Summarizing Utterances from Japanese Assembly Minutes using Political Sentence-BERT-based Method for QA Lab-PoliInfo-2 Task of NTCIR-15

Daiki Shirafuji, Hiromichi Kameya, Rafal Rzepka, and Kenji Araki

Hokkaido University, Japan

Introduction

Concise Instruction for QA LabPoliInfo-2

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)
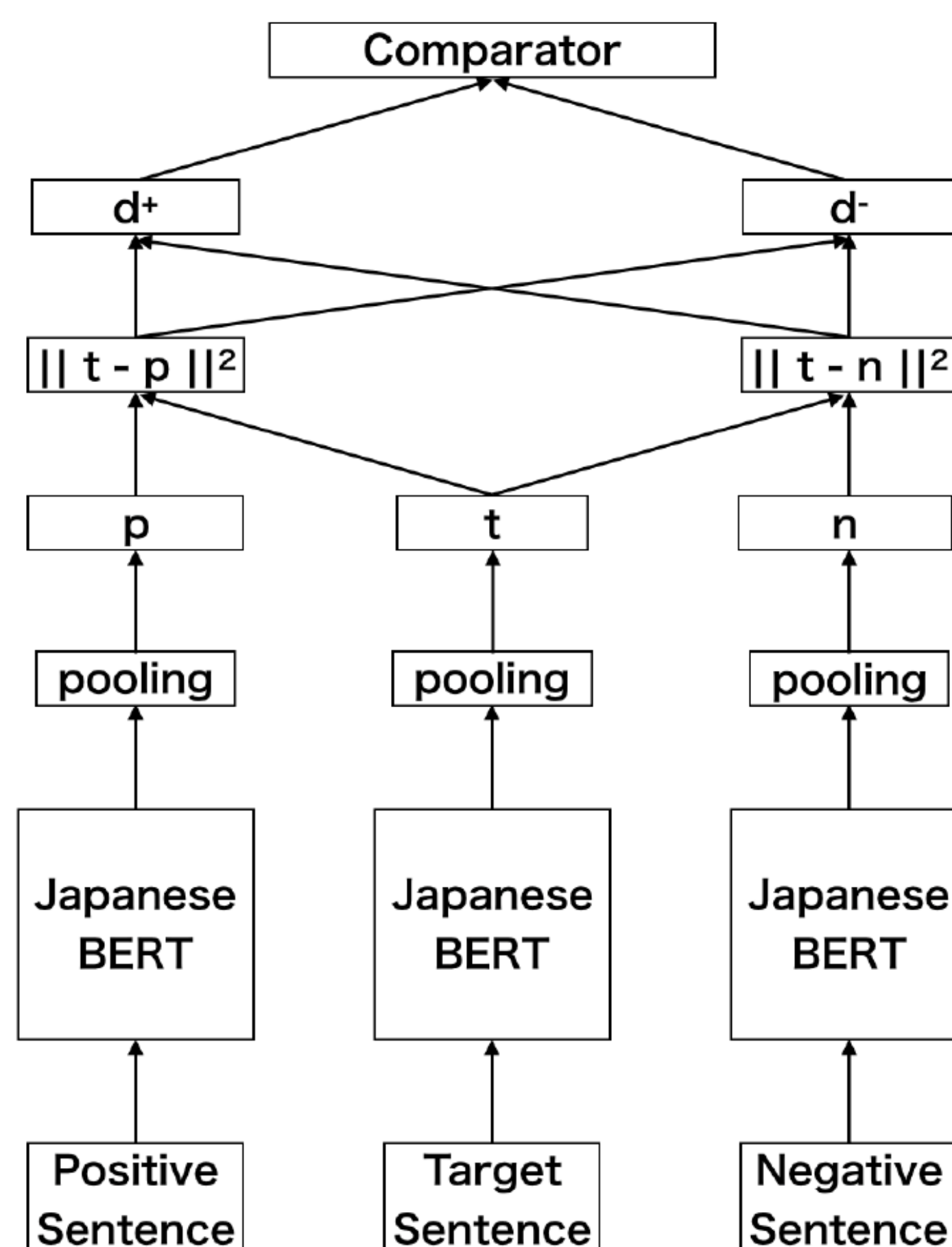
Dataset for Sentence-BERT Training

We utilize utterances from **Tokyo Metropolitan Assembly Minutes dataset**, which are provided by the QA LabPoliInfo-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 *diff* between the cosine similarity of [positive sentence, target sentence] and [negative sentence, target sentence]. The larger *diff* indicates that the model can identify a positive sentence and a negative sentence better.
2. Accuracy

Table: Results of USE and our proposed JPSB evaluated on the test data

	diff	Accuracy
USE	0.2441	0.8674
JPSB	**0.3705**	**0.9849**

Our Proposed Method

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 [*Di, MT*] and [*Di, ST*]**, where D_i is a given document, *MT* is Main Topic for the summaries, and *ST* is SubTopic for the summaries.

$$MMR = \arg\max_{D_i \in R \setminus S} \{ k * \{0.5 * CosSim(D_i, Q) $$
$$ - 0.5 * \max_{D_j \in S} CosSim(D_i, D_j)\} $$
$$ + m * CosSim(D_i, MT) $$
$$ + s * CosSim(D_i, ST)\}, $$

$$ key\ size(Answer) := \frac{Answer\ length}{50}, $$
$$ key\ size(Question) := \frac{Question\ length}{50}. $$

Extract the top-[*key size*] utterances as summaries

Parameters k, m and s are set to **0.2, 0.3, and 0.5**, respectively.
R: the ranked list of documents retrieved by an algorithm
S: the subset of documents in R,
D_i and D_j are retrieved documents,
Q: the averaged vector of inputted all documents
key_size(Answer) and key_size(Question) refer to the number of outputted sentences

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

	ROUGE-1 (Recall)
USE-based EmbedRank++	0.0846
JPSB-based EmbedRank++	0.0696

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

Reference Summary #1	都の産業政策やインフラ整備との一体的取組、川崎港・横浜港との連携に加え、都が責任を持って経営に関わる体制を確保。(We should ensure that the Tokyo Metropolitan Government will take responsibility for management of the Port of Kawasaki and the Port of Yokohama, in addition to coordinated efforts with Tokyo's industrial policies, infrastructure development and cooperation with the ports of Kawasaki and Yokohama.)
Output #1	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ております。/現場の実態を熟知した東京都が責任を持って港湾経営にかかわっていく体制を確保してまいります。(At the same time, we need to work on Tokyo's industrial policies and infrastructure development. We should ensure that the Tokyo Metropolitan Government, which knows well the actual situation in the field, will take responsibility for port management.)
Reference Summary #2	都道について新たな整備目標示す。区市町村支援、国への財源拡充要求、電線事業者等との連携を強化し積極的に推進する。(Let us explain the new purposes for Tokyo metropolitan roads. We will actively promote support for municipalities, also demand expansion of financial resources from the government, and strengthen cooperation with wire-line operators.)
Output #2	一層の支援に努めてまいります。/風格ある都市景観の形成と高度防災都市の実現を目指し世界に誇れる都市空間を創出してまいります。(We will make every effort to support the city. We will create a world-class urban space with the aim of creating a distinctive urban landscape and a highly disaster-resistant city.)

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