Yufei Lin

May 12th 2019

Computer Linguistics

Project 5 Summary

The highest accuracy we obtain from procedure is 14.38 with a 1 1 input meaning we have a normalization and a Manhattan distance. Therefore, it foreshadows the importance of having a normalization of our vectors. It is also because the vector files are normalized and it is easier for us to calculate the distance with higher accuracy if we use a normalized vector.

The “capital-common-countries” one did the best because I think it contains more common words than all other files in the given vector model. Also, I think it is also because the capitals are usually bonded with their own countries and unique of their own kind. Therefore, it is hard to find a much closer relationship in any other relationships.

Furthermore, I think Manhattan distance works better. It usually have the highest total average of accuracy. I think it is because it makes more sense in terms of word similarity. It gives a distance from one place to another in a form of a vector. Then, it means a more coherent result for calculating similarities.