"Adventures With Word Embeddings"

Algorithm	Window Size	Dimension	WordSim	BATS 1	BATS 2	BATS 3	MSR
Word2vec	2	100	0.209	0.022	0.074	0.053	0.671
Word2vec	2	300	0.209	0.032	0.064	0.054	0.671
Word2vec	2	1000	0.209	0.036	0.079	0.052	0.671
Word2vec	5	100	0.209	0.025	0.058	0.041	0.667
Word2vec	5	300	0.205	0.039	0.096	0.045	0.668
Word2vec	5	1000	0.221	0.038	0.072	0.041	0.670
Word2vec	10	100	0.240	0.049	0.072	0.049	0.660
Word2vec	10	300	0.234	0.029	0.045	0.051	0.661
Word2vec	10	1000	0.208	0.030	0.038	0.049	0.660
SVD	2	100	0.083	0.049	0.021	0.031	0.687
SVD	2	300	0.207	0.035	0.003	0.039	0.674
SVD	2	1000	0.313	0.024	0.003	0.025	0.669
SVD	5	100	0.238	0.045	0.026	0.033	0.681
SVD	5	300	0.293	0.031	0.019	0.042	0.678
SVD	5	1000	0.318	0.017	0.000	0.015	0.663
SVD	10	100	0.308	0.029	0.018	0.040	0.678
SVD	10	300	0.305	0.021	0.009	0.042	0.671
SVD	10	1000	0.318	0.004	0.000	0.009	0.663

Results Analysis

➢ Dimensionality is a critical parameter in word embedding. The function of dimensionality is linear or quadratic which influences computational cost and learning time. In theory the high dimension can store larger information, but it would also require complex computation resources. This implies higher dimensions can achieve higher accuracy but that might not be the case always. Higher dimensionality is not recommended as it is expensive in resource usage, decreases the performance and takes more time for learning. The dimension size beyond 500 is considered high and does not benefit. It comes with the problem of over fitting. However, even if the dimensionality is too less it will not work well either. It is because if dimensionality is low, the word embedding will not be able to capture all possible word relations.

Practically, in some application, smaller dimensions work well. The dimension size between 100-300 works nearly perfect and works ideally. After this upper limit, the training will be quite slow, but it won't work significantly well. So larger dimensionality not always equate to better performance.

➢ It is not necessary if performance on one task is better than it will be same on the others. MSR gave almost the same accuracies on both models. BATS1,2,3 showed variation in results with not much differences either. With less dimension SVD didn't perform well whereas when the dimension is larger i.e. 1000 it worked quite well.

➢ As per our observation, the performance is similar across all analogy categories. However, there is some variation in the results for different dimensions. From above table we can clearly see a pattern for different dimensions across all of the analogies. We have considered inflectional_morphology, encyclopedic_semantics and lexicographic_semantics for our comparison. We have concluded following results from observation.

- The values for BATS accuracies have smaller value for each window size with dimension of 100.
- The values of BATS accuracies for each analogy has greater value for a dimension of 300
- The values of BATS accuracies are smaller for dimension of 1000.

Hence, we can say that window size has lower impact on the BATS accuracy as compared to the dimension.