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

The seventh international conference on Language Resources and Evaluation (LREC), Malta, 2010

Naushad UzZaman and James F. Allen
Computer Science Department
University of Rochester
Rochester, NY, USA
TimeML, TimeBank, TempEval-1 and TempEval-2

- TimeML is the scheme for temporal annotation by James Pustejovsky et al.
- TimeBank is the first annotated corpus
- TempEval 1 and 2 are annotated corpus and was shared task on temporal information extraction
Outline

• Our System
• Extension to TimeBank and TimeML
  • Suggest new event
  • Suggest new temporal expressions
  • Suggest ontology type as new event features
  • Suggest improved relations in TimeML
• Future Work
"Advanced Medical paid $106 million in cash for its share in a unit of Henley's Fisher Scientific subsidiary."
"Advanced Medical paid $106 million in cash for its share in a unit of Henley's Fisher Scientific subsidiary."
"Advanced Medical paid $106 million in cash for its share in a unit of Henley's Fisher Scientific subsidiary."
"Advanced Medical paid $ 106 million in cash for its share in a unit of Henley 's Fisher Scientific subsidiary ."
Events and event features extraction using TRIPS parser

Sentence

He fought in the war

TRIPS parser output

(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))

100+ Extraction rules

((THE ?x (? type SITUATION-ROOT))
 -extract-noms>
 (EVENT ?x (? type SITUATION-ROOT) :pos NOUN :class OCCURRENCE ))

Extracted with extraction rules

<Event eid=V2 word=FIGHT pos=VERBAL ont-type=FIGHTING tense=PAST
 class=OCCURRENCE voice=ACTIVE aspect=NONE polarity=POSITIVE nf-morph=NONE>
<RLINK_eventInstanceId=V2_ref-word=HE ref-ont-type=PERSON relType=AGENT>
<SLINK_signal=IN_eventInstanceId=V2_subordinatedEventInstanceId=V5
 relType=SITUATED-IN>
<Event eid=V5 word=WAR pos=NOUN ont-type=ACTION class=OCCURRENCE
 voice=ACTIVE polarity=POSITIVE aspect=NONE tense=NONE>
### Event Extraction Performance

<table>
<thead>
<tr>
<th>System</th>
<th>Precision</th>
<th>Recall</th>
<th>Fscore</th>
<th>(P+R)/2</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRIOS avg</td>
<td>0.8638</td>
<td>0.7074</td>
<td>0.7778</td>
<td>0.7856</td>
</tr>
<tr>
<td>TRIPS avg</td>
<td>0.5801</td>
<td>0.8513</td>
<td>0.6900</td>
<td>0.7157</td>
</tr>
<tr>
<td>STEP</td>
<td>0.82</td>
<td>0.706</td>
<td>0.7587</td>
<td>0.763</td>
</tr>
<tr>
<td>Sim-Evita</td>
<td>0.812</td>
<td>0.657</td>
<td>0.727</td>
<td>0.7345</td>
</tr>
<tr>
<td>IAA</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Table 1: Event Extraction Performance on Bethard and Martin’s test data

<table>
<thead>
<tr>
<th>System</th>
<th>Precision</th>
<th>Recall</th>
<th>Fscore</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRIOS</td>
<td>0.80</td>
<td>0.74</td>
<td>0.77</td>
</tr>
<tr>
<td>TRIPS</td>
<td>0.55</td>
<td>0.88</td>
<td>0.68</td>
</tr>
<tr>
<td>Best (TIPSem)</td>
<td>0.81</td>
<td>0.86</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Table 1: Performance of Event Extraction (Task B) in TempEval-2
## Performance on Temporal Expression Extraction

<table>
<thead>
<tr>
<th></th>
<th>TRIPS</th>
<th>TRIOS</th>
<th>Best</th>
<th>HeidelTime-1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Temp Exp extraction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precision</td>
<td>0.85</td>
<td></td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>Recall</td>
<td>0.85</td>
<td></td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>Fscore</td>
<td>0.85</td>
<td></td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td><strong>Normalization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>type</em></td>
<td>0.94</td>
<td></td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td><em>value</em></td>
<td>0.76</td>
<td></td>
<td>0.85</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Performance on Temporal Expression extraction (Task A)
Outline

- Our System

- Extension to TimeBank and TimeML
  - Suggest new event
  - Suggest new temporal expressions
  - Suggest ontology type as new event features
  - Suggest improved relations in TimeML

- Future Work
What not to tag?

• Not to tag generic interpretations
  
  • *Use of corporate jets for political travel* is legal.

• Complements are generics
  
  • *He said students are prohibited from fighting with each other.*

• Nominalization without extra information
  
  • *Newspaper reports have said* ...
TRIOS minus TimeBank

- Result of wrong parse
- Generic event
- Legitimate event but missed by annotators
Missed legitimate events

1. At least one of the sensitive sites was a barracks of the elite Republican Guard, a well-placed source told The Associated Press.

2. Net interest income for the third quarter declined to $35.6 million from $70.1 million a year ago.

3. About $518 million of debt is affected.

4. If Iraq chooses a simple war of nerves and economic attrition, the Bush administration knows a long stalemate could try the patience of the American public and the West in general, and could open the possibility that moderate Arabs -- even including Saudi Arabia -- might drop out of the effort against Iraq and accept some deal from Saddam Hussein.

5. "It's the whole uncertainty about what's happening around us," said Valentin Von Korff, a trader at Credit Suisse First Boston in Frankfurt.
Suggest new verbal events to TimeBank

Verbal events by TRIPS parser

Extra TRIPS events

MLN classifier

TimeBank events

suggestions

wrong
If Iraq chooses a simple war of nerves and economic attrition, the Bush administration knows a long stalemate could try the patience of the American public and the West in general, …
Suggest new temporal expressions

Total timex: 1414
Accuracy: 50/68 = 73.5%
Suggestion: ~3.5%

- At the end of the broadcast this evening, one more trip around Havana to see what it's been like since the last time.

- Turks feel they have special ties to the whole region, which they ruled for hundreds of years during the Ottoman Empire.

- Weisfield's, based in Seattle, Wash., currently operates 87 specialty jewelry stores in nine states.

- Previously, watch imports were denied such duty-free treatment.
Add ontology-type as new event feature

- TimeML captures event information with coarse-grained class (7) or pos, or fine-grained word
- TRIPS Ontology type
  - more fine-grained than class or pos
  - coarse-grained than word
  - *Fighting* for *fight*, *Action* for *war*
- Few other words with ont-type *Fighting*: contend, defend, struggle, etc.

*TimeML Class:*
1) Occurrence: die, crash, build;
2) State: on board, kidnapped;
3) Reporting: say, report;
4) I-Action: attempt, try, promise;
5) I-State: believe, intend, want;
6) Aspectual: begin, stop, continue;
7) Perception: see, hear, watch, feel.

- Mapping to WordNet
- Freely available
Outline

• Our System
• Extension to TimeBank and TimeML
  • Suggest new event
  • Suggest new temporal expressions
  • Suggest ontology type as new event features
  • Suggest improved relations in TimeML
• Future Work
More SLINK instances

• SLINK or Subordinate Links: relation between two events

• TimeML includes: modal, factive, counterfactive, evidential, negative evidential and conditional

• Many other cases when one event is argument of other
SLINK examples

• "They have to **continue** to **tighten** their belts," said Craig Kloner, an analyst at Goldman, Sachs amp Co. (Purpose)

• **He fought** in the **war**.

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

• **Suggest ~900 SLINKs**
New Relation Link, RLINK

- Dependency information improves temporal relation identification performance (Chambers et al. 2007 and Katsumasa et al. 2009)

- Chambers and Jurafsky (2008)’s narrative chain with AGENT

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

- Suggest ~2000 RLINKs
Table 1: Most common relTypes used in SLINKs and RLINKs

<table>
<thead>
<tr>
<th>Our Role</th>
<th>VerbNet equivalents</th>
<th>Lirics equivalents</th>
<th>SLINK Count</th>
<th>RLINK Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent</td>
<td>Agent, Actor</td>
<td>Agent</td>
<td>19</td>
<td>709</td>
</tr>
<tr>
<td>Theme</td>
<td>Theme, Stimulus</td>
<td>Theme</td>
<td>336</td>
<td>1137</td>
</tr>
<tr>
<td>Affected</td>
<td>Patient</td>
<td>Patient</td>
<td>13</td>
<td>92</td>
</tr>
<tr>
<td>Cause</td>
<td>Cause</td>
<td>Cause</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>Goal-as-Loc</td>
<td>Destination</td>
<td>finalLocation</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>To-Loc</td>
<td>Recipient</td>
<td>Goal</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>At-Loc</td>
<td>Location</td>
<td>Location</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>In-Loc</td>
<td>Location</td>
<td>Location</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>On</td>
<td>Location</td>
<td>Location</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Situated-In</td>
<td>Location?</td>
<td>Location?</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>Purpose</td>
<td>–</td>
<td>Purpose</td>
<td>226</td>
<td></td>
</tr>
</tbody>
</table>
### Approaching TempEval-2

<table>
<thead>
<tr>
<th>Features</th>
<th>TE2 C</th>
<th>TE2 D</th>
<th>TE2 E</th>
<th>TE2 F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TE1 A</td>
<td>TE1 B</td>
<td>TE1 C</td>
<td></td>
</tr>
<tr>
<td>Event Class</td>
<td>YES</td>
<td>YES</td>
<td>$e_1 \times e_2$</td>
<td>$e_1 \times e_2$</td>
</tr>
<tr>
<td>Event Tense</td>
<td>YES</td>
<td>YES</td>
<td>$e_1 \times e_2$</td>
<td>$e_1 \times e_2$</td>
</tr>
<tr>
<td>Event Aspect</td>
<td>YES</td>
<td>YES</td>
<td>$e_1 \times e_2$</td>
<td>$e_1 \times e_2$</td>
</tr>
<tr>
<td>Event Polarity</td>
<td>YES</td>
<td>YES</td>
<td>$e_1 \times e_2$</td>
<td>$e_1 \times e_2$</td>
</tr>
<tr>
<td>Event Stem</td>
<td>YES</td>
<td>YES</td>
<td>$e_1 \times e_2$</td>
<td>$e_1 \times e_2$</td>
</tr>
<tr>
<td>Event Word</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Event Constituent²</td>
<td></td>
<td>YES</td>
<td>$e_1 \times e_2$</td>
<td>$e_1 \times e_2$</td>
</tr>
<tr>
<td>Event Ont-type³</td>
<td>YES</td>
<td>YES</td>
<td>$e_1 \times e_2$</td>
<td>$e_1 \times e_2$</td>
</tr>
<tr>
<td>Event LexAspect⁴ x Tense</td>
<td>YES</td>
<td>YES</td>
<td>$e_1 \times e_2$</td>
<td>$e_1 \times e_2$</td>
</tr>
<tr>
<td>Event Pos</td>
<td>YES</td>
<td>YES</td>
<td>$e_1 \times e_2$</td>
<td>$e_1 \times e_2$</td>
</tr>
<tr>
<td>Timex Word</td>
<td></td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Timex Type</td>
<td>YES</td>
<td>YES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Timex Value</td>
<td>YES</td>
<td>YES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Timex DCT relation</td>
<td>YES</td>
<td>YES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Event’s semantic role⁵</td>
<td>YES</td>
<td>YES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Event’s argument’s ont-type</td>
<td>YES</td>
<td>YES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TLINK event-time signal⁶</td>
<td></td>
<td>YES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLINK event-event relation type⁷</td>
<td>YES</td>
<td>YES</td>
<td>$e_1 \times e_2$</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Features used for TempEval-2 (TE2) Task C, D, E and F or TempEval-1 (TE1) Task A, B and C.
Performance in TempEval-2

<table>
<thead>
<tr>
<th>Task</th>
<th>TRIPS Precision</th>
<th>TRIPS Recall</th>
<th>TRIOS Precision</th>
<th>TRIOS Recall</th>
<th>Best (with corpus features)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task C</td>
<td>0.63</td>
<td>0.52</td>
<td>0.65</td>
<td>0.52</td>
<td>0.63 (JU-CSE, UCFD, NCSU-indi)</td>
</tr>
<tr>
<td>Task D</td>
<td>0.76</td>
<td>0.69</td>
<td>0.79</td>
<td>0.67</td>
<td>0.82 (TIPSem)</td>
</tr>
<tr>
<td>Task E</td>
<td>0.58</td>
<td>0.50</td>
<td>0.56</td>
<td>0.42</td>
<td>0.55 (TIPSem)</td>
</tr>
<tr>
<td>Task F</td>
<td>0.59</td>
<td>0.54</td>
<td>0.60</td>
<td>0.46</td>
<td>0.66 (NCSU-individual)</td>
</tr>
</tbody>
</table>

Table 1: Performance of Temporal Relations on TempEval-2 (Task C-F)

<table>
<thead>
<tr>
<th>Task</th>
<th>Description</th>
<th>Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task A</td>
<td>Temporal expression extraction</td>
<td>TRIOS</td>
</tr>
<tr>
<td>Task B</td>
<td>Event Extraction</td>
<td>TIPSem</td>
</tr>
<tr>
<td>Task C</td>
<td>Event-Timex relationship</td>
<td>TRIOS</td>
</tr>
<tr>
<td>Task D</td>
<td>Event-DCT relationship</td>
<td>TIPSem</td>
</tr>
<tr>
<td>Task E</td>
<td>Main event-event relationship</td>
<td>TRIOS</td>
</tr>
<tr>
<td>Task F</td>
<td>Subordinate event-event relationship</td>
<td>TRIOS</td>
</tr>
</tbody>
</table>

Table 1: Head-to-head comparison of TRIOS, TIPSem and JU-CSE-TEMP (teams that approached all tasks) in TempEval-2 challenge

Outline

• Our System

• Extension to TimeBank and TimeML
  • Suggest new event
  • Suggest new temporal expressions
  • Suggest ontology type as new event features
  • Suggest improved relations in TimeML

• Future Work
Suggest Temporal Links

- Suggest new event-time temporal links that are missed by TimeBank annotators
- Suggest intra-sentence event-event temporal relations, which are ignored in TempEval-1
Automatically Building Larger Temporally Annotated Corpora

- Automatically build larger temporally annotated corpus for news domain, which can be reviewed by human annotators
- Automatically build temporally annotated corpus for other domains like medical domain
Summary

• Suggested New Events in TimeBank
• Added New Event Feature - *Ontology Type*; released TRIPS ontology
• Suggested New Temporal Expressions in TimeBank
• Added Improved Relations in Existing Annotation Scheme
  • SLINK or Subordinate links - relation between events
  • RLINK or Relation link - relation between event and its arguments
• Released TRIOS-TimeBank corpus, TRIPS ontology
• TRIOS-TimeBank Corpus & TRIPS Ontology:
  http://www.cs.rochester.edu/u/naushad/trios-timebank-corpus

• Other temporal related resources:
  http://www.cs.rochester.edu/u/naushad/temporal
Questions?

- Other temporal related resources: http://www.cs.rochester.edu/u/naushad/temporal
# Event Feature Extraction Performance

<table>
<thead>
<tr>
<th>System</th>
<th>TRIPS</th>
<th>TRIOS</th>
<th>Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>0.67</td>
<td>0.77</td>
<td>0.79 (TIPSem)</td>
</tr>
<tr>
<td>Tense</td>
<td>0.67</td>
<td>0.91</td>
<td>0.92 (Edinburgh-LTG)</td>
</tr>
<tr>
<td>Aspect</td>
<td>0.97</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>Pos</td>
<td>0.88</td>
<td>0.96</td>
<td>0.97 (TIPSem, Edinburgh-LTG)</td>
</tr>
<tr>
<td>Polarity</td>
<td><strong>0.99</strong></td>
<td><strong>0.99</strong></td>
<td>0.99</td>
</tr>
<tr>
<td>Modality</td>
<td>0.95</td>
<td>0.95</td>
<td>0.99 (Edinburgh-LTG)</td>
</tr>
</tbody>
</table>

Table 1: Performance of Event Features on TempEval-2 (Task B)
Overview

Text

TRIPS parser

MLN event feature extractor

CRF timex extractor

MLN event filtering

events

event features

timex

timex features

timex normalizer
Benefits of TRIPS ontology

• Superior semantic ontology; better abstraction
• No problem with word sense disambiguation
• Considers semantic roles for disambiguation
• Helps to generate better links
Why RLINK

- RLINKs could be in a separate layer
- needed for complete temporally aware system
- hence included for a complete temporally annotated corpus
Markov Logic Network

• Problems with rule based system and machine learning techniques

• Markov logic = first order logic + markov network (probabilistic graphical model)

• FOL with weights

• weights determine how much penalty for a formula to be violated

Example: It is not going to change

tense(e1, INFINITIVE) & aspect(e1, NONE) => class(e1, OCCURRENCE)  weight = 0.319913