Improving Temporal Relation Extraction with a Globally Acquired Statistical Resource

Qiang (John) Ning
Department of Electrical and Computer Engineering
University of Illinois

Hao Wu
Department of Computer Science
University of Pennsylvania

Haoruo Peng
Department of Computer Science
University of Illinois

Dan Roth
Department of Computer Science
University of Pennsylvania
University of Illinois

Introduction

Time is an important dimension of knowledge representation. In natural language, temporal information is often expressed as relations between verb predicates (e.g., *before/after/includes/equal*).

One challenge in temporal relation extraction is that it requires high-level prior knowledge – the temporal order verbs *usually* follow. For example:

Example 1: Difficulty in understanding TempRels when event content is missing.
More than 10 people have (***e1:***), police said. A car (***e2:***) on Friday in the middle of a group of men playing volleyball.
The first thing I (***e3:***) is that they (***e4:***) writing this column.

When the verbs are missing, it's very difficult, even for humans, to figure out what happens earlier. **However, if we know that** e1=died, e2=exploded, e3=ask, and e4=help, it's obvious that e2<e1 and e3<e4, due to our prior knowledge about these verbs.

Therefore, a probabilistic knowledge-base (KB) representing this prior knowledge is in need. This work constructs TEMporal relation PRObabilistic knowledge Base (TEMPROB).
A sneak peak of TEMPROB (available at http://cogcomp.org/page/resource_view/114):

Example Pairs		Before (%)	After (%)
accept	determine	42	26
ask	help	86	9
attend	schedule	1	82
accept	propose	10	77
die	explode	14	83
…			

Method: Overview

- **Step 1**: Learn a verb temporal relation extractor based on TimeBank-Dense.
- **Step 2**: Extract verb semantic frames from a large corpus (>one million New York Times articles)
- **Step 3**: Apply the temporal relation extractor in Step 1 to all the verb frames from Step 2.
- **Step 4**: Calculate the probabilities of verb_1 and verb_2 being relation *before/after/includes /included/equal/vague*.

Method: Learning and Inference

The verb semantic frames has been extracted via the Illinois-SRL package. Given all the verb frames V in a document, a temporal relation extractor $\mathcal{E}$ is used to predict the temporal relation between any $v_1, v_2 \in V$, i.e.,

$$\mathcal{E}(v_1, v_2, r) \propto \exp\left(w_r^T \phi(v_1, v_2)\right)$$

where $r \in \mathcal{R} = \{before, after, includes, included, vague\}$

Features $\boldsymbol{\phi}(v_1, v_2)$:

(i) The part-of-speech tags
(ii) The token distance
(iii) Modal verbs in between (i.e., will, would, can, could, may and might).
(iv) Temporal connectives in between (e.g., before, after and since).
(v) WordNet based features
(vi) Prepositional head words

Learning of $\boldsymbol{w_r}$: trained and tuned on TimeBank-Dense using the averaged perceptron algorithm.

Global Inference: Temporal relations are transitive, e.g., if (A,B)=(B,C)=before, then (A,C)=before. However, the temporal relation extractor $\mathcal{E}$ alone may produce predictions that break the transitivity requirement. To handle this, we perform global inference via integer linear programming (ILP) on top of $\mathcal{E}$ while enforcing transitivity constraints:

$$\hat{\mathcal{I}} = \underset{\mathcal{I}}{\operatorname{argmax}} \sum_{ij\in\mathcal{E}} \sum_{r\in\mathcal{Y}} f_r(ij)\mathcal{I}_r(ij)$$

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

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

Resulting TEMPROB:

- NYT (1987-2007) >1M articles
- 51K unique verb frames
- 80M relations

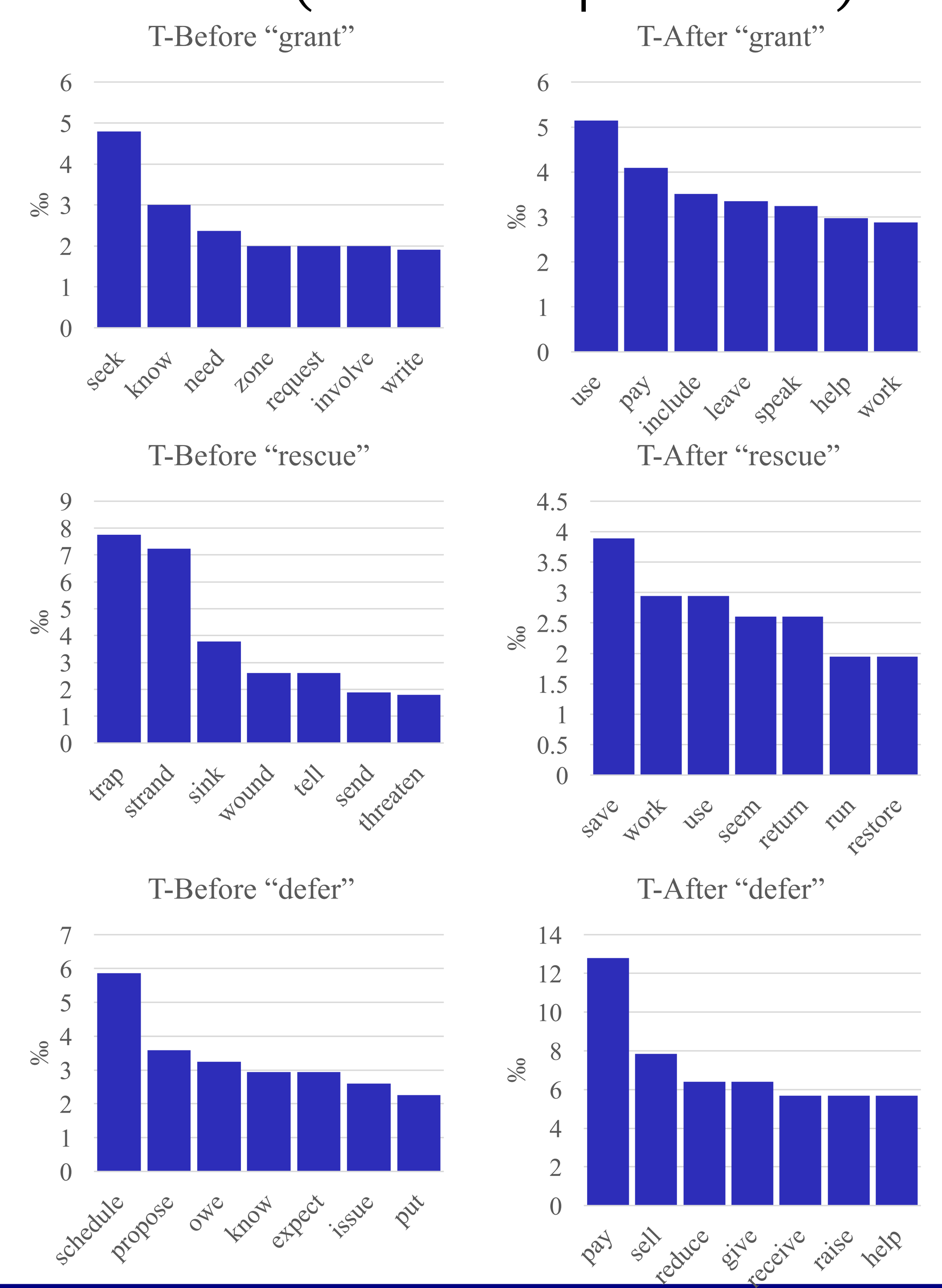
Notation

Let $C(v_1, v_2, r) = \#(v_1, v_2)$ *labeled* $r \in \mathcal{R}$. Define $P\left(v_1 \xrightarrow{r} v_2 \middle| v_1 \xrightarrow{r} V\right) = \frac{C(v_1,v_2,r)}{\sum_{v_2} C(v_1,v_2,r)}$, which characterizes how likely $(v_1, v_2) = r$ given v_1.

Define $f_r = \frac{C(v_1,v_2,r)}{\sum_{r'} C(v_1,v_2,r')}$, representing the prior statistics in TEMPROB.

Result: Temporal Patterns of Verbs

Some exemplar $P\left(v_1 \xrightarrow{before} v_2 \middle| v_1 \xrightarrow{before} V\right)$



Result: Using TEMPROB

Adding TEMPROB priors to the feature set and as an extra regularization **term leads to better relation extraction performance on a benchmark dataset.**

No.	System	P	R	F_1	F_{aware}
	Partial TBDense: Focus of this work.*				
1	CAEVO	52.3	43.7	47.6	46.7
2	Ning et al. (2017)	47.4	56.3	51.5	49.0
3	Proposed	50.0	62.4	55.5	52.8
	Complete TBDense: Naive augmentation.				
4	CAEVO	51.8	32.6	40.0	45.7
5	Ning et al. (2017)	46.2	40.6	43.2	48.5
6	Proposed**	47.2	42.4	44.7	49.2

Conclusion

Constructed from over 1 million documents, TEMPROB is shown to discover temporal patterns of verbs to aid temporal relation extraction. TEMPROB may be a useful resource for many time-aware tasks (e.g., causality and narrative schema extraction).

[0] T. Chklovski and P. Pantel. EMNLP'04. VerbOcean: Mining the Web for Fine-Grained Semantic Verb Relations. [1] N. UzZaman et al. SemEval'13. Task 1: TempEval-3: Evaluating time expressions, events, and temporal relations. [2] T. Cassidy et al. ACL'14. An annotation framework for dense event ordering. [3] N. Chambers et al. TACL' 14. Dense event ordering with a multi-pass architecture. [4] Q. Ning et al. EMNLP'17. A structured learning approach to temporal relation extraction.