Joint Event and Temporal Relation Extraction with Shared Representations and Structured Prediction

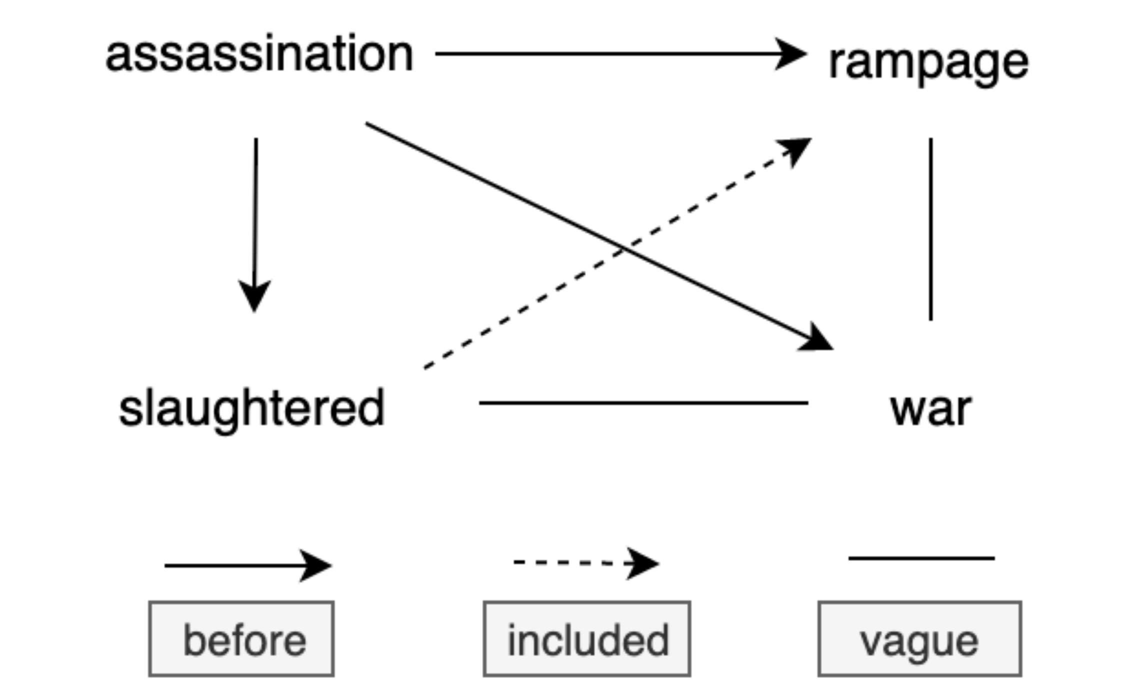
Rujun Han[1,2], Qiang Ning[3], Nanyun Peng[1,2]

[1]Department of Computer Science, University of Southern California
[2]Information Sciences Institute, University of Southern California
[3]Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign

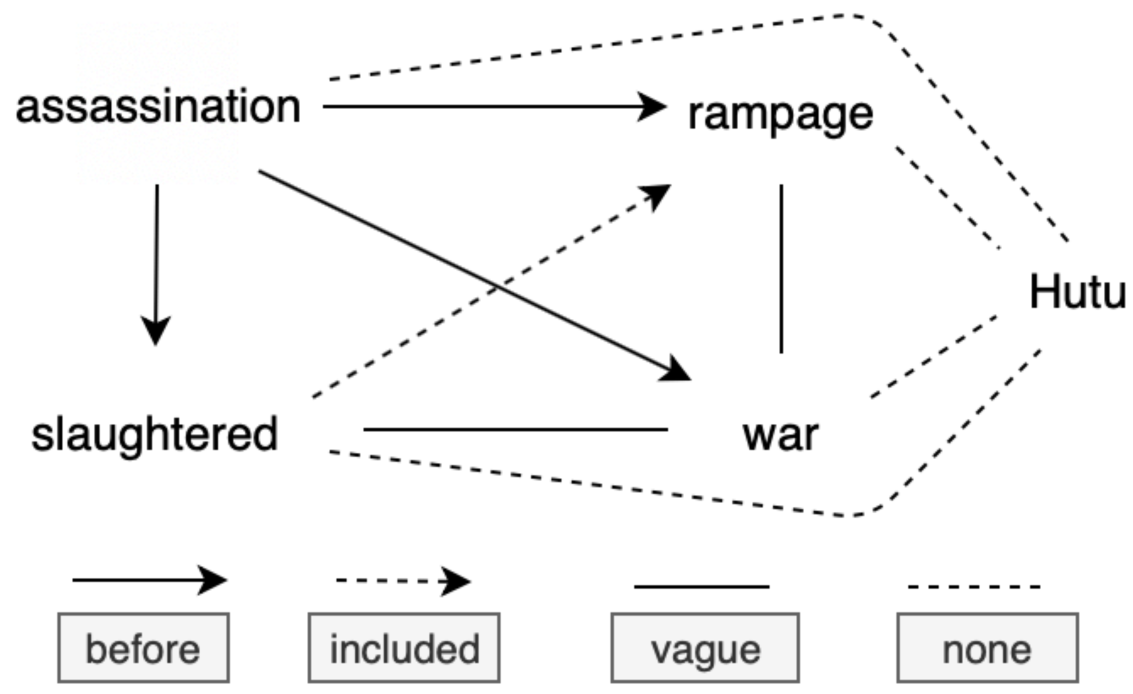
Introduction

The task of temporal relation extraction among events in a text can be modeled as building a graph for a given text, whose nodes represent events and edges are labeled with temporal relations correspondingly.

*The **assassination** touched off a murderous **rampage** by Hutu security forces and civilians , who **slaughtered** mainly Tutsis but also Hutus who favored reconciliation with the minority. It also reignited the civil **war**.*



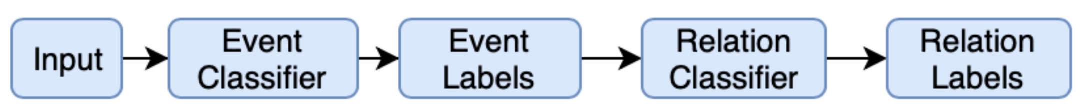
(a) Gold Temporal Relation Graph



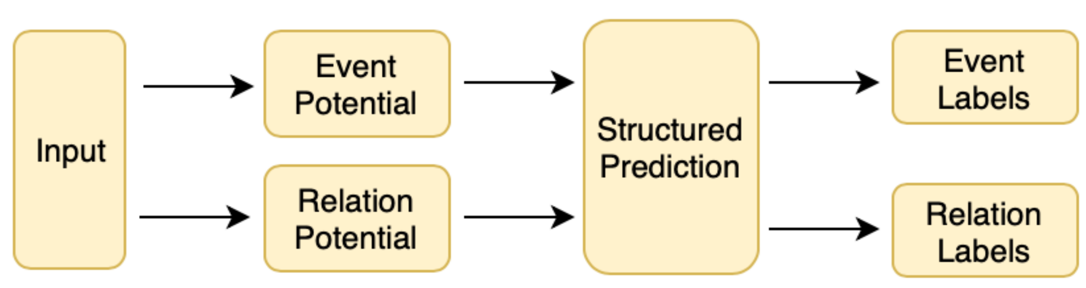
(b) End-to-end Temporal Relation Graph

Contributions

- Built a joint model that extracts both events and temporal relations simultaneously.
- Improved event representations by sharing the same contextualized embeddings and neural representation learner between the event extraction and temporal relation extraction modules.



(a) Pipeline Model



(b) Structured Joint Model

The two flow charts above show the advantage of our model over a traditional pipelined model.

Model Architecture

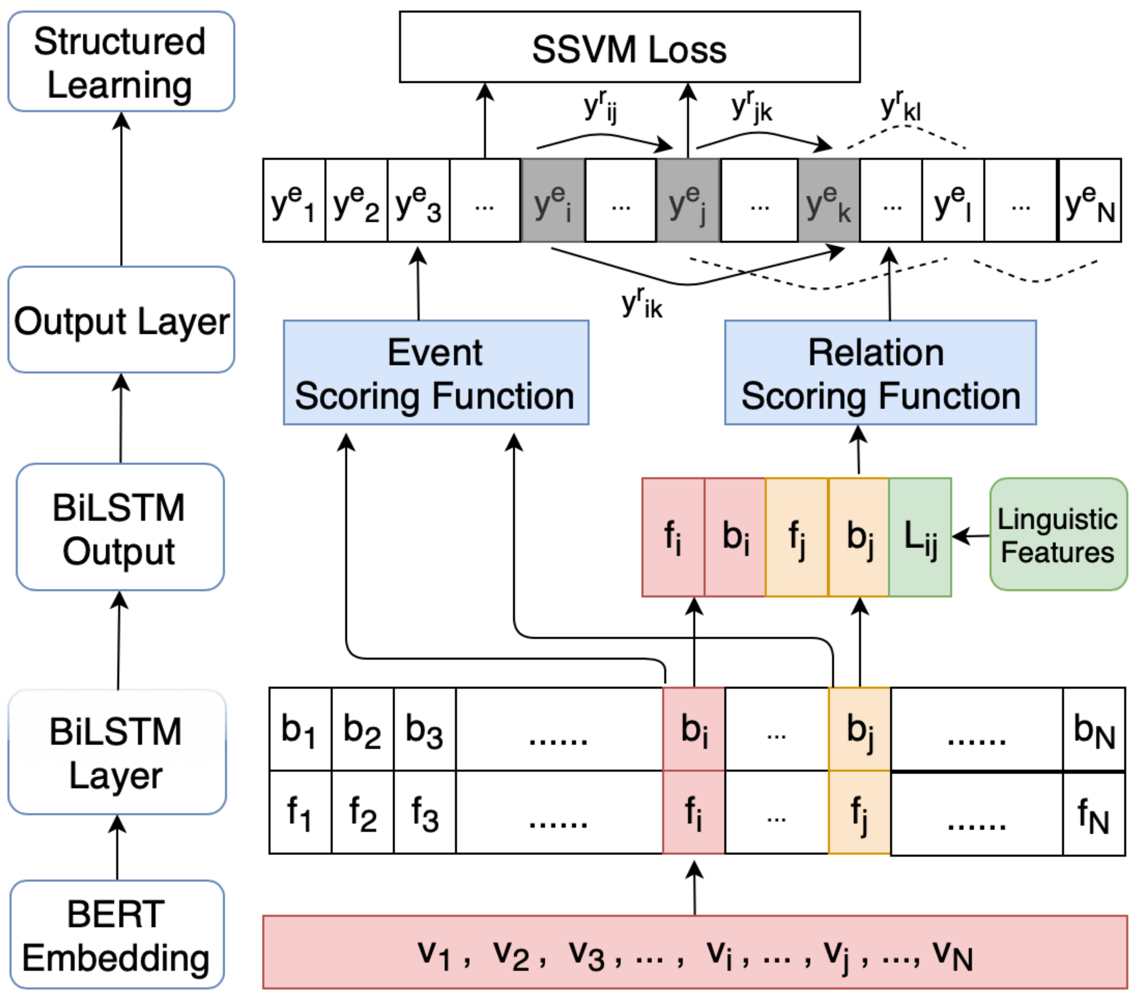


Figure: Deep neural network for joint structured learning

Structured Model Objective Function

$$\mathcal{L} = \sum_{n=1}^{l} \frac{C}{M^n}\Big[\max_{\hat{\boldsymbol{y}}^n\in\mathcal{Y}}\big(0, \Delta(\boldsymbol{y}^n, \hat{\boldsymbol{y}}^n) + \bar{S}^n_{\mathcal{R}} + C_{\mathcal{E}}\bar{S}^n_{\mathcal{E}}\big)\Big] + ||\Phi||^2$$

Learning - SSVM Loss

Details of the model objective function.

- $\bar{S}^n_{\mathcal{E}} = S(\hat{\boldsymbol{y}}^n_{\mathcal{E}}; \boldsymbol{x}^n) - S(\boldsymbol{y}^n_{\mathcal{E}}; \boldsymbol{x}^n)$
- $\bar{S}^n_{\mathcal{R}} = S(\hat{\boldsymbol{y}}^n_{\mathcal{R}}; \boldsymbol{x}^n) - S(\boldsymbol{y}^n_{\mathcal{R}}; \boldsymbol{x}^n)$
- Φ: model parameters
- $\boldsymbol{y}^n$, $\hat{\boldsymbol{y}}^n$: gold / predicted labels

Structured Inference

$$\hat{y} = \arg\max \sum_{(i,j)\in\mathcal{EE}}\sum_{r\in\mathcal{R}} y^r_{i,j} S(y^r_{i,j}, \boldsymbol{x}) + C_{\mathcal{E}}\sum_{k\in\mathcal{E}}\sum_{e\in\{0,1\}} y^e_k S(y^e_k, \boldsymbol{x})$$

$$y^r_{i,j}, y^e_k \in \{0,1\}\ ,\ \sum_{r\in\mathcal{R}} y^r_{i,j} = 1, \sum_{e\in\{0,1\}} y^e_k = 1$$

- $y^e_k, y^r_{i,j}$: binary indicators for event and relation.
- $S(y^e_k, \boldsymbol{x}), \forall e \in \{0,1\}$, $S(y^r_{i,j}, \boldsymbol{x}), \forall r \in \mathcal{R}$ scoring functions

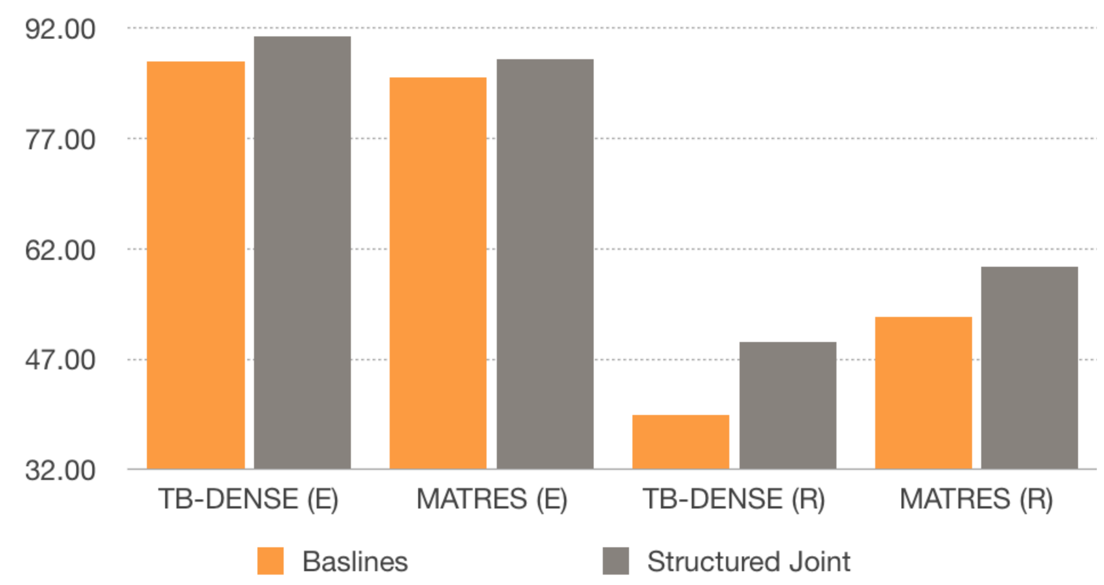
Global Constraints

- **Event-Relation Consistency** a pair of tokens have temporal relation if and only if both are events.
- **Symmetry** two pairs of events with flipping orders should have reversed relations.
- **Transitivity** if (i,j), (j,k) and (i,k) pairs exist, the relation prediction of (i,k) pair in the transitivity set specified by (i,j) and (j,k) pairs.

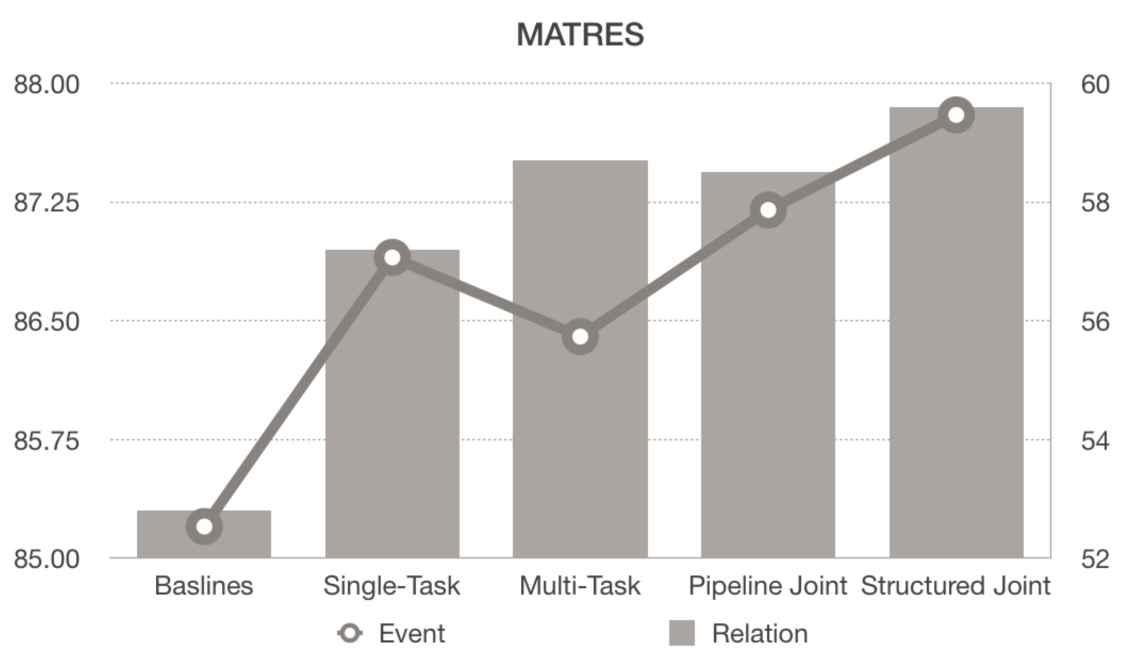
Data and Baselines

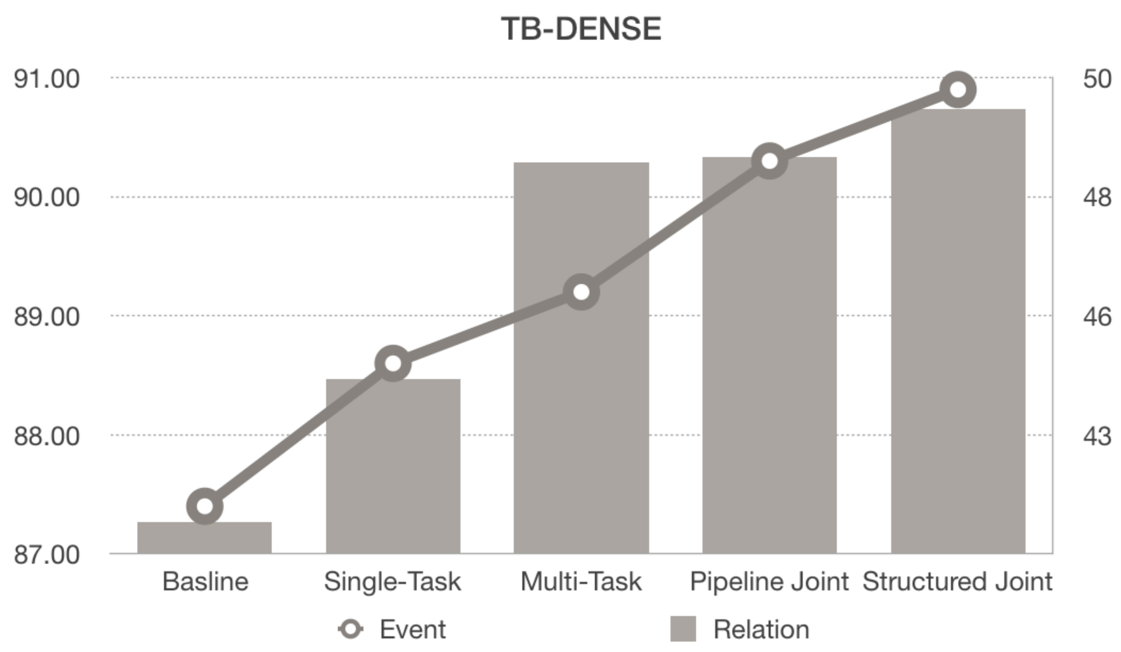
- **TB-Dense** (Chambers et al., 2014) Dense annotations over all event pairs within one or two sentences.
- **MATRES** (Ning at al., 2018) Multi-axis annotation scheme with starting point of events.

Overall Results



Ablation Study





Conclusion

We propose a novel neural structured model with joint representation learning to make predictions on events and relations simultaneously. Results show that the proposed model is effective for end-to-end event temporal relation extraction: we improve the performances of previously published systems by 10% and 6.8% on the TB-Dense and MATRES datasets.