

T5 모델을 이용한 스키마 기반의 한국어 대화 상태 추적

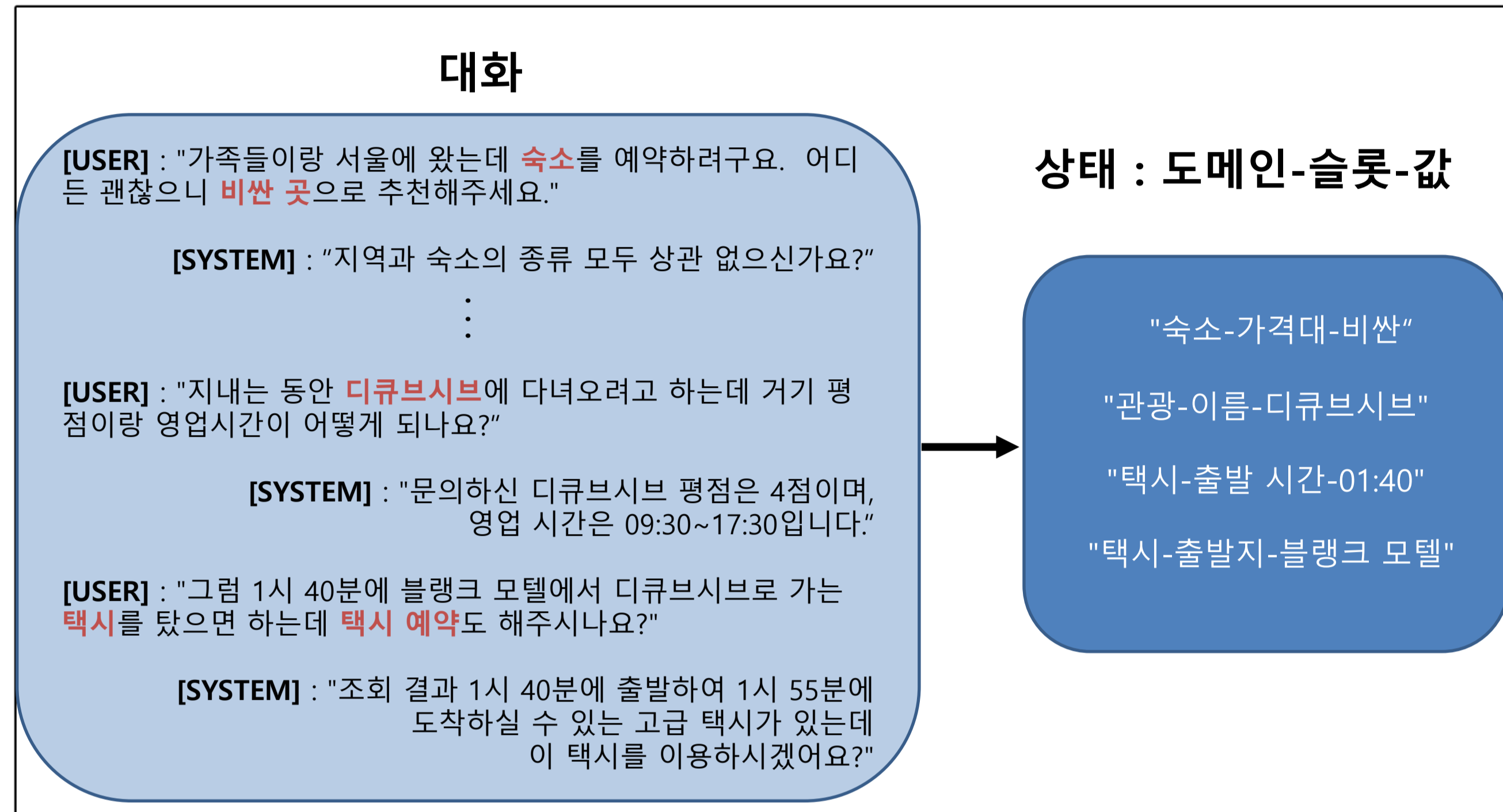
박준범, 이정두, 이종현, 나승훈
전북대학교

{pjb2516, bo0od12}@naver.com, {gus8423, nash}@jbnu.ac.kr

I. 서론

대화 상태 추적(dialogue state tracking)이란 대화 시스템에서 사용자와 시스템 간의 각 대화 문장으로부터 정보를 추출하는 것을 말한다.

이는 대화 중 발생하는 여러 주제 속에서도 시스템이 흐름을 놓치지 않게 해주어 결국 사용자의 질의 의도를 정확하게 파악하는 것에 영향을 줄 수 있기 때문에 매우 중요하다.



II. 제안 방법

Data

실험에 사용된 데이터는 KLUE에서 대화 상태 추적 모델 학습 및 평가를 위해 Wizard-of-Oz 프레임워크를 이용해 한국어로 구성된 WoS-v1.1 데이터셋을 사용하였다.

KLUE-benchmark WoS-v1.1 데이터 셋

	Single domain dialogues	Multi domain dialogues
Train set	1,806	6,194
Dev set	263	737
Test set	226	774
Total	2,295	7,705

WoS-v1.1 데이터 셋에 포함된 스키마 구성

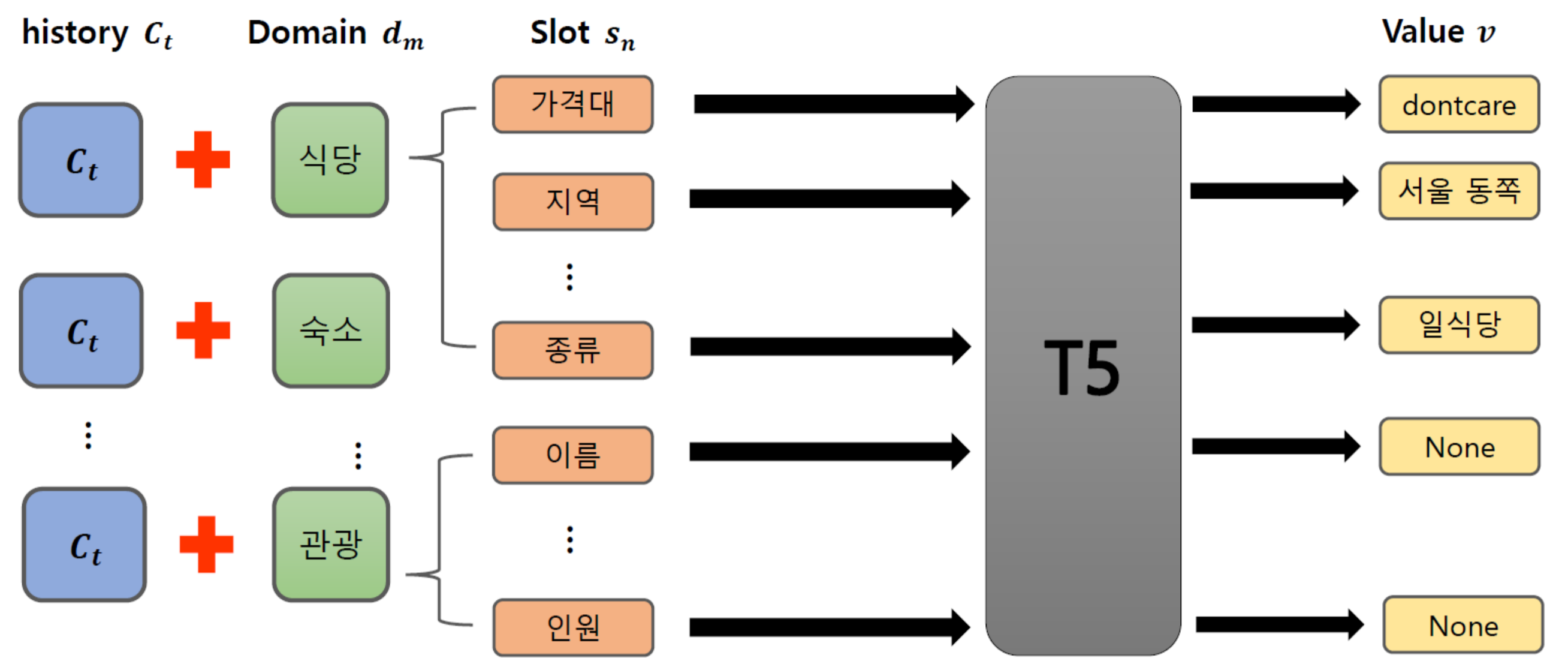
domain	slot	value
숙소	15	105
식당	13	723
관광	9	121
택시	5	808
지하철	3	71

Definition

사용자 입력 : U_t
시스템 답변 : A_t
M개의 도메인 : $D = [d_1, \dots, d_M]$
N개의 슬롯 : $S = [s_1, \dots, s_N]$
입력 컨텍스트 : $C_t = [U_1, A_1, \dots, A_{t-1}, U_t]$

Proposed System

디코더 전용 모델보다 우수한 인코더-디코더 모델의 강점을 갖는 T5 모델을 이용해 스키마 기반의 데이터셋을 학습시켜 대화 상태를 추적하는 성능을 측정한다.



U_{t-1} : 명동 쇼핑 거리에 대해 물어볼게 있는데 영업시간이랑 입장료, 주소를 알려주세요.
 A_{t-1} : 먼저 명동 쇼핑 거리에 영업시간은 09시부터 20시까지입니다. 그리고 입장료는 10000원이고 주소는 서울 중구 38345입니다.
 U_t : 그리고 다음으로 서울 동쪽에 주류를 판매하는 일식당을 찾아주세요. 가격대는 상관 없습니다.

T5 모델을 이용한 학습 예시

입력

컨텍스트 C_t 와 사용자의 목적에 기반한 프롬프트를 입력으로 한다. 이 때 작업별 프롬프트는 주어진 스키마를 기반으로 형성된 도메인 관련 프롬프트 $X(d_m)$ 와 슬롯 관련 프롬프트 $X(s_n)$ 라는 두 가지 유형으로써 학습에 참여하게 된다.

각 요소들은 [user], [system], [domain], [slot]과 같이 시작점을 표시할 수 있는 특수 세그먼트 토큰으로 구분되어 "[user] U_t [system] A_{t-1} ... [user] U_{t-1} [system] A_{t-1} [user] U_t [domain] $X(d_m)$ [slot] $X(s_n)$ "와 같은 형태로 양방향 인코더에 입력되어진다.

인코더는 시퀀스 길이가 L 이고 인코더 hidden size K 를 갖는 $H_t \in \mathbb{R}^{L \times K}$ 를 디코더로 보내

$$H_t = \text{Encoder}(C_t, X(d_m), X(s_n))$$

이후 디코더는 H_t 를 입력으로 하여 각 도메인과 슬롯에 대한 값 B_t 를 출력한다.

$$B_t(d_m, s_n) = \text{Decoder}(H_t)$$

최종적으로 모델은 log-likelihood를 최대화하는 방식으로 학습된다.

$$\sum_{m,n} \log P(B_t(d_m, s_n) | C_t, X(d_m), X(s_n))$$

III. 실험 결과

실험 평가는 joint goal accuracy(JGA)와 micro F1 score 을 이용하여 슬롯 레벨에서의 정확도 뿐만이 아닌 state 레벨에서의 정확도 또한 평가 하였다.

모델	JGA(%)	F1(%)
KLUE-RoBERTaLARGE	49.39	92.19
KE-T5-SMALL(ours)	43.23	89.53
KE-T5-BASE(ours)	62.82	95.46

각 모델의 실험 방법을 요약하면 다음과 같다.

- KE-T5-SMALL, KE-T5-BASE : 평균 14개의 발화로 구성되어있는 총 8,000개의 학습 데이터를 이용하며 모델을 3 epoch 를 학습 시킨후 평가하였다.

KE-T5-SMALL의 경우 43.24% JGA, 89.53% F1score 을 달성 했으며, KE-T5-BASE의 경우 62.82% JGA, 95.46% F1 score을 달성하여 앞서 제시된 모델중 JGA와 F1 score에 대해 최고 점수를 보인 KLUE-RoBERTaLARGE 보다 각각 13.43%p, 3.27%p 높은 최고 성능을 보였다.

향후 스키마에서 제공하는 도메인과 슬롯에 대하여 각각 자연어로 된 설명이 추가된 새로운 스키마를 기반하여 프롬프팅 기법을 활용한 대화 상태 추적 태스크를 연구할 것이다.