Modification To The Scaling Aspect In Gower's Scheme Of Calculating Similarity Coefficient

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 have presented a tiny modification to the Numeric Variables Scaling Aspect In Gower's Scheme of calculating Similarity Coefficient.

Theory

One can note that in Gower's Scheme for computing Similarity Coefficient, the Numeric Variables are Scaled as follows:

$X_{ik} \mapsto \left\{ \frac{X_{ik} - Min(X_{ik})}{Max(X_{ik}) - Min(X_{ik})} \right\}$ where X_{ik} is the i^{th} data point of the k^{th} variable such that $(0 \le X_{ik} \le 1)$. However, if the Data Points are given, ordered along as a Time Series, i.e., the order of points is commensurate with the evolution of time, we can consider scaling in the following fashion:

$X_{ik} \mapsto \left\{ \frac{\int_{m=1}^{f^{-1}(X_{ik})=i} f_{X_{ik}}(m) dm_{X_{ik}}}{\int_{m=1}^{n} f_{X_{ik}}(m) dm_{X_{ik}}} \right\}$ where $f_{X_{ik}}(m)$ is a one-one function from $h = \{1,2,3,4,.....,n-1,n\} \mapsto g = \{X_{1k}, X_{2k}, X_{3k}, ..., X_{(n-1)k}, X_{nk}\}$ wherein g is the given data points k^{th} variable values ordered along as a Time Series.

m	1	2	3	4	.	.	.	*i*	.	.	*(n-1)*	*n*
$f_{X_{ik}}(m)$	X_{1k}	X_{2k}	X_{3k}	X_{4k}	.	.	.	X_{ik}	.	.	$X_{(n-1)k}$	X_{nk}

References

1. http://www.philica.com/advancedsearch.php?author=12897
2. http://www.vixra.org/author/ramesh_chandra_bagadi