Word Similarity & Relatedness

Evaluation of Distributional Models of English using the WordSim353 dataset

Universitat Pompeu Fabra
Class: Semàntica computacional
Lecturer: Gemma Boleda, Lucas Weber

Eva Richter & Nora Graichen
November 24th, 2021

Task and Dataset

Evaluate one or more distributional model of English of choice using the WordSim353 dataset, in the version that distinguishes between similarity and relatedness.
Word Embeddings (300 Dimensions) used:

→ Word2Vec 43'981 Vectors

(nltk_data/models/word2vec_sample/pruned.word2vec.txt)

→ Fasttext 2'000'000 Vectors

(https://dl.fbaipublicfiles.com/fasttext/vectors-crawl/cc.en.300.vec.gz)

WordSim353: test collection for measuring word similarity or relatedness (Gabrilovich et. al 2002)

→wordsim_**relatedness**_goldstandard.txt (Alfonseca et.al 2009)

→wordsim_**similarity**_goldstandard.txt (Alfonseca et.al 2009)

number of similarity pairs: 252

tiger	cat	7.35
tiger	tiger	10.00
word	similarity	4.75
professor	cucumber	0.31

number of relatedness pairs: 203

computer	keyboard	7.62
Jerusalem	Israel	8.46
summer	nature	5.63
king	cabbage	0.23

2

Method

WordSim353:

for each dataset/ embedding:

    calculate for each word pair the cosine in embedding space if possible

    compare scores to goldstandard of dataset

Spearman's rank correlation coefficient:

→   should have a high correlation with
      the human goldstandard scores

Interpretation of Spearman's RHO:
0.0-0.19   'very weak';
0.20-0.39 'weak';
0.40-0.59 'moderate';
0.60-0.79 'strong';
0.80-1.0 'very strong'
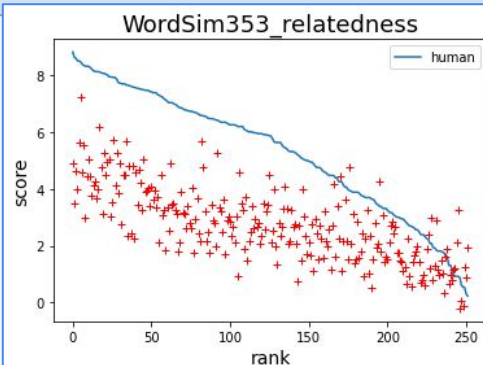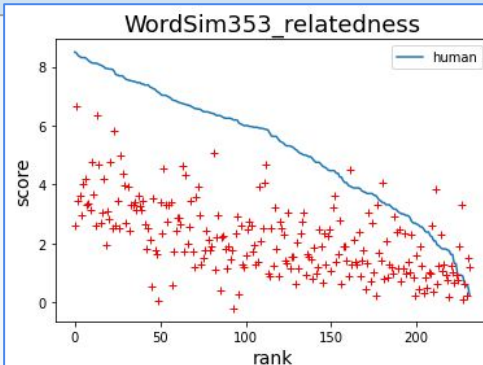
Quantitative Results

Fasttext:

Word2Vec:

Spearman's RHO: 0.7356	**0.6016** (20 pairs not found)

WordSim353_relatedness

WordSim353_relatedness

0.8345	**0.7657** (21 pairs nf)

WordSim353_similarity

WordSim353_similarity

4
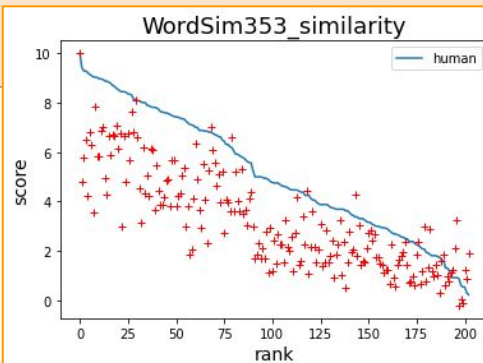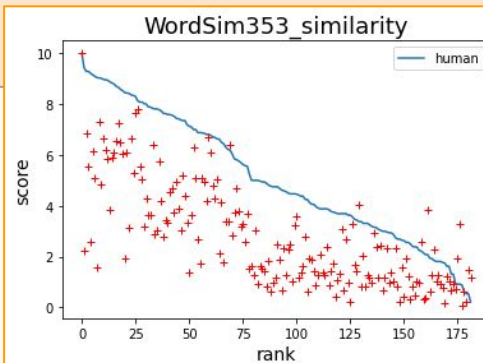
Error Analysis Results

Questions: → *Which notion do word embeddings model <u>better</u>, **similarity**, or **relatedness**?*

→ *For which kinds of semantic phenomena are human vs. model similarities the most dissimilar?*

scores that have a difference greater than 5:

Fasttext:	Word2Vec:
9 word pairs, **top 10:** - - -	27 word pairs, **top 10:** - - -
4 word pairs - - -	8 word pairs - - -

Error Analysis Results

→ *For which kinds of semantic phenomena are human vs. model similarities the most dissimilar?*

scores that have a difference greater than 5:

Fasttext:

word pair	human	score	difference
('network', 'hardware')	8.31	3.0003	5.3097
('investor', 'earning')	7.13	1.83274	5.29726
('money', 'property')	7.57	2.27292	5.29708
('Arafat', 'terror')	7.65	2.35709	5.29291
('preservation', 'world')	6.19	0.925084	5.26492
('money', 'possession')	7.29	2.12168	5.16832
('boxing', 'round')	7.61	2.44675	5.16325
('Maradona', 'football')	8.62	3.47671	5.14329
('money', 'withdrawal')	6.88	1.78997	5.09003

9 word pairs, top 10:
- **being in the**
  semantic field
  how related ?
  → **unknown context,**
  word distributions

Word2Vec:

word pair	human	score	difference
('investor', 'earning')	7.13	0.0340833	7.09592
('money', 'possession')	7.29	0.531828	6.75817
('preservation', 'world')	6.19	-0.214707	6.40471
('money', 'withdrawal')	6.88	0.560263	6.31974
('soap', 'opera')	7.94	1.9558	5.9842
('discovery', 'space')	6.34	0.412564	5.92744
('money', 'bank')	8.5	2.61321	5.88679
('dollar', 'loss')	6.09	0.265669	5.82433
('hundred', 'percent')	7.38	1.80219	5.57781
('summer', 'drought')	7.16	1.63789	5.52211

27 word pairs, top 10:
- **different representation**
  →**static embedding**
- **semantic field; connected**
  through additional word
  (entertainment/numbers/heat..)

4 word pairs

word pair	human	score	difference
('money', 'currency')	9.04	1.60101	7.43899
('fuck', 'sex')	9.44	2.23392	7.20608
('dollar', 'buck')	9.22	2.56212	6.65788
('planet', 'star')	8.45	1.93171	6.51829
('phone', 'equipment')	7.13	1.3774	5.7526
('money', 'dollar')	8.42	3.13566	5.28434
('opera', 'performance')	6.88	1.71675	5.16325
('street', 'avenue')	8.88	3.8327	5.0473

8 word pairs

- **could be used interchangeably ?**
  →but: **static**
  embedding
- universe related ?
- hyponym
- hypernym

word pair	human	score	difference
('money', 'currency')	9.04	3.55105	5.48895
('planet', 'star')	8.45	2.97497	5.47503
('phone', 'equipment')	7.13	1.83728	5.29272
('bird', 'cock')	7.1	2.08169	5.01831

- synonym
- synonym

- hypernym
- hypernym

6

References

WordSim353 - Similarity and Relatedness: Eneko Agirre, Enrique Alfonseca, Keith Hall, Jana Kravalova, Marius Pasca, Aitor Soroa, A *Study on Similarity and Relatedness Using Distributional and WordNet-based Approaches*, In Proceedings of NAACL-HLT 2009.

WordSimilarity-353 Test Collection: Y. Matias E. Rivlin Z. Solan G. Wolfman E. Ruppin L. Finkelstein, E. Gabrilovich. 2002. *Placing search in context: The concept revisited*. ACM Transactions on Information Systems, 20(1):116–131.