PSYCHOMETRIKA—VOL. 73, NO. 4, 729–751
DECEMBER 2008
DOI: 10.1007/S11336-008-9069-9

THE CHIC MODEL: A GLOBAL MODEL FOR COUPLED BINARY DATA

TOM WILDERJANS, EVA CEULEMANS, AND IVEN VAN MECHELEN
KATHOLIEKE UNIVERSITEIT LEUVEN

Often problems result in the collection of coupled data, which consist of different N-way N-mode data blocks that have one or more modes in common. To reveal the structure underlying such data, an integrated modeling strategy, with a single set of parameters for the common mode(s), that is estimated based on the information in all data blocks, may be most appropriate. Such a strategy implies a global model, consisting of different N-way N-mode submodels, and a global loss function that is a (weighted) sum of the partial loss functions associated with the different submodels. In this paper, such a global model for an integrated analysis of a three-way three-mode binary data array and a two-way two-mode binary data matrix that have one mode in common is presented. A simulated annealing algorithm to estimate the model parameters is described and evaluated in a simulation study. An application of the model to real psychological data is discussed.

Key words: coupled data, three-way three-mode data, binary data, hierarchical classes, multi-way data analysis, clustering, data fusion.

1. Introduction

In different fields of research, coupled data sets, consisting of different N-way N-mode data blocks that have one or more modes in common are collected. In such cases, functional relations between the distinctive data blocks often are of key interest to the researcher. For example, in contextualized personality psychology one often wants to investigate how individual differences in specific behaviors in specific situations are related to individual differences in some kind of dispositions. To study this, two data sets are collected: (1) a three-way three-mode person by situation by behavior data array that denotes the extent to which each member of a group of persons displays each behavior from a set of behaviors in each situation from a set of situations and (2) a two-way two-mode person by disposition data matrix. Note that both data blocks have one mode, that is, the person mode, in common.

To disclose the (true) structure underlying a coupled data set, one may represent each N-way N-mode data block by a multi-way decomposition model, with the parameters for each common mode being the same for the decomposition models to which that common mode belongs. To represent the coupled data set in this way, two strategies may be followed (for some examples of applications in which one of both strategies has been used, see, e.g., Coxe, 1986; Boqué & Smilde, 1999; Ten Berge, 1986; Smilde, Westerhuis, & Boqué, 2000). The first strategy consists of two consecutive steps: First, a multi-way model is fitted to the data block that is of primary interest; second, the quantification of the common mode as obtained in the first step is used in the analysis of the other data block(s). In the example above, first the person by situation by behavior data array may be analyzed by means of a three-way three-mode decomposition model. In a second step, a two-way two-mode model may be fitted to the person

T. Wilderjans is a Research Assistant of the Fund for Scientific Research—Flanders (Belgium). The research reported in this paper was partially supported by the Research Council of K.U. Leuven (GOA/2005/04). We are grateful to Kristof Vansteelandt for providing us with an interesting data set. We also thank three anonymous reviewers for their useful comments.

Requests for reprints should be sent to Tom Wilderjans, Department of Psychology, Katholieke Universiteit Leuven, Tiensestraat 102, 3000 Leuven, Belgium. E-mail: Tom.Wilderjans@psy.kuleuven.be

© 2008 The Psychometric Society

by disposition data matrix, the parameters for the common person mode being taken from the first analysis. Note that this strategy implies different loss functions, one for each submodel, that are optimized separately. In the second strategy, a global model, which may consist of different submodels—one for each data block—is fitted to the coupled data set at once. As a consequence, when estimating the parameters for the common mode(s), the information in all data blocks to which that common mode belongs is used. For the example, this implies that the (common) person parameters are estimated on the basis of both the person by situation by behavior data array and the person by disposition data matrix. An integrated strategy implies the optimization of a global loss function, which optionally can be a sum of separate (partial) loss functions, one for each submodel. Note that also a weighted sum of separate loss functions can be used where the weights denote the degree to which each data block is to influence the estimation of the parameters of the common mode(s). In the remainder, the first strategy will be called segmented because different loss functions are optimized separately, while the second one will be called integrated because one global loss function is optimized.[1]

Depending on the research context, each of the two strategies, as explained above, may be appropriate. As such, a segmented strategy may be preferred when the (categorical or dimensional) structure underlying one of the data blocks is of main interest, the other data block(s) being only used to facilitate the interpretation of the main latent structure. For example, when the person by situation by behavior data array is considered to be of primary importance, this data array may be used to induce a latent structure for persons, situations, and behaviors as well as their association; subsequently, the person by disposition data matrix may be only used to further characterize the latent person structure as obtained in the first step, with the person by disposition data matrix not playing any role in the determination of that person structure. A segmented strategy may also be preferred when the structure underlying the common mode(s) is not the same for the different data blocks to which this mode(s) belongs to and when primary interest lies in one of the data blocks. For example, when there is no reason to assume that the individual differences structure underlying the person by situation by behavior data array is the same as the individual differences structure underlying the person by disposition data matrix and when the former structure is of primary interest, a latent structure for the persons can be induced by only taking the former data array into account.

In general, a disadvantage of a segmented strategy is that only information from one data block is used when the structure for the common mode(s) is derived. As a result, error in this data block may have too large an influence on the derived structure for the common mode (which may further also result in representation errors for the other—noncommon—modes). An integrated strategy is most appropriate when the latent structure underlying the common mode(s) in the data is of primary interest. The information in the different data blocks then is used to derive the latent structure for the common mode(s). In this way, different sources of information, which may possibly highlight different aspects of the common mode(s), are integrated (*data fusion*). For example, when in the case of the personality data the latent person structure is of primary interest, the person by situation by behavior data array and the person by disposition data matrix may be both used to determine the individual differences structure of the persons. In general, an integrated strategy may compensate for the disadvantages of a segmented strategy and may lead to more stable and correct inferences about the structure of the common mode(s).

In this paper, we want to develop a global model for an integrated analysis of coupled data sets that consist of one binary two-way two-mode data matrix and one binary three-way three-mode data array that have one mode in common (in Fig. 1, a graphical presentation of the structure of such a coupled data set can be found). More precisely, we want to construct a global

[1]The term segmented is preferred over the term sequential because the former is reserved to denote the type of loss function that is used, while the latter refers to the procedure to optimize the loss function. Note that an integrated loss function may be optimized sequentially.

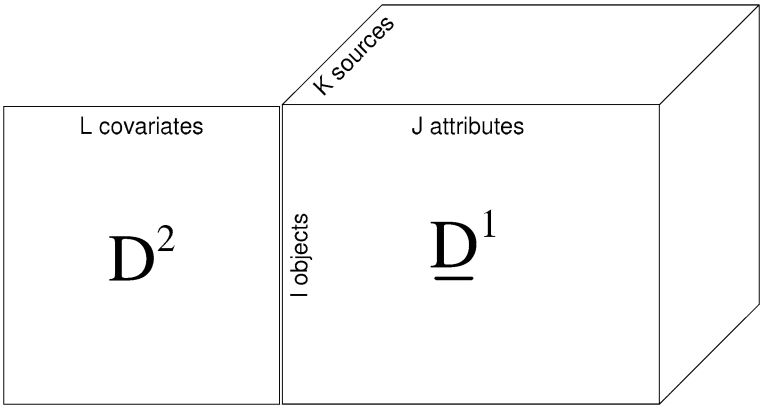


FIGURE 1.
Structure of the coupled data set.

model that consists of two submodels, both being members of the family of hierarchical classes (HICLAS) models (De Boeck & Rosenberg, 1988; Leenen, Van Mechelen, De Boeck, & Rosenberg, 1999; Ceulemans, Van Mechelen, & Leenen, 2003) for binary N-way N-mode data. All hierarchical classes models reduce each mode to a limited number of binary variables, called bundles (or components) that imply an overlapping clustering of the elements of the different modes. Being a member of the family of HICLAS models, the new **c**ombined Hiclas–Indclas model for **c**oupled binary data, denoted by the acronym CHIC, represents the three-way three-mode binary data array by an INDCLAS model (Leenen et al., 1999) and the two-way two-mode binary data matrix by a HICLAS model (De Boeck & Rosenberg, 1988). Both submodels are linked (coupled) to each other by imposing the restriction that the overlapping clustering for the common mode should be the same in both submodels.

The remainder of this paper is organized in five main sections: In Section 2, first the HICLAS and the INDCLAS model are reviewed briefly, and next the new CHIC model is presented. In Section 3, a simulated annealing (SA) algorithm to estimate the parameters of the new model is described, together with some heuristics for rank selection. In Section 4, the performance of the SA algorithm is evaluated in a simulation study. An illustrative application of the CHIC model to real psychological data from the personality domain is presented in Section 5. Section 6 finally contains some concluding remarks.

2. Model

2.1. *The HICLAS Model*

A two-way two-mode HICLAS model approximates an $I \times J$ object by attribute binary data matrix $\mathbf{D}$ by an $I \times J$ binary model matrix $\mathbf{M}$ that can be decomposed into an $I \times P$ binary matrix $\mathbf{A}$ and a $J \times P$ binary matrix $\mathbf{B}$, where P denotes the rank of the model. The P columns of $\mathbf{A}$ and $\mathbf{B}$ define P binary variables that are called object bundles and attribute bundles, respectively; consequently, the matrices $\mathbf{A}$ and $\mathbf{B}$ are called the object and attribute bundle matrix. The HICLAS model represents three types of structural relations in $\mathbf{M}$.

Association The association relation is the binary relation between the objects and the attributes as defined by the 1-entries in the model matrix $\mathbf{M}$. The HICLAS model represents the

association relation between the objects and the attributes by the following decomposition rule:

$$m_{ij} = \bigoplus_{p=1}^{P} a_{ip} b_{jp}, \tag{1}$$

where $\bigoplus$ denotes a Boolean sum. Note that (1) implies a one-to-one correspondence between the respective object and attribute bundles.

Equivalence Two equivalence relations are defined on **M**: one for the object mode and one for the attribute mode. In particular, object i is equivalent to object i' in **M** iff both objects are associated with the same set of attributes. Equivalent objects constitute an object class, with those classes implying a partition of the objects. The HICLAS model represents the equivalence relation among the objects by assigning in **A** the same set of bundles (i.e., identical bundle patterns) to equivalent objects. The equivalence relation among the attributes, represented in **B**, is defined similarly.

Hierarchy Two hierarchical relations are defined; one on each mode of **M**. An object i is hierarchically below an object i' in **M** iff the set of attributes associated with object i is a strict (proper) subset of the set of attributes associated with object i'. Note that the hierarchical relation among the objects implies a hierarchical relation among the object classes. The HICLAS model represents the hierarchical relation among the objects by strict subset-superset relations among their bundle patterns in **A**. The hierarchical relation among the attributes is defined similarly and is represented in **B**.

2.2. The INDCLAS Model

A three-way three-mode INDCLAS model approximates an $I \times J \times K$ object by attribute by source binary data array $\underline{\mathbf{D}}$ by an $I \times J \times K$ binary model array $\underline{\mathbf{M}}$ that can be decomposed into an $I \times P$ binary object bundle matrix **A**, a $J \times P$ binary attribute bundle matrix **B**, and a $K \times P$ binary source bundle matrix **C**, where P denotes the rank of the model. The P columns of **A**, **B**, and **C** define P (possibly overlapping) bundles of objects, attributes, and sources, respectively. As the HICLAS model, the INDCLAS model represents three types of structural relations in $\underline{\mathbf{M}}$.

Association An INDCLAS model represents the ternary relation among the objects, attributes, and sources, as defined by the 1-entries in $\underline{\mathbf{M}}$, by the following decomposition rule:

$$m_{ijk} = \bigoplus_{p=1}^{P} a_{ip} b_{jp} c_{kp}. \tag{2}$$

Note that (2) implies a one-to-one correspondence among the respective object, attribute, and source bundles.

Equivalence Three equivalence relations are defined; one on each mode of $\underline{\mathbf{M}}$. An element i is equivalent to an element i' in $\underline{\mathbf{M}}$ iff both elements are associated with the same set of pairs of elements of the other two modes. The INDCLAS model represents the equivalence relation in that equivalent elements have an identical bundle pattern in the respective bundle matrices.

Hierarchy A hierarchical relation is defined on each of the three modes of $\underline{\mathbf{M}}$. An element(class) i is hierarchically below an element(class) i' in $\underline{\mathbf{M}}$ iff the set of pairs of elements of the other two modes that is associated with the first element(class) is a strict subset of the set of pairs of elements of the other two modes that is associated with the second element(class). The INDCLAS model represents the hierarchical relation among the elements of one mode in terms of strict subset-superset relations among the associated bundle patterns.

2.3. A Global Hierarchical Classes Model for Coupled Binary Data: CHIC

2.3.1. Model The CHIC model approximates an $I \times J \times K$ object by attribute by source binary data array $\underline{\mathbf{D}}^1$ and an $I \times L$ object by covariate binary data matrix $\mathbf{D}^2$ by an $I \times J \times K$ binary model array $\underline{\mathbf{M}}^1$ and an $I \times L$ binary model matrix $\mathbf{M}^2$, where (1) $\underline{\mathbf{M}}^1$ and $\mathbf{M}^2$ can be decomposed into an INDCLAS model and a HICLAS model of rank P, respectively, and (2) the common object bundle matrix $\mathbf{A}$ is the same in both models. Note that, without loss of generality, the object mode is considered the common mode. Note further that the term "covariate", which denotes the distinct mode of $\mathbf{M}^2$, is not to be understood as exogenous, as in the CHIC model both data blocks $\underline{\mathbf{D}}^1$ and $\mathbf{D}^2$ are used to derive the structure underlying the common mode. An important feature of the CHIC model is that it includes a single structure for the common mode in terms of bundles, which makes it possible to relate the underlying structure for $\underline{\mathbf{D}}^1$ to the underlying structure for $\mathbf{D}^2$. This implies that the common objects (object types) can be characterized both in terms of (1) associated attribute-source (-type) pairs and (2) associated covariates (covariate types).

Like the HICLAS and the INDCLAS models, the CHIC model represents three types of structural relations in $\underline{\mathbf{M}}^1$ and $\mathbf{M}^2$. Below, we will successively discuss the representation of each of these three relations more in detail. For this purpose, we will make use of the hypothetical three-way three-mode model array $\underline{\mathbf{M}}^1$ and the hypothetical two-way two-mode model matrix $\mathbf{M}^2$ in Tables 1 and 2 as a guiding example; a CHIC model in rank 3 for $\underline{\mathbf{M}}^1$ and $\mathbf{M}^2$ is presented in Table 3.

Association The CHIC model represents the ternary relation among the objects, attributes, and sources, and the binary relation between the objects and the covariates, by the following decomposition rules, respectively:

$$m^1_{ijk} = \bigoplus_{p=1}^{P} a_{ip} b^1_{jp} c_{kp}, \tag{3}$$

$$m^2_{il} = \bigoplus_{p=1}^{P} a_{ip} b^2_{lp}, \tag{4}$$

where P denotes the rank of the CHIC model. These decomposition rules are equivalent to:

$$m^1_{ijk} = 1 \iff \exists p : a_{ip} = 1 \wedge b^1_{jp} = 1 \wedge c_{kp} = 1, \tag{5}$$

$$m^2_{il} = 1 \iff \exists p : a_{ip} = 1 \wedge b^2_{lp} = 1. \tag{6}$$

The CHIC decomposition rules imply that (1) an object i is associated with an attribute j and a source k in $\underline{\mathbf{M}}^1$ ($m^1_{ijk} = 1$) iff there is at least one bundle to which object i, attribute j, and source k belong, and (2) an object i is associated with a covariate l in $\mathbf{M}^2$ ($m^2_{il} = 1$) iff there exists at least one bundle to which object i and covariate l belong. For example, one can derive from the CHIC model in Table 3 that Object 8, Attribute 4, and Source 1 are associated in $\underline{\mathbf{M}}^1$, because

TABLE 1.
Hypothetical three-way three-mode model matrix $\underline{\mathbf{M}}^1$.

Source 1 Objects	Attributes					Source 2 Objects	Attributes					Source 3 Objects	Attributes				
	A_1	A_2	A_3	A_4	A_5		A_1	A_2	A_3	A_4	A_5		A_1	A_2	A_3	A_4	A_5
O_1	1	1	1	1	1	O_1	1	1	0	1	1	O_1	1	1	1	1	1
O_2	1	1	1	1	1	O_2	1	1	0	1	1	O_2	1	1	1	1	1
O_3	0	0	0	0	0	O_3	1	0	0	0	0	O_3	1	0	0	0	0
O_4	1	1	0	1	1	O_4	1	1	0	1	1	O_4	1	1	0	1	1
O_5	0	0	1	1	0	O_5	0	0	0	0	0	O_5	0	0	1	1	0
O_6	1	1	1	1	1	O_6	1	1	0	1	1	O_6	1	1	1	1	1
O_7	0	0	1	1	0	O_7	1	0	0	0	0	O_7	1	0	1	1	0
O_8	1	1	0	1	1	O_8	1	1	0	1	1	O_8	1	1	0	1	1

TABLE 2.
Hypothetical two-way two-mode model matrix $\mathbf{M}^2$.

Objects	Covariates							
	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
O_1	1	1	1	1	0	1	1	1
O_2	1	1	1	1	0	1	1	1
O_3	0	1	0	0	1	0	1	1
O_4	0	0	1	1	0	0	1	1
O_5	1	1	1	0	0	1	0	1
O_6	1	1	1	1	1	1	1	1
O_7	1	1	1	0	1	1	1	1
O_8	0	1	1	1	1	0	1	1

TABLE 3.
CHIC model in rank 3 for the model matrices $\underline{\mathbf{M}}^1$ and $\mathbf{M}^2$ in Tables 1 and 2.

	A				**B**^{1}				**C**				**B**^{2}			
	Obj B_1	Obj B_2	Obj B_3		Attr B_1	Attr B_2	Attr B_3		Sour B_1	Sour B_2	Sour B_3		Cov B_1	Cov B_2	Cov B_3
O_1	1	1	0	A_1	0	1	1	S_1	1	1	0	C_1	1	0	0
O_2	1	1	0	A_2	0	1	0	S_2	0	1	1	C_2	1	0	1
O_3	0	0	1	A_3	1	0	0	S_3	1	1	1	C_3	1	1	0
O_4	0	1	0	A_4	1	1	0					C_4	0	1	0
O_5	1	0	0	A_5	0	1	0					C_5	0	0	1
O_6	1	1	1									C_6	1	0	0
O_7	1	0	1									C_7	0	1	1
O_8	0	1	1									C_8	1	1	1

all three elements belong to the second bundle (i.e., ObjB_2, AttrB_2, and SourB_2). Further, Object 8 and Covariate 5 are associated in $\mathbf{M}^2$, because both elements belong to the third bundle (i.e., ObjB_3 and CovB_3). Note that (3)–(6) imply a one-to-one correspondence among the respective object, attribute, source, and covariate bundles.

Equivalence On each mode of $\underline{\mathbf{M}}^1$ and $\mathbf{M}^2$ an equivalence relation is defined. An object i is equivalent to an object i' iff both objects are associated with the same set of pairs of attributes and sources in $\underline{\mathbf{M}}^1$ *and* with the same set of covariates in $\mathbf{M}^2$. The equivalence relation for the attributes, sources, and covariates is defined the same as before. The CHIC model represents the

equivalence relation among the objects, attributes, sources, and covariates in terms of identical bundle patterns in $\mathbf{A}$, $\mathbf{B}^1$, $\mathbf{C}$, and $\mathbf{B}^2$, respectively. For example, in Tables 1 and 2, one can see that Object 1 is equivalent to Object 2 both in $\underline{\mathbf{M}}^1$ and $\mathbf{M}^2$; hence, both objects have an identical bundle pattern in Table 3.

Hierarchy On each mode of $\underline{\mathbf{M}}^1$ and $\mathbf{M}^2$ a hierarchical relation is defined. An object i is hierarchically below an object i' iff (1) in $\underline{\mathbf{M}}^1$ the set of pairs of attributes and sources that is associated with object i, is a subset of the set of pairs of attributes and sources that is associated with object i', (2) in $\mathbf{M}^2$ the set of covariates that is associated with object i, is a subset of the set of covariates that is associated with object i', and (3) at least one of the two subset-superset relations as mentioned above is a strict one. The hierarchical relation for the attributes, sources, and covariates is defined the same as before. The CHIC model represents the hierarchical relation among the objects, attributes, sources, and covariates (classes) by strict subset-superset relations among the bundle patterns in the bundle matrices $\mathbf{A}$, $\mathbf{B}^1$, $\mathbf{C}$, and $\mathbf{B}^2$, respectively. For example, one can see in Tables 1 and 2 that Object 3 is hierarchically below Object 7 both in $\underline{\mathbf{M}}^1$ and $\mathbf{M}^2$; hence, the bundle pattern of Object 3 is a strict subset of the bundle pattern of Object 7 in Table 3.

2.3.2. Graphical Representation An overall graphical representation of the CHIC model can be given that links the graphical representation of the INDCLAS model for $\underline{\mathbf{M}}^1$ (in the lower part of the figure) to the graphical representation of the HICLAS model for $\mathbf{M}^2$ (in the upper part of the figure) by the representation of the underlying bundles of the common (object) mode. As a consequence, a double characterization of the common objects is obtained in terms of (1) attribute-source (type) pairs, and (2) covariates (types). Further, the overall graphical representation of the CHIC model also accounts for the three structural relations in the model.

As an example, in Fig. 2, one may consider the overall graphical representation of the CHIC model in Table 3. The hierarchical classifications of the covariates and the attributes are presented at the top and the bottom of Fig. 2, respectively, the hierarchical classification of the attributes being displayed upside down. Classes of equivalent elements are indicated by boxes that enclose the labels of the elements belonging to that class. Note that when a bundle-specific class (i.e., a class of elements that belong to one bundle one) does not contain any elements, this is indicated by a shaded box and that empty classes that are not bundle-specific are omitted from the graphical representation. The hierarchical relation among the elements (classes) of one mode is represented by straight lines that connect the corresponding boxes. Both classifications are linked to each other by the graphical representation of the object bundles, represented by rectangles (i.e., each rectangle corresponds to one object bundle), and the source bundles, represented by hexagons (i.e., each hexagon corresponds to one source bundle); each rectangle (hexagon) further contains the labels of the objects (sources) that belong to the corresponding object (source) bundle.

The association relation between the objects and the covariates is indicated by zigzags that connect corresponding bundle-specific object and covariate classes. The association relation among the objects, attributes, and sources is indicated by zigzags that connect corresponding bundle-specific object and attribute classes, and that include a hexagon, containing the (classes of) sources that belong to the corresponding bundle. The double characterization of the objects in terms of attribute-source (type) pairs and covariates (types) then can be derived immediately from Fig. 2 as follows: An object i can be characterized in terms of (1) an attribute-source pair $(j-k)$, and (2) a covariate l, iff there is a path, consisting of straight lines and zigzags, that connects covariate l with attribute j and that contains object i in a rectangle and source k in a hexagon.

For example, in Fig. 2, one can see that Object 2 can be characterized in terms of (1) Attribute-Source pair (5–3), and (2) Covariate 7, because there is a path between Covariate 7 and Attribute 5 that contains Object 2 and Source 3. Object 3, however, can not be characterized in terms of (1) Attribute-Source pair (3–1), and (2) Covariate 4, because no path exists between

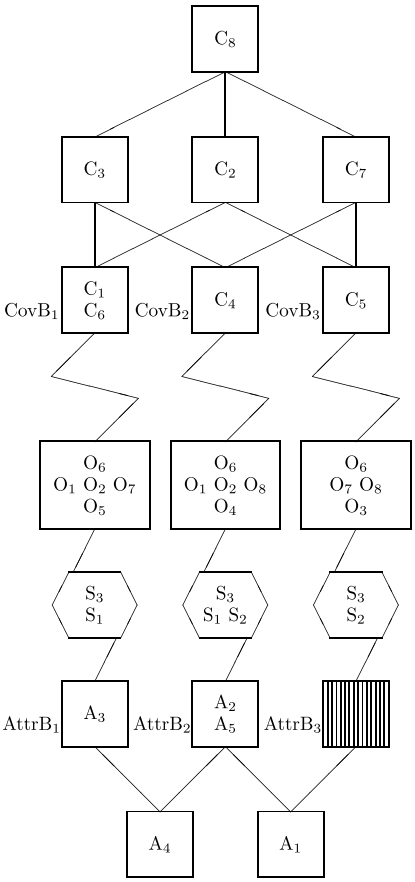


FIGURE 2.
Overall graphical representation of the CHIC model in Table 3.

Covariate 4 and Attribute 3 that contains Object 3 and Source 1. Note that the equivalence and hierarchical relations between the objects and the sources cannot be readily derived from Fig. 2; as such they are displayed separately in Figs. 3 and 4, respectively.

3. Data Analysis

3.1. Aim

The aim of a CHIC analysis in rank P of an $I \times J \times K$ binary data array $\underline{\mathbf{D}}^1$ and an $I \times L$ binary data matrix $\mathbf{D}^2$ is to estimate an $I \times J \times K$ binary model array $\underline{\mathbf{M}}^1$ and an $I \times L$ binary model matrix $\mathbf{M}^2$ such that (1) the value of the loss function

$$f\left(\underline{\mathbf{M}}^1, \mathbf{M}^2\right) = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(d^1_{ijk} - m^1_{ijk}\right)^2 + \sum_{i=1}^{I}\sum_{l=1}^{L}\left(d^2_{il} - m^2_{il}\right)^2 \tag{7}$$

is minimized, and (2) $\underline{\mathbf{M}}^1$ and $\mathbf{M}^2$ can be represented by a rank P INDCLAS model and a rank P HICLAS model, respectively, with a common object bundle matrix $\mathbf{A}$.

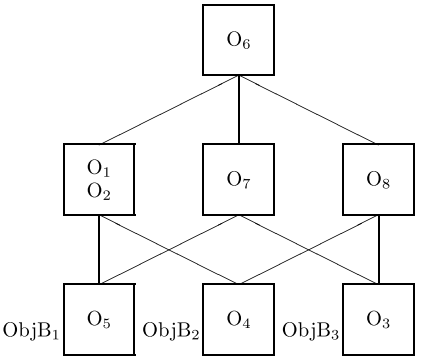


FIGURE 3.
Object hierarchy of the CHIC model in Table 3.

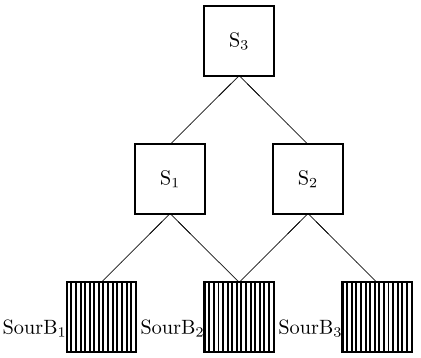


FIGURE 4.
Source hierarchy of the CHIC model in Table 3.

3.2. *Algorithm*

Given an $I \times J \times K$ binary data array $\underline{\mathbf{D}}^1$, an $I \times L$ binary data matrix $\mathbf{D}^2$, and a rank P, the CHIC algorithm goes through two consecutive phases. In the first phase, the bundle matrices $\mathbf{A}$, $\mathbf{B}^1$, $\mathbf{C}$, and $\mathbf{B}^2$ are estimated such as to minimize loss function (7). Since at the end of the first phase the bundle matrices are restricted to represent only the association relation in ($\underline{\mathbf{M}}^1$, $\mathbf{M}^2$), in the second phase, the bundle matrices are adjusted such as to represent the equivalence and hierarchical relations correctly; this can be done without altering the loss function.

First Phase To estimate the optimal bundle matrices $\mathbf{A}$, $\mathbf{B}^1$, $\mathbf{C}$, and $\mathbf{B}^2$, a simulated annealing (SA) algorithm is used. Although being more computational intensive, simulated annealing was preferred over alternating least squares (ALS), the original method to fit models of the HICLAS family. This is because in a recent study of Ceulemans, Van Mechelen, and Leenen (2007), in which SA and ALS were compared for the two-way HICLAS and the three-way TUCKER3-HICLAS models, it appeared that SA outperformed ALS in terms of minimizing the loss function as well as recovering the underlying truth. Moreover, it was shown that these differences in performance can only be attributed in part to differences in computation time.

A general description of the simulated annealing algorithm can be found in Appendix. In the CHIC algorithm, a SA chain is implemented as follows: First, an initial solution is generated by replacing the columns of the bundle matrices $\mathbf{A}$, $\mathbf{B}^1$, $\mathbf{C}$, and $\mathbf{B}^2$ by vectors of objects, attributes,

sources, and covariates, respectively, that are chosen from the data at random. Second, the initial temperature is obtained by running a subchain of solutions and accepting all solutions, and by subsequently dividing the average increase in the loss function f across those links in which worse solutions are accepted, by ln(0.8):

$$T_{\text{initial}} = \frac{\text{Average loss function increase of the worse solutions}}{\ln(0.8)}. \tag{8}$$

Third, a candidate solution is obtained from the current solution by changing the value of one randomly chosen parameter (from zero to one or from one to zero), with each parameter of the CHIC model having an equal probability of being changed, and the parameters of the CHIC model being the (binary) entries of the component matrices $\mathbf{A}$, $\mathbf{B}^1$, $\mathbf{C}$, and $\mathbf{B}^2$; as such, for each pair of solutions of the solution space (with the solution space being defined as the Cartesian product: $\{0,1\}^{I\times P}\times\{0,1\}^{J\times P}\times\{0,1\}^{K\times P}\times\{0,1\}^{L\times P}$) there exists a sequence of solutions, starting from one solution of the pair and ending with the other one, for which each solution in the sequence is a member of the neighborhood of the preceding one. Fourth, a worse candidate solution is accepted if:

$$p < \exp\bigl((f_{\text{current}} - f_{\text{candidate}})/T_{\text{current}}\bigr), \tag{9}$$

where p is a number generated from a uniform(0,1) distribution, T_{current} is the current temperature, and f_{current} and $f_{\text{candidate}}$ are the loss function values of the current and the candidate solution, respectively. Fifth, a subchain stops (1) if a maximum number of $((I+J+K+L)\times 2^P)\times 5$ solutions have been generated, or (2) if 10% of this maximum number have been accepted. Sixth, at the end of each subchain, the temperature is decreased as follows:

$$T_{\text{current}} = 0.9 \times T_{\text{current}}. \tag{10}$$

Seven, a SA chain stops when (1) the current temperature becomes smaller than 0.000001, or (2) for the last five subchains, there is no change in the loss function value f of the last accepted solution in each subchain.

Second Phase In the second phase, the bundle matrices obtained at the end of the first phase are transformed so as to represent the equivalence and hierarchical relations in $\underline{\mathbf{M}}^1$ and $\mathbf{M}^2$. This is accomplished by performing a closure operation (Barbut & Monjardet, 1970; Birkhoff, 1940). In particular, each 0-entry in $\mathbf{A}$, $\mathbf{B}^1$, $\mathbf{C}$, and $\mathbf{B}^2$ is changed to 1 iff this modification does not alter $\underline{\mathbf{M}}^1$ and $\mathbf{M}^2$.

3.3. Rank Selection

To determine the (true) rank P of the CHIC model underlying a particular coupled data set, different rank selection heuristics may be used that aim at finding a rank P CHIC model with a good balance between on the one hand the fit of the model to the data, quantified by the value on the loss function (7), and on the other hand the complexity of the model. Below, four such heuristics are presented. For all heuristics, it is assumed that CHIC analyses in ranks 1 up to R have been performed on a coupled data set at hand. In the first heuristic, which may be called an "elbow criterion" or "scree test" (Cattell, 1966), one searches for an elbow in the scree plot, in which the number of discrepancies between ($\underline{\mathbf{D}}^1$, $\mathbf{D}^2$) and ($\underline{\mathbf{M}}^1$, $\mathbf{M}^2$) is plotted against the rank of the model. This elbow can be determined visually or by computing for each rank P the difference between on the one hand the improvement in the loss function value going from rank $P-1$ to rank P, and on the other hand the improvement going from rank P to rank $P+1$, and selecting that rank P for which this difference is maximized. A second possible heuristic

is a "generalized scree test", which selects that rank P model that maximizes the difference between the mean improvement in the loss function value going from rank 1 to rank P, and the mean improvement going from rank P to rank R (Leenen & Van Mechelen, 2001; Ceulemans et al., 2003). The third heuristic is based on a so-called pseudo-binomial test (McKenzie et al., 1992). According to this heuristic, the model is selected with the lowest rank P for which the number of discrepancies of the model in rank $P + 1$ exceeds the 10th percentile of the binomial distribution Bin(n, Loss), with n denoting the total number of entries in both data blocks and Loss the proportion of discrepancies for the model in rank P (Leenen & Van Mechelen, 2001). Finally, a pseudo-AIC rank selection heuristic (Ceulemans & Van Mechelen, 2005) could be considered, which selects the model with the lowest value on the Akaike information criterion (Akaike, 1973; Bozdogan, 1987, 2000). This heuristic requires a minimal stochastic extension of the CHIC model, which can be obtained by including an additional model parameter π that denotes the probability that for an arbitrary entry of one of the two data blocks the model value differs from the data value.

4. Simulation Study

4.1. Problem

In this section, a simulation study is presented to evaluate the performance of the CHIC algorithm. Two aspects of the algorithm are of main interest: (1) goodness-of-fit, and (2) goodness-of-recovery of the true association, equivalence, and hierarchical relations. With respect to the second aspect, the primary interest lies in the common mode, with focus on the performance of the integrated modeling strategy relative to the segmented one. To clarify this, three types of binary $I \times J \times K$ and $I \times L$ array-matrix couples must be distinguished: (1) the true array-matrix couple ($\underline{\mathbf{T}}^1$, $\mathbf{T}^2$) that can be represented by a CHIC model of rank P, which can be decomposed into $\mathbf{A}^{(T)}$, $\mathbf{B}^{1(T)}$, $\mathbf{C}^{(T)}$, and $\mathbf{B}^{2(T)}$; (2) the data array-matrix couple ($\underline{\mathbf{D}}^1$, $\mathbf{D}^2$) that is obtained by adding noise to ($\underline{\mathbf{T}}^1$, $\mathbf{T}^2$); and (3) the model array-matrix couple ($\underline{\mathbf{M}}^1$, $\mathbf{M}^2$) that is obtained by applying the CHIC algorithm to the data array-matrix couple ($\underline{\mathbf{D}}^1$, $\mathbf{D}^2$) and that can be represented by a CHIC model of rank P, which can be decomposed into $\mathbf{A}^{(M)}$, $\mathbf{B}^{1(M)}$, $\mathbf{C}^{(M)}$, and $\mathbf{B}^{2(M)}$.

Goodness-of-Fit To address the question whether the CHIC algorithm succeeds in finding the global optimum of the loss function (7), the data ($\underline{\mathbf{D}}^1$, $\mathbf{D}^2$) are compared to the model ($\underline{\mathbf{M}}^1$, $\mathbf{M}^2$) as obtained from the algorithm. To this end, a *badness-of-fit* (BOF) statistic is used that is defined as the proportion of discrepancies between ($\underline{\mathbf{D}}^1$, $\mathbf{D}^2$) and ($\underline{\mathbf{M}}^1$, $\mathbf{M}^2$). As ($\underline{\mathbf{T}}^1$, $\mathbf{T}^2$) always is a valid rank P candidate model for the data, the proportion of discrepancies between ($\underline{\mathbf{T}}^1$, $\mathbf{T}^2$) and ($\underline{\mathbf{D}}^1$, $\mathbf{D}^2$) (denoted as *badness-of-data*, or BOD) can be considered an upper bound for the BOF of the global optimum in rank P. Therefore, BOF exceeding BOD implies that the algorithm ended in a suboptimal solution. This, however, does not mean that BOD exceeding BOF implies that the algorithm found the global optimum, since there may exist a couple ($\underline{\mathbf{M}}^{*1}$, $\mathbf{M}^{*2}$) that is closer to ($\underline{\mathbf{D}}^1$, $\mathbf{D}^2$) than ($\underline{\mathbf{M}}^1$, $\mathbf{M}^2$) is to ($\underline{\mathbf{D}}^1$, $\mathbf{D}^2$). Note that for error free data ($\underline{\mathbf{D}}^1$, $\mathbf{D}^2$) equals ($\underline{\mathbf{T}}^1$, $\mathbf{T}^2$) and ($\underline{\mathbf{T}}^1$, $\mathbf{T}^2$) is globally optimal. As a consequence, in that case, BOF also denotes the degree to which ($\underline{\mathbf{M}}^1$, $\mathbf{M}^2$) differs from the global optimum, whereas in general, BOF only measures the degree to which ($\underline{\mathbf{M}}^1$, $\mathbf{M}^2$) differs from ($\underline{\mathbf{D}}^1$, $\mathbf{D}^2$). This implies that BOF values for data sets with error cannot be readily compared with their counterparts for error free data sets.

Goodness-of-Recovery Studying the goodness-of-recovery implies the question to what extent the data-analytic procedure is capable of uncovering the truth underlying the data. With

respect to this truth, a distinction has to be drawn between the true reconstructed data entries (i.e., the true array-matrix couple) and the true parameters (i.e., the true bundle matrices) underlying the data. The former is studied by comparing the entries of the obtained model array-matrix couple ($\underline{\mathbf{M}}^1, \mathbf{M}^2$) to the entries of the true array-matrix couple ($\underline{\mathbf{T}}^1, \mathbf{T}^2$), while the latter is examined by comparing the obtained model parameters (in $\mathbf{A}^{(M)}$, $\mathbf{B}^{1^{(M)}}$, $\mathbf{C}^{(M)}$, and $\mathbf{B}^{2^{(M)}}$) to the true parameters (in $\mathbf{A}^{(T)}$, $\mathbf{B}^{1^{(T)}}$, $\mathbf{C}^{(T)}$, and $\mathbf{B}^{2^{(T)}}$) in terms of the implied equivalence and hierarchical relations. With respect to the latter two relations, the question may be raised whether the structure of the common mode is better recovered than the structure underlying the noncommon modes. Special attention is further to be paid to the question which data-analytic strategy, the integrated or the segmented one, is superior in revealing the underlying truth. Note that this last question can be studied with respect to all three relations.

In the remainder, first, in Section 4.2, the design of the simulation study is outlined. Next, the simulation results are presented in Section 4.3, drawing a distinction between goodness-of-fit (4.3.1) and goodness-of-recovery of the true association, equivalence, and hierarchical relations, with special attention to the recovery of the structure of the common mode (4.3.2). Section 4.4 finally contains some concluding remarks with respect to the simulation study.

4.2. Design and Procedure

Four factors were systematically manipulated in a completely randomized four-factorial design, all factors considered random:

(a) the *Dominance in size*, d, of the common mode over the other modes of the three-way array $\underline{\mathbf{T}}^1$ ($\underline{\mathbf{D}}^1$, $\underline{\mathbf{M}}^1$). This factor was manipulated at three levels: one level with the common mode being the largest mode of the three-way array ($50 \times 20 \times 27$, denoting 50 elements for the common mode and 20 and 27 elements for the other noncommon modes), one level with the common mode being as large as the other modes ($30 \times 30 \times 30$), and one level with the common mode being the smallest mode ($20 \times 50 \times 27$). Note that the size of $\underline{\mathbf{T}}^1$ ($\underline{\mathbf{D}}^1$, $\underline{\mathbf{M}}^1$) was kept constant at 27,000 entries;

(b) the *Array/total ratio*, r, of the size of $\underline{\mathbf{T}}^1$ ($\underline{\mathbf{D}}^1$, $\underline{\mathbf{M}}^1$) to the total size $\underline{\mathbf{T}}^1 + \mathbf{T}^2$ ($\underline{\mathbf{D}}^1 + \mathbf{D}^2$, $\underline{\mathbf{M}}^1 + \mathbf{M}^2$, respectively):

$$r = \frac{I \times J \times K}{(I \times J \times K) + (I \times L)}. \tag{11}$$

This factor was manipulated at four levels: 0.50, 0.90, 0.95, 0.99, implying that the following sizes for the two-way matrix $\mathbf{T}^2$ ($\mathbf{D}^2$, $\mathbf{M}^2$) were used: for $r = 0.50$ (50×540, 30×900, and 20×1350), for $r = 0.90$ (50×60, 30×100, and 20×150), for $r = 0.95$ (50×30, 30×50, and 20×75), and for $r = 0.99$ (50×6, 30×10, and 20×15);

(c) the *True rank*, P, of the CHIC model for ($\underline{\mathbf{T}}^1$, $\mathbf{T}^2$), manipulated at four levels: 2, 3, 4, 5;

(d) the *Error level*, ε, defined as the expected proportion of cells of ($\underline{\mathbf{D}}^1$, $\mathbf{D}^2$) differing from the corresponding cells of ($\underline{\mathbf{T}}^1$, $\mathbf{T}^2$):

$$\varepsilon = E\left[\frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(t^1_{ijk} - d^1_{ijk})^2 + \sum_{i=1}^{I}\sum_{l=1}^{L}(t^2_{il} - d^2_{il})^2}{IJK + IL}\right]. \tag{12}$$

This factor was manipulated at four levels: 0.00, 0.10, 0.20, 0.30.

For each combination of a Dominance in size, d, an Array/total ratio, r, a True rank, P, and an Error level, ε, true bundle matrices $\mathbf{A}^{(T)}$, $\mathbf{B}^{1^{(T)}}$, $\mathbf{C}^{(T)}$, and $\mathbf{B}^{2^{(T)}}$ were constructed by independently drawing entries from a Bernoulli distribution with parameter value 0.50 (implying that for each mode the elements are discrete uniformly distributed over the different element

classes), under the constraint that all bundle-specific classes (i.e., a class of elements belonging to one bundle only) were nonempty. The constraint was imposed to ensure that the true array-matrix couple ($\underline{\mathbf{T}}^1$, $\mathbf{T}^2$), obtained by combining the true matrices $\mathbf{A}^{(T)}$, $\mathbf{B}^{1(T)}$, $\mathbf{C}^{(T)}$, and $\mathbf{B}^{2(T)}$ by the INDCLAS and the HICLAS decomposition rules (3) and (4), could not be perfectly represented by a CHIC model of a lower rank than the True rank. Further, for each true array-matrix couple ($\underline{\mathbf{T}}^1$, $\mathbf{T}^2$), a data array-matrix couple ($\underline{\mathbf{D}}^1$, $\mathbf{D}^2$) was constructed by changing the value in each cell of ($\underline{\mathbf{T}}^1$, $\mathbf{T}^2$) with a probability ε. The whole data generation procedure was repeated 20 times to obtain 20 replications per cell of the design. As a consequence, 20 (replications) $\times$ 3 (Dominance in size) $\times$ 4 (Array/total ratio) $\times$ 4 (True rank) $\times$ 4 (Error level) = 3,840 different array-matrix couples ($\underline{\mathbf{T}}^1$, $\mathbf{T}^2$) and ($\underline{\mathbf{D}}^1$, $\mathbf{D}^2$) were obtained. Subsequently, one CHIC analysis (with 50 SA chains) in the true rank P was performed on each data array-matrix couple ($\underline{\mathbf{D}}^1$, $\mathbf{D}^2$). In addition, in order to compare the integrated to the segmented modeling strategy in terms of the recovery of the true structure underlying the common mode, each data array $\underline{\mathbf{D}}^1$ was subjected to one INDCLAS analysis and each data matrix $\mathbf{D}^2$ to one HICLAS analysis, both in the true rank P. Note that these additional analyses were performed by using the CHIC algorithm (with 50 SA chains) in which the loss function was replaced by the INDCLAS and HICLAS loss function, respectively.

4.3. Results

4.3.1. Goodness-of-Fit To evaluate the degree to which the CHIC algorithm succeeds in minimizing loss function (7), the following *badness-of-fit* statistic (BOF), taking values between 0 (i.e., perfect fit) and 1 (i.e., no fit at all), was used:

$$\text{BOF} = f\left(\underline{\mathbf{M}}^1, \mathbf{M}^2\right) = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(d_{ijk}^1 - m_{ijk}^1)^2 + \sum_{i=1}^{I}\sum_{l=1}^{L}(d_{il}^2 - m_{il}^2)^2}{IJK + IL}. \quad (13)$$

Further, also the BOF $-$ BOD statistic was computed.

Out of the 3,840 analyses, 3,825 (99.61%) result in a solution with a BOF $-$ BOD value smaller than or equal to 0, implying that in only 0.39% of the analyses it is sure that the CHIC algorithm ends in a suboptimal solution; in 1,489 analyses the BOF $-$ BOD value equals 0. An analysis of variance with BOF as dependent variable shows that almost all variance in BOF is explained by the main effect of Error level ($\hat{\rho}_I = 0.99$), with fit increasing when the Error level decreases. When comparing for each data set the 50 solutions obtained by the algorithm (one solution per SA chain) it turns out that on average, 30.53 solutions (61%) take the same optimal loss function value. Further, it appears that this number of optimal solutions increases (1) when the Error level decreases (45% for $\epsilon = 0.30$, while 82% for $\epsilon = 0$), (2) when the True rank of the CHIC model decreases (44% for $P = 5$ against 80% for $P = 2$), and (3) when the three-way array becomes much larger than the two-way matrix (36% for $r = 0.50$, while 81% for $r = 0.99$).

4.3.2. Goodness-of-Recovery When evaluating the extent to which the CHIC algorithm is capable of uncovering the true structure underlying the data, one has to differentiate between reconstructing (1) the true association relation, and (2) the true equivalence and hierarchical relations. Further attention is to be paid to the recovery of the true structure of the common mode in comparison with the other modes. Finally, the integrated modeling strategy is to be compared to the segmented one.

Association Relation To evaluate the recovery of the true association relation of the CHIC model, the kappa coefficient κ (Cohen, 1960) between the entries of the model array-matrix couple ($\underline{\mathbf{M}}^1$, $\mathbf{M}^2$) and the entries of the true array-matrix couple ($\underline{\mathbf{T}}^1$, $\mathbf{T}^2$) was computed. Kappa equals 1 when perfect recovery is observed and 0 when recovery is not better than chance level.

Note that kappa can take negative values, indicating that the recovery is worse than can be expected by chance.

The mean κ across all data sets equals 0.9878 (the standard deviation equals 0.0286), which means that the true array-matrix couple ($\underline{\mathbf{T}}^1$, $\mathbf{T}^2$), on average, is recovered to a very large extent. The true association relation is perfectly recovered ($\kappa = 1$) in 1471 out of 3,840 analyses (38%). Note that perfect recovery of the true association relation is only possible for these data sets in which BOF equals BOD (i.e., 1,489 out of 3,840; see *goodness-of-fit*). An analysis of variance with κ as the dependent variable shows that (only reporting on effects with $\hat{\rho}_I \geq 0.10$) recovery increases when the Array/total ratio increases ($\hat{\rho}_I = 0.16$) and the Error level decreases ($\hat{\rho}_I = 0.12$). Both main effects, however, are qualified by a strong Array/total ratio-Error level interaction ($\hat{\rho}_I = 0.43$): The increase in recovery performance when the three-way array becomes larger relative to the two-way matrix, becomes more pronounced when the amount of Error in the data increases. The main effects of True rank and Dominance can be neglected ($\hat{\rho}_I \leq 0.01$). Studying κ for the INDCLAS and the HICLAS parts of the CHIC model separately further shows that the true INDCLAS association relation is better recovered than the true HICLAS association relation (with the mean κ being equal to 0.9996 for INDCLAS against 0.9283 for HICLAS).

Equivalence and Hierarchical Relations In this paragraph, it is studied to which degree the true equivalence and hierarchical relations are recovered. To study the true equivalence relation, the corrected Rand index (Hubert & Arabie, 1985) between the partition of the set of objects (resp., attributes, sources, covariates) in the CHIC model for ($\underline{\mathbf{T}}^1$, $\mathbf{T}^2$) and its counterpart in the CHIC model for ($\underline{\mathbf{M}}^1$, $\mathbf{M}^2$) was computed (CRI). To study the true hierarchical relation, both for the true object component matrix $\mathbf{A}^{(T)}$ and the object component matrix $\mathbf{A}^{(M)}$ obtained by the model, a dichotomous variable was computed that denotes for each ordered object pair whether or not the first object of the pair is hierarchically below the second object of the pair; subsequently, a *goodness-of-hierarchy-recovery* statistic (GOHR) was calculated by computing the kappa coefficient (Cohen, 1960) between these two dichotomous variables.[2] Similarly, a GOHR statistic for the attributes, sources, and covariates was computed. The CRI and GOHR statistics are equal to 1 when the true equivalence and hierarchical relations are recovered perfectly, respectively, and are equal to 0 when recovery is at chance level. Finally, a combined CRI statistic (cCRI) and a combined GOHR statistic (cGOHR) were computed by, respectively, averaging the CRI and GOHR for the object, attribute, source, and covariate true equivalence and hierarchical relation, weighted by the number of elements in the respective modes.

The mean cCRI equals 0.8684 (standard deviation of 0.1902) and the mean cGOHR equals 0.9132 (standard deviation of 0.1327), implying overall a good recovery of the true equivalence and hierarchical relations. The true equivalence relation is recovered perfectly in 1,515 analyses, while 1,471 analyses resulted in perfect recovery of the true hierarchical relation. Separate analyses of variance with cCRI and cGOHR as dependent variables show very similar results: Only discussing effects with $\hat{\rho}_I \geq 0.10$, these analyses show that the recovery of the true equivalence and true hierarchical relations increases when the Error level decreases ($\hat{\rho}_I = 0.31$) and when the Array/total ratio increases ($\hat{\rho}_I = 0.14$). These main effects, however, are qualified by an Error level by Array/total ratio interaction ($\hat{\rho}_I = 0.21$): When a large amount of error is present in the data, large differences in recovery performance for the different levels of the Array/total ratio are observed, while almost no differences show up for error free data.

[2]Note that also the CRI statistic is a kappa coefficient, i.e., the kappa coefficient between two dichotomous variables, one pertaining to $\mathbf{A}^{(T)}$ and one to $\mathbf{A}^{(M)}$, that indicate for each unordered object pair whether or not both objects of the pair are equivalent.

TABLE 4.
Mean CRI and GOHR value for $\mathbf{A}$, $\mathbf{B}^1$, $\mathbf{C}$, and $\mathbf{B}^2$.

	CRI	GOHR
$\mathbf{A}$	0.9985	0.9993
$\mathbf{B}^1$	0.9975	0.9985
$\mathbf{C}$	0.9982	0.9989
$\mathbf{B}^2$	0.7438	0.8312

TABLE 5.
Mean κ, CRI, and GOHR value for the CHIC model and separate INDCLAS and HICLAS models.

		CHIC	Separate
κ	INDCLAS	0.9996	0.9991
	HICLAS	0.9283	0.8440
CRI	$\mathbf{A}^{\text{IND}}$	0.9985	0.9954
	$\mathbf{B}^1$	0.9975	0.9958
	$\mathbf{C}$	0.9982	0.9966
	$\mathbf{A}^{\text{HI}}$	0.9985	0.7488
	$\mathbf{B}^2$	0.7438	0.6731
GOHR	$\mathbf{A}^{\text{IND}}$	0.9993	0.9973
	$\mathbf{B}^1$	0.9985	0.9975
	$\mathbf{C}$	0.9989	0.9976
	$\mathbf{A}^{\text{HI}}$	0.9993	0.8206
	$\mathbf{B}^2$	0.8312	0.7693

The Common Mode When comparing the recovery of the common bundle matrix (in terms of the implied equivalence and hierarchical relations) to the recovery of the other bundle matrices, it appears that, on average, the structure of the common mode is better recovered than the structure of the other modes, although the difference being small for the bundle matrices associated with the noncommon modes of the three-way three-mode array (see Table 4 in which the mean CRI and GOHR for $\mathbf{A}$, $\mathbf{B}^1$, $\mathbf{C}$, and $\mathbf{B}^2$ are displayed). As an aside, one may note the poor recovery of the structure of the noncommon mode of $\mathbf{D}^2$ (see further Sect. 4.4).

To evaluate the performance of the integrated modeling strategy relative to the segmented one, the recovery of the true association (κ), equivalence (CRI), and hierarchical (GOHR) relations for the CHIC model is compared to the recovery of these relations for separate INDCLAS and HICLAS models, as obtained from a segmented modeling strategy. In Table 5, mean recovery values are presented both for the CHIC model and for separate INDCLAS and HICLAS models. It appears that the integrated modeling strategy outperforms the segmented one: When using the integrated strategy the recovery of all three relations is increased. Note that using the integrated modeling strategy leads to a better recovery of the structure of both the common and the noncommon modes, with the largest increase in recovery being observed for the modes of the HICLAS part of the CHIC model.

4.4. Discussion

In the simulation study, it was demonstrated that the CHIC algorithm succeeds well in minimizing the CHIC loss function. Further, it appeared that the true structure underlying the coupled data was recovered well, with the best recovery performance being observed for the common mode. With respect to the other modes, it was demonstrated that the INDCLAS part of the CHIC model is better recovered than the HICLAS part (i.e., $\mathbf{B}^2$ is poorly recovered in comparison to $\mathbf{B}^1$

TABLE 6.
Mean ratio of badness-of-recovery to badness-of-fit for the integrated an the segmented strategy for both model parts of the CHIC model.

	Integrated	Segmented
INDCLAS	0.0009	0.0017
HICLAS	0.1879	0.6245

and **C**), with the effect being most pronounced when a segmented strategy is used. Finally, the integrated modeling strategy was shown to outperform the segmented one in terms of (1) the recovery of the true association relation of the separate INDCLAS and HICLAS parts of the CHIC model, and (2) the recovery of the true equivalence and hierarchical relations of the common as well as the noncommon modes. Specifically, for the HICLAS part, a larger gain in recovery between both strategies was observed than for the INDCLAS part (i.e., a large recovery gain was observed for $\mathbf{B}^2$, with the gain for $\mathbf{B}^1$ and **C** being small).

When considering these results, one question that arises is why the HICLAS part of the CHIC model is worse recovered than the INDCLAS part, and this especially when a segmented strategy is used. It has to be noted that this asymmetry in recovery performance between the HICLAS and INDCLAS part of the model is not a consequence of the SA algorithm that was used; indeed, the simulation study showed that this algorithm performs well. We hypothesize, however, that this asymmetry is to be attributed to an overfitting tendency of the HICLAS part of the model, with the term overfitting being used to denote that the obtained model is misguided by the data at the expense of the truth.

To quantify the degree to which the HICLAS and INDCLAS parts of the CHIC model are overfitted, one may calculate for each model part, the ratio of the badness-of-recovery to the badness-of-fit. (Badness-of-recovery is defined here as the sum of the squared differences between the model block and the true block, while badness-of-fit is defined as the sum of the squared differences between the model block and the data block, with both sums of squared differences being normalized by the size of the corresponding block.) When the model has been pulled towards the truth, a small ratio will be observed, because in that case the badness-of-recovery will be small at the expense of the badness-of-fit. When, however, the model has been misguided by the data at the expense of the truth, the opposite will be true, implying a large ratio. Note that when no error is added to the data, the data block equals the true block, and hence badness-of-recovery necessarily equals badness-of-fit, implying the ratio of the two being equal to one. In Table 6, the mean ratio of badness-of-recovery to badness-of-fit, computed for the separate model parts, is displayed both for the integrated and the segmented strategy. From this table it appears, indeed, that the HICLAS part of the CHIC model is more misguided by the data at the expense of the truth (namely, is more subject to an overfitting tendency) than the INDCLAS part, with the effect being largest for the segmented strategy.

To understand why this is the case, one may note that the HICLAS and INDCLAS model parts differ with respect to the degree of flexibility with which they can be fitted to the data. As such, the HICLAS part is typically closer to the data than the INDCLAS part (mean BOF for HICLAS is 0.1418, while for INDCLAS it is 0.1502). Differences in degree of flexibility may further also be related to differences in "richness" of the (partial) solution spaces associated with both model parts. One way to quantify such richness is to compute the ratio of the number of parameters in the model to the number of data elements, with this ratio, in general, being larger for the HICLAS than for the INDCLAS part of the CHIC model. One may note that also for other models than HICLAS and INDCLAS an asymmetry in model flexibility may be observed; PCA, for example, is more flexible than PARAFAC/CANDECOMP (Harshman, 1970; Carroll & Chang, 1970), with both component models being counterparts of the HICLAS and INDCLAS model, respectively (see further).

One limitation of the simulation study has to be mentioned: In this study, only CHIC analyses in the true rank have been performed. One advantage of performing analyses in the true rank P is that a solution exists (i.e., the true rank P solution) against which the obtained solution can be compared in order to detect suboptimal solutions. A second advantage is that it allows to study the recovery at the level of the bundle matrices. In practice, however, the true rank of the CHIC model underlying the coupled data is never known. Further research is needed to see how algorithmic performance could be affected in analyses in an incorrect rank.

5. Illustrative Application

In this section, the CHIC model is applied to a coupled data set, gathered by Vansteelandt and Van Mechelen (1998), in a study from the domain of contextualized personality psychology. Within this domain, one tries to capture individual differences beyond differences in general dispositions (e.g., traits). A key concept in this regard is that of a behavioral signature, which is the profile across situations of the extent to which a specific behavior is displayed in each situation. A major challenge is to capture the structure of individual differences in behavioral signatures, and to relate this structure to individual differences in dispositional variables that reflect cognitive-affective processes underlying the signatures in question. In their study, Vansteelandt and Van Mechelen (1998) asked 54 persons to indicate on a 3-point scale the degree to which they would display 15 hostile behaviors in each of 23 frustrating situations (0 = you would not display this behavior in this situation, 1 = to a limited extent, 2 = to a strong extent). Further, they asked the same persons to rate themselves on 21 dispositional process variables making use of a 7-point scale (1 = not applicable at all, 7 = applicable to a strong extent). The $54 \times 15 \times 23$ person by behavior by situation data array $\underline{\mathbf{D}}^1$ was dichotomized by recoding scores of 1 and 2 to 1 and 0 to 0, while the 54×21 person by dispositional variable data matrix $\mathbf{D}^2$ was dichotomized by performing a median split on each variable (with a score above the median being recoded to 1 and a score under and on the median to 0).

CHIC analyses (with 1,000 SA chains) in rank one to six were performed on the dichotomized data ($\underline{\mathbf{D}}^1$, $\mathbf{D}^2$), resulting in the following (optimal) loss function value for the retained solution in rank one to six, respectively: 5,891, 5,489, 5,125, 4,890, 4,715, and 4,555. When applying both the scree test and the generalized scree test (see rank selection heuristics in Sect. 3.3) on these values, the rank three CHIC model was selected. In this rank, however, 12 solutions with the same optimal loss value were found that differ both with respect to some parameter values (i.e., the entries of the bundle matrices $\mathbf{A}$, $\mathbf{B}^1$, $\mathbf{C}$, and $\mathbf{B}^2$) and with respect to some values of the model array-matrix couple ($\underline{\mathbf{M}}^1$, $\mathbf{M}^2$), implying that the optimal solution was not unique. Note that this nonuniqueness is to be distinguished from the case of multiple solutions that differ with respect to some entries of $\mathbf{A}$, $\mathbf{B}^1$, $\mathbf{C}$, or $\mathbf{B}^2$, but that yield the same ($\underline{\mathbf{M}}^1$, $\mathbf{M}^2$). The latter type of nonuniqueness, however, did not show up in the present application. In general, it could be proven that the decomposition of ($\underline{\mathbf{M}}^1$, $\mathbf{M}^2$) into bundle matrices $\mathbf{A}$, $\mathbf{B}^1$, $\mathbf{C}$, and $\mathbf{B}^2$ is unique provided some regularity conditions.

The existence of different optimal solutions does not necessarily mean that the structure underlying the data cannot be revealed and interpreted in a meaningful way, as uncertainty may only apply to a small number of model parameters. For the data set under study, it appeared that the uncertainty was limited to 1 person (out of 54) and to 2 dispositional variables (out of 21) only, while no uncertainty was encountered for the situations or the behaviors. This implies that the structure for the person by behavior by situation data array is (almost) univocal, while uncertainty only exists about a small part of the structure for the person by dispositional variable data matrix. To deal with this, we simply removed the single person and two dispositional variables subject to uncertainty from the model. The resulting graphical representation is shown in

Fig. 5. Note that individual persons, situations, and responses are not displayed in this figure, but that they are represented by labels that give an interpretation of the class to which they belong (with the number of persons, situations and responses belonging to each class being denoted between parentheses). Note further that the bundle-specific situation classes that are empty are not displayed.

Clearly, the three situation classes form a Guttman scale; as such, they can be conceived as points on an underlying quantitative dimension (Gati & Tversky, 1982). This dimension can be interpreted as a frustration dimension, varying from lowly frustrating (e.g., accidently banging your shins against a park bench), over moderately frustrating (e.g., talking to someone who does not answer you), to highly frustrating (e.g., being unfairly accused of cheating). With respect to the responses, five classes can be distinguished, varying from weak (e.g., grimace and turning away) to strong responses (e.g., enragement). With regard to the persons, almost all of them belong to one of three person types P_1, P_2, and P_3 (that contain, respectively, 21%, 11%, and 42% of the persons); therefore, only these three person types will be further discussed and displayed in Fig. 5, resulting in the box representing the first person bundle being empty. From this figure, one may derive that all three person types do not display hostile behaviors in lowly frustrating situations. Further, persons belonging to P_1 differ from persons belonging to P_2 and P_3 in that the former react only to highly frustrating situations in terms of tension and enragement, while the latter display hostile behaviors in both moderately and highly frustrating situations. The latter two person types, however, differ from one another in that persons belonging to P_3, unlike those belonging to P_2, differentiate in terms of their behaviors between moderately and highly frustrating situations, with enragement being displayed by them in highly frustrating situations only. These differences in behavioral signatures can, at least in part, be explained by individual differences in dispositions. For example, the persons from P_1 differ from the persons from P_2 and P_3 in that the former always interpret aversive experiences caused by others as purposefully aggressive and react with enragement, whereas the latter react in an indirect aggressive way (i.e., grimace and turn away) because they think that showing one's anger is not appreciated by others.

6. Concluding Remarks

In this paper, the CHIC model, a global model for an integrated analysis of coupled data, has been proposed. In particular, CHIC can be used to simultaneously model a binary three-way three-mode data array and a binary two-way two-mode data matrix that have one mode in common. This type of data is often encountered in psychology, for example, when subjects rate a set of objects with respect to a set of attributes and when at the same time the same objects are characterized in terms of objective features or when dispositional information is available about the raters. In the context of psychiatric diagnoses, for example, psychiatrists may be asked to rate a set of patient vignettes on a list of symptoms, while at the same time information on the psychiatric diagnoses of the patients under study (in terms of a list of syndromes) is available.

In this paper, the gain of CHIC as an integrated modeling strategy, as compared to a segmented one, has been demonstrated by means of a simulation study and by means of an illustrative application. In particular, the results of our simulation study revealed that the CHIC algorithm performs well with respect to on the one hand minimizing the loss function and on the other hand recovering the true structure underlying the coupled data. Moreover, with respect to the latter, it was shown that the integrated CHIC modeling strategy outperforms its segmented counterpart. In the illustrative application, it has further been shown that the CHIC modeling strategy may lead to a deeper insight into a phenomenon under study in terms of how its different parts are connected to one another. In particular, in the application the CHIC model revealed how

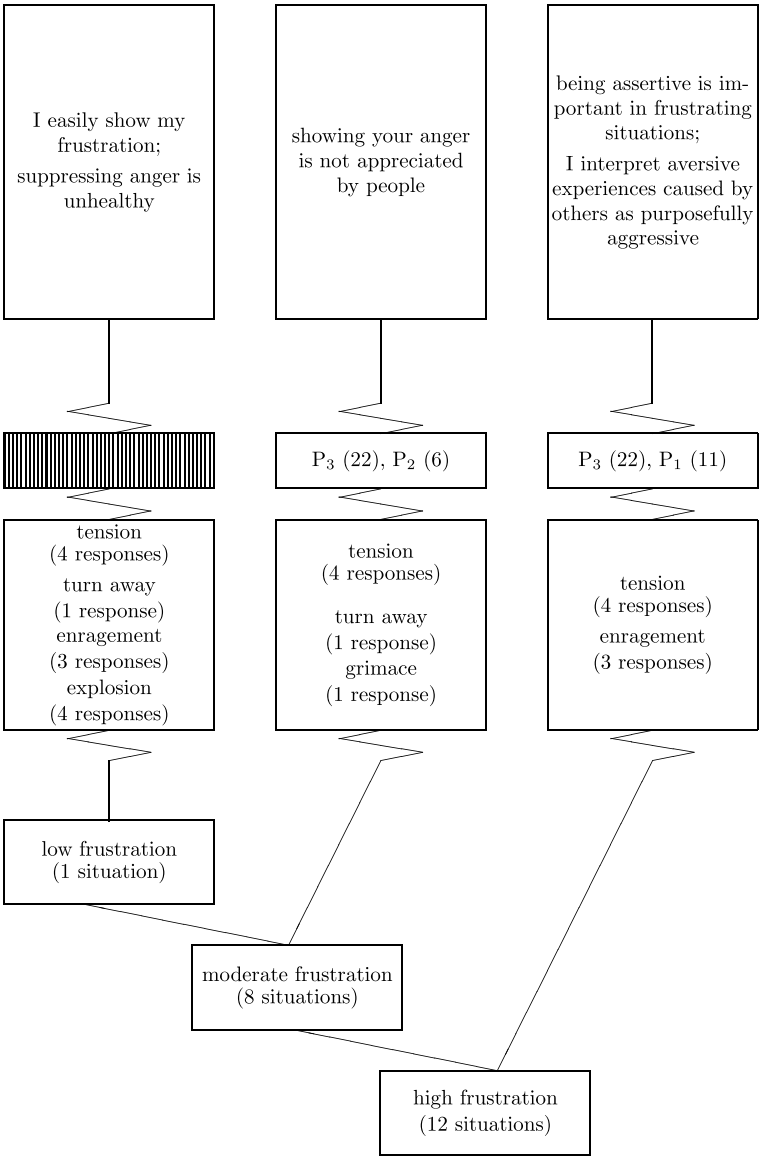


FIGURE 5.
Overall graphical representation of the rank 3 CHIC model for the hostility data.

the structure of individual differences in situation-specific hostile behavior relates to individual differences in process-based dispositional variables.

In the remainder of this section, first the relation of the CHIC model to other models is outlined. Next, generalizations of the proposed integrated modeling strategy are discussed.

6.1. Relations with Other Models

The proposed CHIC model is a novel model, but it bears interesting relationships to several existing models both inside and outside the hierarchical classes family. Inside the hierarchical classes family, the CHIC model can be conceived as an integrated model that includes two mem-

bers of this family as component models (namely, the INDCLAS and the HICLAS model). On the other hand, the CHIC model can also be considered a hierarchical classes counterpart of an integrated multiway covariates regression model consisting of a CANDECOMP/PARAFAC submodel for a three-way three-mode (real-valued) data array and a PCA submodel for a two-way two-mode (real-valued) data matrix, with the two data blocks having one mode in common (see Smilde & Kiers, 1999).[3] In both models a three-mode/two-mode coupled data structure ($\underline{\mathbf{D}}^1$, $\mathbf{D}^2$) is represented by a coupled model structure ($\underline{\mathbf{M}}^1$, $\mathbf{M}^2$) where: (1) each mode of ($\underline{\mathbf{M}}^1$, $\mathbf{M}^2$) is reduced to a limited number of components, with the common mode having the same component scores in each submodel, and (2) ($\underline{\mathbf{M}}^1$, $\mathbf{M}^2$) can be fully reconstructed on the basis of the different sets of component scores and a linking rule. Both models, however, differ in the following aspects: (1) the CHIC model includes binary component matrices, and as such implies a categorical rather than a dimensional reduction of the modes involved in the data; (2) the CHIC model is based on a Boolean decomposition of ($\underline{\mathbf{M}}^1$, $\mathbf{M}^2$), while the multiway covariates regression model is based on a standard real-valued decomposition; and (3) in the CHIC model the component matrices are restricted to represent the quasi-order relations $\leq$ between the elements of the corresponding modes of ($\underline{\mathbf{M}}^1$, $\mathbf{M}^2$).

6.2. Generalizations of the Integrated Modeling Strategy

6.2.1. Generalizations for the Same Type of Data We first consider generalizations of the CHIC strategy that still pertain to coupled data consisting of a binary three-way three-mode data block that is coupled to a binary two-way two-mode data block by one common mode. Within this context, CHIC extensions could be studied both at the level of the model and at the level of the data analysis.

At the *level of the model*, new global models could be developed by representing the separate data blocks by models other than HICLAS and INDCLAS. Below, we will propose two different possibilities in this regard.

A first new global model could be obtained by representing the separate data blocks of the coupled data by a conjunctive variant of the disjunctive HICLAS and INDCLAS models (see Van Mechelen, De Boeck & Rosenberg, 1995 for a discussion of conjunctive HICLAS models). Using a disjunctive or a conjunctive model has implications for the association relation and the way the bundles are interpreted. For example, when a two-way two-mode person by item data matrix is modeled with a disjunctive HICLAS model, the bundles can be interpreted as solution strategies, implying that a person can solve an item if he/she masters at least one strategy through which the item can be solved. With a conjunctive analysis of the same data set, however, the bundles can be interpreted as abilities, implying that a person can solve an item iff that person has all the necessary abilities involved in that item. Of course, also mixed global models consisting of both disjunctive and conjunctive parts could be considered.

As regards a second new global model, one may first note that the INDCLAS model, which is included in CHIC to represent the binary three-way three-mode data block, is rather restrictive because (1) each mode is reduced to the same number of bundles, and (2) the linking structure among the object, attribute, and source bundles is restricted to a one-to-one correspondence. From a substantive point of view, indeed, in several cases, there may be no a priori reason for the object, attribute, and source bundle matrices to have the same rank, let alone for a one-to-one correspondence among the respective bundles. As a way out, the CHIC model may be extended to a global model in which the three-way three-mode binary data block is represented by a less

[3] Note that a distinction can be drawn between two variants of the multiway covariates regression model: A symmetric variant in which both data blocks play an identical role, and an asymmetric variant in which the three-way data block is being considered a predictor block and the two-way block a block of criterion variables, with the CHIC model being closer related to the former variant.

restrictive model than INDCLAS, like, for example, the TUCKER3-HICLAS (Ceulemans et al., 2003) or the TUCKER2-HICLAS model (Ceulemans & Van Mechelen, 2004). The latter two models can be considered generalizations of the INDCLAS model in that in them the relation between the different bundle matrices is allowed to be more complex than a one-to-one correspondence and in that the numbers of bundles for the different modes are allowed to differ.

At the *level of the data analysis*, a weighted version of the global loss function that is optimized could be considered by assigning weights to the partial loss functions, with each weight indicating the importance of the corresponding data block in the analysis. As such, one can compensate for the different amounts of measurement error or noise that could be present in the two data blocks, which may occur when one data block was measured more reliable than the other. A challenging question then reads which weights are optimal with respect to the disclosure of the true underlying structure.

6.2.2. Generalizations for Other Types of Data One may note that CHIC pertains to coupled data that: (1) are binary, and (2) consist of a three-way three-mode data array that is coupled to a two-way two-mode data matrix by one common mode. With respect to the former, in psychology, however, often rating-valued data show up (e.g., as obtained from ratings on Likert-scales). The CHIC model could be extended to accommodate such data, by making use of HICLAS models for rating-valued data that have been developed recently (see Van Mechelen et al., 2007). Otherwise, CHIC extensions could also be developed to accommodate coupled data sets in which one data block is binary and the other one rating-valued.

With respect to the latter, psychology often yields other types of coupled data. For instance, when the same set of variables has been measured in samples from different populations (like when a test has been administered to respondents from two cultures) coupled data consisting of two two-way two-mode data blocks with one mode in common arise. The proposed integrated modeling strategy could easily be extended to accommodate such a coupled data set by modeling both data blocks by a HICLAS model and by restricting the bundle matrix for the common mode to be the same in both submodels. Of course, also global models for coupled data sets consisting of three (or more) two-way two-mode data blocks or two (or more) three-way three-mode data arrays that share one (or more) mode(s) in common could be considered, with the data being binary, rating-valued, or a mixture of both.

Appendix

Simulated annealing (for an elegant introduction, see Aarts et al., 1997) is a local search technique that implies a walk through the solution space. To lower the risk of ending in a suboptimal local minimum, during this walk worse solutions are accepted with a probability that depends on a quantity that varies across the algorithmic process and that is called the temperature. Higher values of the temperature imply a larger acceptance probability. The temperature is being decreased during the algorithm. As a consequence, at the beginning of the algorithm, when the algorithm is searching in a region of the solution space that is far from the optimal solution, worse solutions have a larger probability of acceptance than at the end, when the algorithm is searching near the optimal solution. Note that this procedure is analogous to the physical process of annealing (Kirkpatrick, Gelatt & Vecchi, 1983).

The general flow of a SA algorithm is as follows: Starting from an initial solution and an initial temperature, a chain of solutions is generated, consisting of a number of subchains. In particular, a subchain is generated by repeating the following procedure until a prespecified subchain stop criterion is satisfied: A candidate solution is generated from the current solution, with the current solution being the last accepted solution in the subchain; the first current solution of

each subchain is the last accepted solution of the previous subchain and the first current solution of the first subchain is the initial solution. To generate a candidate solution, the algorithm is based on the concept of the neighborhood of a solution. Although this concept can be defined in many ways, a requirement is that its definition is such that for each pair of solutions that belong to the solution space, there exists a sequence of solutions, starting from the first solution of the pair and ending with the second one, for which each solution in the sequence is a member of the neighborhood of the preceding one. Given the current solution, a candidate solution is generated by drawing one solution out of its neighborhood. Next, the loss function values of the current and the candidate solution are compared. If the candidate solution is associated with a better value on the loss function f, it is accepted, which implies that the current solution is replaced by the candidate solution. If the candidate solution, however, leads to a worse loss function value, it is accepted with a probability that depends on his relative quality (i.e., the difference in loss function value f between the current solution and the candidate one) and the current temperature. At the end of each subchain the temperature is decreased. Subchains are generated until a prespecified chain stop criterion is met. Finally, the best encountered solution in the chain is retained.

Three general comments about SA should further be made. First, SA can be applied with different types of (rational or random) starting configurations. Second, since SA is a heuristic optimization procedure, without a guarantee to yield a globally optimal solution, one could consider to implement SA along with a multistart procedure. This can be achieved by running different SA chains, with each chain starting from a different initial solution (and possibly also a different initial temperature), and retaining the best encountered solution over all chains. Third, because SA includes a stochastic element (i.e., whether or not a particular worse solution is accepted depends on a random number) SA chains that start from the same initial solution and initial temperature will not necessarily yield the same solution.

References

Aarts, E.H.L., Korst, J.H.M., & Van Laarhoven, P.J.M. (1997). Simulated annealing. In E.H.L. Aarts & J.K. Lenstra (Eds.), *Local search in combinatorial optimization* (pp. 91–120). Chichester: Wiley.

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.

Barbut, M., & Monjardet, B. (1970). *Ordre et classification: algèbre et combinatoire*. Paris: Hachette.

Birkhoff, G. (1940). *Lattice theory*. Providence: American Mathematical Society.

Boqué, R., & Smilde, A.K. (1999). Monitoring and diagnosing batch processes with multiway covariates regression models. *AIChE Journal*, *45*, 1504–1520.

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.

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

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*, 375–399.

Ceulemans, E., & Van Mechelen, I. (2005). Hierarchical classes models for three-way three-mode binary data: interrelations and model selection. *Psychometrika*, *70*, 461–480.

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

Ceulemans, E., Van Mechelen, I., & Leenen, I. (2007). The local minima problem in hierarchical classes analysis: An evaluation of a simulated annealing algorithm and various multistart procedures. *Psychometrika*, *72*, 377–391.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37–46.

Coxe, K.L. (1986). Principal components regression analysis. In N.L. Johnson & S. Kotz (Eds.), *Encyclopedia of statistical science* (Vol. 7, pp. 181–186). New York: Wiley.

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

Gati, I., & Tversky, A. (1982). Representations of qualitative and quantitative dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 325–340.

Harshman, R.A. (1970). Foundations of the PARAFAC procedure: models and conditions for an explanatory multi-modal factor analysis. *UCLA Working Papers in Phonetics*, *16*, 1–84.
Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*, 193–218.
Kirkpatrick, S., Gelatt, C.D. Jr., & Vecchi, M.P. (1983). Optimization by simulated annealing. *Science*, *220*, 671–680.
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.
McKenzie, D.M., Clarke, D.M., & Low, L.H. (1992). A method of constructing parsimonious diagnostic and screening tests. *International Journal of Methods in Psychiatric Research*, *2*, 71–79.
Smilde, A.K., & Kiers, H.A.L. (1999). Multiway covariates regression models. *Journal of Chemometrics*, *13*, 31–48.
Smilde, A.K., Westerhuis, J.A., & Boqué, R. (2000). Multiway multiblock component and covariates regression models. *Journal of Chemometrics*, *14*, 301–331.
Ten Berge, J.M.F. (1986). Rotation to perfect congruence and the cross-validation of component weights across populations. *Multivariate Behavioral Research*, *21*, 41–64.
Van Mechelen, I., De Boeck, P., & Rosenberg, S. (1995). The conjunctive model of hierarchical classes. *Psychometrika*, *60*, 505–521.
Van Mechelen, I., Lombardi, L., & Ceulemans, E. (2007). Hierarchical classes modeling of rating data. *Psychometrika*, *72*, 475–488.
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.

Manuscript Received: 22 DEC 2006
Final Version Received: 6 MAY 2008
Published Online Date: 30 JUL 2008