Scoring Lexical Entailment with a Supervised Directional Similarity Network

Marek Rei, Daniela Gerz and Ivan Vulić
Lexical Relations

**Task:** Graded lexical entailment
To what degree is X a type of Y?

<table>
<thead>
<tr>
<th>Relation</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>girl → person</td>
<td>9.85</td>
</tr>
<tr>
<td>guest → person</td>
<td>7.22</td>
</tr>
<tr>
<td>person → guest</td>
<td>2.88</td>
</tr>
</tbody>
</table>

Useful for query expansion, natural language inference, paraphrasing, machine translation, etc.
Lexical Relations

- Distributional vectors are not great for directional lexical relations
  
  \[
  \text{carrot} \sim \text{vegetable} \quad \text{new} \sim \text{old}
  \]

- **Retro-fitting** (Faruqui et al., 2015)
  **Counter-fitting** (Mrkšić et al., 2016)

  BUT these mostly affect words that are in the training data
Main Idea

01 Specialized network for directional lexical relations

02 Off-the-shelf pre-trained embeddings

03 Train the network to discover task-specific regularities in the embeddings
Supervised Directional Similarity Network

Fixed pre-trained word embeddings as input

Predict a score indicating the strength of a specific lexical relation
SDSN: Gating

Conditioning each word based on the other

\[
g_1 = \sigma(W_{g_1}w_1 + b_{g_1}) \\
g_2 = \sigma(W_{g_2}w_2 + b_{g_2}) \\
\tilde{w}_1 = w_1 \odot g_2 \\
\tilde{w}_2 = w_2 \odot g_1
\]
SDSN: Mapping

Mapping the representations to new spaces

\[ m_1 = \tanh(W_{m_1} \tilde{w}_1 + b_{m_1}) \]
\[ m_2 = \tanh(W_{m_2} \tilde{w}_2 + b_{m_2}) \]
SDSN: Sparse Features

Features based on sparse distributional representations

- cosine
- weighted cosine
  (Rei & Briscoe, 2014)
- ratio of shared contexts
SDSN: Scoring

Mapping the representations to a score

Optimize the network with labeled examples

\[ d = m_1 \odot m_2 \]

\[ h = \tanh(W_h d + W_x x + b_h) \]

\[ y = S \cdot \sigma(a(W_y h + b_y)) \]
HyperLex: Graded Lexical Entailment
HypeNet: Hyponym Detection

![Bar chart showing F1 scores for different methods: Shwartz et al. (2016), Glavaš and Ponzetto (2017), Roller and Erk (2016), SDSN, SDSN+.]
Conclusion

01 Can train a neural network to find specific regularities in off-the-shelf word embeddings

02 Traditional sparse embeddings still provide complementary information

03 Achieves state-of-the-art on graded lexical entailment
Thank you! Any questions?
## Examples

<table>
<thead>
<tr>
<th>Premise</th>
<th>Hypothesis</th>
<th>Gold</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>captain</td>
<td>officer</td>
<td>8.22</td>
<td>8.17</td>
</tr>
<tr>
<td>celery</td>
<td>food</td>
<td>9.3</td>
<td>9.43</td>
</tr>
<tr>
<td>horn</td>
<td>bull</td>
<td>1.12</td>
<td>0.94</td>
</tr>
<tr>
<td>wing</td>
<td>airplane</td>
<td>1.03</td>
<td>0.84</td>
</tr>
<tr>
<td>prince</td>
<td>royalty</td>
<td>9.85</td>
<td>4.71</td>
</tr>
<tr>
<td>autumn</td>
<td>season</td>
<td>9.77</td>
<td>3.69</td>
</tr>
<tr>
<td>kid</td>
<td>parent</td>
<td>0.52</td>
<td>8.00</td>
</tr>
<tr>
<td>discipline</td>
<td>punishment</td>
<td>7.7</td>
<td>3.2</td>
</tr>
</tbody>
</table>