Graphical Analysis of Value of Information in Decision Models

Songsong Xu **Kim-Leng Poh**
Department of Industrial & Systems Engineering National University of Singapore
Kent Ridge, Singapore 119260
{ engp8920, isepohkl }@nus.edu.sg

Eric J. Horvitz
Microsoft Research
Redmond, WA 98052-6399
horvitz@microsoft.com

Abstract

We review and extend the qualitative relationships about the informational relevance of variables in graphical decision models based on conditional independencies revealed through graphical separations of nodes from nodes representing utility on outcomes. We exploit these qualitative relationships to generate non-numerical graphical procedures for identifying partial orderings over chance variables in a decision model in terms of their informational relevance. We describe an efficient algorithm based on a consideration of local properties of a property we refer to as u-separation. Finally, we present results of computational efficiencies gained via the application of the new policies, based on analyses of sample networks with different degrees of connectivity.

1 Introduction

The expected value of perfect information (EVPI) is the value of making an observation before taking action under uncertainty. EVPI is an important concept in decision-analytic consultation as well as automated decision-support systems that recommend the best evidence to collect, trading off the cost and benefits of observations and tests. The idea of economic evaluation of information in decision making was first introduced by Howard (1966, 1967).

In recent years, there has been great interest in developing schemes for computing the value of information (VOI). Exact methods for computing the value of information have been explored (Ezawa, 1994; Howard and Matheson, 1981; and Shacher, 1990). Unfortunately, the computational complexity of such exact computation of EVPI in a general decision model with any general utility function is known to be intractable. The intractability of EVPI computation has motivated researchers to explore a variety of quantitative approximations, including myopic, iterative one-step look-ahead procedures (Gorry, 1973; Heckerman, Horvitz & Nathwani, 1992; Dittmer and Jensen, 1997).

We have sought to extend methods for exact and approximate computation of VOI by pursuing opportunities for leveraging qualitative analyses of the value of information. We shall focus on such qualitative evaluation in an influence diagram. In many applications, it is reasonable to bypass the exact numerical computation of the VOI and to instead seek to identify an ordering of variables by their value of information. For example, an ordering over the VOI can be employed in conjunction with cost of information in normative decision systems to determine the most cost effective evidence to collect.

In our earlier related work (Poh & Horvitz, 1996), we have derived qualitative relationships about the information relevance of chance variables in graphical decision models based on the consideration of the topology of the models. We identified dominance relations for the EVPI of chance variables in terms of the position and relationships among variables. We also have found that the EVPIs of chance nodes can be ordered based on conditional independence relationships among the chance nodes and the value node. We outlined an algorithm for obtaining such a partial ordering of EVPI of chance nodes of influence diagrams that are expressed in canonical form.

In this paper we review earlier work and report new results on topological relationships among variables in a graphical decision problem with regards to the VOI. After reviewing earlier work, and presenting extensions, we shall describe studies of the performance of an algorithm that harnesses the new results.

To demonstrate the effectiveness of our algorithm, we have conducted a series of runs on a large number of networks generated randomly. The networks tested are characterized by the number of nodes and the density of the connectivity among nodes. The results of these studies demonstrate that the method can deliver a dramatic improvement in the performance of the algorithm over the previous version. In contrast to previous algorithms, the approach produced tractable runtimes even for large networks of up to 60.

This paper is organized as follows: In Section 2, we review the basic ordering relations for EVPI on chance nodes and show its application to an example. In Section 3, we discuss several extensions to the qualitative relations and present an algorithm which incorporates these extensions. In Section 4, we present the results obtained on a series of networks of varying sizes and density. Finally, in Section 5, we conclude and provide potential directions for extending the method.

2 Value of Information and Conditional Independence

We shall first examine several qualitative relationships about the information relevance of variables in graphical decision models. We will review some results obtained previously and present some extensions. We shall focus on models in *canonical form*. In general, any influence diagrams can be converted to canonical form, a formulation where all chance nodes that are descendants of one or more decision nodes are deterministic nodes (Heckerman & Shachter, 1995).

2.1 Basic information relevance ordering relations

Let $M = (C, D, V, E)$ be a decision model where C is the set of chance nodes, D the set of decision nodes, V the value node, and $E \subset (C \cup D) \times (C \cup D \cup \{V\})$ is the set of directed arcs. We denote the expected value of information for observing the value of chance node $X \in C$ before action $A \in D$ by $\text{EVPI}_M(A \mid X)$.

We have shown previously (Poh & Horvitz, 1996) that chance nodes that are not relevant to the value node given the action have zero value of information.

We have also established the basic relations concerning the possible ordering of EVPI for two chance nodes in a graphical decision model based on conditional independence of the value node of one chance node given the other:

Those results can be generalized to the joint value of perfect information of a set of nodes by replacing X and Y with sets. The conditional independence relations required for identification of the ordering of EVPI can be performed with the notion of d-separation (Pearl 1988, Pearl et al 1990).

An equivalent graphical procedure for identification of conditional independence relations makes use of the notion of u-separation (The notion of u-separation can be found in Castillo et al, 1997). Given a direct acyclic graph and three disjoint sets of nodes X, Y, and Z, we first moralize the smallest subgraph containing X, Y and Z and their ancestral nodes. If Z *u*-separates X and Y in the moralized graph, then Z *d*-separates X and Y in the original directed graph; otherwise Z does not *d*-separate X and Y.

2.2 Example

Figure 1 shows the graphical model of a sample decision problem. The topology of the network is adopted from the car diagnosis example (Norsys 1998). By applying the *d*-separation criterion for the ordering of EVPI values, we obtain the ordering network as shown in Figure 2. Here an arc between two nodes, for example, $A \rightarrow B$, indicates $\text{EVPI}(D \mid A) \leq \text{EVPI}(D \mid B)$.

We shall leverage a concept referred to as a barren node. Shachter (1986) introduced the notion of barren nodes in influence diagram evaluation. Barren nodes are those other than the value node which have directed arcs into them but not out of them. We note that node 10 is a barren node. Hence, its EVPI is bounded by the EVPI on its parent node 4. This is not a very densely linked graphical decision model, and we obtained several EVPI orderings that indicate the relative ranking of the importance of information.

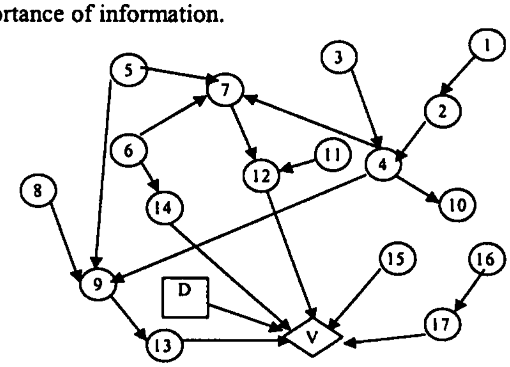

Figure 1: Influence diagram for Example 1.

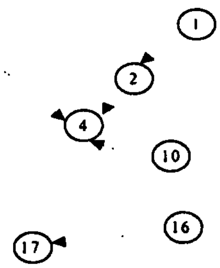

Figure 2: The partial ordering of EVPI for Example 1

2.3 Computational approaches

In practice, we may generate a partial ordering of EVPI by engaging in a pairwise comparison of nodes and checking for *d*-separation of one node from the value node by the other. We call this method the pairwise-comparison approach. This algorithm does not exploit the topological structure of the network to gain efficiency. We shall now introduce a new approach to the identification of partial ordering of EVPI in graphical decision model by identifying barren nodes and extending the *u*-separation relation to more encompassing neighborhoods. We refer to the new algorithm as *u*-separation extension.

3 Efficient Identification of EVPI Orderings

In this section, we first describe a number of extensions on the graphical properties of information relevance for chance nodes. Then, we shall describe an algorithm that exploits these new results.

3.1 Treatment of Barren nodes

Omission of barren nodes from a graphical decision model has no effect on the optimal decision policy.

Furthermore, their VOI is always bounded above by the joint VOI of their direct predecessors.

Theorem 1. In a canonical decision model M, let B be a barren node and $\pi(B)$ be the set of direct predecessors of B, and A be a decision node. Then $\text{EVPI}_M(A \mid B) \leq \text{EVPI}_M(A \mid \pi(B))$.

Proof: The result follows from the fact that since a barren node is a sink node with no arc coming out of it, it follows that it is always d-separated by all its parent nodes from the value node (see Figure 3). We can also infer the result from the so-called *Markov property* of a DAG, since the value is always a non-descendant of any barren nodes and the required conditional independent relation must holds. □

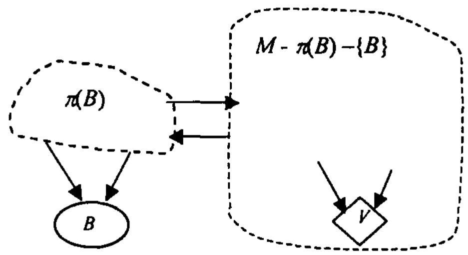

Figure 3: EVPI of barren nodes are always bounded by those of their parents

Hence in trying to obtain an EVPI ordering of the chance nodes in a decision model we may first remove all the barren nodes because their EVPI is always less than those of their respective parents. Furthermore, removing the barren nodes has no influence on the ordering of other nodes since barren nodes are not in the ancestral sets of any other nodes. After the EVPI ordering of all non-barren nodes has been achieved, we may insert the barren nodes into the ordering to complete the analysis.

3.2 Neighborhood Closure property of *u*-separation with the value node

The Neighborhood Closure of u-separation with the value node allows us to infer u-separation relations in a neighborhood thereby eliminating the need to explicitly check for u-separation once u-separation of a single node is established in a neighborhood of a cluster of nodes.

Theorem 2. Let G be the moralized graph of a graphical decision model with the decision node removed. Let node X be chance node, node Y be a neighbor of Z in graph G. Then Y is *u*-separated from the value node V by X if and only if Z is *u*-separated from the value node V by X.

Proof: Referring to Figure 4, suppose Y is *u*-separated from the value node V by X. Then every path from Y to V passes through X, and any path from Z to V must is either pass through both Y and X or only X alone. No path can run from Z to V without going through X for this will violate the *u*-separation of Y from V by X. Hence Z is separated from V by X. The converse is also true by symmetry. That is, if Z is *u*-separated from V by X, then Y is *u*-separated by V by X. □

The above result allows us to check the *u*-separation of any node with V and if it is found to be true, to recursively add the property to all of their direct neighbors. For example, in the network shown in Figure 5, if it is established that Y is u-separated by X from V, then we can infer that all the shaded nodes will also be *u*-separated by X from V. We state this in the following theorem:

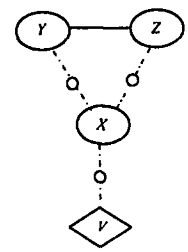

Figure 4: Extension of u-separation from value node to a direct neighbor.

Theorem 3. Let G be the moralized graph of a graphical decision model with the decision node removed. If in G, a chance Y is *u*-separated by another chance node X from the value node, then the maximal connected sub-graph containing Y is also u-separated from V by X.

Proof: The result follows from the recursive application of Theorem 2. □

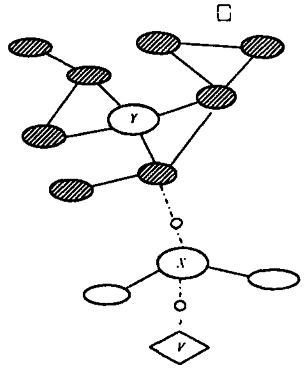

Figure 5. U-separation of Y from V by X can be extended to the maximal connected sub-graph containing Y

3.3 An Algorithm for identifying EVPI orderings

*Input: An influence diagram **M**.*

*Output: An EVPI ordering set **Ω** of the influence diagram.*

1. Convert the network M into canonical form if it is not already so.
2. Drop all the decision nodes in M.
3. Identify the ancestral sub-network of the value node V.
4. Moralize the ancestral sub-network.
5. Let $\Omega = \varnothing$.
6. Let $N \leftarrow C$, the set of chance nodes in M.
7. While $N \neq \varnothing$ do
8. Mark all nodes in N as "unvisited"
9. Pick a node $X \in N$
10. Let $N \leftarrow N \setminus \{X\}$

11. For each node $Y \in Adj(X)$ do
12. If Y is "unvisited" and $Y \neq X$ then
13. Mark node Y as "visited".
14. If Y is u-separated by V given X then
15. Add the ordering $\{X \leq Y\}$ to Ω
16. Recursively add all $\{Z \leq Y\}$ to Ω where $Z \in Adj(Y)$ and Z is "unvisited"
17. Else
18. Mark all nodes $Z \in Adj(Y)$ and $Z \neq X$ as "visited".
19. End if
20. End if
21. End For
22. End While
23. For each barren node B, add $\{B \leq P\}$ to Ω where P is a parent of B in M.
24. Output Ω

Notice that the barren nodes in this algorithm will be excluded from the ancestral sub-network being processed. The algorithm goes through every chance node and considers it as a separator node. If a neighboring node is found to be u-separated by the current node from the value node, the EVPI ordering is added to the list, and Theorem 2 is applied recursively in a depth-first manner to include the ordering of adjacent nodes compared with the current node. Figure 6 shows the adjacent node u-separation probing scheme.

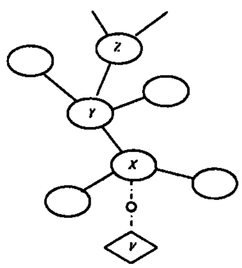

Figure 6. Propagation of EVPI from Y to its neighbourhood.

We shall now provide an estimate of the runtime complexity of *u-separation extension* and compare it to the pairwise-comparison algorithm. For an n-node network (n includes the value node), the pairwise-comparison algorithm requires $(n-1)(n-2)$ checks for undirected conditional independence. The worst-case complexity for depth-first search is $O(n^2)$; by clever implementation the computational time for checking conditional independence thus will be $O(n^2)$. Hence the overall computational time of the pairwise-comparison algorithm is $O(n^4)$.

The new algorithm performs only $(n-1)$ number of u-separation checks. Hence the computational time is $O(n^3)$. We therefore can typically expect a speed up of about n times compared with the pairwise-comparison procedure. Since real world networks are often very large, this speed up proportion to the network scale may be significant.

4 Computational Evaluation of the Algorithm

We implemented the *u-separation extension* algorithm and applied it to a variety of problems.

4.1 Applications of the Algorithm to Sample Problem

Let us first explore the enhanced performance of *u-separation extension* on Example 1. The run time speed up ratio of the *u-separation extension* algorithm over the pairwise-comparison approach is 1.67 for this 18-node simple example. We observed a decrease in run times for the new algorithm over the naïve scheme for this example.

4.2 More Extensive Studies

In order to perform a comprehensive computational evaluation of the algorithm and to study the effect of specific topologies on its performance, we generated a series of networks with different sizes and connectivity. Two parameters that describe the policy for generating networks are (1) the number of nodes in the network which varies between 20 and 60 nodes, and (2) a branching index which is the probability that any two nodes in the network are connected by a directed arc. A branching probability close to zero will generate a very sparse network while a branching probability close to unity will generate a densely connected network. For example, a branching probability of 0.2 on a 20-node network will have, on the average, 19*0.2 or 3.8 number of arcs connected to any node.

One of the experiments is applying both algorithms to a total of 13 networks of 20 nodes each. The branching probability was varied between 0.15 and 0.20 producing an average connectivity of 2.85 to 3.8 representing the typical numbers found in practical networks. In all cases, a significant speed up was obtained by the new algorithm when compared with the naïve scheme; the average speed up ratio for 20-node network is 4.55 times. We also observe that the runtime generally increases with increases in the connectivity density.

We extended the experiment to larger networks of 30 to 60 nodes. Figure 7 shows the plot of the ratio of the runtime for the two algorithms. The graphed data is the average ratio for different numbered networks.

From above we observed that the algorithm generally provides a significant improvement over the naïve approach. However, the performance for individual network depends on the number of nodes, its structure, and the connection density. We also note that the average speed up is roughly proportional to the number of nodes in the network when the latter is large. This is consistent with our earlier analysis on the runtime complexity of algorithm compared to the naïve method.

5 Summary and Conclusion

We have described an algorithm for the identification of partial ordering of EVPI for chance nodes in graphical decision models. The algorithm is based on non-numerical graphical analysis based on the idea of u-separation.

We have tested the algorithm on a number of networks of sizes varying from 20 to 60 nodes and in all cases, satisfactory runtime were obtained. We achieved a significant speed up over a naïve approach proposed previously. Knowledge of EVPI orderings of the chance nodes in a graphical decision network can help decision analysts and automated decision systems weight the importance or information relevance of each node and direct information-gathering efforts to variables with the highest expected payoffs. We believe that the algorithm described in this paper can serve the purpose well.

A limitation of our approach is that it only generates a partial ordering. This is the price for considering only qualitative properties. However, the trade-off in completeness is well spent since the exact numerical computation of EVPI for all nodes is known to be intractable.

We have focused on studies with medium-sized examples. For larger networks, it may be promising to employ methods that decompose the network into several subnets to be individually processed. The partial orderings obtained may then be merged. We also observed that clusters which are densely connected tend to produce very sparse partial ordering graph, i.e., nodes that are densely connected tend to resist yielding an ordering with our method. While this may limit the usefulness of our approach, we can exploit this property by clustering such densely connected nodes as one group and treating it as a single node. We can then use our algorithm to find partial orderings of group of nodes. Another possible extension of our approach is to consider some heuristic classification of decision models based on their network topology and then to apply different types of search strategies based on such a classification.

References

Castillo, E.; Gutierrez, J.M.; and Hadi, A.S. 1997. *Expert Systems and Probabilistic Network Models*. Springer.

Dittmer, S. L.; and Jensen, F. V. 1997. Myopic Value of Information in Influence Diagrams. *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, 142-149, Morgan Kaufmann.

Ezawa, K. J. 1994. Value of Evidence on Influence Diagrams. *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, 212-220, Morgan Kaufmann.

Gorry, G. A., 1973. Computer-assisted clinical decision making. *Computer-assisted clinical decision making*, 12:45-51.

Heckerman, D. E.; Horvitz, E. J.; and Nathwani, B. N. 1992. Toward Normative Expert Systems: Part I The Pathfinder Project. *Methods of Information in Medicine*, 31:90-105.

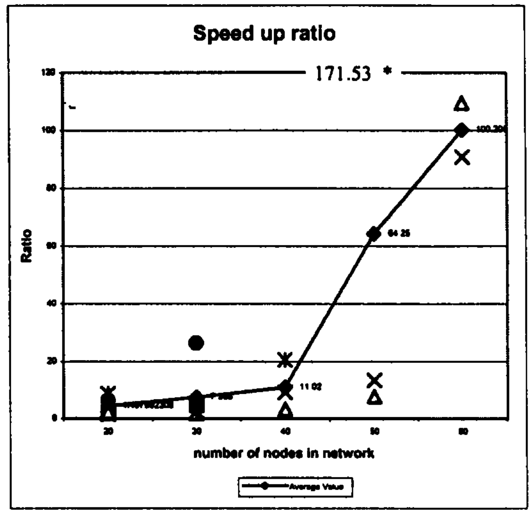

Figure 7. Speed up ratio of the algorithm over range of network size

Heckerman, D.; and Shachter, R. 1995. Decision-Theoretic Foundations for Causal Reasoning. *Journal of Artificial Intelligence Research*, 3: 405-430.

Howard, R. A. 1966. Information value theory. *IEEE Transactions on Systems Science and Cybernetics*, SSC-2(1): 22-26.

Howard, R. A. 1967. Value of information lotteries. *IEEE Transactions on Systems Science and Cybernetics*, SSC-3(1): 54-60.

Howard, R. A.; and Metheson, J. E. 1984. Influence Diagrams. In R.A. Howard and J.E. Metheson (editors), *Readings on the principles and applications of decision analysis*, Vol. 2, 719-762, SDG.

Jensen, F.; Jensen, F. V.; and Dittmer, S. L. 1994. From Influence Diagrams to Junction Trees. *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, 367-373. Morgan Kaufmann.

Pearl, J. 1988. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.

Pearl, J.; Geiger, D.; and Verma, T. 1990. The Logic of Influence Diagram. In *Influence Diagrams, Belief Nets and Decision Analysis*, 67-87. John Wiley.

Poh, K. L.; and Horvitz, E. 1996. A Graph-Theoretic Analysis of Information Value. *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence*, 427-435, .Morgan Kaufmann.

Shachter, R. D. 1990. Evidence Absorption and Propagation through Evidence Reversals. In *Uncertainty in Artificial Intelligence 5*, 173-190, North-Holland.

Shachter, R. D., 1986. Evaluating Influence Diagrams. *Operations Research*, 34: 871-882.

Verma, T.; and Pearl, J. 1990. Causal Networks: Semantics and Expressiveness. *Uncertainty in Artificial Intelligence 4*, 69-76, North Holland

Norsys Software Corp., Netica, 1998. *Application for Belief Networks and Influence Diagrams*, (http://www.norsys.com)