Cody S. Ding

Fundamentals of Applied Multidimensional Scaling for Educational and Psychological Research

Springer

Fundamentals of Applied Multidimensional Scaling for Educational and Psychological Research

Cody S. Ding

Fundamentals of Applied Multidimensional Scaling for Educational and Psychological Research

Springer

Cody S. Ding
Department of Education Science and Professional Program
University of Missouri-St. Louis
St. Louis, MO, USA

Center for Neurodynamics
University of Missouri-St. Louis
St. Louis, MO, USA

ISBN 978-3-319-78171-6 ISBN 978-3-319-78172-3 (eBook)
https://doi.org/10.1007/978-3-319-78172-3

Library of Congress Control Number: 2018935924

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Since I have been conducting research in education and psychology using multidimensional scaling (MDS) models for many years, I read quite a few books as well as some classical articles on MDS. My experiences with these readings are that writings on this topic are often very technical and esoteric, which make it hard to relate to the current research needs in education or psychology. I gradually develop an urge to write a book on MDS that is more understandable and relevant to the current research setting. This is my first effort.

In this round of writing, I focus on MDS concepts that I deem to be more relevant to current research in education or psychology. I try to convey MDS concepts in more understandable terms and focus on aspects of MDS that may be more potentially useful to readers. Although I put some technical aspects of the MDS (such as equations) in a few chapters, they are simple for the purpose of making the discussion more complete. Readers can skip these sections without losing the main ideas of the topic. I made each chapter as short as I can to only cover main points so that readers can focus on the fundamentals. Of course, this is done at the risk of omitting many potentially helpful materials.

MDS has not been often employed in education or psychology research in recent years. Although I did not cover everything that can be done with MDS analysis, I did indicate some potential research that can be done via MDS. I provided some examples of MDS analyses so that readers can get some ideas, with the hope that this could pique reader's interest. MDS analysis has its limitations but it can certainly be useful. Thus, the book is intended for students or researchers who want to know more about MDS but not so technical. The book is not a textbook in a technical sense since it does not teach or show readers how to perform MDS analysis. However, the book provides a comprehensive view of fundamentals of MDS so that readers can understand what MDS is and can do.

This book is intended as a research reference book for graduate students and researchers to get fundamental ideas of multidimensional scaling (MDS) and how this particular analytic method can be used in applied settings. Some of the major problems with the content of existing MDS books are that the discussion on MDS (1) tends to be very technical, (2) covers many topics that are less relevant to current practices in educational or psychological research, and (3) uses language or examples that are less common in today's research setting. As such, graduate students or researchers are not likely to view MDS as a viable method for studying issues at hand.

Before 1985 or so, there were many publications on MDS. But then it somehow fell out of fashion. Today MDS is offered as part of materials on multivariate analysis, usually as a chapter. However, one chapter is not nearly close to covering some unique aspects of MDS, particularly regarding the applications of this method to research in education and psychology. I do not expect dramatic changes in its popularity, but I do believe MDS as a method can offer some interesting applications to research and this is not a popularity contest. This book is an effort to make MDS more accessible to a wider audience in terms of the language and examples that are more relevant to educational research and less technical so that the readers are not overwhelmed by equations and do not see any applications. In addition, it discusses some new applications that have not previously been discussed in MDS literature. My philosophy is that methods are just methods, not bad or good, and it all depends on how you use them and for what purpose. Using popularity to assess the value of academic books will limit the spread of knowledge. In addition, MDS is not one method, but rather it comprises a family of methods that can be used for different purposes.

This book can also be used as a supplemental book for advanced multivariate data analysis on the topic of MDS, which is typically one chapter in such a book of multivariate data analysis for graduate students. As mentioned previously, the main impetus for writing this proposed book is that I hope to have a book that is not so technical for graduate students and researchers who are not interested in the technicality of MDS. I have read quite a few books on MDS and I am struggling with thoughts of why and what these materials in the books are useful for, although they are informative from a purely academic perspective. Therefore, the book is more of response to my own desire to have a book in which I can see the relevancy of MDS in actual research settings. The book does not have exercises or discussion questions since my goal is to have readers learn some fundamentals and start using MDS via available software programs. If they do want to know more technical aspects of MDS, they can always refer to books by Davison (1983) or Borg and Groenen (2005), for example. There is no need for me to replicate what they have already done.

St. Louis, MO, USA Cody S. Ding

Contents

Part I
Basics of MDS Models

Describe basic and fundamental features of MDS models pertaining to applied research applications.

Chapter 1
Introduction

Abstract Discuss fundamental ideas of MDS, particularly MDS as a data visualization tool in the context of big data is highlighted. Similarities and differences between MDS, factor analysis, and cluster analysis are discussed.

Keyword MDS · Visualization · Factor analysis · Cluster analysis

In this chapter, we mainly discuss the concept of multidimensional scaling in the current psychological or education research context. We also discuss some differences and similarities among multidimensional scaling, factor analysis, and cluster analysis. The goal of such a discussion is to have readers obtain a better sense of the concept of multidimensional scaling in relation to other conceptually similar methods, particularly in the language that is more relevant to current educational and psychological research.

1.1 What Is Multidimensional Scaling

In much of the quantitative and statistical literature, multidimensional scaling (MDS) is often referred to as a technique that represents the empirical relationships of data as a set of points in a geometric space, typically in two or higher dimensional spaces. Specifically, multidimensional scaling represents a family of statistical methods or models that portray the structure of the data in a spatial fashion so that we could easily see and understand what the data indicate. This may be the reason that MDS tends to be viewed as a data visual technique, and sometimes it is considered with respect to mapping technique. The unifying theme of different MDS models is the spatial representation of the data structure. In this regard, MDS can be considered as one analytic tool of data visualization in the context of big data.[1] In the context of data visualization, MDS models can be used to investigate a wide range of issues in education and psychology such as perception of school

[1] Big data means that there are lots of data being collected. Visualization is one method for big data analysis.

C. S. Ding, *Fundamentals of Applied Multidimensional Scaling for Educational and Psychological Research*, https://doi.org/10.1007/978-3-319-78172-3_1

climate by various age groups of students, changes in achievement, sensitivity of psychological measures, or individual differences in mental health, to name a few. Moreover, it can also be employed for the purpose of hypothesis testing, like that in structural equation modeling. Although MDS is a powerful tool of studying various behavioral phenomena, it appears to be underused in current educational and psychological research.

In the literature, MDS has been defined in slightly different ways. For example, Davison (1983) defined MDS as a method for studying the structure of stimuli (i.e., variables) or individuals. Borg and Groenen (2005) defined MDS as a technique of representing distances between objects (or variables) in a multidimensional space. In a nutshell, MDS can be defined as a family of analytical methods that use the geometric model (typically in the form of a distance equation) for analysis of interrelationships among a set of variables, people, or combination of variable and people (such as in preference analysis) so that the latent structure of data can be visualized for meaningful interpretation. A distance equation could be the Euclidean distance, the city-block distance, or the Minkowski distance. Thus, an MDS analysis involves employment of a specific model of study, for instance, how people view things in different ways.

More specifically, multidimensional scaling is carried out on data relating objects, individuals, subjects, variables, or stimuli to one another. These five terms are sometimes used interchangeably, which may cause some confusion. Objects, variables, or stimuli usually refer to inanimate things, such as variables or test scores; individuals and subjects referto people. Given the distance measures between variables, MDS models produce a solution that consists of a configuration of patterns of points representing the variables in a space of a small number of dimensions, typically in two or three dimensions. For this reason, it is also called small space analysis (SSA). The following example illustrates this point. This example represents a group of students who take a given reading test five times over a two-month period. The distance matrix of these score is:

$$D = \begin{bmatrix} 0 & & & & \\ 1 & 0 & & & \\ 2 & 1 & 0 & & \\ 3 & 2 & 1 & 0 & \\ 4 & 3 & 2 & 1 & 0 \end{bmatrix}$$

This distance matrix can be thought of as giving information about how similar or dissimilar these test scores are to each other over time. MDS model takes this information and represents these test scores as a point in space, which is shown in Fig. 1.1 In this two-dimensional space, as shown in Fig. 1.1, the more similar the test scores are, the closer they lie to each other. The pattern of points that most accurately represents the information in the data is the MDS solution or configuration.

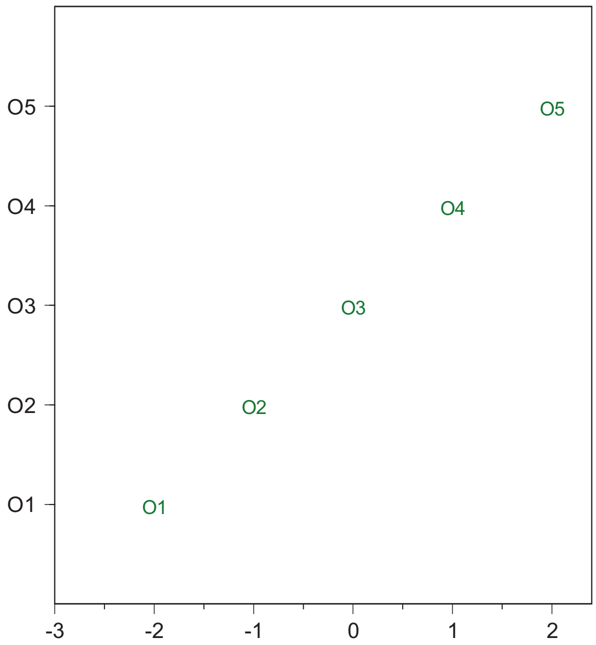

Fig. 1.1 A hypothetical example of a configuration of five reading test scores

In this example, these five reading test scores have a linear configuration, indicating the linear increase of reading achievement over time. As Tukey (1977) says: "A picture is worth a thousand words."Thus, a picture of the data is produced that is much easier to assimilate (visually) than a matrix of numbers, particularly if such a matrix of number is large. It may also bring out features of the data that were obscured in the original matrix of coefficients (i.e., dissimilarity coefficients).

This example, although based on fictitious data, allows a number of points to be noted:

1. MDS is primarily concerned with representation, in this case with the production of a simple and easily assimilated geometrical picture where distances are used to represent the data.
2. MDS models differ in terms of the assumptions they make about how important the quantitative properties of the data are. In the example above, it is in fact only the rank order of the data percentages, which is matched perfectly by the distances of the configuration. This is an example of ordinal scaling or, as it is more commonly termed in MDS literature, non-metric scaling.
3. A wide range of data and measures can be used as input, as will be discussed in Chap. 2. Any data that can be interpreted as measures of similarity or dissimilarity are appropriate for MDS scaling analysis.

Traditionally, there are following issues that we must consider in using multidimensional scaling:

1. *The data*, the information to be represented (discussed further in Chap. 2);
2. *The transformation* how data should be related to the model, such as basic nonmetric MDS model or metric MDS model (discussed in Chap. 3).
3. *The model* how the solution should be interpreted, such as individual differences model or basic scaling model (more on this in later chapters), as giving information about the relationships between the variables.
4. The sample size required for an MDS analysis does not need to be large: it can range from a few people to a few hundred. Since the MDS analysis is more of a descriptive (except for maximum likelihood MDS) and does not involve significance testing, the interpretation and accuracy of the analysis results are not tied to the sample size. Thus, the MDS analysis can be used for studies based on the single-case design such as an investigation of the response of a small group of individuals to a treatment. However, if one wants to make a generalization based on the people in the study, a representative sample is required.
5. MDS models (except for maximum likelihood MDS) do not have distributional requirements such as normality of the coordinates. But the maximum likelihood MDS assumes that the coordinates are normally and independently distributed and each object or variable can have the same variance or different variances (discussed in Chap. 7).

1.2 Differences and Similarities Between MDS, Factor Analysis, and Cluster Analysis

Before we start further discussion on MDS models, it is imperative to discuss differences and similarities between MDS, factor analysis, and cluster analysis. Without a clear conceptual understanding of what MDS models are all about, particularly in relation to these methods, practitioners may have difficulty in utilizing MDS for their work, thus impeding further developments of MDS models in psychological and educational research. In light of this and to remain consistent with the applied orientation of the book, this discussion is focused more on conceptual grounds rather than mathematical aspects.

1.2.1 MDS and Factor Analysis

Conceptually, factor analysis is a technique that discovers latent relationships among a set of variables. The objective is to explain *a number of* observed variables, (m), by a set of latent variables or factors (f), where (f) is much smaller in number than (m). The hypothesis is that only a few latent factors suffice to explain most of the variance of the data. In other words, the relationships among the observed variables exist because of the underlying latent variables. Likewise, the objective of

MDS is to reveal geometrically the structure of data in fewer dimensions. Like MDS, factor analysis yields a quantitative dimensional representation of the data structure. Both have been used to study dimensionality among variables. It is often the case that the term *factor* and *dimension*are used interchangeably in factor analysis literature. Because of this similarity, it is not a surprise that factor analysis and MDS are viewed as very similar if not the same.

Studies have been done to compare the two techniques (e.g., Davison 1985). The differences between the two may be summarized as follows: (1) factor analysis yields more dimensions than does MDS; (2) factor analysis typically represents linear relationships among variables, whereas MDS can represent both linear and nonlinear relationships; (3) MDS is traditionally used more often as a data visualization tool than factor analysis, which is typically a measurement technique of finding a set of latent variables that connect observed variables together; and (4) MDS can employ a variety of kinds of data such as preference ratio data, whose values are coded between 0.0 and 1.0, indicating the degree to which a variable in a variable pair is preferred. But factor analysis usually analyzes the correlation matrix, whose values indicate similarities between variables. Therefore, the applications of MDS can be more diverse than that of factor analysis. For example, MDS preference analysis can be used to study individuals' preferences to a set of coping behaviors (e.g., prefer shouting to talking with friends), whereas factor analysis usually is used in studying how a set of coping behaviors measures a particular coping construct (e.g., withdrawal coping).

The take-home message of the differences between these two methods is that factor analysis is focusing on latent variables that represent some constructs such as anxiety or depression, while MDS analysis is more in line with Network Analysis (McNally et al. 2015), where behaviors (as assessed by variables) are best construed as a system embodied in networks of functionally interconnected fashion. Thus, the configuration of relation between variables is mereological – part to whole – rather than causal as in factor analysis (Borsboom and Cramer 2014; Guyon et al. 2017). For example, typical example used in illustrating MDS analysis is to show the relation between the 50 states as a map. Accordingly, the map is mereological: parts (i.e., 50 states) to whole (i.e., the United States). There is no underlying causal relation between states and the country called the United States; states are part of it. Moreover, MDS as a network analysis is more exploratory, that is, empirically discovered rather than formed by theories. Thus, the difference is ontological as in factor analysis versus mereological as in MDS analysis.

1.2.2 MDS and Cluster Analysis

Another closely related method to MDS is cluster analysis (Kruskal 1977). Traditional cluster analysis, such as hierarchical cluster analysis, is employed to identify individuals who share similar attributes (e.g., high risk adolescents).

While MDS can be used in the same way, Davison (1983) pointed out three differences between MDS and cluster analysis. First, relationships between the observed distance matrix and model derived distance matrix in cluster analysis cannot be expressed in linear or even monotone fashion as in MDS. Second, dimensions in cluster analysis are typically represented in a tree diagram of many simple two-valued dimensions to represent data. As such, the number of dichotomous dimensions needed to represent the data structure become large in practice. Third, MDS defines clusters of individuals in terms of continuous dimensions rather than in either-or fashion. Thus, we can describe a group of individuals who possess more of one attribute (e.g., depression) than the other (e.g., anxiety) rather than having that attribute (e.g., depression) or not. In addition to these three differences, MDS is a model-based approach, while traditional cluster analysis is not. Recently, some researchers have developed model-based cluster analysis (Fraley and Raftery 2007). However, a key difference between model-based cluster analysis and MDS remains in that MDS represents cluster in terms of dimension rather than ina dichotomous fashion.

1.3 Conclusion

In this chapter, we discuss what the MDS models are and their fundamental utilities. We also summarize some fundamental differences between MDS, factor analysis, and cluster analysis. One take-home message is that MDS is not simply a data-reduction method. MDS can be used for many other purposes in education and psychological applications such as the longitudinal study of achievement, treatment preferences, or hypothesis testing of behavioral likings, as we will see in the later chapters.

References

Borg, I., & Groenen, P. J. F. (2005). *Modern multidimensional scaling: Theory and applications* (2nd ed.). New York, NY: Springer.

Borsboom, D., & Cramer, A. O. J. (2014). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology, 9*, 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608.

Davison, M. L. (1983). *Multidimensional scaling*. New York: Wiley.

Davison, M. L. (1985). Multidimensional scaling versus components analysis of test intercorrelations. *Psychological Bulletin, 97*(1), 94–105.

Fraley, R. C., & Raftery, A. (2007). Model-based methods of classification: Using the mclust software in chemometrics. *Journal of Statistical Software, 18*(6), 1–13.

Guyon, H., Falissard, B., & Kop, J.-L. (2017). Modeling psychological attributes in psychology – An epistemological discussion: Network analysis vs. latent variables. *Frontiers in Psychology, 8*, 798. https://doi.org/10.3389/fpsyg.2017.00798.

Kruskal, J. B. (1977). The relationship between multidimensional scaling and clustering. In J. V. Ryzin (Ed.), *Classification and clustering* (pp. 17–44). New York, NY: Academic.

McNally, R., Robinaugh, D. J., Wu, W. G. Y., Wang, L., Deserno, M. K., & Borsboom, D. (2015). Mental disorders as causal systems: A network approach to posttraumatic stress disorder. *Psychological Science, 3*, 836–849.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Chapter 2
Data Issues in MDS

Abstract Data is the first step in process of any statistical analysis. Since MDS has a bit different terms associated with data concepts and it can be confusing, I try to discuss the data used for MDS with terms that are more understandable or relevant to the common research setting.

Keyword Distance measures · Measurement conditionality · Number of ways · Number of mode

In this chapter, we discuss some essential features of data that are typically associated with MDS analysis. Some data are unique to MDS analysis such preference ratio or binary choice data. In addition, some terms used in describing MDS data are a bit confusing. Here we attempt to explain these terms as clearly as a layperson can understand. If we can better understand these features of data, we are more likely to use the MDS analysis in our research or data practices. We also discuss a MDS program that can perform various types of MDS analysis.

2.1 A Look at Data

MDS can be used for various analyses, and therefore different types of data can be involved. Young (1987) provided a thorough discussion of data for MDS models, as did some other authors (e.g., Borg and Groenen 2005; Davison 1983). In here, we will discuss those aspects of data that are most relevant to MDS in the current research context.

Several types of data lend themselves to analysis by multidimensional scaling. Behavioral scientists have adopted several terms relating to data, which often are not familiar to others. Typically, variables can be classified according to their "measurement scale". The four scales that are commonly mentioned in the literature are the nominal scale, the ordinal scale, the interval scale, and the ratio scale. For MDS models, any type of data can be converted into proximity measures as an input for

C. S. Ding, *Fundamentals of Applied Multidimensional Scaling for Educational and Psychological Research*, https://doi.org/10.1007/978-3-319-78172-3_2

MDS analysis. Traditionally, the data used in MDS analysis are typically called proximity measures. The term, *proximity*, is vague, however, since it can indicate similarity data as well as dissimilarity data. For this reason, in this book we use a specific term for a particular kind of data. For instance, if distance matrix is to be used, we will refer to such data as distance data or measure (c.f., dissimilarity or proximities). The most common measure of the relationship of one variable (stimulus, etc.) to another is a distance measure (i.e., distance coefficient). It measures the "dissimilarity" of one object to another, where the distance, δ_{ij}, between the two objects is measured. If the data are given as similarities, s_{ij}, such as correlation coefficients, some monotone decreasing transformation will convert them to dissimilarities or distance coefficients.

$$\delta_{ij} = 1 - s_{ij}$$

$$\delta_{ij} = c - s_{ij} \text{ where c is for some constant}$$

$$\delta_{ij} = \sqrt{2\left(1 - s_{ij}\right)}.$$

MDS analyses assume that distance measures are given. How one collects these distance measures is a problem that is largely external to the MDS models. However, because distance measures are obviously needed and because the way these distance measures are generated has implications for the choice of an MDS model, we discuss some of these issues here.

2.2 Data Source

Traditionally, the data used in MDS analysis usually come from direct judgment of certain stimuli with respect to some attribute. For example, participants are asked to judge which car's color is brighter or to judge which two schools are similar with respect to certain characteristics such as friendliness or orderliness. Such judgment data are generated via four types of judgment tasks: magnitude estimation, category rating, graphic rating, and category sorting. Currently, the judgment data in education or psychology (except for some experimental studies) are not so common because of the practical problems (e.g., time constraints or willingness to participate) and the participant's ability and willingness to perform the various tasks.

Typical data commonly used in today's research is data generated by questionnaires or surveys using a Likert-type scale metric such as from 1 to 4, with 1 being *not at all* and 4 being *always* with respect to a certain event or trait (e.g., how often do you feel happy?). This type of data is typically not discussed in traditional MDS literature; however, data produced by a Likert-type scale can be easily converted into either a distance data matrix by averaging across all participants or individual distance matrices, one for each participant. Such data are called indirect proximity