A Simple and Unified Tagging Model with Priming for Relational Structure Predictions

I-Hung Hsu† Kuan-Hao Huang†‡ Shuning Zhang⋄ Wenxin Cheng‡
Premkumar Natarajan† Kai-Wei Chang‡ Nanyun Peng‡
†Information Science Institute, University of Southern California
‡Computer Science Department, University of California, Los Angeles
⋄Computer Science and Technology Department, Tsinghua University
{ihunghsu, pnataraj}@isi.edu, zhang-sn19@mails.tsinghua.edu.cn
{khhuang, kwchang, violaepeng}@cs.ucla.edu, wenxin0319@ucla.edu

Abstract

Relational structure extraction covers a wide range of tasks and plays an important role in natural language processing. Recently, many approaches tend to design sophisticated graphical models to capture the complex relations between objects that are described in a sentence. In this work, we demonstrate that simple tagging models can surprisingly achieve competitive performances with a small trick — priming. Tagging models with priming append information about the operated objects to the input sequence of pretrained language model. Making use of the contextualized nature of pretrained language model, the priming approach helps the contextualized representation of the sentence better embed the information about the operated objects, hence, becomes more suitable for addressing relational structure extraction. We conduct extensive experiments on three different tasks that span ten datasets across five different languages, and show that our model is a general and effective model, despite its simplicity. We further carry out comprehensive analysis to understand our model and propose an efficient approximation to our method, which can perform almost the same performance but with faster inference speed.

1 Introduction

Many tasks in natural language processing require extracting relational structures from the given context. For example, the event argument extraction task aims to identify event arguments and their corresponding roles to the given event trigger (Huang et al., 2022; Wang et al., 2019). In end-to-end relation extraction, the model recognizes the relationship(s) between tail-entities and head-entities (Wei et al., 2020; Yu et al., 2020). In task-oriented semantic parsing, the model predicts slots and semantic roles of the intent (Tür et al., 2010; Li et al., 2021). These tasks are beneficial for a wide range of applications, such as dialog system (Liu et al., 2018), building knowledge graph (Zhang et al., 2020), question answering (Berant et al., 2014; Han et al., 2021; Yasunaga et al., 2021), and narrative generation (Chen et al., 2019a).

Various task-specific methods have been proposed for different tasks that demand relational structure extraction. Graph-based methods are one of the most commonly used frameworks to deal with such problems (Luan et al., 2019; Lin et al., 2020; Sun et al., 2019; Miwa and Bansal, 2016; Fu et al., 2019). These methods make relational extraction by considering the interaction between the objects (such as event triggers, named entities, etc.) and their relationships. By leveraging the relational interaction, they iteratively update the representation for each object to make its representation more comprehensive and better suit for modeling the whole graph structure, as illustrated in Figure 1(a).

On the other hand, sequential tagging models are conventional methods for addressing the extraction of relational structures (Yang et al., 2019a; Hakkani-Tür et al., 2016; Ma et al., 2020; Zhang et al., 2018; Ramponi et al., 2020). These methods convert the relational structure into a sequential format and make predictions by labeling tokens in the input passage. Compared to graph-based models, tagging models get less attention in recent studies since graph-based methods are expected to capture more complex interaction between objects and thus lead to stronger results.

In this work, we re-examine the effectiveness of sequential tagging methods on relational structure extraction and present a simple and unified, yet
strong model. Inspired by Fincke et al. (2022), we introduce sequential tagging model with priming, which augments the input with the task-specific context and perform sequential labeling on the input passage, as shown in Figure 1(c) & (d). The priming approach that makes the sentential representations task-specific well employs the contextualized pre-trained language models’ nature. Hence, compared to the commonly used feature embedding approach, it better exploits the capability of sequential tagging methods for relational structure extraction, as we will show in Section 4.

To better understand the mechanism of priming, we carry out experiments to analyze whether deeply-stacked layers is essential for priming method to equip the sentential representation with task-specific information. We observe that although deeper layers help, priming sequential models with much shallower layers can achieve similar performance. Based on the study, we further introduce an efficient approximation of the sequential tagging model with priming and the approximating model can achieve similar performance but faster inference speed.

Our contributions can be summarized as follows. First, we present a simple and unified sequential tagging model that can tackle several relational structure extraction tasks and serve as a strong baseline for further studies. The model is verified in three different tasks and in various languages. The results are comparable or achieve a new state-of-the-art on three different natural language processing tasks in ten datasets in total. Second, we perform a thorough analysis on the priming methods, and propose a novel efficient approximation to fasten up models during inference time with only a small performance drop.

2 Relational Structure Extraction with Sequential Tagging

In this section, we first introduce a unified formulation of relational structure extraction (Section 2.1)
and how previous works usually apply sequential tagging model to relational structure extraction (Section 2.2). Then, we present the improved sequential tagging model with the priming technique (Section 2.3).

2.1 A Unified Formulation of Relational Structure Extraction

Given a passage \( x = [x_1, x_2, ..., x_n] \) and a condition \( c \), a relational structure extraction model identifies a list of objects \( \mathbf{o}^c = [o^c_1, o^c_2, ..., o^c_i] \) and their corresponding relationships \( \mathbf{r}^c = [r^c_1, r^c_2, ..., r^c_i] \) to the condition \( c \), where \( r^c_i \in \mathcal{S}^r \) and \( \mathcal{S}^r \) is the set of all possible relationships. Many relational structure extraction tasks can be approached by this formulation and applied to several natural language processing (NLP) tasks. In this paper, we specifically focus on the following three NLP tasks.

**End-to-end event extraction.** End-to-end event extraction aims to extract events, each of which consists of an event trigger and several arguments with their specific roles, from a given passage (Ma et al., 2020; Hsu et al., 2022b; Yang et al., 2019a). We consider a pipelined solution such that, after the event triggers are identified, a relational structure extraction model extracts the event arguments and their corresponding roles for each given event trigger. In this formulation, the condition \( c \) is the given event trigger, and the objects \( \mathbf{o}^c \) and the relationships \( \mathbf{r}^c \) are the event arguments and their argument roles, respectively.

**End-to-end relation extraction.** End-to-end relation extraction extracts entities and their relations from text. We consider the cascading pipelined approach to solve the task (Wei et al., 2020; Yu et al., 2020; Sun et al., 2019). After named entities are extracted, we apply a relational structure extraction model to predict tail-entities and their relationships for each extracted named entity that serve as the “head-entity”. Thus, each given head-entity is the condition \( c \), the extracted tail-entities are \( \mathbf{o}^c \), and their relationships are \( \mathbf{r}^c \).

**Task-oriented semantic parsing.** Task-oriented semantic parsing intends to classify the intent and parse the slots in an utterance (to a task-oriented dialog system) (Li et al., 2021; Gupta et al., 2018). In our framework, we first predict the intent and use a relational structure extraction model to predicts the slots (\( \mathbf{o}^c \)) and the semantic roles (\( \mathbf{r}^c \)) in the utterance for the given intent, which serves as the condition \( c \).

2.2 Sequential Tagging Model for Relational Structure Extraction

We introduce the typical way to apply a sequential tagging model to solve relational structure extraction. The goal of a sequential tagging model for relational structure extraction is to predict BIO-tags \( (y_i) \) for each token \( (x_i) \) in the input passage. The BIO-tag sequence can then be decoded to represent the objects \( \mathbf{o}^c \) (and their relation \( \mathbf{r}^c \)) to be extracted.

Specifically, given an input passage, we obtain the contextualized representation \( z_i \) for each token \( x_i \) by passing the passage to a pre-trained language model.\(^1\) To embed the information of the condition \( c \), one commonly-used technique is to add conditional features to \( z_i \) (Ma et al., 2020; Wei et al., 2020; Yang et al., 2019a; Yu et al., 2020). Following Ma et al. (2020), we consider a condition token feature and a learnable condition type feature, as shown in Figure 1(b). The token feature of \( c \) is the contextualized word representation \( z_j \), if \( x_j \) is the token that represents the condition. The learnable type feature helps the modeling of the attribute of a condition, such as the condition’s event type in event argument extraction. Augmented with these conditional features, the final representation for token \( x_i \) is further fed into multi-layer perceptrons and a conditional random field layer (Lafferty et al., 2001) to predict each token’s corresponding tag \( y_i \).

2.3 Sequential Tagging Model with Priming

**Condition Priming.** Motivated by previous work (Fincke et al., 2022), we consider priming to inject the information of the condition \( c \) to further improve the sequential tagging model. The priming mechanism informs the model the conditional information by directly appending conditional information to the input of pretrained language models. However, unlike Fincke et al. (2022) that uses an integer string to represent features in a categorical style, we use a natural-language-styled indicator to better exploit the semantics of the condition. The indicators can be obtained by simply verbalizing the conditional information.

Take Figure 1(c) as an example, when extracting the tail-entities and the relationships for the “military” head-entity, we first verbalize the entity type

\(^1\)If a token \( x_i \) is split into multiple word pieces, we use the average of all its word pieces embeddings to be \( z_i \).
of “military”, i.e, from “Org” to “Organization”. Then, the string “Organization” will be concatenated to the input sequence together with the token “military”, which serves as the token information about the condition $c$.

Our priming technique leverages the contextualized pre-trained language models’ nature to make the token representation $z_t$ condition-aware. Hence, the representation of every $z_t$ is more task-specific than the one in the model that described in Section 2.2. More precisely, for tagging models without priming, $z_t$ usually captures a more general representation that focuses on the context of input passage. For models with priming, the contextualized representation $z_t$ is affected by the additional verbalized words inherently, hence, becomes task-specific and more suitable for addressing the task (Zheng and Lapata, 2022; Zhong and Chen, 2021). Additionally, the priming method can be easily combined with learnable features, and the discussion on the learnable feature versus priming method will be shown in Section 4.

### Relation Priming
The same idea can also be extended to relation. We first decompose a relational structure extraction task into several extraction sub-tasks, each of that only focuses on relational extraction for one single relation $r$ ($r \in S^r$). Under this setting, the tagging model’s input contains a verbalized relation string for the given relation $r$. Thus, $z_t$ is aware of the queried relation $r$ and suitable for predicting whether any object in the passage forms a relationship with the entity $c$ in the type $r$.

For example, in Figure 1(d), for the given relation “Part-Whole”, we first verbalized it into “is part of”. Then, the “is part of” string is appended to the input sequence for the model. The BIO-tagging sequence can be decoded into those tail-entities $o^r$ that form “Part-Whole” relationship(s) with the head-entity “military”.

### 3 Experiments

To study the effectiveness of the sequential tagging method, we conduct experiments on end-to-end event extraction, end-to-end relation extraction, and task-oriented semantic parsing. All the experimental results are the average of five runs with different random seeds.

#### 3.1 End-to-End Event Extraction

**Datasets.** We consider English and Chinese annotations in ACE-2005 (ACE05-E) (Doddington et al., 2004) and keep 33 event types and 22 roles, as suggested in previous works (Wadden et al., 2019). We also consider English annotations and Spanish annotations in ERE (Song et al., 2015) and follow the preprocessing of Lin et al. (2020) to keep 38 event types and 21 roles.

**Baselines.** We consider the following end-to-end event extraction models, including DyGIE++ (Wadden et al., 2019), TANL (Paolini et al., 2021), Text2Event (Lu et al., 2021), OneIE (Lin et al., 2020), and DEGREE (Hsu et al., 2022b). Since TagPRIME requires trigger predictions, we simply use the trigger predictions made by a sequential tagging model trained with multi-tasking on trigger detection and named entity recognition.

For TagPRIME, DyGIE++, and OneIE, we consider BERT-large for ACE05-E (en) and ERE (en), and consider XLM-R-large for ACE05-E (zh) and ERE (es). For generation-based models, we consider BART-large for DEGREE, T5-base for TANL, and T5-large for Text2Event, as their original papers suggest.

**Evaluation metrics.** Following previous works (Wadden et al., 2019; Lin et al., 2020), we measure the correctness of arguments based on whether the offsets of the argument spans match or not. We consider argument identification F1-score (Arg-I), which cares about only the offset correctness, and

| Model                          | ACE05-E (en) | ACE05-E (zh) | ERE (en) | ERE (es) |
|-------------------------------|-------------|-------------|----------|----------|
|                               | Tri-C Arg-I Arg-C | Tri-C Arg-I Arg-C | Tri-C Arg-I Arg-C | Tri-C Arg-I Arg-C |
| DyGIE++ (Wadden et al., 2019) | 69.7        | 53.0        | 48.8     | 72.3     |
| TANL (Paolini et al., 2021)   | 68.4        | 50.1        | 47.6     | 73.3     |
| Text2Event (Lu et al., 2021)  | 71.9        | -           | 53.8     | 68.4     |
| OneIE (Lin et al., 2020)      | 74.7        | 59.2        | 56.8     | 73.3     |
| DEGREE (Hsu et al., 2022b)    | 73.3        | -           | 55.8     | 73.3     |
| TAGPRIME w/ Condition Priming | 74.6        | 60.0        | 56.8     | 71.9     |
| TAGPRIME w/ Condition & Relation Priming | 74.6 | 59.8 | 58.3 | 71.9 | 64.7 | 62.4 | 57.3 | 52.4 | 49.9 | 66.3 | 55.2 | 52.6 | 66.3 | 53.1 | 53.6 |

Table 1: Results of end-to-end event extraction. All values are micro F1-score, and we highlight highest scores with boldface. *We reproduce the results using their released code.*
### 3.2 End-to-End Relation Extraction

**Datasets.** We test on two popular used end-to-end relation extraction datasets ACE04 and ACE05 (Doddington et al., 2004), denoted as ACE04-R, and ACE05-R. Various domains are covered in both datasets, such as newsire and online forms, and both datasets consider 7 name entity types and 6 different relations. We follow the same procedure in Zhong and Chen (2021) to preprocess and split the datasets. We refer readers to the paper for more dataset details.

**Baselines.** We compare to the following end-to-end relation extraction models: Table-Sequence (Wang and Lu, 2020), Cascade-SRN (Late fusion) (Wang et al., 2022), and Cascade-SRN (Early fusion) (Wang et al., 2022). Additionally, we consider PURE (Zhong and Chen, 2021), which also takes a pipelined approach to solve end-to-end relation extraction, i.e., first train an named entity recognition models and predict relationship(s) among predicted named entities. To fairly compare with prior works, we use the PURE’s named entity predictions on the test set and feed them to TAGPRIME to perform relational structure extraction.² For all the models, we consider the single sentence setting and use ALBERT-xxlarge-v1 as the pretrained language model.

**Evaluation metrics.** We follow the standard evaluation setting with prior works (Bekoulis et al., 2018; Zhong and Chen, 2021) and use micro F1 measure as the evaluation metric. For named entity recognition, a predicted entity is considered as a correct prediction if its span and the entity type are both correct. We denote the score as “Ent” and report the scores even though it is not our main focus for evaluation. For relation extraction, two evaluation metrics are considered: (1) Rel: a predicted relation is considered as correct when the boundaries of head-entity span and tail-entity span are correct and the predicted relation type is correct; (2) Rel+: a stricter evaluation of Rel, where they additionally required that the entity types of head-entity span and tail-entity must also be correct.

### 3.3 Task-Oriented Semantic Parsing

**Datasets.** We choose MTOP (Li et al., 2021), a multilingual dataset on semantic parsing for task-oriented dialog systems. We specifically consider data from English (en), Spanish (es), French (fr), and German (de).

| Model | ACE05-R | ACE04-R |
|-------|---------|---------|
|       | Ent     | Rel     | Rel+  | Ent     | Rel     | Rel+  |
| Table-Sequence (Wang and Lu, 2020) | 89.5    | 67.6    | 64.3  | 88.6    | 63.3    | 59.6  |
| PURE (Zhong and Chen, 2021) | 89.7    | 69.0    | 65.6  | 88.8    | 64.7    | 60.2  |
| PFN (Yan et al., 2021) | 89.0    | -       | 66.8  | 89.3    | -       | 62.5  |
| Cascade-SRN (Late fusion) (Wang et al., 2022) | 89.4    | -       | 65.9  | -       | -       | -     |
| Cascade-SRN (Early fusion) (Wang et al., 2022) | 89.8    | -       | 67.1  | -       | -       | -     |
| TAGPRIME w/ Condition Priming | 89.6    | 69.7    | 67.3  | 89.0    | 65.2    | 61.6  |
| TAGPRIME w/ Condition & Relation Priming | 89.6    | 70.4    | 68.1  | 89.0    | 66.2    | 62.3  |

Table 2: Results of end-to-end relation extraction. All values are micro F1-score with highest value in bold.

argument classification F1-score (Arg-C), which cares about both offsets as well as the role types. We also report trigger classification F1-score, although this is not our main focus.

**Results.** Table 1 shows the results of end-to-end event extraction on various datasets and languages. Although being simple, TAGPRIME surprisingly has a decent performance and achieves better results than the state-of-the-art models in terms of argument F1-scores. We attribute the good performance to the design of priming, which leverages the semantics of the condition and makes the representations more task-specific. Considering relation priming further improves the results, which again shows the importance of task-specific representations.

²We rerun PURE’s released code to get access to their named entity recognition predictions.
We show the superiority of \( T \) across five different languages. The results suggest that \( T \) is a simple but effective model for general relational structure extraction tasks.

### 4 Analysis

In this section, we study two questions: (1) What is the effectiveness of priming techniques compared to learnable features? (2) Whether the deeply-stacked layers to perform priming is essential?

To answer the first question, we conduct ablation experiments on sequential tagging models using different combinations of learnable feature or/and adding information through the priming mechanism (Section 4.1). For the second question, we propose a simple modification to \( T \) so that we can flexibly control the number of layers to fuse priming information to token representations. Moreover, we show that the modified \( T \) can serve as an efficient approximation of \( T \) (Section 4.2).

#### 4.1 Ablation Study

To conduct the ablation studies on sequential tagging models, we focus on the setting that we alter the choice on how to include the type information of the condition \( c \) and the relation information. Table 4 demonstrates our experimental results.

Comparing the first four cases in Table 4, we can observe that the addition of type features is useful in general, and using our priming mechanism is a more effective way to incorporate conditional information. For models in case 5 to case 8 in Table 4, the decomposition formulation described in Section 2.3 is applied. Comparing case 2 to case 5, we can see that simply applying the relation decompo-
situation formulation for solving relational structure extraction does not lead to major improvements if the way to embed the queried relation is only through learnable features. However, comparing case 3 to case 6 and case 4 to case 7, we show that the relation priming approach make the representation $z_i$ well capture the attribute of the queried relation, thus, better exploit the advantage of the relation decomposition formulation and gain larger improvements. It is worth noting that we conduct a preliminary experiments that use pretrained language models’ representations of the same verbalized token to be the initialization of the learnable type feature embedding, but the method show similar results with random initialization, hence, we stick to random initialization on the models with learnable feature embeddings.

4.2 Depth of Layers to Prime Models

The difference between sequential tagging model with only learnable features and TagPrime is whether the contextualized representation $z_i$ is affected by additional verbalized words or not. The deeply-stacked pretrained language model performs several layers of self-attention to gradually embed the verbalized information into $z_i$. We want to study whether the deeply-stacked layers to influence $z_i$ is essential. We perform studies on TagPrime w/ Condition & Relation Priming model.

We now perform two modifications to TagPrime: (1) we first separate the pretrained language model, which contains $L$ layers, into two halves — one with the first $k$ layers, the other one is the remaining layers, and (2) we copy the first half to be another module. When an input passage is fed into the model. We use the original first half to encode the passage and verbalized condition, and use the copied first half to encode the verbalized relation. Finally, the encoded representations will be fed into the second half layers. With this model design, only $(L - k)$ layers is adapted for the verbalized information to influence the contextualized representation $z_i$, as illustrated in Figure 2. We alter the value of $k$, where when $k = 0$, it represents the TagPrime w/ Condition & Relation Priming, and when $k = L$, it is TagPrime w/ Condition Priming model.

We conduct experiments on the ACE05-E (en) dataset (an event extraction dataset). In order to better analyze the results and isolate the influence from pipelined error, we report the results on the event argument extraction, which assumes gold event triggers are given. The performance are shown in Figure 3a. From the figure, we find that when $k \leq 10$, the performance our modified TagPrime is strong in general and is comparable with TagPrime w/ Condition & Relation Priming. Hence, we conclude the answer toward our second question to — as long as the layers to perform priming is deep to an extent, the benefits of priming mechanism can be anticipated.

One additional benefit of our modified TagPrime is its efficiency during inference. Since the encoding stage of the passage and the verbalized relation is separated, we can fasten up the inference time of our modified TagPrime via parallel encoding. More precisely, our modified TagPrime can aggregate instances that share the same passage and verbalized condition. For those instances, TagPrime only needs to perform the encoding for one time on their major part, and paired with several separated embedded verbalized relation, which could be parallely encoded together.

To compare the efficiency of the model, we benchmark the inference time through performing inference on the whole testing dataset fifty times and calculate the average speed, which is measured by checking how many instances can be processed per second. The results are in Figure 3b. From Figure 3, we observe that the for our modified TagPrime with $k = 10$, its inference speed is around 30% faster than the TagPrime w/ Condition & Relation Priming, but they perform similarly.

---

The string of verbalized relation is usually much shorter than the input passage, hence, in most cases, the major part of the input for an instances is the passage.
5 Related Work

Relational structure extraction has been studied for over a decade (Riedel, 2009) together with its wide range of applications in natural language processing tasks, such as event extraction, relation extraction, task-oriented semantic parsing, semantic role labelling, coreference resolution, etc. The prevalence of its application makes us hard to exhaust them in this short summary, hence, in this section, we mainly focus on related works that experiments with the same tasks with us.

Event extraction. Early works in event extraction mostly consider a pipelined approach (Nguyen and Grishman, 2015; Wang et al., 2019; Yang et al., 2019b; Huang et al., 2022; Du and Cardie, 2020) to deal with event extraction. Some followup works argue that pipelined design leads to error propagation issues and hence propose end-to-end approaches to better capture dependencies between each predictions (Lin et al., 2020; Li et al., 2013; Yang and Mitchell, 2016; Nguyen et al., 2016; Hsu et al., 2022b; Lu et al., 2021; Huang and Peng, 2021). However, recently some empirical studies (Hsu et al., 2022b; Zhong and Chen, 2021; Fincke et al., 2022) also show that given abundant amount of data and properly learned distinct contextualized representation for each pipelined tasks, it is hard to conclude that jointly learned approaches always provide a stronger result. This aligns with our discovery and experimental results — even though we apply a pipelined approach with simple sequential tagging framework on event extraction, with the help of priming to learn more condition-aware contextualized representation, we can still achieve very strong performance on multiple datasets in various languages.

Relation extraction. Ways to solve end-to-end relation extraction can be briefly classify into three categories. The first one is directly perform joint inference on named entities and their relationship(s) (Zheng et al., 2017; Wang and Lu, 2020; Katiyar and Cardie, 2017; Sun et al., 2019; Miwa and Bansal, 2016; Fu et al., 2019). Another is to perform a pipeline that first extract named entities first, and then perform relation classification (Wu and He, 2019; Hsu et al., 2022a; Lyu and Chen, 2021; Peng et al., 2020; Zhou and Chen, 2021; Lu et al., 2022), which assume both the head-entity and tail-entity are given. The other is the cascading approaches that we use in the paper (Wei et al., 2020; Yu et al., 2020; Sun et al., 2019). Among them, many sophisticated graph-based methods emerge rapidly recently (Luan et al., 2019; Wadden et al., 2019; Lin et al., 2020; Sun et al., 2019; Miwa and Bansal, 2016; Fu et al., 2019). In contrast, sequence labelling based approach (Miwa and Bansal, 2016; Zhang et al., 2017) are getting less attention. This highlights one of our motivation to re-examine the merit of sequence tagging, the simple and concise approach, and we show that with our simple modification to the model, this classic framework can serve as a strong baseline.

Task-oriented semantic parsing. Most works on task-oriented semantic parsing focus on intent classification and slot filling tasks (Türe et al., 2010; Gupta et al., 2018; Li et al., 2021; Hakkani-Türe et al., 2016; Zhang et al., 2018; Louvan and Magnini, 2020). Recently, some more advanced neural network based approaches have been proposed, such as MLP-mixer (Fusco et al., 2022) or sequence-to-sequence formulation (Desai et al., 2021). Yet, JointBERT (Chen et al., 2019b) a sequence-tagging based model that is jointly trained model to classify intent and extract slots,
still serve as a widely-used baseline due to its simplicity. Our approach enjoys the same benefit on simplicity with the JointBERT and can further enhanced its performance.

6 Conclusion

In this paper, we present TagPrime, a simple but effective and general sequential tagging model. The key idea is applying priming, a small trick to make the representations condition-aware by appending condition-related string to the input passage. Our experimental results demonstrate that TagPrime has consistent improvements over different relational structure extraction tasks and is general to various languages. Finally, we show that TagPrime can be accelerated without much performance loss by reducing the encoding layers.

References

Giannis Bekoulis, Johannes Deleu, Thomas Demeester, and Chris Develder. 2018. Adversarial training for multi-context joint entity and relation extraction. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP).

Jiaao Chen, Jianshu Chen, and Zhou Yu. 2019a. Incorporating structured commonsense knowledge in story completion. 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.

Qian Chen, Zhu Zhuo, and Wen Wang. 2019b. BERT for joint intent classification and slot filling. arXiv preprint arXiv:1902.10909.

Shrey Desai, Akshat Shrivastava, Alexander Zotov, and Ahmed Aly. 2021. Low-resource task-oriented semantic parsing via intrinsic modeling. arXiv preprint arXiv:2104.07224.

George R. Doddington, Alexis Mitchell, Mark A. Pardybk, Lance A. Ramshaw, Stephanie M. Strassel, and Ralph M. Weischedel. 2004. The automatic content extraction (ACE) program - tasks, data, and evaluation. In Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC).

Xinya Du and Claire Cardie. 2020. Event extraction by answering (almost) natural questions. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP).

Steven Fincke, Shantanu Agarwal, Scott Miller, and Elizabeth Boschee. 2022. Language model priming for cross-lingual event extraction. In The Thirty-Sixth AAAI Conference on Artificial Intelligence, (AAAI).

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.

Francesco Fusco, Damian Pascual, and Peter Staar. 2022. pulp-mixer: an efficient all-mlp architecture for language. arXiv preprint arxiv.2202.04350.

Sonal Gupta, Rushin Shah, Mrinal Mohit, Anuj Kumar, and Mike Lewis. 2018. Semantic parsing for task oriented dialog using hierarchical representations. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP).

Dilek Hakkani-Tür, Gökhan Tür, Asli Celikyilmaz, Yun-Nung Chen, Jianfeng Gao, Li Deng, and Ye-Yi Wang. 2016. Multi-domain joint semantic frame parsing using bi-directional RNN-LSTM. In 17th Annual Conference of the International Speech Communication Association (Interspeech).

Rujun Han, I-Hung Hsu, Jiao Sun, Julia Baylon, Qiang Ning, Dan Roth, and Nanyun Peng. 2021. ESTER: A machine reading comprehension dataset for reasoning about event semantic relations. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021.

I-Hung Hsu, Xiao Guo, Premkumar Natarajan, and Nanyun Peng. 2022a. Discourse-level relation extraction via graph pooling. In The Thirty-Sixth AAAI Conference On Artificial Intelligence Workshop on Deep Learning on Graphs: Method and Applications (DLG-AAAI).

I-Hung Hsu, Kuan-Hao Huang, Elizabeth Boschee, Scott Miller, Prem Natarajan, Kai-Wei Chang, and Nanyun Peng. 2022b. Degree: A data-efficient generation-based event extraction model. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT).

Kuan-Hao Huang, I-Hung Hsu, Premkumar Natarajan, Kai-Wei Chang, and Nanyun Peng. 2022. Multilingual generative language models for zero-shot cross-lingual event argument extraction. In Proceedings
of the 60th Annual Meeting of the Association for Computational Linguistics (ACL).

Kung-Hsiang Huang and Nanyun Peng. 2021. Document-level event extraction with efficient end-to-end learning of cross-event dependencies. In The 3rd Workshop on Narrative Understanding (NAACL 2021).

Arzoo Katiyar and Claire Cardie. 2017. Going out on a limb: Joint extraction of entity mentions and relations without dependency trees. 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.

John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the Eighteenth International Conference on Machine Learning (ICML).

Haoran Li, Abhinav Arora, Shuhui Chen, Ankit Gupta, Sonal Gupta, and Yashar Mehdad. 2021. MTOP: A comprehensive multilingual task-oriented semantic parsing benchmark. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume (EACL).

Qi Li, Heng Ji, and Liang Huang. 2013. Joint event extraction via structured prediction with global features. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL).

Ying Lin, Heng Ji, Fei Huang, and Lingfei Wu. 2020. A joint neural model for information extraction with global features. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL).

Shuman Liu, Hongshen Chen, ZhaD-BLP:conf/acl/LiuFCRYL18ochun Ren, Yang Feng, Qun Liu, and Dawei Yin. 2018. Knowledge diffusion for neural dialogue generation. 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.

Samuel Louvain and Bernardo Magnini. 2020. Recent neural methods on slot filling and intent classification for task-oriented dialogue systems: A survey. In Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020.

Keming Lu, I-Hung Hsu, Wenxuan Zhou, Mingyu Derek Ma, and Muhaow Chen. 2022. Summarization as indirect supervision for relation extraction. arXiv preprint arXiv:2205.09837.

Yaojie Lu, Hongyu Lin, Jin Xu, Xianpei Han, Jialong Tang, Annan Li, Le Sun, Meng Liao, and Shaoyi Chen. 2021. Text2event: Controllable sequence-to-structure generation for end-to-end event extraction. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL/IJCNLP).

Yi Luan, Dave Wadden, Luheng He, Amy Shah, Mari Ostendorf, and Hannaneh Hajishirzi. 2019. A general framework for information extraction using dynamic span graphs. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT).

Shengfei Lyu and Huanhuan Chen. 2021. Relation classification with entity type restriction. In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021.

Jie Ma, Shuai Wang, Rishita Anubhai, Miguel Balles-teros, and Yaser Al-Onaizan. 2020. Resource-enhanced neural model for event argument extraction. In Findings of the Association for Computational Linguistics (EMNLP-Findings).

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.

Thien Huu Nguyen, Kyunghyun Cho, and Ralph Grishman. 2016. Joint event extraction via recurrent neural networks. In The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT).

Thien Huu Nguyen and Ralph Grishman. 2015. Event detection and domain adaptation with convolutional neural networks. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL).

Giovanni Paolini, Ben Athiwaratkun, Jason Krone, Jie Ma, Alessandro Achille, Rishita Anubhai, Cícero Nogueira dos Santos, Bing Xiang, and Stefano Soatto. 2021. Structured prediction as translation between augmented natural languages. In 9th International Conference on Learning Representations (ICLR).

Hao Peng, Tianyu Gao, Xu Han, Yankai Lin, Peng Li, Zhiyuan Liu, Maosong Sun, and Jie Zhou. 2020. Learning from Context or Names? An Empirical Study on Neural Relation Extraction. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP).

Alan Ramponi, Rob van der Goot, Rosario Lombardo, and Barbara Plank. 2020. Biomedical event extraction as sequence labeling. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP).
Sebastian Riedel. 2009. Efficient prediction of relational structure and its application to natural language processing. Ph.D. thesis, University of Edinburgh, UK.

Zhili Song, Ann Bies, Stephanie M. Strassel, Tom Riese, Justin Mott, Joe Ellis, Jonathan Wright, Seth Kulick, Neville Ryan, and Xiaoyi Ma. 2015. From light to rich ERE: annotation of entities, relations, and events. In Proceedings of the 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation, (EVENTS@HLP-NAACL).

Changzhi Sun, Yeyun Gong, Yuanbin Wu, Ming Gong, Daxin Jiang, Man Lan, Shiliang Sun, and Nan Duan. 2019. Joint type inference on entities and relations via graph convolutional networks. In Proceedings of the 57th Conference of the Association for Computational Linguistics (ACL).

Gökhan Tür, Dilek Hakkani-Tür, and Larry P. Heck. 2010. What is left to be understood in atis? In 2010 IEEE Spoken Language Technology Workshop (SLT).

David Wadden, Ulme Wennberg, Yi Luan, and Hannaneh Hajishirzi. 2019. Entity, relation, and event extraction with contextualized span representations. 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).

An Wang, Ao Liu, Hieu Hanh Le, and Haruo Yokota. 2022. Towards effective multi-task interaction for entity-relation extraction: A unified framework with selection recurrent network. arXiv preprint arXiv:2202.07281.

Jue Wang and Wei Lu. 2020. Two are better than one: Joint entity and relation extraction with table-sequence encoders. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP).

Xiaozhi Wang, Ziqi Wang, Xu Han, Zhiyuan Liu, Juanzi Li, Peng Li, Maosong Sun, Jie Zhou, and Xiang Ren. 2019. HMEAE: hierarchical modular event argument extraction. 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).

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).

Shanchan Wu and Yifan He. 2019. Enriching pre-trained language model with entity information for relation classification. In Proceedings of the 28th ACM international conference on information and knowledge management, pages 2361–2364.

Zhiheng Yan, Chong Zhang, Jinlan Fu, Qi Zhang, and Zhongyu Wei. 2021. A partition filter network for joint entity and relation extraction. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP).

Bishan Yang and Tom M. Mitchell. 2016. Joint extraction of events and entities within a document context. In The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT).

Sen Yang, Dawei Feng, Linbo Qiao, Zhigang Kan, and Dongsheng Li. 2019a. Exploring pre-trained language models for event extraction and generation. In Proceedings of the 57th Conference of the Association for Computational Linguistics (ACL).

Sen Yang, Dawei Feng, Linbo Qiao, Zhigang Kan, and Dongsheng Li. 2019b. Exploring pre-trained language models for event extraction and generation. In Proceedings of the 57th Conference of the Association for Computational Linguistics (ACL).

Michihiro Yasunaga, Hongyu Ren, Antoine Bosselut, Percy Liang, and Jure Leskovec. 2021. QA-GNN: Reasoning with language models and knowledge graphs for question answering. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.

Bowen Yu, Zhenyu Zhang, Xiaobo Shu, Tingwen Liu, Yubin Wang, Bin Wang, and Sujian Li. 2020. Joint extraction of entities and relations based on a novel decomposition strategy. In 24th European Conference on Artificial Intelligence (ECAI).

Dongjie Zhang, Zheng Fang, Yanan Cao, Yanheng Liu, Xiaojun Chen, and Jianlong Tan. 2018. Attention-based RNN model for joint extraction of intent and word slot based on a tagging strategy. In Artificial Neural Networks and Machine Learning (ICANN).

Hongming Zhang, Xin Liu, Haojie Pan, Yangqiu Song, and Cane Wing-Ki Leung. 2020. ASER: A large-scale eventuality knowledge graph. In The Web Conference 2020 (WWW).

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.

Hao Zheng and Mirella Lapata. 2022. Disentangled sequence to sequence learning for compositional generalization. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL).

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.

Zexuan Zhong and Danqi Chen. 2021. A frustratingly easy approach for entity and relation extraction. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT).

Wenxuan Zhou and Muhao Chen. 2021. An improved baseline for sentence-level relation extraction. arXiv preprint arXiv:2102.01373.