KSU Systems at the NTCIR-14 QA Lab-PoliInfo Task

Tasuku Kimura, Ryo Tagami, Hikaru Katsuyama,
○**Sho Sugimoto**, and Hisashi Miyamori

Kyoto Sangyo University

NTCIR-14 DAY 3: June 12 Wed.

09:30 - 09:50

Segmentation Task

Approach

Pre-processing

1. Single Speech Estimation
2. Speech Type Classification

   QUESTION, **ANSWER**, PROGRESS, GREETING, OPINION, REPORT, and REQUEST

Segmentation Process

1. Document Retrieval of Single Speech
2. Speech Segmentation

Document Retrieval of Single Speech

Approach: obtaining the speech section that contains the sentence most relevant to the query

1. A query is generated from the given inputs

Main Topic: 首都圏の中核なす多摩の実現を 私立高校生の就学支援補充せよ
Sub Topic: 私立高校生への緊急的支援

Question Speaker: 興津秀憲（民主党）
Question Summary: 保護者の失業等により就学が困難とならないよう速やかに助成金が渡る仕組みや授業料軽減等の実施を

Answer Speaker: 生活文化局長
Answer Summary: 家計状況急変には育英資金特別募集等施策を総合的活用し修学機会を確保

Document Retrieval of Single Speech

Approach: obtaining the speech section that contains the sentence most relevant to the query

1. A query is generated from the given inputs
2. Filter: by date, and by speaker(SF) （The effect by the 2nd pre-processing）

   QUESTION type: filtering by agreement of speaker's name

   ANSWER type: filtering by the agreement with the answer immediately after the obtained QUESTION

QUESTION1： Sub Topic❶ Sub Topic❷ Sub Topic❸

Answer1： Sub Topic❶ Sub Topic❷

Answer2： Sub Topic❸

Document Retrieval of Single Speech

Approach: obtaining the speech section that contains the sentence most relevant to the query

1. A query is generated from the given inputs
2. Filter: by date, and by speaker(SF)（The effect by the 2nd pre-processing）

   QUESTION type: filtering by agreement of speaker's name

   ANSWER type: filtering by the agreement with the answer immediately after the obtained QUESTION

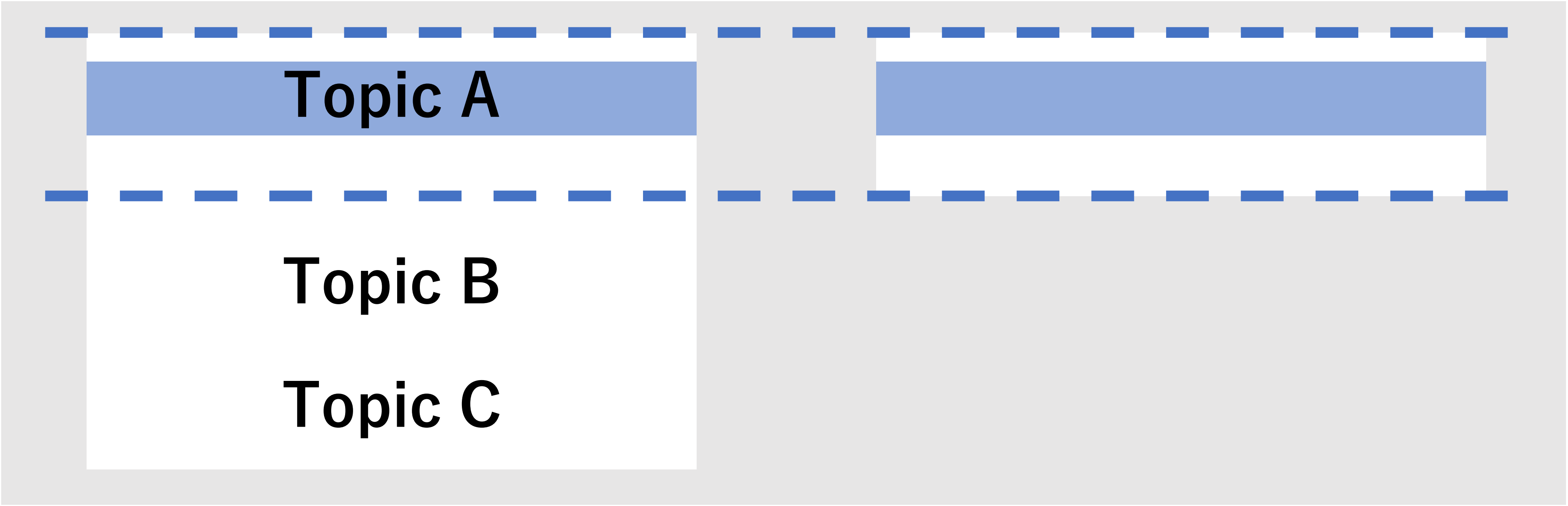
3. the top ranked result based on TF-IDF score is acquired
4. the single speech is acquired as the one to which the top-ranked sentence belongs（The effect by the 1st pre-processing）

Speech Segmentation

<u>It is highly probable that multiple topics exist in one speech</u>

Hypothesis: the similar vocabulary tends to appear frequently in the segment of the same topic.

Approach 1: segmentation based on the distribution of word frequency(WF)

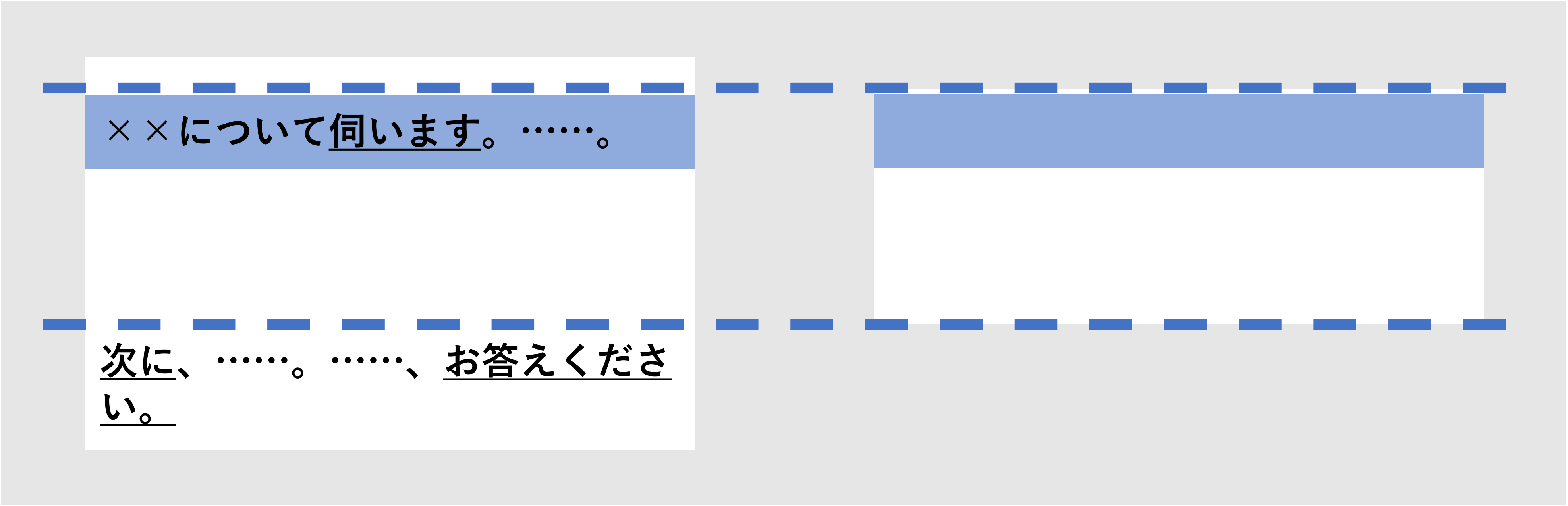


Speech Segmentation

It is highly probable that multiple topics exist in one speech

Hypothesis: the boilerplate language frequently appears at the topic break.

Approach 2: rule-based segmentation(RB)



Results

We tried **eight conditions** corresponding to the different combination of the presence or absence of SF, WF and RB.

#	Priority	SF	WF	RB
C1		—	—	—
C2			—	—
C3	KSU-01	—		—
C4		—	—	
C5	KSU-03			—
C6	KSU-02	—		
C7			—	
C8	KSU-04			

Discussion

The recall is expected to be as close to 1 as possible.
→ **SF improved the recall of "all"**

#	Priority	SF	WF	RB	P	R	F
C1		—	—	—	.087	.940	.160
C2		✔	—	—	.112	.991	.202
C3	KSU-01	—		—	.243	.779	.370
C5	KSU-03	✔		—	.660	.819	.731
C4		—	—		.294	.906	.444
C7		✔	—		.857	.952	.902
C6	KSU-02	—			.267	.759	.395
C8	KSU-04	✔			.922	.796	.854

The appropriate segmentation is expected to improve the precision and F-measure.
→ Both the two segmentation methods contribute to improve the accuracy.
→ **Especially, RB improves F-measure more greatly than WF.**

#	Priority	SF	WF	RB	P	R	F
C1		—	—	—	.087	.940	.160
C3	KSU-01	—		—	.243	.779	.370
C4		—	—		.294	.906	.444
C6	KSU-02	—			.267	.759	.395
C2			—	—	.112	.991	.202
C5	KSU-03			—	.660	.819	.731
C7			—		.857	.952	.902
C8	KSU-04				.922	.796	.854

Summarization Task

Problems of our training data set

1. Unknown words

It is difficult to deal with unknown words, since the data set is constructed from the minutes of the specific Assemblies.

2. The amount of the data set

It is difficult to say that the data set of 19,689 minutes were sufficient amount for deep learning.

Solution using Byte Pair Encoding (BPE)

1. Subword tokenizer treats high frequency words in the training data as one word and divides low frequency words into shorter units such as substrings and characters.

This process can reduce the unknown words.

2. **SentencePiece** which can provide unigram-based tokenizers can output multiple segmentation candidates with confidence degrees for the same input.

The training data can be sampled dynamically from the corpus, to augment data.

Proposed model（1/2）

Extended attention mechanism

- Generating the summary in accordance with the topic
- Global attention mechanism is extended so that the attention of document vector is generated based on the topic vector.

Sub Topic: 私立高校生への緊急的支援

Source: 現在の世界経済状況は日々進化し、人、物、金の交流は激烈な都市間競争時代になっていると考えます。社団法人日本経済調査協議会から二〇一一年三月に発表のあった、強靭な国際競争力をもった東京の実現という調査報告書では、バブル経済崩壊後の九〇年代後半には都心回帰の現象があらわれた。集積が富を生み、それがまた集積を生むということによって、国家の経営がスムーズに行われるという日本の特質が顕在化したのである。東京の国際競争力を高めなければならない理由は、歴史

Proposed model（1/2）

Extended attention mechanism

- Generating the summary in accordance with the topic
- Global attention mechanism is extended so that the attention of document vector is generated based on the topic vector.

Summary: 保護者の失業等により就学が困難とならないよう速やかに助成金が渡る仕組みや授業料軽減等の実施を

Proposed model（1/2）

Extended attention mechanism

- Generating the summary in accordance with the topic
- Global attention mechanism is extended so that the attention of document vector is generated based on the topic vector.

LenEmb mechanism

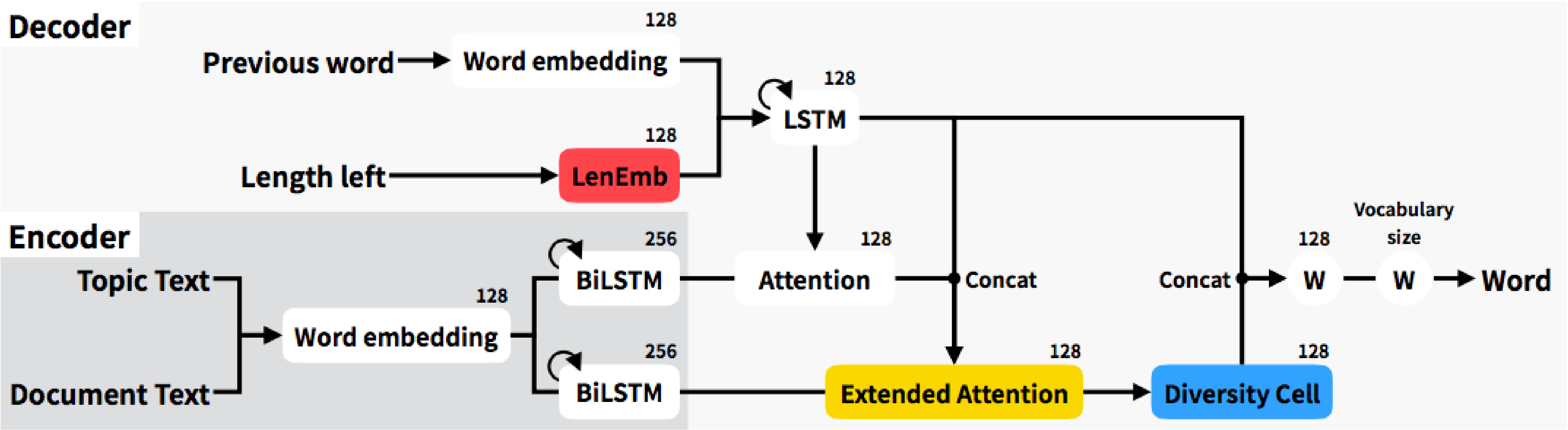
- Controlling the output length
- LenEmb is a method to introduce the length embedding vector to the input of LSTM in the decoder.

Proposed model（2/2）

Diversity cell mechanism

- Solving the problem of the Recurrent Neural Network (RNN)
- The mechanism transforms the input vector into vectors orthogonal to each other in each decoding step by extending the implementation of LSTM.

Model configuration of priority 5



Result

We constructed **six models** by combining the three mechanisms, namely, the tokenizer, the diversity cell, and LenEmb.

Priority	Tokenizer	Diversity cell	LenEmb	ROUGE-N		ROUGE-L	all-topic			
							content		formed	total
				N=1	N=2		X=0	X=2		
KSU-01	MeCab	✔	—	.158	.028	.009	.043	.043	1.955	.048
KSU-02	MeCab	—	—	.185	.043	.021	.076	.121	1.745	.071
KSU-03	BPE	✔	—	.172	.036	.008	.091	.157	1.715	.104
KSU-04	BPE	—	—	.171	.044	.013	.111	.167	1.419	.093
KSU-05	MeCab	✔	✔	.227	.029	.010	.048	.078	1.692	.048
KSU-06	BPE	✔	✔	.221	.038	.013	.078	.169	1.535	.091

Discussion

BPE

content↑: The model could deal with unknown words appropriately.

formed↓: The possibility of outputting a summary with grammatical errors increased.

Diversity cell

content↓: The predicted word vectors should not necessarily be orthogonal in each decoding step.

formed↑: The problem of repeated generation of the same words has been alleviated.

LenEmb

content↓／formed↓:
The content of the summary tends to change according to the remaining length.

Classification Task

Relevance(Re) and Fact-checkability(Fc)

Classification of Relevance(Re)

- Input: A text obtained by concatenating a topic and an utterance
- Output: A probability value
- Configuration: One-layered neural network

Classification of Fact-checkability(Fc)

- Input: An utterance
- Output: A probability value
- Configuration: Two-layered neural network composed of LSTM and fully connected layer

Stance(St)

Two-stage classifiers combining two binary classifiers

1. The classifier to identify “no opinion” or “having opinion”.
2. The classifier to identify “support” or “against”.

Selection of the features

- The occurrence frequency histogram of word N-grams (N=1,2,3) was made from the utterances in the development data per each label.
- The top-K word N-grams (K=200,400,600) having the largest difference in frequency were selected as a feature for each label.

Stance(St)

The combinations of features determined

Model	"no opinion" or "having opinion"		"support" or "against"	
	Features	**Dimension**	**Features**	**Dimension**
St1	1-gram	600	1-gram	600
St2	1-gram	600	1-gram	400
St3	1-gram, 2-gram, 3-gram	200+200+200	1-gram	600
St4	1-gram, 2-gram, 3-gram	200+200+200	1-gram	400

Results

We constructed six models by combining the three classification, namely, Re, Fc, and St.

Priority	RI	FC	St	Acc	P0	P1	P2	R0	R1	R2
1	Re1	Fc1	St1	.932	**.937**	.579	.056	**.995**	**.075**	.008
2	Re1	Fc1	St2	.932	**.937**	.689	.042	.995	.071	.008
3	Re1	Fc1	St3	**.934**	**.937**	**.738**	.083	**.998**	.071	.008
4	Re1	Fc1	St4	**.934**	**.937**	**.738**	.083	**.998**	.071	.008
5	Re2	Fc1	St1	.932	**.937**	.579	**.111**	.995	**.075**	**.019**
6	Re2	Fc1	St2	.932	**.937**	.689	.088	.995	.071	**.019**
7	Re2	Fc1	St3	**.934**	**.937**	**.738**	.100	.997	.071	.011
8	Re2	Fc1	St4	**.934**	**.937**	**.738**	.100	.997	.071	.011

Discussion

It was confirmed that each proposed model has **high ability to correctly estimate the final stance as Other**, whereas they **have low ability to accurately decide whether it is Fact-checkable Support or Fact-checkable Against**.

It is considered that both the recall of **Fact-checkable Support** and that of **Fact-checkable Against** in the final classification results were affected, because both the classification accuracy of “fact checkable” and that of “Support” and “Against” were low.

Conclusion

- In **Segmentation Task**, we proposed a method based on rules and vocabulary distributions. As a result, the team KSU achieved **third in five teams with the f–measure of 0.855**.

- In **Summarization Task**, we tried using a framework of the query–focused abstractive summarization.

- In **Classification Task**, we developed a method combining deep learning and two–stage classifiers. As a result, the team KSU achieved **second place in 11 teams with the accuracy 0.934**.