Report by: Anand Prashar

NLP Homework 5 – q8

Q:

Design two new relation groups and come up with three questions for each group (six total).

All words for all

questions should appear in your word2vec vocabulary. On CrowdMark, below, report how well the two sets

of embeddings perform on your test questions. You're encouraged to be adversarial so that the embeddings

for your relation groups might get an accuracy of zero! Discuss any interesting observations you have made

in the process.

A:

2 new groups-

**: ABBR**

info information thx thanks

dr doctor tv television

mt mountain cuz because

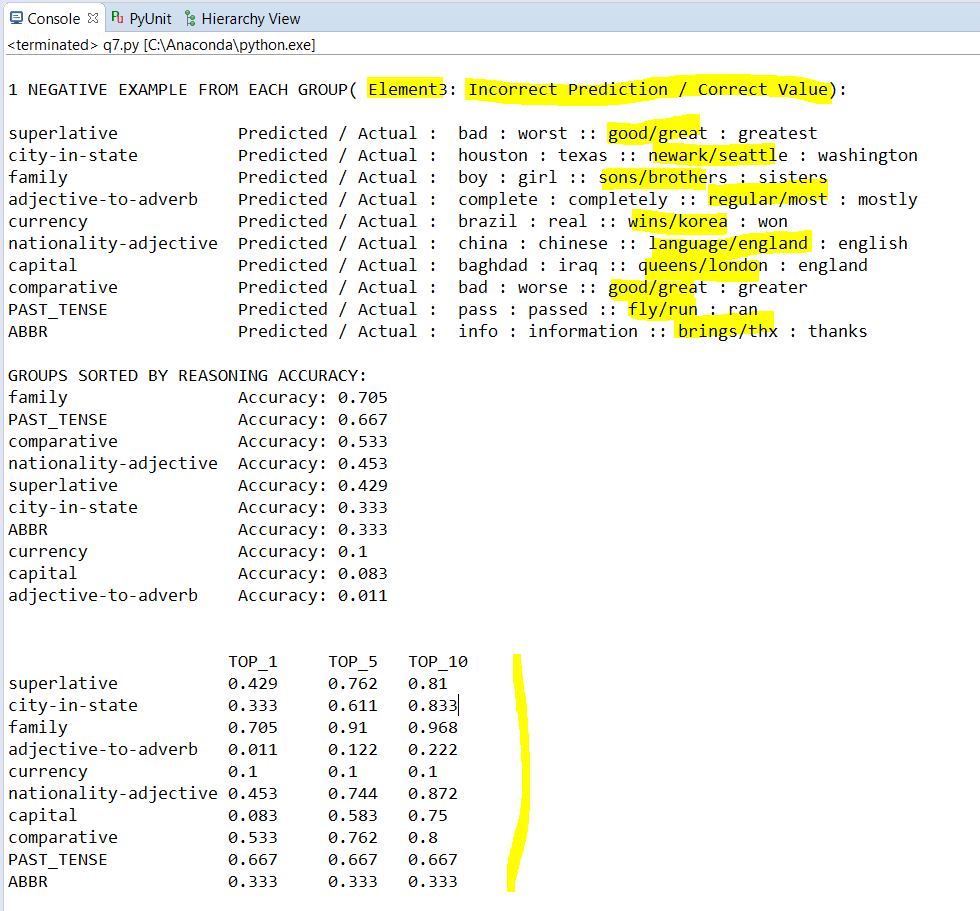
**: PAST\_TENSE**

win won go gone

play played live lived

pass passed run ran

Accuracy-



In last chart, we see the TOP\_1, TOP\_5, TOP\_10 accuracy of new groups: ABBR, PAST\_TENSE.

The accuracies we see for them can be attributed to terms which tend to appear frequently in the same context. It was difficult to be find 0 accuracy in the domain I had chosen due to lack to words in VOCAB file.

However I understand this scenario would arrive when the ACTUAL value does not exist in any of the top 10 words returned by my algorithm; which can mean the

My ACTUAL vector lies out of range of W4 + ( DIFFERENCE zone between W1 AND W2 )