음성 지원 대화형 AI 서비스를 활용한 정서 케어 지원

[1]유영호, [1]김세현, [1]박도현, [2]홍수지, [1]이병정
[1]서울시립대학교 컴퓨터과학부, [2]서울시립대학교 인공지능학과
netface12347@gmail.com, rlatpgus5432580@naver.com, qkrehgus02@uos.ac.kr, sujihong93@gmail.com, bjlee@uos.ac.kr

Towards Supporting Emotion Care by Utilizing Conversational AI Service with Voice

[1]Youngho Yoo, [1]Sehyeon Kim, [1]Dohyeon Park, [2]Suji Hong, [1]Byungjeong Lee
[1]Dept. of Computer Science and Engineering, University of Seoul, [2]Dept. of Artificial Intelligence, University of Seoul

요 약

본 논문은 다중 에이전트 아키텍처를 기반으로 사람의 정서 케어 지원을 돕는 대화형 AI 에어전트 서비스를 소개한다. 본 서비스는 사람의 감정을 반영하기 위하여 LLM 을 미세조정하고, 음성 데이터의 임베딩을 통하여 감정을 분석함으로써 사용자의 감정 맥락을 인지한다. 또한, 사용자의 맞춤형 대화를 지원하기 위해서 클라우드 기억 관리자를 활용한다. 이를 통해 기존의 사실 기반 질의응답에서 벗어나 사용자에게 지인과 대화하는 듯한 보다 친밀한 경험을 제공한다.

Ⅰ. 서 론

최근 대화형 AI 는 빠른 기술적 발전을 이루었으나, 주로 정보의 사실적 정확성을 높이는 데 초점을 맞추어 왔다. 이로 인해 사용자의 감정적 맥락을 인지하거나 과거 대화 이력을 바탕으로 개인화된 상호작용을 제공하는 데에는 명백한 한계가 존재한다. 이러한 한계는 특히 섬세한 접근이 요구되는 정서 케어 지원 분야에서 AI 의 실용성을 제한하는 요인으로 지적된다. 본 논문은 기존 대화형 AI 의 한계를 극복하기 위해, 자율적으로 목표를 수행하는 복수개의 에이전트들이 협력하는 새로운 아키텍처를 제안한다. 본 서비스는 개인화된 정서 지원, 장기 기억, 그리고 실시간 감정 인식의 세 가지 핵심 분야를 기반으로 한다. 서비스 동작 원리는 먼저, 사용자의 음성 입력에서 음향 특징을 추출하여 감정을 분석함과 동시에, 음성 입력을 텍스트로 변환한다. 다음으로 이렇게 변환된 텍스트와 감정 정보는 외부 지식을 활용해 미세조정된 LLM 에 전달된다. LLM 은 클라우드 기억 관리자에 저장된 과거 대화, 토픽, 감정 변화 데이터를 참조하여 개인에게 맞춤화된 일관성 있는 답변을 생성한다. 최종적으로 생성된 텍스트는 음성으로 변환되어 사용자에게 제공된다.

Ⅱ. 관련 연구

최근 AI 연구는 사용자의 정신 건강을 지원하는 개인화된 시스템으로 확장되고 있다. 다중 에이전트 아키텍처를 통해 맞춤형 정신 건강 지원을 제공하는 연구[1]는 AI 의 정서적 지원 역할에 대한 가능성을 제시했다. 이러한 고차원적 지원의 전제 조건은 사용자와의 신뢰이며, 이는 AI 가 과거 대화를 일관성 있게 기억하는 능력에 의존한다. 이를 위해, 과거 대화를 성찰하고 요약하여 깊이 있는 관계 형성을 목표로 하는 '성찰적 기억 관리' 메커니즘이 제안되었다[2]. 나아가, 효과적인 정서적 교감은 사용자의 현재 감정 상태를 정확히 파악하는 능력에 크게 좌우된다. 기존 텍스트 분석의 한계를 넘어, 속삭임과 같은 미세한 음성 신호에서 감정을 직접 인식하는 연구[3]는 음성 데이터가 가진 풍부한 감정 정보의 유용성을 보였다. 이러한 연구들은 정서 지원 프레임워크[1], 기억 메커니즘[2], 음성 감정 인식[3] 등 각 요소 기술에서 중요한 진전을 이루었다. 그러나 이들을 실시간 음성 기반 시스템으로 유기적으로 통합하려는 시도는 부족했다. 따라서 본 연구에서는 이러한 기술적 공백을 참고하여, 세 가지 핵심 개념을 적용하여 대화형 AI 기반 정서 케어 지원 서비스를 제안한다.

Ⅲ. 음성 지원 대화형 AI 서비스

3.1 유즈케이스도

그림 1 은 본 음성 지원 대화형 AI 서비스 유즈케이스도이다. Converse 는 핵심적인 유즈케이스로 사용자가 음성, 텍스트를 통하여 시스템과 데이터를 주고받는다. RequestContents 는 사용자가 자신의 감정 상태에 맞는 케어 컨텐츠를 요청한다. SearchConversation 은 사용자 요청에 맞는 과거 대화 이력을 적절한 형태로 제공해준다.

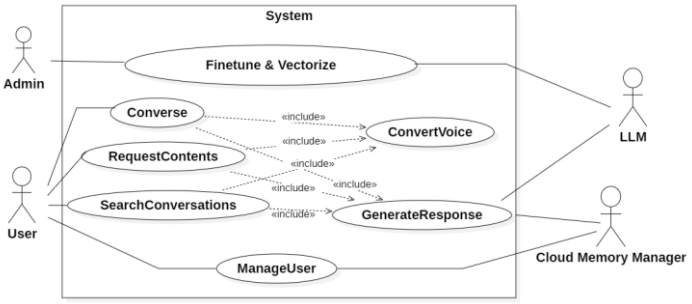


그림 1. 유즈케이스도

GenerateResponse 는 앞선 3 개의 유즈케이스에서 넘어온 입력을 미세조정된 LLM 에 전달하여 답변을 생성한다. ManageUser 는 회원정보에 관한 유즈케이스이며 사용자 별로 맞춤형 대화를 지원하기 위해 사용자 정보를 기억한다. ConvertVoice 는 음성과 텍스트를 서로 변환하고, 음성의 감정 벡터를 추출하는 기능이다. Finetune & Verctorize 은 더 정확한 감정 대화와 컨텐츠 추천을 위해 LLM 을 연관된 전문적인 데이터로 미세조정한다. 또한, 전문적인 데이터를 벡터화 한 후 저장하여 사용자 입력이 들어오면 의미 검색을 통해 정보를 추출하여 LLM 의 입력으로 사용한다.

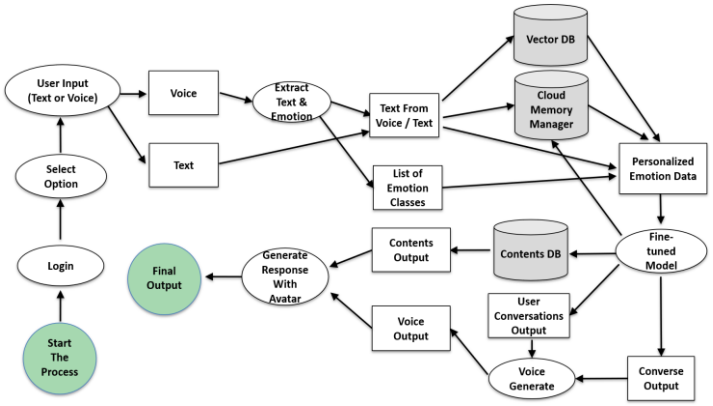


그림 2. 전체 작업흐름

그림 2 는 본 서비스의 전체적인 흐름을 보여주는 작업 흐름도이다. 사용자는 서비스를 사용하기 위해 회원가입하고 로그인한다. 그리고 Converse, RequestContents, SearchConversation 3 가지 옵션 중 하나를 선택하고 음성이나 텍스트로 원하는 입력을 넣는다. 음성 입력이라면 음성으로부터 벡터를 추출해 감정을 판단함과 동시에 음성을 텍스트로 변환한다. 텍스트 입력이라면 별도의 처리 과정 없이 진행된다. 이후 텍스트를 기반으로 외부 상담 데이터와 같은 전문적인 데이터가 들어있는 벡터 DB 에서 의미 검색을 통해 관련 정보를 추출한다. 동

시에 사용자에 대한 기억을 담당하는 클라우드 기억 관리자에서 현재 입력과 관련있는 사용자 정보를 찾는다. 이 모든 정보가 통합되어 미세조정된 LLM 에 입력으로 사용되고 선택한 옵션에 적합한 정보가 출력된다. LLM 의 응답은 사용자 입력과 함께 클라우드 기억 관리자에 저장되어 다음 응답 생성에 활용된다. Converse 와 SearchConversation 의 경우 LLM 의 답변을 그대로 음성으로 변환해 움직이는 캐릭터(아바타)와 함께 출력한다. RequestContents 의 경우 LLM 의 응답을 바탕으로 Contents 가 보관된 DB 에서 컨텐츠를 찾아서 움직이는 캐릭터(아바타)와 함께 출력한다. 본 연구의 작업흐름은 참고문헌[1]의 구조를 기반으로 하되, 음성 중심의 실시간 상호작용에 최적화하기 위해 전체 과정을 간소화하고 개인화된 경험을 위한 클라우드 기억 관리자를 추가하여 재구성했다.

3.2 시스템 아키텍처

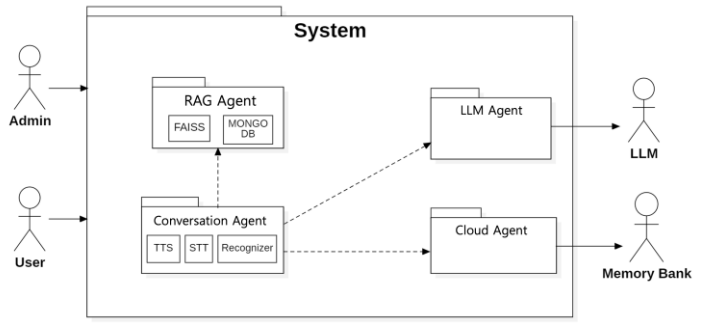
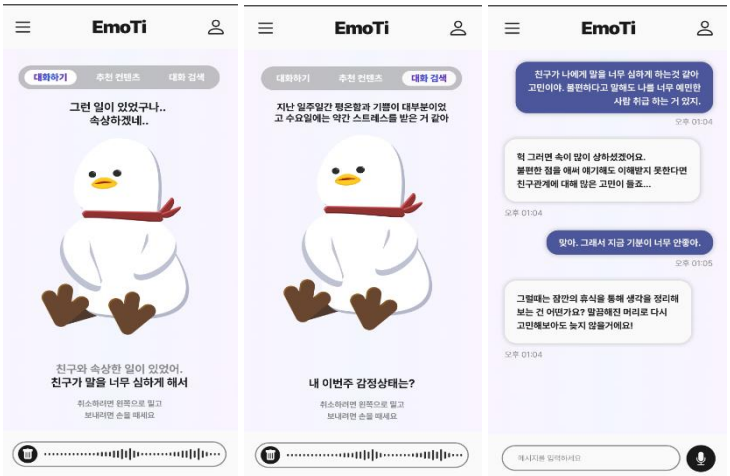


그림 3. 아키텍처도

본 연구에서 제안하는 시스템의 전체 아키텍처는 그림 3 과 같다. 본 시스템은 대화 에이전트(Conversation Agent)를 중심으로 다른 모든 컴포넌트를 유기적으로 호출하고 관리하는 구조이다. 시스템의 모든 상호작용은 대화 에이전트를 통해 시작된다. 먼저 STT(Speech-to-Text)로 사용자의 음성 입력을 인식하고, Recognizer 를 통해 그 안에 담긴 감정 상태를 파악한다. 이렇게 파악된 사용자의 의도와 감정을 기반으로, 대화 에이전트를 답변을 더욱 풍부하게 만들기 위해 필요한 정보를 수집한다. 이를 위해 RAG Agent 를 호출하여 FAISS[i] 및 MongoDB[ii]에서 관련 외부 사실을 검색하고, Cloud Agent 를 호출하여 클라우드 기억관리자, 즉 메모리 뱅크(Memory Bank)[iii]에 저장된 대화 이력이나 과거 감정 상태 같은 장기 기억 데이터를 추출한다. 이렇게 종합된 정보를 바탕으로 LLM Agent 를 호출하여 LLM 을 위한 최적의 프롬프트를 생성한다. 마지막으로, LLM 으로부터 받은 답변을 TTS(Text-to-Speech)를 통해 자연스러운 음성으로 변환하여 사용자에게 최종적으로 전달한다. 한편, 관리자(Admin)는 별도로 시스템의 전반적인 성능 향상을 위한 Finetune 및 Vectorize 작업을 수행한다.

Ⅳ. 대화 서비스 UI



(a) 대화하기 (b) 대화 검색 (c) 채팅 로그

그림 4. 서비스 UI

그림 4 는 본 연구에서 제안하는 음성 지원 AI 대화 서비스 'EmoTi'의 사용자 인터페이스를 보여준다. 본 서비스는 텍스트와 음성 입력을 모두 처리할 수 있는 모델을 기반으로 설계되었으며, 사용자는 자신의 선호나 상황에 따라 입력 방식을 자유롭게 선택하여 시스템과 상호작용한다. 시스템의 주요 기능은 UI 의 상단 탭을 통해 접근 가능하다. '대화하기'는 사용자의 일상 및 감정에 대한 전반적인 대화를 수행하는 핵심 기능으로, 시스템은 사용자의 발화에 공감하고 정서적 지지를 제공하는 상호작용을 목표로 한다. 예를 들어, 그림 4 의 좌측 화면(a)에서는 사용자가 친구와의 갈등으로 인한 속상함을 토로하자, 시스템이 공감적 반응을 보이는 것을 확인할 수 있다. '추천 컨텐츠' 기능은 대화의 맥락과 사용자의 감정 상태를 분석하여 비언어적 콘텐츠를 제공하는 역할을 수행한다. 이는 캐릭터의 제스처가 동반된 노래, 시 낭송 등의 형태로 제공되어 사용자에게 다각적인 위로와 즐거움을 전달한다. '대화 검색'은 축적된 대화 이력을 기반으로 사용자가 특정 정보를 검색하거나 자신의 감정 변화를 회고할 수 있도록 지원한다. 그림 4 의 중앙 화면(b)에 나타난 바와 같이, 사용자가 "내 이번주 감정상태는?"과 같은 질의를 하면, 시스템은 과거 대화 내용을 요약 및 분석하여 유의미한 인사이트를 제공한다. 이러한 핵심 기능들과 더불어, 본 서비스는 UI 측면에서 몰입감 높은 '캐릭터 모드'와 텍스트 기반의 '채팅 모드' 를 모두 지원한다. 그림 4 의 좌측 및 중앙 화면처럼 캐릭터와 음성 또는 텍스트로 상호작용하면, 해당 대화 내용은 우측 화면(c)의 채팅 로그에 실시간으로 기록된다. 이와 같은 유연한 UI 구조는 캐릭터와의 감성적 교감을 깊게 하면서도 대화의 연속성과 정보제공을 보장함으로써 사용자 경험을 극대화하는 효과를 가진다.

Ⅴ. 결론

본 논문에서는 사용자에 대한 심층적 기억을 참고하여 섬세한 감정적 응답으로 정서 케어 지원하는 서비스를 소개하였다. 기존의 단순 사실 기반 응답이 아닌 목소리의 비언어적 특징을 고려한 감정적 소통이 가능하며, 현재의 질문에만 의존한 답장에서 벗어나 사용자와의 대화를 체계적으로 저장하고 참조하여 사용자에게 맞춤형으로 대화를 제공한다. 또한, 각 기술을 자율적인 에이전트로 재정의하고 이들의 협력을 통해 정서 케어라는 복합적인 문제를 해결하는 새로운 아키텍처를 제안했다는 점에서 의의를 가진다. 그러나 표정과 행동 등의 또 다른 비언어적 특징을 고려하지 못한 한계가 있다. 향후에는 실시간 영상 분석을 통한 표정 인식과, 이를 기반으로 한 감정 벡터 표현 기법을 결합하여 보다 현실적이고 몰입감 있는 감정 교류형 AI 서비스로 확장할 것이다. 이러한 발전은 사람과 AI 의 상호작용을 한층 더 자연스럽고 의미 있는 방향으로 진행하게 할 것이다.

참 고 문 헌

[1] Pattar V., Patil T., Dhuri B., Karel A., Deole A., and Kulkarni S., "*Mindcare: A Multi-Agent AI Architecture for Personalized and Responsible Mental Health Support*," International Journal of Research and Scientific Innovation (IJRSI), vol. 12, no. 4, Apr. 2025.

[2] Tan Z., et al., "*In Prospect and Retrospect: Reflective Memory Management for Long-term Personalized Dialogue Agents*," in Proc. of Annual Meeting of the Association for Computational Linguistics (ACL), Aug. 2025.

[3] Qu X., Sun Z., Feng S., Chen C., and Tian T., "*Breaking the Silence: Whisper-Driven Emotion Recognition in AI Mental Support Models*," in Proc. of IEEE Conference on Artificial Intelligence (CAI), 2024.

[i] https://github.com/facebookresearch/faiss

[ii] https://www.mongodb

[iii] https://gemini.google.com/app/08823167abb96296