MRC-based Medical NER with Multi-task Learning and Multi-strategies

Xiaojing Du, Yuxiang Jia[*] and Hongying Zan
School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou, China
zzu_dxj@163.com,{ieyxjia,iehyzan}@zzu.edu.cn

Abstract

Medical named entity recognition (NER), a fundamental task of medical information extraction, is crucial for medical knowledge graph construction, medical question answering, and automatic medical record analysis, etc. Compared with named entities (NEs) in general domain, medical named entities are usually more complex and prone to be nested. To cope with both flat NEs and nested NEs, we propose a MRC-based approach with multi-task learning and multi-strategies. NER can be treated as a sequence labeling (SL) task or a span boundary detection (SBD) task. We integrate MRC-CRF model for SL and MRC-Biaffine model for SBD into the multi-task learning architecture, and select the more efficient MRC-CRF as the final decoder. To further improve the model, we employ multi-strategies, including adaptive pre-training, adversarial training, and model stacking with cross validation. Experiments on both nested NER corpus CMeEE and flat NER corpus CCKS2019 show the effectiveness of the MRC-based model with multi-task learning and multi-strategies.

1 Introduction

With the fast development of medical digitalization, more and more medical documents are generated, including electronic medical records, medical reports, etc. Medical information extraction, including medical named entity recognition (NER), becomes increasingly important to applications like knowledge graph construction, question answering system, and automatic electronic medical record analysis. Medical NER is a task to automatically recognize medical named entities, like body (bod), disease, clinical symptom (sym), medical procedure, medical equipment, drug, medical examination item, etc., from medical texts. Medical named entities are usually long, nested and polysemous, which pose great challenges to medical NER. For example, in Fig 1, the two "bod" entities "延髓"(medulla oblongata) and "脊髓"(spinal cord) are nested in the "sym" entity "延髓和脊髓受损"(damage to the medulla oblongata and spinal cord).

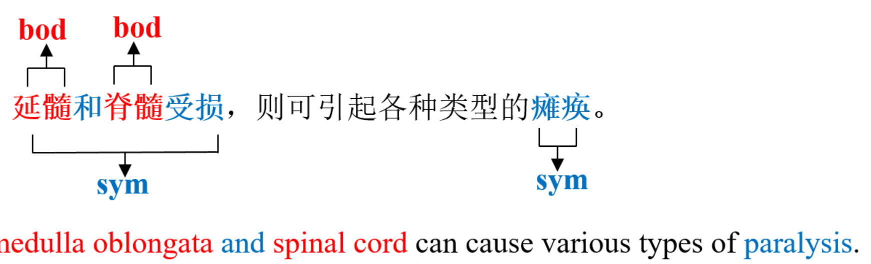

Figure 1: An example with nested entity

To tackle both flat and nested NER, like (Li et al., 2020b), we take NER as a machine reading comprehension (MRC) problem. In addition, from different views, NER can be treated as a

[*]Corresponding author

Proceedings of the 21st China National Conference on Computational Linguistics, pages 836-847, Nanchang, China, October 14 - 16, 2022.
(c) Technical Committee on Computational Linguistics, Chinese Information Processing Society of China

sequence labeling (SL) task or a span boundary detection (SBD) task. We integrate MRC-CRF model for SL and MRC-Biaffine model for SBD into the multi-task learning (MTL) architecture. There is no nested NEs composed of entities of the same type in the datasets, so we select the more efficient MRC-CRF as the final decoder. To further improve the model, we employ multi-strategies (MS), including adaptive pre-training, adversarial training, and model stacking with cross validation. The main contributions of this paper are as follows:

- We improve MRC-CRF for medical NER with Biaffine through a multi-task learning architecture, which is a lighter way than traditional ensemble learning.

- We propose multi-strategies to improve the NER model, including adaptive pre-training, adversarial training, and model stacking with cross validation.

- Experimental results on both the nested NER corpus CMeEE (Zhang et al., 2022) and the flat NER corpus CCKS2019 (Han et al., 2020) show the effectiveness of the proposed model.

2 Related Work

Just like NER in other application domains, medical NER borrows methods from NER in general domain. The methods evolve from rule-based methods, traditional machine learning-based methods, deep learning-based methods, to the present mainstream pre-training-based methods.

Pre-trained models like BERT (Dai et al., 2019; Li et al., 2020a; Qin et al., 2021), ELMo (Li et al., 2019; Li et al., 2020c; Wan et al., 2020), etc., have become a standard module to encode the input texts. To better represent a text, RNN (Chowdhury et al., 2018), LSTM (Dai et al., 2019), GRU (Qin et al., 2021), CNN (Kong et al., 2021) and other neural networks are usually employed after the pre-trained model. Taking the NER as a sequence labeling problem, CRF (Lafferty et al., 2001) is finally used to generate the sequence labels.

For Chinese, characters (Li et al., 2020c), radicals, strokes (Li et al., 2019; Luo et al., 2020) and glyphs (Zhong and Yu, 2021) can provide useful information besides words. Thus such linguistic units are encoded together with words using LatticeLSTM (Zhao et al., 2019), ELMo (Li et al., 2019; Li et al., 2020c; Wan et al., 2020) and other networks. Domain data can be used to improve medical NER. (Liu et al., 2021) pre-train a Med-BERT based on medical texts to boost the performance significantly. (Chen et al., 2020) integrate domain dictionary and rules with Bi-LSTM-CRF.

Multi-task learning is another way to improve the performance. NER model can be enhanced by parameter sharing with models of other tasks. (Chowdhury et al., 2018) take NER and POS tagging as two tasks. (Luo et al., 2020) take NER on two different datasets as two tasks. To tackle nested NER problem and encode knowledge from entity types, NER is formulated as a machine reading comprehension task (Li et al., 2020b), and two binary classifiers are used to detect the span of a named entity. To enhance the information interaction between the head and tail of an entity, (Cao et al., 2021) introduce biaffine to MRC. (Zhu et al., 2021) ensemble sequence labeling and span boundary detection by voting strategies while (Zheng et al., 2021) ensemble CRF and MRC.

3 The MRC-MTL-MS model

MRC model extracts answer fragments from paragraphs by a given question. Suppose X is the input text, for each entity type y, designing a query q_y, extracting a subsequence x of type y from X, and we can get the triple (q_y, x, X), which is exactly the $(question, answer, context)$ a MRC model needs. The model only calculates the loss of context during training, and masks the loss of query and padding.

Proceedings of the 21st China National Conference on Computational Linguistics, pages 836-847, Nanchang, China, October 14 - 16, 2022.

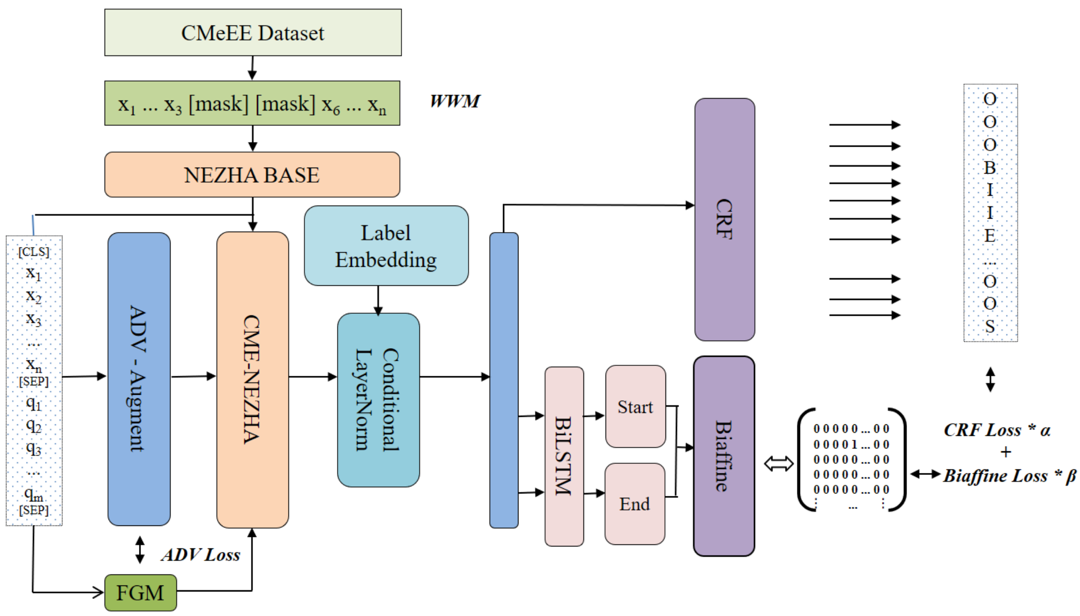

Figure 2: The architecture of the proposed NER model

3.1 Multi-task Learning (MTL)

The overall architecture of the model is shown in Fig 2. The multi-task learning architecture consists of the main task of sequence labeling by CRF and the auxiliary task of span boundary detection by Biaffine. For each entity type y, the input to the model is context X followed by query q_y, which is proved experimentally better than reversed concatenating way. The input is encoded by an adaptive pre-trained model CME-NEZHA, then goes through a Conditional LayerNorm guided by entity label embedding to further untilize entity type knowledge, and finally is decoded by CRF and Biaffine respectively.

3.1.1 Sequence Labeling with CRF

Suppose $h = (h_1, h_2, ..., h_N)$ is the encoded hidden layer sequence after Conditional LayerNorm, and $y = (y_1, y_2, ..., y_N)$ is the tag sequence, as shown in Fig 2. The score of sequence y is computed as follows,

$$s(h, y) = \sum_{n=1}^{N} \mathrm{W}_{n,y_n} + \sum_{n=2}^{N} T_{y_{n-1},y_n} \tag{1}$$

where W is the score matrix of each tag at each time step and T is the transition matrix between tags.

The probability of sequence y is calculated by softmax function, where $Y(h)$ represents all possible tag sequences.

$$p(y \mid h) = \frac{e^{s(h,y)}}{\sum_{\tilde{y} \in Y(h)} e^{s(h,\tilde{y})}} \tag{2}$$

The maximum likelihood loss function is used for training.

$$L_{\mathrm{CRF}} = \log(p(y \mid h)) \tag{3}$$

During inference, the predicted tag sequence with the maximum score is obtained with Viterbi algorithm.

Proceedings of the 21st China National Conference on Computational Linguistics, pages 836-847, Nanchang, China, October 14 - 16, 2022.
(c) Technical Committee on Computational Linguistics, Chinese Information Processing Society of China

$$y^* = \arg \max_{\widetilde{y} \in Y(h)} s(h, \widetilde{y}) \tag{4}$$

3.1.2 Span Boundary Detection with Biaffine

As shown in Fig 2, the hidden layer sequence after Conditional LayerNorm goes through a bidirectional LSTM and two seperate nonlinear layers to learn the representation of start and end of the span. Finally, the score of a span i is calculated by a Biaffine classifier as follows,

$$h_i^s = MLP_{\text{start}} \ (h_i) \tag{5}$$

$$h_i^e = MLP_{\text{end}} \ (h_i) \tag{6}$$

$$r(i) = h_i^{sT} U h_i^e + W \left(h_i^s \oplus h_i^e\right) + b \tag{7}$$

Where U is a $N * C * N$ tensor, W is a $2N * C$ matrix, b is the bias, N is the length of the sentence, C is the number of entity categories +1(non-entity).

We assign span i a NER category y_i:

$$y_i = \arg\max r(i) \tag{8}$$

The learning objective of our named entity recognizer is to assign a correct category to each valid span. Hence it is a multi-class classification problem and we optimise the model with softmax cross-entropy:

$$p\left(i_c\right) = \frac{\exp\left(r\left(i_c\right)\right)}{\sum_{\hat{c}=1}^{C} \exp\left(r\left(i_{\hat{c}}\right)\right)} \tag{9}$$

$$L_{\text{Biaffine}} = -\sum_{i=1}^{N} \sum_{c=1}^{C} y_{ic} \log p\left(i_c\right) \tag{10}$$

3.1.3 The Combined Loss

The final loss function of the model is weighted by the loss function of CRF and the loss function of Biaffine, as shown below:

$$L = \alpha^* L_{\text{CRF}} + \beta^* L_{\text{Biaffine}} \tag{11}$$

Where α and β are positive real number and their sum equals 1. They can be learned and updated iteratively with the training and we initialize both of them as 0.5.

3.2 Multi-strategies (MS)

Three strategies are adopted to enhance the performance, including Adaptive Pre-training (AP), Adversarial Training (AT) and model stacking with Cross Validation (CV). In order to reduce distribution differences between the task data and data used by the pre-trained model, we use CMeEE data for task-adaptive pre-training (Gururangan et al., 2020) based on the pre-trained model NEZHA (Wei et al., 2019) with Whole Word Masking (WWM) strategy to get a new domain adaptive pre-trained model CME-NEZHA. In order to improve the robustness of the model, we employ adversarial training (Kurakin et al., 2016) with Fast Gradient Method (FGM) strategy. Lastly, 5-fold cross validation is adopted to prevent model overfitting and exploit advantages of multi-models. Five models are trained and contribute equally to the final decision.

Proceedings of the 21st China National Conference on Computational Linguistics, pages 836-847, Nanchang, China, October 14 - 16, 2022.

Dataset	Training set	Validation set	Test set	Average sentence length
CMeEE	15000	5000	3000	>50 characters
CCKS2019	800	200	379	>390 characters

Table 1: Statistics of datasets

Entity Type	Entity number	Percent	Average entity length
bod	26589	28%	3.37
dis	24077	26%	5.35
sym	18579	20%	6.70
pro	9610	10%	5.30
dru	6331	7%	4.74
ite	4091	4%	4.37
mic	3019	3%	4.31
equ	1392	1%	4.30
dep	494	1%	2.86
Total	94182	100%	4.91
Anatomy	11520	49%	2.48
Disease	5535	23%	6.98
Drug	2307	10%	3.71
Laboratory	1785	8%	4.00
Image	1317	5%	3.79
Operation	1191	5%	12.85
Total	23655	100%	4.36

Table 2: Entity statistics of CMeEE and CCKS2019

4 Datasets

Two public datasets are used for experiments, CMeEE for nested NER and CCKS2019 for flat NER. Statistics of the two datasets are shown in Table 1, including sizes of the training, validate and test sets. As can be seen, the size of CMeEE is larger while the average text length of CCKS2019 is longer.

The texts of CMeEE are from textbooks of clinical pediatrics, which contain 9 types of entities, including Body (bod), Disease (dis), Symptom (sym), Medical procedure (pro), Medical equipment (equ), Drug (dru), Medical examination item (ite), Department (dep) and microorganism (mic). The texts of CCKS2019 are from electronic medical records, which contain 6 types of entities, including Disease and diagnosis, Image examination, Laboratory examination, Operation, Drug and Anatomy. As show in Table 2, the distributions of entities are imbalanced in both corpora. The top 3 dominant types of entities in CMeEE are bod, dis, and sym, while the top 3 dominant types of entities in CCKS2019 are Anatomy, Disease and Drug. On average, entities of sym and Operation are the longest in the two corpora respectively.

Flat entity	Nested entity	Percent of nested	Percent of nested in sym
84119	10063	10.68%	30.21%

Table 3: Nested entity statistics of CMeEE

As shown in Table 3, 10.68% of all entities in CMeEE are nested entities and 30.21% entities of sym are nested entities. Entities nested in sym entities are shown Table 4. All entity types except dep have entities nested in sym, where bod is the dominant type.

Proceedings of the 21st China National Conference on Computational Linguistics, pages 836-847, Nanchang, China, October 14 - 16, 2022.

Entity type	Number	Percent	Example of nested entity
bod	4706	84.84%	{无色胶冻样 [痰]bod}sym {Colorless jelly like [sputum]bod}sym
ite	486	8.76%	{[胸片]ite 异常}sym {[Chest radiograph]ite Abnormal}sym
dis	229	4.13%	{逐步发生全身弛缓性 [瘫痪]dis}sym {Progressive generalized flaccid [paralysis]dis}sym
pro	59	1.06%	{[肺部听诊]pro 呼吸音减弱}sym {[Lung auscultation]pro respiratory sound is reduced}sym
dru	28	0.50%	{[维生素 A]dru 摄入不足}sym {[vitamin A]dru Insufficient intake}sym
mic	26	0.47%	{气道分泌物 [细菌]mic 培养阳性}sym {Airway secretion [bacteria]mic culture positive}sym
equ	13	0.23%	{长期 [呼吸机]equ 依赖}sym {Long-term [respirator]equ dependence}sym

Table 4: Entities nested in sym

5 Experiments

5.1 Query Generation

As shown in Table 5, for CMeEE, we put example entities into the query, while for CCKS2019, we take the description of the entity type as the query.

5.2 Experimental Settings

We retrain the pre-trained model NEZHA based on the CMeEE corpus by 100 epochs. Then we fine-tune the model for NER by 4 epochs. We set the batch size to 16, dropout to 0.1, NEZHA learning rate to 2.5e-5, other learning rate to 2.5e-3, and maximum text length to 256. NVIDIA GTX2080Ti is used to run the program. Micro average F1 is chosen as the evaluation metric.

5.3 Comparison with Previous Models

5.3.1 Baselines on CMeEE Corpus

(1) MacBERT-large and Human are from (Zhang et al., 2022). MacBERT is variant of BERT, taking a MLM (Masked Language Model) as correction strategy. Human denotes the annotating result of human. (2) BERT-CRF, BERT-Biaffine and RICON are from (Gu et al., 2022). BERT-CRF solves sequence labeling with CRF, BERT-Biaffine detects span boundary with Biaffine, and RICON learns regularity inside entities. (3) Lattice-LSTM, Lattice-LSTM+Med-BERT, FLAT-Lattice, Medical-NER, and Medical NER+Med-BERT are from (Liu et al., 2021). Lattice-LSTM, Lattice-LSTM+Med-BERT and FLAT-Lattice incorporate lexicon to decide entity boundary. Medical NER and Medical NER+Med-BERT introduce big dictionary and pre-trained domain model.

5.3.2 Baselines on CCKS2019 Corpus

(1)BERT-BiLSTM-CRF is from (Dai et al., 2019), taking CRF for sequence labeling. (2)BBC+Lexicon+Glyph is from (Zhong and Yu, 2021), introducing lexicon and glyph information. (3) WB-Transformer+SA is from (Zhang et al., 2021), taking self-attention for semantic enrichment. (4) ELMo-lattice-LSTM-CRF is from (Li et al., 2020c), fusing ELMo and lexicon to improve sequence labeling performance. (5) ACNN is from (Kong et al., 2021), composed of hierarchical CNN and attention mechanism. (6) FS-TL is from (Li et al., 2019), fusing stroke information with transfer learning.

Proceedings of the 21st China National Conference on Computational Linguistics, pages 836-847, Nanchang, China, October 14 - 16, 2022.

Entity type	Query
bod	在文本中找出身体部位，例如细胞、皮肤、抗体 Find body parts in the text, for example, cells, skin and antibodies
dep	在文本中找出科室，例如科、室 Find departments in the text, for example, department and room
dis	在文本中找出疾病，例如癌症、病变、炎症、增生、肿瘤 Find diseases in the text, for example, cancer and pathological changes
dru	在文本中找出药物，例如胶囊、疫苗、剂 Find drugs in the text, for example, capsule, vaccine and agent
equ	在文本中找出医疗设备，例如装置、器、导管 Find medical devices in the text, for example, device and conduit
ite	在文本中找出医学检验项目，例如尿常规、血常规 Find medical test items in the text, for example, urine routine and blood routine
mic	在文本中找出微生物，例如病毒、病原体、抗原、核糖 Find micro organisms in the text, for example, virus and pathogen
pro	在文本中找出医疗程序，例如心电图、病理切片、检测 Find medical procedures in the text, for example, electrocardiogram and pathological section
sym	在文本中找出临床表现，例如疼痛、痉挛、异常 Find clinical manifestations in the text, for example, pain and spasm
Anatomy	找出疾病、症状和体征发生的人体解剖学部位 Find where in the human anatomy the disease, symptoms and signs occur
Disease	找出医学上定义的疾病和医生在临床工作中对病因、病生理、分型分期等所作的判断 Find medically defined diseases and physicians' judgments regarding etiology, pathophysiology, staging, etc., in clinical work-up
Drug	找出用于疾病治疗的具体化学物质 Find specific chemicals for disease treatment
Image	找出影像检查（X 线、CT、MR、PETCT 等）+ 造影 + 超声 + 心电图 Find imaging examinations (X-ray, CT, Mr, PETCT, etc.) + contrast + ultrasound + ECG
Laboratory	找出在实验室进行的物理或化学检查 Find physical or chemical examinations performed in the laboratory
Operation	找出医生在患者身体局部进行的切除、缝合等治疗，是外科的主要治疗方法 Find the main treatment in surgery that doctors perform locally on the patient's body, such as excision, suture, etc.

Table 5: Query for different entity types in CMeEE and CCKS2019

Proceedings of the 21st China National Conference on Computational Linguistics, pages 836-847, Nanchang, China, October 14 - 16, 2022.

Model	Precision/%	Recall/%	F1 score/%
MacBERT-large(Zhang et al., 2022)	-	-	62.40
Human(Zhang et al., 2022)	-	-	67.00
BERT-CRF(Gu et al., 2022)	58.34	64.08	61.07
BERT-Biaffine(Gu et al., 2022)	64.17	61.29	62.29
RICON(Gu et al., 2022)	66.25	64.89	65.57
Lattice-LSTM(Liu et al., 2021)	57.10	43.60	49.44
Lattice-LSTM+Med-BERT(Liu et al., 2021)	56.84	47.58	51.80
FLAT-Lattice(Liu et al., 2021)	66.90	70.10	68.46
Medical NER(Liu et al., 2021)	66.41	70.73	68.50
Medical NER+Med-BERT(Liu et al., 2021)	**67.99**	70.81	69.37
MRC-MTL-MS(Ours)	67.21	**71.89**	**69.47**

Table 6: Comparison with previous models on CMeEE

Model	Precision/%	Recall/%	F1 score/%
BERT-BiLSTM-CRF(Dai et al., 2019)	73.84	75.31	74.53
BBC+Lexicon+Glyph(Zhong and Yu, 2021)	85.17	84.13	84.64
WB-Transformer+SA(Zhang et al., 2021)	-	-	84.98
ACNN(Kong et al., 2021)	83.07	**87.29**	85.13
FS-TL(Li et al., 2019)	-	-	85.16
ELMo-lattice-LSTM-CRF(Li et al., 2020c)	84.69	85.35	85.02
MRC-MTL-MS(Ours)	**85.29**	85.32	**85.31**

Table 7: Comparison with previous models on CCKS2019

As shown in Table 6 and 7, our MRC-MTL-MS model outperforms all comparison models on both the nested NER corpus CMeEE and the flat NER corpus CCKS2019.

5.4 Ablation Experiments

The ablation experiments are shown in Table 8. MRC-Base is the same with (Li et al., 2020b), pointer network is used to detect span boundary. MRC-CRF only uses CRF for decoding. MRC-Biaffine only uses Biaffine for decoding. MRC-MTL integrates CRF and Biaffine with multi-task learning and use CRF as the final decoder. We can see that multi-task learning model outperforms single-task models. Adaptive Pre-training (AP), Adversarial Training (AT), and model stacking with Cross Validation (CV) strategies further improve the performance. Among which, CV contributes the most. Compared with MRC-Base, the improvement of F1 score on the nested NER corpus is 2.56%, which is higher than that of 1.63% on the flat NER corpus.

5.5 Experiments on Different Types of NEs

Experimental results of different types of NEs on the two corpora are shown in Table 9 respectively. As can be seen, on CMeEE, the entity type dru has the highest F1 score 81.17%, while the entity type ite has the lowest F1 score. The averagely longest and most nested entity type sym also has low F1 score and needs further study. The overall F1 scores on CCKS2019 are high and the entity type Drug also has the highest F1 score 95.25%, indicating that Drug entities are easier to recognize. For those entity types with low scores, like ite and Laboratory, constructing related lexicons maybe useful for improvement.

Model	CMeEE/%			CCKS2019/%		
	Precision	Recall	F1 score	Precision	Recall	F1 score
MRC-Base	67.98	65.87	66.91	82.63	84.76	83.68
MRC-CRF	67.17	67.25	67.21	84.40	84.91	84.65
MRC-Biaffine	**70.71**	64.09	67.24	83.22	83.77	83.49
MRC-MTL	64.58	71.76	67.98	84.42	84.97	84.70
+AP	66.28	70.34	68.25	84.23	85.24	84.73
+AP+AT	68.04	69.16	68.59	84.20	**85.39**	84.79
+AP+AT+CV	67.21	**71.89**	**69.47**	**85.29**	85.32	**85.31**

Table 8: Ablation experiments on CMeEE and CCKS2019

Entity type	Precision/%	Recall/%	F1 score/%
bod	62.92	71.33	66.86
dis	76.78	80.69	78.69
dru	75.38	87.93	81.17
dep	54.24	88.89	67.37
equ	74.48	81.20	77.70
ite	51.06	49.23	50.13
mic	76.64	82.16	79.30
pro	61.91	71.50	66.36
sym	58.49	54.68	56.52
Mac-Avg	65.77	74.18	69.72
Anatomy	85.25	87.07	86.15
Disease	85.63	85.56	85.60
Drug	95.45	95.05	95.25
Image	86.65	87.64	87.14
Laboratory	74.54	67.97	71.10
Operation	85.91	79.01	82.32
Mac-Avg	85.57	83.72	84.63

Table 9: Results of different types of NEs on CMeEE and CCKS2019

Proceedings of the 21st China National Conference on Computational Linguistics, pages 836-847, Nanchang, China, October 14 - 16, 2022.

5.6 Case Study

Table 10 gives two examples from CMeEE. In the first example, the MRC-Base model does not correctly detect the boundary of the entity “郎飞结上的补体被激活”(Complement on Ranvier knot is activated), while the MRC-MTL-MS model correctly recognizes the boundary and the entity type. In the second example, the MRC-Base model correctly detects the boundary of the entity “高血压”(hypertension), but predicts a wrong label. The MRC-MTL-MS model correctly recognizes the polysemous entity, indicating its superiority in disambiguating polysemous entities.

Sentence	AMAN 的一个早期表现就是郎飞结上的补体被激活。 An early manifestation of AMAN is that complement on Ranvier knot is activated.
Entity	郎飞结上的补体被激活 Complement on Ranvier knot is activated.
Golden Labels	B-SYM I-SYM I-SYM I-SYM I-SYM I-SYM I-SYM I-SYM I-SYM E-SYM
MRC	B-BOD I-BOD E-BOD O O O O O O O
MRC-MTL-MS	B-SYM I-SYM I-SYM I-SYM I-SYM I-SYM I-SYM I-SYM I-SYM E-SYM
Sentence	患儿情况好，只 1 例发生慢性排异及高血压。 The condition of the child is good, and only one develops chronic rejection and hypertension.
Entity	高血压 hypertension
Golden Labels	B-SYM I-SYM E-SYM
MRC	B-DIS I-DIS E-DIS
MRC-MTL-MS	B-SYM I-SYM I-SYM

Table 10: Two cases with labels BIES

6 Conclusion

This paper proposes a MRC-based multi-task model for Chinese medical NER, enhancing MRC-CRF with Biaffine to recognize the named entities more accurately. To further improve the model, we introduce multi-strategies, including adaptive pre-training, adversarial training and model stacking with cross validation. Our model can cope with both flat NER and nested NER. Experiments on the nested NER corpus CMeEE and the flat NER corpus CCKS2019 show the effectiveness of our model. In the future, we will incorporate domain knowledge to improve the recognition performance on hard named entities.

7 Acknowledgements

We would like to thank the anonymous reviewers for their insightful and valuable comments. This work was supported in part by Major Program of National Social Science Foundation of China (Grant No.17ZDA318, 18ZDA295), National Natural Science Foundation of China (Grant No.62006211), and China Postdoctoral Science Foundation (Grant No.2019TQ0286, 2020M682349).

References

Jun Cao, Xian Zhou, Wangping Xiong, Ming Yang, Jianqiang Du, Yanyun Yang, and Tianci Li. 2021. Electronic medical record entity recognition via machine reading comprehension and biaffine. *Discrete Dynamics in Nature and Society*, 2021.

Xianglong Chen, Chunping Ouyang, Yongbin Liu, and Yi Bu. 2020. Improving the named entity recognition of chinese electronic medical records by combining domain dictionary and rules. *International Journal of Environmental Research and Public Health*, 17(8):2687.

Shanta Chowdhury, Xishuang Dong, Lijun Qian, Xiangfang Li, Yi Guan, Jinfeng Yang, and Qiubin Yu. 2018. A multitask bi-directional rnn model for named entity recognition on chinese electronic medical records. *BMC bioinformatics*, 19(17):75–84.

Zhenjin Dai, Xutao Wang, Pin Ni, Yuming Li, Gangmin Li, and Xuming Bai. 2019. Named entity recognition using bert bilstm crf for chinese electronic health records. In *2019 12th international congress on image and signal processing, biomedical engineering and informatics (cisp-bmei)*, pages 1–5.

Yingjie Gu, Xiaoye Qu, Zhefeng Wang, Yi Zheng, Baoxing Huai, and Nicholas Jing Yuan. 2022. Delving deep into regularity: A simple but effective method for chinese named entity recognition. *arXiv preprint arXiv:2204.05544*.

Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A Smith. 2020. Don' t stop pretraining: Adapt language models to domains and tasks. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8342–8360.

Xianpei Han, Zhichun Wang, Jiangtao Zhang, Qinghua Wen, and Yinnian Lin. 2020. Overview of the ccks 2019 knowledge graph evaluation track: entity, relation, event and qa. *arXiv preprint arXiv:2003.03875*.

Jun Kong, Leixin Zhang, Min Jiang, and Tianshan Liu. 2021. Incorporating multi-level cnn and attention mechanism for chinese clinical named entity recognition. *Journal of Biomedical Informatics*, 116:103737.

Alexey Kurakin, Ian Goodfellow, and Samy Bengio. 2016. Adversarial machine learning at scale. *arXiv preprint arXiv:1611.01236*.

John D Lafferty, Andrew McCallum, and Fernando CN Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, pages 282–289.

Nan Li, Ling Luo, Zeyuan Ding, Yawen Song, Zhihao Yang, and Hongfei Lin. 2019. Dutir at the ccks-2019 task1: improving chinese clinical named entity recognition using stroke elmo and transfer learning. In *Proceedings of the 4th China Conference on Knowledge Graph and Semantic Computing (CCKS 2019)*, pages 24–27.

Xiangyang Li, Huan Zhang, and Xiao-Hua Zhou. 2020a. Chinese clinical named entity recognition with variant neural structures based on bert methods. *Journal of biomedical informatics*, 107:103422.

Xiaoya Li, Jingrong Feng, Yuxian Meng, Qinghong Han, Fei Wu, and Jiwei Li. 2020b. A unified mrc framework for named entity recognition. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5849–5859.

Yongbin Li, Xiaohua Wang, Linhu Hui, Liping Zou, Hongjin Li, Luo Xu, Weihai Liu, et al. 2020c. Chinese clinical named entity recognition in electronic medical records: Development of a lattice long short-term memory model with contextualized character representations. *JMIR Medical Informatics*, 8(9):e19848.

Ning Liu, Qian Hu, Huayun Xu, Xing Xu, and Mengxin Chen. 2021. Med-bert: A pre-training framework for medical records named entity recognition. *IEEE Transactions on Industrial Informatics*.

Ling Luo, Zhihao Yang, Yawen Song, Nan Li, and Hongfei Lin. 2020. Chinese clinical named entity recognition based on stroke elmo and multi-task learning. *Chinese Journal of Computers*, 43(10):1943–1957.

Qiuli Qin, Shuang Zhao, and Chunmei Liu. 2021. A bert-bigru-crf model for entity recognition of chinese electronic medical records. *Complexity*, 2021.

Qian Wan, Jie Liu, Luona Wei, and Bin Ji. 2020. A self-attention based neural architecture for chinese medical named entity recognition. *Mathematical Biosciences and Engineering*, 17(4):3498–3511.

Proceedings of the 21st China National Conference on Computational Linguistics, pages 836-847, Nanchang, China, October 14 - 16, 2022.

Junqiu Wei, Xiaozhe Ren, Xiaoguang Li, Wenyong Huang, Yi Liao, Yasheng Wang, Jiashu Lin, Xin Jiang, Xiao Chen, and Qun Liu. 2019. Nezha: Neural contextualized representation for chinese language understanding. *arXiv preprint arXiv:1909.00204*.

Zhichang Zhang, Xiaohui Qin, Yanlong Qiu, and Dan Liu. 2021. Well-behaved transformer for chinese medical ner. In *2021 3rd International Conference on Natural Language Processing (ICNLP)*, pages 162–167.

Ningyu Zhang, Mosha Chen, Zhen Bi, Xiaozhuan Liang, Lei Li, Xin Shang, Kangping Yin, Chuanqi Tan, Jian Xu, Fei Huang, et al. 2022. Cblue: A chinese biomedical language understanding evaluation benchmark. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7888–7915.

Shan Zhao, Zhiping Cai, Haiwen Chen, Ye Wang, Fang Liu, and Anfeng Liu. 2019. Adversarial training based lattice lstm for chinese clinical named entity recognition. *Journal of biomedical informatics*, 99:103290.

Hengyi Zheng, Bin Qin, and Ming Xu. 2021. Chinese medical named entity recognition using crf-mt-adapt and ner-mrc. In *2021 2nd International Conference on Computing and Data Science (CDS)*, pages 362–365.

Shanhao Zhong and Qingsong Yu. 2021. Improving chinese medical named entity recognition using glyph and lexicon. In *Proceedings of 2021 International Conference on Advanced Education and Information Management (AEIM 2021)*, pages 75–80.

Qinglin Zhu, Zijie Lin, Yice Zhang, Jingyi Sun, Xiang Li, Qihui Lin, Yixue Dang, and Ruifeng Xu. 2021. Hitsz-hlt at semeval-2021 task 5: Ensemble sequence labeling and span boundary detection for toxic span detection. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, pages 521–526.

Proceedings of the 21st China National Conference on Computational Linguistics, pages 836-847, Nanchang, China, October 14 - 16, 2022.