SRCB at the NTCIR-16 Real-MedNLP Task
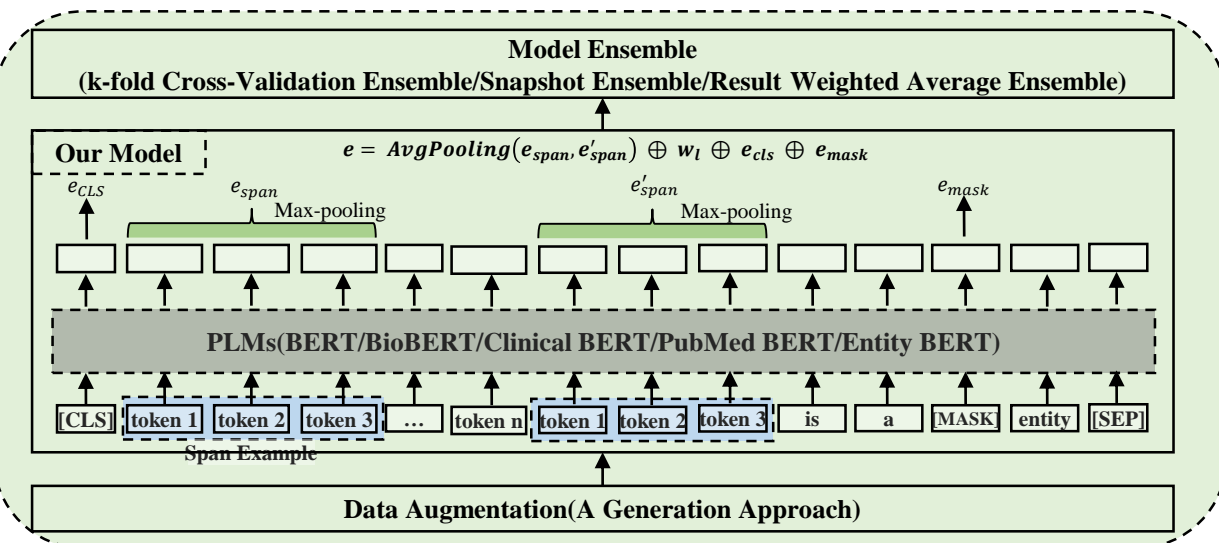
Yongwei Zhang, Rui Cheng, Lu Luo, Haifeng Gao, Shanshan Jiang, Bin Dong

Ricoh Software Research Center (Beijing) Co., Ltd.

RICOH
imagine. change.

Subtask1: Few-resource NER

Methodology

We consider NER as a span-based classification problem, and using the span-based model with prompt learning(PL) in our experiments. In addition, we try different pretrained language models(PLMs), data augmentation(DA) and model ensemble(ME) methods to improve our experiments further.
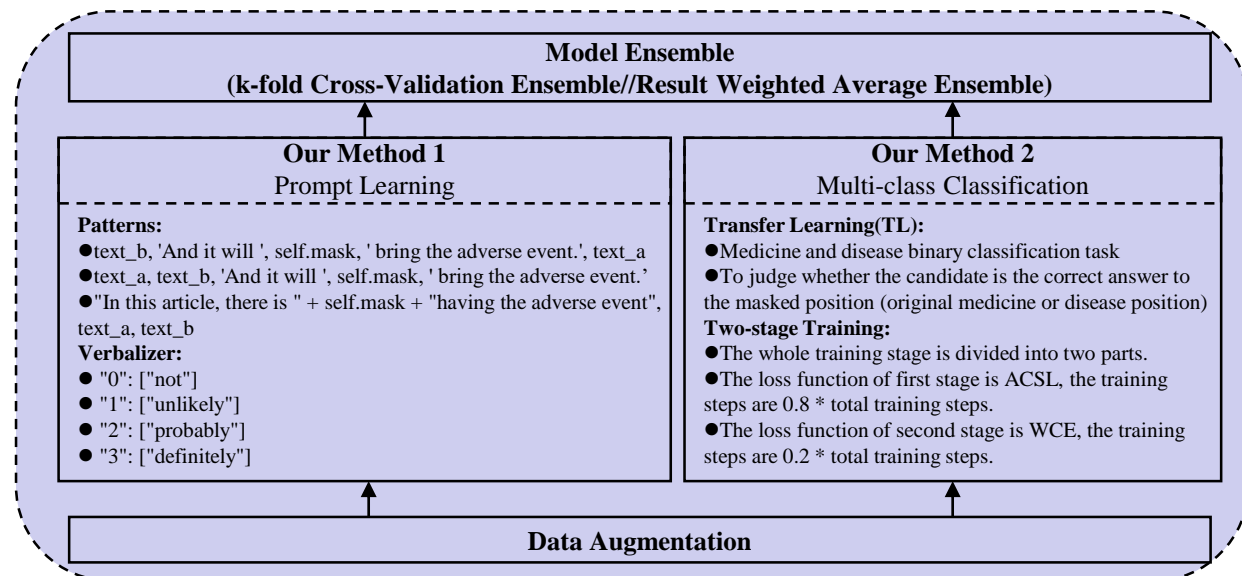

Model Ensemble
(k-fold Cross-Validation Ensemble/Snapshot Ensemble/Result Weighted Average Ensemble)

Our Model

$$e = AvgPooling(e_{span}, e'_{span}) \oplus w_l \oplus e_{cls} \oplus e_{mask}$$

e_{CLS} e_{span} Max-pooling e'_{span} Max-pooling e_{mask}

PLMs(BERT/BioBERT/Clinical BERT/PubMed BERT/Entity BERT)

[CLS] [token 1] [token 2] [token 3] ... [token n] [token 1] [token 2] [token 3] [is] [a] [MASK] [entity] [SEP]

Span Example

Data Augmentation(A Generation Approach)

Experiment

Model (CR)	Sequence tagging				Pointer network	Span+PL	+DA	+ME
	Clinical BERT	BioBERT	Entity BERT	PubMed BERT	PubMed BERT	PubMed BERT		
EntityF	59.232	59.664	60.296	62.284	63.176	64.698	66.760	**69.996**
JointF	55.128	55.392	55.902	57.464	58.946	59.712	62.222	**64.960**

Subtask3: ADE detection

Methodology

We mainly consider the methods fine-tuning on the pre-trained language model, include multi-class classification method and prompt learning(PL) based method. In addition, we also used model ensembles and data augmentation(DA).

Model Ensemble
(k-fold Cross-Validation Ensemble//Result Weighted Average Ensemble)

Our Method 1
Prompt Learning

Patterns:
- text_b, 'And it will ', self.mask, ' bring the adverse event.', text_a
- text_a, text_b, 'And it will ', self.mask, ' bring the adverse event.'
- "In this article, there is " + self.mask + "having the adverse event", text_a, text_b

Verbalizer:
- "0": ["not"]
- "1": ["unlikely"]
- "2": ["probably"]
- "3": ["definitely"]

Our Method 2
Multi-class Classification

Transfer Learning(TL):
- Medicine and disease binary classification task
- To judge whether the candidate is the correct answer to the masked position (original medicine or disease position)

Two-stage Training:
- The whole training stage is divided into two parts.
- The loss function of first stage is ACSL, the training steps are 0.8 * total training steps.
- The loss function of second stage is WCE, the training steps are 0.2 * total training steps.

Data Augmentation

Experiment

Methods	Prompt Learning				Multi-class Classification					
	our	w/o DA	& w/o position information	& w/o PL	our	w/o DA	& w/o cloze test task (TL)	& replace ACSL with CE in two-stage training	& w/o two-stage training	& w/o binary classification task (TL)
F1	**47.4**	43.0	41.8	34.2	**53.7**	52.5	51.7	50.6	47.1	43.6