


Article

Named Entity Identification in the Power Dispatch Domain Based on RoBERTa-Attention-FL Model

Yan Chen ^{1,2,*} , Dezhao Lin ¹, Qi Meng ³, Zengfu Liang ¹ and Zhixiang Tan ¹¹ School of Computer and Electronic Information, Guangxi University, Nanning 530004, China² Guangxi Intelligent Digital Services Research Center of Engineering Technology, Nanning 530004, China³ Guangxi Power Grid Co., Ltd., Nanning 530022, China

* Correspondence: cy@gxu.edu.cn

Abstract: Named entity identification is an important step in building a knowledge graph of the grid domain, which contains a certain number of nested entities. To address the issue of nested entities in the Chinese power dispatching domain's named entity recognition, we propose a RoBERTa-Attention-FL model. This model effectively recognizes nested entities using the span representation annotation method. We extract the output values from RoBERTa's middle 4–10 layers, obtain syntactic information from the Transformer Encoder layers via the multi-head self-attention mechanism, and integrate it with deep semantic information output from RoBERTa's last layer. During training, we use Focal Loss to mitigate the sample imbalance problem. To evaluate the model's performance, we construct named entity recognition datasets for flat and nested entities in the power dispatching domain annotated with actual power operation data, and conduct experiments. The results indicate that compared to the baseline model, the RoBERTa-Attention-FL model significantly improves recognition performance, increasing the F1-score by 4.28% to 90.35%, with an accuracy rate of 92.53% and a recall rate of 88.12%.

Keywords: power dispatching; named entity recognition; RoBERTa; self-attention mechanism; syntactic information



Citation: Chen, Y.; Lin, D.; Meng, Q.; Liang, Z.; Tan, Z. Named Entity Identification in the Power Dispatch Domain Based on RoBERTa-Attention-FL Model. *Energies* **2023**, *16*, 4654. <https://doi.org/10.3390/en16124654>

Academic Editor: Abu-Siada Ahmed

Received: 11 May 2023
Revised: 3 June 2023
Accepted: 9 June 2023
Published: 12 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As smart power systems are put into use, a huge amount of dispatching behavior information is recorded in the process. This is stored in an unstructured form and contains rich knowledge of dispatching behavior [1–3]. Carrying out deep mining of unstructured data in the Chinese power dispatching domain, modeling the empirical knowledge in the power dispatching domain, and constructing domain knowledge graphs have emerged as important tasks in this field. Among these, Named Entity Recognition (NER) is a fundamental task in natural language processing to recognize entities with specific meanings or strong referents in texts, including names of people, places, proper nouns, etc. It is also a key technology for building knowledge graphs with a wide range of applications [4,5]. Upon identifying the entities in unstructured data, the relationships between entities are extracted, and the entities are then connected through their relationships to build the semantic network of the knowledge graph.

The main difficulties of NER technology in the Chinese power dispatching domain are as follows:

- (1) There is a lack of publicly available annotated datasets.
- (2) Power entities are highly specialized, and these are more difficult to identify than general domains. There are also problems, such as nested entities in power entities. For example, “35 kV 北龙线” (35 kV Beilong Line) contains “35 kV” and “35 kV 北龙线”.
- (3) The traditional sequence annotation method requires special processing of the model [6–9] to identify the nested entities.

- (4) Chinese has blurred word boundaries, unlike English, which has separators between words; thus, when conducting NER, the process requires separate words in the text [10,11]. The word separation error generated by the separator affects the accuracy of the named entity recognition. There is no authoritative dictionary for the word splitter in the power dispatching field, and applying a general-purpose domain word splitter to the power domain results in significant errors.

To solve the above problems, a new entity classification method and a new NER model for the power dispatching domain are proposed to improve the recognition of power dispatching entities. The contributions of this paper are as follows:

- (1) We preprocess the dispatching data provided by the Guangxi power grid, extract the unstructured data, annotate them with reference to the national standard electrical terminology specification, and construct a named entity recognition dataset in the power dispatching domain.
- (2) We propose a named entity identification method based on the RoBERTa-Attention-FL model for the power dispatching domain that uses the conventional model to effectively identify nested entities based on the annotation method of span representation. We encode text information in words to avoid errors caused by word segmentation.
- (3) To construct a syntactic information vector, we extract the output values of the middle four to ten layers of RoBERTa. The Transformer Encoder layers' multi-headed attention mechanism is used to learn the information that is important for the model. We fuse the extracted syntactic information with the deep semantic information output from the last layer of RoBERTa for boundary enhancement of the span, and the neural network learns the fusion weights without human tuning. The deep semantic information and syntactic information are computed in parallel, and the predicted power span entities are obtained after the fully connected layer.
- (4) The Focal loss [12] function is used during training to alleviate the problem of sample imbalance.

The experimental results demonstrate that our model outperforms other models, such as BERT-Cross Entropy, BERT-CRF, and BERT-BILSTM-CRF, in terms of recognition performance.

Section 2 of this paper introduces the related work, Section 3 describes the construction of the dataset, and Section 4 introduces the RoBERTa-Attention-FL model. Section 5 verifies the effectiveness of the proposed model through comparative experiments. The final section summarizes the entire text.

2. Related Work

Traditional named entity recognition methods are broadly divided into two groups: unsupervised rule-based learning and supervised feature-based learning [13]. Zhang [14] proposed an unsupervised approach for extracting named entities from biomedical texts. However, the unsupervised rule-based learning approach is highly dependent on rule formulation, lacks transferability, and cannot be easily adapted to other domains.

With the development of machine learning, supervised feature-based learning has gradually replaced unsupervised rule-based learning in named entity recognition. This type of learning can be further divided into traditional machine learning and deep learning approaches. Patil [15] proposed a method for named entity recognition using Conditional Random Fields (CRF) and feature selection. However, traditional machine learning methods require feature selection, whereas deep learning techniques can automatically extract feature information through models; therefore, they are more widely used in various fields. For instance, Srivastava [16] performed named entity recognition based on word embedding and deep learning models for web information security texts. Similarly, Zhang [17] proposed a pre-trained Chinese financial domain named entity recognition model that contains two sub-models for financial entity boundary delineation and financial entity classification. Puccetti [18] provided a patented text named entity recognition system that combines rule-based, gazetteer, and deep learning techniques. Additionally, Li [19] merged

dictionary and Chinese radical features into the BERT-BiLSTM-CRF model for the named entity recognition of clinical terms.

Pre-trained language models are widely used in named entity recognition, and the BERT pre-trained models are the most commonly used. These models can obtain deep contextual representations of text, allowing for the more accurate identification of named entities. Zheng [20] proposed a new NER model, AttCNN-BiGRU-CRF, that combines BERT-based character embedding and word embedding for the recognition of power metering databases. He [21] proposed a progressive multi-type feature fusion entity recognition method based on the BERT preprocessing model to obtain word vectors with contextual information for named entity recognition in electric power maintenance datasets. Tong [22] proposed a named entity approach for power communication planning based on Transformer and BiLSTM-CRF models that improves the efficiency of information extraction in this domain.

BERT is a neural network with a 12-layer structure, but typically only the final layer is used for contextual text representation. Nonetheless, Ganesh [23] performed probing experiments on each layer of the BERT model and discovered that the bottom layer learns phrase-level information representation, the middle layers learn rich syntactic features, and the top layers learn deep semantic information features. RoBERTa is a variant of BERT. Moreover, Zhang [14] proposed an unsupervised method for extracting named entities from biomedical texts using syntactic information to improve the accuracy of named entity recognition. By drawing on the idea of machine learning feature extraction, we fully utilize the RoBERTa middle layers to extract syntactic information and enhance the model's information gain.

Deep-learning-based named entity recognition usually uses the BIOES-style sequence annotation method, where "B" denotes the beginning position of the entity, "I" denotes the internal position of the entity, and "O" denotes the non-entity part. This annotation method enables each entity to be annotated only once. However, in generic domain unstructured data, it is often the case that an entity contains a small entity, and this is called the nested entity problem [24]. Geng [25] proposed a planarized way to represent nested named entities and implemented bidirectional two-dimensional recurrent operations to learn semantic dependencies between spans. Zhong [26] segmented the English vocabulary into word roots, applied span annotation, and generated candidate entities with sentence splicing as a training example, effectively enhancing the accuracy of the relationship extraction for downstream tasks. Ye [27] proposed a neighborhood-oriented packing strategy to pack as many spans as possible with the same starting lexical elements into a training instance to better distinguish entity boundaries. The problem of nested entities also arises frequently in the power domain.

Therefore, we propose the RoBERTa-Attention-FL model to solve the nested entity problem using the annotation method of span representation. First, we generate training data; then, we fuse the RoBERTa mid-level syntactic information as information gain with deep-level semantic information, and finally we perform entity recognition on the power dispatch dataset.

3. Constructing Corpus Datasets

The Guangxi regional smart power system records a vast amount of unstructured behavioral data that includes text data such as accident investigation reports, audit risk statistics, field inspection information, and device operations. A power corpus is built from these data, but the traditional system has low utilization of the information, only simple text queries are supported without deep mining, and the embedded behavioral knowledge remains largely untapped. Manual mining methods are inefficient and costly. In this study, we propose a deep learning approach to develop analytical models. To achieve this goal, we use the power corpus to create named entity recognition datasets in the power dispatching field that will be used to train deep learning models.

The corpus used in this paper contains a large amount of unstructured data. We filtered and removed sentences with unclear meaning, structural mutilation, and semantic repetition. Following this, we extracted 7717 pieces of data for training and testing purposes. Zheng [20] crawled through data concerning electric power metering from several electric power websites to construct the electric power metering corpus and divided them into five categories of entities, including metering data, metering technology, and electric power equipment, to construct the dataset. He [21] used power maintenance records to construct seven categories of power maintenance datasets, and these are voltage level, equipment name, line name, etc. Referring to the above two division schemes and combining the characteristics of the corpus in this paper, the entity types are divided into nine categories, and these are time, voltage level, transmission equipment, etc. (as presented in Table 1). Some transmission equipment, equipment appliances, and stations contain voltage level information, which are labeled according to the nested entities method to enable the model to learn the nested entities' characteristics. This approach facilitates the overall grasp of fine-grained and coarse-grained entities in the knowledge graph construction phase. The dataset contains 48,386 entities with more than 520,000 characters in Chinese and English. The training set, validation set, and test set are divided in a ratio of 7:2:1.

Table 1. Experimental dataset.

Type Name	Entity Example	Number of Entities
Time	2022 年06 月30 日 (30 June 2022)	6471
Voltage level	35 kV	5943
Transmission equipment	35 kV 北龙线 (35 kV Beilong line)	8140
Equipment appliances	903 开关 (903 switch)	10,247
Address	鹿县寨沙镇 (Zaisha Town, Lu County)	5952
Organization	平俪有限公司 (Ping Li Ltd.)	5529
Station	平朗站 (Pinglang Station)	3965
Other	老鼠, 桉树 (Rats, Eucalyptus)	1403
Name	张三 (San Zhang)	737

In this paper, we use a span representation-based annotation method to annotate entities using a visual interface through the Label Studio annotation platform. The span annotation consists of a set of start and end positions, as well as the type of entity in the sentence. An example of an annotation is shown in Table 2.

Table 2. Sample annotation.

Text	1	1	0	k	V	昆	仑	站
Index	0	1	2	3	4	5	6	7
Span entity	(start: 0, end: 4, level)			(start: 0, end: 7, station)				

The meaning of the Text line in the table is "110 kV Kunlun station".

4. Methods

In this paper, we propose a named entity recognition method based on the RoBERTa-Attention-FL model for the power dispatching domain. First, we generate training data and encode them at the character level using an embedding layer to obtain contextual representation information. Next, we input this information into RoBERTa and extract the Encoder output from layers four to ten. We splice this output into syntactic information vectors and use the multi-headed self-attentive mechanism in the Transformer Encoder to learn the weight messages. We then fuse this syntactic information with the last layer of the RoBERTa output and obtain the predicted entity messages via a fully connected layer. The flow chart is shown in Figure 1 and the model architecture is shown in Figure 2.

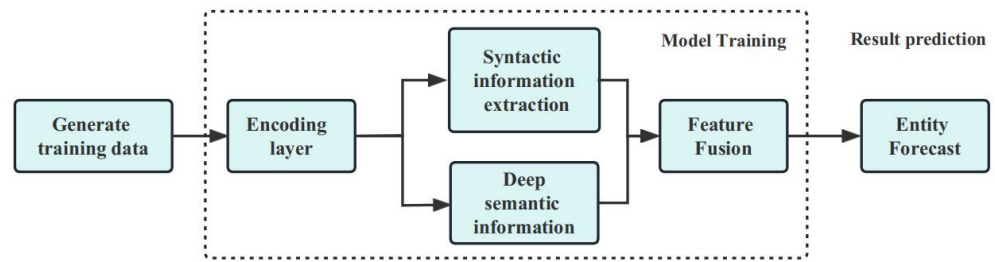


Figure 1. Flow chart. The syntactic information and deep semantic information are obtained through a parallel mechanism.

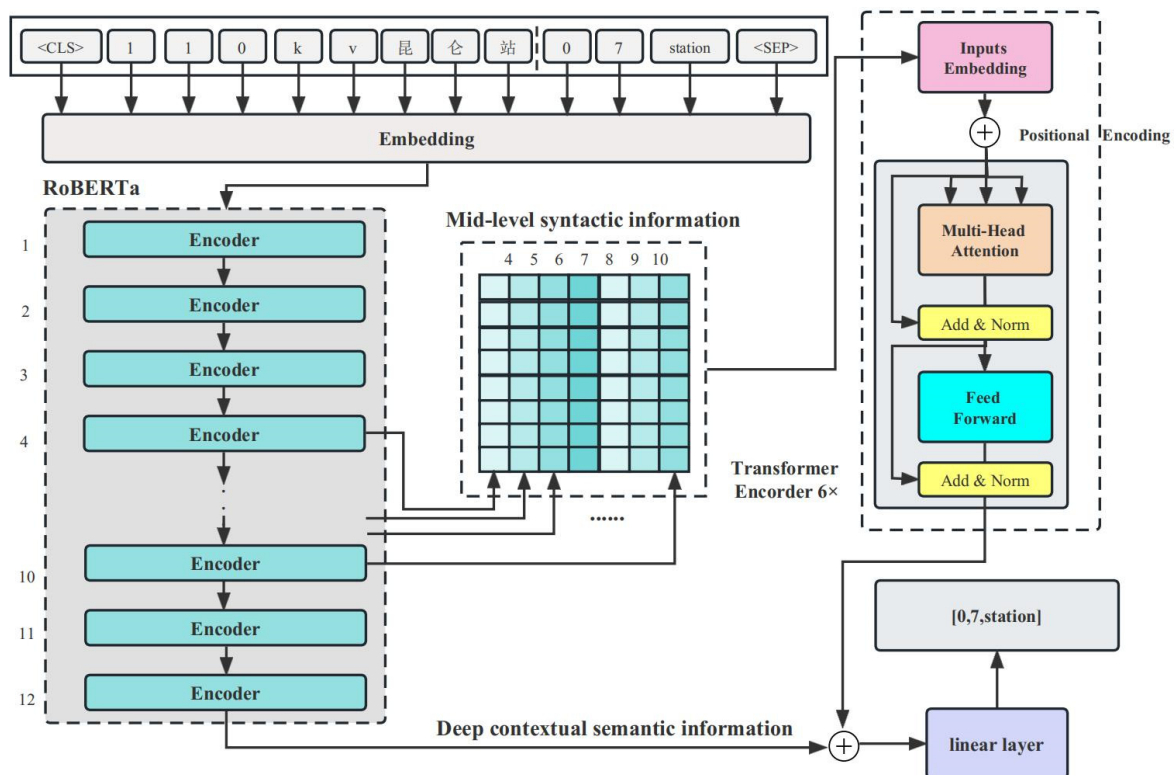


Figure 2. The graph first line “110 kV 昆仑站” corresponds to “110 kV Kunlun Station”. Overall framework of the model including input text examples (the first half of the segmentation line is sentence information and the second half is annotation information), encoding layer, Embedding, RoBERTa 12-layer network, syntactic information, and deep contextual semantic information extraction.

4.1. Generate Training Data

First, the annotation method based on span representation requires the enumeration of each span of a length less than K in the sentence, and for the sentence $S = \{s_1, s_2, \dots, s_n | K < n\}$, the length of the candidate span is set to K . The candidate span set $C(S) = \{(1, 1, t_{1,1}), (1, 2, t_{1,2}), \dots, (i, i, t_{i,i}), \dots, (i, i + K, t_{i,i+k}), \dots, (n - 1, n, t_{n-1,n}), (n, n, t_{n,n}) | i + K < n\}$ is generated, where n is the total length of the sentence. This set contains the start and end positions as well as the entity type $t_{i,i}$ of the candidate span that forms a triplet. The generated candidate span, which is the set of spans with the same start position, is used as a training instance, and multiple training groups are computed in parallel after placing the original text in groups [27].

4.2. Pre-Trained Language Models

A RoBERTa-wwm-ext Chinese pre-training model is used that is pre-trained, unsupervised, on large-scale Chinese text data. It learns rich prior knowledge and achieves excellent

performance in many natural language processing tasks. RoBERTa is a variant of BERT [28], with the following changes from BERT:

- The dynamic masking strategy may have different mask positions for each round of training samples. For the training sample “110 kV 昆仑站” (110 kV Kunlun station), the first round of training replaces the training sample with the special word “110 kV 昆仑<mask>”, the second round of training replaces the training sample with “<mask>10 kV 昆仑站”, and the mask position may change again in the third and fourth rounds. The dynamic strategy improves the randomness of the model input data that eventually improves the learning ability of the model;
- Using whole-sentence input across documents and eliminating next-sentence prediction;
- Using larger training batches and pre-training data to improve the generalization ability of the model.

(1) Encoding layer.

The role of the encoding layer is to convert the input text sequence into a series of high-dimensional vector representations that contain word encoding information, paragraph information, and location information of the input text that can model the long-distance dependencies of the input sequence and better characterize the deep semantic information of the text. The schematic diagram of the encoding layer is shown in Figure 3.

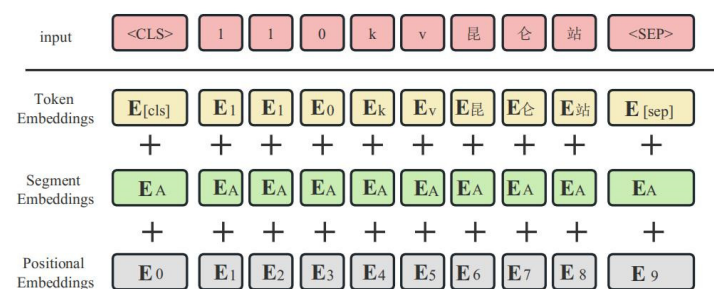


Figure 3. RoBERTa generates the input vector. The token embeddings subscript means “110 kV Kunlun Station”.

The training data input is passed through the RoBERTa encoding layer to obtain the word embedding $X_{embedding}$ that contains location information, paragraph information, and word encoding information:

$$X_{embedding} = X_{word} + X_{segment} + X_{positional} \tag{1}$$

4.3. Deep Contextual Semantic Information

We encode the training data using RoBERTa word embeddings to capture contextual features. The span-based data annotation format differs from traditional sequence annotation. We splice the representation of the start and end positions of the span set, along with the contextual features of the span set, which enhances the boundary features of the candidate spans and strengthens their connection with textual information. As shown in Figure 4, these correspond to the following equations:

$$h_{start} = H_{12}[:, S_{start}] \tag{2}$$

$$h_{end} = H_{12}[:, S_{end}] \tag{3}$$

$$h_{front} = H_{12}[:, S_{start} - 1] \tag{4}$$

$$h_{rear} = H_{12}[:, S_{end} + 1] \tag{5}$$

$$context = Concat(h_{front}, h_{rear}, h_{start}, h_{end}) \tag{6}$$



Figure 4. “6月30日, 110 kV 昆仑站广坤线” corresponds to “On 30 June, 110 kV Kunlun Station Guangkun Line”. Deep contextual semantic information. The solid line matrix box represents the character information corresponding to the start and end positions of the candidate span in the text, and the dashed matrix box represents the character information corresponding to one offset before and after the start and end positions of the candidate span in the text.

H_{12} denotes the output of the last layer of RoBERTa. Equations (2) and (3) obtain the trained features at the span start and end positions, respectively. Equations (4) and (5) obtain the trained features at one offset before and after the span start and end positions, respectively. Equation (6) splices these three features with Concat to obtain deep contextual semantic information.

4.4. Syntactic Information

Traditional machine learning methods use syntactic information features to improve model accuracy, while we use RoBERTa intermediate layers that learn more syntactic information to enhance the performance of the model. The output of RoBERTa’s middle layers four to ten are extracted and spliced into vectors:

$$\text{syntactic} = \text{Concat}(H_4, H_5, \dots, H_{10}) \quad (7)$$

First, weight learning is performed on the syntactic vectors with the help of attention mechanisms:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where $Q, K, V = \text{syntactic}$ (8)

Since the syntactic vector contains output vectors at different levels of RoBERTa, a single attention mechanism cannot focus on the information of multiple location representation sub-spaces [29]; therefore, multiple head attention is used to learn the weights of the syntactic vector:

$$\text{MultiHead}(\text{syntactic}, \text{syntactic}, \text{syntactic}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O$$

where $\text{head}_i = \text{Attention}(\text{syntactic}, \text{syntactic}, \text{syntactic})$ (9)

To prevent gradient explosion and gradient disappearance due to the depth model, a residual connection is made between *syntactic* and *MultiHead*:

$$sy' = \text{syntactic} + \text{MultiHead}(\text{syntactic}, \text{syntactic}, \text{syntactic}) \quad (10)$$

Layer normalization of sy' , calculating the mean and variance for each sample normalizes the hidden layers in the neural network to a standard normal distribution and accelerates the convergence to

$$\text{LN}(sy') = \alpha \times \frac{sy' - \mu_L}{\sqrt{\sigma_L^2 + \epsilon}} + \beta$$

where $\mu_L = \frac{1}{m} \sum_{i=1}^m sy'_i$

where $\sigma_L^2 = \frac{1}{m} \sum_{i=1}^m (sy'_i - \mu_L)^2$ (11)

where the scaling parameters α and β are learnable parameters. ϵ is a very small value that prevents the equation from dividing by a zero value, and m is the number of neurons.

Then, the output of the layer normalization is passed through the feed forward neural network by

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (12)$$

The above equation consists of two linear transformations with a ReLU activation in the middle and x denoting the output of layer normalization $LN(sy')$.

Finally, after residual connection and layer normalization:

$$Att = FFN(x) + x \quad (13)$$

The above is the specific process of syntactic information extraction, and Att denotes the vector of syntactic information learned by weighting.

4.5. Feature Fusion

Deep contextual semantic information and syntactic information are acquired using a parallel mechanism to fuse the two types of information:

$$F = h_c + h_a$$

$$\text{where } h_c = \text{sigmoid}(W_1 \cdot \text{context} + b_1)$$

$$\text{where } h_a = \text{sigmoid}(W_2 \cdot \text{Att} + b_2)$$

$$\text{where } W_1 + W_2 = 1 \quad (14)$$

where context is the deep contextual semantic information obtained in Section 4.3; Att is the syntactic information obtained in Section 4.4; and F denotes the fused vector. We set two learnable weight parameters W_1 and W_2 with initial values of 0.5, and the sum of the two numbers is restricted to 1. The two parameters W_1 and W_2 are learned by the neural network and optimized in the direction of minimizing the loss function in the model training.

Once the layer classification is fully connected, the span and class of predicted entities are obtained as follows:

$$p_t = \sum_{l=1}^g FW + b \quad (15)$$

where F is the above fused features, W is the training parameter, and b is the bias parameter.

As shown in Table 1, the number of different categories in the named entity identification dataset in the power dispatch domain varies widely, the sample categories are unbalanced, and the predicted entities are passed through the focal loss function to obtain the loss values:

$$FL(p_t) = -\alpha(1 - p_t)^\gamma \log(p_t) \quad (16)$$

Here, α is set to 0.25 and γ is set to 2. The larger the parameter p_t , the more accurate the sample classification. Conversely, the closer the $(1 - p_t)$ is to 0, the more accurate the classification is. With the feature of the focal loss function, the name and other categories that have a smaller number of samples take up more weight in the loss function. It makes the model focus more on these two classes of samples and alleviates the sample imbalance problem.

5. Results

The experimental environment uses a Pytorch framework, CUDA version 11.1, Ubuntu system, and an NVIDIA RTX3090 (24 G) graphics card. The length K of the candidate entities is set to 12. A learning rate warm-up strategy is used, where the learning rate increases linearly from 0 to 2×10^{-5} , and then decreases linearly to 0 to prevent model instability caused by too large a learning rate in the initial stage of training and to make the model converge faster. The

model is evaluated every 6000 training steps, and the stage model with high accuracy is saved. The syntactic information is extracted using a six-layer Transformer Encoder, and the number of attention heads is eight. Other parameter settings are detailed in Table 3, and the RoBERTa-wwm-ext pre-training model uses the original parameters.

Table 3. Parameter setting.

Parameter	Value
Learn rate	2×10^{-5}
Batch size	3
Epoch	50
lstm_embedding_size	1024
Hidden size	768
BERT model	RoBERTa-wwm-ext
Embedding size	512
Optimizer	AdamW

5.1. Evaluation Indicators

This experiment uses the criteria of *accuracy*, *recall*, and *F1-score* to measure the performance of the model, and the specific evaluation formula is as follows:

$$precision = correct_{num} / predict_{num}$$

$$recall = correct_{num} / golden_{num}$$

$$F1-score = \frac{2 * precision * recall}{precision + recall} \quad (17)$$

where $correct_{num}$ denotes the number of correct predictions, $predict_{num}$ denotes the total number of predictions, and $golden_{num}$ denotes the number of labeled entities. The *F1-score* is the summed average of precision and recall that can balance the effects of precision and recall, reflecting the performance of the model more comprehensively.

5.2. Results and Analysis

The named entity recognition dataset in the power dispatch domain constructed in Section 1 is used for the training evaluation, F1-score, F1_overlap, accuracy, and recall. These are used as the judging criteria to measure the performance of the model.

Table 4 shows the experimental results for different models. Where model 1 is the baseline model. Model 2 replaces the loss function of the baseline model with a CRF, and the F1-score is improved by 0.8%. Model 3 uses the BERT and BiLSTM (bi-directional long and short time recurrent neural network) models, which are then input into a CRF, where BiLSTM can encode the output values of BERT with bi-directional context and alleviate the bi-directional long-distance dependence problem. The F1-score is improved by 1.8% compared with the baseline model. The conventional model is used for splicing, on the basis of the baseline model, which has a certain improvement on the F1-score.

Model 4, compared with the first three models, no longer uses modular splicing, but extracts syntactic information in parallel with information gain and deep contextual semantic information to span the boundary enhancement by BiGRU (bidirectional gated recurrent neural network), which is a variant of LSTM with a simpler network structure compared to LSTM. The syntactic information is extracted by BiGRU, which improves by 1.21% compared to the baseline model F1-score.

Table 4. Model comparison experiments (“F1_overlap” represents nested entity F1-score; “CE” represents cross-entropy loss; “FL” represents focus loss function, models 1–3 use serial splicing of different modules; models 4 and 5 extract syntactic information and deep contextual semantic information in parallel for fusion).

Index	Model	Precision/%	Recall/%	F1_Overlap/%	F1-Score/%
1	BERT-CE	89.01	83.29	86.15	86.07
2	BERT-CRF	89.98	83.98	86.11	86.87
3	BERT-BiLSTM-CRF	90.01	85.85	86.14	87.87
4	BERT -BiGRU-FL	89.71	84.99	86.38	87.28
5	RoBERTa-Attention-FL	92.53	88.12	89.37	90.35

Model 5 uses the RoBERTa model, replaces the BiGRU module of model 4 with the Encoder encoder based on model 4, and extracts the syntactic information with the help of attention, and the F1-score is improved by 3.07%, demonstrating that the Encoder can focus on more effective information than the BiGRU module and thus improve the accuracy. Compared with model 3, the F1-score is improved by 2.48%. Models 1–3 are conventional models that are serially connected by different modules, and model 5 uses a parallel approach and feature fusion to achieve better results than the conventional model. The syntactic information and the encoder extraction mechanism are proven to be effective in improving the F1-score of the model. The F1-score is improved by 4.28% compared to the baseline model. Through the above comparative experimental analysis, it is proven that the RoBERTa-attention-FL model can effectively improve the accuracy of the named entity recognition.

The F1_overlap column in Table 4 represents the F1-score of the nested entity recognition, and model 5 is improved by 3.22% compared with baseline model 1, which effectively improves the accuracy of the nested entity recognition.

Table 5 shows the results of the ablation experiments with the RoBERTa-attention-FL model. Model 2 uses the BERT model, and the F1-score is reduced by 1.6%. Model 3 replaces the focus loss function of model 2 with the cross-entropy loss function, and the F1-score is reduced by 0.1%, proving that the focus loss function can alleviate the sample imbalance problem. Model 4 removes the syntactic information fusion of model 2, and the F1-score decreases by 2.62%, indicating that the syntactic information can effectively improve the recognition accuracy of the model. Model 5 replaces the focal loss function of model 4 with cross-entropy loss, and the F1-score is reduced by 0.06%.

Table 5. Ablation experiments.

Index	RoBERTa	BERT	Attention	FL	CE	F1_Overlap/%	F1-Score/%
1	✓		✓	✓		89.37	90.35
2		✓	✓	✓		87.77	88.75
3		✓	✓		✓	87.42	88.65
4		✓		✓		86.27	86.13
5		✓			✓	86.15	86.07

This symbol ‘✓’ indicates that the module is used.

Table 6 shows the performance of the model with different layers selected to construct the syntactic information, and it can be seen from the table that the best F1-score can be obtained by selecting layers four to ten.

Table 6. Extracting the effect of different levels of experiments, where “4–7” in the first column means extracting the RoBERTa output values from the 4th to the 7th level to construct the syntactic information vector.

Number of Layers	Precision/%	Recall/%	F1_Overlap/%	F1-Score/%
4–7	92.04	88.23	89.19	90.02
4–8	91.95	88.27	89.31	90.14
4–9	92.09	87.91	89.20	90.03
4–10	92.53	88.12	89.37	90.35
4–11	92.12	88.20	89.38	90.19

The meaning of the *Recognition effect* column in the Table 7 is “28 July 2022 16:49 Furong Substation 10 kV Fucha line 913 switch tripped”. Table 7 shows the entity recognition by different models for the same example, which contains the nested entity “10 kV 芙蓉线” (10 kV Fucha line) and several flat entities. The words with different background colors indicate the entities recognized by the model, and the red font indicates a continuous entity. Model 4 and Model 5, which introduce the syntactic information and deep contextual semantics of BERT, can accurately identify the example, especially for the nested entities, “10 kV”, “芙蓉线” (Fucha Line) and “10 kV 芙蓉线” (10 kV Fucha Line) can be identified at the same time. Grasp both fine-grained and coarse-grained entities at the same time. Models 1-3 have different degrees of entity recognition deficiencies.

Table 7. Recognition effect of different models.

Index	Model	Recognition Effect							
1	BERT-CE	28 July 2022	16:49	芙蓉	变电站	10 kV	芙蓉线	913	开关跳闸
2	BERT-CRF	28 July 2022	16:49	芙蓉	变电站	10 kV	芙蓉线	913	开关跳闸
3	BERT-BiLSTM-CRF	28 July 2022	16:49	芙蓉变电站	10 kV	芙蓉线	913	开关跳闸	
4	Bert-BiGRU-FL	28 July 2022	16:49	芙蓉变电站	10 kV	芙蓉线	913	开关跳闸	
5	RoBERTa-Attention-FL	28 July 2022	16:49	芙蓉变电站	10 kV	芙蓉线	913	开关跳闸	

Overall, the RoBERTa-Attention-FL model constructed in this paper can achieve a better F1-score on the named entity recognition dataset in the power dispatch domain compared to the baseline model.

6. Conclusions

In this study, we designed a RoBERTa-Attention-FL model for named entity recognition in the field of power dispatching. The RoBERTa pre-trained model was used to obtain textual contextual depth representation information. Extraction of the output of the middle four to ten layers of RoBERTa obtained the syntactic information by automatically learning the weights with the help of the multi-headed self-attentiveness mechanism in Transformer Encoder. By fusing the context-depth representation information and syntactic information, the power dispatching domain text was characterized in a deeper way, and finally the predicted entities were obtained from the fully connected layer. A focal loss function was used during training to alleviate the problem of sample category imbalance. In this paper, we constructed a named entity recognition dataset in the grid dispatching domain, and used an annotation method based on span representation to recognize nine entity types. The model can effectively solve the recognition problem of nested entities and improve the recognition accuracy of flat entities. We validated the effectiveness and superiority of the method on a self-built named entity recognition dataset in the field of grid dispatching:

(1) RoBERTa-Attention-FL model is introduced into the RoBERTa pre-training model, and the F1-score is improved by 1.6% compared with the BERT model, indicating that the RoBERTa pre-training model can effectively improve the accuracy rate of named entity recognition;

(2) Introducing mid-level syntactic information and deep contextual semantic information after the Transformer Encoder module to learn the weight information, Bert-Attention-FL model improves F1-score by 2.62% compared with Bert-FL model, which can effectively enhance the depth characterization ability of the model and improve the accuracy rate of named entity recognition.

(3) The RoBERTa-Attention-FL model improves the recognition F1-score of nested entities by 3.22% compared with the benchmark model BERT-CE model, which effectively improves the recognition accuracy of nested entities and the F1-score of flat entities by 4.28%.

In future research, we plan to further extract relationships between entities and study the correlation between relationship extraction and named entity identification to better tap potential knowledge in the field of grid dispatching and further improve the model performance. In addition, we plan to apply the model to other domains to verify the generalization capability of the model.

Author Contributions: Methodology, D.L.; writing—original draft preparation, D.L. and Y.C.; writing—review and editing, D.L.; visualization, Y.C.; software, D.L.; resources, Q.M.; data curation, D.L., Z.L. and Z.T.; supervision, Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Guangxi Scientific Research and Technology Development Plan Project (Grant No. AA20302002-3) and Innovation Project of China Southern Power Grid Co., Ltd. (202201161).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dileep, G. A survey on smart grid technologies and applications. *Renew. Energy* **2020**, *146*, 2589–2625. [[CrossRef](#)]
2. Yin, L.; Gao, Q.; Zhao, L.; Zhang, B.; Wang, T.; Li, S.; Liu, H. A review of machine learning for new generation smart dispatch in power systems. *Eng. Appl. Artif. Intell.* **2020**, *88*, 103372. [[CrossRef](#)]
3. Syed, D.; Zainab, A.; Ghayeb, A.; Refaat, S.S.; Abu-Rub, H.; Bouhali, O. Smart Grid Big Data Analytics: Survey of Technologies, Techniques, and Applications. *IEEE Access* **2020**, *9*, 59564–59585. [[CrossRef](#)]
4. Xu, H.; Fan, G.; Kuang, G.; Wang, C. Exploring the Potential of BERT-BiLSTM-CRF and the Attention Mechanism in Building a Tourism Knowledge Graph. *Electronics* **2023**, *12*, 1010. [[CrossRef](#)]
5. Wu, J.; Xu, X.; Liao, X.; Li, Z.; Zhang, S.; Huang, Y. Intelligent Diagnosis Method of Data Center Precision Air Conditioning Fault Based on Knowledge Graph. *Electronics* **2023**, *12*, 498. [[CrossRef](#)]
6. Žukov-Gregorič, A.; Bachrach, Y.; Coope, S. Named Entity Recognition with Parallel Recurrent Neural Networks. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Melbourne, Australia, 15–20 July 2018; Association for Computational Linguistics: Toronto, ON, Canada, 2018; pp. 69–74.
7. Katiyar, A.; Cardie, C. Nested Named Entity Recognition Revisited. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), New Orleans, LA, USA, 1–6 June 2018; Association for Computational Linguistics: Toronto, ON, Canada, 2018; pp. 861–871.
8. Ju, M.; Miwa, M.; Ananiadou, S. 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), New Orleans, LA, USA, 1–6 June 2018; Association for Computational Linguistics: Toronto, ON, Canada, 2018; pp. 1446–1459.
9. Fisher, J.; Vlachos, A. Merge and Label: A Novel Neural Network Architecture for Nested NER. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; Association for Computational Linguistics: Toronto, ON, Canada, 2019; pp. 5840–5850.
10. Liu, P.; Guo, Y.; Wang, F.; Li, G. Chinese named entity recognition: The state of the art. *Neurocomputing* **2022**, *473*, 37–53. [[CrossRef](#)]
11. Wu, F.; Liu, J.; Wu, C.; Huang, Y.; Xie, X. Neural Chinese Named Entity Recognition via CNN-LSTM-CRF and Joint Training with Word Segmentation. In *The World Wide Web Conference (WWW '19)*; Association for Computing Machinery: New York, NY, USA, 2019; pp. 3342–3348.
12. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2999–3007.
13. Li, J.; Sun, A.; Han, J.; Li, C. A Survey on Deep Learning for Named Entity Recognition. *IEEE Trans. Knowl. Data Eng.* **2022**, *34*, 50–70. [[CrossRef](#)]
14. Zhang, S.; Elhadad, N. Unsupervised biomedical named entity recognition: Experiments with clinical and bio-logical texts. *J. Biomed. Inform.* **2013**, *46*, 1088–1098. [[CrossRef](#)] [[PubMed](#)]
15. Patil, N.; Patil, A.; Pawar, B.V. Named Entity Recognition using Conditional Random Fields. *Procedia Comput. Sci.* **2020**, *167*, 1181–1188. [[CrossRef](#)]

16. Srivastava, S.; Paul, B.; Gupta, D. Study of Word Embeddings for Enhanced Cyber Security Named Entity Recognition. *Procedia Comput. Sci.* **2023**, *218*, 449–460. [CrossRef]
17. Zhang, H.; Wang, X.; Liu, J.; Zhang, L.; Ji, L. Chinese named entity recognition method for the finance domain based on enhanced features and pretrained language models. *Inf. Sci.* **2023**, *625*, 385–400. [CrossRef]
18. Puccetti, G.; Giordano, V.; Spada, I.; Chiarello, F.; Fantoni, G. Technology identification from patent texts: A novel named entity recognition method. *Technol. Forecast. Soc. Chang.* **2023**, *186*, 122160. [CrossRef]
19. Li, X.; Zhang, H.; Zhou, X.-H. Chinese clinical named entity recognition with variant neural structures based on BERT methods. *J. Biomed. Inform.* **2020**, *107*, 103422. [CrossRef] [PubMed]
20. Zheng, K.; Sun, L.; Wang, X.; Zhou, S.; Li, H.; Li, S.; Zeng, L.; Gong, Q. Named Entity Recognition in Electric Power Metering Domain Based on Attention Mechanism. *IEEE Access* **2021**, *9*, 152564–152573. [CrossRef]
21. He, L.; Zhang, X.; Li, Z.; Xiao, P.; Wei, Z.; Cheng, X.; Qu, S. A Chinese Named Entity Recognition Model of Maintenance Records for Power Primary Equipment Based on Progressive Multitype Feature Fusion. *Complexity* **2022**, *2022*, 8114217. [CrossRef]
22. Weiyue, T. Named entity recognition of power communication planning based on transformer. In Proceedings of the IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 17–19 June 2022; pp. 588–592.
23. Jawahar, G.; Sagot, B.; Seddah, D. What Does BERT Learn about the Structure of Language? In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; Association for Computational Linguistics: Toronto, ON, Canada, 2019; pp. 3651–3657.
24. Wang, Y.; Yu, B.; Zhu, H.; Liu, T.; Yu, N.; Sun, L. Discontinuous Named Entity Recognition as Maximal Clique Discovery. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Online; Association for Computational Linguistics: Toronto, ON, Canada, 2021; pp. 764–774. Available online: <https://aclanthology.org/2021.acl-long.63> (accessed on 8 June 2023).
25. Geng, R.; Chen, Y.; Huang, R.; Qin, Y.; Zheng, Q. Planarized sentence representation for nested named entity recognition. *Inf. Process. Manag.* **2023**, *60*, 103352. [CrossRef]
26. Zhong, Z.; Chen, D. A Frustratingly Easy Approach for Entity and Relation Extraction. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Online, 6–11 June 2021; pp. 50–61.
27. Muresan, S.; Nakov, P.; Villavicencio, A. Packed Levitated Marker for Entity and Relation Extraction. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Dublin, Ireland, 22–27 May 2022; Association for Computational Linguistics: Toronto, ON, Canada, 2022; pp. 4904–4917.
28. Burstein, J.; Doran, C.; Solorio, T. 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, Minneapolis, MN, USA, 2–7 June 2019; Association for Computational Linguistics: Toronto, ON, Canada, 2019; Volume 1, pp. 4171–4186.
29. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Long Beach, CA, USA, 4–9 December 2017; Curran Associates Inc.: Red Hook, NY, USA, 2017; pp. 6000–6010.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.