Archives of Current Research International

Volume 24, Issue 6, Page 586-599, 2024; Article no.ACRI.120452
ISSN: 2454-7077

Multidimensional Scaling Method and Some Practical Applications

B. Manjunatha [a*], Appaji Pundalik Naik [a], K. R. Mahendra [a], Manju Prem S. [b], Gunashekhar H. [c], N. R. Kiran [d], Damodhara G. N. [e] and Karthik R. [f]

[a] The Graduate School, ICAR-Indian Agricultural Research Institute, Pusa, New Delhi, India.
[b] College of Agriculture, Vellayani, Kerala Agricultural University, Kerala, India.
[c] ICAR- National Dairy Research Institute, Southern Regional Station, Bengaluru, India.
[d] University of Agricultural Sciences, Bengaluru, India.
[e] Manasa Gangotri Campus, University of Mysore, Mysore, India.
[f] Chaudhary Sarwan Kumar Himachal Pradesh Krishi Vishvavidyalaya, Palampur, India.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: https://doi.org/10.9734/acri/2024/v24i6814

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/120452

Original Research Article

Received: 05/06/2024
Accepted: 07/08/2024
Published: 10/08/2024

ABSTRACT

Multi-Dimensional Scaling (MDS) is a data visualization method that identifies clusters of points by representing the distances or dissimilarities between sets of objects in a lower-dimensional space. This paper explores the theoretical concepts of MDS, various methods of implementation, and the analytical processes involved. Emphasis is placed on the "Stress" function, a goodness-of-fit metric that quantifies the discrepancy between distances in high-dimensional and lower-dimensional spaces. Practical examples and detailed procedures for implementing MDS using MS-Excel and R

**Corresponding author: Email: manju777dsg@gmail.com;*

***Cite as:** Manjunatha, B., Appaji Pundalik Naik, K. R. Mahendra, Manju Prem S., Gunashekhar H., N. R. Kiran, Damodhara G. N., and Karthik R. 2024. "Multidimensional Scaling Method and Some Practical Applications". Archives of Current Research International 24 (6):586-99. https://doi.org/10.9734/acri/2024/v24i6814.*

are provided to enhance understanding. The paper also discusses the use of Scree-plots for determining the optimal number of dimensions. Applications of MDS in different fields, including marketing, ecology, molecular biology, and social networks, are presented with examples on Perceptions of Nations data and Morse code confusion data. Additionally, as a significant contribution, a case study on factors affecting agricultural productivity is included. The versatility and utility of MDS in simplifying complex data and facilitating better decision-making are demonstrated through these practical applications and software implementations.

Keywords: Stress function; proximity; dissimilarities; scree-plot.

1. INTRODUCTION

Multidimensional Scaling (MDS) is a technique used to visualize the distances or dissimilarities between sets of objects, such as colors, faces, or map coordinates [1]. In an MDS plot, objects that are similar (with shorter distances) are placed closer together, while dissimilar objects (with longer distances) are placed further apart. The term "scaling" is derived from psychometrics, where abstract concepts are assigned numerical values based on a specific rule. For instance, an individual's attitude toward global warming might be quantified on a scale from 1 (does not believe in global warming) to 10 (firmly believes in global warming), with intermediate values for varying attitudes. MDS encompasses a range of statistical methods that spatially represent the structure of data, making it easier to visualize and interpret. This method is particularly useful for visualizing complex relationships and is often associated with mapping techniques. Consider a scenario where you have a map of a geographical region with several cities and towns. A table showing the distances between these locations can be created, indicating how close each pair of cities is. The proximity can be defined in various ways, such as straight-line distance or shortest travel distance, or it can represent a measure of association, like the absolute value of a correlation coefficient.

Reversing this process, imagine being given a table of distances and tasked with recreating the original map. This is analogous to the general problem that MDS addresses. MDS creates a spatial representation based on proximity data, even when the number of dimensions required is not known beforehand. Determining the correct number of dimensions is crucial and is typically done using techniques like Scree plots. However, as the number of dimensions increases, the complexity of visualization and interpretation also increases. Even three-dimensional representations can be difficult to display on paper and understand, and using four or more dimensions can make MDS less effective for making complex data comprehensible.

Classical scaling, the traditional MDS method, assumes that dissimilarities are exact Euclidean distances without any transformation. The objective function used in classical scaling commonly referred to as "Stress." To minimize stress, a strategy called Scaling by Majorizing a Complicated Function (SMACOF) is employed, which uses majorization. While majorization itself is not an algorithm, it provides a framework for developing optimization algorithms.

2. LITERATURE REVIEW ON MDS

MDS introduced by Torgerson [2] and further developed by Kruskal and Wish [1], Classical MDS, as described by Torgerson [2], assumes that the input data are dissimilarities that can be directly transformed into Euclidean distances. Non-metric MDS, developed by Kruskal (1964), allows for the analysis of ordinal data, ensuring that the rank order of the distances in the low-dimensional space matches that of the original dissimilarities.

MDS has been widely applied across various fields. In psychology, it is used to map perceptual and cognitive processes [3]. Ecology utilizes MDS for visualizing species distributions and environmental gradients [4]. Additionally, MDS is used in bioinformatics to study protein structures and genetic data [5]. De Leeuw and Mair [6] offer a comprehensive overview of MDS, describing different MDS versions and detailing an R software package named SMACOF that integrates all known MDS procedures.

Several studies have continued to expand the applications of MDS. For example, Pacini et al. [7] combined MDS with cluster analysis to describe the diversity of rural households, while Liu et al. [8] used MDS for information visualization, highlighting its ability to simplify complex datasets for better interpretability.

In marketing, MDS can be used to derive "product maps" of consumer choice and product preference, such as for automobiles and beer, allowing relationships between products to be discerned. In ecology, it provides "environmental impact maps" of pollution, like oil spills and sewage pollution, on local communities of animals, marine species, and insects. This method has been used to study the complex correlations between global temperature time-series, offering a graphical representation of climatic similarities between regions globally [9]. In fisheries, MDS has been applied to study the performance of 18 marine fishery resources in Maharashtra, India [10]. In molecular biology, it helps reconstruct the spatial structures of molecules, such as amino acids, and interpret their interrelations, similarities, and differences, leading to the construction of a 3D "protein map" for a global view of the protein structure universe. In social networks, MDS aids in developing "telephone-call graphs," where vertices represent telephone numbers and edges correspond to calls between them, which can help recognize instances of credit card fraud and detect network intrusions. MDS has been applied in meteorological forecasting [11], content-based retrieval [12], market structure visualization [13], and high-dimensional space exploration [14].

MDS encompasses a range of algorithms designed to find an optimal low-dimensional configuration based on proximity data. Primarily used for data visualization, MDS helps identify clusters of points, where points within the same cluster are closer to each other compared to points in different clusters. Various books provide in-depth discussions on MDS techniques, including works by Kruskal and Wish [1], Coxon [15], Hair et al. [16], Cox and Cox [17], Borg and Groenen [18], Izenman [19], de Leeuw and Heiser [20], Ding [21].

3. MATERIALS AND METHODS

3.1 Correlation Coefficient

A correlation coefficient is a statistical measure that quantifies the strength and direction of the linear relationship between two variables. The most commonly used correlation coefficient is the Pearson correlation coefficient, which ranges from -1 to 1. A value of 1 indicates a perfect positive linear relationship, -1 indicates a perfect negative linear relationship, and 0 indicates no linear relationship. The covariance of two variables divided by the product of their standard deviations gives Pearson's correlation coefficient. It is usually represented by ρ (rho).

$$\rho(X,Y) = \frac{\mathrm{COV(X,Y)}}{\sigma_X\sigma_X} = \frac{\mathrm{E(X-\mu_{\mathit{X}})(Y-\mu_{\mathit{Y}})}}{\sigma_X\sigma_Y} \quad (1)$$

Correlation coefficients are crucial in understanding the degree to which variables move together and are used extensively in various fields. In the context of MDS, incorporating correlation coefficients helps in accurately representing the similarities or dissimilarities between data points, ensuring that the visualization reflects the true relationships within the data.

3.2 Correlation Pairwise Metric

A correlation pairwise metric extends the concept of correlation coefficients by focusing on the relationships between pairs of data points. This metric considers the pairwise correlation between all possible pairs within a dataset, providing a comprehensive view of the interdependencies among variables. By calculating these pairwise correlations, one can construct a similarity or dissimilarity matrix that serves as the foundation for techniques like MDS. This matrix captures the intricate patterns of association between data points, allowing for a more nuanced and accurate representation of the data in a lower-dimensional space. Utilizing a correlation pairwise metric in MDS enhances the fidelity of the resulting configuration, leading to better insights and more meaningful interpretations.

3.3 Proximity Matrices

"The *proximity* measure gives the "closeness of two entities, which can be defined in a number of different ways. In many types of experiments, proximity data are obtained from a group of subjects, each of whom make similarity (or dissimilarity) judgements on all possible unordered pairs of n entities *i.e.* $m = \binom{n}{2} = \frac{1}{2}n(n-1)$. It is irrelevant whether the similarities or dissimilarities are used as our measure of proximity between two entities. In other words, "closeness" of one entity to another could be measured by a small or large value. The only thing that matters when carrying out MDS is that there should be a monotonic relationship (either increasing or decreasing) between the "closeness" of two entities and the corresponding similarity or dissimilarity value" [22]. Anyway, usually similarities are converted into dissimilarities through a monotonically decreasing transformation. Consider a particular collection of *n* entities. Let δ_{ij} represent the dissimilarity of the *i*th entity to the *j*th entity. The *m* dissimilarities, $\{\delta_{ij}\}$, are arranged into $(m \times m)$ square matrix,

$$\Delta = (\delta_{ij}) \tag{2}$$

called a *proximity matrix*. In case of dissimilarities the proximity matrix is usually displayed as a lower-triangular array of non-negative entries, with the understanding that the diagonal entries are all zeroes and that the upper-triangular array is a mirror image of the given lower-triangle (i.e., matrix is symmetric). In other words, for all $i, j = 1, 2, \ldots, n$,

$$\delta_{ij} \geq 0, \qquad \delta_{ii} = 0,\ \delta_{ji} = \delta_{ij}. \tag{3}$$

3.4 Stress Function

So far, the task of MDS was defined as finding a low-dimensional configuration of points representing objects such that the distance between any two points matches their dissimilarity as closely as possible. Of course, it is preferred that each dissimilarity should be mapped exactly into its corresponding distance in the MDS space. But empirical data always contain some component of error given by $f(\delta_{ij}) - d_{ij}$,where d_{ij}'s are the computed Euclidean distances between the objects in the arbitrarily constructed plot. Since positive and negative discrepancies are equally undesirable, the sum of squared errors for all proximities is taken, which yields the formula.

$$Raw\ stress = \sum_i \sum_j (\delta_{ij} - d_{ij})^2 \text{, by taking } f(\delta_{ij}) = \delta_{ij} \tag{4}$$

To counter the effect of scale-dependency, the raw stress is normalised to have the general form,

$$\left\{\sum_{i<j} w_{ij} (\delta_{ij} - d_{ij})^2\right\}^{1/2} \tag{5}$$

where the $\{w_{ij}\}$ are weights chosen by the user. The most popular normalization is where $w_{ij} = \left(\sum_{i<j} d_{ij}^{\ 2}\right)^{-1}$, so that the raw stress become the Stress1 i.e.,

$$Stress1 = S = \left\{\frac{\sum_{i<j}(\delta_{ij} - d_{ij})^2}{\sum_{i<j} d_{ij}^{\ 2}}\right\}^{\frac{1}{2}}, \tag{6}$$

where it is understood that the summations in both the numerator and denominator of *S* are computed for all $i, j = 1, 2, \ldots, n$ such that $i < j$. The stress-1 value (*S*) lies between 0 and 1. “The stress criterion *S* (more commonly known as *Kruskal’s stress formula one* or Stress-1) can be interpreted as a loss function that depends upon the configuration points and the disparities and measures how well a particular configuration fits the given dissimilarities. It is worth noting that certain authors refer to the stress function as S^2. A variant, stress formula 2, differs only in that different weights are used” [22].

3.5 Data Visualization

Data visualization bridges complex data sets and intuitive understanding, revealing patterns and trends that raw data may obscure. Techniques like bar charts, scatter plots, heat maps, and advanced methods such as MDS are crucial in this process. Dimensionality reduction simplifies data while retaining essential features, enhancing visualization [23,24]. Algorithms for dimensionality reduction improve interactive visualization, making data exploration more intuitive [25]. Comprehensive approaches to data visualization, including these methods, are well-documented [26]. Effective visualizations not only aid in analysis but also communicate findings to a broader audience, making them indispensable in research.

3.6 Scree-Plot

A scree-plot is a method for determining the optimal number of components useful to describe the data in the context of MDS. To create a Scree-plot, analysts scale the data several times (with higher dimensionality each time), and plot the stress values as a function of dimensions [27]. Here, the stress values are plotted on y-axis and the number of dimensions are plotted on x axis as shown in Fig. 1. The aim is to evaluate the number of dimensions required to capture most information contained in the data. A point where the slope of the curve changes sharply referred to as the “elbow” of the plot determines the optimal number of dimensions to describe the data.

Normally, a complex set of relationships can be scanned at a glance with the aid of visual representation provided by MDS. Since maps on paper are two-dimensional objects, this translates technically to finding an optimal configuration of points in 2-dimensional space. However, limiting to a two-dimensions may lead

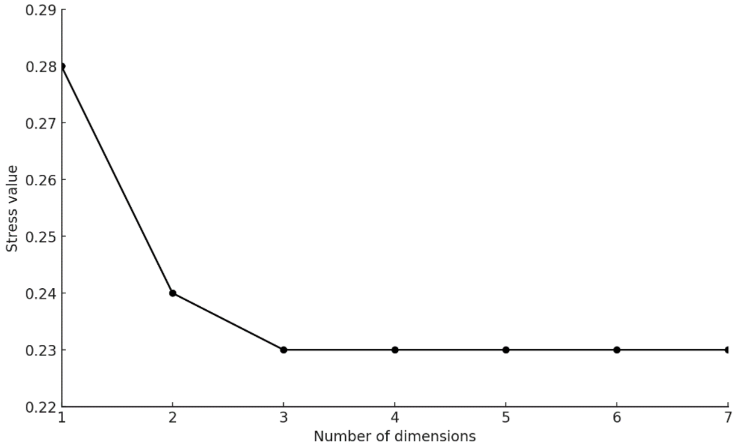

Fig. 1. Scree-plot

to a very poor, highly distorted, representation of the data. In order to overcome this limitation the number of dimensions may be increased (if needed) but there are difficulties in representing, comprehending and estimating the parameters for the higher dimensions. Four or more dimensions render MDS virtually useless as a method of making complex data more accessible to the human mind.

3.7 Smacof

The Stress function that measures the deviance of the distances between points in a geometric space and their corresponding dissimilarities is to be minimised. An easy and powerful minimization strategy is the principle of minimizing a function by iterative majorization. Because for finding the minimum of a function $f(x)$, it is not always enough to compute the derivative $f'(x)$, set it equal to zero, and solve for x. Sometimes the derivative is not defined everywhere, or solving the equation $f'(x) = 0$ is simply impossible. For such cases, other mathematical techniques are referred. A useful method consists of trying to get increasingly better estimates of the minimum. It consists of a set of computational rules that are usually applied repeatedly, where the previous estimate is used as input for the next cycle of computations which outputs a better estimate. In the SMACOF algorithm, the central idea of the majorization method is to replace iteratively the original complicated function f(x) by an auxiliary function g(x, z), where z in g(x, z) is some fixed value. The function g has to meet the following requirements to call g(x, z) a majorizing function of f(x). The auxiliary function g(x, z) should be simpler to minimize than f(x). For example, if g(x, z) is a quadratic function in x, then the minimum of g(x, z)over x can be computed in one step. The original function must always be smaller than or at most equal to the auxiliary function; that is, f(x) ≤ g(x, z). The auxiliary function should touch the surface at the so-called supporting point z; that is, f(z) = g(z, z).

Hence, the iterative majorization algorithm is given by

1. Set $z = z^0$, where z^0 is a starting value.
2. Find update x_u for which g(x_u, z) ≤ g(z, z).
3. If $f(z) - f(x_u) < \varepsilon$, then stop. (ε is a small positive constant)
4. Set $z = x_u$ and go to step 2.

Example 1. Application to Perceptions of Nations:

Now the procedure followed to obtain the two-dimensional MDS plots are discussed with an illustration using excel where the dissimilarity matrix is given. The data reflecting mean scores of 18 respondents' perceptions of overall dissimilarity between twelve nations on a scale ranging from 1 for "very familiar" to 9 for "very different" was ordered as a diagonal matrix of 66 pairs [1] as shown in Table 1. Since it has been decided on a two dimensional representation of the data, a starting configuration for the n objects in the two dimensions has to be set up (i.e. Co-ordinates x_n,, y_n are arbitrarily selected for each object) represented in the Table 2. The next step involves calculating the Euclidean distance between the objects. However the data points arranged within the graph will always be a difference between the actual values in our

Table 1. Representation of the dissimilarity matrix

	Brazil	Congo	Cuba	Egypt	France	India	Israel	Japan	China	Russia	USA	Yugoslavia
Brazil	0.0	4.17	3.72	5.56	4.28	4.5	5.17	5.5	6.61	5.94	3.61	5.83
Congo	4.17	0.0	4.44	4.0	5.0	4.17	5.67	5.61	5.61	5.61	6.61	6.61
Cuba	3.72	4.44	0.0	3.83	4.89	5.17	5.39	6.06	5.61	3.56	5.67	5.35
Egypt	5.56	4.0	3.83	0.0	4.22	5.17	4.33	4.5	5.17	4.61	3.06	4.72
France	4.28	5.0	4.89	4.22	0.0	5.17	4.33	4.78	5.17	3.94	4.28	5.0
India	4.5	4.17	5.17	5.17	5.17	0.0	5.0	5.5	5.33	4.5	3.06	5.0
Israel	5.17	5.67	5.39	4.33	4.33	5.0	0.0	4.45	4.89	4.83	4.72	4.56
Japan	5.5	5.61	6.06	4.5	4.78	5.5	4.45	0.0	5.17	4.83	3.06	5.0
China	6.61	5.61	5.61	5.17	5.17	5.33	4.89	5.17	0.0	4.39	6.44	4.72
Russia	5.94	5.61	3.56	4.61	3.94	4.5	4.83	4.83	4.39	0.0	3.28	3.94
USA	3.61	6.61	5.67	3.06	4.28	3.06	4.72	3.06	6.44	3.28	0.0	2.23
Yugoslavia	5.83	6.61	5.35	4.72	5.0	5.0	4.56	5.0	4.72	3.94	2.23	0.0

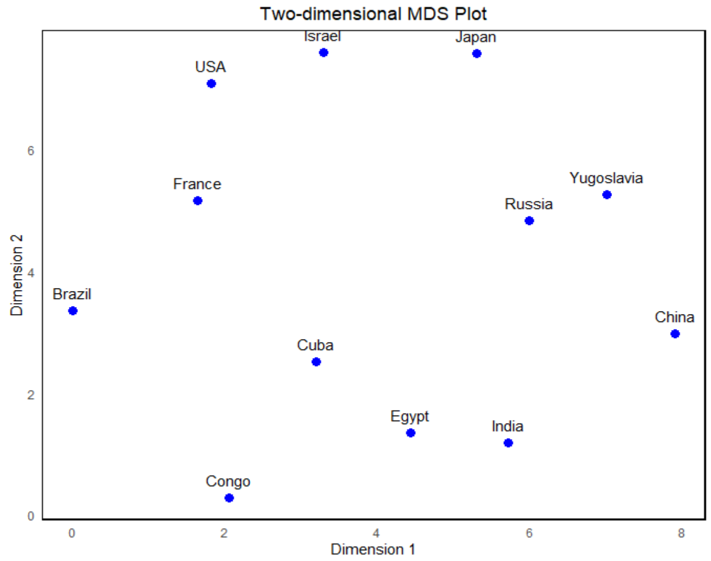

Fig. 2. Two-dimensional MDS plot Perceptions of Nations data

original diagonal matrix and the inter-point distances reflected and measured in the graph. Even after trying thousands of different arrangements there will still be errors and the best option is to minimize the cumulative errors in an arrangement, i.e. minimise the stress and show that as the best representation made out of the data provided. In other words, it would be a trial and error or iterative process of finding the best cumulative error minimising arrangement. Making use of the built in facility within Microsoft Excel (i.e. in Solver Add-in which is part of the Microsoft package) the solutions to such problems are found). Hence the optimised values of the co-ordinates (x and y, shown on Table 3) are obtained by minimising the stress. And finally, these points are plotted in a two-dimensional MDS plot as shown in the Fig. 2.

Table 2. Initial values of the co-ordinates

	x	y
Brazil	0.0	0.0
Congo	4.17	0.0
Cuba	0.0	3.72
Egypt	5.56	0.0
France	0.0	4.28
India	4.5	0.0
Israel	0.0	5.17
Japan	5.5	0.0
China	0.0	6.61
Russia	5.94	0.0
USA	0.0	3.61
Yugoslavia	5.83	0.0

Table 3. Optimised values of the co-ordinates after using SOLVER

	x	y
Brazil	0.001838	3.362963
Congo	2.060104	0.288286
Cuba	3.203355	2.524228
Egypt	4.439001	1.350378
France	1.642403	5.163494
India	5.714004	1.191312
Israel	3.297548	7.593489
Japan	5.302053	7.59025
China	7.910626	2.985486
Russia	5.990066	4.831061
USA	1.818794	7.085297
Yugoslavia	7.00531	5.269065

Example 2. An Application of MDS to Morse Code Confusions Data:

Kruskal and Wish [1] investigated how people unfamiliar with Morse code, a system of dots and dashes representing letters and numbers, perceive and confuse these auditory signals. Participants listened to pairs of Morse code signals and indicated whether they perceived them as the same or different. Morse Code Confusions Data presented in the Figure 3, reveal patterns of similarity and dissimilarity between the signals. This confusion matrix displays the percentage of participants who judged a pair of signals (represented by the row and column) as the same. High percentages along the diagonal indicate that identical signals were correctly perceived as such. Off-diagonal cells, representing different signal pairs, generally exhibit lower percentages, reflecting accurate discrimination. However, some off-diagonal cells show higher percentages, indicating specific pairs of signals were prone to confusion. This study sheds light on the perceptual and cognitive processes involved in auditory pattern recognition. By analyzing the Morse Code Confusions Data, researchers can identify which Morse code signals are more easily confused with each other, providing insights into the underlying mechanisms of auditory perception and the potential challenges in learning and using Morse code.

Applying MDS to the Morse Code Confusions Data yields a visual representation of the perceived similarities among the signals, as shown in Fig. 4. In MDS analysis, proximity signifies similarity: signals perceived as similar are closer together, while dissimilar signals are farther apart. For instance, the signals for "B" and "X" exhibit high similarity values (84% and 64%, respectively) and are positioned close together in the configuration. Conversely, the signals for "E" and "0"display low similarity values (3% and 0%) and are positioned far apart.

Comparing this two-dimensional MDS plot to a three-dimensional MDS plot (Fig. 5) reveals subtle differences. For example, in the two-dimensional plot, "O" appears closer to "I" than to "9", while the reverse is true in the three-dimensional plot. This highlights the nuances that additional dimensions can reveal in complex perceptual data.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	1	2	3	4	5	6	7	8	9	0	
A	82	4	5	13	3	14	10	15	46	5	22	3	25	34	6	0	9	35	23	6	37	13	17	12	7	3	2	7	5	3	6	6	5	6	2	3	A
B	8	84	37	31	5	28	17	21	5	19	34	40	6	10	12	22	25	16	18	2	18	34	8	84	30	42	12	17	14	40	32	74	43	17	4	4	B
C	4	38	87	17	4	29	13	7	11	19	24	35	14	3	9	51	34	24	14	6	6	11	14	32	2	30	13	15	31	14	10	30	28	24	15	12	C
D	6	42	17	88	7	23	40	36	9	13	81	56	8	7	9	27	9	45	29	6	17	20	27	40	15	33	3	9	6	11	9	19	8	10	5	6	D
E	6	13	14	6	97	2	4	4	17	1	5	6	4	4	5	1	5	10	7	67	3	3	2	5	6	5	4	3	5	3	5	2	4	2	3	3	E
F	4	1	33	19	2	90	10	29	5	33	16	50	7	6	10	42	12	35	14	2	21	27	25	19	27	13	6	16	41	25	26	24	21	5	5	5	F
G	9	10	27	38	1	14	90	6	5	22	33	16	14	13	82	52	23	21	5	3	15	14	32	21	23	39	15	14	5	10	4	10	13	23	20	11	G
H	0	45	23	25	9	2	6	7	10	10	9	29	5	0	0	14	6	17	37	4	36	59	8	33	14	11	3	9	15	43	70	35	13	4	3	3	H
I	4	7	7	13	10	8	6	12	93	3	5	16	13	30	7	3	5	19	35	16	10	5	8	2	5	7	2	4	6	9	6	0	5	2	4	5	I
J	7	7	34	9	2	24	10	5	4	5	22	31	0	5	21	63	47	11	2	7	9	9	9	22	32	28	67	66	33	15	7	11	28	29	26	25	J
K	5	24	38	73	1	17	25	11	5	27	91	33	10	12	31	14	31	22	2	2	23	17	33	63	16	18	5	9	17	8	8	18	14	13	3	6	K
L	2	69	43	45	10	24	12	26	9	30	27	56	6	2	9	37	36	28	12	5	16	19	20	31	25	59	12	13	17	15	26	29	36	16	7	3	L
M	24	12	5	14	7	17	27	4	6	11	23	6	96	62	11	10	15	20	7	9	13	4	21	9	10	6	5	7	6	5	5	7	11	7	10	4	M
N	31	4	13	30	0	12	10	16	13	3	16	8	59	83	5	9	5	28	12	10	16	4	12	4	6	11	5	2	3	4	4	6	2	2	10	2	N
O	7	7	20	6	5	9	76	7	2	39	26	10	4	8	86	37	35	10	3	4	11	14	25	35	27	27	19	17	7	7	6	18	14	11	20	12	O
P	6	22	33	12	5	36	22	12	3	78	14	46	9	6	21	3	43	23	9	4	12	19	19	19	41	30	34	44	24	11	15	17	24	25	25	13	P
Q		20	30	11	4	15	10	5	2	27	23	26	7	6	22	51	91	11	2	3	6	14	12	37	50	63	34	32	17	12	9	27	40	58	37	24	Q
R	13	14	16	23	5	34	26	15	7	12	21	37	14	12	12	29	8	87	16	2	23	23	62	14	12	13	7	10	13	4	4	12	7	9	0	2	R
S	17	24	3	30	11	26	5	38	16	3	13	10	5	17	6	6	3	18	96	9	6	24	12	10	6	7	8	2	2	15	26	9	5	5	5	2	S
T	13	10	1	5	46	3	6	6	14	6	14	7	6	5	6	11	4	4	7	96	6	5	4	2	2	6	0	5	3	3	3	8	7	6	14	6	T
U	14	29	12	32	4	32	11	34	21	7	44	32	11	13	6	20	12	40	1	6	93	7	34	17	9	11	6	6	16	34	10	9	9	7	4	3	U
V	6	17	24	16	0	29	6	39	5	11	26	43	4	1	9	17	10	17	11	6	32	92	17	57	35	10	10	14	25	79	44	30	25	10	1	5	V
W	9	21	30	27	9	36	25	15	4	25	29	18	15	6	26	20	25	61	12	4	19	20	86	22	25	22	10	22	19	16	5	9	11	6	3	7	W
X	7	64	45	19	3	28	11	6	1	35	50	42	10	6	24	32	61	10	12	3	12	17	21	91	48	26	12	20	24	27	16	37	29	16	17	6	X
Y	9	23	67	15	4	26	22	9	8	30	12	14	3	6	14	30	2	3	1	4	6	13	21	44	86	23	26	44	40	15	11	26	22	33	23	16	Y
Z	3	16	45	18	2	22	17	10	2	23	21	51	11	2	15	59	72	14	4	3	9	11	12	36	42	7	16	21	27	9	10	29	66	47	15	15	Z
1	2	5	10	3	3	5	13	4	2	29	5	14	9	7	14	30	28	9	4	2	3	12	14	17	19	22	4	63	13	0	10	0	19	32	57	55	1
2	7	14	22	5	4	20	13	3	25	26	9	14	2	0	17	37	28	6	5	3	6	10	11	17	30	13	62	89	34	20	5	14	20	21	16	11	2
3	3	6	21	2	4	32	6	12	8	23	6	13	5	2	5	37	19	9	7	6	4	16	6	22	25	12	18	64	6	31	23	41	16	17	8	10	3
4	6	19	10	12	6	25	14	16	7	21	13	18	3	3	2	17	29	11	9	3	17	55	8	37	24	3	5	26	44	9	42	44	32	10	3	3	4
5	6	45	15	14	2	45	4	47	7	14	4	41	2	0	4	13	1	9	27	2	14	45	7	45	10	10	14	10	30	69	90	42	24	10	6	5	5
6	7	80	30	13	4	23	4	14	8	11	11	27	6	2	7	16	30	11	14	3	12	30	9	58	38	39	15	14	26	24	17	6	68	14	5	14	6
7	6	33	22	14	5	25	6	4	6	24	13	32	7	6	7	36	39	12	6	2	0	13	9	30	30	50	22	29	14	15	12	41	45	70	20	13	7
8	3	23	40	6	3	15	15	0	2	33	10	14	0	6	14	12	45	2	6	4	6	7	5	24	35	50	42	29	16	16	9	30	60	69	41	26	8
9	3	14	23	3	1	4	14	5	2	30	5	7	16	11	10	31	32	0	6	7	6	3	8	11	21	24	57	32	9	12	4	11	42	6	91	18	9
0	2	3	1	2	0	7	14	4	5	30	8	0	2	3	25	21	29	2	3	4	5	3	2	12	15	20	50	26	0	11	5	22	13	52	81	94	0
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	1	2	3	4	5	6	7	8	9	0	

Fig. 3. Representation of Morse code data

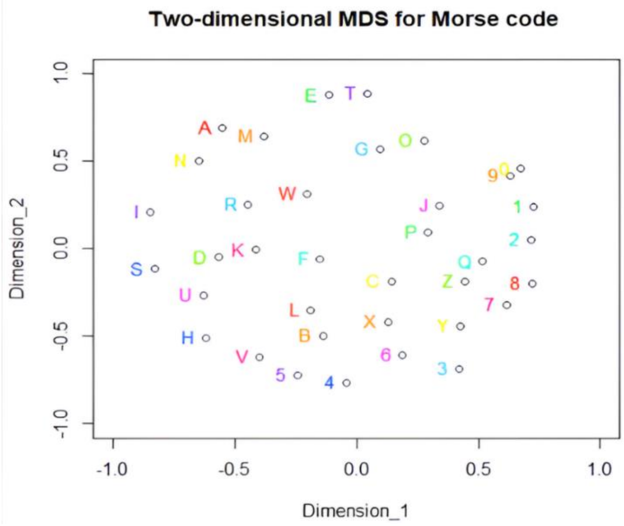

Fig. 4. Result of applying MDS to the proximities of Morse code data

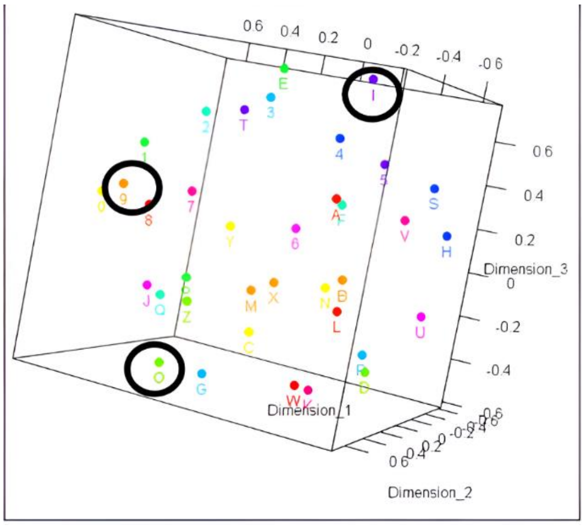

Fig. 5. Three-dimensional MDS plot for Morse code data

4. CASE STUDY: ILLUSTRATION USING A PRACTICAL DATASET RELATED TO AGRICULTURE

In a study, information from experts was obtained through questionnaires for identification of specific technologies / scientific development that need major attention for increasing the productivity of cereals, pulses and oilseeds in India which was then statistically analyzed for prioritizing future technological needs [28]. Attempts are made to analyze the available information using MDS approach. A total of 35 experts responded for ranking the factors responsible for enhancing agricultural productivity. The data is represented in Table 4.

In order to study the experts' perceptions of important factors attributable to agricultural growth, the responses (differing in their levels of importance as viewed by the experts) were considered two at a time ("all-pairs design"). Thus the responses (on a five point score from 0 to 4) of experts for the possible ${}^{10}C_2$ = 45 pairs of factors were collated. The rating for each pair of factors was averaged over all respondents and the result divided by 4 to bring the similarity ratings into the interval (0,1). These mean similarity values were then collected into a (10 x 10) table, which can then be treated as a correlation-like matrix. The similarities were converted into dissimilarities which are tabulated in Table 5.

Table 4. Dataset of factors affecting agricultural productivity

Factors	1 (1.00)	2 (0.75)	3 (0.50)	4 (0.25)	5 (0.00)	Score
F1-Quality seed availability	25	6	2	0	0	30.5
F2-Better varieties	22	8	3	0	0	29.5
F3-Timely availability of inputs	11	20	2	0	0	27.0
F4-Proper research infrastructure	16	11	4	2	0	26.8
F5-Better agronomic practices	11	15	7	0	0	25.8
F6-Adaptation to changing climatic and environmental scenario	12	12	7	2	0	25.0
F7-Marketing facilities	11	11	8	3	0	25.0
F8-Minimum Support Price (MSP)	11	10	10	2	0	24.0
F9-Development of location specific technologies	9	13	8	3	0	23.5
F10-Better extension services	11	8	12	2	0	23.5

Using the below mentioned R codes the MDS of 1,2,3,4,5 dimensions were fitted and the respective stress value vs dimension were plotted to obtain a scree-plot as shown in Fig. 6.

```
ag=read.csv(file.choose())
agg=ag[-1]
head(agg)
rownames(agg)=colnames(agg)
aggmds1 = smacofSym(delta = agg,ndim = 1, type = "ratio" )
aggmds2 = smacofSym(delta = agg,ndim = 2, type = "ratio" )
aggmds3 = smacofSym(delta = agg,ndim = 3, type = "ratio" )
aggmds4= smacofSym(delta = agg,ndim = 4, type = "ratio" )
aggmds5 = smacofSym(delta = agg,ndim = 5, type = "ratio" )
slotNames(summary(aggmds3))
#######______screeplot--------
stress=c(aggmds1$stress,aggmds2$stress,aggmds3$stress,
      aggmds4$stress,aggmds5$stress)
dimensions=c(1:5)
screeplot= plot(dimensions,stress,type = "b")
```

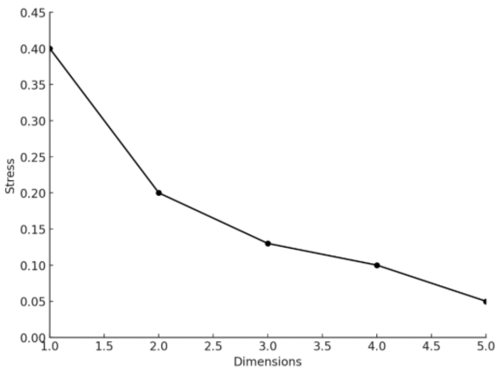

Fig. 6. Scree-plot obtained for the practical dataset

With the aid of the scree-plot it is found that 3 dimensional MDS would be more appropriate. For the comparison's sake, both the 2 dimensional and 3 dimensional MDS plots are obtained. The two-dimensional plot is obtained using the following R codes

```
zz=matrix(c(-0.1721, -0.7945,
          0.6601,-0.1938,
          0.7472,  0.1648,
```

```
            -0.0980, -0.2118,
            -0.4939,  0.7607,
            -0.8062, -0.0477,
            -0.5458, -0.1848,
             0.4331, -0.3045,
             0.3099,  0.3091,
            -0.0344,  0.5025),10,byrow = T)
x <- zz[, 1]
y <- zz[, 2]
plot(x, y,xlab = "Dimension_1",
    ylab = "Dimension_2",
    main = "Two-dimensional MDS ", xlim = c(-1,1),ylim = c(-0.6,0.8))
text(x,y,labels = rownames(agg),col =rainbow(11),pos = 2)
```

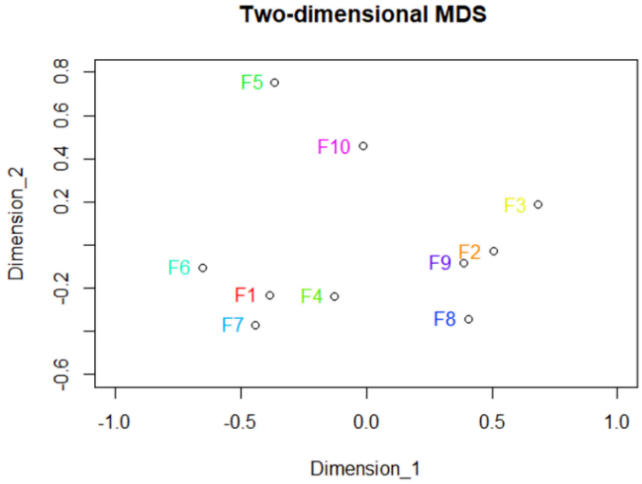

Fig. 7. Two-dimensional MDS plot for the practical dataset

The two-dimensional plot obtained is shown in Fig. 7. Similarly, the three-dimensional MDS was generated using the following R code, and the plots are depicted in Fig. 8.

```
zzz=matrix(c(-0.3870 ,-0.2334 , 0.6767
           , 0.5048 ,-0.0301 ,-0.3832
            ,   0.6853,  0.1862, -0.1379
            ,  -0.1254, -0.2381,  0.1989
            ,  -0.3648,  0.7538, -0.2317
            ,  -0.6523, -0.1045, -0.4210
            ,  -0.4402, -0.3687, -0.2364
            ,   0.4051, -0.3406, -0.2151
            ,   0.3870, -0.0817,  0.4253
            , -0.0124,   0.4571,  0.3244),10,byrow = T)
x <-zzz[, 1]
y <- zzz[, 2]
z= zzz[,3]
#library(rgl)
plot3d(x,y,z,xlab = "Dimension_1",
      ylab = "Dimension_2",
      zlab="Dimension_3",
      col = rainbow(11),size = "10")
text3d(x,y,z,row.names(agg),pos=1,col=rainbow(11))
```

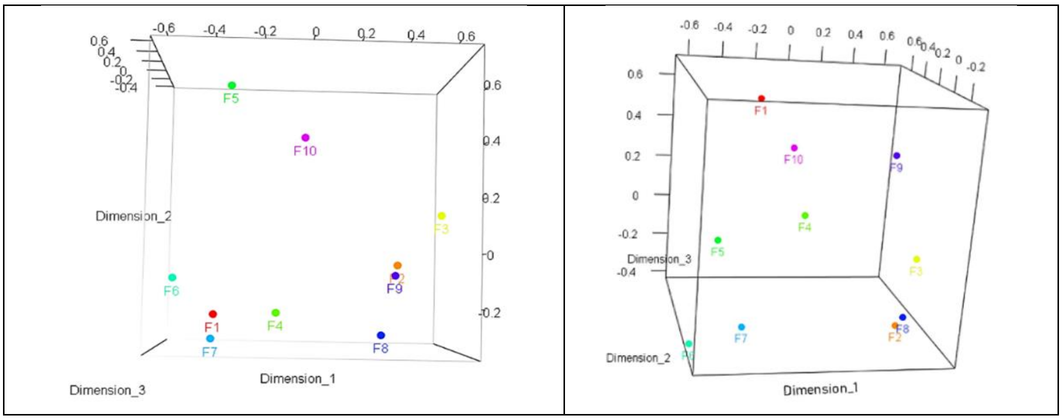

Fig. 8. Three-dimensional MDS plot for the practical dataset

Table 5. Dissimilarity Matrix of Factors Affecting Agricultural Productivity

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
F1	0.00	0.23	0.27	0.15	0.25	0.24	0.22	0.23	0.21	0.22
F2	0.23	0.00	0.09	0.18	0.25	0.26	0.22	0.10	0.19	0.22
F3	0.27	0.09	0.00	0.21	0.26	0.28	0.25	0.13	0.16	0.19
F4	0.15	0.18	0.21	0.00	0.21	0.20	0.14	0.16	0.15	0.16
F5	0.25	0.25	0.26	0.21	0.00	0.23	0.24	0.25	0.25	0.21
F6	0.24	0.26	0.28	0.2	0.23	0.00	0.10	0.22	0.25	0.20
F7	0.22	0.22	0.25	0.14	0.24	0.10	0.00	0.21	0.19	0.20
F8	0.23	0.10	0.13	0.16	0.25	0.22	0.21	0.00	0.16	0.21
F9	0.21	0.19	0.16	0.15	0.25	0.25	0.19	0.16	0.00	0.16
F10	0.22	0.22	0.19	0.16	0.21	0.20	0.20	0.21	0.16	0.00

The three-dimensional plot provides more information as in the two-dimensional plot it can be seen that F1 and F7 and close together which is also represented in the left three-dimensional plot, but the actual distance in three dimension between F1 and F7 can be seen in the right three dimensional plot. In higher dimensions, there are significant challenges in representing, comprehending, and estimating parameters. When extending beyond three dimensions, MDS becomes virtually ineffective as a method for making complex data more accessible to the human mind. Four or more dimensions make it exceedingly difficult to visualize and interpret the results, diminishing the utility of MDS for practical data analysis.

5. CONCLUDING REMARKS

MDS serves as a powerful data visualization technique that simplifies complex data by portraying its structure spatially, making it easier to understand relationships among a set of stimuli. Through various applications in marketing, ecology, molecular biology, social networks, and more, MDS has proven to be a versatile tool for quantifying similarity or dissimilarity between entities. Despite challenges in choosing the number of dimensions and the inherent difficulties in representing higher-dimensional data, tools such as the Scree-plot assist in selecting the optimal number of dimensions. This study provides detailed examples and step-by-step procedures for implementing MDS using MS-Excel and R, enhancing the understanding of the practical aspects of MDS. Additionally, MDS applications in Perceptions of Nations data and Morse code confusion data are presented. Real-world datasets, such as factors affecting agricultural productivity, are analysed to demonstrate the effectiveness of MDS. The practical examples and software implementations provided in this paper illustrate the utility and broad applicability of MDS. By enabling researchers to visualize and interpret complex data, MDS continues to be an essential method in diverse fields, facilitating better decision-making and deeper insights into data patterns and relationships.

DISCLAIMER (ARTIFICIAL INTELLIGENCE)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of manuscripts.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Kruskal JB, Wish M. Multidimensional Scaling, Series: Quantitative applications in the social sciences, Sage University Press, California; 1978.
2. Torgerson WS. Multidimensional scaling: I. Theory and method. Psychometrika. 1952;17(4):401-419.
3. Shepard RN. Multidimensional scaling, tree-fitting, and clustering. Science. 1980;210(4468):390-398.
4. Clarke KR. Non-parametric multivariate analyses of changes in community structure. Australian Journal of Ecology. 1993;18(1):117-143.
5. Mardia KV. Some properties of classical multidimensional scaling. Communications in Statistics-Theory and Methods. 1978;7(13):1233-1241.
6. De Leeuw J, Mair P. Multidimensional scaling using majorization: SMACOF in R, Journal of Statistical Software. 2009;31(3):1-30. Available:https://escholarship.org/uc/item/9z64v481
7. Pacini GC, Colucci D, Baudron F, Righi E, Corbeels M, Tittonell P, Stefanini FM. Combining multi-dimensional scaling and cluster analysis to describe the diversity of rural households. Experimental Agriculture. 2014;50(3):376-397.
8. Liu S, Cui W, Wu Y, Liu M. A survey on information visualization: Recent advances and challenges. The Visual Computer. 2014;30:1373-1393.
9. Saeed N, Nam H, Haq MIU, Muhammad Saqib DB. A survey on multidimensional scaling. ACM Computing Surveys (CSUR). 2018;51 (3):1-25.
10. Adiga MS, Ananthan PS, Kumari HD, Ramasubramanian V. Multidimensional analysis of marine fishery resources of Maharashtra, India. Ocean and Coastal Management. 2016;130:13-20.
11. Asami A, Saika Y. Forecast of meteorological data utilizing state-space model utilizing metric-multidimensional scaling. International Journal Research in Engineering and Technology. 2014;3(17):20-26.
12. Beatty M, Manjunath BS. Dimensionality reduction using multi-dimensional scaling for content-based retrieval. In Proceedings of International Conference on Image Processing IEEE. 1997;(2):835-838.
13. Chen K, Kou G, Shang J, Chen Y. Visualizing market structure through online product reviews: Integrate topic modeling, TOPSIS, and multi-dimensional scaling approaches. Electronic Commerce Research and Applications. 2015;14(1):58-74.
14. Lee JH, McDonnell KT, Zelenyuk A, Imre D, Mueller K. A structure-based distance metric for high-dimensional space exploration with multidimensional scaling. IEEE Transactions on Visualization and Computer Graphics. 2013;20(3):351-364.
15. Coxon APM. The User's guide to Multidimensional Scaling, Heinemann Educational Books, Great Britain; 1982.
16. Hair JF, Anderson RE, Tatham RL, Black WC. Multivariate data analysis, 4th Edition, Prentice Hall, New Jersey; 1995.
17. Cox TF, Cox MAA. Multidimensional Scaling, Second edition. Chapman and Hall/CRC, Boca Raton; 2001.
18. Borg I, Groenen PJF. Modern Multidimensional Scaling: Theory and Applications, Second edition, Springer-Verlag, New York; 2005.
19. Izenman AJ. Modern multivariate statistical techniques: Regression, classification and manifold learning. Springer, New York; 2008.
20. De Leeuw J, Heiser WJ. Convergence of correction matrix algorithms for multidimensional scaling. Geometric Representations of Relational Data. 1977;735-752.
21. Ding CS. Fundamentals of applied multidimensional scaling for educational and psychological research. Springer; 2018.
22. Izenman AJ. Modern multivariate statistical techniques. New York: Springer; 2008.
23. Kaski S, Peltonen J. Dimensionality reduction for data visualization [applications corner]. IEEE Signal

Processing Magazine. 2011;28(2):100-104.

24. Najim SA. Information visualization by dimensionality reduction: A review. Journal of Advanced Computer Science and Technology. 2014;3(2):101.
25. An J, Yu JX, Ratanamahatana CA, Chen YPP. A dimensionality reduction algorithm and its application for interactive visualization. Journal of Visual Languages and Computing. 2007;18(1):48-70.
26. Telea AC. Data visualization: Principles and practice. CRC Press; 2014.
27. Hout MC, Papesh MH, Goldinger SD. Multidimensional scaling. Wiley Interdisciplinary Reviews: Cognitive Science. 2013;4(1):93-103. Available:https://doi.org/10.1002/wcs.1203
28. Ramasubramanian V, Kumar A, Prabhu KV, Bhatia VK, Ramasundaram P. Forecasting technological needs and prioritizing factors in agriculture from a plant breeding and genetics domain perspective: A review. Indian Journal of Agricultural Sciences. 2014;84(3): 311-316.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/120452