

**ANALYST - A KNOWLEDGE-BASED SYSTEM FOR
ANALYZING PANEL DATA**

by **Alf Zimmer**
Josef F. Krems
Stefan Krischker
Armin Hartinger
Heinz Wontorra
University of Regensburg
Germany

1 SUMMARY

This paper describes ANALYST, a knowledge-based system for analyzing panel data. ANALYST supports product managers and market researchers in analyzing market data, helping them in the primary problems of marketing decision making: detecting significant deviations of product sales from long-term expectancies or trends. If deviations are found, the system gives hints for explaining them using rules of thumb.

As a user oriented, intelligent interface to a common database, it provides an overview of data concerning markets and products over time. The core of the system is a market model, derived from year long experience of experts in marketing and market research. Defining market models mathematically is difficult and in many cases computationally untractable (see Lilien & Kottler, 1983). Therefore, the market is modeled heuristically, using expert's rules of thumb. The result of analyzing verbal protocols of about 40 hours of interviews is a dependency network (product model), which covers connections between the main facts that describe the behavior of a product - and actions concerning the product - in the market. ANALYST's architecture is highly modular. The system is programmed in Smalltalk/V and runs under WINDOWS on PCs.

ANALYST was developed as a joint research project between the information management section, subsection market research, of a large European food company and the Department of Applied Psychology of the University of Regensburg, Germany.

2 ANALYZING MARKET DATA

2.1 The task and the problem

The subdivision market research of our partner company – a large European food company with about 6000 employees worldwide – is periodically supplied with data from market research institutes (e.g., Nielsen, Gfk). One essential task of the section is to analyze and interpret the data so that the inferred conclusions (e.g., significant, unexpected deviations in sales) can be used for further decisions or by other divisions (e.g., Marketing) of the company (Sedlmeyer, 1983; Little, 1979; Kiener, 1980).

The available data included here cover three years with about 500 products in a couple of hundred markets (depending on the product) with 25 variables regarding each product. The databases are updated monthly and bi-monthly, respectively. All the facts are on-line available in a PC-based data-base. The database amounts to about 200 MB handled by traditional panel data software. There are three partial databases: advertisement-panel (IMP), retail audit (NLI) and household-panel (G&I).

Routine access to the data base is restricted to a common query-language. Therefore, detecting deviations or surveying single products or submarkets is very time consuming. Everyday work-load in the job prohibits in-depth analyses of *the behavior of products or markets*. One single bi-monthly Nielsen-report for one product group consists of 364 pages, each with 36 rows and 13 columns, that is, about 170.000 numbers. For example, with three product groups a single analyst that could read each number in only one second, the analysis of the reports needs about 140 hours, that is, five man-months per year only for the Nielsen reports.

In general, marketing problems are very badly structured. Normally, only incomplete information about a process which is dynamic, nonlinear, delayed, stochastic and interactive (see Moorthy, 1993) is available. Furthermore, a lot of different variables affect each other and the goal-structure is multidimensional. To provide the product managers (PrMa) and market research groups in the companies with the necessary data, panelists like Nielsen or GfK offer a huge amount of data sets about different products in different markets. Even in the standard reports the information is too much to be effectively used for decision and policy making. Because of the amount of data and their algorithmic untractable nature expert-systems were considered as a major possibility to support decision processes (see Alpar, 1986; Bullinger & Fährich, 1993; aws, 1989, 1990).

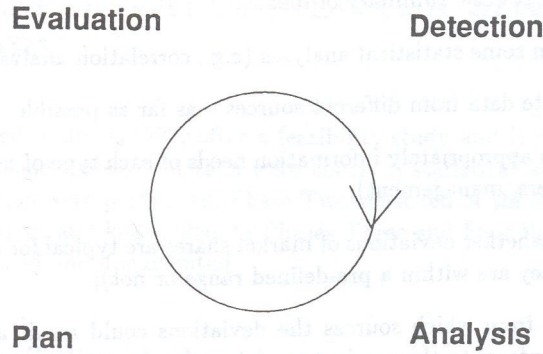


Fig.1: Problem cycle of monitoring

From a problem-solving point of view, analyzing marketing data can be considered a special variant of abductive reasoning. Abduction is reasoning backwards from effects to causes. The very first task in making these decisions is to detect anomalies. Then follows the search for the best explanations. After that, objectives have to be set and means to reach them scheduled. The final stage has to deal with evaluation (are all objectives met? how successful have single strategies proved?). (Fig.1)

2.2 Objectives of ANALYST

The main objective of ANALYST is to support product managers and market researchers in the analysis of market data. The routine job of a pre-analysis should be done automatically by a system, supplied with basic knowledge about data-structure, specifics of the panels, and the standard behavior of products in a certain market. In detail, the system should be able to

- provide a basic summary of means;
- perform some statistical analyses (e.g., correlation analysis);
- integrate data from different sources – as far as possible;
- address appropriately information needs of each type of user (e.g. product managers, management);
- detect whether deviations of market shares are typical for a product or not (i.e., they are within a pre-defined range or not);
- explain from which sources the deviations could result and check if the facts conform to the product model;
- make use of traditional panel data software.

2.3 Constraints on the System

The constraints on the development of the system were derived from the intended application of the system and from the hard- and software infrastructure of the company. The most dominant restrictions were:

- In-depth analyses should be possible in batch-mode (over night); on-line use should cover the already existing data-base features, enlarged by a more specific functionality.
- Modularity and flexibility: It must be possible to add further products or distribution channels easily. It should also be possible to switch between different assumptions based on modified management strategies.
- It should be possible to integrate the system into a larger context (MIS), i.e the system itself should be a module.
- Each user has to know what the system does, i.e. the knowledge should be represented in an obvious and transparent way.
- The system should run on the user's site in this case, under WINDOWS on PCs.

3 SYSTEM DEVELOPMENT: A SHORT HISTORY

The project started in June, 1992, after a feasibility study and is done in four phases: In Phase One technical details were checked, statistical analyses and knowledge elicitation were performed. Phase Two consisted of the implementation of basic functions and knowledge. In Phases Three and Four the prototype will be tested, refined and documented.

3.1 The Feasibility study

For three days a knowledge engineer from the University of Regensburg was at the company and completed an initial analysis of how the experts dealt with the data, what an automatic system should do, and whether it was possible to build a system dealing with that amount of data. The study revealed the following results:

- The problem can be split into subproblems.
- There are distinctive levels of abstraction.
- The experts are available (and highly motivated).
- The number of inferences (logical decisions) in each decision cycle is limited.

In this limited feasibility study the technical details (size of database, size of a dataset, hardware, data structures, desired functions, etc.) were evaluated with regard to the development of an expert system. After the positive results of the feasibility study, the development and implementation of the system ANALYST was started.

3.2 Phase One

As the first stage of development a psychologist with long experience as a knowledge engineer reviewed the literature on the analysis of market panels. Then he began interviewing experts in the market research section of the company. In the first two phases 5 experts from different backgrounds (market research, marketing, product management) were interviewed. Each expert was interviewed on

2-5 days at a rate of 1-4 hours a day. All experts had between two years and 10 years of experience. Although this knowledge provided useful background knowledge, it was inhomogeneous and on different levels of abstraction. Integrating this knowledge would have been a very time consuming process with an uncertain outcome. In order to start with a more homogeneous knowledge-base, we decided to further interview only the main-expert in marketing research and information management. With his background of 10 years in different food companies he was able to structure his ideas in an appropriate way. A "product model" - partly shown in fig. 2 - was developed on the basis of this interview data. For additional analyses, "support modules" were defined, which are intended to perform supporting in-depth analyses of certain facts (e.g. price).

In addition, statistical analyses were performed to detect redundancies in the data and to define threshold values, i.e., statistical significant variances in the data. Regression analyses identified redundancies in the facts and resulted in a reduction of to-be-concerning analyzed facts.

3.3 Phase Two

In Phase Two the essential functionality of the system was implemented into a prototype. The system allows the up-dating of product and market structures, respectively. Furthermore a so-called "event-list", which contains crucial events influencing customer behavior, can be edited. Thresholds for detecting significant deviations from long-term trends on the main facts like sales, prices etc. can be re-adjusted and fine-tuned. Also a relevant set for each product covering the interdependencies with competing products from other companies can be maintained. Finally, the prototype allowed a user to analyze panel data in a full range manner.

3.4 Phases Three to Four

The system will be tested through the end of 1993 at the food company and then refined. At this point, we expect to add some modules like a price module, that will analyze price shiftings in a very detailed way (price differences in regard to the market, the main competitors; price threshold values). If possible, a forecasting module will be added, i.e. a module for which the market analyst can define expected market shares for the next periods. Correspondingly the system will compare these values with the actual ones. If necessary, a further analysis strategy will be implemented. For management reports, the system will be able to provide a top-down analysis. Starting at a product segment, for example,

meat or soups, the system will check market share variations and, if there is no significant change, stop the analysis. If there is a change, it will analyze the market shares of the product groups and then single products within that segment. The reports will be quite short and will only consist of strings like "soups up because of: soup A, B, C up. Warning: soup D down." After that, a detail analysis including all facts can be started by an analyst.

The prototype will be documented, so that it is easily adaptable to other business processes (i.e. other companies).

4 MODELING THE MARKET

4.1 Knowledge About Panels

For the analysis and integration of panels, one has to know how the panels are built. The system does not cover this knowledge in detail, but if there are deviations in facts, and the deviations are not explainable from the data, the system can refer to a short list of possible causes, e.g., if the system detects a rising market share in G&I-data but no causes in the panel, it analyses the NLI-data. If there is no cause to detect, it reports: "increased market share (G&I). possible cause: increased distribution (not yet reported)."

The main facts are price, sales (in comparison to the market-share) and turnover (in comparison to the market-share) (see Rees, 1989). The most recent (actual) data are compared with the last period, the period of one year before and with cumulated values. In the monthly report, consumers (rel.), sales (kg), turnover (DM), buying acts and mean price are reported.

In order to determine in general, if selling in the focussed month is weak or strong, the monthly value is compared to a seasonal curve (monthly average over several years). Trends can also be detected by comparing the actual data with accumulated values for years.

In the IMP-Panel (advertising) 220 newspapers and handouts are analyzed regarding certain products and their promotion. Reported are price per piece and price per 100g, as well as number and package size.

When analyzing these data, the "quality" of the insertion is to be considered. Three or four insertions of the big customers with high turnover are much "better" than ten insertions of a small customer. The same is true for insertions, which are transcending a price threshold value.

Advertisement are highly dependent on the season. So they must decrease more than 25 % within a month to be regarded as critical.

If we start analyzing weekly data from IMP and G&I, possible time lags – depending on the customer – between the data have to be considered. Often advertisements are published on Thursday or Friday but consumers report their purchases only partly on Friday or Saturday, more often in the next week.

In launching new products, other facts are important. Especially the leading distribution is to be observed.

4.2 Rules and Interconnections

The core of the system is a product model, which represents the relevant market factors and their interconnections. The model is represented as a set of rules (contingencies between trends on facts). Part of the model is shown in Figure 2. The following rules are examples of what kind of connections between relevant features are used in ANALYST:

1. If the number of buyers increases, then sales increases as well – if buying intensity and re-buying rate are equal.
2. If average price rises, then sales go down ("Only for a short time").
3. After an advertising campaign secondary product placement and market share will rise and price from the household-panel will lower.

The essential idea of the model is to provide a task-oriented view, i.e. modeling the expectations of PrMas regarding specific effects of marketing activities on products.

As a result, a complete pre-analysis of the data is provided quickly with costs far below that of a human analyst. The result is a more global view on relevant deviations and trends, covering much more than the huge amount of data of a single product,

There is not only a threshold-analysis but also a modeling of expert-knowledge about market behavior and business-function which can be worked upon. That reduces data and provides information. For example: Suppose that there is an advertising campaign for a single product. Of course, the PrMa expects a rise in sales. Traditional analyzing systems would check each product whether there is a significant deviation, and print out a profile with ups and downs of the facts

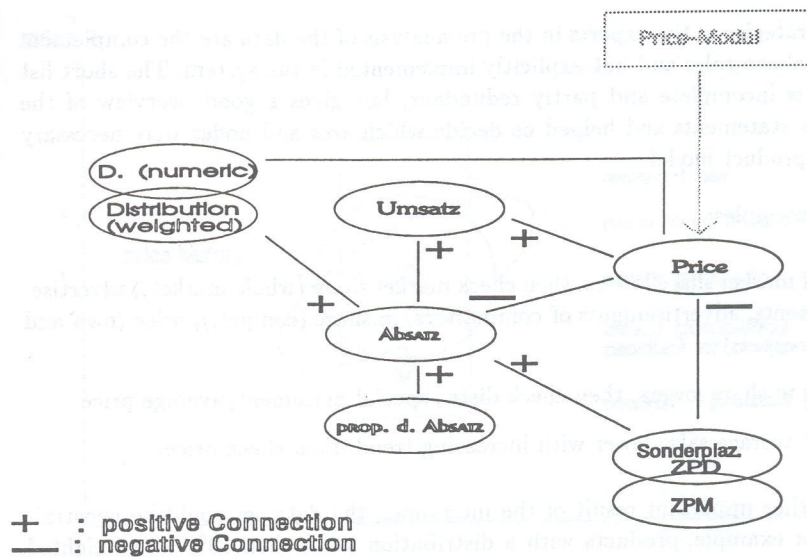


Fig. 2: Part of the Product-Model

for each product. The PrMa would have to search in the (very long!) list of ups and downs by himself in order to find, if his expectation was confirmed.

In contrast, ANALYST is able to check the connections between advertising and sales. If sales do not behave as expected, the system prints out a warning. Thus, only a few pages of the results of the pre-analysis' guide the PrMa to the problematic product (i.e. market) behavior at once. He then can try to find explanations in a more detailed analysis, that the system probably might have missed.

4.3 Strategies

In general, we found that experts analyze the panel-data in two steps. At first, facts are traced in order to detect significant deviations, and the results are stored elsewhere (a piece of paper). Afterwards, the facts are analyzed to detect possible causes, that is, connections to other relevant facts are built. If there are no causes to be found in the data, the experts go back to their knowledge about panels or about the world ("There was a post strike. So we got biased data from GFK, because the consumers could not send in their questionnaires.").

The strategies of the experts in the pre-analysis of the data are the complement of the above rules and not explicitly implemented in the system. The short list below is incomplete and partly redundant, but gives a good overview of the experts statements and helped us decide which arcs and nodes were necessary in the product model.

Some examples:

1. If market share lowers, then check market share (whole market), advertisements, advertisements of competitors, m.share (compet.), price (own and compet.)
2. If m.share lowers, then check distr., special placement, average price
3. If average sales lower with increasing trend, then check price.

As another important result of the interviews, the data set could be constrained. For example, products with a distribution lower than 10% or a weighted distribution lower than 20 % are not considered. These values are set as defaults for the system and can be modified by the user. This reduces the number of to-be-analyzed products.

4.4 Prices

A critical factor in the market dynamics is the price. Therefore, the price itself needs a separate critical analysis. To decide, if a price is attractive, analyses of price distances are necessary. These analyses are an essential part of market observations (development of price and bulk sales). They are the basis for an improved price-promotion-model. The questions are:

1. Are there threshold values of prices, that are important for changes of sales? The consumer thinks in "50-Pfennig-units": a price-shift from 1,99 DM to 2,09 or from 2,49 DM to 2,69 is more important than from 2,29 DM to 2,39. The price-history of the product, the difference to the main competitor, the product group, the segment and the market as a whole must also be considered.
2. Are sales influenced by the price-history? Maybe the consumer accommodates to the new high price or reacts to lowering prices weaker than to rising ones.
3. Is there a threshold value of number of advertisements required before advertising has an effect?

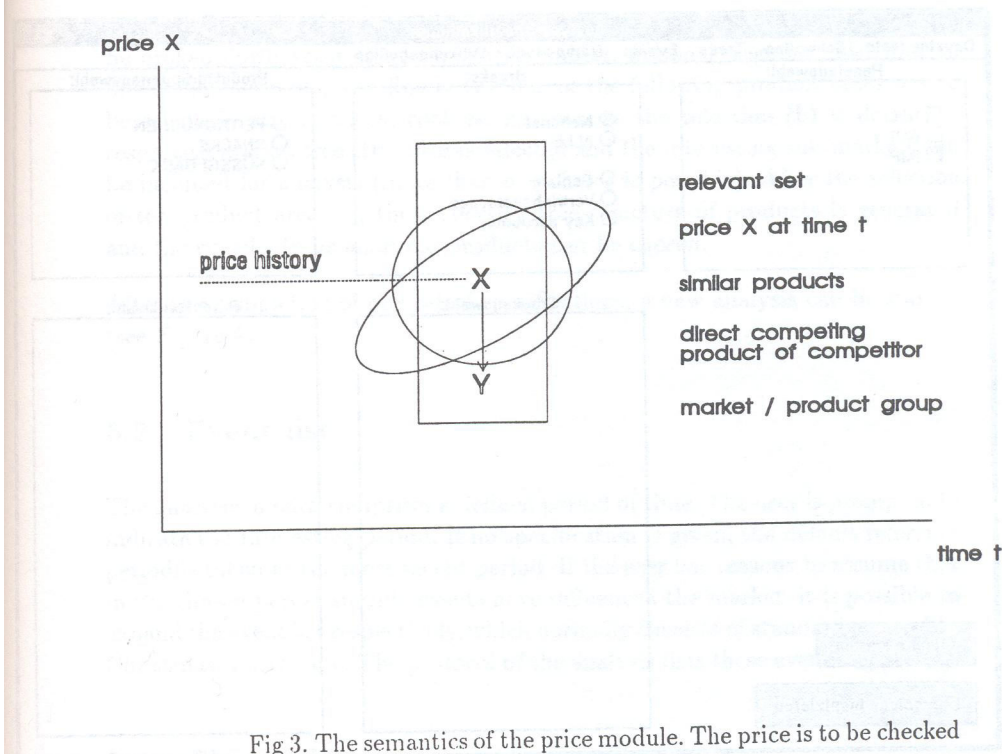


Fig 3. The semantics of the price module. The price is to be checked against the last periods and a relevant set of competitors.

4.5 "Validating" the Model

The existing product model is a kernel of a still more sophisticated one, which is to be developed while using the system. There are, of course, exceptions of our network and the real usefulness can be proved only by using the system over several months or years. But in several statistical analyses (regressions) it could be shown that the connections between the main features are valid in general. A review with experts of the company as well as experts from panel-institutes showed no single severe disagreement with the product model of ANALYST.

5 HOW THE SYSTEM WORKS

The system reports critical deviations of single variables from the the last period and reflects the variation in connection with other variables. If, for example, the

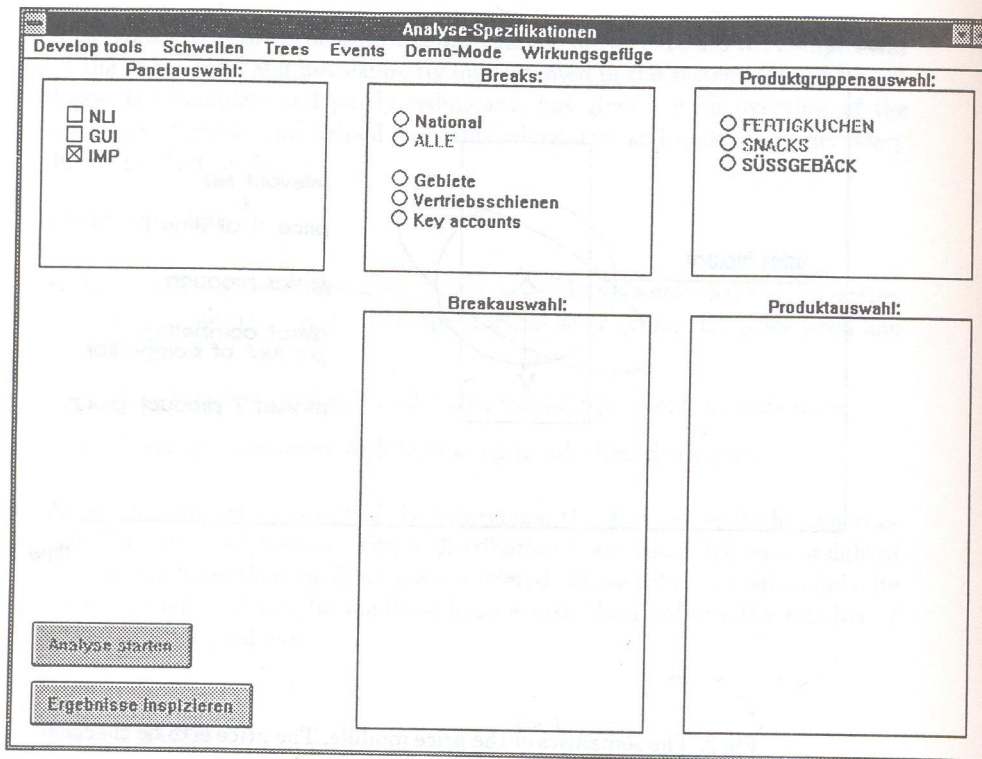


Fig.4: Starting the analysis

market share decreases below a critical point, the system provides possible causes: the prices of the competitors decreased, the distribution of the competitors rose, the product's distribution decreased, or all of these.

5.1 Specification of Analysis

After initialization, the system appears on the screen with a window labeled 'Analyse-Spezifikationen' (Specifications for Analysis). In this top-level view of the system, the analysis is set up. It is possible to select specific kinds of panel data, specific sub-markets, or products. Furthermore, the respective market or product trees can be edited, the thresholds for the event detection can be fine-tuned, and even the 'event list' can be expanded. Additionally, a bitmap graph of the underlying causal structure is provided (see Fig 4).

Specifically, three things are done: (a) the most interesting panel data (IMP, NLI, G&I) can be selected; (b) for the selection of the interesting market, the data can

be broken down according to a specific subset: national, regional, distribution line or key accounts; and finally (c), one of the following product areas has to be selected: cakes, snacks, cookies. As soon as the selection (b) is done, the respective market tree structure is selected and the interesting sub-markets can be included for analysis (more than one choice is possible). After the selection of the product area (c), the respective tree structure of products is generated and the specific to-be-analyzed products can be chosen.

After the completion of one set of specifications, a new analysis can be started (see Figure 4).

5.2 Event list

The analysis of data comprises a defined period of time. The user is prompted to indicate the interesting period. If no specification is given, the default reference period is taken as the most recent period. If the user has reasons to assume that in the chosen period specific events have influenced the market, it is possible to expand the event list respectively, which normally consists of standard events like Christmas, Easter, etc. The protocol of the analysis lists these events separately.

5.3 Thresholds

For the detection of significant changes thresholds of relative changes are used in the data flows. These thresholds are usually deviations from reference values represented as relative changes. The ratio can be fine-tuned by the user before starting analysis.

5.4 Tree Structures for Products and Markets

The tree structures defined by choices (b) and (c) can be edited by the user, that is, vertices (submarkets or products) can be added or deleted. For every vertex in the tree structure for products, a set of relevant other products (relevant set) is defined. This comprises all products that share a comparable market behavior with the product in question. If the relevant set is non-empty, it is automatically analyzed together with the product marked by the user. An additional analysis of products from the relevant set is only performed if a user-defined threshold is surpassed.

5.5 Starting the Analysis

After the analysis is started a verification of the intended analysis is requested. Furthermore, it is asked if the default value for the reference period of analysis is wanted, otherwise a list of plausible or possible starting points or both is provided. The system also requires the analyst to name the current analysis. This name is used for further reference to this analysis.

5.6 The Procedural Characteristics of the Analysis

The submarkets selected in the respective tree structure and the selected products are combined in a Cartesian product, that is, for all products in all market segments the interesting facts (depending on the selected kind of panel) and the corresponding data are available for further analysis.

The specific analysis of a single product in a defined market segment consists of the following steps:

(i) The system checks if a significant change has occurred. For this, the actual period, the preceding period, and the corresponding period of the last year are compared.

(ii) If the selected product is a terminal node in the tree structure, that is, a singular product, the consistency of the actual facts with the underlying causal structure is tested. This test is done by correlating all pairs of facts and by checking if the positive, negative, or zero correlations correspond to the relations in the causal structure.

(iii) If a significant change is observed, it is checked if it corresponds to changes in the next higher level of the tree structure (the 'parent' product). If that is the case, the specific change is explained as caused by a change in a more general segment of products. In the other case, it checks if the behavior of 'sibling' products explains the singular changes. For instance, the introduction of cheap microwaves might cause a significant increase for pop-corn irrespectively of the brand, but in the brand independently from the other products making up the next-higher level in the tree structure. Here again, for all 'sibling' products the analyzes are made correspondingly.

(iv) If the selected product is related to a relevant set (RS, see above), then the entire set is analyzed correspondingly.

5.7 The Inspection of Results

The analyst can inspect a list of all analyses, including the most recent one. Entire or partial results can be viewed by choosing chapters, sections, or charts from each analysis.

(i) A chapter corresponds to the analysis of one product in one market segment. Accordingly, the number of chapters corresponds to the cardinality of the Cartesian set (see above).

(ii) If a chapter is selected, the corresponding set of sections is shown. The number of sections corresponds to the number of products in the associated relevant set plus the selected product itself.

(iii) In a chosen section the charts of analyses in the section are provided, namely the analysis of consistency with the underlying causal structure, the comparison with the next higher level in the tree structure (the 'parent' product) or - if necessary - with the corresponding 'sibling' products. If no significant changes have been observed, only one chart is given. If the behavior of the 'parent' products explains the changes in the specific product, two charts are given. And if comparisons with 'sibling' products have been necessary, the number of charts corresponds to the number of interesting 'sibling' products. Charts are the lowest level for the inspection of results.

If in the inspection window no chapter is selected, the system provides the entire analysis as a default option. Accordingly, the possible selection filters reduce the set of results to be inspected in a top-down manner.

By using the 'Reset' button all selections can be undone. The 'Cancel' button closes the selection window. The button 'Ergebnisse zeigen' (show the results) opens a text window (see Figure 5) according to the selections made. Clicking the button 'Analyse wegschreiben' (store analysis) saves the entire analysis on file.

5.8 Some Technical Details

ANALYST has a highly modular architecture. The main components are the Product-, the Price-, the Event- and the Threshold-Editing-Module. The system extensively uses features from object-oriented programming. For example, the hierarchy of products as well as the hierarchy of markets is represented as a tree structure. All structures can be modified via a graphical tree-editing interface.

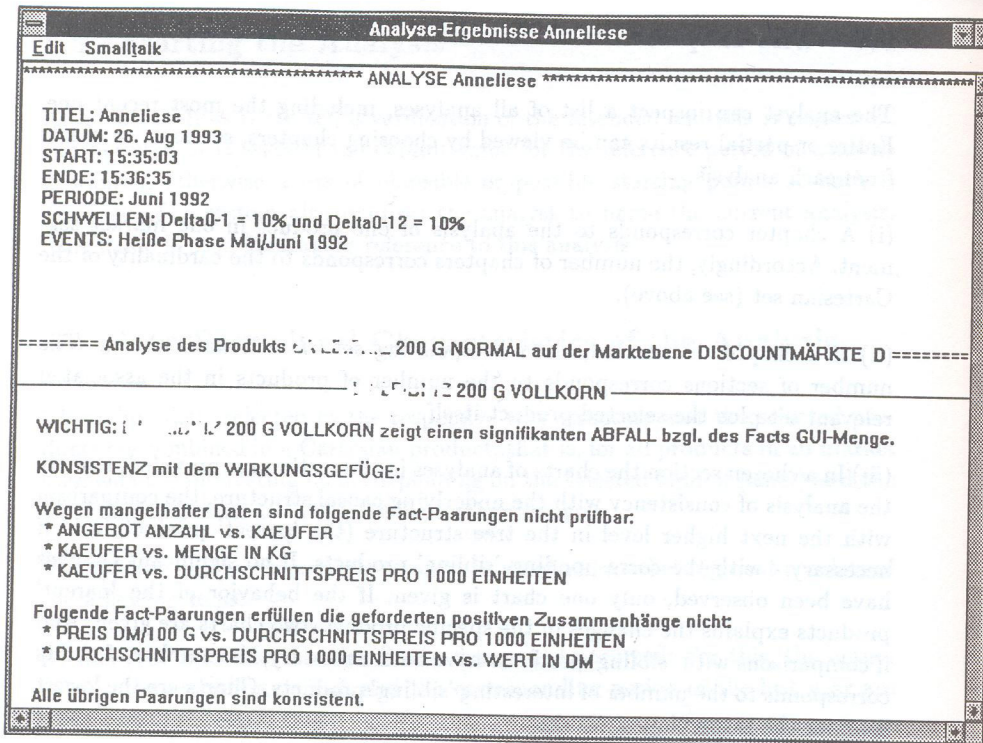


Fig.5: Results of the analysis

ANALYST is implemented in Smalltalk/V under WINDOWS 3.x. It can be used on PCs with at least 8 MB RAM and 33 MHz.

6 RESULTS - FIRST EVIDENCE OF SUCCESS

At the end of the first year of development, a prototype was demonstrated (June 1993) for the top management of the partner company. They were pleased. User interaction is easy and the analysis of the subset of data runs can be performed within acceptable time spans.

Besides mere economic points of view, a more strategic aspect also has to be stressed: developing an expert system requires documenting, analyzing and systemizing the knowledge available in a company. The knowledge-base can be considered a task-specific textbook - in our case on analyzing market data - that

can also be used for training and the resource remains part of the company, even when employees leave.

Another gain of using a computer-based system for data analysis is that users are freed from routine jobs. They can engage in more important and difficult tasks, which should result in an improvement of decisions.

Another result of this kind of analysis is the organizational impact, resulting in a knowledge-based system for this domain. In large organizations a lot of knowledge is not documented and – as only one consequence – there are no standards, e.g. for the analysis of data. This potentially means that information is equally labeled but has a different meaning for different people, which may lead to confusion in decision making. The employees concerned with analyzing data are now supported by a system that reduces their routine job and enforces standards in communication, analysis, and decision making. Additionally, it can be expected that their motivation will rise because they will have more time to extensively analyze data. On the long run, the system offers a good chance for lean management / lean production. With systems like ANALYST, it is possible to cover the whole marketing process and to shorten information paths and decision times.

Of course, there still remains work to be done: The interface must be improved to cover the needs of different user groups, e.g. to set different parameters of the system. Another point is to calibrate the system: top management only needs the essential features of the analysis, and market researchers need different types of result than PrMa.

7 CONCLUSIONS

This project of developing an intelligent tool for the analysis of panel data has been successful. In a comparatively short time a valid product model was developed and implemented. Essential problems of data integration from different sources were solved and a knowledge-based prototype was developed. ANALYST will support product managers, management and market researchers in routine and complete analyzes of huge data sets. It will enable users to focus on decision making rather than on low-level “number crunching”.

8 REFERENCES

- A.C. Nielsen GmbH (Ed.) (o.J.), Nielsen Marketing Forschung, Frankfurt/M.
- Alpar, P. "Expert Systems in Marketing." Dep. of Information and Decision Sciences, College of Business, Univ. of Illinois at Chicago, Working Paper Series, 1986, Number 86-19.
- asw (ed.) "Wissen nutzen. Was leisten Expertensysteme im Marketing?" asw 12/89, 1989, pp. 38-49.
- asw (ed.) "Wissen nutzen. Was leisten Expertensysteme im Marketing?" asw 1/90, 1990, pp. 62-67.
- Bullinger, H.-J., Fähnrich, K.-P. "Informationsarchitekturen im Unternehmen als strategische Herausforderung für das Management". NfD 44, 1993, pp. 174-176.
- GfK. Panel System Forschung, Nürnberg, 1990.
- Kiener, J. Marketing-Controlling. Darmstadt, Toeche-Mittler, 1980.
- Lilien, G.L., Kotler, Ph. Marketing-Controlling and Marketing Decision Making: A Model-Building Approach, New York, Harper & Row, 1983.
- Little, J.D.C. "Entscheidungsunterstützung für Marketingmanager", ZfB, 1979, pp. 982-1007.
- Moynihan, G.P. "An Executive Information System: Planning for Post-Implementation at NASA". Journal of Systems Management, 44, 1993, pp. 8-31
- Moorthy, K.S. "Theoretical Modeling in Marketing". Journal of Marketing, 57, 1993, pp. 92-106.
- Rees, J.v. Glossary of Marketing Research. Amsterdam, E.S.O.M.A.R. 1989.
- Sedlmeyer, K.-J. Panelinformation und Marketing-Entscheidung. München, Florentz, 1983.