

식당 예약 대화 시스템 개발을 위한 한국어 데이터셋 구축

김경민^o, 이동엽, 허윤아, 임희석
고려대학교 컴퓨터학과

totoro4007@gmail.com, dongyub63@gmail.com, yj72722@korea.ac.kr, limhseok@gmail.com

Development of Korean Dialogue Dataset for Restaurant Reservation System

GyeongMin Kim^o, DongYub Lee, YunA Hur, HeuiSeok Lim
Dept. of Computer Science and Engineering, Korea University

요 약

대화 시스템(dialogue system)은 사용자의 언어를 이해하고 그 의도를 분석하여 사용자가 원하는 목적을 달성할 수 있게 도와주는 시스템이다. 인간과 비슷한 수준의 대화를 위해서는 대량의 데이터가 필요하며 데이터의 양질에 따라 그 결과가 달라진다. 최근 페이스북에서 End-to-end learning 방식을 기반으로 한 영어로 구성된 식당 예약 학습 대화 데이터셋(The 6 dialog bAbI tasks)을 구축하여 해당 모델에 적용한 연구가 있다. 대화 시스템에서 활용 가능한 연구가 활발히 진행되고 있지만 영어 기반의 데이터와는 다르게 식당 예약 시스템에서 다른 연구자들의 연구 목적으로 공유한 한국어 데이터셋은 아직까지도 미흡하다. 본 논문에서는 페이스북에서 구축한 영어로 구성된 식당 예약 학습 대화 데이터셋을 이용하여 한국어 기반의 식당 예약 대화 시스템에서 활용 가능한 한국어 데이터셋을 구축하고, 일상생활에서 발생 가능한 발화(utterance)에 따른 형태 변화를 통해 한국어 식당 예약 시스템 데이터셋 구축 방법을 제안한다.

주제어: 식당 예약 대화 시스템, 한국어 데이터셋, 개체, 다양성

1. 서론

대화 시스템(dialogue system)은 컴퓨터가 사용자의 언어를 이해하고 그 의도를 분석하여 사용자가 원하는 목적을 달성할 수 있도록 도와주는 시스템이다. 컴퓨터는 인간과 비슷한 수준의 대화가 가능하도록 대량의 데이터를 가지고 있어야 하기 때문에 대화 시스템은 이러한 데이터의 양질에 따라 결과가 달라진다. 페이스북은 장, 단기 메모리를 통합하는 메모리 네트워크의 지식 기반 연구에서 객관적이고 주관적인 학습효율 향상을 가져온 영화 데이터셋과[1], 대용량의 자연어 및 개체 처리를 위한 간단한 질문 응답 형식의 멀티 태스킹 및 전송 학습의 영향을 수행한 연구[2]에서 사용된 데이터셋을 구축 및 공개하였다. 최근 페이스북에서 end-to-end learning 방식을 기반으로 한 목적 지향형 대화(goal-oriented dialog) 식당 예약 시스템의 모델 학습을 위해 영문으로 구성된 학습 대화 데이터셋(the 6 dialog bAbI tasks)을 구축하여 해당 모델에 적용한 연구가 있다[3]. 대화 데이터를 활용한 연구가 활발히 진행되고 있지만 영어 기반의 데이터와는 달리 식당 예약 시스템에서 활용 가능한 데이터셋을 다른 연구자들이 연구할 목적으로 공유한 한국어 데이터셋은 아직까지 존재하지 않는다. 최근 hybrid code network 구조를 이용하여 영어로 구축된 페이스북의 학습 데이터셋을 한국어로 번역하여 대화 데이터셋을 구축 및 학습시킨 연구[4]가 있으나 단순 번역한 데이터셋을 활용한 정도이다.

본 논문에서는 end-to-end learning 방식의 식당 예약 대화 시스템 개발을 위해 페이스북에서 구축한 영어로 구성된 식당 예약 학습 대화 데이터셋(The 6 dialog

bAbI tasks)을 이용하여 한국어 식당 예약 대화 시스템에서 활용 가능한 한국어 데이터셋을 구축하고, 일상생활에서 발생하는 발화(utterance)에 따라 다양한 방식으로 시도한 질의응답 데이터 구축 방식을 제안한다.

2. 데이터 구성

2.1 식당 예약 대화 시스템 데이터 번역

순서	대화 순서	영어	한국어
1	인사	hello	반가워
2	예약 진행	can you make a restaurant reservation with (food, location, party_size, price)	(음식, 장소, 인원, 가격) 으로 예약 할래
3	필요 예약 정보 추가	for (food, location, party_size, price) please	~ 으로 할래
4	예약 수정	instead could it be with ~	~ 로 바꿀 거야
5	식당 리스트 변경	no, this does not work for me	다른 걸로, 아닌 거 같아
6	추가 변경 거부	no	아니
7	예약 확정	it's perfect	마음에 들어, 완벽해
8	장소 확인	can you provide the address	식당 주소 뭐야?, 주소 좀 알려줘
9	연락처 확인	may I have the phone number of the restaurant	식당 연락처 뭐야?, 연락처 좀 알려줘
10	감사 인사	thank you	고마워, 고마워 넌 최고야
11	끝맺음	no thank you	아니

●(음식, 장소, 인원, 가격),(food, location, party_size, price) = '~' 로 대체

[표 1] 대화 시스템 원문 데이터 번역본

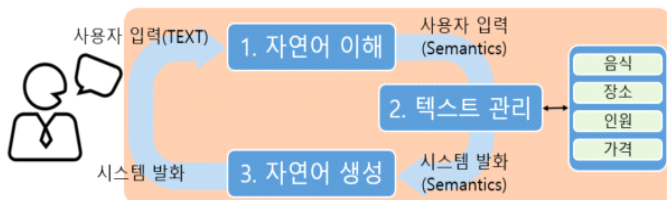
사용자 순서	발화의 목적	발화 형식	다양한 발화 형식의 형태 변화
1	인사	반가워	(생략 가능), 안녕, 안녕하세요, 반갑다, 반가워요, 반갑습니다, 좋은 아침
2	예약	(음식, 장소, 인원, 가격)으로 예약 할래	~ 가능한 식당 예약 할래, ~ 추천해줘, ~ 가고 싶어, ~ 먹고 싶어, ~ 할게, ~ 하려고, N명에서 외식할거야, ~해줄래?, ~예약 하고 싶어, ~찾아줘, 예약하고 싶은데 찾아 줄래?, ~할 만한 식당이 있을까?
3	~ 예약	~로 바꿀 거야 or ~	~로 해줘, ~ 좋겠어, ~가 좋을 것 같아, ~를 찾고 있어
4	예약 수정	~로 바꿔줘 or ~	~말고 ~로 해줘, ~로 바꿀래, 바꾸는 게 좋을 것 같아, 바꾸고 싶어, 바꿀 수 있어? ~로 바꾸자, ~로 바꿔줘, ~ 할게
5	예약 가능성	별로야, 다른 건?	다른 것도 있어?, 마음에 안 들어, 나랑 맞지 않는 것 같아, 너무 별로야 끔찍해, 다른 리스트도 있지?, 내 스타일이 아니야, 다른 건 뭐야, 다른 건 없어?
6	추가 변경 거부	아니	됐어, 없어, 싫어, 충분해, 이 정도면 충분해, 괜찮아, 이게 다예요, 이제 없어 이게 다야, 없는 것 같아
7	예약 확정	좋아, 훌륭해, 마음에 들어, 완벽해	아주, 응(좋아, 훌륭해, 마음에 들어, 완벽해), 진행해줘, 완벽해요, 좋아 보여, 이걸 좋아, 아주 완벽해, 바로 이거야, 이게 좋네, 이거로 가자
8	식당 장소 확인	식당 주소 좀 알려줘, 식당 주소를 알 수 있을까	식당 주소는 뭐야?, 주소 좀 알려줄래?, 식당 주소가 어떻게 되죠? 식당 주소가 필요해, 주소 알려줄 수 있지?, 주소 알아?, 식당 위치가 어디야?
9	식당 연락처 확인	식당 연락처 좀 알려줘, 식당 연락처를 알 수 있을까	식당 연락처는 뭐야?, 연락처 좀 알려줄래?, 식당 전화번호 알아? 전화번호를 알고 싶어, 연락처도 있어?, 연락처도 알 수 있지?, 연락처는?
10	감사 인사	고마워, 고마워 넌 최고야	정말 고마워, 알겠어, 응, 최고예요, 넌 정말 최고야
11	추가 도움 거부	아니	됐어, 없어, 싫어, 충분해, 이 정도면 충분해, 괜찮아, 아니 괜찮아, 이걸로도 충분한 것 같아, 훌륭했어, 이거면 충분해요, 필요 없어

●(음식, 장소, 인원, 가격) = '~' 로 대체

[표 2] 발화 형태에 따른 다양한 데이터 변화

위의 [표 1]에서처럼 대화 순서에 따른 영어 원문을 한국어로 번역하여 학습 데이터셋을 구성하였고 각각의 발화(utterance) 순서는 초기 데이터 형식을 이용하였으나, 본 연구의 최종 목표인 데이터셋 활용에서는 발화(utterance)의 순서와 상관없이 학습이 이루어졌다. 한국어 식당 예약 시스템의 대화 데이터셋 구성을 위해 페이스북에서 구축한 영어로 구성된 식당 예약 학습 대화 데이터셋(the 6 dialog bAbI tasks)을 이용하였으며 연구를 위해 사용된 데이터셋은 깃 허브(Git Hub)를 통해 공유하였다[5].

대화 시스템은 사용자의 발화를 입력받아 사용자로부터 식당 예약에 필요한 속성이 될 수 있는 개체(entity)를 추출하는데, 개체 추출을 위해서는 각 개체들의 해당 개체리스트를 정의하고 문자열 매칭 알고리즘을 통해 문자열 속의 개체를 추출할 수 있다[4]. 대화 시스템은 텍스트 관리 과정에서 필요한 개체 정보를 사용자에게 요구하며 누락된 정보는 반복적인 질의 행위를 통해 획득할 수 있다. 대화 시스템 구조와 각 개체에 해당하는 내용은 [그림 1]과 같다.



[그림 1] 대화 시스템 구조

2.2 발화 방식의 다양한 변화

목적에 따른 발화 형식을 바탕으로 한 대화 시스템에

서의 발화 형태 변화를 [표 2]를 통해 볼 수 있다. 실제 대화는 대화 시스템에서와 다르게 인사 생략 등 단순 의미 전달 만으로도 대화가 가능한 것을 고려하여 발화 형식 별로 6 ~ 13 개의 다양한 형태로 변화시켰으며, [표 2]에 표기된 사용자 순서 필드와 상관없이 실제 대화가 일어날 수 있는 발화의 목적에 따른 상황에서의 순서를 고려하여 데이터셋을 구축하였다.

식당 예약 시스템과 사용자 간 발화 정보는 [표 3]을 통해 확인할 수 있다. 발화의 주체는 사용자(user)와 시스템(system)으로, 시스템은 사용자로부터 원하는 정보 획득을 위해 반복적인 질의를 이용한다.

발화의 주체	발화	발화의 주체	발화
U(User)	안녕	S	이 식당은 어떤가요? <식당 1>
S(System)	안녕하세요.	U	별로야, 다른 곳은 없어?
U	어떻게 도와 드릴까요	S	다른 식당 리스트를 보여 드릴게요
U	파리에서 3명이 인도 요리를 먹을 수 있는 식당 예약해줘	S	이 식당은 어떤가요? <식당 2>
S	알겠습니다. 가격의 범위는 어느 정도로 생각하세요?	S	아주 만족스러워
U	적당한 가격으로	U	예약을 진행해드리도록 하겠습니다
S	좋아요 몇 가지 리스트를 보여 드릴게요	S	식당 연락처 좀 알려줄래?
S	api_call 인도 프랑스 세 명 적당한	S	전화번호는 ... 있습니다
U	저렴한 가격으로 바꿔줘	U	고마워
S	네 또 변경하실 사항이 있나요?	S	또 도와드릴 게 있나요?
U	아니	U	아니 없어
S	api_call 인도 프랑스 세 명 저렴한	S	감사합니다

[표 3] 식당 예약 시스템과 사용자 간 발화 정보

원하는 개체를 입력받은 시스템은 사용자로부터 개체 변경 의사에 대한 질의를 통해 해당 개체를 수정할 수

있으며 사용자에게 원하는 식당에 대한 정보를 제공함으로써 대화를 마무리한다.

3. 결론

본 논문에서는 한국어 식당 예약 시스템 구축을 위한 식당 예약 학습 데이터셋 수집 방식과 실제 대화에서 일어날 수 있는 발화 방식에 따른 데이터 형식의 형태 변화에 대해 제안하였다. 식당 예약 시스템의 영어 기반 데이터가 구축된 것에 비해 식당 예약 시스템의 한국어 데이터 연구는 아직까지 미흡하며 연구를 목적으로 한 데이터 공유가 존재하지 않는다. 이번 연구에서 사용된 한국어 식당 예약 시스템 데이터를 깃 허브(Git Hub)를 통해 공유함으로써 이번 연구를 토대로 해당 분야의 다른 연구자들에게 많은 기여가 될 것으로 기대한다.

Acknowledgement

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2017년도 문화기술 연구개발 지원 사업으로 수행되었음. [2017. 전통문화 융복합 지원을 위한 지능형 검색 플랫폼 구축]

참고문헌

- [1] J. Dodge, A. Gane, Xiang Zhang, A. Bordes, S. Chopra, H. Miller, A. Szlam and J. Weston, Evaluating Prerequisite Qualities for Learning End-to-End Dialog Systems, 2016.
- [2] A. Bordes, N. Usunier, S. Chopra and J. Weston, Large-scale Simple Question Answering with Memory Networks, 2015.
- [3] A. Bordes, Y. Boureau & J. Weston, Learning End-to-End Goal-Oriented Dialog, 2017.
- [4] 이동엽, 허윤아, 임희석, "Hybrid Code Network를 이용한 한국어 식당 예약 시스템 모델", 한국컴퓨터교육학회, 제 21권, 제 2호, pp.57-59, 2017.
- [5] 이동엽, 김경민, "Korean Restaurant Reservation," (2017), GitHub repository, https://github.com/JudeLee19/korean_restaurant_reservation