Overview of CCKS 2020 Task 3: Named Entity Recognition and Event Extraction in Chinese Electronic Medical Records Postprint

Authors: Xia, Li, Qinghua, Wen, Hu, Lin, Zengtao, Jiao, Jiangtao, Zhang, Jiangtao, Zhang

Date: 2022-11-27T00:00:00+00:00

Abstract

The China Conference on Knowledge Graph and Semantic Computing (CCKS) 2020 Evaluation Task 3 addressed clinical named entity recognition and event extraction for Chinese electronic medical records. Two annotated datasets and additional resources for these two subtasks were provided to participants. This evaluation competition attracted 354 teams, of which 46 successfully submitted valid results. Pre-trained language models were widely applied in this evaluation task. Data augmentation and external resources were also helpful.

Full Text

Preamble

Overview of CCKS 2020 Task 3: Named Entity Recognition and Event Extraction in Chinese Electronic Medical Records

Xia Li[1], Qinghua Wen[2], Hu Lin[1], Zengtao Jiao[3] & Jiangtao Zhang[1],[2]†
[1]The 305th Hospital of the Chinese People' s Liberation Army, Wenjin Street, Xicheng District, Beijing 100017, China
[2]Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
[3]Yiducloud Beijing Technology Co., Ltd., Huayuan North Road, Haidian District, Beijing 100089, China

Keywords: Chinese electronic medical records; Event extraction; Named entity recognition; Clinical text; CCKS

Citation: Li, X., et al.: Overview of CCKS 2020 Task 3: Named entity recognition and event extraction in Chinese electronic medical records. Data Intelligence 3(3), 376-388 (2021). doi: 10.1162/dint_a_{00093}

Received: February 11, 2021; **Revised:** March 11, 2021; **Accepted:** March 15, 2021

Abstract

The China Conference on Knowledge Graph and Semantic Computing (CCKS) 2020 Evaluation Task 3 focused on clinical named entity recognition and event extraction for Chinese electronic medical records. Participants were provided with two annotated datasets and additional resources for these subtasks. The competition attracted 354 teams, with 46 successfully submitting valid results. Pre-trained language models were widely applied, and data augmentation techniques along with external resources proved helpful.

1. Introduction

The China Conference on Knowledge Graph and Semantic Computing (CCKS), founded in 2016 and organized by the Chinese Information Processing Society of China, provides evaluation tasks to promote technological development in knowledge graph and semantic computing. In 2020, CCKS offered eight evaluation tasks, with Task 3 focusing specifically on named entity recognition (NER) and event extraction (EE) in Chinese electronic medical records (EMRs).

EMRs represent the core assets of hospitals and typically exist as semi-structured data containing substantial free text. Despite accumulating vast quantities of EMR data, most remain underutilized due to the difficulty of processing free text. NER and EE serve as essential techniques for extracting useful information from such text. NER in EMRs, also known as clinical named entity recognition (CNER), enables identification of diseases, drugs, and other medical entities. The most popular NER methods employ sequence labeling based on long short-term memory (LSTM) [**?**, **?**, **?**] or bidirectional encoder representations from transformers (BERT) [**?**]. Clinical event extraction identifies medical events such as tumor sites, tumor sizes, and metastasis locations, with LSTM, BERT, and other methods applied to this task.

Traditional NER and EE rely on supervised models, yet annotating clinical information proves far more challenging than general domain annotation. While public medical datasets exist for NER tasks, such as i2b2 [**?**], ShARe CLEF eHealth [**?**], and SemEval [**?**], Chinese medical datasets remain scarce. To advance semantic analysis of Chinese EMRs, the Knowledge Engineering Group of Tsinghua University and Yiducloud Beijing Technology Co., Ltd. organized this evaluation challenge at CCKS 2020. The datasets, provided by Yiducloud, are restricted to CCKS evaluation purposes only.

2. Related Work

CCKS 2020 Task 3 centers on NER and EE in Chinese EMRs, which have long been core problems in natural language processing.

2.1 Chinese NER

Named entity recognition involves locating and classifying words or expressions in unstructured text. In English NER, LSTM-CRF (Conditional Random Field) models [**?**, **?**, **?**] represent a classic approach that leverages both character-level and word-level representations to achieve state-of-the-art results. Chinese NER presents greater difficulty than English because Chinese sentences lack natural segmentation. A common approach involves first performing word segmentation with an existing Chinese word segmentation (CWS) system, then applying a word-based NER model. However, this pipeline method suffers from error propagation, as CWS errors inevitably affect NER performance. Consequently, some approaches employ character-based NER models directly [**?**, **?**]. The limitation of purely character-based models is that they fail to fully exploit word information. To incorporate word information in Chinese NER, recent methods such as [**?**, **?**, **?**, **?**, **?**] utilize automatically constructed lexicons.

2.2 Event Extraction

Events represent a common yet crucial knowledge type, making event identification and argument extraction important for many applications. DMCNN [**?**] is a classic EE model that uses convolutional neural networks to learn semantic features from raw text at both lexical and sentence levels. JRNN [**?**] employs recurrent neural networks to integrate discrete features with automatically learned features. JMEE [**?**] utilizes graph convolutional networks to jointly extract multiple event triggers and arguments by introducing syntactic shortcut arcs to enhance information flow and employing attention-based GCNs to model graph information. Recently, event extraction has been explicitly cast as a machine reading comprehension (MRC) problem [**?**], with MRC models used to solve the task.

2.3 NER and EE in Clinical Text

Clinical text information extraction has gained increasing importance in recent years. TREC pioneered shared tasks in clinical natural language processing, focusing on identifying relevant and irrelevant documents. Other evaluation tasks include ImageCLEFmed [**?**] and i2B2 [**?**]. For clinical NER, LSTM units combined with conditional random field classifiers [**?**] are used in NER components. Unsupervised methods [**?**] build clinical NER systems without manual annotation by training models on automatically annotated corpora followed by self-training iterations. For EE in clinical text, bidirectional long short-term memory networks assisted by attention mechanisms [**?**] help uncover important aspects of patients' medical conditions.

3. Task Definition

3.1 Clinical Named Entity Recognition

Given free text from EMRs, this task aims to identify clinical entity mentions and classify them into predefined categories. A novel method has been presented for training clinical NER systems without manual annotations, requiring only a raw text corpus and a resource like the Unified Medical Language System (UMLS) that provides lists of named entities with semantic types. Using these resources, annotations are automatically obtained to train machine learning methods, which were evaluated on the NER shared-task datasets of i2b2 2010 and SemEval 2014.

3.1.1 Formalized Definition We formally define this task as follows. **Input:** (1) A document collection from EMR: $D = \{d_1, \dots, d_i, \dots, d_N\}$, where $d_i = (w_{i1}, \dots, w_{in})$; (2) A set of predefined categories: $C = \{c_1, \dots, c_m\}$. **Output:** Collections of entity mention-category pairs: $\{(m_1, c_{m1}), \dots, (m_i, c_{mi}), \dots, (m_p, c_{mp})\}$.

Here, $m_i = (d_i, b_i, e_i)$ represents the entity mention in document d_i, where b_i and e_i denote the start and end positions of m_i, respectively, and $c_{mi} \in C$ represents the category of m_i. Overlap between mentions is not allowed, requiring $e_i < b_{i+1}$.

3.1.2 Pre-defined Categories Six categories are defined as follows: (1) Disease and diagnose (Dis); (2) Imaging examination (ImgExam); (3) Laboratory examination (LabExam); (4) Operation; (5) Drug; (6) Anatomy.

3.2 Clinical Event Extraction

3.2.1 Formalized Definition This task is formally defined as follows. **Input:** (1) Event entity; (2) A document collection from EMR: $D = \{d_1, \dots, d_N\}$, where $d_i = (w_{i1}, \dots, w_{in})$; (3) A set of predefined attributes: $P = \{p_1, p_2, \dots, p_m\}$. **Output:** Collections of attribute entities: $\{[d_i, (p_j, (s_1, s_2, \dots, s_k))]\}$, where $1 \leq i \leq N$ and $1 \leq j \leq m$.

Here, s_k is the entity of attribute p_j from document d_i. There may be zero or multiple entities for each attribute.

3.2.2 Pre-defined Attributes The three predefined attributes are: (1) Tumor Primary Site; (2) Tumor Size; (3) Tumor Metastatic Site.

4. Datasets

The datasets were provided by Yiducloud Beijing Technology Co., Ltd., which organized a professional medical team for annotation. These datasets are restricted to CCKS evaluation only. Compared with the CNER task in CCKS

2019, the annotated dataset is approximately four times larger. Additionally, Yiducloud provided an entity vocabulary and substantial unannotated data as supplementary resources for participants. The statistics of the CNER dataset are shown in Table 1 .

The clinical event extraction dataset includes a labeled training set, an unlabeled set, and a vocabulary, making this challenge more reflective of real-world scenarios. The statistics of the clinical event extraction datasets are shown in Table 2 .

Table 1. The statistics of clinical named entity recognition data set.

Train	Unlabeled Train	Unlabeled	ImgExam	LabExam	Operation	Anatomy	Total

Table 2. The statistics of clinical event extraction data set.

TumorPrimarySite	TumorSize	TumorMetastaticSite	Total

5. Evaluation Metrics

5.1.1 Strict Metric

Two evaluation metrics are employed: strict metric and relaxed metric. The extracted entities set is denoted as S and the gold entities set as G.

For the strict metric, $s_i \in S$ equals $g_j \in G$ when they are exactly the same: (1) The start position of s_i equals that of g_j; (2) The end position of s_i equals that of g_j; (3) The category of s_i equals that of g_j. The strict Precision, Recall, and F1 can be calculated accordingly.

5.1.2 Relaxed Metrics

The relaxed metric does not require $s_i \in S$ and $g_j \in G$ to be exactly identical; they need only meet the following requirements: (1) The maximum value of the start positions of s_i and g_j is less than or equal to the minimum value of the end positions of s_i and g_j; (2) The category of s_i equals that of g_j. The relaxed Precision, Recall, and F1 can be calculated accordingly.

5.2 Clinical Event Extraction

There may be multiple attribute entities for a single event attribute. Precision, Recall, and F1 are calculated based on attribute entities rather than attributes.

6. Results and Discussion

This evaluation attracted 354 teams, with 46 successfully submitting results. Thirty-two teams submitted results and five evaluation papers for the clinical named entity recognition task, while fourteen teams submitted results and three papers for the clinical event extraction task. The top teams are listed in Table 3 and Table 4 , respectively.

Table 3. The results of clinical named entity recognition.

Affiliation	Score
CASIA&Unisound AI Technology Co., Ltd.	
Medical AI Lab, Tencent Holdings Ltd.	
Lantone	
HFUT&SCUT	
SZU_{IC}	
mAI@pumc	

Table 4. The results of clinical event extraction.

Affiliation	Score
Knowledge Graph Group, Baidu, Inc.	
Medical AI Lab, Tencent Holdings Ltd.	
araloak	
zhjohnchan	
cecbrain	

6.1 Clinical Named Entity Recognition

For the clinical named entity recognition task, the Top 1 and Top 2 teams achieved very close scores, with both focusing on the label inconsistency problem in CNER.

The Top 1 team, from the Institute of Automation, Chinese Academy of Sciences (CASIA) and Unisound AI Technology Co., Ltd., proposed a hybrid system comprising a semi-supervised noisy label learning model based on adversarial training and a rule-based post-processing module. They adopted a five-fold cross-voting mechanism to handle annotation inconsistency in the dataset, used model ensemble and semi-supervised training to address insufficient training data, and applied adversarial training to reduce both aleatoric and epistemic uncertainty.

Based on submitted papers, we draw the following conclusions: (1) Pre-trained language models (PLMs) like BERT or ELMO [**?**] are widely applied, with their use becoming common sense among participants. Most teams did not simply

apply the general BERT model but utilized models pre-trained on Chinese documents, such as RoBERTa-wwm [**?**]. Some teams even collected Chinese medical documents and pre-trained PLMs on these in-domain sources. PLM usage in this year' s evaluation was more diverse than in previous CCKS competitions. (2) Model ensemble was employed by most teams, with ensembled models typically outperforming single models. Two-stage and k-folder ensemble methods proved effective. (3) Feature engineering and rules remain valuable. In the clinical domain, many regular patterns exist while annotated data remains limited, enabling participants to benefit from feature engineering. Some teams incorporated Chinese word features into their models and achieved stable improvements, while also introducing rules to mitigate data noise. (4) Semi-supervised methods were utilized by some participants, who generated pseudo-labels for the 1,000 unlabeled examples using supervised models and trained final models on both supervised and pseudo-labeled data. (5) Adversarial training was employed by some teams to train robust models insensitive to unavoidable label noise in the training data by adding turbulence to word embeddings during training. (6) Domain vocabulary has traditionally been important for CNER, with participants in past evaluations collecting and extending clinical vocabularies from sources like ICD-10, DrugBank, and health websites such as "haodf.com" and "xywy.com" . However, this year' s top three teams did not apply any vocabularies, primarily due to their extensive use of PLMs, suggesting a trend toward replacing vocabularies with PLMs.

6.2 Clinical Event Extraction

For the clinical event extraction task, the Top 1 team achieved an F1 score of 0.76234 and the Top 2 team achieved 0.74597, making the competition fierce.

The Top 1 team, from the Knowledge Graph Group at Baidu, Inc., proposed a system primarily based on pre-trained language models. They applied domain adaptation and task adaptation during pre-training to improve modeling capability. To address insufficient training data, they used back-translation to expand the training set and incorporated entity vocabulary as model input.

Based on submitted papers, we draw the following conclusions: (1) Pre-trained language models are widely used, with the Top 2 teams applying PLMs in their models and both selecting RoBERTa [**?**] as their backbone, confirming the usefulness of PLMs. (2) Data augmentation is crucial due to the difficulty of annotating clinical documents and resulting limited labeled data. Top teams employed various augmentation methods, including randomly rearranging sentence order in each training instance to double the training set, applying back-translation (translating training instances to English and back to Chinese), and randomly replacing key information across entire fields. (3) Feature engineering and rules were utilized by all top teams for pre-processing and post-processing, despite deep learning models being central to their systems. Pre-processing focused on obtaining cleaner data and cutting documents to appropriate lengths, while post-processing rules filtered meaningless results from model outputs.

7. Conclusion

This paper presents a detailed introduction to CCKS 2020 Task 3 for clinical named entity recognition and clinical event extraction from Chinese EMRs. Participants achieved exciting performances, particularly in the CNER task, with more varied models than in previous years' evaluations. Participants approached the evaluation problems from different perspectives. Through this evaluation, we hope to attract more researchers to focus on semantic analysis of Chinese EMRs and encourage greater industry attention to the industrialization of Chinese EMRs.

Author Contributions

All authors contributed equally to this work. X. Li (lixia2012@139.com) summarized the evaluation task and drafted the paper. Q.H. Wen (wtsinghua1@163.com) reviewed method documents submitted by participating teams and conducted code running tests to ensure correct and fair results. H. Lin (laohu20021210@163.com) summarized the results and discussion sections. Z.T. Jiao (zengtao.jiao@yiducloud.cn) was responsible for producing datasets and labeling results by medical experts. J.T. Zhang (zhang-jt13@tsinghua.org.cn) organized this evaluation task, designing and releasing the shared task. All authors made meaningful and valuable contributions to revising and proofreading the manuscript.

Data Availability Statement

The datasets generated and/or analyzed during this study are not publicly available because they were produced by medical expert consultants of Yidu Cloud based on their own experience. Public release requires consent from all expert consultants, but they are available from the corresponding author upon reasonable request.

References

[?] Lample, G., et al.: Neural architectures for named entity recognition. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 260–270 (2016)

[?] Chiu, J.P.C., Nichols, E.: Named entity recognition with bidirectional LSTM-CNNs. In: Transactions of the Association for Computational Linguistics, pp. 357–370 (2016)

[?] Ma, X.Z., Hovy, E.: End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pp. 1064–1074 (2016)

[?] Devlin, J., et al.: BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)

[?] Uzuner, Ö., et al.: 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. Journal of the American Medical Informatics Association 18(5), 552–556 (2011)

[?] Suominen, H., et al.: Overview of the ShARe/CLEF eHealth evaluation lab 2013. In: The Fourth CLEF Conference, pp. 212–231 (2013)

[?] Pradhan, S., et al.: SemEval-2014 task 7: Analysis of clinical text. In: Proceedings of the 8th International Workshop on Semantic Evaluation, SemEval@COLING 2014, pp. 54–62 (2014)

[?] He, J.Z., Wang, H.F.: Chinese named entity recognition and word segmentation based on character. In: Proceedings of the Sixth SIGHAN Workshop on Chinese Language Processing, pp. 128–132 (2008)

[?] Liu, Z.X., Zhu, C.H., Zhao, T.J.: Chinese named entity recognition with a sequence labeling approach: Based on characters, or based on words? In: International Conference on Intelligent Computing, pp. 634–640 (2010)

[?] Zhang, Y., Yang, J.: Chinese NER using lattice LSTM. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, pp. 1554–1564 (2018)

[?] Ding, R.X., et al.: A neural multi-digraph model for Chinese NER with gazetteers. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 1462–1467 (2019)

[?] Liu, W., et al.: An encoding strategy-based word-character LSTM for Chinese NER. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 2379–2389 (2019)

[?] Sui, D.B., et al.: Leverage lexical knowledge for Chinese named entity recognition via collaborative graph network. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 3830–3840 (2019)

[?] Xue, M.G., et al.: Porous lattice transformer encoder for Chinese NER. In: Proceedings of the 28th International Conference on Computational Linguistics, pp. 3831–3841 (2020)

[?] Chen, Y.B., et al.: Event extraction via dynamic multi-pooling convolutional neural networks. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 167–176 (2015)

[?] Nguyen, T.H., Cho, K., Grishman, R.: Joint event extraction via recurrent neural networks. In: Proceedings of the 2016 Conference of the North American

Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 300–309 (2016)

[?] Liu, X., Luo, Z.C., Huang, H.Y.: Jointly multiple event extraction via attention-based graph information aggregation. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 1247–1256 (2018)

[?] Liu, J., et al.: Event extraction as machine reading comprehension. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1641–1651 (2020)

[?] Kalpathy-Cramer, J., et al.: Evaluating performance of biomedical image retrieval systems–an overview of the medical image retrieval task at ImageCLEF 2004–2014. Computerized Medical Imaging and Graphics 39, 55–61 (2015)

[?] Magge, A., Scotch, M., Gonzalez-Hernandez, G.: Clinical NER and relation extraction using Bi-Char-LSTMs and random forest classifiers. In: The International Workshop on Medication and Adverse Drug Event Detection, pp. 25–30 (2018)

[?] Ghiasvand, O., Kate, R.J.: Learning for clinical named entity recognition without manual annotation. Informatics in Medicine Unlocked 13, 122–127 (2018)

[?] Yadav, S., et al.: Exploring disorder-aware attention for clinical event extraction. ACM Transactions on Multimedia Computing, Communications, and Applications 16(1s), Article 31 (2020)

[?] Peters, M., et al.: Deep contextualized word representations. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pp. 2227–2237 (2018)

[?] Cui, Y.M., et al.: Revisiting pre-trained models for Chinese natural language processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pp. 657–668 (2020)

[?] Liu, Y.H., et al.: RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692 (2019)

Author Biography

Xia Li is currently the head of the Clinical Pharmacy Department at the 305th Hospital of the Chinese People' s Liberation Army. She received her Bachelor' s degree from the Second Military Medical University in 2002. Her recent research interests center on clinical knowledge graphs and intelligent decision support for medication. ORCID: 0000-0002-8420-1226

Qinghua Wen is currently a graduate student in the Department of Computer Science and Technology at Tsinghua University. His research interests include

knowledge engineering, relation extraction, and data mining. ORCID: 0000-0002-4116-2140

Hu Lin is currently the head of the Department of Medical Administration at the 305th Hospital of the Chinese People' s Liberation Army. His research interests focus on health management, intelligent follow-up systems, and medical knowledge graphs. ORCID: 0000-0003-1525-5922

Zengtao Jiao is currently the director of the AI lab at Yidu Cloud Technology Co., Ltd. His research interests focus on key challenges in medical artificial intelligence, including medical text information extraction, disease prediction models, and medical knowledge mining. ORCID: 0000-0002-3534-479X

Jiangtao Zhang received his PhD in Computer Science from Tsinghua University in 2018. He currently serves as director of the Information Center at the 305th Hospital of the Chinese People' s Liberation Army. His research interests include knowledge graphs, data mining, and natural language processing in the medical domain. He organized and released a series of shared evaluation tasks for clinical knowledge discovery at CCKS 2017, CCKS 2018, CCKS 2019, and CCKS 2020. ORCID: 0000-0001-8462-3915

Note: Figure translations are in progress. See original paper for figures.