Sequicity: Simplifying Task-oriented Dialogue Systems with Single Sequence-to-Sequence Architectures

Wenqiang Lei, Xisen Jin, Zhaochun Ren, Xiangnan He, Min-Yen Kan, Dawei Yin

{wenqianglei,xisenjin}@gmail.com renzhaochun@jd.com kanmy@comp.nus.edu.sg xiangnanhe@gmail.com yindawei@acm.org

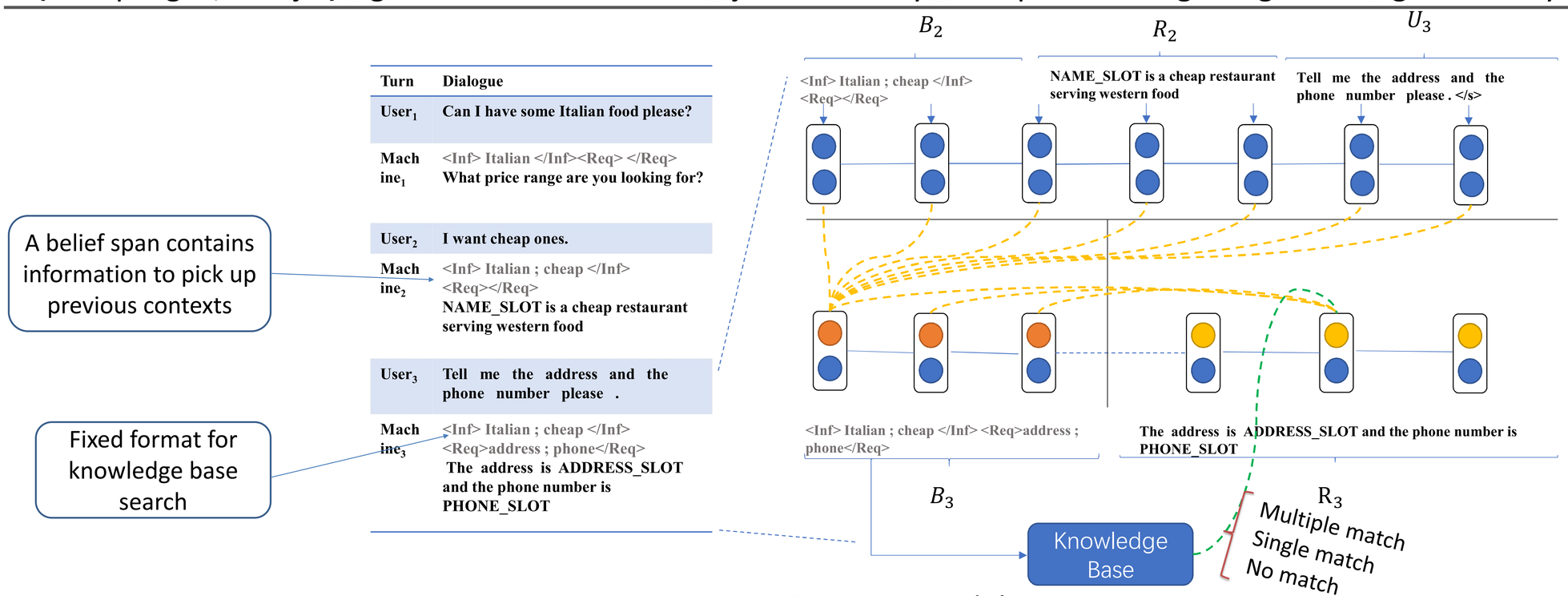

Figure 1. Model overview.

Motivation

Advantages	Sequicity	Neural Pipeline Design	Traditional Pipeline Design
Simple belief tracker	✓	✓	✗
Holistic	✓	✓	✗
Natural response	✓	✗	✗

Belief Trackers

Food style	Price range	…
Chinese	Expensive	…
Japanese	Cheap	…
French	Moderate	…
…	…	…

(a) Informable slots

Address	Phone Number	…
Yes	Yes	…
No	No	…

(b) Requestable slots

Table 1. Examples of slots

Belief Spans

- A text span to record Informable Slots and Requestable Slots to enable a RNN to decode it. For example:

  <Inf>Chinese; Expensive<\Inf> <Req>Address; Phone<\Req>

- Roles: knowledge base search and response conditioning.

Sequicity Formalization

- Source-target sequence pair:

  $\{B_0R_0U_1, B_1R_1\}, \{B_1R_1U_2, B_2R_2\}, \ldots, \{B_{t-1}R_{t-1}U_t, B_tR_t\}$.

- Two-stage decoding:
  - B_t=Seq2seq($B_{t-1}R_{t-1}U_t$)
  - R_t=Seq2seq($B_{t-1}R_{t-1}U_t|B_t$, KB search results)

Notations:
- B: belief spans
- U: user utterances
- R: machine responses

Experiment Settings

- Sequicity instantiation: two-stage CopyNets (Gu et al., 2016) (TSCP) optimized by joint log-likelihood and policy gradient.

Dataset	Cam676		
Size	Train : 408 / Test : 136 / Dev : 136		
Domains	restaurant reservation		
Slot types	price, food style etc.		
Distinct slot values	99		
Dataset	**KVRET**		
Size	Train : 2425 / Test : 302 / Dev : 302		
Domains	calendar	weather info.	POI
Slot types	date, etc.	location, etc.	poi, etc.
Distinct slot values	79	65	140

Table 2. Datasets

Main Experiment Results

	CamRes676					KVRET				
	Mat.	BLEU	Succ. F_1	$Time_{full}$	$Time_{N.B.}$	Mat.	BLEU	Succ. F_1	$Time_{full}$	$Time_{N.B.}$
(1) NDM (Wen etl al., 2017b)	0.904	0.212	0.832	91.9 min	8.6 min	0.724	0.186	0.741	285.5 min	29.3 min
(2) NDM + Att + SS (wen et al., 2016a)	0.904	0.240	0.836	93.7 min	10.4 min	0.724	0.188	0.745	289.7 min	33.5 min
(3) LIDM (Wen et al., 2017a)	0.912	0.246	0.840	97.7 min	14.4 min	0.721	0.173	0.762	312.8 min	56.6 min
(4) KVRN (Eric et al., 2017b)	N/A	0.134	N/A	21.4 min	—	0.459	0.184	0.540	46.9 min	—
(5) TSCP	**0.927**	**0.253**	**0.854**	7.3 min	—	**0.845**	**0.219**	**0.811**	25.5 min	—
(6) Att-RNN	0.851	0.248	0.774	7.2 min	—	0.805	0.208	0.801	23.0 min	—
(7) TSCP \ k_t	**0.927**	0.232	0.835	7.2 min	—	**0.845**	0.168	0.759	25.3 min	—
(8) TSCP \ RL	**0.927**	0.234	0.834	**4.1** min	—	**0.845**	0.191	0.774	17.5 min	—
(9) TSCP \ B_t	0.888	0.197	0.809	22.9 min	—	0.628	0.182	0.755	42.7 min	—

Table 3. Model performances. Mat. is for match rate. Rows 1-4 are baselines where rows 1-3 are a family of neural pipeline-designed models. Row 5 is our final two stage CopyNets. We also performs ablation studies (rows 6-9). Row 6 drops copy mechanism and rows 7-8 drops knowledge base search results, reinforcement learning separately. Row 9 drops belief spans and concatenates all past utterances to record contexts.

OOV Testing

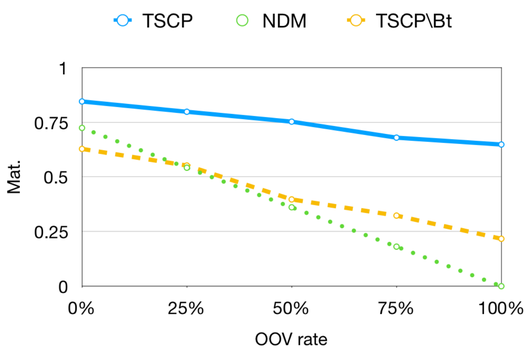

Figure 2. OOV testing with KVRET.

Model Size

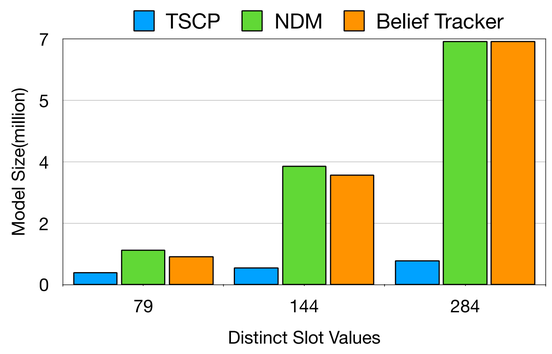

Figure 3: Model size sensitivity with respect to KVRET.

Conclusion

- Proposing Sequicity framework, enabling task-oriented dialogue system to be holistically optimized in a single seq2seq model.
- TSCP, our instantiation of Sequicity demonstrates good properties of Sequicity: better effectiveness, scalability and an additional capability of handling OOV requests.

References

- Tsung-Hsien Wen, David Vandyke, Nikola Mrksic, Milica Gasic, Lina M Rojas-Barahona, Pei-Hao Su, Stefan Ultes, and Steve Young. 2017b. A networkbased end-to-end trainable task-oriented dialogue system. EACL .
- Tsung-Hsien Wen, Yishu Miao, Phil Blunsom, and Steve Young. 2017a. Latent intention dialogue models. ICML .
- Tsung-Hsien Wen, Milica Gasic, Nikola Mrksi˘ c,´ Lina M. Rojas Barahona, Pei-Hao Su, Stefan Ultes, David Vandyke, and Steve Young. 2016b. Conditional generation and snapshot learning in neural dialogue systems. In EMNLP.
- Mihail Eric and Christopher D Manning. 2017b. Keyvalue retrieval networks for task-oriented dialogue. SIGDIAL .
- Jiatao Gu, Zhengdong Lu, Hang Li, and Victor OK Li. 2016. Incorporating copying mechanism in sequence-to-sequence learning. ACL .