Semantic Rule Filtering for Web-Scale Relation Extraction

Andrea Moro$^1$, Hong Li$^2$, Sebastian Krause$^2$, Feiyu Xu$^2$, Roberto Navigli$^1$ and Hans Uszkoreit$^2$
Motivation

- Domain adaptive relation extraction
Motivation

- Domain adaptive relation extraction
  - useful for enriching Knowledge Bases, improving QA, semantic search, ...
Motivation

- Domain adaptive relation extraction
  - useful for enriching Knowledge Bases, improving QA, semantic search, …
- Learn from texts:
Motivation

- Domain adaptive relation extraction
  - useful for enriching Knowledge Bases, improving QA, semantic search, ...

- Learn from texts:

- Semantic relevance of patterns:

  Verb: “meet”
  nsubj PERSON
dobj PERSON

Highly frequent
Motivation

- Domain adaptive relation extraction
  - useful for enriching Knowledge Bases, improving QA, semantic search, ...

- Learn from texts:
  - marriage
  - meet at law court
  - parent—child
  - present movie

- Semantic relevance of patterns:
  - Verb: “meet”
  - nsubj
  - dobj
  - PERSON
  - PERSON

Highly frequent

Low probability of being relevant
Motivation

- Domain adaptive relation extraction
  - useful for enriching Knowledge Bases, improving QA, semantic search, ...

- Learn from texts:

- Semantic relevance of patterns:

  ![Diagram showing examples of verb patterns]

  **Verb: “meet”**
  - nsubj: PERSON
  - dobj: PERSON
  - Highly frequent
  - Low probability of being relevant

  **Verb: “divorce”**
  - nsubj: PERSON
  - dobj: PERSON
  - Might be infrequent
Motivation

- Domain adaptive relation extraction
  - useful for enriching Knowledge Bases, improving QA, semantic search, ...

- Learn from texts:
  - marriage
  - meet at law court
  - parent–child
  - present movie

- Semantic relevance of patterns:
  - **Verb: “meet”**
    - nsubj: PERSON
    - dobj: PERSON
    - Highly frequent
    - Low probability of being relevant
  
  - **Verb: “divorce”**
    - nsubj: PERSON
    - dobj: PERSON
    - Might be infrequent
    - High probability of being relevant
Our Approach

Relation-Specific Lexical Semantic Graphs
Our Approach

World Wide Web

Relation-Specific Lexical Semantic Graphs

Candidate RE Patterns

Unsupervised Classification

High-quality RE Patterns
BabelNet: Concepts, Entities, Semantic Relations

- 5.5 million of concepts and entities
- 140 million of semantic relation instances
- Ready to go API for word sense disambiguation
- babelnet.org

Learning Relation-specific Semantic Graphs

automatically learned RE rules and their source sentences

BabelNet

Word Sense Disambiguation

wife

husband

marriage

marry

divorce

divorce
Building Relation-specific Semantic Graphs

1. **Input**: rules automatically learned from sentences;
Building Relation-specific Semantic Graphs

1. **Input**: rules automatically learned from sentences;

2. We apply **WSD** on the content words of the rules using the source sentences as context;
Building Relation-specific Semantic Graphs

1. **Input**: rules automatically learned from sentences;

2. We apply **WSD** on the content words of the rules using the source sentences as context;

3. **Output**: a frequency distribution over synsets
Building Relation-specific Semantic Graphs

1. **Input**: rules automatically learned from sentences;

2. We apply **WSD** on the content words of the rules using the source sentences as context;

3. **Output**: a frequency distribution over synsets

The 3 most frequent synsets for the *marriage* relation are: *marry\textsubscript{v}^1*, *wife\textsubscript{n}^1*, *husband\textsubscript{n}^1*
The Relation-specific Semantic Graph

1. The top-k most frequent synsets are our **core synsets**;

   \( \text{marry}^1_v, \text{wife}^1_n, \text{husband}^1_n \)
The Relation-specific Semantic Graph

1. The top-k most frequent synsets are our core synsets; 

   \textit{marry}^1_v, \textit{wife}^1_n, \textit{husband}^1_n

2. We augment them with all the synsets that:
   1. have frequency > 1
   2. are connected to at least one of the core synsets by a semantic relation in BabelNet
The Relation-specific Semantic Graph

1. The top-k most frequent synsets are our core synsets:
   \[\text{marry}_v^1, \text{wife}_n^1, \text{husband}_n^1\]

2. We augment them with all the synsets that:
   1. have frequency >1
   2. are connected to at least one of the core synsets by a semantic relation in BabelNet

\[\text{marriage}_n^1\]
The Relation-specific Semantic Graph

1. The top-k most frequent synsets are our core synsets:
   \[ \text{marry}_v^1, \text{wife}_n^1, \text{husband}_n^1 \]

2. We augment them with all the synsets that:
   1. have frequency >1
   2. are connected to at least one of the core synsets by a semantic relation in BabelNet
   \[ \text{marriage}_n^1 \]

3. The **semantic graph** is the induced subgraph of BabelNet on our augmented core synsets
The Relation-specific Semantic Graph

Marriage

wife\textsubscript{1} \rightsquigarrow husband\textsubscript{1} 

marry\textsubscript{1}
The Relation-specific Semantic Graph

Marriage

- wife\textsubscript{1n}
- husband\textsubscript{1n}
- marriage\textsubscript{1n}
- marry\textsubscript{1v}
- divorce\textsubscript{1n}
- divorce\textsubscript{2v}
Semantic Rule Filtering

Rule is accepted \(\iff\) Rule contains a lexicalization of the semantic graph’s synsets

For instance, we filter out:

- \(\text{PERSON} \rightarrow \text{Verb: “lose”} \rightarrow \text{Prep: “to”} \rightarrow \text{PERSON}\)
- \(\text{PERSON} \rightarrow \text{Verb: “meet”} \rightarrow \text{PERSON}\)

While we accept:

- \(\text{PERSON} \rightarrow \text{Noun: “widow”} \rightarrow \text{Prep: “of”} \rightarrow \text{PERSON}\)
- \(\text{PERSON} \rightarrow \text{Verb: “marry”} \rightarrow \text{PERSON}\)
Experiment Setup

- **Intrinsic evaluation**
  - Evaluating the quality of the filtered rules
  - Gold-standard of 100 manually evaluated rules p.R.

- **Extrinsic evaluation**
  - Filter impact on recall and precision of RE task
  - English Gigaword corpus
  - *Precision*: Evaluating a sample (1000 per relation)
  - *Recall*: ~900 Freebase mentions in Gigaword discovered by Web-DARE baseline and NELL patterns

- Experimenting with different values of $k$ (top-$k$ most frequent senses)
Intrinsic Evaluation (Precision and Recall)
Experiment Setup

- **Intrinsic evaluation**
  - Evaluating the quality of the filtered rules
  - Gold-standard of 100 manually evaluated rules p.R.

- **Extrinsic evaluation**
  - Filter impact on recall and precision of RE task
  - English Gigaword corpus
  - *Precision*: Evaluating a sample (1000 per relation)
  - *Recall*: ~900 Freebase mentions in Gigaword discovered by Web-DARE baseline and NELL patterns

- Experimenting with different values of k (top-k most frequent senses)
Extrinsic Evaluation

- Average RE performance (over 7 relations)
Extrinsic Evaluation: NELL Patterns

- Also tested: Filtering on NELL patterns (Carlson et al., 2010)
- **PERSON** “and husband” **PERSON**
- Average RE performance (over 7 relations)
Results and Conclusion

- A novel method for unsupervised building of relation-specific lexical semantic knowledge

- A new semantic rule classifier that improves the precision of large-scale RE

- Relation-specific lexical semantic graphs improve RE performance
  - Web-DARE: precision +45%, recall -15%
  - NELL rules: precision +25%, recall -6%

- A comparison of WordNet and BabelNet from IE perspective
  - BabelNet achieves better recall and F-score than WordNet both in rule filtering and in RE
Thank you! Questions?
Motivation

Relation Extraction good for:

- Enriching Knowledge Bases;
- Improving Question Answering systems;
- Improve semantic search;
- ...

Relation Extraction: Pattern Application

- recognize entities
- grammatical analysis
Pattern Learning: Web-DARE

large number of RE rules are automatically learned by using Freebase as seed knowledge and Web as training corpus

Goal:

- covering most linguistic variants for expressing a relation
- thus solving the notorious long-tail problem of real-world NL applications

Sebastian Krause, Hong Li, Hans Uszkoreit, Feiyu Xu. 2012.

- rules learned for 39 relations
- three domains: business, awards and people
- 2.8 million relation instances retrieved from Freebase to be used as seed
- 20 million web documents as training corpus

However there is a problem of precision and so we need filtering techniques!
Pattern Learning: Web-DARE

- **Queries**
- **Facts**
- **Web pages**
- **Filtered Patterns**
- **Text Examples**
- **Patterns**

Diagram showing the process of pattern learning with Web-DARE, involving queries, facts, web pages, filtered patterns, and text examples.
Pattern Filtering Approaches

- Simple filters
  - Frequency?
  - # arguments?
  - Source sentence?

- More elaborate: e.g., mutual-exclusiveness of relations
  - Siblings
  - Parent-child
  - Marriage
  - Romantic
Other approaches to rule Filtering

- Supervised approaches require a lot of training data

- Feature-based approaches only marginally use Semantic features and consequently obtain only slight improvements
Intrinsic Evaluation (F-score)
Marriage Relation

- Junk patterns in *marriage*:
  - Frequent, but semantically not associated (*meet, know, …*)
  - Easy to filter out with WordNet/BabelNet
**Award Honor Relation**

- False positives in *award honor*:
  - Extraction suffers from positive mentions which are embedded into speculation, negation, ...
  - Pattern filtering not effective here
Experiments

- Impact of utilized lexicalized semantic network (avg. over 12 relations)

<table>
<thead>
<tr>
<th>k</th>
<th>Precision (%)</th>
<th>Rel. Recall (%)</th>
<th>F1 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WN</td>
<td>BN</td>
<td>WN</td>
</tr>
<tr>
<td>(Basel.)</td>
<td>29.58</td>
<td>91.36</td>
<td>44.69</td>
</tr>
<tr>
<td>15</td>
<td>42.18</td>
<td>40.91</td>
<td>59.92</td>
</tr>
<tr>
<td>10</td>
<td>45.12</td>
<td>43.84</td>
<td>56.27</td>
</tr>
<tr>
<td>5</td>
<td>47.79</td>
<td>52.19</td>
<td>50.99</td>
</tr>
<tr>
<td>4</td>
<td>48.75</td>
<td>61.17</td>
<td>49.09</td>
</tr>
<tr>
<td>2</td>
<td><strong>64.47</strong></td>
<td><strong>67.33</strong></td>
<td><strong>38.47</strong></td>
</tr>
<tr>
<td>1</td>
<td>57.57</td>
<td>68.18</td>
<td>20.02</td>
</tr>
</tbody>
</table>

[Diagram showing the relationships between PERSON, Noun: “widow”, Prep: “of”, and PERSON]
Error Analysis

- False Positives:

<table>
<thead>
<tr>
<th></th>
<th>#</th>
<th>P</th>
<th>NER/Coref</th>
<th>Parsing</th>
<th>Wrong Relation</th>
<th>Modality</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>acquisition</td>
<td>13</td>
<td>88.5</td>
<td>15.38 %</td>
<td>7.7 %</td>
<td>61.54 %</td>
<td>15.4 %</td>
<td>0.00 %</td>
</tr>
<tr>
<td>marriage</td>
<td>2</td>
<td>97.8</td>
<td>0.00 %</td>
<td>100.00 %</td>
<td>0.00 %</td>
<td>0.00 %</td>
<td>0.00 %</td>
</tr>
<tr>
<td>birth</td>
<td>24</td>
<td>92.9</td>
<td>4.17 %</td>
<td>16.67 %</td>
<td>79.17 %</td>
<td>0.00 %</td>
<td>0.00 %</td>
</tr>
<tr>
<td>death</td>
<td>32</td>
<td>82.7</td>
<td>9.38 %</td>
<td>6.25 %</td>
<td>90.63 %</td>
<td>0.00 %</td>
<td>3.13 %</td>
</tr>
<tr>
<td>parent–child</td>
<td>3</td>
<td>95.3</td>
<td>0.00 %</td>
<td>66.70 %</td>
<td>0.00 %</td>
<td>33.30 %</td>
<td>0.00 %</td>
</tr>
<tr>
<td>siblings</td>
<td>9</td>
<td>79.0</td>
<td>0.00 %</td>
<td>66.67 %</td>
<td>11.11 %</td>
<td>11.11 %</td>
<td>33.33 %</td>
</tr>
</tbody>
</table>

- Example:
  - In the months after Laci Peterson disappeared from her Modesto home on Christmas Eve 2002, the Peterson case garnered more airtime ...
  - Pryor suffered a heart attack at his home in the San Fernando Valley early Saturday morning.
Error Analysis

■ False Negatives:

<table>
<thead>
<tr>
<th></th>
<th>#</th>
<th>Recall</th>
<th>Semantic Understanding</th>
<th>NER/Coref</th>
<th>Parsing</th>
<th>Rule missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>acquisition</td>
<td>197</td>
<td>32.06</td>
<td>11.67%</td>
<td>23.33%</td>
<td>8.33%</td>
<td>15.00%</td>
</tr>
<tr>
<td>marriage</td>
<td>256</td>
<td>38.6</td>
<td>11.71%</td>
<td>48.43%</td>
<td>16.79%</td>
<td>8.98%</td>
</tr>
<tr>
<td>parent–child</td>
<td>147</td>
<td>38.32</td>
<td>19.04%</td>
<td>34.69%</td>
<td>28.57%</td>
<td>9.52%</td>
</tr>
</tbody>
</table>

■ Example:

- Atom Entertainment joins former rival IFilm in Viacom, which has also acquired such online businesses as Neopets, XFire and GameTrailers.com.
DARE Rule for n-ary Relations

```
person ← nsubj marry → person
dobj

location ← prep in prep on pobj → date
```
Experimental Setup

- Intrinsic evaluation
  - Gold-standard rule set

- Extrinsic evaluation
  - Application of filtered rules to English Gigaword corpus
  - Precision: Manual investigation of samples (1000 facts per relation)
  - Relative Recall:
    - Gold-standard: all Freebase mentions in corpus discovered by Web-DARE baseline and NELL patterns
    - manual validation of extraction coverage of Freebase mentions of various filtering strategies

<table>
<thead>
<tr>
<th>Intrinsic E.</th>
<th>+ 535 rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>- 665 rules</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Extrinsic E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documents</td>
</tr>
<tr>
<td>Sentences w/ ≥ 2 NEs</td>
</tr>
<tr>
<td>Patterns</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Extracted Mentions</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Data Set

- rules learned for 39 relations
- three domains: business, awards and people
- 2.8 million relation instances retrieved from Freebase to be used as seed
- 20 million web documents as training corpus