Abstract

Events are inter-related in documents. Motivated by the one-sense-per-discourse theory, we hypothesize that a participant tends to play consistent roles across multiple events in the same document. However recent work on document-level event argument extraction models each individual event in isolation and therefore causes inconsistency among extracted arguments across events, which will further cause discrepancy for downstream applications such as event knowledge base population, question answering, and hypothesis generation. In this work, we formulate event argument consistency as the constraints from event-event relations under the document-level setting. To improve consistency we introduce the Event-Aware Argument Extraction (EA²E) model with augmented context for training and inference. Experiment results on WIKIEVENTS and ACE2005 datasets demonstrate the effectiveness of EA²E compared to baseline methods.

1 Introduction

Document-level Event Argument Extraction aims at identifying arguments and their roles for multiple events in a document. It is a practically more useful but more challenging task than sentence-level Argument Extraction (Nguyen et al., 2016; Wadden et al., 2019; Lin et al., 2020) because in a typical long input document events usually scatter across multiple sentences and are inherently connected.

Multiple events in one document are usually inter-connected, and thus the arguments are under certain consistency constraints. In Figure 1, the roles of the shared argument Ahmad Khan Rahami in multiple events are constrained because the Attacker in the DetonateExplode event is likely to be the IdentifiedRole in IdentifyCategorize event, the

---

*Equal contribution.

1Our code is released at https://github.com/ZQS1943/DOCIE
et al., 2019), the inconsistency phenomenon addressed in previous research focuses on factual consistency instead of self-contained consistency as in document-level argument extraction. We approach this problem with inspiration from human behavior: while reading, humans subconsciously infer the event-event relations and correctly identify the event arguments under the perceived constraints. Therefore, we refer consistent argument extraction to applying the underlying Event-Event Relations as constraints in multi-event argument extraction.

An intuitive solution to improve consistency is to incorporate explicit Event-Event Relations into the extraction process as additional input. However, the underlying event-event relations are hard to identify and classify due to the lack of reliable resources as supervision signals, especially when the arguments are unknown. In addition, precise event-event relations may not be necessary for argument extraction when the implicit connections can already well support argument extraction.

To avoid explicit modeling of event-event relations, we label the arguments of other events in the context as an implication of event-event relations. We propose an Event-Aware Argument Extraction (EA²E) model, which incorporates alignment-enhanced training and iterative inference. When extracting arguments, the context can be self-augmented by tagging the argument labels of other events. Alignment-enhanced training implicitly introduces event awareness by pulling close the argument representation distributions under regular context and augmented context, where ground-truth argument labels of neighboring events are labeled. Iterative inference explicitly encourages event awareness by augmenting the context with the extracted arguments from the last inference iteration. The advantage of this method is that no predefined Event-Event Relation is required, nor event schema. Our experiments on WIKIEVENTS and ACE2005 datasets show that our EA²E model brings improvement against previous methods.

2 Event-Aware Argument Extraction

Motivated by the observation that introducing event-event relations benefits the consistency of event argument extraction, we propose to incorporate implicit event-event relations with an Event-Aware Argument Extraction (EA²E) model. As shown in Figure 2, EA²E contains alignment-enhanced training and iterative inference with self-augmented context. When extracting the arguments for a target event, the context is augmented by labeling the arguments from neighboring events. During training, an auxiliary training loss pulls close the event argument representations under the regular context and self-augmented context. During inference, iterative inference encourages event awareness by using the extraction arguments from the last inference iteration as inputs.

2.1 Base Encoder-Decoder Model

Following (Li et al., 2021), we formulate event argument extraction as a conditional generation task under the assumption that there exists a pre-defined event ontology describing each event type with an unfilled template with argument placeholders. For example, the template for Attack events is <arg> detonated or exploded <arg> explosive device using <arg> to attack <arg> target at <arg> place. Formally, given a document context c and an event trigger x with template t, the task is to extract a set of arguments y = {a_1, a_2, ..., a_n}, where each a_i corresponds to a role predefined in the ontology.

We base our model on BART (Lewis et al., 2020), an encoder-decoder pretrained model. The input sequence is the concatenation of the document context and an event template, constructed as <s> template </s> context </s>. The output is a filled-in template, where the tokens are all selected from the input context or template.

The model parameter θ is trained by minimizing the argument extraction loss, the conditional probability over all instances:

\[ L_E = - \sum \log p(y|x, t, c) \]  

2.2 Self-augmented Context

We refer event awareness as the implication of event-event relations and reach this goal by labeling the arguments of other neighboring events in the context. Given the arguments of the neighboring events \( \{ j \in N_i \} \), which have small token-wise distances to the target event i, we augment the regular context c by labeling them with <tag>:

\[ c_i' = \mu(c_i, \{ y_j \}), \]  

where \( \mu \) is the tagging operation, and \( y_j \) is the arguments of event j.

For example in Figure 1, when extracting arguments for the target event bombing (Event 1), the augmented context is “The Saturday night’s <trg>
bombing <trg> in New York City, wounded 29 people. <tag> Prosecutor <tag> Prosecutors have charged 28-year-old <tag> Defendant <tag> Ahmad Khan Rahami...", where the two tags highlight the arguments of Event 4.

2.3 Alignment-enhanced Training

An encoder is considered consistent when it is able to understand and encode the underlying relation between events into the text representations. Therefore, we propose to enhance the encoder with an auxiliary training loss \( \mathcal{L}_T \) that pulls close the argument representation distributions under regular context \( c \) and under augmented context \( c' \). During training, \( c' \) is constructed by tagging the ground-truth arguments of neighboring events.

\[
\mathcal{L}_T = \sum \| p(a|c), p(a|c') \|_2 
\]

The final training loss is a weighted sum of argument extraction losses ( \( \mathcal{L}_E \) for regular context \( c \) and \( \mathcal{L}_{E'} \) for augmented context \( c' \)) and alignment-enhanced loss (\( \mathcal{L}_T \)) with weights \( \alpha \) and \( \beta \):

\[
\mathcal{L} = \mathcal{L}_E + \alpha \mathcal{L}_{E'} + \beta \mathcal{L}_T 
\]

2.4 Iterative Inference

Iterative inference explicitly introduces event awareness by utilizing extracted results in multiple inference iterations. In the first iteration, for each target event trigger \( i \), the model obtains the predicted results \( y_1^i \) given the regular context \( c_1^i \). For the \( k \)-th iteration of inference, for each event trigger \( i \) the context \( c_k^i \) is augmented by labeling the extracted arguments \( \{y_j^{k-1}\} \) of neighboring events \( \{j \in \mathcal{N}_i \} \) from the \((k-1)\)-th iteration.

\[
c_k^i = \mu(c, \{y_j^{k-1}\}) 
\]

3 Experiments

3.1 Datasets and Baselines

We evaluate our proposed method on WIKIEVENTS (Li et al., 2021) dataset and ACE 2005 dataset\(^2\). Following previous work (Li et al., 2013), we consider an argument span to be correctly identified when its offsets match any of the reference arguments of the current event (i.e., Argument Identification), and to be correctly classified when its role matches (i.e., Argument Classification). We report the argument extraction performance in terms of Head Word F1 and Coreferential Mention F1. For the latter, full credit will be given when the extracted argument is coreferential with the gold-standard argument.

We compare EA\(^2\)E with document-level BART-Gen (Li et al., 2021), sentence-level ONEIE (Lin et al., 2020) and BERT-CRF (Shi and Lin, 2019).

3.2 Implementation Details

We implement our models with Huggingface (Wolf et al., 2019). We train each model for 4 epochs with a batch size of 4 for baselines and 2 for EA\(^2\)E. The model is optimized with the Adam optimizer with a learning rate of 3e-5, \( \alpha = 1 \) and \( \beta = 0.5 \). We define event neighborhood as trigger distance less than 40 tokens. For inference, the maximum number of iterations is 3. For numerical consistency, all experiment results are averaged across three random runs. The hyper-parameters are selected by grid search based on model performance on development set.\(^3\) On average it takes approximately

\(^2\)https://www.ldc.upenn.edu/collaborations/past-projects/ace

\(^3\) \( \beta \) is chosen from \{0.1, 0.5, 1\}, the trigger distance is chosen from \{20, 40, 60, 80, 100\}, and the learning rate is chosen in \{3e−5, 5e−5 \}.
3.3 Results and Analysis

Table 1 and Table 2 show that our proposed EA^2E consistently performs better than strong baseline methods across datasets and evaluation metrics. In general document-level methods have better performance, especially in terms of recall, because sentence-level methods are more likely to miss cross-sentence arguments.

Alignment-enhanced training brings a significant improvement over BART-Gen but comes with higher training costs since the inputs are doubled. Iterative inference brings unstable improvement. Figure 3 shows that more iterations brings higher performance only to a certain range. Since the only differences among iterations are their inputs, we conclude that labeling the arguments of other events helps the model extract the arguments of the current event. The upper bound of this improvement is limited by the error propagation in the augmented context.

Qualitative Analysis Table 3 presents some representative examples. BART-Gen incorrectly assigns Tsarnaev to the Target role, and police to the attacker role in the first example. It also misses the killer brothers in the second example and the attacker Laden in the third example. The second example shows the advantage of the Alignment-enhanced Training component in EA^2E, which helps extract the killer argument. The third example shows how Iterative Inference works with the augmented input: The tagged attacker in the neighboring bombing event is also the attacker in the target attack event.

Remaining Challenges Though effective, iterative inference may propagate errors among iterations. In addition, the success of event awareness relies on the assumption that events in a neighborhood defined by trigger distance are inter-related to the target event. However, this assumption is not always held true in the case that distant events bring redundant information. It is not necessarily hurting the information but it brings noise by incorrectly implying the relations between the distant events and the target event.
Table 3: Examples of extracted arguments from BART-Gen, EA²E w/o II, and EA²E. We label target event mention with bold, gold arguments with underlines, correct but not annotated arguments with waves, and incorrect arguments with red. In the third example we also present the augmented input for Iterative Inference, in which the arguments of the bombing event are tagged.

4 Related Work

Sentence-level argument extraction approaches, where the event trigger and its arguments are usually located within a single sentence, cannot handle the cross-sentence trigger-argument distribution and the existence of multiple events within one document. Though recent attempts on document-level argument extraction have gone beyond sentence boundaries, they either focus on one-event-per-document setting, or model each event independently. The most related work to ours is (Li et al., 2021), which formulates the task as conditional generation following event templates and extracts arguments for each event independently, while our work focuses on the consistency among arguments for different events. The methods in (Du et al., 2021) and (Du and Cardie, 2020) are designed for Role-filler Entity Extraction (REE) task under the assumption that one generic template is produced for each document, while our work focuses on extracting arguments for multiple events for each document. (Yang et al., 2021) introduces Parallel Prediction Network that generates all possible events in parallel based on the document-aware representations, while we adopt a generative framework. (Xu et al., 2021) model the whole document as graphs and capture the interdependency among events by tracking the extracted events with a global memory, while we introduce event awareness for interdependency without external memory modules.

5 Conclusion

We introduce Event-Aware Argument Extraction (EA²E) model to improve self-contained consistency in document-level event argument extraction. We conclude that in iterative inference brings higher performance only to a certain range of iterations and alignment-enhanced training brings significant improvement with costs.

Acknowledgement

We thank the anonymous reviewers helpful suggestions. This research is based upon work supported by U.S. DARPA AIDA Program No. FA8750-18-2-0014, U.S. DARPA KAIROS Program No, FA8750-19-2-1004, and DARPA INCAS Program No. HR001121C0165. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of DARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

References


