Intelligent Information Systems VII
Proceedings of the Workshop held in
Malbork, Poland, June 15-19 , 1998

A Probabilistic Causal Model for Diagnosis of Liver Disorders

Agnieszka Oniśko[1], Marek J. Druzdzel[2] and Hanna Wasyluk[3]

[1]Institute of Computer Science, Bialystok University of Technology, Bialystok, 15–351, Poland, *aonisko@ii.pb.bialystok.pl*

[2]Decision Systems Laboratory, School of Information Sciences and Intelligent Systems Program, University of Pittsburgh, Pittsburgh, PA 15260, U.S.A., *marek@sis.pitt.edu*

[3]Medical Center of Postgraduate Education, Warsaw, Marymoncka 99, Poland

Abstract. Directed probabilistic graphs, such as Bayesian networks, are useful tools for coherent representation of and reasoning with uncertain knowledge. They are based on the sound foundations of probability theory and they readily combine available statistics with expert judgment. When extended with decision options and measures of desirability of outcomes (utilities), they support decision making. This paper describes our work in progress on a probabilistic causal model for diagnosis of liver disorders that we plan to apply in both clinical practice and medical training. The model, and especially its numerical parameters, is based on patient records at the Gastroenterological Clinic of the Institute of Food and Feeding in Warsaw, collected over the period of several years. We present the model and report initial results of our diagnostic performance tests.

Keywords: Causal models, Bayesian networks, medical diagnosis

1 Introduction

Some of the earliest Artificial Intelligence (AI) approaches to medical diagnosis were based on Bayesian and decision-theoretic schemes. Difficulties in obtaining and representing quantities of numbers and both the computational and representational complexity of probabilistic schemes caused a long-lasting departure from these approaches. Only recently, development of probabilistic graphical models, such as Bayesian networks [5] and

closely related influence diagrams, has caused a renewed interest in applying probability theory in intelligent systems (see [3] for an accessible overview of decision-analytic methods in AI). Today, Bayesian networks are successfully applied to a variety of problems, including machine diagnosis, user interfaces, natural language interpretation, planning, vision, robotics, data mining, and many others (for examples of successful real world applications of Bayesian networks, see March 1995 special issue of the *Communications of ACM*).

In this paper, we describe our work in progress on a probabilistic causal model for diagnosis of liver disorders. Our work is continuation of the HEPAR project [1], conducted in the Institute of Biocybernetics and Biomedical Engineering of the Polish Academy of Sciences in co-operation with physicians at the Medical Center of Postgraduate Education. The HEPAR system contains a database of patient records of the Gastroenterological Clinic of the Institute of Food and Feeding in Warsaw. This database is thoroughly maintained and enlarged with new cases. The system is currently used in the clinic as a diagnostic and training aid. Our model is essentially a Bayesian network modeling causal relations among its variables with its numerical parameters extracted from the HEPAR database. One application of our model, in addition to its diagnostic value, is in training physicians. We present the model and report the initial results of our diagnostic performance tests.

2 Bayesian Networks

A Bayesian network [5] (also referred to as *Bayesian belief network*, *belief network*, *probabilistic network*, or *causal network*) consists of a qualitative part, encoding existence of probabilistic influences among a domain's variables in a directed graph, and a quantitative part, encoding the joint probability distribution over these variables.

Each node of the graph represents a random variable and each arc represents a direct dependence between two variables. Formally, the structure of the directed graph is a representation of a factorization of the joint probability distribution. As many factorizations are possible, there are many graphs that are capable of encoding the same joint probability distribution. Of these, those that minimize the number of arcs are preferred. From the point of view of knowledge engineering, graphs that reflect the causal structure of the domain are especially convenient — they normally reflect expert's understanding of the domain, enhance interaction with a human expert at the model building stage and are readily extendible with new information. Finally, causal models facilitate user insight once a model is employed. This is important in all those systems that aid decisions and fulfill in part a training role, like most diagnostic systems. It is fair to say that Bayesian networks are a high-level language for structuring uncertain knowledge. Their clear semantics allows for them to be used in a variety of tasks in intelligent systems.

Quantification of a Bayesian network consists of prior probability distributions over those variables that have no predecessors in the network and conditional probability distributions over those variables that have predecessors. These probabilities can easily incorporate available statistics and, where no data are available, expert judgment. A probabilistic graph represents explicitly independences among model variables and allows

for representing a full joint probability distribution by a fraction of numbers that would be required if no independences were known. Every independence leads to omitting an arc from the graph and leads to significant reductions of the numbers needed to fully quantify the domain. It should be stressed here that Bayesian networks are capable of representing any independences, not only those assumed to exist in early Bayesian systems. In particular, a domain where no independences exist, will be represented correctly by a Bayesian network that is a complete graph.

The most important type of reasoning in Bayesian networks is known as *belief updating*, and amounts to computing the probability distribution over variables of interest conditional on other, observed variables. In other words, the probability distribution over the model variables is adjusted for a particular case, in which some of the model variables assume given values. While belief updating in Bayesian networks is in the worst case NP-hard [2], there are several very efficient algorithms capable of updating beliefs in networks on the order of hundreds of variables within seconds (this depends strongly on the topology of the network — roughly speaking, the sparser a network, the shorter it takes to update).

3 The Diagnostic Model

The starting point for building our model has been HEPAR's database of patient cases. The database available to us included 570 patient records, each of these records was described by 119 features (binary, denoting presence or absence of a feature or continuous, expressing the value of a feature) and each record belonged to one of 16 liver disorders. One limitation of the HEPAR database is the assumption that a patient appearing in the clinic has at most one disorder. The features can be divided conceptually into three groups: symptoms and findings volunteered by the patient, objective evidence observed by the physician, and results of laboratory tests.

3.1 Model Structure

In our initial effort, we have reduced the number of features from the 119 encoded in the database to 40. We started by eliminating those features that had many missing values — numerical parameters expressing relevance of these features to the diagnosis would not be too reliable. Then we relied on expert's opinion as to which features have the highest diagnostic value. Having selected the total of 40 features, we elicited the structure of dependences among them from our domain experts: Dr. Hanna Wasyluk (third author) of the Medical Center of Postgraduate Education and Dr. Daniel Schwartz, a pathologist at the University of Pittsburgh. Our current network is comprised of 41 variables: the disorder variable with 16 outcomes and 40 feature variables.

The structure of our current model is shown in Figure 1. We believe that it models reasonably causal interactions among the selected variables.

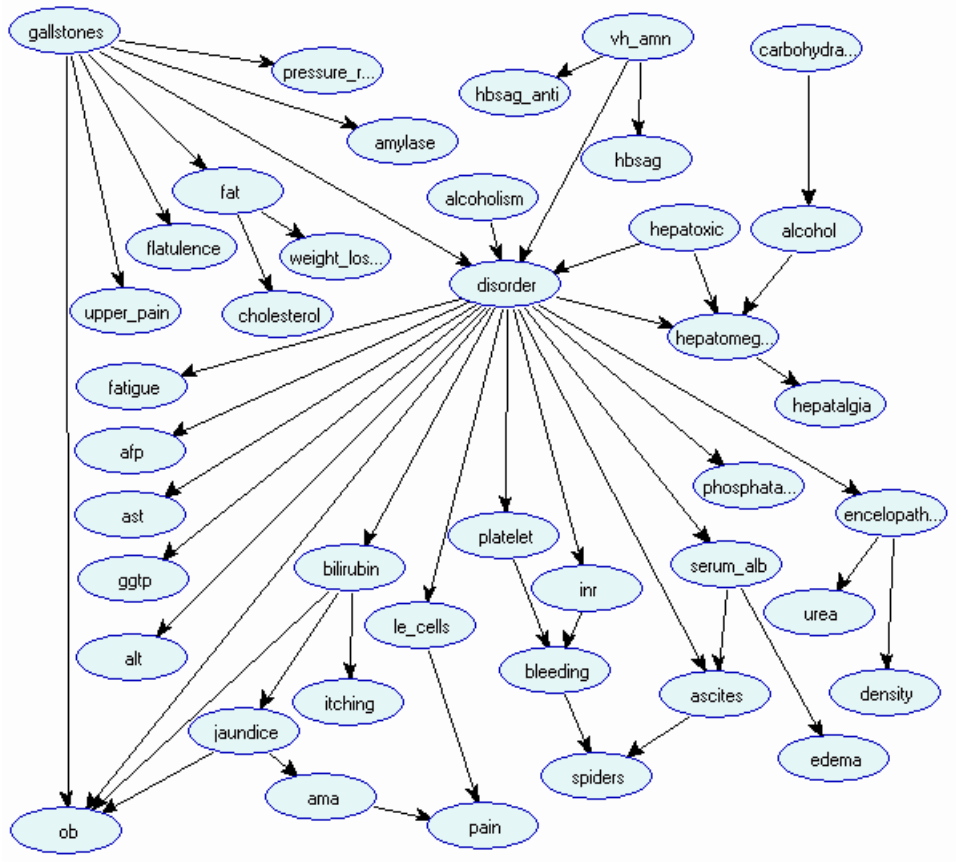

Fig. 1. The structure of the model.

3.2 Model Parameters

While the underlying formalism of Bayesian networks allows both discrete and continuous variables, all general purpose exact algorithms for Bayesian networks deal with models containing only discrete variables. In order to take advantage of these algorithms, we decided to discretize continuous variables. Our discretization is based on expert opinion that variables such as urea, bilirubin, or blood sugar have essentially *low*, *normal*, *high*, and *very high* values. The numerical boundaries of these intervals are based on expert judgment.

Given a structure of the model, the specification of the desired discretization, and the HEPAR database, our program learns the parameters of the network, i.e., prior probabilities of all nodes without predecessors and conditional probabilities of all nodes with predecessors, conditional on these predecessors. Prior probability distributions are

simply relative counts of various outcomes for each of the variables in question. Conditional probability distributions are relative counts of various outcomes in those data records that fulfill the conditions described by every combination of the outcomes of the predecessors.

We would like to make two remarks here. The first is that the HEPAR database contains many missing measurements. We interpreted the missing measurements as possible values of the variables in question. This interpretation requires some care when using our system. We assume namely that the fact that a measurement was not taken is meaningful — the physician did not find taking the measurement appropriate. The meaning of the thus construed outcome *unmeasured* is in this way equivalent to a measured value of the variable.

The second remark concerns the accuracy of the learned parameters. While prior probabilities can be learned reasonably accurately from a database of 570 records, conditional probabilities present more of a challenge. In cases where there are several variables directly preceding a variable in question, individual combinations of their values may be very unlikely to the point of being absent from the data file. In such cases, we assume that the distribution is uniform. In all cases where the counts were zero, and naively interpreted would suggest a zero probability, we inserted a small probability reflecting the fact that almost no probabilities in the domain of medicine are zero or one. We found empirically that a value around 0.1 led to the best diagnostic performance.

Generally, conditional probabilities learned from a data file of this size are not very reliable and need to be verified by an expert. There is much anecdotal and some empirical evidence that imprecision in probabilities has only small impact on the diagnostic accuracy of a system based on a Bayesian network [6]. This remains to be tested in our system.

4 Diagnostic Performance

To evaluate the classification accuracy of our model we performed a standard test in which we used a fraction of the database to learn the network parameters and the remainder of the records to test the network prediction. The results are shown in Table 1. In over 36% of the cases, the most likely disorder indicated the correct diagnosis. In over 74% of the cases, the correct diagnosis was among the first four most likely disorders, as indicated by our model.

The results of the leave-one-out approach [4], i.e., using repeatedly all but one records in the database to learn the parameters and then using the remaining one to test the prediction are presented in Table 2. Here, the performance results are similar.

We would like to note that a naive-Bayes network structure led to a performance of over 92% in terms of the correct diagnosis being among the first four most likely disorders. Still, we prefer a causal model as a diagnostic tool. We believe that pure diagnostic performance, in terms of the percentage of correct diagnoses, is in itself not an adequate measure of quality of a medical decision support system. In the domain of medicine, the physician user carries the ultimate responsibility for the patient and he or she will be unwilling to accept a system's advice without understanding it. While a

Window size	Accuracy	σ
1	36.83%	4.82
2	52.50%	5.00
3	58.33%	4.47
4	74.13%	4.24

Table 1. Summary test results for the diagnostic accuracy test. Window size expresses the number of most likely disorders among which we searched for the correct diagnosis. The number of test records, $n = 100$. The tests were repeated 30 times.

Window size	Accuracy
1	36.20%
2	58.25%
3	61.00%
4	72.00%

Table 2. Summary test results for the diagnostic accuracy test using the leave-one-out approach. Window size expresses the number of most likely disorders among which we searched for the correct diagnosis.

causal model may perform worse in numerical terms than a regression-based model,[1] it offers three important advantages: (1) its intuitive and meaningful graphical structure can be examined by the user, (2) the system can automatically generate explanations of its advice that will follow the model structure and will be reasonably understandable, and (3) the model can be easily enhanced with expert opinion; interactions absent from the database can be added based on knowledge of local causal interactions with the existing parts and can be parameterized by expert judgment. In an interaction with the system, the physician needs to be reminded of the most likely candidate diagnoses and we believe that a performance of over 70% in terms of four top-runners is acceptable for a first cut on the model. We believe that the model performance can be improved in the future.

5 A Diagnostic Tool Based on Our Model

The probabilistic approach based on Bayesian networks allows to query the system with partial observations, something that is not natural for classification systems. To test how intuitive and how useful the model is in practice, we have built a simple user interface to our model that lists all observable features, allows the user to set the values of any of them, updates the probability distribution over different disorders and presents an ordered list of possible diagnoses with their probabilities.

[1] We would like to point out that this parallels the historical fact that Copernican theory of the structure of the Solar System, when introduced, predicted the positions of planets with less precision than the existing Sun-centered theory due to Ptolemy. Still, Copernican theory is preferable because it reflects better the physical structure of the system and explains the movements of planets with fewer free parameters.

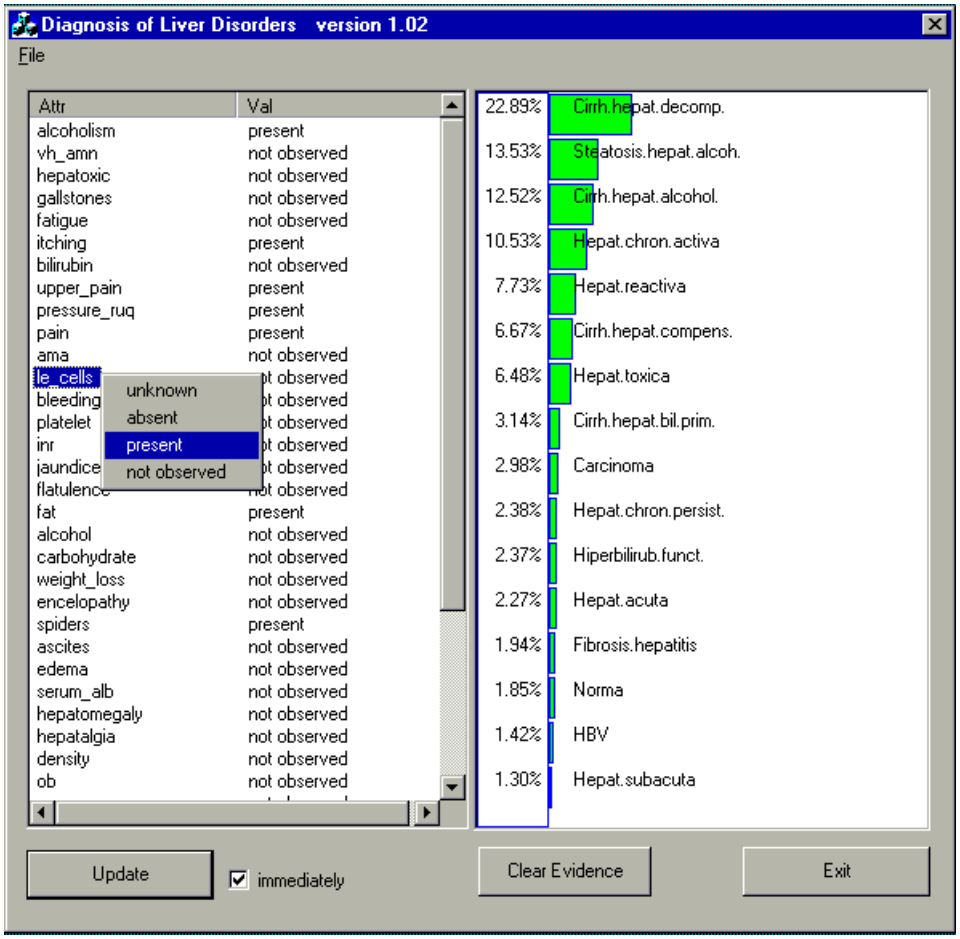

Fig. 2. The front end interface to to our model.

Figure 2 shows a screen shot of this user interface. The left column contains a complete list of all features included in the model. Right-clicking on any of the features brings up a pop-up menu that lists all possible values of the selected variable. By choosing one of these values the user can enter a finding. Please note, that each variable has a possible value defined as *unknown*. We use this value to model the situation common in the data file that there is no information available about the value of that feature (a missing value). *Not observed* denotes simply that the feature has not yet been observed. The right column presents an ordered list of the possible diagnoses along with their associated probabilities, the latter being presented graphically. The probabilities can be updated immediately after entering each finding (the default) or only after the *Update* button is pressed. The

latter feature is meant for models that take longer time to update. Belief updating and presenting a newly ordered list of possible disorders takes in the current version of the model a fraction of a second and is from the point of view of the user instantenous. Our program has been welcomed as a useful interactive diagnostic and training tool by our colleague physicians.

6 Conclusion

We described a probabilistic causal model for diagnosis of liver disorders. The model includes 16 liver disorders and 40 features, such as important symptoms and risk factors. Given a patient's case, i.e., observation of values of any subset of the 40 features, the model computes the posterior probability distribution over the possible 16 liver disorders. This probability can be directly used in diagnostic decisions.

We would like to remark that the model output, probability distribution over the possible disorders, is something that internists are used to and know how to interpret. Since our model follows reasonably the causal structure of the domain, and its output has a sound and unambiguous meaning, we hope that in addition to its value as a diagnostic aid, it will be useful in training beginning diagnosticians.

The current model is our first attempt at capturing the interactions among most essential variables in the domain of liver disorders. Decision analytic approach is usually far from being a one-shot process in which a model is build and used. Usually, the initial model is refined iteratively in which methods such as sensitivity analysis indicate those parts of the model that need further refinement.

In the long run, we plan to enhance our model with an explicit representation of diagnostic decisions and utilities of correct and incorrect diagnoses. This will make our model sensitive to possibly high disutility of missing major disorders that require immediate attention, such as liver carcinomas.

Acknowledgments

The authors gratefully acknowledge financial support from the Committee for Scientific Research under grant number 8T11E00811. In addition, Dr. Druzdzel was supported by the National Science Foundation under Faculty Early Career Development (CAREER) Program, grant IRI–9624629, and by the Air Force Office of Scientific Research under grant number F49620–97–1–0225 to University of Pittsburgh. We would like to thank Dr. Daniel Schwartz for his help in building the structure of our model and Tomek Sowiński for his assistance in programming.

References

1. Leon Bobrowski. HEPAR: Computer system for diagnosis support and data analysis. Prace IBIB 31, Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences, Warsaw, Poland, 1992.

2. Gregory F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42(2–3):393–405, March 1990.
3. Max Henrion, John S. Breese, and Eric J. Horvitz. Decision Analysis and Expert Systems. *AI Magazine*, 12(4):64–91, Winter 1991.
4. A.W. Moore and M.S. Lee. Efficient algorithms for minimizing cross validation error. In *Proceedings of the 11th International Conference on Machine Learning*, San Francisco, 1994. Morgan Kaufmann.
5. Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, Inc., San Mateo, CA, 1988.
6. Malcolm Pradhan, Max Henrion, Gregory Provan, Brendan Del Favero, and Kurt Huang. The sensitivity of belief networks to imprecise probabilities: An experimental investigation. *Artificial Intelligence*, 85(1–2):363–397, August 1996.