Relation Extraction

Presenter: Guoxin Gu
Relation Extraction

Knowledge-Based Approach
- Domain-Dependent
- Domain-Independent
- Low recall & High precision

Supervised Approach
- Features-based

Semi-supervised Approach
- DIPRE
- Snowball

Unsupervised Approach
- OIE

Distant Supervision
Pipeline

Unstructured

- raw text (string)
  - sentence segmentation
    - sentences (list of strings)
    - tokenization
      - tokenized sentences (list of lists of strings)
      - part of speech tagging

Structured

- pos-tagged sentences (list of lists of tuples)
  - entity detection
    - chunked sentences (list of trees)
  - relation detection
    - relations (list of tuples)
Knowledge-Based Approach

- Domain-Dependent
- Domain-Independent
- Low recall & High precision
Knowledge-Based Approach

Domain-Dependent (Riloff et al. 1999; Pasca 2004)
- Domain-specific
- Rely on pattern-matching rules manually crafted for each domain

Domain-Independent (Hearst 1992)
- Lexico-syntactic patterns
- ✔ hyponymy relations
- ✖ meronymy relations and other relations

Low recall & High precision (Humphreys et al. 1998)
- only the rules which never generate errors
- Miss many relations
- Lack of unambiguous rules

Pros
- Perform effectively in well-defined domains

Cons
- Hard to enumerate all possible patterns (a lot of work - zillions, impossible, unpractical)
- Domain-dependent
Supervised Approach

- Features-based
Supervised Approach

Classify the relation of each pair of entities extracted in a sentence.

- A set of relation types
- A named entity tagger
- Lots of **Labeled data** (Break into training set, development set and test set)
- Feature representation
- A classifier (Naïve Bayes, MaxEnt, SVM, ...)

*Evaluation: precision, recall...*

6 relations & 17 sub-relations from Automated Content Extraction (ACE)
Supervised Approach - Features (Zhou et al. 2005)

**Lightweight features – require little pre-processing**
- Words: headwords, bag of words, bigrams (between, before or after)
- Entity type: PERSON, ORGANIZATION, FACILITY, LOCATION & Geo-Political Entity/GPE
- Entity level: NAME, NOMIAL & PRONOUN

**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 entities
- Parse tree paths between entities
Supervised Approach - Experiment (Zhou et al. 2005)

Classifier - SVM

<table>
<thead>
<tr>
<th>ACE corpus</th>
<th>Documents</th>
<th>Words</th>
<th>Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set</td>
<td>674 (155)</td>
<td>~ 300k</td>
<td>9683</td>
</tr>
<tr>
<td>Testing set</td>
<td>97</td>
<td>~ 50k</td>
<td>1386</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Features</th>
<th>P</th>
<th>R</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Words</td>
<td>69.2</td>
<td>23.7</td>
<td>35.3</td>
</tr>
<tr>
<td>+Entity Type</td>
<td>67.1</td>
<td>32.1</td>
<td>43.4</td>
</tr>
<tr>
<td>+Mention Level</td>
<td>67.1</td>
<td>33.0</td>
<td>44.2</td>
</tr>
<tr>
<td>+Overlap</td>
<td>57.4</td>
<td>40.9</td>
<td>47.8</td>
</tr>
<tr>
<td>+Chunking</td>
<td>61.5</td>
<td>46.5</td>
<td>53.0</td>
</tr>
<tr>
<td>+Dependency Tree</td>
<td>62.1</td>
<td>47.2</td>
<td>53.6</td>
</tr>
<tr>
<td>+Parse Tree</td>
<td>62.3</td>
<td>47.6</td>
<td>54.0</td>
</tr>
<tr>
<td>+Semantic Resources</td>
<td>63.1</td>
<td>49.5</td>
<td>55.5</td>
</tr>
</tbody>
</table>

<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>68.1</td>
<td>57.1</td>
<td>62.1</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>PART</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>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>General-Staff</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>Management</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>95</td>
<td>60</td>
<td>21</td>
<td>74.1</td>
<td>63.2</td>
<td>68.5</td>
</tr>
<tr>
<td></td>
<td>Parent</td>
<td>29</td>
<td>16</td>
<td>32</td>
<td>33.3</td>
<td>55.2</td>
<td>41.6</td>
</tr>
</tbody>
</table>

Generally Lower occurrence Lower performance

Diverse lexical, syntactic and semantic features from WordNet
Supervised Approach - Pros & Cons

**Pros**

- Could be adapted to a different domain (compared to knowledge-based)
- High accuracy with enough hand-labeled training data and test similar enough to training

**Cons**

- Have to label a large training set (expensive)
- Could not generalize well to different genres
Semi-supervised Approach

- DIPRE
- Snowball
DIPRE (Brin, 1998) – Pattern-Relation duality

Dual Iterative Pattern Relation Expansion

1. start from a small sample
2. grow the target relation
DIPRE (Brin, 1998) - Algorithm

Given a small seed set of (Author, Title) pairs

1. Find all occurrences on the Web
   7-tuple: \( \text{author, title, order, url, prefix, middle, suffix} \)
   Read *The Adventures of Sherlock Holmes* by *Arthur Conan Doyle* online or in your email.
   [Arthur Conan Doyle, *The Adventures of Sherlock Holmes*, 0, (url), Read, by, online or ]

2. Induce patterns from the labeled data
   Group tuples by matching order and middle.
   5-tuple: \( \text{order, urlprefix, prefix, middle, suffix} \)
   urlprefix: prefix, author, middle, title, suffix
   [Sir, *?*, wrote, *?*, in 1892]

3. Apply the patterns to data
   *Sir* Arthur Conan Doyle *wrote* *Speckled Band* in 1892.

4. To get a new set of (Author, Title) pairs, And add to the seed set
   Iterate until list is large enough
DIPRE (Brin, 1998) – limited Experiment

<table>
<thead>
<tr>
<th>Authors</th>
<th>Books</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>
<tr>
<td>H. P. Lovecraft &amp; August Derleth</td>
<td>The Lurker at the Threshold</td>
</tr>
<tr>
<td>H. P. Lovecraft</td>
<td>At the Mountains of Madness and Other Tales of Terror</td>
</tr>
<tr>
<td>H. P. Lovecraft</td>
<td>The Case of Charles Dexter Ward</td>
</tr>
<tr>
<td>H. P. Lovecraft</td>
<td>The Doom That Came to Sarnath and Other Stories</td>
</tr>
</tbody>
</table>

- **5 books**
- **4047 unique pairs**
- **199 occurrences**
- **3 patterns**
- **24 million web pages 147G**

<table>
<thead>
<tr>
<th>URL Pattern</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/Imann/awards/hugos/1984.html</td>
<td>&lt;i&gt;title&lt;/i&gt; by author (</td>
</tr>
<tr>
<td>dolphin.upenn.edu/ducummins/texts/sf-award.htm</td>
<td>author</td>
</tr>
</tbody>
</table>
Snowball (Agichtein & Gravano, 2000)

Similar to DIPRE, to identify (organization, location) relation.

(CMU, Pittsburgh) ... go to CMU campus in Pittsburgh to meet...

5-tuple:
(prefix, organization, middle, location, suffix)

[(w₁, go), (w₂, to), ORG, (w₁, campus), (w₂, in), LOC, (w₁, to), (w₂, meet)]

Named-entity

Match(tupleᵢ, tupleⱼ) = (prefixᵢ, prefixⱼ) + (suffixᵢ, suffixⱼ) + (middleᵢ, middleⱼ)
Semi-supervised Approach - Pros & Cons

Pros
- avoid labeling manually lots of data
- DIPRE does not rely on NLP tools (other languages)

Cons
- Require seeds for each relation (quality of the original set of seeds is important)
- Big problem of semantic drift at each iteration
- Not high precision

DIPRE
1. Favoring long patterns

Snowball
1. Removing unproductive patterns
2. Confidence score on patterns
3. Confidence score on each seed
Distant Supervision
Distant Supervision (Mintz et al. 2009)

Assumption

If two entities participate in a relation, any sentence that contains those two entities might express that relation.

Use a database of relations Freebase to get lots of training examples

1. Instead of using hand-labeled corpus (supervised)
2. Instead of hand-creating a few seed tuples (semi-supervised)

Advantages of supervised approach
1. Leverage rich, reliable hand-created knowledge
2. Relations have canonical names
3. Can use rich features (e.g. syntactic features)

Advantages of unsupervised approach
1. Leverage unlimited amounts of text data
2. Allows for very large number of weak features
3. Not sensitive to training corpus: genre-independent
Distant Supervision (Mintz et al. 2009)

**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)

**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: ???
  - Feature: X attended Y

**Trained relation classifier**

**Predictions!**
- (Henry Ford, Ford Motor Co.)
  - Label: Founder
- (Steve Jobs, Reed College)
  - Label: CollegeAttended

*Picture from Bill MacCartney*
Distant Supervision - Experiment

- Classifier: multi-class logistic regression optimized using L-BFGS with Gaussian regularization (Manning & Klein 2003)
- Parser: MINIPAR (Lin 1998)
- POS tagger: MaxEnt tagger trained on the Penn Treebank (Toutanova et al. 2003)
- NER tagger: Stanford four-class tagger {PER, LOC, ORG, MISC, NONE} (Finkel et al. 2005)
- 3 configurations: lexical features, syntax features, both
- 1.8 million relation instances used for training
- 800,000 Wikipedia articles used for training
- 400,000 different articles used for testing
- Extract relation instances not already in Freebase
Distant Supervision - 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
- Automatic evaluation on 900k instances of 102 Freebase relations.
- Precision for three different feature sets is reported at various recall levels.

Human evaluation

- Performed by evaluators on Amazon Mechanical Turk
- Calculated precision at 100 and 1000 recall levels for the ten most common relations
- At recall of 100 instances, using both feature sets (lexical and syntax) offers the best performance for a majority of the relations.
- At recall of 1000 instances, using syntax features improves performance for a majority of the relations.
Distant Supervision - Further

**Improve mapping (relations in database to texts)**
Improving matching to make data less noisy
Takamatsu et al. 2012

⊄ **One pattern has only one sense**
‘DIRT’ + LDA (Rel-LDA, Rel-LDA1, Type-LDA)
Yao et al. 2011

⊄ **One pair of entities has only one relation**
Multi-Instance Multi-label (MIML)
Hoffmann et al. 2011
MIML + Bayesian (MIML-RE)
Surdeanu et al. 2012
Future Work

To capture long range relations
Sentence level → over sentences & cross documents

Beyond binary relations
Binary relations → n-ary relations

Evaluation of unsupervised approach
Thanks!

Guoxin Gu
Reference


Reference
