

Modelle und Verfahren der Clusteranalyse auf der Basis von Skalarprodukt-Relationen

(Models and Techniques of Cluster Analysis Based on Scalar Product Relations)

Hartmann, W.: Modelle und Verfahren der Clusteranalyse auf der Basis von Skalarprodukt-Relationen (Models and Techniques of cluster analysis based on scalar product relations)

In: Int. Classif. 9 (1982) No. 3, p. 129–139, 11 refs.

This contribution consists of two main parts. In the first part three different models of cluster analysis are presented based on scalar product relations.

Model I:

Let $T = (t_{ij})$ be a given (0,1)-binary $n \times n$ -proximity-matrix (Incidence Link Matrix, ILM) and further let $X = (x_{ik})$ be a fitting (0,1)-binary $n \times p$ -object-cluster-membership matrix. For disjunctive partitions $X = (x_{ik})$ has the orthogonality property

$$\sum_{i=1}^n x_{ik} x_{il} = X'_k X_l = \begin{cases} n_k, & \text{if } k = l, \\ 0 & \text{otherwise,} \end{cases} \quad k, l = 1, \dots, p, \quad (1)$$

and

$$t_{ij} \simeq s_{ij} = \sum_{k=1}^p x_{ik} x_{jk} = X'_i X_j, \quad i, j = 1, \dots, n, \quad x_{ik} \in \{0, 1\}, \quad (2)$$

describes a model for analysis of p mutual disjunctive clusters $X_k, k = 1, \dots, p$.

Model II:

The logical equivalent of the scalar product relation (2)

$$t_{ij} \simeq s_{ij} = \bigcup_{k=1}^p x_{ik} \cap x_{jk} = X'_i \cap X_j, \quad i, j = 1, \dots, n, \quad x_{ik} \in \{0, 1\}, \quad (3)$$

\cup : logical disjunction, \cap : logical conjunction

(with a given symmetrical, diagonal-dominant, (0,1)-binary ILM) resp. the loss function can be fitted always perfectly

$$\sigma(X) = \sum_{i=1}^n \sum_{j=1}^n (t_{ij} - s_{ij}(X))^2 = \sum_{i=1}^n \sum_{j=1}^n (t_{ij} - \bigcup_{k=1}^p x_{ik} \cap x_{jk})^2 \quad (4)$$

with a sufficiently large number p of overlapping compact clusters $X_k, k = 1, \dots, p$.

Model III:

Let $U = (u_{ij})$ be a given nonnegative symmetrical $n \times n$ -proximity matrix with nondiagonal elements $u_{ij}, i \neq j$, which can be considered as the probabilities for the

events that two objects i and j are commonly included in one of the p clusters. Furthermore let $Y = (y_{ik})$ be a metrical $n \times p$ -object-cluster-membership matrix the elements of which $y_{ik}, 0 \leq y_{ik} \leq 1$, are interpretable as probabilities for the events, that the object i is element of the cluster k . With application of the addition theorem (for disjunctive events) and the multiplication theorem (for independent events) of probabilities the stochastic founded model of cluster analysis can be derivated

$$u_{ij} \simeq r_{ij} = \sum_{k=1}^p y_{ik} y_{jk} = Y'_i Y_j, \quad i \neq j, \quad i, j = 1, \dots, n, \quad 0 \leq y_{ik} \leq 1. \quad (5)$$

In the second part of this contribution different methods and algorithms are presented for the numerical treatment of these three scalar product models.

A: Cluster Modelle

A.1. Einleitung

In letzter Zeit wurden erhebliche Anstrengungen unternommen, Methoden der Clusteranalyse funktional zu begründen in Form von Modellfunktionen, ähnlich denen der klassischen Datenanalysetechniken (vgl. z.B. SHEPARD & ARABIE, 1979). Eine funktionale Modellbegründung erlaubt:

- Schätzmethoden für die Parameter der Modellfunktion anzuwenden (z.B. Least-Squares- oder Maximum-Likelihood-Methoden);
- die mit Hilfe der Modellfunktion erzeugten (reproduzierten) Modellwerte mit den gegebenen Daten zu vergleichen und die Güte der Modellanpassung einer gegebenen Datenmenge zu beurteilen;
- die für eine Analyse nach der Modellfunktion erforderliche Datenstruktur theoretisch zu beschreiben und Bedingungen für die Existenz einer Lösung und deren Eigenschaften anzugeben.

In diesem Beitrag soll, angeregt durch frühere Arbeiten von Raymond B. CATTELL, der Versuch unternommen werden, verschiedene Klassen von Cluster-Analyse-Modellen durch eine Skalarprodukt-Relation, verwandt der der Hauptkomponenten-Analyse, zu begründen.

A.2. Dichotomes Skalarproduktmodell zur disjunktiven Clusteranalyse

Die Zerlegung (Partition) einer Menge von n Objekten in p wechselseitig disjunkte Cluster kann mit einer (0,1)-binären $n \times p$ -Zuordnungs-Matrix $X = (x_{ik})$ beschrieben werden, deren Zeilen zu den Objekten und deren Spalten zu den Clustern korrespondieren, d.h. es gilt

$$x_{ik} = \begin{cases} 1, & \text{falls Objekt } i \text{ zu Cluster } C_k \text{ gehört,} \\ 0, & \text{sonst.} \end{cases}$$

Für disjunkte Partitionen hat die binäre Zugehörigkeits-Matrix $X = (x_{ik})$ die „Orthogonalitäts-Eigenschaft“, d.h. es gilt

$$\sum_{i=1}^n x_{ik} x_{il} = X'_k X_l = \begin{cases} n_k, & \text{falls } k = l, \\ 0, & \text{sonst,} \end{cases} \quad (1)$$

bzw. $X'X = \text{DIAG}(n_1, \dots, n_p)$,

wobei n_k die Anzahl der Elemente des Clusters C_k bezeichnet. Mit der Skalarprodukt-Relation

$$s_{ij} = \sum_{k=1}^p x_{ik}x_{jk} = X_i'X_j \text{ bzw. } S = XX' \quad (2)$$

wird eine $n \times n$ -Matrix $S = (s_{ij})$ von Modellwerten erzeugt, deren Elemente s_{ij} anzeigen, ob die korrespondierenden Zeilen- und Spaltenobjekte in einem gemeinsamen Cluster enthalten sind oder nicht, d.h. es gilt

$$s_{ij} = \begin{cases} 1, & \text{falls Objekt } i \text{ und Objekt } j \text{ in einem Cluster } C_k \\ & \text{gemeinsam enthalten sind,} \\ 0, & \text{sonst.} \end{cases}$$

Falls ein Objekt i Element eines Clusters C_k , $k = 1, \dots, p$, ist, ist das Diagonalelement s_{ii} von S gleich Eins. Bei erschöpfenden (vollständigen) Partitionen, bei denen jedes Objekt Element eines Clusters ist, gilt für alle Diagonalelemente von S

$$s_{ii} = 1, \quad i = 1, \dots, n.$$

Sei $T = (t_{ij})$ eine gegebene $n \times n$ -Incidence-Link-Matrix (ILM), d.h. eine (0,1)-binäre Proximitäts-Matrix, deren Elemente t_{ij} angeben, ob zwei Objekte i und j einander ähnlich sind oder nicht, d.h. es gilt

$$t_{ij} = \begin{cases} 1, & \text{falls die Objekte } i \text{ und } j \text{ ähnlich sind,} \\ 0, & \text{falls die Objekte } i \text{ und } j \text{ unähnlich sind.} \end{cases}$$

Die ILM $T = (t_{ij})$ sei außerdem symmetrisch und habe Eins-Diagonale

$$t_{ij} = t_{ji} \text{ und } t_{ii} = 1, \text{ für alle } i, j = 1, \dots, n.$$

Die ILM $T = (t_{ij})$ kann beispielsweise mit Schwellwert-Dichotomisierung aus einer „metrischen“ Proximitäts- oder Korrelations-Matrix abgeleitet werden. Dann begründet folgende Modellrelation ein Schätzmodell der Clusteranalyse:

Modell I:

$$t_{ij} \approx s_{ij} = \sum_{k=1}^p x_{ik}x_{jk} = X_i'X_j, \quad i, j = 1, \dots, n, \quad (3)$$

$$x_{ik} \in \{0, 1\}, \quad k = 1, \dots, p.$$

Das Least-Squares-Fitkriterium

$$\sigma(X) = \text{Spur} (T - S)'(T - S) = \sum_{i=1}^n \sum_{j=1}^n (t_{ij} - s_{ij}(X))^2$$

$$= \sum_{i=1}^n \sum_{j=1}^n (t_{ij} - X_i'X_j)^2 \quad (4)$$

gibt wegen der binären Form von S und T die Anzahl der Abweichungen korrespondierender Elemente in den Matrizen $T = (t_{ij})$ und $S = (s_{ij}) = XX'$ an. Für einen perfekten Fit, $\sigma(X) = 0$, ist neben einer hinreichend großen Clusteranzahl p ($\ll n$) notwendig, daß sich die Zeilen und Spalten der ILM $T = (t_{ij})$ zu einer Block-Diagonal-Matrix permutieren lassen.

A.3. Dichotomes Skalarproduktmodell zur überlappenden Clusteranalyse

Wenn die Skalarprodukt-Relation (2) durch das logische Äquivalent

$$s_{ij} = \bigcup_{k=1}^p x_{ik} \cap x_{jk}, \quad \cup: \text{Vereinigung, } \cap: \text{Durchschnitt} \quad (5)$$

ersetzt wird, kann mit einer hinreichend großen Anzahl p (i.a. $> n$) nichtdisjunkter (überlappender) kompakter Cluster stets ein perfekter Fit, $\sigma(X) = 0$, zu einer beliebigen gegebenen ILM $T = (t_{ij})$ erreicht werden. Eine überlappende Clusterstruktur erfüllt jedoch nicht die Orthogonalitätseigenschaft (1), auch wenn man die dort verwendete Skalarprodukt-Relation durch das logische Äquivalent (5) ersetzt. Folgende Modellrelation begründet das entsprechende Skalarproduktmodell der Clusteranalyse.

Modell II:

$$t_{ij} \approx s_{ij} = \bigcup_{k=1}^p x_{ik} \cap x_{jk} = X_i' \cap X_j, \quad i, j = 1, \dots, n, \quad (6)$$

$$x_{ik} \in \{0, 1\}.$$

A.4. Metrische Skalarproduktmodelle zur Clusteranalyse von Wahrscheinlichkeitsdaten

Sei $Y = (y_{ik})$ eine nichtnegative metrische $n \times p$ -Zuordnungs-Matrix, deren Elemente y_{ik} , $0 \leq y_{ik} \leq 1$, die Wahrscheinlichkeiten dafür sind, daß ein Objekt i Element eines Clusters C_k ist

$$y_{ik} = \text{Prob}(i \in C_k), \quad 0 \leq y_{ik} \leq 1, \quad (7a)$$

$$i = 1, \dots, n, \quad k = 1, \dots, p.$$

Die Mitgliedschafts-Wahrscheinlichkeiten y_{ik} der Objekte i zu den Clustern C_k können auch als relative (bzw. prozentuale) Mengenanteile interpretiert werden, also als relativer (bzw. prozentualer) Anteil der Gesamtmasse des Objektes i , der zum Cluster C_k gehört (vgl. Theorie der „fuzzy sets“).

Wenn man disjunkte Cluster, d.h. Partitionen, voraussetzt, dann schließt die Mitgliedschaft eines Objektes i zum Cluster C_k die Mitgliedschaft des Objektes i in anderen Clustern aus, d.h. die Ereignisse sind unverträglich und es gilt der Additionssatz für Wahrscheinlichkeiten

$$\sum_{k=1}^p y_{ik} \leq 1 \text{ bzw. } \sum_{k=1}^p y_{ik} = 1, \quad i = 1, \dots, n. \quad (7b)$$

Die zweite Beziehung gilt für vollständige (erschöpfende) Partitionen. Da die Zugehörigkeit eines Objektes i zu einem Cluster C_k , i.a. unabhängig ist von der Zugehörigkeit eines anderen Objektes j zum Cluster C_k (ausgenommen gewisse restriktive Clustermodelle, bei denen z.B. die Zahl der Elemente in den Clustern a priori vorgegeben ist), erhält man als Wahrscheinlichkeit der Zugehörigkeit zweier verschiedener Objekte i und j zum Cluster C_k

$$\text{Prob}(i \in C_k \wedge j \in C_k) = \text{Prob}(i \in C_k) \cdot \text{Prob}(j \in C_k),$$

$$i \neq j, \quad i, j = 1, \dots, n, \quad k = 1, \dots, p.$$

Da die Mitgliedschaft zweier Objekte i und j zu einem Cluster C_k die Mitgliedschaft der beiden Objekte zu anderen Clustern ausschließt, gilt wiederum der Additionssatz für Wahrscheinlichkeiten unverträglicher Ereignisse

$$\begin{aligned} & \text{Prob}\left(\bigcup_{k=1}^p (i \in C_k \wedge j \in C_k)\right) \\ &= \sum_{k=1}^p \text{Prob}(i \in C_k \wedge j \in C_k) = \\ &= \sum_{k=1}^p \text{Prob}(i \in C_k) \cdot \text{Prob}(j \in C_k) \end{aligned}$$

bzw.

$$r_{ij} = \sum_{k=1}^p y_{ik}y_{jk}, \quad i \neq j, \quad i = 1, \dots, n. \quad (8)$$

Die Außerdiagonalelemente r_{ij} , $i \neq j$, der nichtnegativen symmetrischen $n \times n$ -Matrix $R = (r_{ij})$ sind damit interpretierbar als Wahrscheinlichkeiten dafür, daß die korrespondierenden Zeilen- und Spaltenobjekte i und j einem der p Cluster gemeinsam angehören. Man bestätigt leicht arithmetisch, daß $0 \leq r_{ij} \leq 1$ gilt:

Wegen $0 \leq y_{ik} \leq 1$ gilt $y_{ik}^2 \leq y_{ik}$ und daher mit (7b)

$$\sum_{k=1}^p y_{ik}^2 \leq \sum_{k=1}^p y_{ik} \leq 1, \quad i = 1, \dots, n.$$

Mit der CAUCHY-SCHWARZ-Ungleichung erhält man

$$\sum_{k=1}^p y_{ik}y_{jk} \leq \sqrt{\sum_{k=1}^p y_{ik}^2} \sqrt{\sum_{k=1}^p y_{jk}^2} \leq 1, \quad i, j = 1, \dots, n.$$

Zur Interpretation der Diagonalelemente r_{ii} von R bieten sich mindestens drei Alternativen an:

(a) Die Diagonalelemente r_{ii} werden als quadrierte Länge der Wahrscheinlichkeits-Vektoren $Y_i = (y_{i1}, \dots, y_{ip})'$ definiert

$$r_{ii} = \sum_{k=1}^p y_{ik}^2, \quad i = 1, \dots, n. \quad (9a)$$

(b) Die Diagonalelemente r_{ii} werden definiert als Wahrscheinlichkeiten dafür, daß Objekt i einem der p Cluster angehört

$$r_{ii} = \sum_{k=1}^p y_{ik}, \quad i = 1, \dots, n. \quad (9b)$$

(c) Den Diagonalelementen r_{ii} wird keine empirische Bedeutung zugrundegelegt. Sie werden in der Analyse nicht berücksichtigt.

Sei $U = (u_{ij})$ eine gegebene nichtnegative, symmetrische $n \times n$ -Matrix, deren Außerdiagonalelemente u_{ij} , $i \neq j$, interpretierbar sind als Wahrscheinlichkeiten dafür, daß zwei verschiedene Objekte $i \neq j$ einem von p Clustern gemeinsam angehören und deren Diagonalelemente u_{ii}

- (a) als quadrierte Längen von Wahrscheinlichkeits-Vektoren $Y_i = (y_{i1}, \dots, y_{ip})'$ interpretierbar sind;
- (b) als Wahrscheinlichkeiten dafür, daß Objekt i einem der p Cluster angehört, interpretierbar sind;
- (c) nicht näher spezifiziert sind (missing elements).

Obwohl der Skalarprodukt-Relation (8) der Außerdiagonal-Elemente entsprechend, scheint die empirische Erhebung von Diagonalelementen u_{ii} im Falle (a) problematisch zu sein. Im Falle (b) kann häufig $u_{ii} = 1$, $i = 1, \dots, n$, verlangt werden, und im Falle (c) braucht die Diagonale von U nicht gegeben zu sein.

Ziel einer Clusteranalyse ist dann, eine $n \times p$ -Clusterstruktur $Y = (y_{ik})$ mit den Eigenschaften (7) so zu bestimmen, daß ein geeignetes Least-Squares-Fitkriterium $\sigma(Y)$ minimiert wird:

$$\begin{aligned} (a) \quad \sigma_a(Y) &= \text{Spur}(U - R)'(U - R) = \\ &= \sum_{i=1}^n \sum_{j=1}^n (u_{ij} - r_{ij}(Y))^2 = \end{aligned} \quad (10a)$$

$$\begin{aligned} &= \sum_{i=1}^n \sum_{j=1}^n (u_{ij} - Y_i'Y_j)^2; \\ (b) \quad \sigma_b(Y) &= \text{Spur}(U - R)'(U - R) = \\ &= \sum_{i=1}^n \sum_{j=1}^n (u_{ij} - r_{ij}(Y))^2 = \\ &= \sum_{i \neq j}^n \sum_{j=1}^n (u_{ij} - Y_i'Y_j)^2 + \\ &\quad + \sum_{i=1}^n (u_{ii} - Y_i'J)^2 \quad \text{mit } J = (1, \dots, 1)'; \end{aligned} \quad (10b)$$

$$\begin{aligned} (c) \quad \sigma_c(Y) &= \sum_{i \neq j}^n \sum_{j=1}^n (u_{ij} - r_{ij}(Y))^2 = \\ &= \sum_{i \neq j}^n \sum_{j=1}^n (u_{ij} - Y_i'Y_j)^2. \end{aligned} \quad (10c)$$

Die Schätzung der Clusterstruktur $Y = (y_{ik})$ scheint besonders problematisch aufgrund der Restriktionen (7a) und (7b). Im Falle (b) wird eine Lösung $Y = (y_{ik})$, die $\sigma_b(Y)$ minimiert, die Bedingung (7b) zumindest genähert befriedigen, da sie direkt im zweiten Teil des Fitkriteriums verlangt wird. Problematischer scheint die Nichtnegativitäts-Restriktion (7a), die das Problem der Schätzung der Clusterstruktur Y wesentlich von dem der Faktorisierung der Matrix U (Hauptkomponentenanalyse) unterscheidet.

A.5. Metrische Skalarproduktmodelle zur numerischen Datenanalyse

Das im vorigen Abschnitt wahrscheinlichkeitstheoretisch begründete Modell kann für Zwecke der Datenanalyse erheblich abgeschwächt werden. Sei $U = (u_{ij})$ eine gegebene nichtnegative, symmetrische (diagonal dominante) $n \times n$ -Proximitätsmatrix, deren Elemente u_{ij} um so größer sind, je ähnlicher die korrespondierenden Objekte i und j einander sind. Dann wird eine nichtnegative $n \times p$ -Clusterstruktur $Y = (y_{ik})$ derart gesucht, daß die Skalarprodukt-Modellwerte $r_{ij}(Y)$ möglichst gut im Sinne eines der drei Fitkriterien $\sigma_a(Y)$, $\sigma_b(Y)$, $\sigma_c(Y)$ die korrespondierenden gegebenen Proximitäten reproduzieren.

Modell III.a:

$$u_{ij} \approx r_{ij} = \sum_{k=1}^p y_{ik}y_{jk} \quad \text{mit } y_{ik} \geq 0, \quad (11a)$$

für alle $i, j = 1, \dots, n, k = 1, \dots, p$.

$$\text{Modell III.b: } u_{ij} \approx r_{ij} = \begin{cases} \sum_{k=1}^p y_{ik}y_{jk} & \text{für } i \neq j \text{ mit } y_{ik} \geq 0, \\ \sum_{k=1}^p y_{ik} & \text{für } i = j \text{ mit } y_{ik} \geq 0, \end{cases} \quad (11b)$$

für alle $i, j = 1, \dots, n, k = 1, \dots, p$.

Modell III.c:

$$u_{ij} \approx r_{ij} = \sum_{k=1}^p y_{ik}y_{jk} \quad \text{für } i \neq j \text{ mit } y_{ik} \geq 0, \quad (11c)$$

mit $i, j = 1, \dots, n, k = 1, \dots, p$.

Die Skalarprodukt-Werte $r_{ij}(Y)$, $i \neq j$, sind umso größer, je größer die Cluster-, „Ladungen“ y_{ik} und y_{jk} zweier Objekte i und j auf gleichen Clustern C_k , $k = 1, \dots, p$, sind. Die Elemente y_{ik} der $n \times p$ -Matrix Y können dann als Mitgliedschafts-Gewichte bzw. Bedeutungen interpretiert werden, die das Objekt i dem Cluster C_k zuweist. Diese Form eines Clusteranalyse-Modelles ist im wesentlichen äquivalent einem restriktiven Modell der Matrixfaktorisierung (Hauptkomponentenanalyse), wobei die Nichtnegativitäts-Restriktionen der Clusterladungen y_{ik} orthogonale Clusterkomponenten i.a. ausschließt.

Eine obere Schranke für die nichtnegativen Clusterladungen (vgl. Bedingung $0 \leq y_{ik} \leq 1$ beim Wahrscheinlichkeits-Modell) braucht hier nicht beachtet zu werden, da sich die relative Größe der Proximitäten u_{ij} aus der relativen Größe der Clusterladungen y_{ik} ergibt und umgekehrt. Mit einer positiven Konstante K gilt

$$K^2 u_{ij} \approx K^2 r_{ij}(Y) = \sum_{k=1}^p (Ky_{ik})(Ky_{jk}), \quad K > 0, \quad (12)$$

d.h. eine Dilation der Clusterladungen bez. K ist verbunden mit einer Dilation der Skalarprodukt-Modellwerte (und damit der gegebenen Proximitäten) bez. K^2 . Die Restriktion (7b) scheint für Zwecke der Datenanalyse nicht bedeutsam zu sein. Bedingung (7b) wird zudem genähert befriedigt, wenn Kriterium $\sigma_b(Y)$ minimiert wird.

B: Cluster Methoden

B.1. Methode zur Bestimmung maximal kompakter Gruppen

Von CATTELL & COULTER (1966), sowie CATTELL, COULTER & TSUJIOKA (1966) wurde eine Methode zur Aufdeckung sogenannter maximal kompakter Gruppen (MKG), dort genannt Phenomenal Cluster, Ph.Cl.) auf Grundlage einer gegebenen (0,1)-binären ILM entwickelt, die im deutschsprachigen Raum besonders durch das Buch von Urs BAUMANN (1971) bekannt wurde.

Bezeichnung:

Eine Teilmenge $C_k = \{k_1, \dots, n_{nk}\}$ der Objektmenge $Q = \{1, \dots, n\}$ nennt man eine MKG bez. einer gegebenen $n \times n$ -ILM $T = (t_{ij})$, wenn sie folgende Eigenschaften besitzt:

(P1) Alle Elemente der MKG C_k sind bez. der ILM $T = (t_{ij})$ paarweise ähnlich, d.h. für alle Paare $(k_i, k_j) \in C_k \times C_k$ gilt $t_{k_i, k_j} = 1$.

(P2) Alle zur Restmenge $\bar{C}_k = Q \setminus C_k$ gehörenden Elemente müssen mindestens einem der zu C_k gehörenden Elemente bez. der ILM $T = (t_{ij})$ unähnlich sein, d.h. für jedes Element $k_j \in \bar{C}_k$ existiert wenigstens ein Element $k_i \in C_k$, für das gilt $t_{k_i, k_j} = 0$.

Eine Teilmenge C_k , für die nur die Eigenschaft (P1) gefordert wird, nennt man kompakte Gruppe (KG).

Sei $X_k = (x_{1k}, \dots, x_{nk})'$ der (0,1)-binäre Zugehörigkeits-Vektor einer MKG C_k bez. $T = (t_{ij})$, dann sind die Eigenschaften (P1), (P2) auch beschreibbar in der Form:

(P1) Für alle Paare (i, j) für die $x_{ik} = x_{jk} = 1$ gilt, gilt auch $t_{ij} = 1$.

(P2) Für jedes Objekt j mit $x_{jk} = 0$ existiert wenigstens ein Objekt i mit $x_{ik} = 1$ und $t_{ij} = 0$.

Jede symmetrische, diagonaldominante (0,1)-binäre $n \times n$ -Matrix $T = (t_{ij})$ beschreibt eine bestimmte Menge von N verschiedenen, sich im allgemeinen überlappenden MKG C_k , die als Spaltenvektoren X_k in einer $n \times N$ -Zuordnungs-Matrix $X = (x_{ik})$ zusammengefaßt werden können. In ungünstigen Fällen kann dabei die Anzahl N der existierenden MKG sehr groß sein ($N > n$). Wegen $t_{ii} \geq t_{ij} = t_{ji}$, $j = 1, \dots, n$, darf ein Diagonalelement t_{ii} von T nur dann verschwinden, wenn alle anderen Elemente aus Zeile i und Spalte i auch verschwinden. Dann ist das Objekt i in keiner MKG enthalten und kann aus der Clusteranalyse ausgeschlossen werden. Falls $t_{ii} = 1$ und $t_{ij} = 0$ für alle $j \neq i$, dann bildet Objekt i ein singuläres Cluster (separates oder isoliertes Cluster).

Mit der Gesamtmenge der N MKG wird stets ein perfekter Fit, $\sigma(X) = 0$, der gegebenen ILM $T = (t_{ij})$ im Sinne des logischen Skalarprodukt-Modelles erreicht

$$\sigma(X) = \text{Spur}(T - S)'(T - S) =$$

$$= \sum_{i=1}^n \sum_{j=1}^n (t_{ij} - s_{ij}(X))^2 \quad (1)$$

$$\text{mit } s_{ij}(X) = \bigcup_{k=1}^N x_{ik} \cap x_{jk} = X'_i \cap X_j$$

Von HARTMANN (1976a) wurde gezeigt, daß die Methode von CATTELL & COULTER (vgl. BAUMANN, 1971) im allgemeinen nicht fähig ist, sämtliche N MKG aufzudecken. Häufig erreicht sie zwar einen perfekten Fit, $\sigma(X) = 0$, aber die Menge der aufgedeckten Cluster hängt im allgemeinen von der Reihenfolge der N Objekte in der ILM $T = (t_{ij})$ ab. Bei HARTMANN (1976b) wurde ein Algorithmus vorgeschlagen, der zwar stets die vollständige Menge aller N MKG einer ILM $T = (t_{ij})$ aufdeckt, der aber rechentechnisch nicht so effizient ist wie der im folgenden beschriebene und von Peter QUAAS, TU Dresden, entwickelte Algorithmus:

(a) In jedem Schritt i des Algorithmus werden die MKG gebildet, die sich aus den Objekten $1, \dots, i$ ergeben aufgrund der entsprechenden $i \times i$ -Teil-ILM.

Die MKG für die Objekte $1, \dots, i, i+1$ entstehen durch Erweiterung der MKG der Objekte $1, \dots, i$.

Die Analyse beginnt mit den ersten beiden Objekten und endet, wenn alle n Objekte abgearbeitet wurden, besteht also aus insgesamt $n - 1$ Schritten.

(b) Falls die Objekte 1 und 2 ähnlich sind, $t_{21} = 1$, bilden sie eine Gruppe $\{1, 2\}$, falls sie nicht ähnlich sind, $t_{21} = 0$, bilden sie zwei singuläre Gruppen $\{1\}$ und $\{2\}$.

(c) Angenommen, es liegen die MKG für die Objekte $1, \dots, i$ vor. Dann werden die Ähnlichkeiten (Links) des Objektes $i+1$ zu den Objekten $1, \dots, i$ betrachtet.

(c.1) Falls keine Links von Objekt $i+1$ zu den Objekten $1, \dots, i$ bestehen, dann wird Objekt $i+1$ als singuläre Gruppe $\{i+1\}$ notiert.

(c.2) Sonst werden alle vorliegenden Gruppen geprüft:

(c.2.1) Falls Objekt $i+1$ allen Objekten einer vorliegenden Gruppe ähnlich ist, wird diese MKG um Objekt $i+1$ vergrößert.

(c.2.2) Falls Objekt $i+1$ nur zu einem Teil der Elemente einer vorliegenden Gruppe ähnlich ist, wird eine neue Gruppe gebildet, bestehend aus Objekt $i+1$ und den Elementen der vorliegenden Gruppe, die zu $i+1$ ähnlich sind.

(c.3) Jede der mittels (c.2.2) neu entstandenen Gruppen wird geprüft, ob sie Teilmenge einer anderen Gruppe mit Objekt $i+1$ ist, d.h. einer Gruppe, die mittels (c.2.1) oder (c.2.2) gebildet wurde.

Mit Hilfe der logischen Durchschnitts- und Vereinigungs-Relationen zwischen den (0,1)-binären Clusterstrukturen X_k und den Spalten T_j der Matrix T ist der Algorithmus außerordentlich effizient programmierbar. Bis zu etwa $n = 30$ Objekten ist er auch durchaus mit Bleistift und Papier durchführbar. Einer der größten Nachteile des Modells ist jedoch die im allgemeinen sehr große Anzahl N sich teilweise stark überlappender Cluster, die für einen perfekten Fit des Modelles (1) nötig ist.

B.2. Strategien zur Vereinigung maximal kompakter Gruppen

B.2.1. Motivation

Es sind verschiedene Strategien denkbar, ausgehend von der vollständigen Menge der N MKG zu einer praktikablen Menge von p (überlappenden) Clustern zu gelangen, die einerseits einen möglichst guten Fit bez. $\sigma(X)$ garantieren und andererseits eine möglichst sparsame und relevante klassifikatorische Interpretation einer gegebenen ILM erlauben.

1. Durch ersatzloses Weglassen von $N - p$ MKG entsteht eine $n \times p$ -Zuordnungs-Matrix $X = (x_{ik})$, die mit (1) eine Modellwerte-Matrix $S = (s_{ij})$ erzeugt, die im allgemeinen nur eine Teilmenge der Links der gegebenen ILM $T = (t_{ij})$ enthält. Um einen möglichst guten Fit bez. $\sigma(X)$ zu erreichen, sollten die $N - p$ MKG gestrichen werden, die einen möglichst geringen Verlust

an Links in $S = (s_{ij})$ gegenüber den gegebenen Links in $T = (t_{ij})$ bewirken.

2. Durch Vereinigung von MKG entsteht eine $n \times p$ -Zuordnungs-Matrix $X = (x_{ik})$, die mit (1) eine Modellwerte-Matrix $S = (s_{ij})$ erzeugt, die die Links der gegebenen ILM $T = (t_{ij})$ als Teilmenge enthält. Um einen möglichst guten Fit bez. $\sigma(X)$ zu erreichen, sollten solche MKG vereinigt werden, die einen möglichst geringen Zuwachs an Links in $S = (s_{ij})$ gegenüber den gegebenen Links in $T = (t_{ij})$ bewirken.

Natürlich sind auch Strategien denkbar, bei denen sowohl MKG weggelassen als auch vereinigt werden. Da die MKG häufig sehr ähnlich oder gar stark überlappend sind, wurde bisher vor allem die zweite Alternative der Clustervereinigung bevorzugt.

Wenn als einziges Kriterium zur Vereinigung von Gruppen zu Cliques der minimale Link-Zuwachs in der reproduzierten ILM $S = (s_{ij})$ gegenüber der gegebenen ILM $T = (t_{ij})$ verwendet würde, d.h. schrittweise solche Gruppen verwendet würden, die einen minimalen Fitverlust bez. $\sigma(X)$ bewirken, dann hätte diese Strategie den offensichtlichen Nachteil, daß kleine (bzw. singuläre) Gruppen bevorzugt vereinigt würden, auch wenn sie keine oder relativ wenige Links verbinden. Größere Gruppen hätten geringere Chancen vereinigt zu werden, auch wenn sie hoch ähnlich oder gar stark überlappend sind. Häufig wären die so entstehenden Cliques ziemlich inhomogen.

Im folgenden werden drei Strategien beschrieben, mit denen eine größere Anzahl von N MKG nach bestimmten Auswahlkriterien zu einer kleineren Anzahl von Cliques vereinigt werden können. Die damit verbundenen Algorithmen und Prinzipien ähneln einerseits dem von CATTELL & COULTER (1966) und andererseits denen der (sukzessiven) hierarchischen Clusteranalyse (vgl. JOHNSON, 1967).

B.2.2. Strategie der Clusterüberlappung

Kriterium:

Es werden solche MKG (Phenomenal Cluster) zu Cliques (Segregates) vereinigt, die eine Mindestanzahl gemeinsamer Objekte (überlappende Elemente, Nuclear Cluster) enthalten.

Algorithmus:

(a) Konstruktion der $N \times N$ -Überlappungs-Matrix $U = (u_{kl})$ aller Clusterpaare (C_k, C_l) :

$$u_{kl} = \sum_{i=1}^n x_{ik} x_{il} = X'_k X_l, \quad k, l = 1, \dots, N, \quad (2)$$

u_{kl} : Zahl der gemeinsamen Elemente der beiden MKG C_k und C_l . (Für die Diagonalelemente u_{kk} gilt natürlich $u_{kk} = n_k$, wobei n_k die Anzahl der Elemente der MKG C_k , $k = 1, \dots, N$, bezeichnet.)

(b) Schwellwert-Dichotomisierung der Überlappungs-Matrix $U = (u_{kl})$ zur (0,1)-binären Contiguity-Matrix $\bar{U} = (\bar{u}_{kl})$:

$$\bar{u}_{kl} = \begin{cases} 1, & \text{falls } u_{kl} \geq \text{CONT}, \\ 0, & \text{falls } u_{kl} < \text{CONT}, \end{cases} \quad k, l = 1, \dots, N, \quad (3)$$

CONT: gegebener Dichotomisierungs-Parameter, CONT > 0.

- (c) Vereinigung der MKG-Paare mit mindestens CONT gemeinsamen Elementen;

(c.1) Bestimmung einer $N \times p$ -Zugehörigkeits-Matrix $V = (v_{ks})$ der N MKG zu p Segregates durch sukzessive Vereinigung der Spalten von $\bar{U} = (\bar{u}_{kl})$ mit nichtleerem Durchschnitt.

(c.2) Bestimmung einer $n \times p$ -Zugehörigkeits-Matrix $W = (w_{is})$ der n Objekte zu p Segregates durch Vereinigung der entsprechenden Spalten X_k der $N \times p$ -Zugehörigkeits-Matrix $X = (x_{ik})$ der Objekte zu den MKG.

Bemerkungen:

- Das Kriterium zur Vereinigung von MKG zu Cliques berücksichtigt nicht die totale Anzahl der Links, die zwischen den Elementen der beiden Cluster aufgrund der gegebenen ILM $T = (t_{ij})$ bestehen. Die Zahl der überlappenden Elemente zweier Cluster ist daher im allgemeinen kein Maß der Ähnlichkeit der zwei Cluster.
- Das Kriterium zur Vereinigung von MKG zu Cliques berücksichtigt nicht die Zahl der Nicht-Links, die zwischen den Elementen der Clique nach Vereinigung aufgrund der gegebenen ILM $T = (t_{ij})$ bestehen. Die Zahl der überlappenden Elemente zweier Cluster ist daher im allgemeinen kein Maß der Homogenität der entstehenden Clique.
- Aufgrund des fest vorgegebenen ganzzahligen Überlappungs-Schwellwertes CONT werden große Cluster bevorzugt zu Cliques vereinigt.

B.2.3. Strategie der Clusterähnlichkeit

Kriterium:

Es werden (sukzessive) solche Cluster zu Cliques vereinigt, die einander jeweils am ähnlichsten sind, d.h. zwischen denen die relative (bzw. prozentuale) Linkanzahl in der gegebenen ILM $T = (t_{ij})$ maximal ist bzw. einen Schwellwert übersteigt.

Algorithmus:

(a) Konstruktion der $N \times N$ -Cluster-Ähnlichkeits-Matrix $U = (u_{kl})$:

(a.1) Seien C_k und C_l zwei Cluster mit n_k bzw. n_l Elementen und binären Zugehörigkeitsvektoren X_k bzw. X_l . Dann enthält die $(0,1)$ -binäre $n \times n$ -Matrix $X_k X_l'$ Elemente $x_{ik} x_{jl}$, die genau dann gleich Eins sind, wenn Objekt i zu Cluster C_k und Objekt j zu Cluster C_l gehört. Die $n \times n$ -Matrix enthält also genau $n_k n_l$ Links in n_k Zeilen und n_l -Spalten. Die Links in der $(0,1)$ -binären $n \times n$ -Matrix $T \cap (X_k X_l')$ zeigen an, daß zwischen einem zu Cluster C_k gehörenden Objekt i und einem zu Cluster C_l gehörenden Objekt j ein Link in der gegebenen ILM $T = (t_{ij})$ besteht. Ein relatives (bzw.

prozentuales) Maß der Ähnlichkeit der beiden Cluster C_k und C_l wird erhalten, wenn die Linkanzahl von $T \cap (X_k X_l')$ durch die maximal mögliche Linkanzahl $n_k n_l$ zwischen beiden Clustern dividiert wird.

(a.2) In Summations-Notation gilt für die relative Ähnlichkeit zweier Cluster C_k und C_l

$$u_{kl} = \frac{1}{n_k n_l} \sum_{i \in C_k} \sum_{j \in C_l} t_{ij} = \frac{1}{n_k n_l} \sum_{i=1}^n \sum_{j=1}^n t_{ij} x_{ik} x_{jl}, \quad (4)$$

$$k, l = 1, \dots, N.$$

Matrix $U = (u_{kl})$ ist symmetrisch und es gilt $0 \leq u_{kl} \leq 1$. Dabei ist $u_{kl} = 0$, falls zwischen keinem der Objektpaare $i \in C_k$ und $j \in C_l$ ein Link in $T = (t_{ij})$ existiert, und es ist $u_{kl} = 1$, falls zwischen allen Objektpaaren $i \in C_k$ und $j \in C_l$ Links in $T = (t_{ij})$ existieren. Die Diagonalelemente u_{kk} sind nur dann gleich Eins, wenn die korrespondierenden Cluster MKG sind. Die Größe eines Diagonalelementes u_{kk} ist ein Maß der Homogenität des korrespondierenden Clusters C_k aufgrund der gegebenen ILM $T = (t_{ij})$.

(b) Schwellwert-Vereinigungs-Algorithmus:

(b.1) Schwellwert-Dichotomisierung der $N \times N$ -Cluster-Ähnlichkeits-Matrix $U = (u_{kl})$ zur $(0,1)$ -binären Contiguity-Matrix $\bar{U} = (\bar{u}_{kl})$ aufgrund eines gegebenen metrischen Schwellwertes CONT, $0 \leq \text{CONT} \leq 1$, entsprechend Schritt (b) der Strategie von CATTELL & COULTER (1966).

(b.2) Bestimmung einer $N \times p$ -Zugehörigkeits-Matrix $V = (v_{ks})$ der N Cluster zu p Cliques und Bestimmung einer $n \times p$ -Zugehörigkeits-Matrix $W = (w_{is})$ der n Objekte zu p Cliques, entsprechend Schritt (c) der Strategie von CATTELL & COULTER (1966).

(c) Sukzessiver-Vereinigungs-Algorithmus:

(c.1) Es werden die zwei Cluster C_k und C_l , $k \neq l$, vereinigt, deren relative Ähnlichkeit u_{kl} maximal ist.

(c.2) In der Zugehörigkeits-Matrix $X = (x_{ik})$ wird anstelle des Clusters C_k das Cluster $C_k \cup C_l$ notiert und das Cluster C_l gestrichen.

(c.3) In der Cluster-Ähnlichkeits-Matrix $U = (u_{kl})$ werden die k -te Zeile und Spalte neu berechnet und die l -te Zeile und Spalte gestrichen.

(c.4) Die Anzahl N der Cluster wird um Eins verringert und falls N noch größer als die gewünschte Clusteranzahl p ist, wird der Vorgang bei (c.1) wiederholt.

B.2.4 Strategie der Clusterhomogenität

Kriterium:

Es werden (sukzessive) solche Cluster zu Cliques vereinigt, die erwarten lassen, daß die entstehende Clique am homogensten ist, d.h. die nach der Vereinigung eine maximale relative (bzw. prozentuale) Linkanzahl aufgrund der gegebenen ILM $T = (t_{ij})$ aufweisen wird.

Algorithmus:

(a) Konstruktion der $N \times N$ -Cluster-Homogenitäts-Matrix $U = (u_{kl})$:

(a.1) Sei $C_{kl} = C_k \cup C_l$ die Vereinigung zweier Cluster

C_k und C_l mit n_{kl} verschiedenen Elementen und dem binären Zugehörigkeits-Vektor $X_k \cup X_l$. Dann enthält die (0,1)-binäre $n \times n$ -Matrix $(X_k \cup X_l) (X_k \cup X_l)'$ Elemente $(x_{ik} \cup x_{il}) (x_{jk} \cup x_{jl})$, die genau dann gleich Eins sind, wenn die beiden Objekte i und j zur Clique $C_{kl} = C_k \cup C_l$ gehören. Die symmetrische $n \times n$ -Matrix enthält also genau n_{kl}^2 Links in n_{kl} Zeilen und n_{kl} Spalten. Die Links in der (0,1)-binären $n \times n$ -Matrix

$$T \cap ((X_k \cup X_l) (X_k \cup X_l)') \quad (5)$$

zeigen an, daß zwischen zwei zur Clique C_{kl} gehörenden Objekten i und j ein Link in der gegebenen ILM $T = (t_{ij})$ besteht. Ein relatives (bzw. prozentuales) Maß der Homogenität der Clique C_{kl} wird erhalten, wenn die Linkanzahl von (5) durch die maximal mögliche Linkanzahl n_{kl}^2 zwischen den n_{kl} Elementen der Clique dividiert wird.

(a.2) In Summations-Notation gilt für die relative Homogenität einer Clique $C_{kl} = C_k \cup C_l$:

$$u_{kl} = \frac{1}{n_{kl}^2} \sum_{i \in C_{kl}} \sum_{j \in C_{kl}} t_{ij} = \frac{1}{n_{kl}^2} \sum_{i=1}^n \sum_{j=1}^n t_{ij} (x_{ik} \cup x_{il}) (x_{jk} \cup x_{jl})$$

$k, l = 1, \dots, N,$ (6)

wobei $x_{ik} \cup x_{il}$, $i = 1, \dots, n$, die Elemente des Zugehörigkeits-Vektors $X_k \cup X_l$ der Clique $C_{kl} = C_k \cup C_l$ sind. Matrix $U = (u_{kl})$ ist symmetrisch und es gilt $0 < u_{kl} \leq 1$. Dabei ist $u_{kl} = 1$, falls zwischen allen Objektpaaren $i \in C_{kl}$ und $j \in C_{kl}$ Links in $T = (t_{ij})$ bestehen, d.h. falls die Clique C_{kl} kompakte Gruppe ist.

(b) Vereinigungs-Algorithmen:

Entsprechend den Punkten (b) und (c) der Strategie der maximalen Clusterähnlichkeit sind auch hier sowohl ein Schwellwert-Vereinigungs-Algorithmus als auch ein Sukzessiver-Vereinigungs-Algorithmus denkbar. Dabei wird analog zu den dort gegebenen Ausführungen verfahren, wobei der Begriff der Cluster-Ähnlichkeit hier aber durch den der erwarteten Cluster-Homogenität ersetzt wird.

B.3. Methode zur Analyse disjunkter Cluster

Ausgehend von einer vollständigen Menge von MKG einer gegebenen ILM $T = (t_{ij})$, die beispielsweise mit der in B.1. beschriebenen Methode bestimmt wurde, ermittelt der im folgenden vorgeschlagene Algorithmus eine Struktur disjunkter Cluster, die unter bestimmten Voraussetzungen einen relativ guten Fit des Least-Squares-Kriteriums $\sigma(X)$ repräsentiert. Der Algorithmus besteht aus folgenden Teilschritten:

(a) Die größte Menge C_k wird als erstes Cluster D_1 der disjunkten Clusterstruktur definiert.

(b) Im p -ten Schritt des Algorithmus liegen p disjunkte Cluster D_1, \dots, D_p vor, die durch Behandlung von p MKG C_{k1}, \dots, C_{kp} zusammengestellt wurden. Die disjunkten Cluster D_k enthalten m_k , $k = 1, \dots, p$, Elemente. Die Objektmengen D_0 bzw. \bar{D}_0

$$D_0 = \bigcup_{k=1}^p D_k \quad \text{bzw.} \quad \bar{D}_0 = Q \setminus D_0 \quad (7)$$

enthalten die den disjunkten Clustern bereits zugeordneten bzw. noch nicht zugeordneten Objekte der Gesamt-Objekt-Menge $Q = \{1, \dots, n\}$.

(c) Im $(p+1)$ ten Schritt des Algorithmus wird aus den noch nicht behandelten MKG $C_{k,p+1}$ ausgewählt, die die meisten noch nicht D_0 zugeordneten Objekte enthält, d.h. für die die Elementanzahl der Menge

$$D_{p+1} = \bar{D}_0 \cap C_{k,p+1} \quad (8)$$

maximal ist. Der nichtzugeordnete Teil des Clusters $C_{k,p+1}$ bildet also das $(p+1)$ te Cluster D_{p+1} mit m_{p+1} Elementen.

(d) Wenn alle in den MKG enthaltenen Objekte auch Objekte der disjunkten Cluster sind, endet der Algorithmus mit p vorläufigen disjunkten Clustern D_1, \dots, D_p nach p Schritten.

(e) Nachbehandlung der vorläufigen disjunkten Clusterstruktur: Wenn mittels Vereinigung von jeweils zwei disjunkten Clustern der Wert des Least-Squares-Fitkriteriums $\sigma(X)$ verringert werden kann, dann wird durch sukzessive Vereinigung solcher Clusterpaare die Zahl der disjunkten Cluster verringert und der Fit der Clusterstruktur bez. der gegebenen ILM $T = (t_{ij})$ verbessert. Nach Berechnung einer $p \times p$ -Matrix $\Delta = (\delta_{kl})$ der Veränderungen δ_{kl} des LS-Fitwertes $\sigma(X)$ bei Vereinigung zweier disjunkter Cluster D_k und D_l wird ein Algorithmus äquivalent zu dem in B.2. vorgeschlagenen Sukzessiven-Vereinigungs-Algorithmus empfohlen.

Bemerkungen:

1. Veränderung des Wertes von $\sigma(X)$ in den p Auswahl-schritten des Algorithmus:

Da die p Cluster D_k disjunkt und als Teilmengen von MKG kompakte Gruppen sind, gelten die folgenden Relationen

$$\sum_{i \in D_k} \sum_{j \in D_k} (t_{ij} - \sum_{l=1}^p x_{il} x_{jl})^2 = \sum_{i \in D_k} \sum_{j \in D_k} (t_{ik} - 1)^2 = 0,$$

$$\sum_{i \in D_k} \sum_{j \notin D_k} (t_{ij} - \sum_{l=1}^p x_{il} x_{jl})^2 = \sum_{i \in D_k} \sum_{j \notin D_k} t_{ij}^2, \quad (9)$$

$$\sum_{i \in \bar{D}_0} \sum_{j=1}^n (t_{ij} - \sum_{l=1}^p x_{il} x_{jl})^2 = \sum_{i \in \bar{D}_0} \sum_{j=1}^n t_{ij}^2,$$

$$\sum_{j \notin D_k} t_{ij}^2 = \sum_{j=1}^n t_{ij}^2 - \sum_{j \in D_k} t_{ij}^2, \quad i = 1, \dots, n, k = 1, \dots, p.$$

Andererseits ist das Least-Squares-Fitkriterium

$$\sigma(X) = \sum_{i=1}^n \sum_{j=1}^n (t_{ij} - \sum_{l=1}^p x_{il} x_{jl})^2$$

für disjunkte Cluster D_1, \dots, D_p zerlegbar in der Form

$$\sigma(X) = \sum_{k=1}^p \left(\sum_{i \in D_k} \sum_{j \in D_k} (t_{ij} - \sum_{l=1}^p x_{il} x_{jl})^2 \right) \quad (10)$$

$$+ \sum_{i \in D_k} \sum_{j \notin D_k} (t_{ij} - \sum_{l=1}^p x_{il} x_{jl})^2$$

$$\begin{aligned}
& + \sum_{i \in D_0} \sum_{j=1}^n (t_{ij} - \sum_{i=1}^p x_{ii} x_{ij})^2 \\
& = \sum_{k=1}^p \sum_{i \in D_k} \sum_{j=1}^n t_{ij}^2 + \sum_{i \in D_0} \sum_{j=1}^n t_{ij}^2 \\
& - \sum_{k=1}^p \sum_{i \in D_k} \sum_{j \in D_k} t_{ij}^2 \\
& = \sum_{i=1}^n \sum_{j=1}^n t_{ij}^2 - \sum_{k=1}^p \sum_{i \in D_k} \sum_{j \in D_k} t_{ij}^2 \\
& = \sum_{i=1}^n \sum_{j=1}^n t_{ij}^2 - \sum_{k=1}^p m_k^2,
\end{aligned}$$

wobei m_k die Anzahl der Elemente des Clusters D_k bezeichnet. Vor dem ersten Teilschritte des Algorithmus beträgt der Wert des LS-Fitkriteriums

$$\sigma_0(X) = \sum_{i=1}^n \sum_{j=1}^n t_{ij}^2 \quad (11)$$

und nach dem p -ten Auswahlsschritt beträgt der Wert des LS-Fitkriteriums

$$\begin{aligned}
\sigma_p(X) &= \sigma_0(X) - \sum_{k=1}^p \sum_{i \in D_k} \sum_{j \in D_k} t_{ij}^2 = \\
&= \sigma_0(X) - \sum_{k=1}^p m_k^2. \quad (12)
\end{aligned}$$

Bei jedem Auswahlsschritt p des Algorithmus verringert sich der Wert des LS-Fitkriteriums um

$$\sum_{i \in D_p} \sum_{j \in D_p} t_{ij}^2 = m_p^2. \quad (13)$$

Das begründet, warum in jedem Auswahlsschritt p eine kompakte Gruppe D_p mit möglichst vielen Elementen m_p ausgewählt wird.

2. Veränderung des Wertes von $\sigma(X)$ bei Vereinigung zweier disjunkter Cluster D_k und D_l :

a) Sei $\sigma_a(X)$ das LS-Fitkriterium für den Fall, das die beiden Cluster D_k und D_l noch nicht vereinigt sind. Dann ist $\sigma_a(X)$ zerlegbar in der Form

$$\begin{aligned}
\sigma_a(X) &= \sum_{i \in D_k} \left(\sum_{j \in D_k} (t_{ij} - 1)^2 + \sum_{j \in D_l} t_{ij}^2 + \sum_{j \in D_{k \cup D_l}} t_{ij}^2 \right) \\
& + \sum_{i \in D_l} \left(\sum_{j \in D_k} t_{ij}^2 + \sum_{j \in D_l} (t_{ij} - 1)^2 + \sum_{j \in D_{k \cup D_l}} t_{ij}^2 \right) \quad (14a)
\end{aligned}$$

$$+ \sum_{i \in D_{k \cup D_l}} \left(\sum_{j \in D_k} t_{ij}^2 + \sum_{j \in D_l} t_{ij}^2 + \sum_{j \in D_{k \cup D_l}} (t_{ij} - \sum_{s=1}^p x_{is} x_{js})^2 \right).$$

b) Sei $\sigma_b(X)$ das LS-Fitkriterium für den Fall, daß die beiden Cluster D_k und D_l zu einem Cluster vereinigt sind. Dann ist $\sigma_b(X)$ zerlegbar in der Form

$$\begin{aligned}
\sigma_b(X) &= \sum_{i \in D_k} \left(\sum_{j \in D_k} (t_{ij} - 1)^2 + \sum_{j \in D_l} (t_{ij} - 1)^2 + \right. \\
& \left. + \sum_{j \in D_{k \cup D_l}} t_{ij}^2 \right) + \quad (14b) \\
& + \sum_{i \in D_l} \left(\sum_{j \in D_k} (t_{ij} - 1)^2 + \sum_{j \in D_l} (t_{ij} - 1)^2 + \right.
\end{aligned}$$

$$\begin{aligned}
& \left. + \sum_{j \in D_k \cup D_l} t_{ij}^2 \right) + \\
& + \sum_{i \in D_k \cup D_l} \left(\sum_{j \in D_k} t_{ij}^2 + \sum_{j \in D_l} t_{ij}^2 + \right. \\
& \left. + \sum_{j \in D_k \cup D_l} (t_{ij} - \sum_{s=1}^p x_{is} x_{js})^2 \right).
\end{aligned}$$

c) Bei Vereinigung zweier disjunkter Cluster D_k und D_l ändert sich der Wert des LS-Fitkriteriums $\sigma(X)$ um

$$\begin{aligned}
\delta_{kl} &= \sigma_a(X) - \sigma_b(X) = \\
&= \sum_{i \in D_k} \left(\sum_{j \in D_l} t_{ij}^2 - \sum_{j \in D_l} (t_{ij} - 1)^2 \right) + \\
& + \sum_{i \in D_l} \left(\sum_{j \in D_k} t_{ij}^2 - \sum_{i \in D_k} (t_{ij} - 1)^2 \right) \quad (15) \\
&= 2 \sum_{i \in D_k} \sum_{j \in D_l} (2t_{ij} - 1), \quad k, l = 1, \dots, p.
\end{aligned}$$

Das bedeutet: Je mehr Links $t_{ij} = 1$, $i \in D_k$, $j \in D_l$, zwischen den Elementen der beiden Cluster D_k und D_l bestehen, umso größer wird δ_{kl} und falls $\delta_{kl} > 0$, wird mit Vereinigung der beiden Cluster D_k und D_l der Wert des LS-Fitkriteriums $\sigma(X)$ verringert, d.h. der Fit der Clusterstruktur zu den Daten $T = (t_{ij})$ verbessert.

d) Nach Berechnung der symmetrischen Matrix $\Delta = (\delta_{kl})$ sollten sukzessive solche Cluster D_k und D_l vereinigt werden, für die die Fitverbesserung δ_{kl} maximal und positiv ist. Nach Vereinigung zweier Cluster D_k und D_l brauchen nur die δ -Werte zwischen dem neu entstandenen Cluster $D_k \cup D_l$ und den restlichen Clustern bestimmt und in die entsprechende Zeile und Spalte von Matrix Δ eingesetzt zu werden. Die Ordnung der Matrix Δ verringert sich bei jedem Vereinigungsschritt um Eins. Sobald $\max \delta_{kl} < 0$ ist, sollte der Vereinigungsprozeß abgebrochen werden, da keine weitere Fitverbesserung durch Clustervereinigung möglich ist.

3. Die Minimierung des LS-Fitkriteriums $\sigma(X)$ ist äquivalent dazu, eine Struktur $X = (x_{ik})$ von p disjunkten Clustern D_k so zu bestimmen, daß die $n \times n$ -Matrix $S = (s_{ij})$ mit

$$s_{ij} = \sum_{k=1}^p x_{ik} x_{jk} \text{ bzw. } S = XX'$$

der reproduzierten Skalarproduktwerte (Modellwerte) eine gegebene ILM $T = (t_{ij})$ bestmöglich anpaßt.

Bei den p Auswahlsschritten des hier vorgeschlagenen Algorithmus wird die Matrix $S = (s_{ij})$ ausgehend von einer Nullmatrix in jedem Auswahlsschritt k um solche Links $s_{ij} = 1$, $i \in D_k$, $j \in D_k$, angereichert, die auch Links in der gegebenen ILM $T = (t_{ij})$ sind. Die Menge der Links in S nimmt zwar bei jedem Auswahlsschritt an Umfang zu, bleibt aber stets eine Teilmenge der Links in $T = (t_{ij})$.

Bei den Vereinigungsschritten des hier vorgeschlagenen Algorithmus wird die Matrix $S = (s_{ij})$ bei jeder Vereinigung zweier Cluster D_k und D_l auch um solche Links $s_{ij} = s_{ji} = 1$, $i \in D_k$, $j \in D_l$, angereichert, die nicht

Links der gegebenen ILM $T = (t_{ij})$ sind. Wegen der Forderung $\delta_{k1} > 0$ gibt es bei jedem Vereinigungsschritt mehr Positionen (i, j) bzw. (j, i) , $i \in D_k, j \in D_1$, auf denen die hinzukommenden Links in S mit Links in der gegebenen ILM $T = (t_{ij})$ zusammentreffen.

Etwas unscharf kann man das so ausdrücken: Bei den Auswahlritten des Algorithmus werden die Links in S „von unten“ an die Links in T angepaßt und bei den Vereinigungsschritten des Algorithmus werden die Links in S „von oben“ an die Links in T angepaßt.

B.4. Methoden zum Fit der metrischen Skalarproduktmodelle

Sei $U = (u_{ij})$ eine gegebene nichtnegative, symmetrische (diagonaldominante) $n \times n$ -Proximitätsmatrix, deren Außerdiagonalelemente $u_{ij}, i \neq j$, betrachtet werden können als empirisch erhobene Wahrscheinlichkeiten (subjective probabilities) für das Ereignis, daß die beiden Objekte i und j einem von p Clustern gemeinsam angehören, und deren Diagonalelemente u_{ii}

- (a) als quadrierte Längen von Cluster-Zugehörigkeitsvektoren Y_i betrachtet werden können;
- (b) als Wahrscheinlichkeiten dafür, daß ein Objekt i einem der p Cluster angehört, betrachtet werden können;
- (c) nicht näher spezifiziert sind (missing elements).

Das Problem einer Clusteranalyse der im Abschnitt A.5. abgeleiteten Modellrelationen (11.a)–(11.c) besteht dann darin, eine nichtnegative (metrische) $n \times p$ -Matrix $Y = (y_{ik})$ von Objekt-Cluster-Zugehörigkeitsanteilen $y_{ik} \geq 0$ so zu bestimmen, daß je nach der Bedeutung der Diagonalelemente von U eines der folgenden Least-Squares-Fitkriterien minimiert wird:

$$\sigma_a(Y) = \sum_{i=1}^n \sum_{j=1}^n (u_{ij} - r_{ij}(Y))^2, \quad (16.a)$$

$$\sigma_b(Y) = \sum_{i \neq j} (u_{ij} - r_{ij}(Y))^2 + \sum_{i=1}^n (u_{ii} - \sum_{k=1}^p y_{ik})^2, \quad (16.b)$$

$$\sigma_c(Y) = \sum_{i \neq j} (u_{ij} - r_{ij}(Y))^2, \quad (16.c)$$

wobei mit

$$r_{ij}(Y) = \sum_{k=1}^p y_{ik} y_{jk} = Y_i' Y_j \quad (17)$$

die Skalarprodukt-Modellrelation bezeichnet wird.

Dies sind drei Probleme der nichtlinearen Optimierung mit besonders einfachen Schranken-Zwängen (Nichtnegativität) an die Variablen. Mittels der Variablentransformation

$$z_{ik}^2 = y_{ik}, \quad i = 1, \dots, n \quad k = 1, \dots, p, \quad (18)$$

und der veränderten Skalarproduktrelation

$$r_{ij}(Z) = \sum_{k=1}^p z_{ik}^2 z_{jk}^2 \quad (19)$$

werden die drei restriktiven Least-Squares-Fitprobleme (16) in Y zu nichtrestriktiven Least-Squares-Fitkriterien

$\sigma_a(Z), \sigma_b(Z), \sigma_c(Z)$, die für $z_{ik} \in \mathbb{R}$ zu minimieren sind. Mittels einer nichtrestriktiven Optimierungsmethode wird zunächst eine $n \times p$ -Matrix $Z = (z_{ik})$ bestimmt, die ein Least-Squares-Fitkriterium $\sigma(Z)$ lokal minimiert. Über die Variablentransformation (18) wird anschließend eine lokal optimale $n \times p$ -Matrix $Y = (y_{ik})$ erhalten, deren Elemente y_{ik} natürlich nichtnegativ sind. Während die Least-Squares-Probleme $\sigma(Y)$ in (16) die Parameter y_{ik} als Polynom vierten Grades enthalten sind, sind die Parameter z_{ik} in den Least-Squares-Kriterien $\sigma(Z)$ in Form eines Polynoms achten Grades enthalten. Der Vorteil der unkomplizierteren nichtrestriktiven Optimierung der Probleme $\sigma(Z)$ wird also mit einer wesentlichen Erhöhung der Nichtlinearität der objektiven Funktion bezahlt. Dies hat wiederum eine erhöhte Sensibilität der numerischen iterativen Optimierungsverfahren bezüglich lokaler aber nicht globaler Optima zur Folge.

Als notwendige Bedingung dafür, daß eine Parametermenge X^* ein Least-Squares-Fitkriterium minimiert, wird verlangt, daß der an der Stelle X^* bewertete Gradientenvektor $g(\sigma(X))$ der partiellen Ableitungen von $\sigma(X)$ verschwindet. Für das Kriterium $\sigma_a(Y)$ in (16.a) erhält man so

$$\partial \sigma / \partial y_{ik} = -4 \sum_{j=1}^n (u_{ij} - r_{ij}(Y)) y_{jk} = 0 \quad (20)$$

und entsprechend für das Kriterium $\sigma_a(Z)$

$$\partial \sigma / \partial z_{ik} = -8 z_{ik} \sum_{j=1}^n (u_{ij} - r_{ij}(Z)) z_{jk}^2 = 0 \quad (21)$$

jeweils ein System von np Normalgleichungen, wobei die Normalgleichungen (20) algebraische Gleichungen 3. Grades in der Variablen y_{ik} und die Normalgleichungen (21) algebraische Gleichungen 7. Grades in den Variablen z_{ik} sind. (Man erkennt leicht, daß $z_{ik} = 0$ für alle i, k , eine Lösung der Normalgleichungen (21) ist, die $\sigma_a(Z)$ aber nicht minimiert.)

Die bekanntesten Verfahren zur nichtrestriktiven Minimierung von $\sigma(Z)$ sind Gradienten-Methoden, Newton- oder Quasi-Newton-Methoden, konjugierte-Gradienten-Methoden und Levenberg-Marquardt-Methoden. Die Levenberg-Marquardt-Methode ist speziell geeignet zur Minimierung nichtrestriktiver nichtlinearer Least-Squares-Fitprobleme

$$\sigma(X) = \sum_{i=1}^M f_i^2(X) = \|F(X)\|^2, \quad X \in \mathbb{R}^N. \quad (22)$$

In der Umgebung eines Punktes $X_0 \in \mathbb{R}^N$ können die M nichtlinearen Funktionen $f_i(X)$ durch eine Taylor-Reihe approximiert werden

$$f_i(X_0 + P) \approx f_i(X_0) + \sum_{j=1}^N \frac{\partial f_i}{\partial x_j} p_j \quad (23)$$

bzw.

$$F(X_0 + P) \approx F(X_0) + J_0 P$$

wobei $J_0 = \left(\frac{\partial f_i}{\partial x_j}(X_0) \right)$ die $M \times N$ -Jacobi-Matrix der

partiellen Ableitungen der M Funktionen $f_i(X)$ bez. der N Variablen x_j bewertet am Punkt X_0 bezeichnet. Die lineare Taylor-Reihen-Approximation gilt natürlich nur in einer hinreichend kleinen Umgebung des Punktes X_0 , die umso kleiner ist je nichtlinearer die Funktionen $f_i(X)$ sind. Ein hinreichend kleiner Korrekturvektor $P \in \mathbb{R}^N$ der Approximation X_0 des lokalen Minimums X^* von $\sigma(X)$ sollte also so bestimmt werden, daß für eine gegebene diagonale Skalierungsmatrix $D_0 = \text{DIAG}(d_1, \dots, d_N)$ und eine gegebene positive Schranke $\Delta_0 > 0$

$$\varphi(P) = \sum_{i=1}^M (f_i(X_0) + \sum_{j=1}^N \frac{\partial f_i}{\partial x_j} p_j)^2 + \lambda \left(\sum_{j=1}^N (d_j p_j)^2 - \Delta_0^2 \right)$$

bzw. (24)

$$\varphi(P) = \|F(X_0) + J_0 P\|^2 + \lambda (\|D_0 P\|^2 - \Delta_0^2)$$

minimiert wird. Dabei bezeichnet $\lambda \geq 0$ einen Lagrange-Parameter, der die Länge des skalierten Korrekturvektors $D_0 P$ beschränkt mit

$$\|D_0 P\|^2 \leq \Delta_0^2. \quad (25)$$

Die Forderung (25) bedeutet für eine nichtsinguläre diagonale Skalierungsmatrix D_0 , daß P in einem achsenparallelen Hyperellipsoid um den Punkt X_0 liegt und die Länge der j-ten Halbachse beträgt Δ_0/d_j , $j = 1, \dots, N$. Nach Nullsetzen der partiellen Ableitungen von (24) bez. P. erhält man ein System von N Normalgleichungen bez. P

$$\partial \varphi / \partial p_1 = 2 \sum_{i=1}^M (f_i(X_0) + \sum_{j=1}^N \frac{\partial f_i}{\partial x_j} p_j) \frac{\partial f_i}{\partial x_1} + 2 \lambda d_1^2 p_1 = 0$$

bzw.

$$(J_0' J_0 + \lambda D_0' D_0) P = -J_0' F_0,$$

$$J_0 = \left(\frac{\partial f_i}{\partial x_j} (X_0) \right), \quad M \times N\text{-Jacobi-Matrix,}$$

die bei einer nichtsingulären Skalierungsmatrix D_0 mit $\hat{J} = J D_0^{-1}$ in der Form

$$(\hat{J}'_0 \hat{J}_0 + \lambda I) D_0 P = -\hat{J}'_0 F_0 \quad (26)$$

geschrieben werden können. Man zeigt leicht (vgl. MORE, 1978), daß $\|DP(\lambda)\|^2$ für $\lambda \geq 0$ monoton fällt und es gilt

$$\|DP(\lambda)\|^2 \xrightarrow{\lambda \rightarrow \infty} 0. \quad (27)$$

Um möglichst große Korrekturschritte $DP(\lambda)$ zu erhalten, muß $\lambda \geq 0$ möglichst klein gewählt werden. Aus dem linearen System (26) bestimmt man daher λ und P so, daß einerseits $\lambda \geq 0$ so klein wie möglich ist, aber andererseits $DP(\lambda)$ die Bedingung (25) erfüllt.

Vom Autor wurde ein FORTRAN-Programm zur Implementation der Levenberg-Marquardt-Methode geschrieben, das im wesentlichen auf MORE (1978) basiert, aber auch rangdefiziente Jacobi-Matrizen J erlaubt, Rang $(J_0) \leq N$. Dieses Programm wurde an einem so-

genannten synthetischen (konstruiertem) Beispiel zur Minimierung des stark nichtlinearen nichtrestriktiven Least-Squares-Problems $\sigma_c(Z)$ erprobt. Bei einem synthetischen Beispiel ist die global optimale Parametermenge Z^* bekannt. In Tab. 2 ist eine Clusterstruktur $Y^* = (y_{ik}^*)$ von $n = 15$ Objekten in $p = 4$ Clustern gegeben, aus der über die Transformation (18) leicht die Clusterstruktur $Z^* = (z_{ik}^*)$ durch Quadrieren ermittelt wird. Neben einer mehr oder weniger guten Anfangsschätzung $Z^{(0)}$ wird dem FORTRAN-Programm die in Tab. 1 präsente Skalarproduktmatrix als Datenmatrix eingegeben. Das Programm ermittelt dann iterativ eine Folge $\{Z^{(k)}\}$ von sukzessive verbesserten Schätzungen von Z^* , für die $\sigma_c(Z^{(k+1)}) \leq \sigma_c(Z^{(k)})$ gilt.

Tab. 1: Skalarproduktmatrix $U = (u_{ij})$
(Dezimalpunkte weggelassen)

U	
1	
2	0
3	0 0
4	0 0 0
5	90 10 0 0
6	10 80 10 0 17
7	0 10 80 10 1 16
8	0 0 10 90 0 1 17
9	70 20 10 0 65 24 10 1
10	20 60 20 0 24 52 22 2 28
11	0 20 60 20 2 22 52 24 10 24
12	0 10 20 70 1 10 24 65 4 10 28
13	50 30 20 0 48 31 19 2 43 32 18 7
14	10 40 40 10 13 37 37 13 19 34 34 19 25
15	0 20 30 50 2 19 31 48 7 18 32 43 12 25

Tab. 2: Clusterstruktur $Y^* = (y_{ik}^*)$
(Dezimalpunkte weggelassen)

Y*				
1	100	0	0	0
2	0	100	0	0
3	0	0	100	0
4	0	0	0	100
5	90	10	0	0
6	10	80	10	0
7	0	10	80	10
8	0	0	10	90
9	70	20	10	0
10	20	60	20	0
11	0	20	60	20
12	0	10	20	70
13	50	30	20	0
14	10	40	40	10
15	0	20	30	50

Obwohl für die unverzerrte Datenmatrix $U = (u_{ij})$ ein perfekter Fit $\sigma_c(Z^*) = 0$ möglich ist, beendet das Programm den Iterationsprozeß, wenn mit der zuletzt erhaltenen Iterierten keine signifikante Fitverbesserung erreicht werden konnte. Trotz des stark nichtlinearen Problemes zeigte sich das Programm erstaunlich fähig, in relativ wenigen Iterationsschritten die globaloptimale Lösung Z^* mit hoher Genauigkeit zu bestimmen. Je nach der Wahl der Anfangsschätzung $Z^{(0)}$ waren die Spalten der Lösungsmatrix \hat{Z} verschieden permutiert. Beispielsweise wurde mit einer Anfangsschätzung $z_{ik}^{(0)} = 0,25$ für alle $i = 1, \dots, 15$, $k = 1, \dots, 4$, bereits

nach 17 Iterationen eine Lösung $Y^{(17)}$ ermittelt, die in den ersten vier Dezimalstellen mit der exakten Clusterstruktur Y^* , vgl. Tab. 2, übereinstimmt. Das Programm reagierte jedoch empfindlich bei falsch platzierten verschwindenden Anfangsschätzungen $z_{ik}^{(0)} = 0,0$ und führte dann zu einem entsprechenden lokalen (aber nicht globalen) Minimum, bei dem die Anfangsschätzungen $z_{ik}^{(0)} = 0,0$ in den Resultaten wiederkehrten. Die Lösung war auch erstaunlich stabil bezüglich einer mit normalem Zufallsfehler gestörten Datenmatrix $\bar{U} = (\bar{u}_{ij})$

$$\bar{u}_{ij} = u_{ij} + e_{ij}, e_{ij} \in N(0, \sigma). \quad (28)$$

Auf die Rechenresultate kann hier aus Platzmangel leider nicht näher eingegangen werden.

Es kann angenommen werden, daß die iterativen numerischen Methoden bei der Minimierung der stark nichtlinearen nichtrestriktiven Probleme $\sigma(Z)$ empfindlicher sind bezüglich lokaler aber nicht globaler Minima als die Methoden zur Minimierung der wesentlich weniger nichtlinearen Probleme $\sigma(Y)$ mit Schranken-Zwängen $Y \geq 0$. Aus diesem Grunde ist vom Autor die Implementierung einer Kombination der Levenberg-Marquardt-Methode mit der active-set-Methode (vgl. FLETCHER, 1982, Band II) geplant zur restriktiven Minimierung von Least-Squares-Problemen mit linearen Zwängen der Form

$$\begin{aligned} a_i'X &= 0, i \in E, \\ a_i'X &\geq 0, i \in I, \end{aligned} \quad (29)$$

wobei die Vektoren $a_i \in \mathbb{R}^N$ als unabhängig vorausgesetzt werden. Die aktiven Zwänge bei einem Punkt X_0 werden charakterisiert durch die Indexmenge

$$\mathcal{A}_0 = \mathcal{A}(X_0) = \{i: a_i'X_0 = 0\} \quad (30)$$

Für den Fall, daß nur Gleichheits-Zwänge vorkommen, $I = \emptyset$, ist die Aktivmenge \mathcal{A} konstant und in jedem Iterationsschritt wird ein mittels verallgemeinerter Variablenelimination reduziertes (quadratisches oder lineares) Minimierungsproblem gelöst. Wenn Ungleichheits-Zwänge vorkommen, dann ändert sich die Aktivmenge \mathcal{A} im allgemeinen in jedem Iterationsschritt, indem entweder ein Ungleichheits-Zwang p zu \mathcal{A} addiert oder ein Ungleichheits-Zwang q aus \mathcal{A} entfernt wird.

(Dabei sollte Ziggzagging vorgebeugt werden, vgl. FLETCHER, 1981, Band II.) In Verbindung mit der Levenberg-Marquardt-Methode und bei Nichtnegativitäts-Zwängen $X \geq 0$ besteht die Variablenelimination darin, daß in jedem Iterationsschritt die bei der aktuellen Aktivmenge \mathcal{A} angezeigten Variablen p_j in den linearen Problemen (24) bzw. (26) Null gesetzt werden. Darum scheint eine Erweiterung des vorliegenden Levenberg-Marquardt-Programmes mit der active-set-Methode (vgl. (11.2.2) und (11.3.4) bei FLETCHER, 1981, Band II) zur Behandlung der Nichtnegativitätszwänge $X \geq 0$ relativ einfach zu sein. Leider können hier noch keine Resultate präsentiert werden.

References:

- (1) Arabie, P.; Carroll, J.D. (1980): MAPCLUS. A mathematical approach to fitting the ADCLUS model; *Psychometrika* 45, 211–235.
- (2) Baumann, U. (1971): *Psychologische Taxometrie*; H. Huber, Bern.
- (3) Cattell, R.B.; Coulter, M.A. (1966): Principles of behavioral taxonomy and the mathematical basis of the taxonome computer program; *British Journal Mathemat. & Statist. Psychology* 19, 237–269.
- (4) Cattell, R.B.; Coulter, M.A.; Tsujioka, B. (1966): The taxometric recognition of types and functional emergents; in: R.B. Cattell (Ed.): *Handbook of Multivariate Experimental Psychology*, Chicago.
- (5) Fletcher, R. (1981): *Practical Methods of Optimization*, Vol. I: Unconstrained Optimization, Vol. II: Constrained Optimization; J. Wiley & Sons, New York.
- (6) Hartmann, W. (1976a): Über ein Verfahren der numerischen Taxometrie von Cattell und Coulter; *Biometrische Zeitschrift* 18, 273–290.
- (7) Hartmann, W. (1976b): Über einen Algorithmus zur Clusteranalyse maximal kompakter Gruppen und die rechentechnische Realisierung des Verfahrens von Cattell und Coulter; *Biometrische Zeitschrift*, 18, 333–349.
- (8) Johnson, S.C. (1967): Hierarchical clustering schemes; *Psychometrika* 32, 241–254.
- (9) More, J.J. (1978): The Levenberg-Marquardt algorithm: Implementation and theory; in: G.B. Watson (Ed.): *Numerical Analysis, Lecture Notes in Mathematics* 630, Springer Verlag, Berlin.
- (10) Quaas, P. (1974): Die Clusteranalyse und ihre Anwendung in der Psychologie; Technical Report, TU Dresden.
- (11) Shepardt, R.N.; Arabie, P. (1979): Additive Clustering: Representation of similarities as combinations of discrete overlapping properties; *Psychological Review* 86, 87–123.

Dr. Wolfgang Hartmann
Wachsbleichstr. 27
Dresden/German Democratic Republic