CS510 MP2

yuweic3

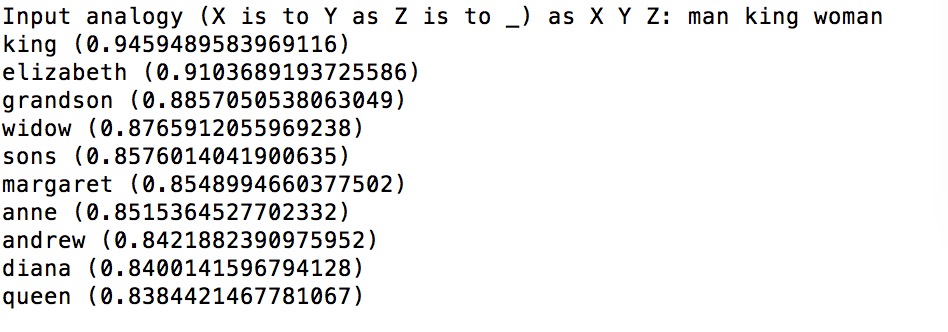
2017/10/7

# 1.c.i

# 1.c.ii

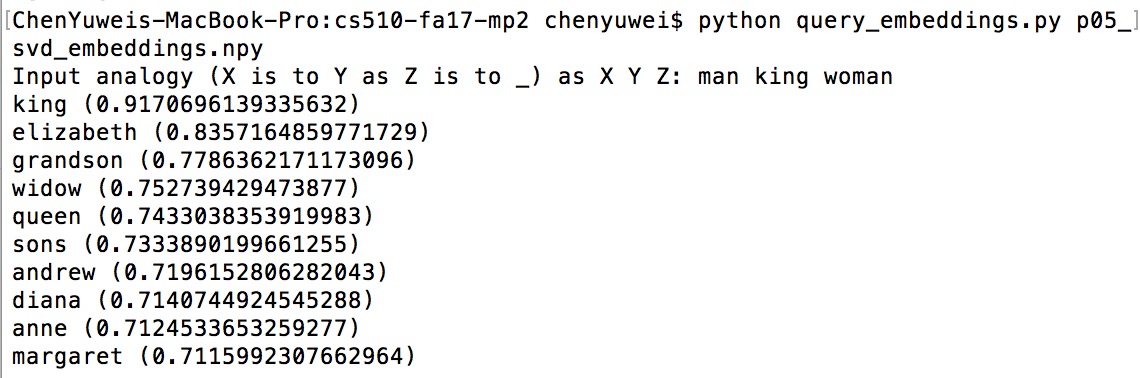
Here are three words I come up for analogy: queen, ladies, princess

With :



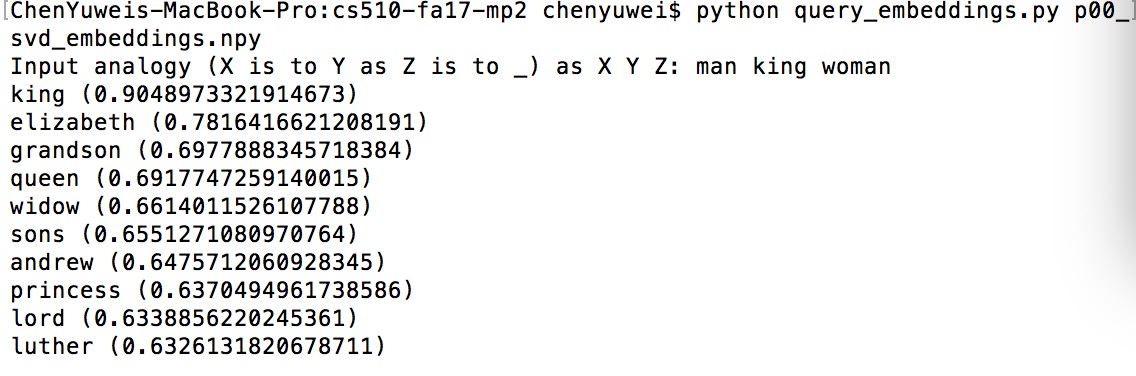
The word “queen” appears in the ranking list but the rank is not high.

With :



This time, the word “queen” has a higher rank with 5.

With :

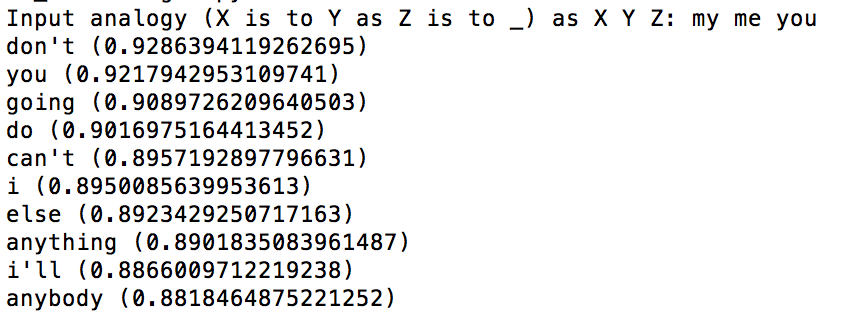


When = 0, the word “queen” has a much higher rank with 4.

In sum, we find that when decreases, the word “queen” has a higher rank.

Here is an example I come up with: **my to me is you to** ?

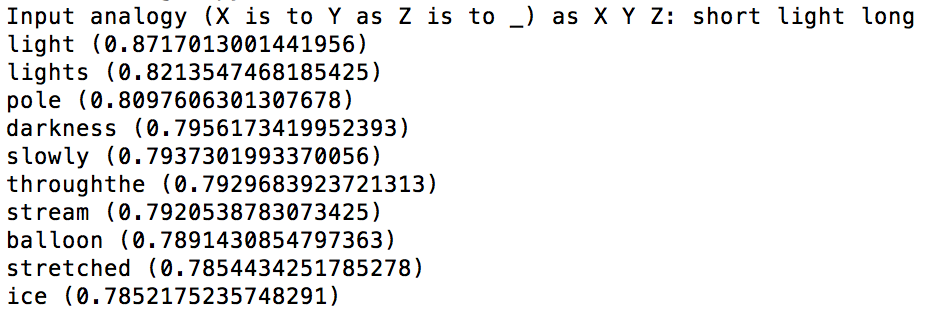
With :



In my expectation, the word should be you. And the result I get back is following my expectation as the word “you” appears in the rank list.

Here is another example: **short to light is long to ?**

With :



In my expectation, the word should be “Heavy”. However, heavy does not appear in the ranking list. I think it is because of dual meaning of “light” and the system treats it as the opposite of darkness.

In sum, if the word given to the system has multiple meaning, the system may have bad performance and give the wrong analogy.

# 2.a

The reason we use MRR is that we only interested in the rank of the first answer. Although other answers are nice, they don’t matter. For example, in the web search, we only care that whether any results have been retrieved or not.

On the other hand, accuracy is the direct and detailed judgment for ranking algorithm performance as we can know scores for each word. For example, when we need to choose top 5 in the ranking result, accuracy evaluation can realize this goal.

# 2.c.i

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| MMR-p | p=1 | p=0.5 | p=0 | GloVe |
| Mean | 0.36016572 | 0.39630103880 | 0.395614706 | 0.4935783685 |
| SD | 0.41878408 | 0.42689735973 | 0.426376241 | 0.4319456254 |

Besides the GloVe, we find that when p = 0.5, we have the largest MMR.

# 2.c.iii

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| T-test | P=1 | P=0.5 | P=0 | GloVe |
| P=1 |  | stat=-14.521581, pv= 5.2475e-47 | stat=-10.591004,  pv= 5.4808e-26 | stat=25.05357, pv= 5.333e-132 |
| P=0.5 |  |  | stat=-0.34607, pv= 0.72930 | stat=-18.5458, pv=9.2602e-75 |
| P=0 |  |  |  | stat=-18.3654, pv=2.1967e-73 |

Since p-value of GloVe and SVD with p=0.5 is 9.2602 and is smaller than 0.05, we can say that GloVe, as the best of four, is significantly better than the SVD with p=0.5, as the second best of four.

# 3.b

|  |  |
| --- | --- |
|  | Mean win ratio |
| P=1 | 0.55172 |
| P=0.5 | 0.51724 |
| P=0 | 0.53448 |
| GloVe | 0.48276 |

From the table, SVD with has the best performance. However, in my expectation GloVe should perform best.

# 3.c

From the table we know that the best two winners are SVD with and SVD with . And their t-test result is

Ttest\_relResult(statistic=0.21641738182341069,

pvalue=0.82943509779099567)

Since the p-value=0.8294 is larger than 0.05, the best performing method is not significantly better than the second performing method.

# 3.d

Yes, such occasion occurs to me. For example, given the word “Miss”, it is difficult to decide the best analogy from the choices of “Jackie”, “William” and “Betty”.

My solution to this situation is to allow user have ratings on each choice and use that rating to decide the best choice.

# 3.e

Yes, the same as the situation above.

The solution can be as similar as above to let user give specific ratings on their choice.