Motivation

- **Distributional Semantic Models** build high-dimensional and sparse representations from co-occurrence statistics
- **Semantic similarity** is measured by vector cosine, which treats all features equally
- **Rank-based metrics** have been successfully applied to DSMs, but not yet on low-dimensional and dense Word Embeddings

APSynP: Rank-Based Metric

- **Hypotheses:**
  - Similarity = sharing a high number of relevant features
  - Dissimilarity = either non-sharing relevant features or sharing non-relevant features
  - Clustering = cluster members share their salient semantic dimensions to increase cluster cohesiveness
  - **APSyn for Sparse and High Dimensional Vectors:**

\[
APSyn(w_x, w_y) = \sum_{i=0}^{N} \frac{1}{AVG(r_{xy}(f_i))}
\]

- Maps the average feature ranks to a non-linear function, emphasizing the contribution of top-ranked feature
- Performs well on synonymy detection and similarity estimation, and SOTA results in thematic fit estimation
- Contribution of lower ranks are negligible

- **APSyn for Dense and Low Dimensional Vectors:**

\[
APSynP(w_x, w_y) = \sum_{i=0}^{|f|} \frac{1}{AVG(r_{xy}(f_i^p))}
\]

- \(N = |f|\) (removing a parameter)
- Adding a smoothing parameter, which can be tuned but tends to be constant \((p=0.1)\) across all experiments
- Preserving the non-linear weight allocation across the average feature ranks during the summation
- Ranks of all features contributing to the final score

Similarity Estimation

- **Spearman Correlation** between system-generated scores and human judgments
- **Benchmark:** WordSim-353, MEN, SimLex-999

![Similarity Estimation Chart]

- **Outlier Detection**
  - **Benchmark:** 8-8-8 dataset
  - 64 sets of 8 words + 1 outlier for the evaluation.
  - \(OPP = \frac{\sum_{w \in D} OP(W)}{D} \times 100\)
  - \(Accuracy = \frac{\sum_{w \in D} OD(W)}{D} \times 100\)
  - **Pairwise Comparisons:** outlier has the lowest average similarity score with the other words in the cluster
  - **Cluster Prototype:** outlier has the lowest similarity score with the average vector of the other N-1 words

<table>
<thead>
<tr>
<th></th>
<th>Skip-Gram</th>
<th>UMBCC</th>
<th>Wiki Acc.</th>
<th>MEN Acc.</th>
<th>GloVe</th>
<th>Wiki Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPP</td>
<td>93.6</td>
<td>64.1</td>
<td>93.8</td>
<td>70.3</td>
<td>81.6</td>
<td>91.8</td>
</tr>
<tr>
<td>Pairwise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APSyn</td>
<td>93.0</td>
<td>67.2</td>
<td>94.0</td>
<td>68.8</td>
<td>78.7</td>
<td>89.3</td>
</tr>
<tr>
<td>APSynP</td>
<td>94.0</td>
<td>68.8</td>
<td>94.5</td>
<td>73.4</td>
<td>81.8</td>
<td>92.8</td>
</tr>
<tr>
<td>Prototype</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AP</td>
<td>93.4</td>
<td>65.6</td>
<td>93.8</td>
<td>68.8</td>
<td>80.3</td>
<td>90.6</td>
</tr>
<tr>
<td>APSyn</td>
<td>92.6</td>
<td>70.3</td>
<td>91.0</td>
<td>62.5</td>
<td>81.6</td>
<td>88.7</td>
</tr>
<tr>
<td>APSynP</td>
<td>94.0</td>
<td>70.3</td>
<td>94.9</td>
<td>73.4</td>
<td>82.2</td>
<td>92.0</td>
</tr>
</tbody>
</table>

Contributions

- **APSynP**, a rank-based similarity measure adapted with a smoothing parameter for word embeddings
- Setting \(N=|f|\) and using a constant parameter makes APSyn unsupervised
- Comparable or better performances than cosine and APSyn on similarity estimation, clustering and outlier detection
- Pilot studies suggest that other rank-based metrics can outperform vector cosine in multiple settings

Reference