



Improving Context Based Thesaurus

Nathaniel Schub, Aniruddh Nautiyal, Jae Il Lee
12/11/18



Objective

Build a model which recommends synonym candidates that best fit in a given context.

| | |
|---------------------------|-------------|
| <i>substantial</i> | 0.39 |
| <i>notable</i> | 0.43 |
| <i>big</i> | 0.64 |
| <i>significant</i> | 0.78 |

We're expecting a **major** change in the organization next year

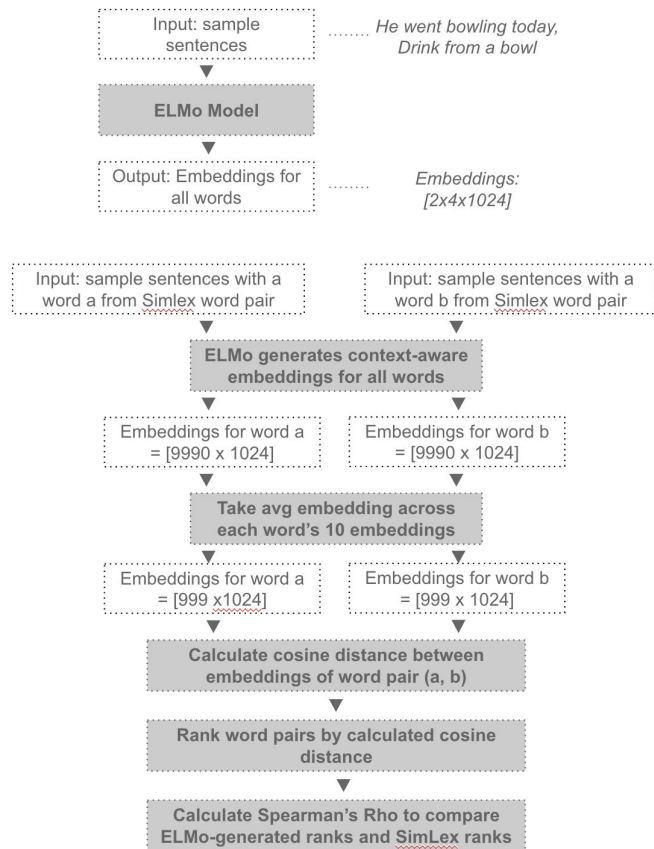
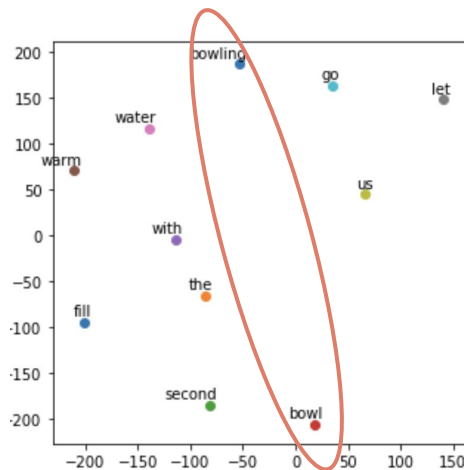


Approach

- Used Word2Vec embeddings as the baseline (not local context-aware), also evaluated GloVe
- Generated context as sentences (10 per word) using a dictionary API
- Evaluated models (NGram and RNN) that attempt to capture semantic differences between words
- Evaluated a final biLSTM model that uses *adaptive word embeddings*
- Aimed to combine the top performing models to build the context-aware thesaurus
- Used '1 Billion Word Benchmark Corpus' throughout, for model training

Models

- ELMo (Embeddings from Language Models)
 - Bi-directional LSTM that incorporates word's context within a sentence to build adaptive embedding
- Evaluated RNN and Trigram as well
- Baseline model- Word2Vec





Evaluation

- Metric: Spearman's Rho score for comparing model performance
- Evaluation datasets: list of word pairs with human annotated scores
 - SimLex-999
 - most challenging to score well
 - differentiates semantic similarity from semantic relatedness
 - WordSim-353
 - MEN 3000
- Model scoring functions:
 - RNN : difference in probabilities of whole sentence
 - Trigram: difference in probabilities of each word occurring with contexts
 - Word2Vec & ELMo: cosine distance between word embeddings

Evaluation

| SimLex Pair | Score (0-10) |
|-----------------|--------------|
| Old, New | 1.58 |
| Happy, Cheerful | 9.55 |
| Stupid, Dumb | 9.58 |
| ... | ... |
| Nice, Generous | 5.00 |
| Bad, Great | 0.35 |



| SimLex Pair | Rank (1-999) | | Ranked Score from each Model | SimLex Pair |
|-----------------|--------------|--|------------------------------|-----------------|
| Old, New | 998 | | 950 | Old, New |
| Happy, Cheerful | 2 | | 1 | Happy, Cheerful |
| Stupid, Dumb | 1 | | 2 | Stupid, Dumb |
| ... | ... | | ... | ... |
| Nice, Generous | 500 | | 400 | Nice, Generous |
| Bad, Great | 999 | | 800 | Bad, Great |

 *Spearman's Rho*



Results: Spearman's Rho

SimLex scores on the models evaluated:

- RNN: **0.03**
- Trigram: **0.10**
- ELMo: **0.43**

Reference - Human annotators : **0.67**

Performance of ELMo across all 3 datasets:

| Dataset | Word2Vec | ELMo | Change(%) |
|---------------|--------------|--------------|--------------|
| <i>SimLex</i> | <i>0.367</i> | <i>0.434</i> | <i>18.17</i> |
| WordSim | 0.591 | 0.570 | - 3.57 |
| MEN | 0.676 | 0.643 | - 4.96 |

Final Model





Conclusions

- Improvement on Word2Vec's context-unaware embeddings
- Model sometimes recommended typos, antonyms, etc.
 - Some of these issues may be fixed by incorporating human-built thesaurus
- Some combinations of sentences/target words worked better than others



Questions