

# 인공지능을 위한 머신러닝 알고리즘

## 11. 메모리 네트워크

# CONTENTS

1

메모리 네트워크

2

종단 메모리 네트워크

3

나와 질의응답을 하는 메모리 네트워크

# 학습 목표

- 메모리 네트워크의 특징과 작동 원리를 이해할 수 있다.
- 메모리 네트워크와 종단 메모리 네트워크의 차이점을 이해할 수 있다.
- bAbI task를 사용한 메모리 네트워크의 언어학습 능력을 이해할 수 있다.



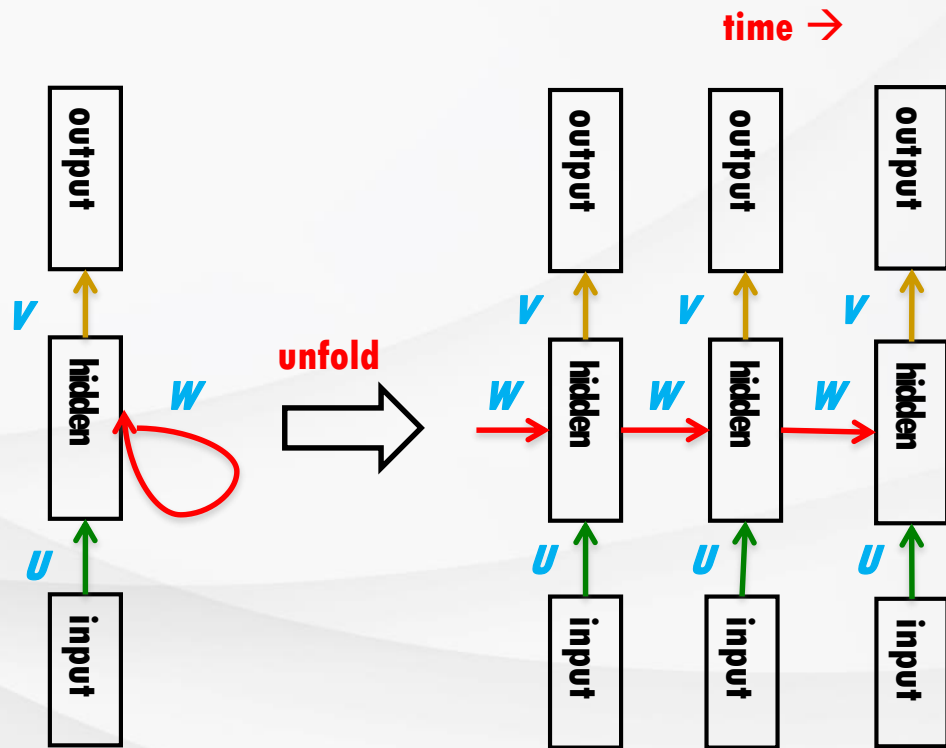
## 1. 메모리 네트워크

## ■ 메모리가 필요한 재현 신경망

$U, W, V$

$$s_t = \tanh(Ux_t + Ws_{t-1})$$
$$y_t' = \text{softmax}(Vs_t)$$

- 은닉 유닛들은 연속된 벡터공간에서 오래전 데이터 정보를 저장
- 길이가 매우 긴 입력 데이터가 주어진다면 은닉 유닛 벡터는 제한된 저장 공간에 많은 입력 정보를 저장해야 함

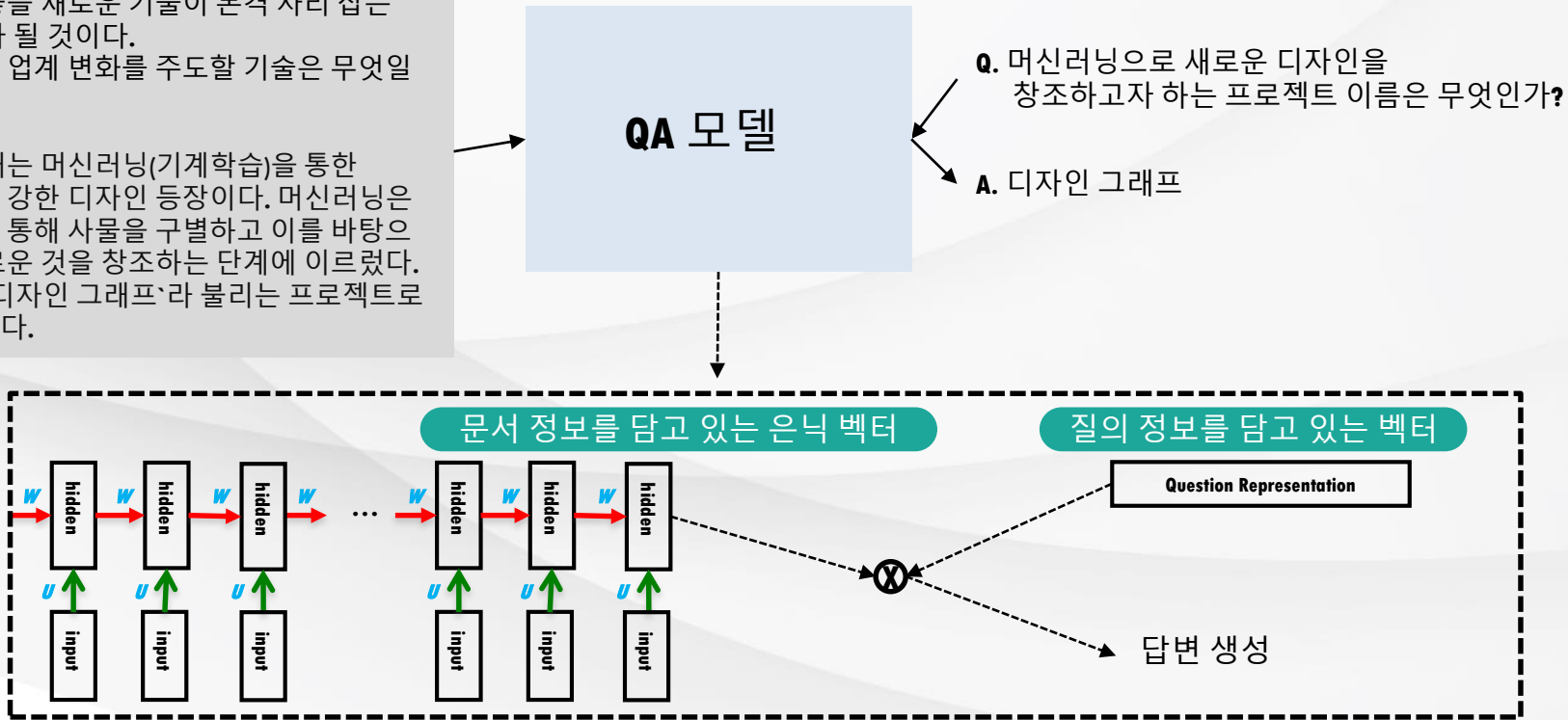


## ■ 예시 : 재현 신경망으로 QA 모델 만들기

**2017**년은 앞으로의 산업계를 완전히 바꿔 놓을 새로운 기술이 본격 자리 잡는 한 해가 될 것이다. 앞으로 업계 변화를 주도할 기술은 무엇일까.

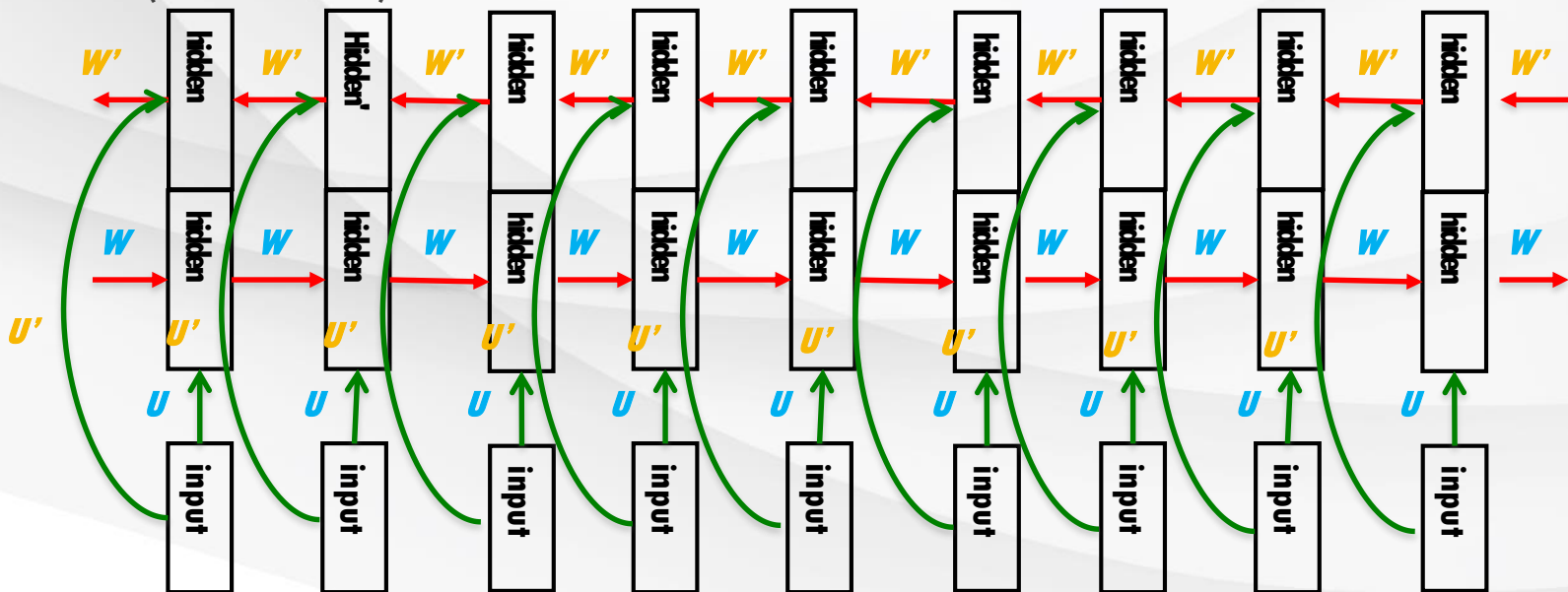
두 번째는 머신러닝(기계학습)을 통한 창조성 강한 디자인 등장이다. 머신러닝은 학습을 통해 사물을 구별하고 이를 바탕으로 새로운 것을 창조하는 단계에 이르렀다. 최근 '디자인 그래프'라 불리는 프로젝트로 구현됐다.

...



## ■ 해결책1: 양방향 재현 신경망 (bi-directional RNN)

- 앞 방향 재현 신경망: 원본 문장의 가장 앞의 단어부터 인코더 재현 신경망의 입력으로 주어짐
- 뒤 방향 재현 신경망: 원본 문장의 가장 뒤의 단어부터 인코더 재현 신경망의 입력으로 주어짐
- 인코더 재현 신경망의 은닉 유닛은 앞 방향 / 뒤 방향 재현 신경망의 은닉 유닛을 단순히 이어붙임 (Concatenation)

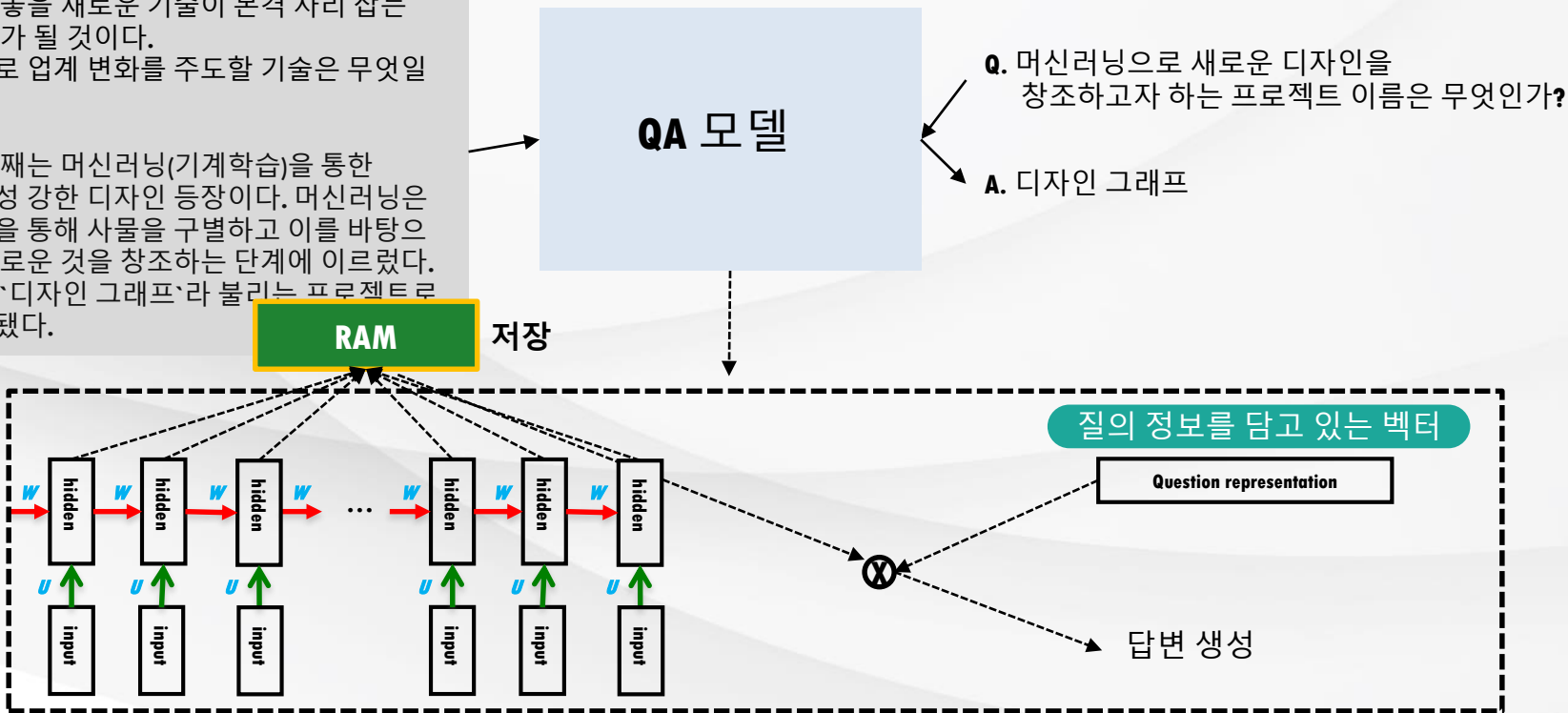


## ■ 해결책2: 히든 유닛 벡터를 저장

2017년은 앞으로의 산업계를 완전히 바꿔 놓을 새로운 기술이 본격 자리 잡는 한 해가 될 것이다.  
앞으로 업계 변화를 주도할 기술은 무엇일까.

...  
두 번째는 머신러닝(기계학습)을 통한 창조성 강한 디자인 등장이다. 머신러닝은 학습을 통해 사물을 구별하고 이를 바탕으로 새로운 것을 창조하는 단계에 이르렀다. 최근 '디자인 그래프'라 불리는 프로젝트로 구현됐다.

...





## ■ 메모리 네트워크

### ❖ 질의 응답 모듈

#### ◎ 스토리 (예> 문장) 선택 모듈

- 질의  $q$ 에 대해 가장 적절한 답변을 포함하고 있는 스토리 (문장)  $s^*$ 를 탐색  
>> 적절성 평가를 통해 가장 높은 점수를 받은 스토리가 선택됨

$$s^* = \arg \max_{s_i} G(q, s_i)$$

- 스토리 선택 모듈의 출력은 질의  $q$ 와 선택된 스토리  $s^*$ 의 결합  
**Ex>** 머신러닝으로 새로운 디자인을 창조하고자 하는 프로젝트 이름은 무엇인가?  
최근 '디자인 그래프'라 불리는 프로젝트로 구현됐다.

#### ◎ 답변 선택 모듈

- 가능한 답변의 집합  $A = \{a_r\}_{1 \dots k}$  중 가장 적합한 답변  $a^*$ 를 선택

$$a^* = \arg \max_{a_r} H(s_a, a_r)$$

## ■ 메모리 네트워크

### ❖ 질의 응답 모듈

#### ◎ 적절성 평가 함수 (Scoring Function)

- $G(x,y) \in [0,1]$  (H 함수도 동일)
- $G(x,y) = \phi_x \mathbf{U}_G^T \mathbf{U}_G \phi_y(y)$
- $H(x,y) = \phi_x \mathbf{U}_H^T \mathbf{U}_H \phi_y(y)$
- $\mathbf{U}$ 는 임베딩 행렬

## ■ 메모리 네트워크의 학습: Triplet Loss

- 학습되어야 하는 함수 :  $G, H$

$$s^* = \arg \max_{s_i} G(q, s_i) \quad a^* = \arg \max_{a_r} H(s_a, a_r)$$

$$G(x, y) = \phi_x U_G^T U_G \phi_y(y)$$
$$H(x, y) = \phi_x U_H^T U_H \phi_y(y)$$

- $G, H$  함수: 긍정 예( $s^*$  또는  $a^*$ )와 부정 예( $s_i$  또는  $a_i$ )의 점수 차이를  $\gamma$ 이상 차이가 나게 만드는 것이 목표
- 손실값 (에러값):  $\max(0, \gamma - G(q, s^*) + G(q, s_i))$ 
  - 긍정 예와 질의의 매치 점수가 부정 예와 질의의 매치 점수보다  $\gamma$ 이상 높을 경우 손실값은 0이 됨

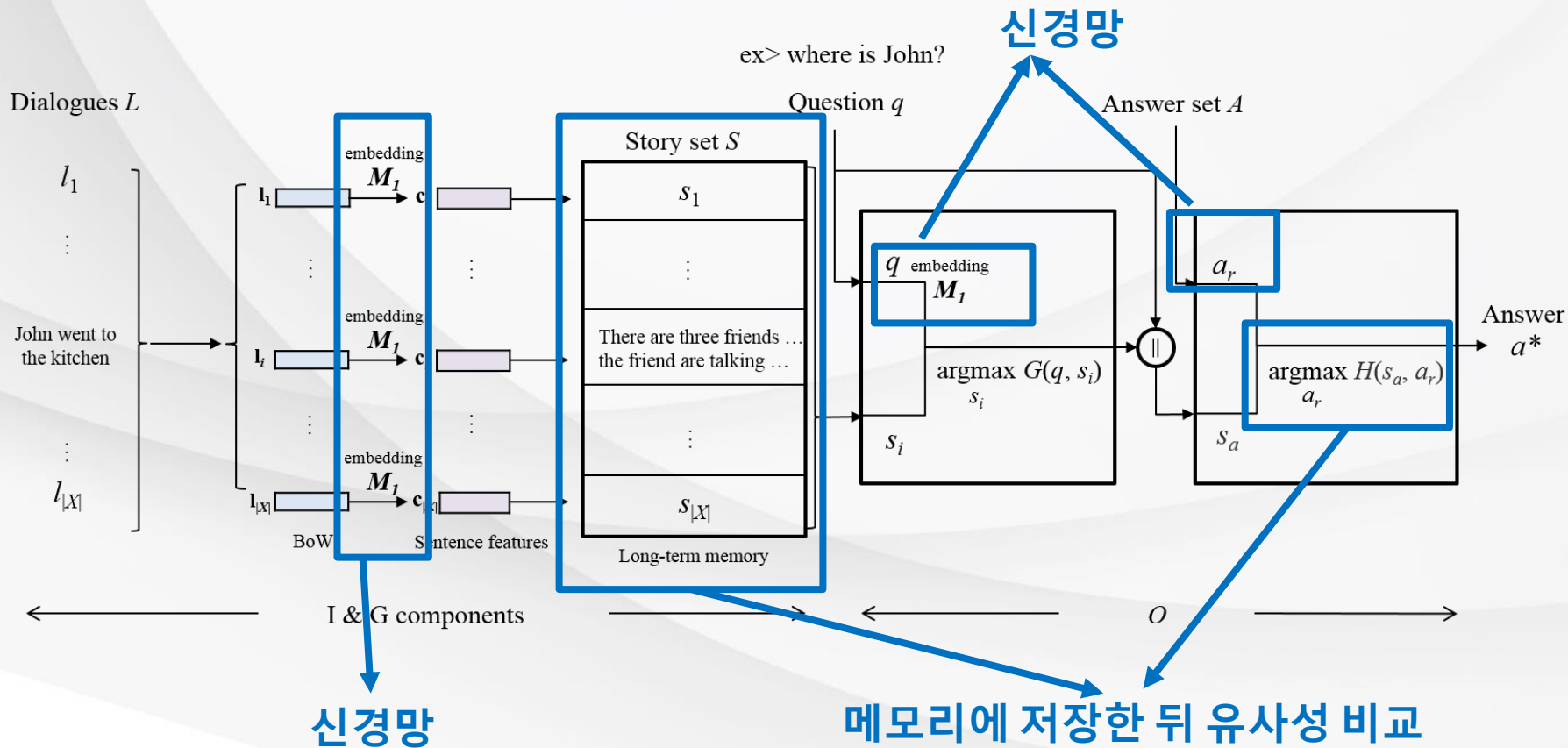
## ■ 메모리 네트워크의 학습: **Triplet Loss**

- 훈련 데이터 전체에 대한 손실 값 :

$$\sum_{s_i} \max(0, \gamma - G(q, s^*) + G(q, si)) + \sum_{a_i} \max(0, \gamma - H([q; s^*], a^*) + H([q; s^*], ai))$$

- 경사 하강법을 사용하여 파라미터  $U_G$ 와  $U_H$ 를 학습

## ■ 메모리 네트워크의 의미





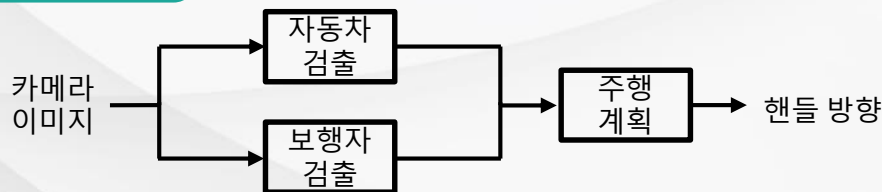
## 2. 종단 메모리 네트워크

### ■ 종단 학습 모델 (End-to-End Learning Model)

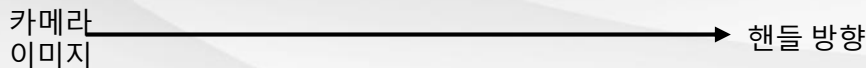
#### ❖ 종단 학습 모델이란

주어진 문제를 풀기 위해서 유용한 내부적 특징 표현들을 스스로 학습하고 다양한 중간 단계의 과정들을 학습 알고리즘 하나로 통합하여 전체적인 작업을 종단으로 수행하는 모델

##### 전통적인 방법



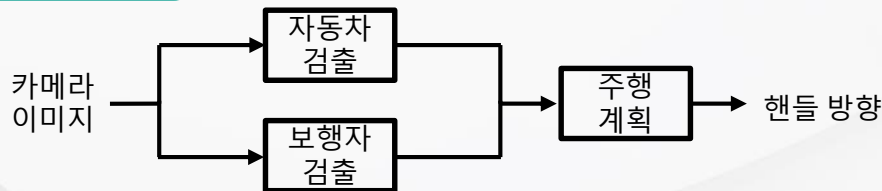
##### 종단 학습



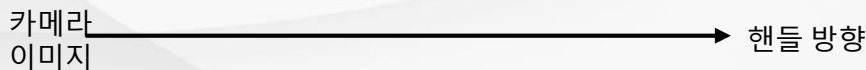
### ■ 종단 학습 모델 (End-to-End Learning Model)

#### ❖ 종단 학습 모델이란

##### 전통적인 방법



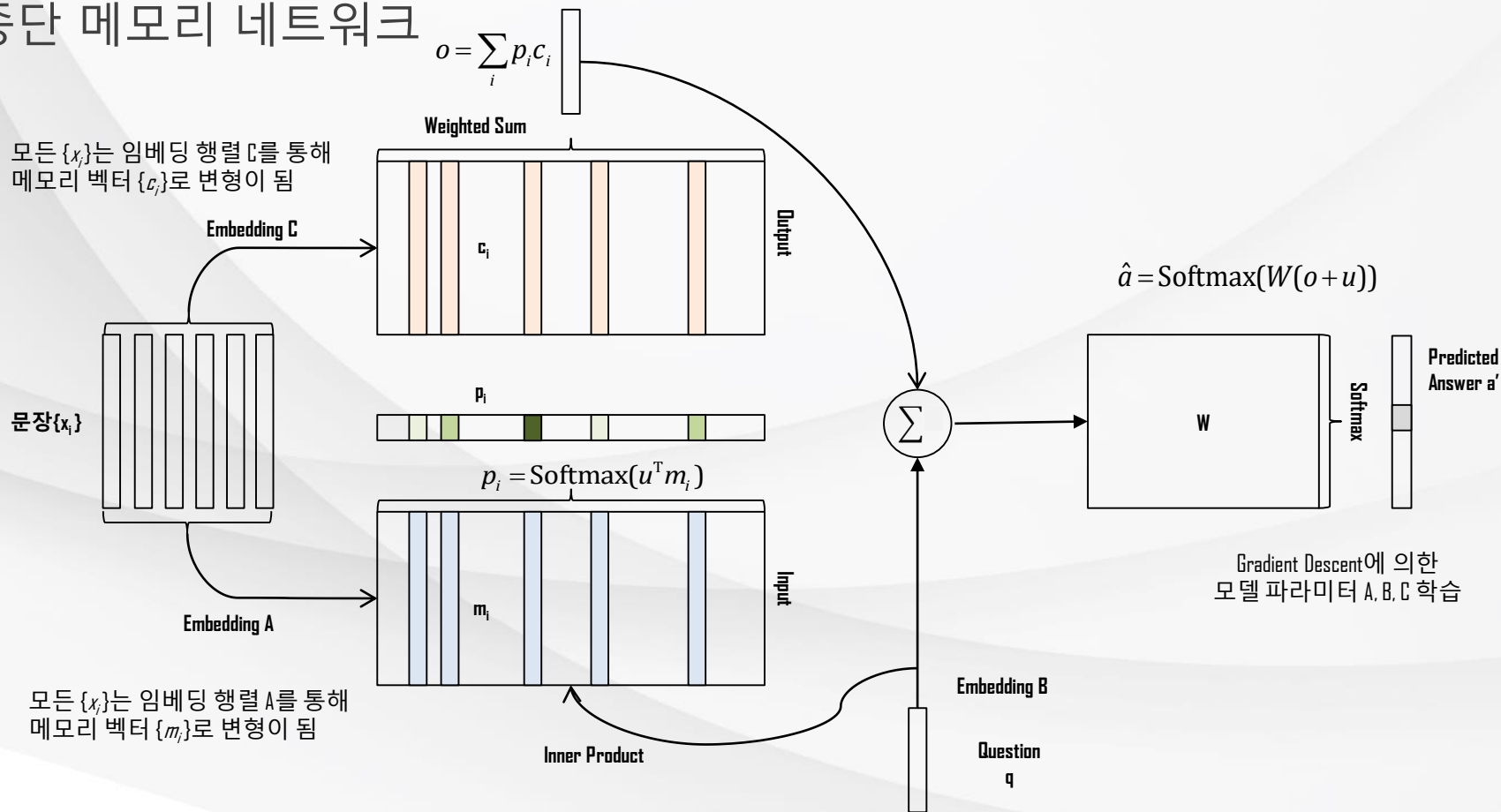
##### 종단 학습



- 기존 메모리 네트워크의 스토리 선택 모듈 (함수 **G**)를 학습시키기 위해서 질의에 대한 적절한 스토리 정보를 사람이 추가로 작성해야 함 (**too expensive!**)
- 종단 학습 메모리 네트워크는(**End-to-End Memory Networks**) 종단 학습 방식을 사용하여 훈련 과정에서 추가적인 감독(**Supervision**)이 필요하지 않음

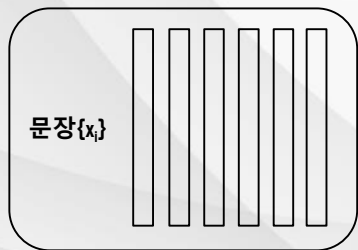


### ■ 종단 메모리 네트워크

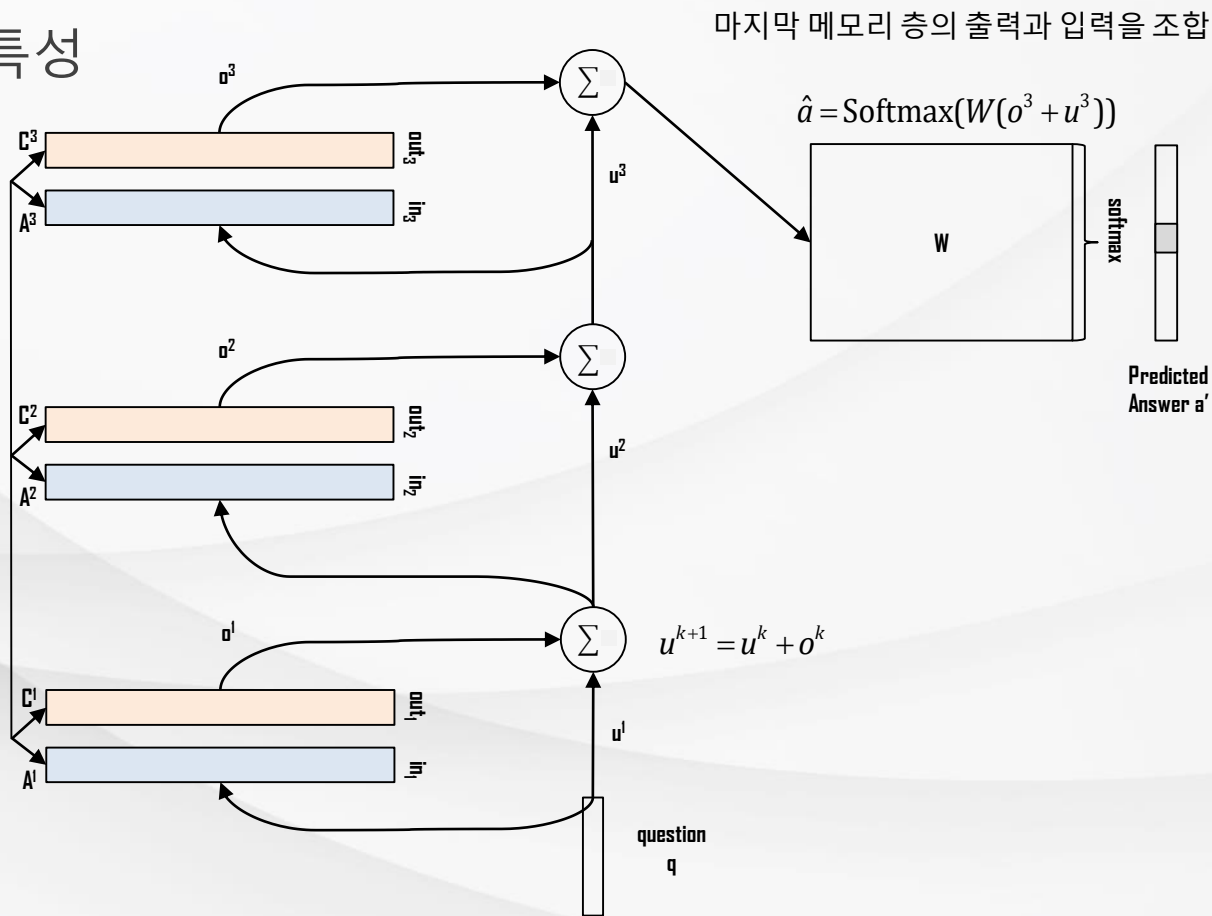



### 재현 네트워크 같은 특성

질의와 관련된 스토리가  
여러 개일 경우,  
 $K$ 번 메모리 접근 연산



임베딩 행렬  $A$ 와  $C$ 를 모든 층에 걸쳐 파라미터를 공유함  
( $A_1=A_2=\dots=A_K$ ,  $C_1=C_2=\dots=C_K$ )



A person is holding a smartphone with both hands, looking at the screen. The background is dark with many out-of-focus, glowing circular lights in shades of yellow, orange, and blue, creating a bokeh effect. The person is wearing a dark jacket.

### 3. 나와 질의응답을 하는 메모리 네트워크

#### ■ bAbI Task

- ◉ 2015년 Facebook AI 연구소에서 제작
- ◉ 모델의 언어 이해 및 추론 능력을 테스트
  - 예전에는 언어 이해/추론 문제를 머신러닝으로 접근하기보다 지식 공학의 형태로 접근하였으나, 신경망을 활용한 해결 방법들이 제시되고 있음

1 John went to the kitchen.  
2 John picked up the milk.  
3 John travelled to the office.  
4 John left the milk there.  
5 John went to the bathroom.  
Where is the milk now? A: office 2 3 4  
Where is John? A: bathroom 5

- ◉ 20개의 테스트로 구성되며, 각 테스트는 별도의 고유한 추론 능력을 요구함
- ◉ 1,000개의 훈련 데이터 쌍 / 1,000개의 테스트 데이터 쌍으로 구성
- ◉ **Supporting fact**라고 불리는 답에 대한 근거 문장 정보가 표시되어 있지만, 이러한 감독 정보를 최소한으로 사용하는 모델이 선호됨

## bAbI Task 종류

- **Supporting fact** 한 개
- **Supporting fact** 두 개
- **Supporting fact** 세 개
- 두 개 객체의 관계
- 세 개 객체의 관계
- 예/아니오
- 개수 세기
- 객체 집합
- 부정 표현
- 불확실한 지식
- 공동 참조 (**coreference**)
- 시간 추론
- 연역 추론
- 귀납 추론
- 장소 추론
- 크기 추론
- 길 찾기
- 의도 추론

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

개수 세기 문제의 예

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

연역 추론 문제의 예

Lily is a swan.  
Lily is white.  
Bernhard is green.  
Greg is a swan.  
What color is Greg? A:white

귀납 추론 문제의 예

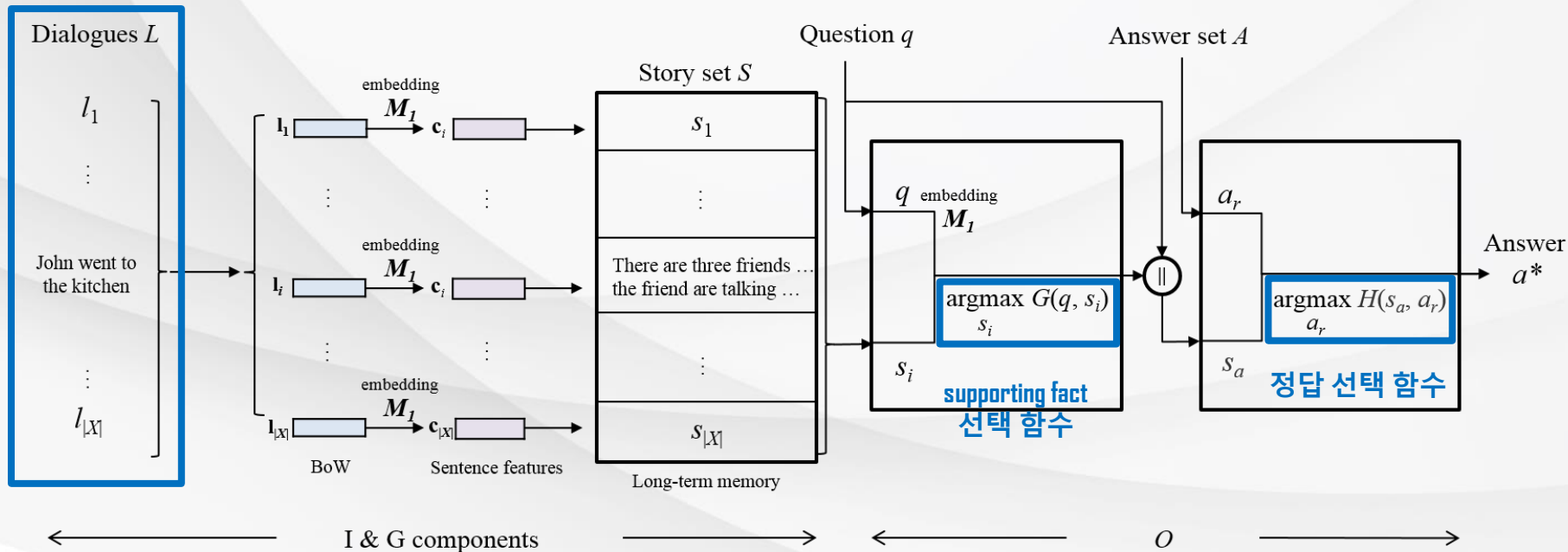
The kitchen is north of the hallway.  
The bathroom is west of the bedroom.  
The den is east of the hallway.  
The office is south of the bedroom.  
How do you go from den to kitchen? A: west, north  
How do you go from office to bathroom? A: north, west

길 찾기 문제의 예

### 3. 나와 질의응답을 하는 메모리 네트워크

#### 메모리 네트워크를 통한 질의 응답

##### 입력: 문장



## ■ bAbI Task에 대한 메모리 네트워크의 성능

|                         | ngram | LSTM | SVM | MemNN | 성공여부 (p>95) | 멀티태스크 학습 |
|-------------------------|-------|------|-----|-------|-------------|----------|
| 1. Supporting fact 한 개  | 36    | 50   | 99  | 100   | 성공          | 100      |
| 2. Supporting fact 두 개  | 2     | 20   | 74  | 100   | 성공          | 100      |
| 3. Supporting fact 세 개  | 7     | 20   | 17  | 100   | 성공          | 98       |
| 4. 두 개 객체의 관계           | 50    | 61   | 98  | 100   | 성공          | 80       |
| 5. 세 개 객체의 관계           | 20    | 70   | 83  | 98    | 성공          | 99       |
| 6. 예/아니오                | 49    | 48   | 99  | 100   | 성공          | 100      |
| 7. 개수 세기                | 52    | 49   | 69  | 85    | 실패          | 86       |
| 8. 객체 집합                | 40    | 45   | 70  | 91    | 실패          | 93       |
| 9. 부정 표현                | 62    | 64   | 100 | 100   | 성공          | 100      |
| 10. 불확실한 지식             | 45    | 44   | 99  | 98    | 성공          | 98       |
| 11. 공동 참조 (Coreference) | 29    | 72   | 100 | 100   | 성공          | 100      |
| 12. Conjunction         | 9     | 74   | 96  | 100   | 성공          | 100      |
| 13. 복합 참조               | 26    | 94   | 99  | 100   | 성공          | 100      |
| 14. 시간 추론               | 19    | 27   | 99  | 99    | 성공          | 99       |
| 15. 연역 추론               | 20    | 21   | 96  | 100   | 성공          | 100      |
| 16. 귀납 추론               | 43    | 23   | 24  | 100   | 성공          | 94       |
| 17. 장소 추론               | 46    | 51   | 61  | 65    | 실패          | 72       |
| 18. 크기 추론               | 52    | 52   | 62  | 95    | 성공          | 93       |
| 19. 길 찾기                | 0     | 8    | 49  | 36    | 실패          | 19       |
| 20. 의도 추론               | 76    | 91   | 95  | 100   | 성공          | 100      |
| 평균                      | 34    | 49   | 79  | 93    |             | 92       |

# 3. 어떻게 기계가 글을 생성 할 수 있을까

## ■ 질의 응답 + 대화 문제

### 1. 영화에 대한 단순 질의 응답

What movies are about open source?

**Revolution OS**

Ruggero Raimondi appears in which movies?

**Carmen**

Can you name a film directed by Stuart Ortiz? **Grave Encounters**

### 2. 영화 추천

Some movies I like are Heat, Kids, Fight Club, Shaun of the Dead, The Avengers, Skyfall, and Jurassic Park. Can you suggest something else I might like? **Ocean's Eleven**

### 3. 영화에 대한 질의 응답 + 추천

I loved Billy Madison, Blades of Glory,

Bio-Dome, Clue, and Happy Gilmore.

I'm looking for a Music movie.

**School of Rock What else is that about?**

**Music, Musical, Jack Black, school, teacher,**

**Richard Linklater, rock, guitar**

### 4. Reddit 대화

I think the Terminator movies really suck, I mean the first one was kinda ok, but after that they got really cheesy.

Even the second one which people somehow think is great.

And after that... forgeddabotit.

**C'mon the second one was still pretty cool.. Arny was still so badass, as was Sararah Connor's character.. and the way they blended real action and effects was perhaps the last of its kind...**



### 3. 어떻게 기계가 글을 생성 할 수 있을까

#### ■ 질의 응답 + 대화 데이터에 대한 메모리 네트워크의 성능

메모리를 활용한 모델이 우수한 성능을 보임

|                       | 단순<br>질의 응답 | 영화추천 | 질의 응답<br>+ 추천 | Reddit 대화 |
|-----------------------|-------------|------|---------------|-----------|
| SVD                   | N/A         | 19.2 | N/A           | N/A       |
| IR                    | N/A         | N/A  | N/A           | 23.7      |
| LSTM                  | 6.5         | 27.1 | 19.9          | 11.8      |
| Supervised embeddings | 50.9        | 29.2 | 65.9          | 27.6      |
| MemN2N                | 79.3        | 28.6 | 81.7          | 29.2      |

#### ■ 기계학습 알고리즘의 메모리 사용에 관한 이슈들

- ◉ 입력으로 들어오는 정보 중 무엇을 저장할지 말지 어떻게 결정할 수 있을까?
- ◉ 메모리에 저장되는 지식을 어떻게 표현해야 할까?
- ◉ 메모리에는 여러 종류 (배열, 스택 구조 등)가 있는데 이들을 어떻게 활용할 수 있을까?
- ◉ 메모리에 저장되는 정보의 크기가 매우 클 때, 어떻게 효율적으로 탐색할 수 있을까?
- ◉ 계층적 메모리를 어떻게 설계할 수 있을까?
- ◉ 합성 함수의 사용과 같은 계층적 추론을 어떻게 할 수 있을까?
- ◉ 정보의 손실 (망각, **unlearning**)을 어떻게 구현할 수 있을까?
- ◉ 추론 모델을 평가하는 가장 좋은 방법은 무엇일까?  
어떤 문제가 가장 좋은 평가가 될 수 있을까?
- ◉ 사람, 동물의 기억 행위에서 얻을 수 있는 영감에는 무엇이 있을까?



학습정리

지금까지 [메모리 네트워크]에 대해서 살펴보았습니다.

## 메모리 네트워크

기존 재현 신경망의 단점: 은닉 유닛 벡터가 제한된 공간에 많은 입력 정보를 저장해야 함

메모리 네트워크: 재현 신경망의 저장 공간 한계 문제를 극복하기 위해 제안되었으며

스토리 선택 모듈 G와 답 선택 모듈 H로 구성됨

$$s_i^* = \arg \max_{s_i} G(q, s_i) \quad a^* = \arg \max_{a_r} H(s_a, a_r)$$

## 종단 메모리 네트워크

종단 학습 모델의 필요성: 기존 메모리 네트워크는 스토리 선택 모듈 G를 학습시키기 위해서 질의에 대한 적절한 스토리 정보를 사람이 추가로 작성해야 함

여러 번의 메모리 접근을 위해 '파라미터 공유'라는 재현 신경망의 특징을 가짐

## 나와 질의응답을 하는 메모리 네트워크

bAbI Task: 모델의 언어 이해 및 추론 능력을 20가지로 나누어 테스트하려는 문제  
기존 룰 기반 방식에서 메모리 네트워크를 활용하는 머신러닝 방식의 해결책이 제시됨