Discourse, Pragmatics, Coreference Resolution

Many slides are adapted from Roger Levy, Chris Manning, Vicent Ng, Heeyoung Lee, Altaf Rahman
A pragmatic issue

• Just how are pronouns and nominals interpreted (resolved) in a discourse?

(1) Jane likes Mary.
(2) She often brings her flowers.
(3) She chats with the young woman for ages
What is Coreference Resolution?

– Identify all noun phrases (mentions) that refer to the same real world entity

Barack Obama nominated Hillary Rodham Clinton as his secretary of state on Monday. He chose her because she had foreign affairs experience as a former First Lady.
What is Coreference Resolution?

– Identify all noun phrases (mentions) that refer to the same real world entity

Barack Obama nominated Hillary Rodham Clinton as his secretary of state on Monday. He chose her because she had foreign affairs experience as a former First Lady.
What is Coreference Resolution?

– Identify all noun phrases (mentions) that refer to the same real world entity

Barack Obama nominated Hillary Rodham Clinton as his secretary of state on Monday. He chose her because she had foreign affairs experience.
What is Coreference Resolution?

– Identify all noun phrases (mentions) that refer to the same real world entity

Barack Obama nominated Hillary Rodham Clinton as his secretary of state on Monday. He chose her because she had foreign affairs experience as a former First Lady.
What is Coreference Resolution?

– Identify all noun phrases (mentions) that refer to the same real world entity

Barack Obama nominated Hillary Rodham Clinton as his secretary of state on Monday. He chose her because she had foreign affairs experience as a former First Lady.
Reference Resolution

• Noun phrases refer to entities in the world, many pairs of noun phrases co-refer, some nested inside others

John Smith, CFO of Prime Corp. since 1986, saw his pay jump 20% to $1.3 million as the 57-year-old also became the financial services co.'s president.
Kinds of Reference

• Referring expressions
  – John Smith
  – President Smith
  – the president
  – the company’s new executive

• Free variables
  – Smith saw his pay increase

• Bound variables
  – The dancer hurt herself.

More common in newswire, generally harder in practice

More interesting grammatical constraints, more linguistic theory, easier in practice

“anaphora resolution”
Not all NPs are referring!

- *Every dancer* twisted *her knee*.
- *(No dancer twisted her knee.)*

- There are three NPs in each of these sentences; because the first one is non-referential, the other two aren’t either.
Supervised Machine Learning
Pronominal Anaphora Resolution

• Given a pronoun and an entity mentioned earlier, classify whether the pronoun refers to that entity or not given the surrounding context (yes/no)

• Usually first filter out pleonastic pronouns like “It is raining.” (perhaps using hand-written rules)

• Use any classifier, obtain positive examples from training data, generate negative examples by pairing each pronoun with other (incorrect) entities

• This is naturally thought of as a binary classification (or ranking) task
Features for Pronominal Anaphora Resolution

• Constraints:
  – Number agreement
    • Singular pronouns (it/he/she/his/her/him) refer to singular entities and plural pronouns (we/they/us/them) refer to plural entities
  – Person agreement
    • He/she/they etc. must refer to a third person entity
  – Gender agreement
    • He → John; she → Mary; it → car
    • Jack gave Mary a gift. She was excited.
  – Certain syntactic constraints
    • John bought himself a new car. [himself → John]
    • John bought him a new car. [him can not be John]
Features for Pronominal Anaphora Resolution

• Preferences:
  – Recency: More recently mentioned entities are more likely to be referred to
    • John went to a movie. Jack went as well. He was not busy.
  – Grammatical Role: Entities in the subject position is more likely to be referred to than entities in the object position
    • John went to a movie with Jack. He was not busy.
  – Parallelism:
    • John went with Jack to a movie. Joe went with him to a bar.
Features for Pronominal Anaphora Resolution

• Preferences:
  – Verb Semantics: Certain verbs seem to bias whether the subsequent pronouns should be referring to their subjects or objects
    • *John* telephoned *Bill*. *He* lost the laptop.
    • *John* criticized *Bill*. *He* lost the laptop.
  – Selectional Restrictions: Restrictions because of semantics
    • *John* parked his *car* in the *garage* after driving *it* around for hours.

• Encode all these and maybe more as features
## Pairwise Features

<table>
<thead>
<tr>
<th>Category</th>
<th>Features</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical</td>
<td>exact_strm</td>
<td>1 if two mentions have the same spelling; 0 otherwise</td>
</tr>
<tr>
<td></td>
<td>left_subsm</td>
<td>1 if one mention is a left substring of the other; 0 otherwise</td>
</tr>
<tr>
<td></td>
<td>right_subsm</td>
<td>1 if one mention is a right substring of the other; 0 otherwise</td>
</tr>
<tr>
<td></td>
<td>acronym</td>
<td>1 if one mention is an acronym of the other; 0 otherwise</td>
</tr>
<tr>
<td></td>
<td>edit_dist</td>
<td>quantized editing distance between two mention strings</td>
</tr>
<tr>
<td></td>
<td>spell</td>
<td>pair of actual mention strings</td>
</tr>
<tr>
<td></td>
<td>ncd</td>
<td>number of different capitalized words in two mentions</td>
</tr>
<tr>
<td>Distance</td>
<td>token_dist</td>
<td>how many tokens two mentions are apart (quantized)</td>
</tr>
<tr>
<td></td>
<td>sent_dist</td>
<td>how many sentences two mentions are apart (quantized)</td>
</tr>
<tr>
<td></td>
<td>gap_dist</td>
<td>how many mentions in between the two mentions in question (quantized)</td>
</tr>
<tr>
<td>Syntax</td>
<td>POS_pair</td>
<td>POS-pair of two mention heads</td>
</tr>
<tr>
<td></td>
<td>apposition</td>
<td>1 if two mentions are appositive; 0 otherwise</td>
</tr>
<tr>
<td>Count</td>
<td>count</td>
<td>pair of (quantized) numbers, each counting how many times a mention</td>
</tr>
<tr>
<td></td>
<td></td>
<td>string is seen</td>
</tr>
<tr>
<td>Pronoun</td>
<td>gender</td>
<td>pair of attributes of {female, male, neutral, unknown}</td>
</tr>
<tr>
<td></td>
<td>number</td>
<td>pair of attributes of {singular, plural, unknown}</td>
</tr>
<tr>
<td></td>
<td>possessive</td>
<td>1 if a pronoun is possessive; 0 otherwise</td>
</tr>
<tr>
<td></td>
<td>reflexive</td>
<td>1 if a pronoun is reflexive; 0 otherwise</td>
</tr>
</tbody>
</table>

[Luo et al. 04]
Machine learning models of coref

• Start with supervised data
  • positive examples that corefer
  • negative examples that don’t corefer
  – Note that it’s very skewed
  • The vast majority of mention pairs don’t corefer
Kinds of Models

• Mention Pair models
  – Treat coreference chains as a collection of pairwise links
  – Make independent pairwise decisions and reconcile them in some way (e.g. clustering or greedy partitioning)

• Mention ranking models
  – Explicitly rank all candidate antecedents for a mention

• Entity-Mention models
  – A cleaner, but less studied, approach
  – Posit single underlying entities
  – Each mention links to a discourse entity [Pasula et al. 03], [Luo et al. 04]
Lee et al. (2010): Stanford deterministic coreference

- Cautious and incremental approach
- Multiple passes over text
- Precision of each pass is lesser than preceding ones
- Recall keeps increasing with each pass
- Decisions once made cannot be modified by later passes
- Rule-based (“unsupervised”)
Approach: start with high precision clumpings

E.g.

Pepsi hopes to take Quaker oats to a whole new level. ... Pepsi says it expects to double Quaker's snack food growth rate. ... the deal gives Pepsi access to Quaker oats' Gatorade sport drink as well as ....

Exact String Match: A high precision feature
Entity-mention model: Clusters instead of mentions
Detailed Architecture

The system consists of seven passes (or sieves):

- Exact Match
- Precise Constructs (appositives, predicate nominatives, ...)
- Strict Head Matching
- Strict Head Matching – Variant 1
- Strict Head Matching – Variant 2
- Relaxed Head Matching
- Pronouns
Cumulative performance of passes

Graph showing the system’s B³ Precision, Recall and F1 on ACE2004-DEV after each additional pass
Evaluation metrics

• MUC Score (Vilain et al., 1995)
  – Link based: Counts the number of common links and computes f-measure
• CEAF (Luo 2005); entity based
• BLANC (Recasens and Hovy 2011) Cluster RAND-index
• ...

• All of them are sort of evaluating getting coreference links/clusters right and wrong, but the differences can be important
CoNLL 2011 Shared task on coref

**Official; Closed track; Predicted mentions**

<table>
<thead>
<tr>
<th>System</th>
<th>MD</th>
<th>MUC</th>
<th>B-CUBED</th>
<th>CEAF$_m$</th>
<th>CEAF$_e$</th>
<th>BLANC</th>
<th>Official $\frac{F_1+F_2+F_3}{3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>lee</td>
<td>70.70</td>
<td>59.57</td>
<td>68.31</td>
<td>56.37</td>
<td>45.48</td>
<td>73.02</td>
<td><strong>57.79</strong></td>
</tr>
<tr>
<td>sapena</td>
<td>43.20</td>
<td>59.55</td>
<td>67.09</td>
<td>53.51</td>
<td>41.32</td>
<td>71.10</td>
<td>55.99</td>
</tr>
<tr>
<td>chang</td>
<td>64.28</td>
<td>57.15</td>
<td><strong>68.79</strong></td>
<td>54.40</td>
<td>41.94</td>
<td>73.71</td>
<td>55.96</td>
</tr>
<tr>
<td>nugues</td>
<td>68.96</td>
<td>58.61</td>
<td>65.46</td>
<td>51.45</td>
<td>39.52</td>
<td>71.11</td>
<td>54.53</td>
</tr>
<tr>
<td>santos</td>
<td>65.45</td>
<td>56.65</td>
<td>65.66</td>
<td>49.54</td>
<td>37.91</td>
<td>69.46</td>
<td>53.41</td>
</tr>
<tr>
<td>song</td>
<td>67.26</td>
<td><strong>59.95</strong></td>
<td>63.23</td>
<td>46.29</td>
<td>35.96</td>
<td>61.47</td>
<td>53.05</td>
</tr>
<tr>
<td>stoyanov</td>
<td>67.78</td>
<td>58.43</td>
<td>61.44</td>
<td>46.08</td>
<td>35.28</td>
<td>60.28</td>
<td>51.92</td>
</tr>
<tr>
<td>sobha</td>
<td>64.23</td>
<td>50.48</td>
<td>64.00</td>
<td>49.48</td>
<td>41.23</td>
<td>63.28</td>
<td>51.90</td>
</tr>
<tr>
<td>kobdani</td>
<td>61.03</td>
<td>53.49</td>
<td>65.25</td>
<td>42.70</td>
<td>33.79</td>
<td>62.61</td>
<td>51.04</td>
</tr>
<tr>
<td>zhou</td>
<td>62.31</td>
<td>48.96</td>
<td>64.07</td>
<td>47.53</td>
<td>39.74</td>
<td>64.72</td>
<td>50.92</td>
</tr>
<tr>
<td>charton</td>
<td>64.30</td>
<td>52.45</td>
<td>62.10</td>
<td>46.22</td>
<td>36.54</td>
<td>64.20</td>
<td>50.36</td>
</tr>
<tr>
<td>yang</td>
<td>63.93</td>
<td>52.31</td>
<td>62.32</td>
<td>46.55</td>
<td>35.33</td>
<td>64.63</td>
<td>49.99</td>
</tr>
<tr>
<td>hao</td>
<td>64.30</td>
<td>54.47</td>
<td>61.01</td>
<td>45.07</td>
<td>32.67</td>
<td>65.35</td>
<td>49.38</td>
</tr>
<tr>
<td>xinxin</td>
<td>61.92</td>
<td>46.62</td>
<td>61.93</td>
<td>44.75</td>
<td>36.23</td>
<td>64.27</td>
<td>48.46</td>
</tr>
<tr>
<td>zhang</td>
<td>61.13</td>
<td>47.28</td>
<td>61.14</td>
<td>44.46</td>
<td>35.19</td>
<td>65.21</td>
<td>48.07</td>
</tr>
<tr>
<td>kummerfeld</td>
<td>62.72</td>
<td>42.70</td>
<td>60.29</td>
<td>45.35</td>
<td>38.32</td>
<td>59.91</td>
<td>47.10</td>
</tr>
<tr>
<td>zhekova</td>
<td>48.29</td>
<td>24.08</td>
<td>61.46</td>
<td>40.43</td>
<td>35.75</td>
<td>53.77</td>
<td>40.43</td>
</tr>
<tr>
<td>irwin</td>
<td>26.67</td>
<td>19.98</td>
<td>50.46</td>
<td>31.68</td>
<td>25.21</td>
<td>51.12</td>
<td>31.28</td>
</tr>
</tbody>
</table>
Remarks

• This simple deterministic approach gives state of the art performance!
• Easy insertion of new features or models
• The idea of “easy first” model has also had some popularity in other (ML-based) NLP systems
  – Easy first POS tagging and parsing

• It’s a flexible architecture, not an argument that ML is wrong
  • Pronoun resolution pass would be easiest place to reinsert an ML model??