

사전학습 기반 생성모델을 이용한 정서적 지지형 디지털 휴먼 프로토타입 구현

송채정*, 이지항*⁺

*상명대학교 지능정보공학과

songchejung98@gmail.com, jeehang@smu.ac.kr

A prototype of digital humans capable of emotionally using deep generative models

Chejung Song*, Jee Hang Lee*⁺

*Department of AI and Informatics, Sangmyung University, Seoul, South Korea

⁺Corresponding author: Jee Hang Lee

요 약

메타버스의 산업적/학술적 가치가 증대되면서, 실세계 인간과 메타버스 내 디지털 휴먼과의 상호작용 시스템 또한 큰 조명을 받고 있다. 본 논문에서는 인간과 디지털 휴먼이 상호작용할 때, 인간의 발화에 대해 감성적 지지가 가능한 디지털 휴먼 프로토타입을 소개한다. 대화의 의미에 따른 동작 생성이 가능한 아바타 구축 공개 프레임워크를 도입하고, 사전학습모델을 바탕으로 감성적 지지가 가능한 심층 대화 생성 모델 기반 대화 시스템을 여기에 통합하여 인간의 감성 상태에 따른 동작과 대화를 진행하는 감성 지지형 디지털 휴먼 프로토타입을 구현하였다. 이러한 프로토타입을 고도화 하면, 향후 메타버스 기반 정신 건강 케어 및 디지털 치료제로의 확장이 가능할 것으로 사료된다.

1. 서론

COVID-19 로 인해, 수많은 사람들의 야외활동, 친목 활동에 제약이 생김에 따라 여가생활을 즐기는 방법으로 메타버스가 떠오르고 있다[1]. 메타버스란 ‘Meta’ 와 ‘Universe’의 합성어로, 일상생활과 경제활동이 가능한 디지털 세계이다. 메타버스가 각광받으면서, 인간-인간, 인간-디지털 휴먼, 디지털 휴먼-디지털 휴먼과 같은 다양한 형태로 상호작용 하는 기회가 급속히 늘고 있다[1, 2].

이러한 추세에 따라, 인간-인공지능으로 만들어진 디지털 휴먼과 대화를 할 수 있는 대화 시스템 또한 주목받고 있다. 이 대화-메시징 애플리케이션-의 주된 사용자는 디지털이 발달한 이후에 태어나 디지털 세상에 익숙한 밀레니엄 세대로 분석되며, 이들은 자기 표현 욕구가 강하고 사람과 직접 소통하는 것보다 화면을 통해 문자를 주고받는 것을 더 선호하는 경향을 보인다[3]. 또한 현재의 대화 에이전트 시장은 2020년 2월 연구개발특구진흥재단(INNOPOLIS)에서 조사하여 발표한 결과, 2019년 25억 7,120만 달러에서 연평균

29.7%씩 증가할 것으로 전망되었다. 이와 같은 배경들을 바탕으로 대화형 시스템은 이미 거대한 시장을 형성하였고, 밀레니엄 세대를 중심으로 더욱 성장할 것으로 예측된다[4].

대화형 시스템 (혹은 챗봇 시스템)은 정신건강 예방, 처치 부문에도 도움이 될 수 있는 것으로 알려져 있다. WHO 에 따르면, 전 세계적으로 정신건강 전문가들의 수가 부족하며, 이러한 문제로 인해 2017년 미국의 경우엔 정신질환을 앓고 있는 성인의 42.6%만이 치료를 받은 것으로 조사되었다[5]. 이러한 상황에서 직접 전문가를 대면하는 대신 대화형 챗봇 시스템을 이용한다면 시간과 공간의 제약 없이 스마트폰, 혹은 컴퓨터로 예방, 처치가 가능하며, 많은 치료 기회를 제공할 수 있다 [6].

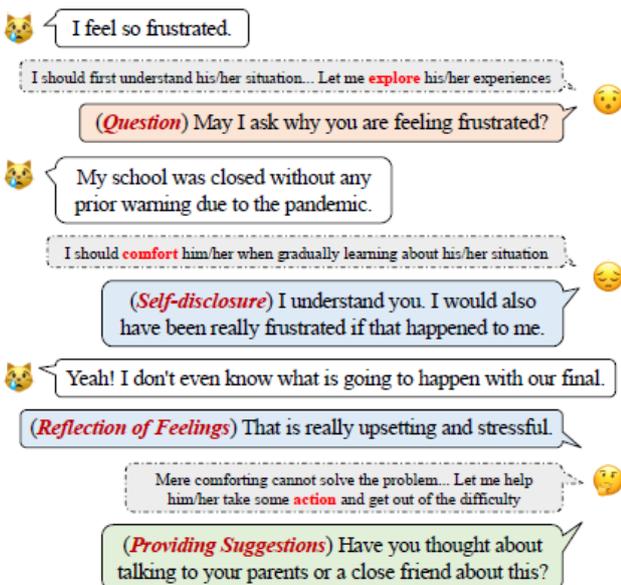
본디 대화를 나누는 것은 불안감과 우울감을 감소시킬 수 있으며, 신체적 조건 개선 효과도 있는데, 연구에 따르면 만성 뇌졸중 환자에게 공감 대화를 시도한 결과, 스트레스와 불안감, 우울증이 저하될 뿐만 아니라 신체적으로 긍정적인 효과를 볼 수 있었다[7].

본 논문에서는 이처럼 사람과 사람 사이에 대화하는 것과 비슷하게, 디지털 휴먼과 상호작용 하는 것이 스트레스와 우울, 불안 등의 정신적 문제를 개선하는 데 효과가 있을 것이라고 가정한다. 이를 위해, 본 논문은 인간의 대화 중 감정적 사건 혹은 맥락이 내포되어 있을 때, 감성적 대화와 행동을 보여주고 지지하는 반응이 가능한 디지털 휴먼 구현 사례를 보이고자 한다.

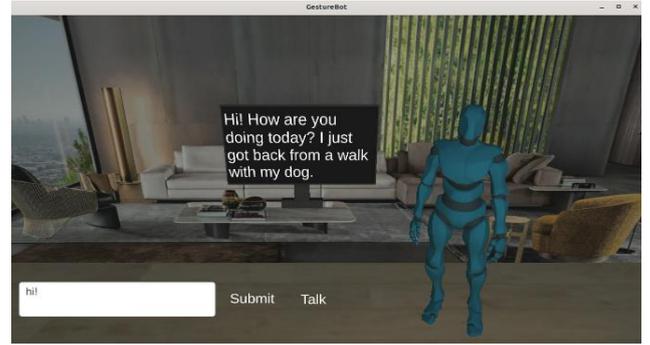
2. 기반 모델

사용자의 정신 건강지원을 수행하는 서비스를 제공하기 위해서는 디지털 휴먼이 상호작용 가운데 정서적 지원 대화를 이어갈 수 있어야 한다[8]. 또한, 인간이 언어적 표현과 더불어 비언어적 대화 행동 (예를 들어, 표정, 제스처)를 함으로써 감정과 의도를 나타내는 것처럼, 디지털 휴먼이 인간과 유사한 형태의 비언어적 정서 반응을 수행할 수 있다면, 더 우수한 형태의 정신 건강 지원을 위한 상호작용이 가능할 것으로 사료된다[9, 10].

언어적 상호작용의 경우, 정서 지원 대화가 가능한 Emotional Support Conversation 챗봇이 있다 (그림 1) [8]. 사전학습기반 심층 텍스트 생성 모델을 사용하여, 대화 중 사용자의 감성적 발화에 대해 공감하는 대화를 이어나갈 수 있는 챗봇이다. 여기서는 정서적 지원 절차를 3 단계, (i) Exploration, (ii) Comforting, (iii) Action 로 구성하였다[8, 11, 12]. 이 연구는 텍스트 기반의 챗봇 엔진을 제시한 것으로, 감성적 대화를 이어나가는 장점이 있으나, 디지털 휴먼의 형태로 사용자와 상호작용하는 시스템으로 구현된 바는 없다.



(그림 1) Emotional Support Conversation 예시 [8]

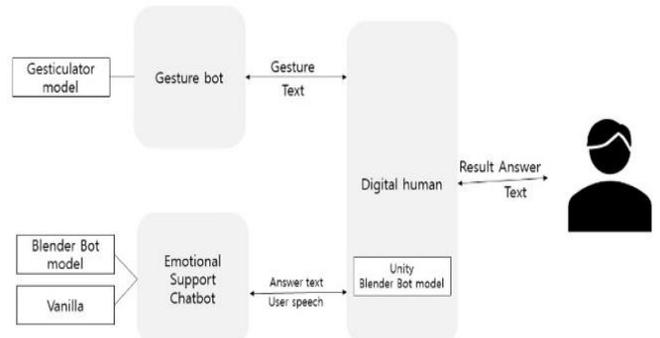


(그림 2) 공개 프레임워크 Gesture Bot 사례 [9]

비 언어적 상호작용의 경우, 최근 아바타 구축 공개 프레임워크로 Gesture bot 이 소개되었다[9]. Gesture bot 은 아바타, 챗봇 및 Text-to-Speech (TTS) 기능이 결합된 아바타 구현 프레임워크로, 언어/음성적 대화를 모두 지원하는 시스템이다. 특히 Gesture bot 프레임워크는 언어적 소통 뿐만 아니라 비 언어적 소통을 통해 대화를 통한 상호작용의 만족도를 높이고자 하였다. 챗봇을 통해 대화를 진행할 때, 대화의 의미에 따라서 아바타의 제스처를 생성하는 딥러닝 모델을 구축하여, 대화 중 비 언어적 추임새를 적극적으로 활용할 수 있도록 하였고, 사용자를 더 깊이 이해하고 지지하는 행동을 구현할 수 있도록 하였다. 다만, 일반적 챗봇 모델을 차용한 바, 대화를 통한 사용자 정서 지지 효과는 미미하고, 제스처 또한 일반적 대화에 국한된 표현으로, 사용자의 감성적 표현에 대한 정서적 지지를 위해서는 감성적 반응이 가능한 챗봇으로 발전할 필요가 있다.

3. 제안 시스템 구조

여기서는 사용자의 감성적 표현을 인지하여 언어적/비언어적으로 사용자의 정서를 지지할 수 있는 표현이 가능한 디지털 휴먼의 프로토타입을 소개한다. 앞서 소개한 두 최신 연구를 통합하는 방법을 통해 선형적 프로토타입을 구현하였다. 크게 정서적 지원 챗봇, 제스처 봇, 그리고 아바타로 이루어진다 (그림 3).



(그림 3) 제안 시스템 구조

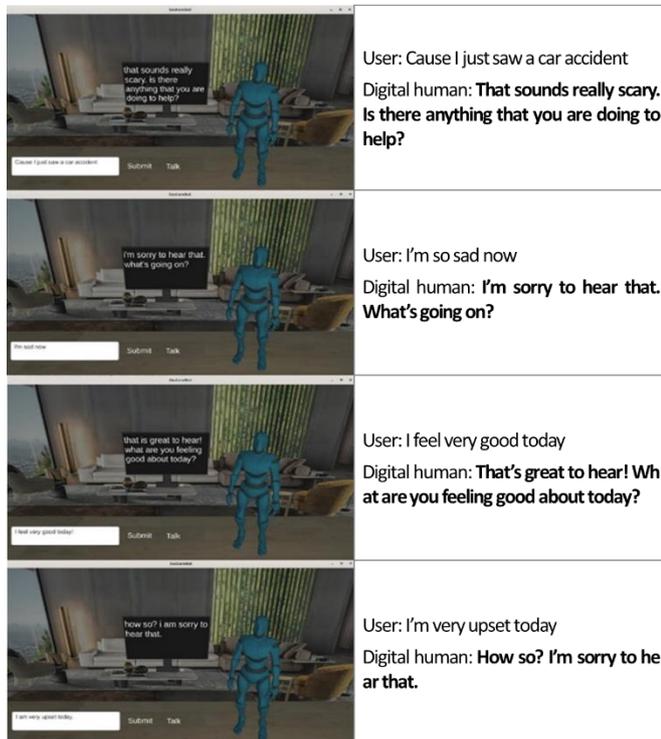
정서적 지원 챗봇은 Emotional Support Chatbot[8]을 사용하여 사용자의 발화에 대해 감성적 반응을 보이는 기능을 수행한다. 예를 들어, 긍정적인 기분 상태에서는 공감대를 해주는 답변을 출력해주고, 부정적인 기분에서는 무슨 일이 있는지를 물어보며 그에 대한 위로를 제공한다.

Gesture bot은 gesticulator 모델[9]을 사용하여 정서적 지원 챗봇의 언어적 반응에 부합하는 제스처를 생성한다. 정서적 지원 챗봇을 통해서 사용자의 발화에 대한 감성적 반응이 추론되면, 그에 어울리는 제스처를 만들어 아바타가 그 제스처를 사용자에게 보여줄 수 있도록 준비한다[9]. 예를 들어, 인사를 할 때는 반가운 제스처를, 때에 따라 사용자의 말에 같이 화를 내주는 상황에서는 동조하는 제스처를 취한다.

아바타 표현은 unity로 구현한 Blender bot[9]를 사용하였다. 정서적 지원 챗봇의 반응과 이에 부합하는 Gesture bot의 행동을 결합하여 디지털 휴먼의 형태로 제공한다.

4. 구현 사례

앞서 밝힌 바와 같이 베이스 모델을 통합하는 접근으로 프로토타입을 구현하였다. Gesture bot과 정서적 대응 챗봇 시스템을 replication한 후, 두 프로그램을 통합하여 사용자의 감정에 대응하는 답변이 나옴과 동시에 그에 어울리는 제스처를 출력할 수 있는 디지털 휴먼 프로토타입을 생성하였다.



(그림 4) 정서적 지지형 디지털 휴먼 프로토타입. 위로부터 당황, 슬픔, 기쁨, 분노 감정으로 대화한 결과.

그림 4는 4가지 감정, 당황/슬픔/기쁨/분노,에 해당하는 감성적 표현을 이용하여 대화를 시도하였을 때 나온 반응들을 보여준다. 각 감정별로 기쁨 상황에서는 내 기쁨의 이유에 공감해주는 것과 동시에 그 일에 대하여 자신의 감정을 표출하는 등의 답변을 주었고, 반대로 부정적인 상황에서는 감정을 위로해주면서 본인의 경험을 이야기하는 등의 방법으로 공감을 전달하였다. 또한, 그 상황을 타개할 수 있는 방법을 전달해 주기도 하는 등의 답변을 주었다.

동시에 대화를 하면서 제스처를 같이 출력하여 사용자의 말에 답변을 할 때 마다 그 상황에 어울리는 제스처를 보여주었다. 이 프로그램에서는 대화를 모두 입력하고 submit 버튼을 누르면, 화면에 대화 응답이 떠오르며 아바타가 움직인다. 하지만 아바타의 움직임이 한정적이고 소극적이라는 단점이 있었다. 제스처는 일반 대화 상황에서 훈련된 결과를 보여주는 바, 정서적 지원에 완전히 부합하는 동작 구현에는 한계를 보였다.

5. 결론 및 추가 연구

많은 사람들이 외출을 삼가며 온라인으로 메타버스와 가상현실에 관심을 가지고 이용하게 되어 챗봇 시장 또한 성장하였다. 이러한 챗봇은 많은 사람들에게 쉽게 접근이 가능하여 개인의 스마트폰, 혹은 컴퓨터를 이용하여 디지털 휴먼과 대화를 나눌 수 있다. 사람들은 대화하는 것으로부터 불안감, 우울, 스트레스 등을 감소시킬 수 있다. 이 논문에서는 이러한 연구들을 바탕으로 사람들과 직접 소통을 하는 것이 아닌 챗봇을 이용하여 대화하는 것에도 효과가 있을 것이라고 가정하였다. 대화를 하는데 비언어적 행동은 대화의 신뢰감을 높일뿐만 아니라 에이전트가 더욱 매력적으로 느껴질 수 있기 때문에 감정이 드러나는 대화를 할 때 텍스트뿐만 아니라 텍스트의 감정에 맞는 제스처를 같이 보여준다면 사용자의 감정에 미치는 효과가 더욱 커질 것이라고 예상된다.

이 논문에서 베이스 모델을 통합하여 언어적/비언어적 정서적 지지를 표현할 수 있는 디지털 휴먼의 프로토타입을 제시하였다. Gesture bot[9]과 정서 지지형 챗봇을 통합하여, 사용자의 감정에 따른 답변을 출력하며 비언어적 감성 지지 제스처 또한 같이 보여줄 수 있는 사례를 구현하였다. 추후 이 시스템을 이용하여 실제 사용자의 만족도를 확인하고, 감정, 심박 상태 등의 변화를 측정하여 감성적 지지의 효과성을 확인하고자 한다.

Acknowledgement

이 성과는 정부(과학기술정보통신부)의 재원으로 한국

연구재단의 지원을 받아 수행된 연구임 (No. 2020R1G1A1102683). 본 연구는 삼성미래기술육성센터의 지원을 받아 수행하였음 (No. SRFC-TC1603-52). 본 결과물은 교육부와 한국연구재단의 재원으로 지원을 받아 수행된 사회맞춤형 산학협력 선도대학 (LINC+) 육성사업의 연구결과임. 본 연구는 환경부 녹색복원 전문인력 양성사업으로 지원을 받아 수행한 과제임.

참고문헌

- [1] 고선영; 정한균; 김종인; 신용태 “문화 여가 중심의 메타버스 유형 및 발전 방향 연구” 정보처리학회 논문지 10 권 8 호, 331-338, 2021
- [2] Davis Alanah; Murpy John; Owens Dawn; Khazanchi Deepak; Zigurse Ilze “Avatars, People, and Virtual Worlds: Foundations for Research in Metaverses” Journal of the Association for Information Systems, 10(2), 2009
- [3] 연세대학교 경영연구소, “문자 기반 인공지능 챗봇 정의 및 활용사례” 4 차 산업혁명 브리프, No.6, 2018
- [4] INNOPOLIS, “챗봇시장 Marketsandmarkets, Chatbot Market, 2019”, 2020
- [5] Dosovitsky G.; Pineda BS.; Chang C.; Bunge EL. “Artificial Intelligence Chatbot for Depression: Descriptive Study of Usage” JMIR Form Res 2020, 4(11):e17065, 2020
- [6] Vijayarani M.; Balamurugan G., “Chatbot in Mental Health Care”, IndianJournal of Psychiatric Nursing 2019, 16:126-8, 2021
- [7] Yu-Won Choe; Myung-Kwon Kim, “Effects of Sympathetic Conversation on Electroencephalogram, stress, Anxiety-Depression, and Muscle Tone in Chronic Stroke Patients*”, Journal of the Korean Society of Physical Medicine, 13(3), pp. 99-111, 2018
- [8] Liu S.; Zheng C.; Damasi O.; Sabour S.; Li Y.; Yu Z.; Jiang Y; Huang M, “Towards Emotional Support Dialog Systems” ArXiv, 2021
- [9] Nagy R.; Kucherenko T.; Moell B.; Kjellstrom H.; Bernardet U., “A Framework for Integrating Gesture Generation Models into Interactive Conversational Agent”, AAMAS, 2021
- [10] Taras Kuncherenko; Patric Jonell; Sanne van Waveren; Gustav Eje Henter; Simon Alexanderson; Iolanda Leite; Hedvig Kjellstrom, “Gesticulator: A framework for semantically-aware speech-driven gesture generation.”, ICMI, 2020
- [11] Monroe B.; Colaresi M; Quinn K; Fightin ‘words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict. Political Analysis 16(4), 372-403, 2017
- [12] Clara E Hill “Helping skills: Facilitating, exploration, insight, and action”, American Psychological Association, 2009