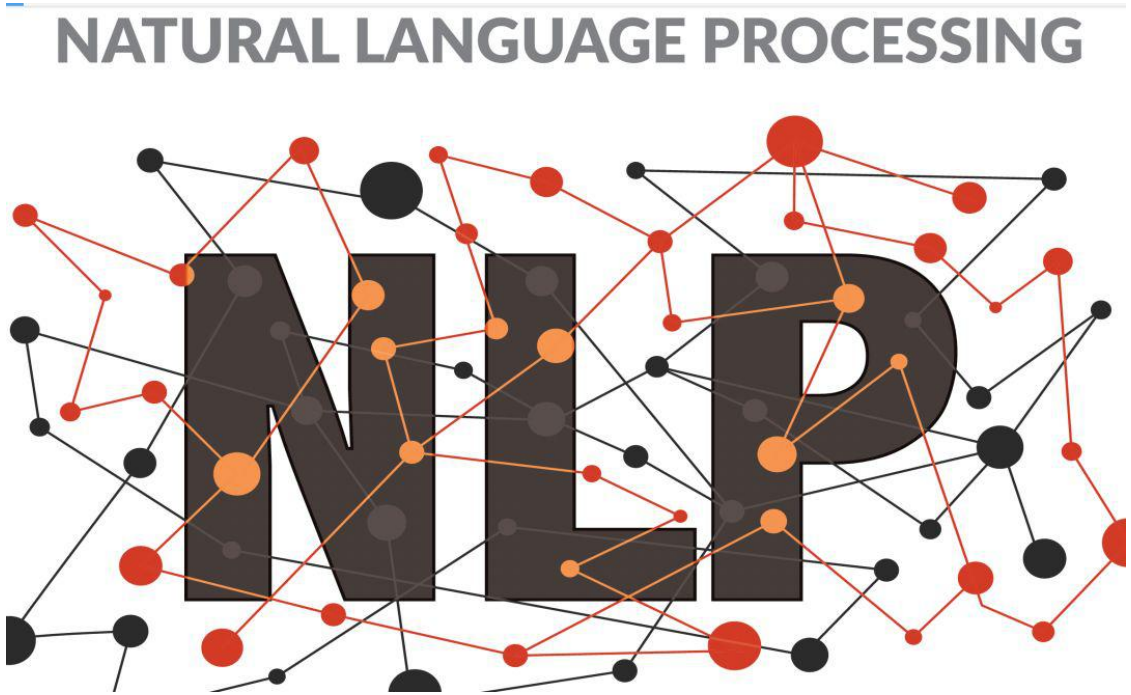
Prompt for Extraction? PAIE: Prompting Argument Interaction for Event Argument Extraction

Yubo Ma[1†*], **Zehao Wang**[2†*], **Yixin Cao**[‡3]
Mukai Li[4†], **Meiqi Chen**[5†], **Kun Wang**[4], **Jing Shao**[4]
[1] Nanyang Technological University [2] KU Leuven [3] Singapore Management University
[4] SenseTime Group Limited [5] Peking University
yubo001@e.ntu.edu.sg, zehao.wang@esat.kuleuven.be

2022.03.27 • ChongQing

ACL_2022

Reported by Yidan Liu



1.Introduction

2.Method

3.Experiments

Introduction

Event type: Transaction.Transfer-Ownership

Sentence-level EAE

Seller Artifact

Cash - strapped Vivendi wants to <t> sell </t> Universal Studios, its Universal theme parks and television production company.

Artifact Artifact

Event type: justice.judicialconsequences.execute

Document-level EAE

Executor Defendant Crime

Dan Sanchez reports : The Saudis Go Full ISIS In Their US - Backed Takfiri War on the Shia Saudi Arabia has perpetrated a mass <t> execution </t> that puts ISIS's beach beheadings to shame. Forty-seven heads rolled on Saturday. One of them belonged to Nimr al - Nimr, a revered Shi'ite cleric who had been sentenced to death for sermons in which he criticized the government (especially for its persecution of the country 's Shi'ite minority).

Figure 1: Examples of (top) sentence-level and (bottom) document-level event argument extraction. Trigger words are included in special tokens <t> and </t>. Underlined words denote arguments and arcs denote roles.

Method

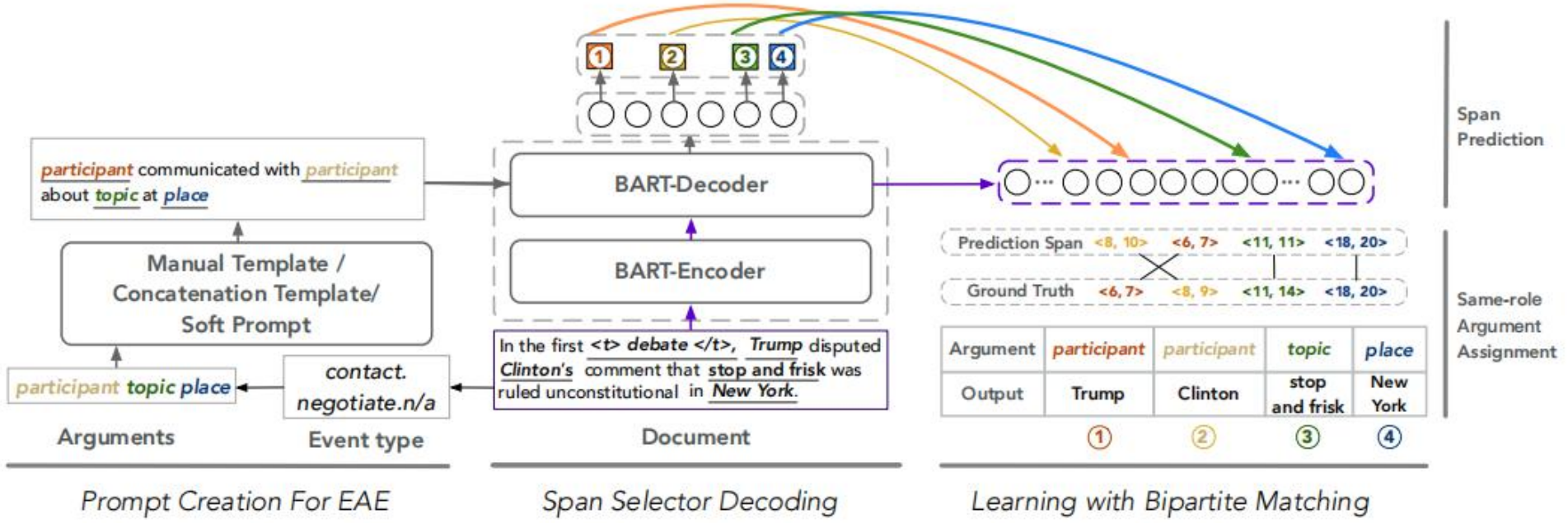
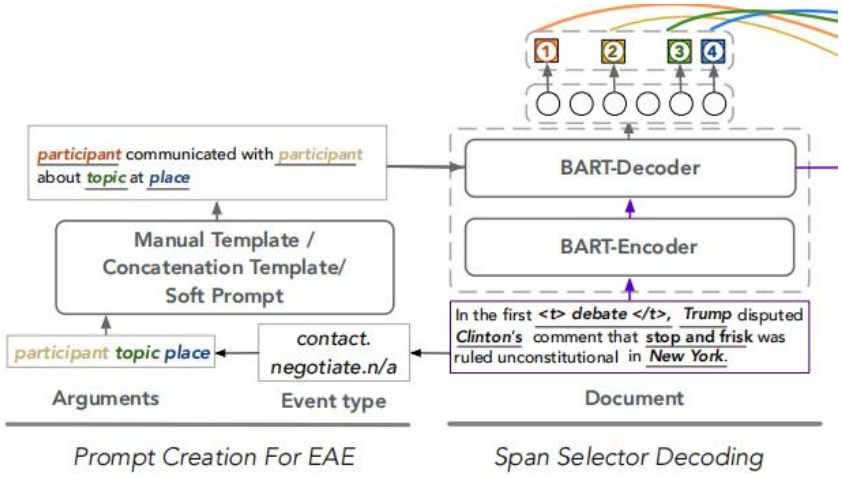


Figure 2: Overall architecture of PAIE. Given a context (about an event), PAIE first creates joint prompt based on its event type. Then the context and prompt are fed into the BART-encoder and BART-decoder to generate context representation and role-specific span selectors. Multiple span selectors extract argument spans from the context simultaneously. A bipartite matching loss finally optimizes the global span assignment.

Method

Prompt Creation for EAE



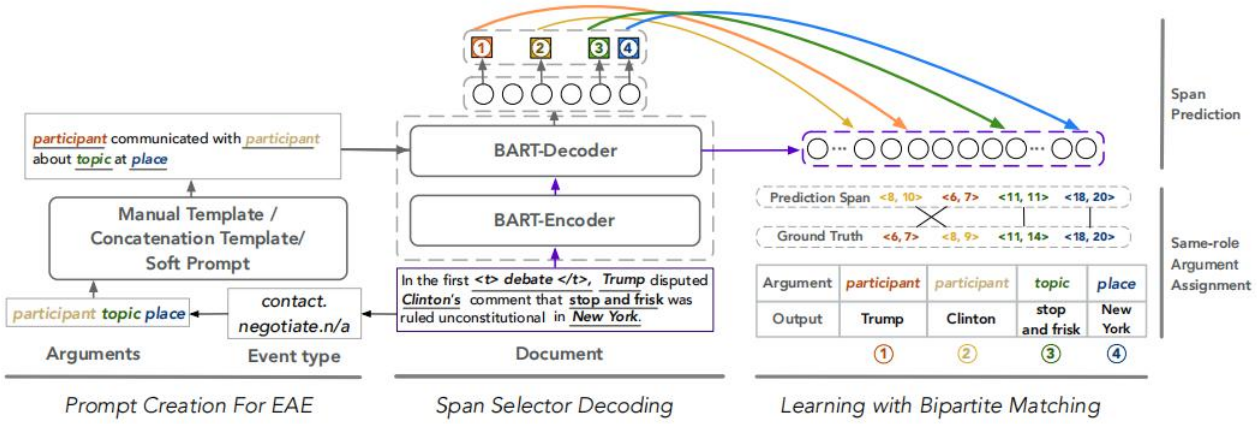
Participant communicated with Participant about topic at Place .

Prompt Type	Prompt Example
MA Template	Victor (and Victor) defeated in ConflictOrElection at Place (and Place)
CA Template	Victor (Victor) ConflictOrElection Place (Place)
SF Prompt	<Vic_left0> Victor <Vic_right0> (<Vic_left0> Victor <Vic_right0>) Defeated <Conf_left0> ConflictOrElection <Conf_right0> <Place_left0> Place <Place_right0> (<Place_left0> Place <Place_right0>)

Table 1: Variants of prompt introduced in section 3.2. **MA**:Manual. **CA**:Concatenation. **SF**:Soft. Words with angle brackets in Soft Prompt denote role-specific pseudo tokens of continuous prompt.

Method

Span Selector Decoding



$$\tilde{X} = [x_1, x_2, ..., \langle \mathbf{t} \rangle, x_{trig}, \langle \mathbf{/t} \rangle, ..., x_n]$$

$$H_X^{(enc)} = \text{BART-Encoder}(\tilde{X})$$
$$H_X = \text{BART-Decoder}(H_X^{(enc)}; H_X^{(enc)}) \quad (1)$$
$$H_{pt} = \text{BART-Decoder}(Pt; H_X^{(enc)})$$

role-specific span selector:

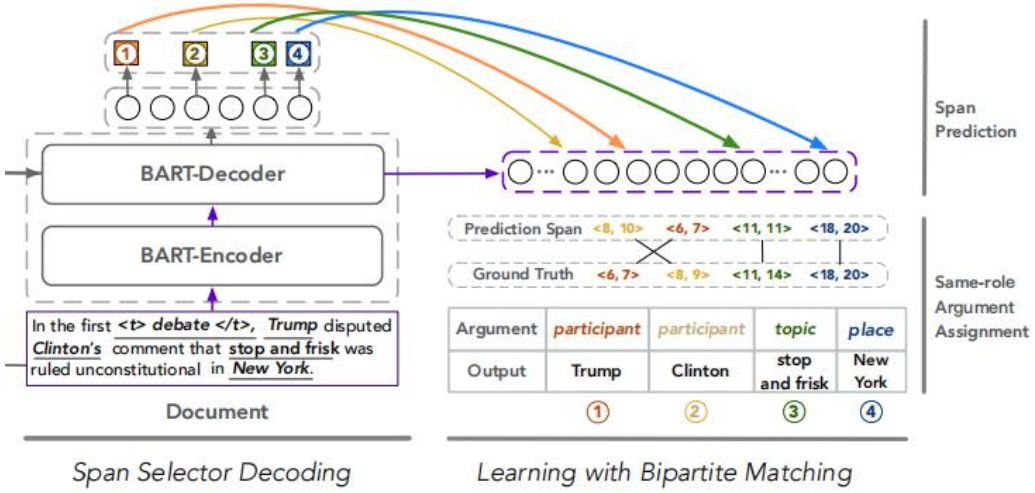
For k-th slot in the joint prompt we mean-pool its corresponding representations from h_{pt} and obtain role feature $\psi_k \in R^h$

$$\psi_k^{(start)} = \psi_k \circ w^{(start)} \in R^h$$
$$\psi_k^{(end)} = \psi_k \circ w^{(end)} \in R^h \quad (2)$$

$\theta = [w^{(start)}; w^{(end)}] \in R^{h \times 2}$ $\theta_k = [\psi_k^{(start)}; \psi_k^{(end)}]$ is exactly the span selector for k-th slot in the prompt.

Method

Learning with Bipartite Matching



Span Prediction:

$$\text{logit}_k^{(start)} = \psi_k^{(start)} H_X \in R^L$$
$$\text{logit}_k^{(end)} = \psi_k^{(end)} H_X \in R^L \quad (3)$$

$$(\hat{s}_k, \hat{e}_k) = \underset{(i,j)\in L^2, i<j}{\arg\max} \ \text{logit}_k^{(start)}(i) + \text{logit}_k^{(end)}(j) \quad (4)$$

Same-role Argument Assignment:

optimal assignment $\hat{\sigma}$,

$$p_k^{(start)} = \text{Softmax}(\text{logit}_{\hat{\sigma}(k)}^{(start)})$$
$$p_k^{(end)} = \text{Softmax}(\text{logit}_{\hat{\sigma}(k)}^{(end)}) \quad (5)$$

$$\mathcal{L}_k = -(\log p_k^{(start)}(s_k) + \log p_k^{(end)}(e_k)) \quad (6)$$

Experiments

Model	PLM	ACE05		RAMS		WIKIEVENTS		
		Arg-I	Arg-C	Arg-I	Arg-C	Arg-I	Arg-C	Head-C
FEAE (Wei et al., 2021)	BERT-b	-	-	53.5*	47.4*	-	-	-
DocMRC (Liu et al., 2021a)	BERT-b	-	-	-	45.7*	-	43.3*	-
OneIE (Lin et al., 2020)	BERT-b	65.9	59.2	-	-	-	-	-
	BERT-l	73.2	69.3	-	-	-	-	-
EEQA (Du and Cardie, 2020)	BERT-b	68.2*	65.4*	46.4	44.0	54.3	53.2	56.9
	BERT-l	70.5	68.9	48.7	46.7	56.9	54.5	59.3
BART-Gen (Li et al., 2021)	BART-b	59.6	55.0	50.9	44.9	47.5	41.7	44.2
	BART-l	69.9*	66.7*	51.2	47.1	66.8	62.4	65.4
EEQA-BART (Our implementation)	BART-b	68.9	67.0	49.4	46.3	60.3	57.1	61.4
	BART-l	73.1	72.2	51.7	48.7	61.6	57.4	61.3
PAIE (Ours)	BART-b	73.0	70.6	53.0	49.8	68.2	63.4	66.4
	BART-l	**75.7**	**73.3**	**55.6**	**53.0**	**69.6**	**65.7**	**69.2**

Table 2: Overall performance. We highlight the best result and underline the second best. * means the value from the original paper. **b** in column **PLM** denotes base model and **l** denotes large model.

Experiments

Model	Arg-C ACE05	Arg-C RAMS	Arg-C WIKI
PAIE	70.6±0.85	49.8 ±0.77	63.4 ±1.04
- bipartite matching	70.3±0.90	49.4±0.69	62.8±0.48
- multi-arg prompt	66.2±1.07	47.7±0.81	61.8±0.85
- role-specific selector	67.0±0.64	46.3±0.77	57.1±0.82
EEQA	65.4	44.0	53.2

Table 3: Ablation study on three benchmarks.

Concatenate	PLM	ACE05	RAMS	WIKIEVENTS
✓	BE-b	65.9	46.3	62.9
✓	BA-b	70.2	49.3	62.8
✓	BA-l	72.3	51.7	65.1
✗	BA-b	70.6	49.8	63.4
✗	BA-l	**73.3**	**53.0**	**65.7**

Table 4: Arg-C F1 of different PLMs (BE and BA denote BERT and BART) and usages of prompt (encoder or decoder). Concatenate stands for concatenating prompt with context as inputs of encoder.

Experiments

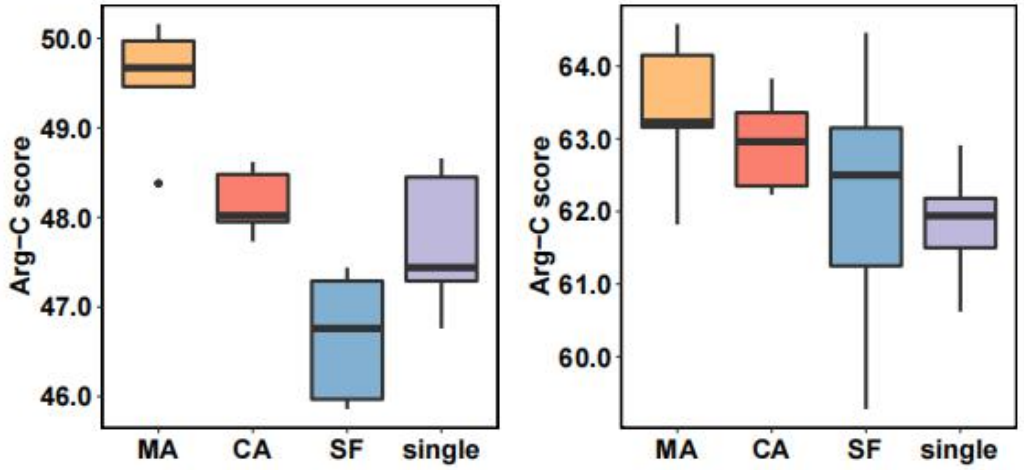


Figure 3: Arg-C F1 using three different types of prompts in Table 1 plus the single template on document-level test set (left) RAMS and (right) WIKIEVENTS.

Model	Trigger-Argument Distance d				
	$-2_{[79]}$	$-1_{[164]}$	$0_{[1811]}$	$1_{[87]}$	$2_{[47]}$
BART-Gen	17.7	16.8	44.8	16.6	9.0
DocMRC	21.0	20.3	46.6	17.2	12.2
FEAE	**23.7**	19.3	49.2	25.0	5.4
EEQA-BART	15.6	24.0	51.7	23.5	8.0
PAIE	21.7	**27.3**	**54.7**	**29.4**	**25.4**

Table 5: Performance (Arg-C F1 score) breakdown by argument-trigger distance d on RAMS development set. The argument number of each case is given in the bracket.

Experiments

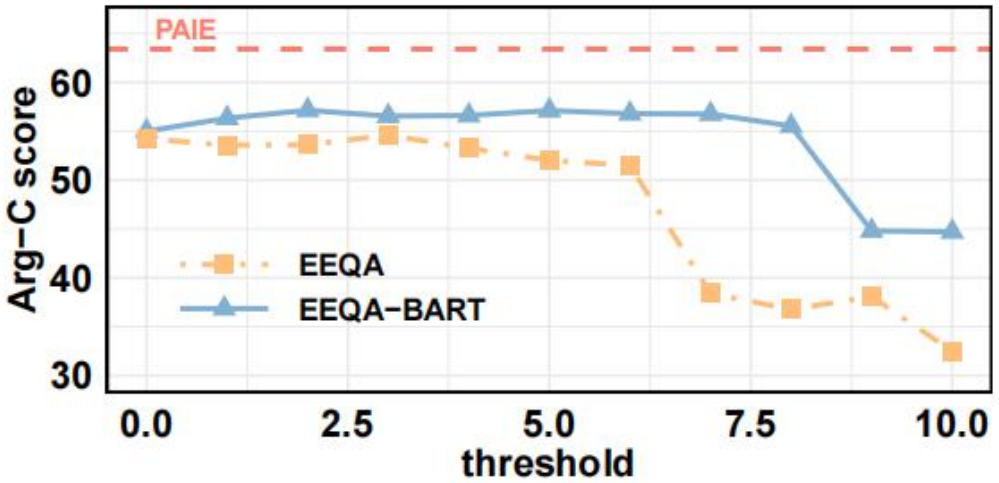


Figure 4: Arg-C F1 w.r.t different thresholds for WIKIEVENTS. We draw the performance of PAIE in red dashed line for comparison (no threshold tuning).

Model	WIKIEVENT Argument Number n			
	$1_{[468]}$	$2_{[66]}$	$3_{[15]}$	$\geq 4_{[17]}$
EEQA-BART	$58.0_{(-6)}$	$59.7_{(-2)}$	$28.6_{(-9)}$	$10.0_{(-18)}$
PAIE (Ours)	**64.7**	**61.4**	**38.1**	**28.6**

Table 6: Arg-C F1 on WIKIEVENTS breakdown by argument number n of one role. The case number is given in the square bracket.

Experiments

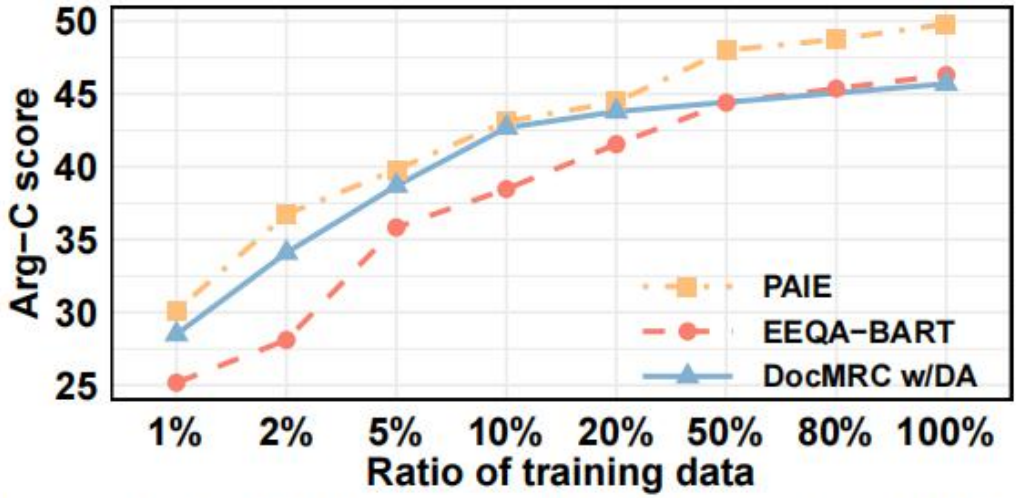


Figure 5: Arg-C F1 score on RAMS test set w.r.t different data ratio. w/DA denotes data augmentation.

Model	ACE05		RAMS		WIKIEVENTS	
	B	L	B	L	B	L
BART-Gen	5.8	12.4	33.2	54.8	19.1	29.0
EEQA-BART	11.8	36.0	66.0	187.4	30.9	83.8
PAIE	**2.9**	**8.4**	**19.0**	**38.6**	**8.4**	**18.3**

Table 7: Inference time in second for different models on the whole test set of ACE05, RAMS, WIKIEVENTS.

Thank you!