Project: Latent Semantic Analysis for Summary Generation

Subject: Artificial Intelligence.

Professor: Dr. Fahad

Group:

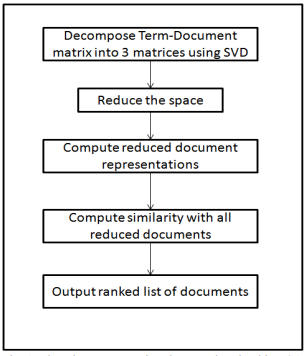
M. Fahad Qureshi (17k-3925) (Team Lead)

Talha Jamal (17k-3877)

1. Naveed (17k-3885)

Latent Semantic Analysis:

It’s a technique that is used to find hidden meaning from a document. It comes under unsupervised learning. We will be using LSA to find topics from a corpus. The way LSA works is that it first creates a bag of words from a corpus then it performs linear algebraic calculations to find three matrices that includes finding Eigen vectors to say the least. This is called decomposition of our term document matrix (bag of words).These matrices represent us with some hidden features which can be analyzed further as shown below in diagram.



Relevance Measure

Another method that is used is Relevance Measure. In this we summarize the text using the relevancy of a particular sentence in the given text. The first challenge is to break the given text into separate sentence and the tokenization of the sentences is completed the next challenge is to remove the stop words from the sentences without affecting the original sentence then we need to build the sentence vector for every sentence and store it in the suitable data structure then we need to take the dot product of the sentence vector Si with document vector D excluding the sentence i where i is the particular sentence of whom the dot product with the document is taken.

                  Score of Sentence (i)= D.Si

Then we after storing the score of each sentence we need rank them in order of the score in descending and return the required no of sentences to the user with the highest score.