PSYCHOMETRIKA— VOL. 70, NO. 3, 461–480
SEPTEMBER 2005
DOI: 10.1007/s11336-003-1067-3

HIERARCHICAL CLASSES MODELS FOR THREE-WAY THREE-MODE BINARY DATA: INTERRELATIONS AND MODEL SELECTION

EVA CEULEMANS AND IVEN VAN MECHELEN

KATHOLIEKE UNIVERSITEIT LEUVEN

Several hierarchical classes models can be considered for the modeling of three-way three-mode binary data, including the INDCLAS model (Leenen, Van Mechelen, De Boeck, and Rosenberg, 1999), the Tucker3-HICLAS model (Ceulemans, Van Mechelen, and Leenen, 2003), the Tucker2-HICLAS model (Ceulemans and Van Mechelen, 2004), and the Tucker1-HICLAS model that is introduced in this paper. Two questions then may be raised: (1) how are these models interrelated, and (2) given a specific data set, which of these models should be selected, and in which rank? In the present paper, we deal with these questions by (1) showing that the distinct hierarchical classes models for three-way three-mode binary data can be organized into a partially ordered hierarchy, and (2) by presenting model selection strategies based on extensions of the well-known scree test and on the Akaike information criterion. The latter strategies are evaluated by means of an extensive simulation study and are illustrated with an application to interpersonal emotion data. Finally, the presented hierarchy and model selection strategies are related to corresponding work by Kiers (1991) for principal component models for three-way three-mode real-valued data.

Key words: three-way three-mode data, binary data, hierarchical classes, multiway analysis, model selection.

1. Introduction

Three-way three-mode binary data are often found in psychology. For some examples, one may think of person by situation by behavior display/not display data or consumer by product by time point select/not select data. To model three-way three-mode binary data, the family of hierarchical classes models, a collection of order preserving Boolean decomposition models, has been developed. More specifically, this family consists of the INDCLAS model (Leenen, Van Mechelen, De Boeck and Rosenberg, 1999), the Tucker3-HICLAS model (Ceulemans, Van Mechelen, and Leenen, 2003), the Tucker2-HICLAS model (Ceulemans and Van Mechelen, 2004), and the Tucker1-HICLAS model that will be introduced in this paper. Each of these hierarchical classes models for three-way three-mode binary data approximates a given data set with a binary reconstructed data or model array of the same size. Furthermore, each of these models includes (1) a Boolean decomposition of the model array, and (2) hierarchically organized classifications of the elements of the three modes (hierarchy to be understood in terms of if-then relations). Both features are of substantive importance as (1) the Boolean decomposition may reveal the structural mechanism underlying the data, and (2) asymmetric, implicational relations between the elements of a mode are often of substantive interest. The latter is, for example, illustrated by successful applications of hierarchical classes modeling of three-way three-mode binary data in contextualized personality and differential emotion psychology (see e.g., Vansteelandt and Van Mechelen, 1998; Kuppens, Van Mechelen, Smits, De Boeck, and Ceulemans, 2005).

Considering that several models have been developed for one and the same type of data, two questions may be raised: (1) how are the different hierarchical classes models for three-way three-

The research reported in this paper was partially supported by the Research Council of K.U. Leuven (GOA/2000/02 and PDM/03/074). Furthermore, the authors are obliged to Kaatje Bollaerts and the three anonymous reviewers for useful comments on an earlier version of this paper.

Requests for reprints should be sent to Eva Ceulemans, Department of Psychology, Tiensestraat 102, B-3000 Leuven, Belgium. Email: Eva.Ceulemans@psy.kuleuven.be

© 2005 The Psychometric Society

mode binary data interrelated, and (2) given a specific data set, which of these models should be selected? Regarding the relationships among the different models, up to now only the relationships between, on the one hand, INDCLAS and Tucker3-HICLAS and, on the other hand, Tucker3-HICLAS and Tucker2-HICLAS have been studied more in depth (Ceulemans, Van Mechelen, and Leenen, 2003; Ceulemans and Van Mechelen, 2004). In this paper, we will discuss the interrelations among all proposed hierarchical classes models for three-way three-mode binary data; the results will show that these models can be organized into a partially ordered hierarchy. With respect to model selection, up to now, only substantive arguments for applying a specific hierarchical classes model have been discussed. In this paper, we will consider model selection from a more formal point of view. In particular, we will present model selection strategies that are based on extensions of the well-known scree test (Cattell, 1966) and on the Akaike information criterion (AIC, see e.g., Bozdogan, 1987).

The remainder of this paper is organized as follows: In Section 2, some notation and terminology that will be used throughout the paper are given. Section 3 recapitulates the INDCLAS, Tucker3-HICLAS, and Tucker2-HICLAS models, presents the Tucker1-HICLAS model for three-way three-mode binary data, and discusses the interrelations of these models. In Section 4, three different model selection strategies are proposed for selecting among hierarchical classes solutions of different types and different ranks for a given data set. In Section 5, the performance of the proposed model selection strategies is evaluated in an extensive simulation study. Section 6 illustrates the model selection strategies with an application to interpersonal emotion data. In Section 7, the proposed hierarchy and model selection strategies are related to the corresponding work of Kiers (1991) for the family of principal component models for three-way three-mode real-valued data.

2. Preliminaries: Notation and Terminology

As the three-way hierarchical classes approach can be situated in the research areas of three-way analysis and Boolean matrix algebra, we introduce, in this section, some of the standard notation and terminology in these research areas that will be used throughout the paper. For more details, we refer to Kiers (2000) for three-way analysis and to Kim (1982) for Boolean matrix algebra.

2.1. Three-way Analysis

2.1.1. Vectors, Matrices, and Arrays

Following the suggestions by Kiers (2000), vectors, two-way matrices, and three-way arrays are respectively indicated by lowercase bold-face symbols (e.g., $\mathbf{x}$), uppercase bold-face symbols (e.g., $\mathbf{X}$), and underlined uppercase bold-face symbols (e.g., $\underline{\mathbf{X}}$). Subvectors and submatrices of two-way and three-way arrays are denoted by using subscript colons for the modes that are retained as a whole and ordinary indices for the elements with which the subarray is associated (e.g., $\mathbf{x}_{i:}$ and $\mathbf{x}_{:j}$ denote the i-th row vector and the j-th column vector of a two-way matrix $\mathbf{X}$).

2.1.2. Unit Superdiagonal Three-Way Three-Mode Array

An $I \times J \times K$ array $\underline{\mathbf{X}}$ is called unit superdiagonal if it holds that $I = J = K$ and $x_{ijk} = 1$ iff $i = j = k$ (Kiers, 2000).

2.1.3. Matricization of a Three-Way Three-Mode Array

Kiers (2000) defines the matricization of a three-way three-mode array as the transformation of a three-way three-mode array into a two-way two-mode matrix by collecting all pairs of elements of two modes of the array in one mode of the matrix. In particular, an $I \times J \times K$ array $\underline{\mathbf{X}}$ can be matricized in three different ways, which respectively yield (1) the $I \times JK$ matrix $\mathbf{X}_a$, (2) the $J \times KI$ matrix $\mathbf{X}_b$, and (3) the $K \times IJ$ matrix $\mathbf{X}_c$.

2.2. Boolean Algebra

2.2.1. Operators

The symbol $\bigoplus$ denotes the Boolean sum. Note that $1 \bigoplus 1 = 1$.

2.2.2. Schein or Decomposition Rank of a Boolean Matrix

The Schein rank of a Boolean matrix is its decomposition rank. In particular, the Schein rank of an $I \times J$ Boolean matrix $\mathbf{X}$ is the minimum number of cross-vectors whose Boolean sum is $\mathbf{X}$, where a cross-vector is defined as the $I \times J$ Boolean matrix $\mathbf{Y}$ that is obtained by calculating the outer product of an $I \times 1$ Boolean vector $\mathbf{a}$ and a $J \times 1$ Boolean vector $\mathbf{b}$ (i.e., $\forall i = 1 \ldots I,\ j = 1 \ldots J : y_{ij} = a_i b_j$). For example, if $\mathbf{X}$ has Schein rank three, then an $I \times 3$ Boolean matrix $\mathbf{A}$ and a $J \times 3$ Boolean matrix $\mathbf{B}$ exist such that $\forall i = 1 \ldots I,\ j = 1 \ldots J : x_{ij} = \bigoplus_{r=1}^{3} a_{ir} b_{jr}$.

It is important to note that the Schein rank of a Boolean matrix can be shown to be smaller than or equal to its row rank and its column rank (which, for Boolean matrices, may differ). Hence, it follows that the Schein rank of $\mathbf{X}$ is smaller than or equal to $\min(I, J)$, or, in other words, that $\mathbf{X}$ can always be decomposed into an $I \times R$ Boolean matrix $\mathbf{A}$ and a $J \times R$ Boolean matrix $\mathbf{B}$ such that $R \leq \min(I, J)$.

3. Interrelations of the Hierarchical Classes Models for 3-way 3-mode Binary Data

3.1. Models

Each hierarchical classes model for three-way three-mode binary data approximates an $I \times J \times K$ object by attribute by source binary data array $\underline{\mathbf{D}}$ by a binary reconstructed data or model array $\underline{\mathbf{M}}$ of the same size. Each model further includes a *Boolean decomposition* of $\underline{\mathbf{M}}$ into (1) up to three binary bundle matrices $\mathbf{A}$ ($I \times P$), $\mathbf{B}$ ($J \times Q$), and $\mathbf{C}$ ($K \times R$) that respectively reduce the objects, attributes, and sources to P, Q, and R overlapping clusters, called bundles, and (2) a three-way three-mode binary core array $\underline{\mathbf{G}}$ that defines a linking structure among the bundles of the reduced modes and, if applicable, the elements of the free modes. Furthermore, each of the models under consideration represents the *quasi-order relation* $\leq$ that is induced by $\underline{\mathbf{M}}$ on the elements of a reduced mode. In particular, if S^x denotes the set of pairs of elements of the other two modes that an element x of a reduced mode is associated with in $\underline{\mathbf{M}}$, it holds that $x \leq y$ iff $S^x \subseteq S^y$. These quasi-order relations are represented in the corresponding bundle matrices in terms of subset-superset relations among the bundle patterns (i.e., the set of bundles to which an element of a reduced mode belongs). The representation of these quasi-order relations implies a hierarchical classification of each reduced mode, in that elements with identical bundle patterns constitute classes, and in that two classes are hierarchically related if the bundle pattern of one of the two classes is a proper subset of the bundle pattern of the other class.

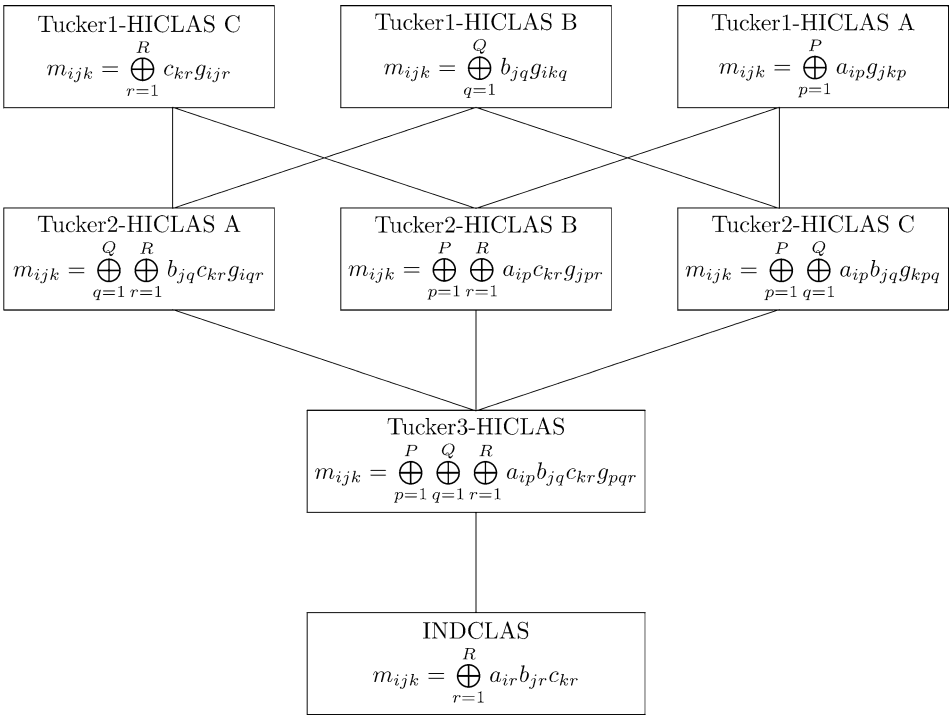


FIGURE 1.
Hierarchical relations among the eight types of hierarchical classes models for 3-way 3-mode binary data.

In this paper, we will consider eight different types of hierarchical classes models for three-way three-mode binary data. All eight are represented in Figure 1. In the following paragraphs, we will discuss their distinctive features.

3.1.1. The Three Types of Tucker1-HICLAS Models

A Tucker1-HICLAS model reduces only one of the three modes of $\underline{\mathbf{M}}$ to bundles. This implies that three different types of Tucker1-HICLAS models can be distinguished: type A that reduces the objects, type B that reduces the attributes, and type C that reduces the sources. Formally, the decomposition rules of the type A, type B, and type C Tucker1-HICLAS models can be stated as follows

$$\text{type A:} \qquad m_{ijk} = \bigoplus_{p=1}^{P} a_{ip} g_{jkp}, \tag{1}$$

$$\text{type B:} \qquad m_{ijk} = \bigoplus_{q=1}^{Q} b_{jq} g_{ikq}, \tag{2}$$

and

$$\text{type C:} \qquad m_{ijk} = \bigoplus_{r=1}^{R} c_{kr} g_{ijr}, \tag{3}$$

with P, Q, and R representing the rank of the respective models.

In addition to the quasi-order relation $\leq$ on the reduced mode, a Tucker1-HICLAS model represents in the core array $\underline{\mathbf{G}}$ the quasi-order relation $\leq$ that can be defined on the pairs of elements of the free modes: $(x,y) \leq (x',y')$ iff $\mathbf{g}_{xy:} \subseteq \mathbf{g}_{x'y':}$.

Note that a rank P type A Tucker1-HICLAS model with model array $\underline{\mathbf{M}}$ is equivalent to a rank P two-way two-mode HICLAS model (De Boeck and Rosenberg, 1988; Van Mechelen, De Boeck, and Rosenberg, 1995) for the matricized $I \times JK$ model array $\mathbf{M}_a$. Similarly, a rank Q type B Tucker1-HICLAS model (resp. a rank R type C Tucker1-HICLAS model) is equivalent to a rank Q (resp. rank R) HICLAS model for $\mathbf{M}_b$ (resp. $\mathbf{M}_c$).

3.1.2. The Three Types of Tucker2-HICLAS Models

A Tucker2-HICLAS model reduces two of the three modes of $\underline{\mathbf{M}}$ to bundles. Hence, three different types of Tucker2-HICLAS models can be considered: type A that reduces the attributes and the sources, type B that reduces the objects and the sources, and type C that reduces the objects and the attributes. The decomposition rules of these three types are given by

$$\text{type A:} \qquad m_{ijk} = \bigoplus_{q=1}^{Q} \bigoplus_{r=1}^{R} b_{jq} c_{kr} g_{iqr}, \tag{4}$$

$$\text{type B:} \qquad m_{ijk} = \bigoplus_{p=1}^{P} \bigoplus_{r=1}^{R} a_{ip} c_{kr} g_{jpr}, \tag{5}$$

and

$$\text{type C:} \qquad m_{ijk} = \bigoplus_{p=1}^{P} \bigoplus_{q=1}^{Q} a_{ip} b_{jq} g_{kpq}, \tag{6}$$

with (Q,R), (P,R), and (P,Q) representing the rank of the respective models.

Apart from the quasi-order relation $\leq$ on the reduced modes, a Tucker2-HICLAS model also represents the quasi-order relation that can be defined on the free mode, in terms of subset-superset relations between the core planes: $x \leq y$ iff $\mathbf{G}_{x::} \subseteq \mathbf{G}_{y::}$.

3.1.3. The Tucker3-HICLAS Model

A Tucker3-HICLAS model reduces each of the three modes of $\underline{\mathbf{M}}$ to bundles. Formally, the Tucker3-HICLAS decomposition rule is given by

$$m_{ijk} = \bigoplus_{p=1}^{P} \bigoplus_{q=1}^{Q} \bigoplus_{r=1}^{R} a_{ip} b_{jq} c_{kr} g_{pqr}, \tag{7}$$

with (P,Q,R) denoting the rank of the model.

3.1.4. The INDCLAS Model

An INDCLAS model reduces each of the three modes of $\underline{\mathbf{M}}$ to the same number of bundles and restricts the core array $\underline{\mathbf{G}}$ to a unit superdiagonal array, implying a one-to-one correspondence among the respective bundles. Hence, the INDCLAS decomposition rule can be stated as follows:

$$m_{ijk} = \bigoplus_{r=1}^{R} a_{ir} b_{jr} c_{kr}, \tag{8}$$

where R denotes the rank of the model.

3.2. Interrelations

The interrelations of all eight hierarchical classes models for three-way three-mode binary data can be described in terms of restrictiveness or constraints, with some models being more restrictive or constrained versions of other models. As such, in terms of restrictiveness, the models can be organized in a partially ordered hierarchy, which is depicted in Figure 1. Figure 1 is to be read as follows: If P and Q are greater than or equal to R, Model 1 is less or equally restrictive than Model 2 iff a downward path of lines exists from Model 1 to Model 2.

Starting from the top of Figure 1, it can be derived that the (Q,R) type A and the (P,R) type B Tucker2-HICLAS models are constrained versions of the rank R type C Tucker1-HICLAS model, with the constraints respectively implying that the $\mathbf{G}_b$ and $\mathbf{G}_a$ matricizations of the Tucker1-HICLAS core array $\underline{\mathbf{G}}$ have Schein (or decomposition) ranks Q and P. In other words, if $Q \geq \min(J, IR)$ and $P \geq \min(I, JR)$, the (Q,R) type A and the (P,R) type B Tucker2-HICLAS models boil down to the rank R type C Tucker1-HICLAS model (see Subsection 2.2.2). Similar relations hold between the (Q,R) type A and the (P,Q) type C Tucker2-HICLAS models and the rank Q type B Tucker1-HICLAS model, on the one hand, and between the (P,R) type B and the (P,Q) type C Tucker2-HICLAS models and the rank P type A Tucker1-HICLAS model, on the other hand.

Furthermore, a (P,Q,R) Tucker3-HICLAS model is more restrictive than a (Q,R) type A Tucker2-HICLAS model if P is smaller than the Schein rank of the $\mathbf{G}_a$ matricization of the Tucker2-HICLAS core array $\underline{\mathbf{G}}$. The latter implies that a (P,Q,R) Tucker3-HICLAS model boils down to a (Q,R) type A Tucker2-HICLAS model if $P \geq \min(I, QR)$. Again, similar relations hold between a (P,Q,R) Tucker3-HICLAS model and (P,R) type B and (P,Q) type C Tucker2-HICLAS models. Finally, a rank R INDCLAS model is a constrained (P,Q,R) Tucker3-HICLAS model, the constraint implying that (1) $P = Q = R$ and (2) the Tucker3-HICLAS core array $\underline{\mathbf{G}}$ is a unit superdiagonal array.

4. Model Selection

For the hierarchical classes analysis of a given three-way three-mode binary data set, algorithms associated with the distinct models of our hierarchy are available. More specifically, to fit the Tucker2-HICLAS, Tucker3-HICLAS, and INDCLAS models to data, one may make use of the algorithms proposed by Ceulemans and Van Mechelen (2004); Ceulemans, Van Mechelen, and Leenen (2003), and Leenen, Van Mechelen, De Boeck and Rosenberg (1999), respectively. Regarding the Tucker1-HICLAS model, we mentioned above that a rank P type A Tucker1-HICLAS model with model array $\underline{\mathbf{M}}$ is equivalent to a rank P HICLAS model for the matricized $I \times JK$ model array $\mathbf{M}_a$; the latter implies that type A Tucker1-HICLAS solutions may be obtained by applying the standard two-way HICLAS algorithm (Leenen and Van Mechelen, 2001) to the matricized $I \times JK$ data array $\mathbf{D}_a$. Similarly, applying this HICLAS algorithm to the $\mathbf{D}_b$ and $\mathbf{D}_c$

matricizations yields type B and C Tucker1-HICLAS solutions. Given a specific rank and a data array, all these hierarchical classes algorithms look for a model array of the same size, which has a minimal number of discrepancies d with the data

$$d = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} |d_{ijk} - m_{ijk}|, \tag{9}$$

and that can be represented by a hierarchical classes model of the pre-specified type and rank.

However, in practice, when applying hierarchical classes analysis to a given three-way three-mode binary data set, one typically faces a model selection problem. In particular, one often does not know which of the eight different types of hierarchical classes models will yield the most useful description of the data set, and, on top of that, each of the eight types of models may be fitted in several ranks. In this section, we will first consider model selection problems and possible solutions in general; subsequently, we will propose two types of solutions for the hierarchical classes model selection problem in particular.

4.1. Model Selection Problems and Possible Solutions

In general, one may distinguish between two types of model selection problems: problems of a nested and of a non-nested type. The nested type, unlike its non-nested counterpart, implies that all the models among which one has to choose can be totally ordered in terms of restrictiveness (i.e., for all pairs (x,y) of models under consideration x is a constrained version of y or vice versa).

Nested model selection problems are relatively easy to solve. In case of deterministic models, one usually compares the goodness of fit of the different models under consideration, while taking the differences in model restrictiveness into account. A well-known example is the scree test of Cattell (1966), which is used to obtain an optimal balance between the fit of principal component models and the number of components involved. In case of stochastic models, one may typically apply a likelihood ratio test (Wilks, 1938), which statistically tests whether a less restrictive model fits the data significantly better than a more restrictive model.

In general, non-nested model selection problems are more difficult to solve. In particular, to our knowledge, no comprehensive strategy has been proposed for directly selecting among non-nested deterministic models of different types and complexities. To be sure, some specific strategies have been developed for selecting among non-nested solutions of one particular model type (e.g., the model selection heuristics for the Tucker3-HICLAS model, Ceulemans, Van Mechelen, and Leenen, 2003, and the DIFFIT method for the Tucker3 model, Timmerman and Kiers, 2000). Furthermore, several authors have suggested to address the problem by inspecting deviance plots with a goodness-of-fit measure of the models under consideration on the Y-axis and a measure of their degrees of freedom on the X-axis (Fowlkes, Freeny, and Landwehr, 1988; Kroonenberg and Van der Voort, 1987; Kroonenberg and Oort, 2003). In such a plot, one selects the models with the best goodness-of-fit/degrees-of-freedom balance by retaining the models on or close by the lower boundary of the convex hull for further consideration. Optionally, one may further reduce the number of selected models by only retaining the models that are closest to an elbow in the lower boundary of the convex hull. Finally, to choose among non-nested stochastic models, one often makes use of information theory based criteria like AIC (Akaike, 1973; Bozdogan, 1987, 2000), which weigh the likelihood of the data under a specific model against the corresponding effective number of parameters (also called the number of free model parameters). In particular, the AIC-based model selection strategy boils down to calculating the AIC value of each model and retaining the model with the lowest value, where the AIC value is defined as:

$$-2\log p(\text{data}|\text{model}) + 2P, \tag{10}$$

with p(data|model) denoting the likelihood of the data given the model and P denoting the effective number of parameters.

4.2. *Two Types of Model Selection Strategies for the Family of Hierarchical Classes Models*

The hierarchical classes model selection problem that we consider in this section is of the non-nested type. More specifically, three kinds of non-nestedness are involved: (1) as discussed in Subsection 3.2, not all the eight types of hierarchical classes models are hierarchically interrelated, (2) the described interrelations only hold in case the respective models have equal corresponding numbers of bundles, and (3) Tucker3-HICLAS models of different ranks are not always nested, e.g., a (2,3,2) Tucker3-HICLAS model is neither less nor more restrictive than a (2,2,3) Tucker3-HICLAS model, and the same holds for Tucker2-HICLAS models of different ranks. Since the hierarchical classes models are deterministic models, it can be derived from the above overview that a solution for the hierarchical classes model selection problem is not readily available.

As a way out, we propose two possible solutions: (1) to extend the above mentioned convex hull approach, and (2) to make use of a minimal stochastic extension of the hierarchical classes models, which allows one to apply an AIC-based strategy. As described above, both the convex hull and AIC strategies take the effective number of parameters P into account. For hierarchical classes models, the determination of P is a complex, open problem, however. Therefore, we will use the best available estimate of P, that is, the raw number of model parameters p. Note that the latter is sometimes problematic, however, because in some cases, a less restrictive (i.e., hierarchically higher) model has fewer model parameters than a more restrictive (i.e., hierarchically lower) model. In this paper, we deal with this problem by discarding all models that include more model parameters than a less or equally restrictive model. Observe that the subset of to be discarded models can be determined beforehand, because it only depends on the size of the data set and on the range of the values for P, Q, and R.

4.2.1. *Extension of the Convex Hull Approach*

As a first solution to the hierarchical classes model selection problem, we propose to apply a variant of the well-known scree test to the hierarchical classes solutions with the best goodness-of-fit/degrees-of-freedom balance, that is, the solutions on the lower boundary of the convex hull (Kroonenberg and Van der Voort, 1987; Kroonenberg and Oort, 2003). To this end, all hierarchical classes solutions for a given data set are displayed in a deviance plot, with the number of discrepancies d (9) on the Y-axis and the number of model parameters p on the X-axis. Subsequently, a procedure to find the solutions on the lower boundary of the convex hull is applied to the deviance plot. Finally, a variant of the scree test is applied to these 'hull' solutions. In the following paragraphs, we present a 'hull boundary finding' procedure and propose two possible variants of the scree test.

The procedure to determine the subset of solutions that are on the lower boundary of the convex hull of the deviance plot starts by retaining for each observed value of p the solution with the lowest d only. Subsequently, a routine that consecutively considers all triplets of adjacent solutions is applied to the n retained solutions, sorted according to p and indicated by s_i ($i = 1 \ldots n$). This routine goes as follows: For the first triplet of solutions (s_1, s_2, s_3), determine whether or not s_2 is situated above the line connecting s_1 and s_3 in the deviance plot. If so, exclude s_2 from the subset of retained models. Next, do the same for the following triplet(s) of adjacent retained solutions, removing the first solution of the previous triplet. Continue until the last considered triplet includes s_n. Repeating this routine until no solution can be excluded anymore, yields the solutions on the lower boundary of the convex hull of the deviance plot.

We further propose two variants of the scree test to select the solution on the lower boundary of the convex hull with the best balance between d and p. The first rule, which we call DiffCH, selects the solution i that maximizes

$$\frac{d_{i-1} - d_i}{p_i - p_{i-1}} - \frac{d_i - d_{i+1}}{p_{i+1} - p_i}. \tag{11}$$

The second rule, called RatioCH, selects the solution i that maximizes

$$\frac{d_{i-1} - d_i}{p_i - p_{i-1}} \Big/ \frac{d_i - d_{i+1}}{p_{i+1} - p_i}. \tag{12}$$

Comparing (11) and (12), one may conclude that RatioCH gives less weight to large $\frac{d_x - d_y}{p_y - p_x}$ values than DiffCH.

4.2.2. Pseudo-AIC Strategy

As a second solution to the hierarchical classes model selection problem, we propose to transform these deterministic models into stochastic models by including an additional model parameter π that represents the probability that for arbitrary i, j, and k, $m_{ijk} \neq d_{ijk}$. Subsequently, one may make use of existing solutions for non-nested stochastic model selection problems. In particular, we propose to use a Pseudo-AIC model selection strategy (i.e., 'pseudo' because, as mentioned above, we use the number of model parameters p as an estimate of the effective number of parameters P). In order to calculate the Pseudo-AIC value of a hierarchical classes solution for a given three-way three-mode binary data set, the likelihood in the left part of (10) can be obtained by considering the likelihood of a specific data entry d_{ijk} given the corresponding model entry m_{ijk} and π:

$$p(d_{ijk}|m_{ijk}, \pi) = \pi^{|d_{ijk} - m_{ijk}|}(1 - \pi)^{1 - |d_{ijk} - m_{ijk}|}. \tag{13}$$

Assuming local independence, the likelihood of the full data set then boils down to

$$p(\underline{\mathbf{D}}|\text{deterministic solution}, \pi) = \pi^d (1 - \pi)^c, \tag{14}$$

with d denoting the number of discrepancies as defined by (9), and with c denoting the number of concordancies, that is:

$$c = IJK - d. \tag{15}$$

Regarding the stochastic extension, it should be noted that this extension is minimal, being based on a homogeneous error process governed by a single parameter. One could advocate the use of more sophisticated error processes. However, in this respect, we want to emphasize that simple error processes as the one outlined above are fairly standard in data analysis (e.g., the assumptions of idd and, therefore, homoscedastic residuals in regression analysis). Furthermore, the minimal stochastic extension has an important property that no longer holds for more sophisticated extensions: Maximizing the likelihood (14) is equivalent to minimizing the number of discrepancies d as defined in (9); this implies that the deterministic hierarchical classes algorithms yield maximum likelihood estimates of the stochastic model parameters, with, in addition, the value $\frac{d}{IJK}$ being the estimate for π.

5. Simulation Study

In this section, we present an extensive simulation study in which we evaluate to which extent the proposed model selection strategies succeed in indicating the type and the rank of the

hierarchical classes model that underlies a given three-way three-mode binary data array. The design of the simulation study is outlined in Subsection 5.1; subsequently, the results are presented and discussed in Subsections 5.2 and 5.3.

5.1. Design

In this simulation study, we distinguish between three different types of $I \times J \times K$ binary arrays: true arrays $\underline{\mathbf{T}}$, data arrays $\underline{\mathbf{D}}$, and model arrays $\underline{\mathbf{M}}$. A true array $\underline{\mathbf{T}}$ can be perfectly represented by a hierarchical classes model; the latter underlying true model is constructed by the simulation researcher. A data array $\underline{\mathbf{D}}$ is a true array $\underline{\mathbf{T}}$ perturbed with error. A model array $\underline{\mathbf{M}}$ can also be perfectly represented by one of the different types of hierarchical classes models in a specific rank, as it is obtained by subjecting $\underline{\mathbf{D}}$ to the associated type of hierarchical classes analysis in the respective rank.

In the simulation study, we used four true array types: in particular, type B Tucker1-HICLAS, type C Tucker2-HICLAS, Tucker3-HICLAS, and INDCLAS. Three parameters were further systematically varied:

1. the *Size*, $I \times J \times K$, of $\underline{\mathbf{T}}$, $\underline{\mathbf{D}}$, and $\underline{\mathbf{M}}$, at 2 levels: $15 \times 15 \times 15$, $30 \times 20 \times 10$;
2. the *True rank* of the hierarchical classes model for $\underline{\mathbf{T}}$; for the type B Tucker1-HICLAS model, 3 levels were used: 2, 3, 4; for the type C Tucker2-HICLAS model, 9 levels were used: (2,2), (2,3), (2,4), (3,2), (3,3), (3,4), (4,2), (4,3), (4,4); for the Tucker3-HICLAS model, 4 levels were used: (2,2,2), (2,3,3), (4,3,2), (4,4,4); for the INDCLAS model, 3 levels were used: 2, 3, 4.
3. the *Error level*, ε, which is the proportion of cells d_{ijk} differing from t_{ijk}, at 5 levels: .00, .05, .10, .20, .30.

The 600 type B Tucker1-HICLAS true arrays $\underline{\mathbf{T}}$ (20×2 (size) $\times$ 3 (true rank) $\times$ 5 (error level)) were constructed as follows: For each combination of size, true rank, and error level, 20 pairs of $J \times Q$ and $I \times K \times Q$ binary arrays $\mathbf{B}$ and $\underline{\mathbf{G}}$ were generated with entries that were independent realizations of a Bernoulli variable with the parameter value for the distribution chosen such that the expected proportion of ones in $\underline{\mathbf{T}}$ equals 0.5. Next, a data array $\underline{\mathbf{D}}$ was constructed from each true array $\underline{\mathbf{T}}$ by randomly altering the values of a proportion ε of the entries of $\underline{\mathbf{T}}$. The 1800 type C Tucker2-HICLAS (20×2 (size) $\times$ 9 (true rank) $\times$ 5 (error level)), 800 Tucker3-HICLAS (20×2 (size) $\times$ 4 (true rank) $\times$ 5 (error level)), and 600 INDCLAS (20×2 (size) $\times$ 3 (true rank) $\times$ 5 (error level)) true arrays $\underline{\mathbf{T}}$ and data arrays $\underline{\mathbf{D}}$, which were constructed similarly, were taken from the simulation studies that have been reported by Leenen, Van Mechelen, De Boeck and Rosenberg (1999), Ceulemans, Van Mechelen, and Leenen (2003), and Ceulemans and Van Mechelen (2004), respectively.

For each $15 \times 15 \times 15$ (resp. $30 \times 20 \times 10$) data array $\underline{\mathbf{D}}$, 147 (resp. 139) model arrays $\underline{\mathbf{M}}$ were obtained; in particular, 15 model arrays resulted from subjecting $\underline{\mathbf{D}}$ to a type A, B, and C Tucker1-HICLAS analysis in ranks 1–5 (with each type yielding five model arrays), 75 model arrays from subjecting $\underline{\mathbf{D}}$ to a type A, B, and C Tucker2-HICLAS analysis in ranks (1,1)–(5,5) (with each type yielding 25 model arrays), 52 (resp. 44) model arrays from analyzing $\underline{\mathbf{D}}$ with the Tucker3-HICLAS algorithm in ranks (1,1,1) to (5,5,5) (see Subsection 4.2), and 5 model arrays from analyzing $\underline{\mathbf{D}}$ with the INDCLAS algorithm in ranks 1 to 5[1]. Subsequently, the three model selection strategies were applied to these 147 (resp. 139) solutions.

[1]Note that the different algorithms were applied as described in the simulation studies reported by Leenen and Van Mechelen (2001); Ceulemans and Van Mechelen (2004); Ceulemans, Van Mechelen, and Leenen (2003), and Leenen, Van Mechelen, De Boeck and Rosenberg (1999), implying 2, 4, 12, and 6 rationally started runs for a Tucker1-HICLAS, Tucker2-HICLAS, Tucker3-HICLAS, and INDCLAS analysis, respectively (the runs differ in that the updating order of the component matrices and core array differs, e.g., $\mathbf{A} \rightarrow \underline{\mathbf{G}}$ or $\underline{\mathbf{G}} \rightarrow \mathbf{A}$).

5.2. Results

For each of the three considered model selection strategies, Table 1 presents how frequently the strategy retained the correct solution out of the 147 or 139 obtained solutions, that is, the hierarchical classes solution that is of the same model type and rank as the true model. From the overall correct selection percentages (i.e., across the four true model types) of respectively 35.9%, 56.8%, and 58.1%, it can be concluded that the RatioCH and Pseudo-AIC strategies perform much better than the DiffCH strategy. Except for the Tucker3-HICLAS data sets, the latter conclusion also holds if one considers the correct selection percentages for each of the four true model types separately. Furthermore, for the RatioCH and Pseudo-AIC strategies, the latter percentages show that correct selection is almost guaranteed in the case of the INDCLAS data sets, with about 95% correct selection, whereas the other types of model selection problems appear to be more difficult to solve, with percentages varying from 45% to 55%.

To further investigate the extent to which the performance of the RatioCH and Pseudo-AIC heuristics depends on the size of the data array, the rank of the underlying true model, and the error level, an analysis of variance was conducted for the data sets generated on the basis of each model type, with the frequency of correct selection as the dependent variable. These frequencies range from 0 to 20, because there were 20 replicates for each combination of size, true rank, and error level. Only considering effects accounting for at least 10% of the dependent variable (i.e., intraclass correlations $\hat{\rho}_I \geq .10$, Haggard, 1958; Kirk, 1982), the latter analyses of variance always yielded important main effects of error level: Except for (almost) error free data, it holds that the higher the error level, the harder it is for the model selection strategies to correctly estimate the type and rank of the true model (see Table 2; note that Table 3 shows exact 95% confidence intervals for 11 equally spaced entries of Table 2). Furthermore, for the RatioCH heuristic, a main effect of true rank was found in the case of the Tucker1-HICLAS and Tucker3-HICLAS data sets: The performance decreases with higher true rank. Finally, except for the Pseudo-AIC-INDCLAS case, true rank $\times$ error level interactions were found, indicating that the effect of error level increases with higher true rank. All these effects, including the results for (almost) error free data, parallel the effects found in the simulation studies set up to evaluate the performance of the different three-way hierarchical classes algorithms (Leenen and Van Mechelen, 2001; Ceulemans and Van Mechelen, 2004; Ceulemans, Van Mechelen, and Leenen, 2003; Leenen, Van Mechelen, De Boeck and Rosenberg, 1999). The latter suggests that the performance of the Pseudo-AIC and RatioCH heuristics depends on the performance of the algorithms.

5.3. Conclusion and Discussion

In general, it may be concluded that the two best heuristics, RatioCH and Pseudo-AIC, which select the same solution in 81% of the cases, yield an overall correct selection percentage of about

TABLE 1.
Percentage of correct selection as a function of the true model type.

	Selection heuristic		
True model type	DiffCH	RatioCH	Pseudo-AIC
Tucker1-HICLAS type B	5.3	46.8	53.7
Tucker2-HICLAS type C	28.2	52.7	54.5
Tucker3-HICLAS	47.3	45.4	45.4
INDCLAS	74	94.5	94.2
Overall	35.9	56.8	58.1

TABLE 2.
Percentage of correct selection at levels of true rank × error.

		Pseudo-AIC						RatioCH					
		Error level						Error level					
True model type	True rank	.00	.05	.10	.20	.30	Overall	.00	.05	.10	.20	.30	Overall
Tucker1-HICLAS	2	55.0	82.5	97.5	87.5	0.0	64.5	55.0	77.5	95.0	97.5	20.0	69.0
type B	3	82.5	90.0	92.5	32.5	0.0	59.5	85.0	95.0	95.0	77.5	5.0	71.5
	4	92.5	80.0	12.5	0.0	0.0	37.0	0.0	0.0	0.0	0.0	0.0	0.0
	Overall	76.7	84.2	67.5	40.0	0.0	53.7	46.7	57.5	63.3	58.3	8.3	46.8
Tucker2-HICLAS	(2,2)	17.5	67.5	77.5	82.5	70.0	63.0	17.5	67.5	77.5	82.5	70.0	63.0
type C	(2,3)	22.5	80.0	87.5	70.0	40.0	60.0	22.5	82.5	82.5	72.5	42.5	60.5
	(2,4)	30.0	72.5	67.5	45.0	15.0	46.0	22.5	65.0	65.0	40.0	30.0	44.5
	(3,2)	30.0	77.5	87.5	62.5	35.0	58.5	30.0	77.5	80.0	45.0	30.0	52.5
	(3,3)	62.5	77.5	92.5	80.0	25.0	67.5	60.0	72.5	90.0	67.5	32.5	64.5
	(3,4)	35.0	77.5	90.0	42.5	5.0	50.0	25.0	75.0	82.5	40.0	20.0	48.5
	(4,2)	32.5	67.5	75.0	52.5	15.0	48.5	25.0	57.5	70.0	55.0	25.0	46.5
	(4,3)	52.5	85.0	67.5	42.5	5.0	50.5	40.0	85.0	62.5	47.5	15.0	50.0
	(4,4)	52.5	87.5	67.5	25.0	0.0	46.5	42.5	82.5	70.0	22.5	2.5	44.0
	Overall	37.2	76.9	79.2	55.8	23.3	54.5	31.7	73.9	75.6	52.5	29.7	52.7
Tucker3-HICLAS	(2,2,2)	35.0	55.0	65.0	72.5	77.5	61.0	32.5	62.5	72.5	75.0	77.5	64.0
	(2,3,3)	27.5	60.0	67.5	42.5	17.5	43.0	27.5	57.5	70.0	40.0	17.5	42.5
	(4,3,2)	42.5	65.0	60.0	32.5	12.5	42.5	40.0	67.5	60.0	27.5	12.5	41.5
	(4,4,4)	55.0	62.5	35.0	20.0	2.5	35.0	55.0	57.5	32.5	20.0	2.5	33.5
	Overall	40.0	60.6	56.9	41.9	27.5	45.4	38.8	61.3	58.8	40.6	27.5	45.4
INDCLAS	2	92.5	100.0	100.0	100.0	97.5	98.0	92.5	100.0	100.0	100.0	97.5	98.0
	3	80.0	100.0	100.0	97.5	97.5	95.0	80.0	100.0	100.0	95.0	97.5	94.5
	4	75.0	100.0	100.0	100.0	72.5	89.5	75.0	95.0	100.0	100.0	85.0	91.0
	Overall	82.5	100.0	100.0	99.2	89.2	94.2	82.5	98.3	100.0	98.3	93.3	94.5

TABLE 3.
Exact 95% confidence intervals for 11 equally spaced entries of Table 2.

Percentage	Lower limit	Upper limit
0	0.0	8.8
10	2.8	23.7
20	9.1	35.7
30	16.6	46.5
40	24.9	56.7
50	33.8	66.2
60	43.3	75.1
70	53.5	83.4
80	64.4	91.0
90	76.3	97.2
100	91.2	100.0

57%. The lower overall correct selection percentage of DiffCH (about 36%) is due to the fact that DiffCH gives more weight to large $\frac{d_x - d_y}{p_y - p_x}$ values (see Subsection 4.2.1), which mostly occur in the left part of the deviance plots; the latter implies that DiffCH often selects very simple models. Considering the results more in detail, the correct selection percentages of RatioCH and Pseudo-AIC depend mostly on the true rank and the error level of a data set, with the effect of these two variables on correct selection paralleling their effect on goodness of fit and goodness of recovery. One may further wonder whether alternative selection procedures may do better (taking into account that a correct selection implies picking out a single solution out of a set of over 100 alternatives). At this point, one should note that it is not unusual for the hierarchical classes algorithms to end in a local minimum (e.g., Ceulemans, Van Mechelen, and Leenen, 2003), which may imply an upper bound for the performance of the selection procedures under study. In particular, if in case of an analysis based on the true model and the true rank the proportion of discrepancies of the obtained solution exceeds ε, this provides evidence for a local minimum. If such cases are removed from our percentage calculations, the overall correct selection percentage increases from about 57% to about 88%. The latter suggests that the performance of the proposed selection procedures is probably near-optimal. Note that Ceulemans and Van Mechelen (2004) indicated that local minima problems may be addressed by an effective but computationally intensive multistart procedure.

6. Illustrative Application

In this section, we will illustrate the presented model selection strategies with data from a longitudinal study on interpersonal emotions. In particular, a subject was asked which persons in her life best fitted six role descriptions, the role descriptions being mother, father, partner, and three unstable relationships. Subsequently, over a period of five months, the subject was asked to indicate every 15 days which of 40 emotions she currently experienced towards the 6 target persons. The resulting 10 time points $\times$ 40 emotions $\times$ 6 target persons binary data array $\underline{\mathbf{D}}$ (i.e., d_{ijk}=1 if target person k elicited emotion j at time point i; 0 otherwise) was analyzed with

1. the type A Tucker1-HICLAS algorithm in ranks 1–5
2. the type B Tucker1-HICLAS algorithm in ranks 1–5
3. the type C Tucker1-HICLAS algorithm in ranks 1–5

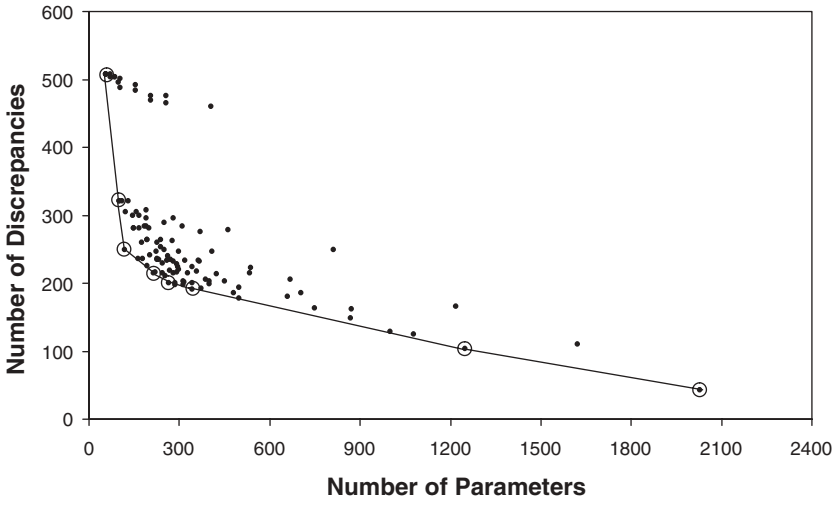


FIGURE 2.
Deviance plot of the 112 hierarchical classes solutions for the interpersonal emotion data, with the line representing the lower boundary of the convex hull and the circles indicating the eight 'hull' solutions.

4. the type A Tucker2-HICLAS algorithm in ranks (1,1)–(5,5)
5. the type B Tucker2-HICLAS algorithm in ranks (1,1) – (5,5)
6. the type C Tucker2-HICLAS algorithm in ranks (1,1)–(5,5)
7. the Tucker3-HICLAS algorithm in ranks (1,1,1) – (5,5,5)
8. the INDCLAS algorithm in ranks 1–5.

The RatioCH and Pseudo-AIC heuristics were applied to the resulting 112 models; note that 57 models could be omitted because they would include more model parameters than a less or equally restrictive model. In particular, Figure 2 displays a deviance plot of these 112 models. Applying the proposed procedure in Subsection 4.2.1 to find the solutions on the lower boundary of the convex hull resulted in the selection of eight 'hull' solutions, which are indicated by circles in Figure 2. The RatioCH values of these eight 'hull' solutions are given in Table 4. From this Table, one may conclude that the RatioCH heuristic indicates the selection of the (2,2,2) Tucker3-HICLAS model. Subsequently, the Pseudo-AIC values of all 112 solutions were calculated. From

TABLE 4.
Numbers of discrepancies d, numbers of parameters p, and RatioCH values of the eight interpersonal emotion solutions on the lower boundary of the convex hull of the deviance plot in Figure 2.

Model	Rank	d	p	RatioCH
INDCLAS	1	507	56	–
Tucker2-HICLAS type C	(1,2)	321	102	1.01
Tucker3-HICLAS	(2,2,2)	249	120	10.74
Tucker3-HICLAS	(4,3,3)	214	214	1.38
Tucker3-HICLAS	(4,4,3)	200	266	2.33
Tucker3-HICLAS	(4,5,4)	191	344	1.19
Tucker1-HICLAS type A	5	103	1250	1.79
Tucker1-HICLAS type C	5	42	2030	–

TABLE 5.
Numbers of discrepancies d, numbers of parameters p, and Pseudo-AIC values for the interpersonal emotion solutions with the five lowest Pseudo-AIC values.

Model	Rank	d	p	Pseudo-AIC
Tucker3-HICLAS	(2,2,2)	249	120	3130.93
Tucker3-HICLAS	(2,3,2)	236	164	3184.17
Tucker3-HICLAS	(3,3,2)	236	180	3216.17
Tucker3-HICLAS	(3,3,3)	225	195	3218.13
Tucker3-HICLAS	(4,3,3)	214	214	3225.04

Table 5, which displays the solutions with the five lowest Pseudo-AIC values, one may derive that applying the Pseudo-AIC heuristic also results in the selection of the (2,2,2) Tucker3-HICLAS model.

A graphical representation of the selected (2,2,2) Tucker3-HICLAS model (Ceulemans, Van Mechelen, and Leenen, 2003) can be found in Figure 3. Note that the upper part of this figure represents the target person hierarchy, whereas the hierarchy of the emotions is represented upside down in the lower part of the figure. The linking structure between the target persons, emotions, and time points is indicated by the lines and the hexagons that contain the time points, which connect the two hierarchies. Target person k then elicits emotion j at time point i if a downward path of lines exists from target person k to emotion j that goes via time point i. As such, Figure 3 can be given the following substantive interpretation: Mother, father, partner, and unstable relationships 2 and 3 are described in positive terms at all time points. At the first and last four time points, the same holds for unstable relationship 1; however, from time points 2 through 6 the latter relationship goes through a temporary crisis that involves a mix of negative and positive emotions.

7. Discussion

In this paper, we described the hierarchical interrelations among hierarchical classes models for three-way three-mode binary data, and we presented three model selection strategies for choosing among hierarchical classes solutions of different types and ranks for a specific data set. In this section, we will discuss this hierarchy and model selection strategy from a viewpoint inside as well as outside the family of hierarchical classes models; as to the latter, we will relate the presented work to the corresponding work by Kiers (1991) for the family of principal component models for three-way three-mode real-valued data.

7.1. Remarks from a Hierarchical Classes Point of View

7.1.1. Hierarchy

The presented partially ordered hierarchy of hierarchical classes models for three-way three-mode binary data is important, because through the hierarchy, one may obtain a more profound understanding of the interrelations between these models. In particular, from the hierarchy one may derive that it may be useful to analyze these interrelations in terms of constraints on the core array of the less restrictive model. Indeed, for the interrelations between models on adjacent hierarchy levels (described in Subsection 3.2), one may derive in a straightforward way that (1) a rank R type C Tucker1-HICLAS model is a (Q,R) type A Tucker2-HICLAS model if the $\mathbf{G}_b$ matricization of the Tucker1-HICLAS core array has Schein rank Q; (2) a (Q,R) type A

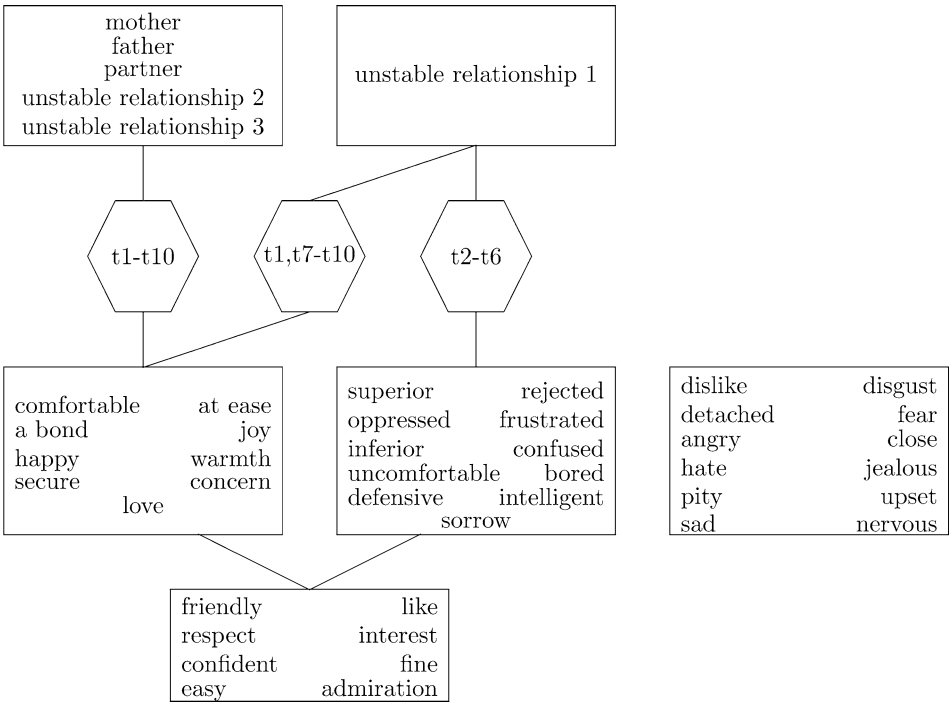


FIGURE 3.
Overall graphical representation of the (2,2,2) Tucker3-HICLAS model for the interpersonal emotion data.

Tucker2-HICLAS model is a (P,Q,R) Tucker3-HICLAS model if the $\mathbf{G}_a$ matricization of the Tucker2-HICLAS core array has Schein rank P; and (3) a (R,R,R) Tucker3-HICLAS model is a rank R INDCLAS model if the Tucker3-HICLAS core array is a unit superdiagonal array. Likewise, for pairs of models that are more than one hierarchy level apart, one may derive that: (1) a rank R type C Tucker1-HICLAS model is a (P,Q,R) Tucker3-HICLAS model if a (P,Q) type C Tucker2-HICLAS model exists for the Tucker1-HICLAS core array; (2) a rank R type C Tucker1-HICLAS model is a rank R INDCLAS model if an $I \times R$ binary matrix $\mathbf{G}^1$ and a $J \times R$ binary matrix $\mathbf{G}^2$ exist such that for each entry of the Tucker1-HICLAS core array holds that $g_{ijr} = g^1_{ir} g^2_{jr}$; and (3) a (R,R) type A Tucker2-HICLAS model is a rank R INDCLAS model if all core planes $\mathbf{G}_{i::}$ of the Tucker2-HICLAS core array are subsets of the identity matrix.

7.1.2. Model Selection

In this paper, model selection was considered from a formal point of view only, without taking into account substantive criteria, that is, which of the different types of hierarchical classes models are relevant for a specific substantive question. However, it is important to note that, as generally holds in data analysis, in practice substantive and formal criteria for hierarchical classes model selection are to be combined: formal criteria, like the Pseudo-AIC value of different hierarchical classes solutions, may be used to choose within the proper subset of hierarchical classes models that are of substantive relevance for the application at hand.

To illustrate the latter, we consider the case of psychiatric diagnosis research. In such research, one may wish to fit hierarchical classes models to a psychiatric patient $\times$ symptom $\times$ clinician data array to grasp how clinicians' symptom judgements come about. From a theoretical point of view, it has been suggested that the latter judgements may be understood by considering the

cognitive structures and processes that underly the symptom judgements (Van Mechelen, 1991; Van Mechelen and De Boeck, 1989). An example of such a cognitive structure is an implicit taxonomy of syndromes; such a taxonomy may be defined in terms of a small number of (possibly overlapping) symptom clusters. As the symptom bundle matrix of a hierarchical classes model would constitute the formalization of such a taxonomy of syndromes, one may therefore decide to only fit those hierarchical classes models that include a symptom bundle matrix (i.e., the type B Tucker1-HICLAS, the type A and C Tucker2-HICLAS, the Tucker3-HICLAS, and the INDCLAS models).

7.2. Relation to Work by Kiers (1991) for the Family of Principal Component Models

The family of hierarchical classes models for three-way three-mode binary data is closely related to the family of principal component models for three-way three-mode real-valued data (see, e.g., Kroonenberg, 1983). More precisely, PARAFAC/CANDECOMP, Tuckals-3, Tuckals-2, and PCA-SUP differ in three respects only from INDCLAS, Tucker3-HICLAS, Tucker2-HICLAS, and Tucker1-HICLAS, respectively: (1) the component matrices and core array of the hierarchical classes models are restricted to be binary, (2) hierarchical classes models are based on Boolean algebra, whereas principal component models involve standard algebra (note that replacing the Boolean sum $\bigoplus$ in the different hierarchical classes decomposition rules by the regular sum $\sum$ yields the respective principal component decomposition rules), and (3) hierarchical classes models represent quasi-order relations among the elements of one or more modes of the model array.

Given this close relationship, in this subsection, we will compare the hierarchy and model selection strategies for hierarchical classes models as proposed in the present paper with corresponding work by Kiers (1991) for principal component models. While making this comparison, an overall difference between Kiers' and our approach should however be taken into account: whereas in the present paper the modeling of each of the three modes is considered equally important, Kiers (1991) describes and selects different principal component models in terms of their adequacy to model the differences between the elements of the third mode; this option is crucial as it influences both his construction of a hierarchy (which does not include some existing types of principal component models) and his model selection strategy (which is of the nested type).

7.2.1. Hierarchy

Kiers (1991) distinguishes between seven different principal component models for three-way three-mode real-valued data. More specifically, he considers: (a) the PCA-SUP, Tuckals-3, and PARAFAC/ CANDECOMP models that were mentioned above; (b) three different types of ORTCP models, which are PARAFAC/CANDECOMP models with orthogonality constraints on the component matrix of the first and/or second mode; and (c) the SUMPCA model, which is an ORTCP model that restricts the component scores of all elements of the third mode to be equal. All these seven models are depicted in Figure 4. From Figure 4, it may easily be derived that the interrelations of these seven models, as indicated by the lines that interconnect them, can all be described in terms of restrictiveness: in particular, the interrelations between the PCA-SUP, the Tuckals-3, and the PARAFAC/CANDECOMP models can be understood in terms of restrictions on the core array; the interrelations between the PARAFAC/CANDECOMP, the ORTCP, and the SUMPCA models, in their turn, all involve restrictions on one or more component matrices of the PARAFAC/CANDECOMP model.

Comparing Kiers' hierarchy with the hierarchy of hierarchical classes models as described in the present paper, it is clear that the two hierarchies, although both based on restrictiveness

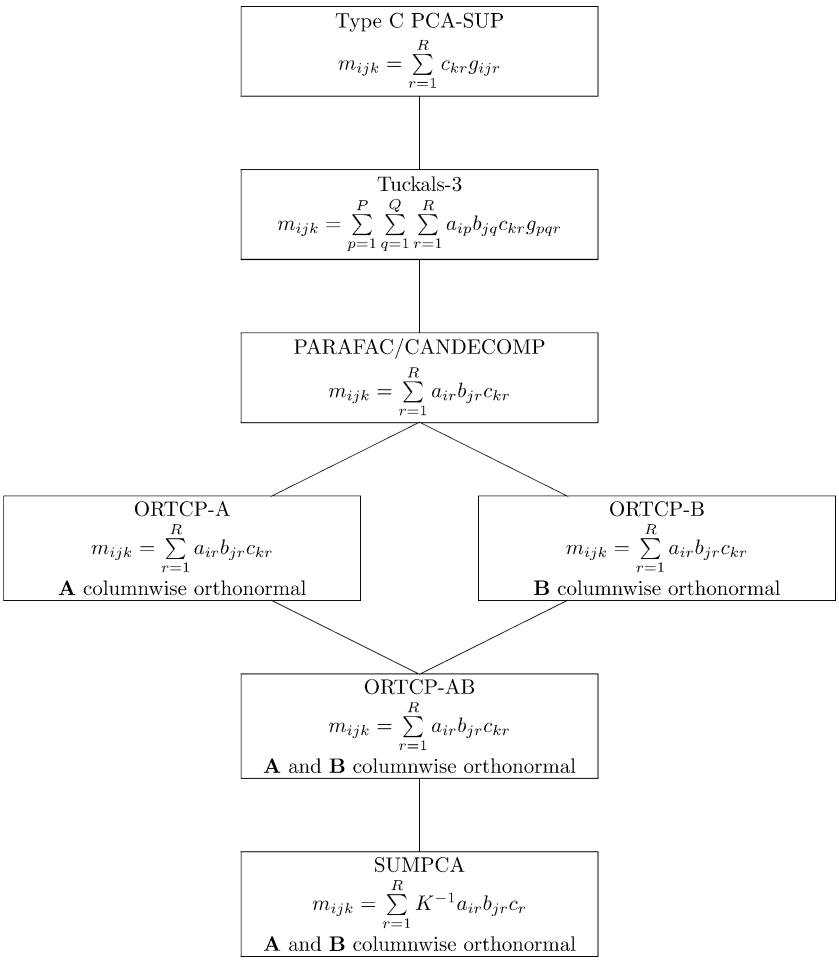


FIGURE 4.
Kiers' (1991) hierarchy of principal component models for 3-way 3-mode real-valued data.

relations, differ in two main respects. Firstly, whereas the hierarchy of hierarchical classes models constitutes a partially ordered hierarchy, the hierarchy of the principal component models is almost totally ordered. The latter is due to the fact that Kiers, unlike us, focuses on the modeling of the third mode, which implies that only one type of PCA-SUP model – the type C model – and none of the three types of Tuckals-2 models is included in his hierarchy. As such, Kiers' hierarchy would obviously also be partially ordered if all three types of PCA-SUP and Tuckals-2 models were to be included, yielding a graphical representation similar to Figure 1. Secondly, not all real-valued PCA models have counterparts in our hierarchy. More in particular, as the orthonormality concept from linear algebra has no direct analogue in Boolean algebra, the principal component models with orthonormality constraints on the component matrices have no hierarchical classes counterpart.

7.2.2. Model Selection

Given that the hierarchy in Figure 4 almost takes the form of a total order, Kiers (1991) proposed the following model selection strategy: first, one determines the number of principal

components R needed to adequately describe the relations between the elements of the third mode; to this end, one should fit type C PCA-SUP solutions with different numbers of components to the matricized data array, and, subsequently, one should determine the number of components R that yields the most useful description by means of the scree test (Cattell, 1966). Next, the other types of principal component models are fitted to the data with the number of components for the third mode equal to R. Finally, one retains the model for which it holds that fitting a less restrictive model does not yield an important gain in fit to the data.

Compared to the hierarchical classes model selection strategies that were presented in the present paper, Kiers' strategy has the advantage that it yields a model selection problem of the nested type, which therefore, in principle, is easier to solve. On the other hand, Kiers' strategy also leaves some issues unresolved. For example, no formal criterion is given for deciding whether or not fitting a less restrictive model yields an important gain in fit to the data. Furthermore, it must also be noted that the final result of Kiers' model selection strategy completely depends on the initial choice of R. Hence, it may be noted that (counterparts of) the model selection strategies proposed in the present paper might also be useful for selecting among different types and ranks of principal component solutions, especially if Kiers' hierarchy is extended to include all three types of PCA-SUP and Tuckals-2 models.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B.N. Petrov & F. Csaki (Eds.), *Second International Symposium on Information Theory* (pp. 267–281). Budapest: Academiai Kiado.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*, 345–370.

Bozdogan, H. (2000). Akaike's information criterion and recent developments in informational complexity. *Journal of Mathematical Psychology*, *44*, 62–91.

Cattell, R.B. (1966). The meaning and strategic use of factor analysis. In R.B. Cattell (Ed.), *Handbook of Multivariate Experimental Psychology* (pp. 174–243). Chicago: Rand McNally.

Ceulemans, E., & Van Mechelen, I. (2004). Tucker2 hierarchical classes analysis. *Psychometrika*, *69*, 413–433.

Ceulemans, E., Van Mechelen, I., & Leenen, I. (2003). Tucker3 hierarchical classes analysis. *Psychometrika*, *68*, 413–433.

De Boeck, P., & Rosenberg, S. (1988). Hierarchical classes: Model and data analysis. *Psychometrika*, *53*, 361–381.

Fowlkes, E.B., Freeny, A.E., & Landwehr, J.M. (1988). Evaluating logistic models for large contingency tables. *Journal of the American Statistical Association*, *83*, 611–622.

Haggard, E.A. (1958). *Intraclass Correlation and the Analysis of Variance.* New York: Dryden.

Kiers, H.A.L. (1991). Hierarchical relations among three-way methods. *Psychometrika*, *56*, 449–470.

Kiers, H.A.L. (2000). Towards a standardized notation and terminology in multiway analysis. *Journal of Chemometrics*, *14*, 105–122.

Kim, K.H. (1982). *Boolean Matrix Theory.* New York: Marcel Dekker.

Kirk, R.E. (1982). *Experimental design: Procedures for the behavioral sciences (second edition).* Belmont, CA: Brooks/Cole.

Kroonenberg, P.M. (1983). *Three-mode Principal Component Analysis: Theory and Applications.* Leiden: DSWO.

Kroonenberg, P.M., & Oort, F.J. (2003). Three-mode analysis of multimode covariance matrices. *British Journal of Mathematical and Statistical Psychology*, *56*, 305–336.

Kroonenberg, P.M., & Van der Voort, T. H.A. (1987). Multiplicatieve decompositie van interacties bij oordelen over de werkelijkheidswaarde van televisiefilms [Multiplicative decomposition of interactions for judgements of realism of television films]. *Kwantitatieve Methoden*, *8*, 117–144.

Kuppens, P., Van Mechelen, I., Smits, D. J.M., De Boeck, P., & Ceulemans, E. (2005). *Individual differences in appraisal and emotion: The case of anger and irritation,* submitted.

Leenen, I., & Van Mechelen, I. (2001). An evaluation of two algorithms for hierarchical classes analysis. *Journal of Classification*, *18*, 57–80.

Leenen, I., Van Mechelen, I., De Boeck, P., & Rosenberg, S. (1999). INDCLAS: A three-way hierarchical classes model. *Psychometrika*, *64*, 9–24.

Timmerman, M.E., & Kiers, H.A.L. (2000). Three-mode principal components analysis: Choosing the numbers of components and sensitivity to local optima. *British Journal of Mathematical and Statistical Psychology*, *53*, 1–16.

Van Mechelen, I. (1991). Symptom and diagnosis inference based on implicit theories of psychopathology: A review. *Cahiers de Psychologie Cognitive*, *11*, 155–171.

Van Mechelen, I., & De Boeck, P. (1989). Implicit taxonomy in psychiatric diagnosis: A case study. *Journal of Social and Clinical Psychology*, *8*, 276–287.

Van Mechelen, I., De Boeck, P., & Rosenberg, S. (1995). The conjunctive model of hierarchical classes. *Psychometrika*, *60*, 505–521.

Vansteelandt, K., & Van Mechelen, I. (1998). Individual differences in situation-behavior profiles: A triple typology model. *Journal of Personality and Social Psychology*, *75*, 751–765.

Wilks, S.S. (1938). The large sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, *9*, 60–62.

Manuscript received 13 MAR 2003
Final version received 08 DEC 2003