Universe Model: A Human-like User Simulator Based on Dialogue Context

Sihong Liu, Keqing He, Weiran Xu

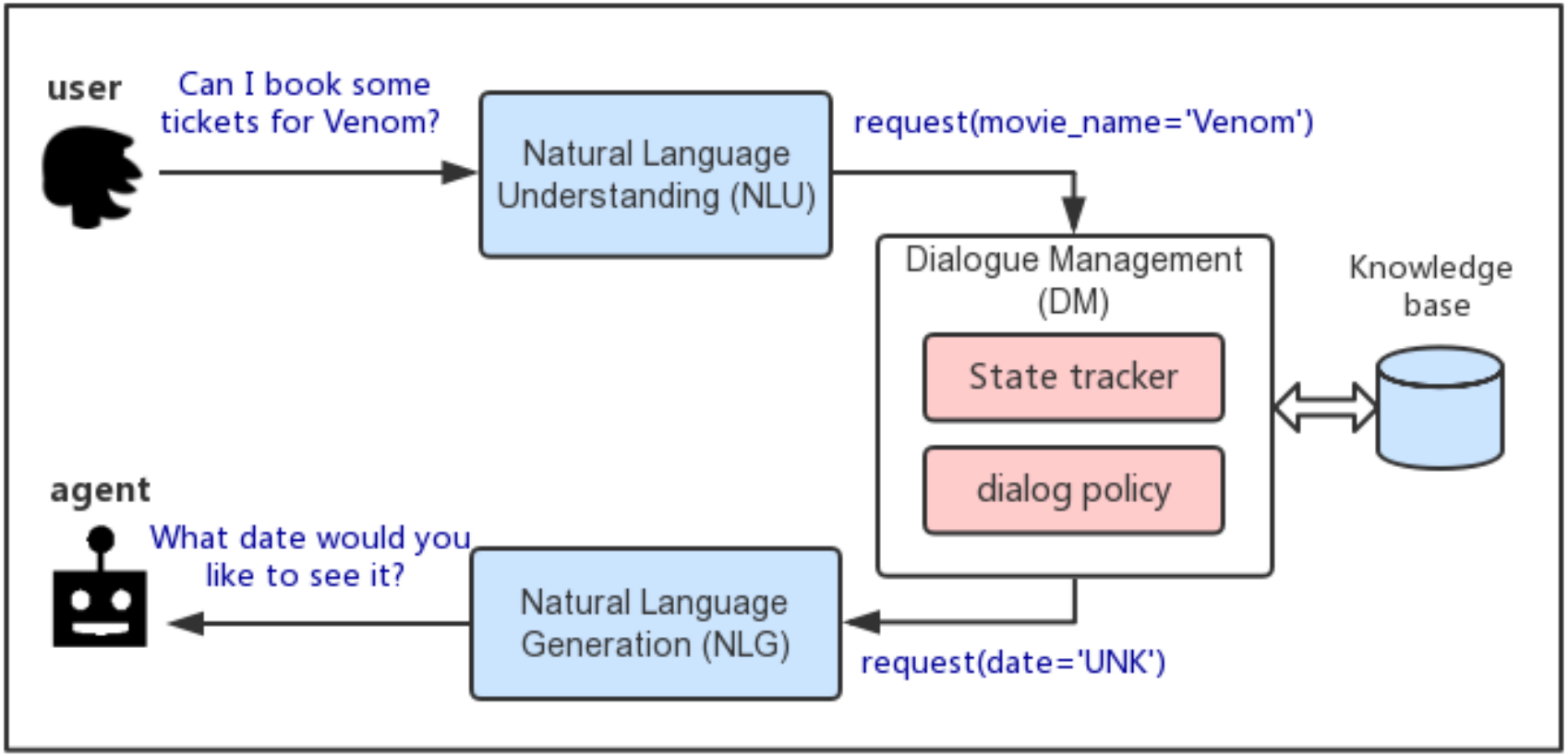
Beijing University of Posts and Telecommunications

Dec 18, SLT 2018

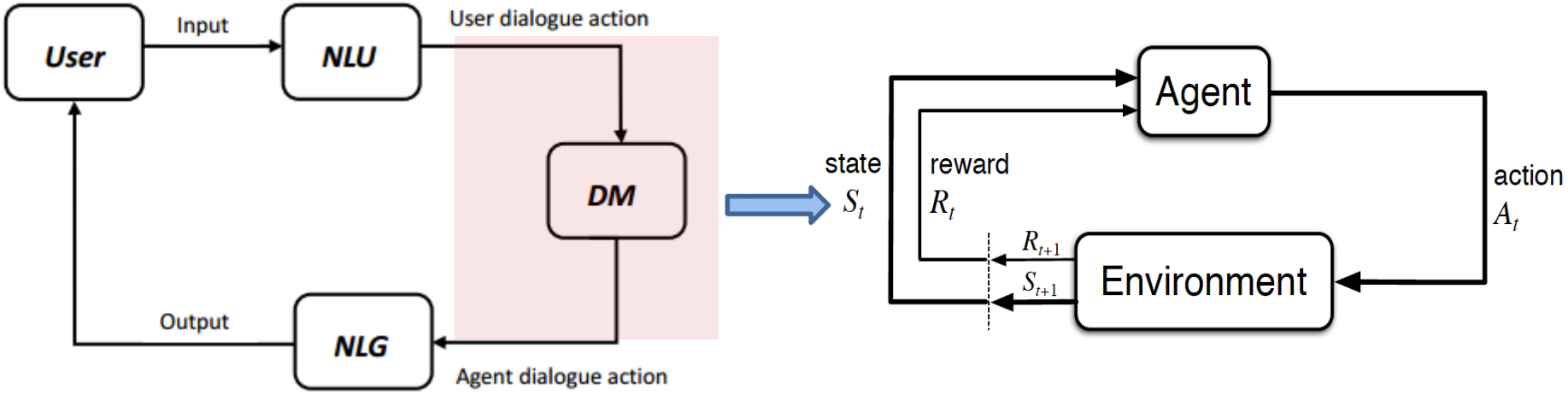
Outline

- Introduction
- Model
- Experiments
- Conculsion and future work

Task-oriented dialogue system



RL in dialog system



RL process : *state, action, reward*

- The agent starts in state S_t
- At step t the agent perform a_t, received reward r and reaches S_{t+1}
- The agent's goal is to maximize the expected accumulate reward.

User simulator

Challenge:

- RL require interaction with the environment.
- The collecting of real experience is expensive and time consuming.

Solution:

- User simulator [Schatzmann et al., 2007]

  Mimic what a real user does in a conversation

  Provide environment for RL agent

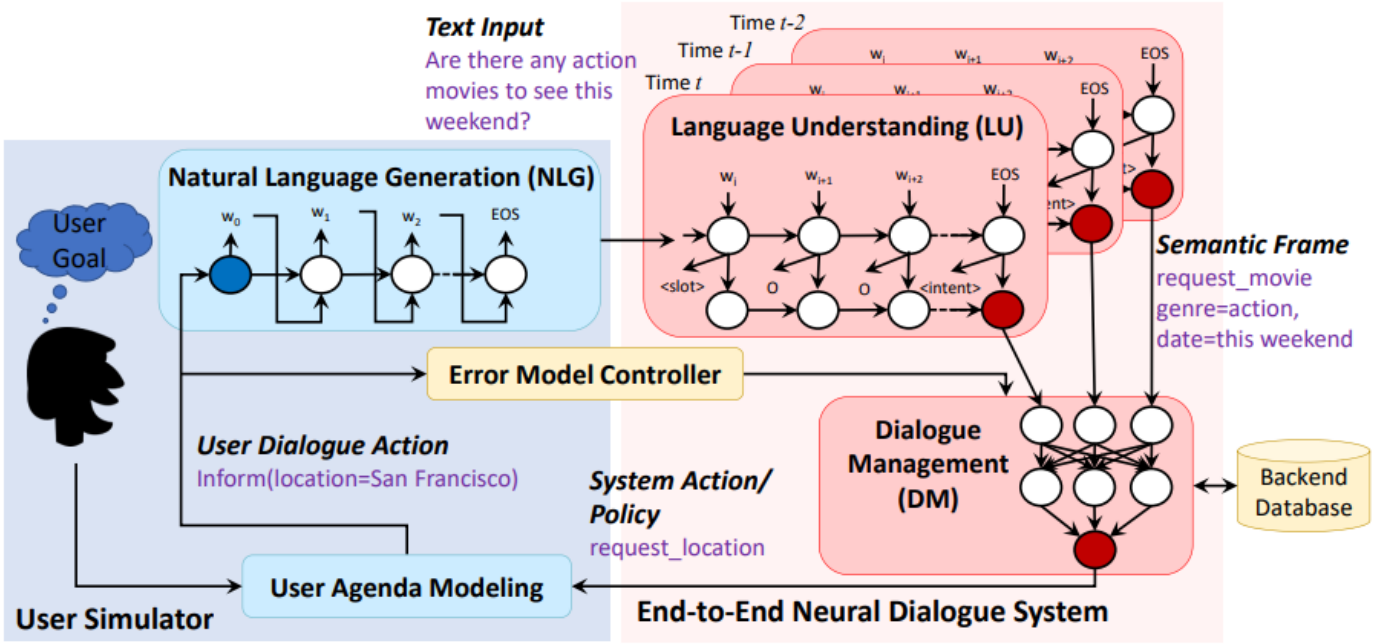


Figure 1 End-to-end neural dialogue system with a user simulator [Li+ 17]

User simulator: exsiting methods

Rule based

Pros

- Unlimited amount of training data for RL agent without any cost.

Cons

- Rough approximation of real users
- Real conversation is more complex

Model based

- Learn from the limited real experience: DDQ - world model [Peng et al.,2017]
- Insufficient context informations

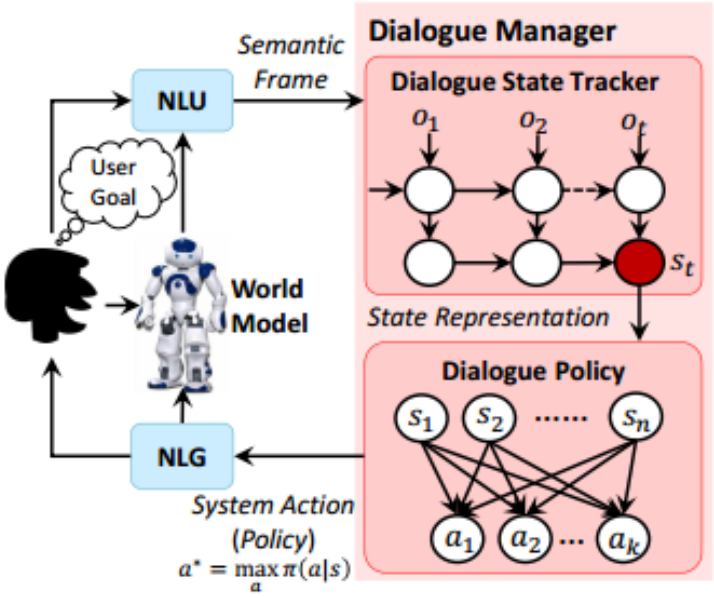
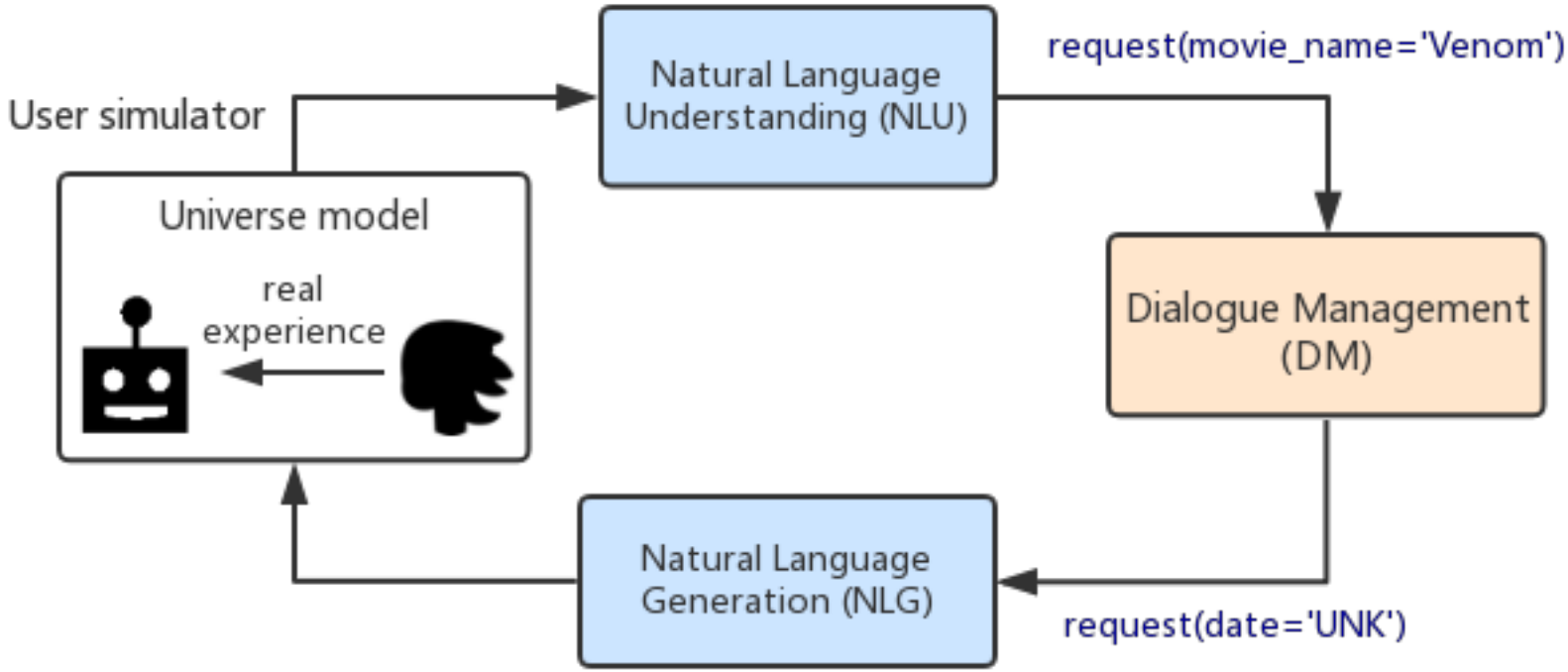
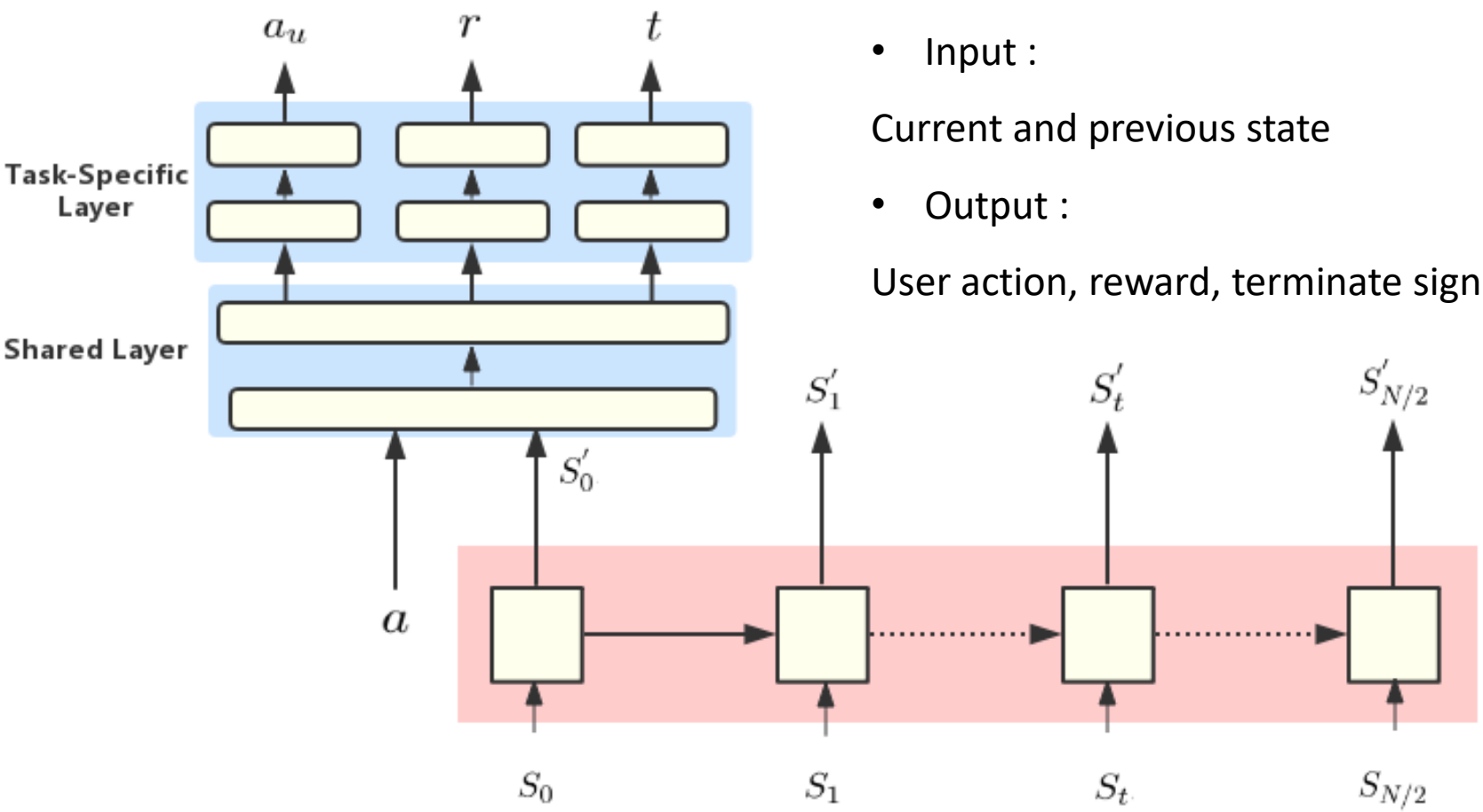


Figure 2: Illustration of the task-completion DDQ dialogue agent. [Peng et al.,2017]

Our proposed model : Universe model



The universe model architecture



RNN + Multi task learning

- Input :

Current and previous state

- Output :

User action, reward, terminate signal

Universe model

User:	Can I book some tickets for Venom?
Agent:	What date would you like to see it?
User:	I want to set it up tomorrow.
Agent:	Which theater would you like?
User:	I want to watch at regal meridian 16.
Agent:	What time would you like to see it?
User:	9:25 pm.
Agent:	How many tickets do you need?
User:	2 tickets please!
Agent:	Great - I was able to purchase 2 tickets for you to see Venom tomorrow at regal meridian 16 theater at 9:25 pm.

S_{t-1}

S_t

- Context information learning

$S_0, S_1, \dots S_{t-1}, S_t \dots, S_{N/2}$

- Multi task learning
  - Input: a, S'_t
  - Output: a_u, r, t

Experimental results

- The simulating performance of different user simulators

User simulators	Reward acc	Term acc	User action acc
Original world model from DDQ(5)	0.419	0.984	0.736
Transferred world model	0.934	0.983	0.891
Universe model	0.951	0.991	**0.95**

- The performance of different simulators on the same dialogue system

User simulator	Dialog avg reward	Dialog avg turns	Dialog success acc
Original world model from DDQ(5)	45.11	19.94	0.784
Transferred world model	53.93	11.73	0.823
Universe model	55.26	11.85	**0.851**

Conclusion and Future work

Universe model :

- Provide a more human-like environment and generate unlimited training data for RL-based dialogue system to interact with.

Future work:

- Narrow the gap between the human and user simulator to improve the human evaluation.

- And It's still a challenge for universe model to predict more complex dialogues.

Thank you!