Semantic similarity of words

- Simlex999(SL) contains 999 word pairs and there were no missing words in both WordNet(WN) and FastText(FT).
- WN's similarity scores were discrete, values like 0.(3), 0.25, 1 and 0.5 covered almost all of the 999 pairs. It is because WN's basic path similarity counts the number of connections between the synsets along the shortest path (the further the synsets, the less similar they are). FT, however, is much less discrete as it uses cosine similarity between vectors.
- WN had larger variance (0.04435063589902128) than FT (0.027196266781775794) and the scores of the latter rarely exceeded 0.7. The variance of SL was 0.06829621454086718, which tells us that WN and FT might have done a less decisive evaluation, since the scores were not as scattered as in SL.
- Kendall's tau scores were close - 0.35344887126870356 for WN and 0.3301400933912036 for FT, so were the Spearman Correlation scores - 0.4756537349907476 and 0.4644247750821859. The scores on their own seem quite low given the range (-1 - 0 or 0 - 1), and similarity between the two makes it harder to make conclusions, however, we can look at existing milestones in the field. The following is quoted from SL github page:
  The best performance of a model trained on running monolingual text is a Spearman Correlation of 0.56 [1].
  A Neural Machine Translation Model (En->Fr) trained on a relatively small bilingual corpus achieves a Spearman Correlation of 0.52 [2].
- The fact that the scores are positive means there is a similar direction of change, and no inverse correlation. If we think of our scoring as a binary mapping of 'similar' and 'not similar', the positive Kendall's tau means we would generally get it right with both WN and FT, but then we have the degree to which they were successful.
- Some instances where the two had differences given SL:
  - Words 'smart' and 'intelligent' should score high, but WN gives them 0.333, while FT gets it right with 0.705. FT's score here is a good explanation to our Kendall's tau scores, we captured the fact that those words are somewhat similar, but maybe did not do the perfect job at telling the degree of similarity (higher than 0.5, but not too close to 0.92).
  - Contrary to the previous example, words 'certain' and 'sure' (0.842) scored 1.000 in WN and 0.419 in FT.
  - Words 'plane' and 'jet' (0.810) scored 0.500 and 0.578. Which shows how both of them failed to some degree, too.

Word1	**Word2**	**SimLex**	**WordNet**	**FastText**
old	new	0.158	0.333	0.442
smart	intelligent	0.920	0.333	0.705
hard	difficult	0.877	1.000	0.631
happy	cheerful	0.955	0.333	0.546
hard	easy	0.095	0.333	0.486
fast	rapid	0.875	0.333	0.526
happy	glad	0.917	1.000	0.674
short	long	0.123	0.333	0.662
stupid	dumb	0.958	0.333	0.871
weird	strange	0.893	0.333	0.851
wide	narrow	0.103	0.333	0.535
bad	awful	0.842	0.333	0.669
easy	difficult	0.058	0.333	0.617
bad	terrible	0.778	0.333	0.732

Table 1: Semantic Similarity Scores

References

[1] Symmetric Pattern Based Word Embeddings for Improved Word Similarity Prediction. Roy Schwarz, Roi Reichart and Ari Rappoport, CoNLL 2015.

[2] Embedding Word Similarity with Neural Machine Translation. Felix Hill, KyungHyun Cho, Sebastien Jean, Coline Devin and Yoshua Bengio. ICLR. 2015.