An Extension of Multiple Correspondence Analysis for Identifying Heterogeneous Subgroups of Respondents

Heungsun Hwang

HEC Montréal

William R. Dillon

Southern Methodist University

and

Yoshio Takane

McGill University

Abstract

An extension of multiple correspondence analysis is proposed that takes into account cluster-level heterogeneity in respondents' preferences/choices. The method involves combining multiple correspondence analysis and k-means in a unified framework. The former is used for uncovering a low-dimensional space of multivariate categorical variables while the latter is used for identifying relatively homogeneous clusters of respondents. The proposed method offers an integrated graphical display that provides information on cluster-based structures inherent in multivariate categorical data as well as the interdependencies among the data. An empirical application is presented which demonstrate the usefulness of the proposed method and how it compares to several extant approaches.

Keywords: Multiple correspondence analysis, k-means, cluster-level respondent heterogeneity, alternating least squares.

1. Introduction

Multiple correspondence analysis (MCA) is a popular descriptive technique to explore the relationships among multiple categorical variables (Benzécri, 1973; Gifi, 1990; Greenacre, 1984; Lebart, Morineau, & Warwick, 1984; Nishisato, 1980). It amounts to a nonlinear principal components analysis that assigns numerical scores to respondents and response categories of dummy-coded categorical variables which results in a graphical map of the interdependencies among the variables. This graphical display is a useful bi-product as it aids in communicating the association structures inherent in multivariate categorical data to practitioners and other researchers.

The parameters of MCA are estimated by pooling the data across respondents under the implicit assumption that all respondents come from a single, homogenous group. However, it often seems more realistic to assume that respondents come from heterogeneous groups, so that they are different with respect to their attitudes and preferences. Such group- or cluster-level respondent heterogeneity has been discussed from several different theoretical and modeling perspectives; for example, consumer belief structures have been hypothesized to vary across different market segments according to the expectancy value models (Bagozzi, 1982), and accounting for cluster-level respondent heterogeneity has been shown to be important in modeling consumer brand choice decisions (cf. Kamakura, Kim & Lee, 1996).

A convenient practice for accommodating cluster-level respondent heterogeneity in MCA is to adopt a two-step sequential, tandem approach (Arabie & Hubert, 1994): In the first step a low-dimensional representation of the categorical variables is obtained via MCA; in the second step some variety of cluster analysis is used to identify a set of

relatively homogenous respondent groups on the basis of the low-dimensional data (e.g., Arimond & Elfessi, 2001; Green & Krieger, 1998; Green, Schaffer, & Patterson, 1988; Lebart, 1994). In addition to the ease in which the two-step sequential approach can be implemented, there can be substantive reasons for adopting this approach as well (see Green & Krieger, 1995).

Despite its popularity, many authors have warned about a critical problem which is inherent to the tandem approach. Specifically, there is no guarantee that the low-dimensional representation of the data obtained in step one is optimal for subsequently identifying cluster structures because data reduction is carried out with no reference to cluster analysis (Arabie & Hubert, 1994; Chang, 1983; DeSarbo, Jedidi, Cool, & Schendel, 1990; De Soete & Carroll, 1994). This suggests that preliminary data reduction may mask or even distort the true cluster structure that is present in the original data. Green and Krieger (1995) offer empirical examples which support the legitimacy of this concern. Similarly, Vichi and Kiers (2001) present a simulation-based example in which tandem analysis failed to identify correct clusters in the context of principal components analysis. Technically, this problem stems from the fact that each step of the tandem approach involves a different optimization criterion (i.e., one criterion for data reduction and another for cluster analysis) and that these criteria are addressed separately.

As a solution to the problem, the combined use of data reduction and cluster analysis in a single framework has been recommended (Bock, 1987; DeSarbo, Howard, & Jedidi, 1991; De Soete & Carroll, 1994; Heiser, 1993; Van Buuren & Heiser, 1989; Vichi & Kiers, 2001). In essence, this amounts to obtaining a low-dimensional representation of variables and classifying respondents into a set of clusters

simultaneously. This simultaneous approach may ensure that low-dimensional data are optimally chosen so as to facilitate the identification of clusters.

In this paper, we extend MCA in such a way that it can explicitly account for cluster-level heterogeneity in respondents' preferences/choices. Specifically, we combine MCA with the k-means algorithm (MacQueen, 1967) in a unified framework. The k-means algorithm is perhaps the most popular method for non-overlapping clustering (Wedel & Kamakura, 1998). When compared to other non-overlapping clustering methods, it appears to be more robust against outliers and is less afflicted by irrelevant variables or dimensions in the data (Punj & Stewart, 1983). It is also efficient in dealing with large data sets (Green, Carmone, & Kim, 1990). More importantly, the k-means algorithm is attractive in the context of the proposed method because it is easily compatible with the distribution-free optimization procedure of MCA. This compatibility is beneficial in combining two separate optimization criteria for MCA and k-means into a single one.

The paper is organized as follows. In Section 2, the proposed method is discussed in full detail. A fitting criterion is presented along with an alternating least squares algorithm developed to optimize this criterion for parameter estimation. In Section 3, a comparison of the proposed method and several related methods appearing in the extant literature is presented. In Section 4, an empirical data set is used to illustrate the usefulness of the proposed method and to demonstrate how it compares to other extant approaches. The final section is devoted to implications with focus on discussing several additional aspects of the proposed method.

2. The Proposed Method

Let $\mathbf{Z}_j$ denote an n by p_j matrix of the j-th dummy-coded categorical variable, where n is the number of respondents, and p_j is the number of response categories in the variable ($j = 1,\ldots, J$). Let $\mathbf{F}$ denote an n by $d\ (\leq p_j)$ matrix of a d-dimensional representation of J categorical variables. Let $\mathbf{W}_j$ denote a p_j by d matrix of weights, also called category quantifications. Let c denote the prescribed number of clusters. Let $\mathbf{U}$ denote an n by c matrix of an indicator variable, allocating respondents into one of c clusters. Let $\mathbf{\Gamma}$ denote a c by d matrix of the centroids or mean values of clusters. Let α_1 and α_2 denote non-negative scalar weights.

At this point, the aim is to combine MCA and k-means into a single framework. This amounts to finding a low-dimensional representation from more than two categorical variables and at the same time identifying clusters of respondents that are relatively homogeneous in terms of the low-dimensional representation. This problem is equivalent to minimizing the following:

$$f = \alpha_1 \sum_{j=1}^{J} \mathrm{SS}(\mathbf{F} - \mathbf{Z}_j \mathbf{W}_j) + \alpha_2 \mathrm{SS}(\mathbf{F} - \mathbf{U\Gamma}), \tag{1}$$

with respect to $\mathbf{F}$, $\mathbf{W}_j$, $\mathbf{U}$, and $\mathbf{\Gamma}$, subject to $\mathbf{F'F} = \mathbf{I}$ and $\alpha_1 + \alpha_2 = 1$, where $\mathrm{SS}(\mathbf{M}) = \mathrm{tr}(\mathbf{M'M})$. When $\alpha_1 = 1$, the first term in (1) amounts to the standard homogeneity criterion for MCA (Gifi, 1990; Yanai, 1998). When $\alpha_2 = 1$, the second term is equivalent to the standard criterion used in the k-means clustering algorithm. However, by minimizing both criteria in (1) simultaneously, the low-dimensional data representation,

F, is obtained in such a way that it recognizes the cluster structure that may be inherent in the multivariate categorical variables.

The values of α_1 and α_2 are a priori specified by the investigator. By specifying $\alpha_1 = \alpha_2 = .5$, we aim to balance out the two terms for MCA and k-means. On the other hand, we may aim to adjust for the relative importance of the two terms by differently weighing them. For example, an investigator may wish to weigh the second term more heavily than the first term (i.e., $\alpha_2 > \alpha_1$) under the belief that classification is of more importance than data reduction in his/her study.

An alternating least squares (ALS) algorithm (de Leeuw, Young, & Takane, 1976) is developed to minimize (1). In the algorithm, the unknown parameters, $\mathbf{F}$, $\mathbf{W}_j$, $\mathbf{U}$, and $\mathbf{\Gamma}$, are updated alternately until convergence. The updates of one parameter matrix are obtained such that they minimize (1) in the least squares sense, while the others are fixed. Specifically, the ALS algorithm repeats the following three main steps until convergence is obtained:

Step 1. $\mathbf{W}_j$ and $\mathbf{\Gamma}$ are updated for fixed $\mathbf{F}$ and $\mathbf{U}$. Since $\mathbf{W}_j$ is only involved in the first term and $\mathbf{\Gamma}$ is only in the second term of (1), their updates are simply given by

$$\hat{\mathbf{W}}_j = (\mathbf{Z}_j'\mathbf{Z}_j)^{-1}\mathbf{Z}_j'\mathbf{F}, \tag{2}$$

and

$$\hat{\mathbf{\Gamma}} = (\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'\mathbf{F}. \tag{3}$$

Step 2. $\mathbf{F}$ is updated for fixed $\mathbf{W}_j$, $\mathbf{U}$ and $\mathbf{\Gamma}$. Let $\mathbf{\Omega}_j = \mathbf{Z}_j(\mathbf{Z}_j'\mathbf{Z}_j)^{-1}\mathbf{Z}_j'$, and $\mathbf{\Psi} = \mathbf{U}(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'$. Note that $\mathbf{\Omega}_j$ and $\mathbf{\Psi}$ are idempotent and symmetric. Putting (2) and (3)

into (1), then, minimizing (1) with respect to $\mathbf{F}$, subject to $\mathbf{F'F} = \mathbf{I}$, amounts to minimizing

$$
\begin{aligned}
f^* &= \alpha_1 \sum_j \mathrm{SS}(\mathbf{F} - \mathbf{\Omega}_j \mathbf{F}) + \alpha_2 \mathrm{SS}(\mathbf{F} - \mathbf{\Psi F}) \\
&= \alpha_1 \sum_j \mathrm{tr}(\mathbf{F'F} - \mathbf{F'\Omega}_j \mathbf{F}) + \alpha_2 \mathrm{tr}(\mathbf{F'F} - \mathbf{F'\Psi F}) \\
&= \alpha_1 \sum_j \mathrm{tr}(\mathbf{F'F}) - \alpha_1 \sum_j \mathrm{tr}(\mathbf{F'\Omega}_j \mathbf{F}) + \alpha_2 \mathrm{tr}(\mathbf{F'F}) - \alpha_2 \mathrm{tr}(\mathbf{F'\Psi F}) \\
&= (\alpha_1 J + \alpha_2)d - \left\{ \alpha_1 \sum_j \mathrm{tr}(\mathbf{F'\Omega}_j \mathbf{F}) + \alpha_2 \mathrm{tr}(\mathbf{F'\Psi F}) \right\} \\
&= (\alpha_1 J + \alpha_2)d - \left\{ \alpha_1 \mathrm{tr}(\mathbf{F'}\left[\sum_j \mathbf{\Omega}_j \right]\mathbf{F}) + \alpha_2 \mathrm{tr}(\mathbf{F'\Psi F}) \right\} \\
&= (\alpha_1 J + \alpha_2)d - \left\{ \mathrm{tr}(\mathbf{F'}\left[\alpha_1 \sum_j \mathbf{\Omega}_j + \alpha_2 \mathbf{\Psi} \right]\mathbf{F}) \right\}.
\end{aligned}
\tag{4}
$$

Minimizing (4) thus reduces to maximizing

$$
\mathrm{tr}\left(\mathbf{F'}\left[\alpha_1 \sum_j \mathbf{\Omega}_j + \alpha_2 \mathbf{\Psi} \right]\mathbf{F} \right),
\tag{5}
$$

with respect to $\mathbf{F}$. This amounts to the eigenvalue decomposition of $\alpha_1 \sum_j \mathbf{\Omega}_j + \alpha_2 \mathbf{\Psi}$

(Yanai, 1998).

Step 3. $\mathbf{U}$ is updated for fixed $\mathbf{F}$, $\mathbf{W}_j$ and $\mathbf{\Gamma}$. This amounts to minimizing the second term

of (1) with respect to $\mathbf{U}$. We employ Lloyd's (1982) k-means algorithm to minimize the

criterion.

The ALS algorithm monotonically decreases the value of criterion (1) which, in

turn, is also bounded from below. The algorithm is therefore convergent. However, it

does not guarantee that the convergence point is the global minimum. In particular, the k-

means algorithm has been shown to be sensitive to local optima (Steinley, 2003). To

safeguard against local minima, we repeat the ALS procedure with a large number (say,

100) of random initial values for $\mathbf{U}$. (The initial values for $\boldsymbol{\Gamma}$ are obtained from $\mathbf{U}$.) We then compare the obtained function values after convergence and subsequently choose the solution associated with the smallest one. Besides $\mathbf{U}$ (and $\boldsymbol{\Gamma}$), MCA is applied to the original data and the resultant low-dimensional data are used as rational starts for $\mathbf{F}$. The initial values for $\mathbf{W}_j$ are obtained on the basis of $\mathbf{F}$.

In the proposed method, we need to decide a priori on the number of clusters, c, as well as the number of dimensions in the data, d. Choices of c and d are closely related to each other since the size of the centroid matrix, $\boldsymbol{\Gamma}$, also depends on the number of dimensions. One approach is to first select d by applying MCA to the data, and then decide c by examining how the values of (1) change with different numbers of clusters (Wedel & Kamakura, 1998). It is recommended that the number of clusters be greater than the number of dimensions (Van Buuren & Heiser, 1989; Vichi & Kiers, 2001). In practice, non-statistical criteria for evaluating the usefulness and relevance of clusters (e.g., cluster size, potential, etc.) will play a crucial role in deciding c as well (Arabie & Hubert, 1994; Wedel & Kamakura, 1998).

3. Comparison to Other Methods

In this section we discuss the proposed approach in the context of three related methods appearing in the literature. The first method is the ScoreCutter algorithm developed by Mucha (2002) which alternates between the steps of dual scaling (Nishisato, 1980) and clustering. A major limitation of this method, however, is that it is restricted to finding only two clusters in data; in contrast, the proposed method has no such a limitation on the number of clusters in data.

The second related method that warrants discussion is due to Nishisato (1984).
This method combines discriminant analysis and MCA and therefore can be viewed as a
forced classification; that is, under this method respondents must be assigned to
clusters/groups in advance of performing MCA, in contrast to the proposed method
which identifies group structure on the basis of the data. Thus, Nishisato's (1984) method
can be viewed as a-priori approach while the proposed method can be as a post-hoc
approach (Wind, 1978; Wedel & Kamakura, 1998).

Perhaps the most closely related method is that of Van Buuren and Heiser (1989).
Their method optimizes the MCA criterion under a specific classification restriction on

object scores: $g = \sum_{j=1}^{J} \text{SS}(\mathbf{X} - \mathbf{Z}_j \mathbf{W}_j)$, subject to $\mathbf{X} = \mathbf{U}\boldsymbol{\Gamma}$, where $\mathbf{X}$ is a matrix of object

scores. As a result, in their optimization algorithm, $\mathbf{U}$ and $\boldsymbol{\Gamma}$ are updated based on the
MCA solution for the object scores (see equation (3) in Van Buuren and Heiser, 1989). In
contrast, the proposed method explicitly combines two criteria for MCA and k-means
into a single criterion, and optimizes them simultaneously. Thus, $\mathbf{U}$ and $\boldsymbol{\Gamma}$ are updated
based on $\mathbf{F}$, which involves both MCA and classification solutions. More importantly, the
Van Buuren and Heiser method does not allow either the MCA or k-means criterion to be
weighted differentially in generating the solution.

In the subsequent section, the proposed method is empirically compared with the
tandem approach and Van Buuren and Heiser's method. The ScoreCutter algorithm is not
considered for comparison because of the limitation on the number of clusters.

4. Empirical Application

The present example is part of a large survey conducted by a South Korean
underwear manufacturer in 1997. In this example, 664 South Korean consumers were

asked to provide responses for 3 multiple-choice items. The first item asked which of eight brands of underwear the consumer most prefers: (b1) BYC, (b2) TRY, (b3) VICMAN, (b4) James Dean, (b5) Michiko-London, (b6) Benetton, (b7) Bodyguard, and (b8) Calvin Klein. Both domestic (b1, b2, b3, b4, and b7) and international (b5, b6, and b8) brands were included. The second item asked the attribute of underwear most sought by the consumers. The number of the attributes was 15: (1) comfortable, (2) smooth, (3) superior fabrics, (4) reasonable price, (5) fashionable design, (6) favorable advertisements, (7) trendy color, (8) good design, (9) various colors, (10) elastic, (11) store is near, (12) excellent fit, (13) design quality, (14) youth appeal, and (15) various sizes. The last item asked the age category of each consumer: (a1) 10 – 29, (a2) 30-49, and (a3) 50 and over.

We first applied MCA to the data so as to determine the number of dimensions. In MCA, the proportions of the total inertia accounted for by the inertias or squared singular values are underestimated because the total inertia is inflated due to fitting both diagonal and off-diagonal blocks of the Burt table (Greenacre, 1984; Hwang & Takane, 2002). To deal with this problem, it is recommended to adjust the inertias greater than $1/J$ using Benzécri's (1979) formula, quoted in Greenacre (1984, p.145). The cut-off point of $1/J$ is equivalent to the value at which Cronbach's alpha becomes zero (Nishisato, 1994, p.151). Table 1 offers the adjusted inertias and their percentages of the adjusted total inertia (i.e., the sum of the adjusted inertias).

[Insert Table 1 about here]

We chose $d = 2$ since the sizes of the adjusted inertias appeared to decrease slowly after the first two or three and also it is typically easier to interpret a two-

dimensional configuration of estimated category points than other higher-dimensional configurations (Rovan, 1994). The first two adjusted inertias explained about 78% of the adjusted total inertia. Next, under the fixed d, we examined the changes in the value of (1) across different numbers of clusters. We found that the optimization criterion values appeared to decrease gradually beyond $c = 3$, suggesting that no substantial changes in the criterion values are obtained by having more than three clusters. Therefore, we decided to choose $c = 3$. Furthermore, we initially specified $\alpha_1 = \alpha_2 = .5$ for the purpose of comparing the proposed method with the tandem approach and Van Buuren and Heiser's method, which do not accommodate any differential weighting of the fitting criteria.

Given the predetermined dimensionality, the number of clusters, α_1, and α_2, we repeated the ALS procedure with 100 random values for initial cluster membership and chose the parameter estimates associated with the smallest optimization criterion value as the final solution. Figure 1 displays the two-dimensional configuration of the final category points and centroids estimated from the proposed method.

[Insert Figure 1 about here]

To render the figure more concise, the centroids of the three clusters are labeled 'C1', 'C2', and 'C3'. The eight brand categories from (b1) to (b8) are labeled 'by', 'tr', 'vm', 'jd', 'ml', 'bn', 'bg', and 'ck', respectively. The 15 attribute categories are represented by their category numbers from 1 to 15. The 3 age categories are labeled 'a1', 'a2', and 'a3'. The order of the labels is equivalent to that of the age categories.

The middle right-hand portion of the display appears related to the first cluster, where its centroid, 'C1', is located. The centroid seems closer to such attributes as

'fashionable design (5)', 'trendy color (7)', 'good design (8)', 'design quality (13)', and 'youth appeal (14)'. This indicates that the respondents in the first cluster appear to prefer the design or style of underwear. These consumers are more likely to purchase such brands as 'Benetton (bn)', Bodyguard ('bg'), 'Calvin Klein ('ck'), 'James Dean (jd)', and 'Michiko-London (ml)'. Also, 'C1' is closer to 'a1', suggesting that the consumers in the first cluster are likely to belong to the young generation of teens and twenties. Approximately 34% (228) of the total respondents were classified into the first cluster.

The upper portion of the display appears related to the second cluster, whose centroid is represented by 'C2'. This centroid is closely located with such attributes as 'comfortable (1)', 'favorable advertisements (6)', 'elastic (10)', 'excellent fit (12)', and 'various sizes (15)'. It is also close to 'VICMAN (vn)'. It suggests that the respondents belonging to the second cluster are attracted to underwear brands that deliver good body fit. Respondents in this cluster are more likely to choose VICMAN than other brands and are more likely to belong to a middle age group, featured by 'a2'. About 32% of all respondents (214) were classified into this cluster.

Finally, the lower left-hand portion of the display appears associated with the third cluster as it embraces the centroid of the third cluster, 'C3'. This centroid is positioned close to such attributes as 'smooth (2)', 'superior fabrics (3)', 'reasonable price (4)', 'various colors (9)', and 'store is near (11)'. It suggests that the customers belonging to the third cluster value the practical aspects of underwear, such as price, textual quality, and store proximity. This cluster is also closely linked to such domestic brands as 'BYC (by)' and 'TRY (tr)'. Furthermore, the centroid of this cluster is also

close to 'a3', indicating that the third cluster is made up of largely older consumers. About 33% of all respondents (222) belonged to this cluster.

Under the same numbers of dimensions and clusters, we also applied the tandem approach and Van Buuren and Heiser's method to this data. For the tandem approach, we repeated the k-means algorithm with 100 different random values for initial cluster membership. For Van Buuren and Heiser's method, we used the MCA solution as rational starts for **X**. We also repeated their ALS algorithm with 100 different random values for initial cluster membership. The results of the tandem approach and Van Buuren and Heiser's method are displayed in Figures 2 and 3, respectively.

[Insert Figures 2 and 3 about here]

In Figure 2, the middle right-hand portion of the display appears related to the first cluster, the upper portion to the second cluster, and the lower left-hand portion to the third cluster. Though similar in spirit to the solution produced by the proposed method, this display is more difficult to interpret; notice it is more difficult to characterize each cluster since the estimated category points are located in a less discriminate way. For example, a number of category points, such as '10', '11', '13', '14', '15', 'ml', etc., have drifted away from others making it difficult to explicitly link a category with a cluster. Similarly, as shown in Figure 3, Van Buuren and Heiser's method also yields more difficult solutions to interpret as compared to the proposed method.

We used this data set to investigate the effects of changing the scalar weights on the final solution of the proposed method. A number of analyses with different scalar weights were performed. Though, in general, most solutions were similar to the solution obtained under no differential weighting there are two interesting cases to consider. The

first scenario specifies values for the scalar weights which are consistent with the actual contribution that MCA and k-means terms had on the final solution. Since the MCA term was responsible for 98% of the optimization criterion value while the k-means term accounted for only 2% we set $\alpha_1 = .98$ and $\alpha_2 = .02$; in other words the scalar weights disproportionately favors the MCA term. Figure 4 displays the two-dimensional configuration of category points and centroid estimates obtained from this differentially weighted analysis. It is interesting that this analysis offers quite similar solutions to that obtained by the tandem analysis. This suggests that this differential weighting scenario largely corresponds to the tandem analysis of the underwear example.

[Insert Figure 4 about here]

In the second scenario we reversed the weighting scheme, that is, we allowed the k-means term to be dominant by setting $\alpha_1 = .02$ and $\alpha_2 = .98$. Figure 5 displays the two-dimensional configuration of parameter estimates obtained. The figure shows that the category point estimates changed little from the no differential weighting scenario (see Figure 1). On the other hand, the centroid estimates seem to move farther away from the configuration origin which aids in characterizing the clusters as their associated centroids are more distinctive. This is likely a direct result of placing more emphasis on the clustering of respondents as opposed to focusing on variable associations. Nonetheless, the interpretation remains essentially the same as in the equally weighted analysis scenario.

[Insert Figure 5 about here]

15

5. Concluding Remarks

In this paper, we proposed an extension of MCA whereby k-means was combined in a unified framework. The proposed method affords an integrated graphical display that offers useful information on cluster-based structures inherent in multivariate categorical data as well as the interrelationships among them. As illustrated with empirical data, the proposed method is shown to provide more easily interpretable solutions than other extant methods. Moreover, the differential weighting feature of this method appears to be an important asset, which allows exploring the data from more diverse perspectives.

The use of categorical variables in surveys is becoming increasingly popular for various reasons (see Arimond & Elfessi, 2001; Dolničar & Leisch, 2001). In particular, they offer respondents faster and less tedious response formats in comparison to Likert-type items. Not only does this increase the likelihood that respondents will complete a survey, it also enables researchers to include more items within survey questionnaires. Furthermore, the use of categorical variables provides a simpler means of data collection/management, thus reducing data entry costs (Javalgi, Whipple, McManamon, & Edick, 1992). Some authors have argued that future developments in classification/segmentation techniques should allow for simpler data collection as well as graphical displays of segmentation findings at the same time (Javalgi et al., 1992; Green at al., 1988). In an effort to capture cluster-level respondent heterogeneity in MCA, the proposed method also appears to be a segmentation tool that satisfies these goals.

The proposed method can be extended in a number of ways to enhance its scope and range of applicability. For example, it may be worthwhile to deal with ordinal variables through data transformation such as a monotonic transformation (Kruskal,

1964; Ramsay, 1988, 1998). The proposed method may also be extended in such a way to accommodate a mixture of nominal and ordinal variables. Recently, Hwang and Takane (2002) proposed a comprehensive approach to constrained MCA, where linear constraints are imposed not only on variables but also on respondents. Incorporating such an approach may further broaden the scope of the method proposed herein. The proposed method employs k-means for cluster analysis, which offers a deterministic classification of respondents. However, it may be more favorable to replace k-means with other types of cluster analysis such as fuzzy clustering (Bezdek, 1974; Manton, Woodbury, & Tolley, 1994), which permits a probabilistic classification of respondents. All of these extensions warrant further attention and provide the fodder for future theoretical and empirical work.

References

Arabie, P., & Hubert, L. (1994). Cluster analysis in marketing research. In R. P. Bagozzi (Ed.), *Advanced Methods of Marketing Research* (pp. 160-189). Oxford: Blackwell.

Arimond, G., & A. Elfessi, A. (2001). A clustering method for categorical data in tourism market segmentation research. *Journal of Travel Research*, **39**, 391-397.

Bagozzi, R. P. (1982). A field investigation of causal relations among cognition, affect, intensions, and behavior. *Journal of Marketing Research*, **19**, 562-584.

Bezdek, J. C. (1974). Numerical taxonomy with fuzzy sets. *Journal of Mathematical Biology*, **1**, 57-71.

Bock, H. H. (1987). On the interface between cluster analysis, principal component analysis, and multidimensional scaling. In Bozdogan, H., & Gupta, A. K. (Eds.). *Multivariate Statistical Modeling and Data Analysis* (pp. 17-34). New York: Reidel.

Benzécri, J. P. (1973). *L'analyse des données. Vol. 2. L'analyse des correspondances*. Paris:Dunod.

Benzécri, J. P. (1979). Sur le calcul des taux d'inertia dans l'analyse d'un questionaire. Addendum et erratum à [BIN.MULT]. *Cahiers de L'analyse des Données*, **4**, 377-378.

Chang, W. (1983). On using principal components before separating a mixture of two multivariate normal distributions. *Applied Statistics*, **32**, 267-275.

de Leeuw, J., Young, F. W., & Takane, Y. (1976). Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika*, **41**, 471-503.

DeSarbo, W. S., Jedidi, K., Cool, K., & Schendel, D. (1990). Simultaneous multidimensional unfolding and cluster analysis: An investigation of strategic groups. *Marketing Letters*, **2**, 129-146.

DeSarbo, W. S., Howard, D. J., & Jedidi, K. (1991). MULTICLUS: A new method for simultaneous performing multidimensional scaling and clustering. *Psychometrika*, **56**, 121-136.

De Soete, G., & Carroll, J. D. (1994). K-means clustering in a low-dimensional Euclidean space. In Diday E. et al. (Eds.). *New Approaches in Classification and Data Analysis* (pp. 212-219). Heidelberg: Springer.

Dolničar, S., & Leisch, F. (2001). Behavioral market segmentation of binary guest survey data with bagged clustering. In Dorffner, G., Bischof, H., & Hornik, K. (Eds.). *ICANN 2001* (pp. 111-118). Berlin: Springer-Verlag.

Gifi, A. (1990). *Nonlinear Multivariate Analysis*. Chichester: Wiley.

Green, P. E., Carmone, F. J., & Kim, J. (1990). A preliminary study of optimal variable weighting in k-means clustering. *Journal of Classification*, **7**, 271-285.

Green, P. E., & Krieger, A. M. (1995). Alternative approaches to cluster-based market segmentation. *Journal of the Market Research Society*, **37**, 221-239.

Green, P. E., & Krieger, A. M. (1998). *User's Guide to HIERMAPR*. The Wharton School. University of Pennsylvania.

Green, P. E., Schaffer, C. M., & Patterson, K. M. (1988). A reduced-space approach to the clustering of categorical data in market segmentation. *Journal of the Market Research Society*, **30**, 267-288.

Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis.* London: Academic Press.

Heiser, W. J. (1993). Clustering in low-dimensional space. In Opitz, O., Lausen, B., & Klar, R. (Eds.). *Information and Classification: Concepts, Methods, and Applications* (pp. 162-173). Heidelberg: Springer-Verlag.

Hwang, H., & Takane, Y. (2002). Generalized constrained multiple correspondence analysis. *Psychometrika*, **67**, 211-224.

Javalgi, R., Whipple, T., McManamon, M., & Edick, V. (1992). Hospital image: A correspondence analysis approach. *Journal of Health Care Marketing*, **12**, 34-41.

Kamakura, W. A., Kim, B., & Lee, J. (1996). Modeling preference and structural heterogeneity in consumer choice. *Marketing Science*, **15**, 152-172.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, **29**, 1-27

Lebart, L. (1994). Complementary use of correspondence analysis and cluster analysis. In Greenacre, M. J., & Blasius, J. (Eds.). *Correspondence Analysis in the Social Sciences* (pp. 162-178). London: Academic Press.

Lebart, L., Morineau, A., & Warwick, K. M. (1984). *Multivariate Descriptive Statistical Analysis*. New York: Wiley.

Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, **28**, 129-37.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Le Cam, L. M. & Neyman, J. (Eds.). *Proceedings of the Fifth*

Berkeley Symposium on Mathematical Statistics and Probability (pp. 281-297). Berkeley: University of California Press.

Manton, K. G., Woodbury, M. A., & Tolley, H. D. (1994). *Statistical Applications Using Fuzzy Sets*. New York: John Wiley & Sons.

Mucha, H.-J. (2002). An intelligent clustering technique based on dual scaling. In S. Nishisato, Y. Baba, H. Bozdogan and K. Kanefuji (Eds.). *Measurement and Multivariate Analysis* (pp. 37-46), Tokyo: Springer Verlag.

Nishisato, S. (1980). *Analysis of Categorical Data: Dual scaling and Its Applications*. Toronto: University of Toronto Press.

Nishisato, S. (1984). Forced classification: A simple application of a quantitative technique. *Psychometrika*, **49**, 25-36.

Nishisato, S. (1994). *Elements of Dual scaling: An Introduction to Practical Data Analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Punj, G. & Stewart, D. W. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research*, **20**, 134-148.

Ramsay, J. O. (1988). Monotone regression splines in action (with discussion). *Statistical Science*, **3**, 425-461.

Ramsay, J. O. (1998). Estimating smooth monotone functions. *Journal of Royal Statistical Society B*, **60**, 365-375.

Rovan, J. (1994). Visualizing solutions in more than two dimensions. In Greenacre, M. J., & Blasius, J. (Eds.). *Correspondence Analysis in the Social Sciences* (pp. 210-229). London: Academic Press.

Steinley, D. (2003). Local optima in K-means clustering: What you don't know may hurt you. *Psychological Methods*, **8**, 294-302.

Van Buuren, S., & Heiser, W. J. (1989). Clustering n objects into k groups under optimal scaling of variables. *Psychometrika*, **54**, 699-706.

Vichi, M., & Kiers, H. A. L. (2001). Factorial k-means analysis for two-way data. *Computational Statistics and Data Analysis*, **37**, 49-64.

Wedel, M., & Kamakura, W. A. (1998). *Market Segmentation: Conceptual and Methodological Foundations*. Boston: Kluwer Academic Publishers.

Wind, Y. (1978). Issues and advances in segmentation research. *Journal of Marketing Research*, **15**, 317-337.

Yanai, H. (1998). Generalized canonical correlation analysis with linear constraints. In C. Hayashi, N. Ohsumi, K. Yajima, Y. Tanaka, H.-H. Bock, & Y. Baba (Eds.). *Data Science, Classification, and Related Methods* (pp. 539-546). Tokyo: Springer-Verlag.

Table 1. The adjusted inertias and their percentages of the adjusted total inertia in the parenthesis obtained from MCA.

	0.0840 (66.1%)
	0.0147 (11.5%)
	0.0113 (8.9%)
	0.0077 (6.1%)
	0.0044 (3.4%)
	0.0026 (2.0%)
	0.0014 (1.1%)
	0.0010 (0.8%)
	0.0000 (0.0%)
	0.0000 (0.0%)
	0.0000 (0.0%)
	0.0000 (0.0%)
	0.0000 (0.0%)
Total	0.1271

Figure 1. The category point and cluster centroid estimates obtained from the proposed method ($d = 2$, $c = 3$, and $\alpha_1 = \alpha_2 = .5$).
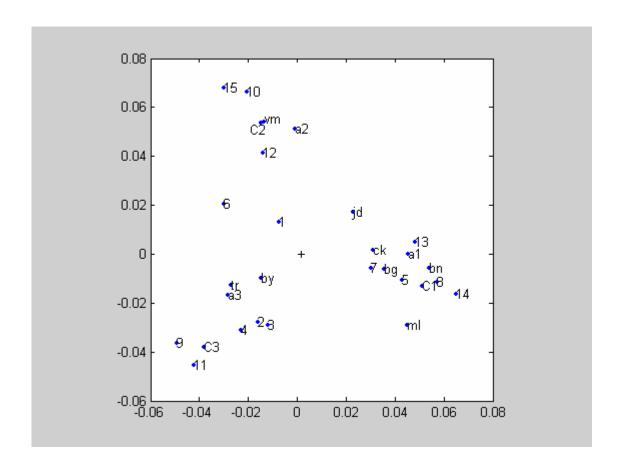
Figure 2. The category point and cluster centroid estimates obtained from the tandem approach ($d = 2$ and $c = 3$).
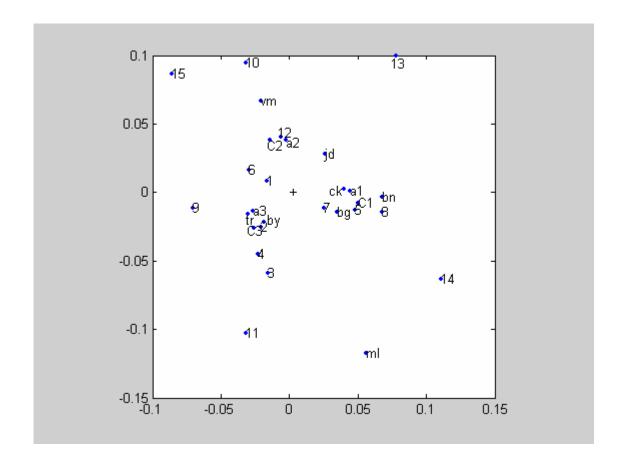
Figure 3. The category point and cluster centroid estimates obtained from Van Buuren and Heiser's method ($d = 2$ and $c = 3$).
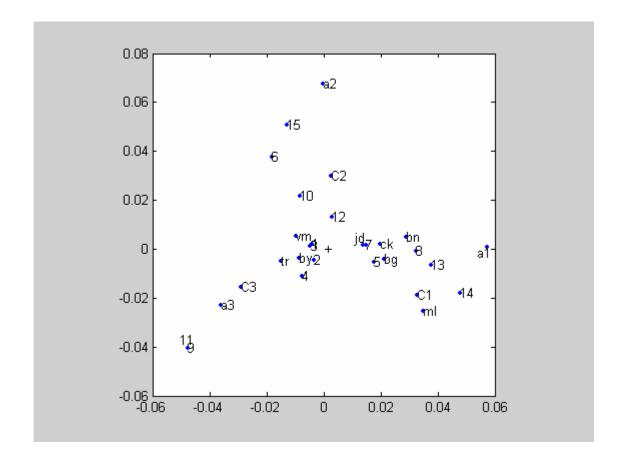
Figure 4. The category point and cluster centroid estimates obtained from the proposed method with a differential weighting ($d = 2$, $c = 3$, $\alpha_1 = .98$, and $\alpha_2 = .02$).
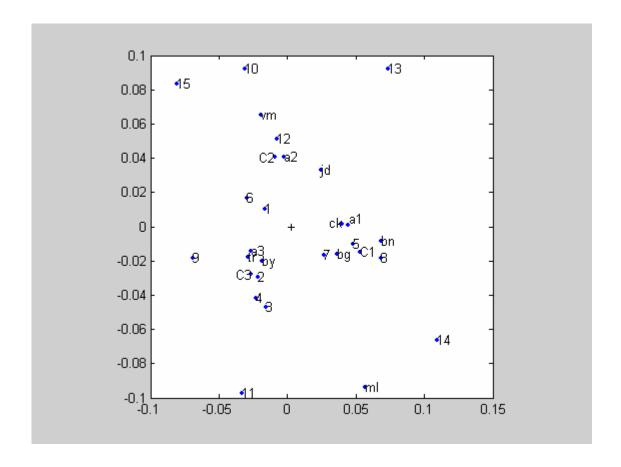
Figure 5. The category point and cluster centroid estimates obtained from the proposed method with a differential weighting ($d = 2$, $c = 3$, $\alpha_1 = .02$, and $\alpha_2 = .98$).