[1] MOE KLINNS Lab, Xi’an Jiaotong University
[2] JIUTIAN Team, China Mobile Research

"Think Before You Speak": Improving Multi-Action Dialog Policy by Planning Single-Action Dialogs

Shuo Zhang [1]   Junzhou Zhao [1]   Pinghui Wang [1]
Yu Li [1]   Yi Huang [2]   Junlan Feng [2]

Paper Link

Contents

Contents

Task-Oriented Dialog System

Goal: order a hot Latte

***X*1**: I would like a cup of coffee.

***Y*1**: What coffee would you like?

***X*2**: What coffee do you serve?

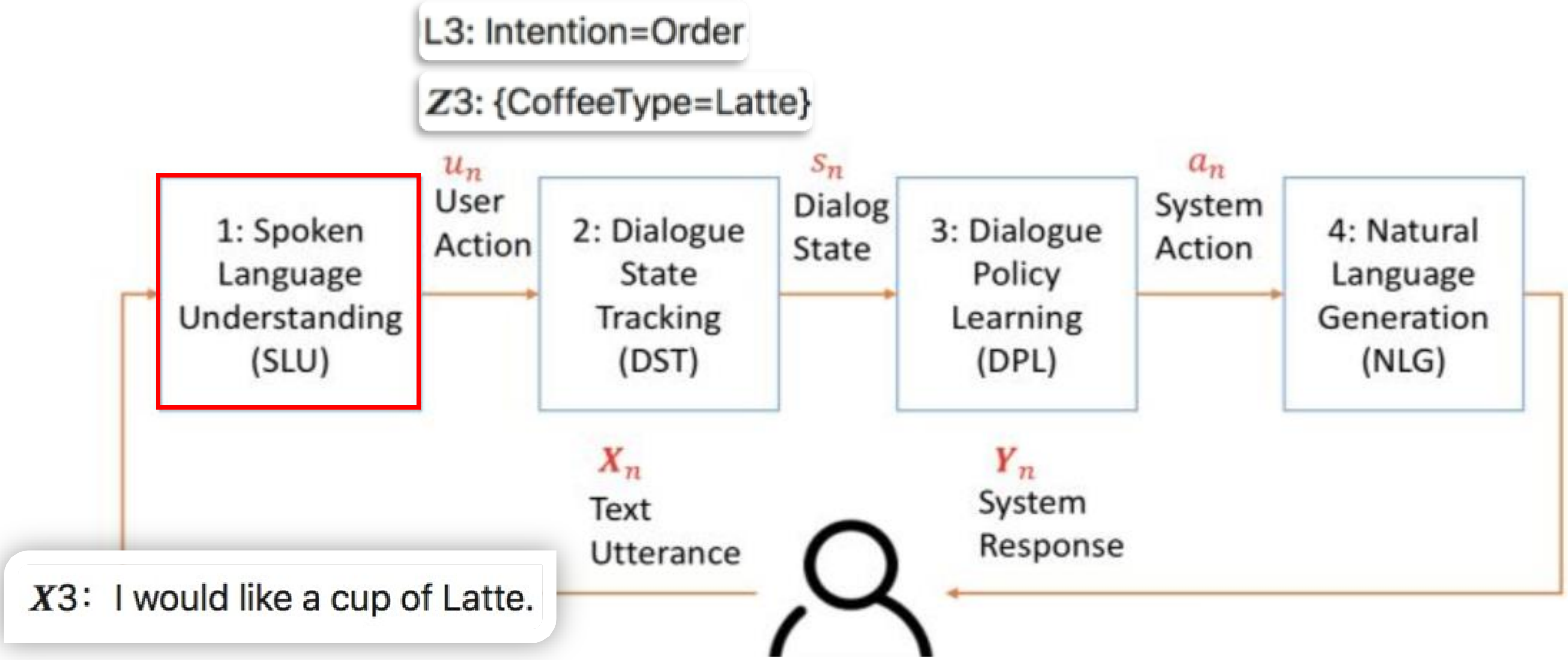
***Y*2**: We serve Espresso, Americano, Latte, Mocha, etc.

***X*3**: I would like a cup of Latte.

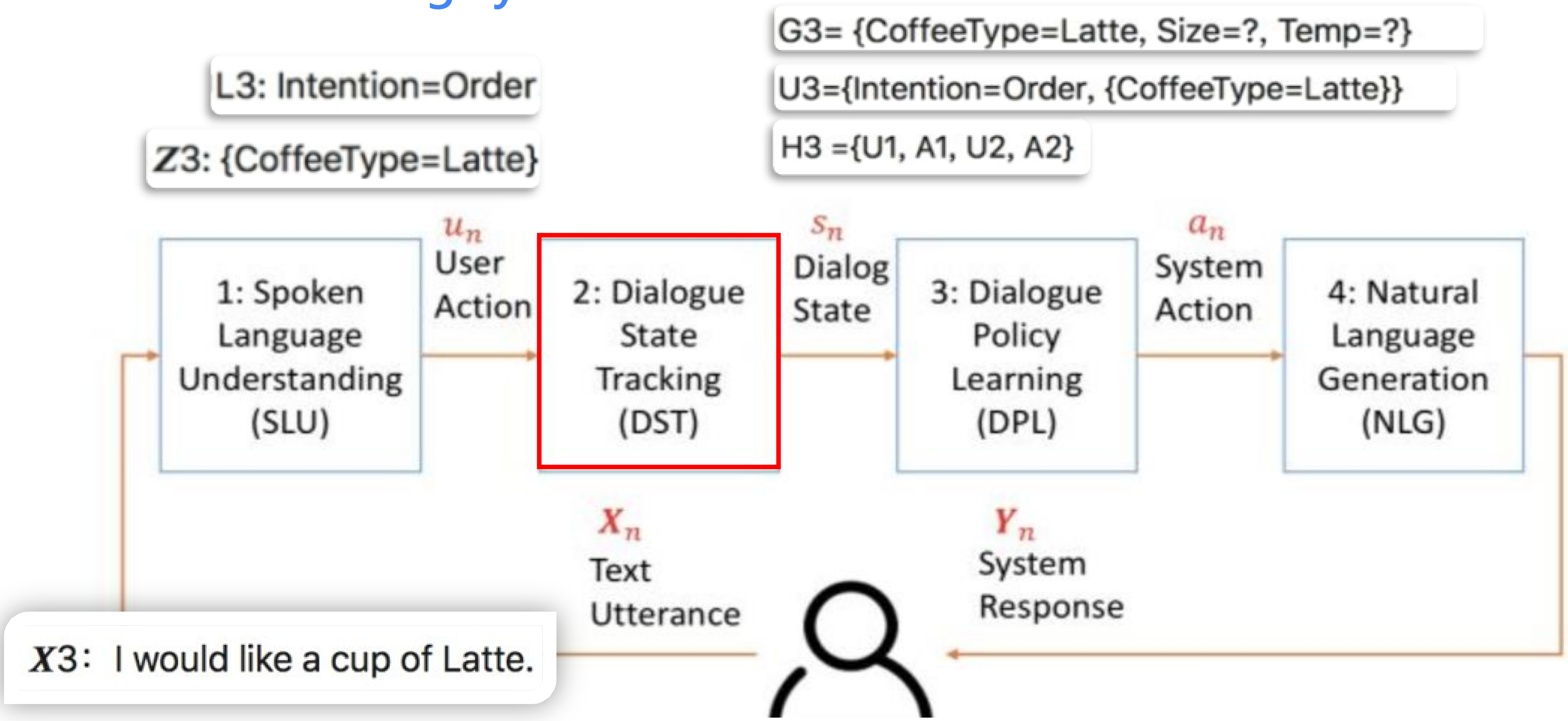
***Y*3**: Hot Latte or Iced Latte?

⋮

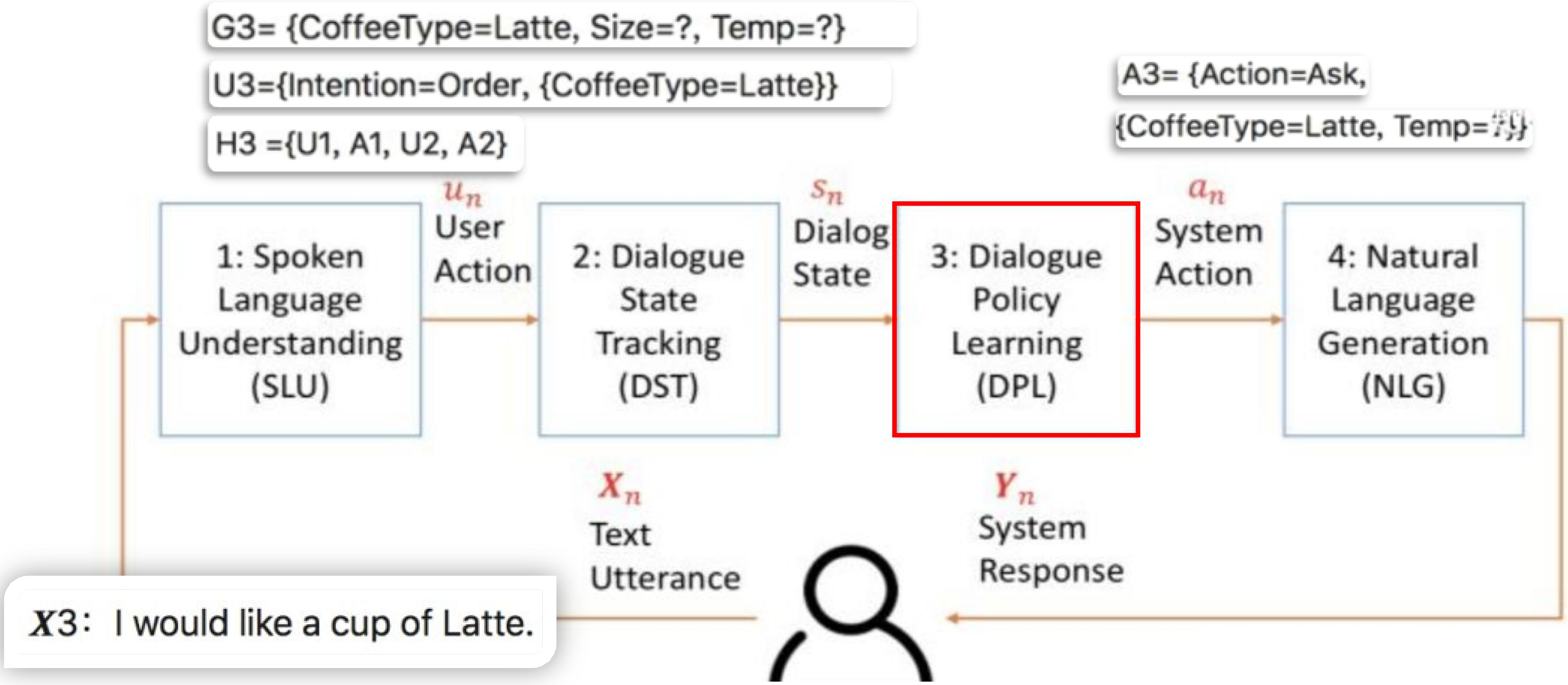
Task-Oriented Dialog System



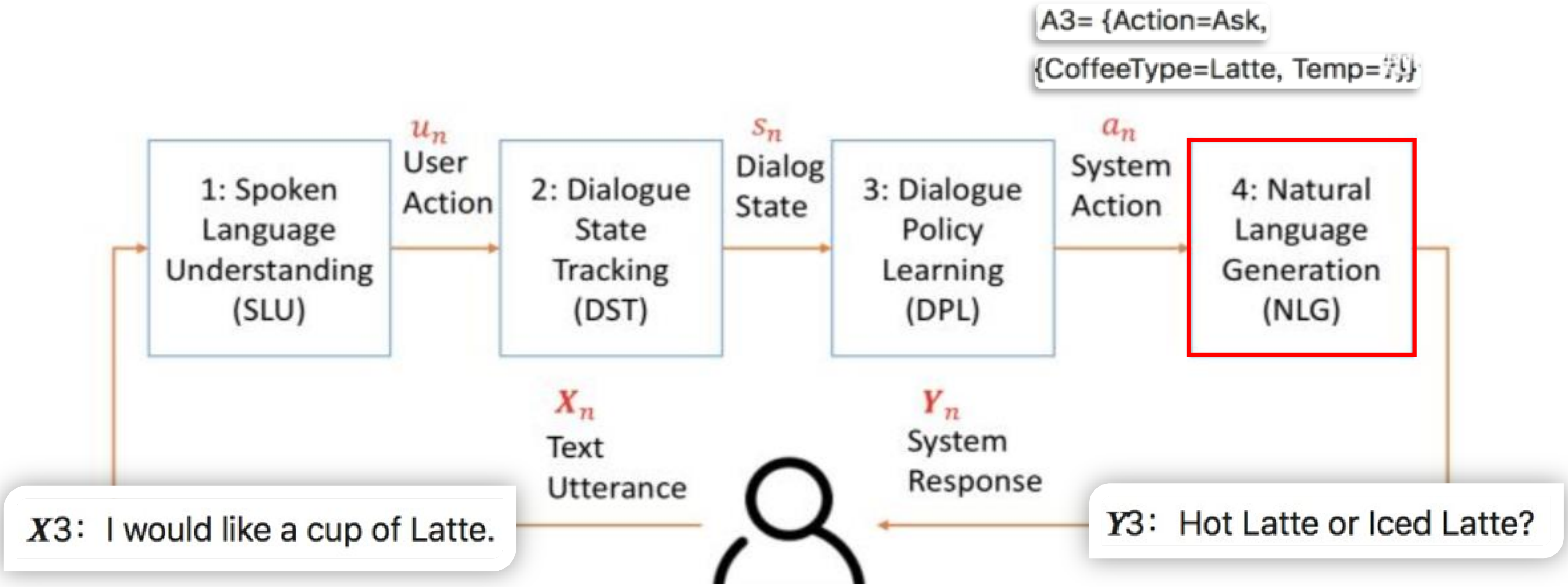
Task-Oriented Dialog System



Task-Oriented Dialog System

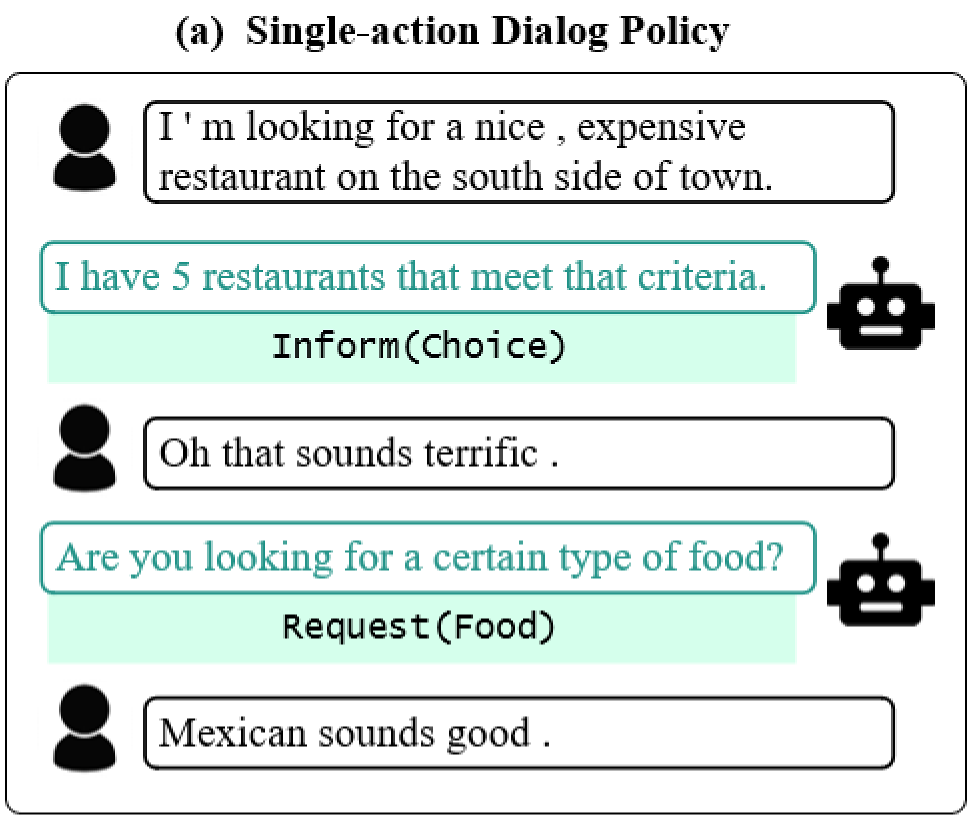


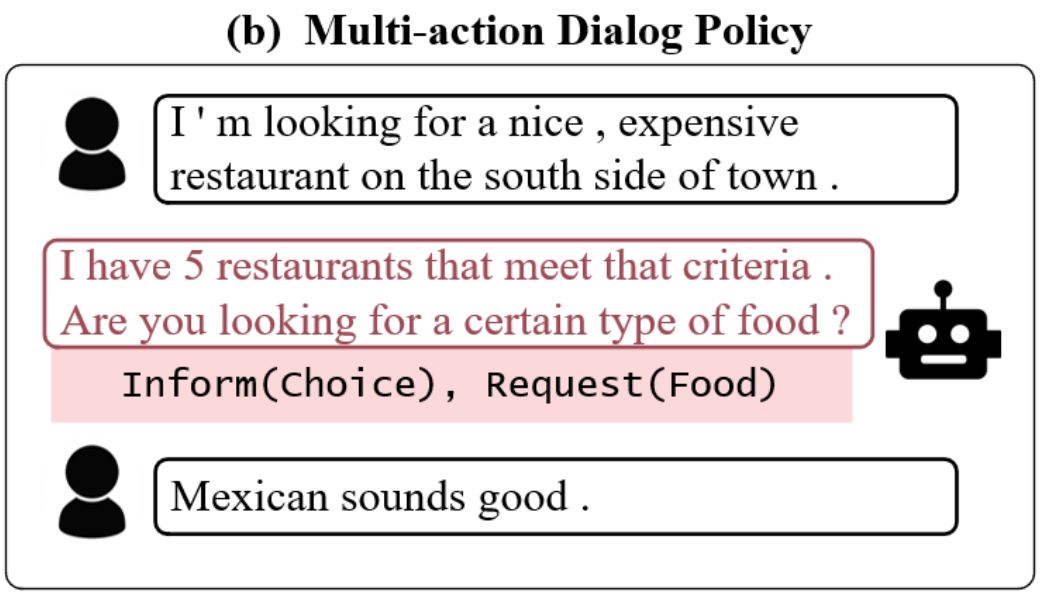
Task-Oriented Dialog System



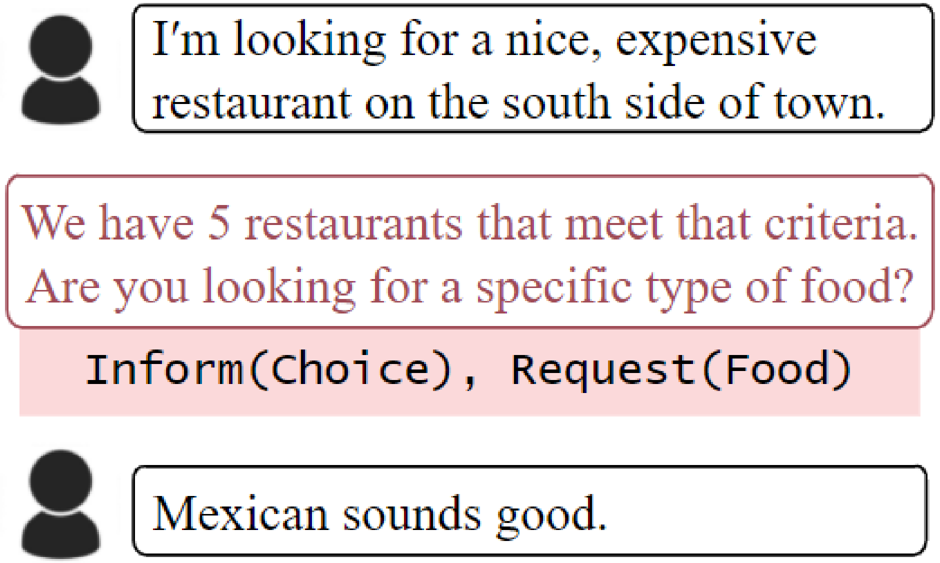
Contents

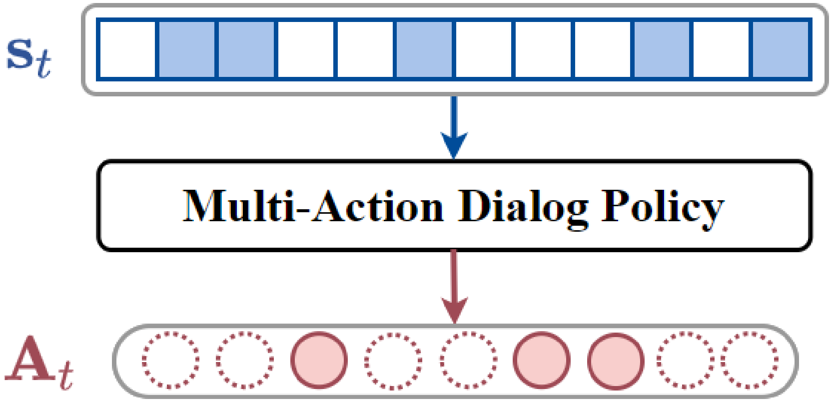
Multi-Action Dialog Policy Learning (MADPL)





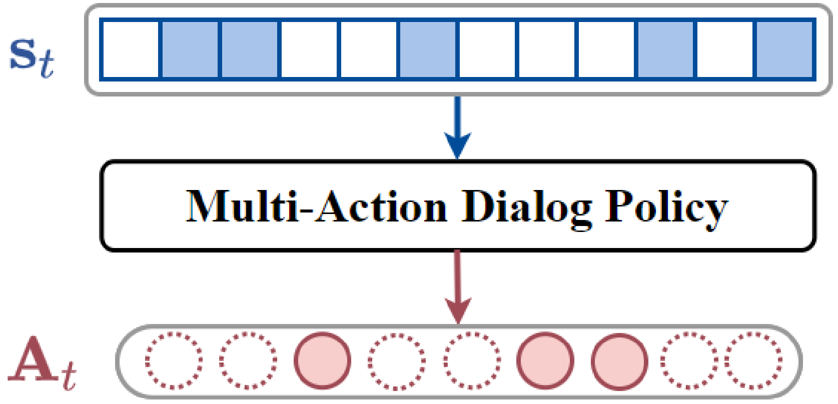
MADPL as Multi-Label Classification





Contents

Existing Methods for MADPL



Supervised Learning-based Imitation

Limited Data → Poor generalization ability

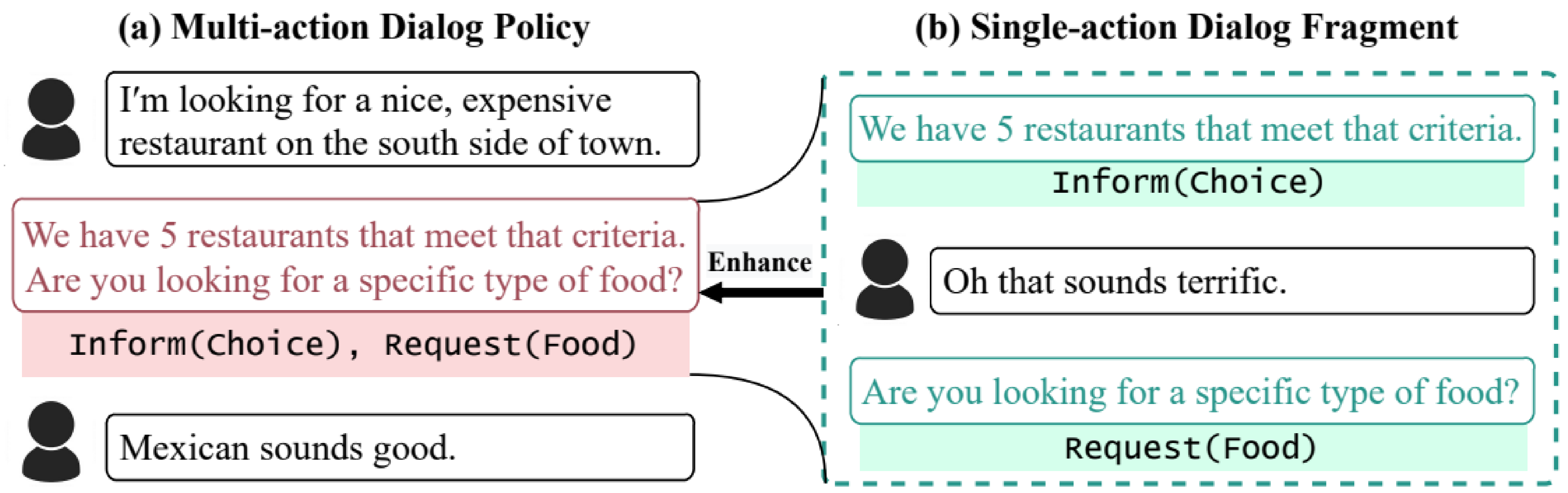
Reinforcement Learning / Adversarial Learning

Costly Real-world Environments & Unstable Training

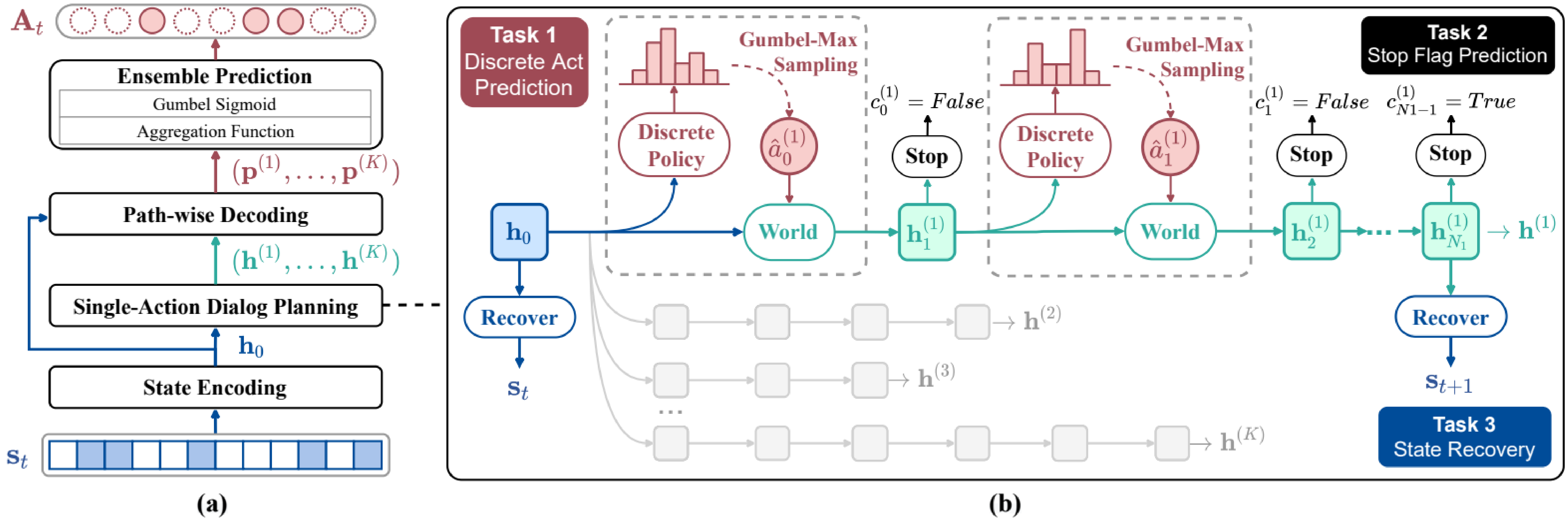
Interactive Learning

Costly Human Annotation

Multi-action Dialog *Turn* ~ Single-action Dialog *Fragment*



Planning-Enhanced Dialog Policy



Contents

Datasets

- MultiWOZ: Standard Benchmark. We use the agenda-based user simulator
- SGD: 2.3x Domains and 8.9x Slots compared to MultiWOZ. For Scaling concerns.

Evaluation Metrics

- Interactive Evaluation: Success Rate Inform Scores Match Rate Turn
- Standard Evaluation: Sample-wise Precision, Recall and F1

Baselines

- SL-based: DiaMultiClass DiaSeq DiaMultiDense gCAS
- RL-based: GP-MBCM ACER PPO
- AL-based: ALDM GDPL DiaAdv

Experiments

Interactive Evaluation

	MultiWOZ				
Agent	**Turn**	**Match**	**Rec**	**F1**	**Success**
DiaMultiClass	11.46 ±0.56	0.68 ±3.9%	0.81 ±3.2%	0.81 ±2.1%	67.3 ±3.69
+ sample	9.23 ±0.2	0.82 ±1.1%	0.90 ±1.8%	0.77 ±1.2%	81.4 ±1.78
DiaSeq (beam)	9.06 ±0.67	0.81 ±0.4%	0.9 ±1.2%	0.86 ±0.9%	81.4 ±0.16
greedy	10.35 ±0.04	0.68 ±1.5%	0.80 ±0.5%	0.77 ±0.5%	67.7 ±1.02
+ sample	8.82 ±0.1	0.86 ±0.6%	0.93 ±0.4%	0.81 ±0.5%	86.9 ±0.49
DiaMultiDense	9.66 ±0.15	0.85 ±0.6%	0.94 ±0.4%	0.87 ±0.6%	86.3 ±0.64
- sample	12.75 ±0.77	0.61 ±6%	0.72 ±5.4%	0.80 ±2.3%	58.4 ±6.05
gCAS	11.69 ±0.53	0.56 ±1.4%	0.72 ±0.4%	0.76 ±1.4%	58.8 ±2.82
GP-MBCM[5]	**2.99**	0.44	-	0.19	28.9
ACER[5]	10.49	0.62	-	0.78	50.8
PPO [5]	15.56	0.60	0.72	0.77	57.4
ALDM[5]	12.47	0.69	-	0.81	61.2
GDPL	7.54 ±0.43	0.84 ±0.9%	0.89 ±2.2%	**0.88** ±1.2%	83.2 ±1.48
DiaAdv	8.90 ±0.18	0.87 ±0.9%	0.94 ±0.75%	0.85 ±0.58%	87.6 ±0.9
- sample	11.9 ±0.88	0.62 ±5.9%	0.73 ±4.6%	0.80 ±2.1%	61.7 ±5.59
PEDP	8.69 ±0.15	**0.88** ±1.3%	**0.97** ±0.4%	0.87 ±1.1%	**90.6** ±0.68
- planning	9.66 ±0.15	0.85 ±0.6%	0.94 ±0.4%	0.87 ±0.6%	86.3 ±0.64
- ensemble	9.25 ±0.43	0.88 ±1.97%	0.96 ±0.8%	0.85 ±2.5%	89.1 ±1.74
- sample	8.85 ±0.22	0.82 ±2.5%	0.93 ±1.4%	0.86 ±1.6%	83.4 ±1.01

Standard Evaluation

	MultiWOZ			SGD (scaling)		
Agent	**F1%**	**Precision%**	**Recall%**	**F1%**	**Precision%**	**Recall%**
DiaMultiClass	39.41 ±1.08	54.59 ±1.71	34.32 ±1.32	58.09 ±0.63	81.29 ±1.13	46.29 ±0.57
+ sample	38.91 ±0.74	47.28 ±0.68	37.56 ±1.08	58.03 ±0.64	81.48 ±0.18	46.14 ±0.80
DiaSeq (beam)	44.64 ±2.08	51.91 ±0.99	43.66 ±2.27	63.13 ±0.18	86.04 ±0.5	50.83 ±0.30
greedy	48.34 ±0.45	54.71 ±0.21	48.84 ±0.84	63.21 ±0.35	86.31 ±0.7	50.85 ±0.40
+ sample	37.82 ±0.45	43.02 ±0.48	38.91 ±0.64	62.64 ±1.03	85.54 ±1.62	50.40 ±0.76
DiaMultiDense	35.92 ±0.54	51.93 ±0.33	30.10 ±0.69	57.85 ±0.68	80.64 ±0.43	46.21 ±0.89
- sample	34.35 ±0.62	52.14 ±0.19	27.74 ±0.74	56.69 ±0.62	79.54 ±0.88	45.19 ±0.75
gCAS	50.01 ±0.62	55.56 ±0.59	51.21 ±1.74	76.37 ±1.60	77.70 ±1.46	79.99 ±1.03
GDPL	31.89±0.96	50.14 ±0.79	24.99 ±1.14	-	-	-
+ sample	34.60 ±0.47	45.01 ±0.24	31.54 ±0.80	-	-	-
DiaAdv	40.97 ±0.95	53.44 ±0.50	36.84 ±1.30	-	-	-
- sample	41.71 ±0.47	56.46 ±0.45	36.28 ±1.48	-	-	-
PEDP	64.63 ±0.16	77.03 ±1.39	61.77 ±1.01	84.12 ±0.38	91.66 ±0.52	81.19 ±0.4
- planning	35.92 ±0.54	51.93 ±0.33	30.10 ±0.69	57.85 ±0.68	80.64 ±0.43	46.21 ±0.89
- ensemble	64.34 ±0.29	77.63 ±2.04	60.85 ±1.54	83.31 ±0.55	91.66 ±0.78	80.10 ±0.55
- sample	**66.95** ±0.45	**78.11** ±3.03	**65.02** ±1.22	**84.74** ±0.55	**92.07** ±0.97	**81.30** ±0.82

Human Evaluation

Dialog pair	Win	Lose	Tie	α
PEDP vs. DiaSeq	41.7	31.3	27.0	0.820
PEDP vs. DiaAdv	36.5	27.6	35.9	0.856
PEDP vs. GDPL	32.6	26.5	40.9	0.839

Hyper-parameter sensitivity

K	1	2	3	4	5	6
Success	89.1	89.3	90.6	89.7	90.2	89.9
Std. Dev.	1.74	1.19	0.68	0.73	0.62	0.57

Case Sample

GDPL

USER: I would prefer Chinese food please. I also need a place to dine that is expensive priced. I ' m so hungry - can you find me a place to eat in the Centre?

SYSTEM: I have 4 options for you!

USER: I would like to know the postcode of the restaurant, if possible?

SYSTEM: The ugly duckling is a nice place would you like to try that one? The restaurant postcode is cb21tw.

PEDP

USER: I would prefer Chinese food please. I also need a place to dine that is expensive priced. I ' m so hungry - can you find me a place to eat in the Centre?

SYSTEM: How about ugly duckling? It is Chinese food. It is located in the Centre. Would you like me to book it?

USER: What is their postcode?

SYSTEM: Their postcode is cb21nw. Is there anything else I can help you with?

Conclusion & Future work

- Multi-Action Dialog Turn ~ Single-Action Dialog Fragment
- Planning-enhanced Dialog Policy for Improved MADPL
- Improved Performance on MultiWOZ and SGD
- Knowledge-enhanced MADPL

Thanks

zs412082986@stu.xjtu.edu.cn

MOE KLINNS Lab, Xi'an Jiaotong University

JIUTIAN Team, China Mobile Research