Meaning and Computation - Exercise 2

Ron Pick - 301643334 , Ron Likvornik - 302574694

Do the models produce similar results for different POS? If not, please hypothesize why this is so

Looking on the results of each model, we can point out few differences:

- The results of the verbs are the lowest in the frequency based models.
- The results of the nouns are the lowest in the PPMI based models.
- The results of the adjectives show almost no correlation in all models.

Our models are based on co occurrences between words, where the co occurrence is defined by the occurring of the exact words. No other features of the words were taken under consideration. Hence it is clear that our model missed for example, co occurrences of similar word pairs with morphological changes, or counted graphonyms with different part of speech. Thus if the exact context words didn't appear beside some similar words (from the simlex collection) then the model couldn't catch the similarity between the pairs. Even if similar context words appeared beside some pair, each context word close to one word of the pair, the similarity haven't taken under consideration.

Our model was trained on large corpus (about 230M words) and still we expected some biased results, as described above. With the above explanation we can explain the inconsistent results of the adjectives in all models, as they can appear close to words from an infinite collection of nouns, but tend to appear beside sub collection due to semantic considerations.
In fact, this can explain that all the results are tend to show no correlation.

The negative relation of the verb scores can be explained by the freq do with the gold standard, in the frequency based models, can be explained by that the frequency models are overfitting the corpus. Even large corpus can't cover all the possible contexts for those verbs. By that we mean to say that different corpuses could produce different results if they contain different topics, and thus the negative relation is considered random. After applying the PPMI method, the normalization is stabling the results and in this case improving them by 0.1 - 0.15.

About nouns case we can propose the following.
Nouns are tend to appear close to stopwords (determiners, prepositions, etc). When looking on the frequency based model, the count of the stopwords are overcome the number of other context words dramatically, and this situation cause too many pairs of nouns to get high score of similarity without taking under significant consideration other context words which implies more about the semantic similarity. This causes to inconsistency and to result of no correlation. After applying the PPMI method, the effect of the high count of the stopwords, becomes smaller, and the similarity score becomes more sensitive to other context words.. Hence the nouns are being affected more by the other context words in the corpus, and in our situation these other context words causing negative correlation.

How does changing the window size alter the results?

By looking at the correlation there is not much difference between one window to the second window in both frequency based methods. We can point out a change of 1 point less for verbs and half a point less for noun, in context window of 5. When looking only on frequency affect without normalizing the count(frequency based methods), very significant context words that appear only few times comparing to less significant ones, is not taken under consideration, moreover, in context window of 5 it is much easier for stop words or other common words to bias the results. In the other hand, in window of 2 it is much harder to significant words to affect the similarity.

But if you go and check the similarity you would witness the different between one frequency window to the other one as explained.

The PPMI method, a method that try to produce a better result than the frequency based model, by normalizing the information. The PPMI similarity produce almost the same result by using window size 2 and size 5 which implies and makes us conclude that window size doesn't alter the results quite deal. The normalizing taking under consideration significant context words.

Guess. Will removing stopwords be helpful in your opinion? Explain your guess.

In general, our answer is Yes, it will be helpful to remove stopwords. Stopwords often have grammatical purpose, and when looking for accurate semantic similarity between many words, stopwords can bias the result of words which not occurred too often. When looking on context window of 2, stop words which are not much related to a word, won't be counted a lot, but comparing to all words, when expecting the exact words to appear to be counted, stopwords would appear more than other words, which can imply much better about similarity. In the better case it won't change the result, but in general it could do so.

In context window of 5 it is difficult not to encounter stopwords. Common words will encounter a lot of stopwords, and hence those can affect the results and say that any pair of common words have some similarity between them.

Of course there are words that stopwords do imply about their semantics, but words are pretty rare. In this case stop words are helpful, but for the general results, they are better not to be counted. To conclude we need to always take into account the fact that by choosing the most frequent words to check into a list, the stop words will always be there and take place of other normal words therefore removing stopwords do change the result.