Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v10n2p432

Multiple Correspondence Analysis and its applications

By Khangar, Kamalja

Published: 14 October 2017

Electronic Journal of Applied Statistical Analysis
Vol. 10, Issue 02, October 2017, 432-462
DOI: 10.1285/i20705948v10n2p432

Multiple Correspondence Analysis and its applications

Nutan Vijay Khangar and Kirtee Kiran Kamalja*

Department of Statistics, School of mathematical Sciences, North Maharashtra University, Jalgaon Maharashtra PIN 425001

Published: 14 October 2017

Correspondence analysis (CA) is a statistical visualization method for picturing the association between the levels of categorical variables (CVs). Specifically, simple and multiple correspondence analysis (MCA) is used to analyze two-way and multiway data respectively. Biplots play an important role in visualization of association. This paper overviews the popular approaches of MCA and discusses the role of biplots in CA. We discuss theoretical issues involved in different methods of MCA and demonstrate each of these methods through examples. The main aim of the present paper is to highlight the importance of MCA based on separate SVDs. We study the association pattern in mother-child behavior over time, using MCA based on separate SVDs.

keywords: Simple Correspondence analysis, Multiple Correspondence Analysis, Biplot, Multi-way contingency table.

1 Introduction

CA have numerous applications in various disciplines, such as archaeology, ecology, medical and health sciences, social sciences, psychological behavior, etc. The use of CA and MCA is well established in behavioral and social science research for understanding relationships between two or more CVs. Nowadays, in many scientific investigations, including sensory evaluation, market research and customer satisfaction evaluations, etc., questionnaires and survey results in a large number of responses to questions with a limited number of answer categories and the aim is to study the behavioral approach

*Corresponding author: kirteekamalja@gmail.com

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

of individuals as per their responses. Virtually every research project categorizes some of its observations into neat, little distinct bins: male or female; marital status; broken or not broken; and so on. The data by categories is recorded as counts, i.e. how many observations fall into a particular bin. The explosion in the development of methods for analyzing such type of categorical data began in the 1960s and has continued apace in recent years. The aim of the categorical data analysis is to study the association between the CVs. Many methodologies have been discussed in literature to study the association between CVs, one such popular method is CA. CA is widely used in different fields such as ecology, archaeology, various disciplines of the sciences, etc. Karl Pearson and R.A. Fisher were central to the original statistical development of the tools needed to perform CA.

The numerical and graphical analysis of the association between CVs has a long and interesting history. Chi-square statistic is used to test the significance of the association between the CVs. If there is a statistically significant association between CVs, then the nature of the association may be studied by performing CA. CA is a technique that allows a user to graphically display the row and column categories and provides a visual inspection of their "correspondences" or associations, at a categorical level. The core of CA was established in 1963 and then combined with clustering methods. It is originally introduced by Bnzecri (1969). A Systematic developments in CA is given by Beh (2004). Beh and Lombardo (2012) described the growth of CA from an international perspective.

There are many introductory discussions on some of the key aspects of CA. Although the literature overview of CA and MCA given by the many researchers is very good, but are frequently too technical. Therefore, our approach is to discuss briefly on the technical details of different approaches to MCA and simultaneously focus on the concepts and applications of MCA for the extensive understanding. The main focus is on conceptual understanding and applications of MCA methods to social and behavioral sciences data to encourage researchers to use MCA. The purpose of this paper is to specifically discuss applications of various methods of MCA to behavioral data.

This paper is organized as follows. In Section 2, we discuss the concept of biplot and its application in CA. In Section 3, The method of SCA and popular approaches to perform MCA are discussed. We demonstrate each approach of MCA through examples in Section 4. Finally, we discuss the applications of CA/MCA with computational issues in Section 5.

2 Biplot

A human can inspect two-dimensional presentations or three-dimensional structures easily. Viewing objects in more than three-dimensions has seemed beyond the scope of human perception. More often, we have the data where the rows of the data matrix are usually observed sampling units such as individuals, countries, demographic groups, locations, cases, objects, etc. and the columns are variables describing the rows such as responses to a questionnaire, economic indicators, product purchased, environmental parameters, genetic markers, etc. In such situations, the researchers are mainly interested to visualize the data in low-dimensional plots, especially in two-dimensional. Biplot is such type of exploratory graph and is a generalization of the simple two-variable scatter plot.

Biplot was originally proposed by Gabriel (1971) in a principal component analysis. It is a graphical display of the rows and columns of a data matrix as points (or, equivalently, as vectors) in a low-dimensional Euclidean space, usually of dimensionality two or three. The points have specific interpretations in terms of scalar products. The idea is to recover individual element of the data matrix approximately through the scalar products. Geometrically, the scalar product of $\underline{x} = (x_1, x_2, ..., x_k)^T$ and $\underline{y} = (y_1, y_2, ..., y_k)^T$ is given as follows.

$$\underline{x}^T \underline{y} = \|\underline{x}\| \|\underline{y}\| \cos\theta$$

where, $\|\underline{x}\|$ is the norm of $\underline{x}$ and θ is angle between $\underline{x}$ and $\underline{y}$.

To clear the concept of biplot in a very simple way, Greenacre (2010) considers the decomposition of a 5×4 matrix (T) into 5×2 left matrix (X) and 2×4 right matrix (Y^T) as $T = XY^T$ and is given as follows.

$$\begin{pmatrix} 8 & 2 & 2 & -6 \\ 5 & 0 & 3 & -4 \\ -2 & -3 & 3 & 1 \\ 2 & 3 & -3 & -1 \\ 4 & 6 & -6 & -2 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 1 & 2 \\ -1 & 1 \\ 1 & -1 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} 3 & 2 & -1 & -2 \\ 1 & -1 & 2 & -1 \end{pmatrix}$$

The rows of the left matrix X and columns of the right matrix Y provide two sets of points $B_P = \{\underline{x}_1, \underline{x}_2, \underline{x}_3, \underline{x}_4, \underline{x}_5\}$ and $B_V = \{\underline{y}_1, \underline{y}_2, \underline{y}_3, \underline{y}_4\}$. B_P considered as a set of biplot points and B_V is considered as a set of biplot vectors/biplot axes. The biplot points when projected onto biplot axes recover the values in target matrix. Calibration of biplot axes can be done to read off the values of the target matrix directly from projections. Different versions of biplot are used as per the applicable area. The biplot for the above decomposition is shown in Figure 1.

The decomposition XY^T of T in the above discussion is not unique. One of the convenient way of obtaining a unique decomposition of T, with convenient properties, is to use the singular value decomposition (SVD). SVD plays a fundamental role underlying the theory and computation of biplot. This decomposition provides the coordinates of the points and vectors in the biplot with respect to the dimensions that are ordered from the most to the least important, so that the reduced dimension of the space that retains the major part of the original data can be selected. Biplot based on SVD is the same concept to visualize the data graphically in two-dimensional space. Decompositions based on SVD with different scaling of biplot point/axes lead to different versions of biplot with different properties. The SVD of $m \times n$ matrix T is as follows.

$$T = U\Sigma V^T$$

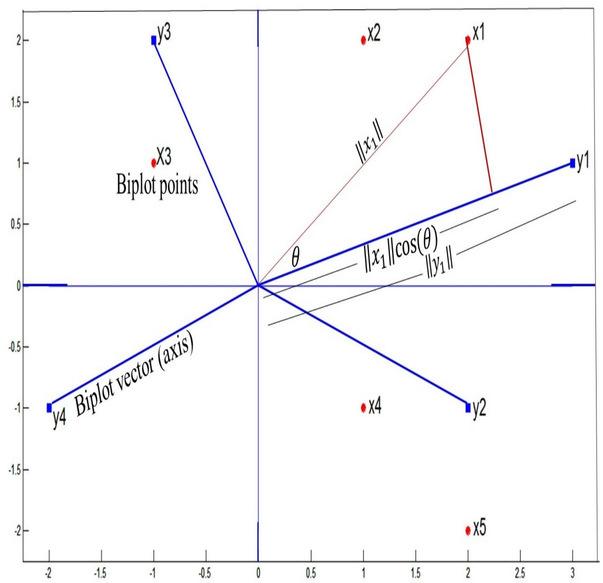

Figure 1: Biplot Corresponding to data matrix T

where, U and V are $m \times m$ and $n \times n$ matrices, respectively such that $UU^T = I_m$ and $VV^T = I_n$, Σ is $m \times n$ rectangular diagonal matrix of singular values $\sigma_1, \sigma_2, ..., \sigma_r$, and $r = rank(T)$. This decomposition of T is used to obtain the symmetric biplot associated with data matrix T. The first two columns of $A = U\Sigma^{1/2}$ and $B = V\Sigma^{1/2}$ provide two sets of points $B_p = \{\underline{x}_1, \underline{x}_2, ..., \underline{x}_m\}$ and $B_V = \{\underline{y}_1, \underline{y}_2, ..., \underline{y}_n\}$ in R^2 of which B_P is considered as a set of biplot points and B_V is considered as a set of biplot vectors/biplot axes.

The basic idea of biplot is very simple and like all simple solutions to complex problems it is both powerful and useful. The biplot makes information in a table of data transparent and revealing the main structures in the data in a methodical way. Biplots show the following quantities of a data matrix in one display.

(i) The variance-covariance structure of the variables, i.e. the inner product between two variables and the cosine of the angle between them approximates their correlation with equality if the fit is perfect.

(ii) It explores the relationship (interrelationship) among (between) rows and columns.

(iii) The Euclidean distances between observations in the multidimensional space is also shown.

Interpretations from biplot

(i) Angle between the biplot points
The cosine of the angle between the lines drawn to each pair of biplot axes and

points show the correlation between the two corresponding variables (i.e. If the angle between two row vectors is small, they have similar response patterns over columns whereas if the angle between two column vectors is small, then they are strongly associated). Thus, a small angle between two vectors indicates that the two variables are highly correlated. While if two vectors form an angle of 90^0 (greater than 90^0) then corresponding variables are uncorrelated (negatively correlated).

(ii) Biplot vector length
The length of the biplot vector indicates how well the variables are represented by the graph with a perfect fit if all vectors have equal lengths. Specifically, more the vector length, better is its discrimination ability.
Biplots play a fundamental role in the theory of CA. For this reason, the use of biplots in a CA context has exponentially grown up over the years. The computational algorithm and some related issues of CA along with biplots are discussed in the next section.

3 Correspondence Analysis (CA)

CA is a statistical visualization method for studying the association between the levels of CVs. We discuss SCA and different methods of MCA in the following.

3.1 Simple CA

SCA is performed to study the association between two categorical variables. Let N be the $j_1 \times j_2$ data matrix (contingency table) with $(i_1, i_2)^{th}$ cell entry $n_{i_1 i_2}$, $i_1 = 1, 2, \cdots, j_1$ and $i_2 = 1, 2, \cdots, j_2$. The data matrix is converted to the correspondence matrix P by dividing N by its grand total $n = \sum_{i_1=1}^{j_1} \sum_{i_2=1}^{j_2} n_{i_1 i_2}$, i.e $P = \frac{N}{n}$. Let $r_{i_1} = \sum_{i_2=1}^{j_2} p_{i_1 i_2}$ and $c_{i_2} = \sum_{i_1=1}^{j_1} p_{i_1 i_2}$ be the row and column masses (marginal), $\frac{p_{i_1 i_2}}{r_{i_1}}$ and $\frac{p_{i_1 i_2}}{c_{i_2}}$ be the i_1^{th} row and i_2^{th} column profile respectively. Let $D_r = diag(\underline{r})$ and $D_c = diag(\underline{c})$ be the diagonal matrices of row and column masses respectively where $\underline{r} = \begin{pmatrix} r_1 & r_2 & ... & r_{j_1} \end{pmatrix}^T$ and $\underline{c} = \begin{pmatrix} c_1 & c_2 & ... & c_{j_2} \end{pmatrix}^T$. The computational algorithm to obtain coordinates of the row and column profile with respect to the principal axes using SVD is as follows.

(i) Calculate the matrix S of standardized residuals of order $j_1 \times j_2$.

$$S = D_r^{-1/2}(P - rc^T)D_c^{-1/2} \text{ i.e. } s_{i_1 i_2} = \frac{p_{i_1 i_2} - r_{i_1} c_{i_2}}{\sqrt{r_{i_1} c_{i_2}}}, i_1 = 1, 2, ..., j_1 \text{ and } i_2 = 1, 2, ..., j_2$$

(ii) Now perform SVD on S as,

$$S = U\Sigma V^T$$

where $UU^T = I$ and $VV^T = I$, $\rho = rank(S)$ and $\Sigma = diag(\sigma_1, \sigma_2, ..., \sigma_\rho, 0, ..., 0)$ and $\sigma_1, \sigma_2, ..., \sigma_\rho$ are non-negative singular values of S in descending order.

(iii) Calculate the total inertia of the data matrix as: $intertia = \sum_{i_1=1}^{j_1} \sum_{i_2=1}^{j_2} s_{i_1 i_2}^2$

The Chi-square statistic is calculated as: $\chi^2 = n \times intertia$

(iv) Obtain the standard, principal and biplot coordinates of rows and columns as follows.

Coordinates	Rows	Columns
Standard Coordinates	$\Phi = D_r^{-1/2} U$	$\Gamma = D_c^{-1/2} V$
Principal Coordinates	$F = D_r^{-1/2} U \Sigma$	$F = D_r^{-1/2} V \Sigma^T$
Biplot Coordinates	$\tilde{F} = U\Sigma^{\gamma}$, $\gamma = 0, 1, 1/2$	$\tilde{G} = V\Sigma^{1-\gamma}$

The columns of Φ, Γ, F ,G, $\tilde{F}$, $\tilde{G}$ matrices are referred as the principal axes, or dimensions, of the solution. The coordinates have different scaling and have different interpretations. For exploring the association between two CVs, the joint map of row and column coordinates is obtained. The variations in plots are as follows.

- **Correspondence plot using standard coordinates:** The correspondence plot obtained by using row and column standard coordinates gives equal weights to each of the dimensions and the weight associated with each dimension of a plot is 1. Hence, a unit principal inertia is associated with each of the dimension.
- **Correspondence plot using principal coordinates:** In this plot the row and column principal coordinates are plotted. The objective is to reflect the strength of the association that exists between the variables. In this case, the m^{th} principal axis is associated with an inertia σ_m^2 instead of unit principal inertia.
- **Biplot:** The biplot coordinates are obtained by the rescaling the principal coordinates, so as to provide a meaningful interpretation of the distance between a row and a column principal coordinates in a low-dimensional space. The advantage of using biplot is that the distance between row and column point makes some sense, unlike the correspondence plot obtained by using principal coordinates. For this reason, the use of biplot in a CA context has exponentially grown up over the years. For more details refer Beh and Lombardo (2014).

The interpretations related to numerical results, correspondence plot and biplot are as follows.
Inertia: The inertia is equivalent to the statistical concept of variance. The higher inertia score indicates a stronger model fit (i.e. large variance). The singular value indicates the relative contribution of each dimension to an explanation of the inertia, or proportion of variation, in the participant and variable profiles. The singular values can be interpreted as the correlation between the rows and columns of the contingency table.
Chi-square statistic: The Chi-square test of independence is used to determine whether

the association between two CVs is significant using the Chi-square statistic. For the significant Chi-square value, the association between the two CVs is confirmed.

Biplot: The biplot visualizes the row and column points in a joint map. From the graphical display if the row and column points are plotted close to one another, and if the same row and column points significantly contribute to the total inertia in the SCA, then the two CVs are concluded to be associated.

The method of SCA, which studies the association between only two CVs, is extended to MCA. MCA explores the association between more than two CVs. We discuss the method of classical MCA in the following.

3.2 Classical MCA

It is the favored approach to MCA in which multi-way contingency table is transformed into an indicator matrix or a Burt matrix and then SCA is applied to one of them. This approach is an extension of the SCA which allows one to analyze the pattern of relationships among several categorical dependent variables. We first introduce notations involved in MCA.

Consider a study which consists of n records on p CVs. Let j_k be the number of categories of k^{th} CV, $k = 1, 2, ..., p$ and X_k be the $n \times j_k$ indicator matrix with $(i_1, i_2)^{th}$ element 1(0) if i_1^{th} individual or unit is (is not) classified into i_2^{th} category of that CV. Then $n \times J$ matrix $X = \begin{bmatrix} X_1 & X_2 & ... & X_p \end{bmatrix}$ is called super-indicator matrix where, $J = \sum_{k=1}^{p} j_k$. The $J \times J$ matrix $B = X^T X$ plays an important role in performing MCA and is called a Burt matrix. For multi-way contingency table, the Burt matrix $B = X^T X$ is as follows.

$$B = \begin{pmatrix} X_1^T X_1 & X_1^T X_2 & ... & X_1^T X_p \\ X_2^T X_1 & X_2^T X_2 & ... & X_2^T X_p \\ \vdots & \vdots & \ddots & \vdots \\ X_p^T X_1 & X_p^T X_2 & ... & X_p^T X_p \end{pmatrix} = \begin{pmatrix} D_1 & X_1^T X_2 & ... & X_1^T X_p \\ X_2^T X_1 & D_2 & ... & X_2^T X_p \\ \vdots & \vdots & \ddots & \vdots \\ X_p^T X_1 & X_p^T X_2 & ... & D_p \end{pmatrix}$$

The Burt matrix B has a square block D_j on the diagonal where D_j is the diagonal matrix with marginal frequencies of j^{th} CV, $j = 1, 2, ..., p$. Each of the off-diagonal submatrix is a rectangular block of a two-way contingency table associated with a pair of CV.

Classical MCA can be performed in two ways, either by performing SCA on super-indicator matrix X or on B. The two forms of classical MCA are discussed in the following.

(a) **Computations based on an indicator matrix**
In this approach, SCA is performed on the super-indicator matrix. It gives $J - p$ nonzero singular values and the squared singular values represent principal inertias.

The column and row standard or principal coordinates for the first two dimensions are to be calculated for graphical display of the results.

(b) **Computations based on Burt matrix** (B)
The computation of MCA in this case is an application of the SCA algorithm to the Burt matrix B. In this approach, the singular values represent principal inertias. Since Burt matrix is symmetric, the standard or principal coordinates for rows and columns are identical and hence any one of them are used to obtain graphical association.
Some of the properties of MCA based on super-indicator matrix are related to that of MCA based on the Burt matrix. These are listed in the following

Greenacre (2007) differentiates the properties of MCA based on super-indicator matrix with that of MCA based on the Burt matrix. These are listed in the following.

- The standard coordinates of the rows (equivalent to columns) of Burt matrix are identical to the standard coordinates of the columns of indicator matrix X.
- The principal inertias of the Burt analysis are the squares of those of the indicator matrix. Hence the percentages of inertia are always going to be higher in the Burt analysis. Actually, in Burt analysis the total inertia ($inertia(B)$) is the average of the inertias of all subtables, including the offensive ones on the diagonal.
- The difference between the computation methods is that the Burt version of MCA gives principal coordinates which are reduced in scale compared to the indicator version, where the reduction is relatively more on the second axis compared to the first.

In classical MCA based on Burt matrix, the percentages of inertia is artificially high since the analysis tries to explain the inertia in the whole table with higher inertias on the diagonal. Hence the inertia explained by the first dimension is severely underestimated. Thus, the inclusion of the tables on the diagonal of the Burt matrix degrades the whole MCA solution. To overcome this drawback Greenacre (2007) introduces JCA. The JCA is the approach where the inflated total inertia in classical MCA is adjusted. The details of the method are discussed in the following section.

3.3 Joint CA

In this approach, adjusted Burt matrix is considered to rectify the problem in the first technique. In the method of classical MCA using Burt matrix, the block diagonals of Burt matrix have extremely high inertias which lead to inflation of inertia. It is possible to improve the calculation of explained inertia by completely ignoring the diagonal blocks in search of an optimal solution. To do this, Greenacre (1988) proposes a special algorithm called JCA. JCA is fitting of the off-diagonal cross-tabulations of all pairs of variables, ignoring the cross-tabulations on the block diagonal of the Burt matrix.

The algorithm to perform JCA is an iterative algorithm which performs SCA on the Burt matrix in such a way that attention is focused on optimizing the fit to the off-diagonal blocks only. The method starts from the classical MCA solution and then replaces the diagonal blocks with values estimated from the solution itself, using the reconstitution formula (Greenacre (2007)). Then the relative frequencies of the diagonal blocks of the Burt matrix are replaced with the estimated values which give the modified Burt matrix. SCA is performed on the modified Burt matrix to get a new solution, from which the diagonal blocks are replaced again with estimates from the new solution to get a new modified Burt matrix. This process is repeated several times until convergence, and at each iteration, the fit to the off-diagonal blocks is improved. Greenacre (1988) summarizes the inertia components of MCA of p CVs through the inertia of diagonal blocks and off-diagonal blocks of the Burt matrix as follows.

Total inertia of the Burt analysis: $inertia(B)$

Chi-square statistic for Burt analysis, $\chi^2_{(B)}$: $n \times inertia(B)$

Sum of inertias of p diagonal blocks: $J - p$

Sum of inertias of all two-way blocks: $p^2 inertia(B) = p^2 \frac{\chi^2_{(B)}}{n}$

Sum of inertias of all off-diagonal blocks: $p^2 \frac{\chi^2_{(B)}}{n-(J-p)}$

Average inertia of off-diagonal blocks: $\frac{p}{(p-1)} \left(\frac{\chi^2_{(B)}}{n} - \frac{(J-p)}{p^2} \right)$

The average inertia of off-diagonal blocks of Burt matrix is nothing but the adjusted inertia in JCA. The total inertia of the Burt matrix is artificially inflated by $(J - p)$ in classical MCA.

For the classical MCA or JCA, the categorical data must be in the form of super-indicator matrix X. When more number of CVs and records are under study, it would be difficult to construct such an indicator matrix and its size would be relatively more. The other approach is MCA based on the separate SVDs in which multi-way contingency table is used directly and is discussed in the following.

3.4 MCA based on Separate SVDs

Kroonenberg (2008) discusses the notion of separate SVDs for analysis of a 3-way array of subjects by variables and by conditions. To study the association of such a 3-way array, Kroonenberg (2008) treats it as a set of matrices and apply two-mode SVD to each of the matrix in a set. We discuss this method in the following.

Let N be a three-way data array (contingency table) of size $j_1 \times j_2 \times j_3$ associated with n records on three CVs A, B and C. Let j_k be the number of categories of k^{th} CV, $k = 1, 2, 3$. Let $n_{i_1 i_2 i_3}$ be the number of records classified into i_k^{th} category of k^{th}

CV, $i_k = 1, 2, ..., j_k, k = 1, 2, 3$. The corresponding 3-way contingency table can be represented as given in Table 1.

Table 1: The representation of 3-way contingency table

Categories of variable	C												
A↓ B⟶	1	2	$\cdots$	j_2	1	2	$\cdots$	j_2	$\cdots$	1	2	$\cdots$	j_2
1	n_{111}	n_{121}	$\cdots$	n_{1j_21}	n_{112}	n_{122}	$\cdots$	n_{1j_22}	$\cdots$	n_{11j_3}	n_{12j_3}	$\cdots$	$n_{1j_2j_3}$
2	n_{211}	n_{221}	$\cdots$	n_{2j_21}	n_{212}	n_{222}	$\cdots$	n_{2j_22}	$\cdots$	n_{21j_3}	n_{22j_3}	$\cdots$	$n_{2j_2j_3}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
j_1	n_{j_111}	n_{j_121}	$\cdots$	$n_{j_1j_21}$	n_{j_112}	n_{j_122}	$\cdots$	$n_{j_1j_22}$	$\cdots$	$n_{j_11j_3}$	$n_{j_12j_3}$	$\cdots$	$n_{j_1j_2j_3}$

To perform classical MCA, a super-indicator matrix X of size $n \times J$ is to be obtained using N where $J = \sum_{k=1}^{3} j_k$. In this method, MCA is performed on N rather than the super-indicator matrix X or the Burt matrix $B = X'X$ of size $J \times J$.

In the method of MCA based on separate SVDs, SVD is performed on each of the frontal slice of the standardized residual matrix. We discuss the slice representation of 3-way array in the following.

A three-way array can be seen as a collection of two-way matrices, often referred as slices or slabs. There are three different types of slices/arrangements for this and are referred to as horizontal slices, lateral slices and frontal slices. Thus, a three-way data array N can be visualized as an array of j_1 horizontal slices (of order $j_2 \times j_3$) as in Figure 2(a) or an array of j_2 lateral slices (of order $j_1 \times j_3$) as in Figure 2(b) or an array of j_3 frontal slices (of order $j_1 \times j_2$) as in Figure 2(c) (Kroonenberg (2008)). We denote i_1^{th} horizontal slice of N by $N(i_1, :, :)$, i_2^{th} lateral slice of N by $N(:, i_2, :)$ and i_3^{th} frontal slice of N is denoted by $N(:, :, i_3)$.

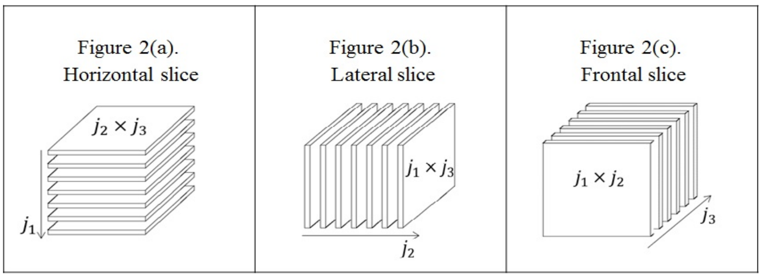

Figure 2: Slices of three-way data array: Horizontal slices, Lateral slices and Frontal slices

Particularly N in Figure 2(c) can be expressed as in Figure 3.

Thus, in the method of MCA based on separate SVDs, SVD is performed either on

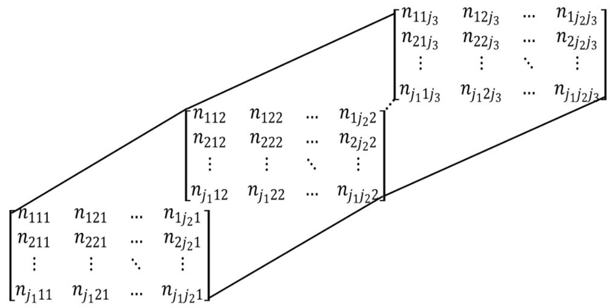

Figure 3: Representation of Frontal slices of three-way array N

the horizontal slices or on lateral slices or on the frontal slice of the standardized residual matrix. When interest is in studying the association of CVs A and B across the categories of C, separate SVDs is to be performed on the frontal slices of standardized residuals array. The algorithm for performing MCA based on separate SVDs on the frontal slices is as follows.

(i) Obtain the three-way array $P = ((p_{i_1i_2i_3}))$ as $P = \frac{N}{n}$ so that $0 \leq p_{i_1i_2i_3} < 1, \forall (i_1, i_2, i_3)$ and $\sum_{i_1=1}^{j_1} \sum_{i_2=1}^{j_2} \sum_{i_3=1}^{j_3} p_{i_1i_2i_3} = 1$.

(ii) Obtain $\hat{P} = ((\hat{p}_{i_1i_2i_3}))$ where $\hat{p}_{i_1i_2i_3} = p_{i_1..}p_{.i_2.}p_{..i_3}$, $i_k = 1, 2, ...j_k, k = 1, 2, 3$.

and $p_{i_1..} = \sum_{i_2=1}^{j_2} \sum_{i_3=1}^{j_3} p_{i_1i_2i_3}, i_1 = 1, 2, ..., j_1,$

$p_{.i_2.} = \sum_{i_1=1}^{j_1} \sum_{i_3=1}^{j_3} p_{i_1i_2i_3}, i_2 = 1, 2, ..., j_2,$

$p_{..i_3} = \sum_{i_1=1}^{j_1} \sum_{i_2=1}^{j_2} p_{i_1i_2i_3}, i_3 = 1, 2, ..., j_3.$

Let $\underline{r} = \begin{pmatrix} p_{1..} & p_{2..} & \cdots & p_{j_1..} \end{pmatrix}^T$ be the vector of masses for CV A,

$\underline{c} = \begin{pmatrix} p_{.1.} & p_{.2.} & \cdots & p_{.j_2.} \end{pmatrix}^T$ be the vector of masses for CV B,

$\underline{t} = \begin{pmatrix} p_{..1} & p_{..2} & \cdots & p_{..j_3} \end{pmatrix}^T$ be the vector of masses for CV C.

(iii) Calculate the array of standardized residuals $S = ((s_{i_1i_2i_3}))$ of size $j_1 \times j_2 \times j_3$ as,

$$s_{i_1i_2i_3} = \frac{p_{i_1i_2i_3} - \hat{p}_{i_1i_2i_3}}{\sqrt{\hat{p}_{i_1i_2i_3}}}, i_k = 1, 2, ..., j_k, k = 1, 2, 3$$

Let ρ_{i_3} be the rank of i_3^{th} frontal slice $S(:,:,i_3)$ (of size $j_1 \times j_2$) of S, $i_3 = 1, 2, ..., j_3$.

(iv) Application of SVD on each of the frontal slice $S(:,:,i_3)$ gives the following decompositions.

$$S(:,:,i_3) = U(:,:,i_3)\Sigma(:,:,i_3)V(:,:,i_3)^T, i_3 = 1, 2, ..., j_3,$$

where $U(:,:,i_3)$, $V(:,:,i_3)$ are the orthogonal matrices of size $j_1 \times j_1$ and $j_2 \times j_2$ respectively and $\Sigma(:,:,i_3)$ is the rectangular diagonal matrix of non-negative singular values $\sigma^2_{i_31}, \sigma^2_{i_32}, ..., \sigma^2_{i_3s}$ of size $j_1 \times j_2$.
The total inertia of the 3-way array is,

$$\sum_{i_3=1}^{j_3}\sum_{s=1}^{\rho_{i_3}} \sigma^2_{i_3s} = \sum_{i_1=1}^{j_1}\sum_{i_2=1}^{j_2}\sum_{i_2=1}^{j_3} S^2_{i_1i_2i_3},$$

and the Chi-square statistic is,

$$\chi^2 = n\sum_{i_3=1}^{j_3}\sum_{s=1}^{\rho_{i_3}} \sigma^2_{i_3s}.$$

(v) Obtain the standard and principal coordinates of biplot associated with CV A and B across C. The standard coordinates can be obtained through the following matrices for $i_3 = 1, 2, ..., j_3$.

$\Phi(:,:,i_3) = D_r^{-\frac{1}{2}}U(:,:,i_3)$, (of size $j_1 \times j_2$),

and $\Gamma(:,:,i_3) = D_c^{-\frac{1}{2}}V(:,:,i_3)$, (of size $j_2 \times j_1$).

While the principal coordinates and biplot coordinates are calculated through the following matrices.

$F(:,:,i_3) = D_r^{-\frac{1}{2}}U(:,:,i_3)\Sigma(:,:,i_3)$, (of size $j_1 \times j_2$),

and $G(:,:,i_3) = D_c^{-frac-12}V(:,:,i_3)\Sigma(:,:,i_3)^T$, (of size $j_2 \times j_1$),

$\tilde{F}(:,:,i_3) = U(:,:,i_3)\Sigma(:,:,i_3)^\gamma$, (of size $j_1 \times j_2$), and $\tilde{G}(:,:,i_3) = V(:,:,i_3)(\Sigma(:,:,i_3)^T)^{\gamma}$, (of size $j_2 \times j_1$).

Now, the first two columns of $\tilde{F}(:,:,i_3)$ and $\tilde{G}(:,:,i_3)$ give j_1 pairs of principal coordinates for CV A and j_2 pairs of principal co-ordinates for CV B across i_3^{th} category of C respectively, $i_3 = 1, 2, ..., j_3$.

(vi) To visualize the association of A and B across categories of C, plot $j_1 j_3$ co-ordinates from $\tilde{F}$ and $j_2 j_3$ co-ordinates from $\tilde{G}$ to get biplot which is equivalent to superimposition of j_3 biplots for CVs A and B across each category of C (but not equivalent to superimposing of j_3 biplots obtained by performing SCA on $N(:,:,i_3)$. Since the coordinates of all the biplots are obtained from a standardized residual array S, we overlay the biplots to visualize the association of any pair of CVs across the categories of other. This plot explores the association of CVs A and B across all the categories of C.

The total inertia obtained through the MCA based on separate SVDs is less than that of obtained through MCA based on Burt matrix. The differences in the MCA using Burt matrix and MCA based on separate SVDs are due to the use of transformation of the multi-way array to indicator matrix/Burt matrix in the former case. It can be observed that the elements of Burt matrix B are sub-totals (i.e. the totals of horizontal, lateral and frontal slices) of 3-way array N. In general, the relation between the Burt matrix $B = (b_{ij})$ of size $J \times J$ and the 3-way contingency table $N = (n_{i_1 i_2 i_3})$ of size $j_1 \times j_2 \times j_3$ is as follows. For $j = 1, 2, ..., J$ and $i \leq j$,

$$b_{ij} = \begin{cases} n_{i..} & \text{if } i, j = 1 : j_1; \\ n_{.i-j_1.} & \text{if } i, j = j_1 + 1 : j_1 + j_2; \\ n_{..i-(j_1+j_2)} & \text{if } i, j = j_1 + j_2 + 1 : j_1 + j_2 + j_3; \\ n_{i(j-i_1).} & \text{if } i = 1 : j_1, j = j_1 + 1 : j_1 + j_2; \\ n_{i.j-(j_1+j_2)} & \text{if } i = 1 : j_1, j = j_1 + j_2 + 1 : j_1 + j_2 + j_3; \\ n_{.i-j_1(j-(j_1+j_2))} & \text{if } i = j_1 + 1 : j_1 + j_2, j = j_1 + j_2 + 1 : j_1 + j_2 + j_3; \\ 0 & \text{elsewhere.} \end{cases}$$

Features of the MCA based on separate SVDs

(a) When interest is in studying the association between CV A and C (B and C) across the categories of $B(A)$, separate SVD on the lateral slices (horizontal slices) of S is to be performed.

(b) It is observed that if MCA is performed for all pairs of CVs across the other CV, then it leads to the exactly equal total inertia and similar interpretations about the association. Hence, it is reasonable to perform MCA based on separate SVDs on any one of the slices as per interest.

(c) In this method, the algorithm utilizes a complete 3-way contingency table data in its original form in the calculation of a 3-way array of standardized residuals and performs separate SVDs on either its horizontal or lateral or frontal slices. Thus, this algorithm is not equivalent to performing SCA for each of the horizontal or lateral or frontal slices of the N.

Overall, the MCA based on separate SVDs allows one to study the association of any pair of CVs across the other. Further, since separate SVDs performed on an array of standardized residuals, the problem of inflation of total inertia as in the case of classical MCA is overcome.

The drawback of this method is that all analyses are independent in the sense that in no way, the SVD of one frontal slice is related to that of another frontal slice. The separate SVDs is advantageous when there are no individual differences or when there is no interest in the modeling terms (Kroonenberg (2008)).

The other approach to study the pairwise association between the CVs is stacking or concatenation of multi-way table in the two-way table and is discussed in the next subsection.

3.5 MCA based on Stacking and Concatenation

Stacking method of MCA is a moderately less popular approach and is discussed by Weller and Romney (1990). In this approach three-way contingency table is stacked to form a two-way contingency table. Performing MCA via stacking involves forming a two-way contingency table from the multi-way table by placing each slice of the array on top of each other. Depending on the association structure that the researcher is interested in exploring, the stacking can be done. We discuss this approach for three-way contingency table.

Let N be $i_1 \times i_2 \times i_3$ be a 3-way contingency table. Now, the stacking can be done in three-ways. To study the association between CVs B and C the horizontal slices $N(i_1, :, :)$ of N are stacked into a two-way table while to study the association between A and C, the lateral slices $N(:, i_2, :)$ of N are stacked into two-way table. If one is interested in studying the association between A and B, the frontal slices $N(:, :, i_3)$ of N are stacked into a two-way table. The stacking of frontal slices can be represented as follows.

$$\begin{pmatrix} N(:,:,1) \\ N(:,:,2) \\ \vdots \\ N(:,:,i_3) \end{pmatrix} = \left(\begin{array}{cccc} n_{111} & n_{121} & \cdots & n_{1j_21} \\ n_{211} & n_{221} & \cdots & n_{2j_21} \\ \vdots & \vdots & \ddots & \vdots \\ n_{j_111} & n_{j_121} & \cdots & n_{j_1j_21} \\ \hline n_{112} & n_{122} & \cdots & n_{1j_22} \\ n_{212} & n_{222} & \cdots & n_{2j_22} \\ \vdots & \vdots & \ddots & \vdots \\ n_{j_112} & n_{j_122} & \cdots & n_{j_1j_22} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline n_{11j_3} & n_{12j_3} & \cdots & n_{1j_2j_3} \\ n_{21j_3} & n_{22j_3} & \cdots & n_{2j_2j_3} \\ \vdots & \vdots & \ddots & \vdots \\ n_{j_11j_3} & n_{j_12j_3} & \cdots & n_{j_1j_2j_3} \end{array} \right)$$

The stacked table is a two-way contingency table to which SCA is performed.

Another type of stacking related to the Burt matrix is known as concatenation, which is discussed by Greenacre and Blasius (1994). Concatenation is the stacking of the bivariate marginal for two particular variables. The Burt matrix considers all three concatenations simultaneously, with the univariate marginals included. The demonstration of this method is given through example in Section 4.

Techniques of MCA discussed so far consists of applying SVD to the standardized residual matrix. The other technique is to use generalized SVD, and is discussed by Kroonenberg (2008), Carlier and Kroonenberg (1996), Beh (1998), etc. We review the modelling approach briefly.

3.6 The modeling approaches to CA

The approach for performing MCA using a generalization of SVD known as a modeling approach. There are many models of decomposition such as the PARAFAC model propsed by Harshman (1970), the CANDECOMP model by Carroll and Chang (1970) and Tucker 3 model by Tucker (1963), Tucker (1966) and are discussed by Beh (1998). Thus, modeling approach is used by performing generalized SVD on three-way contingency table. Beh (1998) discusses the Tucker 3 model, the CANDECOMP model, and PARAFAC model in detail.

3.7 Ordinal CA

The SCA of a two-way contingency table is a very versatile tool to understand the structure of the association among CVs. In cases where the variables consist of ordered categories, there are a number of approaches that can be employed and these generally involve an adaptation of SVD. An alternative decomposition method has been also used for cases where the row and column variables of a two-way contingency table have an ordinal structure. SCA of a two-way contingency table using an amalgamation of SVD and Bivarite Moment Decomposition (BMD) is known as ordinal SCA. A benefit of this technique is that it combines the classical technique with the ordinal analysis by determining the structure of the variables in terms of singular values and location, dispersion and higher-order moments.

The major problems of SCA can be overcome by considering ordinal CA. In ordinal CA, orthogonal polynomial are generated and it require a set of scores which reflect the ordered structure of a set of categories. While showing categories of CV within a variable may or may not be different, there is no clear interpretation from SCA, of how these within-variable categories may or may not be different. The technique or ordinal CA solve this particular problem and it was developed by Beh (1998).

Beh (1998) used Emerson (1968) orthogonal polynomials which require the input of a scoring scheme to reflect the ordered structure of the categories. These orthogonal polynomials used to quantify ordered variables. The method of SCA and MCA using orthogonal polynomials visualizes the relationship between the categories, in terms of the location, dispersion and higher order components, when the data consists of at least one ordered CV.

The ordered CA of the symmetric association between the variables allows for a partitioning of the inertia into sources of variation attributable by specific orthogonal polynomials. Such a partitioning is based on the relevance and/or significance of the generalized correlations between these polynomials. These correlations are also used to evaluate the sources of inertia due to these polynomials.

Now, we demonstrate each of the above discussed techniques to perform MCA through examples. The numerical results, as well as a graphical display (biplot) associated with each of the technique, are given in the next section.

4 Applications

In this section, we apply the different approaches of MCA to the behavioral data. We study attitude of Americans towards life according to marital status, attitude of individuals towards the abortion according to their religion and years of education and association pattern in mother-child behavior across time using different methods of CA. We also compare the results of MCA based on separate SVDs with that of the classical MCA and JCA.

Example 1 A General Social Survey conducted on 995 Americans in 1993. In this survey the individuals are classified according to their marital status and their attitude towards life as dull, routine or exciting. The objective is to study the association between marital status and attitude towards life. The two-way contingency table associated with this survey is given in Table 2. We analyze this data using SCA. We develop Matlab function to perform SCA. The details of nonzero singular values (SV), inertia and %

Table 2: Two-way contingency table classifying 996 Americans according to their Marital Status and Attitude towards life

Marital Status	Attitude about life		
	Dull	Routine	Exciting
Married	21	241	251
Widowed	17	54	40
Divorced	10	74	65
Separated	6	11	8
Never Married	11	79	108

of total inertia led by SCA are summarized in Table 3. While Figure 4 represents the correspondence plot of principal coordinates which shows the pattern of association between the marital status and attitude towards life obtained by performing SCA.

Table 3: Result of SCA

SV	Inertia	% of Inertia
0.1878	0.0353	89.53
0.0642	0.0041	10.47
0.0000	0.0000	0.00
Total	0.0394	100
χ^2 statistic	39.242	
p-value	1.5424E-08	

The principal inertia along the two principal axes is 0.0394. Thus, it is a remarkably good plot for representing the variation of row and column profiles. The association between marital status and attitude towards life from Figure 4 can be interpreted as follows.

- Never married individuals have exciting attitude towards life.
- Separated individuals have dull attitude towards life.
- Divorced individuals have routine attitude towards life.

Now, we use classical method of MCA, JCA, MCA based on separate SVDs and stacking to the data in Example 2. We compare the results of all these methods.

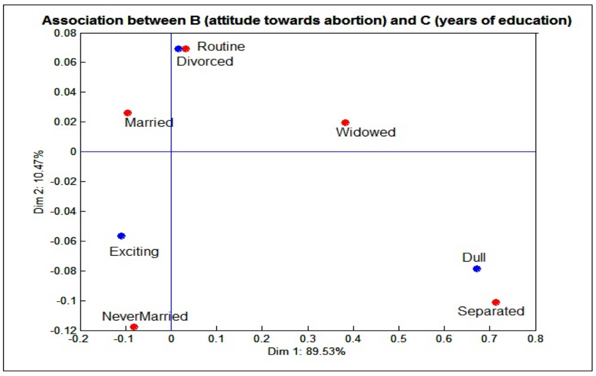

Figure 4: Biplot for SCA

Example 2 Consider the cross-classification of 3181 individuals about attitude towards abortion, according to religion and years of formal education (data used by Böckenholt and Böcknholt (1990) and then by DAmbra and Amenta (2011) and D'Ambra et al. (2012)). The summary of CVs and their categories is given in Table 4 while the associated contingency data given in Table 5.

Table 4: Summary of CVs and their categories for Example 2

Name of the CV	Categories of CV
Religion (A)	Northern Protestant (A_1), Southern Protestant (A_2), Catholic (A_3)
Attitude towards abortion (B)	Positive (B_1), Neutral (B_2), Negative (B_3)
Years of education (C)	$\leq 8(C_1)$, $9-12(C_2)$, $\geq 13(C_3)$

Table 5: Cross-classification of 3181 individuals for Example 2

Categories of variable	C_1			$C \downarrow$ C_2			C_3		
$A \downarrow B \rightarrow$	B_1	B_2	B_3	B_1	B_2	B_3	B_1	B_2	B_3
A_1	49	46	115	293	140	277	244	66	100
A_2	27	34	117	134	98	167	138	38	73
A_3	25	40	88	172	103	312	93	57	135

(a) Analysis based on Classical MCA

We analyze this data using classical MCA and joint CA in (ca-package) of R-software. The details of numerical results obtained by performing classical MCA using an indicator and Burt matrix and JCA are summarized in Table 6, Table 7 and Table 8 respectively. While Figure 5, 6 and 7 represent the biplot which shows the pattern of association between the categories of all the three CVs obtained by performing classical MCA using an indicator and Burt matrix and JCA respectively.

Table 6: Results of classical MCA using indicator matrix

SV	Inertia	% of Inertia
0.9505	0.9034	45.2
0.8255	0.6815	34.1
0.4508	0.2032	10.2
0.4080	0.1664	8.3
0.1738	0.0302	1.5
0.1235	0.0152	0.8
Total	2.0000	100

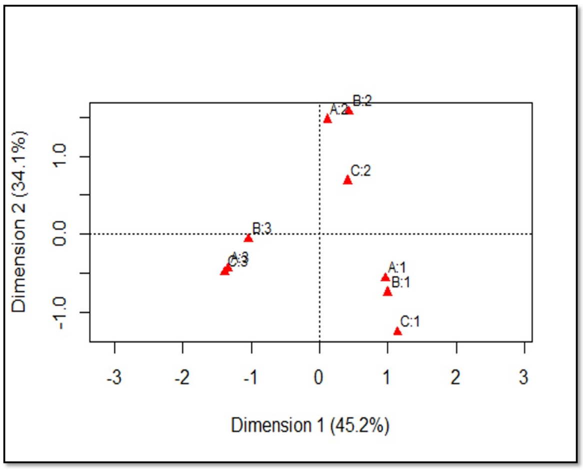

Figure 5: Biplot for MCA using indicator matrix

The total inertia for classical MCA using indicator matrix is 2.00 while using Burt

Table 7: Results of classical MCA using Burt matrix

SV	Inertia	% of Inertia
0.9034	0.8162	60.43
0.6815	0.4644	34.38
0.2032	0.0413	3.06
0.1664	0.0277	2.05
0.0302	0.0001	0.07
0.0152	0.0002	0.02
Total	1.3499	100
χ^2 statistic	4294.032	
p-value	0.00	

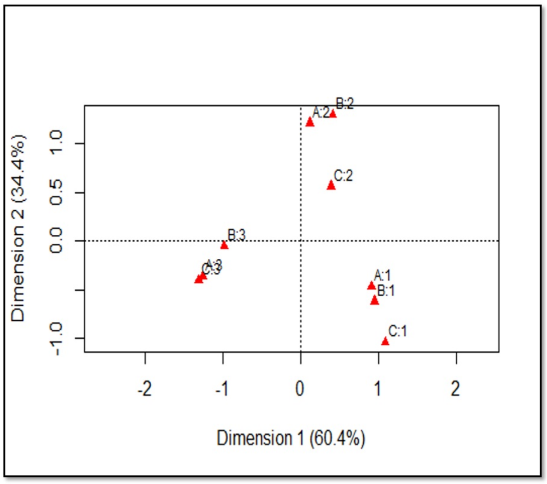

Figure 6: Biplot for MCA using Burt matrix

Table 8: Results of JCA

SV	Inertia	% of Inertia
0.8551	0.7313	71.3
0.5222	0.2727	26.6
Total	1.0261	97.9
χ^2 statistic	3264.024	
p-value	0.00	

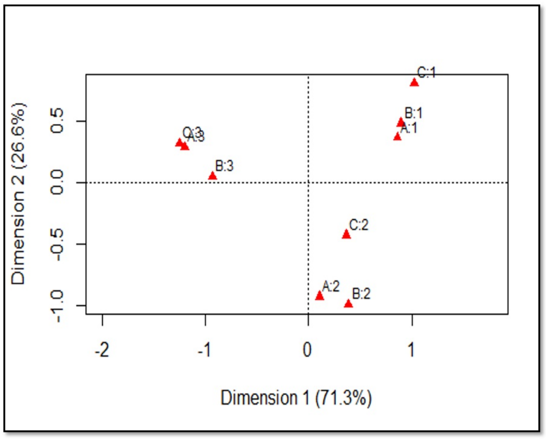

Figure 7: Biplot for JCA

matrix it is 1.35. The percentage of inertia explained by the first principal axis for classical MCA using Burt matrix is 60.43%, which is higher than 45.2% that of indicator matrix. The total inertia for JCA is 1.0261 while the percentage of inertia explained from the first principal axis is 71.3%. The total inertia of classical MCA with indicator matrix seems to be overestimated. Overall, the association pattern in Figure 5, Figure 6 and Figure 7 are similar and summarized as follows.

- North Protestant (A_1) with less than 8 years of formal education (C_1) tend to have positive attitude (B_1) towards abortion.
- South Protestant (A_2) with 9-12 years of education (C_2) tend to have neutral attitude (B_2) towards abortion.
- Catholic (A_3) with more than 12 years of education (C_3) tend to have negative attitude (B_3) towards abortion.

(b) Analysis using MCA based on Separate SVD
A 3-way array N associated with contingency data in Table 4 is of size $3 \times 3 \times 3$. Since here the objective is to study the association between the attitude towards abortion (B) and years formal of education (C) across the religion (A), we perform MCA based on separate SVDs on horizontal slices of N (i.e. along B and C across the categories of A). The horizontal slices of the 3-way array (i.e. cross-classification matrix along its second and third dimension) can be specified as follows.

$$N(:,:,1) = \begin{pmatrix} 49 & 293 & 244 \\ 46 & 140 & 66 \\ 115 & 277 & 100 \end{pmatrix}, N(:,:,2) = \begin{pmatrix} 27 & 134 & 138 \\ 34 & 98 & 38 \\ 117 & 167 & 73 \end{pmatrix},$$

$$N(:,:,3) = \begin{pmatrix} 25 & 172 & 93 \\ 40 & 103 & 57 \\ 88 & 312 & 135 \end{pmatrix}$$

Table 9 shows details of nonzero singular values (SV), inertia and % of inertia obtained by performing MCA based on separate SVDs while Figure 8 shows the corresponding biplot.

Table 9: Results of MCA based on separate SVDs for Example 2 along B and C across the categories of A

Categories of A	SV	Inertia	% of Inertia	Total
	0.1950	0.0380	41.41	
A_1	0.0145	0.0002	0.23	41.71
	0.0084	0.0001	0.08	
	0.1733	0.0300	32.7	
A_2	0.0714	0.0051	5.55	38.34
	0.0088	0.0001	0.09	
	0.1207	0.0146	15.86	
A_3	0.0601	0.0036	3.93	19.95
	0.0120	0.0001	0.15	
Total		0.0918	100	100
χ^2statistic	292.0158			
p-value	4.20 E-60			

The points far from the origin and close to each other are considered for interpretations from the biplot since more the vector length, better is the discrimination ability. The interest would be in pairs of categories of CVs B and C across the same category of A. Such pairs of points are marked in ellipses. It can be seen from Figure 6 that the biplot points associated with the pairs (C_3A_1, B_1A_1), (C_3A_2, B_1A_2), (C_1A_2, B_3A_2), (C_2A_2, B_2A_2) and (B_3A_3, C_2A_3) are close to each other and also have more discrimination ability. For each of these points the association between the CVs B and C across the same category of A can be interpreted as follows.

- (C_3A_1, B_1A_1) and (C_3A_2, B_1A_2): North protestants (A_1) and South protestants (A_2) with more than 13 years of formal education (C_3) tend to have positive attitude (B_1) towards abortion.

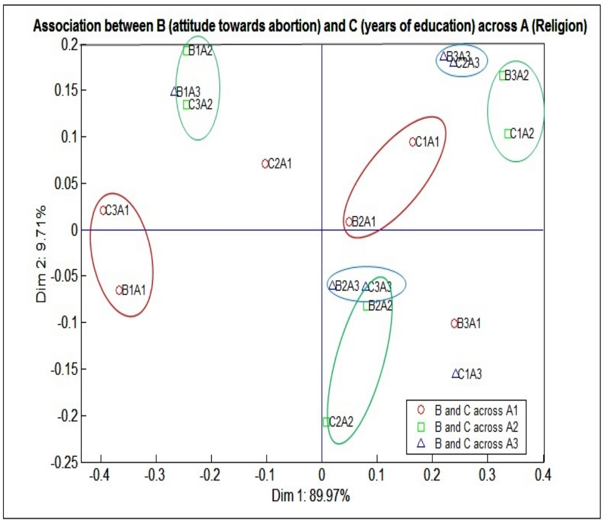

Figure 8: Biplot showing association between B (attitude towards abortion) and C (years of education) across A (Religion) (% of inertia explained is 99.68)

- (C_1A_2, B_3A_2): South protestants (A_2) with less than 8 years of formal education (C_1) tend to have negative attitude (B_3) towards abortion.
- (C_2A_2, B_2A_2): South protestants (A_2) with 9-12 years of formal education (C_2) tend to have neutral attitude (B_2) towards abortion.
- (B_3A_3, C_2A_3): The catholics (A_3) in the study with 9-12 years of formal education (C_2) tend to have negative attitude (B_3) towards abortion.

Thus, for the North Protestant and South Protestant regions, as the number of years of formal education increases, the attitude towards abortion tend to increase (negative to positive). Figure 8 shows the clear association between the attitude towards abortion and years of formal education across the religion.
We perform MCA based on separate SVDs on both lateral slices (i.e. along A and C across categories of B) and frontal slices (i.e. along A and B across categories of C). Table 10 and Table 11 summarize the numerical results for the MCA along A and C across the categories of B and MCA along A and B across categories of C respectively. While Figure 9 and Figure 10 shows the pattern of association for these cases.

The interpretations from the overlaid biplots in Figure 7 and Figure 8 are almost equivalent as from Figure 6. Also, it can be seen that the inertia for MCA along any pair of CVs across the other is the same. Thus, it is reasonable to perform

Table 10: Results of MCA based on separate SVDs for Example 2 along A and C across categories of B

Categories of B	SV	Inertia	% of Inertia	Total
	0.1919	0.0368	40.11	
B_1	0.1059	0.0112	12.21	54.03
	0.0398	0.0016	1.72	
	0.0460	0.0021	2.31	
B_2	0.0169	0.0003	0.32	2.82
	0.0134	0.0002	0.20	
	0.1650	0.0272	29.65	
B_3	0.0942	0.0089	9.66	43.14
	0.0593	0.0035	3.83	
Total		0.0918	100	100
χ^2 statistic	292.0158			
p-value	4.20 E-60			

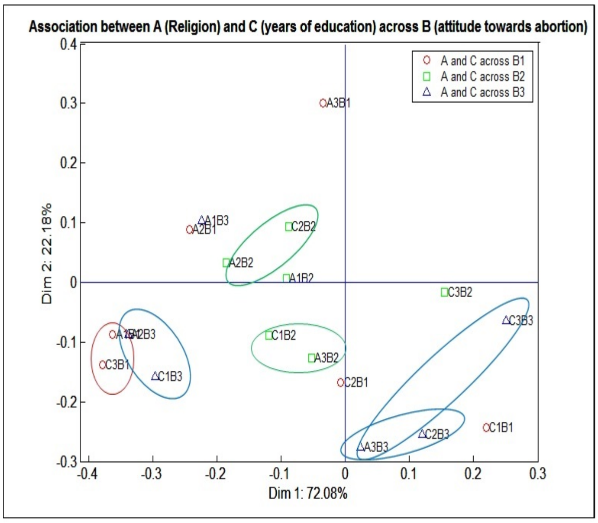

Figure 9: Biplot showing association between A (religion) and C (years of education) across B (attitude towards abortion) (% of inertia explained is 94.26)

Table 11: Results of MCA based on separate SVDs for Example 2 along A and B across categories of C

Categories of C	SV	Inertia	% of Inertia	Total
	0.1714	0.0294	32	
C_1	0.0689	0.0047	5.17	37.24
	0.0081	0.0001	0.07	
	0.1038	0.0108	11.73	
C_2	0.0565	0.0032	3.48	15.25
	0.0056	0.0000	0.03	
	0.2079	0.0432	47.07	
C_3	0.0184	0.0003	0.37	47.52
	0.0084	0.0001	0.08	
Total		0.0918	100	100
χ^2 statistic	292.0158			
p-value	4.20 E-60			

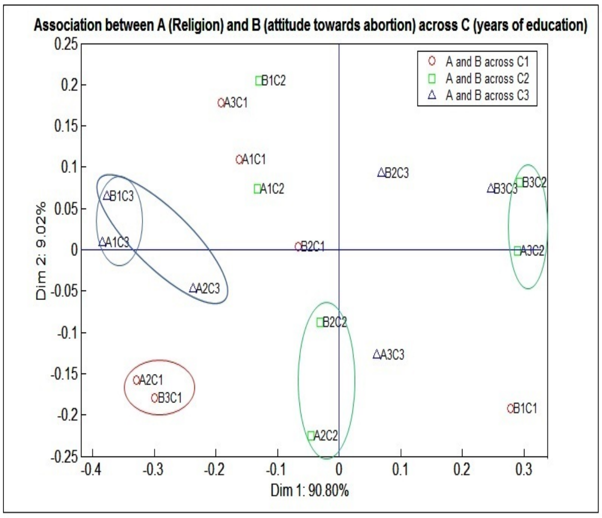

Figure 10: Biplot showing association between A (religion) and B (attitude towards abortion) across C (years of education) (% of inertia explained is 99.82)

MCA along any one pair of CVs. For a given situation the pair of CVs may be chosen as per the interest of the researcher.

(c) MCA with Stacking
Continuing with Example 2, if the objective is to study the association between the attitude towards abortion and years of education across the religion only perform SCA on the stacked two-way contingency table which is obtained by stacking $N(1,:,:)$, $N(2,:,:)$ and $N(3,:,:)$ of N in Table 5. The details of numerical results are given in Table 12 and the corresponding plot is given in Figure 10. The first and second principal axis accounts 86.48% and 13.52% of the total inertia in the data. Thus, the plot in Figure 10 explains 100% of the total inertia. The association between the years of formal education and attitude towards abortion can be interpreted as follows.

Table 12: Results of MCA using stacking

SV	Inertia	% of Inertia
0.2435	0.0593	86.48
0.0963	0.0093	13.52
Total	0.0686	100
χ^2 statistic	218.2166	
p-value	1.61 E-37	

- The individuals above 13 years of formal education (C_3) have a positive attitude (B_1) towards abortion.
- The individuals with less than 8 years of formal education (C_1) have a negative attitude (B_3) towards abortion.
- The individuals with 9-12 years of formal education (C_2) have a negative attitude (B_3) towards abortion.

Here, the attitude of each religion North protestant, south protestant and catholic is investigated towards the abortion.

Example 3 This example considers a data about 30 mother-child pairs' behavior during the first six months of life. This data is analyzed by Carlier and Kroonenberg (1996) and collected by Van den Boom (1988). The data is about infant-mother pairs' behavior with 7 infant categories and 6 behaviors of mother across first 6 months with 143100 respondents. Thus, the data set under consideration forms a three-way contingency table of order $7 \times 6 \times 6$. Out of these three variables, one variable is ordinal but is considered

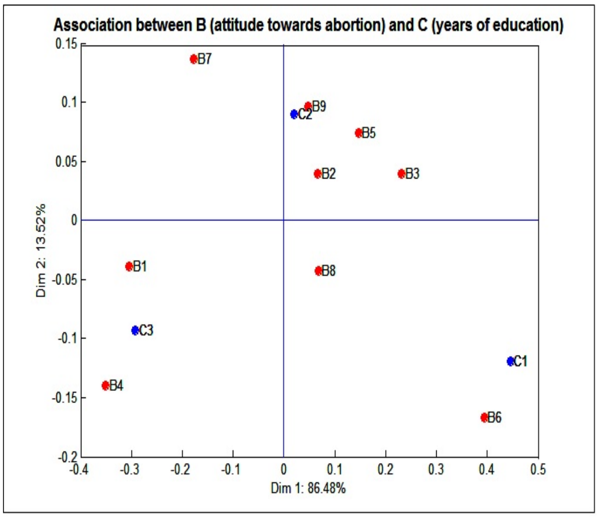

Figure 11: Biplot of MCA using Stacking

nominal. The summary of CVs and their categories is given in Table 13.
We perform MCA based on separate SVDs along A and B across the categories of C.

Table 13: Summary of CVs and their categories for Example 3

Name of the CV	Categories of CV
Infant behavior (A)	Inactive (A_1), Smile (A_2), Look (A_3), Vocalize (A_4), Explore (A_5), Crying (A_6), Sucking (A_7)
Mother's Behavior (B)	Other (B_1), Looking (B_2), Stimulating (B_3), Offering (B_4), Contact (B_5), Soothing (B_6)
Month (C)	First Month (C_1), Second Month (C_2), Third Month (C_3), Fourth Month (C_4), Fifth Month (C_5), Sixth Month (C_6)

Table 14 shows details of nonzero singular values, inertia (I) and % of inertia I, while Figure 7 shows the corresponding biplot. The interpretations of MCA based on separate SVDs are equivalent to those given by Carlier and Kroonenberg (1996). Further, inertia in both the methods is equal and is 0.7513. The pairs of biplot points close to each other, and their interpretations from Figure 7 are as follows.

- $(A_6C_i, B_6C_i), i = 1:6:$ When the infant is crying (A_6), irrespective of month (for all six months), mother is soothing (B_6)(Marked with same colored ellipses).

Table 14: Results of MCA based on separate SVDs for Example 3

Categories of C	SV	I	% of I	Total	Categories of C	SV	I	% of I	Total
C_1	0.3159	0.0998	13.28	25.57	C_4	0.2451	0.0601	8.00	11.61
	0.2738	0.0749	9.97			0.1252	0.0157	2.09	
	0.1062	0.0113	1.50			0.0778	0.0061	0.81	
	0.0619	0.0038	0.51			0.0467	0.0022	0.29	
	0.0438	0.0019	0.26			0.0438	0.0019	0.26	
	0.0181	0.0003	0.04			0.0360	0.0013	0.17	
C_2	0.3222	0.1038	13.82	20.93	C_5	0.2199	0.0483	6.43	12.92
	0.1604	0.0257	3.42			0.1867	0.0348	4.64	
	0.1261	0.0159	2.12			0.0864	0.0075	0.99	
	0.0924	0.0085	1.14			0.0708	0.0050	0.67	
	0.0484	0.0023	0.31			0.0341	0.0012	0.15	
	0.0311	0.0010	0.13			0.0162	0.0003	0.04	
C_3	0.2300	0.0529	7.04	11.32	C_6	0.2705	0.0732	9.74	17.65
	0.1508	0.0227	3.03			0.2047	0.0419	5.58	
	0.0694	0.0048	0.64			0.1087	0.0118	1.57	
	0.0526	0.0028	0.37			0.0615	0.0038	0.50	
	0.0357	0.0013	0.17			0.0407	0.0017	0.22	
	0.0232	0.0005	0.07			0.0181	0.0003	0.04	
Total							0.7513	100	100

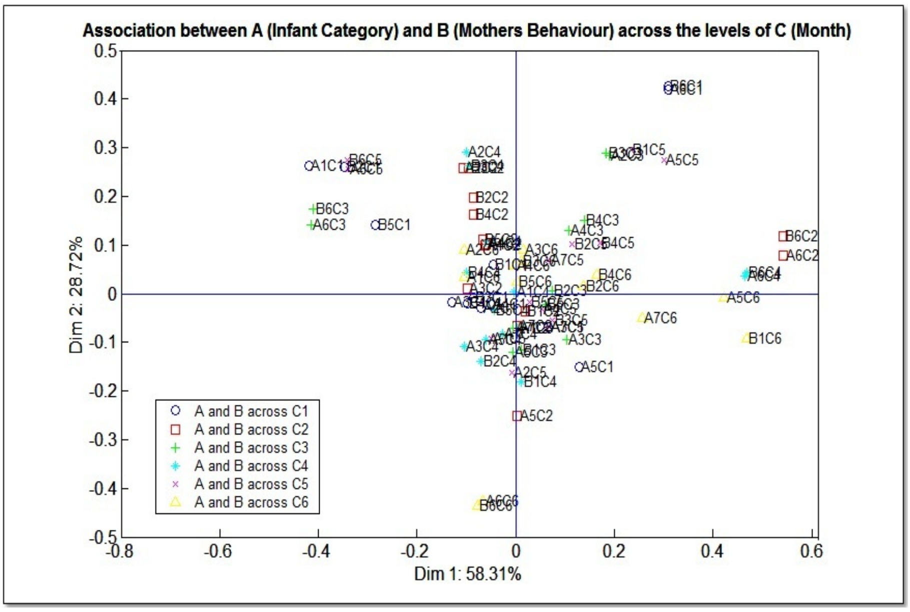

Figure 12: Biplot showing association between Mother-child behavior over Time)

- (A_5C_5, B_1C_5): The infant is exploring (A_5) when mother is engaged in other activities (B_1) in fifth the month (C_5).
- (A_1C_1, B_2C_1) and (A_1C_1, B_5C_1): The infant is inactive (A_1) when mother tends to seek look (B_2) and contact (B_5) at the infant in the first month (C_1).
- (A_2C_2, B_3C_2) and (A_4C_2, B_3C_2): The infant is smiling (A_2) and vocalizing (A_4) when mother stimulates (B_3) in the second month (C_2).
- (A_3C_3, B_1C_3) and (A_3C_3, B_5C_3): The infant is looking (A_3) when mother is doing other child non-related things (B_1) and contact (B_5) at the infant in the third month (C_3).
- (A_2C_4, B_3C_4) and (A_2C_4, B_4C_4): The infant is smiling (A_2) when the mother is stimulating (B_3) and offering (B_4) to the infant in the fourth month (C_4).
- $(A_7C_5B_1C_5), (A_7C_5, B_2C_5)$ and (A_7C_5, B_4C_5): The infant is sucking (A_7) when mother is doing other child non-related things (B_1), looking (B_2) and offering (B_4) to the infant in the fifth month (C_5).
- (A_5C_6, B_1C_6) and (A_7C_6, B_1C_6): The infant is exploring (A_5) and sucking (A_7) when mother does other non-child related things (B_1) in the sixth month (C_6).

The interpretation of biplot points which are in opposite direction are as follows.

- (A_1C_1, B_4C_1): The infant is inactive (A_1) in the first month (C_1) then the mother not offering (B_4).
- (A_1C_5, B_1C_5): The infant is inactive (A_1) in the fifth month (C_5) then the mother not engaged in the other activities (B_1).
- (A_1C_6, B_3C_6): The infant is inactive (A_1) in the sixth month (C_6) then mother is not stimulating (B_3).

Thus, it is observed that the interpretations obtained through the overlaid biplot in Figure 11 for MCA based on separate SVDs are equivalent to the interpretations from joint biplot obtained using a three-way generalization of the SVD as given by Carlier and Kroonenberg (1996).

5 Discussion

The main objective of this paper is to provide a brief overview of different approaches to MCA and discuss its applications in the behavioral studies. We summarize the popular approaches of MCA along with their advantages and disadvantages and demonstrate these through examples. CA is applicable in a diverse range of situations. It has been beneficial in a variety of research areas such as social sciences, engineering, health sciences, medicine, archeology, ecology, software development, market research, etc. Examples of the use of CA can be found in medical research (Greenacre (1992)), students' and teachers' cognitions about good teachers (Beishuizen et al. (2001)), higher education institution image (Ivy (2001)), personalities (Nishisato (2014)), marketing research (Bendixen (1996)). Considering the widespread applicability of CA, we hope that this paper will motivate the researchers to apply this powerful tool of MCA for achieving their research objectives in behavioral sciences.

Acknowledgement

The first author would like to thank UGC, New Delhi for providing support for this work through the Rajiv Gandhi National Fellowship (Award Letter No.: F1-17.1/2011-12/RGNF-SC-MAH-5331).

References

Beh, E. J. (1998). Correspondence analysis using orthogonal polynomials.

Beh, E. J. (2004). Simple correspondence analysis: a bibliographic review. *International Statistical Review*, 72(2):257–284.

Beh, E. J. and Lombardo, R. (2012). A genealogy of correspondence analysis. *Australian & New Zealand Journal of Statistics*, 54(2):137–168.

Beh, E. J. and Lombardo, R. (2014). *Correspondence analysis: theory, practice and new strategies*. John Wiley & Sons.

Beishuizen, J., Hof, E., Putten, C., Bouwmeester, S., and Asscher, J. (2001). Students and teachers cognitions about good teachers. *British Journal of Educational Psychology*, 71(2):185–201.

Bendixen, M. (1996). A practical guide to the use of correspondence analysis in marketing research. *Marketing Research On-Line*, 1(1):16–36.

Bnzecri, J. (1969). Statistical analysis as a tool to make patterns emerge from data [c]. *Methodologies of Pattern Recognition*, pages 35–74.

Böckenholt, U. and Böcknholt, I. (1990). Canonical analysis of contingency tables with linear constraints. *Psychometrika*, 55(4):633–639.

Carlier, A. and Kroonenberg, P. M. (1996). Decompositions and biplots in three-way correspondence analysis. *Psychometrika*, 61(2):355–373.

Carroll, J. D. and Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of eckart-young decomposition. *Psychometrika*, 35(3):283–319.

DAmbra, A. and Amenta, P. (2011). Correspondence analysis with linear constraints of ordinal cross-classifications. *Journal of classification*, 28(1):70–92.

D'Ambra, L., D'Ambra, A., and Sarnacchiaro, P. (2012). Visualizing main effects and interaction in multiple non-symmetric correspondence analysis. *Journal of Applied Statistics*, 39(10):2165–2175.

Emerson, P. L. (1968). Numerical construction of orthogonal polynomials from a general recurrence formula. *Biometrics*, pages 695–701.

Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, pages 453–467.

Greenacre, M. (1992). Correspondence analysis in medical research. *Statistical methods in medical research*, 1(1):97–117.

Greenacre, M. (2007). *Correspondence analysis in practice*. CRC press.

Greenacre, M. J. (1988). Correspondence analysis of multivariate categorical data by weighted least-squares. *Biometrika*, pages 457–467.

Greenacre, M. J. (2010). *Biplots in practice*. Fundacion BBVA.

Greenacre, M. J. and Blasius, J. (1994). *Correspondence analysis in the social sciences: Recent developments and applications*, volume 4. Academic Press London.

Harshman, R. A. (1970). Foundations of the parafac procedure: Models and conditions for an" explanatory" multi-modal factor analysis.

Ivy, J. (2001). Higher education institution image: acorrespondence analysis approach. *International Journal of Educational Management*, 15(6):276–282.

Kroonenberg, P. M. (2008). *Applied multiway data analysis*, volume 702. John Wiley & Sons.

Nishisato, S. (2014). *Elements of dual scaling: An introduction to practical data analysis*. Psychology Press.

Tucker, L. R. (1963). Implications of factor analysis of three-way matrices for measurement of change. *Problems in measuring change*, 122137.

Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311.

Van den Boom, D. C. (1988). *Neonatal irritability and the development of attachment: Observation and intervention*. na.

Weller, S. C. and Romney, A. K. (1990). *Metric scaling: Correspondence analysis*. Number 75. Sage.