Goal:

- Given a natural language sentence, produce a labeled, directed graph that represents its semantic relationships.

Formalisms

- DM
  - Minimal Recursion Semantics
  - 90% Connected
  - 99% deterministic

- PAS
  - Enju predicate-argument-structures
  - 100% deterministic

- PCEDT
  - Tectogrammatical layer of Prague Czech-English Dependency Treebank 2.0
  - 85% deterministic

General approach:

- Learn an arc-factored statistical model that scores each labeled, directed edge.
- Choose subgraph with maximum sum of edge weights, subject to linguistically-motivated graph constraints.
- Similar to MSTParser, but semantic graphs are not trees

Semantic Well-Formedness Constraints

- Only one of each core argument allowed per predicate
- At most one arc between each pair of words.
- Connected?

Edge Models

- Multiclass Logistic Regression
  \[ P(\ell | \phi, x, i, j) = \frac{\exp\{\phi \cdot f(x, i, j, \ell)\}}{\sum_{\ell' \in L} \exp\{\phi \cdot f(x, i, j, \ell')\}} \]
  - best model when edges considered independently

- Structured SVM
  \[ -\phi^T f(x_i, y_i) + \max_y \phi^T f(x_i, y) + \text{cost}(y, y_i) \]
  - easy to incorporate graph constraints

Results

<table>
<thead>
<tr>
<th></th>
<th>LP</th>
<th>LR</th>
<th>LF</th>
<th>LM</th>
</tr>
</thead>
<tbody>
<tr>
<td>DM</td>
<td>0.8446</td>
<td>0.8348</td>
<td>0.8397</td>
<td>0.0875</td>
</tr>
<tr>
<td>PAS</td>
<td>0.9078</td>
<td>0.8851</td>
<td>0.8963</td>
<td>0.2604</td>
</tr>
<tr>
<td>PCEDT</td>
<td>0.7681</td>
<td>0.7072</td>
<td>0.7364</td>
<td>0.0712</td>
</tr>
<tr>
<td>Average</td>
<td>0.8402</td>
<td>0.8090</td>
<td>0.8241</td>
<td>0.1397</td>
</tr>
</tbody>
</table>

Discussion

Optimal choice of the constraints depends on the formalism, and evaluation metric

Missing (important) higher-order features