SRCB at the NTCIR-16 Real-MedNLP Task

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

June 28, 2022

Real-MedNLP

RICOH
imagine. change.

Introduction

- Real-MedNLP is a shared task workshop for medical language processing using actual medical documents (case reports and radiology reports). The goal of this task is to promote the development of practical systems that support various medical services.
- The Real-MedNLP task has two corpus-based tracks (MedTxt-CR Track and MedTxt-RR Track).
- We mainly participated in the evaluation of the following tasks:
  - Subtask1-CR-EN, Subtask1-RR-EN (Few-resource NER)
  - Subtask3-CR-EN (ADE)

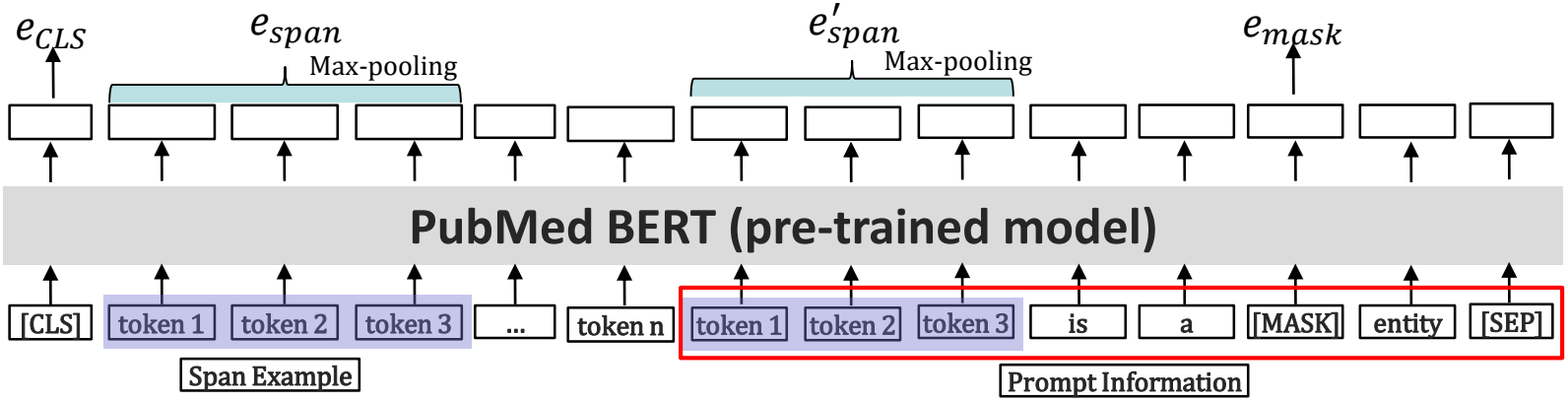
Subtask1:Methods

RICOH
imagine. change.

Methods

- We first tried several popular NER methods, such as sequence tagging, pointer network and span-based method.
- Then we used prompt learning (PL) to improve the span based method.
  - Define a prompt template, like “[SPAN] is a [MASK] entity.”
  - Construct input based on sentences, spans, and templates, “[CLS] [Sentence] [SPAN] is a [MASK] entity. [SEP]”.
  - Obtain the embeddings of the newly input based on pre-trained model.
  - Combine the embeddings of span, span width, [CLS] and [MASK] to get final representation of span.
  - Input final representation to a softmax layer, which yields a posterior for each span.

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



Subtask1:PLMs

RICOH
imagine. change.

Pre-trained language models (PLMs)

- We tried four pre-trained models in the medical field to improve performance, including BioBERT, Clinical BERT, Entity BERT and PubMed BERT.
- In different model structures, PubMed Bert has the best effect, and our model (Span + PL) is better than other models

PMLs	EntityF(CR)	JointF(CR)
Clinical BERT	59.232	55.128
BioBERT	59.664	55.392
Entity BERT	60.296	55.902
Pubmed BERT	62.284	57.464

Subtask1-CR-EN training set 5-fold cross-validation average results of sequence tagging with different pre-training models

Model	EntityF(CR)	JointF(CR)
Sequence tagging	62.284	57.464
Pointer network	63.176	58.946
Span + PL	64.698	59.712

Subtask1-CR-EN training set 5-fold cross-validation average results of different models on PubMed BERT

Subtask1:DA and Model Ensemble

RICOH
imagine. change.

Data Augmentation (DA)

- We used a generation approach to train a language model to learn the distribution of words and tags from the data for generating synthetic training data.

Model Ensemble (ME)

- Use k-fold cross-validation to make full use of training data and select some good models as candidate ensemble models.
- Because we used several different structural models, the simplest result weighted average method was used for final result integration.

Model	EntityF(CR)	JointF(CR)
Span + PL(Base)	64.698	59.712
+DA	66.760	62.222
+ME	69.996	64.960

Subtask1-CR-EN training set 5-fold cross-validation average results on different technical points

Subtask1:Results

RICOH
imagine. change.

Subtask1-CR-EN results for our submitted runs

Run	EntityAcc	EntityP	EntityR	EntityF	JointAcc	JointP	JointR	JointF
Subtask1-CR-EN-1	85.69	65.09	55.38	59.84	84.14	59.69	50.79	54.88
Subtask1-CR-EN-2	88.27	67.65	59.71	63.43	86.58	62.50	55.16	58.60
Subtask1-CR-EN-3	87.77	68.06	57.59	62.39	86.13	62.80	53.14	57.57
Subtask1-CR-EN-4	88.38	58.17	60.71	59.41	86.47	53.46	55.80	54.60
Subtask1-CR-EN-5	88.64	60.90	59.78	60.33	86.66	55.92	54.89	55.40

Subtask1-RR-EN results for our submitted runs

Run	EntityAcc	EntityP	EntityR	EntityF	JointAcc	JointP	JointR	JointF
Subtask1-RR-EN-1	92.23	83.14	82.06	82.60	89.96	79.71	78.67	79.19
Subtask1-RR-EN-2	92.28	83.26	82.06	82.66	89.61	79.31	78.17	78.74
Subtask1-RR-EN-3	92.66	79.90	81.34	80.61	90.37	76.50	77.88	77.19

Subtask3:Methods

RICOH
imagine. change.

- We mainly consider the methods fine-tuning on the pre-trained language model, include prompt learning(PL) based method and multi-class classification method.

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"]

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.

Subtask3:PLMs, DA and Ensemble

RICOH
imagine. change.

- **Pre-trained language models**
  - We tried three pre-trained models in the medical field to improve performance, including PubMed BERT, Clinical BERT and BioBERT.
  - In different methods, PubMed Bert has the best effect.
- **Data Augmentation**
  - **BT**: Back translation (Data augmentation)
  - **KLD**: KL-Divergence (Balanced data proportion)
- **Ensemble**
  - Use k-fold cross-validation to make full use of training data and select some good models as candidate ensemble models.
  - Two-stage ensemble: construct different 5-fold data to obtain the model ensemble results, and another weighted ensemble is used to obtain the finally result .

Subtask3:Experiments

RICOH
imagine. change.

- **Subtask3-CR-EN training set 5-fold cross-validation average results of prompt learning**

Prompt Learning	47.4
w/o data augmentation	43.0
& w/o position information	41.8
& w/o prompt learning	34.2

- **Subtask3-CR-EN training set 5-fold cross-validation average results of prompt learning**

Multi-class Classificatiion	53.7
w/o data augmentation	52.5
& w/o cloze test task (Transfer Learning)	51.7
& replace ACSL with CE in two-stage training	50.6
& w/o two-stage training	47.1
& w/o binary classification task (Transfer Learning)	43.6

Subtask3:Results

RICOH
imagine. change.

- **Subtask3-CR-EN (ADE) results for our submitted runs**

Method	Entity Level (ADEval)				Report Level
	0	1	2	3	
Subtask3-CR-EN-1 (ADE)	97.25	7.69	-	61.54	57.14
Subtask3-CR-EN-2 (ADE)	97.25	0.00	-	63.41	57.14
Subtask3-CR-EN-3 (ADE)	97.18	0.00	-	61.54	52.63
Subtask3-CR-EN-4 (ADE)	97.02	0.00	-	60.00	57.14
Subtask3-CR-EN-5 (ADE)	97.60	0.00	-	66.67	42.86
Subtask3-CR-EN-6 (ADE)	97.11	9.09	-	54.05	47.06

Thank you
Q&A