J. Appl. Numer. Optim. 4 (2022), No. 3, pp. 393-404
Available online at http://jano.biemdas.com
https://doi.org/10.23952/jano.4.2022.3.06

# HIERARCHICAL REINFORCEMENT LEARNING WITH ADVANTAGE FUNCTION FOR ENTITY RELATION EXTRACTION

XIANCHAO ZHU[1,2,*], WILLIAM ZHU[2]

[1]*School of Artificial Intelligence and Big Data, Henan University of Technology, Zhengzhou 450001, China*
[2]*Institute of Fundamental and Frontier Sciences,*
*University of Electronic Science and Technology of China, Chengdu 610054, Chengdu*

**Abstract.** Unlike the traditional pipeline methods, the joint extraction approaches use a single model to distill the entities and semantic relations between entities from the unstructured texts and achieve better performances. A pioneering work, HRL-RE, uses a hierarchical reinforcement learning model to distill entities and relations that decompose the entire extraction process into a high-level relationship extraction and a low-level entity identification. HRL-RE makes the extraction of entities and relations more accurate while solving overlapped entities and relations to a certain extent. However, this method has not achieved satisfactory results in dealing with overlapped entities and relations in sentences. One reason is that learning a policy is usually inefficient, and the other one is the high variance of gradient estimators. In this paper, we propose a new method, Advantage Hierarchical Reinforcement Learning for Entity Relation Extraction (AHRL-ERE), which combines the HRL-RE model with a new advantage function to distill entities and relations from the structureless text. Specifically, based on the reference value of the policy function in the high-level subtask, we construct a new advantage function. Then, we combine this advantage function with the value function of the strategy in the low-level subtask to form a new value function. This new value function can immediately evaluate the current policy, so our AHR-ERE method can correct the direction of the policy gradient update in time, thereby making policy learning efficient. Moreover, our advantage function subtracts the reference value of the high-level policy value function from the low-level policy value function so that AHRL-ERE can decrease the variance of the gradient estimator. Thus our AHRL-ERE method is more effective for extracting overlapped entities and relations from the unstructured text. Experiments on the diffusely used datasets demonstrate that our proposed algorithm has better manifestation than the existing approaches do.

**Keywords.** Advantage function; Entity relation extraction; Hierarchical reinforcement learning; Overlapped entities and relations.

## 1. INTRODUCTION

Entities and relations extraction play an important role in many social computing applications, such as intelligent question answering [1], repository construction [2], and biological text mining [3, 4]. The purpose of entities and relations extraction is to distill the triples with pairs of entities and the probabilistic-lexical relations between them from the structureless textual data.

*Corresponding author.
E-mail addresses: xczhuiffs@163.com (X. Zhu), wfzhu@uestc.edu.cn (W. Zhu).
Received June 29, 2022; Accepted September 3, 2022.

The traditional entities and relations extraction methods use pipeline approaches to divide the entire extraction task into two component tasks: entity identification and relationship extraction [5, 6]. The entity recognition subtask adopts the entity extractor to identify each entity in the sentence text to obtain entity pairs. In the relation extraction subtask, they extract the relation between two input entities through a relation recognizer. These methods make the extraction tasks easier to handle, and each component can be more flexible. In this way, the result of entity identification will influence the manifestation of relation extraction, leading to error propagation [7, 8].

To address this problem, joint extraction methods were proposed to use a single model to train the entity identification and the relation extraction contemporaneously, which can better make use of the relationships between two subtasks, to get better performances [9, 10, 11, 12]. For example, Miwa et al. adopted bi-directional LSTM with syntax information and obtained entity tags in the entity recognition subtask at first. Then they used the LSTM output result of the entity identification subtask to distill the relation between entities in the relation extraction subtask through the bi-tree structure LSTM and Softmax [10]. Zheng et al. regarded entities and relations extraction as a sequence labeling task. They used bi-directional LSTM and uni-directional LSTM to encode and decode, respectively. The output layer simultaneously labels entities and relations to complete entities and relations joint extraction [11]. However, these approaches fail to deal with overlapped entities and relations in unstructured texts generally. There are mainly two types of overlapped entities and relations. One entity may participate in multiple relations in the same sentence. Moreover, the same entity pair in a sentence may be associated with diverse relations [13].

Recently, Takanobu et al. introduced a hierarchical reinforcement learning model, HRL-RE, to jointly extract entities and relations in structureless texts, making entities and relations extraction more accurate and solving overlapped entities and relations to a certain extent [13, 14]. HRL-RE attempts to cope with overlapped entities and relations in sentences by decomposing the entire extraction task into two subtasks: high-level relation detection and low-level entity extraction. In the high-level relation detection subtask, HRL-RE computes the high-level policy function by Bi-LSTM, then detects the relation indicator at a specific position. After identifying the particular relation, the high-level policy will trigger the low-level entity extraction subtask, calculate the corresponding low-level policy function by the Monte Carlo Gradient Estimation method, and extract the corresponding entity pair for that relation. When the current low-level subtask for entity extraction is accomplished, the high-level RL subtask continues its scan to search for the next relation in the sentence. However, this method has not achieved satisfactory results in dealing with overlapped entities and relations in sentences. One reason is that the learning process is tardy and has many invalid attempts, resulting in low efficiency of policy learning. The other one is the high variance of gradient estimators caused by the change of the reference value of the high-level policy function.

To address the above two issues, we present a new joint extraction method, Advantage Hierarchical Reinforcement Learning for Entity Relation Extraction (AHRL-ERE), which combines a new advantage function with HRL-RE. Specifically, inspired by the option-critic method, based on the reference value of the policy function in the high-level subtask, we construct a new advantage function. Then, we incorporate this advantage function with the value function of the strategy in the low-level subtask to form a new value function. When the advantage function value is positive, the current policy is better than the previous policy, so the gradient update direction of the model should be consistent with the gradient direction of the current policy. Conversely, the gradient update direction of the model is the negative gradient direction of the current policy. Evaluating the current policy by the advantage function timely, our AHRL-ERE approach can modify the direction of the policy gradient update, thus improving the efficiency of the policy learning. In addition, by subtracting the reference value of the high-level strategy value function from the low-level policy value function, AHRL-ERE can reduce the variance of the gradient estimator [15]. Therefore, our proposed method can more effectively distill overlapped entities and relations in unstructured texts. Experimental results on two entity relationship extraction datasets, NTY10 and NTY11, show that our method surpasses existing methods by a substantial margin.

This paper is structured as follows. In Section 2, we present a concise review of entities and relations extraction approaches, an introduction to hierarchical reinforcement learning, and a concise review of the HRL-RE method. Subsequently, Section 3 presents our Advantage Hierarchical Reinforcement Learning for Entity Relation Extraction (AHRL-ERE) method to deal with overlapped entities and relations extraction via hierarchical reinforcement learning with an advantage function. Then, Section 4 presents experiments on two publicly available New York Times corpus to demonstrate that our algorithm outperforms other entities and relations extraction methods and achieves significantly higher scores in overlapping entities and relations extraction tasks. Finally, Section 5 concludes the paper.

## 2. Related Work

In this section, we present a concise review of entities and relations extraction algorithms, an introduction to hierarchical reinforcement learing and a concise review of HRL-RE method.

2.1. **Entities and relations extraction.** Entity relationship extraction plays a vital role in social computing applications. One of its primary purposes is extracting meaningful structured information from original unstructured texts for use in social computing applications, such as intelligent question answering [1], repository construction [2], and biological text mining [3, 4, 16, 17, 18].

Traditional pipeline methods regard entity identification and relationship extraction as two independent tasks [19, 20, 21, 22]. In the entity recognition subtask, they adopt the entity extractor to identify each entity in the sentence text to obtain entity pairs. In the relation extraction subtask, they extract the relation between the two input entities by relation recognizer. Pipeline methods have long been explored, but they sustain the propagation of errors issue from entity identification to relation extraction.

To address this issue, various joint extraction methods were proposed [9, 10, 11, 12]. For example, Miwa et al. adopted bi-directional LSTM with syntax information to obtain entity tags in the entity recognition subtask. They used the LSTM output result of the entity identification subtask to distill the relation between entities in the relation extraction subtask through the bi-tree structure LSTM and Softmax [10]. Zheng et al. regarded entities and relationships extraction as a sequence labeling task. They used bi-directional LSTM and unidirectional LSTM to encode and decode, respectively. The output layer simultaneously labels entities and relations to complete entities and relations joint extraction [11]. Bjorne et al. proposed extracting the relation recognizer for the first time, which refers to a phrase that clearly expresses the relation occurring in the sentence and then determines their parameters to reduce the complexity of the task [23]. The open IE system ReVerb [24] uses lexical constraints to recognize the relation phrases, which also follow the "relation"-first, "parameter"-second approach.

Recently, the effectiveness of RL was verified in entities and relations extraction [25, 26, 27, 28]. Feng et al. used RL to joint extract the entities and relation types [29]. Qin et al. presented a deep RL approach for entities and relations extraction [28]. Feng et al. presented a approach for entity relationship extraction which consists of an entity selector with RL and a relationship recognizer with CNN [30].

2.2. **Hierarchical reinforcement learning.** Hierarchical reinforcement learning (HRL) is a branch in the field of RL. Traditional RL uses interaction with the environment to continuously conduct trial and error to optimize the strategy [31, 32]. The most famous theoretical framework of HRL is the options framework (as shown in Figure 1) [14, 33]. The option is a three-element tuple $(I,\pi,\beta)$, where $\pi : S \to [0,1]$ represents the strategy, which is a probability distribution function based on the state space and the motion space; $\beta : S \to [0,1]$ is the termination condition, $\beta(s)$ indicates that the state $s$ has the probability of $\beta(s)$ to terminate and exit the current option; $I \subseteq S$ indicates the initial state of the option. An option-based learning process is as follows: the agent first initializes in a particular state, then selects an option, then chooses an action or other options according to the strategy of this option, then executes the action or enters a new option, and continues to choose or termination.

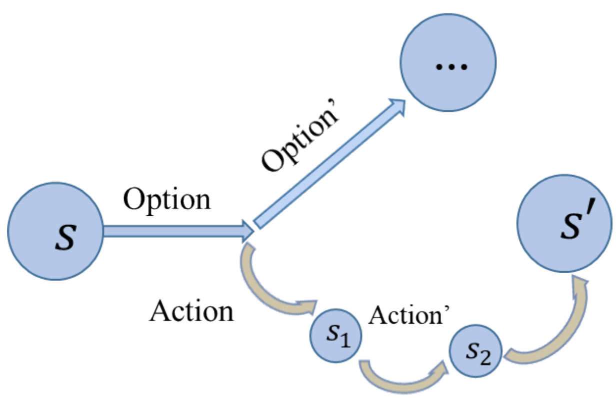


FIGURE 1. The framework of hierarchical reinforcement learning model.

2.3. **Hierarchical reinforcement learning for entitie relation extraction.** The HRL-RE method decomposes the entire entities and relations extraction task into two subtasks [13]. Detection subtask, HRL-RE computes the state's policy processed by the Bi-LSTM, and then acquires the relation type. After obtaining the relation type, this high-level policy will touch off the low-level subtask for entity extraction. In the low-level subtask, HRL-RE calculates the strategy by the Monte Carlo Gradient Estimation method to obtain the entities pair corresponding to this relation. The high-level RL subtask will continue to seek for the next relation in the sentence when the current low-level subtask is accomplished [13].

HRL-RE makes the extraction of entities and relations more accurate while solving overlapped entities and relations to a certain extent. However, this method has not achieved satisfactory results in dealing with overlapped entities and relations in sentences. One reason is that the learning process is tardy and has many invalid attempts, resulting in low efficiency of policy learning, and the other one is the high variance of gradient estimators caused by the change of the reference value of the high-level policy function.

## 3. ADVANTAGE HIERARCHICAL REINFORCEMENT LEARNING FOR ENTITY RELATION EXTRACTION

In this section, we present new joint overlapped entities and relations extraction approach, Advantage Hierarchical Reinforcement Learning for Entity Relation Extraction (AHRL-ERE), to deal with the two problems in Subsection 2.3 via embedding an advantage function into hierarchical reinforcement learning. Specifically, we construct a new advantage function based on the reference value of the policy function in the high-level subtask and incorporate this advantage function with the value function of the strategy in the low-level subtask to form a new value function. Evaluating the current strategy by the advantage function timely, our AHRL-ERE approach can modify the direction of the policy gradient update, thus improving the efficiency of the policy learning. In addition, by subtracting the reference value of the high-level strategy value function from the low-level strategy value function, AHRL-ERE can reduce the variance of the gradient estimator.

3.1. **High-level RL model for relation detection.** We adopt the viewpoint of the HRL-RE method to construct the high-level relationship detection task [13]. The high-level relationship detection subtask scans the sentence line by line and computes the high-level policy $O$ ($O \in NR \cup \mathscr{R}$) of the present time step in the light of the state, where $\mathscr{R}$ denotes all the relationships of the present dataset and $NR$ denotes "no relation".

State: the high-level state $s_t^h \in S$ at present time step $t$ is computed as Equation (3.1). It is computed from the current hide state $h_t$ of the present time step $t$, the relation type vector $v_t^r$ of the recent $non-NR$ high-level policy $o$, and the state $s_{t-1}$ of the anterior time step $t-1$.

$$s_t^h = f^h(W_s^h[h_t; v_t^r; s_{t-1}]), \tag{3.1}$$

where $f^h(\cdot)$ represents a non-linear transformation, and $W_s^h$ represents a weight matrix.

High-level policy: adopting a random policy $\mu : S \longrightarrow O$, the present time step state $s_t^h$, is computed by the softmax layer. The output from softmax layer is stochastically sampled to acquire the present time step $t$ behaviour $o_t$:

$$o_t \sim \mu(o_t|s_t^h) = softmax(W_\mu, s_t^h), \tag{3.2}$$

where $W_\mu$ represents a weight matrix.

Reward: the environment offers a feedback signal $r_t^h$ to estimate the reward when executing $o_t$:

$$r_t^h = \begin{cases} -1 & if\ o_t\ not\ in\ S, \\ 0 & if\ o_t = NR, \\ 1 & if\ o_t\ in\ S. \end{cases} \tag{3.3}$$

3.2. **Low-level RL model for entity extraction.** We adopt the viewpoint of the HRL-RE method to consruct the low-level entity extraction task [13]. The low-level task scans the sentence line by line and computes the present time step action in the light of the state $s_t^l$ and policy $\pi$. If high-level RL strategy predicts a $non-NR$ relation, low-level RL will extract entities in the relation. The high-level policy $o_t$ of high-level RL will be used as an additional input for low-level RL.

Action: the action will allot a entity tag to the word at each time step. The entity tag $A = (\{\mathbf{S},\mathbf{T},\mathbf{O}\} \times \{\mathbf{B},\mathbf{I}\} \cup \{\mathbf{U}\})$, where **S** represents the originating entity, **T** is the objective entity, **O** represents the irrelevant entity, **N** represents a non-entity word, **B** is the beginning of the entity, and **I** represents the internal entity.

State: the formal expression of the low-level state $s_t^l$ is as follows:

$$c_t = g(W_h^l s_t^h),$$

and

$$s_t^l = f^l(W_s^l[h_t; v_t^e; s_{t-1}; c_t]), \tag{3.4}$$

where $c_t$ is the context vector.

Low-level policy: adopt a stochastic policy $\pi : S \longrightarrow A$ to stochastically sample the probabilities output from the softmax layer to acquire the present time step $t$ action.

$$a_t \sim \pi(a_t|s_t^l; o_t) = softmax(W_\pi[o_t s_t^l]), \tag{3.5}$$

where $W_\pi$ represents an array of $R$ matrices.

Reward: the return $r_t^l$ for the present time step $t$ is shown as

$$r_t^l = \lambda(y_t) \cdot sgn(a_t = y_t(o_t)), \tag{3.6}$$

where $sgn$ represents the symbolic function, and $\lambda(y)$ represents a bias weight function for down-weight non-entity tage, defined as follows:

$$\lambda(y) = \begin{cases} 1 & if\ y \neq N, \\ \alpha & if\ y = N. \end{cases}$$

After the sequence labeling of this relation is completed, the overall reward $r_{fin}^l$ is: if all the sequence labels are correct, $r_{fin}^l$ is 1; otherwise, it is $-1$.

3.3. **Hierarchical policy learning models.** Similar to HRL-RE [13], AHRL-ERE optimizes the policy by maximising the expected discounted cumulative return:

$$J(\theta_{\mu,t}) = \mathbb{E}_{s^h,o,r^h \sim \mu(o|s^h)}[\sum_{k=t}^{T} \gamma^{k-t} r_k^h],$$

where $\mu$ is parameterized by $\theta_\mu$, and $\gamma$ represents the discount factor in RL.

Different from HRL-RE, AHRL-ERE uses the following formula to calculate the expected discounted cumulative return of the low-level model:

$$J(\theta_{\pi,t}; o_t) = \mathbb{E}_{s^l,a,r^l \sim \pi(a|s^l;o_t)}(r_k^l + A(\theta_{\pi,t}; o_t)),$$

where $\pi$ is parameterized by $\theta_\pi$, $\varepsilon$ is a hyperparameter, and $A(\theta_{\pi,t}; o_t)$ is the advantage function of the high-level policy $A(\theta_{\pi,t}; o_t) = (1-\varepsilon)J(\theta_{\mu,t}) - \varepsilon \max_{\mu \in O} J(\theta_{\mu,t})$.

We decompose the cumulative rewards into a Bellman equation:

$$R^\mu(s_t^h, o_t) = \mathbb{E}[\sum_{j=0}^{N-1} r_{t+j}^h \gamma^N R^\mu(s_{t+N}^H, o_{t+N})|s_t^h, o_t],$$

and

$$R^\pi(s_t^l, a_t; o_t) = \mathbb{E}[r_t^l + \gamma R^\pi(s_{t+1}^l, a_t; o_t)|s_t^l, o_t],$$

where $N$ represents the number of time step of the entity extraction subtask started under the present high-level policy $o_t$.

The gradients of high-level subtask and low-level subtask are computed by the policy gradient method [34] and the REINFORCE method [35]. The gradient for high-level strategy is defined as follows:

$$\nabla_{\theta_\mu} J(\theta_{\mu,t}) = \mathbb{E}_{s^h,o,r^h \sim \mu(o|s^h)}[R^\mu(s_t^h, o_t)\nabla_{\theta_\mu}\mu(o|s_t^h)]. \tag{3.7}$$

Different from HRL-RE, AHRL-ERE uses the following formula to update the gradient for low-level strategy:

$$\nabla_{\theta_\pi} J(\theta_{\pi,t;o_t}) = \mathbb{E}_{s^l,a,r^l \sim \pi(a|s^l;o_t)}[R^\pi(s_t^l, a_t; o_t)\nabla_{\theta_\pi}\pi(a|s_t^l; o_t)A(\theta_{\pi,t}; o_t)]. \tag{3.8}$$

When the advantage function value is positive, the current policy is better than the previous policy, so the gradient update direction of the model should be consistent with the gradient direction of the current policy. Conversely, the gradient update direction of the model is the negative gradient direction of the current policy. In this way, by timely correcting the update direction of the policy gradient, this advantage function helps learn a policy efficiently. Moreover, by subtracting the reference value of the high-level policy value function from the low-level policy value function, AHRL-ERE can decrease the variance of gradient estimators. Thus our proposed method can more effectively extract overlapped entities and relations in the unstructured texts.

The AHRL-ERE approach is as follows:

**Algorithm 1** Advantage Hierarchical Reinforcement Learning for Entity Relation Extraction (AHRL-ERE)

Compute $h_t$ for each entity in the sentence by Bi-LSTM;
Initialize state $s_0^h \leftarrow 0$ and time step $t \leftarrow 0$;
**for** $i \leftarrow 1$ to $TextLength$ **do**
  $t \leftarrow t+1$;
  Compute $s_t^h$ through Equation (3.1);
  Sample $o_t$ from $s_t^h$ by Equation (3.2);
  Acquire reward $r_t^h$ by Equation (3.3);
  **if** $o_t \neq NR$ **then**
    **for** $j \leftarrow 1$ to $TextLength$ **do**
      $t \leftarrow t+1$;
      Compute $s_t^l$ through Equation (3.4);
      Sample $a_t^l$ from $s_t^l$ by Equation (3.5);
      Acquire reward $r_t^l$ by Equation (3.6);
    **end for**
    Acquire low-level final reward $r_{fin}^l$;
  **end if**
**end for**
Acquire high-level reward $r_{fin}^h$ through Equation (3.3);
Optimize the model with Equation (3.7) and Equation (3.8);

## 4. EXPERIMENTS

4.1. **Experimental setting.** **Dataset** The dataset adopted in this paper is New York Times corpus, which comes frombe from the research on distant supervision and contains noisy relations [36, 37].

**Evalution Criterion** We employ the Precision, Recall, and micro-F1 to assess the method's performance. We compare whether the extracted entities can be precisely matched with those in a relation type.

**Baselines** We select both the pipeline approaches (FCM) and the joint extraction approaches as the baselines.

FCM ([20]): a pipeline model which combines the handcrafted lexicalized linguistic contexts with word embeddings for entities and relations extraction.

MultiR ([37]): a distant supervision method that adopts multiple instances weighting to cope with noisy labels in training data.

CoType ([38]): A single model embeds entities, relations, text characteristics, and type labels into a representation and regards extraction as a global embedding issue.

SPTree ([10]): a joint extraction method that uses bi-directional sequential and bi-directional tree knots LSTM-RNNs to distill entities and relationships in a single model.

Tagging ([11]): an joint extraction approach to extract entities and relationships contemporaneously with a new tagging schema.

CopyR ([39]): a Seq2Seq learning model that adopts multiple decoders to generate triples to distill entities and relations jointly.

HRL-RE ([13]): a HRL based method that decomposes the entire extraction task into a high-level relationship detection subtask and low-level entity extraction subtask.

4.2. **Experimental results.** Table 1 demonstrates the experimental results of entity relation extraction. It is worth noting that there is a significant discrepancy between the manifestation on the noisy dataset (NYT10) and the clean dataset (NYT11) because all models are trained on noisy data. It can be seen that our proposed algorithm AHRL-ERE obtains the better manifestation of other entity relation extraction methods on both data sets. In particular, the score on the NYT10 data set is much higher than the score on the NTY11 data set, which means that proposed algorithm is more robust to noisy data.

**Table 1** Experiment result on entities and relations extraction.

| Model | NTY10 | | | NTY11 | | |
|---|---|---|---|---|---|---|
| | Prec | Rec | $F_1$ | Prec | Rec | $F_1$ |
| FCM | — | — | — | 0.432 | 0.294 | 0.350 |
| MultiR | — | — | — | 0.328 | 0.306 | 0.317 |
| Cotype | — | — | — | 0.486 | 0.386 | 0.430 |
| SPTree | 0.492 | 0.557 | 0.522 | 0.522 | 0.541 | 0.531 |
| Tagging | 0.593 | 0.381 | 0.464 | 0.469 | 0.489 | 0.479 |
| CopyR | 0.569 | 0.452 | 0.504 | 0.347 | 0.534 | 0.421 |
| HRL-RE | 0.714 | 0.586 | 0.644 | 0.538 | 0.538 | 0.538 |
| **AHRL-ERE** | **0.721** | **0.594** | **0.652** | **0.541** | **0.543** | **0.541** |

4.3. **Overlapping entities and relations extraction.** We exhibit the superiority of our approach for extracting overlapping entities and relations on two test sets of NYT11-plus and NYT10-sub. Note that overlapping entities and relations can be divided into two categories.

- Type 1: one entity take part in in multiple relations in the same sentence.
- Type 2: the same entity pair in a sentence is associated with different relations.

Table 2 demonstrates the performance of overlapping entities and relations extracted by different entities and relations extraction methods. The experimental results on the NYT10-sub dataset indicate that our method obtains better performance than the HRL-RE method. In addition, compared with our AHRL-ERE method and HRL-RE method, other entities relations extraction methods perform very poorly in dealing with noisy data in type 2 overlapping entities and relations, which shows that the traditional joint extraction method is essentially unable to deal with overlapping entities and relations effectively. Therefore, our proposed approach is more suitable for dealing with the type 2 overlapping entities and relations problem on noisy data than other entities and relations extraction methods. The results of the experiment on the NYT11-plus dataset indicate that our approach obtains better performance than other entities and relations extraction methods in extracting the type 1 overlapping entities and relations on the clean data. In summary, our method can distill overlapping entities and relationships more effectively.

**Table 2** Performance comparison on extracting overlapping entities and relations.

| Model | NTY10-sub | | | NTY11-plus | | |
|---|---|---|---|---|---|---|
| | Prec | Rec | $F_1$ | Prec | Rec | $F_1$ |
| FCM | — | — | — | 0.234 | 0.199 | 0.219 |
| MultiR | — | — | — | 0.241 | 0.214 | 0.227 |
| Cotype | — | — | — | 0.291 | 0.254 | 0.271 |
| SPTree | 0.272 | 0.315 | 0.292 | 0.466 | 0.229 | 0.307 |
| Tagging | 0.256 | 0.237 | 0.246 | 0.292 | 0.220 | 0.250 |
| CopyR | 0.392 | 0.263 | 0.315 | 0.329 | 0.224 | 0.264 |
| HRL-RE | 0.815 | 0.475 | 0.600 | 0.441 | 0.321 | 0.372 |
| **AHRL-ERE** | **0.821** | **0.480** | **0.607** | **0.449** | **0.328** | **0.376** |

4.4. **Interaction between the two levels of subtasks.** The experimental results in Table 3 exhibit that our approach obtains better manifestation than other entities and relations extraction methods in relation detection on the two data sets. In particular, the improvement on NYT11-plus data is more significant because our method is more suitable for extracting multiple relations from sentences. Therefore, embedding the entity as a kind of relation parameter in the relation detection can make more and better use of the relation information in the text.

When omitting the low-level strategy from the models HRL-RE-Ent and AHRL-ERE-Ent, respectively, the performance of the NYT11 data changes slightly. The reason is that each sentence in this test set contains almost only one relation. In this case, the interaction between the high-level subtask strategy and the low-level subtask strategy has nearly no effect on the results of the relation detection. On the contrary, there are significant differences in the NYT11-plus data set, which means that AHRL-ERE and AHRL-ERE based on hierarchical reinforcement learning can capture the dependencies between multiple extraction tasks. In addition, this interaction can increase the return of high-level subtask strategy. Therefore, the entities and relations extraction method based on hierarchical reinforcement learning does enhance the interaction between relation detection and entity extraction.

**Table 3** Comparison of experimental results of relation prediction.

| Model | NTY11 | | | NTY11-plus | | |
|---|---|---|---|---|---|---|
| | Prec | Rec | $F_1$ | Prec | Rec | $F_1$ |
| FCM | 0.502 | 0.479 | 0.490 | 0.447 | 0.327 | 0.378 |
| MultiR | 0.465 | 0.439 | 0.451 | 0.423 | 0.336 | 0.375 |
| Cotype | 0.558 | 0.558 | 0.558 | 0.491 | 0.413 | 0.449 |
| SPTree | 0.650 | 0.614 | 0.631 | **0.700** | 0.343 | 0.460 |
| CopyR | 0.480 | **0.714** | 0.574 | 0.626 | 0.426 | 0.507 |
| HRL-RE-Env | **0.676** | 0.676 | **0.676** | 0.577 | 0.321 | 0.413 |
| HRL-RE | 0.654 | 0.654 | 0.654 | 0.626 | 0.456 | 0.527 |
| **AHRL-ERE-Env** | **0.676** | 0.676 | **0.676** | 0.577 | 0.321 | 0.413 |
| **AHRL-ERE** | 0.657 | 0.657 | 0.657 | 0.631 | **0.461** | **0.532** |

## 5. Conclusion

In this paper, we proposed a new entities and relationships extraction approach, Advantage Hierarchical Reinforcement Learning for Entity Relation Extraction (AHRL-ERE), which combines a new advantage function with HRL-RE. Different from HRL-RE, based on the reference value of the strategy function in the high-level subtask, we constructed a new advantage function. Then, we combined this advantage function with the value function of the strategy in the low-level subtask to form a new value function. This new value function can immediately evaluate the current policy, so our AHR-ERE method can correct the direction of the policy gradient update in time, thereby making policy learning efficient. Moreover, our advantage function subtracts the reference value of the high-level strategy value function from the low-level strategy value function so that AHRL-ERE can decrease the variance of the gradient estimator. In this way, our AHRL-ERE method is more effective for extracting overlapped entities and relationships from the structureless text. Experiments on the diffusely used datasets exhibit that our approach has better performance than the selected baselines.

**Acknowledgments**

This work was supported by the National Nature Science Foundation of China under Grant No. 61772120.

## References

[1] A. Fader, L. Zettlemoyer, O. Etzioni, Open question answering over curated and extracted knowledge bases, The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, pp. 1156-1165, New York, 2014.

[2] Y. Luan, L. He, M. Ostendorf, H. Hajishirzi, Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction, Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 3219–3232, Brussels, 2018.

[3] C.C. Huang, Z. Lu, Community challenges in biomedical text mining over 10 years: success, failure and the future, Briefings in Bioinformatics, 17 (2016), 132-144. DOI: 10.1093/bib/bbv024.

[4] H. Fei, Y. Ren, Y. Zhang, et al., Enriching contextualized language model from knowledge graph for biomedical information extraction, Briefings in Bioinformatics, 22 (2021). DOI: 10.1093/bib/bbaa110.

[5] D. Nadeau, S. Sekine, A survey of named entity recognition and classification, Lingvisticae Investigationes, 30 (2007), 3-26.

[6] M. Carbonell, A. Fornés, M. Villegas, et al., A neural model for text localization, transcription and named entity recognition in full pages, Pattern Recognition Lett. 136 (2020), 219-227.

[7] F. Li, Me. Zhang, G. Fu, et al., A neural joint model for entity and relation extraction from biomedical text, BMC Bioinformatics 18 (2017), 198.

[8] S. Wang, Y. Zhang, W. Che, Joint extraction of entities and relations based on a novel graph scheme, Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, pp. 4461–4467, Stockholm, S 2018.

[9] M. Miwa, Y. Sasaki, Modeling joint entity and relation extraction with table representation, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, pp. 1858–1869, Doha, 2014.

[10] M. Miwa, M. Bansal, End-to-End Relation Extraction using LSTMs on Sequences and Tree Structures, Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, vol. 1, Berlin, 2016.

[11] S. Zheng, F. Wang, H. Bao, et al., Joint extraction of entities and relations based on a novel tagging scheme, Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, pp. 1227-1236, Vancouver, 2017.

[12] A. Katiyar, C. Cardie, Going out on a limb: Joint extraction of entity mentions and relations without dependency trees, Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, pp. 917-928, Vancouver, , 2017.

[13] R. Takanobu, T. Zhang, J. Liu A hierarchical framework for relation extraction with reinforcement learning, The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, pp. 7072–7079, Honolulu, 2019.

[14] R.S. Sutton, D. Precup, S. Singh, Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning, Artificial Intelligence 112 (1999), 181-211.

[15] Sugiyama, Masashi, Statistical reinforcement learning: modern machine learning approaches, CRC Press, 2015.

[16] X. Shen, X. Zhu, X. Jiang, et al. Visualization of disease relationships by multiple maps t-SNE regularization based on Nesterov accelerated gradient, 2017 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2017, pp. 604-607, Kansas City, 2017.

[17] X. Shen, X. Zhu, X. Jiang, et al. Visualization of non-metric relationships by adaptive learning multiple maps t-SNE regularization, 2017 IEEE International Conference on Big Data, BigData 2017, pp. 3882-3887, Boston, 2017.

[18] X. Zhu, X. Shen, X. Jiang, et al., Nonlinear expression and visualization of nonmetric relationships in genetic diseases and microbiome data, BMC Bioinformatics 19 (2018), 37-48.

[19] M. Mintz, S. Bills, R. Snow, et al., Distant supervision for relation extraction without labeled data, ACL 2009, Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the AFNLP, pp. 1003–1011, Singapore, 2009.

[20] M.R. Gormley, M. Yu, M. Dredze, Improved relation extraction with feature-rich compositional embedding models, Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, pp. 1774–1784, Lisbon, 2015.

[21] J. Tang, M. Qu, M. Wang, et al., Line: Large-scale information network embedding, Proceedings of the 24th International Conference on World Wide Web, WWW 2015, 1067-1077, 2015.

[22] P. Zhou, W, Shi, J. Tian, Attention-based bidirectional long short-term memory networks for relation classification, Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, Berlin, 2016.

[23] J. Björne, J. Heimonen, F. Ginter, Extracting contextualized complex biological events with rich graph-based feature sets, Computational Intelligence 27 (2011), 541-557.

[24] A. Fader, S. Soderland, O. Etzioni, Identifying relations for open information extraction, Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, EMNLP 2011, pp. 1535–1545, Edinburgh, 2011.

[25] K. Narasimhan, A. Yala, R. Barzilay, Improving information extraction by acquiring external evidence with reinforcement learning, Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, pp. 2355-2365, Austin, 2016.

[26] A. Katiyar, C. Cardie, Investigating lstms for joint extraction of opinion entities and relations, Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, pp. 919-929, Berlin, G 2016.

[27] M. Zhang, Y. Zhang, G. Fu, End-to-end neural relation extraction with global optimization, Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, pp. 1730-1740, Copenhagen, 2017.

[28] P. Qin, W. Xu, W.Y. Wang, Robust distant supervision relation extraction via deep reinforcement learning, Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, pp. 2137-2147, Melbourne, 2018.

[29] Y. Feng, H. Zhang, W. Hao, et al;, Joint extraction of entities and relations using reinforcement learning and deep learning, Comput. Intelligence Neurosci. 2017 (2017), 7643065.

[30] J. Feng, M. Huang, L. Zhao, et al., Reinforcement learning for relation classification from noisy data, Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), New Orleans, 2018.

[31] R.S. Sutton, A.G. Barto, Andrew G, Reinforcement learning: An introduction, MIT press, 2018.

[32] Alexander Sasha Vezhnevets and Simon Osindero and Tom Schaul and Nicolas Heess and Max Jaderberg and David Silver and Koray Kavukcuoglu, FeUdal Networks for Hierarchical Reinforcement Learning, Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August, pp. 3540–3549, PMLR, 2017.

[33] X. Zhu and T. Huang, R. Zhang WDIBS: Wasserstein deterministic information bottleneck for state abstraction to balance state-compression and performance, Appl. Intelligence, 52 (2022), 6316-6329.

[34] R.S. Sutton, D.A. McAllester, S.P. Singh, Policy gradient methods for reinforcement learning with function approximation, Advances in Neural Information Processing Systems 12, pp. 1057-1063, Denver, 1999.

[35] R.J. Williams, Simple statistical gradient-following algorithms for connectionist reinforcement learning, Mach. Learn. 8 (1992), 229-256.

[36] S. Riedel, L. Yao, A. McCallum, Modeling relations and their mentions without labeled text, Machine Learning and Knowledge Discovery in Databases, European Conference, ECML PKDD 2010, vol. 6323, pp. 148-163, Barcelona, 2010.

[37] R. Hoffmann, C. Zhang, X. Ling Knowledge-based weak supervision for information extraction of overlapping relations, The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp. 541–550, Portland, 2011.

[38] X. Ren, Z. Wu, W. He, et al., Cotype: Joint extraction of typed entities and relations with knowledge bases, Proceedings of the 26th International Conference on World Wide Web, WWW 2017, pp. 1015-1024, Perth, 2017.

[39] X. Zeng, D. Zeng, S. He, Extracting relational facts by an end-to-end neural model with copy mechanism, Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, pp. 506-514, Melbourne, 2018.