

딥러닝을 이용한 인터랙티브 에이전트 개발

강윤제(*), 김범수(*), 배미경(*), 남경훈(*), 서상현(*)

(*) 중앙대학교 컴퓨터예술학부

{bluejay100@cau.ac.kr ; cmng828rhuyppqq@cau.ac.kr ; doks0607@cau.ac.kr ; kyteris0624@naver.com ; sanghyun@cau.ac.kr}

Interactive Agent using Deep Learning

Kang Yoon-je(*), Kim Bum-soo(*), Bae Mi-kyeong(*), Nam Kyung-hun(*), Seo Sang-hyun(*)

(*) Chung-Ang University, Department of Computer Art

요약

오늘날, 딥러닝을 기반으로 한 인공지능은 사람이 할 일을 대신해주고 자동화가 탑재되어 있다는 편리함 덕분에 일상생활에서 중요한 존재가 되고 있다. 하지만 이러한 인공지능들은 대체로 형태가 없고 음성만을 이용하는 경우가 많아 효과적인 상호작용이 이뤄진다고 보기 어렵다. 본 논문에서는 이러한 인공지능의 단점을 보완하기 위한 학습된 신경망 모델이 사용된 인터랙티브 에이전트를 개발하였다.

1. 서론

최근 딥러닝 기술은 그 성능이 증가함에 따라 기존에 해결하지 못한 문제들의 새로운 대안이 되며 핵심 기술로 주목받고 있다. 딥러닝은 공장의 자동화, 자율주행차량, 안면 인식 등 다양한 분야에 활용될 수 있고 활용 연구 또한 진행되고 있다[1]. 사용자의 생활패턴을 학습하는 인공지능도 그중 하나이며, 일상생활의 일부를 대체할 수 있다는 편리함 때문에 각광받고 있다. 하지만 대부분의 인공지능은 자연어 처리를 통해 음성의 형태로만 출력하는데, 인간은 외부 정보 중 약 70%정도를 시각을 통해 처리한다는 점에서 효과적인 상호작용이 이뤄진다고 보기는 어렵다[2]. 단순히 청각적 환경에서 디바이스를 사용한 것에 비해 시각적, 청각적 요소가 혼합된 복합적 환경에서의 사용자 만족도가 항상 높다고 한다[3]. 따라서 본 연구에서는 기존 인공지능이 가지는 단점을 보완하기 위해 시각적 요소를 추가하여 개발을 진행했다. Unity 게임 엔진을 사용해서 가상 에이전트와 이를 위한 Yolo(You Only Look Once) v4 딥러닝 모델 기반 프레임워크를 제작하여 복합적 상호작용이 가능한 인터랙티브 에이전트를 개발했다[4].

2. 연구내용

2.1 시각적 상호작용 첨가를 위한 딥러닝 모델 연구

가상으로 구현된 환경에 실제 사용자의 모습을 비춰 사용자가 콘텐츠에 몰입할 수 있는 시각적 상호작용이 가능하다는 점에서 웹캠을 사용했다. 웹캠의 사용은 실시간으로 객체 인식이 가능해야한다는 문제를 가지기 때문에 실시간 동작이 지원되는 모델이 필요했다. 이에 적합한 모델로 Yolo v4 모델을 선택하여 학습을 진행했다. Yolo는 다른 모델에 비해 실시간 동작 측면에서 뛰어나며 정확도도 상대적으로 높다는 장점을 가진다.

2.2 사용자 몰입감 증대를 위한 컨셉 및 시나리오 연구

사용자가 몰입하기 위해서는 어떤 목적으로 사용할 것인지와 해당 목적을 위한 시나리오 구현 설정이 우선시된다. 사용자의 범위를 대학생으로 한정했고, 목적은 정보 전달 및 엔터테인먼트로 정했다. 정보 전달을 주된 목적으로 하여 사용자가 이에 몰입할 수 있는 뉴스데스크 컨셉으로 설정했다. 정보 전달을 위해 해당 학과의 학사일정과 날씨 정보를 전달하여 사용자에게 편의를 제공할 수 있게 했다.

2.3 디스플레이 설계 및 프레임워크 구현

〈표 1〉은 사용한 도구의 목록과 생성된 파일에 대해 설명하고 있다. OpenCV for Unity 플러그인과 Yolo v4로 학습한 모델을 이용해 객체인식 기능을 구현했다. 해당 학과 홈페이지에서 얻은 학사 정보와 기상청 Api를 통해 얻은 날씨 정보를 csv파일 형태로 저장하는 응용프로그램과 csv파일에 기록된 텍스트를 Google TTS Api를 이용하여 wav형태의 음성 파일로 변환하여 저장하는 응용프로그램을 제작했다. 웹캠을 통해 사람이 인식되면 해당 인터랙티브 에이전트에 내장된 애니메이션과 저장된 wav파일이 재생되어 사용자로 하여금 인터랙티브 에이전트가 실제 말하는 듯한 느낌을 줄 수 있었다. 전달하는 정보의 최신화와 사용자 편의성을 위해 해당 콘텐츠를 일정 시간 간격으로 자동 업데이트가 되도록 설계했다. 따라서 제작한 인터랙티브 에이전트는 사용자와 목적에 따라 추후에 교체 및 변경이 가능하며 인식의 범주를 늘릴 수 있다는 점에서 프레임워크의 기능을 수행한다고 할 수 있다.

〈표 1〉 기능별 사용 도구 및 생성 파일

Function	Tool	File Format
Object Recognition	Yolo v4, OpenCV for Unity	
Information Collector	Weather Api, Web crawler exe, Google TTS Api,	csv, wav
Display	Unity 2019.4.8f1	Assets

〈Table 1〉 Used Tool Informations by Function
& Generated Files

3. 실험결과

(그림 1)은 인터랙티브 에이전트의 실행 화면이고 기상청에서 얻은 날씨 정보를 이용해 현재 비가 내린다는 정보를 시각적으로 잘 전달했으며 알맞은 음성이 출력되며 복합적인 상호작용이 이뤄졌다. 기존의 인공지능은 일상생활에서 편리함을 제공하는 존재이지만 대부분 청각적 요소의 형태로만 전달하여 효과적인 상호작용이 이뤄지기 어렵다는 단점을 가진다. 따라서 위의 결과를 토대로 기존의 인공지능이 가지는 장점인 편리함과 더불어 시각적 요소의 활용을 통한 단점의 보완으로 효과적인 상호작용이 가능하다는 것을 확인할 수 있었다.

(그림 1) 실행 결과



(Figure 1) Execution Result

4. 결론 및 향후 연구

본 논문에서는 사용자와의 효과적인 상호작용을 위한 인터랙티브 에이전트를 Unity를 사용하여 복합적 요소를 활용해 개발했다. 시각적 요소의 추가로 기존 인공지능의 단점을 극복한 인터랙티브 에이전트를 통해 효과적인 상호작용이 가능했다. 향후 연구 방향으로 사용자 층과 목적을 다양하게 설계하여 인터랙티브 에이전트의 활용 분야 및 가능성에 대한 연구를 진행할 것이다.

감사의 글

“이 논문은 2019년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. NRF-2019R1F1A1058715)”

참고 문헌

- [1] 고광은, 심귀보. (2017). 딥러닝을 이용한 객체 인식 및 검출 기술 동향. 제어로봇시스템학회지, 23(3), 17-24.
- [2] 김옥현, 이미선, 손진우. (1993). 신경회로망과 시각정보처리. [ETRI] 전자통신동향분석, 8(3).
- [3] 오형준, 이주환, 지용구. (2013). 모바일 디바이스의 프로세스 시간에 대한 사용환경의 변화에 따른 사용자의 만족도와 주관적 시간 인지 차이. 한국HCI학회 학술대회, 915-919.
- [4] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.