wfrnt Team at the NTCIR-15 QA Lab-PoliInfo-2 Task

Ken-ichi Yokote
velleykt@gmail.com

ABSTRACT

This paper investigates whether heuristics of conclusion extraction in Japanese is useful to develop a baseline system for summarization. We quantitatively verify the validity of examination of language use such as "English begins with conclusion, Japanese begins with background."

TEAM NAME

wfrnt

SUBTASKS

Dialog-summarization

1 INTRODUCTION

An existing research for summarization extracts the last sentence as a conclusion[3]. This approach is based on the observation that many Japanese texts first present backgrounds and then move on to conclusions. In this study, we apply several heuristics to Dialog-summarization Task[1], which aims at generating a summary of the transcript in the Tokyo Metropolitan Assembly, and conduct a comparison with other systems. We then examine the effectiveness of heuristics of conclusion extraction in Japanese as a baseline system of summarization.

2 SYSTEM OVERVIEW

Figure1 shows an overview of our proposed system. The first step is to assign Utterance Segment ID to a range of contiguous sentences. The second step is to identify the sentences corresponding to the input data using Utterance Segment ID; this step is called *Utterance-Level Segmentation*. The last step is to extract the conclusion statement from the sentences; this step is called *Extracting Conclusion Statement*.

3 UTTERANCE-LEVEL SEGMENTATION

In this study we apply the "Person-Role Relation Detection" and the "Utterance Segment Detection" introduced by Yokote[2] to the assignment of the Utterance Segment IDs and the identification of sentences from the input data.

4 EXTRACTING CONCLUSION STATEMENT

Figure2 shows the process of the extraction of conclusion statement. First, we pick up a sentence from the identified sentences backward. We then check whether it meets the ignore patterns; an example of ignore patterns is whether the sentence contains a particular word such as "Arigatou" (this pattern is motivated from the heuristics that "Arigatou" is likely to be part of greeting exchanged at the end of a conversation, being inappropriate as a conclusion). After finding a sentence that does not meet any ignore patterns, we use word tokenization to it and extract the last 50 words, which constitutes the conclusion statement. The length limitation of 50 is decided by the smallest number of AnswerLength column in the training dataset.

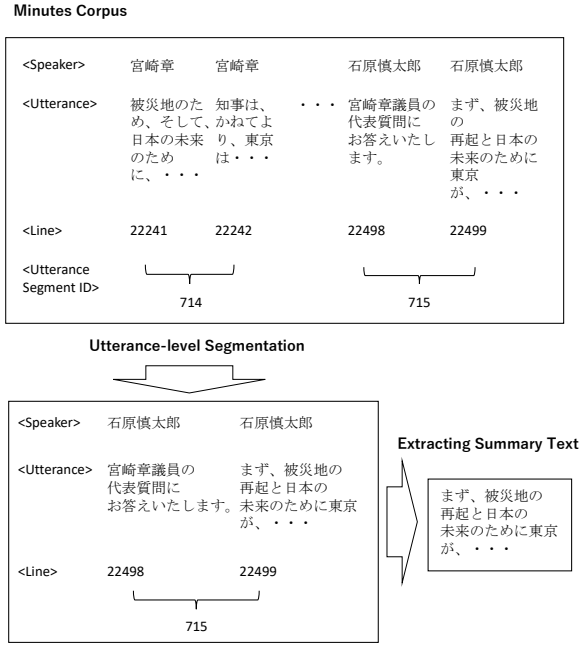

Figure 1: Overview of the proposed method

5 CONCLUSIONS

In the comparison of ROUGE-1-R scores, our system took 13th and 15th places among 19 systems submitted to Formal Run[1]. It would be possible to improve the performance of our system by varying the length limitation and reexamining the ignore patterns. For more detailed error analysis, it is important to consider how appropriate the extracted statement is as not only a conclusion but also a summary this task aims at.

6 ACKNOWLEDGMENTS

The author would like to thank the organizers of QA Lab-PoliInfo-2 for helpful discussions. The author would also like to thank Koji Yokote for proofreading this paper.

REFERENCES

[1] Yasutomo Kimura, Hideyuki Shibuki, Hokuto Ototake, Yuzu Uchida, Keiichi Takamaru, Madoka Ishioroshi, Teruko Mitamura, Masaharu Yoshioka, Tomoyoshi Akiba, Yasuhiro Ogawa, Minoru Sasaki, Kenichi Yokote, Tatsunori Mori, Kenji Araki, Satoshi Sekine, and Noriko Kando. 2020. Overview of the NTCIR-15 QA Lab-PoliInfo-2 Task. *Proceedings of The 15th NTCIR Conference* (12 2020).

Algorithm 1 Extracting Conclusion Statement

Require: Sentences $S = \{S_1, S_2, \cdots S_n\}$
Ensure: Conclusion Statement C
  for $i \leftarrow n, \cdots 3, 2, 1$ **do**
    if ' 登壇' is IN S_i **then**
      continue
    else if ' 拍手' is IN S_i **then**
      continue
    else if ' ありがとう' is IN S_i **then**
      continue
    else if ' 質問' is IN S_i **then**
      continue
    else if ' ——————'is IN S_i **then**
      continue
    end if
    $words \Leftarrow tokenize(S_i)$
    $C \Leftarrow words[-49:]$
    break
  end for

Figure 2: Process of Extracting Conclusion Statement

[2] Ken-ichi Yokote and Makoto Iwayama. [n.d.]. NAMI question answering system at QA Lab-PoliInfo.
[3] 谷川信弘 and 砂山渡. 2009. テキストの結論重視型要約の生成. In **人工知能学会全国大会論文集 第** *23* **回全国大会** *(2009)*. 一般社団法人 人工知能学会, 1B41–1B41.