Improving unsupervised vector-space thematic fit evaluation via role-filler prototype clustering

Clayton Greenberg  Asad Sayeed  Vera Demberg

A multiple prototype framework for unsupervised thematic fit estimation

Thematic fit: Is “donut” a good patient of “eat”? Is “fork” a good instrument for “eating”?

Baroni & Lenci (2010) thematic fit estimation pipeline:
● count verb-role-filler triples

Fig. 1: TypeDM approximates roles from syntax.

Fig. 2: SDDM(X) gets roles directly from SENNA.
● adjust counts by local mutual information (LMI)
● retrieve the 20 highest LMI fillers & compute centroid
● return cosine similarity of test role-filler and centroid

Our improvements:
▶ adding direct links between verb arguments in a semantic parse-based model (SDDMX)
▶ calculating multiple prototypes for a given verb-role to handle polysemy and noise

Results and conclusions

<table>
<thead>
<tr>
<th>Padó (2007) patients</th>
<th>Padó (2007) agents</th>
<th>All from McRaeNN</th>
<th>All datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SDDM</td>
<td>SDDMX</td>
<td>TypeDM</td>
</tr>
<tr>
<td>Centroid</td>
<td>0.511</td>
<td>0.505</td>
<td>0.525</td>
</tr>
<tr>
<td>OneBest</td>
<td>0.447</td>
<td>0.467</td>
<td>0.509</td>
</tr>
<tr>
<td>2Clusters</td>
<td>0.526</td>
<td>0.498</td>
<td>0.551</td>
</tr>
<tr>
<td>kClusters</td>
<td>0.401</td>
<td>0.428</td>
<td>0.555</td>
</tr>
</tbody>
</table>

Fig. 4: Spearman’s $\rho$ for each method on each dataset and on all datasets together, using the 20 highest ranked words per verb-role.
▶ Overall, prototype clustering improves correlation with human judgements of thematic fit.
▶ The Padó (2007) dataset suggests that patients should have multiple prototypes per verb-role, but agents should not.

▶ SDDMX outperforms TypeDM on locations, but not on instruments. So, instruments are more syntactically predictable.
▶ Trends are consistent across distributional memory models and are robust to query size.
▶ Future directions: more sophisticated clustering, $\#$ of senses from WordNet, directly analyzing effect of polysemy on judgements.
▶ For more details, see our paper in the NAACL 2015 proceedings: http://aclweb.org/anthology/N/N15/N15-1003.pdf
▶ Distributional Memory models and datasets are available at: http://rolen.mmci.uni-saarland.de