# CS 4650/7650
# Anaphora and coreference

Jacob Eisenstein

November 5, 2013

# Ripped from the headlines

Apple Inc Chief Executive Tim Cook has jetted into China for talks with government officials as he seeks to clear up a pile of problems in the firm's biggest growth market, from its contested iPad trademark to treatment of local labor. Cook is on his first trip to the country since taking over from late co-founder Steve Jobs in August, keeping to a closely guarded agenda that has included talks on Monday with Beijing's mayor and a visit to one of Apple's two stores in the capital.

# Ripped from the headlines

Apple Inc Chief Executive Tim Cook has jetted into China for talks with government officials as he seeks to clear up a pile of problems in the firm's biggest growth market, from its contested iPad trademark to treatment of local labor. Cook is on his first trip to the country since taking over from late co-founder Steve Jobs in August, keeping to a closely guarded agenda that has included talks on Monday with Beijing's mayor and a visit to one of Apple's two stores in the capital.

- What is Apple's biggest growth market?
- What is the capital of China?
- Who owns the iPad trademark?
- Where is Tim Cook going for the first time since August?

# Resolving anaphoric pronouns

Apple Inc Chief Executive Tim Cook has jetted into China for talks with government officials as **he** seeks to clear up a pile of problems in the firm's biggest growth market, from **its** contested iPad trademark to treatment of local labor. Cook is on **his** first trip to the country...

- he $\stackrel{?}{=}$ Apple Inc, Tim Cook, China, talks, government officials, government, ...
- its $\stackrel{?}{=}$ the firm's biggest growth market, the firm, problems, a pile of problems, ...
- his $\stackrel{?}{=}$ Cook, local labor, its contested iPad trademark, iPad, ...

How can we resolve these pronouns?

# Preferences

- **Recency**:
  Drew made **a nice pie**, but Liz made **an even better pie**.
  **It** had apples and bacon.

# Preferences

- **Recency**:
  Drew made **a nice pie**, but Liz made **an even better pie**.
  **It** had apples and bacon.
- **Repeated mention**

# Preferences

- **Recency**:
  Drew made **a nice pie**, but Liz made **an even better pie**.
  **It** had apples and bacon.
- **Repeated mention**
- **Grammatical role**: subj $>$ direct obj $>$ indirect obj
  **Elmo** went to the bar with **Grover**.
  **He** ordered two vodka tonics.

# Preferences

- **Recency**:
  Drew made **a nice pie**, but Liz made **an even better pie**.
  **It** had apples and bacon.
- **Repeated mention**
- **Grammatical role**: subj > direct obj > indirect obj
  **Elmo** went to the bar with **Grover**.
  **He** ordered two vodka tonics.
- **Parallelism**: **Ellen** went with **Linda** to Providence.
  Jim went with **her** to Boston.

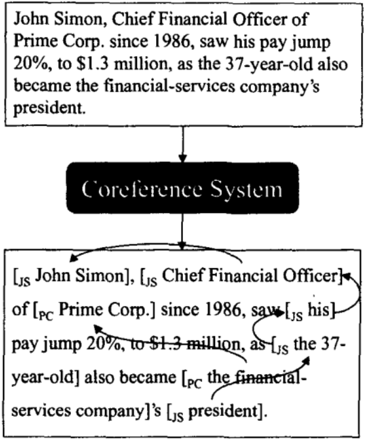
# Preferences

- **Recency**:
  Drew made **a nice pie**, but Liz made **an even better pie**.
  **It** had apples and bacon.
- **Repeated mention**
- **Grammatical role**: subj > direct obj > indirect obj
  **Elmo** went to the bar with **Grover**.
  **He** ordered two vodka tonics.
- **Parallelism**: **Ellen** went with **Linda** to Providence.
  Jim went with **her** to Boston.
- **Selectional preference**:
  They took **the dishes** from **the guests**, and washed **them**.

# Preferences

- **Recency**:
  Drew made **a nice pie**, but Liz made **an even better pie**.
  **It** had apples and bacon.
- **Repeated mention**
- **Grammatical role**: subj > direct obj > indirect obj
  **Elmo** went to the bar with **Grover**.
  **He** ordered two vodka tonics.
- **Parallelism**: **Ellen** went with **Linda** to Providence.
  Jim went with **her** to Boston.
- **Selectional preference**:
  They took **the dishes** from **the guests**, and washed **them**.
- **Semantics**:
  - Elmo telephoned Grover. He had broken the laptop.
  - Elmo yelled at Grover. He had broken the laptop.

# Coreference resolution




(Cardie and Wagstaff 1999)

# Features for classification

(1) Separately, Clinton transition officials said that *Frank Newman*, 50, *vice chairman* and chief financial officer of BankAmerica Corp., is expected to be nominated as assistant Treasury secretary for domestic finance.

**Table 1**
Feature vector of the markable pair ($i$ = *Frank Newman*, $j$ = *vice chairman*).

| Feature | Value | Comments |
|---|---|---|
| DIST | 0 | $i$ and $j$ are in the same sentence |
| I_PRONOUN | – | $i$ is not a pronoun |
| J_PRONOUN | – | $j$ is not a pronoun |
| STR_MATCH | – | $i$ and $j$ do not match |
| DEF_NP | – | $j$ is not a definite noun phrase |
| DEM_NP | – | $j$ is not a demonstrative noun phrase |
| NUMBER | + | $i$ and $j$ are both singular |
| SEMCLASS | 1 | $i$ and $j$ are both persons (This feature has three values: false(0), true(1), unknown(2).) |
| GENDER | 1 | $i$ and $j$ are both males (This feature has three values: false(0), true(1), unknown(2).) |
| PROPER_NAME | – | Only $i$ is a proper name |
| ALIAS | – | $j$ is not an alias of $i$ |
| APPOSITIVE | + | $j$ is in apposition to $i$ |

# Notes on features

Although the overall approach is statistical, a substantial amount of knowledge is encoded in the features.

- **STR_MATCH** tests string equality after stripping determiners: $\text{the data}$ matches $\text{that data}$

# Notes on features

Although the overall approach is statistical, a substantial amount of knowledge is encoded in the features.

- **STR_MATCH** tests string equality after stripping determiners: the data matches that data
- **SEM_CLASS** is based on WordNet, and considers FEMALE, MALE, PERSON, ORGANIZATION, LOCATION, DATE, TIME, MONEY, PERCENT, OBJECT.

# Notes on features

Although the overall approach is statistical, a substantial amount of knowledge is encoded in the features.

- **STR_MATCH** tests string equality after stripping determiners: the data matches that data
- **SEM_CLASS** is based on WordNet, and considers FEMALE, MALE, PERSON, ORGANIZATION, LOCATION, DATE, TIME, MONEY, PERCENT, OBJECT.
- **GENDER** is based on titles and a database of names

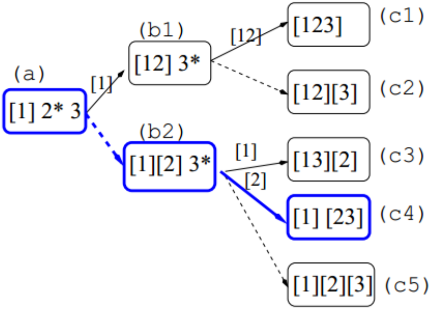
# Notes on features

Although the overall approach is statistical, a substantial amount of knowledge is encoded in the features.

- **STR_MATCH** tests string equality after stripping determiners: the data matches that data
- **SEM_CLASS** is based on WordNet, and considers FEMALE, MALE, PERSON, ORGANIZATION, LOCATION, DATE, TIME, MONEY, PERCENT, OBJECT.
- **GENDER** is based on titles and a database of names
- **ALIAS** is based on a database on known aliases, e.g. IBM/International Business Machines

# Bell Tree clustering for Coreference

We can incrementally build a coreference clustering using the **Bell Tree** representation (Luo et al, 2004)




- Can search over many clusterings, but size of Bell Tree (Bell Number) grows very rapidly, so pruning is needed.
- Approximate $P(y|x_i, e_j) \approx \max_{x' \in e_j} P(y|x_i, x')$

# Coreference as a Markov random field

(Wellner, McCallum, Peng, and Hey, 2003)

- $x_i$, $x_j$: mentions
- $y_{ij}$: coreference label for $x_i$ and $x_j$
- $\mathbf{f}(x_i, x_j, y_{ij})$: features of the mention pair
- $f_*(y_{ij}, y_{jk}, y_{ij})$: agreement feature for the label triple

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z_x} \exp\left(\sum_{i,j} \mathbf{w}^{\mathsf{T}}\mathbf{f}(x_i, x_j, y_{ij}) + \sum_{i,j,k} w_* f_*(y_{ij}, y_{jk}, y_{ij})\right)$$

Can learn weights $\mathbf{w}$ as a conditional random field or structured perceptron, but...

# Inference is NP-hard

- Proof: reduction from graph partitioning with positive and negative edge weights
- Approximate inference:
  - Correlational clustering: allow $y_{ij} \in [0, 1]$, then "round" to binary (Bansal et al, 2002)
  - Integer linear programming (Finkel and Manning 2008). Just add constraints:

$$(1 - y_{ij}) + (1 - y_{jk}) \geq (1 - y_{ik})$$

# Joint anaphoricity determination and coreference resolution

- Some nouns are not anaphoric, like $\text{It's raining}$.
- Anaphoricity determinization is usually applied as a preprocessing step, but coreference information can help.

# Joint anaphoricity determination and coreference resolution

- Some nouns are not anaphoric, like It's raining.
- Anaphoricity determinization is usually applied as a preprocessing step, but coreference information can help.
- Denis and Baldridge (2007) unified them:
  - $P_C(y_{ij}|\langle x_i, x_j \rangle) = \frac{1}{Z_{i,j}} \exp \mathbf{w}_C^\mathsf{T} \mathbf{f}_C(x_i, x_j, y_{ij})$
  - $P_A(z_i|x_i) = \frac{1}{Z_i} \exp \mathbf{w}_A^\mathsf{T} \mathbf{f}_A(x_i, z_i)$

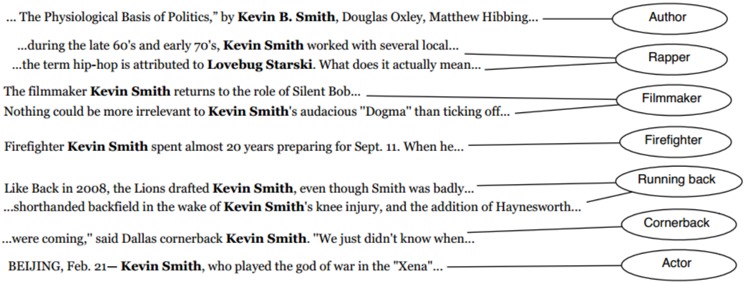
# Joint anaphoricity determination and coreference resolution

- Some nouns are not anaphoric, like It's raining.
- Anaphoricity determinization is usually applied as a preprocessing step, but coreference information can help.
- Denis and Baldridge (2007) unified them:
  - $P_C(y_{ij}|\langle x_i, x_j\rangle) = \frac{1}{Z_{i,j}} \exp \mathbf{w}_C^\mathsf{T} \mathbf{f}_C(x_i, x_j, y_{ij})$
  - $P_A(z_i|x_i) = \frac{1}{Z_i} \exp \mathbf{w}_A^\mathsf{T} \mathbf{f}_A(x_i, z_i)$

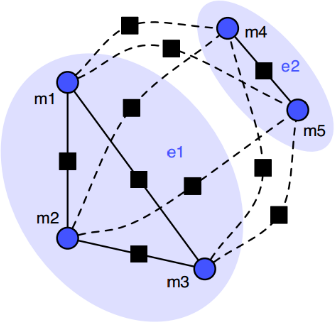
$$
\begin{aligned}
\max_{y,z} \quad & \sum_{\langle i,j\rangle} y_{i,j} \log P(y_{i,j}|x_i, x_j) + \sum_i z_i \log P(z_i|x_i) \\
s.t. \quad & y_{i,j}, z_i \in \{0, 1\} \ \forall i, j \\
& y_{i,j} \leq z_i, \ \forall i, j \\
& z_i \leq \sum_j y_{i,j}
\end{aligned}
$$

# Multi-document coreference resolution

Broaden the task: find all mentions of an entity across a big set of documents.



# A pairwise model



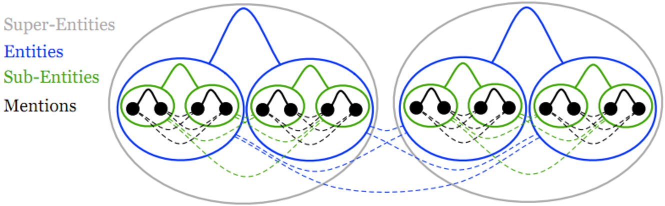

$$p(\mathbf{e}) \propto \exp \sum_{e \in \mathbf{e}} \left\{ \sum_{m,n \in e, n \neq m} \psi_a(m,n) + \sum_{m \in e, n \notin e} \psi_r(m,n) \right\}$$

Singh *et al.*(ACL 2011) estimate $p(\mathbf{e})$ using Metropolis-Hastings, with moves to swap mentions between entities.

# Parallelizing cross-document coreference

A four-level hierarchy:

- Mentions
- Sub-entities (sets of mentions that likely corefer)
- Entities (sets of mentions thought to corefer)
- Super-entities (sets of entities which might corefer)




- Key idea: parallelize by keeping super-entities on separate machines, occasionally shuffling them around.
- Using a very simple mention similarity function, they can scale up to 1.5M mention strings.