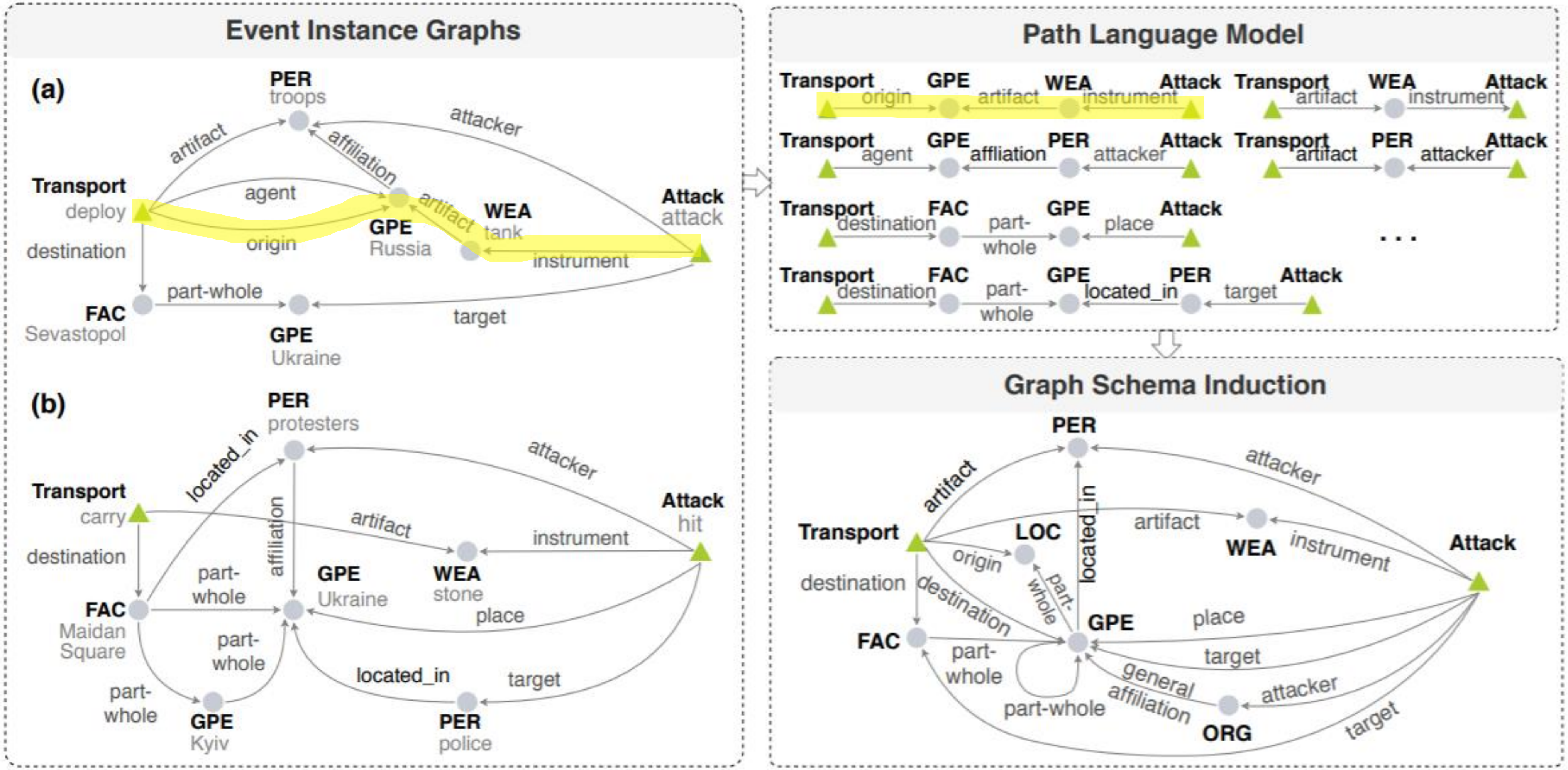
Connecting the Dots: Event Graph Schema Induction with Path Language Modeling

Manling Li, Qi Zeng, Ying Lin, Kyunghyun Cho, Heng Ji, Jonathan May, Nathanael Chambers, Clare Voss

Overview

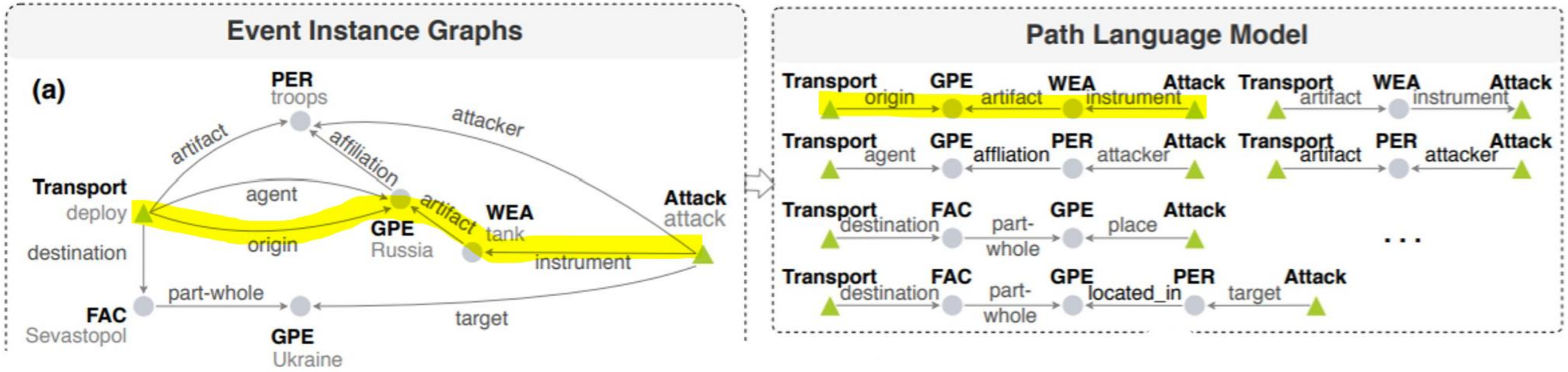
- The paper introduces a method based on path language modeling to induce Event Graph Schema.
- To evaluate the effectiveness of the induction method, the paper proposes two intrinsic evaluation metrics.
- To further demonstrate the quality of the induced event graph schema, they use them to improve their Joint Information Extraction system (i.e., OneIE).

Event Graph Schema Induction



Path Language Model

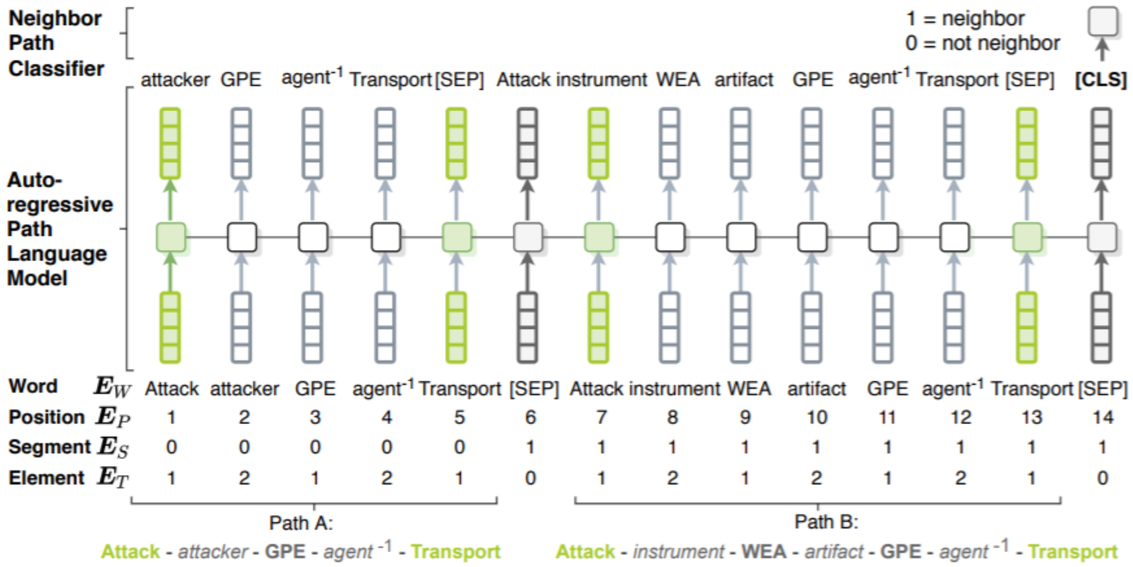
Pretraining data: extracted paths from event instance graphs obtained by either golden or predicted IE graphs.



Path Language Model

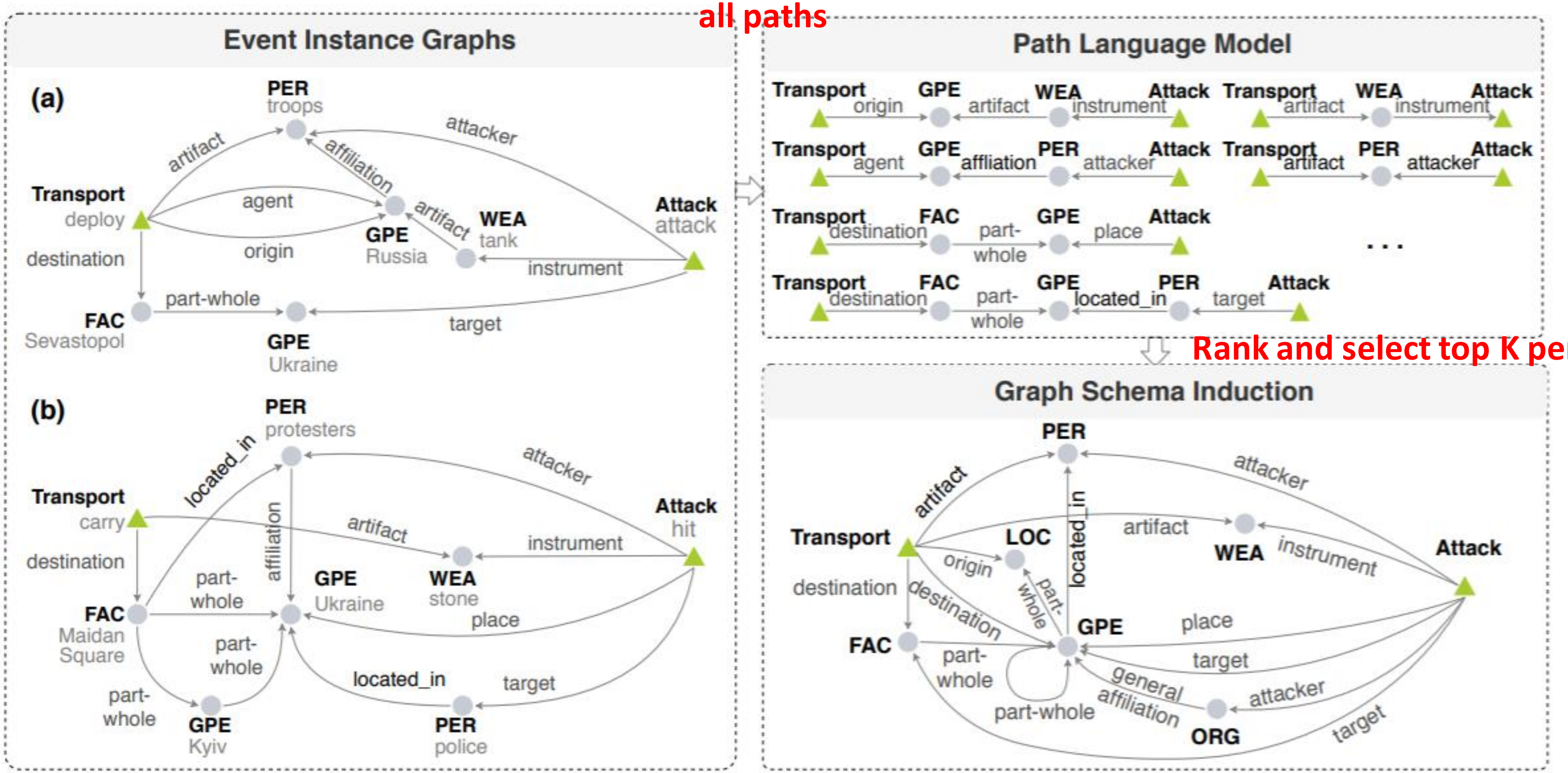
Pretraining tasks:

+ Next path-element prediction.
+ Neighbor path prediction (all orders are considered):
  > Positive pairs: paths from the same event instance graph.
  > Negative pairs: paths from different event instance graphs, however, with the same event types.



Scheme Graph Construction

Extract
all paths



Rank and select top K percent paths

Scheme Graph Construction

- Given two event types ϕ and ϕ', we first retrieve all related paths connecting these two types $\mathcal{P}_{\langle \phi, \phi' \rangle}$
- For each path in $\mathcal{P}_{\langle \phi, \phi' \rangle}$, we compute the following score: $f(p_i) = f_{\mathrm{LM}}(p_i) + \alpha f_{\mathrm{NP}}(p_i)$

Where $f_{\mathrm{LM}}(p)$ captures salience and coherence of the path $f_{\mathrm{LM}}(p_i) = \log P([\phi, \psi_{0;1}, \phi_1, \psi_{1;2}, ..., \phi'])$

And $f_{\mathrm{NP}}(p)$ captures how frequently this path is paired with other paths:

$$f_{\mathrm{NP}}(p_i) = \frac{1}{|\mathcal{P}_{\langle \phi, \phi' \rangle}|} \sum_{p_j \in \mathcal{P}_{\langle \phi, \phi' \rangle}} \log P(p_j \in \mathcal{N}_{p_i})$$

- Next, we select top K *percent* paths and merge them:

  + Nodes of the same type into a single node.

  + Edges are directly reused (loops are allowed)

Evaluate on intrinsic metrics

- After pretraining a path language model, we use the ranking method to induce a set of event **scheme** graphs $\mathcal{S}$.
- Given a set of golden event **instance** graphs $\mathcal{G}$. The quality of the induced scheme graphs is evaluated by two metrics:

  + Instance Coverage.

  + Instance Coherence.

Instance Coverage

- Instance graphs can be considered as partially instantiated graph schema.
- "Instance Coverage" evaluates how well an event scheme graph can capture the structures in instance graphs.
- Given an instance graph $g \in \mathcal{G}$ and a scheme graph $s \in \mathcal{S}$, the interesection between the two graph is evaluated by:

$$|g \cap s|_{\mathbb{I}} = \sum_{\langle \phi_i, \psi_{ij}, \phi_j \rangle \in s} \text{count}(\langle v_m, e_{mn}, v_n \rangle)$$

$$|g \cap s|_{\mathbb{S}} = \sum_{\langle v_m, e_{mn}, v_n \rangle \in g} \text{count}(\langle \phi_i, \psi_{ij}, \phi_j \rangle)$$

- Precision, Recall are then computed as:

$$\text{Precision} = \frac{\sum_{s \in \mathcal{S}} \sum_{g \in \mathcal{G}} |g \cap s|_{\mathbb{S}}}{\sum_{s \in \mathcal{S}} |s|_{\mathbb{S}}},$$

$$\text{Recall} = \frac{\sum_{s \in \mathcal{S}} \sum_{g \in \mathcal{G}} |g \cap s|_{\mathbb{I}}}{\sum_{g \in \mathcal{G}} |g|_{\mathbb{I}}}.$$

Instance Coherence

- A **coherent** schema should have the maximal number of matched instance graphs $g \cap s$
- from a **single document**, but with the minimal number of matched graphs connecting two event instances from **different documents**.

$$\text{Coherence} = \frac{\sum_{s\in\mathcal{S}} \sum_{g\in\mathcal{G}} \sum_{p\in g\cap s} f(p) \cdot \mathbb{I}_g}{\sum_{s\in\mathcal{S}} \sum_{g\in\mathcal{G}} \sum_{p\in g\cap s} f(p)}$$

Where $\mathbb{I}_g$ indicates whether an instance graph g is between event instances from the same document or not.

Performance on intrinsic task

- Dataset: ACE2005. Train/dev/test split follows OneIE's split.

Split	#Docs	#Entities	#Rels	#Events	#Args
Historical$_{ann}$	529	47,525	7,152	4,419	7,888
Historical$_{sys}$	529	48,664	7,018	4,426	6,614
Validation	40	3,422	728	468	938
Target	30	3,673	802	424	897

Table 2: Data statistics.

Performance on intrinsic task

Instance Coverage

Historical Instance Graphs	Model	Schema@10									Schema@20								
		$l=7$			$l=5$			$l=3$			$l=7$			$l=5$			$l=3$		
		P	**R**	**F$_1$**	**P**	**R**	**F$_1$**	**P**	**R**	**F$_1$**	**P**	**R**	**F$_1$**	**P**	**R**	**F$_1$**	**P**	**R**	**F$_1$**
Historical$_{ann}$	**Frequency**	76.7	9.5	16.9	90.5	48.3	63.0	100	37.5	54.6	63.6	17.9	28.0	87.6	70.6	78.2	100	42.6	59.7
	Unigram LM	63.9	7.3	13.1	87.1	35.4	50.3	100	33.7	50.4	55.4	14.8	23.4	86.0	60.8	71.2	100	43.8	60.9
	Bigram LM	75.4	8.5	15.3	92.6	36.8	52.6	100	33.4	50.1	62.6	16.4	26.0	88.1	63.2	73.6	100	43.2	60.3
	Trigram LM	62.7	8.5	15.0	89.4	41.6	56.7	100	39.9	57.0	53.4	17.8	26.7	85.6	68.2	75.9	100	44.6	61.6
	PathLM	54.3	16.6	**25.4**	83.7	63.8	**72.4**	100	41.8	**58.9**	53.8	27.2	**36.1**	83.0	80.0	81.5	100	44.7	**61.8**
	w/o CLS$_{NP}$	71.2	14.5	24.1	90.3	58.3	70.9	100	39.8	56.9	57.8	25.8	35.6	85.7	80.1	**82.8**	100	42.9	60.1
Historical$_{sys}$	**Frequency**	68.6	9.8	17.1	87.0	49.4	63.0	100	37.6	54.7	67.8	19.3	29.9	88.5	70.1	78.2	100	41.6	58.8
	Unigram LM	54.3	7.5	13.1	83.7	36.2	50.5	100	41.0	58.2	52.4	17.9	26.7	83.0	66.4	73.8	100	44.6	61.7
	Bigram LM	61.4	7.9	13.9	88.5	37.7	52.8	100	39.2	56.3	58.3	15.3	24.2	86.5	63.8	73.4	100	43.5	60.6
	Trigram LM	65.2	9.8	17.1	89.6	46.8	61.5	100	37.3	54.4	54.5	17.6	26.6	86.2	68.7	76.5	100	44.1	61.2
	PathLM	51.8	18.5	**27.3**	83.2	68.0	**74.8**	100	41.7	**58.8**	49.6	29.3	**36.9**	81.7	85.4	**83.5**	100	44.8	**61.9**
	w/o CLS$_{NP}$	72.7	14.4	24.1	89.5	55.1	68.2	100	40.1	57.3	54.8	24.7	34.0	83.8	75.9	80.0	100	44.7	61.7

Table 3: Instance coverage (%) by checking the intersection of schemas and instance graphs.

Performance on intrinsic task

Instance Coherence

Historical	**Model**	**Schema@10**	**Schema@20**
Historical$_{\text{ann}}$	**Frequency**	67.8	65.6
	Unigram LM	62.4	69.9
	Bigram LM	59.0	67.5
	Trigram LM	56.6	64.9
	PathLM	**76.0**	**79.9**
	w/o CLS_{NP}	75.3	79.2
Historical$_{\text{sys}}$	**Frequency**	60.1	65.6
	Unigram LM	61.8	70.0
	Bigram LM	59.7	69.6
	Trigram LM	55.8	65.8
	PathLM	**76.4**	**78.5**
	w/o CLS_{NP}	73.9	77.1

Table 4: Instance coherence (%) of schema graphs covering top k percent paths, $k = 10, 20$.

Apply to Joint IE

- Replacing the global feature vectors with path-based vectors.
- Each dimension of the feature vector indicates whether the corresponding path appears.

$$s'(\hat{G}) = \sum_{t \in T} \sum_{i=1}^{N^t} \max \hat{\boldsymbol{y}}_i^t,$$

$$s(G) = s'(G) + \sum_{p_i \in s, s \in \mathcal{S}} n_i * w_i,$$

Performance on ACE2005

Model	Entity	Rel	Event			
			Trig-I	Trig-C	Arg-I	Arg-C
OneIE Baseline	**90.3**	44.7	75.8	72.7	57.8	55.5
+PathLM	90.2	**60.9**	76.0	**73.4**	59.0	**56.6**
w/o $\mathrm{CLS}_{\mathrm{NP}}$	90.1	60.3	75.7	72.8	58.3	55.8

Table 5: F_1 score (%) of schema-guided information extraction, including entity extraction (Entity), relation extraction (Rel), event trigger identification (Trig-I) and classification (Trig-C), event argument identification (Arg-I) and argument role classification (Arg-C).