Query and Extract: Refining Event Extraction as Type-oriented Binary Decoding

Sijia Wang[1], **Mo Yu**[2], **Shiyu Chang**[3], **Lichao Sun**[4], **Lifu Huang**[1]
[1]Virginia Tech [2]WeChat AI [3]University of California Santa Barbara [4]Lehigh University
[1]{sijiawang,lifuh}@vt.edu, [2]moyumyu@tencent.com
[3]chang87@ucsb.edu, [4]lis221@lehigh.edu

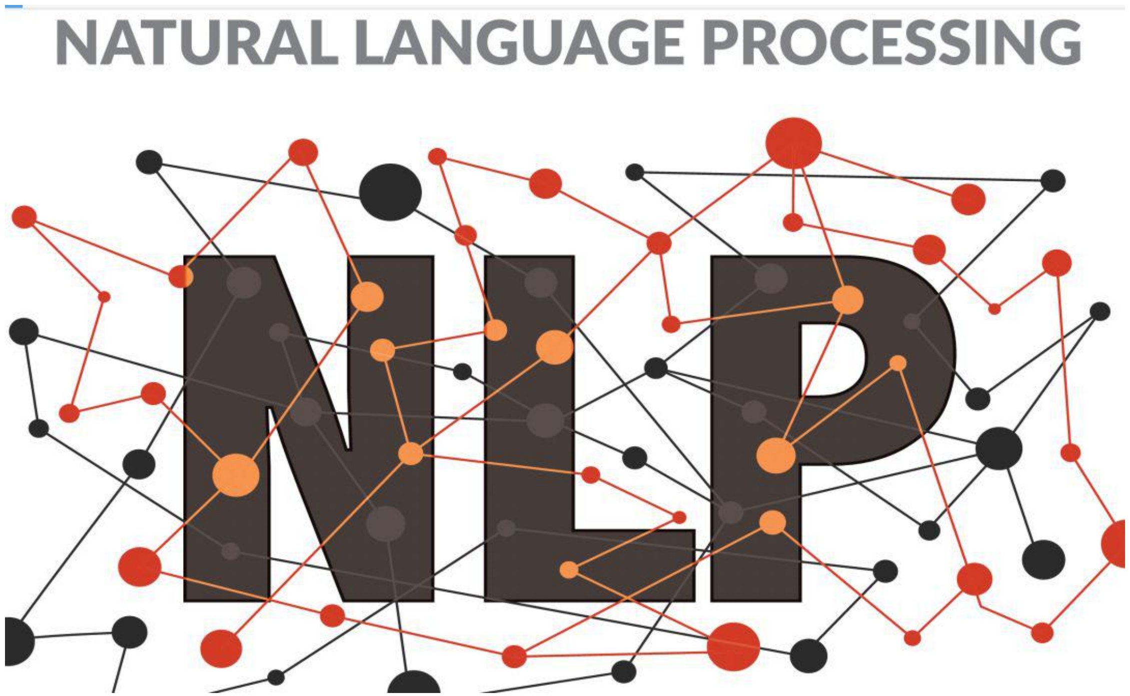
2022. 06. 22 • ChongQing

2022_ACL

Reported by Yidan Liu

Code:https://github.com/VT-NLP/Event_Query_Extract

Leibniz-Institut für Sozialwissenschaften



1.Introduction

2.Method

3.Experiments

Leibniz-Institut für Sozialwissenschaften

Introduction

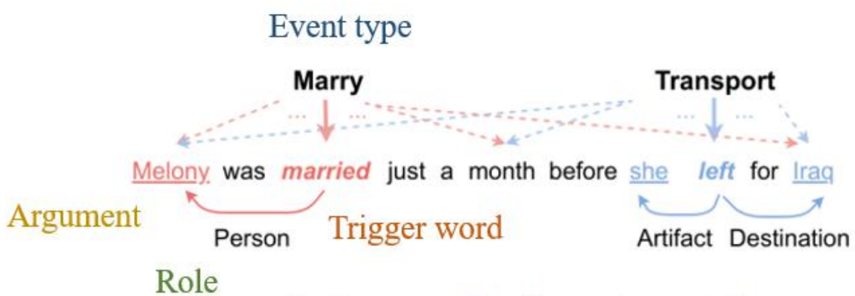


Figure 1: An example of event annotation.

Event extraction is typically modeled as a multi-class classification problem, These approaches are usually limited to a set of pre-defined types.

We propose a novel event extraction framework that uses event types and argument roles as natural language queries to extract candidate triggers and arguments from the input text.

Method

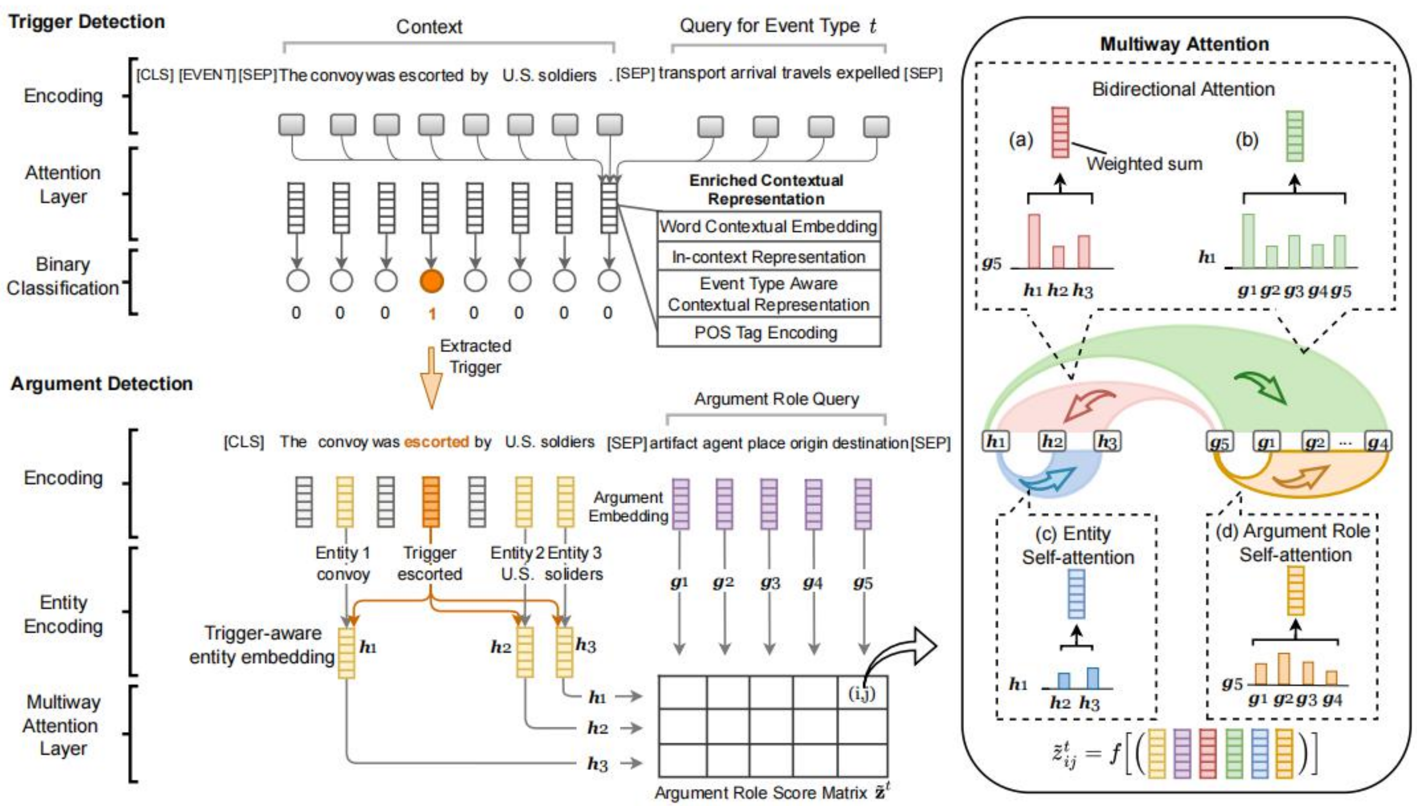
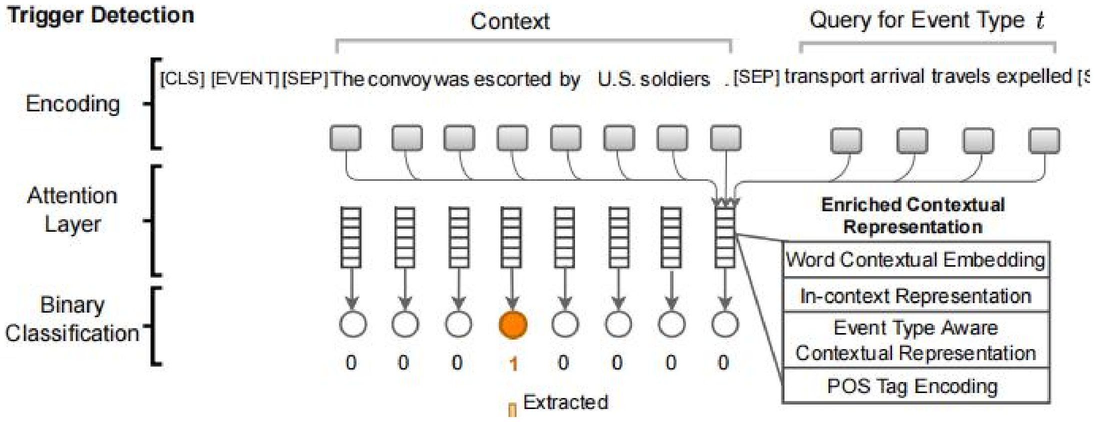


Figure 2: Architecture overview. Each cell in Argument Role Score Matrix indicates the probabilities of an entity being labeled with an argument role. The arrows in Multiway Attention module show four attention mechanisms: (a) entity to argument roles, (b) argument role to entities, (c) entity to entities, (d) argument role to argument roles.

Method



Specifically, for each event type t, we collect a set of annotated triggers from the training examples. For each unique trigger word, we compute its frequency from the whole training dataset as f_o and its frequency of being tagged as an event trigger of type t as f_t, and then obtain a probability f_t/f_o, which will be used to sort all the annotated triggers for event type t. We select the top-K[4] ranked words as prototype triggers $\{\tau_1, \tau_2, \ldots, \tau_K\}$.

Given an input sentence $W = \{w_1, w_2, \ldots, w_N\}$, $T = \{t, \tau_1^t, \tau_2^t, \ldots, \tau_K^t\}$

[CLS][EVENT][SEP] w_1 ... w_N [SEP]
t τ_1^t ... τ_K^t [SEP]

$$\boldsymbol{A}_i^T = \frac{1}{T}\sum_{j=1}^{|T|} \alpha_{ij} \cdot \boldsymbol{T}_j \ ,$$

$$\alpha_{ij} = \cos(\boldsymbol{w}_i, \ \boldsymbol{T}_j) \ ,$$

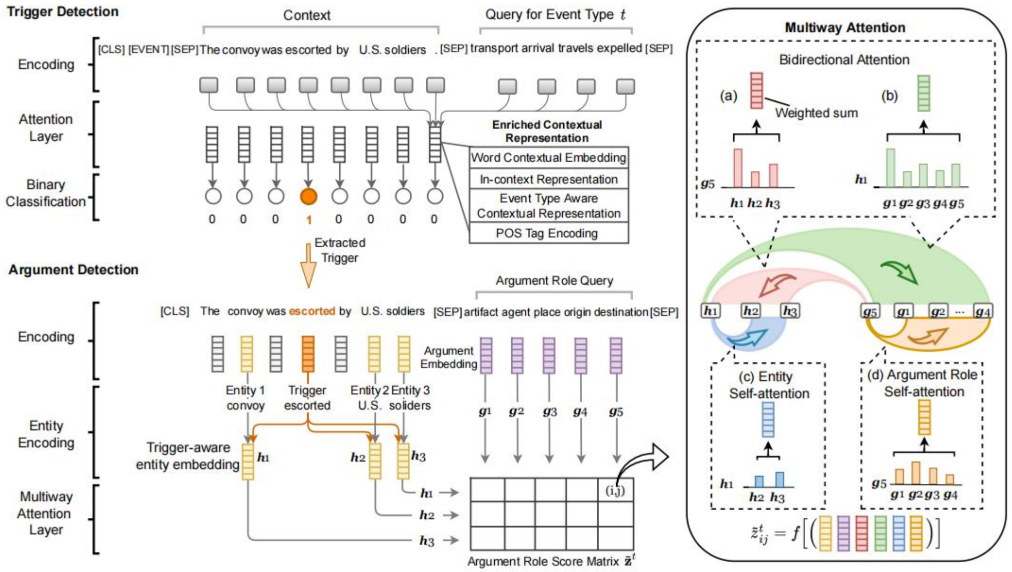
$$\boldsymbol{A}_i^W = \frac{1}{N}\sum_{j=1}^{N} \tilde{\alpha}_{ij} \cdot \boldsymbol{w}_j \ ,$$

$$\tilde{\alpha}_{ij} = \rho(\boldsymbol{w}_i, \ \boldsymbol{w}_j) \ ,$$

$$\tilde{\boldsymbol{y}}_i^t = \boldsymbol{U}_o \cdot ([\boldsymbol{w}_i; \ \boldsymbol{A}_i^W; \ \boldsymbol{A}_i^T; \boldsymbol{P}_i]) \ ,$$

$$\mathcal{L}_1 = -\frac{1}{|\mathcal{T}||\mathcal{N}|}\sum_{t\in\mathcal{T}}\sum_{i=1}^{|\mathcal{N}|} \boldsymbol{y}_i^t \cdot \log \tilde{\boldsymbol{y}}_i^t \ ,$$

Method



Context Encoding:

$$[\text{CLS}]\ w_1\ w_2\ \dots\ w_N\ [\text{SEP}]\ g_1^t\ g_2^t\ \dots\ g_D^t\ [\text{SEP}]$$

$$\tilde{\boldsymbol{W}} = \{\tilde{\boldsymbol{w}}_0, \tilde{\boldsymbol{w}}_2, \dots, \tilde{\boldsymbol{w}}_N\}, \qquad \boldsymbol{G}^t = \{\boldsymbol{g}_0^t, \boldsymbol{g}_1^t, \dots, \boldsymbol{g}_D^t, \boldsymbol{g}_{[\text{Other}]}^t\}.$$

Multiway Attention:

$$\boldsymbol{E} = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_M\}. \qquad \boldsymbol{h}_i = \boldsymbol{U}_h \cdot ([\boldsymbol{e}_i;\ \boldsymbol{r};\ \boldsymbol{e}_i \circ \boldsymbol{r}]),$$

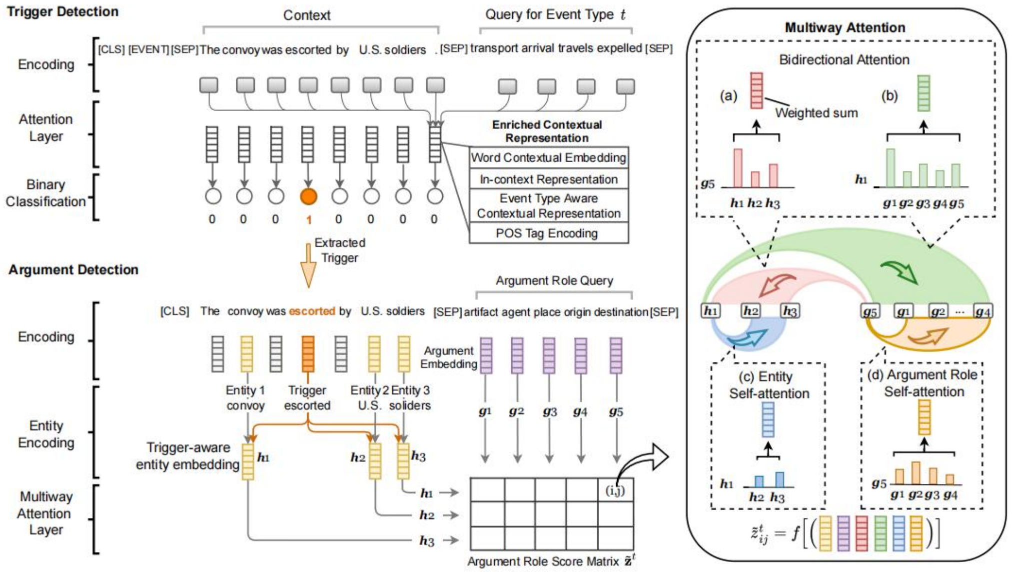
$$S_{ij} = \frac{1}{\sqrt{d}}\sigma(\boldsymbol{h}_i,\ \boldsymbol{g}_j^t),$$

$$\boldsymbol{A}_i^{e2g} = \sum_{j=1}^{D} \boldsymbol{S}_{ij} \cdot \boldsymbol{g}_j^t,$$

$$\boldsymbol{A}_j^{g2e} = \sum_{i=1}^{M} \boldsymbol{S}_{ij} \cdot \boldsymbol{h}_i .$$

Method

Multiway Attention:



$$\mu_{ij} = \frac{1}{\sqrt{d}}\sigma(\boldsymbol{h}_i,\ \boldsymbol{h}_j)\ , \quad \tilde{\boldsymbol{\mu}}_i = \mathrm{Softmax}(\boldsymbol{\mu}_i)\ ,$$

$$\boldsymbol{A}_i^{e2e} = \sum_{j=1}^{M} \tilde{\mu}_{ij} \cdot \boldsymbol{h}_j\ .$$

$$v_{jk} = \frac{1}{\sqrt{d}}\sigma(\boldsymbol{g}_j^t,\ \boldsymbol{g}_k^t)\ , \quad \tilde{\boldsymbol{v}}_j = \mathrm{Softmax}(\boldsymbol{v}_j)\ ,$$

$$\boldsymbol{A}_j^{g2g} = \sum_{k=1}^{D} \tilde{v}_{jk} \cdot \boldsymbol{g}_k^t\ .$$

$$\tilde{\boldsymbol{z}}_{ij}^t = \boldsymbol{U}_a \cdot ([\boldsymbol{h}_i;\ \boldsymbol{g}_j^t;\ \boldsymbol{A}_i^{e2g};\ \boldsymbol{A}_j^{g2e};\ \boldsymbol{A}_i^{e2e};\ \boldsymbol{A}_j^{g2g}]),$$

$$\mathcal{L}_2 = -\frac{1}{|\mathcal{A}||\mathcal{E}|} \sum_{j=1}^{|\mathcal{A}|} \sum_{i=1}^{|\mathcal{E}|} \boldsymbol{z}_{ij} \log \tilde{\boldsymbol{z}}_{ij}\ ,$$

Experiment

Model	ACE05-E$^+$		ERE-EN	
	Trigger Ext.	Argument Ext.	Trigger Ext.	Argument Ext.
DYGIE++ (Wadden et al., 2019)	67.3*	42.7*	-	-
BERT_QA_Arg (Du and Cardie, 2020)	70.6*	48.3*	57.0	39.2
OneIE (Lin et al., 2020)	72.8	54.8	57.0	46.5
Text2Event (Lu et al., 2021)	71.8	54.4	59.4	48.3
FourIE (Nguyen et al., 2021)	73.3	**57.5**	57.9	48.6
Our Approach	**73.6** (0.2)	55.1 (0.5)	**60.4** (0.3)	**50.4** (0.3)

Table 1: Event extraction results on ACE05-E$^+$ and ERE-EN datasets (F-score, %). * indicates scores obtained from their released codes. The performance of BERT_QA_Arg is lower than that reported in (Du and Cardie, 2020) as they only consider single-token event triggers. Each score of our approach is the mean of three runs and the variance is shown in parenthesis.

Experiment

Model	Trigger Ext.	Arg Ext. (GT)
BERT_QA_Arg†	31.6	17.0
Our Approach	**47.8**	**43.0**

Table 2: Zero-shot F-scores on 23 unseen event types. †: adapted implementation from (Du and Cardie, 2020). GT indicates using gold-standard triggers as input.

	Model	ACE	ERE
Trigger	Our Approach	73.6	60.4
	w/o Seed Trigger	72.2	58.2
	w/o In-Context Attention	72.3	57.9
	w/o Event Type Attention	71.1	56.9
Arg.	Our Approach	55.1	50.4
	w/o Entity Detection	53.0	47.6
	w/o Multiway Attention	53.4	42.8
	w/o Entity Self-attention	53.7	48.3
	w/o Arg Role Self-attention	54.1	47.7

Table 4: Results of various ablation studies. Each score is the average of three runs for each experiment.

Experiment

Source	Target	BERT_QA_Arg$_{multi}$		BERT_QA_Arg$_{binary}$ †		**Our Approach**	
		Trigger Ext.	Argument Ext.	Trigger Ext.	Argument Ext.	Trigger Ext.	Argument Ext.
ERE	ACE	48.9 (48.9)	18.5 (18.5)	50.8 (50.8)	20.9 (20.9)	53.9 (52.6)	30.2 (29.6)
ACE	ACE	70.6	48.3	72.2	50.4	73.6	55.1
ACE+ERE	ACE	70.1	47.0	71.3	49.8	74.4	56.2
ACE	ERE	47.2 (47.2)	18.0 (18.0)	47.2 (45.0)	17.9 (17.1)	55.9 (46.3)	31.9 (26.0)
ERE	ERE	57.0	39.2	56.7	42.9	60.4	50.4
ACE+ERE	ERE	57.0	38.6	54.6	37.1	63.0	52.3

Table 3: Cross ontology transfer between ACE and ERE datasets (F-score %). The scores in parenthesis indicate the performance on the ACE and ERE shared event types.

Experiment

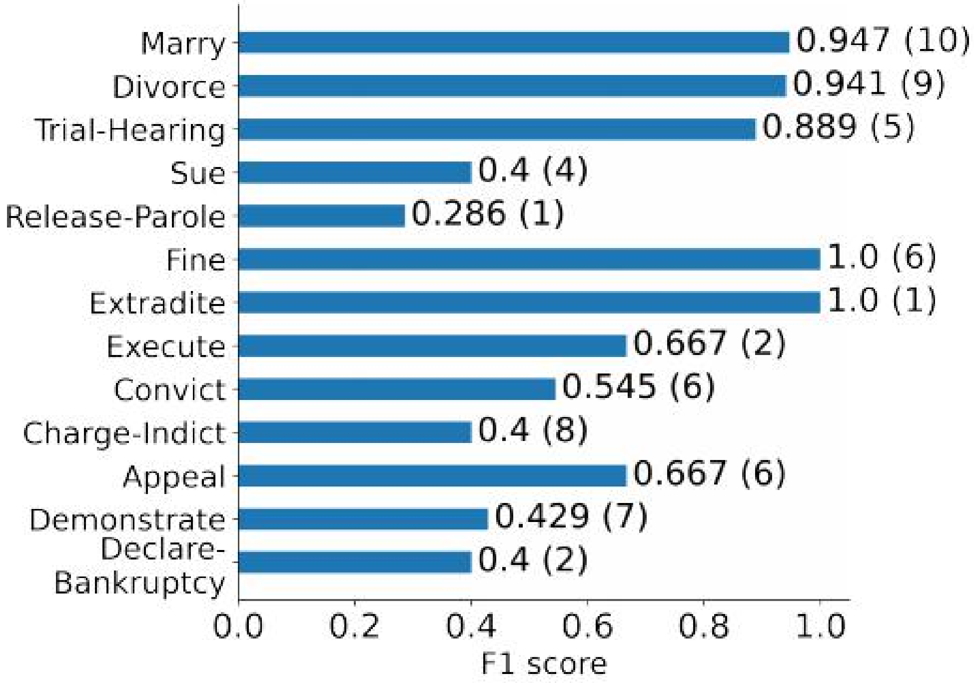


Figure 3: Zero-shot event extraction on each unseen event type. The number in parenthesis indicates # gold event mentions of each unseen type in the test set.

Thank you!

Leibniz-Institut
für Sozialwissenschaften