The 34th Conference on Computational Linguistics and Speech Processing (ROCLING 2022)
Taipei, Taiwan, November 21-22, 2022. The Association for Computational Linguistics and Chinese Language Processing

Overview of the ROCLING 2022 Shared Task for Chinese Healthcare Named Entity Recognition

Lung-Hao Lee, Chao-Yi Chen
Department of Electrical Engineering
National Central University
lhlee@ee.ncu.edu.tw, 110581007@cc.ncu.edu.tw

Liang-Chih Yu
Department of Information Management
Yuan Ze University
lcyu@saturn.yzu.edu.tw

Yuen-Hsien Tseng
Graduate Institute of Library and Information Studies
National Taiwan Normal University
samtseng@ntnu.edu.tw

Abstract

This paper describes the ROCLING-2022 shared task for Chinese healthcare named entity recognition, including task description, data preparation, performance metrics, and evaluation results. Among ten registered teams, seven participating teams submitted a total of 20 runs. This shared task reveals present NLP techniques for dealing with Chinese named entity recognition in the healthcare domain. All data sets with gold standards and evaluation scripts used in this shared task are publicly available for future research.

Keywords: named entity recognition, information extraction, health informatics, Chinese language processing

1 Introduction

Named Entity Recognition (NER) is a traditional and fundamental NLP task in the information extraction domain that locates and identifies mentions of named entities (e.g., person, organization, and location) in unstructured texts. The NER task is usually regarded as a sequence labeling problem, where entity boundaries and category labels are jointed predicted.

Chinese NER is correlated with word segmentation, since named entity boundaries are also word boundaries. Due to a lack of delimiters between characters and a lack of conventional features like capitalization, Chinese NER is more difficult to process than English NER. Incorrect word segmentation will cause error propagation in NER. For example, "思覺失調症" (schizophrenia) is a kind of mental disorder that affects the way a person thinks, feels, perceives reality, and relates to others. This named entity may be incorrectly segmented into three words: "思覺" (thinking and feeling), "失調" (disorder) and "症" (disease), resulting in fail to recognize it as a named entity belonging to disease type. Character-based methods have been found to outperform word-based approaches for breaking through this word segmentation limitation in Chinese NER (He and Wang, 2008; Li et al., 2014, Zhang and Yang, 2018).

Various methods have been proposed to tackle Chinese NER tasks. In addition to machine learning approaches, such as HMM (Hidden Markov Model) (Fu and Luke, 2005), Markov logistic network (Yu, 2007), and CRF (Conditional Random Field) (Chen et al., 2006), deep learning techniques have been widely used, with mostly promising results. A character-based LSTM (Long Short-Term Memory)-CRF model with radical-level features was proposed for Chinese NER (Dong et al., 2016). The BiLSTM (Bidirectional LSTM)-CRF model was trained based on character-word mixed embeddings to improve the recognition effectiveness of Chinese NER (E and Xiang., 2017). A BiLSTM-CRF model with a self-attention mechanism was proposed to integrate part-of-speech labeling information to capture the semantic features of input sequences for Chinese clinical NER (Wu et al., 2019). A residual dilated CNN (Convolution Neural Network) with CRF was also presented to enhance Chinese clinical

Entity Type (Tag)	Description	Examples
Body (BODY)	The whole physical structure that forms a person or animal including biological cells, organizations, organs and systems.	“細胞核” (nucleus), “神經組織” (nerve tissue), “左心房” (left atrium), “脊髓” (spinal cord), “呼吸系統” (respiratory system)
Symptom (SYMP)	Any feeling of illness or physical or mental change that is caused by a particular disease.	“流鼻水” (rhinorrhea), “咳嗽” (cough), “貧血” (anemia), “失眠” (insomnia), “心悸” (palpitation), “耳鳴” (tinnitus)
Instrument (INST)	A tool or other device used for performing a particular medical task such as diagnosis and treatments.	“血壓計” (blood pressure meter), “達文西手臂” (DaVinci Robots), “體脂肪計” (body fat monitor), “雷射手術刀” (laser scalpel)
Examination (EXAM)	The act of looking at or checking something carefully in order to discover possible diseases.	“聽力檢查” (hearing test), “腦電波圖” (electroencephalography; EEG), “核磁共振造影” (magnetic resonance imaging; MRI)
Chemical (CHEM)	Any basic chemical element typically found in the human body.	“去氧核糖核酸” (deoxyribonucleic acid; DNA), “糖化血色素” (glycated hemoglobin), “膽固醇” (cholesterol), “尿酸” (uric acid)
Disease (DISE)	An illness of people or animals caused by infection or a failure of health rather than by an accident.	“小兒麻痺症” (poliomyelitis; polio), “帕金森氏症” (Parkinson’s disease), “青光眼” (glaucoma), “肺結核” (tuberculosis)
Drug (DRUG)	Any natural or artificially made chemical used as a medicine.	“阿斯匹靈” (aspirin), “普拿疼” (acetaminophen), “青黴素” (penicillin), “流感疫苗” (influenza vaccination)
Supplement (SUPP)	Something added to something else to improve human health.	“維他命” (vitamin), “膠原蛋白” (collagen), “益生菌” (probiotics), “葡萄糖胺” (glucosamine), “葉黃素” (lutein)
Treatment (TREAT)	A method of behavior used to treat diseases	“藥物治療” (pharmacotherapy), “胃切除術” (gastrectomy), “標靶治療” (targeted therapy), “外科手術” (surgery)
Time (TIME)	Element of existence measured in minutes, days, years	“嬰兒期” (infancy), “幼兒時期” (early childhood), “青春期” (adolescence), “生理期” (on one’s period), “孕期” (pregnancy)

Table 1: Named entity types with descriptions and examples (Lee and Lu, 2021).

NER in terms of computational performance and training time (Qiu et al., 2019). A BERT-BiLSTM-CRF model was proposed to use BERT embedding for character representation and to train the BiLSTM-CRF model to recognize complex named entities (Lee et al., 2022).

Prior to scheduling a doctor’s appointment for diagnosis and treatment of a perceived medical issues, people frequently seek healthcare-related information online from health-related news articles, digital health services, and medical question-answering forums. Domain-specific healthcare information usually includes many proper names. These often take the form of named entities such as “三酸甘油酯” (triglyceride), “電腦斷層掃描” (computer tomography, CT) and “靜脈免疫球蛋白注射” (intravenous immunoglobulin, IVIG), presenting language processing challenges for healthcare-related applications. Responding to this pronounced challenge in the healthcare domain, the ROCLING-2022 conference features a Chinese healthcare NER task, providing an evaluation platform for the development and implementation of Chinese healthcare NER system. Given a Chinese sentence, the NER system is expected to automatically recognize healthcare entities such as symptoms, chemicals, diseases, and treatments.

The rest of this article is organized as follows. Section 2 provides a description of the Chinese healthcare NER shared task. Section 3 introduces the constructed data sets. Section 4 describes the evaluation metrics. Section 5 compares evaluation results from the various participating teams. Finally, we conclude this paper with findings and offer future research directions in Section 6.

2 Task Description

The goal of this shared task is to develop and evaluate the capability of a Chinese healthcare NER recognizer. A sentence containing at least one named entity is given as the input. The recognizer should predict the named entity’s boundaries and category for each given sentence. We use the common BIO (Beginning, Inside, and Outside) format for the NER task. The B-prefix before a tag indicates that the character is the beginning of a named entity and the I-prefix before a tag indicates that the character is inside a named entity. An O tag indicates that a character belongs to no named entity. We use the same entity types defined in the Chinese HealthNER Corpus (Lee and Lu, 2021). A total of 10 types are described for this Chinese healthcare NER task, and some examples are provided in Table 1.

The input is a sentence consisting of a sequence of character-based tokens including punctuation. The developed NER recognizer returns the corresponding BIO tags aligned to each token as the output. Example sentences are presented below. In Example 1, "肌肉" (muscle) and "骨骼" (skeleton) belong to the body entity type (denoted as BODY). "蛋白質" (protein) and "鈣質" (calcium) are chemicals (denoted as CHEM). In Example 2, we can find a disease "胃食道逆流症" (gastroesophageal reflux disease) (denoted as DISE).

Example 1

- *Input*: 修復肌肉與骨骼罪狀要的便是熱量、蛋白質與鈣質。
- *Output*: O, O, B-Body, I-Body, O, B-Body, I-Body, O, O, O, O, O, O, O, O, O, B-CHEM, I-CHEM, I-CHEM, O, B-CHEM, I-CHEM, O

Example 2

- *Input*: 如何治療胃食道逆流症？
- *Output*: O, O, O, O, B-DISE, I-DISE, I-DISE, I-DISE, I-DISE, O

3 Data Preparation

The Chinese HealthNER Corpus (Lee and Lu, 2021) was used as the training set. It includes 30,692 sentences with a total around 1.5 million characters or 91,700 words. The data was sourced from articles on websites that provide healthcare information, on-line health news and medical question/answer forums. After manual annotation, this corpus consists of 68460 named entities across 10 defined entity types.

We use the existing named entities in the Chinese HealthNER Corpus as the query terms and to find the corresponding articles in Chinese Wikipedia (zh_TW version). The first paragraph in the wiki articles was segmented into sentences for manual annotation. Three graduate students majoring in electrical engineering were trained in the named entity tagging task, producing a Fleiss' Kappa value of inter-annotator agreement of 89%. All annotators were asked to discuss differences and seek consensus. When agreement was reached, each annotator was then asked to process sentences individually. As a result, our constructed test set includes 3,205 sentences with a total of 118,116 characters and 13,369 named entities.

Entity Type	#Train (%)	#Test (%)
Body	26411 (38.58%)	5315 (39.76%)
Symptom	12904 (18.85%)	1944 (14.54%)
Instrument	1089 (1.59%)	250 (1.87%)
Examination	2622 (3.83%)	207 (1.55%)
Chemical	6834 (9.98%)	1718 (12.85%)
Disease	10079 (14.72%)	2609 (19.52%)
Drug	2225 (3.25%)	481 (3.60%)
Supplement	1525 (2.23%)	183 (1.37%)
Treatment	3108 (4.54%)	468 (3.50%)
Time	1663 (2.43%)	194 (1.44%)
Total	68460 (100%)	13,369 (100%)

Table 2: Detailed data statistics.

Table 2 shows detailed statistics of mutually exclusive training and test sets. The entity type distribution is similar in both the training and test sets. The most frequently occurring type was Body, followed by Symptom, Disease and Chemical, collectively accounting for about 83% of all named entity instances, with the remaining 6 types accounting for 17%.

In addition, sentences in the training set may contain named entities or not, each with an average of 49.31 characters and 2.23 named entities. However, all sentences in the test set contained at least one named entity, each with an average of 36.85 characters and 4.17 named entities. In summary, the average sentence length is short in the test set, but named entity density is relatively high.

4 Performance Metrics

Performance is evaluated by examining the difference between the machine-predicted and human-annotated BIO tags. Standard precision, recall and F1-score are the most typical evaluation metrics of NER systems at a character level, and are used here. If the predicted tag of a character in terms of BIO format was completely identical with the gold standard, the character in the testing instances was regarded as correctly recognized.

Rank	Team	Affiliation	Run#	Precision (%)	Recall (%)	F1
1	MIGBasline	National Chengchi University	Run 3	**81.99**	**81.88**	**81.93**
2	SCU-MESCLab	Soochow University	Run 3	80.18	78.3	79.23
3	crowNER	National Taiwan University	Run 1	77.82	78.1	77.96
4	YNU-HPCC	Yunnan University	Run 1	77.22	78.15	77.68
5	NERVE	National Kaohsiung University of Science and Technology	Run 1	79.59	73.09	76.2
6	NCU1415	National Central University	Run 2	74.56	72.81	73.68
7	SCU-NLP	Soochow University	Run 2	64.72	77.92	70.71

Table 3: Testing results of Chinese health named entity recognition task.

Precision is defined as the percentage of named entities found by the NER system that are correct. Recall is the percentage of named entities present in the test set found by the NER system. The F1-score is the harmonic mean of precision and recall.

5 Evaluation Results

The policy of this shared task is an open test. Participating systems are allowed to use other publicly available data for this shared task, but the usage should be specified in their system description paper. Each team was allowed to provide at most three submissions during the evaluation period. Among ten registered teams, seven submitted their testing results, providing a total of 20 submissions, from which the submission with the best F1-score of each team was kept in the leaderboard for performance ranking.

Table 3 summarizes the task testing results. NCU1415 team (Feng et al., 2022) uses BERT (Devlin et al., 2019) to encode sentences, followed by CRF for sequence labeling. SCU-MESCLab (Yang et al., 2022) represents sentences based on RoBERTa (Liu et al., 2019) embeddings, followed by BiLSTM-CRF to recognize named entities. NERVE (Lin et al., 2022) compares three NER frameworks based on BERT transformers and lexicons. SCU-NLP (Chiou et al., 2022) compares experimental results of well-known models, including random forest, HMM, CRF, and BERT and provides error analysis. The crowNER team (Chi et al., 2022) adopts adversarial learning and mixed precision training techniques to improve the performance achieved by MacBERT-CRF. YNU-HPCC (Luo et al. 2022) applies focal loss and regularized dropout mechanisms to enhance BERT-BiLSTM-CRF model performance. MIGBaseline team (Ma et al., 2022) uses PERT (Cui et al., 2022) as embedding representations to train the W2NER model (Li et al., 2022), achieving the best F1 score of 81.93 at this shared task evaluation.

In summary, the overall best results came from the MIGBaseline team (Ma et al., 2022), whose approach achieved the best scores across all the evaluation metrics, followed by SCU-MESCLab (Yang et al., 2022) and crowNER (Chi et al., 2022). The most frequently used neural architecture in this shared task is BiLSTM-CRF, which usually achieved promising results, matching findings from related studies for named entity recognition in the English language (Chiu and Nichols, 2016; Lample et al., 2016; Ma and Hovy, 2016; Liu et al., 2018).

6 Conclusions and Future Work

This paper provides an overview of the ROCLING-2022 shared task for Chinese healthcare named entity recognition, including task design, data preparation, performance metrics and evaluation results. We received a total of 20 testing submissions from seven participating teams. Regardless of actual performance, all submissions contribute to the development of an effective named entity recognition solution in the healthcare domain, and the individual system description papers for this shared task provide useful insights into Chinese language processing.

We hope the data sets collected and annotated for this shared task can facilitate and expedite future development of named entity recognizers. Therefore, in addition to publicly accessed Chinese HealthNER Corpus as the training set, the test set with gold standards and evaluation scripts are available from a public GitHub repository as follows

- Chinese HealthNER Corpus
https://github.com/NCUEE-NLPLab/Chinese-HealthNER-Corpus

- ROCLING-2022 Shared Task
https://github.com/NCUEE-NLPLab/ROCLING-2022-ST-CHNER

Future directions will focus on the development of Chinese healthcare entity-relationship extraction. We plan to build new language resources to develop techniques for the future enrichment of the research topic in open information extraction.

Acknowledgments

We thank all the participants for taking part in our shared task. We appreciate Tzu-Mi Lin, Man-Chen Hung, and Chien-Huan Lu for their efforts in data annotations. This work is partially supported by the National Science and Technology Council, Taiwan, under the grant MOST 111-2628-E-008-005-MY3, MOST 111-2628-E-155-001-MY2, and MOST 109-2410-H-003-123-MY3.

References

Aitao Chen, Fuchun Peng, Roy Shan, Gordon Sun. 2006. Chinese named entity recognition with conditional probabilistic models. In *Proceedings of the 5th SIGHAN Workshop on Chinese Language Processing*. Association for Computational Linguistics, pages 173-176.

Te-Yu Chi, Chiu-Hsia Chang, and Te-Lun Yang. 2022. crowNER at ROCLING 2022 shared task: NER using MacBERT and adversarial learning. In *Proceedings of the 34th Conference on Computational Linguistics and Speech Processing.*

Sung-Ting Chiou, Sheng-Wei Huang, Ying-Chun Lo, Yu-Hsuan Wu, and Jheng-Long Wu. 2022. SCU-NLP at ROCLING 2022 shared task: experiment and error analysis of biomedical entity detection model. In *Proceedings of the 34th Conference on Computational Linguistics and Speech Processing.*

Jason P. C. Chiu, and Eric Nichols. 2016. Named entity recognition with bidirectional LSTM-CNNs. *Transactions of the Association for Computational Linguistics*, 4: 357-370. http://dx.doi.org/10.1162/tacl_a_00104.

Yiming Cui, Ziqing Yang, and Ting Liu. 2022. PERT: pre-training BERT with permuted language model. *arXiv:2203.06906*. https://doi.org/10.48550/arXiv.2203.06906.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, pages 4171–4186. http://dx.doi.org/10.18653/v1/N19-1423.

Chuanhai Dong, Jiajun Zhang, Chengqing Zong, Masanori Hattori, and Hui Di. 2016. Character-based LSTM-CRF with radical-level features for Chinese named entity recognition. *Lecture Notes in Computer Science: Natural Language Understanding and Intelligent Applications*, 10102: 239-250. https://doi.org/10.1007/978-3-319-50496-4_20

Shijia E, and Yang Xiang. 2017. Chinese named entity recognition with character-word mixed embedding. In *Proceedings of the 26th ACM International Conference on Information and Knowledge Management*. Association for Computing Machinery, pages 2055-2058. https://doi.org/10.1145/3132847.3133088

Zhi-Quan Feng, Po-Kai Chen, and Jia-Ching Wang. 2022. NCU1415 at ROCLING 2022 shared task: a light-weight transformer-based approach for biomedical named entity recognition. In *Proceedings of the 34th Conference on Computational Linguistics and Speech Processing.*

Guohong Fu, and Kang-Kwong Luke. 2005. Chinese named entity recognition using lexicalized HMMs. *ACM SIGKDD Explorations Newsletter*, 7(1): 19-25. https://doi.org/10.1145/1089815.1089819.

Jingzhou He, and Houfeng Wang. 2008. Chinese named entity recognition and word segmentation based on character. In *Proceedings of the 6th SIGHAN Workshop on Chinese Language Processing*. Association for Computational Linguistics, pages 128–132.

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer.

2016. Neural architecture for named entity recognition. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, pages 260-270. http://dx.doi.org/10.18653/v1/N16-1030.

Lung-Hao Lee, and Yi Lu. 2021. Multiple embeddings enhanced multi-graph neural networks for Chinese healthcare named entity recognition. *IEEE Journal of Biomedical and Health Informatics*, 25(7): 2801-2810. https://doi.org/10.1109/JBHI.2020.3048700.

Lung-Hao Lee, Chien-Huan Lu, and Tzu-Mi Lin. 2022. NCUEE-NLP at SemEval-2022 task 11: Chinese named entity recognition using the BERT-BiLSTM-CRF model. In *Proceedings of the 16th International Workshop on Semantic Evaluation*. Association for Computational Linguistics, pages 1597-1602.

Haibo Li, Masato Hagiwara, Qi Li, and Heng Ji. 2014. Comparison of the impact of word segmentation on name tagging for Chinese and Japanese. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*. European Language Resources Association, pages 2532-2536.

Jingye Li, Donghong Ji, Jiang Liu, Hao Fei, Meishan Zhang, Shengqiong Wu, Chong Teng, and Fei Li. 2022. Unified named entity recognition as word-word relation classification. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence*.

Bo-Shau Lin, Jian-He Chen, and Tao-Hsing Chang. 2022. NERVE at ROCLING 2022 shared task: a comparison of three named entity recognition frameworks based on language model and lexicon approach. In *Proceedings of the 34th Conference on Computational Linguistics and Speech Processing*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, Veselin Stoyanov. 2019. RoBERTa: a robustly optimized BERT pretraining approach. *arXiv:1907.11692* https://doi.org/10.48550/arXiv.1907.11692

Liyuan Liu, Jingbo Shang, Xiang Ren, Frank F. Xu, Huan Gui, Jian Peng, Jiawei Han. 2018. Empower sequence labeling with task-aware neural language model. In P*roceeding of the 32nd AAAI Conference on Artificial Intelligence*. Association for Computing Machinery, pages 5253-5260.

Xiang Luo, Jin Wang, and Xuejie Zhang. 2022. YNU-HPCC at ROCLING 2022 shared task: a transformer-based model with focal loss and regularization dropout for Chinese healthcare named entity recognition. In *Proceedings of the 34th Conference on Computational Linguistics and Speech Processing*.

Xuezhe Ma, and Eduard Hovy. 2016. End-to-end sequence labeling via Bi-directional LSTM-CNNs-CRF. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, pages 1064-1074. http://dx.doi.org/10.18653/v1/P16-1101.

Hsing-Yuan Ma, Wei-Jie Li, and Chao-Lin Liu. 2022. MIGBaseline at ROCLING 2022 shared task: reports on named entity recognition using Chinese healthcare datasets. In *Proceedings of the 34th Conference on Computational Linguistics and Speech Processing*.

Jiahui Qiu, Yangming Zhou, Qi Wang, Tong Ruan, and Ju Gao. 2019. Chinese clinical named entity recognition using residual dilated convolutional neural network with conditional random field. *IEEE Transactions on NanoBioscience*, 18(3): 306-315. https://doi.org/10.1109/TNB.2019.2908678.

Guohua Wu, Guangen Tang, Zhongru Wang, Zhen Zhang, and Zhen Wang. 2019. An attention-based BiLSTM-CRF model for Chinese clinic named entity recognition. *IEEE Access*, 7: 113942-113949. https://doi.org/10.1109/ACCESS.2019.2935223.

Tsung-Hsien Yang, Ruei-Cyuan Su, Tzu-En Su, Sing-Seong Chong, and Ming-Hsiang Su. 2022. SCU-MESCLab at ROCLING 2022 shared task: named entity recognition using BERT classifier. In *Proceedings of the 34th Conference on Computational Linguistics and Speech Processing*.

Xiaofeng Yu. 2007. Chinese named entity recognition with cascaded hybrid model. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume, Short Papers*. Association for Computational Linguistics, pages 197–200.

Yue Zhang, and Jie Yang. 2018. Chinese NER using lattice LSTM. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, pages 1554–1564. http://dx.doi.org/10.18653/v1/P18-1144.