Name: Nikola Baci
Date: May 5th, 2021
Topic: Word2Vec

Report

This word2vec model is designed following the guidelines of Dr. Ganesan. The model takes in a gzip file and outputs all word vectors and prints the requirements. It uses a neural network with one hidden layer, and takes approximately 5-8 minutes to train.

First run of the model, we use the following parameters:

1. size = 150
2. window = 5
3. min_count = 2
4. iter = 10

The results are:

- Similarity between ***dirty*** and ***clean*** is 39.004794%
- Similarity between ***big*** and ***dirty*** is 26.532975%
- Similarity between ***big*** and ***large*** is 47.743818%
- Similarity between ***big*** and ***small*** 51.39071%

Five most similar words to ***polite*** are:

1. ***respectful*** with similarity score of 71.42271399497986%
2. ***gracious*** with similarity score of 66.87819957733154%
3. ***courteous*** with similarity score of 65.27359485626221%
4. ***attentive*** with similarity score of 61.01087331771851%
5. ***thoughtful*** with similarity score of 60.03850102424622%

Five most similar words to ***orange*** are:

1. ***ventura*** with similarity score of 49.77732300758362%
2. ***emerald*** with similarity score of 49.60307776927948%
3. ***lackawanna*** with similarity score of 49.04917776584625%
4. ***dutchess*** with similarity score of 48.94387423992157%
5. ***peach*** with similarity score of 47.929438948631287%

Second run of the model, we use the following parameters:

1. size = 50
2. window = 2
3. min_count = 2
4. iter = 10

The results are:

- Similarity between ***dirty*** and ***clean*** is 41.984832%
- Similarity between ***big*** and ***dirty*** is 36.361307%
- Similarity between ***big*** and ***large*** is 62.136835%
- Similarity between ***big*** and ***small*** 69.22639%

Five most similar words to ***polite*** are:

1. ***timid*** with similarity score of 78.93770933151245%
2. ***respectful*** with similarity score of 78.45202684402466%
3. ***gracious*** with similarity score of 75.00736713409424%
4. ***forthright*** with similarity score of 72.5963830947876%
5. ***courteous*** with similarity score of 72.43958711624146%

Five most similar words to ***orange*** are:

1. ***fringed*** with similarity score of 65.87183475494385%
2. ***angostura*** with similarity score of 65.85032343864441%
3. ***upturned*** with similarity score of 62.91470527648926%
4. ***alamo*** with similarity score of 62.82759308815002%
5. ***blue*** with similarity score of 62.72799968719482%