Our proposed method are constructed by the following two steps.

**Step1. Alignment step**

Matching the summary of the question utterances with the relevant part of the answer utterances.

- **Paragraph separation method**:
  - ditlab’s method in the QA of QA Lab-Poliinfo-3\(^1\)
- **Model**: BERT\(^2\) (fine-tuned by Sentence-BERT\(^3\))
  - Fine-tuning data: QA Alignment of QA Lab-Poliinfo-3
  - Input: Answer candidates or Question summary
  - Output: Sentence embeddings
- **Matching metrics**: Cosine similarity

**Step2. Summarization step**

Summarizing answer utterances obtained from step 1.

- **Model**: T5 (Text-to-Text Transfer Transformer)\(^4\)
  - Input: Question, SubTopic, and Matched Answer
  - Output: Summary of Answer
- **Same as the baseline method (TO)**

Overview of the alignment step

Overview of the summarization step

**Experimental method**

- Give minutes data to pre-trained BERT\(^2\)
- Fine-tuning to determine correctness
- Split training data: test data = 4 : 1
- Four cross-validation with the training data

**Result**

Result for each input method

<table>
<thead>
<tr>
<th>Input Method</th>
<th>Q + A &amp; Q + Ori</th>
<th>Q + A &amp; Ori</th>
<th>Q + A &amp; Head</th>
<th>Q + A &amp; Sub</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy(_{mean})</td>
<td>0.7483</td>
<td>0.7466</td>
<td>0.6969</td>
<td>0.7021</td>
</tr>
<tr>
<td>Accuracy(_{variance})</td>
<td>4.446 x 10(^{-4})</td>
<td>6.568 x 10(^{-4})</td>
<td>1.994 x 10(^{-4})</td>
<td>9.539 x 10(^{-4})</td>
</tr>
<tr>
<td>F(_{\text{mean}})</td>
<td>0.7983</td>
<td>0.8059</td>
<td>0.6903</td>
<td>0.6820</td>
</tr>
<tr>
<td>F(_{\text{variance}})</td>
<td>7.289 x 10(^{-4})</td>
<td>3.978 x 10(^{-4})</td>
<td>2.893 x 10(^{-4})</td>
<td>9.892 x 10(^{-5})</td>
</tr>
<tr>
<td>Precision(_{mean})</td>
<td>0.7432</td>
<td>0.7153</td>
<td>0.8618</td>
<td>0.8983</td>
</tr>
<tr>
<td>Precision(_{variance})</td>
<td>8.945 x 10(^{-4})</td>
<td>1.583 x 10(^{-3})</td>
<td>6.594 x 10(^{-4})</td>
<td>9.507 x 10(^{-4})</td>
</tr>
<tr>
<td>Recall(_{mean})</td>
<td>0.9594</td>
<td>0.9738</td>
<td>0.9462</td>
<td>0.7695</td>
</tr>
<tr>
<td>Recall(_{variance})</td>
<td>7.452 x 10(^{-4})</td>
<td>6.347 x 10(^{-3})</td>
<td>9.743 x 10(^{-3})</td>
<td>1.565 x 10(^{-3})</td>
</tr>
</tbody>
</table>

Confusion matrix of prediction results at Q + A, Q + Ori (False: Fake data, True : Fact data)

- To train using AnswerOriginal (Ori) improved the models’ performance
- Precision of fake data is tend to be lower than one of fact data

---

\(^1\) Yuuki Tachioka and Atsushi Koyaki. 2022. ditlab at the NTCIR-16 QA Lab-Poliinfo3\(^1\), proceeding of The 16th NTCIR Conference

