Extending the Entity Grid With Entity-Specific Features

Micha Elsner    Eugene Charniak

School of Informatics
University of Edinburgh

Department of Computer Science
Brown University

June 21, 2011
Coherence

Structure of information in a discourse—
Gives readers context they need...
to understand new information

Coherent text

Alice was sitting by her sister.
Suddenly a White Rabbit ran by her.
Alice heard the Rabbit say “I shall be late!”

Incoherent text

Alice heard the Rabbit say “I shall be late!”
The Mouse did not notice this question.
“It isn’t”, said the Caterpillar.
Modeling coherence

**Entity Grid** (Lapata+Barzilay ‘05), (Barzilay+Lapata ‘05)
Extremely popular model...

Used for:
- Multidocument summarization eg (Soricut+Marcu ‘06)
- Essay scoring (Burstein et al. ‘10)
- Readability prediction eg (Pitler et al. ‘10)
- Story generation (McIntyre+Lapata ‘10)
- Chat disentanglement (Elsner+Charniak ‘11)

We improve results on news text
Overview

How the entity grid works

Adding some features

Experiments on news text

Conclusion
Intuitions

A text is about entities: things in the world

Suddenly a White Rabbit ran by her. Alice heard the Rabbit say “I shall be late!” The Rabbit took a watch out of its pocket. Alice started to her feet.

Coherence created by repeated entity mentions
More specific theories, eg Centering (Grosz+Sidner)
<table>
<thead>
<tr>
<th>Text</th>
<th>Syntactic role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suddenly a White Rabbit ran by her. Alice heard the Rabbit say “I shall be late!” The Rabbit took a watch out of its pocket. Alice started to her feet.</td>
<td>subject</td>
</tr>
<tr>
<td></td>
<td>object</td>
</tr>
<tr>
<td></td>
<td>subject</td>
</tr>
<tr>
<td></td>
<td>missing</td>
</tr>
</tbody>
</table>
Representing a text

<table>
<thead>
<tr>
<th>Text</th>
<th>Syntactic role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suddenly <strong>a White Rabbit</strong> ran by her. Alice heard <strong>the Rabbit</strong> say “I shall be late!” <strong>The Rabbit</strong> took a watch out of its pocket. Alice started to her feet.</td>
<td><strong>subject</strong></td>
</tr>
</tbody>
</table>

**Grid**

<table>
<thead>
<tr>
<th><strong>White Rabbit</strong></th>
<th>subj</th>
<th>obj</th>
<th>subj</th>
<th>—</th>
</tr>
</thead>
</table>
Representing a text

<table>
<thead>
<tr>
<th>Text</th>
<th>Syntactic role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suddenly <strong>a White Rabbit</strong> ran by her.</td>
<td><strong>subject</strong></td>
</tr>
<tr>
<td>Alice heard <strong>the Rabbit</strong> say “I shall be late!”</td>
<td><strong>object</strong></td>
</tr>
<tr>
<td><strong>The Rabbit</strong> took a watch out of its pocket.</td>
<td><strong>subject</strong></td>
</tr>
<tr>
<td>Alice started to her feet.</td>
<td><strong>missing</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grid</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>White Rabbit</strong></td>
</tr>
<tr>
<td><strong>Alice</strong></td>
</tr>
</tbody>
</table>
Representing a text

<table>
<thead>
<tr>
<th>Text</th>
<th>Syntactic role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suddenly a White Rabbit ran by her. Alice heard the Rabbit say “I shall be late!” The Rabbit took a watch out of its pocket. Alice started to her feet.</td>
<td>subject (\text{subject})  object (\text{object})  subject (\text{subject})  missing (\text{missing})</td>
</tr>
</tbody>
</table>

**Grid**

<table>
<thead>
<tr>
<th>White Rabbit</th>
<th>subj</th>
<th>obj</th>
<th>subj</th>
<th>–</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>other</td>
<td>subj</td>
<td>–</td>
<td>subj</td>
</tr>
<tr>
<td>watch</td>
<td>–</td>
<td>–</td>
<td>obj</td>
<td>–</td>
</tr>
<tr>
<td>feet</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>other</td>
</tr>
</tbody>
</table>
Modeling (simplified)

Entities treated independently...
Modeled via Markov chain:

**White Rabbit**

Generative and discriminative grids both use these features
Just what is an entity?

Coreference?
We don’t use it!

▸ Only sometimes improves results \(^{(\text{Barzilay+Lapata} \ ‘05)}\)...
▸ Input documents must be fairly coherent
▸ Instead: link mentions with same head noun

Mention detection?
Use all nouns as mentions.

▸ Pick up premodifiers like “a \textbf{Bush} spokesman”
▸ Maximize coreference recall
▸ Improves over NPs as mentions by 4%
Features from previous work

| White Rabbit = subj | obj of previous subj of prev-1 occurs 3x total |

These features aren’t enough...

*White Rabbit* vs *watch*
Features of important entities

**White Rabbit** = subj

- obj of previous subj of prev-1
- occurs 3x total
- is a proper NP
- is named entity class PERSON
- has some modifiers
- is singular

Features separate **White Rabbit** from **watch**

Similar features useful in coref/summary tasks
Coreference features

Spurious entities

Formed around nouns like “care”, “increase”, “percent”
(Elsner+Charniak ‘10)
Don’t throw away, but should distinguish

\[
\text{an increase} = \text{subj} \quad \text{obj of previous}
\]

... in MUC6, but never coreferent rarely has coreferent pronouns

- Automatic pronoun coreference on large dataset
  - (Charniak+Elsner ‘09)
What we learn

Baseline

\[ P(\text{May 25/President Bush} = \text{subj} | \text{missing in previous other in prev-1 occurs 3x total}) = .045 \]
What we learn

Baseline

\[ P(\text{May 25/President Bush} = \text{subj} | \text{missing in previous other in prev-1 occurs 3x total}) = .045 \]

Our model

\[ P(\text{May 25} = \text{subj} | \text{missing in previous } \ldots \text{NE type DATE never corefers in MUC6}) = .001 \]
What we learn

Baseline

\[ P(\text{May 25/President Bush} = \text{subj} \mid \text{missing in previous other in prev-1 occurs 3x total}) = .045 \]

Our model

\[ P(\text{President Bush} = \text{subj} \mid \text{missing in previous} \ldots \text{NE type PERSON proper NP corefers in MUC6 modifiers}) = .133 \]
Standard ordering benchmarks

Discrimination

Insertion

removed sentence
Results on WSJ test

<table>
<thead>
<tr>
<th></th>
<th>Disc.</th>
<th>Ins.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>50</td>
<td>13</td>
</tr>
<tr>
<td>(Elsner+Charniak ’08)</td>
<td>80</td>
<td>23</td>
</tr>
</tbody>
</table>

- (Elsner+Charniak ‘08): previous best on corpus
- Combined model: entity grid and coref-inspired
### Results on WSJ test

<table>
<thead>
<tr>
<th></th>
<th>Disc.</th>
<th>Ins.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>50</td>
<td>13</td>
</tr>
<tr>
<td><em>(Elsner+Charniak ‘08)</em></td>
<td>80</td>
<td>23</td>
</tr>
<tr>
<td>Grid (all nouns mentions)</td>
<td>80</td>
<td>21</td>
</tr>
</tbody>
</table>

- *(Elsner+Charniak ‘08)*: previous best on corpus
  - Combined model: entity grid and coref-inspired
- Mention detection nearly equals baseline
Results on WSJ test

<table>
<thead>
<tr>
<th></th>
<th>Disc.</th>
<th>Ins.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>50</td>
<td>13</td>
</tr>
<tr>
<td><em>(Elsner+Charniak ‘08)</em></td>
<td>80</td>
<td>23</td>
</tr>
<tr>
<td>Grid (all nouns mentions)</td>
<td>80</td>
<td>21</td>
</tr>
<tr>
<td><strong>Extended Grid</strong></td>
<td><strong>84</strong></td>
<td><strong>24</strong></td>
</tr>
</tbody>
</table>

- *(Elsner+Charniak ‘08)*: previous best on corpus
  - Combined model: entity grid and coref-inspired
- Mention detection nearly equals baseline
- Extended grid features +4%, 3%
Our combined models

<table>
<thead>
<tr>
<th></th>
<th>Disc.</th>
<th>Ins.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Elsner+Charniak ‘08)</td>
<td>80</td>
<td>23</td>
</tr>
<tr>
<td>Extended Grid</td>
<td>84</td>
<td>24</td>
</tr>
<tr>
<td>ExtEGrid+combo</td>
<td>86</td>
<td>27</td>
</tr>
</tbody>
</table>

- Reimplemented model combination
  - Using other models from (Elsner+Charniak ‘08)
- Combined adds +2%, 3%
Conclusion

Improving the entity grid:
- Detect all possible mentions
- Entity type features capture importance
- Coreference features without running coreference

Better results on WSJ; other tasks/domains?

Software available
bitbucket.org/melsner/browncoherence