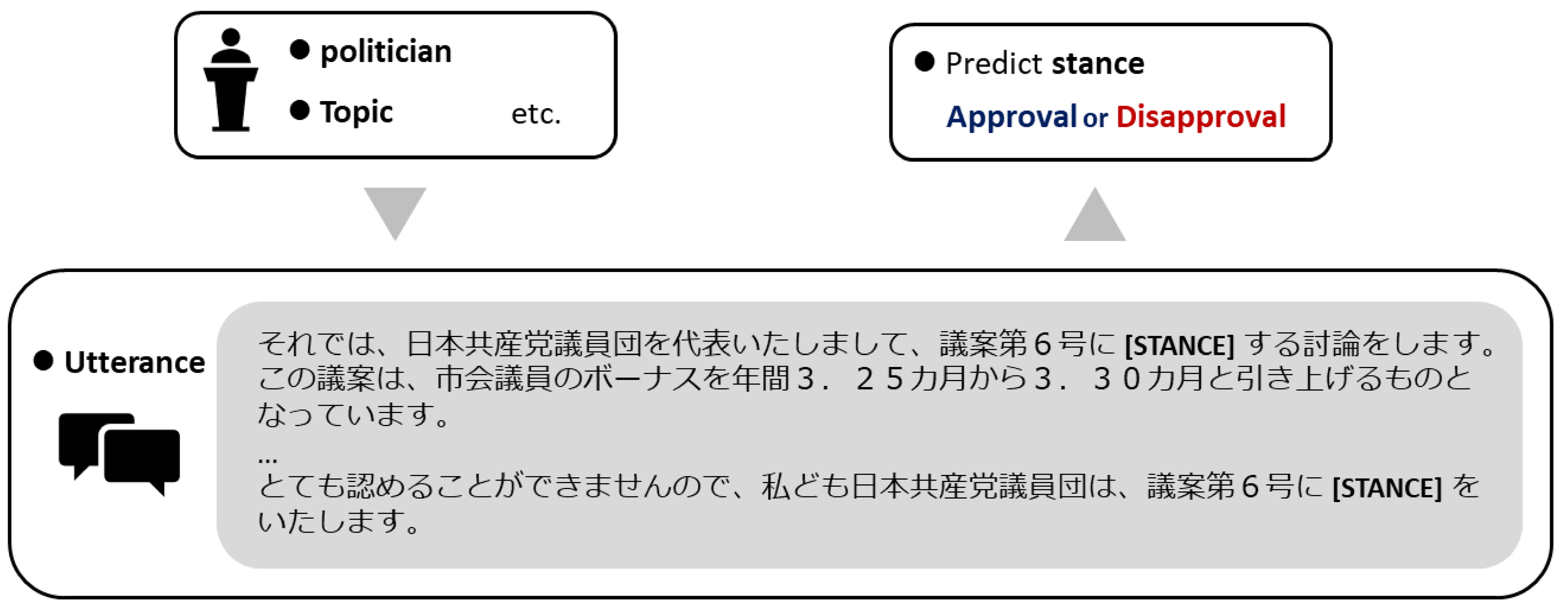
KIS's Stance Classification Model at the NTCIR-17 QA Lab-PoliInfo-4

Akira Nakada, Yoshinobu Kano（Shizuoka University, Japan）

Abstract

- We participated in the **Stance Classification 2 (SC2) subtask** of NTCIR-17 QA Lab-PoliInfo-4 as Team KIS.
- We **incrementally pretrained** the Japanese pretrained LUKE model with a Masked Language Model (MLM) on the **Diet minutes dataset.**
- We found that these methods were effective, achieved the highest score of **97.41% in accuracy** in the formal run of the subtask.

Purpose



Construction of Japanese politics-specific model

Estimate a politicians' stance from their utterances.

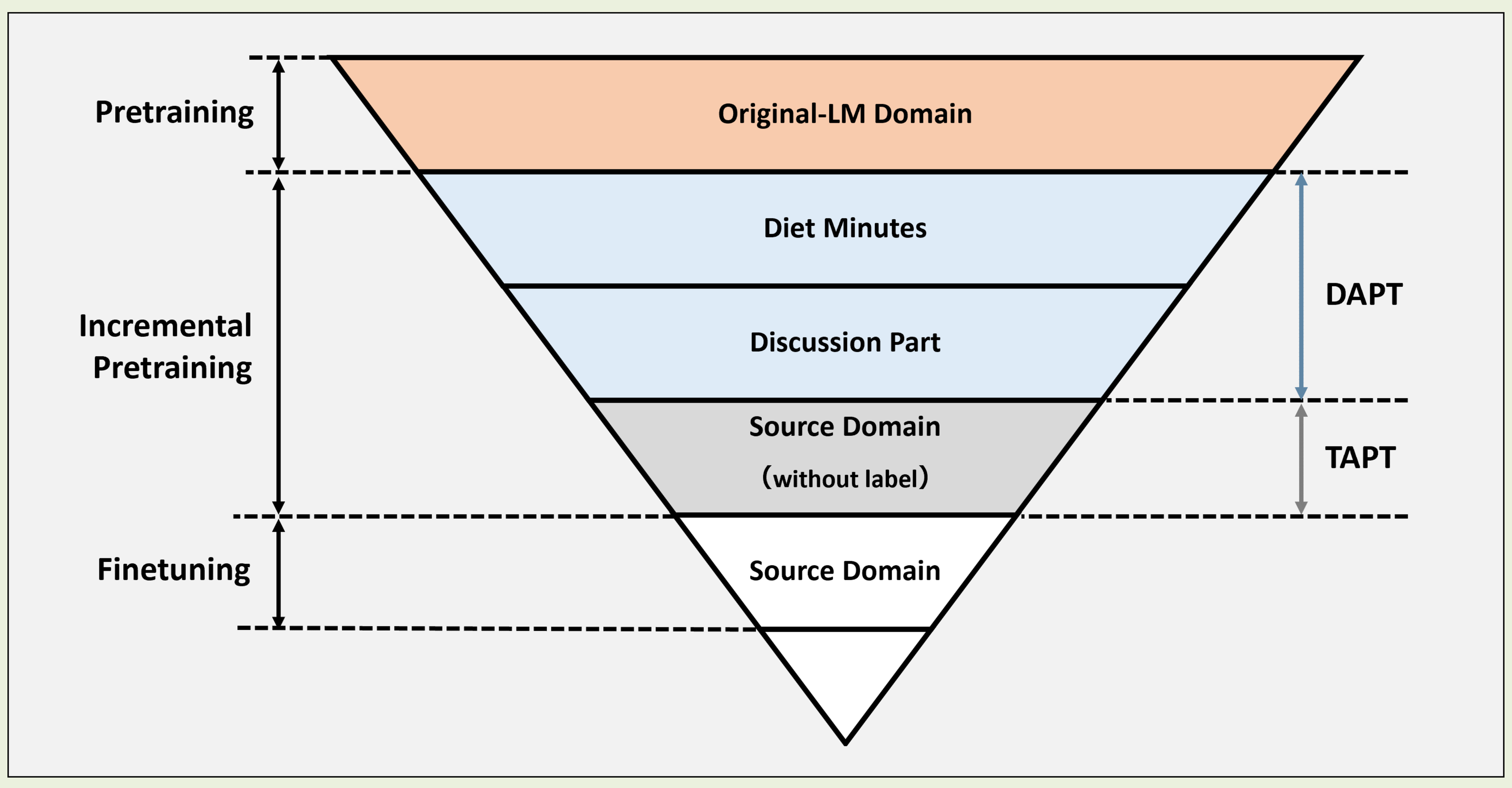
Approach

1. **Dealing with long utterances**
   - Using **head + tail method**
2. **Adaptation to the Japanese political domain**
   - Using **Incremental pretraining**

Incremental pretraining

- **Politics-specific model**
  - **Domain-adaptive pretraining（DAPT）**
    1. **DAPT1（Minutes-specific model）**
       MLM using the full text of the Diet minutes dataset
    2. **DAPT2（Discussion-specific model）**
       MLM using the discussion part of the Diet minutes dataset
- **Task-specific model**
  - **Task-adaptive pretraining（TAPT）**
    1. **TAPT**
       MLM using the utterance text portion of the training data

Conceptual figure of incremental pretraining



Dataset

- **Stance classification 2 (SC2) dataset**
  - **The training data**
    - **8,534** utterances
  - **The test data**
    - **2,240** utterances
- **Diet minutes dataset**
  - **Entire set**
    - **694,907** blocks
    - Each utterances is divided within 512 tokens considering the period ("。").
  - **Discussion part subset**
    - **13,204** blocks
    - Utterances containing the Japanese word "discussion (討論)".

Experiments

Dealing with long sentences

- Extracted head (start) and tail (end) of the sentences with different ratios.

Head + tail ratio	5-fold Cross Validation			Leader board (Acc)
	Acc (max)	Acc (min)	Acc (avg)	
512 / 0	0.9549	0.9349	0.9453	0.9563
384 / 128	**0.9654**	0.9443	**0.9568**	0.9621
256 / 256	0.9555	**0.9496**	0.9531	**0.9652**
128 / 384	0.9596	0.9420	0.9502	**0.9652**
0 / 512	0.9653	0.9436	0.9531	0.9612

The combination of **384 tokens** and **128 tokens** showed a stable performance.

Effectiveness of Domain-Adaptive Pretraining

Model	5-fold Cross Validation			Leader board (Acc)
	Acc (max)	Acc (min)	Acc (avg)	
without DAPT	0.9654	0.9443	0.9568	0.9621
DAPT1	0.9695	**0.9566**	**0.9630**	0.9705
DAPT2	0.9590	0.9490	0.9551	0.9610
DAPT1 + DAPT2	**0.9672**	0.9537	0.9625	**0.9728**

- **DAPT1** showed better performance in both cv and test.
- **DAPT1 + DAPT2** showed a best performance in test.

Effectiveness of Task-Adaptive Pretraining

Model	5-fold Cross Validation			Leader board (Acc)
	Acc (max)	Acc (min)	Acc (avg)	
without DAPT	0.9654	0.9443	0.9568	0.9621
TAPT	**0.9883**	**0.9830**	**0.9852**	0.9629
DAPT1 + TAPT	0.9672	0.9596	0.9633	0.9696
DAPT1 + DAPT2 + TAPT	0.9736	0.9602	0.9652	**0.9741**

- **TAPT** showed slightly better in test.
- **DAPT1 + DAPT2 + TAPT** showed a best performance in test.

Conclusion

- We verified **the effect of incremental pretraining** on a pretrained model with a dataset of the target political domain or task.
- In the future, we would like to apply our incremental pretrained model using the Diet minute dataset **to tasks in other political domains** to evaluate its generalized performance in the political domain.

KANOLAB