Recognizing Named Entities using Automatically Extracted Transduction Rules

D. Nouvel, J.Y. Antoine, N. Friburger, A. Soulet

Université François Rabelais Tours
Laboratoire d'Informatique
Equipe BDTLN
Named Entity Recognition

- Named Entity Recognition (NER) task:
  - **Proper Nouns**: person, location, organization (movie, brand...)
  - **Definite Descriptions**: time expression, amount, function (...)

- Named Entities Recognition (NER) by:
  - **Detecting** / delimiting NEs (determining *frontiers, boundaries*)
  - **Categorizing** / classifying / assigning a type to detected NEs
    ⇒ Finding *markers* as NEs boundaries

**Example**

The `<prod>` iPhone 4 `</prod>` was announced during the `<time>` 7th of June, 2010 `</time>` keynote by `<pers>` Steve Jobs `</pers>`, `<fonc>` chief executive officer `</fonc>` of the `<org>` Apple `</org>` company.
Outline

1. General Context
2. Mining Patterns from Corpus
3. NER using Informative Rules
4. Experimental Results
5. Conclusion
Context of work

- Main approaches of NER:
  - **Knowledge-based** systems (difficult to attain good recall)
  - **Machine learning** systems (generally not easy to customize)
  ⇒ We try to find a common ground for combining/hybriding systems

- Existing system: **CasEN** [Fri06] (transducer/rule-based system)

- Available corpus: **Ester2** [GGC09], corpus of transcription of French radio broadcasts annotated in NEs:

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Tokens</th>
<th>Sentences</th>
<th>NEs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ester2-corr</td>
<td>40 167</td>
<td>1 300</td>
<td>2 798</td>
</tr>
<tr>
<td>Ester2-held</td>
<td>48 143</td>
<td>1 683</td>
<td>3 074</td>
</tr>
</tbody>
</table>

**Table:** Characteristics of Ester2 corpora

⇒ **Our objective:** from Ester2 corpus (as train), *mine pattern* and find *informative rules* that may *enhance CasEN* for NER
Data Flow for NER Learning and Evaluating

- **Learning Corpus (annotated texts)**
  - Mining [AS95]
  - Patterns
  - Filtering
  - Annotation (MaxEnt)
  - Rules
  - **Test Corpus**
  - **Annotated Corpus**

Nouvel et al. (François Rabelais Tours)
Outline

1. General Context

2. Mining Patterns from Corpus

3. NER using Informative Rules

4. Experimental Results

5. Conclusion
Extracting Patterns

- Finding **rules** that help **detecting** and **categorizing** simultaneously by determining **markers** of NEs
  - he *flies to* Poznan → he flies to `<loc>` Poznan `</loc>`
  - *president* Obama → president `<pers>` Obama `</pers>`
  - the *benefits* of Apple → the benefits of `<org>` Apple `</org>`

- **Preprocessings** : tokens, lemmas, POS-tagging (TreeTagger)
  - Regular tokens : we only keep the lemma (generalized patterns)
  - Proper Nouns (PN), we only keep POS (avoids overfitting)

- **Pattern Mining** considerations :
  - Exhaustively looking for patterns on pre-annotated corpus
  - Extracting and filtering patterns correlated to NEs markers
  - Apply patterns on unseen (test) corpus
Building hierarchy of items

- DET
  - the
  - a
  - this

- CN
  - head
  - president
  - officer
  - · · ·

- PN
  - Apple
  - Poznan
  - · · ·

presidents
From Corpus to Patterns: concrete example

Corpus pre-annotated sentence

▸ (…) As he *travels to* Poznan by plane, he thought (…)  
▸ (…) , this time, we *come to* Barcelona with (…)

Extracted Patterns
Filtering Patterns as Informative Rules

Transduction Rule

- A Transduction Rule is a morpho-syntactic pattern (relying on the POS-tagging hierarchy) containing NEs markers for which are defined the standard parameters in pattern mining:
  - **Support**: number of occurrences in corpus
  - **Confidence**: in what proportion pattern appears with its markers

Informative Transduction Rule

- By exhaustively mining corpus, we obtain a very large set of rules
  ⇒ We need to **filter out** rules
  ⇒ For two rules which are generalization one of each other, we keep:
    - The most **specific** one in terms of POS-tagging hierarchy
    - The most **informative** according to markers
Outline

1. General Context

2. Mining Patterns from Corpus

3. NER using Informative Rules

4. Experimental Results

5. Conclusion
Probability model

- Many rules are triggered at a given position
- Define a random variable to define probability of markers

\[ P(M_i = m_{ji}) \]

- Annotation probability for a sentence (assumption: markers are independant):

\[ P(M_1 = m_{j_1}, M_2 = m_{j_2}, \ldots, M_n = m_{j_n}) \approx \prod_{i=1}^{n} P(M_i = m_{ji}) \]

- Probability learned by Maximum Entropy modeling
- Use dynamic programming to search annotation (XML-like / flat)
Dynamic programming
Outline

1. General Context
2. Mining Patterns from Corpus
3. NER using Informative Rules
4. Experimental Results
5. Conclusion
Ester2 Corpus

Pattern extraction results over Ester2-Corr (40K tokens, 3K NEs)

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Sup.</th>
<th>Conf.</th>
<th>Rules</th>
<th>Inf. Rules</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ester2-corr</td>
<td>10</td>
<td>.5</td>
<td>2 270</td>
<td>1 119</td>
<td>2.03</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.5</td>
<td>28 047</td>
<td>3 673</td>
<td>7.63</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.3</td>
<td>458 875</td>
<td>12 653</td>
<td>36.27</td>
</tr>
</tbody>
</table>

Table: Extraction over Ester2 corpus at support and confidence thresholds

Interpretation

- Number of patterns is **very large** when support / confidence thresholds are lowered
- Filtering pattern is effective and allows to keep a **reasonable** number of rules
Predicting Markers

<table>
<thead>
<tr>
<th>Actual markers</th>
<th>Predicted markers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>tot</td>
</tr>
<tr>
<td>0</td>
<td>27803</td>
</tr>
<tr>
<td>&lt;pers&gt;</td>
<td>583</td>
</tr>
<tr>
<td>&lt;/pers&gt;</td>
<td>592</td>
</tr>
<tr>
<td>&lt;loc&gt;</td>
<td>700</td>
</tr>
<tr>
<td>&lt;/loc&gt;</td>
<td>698</td>
</tr>
<tr>
<td>&lt;org&gt;</td>
<td>448</td>
</tr>
<tr>
<td>&lt;/org&gt;</td>
<td>443</td>
</tr>
<tr>
<td>&lt;fonc&gt;</td>
<td>225</td>
</tr>
<tr>
<td>&lt;/fonc&gt;</td>
<td>219</td>
</tr>
<tr>
<td>prec.</td>
<td>0.94</td>
</tr>
</tbody>
</table>

**Table**: Confusion matrix between rule markers using a MaxEnt classifier

**Interpretation**
- Great ambiguity org/pers and org/loc (known problem)
- Beginning of a NE is not necessarily easier to find (cf pers, loc)
Predictions NEs

**Figure**: Evaluating (SER, to be minimized) NER annotations

**Interpretation**

- MaxEnt **accurately weights rules** (even less frequent/confident)
Hybridizing Symbolic and Mining Systems

**TABLE**: Using informative rules to enhance a symbolic system

<table>
<thead>
<tr>
<th></th>
<th>Ins.</th>
<th>Del.</th>
<th>Typ.</th>
<th>Ext.</th>
<th>SER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symbolic</td>
<td>43</td>
<td>348</td>
<td>171</td>
<td>257</td>
<td>29.0</td>
</tr>
<tr>
<td>fonc</td>
<td>0</td>
<td>-1</td>
<td>+1</td>
<td>0</td>
<td>28.8</td>
</tr>
<tr>
<td>loc</td>
<td>+4</td>
<td>-15</td>
<td>+3</td>
<td>+1</td>
<td>16.8</td>
</tr>
<tr>
<td>org</td>
<td>0</td>
<td>-13</td>
<td>+11</td>
<td>0</td>
<td>52.8</td>
</tr>
<tr>
<td>pers</td>
<td>+1</td>
<td>-20</td>
<td>0</td>
<td>+8</td>
<td>15.3</td>
</tr>
<tr>
<td>time</td>
<td>0</td>
<td>-2</td>
<td>0</td>
<td>0</td>
<td>24.6</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>+5</td>
<td>-51</td>
<td>+19</td>
<td>+8</td>
<td><strong>-1.3</strong></td>
</tr>
<tr>
<td>Coupled</td>
<td>48</td>
<td>297</td>
<td>190</td>
<td>265</td>
<td>27.7</td>
</tr>
</tbody>
</table>

Interpretation

- Coupling systems improves system with generic rules
  - *from* `<pers>` PN PN
  - *to* `<loc>` PN
  - *for* `<time>` / years </time> ("for a few years")
Outline

1. General Context
2. Mining Patterns from Corpus
3. NER using Informative Rules
4. Experimental Results
5. Conclusion
Conclusion

Contributions

- **Extracting** rules using a *morpho-syntactic* hierarchy
- Filtering *specific* and *informative* patterns as *rules*
- Using patterns to **annotate** a texte (Named Entities)
- Hybriding systems

Further investigations

- Better **filtering** patterns to be integrated in the *knowledge base*?
- How to **enrich** patterns (syntax, semantics, anaphora)
- Assess performance with other models to **predict markers**
- Involved in NER task of project Etape (French National Research Agency, ANR)
Thank you

Rakesh Agrawal and Ramakrishnan Srikant. 
Mining sequential patterns. 

Nathalie Friburger.
Linguistique et reconnaissance automatique des noms propres. 

Sylvain Galliano, Guillaume Gravier, and Laura Chaubard.
The ester 2 evaluation campaign for the rich transcription of french radio broadcasts. 