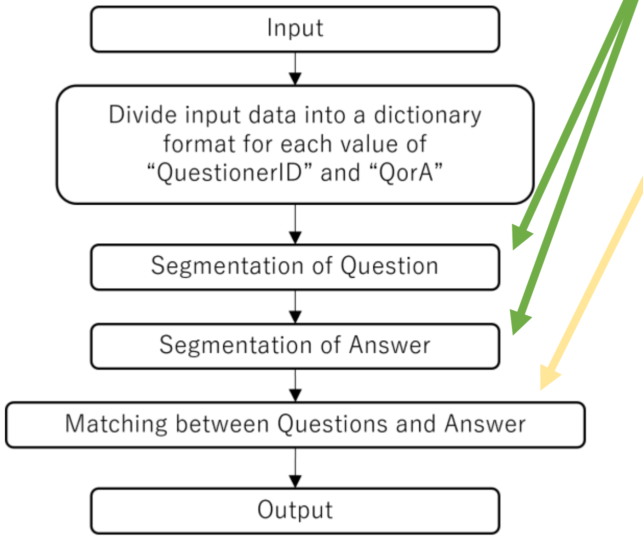
Forst: A Challenge to the NTCIR-16 QA Lab-PoliInfo-3 Task

Naoki Igarashi †[1], Daiki Iwayama †[1], Hideyuki Shibuki †[2], Tatunori Mori †[1]
†[1] Yokohama National University, †[2] Besna Institute Inc.

QA Alignment

Approach

- Our system performs this task in two stages, the **“segmentation”** stage and the **“matching”** stage as shown below.



Input

Divide input data into a dictionary format for each value of “QuestionerID” and “QorA”

Segmentation of Question

Segmentation of Answer

Matching between Questions and Answer

Output

Result

ID	F-measure	Precision	Recall
197 (bf) sw=300	0.77446	0.7854	0.7716
261 (bf+we) sw=200	0.7703	0.7615	0.7837
262 (bf+we) sw=100	0.7699	0.7594	0.7852

sw : Number of stop word

Method

In the segmentation stage, the statements of the questioner and the answerer are divided into individual questions and answers. The segmentation is done by a **rule-based approach** using **cue expressions** such as "伺 います“.

In the matching stage , we used similarity to map the individual questions and answers. The similarity is based on the number of overlaps in the **base form of the words(bf)** . We also used the cosine similarity of the **word embedding (we)** of the top 30 tf-idf values as similarity.

Discussion

Compared to the baseline method, our scores improved, but our scores were not as good as the other participants.

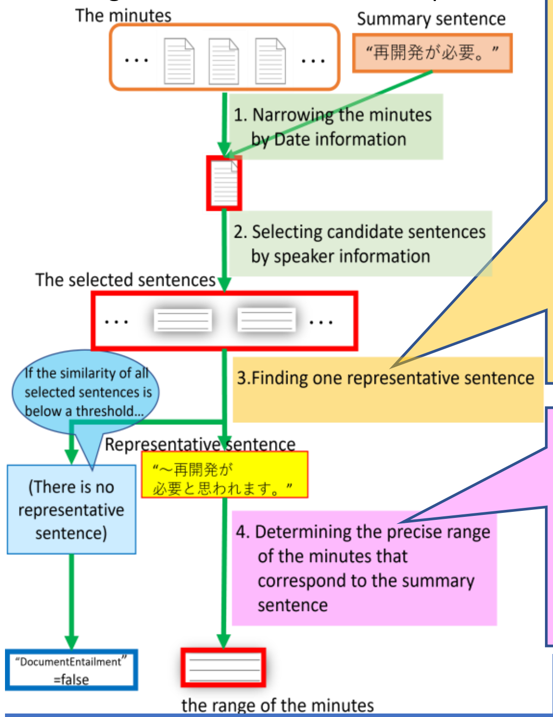
In the segmentation stage, segmentation was done using cue expressions. In terms of the number of segments, the segmentation was relatively correct, especially for answers. In the future, we would like to conduct segmentation that considers not only cue expressions but also topic transitions.

In the matching stage, when the main topic is shared by a group of questions by the same questioner, the main topic is often shared by the corresponding answers. The use of subtopics in determining the degree of similarity may help to identify minor differences.

Fact Verification

Approach

We used a four-step rule-based approach, including sentence-to-sentence comparison.



The minutes

Summary sentence

“再開発が必要。”

1. Narrowing the minutes by Date information

2. Selecting candidate sentences by speaker information

The selected sentences

3.Finding one representative sentence

If the similarity of all selected sentences is below a threshold...

Representative sentence

“～再開発が必要と思われます。”

(There is no representative sentence)

4. Determining the precise range of the minutes that correspond to the summary sentence

“DocumentEntailment” =false

the range of the minutes

Result

ID:method	F-measure	Precision	Recall
257:Shared noun	0.8040	0.8113	0.8110
292:Shared noun + QandA	0.8389	0.8466	0.8451
338:tf-idf similarity(th=0.50)	0.6857	0.6864	0.6925
339:tf-idf similarity + smoothing(th=0.50)	0.7980	0.7989	0.8065
340:tf-idf similarity(th=0.40)	0.7970	0.7964	0.8058
341:tf-idf similarity + smoothing(th=0.40)	0.8563	0.8591	0.8642

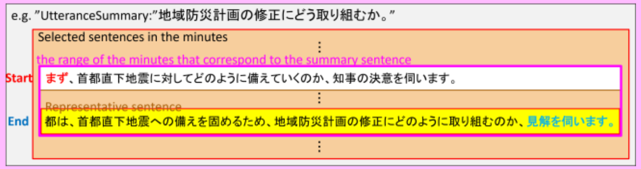
Method

For each selected sentence, We evaluated the similarity to the summary sentence in four viewpoint.

- **Number of shared nouns** between sentences
- Number of shared nouns between **“RelatedUtteranceSummary”** and selected sentences (for the validation of **the relationship between the question and answer** in the representative sentence)
- **Similarity of tf-idf vectors**
- **Smoothing similarity of tf-idf vectors using window function** by considering the similarity of surrounding sentences

The sentence with the largest similarity is adopted as the **representative sentence.**

Find “opening cue expressions” and “ending cue expressions” surrounding the **representative sentences.**



e.g. ”UtteranceSummary:”地域防災計画の修正にどう取り組むか。”

Selected sentences in the minutes

⋮

the range of the minutes that correspond to the summary sentence

Start まず、首都直下地震に対してどのように備えていくのか、知事の決意を伺います。

⋮

Representative sentence

End 都は、首都直下地震への備えを固めるため、地域防災計画の修正にどのように取り組むのか、見解を伺います。

⋮

Discussion

- Relatively good result is obtained even if the number of shared nouns is only taken account of.
- The results of the smoothing showed that the similarity of multiple sentences is more effective than that of single sentence.
- Semantic similarity of words is not considered and will be considered in our future research to improve our method .