A Reducing Sequential Intruction Understanding to Conversational Machine Comprehension

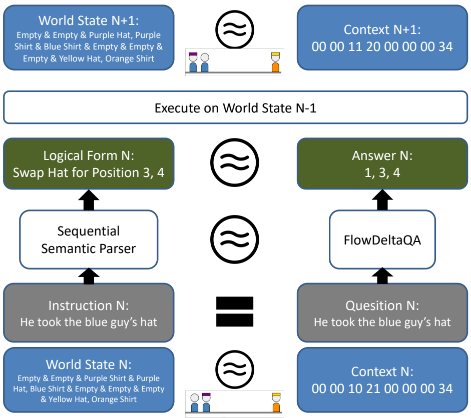
World State N+1:
Empty & Empty & Purple Hat, Purple Shirt & Blue Shirt & Empty & Empty & Empty & Yellow Hat, Orange Shirt
≈
Context N+1:
00 00 11 20 00 00 00 34
Execute on World State N-1
Logical Form N:
Swap Hat for Position 3, 4
≈
Answer N:
1, 3, 4
Sequential Semantic Parser
≈
FlowDeltaQA
Instruction N:
He took the blue guy's hat
=
Quesition N:
He took the blue guy's hat
World State N:
Empty & Empty & Purple Shirt & Purple Hat, Blue Shirt & Empty & Empty & Empty & Yellow Hat, Orange Shirt
≈
Context N:
00 00 10 21 00 00 00 34

Figure 3: Example of the SCONE dataset and its reduction

In SCONE dataset, given the initial world state $\boldsymbol{W}_0$ and a sequence of natural language insturctions $\{\boldsymbol{I_1}, \ldots \boldsymbol{I_K}\}$, the model need to perform the correct sequence of actions on $\boldsymbol{W}_0$ and obtain the correct world states $\{\boldsymbol{W_1}, \ldots, \boldsymbol{W_K}\}$ after each instruction. An example from (Long et al., 2016) is shown in the left-hand side of Figure 3.

Following FlowQA (Huang et al., 2018), for each position in the world state, we encode it as two integers denoting the shirt and hat color in SCENE, image ID and present or not in TANGRAMS, and color of the liquid and number of units in ALCHEMY. Next, the change of world states (i.e., the logical form) is encoded as three or four integers. The first integers is the type of action performed. The second and third integers represent the position of the context (i.e., the encoded world state). Finally, the fourth integer represents the additional property for the action such as the number of units moved.

An example of encoded world states and logical form is shown in the right-hand side of Figure 3. In this example, action $(1, 3, 4)$ means "swap the hat for position 3, 4" and there is no additional property for the action.

B Experimental Details

We reproduce and report the experiment results of FlowQA using the released code except SCONE part since the official released code does not contain it. Authors claim there is further performance improvement on the released version of FlowQA

All hyperparameters are kept the same as recommended one in FlowQA and BERT for CoQA and QuAC datasets. For SCONE, due to the relatively small size of dataset, to prevent overfitting we further tune the hidden size of FLOWDELTAQA in three different domains. The tuned hidden sizes are $50, 60, 70$ for SCENE, ALCHEMY and TANGRAMS respectively.

C Flow Information Gain Variants

We test three different variants of FLOWDELTA on modeling the information flow in the dialog and show results in table 4. The three variants are:

1. SkipDelta: $h_{t-1} - h_{t-3}$
2. DoubleDelta: $[h_{t-1} - h_{t-2}; h_{t-2} - h_{t-3}]$
3. Hadamard Product: $h_{t-1} * h_{t-2}$

The reason to use SkipDelta and DoubleDelta is because we want to see if there is any benefit to incorporate longer (or more) dialog history. Experiment results show while using longer dialog history (i.e., SkipDelta) helps, adding too many dialog history (i.e., DoubleDelta) does not give any improvement.

The intuition behind Hadamard product is to model the similarity of consecutive hidden states. If there are any topic shift in last turns of dialog, we expect Hadamard product can give us useful signal to detect it. Results show although the proposed FLOWDELTA is the best, Hadamard product outperforms SkipDelta and DoubleDelta and proves its effectiveness.

Model	F1
FlowQA	76.7
FlowDeltaQA (SkipDelta)	76.9
FlowDeltaQA (DoubleDelta)	76.7
FlowDeltaQA (Hadamard Product)	77.2
FlowDeltaQA	**77.6**

Table 4: CoQA results of different variants of FLOW interaction. All models are provided with previous 1 gold answer.

D Qualitative Analysis

Here we present an example from CoQA dataset which consists of a passage that the dialog talks about, and a sequence of questions and answers.

Questions	FlowQA	FlowDeltaQA	Gold Answer
Whose house was searched?	Gary Giordano		
In what city?	Gaithersburg		
County?	Montgomery County		
State?	Maryland		
Where is he now?	Aruban jail		
Why?	lack of evidence	6 recent disappearance of an American woman	suspect in the recent disappearance of an American woman

Table 5: Qualitative analysis of FlowDeltaQA.

Table 5 shows the questions, answers and model predicitons. We note the gold answer in CoQA is abstractive and may not be a span in the passage. Only a subset of the dialog is showed to demonstrate the different behaviors of FlowQA and FlowDeltaQA.

Context: (CNN) – FBI agents on Friday night searched the Maryland home of the suspect in the recent disappearance of an American woman in Aruba, an agent said. The search is occurring in the Gaithersburg residence of Gary Giordano, who is currently being held in an Aruban jail, FBI Special Agent Rich Wolf told CNN. Agents, wearing vests that said FBI and carrying empty cardboard and plastic boxes, arrived about 8:40 p.m. Friday. About 15 unmarked cars could be seen on the street, as well as a Montgomery County police vehicle. Supervisory Special Agent Philip Celestini, who was at the residence, declined to comment further on the search, citing the active investigation. Aruban Solicitor General Taco Stein said earlier Friday that the suspect will appear in court Monday, where an investigating magistrate could order him held for at least eight more days, order him to remain on the island or release him outright due to a lack of evidence. Giordano was arrested by Aruban police on August 5, three days after Robyn Gardner was last seen near Baby Beach on the western tip of the Caribbean island. Giordano told authorities that he had been snorkeling with Gardner when he signaled to her to swim back, according to a statement. When he reached the beach, Gardner was nowhere to be found, Giordano allegedly said. The area that Giordano led authorities to is a rocky, unsightly location that locals say is not a popular snorkeling spot. Although prosecutors have continued to identify the 50-year-old American man by his initials, GVG, they also released a photo of a man who appears to be Giordano. His attorney, Michael Lopez, also has said that his client is being held as a suspect in Gardner's death. Lopez has not returned telephone calls seeking comment.

Analysis In this example, to answer the last question ”Why?”, model need to understand the previous conversation correctly to know the actual question is ”Why is Gary Giordano in the Aruban Jail now?”. This example is particularly hard since in order to know ”he” in ”Where is he now?” refers to ”Gary Giordano”, model need to use the information from the very first question ”Whose house was searched”, which requires the ability to utilize full dialog history. While FlowQA fails to hook this question to the correct conversation context and respond reasonable but incorrect answer, our FlowDeltaQA successfully grasps long dialog flow and answers the correct span.