Data & Knowledge Engineering 70 (2011) 167–181

Contents lists available at ScienceDirect

Data & Knowledge Engineering

journal homepage: www.elsevier.com/locate/datak

Dealing with complex queries in decision-support systems

J.A. Fernández del Pozo, C. Bielza *

Departamento de Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid, Boadilla del Monte, Madrid 28660, Spain

ARTICLE INFO

Article history:
Received 19 March 2009
Received in revised form 8 October 2010
Accepted 8 October 2010
Available online 30 October 2010

Keywords:
Decision-support systems
Query system
Information systems
Explanations
Influence diagrams
Medicine

ABSTRACT

In decision-making problems under uncertainty, a decision table consists of a set of attributes indicating what is the optimal decision (response) within the different scenarios defined by the attributes. We recently introduced a method to give explanations of these responses. In this paper, the method is extended. To do this, it is combined with a query system to answer expert questions about the preferred action for a given instantiation of decision table attributes. The main difficulty is to accurately answer queries associated with incomplete instantiations. Incomplete instantiations are the result of the evaluation of a partial model outputting decision tables that only include a subset of the whole problem, leading to uncertain responses. Our proposal establishes an automatic and interactive dialogue between the decision-support system and the expert to elicit information from the expert to reduce uncertainty. Typically, the process involves learning a Bayesian network structure from a relevant part of the decision table and computing some interesting conditional probabilities that are revised accordingly.

1. Introduction

1.1. Decision tables, explanations and queries

Under uncertainty, a modern and useful decision-theoretic model is the *influence diagram* [17]. It consists of an acyclic directed graph with associated probabilities and utilities, respectively modeling the uncertainties and preferences tied in with the stated problem. Nowadays this probabilistic graphical model is frequently adopted as a basis for constructing decision-support systems (DSSs). The results of evaluating an influence diagram are *decision tables* containing the optimal decision alternatives, policies or responses. 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 relevant variables (usually called *attributes* within this context) that are observable before the decision is made. The evaluation algorithm determines which of the observable variables are relevant. These variables are outcomes of random variables and/or other past decisions.

A decision table may have millions of rows and typically more than twenty columns leading to enormous data sets for storage and analysis. Expert DSS users demand such an analysis on mainly two grounds. First, DSS decision tables provide the best decision-making recommendations. However, experts may find such recommendations hard to accept if they come without any *explanation* whatsoever of why the proposed decisions are optimal. Unexplained responses are not good enough for expert users since DSSs operate on a model that is an approximation of the real world. The importance of explanations has been reported in the literature, see e.g. [9,12,13]. Thus, for example, in health-care problems, usually involving difficult trade-offs between the treatment benefits and risks, practitioners may use decision tables to determine the best patient treatment recommendations. For this purpose, they need to understand the underlying reasons or implicit rules.

In medical DSSs, clinical practice guidelines assemble the relevant knowledge gathered through literature review, meta-analysis, expert consensus, etc., and operationalize this information as informal, text documents. This makes the gathered

* Corresponding author.
E-mail addresses: jafernandez@fi.upm.es (J.A. Fernández del Pozo), mcbielza@fi.upm.es (C. Bielza).

0169-023X/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.datak.2010.10.006

information difficult to interpret automatically and the decision-making process hard to guide. Shiffman and Greenes [19] propose translating guideline knowledge into decision table-based rule sets. Shiffman [18] proposes augmenting decision tables by layers, storing collateral information in slots at various levels beneath the logic layer of the conventional decision table. Information relates to table cells, rows and columns. It may include how tests are performed, the benefits/risks of the recommended strategies, costs, literature citations, etc., to help understand the domain. All these decision tables are different than ours. Our knowledge base is the model (influence diagram) and its evaluation, stored in the decision tables. The model (graph with probabilistic dependencies and probability and utility information) is built from clinical practice guidelines, data and expert input. Also, there is no uncertainty in clinical guidelines. Influence diagrams are based on subjective probabilities and utilities, and support learning and reasoning with uncertainty and preferences.

In [6] we introduced *KBM2L* lists to find explanations. The main idea stems from how computers manage multidimensional matrices: computer memory stores and manages these matrices as linear arrays, and each position is a function of the order chosen for the matrix dimensions. KBM2L lists are new list-based structures that optimize this order by putting equal responses in consecutive positions, yielding the target explanations and simultaneously achieving compact storage. These lists implicitly include the probability and utility models, they are simple, and have no added complex layers.

Not only do expert users employ decision tables as a knowledge base (KB) for explanations; they also *query* the DSS about which is the best recommendation for a given set of attributes in different ways. This is the second reason for decision table analysis. In a typical session, experts interact with DSSs to:

(A) formulate a query in the KB domain;
(B) translate the query into the KB formalism;
(C) implement the response retrieval;
(D) build the response efficiently;
(E) communicate the response(s) and/or suggest improvements, and wait for user feedback.

For (A) and (B), we distinguish between two groups of queries (closed/open) depending on whether or not the whole set of attributes is instantiated. A closed query is a specific and well-defined query entered by users that know all the attribute information. An open query is less specific, as it includes attribute values that are undefined either because they are hard or expensive to obtain or they are unreliable. Martinez et al. [15] give a similar classification for GIS (geographical information systems), although they focus on data efficient updating and access from a physical point of view (merely as a database), rather than from a logical point of view (as a KB).

(C) to (E) may be troublesome, especially for open queries, due to imprecise response retrieval failing to satisfy users. Additionally, the DSS may not include the whole decision table, because an exhaustive evaluation of the decision-making problem can be too costly. In this case there will be no response at all. Worse still, both situations could apply at the same time, demanding a methodology to undertake tasks (C)–(E) dealing with ambiguity and ignorance about the response.

1.2. Example: Optimal treatment of gastric non-Hodgkin lymphoma

Let us illustrate these ideas with the following clinical problem. It is a real health-care decision-making problem regarding the optimal treatment of non-Hodgkin lymphoma of the stomach.

Primary gastric non-Hodgkin lymphoma, *gastric NHL* for short, is a relatively rare disorder, accounting for about 5% of gastric tumors. This disorder is caused by a chronic infection by the *Helicobacter pylori* bacterium [5]. Treatment consists of a combination of antibiotics, chemotherapy, radiotherapy and surgery.

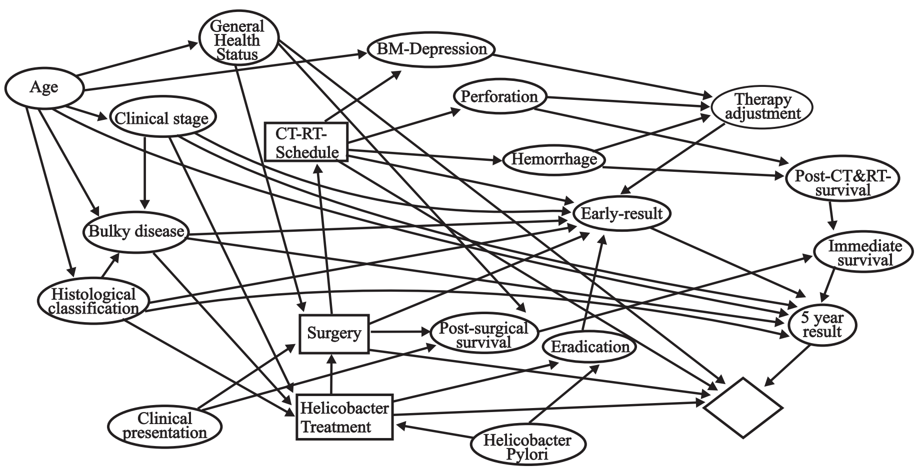

Fig. 1. Influence diagram for the treatment of gastric NHL.