Some Alternative Clustering Methods for Mokken Scale Analysis

van Abswoude, A.A.H.; Vermunt, J.K.

Published in:
New Developments in Psychometrics.

Publication date:
2003

Some Alternative Clustering Methods for Mokken Scale Analysis

A.A.H. van Abswoude[1] and J. K. Vermunt[2]

[1] Tilburg University, P.O. Box 90153, Tilburg, The Netherlands
A.A.H.vanAbswoude@kub.nl
[2] Tilburg University, P.O. Box 90153, Tilburg, The Netherlands
J.K.Vermunt@kub.nl

Summary. In this paper three methods for finding the dimensionality of a data matrix - sequential clustering, hierarchical clustering, and non-hierarchical clustering - are discussed. For each method different measures are suggested to find one or more sets of dichotomous items that satisfy the conditions of a Mokken scale (Mokken, 1971). It is illustrated that non-hierarchical clustering resolves some problems associated with sequential and hierarchical clustering in finding the true dimensionality.

Key words: multidimensionality, item clustering algorithms, nonparametric item response theory.

1 Introduction

In measurement and scaling it is important to measure one single latent trait at one time. Otherwise, unless the exact relationship between the latent variables can be modelled, it will be difficult to make any correct predictions about the subjects' position on a latent trait. In practise, however, researchers are often confronted with data matrices having multiple latent traits. For example, responses to a test measuring *crying* may contain items on subtraits such as *distress*, *sadness*, and *joy*. In these cases, methods like Mokken scale analysis (MSA; Mokken, 1971) may be used to find sets of items that form a single scale.

The MSA-software, MSP (Molenaar & Sijtsma, 2000), uses a sequential clustering algorithm to find sets of items (clusters) that satisfy the conditions of a Mokken scale. A drawback of this algorithm is, however, that it may not yield the *optimal solution*; that is, the solution that maximizes a certain objective function. If the objective function is correct, the optimal solution should reflect the true dimensionality of the data matrix. In this paper, we present two simple new clustering methods, hierarchical and non-hierarchical clustering, for MSA that may do a better job in finding the optimal solution. In the following sections, we will describe how the clustering methods work and discuss how the MSA conditions can be imposed within each of these methods so that solutions may reflect the true dimensionality and satisfy the MSA conditions. In addition, the ability of the methods in finding the optimal solution will be illustrated by means of two small generated examples.

2 Nonparametric IRT Framework

In nonparametric item response theory (IRT) it is assumed that a single underlying latent trait (θ) governs the responses on a set of items (unidimensionality, UD). Further it is assumed that given any value of θ the responses of an individual on a set of items are statistically independent (local independence, LI). Lastly, it is assumed that there is a monotonely nondecreasing relationship between the probability of answering an item correctly and θ (monotonicity, M). A set of items that satisfy UD, LI and M are denoted as monotone homogeneous (MH; Mokken, 1971). The monotone homogeneity model allows the measurement of θ by means of the total test scores of individuals (Grayson, 1988). The total test score is defined by $X_+ = \sum X_i$, with X_i denoting an individual's score on item i ($i = 1, ..., I$). Within the nonparametric IRT framework, the response probability given θ, known as the item response function (IRF), does not need to have a particular shape such as the logistic as long as the items are MH.

Mokken scale analysis, which is a nonparametric IRT method for scale analysis, uses Loevinger's coefficient of homogeneity to quantify the strength of the association within the responses to a pair of items. Let items i and j be two binary items having item scores 0 or 1. Let π_i represent the probability of answering item i correctly, and π_{ij} the probability of answering both item i and j correctly. Items are ordered such that $\pi_i > \pi_j$, $\forall i > j$. The pairwise scalability coefficient H_{ij} for items i and j is defined as

$$H_{ij} = \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i(1 - \pi_j)}. \tag{1}$$

This equals the covariance between items i and j divided by the maximum covariance given the marginal score distributions of items i and j (Mokken, 1971; Molenaar & Sijtsma, 2000). The H_i for item i can be written as

$$H_i = \frac{\sum_{j \neq i}(\pi_{ij} - \pi_i \pi_j)}{\sum_{j \neq i} \pi_i(1 - \pi_j)} = \frac{\sum_{j \neq i} \pi_i(1 - \pi_j) H_{ij}}{\sum_{j \neq i} \pi_i(1 - \pi_j)} \tag{2}$$

as in Mokken (1971, p.150). The scale H is defined as

$$H = \frac{\sum_i \sum_{j \neq i}(\pi_{ij} - \pi_i \pi_j)}{\sum_i \sum_{j \neq i} \pi_i(1 - \pi_j)} = \frac{\sum_i \sum_{j \neq i} \pi_i(1 - \pi_j) H_{ij}}{\sum_i \sum_{j \neq i} \pi_i(1 - \pi_j)}. \tag{3}$$

For more information about the theoretical basis and the sampling distribution of the H coefficient we refer the reader to Mokken (1971) or Molenaar & Sijtsma (2000).

A set of I items are called a *Mokken Scale* (Mokken, 1971, p.184) if all items satisfy the following two conditions,

Condition 1 $\text{cov}(X_i, X_j) > 0$, for all $i \neq j$, and

Condition 2 $H_i \geq c$, for all i, where c is a user-defined constant between 0 and 1.

From the MH model follows that $\text{cov}(X_i, X_j) \geq 0$ (Holland & Rosenbaum, 1986), but the first scaling condition is less restrictive. The second condition serves the practical purpose of only including items into a scale that sufficiently discriminate. Also, when c is sufficiently high -in practice, $c = 0.3$ is used- a single latent trait will be measured.

Below, we present three types of clustering procedures (i.e., sequential clustering, hierarchical clustering, and non-hierarchical clustering) for finding the dimensionality of a data matrix using the H coefficient as a measure of association. Each method uses H_{ij}, π_i, and π_j as input data; and each must satisfy the two conditions of a Mokken scale. In the next sections, we describe how these conditions can be imposed, as well as what the limitations of these methods are when searching for the dimensionality of a set of items.

3 Sequential Clustering

Sequential clustering takes the following stepwise course. In the first step (Step 1), we search for the two items that form the start set for the first scale. This is the item pair with the highest significantly positive H_{ij} (Equation 1) that satisfies Conditions 1 and 2. In the next step (Step 2), items are added one at the time to this start set. More precisely, this is the item with the highest H_i with the previously selected items (Equation 2) that satisfies Conditions 1 and 2. This process of adding items with the highest H_i continues until no item remains. When this happens the first scale has been formed. As long as scalable items remain, subsequent scales may be formed by repeating Steps 1 and 2 using the remaining items. The procedure terminates when no scalable items are left.

The sequential procedure works quite well when one is interested in finding only one single Mokken scale, for example when a data matrix measures one dominant latent trait, and possibly one or more nuisance latent traits. When searching for multiple Mokken scales in a multi-trait context, however, the sequential nature of the procedure may yield suboptimal solutions: the solution that yields the highest overall H_i is not obtained. The reason that suboptimal solutions may be obtained is that sequential clustering forms the clusters one at the time: some items may be collected in Cluster 1, for example, because scaling criteria were still satisfied in Step 2; but it may have a higher H_i when joined with items in Cluster 2.

4 Hierarchical Clustering

Agglomerative hierarchical clustering (Everitt et al., 2001) seems to be a useful alternative because it can yield multiple clusters simultaneously, where

sequential clustering could not. Starting point of a HCA is a data matrix containing proximities between items i and j. The proximities in our case are based on the H coefficient and will be discussed in more detail later on. At each hierarchical step, the two clusters with the highest proximity are joined. This means that at any hierarchical step two single items may be joined to form one new cluster, a single item may be joined with an existing cluster of items, or two clusters may be joined into a single larger cluster. This process continues until some previously defined criterion is met or until all items are in one single cluster.

In the following methods, four types of proximities proposed may be used to form Mokken scales. The first three methods can be reproduced using the H_{ij}-matrix in combination with standard clustering procedures of most statistical packages, including SPSS, SAS, and BMDP. For the fourth method dedicated software was written in PASCAL. Before we present the proximities, however, first we need some notation. Let k and l represent two clusters, let I_k represent the number of items in cluster k, I_l represent the number of items in cluster l, and let H_{kl} represent the proximity between clusters k and l.

In *complete linkage* the proximity between clusters k and l is defined as,

$$H_{kl}^{complete} = \min(H_{ij}), \text{ where } i \in k \text{ and } j \in l.$$

This method joins the two clusters for which the lowest H_{ij} of the two clusters is maximized. This definition of proximity is intuitively attractive because it produces scales for which the $\min(H_{ij})$ satisfies some minimal requirement.

Average linkage defines the proximity between clusters as

$$H_{kl}^{average} = \frac{\sum_{i \in k} \sum_{j \in l} H_{ij}.}{I_k I_l}.$$

As can be seen, $H_{kl}^{average}$ is the unweighted average of the bivariate H_{ij} *between* the items in cluster k and the items in cluster l. This measure can be viewed as a proxy for the average H_i in a cluster.

Within-groups linkage defines the proximity of two clusters k and l as the unweighted average of the H_{ij} of all items *within* k and l; that is,

$$H_{kl}^{within} = \frac{\sum_{i \in k \cup l} \sum_{j \neq i} H_{ij}}{(I_k + I_l)(I_k + I_l - 1)}.$$

This proximity can be seen as a proxy for H as defined in Equation 3.

The fourth method, *scale linkage* uses the scale H (Equation 3) of the possible new cluster that is obtained by joining two clusters as proximity measure. Written in terms of clusters k and l the proximity in scale linkage is defined as,

$$H_{kl}^{scale} = \frac{\sum_{i \in k \cup l} \sum_{j \neq i} \pi_i(1 - \pi_j) H_{ij}}{\sum_{i \in k \cup l} \sum_{j \neq i} \pi_i(1 - \pi_j)}.$$

Different from sequential clustering is that, unless a stopping rule is used, the HCA will continue clustering until all items are joined into one large cluster. For instance, the Mokken scaling conditions (especially, Condition 2) can be used to terminate the clustering process. In that case, clustering stops when the conditions are no longer satisfied. Alternative methods to decide when to stop the HCA can also be proposed, but are beyond the scope of this paper. In this paper we used the Mokken scale conditions as stopping rule.

Unfortunately, hierarchical clustering may also yield suboptimal solutions because clusters that have been formed in previous steps remain intact in subsequent steps. More precisely, a set of items that was clustered at an earlier stage may not be homogeneous with respect to items that were added later.

5 Non-Hierarchical Clustering

Non-hierarchical clustering refers to a class of algorithms where multiple clusters are formed simultaneously and single units within an object (i.e., items) can be moved from one cluster to another. We use a criterion to evaluate the H_i of all items i of a partition $\mathcal{P}_t$ at iteration t.

Let $\delta_{ik}(\mathcal{P}_t) = 1$ if $i \in k$ (where $k = 1, ..., K$), and $\delta_{ik}(\mathcal{P}_t) = 0$ if $i \notin k$ at $\mathcal{P}_t$. In addition, let $H_{i|k}$ be matrix reflecing the conditional homogeneity of each item i with respect to the items in each cluster k. A criterion for evaluating the quality of a K-cluster at partition $\mathcal{P}_t$ may, for instance, be

$$\text{Crt } 1(\mathcal{P}_t) = I^{-1} \sum_{i=1}^{I} \sum_{k=1}^{K} \delta_{ik}(\mathcal{P}_t) H_{i|k}. \tag{4}$$

The goal of the non-hierarchical clustering procedure is to search for that partition that maximizes Crt 1; that is, we intend to join each item into the cluster such that the highest H_i is obtained for all items.

In this paper, we use a k-means type algorithm to assign items to clusters. This clustering method begins with an initial configuration ($t = 0$) in which I items are randomly assigned to K clusters, and the quality of $\mathcal{P}_0$ is evaluated using Crt $1(\mathcal{P}_0)$. In each iteration step, one item i may be moved to another cluster k and Crt $1(\mathcal{P}_1)$ is evaluated. Different rules may be used to move an item i to another cluster k. For example, we could move the item to that cluster for which the improvement in $H_{i|k}$ is the best. In the subsequent iterations this evaluating and maximizing of Crt $1(\mathcal{P}_t)$ is continued until the criterion can no longer be improved.

This procedure can be further refined by adding a random component to the process of assigning items to clusters, thereby reducing the probability of ending up in a local maximum: a stochastic process could be used for the assignment of items to clusters. This means that in the first iterations, Crt 1 may deteriorate from one iteration to the next (e.g., an item is moved to a

cluster for which $H_{i|k}$ is not the highest). For later iterations improvements of Crt 1 are more likely. This random element in the composition of clusters is important because it may yield combinations of items that otherwise would not have been found.

Other criteria could also be used in this context. The following criterion was formulated for finding the partition having the highest H_{ij} within clusters. We adapted Zhang & Stout's statistic for this purpose, which is aimed at finding clusters of items that are locally independent (Zhang & Stout, 1999). Let $\delta_{ijk}(\mathcal{P}_t) = 1$ when items i and j are joined into cluster k at $\mathcal{P}_t$, and -1 otherwise. Then, Crt $2(\mathcal{P}_t)$ is defined as

$$\text{Crt } 2(\mathcal{P}_t) = \frac{2}{I(I-1)} \sum_{1 \leq i < j \leq I} \delta_{ijk}(\mathcal{P}_t) H_{ij}.$$

Criterion Crt 2 can also be maximized using the k-means type procedure which was presented before. In the simulation study, however, we limit ourselves to finding the dimensionality using the k-means type algorithm based on Crt 1 (i.e., without the proposed refinements).

6 Simulation study

The performance of the three general procedures (i.e., sequential, hierarchical, and non-hierarchical clustering) using their specific definitions of dimensionality is shown for two small generated item pools. We used the multidimensional extension of the 2-parameter logistic item response theory model (M2-PLM; e.g., Reckase, 1985) to generate 1000 item responses on a set of six items and a set of ten items. Even though the data were generated, the used item parameter values are representative for true test data. Extending these examples to data matrices having more items will not add much substantial to the simulation study because similar results will be obtained.

Item pool 1 consists of 6 items, $X_1 - X_6$, and two latent traits, θ_1 and θ_2. Items X_1 and X_2 are strongly related to θ_1, X_3 is weakly related to θ_1 and strongly related to θ_2, and $X_4 - X_6$ are moderately related to θ_2 but not related to θ_1.

Item pool 2 consists of 10 items, and three latent traits, $\theta_1 - \theta_3$. Items $X_1 - X_5$ are moderately related to θ_1, X_5 and X_6 are moderately related to θ_2, and $X_6 - X_{10}$ are moderately related to θ_3. The second latent trait, θ_2, has the function of a nuisance trait: that is, it is included in order to make the detection of the items measuring θ_1 and θ_3 more difficult. In both item pools latent traits are assumed to be uncorrelated.

We evaluated the performance of the methods as to whether the results are optimal. Notation $[K : I_1; I_2; ...; I_K]$ is used to reflect the structure of an item pool, where K equals the number of clusters and $I_k (k = 1, ..., K)$ equals the number of items in each cluster. The optimal solution of Item pool 1 using

$c = 0.3$ in Criterion 2 is $[2 : 2; 4]$; that is, items $X_1 - X_2$ are in Cluster 1 and $X_3 - X_6$ in Cluster 2. We look at two optimal solutions for Item pool 2: $[2 : 5; 5]$ using $c = 0.2$ in Condition 2 (i.e., $X_1 - X_5$ and $X_6 - X_{10}$); and $[3 : 4; 2; 4]$ using $c = 0.3$ in Condition 2 (i.e., $X_1 - X_4$, $X_5 - X_6$, and $X_7 - X_{10}$). The different values of c reflect two possible ways of defining a Mokken scale: one is moderately strict ($c = 0.3$) and one is less strict ($c = 0.2$).

We used the default settings for MSP (Molenaar & Sijtsma, 2000) and HCA. For the k-means type algorithm we used the simple version that was portrayed before in combination with Crt 1.

7 Results

Tabel 1 shows the optimal solution and the dimensionality results obtained with sequential clustering, hierarchical clustering, and non-hierarchical clustering. The first column gives the method used for clustering, the other columns give the results for item pools 1 and 2.

Table 1. Number of Clusters and Number of Items per Cluster Using Sequential Clustering, Four Hierarchical Clustering Methods and Non-Hierarchical Clustering for Two Small Generated Item Pools

Method	Item pool 1	Item pool 2
Optimal solution	[2: 2;4]	[2: 5;5], [3: 4;2;4]
Sequential	[2: 3;3]	[2: 5;5], [3: 4;2;4]
Complete linkage	[2: 2;4]	[2: 6;4], [3: 4;2;4]
Average linkage	[2: 2;4]	[2: 6;4], [3: 4;2;4]
Within-groups linkage	[2: 2;4]	[2: 5;5], [3: 1;4;5]
Scale linkage	[2: 2;4]	[2: 5;5], [3: 1;4;5]
Simple + Crt 1	[2: 2;4]	[2: 5;5], [3: 4;2;4]

For item pool 1, all methods, except sequential clustering, yielded the predefined optimal solution. The reason that the first cluster, obtained with sequential clustering, contained one extra item was that this item still satisfied the scaling criteria for the first cluster, although it measures θ_2. Forming multiple clusters simultaneously (i.e., hierarchical and non-hierarchical clustering) was sufficient to find the optimal solution.

In item pool 2, both the $K = 2$ and the $K = 3$ results are presented for each method. Sequential clustering yielded the correct dimensionality for both $K = 2$ and $K = 3$. With the hierarchical procedures either the $K = 2$ or the $K = 3$ solution was correct. One may note that with a hierarchical procedure it is not possible that both the $K = 2$ and the $K = 3$ solution are optimal because the $K = 2$ solution is obtained by combining the two clusters of the $K = 3$ solution. The non-hierarchical method yielded two

optimal solutions because clusters are formed simultaneously and individual items are assigned to the clusters they fit best.

8 Conclusion

Three types of methods for finding the dimensionality of a set of items were presented in this paper. Each method was adapted to yield clusters that satisfy the Mokken scale conditions. As illustrated, non-hierarchical clustering resolves the problems associated with sequential clustering and hierarchical clustering.

The non-hierarchical clustering algorithm we used in the simulation study may yield local maxima. Introducing randomness in the assignment of items to clusters may be the remedy for this problem that deserves further study.

References

Everitt BS, Landau S, Leese M (2001) Cluster Analysis. Arnold/Oxford University Press, London/New York

Grayson DA (1988) Two-group classification in latent trait theory: Scores with monotone likelihood ratio. Psychometrika 53:383–392

Holland PW, Rosenbaum PR (1986) Conditional association and unidimensionality in monotone latent variable models. The Annals of Statistics 14:1523–1543

Mokken RJ (1971) A theory and procedure of scale analysis: With applications in political research. De Gruyter, Berlin

Molenaar IW, Sijtsma K (2000) Users manual MSP5 for Windows. A program for Mokken scale analysis for polytomous items [Software manual]. Iec ProGamma, Groningen, The Netherlands

Reckase MD (1985) The difficulty of test items that measure more than one ability. Applied Psychological Measurement 15:401–412

Zhang J, Stout WF (1999) Conditional covariance structure of generalized compensatory multidimensional items. Psychometrika 64:129–152