Extracting Context-Rich Entailment Rules from Wikipedia Revision History

Elena Cabrio ¹  Bernardo Magnini ²  Angelina Ivanova ³

¹ INRIA - Sophia Antipolis, France
² FBK - Trento, Italy
³ University of Oslo, Norway
Outline

1 Textual Entailment

2 Entailment Rules

3 Automatic rules acquisition from Wiki revisions: experimental setting

4 Results and discussion

5 Conclusion and future work
Textual Entailment

- Generic framework for capturing major semantic inference needs in NLP applications [Dagan and Glickman, 2004]
- Relation between a textual fragment ($T$) and a language expression ($H$). $T \Rightarrow H$ if the meaning of $H$ can be inferred from the meaning of $T$, as interpreted by a typical language user

T: Dr. Thomas Bond established a hospital in Philadelphia for the reception and cure of poor sick persons.

H: Dr. Bond created a medical institution for sick people.
Crucial role of knowledge to support entailment inferences

T: Dr. Thomas Bond established a hospital in Philadelphia for the reception and cure of poor sick persons.

H: Dr. Bond created a medical institution for sick people.
Crucial role of knowledge to support entailment inferences

T: Dr. Thomas Bond established a *hospital* in Philadelphia for the reception and cure of poor sick persons.

H: Dr. Bond created a *medical institution* for sick people.
Crucial role of knowledge to support entailment inferences

LHS: hospital $\Rightarrow$ RHS: medical institution
probability: 1

T: Dr. Thomas Bond established a hospital in Philadelphia for the reception and cure of poor sick persons.

H: Dr. Bond created a medical institution for sick people.
Crucial role of knowledge to support entailment inferences

LHS: hospital $\Rightarrow$ RHS: medical institution
probability: 1

T: Dr. Thomas Bond established a hospital in Philadelphia for the reception and cure of poor sick persons.

H: Dr. Bond created a medical institution for sick people.

LHS: $X$ establish $Y$ $\Rightarrow$ RHS: $X$ create $Y$
probability: 0.8

Entailment rules

Directional relation between two sides of a pattern, corresponding to text fragments with variables.

The left-hand side (LHS) of the pattern entails the right-hand side (RHS) of the same pattern under the same variable instantiation. [Szpektor et al., 2007]

- Typically acquired from structured (e.g. WordNet) or unstructured sources (e.g. DIRT, Microsoft Paraphrase collection, TEASE)

- LIMITATIONS:
  - Lack of an adequate representation of linguistic context
  - Lack of directionality
Entailment rules

Directional relation between two sides of a pattern, corresponding to text fragments with variables.

The left-hand side (LHS) of the pattern entails the right-hand side (RHS) of the same pattern under the same variable instantiation. [Szpektor et al., 2007]

• Typically acquired from structured (e.g. WordNet) or unstructured sources (e.g. DIRT, Microsoft Paraphrase collection, TEASE)

• LIMITATIONS:
  ■ Lack of an adequate representation of linguistic context
  ■ Lack of directionality
Our proposal: rules acquisition from Wikipedia revision history

Cycling in Amsterdam
From Wikipedia, the free encyclopedia

Amsterdam is one of the most bicycle-friendly large cities in the world. In Amsterdam, over 60% of trips are made by bike in the inner city and 38% of trips are made by bike overall in the greater city area. The city is one of the most important centres of bicycle culture worldwide with world-class facilities for cyclists such as bike paths and bike racks, and several guarded bike storage garages (Fietsenstalling) which can be used for a nominal fee. In 2006, there were about 465,000 bicycles in Amsterdam.

Bicycles are used by all socio-economic groups due to their convenience. Amsterdam's small size, the 400 km of bike paths, the flat terrain, and the arguable inconvenience of driving an automobile. Each bike path (Fietspad) is coloured brown, in order to differentiate it from a footpath.

Bike riders in Amsterdam.

Cycling in Amsterdam
From Wikipedia, the free encyclopedia

Amsterdam is one of the most bicycle-friendly large cities in the world. In Amsterdam, over 60% of trips are made by bike in the inner city and 38% of trips are made by bike overall in the greater city area. The city is one of the most important centres of bicycle culture worldwide with world-class facilities for cyclists such as bike paths and bike racks, and several guarded bike storage garages (Fietsenstalling) which can be used for a nominal fee. In 2006, there were about 465,000 bicycles in Amsterdam.

Bicycles are used by all socio-economic groups because of their convenience. Amsterdam's small size, the 400 km of bike paths, the flat terrain, and the arguable inconvenience of driving an automobile. Each bike path (Fietspad) is coloured brown, in order to differentiate it from a footpath.

Bike riders in Amsterdam.
Our proposal: rules acquisition from Wikipedia revision history

Cycling in Amsterdam
From Wikipedia, the free encyclopedia

Amsterdam is one of the most bicycle-friendly large cities in the world. In Amsterdam over 60% of trips are made by bike in the inner city and 38% of trips are made by bike overall in the greater city area. The city is one of the most important centres of bicycle culture worldwide with world-class facilities for cyclists such as bike paths and bike racks, and several guarded bike storage garages (Fietsenstalling) which can be used for a nominal fee. In 2006, there were about 465,000 bicycles in Amsterdam.

Bicycles are used by all socio-economic groups due to their convenience. Amsterdam's small size, the 400 km of bike paths, the flat terrain, and the arguable inconvenience of driving an automobile. Each bike path (Fietstpad) is coloured brown, in order to differentiate it from a footpath.
Benefits of the approach

- **Entailment judgement:** about 95% of revisions preserves entailment [Zanzotto et al., 2010]
- **Monothematic T-H pair:** revision pairs with minimal difference [Bentivogli et al., 2010]
- **Directionality:** $\text{Wiki}_{\text{revised}} \Rightarrow \text{Wiki}_{\text{original}}$
- **Context of the rule:** set of morpho-syntactic constraints over the application of the rule in a specific T-H pair
Wiki-rules acquisition: a 4-step methodology

Step 1: Preprocessing Wikipedia dumps
Step 2: Extraction of entailment pairs
Step 3: Extraction of entailment rules
Step 4: Rules expansion with minimal context
Experiments: step 1 and 2

- **Preprocessing Wikipedia dumps**
  - 1 540 870 documents: are not identical and are present in both versions

- **Extraction of entailment pairs**
  - Documents are sentence splitted, and sentences are aligned
  - To measure the similarity between the sentences: *Position Independent Word Error Rate (PER)* [Tillman et al., 1997]
  - Different thresholds are set to cluster pairs into different sets

<table>
<thead>
<tr>
<th>Set</th>
<th># pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>set a: containment</td>
<td>1 547 415</td>
</tr>
<tr>
<td>set b: minor editing</td>
<td>1 053 114</td>
</tr>
<tr>
<td>set c: major editing</td>
<td>2 566 364</td>
</tr>
</tbody>
</table>

Experiments: step 3

• Extraction of entailment rules
  - Pairs of set $b$ (atomic pairs) are parsed and chunked
  - Algorithm `RULE_EXTRACTOR` compares the chunks in $T$ and $H$ to extract the ones that differ in $T$ and $H$

  $T$: Bicycles are used by all socio-economic groups because of their convenience [...].

  $H$: Bicycles are used by all socio-economic groups due to their convenience [...].

Entailment rule: `causative_1`
Pattern: `because of ⇒ due to`
Experiments: step 3

- Extraction of entailment rules
  - Pairs of set $b$ (atomic pairs) are parsed and chunked
  - Algorithm RULE_EXTRACTOR compares the chunks in $T$ and $H$ to extract the ones that differ in $T$ and $H$

$T$: Bicycles are used by all socio-economic groups because of their convenience [...].

$H$: Bicycles are used by all socio-economic groups due to their convenience [...].

Entailment rule: causative_1
Pattern: because of $\Rightarrow$ due to
Experiments: step 4

• Rules expansion with minimal context
  - Algorithm **EXPAND_RULE** to add minimal context to each rule

```
ruleid="23" docid="844" pairid="15"

(PP(RB because)(IN of)(NP(PPR)(NN))) ⇒ (ADJP(JJ due)(PP(TO to)(NP(PPR)(NN))))
```

Experiments: step 4

- Rules expansion with minimal context
  - Algorithm `EXPAND_RULE` to add minimal context to each rule

(ruleid="23" docid="844" pairid="15"

\[(PP(RB \because)(IN \, of)(NP(PR(P\$)(NN)))) \Rightarrow (ADJP(JJ \, due)(PP(TO \, to)(NP(PR(P\$)(NN)))))\]

Results

- Two large-scale experiments focusing on entailment rules for:
  - causality (seed word: because)
  - temporal expressions (seed word: before)

- Statistics on the data sets of entailment rules:

<table>
<thead>
<tr>
<th></th>
<th>causality</th>
<th>temporal exp.</th>
</tr>
</thead>
<tbody>
<tr>
<td># rules before gen.</td>
<td>1671</td>
<td>813</td>
</tr>
<tr>
<td># rules after gen.</td>
<td>977</td>
<td>457</td>
</tr>
<tr>
<td>rules frequency &gt; 2</td>
<td>66</td>
<td>27</td>
</tr>
</tbody>
</table>

- During the generalization phase, we filtered out
  - antonyms (using WordNet)
  - awkward inconsistencies due to algorithm mistakes
Sample of extracted rules

<table>
<thead>
<tr>
<th>Causality (because)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(PP(RB because)(IN of)(NP(JJ)(NNS))) ⇒ (ADJP(JJ due)(PP(TO to)(NP(JJ)(NNS))))</td>
<td>e.g.: because of contractual conflicts ⇒ due to contractual conflicts</td>
</tr>
<tr>
<td>(SBAR(IN because)(S)) ⇒ (VP(PP(IN on)(NP(DT the)(NNS grounds)))(SBAR(IN that)(S))</td>
<td>e.g.: because it penalized people ⇒ on the grounds that it penalized people</td>
</tr>
<tr>
<td>(PP(RB because)(IN of)(NP(DT)(NN))) ⇒ (PP(IN as)(NP(NP(DT a)(NN result))(PP(IN of)(NP(DT)(NN))))))</td>
<td>e.g.: because of an investigation ⇒ as a result of an investigation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Temporal Exp. (before)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(SBAR(IN before)(S)) ⇒ (ADVP(RB prior)(PP(TO to)(S))</td>
<td>e.g.: before recording them ⇒ prior to recording them</td>
</tr>
<tr>
<td>(ADVP(RB prior)(PP(TO to)(NP(DT)(NN)))) ⇒ (SBAR(IN before)(NP(DT)(NN))))</td>
<td>e.g.: prior to the crash ⇒ before the crash</td>
</tr>
<tr>
<td>(SBAR(IN until)(NP(CD))) ⇒ (SBAR(IN before)(NP(CD))))</td>
<td>e.g.: until 1819 ⇒ before 1819</td>
</tr>
</tbody>
</table>
Evaluation

- **Two annotators** with skills in linguistics
- Five possible values:
  - $entailment=\text{yes}$
  - $entailment=\text{more-phenomena}$
  - $entailment=\text{unknown}$
  - $entailment=\text{unknown:reverse-entailment}$
  - $entailment=\text{error}$
- Manual analysis of a sample of 100 rules
- **Inter-annotator agreement:** 80% causality, 77% temp. expr.
- Rules accuracy:

<table>
<thead>
<tr>
<th></th>
<th>YES</th>
<th>+PHEN</th>
<th>UNK</th>
<th>REV</th>
<th>ERR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>caus.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all</td>
<td>67</td>
<td>2</td>
<td>13</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>fr$\geq$2</td>
<td>80.3</td>
<td>0</td>
<td>16.7</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td><strong>temp.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all</td>
<td>36</td>
<td>6</td>
<td>23</td>
<td>7</td>
<td>28</td>
</tr>
<tr>
<td>fr$\geq$2</td>
<td>52</td>
<td>3.7</td>
<td>37</td>
<td>7.3</td>
<td>0</td>
</tr>
</tbody>
</table>

Evaluation

- **Two annotators** with skills in linguistics
- Five possible values:
  - \( \text{entailment}=\text{yes} \)
  - \( \text{entailment}=\text{more-phenomena} \)
  - \( \text{entailment}=\text{unknown} \)
  - \( \text{entailment}=\text{unknown:reverse-entailment} \)
  - \( \text{entailment}=\text{error} \)
- Manual analysis of a sample of 100 rules
- **Inter-annotator agreement:** 80% causality, 77% temp. expr.
- **Rules accuracy:**

<table>
<thead>
<tr>
<th></th>
<th>YES</th>
<th>+PHEN</th>
<th>UNK</th>
<th>REV</th>
<th>ERR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>caus.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all</td>
<td>67</td>
<td>2</td>
<td>13</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>fr≥2</td>
<td>80</td>
<td>16.7</td>
<td>1.5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>temp.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all</td>
<td>36</td>
<td>6</td>
<td>7</td>
<td>23</td>
<td>28</td>
</tr>
<tr>
<td>fr≥2</td>
<td>52</td>
<td>3.7</td>
<td>7.3</td>
<td>37</td>
<td>0</td>
</tr>
</tbody>
</table>

Error analysis

- **Noisy data**: typos, spam (19% of the rules)
  - spell-checker or dictionary-based filters should be applied
- Use the frequency of the rules in the data, to estimate their reliability
- **entailment=unknown**: the editing of *Wiki10* concerned a change in semantics (*unknown* judgement)
  - *before 1990* ⇒ *1893* for temporal expressions
  - *when x produced* ⇒ *because x produced* for causality
- Two phenomena collapsed on consecutive tokens (about 10% of the rules)
  - *because of the divorce settlement cost* ⇒ *due to the cost of his divorces settlement*
Conclusions

• Methodology for the automatic acquisition of entailment rules from Wikipedia revision pairs

• Main benefits:
  ■ large-scale acquisition
  ■ new coverage
  ■ quality

• Experimental acquisition, on causality (seed because) and temporal expressions (seed before)
  ■ the resource includes, respectively, 977 and 457 rules
  ■ resource can be easily extended and periodically updated

Future work

- Improve **filtering techniques**
- Refine the **rule expansion algorithm**
- Carry out more extended **evaluation**:
  - applying the instance-based approach [Szpektor 2007] on PTB data
  - integrating the extracted rules into existing TE systems

Thanks!
Future work

- Improve filtering techniques
- Refine the rule expansion algorithm
- Carry out more extended evaluation:
  - applying the instance-based approach [Szpektor 2007] on PTB data
  - integrating the extracted rules into existing TE systems

Thanks!