YNU
YOKOHAMA National University

Forst: A Challenge to the NTCIR-15 QA Lab-PoliInfo-2 Task

Dialog Summarization system B

Hiromu Onogi[†1], Kiichi Kondo[†1],
Younghun Lim[†1], **Xinnan Shen**[†1],
Madoka Ishioroshi[†2], Hideyuki Shibuki[†2],
Tatsunori Mori[†1], Noriko Kando[†2†3]
[†1] Yokohama National University, [†2] Nationl Institute of Information, [†3] SOKENDAI

December 10, 2020

The goal of this task

Pref13_tokyo.json

{"ID":"130001_230617_10","Line":10,
"Prefecture":"東京都","Volume":"平成
23年_第2回
","Number":"1","Year":23,"Month":6,"
Day":17,"Title":"平成23年_第2回定例
会(第7号)","Speaker":"白石弥生子
","Utterance":"ご着席願います。"},
……

PoliInfo2-DialogSummarization-JA-Dry-Test.json

{
“AnswerEndingLine”: [
0
],
“AnswerLength”: [
50
],
“AnswerSpeaker”: [
“知事”
],
“AnswerStartingLine”: [
0
],
“AnswerSummary”: [
“”
],
“Date”: “2013-02-26”,
“ID”: “PoliInfo2-DialogSummarization-JA-Dry-Test-00005”,
“MainTopic”: “新知事の東京の将来像を示せ
エネルギー需要側の政
……

{
"AnswerEndingLine": [
"47934"
],
"AnswerLength": [
150
],
"AnswerSpeaker": [
"知事"
],
"AnswerStartingLine": [
"47794"
],
"AnswerSummary": [
"今後とも、現場を持つ強みを生かし、知恵を出し合い
、費用対効果の高い新たな施策を展開することで……
],
"Date": "2013-02-26",
"ID": "PoliInfo2-DialogSummarization-JA-Dry-Test-
00001",
"MainTopic": "新知事の東京の将来像を示せ
エネ
ルギー需要側の政策進化を",
"Meeting": "平成25年第1回定例会",
"Prefecture": "東京都",
"QuestionEndingLine": 47494,
"QuestionLength": 100,
"QuestionSpeaker": "酒井大史（民主党）",
"QuestionStartingLine": 47490,
"QuestionSummary": "まず初めに、都政運営について
伺います、そこで、私は、東京都の新しい知事として…...,
"SubTopic": "都政運営"
},

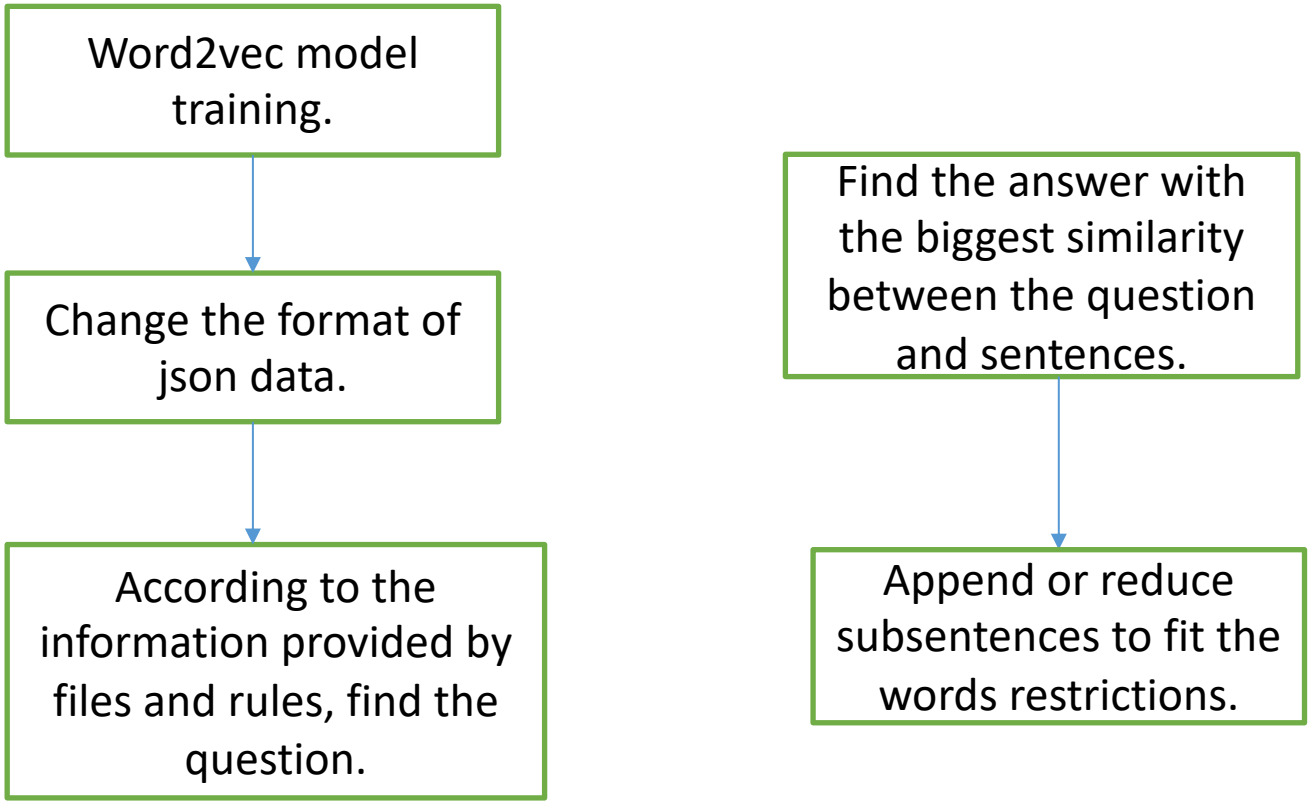
Related work

Kimura et al. (2019) described the tasks of PoliInfo2, and shows the data constructions of the minutes.

Kazuki Terazawa et al. (2019) proposed that we can use clue expressions help us find the questions and answers.For instance, "伺います" is the clue expression of a question, and "思っております" is the clue expression of an answer.

We applied an approach using rule-based method to solve this task.

The approach



Find the question

Feature
1.Limited number of speeches
2.Post all questions in first speaking opportunity
3. “Subtopic” shows the approximate location of question
4. “ ~ を伺います” is the main grammar of questions

Method
1.Look for the first statement that has same meeting time, title, etc, searching questions from here.
2.Using Mecab segment “Subtopic” into words, if all the words appeared, put this sentence into candidate list.
3. If " ～を伺います" appeared, put this sentence into candidate list.

Find the answer

Feature
1. Limited number of speeches
2.We know the speakers' job title instead of his names.
3. "Subtopic" shows the approximate location of answer.

Method
1. Extract information about people's name and job title to create a dictionary.
2. Look for the first statement that has same job and information in minutes file after question.
3. Use mecab to divide the utterance and the question statement and calculate the sentences vector.
4.Increase the subtopic words' weight.
5.Put the sentence of utterance with the biggest cosine similarity into candidate list.

vector = (∑Word vector * tfidf value) / number of words during this meeting

Append or reduce

Append	
Question	Do not process.
Answer	Calculate the similarity between the other sentences and the whole , put the sentences with the highest similarity in the original order until the character limit is exceeded.

Reduce	
Question	Segment the sentences with the signs ‘,’ . Calculate the similarity between the divided part and the whole , put the parts with the highest similarity in the original order until the character limit is exceeded.
Answer	

Discussion

Result		
ID(Forest)	**Modification**	ROUGE
247	Add a second sentence into the question candidate list	0.1471
235	Increase the importance of "Subtopic" words	0.1384
231	TF-idf	0.1219

- In this way, if there is only one answer, the effect of summarization is good. However, even if one person answers one question, maybe he answers from various directions. It's a bit lacking to summarize everything in this method. I will consider this direction in the future.