

기억 회상 현상에 대한 순차적 멀티 모달리티 데이터의 영향 비교

Comparison of Sequential Multi-Modality Data Effects on Memory Recall

윤웅창¹, Umer Fareed¹, 장병탁^{1,2}

¹서울대학교 컴퓨터공학부, ²서울대학교 인지과학협동과정
{wcyoon, ufareed, btzhang}@bi.snu.ac.kr

초 록

본 논문에서는 멀티 모달리티가 인간의 기억 현상 중에서 특히 회상 현상에 도움이 될 것이라는 아이디어를 인간 실험과 컴퓨터 시뮬레이션을 통해서 비교·확인한다. 우리는 순차적 멀티 모달리티 정보가 기억회상 현상에 미치는 영향을 알아보기 위해서 확률 하이퍼그래프 메모리 모델을 사용하여 컴퓨터 시뮬레이션을 수행하였으며, 시뮬레이션 결과를 동일한 데이터를 사용한 인간 실험 결과와 비교하였다. 본 실험에서는 TV 드라마의 비디오 데이터를 이용하여 이전 시점의 사진과 문장 정보가 다음 시점의 문장 생성에 미치는 영향을 확인하였다. 본 실험의 목적은 이전 시점의 문장 정보와 문장/사진 정보를 활용하여 멀티 모달리티의 영향을 확인하는 것이며 다양한 시점의 정보를 활용하여 순차성이 회상에 갖는 영향을 확인하는 것이다. 이를 통해서 기억 회상 현상에 있어서 멀티 모달리티가 미치는 영향과 순차적 데이터가 미치는 영향을 보일 수 있었으며, 기계를 통해서 인간의 기억 회상 현상을 재현할 수 있는 시뮬레이션 모델을 구현 하는데 실마리를 제공하였다.

서 론

인간의 학습에 있어서 단순 모달리티 보다는 멀티 모달리티가 더 나은 결과를 가져다 준다는 사실은 잘 알려져 있다[1]. 일명 ‘멀티미디어 원리(Multimedia principle)’로 알려진 이 현상은 문자로만 학습하는 것 보다는 문자와 이미지를 같이 사용할 때 학습 효과가 더 올라간다는 것이다[2]. 기억과 학습의 관계를 볼 때 기억 현상에서 멀티모달 데이터를 사용한 기억이 단순 모달리티 데이터를 사용한 기억보다 회상에 있어서 더 나은 결과를 보여줄 것이라고 예측할 수 있다[3]. 몇몇의 실험결과는 이러한 가정을 지지하는 결과를 보여주고 있다[4] [5] [6].

우리는 인간의 기억 현상 중에서도 회상 현상에 순차적 멀티 모달리티가 주는 효과를 확인해 보려고 하였다. 이 실험에서는 TV 드라마의 비디오 데이터를 이용하여 이전 시점의

사진과 문장 정보를 이용하여 다음 시점의 문장의 회상 정도를 측정하여 보았다. 먼저 멀티 모달리티가 회상 현상에 미치는 영향을 알아보기 위해서 이전 시점의 문장만 주어졌을 때의 문장 회상 정도와 문장과 이미지 정보가 같이 주어졌을 때의 문장 회상 정도를 알아 보았으며, 순차성에 따른 영향을 알아보기 위해서 이전 시점($t-1$)의 정보만 주어졌을 때의 문장 회상 정도와 그 이전 시점($t-2$)도 이전 시점($t-1$)과 같이 주어졌을 때의 차이를 비교해 보았다. 컴퓨터 시뮬레이션을 통해서 나온 결과는 멀티 모달리티 정보가 단일 모달리티 보다 기억 회상 현상에 도움을 준다는 결과를 얻어 내었으며 인간 실험 결과에서도 동일한 결과를 다시 확인하였다. 순차성에 대해서는 컴퓨터 시뮬레이션과 인간 실험에 있어서 차이가 있었다.

이 논문은 컴퓨터 시뮬레이션과 인간 실험의

결과를 통해서 인간의 기억 회상 현상을 재현할 수 있는 모델을 위한 기초실험이라는 의의를 가진다.

본 론 1

랜덤 하이퍼그래프 메모리 모델(하이퍼네트워크)

이 실험에서 컴퓨터 시뮬레이션은 랜덤 하이퍼그래프 메모리 모델(하이퍼네트워크)을 바탕으로 하고 있다. 랜덤 하이퍼그래프 모델[7]은 여러 개의 정점이 가중치를 가진 하나의 하이퍼에지로 연결될 수 있으며, 이러한 하이퍼에지 들의 집합이라고 볼 수 있다. 랜덤 하이퍼그래프 메모리 모델을 H , 정점을 V , 연결선을 E , 가중치를 W 라고 할 때, $H = (V, E, W)$ 로 나타낼 수 있으며 데이터 $D = \{a^{(i)}\}_i$ 를 저장할 수 있는 확률적 연산 메모리라고 볼 수 있다.

$G = (V, E, W)$
 $V = \{v1, v2, v3, \dots, v7\}$
 $E = \{E1, E2, E3, E4, E5\}$

$E1 = \{v1, v3, v4\}$
 $E2 = \{v1, v4\}$
 $E3 = \{v2, v3, v6\}$
 $E4 = \{v3, v4, v6, v7\}$
 $E5 = \{v4, v5, v7\}$

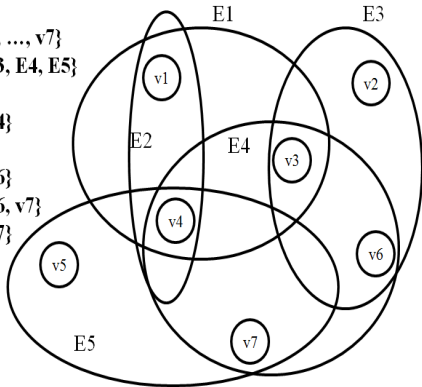


그림 1. 랜덤 하이퍼그래프 메모리 모델(하이퍼네트워크)의 예. 이 모델은 정점들과 하이퍼에지들로 구성되며, 하나의 하이퍼에지는 가중치를 가지고 임의의 정점들로 연결된다.

시뮬레이션을 위한 언어회상 방식

랜덤 하이퍼그래프 메모리 모델을 이용한 전체적인 언어회상 방식은 표1과 같다. 하이퍼네트워크(랜덤 하이퍼그래프 메모리

모델)는 모든 이미지와 텍스트 데이터를 랜덤하게 샘플링 하되 n-gram방식을 사용하여 원래 데이터의 1/3크기를 가지는 일정한 수의 하이퍼에지를 생성하고 원래 데이터의 조합과의 일치여부를 비교하여 가중치를 조정하게 된다. 이러한 과정을 일정횟수 또는 정해진 시간 동안 반복하여 역치 값 이하의 하이퍼에지를 제거하거나 새로운 하이퍼에지를 생성하는 방법으로 학습을 진행한다.

표1. 랜덤 하이퍼그래프 메모리 모델을 통한 학습과 문장 회상 알고리즘

```

H : 하이퍼에지들의 집합,  $h \in H$ 
T : 텍스트 데이터 집합,  $t \in T$ 
I : 이미지 데이터 집합,  $i \in I$ 
C : 후보군들의 집합,  $c \in C$ 
q : 특정 i에 대한 이전 시간의 단서
r : 특정 t에 대한 이전 시간의 단서
S : 생성된 문장,  $s \in S$ 
For 1 to EpochTime do
  -  $H \leftarrow T, I$ 에 대해서 랜덤 샘플링
  - 모든 i, t에 대해서 h가 일치/ 불일치에 따라 하이퍼에지의 가중치를 증가/ 감소
  - 특정 하이퍼에지의 가중치가 역치 이하이면 하이퍼에지를 제거하고 새로 랜덤 샘플링
End For
For 1 to size(H) do
  If q와 r에 대해서 특정 h가 일치하면 then
     $C \leftarrow h$ 
  End If
End For
For 1 to size(q) do
  If q, r에 대한 위치 데이터가 c와 일치 then
    If 복수개의 c가 존재하면 then
       $S \leftarrow$  가장 큰 가중치를 가지는 c
    End If
  End If
End For

```

이렇게 생성된 하이퍼네트워크에서 단서로 주어지는 이미지 또는 텍스트의 일부분과 일치하는 하이퍼에지들을 모두 후보군에 저장한 뒤, 위치정보와 가중치를 사용하여 문장 회상을 수행하게 된다.

Multi-Modal Memory Game(MMG) 인터페이스

컴퓨터 시뮬레이션과 비교하기 위해서 유저친화적이고 쉽게 조작이 가능한 Multi-Modal Memory Game(MMG)라는 인간 실험을 위한 플랫폼을 개발 하였다. 전체적인 구성은 그림 2와 같다.

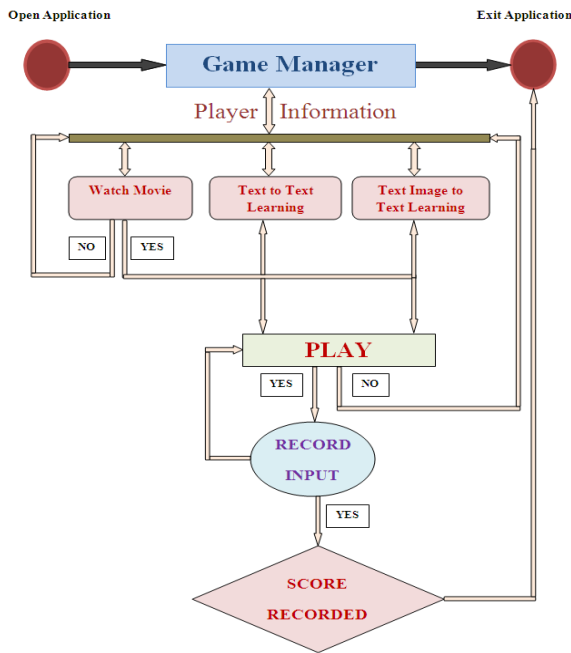


그림2. Multi-modal Memory Game의 구성도

인간 실험자가 Play 버튼을 클릭하여 비디오를 시청한 뒤 실험 방법을 선택이 가능 하도록 하였다. 실험 방법을 선택한 뒤에는 풀이문제 수를 설정하게 된다. 문제풀이를 시작하면 위의 화면에는 이전 시점의 문장 또는 이미지가 주어지게 되며 이를 토대로 기억을 되살려서

다음 문장을 작성하게 된다. 작성된 문장이 원래의 문장과 맞는지 틀리는지를 알려주고 해답에 해당되는 원래의 문장을 알려주어서 유저가 스스로 학습이 가능하도록 하였다. MMG의 유저인터페이스는 그림3과 같다.

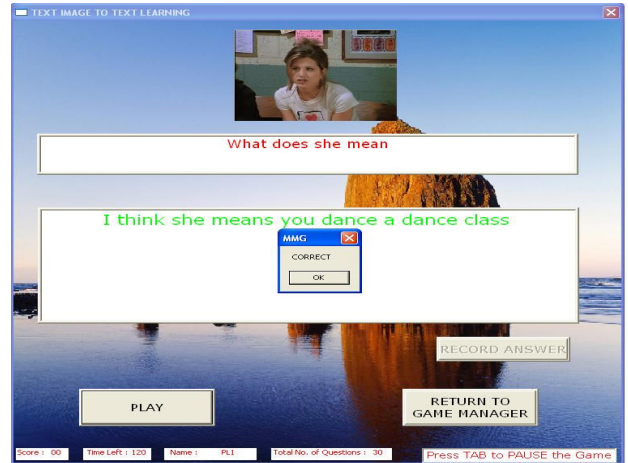


그림3. Multi-modal Memory Game의 UI 화면

본 론 2

실험의 설정

순차적 멀티모달리티 데이터가 기억회상 현상에 미치는 영향을 알아보기 위해서 총 4가지의 실험을 동일한 데이터를 사용하여 각각 컴퓨터 시뮬레이션과 인간을 통해서 실험 하였다.

먼저 멀티 모달리티와 단일 모달리티와의 차이를 알아보기 위해서 이전 시점의 텍스트 단서만을 주고 다음 시점의 문장의 회상의 결과를 보는 실험(T2T)과 이전 시점의 텍스트와 이미지가 같이 주어졌을 때의 문장 회상 정도를 알아보는 실험(TI2T)를 수행하였다.

순차성에 따른 효과를 알아보기 위해서 이전 시점의 텍스트와 이미지 단서를 주고 문장 회상을 알아보는 실험(TI2T)과 바로 전 단계(이전 시점)의 시점과 두 단계 이전 시점의 단서가 같이 주어졌을 때의 문장 회상 정도를

보는 실험(TITI2T)를 수행하였다.

실험에 사용한 데이터는 잘 알려진 미국의 TV 드라마 시리즈인 ‘Friends’를 사용하였다. 이미지는 특정 자막에 해당되는 이미지를 200x150픽셀 크기로 캡처하여 총 294쌍의 텍스트-이미지 쌍을 실험에 사용하였다. 컴퓨터 시뮬레이션의 경우, 이미지를 92개의 그물망 형태로 자른 뒤 k-군집화 기법을 사용하여 2000개의 군집으로 나누었다. 하나의 군집은 비슷한 이미지들로 구성되게 되므로 특정 군집은 특정한 의미를 부여할 수 있다. 이렇게 만들어진 군집을 이용하여 이미지 사전을 만들게 되며 하나의 이미지에서의 각각의 이미지 조각들은 어떠한 군집에 속하는지를 나타내게 된다. 문장의 경우에도 유사하게 텍스트 사전을 만들어서 하나의 문장에 등장하는 단어가 어떻게 구성되었는지를 암호화하여 나타내게 된다.

인간 실험의 경우, 시간과 공간에 제약이 적은 기계와 다르기 때문에 표2와 같은 설정을 사용하였다. 또한 실험자가 하나의 문장에서의 모든 단어를 순서대로 정확하게 기억하는 것은 매우 어려운 일이므로 60%이상 기억해 내면 맞춘 것으로 간주하였다.

표2. 인간 실험의 설정사항

세션당 문제수	총 수행세션	문제당 제한시간
30문항	10번	120초

실험 결과

멀티 모달리티가 기억 회상 현상에 미치는 영향을 보기 위한 이전 시점의 텍스트만 주었을 때의 문장 회상 정도(T2T)와 텍스트와 이미지를 같이 주었을 때의 문장 회상 정도(TI2T)의 차이에 관련된 기계(컴퓨터 시뮬레이션)와

인간의 결과는 그림 5와 같다. 결과를 보면 기계와 인간의 경우 모두 단일 모달리티(텍스트) 보다는 멀티 모달리티(텍스트-이미지쌍)의 단서를 주었을 때 더 나은 성능을 보여 주는 것을 알 수 있다.

순차성이 기억 회상 현상에 미치는 영향을 알아 보기 위한 한 단계 이전 시점의 단서만 주었을 때의 문장 회상 정도(TI2T)와 한 단계와 두 단계 이전 시점의 단서를 같이 주었을 경우의 문장 회상 정도(TI-TI2T)의 기계와 인간에 관련된 결과는 그림 6과 같다. 이전의 멀티 모달리티에 관련된 실험과 달리 기계의 경우에는 두 단계 이전 시점과 한 단계 이전 시점을 모두 주었을 경우에는 성능이 향상 되었으나 인간의 경우에는 오히려 감소한 것을 알 수 있다.

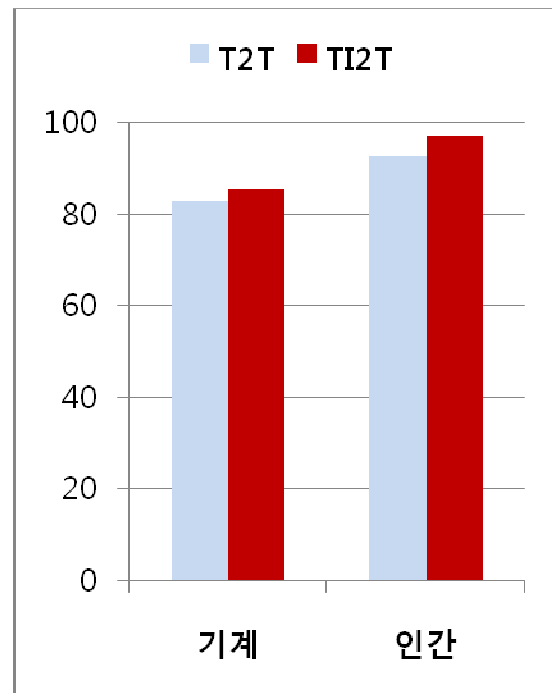


그림4. T2T(text2text)와 TI2T(TextImage2text)의 실험 결과. 이전 시점의 텍스트만 주어졌을 때보다 텍스트와 이미지가 같이 주어졌을 때 더 나은 결과를 보여준다.

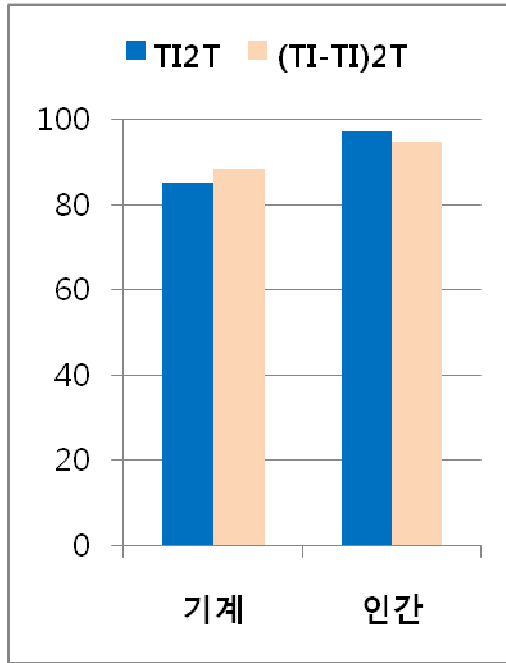


그림5. TI2T(textImage2text)와 TI-TI2T(TextImage-TextImage2text)의 실험 결과. 한 단계 전 시점의 단서만 주어졌을 때와 한 단계와 두 단계 이전 시점의 단서에 따른 기계와 인간의 결과가 다르다는 것을 보여준다.

결론

이 논문은 순차적 멀티모달리티 데이터가 기억 회상 현상에 있어서 어떠한 영향을 미치는지를 문장 회상을 통해서 기계와 인간의 경우를 비교하여 보았다.

모달리티 측면에서는 단일 모달리티 보다는 멀티 모달리티가 기억 회상에 더 좋은 영향을 주지만 시간 측면인 순차성에 있어서는 기계는 더 좋은 영향을 줄지 모르지만 인간에 있어서는 부정적으로 작용할 수도 있다는 것을 알 수 있었다.

추후의 실험에서는 인간의 실험에 있어서 더 많은 참가자를 통해서 통계적으로 신뢰성이 높은 실험이 반드시 필요하며, 컴퓨터 시뮬레이션과 인간 실험의 설정과 여러 속성들을 동일한

환경하에서의 실험을 해야 할 것이다. 또한 기억회상 현상 이외에도 다른 현상들에 대해서도 연구하여 인간의 기억 현상을 포괄적으로 재현하는 모델을 만들어야 할 것이다.

감사의 글

이 논문은 교육과학기술부의 재원으로 국가연구재단의 지원을 받아 수행된 연구(314-2008-1-D00377, Xtran/ No. 2010-0017734)이며, 지식경제부 및 한국산업기술평가관리원의 IT산업원천기술개발사업(K1002138, 차세대 맞춤형 서비스를 위한 기계학습 기반 멀티모달 복합 정보 추출 및 추천 기술 개발, MARS) 및 교육과학기술부의 BK21-IT 사업에 의해 일부 지원되었음.

참고문헌

- [1] Cisco Systems, Inc, Multimodal Learning Through Media: What the Research Says. 2008.
- [2] DA Kobus, Effect of multimodal stimulus presentation on recall, Percept Mot Skills. 1994 Feb;78(1):320-2.
- [3] Serena Mastroberardino, Valerio Santangelo, Fabiano Botta, Francesco S. Marucci and Marta Olivetti Belardinelli, How the bimodal format of presentation affects working memory: an overview, Cognitive Processing, Vol 9, No 1, 69-76, 2008
- [4] Maud Ciekanskial and Thierry Chanier, Developing online multimodal verbal communication to enhance the writing process in an audio-graphic conferencing environment,

ReCALL (2008), 20:162–182

[5] Rigas and Marwan Alseid, Multi-modal aided Presentation of Learning Information: A Usability Comparative Study Dimitris, IADIS International Conference Interfaces and Human Computer Interaction 2008

[6] C. Spence and J. Driver, Editors, Crossmodal Space and Crossmodal Attention, Oxford University Press, Oxford, UK (2004), pp. 179–220.

[7] B.-T. Zhang, Hypernetworks: A molecular evolutionary architecture for cognitive learning and memory, IEEE Computational Intelligence Magazine, 3(3):49–63, 2008