SYNTHESIS OF HIERARCHICAL DECISION SUPPORT MODELS FROM SOCIOECONOMIC DATA

Marjan Krisper (1), Blaž Zupan (2)
(1) Faculty of Computer and Information Sciences, University of Ljubjana, Slovenia
(2) Department of Intelligent Systems, Jožef Stefan Institute, Ljubljana, Slovenia
e-mail: marjan.krisper@fri.uni-lj.si, blaz.zupan@ijs.si

ABSTRACT

The paper presents a data mining approach based on function decomposition to construction of socioeconomic decision support models from domain data. The models are defined in the form of the criteria tree and corresponding utility functions. Given the criteria tree and domain data, the method proposed discovers the utility functions. The approach is interactive as it involves the expert in the process of discovery. We illustrate the benefits of this method on the case of construction of knowledge infrastructure socioeconomic model.

1 INTRODUCTION

A determined orientation of the developed countries to foster the development of information infrastructure that will foster their transition to information society shows that we are undergoing a period that will exert a decisive influence on their future development. This is also or even more true for the Central European countries where the change of the political, economic, and legal system is the basis for their gradual transition to a modern society and their prospective integration within European Union.

In order to monitor and evaluate such transition, compare countries' successfulness, and investigate for the alternative development scenarios, one may benefit from models that assess the value of country's system given a selection of its observable criteria. Crucial to the utility of such models is their ability not only to reach valid and (hopefully) accurate conclusions, but also to explain why such conclusion were obtained. The modeling methodology should provide grounds for explorative analysis of alternatives being evaluated, making the model and decision support environment a valuable tool for decision expert. In these terms, classical numerical decision models that are based on criteria weighting (Chankong and Haimes, 1983) may be inadequate and pose problems where modeling of more complex interdependence of criteria are required (Bohanec et al., 1995). This paper builds on alternative approach for multi-attribute decision making that hierarchically orders the criteria in the criteria tree and introduces new aggregate criteria. The aggregate criteria simplify utility function elicitation and play major role for explorative analysis. The approach was first proposed by (Efstathiou and Rajkovič, 1979) and was subsequently used in over fifity real-world applications. In this paper, we refer to its implementation in an expert system shell for decision support called DEX (Bohanec and Rajkovič, 1990).

Within DEX, the model is manually developed from scratch. Usually, an team of experts would be formed that would typically spend from one to five days to define the criteria tree and the utility functions. In this paper, we advocate an alternative approach which uses a criteria tree as defined by the experts, but utilizes the evaluated options (domain data) to help build the utility functions. The benefit is twofold: (1) it may substantially reduce the time spent by the experts to build the utility functions, and (2) it guarantees the resulting model to be consistent with the existing domain data. We will refer to this approach as a HINT methodology, where HINT (Hierarchy INduction Tool) is also the name of the environment the method is implemented in (Zupan et al., 1998; Zupan 1997).

2 INTERACTIVE DISCOVERY OF UTILITY FUNCTIONS BY HINT

Let us consider a small (and imaginary) socioeconomical problem of finding a quality of knowledge infrastructure (`ki`) from the quality of education (`educat`), telecommunication network (`tel`) and computer deployment (`comp`). Knowledge infrastructure is the *overall utility* or a *target criterion*, and the other three criteria are called *basic (observable) criteria*. Suppose that domain data exists as given in Table 1, that for specific combination of basic criteria gives the value of overall utility. Suppose further that the criteria tree for the target model is as given in Figure 1. The problem is to discover the unknown utility functions G and H (H maps `comp` and `tel` to a new *intermediate criteria* infrastructure (`infra`), and G maps `educat` and `infra` to `ki`). In this way, a domain data can also be regarded as a utility function mapping set of basic criteria to target criterion.

educat	tel	comp	ki
low	med	high	high
low	high	low	high
med	low	high	med
med	med	low	low
med	high	high	high
high	low	low	low
high	med	low	low
high	high	low	high
high	high	high	high

Table 1: An example domain data.

The problem is therefore to transform a utility function F to functions G and H. For this, we use a technique called *function decomposition* that was originally proposed in early fifties for computer aided construction of digital circuits (Ashenhurst, 1952; Curtis, 1962) but recently extended to handle mutli-valued functions (Zupan 1997; Zupan et al. 1998). Function decomposition starts by constructing a partition matrix, which uses all combinations of criteria in H for column labels and all combinations of other basic criteria for row labels. For our example, a partition matrix is given in Table 2.

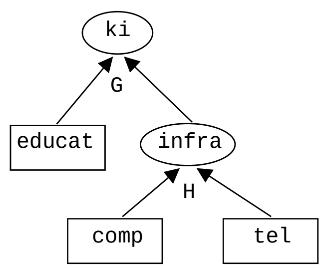

Figure 1: An example criteria tree

Function decomposition attempts to discover utility function H by assigning labels to partition matrix's columns. There are two constraints: the labels have to be assigned such that resulting G and H are consistent with F, and minimal number of different labels should be used. HINT solves this problem with incompatible graph coloring, where graph consist of nodes representing each column and two nodes are connected if, for a pair of columns, there exists two entries of the same row that are different. Once H is defined, it is trivial to obtain G from the domain data (see Zupan et al. 1998 for the details).

The above approach works fine if there is enough domain data. In the cases where partition matrices are sparse, the technique advocated in this paper is to involve the expert in the process of interactive discovery of function H. For our example data, the interface used by HINT is given in Figure 3. There are three important columns in the window shown in the Figure. Column "H" shows the values of H that are agreed or entered by the expert. Column "Proposed" shows a

comp	low	low	med	med	high	high
educat tel	low	high	low	high	low	high
low	-	-	-	high	low	-
med	-	med	low	-	-	high
high	low	-	low	-	high	high

Table 2: A partition matrix for data from Table 1.

single value that is proposed by HINT and is selected from the set of possible column labels given in the column "Candidates". If only a single candidate is proposed, this is a strong indication for the user to consider this suggestion and make it his choice for H.

The approach is interactive, as the value of the above described columns changes by every action of the user. In these terms, HINT guides the user to define H by proposing only the candidate values that would not violate the consistency constraint. For our example, it was found that the expert needs to make only two non-trivial decisions (set the values for two instances of H) to arrive to a fully defined utility function. The resulting model is shown in Figure 2.

In the real-world domains with more complex data sets, the initial data normally comprises several tens of basic criteria. In general, the approach is to still construct H with only a few (say a couple) of criteria, leaving majority of criteria to be incorporated within G. The constructed G is then a new candidate for decomposition, and the process is repeated until all the utility functions in the evolving criteria tree are sufficiently simple. The construction of utility functions for

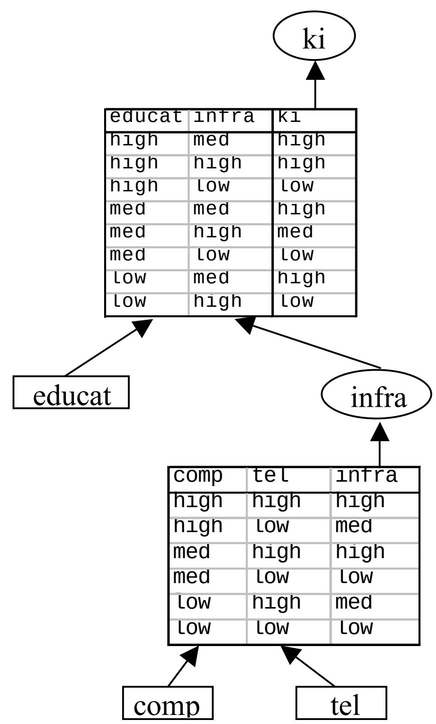

Figure 2: The resulting decision model for data from Table 1 and criteria tree from Figure 1

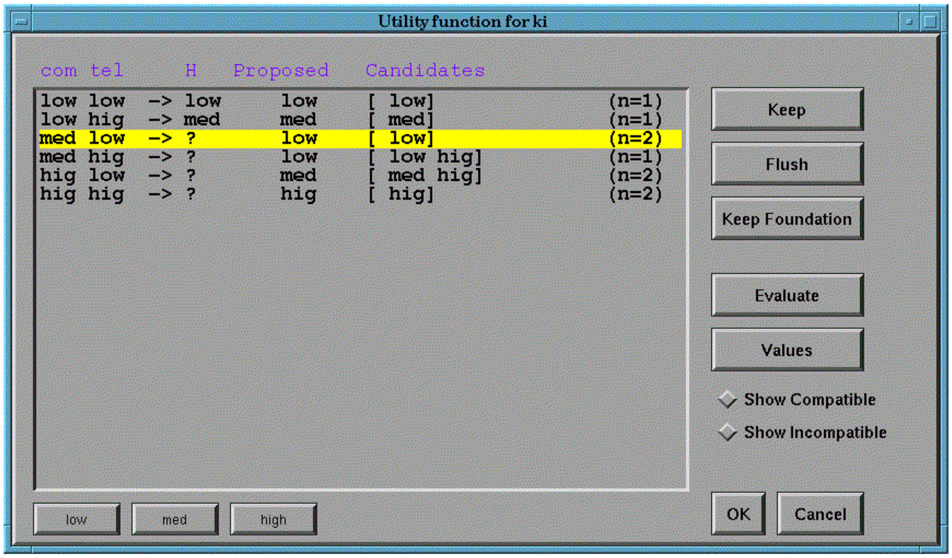

Figure 3: Interactive utility function discovery by HINT

desired criteria tree can therefore be regarded as *divide-and-conquer* technique that, starting from initial complex utility function that comprises no aggregated criteria, constructs a set of smaller, less complex and easier to understand utility functions.

3 SYNTHESIS OF SOCIOECONOMIC MODELS

We have used the socioeconomic data from The World Competitiveness Yearbook 1997 to construct different socioeconomic models using HINT. For each model, an expert chose a set of appropriate criteria from the Yearbook. Due to space limitation we here only present the synthesis of knowledge infrastructure model.

The data used comprised 53 countries and 12 basic criteria:

`inf_req`	Infrastructure requirements
`telecomm`	Telecommunications
`invest`	State investment in telecommunications
`teleph`	Telephones
`tel_cost`	International telephone costs
`c_use`	Computers in use
`c_pc`	Computers per capita
`manag_it`	Management and information technology
`it`	Information technology
`educ_sys`	The educational system
`c_lit`	Computer literacy
`training`	In-company training

First, the data was discretized, so that each criteria used three qualitative values only. Next, a standard hierarchical clustering algorithm from the S-Plus 4.0 program was used to identify four specific clusters of countries. These four groups were found to be also the one with different quality of knowledge infrastructure – in this way, each country obtained a specific value of the overall utility ranging from 1 for countries in the cluster that includes Russia to 4 for the cluster that includes USA (see Figure 4).

Resulting data that comprised 12 basic criteria and associated utility for knowledge infrastructure was then used by HINT. As the knowledge infrastructure model includes 12 criteria, the full coverage of original utility function space would require $3^{12} = 531.441$, which makes the coverage by only 53 instances very sparse and the interactive decomposition method the only viable alternative to automatic decomposition approach.

The criteria tree for knowledge infrastructure model defined by the expert that uses additional 8 aggregated criteria is shown in Figure 5. The decomposition of initial utility function started with criteria for utility functions H that were most familiar to the expert, beginning with that of telecommunication infrastructure and the utility and cost of telephones. It was observed that in the first couple of decomposition steps HINT provided only a minor support for the construction of utility functions – a effect contributed by the very sparse partition matrices. But HINT's support improved with every decomposition step, as constructed functions G covered more and more problem space and included fewer and fewer criteria. The expert especially appreciated HINT's implicit suggestions which value should not be assigned to a specific instance of utility function being interactively constructed. For the last few utility functions, HINT's help in identifying the number of required values for aggregated criteria was also found to be of high value. The complete were all required utility functions were identified lasted two hours and the resulting model was in consistence with the initial set of 53 data instances.

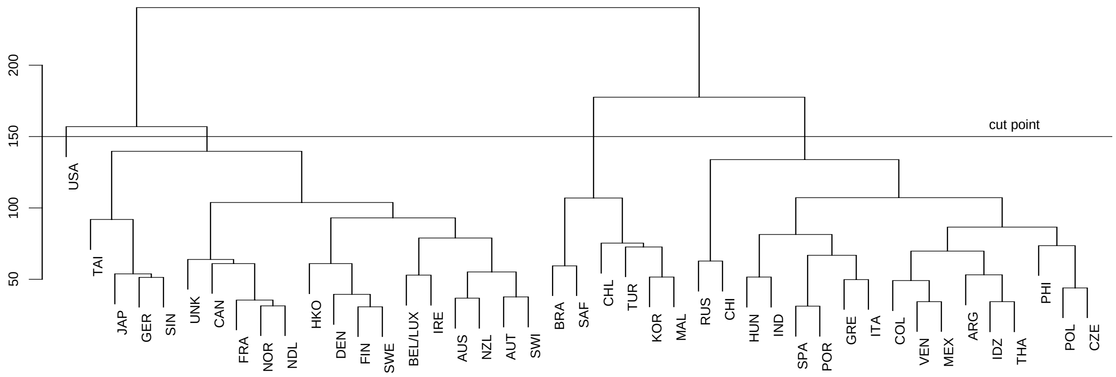

Figure 4: The results of hierarchical clustering of 53 countries

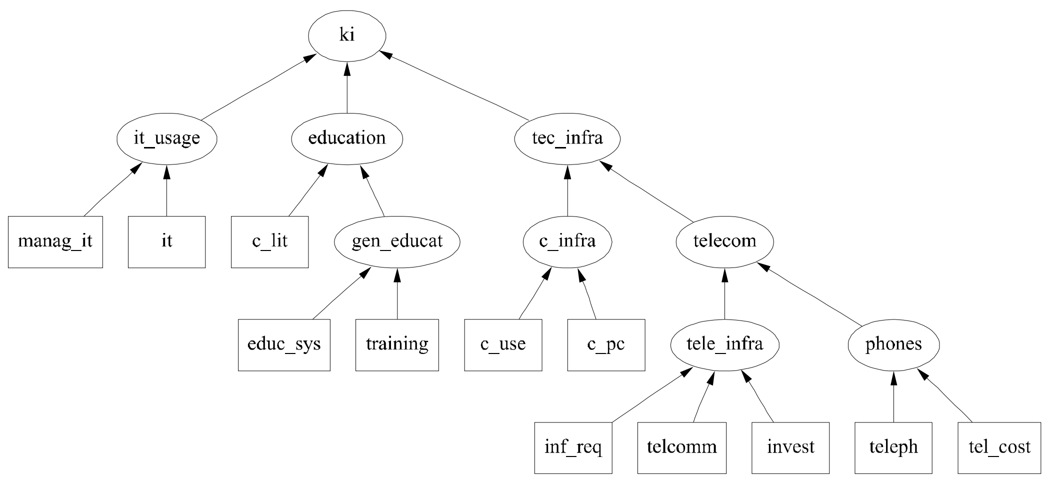

Figure 5: Utility tree for knowledge infrastructure model

4 CONCLUSION AND FURTHER WORK

The case study described above confirmed that HINT is a useful tool that can beneficially be used for construction of hierarchical decision support models. Further research and project work will be devoted to construction of several socioeconomic decision support models, each for its specific criteria group. An expert will assist in model synthesis and model verification. After the completion of this phase, the models will be further integrated into a higher level hierarchical decision models. Furthermore, the models will be used in the portfolio analysis. These stages will be concluded by an application oriented experimental studies.

References

Ashenhurst, R. L. (1952) The Decomposition of Switching Functions, Bell Laboratories BL-1(11), pages 541--602. Also reprinted in (Curtis, 1962).

Bohanec, M., Rajkovič, V. (1990) DEX: An expert system shell for decision support. *Sistemica*, **1**, pp. 145–157.

Chankong, V., and Haimes, Y. Y. (1983) *Multiattribute decision making: Theory and methodology*, North-Holland.

Curtis, H. A. (1962) *A New Approach to the Design of Switching Functions*. Van Nostrand, Princeton, N.J.

Efstathiou, J., and V. Rajkovič (1979) Multiattribute decision making using a fuzzy heuristic approach. *IEEE Trans. on Systems, Man and Cybernetics*, **9**, 326–333.

Zupan, B. (1997*) Machine learning by function decomposition*. Dissertation. University of Ljubljana.

Zupan, B., Bohanec, M., Demsar, J., Bratko, I. (1998) Feature transformation by function decomposition. *IEEE Intelligent Systems*, 13(2): 38-43, March/April 1998.