

Article

A Study on Double-Headed Entities and Relations Prediction Framework for Joint Triple Extraction

Yanbing Xiao ¹, Guorong Chen ^{1,*}, Chongling Du ¹, Lang Li ¹, Yu Yuan ¹, Jincheng Zou ¹ and Jingcheng Liu ²

¹ Department of Intelligent Technology and Engineering, Chongqing University of Science and Technology, Chongqing 401331, China; 2021208022@cqust.edu.cn (Y.X.); 2022208004@cqust.edu.cn (C.D.); 2021208019@cqust.edu.cn (L.L.); 2022208050@cqust.edu.cn (Y.Y.); 2023208002@cqust.edu.cn (J.Z.)

² China Academy of Liquor Industry, Luzhou Vocational and Technical College, Luzhou 646608, China; liujingcheng1980@126.com

* Correspondence: cgr@cqust.edu.cn

Abstract: Relational triple extraction, a fundamental procedure in natural language processing knowledge graph construction, assumes a crucial and irreplaceable role in the domain of academic research related to information extraction. In this paper, we propose a Double-Headed Entities and Relations Prediction (DERP) framework, which divides the entity recognition process into two stages: head entity recognition and tail entity recognition, using the obtained head and tail entities as inputs. By utilizing the corresponding relation and the corresponding entity, the DERP framework further incorporates a triple prediction module to improve the accuracy and completeness of the joint relation triple extraction. We conducted experiments on two English datasets, NYT and WebNLG, and two Chinese datasets, DuIE2.0 and CMeIE-V2, and compared the English dataset experimental results with those derived from ten baseline models. The experimental results demonstrate the effectiveness of our proposed DERP framework for triple extraction.

Keywords: triple extraction; entity recognition; relation extraction; joint extraction

MSC: 68T50 Natural language processing



Citation: 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. <https://doi.org/10.3390/math11224583>

Academic Editor: Florentina Hristea

Received: 23 October 2023

Revised: 7 November 2023

Accepted: 7 November 2023

Published: 8 November 2023



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

1. Introduction

With the development of natural language processing and knowledge graphs, data storage and presentation methods for structured text have become more mature, but there are still many unsolved problems in the processing of unstructured and semi-structured text [1]. Extracting triple groups is crucial in natural language processing and knowledge graph construction. In constructing knowledge graphs, unstructured texts usually extract entities and form correspondences by forming a (head entity, relation, tail entity) triple.

Existing triple extraction methods mainly include two major kinds, pipeline extraction methods and joint extraction methods. Traditional pipeline extraction methods divide knowledge extraction into two subtasks [2]: named entity recognition and relation extraction. However, this approach ignores potential information interactions between entities and relations, leading to incorrect relation extractions or failure to recognize entity relations. Many previous experiments have demonstrated that a joint learning approach greatly improves the effectiveness of entity and relation extraction due to the consideration of the information interactions between the two subtasks, so most of the current research for the task of entity and relation extraction adopts the joint learning approach.

In recent scholarship, there has been a notable surge in research attention directed toward the intricacies of overlapping triples, as shown in Figure 1. This phenomenon is exemplified in sentences wherein there is the potential presence of both entity pair overlap (EPO) triples and single entity overlap (SEO) triples. This burgeoning area of inquiry

underscores the escalating interest in dissecting and comprehending the complexities inherent to overlapping triples in textual data.



Figure 1. Normal, entity pair overlap (EPO) triple, and single entity overlap (SEO) triple cases. In each example, overlapping entities are marked with the same color.

Previous research has revealed several shortcomings in the extraction of multiple relationships (overlapping triples) within the same entity. For example, the NovelTagging method uses a joint decoding of sequence annotations to treat entity and relation extraction as a sequence annotation problem [3]; however, this method only assigns a single label to each token, rendering it incapable of handling overlapping triples in the data. In contrast, the CasRel framework models relations as functions that map subject to object [4], successfully overcoming the issue of poor handling of overlapping triples by previous models. Nevertheless, the CasRel framework suffers from the disadvantage of incorrectly identifying the head entity, leading to failure in identifying the relation and the tail entity. An overview of the CasRel framework structure is shown in Figure 2.

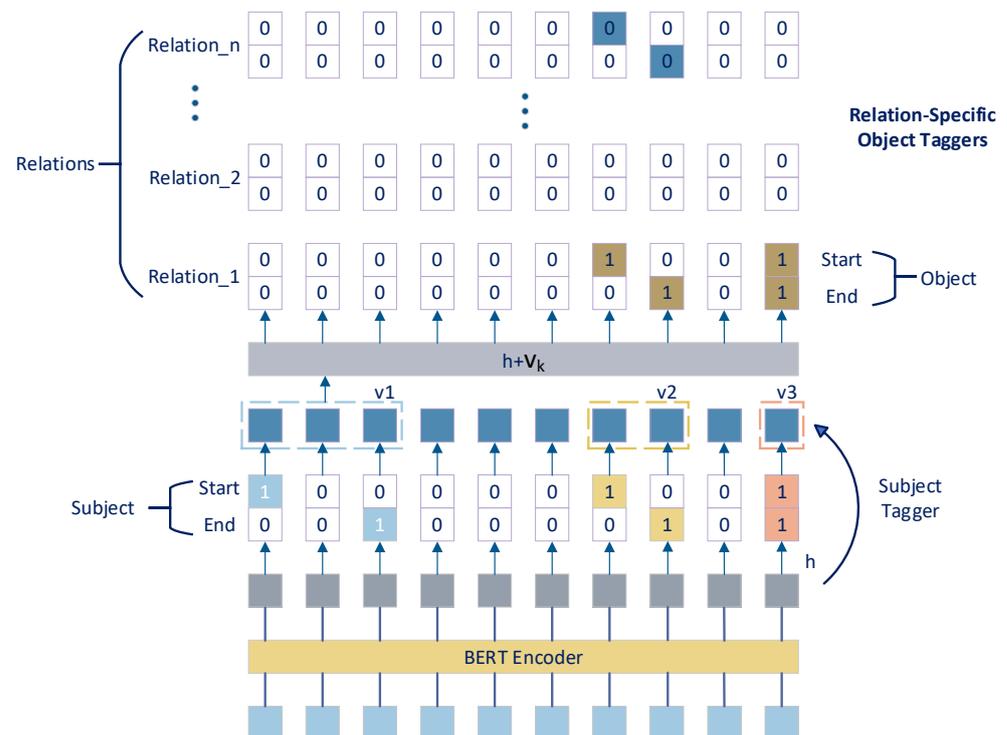


Figure 2. Overview of the CasRel framework structure.

In this study, a head entity recognition module is used to predict the triple related to the head entity and a tail entity recognition module is added to predict the triple related to the tail entity. Combining the information from the two modules results in a triple of higher accuracy. Experimental results show that the performance of the framework is improved by combining the BERT encoder. This work contributes as follows:

1. A double-headed entities and relations prediction framework for joint triple extraction based on the BERT encoder is proposed. The named entity recognition task is decomposed into head entity recognition and tail entity recognition.
2. To ensure recognition accuracy, a triple prediction module, which gives different weights to the triple derived from the head entity recognition and the triple extracted from the tail entity recognition, is set up to improve the accuracy of triple extraction.
3. To validate the method, experiments were conducted on two English public datasets, NYT and WebNLG, and two Chinese datasets, DuIE2.0 and CMeIE-V2, and the proposed framework was compared with ten baselines.

2. Related Work

In recent years, many methods have been proposed to accomplish knowledge extraction that can be categorized into pipeline extraction methods and joint learning methods based on the learning process.

2.1. Pipeline Extraction Methods

Usually, pipeline extraction methods consist of the entity recognition stage and the relation extraction stage, where the output of the previous stage becomes the input of the next stage. This approach has the advantage that a specific model can handle a responding task, but it may also lead to errors accumulating in each stage.

The primary objective of named entity recognition (NER) is to identify and classify named entities within textual content, such as people, places, time, purpose, etc., with specific meanings. It is mainly responsible for automatically extracting the basic element entities in the knowledge graph from the unstructured and semi-structured. In order to uphold the quality of the knowledge graph, it is imperative to ascertain the precision and comprehensiveness of the entities extracted therein. Li et al. proposed a meta-learning method, integrating distributed systems with a meta-learning approach to extract relations among Chinese entities [5]. Through the utilization of machine learning and neural network methodologies, particularly leveraging the attention mechanism within the domain of natural language processing, Li et al. proposed a combination of conditional random fields (CRF) and bidirectional long short-term memory (BiLSTM) for extracting information in a mathematical language [6]. Luo et al. introduced a neural network model, known as the attention-based bidirectional long short-term memory with a conditional random field layer (Att-BiLSTM-CRF), for document-level chemical entity recognition [7]. Li et al. advocated the utilization of distinct layers, specifically long short-term memory (LSTM) for text feature extraction and conditional random field (CRF) for label prediction decoding [8]. Ren proposed a method to enhance entity recognition by transforming text into a vector representation combining contextual and global features through a pretrained model and a graph neural network GCN [9].

Relation extraction refers to extracting relations between connecting basic element entities from the unstructured and semi-structured. The mesh structure of the knowledge graph is similar to the structure of the brain for storing knowledge. Neurons represent entities and record basic information, and the process of extracting relations activates some of the neurons (entities) and adds them to the brain structure (knowledge graph), using relations to connect the entities to the whole knowledge graph. Zeng et al. conducted an analysis of the pivotal role played by the order of relation extraction and employed reinforcement learning techniques to ameliorate the efficiency of relation extraction [10]. Han et al. proposed a one pass model based on BERT, capable of predicting entity relations by processing the text in a single pass [11]. Chen et al. utilized a neutralized feature engi-

neering approach for entity relation extraction, namely, enhancing neural networks with manually designed features [12]. Yuan et al. proposed a relation-aware attention network to construct relation-specific sentence representations [13]. Wan et al. proposed a span-based multi-modal attention network (SMAN) for joint entity and relation extraction [14].

2.2. Joint Learning Methods

In pipeline learning methods of relation extraction, the intrinsic connection between entities and relations is often overlooked, and the federated model is an excellent solution to this problem. Huang et al. suggested using soft label embedding as an effective means to facilitate information exchange between entity recognition and relation extraction [15]. Wei et al. proposed a novel cascade binary tagging framework (CASREL), which models relations as functions that map subjects to objects [4]. Liu et al. introduced an attention-driven integrated model, primarily comprising an entity extraction module and a relation detection module, as a means to effectively confront the prevailing challenges [16]. Yu et al. decomposed the comprehensive extraction task into two mutually interconnected subtasks: one subtask handles the head entities, and the other subtask deals with the tail entities related to the head entities and their respective relations [17]. Guo et al. introduced an integrated model for the extraction of entities and relations pertaining to concepts within the realm of cybersecurity (CyberRel) [18], and they adopted a perspective wherein the triple is conceived as a sequence of entity relations. Subsequently, Lv et al. constructed the joint extraction of entity mentions and relations model, which was based on the bidirectional long short-term memory and maximum entropy Markov model (Bi-MEMM) [19]. Zheng et al. introduced an integrated framework for extracting relational triples, underpinned by the principles of potential relation and global correspondence (PRGC) [20]. Li et al. proposed a relation-aware embedding mechanism (RA) for relation extraction, with attention mechanisms being used to merge relational tags into sentence embeddings, which are used to distinguish the importance of relational tags for each word [21]. Huang et al. proposed a novel translation-based unified framework, which is used to solve redundant predictions, overlapping triples, and relation connections problems [22]. Liu et al. presented a model referred to as the bidirectional encoder representation from transformers–multiple convolutional neural network (BERT–MCNN), which has demonstrated a high level of accuracy and stability [23].

3. CasRel Framework

The goal of triple extraction is to identify all possible triples (head entity, relation, tail entity) in a sentence, which may contain some overlapping and shared entities. The structure of the CasRel framework is shown in Figure 2. The CasRel framework presents a fresh perspective on the task of triple extraction. It introduces a novel cascade binary tagging framework, known as CasRel, that effectively addresses the complex challenge of managing overlapping relations by systematically establishing subject–object mappings within sentences [4]. This framework consists of a set of functions that identify entities and their related relations in an entity tagger and relation-specific object taggers. By employing the CasRel framework, the issue of sharing the same entity in multiple triples is addressed effectively, providing multiple related relations and corresponding entities for each entity. However, in the CasRel framework, if the subject tagger does not recognize an entity, the associated triad will be missed.

To solve the triple extraction omission that occurs in the CasRel framework, we propose an improved DERP framework based on the CasRel framework. Which improves the entity recognition accuracy by adding a tail entity recognition module in the entity tagger, and adding a triple prediction module after relation-specific object taggers. This framework will combine head entities, tail entities and relations to make predictions and comes up with a more accurate triple.

4. The DERP Framework

Entity recognition and relation extraction are the design priorities for triple extraction. The primary objective of this DERP framework is to ascertain the complete set of potential triples within a given sentence, acknowledging the potential existence of entities with overlapping attributes in some instances.

The ultimate prediction of the (head entity, relation, tail entity) triple is achieved through the recognition and forecasting of the acquired triples within the triple prediction layer. The DERP framework is shown in Figure 3.

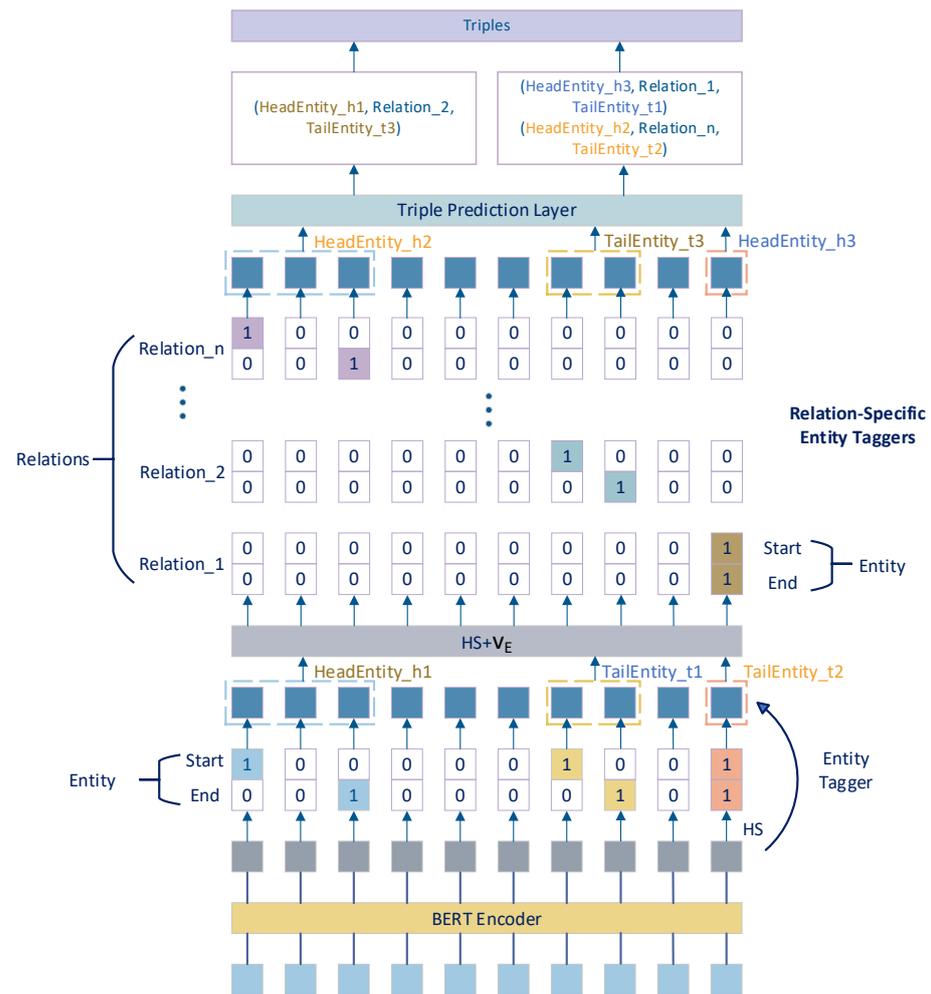


Figure 3. The architecture of the proposed DERP framework. In the framework, the start and end positions of predicted entities and relations are color-marked, with entities belonging to the same group marked with the same color.

In the DERP framework, we model relations as functions that map topics to objects. We optimize the previously commonly used learning relation classifiers $f(E_1, E_2) \rightarrow R$, to learning relation-specific taggers $f_R(E_1) \rightarrow E_2$. Each tagger will identify entities that may exist under a specific relationship, or entities that may not be returned. If the entity is not returned, it indicates that there is no triple in the current entity and relation.

When dealing with overlapping triples, the DERP framework uses an entity tagger for entity recognition and allows multiple relationship representations in relation-specific entity taggers. Within relation-specific entity taggers, multiple relationships and their corresponding entities can be obtained. By using the DERP framework, different types of data structures, including EPO triples and SEO triples, can be effectively handled.

We used an entity tagger to identify head entities at the very beginning of the research on framework development and used the identified head entities to find related relations and tail entities. During the experiments, it was found that if there is a head entity in the entity tagger that is missing, this triple will be missed in the final triple prediction, especially in the case of overlapping triples where a head entity corresponds to more than one related tail entity. There are also cases where some of the tail entities related to this head entity are missed when performing the triple extraction; in this case, we can better find these missing tail entities by adding a tail entity recognition module to the entity tagger. So, two matching entities and accurate relations between entities are achieved by adding a tail entity recognition module to the entity tagger, and by looking up the corresponding relation and another matching entity in the relation-specific entity taggers.

During the experiment, by learning and improving the previous model, we added the tail entity recognition module. If the probability of recognizing the correct triple by the head entity module only is $P(Head)$ and the probability of recognizing the correct triple by the tail entity module only is $P(Tail)$, we will increase the probability of finally recognizing the correct triple by combining the two entity modules with the following probability equation:

$$\begin{aligned} P(Triple) &= P(Head) \cup P(Tail) \\ &= P(Head) + P(Tail) - P(Head \cap Tail) \end{aligned} \quad (1)$$

where $P(Triple)$ is the probability of obtaining the correct triple, $P(Head)$ is the probability of obtaining the correct triple by only using a single head entity recognition module, $P(Tail)$ is the probability of obtaining the correct triple by only using a single tail entity recognition module, and $P(Head \cap Tail)$ is the probability of duplicate triples obtained by the head entity recognition module and tail entity recognition module.

4.1. BERT Encoder

BERT mainly consists of N layers of transformer block. A BERT encoder extracts sentence feature information from sentence S and inputs the feature information into the entity tagger.

$$h_0 = O_{hot}W_n + W_p \quad (2)$$

$$S_{ri} = BERT(ri) \quad (3)$$

where O_{hot} is the one-hot vector matrix indexed in the input sentence, W_n is the word embedding matrix, W_p is the positional embedding matrix, p in W_p denotes the positional index in the input sequence, and S_{ri} is the i-th relation type embedding.

4.2. Entity Tagger

Compared with the CasRel framework, the entity recognition is divided into head entity recognition and tail entity recognition in the entity tagger, which reduces the situation of missing triples due to the omission of the first stage of entity recognition, and also improves the accuracy of the extraction of overlapping triples [24].

The BERT encoded sentence is entered in the entity tagger to extract head and tail entities by the binary method.

Within the entity tagger, the identification of entity positions within sentences encoded by the BERT encoder is achieved. In this module, two binary classifiers are designed to check for the start and end positions of entity words. By setting specific thresholds, if the probability surpasses the designated threshold, the token is marked as 1; otherwise, it is marked as 0. The following is specific to the head entity recognizer and tail entity recognizer:

$$p_i^{HE_start} = \text{sigmoid}\left(W_{start}^{HE}x_i^{HE} + b_{start}^{HE}\right) \quad (4)$$

$$p_i^{HE_end} = sigmoid(W_{end}^{HE} x_i^{HE} + b_{end}^{HE}) \tag{5}$$

$$p_i^{TE_start} = sigmoid(W_{start}^{TE} x_i^{TE} + b_{start}^{TE}) \tag{6}$$

$$p_i^{TE_end} = sigmoid(W_{end}^{TE} x_i^{TE} + b_{end}^{TE}) \tag{7}$$

where $p_i^{HE_start}$, $p_i^{HE_end}$, $p_i^{TE_start}$, and $p_i^{TE_end}$ are the probability of the marker position being predicted to be the start and end positions of the head entity and the tail entity, x_i denotes the i -th marker in sentence S , W_{start}^{HE} , W_{end}^{HE} , W_{start}^{TE} , and W_{end}^{TE} denote the training weights of the head entities and tail entities, and b_{start}^{HE} , b_{end}^{HE} , b_{start}^{TE} , and b_{end}^{TE} denote the bias of the head entities and tail entities. In the use of the model, we need to keep the dimensions of the start binary classifier and the end binary classifier the same.

The entity recognition module uses the following likelihood function to recognize the range of sentences that have been encoded by the encoder:

$$p_{\theta}(E_{Head} | X_{HE}) = \prod_{t \in \{HE_start, HE_end\}} \prod_{i=1}^L (p_i^t)^{I\{y_i^t=1\}} (1 - p_i^t)^{I\{y_i^t=0\}} \tag{8}$$

$$p_{\theta}(E_{Tail} | X_{TE}) = \prod_{t \in \{TE_start, TE_end\}} \prod_{i=1}^L (p_i^t)^{I\{y_i^t=1\}} (1 - p_i^t)^{I\{y_i^t=0\}} \tag{9}$$

where L is the length of the sentence, $I\{z\} = 1$ if z is true and 0 otherwise, $y_i^{HE_start}$, $y_i^{HE_end}$, $y_i^{TE_start}$, and $y_i^{TE_end}$ are the i -th tag in the sequence that marks the start position and the end position.

4.3. Relation-Specific Entity Taggers

In the relation-specific entity taggers, an entity tagger is assigned to each relation word. The relation terms are used to correspond to the head entity or tail entity extracted in the previous layer to extract the entity in satisfying the relations. The calculations are shown below:

$$p_i^{HR_start} = sigmoid(W_{start}^{HR} (x_i^{HE} + v_E^k) + b_{start}^{HR}) \tag{10}$$

$$p_i^{HR_end} = sigmoid(W_{end}^{HR} (x_i^{HE} + v_E^k) + b_{end}^{HR}) \tag{11}$$

$$p_i^{TR_start} = sigmoid(W_{start}^{TR} (x_i^{TE} + v_E^k) + b_{start}^{TR}) \tag{12}$$

$$p_i^{TR_end} = sigmoid(W_{end}^{TR} (x_i^{TE} + v_E^k) + b_{end}^{TR}) \tag{13}$$

where $p_i^{HR_start}$, $p_i^{HR_end}$, $p_i^{TR_start}$, and $p_i^{TR_end}$ are the probabilities that the head entity and the tail entity at the labeled position are predicted to be the entity start position and end position, v_E^k is the relation-specific entity tagger's vector of coded representations of the k th subject detected in the module, W_{start}^{HE} , W_{end}^{HE} , W_{start}^{TE} , and W_{end}^{TE} denote the training weights of the head entities and tail entities, and b_{start}^{HE} , b_{end}^{HE} , b_{start}^{TE} , and b_{end}^{TE} indicate deviations of head entities and tail entities.

Relation-specific entity taggers use the following likelihood function to identify the range of sentences that the encoder has encoded:

$$p_{\theta}(E_{Tail} | E_{Head}, X_{HE}) = \prod_{t \in \{HE_start, HE_end\}} \prod_{i=1}^L (p_i^t)^{I\{y_i^t=1\}} (1 - p_i^t)^{I\{y_i^t=0\}} \tag{14}$$

$$p_{\theta}(E_{Head} | E_{Tail}, X_{TE}) = \prod_{t \in \{TE_start, TE_end\}} \prod_{i=1}^L (p_i^t)^{I\{y_i^t=1\}} (1 - p_i^t)^{I\{y_i^t=0\}} \tag{15}$$

where L is the length of the sentence, $I\{z\} = 1$ if z is true and 0 otherwise, and $y_i^{HE_start}$, $y_i^{HE_end}$, $y_i^{TE_start}$, and $y_i^{TE_end}$ are the i -th tags in the sequence that marks the start position and the end position.

4.4. Triple Prediction

The relation-specific entity taggers identify the head entity, tail entity, and the corresponding relations and use the method of entity relation prediction to match the head entities and tail entities identified in the entity tagger using the following method:

$$f_{HE_start} = \begin{cases} 1, & p_i^{HR_start} \geq \lambda_1 \\ 0, & p_i^{HR_start} < \lambda_1 \end{cases} \tag{16}$$

$$f_{HE_end} = \begin{cases} 1, & p_i^{HR_end} \geq \lambda_2 \\ 0, & p_i^{HR_end} < \lambda_2 \end{cases} \tag{17}$$

$$f_{TE_start} = \begin{cases} 1, & p_i^{TR_start} \geq \lambda_3 \\ 0, & p_i^{TR_start} < \lambda_3 \end{cases} \tag{18}$$

$$f_{TE_end} = \begin{cases} 1, & p_i^{TR_end} \geq \lambda_4 \\ 0, & p_i^{TR_end} < \lambda_4 \end{cases} \tag{19}$$

When f_{HE_start} , f_{HE_end} , f_{TE_start} , and f_{TE_end} equal to 1, the head entity or tail entity corresponding to the entity extracted in entity tagger and the corresponding relation is obtained, and if the value is equal to 0, the triple is excluded. λ_1 , λ_2 , λ_3 , and λ_4 are the set thresholds.

$$g_{HE_TE} = \{HeadEntity, Relation, TailEntity\} \tag{20}$$

$$g_{TE_HE} = \{HeadEntity, Relation, TailEntity\} \tag{21}$$

$$g = g_{HE_TE} \cup g_{TE_HE} \tag{22}$$

where g_{HE_TE} represents the triplets of the tail entity and the relation between entities obtained based on the head entity, g_{TE_HE} represents the triplets of the head entity and the relation between entities obtained based on the tail entity, and g denotes the final predicted triplets.

4.5. Loss Function

The training loss is defined as below:

$$\begin{aligned} \mathcal{L} = \sum_{j=1}^{|D|} & \left[\sum_{E \in T_j} \log p_{\theta}(E_{Head} | X_j^{HE}) + \sum_{E \in T_j} \log p_{\theta}(E_{Tail} | X_j^{TE}) \right. \\ & + \sum_{r \in T_j | E} \log p_{\phi_r}(E_{Tail} | E_{Head}, X_j^{HE}) \\ & + \sum_{r \in T_j | E} \log p_{\phi_r}(E_{Head} | E_{Tail}, X_j^{TE}) \\ & + \sum_{r \in R \setminus T_j | E} \log p_{\phi_r}(E_{Tail\emptyset} | E_{Head}, \mathbf{x}_j) \\ & \left. + \sum_{r \in R \setminus T_j | E} \log p_{\phi_r}(E_{Head\emptyset} | E_{Tail}, \mathbf{x}_j) \right] \tag{23} \end{aligned}$$

where parameters $\theta = \{\theta, \{\emptyset_r\}_{r \in R}\}$, $p_{\theta}(E_{Head} | X_{HE})$, and $p_{\theta}(E_{Tail} | X_{TE})$ are defined in Equations (7) and (8), and $p_{\phi_r}(E_{Tail} | E_{Head}, X_{HE})$ and $p_{\phi_r}(E_{Head} | E_{Tail}, X_{TE})$ are defined in Equations (13) and (14).

5. Experiments

The effectiveness of the proposed framework is validated with experiments. The datasets and evaluation metrics are first introduced, and then the model names are compared with different baseline models.

5.1. Experiment Setup and Experiment Description

As most of the previous studies conducted experiments using English datasets, this study conducted experiments using two publicly English available datasets, NYT [25] and WebNLG [26], and compared the results of the experiments with 10 baseline models. Due to the specificity of the Chinese language, the complexity and difficulty of Chinese triple extraction is considerably greater than that of English relations [27]. We used two Chinese datasets, DuIE2.0 [28] and CMeIE-V2 [29]. DuIE2.0 is the most comprehensive Chinese relational extraction dataset in the industry [30]. CMeIE-V2 is a Chinese medical information extraction dataset, specifically designed for pediatrics and covering more than a hundred common diseases.

This model performs head entity recognition and tail entity recognition in the entity recognition part and performs the corresponding triple extraction based on the experimental results. In the experiments, the head entity recognition model and the tail entity recognition model are used individually for comparison experiments to verify the reliability and validity of the experiments. The schematic diagram of the head entity recognition module and the tail entity recognition module is shown in Figure 4.

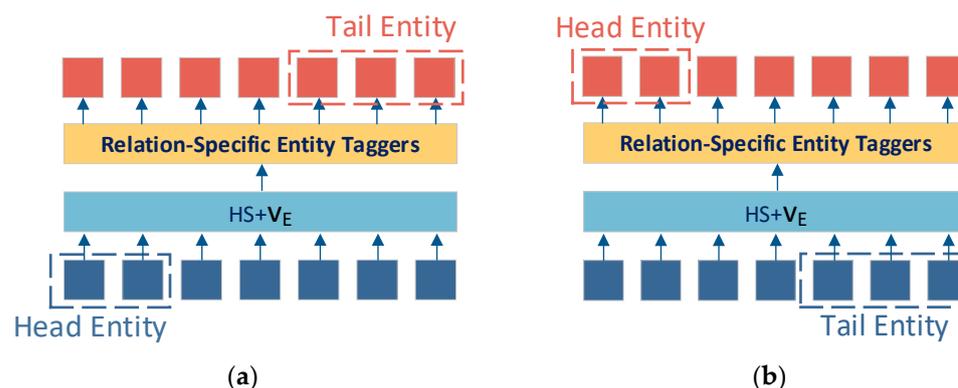


Figure 4. (a) Schematic diagram of the head entity recognition module. (b) Schematic diagram of the tail entity recognition module.

The DERP framework is implemented using TensorFlow. In the BERT encoder section, the framework is implemented on English datasets using the `cased_L-12_H-768_A-12` model and on Chinese datasets using the RoBERTa model. Dropout is applied to word embeddings and hidden states with a rate of 0.1. Network weights are optimized with Adam. The learning rate is set as 1×10^{-5} . The max length of the input sentence is set to 100. The batch size is set as 6. We use 100 epochs and choose the model with the best performance on the validation set to output results on the test set.

In our experimental procedures, for the sake of maintaining consistency with prior research, an extracted triple is deemed accurate if the head entity, the relation, and the tail entity are each validated as correct. The study reports standard metrics, including micro-precision (Prec.), recall (Rec.), and F1 score (f1), in line with the established baselines.

5.2. Baseline

To evaluate the performance of the DERP Framework, it is compared with ten baseline models: NovelTagging [3], CopyRE [31], GraphRel [32], ETL-Span [17], CopyMTL [33], CasRel [4], TPLinker [34], RSAN [13], CGT [35], and RIFRE [36].

Unless otherwise noted, the results of these baseline models were taken from the original papers.

5.3. Results

Table 1 shows the results of our model relative to other baselines extracted from entities and relations on both datasets. On the WebNLG dataset, DERP outperformed all baselines in both recall and F1 score, and on the NYT dataset, DERP achieved the second highest F1 score. These results directly validate the utility of the proposed DERP framework.

Table 1. Precision (%), recall (%) and F1 score (%) of the compared models on the NYT and WebNLG databases. * marks results quoted directly from the original papers.

Model	NYT			WebNLG		
	Prec.	Rec.	f1	Prec.	Rec.	f1
NovelTagging* [3]	61.5	41.4	49.5	-	-	-
CopyRE* [31]	61.0	56.6	58.7	37.7	36.4	37.1
GraphRel* [32]	63.9	60.0	61.9	44.7	41.1	42.9
ETL-Span* [17]	53.8	65.1	59.0	84.3	82.9	83.1
CopyMTL* [33]	75.7	68.7	72.0	58.0	54.9	56.4
CasRel* [4]	89.7	89.5	89.6	93.4	90.1	91.8
TPLinker* [34]	91.3	92.5	91.9	91.8	92.0	91.9
RSAN* [13]	85.7	83.6	84.6	80.5	83.8	82.1
CGT* [35]	94.7	84.2	89.1	92.9	75.6	83.4
RIFRE* [36]	93.6	90.5	92.0	93.3	92.0	92.6
DERP	92.05	89.94	90.98	92.82	92.90	92.86
DERP_HeadEntity	91.12	90.47	90.80	92.10	92.18	92.28
DERP_TailEntity	92.03	72.49	81.10	93.42	86.70	90.35

Table 2 shows the experimental results of DERP on the DuIE2.0 and CMeIE-V2 datasets, which shows an improvement over CasRel in terms of F1 score results. The F1 score of DERP_HeadEntity is also higher than CasRel when experiments are conducted using DERP_HeadEntity.

Table 2. Precision (%), recall (%) and F1 score (%) of the compared models on the DuIE2.0 and CMeIE-V2 databases. * marks results of reproduced experiments.

Model	DuIE2.0			CMeIE-V2		
	Prec.	Rec.	f1	Prec.	Rec.	f1
CasRel*	69.56	65.54	67.49	47.56	42.56	44.91
DERP	71.06	65.35	68.09	47.51	46.11	46.80
DERP_HeadEntity	70.38	65.80	68.01	47.27	45.15	46.19
DERP_TailEntity	73.97	53.50	62.09	49.10	43.01	45.85

We conducted experiments on CasRel under the same experimental conditions as the DERP framework. On the NYT dataset, CasRel* scored precision 88.87%, recall 90.34%, and F1 score 89.60%; on the WebNLG dataset, CasRel* scored precision 91.92%, recall 91.39%, and F1 score 91.65%. Compared with the replicated CasRel* framework, DERP has 1.38 percent improvement in F1 score on the NYT dataset, 1.21 percent improvement in F1 score on the WebNLG dataset, 0.6 percent improvement in F1 score on the DuIE2.0 dataset, and 1.98 percent improvement in F1 score on the CMeIE-V2 dataset. On the four datasets of NYT, WebNLG, DuIE2.0, and CMeIE-V2, in the experiments using head entity recognition and tail entity recognition alone for triple prediction, DERP_HeadEntity has higher precision, recall and F1 score than the original CasRel model in the experiments. In the DERP tail entity experiment, the features of the tail entity are not as easy to recognize as the features of the head entity, resulting in weaker F1 experimental results than DERP_HeadEntity on the four datasets.

Table 1 also presents that in the experiments on the two English datasets, with the existing models compared, a significant gap in processing performance between the models

is found, which proves that DERP performs better in dealing with redundant entities and overlapping triples. In the comparison experiments on four datasets, NYT, WebNLG, DuIE2.0, and CMeIE-V2, it is demonstrated that dividing entity recognition into head entity recognition and tail entity recognition, as in the DERP framework, can effectively improve the accuracy of entity recognition, and can produce more accurate results in relation extraction and triple prediction.

6. Conclusions

In this study, a double-headed entities and relations prediction framework for joint triple extraction is proposed. The entity recognition part is decomposed into head entity recognition and tail entity recognition. Specifically, relation prediction and tail entity recognition are executed for the head entities, and in parallel, relation prediction and head entity recognition are performed for the tail entities. In addition, a triple prediction module is designed to solve the entity overlapping problem in previous joint triple extractions. We systematically conducted experiments across four distinct datasets and compared them with ten baseline models. By proceeding with joint triple extraction, a good foundation is constructed for subsequent natural language processing or knowledge graph construction efforts. The results of these rigorous investigations substantiate that the conceptual framework introduced in this paper exhibits certain improvements when juxtaposed with prior models.

In the DERP framework, we have only improved the case of missing triple extraction, and in future work, we will conduct research on the case of error in triple extraction. We will also conduct research on Chinese text triple extraction to study the special characteristics of Chinese text triple extraction and improve the accuracy and effectiveness of Chinese text triple extraction.

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

Funding: This research was funded by cooperative projects between universities in Chongqing and the Chinese Academy of Sciences, grant number HZ2021015; the Chongqing Technology Innovation and Application Development Special Project, grant number cstc2019jcxmbdxX0016; the General Project of the Chongqing Municipal Science and Technology Commission, grant number cstc2021jcyjmsxm3332; the Sichuan Science and Technology Program 2023JDRC0033; the Young Project of Science and Technology Research Program of the Chongqing Education Commission of China, number KJQN202001513 and number KJQN202101501; the Luzhou Science and Technology Program 2021-JYJ-92; the Chongqing Postgraduate Scientific Research Innovation Project, grant number CYS23752; and the Chongqing University of Science and Technology Master and Doctoral Student Innovation Project, grant number YKJCX2120811.

Data Availability Statement: Not applicable.

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

References

1. Jiang, Z.; Chi, C.; Zhan, Y. Research on Medical Question Answering System Based on Knowledge Graph. *IEEE Access* **2021**, *9*, 21094–21101. [[CrossRef](#)]
2. Ma, L.; Ren, H.; Zhang, X. Effective Cascade Dual-Decoder Model for Joint Entity and Relation Extraction. *arXiv* **2021**. [[CrossRef](#)]
3. 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, Vancouver, BC, Canada, 30 July–4 August 2017; pp. 1227–1236.
4. 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, Online, 5–10 July 2020; pp. 1476–1488.
5. Li, L.; Zhang, J.; Jin, L.; Guo, R.; Huang, D. A Distributed Meta-Learning System for Chinese Entity Relation Extraction. *Neurocomputing* **2015**, *149*, 1135–1142. [[CrossRef](#)]

6. Li, H.; Xu, T.; Zhou, J. Mathematical Subject Information Entity Recognition Method Based on BiLSTM-CRF. In *Machine Learning for Cyber Security, Proceedings of the Third International Conference, MLACS 2020, Guangzhou, China, 8–10 October 2020*; Proceedings, Part III 3; Springer International Publishing: Cham, Switzerland, 2020; pp. 259–268.
7. Luo, L.; Yang, Z.; Yang, P.; Zhang, Y.; Wang, L.; Lin, H.; Wang, J. An Attention-Based BiLSTM-CRF Approach to Document-Level Chemical Named Entity Recognition. *Bioinformatics* **2018**, *34*, 1381–1388. [[CrossRef](#)] [[PubMed](#)]
8. Li, X.; Zhang, H.; Zhou, X.-H. Chinese Clinical Named Entity Recognition with Variant Neural Structures Based on BERT Methods. *J. Biomed. Inform.* **2020**, *107*, 103422. [[CrossRef](#)] [[PubMed](#)]
9. Ren, Z. Joint Entity and Relation Extraction Based on Specific-Relation Attention Mechanism and Global Features. In Proceedings of the Second International Conference on Electronic Information Technology (EIT 2023), Wuhan, China, 31 March–2 April 2023; Volume 12719, pp. 685–691.
10. Zeng, X.; He, S.; Zeng, D.; Liu, K.; Liu, S.; Zhao, J. Learning the Extraction Order of Multiple Relational Facts in a Sentence with Reinforcement Learning. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 367–377.
11. Han, X.; Wang, L. A Novel Document-Level Relation Extraction Method Based on BERT and Entity Information. *IEEE Access* **2020**, *8*, 96912–96919. [[CrossRef](#)]
12. Chen, Y.; Yang, W.; Wang, K.; Qin, Y.; Huang, R.; Zheng, Q. A Neuralized Feature Engineering Method for Entity Relation Extraction. *Neural Netw.* **2021**, *141*, 249–260. [[CrossRef](#)] [[PubMed](#)]
13. 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-20), Yokohama, Japan, 7–15 January 2021; ISBN 978-0-9992411-6-5.
14. Wan, Q.; Wei, L.; Zhao, S.; Liu, J. A Span-Based Multi-Modal Attention Network for Joint Entity-Relation Extraction. *Knowl.-Based Syst.* **2023**, *262*, 110228. [[CrossRef](#)]
15. Huang, W.; Cheng, X.; Wang, T.; Chu, W. BERT-Based Multi-Head Selection for Joint Entity-Relation Extraction. In *Natural Language Processing and Chinese Computing, Proceedings of the 8th CCF International Conference, NLPCC 2019, Dunhuang, China, 9–14 October 2019*; Proceedings, Part II; Springer: Cham, Switzerland, 2019; pp. 713–723.
16. Liu, J.; Chen, S.; Wang, B.; Zhang, J.; Li, N.; Xu, T. Attention as Relation: Learning Supervised Multi-Head Self-Attention for Relation Extraction. In Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20), Yokohama, Japan, 7–15 January 2021; pp. 3787–3793.
17. Yu, B.; Zhang, Z.; Shu, X.; Wang, Y.; Liu, T.; Wang, B.; Li, S. Joint Extraction of Entities and Relations Based on a Novel Decomposition Strategy. In Proceedings of the 24th European Conference on Artificial Intelligence—ECAI 2020, Santiago de Compostela, Spain, 29 August–8 September 2020.
18. Guo, Y.; Liu, Z.; Huang, C.; Liu, J.; Jing, W.; Wang, Z.; Wang, Y. CyberRel: Joint Entity and Relation Extraction for Cybersecurity Concepts. In *Information and Communications Security*; Gao, D., Li, Q., Guan, X., Liao, X., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2021; Volume 12918, pp. 447–463. ISBN 978-3-030-86889-5.
19. Lv, C.; Pan, D.; Li, Y.; Li, J.; Wang, Z. A Novel Chinese Entity Relationship Extraction Method Based on the Bidirectional Maximum Entropy Markov Model. *Complexity* **2021**, *2021*, e6610965. [[CrossRef](#)]
20. Zheng, H.; Wen, R.; Chen, X.; Yang, Y.; Zhang, Y.; Zhang, Z.; Zhang, N.; Qin, B.; Xu, M.; Zheng, Y. 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, Online, 1–6 August 2021; pp. 6225–6235.
21. Li, X.; Li, Y.; Yang, J.; Liu, H.; Hu, P. A Relation Aware Embedding Mechanism for Relation Extraction. *Appl. Intell.* **2022**, *52*, 10022–10031. [[CrossRef](#)]
22. Huang, H.; Shang, Y.-M.; Sun, X.; Wei, W.; Mao, X. Three Birds, One Stone: A Novel Translation Based Framework for Joint Entity and Relation Extraction. *Knowl.-Based Syst.* **2022**, *236*, 107677. [[CrossRef](#)]
23. Liu, C.; Zhang, X.; Xu, Y.; Xiang, B.; Gan, L.; Shu, Y. Knowledge Graph for Maritime Pollution Regulations Based on Deep Learning Methods. *Ocean Coast. Manag.* **2023**, *242*, 106679. [[CrossRef](#)]
24. Zhuang, C.; Zhang, N.; Jin, X.; Li, Z.; Deng, S.; Chen, H. Joint Extraction of Triple Knowledge Based on Relation Priority. In Proceedings of the 2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), Exeter, UK, 17–19 December 2020; pp. 562–569.
25. Riedel, S.; Yao, L.; McCallum, A. Modeling Relations and Their Mentions without Labeled Text. In *Machine Learning and Knowledge Discovery in Databases*; Balcázar, J.L., Bonchi, F., Gionis, A., Sebag, M., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2010; Volume 6323, pp. 148–163. ISBN 978-3-642-15938-1.
26. Gardent, C.; Shimorina, A.; Narayan, S.; Perez-Beltrachini, L. 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.
27. Huang Xun, Y.H. A Review of Relation Extraction. *Data Anal. Knowl. Discov.* **2013**, *29*, 30–39. [[CrossRef](#)]

28. Li, S.; He, W.; Shi, Y.; Jiang, W.; Liang, H.; Jiang, Y.; Zhang, Y.; Lyu, Y.; Zhu, Y. DuIE: A Large-Scale Chinese Dataset for Information Extraction. In *Natural Language Processing and Chinese Computing*; Tang, J., Kan, M.-Y., Zhao, D., Li, S., Zan, H., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2019; Volume 11839, pp. 791–800, ISBN 978-3-030-32235-9.
29. Zhang, N.; Chen, M.; Bi, Z.; Liang, X.; Li, L.; Shang, X.; Yin, K.; Tan, C.; Xu, J.; Huang, F.; et al. CBLUE: A Chinese Biomedical Language Understanding Evaluation Benchmark. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Dublin, Ireland, 22–27 May 2022; pp. 7888–7915.
30. Cheng, D.; Song, H.; He, X.; Xu, B. Joint Entity and Relation Extraction for Long Text. In *Knowledge Science, Engineering and Management*; Qiu, H., Zhang, C., Fei, Z., Qiu, M., Kung, S.-Y., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2021; Volume 12816, pp. 152–162. ISBN 978-3-030-82146-3.
31. Zeng, X.; Zeng, D.; He, S.; Liu, K.; Zhao, J. Extracting Relational Facts by an End-to-End Neural Model with Copy Mechanism. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia, 15–20 July 2018; pp. 506–514.
32. Fu, T.-J.; Li, P.-H.; Ma, W.-Y. GraphRel: Modeling Text as Relational Graphs for Joint Entity and Relation Extraction. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 1409–1418.
33. Zeng, D.; Zhang, H.; Liu, Q. CopyMTL: Copy Mechanism for Joint Extraction of Entities and Relations with Multi-Task Learning. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 9507–9514. [[CrossRef](#)]
34. Wang, Y.; Yu, B.; Zhang, Y.; Liu, T.; Zhu, H.; Sun, L. TPLinker: Single-Stage Joint Extraction of Entities and Relations Through Token Pair Linking. In Proceedings of the 28th International Conference on Computational Linguistics, International Committee on Computational Linguistics, 2020, Online, 8–13 December 2020; pp. 1572–1582.
35. Ye, H.; Zhang, N.; Deng, S.; Chen, M.; Tan, C.; Huang, F.; Chen, H. Contrastive Triple Extraction with Generative Transformer. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 14257–14265. [[CrossRef](#)]
36. Zhao, K.; Xu, H.; Cheng, Y.; Li, X.; Gao, K. Representation Iterative Fusion Based on Heterogeneous Graph Neural Network for Joint Entity and Relation Extraction. *Knowl.-Based Syst.* **2021**, *219*, 106888. [[CrossRef](#)]

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.