Joint part-of-speech and dependency projection from multiple sources

Anders Johannsen*  Željko Agić*  Anders Søgaard

(formerly)* University of Copenhagen
Annotation projection

Parallel corpora

- transfer annotation from source to target
- may have multiple sources
- parse test set
- evaluate by leave-one-out
The many languages of the world

cross-lingual parsing suffers a little from EUROPARLalism

This work extends Agić et al. (2016):

train models for hundreds of languages

evaluate on 26 languages
Our corpora
arg max \( \sum_{(i,j) \in y} \text{score}_T(i, j) \) s.t. \( y \) is a tree

\[ \text{score}_T(\text{word}, \text{was}) = \text{score}_{DA}(\text{ordet}, \text{var}) \]
score_T(\text{the, was}) = \underbrace{\text{score}_{DA}(\text{ordet, var}) a(\text{var, was}) a(\text{ordet, word})}_{\text{PROPOSED EDGE}}$

Yes, but only if "was" is AUX and "the" is N.

---

Projecting layers of annotation.
Projecting layers of annotation

\[ \arg \max_y \sum_{(i,k,j,l) \in y} \text{score}_T(i,k,j,l) \quad \text{s.t. } y \text{ is a tree} \]

the edge \((i, j)\)

tags for \(i\) and \(j\)

more difficult

ILP model

Edges \(e_{i,k,j,l} \in \{0,1\}\)

Vertices \(v_{i,k} \in \{0,1\}\)
Flow $\phi_{i,k,j,l} \in \mathbb{R}^+$

Maximize $\sum_{i,k,j,l} e_{i,k,j,l} w_{i,k,j,l}$

One parent per token
$\sum_{i,k,l} e_{i,k,j,l} = 1 \quad \forall j \neq 0$

The root token (index 0) sends $n$ flow
$\sum_{j,l} \phi_{0,0,j,l} = n$

Each token consumes one unit of flow
$\sum_{i,k,l} \phi_{i,k,x,l} - \sum_{k,j,l} \phi_{x,k,j,l} = 1 \quad \forall x \neq 0$

One POS per token
$\sum_{k} v_{i,k} = 1 \quad \forall i \neq 0$

Active edges choose token POS
$v_{i,k} \geq e_{i,k,j,l} \quad \forall i \neq 0, j, k, l$
$v_{i,l} \geq e_{i,k,j,l} \quad \forall i, j, k, l$

Above, $i, j,$ and $x$ are token indices, while $k$ and $l$ refer to POS. Quantification over these symbols in the equations are always with respect to a given target graph.

(Martins, 2012)
<table>
<thead>
<tr>
<th>Predicted POS</th>
<th>Approach</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ILP</td>
<td>DCA</td>
<td>DELEX</td>
<td></td>
</tr>
<tr>
<td>EBC</td>
<td>51.62 (18)</td>
<td>48.39 (8)</td>
<td>42.44 (1)</td>
<td></td>
</tr>
<tr>
<td>WTC</td>
<td>53.58 (20)</td>
<td>48.40 (0)</td>
<td>47.35 (3)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gold POS</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>EBC</td>
<td>65.43 (25)</td>
<td>59.94 (2)</td>
<td>64.13 (−)</td>
<td></td>
</tr>
<tr>
<td>WTC</td>
<td>66.51 (23)</td>
<td>55.73 (0)</td>
<td>66.68 (−)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>POS tagging</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>EBC</td>
<td>69.40</td>
<td></td>
</tr>
<tr>
<td>WTC</td>
<td>73.05</td>
<td></td>
</tr>
</tbody>
</table>
We extended Agić et al. (2016) to project multiple layers of annotation jointly.

Approach stays simple and heuristics-free.

These initial experiments show promising results.

**Future work**

Project higher/lower layers of annotation, or larger tree parts.

Penalise inconsistent structures instead of disallowing.
Questions?