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

Motivation

Advantages

Sequicity Neural Pipeline Traditional Pipeline

Simple belief tracker ✓ ✓ X

Holistic ✓ ✓ X

Natural response ✓ ✓ X

Sequicity Formalization

(a) Informable slots

(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.

Table 2. Datasets

OOV Testing

Model Size

Figure 1. Model overview.

Figure 2. OOV testing with KVRET.

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 network-based end-to-end trainable task-oriented dialogue system. EACL.

• Tsung-Hsien Wen, Yi Gu, Phua Bloom, and Steve Young. 2017a. Latent attention dialogue models. ICLR.

• Tsung-Hsien Wen, Milica Gasic, Nikola Mrksic, 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.

• Jianan Gu, Zhengdong Lu, Hang Li, and Victor O K Li. 2016. Incorporating copying mechanism in sequence-to-sequence learning. ACL.

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

Model Size

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

Model Size

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

Model Size

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

Model Size

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

Model Size

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

Model Size

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

Model Size

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

Model Size

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.