# HagiwaraLab at the NTCIR-13 QALab-3 Task

### Yuanzhi Ke and Masafumi Hagiwara

### Keio University

enshika8811.a6@keio.jp, hagiwara@keio.jp

## ABSTRACT

- English Term Question Subtask
- RNN-based Approach
- Explicit Summarization to Reduce Computational Cost
- Discussion About the Effects of Explicit Summarization

## CHALLENGES OF THE SUBTASK

- Various format of questions
  - What/Where/Who… for the grand question
  - No What/Where/Who… for the grand question
  - Not for the grand question but another story

Grand Question:
During the first half of the 20th century, the princeples of world peace and a new … (B792W10 The University of Tokyo, 2001)
Instruction:
What is the treaty referred to in the underlined section (2)? Write the name of the treaty.

Grand Question:
Looking back at the history of Western Asia … (B792W10 The University of Tokyo, 2001)
Instruction:
With regard to underlined section (4), from the end of the 19th century there was a political movement among Jews in Europe to return to Palestine. Write the name of this movement.

Figure 1: Question examples.

- Unknown answer length
- Lack of knowledge resource (English Track)
  - Only Wikipedia
    - Not as specified as textbooks
    ➡ More difficult to extract the answer
    ➡ Gold standard may not exists
      = difficult to match exactly
- Lack of Training Data (For neural networks)
- Document (from Wikipedia) Long ( ➡ Memory issues for neural networks)

## METHODOLOGY

**RNN:**

- Works without categorization of the question type
- Works without explicit setting of the output length
- Works without extract entities in advance (though it maybe better)
- Reported effective for other datasets

**For Memory Issues:**

- Long Document ➡ Short Text
- Summary ➡ More Reasonable than Head Tokens



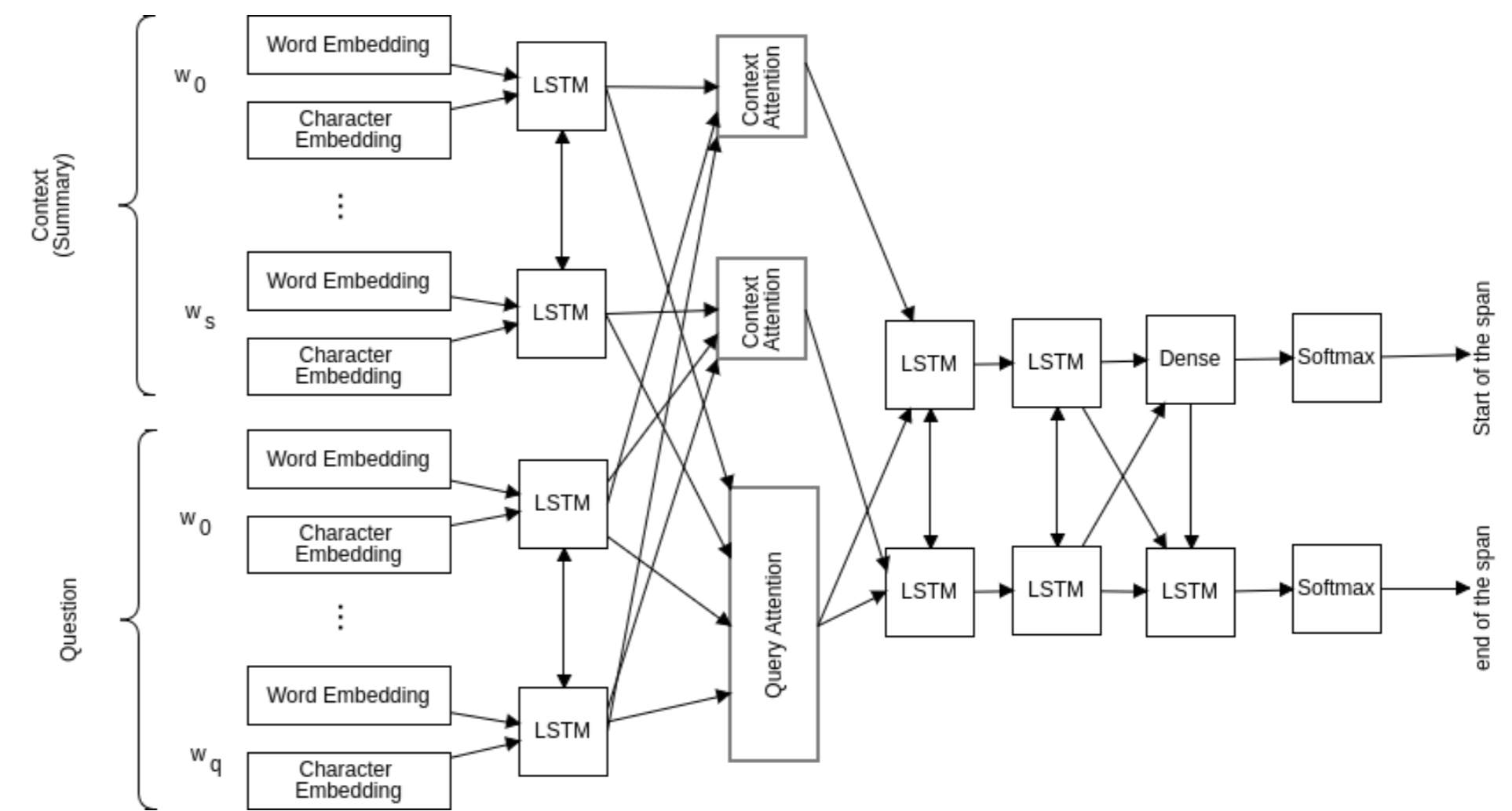Figure 2: The system process flow.



Figure 3: The architecture of the RNN to extract the answer.

**Summarization Method:**

Tf-idf weighted Word Embedding as Sentence Embedding:

$$v_s = \sum_{w \in s} (tfidf(s,w) \circ v_w) \qquad tfidf(s,w) = tf(s,w) \times \left( log\frac{1+n_S}{1+df(S,w)} + 1 \right)$$

Cosine similarity of each sentence and question for ranking:

$$Score = cos \langle v_s, v_q \rangle$$

## VALIDATION EXPERIMENTS

Data: Question Sets of Phase 1 and Phase 2

Metrics:

Correct = the system answer contains any word in the gold standard

Table 1: The number of the correct answer in the validation experiments

| | #Document | Length Limit | #Correct Answer(P1) | #Correct Answer(P1) |
|---|---|---|---|---|
| No Summary | 1 | No limit | Out of Memory | Out of Memory |
| No Summary | 1 | 3200 | 8/69 | **15/78** |
| No Summary | 1 | 1600 | 4/69 | 13/78 |
| Summary(5 Sentences) | 1 | 1600 | 6/69 | 6/78 |
| Summary(10 Sentences) | 1 | 1600 | 5/69 | 7/78 |
| Summary(15 Sentences) | 1 | 1600 | **9/69** | 12/78 |
| Summary(5 Sentences) | 5 | 1600 | 2/69 | 5/78 |
| Summary(10 Sentences) | 5 | 1600 | 3/69 | 9/78 |
| Summary(15 Sentences) | 5 | 1600 | 2/69 | 11/78 |



Figure 3: An example of a summary by the system. We can find less related sentences like the ISBN number are highly ranked.

## CONCLUSIONS

Main Issues:

- Wrongly high-rank less important sentences for question answering
- The answer need to be exists exactly the same in the documents (the bottleneck of the RNN of us)

Secondary Issues:

- Lack of specified knowledge base (another possible reason why the results in Japanese are generally better)
- Lack of training data (we have to use less related external data for traning)

## FUTURE WORKS

- Wrongly high-ranked the meta information
  ➡ Revision of the Ranking Function (e.g. Adding Entity Recognition)
  ➡ Careful preprocessing
- Knowledge Resource
➡ Textbook datasets in the future if possible
- More Powerful Machines