

Unified Framework for Knowledge-Grounding Dialogue Modelling

报告人：陈志

导师：俞凯

2022.10.22



SJTU Cross Media
Language Intelligence Lab
上海交通大学跨媒体语言智能实验室



Background

EXAMPLE



What is the “dialogue”?

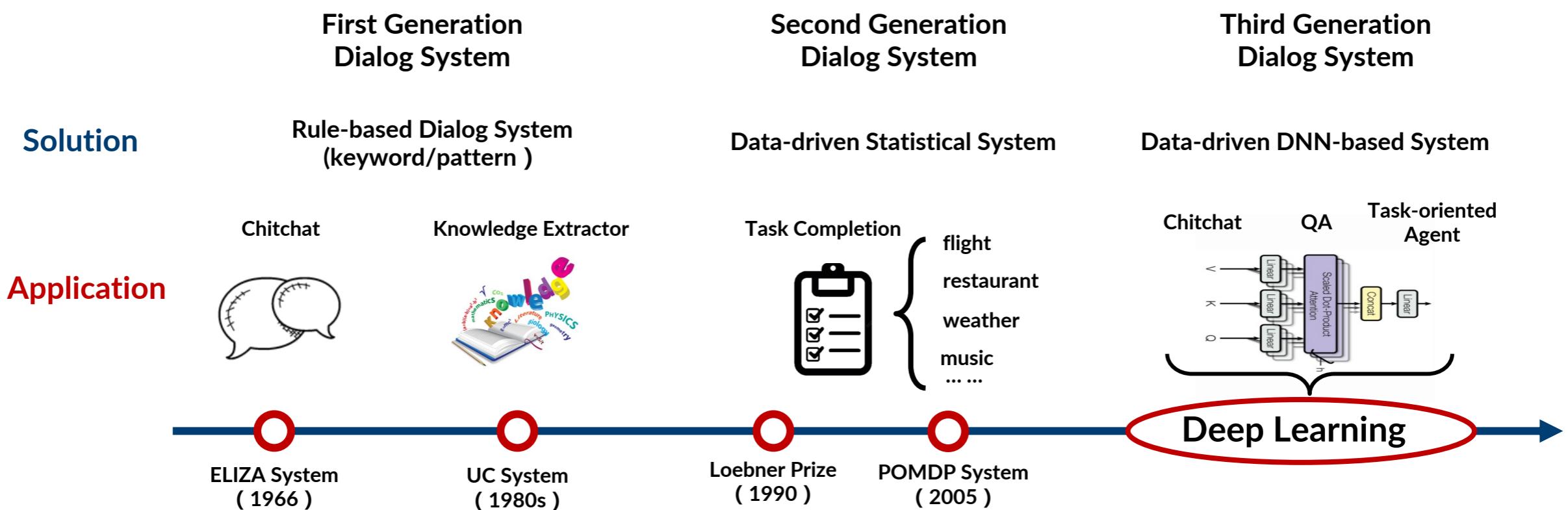
- The word “dialogue” comes from two Greek roots, **dia** and **logos**, suggesting “meaning flowing through.”

Dialogue as QA format to tell you “what is the dialogue?”

New Knowledge + 1

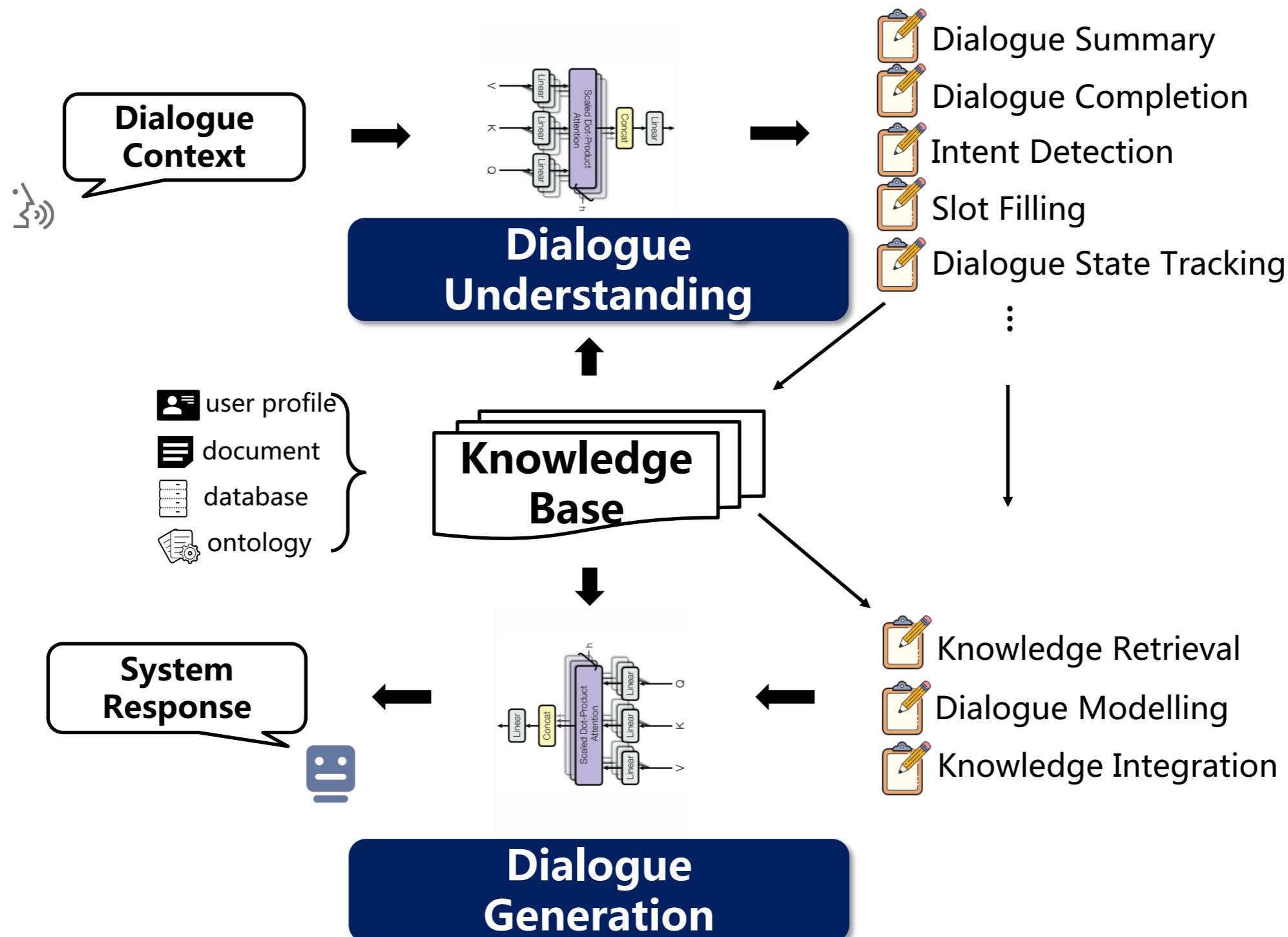


Background





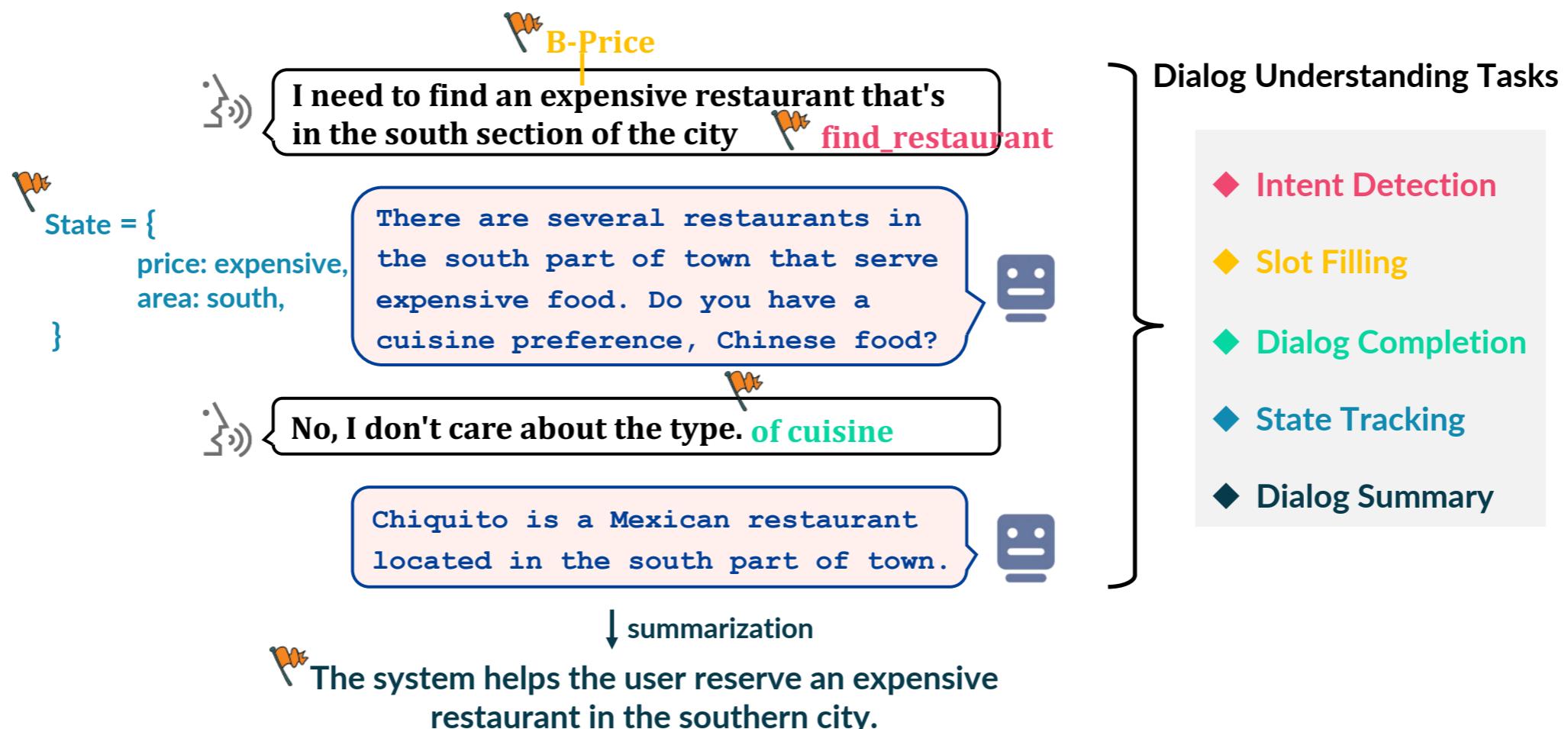
Background





Background

There are five typical dialog understanding tasks.

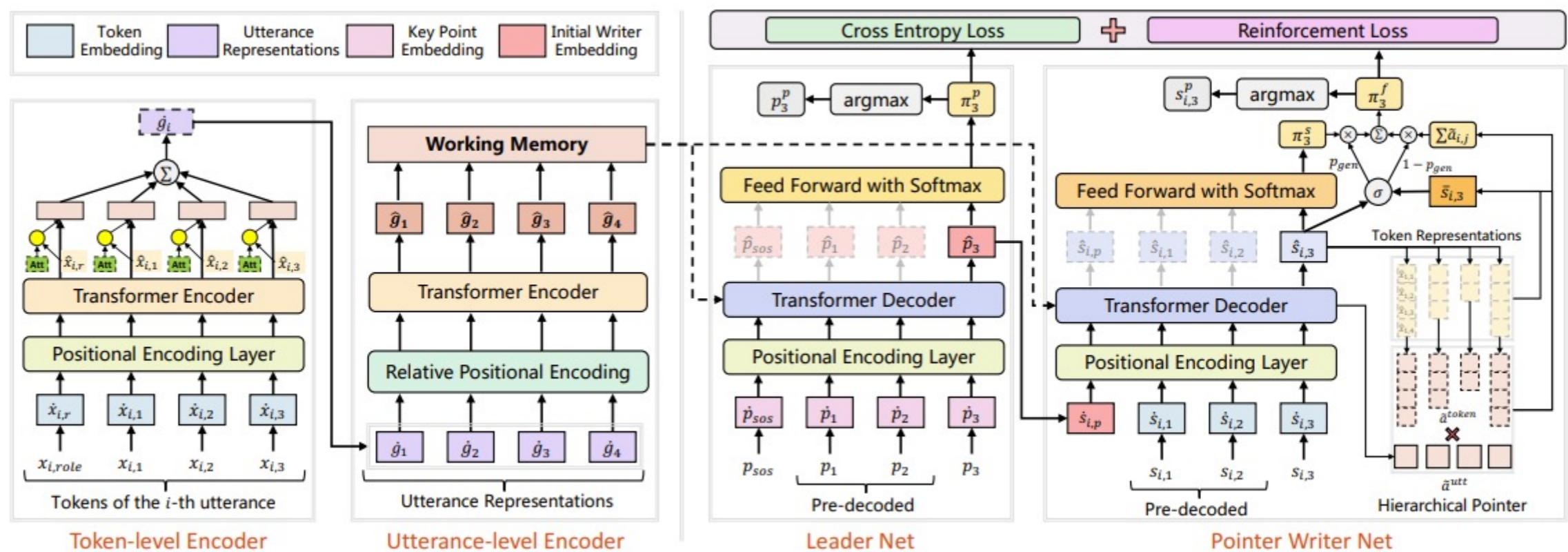




Background

- **Dialog Summary: Hierarchical Dialog Encoding Mechanism**

→ [summary] The system helps the user reserve an expensive ...



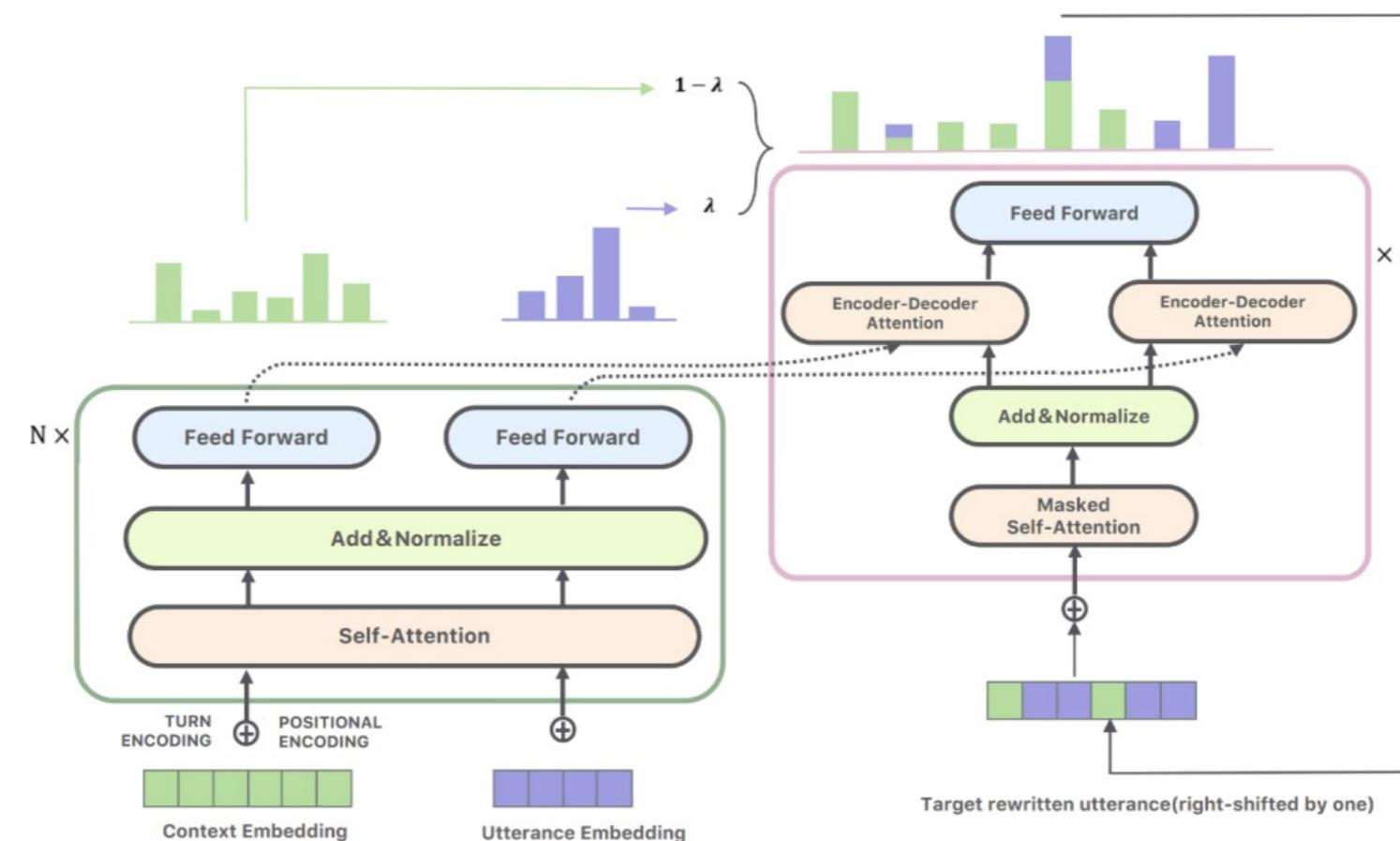
→ [turn-0] I need to find an expensive ...



Background

- Dialog Completion: two-phase copy mechanism

→ [completion] No, I don't care about the type of cuisine.



→ [context] I need to find an expensive ... [T-n] No, I don't care about the type.

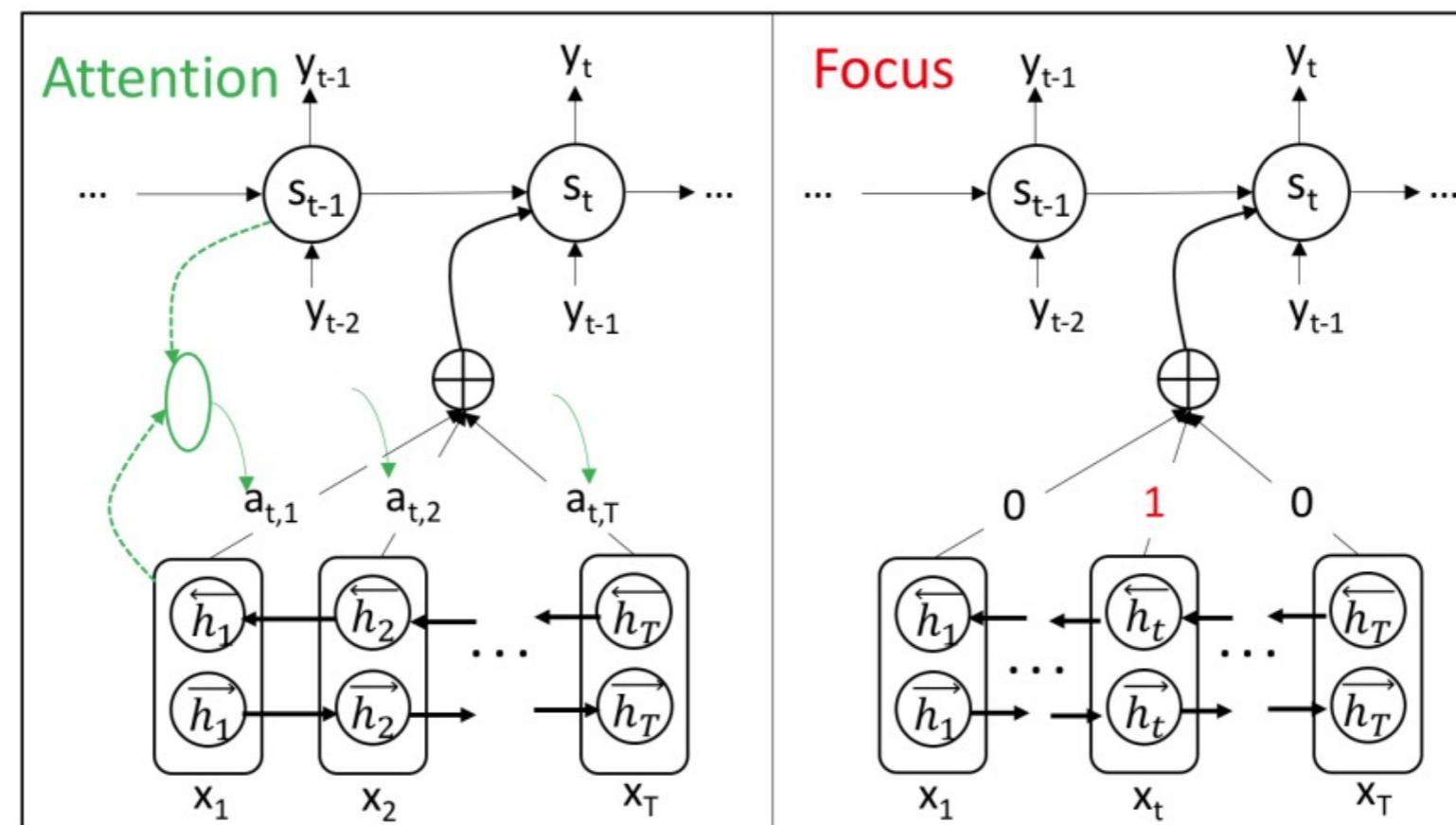


Background

- **Slot Filling: focus attention mechanism**

→ [filling] O O O B-Price O ...

→ [filling] O O O B-Price O ...

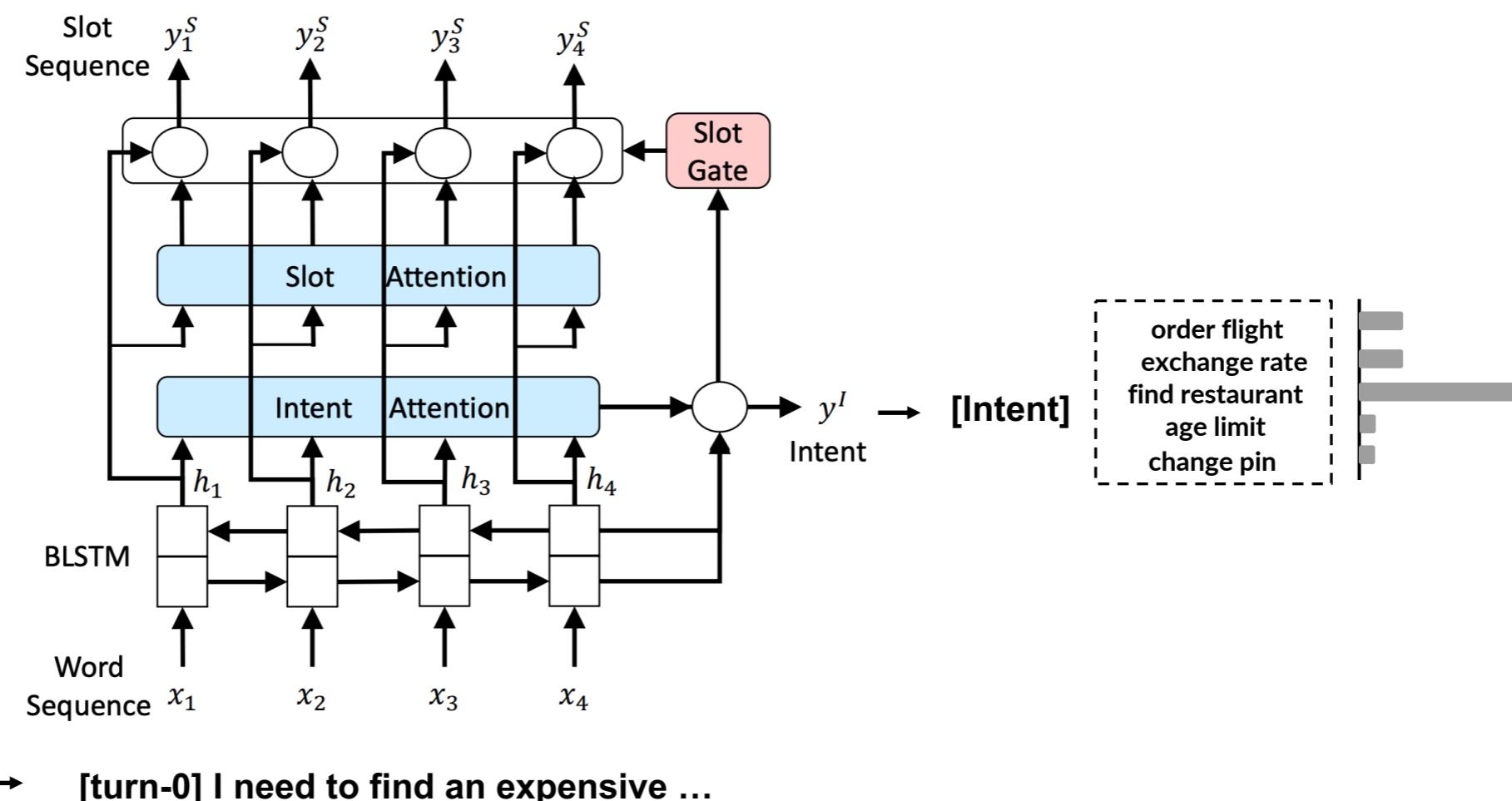


→ [turn-0] I need to find an expensive ... → [turn-0] I need to find an expensive ...



Background

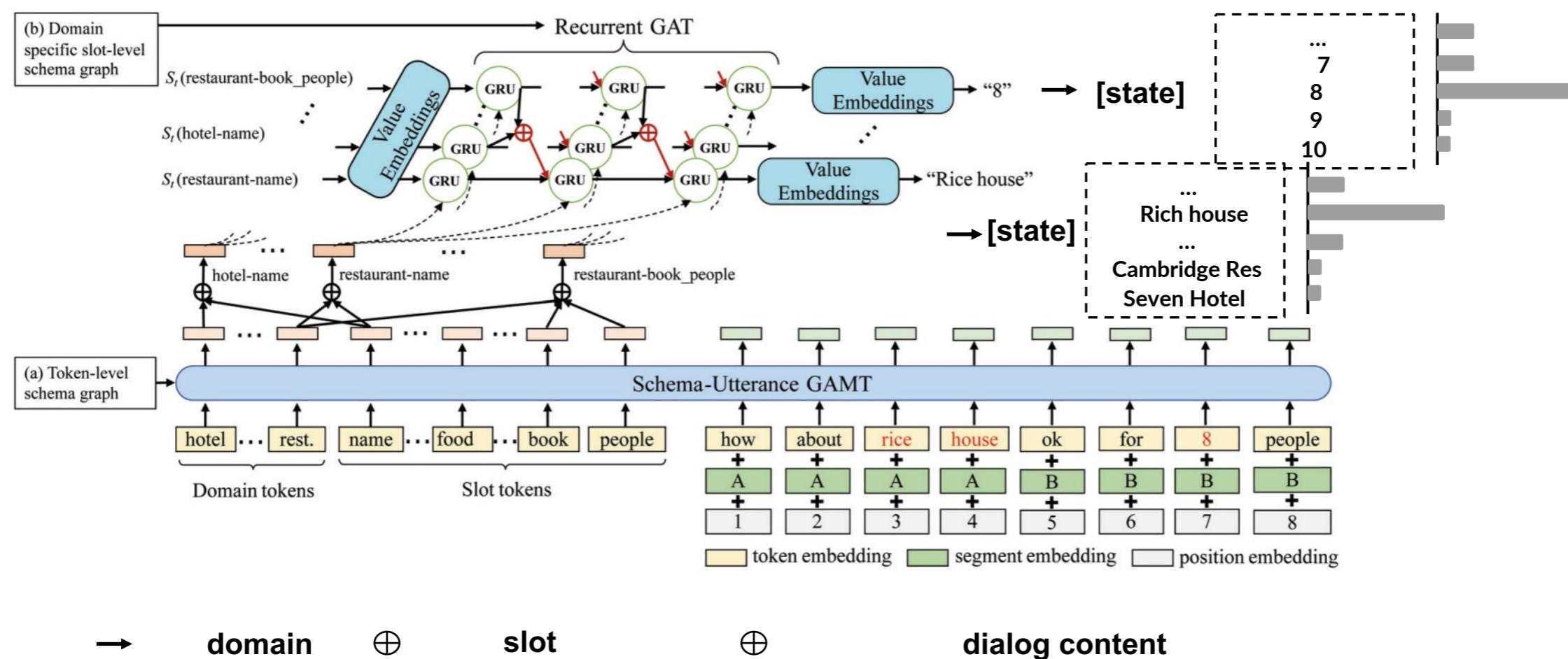
- Intent Detection: classification method





Background

- Dialog State Tracking: graph-based ontology modeling





Motivation



Can we model all the dialog understanding task in a unified framework?



Challenge



Dialog Corpus

- Domain Diversity
- Heterogeneous Scene
- Decentralized Annotation



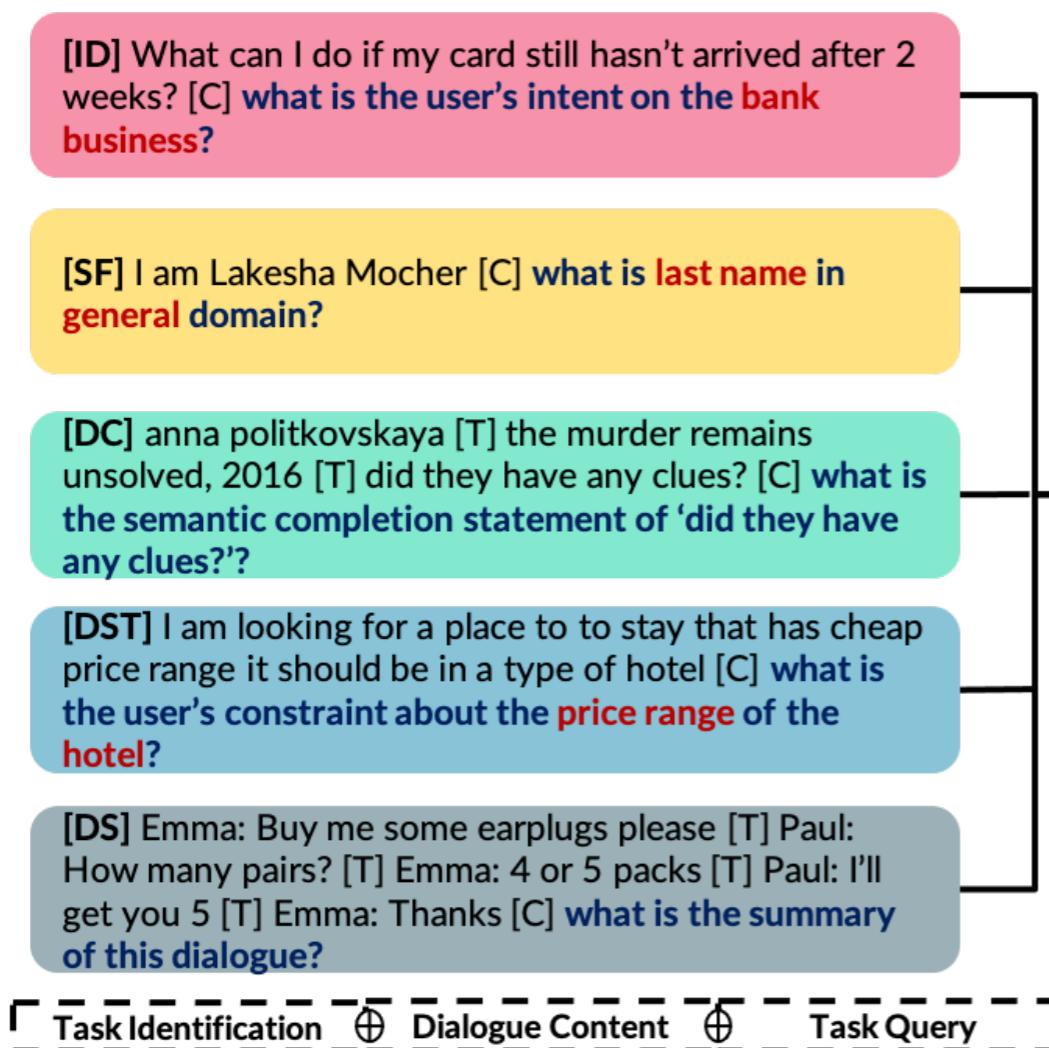
Dialog Task

- Various Definition
- Diversified Vocabulary
- Inconsistent Annotation

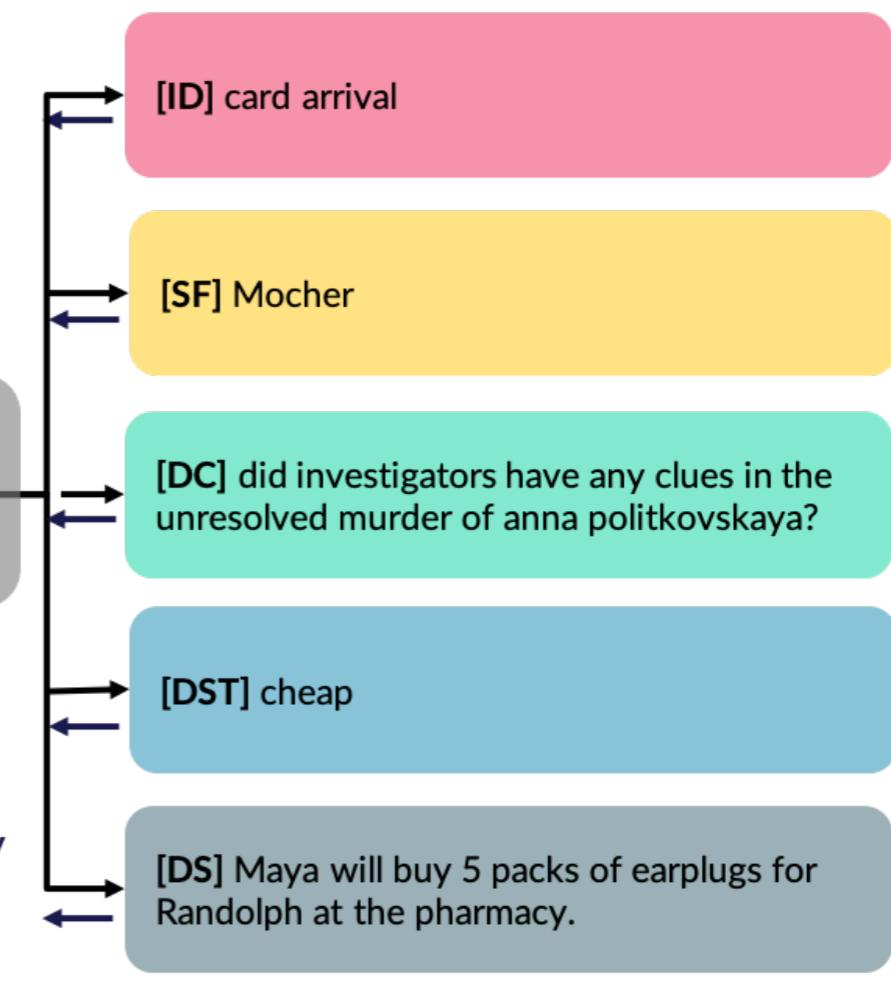


UniDU Framework

● Adaptive Learning Strategies for All Tasks



● Unified Model for All Data





UniDU Framework

- **Dialog Summary & Dialog Completion under UniDU**

For dialog summary:

Input ---> [DS] Emma: Buy me some earplugs please [T] Paul: How many pairs?
[T] Emma: 4 or 5 packs [T] Paul: I'll get you 5 [T] Emma: Thanks [C]
what is the summary of this dialogue?

output ---> [DS] Maya will buy 5 packs of earplugs for Randolph at the pharmacy.

For dialog completion:

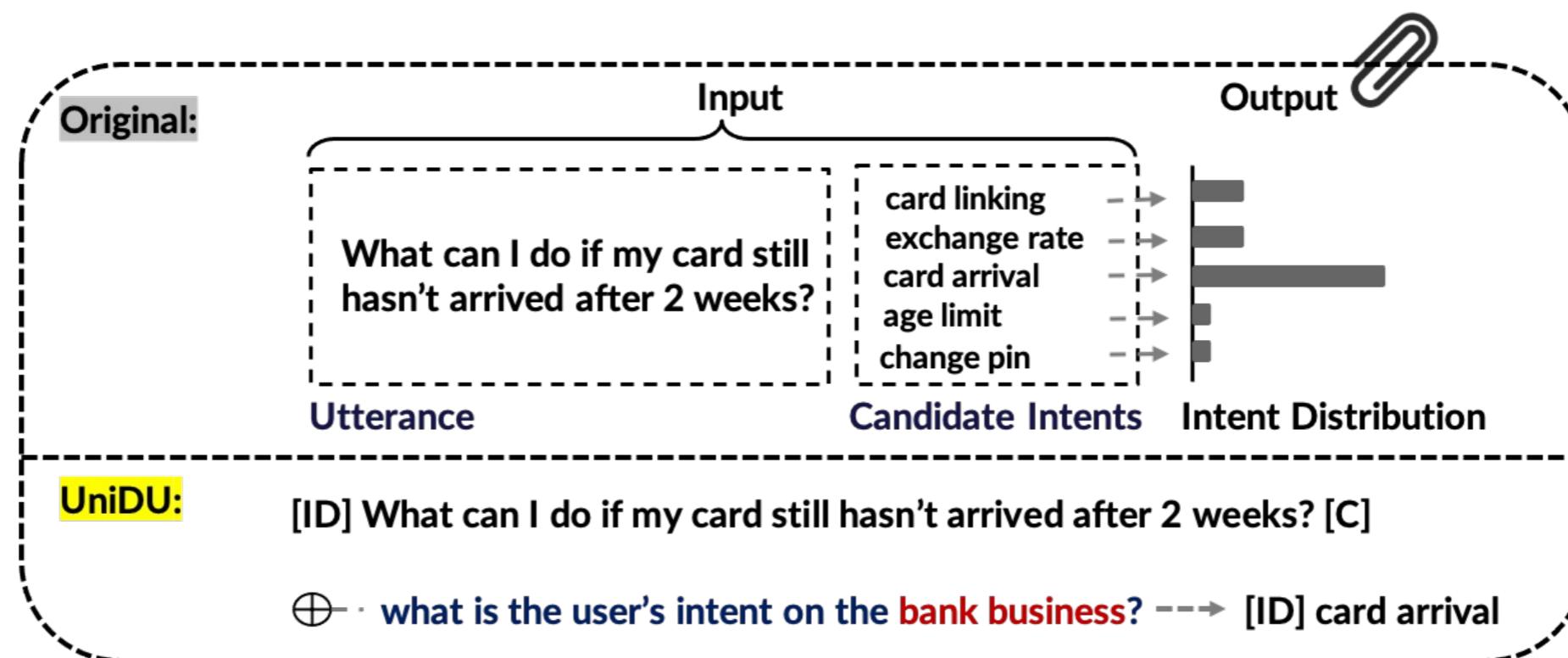
Input ---> [DC] anna politkovskaya [T] the murder remains unsolved, 2016
[T] did they have any clues? [C] **what is the semantic completion statement of 'did they have any clues'?**

output ---> [DC] did investigators have any clues in the unresolved murder of anna politkovskaya?



UniDU Framework

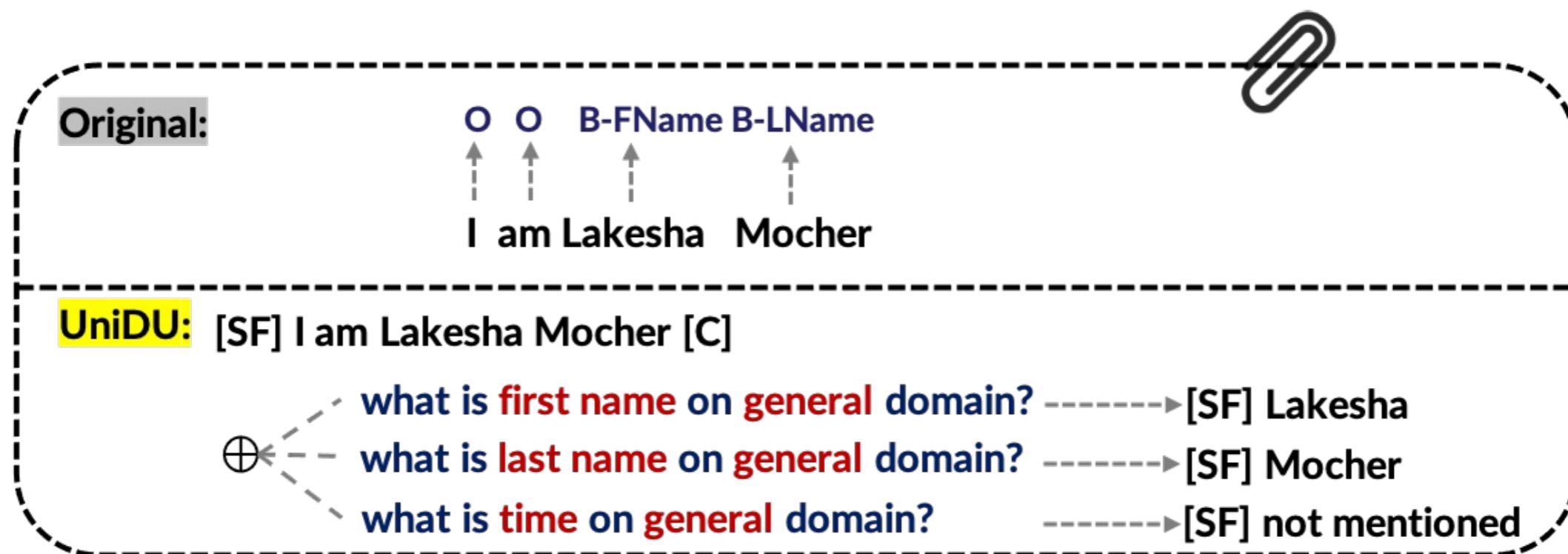
- Intent Detection under UniDU





UniDU Framework

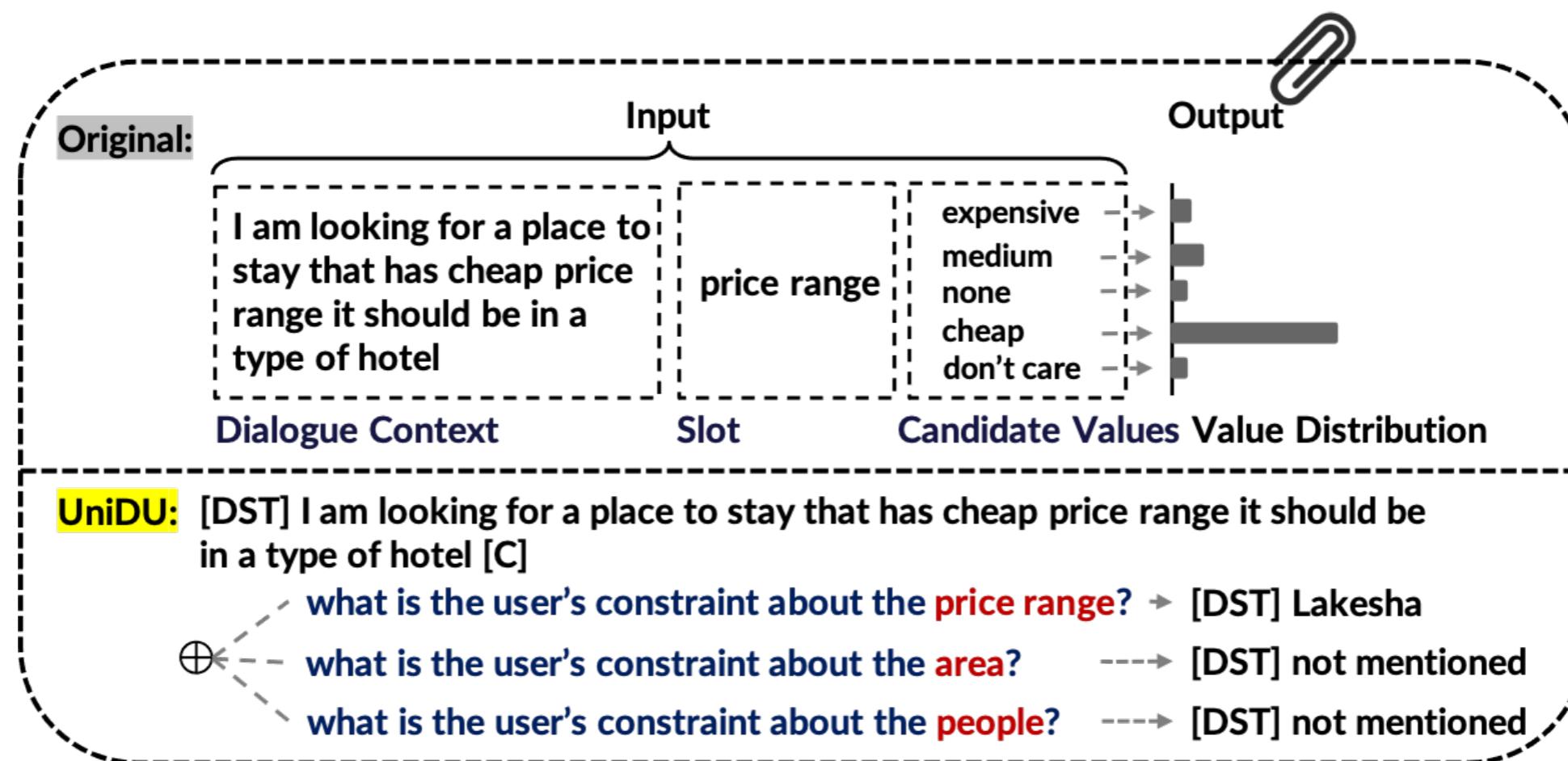
- Slot Filling under UniDU





UniDU Framework

- Dialog State Tracking under UniDU





Adaptive Learning

- **Adaptive Multitask Learning Strategies**

Domain-Transfer Multitask Learning

- **TT (Task Transfer Learning):** like data augmentation where the data is from another auxiliary task
- **MIX (Mixture Learning):** just mix all the available DU corpora together
- **G2S (from general to specific):** training route is from general tasks to domain-specific tasks
- **CL (Curriculum Learning):** training route is according to the input complexity: utterance, turn and dialog

Adaptive Multitask Learning:

- **WL (Weight Learnable Optimization):** losses from different tasks are automatically distributed



Experimental Results

- Datasets

Corpora	#Sample	I _(Token)	I _(Turn)	O _(Token)	Task
SAMS _{UM}	14732	104.95	11.16	20.31	DS
DIALOGS _{UM}	12460	140.48	9.49	22.86	DS
T _{ASK}	2205	34.92	2.75	10.84	DC
C _{ANARD}	31526	102.67	9.80	11.55	DC
B _{ANKING77}	12081	21.64	1	3.14	ID
H _{WU64}	25715	17.69	1	2.05	ID
R _{ESTAURANTS8K}	15270	14.44	1	3.38	SF
S _{NIPS}	35748	15.31	1	1.77	SF
W _{OZ2.0}	7608	78.96	4.63	1.30	DST
M _{ULTI} W _{OZ2.2}	35119	115.80	5.99	1.45	DST

Each task has two corpora.

- ◆ DS: Dialog Summary
- ◆ DC: Dialog Completion
- ◆ ID: Intent Detection
- ◆ SF: Slot Filling
- ◆ DST: Dialog State Tracking



Experimental Results

- Main Results

Methods	DS _(SAMSUM)		DC _(TASK)		ID _(BANKING77) ACC.	SF _(RESTAURANTS8K) F_1	DST _(WOZ2.0) JGA
	R-1	R-L	EM	BLEU			
Baselines	49.67	48.95	74.2	89.4	93.44	96.00	91.4
(Wu et al., 2021) (Chen et al., 2021b) (Mehri et al., 2020) (Coope et al., 2020) (Tian et al., 2021)							
Eight Training Strategies under UniDU Framework							
ST	49.74	47.10	76.4	89.0	91.49	95.76	89.8
TT	51.24	48.59	76.1	89.2	91.94	95.12	91.0
MIX	50.98	48.13	76.2	90.8	91.91	96.43	90.8
G2S	51.13	48.75	76.3	90.1	90.12	94.81	86.8
CL	51.04	48.36	77.2	89.8	92.17	96.02	90.8
GradNorm	51.33	48.69	77.4	90.4	92.07	96.69	90.5
HWU	50.31	47.69	76.2	90.4	93.14	97.43	91.9
MATS	50.53	47.97	76.6	90.6	93.60	97.61	92.3

Adaptive MTL



Experimental Results

• Ablation Study

◆ Effects of PLM

Backbone	DS (R-L)	DC (BLEU)	ID (ACC.)	SF (F_1)	DST (JGA)
around 100M					
Trans.-B	34.84	74.2	86.36	83.01	72.5
BART-B	47.97	90.6	93.60	97.61	92.3
T5-S	41.63	85.9	87.04	96.94	89.9
around 400M					
Trans.-L	34.10	67.4	86.46	71.65	71.0
BART-L	48.89	88.6	93.44	97.12	92.6
T5-B	48.89	90.7	93.90	98.14	92.6

Random
Initialized

◆ Effects of Each Task Corpus

Method	DS (R-L)	DC (BLEU)	ID (ACC.)	SF (F_1)	DST (JGA)
MATS	47.97	90.6	93.60	97.61	92.3
- DS	-	90.2 \downarrow 0.4	93.20 \downarrow 0.4	97.35 \downarrow 0.26	92.8 \uparrow 0.5
- DC	47.77 \downarrow 0.20	-	93.41 \downarrow 0.19	97.39 \downarrow 0.22	91.8 \downarrow 0.5
- ID	47.81 \downarrow 0.16	90.5 \downarrow 0.1	-	97.45 \downarrow 0.16	92.3 <0.0
- SF	47.77 \downarrow 0.20	90.5 \downarrow 0.1	93.60 <0.0	-	92.0 \downarrow 0.3
- DST	47.85 \downarrow 0.12	90.6 <0.0	93.47 \downarrow 0.13	97.58 \downarrow 0.03	-



Experimental Results

• Case Study

Unseen Dialogue Content	UniDUMATS
[DS] USER : I'd like a taxi to take me to ruskin gallery [T] SYSTEM : Sure! What is your departure site? [T] USER : I will depart from saffron brasserie at 7:15 . What is the car type and contact number so I know who and where you will pick me up? [T] SYSTEM : Booking completed! A grey ford will be picking you up. The contact number is 07689877132. [T] USER : That is all I needed, thank you. [C] what's the summary of this dialogue?	[DS] a grey ford will take USER to ruskin gallery at 7:15.
[DC] USER : Please reserve for me a taxi that will pick me up at cambridge arts theatre after 09:30 [T] SYSTEM : And where will you be going? [T] USER : I'm going to restaurant one seven. [T] SYSTEM : Your booking is complete, a black audi will be picking you up . [T] USER : Thank you. I need the contact number , as well. [C] what is the semantic completion statement of "Thank you. I need the contact number, as well."?	[DC] I need the contact number of a black audi to pick me up at cambridge arts theatre
[ID] help me get a taxi to the cambridge museum of technology please. [C] what is the user's intent on the taxi?	[ID] transport taxi
[SF] I need a taxi to pick me up at Ashley Hotel to leave after 10:45 . [C] what is leaving time of taxi?	[SF] 10:45
[DST] USER : I need a taxi. I am going to avalon and I need to leave after 16:15 [C] what is the user's constraint about the destination of the taxi?	[DST] avalon



Motivation

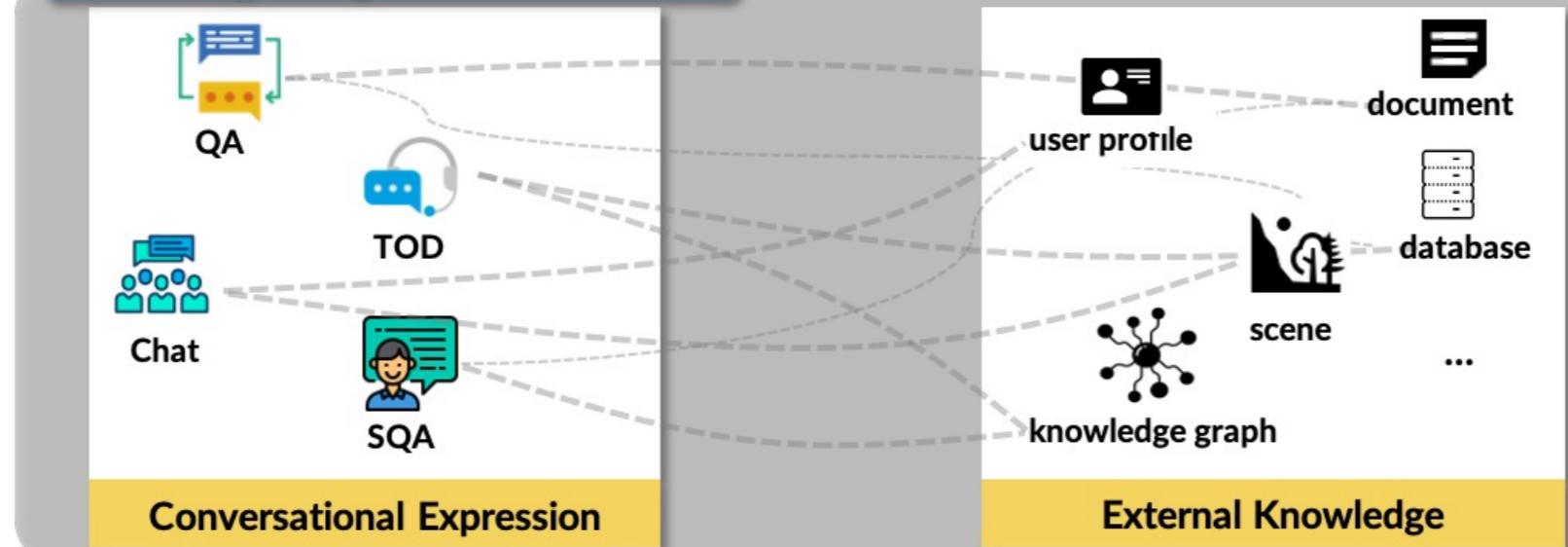


Can we model all the dialog-related task in a unified framework?

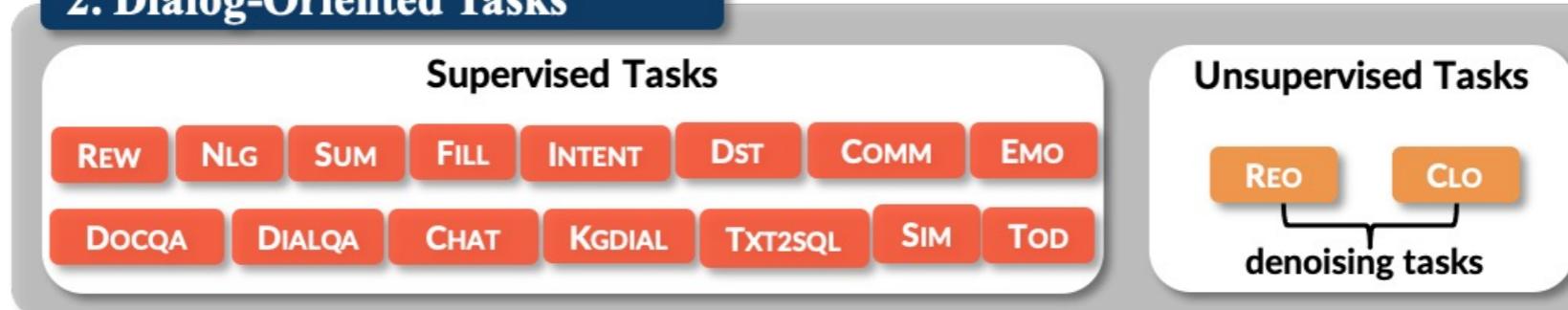


Dialogue Foundation Model

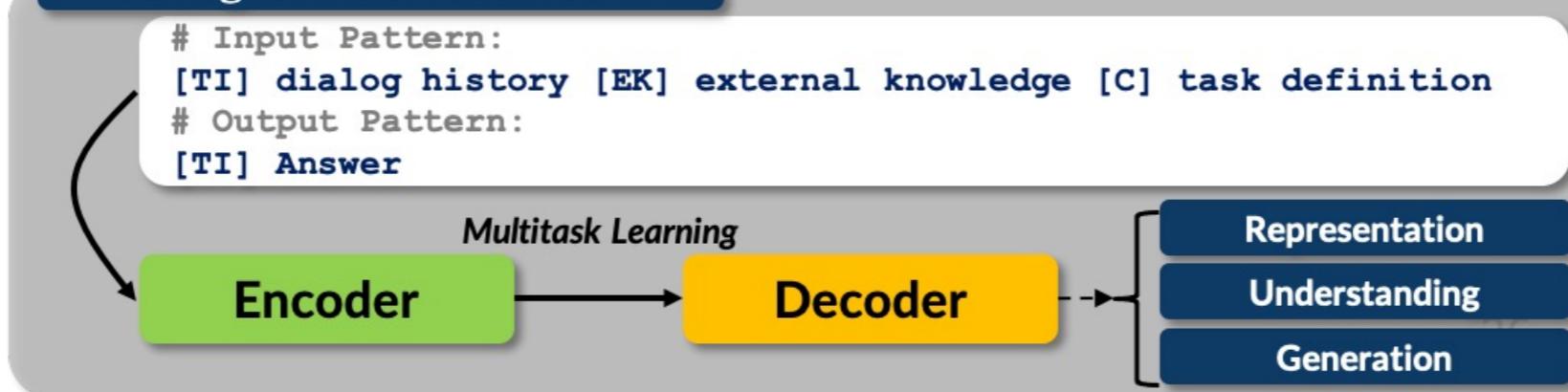
1. Dialog-Corpus Collection



2. Dialog-Oriented Tasks



3. Dialogue Foundation Model





Dialogue Foundation Model

- **Compare UniDU with DFM**

For dialogue state tracking (in UniDU):

Input ---> [dst] USER: I need to find an apartment please. [T] SYSTEM: In what area? [T] USER: Something in San Jose. [C] **What is the user's constraints about the home area?**

output ---> [dst] San Jose



For dialogue state tracking (in DFM):

More General

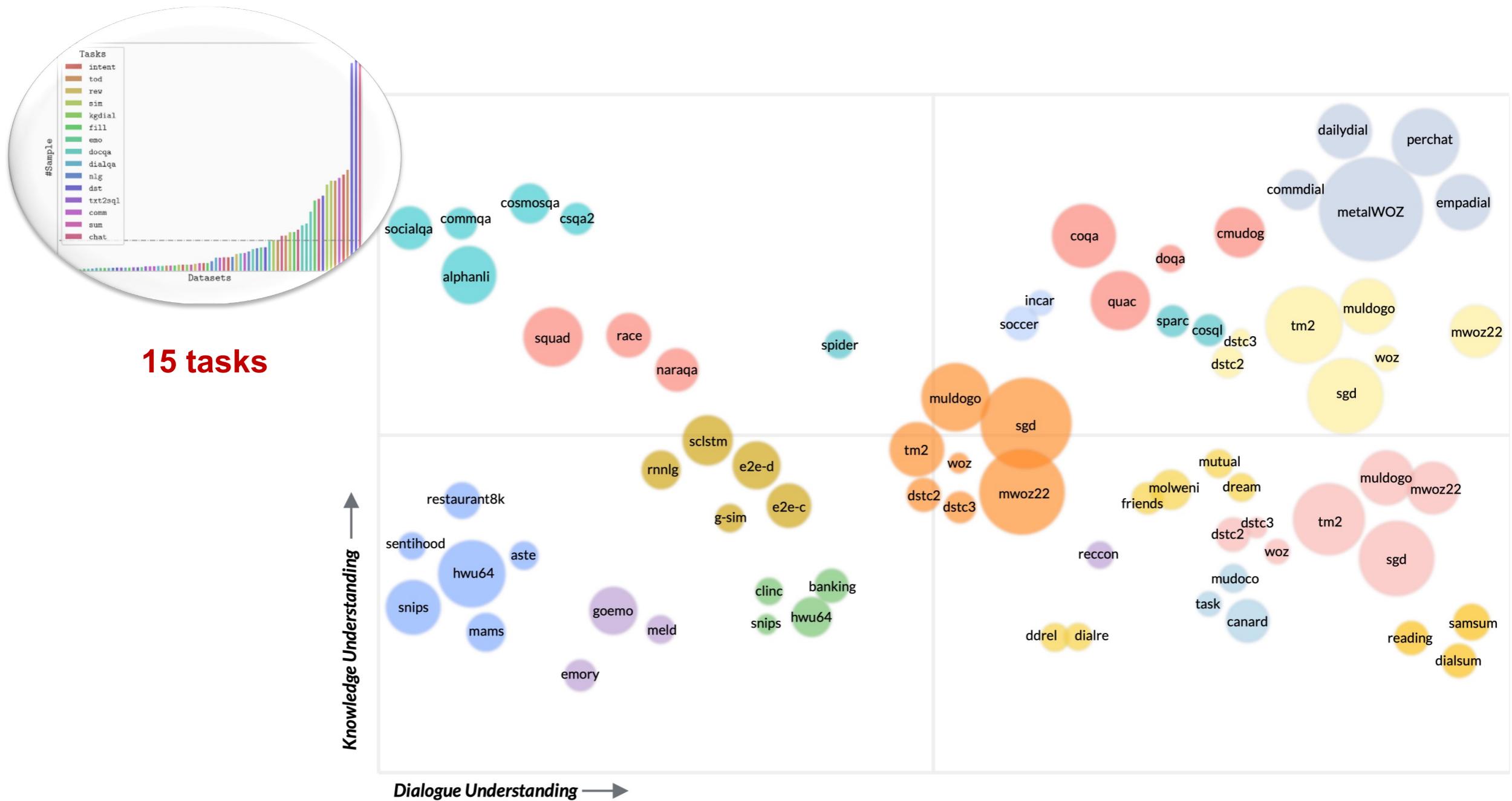
Input ---> [dst] USER: I need to find an apartment please. [T] SYSTEM: In what area? [T] USER: Something in San Jose. [EK] **home.area** [C] Give the user's constraint on the slot.

output ---> [dst] San Jose



Dialogue Foundation Model

Rich Dialog Tasks & Diversity Corpus Scale

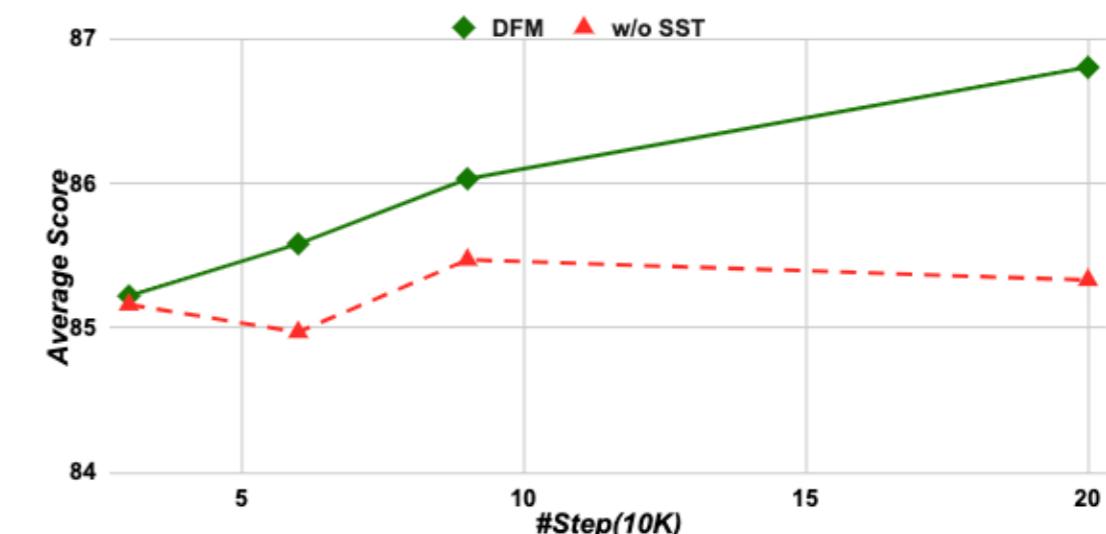




Experimental Results

• Dialogue Representation Ability

DialoGLUE	ConvBERT	ConvBERT*	BART _{base}	PLMs	DFM _{bart}	T5 _{base}	DFM _{t5}
Intent Detection							
BANKING77 (ACC.)	92.89	92.08	93.15	93.32	92.31	93.02	
CLINC150 (ACC.)	97.04	95.58	96.20	97.33	96.49	97.31	
HWU64 (ACC.)	91.82	90.89	91.17	94.42	91.26	93.16	
Slot Filling							
RESTUARANT8K (F1)	96.04	95.29	92.29	93.89	95.98	96.07	
DSTC8 (F1)	89.45	86.71	86.77	88.53	90.56	90.76	
Semantic Parsing							
TOP (EM)	81.43	80.80	80.71	81.24	81.81	82.04	
Dialog State Tracking							
MULTIWOZ2.1 (JGA)	58.11	56.21	56.57	59.19	57.21	60.58	
Average Score	86.68	85.37	85.27	86.85	86.73	87.56	





Experimental Results

- **Dialogue Understanding Ability**

Knowledge Distillation	SOTA	Methods	
		DFM _{t5}	w/ FT
Intent Detection			
BANKING77 (ACC.)	93.86 (2021)	92.82	93.63
CLINC150 (ACC.)	97.16 (2020a)	100	100
HWU64 (ACC.)	92.94 (2020)	98.70	98.42
Slot Filling			
RESTUARANT8K (F1)	98.00 (2021)	95.95	99.28
DSTC8 (F1) [†]	89.45 (2020)	47.85	89.69
Semantic Parsing			
TOP (EM) [‡]	86.67 (2020)	/	87.10
Dialogue State Tracking			
MULTIWOZ2.2 (JGA)	57.70 (2022b)	58.52	59.20

Datasets	Methods		
	T5 _{base}	UNISKG _{base}	DFM _{t5}
With Beam Search Decoding			
Spider (EM)	57.20	59.86	59.96
CoSQL (QEM)	42.30	45.68	46.57
With PICARD Decoding			
Spider (EM)	65.80	/	68.47
CoSQL (QEM)	47.90	/	52.04



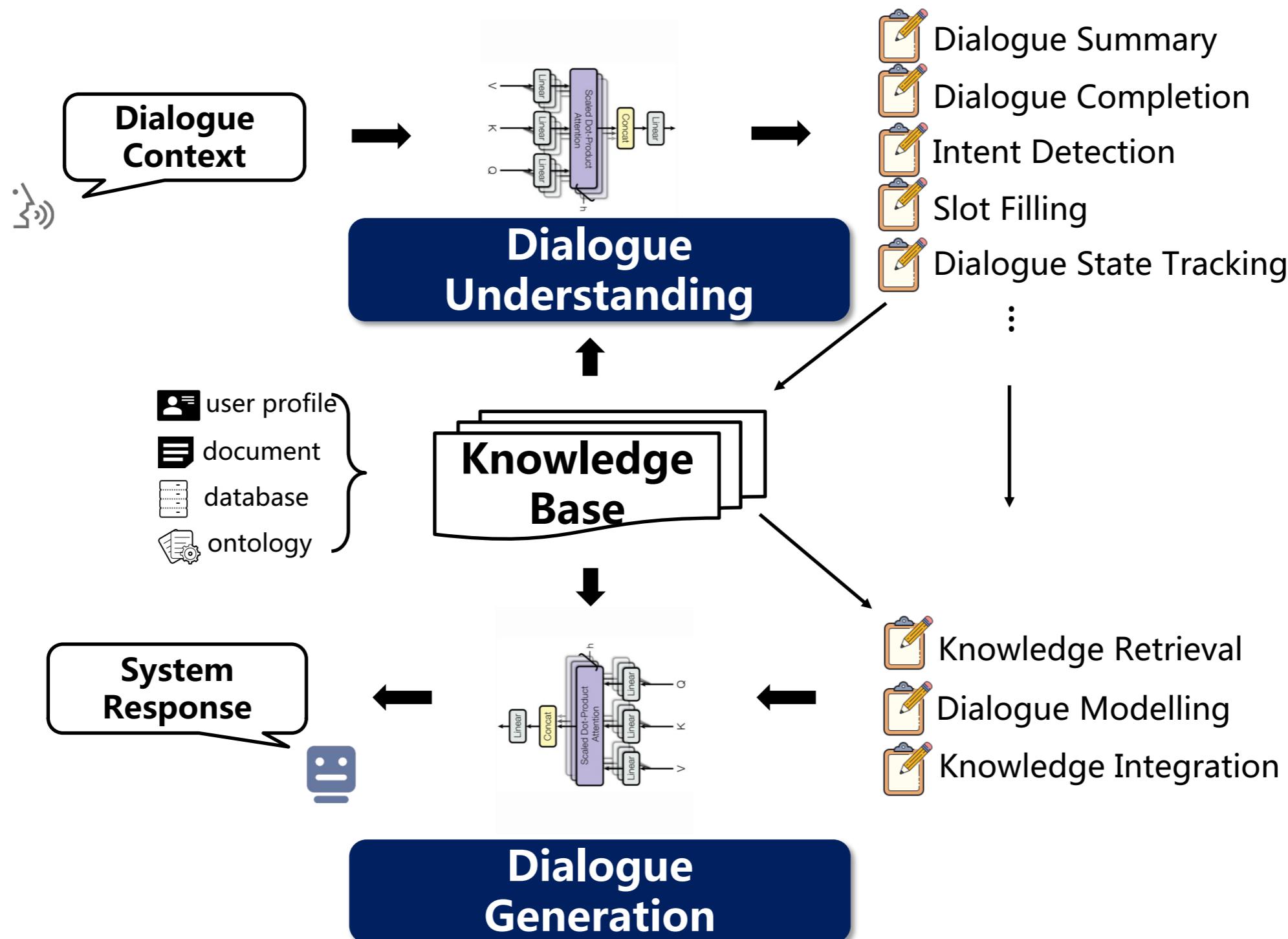
Experimental Results

- **Dialogue Generation Ability**

Model	Inform	Success	BLEU	Combined
With Predicted Dialogue State				
SimpleTOD (2020)	84.40	70.10	15.01	92.26
SOLOIST (2021)	85.50	72.90	16.54	95.74
MinTL (2020)	84.88	74.91	17.89	97.78
UBAR (2021)	91.50	77.40	17.00	101.50
NCM (2021c)	86.90	76.20	20.58	102.13
HTER (2021)	91.72	75.80	19.05	102.81
PPTOD (2021)	89.20	79.40	18.62	102.92
BORT (2022a)	93.80	85.80	18.50	108.30
GALAXY (2022)	94.40	85.30	20.50	110.35
DFM _{t5}	91.50	84.70	22.86	110.96
With Golden Dialogue State				
UBAR (2021)	94.00	83.60	17.20	106.00
GALAXY (2022)	94.80	85.70	19.93	110.18
DFM _{t5}	93.20	85.60	23.38	112.78



Summary





Thanks for attention

Q&A