

페르소나 기반 대화형 에이전트 최신 연구 동향 분석

서영민
(주)브레인덱

ym@braindeck.net

Analysis of The Research Trends in Persona-based Conversational Agents

Seo Youngmin
Braindeck Inc.

요약

대화형 에이전트는 의료, 교육, 금융 등 다양한 분야에서 널리 사용되고 있으며, 특히 사용자에게 개인화된 경험을 제공하는 페르소나 기반 대화형 에이전트에 대한 연구가 주목받고 있다. 페르소나 기반 대화형 에이전트는 사용자와의 상호작용에서 높은 몰입감과 만족감을 제공할 수 있다는 장점이 있다. 최근에는 대규모 언어 모델을 기반으로 프롬프트 엔지니어링, 파인튜닝, 검색 기반 방법 등 다양한 방법론이 활용되고 있다. 본 논문에서는 페르소나 기반 대화형 에이전트의 연구 동향을 분석하고, 페르소나 정의 방법과 평가 지표의 표준화, 그리고 한국어 데이터셋 구축의 필요성을 비롯한 향후 연구 방향을 제시한다.

I. 서론

대화형 에이전트(Conversational Agent) 혹은 챗봇(Chatbot)은 의료, 교육, 금융, 고객 상담, 등 다양한 분야에 널리 사용되고 있다. ChatGPT, Claude, Gemini와 같은 LLM(Large Language Model) 등장 및 발전 이후 많은 연구들이 LLM 기반으로 진행되고 있다.

오늘날 대화형 에이전트는 단순히 질의응답을 뛰어넘어 사용자에게 대한 정보를 기반으로 개인화된 응답을 제공하는 것을 목표로 한다. 이는 개인의 정보가 훨씬 더 민감하고 개인적인 교육, 의료, 금융과 같은 분야에서 특히 중요하게 작용한다[1].

최근의 대화형 에이전트 연구는 화자의 페르소나, 지식 기반 정보와 같은 정보를 이용한 학습 방법을 시도되고 있다[2]. 이 중 페르소나 기반의 에이전트는 의료, 게임, 음악, 영화 등 다양한 분야에서 밀착형 개인 경험을 제공함으로써 사용자에게 훨씬 더 높은 만족감 및 효능을 제공할 수 있다.

본 논문은 개인화 및 페르소나에 대해 설명하고, 학습 방법과 모델의 최신 동향에 대해 살펴본다. 추가적으로 한국어를 포함한 페르소나 기반 대화형 에이전트의 벤치마크 데이터셋과 실험 설계의 문제점을 살펴보고, 현재까지 진행된 연구들의 한계와 전망을 제시하며 결론을 맺는다.

II. 본론

1. 개인화(Personalized) 및 페르소나(Persona)

개인화와 페르소나는 대화형 에이전트 연구에서 중요한 개념으로 자리 잡고 있다. 페르소나 기반 대화형

에이전트는 큰 범주에서 본다면 개인화(Personalized) 대화형 에이전트에 포함되는 것이 맞으나, 많은 연구에서 개인화 대화형 에이전트와 페르소나 기반 대화형 에이전트를 구별해서 사용하고 있다.

개인화 대화형 에이전트는 기존 질의응답 형태의 챗봇에서 개인의 정보에 맞춰 더 세분화된 기능을 제공한다[3].

페르소나 기반 대화형 에이전트의 연구는 통적으로 인간과 흡사하게 동작하는 것을 목표로 한다. 인간과 같은 특성을 구축하고 반영하는 방향으로 연구가 진행되고 있다[2], [3].

2. 페르소나 설정

사람의 성격 설정은 심리학이나 정신병리학에서 널리 사용되던 지표인 Big-5, 16 성격 요인, 대인관계 반응지수 등이 있다. 기존에 정립된 지표를 사용하는 것은 전문적인 지식을 필요로 하고, 데이터 라벨링의 복잡성이 증가하는 한계점을 지닌다. 최근의 연구는 기존 지표가 아닌, 대화의 발화를 이용해 페르소나의 성격 특성을 구축하려는 시도로 이어지고 있다[4].

3. 방법론

프롬프트 엔지니어링 기법 중 하나인 n-shot 학습과 롤플레이는 페르소나 일관성이 있는 대화 몇 가지(n개)를 제공하고, 모델은 해당 예시의 패턴과 스타일을 따라 응답을 생성하는 학습 방법을 의미한다. 학습 방법은 쉬우나 프롬프트 엔지니어링 방법은 다양한 특성과 사람과 같이 행동하는 페르소나가 있는 대화형 에이전트를

표 1 페르소나 기반 대화형 에이전트의 벤치마크 데이터셋

데이터셋 명칭	주제	언어	참고문헌
PERSONA-CHAT	클라우드 소싱으로 수집된 개인 채팅 데이터	영어	[5]
FoCus	위키피디아 지식을 주제로 대화한 데이터	영어	[6]
DuLeMon	자체 제작한 챗봇-사용자 역할 대화 데이터	중국어	[7]
Harry Potter Dialogue	해리포터 영화 및 소설의 대화 데이터	영어	[8]
LaMP	7 가지 작업에 대한 데이터	영어	[9]
RoleBench	영화 대사를 비롯한 다양한 스크립트	영어, 중국어	[4]
RoleEval	다양한 분야의 300 명의 인물과 가상의 인물	영어, 중국어	[10]
ChatHaruhi	32 개의 창작 캐릭터의 대화 데이터	중국어	[11]
페르소나 대화	AI Hub 에서 제공하며 20 개의 주제에 대한 데이터	한국어	-

만들기에는 부적합하다.

파인튜닝 기반 방법은 페르소나 대화 데이터를 기반으로 실행하게 된다. 성능 자체는 더 좋아졌지만 반대로 모델의 환각(Hallucination)이 더 심해지고, 기존 LLM의 영향을 강하게 받는다는 결과가 보고되었다[11].

검색 증강 생성(Retrieval Augmentation Generation; RAG)은 LLM의 환각을 감소시키고, 개인 맞춤형 응답을 제공할 수 있어 최근 활발히 연구되고 있는 학습 방법이다[12]. 페르소나 기반 대화형 에이전트에서도 파인튜닝의 한계점을 극복하기 위해 RAG를 이용한 연구가 시도되고 있다[13].

4. 실험 설계 및 데이터셋

페르소나를 정의하는 방법에 따라 실험 모델을 비롯한 실험의 전체적인 내용도 영향을 받는다. 현재까지 페르소나 기반 대화형 에이전트 연구들은 자체적인 페르소나 정의를 사용하고 있어 연구 결과를 비교하기 어렵다는 한계점이 있으나, 앞서 언급했듯이 발화를 이용해 실험을 설계하고 평가하려는 시도가 이어지고 있다[4].

페르소나 기반 대화형 에이전트의 벤치마크 데이터셋 대부분은 영어와 중국어를 기반으로 하고 있으며, 이를 각 국가의 언어로 번역한 경우가 많다. 한국어의 경우 기존 영어 대화를 기계번역을 거쳐 이를 학습 및 실험용 데이터셋으로 사용한 실험이 진행되었으나[14], 아직까지 페르소나 기반 대화형 에이전트 연구에 있어 한국어에 대한 데이터셋 구축은 미미한 편이다. 대화는 사용자의 습관, 성향, 문화적인 부분과 같은 비정형적 요소가 많으므로 한국어를 기반으로 한 더 많은 데이터셋 구축이 반드시 필요하다.

III. 결론

본 논문에서는 페르소나 기반 대화형 에이전트가 개인화 대화형 에이전트와의 차이점, 그리고 방법론과 실험 설계 및 데이터셋에 대해 살펴보았다.

페르소나 기반 대화형 에이전트는 아직까지 각 연구 별로 데이터셋과 성능 평가 방법이 상이하여 직접적인 모델의 비교가 어렵다는 한계점이 존재하지만, 최근 몇몇 데이터셋을 바탕으로 표준화가 진행되고 있어 향후 귀추가 주목된다.

또한 데이터셋을 비롯해 페르소나 기반 대화형 에이전트는 영어와 중국어를 바탕으로 진행되고 있으므로 데이터셋의 한국어 번역과 더불어 한국어 기반 데이터셋 구축이 반드시 필요하다.

향후 번역된 페르소나 벤치마크 데이터셋을 이용한 실험 설계 및 연구를 진행해 다른 모델과의 성능을 비교해볼 예정이다. 추가적으로 한국어에 기반을 둔 새로운 벤치마크 데이터셋을 다른 기관과 협의해 구축 예정이다. 본 논문이 향후 페르소나 기반 대화형 에이전트 연구의 초석이 되기를 기대한다.

참 고 문 헌

- [1] I. Hendrickx, "Take Back Control: User Privacy and Transparency Concerns in Personalized Conversational Agents".
- [2] S. Gao *et al.*, "PeaCoK: Persona Commonsense Knowledge for Consistent and Engaging Narratives," May 26, 2023.
- [3] R. Sutcliffe, "A Survey of Personality, Persona, and Profile in Conversational Agents and Chatbots," Dec. 31, 2023.
- [4] Z. M. Wang *et al.*, "RoleLLM: Benchmarking, Eliciting, and Enhancing Role-Playing Abilities of Large Language Models," Apr. 24, 2024.
- [5] S. Zhang, E. Dinan, J. Urbanek, A. Szlam, D. Kiela, and J. Weston, "Personalizing Dialogue Agents: I have a dog, do you have pets too?," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Melbourne, Australia: Association for Computational Linguistics, 2018, pp. 2204–2213.
- [6] Y. Jang *et al.*, "Call for Customized Conversation: Customized Conversation Grounding Persona and Knowledge," May 16, 2022.
- [7] X. Xu *et al.*, "Long Time No See! Open-Domain Conversation with Long-Term Persona Memory," Mar. 14, 2022.
- [8] N. Chen *et al.*, "Large Language Models Meet Harry Potter: A Bilingual Dataset for Aligning Dialogue Agents with Characters," Oct. 09, 2023.
- [9] A. Salemi, S. Mysore, M. Bendersky, and H. Zamani, "LaMP: When Large Language Models Meet Personalization," Jan. 09, 2024.
- [10] T. Shen, S. Li, Q. Tu, and D. Xiong, "RoleEval: A Bilingual Role Evaluation Benchmark for Large Language Models," Feb. 16, 2024.
- [11] C. Li *et al.*, "ChatHaruhi: Reviving Anime Character in Reality via Large Language Model," Aug. 18, 2023.
- [12] P. Lewis *et al.*, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," Apr. 12, 2021.
- [13] H. Wang *et al.*, "UniMS-RAG: A Unified Multi-source Retrieval-Augmented Generation for Personalized Dialogue Systems," Jan. 24, 2024.
- [14] J. Yoonna *et al.*, "Persona-based Korean Conversational Model," 한글 및 한국어 정보처리 학술대회 논문집, pp. 453–456, Oct. 2021.