Simple Large-scale Relation Extraction from Unstructured Text

Christos Christodoulopoulos and Arpit Mittal
Amazon Research Cambridge
Alexa Question Answering

“Alexa, what books did Carrie Fisher write?”

“The books that Carrie Fisher is an author of are Delusions of Grandma, Shockaholic, Surrender the Pink, Postcards from the Edge, The Best Awful There Is and Wishful Drinking.”
Alexa Knowledge Base

Named relations between entities

Carrie Fisher is the author of Postcards from the Edge, which is an instance of book.
Alexa Knowledge Base
Alexa Knowledge Base

Sources of knowledge:
1. Human authorship
2. Structured information
3. Unstructured information
Knowledge from Unstructured Text

The Goal:
Carrie Fisher wrote several semi-autobiographical novels, including Postcards from the Edge.
Knowledge from Unstructured Text

The Goal:
Carrie Fisher wrote several semi-autobiographical novels, including Postcards from the Edge.
Knowledge from Unstructured Text

The Goal:
Carrie Fisher wrote several semi-autobiographical novels, including Postcards from the Edge.

Entity Recognition
- Entity Resolution
- Relation Extraction
- Likelihood Estimation
Knowledge from Unstructured Text

The Goal: Carrie Fisher wrote several semi-autobiographical novels, including Postcards from the Edge.

[carrrie_fisher]  [postcards_from_the_edge]
Knowledge from Unstructured Text

The Goal:

Carrie Fisher wrote several semi-autobiographical novels, including Postcards from the Edge.

[carrie fisher] [is the author of] [postcards from the edge]
Knowledge from Unstructured Text

The Goal:

Carrie Fisher wrote several semi-autobiographical novels, including Postcards from the Edge.

[carrrie fisher] [is the author of] [postcards from the edge]

Ontological constraints
Entity embeddings
Distributional information

98% likelihood

Entity Recognition
Entity Resolution
Relation Extraction
Likelihood Estimation
Learning approaches for RE

• Rule-based
• Fully supervised
• Unsupervised
• Distant/weakly supervised
  • Snow, Jurafsky, Ng, 2005
  • Main assumption: if two entities are linked by a relation, any sentence containing both sentences is likely to express that relation
    • [steven spielberg] [is the director of] [saving private ryan]
    • “Spielberg’s film Saving Private Ryan is based on…”
Learning approaches for RE

• Rule-based
• Fully supervised
• Unsupervised
• Distant/weakly supervised
  • Snow, Jurafsky, Ng, 2005
  • Main assumption: if two entities are linked by a relation, any sentence containing both sentences is likely to express that relation
    • [steven spielberg] [is the director of] [saving private ryan]
    • “Spielberg’s film Saving Private Ryan is based on…”

Christodoulopoulos and Mittal (under review)
Distant supervision label generation

Wikipedia → Chunking PoS Tagging → Entity denotations (surface forms) → Gazetteers → Entity pairs (KB IDs)
Distant supervision label generation

Wikipedia → Chunking PoS Tagging → Entity denotations (surface forms) → Gazetteers → Entity pairs (KB IDs)

Ontological Constraints → Check against KB

- YES: Positive Label
- NO: Negative Label
His studies were interrupted by army service and at the end of the war he was forced to return...
[the second world war] [is an instance of] [cause of death]

In the intro to the song, Fred Durst makes reference to...
[intro 15367][is an instance of] [song]

Turner also released one album and several singles under the moniker Repeat.
[the singles the 2011 album] [is an instance of] [album]
Distant supervision label generation

- **Wikipedia page**
  - URL → KB ID lookup
  - Main entity (KB ID)
  - KB
    - Related entities (x)
      - (KB IDs)
    - KB ID → Denotations lookup
      - Entity denotations (x + main strings)

- **Ontological Constraints**
  - Check against KB
    - YES → Positive Label
    - NO → Negative Label

- **Chunking PoS Tagging**
  - Entity denotations (surface forms)
    - Gazetteers
      - Entity pairs (KB IDs)

- **Gazetteers**
  - Entity pairs (KB IDs)
Call Your Girlfriend was written by Robyn, Alexander Kronlund and Klas Ahlund, with the latter producing the song.

[call your girlfriend 3] [is an instance of] [song]

Forget Her is a song by Jeff Buckley.

[forget her] [is an instance of] [song]

The Subei Mongol Autonomous County is an autonomous county within the prefecture-level city of Jiuquan in the northwestern Chinese province of Gansu.

[subei mongol autonomous county] [is an instance of] [Chinese county]
Relation extraction

• HypeNET (Shwartz and Goldberg, 2016)
• Hyponyms [is an instance of] only
  • LexNET extends to multiple relations
Relation extraction

- fastText (Joulin et al., 2016)
- Linear model
  - One hidden layer
  - Rank constraint
Results

HypeNET equally good as the much simpler fastText with the same input features.

<table>
<thead>
<tr>
<th>Relation</th>
<th>HypeNET</th>
<th>fastText</th>
</tr>
</thead>
<tbody>
<tr>
<td>[is an instance of]</td>
<td>94.29 (0.21)</td>
<td>94.31 (0.03)</td>
</tr>
<tr>
<td>[is the birthplace of]</td>
<td>85.57 (0.26)</td>
<td>87.63 (0.01)</td>
</tr>
<tr>
<td>[applies to]</td>
<td>81.98 (1.78)</td>
<td>86.17 (0.01)</td>
</tr>
</tbody>
</table>
## Results

HypeNET **equally good** as the much simpler fastText with the same input features.

### Alexa KB

<table>
<thead>
<tr>
<th>Relation</th>
<th>HypeNET</th>
<th>fastText</th>
</tr>
</thead>
<tbody>
<tr>
<td>[is an instance of]</td>
<td>94.29 (0.21)</td>
<td>94.31 (0.03)</td>
</tr>
<tr>
<td>[is the birthplace of]</td>
<td>85.57 (0.26)</td>
<td>87.63 (0.01)</td>
</tr>
<tr>
<td>[applies to]</td>
<td>81.98 (1.78)</td>
<td>86.17 (0.01)</td>
</tr>
</tbody>
</table>

### Wikidata

<table>
<thead>
<tr>
<th>Relation</th>
<th>HypeNET</th>
<th>fastText</th>
</tr>
</thead>
<tbody>
<tr>
<td>instance of (P31)</td>
<td>93.90 (0.21)</td>
<td><strong>96.44 (0.01)</strong></td>
</tr>
<tr>
<td>birthplace of (P19)</td>
<td>92.06 (0.90)</td>
<td><strong>93.05 (0.07)</strong></td>
</tr>
<tr>
<td>part of (P527)</td>
<td>48.73 (2.59)</td>
<td><strong>72.87 (0.16)</strong></td>
</tr>
</tbody>
</table>
## Results

HypeNET **equally good** as the much simpler fastText with the same input features.

MaxEnt results show that features alone are not enough. Need to create higher-dimensional representations of discrete features.

### Alexa KB

<table>
<thead>
<tr>
<th>Relation</th>
<th>HypeNET</th>
<th>fastText</th>
<th>MaxEnt</th>
</tr>
</thead>
<tbody>
<tr>
<td>[is an instance of]</td>
<td>94.29 (0.21)</td>
<td>94.31 (0.03)</td>
<td>83.93</td>
</tr>
<tr>
<td>[is the birthplace of]</td>
<td>85.57 (0.26)</td>
<td>87.63 (0.01)</td>
<td>80.83</td>
</tr>
<tr>
<td>[applies to]</td>
<td>81.98 (1.78)</td>
<td>86.17 (0.01)</td>
<td>65.27</td>
</tr>
</tbody>
</table>

### Wikidata

<table>
<thead>
<tr>
<th>Relation</th>
<th>HypeNET</th>
<th>fastText</th>
<th>MaxEnt</th>
</tr>
</thead>
<tbody>
<tr>
<td>instance of (P31)</td>
<td>93.90 (0.21)</td>
<td>96.44 (0.01)</td>
<td>58.45</td>
</tr>
<tr>
<td>birthplace of (P19)</td>
<td>92.06 (0.90)</td>
<td>93.05 (0.07)</td>
<td>66.72</td>
</tr>
<tr>
<td>part of (P527)</td>
<td>48.73 (2.59)</td>
<td>72.87 (0.16)</td>
<td>45.13</td>
</tr>
</tbody>
</table>
Summary

• New method for entity resolution
  • Page-specific gazetteers

• Features are important
  • HypeNET vs fastText

• Feature representation is important
  • fastText vs MaxEnt
Future directions

• Enhanced entity recognition
• Use of human annotation for seeding supervision
• Expanding to multiple sources of text
• Coverage of multiple languages
Thanks!
Carrie Fisher wrote several semi-autobiographical novels, including *Postcards from the Edge*.

(shortest) path between entities: X → wrote → several → including → Y
Dependency parsing for RE

Carrie Fisher wrote several semi-autobiographical novels, including Postcards from the Edge.

(shortest) path between entities: X \(\rightarrow\) wrote \(\rightarrow\) several \(\rightarrow\) including \(\rightarrow\) Y

Carrie Fisher, who was friends with Steven Spielberg, wrote several semi-autobiographical novels, including Postcards from the Edge.
Results – feature ablation

<table>
<thead>
<tr>
<th></th>
<th>instance of</th>
<th>applies to</th>
<th>birthplace of</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) 5 supports</td>
<td>94.55</td>
<td>86.26</td>
<td>87.56</td>
</tr>
<tr>
<td>all supports</td>
<td>94.33</td>
<td>85.92</td>
<td>87.63</td>
</tr>
<tr>
<td>(1)-Brown</td>
<td>94.20</td>
<td>85.93</td>
<td>87.51</td>
</tr>
<tr>
<td>(1)-lemma</td>
<td>94.17</td>
<td>84.15</td>
<td>86.65</td>
</tr>
<tr>
<td>(1)-POS</td>
<td>94.15</td>
<td>85.93</td>
<td>87.71</td>
</tr>
<tr>
<td>(1)-dep</td>
<td>93.59</td>
<td>85.42</td>
<td>86.53</td>
</tr>
<tr>
<td>(1)-X/Y entities</td>
<td>93.63</td>
<td>83.89</td>
<td>86.95</td>
</tr>
<tr>
<td>X/Y only</td>
<td>91.15</td>
<td>74.20</td>
<td>81.15</td>
</tr>
<tr>
<td>full sentence</td>
<td>86.70</td>
<td>77.77</td>
<td>87.09</td>
</tr>
</tbody>
</table>
Results

Training data size

F-Score vs Train Data Size for different classifiers:
- HypeNET++
- fastText classifier
- MaxEnt classifier
Results

Using dependency satellite nodes

<table>
<thead>
<tr>
<th>Type</th>
<th>With Satellites</th>
<th>Without Satellites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>is an instance of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>is the birthplace of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>applies to</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results

Grouping supports for each entity pair

Relation
- is an instance of
- is the birthplace of
- applies to

Type
- Grouped Supports
- Ungrouped Supports