

멀티턴 대화를 활용한 레퍼런스 기반의 발화 생성 모델*

박상민^{O*}, 손유리^{*}, 금빛나^{**}, 김홍진^{**}, 김학수^{**}, 김재은^{***}

{주)솔트룩스, AI Labs^{*}, 건국대학교 인공지능학과^{**}

{sangmin.park, yuri.son}@saltlux.com, {beausty23, jin3430, nlpdrkim}@konkuk.ac.kr, jaieun.kim@saltlux.com

Reference-based Utterance Generation Model using Multi-turn Dialogue

Sangmin Park^{*}, Yuri Son^{*}, Bitna Keum^{**}, Hongjin Kim^{**}, Harksoo Kim^{**}, Jaieun Kim^{***}

AI Labs, Saltlux Inc^{*}, Department of Artificial Intelligence, Konkuk University^{**}

요약

디지털 휴먼, 민원 상담, ARS 등 챗봇의 활용과 수요가 증가함에 따라 챗봇의 성능 향상을 위한 다양한 연구가 진행되고 있다. 특히, 오토 인코더(Auto-encoder) 기반의 생성 모델(Generative Model)은 높은 성능을 보이며 지속적인 연구가 이루어지고 있으나, 이전 대화들에 대한 충분한 문맥 정보의 반영이 어렵고 문법적으로 부적절한 답변을 생성하는 문제가 있다. 이를 개선하기 위해 검색 기반의 생성 모델과 관련된 연구가 진행되고 있으나, 현재 시점의 문장이 유사해도 이전 문장들에 따라 의도와 답변이 달라지는 멀티턴 대화 특징을 반영하여 대화를 검색하는 연구가 부족하다. 본 논문에서는 이와 같은 멀티턴 대화의 특징이 고려된 검색 방법을 제안하고 검색된 레퍼런스(준정답 문장)를 멀티턴 대화와 함께 생성 모델의 입력으로 활용하여 학습시키는 방안을 제안한다. 제안 방안으로 학습된 발화 생성 모델은 기존 모델과 비교 평가를 수행하며 Rouge-1 스코어에서 13.11점, Rouge-2 스코어에서 10.09점 Rouge-L 스코어에서 13.2점 향상된 성능을 보였고 이를 통해 제안 방안의 우수성을 입증하였다.

주제어: 생성 모델, 학습데이터 자동 생성, 멀티턴, 적응형 학습

1. 서론

디지털 휴먼, 민원 상담, ARS 등 챗봇 기반 서비스의 증가에 따라 챗봇에 대한 수요가 증가하고 있으며, 챗봇 성능 향상을 위한 다양한 연구가 진행되고 있다. 특히, 오토 인코더(Auto-encoder) 모델을 활용한 생성 모델(Generative Model)이 학계에서 높은 성능을 보이며 지속적인 연구가 이루어지고 있으나 이전 대화들에 대한 충분한 문맥 정보의 반영이 어렵고 디코더(Decoder) 기반의 발화 생성은 문법적, 의미적으로 부적절한 답변을 생성하는 문제가 있다[1,2,3].

최근, 이와 같은 생성 모델의 한계점을 극복하고자 대화 검색 기술을 활용하여 검색한 준정답 문장을 생성 모델에 추가 입력으로 주는 연구가 진행되었고[4], 직전 발화와 예상 답변 벡터를 자동으로 생성하여 유사대화와의 그에 대한 답변을 검색하는 연구가 진행되었다[5].

하지만 멀티턴으로 이루어진 대화는 현재 시점의 문장이 유사해도 이전 문장에 따라 의도와 답변이 달라지는 특징이 있으며 이와 같은 특징을 고려했을 때, 싱글턴 기반의 검색은 명확한 한계가 존재한다.

본 논문에서는 이와 같은 멀티턴 대화의 특징을 반영한 멀티턴 대화 검색 방법을 제안하고 검색된 준정답 문장(레퍼런스)을 멀티턴 대화와 함께 생성 모델의 입력으로

로 활용하는 방안을 제안한다. 또한, 제안한 모델과 기존 모델 간의 비교 평가 및 분석을 통해 제안하는 방법의 우수성을 입증한다.

2. 제안 방안

멀티턴으로 이루어진 대화는 현재 문장이 서로 유사해도 이전 문장들에 따라 의도와 답변이 달라지는 특징을 가지고 있다. 표 1은 이와 같은 멀티턴 대화의 특징에 대한 예시를 나타낸다. 타입 A와 B의 t 시점 문장은 모두 슬프다는 의도를 갖는 동일한 발화이다. 하지만 t-1, t-2 시점에 대한 문장을 함께 보면, 타입 A의 경우는 영화가 슬펐다는 의도를 갖는 발화이며 B의 경우는 시험을 망쳐서 슬펐다는 의도를 갖는 발화이다. 또한, A와 B에 대한 답변 문장 또한 달라지는 것을 볼 수 있다. 이처럼 멀티턴으로 이루어진 대화는 이전 문장들에 대한 문맥이 함께 중요하다는 특징이 있으며, 본 논문에서는 이와 같은 특징을 반영한 검색 방법을 바탕으로 레퍼런스 기반 발화 생성 모델을 학습 방안을 제안한다.

2.1 멀티턴 대화 데이터 구축 방안

레퍼런스 기반 발화 생성 모델의 학습데이터를 구축하기 위해서는 사전에 ‘멀티턴 대화-답변 문장(U-R)’ 쌍에 대한 학습데이터 구축이 필요하다. 본 논문에서는 ‘U-R’ 쌍 데이터를 구축하기 위해 ‘모두의 말뭉치-온라인 대화 말뭉치 2021’ [6]과 사내 보유 중인 대화 말

* 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2013-2-00109, (엑소브레인-2세부) WiseKB: 빅데이터 이해 기반 자가학습형 지식베이스 및 추론 기술 개발)

** 교신저자 : 김재은

표 1 멀티턴 문맥에 따른 마지막 발화의 의미

타입	시점	발화
A	t-2	지금 만나러 갑니다 보고 왔어
	t-1	진짜 그거 슬픈데 어땠어?
	t	너무 슬퍼 ㅠㅠ
	t+1	그치.. 정말 명작이야
B	t-2	오늘 시험 망했어..
	t-1	아 진짜...? 괜찮아 ??
	t	너무 슬퍼 ㅠㅠ
	t+1	힘내.. 다음엔 잘 볼 수 있을 거야!

문치를 활용한다. 그림 1은 'U-R' 쌍 학습데이터 구축 방안을 나타낸다. 각 대화 말문치는 스피커 식별자 (Speaker Id), 스피커의 발화(Form)를 통해 발화하는 스피커 정보를 알 수 있도록 구성되어있으며 스피커 정보가 바뀌기 전까지를 하나의 발화로 정의하고 총 N개의 발화를 멀티턴 대화(U)로 구성한다. 또한, 멀티턴 대화에 대한 응답은(R) N+1번째 발화로 구축한다.

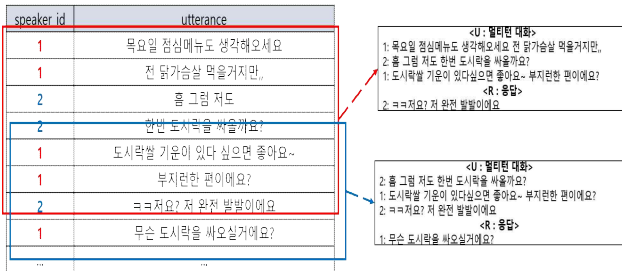


그림 1 'U-R' 쌍 데이터 구축 방안

2.2 레퍼런스 기반 발화 생성 학습데이터 자동 구축방안

레퍼런스 기반의 발화 생성 모델은 멀티턴 대화와 준정답 문장(레퍼런스) 입력으로 함께 활용하여 발화를 생성하는 모델이다. 레퍼런스 기반의 발화 생성 모델 학습을 위해 '멀티턴 대화, 레퍼런스-답변 문장(UR*-R)' 쌍으로 구성된 학습데이터가 필요하지만, 공개된 온라인상 공개된 학습데이터가 없다.

본 논문에서는 이와 같은 학습데이터의 부재에 대한 문제 해결을 위해 'UR*-R' 쌍 학습데이터를 SBERT-IQ 모델[7]을 활용하여 자동으로 구축하는 방안을 제안한다. 제안 방안에서 정의하는 레퍼런스(R*)는 답변 문장(R)과 문맥적, 구조적으로 유사한 문장이며 다음과 같은 방안을 제안 및 활용한다. 첫 번째, R과 문맥적으로 유사한 후보 발화문을 찾기 위해 SBERT-IQ 모델을 활용하여 코사인 유사도가 임계값 이상인 상위 1개의 후보 문장을 찾아온다. 임계값 설정을 위해 총 3명의 연구자가 SBERT-IQ 모델을 활용하여 각 문장의 상위 1개 후보 문장들을 추출하고 이를 분석하였다. 분석 결과 0.85 이상의 유사도를 갖는 문장 쌍들은 문맥적으로 아주 유사하다는 결론을 내렸으며, 이를 바탕으로 임계값을 0.85로 설정하였다.

두 번째, R과 구조적으로 유사한지 문장 확인하기 위해 두 문장 간의 길이를 비교한다. 세 번째, 후보 문장이 R과 문장 타입(평서문, 의문문)이 같은지 확인하기 위해 '?'를 활용하여 문장 타입을 비교한다. 최종적으로, 세 가지 단계를 모두 통과한 후보 문장은 R* 데이터로 활용되며 이를 통해 'UR*-R' 쌍 학습데이터를 구축한다.

2.3 레퍼런스 기반의 발화 생성 입력 정의

레퍼런스 기반의 발화 생성 모델을 학습하기 위해 본 논문에서는 표 2와 같은 입력 포맷에 대해 정의한다. '<user>', '<agent>'라는 스페셜 토큰을 추가하여 스피커 식별자의 전환을 구분하였고, 레퍼런스에 대한 입력이라는 것을 모델에게 알려주기 위해 '<refer>'라는 스페셜 토큰을 사용하였다. 또한, 레퍼런스가 없는 일반 학습데이터일 경우 레퍼런스가 입력되어야 할 위치에 '<empty>'라는 스페셜 토큰을 활용하였다. 이와 같은 입력 정의를 통해 본 논문에서는 KoBART 모델 기반의 발화 생성 모델 학습을 수행한다.

표 2 레퍼런스 기반의 대화 생성 입력 정의 예시

타입		예시
U-R	input	<user> 안녕 반가워! <agent> 응 안녕 만나서 반가워 <user> 오늘 점심 뭐먹었어? <refer> <empty>
	output	나는 김치찌개 먹었엉 ㅋㅋ
UR*-R	input	<user> 안녕 반가워! <agent> 응 안녕 만나서 반가워 <user> 오늘 점심 뭐먹었어? <refer> 응, 나는 김치찌개 먹었어!
	output	나는 김치찌개 먹었엉 ㅋㅋ

2.4 레퍼런스 기반 발화 생성 모델 추론을 위한 멀티턴 기반의 레퍼런스 검색 방안

그림 2는 레퍼런스 기반의 발화 생성 모델이 실제 추론하는 시점에 멀티턴 특성을 고려하여 레퍼런스를 검색하는 방안에 대해 나타낸다. 검색 방안은 다음과 같다. 첫 번째, 보유하고 있는 멀티턴 대화의 각 문장을 SBERT-IQ 모델을 활용하여 임베딩을 수행하고 문장에 대한 임베딩 벡터를 결합하여 색인한다. 두 번째, 입력으로 들어온 멀티턴 발화의 각 문장을 동일한 방법으로 임베딩하고 결합하여 멀티턴 대화가 색인된 데이터베이스에 유사한 벡터 검색을 수행한다. 세 번째, 유사한 벡터를 갖는 멀티턴 대화가 검색되면 해당 멀티턴 대화에 대한 답변 문장을 레퍼런스로 선정 및 활용한다.

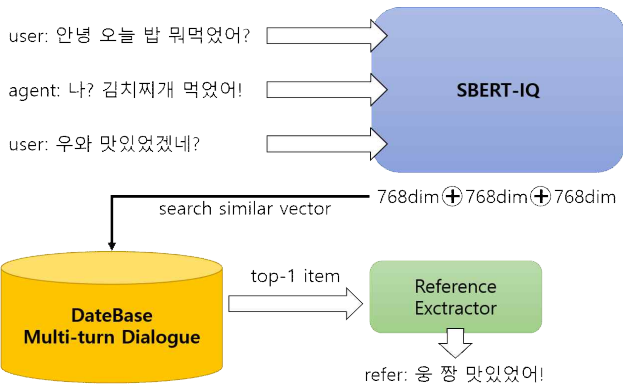


그림 2 멀티턴 기반의 레퍼런스 검색 방안

3. 실험 및 평가

3.1 학습데이터 통계

표 3은 제안한 방안을 통해 자동으로 구축한 학습데이터 통계를 나타낸다. 이와 같은 학습데이터는 레퍼런스 기반의 발화 생성 모델의 학습에 활용되며 KoBART[8] 모델을 기반으로 3 에폭, 16 배치, AdamW의 최적화 기법과 $5e-5$ 의 학습률 적용하여 학습을 수행한다.

표 3 학습데이터 통계

	U-R	UR*-R	전체
학습	664,279	86,629	750,908
테스트	-	1,000	1,000
전체	664,279	87,629	751,908

3.2 발화 생성 모델 간 성능 평가

표 4는 제안 모델에 대한 정량적인 성능 평가를 나타낸다. Baseline 모델은 전체 데이터 세트를 'U-R'로만 구성하여 학습한 KoBART 기반의 발화 생성 모델이며, ours 모델은 'U-R'과 'UR*-R' 학습데이터를 활용한 KoBART 기반의 발화 생성 모델이다. 성능 평가는 Rouge-N 스코어[9]를 활용하였으며, 제안한 모델이 Rouge-1 점수에서 15점, Rouge-2 점수에서 11점, Rouge-L 점수에서 15점 이상의 성능 향상을 보였다. 이를 통해 본 제안 방안의 우수성을 입증하였다.

표 4 모델 성능 비교 평가

모델	Rouge-1	Rouge-2	Rouge-L
baseline	32.29	22.92	32.09
ours	45.40	33.01	45.29

3.3 검색 방법에 따른 비교 평가

표 5는 본 논문에서 제안한 검색 방법과 싱글턴 기반 검색 방법의 검색 결과 예시이다. 검색 결과에 따르면, 제안 방안의 검색 결과와 싱글턴 기반의 검색 결과는 쿼리의 t 시점 문장에 대해 모두 유사한 발화를 가져오는 것을 볼 수 있다. 하지만, 쿼리의 이전 문장들을 모두 비교할 경우 싱글턴 기반의 검색 방법은 문맥이 맞지 않는 결과를 볼 수 있으며, 제안한 검색 방법은 적절한 문맥을 찾아오는 결과를 볼 수 있었다.

이를 통해, 제안한 검색 방법이 멀티턴 대화의 특성을 고려하여 더 적절한 유사대화를 검색해온다는 점을 입증하였다.

표 5 모델 성능 비교 평가

입력	t-2	t-1	t
쿼리	name1님은요?	전 기혼이고 신혼입니다ㅋㅋ	아 신혼이시군요 결혼하신지 얼마나 되셨나요?
baseline	배우자분이랑은 어떻게 만나셨어요?	친구소개요	결혼한지 얼마나 되셨어요?
ours	애인 있으세요? 전 없는데 흑흑	ㅎㅎ저는 신혼이에요	어머 너무 좋으시겠어요 결혼하신지 얼마나 되셨나요?

4. 결론 및 향후 연구

본 논문에서는 멀티턴 대화를 활용한 검색 기반(레퍼런스 기반)의 발화 생성 모델 학습을 위해 레퍼런스 기반 학습데이터 자동 구축방안에 대해 제안하고, 멀티턴 대화의 특징이 고려된 검색 방법을 제안하였다. 발화 생성 모델은 자동 구축한 레퍼런스 기반의 학습데이터 활용하여 학습이 수행되었으며 제안한 검색 방법을 통해 발화 생성을 수행하였다. 제안한 모델은 기존 모델과의 Rouge-N 스코어 기반의 성능 평가에서 약 11~15점 이상 향상된 결과를 보였으며, 제안한 검색 방법의 정성적인 평가를 통해 제안 방안의 우수성을 입증하였다.

향후 연구로는 발화 생성 모델에 추가 입력으로 페르소나 또는 어투 등과 같은 정보 입력하고 이를 바탕으로 적절한 발화를 생성하는 연구를 진행할 예정이며, 레퍼런스에 대한 의존도를 낮추어 발화 생성 모델의 결과에 대한 자유도는 높여 안정적인 발화를 생성할 수 있는 연구를 진행할 예정이다.

참고문헌

- [1] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, L. Zettlemoyer, “BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension”, arXiv preprint arXiv:1910.13461, 2019.
- [2] A. Redford, J. Wu, R. Child, D. Luan, D. Amodei, H. Sutskever, “Language Models are Unsupervised Multitask Learners”, OpenAI blog, 1.8:9, 2019.
- [3] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P. Liu, “Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer”, JMLR: Journal of Machine Learning Research, Vol. 21, No. 140, pp.1-61, 2020.
- [4] 이호동, 이종민, 서재형, 장윤나, 임희석, “생성모델과 검색 모델을 이용한 한국어 멀티턴 응답 생성 연구”, 한국융합학회논문지, 제13권, 제1호, pp.13-21, 2022.
- [5] 김봉민, 홍충선, 박성배, “대화 생성을 위한 고품질 후보 답변 검색 방법”, 한국정보과학회 학술발표논문집, pp.383-385, 2022.
- [6] 국립국어원(2022), 국립국어원 온라인 대화 말뭉치 (버전1.0), Available: <https://corpus.korean.go.kr>, Accessed in 2022.
- [7] 박상민, 이재윤, 손유리, 김재은, “SBERT-IQ: 키워드 정보량을 고려한 Sentence-BERT 기반의 임베딩 모델”, 한국정보과학회 학술발표논문집, pp.1058-1060, 2022.
- [8] SKT-AI/KoBART, <https://github.com/SKT-AI/KoBART>, Accessed in 2022.
- [9] Rouge Metric, ROUGE(metric)Wikipedia, Available : [https://en.wikipedia.org/wiki/ROUGE_\(metric\)](https://en.wikipedia.org/wiki/ROUGE_(metric))