DESED: Dialogue-based Explanation for Sentence-level Event Detection

Yinyi Wei[1*†], Shuaipeng Liu[2*‡], Jianwei Lv[2], Xiangyu Xi[2], Hailei Yan[2], Wei Ye[3‡], Tong Mo[1], Fan Yang[2], Guanglu Wan[2]

[1] Peking University

[2] Meituan Group, Beijing, China

[3] National Engineering Research Center for Software Engineering, Peking University

October, 2022

1. Introduction
2. Methodology
3. Results and Insights
4. Conclusion and Future Work

Event Detection

- Definition: Event detection (ED) is a crucial task in information extraction, which aims to identify event triggers (words or phrases that indicate events) and classify triggers into predefined event types [1].
- Example:

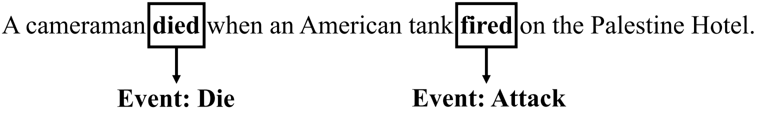

Figure 1: A classic example of event detection.

[1] According to the definition of events in the annotation guideline designed for the ACE2005 dataset

Motivation

- Sentence semantics enhancement.
  - Multi-task Learning: Leveraging annotations from other information extraction tasks.
  - Prompt-based Learning: Exploiting PLMs by retrieving similar instances or adding manual definitions of labels, or by converting information extraction tasks into slot-filling tasks.
- MRC-based methods for event detection.

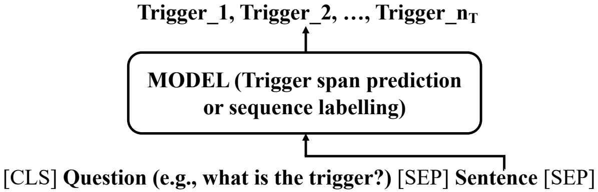

Figure 2: MRC-based methods for event detection.

Our Solution

- We propose to use generative models to generate contextual information for a sentence.
- In order to obtain consistent information with the original sentence, the contexts are generated in the form of a dialogue. We refer the generated dialogue for an event description to dialogue-based explanation.
- We propose three conceptually simple methods to generate dialogue-based explanation and design hybrid attention mechanisms to exploit dialogue information.

Dialogue Generation

Three methods to generate dialogues.

- Direct generation (for casual dialogues).
- Generation with a prompt (for focused dialogues).
- Further training and generation (for domain-specific dialogues).

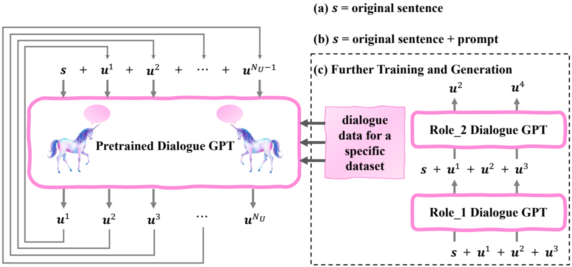

Figure 3: Illustration of dialogue generation methods and an example of dialogue generation with further training on two roles.

Dialogue Generation

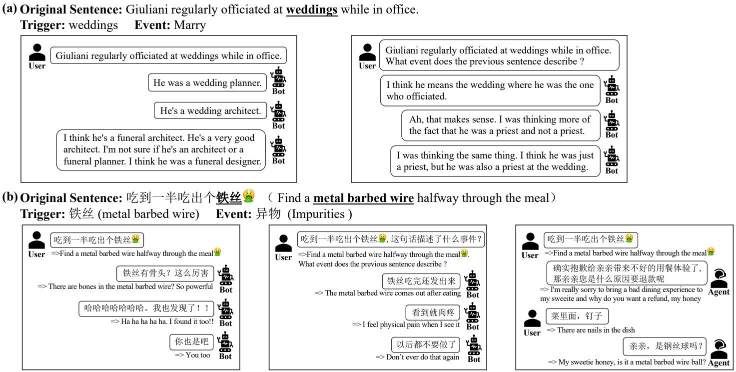

Figure 4: Examples of dialogue generation for a specific sentence with three methods: (1) Direct generation; (2) Generation with a prompt; (3) Further training and generation. Figure (a) shows the dialogue generation using method (1)(2) on ACE05-E+. Figure (b) shows the dialogue generation using method (1)(2)(3) on FOSAED-R.

1 Introduction

2 Methodology

Dialogue Generation

Exploitation of Dialogue Information

3 Results and Insights

4 Conclusion and Future Work

Exploitation of Dialogue Information

Event detection in this work is based on sequence labelling using *BIO* tagging format.

- Token-level attention.
- Utterance-level attention.
- Hybrid attention.

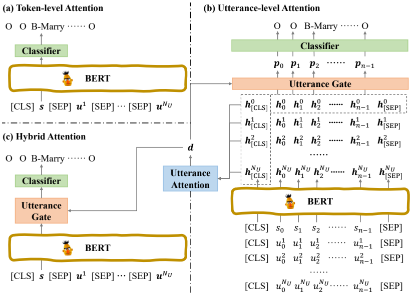

Figure 5: Different attention mechanisms of exploiting dialogue information

Exploitation of Dialogue Information

- Some notations: The original sentence: $\boldsymbol{s}$. Generated utterances $\boldsymbol{u}^1, \ldots \boldsymbol{u}^{N_U}$. Representation of $\boldsymbol{s}$: $\boldsymbol{h}^0$. Representations of utterances: $\boldsymbol{h}^1, \ldots, \boldsymbol{h}^{N_U}$.
- Token-level attention: Taking advantage of the self-attention mechanism in models like BERT. Concatenating the original sentence and generated utterances to form a combined input, $\boldsymbol{c} = \boldsymbol{s}\ [\mathrm{SEP}]\ \boldsymbol{u}^1\ [\mathrm{SEP}]\ \ldots\ [\mathrm{SEP}]\ \boldsymbol{u}^{N_U}$. After obtaining contextual representations of $\boldsymbol{c}$, the token representations corresponding to $\boldsymbol{s}$ are classified into specific tags by a classifier.

Exploitation of Dialogue Information

- Utterance-level attention:
  - Obtaining a dialogue state $\boldsymbol{d}$:

$$\boldsymbol{d} = \sum_{i=0}^{N_U} \alpha_i \boldsymbol{h}^i_{\mathrm{[CLS]}}, \ \ \boldsymbol{d} \in \mathbb{R}^D \tag{1}$$

$$\alpha_i = \frac{\exp(s_i)}{\sum_{j=0}^{N_U} \exp(s_j)} \tag{2}$$

$$s_i = \tanh\left(\boldsymbol{h}^0_{\mathrm{[CLS]}} \cdot (\boldsymbol{W}_a \cdot (\boldsymbol{h}^i_{\mathrm{[CLS]}})^T + \boldsymbol{b}_a)\right) \tag{3}$$

  - Fusing $\boldsymbol{d}$ into token representations of $\boldsymbol{s}$:

$$\boldsymbol{p}_i = \boldsymbol{h}^0_i \parallel \boldsymbol{f}_i \tag{4}$$

$$\boldsymbol{f}_i = \theta_i \circ \boldsymbol{h}^0_i + (1 - \theta_i) \circ \boldsymbol{d} \tag{5}$$

$$\theta_i = \mathrm{sigmoid}((\boldsymbol{h}^0_i \parallel \boldsymbol{d}) \cdot \boldsymbol{W}_g + b_g) \tag{6}$$

- Hybrid attention: Cover both the token-level attention and the utterance-level attention.

Datasets and Evaluation Metrics

- ACE2005: A collection of documents from a diversity of domains, the most widely used dataset for event extraction. For data split and preprocessing, we follow ONEIE (2020), which adds back pronouns and multi-token triggers. The version is denoted as ACE05-E^{+}.
- FOSAED: FOSAED (Food Safety on User Reviews for Event Detection) is a real-world Chinese event detection dataset, consisting of sentence-level user reviews in the domain of food safety based on a leading e-commerce platform for food service. To support further training, a number of unlabelled user-agent conversations are collected, which are also in the domain of food safety.

Datasets and Evaluation Metrics

- Statistics of datasets:

Form	#Docs	#Sents
Labelled User Reviews	4,226	4,226
Unlabelled Conversations	7,155	309,295

Table 1: Statistics of FOSAED. We show the number of documents and sentences for different forms of data.

Dataset	Split	#Sents	#Events
ACE05-E$^+$	Train	19,216	4,419
	Dev	901	468
	Test	676	424
FOSAED-R	Train	3,380	3,893
	Dev	423	494
	Test	423	512

Table 2: Dataset statistics. We show the number of sentences and events for different splits.

- Evaluation metrics: F1-scores of Trig-I and Trig-C.
  - Trig-I: A trigger is correctly identified if its offset match any of the gold triggers.
  - Trig-C: The span of the trigger is correctly identified and its event type is also correctly classified.

Results

Main Results:

Category	Methods		ACE05-E^{+} Trig-I	ACE05-E^{+} Trig-C	FOSAED-R Trig-I	FOSAED-R Trig-C
Basic	BiLSTM+CRF		72.9	69.3	71.5	70.8
	DMBERT		73.5	69.5	72.8	71.4
	BERT		73.4	70.5	73.6	71.5
MRC-based	BERT_QA_TRIGGER		74.6	71.5	72.9	71.8
Multi-task	OneIE*		75.6	72.8	-	-
	FourIE*		76.7	73.3	-	-
Prompt-based	Text2Event*		-	71.8	-	-
	DEGREE*		76.7	72.7	-	-
	PILED*		-	73.4	-	-
Multi-task and Prompt-based	TANL*		71.5	68.4	-	-
	UIE*		-	73.4	-	-
Dialogue-based Explanation	DESED	Direct Generation	76.2	72.3	**75.8**	74.3
		Generation with a Prompt	**76.9**	**73.5**	**75.8**	74.3
		Further Training	-	-	75.6	**74.4**

Table 3: Experimental results of sentence-level event detection on ACE05-E^{+} and FOSAED-R (F1-score, %). The best results are in boldface. * indicates results cited from the original paper.

Attention Mechanisms:

Generation	Att	ACE05-E^{+} Trig-I	ACE05-E^{+} Trig-C	FOSAED-R Trig-I	FOSAED-R Trig-C
Direct	T	74.6	71.6	75.8	74.3
	U	74.9	71.8	75.0	73.4
	H	76.2	72.3	75.7	73.8
Prompt	T	75.2	72.3	75.1	73.7
	U	76.2	73.5	75.8	74.3
	H	76.9	73.3	74.3	72.9
Further	T	-	-	74.3	72.9
	U	-	-	74.9	73.5
	H	-	-	75.6	74.4

Table 4: Different attention mechanisms of DESED on ACE05-E^{+} and FOSAED-R (F1-score, %). T, U and H denote token-level, utterance-level and hybrid attention mechanism respectively.

1. Introduction
2. Methodology
3. Results and Insights
   - Experimental Setup
   - Experimental Results
   - Analytical Insights
4. Conclusion and Future Work

Exploration of Generated Dialogues

Three features to quantify the consistency of generated dialogues:

- Definition of a consistent dialogue: if a sentence contains events, the generated dialogue should contain all events in this sentence; if a sentence has no events, the generated dialogue would also has no events.
- $p(\text{consistent}) = \frac{\text{number of consistent dialogues}}{\text{number of original sentences}}$
- $p(\text{event}) = \frac{\text{number of consistent dialogues having all events}}{\text{number of original sentences with events}}$
- $p(\text{no-event}) = \frac{\text{number of consistent dialogues having no events}}{\text{number of original sentences without events}}$
- A BERT model is employed to detect events in the generated dialogues.

Exploration of Generated Dialogues

- Exploration of different dialogue generation methods:

Generation	Indicator	ACE05-E$^+$	FOSAED-R
Direct	Length	54.6	62.1
	p(event)	11.9	19.5
	p(no-event)	93.2	72.2
	p(consistent)	58.0	30.7
Prompt_3	Length	60.9	79.2
	p(event)	21.2	24.0
	p(no-event)	80.4	71.1
	p(consistent)	54.7	34.0
Further	Length	-	134.6
	p(event)	-	41.1
	p(no-event)	-	26.7
	p(consistent)	-	38.1

Table 6: Heuristic exploration of different dialogue generation methods based on BERT and four indicators. The number of generated utterances is set to five.

Comparison Between Dialogues and Narrative Contexts

- Narrative Contexts vs Dialogues

Generation	Indicator	Context	Dialogue
Direct	Trig-C	70.6	70.9
	p(event)	22.5	11.9
	p(no-event)	50.4	93.2
	p(consistent)	38.3	58.0
Prompt_3	Trig-C	70.6	71.1
	p(event)	23.5	21.2
	p(no-event)	49.1	80.4
	p(consistent)	38.0	54.7

Table 7: Experiments of using plain narrative contexts or dialogues as additional information on ACE05-E^+. Five generated utterances are used for dialogue, and the number of generated tokens is set to the average token length of the five utterances for narrative contexts.

Conclusion and Future Work

- We propose dialogue-based explanation to enhance sentence semantics for sentence-level event detection.
- We propose three conceptually simple methods to generate dialogues for given original sentences, which concentrate on casual dialogues, focused dialogues and domain-specific dialogues respectively. To make effective use of generated dialogues, we design hybrid attention mechanisms at different levels of granularity.
- In the future, we are interested in generating dialogue-based explanation in a more controllable way and extending dialogue-based explanation to other tasks.

Thanks!