drHugin
A system for value of information in Bayesian networks*

Finn V. Jensen & Jianming Liang
Department of Mathematics and Computer Science
Aalborg University
Fredrik Bajers Vej 7 E2
DK-9220 Aalborg, Denmark
e-mail: fvj@iesd.auc.dk; liang@iesd.auc.dk

Abstract

drHugin, which is an extension of Hugin, deals with value of information in Bayesian networks. The approach is constructed for situations where one variable is in the focus of interest, and the probability distribution for that variable drives the system. In this framework both utility based and utility free assessments of information sources are available.

Keywords

Bayesian Networks, Data Request, Utility-based Values, Utility-free Values, Value of information.

1 Introduction

Whenever decisions under uncertainty are to be made, there is a quest for more information to reduce the uncertainty. However, information is rather seldom cost free, and therefore there is also a need for evaluating on beforehand whether it is worthwhile to consult an information source. Furthermore, if several sources are available there is a need to come up with a strategy for a sequence of data requests.

The problem of data request has been formally treated in decision theory (Howard 1966, Lindley 1971, Shachter 1986) where utilities of possible actions are guiding the request decisions. Also utility-free assessments of information sources have been studied (Ben-Bassat 1978, Pearl 1988), where Shannon's measure of mutual information and variance are used as a measure of impact of new information.

The system drHugin is an implementation of a data request technique based on Bayesian networks, and it is an extension of the software system Hugin (Andersen et al. 1988). Readers not familiar with Bayesian networks and the Hugin approach are referred to the literature (e.g. Pearl 1988 and Jensen et al. 1990).

2 A constructed example

A certain treatment against childlessness has a success rate of 0.6, and out of all treatments 15% result in twins, 4% result in 3 embryos, 1% yield 4 embryos. Four weeks after treatment there are three tests for the result (blood, urine, scanning) – the first two with 4 outcomes and the last with 5 outcomes. The results of the blood test as well as the urine test are effects of hormonal changes due to a possible pregnancy. Finally, clinical findings like absence of menstruation, nausea etc. are also indications of pregnancy. See figure 1 for a model. The

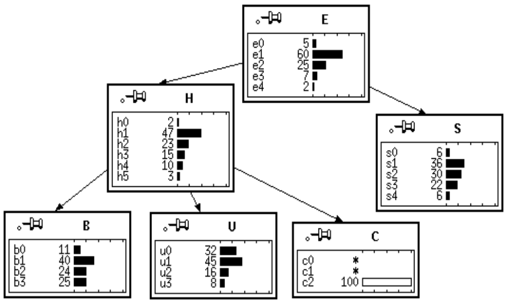

Figure 1: A model for the treatment example with probabilities posterior to the clinical findings indicated

conditional probabilities are listed in Appendix 1.

A woman has been treated, and after four weeks the clinical findings are strongly favoring pregnancy (the variable C is in state **c2**) as shown in

*This work is partially funded by the Danish Research Councils through the PIFT programme.

figure 1. From the resulting probabilities, it can be seen that it is pretty certain that the treatment has been successful. However, the woman is curious to know how many children to expect. Which test should be performed next, if any?

3 Hypothesis driven myopic data request

The *myopic data request* problem can be formulated as follows:

Information may be acquired through various information sources each with a certain cost. If you are allowed to consult at most one information source, which option should be chosen?

Working with Bayesian networks, an exhaustive simulation could give the state of the network for each possible outcome of all options. The number of propagations in such a simulation equals the total number of possible outcomes, and in many practical situations it will require too much time. Furthermore, the evaluation of the state of the network may also be very complex.

In drHugin we constrain ourselves to the following less complex situation, which we call *hypothesis driven*:

The evaluation of the state of the network is based only on the probability distribution of a hypothesis

To be precise: H is a *hypothesis* with a set of mutually exclusive states $h_1, \ldots, h_n$ and V is a *value function* $V(H) : [0;1]^n \rightarrow R$.

$S, \ldots, T$ is a set of *tests* with discrete sets of possible outcomes and with cost $C_S, \ldots, C_T$. We assume that the outcomes of the tests can be entered as evidence to the network and that $P(S), \ldots, P(T)$ are easy to access from the network.

The *expected value* from performance of test T is

$$EV(T) = \sum_{t \in T} V(P(H \mid t))P(t)$$

The *expected benefit* of performing test T is

$$EB(T) = EV(T) - V(P(H))$$

The *expected profit* is

$$EP(T) = EB(T) - C_T$$

The myopic data request task is to calculate the expected profit for the various tests and choose the one with maximal expected profit (if positive).

To calculate expected values, one should determine $P(H), P(S), \ldots, P(T)$ as well as $P(H \mid S), \ \ldots, P(H \mid T)$. Now, instead of propagating all possible outcomes of the tests, Bayes' theorem will yield the necessary probabilities through $n-1$ propagations only:

$$P(h_i \mid S) = \frac{P(S \mid h_i)P(h_i)}{P(S)} \text{ for } i = 1, \ldots, n-1$$

$$P(h_n \mid S) = 1 - (\sum_{i=1}^{n-1} P(h_i \mid S))$$

where $P(S \mid h_i), \ldots, P(T \mid h_i)$ are obtained for $i = 1, \ldots, n-1$ by entering $H = h_i$ as a finding to the network.

By constraining ourselves to hypothesis driven data request we not only focus on situations, where the evaluation of the state of the network is narrowed down to a small part of it; but the real reason for doing so is that in most practical situations, the number of hypothesis alternatives is very small compared to the number of possible test outcomes.

3.1 Joint request

Some tests may yield findings on several variables. In the childlessness example it could be that a laboratory test decision involves both a blood test and a urine test.

If T yields findings on the variables $V_1, \ldots, V_n$ then

$$EV(T) = \sum_{\underline{v} \in V_1 \times \ldots \times V_n} V(P(H \mid \underline{v}))P(\underline{v}).$$

Therefore $P(\underline{v} \mid H)$ and $P(\underline{v})$ are needed for all configurations $\underline{v} \in V_1 \times \cdots \times V_n$.

In Hugin, these probabilities are easy to get if $V_1, \ldots, V_n$ happen to be members of the same clique in the triangulated graph. This is exploited in drHugin by forcing $V_1, \ldots, V_n$ into the same clique.

3.2 Compound hypotheses

It may happen that the hypothesis is a compound expression over several variables. The focus of interest may be whether or not the variable A is in state a and B is in state b.

In drHugin compound hypotheses which are conjunctions or disjunctions of single variable states can be handled. In the example above the hypothesis is the conjunction $(A = a) \wedge (B = b)$. The two findings $(A = a)$ and $(B = b)$ are entered to the network yielding $P(T|A = a, B = b)$ for all tests T. Also $P(A = a, B = b)$ is needed for the analysis, but it is obtained from the propagation procedure as the normalizing constant. In (Jensen 1991) a more detailed treatment of compound hypotheses is given.

4 A Session with drHugin

Pressing the **Analyse** button in the Hugin Browser window causes a specification window (see Figure 2) to pop up. In our example, the hypothesis is specified to be one variable, namely **H**, and the tests are **S, B, U**, and **Lab Test (B&U)**. For a joint request, like Lab Test, the user must create a special variable with only one state and with the requested variables as parents. Then the variables are forced to be present in the same clique. There are several options for value function, we choose **Entropy**. Now we press the **Apply** button and the expected benefits from the tests are calculated (see Figure 2).

5 Value functions

In principle, any value function may be used and it is up to the user to provide it. However, a particular class of functions, namely, *convex* functions are best suited.

Definition: A function $f : R^n \rightarrow R$ is *convex* if the following holds: $\forall t \in [0,1], \forall \underline{x}, \underline{y} \in R^n : tf(\underline{x}) + (1-t)f(\underline{y}) \geq f(t\underline{x} + (1-t)\underline{y})$.

In words it means that for any two points P_1, P_2 on the graph of f, the line segmentP_1P_2 lies above the graph. The reason why a convex function is well suited is due to the following:

Theorem 1: If the value function is a convex function then the expected benefit of perform-

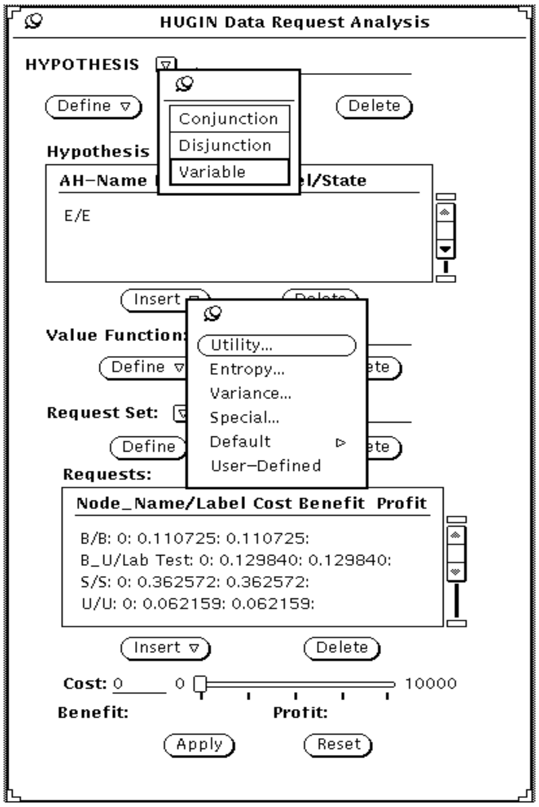

Figure 2: The specification popup window

ing a test is never negative (proof in Jensen et al 1992).

Below we list value functions provided by drHugin. They are all convex (proofs in Jensen et al 1992).

5.1 Entropy

The formula for entropy of a distribution over H is

$$ENT(P(H)) = -\sum_{h \in H} P(h) log_2(P(h))$$

where $plog_2p = 0$ if p vanishes.

Entropy is a measure of how much the probability mass is scattered around on the states. If the distribution is even, and the number of states in H is n, then the entropy is $log_2 n$. In the limit, if the probability mass is concentrated in one state, the entropy is 0. In order to keep the idea to let high values be preferred, we let an entropy based value function be

$$\begin{aligned} V(P(H)) &= -ENT(P(H)) \\ &= \sum_{h \in H} P(h) log_2(P(h)) \end{aligned}$$

For the childlessness example in section 2, $EB(U) = 0.06, EB(B) = 0.11$, $EB(S) = 0.36$ and $EB(lab - test) = 0.13$.

5.2 Variance

If the states of H are numeric, another measure can be used, namely the *variance*. Again, since small variances are preferred, the value function becomes

$$V(P(H)) = -\sum_{h \in H}(h - \sum_{h' \in H} h'P(h'))^2 P(h),$$

For the childlessness example we get $EB(U) = 0.05, EB(B) = 0.08$, $EB(S) = 0.15$ and $EB(lab - test) = 0.09$.

5.3 Utility based functions

In many cases, the value of the hypothesis is due to a set A of *actions* which are taken dependent on the state of the hypothesis. In the childlessness example, the reason for testing may be more than plain curiosity. In the case of 4 or more embryos an induced abortion should be considered. Also, if the treatment is unsuccessful a new treatment should be started soon. In both cases the action is the same: A possible pregnancy is interrupted and the treatment is repeated. The rationale for choosing among the actions is a *utility function* $U : H \times A \rightarrow R$ describing what will be gained by the actions given the various states of the hypothesis. The *expected utility* of action a is

$$EU(a) = \sum_{h \in H} U(h, a)P(h|a),$$

and the action chosen is the one with maximal expected utility.

Under the assumption that the actions have no impact on the hypothesis, the maximal expected utility can be used as a value function:

$$V(P(H)) = \max_{a \in A} \sum_{h \in H} P(h)U(h, a)$$

Theorem 2: The expected benefit of performing test T is 0 if and only if the optimal action is unchanged no matter the outcome of T (proof in Jensen et al 1992).

Notice that the action of selecting the state of H with highest probability can be modelled by utilities yielding a value function equivalent to

$$V(P(H)) = \max_{h \in H} P(h)$$

and $EV(T)$ is then positive if and only if some outcome of T causes a new state to have highest probabilities.

Suppose in the childlessness example that the utilities are as in table 1 (we will not address the rather problematic problem on how to assess the utilities): The expected benefits for B and U are then 0, meaning that no outcome of the laboratory test will change the optimal decision (which is no-action). $EB(S)$ is 7.09, so if the cost of performing scanning is less than 7.09 it is recommendable to do that.

	0	**1**	**2**	**3**	**4**
Action	200	-300	-500	-100	100
No action	-100	500	700	200	-100

Table 1: Utilities for embryos and repeated treatment

5.4 Other value functions

As mentioned already, it is up to the user to specify the value function. If actions with known utilities are attached to the hypothesis then the utility value function should be preferred. If this is not the case, the user will mainly be interested in the precision of a diagnosis.

In e.g. the case of a Boolean hypothesis with states **0** and **1**, the entropy function is $\log p^p(1-p)^{1-p}$ and the variance function is $-p(1-p)$. Theses two functions reflect that the value of p increases as it approaches its borders 0 and 1. The entropy function is rather drastic in the way that the slope is infinite for 0 and 1. Therefore, small changes of p close to 0 and 1 will be highly valued. The variance is of polynomial degree 2, and the slope close to the borders is 1 and -1, giving changes almost even value no matter how close to the border.

Many other convex functions can be used. They can e.g. reflect asymmetry between **1** and **0**, or they can have other kinds of marginal values close to the borders. In drHugin the following options are provided for Boolean hypotheses: $\left|A + B \cdot p\right|$, $B \cdot \left|\frac{1}{A} - p\right|^N$, $\frac{A}{p^N \cdot (1-p)^M}$, $A \cdot \left|\log \frac{p}{1-p}\right|^N$, where p is the probability of the hypothesis, and A,B,N,M are parameters.

6 Non-myopic data request

It may happen that no test has a positive expected profit while the expected benefit for a pair of tests is beyond the costs of them. In that case the myopic approach is misleading. Consider the following simple example:

A disorder H with prior probability 0.02 has two tests T and S each with a 10% risk of giving false positive as well as false negatives. The value function is $\max_{h \in H} P(H)$ (the one originating from selecting the state with highest probability as the diagnoses). The causal relations are given in Figure 3.

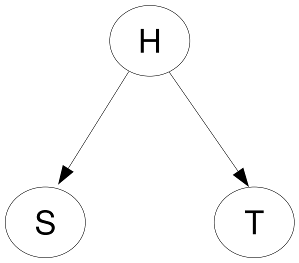

Figure 3: The causal relations between disorder and tests in the non-myopic example

Initially, $V(P(H)) = 0.98$ with *no* as the chosen diagnosis. According to theorem 1 a test will have a positive expected benefit only if some test result gives *yes* highest probability. This may only happen for test result *yes*. However, $P(H = yes|T = yes) = 0.16$. So no single test will have a positive expected benefit.

On the other hand, $P(H = yes|T = yes, S = yes) = 0.62$ which means that the expected benefit from performing both tests is positive. This highlights that the value of performing a test is not only the change of value. Performing a test has the side effect that it changes the expected benefit from performing other tests. Therefore, the options in this example are
T and ?S: Observe T and decide afterwords whether to observe S;
S and ?T: Observe S and decide afterwords whether to observe T;
T & S: Observe T and S.

Unless costs may be saved by performing both test simultaneously, **T & S** can never be better than the first two options. We analyse **T and ?S**.

The expected profit from observing T is
$EP(T) = \sum_{t \in T} V(P(H|t))P(t) - Cost(T) - V(P(H))$.
If the test result is t then the expected profit of observing S is
$EP(S|t) = \sum_{s \in S} V(P(H|s,t))P(s|t) - Cost(S) - V(P(H|t))$.

If $EP(S|t) \leq 0$ then S is not performed. Hence $EP(Tand?S) =$
$EP(T) + \sum_{t \in T} P(t) \max(0, EP(S|t))$.

To be able to calculate $EP(Tand?S)$ we in addition to the probabilities from myopic data request also need $P(H|S,T)$ and $P(S|T)$, and in an analysis of **S and ?T**, we would furthermore need $P(T|S)$.

Since $P(S,T) = \sum_{h \in H} P(S,T|H)P(H)$, we can by Bayes' formula conclude that if $P(S,T|H)$ is available, all sufficient information will be present.

If T and S are not members of the same clique, the number of propagations to achieve $P(T,S|H)$ is the product of the number of states in the two smallest of the variables T, S and H. Furthermore, if more tests are available the number of propagations grows very fast even for a non-myopic analysis involving only two tests at a time.

However, sometimes the situation is not that bad. If, as is the case in our example, S and T are independent given H, then $P(T,S|H) = P(T|H) \times P(S|H)$ and the analysis can be performed through propagation of the states of H.

By setting costs of each test to 0.005 we get $EP(Tand?S) = 0.0014$. To compare, we have $EP(T\&S) = -0.0037$.

In drHugin, the following strategy is used:

- First, select tests through myopic data request;
- If no single test has a positive expected profit then perform a non-myopic analysis for all pairs of tests which are members of the same clique or are independent given the hypotheses;
- If none of the pairs above have a positive expected profit, perform an approximate analysis of the remaining pairs by assuming that they are independent given the hypotheses.

References

Andersen, S.K., Olesen, K.G., Jensen, F.V., Jensen, F. (1989). HUGIN – A shell for building Bayesian belief universes for expert systems. In *Proceedings of the 11th international joint conference on artificial intelligence*, pp. 1080-85.

Ben-Bassat, M. (1978). Myopic policies in sequential classification. *IEEE Transactions of Computing, 27*, pp 170-74.

Howard, R.A. (1966). Information Value Theory. *IEEE Transactions on Systems Science and Cybernetics*, **ssc-2**, pp 22-26.

Jensen, F.V. (1991). Calculation in HUGIN of probabilities for specific configurations – a trick with many applications. In *Scandinavian conference of AI*, pp. 176-86.

Jensen, F.V. and Liang, J (1992). drHugin: A system for hypothesis driven myopic data request. Technical Report R-92-2021, Department of Mathematics and Computer Science, Aalborg University.

Jensen, F.V., Olesen, K.G. and Andersen, S.K. (1990). An algebra of Bayesian belief universes for knowledge-based systems. *Networks*, **20**, pp. 637-59.

Lindley, D.V. (1971). *Making Decisions.* John Wiley & Sons.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann.

Shachter, R.D. (1986). Evaluating Influence Diagrams. *Operation Research,* **34**, pp. 871-82.

Appendix 1 — Tables for the childlessness model

$$P(E) = (0.4, 0.4, 0.15, 0.04, 0.01)$$

$P(H \mid E)$

	h_0	h_1	h_2	h_3	h_4	h_5
e_0	0.95	0.03	0.01	0.01	0	0
e_1	0	0.7	0.15	0.1	0.05	0
e_2	0	0.3	0.4	0.15	0.1	0.05
e_3	0	0.2	0.3	0.2	0.2	0.1
e_4	0	0.1	0.2	0.3	0.2	0.2

$P(B \mid H)$

	b_0	b_1	b_2	b_3
h_0	0.9	0.1	0	0
h_i	0.2	0.7	0.1	0
h_2	0	0.3	0.6	0.1
h_3	0	0	0.3	0.7
h_4	0	0	0.1	0.9
h_5	0	0	0	1

$P(U \mid H)$

	u_0	u_1	u_2	u_3
h_0	0.9	0.1	0	0
h_i	0.5	0.5	0	0
h_2	0.2	0.5	0.2	0.1
h_3	0.1	0.4	0.4	0.1
h_4	0	0.3	0.4	0.3
h_5	0	0.2	0.4	0.4

$P(C \mid H)$

	c_0	c_1	c_2
h_0	0.9	0.08	0.02
h_i	0.2	0.3	0.5
h_2	0.2	0.2	0.6
h_3	0.1	0.2	0.7
h_4	0.1	0.1	0.8
h_5	0.05	0.1	0.85

$P(S \mid E)$

	s_0	s_1	s_2	s_3	s_4
e_0	0.98	0.01	0.01	0	0
e_1	0.01	0.5	0.25	0.2	0.04
e_2	0.01	0.2	0.5	0.2	0.09
e_3	0.01	0.1	0.25	0.5	0.14
e_4	0.01	0.04	0.15	0.4	0.4