Does BERT Know Which Answer Beyond the Question ?

CCKS 2021：面向中文医疗科普知识的内容理解（二）医疗科普知识答非所问识别 冠军方案

1 National Laboratory of Pattern Recognition Institute of Automation, Chinese Academy Sciences,Beijing,100190,China，
xiafei2020@ia.ac.cn,wengsyx@gamil.com,{shizhu.he,kliu,jzhao}@nlpr.ia.ac.cn
2 School of Artificial Intelligence,University of Chinese Academy of Sciences,Beijing,100190,China
3 Beijing Unisound Information Technology,Beijing100028，
Chinaliushengping@unisound.com
4 Qingbo Intelligence Technology Co.LTD,Hefei,China，{xiamaojin,yuqiang}@gsdata.cn

YixuanWeng [1†], Fei Xia [1,2†], Maojin Xia [4], Qiang Yu [4], Jinfeng Huang , ShizhuHe [1,2], KangLiu [1,2], ShengpingLiu [3], and JunZhao [1,2]

† These authors contributed equally to the work.

Motivation

Medical question and answser matching validation (QAMV) aims to construct an intelligent medical question answering classification system. However, the most challenging is that there are few non-ideal answers, and many of them are difficult to understand for the model.

The non-deal answer refers to answer A cannot satisfy the query Q.After the baseline experiment, we analyzed the bad cases in-depth and found hard instances and insufficient model generalization capabilities.

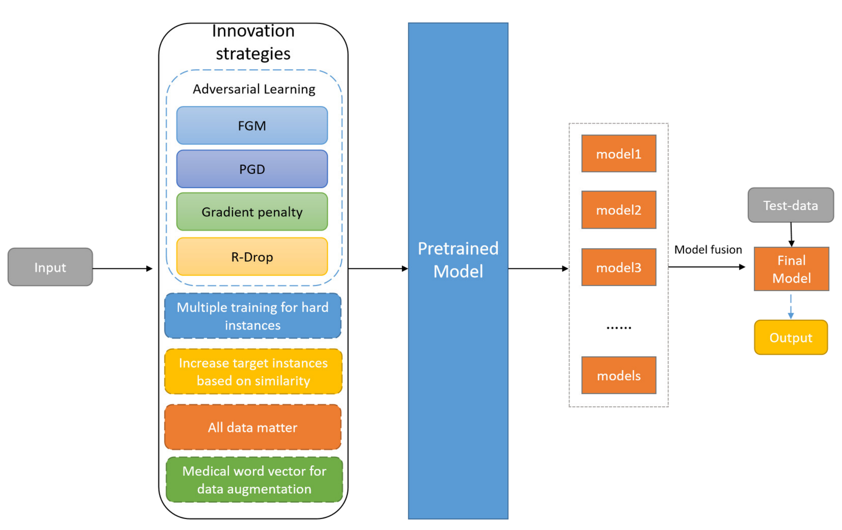

Fig. 1. innovation strategies

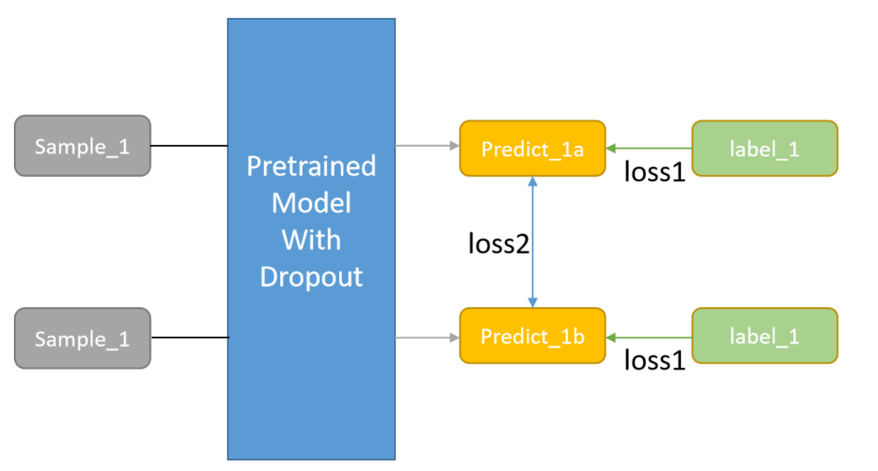

Fig. 2. Schematic diagram of method RDrop.

Method

- Four different adversarial training methods. We researched and compared FGM , PGD , gradient loss , and R-Drop . We also experimented with the combination of FGM and R-Drop and delayed adversarial training. This method improves by 1.64% relative to the baseline. （**Fig.2**）
- Hard instances identification and multi-round training methods. We use external data combined with the pseudo-label method to train a selection model for selecting error-prone and low-confidence instances. After that, we carry out multiple rounds of retraining on selected hard instances according to the ideas of curriculum learning . This method improves by 1.01% relative to the baseline. （**Fig.3**）

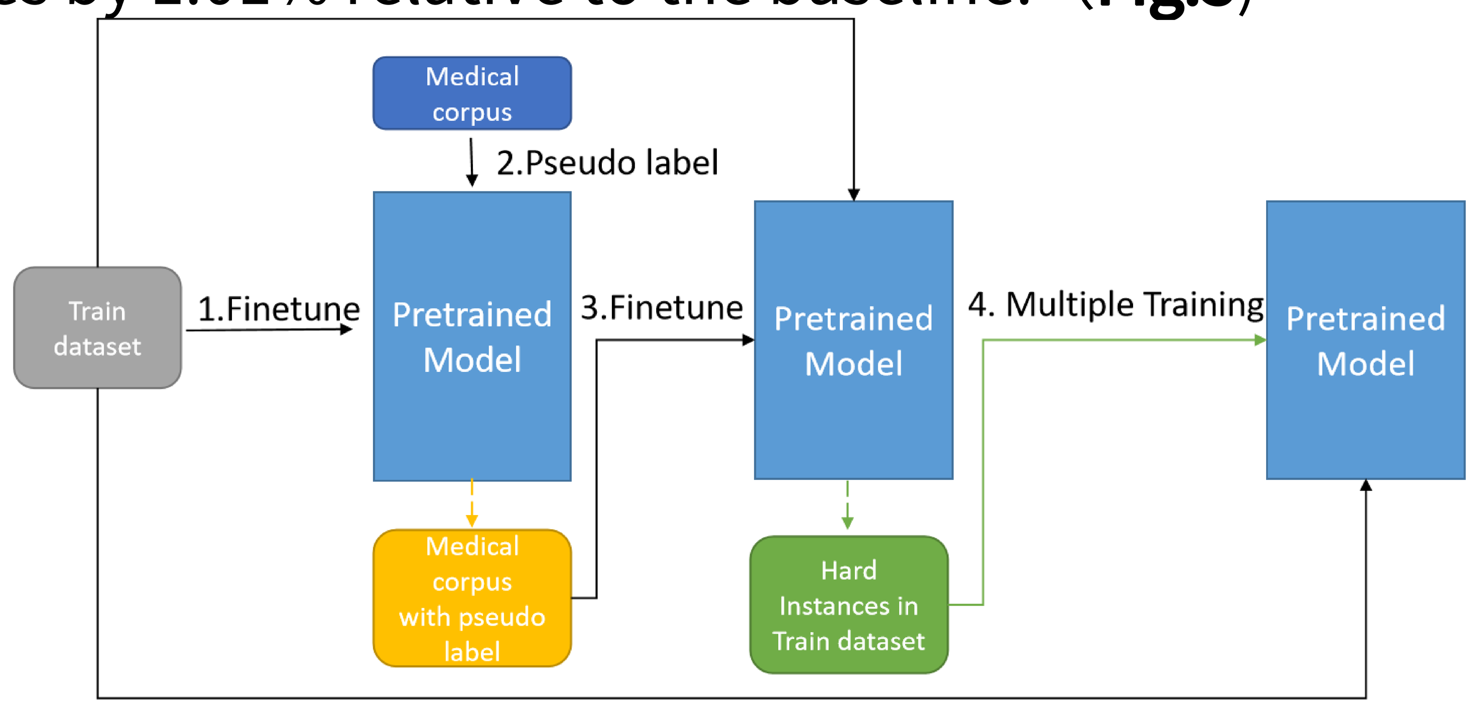

Fig. 3. Multiple Training for Error-prone and Difficult Instances.

Method

- Target instances constructed by similarity. We use the similarity algorithm edit distance to augmented the target instances. This method improves by 0.99% relative to the baseline.(**Fig.4**)
- Validation set retraining with a small learning rate. To make full use of the data, we conducted many experiments on the validation set. Experiments show that retraining the validation set with a small learning rate(2.00E-06) can improve the baseline (0.99%). (**Fig.5**)
- Medical word vector combined with Easy Data Augmentation(EDA) method for text data augmentation. We use medical word vectors combined with EDA for text data augmentation, using RS (replace synonymous medical words), RR (replace random words), SR (swap sentences randomly), ID (insertand delete words randomly). This method improves by 1.19% relative to the baseline.

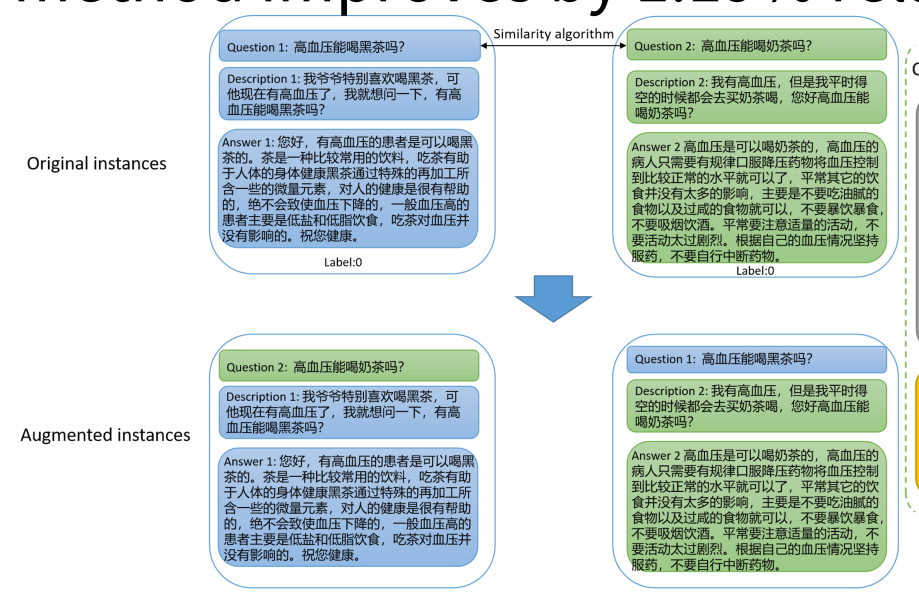

Fig. 4. Minimum Edit Distance Search

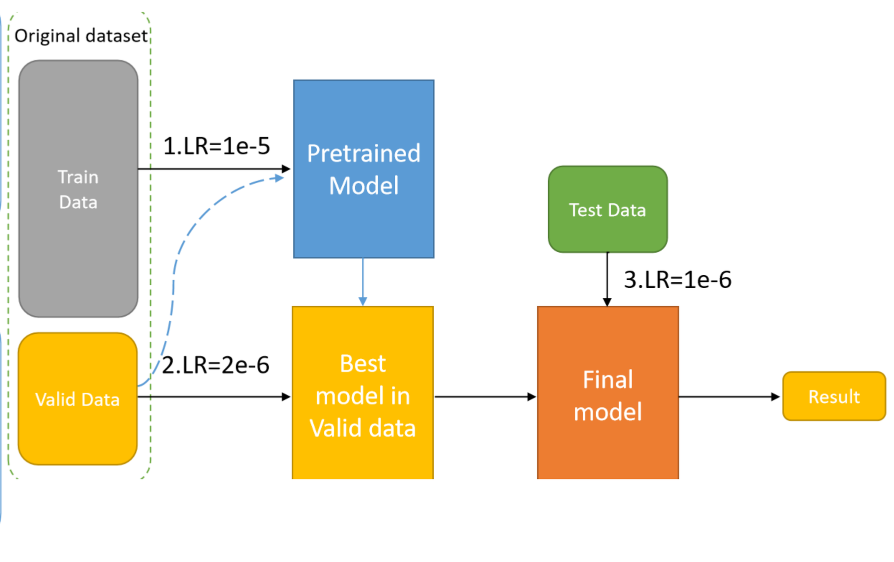

Fig. 5. Small learning rate fine-tuning

Experiments

Baseline	FGM	gradient penalty	R-Drop	FGM&R-Drop	delay FGM&R-Drop
89.61	90.59	90.44	89.86	90.56	91.25

Table 1. Results of adversarial training.

	1/2 predictied wrong and 1/2 low confidence(4000)			all predicted wrong(4000)	all loss confidence(3000)	Edit distance
Baseline	add (1/3) 1333	add(2/3) 2666	add(3/3) 4000			
89.61	90.17	90.52	90.62	90.36	90.15	90.6

Table 2. Multiple Training for hard Instances & Increase target instances

Best Model	2.00E-04	2.00E-05	5.00E-06	2.00E-06	8.00E-07
90.21	67.67	90.32	90.58	90.60	90.32

Table 3. Retraining results with different small learning rates .

Baseline	Adversarial	Hard instances	Edit distance	Retraining on validation	MEDA
89.61	91.25(+1.64)	90.62(+1.01)	90.6(+0.99)	90.60(+0.99)	90.8(+1.19)

Table 4. Results of different innovation methods.

TestA			TestB		
Rank	Team Name	Score	Team Name	Score	Rank
1	DeepBlueAI	84.963	*Our Team*	70.698	1
2	*Our Team*	84.498	DeepBlueAI	69.972	2
3	FREE	84.436	Space Oddity	69.778	3
4	Space Oddity	84.379	united	68.96	4

Contribution

We propose an innovative strategy to solve these problems, which consists of five parts:

1) Four different adversarial training methods 2) Complex instances recognition model and multi-round training methods 3) Target examples constructed by similarity 4) The validation set small learning rate retraining 5) Medical word vector combined with EDA method for text data amplification.

We achieved the best results in the CCKS competition, which proved the effectiveness and practicality of our proposed method. In the future, we will consider using medical word vectors for data amplification and try different fusion meth- ods further to improve the accuracy and quality of the generated responses.