

Team Flow at DRC2022: Pipeline System for Travel Destination Recommendation Task in Spoken Dialogue

Ryu Hirai¹, Atsumoto Ohashi¹, Ao Guo¹, Hideki Shiroma¹, Xulin Zhou¹,
Yukihiko Tone¹, Shinya Iizuka² and Ryuichiro Higashinaka¹

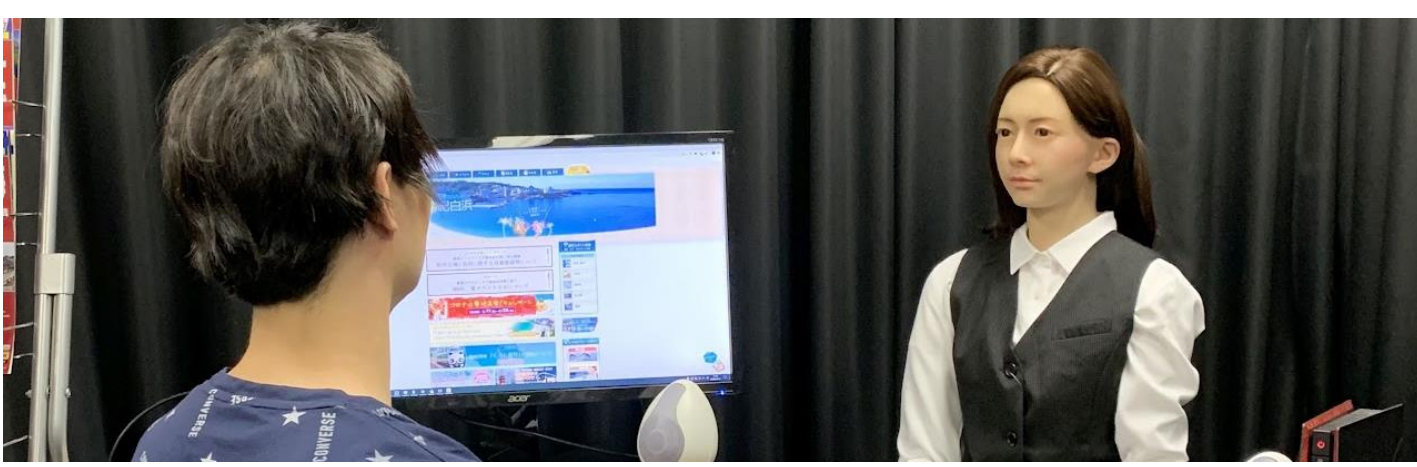
¹Graduate School of Informatics, Nagoya University, Japan

²School of Informatics, Nagoya University, Japan

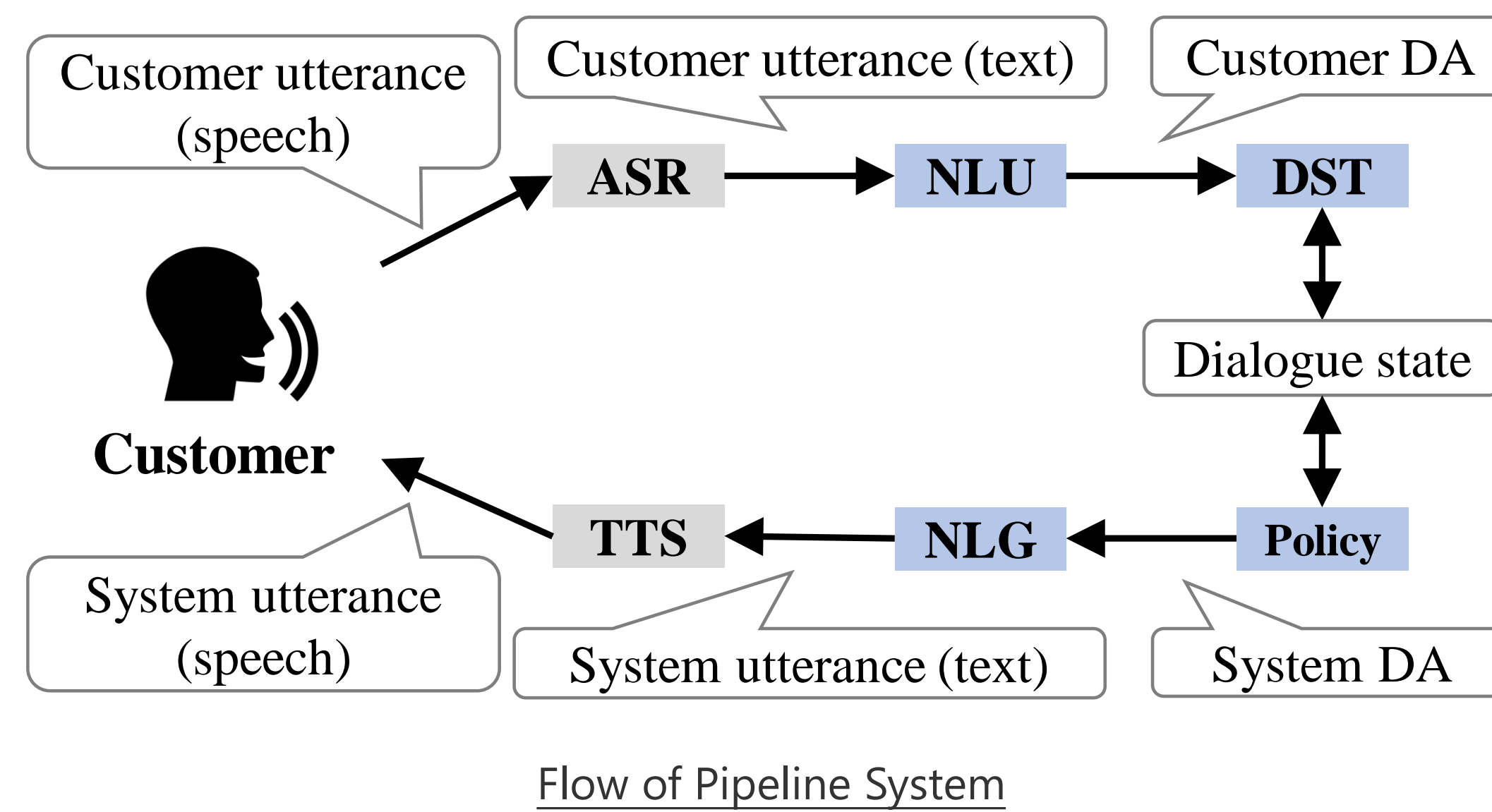
Overview

Dialogue Robot Competition (DRC) 2022

- A humanoid robot role-plays a counter-sales person in a travel agency
- Each team implements a dialogue system which is to be embedded in the robot
- Goal is to help customer choose one of two tourist attractions



Travel Destination Recommendation Task



Flow of Pipeline System

Our System

- Our system has a pipeline structure with four modules connected sequentially
 - Easy to tune the functionality of each module
 - All modules can be optimized simultaneously (future work)
- NLU & NLG are implemented by the GPT-based large-scale language model

Findings

- Our pipeline dialogue system is limited in performance in human evaluation
- The reasons for the limited performance are clarified

Ontology of Dialogue Act

- Dialogue Act (DA)** consists of an *intent*¹ and zero or more *slot-value pairs*²

- Intent* is an intention of customer or system
- Slot-value pairs* are detailed information to supplement the intent

System	Customer
welcome, self_introduction, finish_for_time_limit, thank-you_for_visiting, goodbye, sorry, good, affirm, negate, start_request, request, confirm_attraction, inform, recommend_target, ask_question, start_attraction_introduction	thankyou, goodbye, request, affirm, negate, inform, greet

Intents for System and Customer

Slot	Value type	Description
user_name	String	Customer's name
attraction_parking	Categorical	If parking is available at the (unknown/yes/no) tourist attraction

Example of slot definition

NLU

- NLU** predicts a customer DA from a customer utterance at each turn
- Data Collection via Crowdsourcing**
 - We collected 4,104 customer utterances for NLU
- Generative Model**
 - We used Japanese GPT-2 336M parameters ver. for NLU

DA		Utterance
Intent	Slot-value Pair	
inform	user_accompany = child	I would like to bring my children to see the sights.
		...

Example of collected utterances

I would like to bring my children to see the sights.
[SEP] inform (user_accompany=child)

Example of input and output sequence of NLU

DST

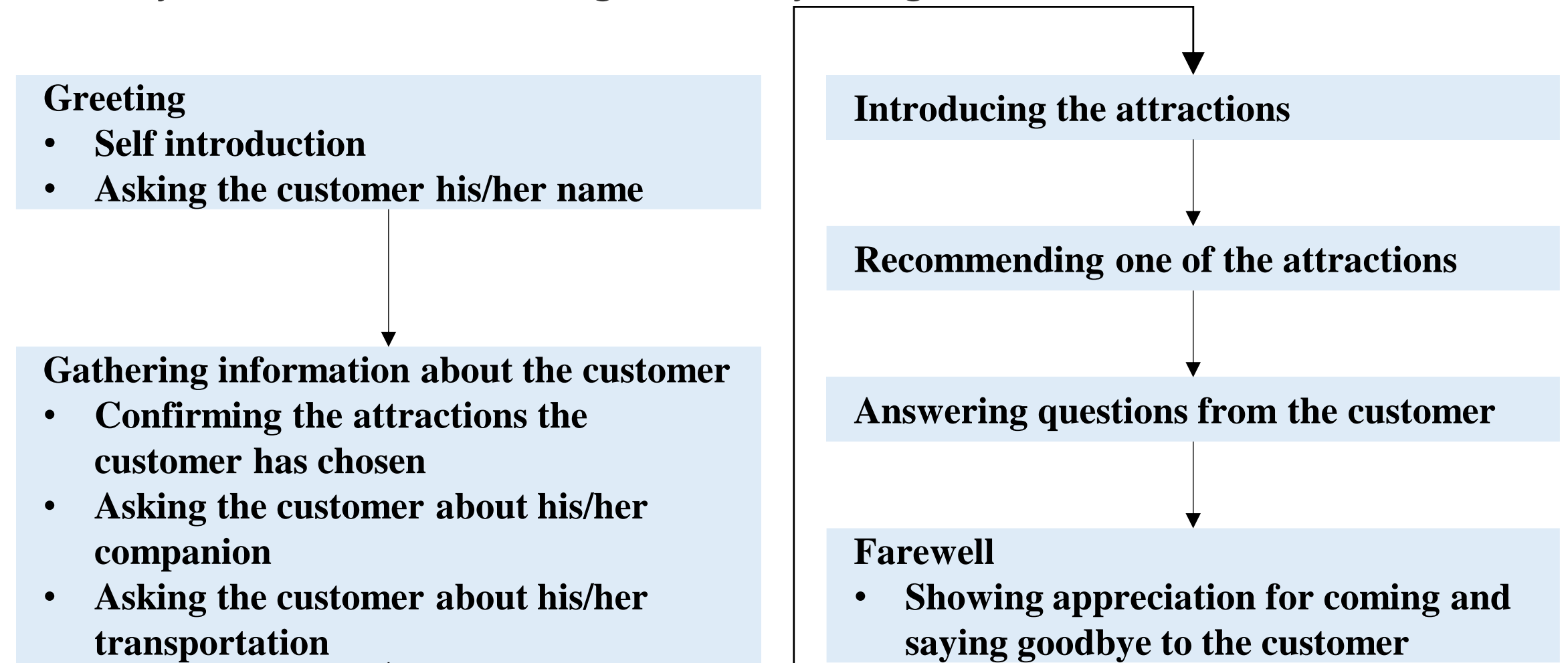
- DST** keeps the dialogue state up-to-date based on the hand-crafted rules for updating the dialogue state on the basis of customer DAs
- The dialogue state consists of:
 - customer's profile
 - attractions focused on by the customer
 - belief state
 - history of the DAs

Customer DA			
inform (user_accompany = family)			
Belief State (t)		Belief State (t+1)	
Slot	Value	Slot	Value
user_name	Taro	user_name	Taro
user_accompany		user_accompany	family
...

Example of procedure for belief state update

Policy

- Policy** determines the next system DA on the basis of the current dialogue state
- The system follows the dialogue flow by using hand-crafted rules:



Flow of dialogue

NLG

- NLG** generates a system utterance from a system DA at each turn
- Data Collection via Crowdsourcing**
 - We collected 9,716 system utterances for NLG
- Generative Model**
 - We used Japanese GPT-2 1.3B parameters ver. for NLG

DA		Utterance
Intent	Slot-value Pair	
request	user_accompany = ?	Who would you like to tour with?
		...

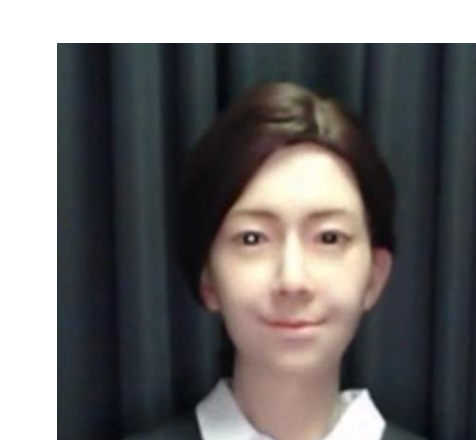
Example of collected utterances

inform (user_accompany=?) [SEP] Who would you like to tour with?

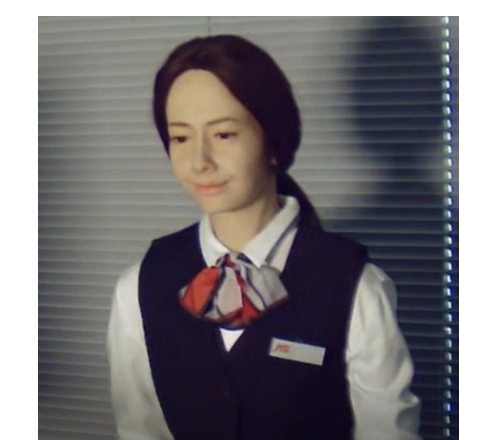
Example of input and output sequence of NLG

Expression Control and Motion Control

- Each system DA was expressed with a combination of one of five expressions and one of six motions
- Expression/motion is controlled **during and after an utterance**
- Example:
 - When system DA includes "good" intent,
 - Robot nods its head with a large smile during an utterance
 - When system DA includes "goodbye" intent,
 - Robot makes a small smile and bows after the utterance



Large Smile

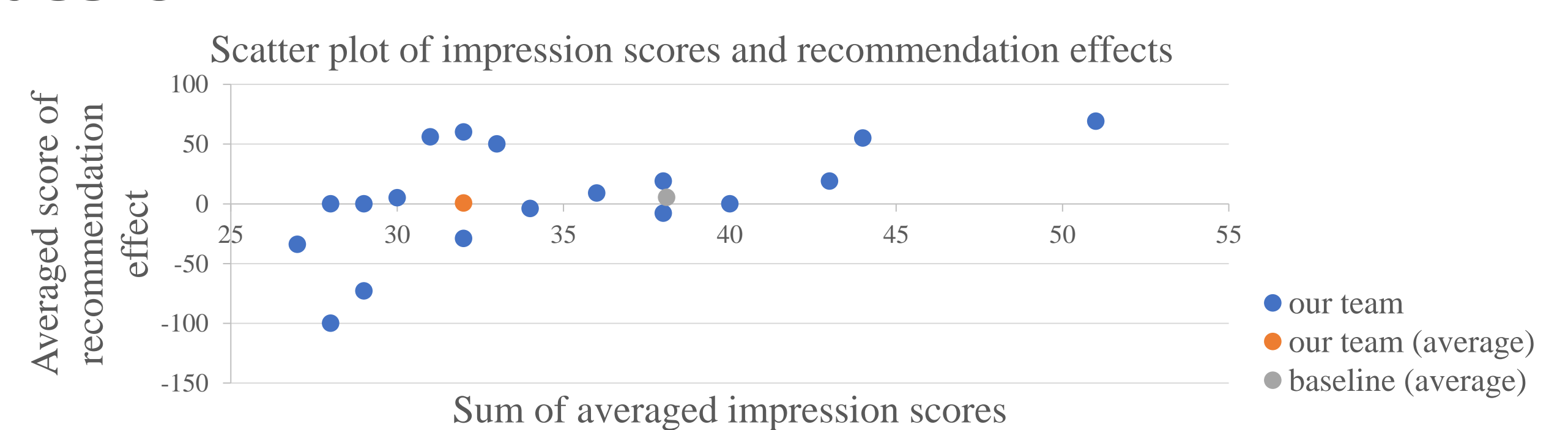


Bow

Results & Discussion

Evaluation & Results

- Impression Evaluation (IE)**: Consists of nine questions on a 7-point Likert
- Robot Recommendation Effect (RRE)**: The change in the degree to which the customer wanted to visit the attraction recommended
- Our system showed a limited performance: 32 for IE and -0.6 for PRE in human evaluation



Case Study

Speaker	Utterance	DA estimated by NLU or determined by policy
System	お客様、どのようなお食事をご希望でしょうか。What kind of meal would you like to have?	request (user_food_type=?)
Customer	子供が食べやすいものかいいです。I prefer something easy for children to eat.	inform (user_accompany=child, user_food_type=steak)
System	ステーキをご希望ですね。かしこまりました。You would like a steak, I understand.	good (user_food_type=steak)
Customer	(silent)	
System	東京トリックアート迷宮館は雨天でも楽しめるのでおすすめです。I recommend the Tokyo Trick Art Museum because you can enjoy it even in the rain.	recommend_target (attraction_name=Tokyo Trick Art Museum, attraction_rain=ok)
Customer	何時から何時までやってるんですか。What are its opening hours?	request (attraction_open_time=?)
System	東京トリックアート迷宮館は11時から21時まで営業しています。東京都水の科学館の営業時間は9時30分～17時です。The Tokyo Trick Art Museum is open from 11:00 am to 9:00 pm. The Tokyo Water Science Museum is open from 9:30 am to 5:00 pm.	inform (attraction_name=Tokyo Trick Art Museum, attraction_open_time=11:00-21:00), inform (attraction_name=Tokyo Water Science Museum, attraction_open_time=9:30-17:00)

Good Points

- The system anticipates and responds to what the customer possibly wants to know by the rules of policy
 - The system correctly answered the opening hours of the Tokyo Trick Art Museum to the customer
 - The system also provided the customer with the opening hours of the Tokyo Water Science ("The Tokyo Water Science Museum is open from 9:30 am to 5:00 pm.")

Bad Points

- NLU training examples lacked variation and the system misrecognized the customer DA
 - From "I prefer something easy for children to eat," NLU incorrectly predicted a DA: inform (user_accompany=child, user_food_type=steak)
- The system failed the recommendation task due to the lack of the rules of policy
 - The system consequently recommended the attraction ignoring the customer's food preferences ("I recommend the Tokyo Trick Art Museum because you can enjoy it even in the rain")

Conclusion

- We implemented a pipeline dialogue system with four modules: *NLU*, *DST*, *Policy* and *NLG*
- We clarified the two reasons why the dialogue system had a limited performance:
 - Low variation in training examples for NLU
 - Failed recommendations due to the lack of the rules of policy

Future Work

- GPT-3 based model for NLU & NLG
- Polish the rules of DST and Policy
- Integrating all the modules and optimize the dialogue performance of the entire system
 - E.g., by using Post-processing Networks (PPN)