An Improved Neural Baseline for Temporal Relation Extraction

Qiang Ning[1,3], Sanjay Subramanian[2,3], Dan Roth[1,2]
[1]Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign
[2]Department of Computer and Information Science, University of Pennsylvania
[3]Allen Institute for Artificial Intelligence

Task Definition

A temporal relation describes the relation between two events with respect to time.

BEFORE

I met with him before leaving for Paris.

Setup: We assume that gold events (e.g., "met" and "leaving") are provided in this work.

Labels: before, after, simultaneous, and vague.

Previous state of the art: CogCompTime [EMNLP'18]

Highlight of this work: 10% absolute improvement in accuracy (25% error reduction) by

- A neuralized model
- Contextualized word-embeddings
- A Siamese network encoding common sense

Recent Progress on Temporal Relation Extraction

Transitivity constraints (specifying a structure):
$$A \rightarrow B, B \rightarrow C \Rightarrow A \rightarrow C$$
Highly interrelated and the decision of a relation often depends on other events.

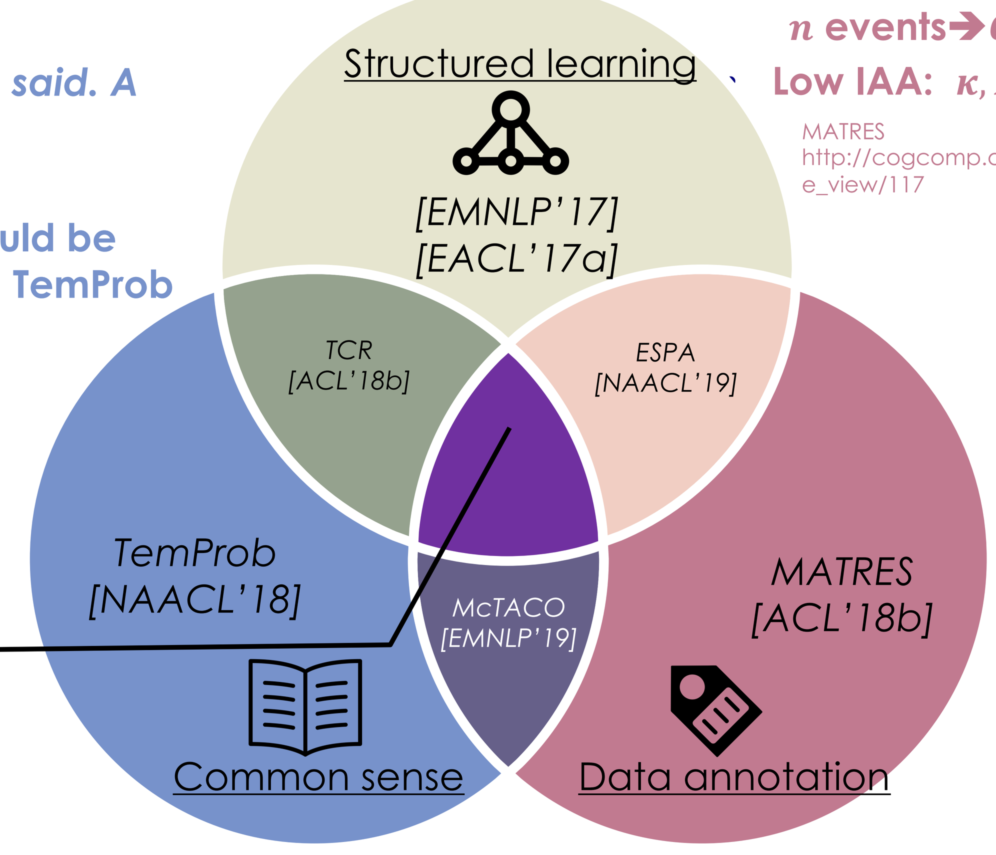
- Labor intensive: n events → $O(n^2)$ pairs
- Low IAA: $\kappa, F_1 \approx 60\%$

MATRES http://cogcomp.org/page/resource_view/117

"More than 10 people have (E1: VBN), police said. A car (E2: VBD) on Friday in a group of men."

What's the relation between E1 and E2? It would be easier if we knew E1=died and E2=exploded. TemProb aims at providing typical temporal ordering knowledge.

TemProb http://cogcomp.org/page/resource_view/114



CogCompTime [EMNLP'18]: A feature-based system that incorporates structured learning, TemProb, and MATRES.
CogCompTime https://github.com/qiangning/CogCompTime

This work: An extension of CogCompTime, which combines neural models with more recent contextualized word-embeddings, and a Siamese common-sense encoder (CSE).
https://cogcomp.seas.upenn.edu/page/publication_view/879

TemProb: Typical Temporal Ordering

When the verbs are missing, it's very difficult even for humans to figure out the relation. However, if we know that E1=died, and E2=exploded, it's obvious that E2->E1 due to our prior knowledge about these verbs.

Example pairs		Before (%)	After (%)
Accept	Determine	42	26
Ask	Help	86	9
Attend	Schedule	1	82
Accept	Propose	10	77
Die	Explode	14	83

TemProb is a probabilistic knowledge base that provides *typical temporal ordering* between verbs (i.e., temporal ordering common sense). CogCompTime adopts the statistics found in TemProb as an additional.

However, TemProb is a simple counting model and fails (or is unreliable) for unseen (or rare) tuples. For example, we may see (ambush, die) less frequently than (attack, die) in a corpus, and the observed frequency of (ambush, die) being before or after is thus less reliable. However, since "ambush" is semantically similar to "attack", the statistics of (attack, die) can actually serve as an auxiliary signal to (ambush, die).

MATRES: Multi-Axis Modeling

Existing data annotation schemes

Scheme 1: General graph modeling
- E.g., TimeBank
- No restrictions on modeling
- Relations are inevitably missed

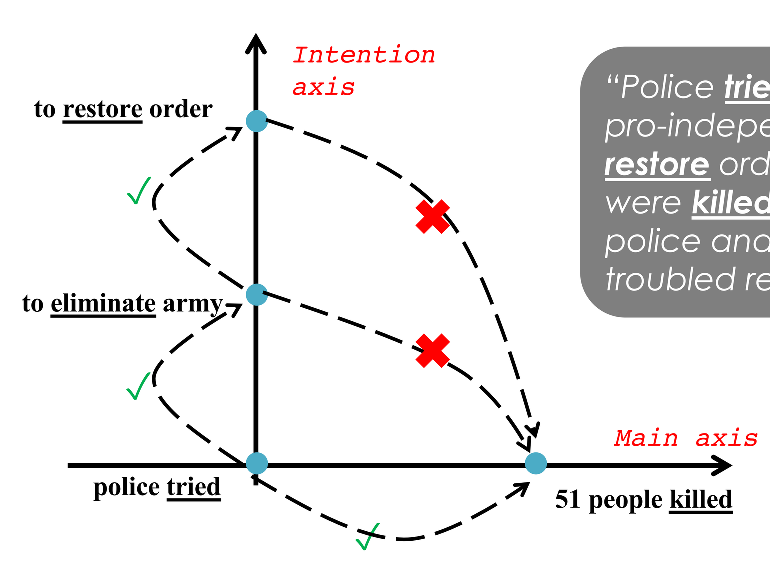
Scheme 2: Chain modeling
- E.g., TimeBank-Dense [ACL'14]
- A strong restriction on modeling
- Any pair is comparable
- But many are confusing

Our approach is a balance between these two, called *Multi-axis modeling:*

We also allow dense modeling, but only within a same axis.

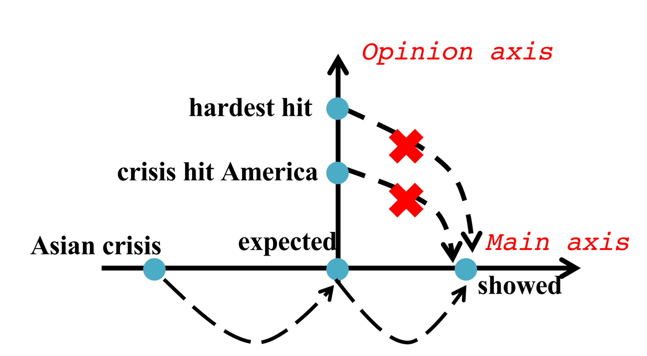
MATRES data statistics

	Purpose	#Doc	#Events	#TempRels
TB+AQ	Train	255	8K	13K
PT	Test	20	537	837
TCR	Test	25	1.3K	2.6K

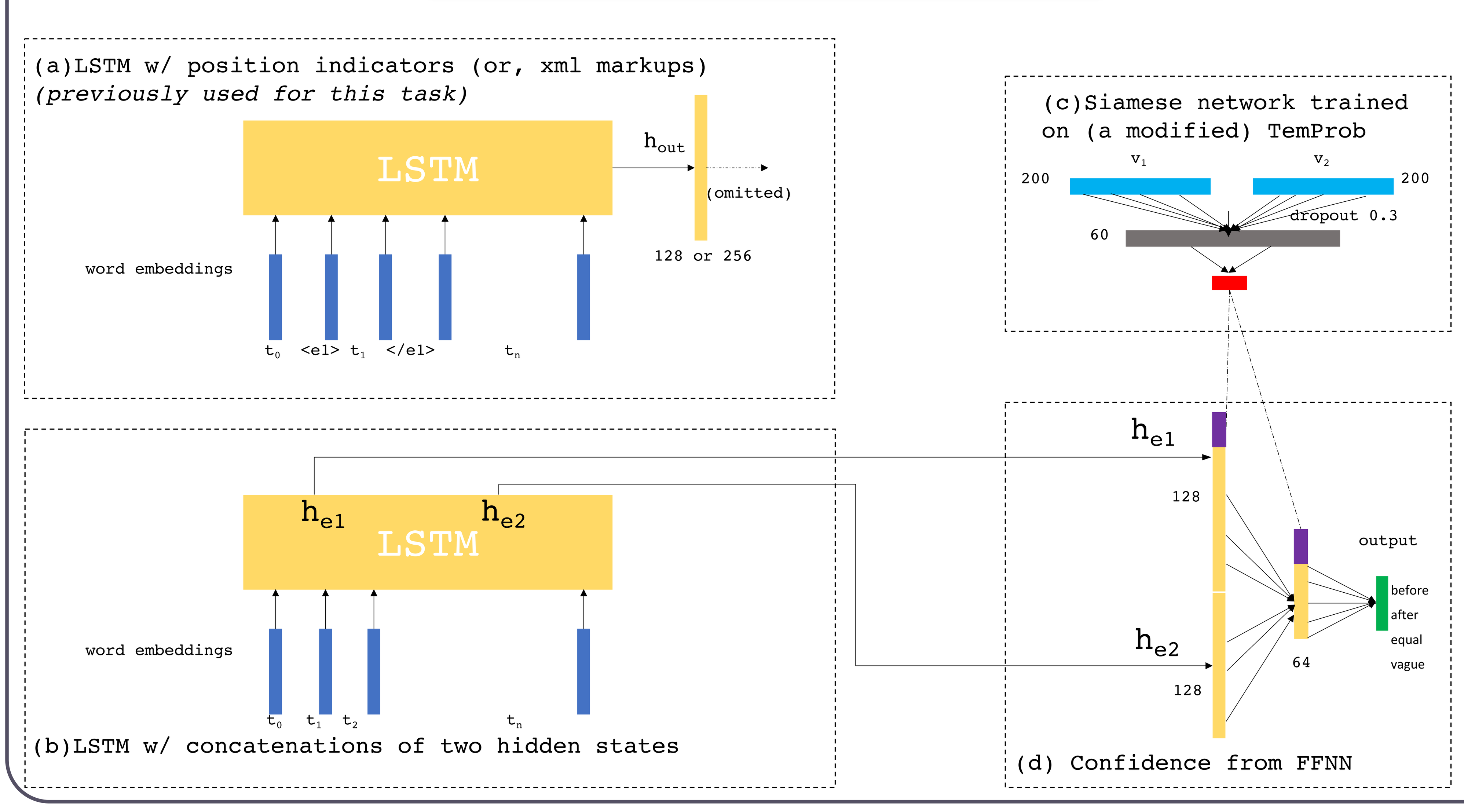


"Police tried to eliminate the pro-independence army and restore order. At least 51 people were killed in clashes between police and citizens in the troubled region."

"Service industries showed solid job gains, as did manufacturers, two areas expected to be hardest hit when the effects of the Asian crisis hit the American economy."



Proposed Neural Structure

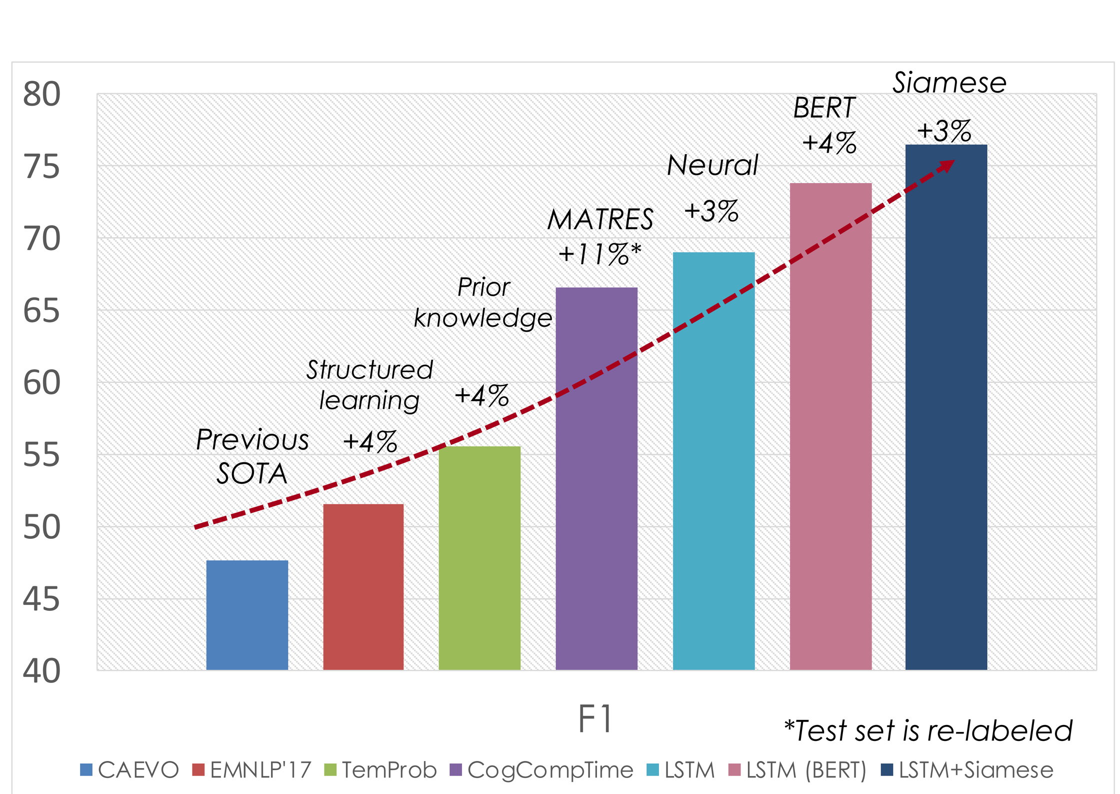


Results on MATRES

System	Emb.	Acc.	F_1	F_{aware}	Avg.
P.I.	word2vec	63.2	67.6	60.5	63.8
	GloVe	64.5	69.0	61.1	64.9
	FastText	60.5	64.7	59.5	61.6
	ELMo	67.5	73.9	63.0	**68.1**
	BERT	68.8	73.6	61.7	68.0
Concat	word2vec	65.0	69.5	59.4	64.6
	GloVe	64.9	69.5	60.9	65.1
	FastText	64.0	68.6	60.1	64.2
	ELMo	67.7	74.0	63.3	68.3
	BERT	69.1	74.4	63.7	**69.1**
Concat+CSE	ELMo	71.7	76.7	66.0	**71.5**
	BERT	71.3	76.3	66.5	71.4
CogCompTime	-	61.6	66.6	60.8	**63.0**

- Concat (i.e., the (b) network on the left) is generally better than position indicator (P.I.; the (a) network).
- Contextualized embeddings expectedly improved over conventional embeddings.
- Improved over CogCompTime significantly.

Summary of Recent Progress



Conclusion

Temporal relation extraction has long been an important yet challenging task in natural language processing. Lack of high-quality data and difficulty in the learning problem defined by previous annotation schemes inhibited performance of neural-based approaches. The discoveries that LSTMs readily improve the feature-based state of the art, CogCompTime, on the MATRES and TCR datasets by a large margin not only give the community a strong baseline, but also indicate that the learning problem is probably better defined by MATRES and TCR. Therefore, we should move along that direction to collect more high-quality data, which can facilitate more advanced learning algorithms in the future.

References

[NIPS'94] J. Bromley, I. Guyon, Y. LeCun, E. Sackinger, and R. Shah. Signature verification using a "siamese" time delay neural network. **[ACL'17a]** F. Cheng and Y. Miyao. Classifying temporal relations by bidirectional LSTM over dependency paths.**[ACL'17b]** J. Tourille, O. Ferret, A. Neveol, and X. Tannier. Neural architecture for temporal relation extraction: A bi-lstm approach for detecting narrative containers. **[EMNLP'17]** Q, Ning, Z. Feng, and D. Roth. A Structured Learning Approach to Temporal Relation Extraction. **[EACL'17a]** A. Leeuwenberg and M.-F. Moens. Structured learning for temporal relation extraction from clinical records. **[EACL'17b]** D. Dligach, T. Miller, C. Lin, S. Bethard, and G. Savova. Neural temporal relation extraction. **[BioNLP'17]** C. Lin, T. Miller, D. Dligach, S. Bethard, and G. Savova. Representations of time expressions for temporal relation extraction with convolutional neural networks. **[NAACL'18]** Q. Ning, H. Wu, H. Peng, and D. Roth. Improving Temporal Relation Extraction with a Globally Acquired Statistical Resource. **[ACL'18a]** Q. Ning, Z. Feng, H. Wu, and D. Roth. Joint reasoning for temporal and causal relations. **[ACL'18b]** Q. Ning, H. Wu, and D. Roth. A multi-axis annotation scheme for event temporal relations. **[ACL'18c]** Y. Meng and A. Rumshisky. Context-aware neural model for temporal information extraction. **[EMNLP'18]** Q. Ning, B. Zhou, Z. Feng, H. Peng, and D. Roth. CogCompTime: A tool for understanding time in natural language. **[NAACL'19]** Q. Ning, H. He, C. Fan, and D. Roth. Partial or Complete, That's The Question. **[EMNLP'19]** B. Zhou, D. Khashabi, Q. Ning and D. Roth. "Going on a vacation" takes longer than "Going for a walk": A Study of Temporal Commonsense Understanding.

Acknowledgement

The IBM-ILLINOIS Center for Cognitive Computing Systems Research (C3SR) - a research collaboration as part of the IBM AI Horizons Network; Contract HR0011-18-2-0052 with the US Defense Advanced Research Projects Agency (DARPA).

ILLINOIS Penn University of Pennsylvania Cognitive Computation Group AI2 Allen Institute for Artificial Intelligence