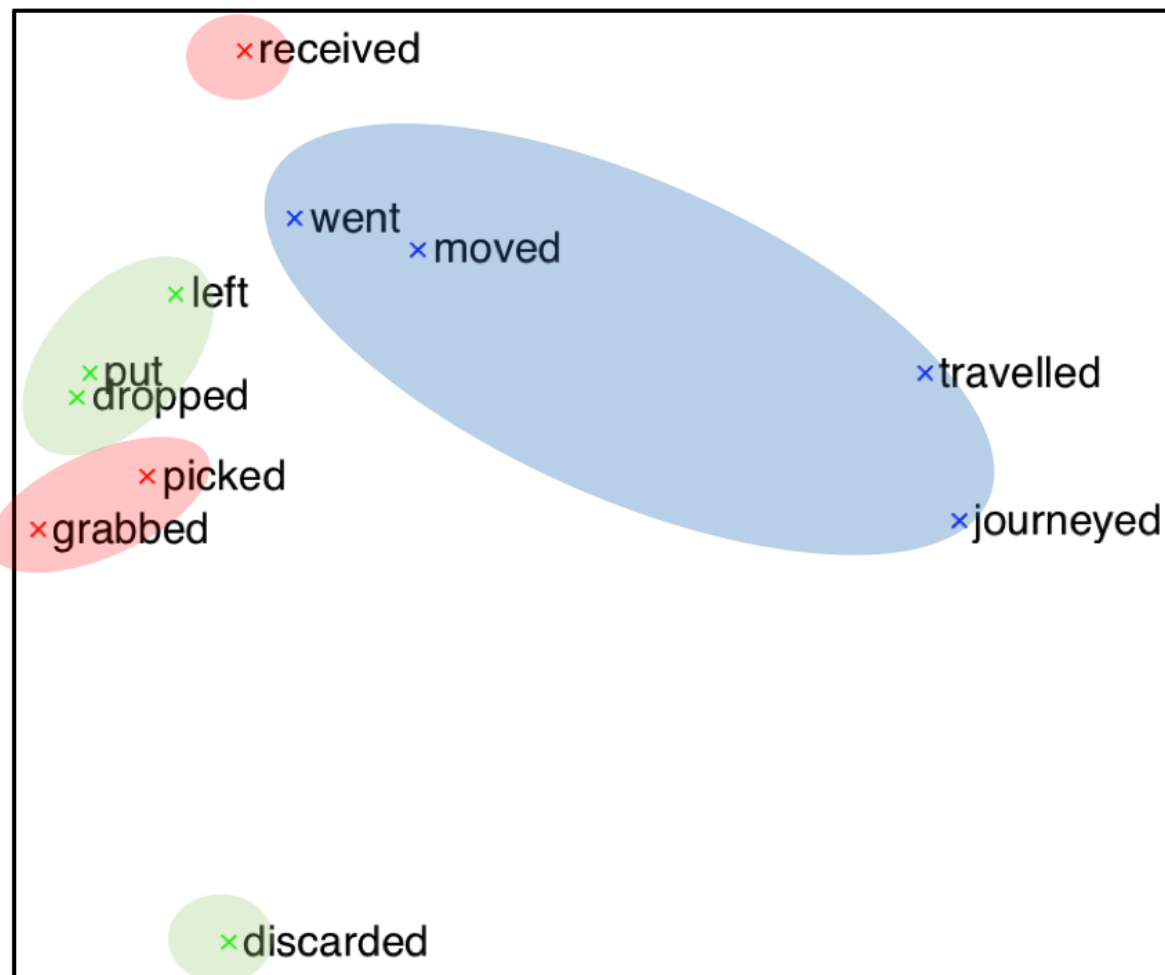


DMN

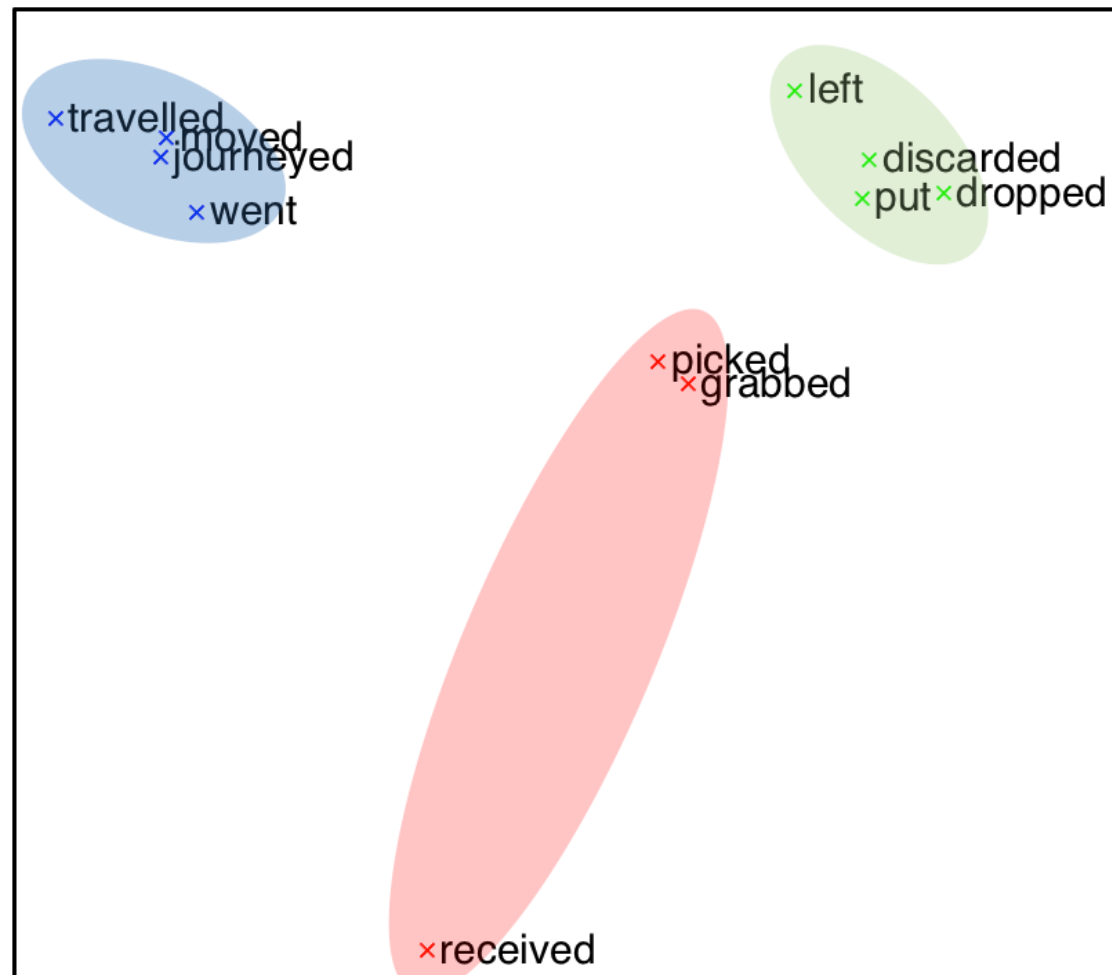


의미 기억 모듈의 동사 단어 지식 분석

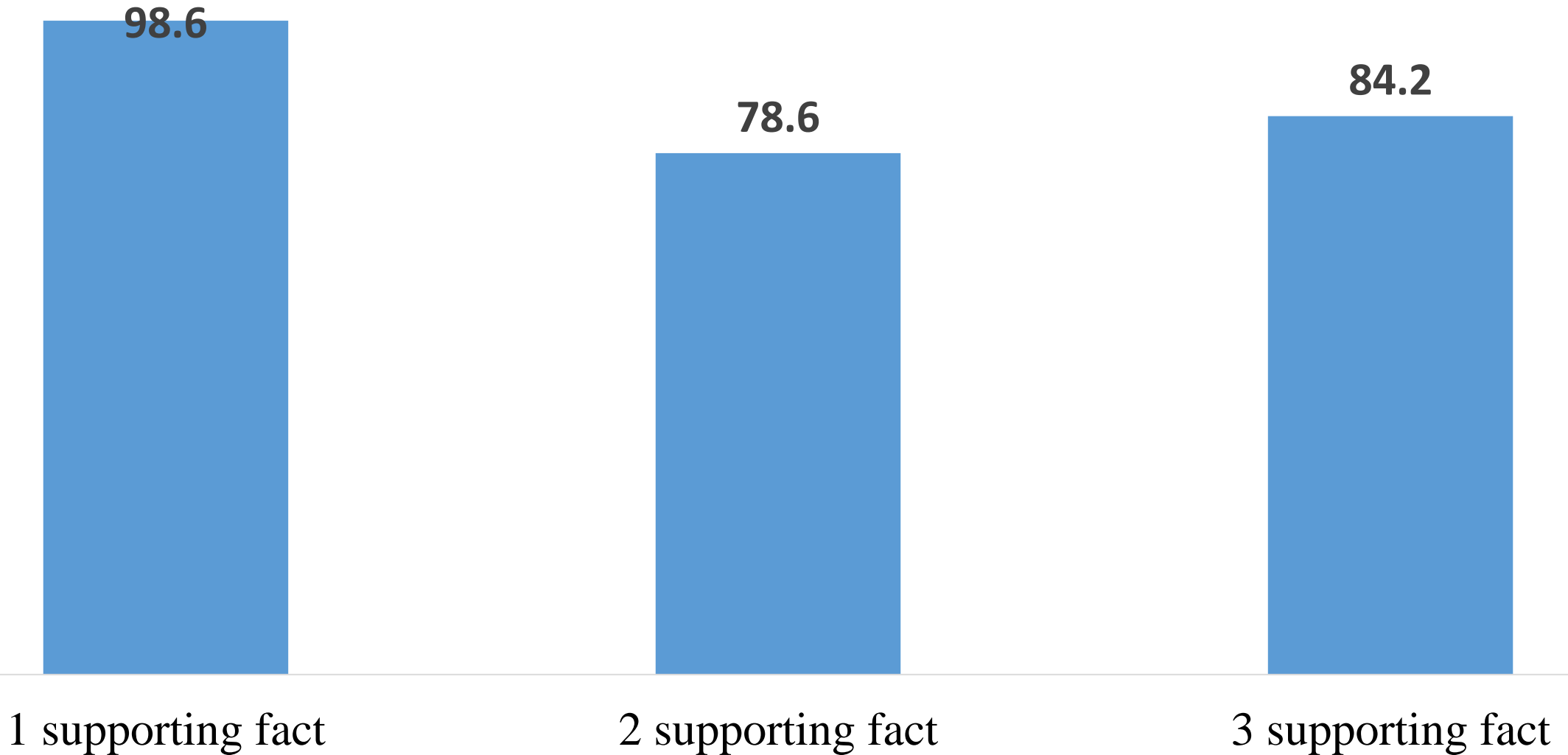
GloVe



DMN



일화 기억 모듈의 주의 기제 분석 (*episodic_pass* = 3)



일화 기억 모듈의 주의 기제 분석

- 모델이 일화적 통과와 회수에 적합한 문제를 더 어려운 문제임에도 불구하고 더 잘 해결하고 있음을 확인할 수 있다.
- 이는 사용되는 주의 기제와 소프트맥스 함수가 단계마다 하나의 추가 근거 사실을 찾는 작업에 적합하기 때문으로 보인다.
- 실험 결과는 DMN이 난이도에 관계없이 특정 과제 유형에 맞는 최적의 구조가 존재한다는 것을 함의하며, 적절한 통과 회수의 조절을 통해 유연성을 갖출 필요성을 나타낸다.

결론 및 향후 연구

- 질의응답 문제를 위한 동적 메모리 네트워크를 다시 소개하였다.
- 최적의 정답률을 얻기 위해 적용할 수 있는 개선들에 대해 고찰하였다.
- 학습된 메모리 구성요소에 대한 분석을 시도하였다.
- 향후 대용량 멀티모달 질의 응답을 위한 외현 메모리 모델 구조 및 관련 알고리즘을 연구할 계획이다.

Q & A

감사합니다.