ULL Practical 1 T. Scheffers (11728191) and S. Knegt (11784261)

Methods In this practical, we have used tests to explore how well various word embedding techniques work. From the three explored word embeddings, two are constructed with a Bag Of Words model with window size k of 2 & 5 and one using a dependency based model which is constructed using word dependencies. We are comparing each embedding type on three tasks: word similarity task where we use SimLex and MEN similarity datasets as comparison and score on Spearman and Pearson correlation, word analogy task where we use the Google Word Analogy and score on Mean Reciprial Rank and finally we look at clustering of words in a lower dimensional space.

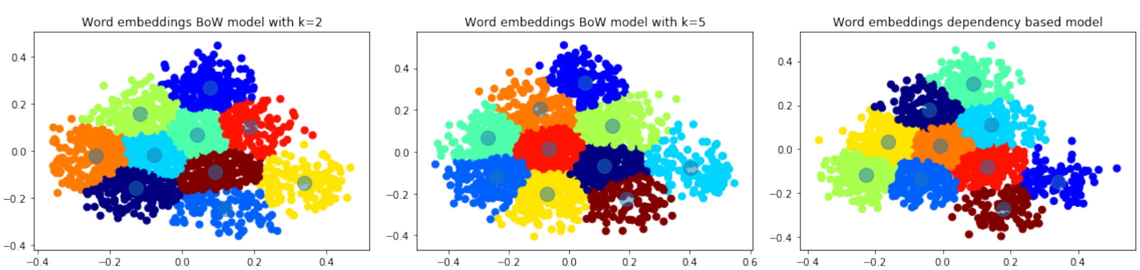
Word Similarity Task The results of the word similarity task are shown in the table below. The highlighted results give the maximum correlation for each benchmark and metric. We can see that for the SimLex similarity dataset, the dependency based embeddings work better, where for the MEN similarity dataset, the BoW with K=5 embeddings worked better. It is also interesting to see that both correlation metrics are much lower on the SimLex dataset for all word embedding models.

Word embedding	Similarity	Spearman	Pearson	Similarity	Spearman	Pearson
BoW k=2	SimLex	0.414	0.428	MEN	0.700	0.678
BoW k=5	SimLex	0.367	0.376	MEN	**0.723**	**0.708**
Dependency based	SimLex	**0.446**	**0.462**	MEN	0.618	0.597

Word Analogy Task For the word analogy task, we used 1000 of the 20000 analogies due to computational issues. The Mean Reciprocal Rank per model is: BoW2: 6.36e-05, BoW5: 6.87e-05 and the Dependency model: 0.24. So the BoW models failed to predict the right analogy, which makes sense as word context does not necessarily capture analogies, where dependencies do. An analogy example for dependency based model is the following. The analogy: eat $\rightarrow$ eats : speak $\rightarrow$ speaks, gives the top 5 results: speaks, speak, spoke, pronounces, brags, which are all words which are very relevant to speaking.

Clustering Word Vectors We first reduced the dimensionality of the word vectors to 2 using PCA and then clustered these reduced word-vectors using $K = 10$ clusters. These results are shown in the figures below.

When clustering the (normalized) word vectors in their normal dimensionality and qualitatively inspecting the clusters we find the following: (1) when we increase the number of clusters from 50 to 200, the qualitative results improve but for $K = 500$ clusters we see that over 60 clusters have no members assigned to them, (2) when using 200 clusters we find that clusters contain very related words like: 'admission', 'entry', 'pass', but also almost always also contain unrelated works an example cluster is 'breakfast', 'dinner', and 'maximum'. When comparing the three models, we see that the BoW2 model outperforms the BoW5 model, indicating that a smaller context leads to better clusters. The DEP model shows less semantically related words but words that have a similar function like 'bank' and 'college'.



(a) BoW2 (b) BoW5 (c) Dependency based

Conclusion In this practical we learned to test different word embedding models on word similarity tasks, where it was interesting to see that different embeddings models work better on different datasets. The dependency based embedding were superior in word analogy task, but less successful in the clustering of words. For the clustering task it was quite vague how to determine success as there is no optimal solution, but could only be inspected by hand. Overall we learned a lot about using and testing embeddings.web