* 1. (see code)

|  |  |  |
| --- | --- | --- |
|  | w = 3 | w = 6 |
| #(chicken, the) | 52 | 103 |
| #(chicken, wings) | 6 | 7 |
| #(chicago, chicago) | 38 | 122 |
| #(coffee, the) | 95 | 201 |
| #(coffee, cup) | 10 | 14 |
| #(coffee, coffee) | 4 | 36 |

1.3

* simlex999: correlation=0.05876135331349779
* MEN: correlation=0.2251396048448754

Overall, we saw a much higher correlation score when the word vectors were evaluated on the MEN dataset compared to the simlex999 dataset. However, the absolute values of correlation were in the low range for both simlex999 and MEN datasets.

2.

IDF without logarithm transformations or other scaling techniques:

* simlex999: correlation=0.1643113945921928
* MEN: correlation=0.47281906258988254

We saw a large increase of correlation scores when the word vectors were evaluated on both simlex999 and MEN datasets. TF-IDF word vectors still performed much better on the MEN dataset than the simlex999 dataset.

3.1

highest PMI (from high to low):

('tea', 8.16600126243293)

('drinking', 7.58797865873193)

('shop', 7.411693771493207)

('costa', 7.350256393786161)

('shops', 7.260751873418467)

('sugar', 6.533949521544205)

('coffee', 6.501977131805925)

('mix', 6.131195903101976)

('seattle', 5.950816325067398)

('houses', 5.868161497268183)

lowest PMI (from low to high):

('he', -2.26033826495274)

('be', -2.1509730526875237)

('had', -1.9875291676196303)

('this', -1.979549817934235)

('not', -1.9115928402014317)

('its', -1.839457915441101)

('after', -1.598505205571959)

('more', -1.4785257922880328)

('when', -1.4043486976803334)

('page', -1.2805627423998573)

3.2

* simlex999: correlation=0.18643183126956037
* MEN: correlation=0.46563240836038006

Overall, PMI-based word vectors performed at a similar level as TF-IDF word vectors; although PMI-based word vectors achieve correlation scores slightly higher than TF-IDF word vectors on the simlex999 dataset, and slightly lower than TF-IDF word vectors on the MEN dataset. For both simlex999 and MEN datasets, PMI-based word vectors performed much better than the original distributional counting-based word vectors.

4.1

A screenshot of a computer screen

Description automatically generated

The figure above shows the EvalWS scores across difference window sizes and vector generating methods, faceted by context vocabulary choices and evaluation datasets with the same color scale across all four subplots. The highest EvalWS score was obtained using IDF method with a window size of 6 on vocab5k context vocabulary and MEN evaluation dataset. The lowest EvalWS score was obtained using raw counts method with a window size of 6 on vocab15k context vocabulary and simlex999 evaluation dataset.

Window size’s effect on EvalWS scores differs between evaluation datasets. For simlex999, as window size increases, EvalWS score decreases for all three methods (counts, IDF, PMI) and both context vocabulary sets (5k and 15k). But for MEN, as window size increases, EvalWS score also increases for all three methods (counts, IDF, PMI) and both context vocabulary sets (5k and 15k). A potential reason for this trend is that

The effect of size of context vocabulary also differs between evaluation datasets. For simlex999, as the size of context vocabulary increases,

Decrease: counts,

4.2