ISSN 2616-5775 Vol. 7, Issue 9: 32-40, DOI: 10.25236/AJCIS.2024.070905

Recognition of News Named Entity Based on Multi-Level Semantic Features Fusion and Keyword Dictionary

Yi Ren[1,a,*], Yu Liu[1,b]

[1]*The Department of Computer Science and Engineering, Shenyang Jianzhu University, Shenyang, Liaoning, China*
[a]*smdoctor@126.com*, [b]*ss10242048@163.com*
[*]*Corresponding author*

Abstract: *Most of the current methods used existing models for named entity recognition tasks, but this could only obtain character vectors and could not solve the problem of polysemy. This study proposed a new model based on multi-level semantic features fusion and dictionary of keyword to solve this problem. This method first uses a keyword dictionary for random entity replacement to achieve data augmentation; then, it utilizes the pre-trained BERT model to transfer prior knowledge to this task to obtain multi-level semantic features; Secondly, in order to obtain more comprehensive sequence information, the vector is input into the multi-semantic feature fusion layer to extract global information; Finally, after correcting the results with the CRF, the output is obtained. Compared with traditional models such as BiLSTM-CRF and BERT-CRF, this model has achieved good results on news domain datasets, with an accuracy rate of 94.95% and an F1 value of 94.99%.*

Keywords: *News Domain, Named Entity Recognition, Multi-level Semantic Features Fusion, Keyword Data Augmentation*

1. Introduction

In most research, names of people, places, and organizational entities are meaningful entities that need to be identified, and this is also true in the field of journalism. From a methodological perspective, there are three main approaches to named entity recognition: statistical machine learning methods, rule and dictionary-based methods, and deep neural network methods. The field of journalistic knowledge is a complex and information-dense mixed domain, encompassing many sub-fields such as politics, economics, humanities, and other types of information. Each sub-field has its specific context and keywords, and the focus of the reports varies, which poses certain difficulties for named entity recognition in this field. Furthermore, as the carrier of news content, digital technology is used to extract key entity information from textual reports and to achieve the transformation from sequential text to concise entity information. This not only facilitates the work of journalists in organizing and categorizing the information but also allows readers to quickly access the content they need.

In the field of named entity recognition, researchers both domestically and internationally have primarily gone through two distinct phases. In the early stage, researchers predominantly employed machine learning methods. Common approaches included the Maximum Entropy Model (MaxEnt), Support Vector Machines (SVMs), and Conditional Random Fields (CRFs). In recent years, with the rapid development of deep learning, neural network models have become very common in named entity recognition research. For instance, pre-training with the BERT model has been introduced to obtain word embeddings; the attention mechanism combined with BiLSTM has been used to enhance the contextual semantic connections. In recent years, with the rapid development of deep learning, neural network models have become very common in named entity recognition research. For instance, pre-training with the BERT model has been introduced to obtain word embeddings; the attention mechanism combined with BiLSTM has been used to enhance contextual semantic connections.

To address the issue of entity recognition within multiple sub-fields in the news domain, this study proposes an entity recognition model based on multi-level semantic feature fusion and keyword enhancement. The model identifies three types of entities in data texts: personal names, locations, and organizational structures. Firstly, the data is enhanced through a keyword dictionary to expand the dataset; secondly, prior knowledge is transferred to the current task using the BERT pre-training model to obtain transfer features; then, a layer for capturing multi-level semantic features is used to derive semantically rich feature vectors; finally, the optimal sequence labeling results are inferred through a conditional random field.

2. Related Works

2.1. Named Entity Recognition

In recent years, the application of deep learning techniques in the field of natural language processing has proven its feasibility and effectiveness in named entity recognition (NER) tasks. Researchers such as Roald Eiselen [1] have developed NER systems and feature sets for South African languages based on named entity annotation resources. Nada Boudjellal et al. [2] proposed a BERT-based model to identify disease and treatment-related named entities in Arabic textual data, outperforming AraBERT and multilingual BERT, confirming the effectiveness of BERT models in enhancing the understanding of Arabic biomedical texts. Xunwei Yin [3], Xiaoyong Tang [4] have suggested using BERT or its variants for word embedding, followed by connecting BiLSTM and CRF to complete Chinese NER tasks, achieving results on the MASR dataset and CLUENER2020 dataset that are superior to other traditional models. Wang et al. [5] used the BERT-BiLSTM-CRF model for named entity recognition of multiple entities such as personal information, educational background, and research directions of scientific researchers, better realizing the understanding and mastery of various aspects of researchers. Hongchao Jiang et al. [6] also employed the BERT-BiLSTM-CRF model for entity recognition and relationship extraction when constructing knowledge graph question answering methods, which, compared to traditional questioning methods, can achieve higher accuracy in entity and relationship identification. These studies reveal that using character-based feature models for NER tasks is now mainstream. Although research results show that, compared to token-based methods, character-based NER models are somewhat lacking in performance for person entity recognition, directly applying such models, like BERT-BiLSTM-CRF, to richly detailed domains like news, which inherently encompasses multiple subfields, often leads to the loss of important information due to the absence of token-level features. This results in an inability to accurately recognize entity boundaries, leading to recognition errors. Therefore, recent research focuses on how to introduce lexical information into named entity recognition without losing character information, thereby improving the overall recognition accuracy of the model.

Zhang et al. [7] incorporated a Lattice structure into the foundation of LSTM, which integrates potential phrase information ending with the current character into the cell state of the current character. Subsequently, a weighted fusion of single-character vectors and phrase vectors is performed to obtain a cell state that includes lexical information. This structure makes full use of Chinese sequence segmentation information and significantly aids named entity recognition. However, because it can only capture segmentation information ending with the current character, information acquisition is insufficient. Additionally, due to the model structure, parallel processing is not possible, preventing the full utilization of GPU resources. In contrast, Jiang Yang et al. [8], when addressing the task of traditional Chinese medicine named entity recognition, proposed a simpler and more direct method: integrating dictionary features into the character-based BiLSTM-CRF model. This approach does not require changing the model structure but involves incorporating lexical information during the embedding process, fully utilizing the text's lexical information. Inspired by this, our research has designed a method based on this foundation that can integrate keyword information. By utilizing phrases from the keyword dictionary to replace similar entities in the dataset, we have enhanced the information of entity phrases to a certain extent. Compared to named entity recognition methods that do not perform this step, our method can obtain boundary information of keyword phrases to a certain extent, improving the effect of entity recognition without affecting its efficiency.

2.2. Multi-head Attention Mechanism

Attention mechanisms are techniques that simulate human visual attention, allowing models to focus on different parts of the input data when processing information. They provide a flexible way for models to capture both local and global information within the input data. By learning to assign different weights to each input element, attention mechanisms enable models to concentrate more on key information when dealing with complex tasks. To address the issue of entity extraction in device configuration recognition, Yang Yang et al. [9] combined attention mechanisms with active learning and introduced an adaptive weighting method based on Laplacian mixture distributions. This enhanced the multi-head self-attention of the Transformer model, enabling it to dynamically adapt to words at different positions and further improving the accuracy of the proposed model. To enhance performance and reduce forward computation, Shengmin Cui et al. [10] proposed a pyramid hierarchical model based on multi-head adjacent attention when dealing with nested named entity recognition (NER) and conducted experiments on multiple nested NER datasets, proving the effectiveness of the model. Kaihong Zheng et al. [11] constructed a new NER model called Att-CNN-BiGRU-CRF, capable of identifying entities in the power domain. The multi-head attention optimized the processing of sentence-level information. Experimental results showed that the model's recall rate was 88.16% and precision was 89.33%, outperforming current mainstream models. Tong Liu et al. [12] utilized three attention networks, two of which were used for word embedding and character embedding respectively, while the other was used to fuse the results of the first two attention networks. This achieved multi-granularity word fusion for Chinese named entity recognition and yielded results on multiple public datasets that were superior to most baseline models, demonstrating good performance.

3. Method

This study proposes a named entity recognition model based on multi-level semantic features fusion and keyword data augmentation. Firstly, rich embedding features are obtained through the pre-training model. Then, the multi-level semantic features layer is used to obtain the contextual semantic information and key local information of the sequence. Finally, the decoding layer outputs the category of entity labels, realizing the recognition of entities. The following will introduce from several aspects: model structure, construction of keyword dictionary and data augmentation, multi-level semantic features fusion, and multi-head attention mechanism.

3.1. Model Structure

The model first adopts a pre-trained BERT to transfer domain prior knowledge to the current task, better capturing the semantic information in the text sequence. Subsequently, perturbation information is added to each embedded feature, which is then fed into the multi-semantic features layer. The positional information of the sequence is initially reinforced, followed by dynamically extracting local features from the sequence, assigning greater weight to more effective features for named entity recognition. Finally, the decoding layer predicts the optimal tagging sequence through observing the sequence. The overall structure is shown in Figure 1.

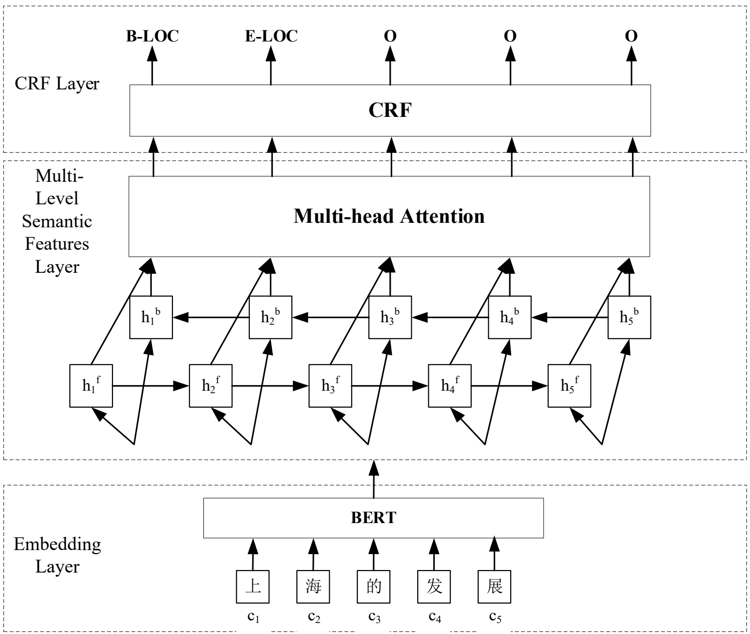

Figure 1: Recognition of the news named entities model based on multi-level semantic features fusion and dictionary of keyword

ISSN 2616-5775 Vol. 7, Issue 9: 32-40, DOI: 10.25236/AJCIS.2024.070905

3.2. Keyword Dictionary

3.2.1. The Basic Vocabulary Set Constructed

Through research, it has been found that the development of the news field has long since transitioned from paper content to digitization. Therefore, in order to construct a more comprehensive keyword dictionary, this study selects words related to the entities studied from publicly available online dictionaries in the Chinese news field, including Tencent's online dictionary and Sogou's news dictionary. These words are added as a basic vocabulary set to the keyword dictionary constructed for the news domain.

3.2.2. The Keyword Dictionary Constructed

The corpora selected for dictionary construction inevitably contain a large number of irrelevant low-frequency words, which can interfere with the construction of domain-specific dictionaries. To address this issue, this study employs the TF-IDF (Term Frequency-Inverse Document Frequency) algorithm to extract keywords from these corpora [13], after preprocessing the corpus collection C, which is represented as $\{c_{ij}\}(i \in [1,2,3,\ldots,n], j \in [1,2,3,\ldots,m])$, where n is the number of valid phrases or word groups after preprocessing, and m is the number of articles in the preprocessed corpus collection. The candidate word set Z extracted by the TF-IDF algorithm is represented as $\{z_i\}(i \in [1,2,3,\ldots,k])$, where k is the number of candidate words in the selected word set.

The extraction steps are specifically as in formulas (1)-(3):

Firstly, the term frequency TF_{ij} of a phrase c_{ij} is calculated:

$$I' = I + \delta \tag{1}$$

m_{ij} is the number of times the phrase appears in the article C_j.

Secondly, the inverse document frequency IDF_{ij} of the phrase c_{ij} is computed:

$$IDF_{ij} = \log \frac{\left|\sum_j C_j\right|}{1 + \left\{ j : c_i \in C_j \right\}} \tag{2}$$

C_j represents the j^{th} article in the corpus, and c_i is the i^{th} word within the article.

Lastly, the $TF - IDF_{ij}$ value of the phrase c_{ij} is determined:

$$TF - IDF_{ij} = TF_{ij} \times IDF_{ij} \tag{3}$$

Another point to note is that the keyword lexicon composed of candidate keywords contains noisy words. Using a lexicon containing noisy words to construct a dictionary will affect the effect of keyword information acquisition. Therefore, two steps are adopted here to reduce the impact of noise. Firstly, the weight of each word is calculated by the TF-IDF algorithm. When the candidate words extracted from the keyword lexicon do not match any high-quality phrases in the domain, these words are considered to be noisy and irrelevant information words and need to be excluded from the keyword dictionary; secondly, some valueless words are filtered out according to the stop word list, and it is stipulated that if a word group contains any element in the stop word list, then the word group also needs to be filtered out. For example, "China is" contains the stop word "is", so this phrase is deleted from the keyword dictionary.

3.2.3. Dictionary Information Expansion

The field of news is highly sensitive to timeliness, necessitating the inclusion of up-to-date terminology in domain-specific dictionaries. Relying on a dictionary that contains only a base set of terms can lead to suboptimal recognition performance due to the absence of timely and relevant new words. To address this issue, this study employs a domain dictionary expansion method using the People's Daily picture and text database as a foundation. The method utilizes Word2vec to facilitate the

ISSN 2616-5775 Vol. 7, Issue 9: 32-40, DOI: 10.25236/AJCIS.2024.070905

discovery of new terms within the domain dictionary.

Inspired by the fact that cosine similarity between word embeddings, which encapsulate the semantic meaning of text, can measure the similarity between two words [14], we can use cosine similarity to calculate the most similar words to each candidate keyword. From these words, we can select new terms to expand our keyword set. Word vectors generated by Word2vec carry a certain degree of contextual semantics, making it reasonable and feasible to use cosine distance to calculate the similarity between the expanded candidate new words and the existing keywords in the set. The cosine similarity between n-dimensional candidate word $a\left(c_{11}, c_{12}, \cdots, c_{1n}\right)$ and n-dimensional domain word $b\left(x_{11}, x_{12}, \cdots, x_{1n}\right)$ is represented as in formula (4):

$$\cos\theta = \frac{\sum_{k=1}^{n} c_{1k} x_{1k}}{\sqrt{\sum_{k=1}^{n} c_{1k}^2}\sqrt{\sum_{k=1}^{n} x_{1k}^2}} \tag{4}$$

By calculating the cosine value between the phrases added to the keyword dictionary and the new words of the expanded terms, we can obtain their similarity with the existing words. Phrases that exceed a set threshold are selected as candidate words. If these words are not already included in the existing keywords, they are considered to meet the screening criteria and can be added as new words to the keyword dictionary.

3.3. Data Augmentation

This study utilizes phrases from a keyword dictionary to randomly replace certain non-stop words in sentences, ensuring that the replacement words are of the same type as the words they replace. For example, replacing "Shenzhen" with "Shanghai," as both words are of the location type. Similarly, names of people and organizations in the original sentences are replaced with corresponding types of words from the keyword dictionary, resulting in a dataset enhanced by the keyword dictionary.

3.4. Multi-Level Semantic Features Fusion

In the process of further features extraction, we consider that the BiLSTM model can effectively obtain the contextual semantic information of the sequence. Meanwhile, the multi-head attention mechanism not only focuses on the local information within each independent space but also takes into account the global information of the sequence after splicing the local information. Therefore, we integrate BiLSTM with the multi-head attention mechanism to construct a multi-semantic feature fusion layer, which can extract features information more deeply and fuse the semantic features extracted from multiple parts. Since the BiLSTM model has a bidirectional network structure and gating mechanism, it can extract global semantic features. Compared with traditional RNNs, BiLSTM has solved the problem of long-term dependencies to a certain extent, allowing for better learning and storage of long-distance contextual information.

In the multi-head attention mechanism, the input sequence is divided into multiple subspaces, each containing a specific part of the original information. Within each subspace, each attention head calculates independently. Thus, each attention head focuses on different parts of the input sequence, concentrating on capturing the local information of that part. Finally, the outputs of all attention heads are concatenated to form a single vector containing all captured information, which captures both local information and combines it with global information.

3.5. Multi-head Attention Mechanism

The attention mechanism typically operates on input consisting of query vectors Q, key vectors K, and value vectors V, with the following computational formula (5):

$$Attention(Q, K, V) = soft\max(\frac{K^T Q}{\sqrt{d_k}})V \tag{5}$$

$K^T Q$ is the attention matrix and $\sqrt{d_k}$ is the adjustment factor.

When the query vector, key vector, and value vector are all the same, it constitutes a self-attention mechanism. The multi-head attention mechanism can be regarded as a multi-task extension of the self-attention mechanism, allowing the model to perform multiple self-attention operations in parallel within the same layer, and then integrate the results of these operations. The primary function of the multi-head attention mechanism is to capture diverse features information through different attention heads, thereby enhancing the model's expressive capability. The core of the multi-head attention mechanism involves performing (h) self-attention computations on the original input sequence, then concatenating the output results from each attention head, and passing them into a linear layer for linear transformation to obtain the final output. The computational formulas are shown in equations (6), (7), and (8):

$$MultiHead(Q,K,V) = Concat(head_1,\ldots,head_h)W^O \tag{6}$$

$$head_i = Attention(Q_i,K_i,V_i), i=1,\ldots,h \tag{7}$$

$$Q_i = QW_i^Q, K_i = KW_i^K, V_i = VW_i^V, i=1,\ldots,h \tag{8}$$

4. Results and Discussion

Experiments were conducted on the People's Daily dataset using the model proposed in this study, and comparisons were made with other mainstream models.

4.1. Experimental Data

The People's Daily Dataset is one of the commonly used datasets for Chinese named entity recognition, with data sourced from China's largest official newspaper, People's Daily. This dataset consists of 45,518 entries of annotated Chinese named entity recognition data, including three categories: personal names, locations, and organizational institutions. In this study, the dataset was further expanded through a method of keyword enhancement, ultimately resulting in 51,865 entries of data.

4.2. Data Processing

The People's Daily dataset is a public dataset that has already been labeled with sequences using the BMESO (Begin Middle End Single Other) tagging scheme. In this scheme, 'B' stands for the first character of the named entity fragment, 'M' for the middle characters, 'E' for the last character, 'S' for single-character entities, and 'O' for non-named entity characters. The three types of entity category labels are shown in Table 1.

Table 1: People's Daily entity labels

entity	the first character	the middle characters	the last character	the single characters
names	B-PER	M-PER	E-PER	S-PER
locations	B-LOC	M-LOC	E-LOC	S-LOC
organizations	B-ORG	M-ORG	E-ORG	S-ORG

The dataset is divided into training, validation, and test sets, with specific allocations as shown in Table 2.

Table 2: Dataset division condition

dataset	train set	validation set	test set
The People's Daily	36,305	5,186	10,374

4.3. Parameter Settings

The experimental parameter settings are shown in Table 3.

ISSN 2616-5775 Vol. 7, Issue 9: 32-40, DOI: 10.25236/AJCIS.2024.070905

Table 3: Model parameter setting

parameter	setting
vector dimension	768
dropout	0.5
lr	0.0005
the optimization algorithm	Adam
the number of LSTM layers	2

4.4. Experimental Environment and Evaluation Metrics

The experimental environment and configuration used in this study are shown in Table 4.

Table 4: Software and hardware environments

items	setting
system	Linux 5.15
gpu	Tesla P100
memory	16GB
Python	3.10
Pytorch	2.12

This study uses Precision, Recall, and F1 score as evaluation metrics to measure the model's recognition performance, calculated as shown in equations (9) to (11).

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{11}$$

Here, TP represents the number of correctly predicted entities; FP represents the number of incorrectly predicted entities; FN represents the number of entities not predicted.

4.5. Experimental Results and Analysis

4.5.1. Comparative Experiment Analysis

To verify the recognition effectiveness of the proposed model in the news corpus domain, comparative experiments were conducted with mainstream models such as BiLSTM-CRF, BERT-CRF, and BERT-BiLSTM-CRF. The results are shown in Table 5.

Table 5: Comparison of experiment results

models	precision	recall	f1
BiLSTM-CRF	87.89%	83.43%	85.60%
BERT-CRF	94.46%	93.90%	94.17%
BERT-BiLSTM- CRF	94.01%	94.85%	94.43%
ours	94.95%	95.03%	94.99%

On the People's Daily dataset, the proposed model improved Precision and F1 score by 7.06% and 9.39%, respectively, compared to BiLSTM-CRF. This improvement can be attributed to the fact that Word2Vec embedding simply maps characters to vectors in space, resulting in static word vectors that cannot resolve polysemy issues. For example, it fails to recognize "Ping An" (the name of an insurance company) as an entity. The term "Ping An" has multiple meanings and can signify safety or the China Ping An Insurance Group depending on the context. By contrast, dynamic word vectors obtained using BERT contain 12 layers of Transformer encoders, which progressively capture semantic features, thus correctly identifying "Ping An" as an organizational entity and effectively resolving polysemy issues.

Moreover, compared to the BiLSTM-CRF model, the BiLSTM-CRF model that uses BERT to obtain word embedding vectors can fully extract monosyllabic features, lexical features, and semantic features through the BERT pre-training model. This allows for better identification of polysemous entities in the news domain, enhancing the model's accuracy. From the evaluation metrics, the news domain entity recognition model proposed in this study, which is based on multi-level features and keyword enhancement, significantly outperforms the other three mainstream models, with the highest recognition accuracy reaching 94.95%, a recall rate of 95.03%, and an F1 score of 94.99%.

4.5.2. Ablation Experiment Analysis

To verify the effectiveness of the multi-level semantic features module and keyword enhancement module in the named entity recognition model presented in this study, an ablation experiment is designed in this section. The "-KW" indicates that compared to our model, the keyword data enhancement is removed; "-BiLSTM" indicates that compared to our model, the bidirectional long short-term memory network is removed; "-ATT" indicates that compared to our model, the multi-head attention mechanism is removed. The experimental results are shown in Table 6.

Table 6: Comparison of ablation experiment results

models	precision	recall	f1
ours	94.95%	95.03%	94.99%
-KW	94.01%	94.85%	94.43%
-BiLSTM & -ATT	94.51%	94.16%	94.33%

1) After the elimination of key words enhancement in this study's model, there was a significant decline in three indicators. The experiment shows that the introduction of BERT pre-training model can fully obtain prior knowledge, making the information contained in word embedding vectors more comprehensive. At the same time, using key words enhancement to randomly replace entities in the dataset expands the training dataset, enabling the model to learn more comprehensive semantic information. This can improve the model's recognition effect to a certain extent, proving that the introduction of key word information in the field of named entity recognition helps the model to better improve the accuracy of recognition.

2) After the elimination of BiLSTM and multi-head attention mechanism in this study's model, all three indicators decreased, indicating that without the BiLSTM part, the model could not fully obtain contextual semantic information. At the same time, the lack of multi-head attention mechanism also reduced the model's ability to acquire global and local information of the sequence to a certain extent. Therefore, in terms of overall effect and performance, the introduction of multi-level features is helpful for improving the overall performance of the model.

5. Conclusion

For the named entity recognition task in the field of news, this study proposes an entity recognition model based on multi-level semantic features fusion and keyword enhancement. By constructing a keyword dictionary, random replacement of entities of the same type is carried out to expand the original dataset; then, the BERT pre-training model is used to transfer existing knowledge to this task, integrating character features, syntactic features, and semantic features, which not only solves the problem of polysemy but also enriches the semantic information; subsequently, BiLSTM and Multi-head attention are used to achieve multi-level semantic features fusion, enhancing the model's ability to acquire global and local information. Experimental verification shows that the model has good recognition results on the selected dataset. The focus of the next step is to improve the training efficiency of the model, because the BiLSTM model, due to its structural characteristics, cannot achieve parallel computing, resulting in the inability to fully utilize the computational power of GPUs. In the future, it may be considered to replace BiLSTM with convolutional neural networks. Additionally, besides multi-level semantic features, font features and phonetic features may also help the model acquire semantic information, and further research will be conducted from these two perspectives in the future.

References

[1] Roald Eiselen, Andiswa Bukula. IsiXhosa Named Entity Recognition Resources. ACM Transactions on Asian and Low-Resource Language Information Processing, Volume 22, Issue 2, pp. 1-19, June 2022.
[2] Nada Boudjellal, Huaping Zhang, Asif Khan, et al. ABioNER: A BERT-Based Model for Arabic

ISSN 2616-5775 Vol. 7, Issue 9: 32-40, DOI: 10.25236/AJCIS.2024.070905

Biomedical Named-Entity Recognition. Complexity, Volume 2021, Issue 3, pp. 1-6, March 2021.

[3] Xunwei Yin, Shuang Zheng, Quanmin Wang. Fine-Grained Chinese Named Entity Recognition Based on RoBERTa-WWM-BiLSTM-CRF Model. Conference: 2021 6th International Conference on Image, Vision and Computing (ICIVC), pp. 408-413, July 2021.

[4] Xiaoyong Tang, Yong Huang, Meng Xia, et al. A Multi-Task BERT-BiLSTM-AM-CRF Strategy for Chinese Named Entity Recognition. Neural Processing Letters, Volume 55, Issue. 3, pp. 1209-1229, July 2022.

[5] Wang Jun, Wang Xiulai, Luan Weixian, et al. Research on named entity recognition of scientific research talents field based on BERT model. Computer Technology and Development, Volume 31, Issue 11, pp. 21-27, 2021.

[6] Hongchao Jiang; Baoqi Yang; Li Jin, et al. A BERT-Bi-LSTM-Based Knowledge Graph Question Answering Method, 2021 International Conference on Communications, Information System and Computer Engineering , CISCE, pp. 308-312, May 2021.

[7] Zhang Yue, Yang Jie. Chinese NER Using Lattice LSTM. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics Volume 1, pp. 1554-1564, 2018.

[8] Jiang Yang, Hongman Wang, Yuting Tang, et al. Incorporating lexicon and character glyph and morphological features into BiLSTM-CRF for Chinese medical NER. 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), pp. 12-17, January 2021.

[9] Yang Yang; Zhenying Qu; Zefan Yan, et al. Network Configuration Entity Extraction Method Based on Transformer with Multi-Head Attention Mechanism. Computers, Materials & Continua, Volume 78, Issue 1, pp. 735-757, January 2024.

[10] Shengmin Cui; Inwhee Joe. A multi-head adjacent attention-based pyramid layered model for nested named entity recognition. Neural Computing and Applications. Volume 35, Issue 4, pp. 2561-2574, September 2022.

[11] Kaihong Zheng; Lingyun Sun; Xin Wang, et al. Named Entity Recognition in Electric Power Metering Domain Based on Attention Mechanism. IEEE Access, Volume 9, pp. 152564-152573, January 2021.

[12] Tong Liu; Jian Gao a; Weijian Ni, et al. A Multi-Granularity Word Fusion Method for Chinese NER. Applied Sciences. Volume 13, Issue 5, pp. 2789, February 2023.

[13] Park Hosik, Lee Minsu, Hwang S, et al. TF-IDF based association rule analysis system for medical data. KIPS Transactions on Software and Data Engineering, Volume 5, Issue 3, pp. 145-154, 2016.

[14] Gao Yan, Liu Chenchen, Zhao Liangyuan, et al. Multi-attribute group decision-making method based on time-series q-rung orthopair fuzzy sets. Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, Volume 41, Issue 1, pp. 2161-2170, 2021.