Chinese Overt Pronoun Resolution: A Bilingual Approach

Chen Chen and Vincent Ng
Human Language Technology Research Institute
University of Texas at Dallas
Pronoun Resolution

- Find an **antecedent** for each anaphoric pronoun
  - a preceding mention in the text the pronoun refers to
Pronoun Resolution

- Find an antecedent for each anaphoric pronoun

  Mary told John that she liked him a lot.
Pronoun Resolution

- Find an **antecedent** for each anaphoric pronoun

  Mary told John that she liked him a lot.
Pronoun Resolution

- Find an antecedent for each anaphoric pronoun

  Mary told John that she liked him a lot.

  玛丽告诉约翰她非常喜欢他。
Pronoun Resolution

- Find an antecedent for each anaphoric pronoun

Mary told John that she liked him a lot.

玛丽告诉约翰她非常喜欢他。
Pronoun Resolution

- Find an **antecedent** for each anaphoric pronoun

  Mary told John that **she** liked **him** a lot.

  玛丽告诉约翰她非常喜欢他。

  Overt pronouns
Pronoun Resolution

- Find an **antecedent** for each anaphoric pronoun

  Mary told John that she liked him a lot.

  玛丽告诉约翰她非常喜欢他。

The English pronoun resolution task is the same as the overt Chinese pronoun resolution task.
Pronoun Resolution

- Find an antecedent for each anaphoric pronoun

Mary told John that she liked him a lot.

玛丽告诉约翰她非常喜欢他。

Can Chinese pronoun resolution be tackled in the same way as English pronoun resolution?
Pronoun Resolution

- Find an antecedent for each anaphoric pronoun

  Mary told John that she liked him a lot.

Can Chinese pronoun resolution be tackled in the same way as English pronoun resolution?

Can we train a resolver on Chinese texts and use it to resolve Chinese pronouns?
Pronoun Resolution

- Find an **antecedent** for each anaphoric pronoun

  Mary told John that **she** liked **him** a lot.

  玛丽告诉约翰**她**非常喜欢**他**。

Can Chinese pronoun resolution be tackled in the same way as English pronoun resolution?

Can we train a resolver on Chinese texts and use it to resolve Chinese pronouns?

Yes, but …
Pronoun Resolution

- Find an **antecedent** for each anaphoric pronoun
  
  Mary told John that she liked him a lot.

Can Chinese pronoun resolution be tackled in the same way as English pronoun resolution?

Can we train a resolver on Chinese texts and use it to resolve Chinese pronouns?

Yes, but … it may not work as well for Chinese
Why?
Why?

- Less coreference-annotated data available in Chinese than in English for training resolvers
Why?

- Less coreference-annotated data available in Chinese than in English for training resolvers
- Lack of publicly-available Chinese resources essential for pronoun resolution, such as Gender and Number wordlists.
Why?

- Less coreference-annotated data available in Chinese than in English for training resolvers.
- Lack of publicly-available Chinese resources essential for pronoun resolution, such as Gender and Number wordlists.

**Goal**: improve Chinese pronoun resolution by addressing these two issues.
Why?

- Less coreference-annotated data available in Chinese than in English for training resolvers

- Lack of publicly-available Chinese resources essential for pronoun resolution, such as Gender and Number wordlists.

**Goal**: improve Chinese pronoun resolution by addressing these two issues
Why?

- Less coreference-annotated data available in Chinese than in English for training resolvers

- Lack of publicly-available Chinese resources essential for pronoun resolution, such as Gender and Number wordlists.

**Goal**: improve Chinese pronoun resolution by addressing these two issues

Exploit English coreference data
Why?

- Less coreference-annotated data available in Chinese than in English for training resolvers
- Lack of publicly-available Chinese resources essential for pronoun resolution, such as Gender and Number wordlists.

**Goal:** improve Chinese pronoun resolution by addressing these two issues
Why?

- Less coreference-annotated data available in Chinese than in English for training resolvers
- Lack of publicly-available Chinese resources essential for pronoun resolution, such as Gender and Number wordlists.

**Goal**: improve Chinese pronoun resolution by addressing these two issues
Why?

- Less coreference-annotated data available in Chinese than in English for training resolvers.
- Lack of publicly-available Chinese resources essential for pronoun resolution, such as Gender and Number wordlists.

**Goal**: improve Chinese pronoun resolution by addressing these two issues.
Why?

- Less coreference-annotated data available in Chinese than in English for training resolvers
- Lack of publicly-available Chinese resources essential for pronoun resolution, such as Gender and Number wordlists.

**Goal**: improve Chinese pronoun resolution by addressing these two issues

**A Bilingual Approach**

- Exploit English coreference data
- Exploit English Gender and Number wordlists
- Exploit Chinese coreference data
Bilingual Approach

- **Training**
  - train the pronoun resolution models

- **Testing**
  - resolve Chinese pronouns using the models
Bilingual Approach

- **Training**
  - train the pronoun resolution models

- **Testing**
  - resolve Chinese pronouns using the models
Model Training

- Train 3 maximum-entropy-based pronoun resolution models
  - The Chinese model
  - The English model
  - The mixed model

- Each model returns the probability that a pronoun and a candidate antecedent are coreferent
The Chinese Model (CM)

- Trained on the Chinese training data
  - Each training instance corresponds to a Chinese anaphoric pronoun and one of its candidate antecedents
    - Class value is 1 if they are coreferent; and 0 otherwise
    - Represented using features designed for Chinese
The Chinese Model (CM)

- Trained on the Chinese training data
  - Each training instance corresponds to a Chinese anaphoric pronoun and one of its candidate antecedents
    - Class value is 1 if they are coreferent; and 0 otherwise
    - Represented using features designed for Chinese

- Pros
  - exploits Chinese training data

- Cons
  - exploits neither English training data nor English wordlists
The English Model (EM)

- Trained on the English training data
  - Each training instance corresponds to an English anaphoric pronoun and one of its candidate antecedents
    - Class value is 1 if they are coreferent; and 0 otherwise
    - Represented using features designed for English
The English Model (EM)

- Trained on the English training data
  - Each training instance corresponds to an English anaphoric pronoun and one of its candidate antecedents
    - Class value is 1 if they are coreferent; and 0 otherwise
    - Represented using features designed for English

- How to use it to resolve Chinese pronouns in the test set?
The English Model (EM)

- Trained on the English training data
  - Each training instance corresponds to an English anaphoric pronoun and one of its candidate antecedents
    - Class value is 1 if they are coreferent; and 0 otherwise
    - Represented using features designed for English

- How to use it to resolve Chinese pronouns in the test set?
  Annotation projection
The English Model (EM)

- Trained on the English training data
  - Each training instance corresponds to an English anaphoric pronoun and one of its candidate antecedents
    - Class value is 1 if they are coreferent; and 0 otherwise
    - Represented using features designed for English

- How to use it to resolve Chinese pronouns in the test set?
  玛丽告诉约翰她非常喜欢他。
The English Model (EM)

- Trained on the English training data
  - Each training instance corresponds to an English anaphoric pronoun and one of its candidate antecedents
    - Class value is 1 if they are coreferent; and 0 otherwise
    - Represented using features designed for English

- How to use it to resolve Chinese pronouns in the test set?
  玛丽告诉约翰她非常喜欢他。

  - **Step 1:** Machine-translate Chinese text into English
The English Model (EM)

- Trained on the English training data
  - Each training instance corresponds to an English anaphoric pronoun and one of its candidate antecedents
    - Class value is 1 if they are coreferent; and 0 otherwise
    - Represented using features designed for English

- How to use it to resolve Chinese pronouns in the test set?
  玛丽告诉约翰她非常喜欢他。
  Mary told John that she liked him a lot.

  **Step 1:** Machine-translate Chinese text into English
The English Model (EM)

- Trained on the English training data
  - Each training instance corresponds to an English anaphoric pronoun and one of its candidate antecedents
    - Class value is 1 if they are coreferent; and 0 otherwise
    - Represented using features designed for English

- How to use it to resolve Chinese pronouns in the test set?

  玛丽告诉约翰她非常喜欢他。

  Mary told John that she liked him a lot.

  **Step 2:** Identify and align the Chinese and English mentions
The English Model (EM)

- Trained on the English training data
  - Each training instance corresponds to an English anaphoric pronoun and one of its candidate antecedents
    - Class value is 1 if they are coreferent; and 0 otherwise
    - Represented using features designed for English

- How to use it to resolve Chinese pronouns in the test set?
  
  [玛丽]告诉[约翰][她]非常喜欢[他]。
  

- Step 2: Identify and align the Chinese and English mentions
The English Model (EM)

- Trained on the English training data
  - Each training instance corresponds to an English anaphoric pronoun and one of its candidate antecedents
    - Class value is 1 if they are coreferent; and 0 otherwise
    - Represented using features designed for English

- How to use it to resolve Chinese pronouns in the test set?

  [玛丽]告诉[约翰][她]非常喜欢[他]。


- **Step 3**: Use EM to resolve English pronouns
The English Model (EM)

- Trained on the English training data
  - Each training instance corresponds to an English anaphoric pronoun and one of its candidate antecedents
    - Class value is 1 if they are coreferent; and 0 otherwise
    - Represented using features designed for English

- How to use it to resolve Chinese pronouns in the test set?
  
  [玛丽]告诉[约翰][她]非常喜欢[他]。


- Step 3: Use EM to resolve English pronouns
The English Model (EM)

- Trained on the English training data
  - Each training instance corresponds to an English anaphoric pronoun and one of its candidate antecedents
    - Class value is 1 if they are coreferent; and 0 otherwise
    - Represented using features designed for English

- How to use it to resolve Chinese pronouns in the test set?
  
  [玛丽]告诉[约翰][她]非常喜欢[他]。


- Step 4: Project annotations from English to Chinese
The English Model (EM)

- Trained on the English training data
  - Each training instance corresponds to an English anaphoric pronoun and one of its candidate antecedents
    - Class value is 1 if they are coreferent; and 0 otherwise
    - Represented using features designed for English

- How to use it to resolve Chinese pronouns in the test set?
  
  [玛丽]告诉[约翰][她]非常喜欢[他]。


- Step 4: Project annotations from English to Chinese
The English Model (EM)

- Trained on the English training data
  - Each training instance corresponds to an English anaphoric pronoun and one of its candidate antecedents
    - Class value is 1 if they are coreferent; and 0 otherwise
    - Represented using features designed for English
- How to use it to resolve Chinese pronouns in the test set?
  
  [玛丽]告诉[约翰][她]非常喜欢[他]。
  
- What are the pros and cons of the English model?
The English Model (EM)

- Trained on the English training data
  - Each training instance corresponds to an English anaphoric pronoun and one of its candidate antecedents
    - Class value is 1 if they are coreferent; and 0 otherwise
    - Represented using features designed for English

- How to use it to resolve Chinese pronouns in the test set?

  [玛丽]告诉[约翰][她]非常喜欢[他]。

- Pros: exploits English training data and English wordlists
- Cons: doesn’t exploit Chinese training data
The Mixed Model (MM)

- Trained on Chinese training data
  - Training instance creation
    - Translate the Chinese training data into English
    - Map Chinese mentions to English mentions in translated text
    - Create an instance between a Chinese pronoun and one of its candidate antecedents if and only if each of them can be mapped to some English mention
The Mixed Model (MM)

- Trained on Chinese training data
  - Training instance creation
    - Translate the Chinese training data into English
    - Map Chinese mentions to English mentions in translated text
    - Create an instance between a Chinese pronoun and one of its candidate antecedents if and only if each of them can be mapped to some English mention
  - Features are computed from both the Chinese pronoun and the candidate antecedent, as well as the mapped English pronoun and the mapped English candidate antecedent
The Mixed Model (MM)

- Trained on Chinese training data
  - Training instance creation
    - Translate the Chinese training data into English
    - Map Chinese mentions to English mentions in translated text
    - Create an instance between a Chinese pronoun and one of its candidate antecedents if and only if each of them can be mapped to some English mention
  - Features are computed from both the Chinese pronoun and the candidate antecedent, as well as the mapped English pronoun and the mapped English candidate antecedent

Feature-augmented model: exploits English and Chinese features
The Mixed Model (MM)

- Trained on Chinese training data
  - **Training instance creation**
    - Translate the Chinese training data into English
    - Map Chinese mentions to English mentions in translated text
    - Create an instance between a Chinese pronoun and one of its candidate antecedents if and only if each of them can be mapped to some English mention
  - **Features** are computed from both the Chinese pronoun and the candidate antecedent, as well as the mapped English pronoun and the mapped English candidate antecedent.

  **Feature-augmented model:** exploits English and Chinese features

- **Pros:** exploits Chinese training data and English wordlists
- **Cons:** doesn’t exploit English training data
Bilingual Approach

- **Training**
  - train the pronoun resolution models

- **Testing**
  - resolve Chinese pronouns using the models
Which of the 3 models should be used to resolve Chinese pronouns?
Which of the 3 models should be used to resolve Chinese pronouns?

- Since each model has its own pros and cons, we adopt an ensemble approach
  - combine their decisions when resolving a Chinese pronoun
  - 4 resolution methods
    - Differ in terms of how the decisions of the models are combined
Resolution Method 1

- Given a Chinese pronoun to be resolved,
  - Map it to the English pronoun $p$ in translated text
  - Use EM to resolve $p$ to the candidate antecedent having the highest coreference probability with $p$ among the candidates
  - Project English resolution result back to Chinese
Resolution Method 1

- Given a Chinese pronoun to be resolved,
  - Map it to the English pronoun $p$ in translated text
  - Use EM to resolve $p$ to the candidate antecedent having the highest coreference probability with $p$ among the candidates
  - Project English resolution result back to Chinese

- If the Chinese pronoun is not mapped to an English pronoun, use CM to resolve it to the most likely candidate antecedent
Resolution Method 1

- Given a Chinese pronoun to be resolved,
  - Map it to the English pronoun $p$ in translated text
  - Use EM to resolve $p$ to the candidate antecedent having the highest coreference probability with $p$ among the candidates
  - Project English resolution result back to Chinese

- If the Chinese pronoun is not mapped to an English pronoun, use CM to resolve it to the most likely candidate antecedent

Use EM for resolution and CM as a backoff model
Resolution Method 2

- Same as resolution method 1 except that EM is replaced with MM
Resolution Method 2

- Same as resolution method 1 except that EM is replaced with MM

Use MM for resolution and CM as a backoff model
Resolution Method 2

- Same as resolution method 1 except that EM is replaced with MM

Use MM for resolution and CM as a backoff model

**Hypothesis**: Method 2 would be better than Method 1, since MM is a feature-augmented model representing an instance using both English and Chinese features
Resolution Method 3

- Same as the previous two resolution methods except that the coreference probability between a pronoun and a candidate antecedent is given by the unweighted average of the probabilities returned by CM, EM, and MM.
Resolution Method 3

- Same as the previous two resolution methods except that the coreference probability between a pronoun and a candidate antecedent is given by the unweighted average of the probabilities returned by CM, EM, and MM

Combine the decisions of all 3 models for resolution and use CM as a backoff model
Resolution Method 3

- Same as the previous two resolution methods except that the coreference probability between a pronoun and a candidate antecedent is given by the unweighted average of the probabilities returned by CM, EM, and MM.

Combine the decisions of all 3 models for resolution and use CM as a backoff model.

**Hypothesis**: Method 3 would be better than Method 2 because it uses three rather than two models.
Resolution Method 4

- Same as resolution method 3 except that we do weighted averaging of coreference probabilities of the three models
- Weights are determined using held-out development data
Resolution Method 4

- Same as resolution method 3 except that we do weighted averaging of coreference probabilities of the three models
  - Weights are determined using held-out development data

Combine the decisions of all 3 models for resolution in a weighted manner and use CM as a backoff model
Resolution Method 4

- Same as resolution method 3 except that we do **weighted averaging** of coreference probabilities of the three models
  - Weights are determined using held-out development data

Combine the decisions of all 3 models for resolution in a weighted manner and use CM as a backoff model

**Hypothesis**: Method 4 would be better than Method 3 because weighted averaging might be better than unweighted averaging
Evaluation

- **Goal**: evaluate our bilingual approach
  - The 4 resolution methods
Experimental Setup

- **Corpus**
  - Coreference data used in the CoNLL 2012 shared task
- **Training**
  - 1391 Chinese documents (750K words)
  - 1940 English documents (1.3M words)
- **Development**
  - 172 Chinese documents (110K words)
- **Test**
  - 166 Chinese documents (90K words)
Experimental Setup

- **Corpus**
  - Coreference data used in the CoNLL 2012 shared task
- **Training**
  - 1391 Chinese documents (750K words)
  - 1940 English documents (1.3M words)
- **Development**
  - 172 Chinese documents (110K words)
- **Test**
  - 166 Chinese documents (90K words)

- **Evaluation measures**
  - recall (R), precision (P), and F-measure (F) on resolving anaphoric pronouns
Three Baseline Systems

- **Monolingual approach**
  - Uses the Chinese model to resolve pronouns

- **Best Chinese resolver in CoNLL-2012 shared task**
  - Chen & Ng (2012): combines rules and machine learning

- **Rahman & Ng (2012)**
  - Annotation projection approach
    - Uses the English model to resolve pronouns in translated text
    - Same as resolution method 1 except that there is no backoff
## Results: Baseline Systems

<table>
<thead>
<tr>
<th>Method</th>
<th>R</th>
<th>P</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monolingual</td>
<td>71.7</td>
<td>65.3</td>
<td>68.4</td>
</tr>
<tr>
<td>Best shared task system</td>
<td>63.8</td>
<td>67.5</td>
<td>65.6</td>
</tr>
<tr>
<td>Rahman and Ng’s (2012) approach</td>
<td>64.3</td>
<td>65.2</td>
<td>64.7</td>
</tr>
</tbody>
</table>

---

- **Method 1**: (EM, but use CM as backoff)
- **Method 2**: (MM, but use CM as backoff)
- **Method 3**: (unweighted averaging of 3 models)
- **Method 4**: (weighted averaging of 3 models)
Results: Baseline Systems

<table>
<thead>
<tr>
<th>Method</th>
<th>R</th>
<th>P</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monolingual</td>
<td>71.7</td>
<td>65.3</td>
<td>68.4</td>
</tr>
<tr>
<td>Best shared task system</td>
<td>63.8</td>
<td>67.5</td>
<td>65.6</td>
</tr>
<tr>
<td>Rahman and Ng’s (2012) approach</td>
<td>64.3</td>
<td>65.2</td>
<td>64.7</td>
</tr>
</tbody>
</table>

- Best baseline: Monolingual Baseline
  - owing to its considerably higher recall
Results: Our Resolution Methods

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>P</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monolingual</td>
<td>71.7</td>
<td>65.3</td>
<td>68.4</td>
</tr>
<tr>
<td>Best shared task system</td>
<td>63.8</td>
<td>67.5</td>
<td>65.6</td>
</tr>
<tr>
<td>Rahman and Ng’s (2012) approach</td>
<td>64.3</td>
<td>65.2</td>
<td>64.7</td>
</tr>
<tr>
<td>Method 1 (EM, but use CM as backoff)</td>
<td>65.6</td>
<td>64.4</td>
<td>65.0</td>
</tr>
<tr>
<td>Method 2 (MM, but use CM as backoff)</td>
<td>73.0</td>
<td>65.1</td>
<td>68.8</td>
</tr>
<tr>
<td>Method 3 (unweighted averaging of 3 models)</td>
<td>71.5</td>
<td>70.5</td>
<td>71.0</td>
</tr>
<tr>
<td>Method 4 (weighted averaging of 3 models)</td>
<td>71.1</td>
<td>71.5</td>
<td>71.3</td>
</tr>
</tbody>
</table>

- Method 4 > Method 3 > Method 2 > Method 1
  - Method 4 outperforms the best baseline by 2.9% in F-score
  - Our bilingual approach improves Chinese pronoun resolution
Summary

- Presented a bilingual approach to Chinese overt pronoun resolution that exploits not only Chinese coreference data but also English coreference data and English wordlists.