Studies in Classification, Data Analysis, and Knowledge Organization

Springer
Berlin
Heidelberg
New York
Hong Kong
London
Milan
Paris
Tokyo

Prof. Dr. Martin Schader
Department of Information Systems
University of Mannheim
Schloss
68131 Mannheim
Germany
martin.schader@uni-mannheim.de

Prof. Dr. Wolfgang Gaul
Institute of Decision Theory
University of Karlsruhe
Kaiserstr. 12
76128 Karlsruhe
Germany
wolfgang.gaul@wiwi.uni-karlsruhe.de

Prof. Maurizio Vichi
Department of Statistics
University of Rome
Piazzale Aldo Moro
00185 Rome
Italy
maurizio.vichi@uniroma1.it

ISSN 1431-8814
ISBN 3-540-40354-X Springer-Verlag Berlin Heidelberg New York

Cataloging-in-Publication Data applied for
A catalog record for this book is available from the Library of Congress.
Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH

http://www.springer.de

Printed in Germany

Softcover-Design: Erich Kirchner, Heidelberg

SPIN 10902882 43/3130/DK – 5 4 3 2 1 0 – Printed on acid-free paper

A Hierarchical Classes Approach to Discriminant Analysis

Luigi Lombardi[1,2], Eva Ceulemans[1], and Iven Van Mechelen[1]

[1] Department of Psychology,
Katholieke Universiteit Leuven, Tiensestraat 102, B-3000 Leuven, Belgium
[2] Department of Developmental Psychology,
Università di Padova, via Venezia 8, 35100, Padova, Italy

Abstract. Given an $I \times J$ object by attribute binary data matrix, and a predefined partition of the I objects into K partition classes, one may wish to characterize each of the K partition classes in terms of a set of singly necessary and jointly sufficient attributes; from the latter sets one may further derive attributes or attribute patterns that discriminate among the partition classes. In this paper, we propose how this goal may be achieved by fitting a new type of constrained hierarchical classes model, called D-HICLAS, to the data matrix. An algorithm for fitting D-HICLAS models is presented and an application to real data is discussed.

1 Introduction

Consider the following situations

- A marketing manager is interested in determining product qualities that best describe a given bunch of product categories.
- A medical researcher is interested in determining symptoms that significantly differentiate among known groups of psychiatric patients.
- A psychologist wants to investigate which cognitive profiles best characterize different given groups of students in primary school.

Each of these examples involves a characterization of given categories in terms of a predefined set of descriptors.

A particular case is the situation in which one wants to characterize each of K possible categories by means of a set of *singly necessary* and *jointly sufficient* binary attributes. From these sets one may further derive attributes or attribute patterns that possibly discriminate among the categories.

More formally, consider an $I \times J$ binary data matrix $\mathbf{D}$ defining a binary relation between a set of I objects and a set of J attributes, and an $I \times K$ binary partition matrix $\mathbf{P}$ defining a K–partition of the object set where, in particular, $p_{ik} = 1$ denotes that object i belongs to class k. Assume further that all classes of the K–partition are homogeneous with respect to each of the J attributes. The *characterization problem* then boils down to finding a $K \times J$ binary matrix $\mathbf{C}$, called the *characterization matrix*, such that the quantity

$$E = \sum_{i=1}^{I} \sum_{j=1}^{J} |d_{ij} - \sum_{k=1}^{K} p_{ik} c_{kj}| \qquad (1)$$

is minimal. It is straightforward to show that (1) is minimal if, given the proportion π_{kj} $(\forall k = 1,\ldots,K; \forall j = 1,\ldots,J)$ of objects in class k which possess attribute j, the entries c_{kj} of $\mathbf{C}$ are replaced with

$$z_{kj} = \begin{cases} 1 & \text{if } \pi_{kj} \geq .50 \\ 0 & \text{if } \pi_{kj} < .50 \end{cases} \qquad (\forall k = 1,\ldots,K;\ \forall j = 1,\ldots,J) \qquad (2)$$

The matrix $\mathbf{Z}$ yields an exact solution to the minimization of (1). However, an exact representation of $\mathbf{C}$ itself may be troublesome in empirical data matrices. In particular, in case of data sets that are not very small and with a large number of attributes, the high complexity of the characterization sets can be very difficult to interpret. A possible way out consists of approximating $\mathbf{C}$ by means of a matrix that has a much simpler structural representation.

The method to be presented here, which is called D-HICLAS (HICLAS model for descriptive Discriminant analysis), is a novel extension of the conjunctive HICLAS model (Van Mechelen et al., 1995) to deal with the problem of approximating the characterization matrix $\mathbf{C}$. A D-HICLAS model yields a simplified approximation of $\mathbf{C}$ by reducing the original set of attributes to a few binary variables, called bundles. Moreover a D-HICLAS analysis guarantees also that the K–partition is well represented in the model.

Section 2 of this paper outlines the general theory of D-HICLAS model and the associated data analysis. Section 3 illustrates the new model with an application to real psychological data. Finally in Section 4 some possible model extensions are discussed.

2 Descriptive discriminant HICLAS approach

2.1 The model

We assume an $I \times J$ object by attribute data matrix $\mathbf{D}$ and an $I \times K$ binary partition matrix $\mathbf{P}$ defining a K–partition on the objects (K–categories). A descriptive discriminant hierarchical classes analysis will approximate $\mathbf{D}$ by an $I \times J$ reconstructed binary matrix $\mathbf{M}$ that can be decomposed into $\mathbf{P}$ and a $K \times J$ binary matrix $\mathbf{M}^*$ that approximates the characterization matrix $\mathbf{C}$. $\mathbf{M}^*$ can be further decomposed into a $K \times R$ binary matrix $\mathbf{A}$ and a $J \times R$ binary matrix $\mathbf{B}$, where R denotes the rank of the model. In particular, $\mathbf{A}$ defines R, possibly overlapping, clusters of the K–categories, whereas $\mathbf{B}$ defines R, possibly overlapping, clusters of attributes. As a guiding example we use the hypothetical matrices shown in Tables 1–3.

2.2 Relations represented in D-HICLAS model

The D-HICLAS model represents three types of relations defined in $\mathbf{M}^*$: association, equivalence and hierarchy.

The *association relation* is the binary relation between the categories and the attributes of $\mathbf{M}^*$ as defined by the 1–entries of $\mathbf{M}^*$. *Equivalence relations*

are defined on the categories and on the attributes of $\mathbf{M}^*$. Categories are equivalent iff they are associated with the same set of attributes. Likewise attributes are equivalent iff they are associated with the same set of categories. A *hierarchical relation* is defined among the categories and among the attributes of $\mathbf{M}^*$. A category is hierarchically below another, iff the respective sets of associated attributes are in a subset/superset relation. Similarly, an attribute is hierarchically below another, iff the respective set of associated categories are in a subset/superset relation.

The matrices $\mathbf{A}$ and $\mathbf{B}$ of a D-HICLAS model represent the three relations as follows

i) *Association relation*:

$$\mathbf{M} = \mathbf{P}\mathbf{M}^* \tag{3}$$

with $\mathbf{M}^* = [\mathbf{A}^c \otimes \mathbf{B}']^c$ (where $\otimes$ denotes the Boolean matrix product (Kim, 1982)). This association rule means that for an arbitrary entry m_{ij} of $\mathbf{M}$,

$$\begin{aligned} m_{ij} &= \sum_{k=1}^{K} p_{ik} m^*_{kj} \\ &= \sum_{k=1}^{K} p_{ik} \left(\bigoplus_{r=1}^{R} a^c_{kr} b_{jr} \right)^c \end{aligned} \tag{4}$$

where m^*_{kj} indicates the (k,j)–entry of $\mathbf{M}^*$ and $\bigoplus$ denotes the Boolean sum (Kim, 1982). Moreover (3) implies that for an arbitrary entry m_{ij} of $\mathbf{M}$,

$$m_{ij} = 1 \quad \Leftrightarrow \quad \exists k : 1, \ldots, K \quad : \quad p_{ik} = 1 \quad \wedge \quad m^*_{kj} = 1 \tag{5}$$

where

$$m^*_{kj} = 1 \quad \Leftrightarrow \quad \forall r : 1, \ldots, R \quad : \quad b_{jr} = 1 \quad \Rightarrow \quad a_{kr} = 1 \tag{6}$$

In (6) the $\mathbf{B}$ row vector of an attribute can be considered to denote a set of formal requisites stemming from that attribute; these requisites are to be jointly met by a category to be associated with the attribute in question. For example, from the model in Table 3, it can be derived that category C_1 is associated with attribute c. This further implies that all objects in category $C_1 = \{1, 2, 3\}$ are also associated with attribute c.

ii) *Equivalence relations*: A category k (resp. an attribute j) is equivalent to another category k' (resp. another attribute j') if and only if $\mathbf{A}_{k:} = \mathbf{A}_{k':}$ (resp. $\mathbf{B}_{j:} = \mathbf{B}_{j':}$). For example, categories C_1, C_2 and C_3 are associated with different set of attributes, and, hence, those categories are not equivalent and have different row vectors in the $\mathbf{A}$ matrix of the D-HICLAS model of Table 3. On the other hand, notice that by (3) all objects belonging to the same category share the same attribute pattern.

iii) *Hierarchical relations*: A category k (resp. an attribute j) is hierarchically below to another category k' (resp. another attribute j') if and only if $\mathbf{A}_{k:} \leq \mathbf{A}_{k':}$ (resp. $\mathbf{B}_{j:} \geq \mathbf{B}_{j':}$). Note that for $\mathbf{B}$ this reversal ($\geq$) reflects the fact that the more requisites stem from an attribute, the fewer categories that attribute is associated with. For example, attribute c is hierarchically above attribute d; consequently, the bundle patterns of attributes c and d are in a subset/superset relation in the D-HICLAS model of Table 3.

Graphic representation. Figure 1 presents a graphic representation for the D-HICLAS model of Table 3. In Figure 1 category and attribute classes are displayed as paired boxes, the upper box of each pair being a category class and the lower box an attribute class. The top of the category hierarchy is at the top of the representation, whereas the attribute hierarchy is represented upside down. The association relation can be read from the representation as a dominance relation: A category is associated with all attributes below it, and an attribute is associated with all categories above it.

	D					**M**			
	Attributes					Attributes			
Objects	*a*	*b*	*c*	*d*	Objects	*a*	*b*	*c*	*d*
Obj 1	1	0	1	0	Obj 1	1	0	1	0
Obj 2	1	1	1	0	Obj 2	1	0	1	0
Obj 3	1	0	1	0	Obj 3	1	0	1	0
Obj 4	0	0	1	1	Obj 4	0	0	1	0
Obj 5	0	0	1	0	Obj 5	0	0	1	0
Obj 6	0	1	1	0	Obj 6	0	1	1	0
Obj 7	0	0	1	0	Obj 7	0	1	1	0

Table 1. Hypothetical data matrix **D** and related D-HICLAS model matrix **M**

	P				**M***				
	Categories					Attributes			
Objects	C_1	C_2	C_3		Categories	*a*	*b*	*c*	*d*
Obj 1	1	0	0		C_1	1	0	1	0
Obj 2	1	0	0		C_2	0	0	1	0
Obj 3	1	0	0		C_3	0	1	1	0
Obj 4	0	1	0						
Obj 5	0	1	0						
Obj 6	0	0	1						
Obj 7	0	0	1						

Table 2. D-HICLAS Model decomposition for matrix **M** in Table 1

	A			B	
	Bundles			Bundles	
Categories	*I*	*II*	Attributes	*I*	*II*
C_1	0	1	*a*	0	1
C_2	0	0	*b*	1	0
C_3	1	0	*c*	0	0
			d	1	1

Table 3. D-HICLAS Model decomposition for matrix $\mathbf{M}^*$ in Table 2

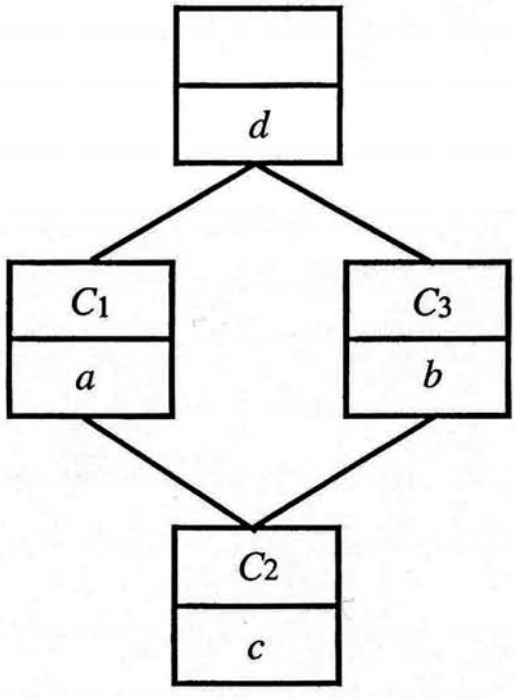

Fig. 1. Graphic representation of the conjunctive D-HICLAS model of Table 3. Empty boxes denote empty classes.

2.3 Data analysis and algorithm

The aim of a D-HICLAS analysis in rank R of a binary data matrix $\mathbf{D}$ and a given partition matrix $\mathbf{P}$ is to approximate $\mathbf{D}$ as closely as possible by a reconstructed binary model matrix $\mathbf{M}$, in terms of the loss function

$$L = \sum_{i=1}^{I} \sum_{j=1}^{J} |d_{ij} - m_{ij}| \tag{7}$$

and such that $\mathbf{M}$ can be represented by a rank R D-HICLAS model.

The routine which looks for matrices $\{\mathbf{A}, \mathbf{B}\}$, such that (7) is minimal is a modified version of the alternating greedy procedure for ordinary HICLAS analysis (Leenen and Van Mechelen, 2001). In particular, the D-HICLAS algorithm successively executes two main routines.

i) Given an initial random configuration, $\mathbf{A}^0$ for $\mathbf{A}$, the procedure looks, conditionally upon $\mathbf{A}^0$, for the optimal matrix $\mathbf{B}^0$, which is such that

(7) is minimal. In the next steps, $\mathbf{A}^w$ is reestimated conditionally upon $\mathbf{B}^{w-1}$, and $\mathbf{B}^w$ is reestimated conditionally upon $\mathbf{A}^w$ ($w = 1, 2 \ldots$). This procedure continues until no further improvement in the loss function (7) is observed.

A conditional estimate for $\mathbf{A}$ (resp. $\mathbf{B}$) is obtained by a greedy heuristic which successively estimates each row of $\mathbf{A}$ (resp. $\mathbf{B}$). In particular, given $\mathbf{A}$, the j–th row $\mathbf{B}_{j:}$ of $\mathbf{B}$ ($\forall j = 1, \ldots, J$) is optimized by means of a Boolean regression that minimizes

$$L_{(j)} = \sum_{i=1}^{I} |d_{ij} - m_{ij}| \tag{8}$$

Whereas, given $\mathbf{B}$, the k–th row $\mathbf{A}_{k:}$ of $\mathbf{A}$ ($\forall k = 1, \ldots, K$) is optimized by means of a generalized form of Boolean regression that minimizes

$$L_{(k)} = \sum_{i:p_{ik}=1} \sum_{j=1}^{J} |d_{ij} - m_{kj}| \tag{9}$$

More precisely, in this regression the values of each predictor variable are $N_k = \sum_i^I p_{ik}$ concatenated copies of each column of $\mathbf{B}$, whereas the values of the criterion vector are the data–entries d_{ij} ($\forall i : p_{ik} = 1; \forall j = 1, \ldots, J$).

ii) In the second main routine the matrices $\mathbf{A}$ and $\mathbf{B}$ obtained at the end of the first routine are modified such as to make them consistent with the equivalence and hierachical relations in the model matrix $\mathbf{M}$, that $\mathbf{A}$ and $\mathbf{B}$ yield by (3). For this, a closure operation (Barbut and Monjardet, 1970) is successively applied to each of the two matrices $\mathbf{A}$ and $\mathbf{B}$. This operation implies that zero–entries in the two matrices are turned into one if this change does not alter $\mathbf{M}$ (and, hence, neither the value of the loss function).

3 An empirical application

In this section we present a D-HICLAS analysis of data from a study on archetypal psychiatric patients (Mezzich and Solomon, 1980). In this study each of 11 psychiatrists was invited to think of a typical patient for each one of four diagnostic categories: manic–depressive depressed (MDD), manic–depressive manic (MDM), simple schizophrenic (SS) and paranoid schizophrenic (PS). These four diagnostic categories are part of the nomenclature of mental disorders (DSM-II) issued in 1968 by the American Psychiatric Association. The 11 psychiatrists characterized each archetypical patient by $0-6$ severity ratings on 17 symptoms from the Brief Psychiatric Rating Scale (BPRS).

In order to yield a D-HICLAS analysis each symptom of the original data base was trichotomized into two dummy variables indicating at least a medium severity rating ($1-6$) and an high severity rating ($3-6$), respectively.

This resulted in a 44×34 patient by symptom data matrix $\mathbf{D}$. Next, $\mathbf{D}$ was analyzed by means of the D-HICLAS algorithm in ranks 1 to 5.

On the basis of a scree test, the rank 3 solution with a proportion of discrepancies of .167 was retained. Figure 2 shows the graphic representation of the conjunctive D-HICLAS rank-3 solution.

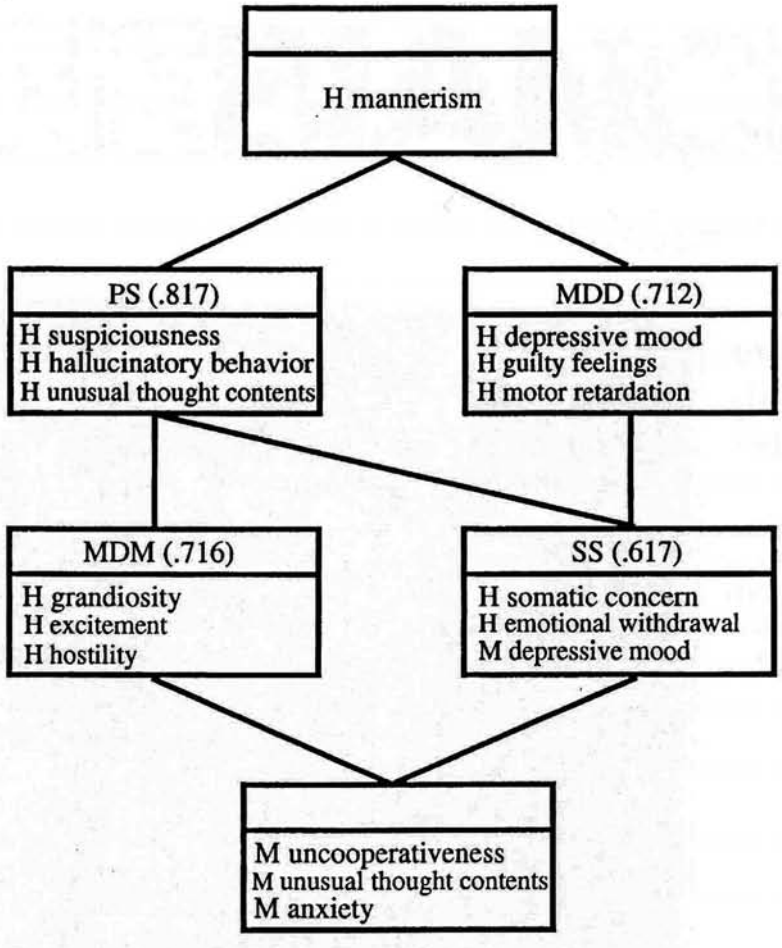

Fig. 2. Graphic representation of the conjunctive D-HICLAS rank-3 solution. Diagnostic categories are displayed in the upper boxes, symptoms in the lower boxes. M and H denote medium rating and high rating, respectively.

The graphic representation reads as follows:

Diagnostic categories and symptom classes are displayed as paired boxes, the upper box of each pair containing diagnostic categories and the lower box containing symptoms. The top of the diagnostic category hierarchy is at the top of the representation, whereas the symptom hierarchy is represented upside down. In order to simplify the graphic representation we inserted in each symptom class the three best fitting symptoms only. Moreover, for each diagnostic category we displayed both the label and the goodness of fit in the upper boxes.

Notice that the association relation can be read from the representation as a dominance relation: A psychiatric category is associated with all symptoms below it, and a symptom is associated with all psychiatric categories above it.

In substantive terms Figure 2 reads as follows: The grouping of symptoms in the graphic representation is in line with the classical distinction between negative and positive psychotic symptoms (Andreasen and Olsen, 1982; Stuart, Malone, Currie, Klimidis and Minas, 1995). In particular, the left side of the symptoms hierarchy contains extreme positive psychotic symptoms and extreme positive affective disorder symptoms that are typical of the paranoid schizophrenic type (PS). Whereas the right side of the symptoms hierarchy contains negative psychotic symptoms and extreme negative affective disorder symptoms that are typical of the manic–depressive depressed type (MDD). Finally, notice that the paranoid schizophrenic type (PS) can be represented as the conjunctive combination of the manic–depressive manic type (MDM) and the simple schizophrenic type (SS), with the latter diagnostic category being characterized by negative psychotic symptoms.

4 Possible extensions

Several possible extensions of the descriptive discriminant hierarchical classes model may be considered.

In the present paper, D-HICLAS has been proposed as a model for two-way Boolean data. However, the current approach can be straightforwardly extended to rating–valued data. In particular, a D-HICLAS model for rating data would be considered as a particular constrained instance of the HICLAS-R model family (Van Mechelen, Lombardi and Ceulemans, 2002). In the rating–valued context the *characterization problem* then would boil down to finding a simplified rating–valued matrix that approximates the characterization matrix $\mathbf{C}$.

Another further extension of the D-HICLAS model could be derived modifying the nature of the association relation (5). In fact, likewise the standard two-way HICLAS model for binary data (De Boeck and Rosenberg, 1988) a disjunctive variant of the conjunctive D-HICLAS model can be formulated by means of the following disjunctive association rule

$$m_{ij} = \sum_{k=1}^{K} p_{ik} \left(\bigoplus_{r=1}^{R} a_{kr} b_{jr} \right) \tag{10}$$

Finally, a possible D-HICLAS model relaxation would imply hierarchical representations in which objects in the same category do not necessarily share the same attribute pattern.

References

ANDREASEN, N.C. and OLSEN, S. (1982): Negative v positive shizophrenia: definition and validation. *Archives of General Psychiatry, 39, 789–794.*

BARBUT, M. and MONJARDET, B. (1970): *Ordre et classification: Algèbre et combinatoire* (2 Vols.) Hachette, Paris.

DE BOECK, P. and ROSENBERG, S. (1988): Hierarchical classes: Model and data analysis. *Psychometrika, 53, 361–381.*

KIM, K.H. (1982): *Boolean matrix theory.* Marcel Dekker, New York.

LEENEN, I. and VAN MECHELEN, I. (2001): An evaluation of two algorithms for hierarchical classes analysis. *Journal of Classification, 18, 57–80.*

MEZZICH, J.E. and SOLOMON, H. (1980): *Taxonomy and behavioral science: comparative performance of grouping methods.* Academic Press, London.

STUART, G.W., MALONE, V., CURRIE, J., KLIMIDIS, S. and MINAS, I.H. (1995): Positive and negative symptoms in neuroleptic–free psychotic inpatients. *Schizophrenia Research, 16, 175–188.*

VAN MECHELEN, I., DE BOECK, P. and ROSENBERG, S. (1995): The conjunctive model of hierarchical classes. *Psychometrika, 60, 505–521.*

VAN MECHELEN, I., LOMBARDI, L. and CEULEMANS, E. (2002): Hierarchical classes modeling of rating data. Submitted for publication.