ZESZYTY NAUKOWE POLITECHNIKI BIAŁOSTOCKIEJ 2007
Informatyka – Zeszyt 2

Agnieszka Oniśko[1]

KNOWLEDGE ACQUISITION FROM HUMAN EXPERTS FOR BUILDING BAYESIAN NETWORK MODELS

Abstract: Knowledge acquisition from experts is a costly and time-consuming task. While domain experts have the necessary knowledge and expertise, they rarely have the experience needed to translate this knowledge into the model. This paper describes typical problems that are encountered by knowledge engineers when building Bayesian network models and illustrates some practical techniques to overcome them. The presented examples capture the problems that occurred during elicitation the numerical parameters of the model for diagnosis of liver disorders.

key words: knowledge acquisition, Bayesian network, parameter elicitation

1. Introduction

Elicitation of probabilities is often pointed out as the major obstacle in building Bayesian networks [1,2]. In addition to tediousness of elicitation, the resulting numbers are not always reliable because of various factors that can skew them. The problems related to biases and poor calibration during elicitation of judgemental probabilities from human experts are well known [7]. There are several techniques that facilitate the process of the elicitation of numerical parameters from human experts.

Probability elicitation essentially involves posing a set of questions that require an expert to either provide direct probability estimates, or to choose between simple alternatives or bets. In case of direct response methods, experts give their estimates either numerically as cumulative probabilities, graphically by plotting density functions, or verbally by expressing their estimates by terms *fifty-fifty*, *certain*, *improbable*, etc. Quantitative interpretation of these different kinds of descriptors are then encoded numerically. Indirect response methods are more sophisticated. Experts are asked here to make choices between simple alternatives in gambles related to the events in question. The subjective probabilities

[1] Wydział Informatyki, Politechnika Białostocka, Białystok

are inferred from these choices. Two classical approaches to indirect probability elicitation are betting and reference lottery [5,14]. However, it is known that in laboratory setting experts tend bet and take a risk more easily than in a real-world setting [8]. Additionally, these methods tend to be infeasible for models that include thousands of probabilities. Other alternatives include a use of ranking, relative likelihood, and interval techniques [10].

Spiegelhalter *et al.* [15] introduced several techniques for assessment, refinement, and improvement of imprecise probabilities that were elicited from human experts. The method involved measuring quality of assessments by scoring rules. The authors observed that reliable probability assessments can be obtained from experts, although the experts tended to be too extreme in their judgements. The method presented by van der Gaag *et al.* [4] allowed to elicit from a domain expert around 150-200 probabilities per hour. The approach was based on the assessment of both verbal probability expressions and numbers.

This paper presents several knowledge acquisition techniques for building Bayesian network models, particularly, for elicitation of numerical parameters from human experts. The paper is structured as follows: Section 2 describes briefly HEPAR II, a Bayesian network model for diagnosis of liver disorders, Section 3 summarizes the interactions between knowledge engineer and human expert, Sections 4 and 5 present the techniques for elicitation of numerical parameters from human experts. Finally, section 6 concludes the paper.

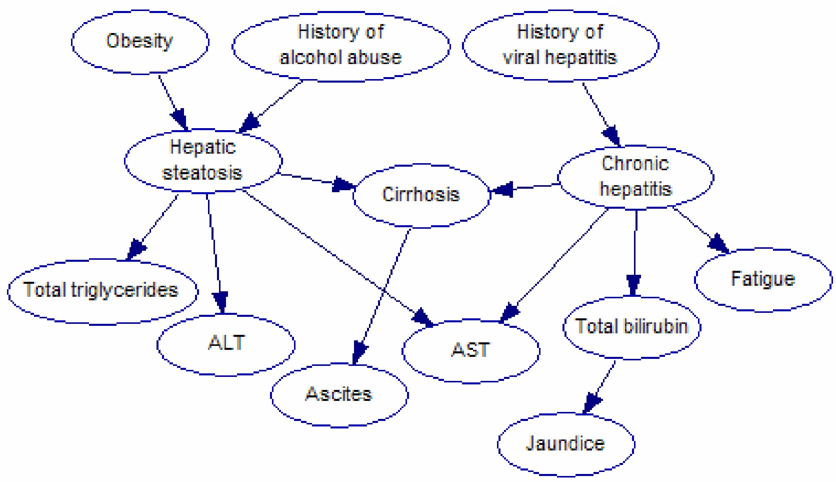

Fig 1. A simplified fragment of the HEPAR II network

2. HEPAR II

HEPAR II is a Bayesian network model for diagnosis of liver disorders [11]. The model consists of 70 variables and covers 11 different liver diseases and 61 medical findings, such as patient self-reported data, signs, symptoms, and laboratory tests results. The structure of the model, (i.e., the nodes of the graph along with arcs among them) was built based on medical literature and conversations with our domain expert, a hepatologist Dr. Hanna Wasyluk and two American experts, a pathologist, Dr. Daniel Schwartz, and a specialist in infectious diseases, Dr. John N. Dowling. The elicitation of the structure took approximately 50 hours of interviews with the experts, of which roughly 40 hours were spent with Dr. Wasyluk and roughly 10 hours spent with Drs. Schwartz and Dowling. This includes model refinement sessions, where previously elicited structure was reevaluated in a group setting. The structure of the model consists of 121 arcs and the average number of parents per node is equal to 1.73. There are on the average 2.24 states per variable. Figure 1 shows a simplified fragment of the HEPAR II network.

There are two versions of HEPAR II: (1) the data model and (2) the expert model. In the former version of the model conditional probability distributions were learned from the HEPAR database.[2] In the latter version, the parameters of HEPAR II were elicited from human expert.

3. Interactions with experts

Building the structure of a Bayesian network typically involves interaction of the knowledge engineer with domain experts. In case of the HEPAR II model, regular, short sessions with the expert worked well. In between these sessions, the knowledge engineer focused on refining the model and preparing questions for the expert. The refinement consisted of analyzing positive and negative influences in the model when the model was fully quantified, i.e., when the numerical parameters of the network were already specified. There are several tools that can be useful in debugging a Bayesian network (e.g., Elvira [16], GeNIe [17]). It helps when the knowledge engineer understands the domain at least at a basic level. It is a good idea to read at least a relevant section of a medical textbook on the topic of the meeting with the expert, so the knowledge engineer is familiar

[2] The HEPAR database was created in 1990 and is thoroughly maintained at the Gastroentorogical Clinic of the Institute of Food and Feeding in Warsaw. Each hepatological case is described by over 160 different medical findings and by a histopathologically verified diagnosis. The version of the HEPAR data set used in HEPAR II consisted of 699 patient records.

with the terminology, the variables, and interaction among them. It is also recommended to record the sessions with the expert because it is often hard to process all the medical knowledge that is provided by a domain expert during a meeting. It is also recommended to organize brainstorming sessions with a participation of knowledge engineers and medical experts who are not directly involved in building the model. With respect to HEPAR II, there were a few such sessions, and they addressed important issues and raised questions about the model.

4. Elicitation of conditional probability distributions

This section presents a set of techniques that were applied during the elicitation of conditional probability distributions of HEPAR II from the domain expert. The expert participating in parameter elicitation was familiar with elementary probability theory, e.g., the expert knew that probability ranges between 0 and 1, and that for a set of mutually exclusive and exhaustive set of events, the probabilities should sum exactly to 1. With respect to probability elicitation for HEPAR II, direct approach was applied.

There were 70 nodes to quantify and 370 independent probabilities to elicit. The elicitation of numerical parameters took around 10 hours, composed of five sessions of roughly two hours each.

4.1 What probability?

Before the process of elicitation begins, it is important to make sure that a common framework between the expert and the knowledge engineer has been established. First of all, the expert should know for which population the probabilities are provided. It is natural that medical experts think in terms of the setting that they work in. Therefore, they often give numerical parameters that reflect their experience in a particular clinic. Since probabilities must refer to a general population, an expert may argue:

> *If I am going to use a model at a hospital, why should I use general population frequencies?*

Hence, it is crucial to agree for which population the probabilities are elicited. Otherwise, the parameters and the entire model can be very wrong [3]. Figure 5 captures risk factors of *Functional hyperbilirubinemia* that were considered in the HEPAR II model. During the elicitation of parameters for the node *Functional hyperbilirubinemia*, one of the values provided by the expert was initially equal to 0.8. This value was supposed to indicate the probability of suffering from *Functional hyperbilirubinemia* by men below 30. However, it appeared that this values represented a hospital population instead of a general population. After clarifying this issue, the value was changed to 0.1.

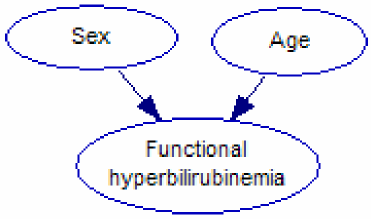

Fig 2. Modeling risk factors of *Functional hyperbilirubinemia*

An example that has been often applied during the quantification of HEPAR II was related to a *tram population* (we assumed here that a tram population is representative of a general population in Poland), i.e., when the knowledge engineer asked the expert about a particular probability, for example, a prevalence of the variable *Alcohol abuse*, the following question was posed:

> *Imagine a tram with 100 people. How many of them would suffer from alcohol abuse?*

The expert found such framing of the question helpful and easy to follow.

4.2 How to get it?

The following section provides the examples of elicitation of numerical parameters for HEPAR II from a human expert. It introduces several types of questions that were posed to the expert during the elicitation sessions for HEPAR II. It was challenging to quantify some of the nodes that were modeled as binary variables. For example, it was hard to estimate the conditional probability distribution for *Steatosis* given *Alcohol abuse*, since the latter was modeled in HEPAR II as a binary variable. It would be easier to quantify this node if alcoholism was represented in the model by several states indicating different degrees of alcoholism. So, during the elicitation of the probability that *Steatosis* occurs given that a patient abuses alcohol, the expert had to average over several groups of patients representing different stages of alcoholism. Similar situation was observed during the quantification of nodes that represented children of the variable *Gallstones* (modeled as a binary variable). It happens that gallstones present with different stages, hence, the expert again needed to average over several groups of patients representing different stages of the disease.

An advisable practice in probability elicitation is verification of the elicited numbers. Monti *et al.* [9] noticed that confronting the expert with inconsistencies in the assessments, after the expert had already gone through the whole elicitation process, is ineffective. With respect to HEPAR II, the expert was asked to confirm a particular probability in terms of a percentage rather than a probability value. For example, when the expert provided a value equal to 0.3, representing a probability that *Amylase* is elevated in patients with *Gallstones*, the following question was posed:

> *Does it mean that 30% of the patients presenting with gallstones have elevated amylase?*

The expert often re-thought the situation and refined the value. It appears that in some cases, it is easier to specify the probabilities for a node that is conditioned on another node. For example, the expert found it difficult to provide the prevalence for the node *Reactive hepatitis* (representing a rare liver disease). However, she found it much easier to quantify this node when it had *Hepatotoxic medications* as a direct predecessor (*Hepatotoxic medications* is a risk factor of *Reactive hepatitis*).

It is challenging to quantify a node with several parents, especially when they consist of several states. For binary nodes with two parents, it is helpful to draw a table that simplifies the process of elicitation. Table 1 captures the table that was introduced to the expert in order to quantify the node *Bleeding*. Each value in this table represents independent probabilities that *Bleeding* occurs given a combination of values of *Platelet* and *INR*. In my experience, experts prefer to start the elicitation from situations that are extreme: either from a normal state or the most abnormal. For example, with respect to elicitation of parameters for the node *Bleeding*, the expert specified the values in bold font first (see Table 1). Similar tables were created for the variables *PBC* and *Functional hyperbilirubinemia*. The expert found these tables useful.

Table 1

Elicitation of probabilities for the node *Bleeding*

Platelet/INR	normal	low	very low
normal	**0.0**	0.01	0.01
low	0.01	0.1	0.2
very low	0.01	0.2	0.5
extremely low	0.01	0.4	**0.9**

Bayesian networks allow to combine different sources of knowledge. For example, they allow to combine expert knowledge with existing clinical data. However, the constructors of Bayesian network models should be aware of biases that can occur during combining different sources of knowledge [3].

5. Assessment of Noisy-OR parameters

Some types of conditional probability distributions can be approximated by canonical interaction models that require fewer parameters. Very often such canonical interactions approximate the true distribution sufficiently well and can reduce the model building effort significantly. One type of canonical interaction, widely used in Bayesian networks, is known as Noisy-OR gate [6,12]. Noisy-OR gates are usually used to describe the interaction between n causes X_1, X_2, …., X_n and their common effect Y. The causes X_i are each assumed to be sufficient to cause Y in absence of other causes and their ability to cause Y is assumed independent of the presence of other causes.

Elicitation of numerical parameters for HEPAR II involved also the assessment of parameters for Noisy-OR gates. This section describes details related to obtaining Noisy-OR parameters. There were 25 nodes identified by the expert that could be approximated by Noisy-OR gates. To obtain the numerical parameters for these nodes, direct approach for elicitation was applied. There was a total of 189 parameters and the assessment took a total of about four hours of expert time.

There were several types of questions posed to the expert to investigate whether a node can be approximated by a parametric distribution. Figure 3 captures the causes of *Cirrhosis* that were modeled in HEPAR II. To check whether *Cirrhosis* can be approximated by a Noisy-OR gate, the following question was posed:

> *Is Cirrhosis the effect of combination of both causes Fibrosis and Steatosis?*

If the expert answered that *Cirrhosis* is the effect of the two modeled causes, the interaction between the node and its parents cannot be modeled by a Noisy-OR gate.

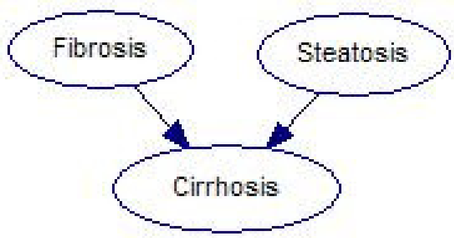

Fig 3. HEPAR II: Modeling the node *Cirrhosis*

Another question that was posed by the knowledge engineer was related to the interaction of mechanisms. Figure 4 captures two causes of the node *Carcinoma* modeled in HEPAR II. There were two mechanisms considered here. The first mechanism represented *PBC* leading to *Carcinoma*. The second mechanism was related to *Cirrhosis* causing *Carcinoma*. In such cases the following question was posed:

Do mechanism-1 and mechanism-2 interact with each other in causing carcinoma?

If the expert is not aware of any interaction between the mechanisms, it is reasonable to assume that such interactions do not exist or are not too strong and the interaction among *Cirrhosis*, *PBC*, and *Carcinoma* can be modeled by a Noisy-OR gate.

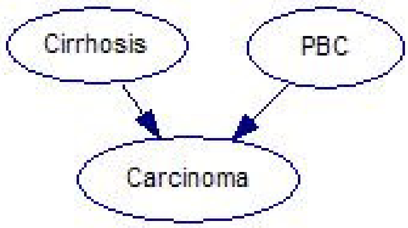

Fig 4. HEPAR II: Modeling the node *Carcinoma*

Another example involved modeling the causes of the variable *Nausea* (see Figure 5). In this case the following question was posed:

> *Imagine that there are two patients: one is abusing alcohol and the other is not. Both of them receive hepatotoxic medications. Will the patient who drinks alcohol have a significantly different probability of being affected by these medications?*

If this probability is reported by the experts as significantly higher or lower, then there is a significant synergy between the causes and *Nausea* should not be modeled as a Noisy-OR gate.

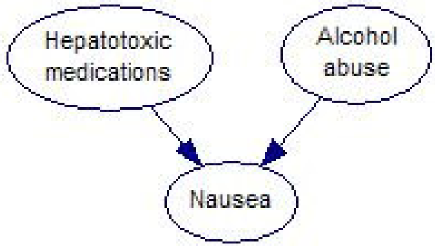

Fig 5. HEPAR II: Modeling the node *Nausea*

Table 2

Noisy-OR parameters for the node Total cholesterol

Cholesterol/parent	Steatosis	ACH	PCH	PBC	Leak
very high	0.02	0.02	0	0	0
high	0.6	0.4	0.25	0.1	0.01
normal	0.38	0.58	0.75	0.9	0.99

Similarly to the CPT (conditional probability table) elicitation, the expert preferred to provide first the values for either normal or the most abnormal states. Table 2 captures the Noisy-OR parameters for the node *Total cholesterol*.[3] With respect to this node, the expert indicated first the probabilities for the states *normal* and *very high*. Then, she specified the values for intermediate range of cholesterol level. The expert often provided the initial values and then adjusted them until reasonably satisfied with them (this behavior is known as "anchoring and adjustment heuristic" [6]). The expert preferred providing a Noisy-OR probability within a particular state, i.e., for a normal state the parameters for each of the auses were specified first (the values in the fourth row of Table 2). The expert often found it difficult to provide a value of the leak probability, i.e., the expert tended to provide this value for a hospital population instead of referring to the general population. I have also observed that the expert preferred giving a leak probability at the very end of the elicitation.

[3] The abbreviations used in Table 2: ACH and PCH, stand for Active Chronic Hepatitis and Persistent Chronic Hepatitis respectively.

6. Conclusions

This paper presented typical problems that are encountered by the knowledge engineers during building Bayesian network models. The author provided examples of problems that have been occured during elicitation of numerical parameters from human experts and then illustrated practical techniques to overcome these problems.

I would like to acknowledge funding from the MNiI grant number 3T10C03529. Contents of this paper is based in part on one of the chapters of my PhD dissertation [11].

References:

[1] Marek J. Druzdzel and Linda C. van der Gaag. Elicitation of probabilities for belief networks: Combining qualitative and quantitative information. In *Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence (UAI-95)*, pages 141-148, San Francisco, CA, 1995. Morgan Kaufmann Publishers, Inc.

[2] Marek J. Druzdzel and Linda C. van der Gaag. Building probabilistic networks: "Where do the numbers come from?" guest editors' introduction. *IEEE Transactions on Knowledge and Data Engineering*, 12(4):481-486, 2000.

[3] Marek J. Druzdzel and F. Javier Díez. Combining knowledge from different sources in causal probabilistic models. *Journal of Machine Learning Research*, 4:295-316, 2003.

[4] Linda van der Gaag, Silja Renooij, Cilia Witteman, Berthe Aleman, and Babs Taal. How to Elicit Many Probabilities, In *Proceedings of the Fifteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-99)*, Morgan Kaufmann Publishers, San Francisco, CA, 1999, 647-654.

[5] J. Hampton, P. Moore, and H. Thomas. Subjective probability and its measurements. *Journal of the Royal Statistical Society*, 21-42, 1973.

[6] Max Henrion. Some practical issues in constructing belief networks. W L.N. Kanal, T.S. Levitt, and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence 3*, 161-173. Elsevier Science Publishers B.V., North Holland. 1989.

[7] Daniel Kahneman, Paul Slovic, and Amos Tversky. Judgment Under Uncertainty: Heuristics and Biases, Cambridge University Press, Cambridge, 1982.

[8] R. Duncan Luce and Howard Raiffa. Games and Decisions: Introduction and Critical Survey. Dover Publications, 1989.
[9] Stefano Monti and Giuseppe Carenini. Dealing with the expert inconsistency in probability elicitation. *IEEE on Knowledge and data engineering*, 12(4):499-508, 2000.
[10] M. Granger Morgan and Max Henrion. Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis, Cambridge University Press, Cambridge, 1990.
[11] Agnieszka Oniśko, Marek J. Druzdzel, and Hanna Wasyluk. Learning Bayesian network parameters from small data sets: Application of Noisy-OR gates. *International Journal of Approximate Reasoning*, 27(2):165-182, 2001.
[12] Agnieszka Oniśko. *Probabilistic Causal Models in Medicine: Application to Diagnosis of Liver Disorders*. Ph.D. dissertation, Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Science, Warsaw, April 2003.
[13] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, Inc., San Mateo, CA. 1988.
[14] L. Savage. The foundations of statistics. Wiley, New York, 1954.
[15] J. Spiegelhalter, David, C. G. Franklin, Rodney, and Kate Bull. Assessment, criticism and improvement of imprecise subjective probabilities for a medical expert systems. *Uncertainty in Artificial Intelligence* 5, pages 285-294. Elsevier Science Publishers B.V., North Holland, 1990.
[16] http://www.ia.uned.es/~elvira
[17] http://genie.sis.pitt.edu

POZYSKIWANIE WIEDZY OD EKSPERTÓW W BUDOWANIU MODELI SIECI BAYESOWSKICH

Streszczenie: Pozyskiwanie wiedzy od ekspertów jest kosztownym i czasochłonnym zadaniem. Pomimo ogromnej wiedzy i doświadczenia, jakie posiadają eksperci, niejednokrotnie nie potrafią ich przenieść na tworzony model. Poniższy artykuł opisuje przykłady problemów, z jakimi może się zetknąć inżynier wiedzy w trakcie budowania modeli sieci bayesowskich, jak również proponuje rozwiązania tych problemów. Prezentowane przykłady dotyczą problemów, jakie pojawiły się w trakcie pozyskiwania od eksperta parametrów numerycznych modelu sieci bayesowskiej do diagnozowania chorób wątroby.

Keywords: pozyskiwanie wiedzy, inżynieria wiedzy, sieci bayesowskie