

도메인 특정 지식을 결합한 End-to-End Learning 방식의

한국어 식당 예약 대화 시스템 모델 개발

이동엽, 김경민, 임희석

고려대학교 컴퓨터학과

judelee93@korea.ac.kr, totoro4007@gmail.com, limhseok@korea.ac.kr

Development of a Dialogue System Model for Korean Restaurant

Reservation with End-to-End Learning Method Combining Domain Specific

Knowledge

Dong-Yub Lee, Gyeong-Min Kim, Heui-Seok Lim

Dept. of Computer Science and Engineering, Korea University

요약

목적 지향적 대화 시스템(Goal-oriented dialogue system)은 텍스트나 음성을 통해 특정한 목적을 수행할 수 있는 시스템이다. 최근 RNN(recurrent neural networks)을 기반으로 대화 데이터를 end-to-end learning 방식으로 학습하여 대화 시스템을 구축하는데에 활용한 연구가 있다. End-to-end 방식의 학습은 도메인에 대한 지식 없이 학습 데이터 자체만으로 대화 시스템 구축을 위한 학습이 가능하다는 장점이 있지만 도메인 지식을 학습하기 위해서는 많은 양의 데이터가 필요하다. 이에 본 논문에서는 도메인 특정 지식을 결합하여 end-to-end learning 방식의 학습이 가능한 Hybrid Code Network 구조를 기반으로 한국어로 구성된 식당 예약에 관련한 대화 데이터셋을 이용하여 식당 예약을 목적으로하는 대화 시스템을 구축하는 방법을 제안한다. 실험 결과 본 시스템은 응답 별 정확도 95%와 대화 별 정확도 63%의 성능을 나타냈다.

주제어: 대화 시스템, 딥러닝, 도메인 지식

1. 서론

목적 지향적 대화 시스템(Goal-oriented dialogue system)은 텍스트나 음성을 통해 특정한 목적을 수행할 수 있는 시스템이다. 목적 지향적 대화 시스템을 구현하기 위한 이전 연구로는 slot-filling의 방식과 [1,2,3] language understanding, action selection과 같이 대화를 이해하고 그에대한 응답을 선택하는 모듈들을 이용하는 연구가 있다. 하지만 이러한 이전 연구들은 도메인에 대해 많은 지식을 포함하는 hand-craft 된 자질(feature)을 필요로 하고 이로인해 새로운 도메인에 대한 확장이 어렵다는 단점이 있다.

최근 RNN(recurrent neural networks)을 기반으로 대화 데이터를 end-to-end learning 방식으로 학습하여 대화 시스템을 구축하는데에 활용한 연구가 있다[4,5]. End-to-end 방식의 학습은 도메인에 대한 지식 없이 학습 데이터 자체만으로 대화 시스템 구축을 위한 학습이 가능하다는 장점이 있다. 하지만 end-to-end learning 방식을 이용하여 도메인 지식을 학습하기 위해서는 많은 양의 데이터가 필요하다. 식당 예약 시스템의 경우, 식당의 위치나 가격과 같은 특정 도메인에 대한 지식 정보를 학습하기 위해서는 해당 정보들을 표현하는 많은 양의 학습 대화 데이터셋이 필요하다.

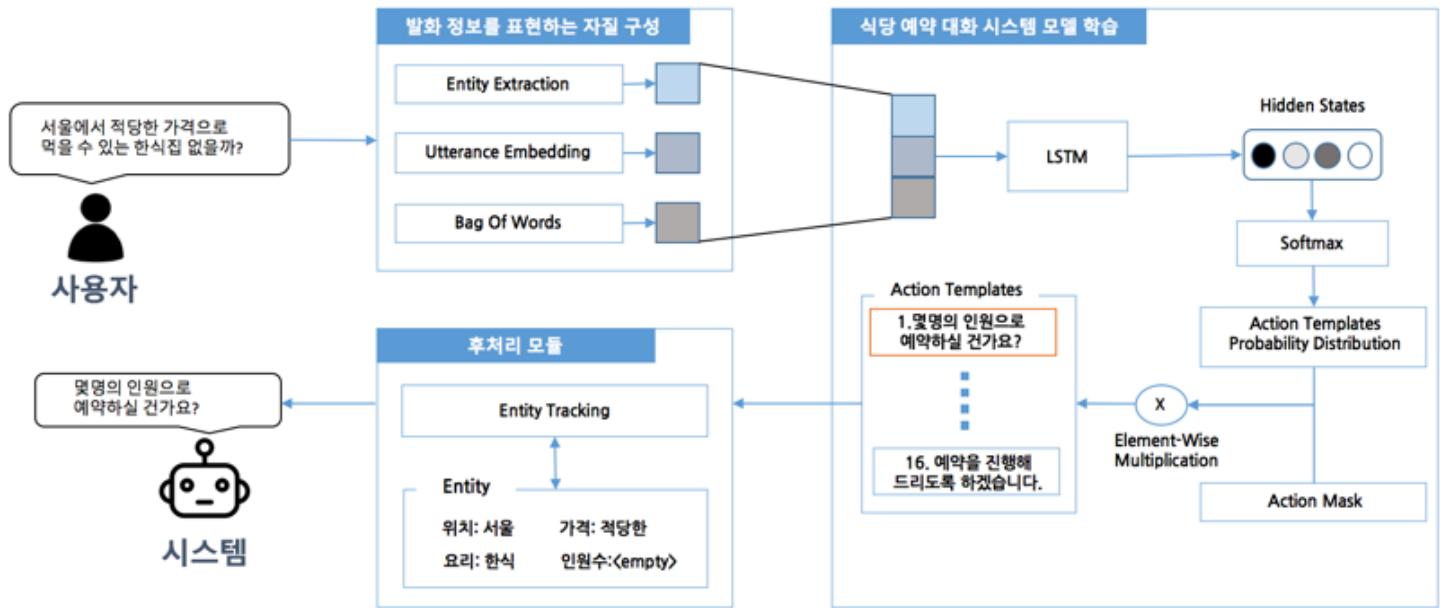
많은 양의 학습 대화 데이터셋이 필요하다는 일반적인 end-to-end learning 기반의 단점을 극복하기 위해 도메인 특정 지식을 결합하며 end-to-end learning 방식의 학습을 가능하게 한 Hybrid Code Network 구조를 제안한 연구가 있다[6]. Hybrid Code Network를 이용하여 구축한 대화 시스템은 도메인에 해당하는 특정 지식을 액션 템플릿(action template)의 정의를 통해 표현함으로써 보다 적은양의 학습 데이터 양으로 도메인 지식을 표현할 수 있다.

목적 지향적 대화 시스템의 학습을 위해 페이스북은 식당 예약에 관련한 학습 데이터셋(The 6 bAbI tasks)을 공개하였다[7]. 영어로 구축된 식당 예약에 관련한 학습 대화 데이터셋을 이용하여 한국어로 이루어진 식당 예약 관련 대화 데이터셋을 구축한 연구가 있다.[8, 9]

본 논문에서는 Hybrid Code Network를 기반으로 한국어로 구성된 식당 예약에 관련한 대화 데이터셋을 이용하여 식당 예약을 목적으로하는 대화 시스템을 구축하는 방법을 제안한다.

2. 제안하는 과정

[그림 1]은 본 논문에서 제안하는 도메인 특정 지식을



[그림 1] 도메인 특정 지식을 결합한 한국어 식당 예약 대화 시스템 모델의 전체 구조도

결합하고 end-to-end learning 방식의 학습이 가능한 한국어 식당 예약 대화 시스템 모델의 전체 구조도를 나타낸다. 본 논문에서는 한국어 식당 예약 대화 시스템 모델의 학습을 위해 학습 대화 데이터 759개와 테스트 대화 데이터 190개를 이용하였다. 사용자의 발화(utterance)로부터 식당 예약 대화 시스템 모델 학습에 필요한 자질(feature)을 구성하기 위해 개체 추출(Entity Extraction), 발화 임베딩(Utterance Embedding), 단어 주머니(Bag of Words) 와 같은 3가지 방법으로 발화 정보를 표현하는 자질을 구성한다. 각각의 방법으로 표현된 발화 정보를 표현하는 자질들은 연결(concatenation)되어 발화 정보에 대한 최종 자질을 형성한다. 형성된 발화 정보를 표현하는 최종 자질은 LSTM(long short term memory)의 입력으로 사용되고 LSTM은 은닉 상태(hidden states)를 계산하여 출력한다. 출력된 은닉 상태는 softmax의 입력으로 사용된다. 본 실험에서는 사용자에게 몇명의 인원으로 식당 예약을 할 것인지, 어떤 요리를 원하는지와 같은 발화를 식당 예약에 관련된 도메인 지식을 표현하기 위한 액션 템플릿(Action Template)로 정의하는데, softmax는 주어진 사용자의 발화에 대한 액션 템플릿들의 확률 분포(probability distribution)를 계산한다. 액션 템플릿들에 대한 확률 분포 값은 사용자의 발화에 대해 어떠한 액션 목록들이 필요한지 표시하는 액션 마스크(Action Mask) 값들과 성분끼리의 곱(element-wise multiplication)을 통해 최종적으로 시스템의 발화를 선택하게 된다. 후처리 모듈에서는 사용자의 발화에서 식당의 위치에 해당하는 서울과 같은 개체(entity)들을 추적(tracking)하여 최종적으로 선택된 시스템의 발화에서 추적중인 개체명이 필요할 시 추적중인 개체를 사용할 수 있도록 한다.

2.1 발화 정보를 표현하는 자질 구성

2.1.1 개체 추출(Entity Extraction)

본 논문에서는 식당의 위치, 음식의 가격, 음식의 종류, 인원 수 와 같이 식당 예약에 필요한 속성이 될 수 있는 개체들을 정의한다. 개체 추출을 진행하기 위해 각 속성들에 해당하는 개체 사전을 정의하고 발화로부터 문자열 매칭 알고리즘을 이용하여 발화 속에 존재하는 개체들을 추출할 수 있다.

2.1.2 발화 임베딩(Utterance Embedding)

발화 임베딩 모듈은 사용자의 발화로부터 의미론(semantics)적 특성을 반영하기 위해 word2vec 모델을 이용하여 사용자의 발화를 임베딩한 후 이를 자질로 구성한다. [식 1]은 발화를 구성하고 있는 전체 단어의 개수가 N, 문서의 i번째 단어를 워드 임베딩 공간에서 벡터 값으로 표현한 것을 v(i)라 할 때, 발화를 구성하고 있는 각 단어들의 벡터 평균을 나타낸다.

$$\frac{1}{N} \sum_{i=1}^N v(i) \quad (1)$$

2.1.3 발화 주머니(Bag of Words)

사용자의 발화를 발화 주머니로 구성하기 위해 학습 대화 데이터셋에 있는 단어들의 집합을 기반으로 사전을 형성하였다. 이후 단어 사전을 기반으로 각 발화를 구성하는 단어들의 등장 여부를 표시하여 사용자의 발화를 표현하는 발화 주머니 자질을 구성하였다.

2.2 LSTM을 이용한 대화 시스템 모델 학습

시스템의 응답에 도메인 특정 지식을 반영하기 위해 [그림 2]와 같이 액션 템플릿을 정의한다. 본 논문에서 제안하는 식당 예약 대화 시스템 모델의 경우 총 16개의 액션 템플릿으로 구성되어 있다. 2.1장에서 구성한 발화 정보를 표현하는 자질들은 연결되어 발화 정보에 대한

최종 자질을 형성하고 이를 입력으로 이용하여 LSTM 은 은닉 상태를 계산하게 된다. 은닉 상태를 입력으로

하이퍼 파라미터에 대한 정보를 나타낸다.

Action Templates
0. api_call <음식종류> <위치> <인원수> <가격>
1. 가격의 범위는 어느정도로 생각하세요
2. 감사합니다
3. 네 또 변경하실게 있나요
4. 다른 리스트를 보여드릴게요
5. 또 도와드릴게 있나요
6. 몇명의 인원으로 예약하실 건가요
7. 안녕하세요 어떻게 도와드릴까요
8. 알겠습니다
9. 어떤 종류의 요리를 좋아하나요
10. 예약을 진행해드리도록 하겠습니다
11. 위치는 <info_address> 입니다
12. 위치는 어디에 있어야 하나요
13. 이 리스트는 어떤가요: <restaurant>
14. 전화번호는 <info_phone> 입니다
15. 좋아요 몇 가지 리스트를 보여드릴게요

[그림 2] 도메인 특정 지식을 반영하기 위한 시스템 응답 액션 템플릿 정의

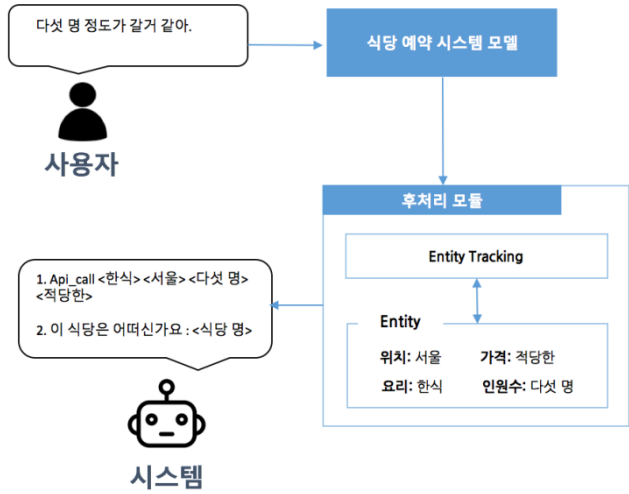
softmax는 각 액션 템플릿들에 대한 확률 분포 값을 계산한다. 이후 각 확률 분포 값들은 액션 마스크와 성분끼리의 곱을 통해 정규화(normalization)되고 이 값들과 실제 학습 데이터셋에서 레이블링 된 정답 시스템 응답과의 cross-entropy 값을 최소화 하도록 모델의 학습이 진행된다.

2.3 후처리 모듈을 통한 개체명 추적

후처리 모듈에서는 [그림 3]과 같이 사용자 발화에 대한 식당 예약 대화 시스템 모델의 예측 결과 선택된 액션 템플릿의 발화에서 개체명이 필요할 시, 해당 개체명을 같이 응답할 수 있도록 이전 발화로부터 추출된 개체들을 추적한다.

2.4 학습 모델 하이퍼 파라미터 설정

[표 1]은 한국어 식당 예약 대화 시스템 학습에 이용한



[그림 3] 이전 발화에서 추출한 개체들을 이용한 시스템 응답

[표 1] 학습에 이용된 모델의 하이퍼 파라미터 값

	Hyper-parameter	Value
word2vec	window size	5
	dimension	300
Bag of words	vocab size	1051
LSTM	initial state	0.0
	state size	128
	forget bias	1.0
	training epoch	20
	action template size	16
	initial learning rate	0.1
	decay rate	0.0

3. 실험 결과

한국어 식당 예약 대화 시스템 모델의 성능을 평가하기 위해 영어로 구성된 식당 예약 대화 데이터셋을 이용하여 실험한 식당 예약 대화 시스템 모델의 연구[6,7]와 같이 응답 별 정확도(per-response accuracy)와 대화 별 정확도(per-dialogue accuracy)을 측정하였다. 응답 별 정확도와 대화 별 정확도는 [식 2], [식 3]와 같이 나타낼 수 있다.

$$\text{per-response accuracy} = \frac{\text{시스템이 올바르게 응답한 발화의 개수}}{\text{모든 대화를 구성하는 발화의 개수}} \quad (2)$$

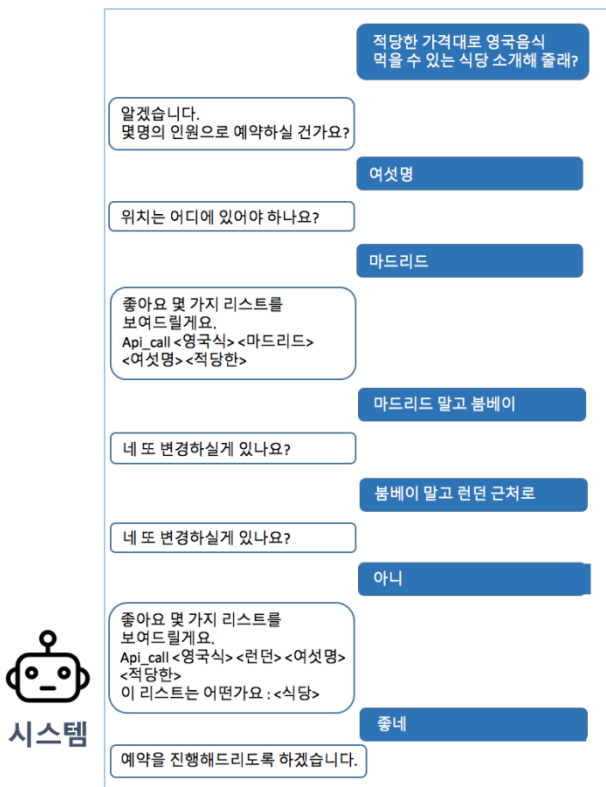
$$\text{per-dialogue accuracy} = \frac{\text{시스템이 모두 올바르게 응답한 대화의 개수}}{\text{모든 대화의 개수}} \quad (3)$$

[표 2]는 본 실험에서 진행한 한국어 식당 예약 대화 시스템의 응답 별 정확도와 대화 별 정확도를 나타낸다.

[표 2] 한국어 식당 예약 시스템의 응답 별 정확도와 대화 별 정확도

Methods	Accuracy
Per-response Accuracy	0.95
Per-dialogue Accuracy	0.64

[그림 4]는 학습된 모델을 이용하여 식당 예약을 진행하는 대화 시스템 모델의 실행 결과를 나타낸다. 사용자의 첫 발화에서 식당 예약을 진행하는데 필요한 속성 중 인원 수와 위치가 부재되어 있어, 시스템은 이후의 대화에서 사용자에게 인원 수와 예약하고자 하는 식당의 위치를 물어본다. 필요한 속성들이 충족되면 시스템은 해당 속성들의 조건을 만족하는 api를 호출한다. 만약 사용자가 속성 값을 변경하면 시스템은 다른 변경사항이 또 있는지 사용자에게 물어보고 변경 사항이 없을 경우 변경된 조건을 만족하는 api를 호출하고 호출 결과 반환되는 식당 리스트를 사용자에게 제안한다. 사용자가 시스템이 제안한 식당 리스트를 만족할 경우, 시스템이 예약을 진행하며 대화가 종료된다.



[그림 4] 식당 예약을 진행하는 대화 시스템 모델

4. 결론

본 논문에서는 Hybrid Code Network를 기반으로 한국어로 구성된 식당 예약에 관련한 대화 데이터셋을 이용하여 식당 예약을 목적으로하는 대화 시스템을 구축하는 방법을 제안하였다. 실험 결과 구축한 대화 시스템은 도메인에 해당하는 특정 지식을 액션 템플릿의 정의를 통해 표현함으로써 보다 적은양의 학습 데이터 양으로 도메인 지식을 표현할 수 있었다. 또한 대화 도중 사용자가 식당 예약에 관련된 속성 값을 변경하고자 할 때에도 시스템이 해당 변경사항을 반영하여 식당 예약을 진행할 수 있었다. 실험 결과 본 시스템은 응답 별 정확도 95%와 대화 별 정확도 63%의 성능을 나타냈다.

본 실험에서는 사용자의 발화에서 개체를 추출하기 위해 각 속성에 해당하는 개체 사전을 정의하여 이를 기반으로 개체 속성을 추출하였는데, 향후 연구에서는 개체 사전에 나타나지 않는 OOV(out of vocabulary) 개체에 대해서도 올바른 속성 추출이 가능할 수 있도록 연구할 필요가 있다.

Acknowledgement

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

참고문헌

[1] Lemon, O , Georgila, K, Henderson, J, and Stuttle, M. "An isu dialogue system exhibiting reinforcement learning of dialogue policies: generic slot-filling in the talk in-car system," In Proceedings of the 11th. Conference of the European Chapter of the ACL: Posters & Demonstrations, pages 119-122. 2006

[2] Wang, Z. and Lemon, O. "A simple and generic belief tracking mechanism for the dialog state tracking challenge: On the believability of observed information," In Proceedings of the SIGDIAL Conference. 2013

[3] Young, S, Gasic, M, Thomson, B , and Williams, J. D. "Pomdp-based statistical spoken dialog systems: A review," Proceedings of the IEEE, 101(5), 1160-1179. 2013

[4] Dzmitry, B, Kyunghyun, C, and Yoshua, B. "Neural machine translation by jointly learning to align and translate," In ICLR. 2015

[5] Oriol, V, Łukasz, K, Terry, K, Slav P, Ilya, S, and Geoffrey, H. "Grammar as a foreign language," In Advances in Neural Information Processing Systems, pages 2755- 2763. 2015

[6] Jason D Williams, Kavosh Asadi, and Geoffrey

Zweig. 2017. Hybrid code networks: practical and efficient end-to-end dialog control with supervised and reinforcement learning. In ACL.

[7] Bordes,A, and Weston,J, “Learning End-to-End Goal-Oriented Dialog,” ArXiv e-prints. 2016

[8] 이동엽, 허윤아, 임희석, “Hybrid Code Network를 이용한 한국어 식당 예약 시스템 모델,” 컴퓨터 교육학회. 2017.08.

[9] 이동엽, 김경민, “Korean Restaurant Reservation” (2017), GitHub repository, https://github.com/JudeLee19/korean_restaurant_reservation