In *Proceedings of the Third Conference on Neural Networks and Their Applications*, pages 730-736, Kule, Poland, October 14-18, 1997

Application of Bayesian Belief Networks to Diagnosis of Liver Disorders

Agnieszka Onisko
Bialystok University of Technology
Institute of Computer Science
Bialystok, 15–351, Poland
aonisko@ii.pb.bialystok.pl
FAX: (085) 422–393

Marek J. Druzdzel
University of Pittsburgh
School of Information Sciences
and Intelligent Systems Program
Pittsburgh, PA 15260, U.S.A.
marek@sis.pitt.edu
FAX: +1–412–624–2788

Hanna Wasyluk
The Medical Center of Postgraduate Education
Warsaw, Marymoncka 99, Poland
FAX: (022) 340–470

Abstract

Probabilistic graphs, such as Bayesian belief networks, are useful tools for coherent representation of uncertain knowledge. They are based on the sound foundations of probability theory and they readily combine available frequency data with expert assessments. When extended with measures of desirability of outcomes, utilities, they support decision making. This paper describes our work in progress on application of Bayesian belief networks to diagnosis of liver disorders. We discuss our initial model and how it was constructed, including both its structure and parametrization.
Keywords: Bayesian belief networks, 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 fairly recently, development of probabilistic graphical models, such as Bayesian belief networks and closely related influence diagrams, has caused a renewed interest in applying probability theory in intelligent systems (see [6] for an accessible overview of decision-analytic methods in AI). Today, Bayesian belief 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 application of Bayesian networks to diagnosis of liver disorders. We discuss different opportunities and difficulties presented by availability of a rich data set of past patient cases. Our work so far focused on creating a simple diagnostic model that we hope to refine and test by the time of the conference.

2 Bayesian Belief Networks

A Bayesian belief network (also referred to as *Bayesian network*, *belief network*, *probabilistic network* or *causal network*) consists of a qualitative part, encoding a domain's variables and the probabilistic influences among them in a directed graph, and a quantitative part, encoding the joint probability distribution over these variables.[1].

Each node of the graph represents a random variable and each arc 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 possible 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.

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 frequency data and, where these are not available, expert judgment. As a probabilistic graph represents explicitly independences among model variables, it allows for representing a full joint probability distribution by a fraction of numbers that would be required if no independences were known. 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. Every independence leads to omiting an arc from the graph and leads to significant reductions of the numbers needed to fully quantify the domain.

The most important type of reasoning in Bayesian networks is knows as belief updating, and amounts to computing the probability distribution over variables of interest conditional on other, observed variables. In other words, the probability

[1] The reader is recommended the books by Pearl [11], Neapolitan [10], or papers by Cooper [3], Charniak [2], or Matzkevich and Abramson [9] as general introductions to Bayesian belief networks.

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 [4], 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).

It is fair to say that Bayesian belief networks are a high-level language for structuring uncertain knowledge. An important advantage of probabilistic networks is their clear semantics, that allows for them to be used in a variety of tasks in intelligent systems.

3 The HEPAR Project

In this paper we analyse the medical database from the computer system HEPAR,[2] which was designed and build a few years ago in the Institute of Biocybernetics and Biomedical Engineering Polish Academy of Sciences in co-operation with doctors from the Medical Center of Postgraduate Education. The system comprises the hepatological patients from the Gastroenterological Clinic of the Institute of Food and Feeding in Warsaw. The HEPAR system is currently used in this Clinic and its database is being steadily enlarged.

4 A Bayesian Belief Network Model for Liver Disorders

The starting point for building our model was the HEPAR project and its database of patient cases.
This database includes the data from 570 patient records, each of these records is described by 119 features (binary or continuous) and each record belongs to one of 17th liver disorders. In this database we can distinguish a few feature groups: symptoms and findings volunteered by the patient, objective evidence observed by the physican and laboratory tests.
On account of a large number of the features, before we constructed our network model, we had conducted the data reduction, i.e. a feature selection and a discretization of the continuous variables. The reduction of the feature dimensionality usually yields the economic benefits, and at the same time lead to removing the redundant and irrelevant features.
We have selected a subset of features present in the data set to construct a Bayesian belief network model for diagnosis of liver disorders. The following sections describe the structure of our model, discretization of continuous variables, and learning the parameters of the model (prior and conditional probability distributions) from the HEPAR database.

[2]Hepar — computer system for diagnosis support and data analysis [1].

4.1 Network Structure

Among 119 features in the database we have selected 37 features. First of all we have eliminated the features with the missing values (or mostly missing). In our feature selection we have considered also the expert knowledge (our expert has indicated these variables, which have a significant, diagnostic weight), and the statistical dependencies between the variables. We have taken into consideration the dependencies between decision variable and the other features.
Our current network structure is comprised of 38 variables: the disorder variable with 17 outcomes and 37 other variables, including both symptoms and risk factors for the different disorders. Please note that similarly to the HEPAR project team we assumed that a patient appearing in the clinic has at most one disorder.

The structure of our current model is shown in Figure 1. It models reasonably interactions among the modeled variables.

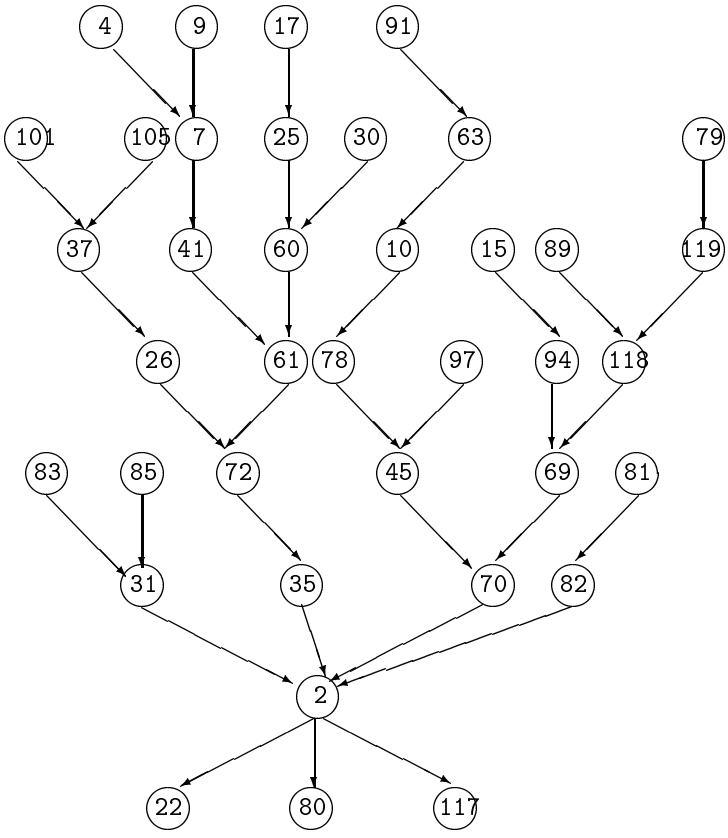

Figure 1: The structure of the model.

4.2 Discretization of Continuous Variables

While the underlying formalism of Bayesian networks allows both discrete and continuous variables, all known diagnostic algorithms for Bayesian networks deal with models containing only discrete variables. In order to take advantage of these algorithms, it is necessary 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. We call this discretization the *gold standard* discretization. Our program for learning the network parameters uses the discretization prepared in a data file that lists the intervals for each of the continuous variables used. This data file can contain *gold standard* discretization obtained from the expert but it can also be prepared with discretizations produced by various machine learning algorithms, such as those implemented in the machine learning library of Kohavi et al. [8]. In fact, we have prepared four different discretizations of all continuous parameters in the HEPAR database: (1) equal width intervals, (2) equal width intervals with S-plus histogram binding algorithm [13], (3) OneR discretizer [7], and (4) the entropy maximization heuristic [5]. We plan to test the effects of different discretizations on the accuracy of our model.

4.3 Learning the Probabilities

Given a structure of the model, the specification of the desired discretization, and the HEPAR data file, 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 case there are several variables directly preceeding a variable in question, individual combinations of their values may be very unlikely to the point of being absent in the data file. 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 ad some empirical evidence [12] that imprecision in probabilities has only small impact on the diagnostic accuracy of a Bayesian belief network system. This remains to be tested in our system.

5 Future Work

Our work is in its initial stage. 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. Our initial model contains several simplifying assumptions that reduced the number of expert elicitations and also allowed to learn the parameters from our data file more reliably. We plan to subject our model to several refinement rounds that will be culminated by a rigorous performance test that should be completed by the time of the conference.

One of the first steps will involve different discretizations of the continuous variables. The current discretization is rather rough and is based on common sense values for the parameters obtained from our expert. We plan to compare the diagnostic accuracy for different discretizations listed in Section 4.2 and our current *gold standard* discretization coming from our expert.

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 carcinomas.

6 Conclusion

We described a Bayesian belief network model for diagnosis of liver disorders. The model includes 17 liver disorders and 37 features, such as important symptoms and risk factors. Given a patient's case, i.e., observation of values of any subset of the 37 features, the model computes the posterior probability distribution over the possible 17 liver disorders. This probability can be directly used in diagnostic decisions. The research and the model described in this paper are in their early stages. The model, in particular, will be refined and tested rigorously on the available data by the time of the conference.

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 interprete. Since our model follows reasonably the 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.

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 Misook Heo for her assistance in coding data conversion and parameter learning functions described in this paper.

References

[1] Leon Bobrowski. Hepar – Computer System for Diagnosis Support and Data Analysis. *Prace IBIB*, 31, pp.27-48, 1992.

[2] Eugene Charniak. Bayesian networks without tears. *AI Magazine*, 12(4):50–63, Winter 1991.

[3] Gregory F. Cooper. Current research directions in the development of expert systems based on belief networks. *Applied Stochastic Models and Data Analysis*, 5(1):39–52, March 1989.

[4] Gregory F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42(2–3):393–405, March 1990.

[5] U.M. Fayyad and K.B. Irani. Multiinterval discretization of continuous-valued attributes for classification learning. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence, IJCAI–93*, pages 1022–1027. Morgan Kaufmann Publishers, Inc., San Mateo, CA, 1993.

[6] Max Henrion, John S. Breese, and Eric J. Horvitz. Decision Analysis and Expert Systems. *AI Magazine*, 12(4):64–91, Winter 1991.

[7] R.C. Holte. Very simple classification rules perform well on most commonly used datasets. *Machine Learning*, 11:63–90, 1993.

[8] R. Kohavi, G. John, R. Long, D. Manley, and K. Pfleger. MLC++: A machine learning library in C++. In *Tools with Artificial Intelligence*, pages 740–743. IEEE Computer Society Press, http://www.sgi.com/Technology/mlc 1994.

[9] Izhar Matzkevich and Bruce Abramson. Decision analytic networks in artificial intelligence. *Management Science*, 41(1):1–22, January 1995.

[10] Richard E. Neapolitan. *Probabilistic Reasoning in Expert Systems: Theory and Algorithms.* John Wiley & Sons, New York, 1990.

[11] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers, Inc., San Mateo, CA, 1988.

[12] 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.

[13] P. Spector. *An Introduction to S and S-Plus.* Duxbury Press, 1994.