Learning to Memorize in Neural Task-Oriented Dialogue Systems

Chien-Sheng (Jason) Wu
Master of Philosophy (MPhil) Thesis Defense Presentation
June 10th, 2019

Supervisor: Prof. Pascale Fung
Outline

▷ Dialogue System Introduction
▷ Copy-Augmented Dialogue State Tracking (Chapter 3, one ACL paper)
▷ End-to-End Retrieval-based Memory-Augmented Dialogue Learning (Chapter 4, one ICASSP paper one DSTC paper)
▷ End-to-End Generation-based Memory-Augmented Dialogue Learning (Chapter 5, one ACL paper one ICLR paper)
▷ Thesis Conclusion (Chapter 6)
Brief History of Dialogue Systems: Language Empowering AI

Task-specific argument extraction (e.g., Nuance, SpeechWorks)
User: “I want to fly from Boston to New York next week.”

Intent Determination (Nuance’s Emily™, AT&T HMIHY)
User: “Uh...we want to move...we want to change our phone line from this house to another house”

Keyword Spotting (e.g., AT&T)
System: “Please say collect, calling card, person, third number, or operator”

Multi-modal systems
e.g., Microsoft MiPad, Pocket PC

TV Voice Search
e.g., Bing on Xbox

2017

Early 2000s

Early 1990s

Material: http://deepdialogue.miulab.tw
Growing Market and Research

Material: https://www.cbinsights.com/research/facebook-amazon-microsoft-google-apple-voice/
Dialogue Systems: Chit-Chat v.s. Task-Oriented

**Chit-Chat Dialogue Systems**
- No Specific goal
- Focus on generating natural responses
- The more turns the better
- Using variants of generation models, ex: Seq2Seq, VAE, etc.

**Task-Oriented Dialogue Systems**
- Help users achieve their goal
- Focus on understanding users, tracking states, and generating next actions.
- The less turns the better
- Combination of rules and statistical components.
## Keywords in Dialogue Systems

| **Domains**       | Topics of the current conversation. Ex: restaurant domain, taxi domain, etc. |
|-------------------|----------------------------------------------------------------------------|
| **Intents**       | Goals of each user utterance. Ex: request_movie, inform_location.          |
| **Slots**         | Predefined variables in dialogues that can be filled with all kinds of values. Ex: location, people, day, etc. |
| **Ontology**      | Predefined slots and their values.                                        |
| **Dialogue History** | Dialogue context in the current multi-turn conversation.                  |
| **Dialogue States** | System belief about user intention as slot-value pairs. Ex: {location: Paris; Price Range: Cheap} |
| **Knowledge Bases (KB)** | Back-end information that could be provided to users.                  |
Challenges in Dialogue Systems

Multi-turn Conversations
- Domain Classification
- Intention Detection
- State Tracking
- Natural Response Generation, etc.

Machine Learning Models

External Knowledge
- Knowledge bases, Common Senses, etc.
Modularized Dialogue Systems

- Domain Identification
- Intent Detection
- Slot Filling

Represent system’s belief of user’s goal as slot-value pairs

Mapping actions and states into natural language.

- Query knowledge base
- Dialogue policy (generate next action)
The Core of User Understanding: Joint SLU+DST

Figure 2.1: Block diagram of modularized task-oriented dialogue systems.
The Core of User Understanding: SLU+DST

Four main problems:

- Full ontology is hard to obtain in real scenarios
- Need to track lots of slot values
- Cannot track unseen slot values
- Missing domain sharing capacities

![Diagram showing the core of user understanding with examples of finding a train and a taxi at 5pm, highlighting the need for an ontology to track slot values and domain sharing capacities.](image-url)
Chapter 3.

Copy-Augmented Dialogue State Tracking

Chien-Sheng Wu, Andrea Madotto, Ehsan Hosseini-Asl, Caiming Xiong, Richard Socher, and Pascale Fung. Transferable Multi-Domain Dialogue State Generators for Task-Oriented Dialogue Systems. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL) 2019.
Usr: Find me a cheap restaurant at 7 pm.
Sys: What cuisine would you like?
Usr: I’d prefer eating sushi or ramen.
Sys: Where should it be?
Usr: Let’s do in causeway bay. Also I need a taxi to go there at 6:30 pm.
Sys: ...?
>User: Find me a cheap restaurant at 7 pm.
System: What cuisine would you like?
>User: I’d prefer eating sushi or ramen.
System: Where should it be?
>User: Let’s do in causeway bay. Also I need a taxi to go there at 6:30 pm.
System: ...?
Usr: Find me a cheap restaurant at 7 pm.
Sys: What cuisine would you like?
Usr: I’d prefer eating **sushi or ramen**
Sys: Where should it be?
Usr: Let’s do in causeway bay. Also I need a taxi to go there at 6:30 pm.
Sys: ...?

**Intuition**

- No need dialogue ontology
- Tracking slot values effectively
- Enable to track unseen slot values
- Have domain sharing capacities

**Dialogue History**

- Usr: Find me a cheap restaurant at 7 pm.
- Sys: What cuisine would you like?
- Usr: I’d prefer eating **sushi or ramen**
- Sys: Where should it be?
- Usr: Let’s do in causeway bay. Also I need a taxi to go there at 6:30 pm.
- Sys: ...?

**State Generator**

**Restaurant & Cuisine**

**Japanese**
Intuition

- No need dialogue ontology
- Tracking slot values effectively
- Enable to track unseen slot values
- Have domain sharing capacities

Usr: Find me a cheap restaurant at 7 pm.
Sys: What cuisine would you like?
Usr: I’d prefer eating sushi or ramen.
Sys: Where should it be?
Usr: Let’s do in causeway bay. Also I need a taxi
to go there at 6:30 pm.
Sys: ...?

State Generator
6:30 pm

Taxi & Time

Dialogue History
Intuition

- No need dialogue ontology
- Tracking slot values effectively
- Enable to track unseen slot values
- Have domain sharing capacities

**Dialogue History**

*Usr:* Find me a cheap restaurant at 7 pm.
*Sys:* What cuisine would you like?
*Usr:* I’d prefer eating sushi or ramen.
*Sys:* Where should it be?
*Usr:* Let’s do in causeway bay. Also I need a taxi to go there at 6:30 pm.
*Sys:* ...?
Sequence-to-Sequence (Seq2Seq)

\[ P_{jk}^{\text{vocab}} = \text{Softmax}(E(h_{jk}^{\text{dec}})^T) \in \mathbb{R}^{V}, \]

Vocabulary Distribution

Chapter 3: Copy-Augmented Dialogue State Tracking
Seq2Seq with Attention

\[ P_{jk}^{\text{history}} = \text{Softmax}(H_t(h_{jk}^{\text{dec}})^T) \in \mathbb{R}^{X_t}. \]
Seq2Seq with (Soft) Copy Mechanism

(See et al. 2017)

\[ p_{jk}^{\text{gen}} = \text{Sigmoid}(W_1[h_{jk}^{\text{dec}}, w_{jk}, c_{jk}]) \in \mathbb{R}^1, \]
**TRAnsferable Dialogue stateE generator** (TRADE) (Wu et al., 2019)

\[ G_j = \text{Softmax}(W_g \cdot (c_{j0})^T) \in \mathbb{R}^3, \]

\[ p_{j0}^{\text{final}} \times (1 - p_{j0}^{\text{gen}}), \]

\[ \times (p_{j0}^{\text{gen}}) \]

### (1) Utterance Encoder

**Bot:** Which area are you looking for the hotel?

**User:** There is one at east town called Ashley Hotel.

### (2) Slot Gate

- Context Vector: Context Vector \( c_{j0} \)
- **Slot Gate** \( G_j \)
  - PTR
  - DONT CARE
  - NONE

### (3) State Generator

**Domains**
- Hotel, Train,
- Attraction,
- Restaurant, Taxi

**Slots**
- Price, Area, Day,
- Departure, name,
- LeaveAt, food, etc.

---

**Chapter 3: Copy-Augmented Dialogue State Tracking**
MultiWOZ Dataset (Budzianowski 2018)

- The largest human-human conversational corpus with DST labels (8438 dialogues with average 13.68 turns).
- 5 domains (Hotel, Train, Attraction, Restaurant, Taxi) and 16 slots (food, leave at, area, etc).
- Total 30 domain-slot pairs and ~4500 slot values.

| Slots               | Hotel        | Train        | Attraction   | Restaurant   | Taxi          |
|---------------------|--------------|--------------|--------------|--------------|---------------|
| price, type, parking, stay, day, people, area, stars, internet, name | destination, departure, day, arrive by, leave at, people | area, name, type | food, price, area, name, time, day, people | destination, departure, arrive by, leave by |
| Train Valid Test    | 3381, 416, 394 | 3103, 484, 494 | 2717, 401, 395 | 3813, 438, 437 | 1654, 207, 195 |
Multi-Domain Joint Training

MDBG (Ramadan et al., 2018)
GLAD (Zhong et al., 2018)
GCE (Nouri et al., 2018)
SpanPtr (Xu et al., 2018)
Multi-Domain Joint Training: Visualization
Table 3: Zero-shot experiments on an unseen domain. In *taxi* domain, our model achieves 60.58% joint goal accuracy without training on any samples from *taxi* domain. *Trained Single* column is the results achieved by training on 100% single-domain data as a reference.
Unseen Domain Testing (Zero-Shot): Correctness Analysis

Chapter 3: Copy-Augmented Dialogue State Tracking
Unseen Domain Testing (Few-Shot 1% data)
TRADE is a simple model that leverages its slot-gate and copy mechanism to track slot values without a predefined ontology. It is the current SOTA model in multi-domain DST because of the domain-sharing ability. It also enables zero-shot and few-shot DST in an unseen domain.
End-to-End Dialogue Systems

▷ Require less human effort, making dataset collection easier
▷ No more module dependency
▷ Learn dialogue states automatically
▷ Better generalization ability

Figure 2.1: Block diagram of modularized task-oriented dialogue systems.
End-to-End Dialogue Systems

Good morning!

Hello, what can I help you today?

Can you find me a pizza restaurant?

| Point of interest (poi)     | Distance | Traffic info     | Poi type           | Address       |
|-----------------------------|----------|------------------|--------------------|---------------|
| The Westin                  | 5 miles  | moderate traffic | rest stop          | 329 El Camino Real |
| Round Table                 | 4 miles  | no traffic       | pizza restaurant   | 113 Anton Ct  |
| Mandarin Roots              | 5 miles  | no traffic       | chinese restaurant | 271 Springer Street |
| Palo Alto Cafe              | 4 miles  | moderate traffic | coffee or tea place| 436 Alger Dr   |
| Dominos                     | 6 miles  | heavy traffic    | pizza restaurant   | 776 Arastradero Rd |
| Stanford Express Care       | 6 miles  | no traffic       | hospital           | 214 El Camino Real |
| Hotel Keen                  | 2 miles  | heavy traffic    | rest stop          | 578 Arbol Dr   |

29
End-to-End Dialogue Systems

Good morning!

Hello, what can I help you today?

Can you find me a pizza restaurant?

There are two nearby, Round Table and Dominos, which one do you like?

The closest one please.

| Point of interest (poi)      | Distance | Traffic info | Poi type                | Address            |
|------------------------------|----------|--------------|-------------------------|--------------------|
| The Westin                   | 5 miles  | moderate traffic | rest stop               | 329 El Camino Real |
| Round Table                  | 4 miles  | no traffic   | pizza restaurant        | 113 Anton Ct       |
| Mandarin Roots               | 5 miles  | no traffic   | chinese restaurant      | 271 Springer Street|
| Palo Alto Cafe               | 4 miles  | moderate traffic | coffee or tea place    | 436 Alger Dr       |
| Dominos                      | 6 miles  | heavy traffic | pizza restaurant        | 776 Arastradero Rd |
| Stanford Express Care        | 6 miles  | no traffic   | hospital                | 214 El Camino Real |
| Hotel Keen                   | 2 miles  | heavy traffic | rest stop               | 578 Arbol Dr       |
End-to-End Dialogue Systems

**Good morning!**

**Hello, what can I help you today?**

**Can you find me a pizza restaurant?**

**There are two nearby, Round Table and Dominos, which one do you like?**

**The closest one please.**

**Round Table is 4 miles away.**
E2E New Challenges

▷ Need more data!

▷ Hard to memorize long dialogue history
  ○ Dependency of a multi-turn conversation

▷ Hard to incorporate KB information
  ○ Usually is large and dynamic
  ○ Structured data format

▷ Hard to generate real entities
  ○ Entities are rare words, but usually are the most important information
Memory-Augmented Neural Network (MANN)

- Input Module sends encoded input (usually called memory query) to Controller
- Controller read or write information from External Memory
- Controller sends results to Inference Module, which does some high-level computation
- Inference Module sends results to Output Module

Figure 2.5: Block diagram of general memory-augmented neural networks.
Chapter 4.
End-to-End Retrieval-based Memory-augmented Dialogue Learning

▷ Chien-Sheng Wu*, Andrea Madotto*, Genta Winata, and Pascale Fung. End-to-end Recurrent Entity Network for Entity-value Independent Goal-oriented Dialog Learning. In Dialog System Technology Challenges Workshop 2017.
▷ Chien-Sheng Wu, Andrea Madotto, Genta Winata, and Pascale Fung. End-to-End Dynamic Query Memory Network for Entity-Value Independent Task-Oriented Dialog. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2018.
E2E Retrieval-based Dialogue Systems

▷ Given dialogue history and KB information, a model needs to predict the correct response from a predefined response candidates.

▷ Evaluation Metrics:
  ○ Per-response accuracy
  ○ Per-dialogue accuracy
  ○ Word error rate

KB Information → Dialogue History → Response Candidate Pool → System Response

Chapter 4: Retrieval-based Memory-augmented Dialogue Learning
Recurrent Entity Networks (REN) (Henaff et al., 2017)

REN is one of the MANNs, it was proposed for QA tasks. We are the first to apply it in dialogue applications and won the 2nd place in the E2E track of Dialogue System Technology Challenge in 2017.

Figure 4.1: Entity-value independent recurrent entity network for goal-oriented dialogues. The graphic on the top right shows the detailed memory block.
Learning End-to-End Goal-Oriented Dialog (Borde et al. 2017)

- It is the first paper for E2E task-oriented dialogue systems.
- It used end-to-end memory networks to retrieve responses.
- It proposed a simulated dataset (bAbI Dialogue) is proposed, with five sub-tasks.

![Diagram of dialogue tasks]

**Chapter 4: Retrieval-based Memory-augmented Dialogue Learning**
End-to-End Memory Network (MN) (sukhbaatar 2015)

- One of the most common MANNs
- It is end-to-end differentiable through backpropagation
- It shows impressive results in QA tasks with the multi-hop reasoning ability.

Figure 2.6: The architecture of end-to-end memory networks
End-to-End Memory Network (MN)
(sukhbaatar 2015)

1. Convert input to memories $x_i \rightarrow m_i$
   \[ m_i = Ax_i \]

2. Transform query $q$ into same representation space
   \[ u = Bq \]

3. Output Vectors $x_i \rightarrow c_i$
   \[ c_i = Cx_i \]
End-to-End Memory Network (MN) (sukhbaatar 2015)

4. Scoring memories against query

\[ p_i = \text{Softmax}(u^T m_i) \]
End-to-End Memory Network (MN) (sukhbaatar 2015)

5. Generate output

\[ o = \sum_{i} p_i c_i \]

Weighted average of all inputs (transformed)
End-to-End Memory Network (MN)
(sukhbaatar 2015)
An example on question-answering task:

| Story (16: basic induction) | Support | Hop 1 | Hop 2 | Hop 3 |
|-----------------------------|---------|-------|-------|-------|
| Brian is a frog.            | yes     | 0.00  | 0.98  | 0.00  |
| Lily is gray.               |         | 0.07  | 0.00  | 0.00  |
| Brian is yellow.            | yes     | 0.07  | 0.00  | 1.00  |
| Julius is green.            |         | 0.06  | 0.00  | 0.00  |
| Greg is a frog.             | yes     | 0.76  | 0.02  | 0.00  |

What color is Greg? Answer: yellow Prediction: yellow
Dynamic Query Memory Networks (DQMN) (Wu et al., 2018)

▷ **Drawbacks of MN**
No dependency between memory slots, but it is crucial in dialogue applications.

▷ **Proposed solution**
Simply add a recurrent component between hops, which can model sequential dependencies, and further updating query vectors.

\[ u^{k+1} = u^k + o^k + h_N^k, \quad q_i^{k+1} = u^{k+1} + h_i^k, \]
Dynamic Query Memory Networks (DQMN) (Wu et al., 2018)
Recorded Delexicalization Copying (RDC)

| LOC1 | NUM1 | CUI1  | LOC2 |
|------|------|-------|------|
| Paris| Two  | British| Rome |

API_call Rome Two British

API_call #LOC2 #NUM1 #CUI1

Dialogue History
Usr: Book a table in Paris for two.
Sys: Any preference on type of cuisine?
Usr: British, actually I’d prefer in Rome.
Sys: ok let me look into some options.

Dialogue History
Usr: Book a table in #LOC1 for #NUM1.
Sys: Any preference on type of cuisine?
Usr: #CUI1, actually I’d prefer in #LOC2.
Sys: ok let me look into some options.

Chapter 4: Retrieval-based Memory-augmented Dialogue Learning
Results on bAbI Dialogue Task 5 (Full Dialogue)
MN can effectively model external knowledge. DQMN extends and outperforms MN with recurrent components to model dialogue sequential dependencies. RDC is able to simplify the task and mitigate the out-of-vocabulary problem.
Problems of Retrieval-based Systems

▶ Retrieved responses may be too regular and limited
  ○ “Let me find some options for you”
  ○ “What do you think about this option?”

▶ Minor entity difference is hard to distinguish
  ○ “API_call Rome Two British”
  ○ “API_call Rome Four British”
  ○ “API_call Rome Two Japanese”

▶ RDC is too idealistic and it could also lose some implied information
  ○ When user asks for a “French” restaurant, it may imply the price of the dinner is expensive.
Chapter 5.

End-to-End Generation-based Memory-augmented Dialogue Learning

▷ Chien-Sheng Wu*, Andrea Madotto*, and Pascale Fung. Mem2Seq: Effectively Incorporating Knowledge Bases into End-to-End Task-Oriented Dialog Systems. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL) 2018.

▷ Chien-Sheng Wu, Richard Socher, and Caiming Xiong. Global-to-local Memory Pointer Networks for Task-Oriented Dialogue. In Proceedings of the 7th International Conference on Learning Representations (ICLR) 2019.
Generation-based Dialogue Systems

▷ Given dialogue history and KB information, a model needs to generate the correct response word by word.
▷ Evaluation Metrics:
  ○ BLEU score
  ○ Entity F1 score
  ○ Human evaluation
Intuition: Naive Way

Encoder

KB Information
Dialogue History

Decoder

System Response

Attention

Too Long Sequence (Slow)
Intuition: **Memory-to-Sequence** (Mem2Seq) (Wu*, Madotto*, and Fung, 2018)
Mem2Seq Encoder

Encoder is a standard memory network, encoding dialogue history.

$O^K$ is the vector that encodes the input sequence, and it is gonna be the initial hidden state for the decoder.
Mem2Seq Decoder

An RNN is used to generate query vectors, denoted with $h_t$

We take the memory attention at the last hop to copy words.

We also generate a distribution over the vocabulary using the first hop.
Mem2Seq Decoder

To select whether to copy from the memory or to generate from the vocabulary, we use a **SENTINEL**.

Sentinel is a special memory that is used as an **hard gate** but it is trained automatically as a soft one.

The model is trained end-to-end by optimizing the sum of two cross-entropy loss: $L_{ptr} + L_{vocab}$.

![Hard-Gate Copying Diagram](image-url)
Memory-to-Sequence (Mem2Seq)
(Wu*, Madotto*, and Fung, 2018)

Figure 5.1: The proposed Mem2Seq architecture for task-oriented dialogue systems. (a) Memory encoder with three hops and (b) memory decoder over two-step generation.
Evaluation: Three Datasets

- bAbI Dialogue (simulated dialogue)
- DSTC2 (human-machine dialogue)
- In-Car Assistant (human-human dialogue)

| Task                | 1   | 2   | 3   | 4   | 5   | DSTC2 | In-Car |
|---------------------|-----|-----|-----|-----|-----|-------|--------|
| Avg. User turns     | 4   | 6.5 | 6.4 | 3.5 | 12.9| 6.7   | 2.6    |
| Avg. Sys turns      | 6   | 9.5 | 9.9 | 3.5 | 18.4| 9.3   | 2.6    |
| Avg. KB results     | 0   | 0   | 24  | 7   | 23.7| 39.5  | 66.1   |
| Avg. Sys words      | 6.3 | 6.2 | 7.2 | 5.7 | 6.5 | 10.2  | 8.6    |
| Max. Sys words      | 9   | 9   | 9   | 8   | 9   | 29    | 87     |
| Pointer Ratio       | .23 | .53 | .46 | .19 | .60 | .46   | .42    |
| Vocabulary          | 3747|     |     |     |     | 1229  | 1601   |
| Train dialogues     | 1000|     |     |     |     | 1618  | 2425   |
| Val dialogues       | 1000|     |     |     |     | 500   | 302    |
| Test dialogues      | 1000 + 1000 OOV |     |     |     |     | 1117  | 304    |
Results: bAbI Dialogue

Chapter 4: Generation-based Memory-augmented Dialogue Learning
Chapter 4: Generation-based Memory-augmented Dialogue Learning

Results: bAbI Dialogue

Non-Copy v.s. Copy

Task 3 OOV (Options Recommendation)

Task 5 OOV (Full Dialogue)

Non-Copy v.s. Copy
Results: DSTC2 & In-Car Assistant

**DSTC2**

| Model                  | Entity F1 |
|------------------------|-----------|
| Seq2Seq                | 69.7      |
| Seq2Seq+Attn           | 67.1      |
| Ptr-Unk                | 71.6      |
| Mem2Seq H1             | 72.9      |
| Mem2Seq H3             | 75.3      |
| Mem2Seq H6             | 72.8      |

**In-Car Assistant**

| Model                  | Entity F1 |
|------------------------|-----------|
| Seq2Seq                | 10.3      |
| Seq2Seq+Attn           | 19.9      |
| Ptr-Unk                | 22.7      |
| Mem2Seq H1             | 32.4      |
| Mem2Seq H3             | 33.4      |
| Mem2Seq H6             | 23.6      |
Results: Training Speed

![Graph showing training speed comparison between Mem2Seq H6, Seq2Seq, Seq2Seq+Attn, and Ptr-Unk models. The x-axis represents the maximum input length (# tokens), and the y-axis represents time per epoch (minutes). The graph compares different model configurations across various data sets and input lengths.](image_url)
What are the directions to the closest parking_garage?

**GOLD:** the closest parking_garage is civic_center_garage located 4_miles away at 270_altaire_walk

**GEN:** the closest parking_garage is civic_center_garage at 270_altaire_walk 4_miles away through the directions
What are the directions to the closest parking garage?

**GOLD:** the closest parking garage is civic_center_garage located 4 miles away at 270_altaire_walk

**GEN:** the closest parking garage is civic_center_garage at 270_altaire_walk 4 miles away through the directions
Mem2Seq is the first model to combine multi-hop memory attention with the idea of the copy mechanism. Mem2Seq can be trained 3-5x faster and achieve SOTA results (Oct, 2018).
Two Main Weaknesses of Mem2Seq

▷ **Wrong entity copying**
  ○ Only 33.4% entity F1 score in In-Car Assistant

▷ **Not fluent response**
  ○ Cannot balance well between generation and copying well enough

---

Table 5.9: Example of generated responses for the In-Car Assistant on the navigation domain.

| distance | traffic info | pol type         | address          | pol  |
|----------|--------------|------------------|------------------|------|
| 3 miles  | no traffic   | chinese restaurant | 593 Arrowhead Way | Chef Chu’s |
| 1 miles  | no traffic   | chinese restaurant | 669 El Camino Real | P.F. Changs |
| 5 miles  | road block nearby | gas station     | 200 Alester Ave  | Valero |
| 3 miles  | no traffic   | pizza restaurant  | 528 Anton Ctr   | Pizza My Heart |
| 4 miles  | heavy traffic | rest stop         | 753 University Ave | Comfort Inn |
| 5 miles  | heavy traffic | chinese restaurant | 842 Arrowhead Way | Panda Express |
| 2 miles  | heavy traffic | pizza restaurant  | 704 El Camino Real | Pizza Hut |
| 6 miles  | no traffic   | friends house     | 864 Almanor Ln   | jacks house |

**Table 5.5: Mem2Seq evaluation on human-human In-Car Assistant dataset.**

|        | BLEU | Ent. F1 | Sch. F1 | Wea. F1 | Nav. F1 |
|--------|------|---------|---------|---------|---------|
| Human  | 13.5 | 60.7    | 64.3    | 61.6    | 55.2    |
| Rule-Based | 6.6  | 43.8    | 61.3    | 39.5    | 40.4    |
| Seq2Seq| 8.4  | 10.3    | 09.7    | 14.1    | 07.0    |
| +Attn  | 9.3  | 19.9    | 23.4    | 25.6    | 10.8    |
| Ptr-Unk| 8.3  | 22.7    | 26.9    | 26.7    | 14.9    |
| Mem2Seq H1 | 11.6 | 32.4    | 39.8    | 33.6    | 24.6    |
| Mem2Seq H3 | 12.6 | 33.4    | 49.3    | 32.8    | 20.0    |
| Mem2Seq H6 | 9.9  | 23.6    | 34.3    | 33.0    | 4.4     |

**Table 5.5 (continued):**

**Human**
- **Seq2Seq**
  - +Attn
  - Ptr-Unk
- **Gold**
  - The nearest gas station is located 5 miles away. Need more info?
Intuition: Mem2Seq

Chapter 4: Generation-based Memory-augmented Dialogue Learning
Intuition: **Global-to-Local Memory Pointer Network (GLMP)** (Wu et al., 2019)

### Encoder
- Dialogue History

### Decoder
- Dialogue History
- System Response

- **External Knowledge**
  - KB Information
  - Dialogue History

- **Response Sketching**
Global-to-local Memory Pointer Network (GLMP) (Wu et al., 2019)

(a) Global memory encoder

(b) Local memory decoder
GLMP: Global Memory Encoder

- **Context RNN**
  - Encode plain text dialogue history
  - Query external knowledge

- **Contextual Dialogue History**
  - Write hidden states into dialogue memory
  - Mitigate OOV copying problem

- **Global Memory Pointer**
  - Point to all the words that may appear in the system response.
  - Multi-label classification

\[ g_i = \text{Sigmoid}\left((q^K)^T c^K_i\right) \]
GLMP: Local Memory Decoder

- **Sketch RNN**
  - Generate sketch response with unfilled slots
  - ex: @poi is @distance away
  - No sentinel needed

- **Local Memory Pointer**
  - Filter memory attention by global memory pointer
  - Copy one single word at each time step

- **Record Function**
  - Mask the copied words

Mathematical equations:

\[ h_t^d = GRU(C^1(\hat{y}_{t-1}^s), h_{t-1}^d), \quad P_t^{vocab} = \text{Softmax}(Wh_t^d) \]

\[ p_t^k = \text{Softmax}((q^k)^T c_t^k), \]

\[ \hat{y}_t = \begin{cases} \arg \max(P_t^{vocab}) & \text{if } \arg \max(P_t^{vocab}) \notin ST, \\ \text{Object}(m_{\arg \max(L_t \odot R)}) & \text{otherwise} \end{cases} \]
GLMP: Workflow

System Response: Valero is 3 miles away

Chapter 4: Generation-based Memory-augmented Dialogue Learning
## Results: In-Car Assistant

### Automatic Evaluation

|                | Rule-Based* | KVR* | S2S | S2S + Attn | Ptr-Unk | Mem2Seq | GLMP H1 | GLMP H3 | GLMP H6 |
|----------------|-------------|------|-----|------------|---------|---------|---------|---------|---------|
| BLEU           | 6.6         | 13.2 | 8.4 | 9.3        | 8.3     | 12.6    | 13.83   | **14.79** | 12.37   |
| Entity F1      | 43.8        | 48.0 | 10.3| 19.9       | 22.7    | 33.4    | 57.25   | **59.97** | 53.54   |
| Schedule F1    | 61.3        | 62.9 | 9.7 | 23.4       | 26.9    | 49.3    | 68.74   | **69.56** | 69.38   |
| Weather F1     | 39.5        | 47.0 | 14.1| 25.6       | 26.7    | 32.8    | 60.87   | **62.58** | 55.89   |
| Navigation F1  | 40.4        | 41.3 | 7.0 | 10.8       | 14.9    | 20.0    | 48.62   | **52.98** | 43.08   |

### Human Evaluation

|                | Mem2Seq | GLMP | Human |
|----------------|---------|------|-------|
| Appropriate    | 3.89    | 4.15 | 4.6   |
| Humanlike      | 3.80    | 4.02 | 4.54  |
Results: Ablation Study

- **Ablation of contextual dialogue history (w/o H)**
  - Hidden states of context RNN are not written into the external knowledge
  - The performance drop is serious in bAbI OOV scenario. (-5.3% in T5)

- **Ablation of global memory pointer (w/o G)**
  - The external knowledge is not filtered by the global memory pointer
  - The performance drop is serious in SMD human-human scenario. (-11.47%)
Results:

Qualitative Study
Results: Qualitative Study
GLMP introduces the concept of response sketching (two-stage generation) and double pointers copying (global and local memory pointers). It is the current SOTA model of end-to-end dialogue response generation.
Chapter 6.
Conclusion
We leverage memory-augmented neural networks and copy mechanism for better information memorization in neural task-oriented dialogue systems.

First

We propose TRADE, a copy-augmented state generator, for multi-domain and unseen-domain dialogue state tracking.

Second

We propose DQMN, a memory-augmented neural network, for end-to-end retrieval-based dialogue systems.

Last

We propose Mem2Seq and GLMP, two memory-augmented sequence generators, for end-to-end generation-based dialogue systems.
Selected Publications:

▷ “Getting To Know You: User Attribute Extraction from Dialogues,” **Chien-Sheng Wu**, A Madotto, Z Lin, X Peng, P Fung. EMNLP 2019 (Under Review).

▷ “Transferable Multi-Domain State Generator for Task-Oriented Dialogue Systems,” **Chien-Sheng Wu**, A Madotto, E Hosseini-Asl, C Xiong, R Socher, P Fung. **ACL 2019** (long).

▷ “Global-to-local Memory Pointer Networks for Task-Oriented Dialogue,” **Chien-Sheng Wu**, C Xiong, R Socher. **ICLR 2019**.

▷ “Mem2Seq: Effectively Incorporating Knowledge Bases into End-to-End Task-Oriented Dialog Systems,” **Chien-Sheng Wu***, A Madotto*, P Fung. **ACL 2018** (long).

▷ “End-to-End Dynamic Query Memory Network for Entity-Value Independent Task-oriented Dialog,” **Chien-Sheng Wu**, A Madotto, G Winata, P Fung. **ICASSP 2018**.

▷ “End-to-End Recurrent Entity Network for Entity-Value Independent Goal-Oriented Dialog Learning,” **Chien-Sheng Wu***, A Madotto*, G Winata, P Fung. **DSTC 2017**.

Check my google scholar for other works: [https://scholar.google.com.hk/citations?user=1G4GV2EAAAAJ&hl=en](https://scholar.google.com.hk/citations?user=1G4GV2EAAAAJ&hl=en)
Thanks!
Any questions?

You can find me at: jason.wu@connect.ust.hk

More information about me: https://jasonwu0731.github.io/