

Article

Joint Entity and Relation Extraction Network with Enhanced Explicit and Implicit Semantic Information

Huiyan Wu ^{1,2}  and Jun Huang ^{1,2,*} 

¹ Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201203, China; wuhuiyan2019@sari.ac.cn

² University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: huangj@sari.ac.cn

Abstract: The main purpose of the joint entity and relation extraction is to extract entities from unstructured texts and extract the relation between labeled entities at the same time. At present, most existing joint entity and relation extraction networks ignore the utilization of explicit semantic information and explore implicit semantic information insufficiently. In this paper, we propose Joint Entity and Relation Extraction Network with Enhanced Explicit and Implicit Semantic Information (EINET). First, on the premise of using the pre-trained model, we introduce explicit semantics from Semantic Role Labeling (SRL), which contains rich semantic features about the entity types and relation of entities. Then, to enhance the implicit semantic information and extract richer features of the entity and local context, we adopt different Bi-directional Long Short-Term Memory (Bi-LSTM) networks to encode entities and local contexts, respectively. In addition, we propose to integrate global semantic information and local context length representation in relation extraction to further improve the model performance. Our model achieves competitive results on three publicly available datasets. Compared with the baseline model on Conll04, EINET obtains improvements by 2.37% in F1 for named entity recognition and 3.43% in F1 for relation extraction.

Keywords: named entity recognition; relation extraction; semantic role labeling; Bi-directional long short-term memory network; pre-trained model



Citation: Wu, H.; Huang, J.

Joint Entity and Relation Extraction Network with Enhanced Explicit and Implicit Semantic Information. *Appl. Sci.* **2022**, *12*, 6231. <https://doi.org/10.3390/app12126231>

Academic Editors: Andrea Prati, Luis Javier García Villalba and Vincent A. Cicirello

Received: 9 May 2022

Accepted: 13 June 2022

Published: 19 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

The main purpose of the joint entity and relation extraction is to extract entities from unstructured texts and extract the relationship between labeled entities. It completes Named Entity Recognition (NER) and Relation Extraction (RE) based on joint learning methods. For example, for the sentence “Leonardo DiCaprio starred in Christopher Nolan’s thriller Inception”. The goal of the joint entity and relation extraction task is to extract triples (Leonardo DiCaprio, Plays-In, Inception) and (Inception, Director, Christopher Nolan), and so on.

Recently, pre-trained models such as BERT [1], Transformer-XL [2], and RoBERT [3] have received great attention in the field of Natural Language Processing (NLP). These models are typically pretrained on large document data, and they are transferred to target tasks with relatively little supervised training data. In many NLP tasks, work based on a pre-trained model achieves the best performance, such as question answering [4], contextual emotion detection [5] and joint entity and relationship extraction. Despite the success of these pre-trained language models, existing joint entity and relation extraction networks only focus on the text representation provided by pre-trained models, while ignoring the introduction of explicit semantic information and enhancements of implicit semantic information.

Semantic Role Labeling (SRL) can build dependencies between the predicates and arguments of a sentence, and this semantic structure information can provide rich semantics

for text representation. At present, there are a number of approaches that incorporate auxiliary explicit semantic information into some NLP tasks [6,7], but there is currently a lack of work on using SRL information for joint entity and relation extraction tasks. If a word or phrase is labeled with one semantic role, it is more likely to be labeled as an entity. For example, the semantic role label *ArgM-LOC* contains location information, which can provide auxiliary information for entity extraction of type *Location*. At the same time, explicit semantic information can cover the semantic relationship between words, which is very useful for relation extraction. Semantic role labeling can provide explicit auxiliary information for NER and RE, and then help the model improve performance. Therefore, we fuse the explicit semantic information obtained by SRL with BERT for joint entity and relation extraction tasks.

In addition to ignoring the introduction of explicit semantic information, many existing models also do not adequately explore implicit semantic information. In most of the existing models, the representation vector of text is shared in both entity recognition and relation extraction. However, named entity recognition focuses on the semantic extraction of entities, while relation extraction focuses on the semantic information of the local context which is drawn from the end of the first entity to the beginning of the second entity. Therefore, in order to fully explore the implicit semantic information, we not only adopt Bi-LSTM to further extract implicit semantic information, but also design a novel separately encoded method which adopts two different Bi-LSTMs to enhance the implicit semantic information of the entity and the local context, respectively.

It is worth noting that, considering the importance of the context information where the entity pair is located for relation extraction, we introduce the global contextual information based on Bi-LSTM, and enhance semantic information of the local context by adding local context length representation. Enriching the semantic information of the context from global and local perspectives further improves the performance of the model on relation extraction. In general, our contributions can be summarized as follows:

- We propose a Joint Entity and Relation Extraction Network with Enhanced Explicit and Implicit Semantic Information (EINET). On the premise of using the pre-trained model, we introduce explicit semantic information and fully explore the implicit semantic information for joint entity and relation extraction.
- As far as we know, we are the first one to use semantic role labeling information for joint entity and relation extraction. Semantic role labeling can not only provide explicit semantic information for NER and RE, but also helps the model to enhance semantic understanding of text.
- While adopting the BERT pre-trained model, we further explore the implicit semantic information of entities and local contexts based on different Bi-LSTMs. By our separate encoding method, the different features of entities and local contexts are fully explored, so as to purposefully improve the performance of named entity recognition and relation extraction.
- We propose to integrate global semantic information and local context length representation in relation extraction to further improve the model performance.
- Our model shows strong competitiveness on three publicly available joint entity and relation extraction datasets (Conll04, SCIERC, ADE), achieving competitive experimental results.

2. Related Work

The pipeline-based method is a typical method to realize NER and RE. However, this type of method ignores the relationship between NER and RE, and it produces errors caused by the cascade relationship [8]. Therefore, methods based on joint learning have emerged. The method based on joint learning can be divided into two categories, the sequence tagging-based method and the span-based method.

2.1. Sequence Tagging Based Method

Zheng et al. [9] first propose the truly joint entity and relation extraction model which uses a sequence tagging scheme for joint extraction. Takanobu et al. [10] propose a hierarchical reinforcement learning-based (RL) deep neural model for joint extraction. High-level reinforcement learning is used to determine relationships based on relation-specific tokens. After determining the relationship, the low-level reinforcement learning extracts the connection between two entities and the relation by using a sequential annotation method. To address the problem that the CopyR [11] model can only extract the last token of an entity, the CopyMTL [12] model improves CopyR and is able to extract the entity name completely using the sequence annotation method. Bowen et al. [13] decompose the joint extraction into two sub-tasks, the first task is head entity extraction and the second task is tail entity and relation extraction. They also use the sequence tagging approach to accomplish these two sub-tasks. Wei et al. [14] added BERT to the model proposed by Bowen et al. [13] to improve the performance of the model. Yuan et al. [15] add relation-specific attention mechanism to the joint extraction model based on sequence tagging and achieve a good improvement. Multi-head+AT [16] treats the relation extraction task as a multi-head selection problem. Each entity is combined with all other entities to form entity pairs, and then the relationships of these entity pairs are predicted. Each relationship is independent to achieve multiple relationship predictions for entity pair. SciIE [17] reduces the error propagation problem between named entity recognition and relation extraction by introducing a multi-task mechanism and disambiguation task. In addition, joint entity and relation extraction based on question answering systems or reading comprehension have emerged. Multi-turn QA [18] treats joint entity and relation extraction as a multi-turn question answering task. MRC4ERE++ [19] proposes a multi-turn question answering-based diversity question answering mechanism and designs two selection strategies to integrate different answers. Zhao et al. [20] propose a unified multi-task learning framework, which decomposes the task into three interactive subtasks, and they present a problem-based method to generate extracted objects.

Sequence tagging-based methods mark words with labels (BIO/BILOU). In this way, each word can only be assigned one label, which leads to the problem that overlapping entities can not be well extracted. In addition, methods based on question-answering depend on the quality of question generation and the performance of the question-answering model. Therefore, span-based joint entity and relation extraction methods have emerged.

2.2. Span-Based Method

The span-based method first identifies the boundary of entities, and then classifies entities according to the boundary identifier, so that overlapping entities can be identified due to different boundaries. Dixit et al. [21] propose a simple Bi-LSTM based model that generates a span representation for each possible entity and performs joint entity and relation extraction. Following this, DyGIE [22] shares span representations between multiple tasks via dynamically constructed span graphs. The MrMep model proposed by Chen et al. [23] uses a variant of the pointer network to generate the boundaries (start/end) of all head and tail entities in turn, and then uses a multi-headed attention mechanism [24] to extract the relationships corresponding to each entity.

With the advent of pre-trained models, some Transformer-based networks such as BERT, Transformer-XL, and RoBERTa have achieved outstanding results in a number of natural language processing tasks. Wadden et al. [25] propose DyGIE++ on the basis of DyGIE, and replace the Bi-LSTM with the BERT pre-trained model, which further improves the performance of DyGIE. Following this, Eberts et al. [26] propose a simple and effective lightweight inference model SpERT, which is a typical span-based joint entity and relation extraction model. Based on SpERT, Ji et al. [27] enhance the semantic representation of candidate entities and relations by attention mechanism, thereby improving the performance of joint entity and relation extract. TriMF is proposed by Shen et al. [28], which employs

a multi-level memory flow attention mechanism to enhance the bidirectional interaction between entity recognition and relation extraction.

However, most of the current models lack the use of explicit semantic information. Specifically, they usually share context representation in name entity recognition and relation extraction, and they do not explore implicit semantic information fully and purposefully. In response to the above problems, we not only introduce explicit semantic information, but also extract more fully implicit semantic information. We propose a simple and effective separate encoding method based on joint learning. This method enhances the implicit semantic information of entity and local context, respectively, while ensuring effective information transfer between NER and RE.

3. Materials and Methods

The overall architecture of our model EINET is shown in Figure 1, which includes three parts: Word Representation, Named Entity Recognition and Relation Extraction.

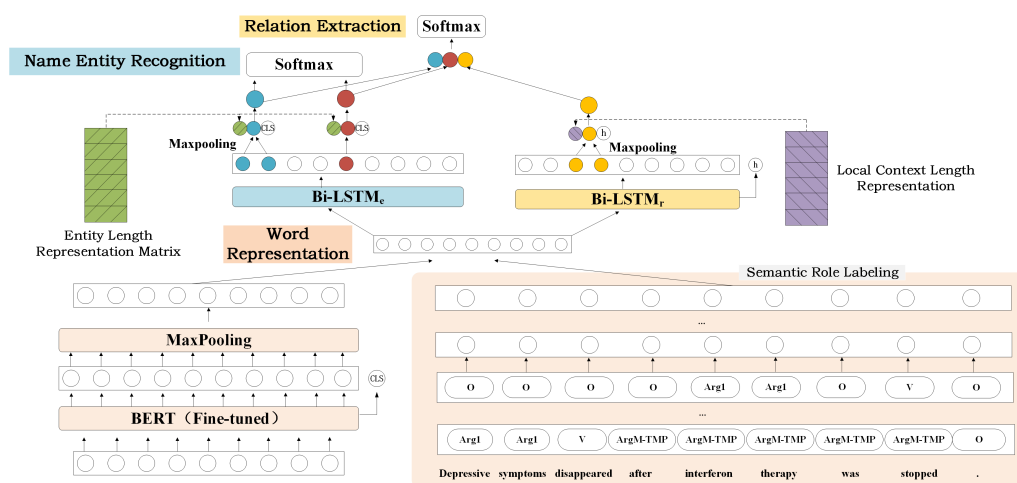


Figure 1. The overall architecture of our model EINET contains three components, Word Representation, Named Entity Recognition, and Relation Extraction. Blue vectors and red vectors are entities representations, and local context is represented by the orange vector.

Word Representation: The purpose of word representation is to convert each word in the sentence into a d -dimensional word embedding. The representation of word vectors consists of two parts, which are word embeddings from the pre-trained model and explicit semantic label embeddings based on SRL.

Named Entity Recognition: Named Entity Recognition is mainly responsible for obtaining candidate entity representations by span-based methods. Then, we can obtain the corresponding entity types by classifying the candidate entity representations.

Relationship Extraction: Relation extraction is mainly responsible for combining entities which are not assigned to the *none* class into entity pairs and then predicting the relation type of these entity pairs. The judgment of the relation type is not only based on the representation of the entity pair, but also is based on the global semantics of the sentence where the entity pair is located and the local semantics of the local context.

3.1. Word Representation

The joint entity and relation extraction task depends on the semantic information of entities, so in order to obtain a richer semantic representation, we not only adopt a pre-trained model to encode sentences, but also introduce explicit semantic information by utilizing semantic role labeling tools.

3.1.1. Pre-Trained Model

Transformer attracts a lot of attention at present. The Transformer-based model is pre-trained on large-scale text data, which has a strong ability to capture language features, and

it can provide relatively high-quality word vector representation for joint entity and relation extraction. The multi-head attention mechanism, the core component of Transformer, is used to capture multi-angle features of language. BERT is a typical Transformer-based pre-trained model. In order to obtain high-quality word vectors, we choose the BERT model to provide the initial word vector representation for EINET.

The initialized sentence representation is passed into the BERT pre-trained model. BERT uses Byte-Pair Encoding. For example, the words “loved”, “loving”, and “loves” are disassembled into “lov”, “ed”, “ing” and “es”. This method can better reduce the number of vocabulary, improve the training speed, and also link the phenomenon of OOV (Out of Vocabulary).

However, the labels of SRL are given to complete words, so in order to fuse the word embedding obtained by BERT with label representations, the subword representations encoded by BERT need to be aggregated into complete word representations. Inspired by Zhang et al. [6], we adopt convolution and maxpooling to aggregate subwords. (s_1, s_2, \dots, s_l) is the subword sequence of x_i^b , where l is the subword sequence length.

First, the subword sequence is passed to a one-dimensional convolution layer:

$$e_i = W_1[\text{BERT}(s_i), \text{BERT}(s_{i+1}), \dots, \text{BERT}(s_{i+k-1})] + b_1 \tag{1}$$

where k is the convolution kernel size, W_1 and b_1 are trainable vectors. $\text{BERT}(\cdot)$ is vector representation from BERT. Then, subword embeddings are aggregated into a word-level representation vector after Maxpooling:

$$x_i^b = \text{Maxpooling}[\text{ReLU}(e_i), \dots, \text{ReLU}(e_{i-k+1})] \tag{2}$$

$$X^b = (x_1^b, x_2^b, \dots, x_n^b) \tag{3}$$

where n is the length of the input text sequence, and ReLU is a common activation function.

3.1.2. Semantic Role Labeling Information

Explicit semantic information is obtained from SRL. We adopt the most commonly used PropBank-style annotator, which considers one single sentence as a unit to analyze the related local semantic structure of each predicate in the sentence. Semantic structure information is very related to named entity recognition and relation extraction tasks. The role information such as *agent*, *theme*, *time* and *location* can help model extract entities. The relation information between predicates and other words can improve relation extraction to a certain extent.

Figure 2 shows an example for semantic role labeling. SRL centers on the predicate to mark the relationship between other words and predicates in the sentence. In this example, *ARG1* represents the theme, *ArgM-TMP* is an adjunct indicating the timing of the action, *O* represents non-argument word, and *V* represents the predicate.



Figure 2. An example for Semantic Role Labeling.

SRL is predicate-centric to assign labels to words in sentences. Since the concerned predicates are different, the resulting semantic label sequences are different. In order to express the semantic structure of sentences as much as possible, we select five semantic label sequences for each sentence and vectorize them, respectively. Their aggregated representations are then concatenated with the sequence of word vectors from BERT to obtain the final word vector representation.

One semantic role label sequence is represented as:

$$T_1 = (t_1^1, t_2^1, \dots, t_n^1) \tag{4}$$

where $t_1^1, t_2^1, \dots, t_n^1$ are labels of the first one semantic role label sequence.

Five semantic label sequences are aggregated through a full connections layer:

$$T_s = W_2(T_1, T_2, \dots, T_5) + b_2 \tag{5}$$

where W_2 and b_2 are trainable vectors.

The final word vector is represented as:

$$X^w = [X^b; T_s] \tag{6}$$

where $[\cdot]$ is vector concatenation across row.

3.2. Named Entity Recognition

Named entity recognition mainly depends on the context in which the entity is located. Because of different contextual information, words with the same letter have different meanings. Therefore, in addition to introducing explicit semantic information to the word vector representation, we design a novel entity representation algorithm based on Bi-LSTM and Maxpooling.

The entity representation algorithm has two advantages: (1) It enhances the implicit semantics of text sequences and the association between entities and contexts. (2) The algorithm mainly encodes the implicit semantics information of the entity. Except for entities, the rest of the vector representations are not shared between NER and RE. Entities and local contexts are encoded separately to extract richer context implicit features.

Compared with the ordinary recurrent neural network, Bi-LSTM alleviates the problems of gradient vanishing and gradient exploding to a certain extent. Compared with LSTM, Bi-LSTM has the characteristics of capturing bidirectional sequence information.

First, the word vector sequence X^w is passed into Bi-LSTM to build the dependencies between entities and contexts. The Bi-LSTM responsible for obtaining the implicit semantic information of the entity is denoted as Bi-LSTM_e.

$$X^t = \text{Bi-LSTM}_e(X^w) \tag{7}$$

Considering the identification of overlapping entities, we adopt the span-based method to construct candidate entity representations. The word vector representation from $X^t = (x_1^t, x_2^t, \dots, x_i^t, \dots, x_n^t)$ is selected as the candidate entity representation according to any length. A candidate entity vector of length f represents:

$$E^t = (x_i^t, x_{i+1}^t, x_{i+2}^t, \dots, x_{i+f-1}^t) \tag{8}$$

Then, we use Maxpooling to obtain the aggregated entity representation:

$$e^t = \text{Maxpooling}(E^t) \tag{9}$$

Similar to SpERT, we take the length of candidate entity sequences as one of the features affecting entity classification. The entity length representation is searched from the entity length representation matrix according to different lengths. At the same time, the global representation vector CLS obtained by BERT contains rich contextual information, so CLS is also one of the influencing factors of candidate entity classification.

Finally, the candidate entity representation is represented by the aggregation of three parts, which are the entity representation vector, the candidate entity sequence length representation vector, and the global semantic vector CLS.

$$y^e = \text{Softmax}(W_3[e^t; w_f^{ent}; c] + b_3) \tag{10}$$

where e^t is entity representation vector, w_f^{ent} is a representation vector of sequence length f , c is the global semantic vector CLS, W_3 and b_3 are trainable parameter vectors, Softmax is the classification function.

3.3. Relation Extraction

Relation extraction is to predict the relations between the entities in addition to those assigned to the *none* class in the candidate entities. Therefore, the essential basis for relation extraction is entity pair representation. In addition, relation extraction also depends on the context information where entity pairs are located, especially the local context.

For local context information, we adopt Bi-LSTM to enhance its implicit semantic information instead of sharing input with name entity recognition. The Bi-LSTM responsible for obtaining the implicit semantic information of local contexts is denoted as Bi – LSTM_r. In addition, the local context length representation is added. The length of the local context reflects the spacing of entities, which affects the judgment of the relationship between entities. The smaller the entity interval is, the greater the degree of association between entity pairs.

For the global context information, we use the last hidden state obtained by Bi – LSTM_r as the global semantics representation instead of CLS in Equation (10). This is because CLS information obtained by BERT has been added in the entity pair representation, and global semantic information is obtained through different methods, which makes the representation of global semantics more abundant.

First, we enhance implicit semantics of the contextual representation by Bi – LSTM_r.

$$X^r = (x_1^r, x_2^r, \dots, x_n^r) = \text{Bi – LSTM}_r(X^w) \tag{11}$$

The local context is the text sequence from the end of the first entity to the beginning of the second entity. We aggregate local context representation by Maxpooling:

$$c_{a-b} = \text{Maxpooling}(x_{a_{end}}^r, x_{a_{end}+1}^r, \dots, x_{b_{start}}^r) \tag{12}$$

where a_{end} denotes the subscript of the end of the first entity, and b_{start} denotes the subscript of the beginning of the second entity.

The method of obtaining the local context length representation and the entity length representation is similar. Local context length representation is also obtained from the local context length representation matrix, and each local context length representation has its corresponding representation vector. Finally, the entity pair representation vector e_a, e_b , the local context length representation vector w_g^c and the last hidden state h from Bi – LSTM_r are concatenated to form the final relation representation vector. Then, we adopt Softmax classification function to classify the relation representation.

$$y_a^r = \text{Softmax}(W_4[e_a; c_{a-b}; e_b; w_g^c; h] + b_4) \tag{13}$$

$$y_b^r = \text{Softmax}(W_5[e_b; c_{b-a}; e_a; w_g^c; h] + b_5) \tag{14}$$

where w_g^c is the representation vector of the local context length g , W_4, W_5, b_4 and b_5 are the trainable parameter vectors.

Due to the asymmetry of the relationship, the possibility of bidirectional relationship is considered in relation extraction. If any one value of y_a^r and y_b^r does not reach the threshold, it is considered that there is no relationship between entity a and entity b .

4. Experiment and Result Analysis

In this section, we introduce the settings and results related to our experiments, including the introduction of datasets, the experimental parameter settings, and the comparison with advanced methods on three datasets to prove the effectiveness of our model. In addition, we conduct ablation experiments to demonstrate the performance of each component in our model.

4.1. Experimental Settings

4.1.1. Datasets

In this paper, we verify the effectiveness of EINET on three publicly available datasets, namely Conll04, SciERC, and ADE.

Conll04: The Conll04 [29] dataset consists of sentences containing entities and relations extracted from news articles. There are four entity types in total, namely *Location*, *Organization*, *People*, and *Other*. There are five relationship types in total, namely *Work-For*, *Kill*, *Organization-based-in*, *Live-In*, and *Location-In*. We choose 1153 sentences as the training set, 288 sentences as the test set, and 20% of the training set as the validation set, which is consistent with Gupta et al [30].

SciERC: The SciERC [17] dataset is derived from 500 abstracts of papers in artificial intelligence-related fields. It contains a total of six scientific entity types, namely *Task*, *Method*, *Metric*, *Material*, *Other-Scientific-Term*, and *Generic*, and seven relation types, namely *Compare*, *Conjunction*, *Evaluate-For*, *Used-For*, *Feature-Of*, *Part-Of*, and *Hyponym-Of*. We use the same train (1861 sentences), validation (275 sentences) and test (551 sentences) split as in [17].

ADE: The ADE [31] dataset comes from medical reports describing adverse drug reactions, and it contains two entity types, *Adverse-Effect* and *Drug*, and one relationship type, *Adverse-Effects*.

4.1.2. Implementation Details

We choose BERT-Base as the pre-trained model on Conll04 and ADE datasets. However, since SciERC is a dataset related to the scientific field, and SciBERT is a BERT model trained on a large amount of scientific corpus, we choose SciBERT as the pre-trained model on the SciERC dataset. In Equation (10), the dimension of the entity length representation vector is set to 20 which is consistent with our baseline model SpERT [26]. In Equations (13) and (14), the dimension of the local context length representation vector is set to 200 which is selected in [25, 50, 100, 150, 200, 250] based on the development set. As shown in Figure 3, the F1 of NER and RE are the highest when the dimension of the local context length representation is 200. Referring to SpERT [26], the relation filtering threshold is set to 0.4, and the upper bound for sample calculation of relation pairs in each sentence is 100. On the Conll04 dataset, the batch size is set to 2, the learning rate is set to 5×10^{-5} , and the dropout rate is set to 0.5. On the SciERC dataset, the batch size is set to 4, the learning rate is set to 5×10^{-5} , and the dropout rate is set to 0.5. On the ADE dataset, the batch size is set to 8, the learning rate is set to 6×10^{-5} , and the dropout rate is set to 0.5.

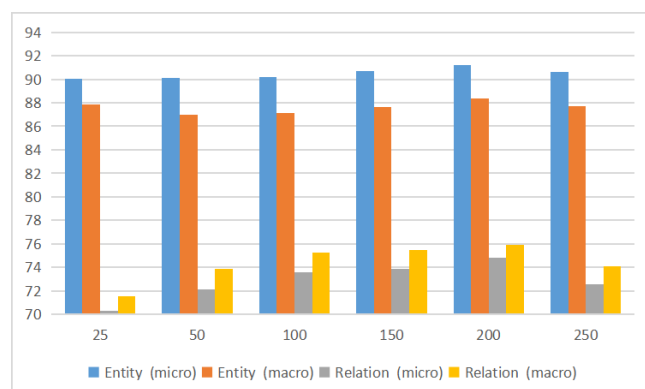


Figure 3. F1 values for named entity recognition (Entity) and relationship extraction (Relation) when the local context length representation vector with dimensions of 25, 50, 100, 150, 200 and 250 (on Conll04 development set).

4.2. Comparison of Results on Datasets

We compare EINET with advanced models on three publicly available datasets. In order to fairly and comprehensively illustrate the effectiveness of models, three evaluation indicators are used: Precision, Recall, and F1. Among them, F1 is the most important evaluation index. For consistency with other models, micro- and macro-average results are compared on Conll04, micro-average results on SciERC, and macro-average results on ADE.

Conll04: We report the average over three runs for the Conll04 dataset. The comparison of methods on the Conll04 dataset is shown in Table 1, where † represents the calculation result of micro-average, ‡ represents the result of macro-average calculation, and * represents that the model does not specify the calculation method.

Table 1. Performance on Conll04 dataset. (The highest F1 is shown in bold. metrics: micro-average = †, macro-average = ‡, not stated = *).

| Model | Entity | | | Relation | | |
|--------------------------------|-----------|--------|--------------|-----------|--------|--------------|
| | Precision | Recall | F1 | Precision | Recall | F1 |
| Relation-Metric [32] ‡ | 84.46 | 84.67 | 84.57 | 67.97 | 58.18 | 62.68 |
| Biaffine Attention [33] ‡ | - | - | 86.20 | - | - | 64.40 |
| Multi-turn QA [18] † | 89.00 | 86.60 | 87.80 | 69.20 | 68.20 | 68.90 |
| Multi-head + AT [16] ‡ | - | - | 83.61 | - | - | 61.95 |
| Multi-head [34] ‡ | 83.75 | 84.06 | 83.90 | 63.75 | 60.43 | 62.04 |
| Hierarchical Attention [35] * | - | - | 86.51 | - | - | 63.32 |
| SpERT [26] † | 88.25 | 89.64 | 88.94 | 73.04 | 70.00 | 71.47 |
| SpERT [26] ‡ | 85.78 | 86.84 | 86.25 | 74.75 | 71.52 | 72.87 |
| MRC4ERE++ [19] * | 89.30 | 88.50 | 88.90 | 72.20 | 71.50 | 71.90 |
| UMT w/ NLGQ [20] * | 88.70 | 88.80 | 88.80 | 72.90 | 71.60 | 72.20 |
| UMT w/ PseudoGQ [20] * | 88.80 | 89.00 | 88.90 | 73.20 | 71.60 | 72.40 |
| TriMF [28] † | 89.26 | 90.34 | 90.30 | 73.01 | 71.63 | 72.35 |
| Two are better than one [36] † | - | - | 90.10 | - | - | 73.80 |
| Two are better than one [36] ‡ | - | - | 86.90 | - | - | 75.80 |
| EINET † | 92.43 | 90.22 | 91.31 | 77.15 | 72.78 | 74.90 |
| EINET ‡ | 90.65 | 86.70 | 88.51 | 77.98 | 74.16 | 75.91 |

Compared with the baseline model SpERT, the F1 of EINET on the entity recognition exceeds 2.37% (micro-average), 2.26% (macro-average), and the F1 on the relation extraction exceeds 3.43% (micro-average), 3.04% (macro-average). EINET is also better than the current advanced model proposed by Wang et al. [36]. The experimental results show that the EINET has achieved relatively advanced experimental results on Conll04 dataset, and the comparison with the baseline model shows that enhanced semantic information can indeed bring obvious benefits.

SciERC: We report the average over three runs for the SciERC dataset. The comparison of methods on the SciERC dataset is shown in Table 2. Compared with the baseline model SpERT, the F1 of EINET on entity recognition is 1.01% higher, and the F1 on relation extraction is 2.09% higher. It is worth noting that, in terms of NER, the performance of EINET is better than the advanced model PL-Marker [37]. The experimental results show that EINET has reached a relatively advanced level on the SciERC dataset. Our proposed method of enriching semantics from explicit and implicit perspectives has a significant positive impact on scientifically relevant dataset.

ADE: The comparison between the proposed model EINET and other methods on the ADE dataset is shown in Table 3. For a fair comparison with existing methods, the final results are averaged over 10 cross-validations. It is worth noting that ADE also contains 120 instances of relations with overlapping entities, which can be discovered by span-based approaches such as EINET and SpERT. These have been filtered in sequence tagging based work [32,34,38]. Compared with the baseline model SpERT, the F1 of EINET is 1.32%

(without overlap) and 1.20% (with overlap) higher in the named entity recognition, and it is 2.80% (without overlap) and 2.54% (with overlap) higher in the relation extraction. This result shows that EINET has advantages in extracting both overlapping and non-overlapping entities.

Table 2. Performance on SciERC dataset. (The highest F1 is shown in bold.)

| Model | Entity | | | Relation | | |
|----------------|-----------|--------|--------------|-----------|--------|--------------|
| | Precision | Recall | F1 | Precision | Recall | F1 |
| SciIE [17] | 67.20 | 61.50 | 64.20 | 47.60 | 33.50 | 39.30 |
| DyGIE [22] | - | - | 65.20 | - | - | 41.46 |
| DyGIE++ [25] | - | - | 67.50 | - | - | 48.40 |
| SpERT [26] | 70.87 | 69.79 | 70.33 | 53.40 | 48.54 | 50.84 |
| UNIRE [39] | 65.80 | 71.10 | 68.40 | 37.30 | 36.60 | 36.90 |
| PFN [40] | - | - | 66.80 | - | - | 38.40 |
| PURE [41] | - | - | 68.90 | - | - | 50.10 |
| TriMF [28] | 70.18 | 70.17 | 70.17 | 52.63 | 52.32 | 52.44 |
| SpERT.PL [43] | 69.82 | 71.25 | 70.53 | 51.94 | 50.62 | 51.25 |
| PL-Marker [37] | - | - | 69.90 | - | - | 53.20 |
| EINET | 71.26 | 71.43 | 71.34 | 55.34 | 50.73 | 52.93 |

Table 3. Performance on ADE dataset. (The highest F1 is shown in bold.)

| Model | Entity | | | Relation | | |
|------------------------------|-----------|--------|--------------|-----------|--------|--------------|
| | Precision | Recall | F1 | Precision | Recall | F1 |
| BiLSTM + SDP [38] | 82.70 | 86.70 | 84.60 | 67.50 | 75.80 | 71.40 |
| Multi-head [34] | 84.72 | 88.16 | 86.40 | 72.10 | 77.24 | 74.58 |
| Multi-head+AT [16] | - | - | 86.73 | - | - | 75.52 |
| Relation-Metric [32] | 86.16 | 88.08 | 87.11 | 77.36 | 77.25 | 77.29 |
| SpERT(without overlap) [26] | 89.26 | 89.26 | 89.25 | 78.09 | 80.43 | 79.24 |
| SpERT(with overlap) [26] | 88.99 | 89.59 | 89.28 | 77.77 | 79.96 | 78.84 |
| PFN [40] | - | - | 89.60 | - | - | 80.00 |
| Two are better than one [36] | - | - | 89.70 | - | - | 80.10 |
| EINET(without overlap) | 90.03 | 91.12 | 90.57 | 80.38 | 83.79 | 82.04 |
| EINET(with overlap) | 89.69 | 91.29 | 90.48 | 79.74 | 83.11 | 81.38 |

4.3. Ablation Analysis

In this section, we investigate the effectiveness of each module in the proposed EINET. The results of the ablation experiment are detailed in Table 4.

Table 4. Ablation study on Conll04 dataset.

| Model | Entity(F1) | | Relation(F1) | |
|--|---------------|---------------|---------------|---------------|
| | Micro-Average | Macro-Average | Micro-Average | Macro-Average |
| 1 EINET | 91.31 | 88.51 | 74.90 | 75.91 |
| 2 w/o SRL | 89.93 | 87.57 | 72.86 | 73.50 |
| 3 w/o Bi – LSTM _e | 90.39 | 87.50 | 73.73 | 74.62 |
| 4 w/o Bi – LSTM _r | 90.72 | 88.01 | 73.46 | 74.63 |
| 5 w/o Bi – LSTM _e and Bi – LSTM _r | 90.19 | 87.33 | 73.16 | 74.39 |
| 6 w/o global semantics (relation) | 91.01 | 88.10 | 74.26 | 75.78 |
| 7 w/o local context length information | 90.85 | 87.87 | 74.07 | 75.52 |
| 8 w/o global semantics (relation) and local context length information | 90.40 | 87.40 | 73.56 | 74.71 |
| 9 Baseline | 88.94 | 86.25 | 71.47 | 72.87 |

To demonstrate the effectiveness of explicit semantic information, we design Model 2 and compare it with the complete model EINET (Model 1). Model 2 is EINET with removal of the vector representation of SRL. Compared with EINET, the F1 of Model 2 in name entity recognition decreased by 1.38% (micro-average) and 0.94% (macro-average), and it is decreased by 2.04% (micro-average) and 2.41% (macro-average) in relation extraction. The obvious drop in model performance after removing the explicit semantic information provided by SRL demonstrates the effectiveness of explicit semantic information. Since the explicit semantic information is introduced in the word representation, removing SRL has a significant impact on both named entity recognition and relation extraction.

To demonstrate the effectiveness of enhancing implicit semantic information, we compare Model 3, Model 4, and Model 5 with the complete model EINET. After removing the named entity recognition Bi-LSTM (Bi-LSTM_e), we obtain Model 3. It is worth noting that Bi-LSTM will double the vector dimension. Therefore, in order to ensure that the dimension of the entity representation vector and the local context vector between entity pairs are consistent, after removing Bi-LSTM_e , we concatenate two identical entity representation vectors to keep the dimensions consistent with the local context representation. Compared with EINET, F1 of Model 3 decreases by 0.92% (micro-average) and 1.01% (macro-average) in named entity recognition, and it decreases by 1.17% (micro-average) and 1.29% (macro-average) in relation extraction. The experimental results show that after removing Bi-LSTM_e , the performance of the model in terms of NER and RE is greatly reduced. A high-quality entity representation will not only improve the performance of NER, but also have a positive impact on RE. After removing the relation extraction Bi-LSTM (Bi-LSTM_r), we obtain Model 4. It is worth noting that in Model 4, the local context representation comes from Bi-LSTM_e , which is shared between NER and RE. The experimental results show that compared with EINET, the F1 of Model 4 decreases by 0.59% (micro average) and 0.50% (macro average) in named entity recognition, and it decreases by 1.44% (micro average) and 1.28% (macro average) in relation extraction. It is proven that enhancing local semantic information has a significant effect on relation extraction. Compared with shared representation, separate encoding can better capture the implicit semantic features of local context. At the same time, NER is positively affected by Bi-LSTM_r . Model 5 removes all implicit semantic enhancement information on the basis of EINET, that is, removes Bi-LSTM_e and Bi-LSTM_r at the same time. The experimental results show that compared with EINET, the F1 of Model 5 decreases by 1.12% (micro average) and 1.18% (macro average) in named entity recognition, and it decreases by 1.74% (micro average) and 1.52% (macro average) in relationship extraction. Removing the enhancement of implicit semantics leads to a large performance drop, which once again illustrates the necessity of the enhancement of implicit semantic information.

To demonstrate the effectiveness of global semantic information and local context length information in relation extraction, we compare Model 6, Model 7, and Model 8 with the complete model EINET. After removing the global semantic information in relation extraction, we obtain Model 6. Compared with EINET, the F1 of Model 6 decreases by 0.30% (micro-average) and 0.41% (macro-average) in named entity recognition, and it decreases by 0.64% (micro average), 0.13% (macro average) in relationship extraction. It shows that global semantic information is helpful to improve the performance of relation extraction, and it can also promote the performance of named entity recognition. Model 7 is obtained after removing the local context length representation information. Compared with EINET, the F1 of Model 7 reduces by 0.46% (micro average) and 0.64% (macro average) in named entity recognition, and it decreases by 0.83% (micro average) and 0.39% (macro average) in relation extraction. It shows that local context length information is helpful to improve the performance of relation extraction, and it can also promote the optimization of named entity recognition. At the same time, compared with the influence of global semantic information in relation extraction, the effect of local context length information is more obvious. Model 8 removes both the global semantic information and local context length information in relation extraction. The experimental results show that compared with

EINET, the F1 of Model 8 reduces by 0.91% (micro-average) and 1.11% (macro-average) in named entity recognition, and it reduces by 1.34% (micro-average) and 1.20% (macro-average) in relation extraction. This experiment further proves that the global semantic information and the local semantic information are effective.

Model 9 is our baseline model. Compared with Model 9, the F1 of EINET is 2.37% (micro-average) higher and 2.26% (macro-average) higher in named entity recognition, 3.43% (micro average) higher and 3.04% (macro average) higher in relation extraction. The experimental results show that the performance gain of EINET in named entity recognition and relation extraction is remarkable, which also confirms that our proposed method of enriching explicit and implicit semantics is effective for joint entity and relation extraction.

4.4. Visualization

In order to better represent the effect of our model, we conduct a visualization on the Conll04 dataset. Some visualizations are shown in Figure 4. As shown in Figure 4a, EINET is able to accurately identify entity-pair relationships that are not identified by the baseline model (*Khrushchev, Live In, Soviet*). As shown in Figure 4b, EINET recognizes the entity type of “DOE” as “Org” and the relation triple (*Steve Wright, Work For, DOE*), but the baseline model can not recognize them. The experimental results show that EINET can extract entities and relationships that cannot be recognized by the baseline model after integrating rich semantic information. The rich semantic information makes the model’s semantic understanding of text more accurate.

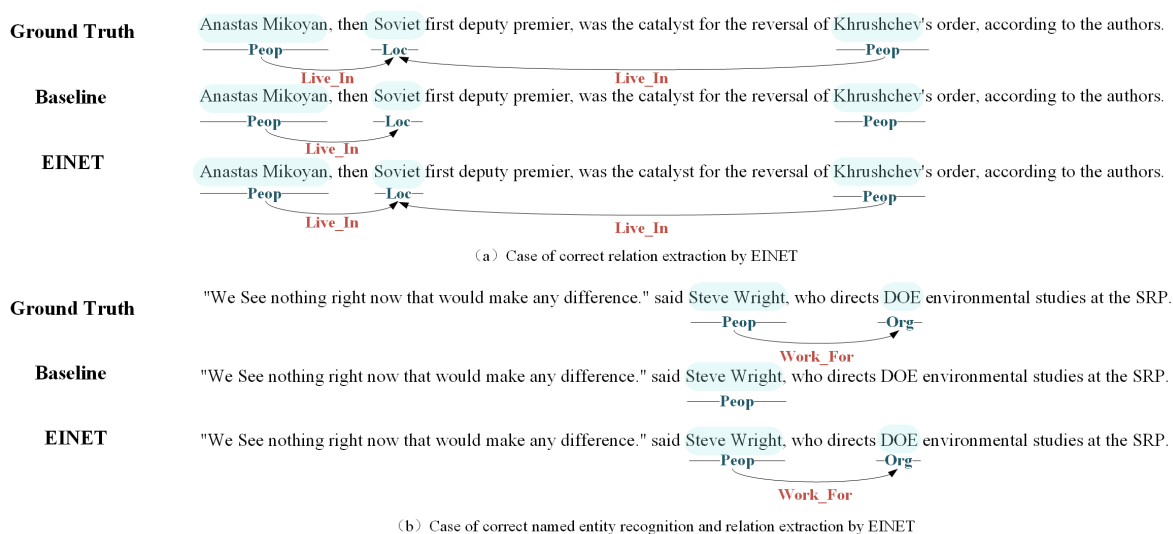


Figure 4. Visualization of joint entity and relation extraction results of different models on Conll04 dataset.

4.5. Error Cases

Although our model achieves competitive results, there are still some errors that leave room for further research. As shown in Table 5, there are a total of three common errors, respectively:

- (1) Incorrect spans: A common error is the prediction of a slightly incorrect entity span, usually with one more or one less word than the ground truth. Here, “interferon alfa” should be marked as an entity but we marke “interferon” as one entity. This error occurs particularly often in domain-specific ADE and SciERC datasets.
- (2) Logical: Sometimes, the relationship between entities is not explicitly described in the sentence, but can be logically inferred from the context. In the case described, the “Work-For” relationship between “Robert Bernero” and “DOE” needs to be inferred from some information (“Robert Bernero , chief of waste disposal for the commission” and “the commission refers to DOE”).

- (3) Missing annotation: There are some cases where a correct prediction is missing in the ground truth. Here, in addition to the correct prediction (Shoshone-Bannock, *Located-In*, Idaho), EIENT also outputs (Hatcher, *Live-In*, Onondaga territory), (Hatcher, *Live-In*, Shoshone-Bannock) and (Hatcher, *Live-In*, Idaho), which are correct but unmarked.

Table 5. Cases of common errors.

| Error Cases | | |
|--------------------|--------------|---|
| Incorrect Spans | Sentences | Cutaneous necrosis after injection of polyethylene glycol—modified interferon alfa. |
| | Ground Truth | Entities: {‘type’: ‘Adverse-Effect’, Cutaneous necrosis} {‘type’: ‘Drug’, interferon alfa} Relations: (interferon alfa, ‘Adverse-Effect’, Cutaneous necrosis) |
| | Our Model | Entities: {‘type’: ‘Adverse-Effect’, Cutaneous necrosis} {‘type’: ‘Drug’, interferon} Relations: (interferon, ‘Adverse-Effect’, Cutaneous necrosis) |
| Logical | Sentences | “NRC has a broad programmatic concern that the pressure to meet unrealistic schedule milestones may leave DOE insufficient time to plan and to execute proper technical information-gathering activities.” said Robert Bernero, chief of waste disposal for the commission. |
| | Ground Truth | Entities: {‘type’: ‘Org’, NRC} {‘type’: ‘Org’, DOE} {‘type’: ‘Peop’, Robert Bernero} Relations: (Robert Bernero, <i>Work-For</i> , DOE) |
| | Our Model | Entities: {‘type’: ‘Org’, NRC} {‘type’: ‘Org’, DOE} {‘type’: ‘Peop’, Robert Bernero} Relations: |
| Missing Annotation | Sentences | Hatcher also fled to the Onondaga territory but has since moved to a Shoshone-Bannock reservation in Idaho. |
| | Ground Truth | Entities: {‘type’: ‘Peop’, Hatcher} {‘type’: ‘Loc’, Onondaga territory} {‘type’: ‘Loc’, Shoshone-Bannock} {‘type’: ‘Loc’, Idaho} Relations: (Shoshone-Bannock, <i>Located-In</i> , Idaho) |
| | Our Model | Entities: {‘type’: ‘Peop’, Hatcher} {‘type’: ‘Loc’, Onondaga territory} {‘type’: ‘Loc’, Shoshone-Bannock} {‘type’: ‘Loc’, Idaho} Relations: (Shoshone-Bannock, <i>Located-In</i> , Idaho) (Hatcher, <i>Live-In</i> , Onondaga territory) (Hatcher, <i>Live-In</i> , Shoshone-Bannock) (Hatcher, <i>Live-In</i> , Idaho) |

5. Conclusions

We propose a novel Joint Entity and Relation Extraction Network with Enhanced Explicit and Implicit Semantic Information (EINET). On the premise of using the pre-trained model, EINET introduces explicit semantic information and fully explores the implicit information for joint entity and relation extraction. Semantic role labelings are vectorized and fused with context representation vectors from BERT. Then, the Bi-LSTM network is used to enhance the implicit semantic information of text. It is worth noting that we adopt different Bi-LSTM networks to capture different features of entities and contexts, respectively. In addition, we introduce the global semantic representation vector based on Bi-LSTM in relation extraction, and add the local context length information to enrich the local semantics. We demonstrate the effectiveness of EINET through comparisons with existing models on three publicly available datasets, and EINET has achieved competitive results. In the next stage, we will further enrich the semantic information of contextual representation based on external knowledge, and explore the common points of joint entity and relation extraction task and other tasks, so as to apply our method to more natural language processing tasks.

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

Funding: This paper was supported by National Key R&D Program of China (2019YFC1521202).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. The data can be accessed at: <https://github.com/lavis-nlp/spert> (accessed on 6 August 2021).

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

Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---------|---|
| NLP | Natural Language Processing |
| NER | Named Entity Recognition |
| RE | Relation Extraction |
| SRL | Semantic Role Labeling |
| BERT | Bidirectional Encoder Representation from Transformers |
| Bi-LSTM | Bi-directional Long Short-Term Memory |
| EINET | Joint Entity and Relation Extraction Network with Enhanced Explicit and Implicit Semantic Information |
| OOV | Out of Vocabulary |

References

- Devlin, J.; Chang, M.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 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; Volume 1, pp. 4171–4186.
- Dai, Z.; Yang, Z.; Yang, Y.; Carbonell, J.G.; Le, Q.V.; Salakhutdinov, R. Transformer-XL: Attentive Language Models beyond a Fixed-Length Context. In Proceedings of the 57th Conference of the Association for Computational Linguistics (ACL 2019), Florence, Italy, 28 July–2 August 2019; Volume 1, pp. 2978–2988.
- Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv* **2019**, arXiv:1907.11692.
- Yang, W.; Xie, Y.; Lin, A.; Li, X.; Tan, L.; Xiong, K.; Li, M.; Lin, J. End-to-End Open-Domain Question Answering with BERTserini. In Proceedings of the 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; pp. 72–77.
- Chatterjee, A.; Narahari, K.N.; Joshi, M.; Agrawal, P. SemEval-2019 Task 3: EmoContext Contextual Emotion Detection in Text. In Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval@NAACL-HLT 2019), Minneapolis, MN, USA, 6–7 June 2019; pp. 39–48.
- Zhang, Z.; Wu, Y.; Zhao, H.; Li, Z.; Zhang, S.; Zhou, X.; Zhou, X. Semantics-Aware BERT for Language Understanding. In Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI 2020), New York, NY, USA, 7–12 February 2020; pp. 9628–9635.
- Zhang, Z.; Wu, Y.; Zhou, J.; Duan, S.; Zhao, H.; Wang, R. SG-Net: Syntax-Guided Machine Reading Comprehension. In Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI 2020), New York, NY, USA, 7–12 February 2020; pp. 9636–9643.
- Geng, Z.; Zhang, Y.; Han, Y. Joint entity and relation extraction model based on rich semantics. *Neurocomputing* **2021**, *429*, 132–140. [[CrossRef](#)]
- Zheng, S.; Wang, F.; Bao, H.; Hao, Y.; Zhou, P.; Xu, B. Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017), Vancouver, BC, Canada, 30 July–4 August 2017; pp. 1227–1236.
- Takanobu, R.; Zhang, T.; Liu, J.; Huang, M. A Hierarchical Framework for Relation Extraction with Reinforcement Learning. In Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI 2019), Honolulu, HI, USA, 27 January–1 February 2019; pp. 7072–7079.
- Zeng, X.; Zeng, D.; He, S.; Liu, K.; Zhao, J. Extracting Relational Facts by an End-to-End Neural Model with Copy Mechanism. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL 2018), Melbourne, VIC, Australia, 15–20 July 2018; Volume 1, pp. 506–514.
- Zeng, D.; Zhang, H.; Liu, Q. CopyMTL: Copy Mechanism for Joint Extraction of Entities and Relations with Multi-Task Learning. In Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI 2020), New York, NY, USA, 7–12 February 2020; pp. 9507–9514.
- Yu, B.; Zhang, Z.; Shu, X.; Liu, T.; Wang, Y.; Wang, B.; Li, S. Joint Extraction of Entities and Relations Based on a Novel Decomposition Strategy. In Proceedings of the ECAI 2020—24th European Conference on Artificial Intelligence, Santiago de Compostela, Spain, 29 August–8 September 2020; pp. 2282–2289.

14. 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 (ACL 2020), Online, 5–10 July 2020; pp. 1476–1488.
15. Yuan, Y.; Zhou, X.; Pan, S.; Zhu, Q.; Song, Z.; Guo, L. A Relation-Specific Attention Network for Joint Entity and Relation Extraction. In Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI 2020), Yokohama, Japan, 11–17 June 2020; pp. 4054–4060.
16. Bekoulis, G.; Deleu, J.; Demeester, T.; Develder, C. Adversarial training for multi-context joint entity and relation extraction. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 2830–2836.
17. 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.
18. Li, X.; Yin, F.; Sun, Z.; Li, X.; Yuan, A.; Chai, D.; Zhou, M.; Li, J. Entity-Relation Extraction as Multi-Turn Question Answering. In Proceedings of the 57th Conference of the Association for Computational Linguistics (ACL 2019), Florence, Italy, 28 July–2 August 2019; Volume 1, pp. 1340–1350.
19. Zhao, T.; Yan, Z.; Cao, Y.; Li, Z. Asking Effective and Diverse Questions: A Machine Reading Comprehension based Framework for Joint Entity-Relation Extraction. In Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI 2020), Yokohama, Japan, 7–15 January 2020; pp. 3948–3954.
20. Zhao, T.; Yan, Z.; Cao, Y.; Li, Z. A Unified Multi-Task Learning Framework for Joint Extraction of Entities and Relations. In Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI 2021), Online, 2–9 February 2021; pp. 14524–14531.
21. Dixit, K.; Al-Onaizan, Y. Span-Level Model for Relation Extraction. In Proceedings of the 57th Conference of the Association for Computational Linguistics (ACL 2019), Florence, Italy, 28 July–2 August 2019; Volume 1, pp. 5308–5314.
22. 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 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; Volume 1, pp. 3036–3046.
23. Chen, J.; Yuan, C.; Wang, X.; Bai, Z. MrMep: Joint Extraction of Multiple Relations and Multiple Entity Pairs Based on Triplet Attention. In Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL 2019), Hong Kong, China, 3–4 November 2019; pp. 593–602.
24. 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, Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
25. Wadden, D.; Wennberg, U.; Luan, Y.; Hajishirzi, H. Entity, Relation, and Event Extraction with Contextualized Span Representations. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Hong Kong, China, 3–7 November 2019; pp. 5783–5788.
26. Eberts, M.; Ulges, A. Span-Based Joint Entity and Relation Extraction with Transformer Pre-Training. In Proceedings of the ECAI 2020—24th European Conference on Artificial Intelligence, Santiago de Compostela, Spain, 29 August–8 September 2020; pp. 2006–2013.
27. Ji, B.; Yu, J.; Li, S.; Ma, J.; Wu, Q.; Tan, Y.; Liu, H. Span-based Joint Entity and Relation Extraction with Attention-based Span-specific and Contextual Semantic Representations. In Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020), Barcelona, Spain, 8–13 December 2020; pp. 88–99.
28. Shen, Y.; Ma, X.; Tang, Y.; Lu, W. A Trigger-Sense Memory Flow Framework for Joint Entity and Relation Extraction. In Proceedings of the WWW '21: The Web Conference 2021, Ljubljana, Slovenia, 19–23 April 2021; pp. 1704–1715.
29. Roth, D.; Yih, W. A Linear Programming Formulation for Global Inference in Natural Language Tasks. In Proceedings of the Eighth Conference on Computational Natural Language Learning (CoNLL 2004), Boston, MA, USA, 6–7 May 2004; pp. 1–8.
30. Gupta, P.; Schütze, H.; Andrassy, B. Table Filling Multi-Task Recurrent Neural Network for Joint Entity and Relation Extraction. In Proceedings of the COLING 2016—26th International Conference on Computational Linguistics, Osaka, Japan, 11–16 December 2016; pp. 2537–2547.
31. Gurulingappa, H.; Rajput, A.M.; Roberts, A.; Fluck, J.; Hofmann-Apitius, M.; Toldo, L. Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports. *J. Biomed. Inform.* **2012**, *45*, 885–892. [[CrossRef](#)] [[PubMed](#)]
32. Tran, T.; Kavuluru, R. Neural Metric Learning for Fast End-to-End Relation Extraction. *arXiv* **2019**, arXiv:1905.07458.
33. Nguyen, D.Q.; Verspoor, K. End-to-End Neural Relation Extraction Using Deep Biaffine Attention. In Proceedings of the Advances in Information Retrieval—41st European Conference on IR Research (ECIR 2019), Cologne, Germany, 14–18 April 2019; pp. 729–738.
34. Bekoulis, G.; Deleu, J.; Demeester, T.; Develder, C. Joint entity recognition and relation extraction as a multi-head selection problem. *Expert Syst. Appl.* **2018**, *114*, 34–45. [[CrossRef](#)]
35. Chi, R.; Wu, B.; Hu, L.; Zhang, Y. Enhancing Joint Entity and Relation Extraction with Language Modeling and Hierarchical Attention. In Proceedings of the Web and Big Data—Third International Joint Conference (APWeb-WAIM 2019), Chengdu, China, 1–3 August 2019; pp. 314–328.

36. Wang, J.; Lu, W. Two are Better than One: Joint Entity and Relation Extraction with Table-Sequence Encoders. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2020), Online, 16–20 November 2020; pp. 1706–1721.
37. 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, Dublin, Ireland, 22–27 May 2022; pp. 4904–4917.
38. Li, F.; Zhang, M.; Fu, G.; Ji, D. A neural joint model for entity and relation extraction from biomedical text. *BMC Bioinform.* **2017**, *18*, 198. [[CrossRef](#)] [[PubMed](#)]
39. Wang, Y.; Sun, C.; Wu, Y.; Zhou, H.; Li, L.; Yan, J. UniRE: A Unified Label Space for Entity Relation Extraction. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics, Online, 1–6 August 2021; pp. 220–231.
40. Yan, Z.; Zhang, C.; Fu, J.; Zhang, Q.; Wei, Z. A Partition Filter Network for Joint Entity and Relation Extraction. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021), Punta Cana, Dominican Republic, 7–11 November 2021; pp. 185–197.
41. Zhong, Z.; Chen, D. A Frustratingly Easy Approach for Entity and Relation Extraction. In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2021), Online, 6–11 June 2021; pp. 50–61.
43. Sai, S.T.Y.S.; 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), Online, 29–30 September 2021; pp. 15–19.
43. Sai, S.T.Y.S.; 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), Online, 29–30 September 2021; pp. 15–19.