유저 페르소나를 위한 인터랙티브 챗봇 시스템

최건*,1, 김강훈*,2, 김현재*,3, 심규환*,4, 구명완†,5
서강대학교 컴퓨터공학과

An Interactive Chatbot System for User Persona, KIRINO

Geon Choi*,1, Ganghun Kim*,2, Hyunjae Kim*,3, Kyuhwan Shim*,4, Myoung-wan Koo†,4
choigochoigun[1]@gmail.com, {iop1091[2], peterhyunjae[3]}@naver.com, {kyuhwan.shim[4], mwkoo[5]}@sogang.ac.kr
Department of Computer Science and Engineering, Sogang University
*equal contribution, † corresponding author

요 약

본 연구는 대화 시스템에서 대규모 언어 모델(LLM)을 활용해 사용자의 페르소나를 학습하고, 이를 정보 검색 기반 증강 생성(RAG) 아키텍처와 결합하여 사용자의 말투와 언어적 특징을 반영한 맞춤형 응답을 생성하는 KIRINO 아키텍처를 개발하였다. 기존 챗봇이 단순히 사용자의 질문에 답변하는 것과는 대조적으로, KIRINO 아키텍처는 사용자의 페르소나와 말투를 반영하여 자연스럽고 일관된 대화를 제공함으로써 대화 품질과 사용자 경험을 향상시킨다. 기존의 일관성 없는 답변이나 맥락에서 벗어난 이야기를 하며, 때로는 재미없는 답변을 하는 기존 챗봇의 문제를 해결하기 위해 본 시스템은 페르소나 기반 대화 시스템 개발에 중점을 두었다. 제안된 알고리즘을 바탕으로 G-Eval과 및 Human Eval을 수행한 결과 각각 26%, 38% 향상되었음을 확인할 수 있었다. 이는 대화의 품질과 자연스러움을 크게 향상시킬 수 있음을 보여주며, 대화 인터페이스의 품질 향상에 기여할 수 있음을 시사한다.

1. 서 론

최근 대화형 인공지능 시스템의 발전은 자연스러운 대화 흐름[1]과 개인화된 응답 제공에 중점을 두고 있으며, LLM은 이러한 시스템의 핵심적인 기술이다[2, 3, 4]. 기존 LLM 기반 대화 시스템은 사용자로부터 질문을 받고 답변하는 형태에 머물러 있다. 본 연구에서는 이를 극복하기 위해, 학습 과정에서 사용자의 대화로부터 페르소나를 구성하고 새로운 대화가 추가될 때마다 페르소나와 말투 정보를 갱신하도록 한다. 그리고 해딩 모델이 상대방의 질문에 응답을 생성할 때 저장된 사용자 페르소나와 말투 정보를 활용하여 보다 적절한 응답을 생성할 수 있도록 한다. 본 연구의 의의는 사용자 페르소나를 명확히 이해하고 반영하며, 개인화된 응답을 생성하는 혁신적인 대화형 프레임워크 및 사용자 경험을 제공하는 데에 있다.

2. 유저 페르소나를 위한 인터랙티브 챗봇 시스템

2.1 연구 목표

본 연구의 목표는 대화형 환경에서 LLM을 활용하여 사용자의 페르소나를 학습하고, 이를 정보 검색 기반 RAG 아키텍처와 결합하여 사용자의 페르소나에 기반한 적절한 응답을 자동으로 생성하는 아키텍처를 설계하는 데 있다. 본 연구는 또한 LLM과 번역 모델을 활용하여 사용자의 말투와 언어적 특징을 추론하고, 이를 기반으로 사용자 고유의 말투를 반영한 응답을 생성하는 방안을 탐구한다. 이러한 접근 방식은 LLM의 학습된 지식과 RAG의 실시간 정보 검색 기능을 통합하여, 보다 자연스럽고 개인화된 대화 경험을 제공하는 AI 시스템을 구현하는 것을 목표로 한다.

2.2 RAG과 대화 데이터 처리 방법론

LLM은 다양한 정보를 내재하나 특정 도메인의 전문 지식에서는 한계가 명확하다. 이를 해결하기 위해 RAG 아키텍처가 도입되어 최신 정보와 도메인 특화된 데이터를 검색하여 보다 문맥에 적합하고 정확한 답변을 제공할 수 있도록 한다. 채팅과 같은 대화 시스템에서는 메시지를 로딩하고 처리하는 과정도 매우 중요하다. 이 과정에서는 사용자의 대화 코퍼스가 형성되어 페르소나를 구성하고, 대화 상대와의 특징를 파악할 수 있드. 또한, 대화 상대와 상황에 따른 사용자의 말투를 추론하는 데에도 활용된다. 이와 같은 처리 과정을 통해 KIRINO 시스템은 보다 정교하고 개인화된 응답을 생성할 수 있게 된다.

2.2.1 페르소나 정의 및 추론

새로운 대화가 추가되었을 때, 수식 (1)과 같이 기존 페르소나($P_{current}$)와 새로운 대화($\{(Q, A)\}$)를 $LLM_{persona}$의 입력으로 제공한다.

$$LLM_{persona} : (P_{current}, Q, A) \rightarrow P_{new} \quad (1)$$

$LLM_{persona}$는 프롬프트 엔지니어링을 통해 대화 내용에서 상대와의 관계, 성격, 관심사, 일정 등과 같은 요소를 추출하여 사용자의 새로운 페르소나(P_{new})를 생성하고 저장한다. 이를 통해 대화 시스템은 지속적으로 사용자의 페르소나를 갱신하여 보다 개인화된 응답을 제공할 수 있게 된다.

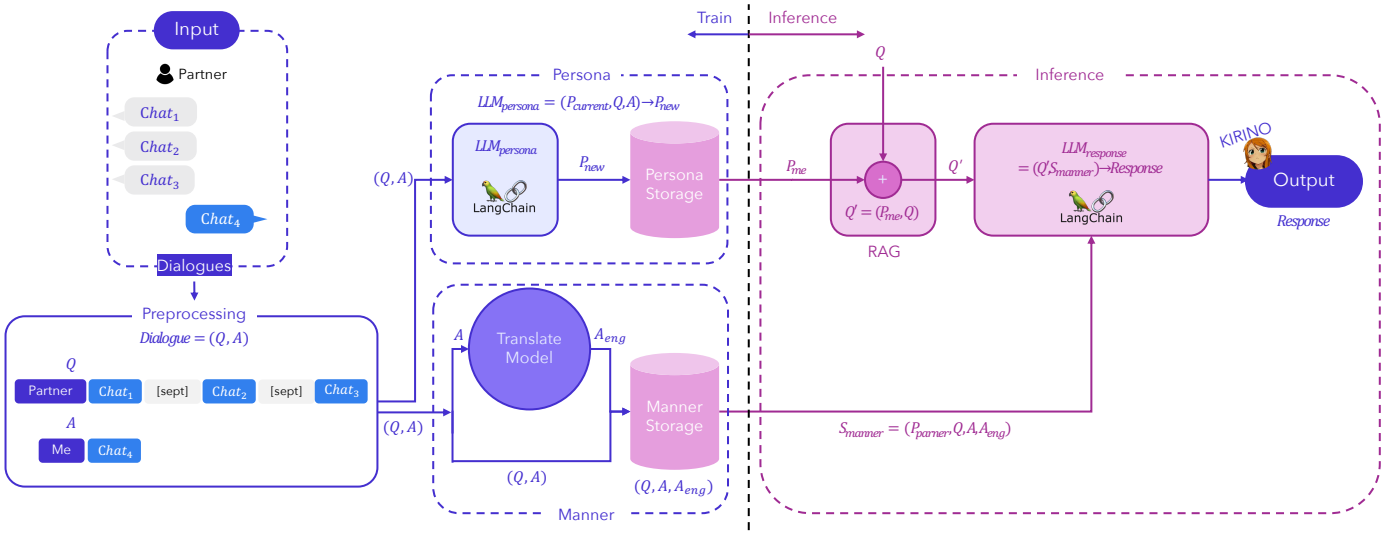

그림 1: KIRINO RAG 아키텍처

2.2.2 사용자 말투 추론

대화를 전처리하면 (상대 페르소나, 질문, 사용자 응답, 영어로 번역된 사용자 응답)로 구성된 집합이 생성된다. 사용자를 대신하여 $LLM_{response}$가 답변을 생성할 때, "{상대 페르소나 $P_{partner}$}를 지닌 상대방의 {질문 Q}에 대한 나의 대답인 {영어로 번역된 나의 응답 A_{eng}}은 {실제 응답 A}이다"의 의미로 RAG가 이루어진다. 이로써 정의되는 집합 S_{manner}는 (2)과 같고 사용자의 적절한 응답 말투를 추론하는 데 사용된다.

$$S_{manner} : \{(P_{partner},\ Q,\ A,\ A_{eng})\} \tag{2}$$

2.2.3 $LLM_{response}$에 의한 사용자 응답 생성

$LLM_{response}$가 상대방의 질문(Q)에 대한 응답을 생성할 때, 저장된 내 페르소나(P_{me})와 말투(S_{manner})를 가져와 RAG를 활용해 사용자를 대신하여 응답을 생성한다. 이는 "{상대 페르소나 $P_{partner}$}를 지닌 상대방의 {질문 Q}에 대해, 나의 대답 {영어로 번역된 응답 A_{eng}}은 {실제 응답 A}일 때, {상대 페르소나 $P_{partner}$}에 기반하여 해당 질문에 대한 적절한 답변은 무엇인가?" 로 엔지니어링되어 사용자 말투로 답변을 생성한다.

$$LLM_{response} : (P_{me},\ S_{manner},\ Q) \rightarrow Response \tag{3}$$

3. 실험

3.1 실험 설계

기본적인 페르소나의 요소로 직업, 성별, 결혼 여부, 나이 등을 정의하였다[5]. 또한, LLM이 생성한 요약이 유창성, 일관성, 사실적 정확성 면에서 기존 자연어 모델과 인간의 요약을 모두 능가했음을 확인하였고[6], 각 대화로부터 추가된 사용자의 특징을 LLM으로 요약하여 페르소나로 업데이트했다. 또한, LLM API와 CO-STAR Framework를 활용하여 다양한 페르소나의 대화 상대와 다양한 상황을 가정하여 대화 코퍼스를 생성하였다.

실험은 Python 환경에서 OpenAI의 GPT-3.5-turbo, GPT-4 모델을 langchain을 활용한 파이프라인을 구축하여 진행했고, FAISS 벡터 스토어를 활용하였다.

3.2 실험 진행

3.2.1 데이터 전처리

대화의 입력은 특정 대화 상대 $P_{Partner}$와 나눈 여러 차례의 대화($Chat_1$, $Chat_2$ 등)로 구성된다. 대화를 질문(Q)과 응답(A)의 형태로 분리한다. 이때, Q은 상대방의 여러 대화를 하나의 질문으로 결합하고, 각 대화 사이에는 구분자[sept]를 추가하여 각 대화를 구별한다. A은 사용자의 대화에 대한 답변을 추출하여 질문에 대한 응답으로 저장하여 일관된 대화 세트로 묶는다.

3.2.2 페르소나 정의 및 추론

새로운 대화가 추가될 때, 기존 페르소나와 새로운 대화(질문 Q와 응답 A)를 LLM 기반 시스템에 입력으로 제공한다. 이 시스템은 대화 내용을 분석하여 상대와의 관계, 성격, 관심사, 일정 등과 같은 요소를 추출하고, 이를 바탕으로 사용자의 새로운 페르소나를 생성하고 저장한다. 데이터 전처리를 통해 각 대화에 페르소나 태그를 부여하고, 다양한 상황을 가정한 대화를 수집하여 코퍼스를 구축한다. 각 대화 내용을 LLM으로 요약하여 기존 페르소나에 반영하며, 새로운 대화가 추가될 때마다 페르소나 추론 모델을 사용하여 정보를 갱신한다. 이를 통해 사용자의 최신 정보를 반영한 개인화된 응답을 지속적으로 제공할 수 있다.

3.2.3 사용자 말투 추론

사용자의 말투를 추론하기 위해, 수집된 대화 데이터를 전처리하고 LLM을 활용하여 자연스러운 응답을 생성하는 시스템을 구축하였다. 먼저, 각 대화 데이터는 상대방과의 질문 및 사용자의 응답으로 구성된다. LLM 기반의 번역 시스템을 이용하여 사용자의 응답을 영어로 번역하고, 존댓말과 같은 특정 톤을 제거하여 중립적인 영어 응답을 생성한다. 이후, 전처리된 데이터를 바탕으로 LLM을 사용하여 최종 응답을 생성한다. 이 과정에서 LLM은 상대방의 페르소나와 질문, 그리고 사용자의 평소 응답 패턴을 고려하여 사용자의 말투가 고려된 응답을 생성한다.

3.2.4 실험 결과

연구에서는 생성된 응답의 품질을 평가하기 위해 G-Eval[7]을 활용해 생성된 응답을 자동으로 평가했다. 대화 상대의 페르소나와 일치하는 자연스러운 응답을 생성하는 **(1) 응답 내용의 적절성**, 사용자의 페르소나와 일치하는 응답을 생성하여 대화의 질을 향상해서 **(2) 페르소나 적합성**, 그리고 대화 흐름을 유지하며 일관성 있는 응답을 생성하는 **(3) 자연스러운 말투**를 평가 지표로 설정했다. GPT에게는 각 지표를 기준으로 1점(매우 나쁨)에서 5점(매우 좋음)까지 점수를 매기도록 80개의 예시를 바탕으로 프롬프트 엔지니어링했다. 또한 생성된 응답의 품질을 평가하기 위해 30명의 평가자가 참여하여 Human Evaluation을 진행하였다. 각 평가자는 G-Eval의 세 가지 지표를 활용하여 동일한 방식으로 점수를 매긴 후 평균을 취하였다. 실험에 대한 평가 결과는 1과 2와 같다.

Criterion	G-Eval	Human Eval
응답 내용의 적절성	5	4.2
페르소나 적합성	2.8	2.4
자연스러운 말투	3.3	2.7

표 1: 페르소나 시스템이 없는 기존 시스템

Criterion	G-Eval	Human Eval
응답 내용의 적절성	5	4.6
페르소나 적합성	4.5	4.0
자연스러운 말투	4.5	4.3

표 2: KIRINO 실험 결과

페르소나 시스템이 반영되지 않은 챗봇 시스템과 KIRINO 시스템을 비교한 결과, G-Eval에서는 26%, Human Eval에서는 38% 향상되었음을 확인했다. 일반적으로 페르소나가 충분한 경우, 가령 '자바 개발자'라는 페르소나를 가진 사람이 '너 자바 할 줄 알아'라는 친구의 질문에 '응. 나 자바 잘 해.'라는 페르소나가 잘 반영됐으며, 이를 반영해 상대방을 고려한 말투의 답변을 잘 생성함을 확인할 수 있었다. 그러나 적절한 답변을 생성하기 위한 배경지식이 페르소나에 없는 경우에는 지표 중 (2)와 (3)은 잘 반영함을 확인하였으나, (1)에서 너무 일반적인 답변을 내놓은 경향을 보여 낮은 점수를 받는 경우가 있었다.

3.3 결론

본 연구에서 사용자의 페르소나와 말투를 학습한 챗봇 시스템인 KIRINO를 설계하였다. 페르소나 추론 및 말투 추론 모듈을 통해 본 모델은 페르소나에 적합하게 응답을 잘 생성하며, 자연스러운 말투를 지닌 챗봇 능력을 보일 수 있었다. 이러한 챗봇은 고객 상담에 활용되거나 웨어러블 기기에 도입되어 운전과 같은 특수상황에서 사용자 대신 답변을 보내는 시스템으로 사용될 수 있음을 시사한다. 대화 시스템의 생성 능력을 유지하면서도 프라이버시 유출 문제를 크게 줄이는 전략을 설계하는 데에 중점[8]을 두고 후속 연구를 진행하고 있다.

참고 문헌

[1] R. Liu, A. Rashid, I. Kobyzev, M. Rezagholizadeh, and P. Poupart, "Attribute controlled dialogue prompting," in *ACL Findings*, 2023.

[2] B. P. Majumder, H. Jhamtani, T. Berg-Kirkpatrick, and J. McAuley, "Like hiking? you probably enjoy nature: Persona-grounded dialog with commonsense expansions," in *EMNLP*, 2020.

[3] H. Kim, B. Kim, and G. Kim, "Will i sound like me? improving persona consistency in dialogues through pragmatic self-consciousness," in *EMNLP*, 2020.

[4] D. Kwon, S. Lee, K. H. Kim, S. Lee, T. Kim, and E. Davis, "What, when, and how to ground: Designing user persona-aware conversational agents for engaging dialogue," in *ACL Industry Track*, 2023.

[5] A. Tigunova, A. Yates, P. Mirza, and G. Weikum, "Listening between the lines: Learning personal attributes from conversations," in *WWW Conference*, 2019.

[6] X. Pu, M. Gao, and X. Wan, "Summarization is (almost) dead," *ArXiv*, vol. abs/2309.09558, 2023.

[7] Y. Liu, D. Iter, Y. Xu, S. Wang, R. Xu, and C. Zhu, "G-eval: NLG evaluation using gpt-4 with better human alignment," ACL, 2023.

[8] H. Li, Y. Song, and L. Fan, "You don't know my favorite color: Preventing dialogue representations from revealing speakers' private personas," ACL, 2022.