Current Science
CS 5 (4), 4075-4083 (2025)
Received 15 May 2025 | **Revised 13 June** 2025 | **Accepted 15 July**. 2025 | **Online Available 22 August** 2025
https://doi.org/ 10.52845/CS/2025-5-4-58

Current Science

Current Science

OPEN ACCESS

ISSN(O)2795-8639

Original Article

Ribert: Leveraging Rule Information for Chemical Production Named Entity Recognition via Pretrained BERT

Jingyang Ge[1], Jianbin Wang[2,3,4¶], Qingtao Chang[2,3,4], Yali Chen[2,3,4], Xun Wang[1], Hongxun Shi[2.3.4*]

[1]College of Computer Science and Technology, China University of Petroleum (East China), Qingdao 266580, China

[2]SINOPEC Research Institute of Safety Engineering Company Ltd., Qingdao, Shandong 266000, China

[3]Key Laboratory of Risk Warning and Intelligent Management Technology for Hazardous Chemical Safety, Ministry of Emergency Management, Qingdao,Shandong 266000, China

[4]State Key Laboratory of Chemical Safety, Qingdao, Shandong 266000, China

[*]Corresponding Author: Hongxun Shi

Abstract:
Chemical Named Entity Recognition (NER) faces critical challenges in accurately identifying domain-specific entities such as complex compound nomenclature and safety protocols. To address the underutilization of structural rules in chemical literature, this study proposes RiBERT, a novel framework integrating rule-informed boundary definitions with a hybrid BERT-KNN architecture. We systematically define 15 entity types across 4 categories based on regulatory documents. The model synergizes BERT's contextual embeddings with KNN's non-parametric retrieval to dynamically adapt to sparse entities and lexical variations. Evaluations on specialized chemical and public maintenance corpora demonstrate F1 scores of 69.77% and 88.12%, outperforming state-of-the-art baselines by up to 3.91%. The results indicate that incorporating rule-based information significantly enhances NER performance for chemical data.

Key Words:Named Entity Recognition, Chemical Production, Rule Information, Pre-trained large models, BERT

1. Introduction

The annual output of chemical research exceeds 1.5 million publications [1], yet leveraging these discoveries encounters severe bottlenecks from exponential growth in unstructured data. This paradigm necessitates Automated Information Extraction (IE) systems, wherein Named Entity Recognition (NER) constitutes a core computational infrastructure for distilling actionable insights from chemical literature. Crucially for hazardous material management, NER provides the foundational mechanism to identify safety-critical entities (e.g., concentration thresholds, reactive intermediates), thereby enabling rigorous risk quantification and compliance with standards such as the Globally Harmonized System (GHS).

Chemical hazard NER exhibits domain-specific complexities that fundamentally differentiate it from general entity recognition. First, terminological heterogeneity stems from: (i) structural polymorphism in chemical nomenclature (e.g., formulaic "NaCl" vs. systematic "sodium chloride" [3]), and (ii) hierarchical nesting under IUPAC conventions [4]. Second, contextual polysemy arises when lexically common terms (e.g., "solution") acquire domain-dependent interpretations distinct from colloquial usage [5]. Third, safety-critical

granularity necessitates exact quantification of hazard-determinative attributes, particularly concentration thresholds distinguishing "5% HCl" (corrosive) from "dilute HCl" (irritant) [6]. Fourth, regulatory alignment mandates strict adherence to standardized lexicons prescribed by frameworks such as GHS [7]. Finally, long-tail distributions dominate real-world entity occurrences—where safety-critical categories (e.g., emergency protocols) exhibit extreme sparsity [8]. These characteristics collectively impede conventional NER architectures; notably, BERT's fixed parametric representations prove inadequate for dynamically evolving regulatory schemata and low-frequency entity patterns.

State-of-the-art hazardous chemical NER predominantly employs fine-tuned BERT architectures, where contextual embeddings drive downstream classifiers [9]. This paradigm capitalizes on two intrinsic strengths: (i) WordPiece tokenization robustly processes complex chemical lexicons [10], and (ii) bidirectional attention mechanisms model intricate semantic dependencies [11]. However, the long-tail entity distribution—exhibiting extreme sparsity in safety-critical categories like emergency procedures—demands hybrid methodologies beyond parametric fine-tuning [12]. To bridge this gap, RiBERT synergizes BERT with k-nearest neighbors (KNN) through three innovative mechanisms: (1) Dynamic Adaptation: KNN's non-parametric formulation complements BERT's static parameterization, enabling real-time adjustment to novel entity patterns without model retraining. (2) Lexical Invariance: Distance metrics in BERT's embedding space resolve lexical variations (e.g., "ethanol" vs. "C_2H_5OH") by clustering semantically equivalent entities. (3) Imbalance Mitigation: Minority-class entities (e.g., <5% occurrence) propagate classification influence via neighborhood propagation in the latent space. This hybrid framework explicitly addresses the domain's entity hierarchy: high-frequency categories (chemical compounds, containment protocols, environmental media) establish dense representation bases, while KNN amplifies discriminative power for sparse safety labels. Empirical optimization (k=5, cosine similarity) demonstrates superior recall-precision tradeoffs, particularly for low-prevalence entities, outperforming pure BERT baselines by 3.2-8.7% F1 in preliminary studies.

Materials and methods

1. Datasets and Boundary Definition

Referring to the MSDS chemical corpus and safety accident plan text of the "Shanghai Organic Institute Chemical Database", it is stipulated that the identification label categories will be classified into 4 categories, totaling 15 types, as shown in Table 1:

Table 1. NER boundary division

Boundaries		Entity	Begin-point	Inner	End-point
Entity boundaries	Chemicals	ID	B-num	I-num	O-num
		Name	B-entity	I-entity	O-entity
		SecondName	B-pronoun	I-pronoun	O-pronoun
		Code	B-ref	I-ref	O-ref
	Emergency methods	Target	B-entity	I-entity	O-entity
		UseCase	B-case	I-case	O-case
		Method	B-method	I-method	O-method
		Sequela	B-seq	I-seq	O-seq
	Management plan	Manage Target	B-manage	I-manage	O-manage
		Scenario	B-scen	I-scen	O-scen
		Plan	B-plan	I-plan	O-plan
Scene boundaries		Field	B-field	I-field	O-field
		Temperature	B-tem	I-tem	O-tem
		Humidity	B-hum	I-hum	O-hum
		Container	B-cont	I-cont	O-cont

2. Model Structure

Building upon Wang et al.'s kNN-NER framework [1], we redesigned the label prediction module to better integrate rule knowledge. As illustrated in Figure 1, RiBERT comprises four sequentially connected layers: (1) an embedding layer that maps each token into a dense vector via the domain-specific ChemBERT; (2) a KNN retrieval layer that queries a key-value memory, where keys are training-set token embeddings and values are their gold labels, to produce a soft label distribution based on RBF-weighted nearest neighbors; (3) a CRF decoding layer that models label dependencies and generates a globally optimal tag sequence; and (4) an output layer that fuses the CRF distribution with the KNN distribution through simple averaging (voting), yielding the final predicted tags.

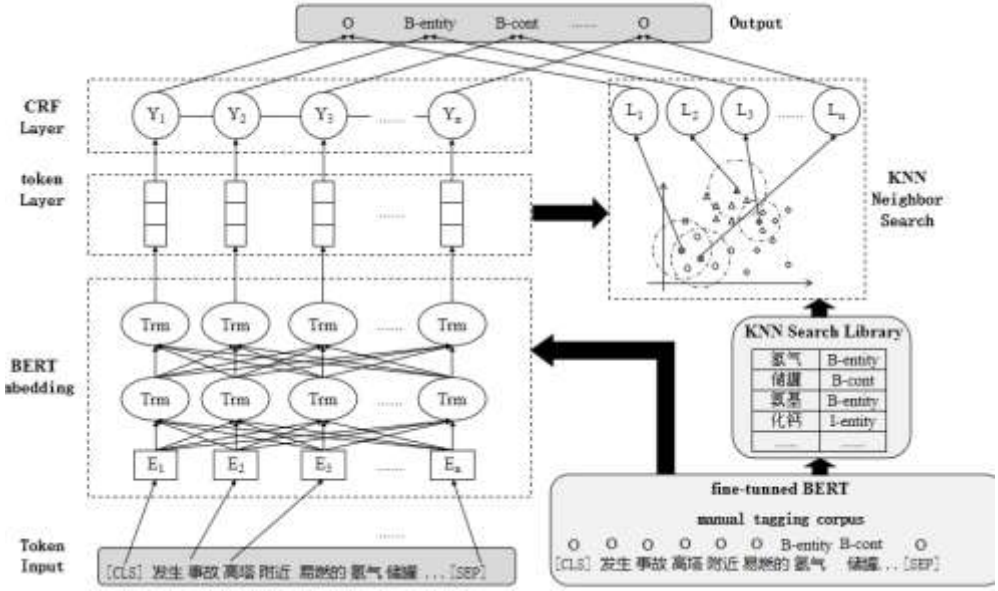

Figure 1. Overview structure of RiBERT

Embedding Layer

The BERT component encompasses two modules. Initially, the segmented sentences undergo encoding to derive contextual semantics at both the word and sentence levels. The input text is preprocessed to generate dynamic token vectors, from which a wealth of semantic features at both word and sentence levels are automatically extracted. Ultimately, the CRF layer constraints are employed to predict the dependency relationships between labels and derive the globally optimal label sequence.

We utilize words from sentence segmentation as input tokens for the model. The model is fine-tuned using chemical entry texts and labels from structured corpora, supplemented by manually annotated texts and labels from safety accident emergency plans. To obtain the contextual representation of each label in the input sentence, we first apply BERT and construct the input sequence by adding the special tag [CLS] at the beginning of the sequence and the inter-sentence tag [SEP], as follows:

$$X = \{[CLS], t_2, t_3, \dots, t_{n-1}, [SEP]\}$$

Where t denotes the words after sentence segmentation. Since we use words instead of characters as tokens, t also represents tokens. Subsequently, the sequence is input into the Chinese BERT for feature extraction to obtain semantic representations (i.e., embedding vectors), as shown below:

$$\mathrm{S} = \mathrm{BERT}(\mathrm{X} = \{\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_\mathrm{n}\})$$

Where n represents the number of tokens in the sequence, $S = \{s_1, s_2, \ldots, s_n\}$

KNN Retrieval Layer

As noted by Wang et al. [1], since we use segmented words as tokens for input to the BERT model, the embedding vectors of these tokens can be directly employed to construct the retrieval library D' for the KNN model. Initially, we match the labeled token embedding vectors S from the training set with their corresponding labels Y to form key-value pairs, which are then aggregated into the retrieval library as follows:

$$D' \stackrel{\text{def}}{=} \{\text{key}, \text{value}\} = \{(s_i, y_i) | \forall s_i \in S, \forall y_i \in Y\}$$

Here, key denotes the set of keys, and value represents the corresponding set of values. For a given text sequence X, token embedding vectors $S = \{s_1, s_2, \ldots, s_n\}$ are derived through BERT processing. Subsequently, k-nearest neighbor retrieval is performed against the tokens stored in D' using L_2 Euclidean distance as the metric. The resulting neighbor set N is converted into an output distribution via the RBF kernel method [x], formulated as:

$$P'(Y|X) = (y_i = e_j | X, x_i) \propto \sum_{(k,v) \in N} \mathbb{I}_{v=e_j} \exp\left(\frac{-d(s_i, k)}{T'}\right)$$

Where N represents the set of token tags obtained from k-nearest neighbor retrieval, and (k, v)represents the search key value pairs in N; e_j represents the jth entity in the label category table, and the probability of predicting the label as an entity e_j is proportional to the sum of the kernel outputs of all values equal to e_j in N; $\mathbb{I}$ is the indicator function, with a value of 1 when $v = e_j$, otherwise 0; T' is a scaling parameter used to smooth the distribution.

CRF Decoding Layer

Using a conditional random field CRF for label prediction task of embedding sentences, given input sentences $X = \{x_1, x_2, \ldots, x_n\}$, the probability of label sequence $Y = \{y_1, y_2, \ldots, y_n\}$ is:

$$P(Y|X) = \frac{\exp(\sum_i (W^{y_i} s_i + T(y_{i-1}, y_i)))}{\sum_{\tilde{y} \in D} \sum_i (W^{y_i} s_i + T(\tilde{y}_{i-1}, \tilde{y}_i))}$$

Where D represents all possible tag sequences input, W^{y_i} represents the emission weight of y_i for s_i, T represents the transfer matrix weight of two adjacent tags, and $\tilde{y}_i$ represents the true tag of the token. The negative logarithmic likelihood function is used as the loss function as follows:

$$\text{loss} = -\sum_{i=1}^{n} \log(P(y_i | x_i))$$

Output layer

Finally, merge the two predicted distributions as the final output, as shown below:

$$P_{\text{final}}(Y|X) = \lambda P(Y|X) + (1 - \lambda) P'(Y|X)$$

The result is the label with the highest score after decoding:

$$y_{\text{final}} = \arg_{y \in D} \max(P_{\text{final}}(y_i | x_i))$$

Results

Datasets

In our experiments, we utilized two public corpora and a professional dataset: the Traffic Accident Corpus, the Vehicle Maintenance Corpus, and the Chemical Production Dataset. Similar to many teams participating in the challenge, we employed the original training set and development set for training. We then randomly selected 10% of the training set as the validation set to fine-tune the hyperparameters. The chemical NER performance was evaluated using an F-score, which assigns equal importance to precision and recall (F1 score), on the test set.

Experimental settings

We employ the "BIO" (B-begin, I-inside, O-outside) labeling scheme to delineate the boundaries of entity mentions. The Adam optimization algorithm serves as the optimizer, with a learning rate set to 1e-5. We set the maximum sentence length to 175 for the two public corpora and 300 for the Chemical Production dataset, which encompasses over 99% of sentences. The hidden size of the BERT model is set to 512. The dropout rate in the BERT layer is set to 0.2. The number of epochs is set to 10, with a batch size of 35 for each epoch. The model, bert-base-chinese-accidentreason-classifier, pre-trained from scratch on the same corpus, is used to embed the sentences. We train the entity embeddings on the Chemical Production dataset using bert-base-chinese-accidentreason-classifier. All hyperparameters in bert-base-chinese-accidentreason-classifier for entity embedding training are set to their default values, except for the following: the number of epochs is set to 10, and the batch size is set to 35. All experiments are conducted on an RTX 3090.

Evaluation metrics

We use several metrics to evaluate model performance, including precision, recall and F1 score. Precision is the ratio of true positives in the identified positive samples which reflects the classification accuracy of the model for BioNEs tokens:

$$Precision = \frac{TP}{TP + FP}$$

Where TP indicates the number of positive classes predicted as positive classes, FP indicates the number of negative classes predicted as positive classes. Recall represents the proportion of all positive samples in the test set that are correctly identified as positive samples which reflects the ability of the model to distinguish BioNEs：

$$Recall = \frac{TP}{TP + FN}$$

Where FN indicates the number of positive classes predicted as negative classes. F1 score is one of the most commonly used metrics in classification and information retrieval, reflecting the average performance of model precision and recall:

$$F_1 = \frac{2 * Precision * Recall}{Precison + Recall}$$

Comparison with existing methods

We compare our model with existing methods on two datasets (Table 1). Initially, we tested the NB-CRF model without considering rule information. We then compare our method with state-of-the-art (SOTA) methods that utilize BERT as the embedding layer, including BERT-base-chinese and BERT-base-chinese-accidentreason-classifier, which have already been fine-tuned on the traffic accident corpus.

NB-CRF is a machine learning model that employs the naive Bayes method for named entity recognition (NER).

Layered-BiLSTM-CRF is a neural network model designed for nested NER.

MLP+Softmax treats the NER sequence annotation task as a multi-class classification problem. Each word's label is independently predicted based on contextually relevant representations, without taking neighboring label results into account.

MLP+CRF also treats the NER sequence annotation task as a multi-class classification problem. Each word's label is independently predicted based on contextually relevant representations, without considering neighboring label results.

Table 2. Comparison of different methods on the Chemical production vehicle maintenance datasets

Methods	Datasets					
	the Chemical production dataset			the vehicle maintenance corpus		
	Precision	Recall	F1	Precision	Recall	F1
NB-CRF	64.12	57.19	60.46	77.42	72.1	74.66
MLP+softmax	63.63	66.26	64.92	79.43	77.25	78.32
MLP-CRF	65.43	64.71	65.07	81.17	82.42	81.79
RNN-CRF	67.78	58.85	63	85.04	83.26	84.14
Transformer-CRF	67.24	59.89	63.35	84.13	84.95	84.54
layered-BiLSTM-CRF	68.01	70.76	69.36	85.05	85.38	85.21
RiBERT	69.04(±0.12)	70.52(±0.63)	69.77	88.19(±0.72)	88.05(±1.19)	88.12

RiBERT outperforms NB-CRF, layered-BiLSTM-CRF, MLP+Softmax, MLP-CRF, RNN-CRF, Transformer-CRF, BERT-base-chinese, and BERT-accidentreason on both the Chemical Production dataset and the Vehicle Maintenance corpus (Table 2). RiBERT achieves F1 scores of 69.77% on the Chemical Production dataset and 88.12% on the Vehicle Maintenance corpus. Compared to BERT-accidentreason, RiBERT achieves higher F1 scores by 0.7% and 0.42% on the Chemical Production dataset and the Vehicle Maintenance corpus, respectively. The improvements resulting from the rule information encoding and fusion layer are significant. Methods that can simultaneously utilize rule information and local dependency may be more effective. For instance, RiBERT performs better than layered-BiLSTM-CRF on the Vehicle Maintenance corpus, but sometimes worse on the Chemical Production dataset (layered-BiLSTM-CRF scores higher than RiBERT-min). This result indicates that leveraging local dependency from external knowledge for NER is complex. This may be attributed to the shorter text length of the Chemical Production dataset compared to that of the Vehicle Maintenance corpus, which reduces the impact of long-distance dependencies on token representation. As observed, RiBERT performs better on the Vehicle Maintenance corpus than on the Chemical Production dataset, similar to other models (Figure 2). This may be due to the dataset's scale and the length of the entities within them. RiBERT employs an effective attention mechanism to leverage local dependency for NER and also has the potential to be applied to other tasks.

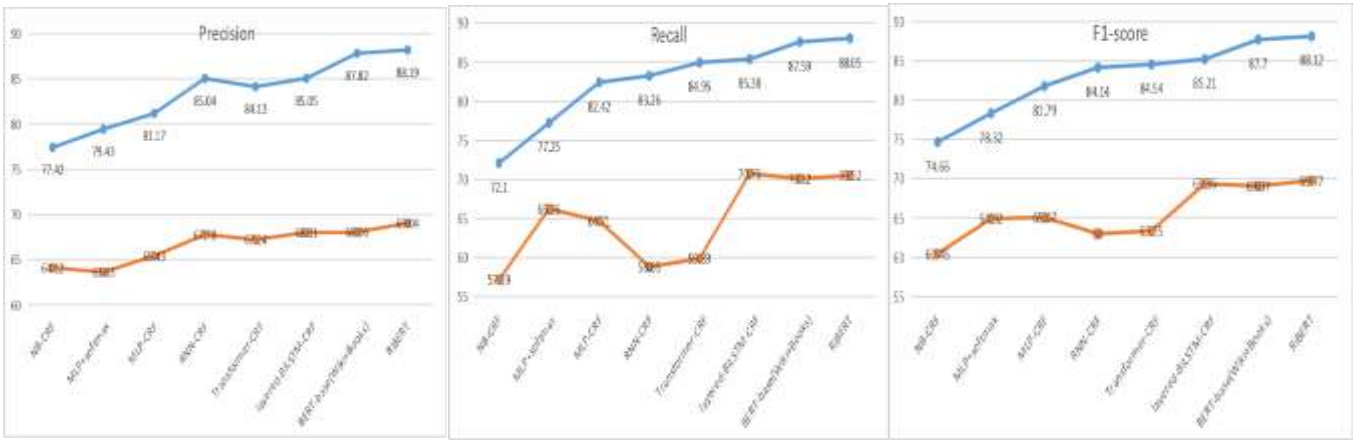

Figure 2 Three indicator results of RiBERT

Ablation Study

We conduct an ablation study on the two datasets to evaluate the performance of embedding mode and knowledge from rule information on the RiBERT. “BERT-base-chinese” and “BERT-accidentreason” are the models which replacing the BERT basic model, the models with “w/o KNN” are the models using BERT to embedding tokens without KNN, the models with “w/o CRF” are the models directly using BERT for prediction, the models with “w/o MLP” are the models replacing CRF with MLP(Table 3).

Table 3. Ablation experiments for different embeddings and network structures

Methods	the Chemical production dataset			the vehicle maintenance corpus		
	Precision	Recall	F1	Precision	Recall	F1
BERT-base-chinese	67.03	69.12	68.06	86.19	88.12	87.14
BERT-accidentreason	67.82	70.05	68.92	86.91	86.73	86.82
RiBERT w/o KNN	68.06	70.12	69.07	87.82	87.59	87.7
RiBERT w/o CRF	68.12	70.53	69.3	85.05	85.38	85.21
RiBERT w/o MLP	68.69	70.49	69.58	87.91	87.23	87.57

It is worth mentioning that the effect of the structure of RiBERT w/o MLP is compared. The results show that replacing the MLP module can still achieve better results, which may be related to our concern about the relationship from additional knowledge rather than context. Comparing the results of RiBERT with that of RiBERT w/o KNN, we can see both embedding mode and KNN can bring improvements. The performence that BERT-base-chinese and BERT-accidentreason gets significantly decreased scores on precision, recall and F1 shows that indicating that considering the type of entity can help to find out a better representation space for all words in articles so as to identify entities more accurately.

Discussion

Because the published literature far exceeds people's imagination, how to extract key chemical information from a large amount of text has become an urgent problem to be solved. Chemical Named Entity Recognition (NER) is the foundation of information extraction tasks in the field of chemistry. However, the current extraction of key information on chemicals is not effective, especially when complex information such as chemical names, response plans, and storage environments are included in the literature. However, these information categories are clear and related to the chemicals themselves, and these rules have not been fully utilized.

In this work, we investigated the impact of rule information on chemical NER tasks, defined boundaries for chemical names, response plans, etc., and combined KNN and BERT pre-trained large models to complete the extraction. There are two public corpora and a professional dataset were used: the Traffic accident corpus, the vehicle maintenance corpus and the Chemical production dataset. Like many teams in the challenge, the original training set and development set were used as the trainingThen we randomly selected 10% of the training set as the validation set toset.tune the hyper-parameters. The chemical NER performance was measured with an F-score which attributes equal importance to precision and recall (Fl score) on the test set.

And we selected 6 classic algorithms with the best performance to compare their effectiveness with our method. RiBERT performs better on the vehicle maintenance corpus than the Chemical production dataset as same as other models. At the same time, we also validated the rule information and algorithm structures of each part. It is worth mentioning that the effect of the structure of RiBERT w/o MLP is compared. The results show that replacing the MLP module can still achieve better results, which may be related to our concern about the relationship from additional knowledge rather than context. Comparing the results of RiBERT with that of RiBERT w/o KNN, we can see both embedding mode and KNN can bring improvements. The performence that BERT-base-chinese and BERT-accidentreason gets significantly decreased scores on precision, recall and F1 shows that indicating that considering the

type of entity can help to find out a better representation space for all words in articles so as to identify entities more accurately.

Data Availability Statement

The original contributions presented in the study are included in the article/github, further inquiries can be directed to the corresponding authors.

Author Contributions

JG conceived the study. JG, JW, QC, and YC contributed to the study design. Data analysis was curated by JG, XW and HS. The first draft of the manuscript was written by JG. JG critically revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by Open Project of Institute of Safety Engineering(HX20240202).

Reference

1. Bornmann, L., & Haunschild, R. (2022). Empirical analysis of recent temporal dynamics of research fields: Annual publications in chemistry and related areas as an example. Journal of Informetrics, 16(2), 101253.
2. Wang, Shuhe, et al. (2022) kNN-NER: Named Entity Recognition with Nearest Neighbor Search." arxiv preprint arxiv:2203.17103.
3. Xuezhe Ma and Eduard Hovy. (2016) End-to-end sequence labeling via bi-directional LSTM-CNNsCRF. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1064–1074, Berlin, Germany. Association for Computational Linguistics.
4. Jie Yang and Yue Zhang. (2018) Ncrf++: An opensource neural sequence labeling toolkit. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics.
5. Alan Akbik, Duncan Blythe, and Roland Vollgraf. 2018. Contextual string embeddings for sequence labeling. In Proceedings of the 27th International Conference on Computational Linguistics, pages 1638–1649.
6. 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), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
7. Li L, Jiang Y. (2017) Biomedical named entity recognition based on the two channels and sentence-level reading control conditioned lstm-crf. In: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 380–5.
8. Ju M, Miwa M, Ananiadou S. (2018) A neural layered model for nested named entity recognition. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), 1446–59.
9. Hemati W, Mehler A. (2019) LSTMVoter: chemical named entity recognition using a conglomerate of sequence labeling tools. J Cheminformatics 2019; 11(1): 1–7.
10. Zhu Q, Li X, Conesa A, et al. (2018) GRAM-CNN: a deep learning approach with local context for named entity recognition in biomedical text. Bioinformatics 34(9):1547–54.
11. Korvigo I, Holmatov M, Zaikovskii A, et al. (2018) Putting hands to rest: efficient deep CNN-RNN architecture for chemical named entity recognition with no hand-crafted rules. J Cheminformatics 10(1): 1–10.
12. B. Nie, et al. (2021) Knowledge-aware Named Entity Recognition with Alleviating Heterogeneity. in Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 13595–13603.
13. Z Wang, Y Qu, L Chen, et al. (2018) Label-aware double transfer learning for cross-specialty medical named entity recognition. arXiv preprint arXiv:1804.09021.
14. Yansen Su, Chunlong Liu, Yunyun Niu et al. (2021) A community structure enhancement based community detection algorithm for complex networks. IEEE Transactions on Systems, Man and Cybernetics: Systems, 51 (5): 2833-2846.
15. Ye Tian, Xiaochun Su, Yansen Su* et al.

(2020) EMODMI:A Multi-Objective Optimization Based Method to Identify Disease Modules. IEEE Transactions on Emerging Topics in Computational Intelligence, 5(4): 570-582.
16. Yansen Su, Sen Li, Chunhou Zheng et al. (2020) A heuristic algorithm for identifying molecular signatures in cancer, IEEE Transactions on NanoBioscience, 19(1): 132-141.
17. Li Lei, Gao Zhen, Wang Yu-Tian et al. (2021) SCMFMDA: Predicting microRNA-disease associations based on similarity constrained matrix factorization. PLOS COMPUTAT IONAL BIOLOGY. 17(7):e1009165.
18. Wu Qingwen, Cao Rui-Fen, Xia Junfeng et al. (2021) Extra Trees Method for Predicting LncRNA-Disease Association Based on Multi-layer Graph Embedding Aggregation. IEEE/ACM transactions on computational biology and bioinformatics. DOI10.1109/ TC BB.2021.3113122.
19. Wang Haiyun, Zhao Jianping, Su Yansen et al. (2021) scCDG: A Method based on DAE and GCN for scRNA-seq data Analysis. IEEE/ ACM transactions on computational biology and bioinformatics. DOI10.1109/TCBB.2021 .3126641.
20. Cai Lijun, Lu Changcheng, Xu Junlin et al. (2022) Drug repositioning based on the heterogeneous information fusion graph convolutional network. Briefings in Bioinformatics. 22(6): bbab319.
21. B. Tang, et al. (2019) De-identification of Clinical Text via Bi-LSTM-CRF with Neural Language Models. in AMIA 2019, American Medical Informatics Association Annual Symposium, Washington, DC, USA, November 16-20, 2019.
22. Q. Zhang, et al. (2018) Multitask Learning for Chinese Named Entity Recognition. in Advances in Multimedia Information Processing - PCM 2018 - 19th Pacific-Rim Conference on Multimedia, Hefei, China, September 21-22, 2018, Proceedings, Part II vol. 11165 653–662.
23. Kocaman V, Talby D. Biomedical named entity recognition at scale[C]//International Conference on Pattern Recognition. Springer, Cham, 2021: 635-646.