





# **EventPlus: A Temporal Event Understanding Pipeline**

Mingyu Derek Ma\*, Jiao Sun\*, Mu Yang, Kung-Hsiang Huang, Nuan Wen, Shikhar Singh, Rujun Han, Nanyun Peng

## Introduction



Example of event understanding result

- Event understanding is a core problem for Natural Language Understanding
- Existing NLP analysis tools are mostly in token and sentence-level
- Semantic understanding of events and their temporal information is under-explored

#### **Contributions**:

- The first event pipeline system providing comprehensive event understanding capabilities
- Each component in EventPlus has comparable performance to the state-of-the-art

#### Annotation as convicted in November of seven charges , which included lying to Description state-of-the-art event-related knowledge extraction models, EventPlus extracts and integrates event triggers, corresponding arguments and roles, event al relation between events, and etc. Please click on the "Feature & Tas o" button on the top right to know how to interpret the resu **Temporal Relation** E Text Input ovember of seven charges, which included lying to Congres and witness tampering, as part of former special counsel Robert Mueller's Russia ong the things he misled Congress about were his communications

## **EventPlus interface**

## **Components and Pipeline Design**

**Event Duration Detection** 

#### raw text

A powerful ice storm continues to maintain its grip. Yesterday New York governor George Pataki toured five counties that have been declared under a state of emergency.



## **Component 1: Multi-task** Learning of Event Trigger and **Temporal Relation Extraction**

- BERT embedding + BiLSTM layer (Han et al., 2019)
- SOTA on both tasks

Corpus		Model	F1	
		Chambers et al. (2014)	87.4	
Т	B-Dense	Han et al. (2020a)	90.3	
		Ours	90.8	
N	ATDES	Ning et al. (2018b)	85.2	
IV	IAIKES	Ours	87.8	

Performance of event trigger extraction



Corpus	Model	F1
	Vashishtha et al. (2019)	56.6
<b>TB-Dense</b>	Meng and Rumshisky (2018)	57.0
	Ours	64.5
	Ning et al. (2018b)	65.9
MATRES	Ning et al. (2018a)	69.0
	Ours	75.5

Performance of temporal relation extraction

## **Component 2: Event Extraction on ACE Ontology**

• Raw text -> semantic rich information of events

O, B-Ag, O, ... B-Pat, O, O, O, O, O, ... O, O, B-Tri, ... O, O, B-Tri, O, O, O,... O, B-Ag, O, ... O, O, O, O, B-Ag, I-Ag ,... O, O, B-Tri, ... O, O, O, O, O, O, ... Trigger O, O, O, ... O, O, B-Tri, O, O, O,...

### **Component 3: Event Duration Detection**

- Model 1: Fine-tuned BERT
- Model 2: ELMo embeddings followed by attention layers (Vashishtha et al., 2019)

	Typical-Duration			<b>ACE-Duration</b>		
Model	Acc	Acc-c	Corr	Acc	Acc-c	Corr
UDS-T (U)	0.20	0.54	0.59	0.38	0.68	0.62
UDS-T (T)	0.52	0.79	0.71	0.47	0.67	0.50
UDS-T (T+U)	0.50	0.76	0.68	0.49	0.74	0.66
BERT (T)	0.59	0.81	0.75	0.31	0.67	0.64
BERT (T+U)	0.56	0.81	0.73	0.45	0.79	0.70

- Dataset 1: UDS-T (Vashishtha et al., 2019) over 11 duration categories(instant, ..., centuries)
- Dataset 2: Typical-Duration (Pan et al., 2006) over 7 duration categories (seconds, minutes, hours, days, weeks, months, years)
- Dataset 3: ACE-Duration (new) over 7 duration categories
- SOTA on event duration detection

## **Component 4: Negation and Speculation Cue Detection and Scope Resolution**

The United States is **not** considering sending troops to Mozambique The United States *might* send troops to Mozambique

- Trained with ACE2005
- Multi-task learning with shared BERT encoder



- Inference constraints to keep valid predictions sentence
  - Entity-Argument, Entity-Trigger, Valid Trigger-Argument constraints Ο

Entity Trig-I Trig-C Arg-I Arg-C • Outperforms OneIE on entity detection Model OneIE 90.2 78.2 74.7 59.2 56.8 • Worse on trigger and argument 72.5 57.7 55.7 91.3 75.8 Ours Evaluation on ACE 2005 dataset



Try live demo at: kairos-event.isi.edu

Code available at:

github.com/pluslabnlp/eventplus

detection performance

- Identifies speculation and negation events and mark with special labels
- BERT-based models trained on SFU Review dataset (Khandelwal and Sawant, 2020)

### **Extension to Biomedical Domain**

- Two approaches to extend event extraction to biomedical domain
- Multi-domain training with GENIA
- Replace current component with an in-domain event extraction component: SciBERT-FT (Huang et al., 2020)

- 0.92 F1 score for cue detection model
- 0.88 F1 score for tokenlevel scope resolution



**Evaluation on GENIA dataset** 

Experiments show multi-domain training improves the performance, though SciBERT-FT is still the best