ROBUST TREE-STRUCTURED NAMED ENTITIES RECOGNITION FROM SPEECH
Christian Raymond
INSA de Rennes, IRISA, France
christian.raymond@irisa.fr

INTRODUCTION
An approach to build a named entities recognition system:
- efficient on speech
- robust to noise
- with tree structured outputs

Main idea: avoid cascade approach
- make node hypothesis with standard CRF
- retrieve the tree-structure from nodes

PROBLEM AND SOLUTION PROPOSED

- Higher order CRF models:
  - direct tree modification
  - immutability
- First-order Conditional Random Field
  - efficient
  - robust to noise
  - flat output model
- Grammar based models
  - direct tree modification
  - inefficient on speech
  - not robust to noise

GLOBAL STRATEGY

LEARN NODES
structure information

NEW MINISTER OF THE BUDGET

STEP 1: DE-STRUCTURING

1. forget structure
2. consider tree nodes independent
3. learn each of them separately

- avoid cascade processes and the impact of error
- binary problems fast to learn
- loose node dependencies information during node hypothesization

STEP 2: LEARN NODES: A BINARY CRF PER NODE

Example with “loc.adm.town” CRF:

<table>
<thead>
<tr>
<th>CRF (CRF)</th>
<th>O</th>
<th>O</th>
<th>O</th>
<th>O</th>
<th>O</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>class</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>word</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>position</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Extract word = robust multilevel 3grams in [−2, +2] :
1. a priori knowledge (cities, countries, etc.)
2. or the word itself, if W has a high Mutual Information with label
3. otherwise Part-OF-Speech

STEP 3: STRUCTURING

Evaluation

- 8 participants
- evaluated in Slot Error Rate (Word Error Rate) +2
- the lower is the better
- on manual transcriptions
- +4 automatic transcriptions with different WER
- + a rover combination of them

CONCLUSION
I proposed a NER recognition system:

- able to deal with tree-structured output
- to process automatic transcriptions
- to be robust to noise

REFERENCES


ETAPE DATA

ETAPE data is composed of transcription with NE annotation:
1. 42h of TV Shows
2. 250h of radio
3. NEs have tree structure (compositional-hierarchical) [1]

ETAPE NE CAMPAIGN RESULTS IN SER

<table>
<thead>
<tr>
<th>sys</th>
<th>Manual Trans.</th>
<th>Automatic transcriptions with WER of</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trans.</td>
<td>23% 24% 25% 30% Rover</td>
</tr>
<tr>
<td>sys0</td>
<td>33.81</td>
<td>58.35 63.40 62.53 52.71 55.51</td>
</tr>
<tr>
<td>sys1</td>
<td>36.44</td>
<td>68.57 67.73 75.02 60.44 67.16</td>
</tr>
<tr>
<td>sys2</td>
<td>43.58</td>
<td>74.55 71.93 85.60 69.24 69.54</td>
</tr>
<tr>
<td>sys3</td>
<td>42.89</td>
<td>74.93 70.77 86.10 66.23 68.65</td>
</tr>
<tr>
<td>sys4</td>
<td>41.01</td>
<td>71.01 66.89 90.32 65.37 65.97</td>
</tr>
<tr>
<td>sys5</td>
<td>55.63</td>
<td>107.71 82.67 142.96 97.19 94.24</td>
</tr>
<tr>
<td>sys6</td>
<td>62.76</td>
<td>80.84 77.97 82.71 76.63 76.45</td>
</tr>
<tr>
<td>sys7</td>
<td>84.78</td>
<td>101.45 95.03 100.72 97.28 98.82</td>
</tr>
</tbody>
</table>

4 BEST ETAPE SYSTEMS

sys0: bottom-up cascade approach
1. CRF for labelling the components
2. PCFG for semantic tree reconstruction

sys1: top-down cascade approach
1. 2 CRF for labelling the entities (one for each tree level)
2. local boosting classifier to retrieve components knowing the entities

sys3: rule-based system with knowledge sources

sys4: use a data mining approach to extract NE annotation rules