UNIVERSITY of PENNSYLVANIA

C3SR center for cognitive computing systems research

IBM | ILLINOIS

JOINT REASONING FOR TEMPORAL AND CAUSAL RELATIONS




Qiang Ning, Zhili Feng, Hao Wu, Dan Roth

07/18/2018

University of Illinois, Urbana-Champaign & University of Pennsylvania

Time is Important

- Understanding time is key to understanding events
  - Timelines (in stories, clinical records), time-slot filling, Q&A, common sense
- *[June, 1989] Chris Robin lives in England and he is the person that you read about in Winnie the Pooh. As a boy, Chris lived in Cotchfield Farm. When he was three, his father wrote a poem about him. His father later wrote Winnie the Pooh in 1925.*
  - Where did Chris Robin live? Clearly, time sensitive.
  - When was Chris Robin born? poem [Chris at age 3] $\xrightarrow{\text{before}}$ Winnie the Pooh [1925]
    - Based on text: <=1922 (Wikipedia: 1920)
  - Requires identifying **relations** between events, and temporal reasoning.
- Temporal relation extraction
  - Events are associated with time intervals: $[t^1_{start}, t^1_{end}], [t^2_{start}, t^2_{end}]$
  - "A" happens BEFORE/AFTER "B"; "Time" is often expressed **implicitly**
  - 2 explicit time expressions per 100 tokens, but **12 temporal relations**

EXAMPLE

- More than 10 people ***(e1:   )***, he said. A car ***(e2:   )*** Friday in the middle of a group of men playing volleyball.

- Temporal question: Which one happens first?
  - ”e1” appears first in text. Is it also earlier in time?
  - “e2” was on “Friday”, but we don’t know when “e1” happened.
  - No explicit lexical markers, e.g., “before”, “since”, or “during”.

EXAMPLE: TEMPORAL DETERMINED BY CAUSAL

- More than 10 people ***(e1: died)***, he said. A car ***(e2: exploded)*** Friday in the middle of a group of men playing volleyball.

- Temporal question: Which one happens first?
- Obviously, “e2:exploded” is the <u>cause</u> and “e1:died” is the <u>effect</u>.
- So, “e2” happens first.

- In this example, the <u>temporal</u> relation is determined by the <u>causal</u> relation.
- Note also that the **lexical information** is important here; it’s likely that **explode** BERORE **die,** irrespective of the context.

Example: Causal determined by temporal

- People ***raged*** and took to the street the government ***stifled*** protesters.
- Causal question:
  - Did the government stifle people <u>because</u> people raged?
  - Or, people raged <u>because</u> the government stifled people?
  - Both sound correct and we are not sure about the causality here.

Example: Causal determined by temporal

- People ***raged*** and took to the street (after) the government ***stifled*** protesters.
- Causal question:
  - Did the government stifle people <u>because</u> people raged?
  - Or, people raged <u>because</u> the government stifled people?
  - Since "stifled" happened earlier, it's obvious that the cause is "stifled" and the result is "raged".
- In this example, the <u>causal</u> relation is determined by the <u>temporal</u> relation.

THIS PAPER

- **Event relations**: an essential step of event understanding, which supports applications such as story understanding/completion, summarization, and timeline construction.
  - [There has been a lot of work on this; see Ning et al. ACL'18, presented yesterday. for a discussion of the literature and the challenges.]

- This paper focuses on the joint extraction of **temporal** and **causal** relations.
  - A **temporal relation (T-Link)** specifies the relation between two events along the temporal dimension.
    - Label set: before/after/simultaneous/…
  - A **causal relation (C-Link)** specifies the [cause – effect] between two events.
    - Label set: causes/caused_by

- **T-Link Example:** *John* ***worked*** *out after* ***finishing*** *his work.*
- **C-Link Example:** *He was* ***released*** *due to* ***lack*** *of evidence.*

- Temporal and causal relations **interact** with each other.
  - For example, there is also a T-Link between ***released*** and ***lack***

- The decisions on the T-Link type and the C-link type depend on each other, suggesting that joint reasoning could help.

Related Work

- Obviously, temporal and causal relations are closely related (we're not the first who discovered this).
- NLP researchers have also started paying attention to this direction recently.
  - **CaTeRs**: Mostafazadeh et al. (2016) proposed an *annotation* framework, CaTeRs, which captured both temporal and causal aspects of event relations in common sense stories.
  - **CATENA**: Mirza and Tonelli (2016) proposed to extract both temporal and causal relations, but only by "*post-editing*" temporal relations based on causal predictions.
  - …

CONTRIBUTIONS

1. Proposed a novel joint inference framework for temporal and causal reasoning
   - Assume the availability of a temporal extraction system and a causal extraction system
   - Enforce declarative constraints originating from the physical nature of causality

2. Constructed a new dataset with both temporal and causal relations.
   - We augmented the EventCausality dataset (Do et al., 2011), which comes with causal relations, with new temporal annotations.

Temporal Relation Extraction: An ILP Approach [Do et al. EMNLP'12]

- Notations
  - $\mathcal{E}$--Event node set. $i, j, k \in \mathcal{E}$ are events.
  - $r \in \mathcal{R}$--temporal relation label
  - $\boldsymbol{I_r(ij)}$—Boolean variable – is there a of relation r between i *and* j? (Y/N)
  - $\boldsymbol{f_r(ij)}$--score of event pair (i, j) having relation r

Global assignment of relations:

$$\hat{I} = arg\max_{I} \sum_{ij \in \mathcal{E}} \sum_{r \in \mathcal{R}} f_r(ij) I_r(ij)$$

The sum of all softmax scores in this document

$$such\ that\ \forall i, j, k \in \mathcal{E}, \forall r_1, r_2 \in \mathcal{R}$$

$$\sum_{r} I_r(ij) = 1$$

Uniqueness

$$I_{r1}(ij) + I_{r2}(jk) - I_{r3}(ik) \leq 1$$

Transitivity

r_3--the relation dictated by r_1 and r_2

- Notations
  - $\mathcal{E}$--Event node set. $i, j, k \in \mathcal{E}$ are events.
  - $r \in \mathcal{R}$--temporal relation label
  - $\boldsymbol{I_r(ij)}$—Boolean variable – is there a of relation r between i and j? (Y/N)
  - $\boldsymbol{f_r(ij)}$--score of event pair (i, j) having relation r
  - $c \in \mathcal{C}$--causal relation; with corresponding variables $\boldsymbol{J_c(ij)}$ and $\boldsymbol{h_c(ij)}$

Global assignment of T & C relations

$$\hat{I}, \hat{J} = arg \max_{I,J} \sum_{ij \in \mathcal{E}} \left(\sum_{r \in \mathcal{R}} f_r(ij) I_r(ij) + \boxed{\sum_{c \in \mathcal{C}} h_c(ij) J_c(ij)} \right)$$

$$such\ that\ \forall i, j, k \in \mathcal{E}, \forall r_1, r_2 \in \mathcal{R}$$

$$\sum_r I_r(ij) = 1$$

$$I_{r1}(ij) + I_{r2}(jk) - I_{r3}(ik) \leq 1$$

$$\boxed{J_{causes}(ij) \leq I_{before}(ij)}$$

The “causal” part

“Cause” must be before “effect”

COGNITIVE COMPUTATION GROUP

Scoring Functions

$$\hat{I} = arg\max_{I} \sum_{ij \in \mathcal{E}} \left(\sum_{r \in \mathcal{R}} f_r(ij) I_r(ij) + \sum_{c \in \mathcal{C}} h_c(ij) J_c(ij) \right)$$

- Two scoring functions are needed in the objective above
  - $\boldsymbol{f_r(ij)}$--score of event pair (i, j) having temporal relation r
  - $\boldsymbol{h_c(ij)}$--score of event pair (i, j) having causal relation c
- Scoring functions
  - We use the soft-max scores from temporal/causal classifiers (or the log of the soft-max scores)
  - Choose your favorite model for the classifiers; here: sparse averaged perceptron
  - Features for a pair of events:
    - POS, token distance
    - modal verbs in-between (i.e., will, would, can, could, may and might)
    - temporal connectives in-between (e.g., before, after and since)
    - Whether the two verbs have a common synonym from their synsets in WordNet
    - The head word of the preposition phrase that covers each verb

Can we use more than just this "local" information?

Back to the Example: Temporal determined by causal

- More than 10 people ***(e1: died)***, he said. A car ***(e2: exploded)*** Friday in the middle of a group of men playing volleyball.

- Temporal question: Which one happens first?
- Obviously, "e2:exploded" is the cause and "e1:died" is the effect.
- So, "e2" happens first.

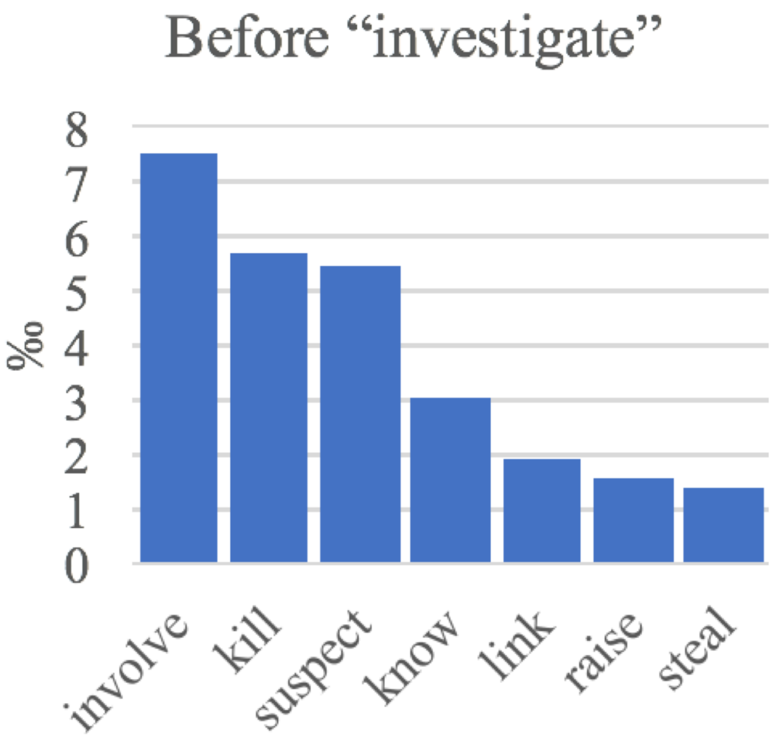
- In this example, the temporal relation is determined by the causal relation.
- Note also that the **lexical information** is important here; it's likely that **explode** BERORE **die,** irrespective of the context.

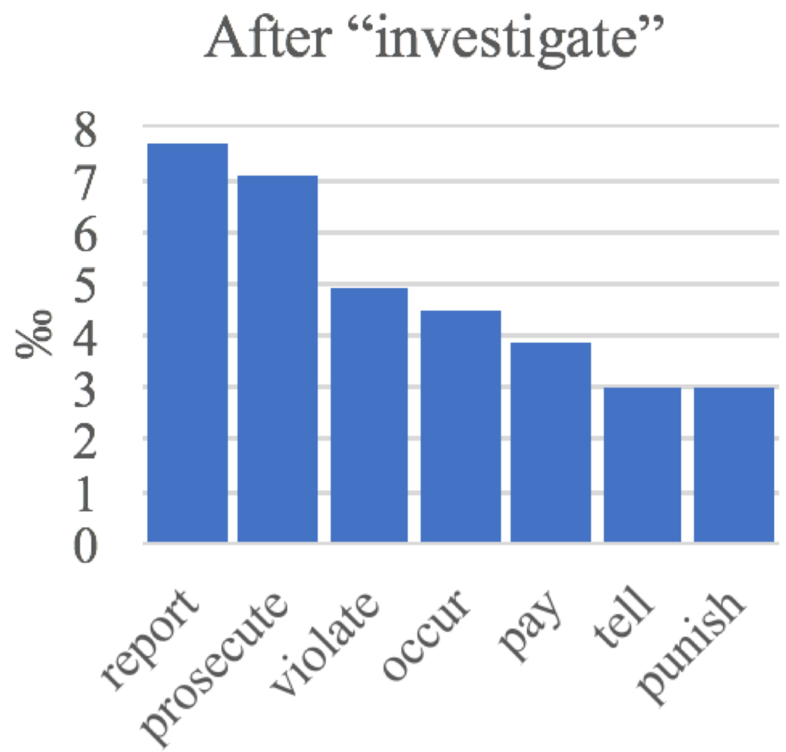
TEMPROB: PROBABILISTIC KNOWLEDGE BASE

- Source: New York Times 1987-2007 (#Articles~1M)
- Preprocessing: Semantic Role Labeling & Temporal relations model
- Result: 51K semantic frames, 80M relations
- Then we simply count how many times one frame is before/after another frame, as follows. http://cogcomp.org/page/publication_view/830

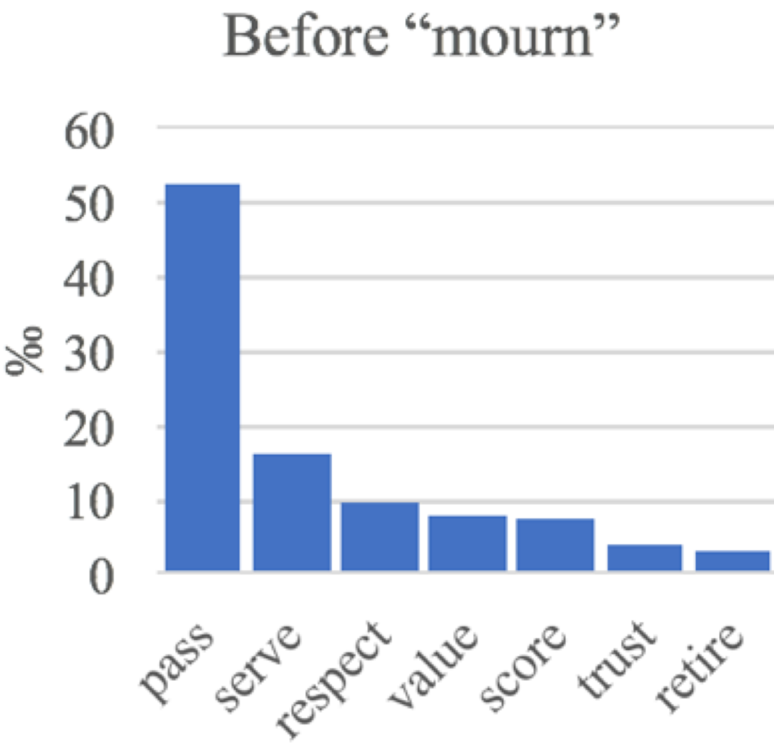
Frame 1	Frame 2	Before	After
concern	protect	92%	8%
conspire	kill	95%	5%
fight	overthrow	92%	8%
accuse	defend	92%	8%
crash	die	97%	3%
elect	overthrow	97%	3%
…			

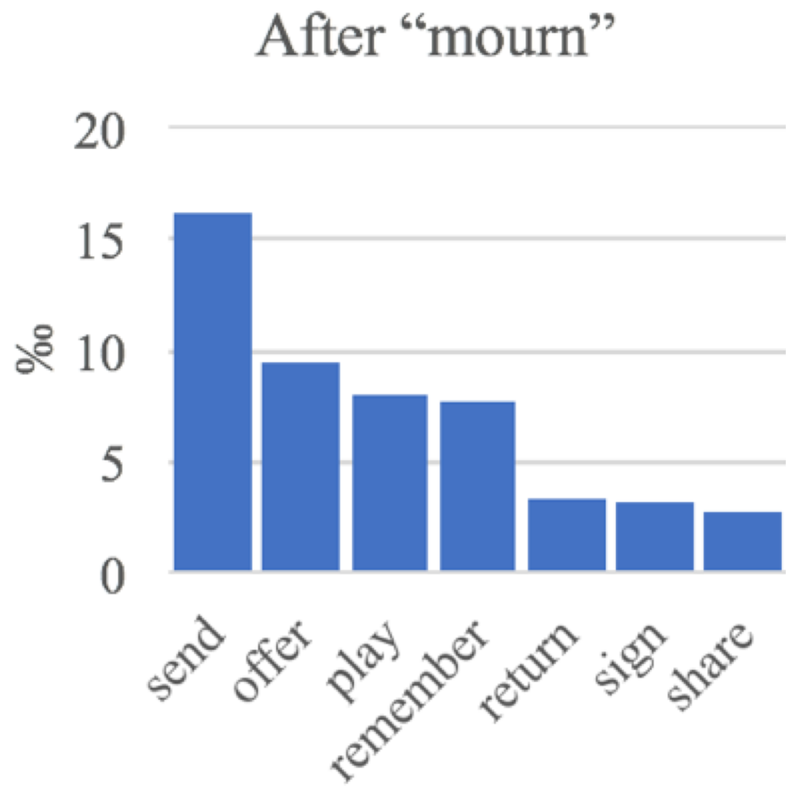
Some Interesting Statistics In TemProb





Some Interesting Statistics In TemProb





Scoring Functions: Additional Feature For Causality

$$\hat{I} = arg \max_{I} \sum_{ij \in \mathcal{E}} \left(\sum_{r \in \mathcal{R}} f_r(ij) I_r(ij) + \sum_{c \in \mathcal{C}} h_c(ij) J_c(ij) \right)$$

- Two scoring functions are needed in the objective above
  - $\boldsymbol{f_r(ij)}$--score of event pair (i,j) having <u>temporal</u> relation r
  - $\boldsymbol{h_c(ij)}$--score of event pair (i,j) having <u>causal</u> relation c
- How to obtain the scoring functions
  - We argue that this prior distribution based on ***TemProb*** is correlated with causal directionality, so it will be a useful feature when training $\boldsymbol{h_c(ij)}$.

Result on TimeBank-Dense

- TimeBank-Dense: A Benchmark Temporal Relation Dataset
- The performance of temporal relation extraction:
  - CAEVO: the temporal system proposed along with TimeBank-Dense
  - CATENA: the aforementioned work "post-editing" temporal relations based on causal predictions, retrained on TimeBank-Dense.

System	P	R	F1
ClearTK (2013)	53	26	35
CAEVO (2014)	56	42	48
CATENA (2016)	63	27	38
Ning et al. (2017)	47	53	50
This work	**46**	**61**	**52**

A New Joint Dataset

- TimeBank-Dense has only temporal relation annotations, so in the evaluations above, we only evaluated our temporal performance.
- EventCausality dataset has only causal relation annotations.
- To get a dataset with both temporal and causal relation annotations, we choose to augment the EventCausality dataset with temporal relations, using the annotation scheme we proposed in our paper *[Ning et al., ACL'18. A multi-axis annotation scheme for event temporal relation annotation.]*

	Doc	Event	T-Link	C-Link
TimeBank-Dense	36	1.6K	5.7K	-
EventCausality	25	0.8K	-	0.6K
Our new dataset	**25**	**1.3K**	**3.4K**	**0.2K***

- **due to re-definition of events*

Result On Our New Joint Dataset

	Temopral			Causal
	P	R	F	Acc.
Temporal Scoring Fn.	67	72	69	-
Causal Scoring Fn.	-	-	-	71
Joint Inference	**69**	**74**	**71**	**77**

- The temporal performance got strictly better in P, R, and F_1.
- The causal performance also got improved by a large margin.
- Comparing to when gold temporal relations were used, we can see that there's still much room for causal improvement.
- Comparing to when gold causal relations were used, we can see that the current joint algorithm is very close to its best.

Conclusion

Thank you!

- We presented a novel joint inference framework, Temporal and Causal Reasoning (TCR)
  - Using an Integer Linear Programming (ILP) framework applied to the extraction problem of temporal and causal relations between events.

- To show the benefit of TCR, we have developed a new dataset that jointly annotates temporal and causal annotations
  - Showed that TCR can improve both temporal and causal components