
스키마 기반 자연어 증강 프롬프트를 이용한 한국어 대화 상태 추적

이서준°, 나승훈
신명철, 이동훈, 김응균
sjl_@naver.com°

전북대학교°
카카오 엔터프라이즈

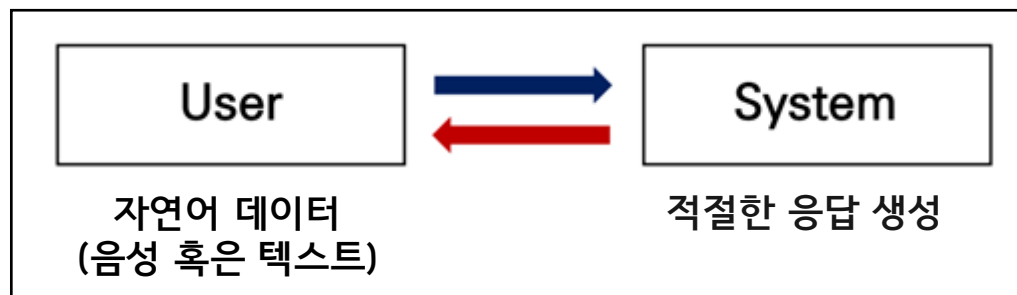


kakaoenterprise

1. 서론

목적 지향 대화 시스템 Task-Oriented Dialogue System

: 특정 태스크에서 사용자 요구사항을 이해하고 이를 해결하기 위한 적절한 대화 시스템

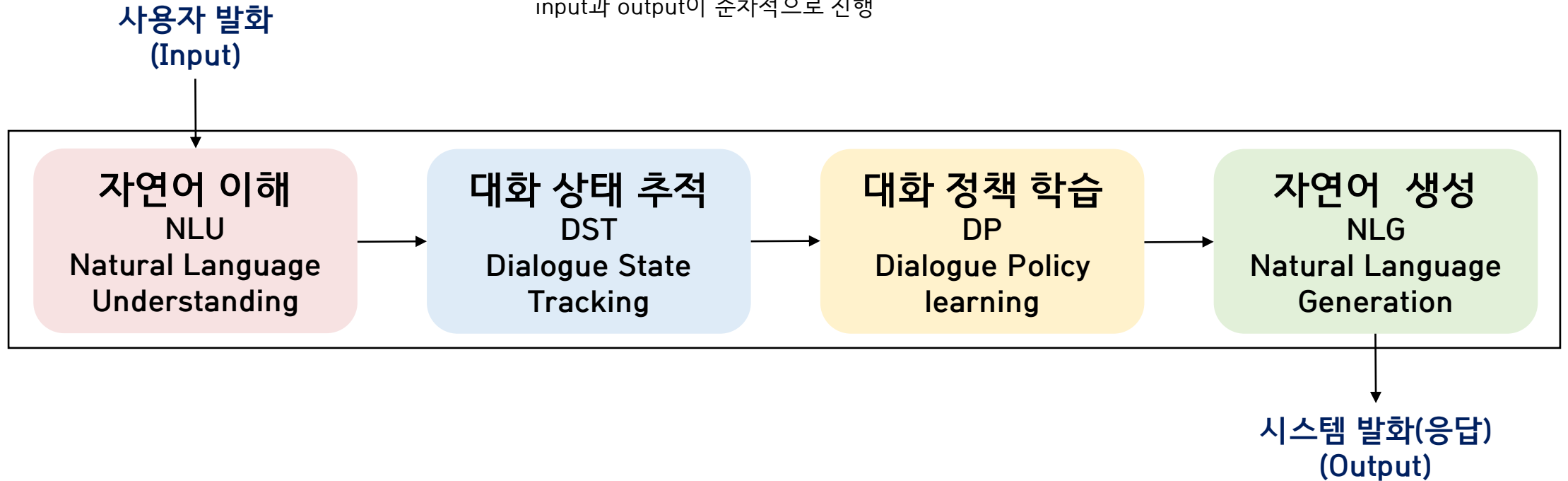


- User : 사용자
- System : Chat-bot, Siri, ChatGPT 등
- Data : 특정 시나리오를 가정하여 구성된 데이터
(호텔 예약, 식당 예약, 관광지 찾기 등)

1. 서론

목적 지향 대화 시스템 Task-Oriented Dialogue System

- 크게 4개의 모듈로 나누어짐
- 각각의 태스크는 독립적인 모델로 수행되는 분야이지만 input과 output이 순차적으로 진행

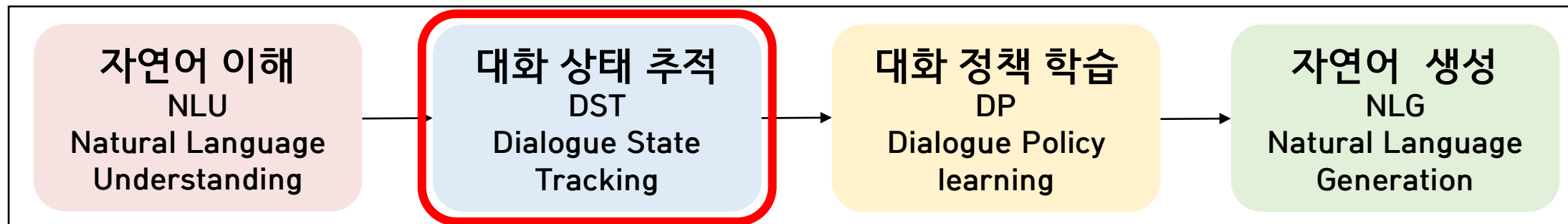


1. 서론

대화 상태 추적 Dialogue State Tracking

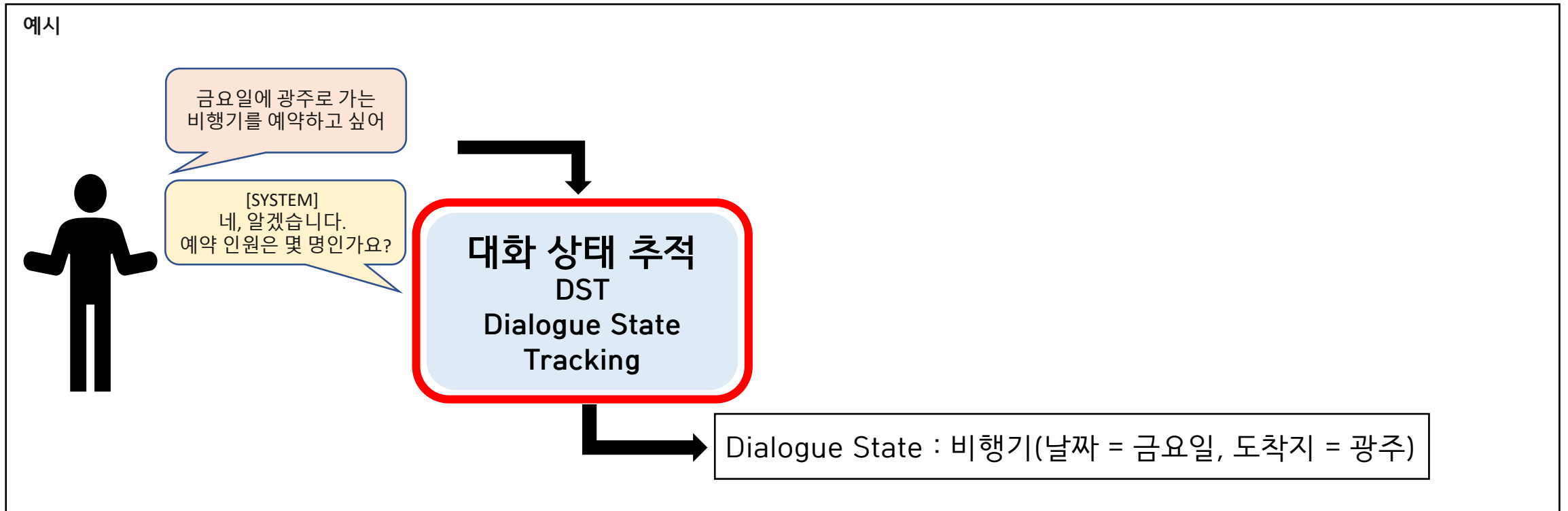
: 대화 시스템에서 사용자와 시스템 간의 각 대화 문장에서 정보 추출

- Data : 사용자와 시스템 간에 교대로 발화 시퀀스를 구성
- Input : 대화 (턴에 따라 진행됨)
- Output : 대화 상태 → (slot, value) 쌍



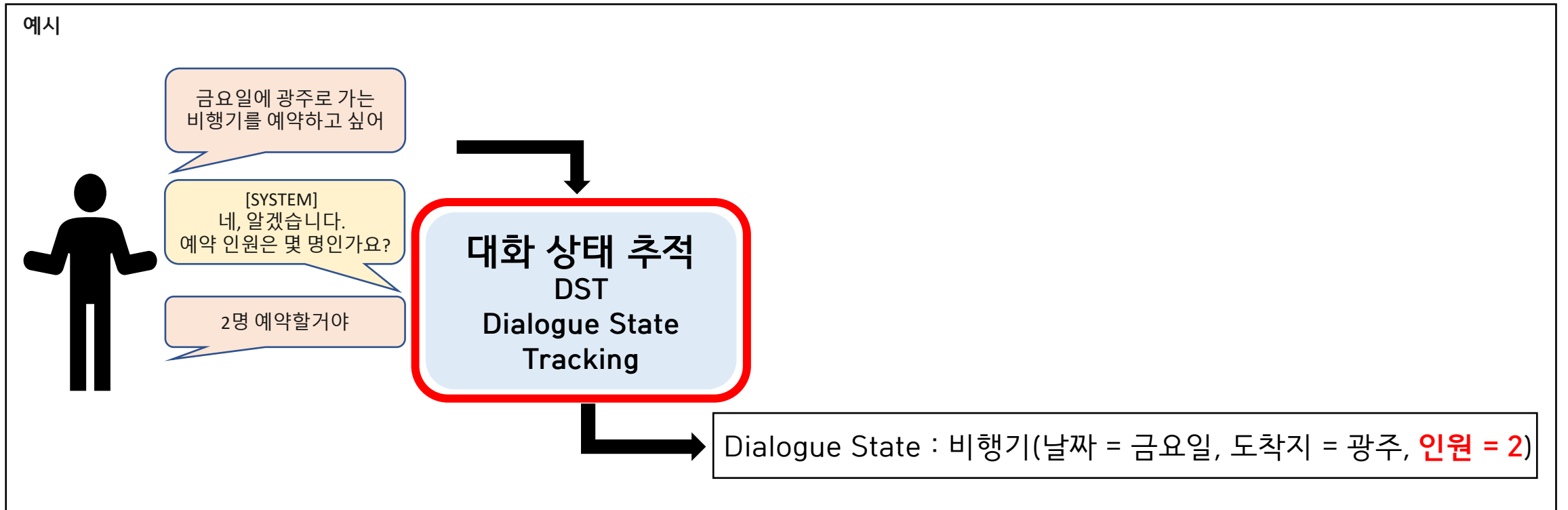
1. 서론

대화 상태 추적 Dialogue State Tracking



1. 서론

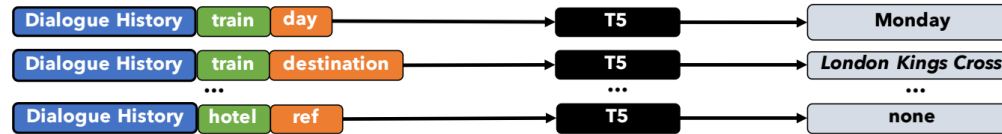
대화 상태 추적 Dialogue State Tracking



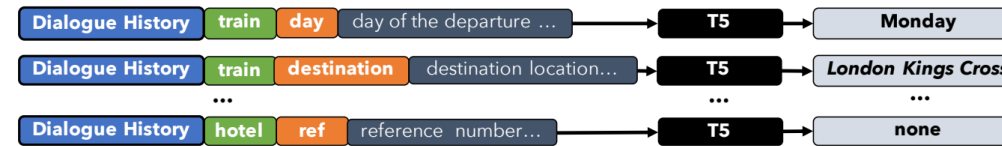
2. 제안 모델



(a) Generation-based DST w/ Sequential Decoding



(b) Schema-Based Prompt DST w/ Independent Decoding



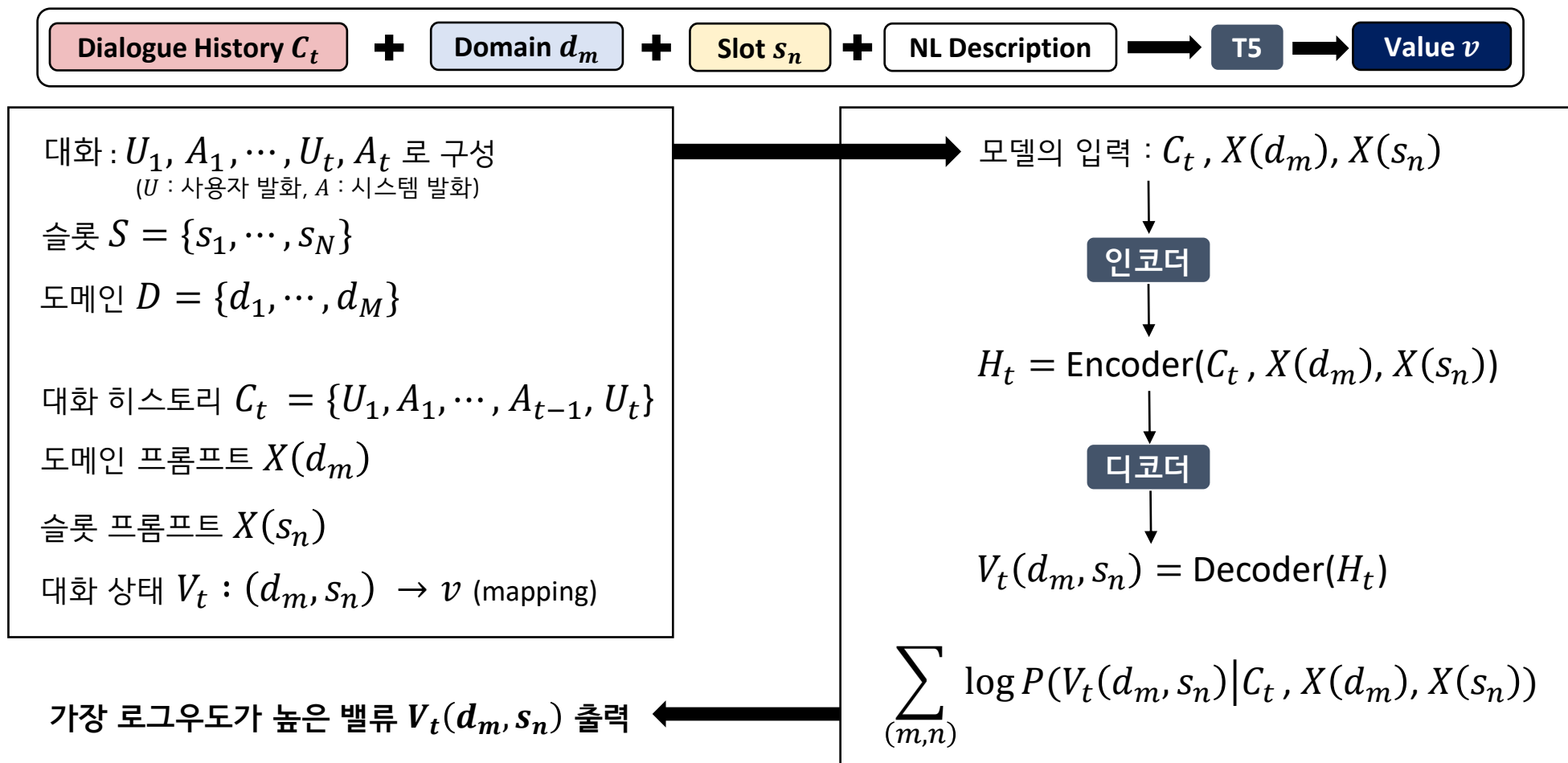
(c) Natural Language Augmented Prompt DST w/ Independent Decoding

- [1]의 방법을 이용
- T5를 백본 모델로 활용하여 대화 상태 추적 모델을 제안
- 프롬프트 기반 입력을 이용함으로써 좋은 성능 도출

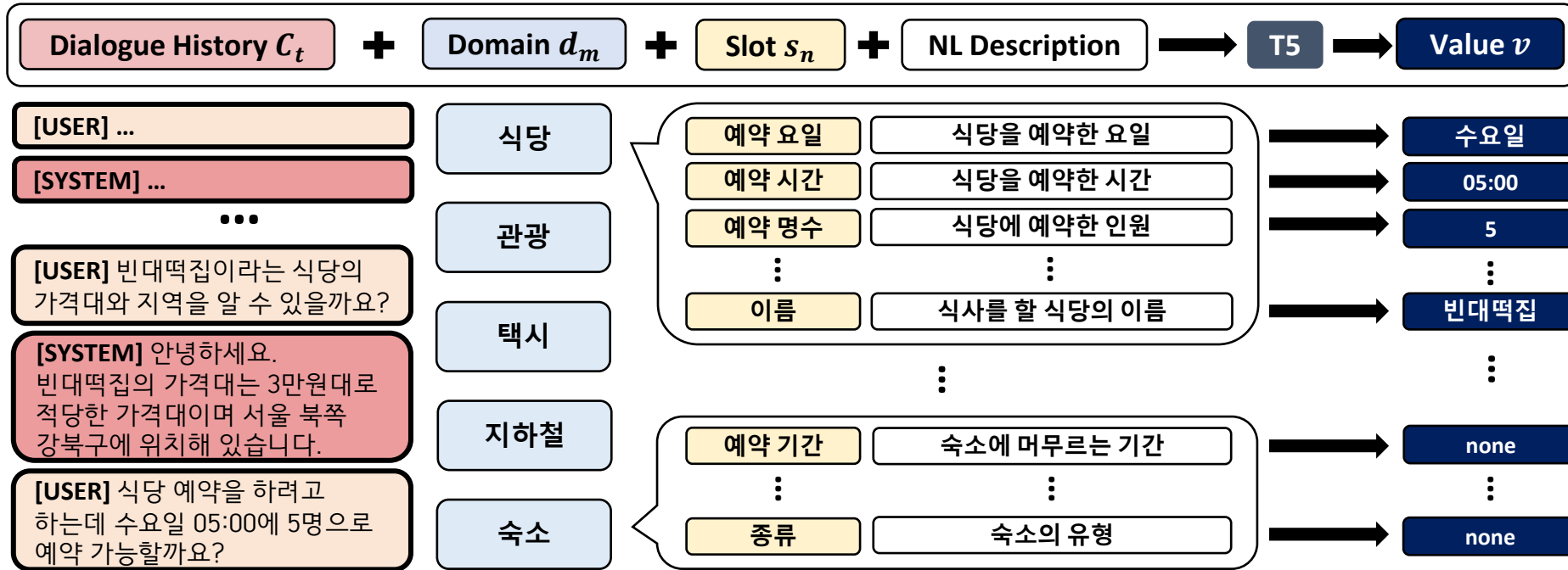


한국어 대화 상태 추적 태스크로 확장 및 적용

2. 제안 모델



2. 제안 모델



(현재 대화 상태가 아닌 경우, $V_t(d_m, s_n) = \emptyset$
즉, "none" 할당)

3. 실험

(1) 데이터셋

데이터셋 : WoS-v1.1 (KLUE-benchmark)

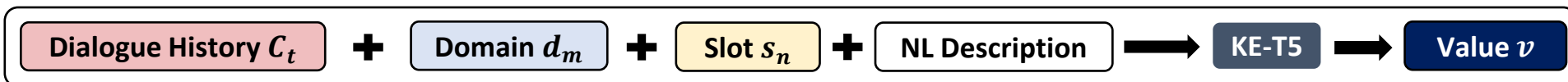
- 한국어 대화상태추적 데이터셋
- 5개의 도메인에 걸친 총 10,000개의 대화로 구성 (숙소, 식당, 관광, 택시, 지하철)
- 각 대화는 평균 14개의 발화로 구성

표 1: WoS-v1.1 데이터셋 구성

Dialogues	Single Domain	Multi Domain	Total
Train set	1,806	6,194	8,000
Dev set	263	737	1,000
Test set	226	774	1,000
Total set	2,295	7,705	10,000

3. 실험

(2) 실험 구성 및 평가 방법



실험 구성

- 언어모델 : KE-T5 (small, base)
- Prompt 기반 fine-tuning
- Epoch = 3

평가 방법

- Joint Goal Accuracy(JGA)
- Slot F1 score

(예시) (식당 이름, 식당 주소)
올바른 예측 : (제주흑돼지, 애월)
모델의 예측 : (제주흑돼지, 구좌)
→ JGA = 0, Slot F1 score = 0.5

4. 실험 결과

Model	JGA(%)	Slot F1(%)
KLUE-BERT-base	46.64	91.61
KLUE-RoBERTa-small	46.62	91.44
KLUE-RoBERTa-base	47.49	91.64
KLUE-RoBERTa-large	50.22	92.23
KR-BERT-base	45.33	90.70
koELECTRA-base	41.58	89.60
mBERT-base	35.46	88.63
XLM-R-base	39.82	89.61
XLM-R-large	41.20	89.80
KE-T5-SMALL w/o desc	36.68	84.14
KE-T5-SMALL w. desc	38.65	85.81
KE-T5-BASE w/o desc	65.09	95.61
KE-T5-BASE w. desc	66.14	95.67

KLUE WoS-v1.1
Baseline Scores¹

Our result

w. desc : 자연어 설명과 함께 학습

w/o desc : 자연어 설명을 포함하지 않고 학습

KLUE-RoBERTa-large

비교 모델 중 가장 높은 성능
50.22 % JGA, 92.23 % F1 score

KE-T5-SMALL

w/o desc, w. desc 모두 다른 비교 모델들에 비해
좋은 성능을 보이지 못함

KE-T5-BASE

- w/o desc : 65.09 % JGA, 95.61 % F1 score
- w. desc : 66.14 % JGA, 95.67 % F1 score

→ 두 모델 모두 KLUE-RoBERTa-large 보다 좋은 성능을 보임

¹ <https://github.com/KLUE-benchmark/KLUE>

5. 결론

실험 결과

KE-T5-SMALL

w/o desc, w. desc 모두 다른 비교 모델들에 비해 좋은 성능을 보이지 못했지만, w. desc 가 w/o desc 보다 더 나은 성능을 보임

KE-T5-BASE

w/o desc, w. desc 모두 좋은 성능을 보였으며, 특히 w. desc 는 가장 좋은 성능을 보임

결론

1. 스키마를 기반으로 한 프롬프트 방식의 입력은 언어 모델의 예측에 있어 효과적이다.
2. 슬롯에 대한 자연어 설명이 대화 상태 추적 태스크에 효과적으로 활용될 수 있다.

감사합니다.
