Identifying Anaphoric and Non-Anaphoric Noun Phrases to Improve Coreference Resolution

Vincent Ng and Claire Cardie
Department of Computer Science
Cornell University
Plan for the Talk

- Noun phrase coreference resolution
  - general machine learning approach
  - baseline coreference resolution system

- Identification of anaphoric/non-anaphoric noun phrases (Anaphoricity determination)
  - why anaphoricity info can help coreference resolution
  - general machine learning approach
  - anaphoricity determination system

- Using anaphoricity information in coreference resolution
Noun Phrase Coreference

Identify all noun phrases that refer to the same entity

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment...
Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment...
Noun Phrase Coreference

Identify all noun phrases that refer to the same entity

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment...
Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment...
Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment...
A Machine Learning Approach

§ Classification

– given a description of two noun phrases, $NP_i$ and $NP_j$, classify the pair as coreferent or not coreferent

Aone & Bennett [1995]; Connolly et al. [1994]; McCarthy & Lehnert [1995]; Soon, Ng & Lim [2001]

[Queen Elizabeth] set about transforming [her] [husband], ...

not coref?
A Machine Learning Approach

§ Clustering

- coordinates pairwise coreference decisions

Queen Elizabeth

[Queen Elizabeth], set about transforming
[her]
[husband]
...

King George VI

husband

the King

his

Logue

a renowned speech therapist

Clustering Algorithm

not coref

not coref

coref

coref
Machine Learning Issues

- Training data creation
- Instance representation
- Learning algorithm
- Clustering algorithm

[Ng and Cardie, ACL’02]
Baseline System: Training Data Creation

§ Creating training instances

– texts annotated with coreference information

– one instance for each pair of noun phrases
  » feature vector: describes the two NPs and context
  » class value:
    coref pairs on the same coreference chain
    not coref otherwise

– use sampling to deal with skewed class distributions
## Baseline System: Instance Representation

### 53 features per instance

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical (9)</td>
<td>NP string matching operations</td>
</tr>
<tr>
<td>Semantic (6)</td>
<td>Semantic compatibility tests, aliasing</td>
</tr>
<tr>
<td>Positional (2)</td>
<td>Distance in terms of number of sentences/paragraphs</td>
</tr>
<tr>
<td>Knowledge-based (2)</td>
<td>Naïve pronoun resolution, rule-based coref resolution</td>
</tr>
<tr>
<td>Grammatical (34)</td>
<td>NP type, Grammatical role, Linguistic constraints, Linguistic preferences, Heuristics</td>
</tr>
</tbody>
</table>
Baseline System: Learning Algorithm

- C4.5 (Quinlan, 1993): decision tree induction

- Classifier outputs coreference likelihood
Baseline System: Clustering Algorithm

- Best-first single-link clustering algorithm
  - selects as antecedent the NP with the highest coreference likelihood from among preceding coreferent NPs for each noun phrase
Baseline System: Evaluation

- MUC-6 and MUC-7 coreference data sets
- Documents annotated w.r.t. coreference
- MUC-6: 30 training texts + 30 test texts
- MUC-7: 30 training texts + 20 test texts
- MUC scoring program
  - recall, precision, F-measure
## Baseline System: Results

<table>
<thead>
<tr>
<th></th>
<th>MUC-6</th>
<th>MUC-7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>P</td>
</tr>
<tr>
<td>Baseline</td>
<td>70.3</td>
<td>58.3</td>
</tr>
<tr>
<td>Best MUC System</td>
<td>59</td>
<td>72</td>
</tr>
<tr>
<td>Worst MUC System</td>
<td>36</td>
<td>44</td>
</tr>
</tbody>
</table>
## Baseline System: Results

<table>
<thead>
<tr>
<th></th>
<th>MUC-6</th>
<th></th>
<th>MUC-7</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>P</td>
<td>F</td>
<td>R</td>
</tr>
<tr>
<td>Baseline</td>
<td>70.3</td>
<td>58.3</td>
<td><strong>63.8</strong></td>
<td>65.5</td>
</tr>
<tr>
<td>Best MUC System</td>
<td>59</td>
<td>72</td>
<td><strong>65</strong></td>
<td>56.1</td>
</tr>
<tr>
<td>Worst MUC System</td>
<td>36</td>
<td>44</td>
<td><strong>40</strong></td>
<td>52.5</td>
</tr>
</tbody>
</table>
Plan for the Talk

§ Noun phrase coreference resolution
  – general machine learning approach
  – baseline coreference resolution system

§ Identification of anaphoric/non-anaphoric noun phrases (Anaphoricity determination)
  – why anaphoricity info can help coreference resolution
  – general machine learning approach
  – anaphoricity determination system

§ Using anaphoricity information in coreference resolution
Motivation

- Baseline coreference system
  - single-link clustering algorithm attempts to find an antecedent for *each* noun phrase
Motivation

§ Baseline coreference system
   – single-link clustering algorithm attempts to find an antecedent for each noun phrase

§ What we really want
   – single-link clustering algorithm attempts to find an antecedent for each anaphoric noun phrase
Motivation

- Baseline coreference system
  - single-link clustering algorithm attempts to find an antecedent for each noun phrase

- What we really want
  - single-link clustering algorithm attempts to find an antecedent for each anaphoric noun phrase

- Availability of anaphoricity info can increase the precision of the coreference system
Anaphoricity Determination

For each noun phrase in a text, determine whether it is part of a coreference chain but is not the head of the chain.

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment...
Anaphoricity Determination

For each noun phrase in a text, determine whether it is part of a coreference chain but is not the head of the chain.

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment...
Anaphoricity Determination

For each noun phrase in a text, determine whether it is part of a coreference chain but is not the head of the chain.

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment...
Anaphoricity Determination

For each noun phrase in a text, determine whether it is part of a coreference chain but is not the head of the chain.

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment...
Anaphoricity Determination

For each noun phrase in a text, determine whether it is part of a coreference chain but is not the head of the chain.

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment...
Anaphoricity Determination

For each noun phrase in a text, determine whether it is part of a coreference chain but is not the head of the chain.

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment...
Anaphoricity Determination

For each noun phrase in a text, determine whether it is part of a coreference chain but is not the head of the chain.

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment...
A Machine Learning Approach

Classification

- given a description of a noun phrases, $NP_i$, classify $NP_i$ as anaphoric or not anaphoric

[Queen Elizabeth] set about transforming [her] [husband], ...
Anaphoricity Determination System

§ Training data creation
   – texts annotated with coreference information
   – one instance for each noun phrase

§ Learning algorithm
   – C4.5
## Instance representation

- 37 features per instance

<table>
<thead>
<tr>
<th>Category</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical (4)</td>
<td>case, string matching, head matching</td>
</tr>
<tr>
<td>Positional (3)</td>
<td>header, first sentence, first paragraph</td>
</tr>
<tr>
<td>Semantic (4)</td>
<td>title, aliasing, semantic compatibility</td>
</tr>
<tr>
<td>Grammatical (35)</td>
<td>NP type: definite, indefinite, bare plural&lt;br&gt;NP property: pre-modified, post-modified, number&lt;br&gt;Syntactic pattern: THE_N, THE_PN, THE_ADJ_N</td>
</tr>
</tbody>
</table>
Anaphoricity Determination System: Evaluation

- MUC-6 and MUC-7 coreference data sets

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Instances</th>
<th>% Negatives</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUC-6 test</td>
<td>4565</td>
<td>66.3</td>
<td>86.1</td>
</tr>
<tr>
<td>MUC-7 test</td>
<td>3558</td>
<td>73.2</td>
<td>84.0</td>
</tr>
</tbody>
</table>
Existing Approaches to Anaphoricity Determination

- Heuristic-based approaches

- Machine learning approaches
  - Unsupervised: Bean and Riloff (1999)
  - Supervised: Evans (2001)
Comparison with Previous Work (I)

Approaches to anaphoricity determination

<p>| Our Approach | Previous Approaches |</p>
<table>
<thead>
<tr>
<th>Our Approach</th>
<th>Previous Approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>focuses on common nouns</td>
<td></td>
</tr>
</tbody>
</table>

Comparison with Previous Work (I)

- Approaches to anaphoricity determination

- focuses on common nouns
Comparison with Previous Work (I)

Approaches to anaphoricity determination

<table>
<thead>
<tr>
<th>Our Approach</th>
<th>Previous Approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>§ focuses on common nouns</td>
<td></td>
</tr>
<tr>
<td>§ can operate on all types of noun phrases</td>
<td></td>
</tr>
</tbody>
</table>
## Comparison with Previous Work (I)

<table>
<thead>
<tr>
<th>Our Approach</th>
<th>Previous Approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>focuses on common nouns</td>
<td>handle specific types of noun phrases only</td>
</tr>
<tr>
<td>can operate on all types of noun phrases</td>
<td></td>
</tr>
</tbody>
</table>

Approaches to anaphoricity determination.
Existing anaphoricity determination algorithms address only specific types of NPs:

- pleonastic pronouns

- definite descriptions

- anaphoric and non-anaphoric uses of *it*
  - Evans (2001)
Comparison with Previous Work (II)

§ Using anaphoricity information in coreference resolution

| Our Coref System | Previous Coref Systems |
Comparison with Previous Work (II)

§ Using anaphoricity information in coreference resolution

<table>
<thead>
<tr>
<th>Our Coref System</th>
<th>Previous Coref Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>§ employs anaphoricity determination as a separate component</td>
<td></td>
</tr>
</tbody>
</table>
Using anaphoricity information in coreference resolution

<table>
<thead>
<tr>
<th>Our Coref System</th>
<th>Previous Coref Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>employs anaphoricity determination as a separate component</td>
<td>perform anaphoricity determination within the coreference system</td>
</tr>
</tbody>
</table>
Comparison with Previous Work (II)

Most previous work performs anaphoricity determination *implicitly*
- e.g. via a specific feature in the coreference system

- One exception:
  » Harabagiu *et al.* (2001)
  » assumes perfect anaphoricity information
  » effectively employs a separate (manual) anaphoricity determination component
Comparison with Previous Work (III)

|$\quad$ Evaluation of anaphoricity determination system

| Our System | Previous Systems |
Comparison with Previous Work (III)

- Evaluation of anaphoricity determination system

<table>
<thead>
<tr>
<th>Our System</th>
<th>Previous Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>§ evaluated as a standalone component</td>
<td></td>
</tr>
</tbody>
</table>
### Comparison with Previous Work (III)

<table>
<thead>
<tr>
<th>Our System</th>
<th>Previous Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>§ evaluated as a standalone component</td>
<td>§ evaluated in the context of coreference resolution</td>
</tr>
</tbody>
</table>

| § Evaluation of anaphoricity determination system |
Comparison with Previous Work (III)

- Evaluation of anaphoricity determination system

<table>
<thead>
<tr>
<th>Our System</th>
<th>Previous Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>§ evaluated as a standalone component</td>
<td>§ evaluated as a standalone component</td>
</tr>
<tr>
<td>§ evaluated in the context of coreference resolution</td>
<td></td>
</tr>
</tbody>
</table>
Comparison with Previous Work (III)

- Evaluation of anaphoricity determination system
  - Our System
    - evaluated as a standalone component
    - evaluated in the context of coreference resolution
  - Previous Systems
    - evaluated as a standalone component
    - contribution to coreference resolution not evaluated
Comparison with Previous Work (III)

Little previous work evaluates the effects of anaphoricity determination in anaphora/coreference resolution.

<table>
<thead>
<tr>
<th>Anaphoricity Determination System</th>
<th>Effects on Coref Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bean and Riloff (1999)</td>
<td>?</td>
</tr>
<tr>
<td>Denber (1998)</td>
<td>?</td>
</tr>
<tr>
<td>Kennedy and Boguraev (1996)</td>
<td>?</td>
</tr>
<tr>
<td>Lappin and Leass (1994)</td>
<td>?</td>
</tr>
<tr>
<td>Paice and Husk (1987)</td>
<td>?</td>
</tr>
<tr>
<td>Vieira and Poesio (2000)</td>
<td>➤</td>
</tr>
</tbody>
</table>
Plan for the Talk

§ Noun phrase coreference resolution
  – general machine learning approach
  – baseline coreference resolution system

§ Identification of anaphoric/non-anaphoric noun phrases (Anaphoricity determination)
  – why anaphoricity info can help coreference resolution
  – general machine learning approach
  – anaphoricity determination system

§ Using anaphoricity information in coreference resolution
How can anaphoricity information be used?

$\diamond$ The clustering algorithm will only search for an antecedent for anaphoric noun phrases.

$\diamond$ Hypothesis
- Anaphoricity information will improve precision
Anaphoricity Determination for Coref Resolution

<table>
<thead>
<tr>
<th></th>
<th>MUC-6</th>
<th></th>
<th>MUC-7</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>P</td>
<td>F</td>
<td>R</td>
</tr>
<tr>
<td>Baseline</td>
<td>70.3</td>
<td>58.3</td>
<td>63.8</td>
<td>65.5</td>
</tr>
</tbody>
</table>

coreference system has fairly low precision
Results (Perfect Anaphoricity Information)

<table>
<thead>
<tr>
<th></th>
<th>MUC-6</th>
<th></th>
<th>MUC-7</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>P</td>
<td>F</td>
<td>R</td>
</tr>
<tr>
<td>Baseline</td>
<td>70.3</td>
<td><strong>58.3</strong></td>
<td>63.8</td>
<td>65.5</td>
</tr>
<tr>
<td>With perfect anaphoricity info</td>
<td>66.3</td>
<td><strong>81.4</strong></td>
<td>73.1</td>
<td>61.5</td>
</tr>
</tbody>
</table>

perfect anaphoricity information can improve precision
## Results (Learned Anaphoricity Information)

<table>
<thead>
<tr>
<th></th>
<th>MUC-6</th>
<th></th>
<th>MUC-7</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>P</td>
<td>F</td>
<td>R</td>
</tr>
<tr>
<td><strong>Baseline</strong></td>
<td>70.3</td>
<td><strong>58.3</strong></td>
<td>63.8</td>
<td>65.5</td>
</tr>
<tr>
<td><strong>With learned anaphoricity info</strong></td>
<td>57.4</td>
<td><strong>71.6</strong></td>
<td>63.7</td>
<td>47.0</td>
</tr>
</tbody>
</table>

An improvement in precision comes at the expense of significant loss in recall.
What went wrong?

Hypothesis 1

- drop in recall and overall performance is caused by poor accuracy of anaphoricity classifier on *positive* instances
What went wrong?

§ Hypothesis 1
- drop in recall and overall performance is caused by poor accuracy of anaphoricity classifier on positive instances

§ Accuracy of anaphoricity classifier
- overall: 86.1% (MUC-6) and 84.0% (MUC-7)
- positives only: 73.1% (MUC-6) and 66.2% (MUC-7)

§ Anaphoricity classifier misclassifies 414 and 322 anaphoric entities as non-anaphoric for the MUC-6 and MUC-7 data sets, respectively
Need more accuracy?

§ Hypothesis 1.1

- accuracy levels of 66-73% on positive instances for anaphoricity determination are not adequate for improving coreference resolution
Need more accuracy?

§ Hypothesis 1.1
- accuracy levels of 66-73% on positive instances for anaphoricity determination are not adequate for improving coreference resolution

§ Goal
- improve the accuracy on positive instances
Need more accuracy?

- Hypothesis 1.1
  - accuracy levels of 66-73% on positive instances for anaphoricity determination are **not** adequate for improving coreference resolution

- Goal
  - improve the accuracy on positive instances

- How?
Improving Accuracy on Positive Instances

Observations
- **string matching** and **aliasing** are strong indicators of coreference
Improving Accuracy on Positive Instances

§ Observations

– **string matching** and **aliasing** are *strong* indicators of coreference

– **string matching** and **aliasing** are *weaker* indicators of anaphoricity
Improving Accuracy on Positive Instances

§ Observations
- **string matching** and **aliasing** are *strong* indicators of coreference
- **string matching** and **aliasing** are *weaker* indicators of anaphoricity

§ Goal
- ensure that anaphoric NPs involved in these two types of relations are correctly classified
Assume that an NP is anaphoric (and bypass the anaphoricity classifier) if anaphoricity is indicated by either the **string matching** or the **aliasing** constraint.

Accuracy on positive instances
- no constraints: 73.1% (MUC-6) and 66.2% (MUC-7)
- with constraints: 82.0% (MUC-6) and 80.8% (MUC-7)
### Results (Classification with Constraints)

<table>
<thead>
<tr>
<th></th>
<th>MUC-6</th>
<th>MUC-7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>P</td>
</tr>
<tr>
<td>Baseline</td>
<td>70.3</td>
<td>58.3</td>
</tr>
<tr>
<td>With anaphoricity (no constraints)</td>
<td>57.4</td>
<td>71.6</td>
</tr>
<tr>
<td>With anaphoricity (with constraints)</td>
<td>63.4</td>
<td>68.3</td>
</tr>
</tbody>
</table>

- Large gains in precision and smaller drops in recall.
- Automatically acquired anaphoricity info can be used to improve the performance of coreference resolution.
## Results (Comparison with Best MUC Systems)

<table>
<thead>
<tr>
<th></th>
<th>MUC-6</th>
<th></th>
<th></th>
<th>MUC-7</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>P</td>
<td>F</td>
<td>R</td>
<td>P</td>
<td>F</td>
</tr>
<tr>
<td>With anaphoricity (with constraints)</td>
<td>63.4</td>
<td>68.3</td>
<td><strong>65.8</strong></td>
<td>59.7</td>
<td>69.3</td>
<td><strong>64.2</strong></td>
</tr>
<tr>
<td>Best MUC System</td>
<td>59</td>
<td>72</td>
<td><strong>65</strong></td>
<td>56.1</td>
<td>68.8</td>
<td><strong>61.8</strong></td>
</tr>
</tbody>
</table>
Results (Comparison with Perfect Anaphoricity)

<table>
<thead>
<tr>
<th></th>
<th>MUC-6</th>
<th>MUC-7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>P</td>
</tr>
<tr>
<td>With anaphoricity (with constraints)</td>
<td>63.4</td>
<td>68.3</td>
</tr>
<tr>
<td>With perfect anaphoricity info</td>
<td>66.3</td>
<td>81.4</td>
</tr>
</tbody>
</table>

§ substantial room for improvement in anaphoricity determination
Summary

- Presented a supervised learning approach for anaphoricity determination that can handle all types of NPs
- Investigated the use of anaphoricity information in coreference resolution
- Showed automatically acquired knowledge of anaphoricity can be used to improve the performance of a learning-based coreference system