CSE 554/484
Homework 03

Word2vec yazılımını kullanarak her word için vektörleri oluşturdum. Bu vektörler dense vektörler olarak oluşturulmuştur, boyutu 200'dür her vektörün. Kullanıcıdan min, max, avg gibi metodlar girmesini isteyerek girilen değere göre ve her kelimenin vektöründen bu metodlara göre bir değer elde ettim. Ve bu değerleri Rocchio ve KNN algoritmaları için kullandım.

Class	Magazin	Siyasi	Spor	Sağlık	Ekonomi
Number	1	2	3	4	5

1-)Rocchio

Protype vektörü oluşturdum ve query vektörü ile cosSim(d,pi) hesaplaması yaptım. En büyük cosSim değeri olan classification değerinin buldum.Similarity score ise cosSim değeridir.

File Name	9.txt	421.txt	276.txt
MIN Method	5	4	1
MAX Method	5	4	1
AVG Method	5	4	5
True Class	5	4	1

2-)K Nearest-Neighbour

Data set içerisindeki tüm dökümanların vektörleri ile query vektörünü cosSim işlemine soktum. Ardından bu değerleri sort ettim. Sonra sort edilmiş değerlere göre ilk k değerinin içerisinde en çok hangi classification var ise o classification değerinin döndürdüm.

File Name	9.txt	421.txt	276.txt
KNN3 MIN Method	5	4	1
KNN5 MAX Method	5	1	1
KNN3 AVG Method	5	5	1
True Class	5	4	1

Gereklilikler

1. Word2vec kodu make edilmeli.
2. Ardından run edilmeli fakat ben bu işlemleri önceden yaptığım için output olarak oluşturduğum trwikivectors.txt doyasını boyutunun yüksek olmasından da dolayı demo da yanımda getireceğim.
3. 2 adet data set'imiz bulunmaktadır. 1150Haber.rar dosyasını ödev dosyasında paylaşacağım fakat diğerini yüksek boyutlu olmasından dolayı demoda yanımda getireceğim.

Süleyman Balaban
121044014