

TRIOS-TimeBank Corpus: Extended TimeBank corpus with help of Deep Understanding of Text

Naushad UzZaman and James Allen

Department of Computer Science, University of Rochester, Rochester, NY 14627
{naushad, james}@cs.rochester.edu

1 Introduction

TimeBank (Pustejovsky et al, 2003a), a reference for TimeML (Pustejovsky et al, 2003b) compliant annotation, is widely used temporally annotated corpus in the community. It captures time expressions, events, and links between events, as suggested by TimeML; but there are room for improvements. This work is one such effort to extend the TimeBank corpus. Our first goal is to include missing TimeBank events, i.e. events that are missed by TimeBank annotators. Along with that this paper also suggests some additions to TimeML language by adding new event features (ontology type) and also relations between events or event with words, which we call RLINK (relation link). With our new suggestions we present the **TRIOS-TimeBank** corpus, extending TimeBank corpus.

2 TRIOS system

We developed a system **TRIOS** (**TRIPS**¹ Temporal Reasoning System), which uses deep language understanding (semantic parsing with TRIPS parser² (Allen et al, 2008)) to extract events, their linguistic features, and relations in the text.

The TRIPS parser uses an extensive grammar of English and a large semantic lexicon to produce logical form representations in terms of linguistically-motivated ontology³. As an example, the result of parsing the sentence “*He fought in the war*”, is expressed as set of expressions in an unscoped logical formalism with reified events and semantic roles.

```
(SPEECHACT V1 SA-TELL :CONTENT V2)
(F V2 (:* FIGHTING FIGHT) :AGENT V3 :MODS
(V4) :TMA ((TENSE PAST)))
(PRO V3 (:* PERSON HE) :CONTEXT-REL HE)
(F V4 (:* SITUATED-IN IN) :OF V2 :VAL V5)
(THE V5 (:* ACTION WAR))
```

¹TRIPS: The Rochester Interactive Planning System: <http://www.cs.rochester.edu/research/trips/>

²TRIPS parser demo: <http://www.cs.rochester.edu/research/cisd/projects/trips/parser/cgi/parserinterface.pl>

³TRIPS ontology browser: <http://www.cs.rochester.edu/research/trips/lexicon/browse-ont-lex.html>

The main event produced is an event of type FIGHTING, which is a sense corresponding to the fifth WordNet sense of contend, and shares verbs such as fight, defend, contend and struggle. For words not in the TRIPS core lexicon, the system looks up the WordNet senses and maps them to the TRIPS ontology. The word “war” is not in the core lexicon, and via WordNet is classified into the ontology as a type of ACTION.

Given the logical form of a sentence, we then use a set of extraction patterns, which match logical form expressions in terms of the ontology, to identify relevant events, temporal expressions and other relations of relevant to TimeML. Because of the ontology, we can usually express general rules that capture a wide range of events. For instance, all noun-phrases describing objects that fall under the TRIPS Ontology’s top-level type SITUATION-ROOT are extracted as described events. The result of matching such rules to the sentence above is:

```
<EVENT eid=V2 word=FIGHT pos=VERBAL
  ont-type=FIGHTING class=OCCURRENCE
  tense=PAST voice=ACTIVE
  polarity=POSITIVE aspect=NONE
  nf-morph=NONE>
<RLINK eventInstanceID=V2 ref-word=HE
  ref-ont-type=PERSON relType=AGENT>
<SLINK signal=IN eventInstanceID=V2
  subordinatedEventInstance=V5
  relType=SITUATED-IN>
<EVENT eid=V5 word=WAR pos=NOUN
  ont-type=ACTION class=OCCURRENCE
  voice=ACTIVE polarity=POSITIVE
  aspect=NONE tense=NONE>
```

Readers can see some differences (in bold) from usual TimeML annotations; these differences are our new suggestions that we will describe in next sections. Detail description on how TRIPS parser is used to extract events and its features, could be found in (UzZaman, 2009).

TimeBank has an inter-annotator agreement of 78% (average of precision and recall on subset of 10 documents) on event extraction. Our system performed with 65.08% precision and 73.96% recall on

TimeBank. With an additional filtering⁴, we can improve our precision to 91.3%, compromising 3% decrease in recall. For temporally annotating new documents, this second level filtering might be useful. But for the task of extending TimeBank, we refer to TimeBank to add event features for events that are missed by TRIOS system. So, we skip second level filtering to keep higher recall, so that we can add our additional features and relations for more event instances.

3 Extensions to TimeBank and TimeML

3.1 Suggesting new events in TimeBank

The low inter-annotator agreement suggests that there should be some effort to refine TimeBank events. It is hard to automatically suggest that some annotated event in TimeBank is wrong; so we only **suggest new events that are missing in TimeBank**.

The TimeML (Pustejovsky et al, 2003b) specification says not to tag Generic interpretations, even though capturing them could be of use in question answering. By generics, they mean, events that are not positioned in time, or in relation to other temporally located events in the document. For example, they won't annotate *use* and *travel* in the sentence: *Use of corporate jets for political travel is legal*.

It also suggests not to tag subordinate verbs that express events which are clearly temporally located, but whose complements are generics. For example, *He said Jews are prohibited from killing one another*. Even though the verb *said* is temporally located, it isn't tagged because its complement, *Jews are prohibited from killing one another*, is generic.

And finally an event nominalization that doesn't provide any extra information than the supplied verbs, are also not tagged.

Many of the extra events generated by TRIOS that are not in TimeBank fall into these categories. We made a decision to only suggest verbal event, so we don't have to worry about the last case. For verbal events, our task would be to keep the events that matches with TimeML specification.

There is one case where we decided to tag events even though they don't meet the TimeML specification. An example is, *He said the earth is round. They killed him.*. Here, "*the earth is round*" is generic by their definition. According to their scheme, "*said*"

⁴We implemented a MLN classifier to classify TRIOS events into TimeBank type event and wrong extraction. This extra step was to remove generics and wrong extraction from TRIOS generated output.

won't be annotated. But this saying event might explain rest of the story, i.e. in this case, why he was killed. If we don't annotate these kind of events, we are removing the information that we are interested in. So, we keep these kind of events in our event suggestion list, but will try to distinguish and eliminate other generic events.

The extra TRIOS events that don't exist in TimeBank could be categorized as follows: i. the result of wrong parse, ii. a generic event and iii. a legitimate event but missed by annotators.

Here are few examples that we think are legitimate events and also missed by TimeBank annotators:

*An intense manhunt **conducted** by the FBI and the bureau of alcohol, ...*

*If Iraq **chooses** a simple war of nerves and economic attrition, the Bush administration knows...*

*American strategists are calculating, though, that the trade sanctions – enforced by an effective though perhaps undeclared naval blockade – will **hold** tightly enough to **convince** Iraq that it will **lose** in the long run by simply standing pat.*

We are interested in suggesting these legitimate events and filter out generic and wrong events. The first level filtering is done by only keeping the events that are suggested as Verbs by both TRIPS parser and also Stanford POS tagger. Then the final filtering is done by classifying these extra events into "suggestion", "generic" and "wrong" categories. To do this task, we implemented a MLN (markov logic network) classifier using TheBeast tool⁵. We generated the formulas for MLN from TRIOS generated event features. Flowchart for suggesting new TRIOS events is shown in Figure 1.

To analyze, we picked 40 TimeBank documents and annotated the extra TRIOS events with "suggestion", "generic" and "wrong" categories. We used 20 documents as test data and other 20 documents as training data. Since we didn't have enough annotated data, we randomly picked some TimeBank events as "suggestion" instance and some non-verbal extra TRIOS events, which were filtered out in our first level filtering, as "wrong" instances. We added these new instances with our training data and then tested on our unseen 20 documents.

In these 20 documents test data, we had 90 extra events after our first level filtering, where we

⁵MLN Tool TheBeast: <http://code.google.com/p/thebeast/>. All MLN classifier in this paper are implemented using TheBeast

had 14 events which were result of wrong parse, 41 generic events and 35 events that we think are legitimate events but missed by TimeBank that we want to suggest. The performance of classifying these categories are reported in Table 1.

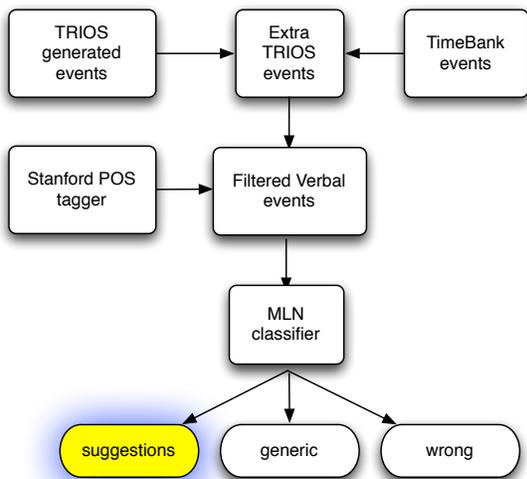


Figure 1: Flowchart to classify extra TRIOS events to “suggestion”, “generic” and “wrong” categories

In full TimeBank, we nominate 484 events as new event. From system’s performance, we expect that we extracted almost 90% of probable missing events and around 50% of these events are legitimate. These nominated new events are small in numbers. Hence, suggestions from this tool will help significantly the annotator to add new events to TimeBank.

Table 1: Performance on classification of extra TRIOS events into “suggestion”, “generic” and “wrong” category

Category	# Instance	Precision	Recall
suggestion	35	50.00%	88.57 %
generic	41	70.00%	34.14%
wrong	14	87.57%	50.00%

3.2 Adding ontology type as new event features in TimeML

TimeML comprises of event features, *class*, *tense*, *aspect*, *nf-morph*, *pos*, *modality*, and *polarity*. TRIOS system generates these features and also add ontology type as event feature.

Ontology type is the semantic type of word, par-

ticular word sense in the context, in TRIPS ontology⁶. TimeML tries to capture event information by very high level *class* or *pos*. Ontology type feature will try to capture more fine grained information about the event, but in higher level than event word.

Ontology type instances from our initial example are, FIGHTING for word FIGHT and ACTION for word WAR. Few other words with ontology type FIGHTING would be: *contend*, *defend*, and *struggle*, i.e. these words with similar meaning will get the same ontology type, in this case FIGHTING.

TRIPS Ontology will be available for public use, so people can use the ontology for their system. It also has mapping to WordNet. So converting it to WordNet classes, someone can take benefit of this ont-type feature.

TRIOS system generates event features from TRIPS parser output. For classifying feature *class*, we implemented a MLN classifier and used rest of the event features as features for classifier. In this classification problem, our system with TRIOS generated features performed with 77.3%⁷ and the same system with TimeBank’s event features performed with 77.47% accuracy. It is worth mentioning that TimeBank’s inter-annotator agreement (IAG) on *class* is 77%. Comparing IAG and TimeBank’s feature, our performance suggests that TRIOS generated features are equally good, which is an indication of ontology-type’s performance as well.

3.3 Adding improved relations in TimeML

Our next contribution is adding richer relations in TimeML. TimeML captures the relations between different events with TLINK (temporal links), SLINK (subordinate link), and ALINK (aspectual link).

3.3.1 New Relation Link, RLINK

Many researchers (Chambers et al, 2007), (Katsumasa et al, 2009) showed that having the dependency information improves the performance for extracting temporal relations. They tried to capture the dependency relation with dependency parsers like Stanford dependency parser. This gives a hint that capturing how other dependent words are connected with the event will enrich the information about event.

We introduce new relation link, RLINK, to capture what is related to event (other than another

⁶TRIPS ontology browser: <http://www.cs.rochester.edu/research/trips/lexicon/browse-ont-lex.html>

⁷All performance for *class* identification are reported using 10-fold cross-validation.

event, which is captured with SLINKs).

In our initial example, for event FIGHT, we try to capture the information that the AGENT of that fighting event is HE, which is a PERSON. This relations give us information what are the dependent words of event and how they are connected.

```
<RLINK eventInstanceID=V2 ref-word=HE
  ref-ont-type=PERSON relType=AGENT>
```

Another example of RLINK’s importance could be explained with (Chambers and Jurafsky, 2008). They learned narrative event chains considering the idea of protagonist (central actor). They are basically considering the events performed by same agent. We are trying to capture agent and other different thematic roles (or semantic roles) using RLINK, which would help many other applications like (Chambers and Jurafsky, 2008).

3.3.2 More SLINK instances

SLINK or Subordinate Link is used for contexts introducing relations between two events. They classify SLINKs into *Modal, Factive, Counter-Factive, Evidential, Negative evidential and Conditional*. This classification leave out many instances where two events are related with each other. We try to capture all possible relations when one event is related with another event.

An instance from our initial example is:

```
<SLINK signal=IN eventInstanceID=V2
  subordinatedEventInstance=V5
  relType=SITUATED-IN>
```

This is very similar to usual SLINK in TimeML. If possible, we also try to capture the signal (connectives, that connects two events). The problem is identifying the relation type (relType). We decided to use the ontology type of our signal (connective) as the relType for these kind of extra SLINKs.

3.4 Building TRIOS-TimeBank corpus

Our final task is to include all this new information in a new version of TimeBank. First of all, the TRIOS system misses some events that are already in TimeBank. For these missing events, we keep the TimeBank event features for that event, we will just copy the *class* as *ontology type* for this case. Our system’s event extraction recall is 73.96%, i.e. we take 26.04% from TimeBank. For other events that we are able to extract, we will replace the TimeBank event entry with TRIOS generated features. We will add extra SLINK and RLINKs in the document as well.

For the new event suggestions, we will add these events in the TimeBank document as well with event

feature “source”, which will make users easily identify if this event is a new suggested event or not. This “source” feature will also distinguish TimeBank event from TRIOS generated event as well.

4 Conclusion

In this paper, we presented the TRIOS-TimeBank corpus, an extended TimeBank corpus with additional events. We also proposed an extension to TimeML language with richer event feature, and event relations, which we generated with help of deep understanding of text using semantic parsing. This resource, TRIOS-TimeBank corpus, with newly added events, event feature and relations, is available to the community for further research on temporal reasoning. As a future work, we want to improve our TRIPS parser in order to extract better events, generate better event features and relations; and with help of that, we want to automatically temporally annotate more newswire texts and build a much bigger temporally annotated corpus for the community.

5 Acknowledgment

Sebastian Reidel for help on MLN TheBeast. Mary Swift and Will for helping with different TRIPS related issues. Benjamin van Durme for many useful suggestions.

References

- James Allen, Mary Swift, and Will de Beaumont. 2008. *Deep semantic analysis of text*. Symposium on Semantics in Systems for Text Processing (STEP), Venice, Italy, 2008.
- Nathanael Chambers, Shan Wang, and Daniel Jurafsky. 2007. *Classifying temporal relations between events*. Proceedings of the 45th Annual Meeting on Association for Computational Linguistics. Association of Computational Linguistics, 2007.
- Nathanael Chambers and Daniel Jurafsky. 2008. *Un-supervised Learning of Narrative Event Chains*. Proceedings of the 46th Annual Meeting on Association for Computational Linguistics. Association of Computational Linguistics, 2008.
- Katsumasa Yoshikawa, Sebastian Riedel, Masayuki Asahara and Yuji Matsumoto. 2009. *Jointly Identifying Temporal Relations with Markov Logic*. Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2009.
- James Pustejovsky, Jos M. Castao, Robert Ingria, Roser Sauri, Robert J. Gaizauskas, Andrea Setzer, Graham Katz, and Dragomir R. Radev. 2003. *TimeML: Robust Specification of Event and Temporal Expressions in Text*. In Mark T. Maybury, editor, *New Directions in Question Answering*, pages 2834. AAAI Press, 2003.

J. Pustejovsky, P. Hanks, R. Saur, A. See, R. Gaizauskas, A. Setzer, D. Radev, B. Sundheim, D. Day, L. Ferro, and M. Lazo. 2003. *The TIMEBANK corpus*. Corpus Linguistics, 2003, Lancaster, March 2003.

Naushad UzZaman. 2009. *Interpreting the Temporal Aspects of Language*. Area Paper, Department of Computer Science, University of Rochester, 2009.