Article

RoUIE: A Method for Constructing Knowledge Graph of Power Equipment Based on Improved Universal Information Extraction

Zhenhao Ye [1] , Donglian Qi [1,*], Hanlin Liu [2], Yunfeng Yan [1], Qihao Chen [3] and Xiayu Liu [1]

1 College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China
2 School of Computer Science, Guangdong University of Technology, Guangzhou 510006, China
3 Faculty of Electrical and Control Engineering, Liaoning Technical University, Huludao 125105, China
* Correspondence: qidl@zju.edu.cn

Abstract: The current state evaluation of power equipment often focuses solely on changes in electrical quantities while neglecting basic equipment information as well as textual information such as system alerts, operation records, and defect records. Constructing a device-centric knowledge graph by extracting information from multiple sources related to power equipment is a valuable approach to enhance the intelligence level of asset management. Through the collection of pertinent authentic datasets, we have established a dataset for the state evaluation of power equipment, encompassing 35 types of relationships. To better suit the characteristics of concentrated relationship representations and varying lengths in textual descriptions, we propose a generative model called RoUIE, which is a method for constructing a knowledge graph of power equipment based on improved Universal Information Extraction (UIE). This model first utilizes a pre-trained language model based on rotational position encoding as the text encoder in the fine-tuning stage. Subsequently, we innovatively leverage the Distribution Focal Loss (DFL) to replace Binary Cross-Entropy Loss (BCE) as the loss function, further enhancing the model's extraction performance. The experimental results demonstrate that compared to the UIE model and mainstream joint extraction benchmark models, RoUIE exhibits superior performance on the dataset we constructed. On a general Chinese dataset, the proposed model also outperforms baseline models, showcasing the model's universal applicability.

Keywords: state evaluation; power equipment; universal information extraction; knowledge graph construction

1. Introduction

With the continuous expansion of the scale of power equipment and the increasing operational requirements, carrying out intelligent operation and maintenance of power equipment has become an important development trend in the field of power production. In recent years, many scholars have proposed using methods such as hierarchical evaluation [1,2], association analysis [3], the entropy weight method [4], etc., to reveal the relationship between operational indicators of power equipment and equipment health for equipment state evaluation. The primary challenges in the current methods for assessing the health and performance of electrical equipment can be identified in two aspects. Firstly, scholars often tend to place excessive emphasis on the electrical performance aspects when evaluating the overall condition of equipment. While real-time monitoring data have been considered, there has been a notable absence of comprehensive attention paid to the equipment state evaluation guidelines. The guidelines should not only focus on electrical data such as Dissolved Gas Analysis (DGA) in transformers, but also encompass pre-commissioning, operational, maintenance, and testing information [5]. Secondly, the methods proposed by the aforementioned scholars involve analyzing and calculating to directly obtain equipment health scores without adhering to specific guidelines purposed by enterprises. Power grid enterprises have established numerous information systems in

the field of power production and have implemented a state evaluation process based on guidelines within the system. The guidelines encompass a range of text-based criteria for deducting scores, such as equipment categories, components, status indicators, deduction content, and corresponding deduction values. The current process for conducting state evaluations within the enterprise involves initiating evaluation tasks triggered by defect reports generated by other modules of the system, followed by manual corrections by operations personnel in accordance with the guidelines. Alternatively, operations personnel manually correlate device anomalies with deduction criteria and subsequently input the deduction details into the system. Many academic models developed by researchers primarily provide scores and status categories without clearly identifying the specific defective components and associated criteria. Consequently, maintenance staff face challenges in efficiently performing system-based state evaluation, including recording and defect tracking procedures. To gain a comprehensive understanding of equipment status, individuals often need to navigate through multiple systems to collect fragmented data. Therefore, acquiring information from various systems and integrating device-related multi-source information on a data platform to make it convenient for equipment management personnel to query and analyze common characteristics of devices is a pressing issue that needs to be addressed in order to carry out dynamic state evaluation and enhance the efficiency of power equipment management.

Knowledge graph, as a technology for data storage, updating, and processing, can visualize the complex information and relationships contained in the massive operational data of power equipment [6]. Constructing a knowledge graph is a valuable approach for creating the aforementioned data platform, which will establish the groundwork for conducting a comprehensive state evaluation. Specifically, when a system generates abnormal information and transfers it to the data platform, the platform can quickly identify the corresponding equipment and abnormal description by adopting information extraction methods. Subsequently, by employing text matching methods to match the abnormal description with the deduction criteria from guidelines, adjustments can be made to the health score of the specific equipment. This article focuses on the former, which is the method of constructing a knowledge graph based on information extraction. Some scholars have proposed constructing a knowledge graph of power equipment in the field of power production [7], but there has not yet been effective extraction of multi-source information related to equipment condition assessment. According to the construction of the information system in the field of power production, the evaluation of the current equipment state mainly relies on three types of data: first, the equipment inventory information, also known as basic information; second, the monitoring information generated by the equipment's online monitoring system; and third, records of equipment testing, maintenance operations, corrective action implementation, defect reports, accident event reports, and other management information. To achieve unified maintenance, it is essential to utilize information extraction technology to extract the three types of information separately and form valuable triplets. Additionally, it is crucial to link the extracted information with a unified inventory database. Due to the fact that various types of online monitoring systems currently maintain their databases independently, the abnormal monitoring information generated is often manually combined. The formats of abnormal alarms for the same device may vary among different systems. This article combines expert experience to propose a standardized short-text format for monitoring information on power primary equipment, which is applicable to monitoring information on all power primary equipment. Furthermore, we collect a series of real-text corpora, such as a power equipment inventory, and monitoring and management information, and select those related to state evaluation for annotation to create a dataset for evaluating the state of power equipment. For a specific device, we summarize and organize the relationships between its related labeling information, and use a combination of top-down and bottom-up approaches to construct the ontology layer, based on which we build a knowledge graph centered around devices. The graph helps operations and maintenance personnel to have a more comprehensive understanding of

operating conditions for specific equipment. It also facilitates the subsequent association and dynamic updating of equipment information from different systems. This is also a prerequisite for carrying out tasks such as equipment state evaluation, equipment reliability and life analysis, and common-mode fault diagnosis in the next step. Adopting information extraction technology to extract triplets from text is the foundation for constructing large-scale knowledge graphs. Information extraction typically involves two sub-tasks, namely Named Entity Recognition and Relation Extraction [8]. The construction of a knowledge graph for power equipment is still in its early stages. The most mainstream technique in this research field is the use of pipeline extraction methods [9,10]. Typically, entities are first identified in sentences, followed by the classification of relationships between entities, thus forming triplets of subject, relationship, and object (SRO). Recently, some scholars have referred to the concept of large language models and proposed generative information extraction universal models such as UIE [11], USM [12], rexUIE [13], lasUIE [14], etc., achieving state-of-the-art results in a series of rankings. However, most of the aforementioned models are trained and evaluated using general corpora.

To better adapt to the characteristics of text for evaluating the state of power equipment, this article proposes an improved UIE model named RoUIE. The RoFormer pre-training language model, based on the enhanced Transformer model with rotary position embedding [15] is utilized as the text encoder in the fine-tuning stage of the model. In addition, inspired by the Generalized Focal Loss [16] in the field of Computer Vision (CV), we replaced the Binary Cross-Entropy Loss (BCE) of the UIE model with Distribution Focal Loss (DFL) as the loss function. This change aims to measure the deviation in predicting the starting and ending positions of entities in text from their true labels in the text. Upon comparing the proposed method with the baseline model for information extraction, it has been confirmed that our method demonstrates enhanced adaptability and superior performance on both the power equipment condition evaluation dataset and a general Chinese dataset for information extraction. We also conduct ablation experiments to confirm the effectiveness of various improvements. The contributions of this paper can be summarized as follows:

- By collecting authentic data from three categories, including inventory information, monitoring information, and management information, we have constructed a dataset for the state evaluation of power equipment. Specifically, we have introduced a standardized short-text format for monitoring information that is applicable across various types of primary power equipment.
- Based on the characteristics of the three categories of information mentioned above, we have developed an ontology layer and proposed the construction of a device-centric knowledge graph, thereby addressing the current research gap in multi-source information fusion for power equipment state evaluation and assisting in resolving operation and maintenance challenges on site.
- In order to better accommodate the concentrated relationship representations and varying lengths of device state evaluation texts, this paper proposes an improved UIE model called RoUIE, which utilizes a RoFormer pre-trained language model as the text encoder during the fine-tuning stage. Additionally, the Distribution Focal Loss is employed to replace Binary Cross-Entropy Loss as the loss function, further enhancing the extraction performance of the model.
- By comparing our proposed model with the UIE model and other mainstream joint information extraction baseline models, we demonstrate its superiority and general applicability on both a Chinese general dataset and a specific dataset for power equipment state evaluation.

2. Related Work

2.1. Current Methods of Information Extraction

Contemporary scholars' research on information extraction can be categorized into extractive models and generative models, with extractive models further divided into

pipeline methods and joint methods. Although pipeline methods offer advantages in flexibility and independence, they are associated with several challenges: (1) accumulation of errors; (2) lack of information exchange between sub-tasks; (3) redundancy in named entity recognition; (4) difficulty in identifying long-range dependencies between entities; and (5) multiple overlapping relationships [17].

The joint learning methods simultaneously address the tasks of named entity recognition and entity pair relationship classification, and have shown research advancements and practical applications in various domains, such as finance [18] and healthcare [19], in recent years. Due to the consideration of information interaction between two sub-tasks, such methods have significantly enhanced the effectiveness of entity relation extraction. There are three main modeling approaches, with the multi-module multi-step modeling method [20–22] being more commonly used in recent years. This method utilizes different sub-modules to extract relatively rich feature information, but it may lead to insufficient information interaction. Wei et al. [23] introduced a Casrel model that addresses the issue of multiple entity relationships and overlapping relationship triplets within a single sentence by establishing a Hierarchical Binary Tagging framework. Another common approach is the multi-module single-step modeling method [24–26], which addresses the issue of insufficient interaction of entity relationship information by establishing a joint decoder. One typical model is TPLinker [27], which considers entity relation joint extraction as a token pair linking problem. However, this model employs a complex joint decoding algorithm, leading to inefficiencies in extracting triplets. The third approach involves a single-module, single-step method. For instance, the OneRel model proposed by Shang et al. [28] constructs a triplet classification model at the token level. It directly extracts triplets from textual sentences to reduce redundant errors and address issues present in pipeline methods. However, this approach may lead to conflicts between features of entity relationships.

In the past, traditional Seq2Seq frameworks based on RNN for generative models did not exhibit significant advantages in terms of accuracy and efficiency compared to extractive models. It was not until the recent widespread adoption of generative pre-trained models such as UniLM [29], BART [30], T5 [31], and GPT [32] that the development of effective generative information extraction models has gradually emerged as a forefront research direction. Extractive models are more susceptible to schema limitations, while generative models exhibit greater strength in terms of transferability and scalability compared to extractive models. Additionally, generative models make it possible to achieve unified information extraction tasks across various scenarios, tasks, and schemas. Lu et al. [11] introduced the UIE model in 2022, which consists of three main components: (1) text-to-structure for generating architecture, (2) a structure-constrained decoding mechanism based on prompts, and (3) a large pre-trained information extraction model. Serving as a general model for information extraction, the structure of UIE is determined by the extraction structure and demand schema, making it applicable to tasks such as entity extraction, relation extraction, and event extraction. The RoUIE model proposed in this paper represents an improvement over the generative model UIE during the fine-tuning phase of model training.

2.2. Characteristics of Texts Related to State Evaluation

Although the current production management systems have established equipment inventory information, many other state monitoring systems or manually written reports cannot accurately link to the existing inventory information. There are varying degrees of differences in inventory information for different equipment within the same enterprise. These differences are manifested in variations such as different hierarchies corresponding to the same type of equipment at different sites, as well as inconsistencies in descriptions for the same type of equipment at different sites. Drawing on the existing equipment inventory information in the production management system and expert knowledge, this paper employs text augmentation to handle dual numbering, a common practice in dispatch

departments, aiming to reduce unnecessary paths, and eliminate redundant information. This method aims to establish a unified representation of equipment inventory information.

The current monitoring information of equipment primarily originates from online monitoring systems that generate data. Due to variations in the operational teams of these systems, the alert formats for describing abnormal conditions of the same equipment differ. In addition, the current equipment management data primarily comes from various modules of the production management system, constituting semi-structured data. This includes operational records such as installation, acceptance, routine inspections, tests, and periodic maintenance, as well as reports on defects and accident investigations. The diverse nature of the tasks performed and the different personnel involved in each recording result in inconsistencies in the formats of operational records and significant variations in descriptive content.

The relationship among inventory information, monitoring information, and management information is illustrated in Figure 1. By considering the intersection of these three elements, a unified short-text format for power equipment monitoring information is proposed. The text structure is denoted as date and time of action or return, substation, bay, equipment, component, and description of abnormal information, as shown in the example provided in Table 1. This format is utilized to standardize the textual representation of monitoring and management information generated by systems in the production field. It is applicable to various primary equipment such as transformers, circuit breakers, disconnecting switches, earthing disconnectors, and surge arresters. The format is similar to mainstream monitoring information formats, with the initial part of sentences presenting relatively concentrated basic equipment information. The latter part mainly describes abnormal equipment conditions or operational records, which exhibit a strong colloquial and varied length characteristic. The text, after being processed into a standardized format, will be included as labeled text in the dataset for the state evaluation of power equipment.
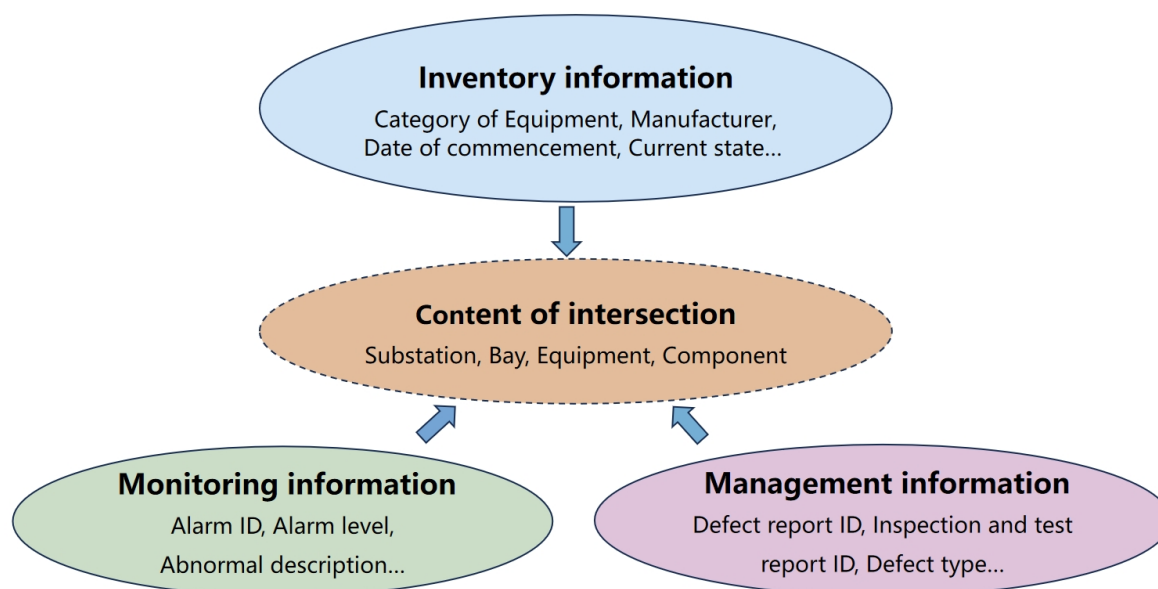
Figure 1. Relationship among inventory information, monitoring information, and management information.

Table 1. An illustration of the unified short-text format for power equipment monitoring information.

Unified Short Text Format	Date and Time of Action	Substation	Bar	Equipment	Component	Abnormal Description
Entities	18 December 2023 06:35:25	220 kV Test Substation	Test Line	2518	Hydraulic Mechanism	... not reached the interlocking threshold
Sample Text	18 December 2023 06:35:25, the pressure of the hydraulic mechanism of the 2518 circuit breaker of the Test Line of 220 kV Test Substation below the required level, yet it has not reached the interlocking threshold (in Chinese).					

3. Construction Method of a Device-Centric Knowledge Graph

This article focuses on the task of equipment status evaluation and proposes the construction of a device-centric knowledge graph ontology layer, as illustrated in Figure 2 below. Typically, a knowledge graph structure can be divided into a schema layer and a data layer. The schema layer, also known as the ontology layer, serves as a descriptive framework for entities and relationships. Based on the aforementioned three types of information, a combined approach of top-down and bottom-up methods is employed to delineate 35 types of relationships that directly impact dynamic status evaluation, forming the corresponding schema, as detailed in Appendix A. The subject type refers to a set of entities that need to be extracted and manually summarized. A sentence to be processed must necessarily contain a specific entity in its textual content, but it may not include the subject type. For instance, specific equipment names included in the subject type of "equipment" could be "transformer #1", "circuit breaker 2518", and "disconnecting switch 11011", and so on. Similarly, the object type represents a set of attributes to be extracted. Predicates appearing in the schema are relationships defined by us, which are represented in the ontology layer diagram by connecting lines between entities and attributes.
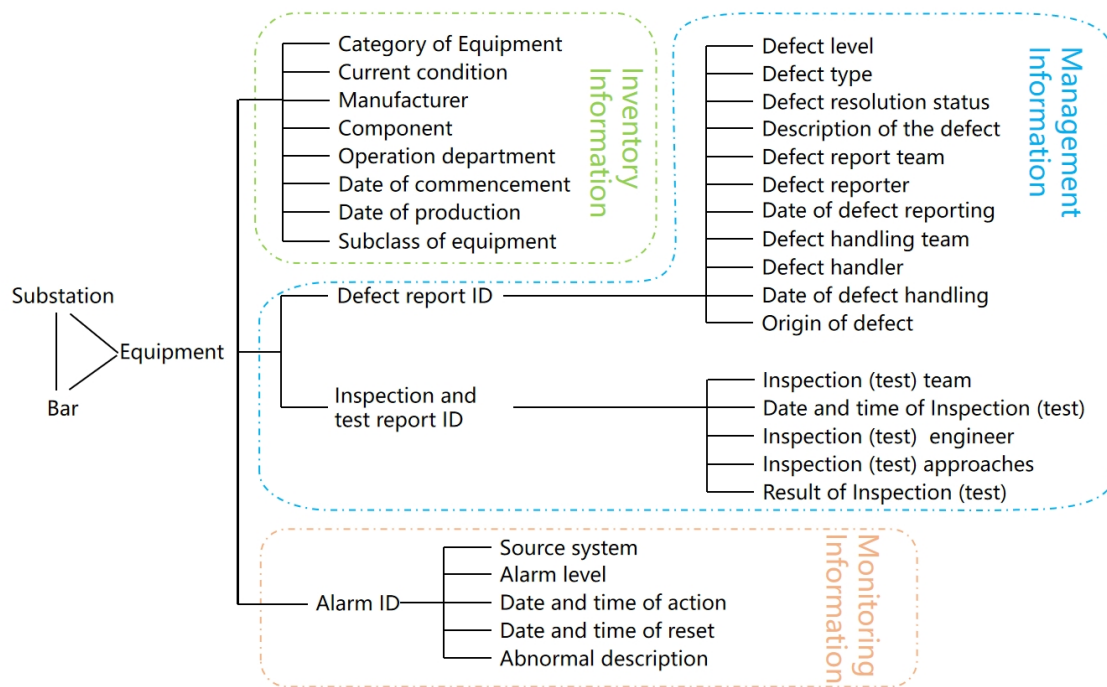
Figure 2. Ontology layer of the device-centric knowledge graph.

We have observed that some texts do not contain interval information; instead, they directly associate the equipment with the substation. Although deleting bay information, in most cases, one substation can correspond to a unique device, but in practice, it will appear messy and scattered after constructing the knowledge graph, which does not align with typical querying habits. This situation falls under Single-Entity-Overlap (SEO), which represents one of the scenarios of overlapping relations [33]. In such instances, the utilization of pipeline information extraction methods frequently does not yield satisfactory

outcomes. That is why we prefer joint extraction methods over pipeline methods for our baseline models.

4. Improved UIE Model with Rotary Position Embedding

4.1. Model Architecture

The core architecture of UIE is based on a transformer encoder–decoder framework. The authors have introduced a task output format known as Structural Extraction Language (SEL), which is jointly composed of the Structural Schema Instructor (SSI) and the text. Given the SSI and text as input, UIE first computes the hidden representation as the encoder state:

$$\mathbf{H} = \text{Encoder}\left(s_1, \ldots, s_{|s|}, x_1, \ldots, x_{|x|}\right) \tag{1}$$

where $\mathbf{H}$ denotes the hidden representation. At the step i of decoding, UIE generates the i-th token y_i in the linearized SEL sequence. The corresponding decoder state h_i^d is as follows:

$$y_i, \mathbf{h}_i^d = \text{Decoder}\left(\left[\mathbf{H}; \mathbf{h}_1^d, \ldots, \mathbf{h}_{i-1}^d\right]\right) \tag{2}$$

where the decoder predicts the condition probability $p(y_i \mid y_{<i}, x, s)$ of token y_i. The predicted SEL expression will then be converted into the extracted information record. In other words, the model converts the schema to be extracted into a string SSI, which is appended to the text to be extracted, forming SEL that is input together into the encoder. Utilizing a prompt-based target information guidance mechanism, the SSI transforms the schema information into prompts for the generation process, guiding the UIE model to locate the correct information from the original text. The model represents the category framework and text segments as a generating prefix tree. The generation process involves treating it as a search on the prefix tree to create a valid structural extraction expression, which enables the implementation of constrained decoding. Based on the generation process, a valid vocabulary is dynamically provided. Schema constraints ensure both structural and semantic validity while also limiting the decoding space and reducing decoding complexity.

The RoUIE model proposed in this paper is an improved UIE model with rotary position embedding, and its overall framework is illustrated in Figure 3. By constructing a series of schemas within the field of power production, we leverage knowledge extraction techniques to generate 35 types of triplets associated with equipment from input inventory information, monitoring information, and management information. Unlike the original model, RoFormer is adopted during the model fine-tuning phase as a pre-trained language model, so as to more effectively acquire semantic features. In addition, we leverage DFL instead of BCE as a loss function to further improve the precision of information extraction.

4.2. Rotary Position Embedding

Pre-trained language models typically leverage self-attention mechanisms to semantically capture contextual representations of a given corpus [34]. According to the explanation provided by the original author, the UIE base and UIE-large models are initialized using T5-v1.1-base and T5-v1.1-large models, respectively [11]. It is noted that while T5 also employs an Encoder–Decoder architecture, it does not append position embedding after the input embedding [31]. In contrast to the fine-tuning stage of the original UIE model, where the encoder utilizes the absolute position encoding method of the transformer module, RoUIE encodes the input text by RoFormer v2. The primary modifications in the v2 version compared to RoFormer include simplifying model parameters and structure, increasing training data, and incorporating supervised training. Despite these changes, the core concept remains unchanged. Therefore, the focus here will be on discussing the RoFormer model. The RoFormer model was introduced by Su et al. in 2021, who presented a novel positional embedding method called RoPE [15]. This method encodes absolute positions

using rotation matrices and incorporates explicit relative positional dependencies into the self-attention mechanism to enhance the performance of the transformer architecture. The utilization of rotation matrices to encode the position of words within a sentence takes into account both the absolute and relative positions of the words. This method offers a dynamic approach to represent positional relationships, enabling models to more accurately comprehend word order and dependencies between words.
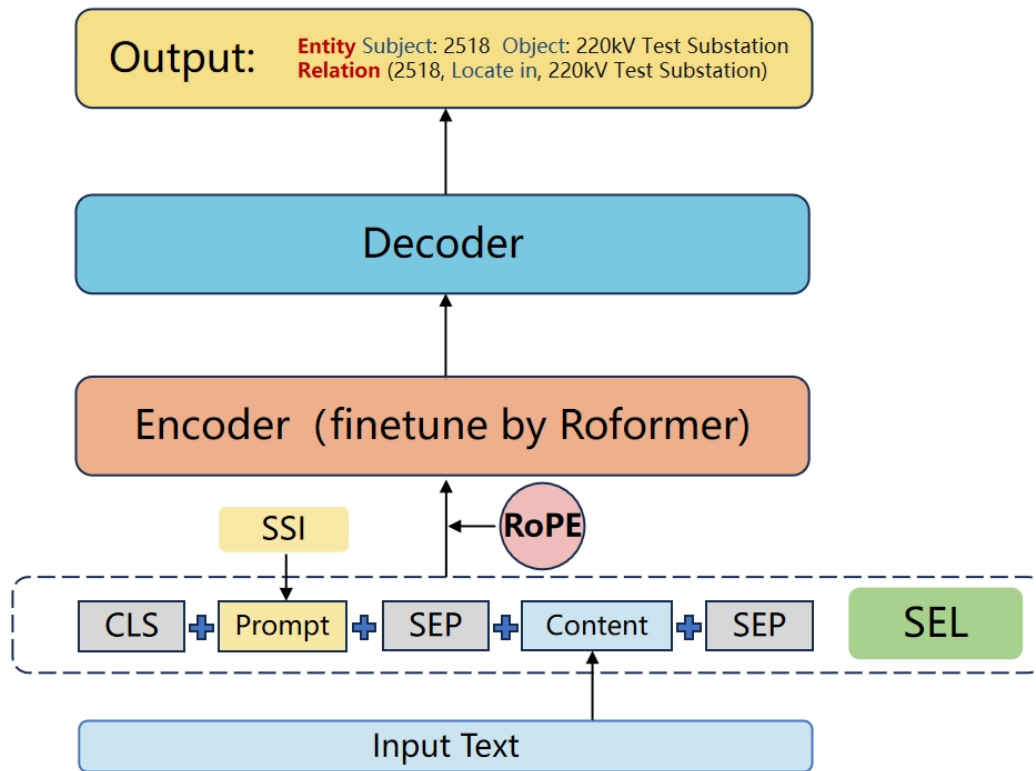
Figure 3. Model architecture of RoUIE.

The rationale behind this approach is determined by the textual characteristics of monitoring and management information in the field of power production. Such texts often exhibit a concentration of relatively fixed entity information at the beginning, with longer descriptions of anomalies provided towards the end. Due to the lengthy description in the latter part of the text, it may introduce noise to the preceding entity information. RoFormer exhibits favorable long-term decay characteristics, which lead to a reduction in the dependency relationships between tokens with relatively longer distances. This enables the attention model to focus more on the weights between the first few entities with shorter relative distances, facilitating faster model convergence. Furthermore, whether it is monitoring or managing information, the lengths of their corresponding abnormal description information are generally inconsistent, with the total length of the original samples may potentially exceed the maximum input length of the original UIE model. RoFormer encodes positional information by rotation matrices, demonstrating flexibility in accommodating text inputs of varying lengths without the need for adjusting encoding strategies. This feature enables its application in managing information texts, such as patrol operation or fault elimination operation records, that exceed the maximum input length of the original UIE model.

Within the mechanism of the Transformer model, in order to incorporate relative positional information, a function g is required to represent the inner product $q_m^T k_n$ between the query q and the key k. The authors aim for the inner product function encoding spatial information to manifest in the following form:

$$\langle f_q(x_m, m), f_k(x_n, n) \rangle = g(x_m, x_n, m - n) \tag{3}$$

where x_m and x_n denote the word embedding information to be processed, and their relative position is $m - n$. The authors propose that word embeddings (like x_m and x_n) are rotated using rotation matrices based on the positions of the words in the sequence (m and n), thereby encoding relative positional information. Consider a basic scenario with a dimension of $d = 2$ as an illustration.

$$f_q(x_m, m) = (W_q x_m) e^{im\theta}$$
$$f_k(x_n, n) = (W_k x_n) e^{in\theta} \tag{4}$$
$$g(x_m, x_n, m - n) = \text{Re}\left[(W_q x_m)(W_k x_n)^* e^{i(m-n)\theta} \right]$$

where $\text{Re}[\cdot]$ represents the real part of a complex number, and $(W_k x_n)^*$ denotes the conjugate complex number of $(W_k x_n)$. Symbol θ is defined as a preset nonzero constant. The authors further write $f_{\{q,k\}}$ in a multiplication matrix:

$$f_{\{q,k\}}(x_m, m) = \begin{pmatrix} \cos m\theta & -\sin m\theta \\ \sin m\theta & \cos m\theta \end{pmatrix} \begin{pmatrix} W_{\{q,k\}}^{(11)} & W_{\{q,k\}}^{(12)} \\ W_{\{q,k\}}^{(21)} & W_{\{q,k\}}^{(22)} \end{pmatrix} \begin{pmatrix} x_m^{(1)} \\ x_m^{(2)} \end{pmatrix} \tag{5}$$

where $\left(x_m^{(1)}, x_m^{(2)} \right)$ is x_m expressed in 2D coordinates. The essence of the rotational positional encoding algorithm lies in the formation of a rotation matrix through a predefined constant angle θ, which rotates the word embedding vectors of affine transformations by multiples of their positional indices. Consequently, each word vector is endowed with a unique rotation, reflecting its position within the sentence. As depicted in Figure 4a,b, the relative distance between two words remains constant; hence, $\theta' = \theta_1 - \theta_2$ remains unchanged. However, the positions of the two words within the sentence have shifted, resulting in a mutual rotation by a certain angle. In contrast, the relative distance between the two words has altered, leading to a change in θ' as illustrated in Figure 4c.

In order to adapt word embedding representations for practical use in Natural Language Processing (NLP), the authors extend the formula from a 2-dimensional plane to a more general form. Specifically, the authors partition the d-dimensional space into d/2 subspaces, where d is an even number, and utilize the linearity of inner products to combine them, transforming fq, k into:

$$f_{\{q,k\}}(x_m, m) = R_{\Theta,m}^d W_{\{q,k\}} x_m \tag{6}$$

where

$$R_{\ominus,m}^d = \begin{pmatrix} \cos m\theta_1 & -\sin m\theta_1 & 0 & 0 & \cdots & 0 & 0 \\ \sin m\theta_1 & \cos m\theta_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cos m\theta_2 & -\sin m\theta_2 & \cdots & 0 & 0 \\ 0 & 0 & \sin m\theta_2 & \cos m\theta_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \cos m\theta_{d/2} & -\sin m\theta_{d/2} \\ 0 & 0 & 0 & 0 & \cdots & \sin m\theta_{d/2} & \cos m\theta_{d/2} \end{pmatrix} \tag{7}$$

is a rotary matrix with predefined parameters $\Theta = \left\{ \theta_i = 10{,}000^{-2(i-1)/d}, i \in [1, 2, \ldots, d/2] \right\}$. The authors apply a Rotary Position Embedding matrix to the self-attention formula, resulting in:

$$q_m^\top k_n = \left(R_{\Theta,m}^d W_q x_m \right)^\top \left(R_{\Theta,n}^d W_k x_n \right) = x^\top W_q R_{\Theta,n-m}^d W_k x_n \tag{8}$$

where $R_{\Theta,n-m}^{d} = \left(R_{\Theta,m}^{d}\right)^{\top} R_{\Theta,n}^{d}$. This is an orthogonal matrix, ensuring stability in the process of encoding positional information. When applied to the self-attention formula, the RoFormer model naturally incorporates relative positional information through the product of rotation matrices, rather than utilizing additive positional encoding methods based on altering terms in the expanded formulation.
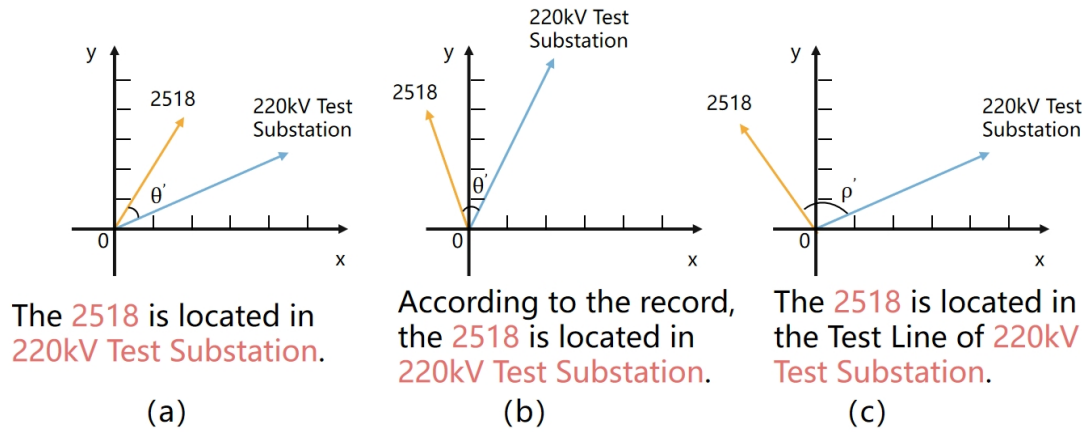
Figure 4. An illustration of rotary position. Subfigure (**a**) illustrates the first set of positions of two entities in a sentence, with the relative distance corresponding to an angle denoted as θ'. Subfigure (**b**) depicts the second set of positions of the two entities in the sentence, both of which have rotated by a certain angle simultaneously. As their relative distance remains the same, the corresponding angle is still denoted as θ'. Subfigure (**c**) represents the third set of positions of the two entities in the sentence, where their relative distance differs, resulting in the corresponding angle becoming ρ'.

4.3. Distribution Focal Loss

The original authors fine-tuned the UIE model using teacher-forcing cross-entropy loss [11]. We propose an innovative approach of utilizing DFL in place of BCE as the loss function in the fine-tuning stage. Distribution Focal Loss, derived from the Generalized Focal Loss, is a loss function utilized in CV for dense object detection tasks. The purpose is to minimize the relative offset between the predicted position and the four sides of the bounding box. This paper extends its application to the field of NLP. Based on the utility of the BCE function in the original model, we propose employing the DFL function to calculate the loss between the predicted entity's start and end positions in the text and the corresponding true labels' start and end positions in the text. This approach can be considered a general method for optimizing various models that utilize the BCE function.

Initially, the Focal Loss (FL) was introduced to address the issue of extreme imbalance in the number of positive and negative samples during the training process. The formula is as follows:

$$\mathrm{FL}(p_{\mathrm{t}}) = -\alpha_{\mathrm{t}}(1 - p_{\mathrm{t}})^{\gamma} \log(p_{\mathrm{t}}) \tag{9}$$

The Focal Loss is composed of two parts: the standard cross-entropy component $-\log(pt)$ and the dynamic scaling factor component $(1 - pt)\gamma$. The parameter α_{t} denotes the class weight of the t-th sample, which is utilized to address issues related to class imbalance. γ serves as an adjustable focusing parameter that dynamically adjusts the rate at which weights of simple samples are reduced using the rapid scaling properties of power functions. A larger γ results in a greater degree of scaling by the power function, thereby increasing the emphasis on difficult samples. In the context of general prediction problems, let us denote the predicted label as $\hat{y}$ and the true label as y. A generalized distribution can be represented by a function $P(x)x$, where $P(x)x$ can encompass various values such that the integral results in y. Assuming the true label y falls within the range $[y_0, y_n]$, there exists a predicted value $\hat{y}$ that satisfies:

$$\hat{y} = \int_{-\infty}^{+\infty} P(x)x \, \mathrm{d}x = \int_{y_0}^{y_n} P(x)x \, \mathrm{d}x \tag{10}$$

As shown in Figure 5, various flexible distributions can achieve the same integral target. In other words, the sample distribution for outcomes predicted as label y may be imbalanced. Intuitively, when compared with sub-graph (1) and sub-graph (2), the distribution of samples in sub-graph (3) appears to be more compact. In the field of CV, this function is utilized for optimizing the coordinates of a bounding box, with the aim of minimizing relative offsets from the location to the four sides of the bounding box. When the predictions are more concentrated around the target label y, the resulting bounding boxes tend to convey a greater sense of confidence and precision. Therefore, in order to improve the predictability of accurate outcomes and achieve faster convergence, it is essential to optimize the shape of $P(x)x$ to increase the probability of it being closer to the target label y. Considering all the factors mentioned above, the original authors propose the concept of Distribution Focal Loss (DFL). This concept aims to expedite the model's prediction values $\hat{y}$ to converge around the label y by explicitly increasing the probability of the two nearest values y_i and y_{i+1} to y, where $y_i \le y \le y_{i+1}$. The DFL can be expressed as the following equation:

$$\mathbf{DFL}(\mathbf{S}_i, \mathbf{S}_{i+1}) = -((y_{i+1} - y)\log(\mathbf{S}_i) + (y - y_i)\log(\mathbf{S}_{i+1})) \tag{11}$$

where $\log(\mathbf{S}_i)$ is defined as the Binary Cross-Entropy Function. The global minimum solution of the DFL is $\mathbf{S}_i = \frac{y_{i+1}-y}{y_{i+1}-y_i}, \mathbf{S}_{i+1} = \frac{y-y_i}{y_{i+1}-y_i}$. It can drive the estimated regression target $\hat{y}$ to approach the corresponding true label y infinitely, thereby ensuring the accuracy of its use as a loss function. For example, $\hat{y} = \sum_{j=0}^{n} P(y_j)y_j = \mathbf{S}_i y_i + \mathbf{S}_{i+1} y_{i+1} = \frac{y_{i+1}-y}{y_{i+1}-y_i}y_i + \frac{y-y_i}{y_{i+1}-y_i}y_{i+1} = y$.

(a)   (b)   (c)
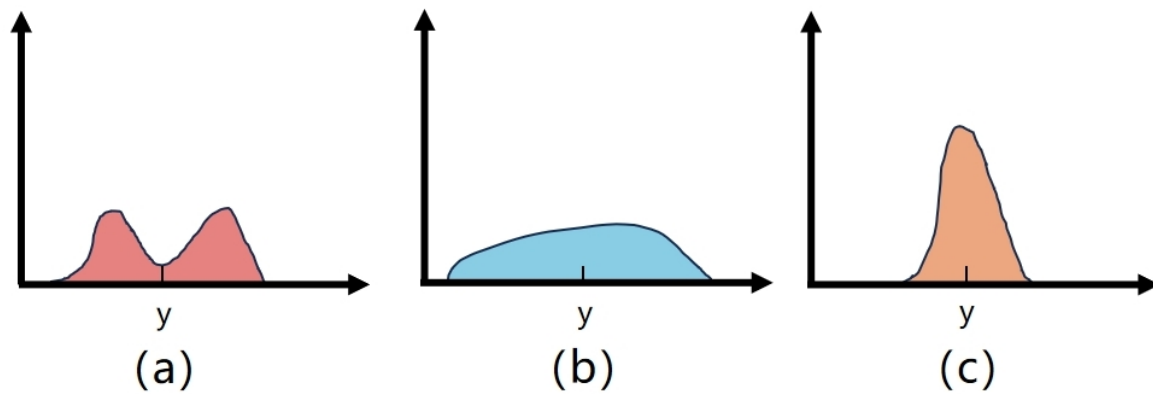
Figure 5. Different distributions but the same integral target. Subfigures reflect the sample distribution for outcomes predicted as label y. Subfigure (**a**) displays a bimodal distribution. Subfigure (**b**) shows a uniform distribution. Subfigure (**c**) exhibits a compact distribution.

5. Experiments

5.1. Datasets

In order to evaluate the general applicability of the model proposed in this paper, we have selected the large-scale high-quality Chinese open dataset DUIE from Baidu as a universal dataset for model comparison. This dataset comprises 210,000 sentences and 450,000 instances encompassing 49 commonly used relations, thus providing a representation of real-world scenarios [35].

At present, there is a lack of comprehensive and reliable publicly available datasets in the field of power production. In order to closely reflect the actual on-site environment, we collect original textual data, including equipment inventory records, system alert logs, and defect or test reports spanning the years 2016 to 2023. As previously mentioned, the original

corpus originates from various regions and systems, maintained by different personnel. Certain data is presented in JSON format files, covering details such as inventory, defect descriptions, and other relevant information. Consequently, it is necessary to adhere to a standardized short-text format for power equipment monitoring information, as discussed earlier, to systematically arrange the textual content into a consistent and concise style. In accordance with the DUIE format and after thorough pre-processing and data cleaning, we have constructed a dataset for information extraction named the Power Equipment State Evaluation Dataset (PESED). Based on the ontology layer of the knowledge graph proposed in Figure 2, 35 connecting lines correspond to 35 types of relationships. PESED encompasses all inventory, management, and monitoring information, similarly covering the 35 types of relationships. The schema and instances corresponding to this dataset are detailed in Appendix A. The dataset consists of 34,043 distinct entities and attributes, encompassing 35 different types of relationships, which are represented in 56,738 sentences. It comprises processed text data divided into a training set (80%), testing set (10%), and validation set (10%).

5.2. Model Parameter Settings

Due to the greater computational resources required by generative models compared to extractive models, this paper conducted experiments using three NVIDIA 3090 24G GPUs. In contrast to utilizing the Paddle deep learning training framework in the UIE model, the RoUIE model and other baseline models in this paper are implemented on the PyTorch deep learning framework for the purpose of facilitating comparisons within the same training framework. Training and evaluation of all types of models are conducted on the PyCharm programming platform. The detailed hyper-parameter settings of RoUIE are presented in Table 2. According to the UIE model, RoUIE collects positive and negative samples and converts data into the doccano format during the data processing stage. For the training set of PESED, a total of 983,745 samples are generated. With a batch size set at 96, each iteration comprises 10,248 steps, where each step involves loading and training 96 data entries. Similarly, for the DUIE dataset, a total of 989,863 samples are constructed, resulting in 10,312 steps executed per iteration. The plots illustrating the relationship between error rates and the number of epochs are presented in Figures 6 and 7, showcasing the convergence situation of the UIE and RoUIE models across both the DUIE and PESED datasets. It can be observed that by the tenth iteration of training, both the UIE and RoUIE model have essentially converged.

Table 2. Detailed hyper-parameter settings of RoUIE.

Hyper-Parameter	Setting
Learning rate	1×10^{-5}
Batch size	96
Num of epochs	10
max length	512
Random seed for initialization	1000
optimizer	AdamW

Referring to other evaluation methods for relational extraction models, this paper employs the F1 score as the evaluation index. The F1 score is a metric used for binary classification problems, which combines precision and recall. It is a single value used to assess the performance of a classifier, with a range between 0 and 1, where 1 indicates a perfect classifier and 0 represents the poorest classifier. The relevant formula is as follows:

$$P = \frac{T_P}{T_P + F_P}$$
$$R = \frac{T_P}{T_P + F_N}$$
$$F_1 = \frac{2PR}{P + R}$$

(12)

where T_P, F_P, and F_N are elements in the confusion matrix. They refer to true positive, false positive, and false negative, respectively. In other words, P represents the ratio of correctly predicted triplets in all sentences of the test set to the total number of predicted triplets, while R denotes the ratio of correctly predicted triplets in all sentences of the test set to the total number of manually annotated triplets in the test set. The F1 measure is commonly leveraged to assess the performance of relation extraction tasks on datasets with known relationships. This metric can be employed to evaluate the accuracy of the triplets generated by the proposed model. A predicted triplet is considered to have a valid relation with a boundary match when the relation type is accurate and the subject/object strings are also correct simultaneously. A higher number of correct triplets generated by the model indicates a more accurate and comprehensive extraction of state evaluation-related textual information obtained from various systems.
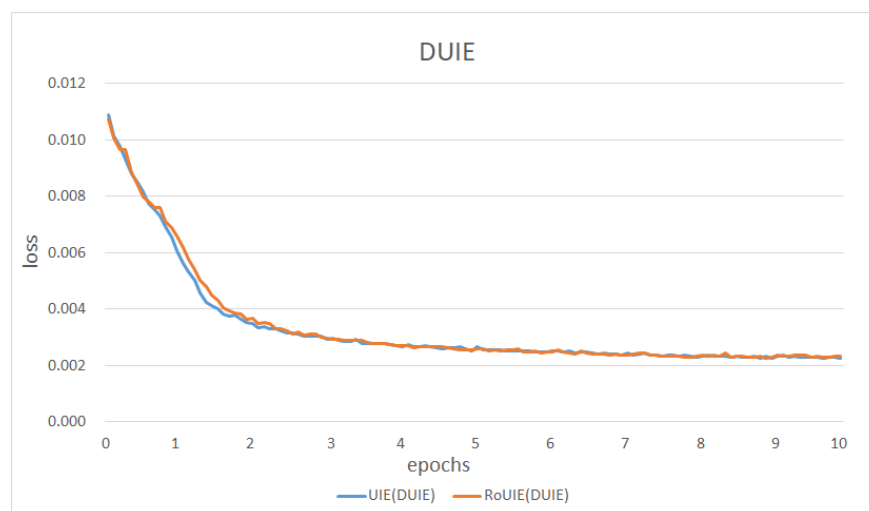
Figure 6. Convergence situation of UIE and RoUIE model in the DUIE dataset.
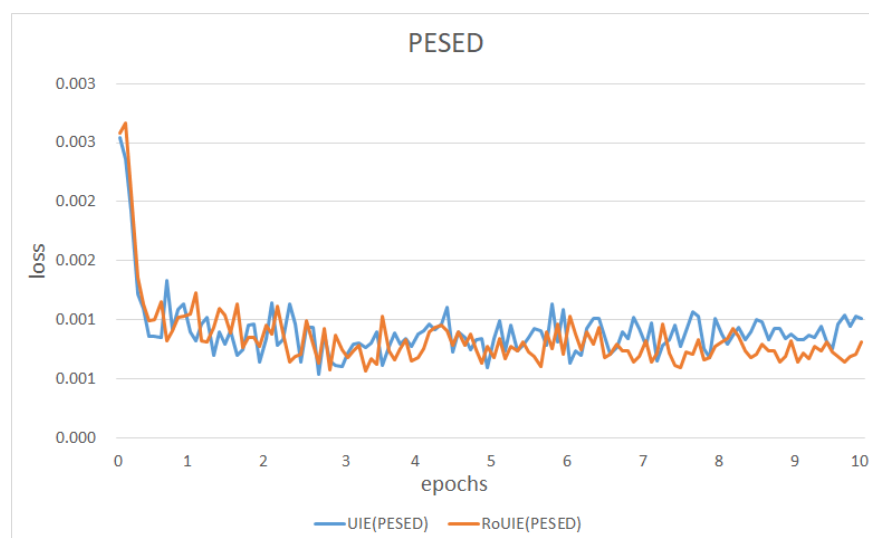
Figure 7. Convergence situation of UIE and RoUIE model in the PESED dataset.

5.3. Experimental Results

In this article, we adopt mainstream joint extraction models proposed in recent years as comparative baselines, including Casrel, Onerel, TPlinker, and GPlinker [36], with the results presented in Table 3. To ensure experimental fairness and considering the characteristics of Chinese text, we leverage the Chinese-bert-wwm-ext [37] as the unified base pre-training language model for our baseline models. We can observe that our proposed model outperforms all the baseline models in terms of F1 score. In the DUIE dataset, RoUIE achieves an F1 score of 81.97, demonstrating an improvement ranging from 6.27 to 11.99 compared to other mainstream joint extraction models. When compared to the UIE model, RoUIE exhibits a modest increase of 0.82 points. In the PESED dataset, RoUIE demonstrates a substantial average improvement of approximately 21 points in the F1 score over other joint extraction models, attributed to its architecture being well-designed to the characteristics of the dataset. Furthermore, compared to the UIE model, RoUIE shows a relatively significant improvement with an increase of 2.92 points. The generative models significantly outperformed other models in terms of recall rate in the PESED, thereby further highlighting the advantages of generative models in the application to domain-specific datasets.

Table 3. Comparison experiment results with mainstream joint extraction models and UIE.

Model	DUIE			PESED		
	Prec.	Rec.	F1	Prec.	Rec.	F1
Casrel [23]	74.47	66.00	69.98	99.15	52.87	68.97
TPLinker [27]	76.66	76.61	76.63	98.85	52.89	68.91
GPLinker [36]	76.77	74.55	75.65	94.56	48.79	64.37
Onerel [28]	76.49	74.93	75.70	94.45	54.71	69.29
UIE$_{base}$ [11]	82.30	80.02	81.15	76.14	98.72	85.97
RoUIE	83.41	80.58	81.97	80.50	99.23	88.89

Several recently proposed advanced pre-trained language models, such as RocBERT [38], are utilized as comparative pre-trained language models for the UIE model. Ablation experiments are carried out to evaluate the two enhancements proposed in Sections 4.2 and 4.3. The detailed results are shown in Table 4. Specifically, when RoFormer v2 is utilized as the encoder without altering the model's loss function, the model demonstrates F1 scores of 81.79 in the DUIE dataset and 88.04 in the PESED dataset. In comparison to the UIE base model and the model employing RocBERT as the encoder, there is a slight increase of 0.64 and 0.14 in the DUIE dataset, and a growth of 2.07 and 0.87 in the PESED dataset, respectively. Furthermore, in the context of the UIE model with DFL where only the loss function is modified without adjusting the encoder, the F1 scores show improvements of 0.51 in the DUIE dataset and 0.94 in the PESED dataset compared to the UIE base model with BCE loss. The comparison results and our ablation study findings indicate that the two improvements we proposed for the model have produced positive effects.

Table 4. Results of ablation experiments and comparisons with various pre-trained language models.

Model	DUIE			PESED		
	Prec.	Rec.	F1	Prec.	Rec.	F1
UIE$_{RocBERT}$	83.19	80.16	81.65	78.85	97.45	87.17
UIE$_{RoFormer\ v2}$	83.51	80.14	81.79	79.34	98.88	88.04
UIE$_{base+DFL}$	82.70	80.64	81.66	77.30	99.25	86.91
UIE$_{RoFormer\ v2+FL}$	84.57	77.90	81.10	79.49	99.25	88.28
RoUIE (UIE$_{RoFormer\ v2+DFL}$)	83.41	80.58	81.97	80.50	99.23	88.89

6. Conclusions

In this article, we propose RoUIE, a generative information extraction model, as an improved model for the fine-tuning stage of UIE, aimed at effectively extracting textual information related to the state evaluation of power equipment. To better align real-time monitoring information or operational records generated by the system with existing inventory information, we introduce a unified short-text format for power equipment monitoring information and use it as part of the annotated text to create a dataset for evaluating the status of electrical equipment. In this article, we propose RoUIE, a generative information extraction model, as an improved model for the fine-tuning stage of UIE, aimed at effectively extracting textual information related to the evaluation of electrical equipment status. To better align real-time monitoring information or operational records generated by the system with existing equipment ledger information, we introduce a unified short-text format for monitoring the information of power equipment, which is utilized as part of the annotated text to construct a dataset for state evaluation. Considering the characteristics of the textual evaluation of power equipment status, RoUIE adopts RoFormer as the text encoder, which exhibits good long-term decay properties and sequence flexibility. Additionally, we creatively incorporate the Distribution Focal Loss, originally used in the field of CV, to replace the Binary Cross-Entropy Loss as the model's loss function. Experimental results demonstrate that the RoUIE model exhibits superior performance and general applicability compared to baseline relation extraction models and the original UIE model, across the specific dataset in the field of power production and the general Chinese dataset.

Author Contributions: Conceptualization, Z.Y. and D.Q.; methodology, Z.Y.; software, Z.Y. and H.L.; validation, Z.Y., H.L. and X.L.; formal analysis, Z.Y.; investigation, Z.Y.; resources, Z.Y. and Y.Y.; data curation, Z.Y.; writing—original draft preparation, Z.Y.; writing—review and editing, Z.Y.; visualization, Q.C.; supervision, D.Q.; project administration, Z.Y. and D.Q.; funding acquisition, Y.Y. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Research data are not shared.

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

Abbreviations

The following abbreviations are used in this manuscript:

UIE	Universal Information Extraction
BCE	Binary Cross-Entropy Loss
DFL	Distribution Focal Loss
DGA	Dissolved Gas Analysis
SRO	triplets of subject, relationship, and object
SEO	Single-Entity-Overlap
SEL	Structural Extraction Language
SSI	Structural Schema Instructor
FL	Focal Loss
CV	Computer Vision
NLP	Natural Language Processing
DUIE	A large-scale Chinese dataset for information extraction
PESED	Power Equipment State Evaluation Dataset

Appendix A

Table A1. Detailed schema of device-centric knowledge graph.

Object Type (Combination of Attributes)	Predicate (Relation)	Subject Type (Combination of Entities)
Substation	Locate in	Equipment
Substation	Locate in	Bar
Bar	Corresponding bar	Equipment
Current condition	Current condition	Equipment
Component	Consist of	Equipment
Operation department	Operation department	Equipment
Subclass of equipment	Corresponding Subclass	Equipment
Date and time	Date of commencement	Equipment
Date and time	Date of production	Equipment
Equipment	Corresponding equipment	Defect report ID
Defect level	Corresponding defect level	Defect report ID
Defect type	Corresponding defect type	Defect report ID
Defect resolution status	Defect resolution status	Defect report ID
Description of the defect	Description of the defect	Defect report ID
Personnel	Defect reporter	Defect report ID
Date and time	Date of defect reporting	Defect report ID
Team	Defect handling team	Defect report ID
Personnel	Defect handler	Defect report ID
Date and time	Date of defect handling	Defect report ID
Origin of defect	Origin of defect	Defect report ID
Equipment	Corresponding equipment	Inspection and test report ID
Team	Inspection (test) team	Inspection and test report ID
Date and time	Date and time of Inspection (test)	Inspection and test report ID
Personnel	Inspection (test) engineer	Inspection and test report ID
Approaches	Inspection (test) approaches	Inspection and test report ID
Result of Inspection (test)	Result of Inspection (test)	Inspection and test report ID
Equipment	Corresponding equipment	Alarm ID
Source system	Source system	Alarm ID
Alarm level	Alarm level	Alarm ID
Abnormal description	Abnormal description	Alarm ID

Table A2. An instance of input data from the PESED training set (the original sample is represented in Chinese).

Datase	Input Data from the PESED
PESED	{"spo_list":[{"subject_*type*":"Bar", "object_*type*":"Substation", "subject":"Test Line", "predicate":"Locate in", "object":"110kV Test A Substation"}, {"subject_*type*":"Equipment", "object_*type*":"Bar", "subject":"1830", "predicate":" Corresponding bar", "object":"Test Line"}, {"subject_*type*":" Corresponding bar", "object_*type*":"Operation department", "subject":"1830", "predicate":" Operation department", "object":"Substation Management Department 1"}, {"subject_*type*":"Equipment", "object_*type*":"Date and time", "subject":"1830", "predicate":"Date of production", "object":"1 Mar 2011 00:00:00"}, {"subject_*type*":"Equipment", "object_*type*":"Date and time", "subject":"1830", "predicate":"Date of commencement", "object":"30 Nov 2011 00:00:00"}, {"subject_*type*":"Equipment", "object_*type*":" Current condition", "subject":"1830", "predicate":" Current condition", "object":"hot standby"}], "text":"Incident Report: The 110kV Test A Substation Test Line 1830 tripped and failed reclosing. Upon investigation, it was found that the equipment is managed and maintained by ubstation Management Department 1, with a production date of 1 Mar 2011 00:00:00, and was officially put into operation on 30 Nov 2011 00:00:00. Its current status is hot standby."}

Table A3. Based on the text of the sample provided in Table A2, applying the RoUIE model for inference yields the following corresponding result (In Chinese).

Input Text	Corresponding Output Result Generated from the Model Inference
As shown in the input text from Table A2	[{'Equipment': [{'end': 21, 'probability': 0.8385402622754041, 'relations': {'Corresponding bar': [{'end': 17, 'probability': 0.8421710521379353, 'start': 14, 'text': 'Test Line'}], 'Current condition': [{'end': 109, 'probability': 0.9079101521374563, 'start': 106, 'text': 'hot standby'}], 'Operation department': [{'end': 46, 'probability': 0.9754414623456341, 'start': 40, 'text': 'Substation Management Department 1'}], 'Date of production': [{'end': 75, 'probability': 0.8292706618236352, 'start': 56, 'text': '1 Mar 2011 00:00:00'}, {'end': 96, 'probability': 0.4456306614587452, 'start': 77, 'text': '30 Nov 2011 00:00:00'}], 'Date of commencement': [{'end': 96, 'probability': 0.6671295049136148, 'start': 77, 'text': '30 Nov 2011 00:00:00'}]}, 'start': 17, 'text': '1830'}], 'Bar': [{'end': 17, 'probability': 0.7685434552767871, 'relations': {'Locate in': [{'end': 14, 'probability': 0.9267710521345789, 'start': 5, 'text': '110kV Test A Substation'}, {'end': 14, 'probability': 0.5364710458655789, 'start': 12, 'text': 'A Substation'}]}, 'start': 14, 'text': 'Test Line'}] }]

Table A4. Based on the output results from Table A3, the meanings of the model evaluation metrics T_P, F_P, and F_N can be demonstrated using triplets (In Chinese).

True Positive T_P	False Positive F_P	False Negative F_N
T_P denotes the number of correctly predicted triplets in all sentences. It consists of the corresponding triplets as follows: (Test Line, Locate in, 110 kV Test A Substation), (1830, Corresponding bar, Test Line), (1830, Current condition, hot standby), (1830, Operation department, Substation Management Department 1), (1830, Date of production, 1 March 2011 00:00:00), (1830, Date of commencement, 30 November 2011 00:00:00)	F_P denotes the number of triplets extracted by the model that are incorrectly predicted. It consists of the corresponding triplet as follows: (Test Line, Locate in, A Substation)	F_N denotes the number of manually annotated triplets that were not successfully extracted by the model. In the aforementioned inference stage, no manually annotated triplet was missed during extraction. Hence, it is unable to provide the corresponding triplets for F_N here.

References

1. Wang, J.; Liu, H.; Pan, Z.; Zhao, S.; Wang, W.; Geng, S. Research on Evaluation Method of Transformer Operation State in Power Supply System. In *Proceedings of the Wireless Technology, Intelligent Network Technologies, Smart Services and Applications: Proceedings of 4th International Conference on Wireless Communications and Applications (ICWCA 2020)*; Springer: Singapore, 2022; pp. 249–256.
2. Chen, L. Evaluation method of fault indicator status detection based on hierarchical clustering. *J. Phys. Conf. Ser.* **2021**, *2125*, 012002. [CrossRef]
3. Qi, B.; Zhang, P.; Rong, Z.; Li, C. Differentiated warning rule of power transformer health status based on big data mining. *Int. J. Electr. Power Energy Syst.* **2020**, *121*, 106150. [CrossRef]
4. Jin-qiang, C. Fault prediction of a transformer bushing based on entropy weight TOPSIS and gray theory. *Comput. Sci. Eng.* **2018**, *21*, 55–62. [CrossRef]
5. Xie, C.; Zou, G.; Wang, H.; Jin, Y. A new condition assessment method for distribution transformers based on operation data and record text mining technique. In Proceedings of the 2016 China International Conference on Electricity Distribution (CICED), Xi'an, China, 10–13 August 2016; pp. 1–7. [CrossRef]
6. Liu, R.; Fu, R.; Xu, K.; Shi, X.; Ren, X. A Review of Knowledge Graph-Based Reasoning Technology in the Operation of Power Systems. *Appl. Sci.* **2023**, *13*, 4357. [CrossRef]
7. Meng, F.; Yang, S.; Wang, J.; Xia, L.; Liu, H. Creating knowledge graph of electric power equipment faults based on BERT–BiLSTM–CRF model. *J. Electr. Eng. Technol.* **2022**, *17*, 2507–2516. [CrossRef]
8. Yang, Y.; Wu, Z.; Yang, Y.; Lian, S.; Guo, F.; Wang, Z. A survey of information extraction based on deep learning. *Appl. Sci.* **2022**, *12*, 9691. [CrossRef]
9. Yang, J.; Meng, Q.; Zhang, X. Improvement of operation and maintenance efficiency of power transformers based on knowledge graphs. *Iet Electr. Power Appl.* **2024**, *1*, 1–13. [CrossRef]
10. Xie, Q.; Cai, Y.; Xie, J.; Wang, C.; Zhang, Y.; Xu, Z. Research on Construction Method and Application of Knowledge Graph for Power Transformer Operation and Maintenance Based on ALBERT. *Trans. China Electrotech. Soc.* **2023**, *38*, 95–106.
11. Lu, Y.; Liu, Q.; Dai, D.; Xiao, X.; Lin, H.; Han, X.; Sun, L.; Wu, H. Unified structure generation for universal information extraction. *arXiv* **2022**, arXiv:2203.12277.
12. Lou, J.; Lu, Y.; Dai, D.; Jia, W.; Lin, H.; Han, X.; Sun, L.; Wu, H. Universal information extraction as unified semantic matching. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington DC, USA, 7–14 February 2023; Volume 37, pp. 13318–13326.
13. Liu, C.; Zhao, F.; Kang, Y.; Zhang, J.; Zhou, X.; Sun, C.; Wu, F.; Kuang, K. Rexuie: A recursive method with explicit schema instructor for universal information extraction. *arXiv* **2023**, arXiv:2304.14770.
14. Fei, H.; Wu, S.; Li, J.; Li, B.; Li, F.; Qin, L.; Zhang, M.; Zhang, M.; Chua, T.S. Lasuie: Unifying information extraction with latent adaptive structure-aware generative language model. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 15460–15475.

15. Su, J.; Ahmed, M.; Lu, Y.; Pan, S.; Bo, W.; Liu, Y. Roformer: Enhanced transformer with rotary position embedding. *Neurocomputing* **2024**, *568*, 127063. [CrossRef]

16. Li, X.; Wang, W.; Wu, L.; Chen, S.; Hu, X.; Li, J.; Tang, J.; Yang, J. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 21002–21012.

17. Zhang, Y.S.; Liu, S.K.; Liu, Y.; Ren, L.; Xin, Y.H. Joint Extraction of Entities and Relations Based on Deep Learning:A Survey. *Acta Electron. Sin.* **2023**, *51*, 1093–1116.

18. Liu, W.; Yin, M.; Zhang, J.; Cui, L. A Joint Entity Relation Extraction Model Based on Relation Semantic Template Automatically Constructed. *Comput. Mater. Contin.* **2024**, *78*, 975–997. [CrossRef]

19. Fei, H.; Ren, Y.; Zhang, Y.; Ji, D.; Liang, X. Enriching contextualized language model from knowledge graph for biomedical information extraction. *Briefings Bioinform.* **2021**, *22*, bbaa110. [CrossRef] [PubMed]

20. Katiyar, A.; Cardie, C. Going out on a limb: Joint extraction of entity mentions and relations without dependency trees. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vancouver, BC, Canada, 30 July–4 August 2017; pp. 917–928.

21. Bekoulis, G.; Deleu, J.; Demeester, T.; Develder, C. Adversarial training for multi-context joint entity and relation extraction. *arXiv* **2018**, arXiv:1808.06876.

22. Zeng, X.; Zeng, D.; He, S.; Liu, K.; Zhao, J. Extracting relational facts by an end-to-end neural model with copy mechanism. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia, 15–20 July 2018; pp. 506–514.

23. Wei, Z.; Su, J.; Wang, Y.; Tian, Y.; Chang, Y. A novel cascade binary tagging framework for relational triple extraction. *arXiv* **2019**, arXiv:1909.03227.

24. Wang, Y.; Sun, C.; Wu, Y.; Zhou, H.; Li, L.; Yan, J. UniRE: A unified label space for entity relation extraction. *arXiv* **2021**, arXiv:2107.04292.

25. Li, X.; Yin, F.; Sun, Z.; Li, X.; Yuan, A.; Chai, D.; Zhou, M.; Li, J. Entity-relation extraction as multi-turn question answering. *arXiv* **2019**, arXiv:1905.05529.

26. Sui, D.; Zeng, X.; Chen, Y.; Liu, K.; Zhao, J. Joint entity and relation extraction with set prediction networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, 1–12. [CrossRef]

27. Wang, Y.; Yu, B.; Zhang, Y.; Liu, T.; Zhu, H.; Sun, L. TPLinker: Single-stage joint extraction of entities and relations through token pair linking. *arXiv* **2020**, arXiv:2010.13415.

28. Shang, Y.M.; Huang, H.; Mao, X. Onerel: Joint entity and relation extraction with one module in one step. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 22 February–1 March 2022; Volume 36, pp. 11285–11293.

29. Dong, L.; Yang, N.; Wang, W.; Wei, F.; Liu, X.; Wang, Y.; Gao, J.; Zhou, M.; Hon, H.W. Unified language model pre-training for natural language understanding and generation. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 1–13 .

30. Lewis, M.; Liu, Y.; Goyal, N.; Ghazvininejad, M.; Mohamed, A.; Levy, O.; Stoyanov, V.; Zettlemoyer, L. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv* **2019**, arXiv:1910.13461.

31. Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; Liu, P.J. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.* **2020**, *21*, 1–67.

32. Katikapalli, K. A survey of GPT-3 family large language models including ChatGPT and GPT-4. *Natural Language Processing Journal.* **2024**, *6*, 100048. [CrossRef]

33. Hang, T.; Feng, J.; Wu, Y.; Yan, L.; Wang, Y. Joint extraction of entities and overlapping relations using source-target entity labeling. *Expert Syst. Appl.* **2021**, *177*, 114853. [CrossRef]

34. Rosin, G.D.; Radinsky, K. Temporal attention for language models. *arXiv* **2022**, arXiv:2202.02093.

35. Li, S.; He, W.; Shi, Y.; Jiang, W.; Liang, H.; Jiang, Y.; Zhang, Y.; Lyu, Y.; Zhu, Y. Duie: A large-scale chinese dataset for information extraction. In *Natural Language Processing and Chinese Computing, Proceedings of the 8th CCF International Conference, NLPCC 2019, Dunhuang, China, 9–14 October 2019*; Proceedings, Part II 8; Springer: Cham, Switzerland, 2019; pp. 791–800.

36. Su, J.; Murtadha, A.; Pan, S.; Hou, J.; Sun, J.; Huang, W.; Wen, B.; Liu, Y. Global pointer: Novel efficient span-based approach for named entity recognition. *arXiv* **2022**, arXiv:2208.03054.

37. Cui, Y.; Che, W.; Liu, T.; Qin, B.; Yang, Z. Pre-Training With Whole Word Masking for Chinese BERT. *IEEE/ACM Trans. Audio, Speech, Lang. Process.* **2021**, *29*, 3504–3514. [CrossRef]

38. Su, H.; Shi, W.; Shen, X.; Xiao, Z.; Ji, T.; Fang, J.; Zhou, J. Rocbert: Robust chinese bert with multimodal contrastive pretraining. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Dublin, Ireland, 22–27 May 2022; pp. 921–931.