Optimal Decision Explanation by Extracting Regularity Patterns

Concha Bielza
Juan A. Fernández del Pozo
Decision Analysis Group, Technical University of Madrid
Campus de Montegancedo, Boadilla del Monte, 28660 Madrid, Spain

Peter Lucas
Institute for Computing and Information Sciences, University of Nijmegen
Toernooiveld 1, 6525 ED Nijmegen, The Netherlands

Abstract

When solving decision-making problems with modern graphical models like influence diagrams, we obtain the decision tables with optimal decision alternatives. For real-life clinical problems, these tables are often extremely large, hindering the understanding of the reasons behind their content. *KBM2L* lists are new structures that simultaneously minimise memory storage space of these tables, and search for a better knowledge organisation. In this paper, we study the application of *KBM2L* lists in finding and thoroughly studying the optimal treatments for gastric non-Hodgkin lymphoma. This is a difficult clinical problem, mainly because of the uncertainties involved. The resultant lists provide high-level explanations of optimal treatments for the disease, and are also able to find relationships between groups of variables and treatments.

1 Introduction

An *influence diagram* (ID) is a modern decision-theoretic formalism, nowadays frequently adopted as a basis for the construction of decision-support systems (DSS), and used to structure and solve decision-making problems [1]. It consists of a graph structure, which is directed and acyclic, and probabilities and utilities modelling the uncertainties and preferences that come with the problem concerned. The result of solving, or evaluating as is the technical term, an ID are *decision tables* containing the optimal decision alternatives. Thus, for every decision, there is an associated decision table with the best alternative, i.e. the alternative with the maximum expected utility, for every combination of variables. The evaluation algorithm determines which these variables are.

For some medical problems, usually problems that involve difficult trade-offs between benefits and risks of a treatment, doctors may use decision tables to determine the best recommendations in treating patients. However, medical doctors may find it difficult to accept such recommendations if they do not understand the reasons behind the table content. Questions a medical doctor is likely to ask are then: "why is the proposed decision optimal?", and "what are the implicit rules underlying the modelled decision problem?" Answering these

questions can be seen as providing *explanations* to medical decisions; these may provide new insights into the problem, and, as a knowledge synthesis, may also serve to validate a system.

Considering that the table sizes are exponential in the number of variables, finding explanations is a hard task, also from a purely computational viewpoint. Turning the huge tables into more compact tables will bring out memory savings. If the resulting compact tables offer insight into the original tables, then finding explanations and optimising the storage space of the decision tables are to some extent the same problem.

In [2], we introduced *KBM2L* lists to address this problem. Finding explanations is a goal pursued by a number of disciplines such as knowledge-based systems and machine learning. Thus, our approach bears some resemblance to techniques for knowledge extraction such as used in the construction of tree-based classifiers [3], oblivious read-once decision graphs [4], and rough sets [5], and to the identification of relevant nodes for each decision node in an ID [6, 7]. As explained in detail in [2], the *KBM2L* method tries to reorganise a knowledge structure by a global search for good, representative candidates. At the start of the algorithm there are already correctly classified cases represented in table form, which may be interpreted as representing types of patients, and we try to extract reasons underlying this classification. Unlike the typical situation in machine learning, these cases are unique as they correspond to configurations of variables in the original ID. Note that our method is applied after the ID has been solved, whereas most of the work on the ID framework concerns operations on the graph structure before evaluating the ID [6].

In order to investigate the usefulness of this method within a medical setting, an ID regarding the treatment of *non-Hodgkin lymphoma of the stomach*, abbreviated to *gastric NHL*, previously developed by one of the authors in collaboration with expert clinical oncologists, was chosen as an experimental vehicle [8]. This is a realistic clinical model, reflecting the current scientific evidence in medical literature about this disorder. It can be used to determine the optimal treatment for individual patients with gastric NHL. A treatment consists of the prescription of either or not antibiotics, followed by curative or palliative surgery, or no surgery at all, and finally by a combination of chemotherapy, radiotherapy, or none of these two. The probabilistic part of the model, a Bayesian network, can be used independently to predict prognosis and to generate patient-specific risk profiles. The model exceeds common prognostic models based on logistic regression, as it is part of a DSS that can answer many different clinical questions.

In this paper, we analyse this ID with the aim of extracting patterns from it; these are then used to explain the optimal treatment alternatives that can be generated by evaluating the diagram. The paper is organised as follows. Section 2 reviews the technique of *KBM2L lists*. Section 3 summarises the structure and content of the gastric NHL model. Section 4 discusses the results of applying *KBM2L* lists to the NHL model. Section 5 finishes off with some conclusions. This paper is an extended version of [9].