CogCompTime: A Tool for Understanding Time in Natural Language

Qiang Ning[1], Ben Zhou[1], Zhili Feng[2], Haoruo Peng[1], Dan Roth[1,3]

[1]University of Illinois at Urbana-Champaign, [2]University of Wisconsin-Madison, [3]University of Pennsylvania

Highlights

Time is an important dimension. In NLP, there are two fundamental components related to time.

- **Time expressions (Timex; e.g., "yesterday")**
- **Temporal relations (TempRel; e.g., before/after)**

BEFORE BE_INCLUDED

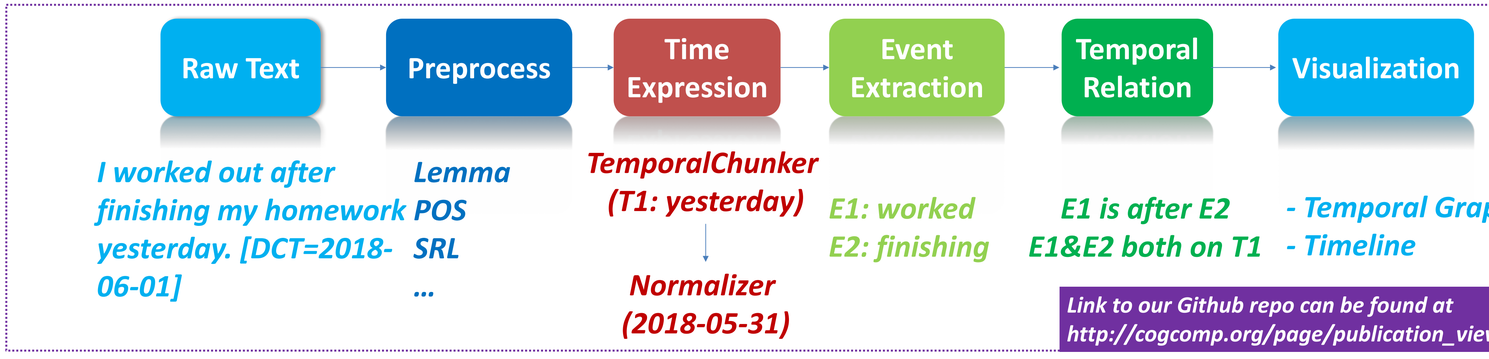
Example

I met with him before leaving for Paris on [Thursday](2018-10-25)

***CogCompTime* extracts these two components from raw text, with the most recent research progress incorporated.**

- **Timex component: x2 faster with comparable performance to HeidelTime**
- **TempRel component: F1~70%, roughly 20% improvement from previous SOTA**

System Overview



Link to our Github repo can be found at http://cogcomp.org/page/publication_view/844

System Prerequisite: JAVA1.8, maven, Gurobi (>6.5.2)

Preprocessing: The pipeline module in CogCompNLP (available at https://github.com/CogComp/cogcomp-nlp)

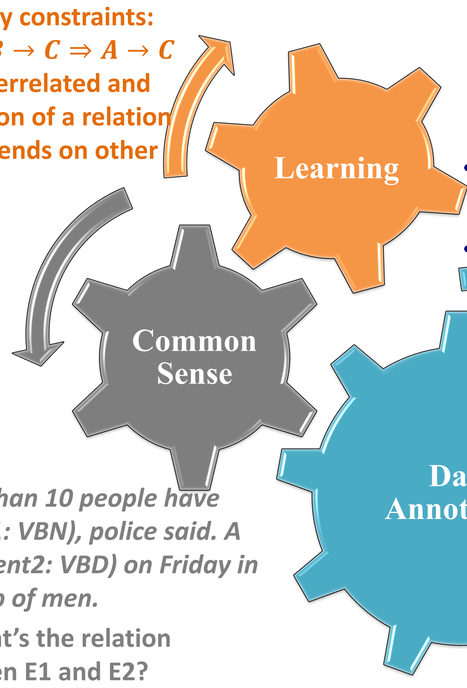
Time Expression & Temporal Relation components explained in detail below.

Event Extraction: A Binary classifier (EVENT/NOT_EVENT) following the definition of main-axis events in [ACL'18]

Temporal Relation Component

Challenges

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

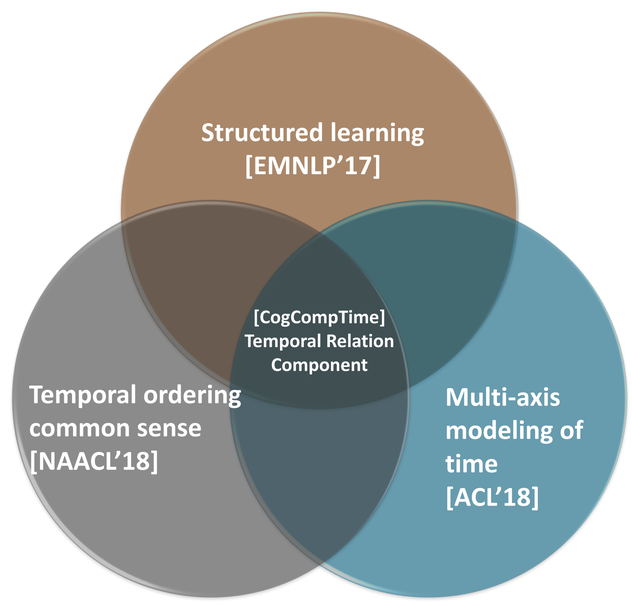


- **Labor intensive: n events➔$O(n^2)$ pairs**
- **Low IAA: $\kappa, F_1 \approx 60\%$**

More than 10 people have (event1: VBN), police said. A car (event2: VBD) on Friday in a group of men.

Q: What's the relation between E1 and E2?

Method Overview



Common Sense on Temporal Ordering [NAACL'18]

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

The TEMporal relation PRObabilistic knowledge Base (TEMPROB) is a probabilistic KB that provides the *typical temporal ordering* between verbs (i.e., temporal ordering common sense). CogCompTime adopts the statistics found in TEMPROB as an additional feature for the temporal relation classifier.

TEMPROB is available at http://cogcomp.org/page/resource_view/114

Multi-axis Modeling of Temporal Structures [ACL'18]

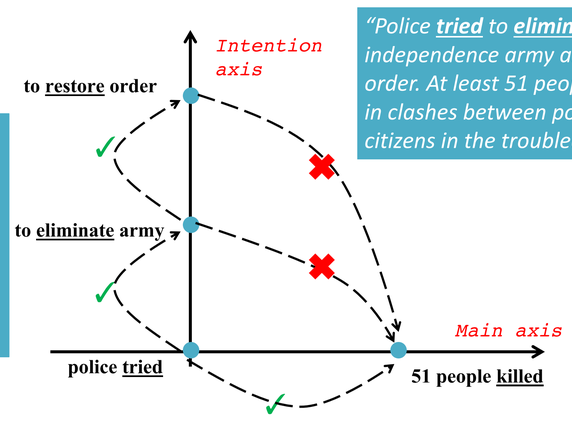
Existing data annotation schemes

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

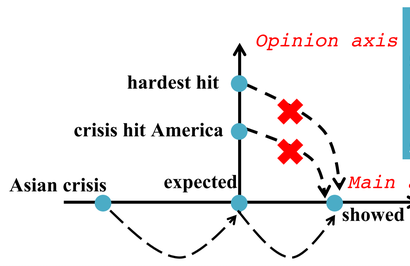
Scheme 2: Chain modeling
- E.g., TimeBank-Dense [ACL'14]
- A strong restriction on modeling
- Any pair is comparable
- But many are confusing

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

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



*"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."*

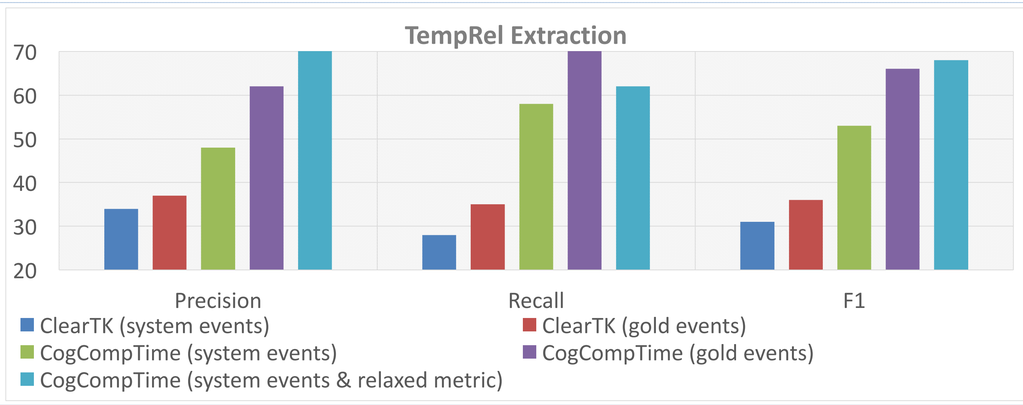
What we achieved:
- 276 docs: Annotated the 276 documents from TempEval3
- 1 week: Finished in about one week (using crowdsourcing)
- 80%: IAA improved from literature's 60% to 80%

Our dataset is available at http://cogcomp.org/page/resource_view/117

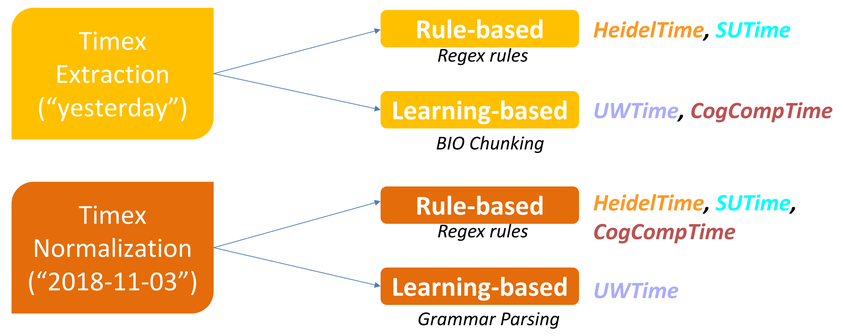
Benchmark Evaluation

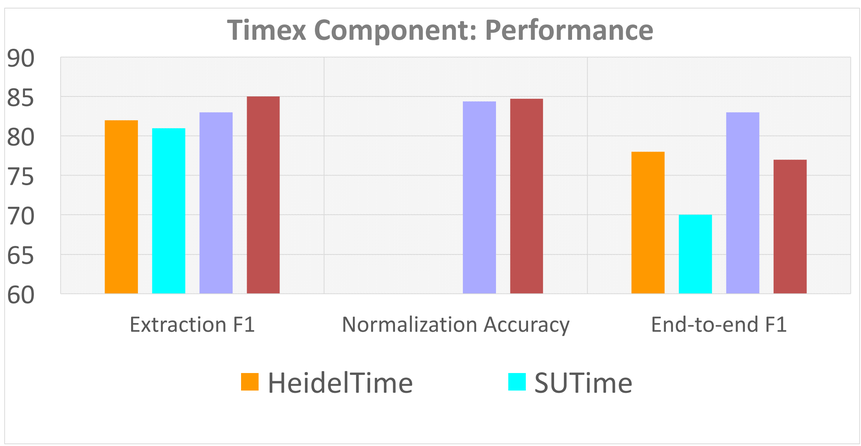
The overall performance on the proposed dataset is much better than literature values. *(Note the figure is "unfair" since we are using our own dataset, while ClearTK was using TempEval3.)*

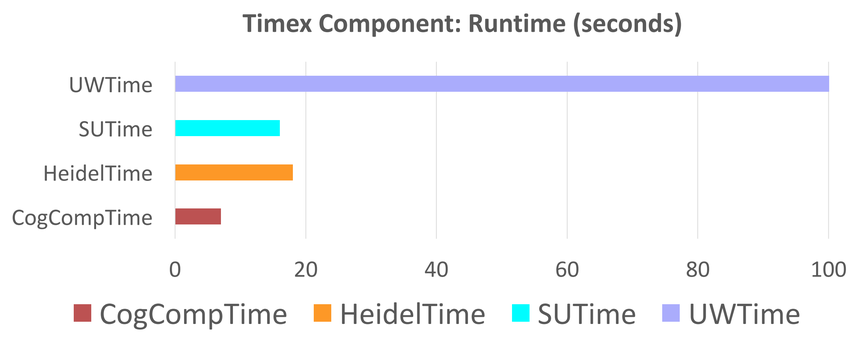
Indication: We can extract robust temporal relation signals from natural language.



Time Expression Component







Conclusion

This paper presents **CogCompTime**, a state-of-the-art package that extracts from natural language text:
- time expressions (Timex) and their normalized values
- events on the main-axis of a story
- temporal relations (TempRel) between events and Timexes.

This demo may be useful for
- identifying the shortcomings of existing methods
- applications that need temporal understanding

Reference

[SEMEVAL'10] J. Strotgen and M. Gertz. HeidelTime: High quality rule-based extraction and normalization of temporal expressions. **[NAACL'12]** R. Zhao, Q. Do, and D. Roth. A robust shallow temporal reasoning system. **[LREC'12]** A. Chang and C. Manning. SUTIME: A library for recognizing and normalizing time expressions. **[*SEM'13]** S. Bethard. 2013. ClearTK-TimeML: A minimalist approach to TempEval 2013. **[ACL'14]** Kenton Lee, Yoav Artzi, Jesse Dodge, and Luke Zettlemoyer. Context-dependent semantic parsing for time expressions. **[TACL'14]** N. Chambers, T. Cassidy, B. McDowell, and S. Bethard. Dense event ordering with a multi-pass architecture. **[EMNLP'17]** Q. Ning, Z. Feng, and D. Roth. A structured learning approach to temporal relation extraction. **[NAACL'18]** Q. Ning, H. Wu, H. Peng, and D. Roth. Improving temporal relation extraction with a globally acquired statistical resource. **[ACL'18]** Q. Ning, H. Wu, and D. Roth. A multi-axis annotation scheme for event temporal relations.

Acknowledgement: The Allen Institute for Artificial Intelligence (allenai.org); the IBM-ILLINOIS Center for Cognitive Computing Systems Research (C3SR) – a research collaboration as part of the IBM AI Horizons Network; the Army Research Laboratory (ARL).

ILLINOIS

COGNITIVE COMPUTATION GROUP

Penn University of Pennsylvania

C3SR center for cognitive computing systems research

IBM | ILLINOIS