

StereoRel: Relational Triple Extraction from a Stereoscopic Perspective

Xuetao Tian¹, Liping Jing^{1,*}, Lu He², Feng Liu²

¹ Technology Beijing Key Lab of Traffic Data Analysis and Mining

² Engineering Research Center of Network Management Technology for High Speed Railway

School of Computer and Information Technology, Beijing Jiaotong University,

Beijing, China, 100044

{xttian, lpjing, helu19, fliu}@bjtu.edu.cn

Abstract

Relational triple extraction is critical to understanding massive text corpora and constructing large-scale knowledge graph, which has attracted increasing research interest. However, existing studies still face some challenging issues, including information loss, error propagation and ignoring the interaction between entity and relation. To intuitively explore the above issues and address them, in this paper, we provide a revealing insight into relational triple extraction from a stereoscopic perspective, which rationalizes the occurrence of these issues and exposes the shortcomings of existing methods. Further, a novel model is proposed for relational triple extraction, which maps relational triples to a three-dimension (3-D) space and leverages three decoders to extract them, aimed at simultaneously handling the above issues. Extensive experiments are conducted on five public datasets, demonstrating that the proposed model outperforms the recent advanced baselines.

1 Introduction

Relational triple is a common structural representation of semantic facts. A triple is always in form of (subject, relation, object), where subject and object are two entities connected by a type of predefined semantic relation. Relational triple extraction from unstructured texts is critical to understanding massive text corpora and constructing large-scale knowledge graph (Ren et al., 2017; Wei et al., 2020), which is widely concerned in recent years.

Early researches (Zhou et al., 2005; Chan and Roth, 2011; Zhang et al., 2017) first recognize entities and predict the relations for each entity pair. Such approaches suffer from error propagation problem and thus recent researches (Zheng et al., 2017; Zeng et al., 2018; Fu et al., 2019;

Nayak and Ng, 2020; Wei et al., 2020; Liu et al., 2020) try to build a jointly-decoding schema for entities and relations. However, relational triple extraction still faces the following challenging issues:

- Information loss (**I-IL**). Information loss includes entity incompleteness (Zeng et al., 2020) and entity overlapping (Zeng et al., 2018; Wei et al., 2020). Entity incompleteness (I-IL-EI) refers to that only head or tail token rather than completed entity is recognized, while entity overlapping (I-IL-EO) is that one entity belonging to multiple triples cannot be marked.
- Error propagation (**I-EP**). Error propagation comes from the prediction process with strict order. For examples, pipeline models (Zhang et al., 2017; Takanobu et al., 2019) recognize entities first and predict relations based on each specific entity pair. Generative models (Zeng et al., 2018, 2019) extract subject, object and relation with a predetermined order.
- Ignoring the interaction between entity and relation (**I-II**). Subjects (or objects) in different predefined relations should have different recognition patterns, which are not modelled when ignoring the interaction between entity and relation.

To intuitively explore the above issues and address them, from a stereoscopic perspective, we map the relational triples of a text to a three-dimensional (3-D) space, which is like a cube as Figure 1. The relational triples are actually some small cubes in the whole cube. Existing researches are actually to model the cube from different perspectives and further extract the triples. Based on the representation of triples in 3-D space, three operations (i.e. slice, projection and shrinkage)

*Corresponding author: Liping Jing.

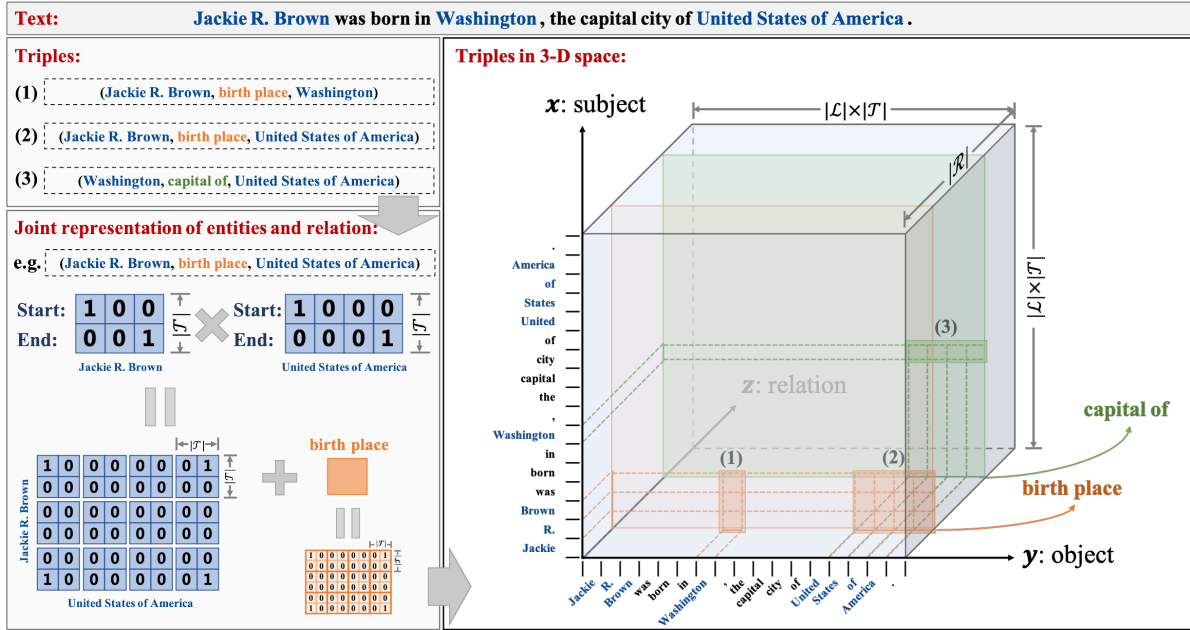


Figure 1: The representation of triples in 3-D space, where a text corresponds to a cube with size being $(|\mathcal{L}| \times |\mathcal{T}|) \times (|\mathcal{L}| \times |\mathcal{T}|) \times |\mathcal{R}|$, while each triple is mapped to a small cube with size being $(m \times |\mathcal{T}|) \times (n \times |\mathcal{T}|) \times 1$.

are defined as Figure 2, to understand why existing methods suffer from the above issues. Furthermore, we propose a novel model for relational triple extraction, which can simultaneously handle the above issues, named **StereoRel**. More precisely, the cube is modelled from three perspectives, including (x, z) -plane projection, (y, z) -plane projection and z -slices, which indicates the subjects, objects and their correspondences for each predefined relation. Correspondingly, the proposed method leverages three decoders to extract relational triples in a unified model.

This work has the following main contributions:

- We provide a revealing insight into relational triple extraction from a stereoscopic perspective, where the occurrence of several challenging issues and shortcomings of existing methods are rationalized.
- We propose a novel **StereoRel** model for relational triple extraction, which can simultaneously reduce information loss, avoid error propagation and not ignore the interaction between entity and relation.
- Extensive experiments are conducted on five public datasets, demonstrating that the proposed model outperforms the recent advanced baselines.

2 Relational Triple Extraction from 3-D Perspective

In form of (subject, relation, object), triples can naturally be mapped to a three-dimensional (3-D) space, which is elaborated in this section. Meanwhile, we define three operations (i.e. slice, projection and shrinkage) in 3-D space, to make it easy to understand the strengths and shortcomings of previous researches.

2.1 Triple Representation in 3-D Space

Given a text \mathcal{L} with length being $|\mathcal{L}|$ and a predefined relation set \mathcal{R} having $|\mathcal{R}|$ relations, \mathcal{L} may have several triples, that is, $p([s, r, o]|\mathcal{L})$ where $[\cdot]$ represents a collection. Each triple consists of a subject (s), an object (o) and one relation (r) belonging to \mathcal{R} . Subject is one entity, that is, n-gram in \mathcal{L} and so does object. To model $p([s]|\mathcal{L})$ or $p([o]|\mathcal{L})$, the common strategy is to leverage sequence tagging on \mathcal{L} , which has some existing strategies, such as BMES tagging (Zhang and Yang, 2018; Li et al., 2020) and start-and-end binary tagging (Wei et al., 2020; Sui et al., 2020). Anyway, there is a tag set \mathcal{T} , and thus $p([s]|\mathcal{L})$ or $p([o]|\mathcal{L})$ can be represented by a vector with length being $|\mathcal{L}| \times |\mathcal{T}|$. Meanwhile, due to $r \in \mathcal{R}$, modeling $p([r]|\mathcal{L})$ can be taken as a classification task, which requires a vector with length being $|\mathcal{R}|$ to represent. Therefore, when modeling $p([s, r, o]|\mathcal{L})$ by consid-

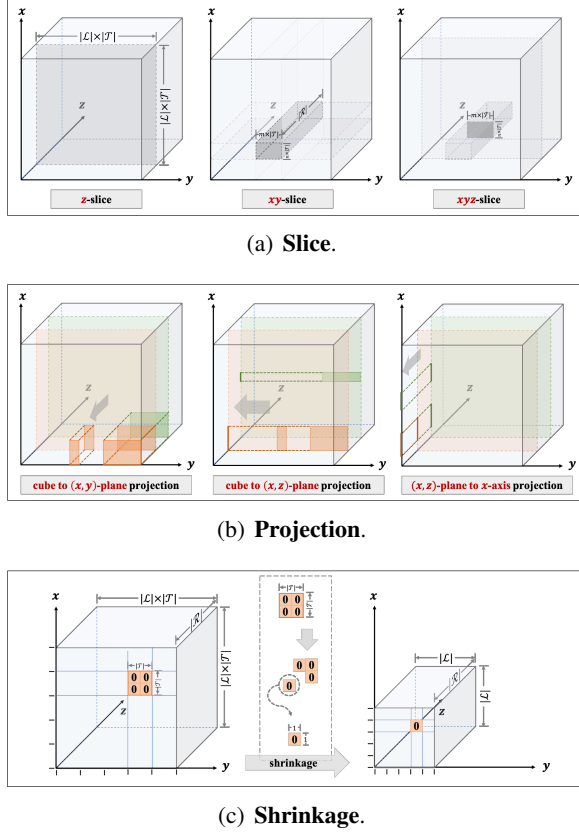


Figure 2: Our defined operations for relational triple extraction in 3-D space.

ering all possible connections, it should be equivalent to a cube with size being $(|\mathcal{L}| \times |\mathcal{T}|)^2 \times |\mathcal{R}|$ in a 3-D space.

As shown in Figure 1, the line segments of the cube mapping on x -axis, y -axis and z -axis are respectively regarded as the representations of subjects, objects and relations, that is, $p([s]|\mathcal{L})$, $p([o]|\mathcal{L})$ and $p([r]|\mathcal{L})$. Similarly, the rectangles of the cube mapping on (x, y) -plane, (x, z) -plane and (y, z) -plane are respectively regarded as $p([s, o]|\mathcal{L})$, $p([s, r]|\mathcal{L})$ and $p([o, r]|\mathcal{L})$. Further, each triple is mapped to a small cube in the space. Based on the stereoscopic representation of relational triples, we define the following operations.

Slice, denoted as $sl_i(\cdot)$. As shown in Figure 2(a), when some elements (i.e. subject, object or relation) are specified, the representation space will be reduced. The operation is like slicing the cube. For instances, a specific relation corresponds to a z -slice with size being $(|\mathcal{L}| \times |\mathcal{T}|)^2 \times 1$. Both subject and object being specified leads to an xy -slice with size being $(m \times |\mathcal{T}|) \times (n \times |\mathcal{T}|) \times |\mathcal{R}|$, which can be seen as the intersection of an x -slice and a y -slice. If subject, object or relation are

Table 1: The correspondence between the issues of relational triple extraction and the operations in 3-D space.

operation	I-IL		I-EP	I-II
	I-IL-EI	I-IL-EO		
$sl_i(\cdot)$			+	
$pro_{x/y}(\cdot)$				+
$pro_{xy}(\cdot)$		+		+
$shr(\cdot)_{only}$	+			

all specified, there is an xyz -slice with size being $(m \times |\mathcal{T}|) \times (n \times |\mathcal{T}|) \times 1$, that is, a triple.

Projection, denoted as $pro(\cdot)$. As depicted in Figure 2(b), two types of projection are defined, *cube-to-plane* and *plane-to-axis*. The former seems to look at the whole cube from a certain plane. For example, in the projection from cube to (x, y) -plane, two triples with the same subject and object are indistinguishable. Similarly, in the projection from cube to (x, z) -plane, there is only subject and relation information but no object information. The later seems to look at a plane from a certain axis, such as (x, z) -plane to x -axis projection, where the subjects in different z -slices may have the same representation on x -axis. Hereafter, for easy reading, (x, z) -plane to x -axis and (y, z) -plane to y -axis projections are denoted as $pro_x(\cdot)$ and $pro_y(\cdot)$ respectively. The projection from cube to (x, y) -plane is denoted as $pro_{xy}(\cdot)$. The rest ones are similar.

Shrinkage, denoted as $shr(\cdot)$. In the cube representation, each token pair is represented by an xy -slice with size being $|\mathcal{T}| \times |\mathcal{T}| \times |\mathcal{R}|$. Such an xy -slice can reflect all possible entity-tagging combinations of a token pair. As described in Figure 2(c), a shrinkage over a cube only represent whether the token pair satisfies one specific entity-tagging combination, such as (start, start). Thus, the size of a shrinkage is $|\mathcal{L}| \times |\mathcal{L}| \times |\mathcal{R}|$.

2.2 Analysis of Previous Researches

As aforementioned, relational triple extraction faces three challenging issues: information loss (**I-IL-EI** or **I-IL-EO**), error propagation (**I-EP**) and ignoring the interaction between entity and relation (**I-II**). It is clear to match them to the three operations in 3-D space as Table 1. $sl_i(\cdot)$ corresponds to the prediction process with strict order, and thus leads to error propagation. Without considering the nested entities in a text, $pro_{xz/yz/z}(\cdot)$ do not result in any problems, while $pro_{x/y/xy}(\cdot)$ is the

Table 2: The analysis of previous researches from the stereoscopic perspective. For the mentioned issues, \checkmark indicates that a model can handle a certain issue and \times is the opposite.

Models	Perspective in 3-D space (for modeling $p([s, r, o] \mathcal{L})$)	I-IL		I-EP	I-II
		I-IL-EI	I-IL-EO		
Pipeline	$p([s] \cup [o] \mathcal{L}) + \sum_{s_i}^{[s] \cup [o]} \sum_{o_j \neq s_i}^{[s] \cup [o]} p([s_i, r, o_j] \mathcal{L}, \mathbf{sli}_{xy}(s_i, o_j))$, where $p([s] \cup [o] \mathcal{L}) = p(\mathbf{pro}_x(\mathbf{pro}_{xz}([s, r, o])) \cup \mathbf{pro}_y(\mathbf{pro}_{yz}([s, r, o])) \mathcal{L})$.	\checkmark	\checkmark	\times	\times
MHS	$p([s] \cup [o] \mathcal{L}) + \sum_{s_i}^{[s] \cup [o]} p([s_i, r, o] \mathcal{L}, \mathbf{sli}_x(s_i))$, where $p([s] \cup [o] \mathcal{L}) = p(\mathbf{pro}_x(\mathbf{pro}_{xz}([s, r, o])) \cup \mathbf{pro}_y(\mathbf{pro}_{yz}([s, r, o])) \mathcal{L})$.	\checkmark	\checkmark	\times	\times
NovelTagging	$p(\mathbf{pro}_x(\mathbf{pro}_{xy}([s, r, o])) \cup \mathbf{pro}_y(\mathbf{pro}_{xy}([s, r, o])) \mathcal{L})$.	\checkmark	\times	\checkmark	\times
PA-Tagging	$p(\mathbf{pro}_{xy}([s, r, o] \mathcal{L}))$.	\checkmark	\times	\checkmark	\times
CopyRE CopyRRL	$p(\mathbf{shr}([s, r, o] \mathcal{L})) = \prod_i p(\mathbf{sli}_{xyz}(s_i, r_i, o_i) \mathcal{L}, [\mathbf{sli}_{xyz}(s_j, r_j, o_j)]_{j < i})$, where $p(\mathbf{sli}_{xyz}(s_i, r_i, o_i)) = p(\mathbf{sli}_z(r_i)) \cdot p(\mathbf{sli}_{xz}(s_i, r_i) \mathbf{sli}_z(r_i)) \cdot p(\mathbf{sli}_{xy}(s_i, r_i, o_i) \mathbf{sli}_{xz}(s_i, r_i))$.	\times	\checkmark	\times	\checkmark
GraphRel	$p(\mathbf{pro}_{xy}(\mathbf{shr}([s, r, o] \mathcal{L})))$.	\times	\times	\checkmark	\times
CopyMTL	$p(\mathbf{shr}([s, r, o] \mathcal{L})) + p([s] \cup [o] \mathcal{L})$, where $p(\mathbf{shr}([s, r, o] \mathcal{L}))$ is the same as CopyRE, and $p([s] \cup [o] \mathcal{L}) = p(\mathbf{pro}_x(\mathbf{pro}_{xz}([s, r, o])) \cup \mathbf{pro}_y(\mathbf{pro}_{yz}([s, r, o])) \mathcal{L})$.	\checkmark	\checkmark	\times	\times
WDec PNDec	$\prod_i p(\mathbf{sli}_{xyz}(s_i, r_i, o_i) \mathcal{L}, [\mathbf{sli}_{xyz}(s_j, r_j, o_j)]_{j < i})$, where $p(\mathbf{sli}_{xyz}(s_i, r_i, o_i)) = p(\mathbf{sli}_x(s_i)) \cdot p(\mathbf{sli}_{xy}(s_i, o_i) \mathbf{sli}_x(s_i)) \cdot p(\mathbf{sli}_{yz}(s_i, r_i, o_i) \mathbf{sli}_{xy}(s_i, o_i))$.	\checkmark	\checkmark	\times	\checkmark
CasRel	$p([s] \mathcal{L}) + \sum_{s_i}^{[s]} \sum_j^{[R]} p([s_i, r_j, o] \mathcal{L}, \mathbf{sli}_x(s_i))$, where $p([s] \mathcal{L}) = p(\mathbf{pro}_x(\mathbf{pro}_{xz}([s, r, o])) \mathcal{L})$.	\checkmark	\checkmark	\times	\times
Att-as-Rel	$p(\mathbf{shr}([s, r, o] \mathcal{L})) + p([s] \cup [o] \mathcal{L})$, where $p(\mathbf{shr}([s, r, o] \mathcal{L})) = \sum_i^{[R]} p(\mathbf{shr}([s, r_i, o] \mathcal{L}))$, and $p([s] \cup [o] \mathcal{L}) = p(\mathbf{pro}_x(\mathbf{pro}_{xz}([s, r, o])) \cup \mathbf{pro}_y(\mathbf{pro}_{yz}([s, r, o])) \mathcal{L})$.	\checkmark	\checkmark	\checkmark	\times
TPLinker	$2 \times p(\mathbf{shr}([s, r, o] \mathcal{L})) + p([s] \cup [o] \mathcal{L})$, where one $\mathbf{shr}([s, r, o] \mathcal{L})$ is (start, start) $\mathbf{shr}(\cdot)$, the other is (end, end) $\mathbf{shr}(\cdot)$, and $p([s] \cup [o] \mathcal{L}) = p(\mathbf{pro}_x(\mathbf{pro}_{xz}([s, r, o])) \cup \mathbf{pro}_y(\mathbf{pro}_{yz}([s, r, o])) \mathcal{L})$.	\checkmark	\checkmark	\checkmark	\times
Ours	$\sum_i^{[R]} [p([s, r_i] \mathcal{L}) + p([r_i, o] \mathcal{L}) + p(\mathbf{shr}([s, r_i, o] \mathcal{L}))]$, where $\sum_i^{[R]} p([s, r_i] \mathcal{L}) = p(\mathbf{pro}_{xz}([s, r, o] \mathcal{L}))$, $\sum_i^{[R]} p([r_i, o] \mathcal{L}) = p(\mathbf{pro}_{yz}([s, r, o] \mathcal{L}))$, and $\mathbf{shr}([s, r, o] \mathcal{L})$ is one specific $\mathbf{shr}(\cdot)$.	\checkmark	\checkmark	\checkmark	\checkmark

opposite. Both $\mathbf{pro}_{x/y}(\cdot)$ and $\mathbf{pro}_{xy}(\cdot)$ lead to ignoring the interaction between entity and relation. Meanwhile, $\mathbf{pro}_{xy}(\cdot)$ makes the triples with overlapped entities indistinguishable. The cube can be disassembled into $|\mathcal{T}| \times |\mathcal{T}| \mathbf{shr}(\cdot)$. Modeling only one $\mathbf{shr}(\cdot)$ will cause entity incompleteness. To get deep insights on relational triple extraction, based on the correspondence between the operations and issues, we analyze previous researches as shown in Table 2.

Early researches (Zelenko et al., 2002; Zhou et al., 2005; Chan and Roth, 2011) adopt pipeline approaches, where the entities are recognized first and the relations for each entity pair are predicted. Arguing that such approaches neglect the inherent relevance between entity recognition and relation extraction, some solutions (Miwa and Bansal, 2016; Zhang et al., 2017; Takanobu et al., 2019) still extract entities and relations sequentially, but make two tasks share the same encoder. These methods model $p([s, r, o]|\mathcal{L})$ as $p([s] \cup [o]|\mathcal{L})$ and $p([s_i, r, o_j]|\mathcal{L}, \mathbf{sli}_{xy}(s_i, o_j))$,

where $p([s] \cup [o]|\mathcal{L}) = p(\mathbf{pro}_x(\mathbf{pro}_{xz}([s, r, o])) \cup \mathbf{pro}_y(\mathbf{pro}_{yz}([s, r, o]))|\mathcal{L})$. Therefore, pipeline paradigm suffers from **I-EP** and **I-II** issues. MHS (Bekoulis et al., 2018) is another two-stage method. The model recognizes entities firstly and extracts relational triples with a multi-head selection strategy on each subject, where $\mathbf{pro}_{x/y}(\cdot)$ and $\mathbf{sli}_x(\cdot)$ lead to **I-EP** and **I-II** issues respectively.

In the following researches on relational triple extraction, several methods with joint decoding schema are proposed. Specifically, NovelTagging (Zheng et al., 2017) and PA-Tagging (Dai et al., 2019) achieve joint decoding by designing a unified tagging scheme and convert relational triple extraction to an end-to-end sequence tagging problem. Such a tagging schema has to model $p(\mathbf{pro}_{xy}([s, r, o]|\mathcal{L}))$ and thus suffers from **I-IL-EO** and **I-II** issues. CopyRE (Zeng et al., 2018) and CopyRRL (Zeng et al., 2019) leverage sequence-to-sequence model with copy mechanism. GraphRel (Fu et al., 2019) introduces graph convolutional network jointly learn entities

and relations. Despite their initial success, the three methods only model $p(\mathit{shr}([s, r, o])|\mathcal{L})$ and thus suffer from **I-IL-EI** issue. Sequence generation models, CopyRE and CopyRRL, predict triples one by one and model $p(\mathit{sl}_{xyz}(s_i, r_i, o_i))$ via $p(\mathit{sl}_z(r_i)) \times p(\mathit{sl}_{xz}(s_i, r_i)|\mathit{sl}_z(r_i)) \times p(\mathit{sl}_{xyz}(s_i, r_i, o_i)|\mathit{sl}_{xz}(s_i, r_i))$, which leads to **I-EP** issue. GraphRel cannot avoid **I-IL-EO** and **I-II** issues due to its utilizing $\mathit{pro}_{xy}(\cdot)$.

Recently, to address **I-IL** issue, CopyMTL (Zeng et al., 2020) proposes a multi-task learning framework based on CopyRE, to simultaneously predict completed entities and capture relational triples. However, the model still does not solve **I-EP** issue. Meanwhile, entity recognition is implemented by modeling $p([s] \cup [o]|\mathcal{L})$ via a standalone module, which leads to **I-II** issue. Following sequence-to-sequence schema, WDec and PNDec (Nayak and Ng, 2020) design specific decoder block which can generate triples with completed entities. Such models ease **I-II** issue, but still suffers from **I-EP** issue since that it models $p(\mathit{sl}_{xyz}(s_i, r_i, o_i))$ via $p(\mathit{sl}_x(s_i)) \times p(\mathit{sl}_{xy}(s_i, o_i)|\mathit{sl}_x(s_i)) \times p(\mathit{sl}_{xyz}(s_i, r_i, o_i)|\mathit{sl}_{xy}(s_i, o_i))$. CasRel (Wei et al., 2020) regards relations as functions that map subjects to objects in a text. It is necessary to recognize subjects first and then objects, which leads to **I-EP** issue. To recognize subjects, $p([s]|\mathcal{L})$ is modelled via $p(\mathit{pro}_x(\mathit{pro}_{xz}([s, r, o])|\mathcal{L}))$, where $\mathit{pro}_x(\cdot)$ leads to **I-II** issue. Att-as-Rel (Liu et al., 2020) models the triples by multi-head attention, where completed entities are recognized by modeling $p([s] \cup [o]|\mathcal{L})$ separately and thus there is **I-II** issue. Similarly, TPLinker (Wang et al., 2020b) regards joint extraction as a token pair linking problem, where entity recognition is also modelled separately via $p([s] \cup [o]|\mathcal{L})$.

3 The Proposed StereoRel Model

To handle the above three issues simultaneously, we avoid to make the operations in Table 1 and try to model $p([s, r, o]|\mathcal{L})$ via $\sum_i^{|\mathcal{R}|} [p([s, r_i]|\mathcal{L}) + p([r_i, o]|\mathcal{L}) + p(\mathit{shr}([s, r_i, o])|\mathcal{L})]$. As depicted in Figure 3, the proposed StereoRel model first leverages BERT encoder to extract the text representation for the original text. Then, for each predefined relation, the text representation is transformed to its subject and object spaces. Based on them, three decoders are built to separately model $p(\mathit{pro}_{xz}([s, r, o])|\mathcal{L})$, $p(\mathit{pro}_{yz}([s, r, o])|\mathcal{L})$ and

$p(\mathit{shr}([s, r, o])|\mathcal{L})$. The first two will cause no issue and provide complete entities for the last $\mathit{shr}(\cdot)$ operation.

3.1 BERT Encoder

To sufficiently capture the textual information, the encoder is built by a pre-trained language model, BERT (Devlin et al., 2019). BERT encoder tokenizes a text \mathcal{L} using a predefined vocabulary and generates a corresponding sequence $\hat{\mathcal{L}}$ by concatenating a [CLS] token, the tokenized text and a [SEP] token. The detailed steps can be referred to (Devlin et al., 2019). BERT encoder will embed a text \mathcal{L} into a matrix $\mathbf{T} \in \mathbb{R}^{(|\mathcal{L}|+2) \times d_b}$, where d_b is the hidden size of BERT, and \mathbf{T}_j can be seen as the word embedding of j -th token, $j \in [0, |\mathcal{L}| + 1]$. After this, for each relation r_i , \mathbf{T} is transformed to a new text representation $\mathbf{T}_i \in \mathbb{R}^{(|\mathcal{L}|+2) \times d_r}$ by:

$$\mathbf{T}_i = \phi(\mathbf{T} \mathbf{W}_i + \mathbf{b}_i), \quad (1)$$

where $\{\mathbf{W}_i\}_{i=1}^{|\mathcal{R}|} \in \mathbb{R}^{d_b \times d_r}$, $\{\mathbf{b}_i\}_{i=1}^{|\mathcal{R}|} \in \mathbb{R}^{1 \times d_r}$ are trainable parameters and $\phi(\cdot)$ is predetermined activation function.

3.2 Subject Decoder

Subject decoder is to model $\sum_i^{|\mathcal{R}|} p([s, r_i]|\mathcal{L})$, that is, (x, z) -plane projection $p(\mathit{pro}_{xz}([s, r, o])|\mathcal{L})$, which recognizes the subjects for each predefined relation. For one specific relation r_i , we transform its text representation to $\mathbf{T}_i^{sub} \in \mathbb{R}^{(|\mathcal{L}|+2) \times d_e}$ in r_i 's subject space with d_e being the hidden size. The transformation is implemented by

$$\mathbf{T}_i^{sub} = \mathbf{T}_i^{sub-q} + \mathbf{T}_i^{sub-k} + \mathbf{T}_i^{sub-b}, \quad (2)$$

$$\mathbf{T}_i^{sub-q/k/b} = \phi(\mathbf{T}_i \mathbf{W}_i^{sub-q/k/b} + \mathbf{b}_i^{sub-q/k/b}), \quad (3)$$

where \mathbf{T}_i^{sub-q} , \mathbf{T}_i^{sub-k} , \mathbf{T}_i^{sub-b} are linear transformations on top of \mathbf{T}_i . \mathbf{T}_i^{sub-q} , \mathbf{T}_i^{sub-k} will be used by shrinkage decoder, while \mathbf{T}_i^{sub-b} only works for subject decoder. $\{\mathbf{W}_i^{sub-q}\}_{i=1}^{|\mathcal{R}|}$, $\{\mathbf{W}_i^{sub-k}\}_{i=1}^{|\mathcal{R}|}$, $\{\mathbf{W}_i^{sub-b}\}_{i=1}^{|\mathcal{R}|} \in \mathbb{R}^{d_r \times d_e}$, $\{\mathbf{b}_i^{sub-q}\}_{i=1}^{|\mathcal{R}|}$, $\{\mathbf{b}_i^{sub-k}\}_{i=1}^{|\mathcal{R}|}$, $\{\mathbf{b}_i^{sub-b}\}_{i=1}^{|\mathcal{R}|} \in \mathbb{R}^{1 \times d_e}$ are trainable parameters and $\phi(\cdot)$ is predetermined activation function. Based on \mathbf{T}_i^{sub} , all possible subjects in relation r_i 's subject space are recognized by a sequential conditional random field (CRF) (Lafferty et al., 2001) layer with

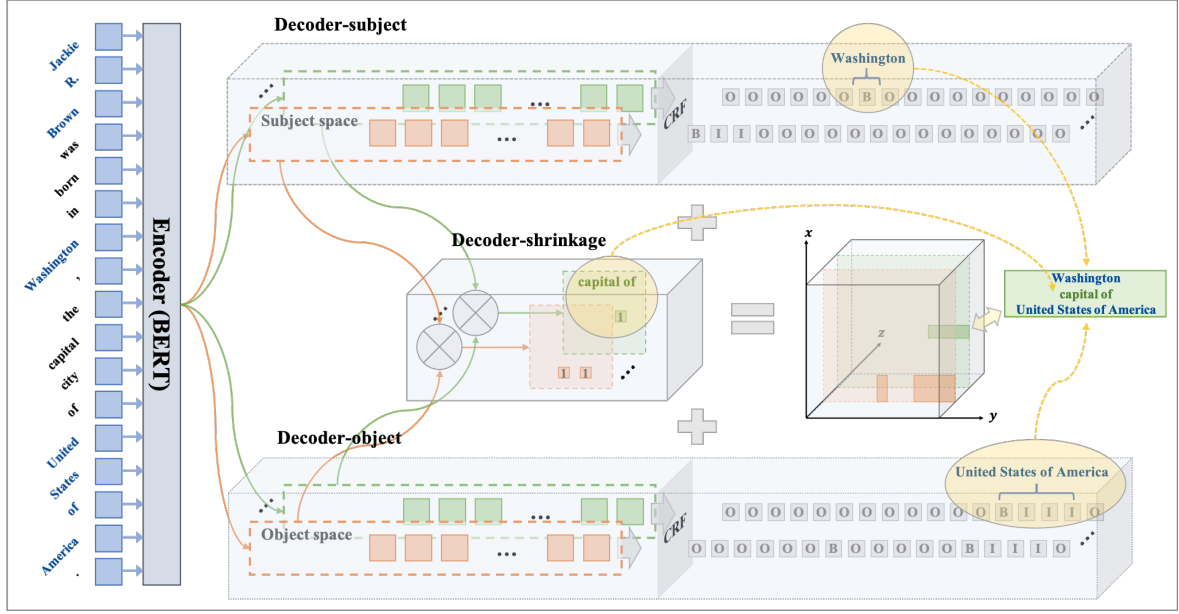


Figure 3: The proposed StereoRel model. There are three decoders, respectively extracting subjects, objects and their correspondences for each predefined relation.

[Begin, Inside, Outside] tagging schema, where the probability of the final label sequence, $\mathbf{y}_i^{sub} = [y_{i1}^{sub}, y_{i2}^{sub}, \dots, y_{i|\mathcal{L}|}^{sub}]$, is modeled as follows:

$$P(\mathbf{y}_i^{sub} | \mathcal{L}) = \frac{\prod_{j=1}^{|\mathcal{L}|} \phi_j(y_{i(j-1)}^{sub}, y_{ij}^{sub} | \mathcal{L})}{\sum_{\mathbf{y}' \in Y} \prod_{j=1}^{|\mathcal{L}|} \phi_j(y'_{j-1}, y'_j | \mathcal{L})}, \quad (4)$$

$$\phi_j(y, \hat{y} | \mathcal{L}) = \exp(\mathbf{T}_i^{sub} \mathbf{w}_{y, \hat{y}}^{crf_{sub}} + b_{y, \hat{y}}^{crf_{sub}}), \quad (5)$$

where Y denotes all possible label sequence of \mathcal{L} . $\mathbf{w}_{y, \hat{y}}^{crf_{sub}}$ and $b_{y, \hat{y}}^{crf_{sub}}$ are trainable parameters corresponding to the label pair (y, \hat{y}) .

3.3 Object Decoder

Object decoder is to model $\sum_i^{|\mathcal{R}|} p([r_i, o] | \mathcal{L})$, that is, (y, z) -plane projection $p(\mathbf{proj}_{yz}([s, r, o]) | \mathcal{L})$, which recognizes the objects for each predefined relation. Similar to subject decoder, the text representation in object space, $\mathbf{T}_i^{obj} \in \mathbb{R}^{|\mathcal{R}| \times (|\mathcal{L}|+2) \times d_e}$, is obtained in object decoder as:

$$\mathbf{T}_i^{obj} = \mathbf{T}_i^{obj-q} + \mathbf{T}_i^{obj-k} + \mathbf{T}_i^{obj-b}, \quad (6)$$

$$\mathbf{T}_i^{obj-q/k/b} = \phi(\mathbf{T}_i \mathbf{W}_i^{obj-q/k/b} + \mathbf{b}_i^{obj-q/k/b}), \quad (7)$$

where \mathbf{T}_i^{obj-q} , \mathbf{T}_i^{obj-k} , \mathbf{T}_i^{obj-b} are linear transformations on top of \mathbf{T}_i . $\{\mathbf{W}_i^{obj-q}\}_{i=1}^{|\mathcal{R}|}$, $\{\mathbf{W}_i^{obj-k}\}_{i=1}^{|\mathcal{R}|}$, $\{\mathbf{W}_i^{obj-b}\}_{i=1}^{|\mathcal{R}|} \in \mathbb{R}^{d_r \times d_e}$, $\{\mathbf{b}_i^{obj-q}\}_{i=1}^{|\mathcal{R}|}$, $\{\mathbf{b}_i^{obj-k}\}_{i=1}^{|\mathcal{R}|}$, $\{\mathbf{b}_i^{obj-b}\}_{i=1}^{|\mathcal{R}|} \in \mathbb{R}^{1 \times d_e}$ are trainable parameters. In like wise, objects of the i -th predefined relation are tagged as $\mathbf{y}_i^{obj} = [y_{i1}^{obj}, y_{i2}^{obj}, \dots, y_{i|\mathcal{L}|}^{obj}]$ via another CRF layer as:

$$P(\mathbf{y}_i^{obj} | \mathcal{L}) = \frac{\prod_{j=1}^{|\mathcal{L}|} \phi_j(y_{i(j-1)}^{obj}, y_{ij}^{obj} | \mathcal{L})}{\sum_{\mathbf{y}' \in Y} \prod_{j=1}^{|\mathcal{L}|} \phi_j(y'_{j-1}, y'_j | \mathcal{L})}, \quad (8)$$

$$\phi_j(y, \hat{y} | \mathcal{L}) = \exp(\mathbf{T}_i^{obj} \mathbf{w}_{y, \hat{y}}^{crf_{obj}} + b_{y, \hat{y}}^{crf_{obj}}), \quad (9)$$

where $\mathbf{w}_{y, \hat{y}}^{crf_{obj}}$ and $b_{y, \hat{y}}^{crf_{obj}}$ are trainable parameters.

3.4 Shrinkage Decoder

To extract the correspondences between subjects and objects, shrinkage decoder is leveraged to model $\sum_i^{|\mathcal{R}|} p(\mathbf{shr}([s, r_i, o]) | \mathcal{L})$, where each element of $\mathbf{shr}(\cdot)$ denotes whether the corresponding token pair is one specific position of a (subject, object) pair, such as (start, start) or (end, end). To model this, a pair-wise classification function f is established as:

$$p_{ijj'}^{shr} = f(\mathbf{T}_{ij}^{sub}, \mathbf{T}_{ij'}^{obj}), \quad (10)$$

indicating the probability that j -token and j' -token is the specific position of a (subject, object) pair, which satisfies the i -th predefined relation. We design the function as follows:

$$p_{ijj'}^{shr} = \xi(p_{ijj'}^{sub \rightarrow obj}, p_{ijj'}^{obj \rightarrow sub}), \quad (11)$$

$$p_{ijj'}^{sub \rightarrow obj} = \text{softmax}_j(\psi(\mathbf{T}_{ij}^{sub-q}, \mathbf{T}_{ij'}^{obj-k})), \quad (12)$$

$$p_{ijj'}^{obj \rightarrow sub} = \text{softmax}_{j'}(\psi(\mathbf{T}_{ij'}^{obj-q}, \mathbf{T}_{ij}^{sub-k})), \quad (13)$$

where $\psi(\cdot)$ is implemented by dot product or neural network to provide an initial probability. $p_{ijj'}^{sub \rightarrow obj}$ and $p_{ijj'}^{obj \rightarrow sub}$ respectively indicate the probability distributions for a subject searching for its objects and an object searching for its subjects, which are integrated via a predetermined function $\xi(\cdot)$, such as minimum, maximum and multiplication.

3.5 Learning and Inference

Subject and object decoders are learned by text-level log-likelihood loss, while shrinkage decoder is learned by token-level binary cross-entropy loss. Thus, the unified model is learned by a combined loss function $L_{total} = L_{sub} + L_{obj} + L_{shr}$, where

$$L_{sub} = - \sum_i^{|\mathcal{R}|} \log(P(\mathbf{y}_i^{sub} | \mathcal{L})), \quad (14)$$

$$L_{obj} = - \sum_i^{|\mathcal{R}|} \log(P(\mathbf{y}_i^{obj} | \mathcal{L})), \quad (15)$$

$$L_{shr} = - \sum_i^{|\mathcal{R}|} \sum_j^{|\mathcal{L}|} \sum_{j'}^{|\mathcal{L}|} \left[\hat{p}_{ijj'}^{shr} \log(p_{ijj'}^{shr}) + (1 - \hat{p}_{ijj'}^{shr}) \log(1 - p_{ijj'}^{shr}) \right]. \quad (16)$$

The relational triples can be inferred based on the three decoders. Concretely, for each predefined relation r_i , the subjects and objects can be obtained by \mathbf{y}_i^{sub} and \mathbf{y}_i^{obj} respectively. For the subject s_{ij} and object $o_{ij'}$ with (j -th token, j' -th token) satisfying the specific position, if $p_{ijj'}^{shr}$ is greater than a predetermined threshold δ , $(s_{ij}, r_i, o_{ij'})$ will be extracted as a relational triple.

Table 3: Statistics of datasets.

Dataset	Rel-num	Train	Valid	Test
NYT	24	56196	5000	5000
WebNLG	171	5019	500	703
NYT10	29	70339 (0.5%)		4006
NYT11	12	62648 (0.5%)		369
Wiki-KBP	13	79934	(10%)	289

4 Experiments

To evaluate the proposed StereoRel model, we conduct a performance comparison on five public datasets in this section.

4.1 Experimental Settings

Evaluation Metrics and Datasets. Generally, the performance on relational triple extraction is evaluated by *precision* (Pre.), *recall* (Rec.) and *F1-score* (F1), where a triple is regarded as correct if subject, relation and object are all matched. Notably, in previous works, there are two evaluation modes: *Partial Match* and *Exact Match*. The former holds that subject (or object) is correct as long as its head or tail is correct, while the latter requires it to be recognized completely. To properly compare our model with various baselines, benchmark datasets are selected for the two modes separately. Concretely, we utilize NYT (Riedel et al., 2010), WebNLG (Gardent et al., 2017), NYT10 (Takanobu et al., 2019) and NYT11 (Takanobu et al., 2019) datasets for *Partial Match*, while NYT (Riedel et al., 2010) and Wiki-KBP (Dai et al., 2019) datasets for *Exact Match*. The details are shown in Table 3. The splits of validation set are the same as previous researches.

Implementation Details. For making a fair comparison, we utilize the cased BERT-base¹ model in our experiments, which is the same as CasRel (Wei et al., 2020) and TPLinker (Wang et al., 2020b), and thus $d_b = 768$. Adam optimizer (Kingma and Ba, 2015) is utilized to train the proposed method with initial learning rate being $1e-5$. The hidden size d_r, d_e are set as 64, 32. The threshold δ is tuned for each relation and determined by the validation set. $\phi(\cdot)$ is set as *relu* activation function. $\psi(\cdot)$ is set as dot product. $\xi(\cdot)$ is set as the multiplication function.

¹https://storage.googleapis.com/bert_models/2018_10_18/cased_L-12_H-768_A-12.zip

Table 4: Performance comparison by *Partial Match* on NYT

Model	Pre.	Rec.	F1
NovelTagging	62.4	31.7	42.0
CopyRe-one	59.4	53.1	56.0
CopyRe-mul	61.0	56.6	58.7
GraphRel-1p	62.9	57.3	60.0
GraphRel-2p	63.9	60.0	61.9
CopyRRL	77.9	67.2	72.1
WDec	94.5	76.2	84.4
CasRel	89.7	89.5	89.6
TPLinker	91.3	92.5	91.9
StereoRel (ours)	92.0	92.3	92.2

Table 5: Performance comparison by *Partial Match* on WebNLG

Model	Pre.	Rec.	F1
NovelTagging	52.5	19.3	28.3
CopyRe-one	32.2	28.9	30.5
CopyRe-mul	37.7	36.4	37.1
GraphRel-1p	42.3	39.2	40.7
GraphRel-2p	44.7	41.1	42.9
CopyRRL	63.3	59.9	61.6
WDec	88.6	51.3	65.0
CopyMTL-one	57.8	60.1	58.9
CopyMTL-mul	58.0	54.9	56.4
Att-as-Rel	89.5	86.0	87.7
CasRel	93.4	90.1	91.8
TPLinker	91.8	92.0	91.9
StereoRel (ours)	91.6	92.6	92.1

4.2 Performance Comparison

We employ some recent advanced methods as baselines, mainly including the models analyzed in Table 2. Table 4, 5, 6 and 7 report the results of our method against the baselines for *Partial Match* evaluation mode, and Table 8 and 9 report the results for *Exact Match*. The models before CasRel do not employ BERT encoder and the rest does.

As aforementioned in Table 2, existing models do not handle three challenging issues simultaneously, while our proposed StereoRel model does. Among the baselines, Att-as-Rel is the first work to extract triples for each predefined relation with no **I-IL** and **I-EP** issues, and thus achieves a huge performance improvement compared with previous methods. Based on BERT encoder, the performance on relational triple extraction has been further improved by CasRel and TPLinker. Due

Table 6: Performance comparison by *Partial Match* on NYT10

Model	Pre.	Rec.	F1
NovelTagging	59.3	38.1	46.4
CopyRe-mul	56.9	45.2	50.4
HRL	71.4	58.6	64.4
WDec	84.6	62.1	71.6
PDec	81.5	63.9	71.6
CasRel	77.7	68.8	73.0
StereoRel (ours)	80.0	67.4	73.2

Table 7: Performance comparison by *Partial Match* on NYT11

Model	Pre.	Rec.	F1
NovelTagging	46.9	48.9	47.9
CopyRe-mul	34.7	53.4	42.1
HRL	53.8	53.8	53.8
CasRel	50.1	58.4	53.9
StereoRel (ours)	53.8	55.4	54.6

to no **I-EP** issue, TPLinker outperforms CasRel. However, TPLinker still suffers from **I-II** issue. Our proposed StereoRel model further considers it and achieves a better performance. From the results, comparing with the second best baseline, the performance improvement of the existing best baseline on the five datasets were 2.5%, 0.1%, 1.4%, 0.1% and 1.5% respectively, in terms of *F1-score*. Our model obtains performance gain about 0.3%, 0.2%, 0.2%, 0.7% and 0.6% in terms of the best baseline. It can be seen that the improvement is satisfied.

5 Discussions and Perspectives

For relational triple extraction, from the stereoscopic perspective, there are the following two aspects worthy of discussion. The first one is about learning strategy. Most of previous studies and ours employ binary cross-entropy loss to learn the models. However, since the label space of relational triple in 3-D space is huge, binary cross-entropy is available but not necessarily optimal. Meanwhile, cross-entropy is permutation-sensitive loss function (Sui et al., 2020), which is incompatible with generative models (Zeng et al., 2018, 2019, 2020) since it is necessary to predetermine extraction order of multiple triples. To this question, CGT (Ye et al., 2020) incorporates contrastive learning strategy and SPN (Sui et al., 2020) transforms relational

Table 8: Performance comparison by *Exact Match* on NYT

Model	Pre.	Rec.	F1
NovelTagging	32.8	30.6	31.7
HRL	78.1	77.1	77.6
MHS	60.7	58.6	59.6
WDec	88.1	76.1	81.7
PNDec	80.6	77.3	78.9
CopyMTL-one	72.7	69.2	70.9
CopyMTL-mul	75.7	68.7	72.0
Att-as-Rel	88.1	78.5	83.0
TPLinker	91.4	92.6	92.0
StereoRel (ours)	92.0	92.3	92.2

Table 9: Performance comparison by *Exact Match* on Wiki-KBP

Model	Pre.	Rec.	F1
NovelTagging	53.6	30.3	38.7
PA-Tagging	51.1	39.3	44.4
CasRel	49.8	42.7	45.9
StereoRel (ours)	50.8	42.9	46.5

triple extraction into set prediction problem learned by bipartite matching loss. These ideas may be introduced in the future.

The second one is to recognize nested entities in relational triples. Nested entities are the entities among which there are substring relationships, like “U.N.” being a substring of “U.N. Ambassador”. Such entities definitely affect the overall performance on *Exact Match* mode. Take NYT dataset as an example, there are about 2.5% sentences containing nested entities. Nested entity recognition has been widely studied (Li et al.; Wang et al., 2020a), but most studies on relational triple extraction have not considered it. TPLinker (Wang et al., 2020b) provides a solution to recognize nested entities via a token pair tagging, but it ignores the interaction between entity and relation. For StereoRel model, although not focusing on nested entities, it has not been much affected. The reason is that StereoRel recognizes subjects and objects for each predefined relation separately. In this case, only 0.06% of nested entities in NYT cannot be marked. Anyway, modeling nested entities from the stereoscopic perspective is worth exploring in the future.

6 Conclusions

Relational triple extraction is critical to understanding massive text corpora. However, existing studies

face some challenging issues, including information loss, error propagation and ignoring the interaction between entity and relation. In this paper, aiming at simultaneously handling the above issues, we provide a revealing insight into relational triple extraction from a stereoscopic perspective, which rationalizes the occurrence of these issues and exposes the shortcomings of existing methods. Further, we propose a novel model leveraging three decoders to respectively extract subjects, objects and their correspondences for each predefined relation. Extensive experiments are conducted on five public datasets, demonstrating that the proposed model outperforms the recent advanced baselines.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61822601 and 61773050; the Beijing Natural Science Foundation under Grant Z180006; the Open Project Program Foundation of the Key Laboratory of Opto-Electronics Information Processing, Chinese Academy of Sciences(OEIP-O-202004); National Key Research and Development Project No. 2019YFB1405202.

References

- Giannis Bekoulis, Johannes Deleu, Thomas Demeester, and Chris Develder. 2018. Joint entity recognition and relation extraction as a multi-head selection problem. *Expert Syst. Appl.*, 114:34–45.
- Yee Seng Chan and Dan Roth. 2011. Exploiting syntactico-semantic structures for relation extraction. In *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, USA*, pages 551–560.
- Dai Dai, Xinyan Xiao, Yajuan Lyu, Shan Dou, Qiaoqiao She, and Haifeng Wang. 2019. Joint extraction of entities and overlapping relations using position-attentive sequence labeling. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 6300–6308.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 17th Conference of the*

- North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186.
- Tsu-Jui Fu, Peng-Hsuan Li, and Wei-Yun Ma. 2019. Graphrel: Modeling text as relational graphs for joint entity and relation extraction. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 1409–1418.
- Claire Gardent, Anastasia Shimorina, Shashi Narayan, and Laura Perez-Beltrachini. 2017. Creating training corpora for NLG micro-planners. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 179–188.
- Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations*.
- John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML 2001, Williams College, Williamstown, MA, USA, June 28 - July 1, 2001*, pages 282–289.
- Bing Li, Shifeng Liu, Yifang Sun, Wei Wang, and Xiang Zhao. Recursively binary modification model for nested named entity recognition. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 8164–8171.
- Xiaonan Li, Hang Yan, Xipeng Qiu, and Xuanjing Huang. 2020. FLAT: Chinese NER using flat-lattice transformer. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 6836–6842.
- Jie Liu, Shaowei Chen, Bingquan Wang, Jiaxin Zhang, Na Li, and Tong Xu. 2020. 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 2020*, pages 3787–3793.
- Makoto Miwa and Mohit Bansal. 2016. End-to-end relation extraction using LSTMs on sequences and tree structures. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*.
- Tapas Nayak and Hwee Tou Ng. 2020. Effective modeling of encoder-decoder architecture for joint entity and relation extraction. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 8528–8535.
- Xiang Ren, Zeqiu Wu, Wenqi He, Meng Qu, Clare R. Voss, Heng Ji, Tarek F. Abdelzaher, and Jiawei Han. 2017. CoType: Joint extraction of typed entities and relations with knowledge bases. In *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017*, pages 1015–1024.
- Sebastian Riedel, Limin Yao, and Andrew McCallum. 2010. Modeling relations and their mentions without labeled text. In *Machine Learning and Knowledge Discovery in Databases, European Conference, ECML PKDD 2010, Barcelona, Spain, September 20-24, 2010, Proceedings, Part III*, pages 148–163.
- Dianbo Sui, Yubo Chen, Kang Liu, Jun Zhao, Xiangrong Zeng, and Shengping Liu. 2020. Joint entity and relation extraction with set prediction networks. *CoRR*, abs/2011.01675.
- Ryuichi Takanobu, Tianyang Zhang, Jiexi Liu, and Minlie Huang. 2019. A hierarchical framework for relation extraction with reinforcement learning. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 7072–7079.
- Jue Wang, Lidan Shou, Ke Chen, and Gang Chen. 2020a. Pyramid: A layered model for nested named entity recognition. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 5918–5928.
- Yucheng Wang, Bowen Yu, Yueyang Zhang, Tingwen Liu, Hongsong Zhu, and Limin Sun. 2020b. Tplinker: Single-stage joint extraction of entities and relations through token pair linking. In *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020*, pages 1572–1582.
- Zhepei Wei, Jianlin Su, Yue Wang, Yuan Tian, and Yi Chang. 2020. 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, July 5-10, 2020*, pages 1476–1488.
- Hongbin Ye, Ningyu Zhang, Shumin Deng, Mosh Chen, Chuanqi Tan, Fei Huang, and Huajun Chen.

2020. Contrastive triple extraction with generative transformer. *CoRR*, abs/2009.06207.
- Dmitry Zelenko, Chinatsu Aone, and Anthony Richardella. 2002. Kernel methods for relation extraction. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing, EMNLP 2002, Philadelphia, PA, USA, July 6-7, 2002*, pages 71–78.
- Daojian Zeng, Haoran Zhang, and Qianying Liu. 2020. CopyMTL: Copy mechanism for joint extraction of entities and relations with multi-task learning. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 9507–9514.
- Xiangrong Zeng, Shizhu He, Daojian Zeng, Kang Liu, Shengping Liu, and Jun Zhao. 2019. 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 2019, Hong Kong, China, November 3-7, 2019*, pages 367–377.
- Xiangrong Zeng, Daojian Zeng, Shizhu He, Kang Liu, and Jun Zhao. 2018. 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, Australia, July 15-20, 2018, Volume 1: Long Papers*, pages 506–514.
- Meishan Zhang, Yue Zhang, and Guohong Fu. 2017. End-to-end neural relation extraction with global optimization. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, pages 1730–1740.
- Yue Zhang and Jie Yang. 2018. Chinese NER using lattice LSTM. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pages 1554–1564.
- Suncong Zheng, Feng Wang, Hongyun Bao, Yuexing Hao, Peng Zhou, and Bo Xu. 2017. 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, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 1227–1236.
- Guodong Zhou, Jian Su, Jie Zhang, and Min Zhang. 2005. Exploring various knowledge in relation extraction. In *ACL 2005, 43rd Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, 25-30 June 2005, University of Michigan, USA*, pages 427–434.