1. Additional Results

In Table 1, we compare FOREST with three additional baselines:

- **Murphy et al. (2012):** a word representation trained using non-negative sparse embedding (NNSE) on our corpus. Similar to the authors, we use an NNSE implementation from http://spams-devel.gforge.inria.fr/ (Mairal et al., 2010).

- **Mikolov et al. (2013):** a log bilinear model that predicts a word given its context, trained using hierarchical softmax with a binary Huffman tree (continuous bag of words, CBOW-HS). We use an implementation from https://code.google.com/p/word2vec/.

- **Mikolov et al. (2013):** a log bilinear model that predicts context words given a target word, trained using hierarchical softmax with a binary Huffman tree (skip gram, SG-HS). We use an implementation from https://code.google.com/p/word2vec/.

We train these models on our corpus using the same setup as experiments in our paper.

2. Additional Two-Dimensional Projections

For FOREST, SG, and NCE with $M = 520$, we project the learned word representations into two dimensions using the t-SNE tool (van der Maaten and Hinton, 2008) from http://homepage.tudelft.nl/19j49/t-SNE.html. We show projections of words related to the concept “good” vs. “bad” and “man” vs. “woman” in Figure 1.

3. List of Word Similarity Datasets

We use the following word similarity datasets in our experiments:

- **Finkelstein et al. (2002):** WordSimilarity dataset (353 pairs).
- **Agirre et al. (2009):** a subset of WordSimilarity dataset for evaluating similarity (203 pairs).
- **Agirre et al. (2009):** a subset of WordSimilarity dataset for evaluating relatedness (252 pairs).
- **Miller and Charles (1991):** semantic similarity dataset (30 pairs)
- **Rubenstein and Goodenough (1965):** contains only nouns (65 pairs)
- **Luong et al. (2013):** rare words (2,034 pairs)
- **Bruni et al. (2012):** frequent words (3,000 pairs)
- **Radinsky et al. (2011):** MTurk-287 dataset (287 pairs)
- **Halawi and Dror (2014):** MTurk-771 dataset (771 pairs)
- **Yang and Powers (2006):** contains only verbs (130 pairs)

References


Table 1. Summary of results for non-negative sparse embedding (NNSE), continuous bag-of-words and skip gram models trained with hierarchical softmax (CBOW-HS and SG-HS). Higher number is better (higher correlation coefficient or higher accuracy).

<table>
<thead>
<tr>
<th>( M )</th>
<th>Task</th>
<th>NNSE</th>
<th>CBOW-HS</th>
<th>SG-HS</th>
<th>FOREST</th>
</tr>
</thead>
<tbody>
<tr>
<td>52</td>
<td>Word similarity</td>
<td>0.04</td>
<td>0.38</td>
<td>0.47</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>Syntactic analogies</td>
<td>0.10</td>
<td>19.50</td>
<td>24.87</td>
<td>24.38</td>
</tr>
<tr>
<td></td>
<td>Semantic analogies</td>
<td>0.01</td>
<td>5.31</td>
<td>14.77</td>
<td>9.86</td>
</tr>
<tr>
<td></td>
<td>Sentence completion</td>
<td>0.01</td>
<td>22.51</td>
<td>28.78</td>
<td>28.88</td>
</tr>
<tr>
<td></td>
<td>Sentiment analysis</td>
<td>61.12</td>
<td>68.92</td>
<td>71.72</td>
<td>75.83</td>
</tr>
<tr>
<td>520</td>
<td>Word similarity</td>
<td>0.05</td>
<td>0.50</td>
<td>0.57</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>Syntactic analogies</td>
<td>0.81</td>
<td>46.00</td>
<td>50.40</td>
<td>48.00</td>
</tr>
<tr>
<td></td>
<td>Semantic analogies</td>
<td>0.57</td>
<td>8.00</td>
<td>31.05</td>
<td>41.33</td>
</tr>
<tr>
<td></td>
<td>Sentence completion</td>
<td>22.81</td>
<td>25.80</td>
<td>27.79</td>
<td>35.86</td>
</tr>
<tr>
<td></td>
<td>Sentiment analysis</td>
<td>67.05</td>
<td>78.50</td>
<td>79.57</td>
<td>81.90</td>
</tr>
</tbody>
</table>

Figure 1. Two dimensional projections of the FOREST (left), SG (middle), and NCE (right) word representations using the t-SNE tool (van der Maaten and Hinton, 2008). Words associated with “good” (top) and “man” (bottom) are colored in blue, words associated with “bad” (top) and “woman” (bottom) are colored in red. The two plots on the top left are the same plots shown in the paper.


