Data Analysis and Operations Research

Wolfgang Gaul

Institut für Entscheidungstheorie und Unternehmensforschung
Universität Karlsruhe (TH), D-76128 Karlsruhe, Germany
wolfgang.gaul@wiwi.uni-karlsruhe.de

Abstract. Data Analysis and Operations Research are two overlapping sciences as there are, e.g., many data problems in which optimization techniques from Operations Research have to be applied to detect best fitting structures (under suitable constraints) in the underlying data. On the other hand, Operations Research is often based on model formulations for which some model parameters might be unknown or even unobservable. In such cases Operations Research problems consist of a data collection and analysis part and an optimization part in which solutions dependent on model parameters (derived from available information via Data Analysis techniques) are calculated.

We give typical examples for research directions where Data Analysis and Operations Research overlap, start with the topic of pyramidal clustering as one of the fields of interest of Edwin Diday, and present methodology how selected problems can be tackled via a combination of techniques from both scientific areas.

1 Introduction

When the data analysis community is the target group for a contribution concerning Data Analysis (DA) and Operations Research (OR) it is not necessary to present a list of topics that describe which kinds of data problems are of interest (see, e.g., the Springer series "Studies in Classification, Data Analysis, and Knowledge Organization" the articles of which cover nearly all aspects in this context). From a methodology-oriented point of view most textbooks in OR deal with topics as Linear/Nonlinear (Convex) Programming, Integer/Combinatorial Programming, Multicriteria Decision Making/Goal Programming and the Analytic Hierarchy Process (AHP), Dynamic Programming, Stochastic Programming, Stochastic Processes' Applications (e.g., Markov Decision Processes, Queueing Theory), Simulation and Sensitivity Analysis, Forecasting as well as Graph Theory and Network Models (see, e.g., the 6$^{\text{th}}$ edition of Domschke, Drexl (2004) or the 8$^{\text{th}}$ editions of Hillier, Lieberman (2004) or Taha (2007)). Sometimes, questions concerning problem definition, data gathering, and OR model formulation (and dependencies between these tasks) are addressed but a combination of tools from DA and OR is rarely described, explicitly. Against this background the underlying paper emphasizes situations where DA and OR overlap and presents methodology how selected problems can be tackled via a combination of techniques from both scientific areas.

2 Situations for combining data analysis and operations research

2.1 Mixed integer programming for pyramidal clustering

As starting point pyramidal clustering is selected because of the contributions of Edwin Diday to this area (see, e.g., Diday(1986, 1987), Diday, Bertrand (1986)).

The pyramidal generalization of hierarchical classification allows a certain kind of overlapping of clusters (which the hierarchical counterpart does not) where – based on a total order on the set of objects to be clustered – those objects of different clusters that are minimal or maximal with respect to the given order are candidates for overlapping.

Let $I = \{1, ..., m\}$ denote the index set of objects of interest and δ_{ij} given non-negative empirical dissimilarities between pairs (i, j) of objects (a transformation of any measure of association between pairs of objects to non-negative dissimilarities is possible in all realistic empirical situations).

Empirical dissimilarities may not be available for all pairs of objects and don't have to fulfill conditions needed for representation of the objects via, e.g., hierarchies (the ultrametric condition) or pyramids (the pyramidal condition).

The PLSC (Pyramidal Least-Squares Classification) technique is based on the following mixed integer optimization problem:

Denote by $M \subset I^2$ the set of pairs of objects for which the empirical dissimilarities are missing. Choose an initial total order $\preceq$ on I. Describe this total order and the total orders generated in subsequent steps of the solution procedure by a vector $x = (..., x_{ij}, ...)$, with

$$\begin{array}{lll} x_{ij} \in \{0,1\}, & \forall i,j \in I & \\ x_{ii} = 1, & \forall i \in I & \text{(reflexivity)} \\ x_{ij} + x_{ji} = 1, & \forall i,j \in I & \text{(antisymmetry and completeness)} \\ x_{ij} + x_{jk} - x_{ik} \leq 1, & \forall i,j,k \in I & \text{(transitivity)} \end{array} \tag{1}$$

and solve the problem

$$F(d^x) = \sum_{(i,j) \in I^2 - M} (\delta_{ij} - d^x_{ij})^2 = min \tag{2}$$

$$\begin{array}{l} d^x_{ik} \geq max\{d^x_{ij} x_{ij} x_{jk} \ , \ d^x_{jk} x_{ij} x_{jk}\} \ , \qquad \forall i,j,k \in I \\ d^x_{ij} = 0 \Leftrightarrow i = j \ , \ d^x_{ij} = d^x_{ji} \ , \ d^x_{ij} \geq 0 \ , \forall i,j \in I \end{array} \tag{3}$$

The procedure suggested in Gaul, Schader(1994) to tackle (2) under the constraints (1) and (3) can be described as follows:

Select a total order x, set $y = x, F = \infty$. Step 1: Solve (2), (3). If $F(d^x) < F$, update $y = x, d^y = d^x, F = F(d^x)$, and go to Step 2; otherwise got to Step 3. Step 2: Take y and create a new total order x_{new} from y by using the DD (Doubles Décalages) method (the DD method updates an underlying total order, for a description see, e.g., Gaul, Schader (1994), appendix c), set $x = x_{new}$, and go to Step 1. Step 3: Take x and check whether the DD method can be continued. If not, STOP with the results y and d^y; otherwise create a new total order x_{new} from x by using the DD method, set $x = x_{new}$, and go to Step 1.

Remark:

The data problem of subsection 2.1 is to find pyramidal dissimilarities – that allow visualization of clustering structures in the set of underlying objects – which best fit given empirical dissimilarities (perhaps with missing values). For the solution OR methodology based on a mixed integer programming formulation is suggested. The situation can be explained by Tables 1a, b and Figures 1a, b taken from Gaul, Schader (1994).

	1	2	3	4	5	6
1	0					
2	7	0				
3	1	7	0			
4	5	4	5	0		
5	7	2	7	5	0	
6	4	6	4	3	6	0

Table 1a: Dissimilaritiy Data Between Pairs of Objects for $I = \{1, ..., 6\}$.

	3	1	6	4	2	5
3	0					
1	1	0				
6	4	4	0			
4	5	5	3	0		
2	7	7	6	4	0	
5	7	7	6	5	2	0

Table 1b: Rearranged Dissimilarity Data of Table 1a according to the Total Order $3 \prec 1 \prec 6 \prec 4 \prec 2 \prec 5$.

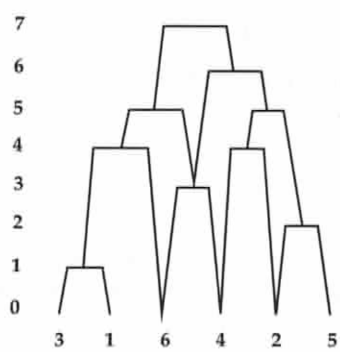

Figure 1a: Pyramidal Classification (Indexed Pyramid) of the Dissimilarity Data of Tables 1a,b.

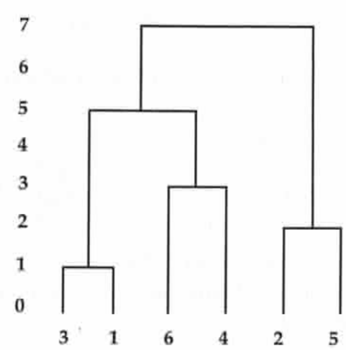

Figure 1b: Hierarchical Classification (Indexed Hierarchy, Complete-Linkage) of the Dissimilarity Data of Tables 1a,b.

Notice that the dissimilarities of Table 1a don't fulfill the ultrametric condition and that the rearrangement of these dissimilarities in Table 1b (according to the total order $3 \prec 1 \prec 6 \prec 4 \prec 2 \prec 5$) can be represented by the (indexed) pyramid of Figure 1a without any loss of information while the (indexed) dendrogram of Figure 1b gives a "poor" fit (of the dissimilarities of Table 1a (or Table 1b)).

2.2 Clustering of relations via combinatorial optimization

Let, again, $I = \{1, ..., m\}$ denote the index set of objects under consideration, and $S = \{1, ..., p\}$ the index set of given binary relations $R_1, ..., R_p$ on I. Different situations can be handled within this framework:

If S is a set of judging subjects, then $R_1, ..., R_p$ could be individual relations which result from paired comparisons with respect to the elements of I or $R_1, ..., R_p$ could be individual total orders or preorders – in other words: rankings – on the elements of I. $R_1, .., R_p$ could also be derived from a mixed data matrix $A = (a_{is}), i \in I, s \in S$, where a_{is} is the value of variable s with respect to object i. Here, S denotes a set of variables used to describe the elements of I. In this case the relations $R_s, s \in S$, are usually defined by

$$iR_sj :\Leftrightarrow a_{is} = a_{js} \text{ for a nominal variable } s,$$
$$iR_sj :\Leftrightarrow a_{is} \leq a_{js} \text{ for an ordinal or a cardinal variable } s,$$

where R_s is an equivalence relation or a complete preorder on I.

If for a relation R one uses the graph G_R with node set $N(G_R) = I$ and arc set $A(G_R) = \{(i,j) : i,j \in I \text{ and } iRj\}$, a well-known distance function for two relations R_1, R_2 is

$$d(R_1, R_2) := |A(G_{R_1}) \cup A(G_{R_2})| - |A(G_{R_1}) \cap A(G_{R_2})|.$$

With $T = \{1, ..., q\}$ as index set of target segments, i.e.,

$$S_t = \{t_1, ..., t_{p_t}\} \subset S, \quad t \in T,$$

and C_t as so-called central relation that best represents the relations contained in S_t one can now solve the problem

$$\sum_{t=1}^{q} \sum_{s \in S_t} d(R_s, C_t) = min \tag{4}$$

subject to constraints that, e.g., $\{S_1, ..., S_q\}$ is a partition of S and $C_1, ..., C_q$ are central relations on I of some specific type(s) (described by constraints similar to (1)).

Remark:

The data problem of subsection 2.2 is to find segments of similar relations and segment-specific central relations – that allow visualization of important relational structures – which best explain the information contained in a set of given relations. For the solution OR methodology based on combinatorial programming is suggested. A more detailed description and examples can be found in Gaul, Schader (1988).

2.3 Optimal positioning

Again, let $I = \{1, ..., m\}$ denote the index set of objects, $S = \{1, ..., p\}$ the index set of judging subjects, and $T = \{1, ..., q\}$ the index set of target segments to which similar subjects are assigned.

In an r-dimensional perceptual space in which the objects are represented by deterministic coordinate vectors $x_i = (x_{i1}, ..., x_{ir})', i \in I$, the target segments are described by stochastic ideal points $v_t = (v_{t1}, ..., v_{tr})', t \in T$, which are assumed to follow multivariate normal distributions $N(\mu_t, \sum_t)$. As it may be difficult for subjects to report about their ideal objects, the idea behind the presented perceptual space model is that subjects from a target segment sample an ideal point from their corresponding segment-specific ideal point distribution and give greater preferences to those objects that are nearer to their ideal points. Consequently, the notation

$$R_{i|I} = \{z \in \boldsymbol{R}^r : (z - x_i)'(z - x_i) \leq (z - x_j)'(z - x_j) \quad \forall j \in I\} \quad (5)$$

describes what could be called *preference subset* for object i (which contains all points in $\boldsymbol{R}^r$ for which i is the closest object with respect to I) and

$$p_{ti|I} = Pr(v_t \in R_{i|I}) \quad (6)$$

gives the probability that subjects from segment t prefer object i to any other object from I.

Using λ_t as a relative size of segment t $\left(\sum_{t=1}^{q} \lambda_t = 1\right)$

$$p_{i|I} = \sum_{t=1}^{q} \lambda_t p_{ti/I} \quad (7)$$

is the so-called overall share of choices for object i.

For $|I| = 2$ a closed form solution of the probability $p_{ti|I}$ of (6) is

$$p_{ti|\{i,j\}} = Pr(v_t \in R_{i|\{i,j\}}) = \Phi(\frac{x_j'x_j - x_i'x_i - 2(x_j - x_i)'\mu_t}{4(x_j - x_i)'\sum_t(x_j - x_i)}),$$

where Φ denotes the standard normal distribution,

for $|I| > 2$ an analytical solution of the probability expression (6) is not known (see, e.g., Baier, Gaul (1999), appendix, for hypercube approximation).

With $\Theta_t = (\mu_{t1}, ..., \mu_{tr}, \sigma_{t11}, \sigma_{t12}, ..., \sigma_{trr})'$ as parameter vector for $N(\mu_t, \Sigma_t)$, $t \in T$, and $\Theta = (\Theta_1', ..., \Theta_q')'$ as overall parameter vector the data collection and parameter estimation part of the optimal positioning problem can be described as follows:

Paired comparisons $Y = (y_{sij}), s \in S, i, j \in I$, with

$$y_{sij} = \begin{cases} 1 & , \text{ if subject } s \text{ prefers object } i \text{ to object } j, \\ 0 & , \text{ otherwise,} \end{cases}$$

are collected (Note that this notation allows for missing values in the data.). As segment-specific model parameters Θ_t are needed, an additional segmentation matrix $H = (h_{ts}), t \in T, s \in S$, with

$$h_{ts} = \begin{cases} 1 & , \quad \text{if subject } s \text{ belongs to segment } t, \\ 0 & , \quad \text{otherwise,} \end{cases}$$

is introduced (from which one gets the relative segment sizes, $\lambda_t = \sum_{s=1}^{p} h_{ts}/p$).

The parameter estimation part will not be explained in detail. A simultaneous technique for jointly determining Θ and H (based on the classification maximum likelihood method which incorporates a quasi-Newton procedure) is used. Notice that the negative log-likelihood function

$$\begin{aligned} -lnL(\Theta, H|Y) &= -\sum_{t=1}^{q}\sum_{i=1}^{m}\sum_{j\in I\setminus\{i\}}\Big(\sum_{s=1}^{p} h_{ts}y_{sij}\Big)ln(p_{ti|\{i,j\}}) \\ &= -\sum_{s=1}^{p}\sum_{t=1}^{q} h_{ts}\Big(\sum_{i=1}^{m}\sum_{j\in I\setminus\{i\}} y_{sij}ln(p_{ti|\{i,j\}})\Big) \\ &= -\sum_{s=1}^{p}\sum_{t=1}^{q} h_{ts}L_{ts}(\Theta_t|Y) \end{aligned} \tag{8}$$

allows simplifications for given H. The determination of H is improved by allocating subjects to segments in such a way that (8) is minimized.
Based on the estimated parameters overall shares of choices for the given objects (see (7)) can be predicted.
For optimal positioning of a new object assume that I is enlarged to $I_0 = I \cup \{0\}$ where 0 describes the additional alternative.
If the new object is positioned at $x_0 = (x_{01}, ..., x_{0r})'$ one gets the preference subset

$$R_{0|I_0}(x_o) = \{z \in \mathbb{R}^r : (z - x_0)'(z - x_0) \leq (z - x_j)'(z - x_j) \quad \forall j \in I_0\}$$

and

$$p_{0|I_0}(x_o) = \sum_{t=1}^{q} \lambda_t p_{t0|I_0}(x_o) \tag{9}$$

as overall share of choices for the new object (dependent on x_0).
Now, optimal positioning options for the new object can be obtained through maximizing (9) by one of the adequate positioning techniques listed in Baier, Gaul (1999), Table 3, which gives a quite complete overview concerning references up to the end of the nineties of the last century.

Remark:

The data problem of subsection 2.3 is to find segment-specific stochastic ideal points described by multivariate normal distributions – that allow visualization of the underlying choice situation in corresponding perceptual spaces

– which best explain the preferences of segments of subjects contained in given individual paired comparisons. OR methodology (e.g., a quasi-Newton procedure) is already incorporated in the classification maximum likelihood method for the estimation of the model parameters. For the generation of positioning options of a new object in the given perceptual space a standard hill-climbing algorithm of nonlinear programming was applied. A more detailed description with Monte Carlo experiment and application can be found in Baier, Gaul(1999).

2.4 Random variables in operations research models

This time, the starting point is an OR model – a linear program, say – in which some model parameters have to be viewed as random variables, which is the basic assumption of stochastic programming (see, e.g., Kall (1979) for an early and Kall, Wallace (1994) for a more recent textbook concerning this OR field). Let

$$\begin{aligned} c'x &= min \\ Ax &= b \\ x &\geq 0 \end{aligned} \tag{10}$$

be a standard linear program with known $m \times n-$matrix $A = (a_{ij}), b = (..., b_i, ...)' \in \boldsymbol{R}^m$, and $c = (..., c_j, ...)' \in \boldsymbol{R}^n$ for which the decisions $\{x : Ax = b, x \geq 0\}$ form a closed convex set. Assume there exist additional constraints

$$Bx = d \tag{11}$$

with $\widetilde{m} \times n$-matrix $B = (b_{ij})$ and $d = (..., d_i, ...)' \in \boldsymbol{R}^{\widetilde{m}}$ where – for simplicity – only d is assumed to be a random vector on a probability space $(\Omega, \mathfrak{S}, Pr)$ for which the expectation E_d exists. If the realization $d(\omega), \omega \in \Omega$, is known before the decision x has to be calculated, the problem is "easy". If x has to be determined before the realization of d is known

$$Q(x, d) = \inf\{cc'_+ y_+ + cc'_- y_- : y_+ - y_- = d - Bx, y_+ \geq 0, y_- \geq 0\}$$

describes a possibility for compensation (a so-called simple recourse compensation) with compensation costs $cc_+, cc_- \in \boldsymbol{R}^{\widetilde{m}}$. If $cc_+ - cc_- \geq 0$ the so-called two-stage stochastic programming problem with simple recourse

$$\begin{aligned} c'x + E_d[Q(x, d)] &= \min \\ Ax &= b \\ x &\geq 0 \end{aligned} \tag{12}$$

solves (10) and (11) in the sense that x is selected in such a way that non-conformity of Bx with d in (11) is optimally compensated.

Now, from the data problem point of view the probability distribution of d is of importance for the solution of (12). Here, it is assumed that the components d_i have finite discrete probability distributions (or that the corresponding distributions are approximated by finite discrete probability distributions), i.e.,

$$p_{ik} = Pr(d_i = d_{ik}) \quad , \quad k = 1, ..., r_i \quad , \quad i = 1, ..., \widetilde{m} \quad ,$$

(with lower (upper) bounds d_{i0} (d_{ir_i+1}) with $p_{i0} = 0$ $(p_{ir_i+1} = 0)$) are taken into consideration.

For a selected realization $d_{k^*} = (d_{1k_1^*}, ..., d_{\widetilde{m}\,k_{\widetilde{m}}^*})'$ of the vector d one solves the following dual problems

PRIMAL (k^*)

$$\begin{aligned}
&\sum_{j=1}^{n}\Big(c_j + \sum_{i=1}^{\widetilde{m}}\big(-(cc_+)_i + ((cc_+)_i + (cc_-)_i)\sum_{k=1}^{k_i^*} p_{ik}\big)b_{ij}\Big)x_j = min \\
&\sum_{j=1}^{n} a_{ij}x_j \qquad\qquad\qquad = b_i \qquad , \quad i = 1, ..., m \\
&\sum_{j=1}^{n} b_{ij}x_j \quad - s_{1i} \qquad\quad = d_{ik_i^*} \qquad , \quad i = 1, ..., \widetilde{m} \\
-&\sum_{j=1}^{n} b_{ij}x_j \qquad\quad - s_{2i} \quad = -d_{i(k_i^*+1)} \; , \quad i = 1, ..., \widetilde{m} \\
&\qquad x_{ij} \geq 0, s_{1i} \geq 0, s_{2i} \geq 0
\end{aligned}$$

DUAL (k^*)

$$\begin{aligned}
&\sum_{i=1}^{m} b_i u_i + \sum_{i=1}^{\widetilde{m}} d_{ik_i^*} v_{1i} - \sum_{i=1}^{\widetilde{m}} d_{i(k_i^*+1)} v_{2i} = max \\
&\sum_{i=1}^{m} u_i a_{ij} + \sum_{i=1}^{\widetilde{m}} v_{1i} b_{ij} - \sum_{i=1}^{\widetilde{m}} v_{2i} b_{ij} \qquad \leq \mathfrak{c}_j \quad , \quad j = 1, ..., n \\
&\qquad\qquad v_{1i} \geq 0 \quad , \quad v_{2i} \geq 0
\end{aligned}$$

$$\text{with } \mathfrak{c}_j = c_j + \sum_{i=1}^{\widetilde{m}}\big(-(cc_+)_i + ((cc_+)_i + (cc_-)_i)\sum_{k=1}^{k_i^*} p_{ik}\big)b_{ij} \; .$$

Notice that PRIMAL (k^*) and DUAL (k^*) are optimization problems on the grid given by the realizations of the vector d. For a selected k^* optimization is performed between d_{k^*} and d_{k^*+e}, $e = (1, ..., 1)'$.

If $\widetilde{x}$, $\widetilde{s_1}$, $\widetilde{s_2}$ (for PRIMAL (k^*)) and $\widetilde{u}$, $\widetilde{v_1}$, $\widetilde{v_2}$ (for DUAL (k^*)) are complementary optimal solutions and

$$\widetilde{v_{1i}} \leq \big((cc_+)_i + (cc_-)_i\big)p_{ik_i^*} \quad , \quad i = 1, ..., \widetilde{m} \tag{13}$$

$$\widetilde{v_{2i}} \leq \big((cc_+)_i + (cc_-)_i\big)p_{ik_i^*+1} \quad , \quad i = 1, ..., \widetilde{m} \tag{14}$$

then $\widetilde{x}$ is optimal for the two-stage stochastic programming problem with simple recourse (12), otherwise update

$$k_i^{*(new)} = k_i^{*(old)} + \begin{cases} (-1)\,, & \text{if (13) is violated,} \\ 1\,, & \text{if (14) is violated,} \quad i = 1, ..., \widetilde{m}\,, \\ 0\,, & \text{otherwise,} \end{cases} \tag{15}$$

and solve PRIMAL ($k^{*(new)}$), DUAL ($k^{*(new)}$). Under reasonable assumptions an optimal solution is obtained after a finite number of iterations.

Remark:

A stochastic programming problem is solved by a finite sequence of "easier to handle" non-stochastic PRIMAL/DUAL problems. It is, of course, advantageous, when the PRIMAL/DUAL problems are of a special form for which fast solutions are already available.

Notice, that sometimes the dual problem of an initial linear program is of the form described by (10), (11). For an application of the described OR methodology to project scheduling via stochastic programming (in which project activity times are random variables) see Cleef, Gaul (1982) where the "easier to handle" PRIMAL/DUAL problems are based on network models (e.g., solving minimal cost flow problems by the "out-of-kilter" algorithm).

2.5 Web mining and recommender systems

Nowadays, contributions concerning DA and OR have to cope with web mining because the web as one of the fastest growing sources of information is a challenge for data analysts. Here, a recent reference is Gaul (2006) (see also Gaul (2004)) in which certain topics (concerning web data, data analysis techniques, and web mining applications) are presented that will not be repeated in this paper. However, at least recommender systems, e.g., for clickstream analysis (see, e.g., Gaul, Schmidt-Thieme (2000, 2002)), should be mentioned, explicitly, as – in the narrow sense – these systems tackle data problems. Here, data (input) has to be analysed in such a way by DA and/or OR techniques that recommendations (output) for target segments can be provided.

3 Conclusion

DA (Data Analysis) and OR (Operations Research) techniques are needed in quite a number of situations in which on the basis of underlying data (sometimes with missing values) "optimal" solutions for target groups have to be calculated. Thus, it seems to be worth while to consider research directions where DA and OR overlap. In this paper, a constrained optimization formulation for pyramidal clustering was the starting point for a collection of examples (in which, e.g., combinatorial programming, optimization techniques to calculate maximum likelihood estimates, algorithms for optimal positioning, and stochastic programming were applied) that describe situations where a combination of DA and OR has to be used to solve the underlying problems. As KIT (Karlsruhe Institute of Technology, a merger of the Forschungszentrum Karlsruhe and the Universität Karlsruhe) was elected as one of the best German universities in 2006, new courses will be established in 2007 and one of it is "Data Analysis and Operations Research". Here, hints and remarks concerning additional examples, situations, and solutions are welcome.

References

BAIER, D. and GAUL, W. (1999): Optimal product positioning based on paired comparisons data. *Journal of Econometrics, 89, 365-392.*

CLEEF, H.J. and GAUL, W. (1982): Project scheduling via stochastic programming. *Mathematische Operationsforschung und Statistik, Ser. Optimization, 13, 449-468.*

DIDAY, E. (1986): Orders and overlapping clusters by pyramids. In: J. De Leeuw, W. Heiser, J. Meulman and F. Critchley (Eds.): *Multidimensional Data Analysis.* DSWO, Leiden, 201-234.

DIDAY, E. (1987): *Orders and Overlapping Clusters by Pyramids.* Rapport de Recherche No 730. INRIA, Paris.

DIDAY, E. and BERTRAND, P. (1986): An extension of hierarchical clustering: the pyramidal representation. In: E.S. Gelsema and L.N. Kanal (Eds.): *Pattern Recognition in Practice II.* North-Holland, 411-424.

DOMSCHKE, W. and DREXL, A. (2004): *Einführung in das Operations Research (6. Auflage).* Springer.

GAUL, W. (2004): Market research and the rise of the web: the challenge. In: Y. (Jerry) Wind and P.E. Green (Eds.): *Market Research and Modeling: Progress and Prospects: A Tribute to P.E. Green. International Series in Quantitative Marketing.* Kluwer, 103-113.

GAUL, W. (2006): Challenges concerning web data mining (Invited Paper). In: A. Rizzi and M. Vichi (Eds.): *COMPSTAT 2006.* Physica, 403-416.

GAUL, W. and SCHADER, M. (1988): Clusterwise aggregation of relations. *Applied Stochastic Models and Data Analysis, 4, 273-282.*

GAUL, W. and SCHADER, M. (1994): Pyramidal classification based on incomplete dissimilarity data. *Journal of Classification, 11, 171-193.*

GAUL, W. and SCHMIDT-THIEME, L. (2000): Frequent generalized subsequences – a problem from web mining. In: W. Gaul, O. Opitz, and M. Schader (Eds.): *Data Analysis: Scientific Modeling and Practical Application. Studies in Classification, Data Analysis, and Knowledge Organization.* Springer, 429-445.

GAUL, W. and SCHMIDT-THIEME, L. (2002): Recommender systems based on user navigational behavior in the internet. *Behaviormetrika, 29, 1-29.*

HILLIER, F.S. and LIEBERMAN, G.J. (2004): *Introduction to Operations Research (8^{th} Edition).* McGraw-Hill.

KALL, P. (1979): *Stochastic Linear Programming.* Springer.

KALL, P. and WALLACE, S. (1994): *Stochastic Programming.* Wiley.

TAHA, H.A.(2007): *Operations Research: An Introduction (8^{th} Edition).* Prentice Hall.