

---

# Learning Relational Structure for Temporal Relation Extraction

---

Tushar Khot, Siddharth Srivastava\*, Sriraam Natarajan<sup>+</sup>, Jude Shavlik

University of Wisconsin-Madison, USA

<sup>+</sup> Wake Forest University, USA

## Abstract

Recently there has been a lot of interest in using Statistical Relational Learning (SRL) models for Information Extraction (IE). One of the important IE tasks is extraction of temporal relations between events and time expressions (timex). SRL methods that use hand-written rules have been proposed for various IE tasks. In contrast, we propose an approach that employs structure learning in SRL to learn such rules. Although not required, our method can also incorporate expert advice either as features or initial theory to learn a more accurate model. We present preliminary results on the TempEval-2 task of classifying relations between events and timexes.

## 1 Introduction

Information extraction (IE) has been an important problem in the Natural Language Processing (NLP) community. One specific challenging IE problem is extraction of temporal ordering between events and temporal expressions. The introduction of corpora such as the TimeBank and TimeML makes it possible to use machine learning methods to learn ordering relations between events and time expressions (timex). For example, for the sentence “He met the ambassador on June 3rd.”, we should extract the relations `OVERLAP("met", "June 3rd")` and `BEFORE("met", DOCTIME)` where `DOCTIME` corresponds to the document creation time.

The TempEval dataset [1] simplified the TimeML annotations by using six coarse-grained temporal ordering relations between events and timexes; events and document creation time; and between events.

TempEval-2 [2] extended this dataset to six tasks including the three tasks from the original dataset.

Most of the approaches applied to the TempEval tasks use propositional features and independently learn relations for each task. However, learning to predict each task independently can lead to inconsistencies in the final prediction. For example, predicting event A happened before time T ( $A < T$ ) and event B happened after time T ( $T < B$ ) is inconsistent with predicting event A happened after event B ( $A > B$ ).

There have been approaches to handle these global inconsistencies for propositional models, such as creating a globally consistent set of joint predictions by selecting from the individual predictions during inference [3]. In this work we concentrate on employing the relational approach to address this issue. Relational approaches have the advantage of focusing on the joint set of predictions during learning, rather than deferring the consideration of interaction among predictions to the inference step

Using SRL models such as MLNs would allow joint inference across various examples and tasks. As shown above, in the TempEval task we need to ensure consistent ordering between events and timexes. Also, the i.i.d. assumption made by most propositional methods is not valid as the events are not independent, further making the case for using SRL models. Furthermore, many constraints are not necessarily hard constraints in this task. For example, if event A occurred before event B and event B overlapped with C then it is *likely* that event A occurred before event C. Hence, Yoshikawa et al. [4] and more recently UzZaman and Allen [5] used Markov Logic Networks (MLNs) to specify the model as well as the global constraints as weighted first-order logic rules.

We propose using structure-learning approaches from SRL to learn rules in the absence of expert advice by using boosted Relational Dependency Networks (RDNs) [6]. We also propose two extensions to lever-

---

\* Currently at UC Berkeley

In [September]<sup>t239</sup>, it [announced]<sup>e133</sup> [plans]<sup>e134</sup> to [acquire]<sup>e135</sup> the tropical-fruit business of RJR Nabisco Inc. 's Del Monte foods unit for #557 million ( \$878 million ).

Task C: e134 OVERLAP t239

Task E: e133 BEFORE e136

Task C: e133 OVERLAP t239

Task F: e134 BEFORE e135

Task C: e135 OVERLAP-OR-AFTER t239

Task F: e133 OVERLAP e134

Figure 1: Sample TempEval-2 annotations

age expert advice whenever available. Preliminary results of our approach show promise for structure learning approaches in IE and other NLP tasks.

## 2 Background

**RDN-Boost:** RDNs are relational extensions of dependency networks, which are directed graphical models that may contain cycles. The joint distribution can be factored as product of individual conditional distributions. Natarajan et al. [6] proposed a method based on functional gradient boosting where each conditional distribution  $P(x|Pa(x))$  is approximated by a sum of relational regression trees (RRT) that are grown in a stage-wise manner. We refer to previous work [6] for more details. We chose this approach due to its competitive results across a variety of tasks in SRL [6].

**TempEval Tasks:** The TempEval task [1] in SemEval 2007 used the TimeBank corpus to create three separate relation-extraction tasks: 1) identify relations between event and timex, 2) identify relations between event and document time and 3) identify relations between events. The TempEval-2 task extended this dataset to include the problem of identifying timexes and events along with their properties. It also modified task 3 of the previous TempEval into temporal ordering tasks between: 1) events in consecutive sentences and 2) events where one event syntactically dominates the other. We show preliminary results in Section 4 for identifying relations between events and temporal expressions (called task C in TempEval-2). Figure 1 shows a sample TempEval-2 annotation. e133, e134 and e135 are the event words whereas t239 marks a timex. In this example, since the announcement happened in September, the annotations marked an OVERLAP relation between e133 and t239.

## 3 Structure Learning for TempEval-2

We first use the Stanford NLP toolkit<sup>1</sup> to convert the documents into first-order logic facts. We then use these raw features to create richer features based on our analysis of the domain. If provided, we can also use expert advice such as the rules written by previous work in this domain as the initial model. Given the initial model and the set of facts, we use RDN-Boost to learn a joint model for the target relations. Figure

<sup>1</sup><http://nlp.stanford.edu/software/corenlp.shtml>

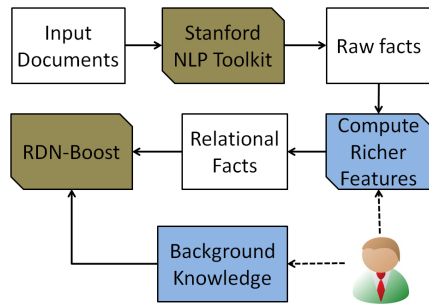


Figure 2: Flowchart describing our approach for relation-extraction

Example	Definition
wordText(W3, occurred)	Word W3 corresponds to the token occurred in the article
wordLoc(S1, W1, 1)	Word W1 is the first word of the sentence S1
wordType(W5, NN)	Word W5 is a noun (NN)
phraseType(P3, NP)	Phrase P3 is a noun phrase (NP)
phrHasWord(P3, W5)	Phrase P3 contains the word W5
headWord(P5, W11)	Word W11 is the head word of P5
depType(W3, W7, CCOMP)	Dependency graph contains an edge of type CCOMP between W3 and W7

Table 1: Sample facts generated using the Stanford toolkit

2 presents our approach.

**Raw Facts:** For each sentence, the Stanford NLP toolkit returns the tokenization, parse tree, dependency graph and named entity information. We create a word object for each token in the sentence and a phrase object for each phrase in the parse tree. Table 1 presents a subset of the generated facts. Dependency paths are considered to be important features for relation extraction and hence we create a special predicate to store the dependency path between every pair of words. Since there can be many such paths, we create the dependency path facts only if the path length is smaller than 7. For TempEval, we also convert the event and timex properties to relational facts such as `eventHasProperty(Event, Property, Value)`.

**Domain Advice:** We allow the provision of two forms of domain knowledge:

(1) **Specialized Features.** We noticed that for most of the valid event-timex pairs (i.e. having some relation), the event word is present in the dependency path (DP) from the timex to the root of the dependency graph (DG). Hence, if the DP goes up the tree and then goes down i.e. if there is a  $\swarrow \searrow$  in the DP, then it is a strong signal that the event and timex are not related. We added a predicate `veeInDepPath(W1, W2)` which is true if neither W1 nor W2 is the ancestor of the other word. For example, in Figure 3 we would create a fact: `veeInDepPath("be", "2002")`.

Typically, a timex  $t$  is related to the first verb that ap-

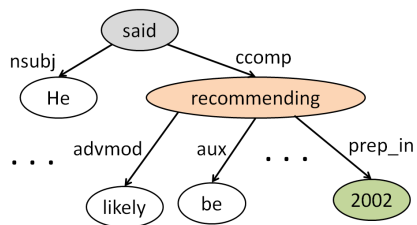


Figure 3:  
Dependency graph for a sentence where OVERLAP relation exists between “said” and “2002”

pears in the DP from  $t$  to the root of the DG. However, additional verbs in the path to the root can also be related to  $t$  if they are preceded by special dependency tags (e.g. CCOMP). In order to learn such tags, we included a predicate `verbAlongDependencyPath(word, word, verb, depType)` to represent this feature. We now let RDN-Boost discover which dependency types could be present for valid relations. Figure 3 shows a snippet of a DG. Although there is a verb in the DP from “2002” to “said” since “recommending” is connected by a CCOMP dependency type, “2002” applies to “said” too.

**(2) Expert Rules.** For the TempEval task, Yoshikawa et al. [4] designed rules to encode the constraints for consistent ordering between events, timexes and document times. Similar rules were also used by the TRIPS/TRIOS system [5] for the TempEval-2. We can use these rules as the initial model for RDN-Boost. Each Horn clause is used as a part of the initial model for the predicate that appears in the head of the clause. For example, a sample rule used by previous approaches was  $\text{relE2T}(e1, t, \text{“BEFORE”}) \wedge \text{relE2T}(e2, t, \text{“AFTER”}) \rightarrow \text{relE2E}(e1, e2, \text{“BEFORE”})$ . This rule can be used as the initial model for predicting `relE2E`. While `relE2T` represents relations between events and timexes, `relE2E` represents relations between events.

## 4 Preliminary Results

We present the preliminary results of our approach on task C of TempEval-2. We did not use any cross-task rules in the initial model, since we learn a model for a single task. When not using any domain-specific features, RDN-Boost is able to achieve an accuracy of 0.56 on the test set. Including the domain-specific features improved the testset accuracy of the system to 0.60. Most of the systems that competed in TempEval-2 had an accuracy ranging between 0.62-0.65. We believe with better features and simultaneously using the data from all the TempEval tasks to learn a joint model would further improve the results.

## 5 Discussion and Future Work

Temporal relation-extraction is an important IE task where SRL methods have shown promise. SRL allows one to learn a joint model to find a globally consis-

tent relation-extraction system, but prior work used hand-written rules. We propose an approach to learn the rules for the SRL model, while taking advantage of any domain-specific knowledge that is available. Our preliminary results for structure learning for temporal relation-extraction are promising. We expect the accuracy to improve as we utilize the data from all the TempEval tasks to learn a joint model.

We will work on using the other TempEval tasks such as relation between events and document creation time to perform joint inference across tasks. We can also use the cross-task MLN rules as the initial model. We plan to incorporate more features based on previous work on relation extraction such as the words between the event and timex. Previous approaches [4] have also used the annotations from other tasks to increase the training data for relation extraction tasks. For example, if the annotations have marked event A occurred before the document creation time and event B occurred after the document time, we can add annotations to mark event A occurred before event B.

## Acknowledgments

The authors gratefully acknowledge the support of Defense Advanced Research Projects Agency (DARPA) Machine Reading Program under Air Force Research Laboratory (AFRL) prime contract no. FA8750-09-C-0181. Any opinions, findings, and conclusion or recommendations expressed in this material are those of the authors and do not necessarily reflect the view of the DARPA, AFRL, or the US government.

## References

- [1] M. Verhagen, R. Gaizauskas, F. Schilder, M. Happle, G. Katz, and J. Pustejovsky. SemEval-2007 task 15: TempEval temporal relation identification. In *SemEval*, 2007.
- [2] M. Verhagen, R. Sauri, T. Caselli, and J. Pustejovsky. SemEval-2010 task 13: Tempeval-2. In *SemEval*, 2010.
- [3] N. Chambers and D. Jurafsky. Jointly combining implicit constraints improves temporal ordering. In *EMNLP*, 2008.
- [4] K. Yoshikawa, S. Riedel, M. Asahara, and Y. Matsumoto. Jointly identifying temporal relations with Markov Logic. In *ACL*, 2009.
- [5] N. UzZaman and J. F. Allen. TRIPS and TRIOS system for TempEval-2: Extracting temporal information from text. In *SemEval*, 2010.
- [6] S. Natarajan, T. Khot, K. Kersting, B. Guttman, and J. Shavlik. Gradient-based boosting for statistical relational learning: The relational dependency network case. *Machine Learning*, 2011.