APPLIED STOCHASTIC MODELS AND DATA ANALYSIS, VOL. 5, 341–356 (1989)

Invited paper

DATA ANALYSIS AND DECISION SUPPORT

WOLFGANG GAUL

Institut für Entscheidungstheorie und Unternehmensforschung, Universität Karlsruhe, Kaiserstr. 12, D-7500 Karlsruhe 1, West Germany

AND

MARTIN SCHADER

Institut für Informatik, Fachbereich WOW, Universität der Bundeswehr Hamburg, Holstenhofweg 85, D-2000 Hamburg 70, West Germany

SUMMARY

In this paper, we consider developmental lines of computer-assisted decision support (with consideration of knowledge-based approaches) for data analysis problems. First, we discuss some situations where it is obviously appropriate to apply computer-assisted decision support in connection with data analysis tasks. Then, a brief historical retrospect is given viewing the development of this area of research and its interfaces to knowledge-based approaches. Against this background we illustrate two prototypes of knowledge-based decision support systems for specific data-analysis problems related to fields of interest of our own. Finally, we indicate possible progress and future activities in this area.

KEY WORDS Data analysis Computer-assisted decision support Statistical expert systems

1. MOTIVATION

The early (scientific) activities of mankind already included tasks of gathering and analysing data and information (concerning, for instance, weather conditions and cropping results and whatever omens to indicate possible future events) which could be used in order to prepare and/or support particular decisions.

Therefore, at the beginning of this article one of the first questions arising is whether it is really necessary to write another paper about such an 'old' subject as *data analysis and decision support* or if, on the contrary, one has to do so due to the 'topicality' of this theme. Obviously, we decided in favour of the second possibility and we hope to be able to demonstrate trends of new developments in this area, especially by including knowledge-based approaches. Consequently, we will limit the historical retrospect—with emphasis on computer-assisted decision support and consideration on possible applications of knowledge-based approaches—to a time period where the interdependencies between 'computer science' and 'data analysis and statistics' as well as the state of the art in the respective domains admitted the development of software already suitable to some extent for business applications.

1.1. Some potential fields of application of computer-assisted decision support in data analysis

The existence of a close association between the concepts of 'data analysis' and 'decision support' is quite obvious, as shown in the historical remark at the outset. In a narrow sense

8755–0024/89/040341–16$08.00

Received 8 December 1988

data analysis itself can be viewed as a tool of decision support; in a broad sense one has to consider which aids are desirable for reasonable computer-aided decision support with respect to data analysis tasks. At the same time—if we regard the processing of data analysis problems as a procedure with the phases 'problem identification' $\overset{(1)}{\rightarrow}$ 'data acquisition and preparation' $\overset{(2)}{\rightarrow}$ 'data analysis' $\overset{(3)}{\rightarrow}$ 'decision'—one should reflect on possibilities to support each of these phases appropriately.

Obviously, one desirable characteristic of decision-support software in the data analysis area would be the ability to offer—on the basis of suitably prepared data (e.g. by application of appropriate data analysis techniques)—a choice of alternatives of decisions (or at least references). Here, mainly aid with respect to the transition $\overset{(3)}{\rightarrow}$ would be given. (A prototype development which copes with this task is described in Section 2.1. The prototype gives support for track selection and traffic control planning in train stations and signal cabins.[1])

An additional desirable feature of decision support software in the data analysis area would be the ability to provide reasonably applicable data analysis techniques (or at least references) if adequate data are presented. Here, mainly aid with respect to the transition $\overset{(2)}{\rightarrow}$ would be given. (A prototype development—called DANEX (Data ANalysis EXpert)—which copes with this task is described in Section 2.2. The present implementation of DANEX supports the selection of methods from 'cluster analysis' and 'multidimensional scaling'.[2,3] An object-oriented approach supporting method selection in the case of qualitative or mixed data problems is described by Tüshaus.[4])

Finally, a further desirable characteristic of decision support software in the data analysis area would be the ability to provide information on the nature of data required (or at least references) if the problem to be handled is suitably presented. Here, mainly aid with respect to the transition $\overset{(1)}{\rightarrow}$ would be given.

In Figure 1 the aspects of decision support possibilities just discussed are shown again. In addition, Figure 1 includes references to knowledge domains which might be considered useful when computer-based support systems have to be designed.

First reflections and questions connected with selecting and preparing appropriate data arise in the 'problem identification' phase. Here, proper decision support software could initiate an inquiry within the system-inherent database (if existing) with regard to useful secondary data material or list references concerning the availability of potentially suitable secondary data or sources of required primary data material and indications of their accessibility.

In addition to possible support for data collection (questions of measurement, observation, conduct of interviews, preparation of questionnaires or parts of them, etc.) appropriate methods for data cleaning, aggregation, transformation and examination (e.g. problems of outliers, missing values, assumptions concerning distributions, correlations, etc.) are needed in the 'data acquisition and preparation' phase.

Here the field of applications of classical statistical techniques commences, although the main area of its use will be the 'data analysis' phase. However, statistical and data analysis software often constitutes just a small part of the existing software offered. This variety of software tools may be already confusing to a non-(data analysis software) expert. In addition to the many 'stand-alone programs' also 'program libraries', 'program systems', 'method base systems', 'planning systems', 'spreadsheet programs' and 'database systems'—which often comprise software for data analysis and statistical computation to a more or less great extent—are to be mentioned (see, e.g., Reference 5, for PC-software for statistical analysis).

In order to be able to differentiate (at least roughly) between the software concepts shown in Figure 1 and to make allowance for the emphasis on computer-assisted decision support noted at the outset, the short descriptions of the above-mentioned software tools as shown in Table I could, perhaps, be useful.

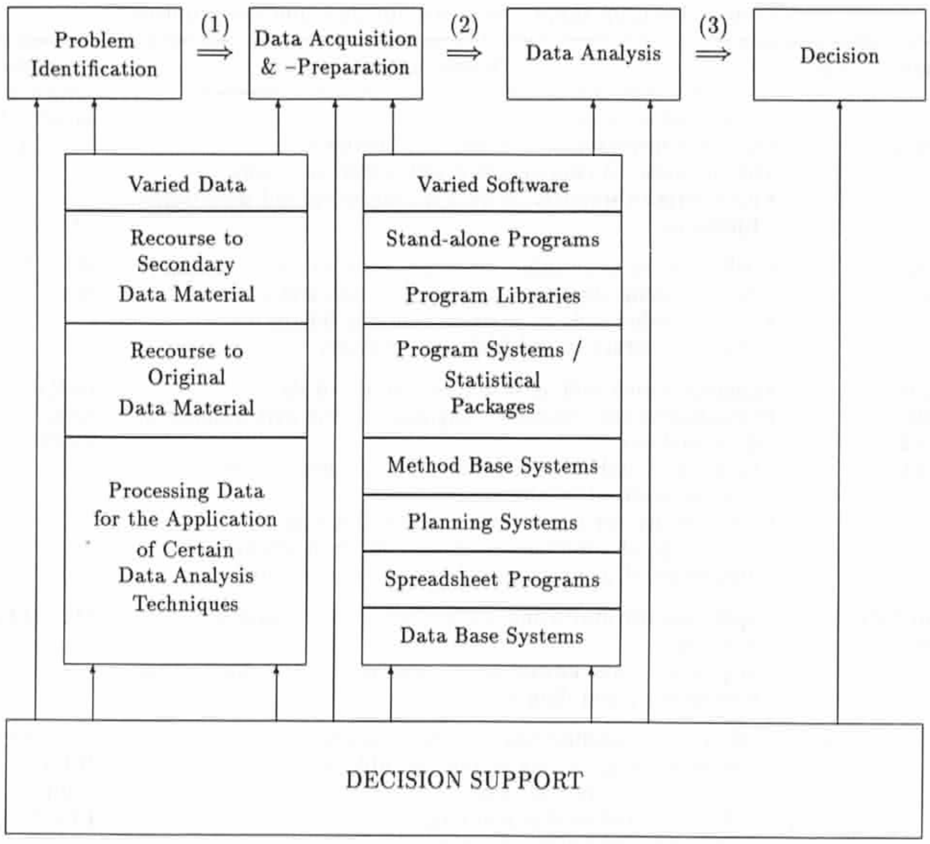

Figure 1. Possibilities of decision support in data analysis

Decision support software for data analysis should provide knowledge of existing software (Table I can only give a glimpse of it) and the reasonableness of its application.

If possible, the best qualified techniques should be selected from the available software and applied to the data under consideration. Then, the results obtained should be prepared for the 'decision' phase in such a way that inspection of different alternatives and determination of a final decision is supported.

So far, the discussion of some aspects where computer-assisted decision support in the data analysis area might be applied reasonably has described some outlines of a concept of a desirable software tool. It has—at the same time—shown that for obtaining an adequate realization of such a concept there is a long way to go yet. To end with, some historical remarks concerning the present lines of development of computer-assisted decision support in the data analysis area with regard to knowledge-based approaches have to be given; this discussion was deliberately omitted in the introductory part of this paper.

1.2. Historical retrospect and trends incorporating knowledge-based approaches

During the discussion of computer-based decision support the term 'expert system' came into vogue. The problem is that this expression seems so clear to the non-specialist, and that meanwhile expectations have become attached to it that it starts to trouble the researcher who

Table I. Decision support software for data analysis problems

Software concepts	Description	Examples
Stand-alone programs	• oldest software concept • possibly novel types of model and methods • no intention to communicate with other programs • no efforts to standardize data management and data-transfer functions	MAPCLUS[6] PENCLUS[7]
Program libraries	• collection of subroutines/program units written in a specific programming language; satisfying certain conventions • user-provided code of drivers/programs linking up the available library subroutines are necessary	IMSL[8] NAG[9]
Program systems/ Statistical packages	• concept rather aiming at the evaluation of data • system-inherent command language for the performance of data analyses • procedures and functions for data management and interpretation of results are provided • no necessity for user-provided driver routines • knowledge of command language, of implications and modalities of using built-in procedures is required	BMDP[10] SAS[11] SPSS[12]
Method base systems	• approach for improving the performance of program libraries • attempts to take advantage of analogies concerning method management and data management	METHAPLAN[13]
Planning systems	• support of planning and reporting processes • classification according to the available functions into —cost and budget planning —financial and capital planning —sales and marketing planning —strategic planning	FCS–EPS IFPS INFPLAN PLANCODE[14,15]
Spreadsheet programs	• software concept which has profited by spreading of personal computers • system kernel consists of a program handling data in tabulated form • system-inherent programming language if more complex problems are to be solved	VISICALC MULTIPLAN JAZZ SYMPHONY
Database systems	• most important feature is data management • suitable interfaces to statistical software packages not always available	SIR ORACLE

is interested in future development in this area. As is known from similar events in connection with 'management information systems (MIS)' the failure to comply with such expectations can have unfavourable consequences. The MIS discussion, for instance, died away fast after a short 'euphoric' phase. It is hoped that a similar process concerning 'expert systems' will not be repeated.

Attempts to answer the question 'What *are* expert systems?' can be found in many sources. The following text is quoted from Reference 16, p. 1: 'expert systems can thus be said to grasp fundamental domain principles ... to solve complex problems and to interact intelligibly with the user'. Hahn, (Reference 17, p. 1), puts it in a similar way: 'An expert system is one in which knowledge of an expert is built into a computer program, enabling it to emulate the process that the expert follows in attacking a problem'.

As a more detailed quotation we include the summary of chapter 2, 'What are expert systems?' by Brachman, Amarel, Engelman, Engelmore, Feigenbaum and Wilkins, from Reference 18, p. 50: 'An expert system is one that has expert rules and avoids blind search, performs well, reasons by manipulating symbols, grasps fundamental domain principles, and has complete weaker reasoning methods to fall back on when expert rules fail and to use in producing explanations. It deals with difficult problems in a complex domain, can take a problem description in lay terms and convert it to an internal representation appropriate for processing with its expert rules, and it can reason about its own knowledge (or lack thereof), especially to reconstruct inference paths rationally for explanation and self-justification. An expert system works on (generally at least) one of these types of task: interpretation, diagnosis, prediction, instruction, monitoring, planning, and design'.

These attempts to provide a definition of the term 'expert system' show in which framework the field of research wants to progress. They also show that not all the software offered under

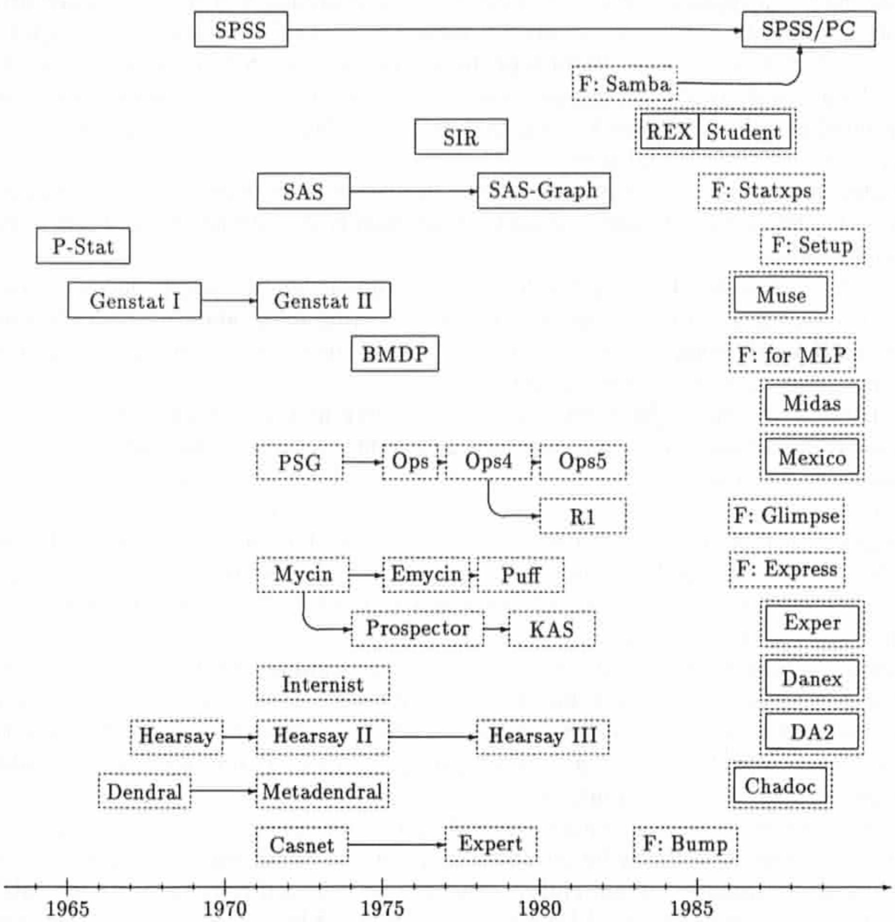

Figure 2. Evolution of statistical and knowledge-based decision support software (F: front ends)

the expert system label may be judged by such set standards. One essential feature is the existence of a knowledge base comprising the expert knowledge which has to be consulted for elaborating and controlling computer-assisted decision support. In the following we will therefore use the less attractive term 'knowledge-based approaches', thus laying stress on the fundamental feature of such software.

Figure 2 shows the chronology of some of the known knowledge-based approaches (boxes with dashed lines) together with some data analysis systems (boxes with straight lines) which will be discussed later. DENDRAL supports the analysis of mass spectrographic data of unknown material, possible molecular structures are generated from the input data and subsequently tested; HEARSAY is able to interact with the user in natural language on the basis of limited knowledge of vocabulary and grammatical rules; INTERNIST supports differential diagnosis of organ system diseases; MYCIN assists physicians in diagnosing and therapy selection of infectious blood diseases; from PSG, a production system language for investigating human memory and cognition, the OPS series of production systems emerged and led to R1, a system for configuring computer systems; EXPERT is a tool for building expert systems which was derived from CASNET, a knowledge-based system for diagnosis and treatment of eye diseases (for more detailed descriptions see, e.g., References 18–21; in Reference 22 potential applications of knowledge-based approaches in economics are covered).

This short list may be sufficient to indicate some features of knowledge-based approaches that have proved to be successful. Besides problem-oriented expert knowledge being recorded in the knowledge base, tools are needed which are able to examine current input data, to perform an intelligent search in the knowledge base, to extract, compare and connect relevant knowledge, to acquire additional information, etc.

In the sixties these systems were mainly written in Lisp or even in procedural languages such as Fortran or Cobol. Since the early seventies (and mainly in Europe) Prolog has come into use more and more.

Thus, within the framework of the development of knowledge-based approaches, it was only a matter of time until a closer connection with the existing data analysis software would be established; particularly since at the end of the sixties a trend had started off which Victor[23] defined by the term 'computational statistics'.

Figure 3 (similar to one in Reference 23) shows some activities that may illustrate these developments (the publication of new journals or special sections in journals, the holding of scientific conferences—even the foundation of SFC, the French Classification Society, took place at that time).

Additionally, together with the course of development of data analysis software, such as BMDP, SAS, SPSS, the prerequisites for interlocking the fields 'computer-assisted data analysis' and 'knowledge-based systems' were met not later than in the early eighties (see Table I and Figure 2 for examples).

Reference 24 is one of the early references stating—among other things—an enormous amount of uncritical use of standard data analysis methods and therefore demanding guidance by intelligent software support; Chambers[25] argues similarly since the increasing availability of statistical software would result in a corresponding increase of uninformed, unguided and incorrectly performed statistical analyses.

Inspecting the attempts now starting to provide guidance for an inexperienced user of data analysis software, two trends may be observed. On the one hand, user interfaces for complex statistical packages—so-called 'front ends'—have been developed (see, for instance, BUMP, a front end for MULTIVARIANCE,[26] EXPRESS and STATXPS, respectively, both supporting data analysis using BMDP,[27,28] GLIMPSE, a front end for GLIM,[29] a front end for MPL,[30] SETUP, supporting P-STAT,[31] as well as SAMBA, an interface to SPSS/PC+ [32]). On the other

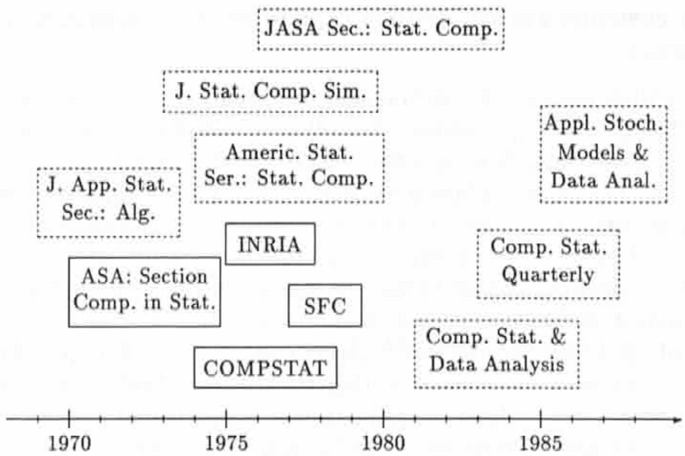

Figure 3. First conferences and journals with sections/series on computer-assisted decision support and data analysis

hand, programming of prototypes began where one tried to combine user-guidance for specified data analysis tasks with some kind of automatic carrying out of analyses [see, e.g., REX (with STUDENT as its successor) probably the best known system of this kind, supporting the user in applying regression techniques[33]].

Several of these references are drawn from the COMPSTAT'84 and COMPSTAT'86 proceedings volumes (in this context also MUSE,[34] CHADOC[35] and Reference 36 should be mentioned), for some of the still more recent papers partly shown in Figure 2 we refer to Reference 37. Here, among introductory papers (e.g., Reference 38 on the interfaces among data analysis, knowledge representation and applications) and a contribution concerning the DANEX prototype described in Section 2.2, papers are to be found which deal with, e.g., rule-based credit checking in a mail-order firm,[39] decision support for personnel disposition in a garage,[40] expert debt management in a public company,[41] incorporating knowledge for decision support using EXPER,[42] supporting new product introduction via MEXICO,[43] generating rules from data,[44] fuzzy concepts in knowledge processing,[45] giving buy/sell recommendations in stock and options markets[46] or linking Fortran programs to the expert system shell KEE.[47]

Moreover, concepts concerning the design philosophy of decision support systems which can be found in Reference 48 together with reflections of Hand[49–51] on statistical expert systems, of Thisted,[52] on representing statistical knowledge and of Hand,[53] on strategies for data analysis, will be adopted in joint research of the authors to implement software which will support the transition $\overset{(2)}{\rightarrow}$ of Figure 1. Preliminary work (see, e.g., Reference 54 on computer-based decision support and References 55 and 56 on methods for analysing qualitative data) will be combined for such a joint prototype of computer-assisted decision support for problems from the data analysis area. In the following, we will report on some of our first attempts to provide such support on the basis of knowledge-based approaches.

2. EXAMPLES FROM OWN FIELDS OF INTEREST

In this section, we will discuss two prototype developments from different research areas we are interested in to indicate how concrete realizations of further development in the data analysis and decision support field might appear.

2.1. Prototype of computer-assisted decision support for track selection in train stations and signal cabins

Trains running—with considerable differences in speed—on the transport way connections of the 'Deutsche Bundesbahn' are today controlled by graphic timetables. The schedule for running through stations and passing other trains is also described in the form of such route–time diagrams. Figure 4 displays part of such a timetable (approximately 2 per cent of its actual size). This case is concerned with a double-track north–south connection where passenger as well as freight traffic is handled. Lines drawn from upper left to bottom right represent trains going north (and vice versa). In the original plan passenger trains and freight trains are distinguished by black and blue lines, respectively.

Time is displayed vertically, so that fast trains correspond to flat slopes. From the timetable shown in Figure 4 one can see for instance, that freight train 51888 arrives in station Bad Be at 20.10 hours and leaves at 20.14. At the same time it happens that express D780 runs through the same station in the same direction at 20.12 hours. To enable the passing of D780 the operator responsible has to switch 51888's route to a side-track and to set the matching departure signal to red. Essential for the reasonableness of this arrangement is that both trains are on time: if, for instance D780 is 3 minutes late it would be pointless to allow freight train

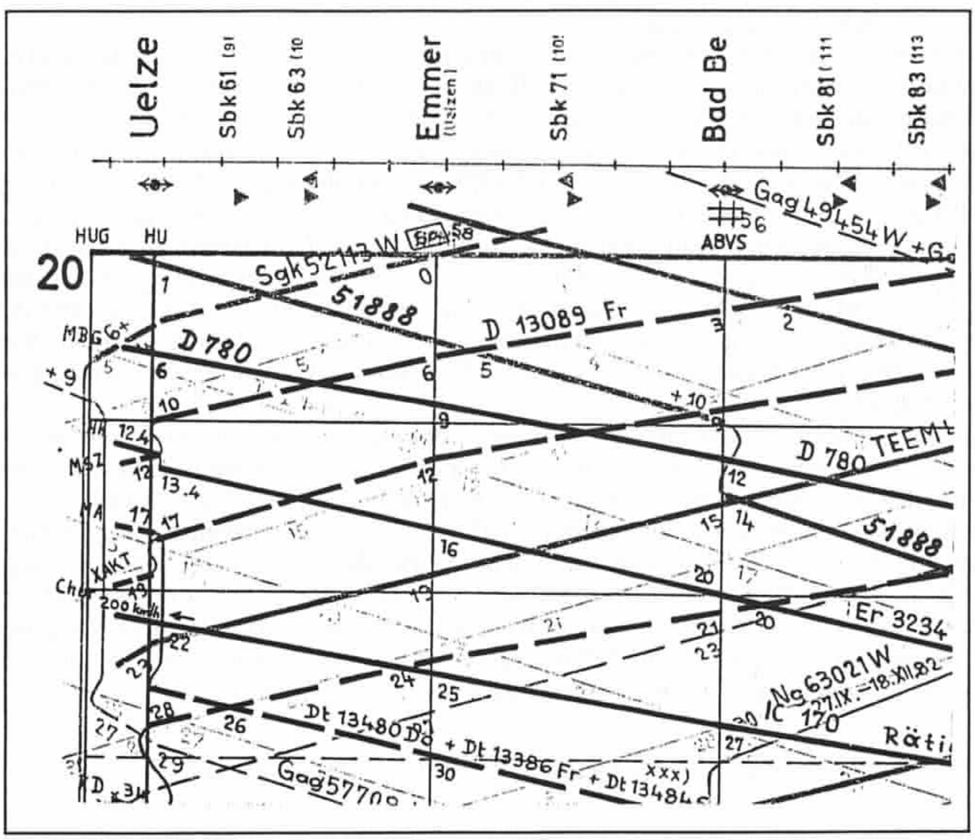

Figure 4. Part of a timetable

51888 to depart at 20.14 hours since D780 would catch up with it very quickly and then would have to follow slowly (and would be 6 minutes late in the next station).

In addition to such interventions due to trains being early or late, operators have to act differing from the time tables particularly if parts of tracks are closed (due to repair or maintenance work or train break-down), if switches are blocked (due to weather conditions), if signals are down, etc.

The situation just described is concerned with decision support in a quickly changing database. Recurring to the presentation of the processing of data analysis problems as a phase procedure as described in Figure 1, here the phases 'data acquisition and preparation' and 'data analysis' are less significant, since we can assume that delays of trains, closing of tracks, etc. are already known, so that mainly the transition $\overset{(3)}{\rightarrow}$ of Figure 1 will need support.

By way of example, we display in Figure 5 the station Stelle with its five tracks—in the way the operator sees it from his signal cabin.

Tracks 1 and 4 have platforms where public transit trains stop. Express trains usually run through Stelle on track 2 heading south and on track 3 heading north, respectively. On these tracks trains can move at a speed of 160 km/h. Track 5 runs into a viaduct, enabling the passage of freight trains from Hannover towards Maschen (Germany's largest freight depot) without blocking tracks 2 and 3. Stelle is controlled by 13 signals labelled A, F, G, Lsf1, Lsf2, Lsf3, N1, N3, N4, P1, P2, P4, P5. A, F and G are signals on the station's inbound tracks. Lsf1, Lsf2 and Lsf3 are departure signals. Signals beginning with an N or a P allow trains to halt in the station itself.

In order to support track selection a prototype was developed in Turbo-Prolog. A typical screen dump of this program is shown in Figure 6.

One can see that track 2 is interlocked for the passage of express 677 (on the screen this route is highlighted and coloured differently). Since 677 arrives 6 minutes late (this information is

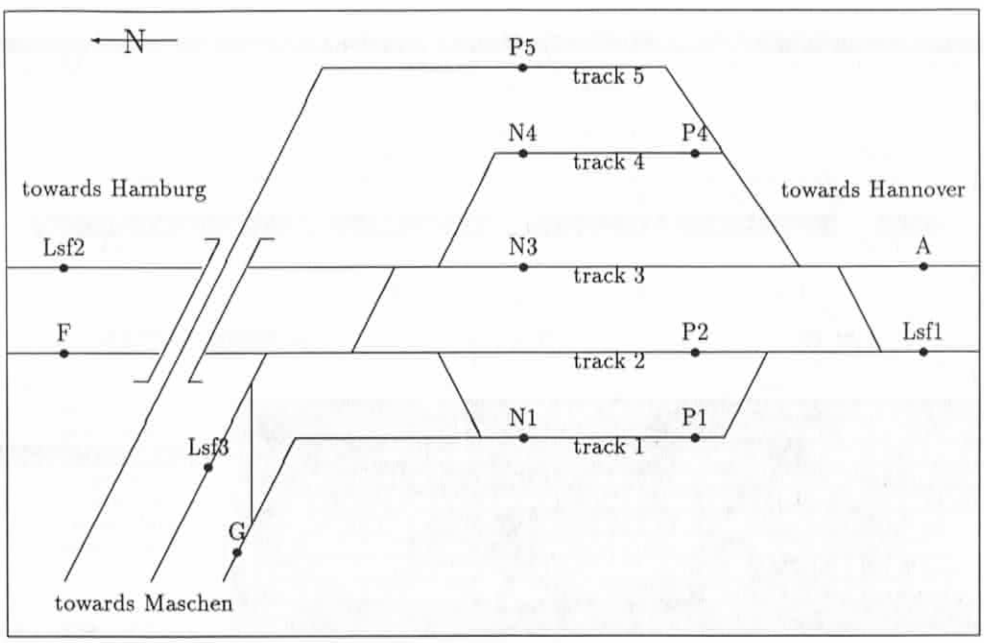

Figure 5. Track layout plan of the station Stelle

given by '17.17h + 6'), it is proposed to halt freight train 47431 for one minute at signal G until 677 has passed Stelle (according to schedule 47431 should run through Stelle at 17.23 hours without halt—going from Maschen to Ashausen, the next station in the direction of Hannover). Owing to a shutdown of track 5, freight train 47974, usually running towards Maschen via track 5, should be redirected to track 3. This implies a crossing over track 2 but does not require further interventions at the current planning stage (also, passage on track 3 is faster than using track 4).

In the input window (bottom right) of Figure 6 the main menu is shown; depending on the chosen item other windows are pulled down. Following each user input the screen and diverse program stacks must be updated. These updates always follow the same scheme:

1. Determining the current list of the next eight trains. Here, at first, it is to be analysed if previously halted trains can depart.
2. Planning of routes for these eight trains (to be displayed in the bottom left window); the routes are not yet interlocked.
3. Updating the track layout window, for instance, by emphasizing trains running through or halting at a signal, indicating closed tracks, etc.
4. Displaying the main menu in the input window and reading the next input.

As an example for a rule from the knowledge base used we will describe the selection of track 1 for freight train 47431 coming from signal G (that is, starting in Maschen) and running towards Ashausen (in the direction of Hannover) via Stelle. This is permissible if:

(a) track 1 is not blocked *and*
(b) track 1 is not interlocked for another train in the current or following minute *and*
(c) track 2 is not interlocked for another train in the current or following minute *and*

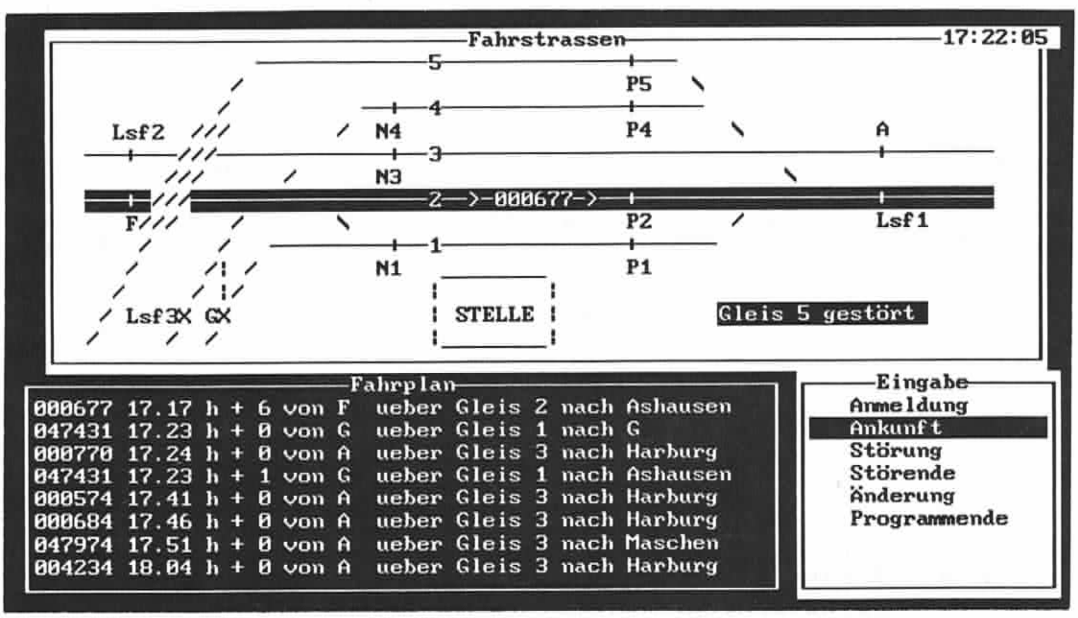

Figure 6. Screen of the program 'at work'

(d) no train from Hamburg (only Intercity or express trains, i.e. high priority trains, are coming directly from Hamburg) is scheduled to pass through Stelle in the current or following minute *and*
(e) running into the track heading south (at the back of Lsf1) will not slow down following trains (this condition will fire further rules)

For a more detailed description see Reference 1.

2.2. Prototype of computer-assisted decision support for selection of data analysis techniques in marketing

Meanwhile there exists a broad spectrum of tools for processing available data in the field 'data analysis/statistics'. Scientists with various research interests have developed a multitude of models and methods suited for their respective problems. Nevertheless, the conditions for correct application (sometimes even the existence) of such methods are not always known to a desirable extent.

Empirical studies on the usage of corresponding software for selected target groups (see e.g., References 57 and 58 for forecasting techniques, References 59–63 for the application of data analysis methods in market research) have revealed a deficit, particularly in the application of new and sophisticated techniques (cf. also the remarks in Section 1.2). Therefore, the DANEX prototype was developed.

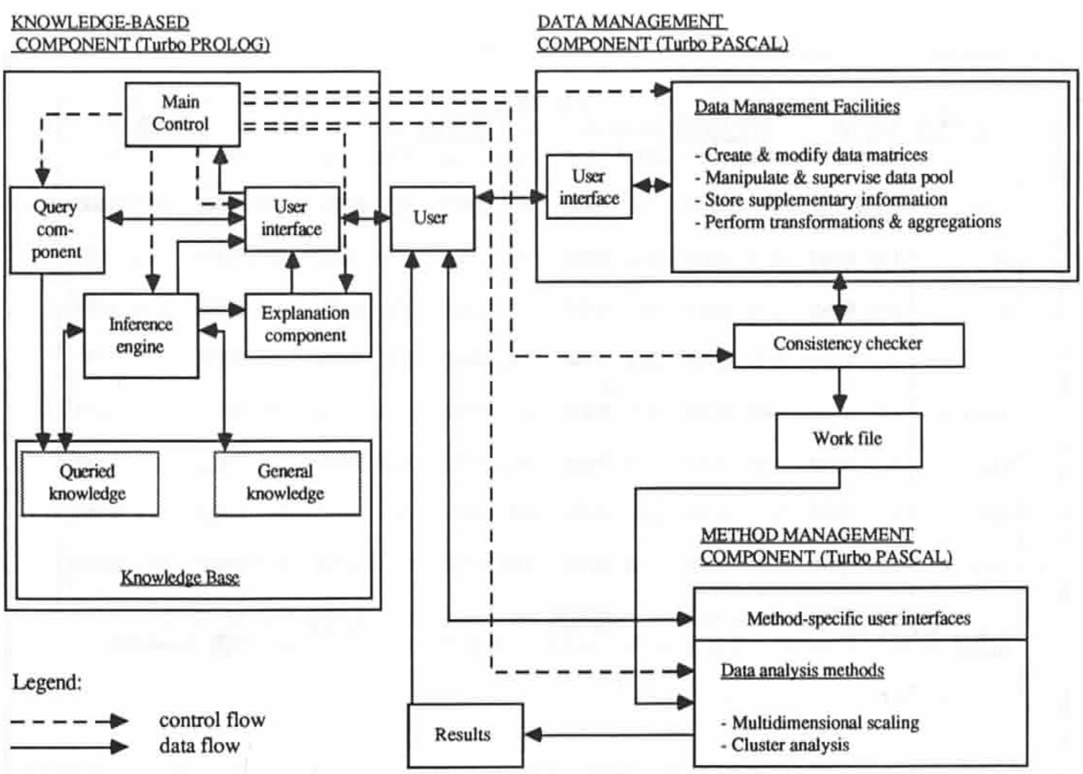

Figure 7. Architecture of DANEX

Recurring to the presentation of the processing of data analysis problems as a phase procedure as described in Figure 1, the present implementation mainly supports the transition $\overset{(2)}{\rightarrow}$ of Figure 1 since after the characterization of the structures of the underlying data, specific data analysis techniques are proposed. Here, techniques are preferred which are directly accessible by the system so that in a second step suitable solutions can be provided. Figure 7 shows the architecture of the current version.

DANEX is divided into three components, the knowledge base, implemented in Turbo-Prolog, and the data management as well as the method management components, both implemented in Turbo-Pascal. During a DANEX consultation session the user is queried about certain data characteristics (query component) which might also partly be set implicitly by the prototype if sufficient information was submitted. This knowledge is subsequently stored in the knowledge base and leads—in combination with general knowledge on usage of software—to propositions of data analysis methods which could be employed (inference engine). Here, the user can ask (explanation component) why a specific method was selected or not—and will in return receive a list of facts and rules which influenced the recommendations regarding a certain technique.

The data-management component can be invoked independently of the knowledge-based component. Here, typical functions for data management (transformation and aggregation methods, including/excluding of data values, processing of missing values, etc.) are available. The method management component can only be used after a consistency check of the given data. At the moment, especially methods for cluster analysis or multidimensional scaling are supported.

Matrix Edit

Name CIG_MA Matrix Association Data Similarity Mode 2

	full fla	adventur	good moo	activity	sympathy	relaxat	sociabil
Camel	411.0000	590.0000	314.0000	487.0000	314.0000	320.0000	269.0000
HB	330.0000	234.0000	466.0000	346.0000	337.0000	420.0000	393.0000
R6	215.0000	194.0000	344.0000	312.0000	326.0000	353.0000	370.0000
Ernte23	372.0000	247.0000	321.0000	290.0000	299.0000	324.0000	350.0000
Stuyve	370.0000	409.0000	395.0000	430.0000	370.0000	330.0000	369.0000
Kim	223.0000	194.0000	343.0000	318.0000	317.0000	310.0000	350.0000
West	427.0000	519.0000	345.0000	449.0000	286.0000	303.0000	314.0000
Lord	300.0000	227.0000	380.0000	323.0000	343.0000	367.0000	403.0000

↑↓→←, Home Move Cursor ↵ Save Matrix Esc End of Edit
Del Delete Element m Missing Value z Fill with Zero s Symmetrify

Figure 8. Example screen of data management part

In Figure 8 a typical screen of the data management component is shown. The data matrix to be analysed is a two-mode matrix (Mode: 2) of object similarities (Data: Similarity) reflecting associations (Matrix: Association) between elements of sets of different modes (elements of mode 1: brands of cigarettes; elements of mode 2: features for judging dimensions of product perceptions).

Before applying a two-mode cluster analysis (see e.g., Reference 7 for a comparison of cluster methods for two-mode data) the association matrix of Figure 8 has to be transformed into a two-mode dissimilarity matrix (a function already incorporated in the data management component). In the following, the two-mode version of MVAL (Missing Values Average Linkage, see, e.g. Reference 64) might be applied and would yield the result shown in Figure 9.

According to the data (only partially shown in Figure 8) the consumers interviewed see two clearly separated clusters. One cluster comprising Lord, HB, R6, Kim, Ernte and Krone is associated with 'sociability', 'good mood', 'relaxation' and—less strongly—with 'harmony'. Camel and 'adventure' show the highest overall association. In this second cluster also Marlboro, West and Stuyvesant as well as the dimensions 'strongness', 'activity', 'self confidence', 'full flavour', 'enjoyment' and 'sympathy' are to be found. This cluster also contains a four-element subgroup consisting of Dunhill, 'exclusiveness', 'tradition' and Philip Morris. After inspection of this solution which gives a first impression with respect to the

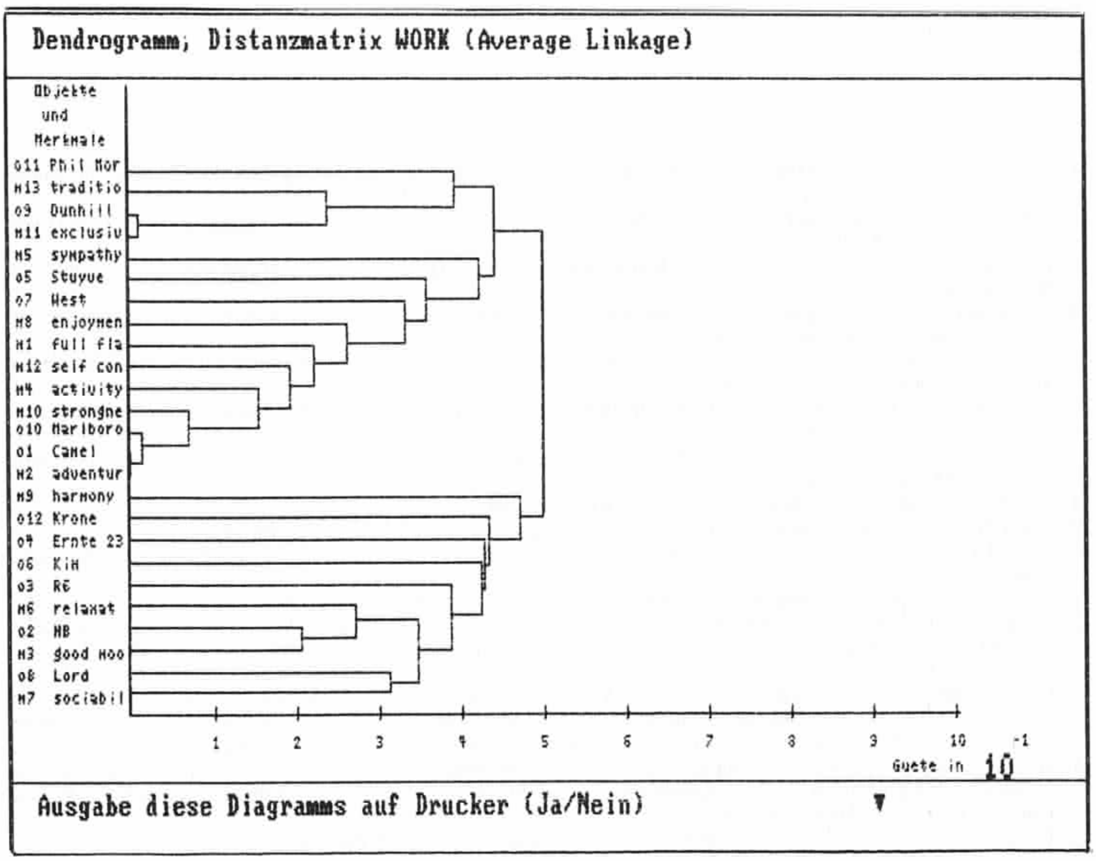

Figure 9. Dendrogram of a two-mode cluster analysis

problem 'brands and perceived product characteristics' control can, again, be transferred to the other system components, e.g., to perform different transformations or aggregations (possibly according to sex (males, females) or according to smoking habits (smokers, non-smokers)) or to select and apply further techniques. For a more detailed description see References 2 and 3.

3. OUTLOOK

The prototypes described have shown different situations where data analysis problems can be successfully tackled by computer-assisted decision support based on knowledge processing techniques. At the same time, the prototype descriptions should have revealed the magnitude of possible extensions and the challenges waiting for the researcher. Stand-alone developments of software for specific domains are often determined by fields of application. Here, the possibility of joining together individual routines to a more comprehensive system should be taken into consideration. Recurring to the presentation of the processing of data analysis problems as a phase procedure as described in Figure 1, the transitions $\overset{(1)}{\rightarrow}$ and $\overset{(3)}{\rightarrow}$ are more application-dependent than transition $\overset{(2)}{\rightarrow}$. Concerning the transition $\overset{(2)}{\rightarrow}$ of Figure 1 which represents the classical link of data analysis and decision support, joint development of decision support software is being advanced. Here, additional methods (and classes of methods) as described, for instance, in References 65 or 66 will be integrated.

REFERENCES

1. M. Schader 'PROLOG-basierte Entscheidungsunterstützung bei der Festlegung von Fahrstraßen in einem Bahnhof', *Computer Anwendungen Universität Karlsruhe*, **5**, 19–26 (1988).
2. I. Böckenholt, M. Both and W. Gaul, 'A knowledge-based system for supporting data analysis problems', *Decision Support Systems*, **5**, (1989).
3. I. Böckenholt, M. Both and W. Gaul, 'PROLOG-based decision support for data analysis in marketing', in Reference 37, pp. 19–34.
4. U. Tüshaus, 'Entscheidungsunterstützung bei der Analyse qualitativer Daten', in R. Wille (ed.), *Klassifikation und Ordnung*, Indeks, 1989, pp. 295–300.
5. W. A. Woodward and A. C. Elliot 'Observations on microcomputer programs for statistical analysis', *Statistical Software Newsletter*, **9**, 52–60 (1983).
6. P. Arabie and J. D. Carroll, 'MAPCLUS: a mathematical programming approach to fitting the ADCLUS model', *Psychometrika*, **45**, 211–235 (1980).
7. M. Both and W. Gaul, 'Ein Vergleich zweidimensionaler Clusteranalyseverfahren', *Methods of Operations Research*, **57**, 593–605 (1986).
8. *IMSL Library Reference Manual*, Vol 5. 1–3, Edition 8, 1981.
9. *NAG Fortran Library Manual (Mark 8)*, Numerical Algorithms Group Ltd., Oxford, 1983.
10. W. J. Dixon and M. B. Brown, (eds) *BMDP Statistical Software 1983*, University of California Press, Berkeley, 1983.
11. *SAS User's Guide: Basics and Statistics*, SAS Institute, Cary, North Carolina, 1982.
12. N. H. Nie, C. H. Hull, J. G. Jenkins, K. Steinbrenner and D. H. Bent, *SPSS Statistical Package for the Social Sciences*, McGraw-Hill, 1975.
13. *METHAPLAN Methodenbankablaufsystem für Planung und Analyse*, Siemens, München, 1976.
14. S. Schneider, P. Schwab and W. Renninger, 'Wesen, Vergleich und Stand von Software zur Produktion von Systemen der computergestützten Unternehmensplanung, Arbeitsbericht des Institut für Mathematische Maschinen und Datenverarbeitung (Informatik)', **16**, (5), Universität Erlangen Nürnberg, 1983.
15. H. Siegmann, 'Verleich und Stand von zehn ausgewählten kleinrechnerorientierten Planungssprachen, Arbeitsberichte des Instituts für Mathematische Maschinen und Datenverarbeitung (Informatik)', **18**, (1), Universität Erlangen Nürnberg, 1985.
16. L. Johnson and E. T. Keravnou, *Expert Systems Technology*, Abacus Press, 1986.
17. G. J. Hahn, 'More intelligent statistical software and statistical expert systems: future directions', *The American Statistician*, **39** (1), 1–16 (1988). (Including Comments and Reply).

18. F. Hayes-Roth, D. A. Waterman, and D. B. Lenat, (eds), *Building Expert Systems*, Addison-Wesley, 1983.
19. P. Harmon and D. King, *Expert Systems—Artificial Intelligence in Business*, Wiley, 1985.
20. W. A. Gale, 'Statistical applications of artificial intelligence and knowledge engineering', *Knowledge Engineering Review*, **1**, (1988).
21. E. Lehmann, *Expertensysteme: Überblicke über den aktuellen Wissensstand (1983)*, Siemens AG, 1984.
22. L. F. Pau, (ed.), *Artificial Intelligence in Economics and Management*, 2nd edn, North-Holland, 1987.
23. N. Victor, 'Computational statistics: tools or science?', *Statistical Software Newsletter*, **10**, 105–116 (1984).
24. J. A. Nelder, 'Intelligent programs, the next stage in statistical computing', in J. R. Barra, F. Brodeau, G. Ronner and B. Van Cutsem, (eds), *Recent Developments in Statistics*, North-Holland, 1977, pp. 79–86.
25. J. M. Chambers 'Some thoughts on expert software', *Proceedings of the 13th Symposium on the Interface of Computer Science and Statistics*, Springer, 1981, pp. 36–40.
26. A. M. R. Smith, L. S. Lee and D. J. Hand, 'Interactive user-friendly interfaces to statistical packages', *The Computer Journal*, **26**, 199–204 (1983).
27. F. Carlson and I. Heuch, 'EXPRESS—an expert system utilizing standard statistical packages', in *COMPSTAT*, 1986, pp. 265–270.
28. A. Prat, M. Marti and J. M. Catot, 'Incorporating expertise in time series modelling: the STATXPS system', *Statistical Software Newsletter*, **11**, 55–62 (1985).
29. D. F. Wolstenholme and J. A. Nelder, 'A front end for GLIM', in R. Haux (ed.), *Expert Systems in Statistics*, Fischer, 1986, pp. 155–177.
30. C. Berzuini, G. Ross and C. Larizza, 'Developing intelligent software for non-linear model fitting as an expert system', in *COMPSTAT*, 1986, pp. 259–264.
31. P. Naeve and J. Steinecker, 'SETUP Statistisches Expertensystem mit TWAICE und P-STAT', *Research Paper 172/1987*, University of Bielefeld, 1987.
32. F. Bodendorf and U. Osiander, 'Eine Methodenbankhülle um das Programmpaket SPSS/PC + für Gelegenheitsanwender', *Angewandte Informatik*, **1/88**, 3–8 (1988).
33. D. Pregibon and W. A. Gale, 'REX: an expert system for regression analysis', in *COMPSTAT*, 1984, pp. 242–248.
34. E. Dambroise and P. Masotte, 'MUSE: an expert system in statistics', in *COMPSTAT*, 1986, pp. 271–276
35. J. Jida and J. Lemaire, 'Expert systems and data analysis package management', in *COMPSTAT*, 1986, pp. 251–258.
36. P. L. Darius, 'Building expert systems with the help of existing statistical software: an example', in *COMPSTAT*, 1986, pp. 277–282.
37. W. Gaul and M. Schader, (eds), *Data, Expert Knowledge and Decisions: An Interdisciplinary Approach with Emphasis on Marketing Applications*, Springer, 1988.
38. P. Arabie and J. Daws, 'The interface among data analysis, marketing and representation of knowledge', in Reference 37, pp. 10–15.
39. O. Büring and M. Schader, 'KREDIT: KEE-based support for credit decisions in a mail-order-firm', in Reference 37, pp. 83–92.
40. M. Demann, M. Schader and U. Tüshaus, 'Developing a decision support system for personnel disposition in a garage—experiences using the KEE shell', in Reference 37, pp. 103–115.
41. M. A. H. Dempster and A. M. Ireland, 'MIDAS: an expert debt management advisory system', in Reference 37, pp. 116–127.
42. F. Esposito, 'Incorporating knowledge in decision support systems, in Reference 37, pp. 128–144.
43. W. Gaul and A. Schaer, 'A PROLOG-based PC-implementation for new product introduction', in Reference 37, pp. 42–53.
44. M. Gettler-Summa, H. Ralambondrainy and E. Diday, 'Data analysis and expert systems: generating rules from data', in Reference 37, pp. 161–173.
45. C. Granger, 'Fuzzy reasoning for classification: an expert system approach', in Reference 37. pp. 174–184.
46. S. Trautmann, 'OPTRAD: a decision support system for portfolio management in stock and options markets', in Reference 37, pp. 185–203.
47. U. Tüshaus, 'Decision support for qualitative data—KEE shell linked to Fortran programs', in Reference 37, pp. 204–212.
48. P. Dehio and A. Kieser, 'Die Gestaltung von Entscheidungsunterstützungssystemen', *Angewandte Informatik*, **25**, 371–382 (1983).
49. D. J. Hand, 'Statistical expert systems: design', *The Statistician*, **37**, 351–369 (1984).
50. D. J. Hand, 'Statistical expert systems: necessary attributes', *Journal of Applied Statistics*, **12**, 19–27 (1985).
51. D. J. Hand, 'A statistical knowledge enhancement system', *Journal of the Royal Statistical Society*, **150**, 334–345 (1987).
52. R. A. Thisted, 'Representing statistical knowledge for expert data analysis systems', in W. A. Gale, (ed.), *Artificial Intelligence and Statistics*, Addison-Wesley, 1986, pp. 267–284.
53. D. J. Hand, 'Patterns in statistical strategy', in W. A. Gale, (ed.), *Artificial Intelligence and Statistics*, Addison-Wesley, 1986, pp. 355–387.
54. W. Gaul, I. Böckenholt, M. Both, E. Espejo, Ch. Homburg and A. Schaer, 'Computer-based decision support in marketing', in *HECTOR (HEterogeneous Computers TOgetheR), Vol. I: New Ways in Education and Research*, Springer 1988, pp. 360–396.

55. M. Schader and U. Tüshaus, 'Subgradient methods for analyzing qualitative data', in Reference 67, pp. 397–403.
56. M. Schader and U. Tüshaus, 'Analysis of qualitative data—a heuristic for finding a complete preorder', in H. Bock, (ed.), *Classification and Related Methods of Data Analysis*, North-Holland, 1988, pp. 341–346.
57. J. T. Mentzler and J. E. Cox, Jr. 'Familiarity, application and performance of sales forecasting techniques', *Journal of Forecasting*, **3**, 27–36 (1984).
58. J. R. Sparkes and A. K. McHugh, 'Awareness and use of forecasting techniques in British Industry', *Journal of Forecasting*, **3**, 37–42 (1984).
59. B. A. Greenberg, J. L. Goldstucker and D. N. Bellenger, 'What techniques are used by marketing researchers in business?', *Journal of Marketing*, **41**, 62–68 (1977).
60. W. Gaul 'Zum Einsatz von Datenanalysetechniken in der Marktforschung', in O. Opitz and B. Rauhut, (eds), *Ökonomie und Mathematik*, Springer, 1987, pp. 396–408.
61. W. Gaul, F. Förster and K. Schiller, 'Typologisierung deutscher Marktforschungsinstitute: Ergebnisse einer empirischen Studie', *Marketing-ZfP*, Heft 3, 163–172 (1986).
62. W. Gaul, F. Förster and K. Schiller, 'Empirische Ergebnisse zur Verbreitung und Nutzung von Statistiksoftware in der Marktforschung', in *Tagungsband zur 3. Konferenz über die wissenschaftliche Anwendung von Statistik-Software*, Fischer, 1986, pp. 323–332.
63. W. Gaul and Ch. Homburg, 'The use of data analysis techniques by German market research agencies: a causal analysis', *Journal of Business Research*, **17**, 67–79 (1988).
64. E. Espejo and W. Gaul 'Two-mode hierarchical clustering as an instrument for marketing research', in Reference 67, pp. 121–128.
65. W. Gaul and M. Schader, 'Clusterwise aggregation of relations', *Applied Stochastic Models and Data Analysis*, **4**, 273–282 (1988).
66. S. Nishisato and W. Gaul, 'Marketing Data analysis by dual scaling', *International Journal of Research in Marketing*, **5**, 151–170 (1988).
67. W. Gaul and M. Schader, (eds), *Classification as a Tool of Research, Proceedings of the 9th Annual Meeting of the German Classification Society*, North-Holland, 1986.