The Influence of Down-Sampling Strategies on SVD Word Embedding Stability

Johannes Hellrich, Bernd Kampe & Udo Hahn

Jena University Language & Information Engineering (JULIE) Lab
Friedrich Schiller University Jena,
Jena, Germany
www.julielab.de
Typical Word Embeddings are Unstable

Random embeddings → Random processing → Final embeddings

dog, cat, tiger
Typical Word Embeddings are Unstable

- **tiger**
- **dog**
- **cat**

Random embeddings → Random processing → Final embeddings
Measuring Stability

\[ j_{@n} := \frac{1}{|A|} \sum_{a \in A} \frac{\left| \bigcap_{m \in M} \text{msw}(a, n, m) \right|}{\left| \bigcup_{m \in M} \text{msw}(a, n, m) \right|} \]
Why SVD Embeddings?

corpus

counting

<table>
<thead>
<tr>
<th></th>
<th>food</th>
<th>roar</th>
</tr>
</thead>
<tbody>
<tr>
<td>dog</td>
<td>475</td>
<td>156</td>
</tr>
<tr>
<td>cat</td>
<td>823</td>
<td>492</td>
</tr>
<tr>
<td>tiger</td>
<td>51</td>
<td>19</td>
</tr>
</tbody>
</table>

SVD

final embeddings

dog
cat

tiger
Why SVD Embeddings?

- Corpus
- Counting & down-sampling
- Replaced with association values in SVD_{PPMI} (Levy et al., TACL 2015)
Why Down-Sampling?

- Avoids over-representing frequent words
- Closer context words are more salient than distant ones
→ Increased Performance (Mikolov, NIPS 2013)
Down-Sampling Mechanism

Probabilistic
- word2vec
- $\text{SVD}_{\text{PPMI}}$

Weighting
- GloVe
- New: $\text{SVD}_{\text{wPPMI}}$
Experimental Design I/II

• Three Corpora:
  • Corpus of Historical American English 2000s decade (COHA; 28M tokens.)
  • English News Crawl Corpus (NEWS; 550M tokens)
  • Wikipedia (WIKI; 1.7G tokens)
→ Other studies used mostly COHA-sized corpora!
Experimental Design II/II

• Train 10 models each with SGNS, GloVe, SVD\textsubscript{PPMI} (none / prob. down-sampling), SVD\textsubscript{wPPMI}
• Evaluate intrinsically with four word similarity & two analogy test sets
• Measure stability with j@10 for 1k most frequent words
Stability Results

GloVe’s high stability (Antoniak & Mimno, TACL 2018; Wendlandt et al., NAACL 2018) is true only for small corpora.
Exemplary Accuracy Results

Wilcoxon rank-sum test shows $SVD_{wPPMI}$ and SGNS to be indistinguishable in accuracy over all test sets and corpora.
Conclusion

• Typical word embeddings are unstable
• Down-sampling details greatly affect stability
• GloVe’s stability is worse than claimed in literature
• $\text{SVD}_{\text{wPPMI}}$ embeddings provide SGNS-like performance and perfect stability
• See paper for additional results (and bootstrapping)
The Influence of Down-Sampling Strategies on SVD Word Embedding Stability

Johannes Hellrich, Bernd Kampe & Udo Hahn

Jena University Language & Information Engineering (JULIE) Lab
Friedrich Schiller University Jena,
Jena, Germany
www.julielab.de