Komplexitätsreduktion im Data- und e-Mining: Dependente Q-Techniken

Joachim Hartung, Guido Knapp

Fachbereich Statistik*, Universität Dortmund,
44221 Dortmund
E-Mail: hartung@statistik.uni-dortmund.de

Abstract

Im Rahmen der Wissenserkennung in großen Datenbeständen (KDD) werden Dimensionsreduzierungen sowohl hinsichtlich der Items als auch bezüglich der Transaktionen diskutiert. Dabei werden passende Partitionierungen der Datenkonfiguration erstellt und analysiert. Die gewonnenen Teilergebnisse werden dann unter Berücksichtigung von Varianzen und insbesondere der Korrelationen der Items optimal zu einem Gesamtergebnis kombiniert. Als unmittelbare Anwendung der vorgestellten Strategien wird die Personalisierung von Nutzern im Internet angesprochen.

Summary:
In the framework of Knowledge Discovery in (large) Databases (KDD) dimension reductions are considered with respect to the number of items as well as the number of transactions. Hereby suitable partitions of the database are performed and analysed. The obtained partial results then are combined to a global result in an optimal manner taking into consideration variances and in particular the correlations of the items. As a direct application of the presented strategies personalization of users in the world wide web is discussed.

Key words:
Data Mining, Web Mining, Dimension Reduction, Combining Information, Meta-Analysis, KDD, eCRM, Personalization, Pattern Recognition, Artificial Neural Networks.

* und SFB 475 „Komplexitätsreduktion in multivariaten Datenstrukturen“ der DFG an der Universität Dortmund sowie „Kompetenzzentrum Unternehmensberatung und -entwicklung“, (KUBE e.V.).

1. Einführung

Data-Mining und speziell e- oder Web-Mining wird in Datenkonfigurationen betrieben, deren Komplexität sich zur Zeit in einem rasanten Wachstum befindet. Dies stellt die bewährten Analysemethoden zunehmend vor beträchtliche Dimensionsprobleme, sowohl bezüglich der Anzahl an Items, Attributen, Variablen oder Merkmalen, als auch hinsichtlich der Anzahl an Transaktionen oder Objekten, die der Datenbasis zugrunde liegen.

Daher verfolgen wir ein Zerlegungs-Vereinigungs-Prinzip, welches zunächst die Datenbasis derart in Stücke zerlegt, daß diese den passenden Einsatz des jeweils zugedachten Analyseverfahrens erlauben, und dann die so erhaltenen dezentralen Ergebnisse adäquat zu einem Gesamtergebnis für das Ausgangsproblem vereinigt. Dabei beschränken wir uns hier auf dependente Q-Techniken der Multivariaten Statistik (Hartung, Elpelt, 1999), zu denen insbesondere auch Verfahren zur Mustererkennung bzw. Pattern-Recognition etwa mittels Künstlicher Neuronaler Netze unter überwachtem Lernen (Berthold, Hand, 1999) gehören.

An $i = 1,\dots,n$ Objekten mögen wir die Ausprägungen $x_i^T = (\xi_{i1},\dots,\xi_{ip})$ von p Merkmalen sowie den Wert η_i einer Kriteriumsvariablen y erhoben haben, so dass wir die Daten darstellen können in Form einer Datenmatrix D folgender Anordnung:

$$D = (y \mid X) = \left(y \,\middle|\, \begin{matrix} x_1^T \\ \vdots \\ x_n^T \end{matrix} \right) = \left(\begin{array}{c|ccc} \eta_1 & \xi_{11} & \cdots & \xi_{1p} \\ \vdots & \vdots & & \vdots \\ \eta_n & \xi_{n1} & \cdots & \xi_{np} \end{array} \right)_{n\times(p+1)}$$

2. Dimensionsreduktion auf der Merkmalsebene

Die p Merkmale werden kanonisch in P Merkmalsgruppen (P < p) der Umfänge p_j zerlegt, $\sum_{j=1}^{P} p_j = p$. Des weiteren seien die Indexmengen

$$P_j = \left\{ \sum_{k=0}^{j-1} p_k + 1, \dots, \sum_{k=0}^{j-1} p_k + p_j \right\}, \; j = 1,\dots,P \,;\; p_0 = 0 ,$$

eingeführt, so dass mit

$$X_{*j} = \begin{pmatrix} x_{1j}^T \\ \vdots \\ x_{nj}^T \end{pmatrix} = \begin{pmatrix} \{\xi_{1k}\}_{k\in P_j} \\ \vdots \\ \{\xi_{nk}\}_{k\in P_j} \end{pmatrix}, \; j = 1,\dots,P ,$$

die Datenmatrix sich schreiben lässt als

$$D = (y \mid X) = (y \mid X_{*1},\dots,X_{*p}) .$$

Das j-te partitionierte Problem sei nun charakterisiert durch die Datenmatrix

$$D_{*j} = (y \mid X_{*j}) , \; j = 1,\dots,P .$$

Für einen passend dimensionierten Parametervektor β_j wird sodann gemäß eines Ausgangskriteriums zur Prognose der Kriteriumsvariablen y eine Ziel- oder Entscheidungsfunktion

$$g_j(\beta_j; x_{ij}) = g_j\left(\beta_j; \{\xi_{ik}\}_{k\in P_j}\right),\ i = 1,\dots,n\ ,$$

bezüglich β_j an Hand der Beobachtungen $\eta_1,\dots,\eta_n$ von y optimiert oder trainiert. Bezeichnet $\hat{\beta}_j$ die Lösung, so lassen sich mit dem Prognoseresiduum $e_j^T = (\varepsilon_{1j},\dots,\varepsilon_{nj})$ die Komponenten von y schreiben als

$$\eta_i = g_j(\hat{\beta}_j; x_{ij}) + \varepsilon_{ij},\ i = 1,\dots,n\ ;\ j = 1,\dots,P\ .$$

Beobachten wir nun für ein „neues" Objekt o die p Merkmalswerte $(\xi_{01},\dots,\xi_{0p}) = (x_{01}^T,\dots,x_{0P}^T)$, so erhalten wir für dessen Wert η_0 in der y-Variablen die P Prognosen

$$\begin{aligned}\hat{\eta}_{0j} &= g_j(\hat{\beta}_j; x_{0j}) + \hat{\varepsilon}_{0j}\\ &= \hat{g}_j + \hat{\varepsilon}_{0j} \qquad ,\quad j = 1,\dots,P,\end{aligned}$$

wobei üblicherweise $\hat{\varepsilon}_{0j}$ gesetzt wird als

$$\hat{\varepsilon}_{0j} = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{ij}\ ,\ j = 1,\dots,P\ ,$$

welches sich im Falle unverzerrter Prognosen zu 0 ergibt (zumindest angenähert).

Gesucht sind dann Gewichte w_j für eine „optimale" Kombination dieser Prognosen zu einer Gesamtprognose von η_0 der Form

$$\hat{\eta}_0 = \sum_{j=1}^{P} w_j \cdot \hat{\eta}_{0j}\ , \text{ mit } \sum_{j=1}^{P} w_j = 1\ .$$

Wählen wir als Kriterium das der minimalen Varianz, so bezeichne V die Kovarianzmatrix der Einzelprognosen,

$$V = \{\mathrm{Cov}(\hat{\eta}_{0j}, \hat{\eta}_{0k})\}_{\substack{j=1,\dots,P\\ k=1,\dots,P}}\ ,$$

die sich i.d.R. reduzieren lässt auf

$$V = \left(\{\mathrm{Cov}(\hat{g}_j, \hat{g}_k)\}_{\substack{j=1,\dots,P\\ k=1,\dots,P}} + \mathrm{Diag}\{\mathrm{Var}(\hat{\varepsilon}_{0j})\}_{j=1,\dots,P}\right).$$

Nehmen wir an, es existiert die Inverse

$$V^{-1} = \{v_{jk}^*\}_{\substack{j=1,\dots,P\\ k=1,\dots,P}}\ ,$$

so lassen sich die optimalen Gewichte explizit angeben als

$$w_j = \frac{1}{\sum_{\upsilon=1}^{P}\sum_{k=1}^{P} v_{\upsilon k}^*} \cdot \sum_{k=1}^{P} v_{jk}^*\ ,\ j = 1,\dots,P\ .$$

Die optimale Prognose $\hat{\eta}_0$ hat dann die minimale Varianz

$$\mathrm{Var}(\hat{\eta}_0) = \left(\sum_{\upsilon=1}^{P} \sum_{k=1}^{P} v^*_{\upsilon k} \right)^{-1} .$$

Bei diesem Kriterium werden sowohl die Varianzen der Prognosefunktionen selbst als auch deren Treffsicherheit in Form der Residualvarianzen zur Gewichtung herangezogen. Legen wir beispielsweise als Kombinationskriterium das der besten quadratischen Approximation zu Grunde, so erhalten wir mit dem P-dimensionalen Vektor

$$\alpha^T = (\alpha_1, \ldots, \alpha_p) \,,\; \sum_{j=1}^{P} \alpha_j = 1 \,,\; \alpha_j \ge 0 \,,\; j = 1, \ldots, P \,,$$

wegen

$$A(\alpha) := \left[\sum_{i=1}^{n} \left(\eta_i - \sum_{j=1}^{P} \alpha_j \cdot g_j(\hat{\beta}_j ; x_{ij}) \right)^2 \right]^{1/2}$$
$$\le \sum_{j=1}^{P} \alpha_j \left[\sum_{i=1}^{n} \left(\eta_i - g_j(\hat{\beta}_j ; x_{ij}) \right)^2 \right]^{1/2}$$

auf Grund eines Konvexitätsargumentes (Hartung, 1976), folglich das Minimierungsproblem

$$\min \left\{ A(\alpha) \middle| \, \alpha \ge 0, \sum_{j=1}^{P} \alpha_i = 1 \right\} .$$

Mit den Bezeichnungen

$$G = \left(\{ g_1(\hat{\beta}_1 ; x_{i1}) \}_{i=1,\ldots,n}, \ldots, \{ g_P(\hat{\beta}_P ; x_{iP}) \}_{i=1,\ldots,n} \right)_{n \times P} ,$$

$\iota = (1, \ldots, 1)^T$ der P-dimensionale Einservektor, I die P-dimensionale Einheitsmatrix und $J = I - (1/P) \cdot \iota\iota^T$, ist $\hat{\alpha}$ die gesuchte Lösung, wenn für einen P-dimensionalen Vektor λ und $\hat{\alpha}$ die folgenden Bedingungen erfüllt sind:

$$J(G^T G \hat{\alpha} - G^T y) = J\lambda$$
$$\iota^T \hat{\alpha} = 1$$
$$\lambda^T \hat{\alpha} = 0$$
$$\hat{\alpha} \ge 0 \,,\; \lambda \ge 0 \,.$$

Entsprechend ergibt sich hiermit für unser neues Objekt o so die kombinierte Prognose zu

$$\hat{\eta}_0(\hat{\alpha}) = \sum_{j=1}^{P} \hat{\alpha}_j \cdot \hat{\eta}_{0j} \,.$$

3. Dimensionsreduktion auf der Objektebene

Jetzt werden die n Objekte kanonisch in N Objektgruppen (N < n) der Umfänge n_i aufgeteilt, $\sum_{i=1}^{N} n_i = n$. Entsprechend definieren wir die Indexmengen

$$N_i = \left\{ \sum_{\nu=0}^{i-1} n_\nu + 1, \ldots, \sum_{\nu=0}^{i-1} n_\nu + n_i \right\},\ i = 1, \ldots, N;\ n_0 = 0,$$

sowie die zugehörigen Zerlegungen

$$y_i = \{\eta_k\}_{k \in N_i},\ X_{i*} = (x_k^T)_{k \in N_i},\ i = 1, \ldots, N,$$

so dass sich die anfängliche Datenmatrix darstellen lässt als

$$D = (y \mid X) = \left(\begin{array}{c|c} y_1 & X_{1*} \\ \vdots & \vdots \\ y_N & X_{N*} \end{array} \right),$$

und das i-te partitionierte Problem hier charakterisiert wird durch die Datenmatrix

$$D_{i*} = (y_i \mid X_{i*}),\ i = 1, \ldots, N.$$

Mit einem passend dimensionierten Parametervektor γ_i wird wieder nach dem Ausgangskriterium für die Prognose der y-Variablen eine Zielfunktion

$$h_i(\gamma_i; x_k) = h_i(\gamma_i; \xi_{k1}, \ldots, \xi_{kp}),\ k \in N_i,$$

bezüglich γ_i optimiert bzw. trainiert, allerdings hier jeweils nur an Hand der Beobachtungen in y_i, für $i = 1, \ldots, N$. Mit der Lösung $\hat{\gamma}_i$ und dem Residuumvektor $d_i = \{\delta_k\}_{k \in N_i}$ erhält man für die Komponenten von y_i

$$\eta_k = h_i(\hat{\gamma}_i, x_k) + \delta_k,\ k \in N_i,\ i = 1, \ldots, N.$$

3.1 Prognose

Für die Merkmalswerte $x_0^T = (\xi_{01}, \ldots, \xi_{0p})$ eines „neuen" Objektes o ergeben sich somit bezüglich der y-Variablen die N Prognosen

$$\begin{aligned} \hat{\eta}_{i0} &= h_i(\hat{\gamma}_i; x_0) + \hat{\delta}_{i0} \\ &= \hat{h}_i + \hat{\delta}_{i0} \qquad ,\quad i = 1, \ldots, N, \end{aligned}$$

wobei wiederum $\hat{\delta}_{i0}$ gesetzt wird als

$$\hat{\delta}_{i0} = \frac{1}{n_i} \cdot \sum_{k \in N_i} \delta_k .$$

Mit der Kovarianzmatrix

$$U = \left(\{\mathrm{Cov}(\hat{h}_i, \hat{h}_\upsilon)\}_{\substack{i=1,\ldots,N \\ \upsilon=1,\ldots,N}} + \mathrm{Diag}\{\mathrm{Var}(\hat{\delta}_{i0})\}_{i=1,\ldots,N} \right),$$

bzw. ihrer Inversen

$$U^{-1} = \{u^*_{i\upsilon}\}_{\substack{i=1,\ldots,N \\ \upsilon=1,\ldots,N}} ,$$

und den daraus gewonnenen Gewichten

$$w_i^* = \frac{1}{\sum_{k=1}^{N}\sum_{\upsilon=1}^{N} u^*_{k\upsilon}} \cdot \sum_{\upsilon=1}^{N} u^*_{i\upsilon} ,\ i = 1,\ldots, N$$

ergibt sich analog zu Abschnitt 2 die optimale Kombination minimaler Varianz zur Gesamtprognose

$$\hat{\eta}_0^* = \sum_{i=1}^{N} w_i^* \cdot \hat{\eta}_{i0} ,$$

wenn wir davon ausgehen, dass bei der Gruppierung der Objekte kein Auswahlfehler zu berücksichtigen ist. Die zugehörige Varianz ist dann gegeben durch

$$\mathrm{Var}(\hat{\eta}_0^*) = \left(\sum_{k=1}^{N}\sum_{\upsilon=1}^{N} u^*_{k\upsilon} \right)^{-1} .$$

3.2 Parameterschätzung

Mitunter steht nicht die Prognose der Kriteriumsvariablen für zukünftige Objekte im Vordergrund der Analyse, sondern etwa die Schätzung eines ausgezeichneten Parameters μ in allen Objektgruppen, d.h. wir nehmen an, dass mit passend dimensionierten gegebenen Vektoren c_i unverzerrte Schätzungen $c_i^T\hat{\gamma}_i$ für μ vorliegen,

$$\hat{\mu}_i = c_i^T\hat{\gamma}_i \text{, mit } E\hat{\mu}_i = \mu ,\ i = 1,\ldots, N .$$

Berücksichtigt man einen möglichen Auswahlfehler bei der Objektgruppierung, so kommen wir zum hier dann passenden Random Effects Modell der Meta-Analyse (Draper et al., 1992; Hartung, 1999a), in dem sich die Varianz von $\hat{\mu}_i$ ergibt aus der Zwischengruppenvarianz σ_z^2 und der Varianz σ_i^2 innerhalb der i-ten Objektgruppe,

$$\mathrm{Var}(\hat{\mu}_i) = \sigma_z^2 + \sigma_i^2 ,\ i = 1,\ldots, N ,$$

wobei eine Schätzung $\hat{\sigma}_i^2$ mittels der i-ten Datenmatrix D_{i*} zu gewinnen ist. Übliche Schätzer für σ_z^2 haben den gravierenden Nachteil, dass sie mit positiver und mitunter recht großer Wahrscheinlichkeit negative Werte annehmen können (Draper et al., 1992). Bezeichne

$$\tau_i = 1/\hat{\sigma}_i^2 ,\ \tau_\Sigma = \sum_{i=1}^{N} \tau_i ,\ \tau = \tau_\Sigma \left(\tau_\Sigma^2 - \sum_{i=1}^{N} \tau_i^2 \right)^{-1}$$

und

$$Z = \tau \cdot \sum_{i=1}^{N} \tau_i \left(\hat{\mu}_i - \sum_{\upsilon=1}^{N} \frac{\tau_i}{\tau_\Sigma} \cdot \hat{\mu}_\upsilon \right)^2 ,$$

so schlagen wir zur Schätzung von σ_z^2 vor:

$$\hat{\sigma}_z^2 = \frac{Z}{Z + 2 \cdot (N-1) \cdot \tau} \cdot Z .$$

Dieser Schätzer ist ein stets (mit Wahrscheinlichkeit 1) positiver minimal verzerrter Schätzer (Hartung, 1981) mit einer anschließenden Bias-Adjustierung, vgl. auch Hartung (1999b). Mit

$$\rho_i = \frac{1}{(\hat{\sigma}_z^2 + \hat{\sigma}_i^2)} , \; \rho_\Sigma = \sum_{i=1}^{N} \rho_i ,$$

ist nach dem Kriterium der minimalen Varianz die optimale kombinierte Schätzung für μ dann gegeben als

$$\hat{\mu} = \sum_{i=1}^{N} \frac{\rho_i}{\rho_\Sigma} \cdot \hat{\mu}_i .$$

Die Varianz von $\hat{\mu}$ empfiehlt sich jedoch direkt zu schätzen und nicht über die Schätzungen der einzelnen Varianzparameter. Eine solche optimale (positiver Minque; Hartung, 1981) Direktschätzung hat die Gestalt (Hartung, 1999a)

$$\hat{\mathrm{Var}}(\hat{\mu}) = \frac{1}{N-1} \cdot \sum_{i=1}^{N} \frac{\rho_i}{\rho_\Sigma} \cdot \left(\hat{\mu}_i - \sum_{\upsilon=1}^{N} \frac{\rho_\upsilon}{\rho_\Sigma} \cdot \hat{\mu}_\upsilon \right)^2 .$$

Darüber hinaus können wir bei Vorliegen von (approximativer) Normalverteilung der $\hat{\mu}_i$ ein Konfidenzintervall zum Niveau $1-\kappa$ angeben als

$$\hat{\mu} \pm \sqrt{\hat{\mathrm{Var}}(\hat{\mu})} \cdot t_{N-1;1-\kappa/2} ,$$

wobei $t_{N-1;1-\kappa/2}$ das $(1-\kappa/2)$-Quantil der (zentralen) t-Verteilung mit $(N-1)$ Freiheitsgraden bezeichnet, vgl. auch Hartung, Knapp (2001a).

4. Simultane Dimensionsreduktion auf der Merkmalsebene und auf der Objektebene

Liegt nun eine Datenmatrix D derart vor, dass sowohl bezüglich der Anzahl an Merkmalen als auch bezüglich der Anzahl an Objekten Dimensionsreduzierungen angebracht erscheinen, so gehen wir sukzessive vor. Mit den Notationen aus den voranstehenden Abschnitten sei definiert

$$X_{ij} = \left(\{\xi_{k\upsilon}\}_{\substack{k \in N_i \\ \upsilon \in P_j}} \right)_{n_i \times p_i} ,$$

so dass wir die Datenmatrix D schematisch darstellen können als

$$D = (y \mid X) = \left(\begin{array}{c|ccc} \vdots & & \vdots & \\ y_i & \cdots & X_{ij} & \cdots \\ \vdots & & \vdots & \end{array} \right) .$$

Das ij-te partitionierte Problem wird dann charakterisiert durch die Datenmatrix

$$D_{ij} = (y_i \mid X_{ij}),\ i = 1,...,N;\ j = 1,...,P .$$

Etwa für festgehaltenes i werden sodann die D_{ij} für $j = 1,...,P$ analysiert und gemäß Abschnitt 2 die Ergebnisse $\hat{\eta}_{i0j}$ für das neue Objekt o mittels der zugehörigen Gewichte w_{ij} kombiniert zu

$$\hat{\eta}_{i0*} = \sum_{j=1}^{P} w_{ij} \cdot \hat{\eta}_{i0j} .$$

Dies geschieht für alle $i = 1,...,N$, so dass sich mit $\hat{\eta}_{i0*}$ anstelle von $\hat{\eta}_{i0}$ in Abschnitt 3.1 die finale Kombination ergibt zu

$$\hat{\eta}^*_{0*} = \sum_{i=1}^{N} w^*_{i*} \cdot \hat{\eta}_{i0*} .$$

5. Online Personalisierung im Web-Mining

Als eine direkt schon von der Problemstellung her unmittelbare Anwendung obiger Überlegungen sei die online Personalisierung von Nutzern im Internet hier kurz aufgeführt. Personalisierung stellt auch ein strategisches Instrument im sogenannten Customer Relationship Management, kurz CRM oder speziell eCRM, dar.

HTML-Adressen selbst, welche der Nutzer aufruft, und/oder gegebenenfalls die Response-Funktionen, die vom Nutzer bedient werden, stellen die Items oder Merkmale dar, auf Grund derer ein Nutzer einem bestimmten Profil oder Cluster zugeordnet werden soll, bzw. seine Interessen und somit zukünftiges Internetverhalten prognostiziert werden sollen, um seitens des Internetbetreibers passende Empfehlungen geben zu können. Wir haben es in dieser Form also mit einem erweiterten Problem der Diskriminanzanalyse (Hartung, Elpelt, 1999) zu tun.

Nun liegt es in der Natur des Internets, dass man die Profilerkennung so früh wie möglich beginnen möchte, d.h. schon nach dem ersten seitens des Nutzers angesprochenen Item, um online in der laufenden Session reagieren zu können. Mit Bezug auf Abschnitt 2 haben wir hier also den Fall vorliegen, dass beginnend mit einem Merkmal sukzessive Werte von weiteren Merkmalen vorliegen, und nach jedem weiteren Merkmal sind die bisherigen Entscheidungen zu kombinieren mit derjenigen, die sich auf Grund des zuletzt angesprochenen Merkmals ergibt. Dieses Vorgehen wird auch automatische Personalisierung genannt, wobei es sich empfiehlt, ein ständiges Update der Entscheidungsregeln in einem parallelen Batch-Prozess vornehmen zu lassen (Hartung, Hartung, 2001).

6. Abschließende Bemerkungen

Zur Realisierung der vorgeschlagenen Partitionierungen mit anschließenden Kombinationen bietet sich etwa der gemeinsame Einsatz verschiedener Optionen im SAS Enterprise Miner (2000) an, von denen hier der Ensemble Knoten hervorgehoben sei, der bei entsprechender Steuerung schon einen großen Teil der anfallenden Analyse- und Organisationsarbeiten übernehmen kann (Hartung, Knapp, 2001b).

Probleme bereitet mitunter die Gewinnung der zur optimalen Kombination benötigten Kovarianzmatrizen, insbesondere wenn die Daten bereits in partitionierter und zudem anonymisierter Form vorliegen. Hier gilt es dann, die Varianz einer etwa invers zum quadratischen Prognosefehler gewichteten Kombination mittels einer passenden

Schätzfunktion direkt aus den Prognosen selbst zu schätzen (Hartung, Knapp, 2001b). Speziell in der Diskriminanzanalyse etwa zur Betrugs-Früherkennung kann ein Ansatz über verallgemeinerte p-Werte vorgenommen werden (Hartung, Weimann, 2000).

Literatur

Berthold, M., Hand, D.J. (1999). *Intelligent Data Analysis*. Springer, Berlin.

Draper, D., Gaver, D.P. Jr., Goel, P.K., Greenhouse, J.B., Hedges, L.V., Morris, C.N., Tucker, J.R., Waternaux, C.M. (1992). *Combining Information*. National Academy Press, Washington, D.C..

Hartung, J. (1976). Some Inequalities for a Random Function. *Teorija Verojatnostei i ee Primenenja* **21**, 661-665 (und: *Theory of Probability and its Applications* **21**, SIAM, 1977).

Hartung, J. (1981). Nonnegative Minimum Biased Invariant Estimation in Variance Component Models. *The Annals of Statistics* **9**, 278-292.

Hartung, J. (1999a). An Alternative Method for Meta-Analysis. *Biometrical Journal* **41**, 901-916.

Hartung, J. (1999b): A Short-Cut Method for Computing Positive Variance Component Estimates. In: *Quo vadis geodesia ...?* – Festschrift for Erik W. Grafarend on the occasion of his 60th birthday (Part 1), F. Krumm und V.S. Schwarze (eds.), Schriftenreihe der Institute des Studiengangs Geodäsie und Geoinformatik – Technical Reports Department of Geodesy and Geoinformatics (Report Nr. 1999.6-1), Universität Stuttgart, 151-154.

Hartung, J., Elpelt, B. (1999). *Multivariate Statistik: Lehr- und Handbuch der angewandten Statistik.* Oldenbourg, München, 6. Auflage.

Hartung, J., Hartung, J.E. (2001). WebUp: Web Usage Predictor. Ein automatischer Personalisierer im World Wide Web. *Imperia Software Solutions*, Aachen, San Francisco.

Hartung, J., Knapp, G. (2001a). On Tests of the Overall Treatment Effect in the Meta-Analysis with Normally Distributed Responses. *Statistics in Medicine*, to appear.

Hartung, J., Knapp, G. (2001b). FeDeM: Features in Data and e Mining. Ein KDD-Tool für komplexe Datenstrukturen. *Hartungs.net*, Königswinter.

Hartung, J., Weimann, B. (2000). Combining Information im Data Mining. Ein Meta-Analyse Tool zum SAS Enterprise Miner. *Information Works Unternehmensberatung und Informationssysteme*, Köln, Barcelona.

SAS Enterprise Miner (2000). *SAS Institute Inc.*, Cary.