

Soft Similarity & Soft Cosine Measure.

Differences.

Soft Similarity (so called) is proposed similarity measure.

Soft cosine measure is kind of generalizing the cosine similarity, which are well-known, but Soft Similarity measure is equal to the standard similarity. They are both take their place at VSM (Vector Space Module), which features the traditional cosine measure and traditional similarity measure consider as independent. But it often happens when dimensions are not independent.

In this case authors of the article propose JS and SC measures to use, which generalize the concepts of cosine measure and similarity.

Broadly speaking JS and SC are hardly to compare, because Soft Similarity is ~~not~~ modified version of the Soft Cosine measure.

SC is transformed way of calculation of similarity. In VSM, it's not ignoring similarity of features.

So, if all this (representing new dimension in VSM by making a matrix of similarity between pairs of features) will be applied to similarity while

using ML algorithms in VSM, then it'll transform into Soft Similarity.

So, we take vectors, which are representing the frequencies of the words we are interested in (belonging to some exact field) in some documents and measure the similarity between those, e.g.:

$$a = (1, 3, 1, 2, 0, 1, 1, 0).$$

$$b = (2, 1, 1, 0, 1, 0, 0, 2).$$

Accordingly to the words in some texts about Christmas:

tree, happy, Santa, gift, eve,
snow, lights, tangerines.

for demonstration how it

works let's take ~~the first~~
~~first element~~. cosine(a, b)
to measure the similarity
between those texts.

$$\underline{\text{cosine}(e_i, e_j) = \text{sim}(f_i, f_j) = S_{ij}}$$

So, to calculate soft cosine matrix, we consider each
pair of features as a new one,
which include a weight of
significance, at the same
time presenting the similarity
of the 2 features.

soft cosine
(run out of time n^4)