A Generalized Similarity Measure ISSN 1751-3030

Author:
Ramesh Chandra Bagadi
Data Scientist
INSOFE (International School Of Engineering), Hyderabad, India.
rameshcbagadi@uwalumni.com
+91 9440032711

Technical Note

Abstract

In this research Technical Note the author has presented a novel method of finding a Generalized Similarity Measure between two Vectors or Matrices or Higher Dimensional Data of different sizes.

Theory

Considering two different vectors of different sizes namely

A_{1xm} and B_{1xn}, we first find the Proximity Matrix between elements of the given vectors wherein the Proximity Matrix is given by

$$P_A = \begin{bmatrix} d(1,1) & d(1,2) & d(1,3) & .\ .\ .\ . & d(1,(m-1)) & d(1,m) \\ d(2,1) & d(2,2) & d(2,3) & .\ .\ .\ . & d(2,(m-1)) & d(2,m) \\ d(3,1) & d(3,2) & d(3,3) & .\ .\ .\ . & d(3,(m-1)) & d(3,m) \\ . & . & . & .\ .\ .\ . & . & . \\ . & . & . & .\ .\ .\ . & . & . \\ . & . & . & .\ .\ .\ . & . & . \\ . & . & . & .\ .\ .\ . & . & . \\ d((m-1),1) & d((m-1),2) & d((m-1),3) & .\ .\ .\ . & d((m-1),(m-1)) & d((m-1),m) \\ d(m,1) & d(m,2) & d(m,3) & .\ .\ .\ . & d(m,(m-1)) & d(m,m) \end{bmatrix}..$$

and

$$P_B = \begin{bmatrix} d(1,1) & d(1,2) & d(1,3) & .\ .\ .\ . & d(1,(n-1)) & d(1,n) \\ d(2,1) & d(2,2) & d(2,3) & .\ .\ .\ . & d(2,(n-1)) & d(2,n) \\ d(3,1) & d(3,2) & d(3,3) & .\ .\ .\ . & d(3,(n-1)) & d(3,n) \\ . & . & . & .\ .\ .\ . & . & . \\ . & . & . & .\ .\ .\ . & . & . \\ . & . & . & .\ .\ .\ . & . & . \\ . & . & . & .\ .\ .\ . & . & . \\ d((n-1),1) & d((n-1),2) & d((n-1),3) & .\ .\ .\ . & d((n-1),(n-1)) & d((n-1),n) \\ d(n,1) & d(n,2) & d(n,3) & .\ .\ .\ . & d(n,(n-1)) & d(n,n) \end{bmatrix}..$$

d indicates the distance measured in some metric (default = Eucleadean)

We then find the Norm Of P_A as $\|P_A \cdot P_A\|$. For the Euclidean case, it is given by $\|P_A \cdot P_A\| = \sum_{j=1}^{m}\sum_{i=1}^{m} P(i,j) \cdot P(i,j)$. Also, $m < n$. Similarly, we compute the Norm of P_B as $\|P_B \cdot P_B\|$. For the Euclidean case, it is given by $\|P_B \cdot P_B\| = \sum_{j=1}^{n}\sum_{i=1}^{n} P(i,j) \cdot P(i,j)$. We then find the ratio $k_1 = \frac{\|P_B \cdot P_B\|}{\|P_A \cdot P_A\|}$. Similarly, we find another ratio $k_2 = \frac{\|P_B{}^2 \cdot P_B{}^2\|}{\|P_A{}^2 \cdot P_A{}^2\|}$ and so on so forth, till $k_n = \frac{\|P_B{}^{g/2} \cdot P_B{}^{g/2}\|}{\|P_A{}^{g/2} \cdot P_A{}^{g/2}\|}$ where $g = 2^n$.

We also find We now consider a fictitious Vector $A_{B_{1xn}}$, i.e., Vector A in the basis of Vector B, colloquially speaking. Let this be $A_{B_{1xn}} = [c_1 \quad c_2 \quad c_3 \quad . \quad . \quad c_{n-1} \quad c_n]$. Now, for this, vector, we find the Proximity Matrix $P_{A_{B_{1xn}}}$ and now assert that $k_n = \frac{\|P_B{}^{g/2} \cdot P_B{}^{g/2}\|}{\|P_{A_{B_{1xn}}}{}^{g/2} \cdot P_{A_{B_{1xn}}}{}^{g/2}\|}$. This gives us n number of equations from which we can solve for elements of $A_{B_{1xn}}$. Now, we can find distance between $A_{B_{1xn}}$ and B_{1xn} and can also consequently find the Similarity co-efficient between them. We can also, repeat this procedure using the normalized values of the vectors A_{1xm} and B_{1xn}. The motivation to cook this kind of procedure is that Norms are usually invariant under Dimension Upgrading or/ and Downgrading Transformations {Hadamard Theorem}. In the same fashion as detailed above, we can repeat this procedure for Matrices or Higher Dimensional Data of differing sizes.

References

http://www.philica.com/advancedsearch.php?author=12897

http://www.vixra.org/author/ramesh_chandra_bagadi