Unsupervised Entity Linking with Abstract Meaning Representation

Xiaoman Pan, Taylor Cassidy, Ulf Hermjakob, Heng Ji, Kevin Knight
Intro

• Natural Language Processing (NLP)
  o Interactions between computers and human languages
  o Machine Translation
  o Speech Recognition
  o Information Extraction
  o Information Retrieval
  o Dialog
I am cautiously anticipating the GOP nominee in 2012 not to be Mitt Romney.

Romney was the Governor of Massachusetts...

Romney is the great-great-grandson of a Mormon pioneer...

Republican candidates like Romney, Paul, and Johnson...
Challenges

- Ambiguity:
  - An entity mention could have multiple meanings
- Variability:
  - An entity could be expressed in many ways
A Typical Pipeline

• Entity candidates retrieval
  • **Salience**: The retrieved entity candidates should be salient and popular in KB

• Matching between the contexts of mention and the contexts of entity candidate
  o **Similarity**: The mention and the entity should have highly similar contexts
  o **Coherence**: The entity and its collaborators decided by the mention’s collaborators should be strongly connected in KB

  o Both require *context representation* and *context comparison*
Our Basic Idea

• Construct a Knowledge Network for mentions from Source

• Construct a Knowledge Network for entities from KBs

• Each Knowledge Network contains a thematically homogeneous coherent story/context

• Semantic Comparison between Knowledge Networks to match three criteria (Salience, Similarity and Coherence)
Abstract Meaning Representation (Banarescu et al., 2013)

(a) I am cautiously anticipating the GOP nominee in 2012 not to be Mitt Romney.

(b) Romney was the Governor of Massachusetts...

(c) Romney is the great-great-grandson of a Mormon pioneer...

(d) Republican candidates like Romney, Paul, and Johnson...
Core Semantic Roles

Did **Palin** apologize to **Giffords**?

- **Basic idea**: use all other concepts which played certain semantic roles in the same event as neighbors for the target entity mention.
Special Roles

• have-org-role-91
  o :ARG0 of have-org-role-91 is the office holder, typically a person
  o :ARG1 of have-org-role-91 is the organization, could also be a GPE
  o :ARG2 of have-org-role-91 is the title of the office held, e.g. president

Romney was the Governor of Massachusetts...
Special Roles

• have-rel-role-91
  o :ARG0 of have-rel-role-91 entity A
  o :ARG1 of have-rel-role-91 entity B
  o :ARG2 of have-rel-role-91 role of entity A (must be specified)
  o :ARG3 of have-rel-role-91 role of entity B (often left unspecified)

Romney is the great-great-grandson of a Mormon pioneer...
Coherent Mentions

• Entity mentions involved in AMR conjunction relations should be linked jointly to KB; their candidates in KB should also be strongly connected to each other with high semantic relatedness
  - “and”, “or”, “contrast-01”, “either”, “compared to”, “prep along with”, “neither”, “slash”, “between” and “both”

Republican candidates like *Romney, Paul, and Johnson...*
Putting Everything Together: Knowledge Network for Mentions in Source

- Johnson
- Romney
- Massachusetts
- Mormon pioneer
governor
grandson

- Republican

- Paul
- Mitt Romney
- GOP

-modify
- conjunction: and

- nominate:01: polarity
Construct Knowledge Network for Entities in KB

- Wikipedia
  - Infoboxes, Templates, Categories
  - Untyped hyperlinks within Wikipedia article text
- Typed relations within DBPedia and Freebase
- Google’s “people also search for” list
Construct Knowledge Network for Entities in KB
Commonness($m,e$) = \frac{\text{count}(m,e)}{\sum_{e'}\text{count}(m,e')} 

moderate for their political views. As a result, a number of potential "anti-Romney" candidates were put forward,\textsuperscript{[23]}\textsuperscript{[24]} including Donald Trump,\textsuperscript{[25]} Sarah Palin,\textsuperscript{[26]} Chris Christie,\textsuperscript{[27]} and Texas Governor Rick Perry,\textsuperscript{[28]} the last of

As in 2008, in 2012 Paul ultimately refused to endorse the ticket selected by the Republican Party. He said that there was no essential difference between Romney and his Democratic opponent, President Obama, on the
Salience based Ranking

- **Mitt Romney**
  - Mitt Romney presidential campaign, 2012
  - George W. Romney
  - Romney, West Virginia
  - New Romney
  - George Romney (painter)
  - HMS Romney (1708)
  - New Romney (UK Parliament constituency)
  - Romney family
  - Romney Expedition

- **Ron Paul**
  - Paul McCartney
  - Paul the Apostle
  - St Paul's Cathedral
  - Paul Martin
  - Paul Klee
  - Paul Allen
  - Chris Paul
  - Pauline epistles
  - Paul I of Russia

- **Gary Johnson**
  - Lyndon B. Johnson
  - Andrew Johnson
  - Samuel Johnson
  - Magic Johnson
  - Jimmie Johnson
  - Boris Johnson
  - Randy Johnson
  - Johnson & Johnson
  - Robert Johnson
**Similarity**

- $g(m)$: knowledge network for mention $m$
- $g(e_i)$: knowledge network for each entity candidate $e_i$ of $m$

- Compute similarity between $g(m)$ and $g(e_i)$ based on Jaccard Index

$$J(g(m), g(e_i)) = \frac{|g(m) \cap g(e_i)|}{|g(m) \cup g(e_i)|}$$

- Note that the edge labels are ignored
  Two elements are considered equal if and only if they have one or more token in common.
Similarity based Re-ranking

- **Mitt Romney**
  - George W. Romney
  - Mitt Romney presidential campaign, 2012
  - Ann Romney
  - Lenore Romney
  - Ronna Romney
  - Tagg Romney
  - G. Scott Romney
  - Vernom B. Romney
  - New Romney

- **Ron Paul**
  - Paul Ryan
  - Rand Paul
  - Paul McCartney
  - Paul Krugman
  - Paul Wellstone
  - Paul Broun
  - Paul Laxalt
  - Paul Coverdell
  - Paul Cellucci

- **Gary Johnson**
  - Lyndon B. Johnson
  - Andrew Johnson
  - Hiram Johnson
  - Sam Johnson
  - Tim Johnson (U.S. Senator)
  - Ron Johnson (U.S. politician)
  - Walter Johnson
  - Samuel Johnson
  - Magic Johnson
Coherence

- $R_m$: a set of coherent entity mentions
  - [Romney, Paul, Johnson]

- $R_E$: the set of corresponding entity candidate lists

- $C_m$: all the possible combinations of top candidate lists from $R_E$
  - [Mitt Romney, Ron Paul, Gary Johnson]
  - [Mitt Romney, Paul McCartney, Lyndon Johnson]
  - etc.

- Compute coherence for each combination $c \in C_m$ as Jaccard similarity, taking any number of arguments to the set of knowledge networks for all entities in $C$
Knowledge Network for Entities in KB

Mitt Romney

United States presidential candidates, 2012

Ron Paul

George W. Romney

Mitt Romney presidential campaign, 2012

Salt Lake City, Utah

Medicare

Author

Freebase object type

Freebase object type

Gary Johnson

Lyndon B. Johnson

Andrew Johnson

Governor of Massachusetts

Mormon

426208

Persondata

Infobox: religion

Infobox: political party

Infobox: other party

Living People

共和党

Paul McCartney

Paul the Apostle

Google people also search for

Wiki Template type

Dobedia wikiPageId
Coherence based Re-Ranking

- Mitt Romney
  - George W. Romney
  - Mitt Romney presidential campaign, 2012
  - Mitt Romney presidential campaign, 2008
  - List of Mitt Romney presidential campaign endorsements, 2012
  - Governorship of Mitt Romney
  - Ann Romney
  - Lenore Romney
  - Ronna Romney

- Ron Paul
  - Paul Ryan
  - Paul Krassner
  - Chris Paul
  - Paul Harvey
  - Ron Paul presidential campaign, 2008
  - Paul Samuelson
  - Rand Paul
  - Ron Paul presidential campaign, 2012
  - Paul McCartney

- Gary Johnson
  - Lyndon B. Johnson
  - Andrew Johnson
  - Magic Johnson
  - Woody Johnson
  - Boris Johnson
  - Jimmie Johnson
  - Dwayne Johnson
  - Donald Johnson
  - Hiram Johnson
Data

- AMR R3 Corpus (LDC2013E11) that includes manual EL annotations for all entity mentions (LDC2014E15)

- All discussion forum posts and 1/10 news documents

<table>
<thead>
<tr>
<th></th>
<th>PER</th>
<th>ORG</th>
<th>GPE</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>News</td>
<td>159</td>
<td>187</td>
<td>679</td>
<td>1,025</td>
</tr>
<tr>
<td>Discussion Forum</td>
<td>235</td>
<td>129</td>
<td>224</td>
<td>588</td>
</tr>
<tr>
<td>All</td>
<td>394</td>
<td>316</td>
<td>903</td>
<td>1,613</td>
</tr>
</tbody>
</table>

# of Entity Mentions
## Compare with Baseline and State-of-the-art

<table>
<thead>
<tr>
<th>Approach</th>
<th>News</th>
<th>DF</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Popularity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commonness</td>
<td>89.76</td>
<td>68.99</td>
<td>82.2</td>
</tr>
<tr>
<td>Google Search</td>
<td>88.10</td>
<td>77.17</td>
<td>84.12</td>
</tr>
<tr>
<td><strong>Supervised</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State-of-the-art supervised re-ranking using multi-level linguistic features for collaborators and collective inference, trained from 20,000 entity mentions from TAC-KBP2009-2014 (Chen and Ji, 2011; Cheng and Roth, 2013).</td>
<td>93.07</td>
<td>87.41</td>
<td>91.01</td>
</tr>
<tr>
<td><strong>Non-Collective</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System AMR (Flanigan et al., 2014)</td>
<td>90.15</td>
<td>85.69</td>
<td><strong>88.52</strong></td>
</tr>
</tbody>
</table>
Thank you!

(t / thank-01
  :ARG1 (y /you))

Entity Linking

en.wikipedia.org/wiki/Audience