Focused quantification of a belief network using sensitivity analysis*

Niels Peek[1], Veerle M.H. Coupé[2], Jaap Ottenkamp[3]

[1] Department of Computer Science, Utrecht University
P.O. Box 80.089, 3508 TB Utrecht, The Netherlands

[2] Center for Clinical Decision Sciences,
Department of Public Health, Erasmus University Rotterdam

[3] Department of Paediatric Cardiology,
Leiden University Medical Center

Abstract

The quantification of belief networks is known to be a laborious and difficult task, which hampers their application in practice. However, sensitivity analyses generally reveal that the influences of individual parameters on a network's performance differ considerably. This suggests that the quantification effort can be focused on the most influential parameters, as for less influential parameters, rough estimates may suffice. The paper presents an empirical investigation of the viability of this approach, by comparing several belief-network quantifications of different levels of informedness. It was established that refining a limited number of highly influential parameters in a poorly-informed quantification may be sufficient to obtain satisfying network performance.

1 Introduction

In building a Bayesian belief network [8], two closely related tasks can be discerned: the construction of the graphical part of the network, and subsequent quantification of the local conditional probability distributions associated with the variables in the network. Especially quantifying a belief network is a difficult and time-consuming task; it is often considered to be a bottleneck in belief-network construction. Several methods have therefore been proposed in the literature to facilitate the task [4, 11]. It has also been claimed however that once the graphical part of the network is correct, then the behaviour of the network is insensitive to the quality of the majority of quantification parameters, [10]. If this conjecture is true, then a satisfactory network quantification can be obtained by only estimating a small set of highly influential parameters from well-informed sources, and taking rough estimates for the others. Recently, Coupé et al. [2] have proposed a quantification procedure for belief networks that is based on the above conjecture of variational sensitivity. The procedure consists of iteratively

*Presented at: *Eleventh Belgium-Netherlands Conference on Artificial Intelligence*, Maastricht, November 1999.

performing *sensitivity analyses* of an initially roughly quantified network, in order to stepwise refine the quantification. Sensitivity analyses can reveal which network parameters have a large effect on posterior probabilities, and therefore, on which parameters the quantification effort should be focused.

This paper presents an empirical investigation regarding the viability of the quantification procedure proposed in [2]. As a case study, we selected a belief network that models the clinical pathophysiology of ventricular septal defect (VSD), a congenital heart disease. Three quantifications were obtained for this network, differing in level of underlying informedness (i.e. amount of domain knowledge used); in the two least-informed quantifications, the estimates of highly influential parameters were stepwise replaced with the estimates from the third, best-informed, quantification. Our results show that the procedure contributes to efficient quantification of a belief network: after selective replacements of probability estimates, the poorly-informed quantifications give predictions that are comparable to the network that is completely quantified with well-informed estimates.

The paper is organised as follows. In Section 2, we briefly discuss the belief network for VSD. Section 3 gives a description of the method of investigation that was used. Then, in Section 4, the results of the sensitivity analyses and subsequent refinements of the network quantification are presented. We conclude with a discussion in Section 5.

2 The VSD network

Ventricular septal defect (VSD) is an abnormal opening in the wall that separates the heart's two ventricles [5]; it is the most frequently occurring congenital heart disease. The main pathophysiological consequence of a VSD is blood flow, called a *shunt*, from the left to the right ventricle. The shunting of blood accounts for most of the symptoms associated with VSD, such as shortness of breath, growth arrearage, cardiomegaly (enlarged heart), hepatomegaly (enlarged liver), and pulmonary infections. For the clinician, the main dilemma when treating a VSD patient is to decide if and when to submit a patient to surgery. About 70% of all VSDs close spontaneously by normal tissue growth during the first years of life; this development precludes the need for surgical intervention. Unfortunately, when the defect does not close, the continual shunting may cause severe, irreversible damage to the lung circulation; surgical intervention is indispensable to prevent this reaction.

A belief network that models VSD pathophysiology was developed in co-operation with a field expert, a senior paediatric cardiologist [9]. The independency graph for the network was developed by hand; it consists of 38 nodes and 50 arcs; the number of parameters needed to quantify the belief network is 1298. Prior to quantification three types of qualitative information on the probability distributions were collected: (i) *functional dependencies*, modelling deterministic relations between variables, (ii) *consistency constraints*, describing impossible combinations of values, and (iii) *qualitative influences*, expressing the sign of probabilistic interactions. Each of these information types can be interpreted as expressing a constraint on one of the local probability distributions in the quantification [4]. The functional dependencies and consistency constraints jointly reduced the number of parameters that need estimation from 1298 to 738. A total of 24 positive and 5 negative qualitative influences was found, each inducing a number of inequalities between quantification parameters.

3 Quantification and sensitivity analysis

This section reviews the quantification procedure under study, discusses the three quantifications that were obtained for the VSD network, and describes 5 *case profiles*, typical patterns of observations on VSD patients, that were used in the experiments.

3.1 A focused quantification procedure

Sensitivity analysis is a technique to study the effects of varying a model's parameters on its predictions. It is widely used in the fields of decision theory and mathematical modelling, [6, 7]. For a Bayesian belief network, sensitivity analysis provides for studying the effects of variations in the estimates of the network's quantification parameters on one or more posterior probabilities of interest. In this investigation, we restrict ourselves to *one-way sensitivity analyses*, where estimates of the network's parameters are varied one at a time.

In a belief network, the relationship between a probabilistic parameter θ and a posterior probability $y = \Pr(V = v \mid \xi)$ can be expressed as

$$f(x) \quad = \quad \frac{a \cdot x + b}{x + c}, \tag{1}$$

where a, b, and c are real-valued constants, and $0 \leq x \leq 1$ is the value of θ, [1]. Now, if x_0 is the estimate for θ in a given network quantification Q, then $f'(x_0)$ provides a quantitative impression of the effect of (small) variations in the estimate of θ on the posterior in that quantification. We have chosen to use this quantity as a measure of influence of parameter θ on $\Pr(V = v \mid \xi)$. It is often found that $b = ac$ for many network parameters; then $f'(x) = 0$ for all x and these parameters are therefore uninfluential. Note that the shape of the function f may differ with the observed evidence ξ; sensitivity analyses should therefore be performed for several evidence sets. We will refer to these sets as *case profiles*. It should also be established which variable V is indicative for the performance of the belief network. It depends on the envisioned application of the belief network under consideration; we will refer to it as the *measurement variable*.

To facilitate the quantification of a belief network, the following two-stage procedure is proposed in [2]. After the graphical part of the network has been assessed, a rough quantification is established. Such a rough quantification can be based, for instance, on a small collection of statistical data, or order-of-magnitude estimates. The second stage consists of performing sensitivity analyses on the network. The estimates of parameters that turn out to be highly influential are refined, where the refined estimates are obtained, for instance, by gathering more statistical data on the variables involved. The effort of obtaining highly accurate parameter estimates is thus limited to a subset of quantification parameters.

3.2 Method of investigation

The main question now is: how does a network quantification obtained by following the procedure described above compare to a network quantification that consists completely of accurate parameter estimates? To answer this question, we have acquired three different quantifications of the VSD network, Q_1, Q_2, and Q_3, each consisting of estimates for the 560 parameters that were not determined by a functional dependency or a consistency constraint. Quantification Q_1 consists of completely uninformative estimates: a uniform probability distribution

was used for each variable in the network. Quantification Q_2 consists of estimates supplied by a non-medical researcher involved in the project, based on the available information on qualitative probabilistic influences between network variables. These qualitative influences were translated into linear models. For quantification Q_3 finally, each of the parameters was estimated by the expert physician. We assumed quantification Q_3 to be more accurate than quantifications Q_1 and Q_2. Our objective was to assess whether it is possible to improve quantifications Q_1 and Q_2 up to the level of Q_3, where the improvements consist of selective revisions of influential parameters.

The next step in the investigation was to select an appropriate measurement variable and case profiles. The field expert indicated that he would normally base his management decisions on his expectation of the shunt size, the amount of blood that flows through the VSD. Shunt size is expressed as ratio of pulmonary and systemic blood flows; of particular interest is whether this ratio exceeds the value $2:1$. We have therefore focused on the value $\texttt{shunt} \geq 2:1$ to measure the performance of the VSD network under different quantifications. Furthermore, five case profiles $\xi_1, \ldots, \xi_5$ were composed in cooperation with the field expert. We have chosen to use profiles that consist of values for network variables that are measurable in clinical circumstances. The profiles differ with respect to the (suspected) severity of the underlying disease and their unequivocalness. The first profile, ξ_1, consists of symptoms that are for a VSD patient, but the necessary heart murmurs are absent; this profile corresponds to a patient not having a VSD, but some other disease. Profiles ξ_2 and ξ_5 consist of related symptoms and do include heart murmurs; they provide compelling evidence for a small and a large VSD, respectively. Profiles ξ_3 and ξ_4 also clearly relate to VSD, but the severity is now harder to assess, as the findings are more equivocal.

4 Results

In this section we compare the predictions of the three network quantifications for the five case profiles, discuss the results of the sensitivity analyses, and the effects of refining network quantifications Q_1 and Q_2.

4.1 Posteriors of the original network quantifications

Let $P_{i,j}$ denote the posterior probability $\Pr(\texttt{shunt} \geq 2:1 \mid \xi_i)$, $1 \leq i \leq 5$, under quantification Q_j, $j = 1, 2, 3$; these posteriors are shown in Figure 1. We see $P_{1,j}$ and $P_{2,j}$ are zero for each quantifiction Q_j. This is explained by the consistency constraints that were established prior to quantification; the first two profiles therefore provide no basis for comparison. For the other profiles, we find that the quantifications Q_1 through Q_3 have increasingly more discriminative power; this is in line with the increasing level of informedness of the quantifications. Quantification Q_1 gives the same posterior for each of these profiles, due to the uniform distributions used in this quantification. The posteriors of quantifications Q_2 and Q_3 are more pronounced. Notably, a large difference is seen in the posteriors for profile ξ_3. This corresponds well to the fact that this profile provides equivocal observations and is therefore hard to interpret. The field expert indicated, however, that quantification Q_3's posterior was best in line with his intuition for this profile.

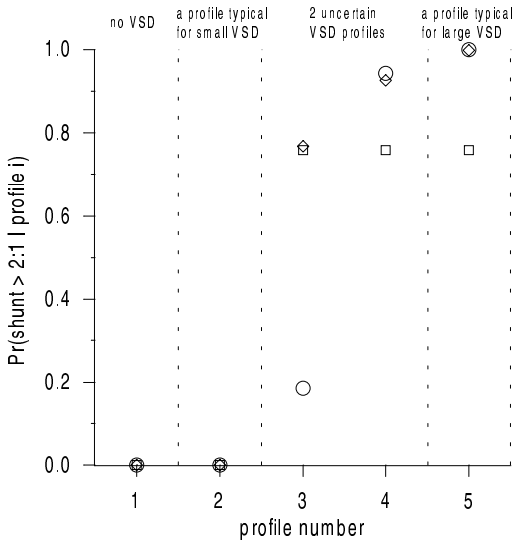

Figure 1: Posterior probabilities, $\Box$ for Q_1, $\Diamond$ for Q_2 and $\circ$ for Q_3.

4.2 Results of the sensitivity analyses

A one-way sensitivity analysis of the posterior $P_{i,j}$ was performed for all three quantifications Q_j and with each of the five case profiles ξ_i; the 560 parameters that were determined by a functional relationship or a consistency constraint were excluded from the analysis. Table 1 shows the average influences of the remaining 738 parameters. Between parentheses, the percentage of uninfluential network parameters is written; note that this percentage is often considerable.

	Q_1	Q_2	Q_3
ξ_1	0 (100%)	0 (100%)	0 (100%)
ξ_2	0 (100%)	0 (100%)	0 (100%)
ξ_3	0.0117 (87%)	0.0169 (84%)	0.0145 (83%)
ξ_4	0.0122 (83%)	0.0072 (83%)	0.0088 (85%)
ξ_5	0.0126 (82%)	0.0007 (78%)	0.0000 (77%)

Table 1: The average sensitivity of $P_{i,j}$ to parameter variations.

We recall from the previous section that for profiles ξ_1 and ξ_2, the posterior is determined to be zero by consistency constraints, and varying parameters estimates will therefore not affect it; we restrict the discussion to profiles ξ_3, ξ_4 and ξ_5. We find that in quantifications Q_2 and Q_3 the posterior is significantly more sensitive to parameter variation for profiles ξ_3 and ξ_4 than for profile ξ_5. An explanation for this pattern exists in the fact that profile ξ_5 provides several independent pieces of evidence indicating a large shunt; varying individual parameters therefore hardly influences that posterior. In contrast, profiles ξ_3 and ξ_4 comprise contradicting observations; in these cases, varying a single parameter can change the posterior considerably. The result is not found for the quantification Q_1, due to the uniform probability distributions it comprises.

Using the results of the sensitivity analyses, quantifications Q_1 and Q_2 were stepwise

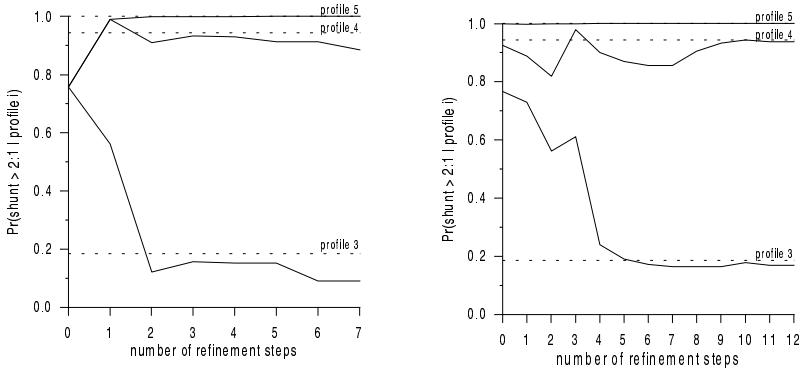

Figure 2: The posteriors of (a) quantification Q_1 and (b) quantification Q_2 after refinement. The dotted horizontal lines correspond to the posteriors of quantification Q_3.

refined with parameter estimates from Q_3. Although sensitivity analysis reveals the influence of individual parameters, we chose to substitute, at every refinement step, all parameter estimates pertaining to a network variable. The motivation is that a parameter often has little meaning in isolation: it is its relation to other parameters from the same local distribution that matters. Furthermore, there usually exists a substantial overlap in the variables to which highly influential parameters pertain; this was also found in the present analyses.

The following procedure was now used to select variables whose parameters were eligible for substitution. First, for each case profile ξ_i, $i = 3, 4, 5$, we identified the thirty parameters showing the largest influence on $P_{i,j}$, and the variables to which these parameters pertained. Subsequently, these sets of influential variables per profile were compiled to a single, ordered set; the order was determined by averaging the positions of the variables in the original sets. In the resulting selection for quantification Q_1, consisting of 7 variables, only the measurement variable itself (`shunt`), and observed, direct descendants of this variable were selected. This is not surprising, as the uniform distributions used in this quantification eliminate all influences through longer pathways in the graph when only one parameter estimate is varied at a time. The selection for quantification Q_2 (12 variables) also comprised variables at a greater distance of the `shunt` variable, and ascendants of `shunt` in the graph. As such, the selection for quantification Q_2 seemed to provide a more realistic pattern of influential variables.

4.3 Predictions of the refined network quantifications

The next stage in our investigation was to refine quantifications Q_1 and Q_2, by stepwsie replacing the parameter estimates of variables in the ordered selections with estimates from Q_3. With each of the refined quantifications, the posterior $P_{i,j}$ was computed for case profiles ξ_3, ξ_4, and ξ_5; these posteriors are plotted in Figure 2. The plots indicate that for both quantifications, the posteriors rapidly shift towards the posteriors of quantification Q_3. However, the difference between posteriors does generally not decrease monotonically. This is most notable with the refinements of quantification Q_2: the posteriors for profiles 3 and 4 show considerable fluctuations in the first seven steps. Thereafter, they quickly converge to the desired level.

Unfortunately, this convergence is not obtained for quantification Q_1: even after replacing the distributions of all seven variables with influential parameters, the posteriors still deviate from the posteriors of Q_3.

To investigate whether the results generalise over more cases, the effects of refinements were also tested on real-world data: 36 cases were selected from a clinical database of VSD patients collected at the Leiden University Medical Center in The Netherlands. For each case, the posteriors of quantifications Q_1 and Q_2 both before and after the refinements were compared with the posteriors of quantification Q_3. Unfortunately, the results of the refinements were not unequivocal. In short, it was found that for some cases in the database, our procedure selected inappropriate variables to refine; this was probably due to the small set of case profiles used in the sensitivity analyses. Due to space limitations, we cannot elaborate on this investigation here; the interested reader is referred to [3].

5 Discussion

We have presented an empirical investigation of a quantification method for belief networks, where the effort is focused on influential network parameters. These parameters were found by performing sensitivity analyses. The results of our investigation suggest that the method is promising: partially refined, poorly-informed quantifications give predictions that are comparable to a well-informed quantification. However, the procedure showed better results for the small set of case profiles that was used to identify influential parameters, than for the larger set of case from a real-world clinical database. We conclude that it is preferable to also use a large set of real-world cases when identifying influential parameters.

For the quantification, we have used subjective probability estimates. The proposed procedure is not limited to subjective estimates though; it is equally well applicable to statistics from data sets, frequencies reported in the literature, or a combination of sources. The method is however limited to finding the effects of *individual* parameter variations on a network's predictions, as the influential parameters are found by performing one-way sensitivity analyses. To reveal synergetic effects of varying multiple parameters, higher-order sensitivity analyses, or *uncertainty analyses*, which investigate the effect of varying all network parameters simultaneously, are required.

The main shortcoming of the current procedure is that the effects of refinements are non-monotonic; it is therefore difficult to establish a stopping criterion. In the future, we plan to investigate whether monotonic improvement of network behaviour is feasible, and if so, under what conditions. Further sophistication of the procedure is envisioned in alternating sensitivity analyses and quantification refinement, and in using the expected accuracy of parameter estimates as expressed by confidence intervals.

Acknowledgements

The investigations were (partly) supported by the Netherlands Computer Science Research Foundation with financial support from the Netherlands Organisation for Scientific Research (NWO). The authors wish to thank Robert Cannon for valuable discussions, and Linda van der Gaag, Dik Habbema, Peter Lucas and Silja Renooij for their useful comments on the manuscript.

References

[1] V. Coupé and L. van der Gaag. Practicable sensitivity analysis of Bayesian belief networks. *Technical Report UU-CS-1998-10*, Utrecht University, 1998.

[2] V. Coupé, L. van der Gaag, and D. Habbema. Sensitity analysis: an aid for belief-network quantification. *Technical Report UU-CS-1999-13*, Utrecht University, 1999.

[3] V. Coupé, N. Peek, J. Ottenkamp, and D. Habbema. Using sensitivity analysis for efficient quantification of a belief network. *Artificial Intelligence in Medicine*, accepted for publication.

[4] M. Druzdzel and L. van der Gaag. Elicitation of probabilities for belief networks: combining qualitative and quantitative information. In *Proc. 11th Conf. Uncertainty in Artificial Intelligence*, pp. 141–148, 1995.

[5] T. Graham and H. Gutgesell. Ventricular septal defects. In A. Moss, F. Adams et al. (eds), *Heart Disease in Infants, Children, and Adolescents*, pp. 724–746. Williams & Wilkins, 1995.

[6] D. Habbema, P. Bossuyt, and D. Dippel. Analysing clinical decision analyses. *Statistics in Medicine*, 9:1229–1242, 1990.

[7] M. Morgan and M. Henrion. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis.* Cambridge Univ. Press, 1990.

[8] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, 1988.

[9] N. Peek and J. Ottenkamp. Developing a decision-theoretic network for a congenital heart disease. In E. Keravnou, C. Garbay, R. Baud, and J. Wyatt (eds), *Proc. 6th Conf. Artificial Intelligence in Medicine Europe*, pp. 157–168, Springer, 1997.

[10] M. Pradhan, M. Henrion, G. Provan, B. Del Favero, and K. Huang. The sensitivity of belief networks to imprecise probabilities: an experimental investigation. *Technical Report SMI-95-0595*, Stanford University, 1995.

[11] D. Spiegelhalter, R. Franklin, and K. Bull. Assessment, criticism and improvement of imprecise subjective probabilities for a medical expert system. In M. Henrion, R. Shachter, L. Kanal, and J. Lemmer (eds), *Uncertainty in Artificial Intelligence* 5, pp. 285–294, North-Holland, 1990.