

Assignment-3 : Word Similarity

Bhavy Khatri(150186)
Shubham Kumar Bharti(15807702)

5th April 2019

Question 1:

To calculate Normalized Google Distance between a pair of words we made successive word queries to Google search engine and retrieved the search frequency result of the query words from the search result. For each word pairs we perform three search queries to google(word1, word2, word1+word2). We places a fixed timer between our search requests to avoid blocking of our IP.

Question 2:

We implemented the Normalised Google Distance(NGD) function based on the formula discussed in the class. We further min-max scaled the NGD results to a range of 10. Finally we plotted the NGD vs Human Similarity rating score for each pair of words in the dataset to reproduce the scatter plot.

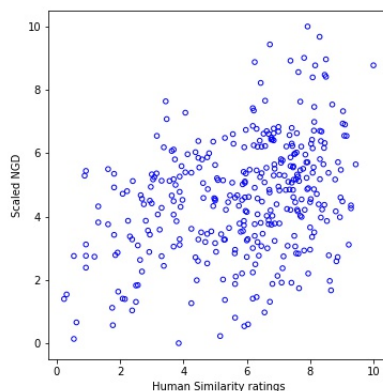


Figure 1: NGD vs Human Similarity Plot

Question 3:

In this part, we used the word2vec-api to query for the semantic similarity between two words based on the word2vec embedding. We further plotted the obtained word2vec similarity score of each word pair in the dataset against the human similarity scores of that word pair.

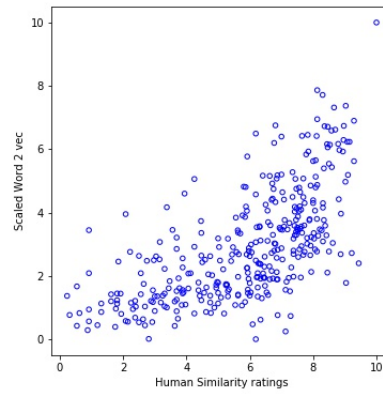


Figure 2: Word2Vec vs Human Similarity Plot