Fuzzy Cluster Multiple Correspondence Analysis

Heungsun Hwang

McGill University

William R. Dillon

Southern Methodist University

Yoshio Takane

McGill University

May 15, 2010

The work reported in this paper was supported by Grant 290439 and Grant 10630 from the Natural Sciences and Engineering Research Council of Canada to the first and third authors, respectively. The authors thank the two anonymous reviewers for their constructive comments. Requests for reprints should be sent to: Heungsun Hwang, Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, QC, H3A 1B1, Canada. Tel: 514-398-8021, Fax: 514-398-4896, Email: heungsun.hwang@mcgill.ca

Abstract

Multiple correspondence analysis (MCA) is a useful tool for exploring the interdependencies among multiple-choice variables. However, MCA is not geared for explicitly investigating whether or not heterogeneous subgroups of respondents exist in the population with qualitatively distinct patterns of choice behaviour. In this paper, we extend MCA to capture such cluster-level heterogeneity. Specifically, the proposed method combines MCA with fuzzy *k*-means simultaneously. Consequently, it can provide a single map of displaying variable-level and cluster-level structures so as to facilitate the interpretation of the underlying structures. The performance of the proposed method in recovering true coordinates is investigated based on a Monte Carlo study involving synthetic data. In addition, two empirical applications are presented which compare the proposed method to two extant approaches that combine MCA and cluster analysis.

Keywords: Multiple correspondence analysis, fuzzy *k*-means, alternating least squares, cluster-level heterogeneity

1. Introduction

Multiple correspondence analysis (MCA) (Benzécri, 1973; Gifi, 1990; Greenacre, 1984; Lebart, Morineau, & Warwick, 1984; Nishisato, 1980) is a simple and efficient tool for exploring interrelationships in multiple-choice data. Technically, MCA is a categorical principal components analysis that assigns numerical scores to respondents and response categories of dummy-coded multiple-choice data resulting in a low-dimensional graphical map of category interdependencies. By representing complex interrelationships in low-dimensionality, MCA has proved useful to both practitioners and academics alike (Hoffman, de Leeuw, & Arjunji, 1994). Moreover, MCA is a non-parametric approach that does not require the a priori and correct specification of the distribution underlying multiple-choice data. Thus, MCA has been a popular mapping method that is capable of describing the association structures in multiple-choice data without recourse to stringent distribution assumptions (Green, Krieger, & Carroll, 1987).

MCA results in a single set of solutions for all respondents under the assumption that the entire sample of respondents comes from a single population. Hence, MCA is not ideally suited for investigating the existence of subgroups of respondents who exhibit distinctive choice patterns. Accommodating such cluster-level heterogeneity has been considered important in various areas including developmental psychology (Moffitt, 1993), cognitive and language development (Rescorla, Mirak, & Singh, 2000), aging (Aldwin, Spiro, Levenson, & Cupertino, 2001), and consumer psychology and decision making (Bagozzi, 1982; Kamakura, Kim, & Lee, 1996).

In practice, the most common approach to recognizing cluster-level heterogeneity in MCA is the so-called *tandem analysis* (e.g., Arimond & Elfessi, 2001; Green &

Krieger, 1998; Green, Krieger, & Carroll, 1987; Green, Schaffer, & Patterson, 1988; Lebart, 1994). This is a two-step sequential approach: In the first step a low-dimensional solution of multiple-choice data is obtained via MCA; in the second step some cluster analysis technique is used to identify a set of relatively homogenous clusters on the basis of the MCA solution. In spite of the ease in which the tandem analysis can be implemented, this approach suffers from a serious drawback in that it does not guarantee that the low-dimensional data solution obtained in step one is optimal for subsequently identifying clusters of respondents. The reason for this is that each step of the tandem analysis involves a different optimization criterion (i.e., one criterion for data reduction and another for cluster analysis) and that these criteria are addressed separately; in other words, the data reduction effected by MCA is conducted with no reference to the clustering technique used to group respondents (Arabie & Hubert, 1994; Chang, 1983; DeSarbo, Jedidi, Cool, & Schendel, 1990; De Soete & Carroll, 1994).

As an alternative to the tandem approach, the combined use of MCA and cluster analysis in a single framework has been proposed (Hwang, Dillon, & Takane, 2006; Mucha, 2002; van Buuren & Heiser, 1989). In essence, this is equivalent to obtaining a low-dimensional representation of multiple-choice data and classifying respondents into a set of clusters simultaneously. The simultaneous approach yields a joint-space map in which both variable-level and cluster-level structures inherent to multiple-choice data are integrated into a single display. In particular, the method proposed by Hwang *et al*. (2006) is shown to be superior to other related methods (e.g., Mucha, 2002; van Buuren & Heiser, 1989) in term of data-analytic flexibility and interpretability.

This simultaneous approach has thus far adopted a non-overlapping clustering method such as *k*-means for the classification of respondents. The non-overlapping clustering method is based on a hard classification -- respondents are assumed to belong to one and only one cluster with rigidly defined boundaries. However, this hard classification appears to be too restricted because it is often difficult to identify a clear boundary between clusters in real-world problems (Arabie, 1977; McBratney & Moore, 1985; Wedel & Kamakura, 1998). Instead, a partial classification can potentially provide more insights with respondents assigned to multiple clusters with differential degrees of cluster membership.

Fuzzy clustering is an overlapping clustering method that is based on partial classification (Bezdek, 1974a; Dunn, 1974; Bezdek, Coray, Gunderson, & Watson, 1981; Hathaway & Bezdek, 1993; Wedel & Steenkamp, 1989) and therefore offers several advantages. First, it is likely to produce better fits to a data set than non-overlapping clustering since it is less restrictive -- the binary assignment restriction (i.e., 1 = member and 0 = non-member) is relaxed. Second, compared to non-overlapping clustering methods, fuzzy clustering is computationally more efficient because dramatic changes in the value of cluster membership (e.g., 1 (member) → 0 (non-member)) are less likely to occur in estimation procedures (McBratney & Moore, 1985). Third, fuzzy clustering has been shown to be less afflicted by local optima problems (Heiser & Groenen, 1997); in fact, the superiority of fuzzy clustering methods to a non-overlapping clustering method has been empirically demonstrated (Hruschka, 1986). Finally, the partial memberships for any given set of respondents derived from fuzzy clustering open up additional insights into the phenomenon under study by identifying the second best cluster that is almost as good as the best cluster, which non-overlapping clustering methods cannot speak to (Everitt, Landau & Leese, 2001).

In this paper, MCA is extended to accommodate fuzzy clustering. Specifically, MCA is combined with fuzzy *k*-means (Bezdek, 1974a; Dunn, 1974) in a unified framework. The proposed method provides a joint-space map that displays the relationships in multiple-choice data while simultaneously accommodating cluster-level heterogeneity. As we discuss, this is accomplished by developing a single optimization criterion for MCA and fuzzy *k*-means without imposing distributional assumptions on the data.

The paper is organized as follows. Section 2 begins with a brief review of the technical underpinnings of MCA and fuzzy *k*-means so as to facilitate the derivation of the proposed method. Then, the proposed method is discussed in detail. In Section 3, a Monte Carlo simulation study is carried out to evaluate the performance of the proposed method and the tandem approach in recovering known structures. In Section 4, two applications are provided to illustrate the empirical usefulness of the proposed method as compared to extant approaches. The final section is devoted to discussing the implications of the proposed method as well as directions for future research.

2. Method

2.1. Multiple Correspondence Analysis

MCA is a data-analytic tool for exploring the interdependencies among a set of multiple-choice variables. MCA aims to construct a set of weighted sums of dummy-coded multiple-choice variables in such a way that they maximize the association or homogeneity among the variables, or equivalently they minimize the departure from homogeneity among them (Gifi, 1990; Hoffman, de Leeuw & Arjunji, 1994).

MCA can be formulated as follows: Let $\mathbf{X}_j$ denote an N by p_j matrix of the j-th dummy-coded multiple-choice variable, where N is the number of respondents, and p_j is the number of response categories in the variable ($j = 1, \cdots, J$). Let $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N]'$ denote an N by d ($\leq \sum_{j=1}^{J} p_j - J$) matrix of a d-dimensional representation (also called object scores) of J variables, where $\mathbf{y}_i$ is a d by 1 vector of object scores for the i-th respondent ($i = 1, \cdots, N$). Let $\mathbf{W}_j$ denote a p_j by d matrix of weights, also called category quantifications. The aim of MCA can be achieved by minimizing the following optimization criterion:

$$f_1 = \sum_{j=1}^{J} \mathrm{SS}(\mathbf{Y} - \mathbf{X}_j \mathbf{W}_j), \tag{1}$$

with respect to $\mathbf{Y}$ and $\mathbf{W}_j$, subject to $\mathbf{Y}'\mathbf{Y} = \mathbf{I}$, where SS(**M**) is shorthand for the sum of squares of **M**, i.e., SS(**M**) = trace($\mathbf{M}'\mathbf{M}$). In the MCA literature, this criterion is often called the homogeneity loss function (e.g., Gifi, 1990). The category weights $\mathbf{W}_j$ obtained based on (1) subject to the identification constraint are called the principal coordinates (Greenacre, 1984, p. 90).

Minimization of (1) subject to the identification constraint can be obtained by either an analytic or iterative procedure. The former procedure comes down to calculating the eigenvalue decomposition of $\sum_{j=1}^{J} \mathbf{X}_j (\mathbf{X}_j'\mathbf{X}_j)^{-1} \mathbf{X}_j'$ (e.g., Hwang & Takane, 2002; Yanai, 1998). On the other hand, the latter procedure is based on an alternating least square (ALS) algorithm (de Leeuw, Young, & Takane, 1976), in which **Y** and $\mathbf{W}_j$ are updated alternately such that the update of one matrix minimizes (1) in the least squares sense, while the other is fixed (see Gifi, 1990, p. 107).

2.2. Fuzzy *k*-means

Fuzzy *k*-means is based on the fuzzy-set theory (Zadeh, 1965) which permits respondents to belong totally, partially, or not at all to a cluster due to the vagueness of cluster boundaries. More specifically, two assumptions underlie fuzzy *k*-means: (1) a respondent can be assigned to more than one cluster where his degree of membership in a cluster lies between 0 and 1, and (2) the sum of the memberships of a respondent across all clusters must be equal to one (Bezdek, 1981; Wedel & Kamakura, 1998).

Let u_{ki} denote a membership value for respondent *i* in the *k*-th cluster (k = 1, ···, *K*), which satisfies the above two assumptions, i.e., (1) $0 \le u_{ki} \le 1$ and (2) $\sum_{k=1}^{K} u_{ki} = 1$. Let $\mathbf{r}_k$ denote a *d* by *1* vector of the mean values or centroids of the *k*-th cluster in *d* dimensions. Let *m* indicate the prescribed fuzzy weight scalar, often called the 'fuzzifier' (Bezdek, 1974a), which controls for the degree of fuzziness of the solution. Suppose for now that **Y** is a known data matrix. Then the fuzzy *k*-means algorithm is equivalent to minimizing the following criterion:

$$f_2 = \sum_{k=1}^{K} \sum_{i=1}^{N} u_{ki}^{m} \mathrm{SS}(\mathbf{y}_i - \mathbf{r}_k), \qquad (2)$$

with respect to $\mathbf{r}_k$ and u_{ki}, subject to the fuzzy membership constraints (Bezdek, 1974a; Dunn, 1974). Note that in (2), **Y** can be any data matrix which is used for the classification of respondents, although the same matrix as in (1) is used to facilitate the derivation of the proposed method in the next section. When m = 1 and u_{ki} = 1 or 0 (binary membership), the fuzzy *k*-means criterion reduces to the *k*-means criterion (e.g., Rocci & Vichi, 2005). Minimization of (2) for $m > 1$ can be accomplished using a two-step optimization procedure of updating u_{ki} and $\mathbf{r}_k$ repeatedly.

In fuzzy *k*-means, the value of *m* should be selected in advance. However, values of *m* close to 1 will result in a near hard partition with all memberships close to 0 or 1, while excessively large values will lead to disproportionate overlap with all memberships close to *1/K* (Wedel & Steenkamp, 1989). Consequently, neither of these values of *m* is recommended (Arabie, Carroll, DeSarbo, & Wind, 1981). Although there have been some heuristic procedures to determine the value of *m* (e.g., McBratney & Moore, 1985; Okeke & Karnieli, 2006; Wedel & Steenkamp, 1989), there seems to exist no formal way of choosing *m.* In practice, m = 2 is the most popular choice in fuzzy clustering algorithms (Bezdek, 1981; Gordon, 1999; Hruschka, 1986; Wedel & Steenkamp, 1991).

A number of so-called cluster validity measures (Bezdek, 1981; Roubens, 1982) are used to determine *K*. Based on the analysis of synthetic data, Roubens (1982) concluded that the Fuzziness Performance Index (FPI) and the Normalized Classification Entropy (NCE) are the most useful cluster validity measures for fuzzy clustering. The FPI and NCE are given by:

$$\mathrm{FPI} = 1 - (K\times\mathrm{PC} - 1)/(K - 1),$$

where PC is the Partition Coefficient (Bezdek, 1974b), defined as $\mathrm{PC} = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} u_{ki}^2$, and

$$\mathrm{NCE} = \mathrm{PE}/\log K,$$

where PE is the Partition Entropy (Bezdek , 1974b), defined as $\mathrm{PE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} u_{ki}\log u_{ki}$.

These measures indicate the separation status of clusters, i.e., how well the derived clusters are separated from each other. The smaller the values of FPI and NCE are, the more distinctly separated the clusters are from each other. Thus, the appropriate number of clusters should result in small values of FPI and NCE. Note that NCE and FPI can be used when $K > 1$. Moreover, the changes in the values of (2) or other related measures (e.g., the average adjusted R^2 across clusters) may be graphically examined against the different number of clusters in order to determine the number of clusters (Wedel & Steenkamp, 1989). The number of clusters may be chosen as an elbow point in the trajectory of the value of (2) over clusters, beyond which no substantial changes in the value occur.

2.3. Fuzzy Cluster Multiple Correspondence Analysis

Our objective is to combine MCA and fuzzy *k*-means into a unified framework, resulting in a single graphical display of variable categories and cluster centroids. In other words, the objective is to find a low-dimensional solution of multiple-choice data while at the same time identifying fuzzy clusters of respondents that are homogeneous in terms of the low-dimensional solution. This problem is equivalent to minimizing the following optimization criterion:

$$\begin{aligned} f &= f_1 + f_2 \\ &= \sum_{j=1}^{J} \mathrm{SS}(\mathbf{Y} - \mathbf{X}_j \mathbf{W}_j) + \sum_{k=1}^{K} \sum_{i=1}^{N} u_{ki}^{m} \mathrm{SS}(\mathbf{y}_i - \mathbf{r}_k), \end{aligned} \tag{3}$$

with respect to $\mathbf{Y}$, $\mathbf{W}_j$, $\mathbf{r}_k$, and u_{ki}, subject to the identification and membership constraints in MCA and fuzzy *k*-means. This criterion is the sum of the two optimization criteria for MCA and the fuzzy *k*-means algorithm. By minimizing (3), therefore, the low-dimensional data representation is obtained in such a way that it recognizes the fuzzy cluster structure that may be inherent in multiple-choice data. We shall call this new approach *Fuzzy Cluster Multiple Correspondence Analysis* (FCMCA).

An alternating least squares (ALS) algorithm is developed to minimize (3). In the algorithm, the unknown coordinates and fuzzy memberships ($\mathbf{Y}, \mathbf{W}_j$, $\mathbf{r}_k$, and u_{ki}) are updated alternately until convergence of (3) is reached. A detailed description of the ALS algorithm is provided in the Appendix.

The proposed algorithm monotonically decreases the value of the objective function 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. To safeguard against local minima, we operationalize two alternative methods of obtaining starting values. With rational starts, MCA can be first applied to the data and the resultant object scores are used as the rational starts for $\mathbf{Y}$. Then the fuzzy k-means algorithm is applied to $\mathbf{Y}$, and the resultant memberships are used as initial values for u_{ki}. In other words, we use the solutions of the tandem approach as rational starts. We repeat the ALS algorithm with many random initial starts, compare the obtained criterion values after convergence, and choose the solution that yields the smallest criterion value.

In the proposed method, the number of dimensions (d) needs to be decided in advance. As in the extant (hard-cluster) simultaneous approach (e.g., Hwang *et al*., 2006), d may be selected by applying MCA to the data. In MCA, d is usually chosen to be less than or equal to 3 (typically 2) to facilitate the visualization and interpretation of the solutions (Rovan, 1998). Subsequently, in the proposed method, the number of clusters (K) may be determined by examining how the values of (3) change with different numbers of clusters. As in fuzzy clustering, an elbow point in the trajectory of the value of (3) over clusters may be chosen as the number of clusters. In addition, FCMCA may utilize cluster validity measures such as FPI and NCE in order to investigate whether a chosen set of clusters are sufficiently separated from each other compared to other sets of clusters. It is also recommended that the number of clusters be greater than the number of dimensions (van Burren & Heiser, 1989; Vichi & Kiers, 2001). As discussed earlier, we set $m = 2$ as the default value of the fuzzy weight.

3. Synthetic Data Analysis

In this section we report on a Monte Carlo simulation study designed to evaluate the performance of the proposed method and the tandem approach that applies MCA and fuzzy k-means sequentially.[1] The simulation study focuses on how well FCMCA and the tandem approach recover the original variable category points and the underlying clustering structure of the data for varying sample sizes.

In the simulations which follow we set $d = 2$ and $K = 4$ and considered the case of three multiple-choice variables each having four categories (i.e., $J = 3$ and $p_j = 4$). The coordinate values of the variable categories ($\mathbf{W}_j$) and the centroids ($\mathbf{r}_k$) were chosen to be located on a 4 x 4 equally spaced grid on [-1, 1]. Figure 1 shows the two-dimensional plot of the variable category points and centroids. In Figure 1, the centroids of the four

[1] We chose not to compare the performance of FCMCA to that of the extant simultaneous approach because the two approaches are built on theoretically different schemes of classification (fuzzy vs. hard), and consequently it is difficult to generate synthetic data (i.e., membership values) that can be applicable to both approaches equally. FCMCA will be compared to the extant simultaneous approach proposed by Hwang et al. (2006) in the empirical application section to follow.

clusters are labelled ‘C1’, ‘C2’, ‘C3’, and ‘C4’ and each variable category is represented by a two-digit number in which the first digit indicates item number while the second denotes category number. The symbol ‘+’ is used to indicate the origin of the configuration. Finally, three sample sizes were considered for the study: N = 50, 100, and 200. The number of respondents per cluster for each sample size also varied as follows: When N = 50, cluster 1 = 5, cluster 2 = 10, cluster 3 = 15, cluster 4 = 20; when N = 100, cluster 1 = 10, cluster 2 = 20, cluster 3 = 30, cluster 4 = 40; and when N = 200, cluster 1 = 20, cluster 2 = 40, cluster 3 = 60, cluster 4 = 80.

Insert Figure 1 about here

To generate a synthetic data set, we first set **Y** = **UR** + **E**, where **U** is an N by 4 matrix of fuzzy memberships, **R** is a 4 by 2 matrix consisting of the prescribed centroids, and **E** is an N by 2 matrix of random errors. Specifically, the fuzzy membership matrix **U** was generated as follows: hard classifications of respondents to only one of the four clusters were made based on the predetermined cluster sizes, and then random errors uniformly drawn from [0, .5] were added to the hard classifications (the resultant membership values of each respondent were re-scaled to satisfy the fuzzy membership constraint $\sum_{k=1}^{K} u_{ki} = 1$). Moreover, the random errors in **E** were drawn from a normal distribution with mean 0 and variance of about 40% of the variance of **UR**, although there is no justifiable rule available for deciding the range of the variance. Then, a single synthetic data set was generated from $\mathbf{X}_j = \mathbf{Y}\mathbf{W}_j'(\mathbf{W}_j'\mathbf{W}_j)^{-1}$. Five hundred synthetic data sets were drawn for each sample size.

The synthetic data were analyzed by the proposed method and the tandem approach. The proposed FCMCA algorithm was found to converge in all cases. Table 1 provides the mean, standard errors, and biases of the coordinates of variable categories ($\mathbf{W}_j$) and centroids ($\mathbf{r}_k$) obtained from the two approaches across different sample sizes. We first applied Procrustes rotation to the estimated coordinates toward the true configuration, and then evaluated their biases. Table 1 also provides the mean congruence coefficient (Tucker, 1951) between true membership values and their estimates obtained from the two approaches for each sample size as an overall measure of membership recovery. The congruence coefficient is calculated as follows: Let $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$ denote the vectors of the true membership values and estimates obtained from a single synthetic data set, respectively. Then, the congruence coefficient is $(\boldsymbol{\theta}'\boldsymbol{\rho})/(\sqrt{\boldsymbol{\theta}'\boldsymbol{\theta}}\sqrt{\boldsymbol{\rho}'\boldsymbol{\rho}})$. The congruence coefficient indicates the degree of similarity between true and estimated values.

Insert Table 1 about here

As shown in Table 1, the biases of the estimates of both variable categories and cluster centroids obtained from FCMCA appear to be quite small across all sample sizes. Also, they tend to gradually decrease with increased sample size. On the other hand, in all sample sizes, the biases of the estimates of variable categories obtained from the tandem approach are much larger that those from FCMCA, while the biases of centroid estimates appear similar to those from FCMCA. Under all sample sizes, the mean congruence coefficients for the membership values are quite similar across the two approaches. Moreover, all of them are much larger than .90, which is a conventional rule of thumb criterion as an acceptable degree of congruence (Mulaik, 1972). In both approaches, the mean congruence coefficients tend to gradually increase as sample size increases.

The present simulation study was designed to generate synthetic data based on two-dimensional data (**Y**) under clear and simple structures between variable categories and cluster centroids. In such a situation, the tandem approach is expected to recover true coordinates well because two-dimensional data possess all necessary information on the relationships among variable categories and cluster centroids. Nonetheless, in this simulation study, the proposed method was found to recover the coordinates of variable categories better than the tandem approach while capturing fuzzy cluster centroids and memberships equally well. Thus, the proposed approach is shown to perform better than the tandem approach in terms of overall coordinate recovery.

4. Empirical Data Analyses

In this section, we present two empirical applications that focus on the usefulness of the proposed method as compared to that of the tandem approach and the extant simultaneous approach by Hwang *et al*. (2006).

4.1. The Korean Underwear Data

The first example comes from a large survey conducted by a Korean underwear manufacturer (Yang, 1997). 664 consumers were asked to answer two multiple-choice questions: 1) “Which of the following brands is your first choice (i.e., favourite brand)?” and 2) “In deciding which brand of underwear to buy which of the following attributes is most important?” Eight brands of underwear and 15 attributes were included in the study.

Table 2 provides brand names and attributes. As reported by Yang (1997) these brands have markedly different positioning within the Korean market; for example, three of the brands (BYC, TRY, and VICMAN) are popular, economical domestic brands; James Dean is positioned as a high quality domestic brand; Bodyguard is perceived as a more fashionable domestic brand; and Michiko-London, Benetton, and Calvin Klein are multinational brands.

Insert Table 2 about here

MCA was first applied 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). 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). Table 3 shows the adjusted inertias and their percentages of the adjusted total inertia (i.e., the sum of the adjusted inertias).

Insert Table 3 about here

We chose $d = 2$ because the adjusted inertias tended to decrease gradually after the first two or three and, at the same time, interpretation of a two-dimensional configuration of parameter estimates is generally easier (Rovan, 1994). The first two adjusted inertias explained about 57% of the adjusted total inertia. Next, with d fixed, we investigated changes in the value of (3) by varying numbers of clusters. Figure 2 provides the plot of the optimization criterion values against different numbers of clusters. It also displays the values of FPI and NCE as the number of clusters is increased. The values of (3) were re-scaled by dividing each of them by the maximum of the optimization criterion values. This re-scaling was performed to set the range in the criterion values within the same unit as the FPI and NCE measures (i.e., [0 1]), thereby allowing all measures to be displayed in a single plot. In the figure, the (re-scaled) values of (3) appear to decrease slowly beyond three clusters, suggesting that no substantial changes in the criterion values are obtained by having more than three clusters. Moreover, the values of FPI and NCE for $K = 3$ seem to indicate a sufficient level of separation among clusters compared to other solutions (i.e., different numbers of clusters). Thus, $K = 3$ was adopted for our analysis.

Figure 3 displays the two-dimensional map of the category points and the centroids of three clusters estimated from the proposed method. To render the figure more concise, the centroids of clusters 1, 2, and 3 are labelled ‘C1’, ‘C2’, and ‘C3’, respectively. The eight brands are labelled: ‘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 symbol ‘+’ indicates the origin of the display.

Insert Figure 3 about here

In Figure 3, the top portion seems to be related to the first cluster, at which its centroid, C1, is located. C1 is closely located with such attributes as ‘fashionable design

(5)’, ‘trendy colour (7)’, ‘good design (8)’, and ‘youth appeal (14)’. This indicates that the respondents in the first cluster appear to prefer the design or style of underwear and appears to be consistent with C1 being positioned close to such brands as ‘Benetton (bn)’ and ‘Bodyguard (bg)’. As mentioned earlier, Bodyguard is positioned as a fashionable, stylistic underwear brand in Korea (Yang, 1997).

On the other hand, the middle right-hand side seems to be associated with the second cluster because it embraces the centroid of the second cluster, C2. This centroid is positioned close to such attributes as ‘superior fabrics (3)’, ‘excellent fit (12)’, and ‘design quality (13)’. It thus appears that the respondents in the second cluster were more inclined toward high quality varieties of underwear. This cluster appears to be close to such brands as ‘Calvin Klein (ck)’, ‘Michiko-London (ml)’, and ‘James Dean (jd)’. These brands are recognized as being of high-quality in Korea (Yang, 1997).

Finally, the middle left-hand portion of the display seems to represent the characteristics related to the third cluster, whose centroid is represented by C3. This centroid is positioned close to attributes: ‘comfortable (1)’, ‘smooth (2)’, ‘reasonable price (4)’, ‘favourable advertisements (6)’, ‘various colours (9)’, ‘elastic (10)’, ‘store is near (11)’, and ‘various sizes (15)’, suggesting that the respondents belonging to the third cluster were more likely to emphasize practical aspects of underwear in their purchase. This cluster is also closely linked to such domestic brands as ‘BYC (by)’, ‘TRY (tr)’, and ‘VICMAN (vm)’. When all respondents are assigned to the cluster associated with the highest membership value, the sizes of clusters 1, 2, and 3 arrive at 199 (30%), 156 (23.5%) and 309 (46.5%), respectively.

Using the same number of dimensions and clusters, we also applied the tandem approach and the extant simultaneous approach by Hwang *et al*. (2006) to these data. For the tandem approach, we repeated the fuzzy *k*-means algorithm with 100 different random values for initial cluster membership. For the extant simultaneous approach, we repeated its algorithm with 100 different random values for initial cluster membership. The results of the tandem approach and the extant simultaneous approach are displayed in Figures 4 and 5, respectively.

Insert Figures 4 and 5 about here

In Figure 4, the top left-hand portion seems to be related to the first cluster, the lower right-hand side to the second cluster, and the lower left-hand side to the third cluster. However, it is much more difficult to characterize the clusters in the tandem approach because the estimated attribute and brand category points are located in a quite indiscriminate way, compared to the proposed method. When all respondents are assigned to the cluster associated with the highest membership value, the sizes of clusters 1, 2, and 3 (30%, 22%, and 48%, respectively) appear similar to those obtained with the proposed method.

At first, the three clusters in Figure 5 appear to have a similar interpretation to the FCMCA solution. However, compared to the FCMCA solution note the following: 1) the first cluster is much to be closer to ‘James Dean (jd)’ and ‘Calvin Klein (ck)’; 2) the second cluster in Figure 5 is located close to only one brand, ‘Michiko-London (ml)’; and 3) the C1 and C2 cluster sizes are very different (10% vs. 30% and 43% vs. 24%, respectively). Although, admittedly, it is difficult to tell which method provides more accurate solutions because true coordinate values are unknown, the result of the extant simultaneous approach seems to be inconsistent with what is known about this market which categorizes Calvin Klein, Michiko-London and James Dean as high-quality brands (Yang, 1997).

4.2. The French Worker Survey Data

The second example is taken from the French Worker Survey (Adam, Bon, Capdevielle, & Mouriaux, 1970). This survey, conducted in 1969 on 1049 French workers, focused on political and social opinions and preferences. Among the original 70 survey questions, we use, for the purposes of this application, the three questions reported in Le Roux and Rouanet (1998): [Q1] "In the professional elections in your firm, would you rather vote for a list supported by?", [Q2] "On the last presidential election (1969), can you tell me the candidate for whom you have voted?" , and [Q3] "Which political party do you feel closest to, as a rule?" Table 4 provides the response categories for each question. Among the trade unions listed in the first question, CGT had a strong link with the Communist Party; CFDT and FO were loosely classified as non-communist left unions; and Autonomous was inclined toward right wing. The four major candidates of the 1969 presidential election may be placed from politically left to right as follows: Communist (Duclos), Socialist (Defferre), Center (Poher), and Gaullist (Pompidou) (see Le Roux & Rouanet, 1998).

Insert Table 3 about here

As in the previous application, we again use MCA to determine the number of dimensions. Table 3 provides the adjusted inertias and their percentages of the adjusted total inertia obtained from the MCA. Although the third dimension seemed to explain a sizeable portion of the adjusted total inertia, we decided to choose $d = 2$ because the first two adjusted inertias accounted for a majority of the adjusted total inertia (76%) and to aid the interpretation. Next, with d fixed, we investigated changes in the value of (3) by varying numbers of clusters. Figure 6 provides the plot of the re-scaled optimization criterion values and the values of FPI and NCE against different numbers of clusters. In the figure, the re-scaled values of (3) appear to decrease gradually beyond three clusters. Moreover, the values of both FPI and NCE were minimized at $K = 3$. Thus, we set $K = 3$ in all of the analyses that follow.

Insert Figure 6 about here

Figure 7 exhibits the two-dimensional map of the category points and the centroids of three clusters obtained from the proposed method. In this map, the estimated response categories for each item are represented by a two-digit label, in which the first digit indicates item number and the second corresponds to category number (e.g., ‘11’ = category 1 in item 1, ‘12’ = category 2 in item 1, ‘21’ = category 1 in item 2, and so forth). As in the previous application, the three centroids are labelled ‘C1’, ‘C2’, and ‘C3’.

Insert Figure 7 about here

In Figure 7, the centroid of the first cluster, C1, is located on the middle left-hand portion of the map relatively close to such categories as ‘CGT (11)’, ‘Duclos (21)’, and ‘Communist Party (31)’; this suggests that respondents belonging to this cluster manifest a politically communist position. On the other hand, the centroid of the second cluster, C2, is located on the bottom side of the figure close to the categories: ‘No response to question 1 (18)’, ‘No response to question 2 (28)’, ’No response to question 3 (38)’, ‘Abstention (16)’, and so on. This suggests that these respondents are either indifferent or, for whatever reasons, circumspect. Finally, the centroid of the third cluster, C3, is located in the top right corner of the figure close to most of the remaining categories. Thus, this cluster is likely to represent a politically non-communist group of respondents among those who tended to express their opinions in the survey. When respondents are assigned to the cluster associated with the highest membership value, the sizes of clusters 1, 2, and 3 are 243 (23%), 344 (33%) and 462 (44%), respectively.

Given the same numbers of dimensions and clusters, we also applied the tandem approach and the extant simultaneous approach to the same data. Again, for the tandem approach, we repeated the fuzzy *k*-means algorithm with 100 different random values for initial cluster membership. For the extant simultaneous approach, we repeated its algorithm with 100 different random values for initial cluster membership. The results of the tandem approach and the hard cluster method are displayed in Figures 8 and 9, respectively.

Insert Figures 8 and 9 about here

As shown in Figure 8, the two-dimensional map obtained from the tandem approach appears similar to that from FCMCA, and hence leads to essentially the same interpretations. This may be because the retained dimensions from MCA account for a majority of the variance (76%) in the data, so that they contain sufficient information on the underlying data structure. Nonetheless, many category points estimated by FCMCA are more compactly grouped, in particular, those related to C3, thereby providing a cleaner interpretation. This seems to be consistent with the findings of the simulation study in the previous section; that is, the category estimates from the tandem approach were shown to have larger biases than those obtained with FCMCA even in cases of two dimensional data with simple structure. When all respondents are assigned to the cluster associated with the highest membership value, the sizes of clusters 1, 2, and 3 arrive at 260 (25%), 413 (39%) and 376 (36%), respectively.

On the other hand, the middle left-hand portion of Figure 9, obtained from the extant simultaneous approach, appears related to the first cluster, representing 25% of the sample, the bottom portion to the second cluster, representing 41% of the sample, and the upper portion to the third cluster, representing 34% of the sample. Though similar in spirit to the solution produced by the proposed method, it is more difficult to characterize each cluster because the estimated category points are less compactly clustered close to a centroid; for example, 'CGT (11)' has drifted away from C1 toward the origin of the display which is inconsistent with the very strong link of CGT to the Communist Party (Le Roux & Rouanet, 1998).

5. Concluding Remarks

A new method for simultaneously combining MCA with fuzzy *k*-means was proposed to take into account cluster-level heterogeneity in multiple-choice data. Technically, the proposed method may be viewed as a generalization of the extant simultaneous approach (e.g., Hwang et al., 2006) which captures this heterogeneity based on a non-overlapping clustering method. The proposed method relaxes the assumption of dichotomous memberships in the extant simultaneous approach.

The performance of the proposed method was evaluated and compared to the tandem approach with synthetic data with different sample sizes. The analyses of synthetic data showed the effectiveness of the proposed method in recovering known structures compared to the tandem approach. In addition, the usefulness of the proposed method was demonstrated in two empirical applications. The proposed method was shown to be useful and generally superior to the tandem and extant simultaneous approaches in studying qualitatively different patterns of choice behaviour in heterogeneous subgroups of respondents in both applications.

The proposed method may be further refined and extended so as to enhance its data-analytic capability and the scope of its applicability. For example, although it is not a problem unique to this method, more formal rules are needed for determining the fuzzy weight. Similarly, a fruitful research area would be to develop a more confirmatory way of selecting the number of clusters as some arbitrary decisions still need to be made with current cluster validity measures. For instance, the graphical judgment on the elbow point of the optimization criterion values may often be subjective. Non-parametric procedures such as bootstrap scree tests (Hong, Mitchell, & Harshman, 2006) may be adopted for addressing this issue. Moreover, the proposed method may be viewed as a post-hoc classification approach in that it identifies clusters of respondents through the analysis of the data (Wind, 1978; Wedel & Kamakura, 1998). In many cases, cluster-level heterogeneity can also be addressed by classifying respondents into clusters *a priori* on the basis of demographic variables or other grouping variables. This *a priori* classification approach may be incorporated into the proposed method for more sophisticated analyses. This extension may involve combining the proposed method with multilevel MCA (Michailidis & de Leeuw, 2000).

Appendix: The proposed alternating least squares algorithm

Let $\mathbf{U}_k$ denote an N by N diagonal matrix consisting of u_{ki} as an element. Let $\mathbf{e}$ denote an N by 1 vector of ones. Then, criterion (3) can be re-written as:

$$\begin{aligned} f &= \sum_{j=1}^{J} \mathrm{SS}(\mathbf{Y} - \mathbf{X}_j \mathbf{W}_j) + \sum_{k=1}^{K} \mathrm{SS}(\mathbf{Y} - \mathbf{er'}_k)_{\mathbf{U}_k^m} \\ &= \sum_{j=1}^{J} \mathrm{SS}(\mathbf{Y} - \mathbf{X}_j \mathbf{W}_j) + \sum_{k=1}^{K} \mathrm{SS}(\mathbf{\Delta}_k \mathbf{Y} - \mathbf{\Delta}_k \mathbf{er'}_k), \end{aligned} \tag{A1}$$

where $\mathrm{SS}(\mathbf{M})_{\mathbf{H}} = \mathrm{trace}(\mathbf{M'HM})$ and $\mathbf{\Delta}_k = \left(\mathbf{U}_k^m\right)^{1/2}$. The proposed ALS algorithm repeats the following three main steps until convergence:

Step 1. $\mathbf{W}_j$ and $\mathbf{r}_k$ are updated for fixed $\mathbf{Y}$ and u_{ki}. Since $\mathbf{W}_j$ is only involved in the first term and $\mathbf{r}_k$ is only in the second term of (A1), their updates are simply given by:

$$\hat{\mathbf{W}}_j = (\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'\mathbf{Y}, \tag{A2}$$

and

$$\hat{\mathbf{r}}'_k = (\mathbf{e}'\mathbf{U}_k^m\mathbf{e})^{-1}\mathbf{e}'\mathbf{U}_k^m\mathbf{Y}. \tag{A3}$$

Step 2. $\mathbf{Y}$ is updated for fixed $\mathbf{W}_j$, $\mathbf{r}_k$, and u_{ki}. Let $\mathbf{\Omega}_j = \mathbf{X}_j(\mathbf{X'}_j \mathbf{X}_j)^{-1}\mathbf{X'}_j$, and $\mathbf{\Psi}_k = \mathbf{\Delta}_k \mathbf{e}(\mathbf{e'U}_k^m\mathbf{e})^{-1}\mathbf{e'}\mathbf{\Delta}_k$. Note that $\mathbf{\Omega}_j$ and $\mathbf{\Psi}_k$ are idempotent and symmetric. Putting (A2) and (A3) into (A1), then, minimizing (A1) with respect to $\mathbf{Y}$, subject to $\mathbf{Y'Y} = \mathbf{I}$, is equivalent to minimizing:

$$
\begin{aligned}
f^* &= \sum_j \mathrm{SS}(\mathbf{Y} - \mathbf{\Omega}_j \mathbf{Y}) + \sum_k \mathrm{SS}(\mathbf{\Delta}_k \mathbf{Y} - \mathbf{\Psi}_k \mathbf{\Delta}_k \mathbf{Y}) \\
&= \sum_j \mathrm{trace}(\mathbf{Y'Y} - \mathbf{Y'}\mathbf{\Omega}_j \mathbf{Y}) + \sum_k \mathrm{trace}(\mathbf{Y'U}_k^m \mathbf{Y} - \mathbf{Y'}\mathbf{\Delta}_k \mathbf{\Psi}_k \mathbf{\Delta}_k \mathbf{Y}) \\
&= \sum_j \mathrm{trace}(\mathbf{Y'Y}) - \sum_j \mathrm{trace}(\mathbf{Y'}\mathbf{\Omega}_j \mathbf{Y}) + \sum_k \mathrm{trace}(\mathbf{Y'U}_k^m \mathbf{Y}) - \sum_k \mathrm{trace}(\mathbf{Y'}\mathbf{\Delta}_k \mathbf{\Psi}_k \mathbf{\Delta}_k \mathbf{Y}) \\
&= \sum_j \mathrm{trace}(\mathbf{Y'Y}) - \left\{ \mathrm{trace}(\mathbf{Y'}\left[\sum_j \mathbf{\Omega}_j\right]\mathbf{Y}) + \mathrm{trace}(\mathbf{Y'}\left[\sum_k \mathbf{\Delta}_k \mathbf{\Psi}_k \mathbf{\Delta}_k\right]\mathbf{Y}) - \mathrm{trace}(\mathbf{Y'}\left[\sum_k \mathbf{U}_k^m\right]\mathbf{Y}) \right\} \\
&= Jd - \left\{ \mathrm{trace}(\mathbf{Y'}\left[\sum_j \mathbf{\Omega}_j + \sum_k \mathbf{\Delta}_k \mathbf{\Psi}_k \mathbf{\Delta}_k - \sum_k \mathbf{U}_k^m\right]\mathbf{Y}) \right\}.
\end{aligned}
\tag{A4}
$$

Minimizing (A4) thus reduces to maximizing:

$$
\mathrm{trace}\left(\mathbf{Y'}\left[\sum_j \mathbf{\Omega}_j + \sum_k \mathbf{\Delta}_k \mathbf{\Psi}_k \mathbf{\Delta}_k - \sum_k \mathbf{U}_k^m\right]\mathbf{Y} \right) \tag{A5}
$$

with respect to **Y**. Maximizing (A5) over **Y** comes down to calculating the eigenvalue decomposition of $\left[\sum_j \mathbf{\Omega}_j + \sum_k \mathbf{\Delta}_k \mathbf{\Psi}_k \mathbf{\Delta}_k - \sum_k \mathbf{U}_k^m\right]$ (e.g.,Yanai, 1998).

<u>Step 3</u>. The membership parameter u_{ki} is updated for fixed **Y**, $\mathbf{W}_j$ and $\mathbf{r}_k$. It is equivalent to minimizing the second term of (A1) with respect to u_{ki}. We employ the standard fuzzy *k*-means algorithm (Bezdek, 1974a; Dunn, 1974): Let $d_{ki} = (\mathbf{y}_i - \mathbf{r}_k)'(\mathbf{y}_i - \mathbf{r}_k)$. Then, u_{ki} is updated by:

$$
\hat{u}_{ki} = \left[\sum_c^K \left(\frac{d_{ki}}{d_{ci}} \right)^{1/(m-1)} \right]^{-1}. \tag{A6}
$$

Formula (A6) can be derived as follows: Minimizing the second term of (A1) under the membership constraint is equivalent to minimizing:

$$
L = \sum_{k=1}^K \sum_{i=1}^N u_{ki}^m d_{ki} - \lambda(\sum_{k=1}^K u_{ki} - 1), \tag{A7}
$$

where λ is a Lagrangian multiplier. Solving $\frac{\partial L}{\partial u_{ki}} = m u_{ki}^{m-1} d_{ki} - \lambda = 0$ for u_{ki} yields:

$$\hat{u}_{ki} = \left(\frac{\lambda}{m d_{ki}} \right)^{1/(m\text{-}1)} \quad . \tag{A8}$$

Using $\frac{\partial L}{\partial \lambda} = \sum_{k=1}^{C} u_{ki} - 1 = 0$ and (A8) leads to:

$$\hat{\lambda} = \left(\left(\sum_{k=1}^{K} 1/(m d_{ki}) \right)^{1/(m\text{-}1)} \right)^{1-m} \quad . \tag{A9}$$

Then, (A6) is obtained by inserting (A9) in (A8).

References

Adam, G., Bon, F., Capdevielle, J., & Mouriaux, R. (1970). *L'ouvrier français en* 1970. Paris: FNSP.

Aldwin, C. M., Spiro, A. I., Levenson, M. R., & Cupertino, A. P. (2001). Longitudinal findings from the Normative Aging Study: III. Personality, individual health trajectories, and mortality. *Psychology and Aging*, 16, 450-465.

Arabie, P. (1997). Clustering representations of group overlap. *Journal of Mathematical Sociology*, 5, 113-128.

Arabie, P., Carroll, J. D., DeSarbo, W. S., & Wind, J. (1981). Overlapping clustering: A new method for product positioning. *Journal of Marketing Research*, 18, 310-317.

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.

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.

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

Bezdek, J. C. (1974b). Cluster validity with fuzzy set. *Journal of Cybernetics*, 3, 58-72.

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press.

Bezdek, J. C., Coray, C., Gunderson, R., & Watson, J. (1981). Detection and characteristics of cluster substructure. II. Fuzzy c-varieties and convex combinations thereof. *SIAM Journal on Applied Mathematics*, 40, 358-372.

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.

De Soete, G., & Carroll, J. D. (1994). K-means clustering in a low-dimensional Euclidean space. In E. Diday, Y. Lechevallier, M. Schader, P. Bertrand and B. Burtschy (Eds.). *New Approaches in Classification and Data Analysis* (pp. 212-219). Heidelberg: Springer.

Dunn, J. C. (1974). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3, 32-57.

Everitt, B. S., Landau, S., & Leese, M. (2001). *Cluster Analysis, 4th ed.* London: Arnold Press.

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

Gordon, A. D. (1999). *Classification*. London: Chapman & Hall/CRC.

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

Green, P. E., Krieger, A. M., & Carroll, D. J. (1987). Conjoint analysis and multidimensional scaling: A complementary approach. *Journal of Advertising Research*, 27, 21-27.

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.

Hathaway, R. J., & Bezdek, J. C. (1993). Switching regression models and fuzzy clutering. *IEEE Transactions on Fuzzy Systems*, 1, 195-204.

Heiser, W. J., & Groenen, P. J. F. (1997). Cluster differences scaling with a within-clusters loss component and a fuzzy successive approximation strategy to avoid local minima. *Psychometrika*, 62, 63-83.

Hoffman, D. L., de Leeuw, J., & Arjunji, R. V. (1994). Multiple correspondence analysis. In Bagozzi, R. P. (ed.), *Advanced Methods of Marketing Research* (pp. 260-294). Oxford: Blackwell.

Hong, S., Mitchell, S., & Harshman, R. A. (2006). Bootstrap scree tests: A Monte Carlo simulation and applications to published data. *British Journal of Mathematical and Statistical Psychology*, 59, 35-57.

Hruschka, H. (1986). Market definition and segmentation using fuzzy clustering methods. *International Journal of Research in Marketing*, 3, 117-134.

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

Hwang, H., Dillon, W. S., & Takane, Y. (2006). An extension of multiple correspondence analysis for identifying heterogeneous subgroups of respondents. *Psychometrika*, 71, 161-171.

Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science*, 16, 39-59.

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

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.

Le Roux, B., & Rouanet, H. (1998). Interpreting axes in multiple correspondence analysis: Method of the contributions of points and deviations. In Greenacre, M. J., & Blasius, J. (Eds.). *Visualization of categorical data* (pp. 197-220). Chestnut

Hill, MA: Academic Press.

McBratney, A. B., & Moore, A. W. (1985). Application of fuzzy sets to climatic classification. *Agricultural and Forest Meteorology*, 35, 165-185.

Moffitt, T. E. (1993). Adolescent-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review*, 100, 674-701.

Michailidis, G., & de Leeuw, J. (2000). Multilevel homogeneity analysis with differential weighting. *Computational Statistics and Data Analysis*, 32, 411-442.

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.

Mulaik, S. A. (1972). *The Foundations of Factor Analysis*. New York: McGraw-Hill.

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

Oh, C.-H., Honde, K., & Ichihashi, H. (2005). Quantification of multivariate categorical data considering clusters of items and individuals. *Modeling decisions for artificial intelligence; Proceedings of the second international conference, MDAI 2005*, 164-171.

Oh, C.-H., Honde, K., & Ichihashi, H. (2007). Quantification of multivariate categorical data considering typicality of item. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 11, 35-39.

Okeke, F., & Karnieli, A. (2006). Linear mixture model approach for selecting fuzzy exponent value in fuzzy c-means algorithm. *Ecological Informatics*, 1, 117-124.

Rescorla, L., Mirak, J., & Singh, L. (2000). Vocabulary growth in late talkers: Lexical development from 2;0 to 3;0. *Journal of Child Language*, 27, 293-311.

Rocci, R., & Vich, M. (2005). Three-mode component analysis with crisp or fuzzy partition of units. *Psychometrika*, 70, 715-736.

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.

Roubens, M. (1982). Fuzzy clustering algorithms and their cluster validity. European *Journal of Operational Research*, 10, 294-301.

Tucker, L. R. (1951). A method for synthesis of factor analysis studies (Personnel Research Section Report No. 984). Washington, D.C.: U.S. Department of the Army.

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.

Wedel, M., & Steenkamp, J.-B. E. M. (1989). Fuzzy clusterwise regression approach to benefit segmentation. *International Journal of Research in Marketing*, 6, 241-258.

Wedel, M., & Steenkamp, J.-B. E. M. (1991). A clusterwise regression method for simultaneous fuzzy market structuring and benefit segmentation. *Journal of Marketing Research*, 28, 385-396.

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

Yamakawa, A., Kanaumi, Y., Ichibashi, H., & Miyoshi, T. (1999). Simultaneous application of clustering and correspondence analysis. *IJCNN'99. International Joint Conference on Neural Networks*, 6, 4334-4338.

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.

Yang, B. (1997). A consumer research report on underwear in South Korea. Unpublished report, Chung-Ang University, Seoul, South Korea.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8, 338-353.

Table 1. Mean coordinate estimates, mean standard errors, and biases obtained from the simulation study.

		N = 50						*N = 100*						*N = 200*					
		FCMCA			Tandem			FCMCA			Tandem			FCMCA			Tandem		
True		est	s.e.	bias	est	s.e.	bias	est	s.e.	bias	est	s.e.	bias	est	s.e.	bias	est	s.e.	bias
	- .4	-.39	.03	.01	-.48	.02	-.08	-.39	.02	.01	-.48	.02	-.08	-.39	.01	.01	-.48	.01	-.08
	.6	.59	.03	-.01	.49	.03	-.11	.59	.02	-.01	.48	.02	-.12	.59	.02	-.01	.48	.02	-.12
	-.4	-.39	.03	.01	-.48	.02	-.08	-.39	.02	.01	-.48	.01	-.08	-.39	.02	.01	-.48	.01	-.08
	.6	.59	.04	-.01	.48	.02	-.12	.59	.03	-.01	.49	.02	-.11	.59	.02	-.01	.49	.01	-.11
	-.5	-.51	.03	-.01	-.40	.02	.10	-.51	.02	-.01	-.40	.01	.10	-.51	.02	-.01	-.40	.01	.10
	.5	.51	.03	.01	.40	.02	.10	.51	.02	.01	.40	.01	.10	.51	.02	.01	.40	.01	.10
	-.5	-.51	.03	-.01	-.60	.03	-.10	-.51	.02	-.01	-.60	.02	-.10	-.51	.02	-.01	-.60	.02	-.10
	.5	.51	.03	.01	.60	.03	.10	.51	.02	.01	.60	.02	.10	.51	.01	.01	.60	.02	.10
	-.6	-.59	.03	.01	-.48	.03	.12	-.59	.02	.01	-.49	.02	.11	-.59	.02	.01	-.49	.02	.11
	.4	.39	.03	-.01	.48	.02	.08	.39	.02	-.01	.48	.02	.08	.39	.01	-.01	.48	.01	.08
	-.6	-.59	.04	.01	-.48	.02	.12	-.59	.03	.01	-.48	.02	.12	-.59	.02	.01	-.48	.01	.12
$\mathbf{W}_j$	.4	.39	.03	-.01	.48	.02	.08	.39	.02	-.01	.48	.01	.08	.39	.01	-.01	.48	.01	.08
	.5	.51	.03	.01	.63	.03	.13	.51	.02	.01	.63	.02	.13	.51	.01	.01	.63	.02	.13
	.5	.51	.03	.01	.43	.02	-.07	.51	.02	.01	.43	.01	-.07	.51	.02	.01	.43	.01	-.07
	-.5	-.50	.03	.00	-.56	.03	-.06	-.50	.02	.00	-.56	.02	-.06	-.51	.02	-.01	-.56	.02	-.06
	-.5	-.50	.04	.00	-.37	.02	.13	-.51	.03	-.01	-.37	.01	.13	-.51	.02	-.01	-.37	.01	.13
	.6	.58	.03	-.02	.51	.03	-.09	.59	.02	-.01	.51	.02	-.09	.59	.02	-.01	.51	.01	-.09
	.6	.58	.03	-.02	.51	.03	-.09	.58	.02	-.02	.51	.02	-.09	.58	.02	-.02	.51	.01	-.09
	-.4	-.39	.03	.01	-.44	.02	-.04	-.39	.02	.01	-.44	.01	-.04	-.39	.01	.01	-.45	.01	-.05
	-.4	-.39	.03	.01	-.44	.02	-.04	-.39	.02	.01	-.45	.01	-.05	-.39	.02	.01	-.43	.01	-.03
	.5	.51	.03	.01	.43	.02	-.07	.51	.02	.01	.43	.01	-.07	.51	.02	.01	.43	.01	-.07
	.5	.51	.03	.01	.63	.03	.13	.51	.02	.01	.63	.02	.13	.51	.01	.01	.63	.02	.13
	-.5	-.51	.03	-.01	-.37	.02	.13	-.50	.03	.00	-.37	.01	.13	-.50	.02	.00	-.37	.01	.13
	-.5	-.50	.03	.00	-.56	.03	-.06	-.51	.02	-.01	-.56	.02	-.06	-.51	.02	-.01	-.56	.02	-.06
	-.5	-.47	.09	.03	-.46	.09	.04	-.48	.06	.02	-.48	.07	.02	-.48	.04	.02	-.48	.04	.02
	.5	.49	.07	-.01	.49	.07	-.01	.50	.04	.00	.50	.04	.00	.51	.03	.01	.51	.03	.01
	-.5	-.52	.06	-.02	-.52	.06	-.02	-.52	.05	-.02	-.52	.05	-.02	-.52	.03	-.02	-.53	.03	-.03
$\mathbf{r}_c$	.5	.49	.05	-.01	.49	.05	-.01	.50	.03	.00	.50	.03	.00	.50	.02	.00	.50	.02	.00
	.5	.46	.11	-.04	.45	.11	-.05	.47	.07	-.03	.46	.08	-.04	.47	.05	-.03	.46	.05	-.04
	.5	.51	.08	.01	.51	.08	.01	.51	.06	.01	.52	.05	.02	.52	.04	.02	.52	.04	.02
	-.5	-.48	.07	.02	-.48	.06	.02	-.49	.04	.01	-.48	.05	.02	-.49	.03	.01	-.49	.03	.01
	-.5	-.48	.05	.02	-.48	.05	.02	-.49	.03	.01	-.49	.03	.01	-.49	.02	.01	-.49	.02	.01
		CC(U) = .91			CC(U) = .92			CC(U) = .92			CC(U) = .93			CC(U) = .93			CC(U) = .93		

CC(U): Congruence coefficient between true fuzzy memberships and estimates.

Table 2. Brand and attribute descriptions for the Korean underwear data

Brands	Attributes
BYC ('by')	Comfortable (1)
TRY ('tr')	Smooth (2)
VICMAN ('vm')	Superior fabrics (3)
James Dean ('jd')	Reasonable price (4)
Michiko-London ('ml')	Fashionable design (5)
Benetton ('bn')	Favourable advertisements (6)
Bodyguard ('bg')	Trendy colours (7)
Calvin Klein ('ck')	Good design (8)
	Various colours (9)
	Elastic (10)
	Store is near (11)
	Excellent fit (12)
	Design quality (13)
	Youth appeal (14)
	Various sizes (15)

Table 3. The adjusted inertias and their percentages of the adjusted total inertia in the parenthesis obtained from MCA for the two empirical data sets.

Korean underwear data	French worker survey data
0.0734 (35%)	0.2667 (55%)
0.0443 (21%)	0.1031 (21%)
0.0387 (19%)	0.0742 (15%)
0.0220 (11%)	0.0306 (6%)
0.0142 (7%)	0.0064 (1%)
0.0098 (5%)	0.0023 (0.5%)
0.0047 (2%)	0.0016 (0.3%)
0.0000 (0.0%)	0.0004 (0.1%)
0.0000 (0.0%)	0.0002 (0.0%)
0.0000 (0.0%)	

Table 4. Question category descriptions for the French worker survey data

Q1	Q2	Q3
CGT (11)	Jacques Duclos (21)	Communist (31)
CFDT (12)	Gaston Defferre (22)	Socialist (32)
FO (13)	Alain Krivine (23)	Left (33)
CFTC (14)	Michel Rocard (24)	Centre (34)
Autonomous (15)	Alain Poher (25)	RI (35)
Abstention (16)	Louis Ducatel (26)	Right (36)
Non-affiliated (17)	Georges Pompidou (27)	Gaullist (37)
No response (18)	No response (28)	No response (38)

Figure 1. The true coordinates of variable categories and cluster centroids in the simulation study.

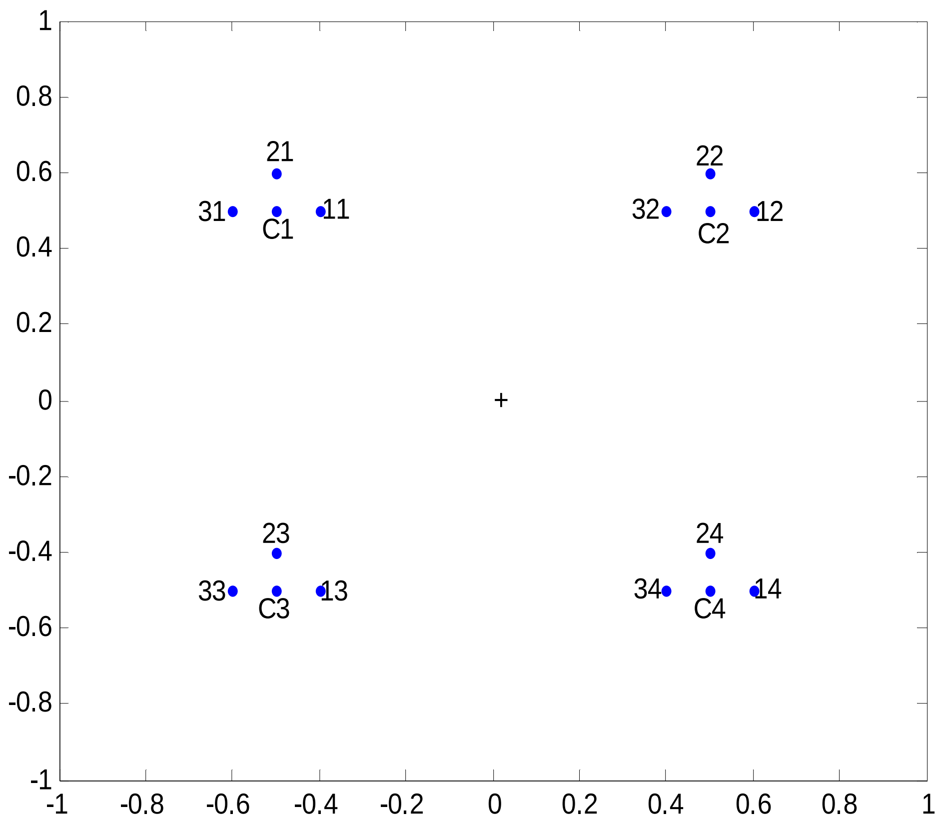

Figure 2. The plot of the values of the optimization criterion (*f*) and two cluster validity measures (FPI and NCE) against different numbers of clusters for the Korean underwear data.

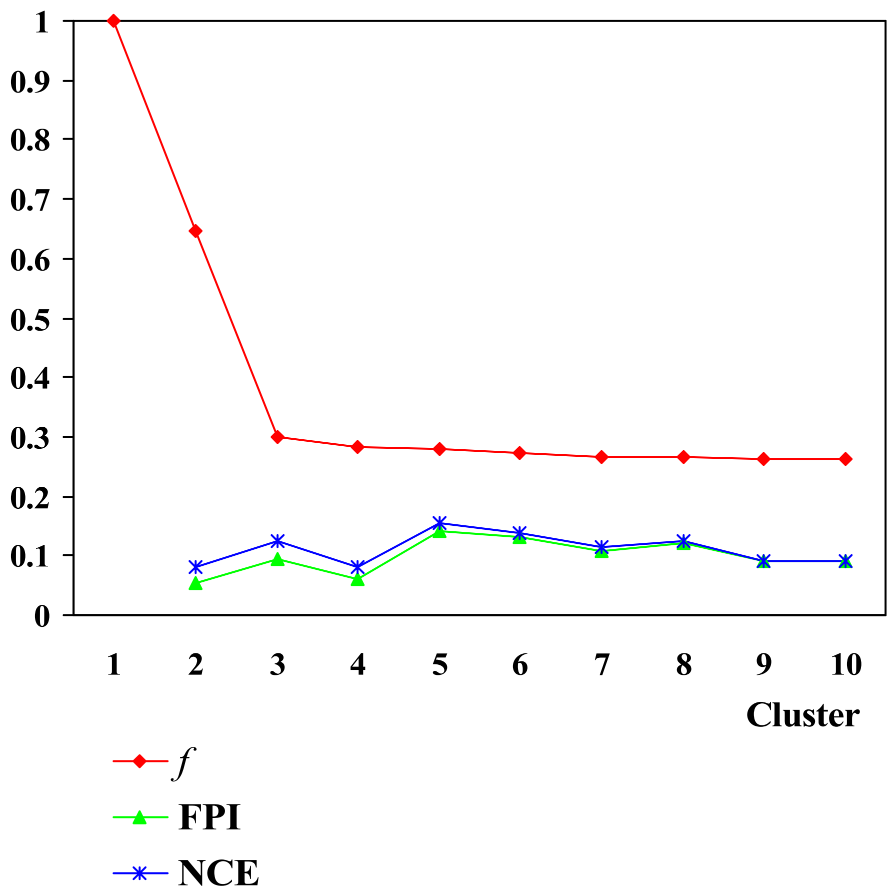

Figure 3. FCMCA - The principal coordinates of variable categories and the centroids of three clusters for the Korean underwear data.

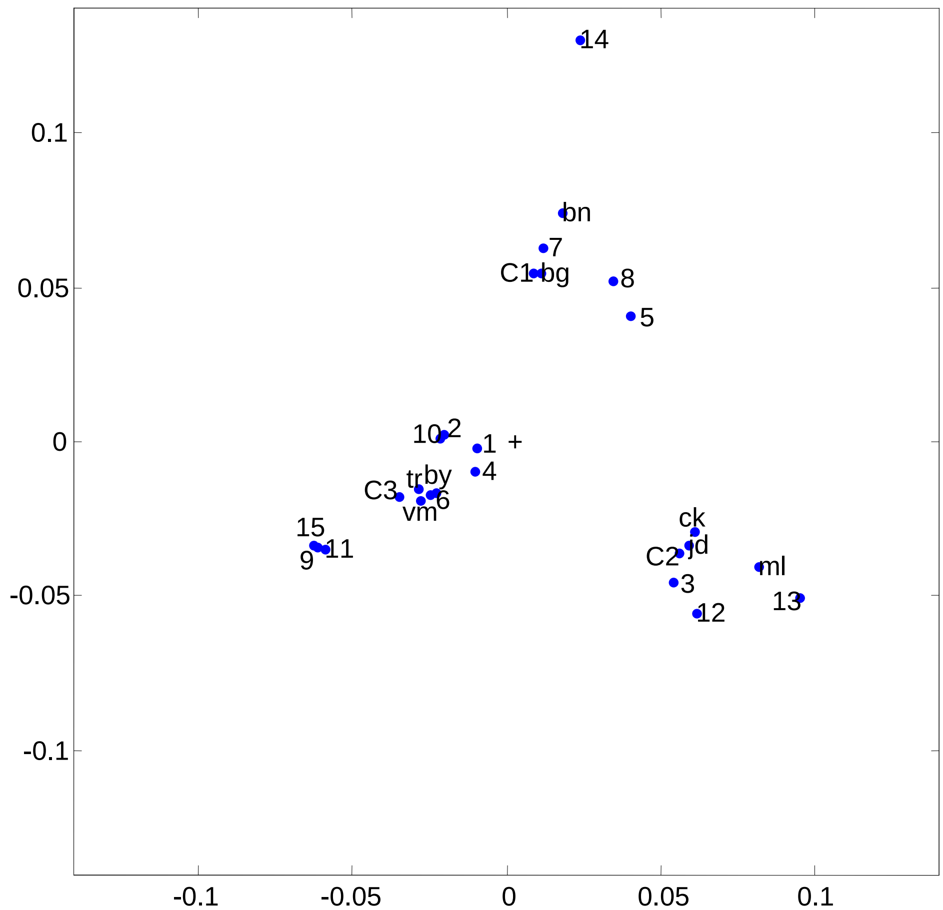

Figure 4. The tandem approach - The principal coordinates of variable categories and the centroids of three clusters for the Korean underwear data.

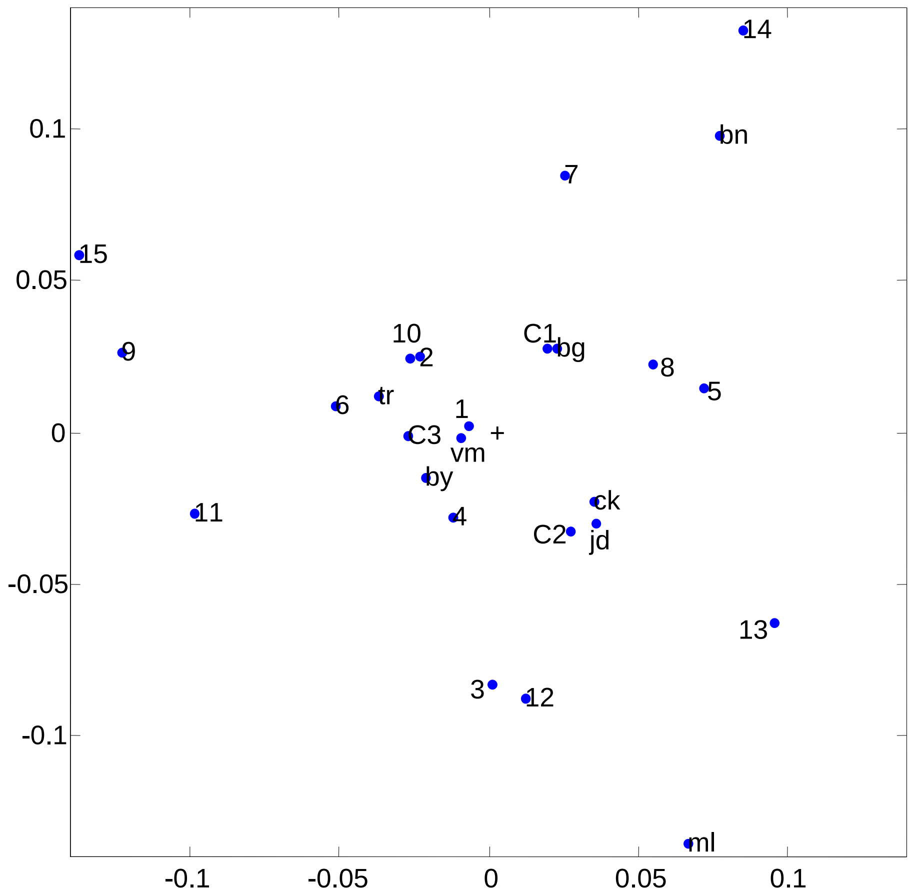

Figure 5. Hwang et al.'s (2006) approach - The principal coordinates of variable categories and the centroids of three clusters for the Korean underwear data.

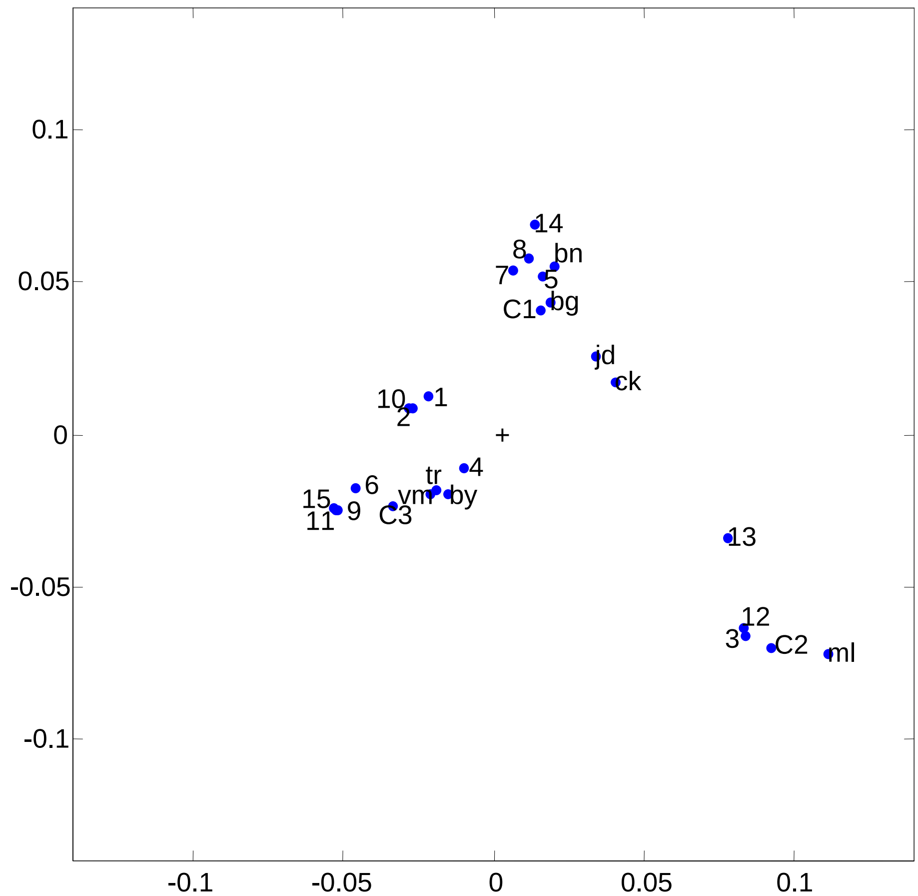

Figure 6. The plot of the values of the optimization criterion (*f*) and two cluster validity measures (FPI and NCE) against different numbers of clusters for the French worker survey data.

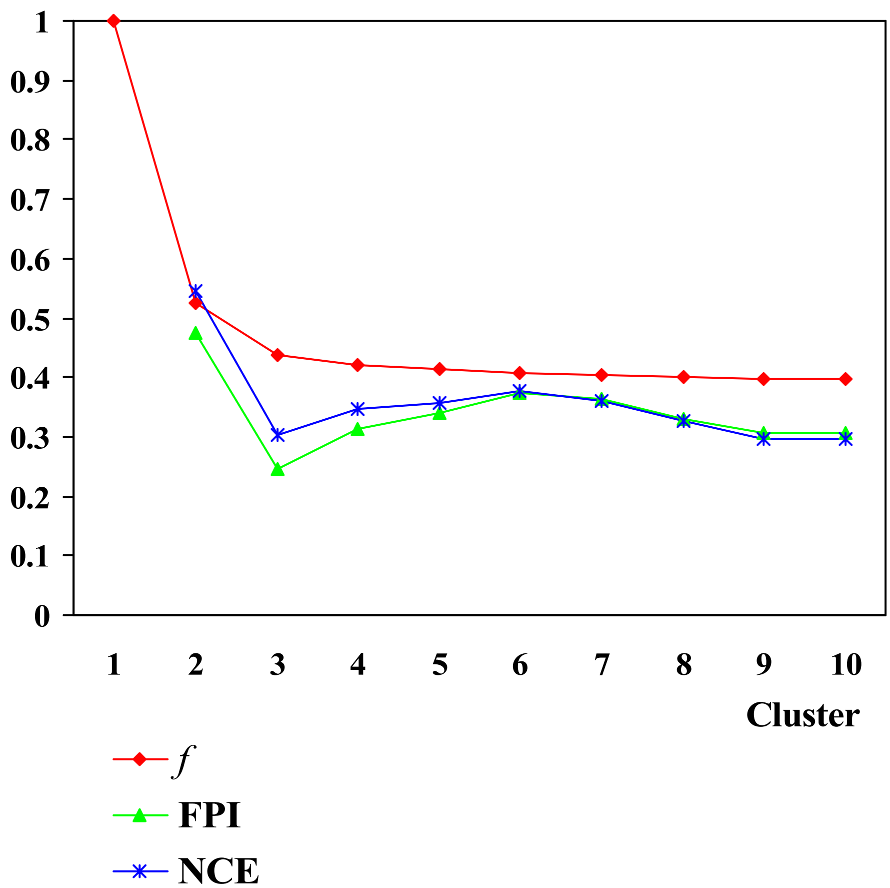

Figure 7. FCMCA - The principal coordinates of variable categories and the centroids of three clusters for the French worker survey data.

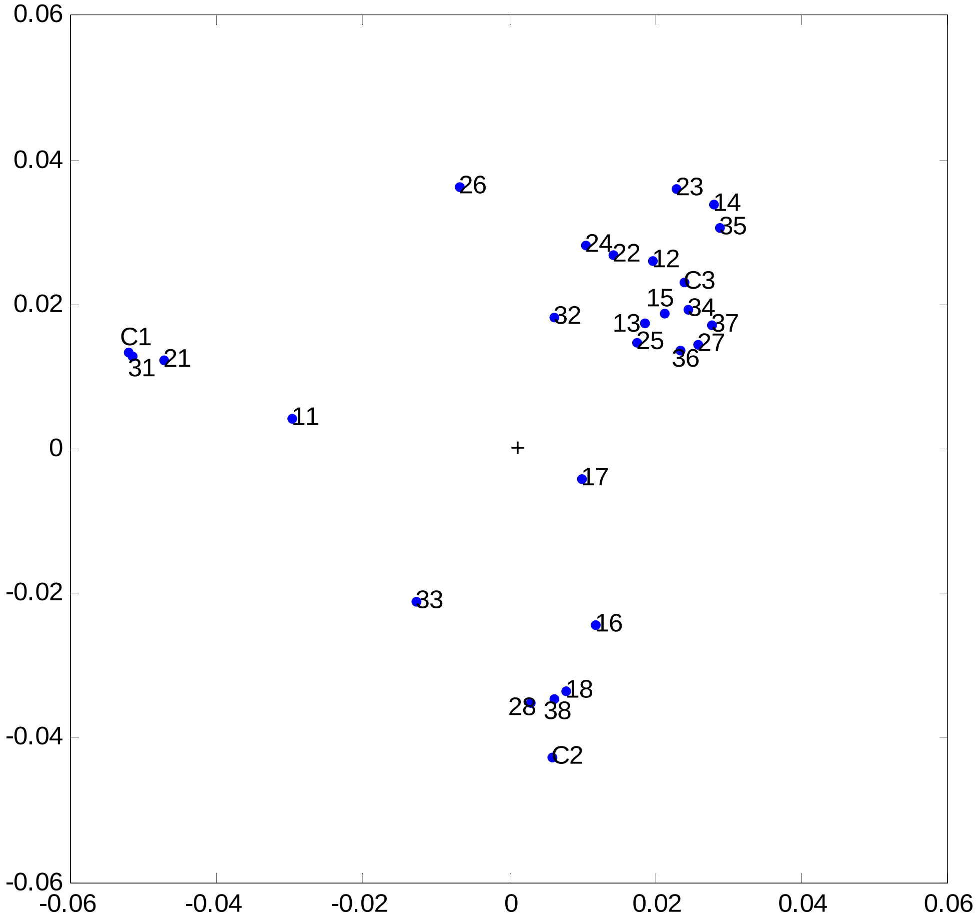

Figure 8. The tandem approach - The principal coordinates of variable categories and the centroids of three clusters for the French worker survey data.

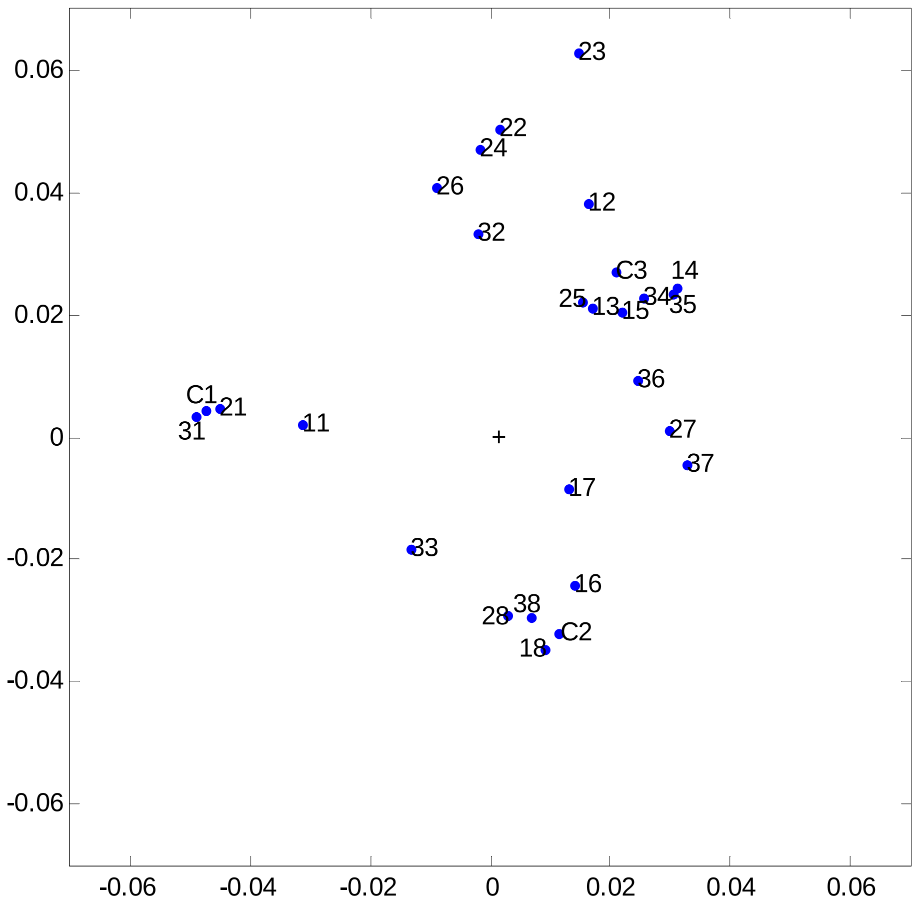

Figure 9. Hwang et al.'s (2006) approach - The principal coordinates of variable categories and the centroids of three clusters for the French worker survey data.