Asking Effective and Diverse Questions: A Machine Reading Comprehension based Framework for Joint Entity-Relation Extraction

Tianyang Zhao[1*] , **Zhao Yan**[2] , **Yunbo Cao**[2] , **Zhoujun Li**[1†]

[1]State Key Lab of Software Development Environment, Beihang University, Beijing, China
[2]Tencent Cloud Xiaowei, Beijing, China

{tyzhao,lizj}@buaa.edu.cn, {zhaoyan,yunbocao}@tencent.com

Abstract

Recent advances cast the entity-relation extraction to a multi-turn question answering (QA) task and provide an effective solution based on the machine reading comprehension (MRC) models. However, they use a single question to characterize the meaning of entities and relations, which is intuitively not enough because of the variety of context semantics. Meanwhile, existing models enumerate all relation types to generate questions, which is inefficient and easily leads to confusing questions. In this paper, we improve the existing MRC-based entity-relation extraction model through diverse question answering. First, a diversity question answering mechanism is introduced to detect entity spans and two answering selection strategies are designed to integrate different answers. Then, we propose to predict a subset of potential relations and filter out irrelevant ones to generate questions effectively. Finally, entity and relation extractions are integrated in an end-to-end way and optimized through joint learning. Experiment results show that the proposed method significantly outperforms baseline models, which improves the relation F1 to 62.1% (+1.9%) on ACE05 and 71.9% (+3.0%) on CoNLL04. Our implementation is available at https://github.com/TanyaZhao/MRC4ERE.

1 Introduction

Identifying entity mentions and their relations from unstructured texts is a fundamental and challenging task in information extraction, which has received growing interests recently. Given an input context, the task aims to recognize the entity spans and detect the relations between every head and tail entity pairs, i.e.,(*New York*, PART-WHOLE, *U.S.*).

Existing advances in entity-relation extraction fall into two groups: pipeline approaches and joint approaches. Traditional methods employ a pipelined structure which divides the task into two sub-tasks, recognizing entity spans and predicting relations of any entity pairs. The limitation of these
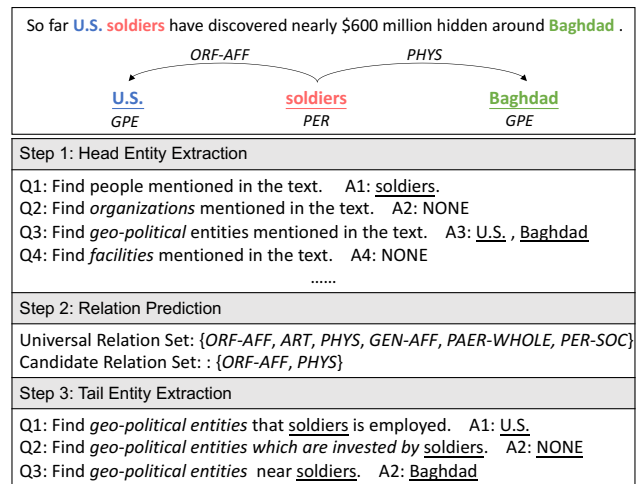

Figure 1: An illustration of the entity-relation extraction of the proposed MRC-based framework.

approaches is obvious - they neglect the potential interactions of the sub-tasks and may suffer from error propagation. Joint approaches integrate entity extraction and relation extraction into a unified model. Various mechanisms for joint learning have been explored, such as parameters sharing [Katiyar and Cardie, 2017], global normalization [Zhang *et al.*, 2017] and joint type decoding [Sun *et al.*, 2019]. They treat relation extraction as a multi-classification task and use multi-classification models to predict the relation of each pair of entities. However, as stated in [Zeng *et al.*, 2018], these models capture only the features based on the input contexts and the entity pairs, which are insufficient to extract all lexical and semantic information.

Recently, with the boost of machine reading comprehension (MRC), several works propose to address entity-relation extraction task with MRC-based method. Levy et al. [2017] firstly reduce relation extraction to the problem of answering simple questions. Later, Li et al. [2019a] improve the framework and propose to transform entity-relation extraction into a multi-turn question answering (QA) task. Their method first detects head entities from the context by answering entity-specific questions using the machine reading comprehension model. Then, it goes through the universal rela-

tion set to generate a relation-specific question based on the head entity. Finally, tail entities are obtained by answering the question, as shown in step 3 of Figure 1. Advantages of the MRC-based framework are as follows. (1) The question provides external prior evidences, i.e.,entity and relation types. (2) The MRC model can better capture the semantic information based on the interaction between question and context. Both above contribute to entity and relation extraction. In this paper, we also focus on the MRC-based framework.

Although previous works for MRC-based entity-relation extraction have achieved great success, these methods still face two challenges. Intuitively, due to the variety of context semantics, using just one question can not well characterize the exact meaning of entities and relations. Meanwhile, it easily leads to confusing questions. For example, the relation ORG-AFF (organization-affiliation) contains a broad sub-types, such as investor-shareholder, ownership, employment, etc. Consider the example in Figure 1, if we represent ORG-AFF only with one question as in [Li *et al.*, 2019b]: *Find geo-political entities which is invested by soldiers*, it is difficult for the MRC model to capture the meaning of *employment* in this case. Generally, multiple explanations can make a complex problem clearer. Therefore, it is necessary to introduce diverse questions to better formalize entities and relations. Second, existing work needs to enumerate all relation types when generating relation-specific questions. That would lead to a large set of question samples. In addition, most of the questions are negative samples and thus result in a serious bias issue and make the extraction less inefficient.

In this paper, we present a novel end-to-end solution to enhance the existing MRC-based entity-relation extraction. To address the first issue, we design a diverse question answering (DQA) mechanism. It exploits multiple simple questions to extract corresponding answers successively. Then, an answer ensemble strategy based on weighted voting is proposed to combine the different answers. As for the second issue, we present to conduct relation prediction (RP) as a prior to obtain a subset of most relevant relations and filter out the useless ones. Hence, questions can be generated effectively based on the relevant relations, rather than traversing the entire set of relations exhaustively, as shown in step 2 of Figure 1. Additionally, to better capture the inherent interaction among the proposed procedures, we combine all the components into an end-to-end structure and optimize the model jointly. Extensive experiments on ACE05 and CoNLL04 datasets demonstrate the effectiveness of the proposed method. To summarize, the main contributions of this work are:

- We design a diverse question answering mechanism to better characterize entities and relations, and obtain the proper answer based on the answer ensemble strategy.

- We propose to apply relation prediction to select most potential relations and filter out irrelevant ones so as to generate relation questions in an effective way.

- By training jointly, the proposed method significantly outperforms the baseline models on both ACE05 and CoNLL04 datasets, which well demonstrates its effectiveness.

2 Related Work

This work relates to three lines of research: relation extraction, machine reading comprehension (MRC) and MRC-based methods for NLP tasks.

2.1 Relation Extraction

Traditional works on relation extraction (RE) adopt pipelined methods that recognize entities first and then predict their relations [Miwa *et al.*, 2009; Chan and Roth, 2011; Lin *et al.*, 2016]. This separation makes the RE task easy to handle, but ignores the inherent interaction between the sub-tasks and is affected by error propagation. To alleviate this limitation, later works propose to extract entities and relations jointly. Earlier joint models are built on hand-crafted features or external parsers which and thus introduce additional complexity. With the success of deep learning models, several neural-network based methods have been presented to address this issue. For example, Miwa et al. [2016] propose to extract entities and relations with tree-structured BiLSTM based on parameter sharing. Then Katiyar and Cardie [2017] replace it with a attention-based network to better model the semantic relations between entities. Later, Zeng et al. [2018] introduce a seq2seq structure to generate entity-relation triples naturally. However, as stated in [Sun *et al.*, 2019], existing joint models can predict entity spans correctly, but predict their types less correctly. So far, to better tackle the joint inference on entity types and relation types, the graph convolutional network is used in [Fu *et al.*, 2019] and [Sun *et al.*, 2019]. Specifically, the proposed model predicts entities and all corresponding relations jointly through a MRC-based method in an end-to-end way.

2.2 Machine Reading Comprehension

In recent years, the boost of large-scale corpora [Rajpurkar *et al.*, 2016; Joshi *et al.*, 2017; Dua *et al.*, 2019] have led to the rapid progress on machine reading comprehension. SQUAD [Rajpurkar *et al.*, 2016] is an extractive MRC benchmark to detect the answer span from the context. A majority of neural-based models tackle the task by predicting the start and the end position of the answer based on the attention mechanism, such as BiDAF [Seo *et al.*, 2017], QANet [Yu *et al.*, 2018] and SAN [Liu *et al.*, 2018]. More recently, the work by Hu et al. [2019] achieves excellent performance by enhancing with the pre-trained contextual embeddings like BERT [Devlin *et al.*, 2019]. However, these models are defective for multi-answer-typed MRC. To address this issue, some existing works propose to validate the answerability of a question first and then predict the candidate answers [Clark and Gardner, 2017]. In this work, instead of extracting the start and end position from context, we predict the answer boundary of every token based on the BIOES tagging scheme. In this way, multiple entity spans can be detected from the context.

2.3 MRC-based Methods for NLP Tasks

Recently, several attempts of addressing NLP tasks with MRC-based methods have been made. For example, Levy et al. [2017] firstly reduce the relation extraction to the problem of reading comprehension and effectively generalize to zero-shot scenarios. McCann et al. [2018] transform ten tasks,

such as summarization, sentiment analysis and relation extraction, into a question answering (QA) paradigm and propose to train all tasks jointly. Different from works above, Li et al. [2019a] present a unified MRC framework and apply it to named entity recognition problem. The latest work [Li *et al.*, 2019b] introduces a multi-turn QA formalization for entity and relation extraction. Our work is significantly inspired by [Li *et al.*, 2019b], but enjoys new features as follows. First, instead of using one question template for extraction, we consider diverse questions to obtain answers from multiple perspectives. Second, during the question generation step, Li et al.'s work traverses all relation types, while we manage to select a subset of relations by filtering out the irrelevances. Finally, we propose an answer ensemble strategy to select the most proper answer. Together these new features improve the extraction performance remarkably.

3 Method

In this section, we introduce each component of the proposed method in detail. A key motivation behind this is that, generally different perspectives of descriptions can make a complex problem clearer. And we find that exploiting different questions helps to extract multiple answers. Additionally, enumerating all relations for question generation is not only costly but can also lead to confusing samples. Therefore, we consider generating effective questions by predicting potential relations and filtering out irrelevant ones.

3.1 Problem Definition

Formally, denote $\mathcal{E}$ and $\mathcal{R}$ as the set of pre-defined entity types and relation categories, respectively. Given an input context with N_c tokens $\boldsymbol{c} = \{c_1, c_2, \ldots, c_{N_c}\}$, the entity-relation extraction task aims to extract a set of entities mentions $\boldsymbol{e} = \{e_1, e_2, \ldots, e_M\}$ with specific types $\boldsymbol{y} = \{y_1, y_2, \ldots, y_M\}$, and predict the relation r_{ij} for each entity pair (e_i, e_j), where $y_i \in \mathcal{E}$ and $r_{ij} \in \mathcal{R}$. Triplets such as (e_i, r_{ij}, e_j) are formulated as the output, where e_i is the head entity and e_j is the tail entity, e.g.,(*New York*, PART-WHOLE, *U.S.*).

In this work, we reduce the entity-relation extraction to the problem of answering simple questions. Specifically, the extraction of triplet (e_i, r_{ij}, e_j) is transformed into two QA steps as follows. First, the head entity e_i is detected from the context by answering questions such as "*Find y_i that mentioned in the text*". Then, relation r_{ij} is mapped to questions that are parameterized by e_i, y_j, and with e_j as the answer. For example, the relation PART-WHOLE corresponds to questions like "*Find y_j that e_i geographically relates to*". In this way, answer spans can be extracted from the context based on the the MRC model.

Overall, the MRC-based entity-relation extraction consists of three steps as follows.

(1) The head entity extraction step. As shown in Figure 2, we generate diverse questions for every entity type. Then, each question is combined with the context and is fed into the MRC-based entity extractor successively. After obtaining corresponding answers, we select the final answer based on the answer ensemble strategy. If no answer is detected, it means this type of entity is not included in the context.
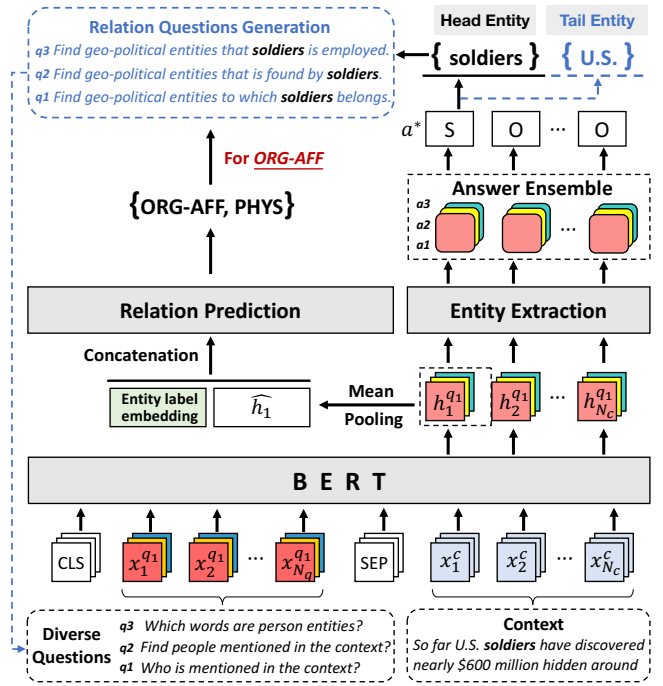

Figure 2: The framework of the proposed MRC-based model for entity-relation extraction. T questions are combined with the context successively as the input of the BERT. Then, T answers are generated and integrated to extract head entities using the answer ensemble strategy. After that, the relation predictor is used to predict the relation prior. Finally, relation questions are generated for tail entity extraction based on the head entities and predicted relations. The blue dotted lines indicate the relation extraction process.

(2) The relation prediction step. In this step, for each extracted head entity e_i, we filter out the low probability relations irrelevant to e_i and predict a potential subset $\mathcal{R}_i \in \mathcal{R}$ to keep the useful ones. In this way, most of the negative samples can be discarded.

(3) The tail entity extraction. Given the extracted head entity, we generate diverse relation-specific questions for each relation $r_{ij} \in \mathcal{R}_i$. Examples are shown in the left part of Figure 2. Then, similar to the head entity extraction step, the questions are integrated with the context and fed into the MRC model to extract potential tail entity e_j. Therefore, the entity-relation extraction problem can effectively be addressed by the MRC-based framework.

3.2 BERT-based MRC model for Entity Extraction

BERT [Devlin *et al.*, 2019] is known as a language representations built on the deep bidirectional transformers. It outperforms state-of-the-art models on a wide-range of NLP tasks, including machine reading comprehension. We use the pretrained BERT as the main structure for the MRC model. As illustrated in Figure 2, given a question $\boldsymbol{q} = \{q_1, q_2, \ldots, q_{N_q}\}$ and an context $\boldsymbol{c} = \{c_1, c_2, \ldots, c_{N_c}\}$, the input of the MRC model are the concatenation as

$$\boldsymbol{x} = [\text{CLS}, x_1^q, \ldots, x_{N_q}^q, \text{SEP}, x_1^c, \ldots, x_{N_c}^c, \text{SEP}], \quad (1)$$

where $\{x_1^q, \ldots, x_{N_q}^q\}$ and $\{x_1^c, \ldots, x_{N_c}^c\}$ are the word piece embeddings of the question Q and the context C, respectively. CLS denotes a special token and SEP denotes a separator. Encoded by the multi-layer self-attention structure, BERT outputs the contextual representation for each context token as $\boldsymbol{h} = \{h_1, h_2, \ldots, h_{N_c}\}$, $h_i \in \mathbb{R}^{d_h}$, where d_h denotes the dimension of the last hidden layer of BERT.

Then, considering that the context might have multiple answers, we apply a softmax classification layer to the hidden outputs $\boldsymbol{h}$ and predict the BIOES boundary labels. For each input x_i, the probability of the candidate BIOES label can be calculated as

$$\Pr(\text{label} = \hat{a}|x_i) = \text{softmax}(W^b \cdot h_i + b^b), \qquad (2)$$

where $W^b \in \mathbb{R}^{d_h \times d_b}$, $b^b \in \mathbb{R}^{d_b}$ are learned parameters, d_b is the size of boundary label set $\mathcal{B}$, and $\hat{a}$ denotes the predicted boundary label. Consequently, candidate entities $\boldsymbol{e} = \{e_1, e_2, \ldots, e_M\}$ can be extracted from the label sequence by identifying the boundaries.

3.3 Diverse Question Answering

Intuitively, explaining a problem from different perspectives can make it more clear. As an inspiration, we generate a group of questions for each entity and relation type based on the pre-defined question templates. Questions within a group share the same meanings but they are expressed in different ways. For example, to identify the PER (person) entities in a context, three questions with the same semantics but diverse expressions can be generated as follows.

- *q1*: Who is mentioned in the context?
- *q2*: Find people mentioned in the context?
- *q3*: Which words are person entities?

Specifically, we use T questions as $\boldsymbol{Q} = \{\boldsymbol{q_1}, \boldsymbol{q_2}, \ldots, \boldsymbol{q_T}\}$ for entity extraction. As shown in Figure 2, T questions are combined with the context $\boldsymbol{c}$ as the input to the MRC model. Then T corresponding answers are obtained as $\boldsymbol{A} = \{\boldsymbol{a_1}, \boldsymbol{a_2}, \ldots, \boldsymbol{a_T}\}$, where $\boldsymbol{a_t} = \{a_{t1}, a_{t2}, \ldots, a_{tN_c}\}$ is the boundary sequence obtained by the MRC model (Eq.2).

Answer Ensemble Strategy. To ensemble different answers, we propose a weighted voting scheme to obtain the proper answer dynamically. Consider w_t as the weight for each question $\boldsymbol{q_t} \in \boldsymbol{Q}$, which is initialized as 1.0. At the end of each training epoch, we calculate the F1 score f_t for question $\boldsymbol{q_t}$ on the development set and update the weight w_t as

$$w_t = \sigma(f_t) * T, \qquad (3)$$

where $\sigma(\cdot)$ is the sigmoid function and $*$ is the element-wise multiplication. Note that, the higher the F1 score, the higher the weight. Hence, the weight w_t showcases the quality of the question $\boldsymbol{q_t}$. Based on the learned weight, the final ensemble answer $\boldsymbol{a^\star} = \{a_1^\star, a_2^\star, \ldots, a_N^\star\}$ is obtained by weighted voting on the token-level. Specifically, the boundary label of the i-th token is selected as

$$a_i^\star = \arg\max_{\mathcal{B}} \sum_t w_t \cdot a_{ti}. \qquad (4)$$

To this end, the final extracted entities can be inferred based on the ensemble answer $\boldsymbol{a^\star}$.

3.4 Relation Prediction

Relation prediction aims to identify for each extracted head entity e_i, $i \in \{1, 2, \ldots, M\}$ the set of most probable relation types $R_i \subseteq \mathcal{R}$. This is different from previous work [Li *et al.*, 2019b] which need to enumerate every relation type in $\mathcal{R}$ to generate relation questions. We predict a prior relation types for each head entity and thus those irrelevant relations will be filtered out.

Specifically, denote $h_i^{q_t}$ as the BERT contextual representation for the start token of the head entity e_i, q_t denotes the t-th question, and x_i^l the corresponding entity label embedding, the input to the relation predictor is the concatenation of $\hat{h}_i$ and x_i^l as

$$\begin{aligned} \hat{h}_i &= \text{avg}_{t \in T} h_i^{q_t} \\ l_i &= [\hat{h}_i, x_i^l], \end{aligned} \qquad (5)$$

where x_i^l is initialized by random sampling and will be fine-tuned during training. Then, this input is fed to a softmax classifier to yield the probability for entity e_i with each relation type $r_k \in \mathcal{R}$ as

$$\Pr(\text{label} = r_k|e_i) = \sigma(W^r \cdot l_i + b^r), \qquad (6)$$

where $\sigma(\cdot)$ is the sigmoid function, $W^r \in \mathbb{R}^{(d_h+d_l) \times |\mathcal{R}|}$, $b^r \in \mathbb{R}^{|R|}$, d_l is the dimension of the entity label embedding and $|\mathcal{R}|$ is the size of the relation set. The high score in the classifier denotes the corresponding relation holds for entity e_i. Consider a confidence threshold δ, we keep any relation label with a score exceeding δ. And those labels with scores lower than δ will be discarded.

3.5 Joint Training

To train the model jointly, we optimize the combined objective function during training:

$$\mathcal{L} = \mathcal{L}_{\text{head}} + \mathcal{L}_{\text{rel}} + \mathcal{L}_{\text{tail}}, \qquad (7)$$

where $\mathcal{L}_{\text{head}}$ and $\mathcal{L}_{\text{tail}}$ denote the cross-entropy loss for head entity and tail entity extraction, respectively. $\mathcal{L}_{\text{rel}}$ denotes the binary cross-entropy loss over relation types for relation prediction. The head and tail entity extractor are built on the standard BERT model and share the parameters for training. $\mathcal{L}$ is averaged over samples for each batch.

4 Experiment

In this section, we empirically demonstrate the effectiveness of diverse question answering and relation prediction strategies for MRC-based entity-relation extraction.

4.1 Datasets

We evaluate the proposed method on two widely-used benchmarks for entity relation extaction: ACE05 and CoNLL04.

- **ACE05** contains 7 entity types LOC, ORG, PER, GPE, VEH, FAC, WEA and 6 relation types ORG-AFF, PER-SOC, ART, PART-WHOLE, GEN-AFF, PHYS. We adopt the same data splits as previous work [Miwa and Bansal, 2016].

	Model	Entity			Relation		
		P	R	F1	P	R	F1
ACE05	[Li and Ji, 2014]	85.2	76.9	80.8	**65.4**	39.8	49.5
	[Miwa and Bansal, 2016]	82.9	83.9	83.4	57.2	54.0	55.6
	[Katiyar and Cardie, 2017]	84.0	81.3	82.6	55.5	51.8	53.6
	[Zhang *et al.*, 2017]	-	-	83.5	-	-	57.5
	[Sun *et al.*, 2019]	83.9	83.2	83.6	64.9	55.1	59.6
	[Li *et al.*, 2019b]*	84.7	84.9	84.8	64.8	56.2	60.2
	MRC4ERE	85.1 ($\pm$0.4)	84.2 ($\pm$0.2)	84.6 ($\pm$0.2)	57.8 ($\pm$0.3)	61.9 ($\pm$0.2)	59.8 ($\pm$0.2)
	MRC4ERE++	85.9 ($\pm$0.4)	85.2 ($\pm$0.3)	**85.5 ($\pm$0.2)**	62.0 ($\pm$0.4)	62.2 ($\pm$0.4)	**62.1 ($\pm$0.2)**
CoNLL04	[Miwa and Sasaki, 2014]	81.2	80.2	80.7	76.0	50.9	61.0
	[Adel and Schütze, 2017]	-	-	82.1	-	-	62.5
	[Zhang *et al.*, 2017]	-	-	85.6	-	-	67.8
	[Bekoulis *et al.*, 2018]	83.8	84.1	83.9	63.8	60.4	62.0
	[Li *et al.*, 2019b]*	89.0	86.6	87.8	69.2	68.2	68.9
	MRC4ERE	87.0 ($\pm$ 0.2)	88.6 ($\pm$ 0.3)	87.8 ($\pm$ 0.1)	65.0 ($\pm$ 0.2)	72.4 ($\pm$ 0.3)	68.5 ($\pm$ 0.2)
	MRC4ERE++	89.3 ($\pm$ 0.2)	88.5 ($\pm$ 0.5)	**88.9 ($\pm$ 0.3)**	72.2 ($\pm$ 0.4)	71.5 ($\pm$ 0.3)	**71.9 ($\pm$ 0.2)**

Table 1: Performance comparisons on ACE05 and CoNLL04. Here, we report the average result and the standard deviation (scores in brackets) when re-training with 5 random seeds. **MRC4ERE** uses only one question for entity and relation extraction and enumerates universal relation set to generate questions. **MRC4ERE++** is the full model presented in Section 3 embedded with the diverse question answering and relation prediction mechanisms. * denotes baselines based on the MRC framework.

- **CoNLL04** defines 4 entity types including `LOC`, `ORG`, `PER` and `Other` and 5 relation categories as `Located-In`, `OrgBased-In`, `Live-In`, `Kill` and `Work-For`. We use the data split by Gupta et al. [2016]. To tune hyperparameters, 20% of the training set is selected as the development set.

4.2 Implementation Details

We adopt the $BERT_{base}$ (cased) [Devlin *et al.*, 2019] as the MRC model for our experiments. We apply the BIOES tagging scheme for boundary classification. The entity type embedding is initialized randomly with a uniform distribution and the size d_l is set as 50. We generate 3 different questions with a simple natural language-based template for the diverse question answering. The threshold δ for the relation prediction is set as 0.3. During training, we use the early stopping to avoid overfitting based on the performance on the development set. For evaluation, precision (P), recall (R) and micro-F1 score (F1) are used as metrics in our experiments.

4.3 Result and Analysis

Baselines. We consider the following strong baselines for comparison. Model [Li and Ji, 2014] adopts an incremental beam-search framework to extract entities and relations. Models [Miwa and Sasaki, 2014] and [Zhang *et al.*, 2017] treat relation extraction as a table-filling problem, the later enhance it with global optimization. Miwa and Bansal [2016] present a tree-based LSTM to capture dependency information. Models [Katiyar and Cardie, 2017] and [Adel and Schütze, 2017] replace the tree structure with the attentional LSTMs and the globally normalized CNNs, respectively. Bekoulis et al. [2018] address the relation extraction task with a multi-head selection model. Sun et al. [2019] explore the graph convolutional network for entity relation extraction Model [Li *et al.*, 2019b] is MRC-based but uses one question for extraction and enumerates all the relation types.

Experimental Results. Table 1 illustrates the performance of the proposed method against previous state-of-the-arts on ACE05 and CoNLL04 dataset. The first block of subtable lists the published results of previous models. As shown in Table 1, **MRC4ERE++** significantly outperforms all the baselines for both entity and relation extraction on the two datasets. Specifically, the relation F1 scores of **MRC4ERE++** advances the best model [Li *et al.*, 2019b] by +1.9% and +3.0% on ACE05 and CoNLL04, respectively. Meanwhile, our method is stable with all F1 standard deviations are no more than 0.3. We perform a significant test with the best baseline suggesting that performance is statistically significant ($p < 0.05$). We highlight the improvement benefits from two scientific contributions: the diverse question answering and the relation prediction, which enhance the formation of MRC-based entity-relation extraction obviously. In addition, the performance of MRC-based models is remarkably superior to non-MRC-based baselines, which verifies its effectiveness. We consider the reasons are: (1) The question provides important prior type information. (2) The MRC model can better capture the interaction between the question and the context based on the self-attention structure.

Ablation Study. We further study the effects of each proposed components: the diverse question answering (DQA), the relation prediction (PR) and the weighted voting strategy. Results are listed in Table 2. Specifically, using diverse questions can significantly improve the relation F1 scores by +0.7% and +3.5% on the two datasets comparing **MRC4ERE + DQA** with **MRC4ERE**. We consider that, by asking diverse questions, the model can extract multiple entities from different perspective. Meanwhile, under the single question scenario, filtering out irrelevant relations increases the relation F1 from 59.8% to 60.4% on the ACE05 (**MRC4ERE + RP** v.s. **MRC4ERE**). The performance of this setting is comparable with the full model on the CoNLL04. We attribute the results to that, the entity-

Model	Entity		Relation	
	F1	**Δ**	**F1**	**Δ**
ACE05				
MRC4ERE	84.6		59.8	
+ DQA	85.7	+1.1	60.5	+0.7
+ RP	84.8	+0.2	60.4	+0.6
+ RP + OBQ	85.6	+1.0	61.5	+1.7
MRC4ERE++	85.5	+0.9	62.1	+2.3
CoNLL04				
MRC4ERE	87.8		68.5	
+ DQA	88.3	+0.5	72.0	+3.5
+ RP	88.0	+0.2	69.0	+0.5
+ RP + OBQ	88.6	+0.8	71.7	+3.2
MRC4ERE++	88.9	+1.1	71.9	+3.4

Table 2: Ablation Study on ACE05 and CoNLL04. **MRC4ERE** is the simplified model with neither diverse question answering (DQA) and relation prediction (RP) included. **MRC4ERE + DQA** adopts DQA but without RP. **MRC4ERE + RP** uses RP but without DQA. **MRC4ERE + RP + OBQ** replace the weighted voting answer ensemble strategy by selecting answers corresponding to the one best question with highest weight. **MRC4ERE++** is the full model.

relation mapping is much simpler for CoNLL04 than ACE05, e.g.,almost one-to-one mapping, and thus there would be much less noise. Hence, the relation prediction mechanism plays an essential role for complicated datasets. Furthermore, with both DQA and RP integrated, **MRC4ERE++** achieves further $+2.3\%$ and $+3.4\%$ boosts for the relation F1 on the two datasets. Finally, simply selecting the best answer is not as effective as the proposed weighted voting strategy (**MRC4ERE + RP + OBQ** v.s. **MRC4ERE++**). By voting dynamically, the model can integrate more confidential answers, which is also crucial to entity-relation extraction.

4.4 Effects of Question Generating Template

Questions used in the full model are generated with a simple natural language-based template. It is also possible to obtain them using auto-generated pseudo-questions. Therefore, we study the effects of the two question generation ways.

Specifically, we generated the entity-specific pseudo-questions based on different descriptions of the entity type. Examples for the person entity are (1) *person*; (2) *entity:person*, (3) *find person*. The relation-specific pseudo-questions are the combinations of head entity text, relation type and tail entity type such as (1) *soldier;organization-affiliation;geo-political entity*; (2) *soldier;employment;geo-political entity*; (3) *soldier;ownership;geo-political entity*.

In Table 3, models using pseudo-questions are obviously inferior to models using natural language. The reason is that natural language can provide more semantic information. But the pseudo-questions lack auxiliary words that contain structural information of entities so that it is difficult for models to understand. However, by using diverse pseudo-questions, the performance of **MRC4ERE++**$_P$ is improved and the results are strongly competitive to model **MRC4ERE**$_N$ that uses one natural language question.

Model	Entity P / R / F1	Relation P / R / F1
[Li *et al.*, 2019b]$_P$	83.6 / 84.7 / 84.2	60.4 / 55.9 / 58.1
MRC4ERE$_P$	83.3 / 85.0 / 84.2	57.8 / 59.8 / 58.8
MRC4ERE++$_P$	84.5 / 85.6 / 85.0	58.8 / 61.1 / 59.9
[Li *et al.*, 2019b]$_N$	84.7 / 84.9 / 84.8	64.8 / 56.2 / 60.2
MRC4ERE$_N$	85.4 / 84.2 / 84.7	57.8 / 61.9 / 60.0
MRC4ERE++$_N$	85.9 / 85.2 / 85.5	62.0 / 62.2 / 62.1

Table 3: Comparison between models using natural language questions (N) and models using pseudo-questions (P) on ACE05.

Model	Relation		
	P	**R**	**F1**
[Miwa and Bansal, 2016]	70.1	61.2	65.3
[Christopoulou *et al.*, 2018]	69.7	59.5	64.2
[Sun *et al.*, 2019]	68.7	65.4	67.0
MRC4ERE++	67.7	70.1	**68.7**

Table 4: Relation Extraction results on ACE05 with golden entity.

4.5 Golden Entity Results on ACE05

To better evaluate the performance of our model in relation extraction, we conduct test with golden head entities on ACE05 datasets, which showcases the upper bound result that our model can achieve for relation extraction. Specifically, we keep the same experimental setting with the baseline models [Miwa and Bansal, 2016; Sun *et al.*, 2019]. The result in Table 4 shows that, the proposed method outperforms existing relation classification models by a large margin. This indicates that our method is able to capture the relevant information of given head entities which helps to improve the performance of relation extraction.

4.6 Conclusion

In this paper, we propose an end-to-end solution to improve the existing MRC-based entity-relation framework. First, we present a diverse question answering mechanism and an answer ensemble strategy to extract the proper answer from different perspectives. Then, we introduce the relation prediction method to obtain useful relations for question generation. Finally, the model is combined and optimized through joint learning. Extensive experiments show that the proposed model is effective for entity-relation extraction and achieves significant improvement on ACE05 and CoNLL04 datasets.

Acknowledgments

References

[Adel and Schütze, 2017] Heike Adel and Hinrich Schütze. Global normalization of convolutional neural networks for joint entity and relation classification. In *EMNLP*, pages 1723–1729, 2017.

[Bekoulis *et al.*, 2018] Giannis Bekoulis, Johannes Deleu, Thomas Demeester, and Chris Develder. Joint entity recognition and relation extraction as a multi-head selection problem. *Expert Systems with Applications*, 114:34–45, 2018.

[Chan and Roth, 2011] Yee Seng Chan and Dan Roth. Exploiting syntactico-semantic structures for relation extraction. In *ACL*, pages 551–560, 2011.

[Christopoulou *et al.*, 2018] Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou. A walk-based model on entity graphs for relation extraction. In *ACL*, 2018.

[Clark and Gardner, 2017] Christopher Clark and Matt Gardner. Simple and effective multi-paragraph reading comprehension. *arXiv preprint arXiv:1710.10723*, 2017.

[Devlin *et al.*, 2019] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL*, pages 4171–4186, 2019.

[Dua *et al.*, 2019] Dheeru Dua, Yizhong Wang, Pradeep Dasigi, Gabriel Stanovsky, Sameer Singh, and Matt Gardner. Drop: A reading comprehension benchmark requiring discrete reasoning over paragraphs. In *NAACL*, pages 2368–2378, 2019.

[Fu *et al.*, 2019] Tsu-Jui Fu, Peng-Hsuan Li, and Wei-Yun Ma. Graphrel: Modeling text as relational graphs for joint entity and relation extraction. In *ACL*, pages 1409–1418, 2019.

[Gupta *et al.*, 2016] Pankaj Gupta, Hinrich Schütze, and Bernt Andrassy. Table filling multi-task recurrent neural network for joint entity and relation extraction. In *COLING*, pages 2537–2547, 2016.

[Hu *et al.*, 2019] Minghao Hu, Furu Wei, Yuxing Peng, Zhen Huang, Nan Yang, and Dongsheng Li. Read+ verify: Machine reading comprehension with unanswerable questions. In *AAAI*, volume 33, pages 6529–6537, 2019.

[Joshi *et al.*, 2017] Mandar Joshi, Eunsol Choi, Daniel Weld, and Luke Zettlemoyer. Triviaqa: A large scale distantly supervised challenge dataset for reading comprehension. In *ACL*, pages 1601–1611, 2017.

[Katiyar and Cardie, 2017] Arzoo Katiyar and Claire Cardie. Going out on a limb: Joint extraction of entity mentions and relations without dependency trees. In *ACL*, pages 917–928, 2017.

[Levy *et al.*, 2017] Omer Levy, Minjoon Seo, Eunsol Choi, and Luke Zettlemoyer. Zero-shot relation extraction via reading comprehension. *arXiv preprint arXiv:1706.04115*, 2017.

[Li and Ji, 2014] Qi Li and Heng Ji. Incremental joint extraction of entity mentions and relations. In *ACL*, pages 402–412, 2014.

[Li *et al.*, 2019a] Xiaoya Li, Jingrong Feng, Yuxian Meng, Qinghong Han, Fei Wu, and Jiwei Li. A unified mrc framework for named entity recognition. *arXiv preprint arXiv:1910.11476*, 2019.

[Li *et al.*, 2019b] Xiaoya Li, Fan Yin, Zijun Sun, Xiayu Li, Arianna Yuan, Duo Chai, Mingxin Zhou, and Jiwei Li. Entity-relation extraction as multi-turn question answering. *arXiv preprint arXiv:1905.05529*, 2019.

[Lin *et al.*, 2016] Yankai Lin, Shiqi Shen, Zhiyuan Liu, Huanbo Luan, and Maosong Sun. Neural relation extraction with selective attention over instances. In *ACL*, pages 2124–2133, 2016.

[Liu *et al.*, 2018] Xiaodong Liu, Yelong Shen, Kevin Duh, and Jianfeng Gao. Stochastic answer networks for machine reading comprehension. In *ACL*, pages 1694–1704, 2018.

[McCann *et al.*, 2018] Bryan McCann, Nitish Shirish Keskar, Caiming Xiong, and Richard Socher. The natural language decathlon: Multitask learning as question answering. *arXiv preprint arXiv:1806.08730*, 2018.

[Miwa and Bansal, 2016] Makoto Miwa and Mohit Bansal. End-to-end relation extraction using lstms on sequences and tree structures. *arXiv preprint arXiv:1601.00770*, 2016.

[Miwa and Sasaki, 2014] Makoto Miwa and Yutaka Sasaki. Modeling joint entity and relation extraction with table representation. In *EMNLP*, pages 1858–1869, 2014.

[Miwa *et al.*, 2009] Makoto Miwa, Rune Sætre, Yusuke Miyao, and Jun'ichi Tsujii. A rich feature vector for protein-protein interaction extraction from multiple corpora. In *EMNLP*, pages 121–130, 2009.

[Rajpurkar *et al.*, 2016] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. Squad: 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv:1606.05250*, 2016.

[Seo *et al.*, 2017] Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hannaneh Hajishirzi. Bidirectional attention flow for machine comprehension. In *ICLR*, 2017.

[Sun *et al.*, 2019] Changzhi Sun, Yeyun Gong, Yuanbin Wu, Ming Gong, Daxin Jiang, Man Lan, Shiliang Sun, and Nan Duan. Joint type inference on entities and relations via graph convolutional networks. In *ACL*, pages 1361–1370, 2019.

[Yu *et al.*, 2018] Adams Wei Yu, David Dohan, Minh-Thang Luong, Rui Zhao, Kai Chen, Mohammad Norouzi, and Quoc V Le. Qanet: Combining local convolution with global self-attention for reading comprehension. In *ICLR*, 2018.

[Zeng *et al.*, 2018] Xiangrong Zeng, Daojian Zeng, Shizhu He, Kang Liu, Jun Zhao, et al. Extracting relational facts by an end-to-end neural model with copy mechanism. In *ACL*, 2018.

[Zhang *et al.*, 2017] Meishan Zhang, Yue Zhang, and Guohong Fu. End-to-end neural relation extraction with global optimization. In *EMNLP*, pages 1730–1740, 2017.