Semi-Supervised Frame-Semantic Parsing for Unknown Predicates

Dipanjan Das and Noah A. Smith

Frame-Semantic Parsing: Shallow, robust semantic analysis

Two subtasks:
- frame identification
- argument identification

Frame-Semantic Parsing:
A mulberry is very similar in shape to a loganberry.

Entity_1
Degree
SIMILARITY
Similar

Entity_2
Dimension

Limited coverage of (FrameNet) predicate lexicon

Solution:
Soft expansion of lexicon using graph-based semi-supervised learning

Graph Construction from labeled and unlabeled data

Label Propagation: Graph Objective Function

\[
\sum_{t \in V_l} \left\| q_t - r_t \right\|_2^2 + \mu \sum_{t \in V} \sum_{u \in \mathcal{N}(t)} w_{tu} \left\| q_u - q_t \right\|_2^2 + \lambda \sum_{t \in V} \left\| q_t - \frac{1}{|F|} \sum_{F} q_t \right\|_2^2
\]

labeled vertices
Supervised distribution for predicate \( t \)
Nearest neighbors of \( t \)
Induced distribution over predicate \( u \)
Set of FrameNet frames

Constrained Frame Identification:
For a target \( t \), selecting the best frame among the \( M \)-best frames from \( q_t \) under a probabilistic model

Example induced frame distributions

<table>
<thead>
<tr>
<th>( t = \text{discrepancy.N} )</th>
<th>( t = \text{contribution.N} )</th>
<th>( t = \text{print.N} )</th>
<th>( t = \text{mislead.N} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>#SIMILARITY</td>
<td>0.076</td>
<td>#SIMILARITY</td>
<td>0.167</td>
</tr>
<tr>
<td>#SIMILARITY</td>
<td>0.167</td>
<td>#SIMILARITY</td>
<td>0.081</td>
</tr>
<tr>
<td>#SIMILARITY</td>
<td>0.046</td>
<td>#SIMILARITY</td>
<td>0.054</td>
</tr>
<tr>
<td>#SIMILARITY</td>
<td>0.046</td>
<td>#SIMILARITY</td>
<td>0.054</td>
</tr>
<tr>
<td>#SIMILARITY</td>
<td>0.040</td>
<td>#SIMILARITY</td>
<td>0.042</td>
</tr>
<tr>
<td>#SIMILARITY</td>
<td>0.024</td>
<td>#SIMILARITY</td>
<td>0.028</td>
</tr>
<tr>
<td>#SIMILARITY</td>
<td>0.152</td>
<td>#SIMILARITY</td>
<td>0.130</td>
</tr>
<tr>
<td>#SIMILARITY</td>
<td>0.046</td>
<td>#SIMILARITY</td>
<td>0.046</td>
</tr>
<tr>
<td>#SIMILARITY</td>
<td>0.041</td>
<td>#SIMILARITY</td>
<td>0.038</td>
</tr>
</tbody>
</table>

Performance on unknown predicates

Baseline (Das et al., 2010)
Self-Training
Graph-Based

Parser freely available at:
http://www.ark.cs.cmu.edu/SEMAFOR