A Knowledge-Based System for Supporting Data Analysis Problems *

I. BÖCKENHOLT, M. BOTH and W. GAUL
University of Karlsruhe, D-7500 Karlsruhe, FRG

Scientists from different research areas have developed a variety of models and methods to support data analysis problems in their specific fields of interest. However, the preconditions for a proper usage of the corresponding software which is often provided in the shape of software packages or individual programs may cause a severe problem. These preconditions preponderantly demand knowledge as well about theoretical and software specific aspects of algorithms used as about essentials of the area of application. The prototype to be presented in this paper aims at extending the capabilities of conventional software approaches by incorporating knowledge concerning the suitability of certain data analysis methods. The knowledge-based part is realized in PROLOG. The application area is market research. The prototype already comprises several data analysis procedures, especially from multidimensional scaling and cluster analysis, and data management facilities which can be invoked by the user based on the recommendations given by the system.

Keywords: Knowledge-Based Decision Support, Data Analysis, Combinations of Data Management and Method Management.

Ingo Böckenholt is an Assistant Professor of Marketing at the Faculty of Economics, University of Karlsruhe, Germany. He received his Ph.D. at this University and also holds a Master Degree in Industrial Engineering which includes topics such as Economics, Operations Research and Computer Science. His latest articles have been on the subject of new product planning, measurement of advertisement responses and choice modelling.

Martin Both is an Assistant Professor of Marketing at the Faculty of Economics, University of Karlsruhe. He received his Ph.D. at this University and also holds a Master Degree in Industrial Engineering with emphasis on Marketing, Computer Science and Operations Research. He is the co-author (with W. Gaul) of a book on "Computergestütztes Marketing" (Springer, in press). His research interests include computer-based support of managerial activities and integration of different concepts for decision support, especially of Decision Support Systems and of Expert Systems.

* Research has been supported by the Deutsche Forschungsgemeinschaft.

North-Holland
Decision Support Systems 5 (1989) 345–354

1. Motivation

Data analysts are faced with an abundance of models and methods which can be potentially applied to a given data analysis problem. Quite a number of those methods have been made available by means of software packages (see e.g. SAS ([28]) or SPSS ([30])) but nonetheless there is empirical evidence that some of the potential users of data analysis techniques don't apply the existing tools to an extent desirable (see e.g. [15], [16], [17], [18] for results on a survey conducted among market research institutes concerning their usage of data analysis techniques).

Whereas some straightforward standard data analysis methods can be applied routinely even by a non-expert user, more advanced techniques presuppose specific knowledge concerning the preconditions that have to be fulfilled in order to ensure their proper application. To say it in other words, the conventional procedurally oriented software concept which is appropriate for implementing data analysis techniques incorporates algorithmic expertise; however, knowledge of a more qualitative kind and experiences being helpful in the process of data analysis cannot easily be made available by means of conventional programming approaches. An appropriate "knowledge base" would have to comprise e.g. knowledge about dependencies between data structures to be analyzed and specific data analysis procedures as well as knowledge about relationships between

Wolfgang Gaul is a Professor of Business Administration and Economics at the Faculty of Economics, University of Karlsruhe. He received his Ph.D. and his "Habilitation" in Mathematics at the University of Bonn. He has co-edited (with M. Schader) two books, "Classification as a Tool of Research" (North Holland, 1986), and "Data, Expert Knowledge and Decisions" (Springer, 1988) and co-authored (with M. Both) a book on "Computergestütztes Marketing" (Springer, in press.) His research interests include mathematical modeling in economics, especially in marketing, operations research, and data analysis and statistics.

users' objectives and the suitability of certain techniques or clues for interpreting results obtained by (different) data analysis methods.

In case a user lacks essential parts of this knowledge, there is the danger of misuse of the appertaining algorithms.

Thus, in order to support an (inexperienced) user of data analysis techniques, the purely algorithmic expertise has to be enriched by symbolic and qualitative knowledge alluded to previously.

Of course, tools for handling situations in which decision support on the basis of expert knowledge about the problem area is needed would be of great importance, and first efforts from the area of artificial intelligence have gained popularity, recently (see e.g. [22] and [23] for an introduction into the field of expert systems).

This paper will give a short survey on some developments concerning expert system/decision support system approaches for data analysis problems in the next section. Then, a prototype of our own, intended to support positioning and segmentation, two important objectives within marketing planning problems which often need sophisticated data analysis assistance, will be illustrated. Part of the architecture of the prototype will be outlined, and some functions and interconnections of its components will be described. Aspects concerning the implementation of the expert system shell, which is in control of the knowledge-based component of the prototype, will be discussed and hints how the user is supported will be given.

Conclusions and an outlook on further research terminate the paper.

2. Knowledge-Based Approaches for Handling Data Analysis Problems

Since ideas of artificial intelligence have entered the field of data analysis and statistics (see e.g. [24]) the number of theoretical papers as well as prototype developments has increased rapidly so that a complete survey will not be given here (see e.g. the surveys [10] and [13]). Instead, only some directive contributions characterizing the current main streams of research and some approaches which have proved to be helpful in developing the prototype to be presented in the next section will be shortly reviewed. Knowledge-based approaches for supporting data analysis problems which, at least, have led to prototypical implementations can be classified into two categories.

On the one hand, there are front-end developments to complex statistical software systems which serve as user interfaces. Their main objective is to support users in operating the software, e.g. by generating input parameter files on the basis of users' requirements or by preparing or transforming the output in such a way that it can be used for a proper presentation of solutions or as input for additional analyses.

Representatives of this kind of knowledge-based data analysis support systems (closely connected with the respective conventional software)are e.g. the front-end to GLIM ([32]), the front-end to MULTIVARIANCE ([29]), and the front-end to BMDP ([5]). Such front-ends need not necessarily be implemented by means of specific artificial intelligence tools, e.g. the latter two of the just mentioned front-ends were written in FORTRAN.

On the other hand, there are prototypes which either aim at giving advice with respect to certain stages of data collection (see e.g. [8] where a system is proposed for an automatic validation check in the process of data gathering) or which focus on specific (classes of) data analysis techniques (see e.g. [11], [12], [27] for the regression expert system REX, in which a strategy for performing regression analysis is encoded). Of course, it could be argued that support should be made available rather for more complex statistical models than e.g. for regression (see e.g. [7] for a prototype which aims at supporting principal component analysis and correspondence analysis) but, at the moment, such attempts have not yet led to fully implemented systems. This state is, among other things, due to shortcomings of the knowledge bases and the insufficient integration of expert system components and conventional software components. More theoretically oriented contributions concerning the role of knowledge-based systems for data analysis purposes have dealt with e.g. design aspects (see e.g. [20]), necessary attributes (see e.g. [21]), and with methods for knowledge representation (see e.g. [26] and [31]).

At the end of this section some features a knowledge-based system in the area of data analysis should exhibit are outlined. General characteristics, such as capabilities explaining the deductive processes used or requirements of rapid modifications of the knowledge bases used, which have already been extensively discussed in the literature (see e.g. [23]), will not be repeated, here. Instead, emphasis is laid on specific requirements caused by the peculiarities of the data analysis problems for which support is to be provided.

First, rather than in other application domains one is confronted with two kinds of information. One kind is provided by the users (their objectives and different background knowledge about how the data were collected and should be analyzed). The other kind is made up by the data themselves (how they could be examined automatically, how they should be aggregated or transformed in order to assure that the preconditions required for the proper usage of specific data analysis techniques are fulfilled).

Next, the exploratory character of data analysis tasks necessitates that frequently not only a single method but a whole set of techniques which complement one another should be proposed by the system.

Furthermore, data analysts often pursue multiple objectives so that a system should be capable of providing several answers in order to support the whole range of aspects which are wished to be studied.

Additionally, specific requirements concerning the explanation facilities of the system arise. Besides the common explanation component, which e.g. provides help functions and justifies proposals or solutions given by the system, explanations why a certain data analysis technique was not selected are certainly of interest. The implementation of such a kind of support will help the user to learn more about the underlying relationships incorporated in the system and increase the confidence in the system.

3. The DANEX (*D*ata *AN*alysis *EX*pert) Prototype

In this section, a prototype development of a knowledge-based system, disposing some of the features mentioned previously, will be presented. Some objectives pursued with this prototype can be summarized as follows.

First, the prototype aims at supporting the analysis of such data structures which are especially important in the context of positioning and segmentation problems. These data structures are usually aggregated in the shape of matrices reflecting e.g. consumer preferences, perceived similarities or dissimilarities.

Secondly, with respect to the variety of data analysis techniques available to analyze relationships described by such matrices, for the implementation of the current version of the prototype focus is laid upon two classes of techniques, namely multidimensional scaling and cluster analysis.

Thirdly, within the mentioned classes emphasis is given to advanced techniques most of which have not yet been integrated in statistical software systems (see e.g. [1] for sophisticated probabilistic multidimensional scaling methods and [3], [9] for advanced non-hierarchical respectively hierarchical cluster analysis techniques).

Fourthly, in order to make such advanced techniques available to a nonexpert user, knowledge required for their proper usage is to be provided in addition to the purely algorithmic expertise of the procedures. By means of this knowledge, relevant features of the data to be analyzed should automatically be related to the requirements of the data analysis procedures, and adequate procedures should be identified by the system.

Next, besides such a knowledge-based component, the system should comprise facilities for performing data management functions as well as executable versions of the mentioned data analysis procedures.

Finally, in view of the wide dissemination of microcomputers, a further objective is that the system should run on a computer of this type. This requirement will inevitably lead to restrictions concerning the data analysis procedures to be incorporated, especially with respect to their execution time.

The objectives just formulated resulted in the development of a prototype called DANEX (Data ANalysis EXpert) the architecture of which is depicted in fig. 1.

The current version of DANEX consists of three components which are not only functionally separated but also differ with respect to their implementation. The *knowledge-based component* – including the main control which provides access to all three components – is embedded in an

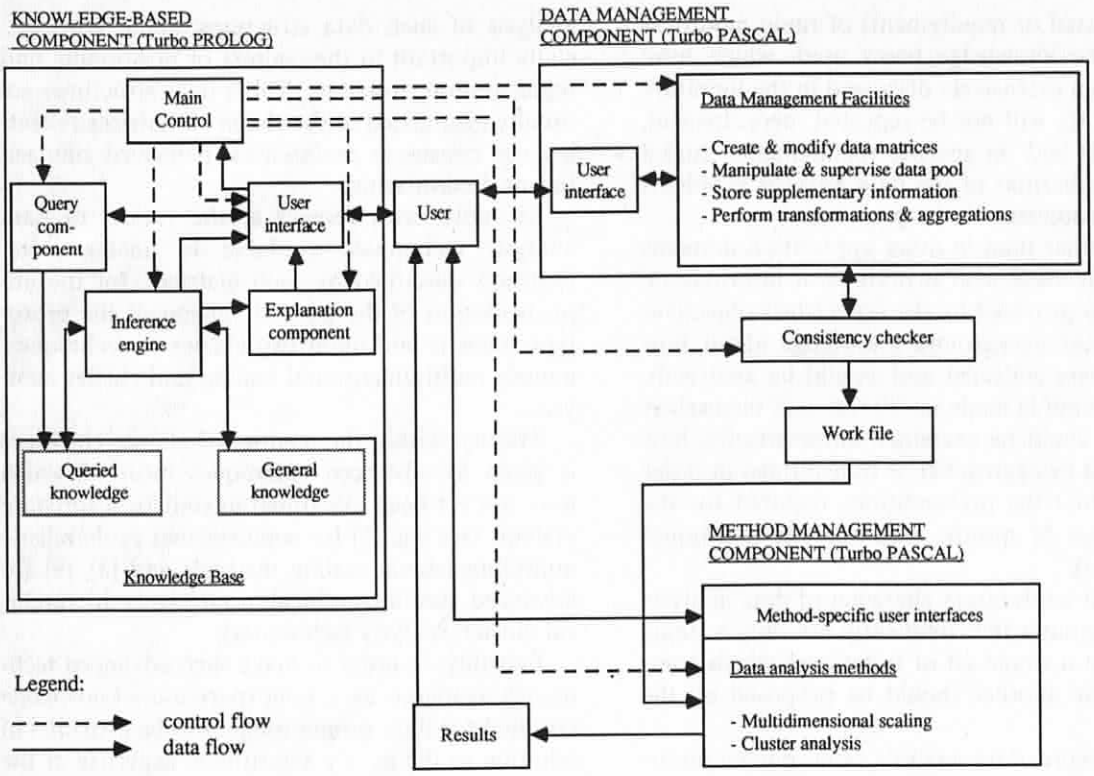


Fig. 1. Architecture of the Current DANEX Version.

expert system shell written in TURBO-PROLOG [1] (see e.g. [4]). PROLOG has recently gained popularity as a language for implementing knowledge-based systems due to its nonprocedural character. This popularity is reflected by an increasing quantity of various dialects which differ in e.g. the number of built-in predicates, the types of development environments, the possibilities to handle data structures, the input/output facilities, and the arrangements of interfaces to other programs (see e.g. [6] for a comparison of three PROLOG compilers).

Although TURBO-PROLOG disposes of less built-in predicates than e.g. ARITY-PROLOG, its advantages with respect to execution speed and user interfaces are considerable. Therefore, TURBO-PROLOG was preferred as implementation language for the underlying prototype.

[1] A preceding version was written in ARITY-PROLOG and developed together with H. KLEINE BÜNING and co-workers from the Institut für Angewandte Informatik und Formale Beschreibungsverfahren, Fakultät für Wirtschaftswissenschaften, Universität Karlsruhe (TH).

When a user consults DANEX, a characterization of the data structures to be analyzed has to be given (Knowledge about relationships between characteristics such as matrix type, modality, data type, scale type and adequate data analysis techniques is represented in the knowledge base; explanations concerning these characteristics are, of course, available in an explanation component.). Needed specifications are asked by a query component and are stored in the temporary section of the knowledge base, while internal knowledge about relationships between characteristics of data and suitable data analysis methods is represented in the permanent section of the knowledge base.

After the required specifications have been completed, the actual information provided by the user as well as the general knowledge which is independent from a specific data analysis problem will be processed by the inference engine. The search for suitable data analysis techniques is performed in a goal driven way according to the resolution mechanisms inherent in PROLOG.

After termination of the inference process, the explanation component comprising the following features can be invoked. As a preliminary result, the user is presented a set of appropriate data analysis techniques recommended for the analysis of the data to be studied. The way this result was achieved can be made transparent by means of several explanation facilities. First, a list of all questions and answers can be requested, distinguishing between answers explicitly given by the user and specifications derived implicitly. Secondly, a verbalized description of all successful rules which led to the proposal of a specific data analysis technique can be asked for (analogously, a description of all falsified rules can be presented, too). Thirdly, a "why-not" facility may be invoked which is able to explain why a certain data analysis technique, the user might have in mind, was not proposed by DANEX.

By means of the main control mechanism, included in the knowledge-based component, the user can invoke a data management component and a method management component both written in TURBO-PASCAL.

The *data management facilities* which have a user interface of their own have been selected with respect to the specific data structures and data to be analyzed. As depicted in the corresponding part of fig. 1, various tools which allow for comfortable editing and different types of aggregations and transformations of the data under study are provided. To start with the first task, a matrix editing environment can be invoked (see fig. 3a, b in the next section). There are options for e.g. deleting elements of the data matrix, indicating missing values, filling up an incomplete matrix with zeros, and symmetrifying a matrix on the basis of the upper triangle matrix. Matrix name, matrix type, data type and number of modes are automatically shown together with the actually interesting part of the data matrix (browsing around in the whole matrix is possible in conventional way). Additionally, objects and/or subjects the relations of which are comprised in the matrix can be labelled, and a short comment characterizing the underlying data set can be given.

Besides these editing features, the user can choose between different possibilities to aggregate the data under study, e.g. compute (dis)similarities, distances or correlation measures (see fig. 3a, b in the next section which give an example how a dissimilarity matrix is yielded from a two-mode association matrix), and between various options to perform transformations of given data matrices, e.g. change similarities to dissimilarities or vice versa.

The user can also access the data management facilities directly and independently from a previous consultation of the knowledge-based component. Nonetheless, several checks concerning consistency and plausibility are conducted automatically when a data matrix is being edited, aggregated or transformed.

The *method management component* is activated when the user decides to carry out one of the data analysis procedures proposed by the knowledge-based component. In such a case, a consistency check whether the data stored fulfil specific requirements is performed, automatically. If the result of this check is positive, the data are transferred into a workfile which is made accessible to the data analysis procedure. When the data analysis procedure is executed, control is being transferred to a corresponding subroutine because the user communicates with the various data analysis techniques by means of method-specific interfaces.

After the computations within a special method have been completed the results are displayed graphically or in tabular form and control is, again, taken over by the knowledge-based component of the system.

Of course, different viewpoints for selecting methods to be incorporated into the method management component can be stressed. In this prototype version, we started with cluster analysis and multidimensional scaling techniques (see e.g. [1], [3], [9] for references containing illustrative examples and/or mathematical background with respect to the methodology used) which can be described as own implementations of moderate size and which can be operated without recourse to other (commercial) software on a PC. Other research directions for which implementations are available comprise e.g. correspondence analysis and dual scaling ([25]), clusterwise aggregation of relations ([19]), and – beyond data analysis support – special applications, e.g. in the area of new product introduction ([2]).

4. Usage of DANEX

In the beginning of a consultation by DANEX – among other things – the data structures to be

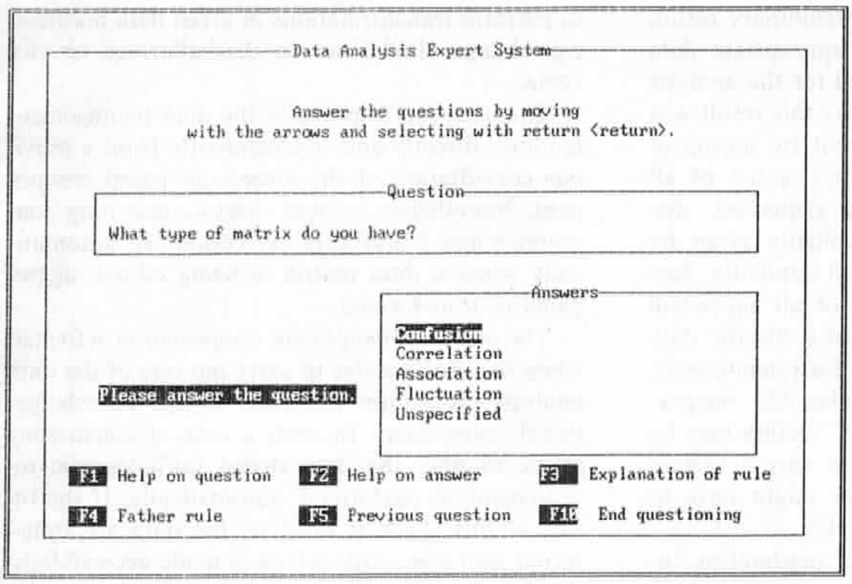


Fig. 2. Example Screen of a DANEX Consultation.

analyzed have to be characterized by the user. In order to support this task, a menu-driven dialog has been implemented in the course of which the user can call several help functions concerning requests for e.g. explanations of the terms used in the questions, explanations of the respective answer categories or verbalized descriptions of the rule(s) which caused specific questions to be formulated. An example screen of the dialog in which the type (if known) of the matrix to be analyzed has to be indicated is depicted in fig. 2.

If the user is able to specify the matrix type, some subsequent questions can implicitly be answered by DANEX and, thus, the process of characterizing the data structures can be abbreviated.

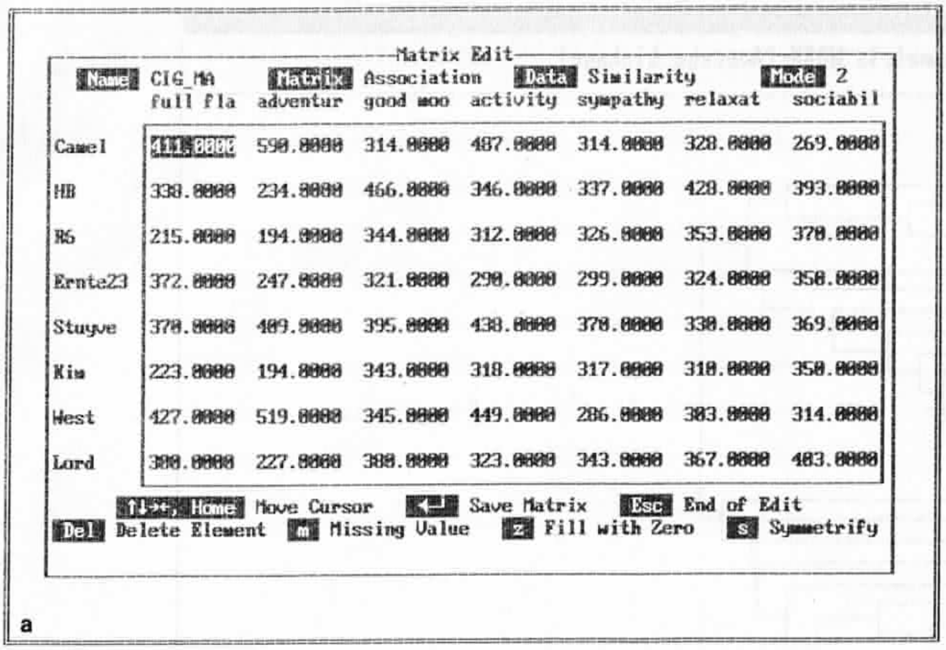
In some cases, data have to be transformed to meet the requirements of certain data analysis techniques which have been suggested by DANEX. The following example may help to clarify how DANEX proceeds: Suppose, the matrix type is known as two-mode association matrix. Part of this data matrix which is to be analyzed in the following is depicted in fig. 3a which also illustrates some of the facilities the user is offered for the purpose of data editing.

In this aggregated two-mode association matrix, the rows describe cigarette brands (row objects) while the columns reflect features (column objects) used for characterizing the brands. The single matrix elements contain the aggregated values of the features – measured on a seven point scale – which were assigned to the different brands by the male subgroup of the participants of an experiment (see e.g. [3] for a more detailed description of the experimental setting used for data collection). For purposes of evaluating the relationships within this association matrix, the data can be transformed into dissimilarities by means of procedures provided by the data management component. Fig. 3b shows the result of such a transformation.

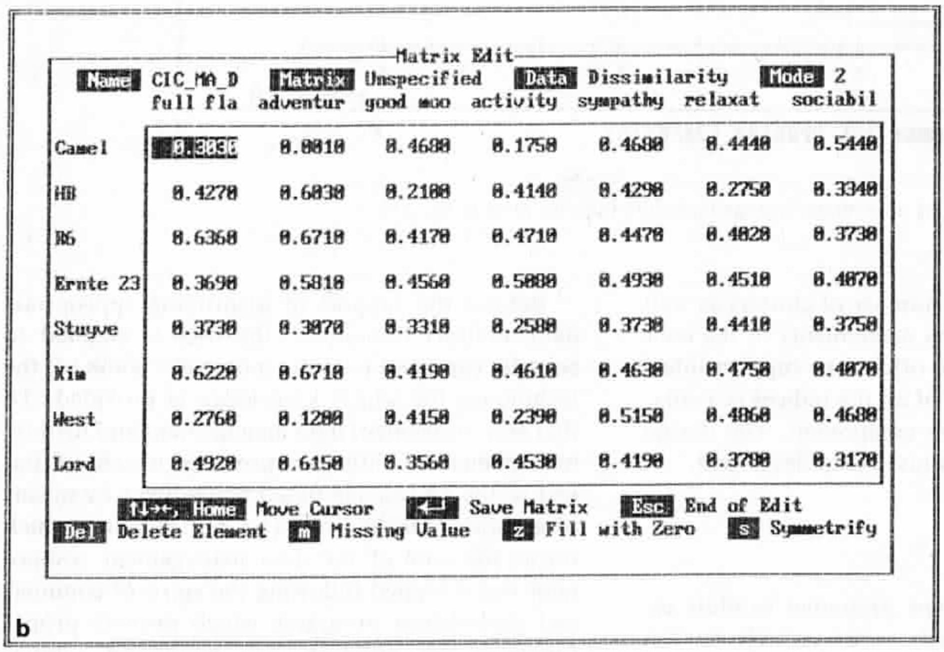
In order to analyze such a two-mode dissimilarity matrix, DANEX proposes different data analysis techniques, e.g. two-mode hierarchical clustering (for a two-mode non-hierarchical overlapping clustering a further matrix transformation to similarity data is needed) and probabilistic unfolding and/or probabilistic vector approaches.

In the following, just part of the results obtainable by two-mode hierarchical clustering will be illustrated. The method management component of DANEX will use the two-mode dissimilarity matrix partly described in fig. 3b to construct a best fitting so-called "grand matrix" (see e.g. [9]) the elements of which fulfil a so-called "ultrametric inequality" condition. From this ultrametric grand matrix a dendrogram as illustrated in fig. 4 can be drawn.

In this dendrogram the relationships between



	full fla	adventur	good moo	activity	sympathy	relaxat	sociabil
Camel	411.0000	590.0000	314.0000	487.0000	314.0000	328.0000	269.0000
HB	330.0000	234.0000	466.0000	346.0000	337.0000	428.0000	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	438.0000	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	300.0000	323.0000	343.0000	367.0000	403.0000



	full fla	adventur	good moo	activity	sympathy	relaxat	sociabil
Camel	0.3030	0.0010	0.4600	0.1750	0.4600	0.4440	0.5440
HB	0.4270	0.6030	0.2100	0.4140	0.4290	0.2750	0.3340
R6	0.6360	0.6710	0.4170	0.4710	0.4470	0.4020	0.3730
Ernte 23	0.3690	0.5810	0.4560	0.5000	0.4930	0.4510	0.4070
Stuyve	0.3730	0.3070	0.3310	0.2580	0.3730	0.4410	0.3750
Kim	0.6220	0.6710	0.4190	0.4610	0.4630	0.4750	0.4070
West	0.2760	0.1200	0.4150	0.2390	0.5150	0.4860	0.4600
Lord	0.4920	0.6150	0.3560	0.4530	0.4190	0.3700	0.3170

Fig. 3. (a) Part of a Two-Mode Association Matrix Reflecting Relationships Between Cigarette Brands (Row Objects) and Associated Features (Column Objects), (b) Part of a Two-Mode Dissimilarity Matrix Generated from the Matrix Depicted in (a).

the row objects (12 cigarette brands numbered from O1 to O12) and the column objects (13 features numbered from M1 to M13) are visualized and can be interpreted meaningfully (see e.g. [3] for an interpretation from the marketing point of view and for a comparison with results obtained by application of an additional non-hierarchical overlapping two-mode clustering version developed at our institute).

Besides the dendrogram of fig. 4, several indices explaining homogeneity within clusters respectively heterogeneity between clusters, which

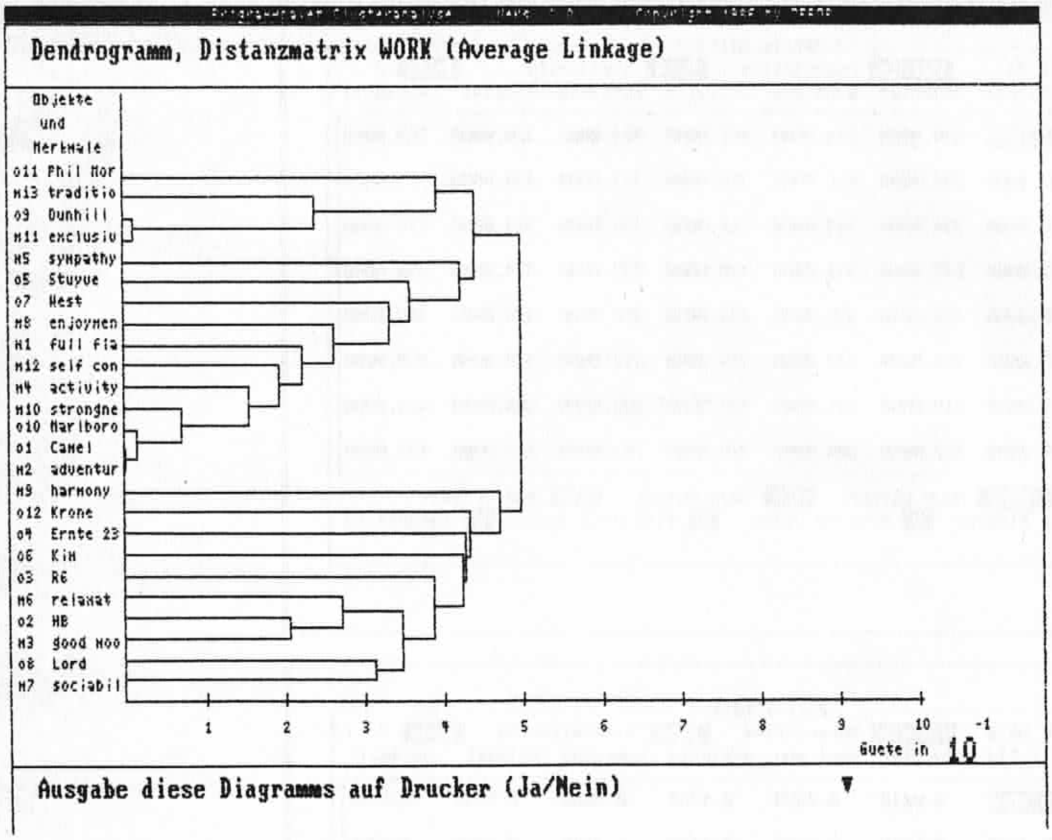


Fig. 4. Dendrogram Obtained by an Average Linkage Procedure from the Data in fig. 3(b).

give hints on the proper number of clusters as well as fit indices which allow assessments of the configurations obtained, are offered to support interpretation. A description of all the indices is available from the explanation component. The design of a formatted output is just being developed.

5. Conclusions

The DANEX prototype presented exhibits essential features which are called for a knowledge-based system designed for supporting data analysis purposes. At the current implementation stage, the analysis of data structures important in the context of positioning and segmentation problems is supported by advanced techniques from cluster analysis and multidimensional scaling. The inferences drawn by the system can be made transparent to the user by means of explanation facilities orientated by the characteristics of the consultation task.

Besides the support of identifying appropriate data analysis techniques, the user is enabled to actually carry out (–at the moment – some of) the techniques for which knowledge is provided. To that end, convenient data management and method management facilities are provided which are linked to the knowledge-based component by means of a control mechanism. The matrix editor which forms the core of the data management component was designed following the spirit of commercial spreadsheet programs which provide proper functions for creating and modifying tabular structures.

Future extensions of the prototype will take into account requirements which are needed in further steps of the data analysis process, such as proposals for appropriate data collection and hints for substantial interpretation of the results yielded by specific data analysis procedures. To that end, subject-matter knowledge will have to be incorporated to a higher degree.

the knowledge base and may entail modifications of the functions the expert system shell currently provides. Moreover, a mechanism for storing and retrieving results obtained in previous consultation sessions will have to be installed to allow for an immediate comparison of results obtained by different alternatively applicable techniques.

Given the performance of the present version of the prototype just described and the potential that might be realized by means of the extensions just mentioned, DANEX can be expected to become a helpful instrument for guiding an inexperienced user during important steps of the process of data collection, preparation, analysis and interpretation.

References

[1] I. Böckenholt, W. Gaul, Analysis of Choice Behaviour via Probabilistic Ideal Point and Vector Models, Applied Stochastic Models and Data Analysis 2 (1986) 209–226.

[2] I. Böckenholt, W. Gaul, New Product Introduction based on Pre-Test-Market Data, in: Proceedings of EMAC/ESOMAR Symposium on Micro and Macro Modelling: Research on Prices, Consumer Behaviour and Forecasting, (ESOMAR (1987) 77–96; ISBN: 92-831-1127-3).

[3] M. Both, W. Gaul, Ein Vergleich zweimodaler Clusteranalyseverfahren, Methods of Operations Research 57 (1987) 593–605.

[4] Borland, Turbo Prolog (Borland International Inc., Scotts Valley, USA 1986; ISBN: 0-87524-150-6).

[5] F. Carlsen, I. Heuch, EXPRESS – An Expert System Utilizing Standard Statistical Packages, in: F. De Antoni, N. Lauro, A. Rizzi, Eds., COMPSTAT 86 (Physica, Heidelberg, Wien (1986) 265–270; ISBN: 3-7908-0355-3).

[6] M. Covington, A. Vellino, Prolog Arrives, Tech Journal, November 1986.

[7] E. Dambroise, P. Masotte, MUSE: An Expert System in Statistics, in: F. De Antoni, N. Lauro, A. Rizzi, Eds., COMPSTAT 86 (Physica, Heidelberg, Wien (1986) 271–276; ISBN: 3-7908-0355-3).

[8] J.M. Dickson, M. Talbot, Statistical Data Validation and Expert Systems, in: F. De Antoni, N. Lauro, A. Rizzi, Eds., COMPSTAT 86 (Physica, Heidelberg, Wien, (1986) 283–288; ISBN: 3-7908-0355-3).

[9] E. Espejo, W. Gaul, Two-Mode Hierarchical Clustering as an Instrument for Marketing Research, in: W. Gaul, M. Schader, Eds., Classification as a Tool of Research (North Holland, Amsterdam (1986) 121–128; ISBN: 0-444-87980-3).

[10] W.A. Gale, Overview of Artificial Intelligence and Statistics, in: W.A. Gale, Ed., Artificial Intelligence and Statistics (Addison-Wesley, Reading, Mass. (1986) 1–16; ISBN: 0-201-11569-7).

[11] W.A. Gale, REX Review, in: W.A. Gale, Ed., Artificial Intelligence and Statistics (Addison-Wesley, Reading, Mass. (1986) 173–227; ISBN: 0-201-11569-7).

[12] W.A. Gale, D. Pregibon, An Expert System for Regression Analysis, in: Proceedings of Computer Science and Statistics, Annual Symposium on the Interface, (North Hollywood, Cal., 14, (1982) 110–117).

[13] W.A. Gale, D. Pregibon, Artificial Intelligence Research in Statistics, The AI Magazine 4 (1985) 72–75.

[14] W. Gaul, Probabilistic Choice Behaviour Models and their Combination with Additional Tools needed for Applications to Marketing, Communication & Cognition 20, No. 1 (1987) 77–92.

[15] Gaul, W., Zum Einsatz von Datenanalysetechniken in der Marktforschung, in: Opitz, D., Rauhut, B. Eds., Ökonomie und Mathematik, (Springer, Berlin, Heidelberg, (1987) 396–408; ISBN: 3-540-17819-8).

[16] W. Gaul, F. Förster, K. Schiller, Empirische Ergebnisse zur Verbreitung und Nutzung von Statistiksoftware in der Marktforschung, in: W. Lehmacher, A. Hörmann, Eds., Statistik-Software, 3. Konferenz über die wissenschaftliche Anwendung von Statistik-Software, (Gustav Fischer, Stuttgart, New York, (1986) 323–332; ISBN: 3-437-40170-X).

[17] W. Gaul, F. Förster, K. Schiller, Typologisierung deutscher Marktforschungsinstitute: Ergebnisse einer empirischen Studie, Marketing ZFP 8 (1986) 163–172.

[18] W. Gaul, C. Homburg, The Use of Data Analysis Techniques by German Market Research Agencies: A Causal Analysis, Journal of Business Research 17 91988) 67–79.

[19] W. Gaul, M. Schader, Clusterwise Aggregation of Relations, Applied Stochastic Models and Data Analysis 4 (1988) 273–282.

[20] D.J. Hand, Statistical Expert Systems: Design, The Statistician 33 (1984) 351–369.

[21] D.J. Hand, Statistical Expert Systems: Necessary Attributes, Journal of Applied Statistics 12 (1985) 19–27.

[22] P. Harmon, D. King, Expert Systems (Wiley, New York, 1985; ISBN: 0-471-81544-3).

[23] F. Hayes-Roth, D. Waterman, D. Lenat, Eds., Building Expert Systems (Addison-Wesley, Reading, Mass., 1983; ISBN: 0-201-10686-8).

[24] J.A. Nelder, Intelligent Programs, the Next Stage in Statistical Computing, in: J.R. Barra, F. Brodeau, G. Romier, B. Van Cutsem, Eds., Recent Developments in Statistics (North Holland, Amsterdam (1977) 79–86; ISBN: 0-7204-0751-6).

[25] S. Nishisato, W. Gaul, Marketing Data Analysis by Dual Scaling, International Journal of Research in Marketing 5 (1988) 151–170.

[26] R.W. Oldford, S.C. Peters, Implementation and Study of Statistical Strategy, in: W.A. Gale, Ed., Artificial Intelligence and Statistics (Addison-Wesley, Reading, Mass. (1986) 159–171; ISBN: 0-201-11569-7).

[27] D. Pregibon, W.A. Gale, REX: An Expert System for Regression Analysis, in: T. Havranek, Z. Sidak, M. Novak, Eds., COMPSTAT 84 (Physica, Wien, (1984) 242–248; ISBN: 3-7908-0007-4).

[28] SAS Institute, SAS User's Guide: Basics and Statistics (Cary, North Carolina 1982; ISBN: 0-917382-66-8).

[29] A.M.R. Smith, L.S. Lee, D.J. Hand, Interactive User-Friendly Interfaces to Statistical Packages, The Computer Journal 26 (1983) 199–204.
[30] SPSS, SPSS-X User's Guide, (McGraw Hill, New York, 1983).
[31] R.A. Thisted, Representing Statistical Knowledge and Search Strategies for Expert Data Analysis Systems, in: W.A. Gale, Ed., Artificial Intelligence and Statistics (Addison-Wesley, Reading, Mass. (1986) 267–284; ISBN: 0-201-11569-7).
[32] D.E. Wolstenholme, J.A. Nelder, A Front End for GLIM, in: R. Haux, Ed., Expert Systems in Statistics (Gustav Fischer, Stuttgart, New York, (1986) 155–177; ISBN: 3-437-40171-8).