Evaluating Co-reference Chains based Conversation History in Conversational Question Answering

Angrosh Mandya, Danushka Bollegala, Frans Coenen

Department of Computer Science, University of Liverpool, Liverpool, UK

Focus

- Evaluate co-reference chains based conversational history vs. using entire conversation history for Conversational Question Answering (CoQA).
- Modified QANet model to include conversational history and used NeuralCoref to obtain co-reference chains based conversation history.

Key Results

- Even though large proportion of co-reference links is available in CoQA, the abstract nature of questions in CoQA renders it difficult to obtain correct mapping of co-reference related conversation history
- The effect of co-reference resolution examined on various domains and different conversation length, shows that co-reference resolution across questions is helpful for certain domains and medium-length conversations.

Introduction

- CoQA dataset has large amounts of co-reference links between questions - almost half of the CoQA questions (49.7%) contain explicit co-reference markers such as *he, she, it*.
- Given context passage c, a question q_i and conversational history $(q_1, a_1, ...q_{i-1}, a_{i-1})$, the task is to predict answer $\hat{a}_i$:

$$p(\hat{a}_i|q_i) = f(c_i, q_1, a_1, ..., q_{i-1}, a_{i-1}) \quad (1)$$

- However, instead of using $(q_1, a_1, ...q_{i-1}, a_{i-1})$, this study proposes to use co-reference chains based conversation history $(q_k, a_k, ...q_{k-1}, a_{k-1})$, defined as the set of previous question-answer pairs that have co-reference links to the current question q_i.

$$p(\hat{a}_i|q_i) = f(c_i, q_k, a_k, ..., q_{k-1}, a_{k-1}) \quad (2)$$

- Given two questions q_i and q_j, we say that there exists a co-reference link between q_i and q_j, if a word $u \in q_i$ refer to the same *person* or *thing* $v \in q_j$.

Modified QANet Model for CoQA

Input Embedding layer

- Concatenate word and character embeddings

Attention layer

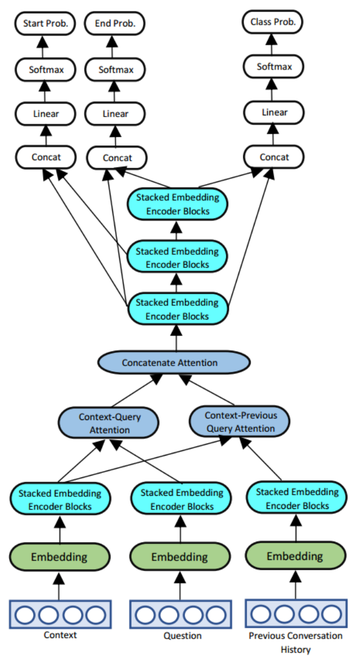
- Compute context-query attention and context-conversation history attention

Model Encoding layer

- The encoder uses attention combined with context to predict output.

Output layer

- Predict start and end probabilities of each position in the context.
- Also predict yes/no type and unknown answers.



Examined Models

- **QANET-1-CCQ** and **QANET-2-CCQ**, model that uses previous one and two co-reference chain linked questions, respectively;
- **QANET-1-CCQA** and **QANET-2-CCQA**, model that uses previous one and two co-reference chain linked questions and answers, respectively;
- **QANET-1-PQA** and **QANET-2-PQA**, that uses previously available one and two questions and answers, respectively;

Results

Performance of Different Models

	Child.	Liter.	Mid-High.	News	Wiki.	Overall
QANET-1-CCQ	62.4	56.7	63.1	66.9	67.4	63.4
QANET-2-CCQ	61.3	57.4	63.5	68.5	69.2	63.9
QANET-1-CCQA	65.7	59.3	64.6	70.2	68.2	65.3
QANET-2-CCQA	**66.8**	**60.1**	62.8	71.5	70.2	66.2
QANET-1-PQA	64.9	57.8	65.8	74.1	73.7	67.2
QANET-2-PQA	65.2	58.9	**66.2**	**75.5**	**73.9**	**67.9**

Table: F1 scores of QANet based models for different domains in CoQA Development Set.

- Using entire previous conversation history is useful compared to co-reference chains based conversation history (performance of QANET-1-PQA and QANET-2-PQA is better than QANET-1-CCQ, QANET-2-CCQ, QANET-1-CCQA, QANET-2-CCQA).
- Co-reference chains helps certain domains E.g.: domains of "Children Stories" and "Literature".

Absence of Contextual Information

	Child.	Liter.	Mid-High.	News	Wiki.	Total
TQ	1425	1630	1653	1649	1626	7983
TQ_coref_links	1181	1274	1385	1313	1223	6376
(%)	82.87	78.15	83.78	79.62	79.33	80.70

Table: Number of co-reference chain linked questions for various domains in CoQA Development Set

- Absence of conversation history. Inability to predict previous questions for nearly 20% of the questions in CoQA dataset.

Incorrect contextual Information

	Questions in sequence	Co-reference chains based questions
1.	What was the name of the fish?	-
2.	What looked like a birds belly?	-
3.	Who said that?	-
4.	Was Sharkie a friend?	-
5.	Did they get the bottle?	-
6.	What was in it?	Did they get the bottle?
7.	Did a little boy write the note?	Did they get the bottle?
8.	Who could read the note?	Did they get the bottle?
9.	What did they do with the note	Did they get the bottle?
10.	Did they write back?	Did a little boy write the note? Did they get the bottle?
11.	Were they excited ?	Did a little boy write the note? Did they get the bottle?

Table: Co-reference chains based questions obtained using NeuralCoref for a sample paragraph in domain "Children Stories" in CoQa development set.

- Wrong conversation history. Wrong questions used as conversation history or previous questions are absent.

Conclusion

- Results indicate that resolving co-reference chains alone does not help much in answering CoQA questions.