Annals of Operations Research 55(1995)1–7 

# Data, expert knowledge, and decisions: an introduction to the volume

W. Gaul, F.J. Radermacher, M. Schader and D. Solte

## Background

Important goals in science have always been to understand how human beings handle their affairs and to assist them in their various activities. Thus, support for the process of decision making, as one of the central challenges of mankind, is a very important direction for scientific research.

Seen from a theoretical perspective, humans have to cope with rather complex decision environments (e. g., resulting from lack of information or the dynamic, interdependent, and stochastic nature of situations) and difficulties in modeling decision behavior (e. g., because of changing preferences or game-theoretical restrictions).

Stressing these points of view – and given that decision making is an activity centrally related to human intelligence – up to now AI (Artificial Intelligence) research has not fully captured the human capacity for making decisions in complicated situations. Classical decision theory has, however, provided means to describe and use the frameworks, options, and practical outcomes of human decision behavior.

In the last decade, ongoing AI research has focused on understanding problem-solving tasks and obtaining, processing, and storing as well as enlarging and improving many kinds of knowledge and meta-knowledge. This work has resulted in techniques for learning and for knowledge handling. As a consequence, it is now widely accepted that so-called expert knowledge should be an integral part of application systems for real-world problems.

In addition to knowledge-oriented aspects, most decision situations also present a data problem. Adequate information has to be collected, processed via appropriate methods, and, finally, translated into suggestions for decision alternatives. Research activities related to data analysis and statistics have a long history and include the handling of different data types (e. g., quantitative, qualitative, mixed, symbolic). When one tries to provide a verifiable and scientifically well-founded basis for the transformation of problems into decisions, the importance of models becomes evident. Here, in addition to data analysis and statistics, the field of operations research, with its long tradition of modeling and problem solving, plays an important role.

In our opinion, in most applications related to the research directions described above, a typical pattern of sequential tasks to be solved can be

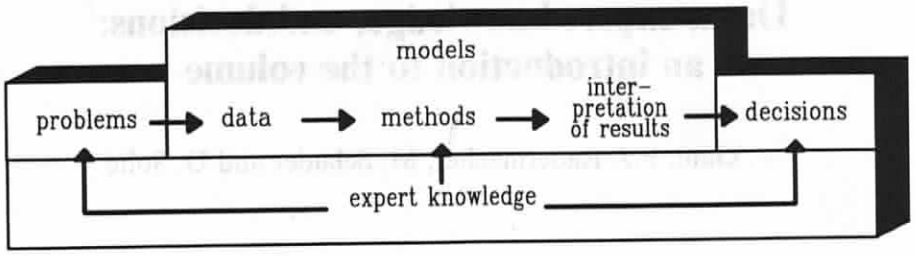


Figure 1.

observed. The figure outlines the main phases of such a problem solving process; in addition, repeating of previous phases is also possible.

As indicated in the figure, expert knowledge is the basis for and is needed within as well as between the phases of the sequence. Here, in order to successfully transform problems into decisions, cross-disciplinary research must be combined where, often, computer assistance is indispensable because of the enormous amount of data, the size of the solution spaces, and the magnitude of the results to be interpreted, which real-world problems often show.

Against this background, we have attempted to bring contributions together in this volume that are related in a variety of scopes to the topics discussed above and that describe computer support for the problems they tackle.

**Contributions in this volume**

Since the selected papers contain material from different areas, we have divided the volume into three parts: Knowledge and Structures, Data Analysis, and Applications. In cases where a contribution emphasizes questions from more than one area, it has been assigned to that part we consider most appropriate.

**I. Knowledge and Structures: how to represent, handle, and find knowledge and insight into structure**

*"COLAB: A hybrid knowledge representation and compilation laboratory" by H. Boley, P. Hanschke, K. Hinkelmann and M. Meyer*

This paper gives a broad overview of topics arising in the fields of knowledge representation and the use and combination of various methods of inferences. Quite different approaches from logic programming, object orientation, functional approaches, constraint mechanisms and others are described. A homogeneous language for the purpose of integration and modeling environments is also presented. The authors also discuss concrete examples from the field of construction using computer design tools. For these examples, a concrete knowledge base is given that can give uninformed readers a feeling about what such knowledge

bases look like, what kind of information is encoded, and how much work in designing and representing might eventually be involved.

*"A unifying approach to heuristic search" by A.E. Eiben, E.H.L. Aarts, K.M. van Hee and W.P.M. Nuijten*

In this paper, search is discussed on a very abstract, general level as one of the unifying elements in inference and optimization systems. This gives us better insight into how heuristics work and how many systems that are made to look for good solutions are organized on an abstract level. In particular, connections are described to fields such as artificial intelligence and operations research, in which search procedures play a central role, which at the same time indicates a certain integrating aspect between the two fields (which also extends to other fields, such as simulated annealing).

*"On the usage of qualitative reasoning as an approach towards enterprise modelling" by A. Hinkkanen, K.R. Lang and A.B. Whinston*

This paper generally aims at enterprise integration via suitable information and decision support systems and builds strongly on the background of qualitative reasoning. With this in mind, the paper gives an overview of issues in enterprise integration, as well as in the field of qualitative reasoning, and describes many of the methods, algorithms, and motivations in this context. Also, the authors develop a new and original framework called "rules construction reasoning", and show to what extent it might be usefully applied.

*"Running time experiments on some algorithms for solving propositional satisfiability problems" by J. Mayer, I. Mitterreiter and F.J. Radermacher*

The satisfiability problem is of basic importance in complexity theory. Its particular prominence results from the fact that an enormous number of real-life problems can reasonably be described as satisfiability problems. Also, it is interesting to note that although the problem is known to be NP-hard, many special cases of the problem can be handled efficiently (we mention here Horn clauses and the connection to the PROLOG language). Also, many heuristics have been discovered that work well on most satisfiability problems, given certain probability distributions. From this, several powerful general-purpose algorithms have resulted. The paper gives some hints for understanding the whole topic and reports on the evaluation of some of the available algorithms for test sets that have been compiled for this study.

*"Autonomous theory building systems" by W.J. Paul and R. Solomonoff*

Machine learning is among the most interesting, promising but also difficult topics considered in AI and related fields. The number of methods that can be applied has grown in recent years, and today several approaches are available, be it neural networks, genetic algorithms, machine learning in the AI framework, in

particular learning from examples, and others. The present paper gives another approach that has close relations to the Kolmogorov complexity theory. This means that one tries to code information as compactly as possible and to find, if possible, small encoding algorithms for (Turing) machines to produce a given data. This idea is applied to try to make machines invent programs, and to achieve this by presenting the right examples to those machines, and to have them invent certain basic constructs. The paper gives some indication that it might be possible to have systems learning basic constructs in programming in a reasonable time and offers in this respect interesting perspectives for the future.

*"On cautious probabilistic inference and default detachment" by H. Thöne, W. Kießling and U. Güntzer*

This paper addresses the handling of fuzzy or weak knowledge within the framework of symbolic knowledge processing, for instance, as an alternative to other methods such as neural networks. The authors give a nice review of many issues and approaches in this field and do this with the aim of developing and transforming other approaches to a unified formalism called DUCK. They can show that for certain interesting problem types it is possible within DUCK to deal with uncertain knowledge in a homogeneous way that is also algorithmically feasible.

## II. Data Analysis: progress and recent trends

*"Probabilist, possibilist and belief objects for knowledge analysis" by E. Diday*

In this survey paper, the author extends classical notions of data analysis to the symbolic objects approach. Symbolic objects are well-suited to incorporating knowledge representations of various types. Boolean, possibilistic, probabilistic, and belief objects are introduced and discussed with some of their main characteristics and properties. Four data analysis problems and their symbolic extensions are used to illustrate the process of deducing knowledge from data.

*"Symbolic objects: Order structure and pyramidal clustering" by P. Brito*

In Brito's contribution, the concept of symbolic objects which generalize the classical tabular model of data analysis is formalized. A particular class – assertion objects – is studied to express the duality extension-intension. A pyramidal clustering method is presented to solve the problem of clustering a set of assertion objects where each cluster is a complete assertion object whose extension is the cluster itself. The pyramidal structure then leads to the generation of rules.

*"Histograms in symbolic data analysis" by F.A.T. de Carvalho*

Assertion objects are also dealt with in the paper by de Carvalho. Here, an approach to describe a knowledge base of Boolean assertion objects by histograms is developed. Such Boolean assertion objects are well adapted to describe a class of individuals, simultaneously taking into account variables that may assume zero,

one, or several different values, and that may be logically dependent on other variables. These dependencies are expressed by rules.

*"Discrimination decisions for 100,000-dimensional spaces" by W.A. Gale, K.W. Church and D. Yarowsky*

High-dimensional discrimination analysis problems in natural language processing are discussed in this contribution. Information retrieval (discriminating documents by their relevance to some query), author identification (discriminating texts by their authors), sense discrimination (discriminating multi-meaning words by their meanings), restoration of capitals and accents (avoiding ambiguities in texts where no capitals are used or accents are deleted), and the "person or place" problem (distinguishing nouns for persons from those for places) are used as examples for the demonstration of difficulties arising in natural-language processing. A method introduced by the authors for the sense discrimination problem is shown to be useful for other text discrimination questions, too.

*"Optimization and data structure: Seven faces of dual scaling" by S. Nishisato*

After some historical remarks and a description of optimization aspects, seven key characteristics of dual scaling are presented: duality, disparity between row and column space, data types and multidimensional spaces, expanded dimensionality, over-quantification, the principle of equivalent partitioning, and multiway principal component analysis. Examples are used to combine old and new faces of dual scaling for an enhanced understanding of the method.

*"When not to analyze data: Decision making on missing responses in dual scaling" by S. Nishisato and H. Ahn*

A well-known problem in data analysis/statistics is related to the question of what to do when data are missing. This paper describes effects of missing responses on dual scaling results. Deletion of data belonging to the missing values, introduction of an extra category for missing responses, and imputation of responses for missing values are discussed, the latter in more detail. Numerical examples are used to show how decision making might look like when missing responses have to be taken into consideration in dual-scaling applications.

*"A numerical strategy to defectuous knowledge using" by M. Sebag and M. Schoenauer*

The authors deal with the problem of inconsistent, incomplete, or redundant knowledge contained in a rule base. A numeric-symbolic strategy to enable a knowledge-based system to perform properly – even when based on such a defectuous set of rules – is proposed. The algorithm is related to the notation of similarity measurement and $k$-nearest neighbors clustering, which is well-known in numerical data analysis.

*"A system on categorical data analysis: SANI" by S. Adamov*

SANI is a computer system that documents activities in data analysis in Russia. Besides techniques from classical research areas (correspondence analysis, measures of association, contingency table analysis, loglinear analysis) mentioned in the paper, the author gives examples for determination analysis, the handling of block and quasi-symmetric structures in contingency tables, and a classification application to stochastic processes with segment-specific estimation of transition matrices, all of them supported by SANI.

## III. Applications: recent steps to new and powerful forms of knowledge-based decision support systems

*"Different kinds of models to integrate the OR-expert's knowledge into decision support systems" by M. Lachmann*

The paper shows how expert knowledge in operations research can be reasonably integrated into the context of decision support systems (DDS). The guiding principle is a model of different phases of interaction between a DSS and its user. The different kinds of models necessary to enable a DSS to be supportive for its user in this context are shown. Particularly, the question of how operations-research knowledge can be reasonably represented is addressed. The paper also gives hints for the implementation of an existing prototype.

*"How to make OR-results available: A proposal for project scheduling" by R. Müller and D. Solte*

This paper addresses the question of how to make available and how to get realistic insight into the behavior of hard algorithms for application fields such as scheduling. A typical situation in this context is the availability of hundreds of algorithms for special subclasses of certain problems where one of the main difficulties is to handle these algorithms simultaneously and make available knowledge about when and how to apply a particular algorithm, using statistical information on the suitability of certain algorithms under appropriate conditions. Major questions concern "accounting problems" when particular algorithmic tasks are required to apply special methods but might also be useful for other methods. In particular, the question of finding adequate software-engineering strategies and software-development environments for such work is addressed. Also, the paper reviews recent developments at the Research Institute for Applied Knowledge Processing (FAW) in Ulm and describes a method base for scheduling algorithms built within the framework of a cooperative effort between FAW and the Technical University of Berlin.

*"Advanced DSS for scheduling: Software engineering aspects and the role of Eigen-models" by R.H. Möhring, R. Müller and F.J. Radermacher*

As mentioned in the previous paper, for a number of years now, a new type of

scheduling system has been under development in a cooperative effort between FAW in Ulm and the Technical University of Berlin, using sophisticated software-engineering technologies to integrate hundreds of algorithms from the field of scheduling and to make knowledge available of when and how to best use these algorithms to treat certain problems. In this context, the concept of an Eigenmodel was incorporated as a way for a system to describe part of its internal functions for its own use and as a way to be able to monitor the system's own behavior in order to eventually modify it and learn better behavioral patterns. The paper, which is a reprint from the 1994 HICSS conference, gives an introduction to the subject and reports experience from the realization of such a system in Berlin and Ulm.

*"Positioning analysis using knowledge-based support" by D. Baier, W. Gaul and F. Wartenberg*

In this contribution, a knowledge-based marketing data analysis system is presented. Assistance for decision making by automated analyses of data is exemplified by a sample consultation session dealing with a real world positioning problem. The approach is based mainly on knowlege about user wishes, positioning analysis objectives, and the input/output behavior of methods. Suggestions for possible sequences of applications of methods are given by the system. For selected alternatives, solutions are computed and visualized by graphical outputs (dendrograms, multidimensional space representation, etc.).

*"Robust vehicle routing DSS and road networks on a European scale?" by A.J.M. Beulens*

Traffic management in a united Europe is becoming an increasingly important topic. At the same time, new technical options are substantially changing the environment in which such planning tools are needed. Examples include driver assistance systems, traffic management and influence via road pricing, satellite position monitoring, and so forth. From a mathematical point of view, we have to deal with problems that have a combinatorial structure depending on the constraints given in a particular situation, where, once certain parameters are fixed, there are quite complicated versions of travelling salesman problems and variants to be solved. Further, nasty side aspects have to do with the nature of road systems and the need to use information on such road systems with different levels of granularity. This reflects the actual situation with digital maps and the possible use of geo-information systems. This is an extremely difficult topic, and this paper gives a description of developments that have taken place and of further work that will have to be done and puts these into the more general context of our responsibilities to society and nature.