The Value of Relevance

Robert J. Korsan
Decisions, Decisions!
2136 Lyon Avenue
Belmont, California 94002-1639
bobk@Decide.com

Abstract

The highest cost technical activity in decision analysis, knowledge engineering and expert systems development is probabilistic assessment. Influence diagram (ID) technology has reduced assessment and improved decision models dramatically. However, once an ID has been assessed from the decision-maker (DM) and deterministic sensitivity analysis performed, a method to price the relevance arrows in the ID would allow the DM to balance the cost of assessment against the improvement in the authenticity of the value lottery and accuracy of the recommended action.

This paper outlines such a method to price relevance arrows in an ID contianing discrete uncertainties and discrete decisions. The method also determines the sensitivity of a decision to the probability of particular outcomes of any uncertain event.

Introduction

Rule-based schemes for knowledge-based expert systems have been shown to have critical flaws in their treatment of uncertainty, decisions and values. Bayesian methods such as decision analysis offer an appealing basis for solving these problems. However, Bayesian methods are often dismissed as intractable because they require extensive assessment of expert knowledge, usually not available from experimental data, and because of the complexity of Bayesian calculations based on models with realistic numbers of variables. This paper shows how we can determine which information must be preserved and which information can be treated as irrelevant, thus making Bayesian methods more tractable.

This paper outlines the steps of the procedure and presents an application to a small decision problem. The price of a relevance arrow is obtained in 4 steps:

1. Assess an influence diagram, A, from the decision-maker. This influence diagram bounds the complexity of the decision problem. For all chance nodes in A assess the *marginal* probability density functions of the uncertain quantity.
2. Systematically generate all influence diagrams which are simpler than A by removing relevance arrows. This collection of influence diagrams is called the *family of A*.
3. For each influence diagram in the family of A, compute the *merit of the influence diagram.* The merit of an influence diagram is the difference between the upper bound and lower bound on the value of the decision to the decision-maker.
4. To compute the *value of relevance* of a set of relevance arrows in A: remove the arrows from A yielding A' in the family of A and compute the merit of A minus the merit of A'.

The following sections will provide the details of this procedure.

The Value of Relevance

The usual steps in decision analysis are to 1) structure the decision problem and determine an influence diagram (ID)[Howard (1990)], 2) build a deterministic model of the decision problem, 3) perform sensitivity analysis, 4) assess the joint distribution over all uncertain events and quantities 5) determine the value lottery and recommend action. Step 4 is one of the highest cost technical activities in decision analysis. If the ID is complex, then the decision-maker (DM) would like to know, before assessing anything, that the costly assessment of conditional relevance will develop a more authentic value lottery and a better decision.

If a DM knows that assessing a particular joint distribution does not change the recommended action or increase the authentic value lottery by more than an amount \$ δv, then the cost of assessing and determining the optimal policy given the assessment can be balanced against this potential gain. If δv is small enough, the DM will decide to ignore the relevance arrow in the ID.

I adopt the view that computational expense is negligible compared to the cost of assessment. The method developed in this paper minimizes the amount of information which must be assessed in order to compute the value of relevance (VOR) for a set of arrows in an ID.

Although the algorithm has been extended to continuous distributions and continuous decisions, this paper is restricted to discrete uncertainties and discrete decisions and extends previous work on better sensitivity methods [Korsan (1990)]. A comparison with [Lowell (1994)] will be the topic of a future paper.

1: Information Requirements

What is a minimal set of requirements for computing VOR? The following procedure assumes 1) an assessed ID, denoted A, describing the decision problem, 2) a value model, $v(s,d)$, where s are the state variables $(s_1,\ldots,s_n)$ and d are the decision variables $(d_1,\ldots,d_m)$, 3) the DM's utility function $u(v)$ and 4) the marginal probability density functions (PDFs) of the state variables, i.e. $\{s_1|\mathcal{S}\},\ldots,\{s_n|\mathcal{S}\}$. $\mathcal{S}$ is the DM's current state of information.

Assuming that we have performed sensitivity analysis, the only added cost is the assessment of marginal distributions for each of the state variables. This appears to be the minimal burden. When conditional distributions must be assessed, the marginal distributions may be used to reduce the number of assessments needed. Thus the assessment order is changed but not the amount of assessment. When joint distributions are eliminated from assessment, there is a net cost savings.

2: A *Family* of Influence Diagrams

The ID, A, is called the *Bounding Influence Diagram* (BID). The BID represents an upper bound on the complexity of the decision models considered. The simplest possible ID treats all uncertainties as independent. The collection of all possible IDs obtained from A by removing some subset of arrows is called the BID-Family and denoted $Fam(A)$. The collection of IDs, $Fam(A)$, can be generated as follows:

1. Number all relevance arrows in an ID as $1,\ldots,n$.
2. For $m=0,\ldots,2^n-1$ generate the binary expansion $m=a_0 2^0+\ldots+a_{n-1}2^{n-1}$.
3. $A_m \in Fam(A)$ is obtained by removing the i-th arrow if $a_i=0$ from the BID.

Remark 1 *a) A fast algorithm to generate the BID-Family is based on the Gray Code. b) $A_0 \in Fam(A)$ treats all uncertain events as independent. c) $A_{2^n-1} \equiv A$.*

The BID-Family is an exhaustive description of consistent decision models with simpler relevance structure than the BID.

3: The *Merit* of an Influence Diagram

Given the current state of information of the decision analyst (prior to further assessment), how can we associate a value or range of values with each member of the BID-Family? The problem is that the joint distribution is unassessed. However, we can assume that the DM is coherent and restrict the space of consistent joint distributions.

The Space of Joint PDFs Given $A_m \in Fam(A)$ the state variables may be partitioned into two sets: 1) the collection of independent variables (relabeled as $s_1,\ldots,s_k$) and 2) the remaining state variables $(s_{k+1},\ldots,s_n)$ for which the joint distribution is unknown. I assume that the DM is coherent, i.e. if I assess a PDF from a DM and then perform various probabilistic operations on the assessed distribution, then the resulting PDF obtained will be identical to the PDF obtained by assessing the tranformed object directly from the DM. In particular, I assume that marginalizing an assessed joint distribuition would produce marginal PDFs identical with those assessed from the DM. From the collection of all possible joint distributions over $s_{k+1},\ldots,s_n$ we restrict ourselves to those which satisfy coherence constraints, i.e.

$$\{s_\nu|\mathcal{S}\} = \sum_{i\neq\nu}\{s_{k+1},\ldots,s_n|\mathcal{S}\}\ \nu=k+1,\ldots,n.$$

With each $A_m \in Fam(A)$ I associate a space of joint distributions over the state variables subject to the coherence constraints. This space is denoted $\Gamma(A_m,s_1,\ldots,s_n)$. This space consists of all distributions consistent with the desiderata of the decision problem and the information obtained from the DM.

ID Value Bounds For any $A_m \in Fam(A)$, consider the triple $(\Gamma(A_m,s_1,\ldots,s_n),u(v),v(s,d))$. For each joint distribution in $\Gamma(A_m,s_1,\ldots,s_n)$, the decision problem may be solved and an expected utility found. This collection of expected utility values is contained in an interval $[\underline{u}_m,\overline{u}_m]$. There are two joint distributions which are of particular interest. These are the distributions which prduce the extremal values $\underline{u}_m,\overline{u}_m$.

Using Game Theory We now proceed to isolate the distributions and determine the values $\underline{u}_m$ and $\overline{u}_m$. Although these distributions and the values themselves are unknown, we know that they are the stochastic strategies of an opponent in the solution of two games. The upper bound, $\overline{u}_m$, corresponds to the value of a game against a cooperative opponent. The highest value attained by cooperation is the upper bound [Luce and Raiffa (1957)]. The opponent's strategy must match the coherence constraints. The joint distribution over all uncertain events corresponds to the stochastic strategy which the cooperating opponent employs. The information we have required is sufficient to specify and solve this game. Thus we can determine both the upper bound and the extremal distribution in $\Gamma(A_m,s_1,\ldots,s_n)$. Denote the upper extremal distribution as $\overline{\{s|\mathcal{S}\}}_m$.

Similarly, the lower bound, $\underline{u}_m$, is the value of a game against the DM's worst enemy. The strategy is similarly constrained by the coherence constraints.

The joint distribution over the uncertain events corresponds to the stochastic strategy employed by the opponent. Thus we can also determine both the lower bound and the other extremal distribution in $\Gamma(A_m, s_1, \ldots, s_n)$. Denote the lower extremal distribution as $\underline{\{s|S\}_m}$.

For any influence diagram in the BID-Family we can determine the upper and lower bounds on the value of that ID.

Defining the *Merit* of an ID For any ID, $A_m \in Fam(A)$, in the BID-Family we define the *merit* of the ID to be

$$\mu(A_m) = CE(\overline{u}_m) - CE(\underline{u}_m),$$

where $CE(\bullet)$ is the *certain equivalent* of the expected utility. If the utility function does not map a denominated value function or numerare, then $\mu(A_m)$ is defined directly as the difference in utilities. The merit of an ID represent the range of value/utility which may be obtained using the decision model represented by the ID and which satisfies the coherence constraints.

4: Computing Value of Relevance

The VOR is defined as the amount to be gained/lost by using a more complex/simpler model of the decision problem. Complexity means the addition or subtraction of a set of arrows from the reference ID. Given two IDs, $A_j, A_k \in Fam(A)$ the value of relevance of A_j relative to A_k is $\delta_j^k v = \mu(A_k) - \mu(A_j)$.

The difference in merit of the two IDs is the value of relevance for the arrows removed (added) from the reference ID.

Collections of Relevance Arrows Given two IDs, $A_j, A_k \in Fam(A)$, we say $A_j \sqsubseteq A_k$ if all arrows of A_j are contained in A_k. $A_j \equiv A_k$ if $A_j \sqsubseteq A_k$ and $A_k \sqsubseteq A_j$. $A_j \sqsubset A_k$ if $A_j \sqsubseteq A_k$ and some arrows in A_k are not in A_j. When $A_j \sqsubset A_k$, the arrows in A_k not in A_j are denoted $Arrows(A_k \setminus A_j)$.

Pricing Collections of Arrows There are two possible ways to price collections of arrows in the BID-Family. The first, and less useful, is to compute VOR for all possible collections of arrows relative to A_0, i.e. the ID in which all uncertainties are treated as independent. This views added arrows as providing added value. Since the DM has asserted that relevance exists, this reference ID has little value.

The second way to price collections of arrows in the BID-Family is relative to the BID itself. Suppose removing some collection of arrows from the BID, i.e. from A, produces the ID A_m. Then the arrows removed are $Arrows(A \setminus A_m)$ and the VOR of those arrows is $VOR(A \setminus A_m) = \mu(A) - \mu(A_m)$.

Since all possible simplifications of the decision model are represented by $Fam(A)$, the amount potentially lost by using any simpler model is captured by this calculation.

A Particular Ordering We can sort $Fam(A)$ in increasing order. Removing no relevance arrows will produce zero loss. Removing arrows one at a time will produce small losses. Removing arrows two at a time will produce larger losses. And so forth.

The Irrelevance of Relevance

If the DM is willing to tolerate some error, \$ ϵ, in the authenticity of the value lottery, then we can use this amount to perform a sensitivity to relevance. Using the sorted order of the last section, choose the ID, A_ν, which maximizes $VOR(A\backslash A_\nu)$ and has $VOR(A\backslash A_\nu) \leq \epsilon$. The ID so chosen represents the simplest model of the decision subject to the coherence constraints and the willingness of the DM to tolerate error in the authenticity of the value lottery.

The space of joint distributions of the state variables is convex. Hence we can also determine the stability of the recommended policy as we traverse convex combinations of the extremal distributions. If the policy is unstable during this traversal we should be hesitant to use the indicated simplification.

An Example

Suppose we are building a diabetes diagnosis and treatment decision support system. We don't know if the patient has diabetes. The symptoms are "blue toe" and glucose present in the urine of the patient. We have assessed the BID shown in figure 1.

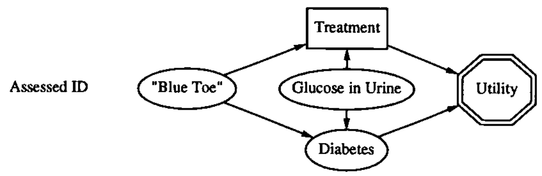

Figure 1: The BID shows the most complex model of the decision.

Suppose we also assess the marginal distributions

Diabetes		"Blue Toe"		Glucose	
d	0.10	b	0.0068	g	0.099
¬d	0.90	¬b	0.9932	¬g	0.901

and the doctor's utility function.

$U(t,d)$	Action: Treat t	Action: Don't treat ¬t
has diabetes (d)	10	0
no diabetes (¬d)	5	10

The BID-family is shown in figure 2.

Using the procedure developed in this paper we have the following results.

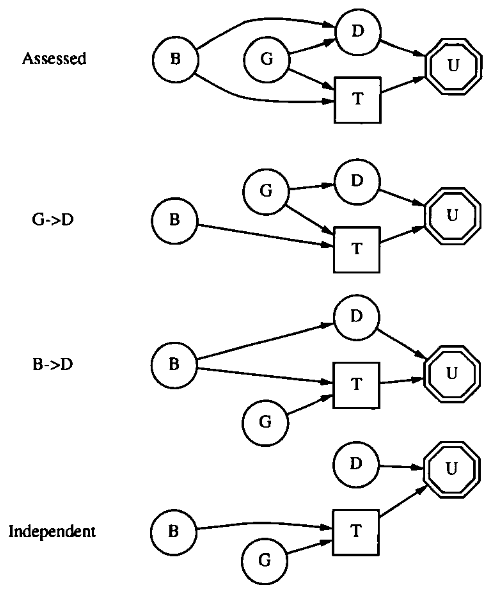

Figure 2: The BID-Family shows all possible simplifications of the decision.

$Fam(A)$	$\underline{u}_n$	$\overline{u}_n$	Merit $\mu(A_n)$	VOR	Arrow(s) Priced
A_3	9.00	10.00	1.00	0	None
A_2	9.00	9.99	0.99	0.01	$B \to D$
A_1	9.00	9.07	0.07	0.97	$G \to D$
A_0	9.00	9.00	–	1.00	Both

We see that there is nothing to be gained by including both relevance arrows. Almost all of the value can be obtained by using the simpler model based on glucose in the urine alone. The relevance arrow from blue toe to diabetes is priced at 1% of the value of all relevance and 0.01% of the maximum value of the decision overall. This example was taken from the literature [Shenoy (1990)] where the full joint distribution was available. Each member of the BID-Family was solved using the appropriate marginalization of the full joint distribution. The value of the strategy based on the full joint is not measurably different from the strategy based on removing the the blue toe relevance. Thus the VOR recommendations are confirmed when the authentic value lottery is produced.

Finally, we need only assess one more probability, either the probability of diabetes given glucose or the probability of diabetes given no glucose. Assuming coherence allows us to infer the other probability. The result is a 20% savings in assessment effort, despite the fact that we assessed the marginal on "Blue Toe" and then decided not to use it.

Conclusions

The algorithm developed in this paper is computationally intensive. However, the point of view taken is that assessment effort is much more expensive than computation, especially since the sensitivities performed have the potential to simplify the model of the decision significantly. If an expert system, decision support system or knowledge base is being developed, this simplification will generate cost savings over the life of the product.

In particular, sensitivity to *Value of Relevance*

- indicates the simplest ID within the DM's tolerance for the authenticity of the value lottery
- indicates the potential gain/loss in value due to the addition/deletion of any set of arrows in the BID
- provide a manifold of joint PDFs to explore the sensitivity of decisions to a specific probability in a decision problem
- guides the decision analyst, decision support system builder or knowledge engineer in the most cost effective model of uncertainty

The value of relevance has also been extended to decision problems involving continuous decisions and continuous uncertainties. This extension will be the topic of another paper.

Future Research Future research will focus on improving the computational efficiency of the algorithm and validating the technique on "real-world" problems drawn from consulting practice. Another area of research will be the development of a procedure to obtain VOR when a expert system or knowledge base exists and an extension is proposed. The ability to reduce the amount of work needed to extend an extant system may produce the greatest practical savings.

References

Howard, Ronald A. 1990. From Influence to Relevance to Knowledge. In *Influence Diagrams, Belief Nets and Decision Analysis*, 3-23. Edited by Robert M. Oliver and James Q. Smith. John Wiley & Sons, Chichester, England

Korsan, Robert J. 1990. Towards Better Assessment and Sensitivity Procedures. In *Influence Diagrams, Belief Nets and Decision Analysis*, 427-455. Edited by Robert M. Oliver and James Q. Smith. John Wiley & Sons, Chichester, England

Luce, R. Duncan and Raiffa, Howard 1957. *Games and Decisions.* John Wiley & Sons, Inc. New York, NY

Lowell, David 1994. Sensitivity to Relevance in Decision Analysis, Ph.D. Dissertation, Department of Engineering Economic Systems, Stanford University, Stanford, CA

Shenoy, P. P. and Shafer, G. 1990. Axioms for Probability and Belief Function Propagation. In *Uncertainty in Artificial Intelligence - 4* 169-198. Edited by Schacter, Levitt, Lemmer and Kanal. North-Holland, Amsterdam