At least 11 people have died in new clashes with security forces in Tunisia after four weeks of unrest, it was reported today... 

Rioting against joblessness and other social ills has scarred many cities in the country since 17 December, when a 26-year-old graduate set himself on fire when police confiscated his fruits and vegetables for selling without a permit. . .

Mobs have since attacked public buildings and the local office of the party of President Zine El Abidine Ben Ali.
A 32-year-old woman was admitted to the hospital because of left subcostal pain. Three months before admission an evaluation elsewhere included a CT scan of the abdomen. She had a history of eczema and of asthma. She had lost 18 kg in weight during the preceding 18 months. An abdominal examination revealed a soft systolic bruit...
Outline

1 Introduction

2 Timeline extraction as supervised learning
   - Identifying events and times
   - Normalizing times
   - Linking events and times

3 Improving timelines with unannotated data
   - Temporal information via web queries
   - Latent structure from unsupervised models

4 Improving the model of temporal links
   - Links from linguistic constructions
   - Links from narrative containers
   - Links as dependency trees
   - Dense links through transitivity

5 Conclusions
At least 11 people have died in new clashes with security forces in Tunisia after four weeks of unrest, it was reported today. Rioting against joblessness and other social ills has scarred many cities in the country since 17 December, when a 26-year-old graduate set himself on fire...
At least 11 people have died in new clashes with security forces in Tunisia after four weeks of unrest, it was reported today. Rioting against joblessness and other social ills has scarred many cities in the country since 17 December, when a 26-year-old graduate set himself on fire...
At least 11 people have died in new clashes with security forces in Tunisia after four weeks of unrest, it was reported today. Rioting against joblessness and other social ills has scarred many cities in the country since 17 December, when a 26-year-old graduate set himself on fire when police confiscated his fruits and vegetables for selling without a permit...
At least 11 people have died in new clashes with security forces in Tunisia after four weeks of unrest, it was reported today. Rioting against joblessness and other social ills has scarred many cities in the country since 17 December, when a 26-year-old graduate set himself on fire when police confiscated his fruits...
Events and times as word classification

<table>
<thead>
<tr>
<th>x</th>
<th>$h_{\text{time+event}}(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>after</td>
<td>O</td>
</tr>
<tr>
<td>four</td>
<td>B-TIME</td>
</tr>
<tr>
<td>weeks</td>
<td>I-TIME</td>
</tr>
<tr>
<td>of</td>
<td>O</td>
</tr>
<tr>
<td>unrest</td>
<td>B-EVENT</td>
</tr>
<tr>
<td>,</td>
<td>O</td>
</tr>
<tr>
<td>it</td>
<td>O</td>
</tr>
<tr>
<td>was</td>
<td>O</td>
</tr>
<tr>
<td>reported</td>
<td>B-EVENT</td>
</tr>
<tr>
<td>today</td>
<td>B-TIME</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

Features $x = (f_1, f_2, \ldots, f_m)$:
- Word itself, e.g. weeks
- Part-of-speech, e.g. VERB
- Character category patterns, e.g. Dec $\rightarrow$ LULLLL
- Preceding/following features
- Preceding $h_{\text{time+event}}(x)$ values

Learning $h_{\text{time+event}}$:
- Logistic regression:
  $$h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^T x}}$$
- Support vector machines
- Conditional random fields

Steven Bethard (UAB)  
http://bethard.cis.uab.edu/
Events and times as word classification

\[ x \quad h_{\text{time+event}}(x) \]

...after four weeks of unrest, it was reported today...

Features \( x = (f_1, f_2, \ldots, f_m) \):
- Word itself, e.g. *weeks*
- Part-of-speech, e.g. *VERB*
- Character category patterns, e.g. *Dec \rightarrow LuLLLL*
- Preceding/following features
- Preceding \( h_{\text{time+event}}(x) \) values

Learning \( h_{\text{time+event}} \):
- Logistic regression:
  \[ h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^T x}} \]
- Support vector machines
- Conditional random fields

Steven Bethard (UAB)  http://bethard.cis.uab.edu/
Events and times as word classification

<table>
<thead>
<tr>
<th>x</th>
<th>( h_{\text{time+event}}(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>\ldots</td>
<td>\ldots</td>
</tr>
<tr>
<td>after</td>
<td>O</td>
</tr>
<tr>
<td>four weeks of unrest , it was reported today</td>
<td>( \text{B-TIME} ), ( \text{I-TIME} ), ( \text{O} ), ( \text{O} ), ( \text{O} ), ( \text{B-EVENT} ), ( \text{B-TIME} )</td>
</tr>
</tbody>
</table>

Features \( x = (f_1, f_2, \ldots, f_m) \):
- Word itself, e.g. \textit{weeks}
- Part-of-speech, e.g. \textit{VERB}
- Character category patterns, e.g \( \text{Dec} \rightarrow \text{LuLLLL} \)
- Preceding/following features
- Preceding \( h_{\text{time+event}}(x) \) values

Learning \( h_{\text{time+event}} \):
- Logistic regression:
  \[
  h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^T x}}
  \]
- Support vector machines
- Conditional random fields
Events and times as word classification

\[
x \quad h_{\text{time+event}}(x)
\]

... after four weeks of unrest, it was reported today...

Features \( x = (f_1, f_2, \ldots, f_m) \):
- Word itself, e.g. weeks
- Part-of-speech, e.g. VERB
- Character category patterns, e.g. \( Dec \rightarrow \text{LuLLLL} \)
- Preceding/following features
- Preceding \( h_{\text{time+event}}(x) \) values

Learning \( h_{\text{time+event}} \):
- Logistic regression:
  \[
  h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^T x}}
  \]
- Support vector machines
- Conditional random fields

Steven Bethard (UAB)
http://bethard.cis.uab.edu/
Events and times as word classification

<table>
<thead>
<tr>
<th>x</th>
<th>$h_{\text{time+event}}(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>after</td>
<td>O</td>
</tr>
<tr>
<td>four</td>
<td>B-TIME</td>
</tr>
<tr>
<td>weeks</td>
<td>I-TIME</td>
</tr>
<tr>
<td>of</td>
<td>O</td>
</tr>
<tr>
<td>unrest</td>
<td>B-EVENT</td>
</tr>
<tr>
<td>,</td>
<td>O</td>
</tr>
<tr>
<td>it</td>
<td>O</td>
</tr>
<tr>
<td>was reported</td>
<td>B-EVENT</td>
</tr>
<tr>
<td>today</td>
<td>B-TIME</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

Features $x = (f_1, f_2, \ldots, f_m)$:

- Word itself, e.g. *weeks*
- Part-of-speech, e.g. *VERB*
- Character category patterns, e.g. *Dec → L_uLLL_L*
- Preceding/following features
- Preceding $h_{\text{time+event}}(x)$ values

Learning $h_{\text{time+event}}$:

- Logistic regression:
  $$h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^Tx}}$$
- Support vector machines
- Conditional random fields

Steven Bethard (UAB)
http://bethard.cis.uab.edu/
Events and times as word classification

<table>
<thead>
<tr>
<th>$x$</th>
<th>$h_{\text{time+event}}(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>after</td>
<td>O</td>
</tr>
<tr>
<td>four</td>
<td>B-TIME</td>
</tr>
<tr>
<td>weeks</td>
<td>I-TIME</td>
</tr>
<tr>
<td>of</td>
<td>O</td>
</tr>
<tr>
<td>unrest</td>
<td>B-EVENT</td>
</tr>
<tr>
<td>,</td>
<td>O</td>
</tr>
<tr>
<td>it</td>
<td>O</td>
</tr>
<tr>
<td>was</td>
<td>O</td>
</tr>
<tr>
<td>reported</td>
<td>B-EVENT</td>
</tr>
<tr>
<td>today</td>
<td>B-TIME</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

Features $x = (f_1, f_2, \ldots, f_m)$:
- Word itself, e.g. *weeks*
- Part-of-speech, e.g. *VERB*
- Character category patterns, e.g. *Dec → LuLlLl*
- Preceding/following features
- Preceding $h_{\text{time+event}}(x)$ values

Learning $h_{\text{time+event}}$:
- Logistic regression:
  \[
  h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^\top x}}
  \]
- Support vector machines
- Conditional random fields
Events and times as word classification

\[
x \quad h_{\text{time+event}}(x)
\]

\[
\begin{array}{ll}
\text{after} & \text{O} \\
\text{four} & \text{B-TIME} \\
\text{weeks} & \text{I-TIME} \\
\text{of} & \text{O} \\
\text{unrest} & \text{B-EVENT} \\
, & \text{O} \\
\text{it} & \text{O} \\
\text{was} & \text{O} \\
\text{reported} & \text{B-EVENT} \\
\text{today} & \text{B-TIME} \\
\end{array}
\]

Features \( x = (f_1, f_2, \ldots, f_m) \):
- Word itself, e.g. \textit{weeks}
- Part-of-speech, e.g. \texttt{VERB}
- Character category patterns, e.g. \textit{Dec} \rightarrow \texttt{LULLLL}
- Preceding/following features
- Preceding \( h_{\text{time+event}}(x) \) values

Learning \( h_{\text{time+event}} \):
- Logistic regression:
  \[
  h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^Tx}}
  \]
- Support vector machines
- Conditional random fields

Steven Bethard (UAB)  http://bethard.cis.uab.edu/
Events and times as word classification

<table>
<thead>
<tr>
<th>x</th>
<th>$h_{time+event}(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>after</td>
<td>O</td>
</tr>
<tr>
<td>four</td>
<td>B-TIME</td>
</tr>
<tr>
<td>weeks</td>
<td>I-TIME</td>
</tr>
<tr>
<td>of</td>
<td>O</td>
</tr>
<tr>
<td>unrest</td>
<td>B-EVENT</td>
</tr>
<tr>
<td>,</td>
<td>O</td>
</tr>
<tr>
<td>it</td>
<td>O</td>
</tr>
<tr>
<td>was</td>
<td>O</td>
</tr>
<tr>
<td>reported</td>
<td>B-EVENT</td>
</tr>
<tr>
<td>today</td>
<td>B-TIME</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

Features $x = (f_1, f_2, \ldots, f_m)$:
- Word itself, e.g. weeks
- Part-of-speech, e.g. VERB
- Character category patterns, e.g. $Dec \rightarrow LLLL $
- Preceding/following features
- Preceding $h_{time+event}(x)$ values

Learning $h_{time+event}$:
- Logistic regression:
  $$ h_{time+event}(x) = \frac{1}{1 + e^{-w^T x}} $$
- Support vector machines
- Conditional random fields
Events and times as word classification

<table>
<thead>
<tr>
<th>$x$</th>
<th>$h_{\text{time+event}}(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>after</td>
<td>O</td>
</tr>
<tr>
<td>four</td>
<td>B-TIME</td>
</tr>
<tr>
<td>weeks</td>
<td>I-TIME</td>
</tr>
<tr>
<td>of</td>
<td>O</td>
</tr>
<tr>
<td>unrest</td>
<td>B-EVENT</td>
</tr>
<tr>
<td>,</td>
<td>O</td>
</tr>
<tr>
<td>it</td>
<td>O</td>
</tr>
<tr>
<td>was</td>
<td>O</td>
</tr>
<tr>
<td>reported</td>
<td>B-EVENT</td>
</tr>
<tr>
<td>today</td>
<td>B-TIME</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

Features $x = (f_1, f_2, \ldots, f_m)$:
- Word itself, e.g. *weeks*
- Part-of-speech, e.g. *VERB*
- Character category patterns, e.g. $Dec \rightarrow L_uL_{LL}$
- Preceding/following features
- Preceding $h_{\text{time+event}}(x)$ values

Learning $h_{\text{time+event}}$:
- Logistic regression:
  $$h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^T x}}$$
- Support vector machines
- Conditional random fields

Steven Bethard (UAB) 
http://bethard.cis.uab.edu/
Events and times as word classification

<table>
<thead>
<tr>
<th>x</th>
<th>$h_{\text{time+event}}(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>after</td>
<td>O</td>
</tr>
<tr>
<td>four</td>
<td>B-TIME</td>
</tr>
<tr>
<td>weeks</td>
<td>I-TIME</td>
</tr>
<tr>
<td>of</td>
<td>O</td>
</tr>
<tr>
<td>unrest</td>
<td>B-EVENT</td>
</tr>
<tr>
<td>,</td>
<td>O</td>
</tr>
<tr>
<td>it</td>
<td>O</td>
</tr>
<tr>
<td>was</td>
<td>O</td>
</tr>
<tr>
<td>reported</td>
<td>B-EVENT</td>
</tr>
<tr>
<td>today</td>
<td>B-TIME</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

Features $x = (f_1, f_2, \ldots, f_m)$:
- Word itself, e.g. weeks
- Part-of-speech, e.g. VERB
- Character category patterns, e.g. $Dec \rightarrow \text{LULLL}$
- Preceding/following features
- Preceding $h_{\text{time+event}}(x)$ values

Learning $h_{\text{time+event}}$:
- Logistic regression:
  $$h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^T x}}$$
- Support vector machines
- Conditional random fields
Events and times as word classification

<table>
<thead>
<tr>
<th>x</th>
<th>$h_{\text{time+event}}(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>after</td>
<td>O</td>
</tr>
<tr>
<td>four</td>
<td>B-TIME</td>
</tr>
<tr>
<td>weeks</td>
<td>I-TIME</td>
</tr>
<tr>
<td>of</td>
<td>O</td>
</tr>
<tr>
<td>unrest</td>
<td>B-EVENT</td>
</tr>
<tr>
<td>,</td>
<td>O</td>
</tr>
<tr>
<td>it</td>
<td>O</td>
</tr>
<tr>
<td>was</td>
<td>O</td>
</tr>
<tr>
<td>reported</td>
<td>B-EVENT</td>
</tr>
<tr>
<td>today</td>
<td>B-TIME</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

Features $x = (f_1, f_2, \ldots, f_m)$:
- Word itself, e.g. weeks
- Part-of-speech, e.g. VERB
- Character category patterns, e.g. $Dec \rightarrow LuLLLl$
- Preceding/following features
- Preceding $h_{\text{time+event}}(x)$ values

Learning $h_{\text{time+event}}$:
- Logistic regression:
  $$h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^T x}}$$
- Support vector machines
- Conditional random fields
Events and times as word classification

<table>
<thead>
<tr>
<th>$x$</th>
<th>$h_{\text{time+event}}(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>after</td>
<td>O</td>
</tr>
<tr>
<td>four</td>
<td>B-TIME</td>
</tr>
<tr>
<td>weeks</td>
<td>I-TIME</td>
</tr>
<tr>
<td>of</td>
<td>O</td>
</tr>
<tr>
<td>unrest</td>
<td>B-EVENT</td>
</tr>
<tr>
<td>,</td>
<td>O</td>
</tr>
<tr>
<td>it</td>
<td>O</td>
</tr>
<tr>
<td>was</td>
<td>O</td>
</tr>
<tr>
<td>reported</td>
<td>B-EVENT</td>
</tr>
<tr>
<td>today</td>
<td>B-TIME</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

Features $x = (f_1, f_2, \ldots, f_m)$:
- Word itself, e.g. *weeks*
- Part-of-speech, e.g. *VERB*
- Character category patterns, e.g. $Dec \rightarrow \text{LULLLL}$
- Preceding/following features
- Preceding $h_{\text{time+event}}(x)$ values

Learning $h_{\text{time+event}}$:
- Logistic regression:
  $$h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^T x}}$$
- Support vector machines
- Conditional random fields

Steven Bethard (UAB)  
http://bethard.cis.uab.edu/
Events and times as word classification

<table>
<thead>
<tr>
<th>x</th>
<th>( h_{\text{time+event}}(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>after</td>
<td>O</td>
</tr>
<tr>
<td>four</td>
<td>B-TIME</td>
</tr>
<tr>
<td>weeks</td>
<td>I-TIME</td>
</tr>
<tr>
<td>of</td>
<td>O</td>
</tr>
<tr>
<td>unrest</td>
<td>B-EVENT</td>
</tr>
<tr>
<td>,</td>
<td>O</td>
</tr>
<tr>
<td>it</td>
<td>O</td>
</tr>
<tr>
<td>was</td>
<td>O</td>
</tr>
<tr>
<td>reported</td>
<td>B-EVENT</td>
</tr>
<tr>
<td>today</td>
<td>B-TIME</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

Features \( x = (f_1, f_2, \ldots, f_m) \):

- Word itself, e.g. *weeks*
- Part-of-speech, e.g. *VERB*
- Character category patterns, e.g. 
  \( Dec \rightarrow L\underline{u}L\underline{L}L\underline{L}L \)
- Preceding/following features
- Preceding \( h_{\text{time+event}}(x) \) values

Learning \( h_{\text{time+event}} \):

- Logistic regression:
  \[
  h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^T x}}
  \]
- Support vector machines
- Conditional random fields
Events and times as word classification

<table>
<thead>
<tr>
<th>$x$</th>
<th>$h_{\text{time+event}}(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>after four weeks of unrest , it was reported today ...</td>
<td>O B-TIME I-TIME O O B-EVENT O B-TIME</td>
</tr>
</tbody>
</table>

Features $x = (f_1, f_2, \ldots, f_m)$:

- Word itself, e.g. weeks
- Part-of-speech, e.g. VERB
- Character category patterns, e.g. $Dec \rightarrow LULLLL$
- Preceding/following features
- Preceding $h_{\text{time+event}}(x)$ values

Learning $h_{\text{time+event}}$:

- Logistic regression:
  
  $$h_{\text{time+event}}(x) = \frac{1}{1 + e^{-w^T x}}$$
- Support vector machines
- Conditional random fields

Steven Bethard (UAB)  
http://bethard.cis.uab.edu/
Normalizing times using rules

Normalizing time expressions:

- *December 5, 2007* ⇒ *2007-12-05*
- *the day before yesterday* ⇒ *2014-07-07*

Modeled with hand-constructed grammars (e.g. Bethard 2013):

- *the week of March 6*

```
FindEnclosing
  ┌─ TimeSpan ─┐
  │           │
  │      ┌─ TimeSpan ─┐
  │      │      ┌─ Unit ─┐
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  │      │      │      │
  └─ 2014-W10 ┘
```

- *today=* *2014-07-09*
Normalizing times using rules

Normalizing time expressions:

- *December 5, 2007* ⇒ *2007-12-05*
- *the day before yesterday* ⇒ *2014-07-07*

Modeled with hand-constructed grammars (e.g. Bethard 2013):

\[
\textit{the week of March 6}
\]

\[
\begin{array}{c}
\text{FindEnclosing} \\
\hline
\text{FindEarlier} \\
\hline
\text{Present} \\
\hline
\text{MonthOfYear} 3 \\
\hline
\text{DayOfMonth} 6 \\
\hline
\text{Weeks} 2014-W10
\end{array}
\]

\[
\text{today}=2014-07-09
\]
Linking events and times as pair classification

\[
\begin{array}{ll}
  x & h_{\text{link}}(x) \\
  \text{(died, clashes)} & \text{INCLUDED} \\
  \text{(died, four weeks)} & \text{AFTER} \\
  \text{(died, unrest)} & \text{AFTER} \\
  \text{(died, today)} & \text{INCLUDED} \\
  \text{(clashes, four weeks)} & \text{AFTER} \\
  \text{(clashes, unrest)} & \text{AFTER} \\
  \text{(clashes, today)} & \text{INCLUDED} \\
  \text{(four weeks, unrest)} & \text{INCLUDES} \\
  \text{(four weeks, today)} & \text{BEFORE} \\
  \text{(unrest, today)} & \text{BEFORE} \\
\end{array}
\]
Linking events and times as pair classification

<table>
<thead>
<tr>
<th>x</th>
<th>$h_{\text{link}}(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(died, clashes)</td>
<td>INCLUDED</td>
</tr>
<tr>
<td>(died, four weeks)</td>
<td>AFTER</td>
</tr>
<tr>
<td>(died, unrest)</td>
<td>AFTER</td>
</tr>
<tr>
<td>(died, today)</td>
<td>INCLUDED</td>
</tr>
<tr>
<td>(clashes, four weeks)</td>
<td>AFTER</td>
</tr>
<tr>
<td>(clashes, unrest)</td>
<td>AFTER</td>
</tr>
<tr>
<td>(clashes, today)</td>
<td>INCLUDED</td>
</tr>
<tr>
<td>(four weeks, unrest)</td>
<td>INCLUDES</td>
</tr>
<tr>
<td>(four weeks, today)</td>
<td>BEFORE</td>
</tr>
<tr>
<td>(unrest, today)</td>
<td>BEFORE</td>
</tr>
</tbody>
</table>

Steven Bethard (UAB)  http://bethard.cis.uab.edu/
Linking events and times as pair classification

Features $x = (f_1, f_2, \ldots, f_m)$

- Event\textsubscript{1} part-of-speech
- Event\textsubscript{2} part-of-speech
- Bag of words between
- ...
- Path in syntactic tree

Learning $h_{\text{link}}$:

- Support vector machine
- Logistic regression
- ...

Steven Bethard (UAB)  
http://bethard.cis.uab.edu/
How well does it work?


- Common annotated training set: events, times, links
- Common unannotated test set for system predictions

Evaluation metrics:

- **Precision** = \( \frac{\text{# correct predictions}}{\text{# predictions}} \)
- **Recall** = \( \frac{\text{# correct predictions}}{\text{# actual}} \)
- **\( F_1 \)** = \( \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \)
How well does it work?

TempEval 2007:
- Link classification: 55-80% accuracy

TempEval 2010:
- Event identification: 81% precision, 86% recall
- Time identification: 90% precision, 82% recall
- Time normalization: 85% accuracy
- Link classification: 55-81% accuracy

TempEval 2013:
- Event identification: 81% precision, 81% recall
- Time identification: 86% precision, 80% recall
- Time normalization: 86% accuracy
- Link identification: 37% precision, 35% recall
Some remaining challenges

Current features aren’t predictive enough:

- *He waited* there and *looked* around. (INCLUDES)
- *He arrived* there and *looked* around. (BEFORE)

TempEval data is small and sparse:

- no examples of *autumn*
- only 1 example of *winter*

TempEval links are incomplete:

- E.g., *Farkas was ordered* home and *retired.*
Some remaining challenges

Current features aren’t predictive enough:

- *He waited* there and *looked* around. *(INCLUDES)*
- *He arrived* there and *looked* around. *(BEFORE)*

TempEval data is small and sparse:

- no examples of *autumn*
- only 1 example of *winter*

TempEval links are incomplete:

- E.g., Farkas was *ordered* home and *retired*. 
Some remaining challenges

Current features aren’t predictive enough:

- He \textcolor{blue}{waited} there and \textcolor{blue}{looked} around. (INCLUDES)
- He \textcolor{blue}{arrived} there and \textcolor{blue}{looked} around. (BEFORE)

TempEval data is small and sparse:

- no examples of \textit{autumn}
- only 1 example of \textit{winter}

TempEval links are incomplete:

- E.g., \textit{Farkas was ordered} home and \textcolor{blue}{retired}. 
Intuition: Duration information should help

- He \textcolor{blue}{waited} there and \textcolor{blue}{looked} around. (INCLUDES)
- He \textcolor{blue}{arrived} there and \textcolor{blue}{looked} around. (BEFORE)

Approach: ask the web how long \textcolor{blue}{waiting} takes
Developing features that capture semantics
(Gusev, Chambers, et al. 2011)

Intuition: Duration information should help
- He waited there and looked around. (INCLUDES)
- He arrived there and looked around. (BEFORE)

Approach: ask the web how long waiting takes

\[ P(\text{"spent * unit waiting"}) \]

---

Steven Bethard (UAB)
http://bethard.cis.uab.edu/
Developing features that capture semantics

(Gusev, Chambers, et al. 2011)

Intuition: Duration information should help

- He \textit{waited} there and \textit{looked} around.  \textit{(INCLUDES)}
- He \textit{arrived} there and \textit{looked} around.  \textit{(BEFORE)}

Approach: ask the web how long \textit{waiting} takes

\begin{equation}
P(\text{“spent * \text{UNIT} waiting”} | \text{“spent * \text{UNIT}”})
\end{equation}

![Bar chart showing the distribution of time units for looking and waiting](chart.png)

Web patterns as classifier $>$ training on 1700 annotations
Intuition: Duration information should help

- He waited there and looked around. (INCLUDES)
- He arrived there and looked around. (BEFORE)

Approach: ask the web how long waiting takes

Web patterns as classifier > training on 1700 annotations
Developing features that capture semantics

(Gusev, Chambers, et al. 2011)

Intuition: Duration information should help

- He waited there and looked around. (INCLUDES)
- He arrived there and looked around. (BEFORE)

Approach: ask the web how long waiting takes

Web patterns as classifier > training on 1700 annotations
Managing sparsity with unsupervised models

(Kolomiyets, Bethard, et al. 2011)

TempEval data is small and sparse:
- no examples of *autumn*
- only 1 example of *winter*

Latent Words Language Model (Deschacht and Moens 2009)

Generate training examples
- TempEval train, TempEval test: 0.820 $\rightarrow$ 0.861 $F_1$
- TempEval train, Reuters test: 0.773 $\rightarrow$ 0.826 $F_1$
- Better than synonym expansion with WordNet
Managing sparsity with unsupervised models

(Kolomiyets, Bethard, et al. 2011)

TempEval data is small and sparse:
- no examples of *autumn*
- only 1 example of *winter*

Latent Words Language Model (Deschacht and Moens 2009)

Generate training examples
- TempEval train, TempEval test: 0.820 $\rightarrow$ 0.861 $F_1$
- TempEval train, Reuters test: 0.773 $\rightarrow$ 0.826 $F_1$
- Better than synonym expansion with WordNet
Managing sparsity with unsupervised models

(Kolomiyets, Bethard, et al. 2011)

TempEval data is small and sparse:
- no examples of *autumn*
- only 1 example of *winter*

**Latent Words Language Model** (Deschacht and Moens 2009)

Generate training examples
- TempEval train, TempEval test: 0.820 → 0.861 $F_1$
- TempEval train, Reuters test: 0.773 → 0.826 $F_1$
- Better than synonym expansion with WordNet
Managing sparsity with unsupervised models

(Kolomiyets, Bethard, et al. 2011)

TempEval data is small and sparse:

- no examples of *autumn*
- only 1 example of *winter*

Latent Words Language Model (Deschacht and Moens 2009)

![Diagram of Latent Words Language Model](image)

Generate training examples

- TempEval train, TempEval test: 0.820 → 0.861 $F_1$
- TempEval train, Reuters test: 0.773 → 0.826 $F_1$
- Better than synonym expansion with WordNet

Steven Bethard (UAB) http://bethard.cis.uab.edu/
TempEval relation annotations are unintuitive

Arbitrary links in TempEval 2007:

*Turning its back on* 210 years of loyalty to the British royal family, a constitutional convention voted overwhelmingly Friday to *make* Australia a republic under its own president.

Missing links in TempEval 2010

The World Court *Friday* rejected U.S. and British objections to a Libyan World Court case that has blocked the trial of two Libyans suspected of blowing up a Pan Am jumbo jet over Scotland in 1988.
More linguistically plausible links

(Bethard, Martin, et al. 2007)

More natural links using syntax:

*Turning* its back on
210 years of loyalty to the
British royal family, a
constitutional convention
voted overwhelmingly
*Friday* to *make* Australia a
republic under its own
president.
More linguistically plausible links

(Bethard, Martin, et al. 2007)

More natural links using syntax:

**Turning** its back on **210 years** of loyalty to the **British royal family**, a **constitutional convention** **voted** overwhelmingly on **Friday** to **make** Australia a republic under its own **president**.
More linguistically plausible links

(Bethard, Martin, et al. 2007)

More natural links using syntax:

Turning its back on 210 years of loyalty to the British royal family, a constitutional convention voted overwhelmingly Friday to make Australia a republic under its own president.

Case study on 132 document corpus:

- Easier for annotators (90% agreement)
- Easier for models (89% accuracy)
President Obama paid tribute Sunday to 29 workers killed in an explosion at a West Virginia coal mine earlier this month, saying they died “in pursuit of the American dream.” The blast at the Upper Big Branch Mine was the worst U.S. mine disaster in nearly 40 years.

Case study re-annotating the TimeBank:
- Easier for annotators (Kappa 0.74)

Case study with clinical narratives:
- Tractable for models (Miller, Bethard, et al. 2013)
President Obama paid tribute Sunday to 29 workers killed in an explosion at a West Virginia coal mine earlier this month, saying they died “in pursuit of the American dream.” The blast at the Upper Big Branch Mine was the worst U.S. mine disaster in nearly 40 years.
President Obama paid tribute Sunday to 29 workers killed in an explosion at a West Virginia coal mine earlier this month, saying they died “in pursuit of the American dream.” The blast at the Upper Big Branch Mine was the worst U.S. mine disaster in nearly 40 years.

Case study re-annotating the TimeBank:
- Easier for annotators (Kappa 0.74)

Case study with clinical narratives:
- Tractable for models (Miller, Bethard, et al. 2013)
Timelines as dependency trees
(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them
(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other ... threw himself on the ground and pretended to be dead.
Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other ... threw himself on the ground and pretended to be dead.
Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnon-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other . . . threw himself on the ground and pretended to be dead.
Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other . . . threw himself on the ground and pretended to be dead.
Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other ... threw himself on the ground and pretended to be dead.
Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other ... threw himself on the ground and pretended to be dead.
Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other ... threw himself on the ground and pretended to be dead.
Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other ... threw himself on the ground and pretended to be dead.
Timelines as dependency trees
(Kolomiyets, Bethard, et al. 2012)

Intuition:  link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on
the scene. Before he observed them, one made for a tree at the side of the
road, and climbed up into the branches and hid. The other . . . threw himself
on the ground and pretended to be dead.
Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other ... threw himself on the ground and pretended to be dead.
Timelines as dependency trees

(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other ... threw himself on the ground and pretended to be dead.
Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other . . . threw himself on the ground and pretended to be dead.
Timelines as dependency trees
(Kolomiyets, Bethard, et al. 2012)

Intuition: link events/times as you read them
(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were **on** the road together, when a bear suddenly **appeared** on the scene. Before he **observed** them, one **made** for a tree at the side of the road, and **climbed** up into the branches and **hid**. The other ... **threw** himself **on** the ground and **pretended** to be dead.
Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other ... threw himself on the ground and pretended to be dead.
Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other ... threw himself on the ground and pretended to be dead.
Intuition: link events/times as you read them

(Johnson-Laird 1980; Brewer and Lichtenstein 1982)

Two travelers were on the road together, when a bear suddenly appeared on the scene. Before he observed them, one made for a tree at the side of the road, and climbed up into the branches and hid. The other ... threw himself on the ground and pretended to be dead.
Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing
Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing
Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing
Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

Shift-reduce timeline parsing outperforms:
- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing
Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

Shift-reduce timeline parsing outperforms:
- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing
Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

<table>
<thead>
<tr>
<th>Stack</th>
<th>Queue</th>
</tr>
</thead>
<tbody>
<tr>
<td>appeared</td>
<td>observed</td>
</tr>
<tr>
<td>on</td>
<td>RE/DUCE INCLUDES</td>
</tr>
</tbody>
</table>

Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing
Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

Shift-reduce timeline parsing outperforms:
- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing
Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree tree parsing
Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

Shift-reduce timeline parsing outperforms:
- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing
Timeline extraction as shift-reduce parsing
(Kolomiyets, Bethard, et al. 2012)

Shift-reduce timeline parsing outperforms:
- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing
Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

Shift-reduce timeline parsing outperforms:
- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing
Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing
Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

<table>
<thead>
<tr>
<th>Stack</th>
<th>Queue</th>
</tr>
</thead>
<tbody>
<tr>
<td>made→observed</td>
<td>climbed</td>
</tr>
<tr>
<td>on→appeared</td>
<td>hid</td>
</tr>
</tbody>
</table>

Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing

Steven Bethard (UAB) [http://bethard.cis.uab.edu/](http://bethard.cis.uab.edu/)
Timeline extraction as shift-reduce parsing
(Kolomiyets, Bethard, et al. 2012)

Shift-reduce timeline parsing outperforms:
- Simple pair-wise classification
- Graph-based maximum-spanning-tree tree parsing
Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

Shift-reduce timeline parsing outperforms:
- Simple pair-wise classification
- Graph-based maximum-spanning-tree tree parsing
Timeline extraction as shift-reduce parsing

(Kolomiyets, Bethard, et al. 2012)

<table>
<thead>
<tr>
<th>Stack</th>
<th>Queue</th>
</tr>
</thead>
<tbody>
<tr>
<td>on → appeared → made → observed</td>
<td>climbed hid threw pretended ...</td>
</tr>
</tbody>
</table>

Shift-reduce timeline parsing outperforms:

- Simple pair-wise classification
- Graph-based maximum-spanning-tree parsing
There were four or five people inside and they just started firing. Ms. Sanders was hit several times and was pronounced dead at the scene.

Dense annotation:
There were four or five people inside and they just started firing. Ms. Sanders was hit several times and was pronounced dead at the scene.

Dense annotation:

- **firing** INCLUDES **hit**

There were four or five people inside and they just started **firing**. Ms. Sanders was **hit** several times and was **pronounced** **dead** at the scene.

Dense annotation:
There were four or five people inside and they just started firing. Ms. Sanders was hit several times and was pronounced dead at the scene.

Dense annotation:
There were four or five people inside and they just started **firing**. Ms. Sanders was **hit** several times and was **pronounced** **dead** at the scene.

Dense annotation:

- **firing** → INCLUDES → **hit**
- **firing** → **pronounced**
- **hit** → **pronounced**
- **firing** → **dead**
There were four or five people inside and they just started firing. Ms. Sanders was hit several times and was pronounced dead at the scene.

Dense annotation:
There were four or five people inside and they just started **firing**. Ms. Sanders was **hit** several times and was **pronounced dead** at the scene.

Dense annotation:
Dense links through transitivity

There were four or five people inside and they just started firing. Ms. Sanders was hit several times and was pronounced dead at the scene.

Dense annotation:

Cascade model:
- Rule-based & learned classifiers
- Classifiers sorted by precision
- Transitivity after each classifier
- Earlier classifiers constrain later

On dense link annotations, the cascade model outperforms the top system from TempEval 2013
Conclusions

Summary:

- Timelines represented as event-time graphs
- Events and times accurately identified with supervised classifiers
  - Times normalized using rules
  - Web patterns, unsupervised models for semantics
- Linking events and times is challenging
  - Syntax, narrative containers suggest better links
  - Dependency parsing and using transitivity improve coverage

Open issues:

- Define “event” in a domain-independent way
- Normalize event-relative times, e.g. “3 weeks postop”
- Achieve high agreement + coverage in relation annotation
Thanks!

University of Colorado:
- Jim Martin
- Martha Palmer
- Tammy Sumner
- Sara Klingenstein
- Will Corvey
- Will Styler

Stanford University:
- Dan Jurafsky
- Andrey Gusev

KU Leuven, Belgium:
- Marie-Francine Moens
- Oleksandr Kolomiyets

United States Naval Academy:
- Nate Chambers

Brandeis University:
- James Pustejovsky
- Marc Verhagen

Children’s Hospital Boston:
- Guergana Savova
- Tim Miller
- Dima Dligach
- Chen Lin
- Sameer Pradhan

And to i2b2 AUG NLP Workshop for inviting me here!

Steven Bethard (UAB)
http://bethard.cis.uab.edu/
Event durations: looking, waiting, arriving

- looking: seconds > minutes > hours > days > weeks > months > years
- waiting: minutes > hours > days > years
- arriving: seconds > minutes > days > weeks > months > years


