Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 6507719, 11 pages
https://doi.org/10.1155/2022/6507719

Research Article

Named Entity Recognition of Ancient Poems Based on Albert-BiLSTM-MHA-CRF Model

Faguo Zhou, Chao Wang, and Jipeng Wang

School of Mechanical Electronic & Information Engineering, China University of Mining & Technology, Beijing, China

Correspondence should be addressed to Faguo Zhou; zhoufaguo@cumtb.edu.cn

Received 25 August 2021; Revised 24 March 2022; Accepted 6 April 2022; Published 21 April 2022

Academic Editor: Jiliang Zhang

The extraction and construction of the knowledge graph related to the entity of ancient poems are helpful to excavate the connection between ancient poems, and it is of great significance to inherit the traditional Chinese culture. This paper proposes an Albert-BiLSTM-MHA-CRF model for entity extraction in ancient poems. Based on the BiLSTM-CRF model, the author introduces the Albert pretraining model and the multihead self-attention mechanism to extract character vectors and enhance the generalization ability of word embedding vectors and the potential semantics between characters of the model, depending on weight and other feature extraction capabilities. The experiment is carried out in the corpus of ancient poetry, and the model is compared with Bert-BiLSTM-CRF, BiLSTM-CRF, and CRF model. The results show that the entity extraction effect of ancient poetry is significantly improved, and the harmonic average value is 97.17%. Compared with Bert model as the pretraining model, Albert model reduces the time by 19.56%.

1. Introduction

Poem as an important part of ancient Chinese traditional culture, has a long history. To study and carry forward the traditional culture in ancient China, ancient Chinese poem is of great significance to the construction of beautiful China. The main emotion in ancient poetry is often related to the imagination in ancient poetry. Poets often use these imaginative artistic methods to express their emotions and inner feelings. These imagination are often the entities of poetry, such as "兰" (orchid), "芭蕉" (banana), and "梧桐" (sycamore). The author crawled 1983 poems from the ancient poetry website, including the patriotic poems, spring, summer, autumn, winter, and other categories of poems. Extracting entities from it, analyze the most important entities in various poems, and displaying them in the form of knowledge graph [1] is of great significance for the study of ancient poetry. In the past, the research on ancient poems was mainly the artificial research on ancient books without the assistance of computer, while the knowledge graph and other technologies provided new ideas for the study of ancient poems [2–4].

At present, in the field of Chinese named entity recognition (CNER), there are still a lot of studies on ancient classics, such as the research on some pre-Qin classics. But most of them are still named entity recognition in the field of modern Chinese, such as medicine [5] and military [6] field. The difference between the field of ancient poetry and other fields of named entity recognition mainly shows the difference between ancient Chinese and modern Chinese. As an ancient Chinese, ancient poetry has both similarities and differences compared with modern Chinese. The similarities are manifested in the same sentence components, which all have six components: adverbial, attribute, subject, predicate, complement, and object, and their relative positions are basically the same. The differences are embodied in grammar and sentence patterns. The language of ancient poetry is often short, subject-verb-object structure is bound, and words change greatly. For example, in the poem "为赋新词强说愁," it is expressed in modern Chinese as "to describe the sad artistic conception in order to write a new poem." It can be seen that there are many modifiers in modern Chinese, which makes the model easy to extract entities. In this paper, on the premise of establishing its own corpus, the

pretraining model in NLP is used to construct the knowledge graph based on ancient poems, which deepens the computer-aided research on ancient poems.

The main results of this paper are as follows: (1) to construct the entity recognition model of a variety of ancient poems based on the lite version of BERT (Albert) pretraining model. (2) Albert, bidirectional long short-term memory, multihead self-attention model, and conditional random field (Albert-BiLSTM-MHA-CRF), the best model for entity extraction of ancient poems, was selected to improve the effect of entity recognition of ancient poems. (3) A knowledge graph of ancient poetry was created according to the dynasty, author, title information, entity, etc.

2. Related Work

Named entity recognition, as a basic task of natural language processing (NLP), was proposed on MUC-6 [7] and is a basic technology of question answering system [8, 9], information extraction [10], and knowledge graph. The task is also related to relation 57 extraction [11–13] and event extraction [14–16]. Its purpose is to extract the names of people, place names, time, and other entities in specific fields. Entities in ancient poetry, such as imagination and time entities, are of great significance for the study of traditional ancient poetry and are also conducive to constructing the knowledge graph related to ancient poetry. Because ancient poetry is different from modern Chinese and is a style of ancient Chinese, which is similar to classical Chinese, at present, there is less research on ancient poetry, and most of them focus on the recognition of named entities in modern Chinese related fields.

Named entity recognition generally has three main research methods: firstly, rule-based method; secondly, statistical machine learning-based methods; and thirdly, deep learning-based methods. In the rule-based named entity extraction, it is time-consuming to build rules, knowledge base and dictionary manually, and the effect is poor. Hidden Markov model (HMM) and CRF model are widely used among the statistical machine learning models. For example, Y. Zhang [17] et al. used CRF to extract place names in Tang poems, and the harmonic average value reached 82.33%. However, statistical machine learning needs to design feature templates to extract features artificially, and its effect depends on artificially designed features to achieve good results. Compared with statistical machine learning, deep learning can automatically extract features from the data. In addition, deep learning has achieved the best state-of-the-art (SOTA) record in a series of downstream tasks of NLP. Some classical deep learning models, such as recurrent neural network (RNN) and long short-term memory (LSTM), have been used for named entity recognition [18–20]. For example, Limsopatham and Collier [18] used the BiLSTM model for named entity recognition in Twitter messages. After 2018, as bidirectional encoder representations from transformers (Bert), Taher et al. proposed a series of pretraining models and introduced them into various tasks in NLP, and named entity recognition based on the pretraining model has been widely applied [21–23]. For example, Zhang [24] et al. proposed the Bert-BiLSTM-CRF model. Based on BiLSTM-CRF, the Bert pretraining model was introduced, which was superior to other models in extraction effect of Chinese medicine. Moreover, Lv [25] et al. used Albert to improve recognition effect in Chinese named entity recognition for less training time and better effect. Therefore, this paper proposed a kind of ancient poetry entity recognition based on pretraining model and further improved the effect of ancient poetry entity extraction by using the latest NLP pretraining model Albert.

3. Methodology

1In order to explore the effect of Albert model on named entity recognition in classical poetry texts, the research ideas of this paper mainly include the following: firstly, corpus collection: collect corpus of ancient poetry through multisource channels such as websites and books; secondly, corpus construction: preprocess the collected classical poetry corpus, analyze and label named entities, and construct experimental corpus; and thirdly, named entity recognition experiment: CRF [26] model, BiLSTM-CRF [27] model, Bert-BiLSTM-CRF [28] model, and Albert-BiLSTM-MHA-CRF model were used to carry out named entity recognition experiment of ancient poetry texts, to test and compare the performance of various models in accuracy and recall rate and $F1$ value. The most 100 suitable named entity recognition model for poem text was obtained by analyzing the 101 experimental results, and the model was applied in the test set for verification.

3.1. Construction of Ancient Poetry Dataset. At present, there are few studies on ancient poetry and there is no relevant dataset. Therefore, this paper crawls 1983 ancient poems, and these poems from relevant ancient poetry websites, then, build a dataset related to ancient poems. The specific construction steps are as follows:

Firstly, 80% of the crawling ancient poetry are used as the training set, 10% of the training set is used as the validation set, and 20% of the other corpus is used as the testing set and YEDDA [29] is used for labeling, as shown in Figure 1. Label the four types of entities involved in the ancient poetry dataset. The labeling method uses "[@" and "∗]" to represent the left and right boundaries of the entity, and "Time," "Scene," "Person," "Location" represent the entity classes, for example, "[@雨#scene∗]" and "[@洞庭#location∗]" in Figure 1, where "雨" is the scene entity and "洞庭" is the location entity. In the process of labeling, if an ambiguous entity is encountered, it will be documented, and the final labeling conclusion will be determined through multiperson discussion. According to the above rules, the annotated dataset is processed into sequence annotation form in python. Each character and corresponding label are on one line, and there is a blank line at the end of each sentence. Finally, the ancient poetry dataset constructed in this paper is obtained.

Second is the establishment of entity category, because the poet has always studied poetry in terms of time, place name, person name, and imagination. Therefore, four basic entity types are identified in this study, as shown in Table 1.

Thirdly is the labeling system: the labeling system used in the experiment is the BIO labeling system, where "B" represents the initial position of the entity and "I" represent other

(a) Before labeling (b) After labeling

Figure 1: Poetry corpus labeling by YEDDA.

Table 1: The entity of ancient poetry.

Entity category	Entity instance	Entity symbol
Imagination	春雨, 春风 (spring rain, spring breeze)	Scene
Person name	赵飞燕, 相如, 司马相如 (Zhao Feiyan, Xiangru, Sima Xiangru)	Person
Place name	中国, 东海 (China, East China Sea)	Location
Time	春, 夏, 秋, 冬 (spring, summer, autumn, winter)	Time

positions of the entity except the initial position of the entity. "O" means not an entity location. Annotation examples are shown in Table 2.

3.2. Albert-BiLSTM-MHA-CRF Model. This model is constructed by Albert, BiLSTM, and CRF models. First, the sum of the word embedding, position embedding, and segment embedding of input characters is taken as the input vector of Albert. The vector output from Albert is input into BiLSTM model to encode and learn the features of the text. Then, the mined features, the hidden state (h_t) at time t, were used as the output to decode and predict the rational relationship between tags and output the optimal tag sequence. The model structure was shown in Figure 2.

The first layer of the model is the Albert layer. When input to the Albert layer of the model, according to the Vocab file in Albert, the input characters are vectorized to represent the poems, and the poems are converted into data that can be processed by the computer. Then, the poetry represented by the vector should be output by Albert training and recorded as a sequence $X = (x_1, x_2, x_3, \cdots, x_n)$. Compared with word vectors such as Word2vec and global vectors for word representation (Glove), the character vectors generated by Albert can effectively generate different character vectors according to the context, effectively solving the problem of polysemy.

The second layer of the model is the BiLSTM layer. The word vector X generated by Albert is used as input to the BiLSTM layer to obtain the forward hidden layer state $\overleftarrow{h}_t$ and the reverse hidden layer state $\overrightarrow{h}_t$. The resulting hidden layer state is $\overleftarrow{h}_t$ and $\overrightarrow{h}_t$. It is stitched together according to its position and denoted as $h_t = (h_1, h_2, h_3, \cdots, h_n)$.

Table 2: Entity annotation of ancient poetry.

Word	Tagging	Word	Tagging
庆	B-time	子	I-person
历	I-time	京	I-person
四	B-time	谪	O
年	I-time	守	O
春	B-time	巴	B-location
,	O	陵	I-location
滕	B-person	郡	I-location

Chinese: 庆历四年春, 滕子京谪守巴陵郡. English: In the spring of the fourth year of Qingli (A.D.1044), Teng Zijing was demoted to Yuezhou prefecture chief.

The third layer of the model is the multihead self-attention mechanism. The sequence Y generated by the BiLSTM layer is input to the MHA model. Through three different mapping operations, transform to matrix queries Q, key value K, and value V with both dimensions of d_k, respectively. Then, do H times of parallel self-attention between sequence Y to get head_i, and continue to integrate all semantic information of the head and define it as MultiHead. Secondly, MultiHead is mapped to the s dimension (s is the number of entity classes), and the sequence after mapping is $Y = (y_1, y_2, y_3, \cdots, y_n)$; $y_{i,j}$ are the scores of x_i, corresponding to each entity type t_j. The fourth layer of the model is the CRF layer, which can consider

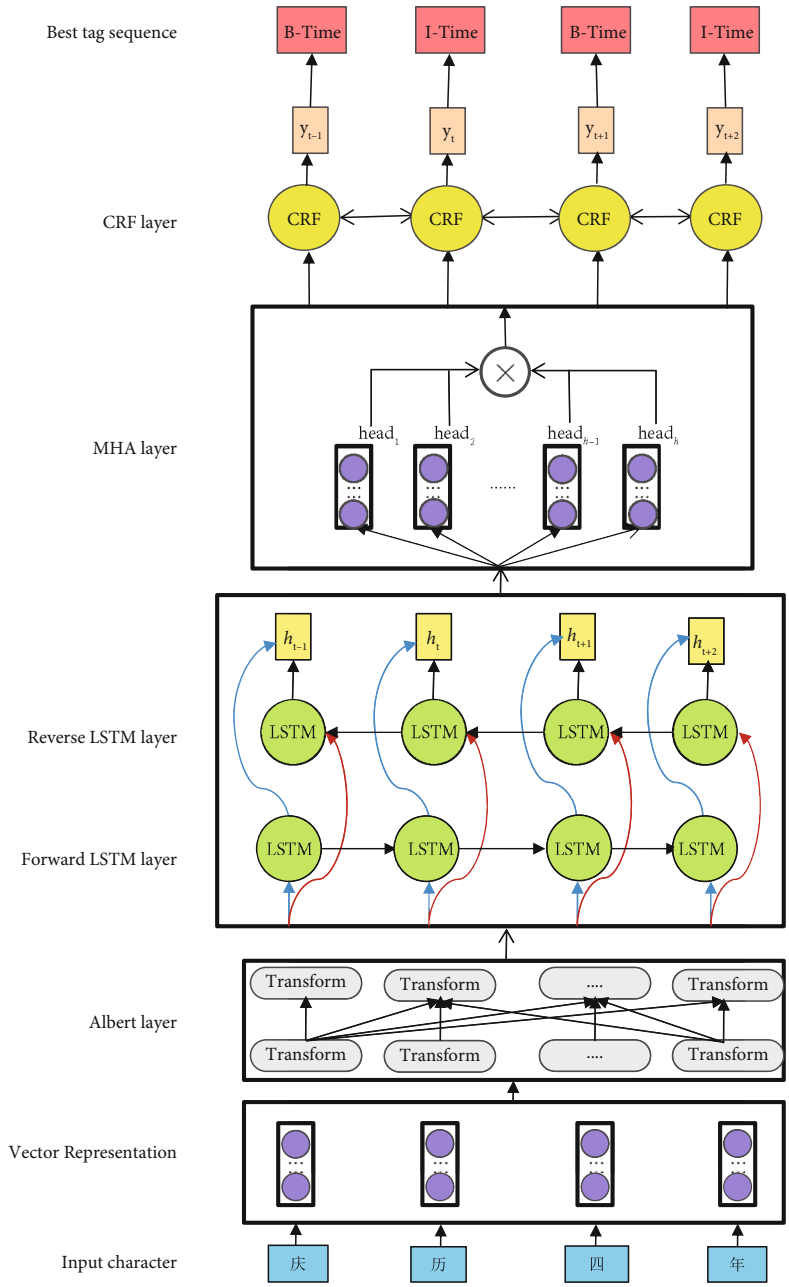

Figure 2: Overall architecture of Albert-BiLSTM-MHA-CRF model.

the order of output labels according to the transfer characteristics, so it is the last layer.

3.3. CRF Model. Although the long- and short-term memory neural network and the multihead self-attention mechanism can learn contextual labels, to output the label with the highest probability, they did not take into account the dependencies between labels, which may cause the same label together. This is unlikely to happen in reality; however, the conditional random field model can consider the order of the tags. Therefore, the CRF layer is selected as the final output layer. The commonly used first-order chain structure CRF is shown in Figure 3.

For character sequence $(x_1, x_2, x_3 \cdots \cdots x_n)$, the prediction label sequence $(y_1, y_2, y_3 \cdots \cdots \cdots y_n)$ can be obtained by using the linear link conditional random field. Its predicted score is

$$s(x, y) = \sum_{i=0}^{n} A_{y_i, y_{i+1}} + \sum_{i=1}^{n} P_{i, y_i}. \quad (1)$$

P_{i,y_i} is the i position, the probability that the output is y_i, and $A_{y_i,y_{i+1}}$ is the probability of transferring from y_i to y_{i+1}. The best predicted tag sequence can be obtained by using viterbi algorithm:

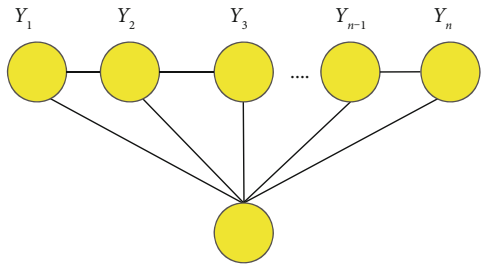

Figure 3: Conditional random field first-order chain structure.

$$y^* = \arg\max\left(s(x, y)\right). \tag{2}$$

Viterbi algorithm can obtain the maximum state path by dynamic programming algorithm.

3.4. MHA Model. Although the BiLSTM model can obtain the current context information, as the length of the sentence continues to increase, BiLSTM model will also lose some important information. However, the MHA model can fully capture the characteristics of long distance and obtain global information. Acquisition of various features from character, word, and sentence level can improve the effect of entity recognition. The matrix output from BiLSTM gets Q, K, and V ($Q = K = V$) of self-attention through cubic matrix mapping. By doing attention on Q and K, the calculation of attention is as shown in

$$\text{attention}(Q, K, V) = \text{soft}\ \max\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \tag{3}$$

By doing h self-attention operations in parallel, d_k is the dimension of Q, K and V. Every time the attention function operates, head_i is obtained; finally, the $H\text{head}_i$ are spliced to obtain $\text{Multihead}(Q, K, V)$, and the specific calculation formula is shown in

$$\text{head}_i = \left(QW_i^Q, KW_i^K, VW_i^V\right), \tag{4}$$

$$\text{Multihead}(Q, K, V) = \text{concat}(\text{head}_1, \text{head}_2, \cdots, \text{head}_h)W^O. \tag{5}$$

Thereunto, W_i^Q, W_i^K, W_i^V is the matrix of linear transformation of Q, K and V, respectively, and W^O is also the parameter matrix to be used.

3.5. LSTM Model. Long short-term memory neural network is the most popular recurrent neural network. Compared with the general recurrent neural network, there are three more gate states: input gate, output gate, and forget gate. Input gate and output gate control the input and output of the unit, and the forget gate controls whether to save the previous unit state to the current unit status. The calculation formula is shown in

$$\text{f}_\text{t} = \sigma\left(\text{w}_{fh} * h_{t-1} + w_{fx} * x_t + b_f\right). \tag{6}$$

The input gate determines whether the current input is saved to the state of the unit. The calculation formula is shown in

$$i_\text{t} = \sigma(\text{w}_i * [h_{t-1}, x_t] + b_i). \tag{7}$$

Output gate and unit state determine the output of LSTM. The calculation formula is shown in

$$o_\text{t} = \sigma(\text{w}_o * [h_{t-1}, x_{t+1}] + b_c). \tag{8}$$

LSTM automatically extracts the features from the character vectors output in Bert and then uses the tags predicted in the context at the CRF layer to get the optimal sequence.

3.6. Albert Pretraining Model. Albert model is derived from the encoder of transformers and is considered as a lightweight Bert with few parameters and has been optimized in two aspects:

First is the pretraining tasks: the two pretraining tasks of Bert are MLM (masked languages model) and NSP (next sentence predication), both of which have certain defects. MLM uses the Cbow and Skip Gram methods in Word2vec to mask the token in sentence. The MLM task selects 15% of the tokens, replaces those words with masks, and then predicts those tokens. However, in order to prevent overfitting, the general model will choose to dropout. Once these masks are ignored, the information will be lost. But MLM is already hard to fit, so Albert does not do dropout. Deleting dropout can also reduce the number of arguments and save memory. And the experiment verified that after deleting dropout, the effect of the downstream tasks was enhanced, and the best result of SOTA was achieved. Albert also improved MLM by predicting n-gram fragments, which contain more semantic information, rather than random 15% tokens. However, NSP task is too simple. Positive samples are two sentences adjacent to each other, while negative samples are randomly selected from the training set, resulting in less semantic information of the trained vector. Therefore, an improvement is made on Albert, replacing NSP task with SOP (sentence order prediction). SOP task can capture more context semantic information than NSP task. The positive sample of NSP task is two sentences in normal order, and the negative sample is two sentences in transposition order. In a single task of the SOP, there is a mix of topic prediction and coherence prediction. Topic prediction can be learned in NSP task, but coherence prediction cannot be learned. SOP task is needed to learn coherence between sentences.

Albert reduced the number of parameters from the following two aspects: first is the parameterized factorization of the embedding vector. In Bert, the word embedding size E and the hidden layer size H are equal, and the dictionary V is large, so that $O(V * E)$ is also large, where E is much less than H. Second is the parameter sharing across layers. Albert shares all parameters of all layers, instead of just sharing parameters of the full connected layer and the attention layer, which can greatly reduce the number of required parameters. As can be seen from the above improvements,

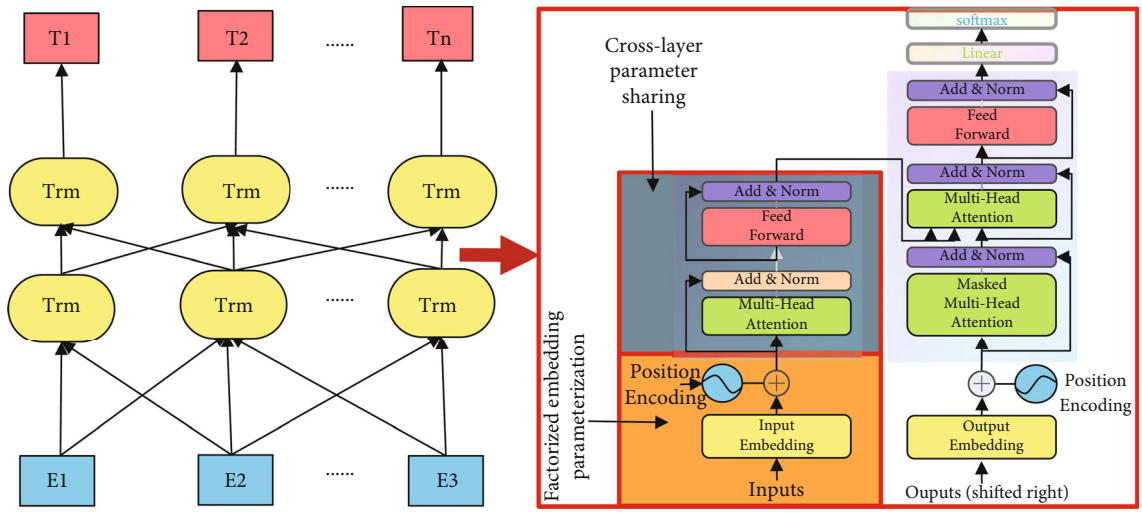

Figure 4: Albert model pretraining structure.

Albert is not only a lightweight Bert but also has been optimized.

Secondly, the general named entity recognition models cannot make use of the relationship between sentences effectively. Albert has an advantage in this respect. Compared with Word2vec, the character vector generated by Albert is dynamic, while the word vector generated by Word2vec is fixed, which cannot effectively solve the problem of polysemy. One of the advantages of Albert is that it can be used for transfer learning. The features extracted from the Albert pretraining model can be directly applied to the new task, or it can be fine-tuned and then applied to the new task. Albert is shown in Figure 4.

$E1$ and $E2$ are the input word vectors, Trm is the transformer encoder, and $E1$ and $E2$ are output by the encoder as $T1$ and $T2$.

4. Experiments and Analysis

4.1. Collection and Processing of the Original Corpses. At present, there are no tagged poems. The author crawled 1983 poems from the ancient Chinese poetry website, obtained the titles, contents and authors of the poems with crawlers, and stored them in the files.

The author found duplicates in the ancient poems, so it was necessary to search and delete the duplicated poems; the distribution of number of entity categories is shown in Table 3.

Entity statistics is as follows: in the ancient poetry dataset, there are 1983 poems, among which 1925 poems contain entity and 58 poems have no entity. In the content of poems, there are a total of 11,428 entities of 4 types, of which there are 143 types of time entities, totaling 2041, imagination entities have 295 types, totaling 8,045, place name entities have 269 types, totaling 972, and person name entities totaling 190 types, totaling 370.

The four types of high-frequency entities are shown in Table 4. Analysis shows that poets like to travel in spring in some scenic spots, such as "江南" (Jiangnan), "洞庭湖" (Dongting Lake), "长安" (Chang'an City), and "西湖" (West Lake), to write poems about the wind, flowers, snow, and moon, which is also in line with the actual situation.

Table 3: Distribution of number of entity categories.

Entity symbol	Train set	Test set
Time	1756	285
Scene	6614	1431
Person	297	73
Location	790	182

4.2. Universal Experimental Dataset. In order to further verify the generality of the model in this paper, the author conducts experiments on public datasets in the field of news, social media, and finance. (1) MSRA: datasets in the field of news, including three entities: place name (LOC), organization name (ORG), and person name (PER). (2) Weibo: the field of social media, including place name (LOC), organization name (ORG), administrative region name (GPE), and person name (PER). (3) Resume: financial domain, including place name (LOC), organization name (ORG), and person name (PER) and other entities.

The annotation methods of the three datasets are all BIO annotation. The detailed information of the datasets is shown in Table 5.

4.3. Experimental Environment and Parameter Setting. Firstly, the model was trained and tested under the framework of Python3.6 and Tensorflow1.14. The experimental hardware was 1080Ti, and the video memory was 11G. The Albert-based model was used in the experiment, with 64 multihead attention mechanism and 768 layers of hidden layers. The state of LSTM network hidden layer is set to 200 dimensions from front to back. The maximum sequence length was set to 64, and Adam was selected as the optimization function to reduce the loss each time. The learning rate of the model is set to 0.001, and the dropout is set to 0.5 to prevent

Table 4: High-frequency entities of each entity type.

Time	Imagination	Person name	Place name
春 (spring)	花 (flower)	宋玉 (Song Yu)	江南 (Jiangnan)
秋 (autumn)	风 (wind)	禹 (Yu)	洞庭 (Dongting Lake)
今日 (today)	云 (cloud)	相如 (Xiangru)	长安 (Chang'an City)
夏 (summer)	雨 (rain)	刘郎 (Liu Lang)	西湖 (West Lake)
三月 (march)	月 (moon)	匈奴 (Huns)	钱塘 (Qiantang)
昨夜 (last night)	酒 (liquor)	飞燕 (Zhao Feiyan)	南山 (Nanshan)
清明 (Ching Ming Festival)	雪 (snow)	大铁椎 (big iron vertebrae)	长江 (Yangtze)
去年 (last year)	柳 (willow)	尧 (Yao)	洛阳 (Luoyang City)
今夜 (tonight)	马 (horse)	舜 (Shun)	潇湘 (Xiaoxiang)

Table 5: Details of the three datasets.

Dataset	Type	Train set	Validation set	Test set
MSRA	Sentence	46.4 k	—	4.4 k
	Word	2169.9 k	—	172.6 k
Weibo	Sentence	1.4 k	0.27 k	0.27 k
	Word	73.8 k	14.5 k	14.8 k
Resume	Sentence	3.8 k	0.46 k	0.48 k
	Word	124.1 k	13.9 k	15.1 k

overfitting. The batch size of the training set, validation set, and the test set was selected as 64, and the maximum number of iterations of the model was 500. The best model was saved each time.

Second is the experimental comparison based on CRF model. As a traditional statistical machine learning method, conditional random field is more classical, so the CRF model is used as baseline model in this paper.

4.4. Evaluation Indexes and Experimental Results. First is the evaluation indexes: the evaluation indexes in this paper adopted the classical accuracy P, recall rate, and the harmonic mean of the two, namely, $F1$ value.

Second is the experimental results: CRF, BiLSTM-CRF, and Bert-BiLSTM-CRF models and their comparison with Albert-BiLSTM-MHA-CRF models were conducted in this experiment. The accuracy, recall rate, and $F1$ value of named entity recognition of each model were shown in Table 6.

It can be seen from Table 6 that CRF can identify a considerable number of entities. Compared with the CRF model, the $F1$ value of BiLSTM-CRF increases by 0.63%, indicating that the ability of extracting features and the effect of entity recognition are improved after the addition of BiLSTM. Followed by the Bert-BiLSTM-CRF model, compared with BiLSTM-CRF and CRF models, accuracy, recall rate, and $F1$ value of Bert-BiLSTM-CRF have been greatly improved. Compared with BiLSTM-CRF model and CRF model, $F1$ value was increased by 4.53% and 5.16%, respectively. It shows that Bert uses bidirection transformer to extract features based on contextual semantic depth and can learn character-level, word-level, and sentence-level features through the two tasks of MLM and NSP during pretraining. Overall, the optimal model is the Albert-BiLSTM-MHA-CRF model.

The $F1$ value of the Albert-BiLSTM-MHA-CRF model was 1.27% higher than that of the Bert-BiLSTM-CRF model. Under the same hyperparameter setting, when the Epoch number was set to 500, the running time of Albert was 6 hours 43 minutes, while the running time of Bert was 8 hours 21 minutes. Albert's training time is 98 minutes less than Bert's. This is because the number of parameters is much smaller than that of Bert due to cross-layer parameter sharing and factorization of embedded vector parameterization, leading to Albert's faster speed than that of Bert's training. The results on the $F1$ value can be explained from both Albert and MHA models. On the one hand, Bert has more parameters than Albert, and to a certain extent, it can train a better model than Albert. However, Albert has also improved based on Bert, and the amount of data is relatively small. The improvement of Albert over Bert lies in the improvement of the two tasks during pretraining. During pretraining, the mask on the MLM task is the N-gram segment, and the N-gram segment contains more semantic information, and the simpler NSP task is replaced by the SOP task, resulting in a better effect of Albert than Bert. On the other hand, although the BiLSTM model can obtain the current context information, as the length of the sentence increases, BiLSTM will also lose some more important information. The MHA model can fully capture long-distance features and obtain global information. Obtaining multiple features from the character, word, and sentence level improves the effect of entity relationship extraction, assigns more weight to important content in the text, and reduces the attention to nonimportant features, so that it is easier to capture long-distance important features. In summary, the Albert-BiLSTM-MHA-CRF model performs better on the ancient poetry dataset.

In addition, the $F1$ values of entity recognition of the method proposed in this paper and other methods based on deep learning on public datasets in three different domains are shown in Table 7. These models are lattice LSTM [30], a lstm model that fully considers word and character information; CAN_NER [31], a character-based local attention layer convolutional neural network (CNN) and gated recurrent unit (GRU) with global self-attention layer; CWPC_BiAtt [32], a attention-based bilstm model combining character and word position information; and ACNN [33], a model combining

Table 6: Experimental results of different models of ancient poetry dataset.

Model	Accuracy	Recall	$F1$ value
CRF	92.05%	85.53%	88.67%
BiLSTM-CRF	90.79%	87.85%	89.30%
Bert-BiLSTM-CRF	93.89%	93.77%	93.83%
Albert-BiLSTM-MHA-CRF	96.73%	97.62%	97.17%

Table 7: $F1$ values of different models in three public datasets.

Model	Weibo	MSRA	Resume
CRF	48.76%	84.22%	93.92%
BiLSTM-CRF	52.35%	90.25%	94.31%
Lattice-LSTM	58.79%	93.18%	94.46%
CAN_NER	59.31%	92.97%	94.94%
CWPC_BiAtt	59.5%	92.99%	—
ACNN	—	93.01%	94.45%
Bert-BiLSTM-CRF	67.33%	**94.83%**	95.51%
Albert-BiLSTM-MHA-CRF	**72.84%**	94.60%	**95.92%**

Table 8: $F1$ values for different scales of the dataset.

Model	MSRA_ 1.1 k	MSRA_ 5.5 k	MSRA_ 11k
Bert-BiLSTM-CRF	74.56%	89.46%	92.51%
Albert-BiLSTM-MHA-CRF	76.30%	89.80%	91.87%

multilevel CNN and attention mechanism. As can be seen from Table 7, Albert-BiLSTM-MHA-CRF has improved the effect of entity recognition on Weibo and Resume datasets. Compared with Bert-BiLSTM-CRF, the entity recognition effect of Albert-BiLSTM-MHA-CRF is increased by 5.51% and 0.41% on $F1$ value, respectively. The improvement effect of $F1$ value on Weibo dataset is the most obvious. The results are poor on MSRA dataset. This is because the MSRA data set is large, and the Bert-BiLSTM-CRF model has been well modeled for MSRA dataset, so there is little need for multihead self-attention mechanism. In terms of Weibo and Resume datasets, the model proposed in this paper can make full use of extracted words and sentence features to improve the accuracy of entity recognition. Compared with other models, the model proposed in this paper has achieved a better recognition effect on datasets in multiple fields, is more stable on different data sets, and has certain robustness. Therefore, the model proposed in this paper is proved to be effective.

This paper also conducts experimental analysis on the performance of Bert-BiLSTM-CRF and Albert-BiLSTM-MHA-CRF on different datasets and different data volumes. According to Tables 6 and 7, it can be seen that Albert-BiLSTM-MHA-CRF performs well on the Weibo and Resume datasets, but slightly worse on the MSRA model. The sentences of Weibo, Resume, and MSRA datasets are 1.94 k, 4.74 k, and 50.8 k, respectively, and Weibo has 4 types of entities, Resume has 8 types of entities, and MSRA has 3 types of entities. It can be concluded that the Weibo and Resume datasets are small and have many entity categories, while the MSRA dataset is large and has relatively few entity categories. Therefore, this paper focuses on the experimental analysis of the training set of the MSRA data set. The experimental results are shown in Table 8. With the increase of the amount of MSRA data, it can be seen that the entity extraction effect of Bert as the pretraining model is better than that of Albert, getting better and better. Combining Tables 7 and 8, the basic conclusion can be drawn: in these three datasets, when the number of entity types is greater than or equal to 3, and the number of sentences in the dataset is less than 5.5 k, the Albert-BiLSTM-MHA-CRF model is better than the Bert-BiLSTM-CRF model.

Table 9: Experimental results of various entity types in Albert-BiLSTM-MHA-CRF model.

Entity symbol	Accuracy	Recall	$F1$ value
Location	93.75%	93.75%	93.75%
Person	54.55%	75.00%	63.16%
Scene	97.28%	98.11%	97.70%
Time	97.95%	97.70%	97.82%

At the same time, the experimental results of each entity type in the optimal model Albert-BiLSTM-MHA-CRF are analyzed, as shown in Table 9. The $F1$ value of the location entity is 11.42% higher than that of Y. Zhang [17] et al. The best effect of entity recognition is the time entity, whose $F1$ value reaches 97.82%, while person name entity has the worst effect, whose $F1$ value is 63.16%. Its recall rate is significantly lower than the accuracy rate. The low recall rate indicates that the original corpus's person name entity is predicted as nonperson name entity in more cases, while the accuracy rate is high, which means few cases in the corpus where nonperson name entities are predicted to be person name entities. Since there are 190 categories of person name entity in the corpus and there are 370 person name entities in total, this results in a small number of entities per entity type, while the other three types of entities have few entity types, which leads to more features identified by the other three types of entities. This increases the number of situations in which person name entities are predicted to be nonperson name entities during training, resulting in a lower recall rate for person name entities during training.

4.5. Knowledge Graph Display for Named Entity Recognition of Ancient Poems. The knowledge graph as shown in Figure 5 is constructed by establishing connections among the four identified entities and the name of the poem, the author of the poem and the dynasty of the poem. The nodes in the knowledge graph of ancient poetry include poetry names, four types of entities, poets, and the dynasty of poets. The three kinds of edges include "has_poetry," "entity_is," and "dynasty_is," and edges are the connection between entities. For example, "纳兰性德" (Nalanxingde) established a connection with the entity "花" (flower) through his poem "如梦令" (Like A Dream), and other poets also have the entity "花" (flower) in their poems. It can be seen that the connection between poems is

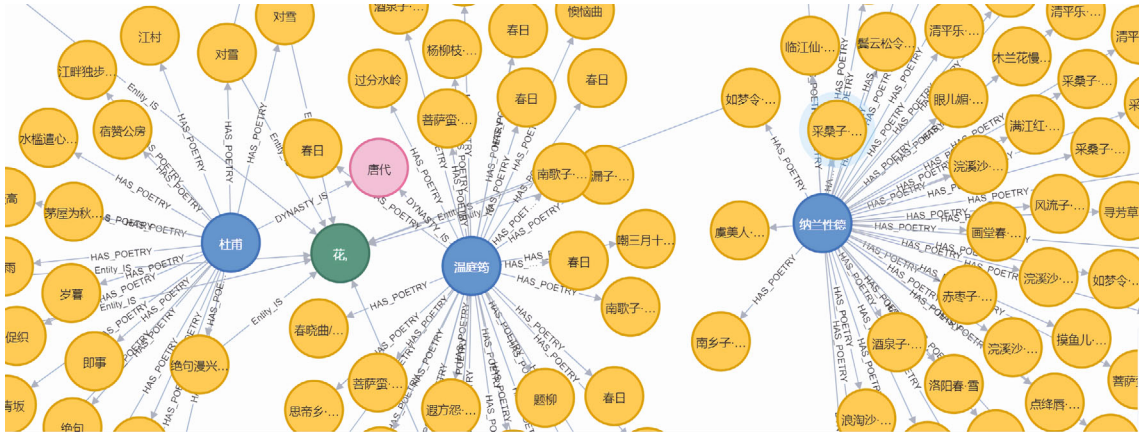

Figure 5: Knowledge graph of ancient poetry entity and relation.

established through various entities, and the imagination, time, and place name that poets like can also be found through the statistics of entities in their poems. The more frequently the characters and entities appear in the poem, the closer the relationship with the poet is.

4.6. Reasons for Identification Errors of the Four Types of Named Entities. For the best Albert-BiLSTM-MHA-CRF model, the *F*1 value is 97.17%, but there are still misidentified entities. After analysis, there are four reasons as follows:

Firstly, the design of a single entity is too complicated. For example, the time entity can not only represent a specific time in a certain year on a certain day, but also the name of a dynasty such as "越明年" (next year), which can represent a specific period of time in history, as well as the "甲子" (Jiazi) time of the lunar calendar, such as "庚戌" (Gengxu) and other lunar timing, which increases the difficulty of identifying the time entity.

Secondly, the entity of a single word is difficult to recognize. Compared with modern Chinese, ancient poetry has more entities of a single word. For example, the "蕙" (Hui) and "芷" (angelica) entities, in the imagination entity, are plants in ancient Chinese, far from the general imagination entity, so this entity of a single word is difficult to identify.

Thirdly, there will also be crossover between the four types of entities, such as "秋 风" (autumn wind) and "秋" (autumn). "秋风" (autumn wind) is an imagination entity, while "秋" (autumn) is a time entity representing time, which is easy to mistakenly identify the entity to be identified as another entity.

Fourthly, the data is unbalanced. For example, the number of time entities "春" (spring), "夏" (summer), "秋" (autumn), and "冬" (winter) and a certain day in a certain year are relatively large. However, such as "七夕" (Qixi Festival) and "寒食" (Cold Food Festival), the number of time entities representing festivals is usually relatively small, which brings difficulties to identifying time entities representing festivals.

Fifthly, the model falls into saddle point. As the network structure becomes more complex and the network training parameters increase, the model may fall into local optimization, that is, saddle point, resulting in errors in the recognition of four types of entities.

4.7. Improvement of Named Entity Recognition Errors. In view of the above four types of entity recognition errors, the author believes that improvements can be made from the following aspects:

Firstly, build a dictionary related to ancient poetry. The new words in ancient poetry can be found by using the new word discovery algorithm based on mutual information and left-right entropy. For the problem of confusion between unrecognizable words and entity names in ancient poetry, characters and words can be added to the dictionary, such as "秋风" (autumn wind) and "秋" (autumn), which is conducive to improving the effect of entity recognition.

Secondly, continue to subdivide complex entity types. For example, the time entity can be further subdivided into Jiazi time, festivals, etc. The model can more easily extract the same type of features.

Thirdly, expand the scale of corpus. Expanding the corpus can increase the frequency of each entity type, make the model training and extract relevant features more fully, and improve the training effect.

Fourthly, using the gradient activation function (GAF) proposed by Liu [34] et al., the function can make the model alleviate the saddle point problem and obtain a global optimal solution to improve the effect of entity recognition.

4.8. Application in the Field of Password Guessing. Passwords in text form are still commonly used authentication mechanisms in various computer systems [35], and passwords are essentially short texts with rich semantics, which contain the user's personal information, such as name, birthday, and mobile phone number. The previous method is to use PCFG (probabilistic context free grammar) to obtain all possible password combinations and corresponding probabilities of a user. For example, D. Wang [36] et al. proposed TarGuess, on the basis of PCFG; the user's personal information and the password leaked by the user on other similar websites are added. With the continuous development of deep learning, password guessing based on neural networks has gradually emerged. Melicher [37] et al. found that neural networks

are better at guessing passwords than PCFG at higher guessing times and for more complex or longer passwords. However, PCFG still has advantages. For example, Veras [38] et al. applied LSTM to password guessing and found that compared with the LSTM model, PCFG is still a competitive model, which is more important for the security of passwords, and PCFG is more explanatory.

The model proposed in this paper can be applied in the field of password guessing, as text, passwords are mainly composed of numbers, letters, and symbols, and both passwords and ancient poems are short texts. Albert-BiLSTM-MHA-CRF can more accurately identify entities such as person entity, birthday date entity, mobile phone number, and other entities according to the context in the constructed leaked password dataset, not just based on the user's personal information in the dataset. And it can build a knowledge graph in the field of passwords and assist PCFG to guess passwords with fewer times through the entity relationship between passwords.

5. Conclusion

With the continuous development of entity name recognition technology and deep learning-related models, the recognition accuracy of modern Chinese has been greatly improved. However, few researches on the entity recognition are related to ancient Chinese, such as the entity extraction related to ancient poetry. In this paper, the entity extraction of ancient Chinese is carried out on the basis of Albert pretraining model. In the following research, we will expand the scale of the corpus of ancient poems and words and explore more brand-new models to further improve the effect of entity recognition of ancient poems.

Data Availability

All data included in this study are available upon request by contact with the corresponding author.

Conflicts of Interest

The authors declare no conflict of interest.

Authors' Contributions

C.W., F.Z., and J.W. conceptualized the study; methodology was carried out by C.W. and F.Z.; C.W. was responsible for the software; C.W. and F.Z. were responsible for the formal analysis; investigation was carried out by C.W. and J.W.; C.W. was responsible for the data curation; C.W. wrote the original draft; C.W. and J.W. wrote, reviewed, and edited the manuscript; C.W. visualized the study; F.Z. supervised the study; F.Z. was responsible for the project administration; F.Z. was responsible for the funding acquisition. All authors have read and agreed to the published version of the manuscript.

References

[1] S. Ji, S. Pan, E. Cambria, P. Marttinen, and P. S. Yu, "A survey on knowledge graphs: representation, acquisition and applications," 2002, CoRR 2020, abs/2002.00388.

[2] Y. Wei, H. Wang, J. Zhao, Y. Liu, Y. Zhang, and B. Wu, "GeLaiGeLai: a visual platform for analysis of classical Chinese poetry based on knowledge graph," in *2020 IEEE International Conference on Knowledge Graph, ICKG 2020*, pp. 513–520, 2020.

[3] W. Jiang, C. Li, and C. Wu, "The visualization of cross-media knowledge graph of tang and song poetry," in *21st ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD Winter 2021*, pp. 69–73, Ho Chi Minh City, Vietnam, 2021.

[4] Z. Chen, S. Yin, and X. Zhu, "Research and implementation of QA system based on the knowledge graph of Chinese classic poetry," in *2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, pp. 495–499, 2020.

[5] N. Deng, H. Fu, and X. Chen, "Named entity recognition of traditional Chinese medicine patents based on BiLSTM-CRF," *Wireless Communications and Mobile Computing*, vol. 2021, 12 pages, 2021.

[6] X. Wang, R. Yang, Y. Lu, and Q. Wu, "Military named entity recognition method based on deep learning," in *5th IEEE International Conference on Cloud Computing and Intelligence Systems, CCIS 2018*, pp. 479–483, Nanjing, China, 2018.

[7] R. Grishman and B. Sundheim, "Message understanding conference-6: a brief history," in *Message understanding Conference-6: a brief History*, 1996.

[8] A. Bouziane, D. Bouchiha, N. Doumi, and M. Malki, "Question answering systems: survey and trends," *Procedia Computer Science*, vol. 73, pp. 366–375, 2015.

[9] B. Zhong, W. He, Z. Huang, P. E. D. Love, J. Tang, and H. Luo, "A building regulation question answering system: a deep learning methodology," *Advanced Engineering Informatics*, vol. 46, article 101195, 2020.

[10] C. Niklaus, M. Cetto, A. Freitas, and S. Handschuh, "A survey on open information extraction," in *Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018*, pp. 3866–3878, Santa Fe, New Mexico, USA, 2018.

[11] J. Zhu, Z. Nie, X. Liu, B. Zhang, and J. R. S. S. Wen, "A statistical approach to extracting entity relationships," in *Proceedings of the 18th international conference on World Wide Web*, pp. 101–110, 2009.

[12] S. Di, Y. Shen, and L. Chen, "Relation extraction via domain-aware transfer learning," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019*, pp. 1348–1357, Anchorage, AK, USA, 2019.

[13] S. Jia, E. Shijia, M. Li, and Y. Xiang, "Chinese open relation extraction and knowledge base establishment," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 17, no. 3, pp. 1–22, 2018.

[14] J. Björne and T. Salakoski, "Generalizing biomedical event extraction," in *Proceedings of BioNLP Shared Task 2011 Workshop*, pp. 183–191, Portland, Oregon, USA, 2011.

[15] A. K. Cybulska and P. Vossen, "Historical event extraction from text," in *Proceedings of the 5th ACL Workshop on Language Technology for Cultural Heritage, Social Sciences, and*

Humanities, LaTeCH@ACL 2011, pp. 39–43, Portland, Oregon, USA, 2011.

[16] T. Zhang, S. Whitehead, H. Zhang et al., "Improving event extraction via multimodal integration," in *Proceedings of the 2017 ACM on Multimedia Conference, MM 2017*, pp. 270–278, Mountain View, CA, USA, 2017.

[17] Y. Zhang, Y. Li, J. Zhang, and Y. Ye, "A method for place name recognition in Tang poetry based on feature templates and conditional random field," in *Web and Big Data -4th International Joint Conference, APWeb-WAIM 2020*, pp. 627–635, Tianjin, China, 2020.

[18] N. Limsopatham and N. Collier, "Bidirectional LSTM for named entity recognition in Twitter messages," in *Proceedings of the 2nd Workshop on Noisy User-generated Text, NUT@-COLING 2016*, pp. 145–152, Osaka, Japan, 2016.

[19] L. Simeonova, K. Simov, P. Osenova, and P. Nakov, "A morpho-syntactically informed LSTM-CRF model for named entity recognition," in *Proceedings of the International Conference on Recent Advances in Natural Language Processing, RANLP 2019*, pp. 1104–1113, Varna, Bulgaria, 2019.

[20] X. Yang, Z. Gao, Y. Li et al., "Bidirectional LSTM-CRF for biomedical named entity recognition," in *14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, ICNC-FSKD 2018*, pp. 239–242, Huangshan, China, 2018.

[21] E. Taher, S. A. Hoseini, and M. Shamsfard, "Beheshti-NER: Persian named entity recognition using BERT," 2003, CoRR 2020, abs/2003.08875.

[22] K. Hakala and S. Pyysalo, "Biomedical named entity recognition with multilingual BERT," in *Proceedings of The 5th Workshop on BioNLP Open Shared Tasks, BioNLP-OST@EMNLP-IJNCLP 2019*, pp. 56–61, Hong Kong, China, 2019.

[23] W. Zhang, S. Jiang, S. Zhao, K. Hou, Y. Liu, and L. Zhang, "A BERT-BiLSTM-CRF model for Chinese electronic medical records named entity recognition," in *2019 12th International Conference on Intelligent Computation Technology and Automation (ICICTA)*, pp. 166–169, 2019.

[24] M. Zhang, Z. Yang, C. Liu, and L. Fang, "Traditional Chinese medicine knowledge service based on semi-supervised BERT-BiLSTMCRF model," in *2020 International Conference on Service Science, ICSS 2020*, pp. 64–69, Xining, China, 2020.

[25] H. Lv, Y. Ning, and K. Ning, "ALBERT-based Chinese named entity recognition," in *Cognitive Computing-ICCC 2020-4th International Conference, Held as Part of the Services Conference Federation, SCF 2020*, pp. 79–87, Honolulu, HI, USA, 2020.

[26] J. D. Lafferty and A. McCallum, Eds.F. C. N. Pereira, "Conditional random fields: probabilistic models for segmenting and labeling sequence data," in *Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001)*, pp. 282–289, Williams College, Williamstown, MA, USA, 2001.

[27] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," 2015, CoRR 2015, abs/1508.01991.

[28] Z. Dai, X. Wang, P. Ni, Y. Li, G. Li, and X. Bai, "Named entity recognition using BERT BiLSTM CRF for Chinese electronic health records," in *12th International Congress on Image and Signal Processing, Bio Medical Engineering and Informatics, CISP-BMEI 2019*, pp. 1–5, Suzhou, China, 2019.

[29] J. Yang, Y. Zhang, L. Li, and X. Y. E. D. D. A. Li, "A lightweight collaborative text span annotation tool," in *Proceedings of ACL 2018*, pp. 31–36, Melbourne, Australia, 2018.

[30] Y. Zhang and J. Yang, "Chinese NER using lattice LSTM," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018*, pp. 1554–1564, Melbourne, Australia, 2018.

[31] Y. Zhu and G. Wang, "CAN-NER: convolutional attention network for Chinese named entity recognition," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019*, pp. 3384–3393, Minneapolis, MN, USA, 2019.

[32] S. Johnson, S. Shen, and Y. Liu, "CWPC_BiAtt: character-word-position combined BiLSTM-attention for Chinese named entity recognition," *Information*, vol. 11, no. 1, p. 45, 2020.

[33] J. Kong, L. Zhang, M. Jiang, and T. Liu, "Incorporating multi-level CNN and attention mechanism for Chinese clinical named entity recognition," *Journal of Biomedical Informatics*, vol. 116, article 103737, 2021.

[34] M. Liu, L. Chen, X. Du, L. Jin, and M. Shang, "Activated gradients for deep neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. PP, pp. 1–13, 2021.

[35] D. Wang, P. Wang, D. He, and Y. Tian, "Birthday, name and bifacial-security: understanding passwords of Chinese web users," in *28th USENIX Security Symposium, USENIX Security 2019*, pp. 1537–1555, Santa Clara, CA, USA, 2019.

[36] D. Wang, Z. Zhang, P. Wang, J. Yan, and X. Huang, "Targeted online password guessing: an underestimated threat," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security; Association for Computing Machinery*, pp. 1242–1254, New York, NY, USA, 2016.

[37] W. Melicher, B. Ur, S. M. Segreti et al., "Fast, lean, and accurate: modeling password guessability using neural networks," in *25th USENIX Security Symposium, USENIX Security 16*, pp. 175–191, Austin, TX, USA, 2016.

[38] R. Veras, C. Collins, and J. Thorpe, "A large-scale analysis of the semantic password model and linguistic patterns in passwords," *ACM Transactions on Privacy and Security*, vol. 24, no. 3, pp. 1–21, 2021.