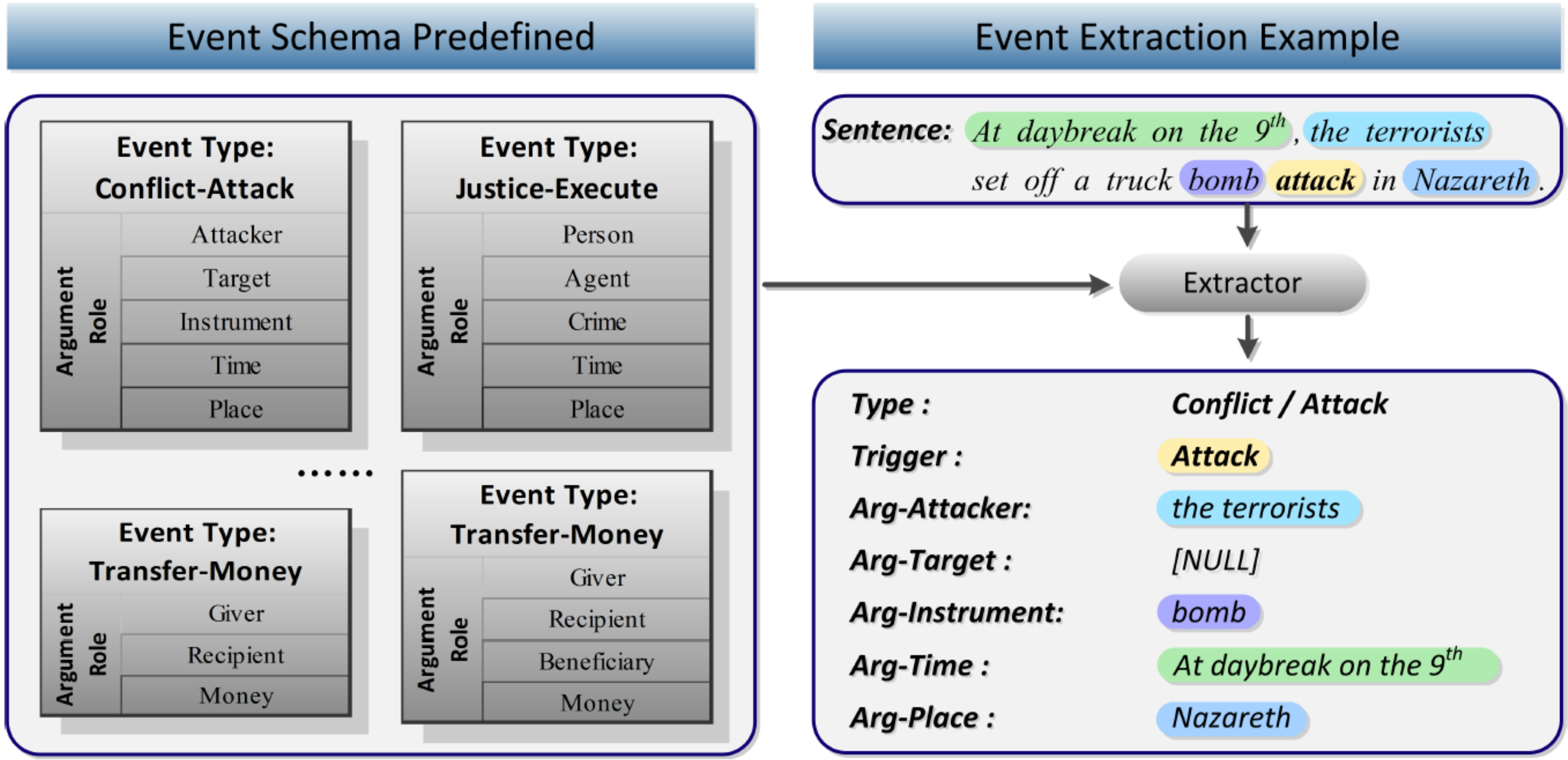
A Joint Neural Model for Information Extraction with Global Features

Ying Lin, Heng Ji, Fei Huang, Lingfei Wu

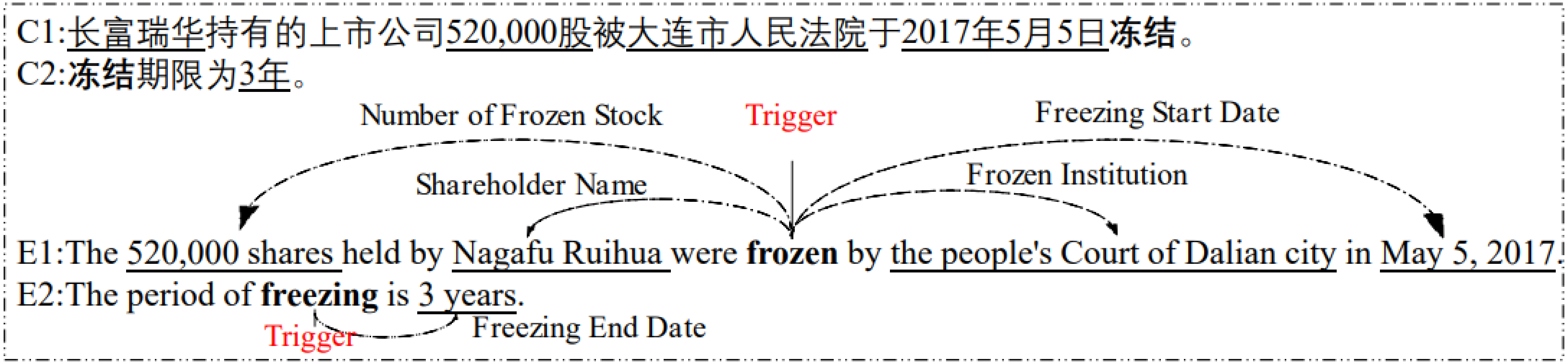
ACL2020

Event Extraction Template



Wei Xiang and Bang Wang. 2019. A Survey of Event Extraction From Text. *IEEE Access*, 7:173111–173137.

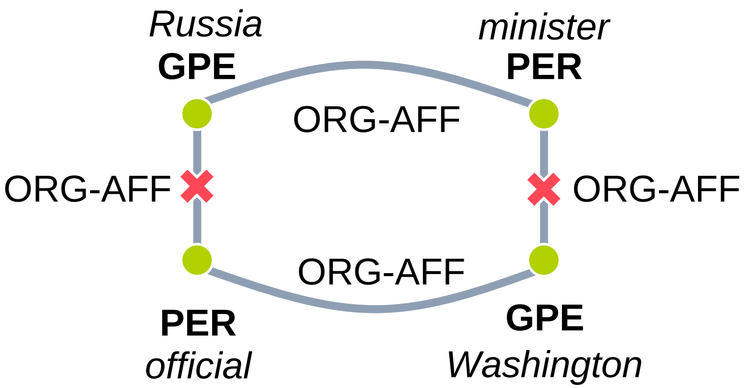
Sentence-level Event Extraction



Hang Yang, Yubo Chen, Kang Liu, Yang Xiao, and Jun Zhao. 2018. DCFEE: A Document-level Chinese Financial Event Extraction System based on Automatically Labeled Training Data. In *Proceedings of ACL 2018, System Demonstrations*, pages 50–55, Melbourne, Australia, July. Association for Computational Linguistics.

Motivation

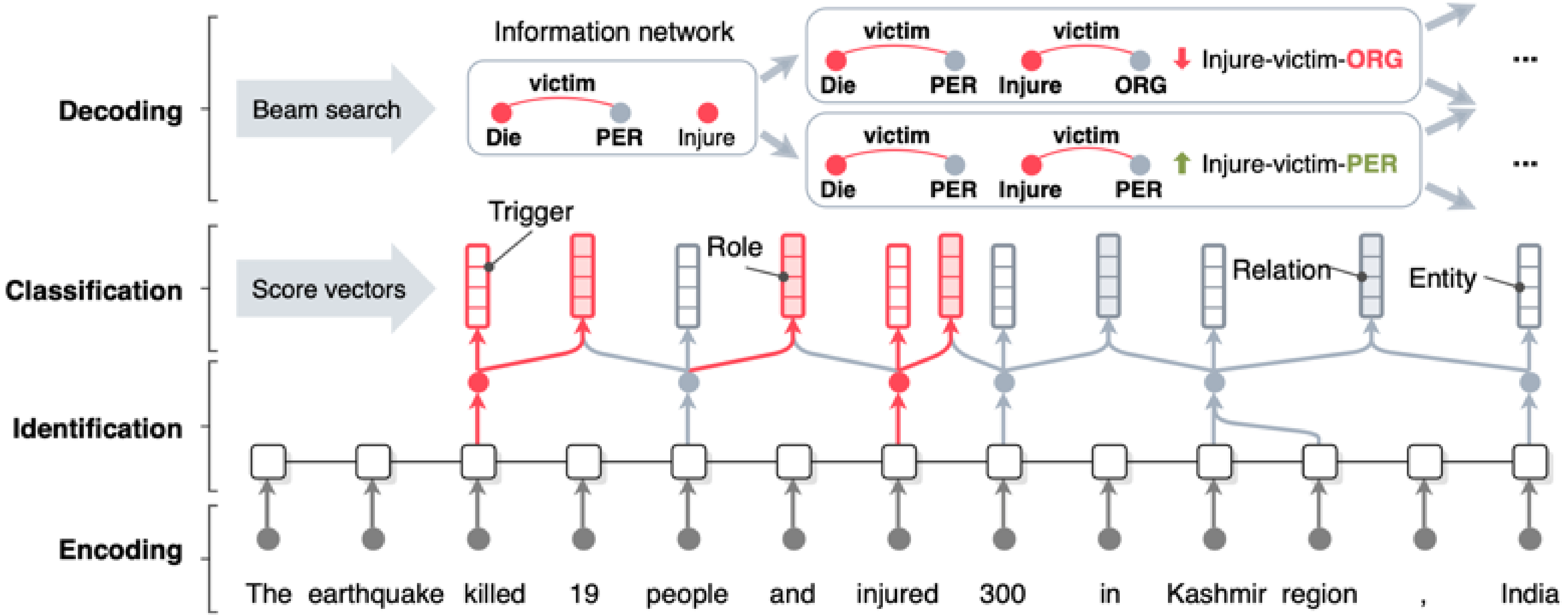
- Pipeline models suffer from the **error propagation problem** and disallow interactions among components.
- Existing neural models do not explicitly model **cross-subtask and cross-instance interactions** among knowledge elements.



***Russia**'s foreign **minister** expressed outrage at suggestions from a top **Washington official** last week...*

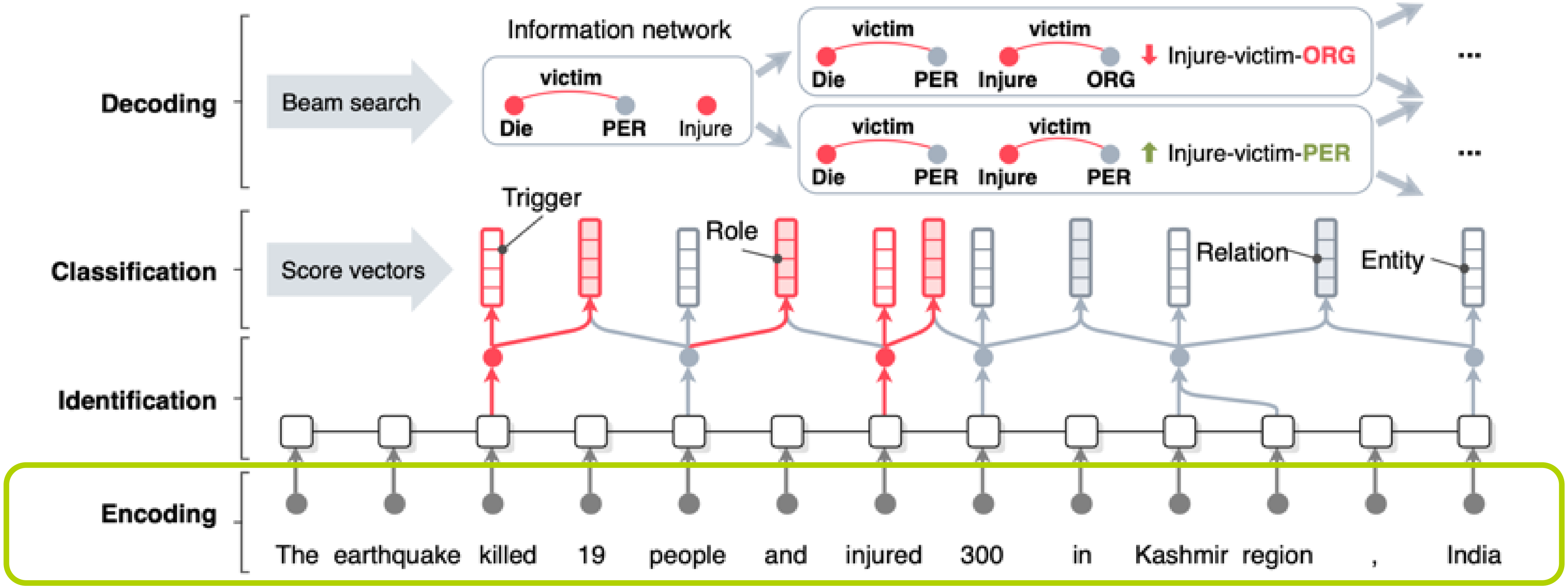
➔ A **single joint neural model** for Information Extraction.

OneIE: An End-to-end Neural Model for IE



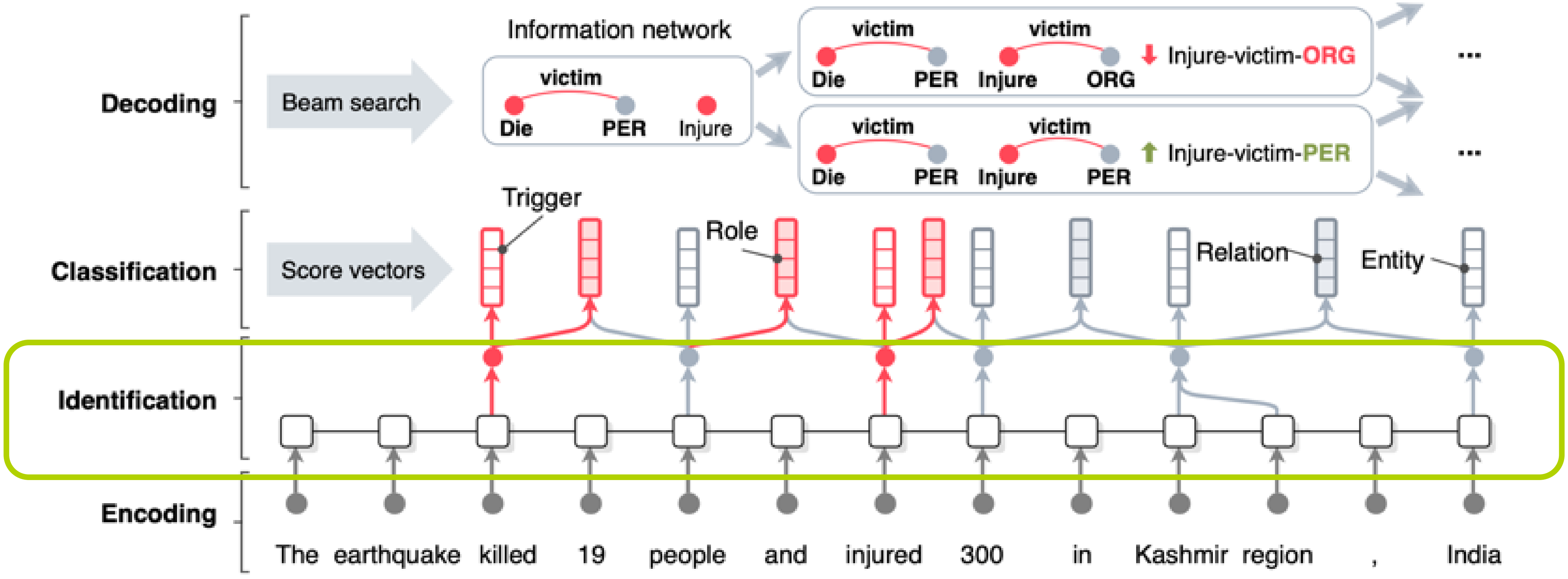
- Our OneIE framework extracts the information graph from a given sentence in four steps: encoding, identification, classification, and decoding

OneIE: An End-to-end Neural Model for IE



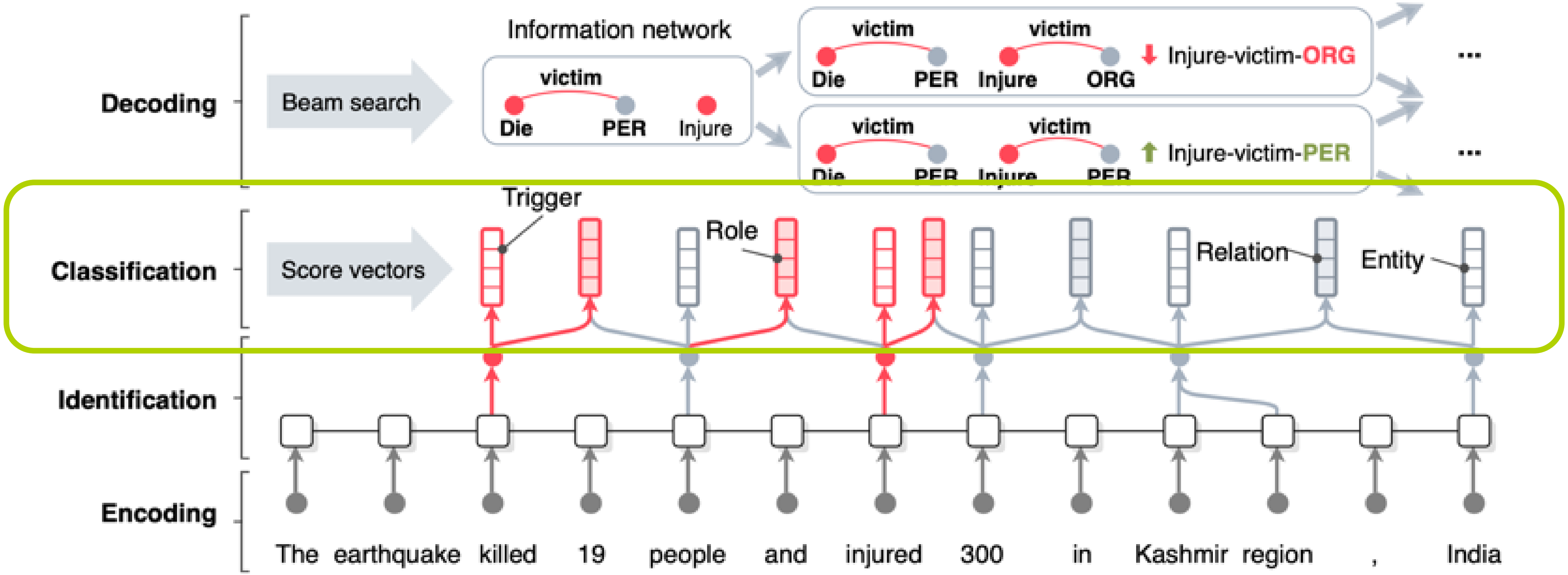
- **Encoding**: We use a BERT encoder to obtain the contextualized embedding of each token

OneIE: An End-to-end Neural Model for IE



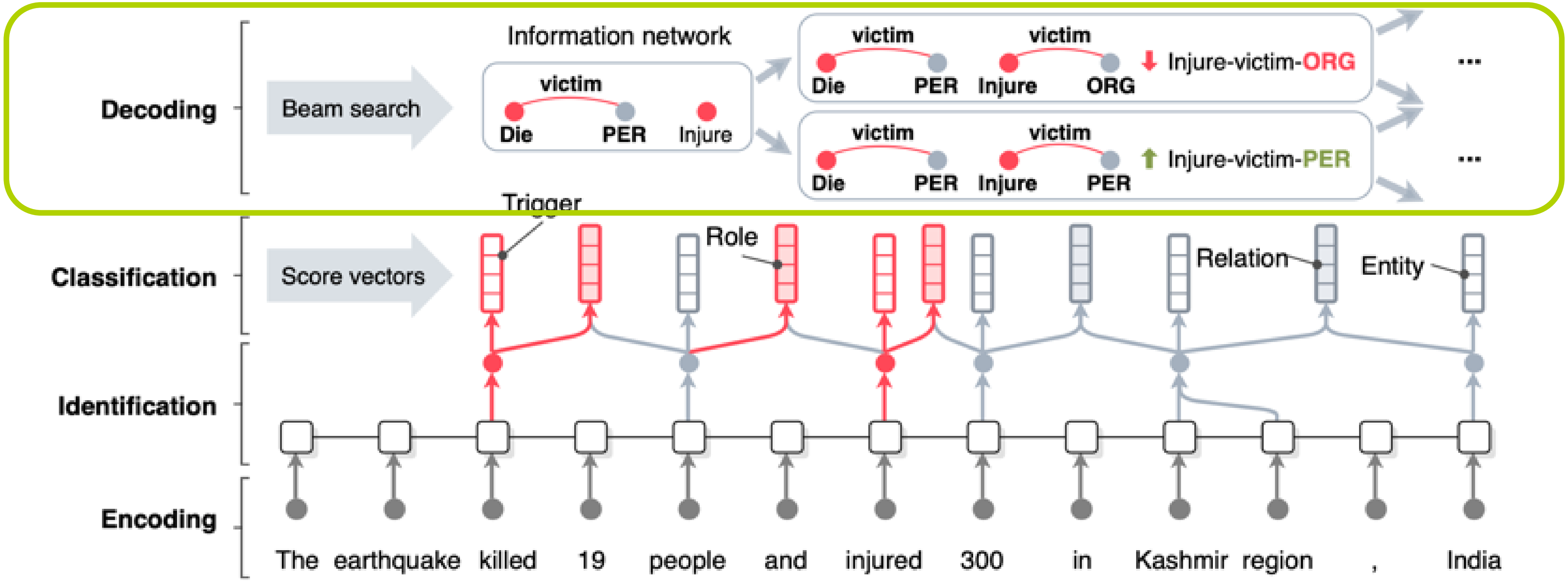
- **Identification**: We use CRF taggers to identify entity mentions and event triggers
- We define the identification loss as $\mathcal{L}^{\mathrm{I}} = -\log p(\boldsymbol{z}|\boldsymbol{X})$

OneIE: An End-to-end Neural Model for IE



- **Classification**: We use task-specific feed-forward networks to calculate label scores for each node or edge
- We define the classification loss as $$\mathcal{L}^{t} = -\frac{1}{N^t}\sum_{i=1}^{N^t} \boldsymbol{y}_i^t \log \hat{\boldsymbol{y}}_i^t$$

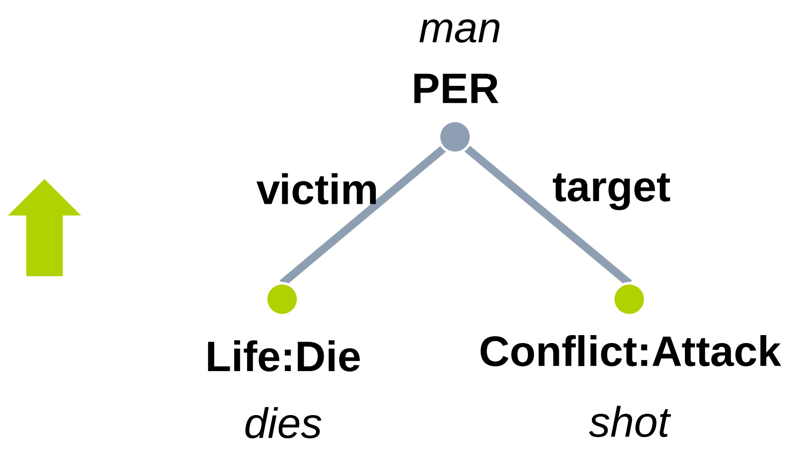
OneIE: An End-to-end Neural Model for IE



- **Decoding**: In the test phase, we use a beam search decoder to find the information graph with the highest global score

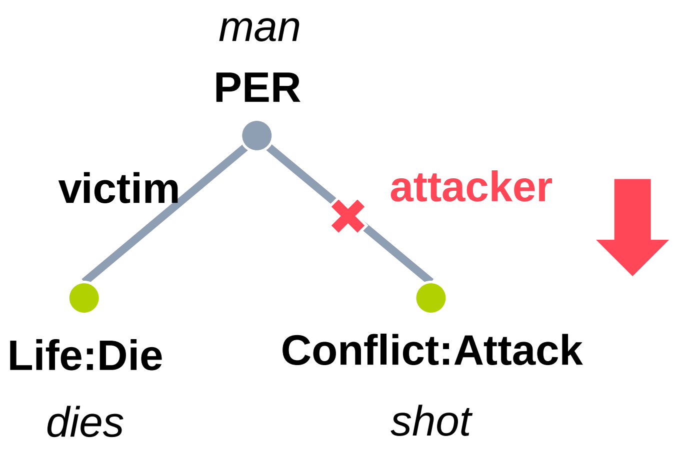
Incorporating Global Features

- We design a set of *global feature templates* (e.g., event_type_1 – role_1 – role_2 – event_type_2 : an entity acts a role_1 argument for an event_type_1 event and a role_2 argument for an event_type_2 event in the same sentence)
- The model learns the *weight* of each feature during training



Life:Die – victim – target – Conflict:Attack

Positive weight



Life:Die – victim – attacker – Conflict:Attack

Negative weight

Incorporating Global Features

- Given a graph G, we generate its global feature vector as $\boldsymbol{f}_G = \{f_{1(G)}, \dots, f_M(G)\}$, where $f_i(\cdot)$ is a function that evaluates a certain feature and returns a scalar. For example,

$$f_i(G) = \begin{cases} 1, & G \text{ has multiple } \textsc{Attack} \text{ events} \\ 0, & \text{otherwise.} \end{cases}$$

- Next, we learn a weight vector $\boldsymbol{u} \in \mathbb{R}^M$ and calculate the global feature score of G as the dot production of $\boldsymbol{f}_G$ and $\boldsymbol{u}$.
- **Global score** of a graph: local graph score + global feature score:

$$s(G) = s'(G) + \boldsymbol{u}\boldsymbol{f}_G$$

- We assume that the gold-standard graph for a sentence should achieve the highest global score and minimize the following loss function:

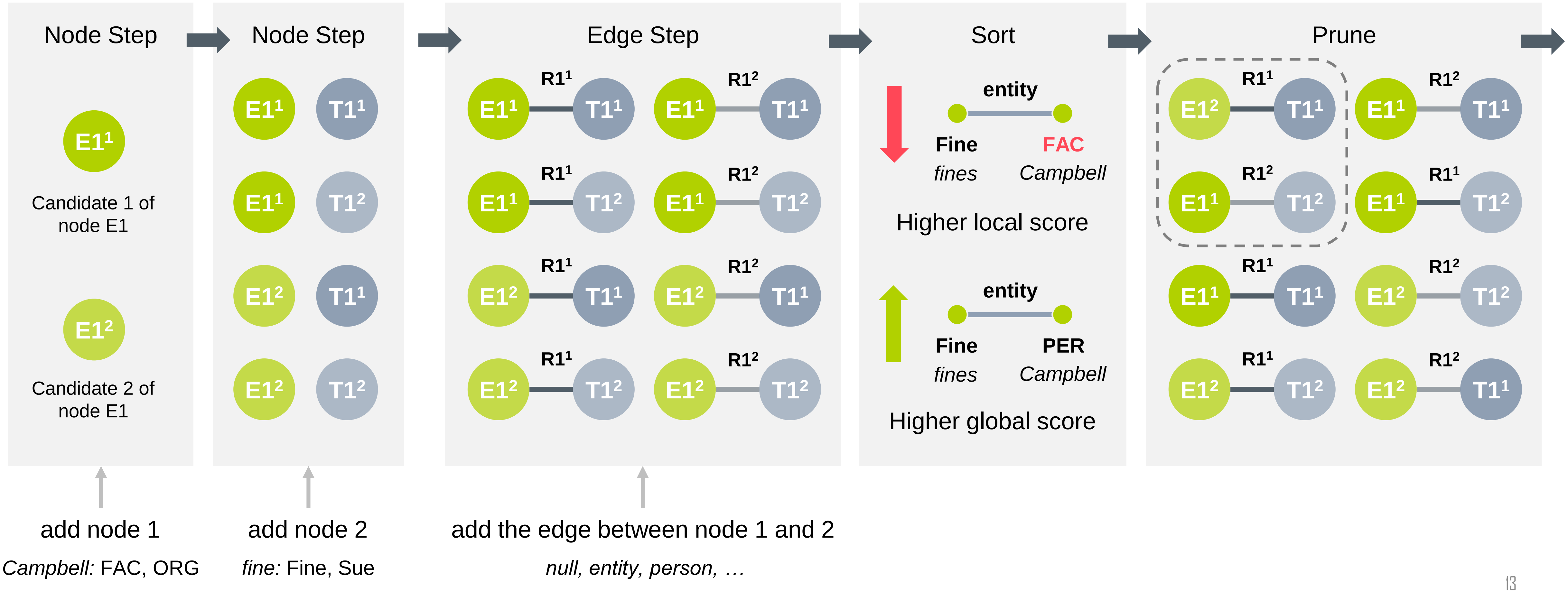
$$\mathcal{L}^{\mathrm{G}} = s(\hat{G}) - s(G)$$

Global Features

Categary	Description
Role	1. The number of entities that act as <role$_i$> and <role$_j$> arguments at the same time.
	2. The number of <event_type$_i$> events with <number> <role$_j$> arguments.
	3. The number of occurrences of <event_type$_i$>, <role$_j$>, and <entity_type$_k$> combination.
	4. The number of events that have multiple <role$_i$> arguments.
	5. The number of entities that act as a <role$_i$> argument of an <event_type$_j$> event and a <role$_k$> argument of an <event_type$_l$> event at the same time.
Relation	6. The number of occurrences of <entity_type$_i$>, <entity_type$_j$>, and <relation_type$_k$> combination.
	7. The number of occurrences of <entity_type$_i$> and <relation_type$_j$> combination.
	8. The number of occurrences of a <relation_type$_i$> relation between a <role$_j$> argument and a <role$_k$> argument of the same event.
	9. The number of entities that have a <relation_type$_i$> relation with multiple entities.
	10. The number of entities involving in <relation_type$_i$> and <relation_type$_j$> relations simultaneously.
Trigger	11. Whether a graph contains more than one <event_type$_i$> event.

Decoding

- We use beam search to decode the information graph
- Example: *He also brought a check from **Campbell** to pay the **fines** and fees.*



Experiment: Datasets

- We conduct our experiments on four datasets derived from ACE (Automatic Content Extraction) 2005 and two datasets derived from ERE (Entities, Relations and Events).
- Existing datasets developed by Wadden et al. (2019):
  - ACE05-R: named entity and relation annotations.
  - ACE05-E: entity, relation, and event annotations.
- We created the following datasets:
  - ACE05-CN: Chinese entity, relation, and event annotations.
  - ACE05-E$^+$ follows the split of ACE05-E and has richer annotations.
  - ERE-EN is derived from LDC205E29, LDC2015E68, and LDC2015E78.
  - ERE-ES: Spanish entity, relation, and event annotations.

Experiment: Results

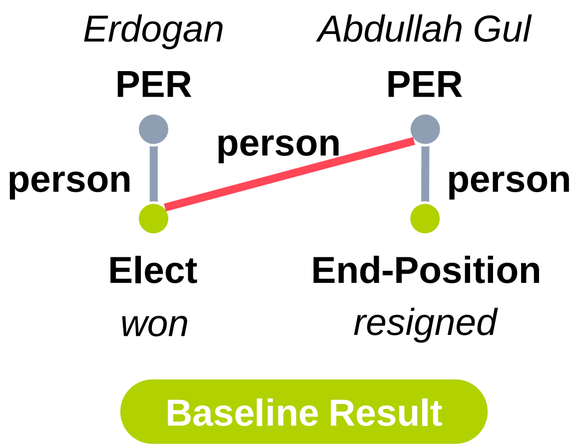
- We compare our model to DyGIE++, the state-of-the-art end-to-end IE model, on ACE05-R and ACE05-E.

Model	ACE05-R		ACE05-E				
	Entity	Relation	Entity	Trigger Identification	Trigger Classification	Argument Identification	Argument Classification
DyGIE++	88.6	63.4	89.7	-	69.7	53.0	48.8
DyGIE++*	-	-	**90.7**	76.5	73.6	55.4	52.5
OneIE	88.8	67.5	90.2	78.2	74.7	59.2	56.8
OneIE*	-	-	90.3	**78.6**	**75.2**	**60.7**	**58.6**

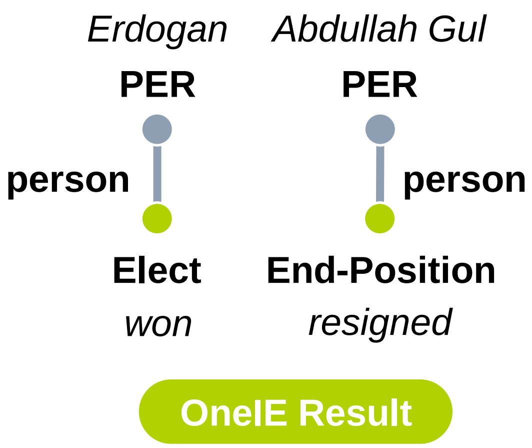
- DyGIE++* and OneIE* use a four-model ensemble optimized for trigger detection.
- We hold the opinion that single-model scores better reflect the actual performance of OneIE and should be used for future comparison

Experiment: Example

- Example: *Prime Minister Abdullah Gul* ***resigned*** *earlier Tuesday to make way for Erdogan, who* ***won*** *a parliamentary seat in by-elections Sunday.*
- The local argument role classifier predicts a redundant Person edge between “won” and “Abdullah Gul”.



1. An **Elect** event usually has only one **Person** argument
2. An entity is unlike to act as a **Person** argument for **End-Position** and **Elect** events at the same time



New Benchmark Results

- In order to reinstate some important elements absent from ACE05-R and ACE05-E, we construct two new benchmark datasets, ACE05-E$^+$ and ERE-EN by adding back:
  - Entity: pronouns
  - Relation: directions
  - Event: multi-token triggers

Dataset	**Entity**	**Trigger** Identification	**Trigger** Classification	**Argument** Identification	**Argument** Classification	**Relation**
ACE05-E$^+$	89.6	75.6	72.8	57.3	54.8	58.6
ERE-EN	87.0	68.4	57.0	50.1	46.5	53.2

Porting to Another Language

- We evaluate our proposed framework on ACE05-CN (Chinese) and ERE-ES (Spanish).
- OneIE works well on Chinese and Spanish data without any special design for the new language.
- We observe that adding English training data can improve the performance on Chinese and Spanish.

Dataset	Training	Entity	Trigger Classification	Argument Classification	Relation
ACE05-CN	CN	88.5	65.6	52.0	62.4
	CN+EN	89.8	67.7	53.2	62.9
ERE-ES	ES	81.3	56.8	40.3	48.1
	ES+EN	81.8	59.1	42.3	52.9

- For ACE05-CN, EN refers to ACE05-E^{+}. For ERE-ES, EN refers to ERE-EN.

Salient Global Features

- Salient positive and negative global features learned by the model
- Our global features are explainable

	Features	Weight
1	A **Transport** event has only one **Destination** argument	2.61
2	An **Attack** event has only one **Place** argument	2.31
3	A **PER-SOC** relation exists between two **PER** entities	1.51
4	A **Beneficiary** argument is a **PER** entity	0.93
5	An entity has an **ORG-AFF** relation with multiple entities	-3.21
6	An event has two **Place** arguments	-2.47
7	A **Transport** event has multiple **Destination** argument	-2.25
8	An entity has a **GEN-AFF** relation with multiple entities	-2.02

Remaining Errors

- We have analyzed 75 of the remaining errors. In this figure, we present the distribution of various error types which need more features and knowledge acquisition to address in the future.