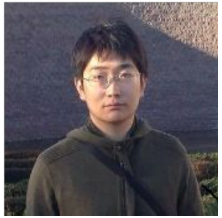
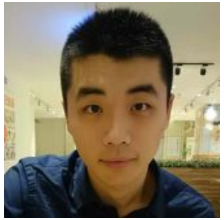
Joint Constrained Learning for Event-Event Relation Extraction

Haoyu Wang[1], Muhao Chen[1], Hongming Zhang[2*], Dan Roth[1]

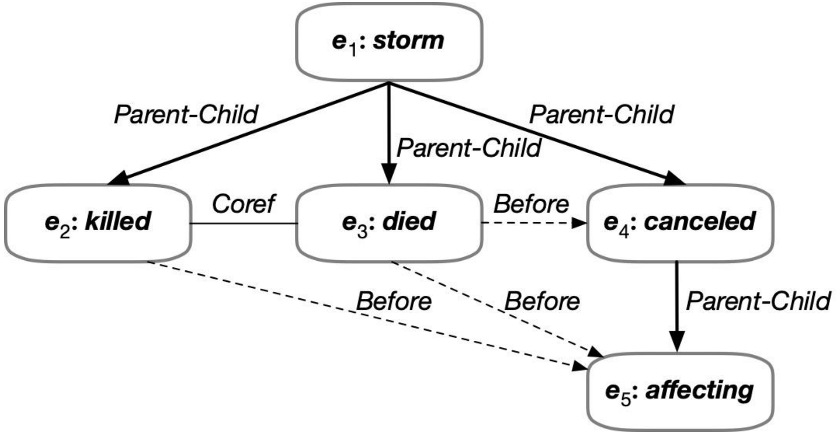
[1]Department of Computer and Information Science, UPenn
[2]Department of Computer Science and Engineering, HKUST

* This work was done when the author was visiting the University of Pennsylvania

Event-Event Relation Extraction

- Events are not simple, stand-alone predicates
  - Described at different granularities & may form complex structures
- Event-event relations are needed to induce such "event complex"
  - Temporal Relations
  - Membership
  - Coreference

On Tuesday, there was a typhoon-strength (***e_1:storm***) in Japan. One man got (***e_2:killed***) and thousands of people were left stranded. Police said an 81-year-old man (***e_3:died***) in central Toyama when the wind blew over a shed, trapping him underneath. Later this afternoon, with the agency warning of possible tornadoes, Japan Airlines (***e_4:canceled***) 230 domestic flights, (***e_5:affecting***) 31,600 passengers.



Problem Definition

Input: document, event triggers
Output: event-event relations

Tasks we study:

- Temporal Relations: Before, After, Equal, Vague
- Membership Relations: Parent-Child, Child-Parent, Coreference, NoRel

Challenges

- Lack of learning resources
  - No large-scale single resource contains annotation for these interrelated tasks
- Global consistency among interrelated predictions
  - Symmetry constraints
  - Transitivity constraints
  - Conjunction constraints

Problem Definition

e_2: *killed* e_3: *died* e_4: *canceled*

e_5: *affecting*

Input: document, event triggers
Output: event-event relations

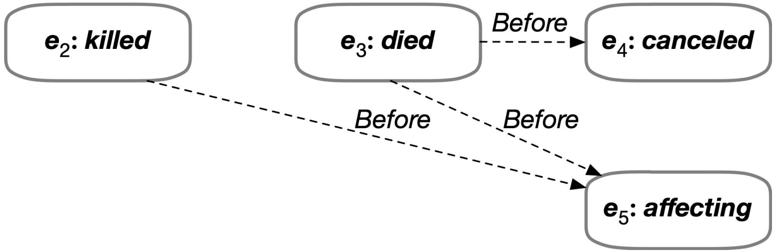
Tasks we study:

- Temporal Relations: Before, After, Equal, Vague
- Membership Relations: Parent-Child, Child-Parent, Coreference, NoRel

Challenges

- Lack of learning resources
  - No large-scale single resource contains annotation for these interrelated tasks
- Global consistency among interrelated predictions
  - Symmetry constraints
  - Transitivity constraints
  - Conjunction constraints

Problem Definition



Input: document, event triggers
Output: event-event relations

Tasks we study:

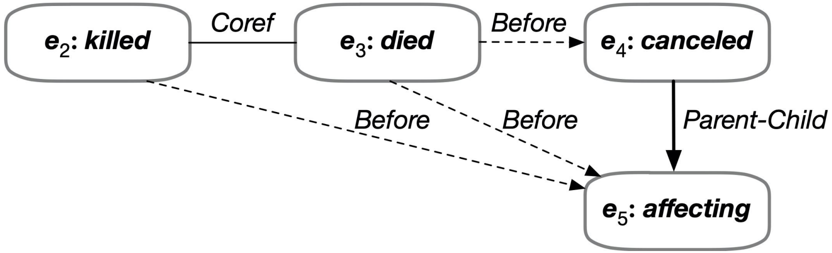
- Temporal Relations: Before, After, Equal, Vague
- Membership Relations: Parent-Child, Child-Parent, Coreference, NoRel

Challenges

- Lack of learning resources
  - No large-scale single resource contains annotation for these interrelated tasks
- Global consistency among interrelated predictions
  - Symmetry constraints
  - Transitivity constraints
  - Conjunction constraints

Problem Definition

Input: document, event triggers
Output: event-event relations



Tasks we study:

- Temporal Relations: Before, After, Equal, Vague
- Membership Relations: Parent-Child, Child-Parent, Coreference, NoRel

Challenges

- Lack of learning resources
  - No large-scale single resource contains annotation for these interrelated tasks
- Global consistency among interrelated predictions
  - Symmetry constraints
  - Transitivity constraints
  - Conjunction constraints

Constraints

Symmetry

e3:died is BEFORE ***e4:canceled***

=> ***e4:canceled*** is AFTER ***e3:died***

Transitivity

e1:storm is PARENT of ***e4:canceled***

\+ ***e4:canceled*** is a PARENT of ***e5:affecting***

=> ***e1:storm*** is a PARENT of ***e5:affecting***

Conjunction

e3:died is BEFORE ***e4:canceled***

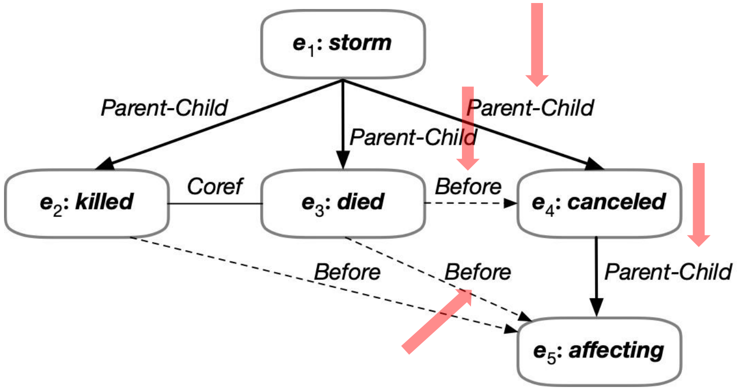
\+ ***e4:canceled*** is a PARENT of ***e5:affecting***

=> ***e3:died*** BEFORE ***e5:affecting***

α(e3, e4) = BEFORE

β(e4, e5) = Parent-Child

=> γ(e3, e5) = BEFORE



α \ β	PC	CP	CR	NR	BF	AF	EQ	VG
PC	PC, ¬AF	–	PC, ¬AF	¬CP, ¬CR	BF , ¬CP, ¬CR	–	BF , ¬CP, ¬CR	–
CP	–	CP, ¬BF	CP, ¬BF	¬PC, ¬CR	–	AF, ¬PC, ¬CR	AF, ¬PC, ¬CR	–
CR	PC, ¬AF	CP, ¬BF	CR, EQ	NR	BF , ¬CP, ¬CR	AF, ¬PC, ¬CR	EQ	VG
NR	¬CP, ¬CR	¬PC, ¬CR	NR	–	–	–	–	–
BF	BF , ¬CP, ¬CR	–	BF , ¬CP, ¬CR	–	BF , ¬CP, ¬CR	–	BF , ¬CP, ¬CR	¬AF, ¬EQ
AF	–	AF, ¬PC, ¬CR	AF, ¬PC, ¬CR	–	–	AF, ¬PC, ¬CR	AF, ¬PC, ¬CR	¬BF , ¬EQ
EQ	¬AF	¬BF	EQ	–	BF , ¬CP, ¬CR	AF, ¬PC, ¬CR	EQ	VG, ¬CR
VG	–	–	VG, ¬CR	–	¬AF, ¬EQ	¬BF , ¬EQ	VG	–

Constrained Learning

Earlier work: Incorporate external knowledge as hard and soft constraints via Constrained Conditional Models (CCMs, Chang et al., 2012); usually post-learning correction

$$\hat{\mathcal{I}} = \operatorname*{argmax}_{\mathcal{I}} \sum_{ij \in \mathcal{E}} \sum_{r \in R} (x_r(ij) + \lambda f_r(ij)) \mathcal{I}_r(ij)$$

$$\text{s.t.} \quad \underbrace{\Sigma_r \mathcal{I}_r(ij) = 1}_{\text{(uniqueness)}}, \ \underbrace{\mathcal{I}_r(ij) = \mathcal{I}_{\bar{r}}(ji)}_{\text{(symmetry)}},$$

$$\underbrace{\mathcal{I}_{r_1}(ij) + \mathcal{I}_{r_2}(jk) - \Sigma_{m=1}^{M} \mathcal{I}_{r_3^m}(ik) \leq 1}_{\text{(transitivity)}},$$

Constrained learning (Li et al., EMNLP'19): convert declarative rules to differentiable learning objectives using t-norm

- L_A Annotation Loss: $\top \rightarrow r(e_1, e_2)$ ➡️ $-w_r \log r_{(e_1,e_2)}$
- L_S Symmetric Loss: $\alpha(e_1, e_2) \leftrightarrow \bar{\alpha}(e_2, e_1)$ ➡️ $|\log \alpha_{(e_1,e_2)} - \log \bar{\alpha}_{(e_2,e_1)}|$
- L_C Conjunctive Loss: $\alpha(e_1, e_2) \wedge \beta(e_2, e_3) \rightarrow \gamma(e_1, e_3)$ ➡️ $\log \alpha_{(e_1,e_2)} + \log \beta_{(e_2,e_3)} - \log \gamma_{(e_1,e_3)}$

  $\alpha(e_1, e_2) \wedge \beta(e_2, e_3) \rightarrow \neg\delta(e_1, e_3)$ ➡️ $\log \alpha_{(e_1,e_2)} + \log \beta_{(e_2,e_3)} - \log(1 - \delta_{(e_1,e_3)})$
- Training Objective: $L = L_A + \lambda_S L_S + \lambda_C L_C$

Exact Inference at decision time via CCMs

- Formulate as an ILP problem: Gurobi

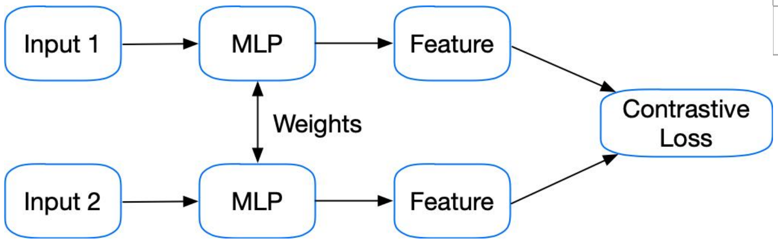
Commonsense Knowledge

- TemProb (Ning et al. 2018a)
  - Source: New York Times 1987-2007 (#Articles~1M)
  - Preprocessing: Semantic Role Labeling & Temporal relation model
  - 51K semantic frames, 80M relations
- ConceptNet
  - "HasSubevent", "HasFirstSubevent"
  - 30K Positive & 30K Negative

en kill people — HasSubevent ⟶ en attack
Weight: 1.0

en commit murder — HasFirstSubevent ⟶ en find a reason
Weight: 2.0

- MLP Encoders



More than 10 people ***died*** on their way to the nearest hospital, police said. A suicide car bomb ***exploded*** on Friday in the middle of a group of men playing volleyball in northwest Pakistan.

Frame1	Frame2	Before	After
concern	protect	92%	8%
conspire	kill	95%	5%
fight	overthrow	92%	8%
accuse	defend	92%	8%
crash	die	97%	3%
explode	die	83%	17%

Training Process

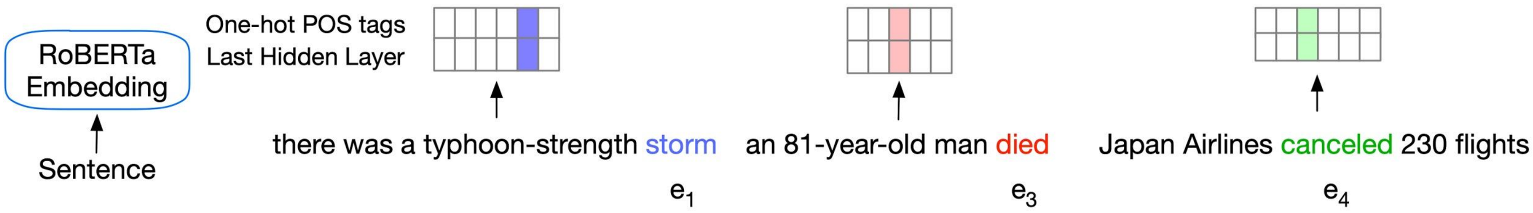
Input sentences

On Tuesday, there was a typhoon-strength (e_1:***storm***) in Japan. One man got (e_2:***killed***) and thousands of people were left stranded. Police said an 81-year-old man (e_3:***died***) in central Toyama when the wind blew over a shed, trapping him underneath. Later this afternoon, with the agency warning of possible tornadoes, Japan Airlines (e_4:***canceled***) 230 domestic flights, (e_5:***affecting***) 31,600 passengers.

Sentence	there was a typhoon-strength storm	an 81-year-old man died	Japan Airlines canceled 230 flights
	e_1	e_3	e_4

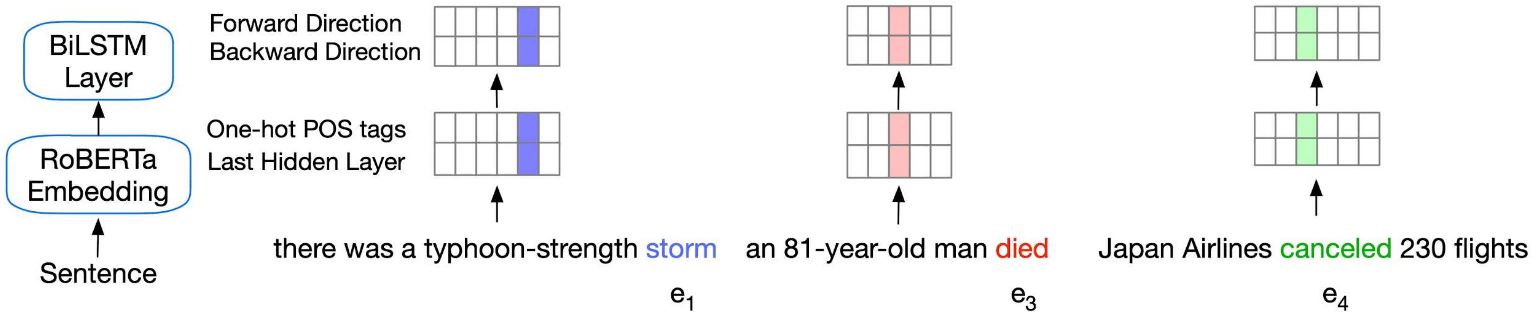
Training Process

Contextualized Representation



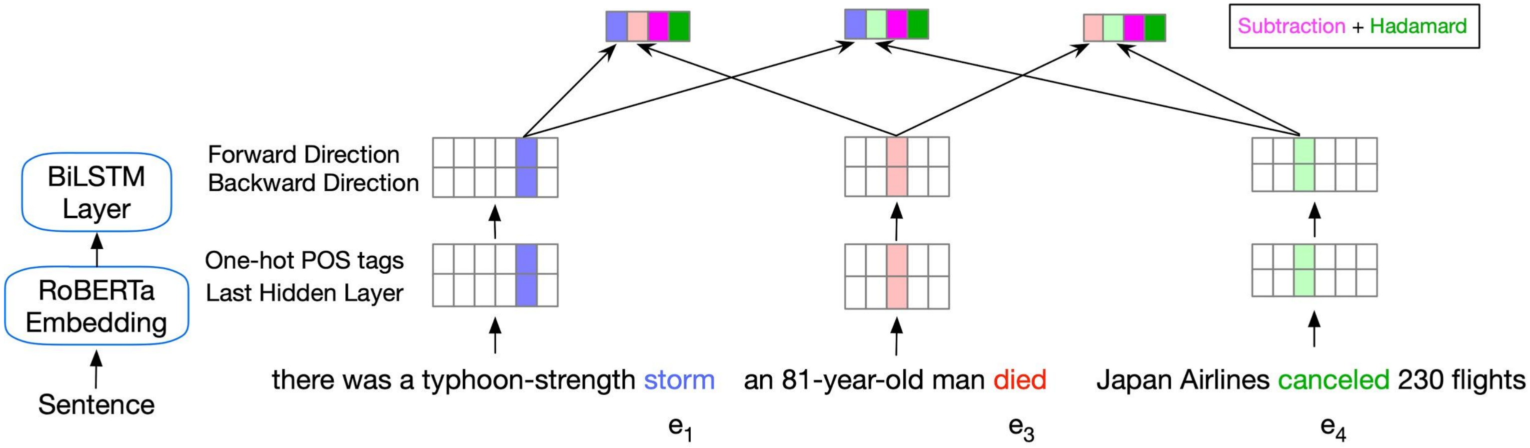
Training Process

BiLSTM Encoder



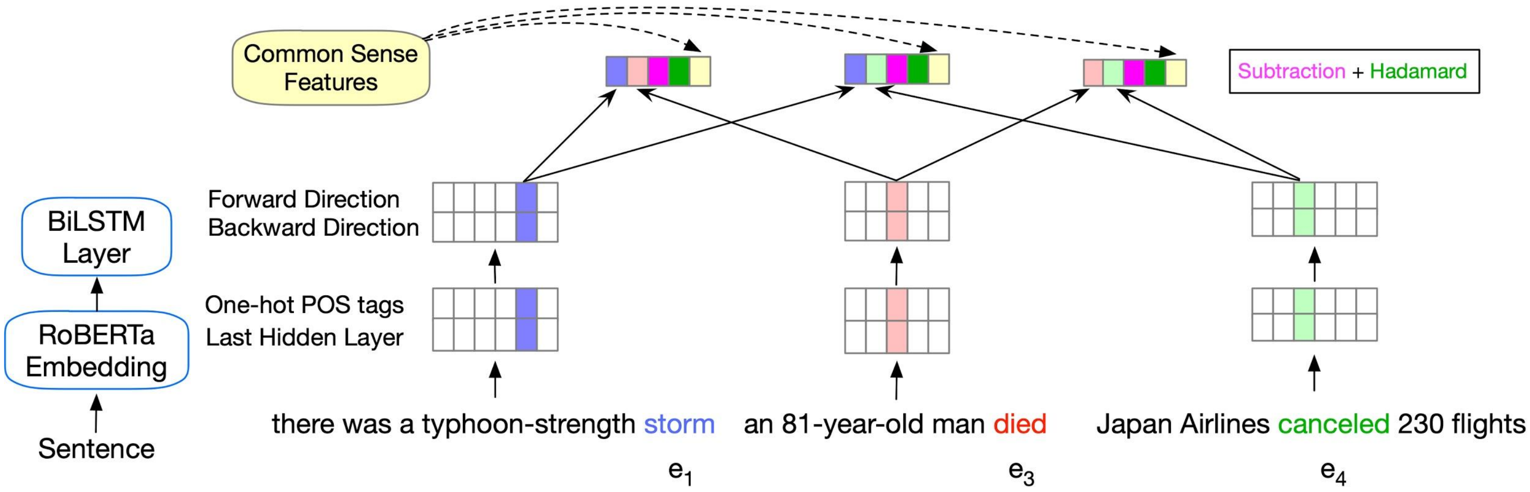
Training Process

Event Pair Representation



Training Process

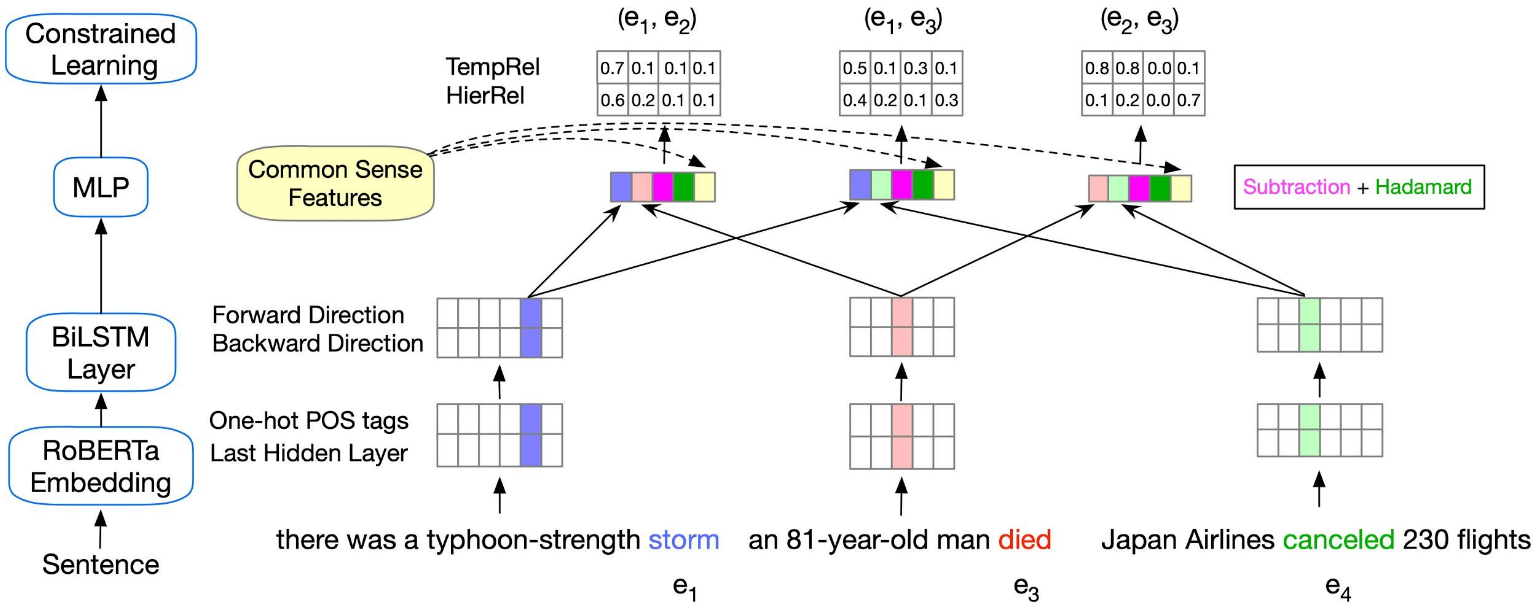
Adding Common Sense Features



Training Process

Constrained Learning

Loss Function: $L = L_A + \lambda_S L_S + \lambda_C L_C$



Evaluation on Benchmarks

Benchmark Datasets

- Membership relations: HiEve
- Temporal relations: MATRES
- Case study: RED

Dataset Statistics

	HiEve	MATRES	RED
# of Documents			
Train	80	183	-
Dev	-	72	-
Test	20	20	35
# of Pairs			
Train	35001	6332	-
Test	7093	827	1718

- The proposed method surpasses the SOTA TempRel extraction method on MATRES by relatively 3.27% F_1.

Model	P	R	F_1
CogCompTime (Ning et al., 2018c)	0.616	0.725	0.666
Perceptron (Ning et al., 2018b)	0.660	0.723	0.690
BiLSTM+MAP (Han et al., 2019b)	-	-	0.755
LSTM+CSE+ILP (Ning et al., 2019)	0.713	0.821	0.763
Joint Constrained Learning (ours)	**0.734**	**0.850**	**0.788**

- It also offers promising performance on the HiEve dataset for subevent relation extraction, relatively surpassing previous methods by at least 3.12% in F_1.

	F_1 score		
Model	PC	CP	Avg.
StructLR (Glavaš et al., 2014)	0.522	**0.634**	0.577
TacoLM (Zhou et al., 2020a)	0.485	0.494	0.489
Joint Constrained Learning (ours)	**0.625**	0.564	**0.595**

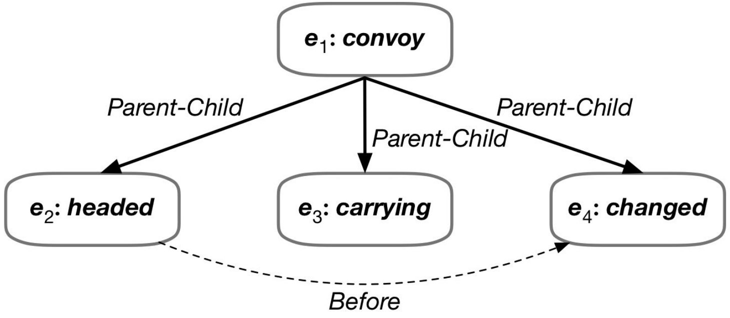
Case Study on RED dataset

Mapping for relations

Original labels in RED	Mapped labels
BEFORE, BEFORE/CAUSES, BEFORE/PRECONDITION, ENDS-ON, OVERLAP/PRECONDITION	BEFORE
SIMULTANEOUS	EQUAL
OVERLAP, REINITIATES	VAGUE
CONTAINS, CONTAINS-SUBEVENT	PARENT-CHILD & BEFORE
BEGINS-ON	AFTER

Example Output

A (***e1:convoy***) of 280 Russian trucks (***e2:headed***) for Ukraine, which Moscow says is (***e3:carrying***) relief goods for war-weary civilians, has suddenly (***e4:changed***) course, according to a Ukrainian state news agency.



Performance (F_1)

Model	TEMPREL	SUBEVENT
Joint Constrained Learning (ours)	**0.72**	**0.54**

Ablation Study

- Single-task Training vs Joint Training
- Constrained Learning
  - Task-specific constraints
  - Cross-task constraints
- Commonsense Knowledge
- Post-learning Correction: ILP

Model	SUBEVENT P	SUBEVENT R	SUBEVENT F_1	TEMPREL P	TEMPREL R	TEMPREL F_1
Single-task Training	32.5	**73.1**	45.0	67.7	80.3	73.5
Joint Training	50.4	43.1	46.5	68.4	82.0	74.6
+ Task-specific constrained learning	51.6	59.7	55.4	71.3	82.7	76.6
+ Cross-task constrained learning	51.1	67.0	58.0	72.2	83.8	77.6
+ Commonsense knowledge	56.9	61.6	59.2	73.3	84.2	78.4
+ Global inference (ILP)	**57.4**	61.7	**59.5**	**73.4**	**85.0**	**78.8**
All but constrained learning	54.2	41.8	47.2	72.1	80.8	76.2

Conclusion

Thanks!

- A new paradigm for joint constrained learning
  - Bridge temporal and subevent relations with a comprehensive set of logical constraints
  - Convert constraints into differentiable objective functions
- Address the lack of jointly annotated data
  - Complementary supervision signals from two tasks
- Outperform previous methods on benchmark datasets
  - MATRES: brings about 3.27% F_1 improvement
  - HiEve: brings about 3.12% F_1 improvement
- Present promising results on RED
- Generalizable to other relations / constraints