**Food recommendation according to nutrients**

For recommendation use case we are planning to recommend user food items as per nutrition he is following or meal plan he is taking.

We have collected data from USDA national nutrient data and made it in required format after some clean up. We dumped it to mysql as structured format. We have over 9,000 food items and 25 different nutrients.

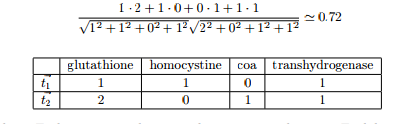
We are using cosine similarity function for finding the similarity between users’ selected or consumed item and the other items available to the database. We compare individual item with other item to make pair of two rows. We calculate cosine similarity between these two rows and store result as (key, pair) value where key is id of two food items and value is cosine similarity which lies between 0 to 1. Here, we don’t have any negative values so it lies between 0 to 1 else it would be between -1 to 1.

Below is the simple example of cosine similarity.

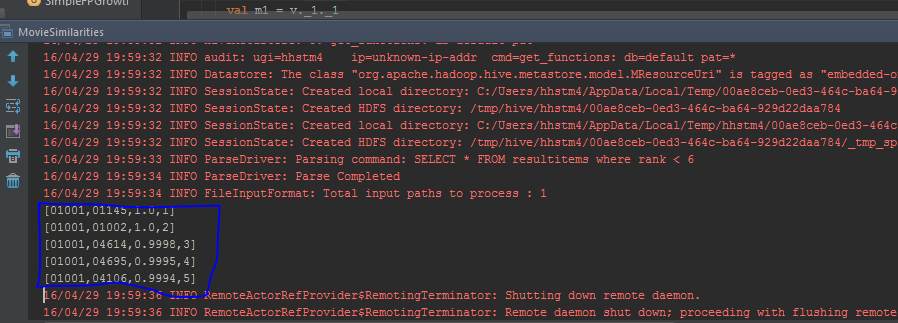


To calculate cosine similarity between two texts t1 and t2, they are transformed in vectors as shown in the Table 1. Each word in texts defines a dimension in Euclidean space and the frequency of each word corresponds to the value in the dimension. Then, the cosine similarity is measured by using the word vectors as in equation

For example, a cosine similarity can be computed as below for two texts



We ran this using apache spark core and spark SQL. We used JDBC connecter for java to store result back to the mysql database. Below is the snap shot for one item with its most similar 5 items.



In result, first column is the ID of item for which we are finding the most similar items. Second column is the Id of matching item, 3rd column is similarity and 4th column is rank.

The main idea is that its huge database with bunch of items. User is not able to go through each item and check nutrient contains inside it. So this is the way of providing him matching items of his interest of nutrients to make him healthy.