A Corpus-free State2Seq User Simulator for Task-oriented Dialogue

Yutai Hou†, Meng Fang‡, Wanxiang Che†, Ting Liu†
†Harbin Institute of Technology, ‡ Tencent Robotics X

1. Introduction

- **Why User Simulator (US)?**

Training of multi-turn dialog agent faces unavoidable obstacle:
(1) It's hard to decide the best action for a single turn.
(2) Annotated dialogue corpora are often unavailable.
To addressing these, it's popular to train an agent with *RL algorithm* and *a user simulator* as training environment.

- **What are the challenges for User Simulator (US)?**

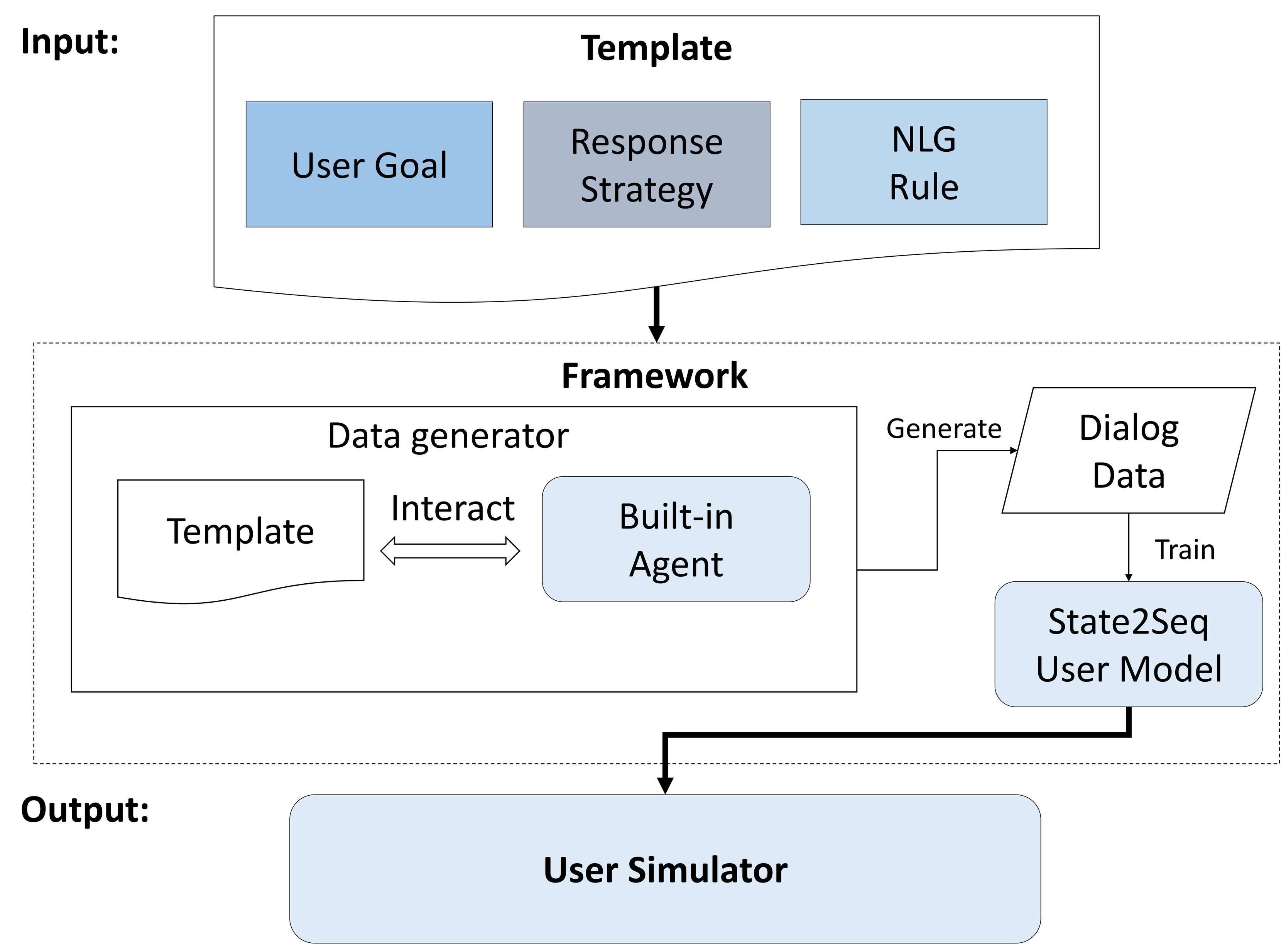
(1) Traditional US:
Based on rules and lack response diversity.
(2) Data-driven US:
Better in diversity but suffer from data scarcity.

- **How do we solve the challenges?**

(1) We design a new corpus-free framework that taking advantage of both two above user simulators' benefits.
(2) We propose a State2Seq model by introducing the attention mechanism and to model user behavior better.

2. Framework Overview

The framework builds a new State2Seq user simulator with no data but only a few templates.

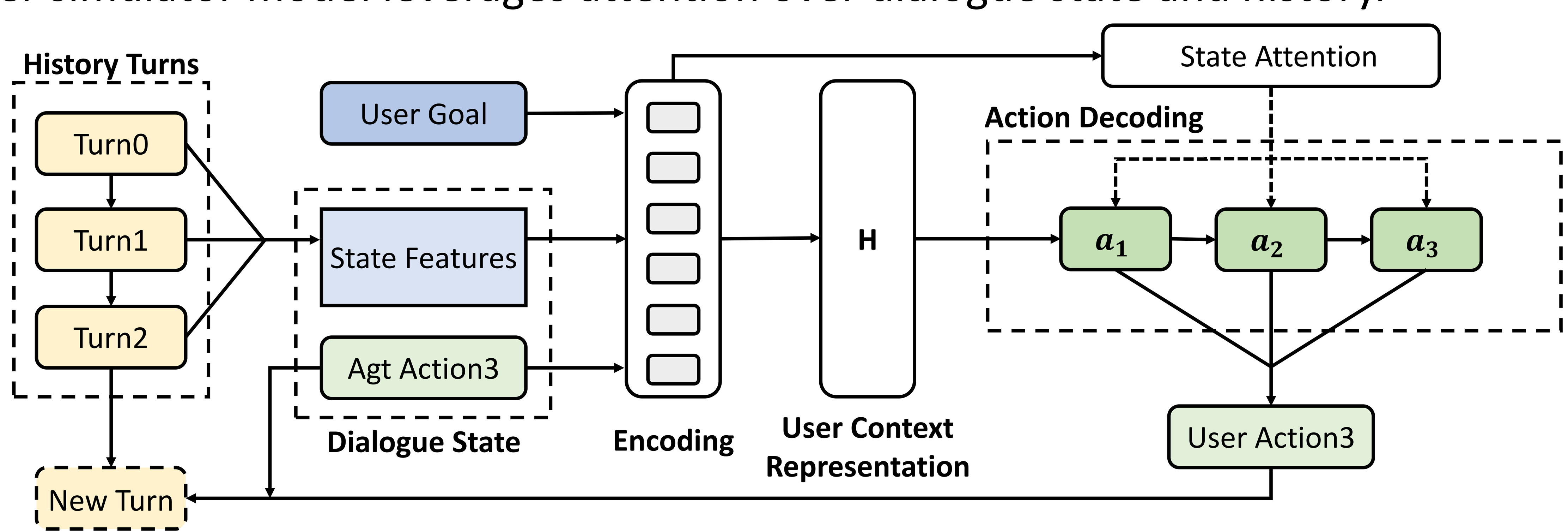


3. State2Seq User Simulator

To generate better response, our State2Seq user simulator model leverages attention over dialogue state and history.

Feature	Description
Constraint Status	whether user constraint slots have been informed by user
Request Status	whether user request slots have been satisfied by agent
Slot Consistency	whether the slot values provided by agent are consistent to user constraints
Dialog Status	Dialogue status of success, failed and no outcome yet

State Features



4 Experiment

- **Evaluation of trained agent policy.**

Model	Movie Booking			Restaurant Reservation		
	Avg. Succ.	Avg. Rwd.	Avg. Turns	Avg. Succ.	Avg. Rwd.	Avg. Turns
State2MLC	0.487	8.55	**21.82**	0.305	-17.22	29.69
State2Seq	**0.551**	**14.17**	25.85	**0.524**	**11.77**	24.21
Seq2Seq	0.412	-3.49	27.77	0.501	6.23	29.83
Seq2Seq-Att	0.430	-2.59	30.39	0.514	9.67	26.20
Agenda	0.438	-2.88	32.88	0.508	10.88	**22.17**

Evaluation of agent policy trained by different user model. Results above dash-line are from our model, which achieve best performance in most task.

Model	Avg. Succ.	Avg. Rwd.	Avg. Turns
State2Seq (Ours)	**0.778**	**53.88**	**11.88**
Agenda	0.571	22.29	14.57

Human evaluation of trained agent

Agents are likely to overfit to rule-based user simulator environment.

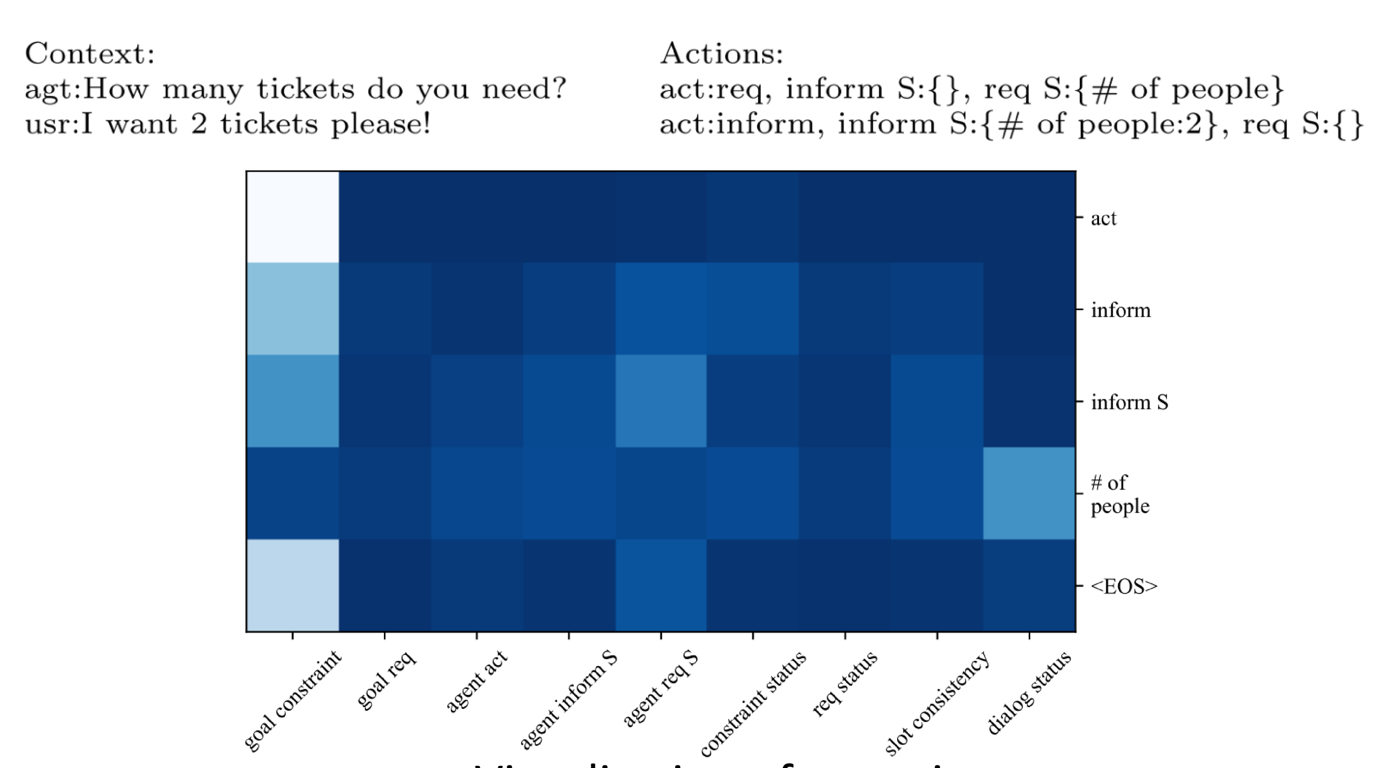
Model	Succ.	Avg. Rwd.	Avg. Turns
State2MLC (Ours)	0.628	26.19	16.36
State2Seq (Ours)	0.800	48.39	17.23
Seq2Seq	0.462	2.98	28.92
Seq2Seq-Att	0.480	2.11	32.98
Agenda	**0.814**	**50.36**	**15.66**

Analysis of agent models' overfitting to training environment.

- **Evaluation of trained agent policy.**

Model	Action Accuracy		Generalization Ability Test			
	Movie	Restaurant	Movie		Restaurant	
	F1	F1	Avg. Succ.	Avg. Rwd.	Avg. Succ.	Avg. Rwd.
State2MLC (Ours)	0.704	**0.695**	**0.442**	**6.06**	0.436	5.34
State2Seq (Ours)	**0.711**	0.683	0.400	5.51	**0.484**	**11.09**
Seq2Seq	0.692	0.662	0.063	-31.87	0.126	-31.87
Seq2Seq-Att	0.705	0.677	0.000	-46.99	0.000	-46.99
Agenda	N/A	N/A	0.392	0.04	0.410	2.20

Evaluation of user simulator model



Visualization of attention.

Get paper here