

Article

Conditional Knowledge Extraction Using Contextual Information Enhancement

Zhangbiao Xu ^{1,2,3,†}, Botao Zhang ^{1,2,3,†} , Jinguang Gu ^{1,2,3,†}  and Feng Gao ^{1,2,3,*,†} 

¹ College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China

² Institute of Big Data Science and Engineering, Wuhan University of Science and Technology, Wuhan 430065, China

³ Key Laboratory of Rich-media Knowledge Organization and Service of Digital Publishing Content, National Press and Publication Administration, Beijing 100038, China

* Correspondence: feng.gao86@wust.edu.cn

† These authors contributed equally to this work.

Abstract: Conditional phrases provide fine-grained domain knowledge in various industries, including medicine, manufacturing, and others. Most existing knowledge extraction research focuses on mining triplets with entities and relations and treats that triplet knowledge as plain facts without considering the conditional modality of such facts. We argue that such approaches are insufficient in building knowledge-based decision support systems in vertical domains, where specific and professional instructions on what facts apply under given circumstances are indispensable. To address this issue, this paper proposes a condition-aware knowledge extraction method using contextual information. In particular, this paper first fine-tunes the pre-training model to leverage a local context enhancement to capture the positional context of conditional phrases; then, a sentence-level context enhancement is used to integrate sentence semantics; finally, the correspondences between conditional phrases and relation triplets are extracted using syntactic attention. Experimental results on public and proprietary datasets show that our model can successfully retrieve conditional phrases with relevant triplets while improving the accuracy of the matching task by 2.68%, compared to the baseline.

Keywords: phrase extraction; pre-training model; information matching; conditional knowledge graph



Citation: Xu, Z.; Zhang, B.; Gu, J.; Gao, F. Conditional Knowledge Extraction Using Contextual Information Enhancement. *Appl. Sci.* **2023**, *13*, 4954. <https://doi.org/10.3390/app13084954>

Academic Editor: Yu-Dong Zhang

Received: 15 March 2023

Revised: 9 April 2023

Accepted: 11 April 2023

Published: 14 April 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

Information extraction [1] using deep learning has been discussed in depth in both scientific research and industrial application across various domains. Information extraction enables the automatic extraction of entities, relations, events, and others from natural language text. It plays a pivotal role in domain knowledge constructions, which is crucial for knowledge-based intelligent applications.

Among various information extraction tasks, relation extraction [2] is designed to extract the relations contained in the text between entities. A complete relation extraction consists of entity extraction and relation classification. The entity extraction sub-process, also known as named entity recognition, detects and classifies the entities in a sentence; the relation classification sub-process determines the semantic relation between two entities in a given sentence, which is a multi-category classification problem. Sometimes these two tasks can be integrated into a joint relation extraction framework [3].

In the industrial field, the knowledge contained in device-related documents is crucial for device manufacturing and maintenance. Therefore, analyzing the language structure of documents and extracting device knowledge can provide insight into related information. Within this device knowledge, conditional information plays a pivotal role. We conducted a statistical analysis of industrial documents, which shows that 6.25% of all sentences contain

conditional phrases, and the proportion reaches 16.5% among sentences with important relations. Relation extraction has been discussed in depth in different domains using different methods. However, most existing research treats extracted triplets as simple, static facts without considering the condition of the facts. We argue that facts are relative to the conditions they apply to. Without knowing the relevant conditions, facts can be incomplete, contradictory, or erroneous. These errors and this missing information will propagate to the downstream tasks and cause failures. For example, given the sentence “fishing vessels should travel at speeds greater than 18 km/h under normal conditions and travel at speeds less than 22 km/h under emergencies,” conventional relation extract will find two triplets: “travel_at_speeds-greater_than-18 km/h” and “travel_at_speeds-less_than-22 km/h.” However, the conditional phrases “under_normal_conditions” and “under_emergency_conditions” are not recognized, nor can they figure out which speed is allowed under which condition; decision-making without such information would be impossible.

To address this issue, we propose to extract conditions together with relations by adding a condition recognition process on top of traditional entity relation extraction. More specifically, for the conditional phrase extraction problem, this paper first uses a hybrid approach of rule-based extraction and similarity-based matching to identify conditional phrases. After extracting conditional phrases, the correspondences between conditions and relational triples are constructed. In the example presented in Figure 1, “under_normal_conditions” needs to correspond to the relation triplets generated from “travel_at_speeds-greater_than-18 km/h,” and “under_emergencies” needs to correspond to the relation triplets from “travel_at_speeds-less_than-22 km/h.”

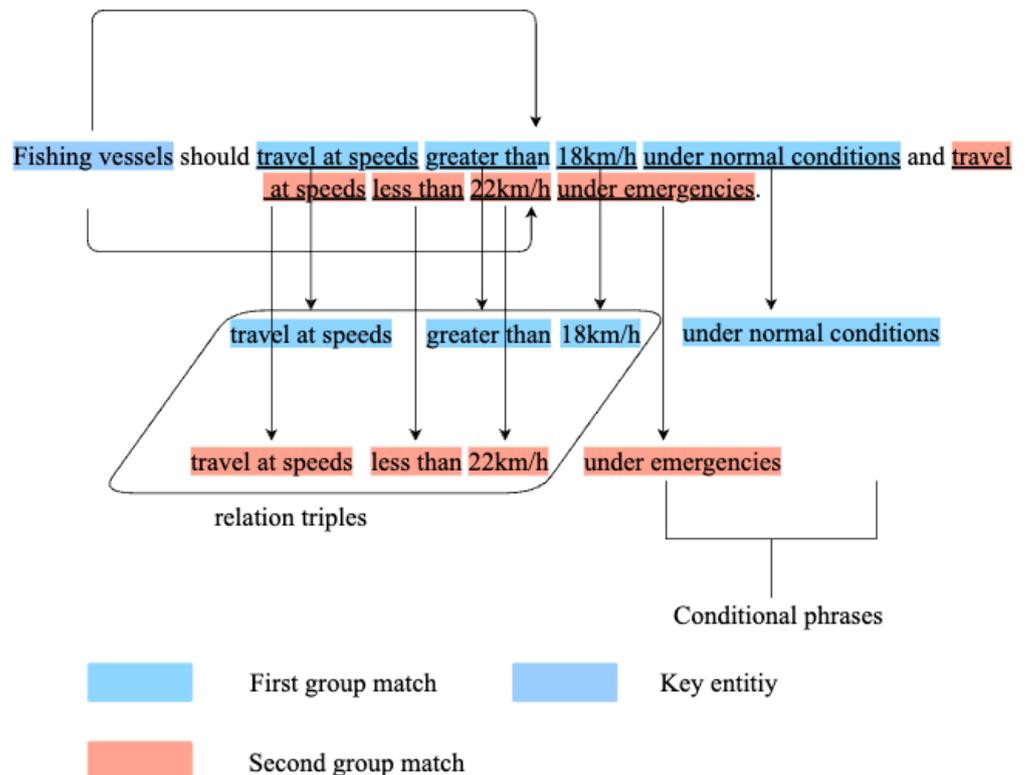


Figure 1. Sentence information matching results: “fishing vessels should travel at speeds greater than 18 km/h under normal conditions and travel at speeds less than 22 km/h under emergencies”.

For the conditional-phrase and triplet-matching problem, this paper uses dependency parsing techniques to assign weights based on the existence of dependencies in the sentence and the semantic information of the sentence as the attention mechanism. Moreover, this paper fine-tuned the pre-training model to allow sentences to learn more conditional information.

The main contributions of this paper to the enhancement effect compared to the baseline model are:

1. This paper proposes a task framework for conditional phrase extraction and conditional phrase matching. It solves the information-matching task.
2. The MLM task is modified based on the baseline model using domain-specific datasets, with the addition of conditional boundaries and covering conditional-phrase keywords to continue the pre-training and enhance the model effect.
3. This paper conducts experiments on the industrial domain and English datasets, respectively. This paper compares the experimental results and verifies the generalization effect of the model.

2. Problem Definition

In the following, we give the formal definition of the problem. We will give the notion of the key concepts, including sentences, conditional phrases, relational triples, and conditional triples. Moreover, we will give the definition of the processes of generating the concepts and their relations.

Sentence: a sentence S is a chronological set of tokens: $S = \{t_1, t_2, \dots, t_n\}$, where t_i is the i th token in the sentence.

Conditional Phrase: a conditional phrase S_{cond} is a subsequence of a sentence S , i.e., $S_{cond} = t_i, \dots, t_j$, where $1 \leq i \leq j \leq n$, so that $S_{cond} \subseteq S$.

A Triple $t = \{s, p, o\}$ is 3-ary statement, where s is the subject, p is the predicate and o is the object.

A mapping function $G_t: S \rightarrow T$ gives the correlation between a sentence S and a set of triples T generated from the sentence.

A mapping function $G_{cond}: S \rightarrow CP$ extracts all conditional phrases (CP) from the sentence S .

A mapping function $G_{ct}: S_{cond} \rightarrow T_{cond}$ generates a set of triples as Conditional Triples (T_{cond}) from a conditional phrase S_{cond} , the set of all Conditional Triples from the sentence S is thus given by $\Sigma G_{ct}(G_{cond}(S))$, and denoted as a superset of CT: Γ_{CT} .

A mapping function map: $(T - \Gamma_{CT}) \times \Gamma_{CT} \rightarrow T_{cor}$ gives the correspondence (in the form of triple set T_{cor}) between relational triples (all triples that do not represent conditions) and conditional triples.

Using the above definition, the entire process of creating conditional knowledge graphs can be defined as a mapping function G for a given sentence S , where $G = G_t(S) \cup \text{map}(G_t(S) - \Sigma G_{ct}(G_{cond}(S)), \Sigma G_{ct}(G_{cond}(S)))$.

From the above definition, it is observable that the conditional KG generation process is a generic framework, and the mapping of relational triplets and conditional triplets are orthogonal to the specific labels of the predicates of the relational triplets. In practice, we can model the correlation between conditional triplets and relation triplets with RDF reification. Figure 2 gives the KG of conventional triplets (without capturing the conditions) from the example in Figure 1, and Figure 3 gives the condition-aware KG of the same sentence using reification. In Figure 3, we have used the predicate "hasTravelSpeed" to indicate that the vessel has speed, which is different from the meaning of "should travel at speeds" in the example in Figure 1.

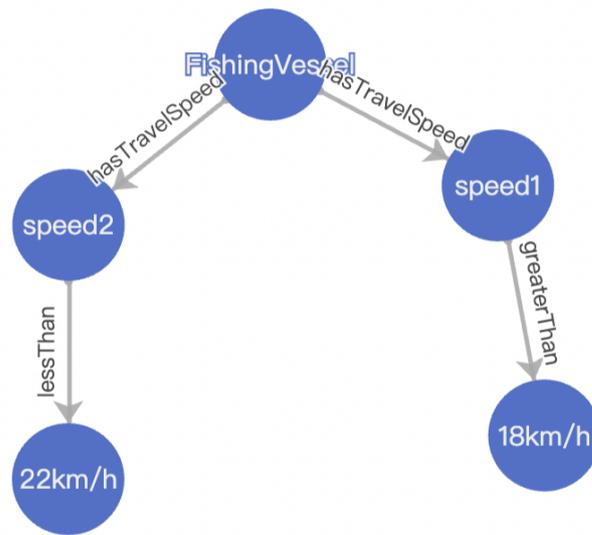


Figure 2. “Fishing vessels should travel at speeds greater than 18 km/h under normal conditions and travel at speeds less than 22 km/h under emergencies” of the conventional knowledge graph statement.

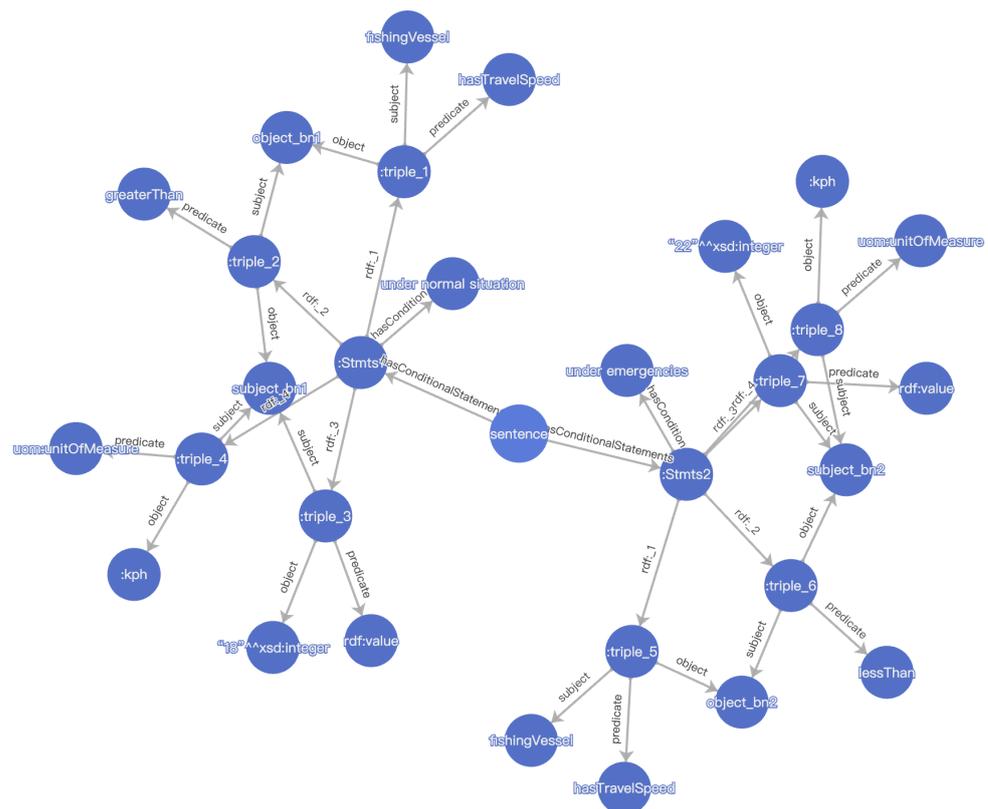


Figure 3. “Fishing vessels should travel at speeds greater than 18 km/h under normal conditions and travel at speeds less than 22 km/h under emergencies” of the conditional knowledge graph statement.

3. Related Work

3.1. Information Extraction

Recent years have significantly focused on extracting information from technical industrial documentation. Early studies used rule-based methods [4,5], but statistical-based approaches have gained popularity with the mainstream adoption of statistical thinking [6]. However, traditional methods have been outperformed by newer neural network models.

Zeng et al. introduced convolutional neural networks (CNNs) [7] that utilize word vectors and positions as inputs and employ convolutional and pooling layers to derive sentence representations. This method has yielded better results in relation extraction tasks by considering location information and lexical features. However, CNNs cannot capture long-term semantic information, which led to the development of long short-term memory (LSTM) models. Peng Z et al. [8] proposed a BiLSTM approach based on the Attention mechanism, which uses a neural attention mechanism to capture important semantic information in sentences. Lin et al. [9] developed a CNN model with a sentence-level attention mechanism that assigned higher weights to critical information in sentences. This paper incorporates dependency relations between words as a critical element of the task and uses sentence dependencies [10,11] to assign attention weights to obtain critical sentence information. Vaswani A et al. [12] introduced the Transformer model, which used an attention mechanism as the entire network structure, producing the best experimental results at that time. Devlin J et al. [13] proposed the pre-training model BERT, which used a bi-directional Transformer to enhance relation extraction and ushered in the era of large-scale pre-training. Thus, this paper uses pre-training models for information extraction and conditional matching tasks.

In summary, these studies have made some progress in information extraction but have yet to propose solutions to specific problems in specific areas. This paper builds on them with pre-training fine-tuning, adding possible solutions tailored to the problem factors.

3.2. Conditional Knowledge Graph

In recent years, knowledge graphs have gradually become a core technology driving the development of artificial intelligence and play a crucial role in various applications such as question and answer [14] and information retrieval [15]. This paper observed the presence of conditional sentences in a certain percentage of sentences in a large amount of scientific text data. The construction of traditional knowledge graphs relies only on the plain text extracted from the text and lacks correspondence between the extracted information. As a result, important conditional information is lost, which limits the expressive power of knowledge graphs and potentially affects the exploration of downstream tasks, such as information retrieval. Conditional sentences are widely available in the scientific literature. Most existing knowledge graphs ignore the impact of conditionals. To address this problem, Jiang et al. [16] proposed a model for representing and constructing scientific conditional knowledge graphs, but it needs to handle the overlapping tuple problem. As far as we know, Jiang et al. [17] were the first to attempt to extract conditional information from text. They proposed a MIMO model to extract factual and conditional triplets but ignored several problems. MIMO has two modules: (1) a multi-input module that harnesses recent NLP development to process the text for input sequences from multiple tasks and feeds them into a multi-head encoder-decoder model with multi-input gates; (2) a multi-output module that generates multiple tuple tag sequences for fact and condition tuples, which consists of a relation name-tagging layer and a tuple-completion tagging layer. However, MIMO cannot clearly distinguish the inter-constraint relations between triplets and does not consider the logical relations between conditional and factual triplets. Zheng et al. [18] proposed a new conditional knowledge graph representation based on Jiang's, which is in the form of nested hierarchical triplets, and also designed a text structure hierarchy analysis module to derive the hierarchical structure of conditional sentences. However, that article focuses on the task of extracting triplets, whereas this paper focuses on the task of matching between conditional phrases and triplets. Moreover, this paper is an innovation in pre-training and model task optimization, which differs from its innovation points. Zheng's article does not show good results on complex sentences, for example: "If symptomatic treatment fails, pharyngeal airway obstruction is possible, and tonsillectomy may be required." This article would make "If symptomatic treatment fails" a condition for "pharyngeal airway obstruction is possible," and "a tonsillectomy may be necessary" incorrectly as a result of "If symptomatic treatment fails" and "pharyngeal airway obstruction is possible" in

juxtaposition, thus losing the “If symptomatic treatment fails” as the “a tonsillectomy may be necessary” conditional relation.

In summary, while these studies have progressed in the conditional knowledge graph, they lack a general framework for matching conditional phrases with relation triplets. Furthermore, certain domains have a high proportion of texts with conditional phrases, necessitating the creation of an efficient model to map these associations.

4. Approach

In this paper, we propose a three-step approach for conditional knowledge extraction. (1) Pre-processing: first, we employ an NER model [19] to extract all entities; second, the entities are individually paired, tagged, and fed into the relation extraction model [20] as an input dataset; and third, the relation extraction model predicts relations among the entities, and the data predicted to have relations undergoes conditional extraction. (2) Extract conditions: the current approach relies on creating an annotated corpus and using rule extraction to identify sentences that match a set similarity threshold. (3) Information matching: the matching relation between conditions and triplets is then established. Alternatively, we propose an improved method that involves calculating attention weights based on syntactic and semantic information, generating hidden layer vectors, and performing binary classification to identify the relevant conditions for each triplet. This paper focuses on the third step, which is highlighted below.

The framework for matching conditional phrases and triplets is depicted in Figure 4; it consists of three sub-steps: (1) We fine-tune the pre-training model using the method in Section 4.1 and use the fine-tuned pre-training model to represent the sentences as hidden embeddings. (2) We use the method in Section 4.2 to obtain the semantic information of the sentence and combine it with the dependency distance vector generated by the dependency syntax matrix as input to the attention layer. (3) We used the method in Section 4.3 to calculate the final sentence vectors with the attention layer output. Then we extracted the vectors of entities and conditional phrases to evaluate the information-matching probability after calculation.

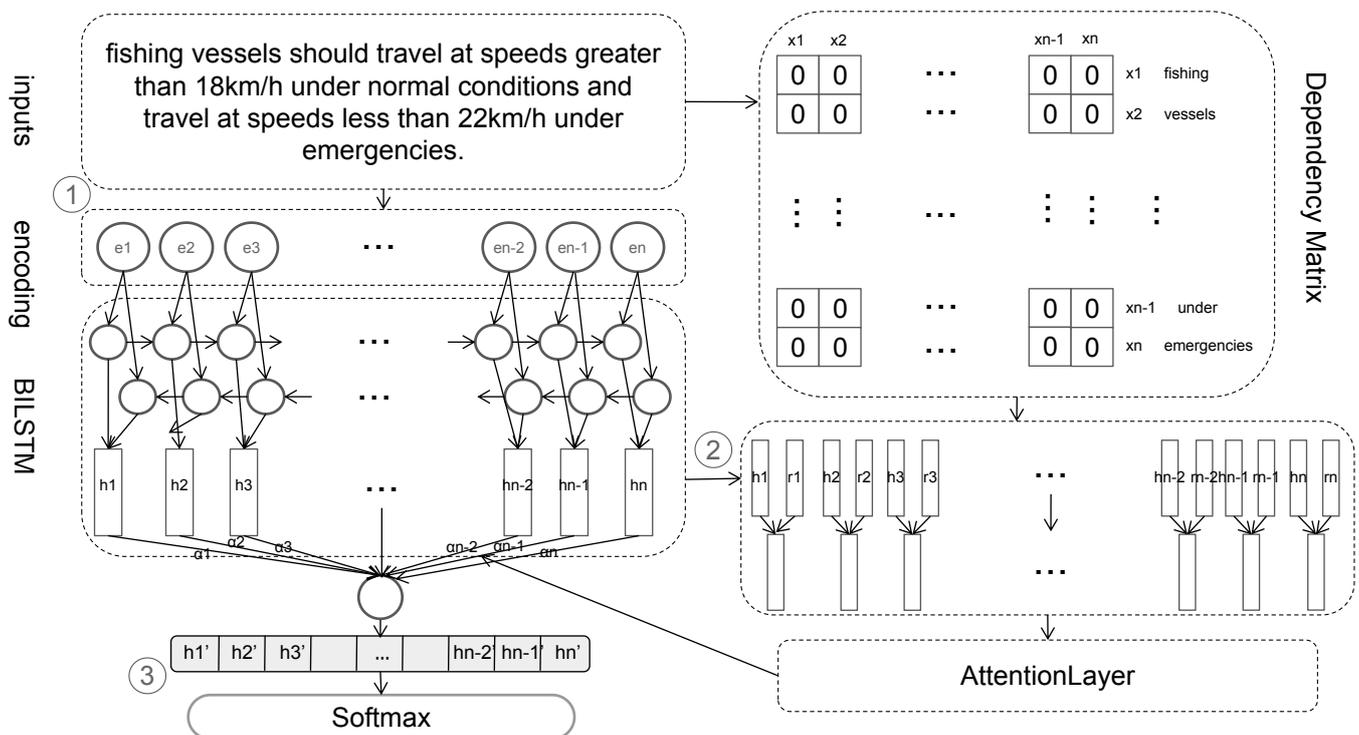


Figure 4. The framework for matching conditional phrases with triplets.

4.1. Local Conditional Context Enhancement

WoBERT [21] is a pre-training language model based on lexical refinement, reducing uncertainty in lexical meaning and modeling complexity. To learn more contextual information [22] for conditional phrases, this paper extends WoBERT as a novel pre-training model, condition-BERT (C-BERT). A detailed depiction of Step 1 in Figure 4 is shown in Figure 5. The “MLM” task below involves randomly selecting a certain percentage of tokens in a sentence and replacing these tokens with “MASK.” A classification model is then used to predict the word “MASK.” C-BERT modifies the mask mechanism of the “MLM” task [23] on top of the WoBERT full-word mask for the pre-training model task. Our approach is divided into two steps: first, we define the boundaries for the conditional phrases by marking the start and end positions; second, we mask them at the phrase and entity granularity, respectively.

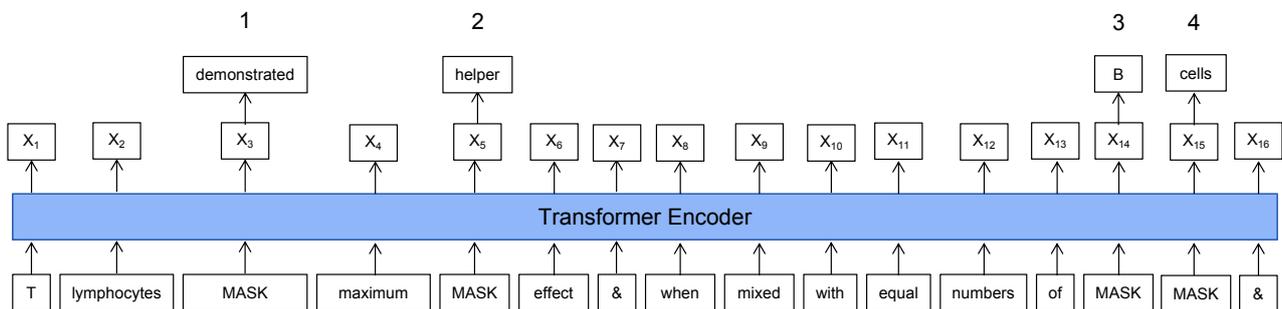


Figure 5. Example of a modified MLM task.

To mark the phrase granularity, we mask all the words in the conditional phrase except for deactivation. As shown in the Figure 6, given the original text as “under normal conditions,” the participle process with divide the text into three sequential words: “under,” “normal,” and “conditions,” after applying the mask mechanism in C-BERT, it becomes: “under **mask mask**.”

Original text	under	normal	conditions
Participle	under	normal	conditions
mask result	under	mask	mask

Figure 6. Sample sentence processing results: “under normal conditions”.

However, the long length of some complex conditional phrases may impact model training. So this paper adds entity granularity to the phrase granularity. It masks the entities attached to the conditional phrases. As shown in the Figure 7, given the original text as “when mixed with equal numbers of B cells,” the participle process with divide the text into nine sequential words: “when” “mixed” “with” “equal” “numbers” “of” “B” “cells,” after applying the mask mechanism in C-BERT it becomes: “when mixed with equal numbers of **mask mask**.”

Original text	when	mixed	with	equal	numbers	of	B	cells
Participle	when	mixed	with	equal	numbers	of	B	cells
mask result	when	mixed	with	equal	numbers	of	MASK	MASK

Figure 7. Sample sentence processing results: “when mixed with equal numbers of B cells”.

4.2. Sentence-Level Conditional Context Enhancement

Longer distance dependencies can be better captured using LSTM models, as LSTMs can learn which information should be remembered and which information should be forgotten through the training process. However, a further problem with modeling sentences using LSTM is the inability to encode information from back to front. At the finer granularity of classification, the semantic dependencies in both directions can be better captured using BiLSTM. Therefore, in this paper, this paper uses the BiLSTM network to extract semantic features [24] from the text and obtain the semantic features of each part of the text $H = (h_1, h_2, \dots, h_n)$, i.e.,

$$\vec{h}_t = LSTM(\vec{h}_{t-1}, c_t), \tag{1}$$

$$\overleftarrow{h}_t = LSTM(\overleftarrow{h}_{t-1}, c_t), \tag{2}$$

$$h_t = \vec{h}_t \oplus \overleftarrow{h}_t, \tag{3}$$

where \vec{h}_t indicates that the forward LSTM encodes the text information before moment t , and \overleftarrow{h}_t indicates that the reverse LSTM encodes the text information before moment t . Concatenate the features \vec{h}_t with \overleftarrow{h}_t to obtain the feature vector h_t obtained at moment t , and the text features H at each moment are obtained by calculation.

After obtaining the semantic features H of each part of the text, to highlight the role of the features of the semantic structure, this paper uses the “ltp” tool [25] to obtain the dependent syntactic structure of the sentences. The “ltp” tool supports basic Chinese natural language processing tasks, including Chinese word separation, lexical annotation, named entity recognition, and dependent syntactic analysis. The vector H is spliced with the dependency vectors r , obtained from the dependent grammatical relations, i.e.,

$$C = h_i \oplus r_i, \tag{4}$$

considering that different dependencies should have different weights, the attention mechanism is used to find the weight of the current dependency by using the dependencies present in the sentence [10] and the semantic information of the sentence as the basis for assigning weights by the attention mechanism, i.e.,

$$score_t = u^T \tanh(W_1[h_t \oplus r_t] + b_1), \tag{5}$$

$$\alpha_t = \exp(score_t) \left(\sum_i \exp(score_i) \right)^{-1}, \tag{6}$$

where h_t denotes the semantic features extracted at moment t . h_t is fused with r_t in the form of vector splicing; α_t denotes the attention weights at moment t . Where u , W_1 , and b_1 are all on-the-fly initialization parameters, parameter learning is performed based on error feedback from the output layer.

The attention weights for each component of $A=(\alpha_1, \alpha_2, \dots, \alpha_n)$ are obtained through calculation. Then the model adjusts the weight of semantic features in the feature vector

according to the attention weights, i.e., by weighting and summing the final feature vector to obtain the feature representation Z of the final sentence.

$$Z = \sum_i \alpha_i h_i, \tag{7}$$

4.3. Condition-Entity Correspondance Extraction with Syntactical Attention

For a text performing an entity relation extraction task, this paper aims to enable the BERT model to capture the semantic information of entity pairs with conditional phrases. In this paper, special positional markers "\$" are inserted at the start and end of entity 1, special positional markers "#" are inserted at the start and end of entity 2, and special positional markers "&" are inserted at the start and end of conditional phrases. A closer depiction of the step 3 method in Figure 4 is shown in Figure 8.

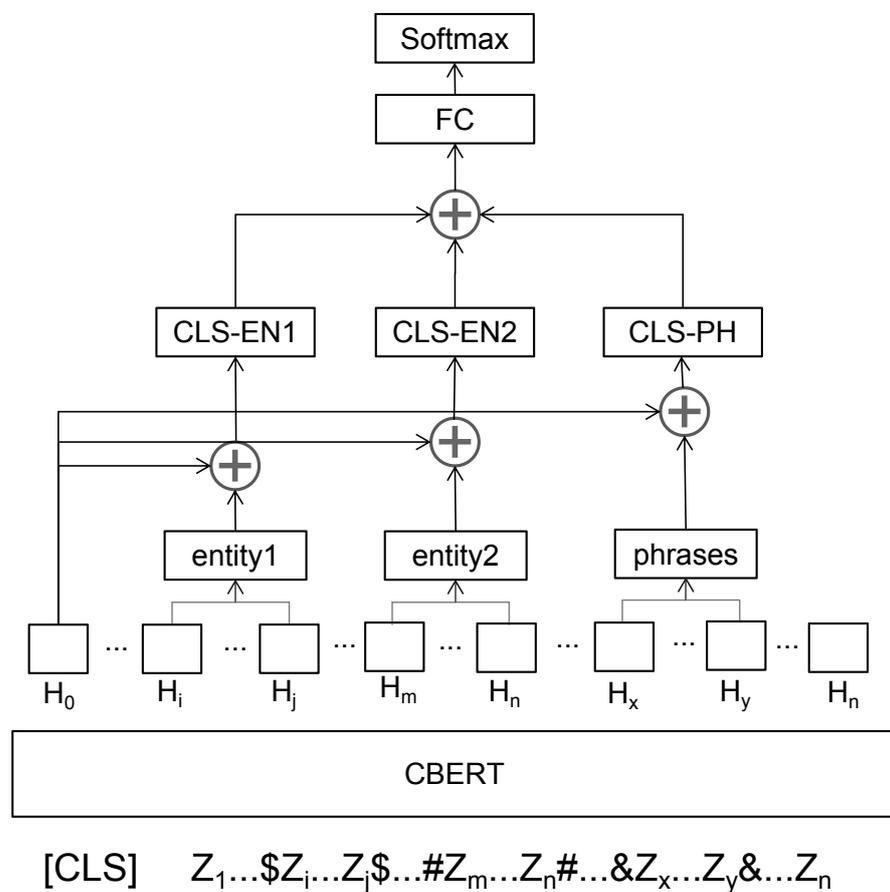


Figure 8. Conditional-entity vector extraction methods.

For a given sentence containing entity 1, entity 2, and conditionals, vectors Z_i to Z_j , vectors Z_m to Z_n , and vectors Z_x to Z_y are the vectors of the hidden layers of entity 1, entity 2, and conditional phrases, respectively, and then the corresponding vectors are taken out for averaging:

$$Z_{avg1} = \frac{1}{j - i + 1} \sum_{t=i}^j Z_t, \tag{8}$$

$$Z_{avg2} = \frac{1}{n - m + 1} \sum_{k=m}^n Z_k, \tag{9}$$

$$Z_{avg3} = \frac{1}{y - x + 1} \sum_{l=x}^y Z_l, \tag{10}$$

the vector is then vector summed with the vector Z_0 corresponding to “[CLS]” so that the final training hidden layer combines the head vector and the vector itself, using this information to make the BERT model encoding the output vector information enhanced:

$$Z_1 = Z_0 + Z_{avg1}, \quad (11)$$

$$Z_2 = Z_0 + Z_{avg2}, \quad (12)$$

$$Z_3 = Z_0 + Z_{avg3}, \quad (13)$$

the vectors of entities and conditions in the sentence are obtained from the above equations, and then the BERT-encoded CLS vectors are stitched with these three enhanced vectors to obtain the final vector Z_{final} , which is then fed directly into the fully connected layer for fine-tuning, and the corresponding relation is finally output via Softmax:

$$Z_{final} = Z_1 + Z_2 + Z_3, \quad (14)$$

$$Z_f = W \left[\tanh(Z_{final}) \right] + b, \quad (15)$$

$$p = \text{Softmax}(Z_f), \quad (16)$$

where W is the weight, b is the bias, and p is the probability corresponding to the final label. Finally, this model uses cross entropy as the loss function:

$$\text{Loss} = -\log p_k, \quad (17)$$

where the sample Label is k . In order to make the model converge more quickly, this paper uses the Adam method to train the model parameters. At the same time, to avoid overfitting, this paper uses dropout for processing.

5. Experiment

To validate the efficacy of our method, this paper tested its performance on different methods using both the Chinese industrial domain dataset and the English open dataset with similar parameter settings to the baseline model. In this section, this paper describes our experiments in detail.

5.1. Experimental Setup

5.1.1. Dataset Construction

The Chinese dataset “cpd-cn” is first separated into four components: title, content, table, and picture. The table and pictures are further processed to match the document’s structure. The “python-docx” tools identify and extract headings, body text, and tables, providing structural and content information. The content section is split into sentence-level information using special symbols. For vector extraction, position marks such as “<e1>,” “</e1>,” “<e2>,” “</e2>,” “<e3>,” and “</e3>” are inserted at the beginning and end of the entities and conditional phrases. For Chinese and English, the process of inserting labels is the same, and the specific algorithm process is as follows:

1. Take out the entity of the sentence in the output of the named entity recognition result;
2. Find the position of the entity in the sentence;
3. Insert the label corresponding to the entity before and after the position in step 2;
4. Input the sentence obtained in step 3 into the relationship extraction model;
5. Take out the sentence with the relationship in step 4, and perform conditional phrasing Extraction;
6. For the sentence in step 5 where the conditional phrase exists, find the position of the conditional phrase and perform label insertion.

Since BERT’s glossary does not recognize these marks, they are converted to “\$,” “#,” and “&.” Due to various structural problems in the document, some headings need to be

recognized, and manual help is needed for the modification. A rule-based plus manual annotation approach [26] extracts the corresponding conditions from the text. The process begins with manual annotation to build a condition corpus. The conditions extracted with rules are then matched with all the data in the condition corpus using text similarity. A threshold is set, and any data exceeding this threshold is considered the conditions for that text.

The English dataset “cpd-en” pairs factual triplets extracted from the published dataset in [17] with conditional triplets. The data are processed to clean up any factual and conditional-triplet noise, and pairwise labels are assigned. The resulting data include fact triplets, conditional triplets, and pairwise labels.

The cpd-cn and cpd-en datasets were divided into training, validation, and test sets using specific ratios, as outlined in Table 1, respectively. The experiment was repeated five times with small validation and testing sets to ensure accuracy, and the average results were reported.

Table 1. Percentage of data in the Chinese dataset cpd-cn and English dataset cpd-ed.

Dataset Name	Dataset Type	Amount (Sentence)	Percentage
cpd-cn	Training	413	80%
	Validation	103	10%
	Test	103	10%
cpd-en	Training	319	80%
	Validation	79	10%
	Test	79	10%

5.1.2. Baseline Method

This paper compares our optimized pre-training model to three other models: WoBERT on a cpd-cn dataset, BERT on a cpd-en dataset, and the MIMO + MATCH baseline model on the cpd-en dataset. The primary implementation method of MIMO focuses on extracting factual and conditional triplets. However, it does not include a method for matching triplets. To address this, we propose that MIMO + MATCH match conditional and factual triplets by computing a distance between the corresponding vectors. By using this distance, we can determine the matching relation between the two triplets.

5.1.3. Parameter Settings

The model parameters on the Chinese dataset cpd-cn and the English dataset cpd-en are set as shown in Table 2.

Table 2. cpd-cn and cpd-en corresponding model parameter settings.

Dataset Name	Parameter	Amount
cpd-cn	Batch size	16
	Max sentence length	256
	Warmup_proportion	0.1
	Dropout_rate	0.2
	Learning_rate	$1 \times 10^{-7} \sim 1 \times 10^{-5}$
	Epochs	10
cpd-en	Batch size	16
	Max sentence length	256
	Warmup_proportion	0.1
	Dropout_rate	0.1
	Learning_rate	$1 \times 10^{-7} \sim 1 \times 10^{-5}$
	Epochs	15

5.1.4. Evaluation Methods

For the conditional-phrase and triplet-matching task, performance is measured using the standard metrics, Accuracy (Acc), Precision (P), Recall (Rec), and F1 value (F1). The evaluation metrics for the pure relation extraction task are Equations (19)–(21), where TP indicates the number of positive samples correctly predicted by the model, FP indicates the number of positive samples incorrectly predicted by the model, and FN indicates the number of negative samples correctly predicted by the model. The evaluation metrics in the information matching task are Equations (18)–(21), where FP indicates that the classifier is identified as correct, but the condition has no matching relation with the triplet, and FN indicates that the classifier is identified as incorrect, but the condition has a matching relation with the triplet. Only the correctly extracted relationships form the basis for the evaluation of conditions.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%, \quad (18)$$

$$Precision = \frac{TP}{TP + FP} \times 100\%, \quad (19)$$

$$Recall = \frac{TP}{TP + FN} \times 100\%, \quad (20)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\%, \quad (21)$$

5.2. Results Analysis

In this section, this paper analyzes the results of each part of the experiment. It shows that the conditional phrase extraction task is accurate enough for subsequent tasks and that the overall performance of the optimized model in this paper outperforms other baselines. This paper has also carried out ablation experiments to verify the effectiveness of our method.

5.2.1. Phrase Extraction Results Analysis

The final results of this paper, obtained using the rule extraction phrase, phrase by phrase, and conditional phrase corpus [27] for the similarity matching method, are presented in Table 3. The table includes the correct rates of conditional phrase extraction for different similarity thresholds. Based on the experimental results, a threshold of 85% is selected for subsequent work, as it achieved the highest correct rate of 85.5% for conditional phrase extraction.

Table 3. Correct extraction rates corresponding to thresholds in the conditional phrase corpus. The bolded data indicates the best row of results.

Threshold	Accuracy
82.5%	84.5%
85%	85.5%
87.5%	82.4%

5.2.2. Model Results Analysis

Our optimized model was evaluated against the existing WoBERT model on the cpd-cn dataset, as shown in Table 4. The local conditional context enhancement proposed in this paper significantly improves the effectiveness of the information-matching method compared to the WoBERT model. We observed an improvement in Acc, F1, Recall, and Precision values by several percentages, further demonstrating the effectiveness of our approach.

Additionally, Table 4 also presents the evaluation of our optimized model against the existing and baseline models on the English dataset. The experimental data in Table 4 demonstrates that our method enhances the effectiveness of the information matching method through local conditional context enhancement on the cpd-en dataset. Moreover, we compared our optimized model with the baseline model MIMO-MATCH and achieved significant improvements in Acc, F1, Recall, and Precision values, thus proving the effectiveness of our method.

Table 4. Compares existing WoBERT models on the cpd-cn dataset with our optimized model, demonstrating the effectiveness of including an attention mechanism based on dependent syntactic analysis and semantic sentence information and comparison of the BERT model and baseline model with our optimized model on the cpd-en dataset. The bolded data indicates the best row of results.

Dataset Name	Models	Acc	F1	Rec	P
cpd-cn	<i>CBERT</i> + attention + <i>BiLSTM</i> (ours)	86.61%	86.65%	86.14%	87.17%
	<i>WoBERT</i> [21] + attention + <i>BiLSTM</i>	83.93%	83.21%	83.17%	83.24%
cpd-en	<i>CBERT</i> + attention + <i>BiLSTM</i> (ours)	82.29%	79.48%	79.27%	79.69%
	<i>BERT</i> + attention + <i>BiLSTM</i>	76.04%	73.17%	73.17%	73.17%
	<i>MIMO</i> [17]- <i>MATCH</i>	68.75%	66.33%	66.25%	66.41%

In this paper, the reasons for the different F1 values on the English and Chinese data sets are described below:

1. Because of the difference between Chinese and English word splitting, there is no separator in the middle of each sentence in Chinese compared to English; instead, there is a sequence of consecutive Chinese characters connected to form a sentence, which leads to inconsistencies in the word splitting tools. In the Chinese dataset, this paper adds some specialized words to the dictionary to ensure the correctness of the word splitting. The inconsistency of Chinese word separation may have a negative impact on the model, but the inclusion of some specialized words can improve the accuracy of the separation and have a positive impact on the model.
2. English utilizes a multitude of conjunctions and prepositions to supplement gaps between actual words, creating dependent and independent clauses to reflect various relations in a sentence. On the other hand, Chinese emphasizes meaning more than form, utilizing a looser sentence structure that relies on the relations between the order of words before and after rather than as much semantic glue between actual words.
3. The WoBERT model utilizes the Jieba split to remove redundant parts of BERT's word list while incorporating 20,000 Chinese words; it can improve the model's performance and positively impact the model.

5.2.3. Ablation Experiment

In our proposed approach, we conducted an ablation experiment to evaluate the effectiveness of combining dependent syntactic analysis with the attention mechanism and semantic information of sentences. Specifically, we evaluated our framework's performance with an attention mechanism and contextual sentence information. Moreover, we evaluated our fine-tuning methods. Results indicate that our proposed method significantly improves the performance of syntactic analysis in both settings, and fine-tuning methods are effective.

On the cpd-cn dataset, the result of Table 5 compared to Table 4 demonstrates that using sentence dependency grammatical relations to assign weights to the attention mechanism based on WoBERT significantly improves all indicators. It highlights the usefulness of pre-including sentence dependency grammatical relations followed by information matching. However, after we added BiLSTM to *WoBERT* + attention, the F1-value dropped by 0.95%. It is because WoBERT has difficulty identifying conditional contextual information. However, CBERT initially underperforms WoBERT with only the attention mechanism, but after adding the BiLSTM layer, the model's F1-value improved by 4.97%. It is because

the attention mechanism alone does not capture enough contextual information, but the BiLSTM layer enhances the model's ability to learn and leverage context information, demonstrating the effectiveness of the local conditional context enhancement.

Table 5. Results of ablation experiments with the WoBERT model on the cpd-cn dataset, and results of ablation experiments with the CBERT model on the cpd-cn dataset. The bolded data indicates the best row of results.

Dataset Name	Models	Acc	F1	Rec	P
cpd-cn	WoBERT	78.57%	79.40%	79.21%	79.60%
	WoBERT + attention	85.71%	84.16%	84.16%	84.16%
	WoBERT + attention + BiLSTM	83.93%	83.21%	83.17%	83.24%
cpd-cn	CBERT + attention	82.14%	81.18%	81.19%	81.18%
	CBERT + attention + BiLSTM	86.61%	86.65%	86.14%	87.17%

On the cpd-en dataset, Table 6's experimental results show that using BiLSTM to add semantic information leads to an F1-value drop by 7.47%, possibly due to the difficulty for BERT to recognize conditional contextual information. However, CBERT initially performed less well than BERT when using only the attention mechanism. However, with the addition of the BiLSTM layer, the model's F1-value improved by 2.6%. These results emphasize the importance of our local conditional context enhancement approach.

Table 6. Results of ablation experiments with the BERT model on the cpd-en dataset, and results of ablation experiments with the CBERT model on the cpd-en dataset. The bolded data indicates the best row of results.

Dataset Name	Models	Acc	F1	Rec	P
cpd-en	BERT + attention	81.25%	80.78%	79.27%	82.35%
	BERT + attention + BiLSTM	76.04%	73.31%	73.17%	73.45%
cpd-en	CBERT + attention	80.21%	76.88%	76.83%	76.94%
	CBERT + attention + BiLSTM	82.29%	79.48%	79.27%	79.69%

In addition, we conducted an ablation experiment to assess the effectiveness of model fine-tuning using phrase and entity granularity methods. The experiment compared the performance of the model using only phrase granularity with that using both phrase and entity granularity on the cpd-cn and cpd-en datasets, respectively. The results in Table 7 demonstrate that the model performs better when entity granularity is added, proving the effectiveness of our approach.

Table 7. The WoBERT model was fine-tuned on the cpd-cn dataset using phrase and entity granularity, while the BERT model was fine-tuned on the cpd-en dataset using both phrase and entity granularity. The bolded data indicates the best row of results.

Dataset Name	Models	Acc	F1	Rec	P
cpd-cn	WoBERT + Phrase granularity	83.04%	81.18%	81.19%	81.18%
	WoBERT + Phrase and entity granularity	86.61%	86.65%	86.14%	87.17%
cpd-en	BERT + Phrase granularity	78.13%	76.74%	76.83%	76.67%
	BERT + Phrase and entity granularity	82.29%	79.47%	79.27%	79.69%

5.2.4. Case Study

This paper presents a case study that validates the weighting of attention assigned by the proposed method. Figure 9 illustrates the attention score of our framework on a

random sentence: “fishing vessels should \$ travel at speeds \$ greater than # 18 km/h # & under normal conditions & and travel at speeds less than 22 km/h under emergencies.” After incorporating syntactic analysis of dependency and sentence context information, the experimental results demonstrate that high attention scores are primarily distributed among words with dependency relations. Adding markers makes some vocabulary clearer, even though not all are apparent. The implication of the attention score is to allow the model to adjust the weight of the semantic features at each moment in the final feature vector of the sentence according to the attention weights, which can highlight the role of words with high relevance in the matching results. From the results, it can be inferred that the model can focus on words with specific dependency relations in the classification process.

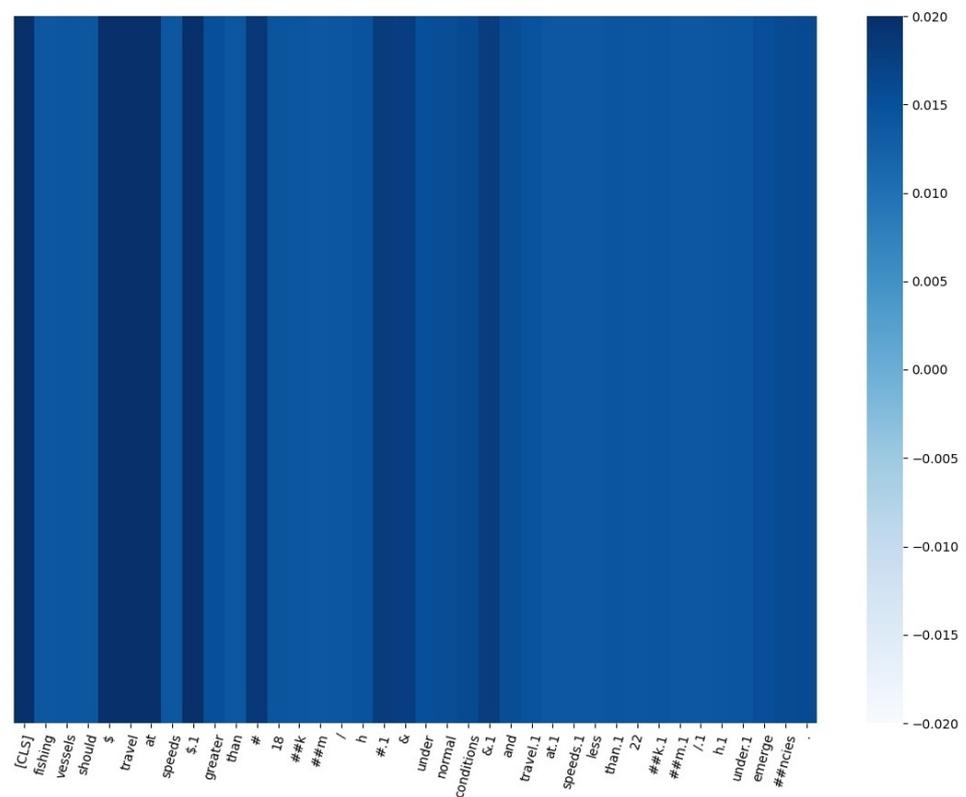


Figure 9. The attention scores for the keywords, random sentence is “fishing vessels should \$ travel at speeds \$ greater than # 18 km/h # & under normal conditions & and travel at speeds less than 22 km/h under emergencies.”

5.2.5. Error Analysis

We analyze the errors based on the matching results on the test set. On the one hand, the model achieves better performance for conditional phrase patterns that frequently appear in the training corpus, such as “when . . .” For less frequent conditional phrase patterns, such as “compared with . . .,” the model’s results are unsatisfactory. On the other hand, the model’s results in this paper are somewhat unsatisfactory for cases where the sentences contain too many triplets and conditions. We will try to use more effective methods to solve this problem in our future work.

6. Conclusions and Future Work

This paper proposes a framework for constructing conditional knowledge by adding a novel conditional-phrase extraction and conditional-phrase matching task on top of the conventional relation extraction. In the conditional phrase extraction task, a phrase-matching corpus is built, which uses phrase similarity matching to extract the conditional phrases to be extracted from the sentences. After the conditional phrases are extracted, this paper

uses conditional boundaries to fine-tune the existing pre-training model and later uses the dependency grammar analysis technique combined with the attention mechanism to match the conditional phrases with the triplets extracted from the information-extraction technique. Experiment results show that our fine-tuning approach has to boost implications in downstream tasks while improving the accuracy of the matching task by 2.68%, compared to the baseline. The dependency grammar analysis technique combined with the attention mechanism has boosting implications for matching conditional phrases to triplets while improving the accuracy of the matching task by 4.47%, compared to the baseline.

While the method proposed in this paper improves the model's effectiveness, there is still room for improvement. For the conditional phrase extraction task in this paper, the paper uses rule extraction, manual annotation of a conditional phrase corpus, and sentence similarity matching, which consumes many human resources. To address this problem, this paper will later incorporate artificial intelligence to extract highly accurate conditional phrases without using as many human resources. As the task in this paper is based on a relation extraction task, the task accuracy also depends on the relation extraction task accuracy. To address this problem, multi-task learning of the relationship extraction task and the information matching task will be implemented later in this paper, hoping to optimize the relation extraction task and the information matching task in this paper. In the subsequent task knowledge graph construction, as the conditional phrases in the sentences are extracted in this paper, it is equivalent to adding a new element to the ternary group to form a quadruple. Therefore, it creates information storage problems during the knowledge graph construction. In response to this problem, the paper will address the problems in the subsequent work on the knowledge graph construction task.

Author Contributions: Conceptualization, Z.X., B.Z., J.G. and F.G.; methodology, Z.X.; software, Z.X.; validation, Z.X., J.G. and F.G.; formal analysis, Z.X.; investigation, B.Z., J.G. and F.G.; data curation, Z.X.; writing—original draft preparation, Z.X.; writing—review and editing, J.G., B.Z. and F.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Natural Science Foundation of China (Grant No. U1836118), the National key research and development program (Grant No. 2020AAA0108500), and the Key Research and Development Program of Wuhan (Grant No. 2022012202015070).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Reference

1. Zhou, S.; Yu, B.; Sun, A.; Long, C.; Li, J.; Sun, J. A Survey on Neural Open Information Extraction: Current Status and Future Directions. In Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23–29 July 2022; pp. 5694–5701. [\[CrossRef\]](#)
2. Yang, J.; Han, S.C.; Poon, J. A survey on extraction of causal relations from natural language text. *Knowl. Inf. Syst.* **2022**, *64*, 1161–1186. [\[CrossRef\]](#)
3. Wei, Z.; Su, J.; Wang, Y.; Tian, Y.; Chang, Y. A Novel Cascade Binary Tagging Framework for Relational Triple Extraction. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics 2020, Online, 5–10 July 2020; Jurafsky, D., Chai, J., Schluter, N., Tetreault, J.R., Eds.; Association for Computational Linguistics: Cedarville, OH, USA, 2020; pp. 1476–1488. [\[CrossRef\]](#)
4. Xu, W.; Zhang, C. Trigger word mining for relation extraction based on activation force. *Int. J. Commun. Syst.* **2016**, *29*, 2134–2146. [\[CrossRef\]](#)
5. Culotta, A.; Sorensen, J.S. Dependency Tree Kernels for Relation Extraction. In Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, Barcelona, Spain, 21–26 July 2004; Scott, D., Daelemans, W., Walker, M.A., Eds.; Association for Computational Linguistics: Cedarville, OH, USA, 2004; pp. 423–429. [\[CrossRef\]](#)
6. Yu, X.; Jin, Z. Web content information extraction based on DOM tree and statistical information. In Proceedings of the 2017 IEEE 17th International Conference on Communication Technology (ICCT), Chengdu, China, 27–30 October 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1308–1311.

7. Zeng, D.; Liu, K.; Lai, S.; Zhou, G.; Zhao, J. Relation Classification via Convolutional Deep Neural Network. In Proceedings of the COLING 2014, 25th International Conference on Computational Linguistics, Dublin, Ireland, 23–29 August 2014; Hajic, J., Tsujii, J., Eds.; Association for Computational Linguistics: Cedarville, OH, USA, 2014; pp. 2335–2344.
8. Zhou, P.; Shi, W.; Tian, J.; Qi, Z.; Li, B.; Hao, H.; Xu, B. Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, Berlin, Germany, 7–12 August 2016; Volume 2. [\[CrossRef\]](#)
9. Lin, Y.; Shen, S.; Liu, Z.; Luan, H.; Sun, M. Neural Relation Extraction with Selective Attention over Instances. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, Berlin, Germany, 7–12 August 2016; Volume 1, Long Papers. [\[CrossRef\]](#)
10. Li, D.; Yan, L.; Yang, J.; Ma, Z. Dependency syntax guided BERT-BiLSTM-GAM-CRF for Chinese NER. *Expert Syst. Appl.* **2022**, *196*, 116682. [\[CrossRef\]](#)
11. Ye, Z.; Zhao, H. Syntactic word embedding based on dependency syntax and polysemous analysis. *Front. Inf. Technol. Electron. Eng.* **2018**, *19*, 524–535. [\[CrossRef\]](#)
12. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All you Need. In Proceedings of the Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
13. Devlin, J.; Chang, M.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2018**, arXiv:1810.04805.
14. Tong, P.; Zhang, Q.; Yao, J. Leveraging Domain Context for Question Answering Over Knowledge Graph. *Data Sci. Eng.* **2019**, *4*, 323–335. [\[CrossRef\]](#)
15. Maksimov, N.; Golitsyna, O.; Lebedev, A. Knowledge Graphs in Text Information Retrieval. In *Biologically Inspired Cognitive Architectures 2021, Proceedings of the 12th Annual Meeting of the BICA Society, Online, 19–21 September 2021*; Springer: Cham, Switzerland, 2022; pp. 268–274.
16. Jiang, T.; Zhao, T.; Qin, B.; Liu, T.; Chawla, N.V.; Jiang, M. The role of “condition” a novel scientific knowledge graph representation and construction model. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 1634–1642.
17. Jiang, T.; Zhao, T.; Qin, B.; Liu, T.; Chawla, N.; Jiang, M. Multi-input multi-output sequence labeling for joint extraction of fact and condition tuples from scientific text. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019.
18. Zheng, T.; Xu, Z.; Li, Y.; Zhao, Y.; Wang, B.; Yang, X. A Novel Conditional Knowledge Graph Representation and Construction. In Proceedings of the Artificial Intelligence—First CAAI International Conference, CICA 2021, Hangzhou, China, 5–6 June 2021; Proceedings, Part II; Fang, L., Chen, Y., Zhai, G., Wang, Z.J., Wang, R., Dong, W., Eds.; Springer: Cham, Switzerland, 2021; Volume 13070, pp. 383–394; Lecture Notes in Computer Science. . .32. [\[CrossRef\]](#)
19. Syafiq, M.I.; Talib, M.S.; Salim, N.; Haron, H.; Alwee, R. A Concise Review of Named Entity Recognition System: Methods and Features. In Proceedings of the IOP Conference Series: Materials Science and Engineering, Bangkok, Thailand, 4–5 February 2019; IOP Publishing: Bristol, UK, 2019; Volume 551, p. 012052.
20. Wu, S.; He, Y. Enriching Pre-trained Language Model with Entity Information for Relation Classification. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM 2019, Beijing, China, 3–7 November 2019; Zhu, W., Tao, D., Cheng, X., Cui, P., Rundensteiner, E.A., Carmel, D., He, Q., Yu, J.X., Eds.; ACM: New York, NY, USA, 2019; pp. 2361–2364. [\[CrossRef\]](#)
21. Su, J. Wobert: Word-Based Chinese Bert Model-Zhuyiyai. Technical Report. 2020. Available online: <https://github.com/ZhuyiTechnology/WoBERT> (accessed on 14 March 2023).
22. Zhao, H.; Lu, J.; Cao, J. A short text conversation generation model combining BERT and context attention mechanism. *Int. J. Comput. Sci. Eng.* **2020**, *23*, 136–144. [\[CrossRef\]](#)
23. Devlin, J.; Chang, M.; Lee, K.; Toutanova, K. 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, NAACL-HLT 2019, Minneapolis, MN, USA, 2–7 June 2019; Burstein, J., Doran, C., Solorio, T., Eds.; Association for Computational Linguistics: Cedarville, OH, USA, 2019; Volume 1; pp. 4171–4186. [\[CrossRef\]](#)
24. Turton, J.; Smith, R.E.; Vinson, D.P. Deriving Contextualised Semantic Features from BERT (and Other Transformer Model) Embeddings. In Proceedings of the 6th Workshop on Representation Learning for NLP, RepL4NLP@ACL-IJCNLP 2021, Online, 6 August 2021; Rogers, A., Calixto, I., Vulic, I., Saphra, N., Kassner, N., Camburu, O., Bansal, T., Shwartz, V., Eds.; Association for Computational Linguistics: Cedarville, OH, USA, 2021; pp. 248–262. [\[CrossRef\]](#)
25. Che, W.; Feng, Y.; Qin, L.; Liu, T. N-LTP: An Open-source Neural Language Technology Platform for Chinese. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, EMNLP 2021, Online and Punta Cana, Dominican Republic, 7–11 November 2021; Adel, H., Shi, S., Eds.; Association for Computational Linguistics: Cedarville, OH, USA, 2021; pp. 42–49. [\[CrossRef\]](#)

26. Achlow, I.; Zeichner, N.; Kneller, H. System and Method for Automatic Key Phrase Extraction Rule Generation. U.S. Patent 11,507,743, 22 November 2022.
27. Barkan, O.; Razin, N.; Koenigstein, N. Sentence Similarity Scoring Using Neural Network Distillation. U.S. Patent 11,392,770, 19 July 2022.

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.