Retrospective and Prospective Mixture-of-Generators for Task-oriented Dialogue Response Generation

Jiahuan Pei Pengjie Ren Christof Monz Maarten de Rijke

University of Amsterdam

{*j.pei, p.ren, c.monz, derijke*}*@uva.nl*

ECAI 2020 Santiago de Compostela

Task: dialogue response generation (DRG)

Task definition

- **Input**: dialogue context
  ⇒ historical utterances,
  ⇒ belief states,
  ⇒ retrieved database results.
- **Output**: response – a sequence of words generated token by token
  ⇒ *Task-completion* – e.g., assist a user to book a flight successfully,
  ⇒ *Language fluency* – it sounds like a natural language from a real person.

Motivation: statistical studies

- Exiting assumption: each token is drawn from a *single* distribution over the vocabulary.

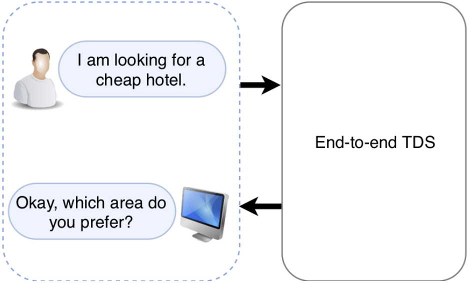


Figure 1: Dominant end-to-end TDS.

- This is **unreasonable**: responses vary greatly with different intents.

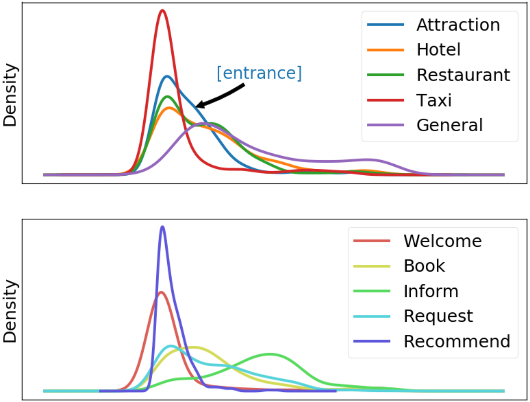
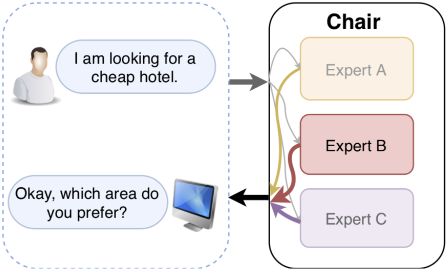


Figure 2: Relative token frequency distribution of different domains (top), system actions (bottom).

Solution: 1 chair – k experts framework



- *k* **expert generators**, each of which is specialized for a particular *intent*, e.g., a domain, a type of action of a system, etc.
- a **chair generator**, which learns to coordinate a group of experts to make an optimal decision.
- Each expert follows Seq2Seq under an encoder-decoder architecture.
- Mixture-of-Experts adopted to decoder part.

Main research question

Does DRG performance improves if a response be drawn from a **mixture of distributions** for *multiple intents* compared with from a single distribution for a general intent?

Dataset

Multi-domain Wizard-of-Oz (MultiWOZ)

- ∼10k natural conversations between a tourist and a clerk.
- 8,438/1,000/1,000 dialogues for training, validation, and testing, respectively.

Table: Two groups of intents that are divided by domains and the type of system actions.

Type	**Intents**
Domain	Attraction, Booking, Hotel, Restaurant, Taxi, Train, General, UNK.
Action	Book, Inform, NoBook, NoOffer, OfferBook, OfferBooked, Select, Recommend, Request, Bye, Greet, Reqmore, Welcome, UNK.

Main results: automatic evaluation

- MoGNet improves **5.64%** of the overall performance automatically evaluated by balancing both task-completion and language fluency.

	BLEU	**Inform**	**Success**	**Score**
S2SAttnLSTM	18.90%	71.33%	60.96%	85.05
S2SAttnGRU	18.21%	81.50%	68.80%	93.36
Structured Fusion	16.34%	82.70%	72.10%	93.74
LaRLAttnGRU	12.80%	82.78%	**79.20%**	93.79
MoGNet	**20.13%***	**85.30%***	73.30%	**99.43***

$$Score = (0.5 * Inform + 0.5 * Success + BLEU) * 100$$

Main results: human evaluation

- MoGNet improves ⩾**12.5%** of the overall performance manually evaluated by *Satisfactory* of workers on Amazon Mechanical Turk.

	S2SAttnGRU		LaRLAttnGRU		MoGNet	
	⩾ 1	⩾ 2	⩾ 1	⩾ 2	⩾ 1	⩾ 2
Informativeness	56.79%	31.03%	76.54%	44.83%	**80.25%**	**53.45%**
Consistency	45.21%	23.53%	71.23%	39.22%	**80.82%**	**50.98%**
Satisfactory	26.79%	25.00%	44.64%	21.88%	**60.71%**	**37.50%**

Main results: case study

- Example:

User	**System**
i am looking for a train departing from [value_place] that will arrive at [value_place].	**S2SAttnGRU**: what day would you like to travel? **LaRLAttnGRU**: there are [value_count] trains leaving [value_place] on [value_day] . what day would you like to travel?
	MoGNet: what day and time would you like to travel?

Main results: ablation study

- PR coordination mechanisms (-2.62%); GL learning scheme (-6.95%).

	BLEU	**Inform**	**Success**	**Score**
MoGNet	20.13%	85.30%	73.30%	99.43
MoGNet-P-R	<u>18.16%</u>	85.10%	72.20%	96.81
MoGNet-GL	19.33%	<u>78.40%</u>	<u>67.90%</u>	<u>92.48</u>

Conclusion

- **Contribution**:
  - A mixture-of-generators network (MoGNet) model that firstly devises chair and expert generators for DRG.
  - Two coordination mechanisms (RMoG & PMoG) to help the chair make better decisions.
  - A global-and-local (GL) learning scheme to differentiate experts and fuse data efficiently.
- **Future work**:
  - Explorations of coordination mechanisms between chair and expert generators.
  - Studies on how to do intent partition automatically.

Thanks for your attention!

- **Paper**: Jiahuan Pei, Pengjie Ren, Christof Monz, and Maarten de Rijke. "Retrospective and Prospective Mixture-of-Generators for Task-oriented Dialogue Response Generation." In Proceedings of ECAI 2020.
- **Emails**: {*j.pei, p.ren, c.monz, derijke*}*@uva.nl*
- **Code**: https://github.com/Jiahuan-Pei/multiwoz-mdrg