

페르소나 기반 한국어 대화 모델링을 위한 데이터셋

이요한[○], 김현, 신중훈, 조민수, 권오욱, 김영길

한국전자통신연구원

carep@etri.re.kr, h.kim@etri.re.kr, jhshin82@etri.re.kr, mscho@etri.re.kr, ohwoog@etri.re.kr, kimyk@etri.re.kr

A Dataset for Persona-based Korean Dialogue Modeling

Yohan Lee[○], Hyun Kim, Jonghun Shin, Minsoo Cho, Ohwoog Kwon, Youngkil Kim
Electronics and Telecommunications Research Institute

요약

페르소나 기반의 대화 시스템은 일관적인 대화를 수행할 수 있어 많은 관심을 받고 있다. 영어권에서 구축된 페르소나 대화 데이터셋은 서로의 페르소나를 알아가기 보다는 자신의 페르소나에 대해서만 말하는 경향을 보이며 이는 상대방의 말을 이해하여 관련 대화를 진행하는 대화의 특성을 반영하지 못한다. 본 연구에서는 회사 방문객이 안내 시스템과 대화하는 상황을 가정하여 안내 시스템이 주도적으로 방문객의 페르소나를 묻고 관련 대화를 수행하는 데이터셋을 구축함과 동시에 목적지향 대화 시스템의 대화 관리 프레임워크를 기반으로 시스템 주도적인 대화를 모델링하는 페르소나 대화 관리 모델을 제안한다. 실험을 통해 제안한 대화 관리 모델의 대화 이해 및 정책 성능을 검증하고 방문객의 페르소나를 예측할 때 대화 정책의 성능이 향상됨을 보임으로써 구축한 데이터셋이 이해와 정책이 포함된 대화의 특성을 반영하는 것을 확인한다.

주제어: 한국어 페르소나 대화, 대화 이해, 대화 정책, 대화 관리

1. 서론

대화 시스템은 정해진 태스크 없이 대화를 수행하는 자유 대화 시스템과 정의된 태스크를 바탕으로 사용자의 목적을 달성하기 위해 대화를 수행하는 목적 지향 대화 시스템으로 구분된다. 자유 대화 시스템은 대화의 범위가 한정되어 있지 않아 다양한 주제에 대해 대화가 가능한 특징을 갖는다. 초기 자유 대화 시스템은 대화자가 자연어 발화를 주고받는 대화 데이터를 바탕으로 대화 이력으로부터 응답을 생성하는 방식으로 구현되었다 [1]. 이러한 대화 시스템은 일관성 없는 응답을 하는 문제를 보였고 이를 해결하기 위해 페르소나에 기반하는 대화가 제안되었다 [2]. 페르소나란 개인의 정보, 상황, 경험 등으로 이루어진 프로필로, 이를 활용하여 대화 시스템에게 인격을 부여할 수 있고 일관된 응답을 유도할 수 있다. 예를 들어, ‘내 직업은 학생이다.’ 라는 페르소나를 대화 시스템에게 부여한다면 직업을 물어보는 상대방의 질문에 대화 시스템은 학습 데이터에 존재하는 여러 가지 직업으로 응답하지 않고 페르소나에 있는 정보를 활용하여 일관성 있는 응답을 할 수 있게 된다.

영어권에서 구축된 페르소나 데이터셋은 두 명의 대화자에게 서로 다른 임의의 페르소나를 부여하고 대화를 진행하는 방식으로 구축되었고, 대화 이력과 주어진 페르소나에 기반하여 응답을 생성하는 대화 시스템이 제안되었다 [2]. 구축된 대화를 살펴보면, 대화자들은 서로의 정보를 묻고 알아가기보다는 자신의 정보를 계속해서 이야기해 나가는 경향을 보인다. 저자의 실험에서 자신의 페르소나를 기반으로 응답을 생성하는 것이 상대방의 페르소나를 포함하여 응답을 생성하는 것보다 높은 성능을 보이는 결과가 이를 뒷받침한다고 볼 수 있다. 이는

상대방의 말을 이해하고 관련된 정보를 활용하여 대화를 수행하는 사람의 대화 특성을 반영하지 못한다.

본 연구에서는 주어진 페르소나를 바탕으로 서로를 알아가는 한국어 대화 데이터셋을 구축한다. 대화자들이 서로의 페르소나에 대해 알지 못하는 것은 동일하지만 보다 현실적인 대화를 위해 자신의 페르소나를 위주로 말하기보다는 상대방의 페르소나에 대해 묻고 대화를 이끌어 가는 것에 목적을 둔다. 중요한 점은 단순히 페르소나만을 묻는 것이 아니고 대화를 이끌어 간다는 것이다. 대화를 이끌어 가기 위해서는 상대방의 말을 이해한 것을 바탕으로 앞으로 어떻게 대화를 수행할 지 결정해야 한다. 따라서 본 연구는 사용자의 발화를 이해하여 대화 상태를 추적하고 이를 바탕으로 사용자의 목적을 달성하기 위한 대화 정책을 결정하는 것을 포함하는 목적 지향 대화 시스템의 대화 관리 프레임워크를 활용한다.

위 목적을 달성하기 위해 방문객이 회사를 방문하는 상황을 가정하여 안내 시스템이 방문객의 페르소나를 묻고 관련 대화를 주도하는 형식의 대화 4,300개를 구축하였다. 모든 대화는 방문객 역할자가 실제 시스템과 대화한다는 설정 하에 사람이 시스템 역할을 대신하는 wizard-of-oz 방식으로 수집되었으며 방문객과 안내 시스템의 각 발화에 대해 관련된 페르소나 정보를 부착하였다. 이를 활용하여 방문객의 발화를 이해하고 대화를 이끌어 가기 위한 페르소나 대화 모델링 방법을 제안하고, 구축한 대화가 상대방의 말을 이해하고 관련된 정보를 활용하는 특성을 가지고 있음을 실험적으로 확인한다.

2. 관련 연구

페르소나 기반 한국어 대화 데이터셋은 기계 번역을 통해 영어로 구축된 페르소나 데이터셋 [2]을 번역하여 구축되었다 [3]. 이를 기반으로 대화 이력 및 페르소나와 응답 후보 중 적합한 응답을 선택하는 방식의 페르소나 기반 한국어 대화 모델이 제안되었다 [4]. 본 연구에서 제안하는 페르소나 기반 한국어 대화 데이터셋은 응답 후보를 포함하고 있지 않지만 각 발화에 부착된 페르소나 정보를 통해 응답 후보 샘플링이 가능하므로 대화 이력 및 페르소나를 기반으로 응답 후보 중 적합한 응답을 선택하는 태스크 역시 수행할 수 있다.

시스템 주도적인 대화를 위해서는 교육을 위한 한국어 목적지향 대화 데이터셋 [5]이 제안되었고, AI-Hub에서는 콜센터, 심리 상담 대화 데이터셋 등을 제공하고 있다. 최근에는 대화 상태 정보가 부착된 심리 상담 대화 데이터셋을 활용하여 대화 이력으로부터 대화 상태를 추적하기 위한 모델링 연구가 진행되었다 [6].

3. 한국어 페르소나 기반 대화 모델링

3.1 페르소나 기반 한국어 대화 데이터셋

본 연구에서는 방문객이 회사를 방문했을 때 안내 시스템과 대화하는 상황을 가정하여 시스템이 페르소나에 관련한 대화를 주도하는 데이터셋을 제안한다. 페르소나는 날씨, 시간 등 5개의 환경 정보 카테고리를 포함하는 환경 프로파일, 방문객의 성별, 마스크 착용 여부 등 13개의 시각 정보 카테고리를 포함하는 시각 프로파일, 방문객 및 안내 시스템의 개인 프로파일로 구성된다. 환경 프로파일과 시각 프로파일은 방문객과 안내 시스템이 공유하는 정보이며 개인 프로파일은 서로 알지 못하고 대화를 통해 알아가야 하는 정보이다. 안내 시스템의 개인 프로파일은 안내 시스템에게 물어보고 싶은 내용에 대해 회사 방문객을 대상으로 설문 조사를 시행하여 그 결과를 바탕으로 이름, 나이와 같은 35개의 프로파일을 선정하여 구성하였다. 방문객의 개인 프로파일은 방문, 직업, 여행 등 6개의 주제와 각 주제에 해당하는 프로파일을 포함하여 총 25개의 프로파일로 구성하였다. 구체적인 페르소나의 구성은 부록에서 확인할 수 있다.

표 1. 데이터 통계

	시스템 발화 분포 (%)	방문객 발화 분포 (%)
대화 수	4,300	
발화 수	113,430	
평균 발화 수	26.4	
자유 대화	45.5	34.1
환경 프로파일	5.6	3.0
시각 프로파일	2.7	1.8
방문객 프로파일	42.0	58.7
시스템 프로파일	4.2	2.4

데이터 구축 작업자는 각각 안내 시스템과 방문객 역할

을 맡아 대화를 수행한다. 시스템이 주도하는 대화를 수집하기 위해 안내 시스템 역할자에게는 대화를 통해 알아야 할 방문객 프로파일 목록 (미션 프로파일)을 주어 자연스럽게 대화를 진행하되 미션 프로파일에 대해서는 반드시 언급하도록 하였다. 미션 프로파일은 2개 또는 3개의 주제를 포함하도록 하였다. 방문객 역할자 또한 안내 시스템의 프로파일에 대해 물어볼 수 있으나 알아야 할 목록을 제한하지 않았으며 두 대화자 모두 대화를 진행하면서 환경 및 시각 프로파일을 자연스럽게 언급하거나 정해진 프로파일 외의 자유 대화를 수행하도록 하였다. 안내 시스템 역할자는 안내 시스템 프로파일을 숙지한 숙련된 작업자 2명이 담당하고 방문객 역할자는 50명 이상의 작업자가 담당하여 다양한 회사 방문객이 안내 시스템과 대화하는 실제 상황을 모방하였다. 대화를 수집한 후 방문객과 안내 시스템의 각 발화에 대해 문장 단위로 관련된 프로파일 정보를 부착하였고 관련된 프로파일 없다면 자유 대화로 표시하였다. 표 1의 통계 정보에서 시스템이 자신의 프로파일보다는 방문객의 프로파일에 대해 더 많이 언급하는 것을 알 수 있다. 본 연구를 통해 구축한 데이터셋의 샘플은 부록에서 확인할 수 있다.

3.2 페르소나 기반 대화 관리 모델

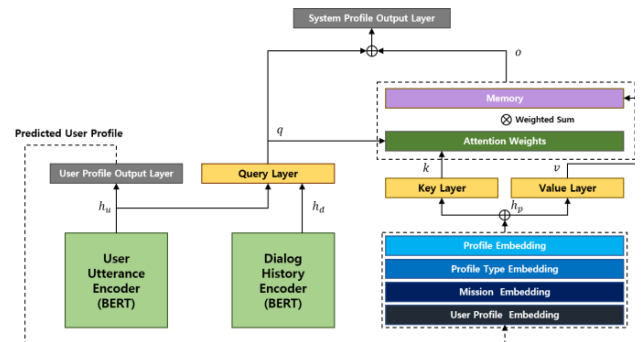


그림 1. 페르소나 기반 대화 관리 모델

대화를 주도하는 안내 시스템은 방문객의 발화를 이해하고 앞으로의 대화 방향, 즉 대화 정책을 결정해야 한다. 따라서 본 연구에서는 프로파일 정보와 대화 이력을 바탕으로 다음 언급할 프로파일을 결정하기 위해 Memory Network [7] 구조의 페르소나 대화 관리 모델을 그림 1과 같이 제안한다. 자유 대화 유형을 포함하는 각 프로파일 $p^i (i = 1, \dots, N)$ 에 대한 특징 임베딩 벡터는 다음과 같이 정의된다.

- Profile Embedding (e_x^i): i 번 째 프로파일의 갖는 고유의 임베딩 벡터
- Profile Type Embedding (e_t^i): i 번 째 프로파일의 유형 (자유 대화, 환경, 시각, 방문객, 시스템)에 대한 임베딩 벡터
- Mission Embedding (e_m^i): i 번 째 프로파일의 미션 및 활용 여부를 나타내는 임베딩 벡터

- User Profile Embedding (e_u^i): i 번 째 프로필의 방문객 언급 여부를 나타내는 임베딩 벡터

프로필 특징 임베딩 벡터에서 e_u^i 는 방문객 발화가 어떤 프로필을 언급하고 있는지 예측하는 것으로 자연어 이해에 해당하며 미션 및 활용 여부를 나타내는 e_m^i 는 대화 상태 추적에 해당한다고 볼 수 있다. 위 특징 임베딩 벡터를 통해 각 프로필에 대한 상태 벡터 h_p^i 를 구한다.

$$h_p^i = e_x^i + e_t^i + e_m^i + e_u^i \quad (1)$$

방문객의 발화와 이전 대화 이력은 BERT [8]를 통해 각 CLS 토큰의 상태 벡터 h_u, h_d 를 구하고 프로필 상태 벡터와의 주의 집중 메커니즘을 통해 대화와 프로필 간의 연관성을 모델링한다.

$$q = \text{query_layer}([h_u; h_d]) \quad (2)$$

$$k^i = \text{key_layer}(h_p^i), v^i = \text{value_layer}(h_p^i) \quad (3)$$

$$a^i = \text{softmax}(q \cdot k^i) \quad (4)$$

$$o = \sum_i a^i * v^i \quad (5)$$

이를 바탕으로 앞으로 발화할 프로필 확률 p_{sys}^i 을 출력하고 동시에 방문객 발화 상태 벡터 h_u 를 바탕으로 방문객이 발화한 프로필 확률 p_{usr}^i 에 대해 출력한다. 모델 학습을 위한 손실 함수는 추측한 각 프로필 확률과 정답 프로필 y_{sys}^i, y_{usr}^i 간의 교차 엔트로피 (Cross Entropy, CE)로 정의한다.

$$L = \frac{1}{N} (\sum_i CE(p_{sys}^i, y_{sys}^i) + CE(p_{usr}^i, y_{usr}^i)) \quad (6)$$

4. 실험

3.2절에서 설계한 대화 관리 모델을 기반으로 ablation 실험을 수행하여 구축한 데이터셋이 상대방의 말을 이해하고 응답하는 대화의 특성을 반영하였는지 확인한다. 실험을 위해 구축한 데이터셋에서 임의의 10%를 추출하여 430 대화를 평가 데이터로, 나머지 3,870 대화를 학습 데이터로 활용하였다.

방문객 발화와 대화 이력을 인코딩하는 BERT는 KLUE-BERT [9]를 활용하고 대화 이력은 이전 3턴까지의 시스템과 방문객 발화로 구성하였다. 모든 은닉 층의 크기는 BERT와 동일한 크기인 768로 설정하였고 32의 배치 사이즈로 총 10회 학습하였다. 옵티마이저로는 AdamW [10]을 활용하고 초기 학습률을 $1e-4$ 으로 설정하여 학습을 마칠 때까지 선형 감소하였다.

추론 시에는 방문객의 발화 프로필을 먼저 예측한 후 이를 활용하여 User Profile Embedding을 계산하고 최종 시스템 발화 프로필을 예측한다. 최종 예측된 프로필로는 프로필에 대한 예측 확률이 임계 값을 넘는 프로필을

선택하였다. 따라서 다수의 프로필이 한 번에 선택될 수 있다.

$$\hat{p} = \mathbb{I}[p^i > \tau] \quad (7)$$

방문객과 시스템 프로필 예측에 대한 임계 값 τ 는 모두 0.5로 설정하였다.

시스템의 성능은 예측한 프로필과 정답 프로필 간의 precision, recall, F1을 대화 단위로 측정한다.

표 2. 프로필 예측 성능 비교

Model	시스템 프로필			방문객 프로필		
	Prec.	Rec.	F1	Prec.	Rec.	F1
	72.70	48.6	58.26	88.53	89.66	89.09
- Pred. user profile	71.06	48.77	57.84	-	-	-

표 2는 제안 모델과 제안 모델에서 방문객 프로필 예측에 대한 손실 함수와 특징을 제거하였을 때의 시스템 프로필 예측 성능을 나타낸다. 제안 모델은 방문객이 발화한 프로필에 대해 비교적 잘 예측하고 있고 이를 제거하게 되는 경우 시스템 프로필 예측, 즉 대화 정책 성능이 F1기준 약 0.4 감소하는 것을 확인할 수 있다.

5. 결론

본 연구에서는 상대방의 페르소나를 알기 위해 시스템이 대화를 주도하는 페르소나 기반 한국어 대화 데이터셋을 설계하여 구축하였고 사용자의 발화를 이해하여 대화 상태를 추적하고 이를 바탕으로 대화 정책을 결정하는 것을 포함하는 목적 지향 대화 시스템의 대화 관리 프레임워크를 기반으로 페르소나 기반 대화 모델링 방법을 제안하였다. 또한 대화의 이해에 해당하는 특징을 제거하는 실험을 통해 대화 정책 성능이 감소함을 보여 상대방의 말을 이해하고 응답하는 대화의 특성이 구축한 데이터셋에 잘 반영되었다는 것을 확인하였다.

본 연구의 페르소나 데이터셋은 한정적인 페르소나를 바탕으로 구축되었기 때문에 다양한 페르소나를 가질 수 있는 자유 대화의 특성을 반영하지 못하는 한계점을 갖는다. 따라서 향후에는 페르소나의 범위를 넓히고 학습 페르소나와 평가 페르소나를 분리하여 학습되지 않은 페르소나로도 대화가 가능한 모델에 대한 연구가 필요할 것이다.

감사의 글

이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (2019-0-00004, 준지도학습형 언어지능 원천기술 및 이에 기반한 외국인 지원용 한국어 튜터링 서비스 개발)

참고문헌

[1] Y. Li, H. Su, X. Shen, W. Li, Z. Cao & S. Niu, "DailyDialog: A Manually Labelled Multi-turn Dialogue Dataset", In Proceedings of the Eighth International Joint Conference on Natural Language Processing, vol.1, pp.986-995, 2017.

[2] S. Zhang, E. Dinan, J. Urbanek, A. Szlam, D. Kiela & 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, vol.1, pp.2204-2213, 2018.

[3] Z. Lin, Z. Liu, G. I. Winata, S. Cahyawaijaya, A. Madotto, Y. Bang, E. Ishii & P. Fung, "XPersona: Evaluating Multilingual Personalized Chatbot", In Proceedings of the 3rd Workshop on Natural Language Processing for Conversational AI, pp.102-112, 2021.

[4] 장윤나, 임정우, 허윤아, 양기수, 박찬준, 서재형, 이승준 & 임희석, "페르소나 기반 한국어 대화 모델", 제33회 한글 및 한국어 정보처리 학술발표 논문집, pp.625-628, 2021.

[5] 최승권, 이요한 & 권오욱, "외국어 교육용 대화 시스템을 위한 목적지향 대화 데이터 구축 연구: 한국어 대화 데이터를 중심으로", Foreign Language Education, vol.29, no.1, pp.105-124, 2022.

[6] 장윤정 & 이근배, "상담 대화를 위한 사용자 정보 추출 대화 상태 추적 시스템", 제33회 한글 및 한국어 정보처리 학술발표 논문집, pp.453-456, 2021.

[7] S. Sukhbaatar, A. Szlam, J. Weston & R. Fergus, "End-To-End Memory Network", arXiv preprint arXiv:1503.08895, 2015.

[8] J. Devlin, M.-W. Chang, K. Lee & K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", In Proceedings of the 2019 conference of the North-American Chapter of the Association for Computational Linguistics: Human Language Technologies, vol.1, pp.4171-4186, 2019.

[9] S. Park, J. Moon, S. Kim, W. I. Cho, J. Han, J. Park, C. Song, J. Kim, Y. Song, T. Oh, J. Lee, J. Oh, S. Lyu, Y. Jeong, I. Lee, S. Seo, D. Lee, H. Kim, M. Lee, S. Jang, S. Do, S. Kim, K. Lim, J. Lee, K. Park, J. Shin, S. Kim, L. Park, A. Oh, J.-W. Ha & K. Cho, "KLUE: Korean Language Understanding Evaluation", arXiv preprint arXiv:2105.09680, 2021.

[10] I. Loshchilov and F. Huter, "Decoupled weight decay regularization", arXiv preprint arXiv:1711.05101, 2017.

부록

A. 페르소나 구성

프로필 유형	프로필 카테고리
환경 프로필 (EP)	기온
	날씨
	미세먼지
	요일
시각 프로필 (VP)	시간
	마스크 착용 여부
	성별
	연령
	헤어 밝기
	헤어 길이
	헤어 스타일
	목걸이 착용 여부
	귀걸이 착용 여부
	안경 착용 여부
시스템 프로필 (SP)	상의 색상
	상의 패턴
	상의 종류
	상의 길이
	이름
	성별
	가족관계
	직업
	거주지
	취미
	부모님
	생일
	친구관계
	좋아하는 영화 장르
	감명 깊게 본 영화
	좋아하는 음악 장르
	좋아하는 연예인
	좋아하는 드라마
	좋아하는 음식
	취미 생활
	출근 시간
	퇴근 시간
	IQ
	체중
	싫어하는 것
좋아하는 것	
가고 싶은 곳	
잘하는 일	
나이	
꿈	
관심사	
할 수 있는 외국어	
MBTI	
운동 여부	
관심 연구 분야	
업무 만족도	

제34회 한글 및 한국어 정보처리 학술대회 논문집 (2022년)

프로필 유형	프로필 카테고리	
시스템 프로필 (SP)	스트레스	
	직장 동료 관계	
	여행가본 적	
방문객 프로필 (UP)	방문 목적	
	방문 목적 달성 여부	
	재방문 여부	
	관심 연구분야	
	식사 여부	
	커피 선호	
	직장	직업
		직업 만족도
		스트레스
	건강	직장 동료 관계
		운동 여부
		운동 종류
		운동 빈도
		운동 파트너
		건강 유지 방법
	반려 동물	반려 동물 키우는 지
		반려 동물 종류
		반려 동물 관련 걱정
	음식	선호 음식
		자주 먹는지
		요리 선호 여부
	여행	여행 빈도
		기억에 남는 여행지
		가보고 싶은 여행지
여행 파트너		

B. 샘플데이터 (G=자유 대화)

발화자	발화	프로필 정보
시스템	환영합니다, 방문자님. 오늘 춥다고 하던데 오는 길은 편찮았나요?	G, EP(temperature)
방문객	코트를 걸쳤는데도 좀 춥네.	VP(upper_type), EP(temperature)
시스템	방문자님은 어떤 프로젝트때문에 오셨나요?	UP(research)
방문객	반도체 연구를 하려고 왔어.	UP(research)
시스템	그럼 직업이 연구원인가요?	UP(job)
방문객	아니, 나는 대학원생이야.	UP(job)
시스템	대학원은 어떤가요?	UP(job_satisfaction)
방문객	별로야. 전혀 만족스럽지 않아. 너는 네 일이 만족스러워?	UP(job_satisfaction), SP(job_satisfaction)
시스템	네, 저는 제 일이 좋아요. 방문자님은 일이 많이 힘든신가봐요.	SP(job_satisfaction)
방문객	대학원이 힘들다기보단 동료들과 사이가 좀 안좋아.	UP(job_relation)
시스템	저런, 힘드시겠어요. 반려동물은 키우시나요?	G, UP(pet)
방문객	아니, 키우지 않아.	UP(pet)
시스템	반려동물을 키우면 스트레스가 줄어든대요. 키워보시는 게 어때요?	G
방문객	좋은 생각같아. 전부터 강아지를 키워보고 싶었어.	G, UP(pet_type)
시스템	도움이 됐다니 기쁘네요. 오늘 금요일인데 뭔가 계획이 있나요?	G, EP(day)
방문객	집에서 드라마를 볼까 해. 재밌게 본 드라마가 있니?	G, SP(drama)
시스템	저는 오징어 게임을 재밌게 봤어요.	SP(drama)
방문객	나도 오늘 그걸 한 번 봐야겠어.	G
시스템	좋아요. 저는 이만 가볼게요. 프로젝트 잘 마무리하시길 바라요.	G