TOOLS FOR EXPLANATION IN BAYESIAN NETWORKS WITH APPLICATION TO AN AGRICULTURAL PROBLEM

http://www.lr.dk/dina/papers/tools.htm

Søren L. Dittmer[1] and Finn V. Jensen[2]

[1]*DINA Skejby*
Danish Agricultural Advisory Centre
Udkærsvej 15, Skejby
DK – 8200 Århus N
Denmark
E-mail: dittmer@lr.dk

[2]*DINA Aalborg*
Department of Computer Science
Aalborg University
Fredrik Bajers Vej 7E
DK – 9220 Aalborg Øst
Denmark
E-mail: fvj@cs.auc.dk

Abstract: We give an overview of a minimal set of tools for explanation in decision problems formulated as Bayesian networks. After an introduction to the Bayesian network paradigm we introduce sensitivity analysis and data conflict analysis as applied to Bayesian networks. Finally, we apply these tools to BOBLO, a large Bayesian network for determining the blood group of Jersey cattle.

Keywords: Bayesian networks, probabilistic inference, sensitivity analysis, conflict analysis, explanation

1 Introduction

In almost any production system in a modern society, we find decision makers facing complex scenarios where the decisions made have a considerable effect on the feasibility of the production. A very similar situation is applicable to physicians facing a diagnostic problem where the goal is to diagnose correctly at minimal cost and discomfort for the patient.

From cognitive science it is known that a decision maker is able to process up to approximately nine independent input factors before the quality of the decision will decrease due to information overload[1] (Horvitz & Barry, 1995). This fact has caused an increasing number of decision support systems to be used in the solution of complex decision problems with some uncertainty involved.

[1] The number of independent inputs may become as low as two (2) when the decision maker is in a "noisy" environment, that is, when the decision maker is stressed, disturbed, under (time) pressure etc.

If there is a risk of a considerable loss upon making a wrong decision, the decision maker is particularly interested in confirming the information on which she bases her decision. This means that any information produced by automated and modeled systems should be subject to an explanatory procedure in which the recommendations are scrutinized. A set of explanatory actions would at least include *sensitivity analysis* and *conflict analysis*.

2 Bayesian networks

Bayesian networks are an increasingly popular paradigm for representing decision problems. A Bayesian network (Pearl, 1988; Jensen, 1996) as the one shown in Figure 1, is a directed, acyclic graph (DAG) with a set of vertices representing the random variables in the decision problem, and a set of directed edges connecting the vertices and thereby defining dependencies between variables. The quantification of these dependencies are conditional probability tables associated with each variable. Consider edges pointing into the vertex a from every vertex in the (possibly empty) set B. Then the set of vertices in B is called *parents* of a, and is denoted pa(a). The probability table associated with variable a holds $P(a|\text{pa}(a))$, that is the conditional probabilities of variable a given pa(a).

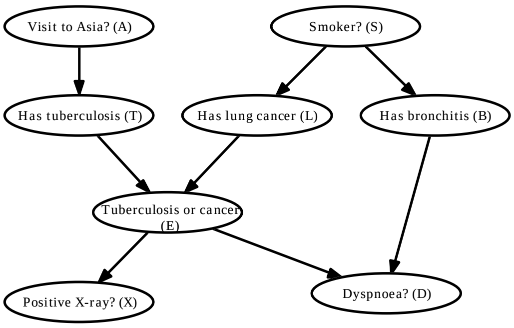

Figure 1: A sample Bayesian network (from Lauritzen & Spiegelhalter (1988)).

Bayesian networks are primarily used for *inference*: Evidence e in the form of statements on the state of some variables (such statements are also called *findings*) is used to infer the posterior probabilities $P(a|e)$ for all remaining variables. Inference for Bayesian networks is carried out by *inserting* the evidence in a derived structure called a *junction tree* and performing a *propagation* in the junction tree, see for example Jensen (1996) for a description. A detailed study of propagation is given in Shafer (1996).
It is beyond the scope of this paper to give further details on the inference method. we shall only state that propagation requires considerable computer resources and usually it is not a task which should be repeated hundreds of times.

In this paper we shall frequently use *Bayes' rule*:

$$P(a|b,c) = \frac{P(b|a,c) \cdot P(a|c)}{P(b|c)}$$

3 Explanation in Bayesian networks

Several means of explanatory tools have been proposed for decision problems formulated by Bayesian networks. In this paper we will concentrate on the application of only two, namely *sensitivity analysis* and *conflict analysis*. Sensitivity analysis in an analysis of how sensitive the conclusion is to variations in the evidence. Conflict analysis is concerned with conflicting evidence in order to identify erroneous data.

3.1 Sensitivity analysis

Sensitivity analysis consists of several tasks within decision support systems. A usual task is to determine e.g. which subsets of evidence are irrelevant to some hypothesis h_i, or which subsets are crucial to discriminate between hypotheses h_j and h_k. Jensen, Aldenryd & Jensen (1994) report on subset analysis and determine five categories of subsets of evidence for some hypothesis h, evidence e, and evidence subset $e' \subseteq e$. The subset e' is *important* if the exclusion of e' reduces the ratio $P(h|e\backslash e')/P(h|e)$ below some threshold θ_1, it is *sufficient* if $P(h|e')/P(h|e)$ rises above some threshold θ_2,, it is *minimal sufficient* if e' is sufficient but no proper subset of e' is, it is *crucial* if e' is a subset of any sufficient set, and it is *decisive* if $P(h|e')$ is greater than some threshold θ_3. Also *redundancy* in data can be found from sensitivity analysis. A subset e' is said to be redundant if $P(h|e\backslash e')/P(h|e)$ is greater than some threshold θ_4. Hence, the crucial thing in subset analysis is to calculate $P(h|e')$ for a large amount of subsets $e' \subseteq e$. They can be obtained from the propagation in the following way. $P(e)$ is obtained as the normalization constant after a propagation with all evidence inserted, and $P(h|e)$ is found by applying Bayes' rule as follows. We have that

$$P(e|h) \cdot P(h) = P(e,h) = P(h|e) \cdot P(e) ,$$

and therefore we can find the desired values by entering h along with the evidence and propagate, yielding $P(e,h)$ as normalization constant. $P(h|e)$ is then simply found by division[2], that is

$$P(h|e) = \frac{P(e,h)}{P(e)} .$$

In principle we need one propagation for calculating $P(e',h)$ for each $e' \subseteq e$, and if the model is large and e consists of a large data set then a thorough subset analysis is intractable. Various tricks exist to reduce the calculation burden (Jensen *et al.*, 1994) but in general it is not possible to perform a thorough subset analysis.

Another task of sensitivity analysis is a *what-if* analysis, that is, a series of questions about the impact on the hypothesis h, when one or more findings are changed. This is a means for identifying critical findings that – if altered – may change the hypothesis, and thereby rendering a decision useful or maybe even harmful. If findings are found through some data sensor (or are typed in manually), there is a possibility that the entered findings are erroneous. If a critical decision is dependent on this finding, a *what-if* analysis may prevent losses. Contrary to subset analysis, a what-if analysis turns out to be tractable.

Assume that the evidence contains the statement "*the variable* A *is in the state* a", and we are interested in knowing the effect on the hypothesis h if A had been in some other state. Let e' denote the evidence without the statement above. Hence we are interested in calculating $P(h|e',A)$. Bayes' rule yields

$$P(h|e', A) = \frac{P(A|e',h) \cdot P(h|e')}{P(A|e')} .$$

[2] Of course, assuming that $P(e) > 0$, that is the indivual parts of the evidence have a strictly positive probability of occuring simultaneously.

A propagation trick (Dawid, 1992; Jensen, 1995) yields that 3 propagations usually will be sufficient for obtaining the required probabilities for all variables in the network. It is worth noting that this technique also can be used for calculating the effect of observing an unknown variable.

3.2 Conflict analysis

Let e consist of a set $\{X_1, \dots, X_n\}$ of statements on single variables. Several methods have been proposed for analyzing e for internal conflicts (Habbema, 1976), but most of them are computationally inefficient. Jensen *et al.* (1991) define a measure of data conflict which is easily obtainable through one propagation. This conflict measure is defined as

$$\mathrm{conf}(X_1, \dots, X_n) = \log \frac{\mathrm{P}(X_1) \times \dots \times \mathrm{P}(X_n)}{\mathrm{P}(X_1, \dots, X_n)},$$

where $e = \{X_1, \dots, X_n\}$.

The idea behind this conflict measure is that if a data conflict is present, the joint probability will decrease in comparison with the product of the probabilities of each of the findings, and the conflict measure becomes positive. Coherent data will yield a high joint probability compared to the product of the finding probabilities, and a negative conflict measure is found.

It may happen that typical data from a *rare case* cause a high conflict value. If the case for example is a certain variable q being in a very rare state then $\mathrm{P}(e|q)$ should be much bigger than $\mathrm{P}(e)$. From Bayes' rule we get

$$\frac{\mathrm{P}(e|q)}{\mathrm{P}(e)} = \frac{\mathrm{P}(q|e)}{\mathrm{P}(q)}$$

and, since both $\mathrm{P}(q|e)$ and $\mathrm{P}(e)$ are achieved in the propagation, it is easy to monitor the system for possible rare cases.

4 The BOBLO example

The BOBLO system (Rasmussen, 1992) is a Bayesian network for determining the blood group of Danish Jersey cattle in the F-blood group system. In fact, the network tests whether the alleged parentship is correct. The network holds 221 variables each containing between two and 63 states, and the resulting junction tree contains more than 1.2 million table entries. The hypothesis node is the node [*Parental error*] whith four states, (*both incorrect*), (*sire incorrect*), (*dam incorrect*), and (*no error*). The experiment consists of nine test cases each with more than 70 findings involved.

4.1 What-if analysis and results

We have applied a *what-if* analysis to the same nine test cases. This analysis examines what the effect is on the hypothesis if a finding is changed. We have performed this analysis on single findings only, that is we do not examine the effects of two or more findings changing simultaneously as this may become an intractable task for even small sizes of finding subsets. The number of findings (the size of the subset) examined simultaneously should however be adjusted according to the hypothesis. [The node [*Parental error*] contains four states, but for our purposes it might have been two separate nodes each containing two states. The findings for the parents are independent of each other, and therefore we need at least two findings (one for each of the parents) to change the hypothesis into (*both incorrect*). For our example we are satisfied with the change from (*no error*) to either (*sire incorrect*) or (*dam incorrect*).]

The results from the analysis actually *does* show that in some cases a change in a single finding may yield a different result. Initially these cases all have (*no error*) as the prevalent hypothesis and a conflict measure indicating coherent data, and therefore without this analysis these cases would pass without further examination. The analysis does show for one case (3) that for *any other state of a specific single variable*, the hypothesis will change from (*no error*) into an error condition. For other cases (3, 4, 7) the analysis shows that *most other states of a single variable* will cause the hypothesis to change from (*no error*) to one of the error conditions.

	few other states cause change in hypothesis	*most other states cause change in hypothesis*	*all other states cause change in hypothesis*
Case set 1 (75 findings)	csts1, cstd1		
Case set 2 (71 findings)	csts1, bsts2		
Case set 3 (91 findings)	fsts1, csts2	bstd2	bsts2
Case set 4 (86 findings)		csts1, bsts1	
Case set 5 (71 findings)			
Case set 6 (71 findings)			
Case set 7 (72 findings)		bsts1	
Case set 8 (72 findings)			
Case set 9 (72 findings)	l18 (+ conflict)		

Table 1: Identified single variables from the BOBLO systems where change of inputs can cause change in the hypothesis.

When we take the conflict measure into account, we see that for all cases but one, the states causing a shift to an error condition in the hypothesis are coherent with the other findings. Only in one case (9) does the conflict measure signal a data conflict when we enter a state which causes the hypothesis to indicate a parental error. Such a possible finding should perhaps not cause an alarm for the hypothesis, but the source of the finding should perhaps be consulted. In the same case we find that some states in an additional two variables are associated with large, positive conflict measures, Such findings would cause the evidence to be less decisive but would not alone cause a change in the hypothesis.

This analysis should of course be accompanied by an assessment of the possibility for the values to be in error, that is the possibility of sensor error, typos, and how difficult it is to discriminate between the states of a variable. If the finding is correlated with other findings (derived from the same sample, etc.) the analysis should probably be extended to examine these findings together. This does, however, present a very large (and most likely computationally unacceptable) task if an exhaustive analysis is needed.

We see that the analysis does pin-point single findings that may change the prevalent hypothesis should they be in error. For critical decisions this may be a very valuable tool, even though the computational task could be considerable.

5 References

Dawid, A. P. (1992): Applications of a general propagation algorithm in a probabilistic expert system. *Statistics and Computing* **2**, 25–36.

Habbema, J. D. F. (1976): Model diagnosis and detection of diseases. *In* de Dombal *et al.* (eds.) Decision making and medical care. Amsterdam: Elsevier Science; pp. 399–411.

Horvitz, E. J. & B. Barry (1995): Display of information for time-critical decision making. *In* Besnard, P. & S. Hanks (eds.) *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence.* San Francisco, CA: Morgan Kaufman; pp. 296–305.

Jensen, F. V., B. Chamberlain, T. Nordahl & F. Jensen (1991): Analysis in HUGIN of data conflict. *In* N.-H. Bonnisone *et al.* (eds.), *Uncertainty in Artificial Intelligence 6*, pp. 519–528.

Jensen, F. V. (1996): *An Introduction to Bayesian Networks*. London: UCL Press.

Jensen, F. V. (1995): Cautious propagation in Bayesian networks. *In* Besnard, P. & S. Hanks (eds.) *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*. San Francisco, CA: Morgan Kaufman; pp. 323–328.

Jensen, F. V., S. H. Aldenryd & K. B. Jensen (1995): Sensitivity analysis in Bayesian networks. *In* Froidevaux, C. and J. Kohlas. (eds.) *Symbolic and Quantitative Approaches to Reasoning and Uncertainty*. Springer Lecture Notes in Artificial intelligence 946, Springer, Berlin; pp. 243–250.

Lauritzen, S. L. & D. J. Spiegelhalter (1988): Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, Series B (Methodological),* **50**(2):157–224.

Pearl, J. (1988): *Probabilistic Reasoning In Intelligent Systems: Networks of Plausible Inference.* San Mateo, CA: Morgan Kaufman Publishers.

Rasmussen, L. K. (1992): Blood groups determination of Danish Jersey cattle in the F-blood group system, *Technical Report Dina Res. Rep. No. 8*, Research Centre Foulum, Denmark.

Shafer, G. (1996): *Probabilistic Expert Systems*. CBMS – NSF regional Conference Series in Applied Mathematics; SIAM.