

HagiwaraLab at the NTCIR-13 QALab-3 Task

Keio University



Yuanzhi Ke and Masafumi Hagiwara

Keio University

@NTCIR-13

Dec. 7th ,2017



Abstract

- English Term Question Subtask
- RNN-based Approach
- Explicit Summarization to Reduce Computational Cost

Challenges of the English Term Question Subtask

- Various format of questions
 - What/Where/Who... for the grand question
E.g. What is the treaty referred to in the underlined section (2) ?
(B792W10 - The University of Tokyo, 2001)
 - No What/Where/Who... for the grand question
E.g. Write the name of the accord in underlined section (6)
(B792W10 - The University of Tokyo, 2001)
 - Not for the grand question
E.g. India, which advocated peaceful foreign diplomacy, took a leadership role in the Afro Asian Conference. However, India had violent disagreements, leading to war, with neighboring Pakistan immediately following its independence regarding a certain territory. Write the name of this territory (B792W10 - The University of Tokyo, 2001)

Challenges of the English Term Question Subtask

- 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** exist
 - = difficult to match exactly
- Lack of Training Data (For neural networks)
- Document (from Wikipedia) Long (➡ Memory issues for neural networks)

Our Solution

Various format of questions
Unknown answer length

Recurrent
Neural Network

Long Documents
(Memory Issues)



Summarization

Lack of Training Data



External
Training Data
(with SQuAD)

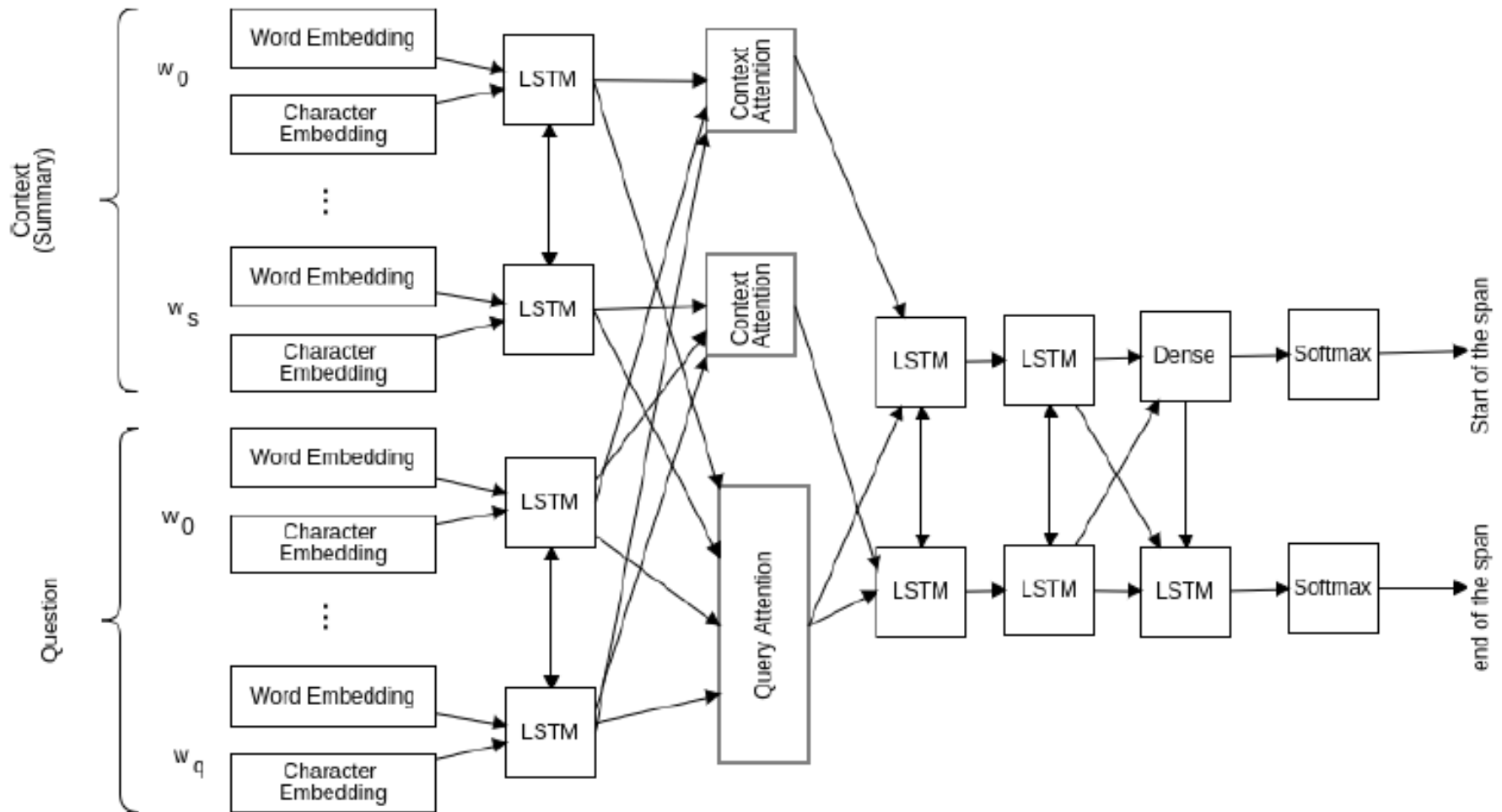
Lack of knowledge resource



Why we used RNN

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

The RNN to Extract Answers



Memory Issues

GPU: GTX980, 4GB GDDR (During the Track)

- Input $>$ 1600 tokens
➔ **Out of Memory**

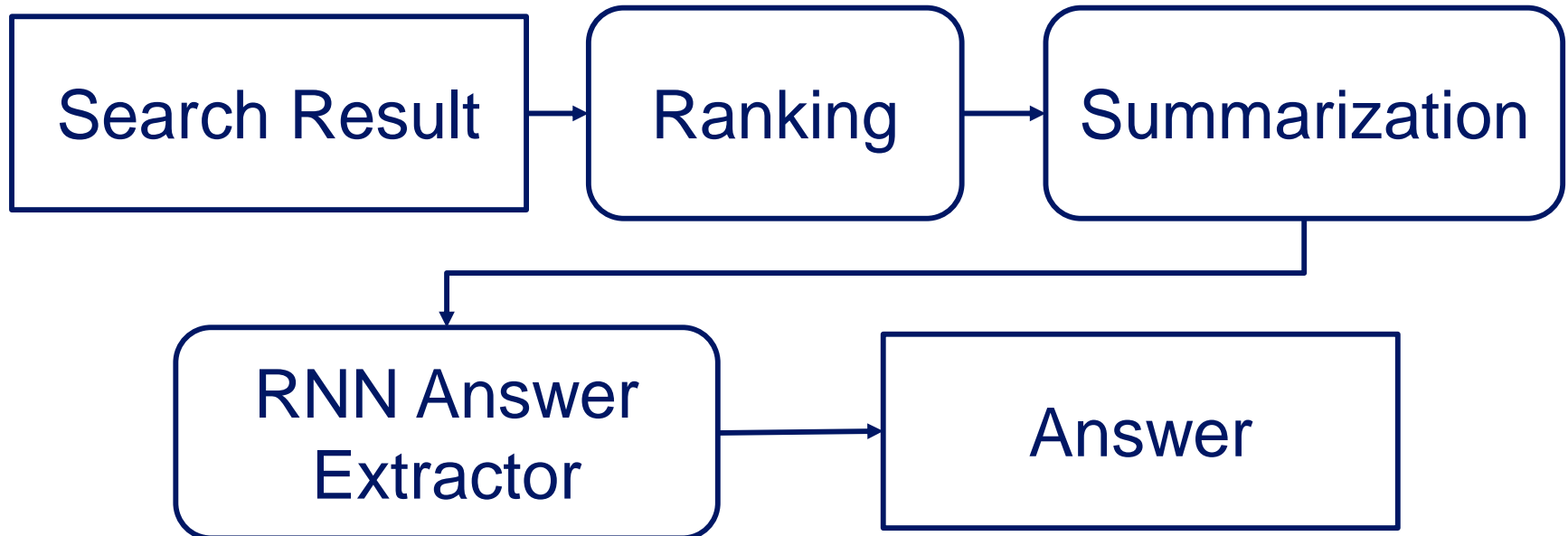
GPU: GTX1080, 8GB GDDR (Now)

- Input $>$ 3200 tokens
➔ **Out of Memory**

Summarization for Memory Issue

- Idea:

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



Summarization Method

- Tf-idf weighted Word Embedding-based:

$$tfidf(s, w) = tf(s, w) \times \left(\log \frac{1 + n_s}{1 + df(S, w)} + 1 \right)$$

$$v_s = \sum_{w \in s} (tfidf(s, w) \circ v_w)$$

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

Effect of Explicit Summarization

Table 1: The number of the correct answers of each runs in the validation.

	#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

For Phase 1:

- ✓ Outperform Head Tokens of the Same
- ✓ Similar to Two Times Longer Head Tokens

For Phase 2:

- × Similar to Head Tokens of the Same Length
- × Bad than Two Times Longer Head Tokens

Note: In our experiments, containing any word in the ground answer = correct

Effect of Explicit Summarization

Table 1: The number of the correct answers of each runs in the validation.

	#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

Other Findings:

- Longer is better
- × Bad for Multiple Documents

Note: In our experiments, containing any word in the ground answer = correct

Weakness on Multi-document

ISBN 0-19-826463-1 ==Reform Judaism in the UK== UK Reform Progressive movements in the UK. migration from the villages to cities by the countrys younger demographic. The government first selected 33,267 villages and provided 335 sacks of cement. ==Reform Judaism in North America== Reform Judaism denomination of American Jews today. Official bodies of the Reform Movement in North America include the Union for Reform Judaism, the Central Conference of American Rabbis, and Hebrew Union College-Jewish Institute of Religion. with a fellow miner]] Soviet history Taylorist efficiencies to over-achieve at work. The Stakhanovite movement was named after Aleksei Stakhanov, who had mined 102 tons of coal in less than 6 hours (14 times his quota). 16,600 villages that demonstrated success were then granted additional resources of 500 sacks of cement and a ton of iron bars. ==Reform movement in Judaism==

The end of the summary of 5 sentences

- **Wrongly** High-ranked Meta Information (ISBN, etc.)
 - ➔ **Too many** of them in the case of multi-document summarization

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 training)

Future Works

- Wrongly high-ranked meta information
 - ➔ Revision of the Ranking Function (e.g. Adding Entity Recognition)
 - ➔ Careful preprocessing
- Knowledge Resource
 - ➔ Access the Japanese textbook data in the future if possible
- More Powerful Machines

Thank you for Listening.