Relation Extraction

Bill MacCartney
CS224U
23 April 2018

[with slides adapted from many people, including Dan Jurafsky, Rion Snow, Jim Martin, Chris Manning, William Cohen, Michele Banko, Mike Mintz, Steven Bills, and others]
Goal: “machine reading”

Reading the Web: A Breakthrough Goal for AI

I believe AI has an opportunity to achieve a true breakthrough over the coming decade by at last solving the problem of reading natural language text to extract its factual content. In fact, I hereby offer to bet anyone a lobster dinner that by 2015 we will have a computer program capable of automatically reading at least 80% of the factual content on the web, and placing those facts in a structured knowledge base. The significance of this AI achievement would be tremendous: it would immediately increase by many orders of magnitude the volume, breadth, and depth of ground facts and general knowledge accessible to knowledge based AI programs. In essence, computers would be harvesting in structured form the huge volume of knowledge that millions of humans are entering daily on the web in the form of unstructured text.

— Tom Mitchell, 2004
CHICAGO (AP) — Citing high fuel prices, United Airlines said Friday it has increased fares by $6 per round trip on flights to some cities also served by lower-cost carriers. American Airlines, a unit of AMR, immediately matched the move, spokesman Tim Wagner said. United, a unit of UAL, said the increase took effect Thursday night and applies to most routes where it competes against discount carriers, such as Chicago to Dallas and Atlanta and Denver to San Francisco, Los Angeles and New York.

Relation extraction example

<table>
<thead>
<tr>
<th>Subject</th>
<th>Relation</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Airlines</td>
<td>subsidiary</td>
<td>AMR</td>
</tr>
<tr>
<td>Tim Wagner</td>
<td>employee</td>
<td>American Airlines</td>
</tr>
<tr>
<td>United Airlines</td>
<td>subsidiary</td>
<td>UAL</td>
</tr>
</tbody>
</table>
Example: company relationships

Microsoft is working with Intel to improve laptop touchpads ...

Anobit Technologies was acquired by Apple for $450M.

Volkswagen partners with Apple on iBeetle ...
Example: gene regulation

textual abstract: summary for human

structured knowledge extraction: summary for machine

<table>
<thead>
<tr>
<th>Subject</th>
<th>Relation</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>p53</td>
<td>is_a</td>
<td>protein</td>
</tr>
<tr>
<td>Bax</td>
<td>is_a</td>
<td>protein</td>
</tr>
<tr>
<td>p53</td>
<td>has_function</td>
<td>apoptosis</td>
</tr>
<tr>
<td>Bax</td>
<td>has_function</td>
<td>induction</td>
</tr>
<tr>
<td>apoptosis</td>
<td>involved_in</td>
<td>cell_death</td>
</tr>
<tr>
<td>Bax</td>
<td>is_in</td>
<td>mitochondrial outer membrane</td>
</tr>
<tr>
<td>Bax</td>
<td>is_in</td>
<td>cytoplasm</td>
</tr>
<tr>
<td>apoptosis</td>
<td>related_to</td>
<td>caspase activation</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Lexical semantic relations

Many NLP applications require understanding relations between word senses: synonymy, antonymy, hyponymy, meronymy.

WordNet is a machine-readable database of relations between word senses, and an indispensable resource in many NLP tasks.

http://wordnetweb.princeton.edu/perl/webwn
WordNet is incomplete

But WordNet is manually constructed, and has many gaps!

<table>
<thead>
<tr>
<th>In WordNet 3.1</th>
<th>Not in WordNet 3.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>insulin</td>
<td>leptin</td>
</tr>
<tr>
<td>progesterone</td>
<td>pregnenolone</td>
</tr>
<tr>
<td>combustibility</td>
<td>affordability</td>
</tr>
<tr>
<td>navigability</td>
<td>reusability</td>
</tr>
<tr>
<td>HTML</td>
<td>XML</td>
</tr>
<tr>
<td>Google, Yahoo</td>
<td>Microsoft, IBM</td>
</tr>
</tbody>
</table>

Esp. for specific domains: restaurants, auto parts, finance
Esp. neologisms: iPad, selfie, bitcoin, twerking, Hadoop, dubstep
Mirror ran a headline questioning whether the killer’s actions were a result of playing Call of Duty, a first-person shooter game...

Melee, in video game terms, is a style of elbow-drop hand-to-hand combat popular in first-person shooters and other shooters.

Tower defense is a kind of real-time strategy game in which the goal is to protect an area/place/locality and prevent enemies from reaching...

Example: extending WordNet

- video game
- action game
- ball and paddle game
  - Breakout
- platform game
  - Donkey Kong
- shooter
  - arcade shooter
  - Space Invaders
- first-person shooter
  - Call of Duty
- third-person shooter
  - Tomb Raider
- adventure game
  - text adventure
  - graphic adventure
- strategy game
  - 4X game
  - Civilization
- tower defense
  - Plants vs. Zombies
Example: extending Freebase

Freebase: 20K relations, 40M entities, 70B facts
Curation is an ongoing challenge — things change!
Relies heavily on relation extraction from the web

/film/film/starring

<table>
<thead>
<tr>
<th>Wonder Woman</th>
<th>Gal Gadot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dunkirk</td>
<td>Tom Hardy</td>
</tr>
<tr>
<td>Tomb Raider</td>
<td>Alicia Vikander</td>
</tr>
</tbody>
</table>

/organization/organization/parent

<table>
<thead>
<tr>
<th>tbh</th>
<th>Facebook</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaggle</td>
<td>Google</td>
</tr>
<tr>
<td>LinkedIn</td>
<td>Microsoft</td>
</tr>
</tbody>
</table>

/music/artist/track

<table>
<thead>
<tr>
<th>Frank Ocean</th>
<th>Chanel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardi B</td>
<td>Bodak Yellow</td>
</tr>
<tr>
<td>Selena Gomez</td>
<td>Bad Liar</td>
</tr>
</tbody>
</table>

/film/film/starring

<table>
<thead>
<tr>
<th>Wonder Woman</th>
<th>Gal Gadot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dunkirk</td>
<td>Tom Hardy</td>
</tr>
<tr>
<td>Tomb Raider</td>
<td>Alicia Vikander</td>
</tr>
</tbody>
</table>

/people/person/date_of_death

<table>
<thead>
<tr>
<th>Barbara Bush</th>
<th>2018-04-17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milos Forman</td>
<td>2018-04-14</td>
</tr>
<tr>
<td>Winnie Mandela</td>
<td>2018-04-11</td>
</tr>
</tbody>
</table>
Approaches to relation extraction

1. Hand-built patterns
2. Bootstrapping methods
3. Supervised methods
4. Distant supervision
5. Other related work
Approaches to relation extraction

1. Hand-built patterns
2. Bootstrapping methods
3. Supervised methods
4. Distant supervision
5. Other related work
Patterns for learning hyponyms

- Intuition from Hearst (1992)
  
  *Agar is a substance prepared from a mixture of red algae, such as Gelidium, for laboratory or industrial use.*

- What does *Gelidium* mean?

- How do you know?
Patterns for learning hyponyms

- Intuition from Hearst (1992)
  
  Agar is a substance prepared from a mixture of red algae, such as Gelidium, for laboratory or industrial use.

- What does Gelidium mean?

- How do you know?
Hearst’s lexico-syntactic patterns

Ys such as X ((, X)* (, and/or) X)
such Ys as X...
X... or other Ys
X... and other Ys
Ys including X...
Ys, especially X...

Examples: “Ys, especially X”

The best part of the night was seeing all of the tweets of the performers, especially Miley Cyrus and Drake. ✓

Those child stars, especially Miley Cyrus, I feel like you have to put the fault on the media. ✓

Kelly wasn’t shy about sharing her feelings about some of the musical acts, especially Miley Cyrus. ✓

Rihanna was bored with everything at the MTV VMAs, especially Miley Cyrus. ✗

The celebrities enjoyed themselves while sipping on delicious cocktails, especially Miley Cyrus who landed the coveted #1 spot. ✗

None of these girls are good idols or role models, especially Miley Cyrus. ✗
Examples: “X was founded by Y”

**NeXT** was founded by **Steve Jobs** in 1985, after he was ousted from Apple Computers by John Sculley. ✓

Since 2002, when Blue Origin rival **SpaceX** was founded by **Elon Musk**, venture investment in the sector has increased markedly. ✓

**Microsoft** was founded by **Paul Allen** and Bill Gates on April 4, 1975, to develop and sell BASIC interpreters for the Altair 8800. ✓

The first successful commercial winery in **New York** was founded by **Jean Jacques** in 1839, at Washingtonville, on the west bank of the Hudson. ✗

One of the most obscure and fascinating companies implicated in the **Panama Papers** was founded by **Jürgen Mossack** in 1977. ✗

The largest annual space event on **Earth** was founded by **the United Nations General Assembly** and has been running every year since 1999. ✗
Examples: founder patterns

Elon Musk, the creator and founder of SpaceX, poked fun at the chaos his rocket launch caused for Californians on social media Friday night.

The co-founder of PayPal, Elon Musk, established SpaceX in 2002 with the goal of increasing space travel by reducing the cost of space launches.

Elon Musk co-founded PayPal and Tesla Motors, and created the space corporation SpaceX, which is credited with sending the first ...

SpaceX was founded in 2002 by entrepreneur Elon Musk with the goal of reducing space transportation costs.

Elon Musk is the founder, CEO and lead designer at Space Exploration Technologies (SpaceX), where he oversees ...

When Elon Musk first founded his rocket-ship company SpaceX, he had no idea how it would make a profit.
Problems with hand-built patterns

- Recall is not that great
  - Any finite set of patterns will fail to match many potential extractions

- Precision is not great either!
  - Many pattern-driven extractions are just wrong
  - Hearst: 66% accuracy on hyponym extraction

- Requires hand-building patterns for each relation!
  - And for every language!
  - Hard to write; hard to maintain
Approaches to relation extraction

1. Hand-built patterns
2. Bootstrapping methods
3. Supervised methods
4. Distant supervision
5. Other related work
Bootstrapping approaches

- If you have:
  - some *seed instances* of the relation
  - (or some patterns that work pretty well)
  - and lots & lots of *unannotated text* (e.g., the web)

- … can you use those seeds to do something useful?

- Bootstrapping can be considered *semi-supervised*
Bootstrapping example

- Target relation: *burial place*
- Seed tuple: \([\text{Mark Twain, Elmira}]\)
- Grep/Google for “Mark Twain” and “Elmira”
  
  “Mark Twain is buried in Elmira, NY.”
  \(\rightarrow\)  
  X is buried in Y

  “The grave of Mark Twain is in Elmira”
  \(\rightarrow\)  
  The grave of X is in Y

  “Elmira is Mark Twain’s final resting place”
  \(\rightarrow\)  
  Y is X’s final resting place

- Use those patterns to search for new tuples
Bootstrapping example

A bootstrapping example is shown in the image. It involves using search results to infer information about buried individuals. The example includes searches for people buried in Springfield, Illinois, the Hoover Dam, and a Pakistani singer named Mehnaz Begum.
The bootstrapping loop

slide adapted from Jim Martin
DIPRE (Brin 1998)

Extract (author, book) pairs
Start with these 5 seeds:

<table>
<thead>
<tr>
<th>Author</th>
<th>Book</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isaac Asimov</td>
<td>The Robots of Dawn</td>
</tr>
<tr>
<td>David Brin</td>
<td>Startide Rising</td>
</tr>
<tr>
<td>James Gleick</td>
<td>Chaos: Making a New Science</td>
</tr>
<tr>
<td>Charles Dickens</td>
<td>Great Expectations</td>
</tr>
<tr>
<td>William Shakespeare</td>
<td>The Comedy of Errors</td>
</tr>
</tbody>
</table>

Learn these patterns:

<table>
<thead>
<tr>
<th>URL Prefix</th>
<th>Text Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.sff.net/locus/c">www.sff.net/locus/c</a>.*</td>
<td>`&lt;LI&gt;&lt;B&gt;title&lt;/B&gt; by author (</td>
</tr>
<tr>
<td>dns.city-net.com/~lmann/awards/hugos/1984.html</td>
<td>`&lt;i&gt;title&lt;/i&gt; by author</td>
</tr>
<tr>
<td>dolphin.upenn.edu/~dcummins/texts/sf-award.htm</td>
<td>author</td>
</tr>
</tbody>
</table>
Bootstrapping problems

- Requires that we have seeds for each relation
  - Sensitive to original set of seeds
- Big problem of semantic drift at each iteration
- Precision tends to be not that high
- Generally have lots of parameters to be tuned
- No probabilistic interpretation
  - Hard to know how confident to be in each result
Approaches to relation extraction

1. Hand-built patterns
2. Bootstrapping methods
3. **Supervised methods**
4. Distant supervision
5. Other related work
Supervised relation extraction

For each pair of entities in a sentence, predict the relation type (if any) that holds between them.

The supervised approach requires:

- Defining an inventory of relation types
- Collecting labeled training data (the hard part!)
- Designing a feature representation
- Choosing a classifier: Naïve Bayes, MaxEnt, SVM, ...
- Evaluating the results
An inventory of relation types

<table>
<thead>
<tr>
<th>Type</th>
<th>Subtype</th>
</tr>
</thead>
<tbody>
<tr>
<td>ART (artifact)</td>
<td>User-Owner-Inventor-Manufacturer</td>
</tr>
<tr>
<td>GEN-AFF (General affiliation)</td>
<td>Citizen-Resident-Religion-Ethnicity, Org-Location</td>
</tr>
<tr>
<td>METONYMY*</td>
<td>None</td>
</tr>
<tr>
<td>ORG-AFF (Org-affiliation)</td>
<td>Employment, Founder, Ownership, Student-Alum, Sports-Affiliation, Investor-Shareholder, Membership</td>
</tr>
<tr>
<td>PART-WHOLE (part-to-whole)</td>
<td>Artifact, Geographical, Subsidiary</td>
</tr>
<tr>
<td>PER-SOC* (person-social)</td>
<td>Business, Family, Lasting-Personal</td>
</tr>
<tr>
<td>PHYS* (physical)</td>
<td>Located, Near</td>
</tr>
</tbody>
</table>

Relation types used in the ACE 2008 evaluation
Labeled training data

Datasets used in the ACE 2008 evaluation

<table>
<thead>
<tr>
<th>Source</th>
<th>Training epoch</th>
<th>Approximate size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>English Resources</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Broadcast News</td>
<td>3/03 – 6/03</td>
<td>55,000 words</td>
</tr>
<tr>
<td>Broadcast Conversations</td>
<td>3/03 – 6/03</td>
<td>40,000 words</td>
</tr>
<tr>
<td>Newswire</td>
<td>3/03 – 6/03</td>
<td>50,000 words</td>
</tr>
<tr>
<td>Weblog</td>
<td>11/04 – 2/05</td>
<td>40,000 words</td>
</tr>
<tr>
<td>Usenet</td>
<td>11/04 – 2/05</td>
<td>40,000 words</td>
</tr>
<tr>
<td>Conversational Telephone Speech</td>
<td>11/04-12/04 (differentiated by topic vs. eval)</td>
<td>40,000 words</td>
</tr>
<tr>
<td><strong>Arabic Resources</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Broadcast News</td>
<td>10/00 – 12/00</td>
<td>30,000+ words</td>
</tr>
<tr>
<td>Newswire</td>
<td>10/00 – 12/00</td>
<td>55,000+ words</td>
</tr>
<tr>
<td>Weblog</td>
<td>11/04 – 2/05</td>
<td>20,000+ words</td>
</tr>
</tbody>
</table>
Feature representations

- Lightweight features — require little pre-processing
  - Bags of words & bigrams between, before, and after the entities
  - Stemmed versions of the same
  - The types of the entities
  - The distance (number of words) between the entities

- Medium-weight features — require base phrase chunking
  - Base-phrase chunk paths
  - Bags of chunk heads

- Heavyweight features — require full syntactic parsing
  - Dependency-tree paths between the entities
  - Constituent-tree paths between the entities
  - Tree distance between the entities
  - Presence of particular constructions in a constituent structure
Classifiers

Now use any (multiclass) classifier you like:

- multiclass SVM
- MaxEnt (aka multiclass logistic regression)
- Naïve Bayes
- etc.
Zhou et al. 2005 results

<table>
<thead>
<tr>
<th>Type</th>
<th>Subtype</th>
<th>#Testing Instances</th>
<th>#Correct</th>
<th>#Error</th>
<th>P</th>
<th>R</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>Based-In</td>
<td>85</td>
<td>39</td>
<td>10</td>
<td>79.6</td>
<td>45.9</td>
<td>58.2</td>
</tr>
<tr>
<td></td>
<td>Located</td>
<td>241</td>
<td>132</td>
<td>120</td>
<td>52.4</td>
<td>54.8</td>
<td>53.5</td>
</tr>
<tr>
<td></td>
<td>Residence</td>
<td>66</td>
<td>19</td>
<td>9</td>
<td>67.9</td>
<td>28.8</td>
<td>40.4</td>
</tr>
<tr>
<td>NEAR</td>
<td>Relative-Location</td>
<td>35</td>
<td>8</td>
<td>1</td>
<td>88.9</td>
<td>22.9</td>
<td>36.4</td>
</tr>
<tr>
<td></td>
<td>Part-Of</td>
<td>136</td>
<td>76</td>
<td>32</td>
<td>70.4</td>
<td>55.9</td>
<td>62.3</td>
</tr>
<tr>
<td></td>
<td>Subsidiary</td>
<td>27</td>
<td>14</td>
<td>23</td>
<td>37.8</td>
<td>51.9</td>
<td>43.8</td>
</tr>
<tr>
<td>ROLE</td>
<td>Citizen-Of</td>
<td>36</td>
<td>25</td>
<td>8</td>
<td>75.8</td>
<td>69.4</td>
<td>72.6</td>
</tr>
<tr>
<td></td>
<td>General-Staff</td>
<td>201</td>
<td>108</td>
<td>46</td>
<td>71.1</td>
<td>53.7</td>
<td>62.3</td>
</tr>
<tr>
<td></td>
<td>Management</td>
<td>165</td>
<td>106</td>
<td>72</td>
<td>59.6</td>
<td>64.2</td>
<td>61.8</td>
</tr>
<tr>
<td></td>
<td>Member</td>
<td>224</td>
<td>104</td>
<td>36</td>
<td>74.3</td>
<td>46.4</td>
<td>57.1</td>
</tr>
<tr>
<td>SOCIAL</td>
<td>Other-Professional</td>
<td>29</td>
<td>16</td>
<td>32</td>
<td>33.3</td>
<td>55.2</td>
<td>41.6</td>
</tr>
<tr>
<td></td>
<td>Parent</td>
<td>25</td>
<td>17</td>
<td>0</td>
<td>100</td>
<td>68.0</td>
<td>81.0</td>
</tr>
</tbody>
</table>

Table 4: Performance of different relation types and major subtypes in the test data
Supervised RE: summary

- Supervised approach can achieve high accuracy
  - At least, for some relations
  - If we have lots of hand-labeled training data

- But has significant limitations!
  - Labeling 5,000 relations (+ named entities) is expensive
  - Doesn’t generalize to different relations, languages

- Next: beyond supervised relation extraction
  - Distantly supervised relation extraction
  - Unsupervised relation extraction
Approaches to relation extraction

1. Hand-built patterns
2. Bootstrapping methods
3. Supervised methods
4. Distant supervision
5. Other related work
Distant supervision paradigm

Snow, Jurafsky, Ng. 2005. Learning syntactic patterns for automatic hypernym discovery. NIPS 17


- Hypothesis: If two entities belong to a certain relation, any sentence containing those two entities is likely to express that relation
- Key idea: use a database of relations to get lots of training examples
  - instead of hand-creating a few seed tuples (bootstrapping)
  - instead of using hand-labeled corpus (supervised)
We construct a noisy training set consisting of occurrences from our corpus that contain a hyponym-hypernym pair from WordNet.

This yields high-signal examples like:

“...consider authors like Shakespeare...”

“Some authors (including Shakespeare)...”

“Shakespeare was the author of several...”

“Shakespeare, author of The Tempest...”
Hypernyms via distant supervision

We construct a noisy training set consisting of occurrences from our corpus that contain a hyponym-hypernym pair from WordNet.

This yields high-signal examples like:

“...consider authors like Shakespeare...”
“Some authors (including Shakespeare)...”
“Shakespeare was the author of several...”
“Shakespeare, author of The Tempest...”

But also noisy examples like:

“The author of Shakespeare in Love...”
“...authors at the Shakespeare Festival...”
Learning hypernym patterns

1. Take 6M newswire sentences
   ...doubly heavy hydrogen atom called deuterium...
2. Collect noun pairs
3. Is pair a hypernym in WordNet? e.g. (atom, deuterium)
   752,311 pairs from 6M sentences of newswire
   14,387 yes; 737,924 no
4. Parse the sentences
5. Extract patterns
6. Train classifier on patterns
   logistic regression with 70K features
   (converted to 974,288 bucketed binary features)

...doubly heavy hydrogen atom called deuterium...

69,592 dependency paths with >5 pairs
One of 70,000 patterns

Pattern: `<superordinate>` called `<subordinate>`
or: `<Y>` called `<X>`

Learned from cases such as:

(*sarcoma*, *cancer*)  ...an uncommon bone *cancer* called osteogenic *sarcoma* and to...

(*deuterium*, *atom*)  ...heavy water rich in the doubly heavy hydrogen *atom* called *deuterium*.

New pairs discovered:

(*efflorescence*, *condition*)  ...and a *condition* called *efflorescence* are other reasons for...

(*O’neal_inc*, *company*)  ...The *company*, now called O’Neal Inc., was sole distributor of...

(*hat_creek_outfit*, *ranch*)  ...run a small *ranch* called the Hat Creek Outfit.

(*hiv-1*, *aids_virus*)  ...infected by the *AIDS virus*, called *HIV-1*.

(*bateau_mouche*, *attraction*)  ...local sightseeing *attraction* called the Bateau Mouche...
Syntactic dependency paths

Patterns are based on paths through dependency parses generated by MINIPAR (Lin, 1998)

Example word pair: (Shakespeare, author)
Example sentence: “Shakespeare was the author of several plays...”

Minipar parse:

Extract shortest path: -N:s:VBE, be, VBE:pred:N
P/R of hypernym extraction patterns

Individual Feature Analysis

Y including X

slide adapted from Rion Snow
P/R of hypernym extraction patterns

Individual Feature Analysis

- Y including X
- Y such as X

slide adapted from Rion Snow
Individual Feature Analysis

- Y including X
- Y such as X
- X and/or other Y
P/R of hypernym extraction patterns

Individual Feature Analysis

- X and/or other Y
- Y such as X
- such Y as X
- Y including X
- Y, especially X
- Y like X
- Y called X
- X is Y
- X, a Y (appositive)
P/R of hypernym classifier

logistic regression

$P(R|E) = \frac{1}{1 + e^{-\sum w_i x_i}}$

10-fold Cross Validation on 14,000 WordNet-Labeled Pairs
P/R of hypernym classifier

logistic regression

\[ P(R|E) = \frac{1}{1 + e^{-\sum w_i x_i}} \]

10-fold Cross Validation on 14,000 WordNet-Labeled Pairs
What about other relations?

Mintz, Bills, Snow, Jurafsky (2009).
Distant supervision for relation extraction without labeled data.

**Training set**
- 102 relations
- 940,000 entities
- 1.8 million instances

**Corpus**
- 1.8 million articles
- 25.7 million sentences
<table>
<thead>
<tr>
<th>Relation name</th>
<th>Size</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>/people/person/nationality</td>
<td>281,107</td>
<td>John Dugard, South Africa</td>
</tr>
<tr>
<td>/location/location/contains</td>
<td>253,223</td>
<td>Belgium, Nijlen</td>
</tr>
<tr>
<td>/people/person/profession</td>
<td>208,888</td>
<td>Dusa McDuff, Mathematician</td>
</tr>
<tr>
<td>/people/person/place_of_birth</td>
<td>105,799</td>
<td>Edwin Hubble, Marshfield</td>
</tr>
<tr>
<td>/dining/restaurant/cuisine</td>
<td>86,213</td>
<td>MacAyo’s Mexican Kitchen, Mexican</td>
</tr>
<tr>
<td>/business/business_chain/location</td>
<td>66,529</td>
<td>Apple Inc., Apple Inc., South Park, NC</td>
</tr>
<tr>
<td>/biology/organism_classification_rank</td>
<td>42,806</td>
<td>Scorpaeniformes, Order</td>
</tr>
<tr>
<td>/film/film/genre</td>
<td>40,658</td>
<td>Where the Sidewalk Ends, Film noir</td>
</tr>
<tr>
<td>/film/film/language</td>
<td>31,103</td>
<td>Enter the Phoenix, Cantonese</td>
</tr>
<tr>
<td>/biology/organism_higher_classification</td>
<td>30,052</td>
<td>Calopteryx, Calopterygidae</td>
</tr>
<tr>
<td>/film/film/country</td>
<td>27,217</td>
<td>Turtle Diary, United States</td>
</tr>
<tr>
<td>/film/writer/film</td>
<td>23,856</td>
<td>Irving Shulman, Rebel Without a Cause</td>
</tr>
<tr>
<td>/film/director/film</td>
<td>23,539</td>
<td>Michael Mann, Collateral</td>
</tr>
<tr>
<td>/film/producer/film</td>
<td>22,079</td>
<td>Diane Eskenazi, Aladdin</td>
</tr>
<tr>
<td>/people/deceased_person/place_of_death</td>
<td>18,814</td>
<td>John W. Kern, Asheville</td>
</tr>
<tr>
<td>/music/artist/origin</td>
<td>18,619</td>
<td>The Octopus Project, Austin</td>
</tr>
<tr>
<td>/people/person/religion</td>
<td>17,582</td>
<td>Joseph Chartrand, Catholicism</td>
</tr>
<tr>
<td>/book/author/works_written</td>
<td>17,278</td>
<td>Paul Auster, Travels in the Scriptorium</td>
</tr>
<tr>
<td>/soccer/football_position(players)</td>
<td>17,244</td>
<td>Midfielder, Chen Tao</td>
</tr>
<tr>
<td>/people/deceased_person/cause_of_death</td>
<td>16,709</td>
<td>Richard Daintree, Tuberculosis</td>
</tr>
<tr>
<td>/film/film/music</td>
<td>14,070</td>
<td>Stavisky, Stephen Sondheim</td>
</tr>
<tr>
<td>/business/company/industry</td>
<td>13,805</td>
<td>ATS Medical, Health care</td>
</tr>
</tbody>
</table>
Collecting training data

Corpus text

Bill Gates founded Microsoft in 1975.
Bill Gates, founder of Microsoft, ...
Bill Gates attended Harvard from ...
Google was founded by Larry Page ...

Freebase

Founder: (Bill Gates, Microsoft)
Founder: (Larry Page, Google)
CollegeAttended: (Bill Gates, Harvard)
Collecting training data

Corpus text

Bill Gates founded Microsoft in 1975.
Bill Gates, founder of Microsoft, ...
Bill Gates attended Harvard from ...
Google was founded by Larry Page ...

Freebase

Founder: (Bill Gates, Microsoft)
Founder: (Larry Page, Google)
CollegeAttended: (Bill Gates, Harvard)

Training data

(Bill Gates, Microsoft)
Label: Founder
Feature: X founded Y
Collecting training data

Corpus text

Bill Gates founded Microsoft in 1975. Bill Gates, founder of Microsoft, ... Bill Gates attended Harvard from ... Google was founded by Larry Page ...

Freebase

Founder: (Bill Gates, Microsoft)
CollegeAttended: (Bill Gates, Harvard)

Training data

(Bill Gates, Microsoft)
Label: Founder
Feature: X founded Y
Feature: X, founder of Y
Collecting training data

Corpus text

Bill Gates founded Microsoft in 1975.
Bill Gates, founder of Microsoft, ...
Bill Gates attended Harvard from...
Google was founded by Larry Page ...

Freebase

Founder: (Bill Gates, Microsoft)
Founder: (Larry Page, Google)
CollegeAttended: (Bill Gates, Harvard)

Training data

(Bill Gates, Microsoft)
Label: Founder
Feature: X founded Y
Feature: X, founder of Y

(Bill Gates, Harvard)
Label: CollegeAttended
Feature: X attended Y
Collecting training data

Corpus text

Bill Gates founded Microsoft in 1975.
Bill Gates, founder of Microsoft, ...
Bill Gates attended Harvard from ...

Google was founded by Larry Page ...

Freebase

Founder: (Bill Gates, Microsoft)
Founder: (Larry Page, Google)
CollegeAttended: (Bill Gates, Harvard)

Training data

(Bill Gates, Microsoft)
Label: Founder
Feature: X founded Y
Feature: X, founder of Y

(Bill Gates, Harvard)
Label: CollegeAttended
Feature: X attended Y

(Larry Page, Google)
Label: Founder
Feature: Y was founded by X
Can’t train a classifier with only positive data! Need negative training data too!

Solution?
Sample 1% of unrelated pairs of entities.

Result: roughly balanced data.

Corpus text
Larry Page took a swipe at Microsoft... ...after Harvard invited Larry Page to... Google is Bill Gates' worst fear ...

Training data

(Larry Page, Microsoft)
Label: NO_RELATION
Feature: X took a swipe at Y

(Larry Page, Harvard)
Label: NO_RELATION
Feature: Y invited X

(Bill Gates, Google)
Label: NO_RELATION
Feature: Y is X's worst fear

Can’t train a classifier with only positive data! Need negative training data too!
The experiment

Positive training data

(Bill Gates, Microsoft)
Label: Founder
Feature: X founded Y
Feature: X, founder of Y

(Bill Gates, Harvard)
Label: CollegeAttended
Feature: X attended Y

(Larry Page, Google)
Label: Founder
Feature: Y was founded by X

Negative training data

(Larry Page, Microsoft)
Label: NO_RELATION
Feature: X took a swipe at Y

(Larry Page, Harvard)
Label: NO_RELATION
Feature: Y invited X

(Bill Gates, Google)
Label: NO_RELATION
Feature: Y is X’s worst fear

Learning: multiclass logistic regression

Test data

(Henry Ford, Ford Motor Co.)
Label: ???
Feature: X founded Y
Feature: Y was founded by X

(Steve Jobs, Reed College)
Label: NO_RELATION
Feature: X attended Y

Test data

(Henry Ford, Ford Motor Co.)
Label: Founder

(Steve Jobs, Reed College)
Label: CollegeAttended
Benefits of distant supervision

- Has advantages of supervised approach
  - leverage rich, reliable hand-created knowledge
  - relations have canonical names
  - can use rich features (e.g. syntactic features)

- Has advantages of unsupervised approach
  - leverage unlimited amounts of text data
  - allows for very large number of weak features
  - not sensitive to training corpus: genre-independent
Astronomer Edwin Hubble was born in Marshfield, Missouri.
### High-weight features

<table>
<thead>
<tr>
<th>Relation</th>
<th>Feature type</th>
<th>Left window</th>
<th>NE1</th>
<th>Middle</th>
<th>NE2</th>
<th>Right window</th>
</tr>
</thead>
<tbody>
<tr>
<td>/architecture/structure/architect</td>
<td>LEX -</td>
<td>ORG</td>
<td>the designer of the</td>
<td>PER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/book/author/works_written</td>
<td>SYN</td>
<td>designed</td>
<td>ORG</td>
<td>↑<em>a designed $\downarrow</em>{beg} \downarrow_{subj}$ by $\downarrow_{pen}$</td>
<td>PER</td>
<td>↑_a designed</td>
</tr>
<tr>
<td>/business/company/founders</td>
<td>SYN</td>
<td>PER</td>
<td>co-founder</td>
<td>PER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/business/company/place.founded</td>
<td>LEX -</td>
<td>ORG</td>
<td>an owner</td>
<td>PER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/film/film/country</td>
<td>SYN</td>
<td>ORG</td>
<td>released</td>
<td>LOC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/geography/river/mouth</td>
<td>LEX</td>
<td>LOC</td>
<td>flows into</td>
<td>LOC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/government/political_party/country</td>
<td>SYN</td>
<td>the $\downarrow_{det}$</td>
<td>LOC</td>
<td>$\downarrow_{beg}$ is $\downarrow_{pred}$ tributary $\downarrow_{mod}$ of $\downarrow_{pen}$</td>
<td>LOC</td>
<td>$\downarrow_{beg}$ the</td>
</tr>
<tr>
<td>/influences/influence_node/influenced</td>
<td>LEX -</td>
<td>ORG</td>
<td>politician of the</td>
<td>LOC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/language/human_language/region</td>
<td>SYN</td>
<td>candidate</td>
<td>ORG</td>
<td>$\uparrow_{nn}$ candidate $\downarrow_{mod}$ for $\downarrow_{pen}$</td>
<td>LOC</td>
<td>$\uparrow_{nn}$ candidate</td>
</tr>
<tr>
<td>/music/artist/origin</td>
<td>LEX -</td>
<td>ORG</td>
<td>speaking areas of</td>
<td>LOC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/people/deceased.person/place.of.death</td>
<td>SYN</td>
<td>of $\downarrow_{pen}$</td>
<td>PER</td>
<td>$\downarrow_{beg}$ of $\downarrow_{mod}$ student $\downarrow_{appo}$</td>
<td>PER</td>
<td>$\downarrow_{beg}$ of</td>
</tr>
<tr>
<td>/people/person/nationality</td>
<td>LEX</td>
<td>LOC</td>
<td>died in</td>
<td>LOC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/people/person/patients</td>
<td>SYN</td>
<td>of $\downarrow_{pen}$</td>
<td>PER</td>
<td>is a citizen of</td>
<td>LOC</td>
<td></td>
</tr>
<tr>
<td>/people/person/patients</td>
<td>LEX -</td>
<td>ORG</td>
<td>based band</td>
<td>LOC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/people/person/place.of.birth</td>
<td>SYN</td>
<td>the $\downarrow_{beg}$</td>
<td>PER</td>
<td>$\downarrow_{beg}$ born $\downarrow_{pred}$</td>
<td>LOC</td>
<td></td>
</tr>
<tr>
<td>/people/person/religion</td>
<td>LEX -</td>
<td>ORG</td>
<td>embraced</td>
<td>LOC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Experimental set-up

● 1.8 million relation instances used for training
  ○ Compared to 17,000 relation instances in ACE

● 800,000 Wikipedia articles used for training, 400,000 different articles used for testing

● Only extract relation instances not already in Freebase
It works!

Ten relation instances extracted by the system that weren’t in Freebase

<table>
<thead>
<tr>
<th>Relation name</th>
<th>New instance</th>
</tr>
</thead>
<tbody>
<tr>
<td>/location/location/contains</td>
<td>Paris, Montmartre</td>
</tr>
<tr>
<td>/location/location/contains</td>
<td>Ontario, Fort Erie</td>
</tr>
<tr>
<td>/music/artist/origin</td>
<td>Mighty Wagon, Cincinnati</td>
</tr>
<tr>
<td>/people/deceased_person/place_of_death</td>
<td>Fyodor Kamensky, Clearwater</td>
</tr>
<tr>
<td>/people/person/nationality</td>
<td>Marianne Yvonne Heemskerk, Netherlands</td>
</tr>
<tr>
<td>/people/person/place_of_birth</td>
<td>Wavell Wayne Hinds, Kingston</td>
</tr>
<tr>
<td>/book/author/works_written</td>
<td>Upton Sinclair, Lanny Budd</td>
</tr>
<tr>
<td>/business/company/founders</td>
<td>WWE, Vince McMahon</td>
</tr>
<tr>
<td>/people/person/profession</td>
<td>Thomas Mellon, judge</td>
</tr>
</tbody>
</table>
Evaluation

- Held-out evaluation
  - Train on 50% of gold-standard Freebase relation instances, test on other 50%
  - Used to tune parameters quickly without having to wait for human evaluation

- Human evaluation
  - Performed by evaluators on Amazon Mechanical Turk
  - Calculated precision at 100 and 1000 recall levels for the ten most common relations
Held-out evaluation

Automatic evaluation on 900K instances of 102 Freebase relations. Precision for three different feature sets is reported at various recall levels.
Distant supervision: takeaways

- The distant supervision approach uses a database of known relation instances as a source of supervision.
- We’re classifying pairs of entities, not pairs of entity mentions.
- The features for a pair of entities describe the patterns in which the two entities have co-occurred across many sentences in a large corpus.
- Can make use of 100x or even 1000x more data than in the supervised paradigm.
Approaches to relation extraction

1. Hand-built patterns
2. Bootstrapping methods
3. Supervised methods
4. Distant supervision
5. Other related work
What else is out there?

- **Open information extraction** (OpenIE) aims to extract *all* relations from text, without supervision or any fixed set of relations.

  (Google, is based in, Mountain View)
  (Mountain View, is home to, Google)

- **Knowledge base completion** (KBC) aims to use information in a KB to fill in missing entries.

  (AB, country_of_birth, Iceland)
  => (AB, speaks_language, Icelandic)
OpenIE demo

http://openie.allenai.org/
For next time

- Read Mintz et al. 2009
- Start working through the relation extraction codebook, rel_ext.ipynb