Introduction

The purpose of QA Lab-PoliInfo-2 Summarization subtask is to give a summary for Tokyo Metropolitan Assembly Minutes in order to avoid reading long utterances. To create summaries, the organizers prepared a summary for each assembly speaker. Except the main text, task participants are provided with main topic of a meeting, a subtopic of a meeting, and the speaker name.

Our Assumption

1. Sentence embedding model which is pre-trained on political domain could be useful for this task;
2. MMR score could be useful for extracting diverse sentences from documents

Main Contribution

1. Proposing Japanese Political Sentence-BERT;
2. Adapting an embedding-based unsupervised key-phrase extraction, EmbedRank++, to summarization
3. Adding two similarity functions to the MMR score which is used in EmbedRank++

Japanese Political Sentence-BERT (JPSB) Creation

JPSB Algorithm

1. Insert Japanese BERT into a triplet network
2. Input target and positive / negative sentence
3. Embed inputted sentences with the BERT modules, inside a triplet network, using MEAN pooling;
4. Compute cosine similarities
5. Training on dataset via making the distance d+ as 1 (positive), and the distance d- as 0 (negative)

Dataset for Sentence-BERT Training

We utilize utterances from Tokyo Metropolitan Assembly Minutes dataset, which are provided by the QA Lab-PoliInfo-2 organizers. 27,078 triplets of [target sentence, positive sentence, negative sentence] are created via the following steps:
1. We assumed that utterances that are adjacent to each other in terms of time series are semantically related, so that we treat an utterance which follows a target sentence as a positive sentence
2. Take a utterance as a negative sentence, as far as it is spoken on a different day or in a different meeting

These triplets are divided into 21,662 (80%) for training, 2,708 (10%) for development, and 2,708 (10%) for testing.

Model Evaluation

1. Take the difference \( \text{diff} \) between the cosine similarity of [positive sentence, target sentence] and [negative sentence, target sentence]. The larger \( \text{diff} \) indicates that the model can identify a positive sentence and a negative sentence better.
2. Accuracy

<table>
<thead>
<tr>
<th>Model</th>
<th>diff</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>USE</td>
<td>0.2441</td>
<td>0.8674</td>
</tr>
<tr>
<td>JPSB</td>
<td>0.3705</td>
<td>0.9849</td>
</tr>
</tbody>
</table>

EmbedRank++ based Method

To generate various sentences as summaries of utterances, we adopt EmbedRank++, which is based on MMR score. We newly add cosine similarities between \([D_i, MT]\) and \([D_i, ST]\), where \(D_i\) is a given document, \(MT\) is Main Topic for the summaries, and \(ST\) is SubTopic for the summaries.

\[
\text{MMR} = \max \{k, q\} + 0.5 \cdot \cos \text{Sim}(D_i, Q) - 0.5 \max_{D_j \in R} \cos \text{Sim}(D_j, D_i) + m \cdot \cos \text{Sim}(D_i, MT) + s \cdot \cos \text{Sim}(D_i, ST),
\]

\[
\text{key size (Answer)} \approx \frac{\text{Answer length}}{50},
\]

\[
\text{key size (Question)} \approx \frac{\text{Question length}}{50}.
\]

Extract the top-[key size] utterances as summaries

Results and Analysis

Regarding ROUGE-1 scores, Sentence-BERT-based approach does not outperform USE-based one.

Table: ROUGE-1 score of USE-based and JPSB-based methods tested on the test data

<table>
<thead>
<tr>
<th>Method</th>
<th>ROUGE-1 (Recall)</th>
</tr>
</thead>
<tbody>
<tr>
<td>USE-based EmbedRank++</td>
<td>0.0846</td>
</tr>
<tr>
<td>JPSB-based EmbedRank++</td>
<td>0.0696</td>
</tr>
</tbody>
</table>

Considering the output, the generated summary #1 expresses the theme and contents similar to the reference summary. On the other hand, the summary #2 refers to a totally different theme from the reference summary.

Table: Example of the outputted summaries

<table>
<thead>
<tr>
<th>Reference Summary #1</th>
<th>Output #1</th>
<th>Reference Summary #2</th>
<th>Output #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>新たな整備目標は、区市町村支援、国への財源拡充、電線事業者等との連携を強化し積極的に推進する。</td>
<td>東京の産業政策やインフラ整備と一体的に取り組むことが求められます。</td>
<td>都道府県の産業政策やインフラ整備と一体的に取り組むことが求められます。</td>
<td>一層の改善に努めてまいります。</td>
</tr>
</tbody>
</table>

Conclusion

1. We created Sentence-BERT for Japanese political texts;
2. We adopt the Japanese Political Sentence-BERT to utterances included in the Japanese Minute data summarization task;
3. MMR did not work well for this subtask (both ROUGE and human evaluation)
4. It is necessary to tune the parameters of score function in our method