GuessWhat?! 문제에 대한 분석과 파훼

Analyzing and Solving GuessWhat?!

저자 (Authors)

이상우, 한철호, 허유정, 강우영, 전재현, 장병탁

Sang-Woo Lee, Cheolho Han, Yujung Heo, Wooyoung Kang, Jaehyun Jun, Byoung-Tak Zhang

출처 (Source)

정보과학회논문지 45(1), 2018.1, 30-35 (6 pages)

Journal of KIISE 45(1), 2018.1, 30-35 (6 pages)

발행처 (Publisher)

한국정보과학회

KOREA INFORMATION SCIENCE SOCIETY

URL

http://www.dbpia.co.kr/Article/NODE07319396

APA Style

이상우, 한철호, 허유정, 강우영, 전재현, 장병탁 (2018). GuessWhat?! 문제에 대한 분석과 파훼. 정보과학회논문지, 45(1), 30-35.

ISSN 2383-630X(Print) / ISSN 2383-6296(Online)
Journal of KIISE, Vol. 45, No. 1, pp. 30-35, 2018. 1
https://doi.org/10.5626/JOK.2018.45.1.30

GuessWhat?! 문제에 대한 분석과 파훼
(Analyzing and Solving GuessWhat?!)

이 상 우 [†] 한 철 호 [††] 허 유 정 [††]
(Sang-Woo Lee) (Cheolho Han) (Yujung Heo)

강 우 영 [†††] 전 재 현 [††††] 장 병 탁 [†††††]
(Wooyoung Kang) (Jaehyun Jun) (Byoung-Tak Zhang)

요 약 GuessWhat?!은 질문자와 답변자로 구성된 두 플레이어가 이미지를 보고 질문자에게 비밀로 감추어진 정답 물체에 대해 예/아니오/잘 모르겠음 셋 중 하나로 묻고 답하며, 정답 물체를 추려 나가는 문제이다. GuessWhat?!은 최근 컴퓨터 비전과 인공지능 대화 시스템의 테스트베드로서 컴퓨터 비전과 인공지능 학계의 많은 관심을 받았다. 본 논문에서, 우리는 GuessWhat?! 게임 프레임워크가 가지는 특성에 대해 논의한다. 더 나아가, 우리는 제안된 틀을 기반으로 GuessWhat?!의 간단한 solution을 제안한다. 사람이 평균 4~5개 정도의 질문을 통하여 맞추는 이 문제에 대하여, 우리가 제안한 방법은 2개의 질문만으로 기존 딥러닝 기반 기술의 성능을 상회하는 성능을 보이며, 5개의 질문이 허용되면 인간 수준의 성능을 능가한다.

키워드: 대화 시스템, 이미지 질의 응답, GuessWhat?!, 인공지능 게임

Abstract GuessWhat?! is a game in which two machine players, composed of questioner and answerer, ask and answer yes-no-N/A questions about the object hidden for the answerer in the image, and the questioner chooses the correct object. GuessWhat?! has received much attention in the field of deep learning and artificial intelligence as a testbed for cutting-edge research on the interplay of computer vision and dialogue systems. In this study, we discuss the objective function and characteristics of the GuessWhat?! game. In addition, we propose a simple solver for GuessWhat?! using a simple rule-based algorithm. Although a human needs four or five questions on average to solve this problem, the proposed method outperforms state-of-the-art deep learning methods using only two questions, and exceeds human performance using five questions.

Keywords: dialogue system, visual question answering, GuessWhat?!, artificial intelligence game

· 이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터(R0126-16-1072-SW스타랩, 2017-0-01772-VTT), 한국산업기술평가관리원(10044009-HRI.MESSI, 10060086-RISF)의 지원을 받았음
· 이 논문은 2017 한국컴퓨터종합학술대회에서 'GuessWhat?! 문제에 대한 분석과 파훼'의 제목으로 발표된 논문을 확장한 것임

† 학생회원 : 서울대학교 컴퓨터공학부
slee@bi.snu.ac.kr
†† 비 회 원 : 서울대학교 컴퓨터공학부
chhan@bi.snu.ac.kr
yjheo@bi.snu.ac.kr
††† 학생회원 : 서울대학교 컴퓨터공학부
wykang@bi.snu.ac.kr
†††† 학생회원 : 서울대학교 뇌과학협동과정
jhjun@bi.snu.ac.kr
††††† 종신회원 : 서울대학교 컴퓨터공학부 교수(Seoul Nat'l Univ.)
btzhang@bi.snu.ac.kr
(Corresponding author임)

논문접수 : 2017년 7월 11일
(Received 11 July 2017)
논문수정 : 2017년 11월 3일
(Revised 3 November 2017)
심사완료 : 2017년 11월 9일
(Accepted 9 November 2017)

정보과학회논문지 제45권 제1호(2018. 1)

1. 서 론

GuessWhat?!은 질문자와 답변자로 구성된 두 플레이어가 이미지를 보고 질문자에게 비밀로 감추어진 정답 물체에 대해 예/아니오로 묻고 답하며, 정답 물체를 추려 나가는 문제이다[1]. GuessWhat?!은 최근 컴퓨터 비전과 인공지능 학계의 많은 관심을 받고 있다.

최근 많은 연구들이 두 인공지능이 공진화하여, 특정 문제에서 인간과 유사한 성능의 상호 작용 시스템을 만드는 문제를 다루었다. 가장 성공적인 공진화 시스템의 예로 generative adversial network (GAN)가 있다[2]. GAN은 무감독 학습을 위하여 생성기와 분류기를 경쟁시키는 모델이다. 이러한 경쟁 기작을 바탕으로 GAN은 이미지 자동 생성과 관련하여 기존 generative model을 압도하는 성능을 내었다. 서로 대화하는 chatbot 시스템은 또 다른 대표적인 예 중의 하나이다[3]. 두 개의 구글 로봇이 며칠에 걸쳐서 철학적인 담론을 주고받는 모습이 인터넷에 올라와서 사람들의 주목을 받았다. 한 에이전트가 이미지의 물체에 대해 한 문장으로 언급하고, 다른 에이전트가 그 물체가 무엇인지 맞추는 ReferIt!도 이와 관계가 있다[4]. GuessWhat?!도 이러한 연구의 연장선 상에 있다.

하지만, 상호 진화하는 시스템을 계속 학습시키는 것은 어려운 일이다. 지금은 많이 개선되었지만, GAN 모델을 의미 있게 학습하는 것은 매우 어려운 일이었고 아주 최근에 이르러서야 응용에 사용될 수 있을 수준의 의미 있는 성능 개선이 있었다[5]. 또한, 두 인공지능 대화 시스템이 특정 목표를 위하여 묻고 답하는 과정에서 자기네들끼리의 은어로 대답하는 문제가 생긴다[6]. Refer it의 경우, 좌표를 주는 경우, 문제가 바로 풀린다. 이는 기계학습 관점에서 보면 목표 함수에 문제가 있는 것으로, 인공지능 관점에서 보면 풀고자 하는 과제의 정의에 문제가 있는 것이다.

우리는 본 논문에서, GuessWhat?!에서 두 플레이어의 정답 물체에 대한 정확도 자체만으로는 완결성 있게 문제를 정의하는 데 부족함이 있음을 논증한다. 비록 GuessWhat?!이 ReferIt!의 대안으로서 생각될 수 있음에도 불구하고, 우리는 GuessWhat?! 문제가 복잡한 딥러닝이 아닌 아주 쉬운 방법으로 해결 될 수 있음을 보인다. 우리는 위치에 대해 질문하는 아주 간단한 알고리즘으로 기존 state-of-the-art 뿐 만이 아니라 인간 수준을 넘는 성능을 보일 수 있음을 논증한다.

이 논문은 단순히 GuessWhat?! 문제에서 제안하는 방법의 압도적인 성능을 보고하기 위하여 쓰여진 것이 아니다. 이 논문에서는 GuessWhat?!의 문제 특성에 대하여 논의하며, 더 나아가 목표 지향적인 대화 시스템의 개발에 대하여 생각할 지점을 제공하고자 한다.

2. GuessWhat?!

GuessWhat?!에서 질문자는 이미지에서 어떤 물체가 답변 문제인지를 맞추기 위하여, 답변자에게 언어 형태로 질문을 한다. 답변자는 이미지에서 어떤 물체가 답변 물체인지를 알고 있다. 하지만, 답변자는 질문자에게 예, 아니오, 잘 모르겠음 세 가지 중 하나의 답변만을 할 수 있다. 질문자가 어떤 물체가 목표 물체인지 알게 되었으면, 답변 물체를 맞추겠다고 선언한다. 그러면, 이미지에서 후보 물체들이 segment 형태로 나오게 되고 그 중 하나를 선택하여 답을 맞추면 된다. 평가가 객관식으로 이루어지기에, 정확도라는 평가 지표로 평가하기에 좋다. 이 질문자와 답변자를 모두 인공지능으로 만드는 것이 GuessWhat?! 인공지능 문제의 목표이다. GuessWhat?!을 제안한 논문에서는 이를 위하여서는 컴퓨터 비전, 자연어 처리, 의사 결정, 계획 수립 등 다양한 인공지능의 과제들을 해결하는 것이 필요하다고 주장되었다.

GuessWhat?!은 기본적으로 두 인공지능을 질문자가 정답을 맞출 수 있도록 같이 학습 시키고, 이를 바탕으로 서로 대화 할 수 있는 인공지능 에이전트를 만드는 것을 가정하고 있다. 이를 위하여, 이미지와 후보 물체 및 답변에 대한 정보 뿐 아니라, GuessWhat?! 문제에 대하여 실제 사람들이 질의 응답한 데이터셋을 제공한다. GuessWhat?! 데이터셋은 은 66,537개의 이미지와 155,280개의 게임, 그리고 831,889개의 질문 답변 페어로 구성되어 있다.

GuessWhat?! 연구, 더 나아가 일반적인 딥러닝 기반 visual dialogue 연구의 궁극적인 목표가 단순히 두 에이전트의 게임에 대한 성능을 높이는 것은 아닐 것이다. 예컨대, 대안이 되는 목표는 두 에이전트가 만드는 대화가 그럴 듯 한지 여부일 수도 있고, 학습된 에이전트가 사람과 대화할 수 있는 지 여부일 수도 있다. 하지만, 기존 GuessWhat?! 연구들은 그러한 다른 목표들에 대해 명시적으로 기술하지 않았다[1,7].

GuessWhat?! 게임에서는 또한, 어떠한 정보가 질문자와 답변자가 공유될 수 있는 지가 논란이 될 여지가 있다. 명백하게도 두 인공지능 에이전트는 이미지 정보를 공유한다. 또한 명백하게도 질문자는 답변자와 달리 이미지 내의 정답 물체의 위치를 알지 못한다. 하지만 그 외에 대해서는 질문자와 답변자는 어떤 정보든 공유할 수 있다. 예컨대, 정답 물체가 사람인 경우, 문자열 “CODE_HUMAN”로 이루어진 질문에 대하여 yes로 답하도록 하는 사전 협의가 질문자와 답변자 사이에 이루어질 수 있다. 이것은 사람이 해석할 수 없는 암호를 만들 수 있다는 점에서 직관에 반한다. 하지만, 딥러닝 알고리즘이 서로의 질문과 답변 패턴을 숙지하도록 학습한다는 점에서 생각해보았을 때, 이상한 것은 아닐 수 있다.

알고리즘 1 제안된 위치 기반 질의 응답 시스템의 예시
Algorithm 1 Example of proposed coordinate-based question-answering system

First Question: “Is it left?”
Answer: If it is the first left side of the thirds, yes. If it is the first right side of the thirds, no. Otherwise, N/A.
Second Question: “Is it at the top?”
Answer: If it is the first top side of the thirds, yes. If it is the first bottom side of the thirds, no. Otherwise, N/A.
Third Question: “Imagine that it is at the left, is it the left side?”
...

3. 실험: 좌표 기반 방법

우리는 질문자와 답변자가 특정 물체에 대한 질문의 답변을 공유하고 있다면 GuessWhat?! 문제가 나무 탐색 문제와 유사하다는 점을 주목한다. 특히, 정답을 선택하는 상황에서는 정답 물체의 후보가 평균 9개 전후로 적어서, 현재 수준의 기술이나 인간 수준의 기술에 비하여, 알고리즘이 마주하게 되는 문제의 난이도가 낮다. 이러한 특성들은 인공지능의 깊은 사고와 딥러닝의 다양한 처리 없이도, GuessWhat?! 문제를 쉽게 해결할 수 있게끔 한다.

우리는 가장 간단한 규칙 기반 방법 중 하나로 이 문제를 해결한다. 이는 기본적으로 물체의 위치에 대해 질문하는 것이다. 실제 알고리즘은 물체의 좌표에 대해 묻는 형식이 된다. 그림 1과 알고리즘 1은 우리의 방법의 도식을 표 1은 기존 방법과 우리의 방법의 성능 차이를 보여준다. 우리의 방법은 2번의 질문만으로 56.3%의 성능을 얻어 기존 딥러닝 모델들보다 좋은 성능을 얻었으며, 5번의 질문으로 94.3%의 성능을 얻어 사람의 성능 수준에 도달하였다.

GuessWhat?!을 제안한 연구[1]에서는 baseline으로서 질문자와 답변자 모두 신경망 기반 방법을 제안했다. 이것은 컨볼루션 신경망과 순환 신경망을 기반으로 하여, 질문 문장 생성도 순환 신경망으로 하고, 그 답변도 신경망 기반 분류 모델로 수행하는 것이다. 하지만, 그 성능은 46.8%로 상당히 낮았으며, 심지어는 이미지를 사용하는 것이 큰 개선을 가지고 오지 않았다는 결론을 보고 하였다. 후속 연구로 강화학습을 사용한 방법이 제안되었으나, 그 성능은 53.1%로 큰 개선이 있지는 않았다[7]. 우리의 관점에 비추어 보면, 개념적으로 GuessWhat?! 문제가 네 다섯 번 정도의 질문 기회로 충분하기 때문에, 복잡한 강화학습의 탐색을 필요로 하지는 않는다.

기존 연구자들은 딥러닝 기반 시스템의 질문을 만드는 시스템의 성능에 문제가 있는 것으로 보고, 각 이미지에 대하여 사람이 사전이 질의 응답한 문장들을 가지고

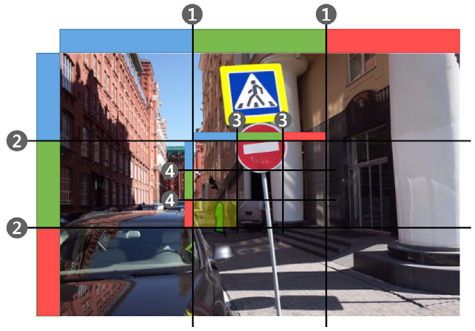


그림 1 제안된 좌표 기반의 질의 응답 시스템의 동작에 대한 도식
Fig. 1 A sequence of divisions of an image by coordinate-based search systems

표 1 GuessWhat?!에 대한 성능. 이미지 정보가 사용되지 않았음에도, 제안된 방법은 2번의 질문만으로 기존 딥러닝 기반의 모델보다 더 나은 성능을 보이며, 5번의 질문이 허용되는 경우 사람보다 더 나은 성능을 보인다.
Table 1 Performance on the GuessWhat?! dataset. Even if the image information is not used, the proposed method outperforms state-of-the-art deep learning methods in two turns and exceeds human performance in five turns

Model	Test Accuracy
Baseline	0.1604
1 Question	0.3896
2 Questions	0.5625
3 Questions	0.7661
4 Questions	0.8585
5 Questions	**0.9434**
1 Question w/ Fine-tune	0.3982
2 Questions w/ Fine-tune	0.5940
1 Question w/ Oracle Segment	0.4812
2 Questions w/ Oracle Segment	0.8767
LSTM-based System [1]	0.468
Deep RL System [7]	0.531
Using Human QAs [1]	0.618
Using Human QAs [7]	0.638
Human Performance [1]	0.908

LSTM 모델 기반 모델을 구축, 성능을 평가하여 60% 대 초반의 성능을 얻었다. 하지만, 이는 인간 수준의 성능 혹은 우리가 제안한 방법의 성능에 크게 미치지 않는 것이다.

우리는 왼쪽과 오른쪽의 경계에 대하여, 학습을 할 수도 있다. 물체는 주로 왼쪽보다는 가운데에 더 많이 있다. 우리는 학습 데이터의 통계치를 통하여, 왼쪽에서 첫 41%의 공간에 전체 물체의 1/3이 존재한다는 사실

을 알아냈다. 또한, 59% 이후 공간은 전체 물체의 1/3이 존재한다. 이러한 정보를 활용하여, 더 효율적으로 물체의 위치를 탐색할 수 있다. 위치에 대해서만 질문하는 경우, 이미지에 대한 정보를 사용하지 않는다고 하였을 때, 이러한 탐색은 최적 탐색을 보장하게 된다. 우리는 이 방법을 fine-tune 방법이라고 지칭한다. fine-tune에 대한 성능 보고는, 우리의 간단한 방법이 학습과 융합될 수 있음을 상징적으로 보여주는 것으로, 아주 약간의 성능 개선이 있었다.

이미지와 segmentation에 대한 정보가 없는 경우, 위치 기반 질의 응답은 아주 강력한 성능을 보고한다. segmentation이 있는 경우 없는 경우보다 더 좋은 성능을 만들 수 있을 것이다. 하지만, segmentation에 대해 완벽히 정보를 가지고 있는 경우에도, 위치 기반 질의 응답은 아주 강력한 성능을 보고한다. 정보를 질문자와 답변자가 모두 가지고 있기 때문에, 아주 정교하게 픽셀 단위로 물어보는 경우, optimal한 search가 가능하기 때문이다. Oracle segment는 이와 같이, 사전에 정답 물체의 후보를 정확히 알고 있다고 가정했을 때의 본 방법의 성능이며, 또한 GuessWhat?! 문제의 성능의 이론적인 상한선이다.

4. 논의

기존의 복잡다단한 딥러닝 방법과 달리, 우리는 좌표 기반으로 간단하게 GuessWhat?! 문제를 해결하였다. 하지만, 이러한 문제 해결 방법은 GuessWhat?! 연구자들이 원하는 방향과 일치하지 않는 것이다. 혹자는 우리의 접근 방법이 GuessWhat?! 문제에 대한 부정 행위이며, 인공지능 연구에 제한적인 기여밖에 할 수 없다고 주장할 수 있다. 기존 방법에 대한 예상되는 대표적인 반론은 두 가지이다. 첫째, 인공지능의 질문과 답변의 형태가 사람이 보기에 사람답지 않다는 점이다. 그림 2가 도식화하듯이, 질문자가 던지는 질문이 사람의 상식에 크게 벗어날 수 있다는 점에서 제안된 시스템은 문제가 있다는 것이다. 둘째, 이 인공지능은 사람과 같이 문제를 풀 수 없다. 비록 본 논문에서 제안한 질문자와 답변자는 GuessWhat?!에서 좋은 성능을 내겠지만, 인간

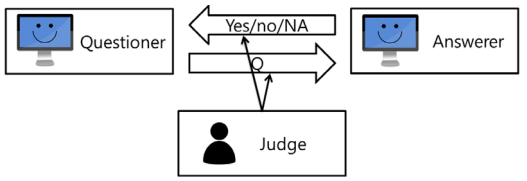


그림 2 사람 판단자가 포함된 GuessWhat?! 문제 정의 세팅의 도식

Fig. 2 GuessWhat?! problem setting with a human judge

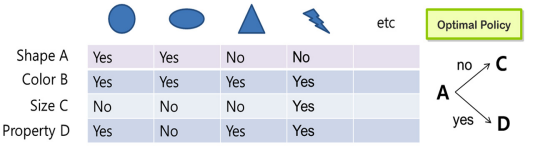


그림 3 정보이론적 관점에서 본 대화 질의응답의 도식. 후보 물체들을 잘 구분하는 질문을 선택하는 것이 정확도 향상에 유리하다.

Fig. 3 Illustration of the proposed approach on dialogue question answering. It is advantageous to select questions that can separate candidate objects evenly

과는 대화할 수 없다는 것이다.

첫째 반론에 대한 우리의 재반론은, 우리의 접근에서 위치 정보에 대한 질문이 다른 특징에 대한 질문으로 대체될 수 있다는 것이다. 물론 왼쪽에 있는 지 오른쪽에 있는 지, 위치에 대한 질문 그 자체는 사람들이 이미지에 대해 흔히 질문하는 일반적인 질문이 될 수 있다. 더 나아가, 위치에 대한 질문만으로 반복하여 보기를 좁혀가는 대신, 크기 기반 질문, 혹은 색에 대한 질문 등을 추가로 질문할 수 있다. 또한 강력한 딥러닝 물체 인식 성능을 바탕으로, 특정 물체인가 아닌가? 에 대해 질문할 수도 있다. 이러한 질문들은 사람들이 보기에 이상하지 않으면서도, 기계적으로 경우의 수를 줄이는 데에 도움을 준다. 그림 3은 일반 특징에 대한 적절한 질문들이 정답 물체 후보들의 선택지를 줄일 수 있음을 도식하고 있다.

둘째 반론에 대한 우리의 재반론은 retrieval 기반 시스템에 대한 제안으로 해결될 수 있다. 제안된 위치 기반 인공지능은 실제 사람과 잘 대답하지 못할 것이다. 사람들은 종종 약간 왼쪽에 치우친 물체에 대해서도 왼쪽에 있느냐는 질문에 '잘 모르겠음'보다는 '예' 라고 대답할 가능성이 높다. 하지만, 제안된 방법의 철학을 공유하는 다른 형태의 규칙들을 생각해볼 수 있다. 질문자와 답변자 중 답변자를 기존 딥러닝 방법처럼 만들었다고 가정하자. 이미지 질의응답 문제에 대하여 최근 상당한 진척이 있었으며, 한편으로는 답변자 시스템 구축은 질문자 시스템 구축과 별개의 문제로 볼 수도 있으므로, 답변자를 딥러닝 시스템으로 만드는 것 자체는 큰 문제가 안된다. 그 뒤에 후보 segment들에 대하여 답변자의 응답이 가장 크게 갈리는 문장들을 GuessWhat?! 학습 데이터의 질문 문장에서 하나 골라서 질문하는 형태로 질문자 시스템을 만드는 것을 생각하자. 계속하여 질문하다보면 앞서 설명된 규칙 기반 방법과 마찬가지로 질문들이 segment 후보들의 선택지를 줄일 수 있게 된다.

GuessWhat?! 문제를 단순한 딥러닝 기반 문제로 정의할 때, 가장 간과되는 지점은, 탐색에 대한 것이다.

GuessWhat?! 문제는 기본적으로 스무고개 문제에 이미지가 포함된 버전으로, 그림 3에서 도식하듯이, property map이 있으면, 이진 탐색 방법으로 문제가 해결될 수 있다. 이러한 성질은 GuessWhat?! 문제의 특수성이라기 보다는 대화 질의 응답 시스템의 본질적인 특성이라는 것이 우리의 생각이다. 우리는 사람들의 일상 대화가 종종 불확실성을 최소화하는 방향으로 이루어진다고 본다. 예컨대, 레스토랑에서 주문을 받는 상황에서 하는 질문들은 이전 질문들로도 확실해지지 않는 답변자의 의도를 더 명확하게 하기 위하여 주어지는 것들로 이해될 수 있다.

우리는 GuessWhat?! 문제의 분석과 이를 해결하는 알고리즘의 개발이 정보 이론적인 접근 속에서 이루어져야 한다고 생각한다. 문제 자체와 딥러닝 시스템 모두 답변 물체의 불확실성 혹은 엔트로피를 최소화하는 방향으로 추론을 수행하여야 한다. 하지만, 단순한 신경망을 그대로 사용하기에, 이러한 추론은 부자연스러우며, 실제로도 낮은 성능을 보인다.

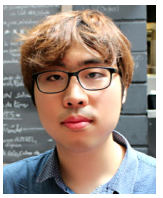
5. 결론 및 후속연구

GuessWhat?!은 인공지능 게임을 만들기 위한 훌륭한 새로운 시도였지만, 문제를 가지고 있었다. 본 논문에서는 GuessWhat?! 문제를 분석하고, 이를 바탕으로 두 번의 질문 만에 state-of-the-art 성능을 넘고, 다섯 번의 질문 만에 인간 수준의 성능에 도달하는 규칙 기반 solver를 개발하였다. 더 나아가, GuessWhat?!의 정보이론적인 접근에 대하여 논의하고, 예상되는 반론에 대하여 두 가지 추가적인 알고리즘들을 제안하는 방식을 통하여 논증하였다.

우리의 후속 연구에서 다룰 주제는 세 가지이다. 첫째, 우리는 정보이론적인 관점에서 GuessWhat?! 문제가 어떻게 이해될 수 있는 지 탐구하며, 이를 바탕으로 GuessWhat?! 문제가 어떠한 형태로 보완될 수 있을 지 논의할 것이다. 둘째, 이를 바탕으로, 질의응답형 대화를 불확실성을 최소화하는 방향으로 수행하는 프레임워크를 제안할 것이다. 셋째, 이러한 프레임워크를 기존 신경망 접근과 결합하여 다양한 질의응답 인공지능 대화 시스템으로 확장할 것이다.

References

[1] Ham de vries et al., "GuessWhat?! Visual Object Discovery through Multi-modal Dialogue," *CVPR*, 2017.
[2] Ian J Goodfellow et al., "Generative Adversial Nets," *NIPS*, 2014.
[3] Oriol Vinyals and Quoc Le, "A Neural Conversation Model," *ICML deep learning workshop*, 2015.
[4] Junhua Mao et al., "Generation and comprehension of unambiguous object descriptions," *CVPR*, 2016.
[5] Martin Arjovsky et al., "Wasserstein Gan," *arXiv*, 2017.
[6] Mike Lewis et al., "Deal or no deal? end-to-end learning for negotiation dialogues," *arXiv*, 2017.
[7] Florian Strub et al., "End-to-end optimization of goal-driven and visually grounded dialogue systems," *IJCAI*, 2017.



이 상 우

2012년 서울대학교 컴퓨터공학부 학사 2012년~현재 서울대학교 컴퓨터공학부 석박사 통합과정. 관심분야는 딥러닝, 베이지안, 평생학습, 멀티모달, 질의 응답, 대화 시스템

한 철 호

2012년 포항공과대학교 전자전기공학과 학사. 2012년~2013년 포항공과대학교 전자컴퓨터공학부 석박사 통합과정. 2014년~현재 서울대학교 컴퓨터공학부 석박사 통합과정. 2017년~현재 네이버 재직. 관심분야는 인공지능

허 유 정

2015년 인하대학교 컴퓨터공학과&IT경영학과 학사. 2015년~현재 서울대학교 컴퓨터공학부 석박사 통합과정. 관심분야는 인공신경망, 기계학습, 멀티모달학습, 컴퓨터 비전

강 우 영

2014년 전북대학교 컴퓨터공학부 학사 2015년~2017년 서울대학교 컴퓨터공학부 석사. 2017년~현재 써로마인드 로보틱스 재직. 관심분야는 컴퓨터비젼, 딥러닝, 인공지능

전 재 현

2013년 한양대학교 생체공학과 학사. 2013년~2015년 삼성전자 의료기기사업부 재직. 2017년~현재 서울대학교 뇌과학 협동과정 석사과정. 관심분야는 딥러닝, 컴퓨터 비전, 자연어 처리

장 병 탁

1997년~현재 서울대학교 컴퓨터공학부 조교수, 부교수, 교수. 2014년~현재 한국인지과학산업협회 회장. 2012년~2016년 서울대 인지과학연구소 소장. 2010년~2013년 한국정보과학회 인공지능소사이어티 회장. 2007년~2008년 삼성전자 종합기술연구원 초빙교수. 2003년~2004년 MIT 인공지능연구소(CSAIL) 초빙교수. 1992년~1995년 독일국립정보기술연구소(현 프라운호퍼) 선임연구원. 1992년 독일 Bonn 대학교 컴퓨터과학 박사. 1982년~1988년 서울대학교 컴퓨터공학과 학사 및 석사