



Yujiang Liu^{1,2,*}, Lijun Fu², Xiaojun Xia^{1,2} and Yonghong Zhang³

- ¹ University of Chinese Academy of Sciences, No.1 Yanqihu East Rd., Huairou District, Beijing 101408, China; xiaxj@sict.ac.cn
- ² Shenyang Institute of Computing Technology Co., Ltd., Chinese Academy of Sciences, No. 16 Nanping East Rd., Dongling District, Shenyang 110168, China; fu_lijun@ucas.ac.cn
- ³ Laboratory of Big Data and Artificial Intelligence Technology, Shandong University, Jinan 250100, China; 202084000165@email.sdu.edu.cn
- * Correspondence: liuyujiang16@mails.ucas.ac.cn; Tel.: +86-17604010961

Featured Application: This paper proposes a method that can increase the connectivity of the knowledge graph when it is constructed and improve usability, e.g., retrieving more knowledge at the time of query.

Abstract: Extracting relation quintuple and feature words from unstructured text is a prelude to the construction of the scientific knowledge base. At present, the prior works use explicit clues between entities to study this task but ignore the use and the association of the feature words. In this work, we propose a new method to generate self-adaptive feature words from the original text for every single sample. These words can add additional correlation information to the knowledge graph. We allow the model to generate a new word representation and apply it to the original sentence to judge the relation type and locate the head and tail of the relation quintuple. Compared with the previous works, the feature words increase the flexibility of relying on information and improve the explanatory ability. Extensive experiments on scientific field datasets illustrate that the self-adaptive feature words method (SAFW) is good at ferreting out the unique feature words and obtaining the core part for the quintuple. It achieves good performance on four public datasets and obtains a markable performance improvement compared with other baselines.

Keywords: relation quintuple; scientific knowledge; feature words

1. Introduction

Automatically extracting relational quintuples is a typical task in information extraction, which is widely used in vertical domains. Especially in various specialized fields, it can help nonspecialists to understand how cleaning evolved and remains relevant. Quintuples are considered meta-information for knowledge-base construction, including the type of head entity, the name of the head entity, the type of relation, the type of tail entity, and the name of the tail entity. This kind of information can form a knowledge graph by the merging of the same entity and relation, such as the vertex and the edge. This task has been labeled "Rel+" in some papers, in contrast to the entity–relationship joint extraction task, which does not consider unrelated entities when extracting quintuples. This can be seen in the recent datasets, such as scientific papers [1], the financial domain [2], biomedical information [3], and energy data [4].

Feature words play an important role in the relation quintuple extraction task. These words act as triggers to guide relational ternary extraction [5]. We define a feature word as a key word that assists in the extraction of relational quintuples and establishes the relevance of the original sentence after the construction of a knowledge graph. Thus, the feature words decide which contribute to the specific type and act as a kind of relation



Citation: Liu, Y.; Fu, L.; Xia, X.; Zhang, Y. Exploring the Role of Self-Adaptive Feature Words in Relation Quintuple Extraction for Scientific Literature. *Appl. Sci.* **2024**, *14*, 4020. https:// doi.org/10.3390/app14104020

Academic Editor: David Ruano Ordás

Received: 12 April 2024 Revised: 3 May 2024 Accepted: 7 May 2024 Published: 9 May 2024



Copyright: © 2024 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/). prompt. These feature words are added to the knowledge graph interface to increase the degree of connectivity, allowing users to obtain more relevant information when querying the knowledge graph. Thus, the task is defined as the extraction of relational quintuples and feature words from each sentence in the dataset. Similar to this idea, Yan et al. [6] used feature words as boundaries to restrict the position of entities during extraction. However, these labeling processes consume resources due to boundary labeling. Zhu et al. [7] used feature words for relational ternary extraction in the urban rail domain, but they still used the hidden semantics of the feature words without feature word labeling, instead of outputting the feature words themselves with the supervision method. As a result, there was no improvement in graph connectivity at the time of graph construction. From the above, it is known that the feature words can provide self-enhancing information for the extraction process. However, since these feature words are unlabeled, we consider using "associative supervision" [8] to extract these words during quintuple extraction, unlike existing semi-supervised schemes. Using such a supervised method, we are the first to propose the direct output of feature words.

The word-based approach provides a model design perspective that influences the relationship classification results. Because of the pre-trained model and the large language model, the semantic features are no longer the bottleneck; therefore, it is very important to retrieve the text expression or separate the targets from the sentence. In Li's work, the relation results of the expression and the predefined templates were encoded to generate a fine-grained semantic representation [9], which injects more link features into the training model. Learning the features from existing results [10] can generate global features and enrich the pattern information in the current sample. In addition, when the training data change and the head and tail entities are arranged in order and disorder, the ordering of the two entities in the quintuple can be learned [11]. Remarkable improvements in the word-based solution show that both the modified model and the refined data are required. This approach can be of use in the extraction of feature words in conjunction with relational quintuples.

Despite its success in using words, this approach still has a number of shortcomings in setting fixed words as feature words. The fixed words, regarded as templates, are not flexible enough in their meaning, which may result in redundant or missing information. This problem is simply compensated for by the feature words drawn from the original sample, considered as self-adaptive feature words. Figure 1 shows the fixed feature words improved by these words. In order to increase the relevance of the knowledge graph, we link the feature words to the origin graph. Figure 2 shows the comparison of those with and those without feature words. Here, we use examples of extracting relational quintuples in common scenarios to help us to understand them. In fact, in the knowledge graph constructed from the scientific texts, these feature words can play the same role as informational cues. They can establish connections with professional knowledge and help users to understand the relevant work in the field. Therefore, using the associative supervision method to improve the recognizability of the samples is an appropriate solution.

Sample Text	The owner of the FC Magdeburg is Uderzo Rene who play making Berlin AK 07 also belong to the football 1. FC Union Jakarta.						
	Name of Head Entity	Type of "	Type of "head entity/relation/tail entity" Name of Tail Entity				
Relation	1.FC	org	organization/manager/person Uderzo				
Quintuple	Uderzo Rene	р	Berliner AK 07				
	Uderzo Rene	р	erson/club/organization	1 . FC Union Jakarta			
Fixed Feature	Name of the Ty	pes	Fixed Words of the Types				
Words	organization/manage	er/person	have a organization/manager/person relationship with				
vvorus	person/club/organ	ization	have a person/club/organization relationship with				
Self-Adaptive	owner belong playing making football						
Feature Words							

Figure 1. Cases with the improvement of self-adaptive feature words. The fixed words (purple areas) do not provide enough clues for extracting relation quintuples. However, the combination with self-adapting words (red areas) improves the situation.



Figure 2. The comparison with and without feature words. The upper part shows a knowledge graph without feature word links. If there is no feature word, the graph is divided into 3 subgraphs. The bottom part shows a knowledge graph with several feature words. The number of sub-graphs becomes 1 after the feature words are added. This indicates that, when the graph is constructed, the extracted feature words can increase the correlation.

In this paper, we begin our work with the goal of enabling the self-adaptive feature words method (SAFW) to find descriptions of problem solutions. SAFW aims to explore

how to extract both relational quintuple and sample feature words. To the best of our knowledge, this is the first hybrid supervised method for the extraction task of multiple targets. Our approach is to allow SAFW to generate specific feature words for each sample by extracting words from the source sentences and guiding the extraction of these feature words with the help of the quintuple results. To extract useful information as keywords from the distorted sentences, we propose a multi-turn vector resampling module to redistribute the weights of the words and combine them with the initial weights to obtain a unique set of candidates. As a result, SAFW can extract adaptive feature words for each input and compute them for target discrimination. Specifically, after sentence encoding, we use a pointer network [12] to iteratively down-sample the words and generate combinatorial vectors to represent them. Then, we use a fixed template filling strategy to find the head and tail entities, as in the work of Li et al. [9]. Since these feature words obtained after sampling are again used as inputs for relational quintuple extraction, we can assume that they are supervised by the quintuples. Such re-arrangement, re-screening, and the application of external data [13] inspire our method. This improves the generalization performance and removes noise from our training data. Lexical applications such as guiding words [14,15], robustness [16,17], and low data resources [18,19] have inspired similar approaches of this type.

In short, SAFW can be summarized as the generation of feature words by each training sample first, and then combining them with fixed template matching to find the relation and entity parts. The experiments show that SAFW significantly outperforms the existing methods. The contributions of this paper are as follows:

We introduce a novel viewpoint that generates feature words from all of the dataset and its derived data to guide the model to extract the relation quintuple. For training data, this is a simple and data-saving supplementary mechanism.

We propose the idea of associative supervision, which combines unsupervised learning with supervised learning. The supervision of unlabeled results can be achieved by a supervised process. This is a source reduction idea that can effectively reduce labeling labor and resources.

We propose a new method to select some down-sampled keywords and redistribute their weight to increase the differences for the same words or phrases in different samples. This hybrid supervising strategy, which incorporates two objectives, can eliminate the problem of conflicting optimizing directions that arise from multiple supervising in traditional models.

Extensive experiments on four public datasets of three scientific ones and two general ones show that the proposed method achieves state-of-the-art results. The ablation studies prove that our proposal is feasible.

In this paper, we introduce the concept and the innovative approach of relation quintuple extraction for the scientific literature in Section 1; explain the algorithmic flow in Section 2; show the results of the basic experiments, the ablation experiments, the case study, and the quantitative analysis in Section 3; and summarize the full text in Section 4.

2. Materials and Methods

2.1. Problem Overview

We regard this problem as a union task of feature words and a relation quintuple extraction task, with a sentence of $S = \{w1, w2, ..., wn\}$ as an input, feature word indexes of $I(S) = \{i1, i2, ..., it\}$, and relation quintuples of $T(S) = \{(s, t, o)\}$ as outputs, where $s, o \in E$ are the tokens of the head entities and tail entities in the sentence, connected by the union types $t \in T$. The length of the sequence is n. The total t feature words are extracted. The E is the set of the entities. The T contains the type of head and tail entities and the type of relationship between them. An overview of our method is shown in Figure 3.



Figure 3. The structure of SAFW. The sequence $S = \{w1, w2, ..., wn\}$ is converted into relationencoded sequence V^r and entity encoded sequence V^e by the multi-turn vector resampling module and linear mapping function. Then, the two sequences are merged with the union-type candidates to select the entities used to form the relation quintuples. The resampling selects *t* feature words within *t* times. The mask only contains 1 and 0, indicating whether the corresponding words are available or not (green means available, while red means not available). The available indexes are all tokens, except [CLS] and [SEP], and are masked one by one until the time *t* or all-covered.

2.2. Encoder

The encoder gives the token representation of the samples in the dataset. We chose the BERT model [20] for the sake of the prior knowledge. It is designed to obtain deep representation and outputs a sequence of token vector, $V = \{v_{cls}, v_1, \ldots, v_n, v_{sep}\}$, which can be used in the re-sampling process. For BERT, the v_{cls} and the v_{sep} are the vectors of the special tokens [CLS] and [SEP] located at the beginning and end of the sentence.

2.3. Multi-Turn Vector Resampling Module

The multi-turn vector resampling module is designed to extract feature words and optimize the capturing effect of the input by changing the output scale of all of the vectors in V. This module selects a number of k indexes for the feature words with loops of $t(0 \le t < k)$. These positions can represent the corresponding feature words, so as to regard them as self-information enhancement. Using a brand-new multi-loop method, the encoder can collect different semantic states in all t steps as the next inputs. When t = 0, we choose v_{cls} as v_{cur} and a zero vector as v_{pre} . These vectors are combined into a new sequence, which is the input of a two-layer transformer network. It outputs a re-encoded fusion result h_T from the source representing the semantic information that can be viewed

as a transition from the previous state to the current state. Note that the equations with (t) are calculated in the loop and with the process of the t turn, as follows:

$$h_T(t) = Transformer\left(V_{pre(t)}, V_{cur(t)}\right)$$
(1)

$$V_{pre(t+1)} = h_T(t) \tag{2}$$

where h_T is copied n times as nh_T , so that they can be merged and keep the same length as the input data. In the t + 1 stage, the value will be fed into v_{pre} . In time t = 0, we also set a 0 vector within V length as the initial sequence mask M(t). If a word is selected, the mask at the corresponding position changes to 1 in the next round, which means that the word is masked in the calculation.

The pointer attention decoder reallocates weights for all of the vectors in the sequence, which is the key solution for feature word extraction rendered by the samples. With the progress of training, the weights of SAFW become more and more reasonable, and the vector selection is adjusted appropriately at each time step. This novel strategy ensures that the newly generated probability distribution depends on the previous step value, thus improving the fault tolerance of capturing an uncertain number of keywords. We employ a linear mapping function to transform all of the inputs into the same shape. In this way, the result will be bound with V (without v_{cls} and v_{sep}) and sent into a nonlinear activation function to obtain the possibility. The maximum value of the possibility points to the selected position in the current round, leaving the next stage result, and adding it to the probability group α as $\alpha(t)$. The max-pointer position represents the extracted words, while the source represents all of the basic vectors, implemented by scatter and max functions and changing the α to one-hot form, as follows:

$$h_p(t) = W_p concat(ch_T(t), V) + b_p$$
(3)

$$\alpha(t) = Softmax(M(t) + W_x(tanh(h_p(t))) + b_x)$$
(4)

$$onehot\alpha(t) = scatter(max(\alpha(t)))$$
(5)

$$M(t+1) = M(t) + onehot\alpha(t)$$
(6)

where W_p is the mapping trainable parameters of the linear function and W_x is the trainable weights for integrating the sum of the two source vectors. Here, M(t) is updated by onehota(t). This mask growth mechanism is an unsupervised way to prevent the model from selecting the words that it has selected before. The maximum value in time t will no longer be the maximum value in time t + 1, because we add -1×10^9 for every value of 1, illustrating that the mask can cover the chosen ones in all time steps. We designed the mask in the following way, referring to the work of Liao et al. [21] and improving it. They proposed a variable mask in a generative model. The mask on a certain position is drawn from a probabilistic distribution and exchanged by specific words in the generation phase. In addition, they prepared a set to mark the pointed positions and ignore them afterwards. We consult these two features of the variable mask.

The output of the final distribution is the sum of the max-pointer position probability and the source probability. We enhance the contribution of the selected words and keep the source part smoother, reconciling the local importance with the global sequence. In addition, the result is also used to calculate the next current state $v_{cur}(t + 1)$, as follows:

$$p(t) = \alpha(t) + onehot\alpha(\alpha(t))$$
(7)

$$v_{cur}(t+1) = Maxpool(p(t) * V)$$
(8)

where the p(t) is the output probability in step t. We adopt max-pooling to obtain the $v_{cur}(t+1)$ value for the multi-round process.

At the end of the iteration, t=k, the entity representation $V^{(e)} = \left\{ v_{cls}^{(e)}, v_1^{(e)}, \ldots, v_p^{(e)}, v_{sep}^{(e)} \right\}$, and the relation representation $V^{(r)} = \left\{ v_{cls}^{(r)}, v_1^{(r)}, \ldots, v_p^{(r)}, v_{sep}^{(r)} \right\}$ are multiplied by the probability with different linear mapping. Because we use a batched index selection function to find all possible entity spans to combine into this form, the end index of the sequence is p, not n. Thus, the vectors in $V^{(e)}$ and $V^{(r)}$ are converted into span-based ones, such as $v_i^{(e)} = \left[v_{start}^{(e)}(i); v_{end}^{(e)}(i); \theta^{(e)}(v_i) \right]$, where the $v_{start}^{(e)}(i)$ and $v_{end}^{(e)}(i)$ are the boundary vector of entity i and $\theta^{(e)}(v_i)$ is the length of it. The feature word indexes I are collected by a container and output after concatenating, as follows:

$$V^{(e)} = ent(W_e(p(k) * V) + b_e)$$
⁽⁹⁾

$$V^{(r)} = ent(W_r(p(k) * V) + b_r)$$
(10)

$$I = max(concat(onehot\alpha(\alpha(1)), onehot\alpha(\alpha(2)), \dots, onehot\alpha(\alpha(k-1))))$$
(11)

where ent is the entity selection function and the W_e and W_r are the trainable weights for generating the entity and relation. With this method, we extract the feature words by indexing the mapping to the original text. In addition, we modify the vector distribution for subsequent calculations. This result is supervised with the help of relational quintuples.

2.4. Relation Matching Module

The priority of union-type extraction can reduce the result space of relation quintuples, which inspired us to design a candidate set to find all possible results. We allow all of the union types U to be inputs and obtain the representation of self-attention. Another self-attention layer is used to integrate and enhance the semantic information in $V^{(r)}$. After encoding, we introduce $V^{(r)}$ as a key and a value into the cross attention module. The module's query is a union-type representation, which keeps the same shape as the output. After the calculation of the attentions, we expand both of the results and form an interactive matrix $X^{(r)}$, which represents full mapping of the union-type candidates and the vectors in the representations. Here, we denote the result of the expanded as $X^{(r)} \in \mathbb{R}^{(m \times d)}$, where m is the number of the total union types, as follows:

$$U^{(r)} = crossatt\left(selfatt(U), selfatt\left(V^{(r)}\right)\right)$$
(12)

$$X^{(r)} = expand\left(U^{(r)}\right) + expand\left(\left[V^{(r)}\right]\right)$$
(13)

$$T = Sigmoid\left(W_T X^{(r)} + b_T\right) \tag{14}$$

where $U^{(r)}$ is the result of the cross attention and T is the sigmoid mapping value, which indicates whether the type is true by a threshold. This will guide the feature word generation based on union types in the next module.

2.5. Entity Matching Module

Since the results T are known, we use them as inputs and obtain entity pairs in this module. The results are prepared in the training phase, but they must be converted from T during the test. Similar to the union type, we choose the same method to obtain the pairs. Assuming that we find a number of j union types in m, we will supervise the two matrices that represent the head entity record $X^{(eh)} \in \mathbb{R}^{(j \times d)}$ and the tail entity record $X^{(et)} \in \mathbb{R}^{(j \times d)}$. If the corresponding positions of both are activated, it means that they can form a relationship, as follows:

$$T^{(e)} = crossatt\left(selfatt(T), selfatt\left(V^{(e)}\right)\right)$$
(15)

$$X^{(eh)} = W_{eh}\left(expand\left(U^{(r)}\right) + expand\left(\left[V^{(r)}\right]\right)\right) + b_{eh}$$
(16)

$$X^{(et)} = W_{et}\left(expand\left(U^{(r)}\right) + expand\left(\left[V^{(r)}\right]\right)\right) + b_{et}$$
(17)

$$Eh, Et = Sigmoid\left(tanh\left(X^{(eh)}\right)\right), Sigmoid\left(tanh\left(X^{(et)}\right)\right)$$
(18)

where W_{eh} and W_{et} are trainable weights and b_{eh} and b_{et} are the union training bias. We obtain Eh and Et from the sigmoid function to describe the probability after employing the tanh function to limit the uncertain boundary value.

2.6. Loss Function

Because the extraction is a joint task, we combine two available losses of union type and entity in the training stage, as follows:

$$L_{entity} = -\frac{1}{2j} \sum_{z=1}^{j} \left(y_z^{(eh)} \log(Eh_z) + y_z^{(et)} \log(Et_z) \right)$$
(19)

$$L_{relation} = -\frac{1}{j} \sum_{z=1}^{j} y_{z}^{(r)} \log(T_{z})$$
(20)

$$L_{total} = L_{entity} + L_{relation} \tag{21}$$

where *j* is the maximum number of entity pairs in a sample. Since we extract the type and entity sequentially and keep symmetry in the number of entity pairs and union types, we keep the ratio of the two losses of 1:1.

3. Results

3.1. Dataset

For the goal of a fair and meaningful comparison, we follow the previous works [22] to evaluate SAFW on two widely used scientific datasets, namely SciERC [23] and Semeval 2017 task 10 [24]. SciERC contains text from 500 scientific abstracts. Semeval 2017 task 10 is a corpus for the information retrieval task, which was built from Science Direct open access publications. In order to validate the effectiveness of our proposed method on different types of relation extraction datasets, these two datasets are used. Both of them contain entity and relation type. The experiments for each dataset are divided into two types, namely "Rel" and "Rel+". Our focus is on the Rel+ section of the results, the relational quintuple result. It is a strict constraint on the entity types. The Rel section is the traditional triple extraction of the relation, the results of which are for reference only. We did not select certain typical datasets (such as ADE) because their entities or relations are not labeled with categories, making it meaningless to count the results of Rel+.

In addition, we make use of two other datasets, NYT [25] and WebNLG [26], which are both divided into two different versions. NYT* and WebNLG* annotate the last word of the entity span. On the contrary, NYT and WebNLG annotate the whole span. These two datasets have shown good performance in previous work and have not been separately labeled with entity categories. Therefore, our purpose in using them is to test whether SAFW can still make breakthroughs under such conditions. We compare the total testing data with other baselines in the supplementary experimental section.

3.2. Evaluation

We compare SAFW with other joint extraction methods. The experimental results of the baseline models come from the original paper records. According to the basic evaluation system, we use micro precision, recall, and F1 value to evaluate the result in order to be consistent with the previous works. As the feature words are not annotated, we do not compare them with the other baselines to produce the statistical characteristics. The visualization of the feature words will be discussed in Section 3.8. We provide math statistics for them, as they are indirectly supervised by the results.

3.3. Implementation Details

We use BERT-base-cased or SciBERT-base as the encoder, which is the pre-trained model from a hugging-face format file. The batch size of all of the datasets is 16 for BERT. We set the epoch to 150 for all of the training processes with the AdamW optimizer for tuning the parameters. The learning rate is 1×10^{-5} . We use Nvidia RTX 3090 (NVIDIA, Santa Clara, CA, USA) to train the model and choose the model parameters with the best performance on the validation set to output the test results.

3.4. The Results of the Experiment

Table 1 shows the results of SAFW on the scientific datasets. There are two types of experiments. Compared with the other baseline methods in the Rel+ mode, SAFW outperforms them. The result of Rel+ improves by 8.2% in SciERC and by 7.2% in Semeval 2017 task 10. Table 2 reports the results of SAFW on two basic datasets compared with the other baseline methods. SAFW improves by 0.8% F1 on NYT, 0.8% F1 on NYT*, and 1.4% F1 on WebNLG*. Note that the F1 values of NYT and WebNLG are over 90%, so our improvements show an effective promotion. All of these have proved the effectiveness of SAFW in performance improvement.

Table 1. The results of the two scientific datasets.

Dataset	Extraction Mode	Model Name	Р	R	F
		DyGIE [27]	-	-	41.6
		SPEKI.PL [28]	51.94	50.62	51.25
		Specific provide the second s			52.6 E0.1
		PL-Marker [31]			52
		PL-Marker (Entity Neighbor)			53.2
		PL-Marker + OVO			56.1
		SpERT (SciBERT) [32]	53.4	48.54	50.84
	Rel	Trimf (SciBERT) [33]	52.63	52.32	52.44
		PURE (SciBERT) [28]	-	-	50.1
		PURE (SciBERT) + OVO			51.6
		PERA (SciBERT) [22]	-	-	55.3
		MCER [34]	-	-	53.3
		HGERE [35]	-	-	55.7
		ATG [36]	-	-	51.1
0.15D.0		SAFW (SciBERT)	53.1	46.3	49.5
SciERC		SpERT.PL	39.94	38.98	39.41
		SpERT.PL + OVO			41.5
		cross-sentence ALB			36.7
		PL-Marker			40.6
		cross-sentence ALB PL-Marker PL-Marker (Entity Neighbor) PL-Marker + OVO SpERT (SciBERT) 36.25 40.04	41.6		
		PL-Marker + OVO		40.04 39.01	44.5
		SpERT (SciBERT)	36.25	40.04	38.05
	Rel+	Trimf (SciBERT)	42.27	39.01	40.58
	Ref	PURE (SciBERT)	-		36.8
		PURE (SciBERT) + OVO			40.1
		PERA (SCIBERT)	-	-	35.7
		MCER	-	- 25.46	42.8
		MIL [57] HCERE	44.55	55.40	39.00 43.6
		ATC		_	38.6
		SAFW (SciBERT)	52.6	50.5	51.5
			02.0		0110
		DyGIE (BERT)	46.8	30.5	36.9
	Rel	SpERT (BERT)	56.52	31.1	40.12
		Irimf (BERI)	52.86	35.41	42.41
Somoval 2017		SAFW (BER1)	42.6	39.7	41.1
task 10		SCIIE (BERT) [23]	40.4	21.2	27.8
105K 10		DyGIE (BERT)	35.11	24.23	28.7
	Rel+	SpERT (BERT)	45.22	24.88	32.1
		Trimf (BERT)	45.22	24.88	32.1
		SAFW (BERT)	40.4	38.3	39.3

M. 1.1		NYT*		I I	WebNLG*			NYT		WebNLG		
Model	Р	R	F	Р	R	F	Р	R	F	Р	R	F
EmRel (BERT) [38]	91.7	92.5	92.1	92.7	93	92.9	92.6	92.7	92.6	90.2	87.4	88.7
GRTE (BERT) [10]	92.9	93.1	93	93.7	94.2	93.9	93.4	93.5	93.4	92.3	87.9	90
ConCasRTE (BERT) [39]	92.9	92.3	92.6	93.8	92.5	93.1	92.9	92.1	92.5	90.6	88.1	89.3
BiRTE (BERT) [11]	92.2	93.8	93	93.2	94	93.6	91.9	93.7	92.8	89	89.5	89.3
CasRel (BERT) [40]	89.7	89.5	89.6	93.4	90.1	91.8	-	-	-	-	-	-
TP-Linker (BERT) [41]	91.3	92.5	91.9	91.8	92	91.9	91.4	92.6	92	88.9	84.5	86.7
PRGC (BERT) [42]	93.3	91.9	92.6	94	92.1	93	93.5	91.9	92.7	89.9	87.2	88.5
RFBFN (BERT) [9]	93.4	93.2	93.3	93.9	94.1	94	93.7	93.6	93.6	91.5	89.4	90.4
RPSS (BERT) [43]	93.5	93.2	93.3	94.7	95.1	94.9	-	-	-	-	-	-
PRE-Span (BERT) [44]	90	85.3	88	95.5	92.9	94.2	88.6	84.7	86.6	83.4	82.7	83
BTDM (BERT) [45]	93	92.5	92.7	94.1	93.5	93.8	93.1	92.5	92.7	90.9	90.1	90.5
SPN (BERT) [46]	93.3	91.7	92.5	93.1	93.6	93.4	-	-	-	-	-	-
NEDORT (BERT) [47]	91.8	89.7	90.7	92.2	91.5	91.9	-	-	-	-	-	-
DERP (BERT) [48]	92.1	90	91	92.8	92.9	92.9	-	-	-	-	-	-
ODRTEBERT [49]	93.5	93.9	93.7	94.6	95.1	94.9	94.2	93.6	93.9	92.8	92.1	92.5
SAFW (BERT) (Ours)	94.4	94.6	94.5	96.4	96.3	96.3	94.4	95	94.7	91.9	92.1	92.0

Table 2. Comparison of SAFW with other baselines (all Rel-type results).

We analyze the reasons for the improvement of "Rel+" and summarize two key points. First, the multiple rounds of resampling mines more accurate semantic information in the sequence encoded by the pre-trained encoding model and weakens the influence of irrelevant words on relation prediction. As relation prediction has priority over entity extraction, it is susceptible to the space of the result. The contribution ratio of several feature words is boosted by the re-sampling module, leading the model to find the correct relation types. Second, since the head and tail entity categories are encoded in common categories, the resulting template is set up to provide guidelines for both entity and relation categories. In the subsequent fusion process, the semantic information in the template is greatly enriched, due to the enhancing effect of the feature words, thus providing sufficient information to the relation quintuple. This kind of information is more explicit for type selection and entity extraction. Our SAFW combines the two improvements, in line with the end-to-end principle, so that we can achieve a better performance. In contrast, the other baselines do not consider reducing the result space by vector sampling in the encoding stage and reuse cross attention in mixed coding of hidden states. Therefore, their performances suffer.

By the analysis of "Rel", their performance may not increase uniformly. In the first three datasets, there was a slight decrease in the results of this section; however, in the next two datasets, only NYT, NYT*, and WebNLG* showed a slight improvement. This suggests that the performance of the current scheme is unpredictable when using only relation information as a template. The uniqueness of the NYT data lies in the fact that the relation category results include some information that may indicate entity categories (such as "/location/location/contents" for location entities). WebNLG* only considers the last word of the entity, and missing entity categories have a relatively small impact on such entity determination. As a comparison, missing entity categories will affect the performance of entity region determination in the other datasets. However, predicting the entity categories at the same time is more in line with the practicalities of knowledge graph construction. Thus, we can still demonstrate the feasibility of SAFW.

3.5. Ablation Study on the Components of the Framework

We remove the multi-turn vector resampling module to detect the contribution of this part. As shown in Table 3, all of the evaluation indexes, including precision, recall, and the F1 value, significantly decrease without this component, taking some irrelevant words into consideration. Since the change in the probability distribution occurs before the extraction stage, it acts on the union type and entity. In addition, the component transforms the

output of the pre-trained model to a value more suitable for prediction, especially at a low learning rate. The results of the experiment prove that the multi-turn vector resampling module is effective and outperforms the direct use of the encoder output.

Table 3. Ablation study on the SciERC dataset by removing the components of SAFW.

Model Name	Р	R	F
SAFW	52.6	50.5	51.5
without multi-turn vector resampling module	50.2	48.3	49.3
delete position max of α only	52.5	47.3	49.8
without self-attention of encoded relation sequence $V^{(r)}$	52.3	50.3	51.3
without self-attention of encoded entity sequence $V^{(e)}$	52.2	50.4	51.3

Another improvement comes from the double cross attention layers, so we use single cross attention instead as a comparison. The F1 values drop with the recalls, because the design can mine more information from the encoded sequence. The input query may not be able to retrieve enough information from an interaction with it, because the information density of the former is lower than that of the latter. The candidates are from the description of the union types in the training datasets, and the vector sequences represent all of the sentences in the samples. Therefore, one time calculation may lose some important information for union-type finding or entity extraction in the fixed templates.

3.6. Ablation Study on Union-Type Prediction Only

In this section, we only examine the results of the extraction of the union types, so the results presented do not include the locations of the entities, but only the union types of the relation and the entity. We performed these experiments on each dataset, and the results are shown in Table 4. The results in the table show a significant improvement in the F value over the quintuple results when only the union types are output, indicating that the performance improvement of SAFW is mainly due to the use of self-adaptive feature words to strengthen the focus on certain parts of the sentence. This method forces the feature vector with the maximum weight at each moment to have another $1 \times$ weight. Instead, after the self-attention calculation, the other non-maximum weight vectors are multiplied by a coefficient of between 0 and 1, close to 0, thus maximizing the distance between the self-adaptive feature words and the irrelevant words. Such a calculation increases the F value for multi-category classification. This is reflected in the union types.

Table 4. Ablation study on the prediction of only the union types.

	SciERC	Semeval 2017 Task 10	NYT	NYT*	WebNLG	WebNLG*
Only Union Type	81.0	60.6	97.2	96.7	94.4	97.5
Full Result	51.5	39.3	94.7	94.5	91.8	96.3

3.7. Additional Experiments by LLMs

We also employ LLMs such as ChatGLM2-6B, Llama-7B, and Baichuan2-7B and modify the task into a generative one. We employ the results of the LLMs with Lora fine-tuning [50] as a comparison. As shown in Table 5, they obtain a similar or a slightly worse performance, indicating that the performance may not be improved by these models. Although the LLMs have emergent and reasoning abilities, they are unable to accept the restrictions of the specific knowledge system in the datasets well. Furthermore, they tend to produce results that should not occur in the current situation, because they have an excessive amount of prior knowledge. Therefore, large language models do not necessarily work better for certain tasks.

Extraction Mode	Dataset	ChatGLM2-6B	Llama-7B	Baichuan2-7B	SAFW
	SciERC	48.8	48.7	49.2	49.5
	2017t10	40.4	40.5	41.1	41.1
Rel	NYT	94.6	94.5	94.7	94.7
	NYT*	94.4	93.7	94.6	94.5
	WebNLG	91.7	91.7	91.6	92.0
	WebNLG*	95.4	95.5	96.1	96.3
D 1	SciERC	42.4	43.1	43.3	51.5
Kel+	2017t10	36.2	35.4	35.9	39.3

Table 5. The results from LLMs.

3.8. Case Study

We show two correct cases and one incorrect case in Figure 4. These cases are from the WebNLG* dataset. We show three cases, including three parts in vertical order, as follows: source sentence, BERT tokens with probabilities, and relation quintuples. The words presented in bold, red font can be viewed as the feature words for the sentences in these samples. We set five keywords for each case and make it clear in the figure.

For the correct cases, we find that the feature words are given greater weights by the multi-turn vector resampling module. The selected words "created" and "publisher" are mapped to the union types "Character/Creator/Person" and "Publication/Publisher/ Organization". The types of "state" and "city" are predicted by the influence of the phrase "New York", which is the interaction of "publisher" and "New York". The feature words in the second case are not capable of inferring the union type "affiliation", as this is decided by the two entities "university" and "universities".

For the incorrect case, although the correct prediction of the first two quintuples relies on the words "Museum" and "ethnic", they do not cover the hints of all of the quintuples. The entity "Prefecture" is not considered in the prediction, due to no prompting for "Location/isPartOf/Area" and the inherent meaning between "Prefecture" and "Japan". In addition, "Prefecture" is next to "Akita", which leads to ignoring the existence of "Akita" in the text. On the contrary, the missed quintuple contains the correct type, but no target entity is found, which illustrates that feature words cannot redirect entities when they do not clearly reflect the relationship. The above failures illustrate that SAFW can lead to insufficient or incorrect information about the selected feature words if sufficient feature words are not obtained or if the feature words are implicit semantic representations, which, in turn, misleads the final extraction results of the relational quintuple.

3.9. POS Distribution

Figure 5 shows the POS distribution of the words in the feature words. By the analysis of the part-of-speech, we find that the highest proportion of the selected words are nouns. This fact proves that SAFW can provide key indicators for relation quintuple extraction and give a meaningful explanation result. Compared with the others, this is one of the differences of SAFW. We can collect them and generate a report for each different dataset.

Baymax, is a character in Big Hero 6 that stars Damon Wayans Jr. The character							
was <mark>created</mark> by the American, Duncan Rouleau.							
[CLS](0.066), Bay(0.008), ##max(0.004), is(0.261), a(0.264), character(1.114),							
in(0.241), Big(1.246) , Hero(0.208), 6(0.231), that(0.239), stars(1.177) ,							
Damon(0.262), Way(0.041), ##ans(0.25), Jr(0.091), .(0.062), The(0.314),							
character(1.052), was(0.294), created(1.317), by(0.294), the(0.211),							
American(0.228), ,(0.222), Duncan(0.24), R(0), ##ou(0), ##lea(0), ##u(0), .(0.063)							
Right case(head-union type-tail):							
Baymax-Character/Creator/Person-Rouleau							
Cornell University in Ithaca, New York is the publisher of Administrative Science							
Quarterly. The University is affiliated with the Association of American							
Universities.							
[CLS](0.076), Cornell(1.4), University(0.252), in(0.225), lt(0.149), ##ha(0.142),							
##ca(0.192), ,(0.095), New(0.216), York(0.002), is(1.233), the(1.14),							
publisher(1.067), of(1.312), Administrative(0.265), Science (0.234),							
Quarterly(0.013), .(0.149), The(0.28), University(0.002), is(0.23),							
affiliated (0.121) , with (0.106) , the (0.18) , Association (0.289) , of (0.189) ,							
American(0.295), Universities(0), .(0.148)							
Right case (head-union type-tail):							
Quarterly-Publication/Publisher/Organization-University							
University-Organization/Affiliation/Organization-Universities							
University-Organization/State/Location-York							
University-Organization/City/Location-York							
othering groups of Japan is the Filinings							
$[C S](0.096)$ The (0.174) A (0.088) ##b $i(0.077)$ ##t $_2(0.11)$ Museum (1.328)							
f(0.236) Art(0.022) is(0.322) located(0.271) in(0.146) A(0.071) ##ki(0.081)							
(0.0230), A((0.022) , (0.022) , (0.022) , (0.069) , (0.001) , (0.01) , (0.01) , (0.029)							
(0.003), (0.003) , (0.003) , (0.003) , (0.003) , (0.003) , (0.001) , (0.001) , (0.023) , (0.023) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) , (0.025) ,							
$groups(0.382) \circ of(0.165) = Japan(0.001) is(0.354) the(1.339) Filipino(1.28)$							
##s(1,120) (0,094)							
Right case(head-union type-tail):							
Art-Object/Location/Location-Akita							
Japan-Location/ethnicGroup/Location-Japan							
Wrong case(head-union type-tail):							
Akita-Location/isPartOf/Area-Prefecture							
Prefecture-Area/Country/Country-Japan							
Missing case(head-union type-tail):							
Prefecture-Area/isPartOf/Area-Prefecture							
Art-Object/Country/Country-Japan							
Art-Object/Location/Area-Prefecture							

Figure 4. The correct and incorrect cases. The first two are correct and the last one is incorrect. The red characters are the highest probability in every turn, which is added by 1 and masked by the growing mask in order to block next step of the calculation. The words starting with "##" represent the subword obtained by further splitting the word by the WordpieceTokenizer in the BERT model.



Figure 5. The POS distribution of the words in feature words. The data come from all four groups of the datasets. A comparison is shown with bar chart.

4. Conclusions

In this paper, we present a framework for assisting the extraction of relational quintuples through the use of unlabeled adaptive feature words. By summarizing our framework, we find that the use of relational quintuples to simultaneously supervise feature words, relations, and entities can extract interpretable words for each quintuple while improving the performance. Extensive experiments show that our framework is effective and can be used when BERTs act as encoders. This strategy provides additional information for knowledge graph construction by retrieving words that are considered useless and are discarded after information extraction. Thus, it enhances the relationship-to-relationship links in the construction of a knowledge graph of a single corpus. It also shows that the current work can be applied to hidden knowledge discovery scenarios, as knowledge graphs constructed in this way are not bound to entities and relationships. In the future, we have plans for the application of this approach to more vertical domains for novel knowledge discovery.

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

Funding: This research was funded by the National Social Science Foundation of China, grant number 21BTQ06.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available at [SciERC] https://nlp.cs.washington.edu/sciIE/, accessed on 29 August 2018, [Semeval 2017 task 10] https://scienceie.github.io/, accessed on 17 April 2017, [NYT&NYT*] https://catalog.ldc.upenn.edu/LDC2008T19, accessed on 1 December 2010, and [WebNLG&WebNLG*] https://synalp.gitlabpages. inria.fr/webnlg-challenge/, accessed on 1 October 2017. The codes are available at https://github.com/LewisYJohnson/SAF_W (accessed on 30 April 2024).

Acknowledgments: Upon the completion of this thesis, I would like to take this opportunity to express my sincere gratitude to my two supervisors, Fu Lijun and Xia Xiaojun, who have given me a research environment, equipment support, and important guidance on this thesis. We are also indebted to Tong Li of Guangming Net, who helped us to improve the presentation of the figures and tables and managed the related project. We would also like to thank Elaine Wang for assisting us in resolving manuscript-related issues.

Conflicts of Interest: Author Yujiang Liu, Lijun Fu, Xiaojun Xia were employed by the company Shenyang Institute of Computing Technology Co., Ltd. The remaining authors declare that the re-search was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Cheng, Q.; Liu, J.; Qu, X.; Zhao, J.; Liang, J.; Wang, Z. HacRED: A Large-Scale Relation Extraction Dataset Toward Hard Cases in Practical Applications. In Proceedings of the 10th ACL-IJCNLP, Bangkok, Thailand, 1–6 August 2021; pp. 2819–2831.
- Sharma, S.; Nayak, T.; Bose, A.; Meena, A.K.; Dasgupta, K.; Ganguly, N. FinRED: A Dataset for Relation Extraction in Financial Domain. In Proceedings of the Companion Proceedings of the Web Conference, Lyon, France, 25–29 April 2022; pp. 595–597.
- Al-Sabri, R.; Gao, J.; Chen, J.; Oloulade, B.M. Multi-View Graph Neural Architecture Search for Biomedical Entity and Relation Extraction. In *IEEE/ACM Transactions on Computational Biology and Bioinformatics*; IEEE: Atlanta, GA, USA, 2022; Volume 1, pp. 1–13.
- 4. Zhou, B.; Gao, D.; Yan, L.; Cao, J.; Zhang, S. Research on key technologies for fault knowledge acquisition of power communication equipment. In *Procedia Computer Science*; Zeng, X., Ed.; Elsevier: Manchester, UK, 2021; Volume 183, pp. 479–485.
- Li, Z.; Fu, L.; Wang, X.; Zhang, H. RFBFN: A Relation-First Blank Filling Network for Joint Relational Triple Extraction. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, Dublin, Ireland, 22–27 May 2022; pp. 10–20.
- 6. Ahmed, K.; Khurshid, S.K.; Hina, S. CyberEntRel: Joint extraction of cyber entities and relations using deep learning. *Comput. Secur.* 2024, 136, 103579. [CrossRef]
- Yan, F.; Shen, B.; Dai, C. Causality Extraction Cascade Model Based on Dual Labeling. J. Adv. Comput. Intell. Intell. Inform. 2023, 27, 421–430. [CrossRef]
- Tse, T.H.E.; Kim, K.I.; Leonardis, A.; Chang, H.J. Collaborative learning for hand and object reconstruction with attention-guided graph convolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 1664–1674.
- 9. Zhu, G.; Huang, X.; Yang, R.; Sun, R. Relationship Extraction Method for Urban Rail Transit Operation Emergencies Records. *IEEE Trans. Intell. Veh.* **2022**, *8*, 520–530. [CrossRef]
- Ren, F.; Zhang, L.; Yin, S.; Zhao, X.; Liu, S.; Li, B. A novel global feature-oriented relational triple extraction model based on table filling. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Punta Cana, Dominican Republic, 7–11 November 2021; pp. 2646–2656.
- Ren, F.; Zhang, L.; Zhao, X.; Yin, S.; Liu, S. A Simple but Effective Bidirectional Framework for Relational Triple Extraction. In Proceedings of the 15th ACM International Conference on Web Search and Data Mining, Phoenix, AZ, USA, 21–25 February 2022; pp. 824–832.
- 12. Vinyals, O.; Fortunato, M. Pointer Networks. Adv. Neural Inf. Process. Syst. 2015, 28, 2692–2700.
- Lee, J.; Lee, M.J.; Yang, J.Y. Does it Really Generalize Well on Unseen Data? Systematic Evaluation of Relational Triple Extraction Methods. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Seattle, WA, USA, 10–15 July 2022; pp. 3849–3858.
- Liu, Y.; Zhang, L.; Yin, S.; Zhao, X. An Effective System for Multi-format Information Extraction. In Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing, Qingdao, China, 13–17 October 2021; pp. 460–471.
- Xie, Y.; Shen, J.; Li, S.; Mao, Y. EIDER: Empowering Document-level Relation Extraction with Efficient Evidence Extraction and Inference-stage Fusion. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, Dublin, Ireland, 22–27 May 2022; pp. 10–20.
- Li, L.; Chen, X.; Bi, Z.; Xie, X.; Deng, S.; Zhang, N. Normal vs. adversarial: Salience-based analysis of adversarial samples for relation extraction. In Proceedings of the 10th International Joint Conference on Knowledge Graphs, Bangkok, Thailand, 6–8 December 2021; pp. 115–120.
- Zhou, W.; Chen, M. An improved baseline for sentence-level relation extraction. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, Dublin, Ireland, 22–27 May 2022; pp. 161–168.
- Chen, X.; Zhang, N.; Xie, X.; Deng, S.; Yao, Y. Knowprompt: Knowledge-aware prompt-tuning with synergistic optimization for relation extraction. In Proceedings of the ACM Web Conference, Lyon, France, 25–29 April 2022; pp. 2778–2788.
- Ding, N.; Wang, X.; Fu, Y.; Xu, G.; Wang, R.; Xie, P. Prototypical representation learning for relation extraction. In Proceedings of the International Conference on Learning Representations, Vienna, Austria, 3–7 May 2021; pp. 7970–7986.

- Devlin, J.; Chang, M.; Lee, 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, Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186.
- Liao, Y.; Jiang, X.; Liu, Q. Probabilistically Masked Language Model Capable of Autoregressive Generation in Arbitrary Word Order. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 263–274.
- Zhang, C.; Gao, S.; Wang, H.; Zhang, W. Position-aware Joint Entity and Relation Extraction with Attention Mechanism. In Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23–29 July 2022; pp. 4496–4502.
- Luan, Y.; He, L.; Ostendorf, M.; Hajishirzi, H. Multi-Task Identification of Entities, Relations, and Coreference for Scientific Knowledge Graph Construction. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 3219–3232.
- Augenstein, I.; Das, M.; Riedel, S.; Vikraman, L.; McCallum, A. SemEval 2017 Task 10: Scienceie-Extracting Keyphrases and Relations from Scientific Publications. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), Vancouver, BC, Canada, 3–4 August 2017; pp. 546–555.
- Riedel, S.; Yao, L. Modeling relations and their mentions without labeled text. In Proceedings of the Joint European Coreference on Machine Learning and Knowledge Discovery in Databases, Barcelona, Spain, 20–24 September 2010; pp. 148–163.
- Gardent, C.; Shimorina, A.; Narayan, S. Creating training corpora for NLG micro-planning. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, BC, Canada, 30 July–4 August 2017; pp. 179–188.
- 27. Luan, Y.; Wadden, D.; He, L.; Shah, A.; Ostendorf, M.; Hajishirzi, H. A General Framework for Information Extraction using Dynamic Span Graphs. In Proceedings of the NAACL-HLT, Minneapolis, MN, USA, 2–7 June 2019; pp. 3036–3046.
- Zhong, Z.; Chen, D. A frustratingly easy approach for entity and relation extraction. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Online, 6–11 June 2021; pp. 50–61.
- 29. Ren, L.; Liu, Y.; Cao, Y.; Ouyang, C. CoVariance-based Causal Debiasing for Entity and Relation Extraction. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2023, Singapore, 6–10 December 2023; pp. 2627–2640.
- Santosh, T.; Chakraborty, P.; Dutta, S.; Sanyal, D.K.; Das, P.P. Joint entity and relation extraction from scientific documents: Role of linguistic information and entity types. In Proceedings of the 2nd Workshop on Extraction and Evaluation of Knowledge Entities from Scientific Documents (EEKE 2021) Co-Located with JCDL 2021, Virtual Event, 30 September 2021; pp. 15–19.
- Ye, D.; Lin, Y.; Li, P.; Sun, M. Packed Levitated Marker for Entity and Relation Extraction. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Dublin, Ireland, 22–27 May 2022.
- Eberts, M.; Ulges, A. Span-based joint entity and relation extraction with transformer pre-training. In Proceedings of the ECAI 2020, Santiago de Compostela, Spain, 29 August–8 September 2020; IOS Press: Amsterdam, The Netherlands, 2020; pp. 2006–2013.
- Shen, Y.; Ma, X.; Tang, Y.; Lu, W. A triggersense memory flow framework for joint entity and relation extraction. In Proceedings of the Web Conference 2021, Ljubljana, Slovenia, 19–23 April 2021; pp. 1704–1715.
- 34. Wu, Y.; Chen, Y.; Qin, Y.; Huang, R.; Tang, R. A marker collaborating model for entity and relation extraction. *J. King Saud Univ.-Comput. Inf. Sci.* 2022, 34, 9163–9172. [CrossRef]
- 35. Yan, Z.; Yang, S.; Liu, W.; Tu, K. Joint Entity and Relation Extraction with Span Pruning and Hypergraph Neural Networks. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Singapore, 6–10 December 2023.
- Zaratiana, U.; Tomeh, N.; Holat, P.; Charnois, T. An Autoregressive Text-to-Graph Framework for Joint Entity and Relation Extraction. In Proceedings of the ICML 2023 Workshop on Structured Probabilistic Inference & Generative Modeling, Honolulu, HI, USA, 5 July 2023.
- Zaratiana, U.; Tomeh, N.; Holat, P.; Charnois, T. Solving Label Variation in Scientific Information Extraction via Multi-Task Learning. In Proceedings of the 37th Pacific Asia Conference on Language, Information and Computation, Hong Kong, China, 2–4 December 2023; pp. 243–256.
- Xu, B.; Wang, Q.; Lyu, Y.; Shi, Y.; Zhu, Y.; Gao, J.; Mao, Z. EmRel: Joint Representation of Entities and Embedded Relations for Multi-triple Extraction. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Seattle, WA, USA, 10–15 July 2022; pp. 659–665.
- Ren, F.; Zhang, L.; Yin, S.; Zhao, X.; Liu, S.; Li, B. A Conditional Cascade Model for Relational Triple Extraction. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management, Gold Coast, QLD, Australia, 1–5 November 2021; pp. 5585–5590.
- 40. Wei, Z.; Su, J.; Wang, Y.; Wang, Y. A novel cascade binary tagging framework for relational triple extraction. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Seattle, WA, USA, 5–10 July 2020; pp. 1476–1488.
- Wang, Y.; Yu, B.; Zhang, Y.; Wang, Y.; Liu, T.; Zhu, H. TPLinker: Single-stage joint extraction of entities and relations through token pair linking. In Proceedings of the 28th International Conference on Computational Linguistics, Barcelona, Spain, 8–13 December 2020; pp. 1572–1582.
- Zheng, H.; Wen, R.; Chen, X.; Yang, Y.; Zhang, Y.; Zhang, Z. PRGC: Potential Relation and Global Correspondence Based Joint Relational Triple Extraction. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Bangkok, Thailand, 1–6 August 2021; pp. 6225–6235.

- 43. Zhang, Z.; Liu, H.; Yang, J.; Li, X. Relational prompt-based single-module single-step model for relational triple extraction. *J. King Saud Univ.-Comput. Inf. Sci.* 2023, 35, 101748. [CrossRef]
- 44. Liu, S.; Lyu, W.Q.; Ma, X.; Ge, J.K. An Entity-Relation Joint Extraction Method Based on Two Independent Sub-Modules from Unstructured Text. *IEEE Access* 2023, *11*, 122154–122163. [CrossRef]
- 45. Zhang, Z.; Yang, J.; Liu, H.; Hu, P. BTDM: A Bi-Directional Translating Decoding Model-Based Relational Triple Extraction. *Appl. Sci.* **2023**, *13*, 4447. [CrossRef]
- 46. 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. [CrossRef]
- 47. Zhang, Z.; Hu, X.; Zhang, H.; Liu, J. NEDORT: A novel and efficient approach to the data overlap problem in relational triples. *Complex Intell. Syst.* **2023**, *9*, 5235–5250. [CrossRef]
- 48. Xiao, Y.; Chen, G.; Du, C.; Li, L.; Yuan, Y.; Zou, J.; Liu, J. A Study on Double-Headed Entities and Relations Prediction Framework for Joint Triple Extraction. *Mathematics* **2023**, *11*, 4583. [CrossRef]
- Ning, J.; Yang, Z.; Sun, Y.; Wang, Z.; Lin, H. Od-rte: A onestage object detection framework for relational triple extraction. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Toronto, ON, Canada, 9–14 July 2023; pp. 11120–11135.
- 50. Hu, E.J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W. LoRA: Low-Rank Adaptation of Large Language Models. In Proceedings of the International Conference on Learning Representations, Vienna, Austria, 4 May 2021.

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.