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

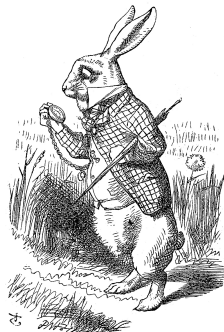How the entity grid works

Adding some features

Experiments on news text

Conclusion

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)

Representing a text

Text	Syntactic role
Suddenly **a White Rabbit** ran by her.	**subject**
Alice heard **the Rabbit** say "I shall be late!"	**object**
The Rabbit took a watch out of its pocket.	**subject**
Alice started to her feet.	**missing**

Representing a text

Text	Syntactic role
Suddenly **a White Rabbit** ran by her.	**subject**
Alice heard **the Rabbit** say "I shall be late!"	**object**
The Rabbit took a watch out of its pocket.	**subject**
Alice started to her feet.	**missing**

Grid

White Rabbit │ subj │ obj │ subj │ –

Representing a text

Text	Syntactic role
Suddenly **a White Rabbit** ran by her.	**subject**
Alice heard **the Rabbit** say "I shall be late!"	**object**
The Rabbit took a watch out of its pocket.	**subject**
Alice started to her feet.	**missing**

Grid

White Rabbit	subj	obj	subj	–
Alice	other	subj	–	subj

Representing a text

Text	Syntactic role
Suddenly **a White Rabbit** ran by her.	**subject**
Alice heard **the Rabbit** say "I shall be late!"	**object**
The Rabbit took a watch out of its pocket.	**subject**
Alice started to her feet.	**missing**

Grid

White Rabbit	subj	obj	subj	–
Alice	other	subj	–	subj
watch	–	–	obj	–
feet	–	–	–	other

Modeling (simplified)

Entities treated independently...
Modeled via Markov chain:

White Rabbit
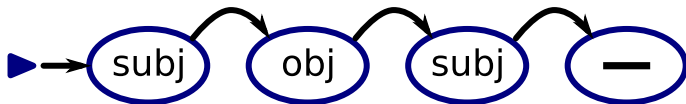
Generative and discriminative grids
both use these features

Just what is an entity?

Coreference?
We don't use it!
- Only sometimes improves results (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 **Bush** spokesman"
- Maximize coreference recall
- Improves over NPs as mentions by 4%

Modeling redux

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

an increase = subj | 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(**May 25/President Bush** = subj | missing in previous
other in prev-1
occurs 3x total)

= .045

What we learn

Baseline

P(**May 25/President Bush** = subj | missing in previous
other in prev-1
occurs 3x total)

= .045

Our model

P(**May 25** = subj | missing in previous
...
NE type DATE
never corefers in MUC6)

= .001

What we learn

Baseline

P(**May 25/President Bush** = subj | missing in previous
other in prev-1
occurs 3x total)

$$= .045$$

Our model

P(**President Bush** = subj | missing in previous
...
**NE type PERSON
proper NP
corefers in MUC6
modifiers**)

$$= .133$$

Standard ordering benchmarks

Discrimination
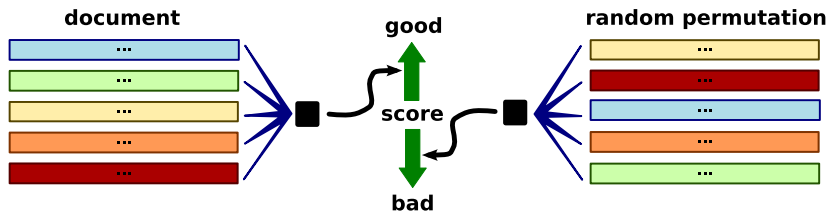
Insertion
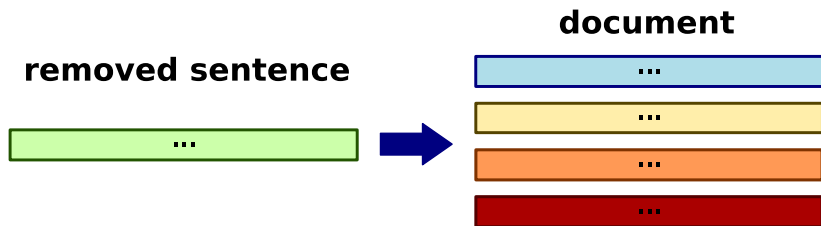
Results on WSJ test

	Disc.	Ins.
Random	50	13
(Elsner+Charniak '08)	80	23

- (Elsner+Charniak '08): previous best on corpus
  - Combined model: entity grid and coref-inspired

Results on WSJ test

	Disc.	Ins.
Random	50	13
(Elsner+Charniak '08)	80	23
Grid (all nouns mentions)	80	21

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

Results on WSJ test

	Disc.	Ins.
Random	50	13
(Elsner+Charniak '08)	80	23
Grid (all nouns mentions)	80	21
Extended Grid	**84**	**24**

- (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

	Disc.	Ins.
(Elsner+Charniak '08)	80	23
Extended Grid	84	24
ExtEGrid+combo	**86**	**27**

- 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