Raven: Bayesian Networks for Human-Computer Intelligent Interaction

Ole J. Mengshoel and David C. Wilkins

Abstract. *Bayesian Networks are a promising computational formalism for drawing conclusions from large amounts of intelligence analysis data. They are suited to this human-computer intelligent interaction task because they're easily mapped onto a comprehensible graphical network representation; and because they are superb in domains with large amounts of uncertain data. This paper shows how the intelligence analysis task can be mapped into Bayesian networks. And it overviews two research contribution: two new heuristic algorithms for efficient inference and approximation of Bayesian networks; and algorithms to generate representative test sets for evaluation of faster methods of inference.*

1. Introduction

Bayesian networks (BNs) are an essential knowledge representation technique in artificial intelligence [Pearl 1988]. Substantial progress has been made in recent years in all areas of BN research. However, there are still important problems to solve, and the contributions presented in this chapter fall in three categories. First, we describe how the challenging task of analyzing sensor intelligence reports can be solved using Bayesian network representation and inference methods. Second, to address the slowness of Bayesian networks with respect to real-time intelligence analysis, we present approximation methods that deliver significant inference speedup. These methods are the stochastic search techniques of probabilistic crowding and stochastic greedy search as well as abstraction methods, which can be used when computing the most probable explanation in BNs. Third, to facilitate the development of more efficient inference techniques on large Bayesian networks, we have developed techniques to construct large synthetic BNs for empirical experimentation. In particular, two constructions are presented: deceptive BNs and satisfiability-like BNs.

Intelligence analysis is a complex reasoning task where uncertain and incomplete data and knowledge is pervasive, in particular with respect to knowledge about the other players intentions and actions. A BN can be used to address these issues, and this chapter describes how this can be accomplished. The rest of this chapter is organized as follows. In Section 2 we present BNs. Section 3 discusses how BNs can be used for intelligence analysis. Section 4 summarizes the novel research results on heuristic algorithms, Section 5 focuses on construction of synthetic BNs for systematic experimentation, and Section 6 concludes.

2. Bayesian Networks

Bayesian networks (BNs) are used for reasoning and learning under uncertainty. Probability theory and graph theory form their basis: random variables are represented as nodes, conditional dependencies are represented as edges, and structured as a directed acyclic graph. In a discrete BN with n nodes $\{X_1,...,X_n\}$ and instantiations $\{X_1 = x_1,..., X_n = x_n\}$, joint probability is

$$\Pr(x_1,..., x_n) = \Pr(\boldsymbol{x}) = \prod_{i=1}^{n} \Pr(x_i \mid \boldsymbol{p}_{x_i}) , \tag{1}$$

where $\boldsymbol{p}_{x_i}$ is the instantiation of the parents of node X_i, which has instantiation x_i. Inference in BNs, in the form of belief updating revision and belief revision, is the focus of our research. *Belief updating* amounts to the following: Given query node Q and evidence $\boldsymbol{e}$ at nodes $\boldsymbol{E}$, infer posterior probability $Pr(Q \mid e)$. Note that any nodes in the network can be evidence or query nodes. *Belief revision* amounts to computing a most probable explanation x_{MPE} in a BN, namely

$$\boldsymbol{x}_{\text{MPE}} = \arg\max_{\boldsymbol{x} \text{ an explanation}} \Pr(\boldsymbol{x} \mid \boldsymbol{e}) , \tag{2}$$

where an explanation $\boldsymbol{x}$ instantiates all nodes except the evidence nodes. A variety of approaches to BN inference have been investigated by Pearl, Lauritzen, Lin, Rojas-Guzman, and Mengshoel, among others. These inference algorithms vary in many respects – they are exact, approximate, or heuristic; work on singly or multiply connected graphs; and are used on different inference tasks including belief propagation and belief revision. For both of these computational tasks, *NP*-hardness has been proven [Cooper, 1990], and unless *P* = *NP*, solving these tasks approximately or heuristically is therefore important, and is discussed below.

3. Intelligence Analysis Using Bayesian Networks

In this section we discuss the intelligence analysis task, and in particular what the problem involves in terms of human decision making, the input given to a human intelligence analyst and the output that is required.

3.1 The Intelligence Analysis Problem

The role of intelligence analysts is to provide assessments of the situation, in particular the activity of the opposing side [Mengshoel, 1997a] [Mengshoel, 1998a]. Specifically, the purpose of intelligence analysis is to answer priority intelligence requests (PIRs) such as "Position of opposing force". There is a doctrinal one to one mapping between PIRs and decisions that are made during an engagement, so PIRs play a central role in the reasoning process. Based on a terrain map, and tactical and strategic knowledge, the analyst identifies named areas of interest (NAIs). NAIs are areas where one is more likely to find opposing forces given the terrain. For instance, an NAI might be a choke point. After identifying NAIs, the analyst allocates intelligence assets to investigate a number of these NAIs, resulting in intelligence reports. Figure 1 shows how these intelligence reports are input to the Bayesian reasoning system. Table 2 provides examples of the intelligence reports, called SALUTE reports. SALUTE is an abbreviation for the header fields of a SALUTE report, namely, Sensor, Activity, Location, Unit, Time, and Equipment.

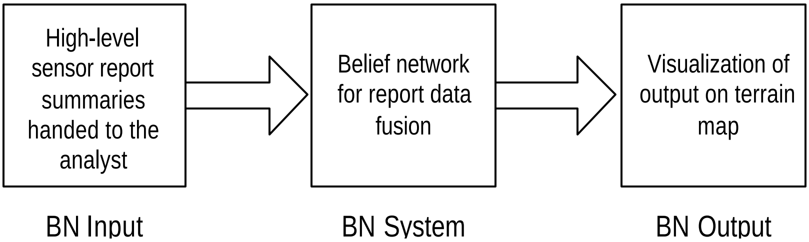

Figure 1. Bayesian belief networks for intelligence analysis. The intelligence analyst and the Bayesian Network are given the same input, namely SALUTE reports. Thus the analyst conclusions can be compared and contrasted with those of the Bayesian Network

What makes this task of interpreting SALUTE intelligence reports difficult for intelligence analysts? The first major challenge is the amount of uncertainty in each of the SALUTE data fields. For example, Sensor types have different levels of reliability and credibility. The second major challenges in interpreting SALUTE reports is their large volume. At present, 100s to 1000s of intelligence reports per hour is common, and as the degree of automation in data collection increases, this information rate is going to increase, too. Experience has shown that it is easy to for an analyst to over-look a small, but crucial, piece of information that might be inconsistent with other observations received so far. Clearly, BNs can help with both of these challenges.

Table 1 Examples of three intelligence reports in SALUTE format. The intelligence analyst and the BN infer conclusions based on this information.

Time	Sensor	Location	Size	Equipment	Unit	Activity
200917	SIGINT	NK285105	?	artillery	?	traffic
200943	JSTARS	NK4300 to NK3000	200+	MTI	?	moving
201203	REMBASS strings	R2-R5	200+	vehicles	?	moving

3.2 Mapping Intelligence Analysis Tasks into Bayesian Networks

Due to the complexity of manual intelligence analysis, some type of automated decision aid is desirable. When integrated into a software system as shown in Figure 1, an intelligence analysis BN as shown in Figure 2 can be used to perform the intelligence analysis task automatically. The BN-based approach is to perform data fusion and filtering using BNs and output and visualize the results. As new information is received, the BN probabilities change, and we give examples showing how this takes place.

As an example of mapping intelligence task onto BNs, we focus on enemy defensive scenarios, where the enemy is defending and friendly forces are attacking. Here, a BN is used to model the enemy and to filter uncertain information. An example BN model for battlefield reasoning is shown in Figure 3. The intelligence analysis task is analogous to medical diagnosis, where the PIRs are diagnoses and intelligence reports are symptoms. Consequently, these networks consist of several layers. The diagnosis layer corresponds to priority intelligence requests (PIRs). The symptom layer corresponds to SALUTE reports. Such reports typically affect the probabilities in other BN nodes, and in particular PIR nodes, as illustrated in Figure 2.

3.3 Advantages of Using Bayesian Networks

There are several advantages of using BNs in the Intelligence Analysis domain. First, BNs are ideally suited for problems that involve massive amounts of uncertainty, and as reflected in the phrase “fog of war”, this is certainly the situation in intelligence analysis. Second, exploiting belief updating and belief revision algorithms, changes to belief are modified in real-time as each new piece of information is received. Third, BNs provide decision makers with a graphical representation of the casual reasoning structure; this makes knowledge

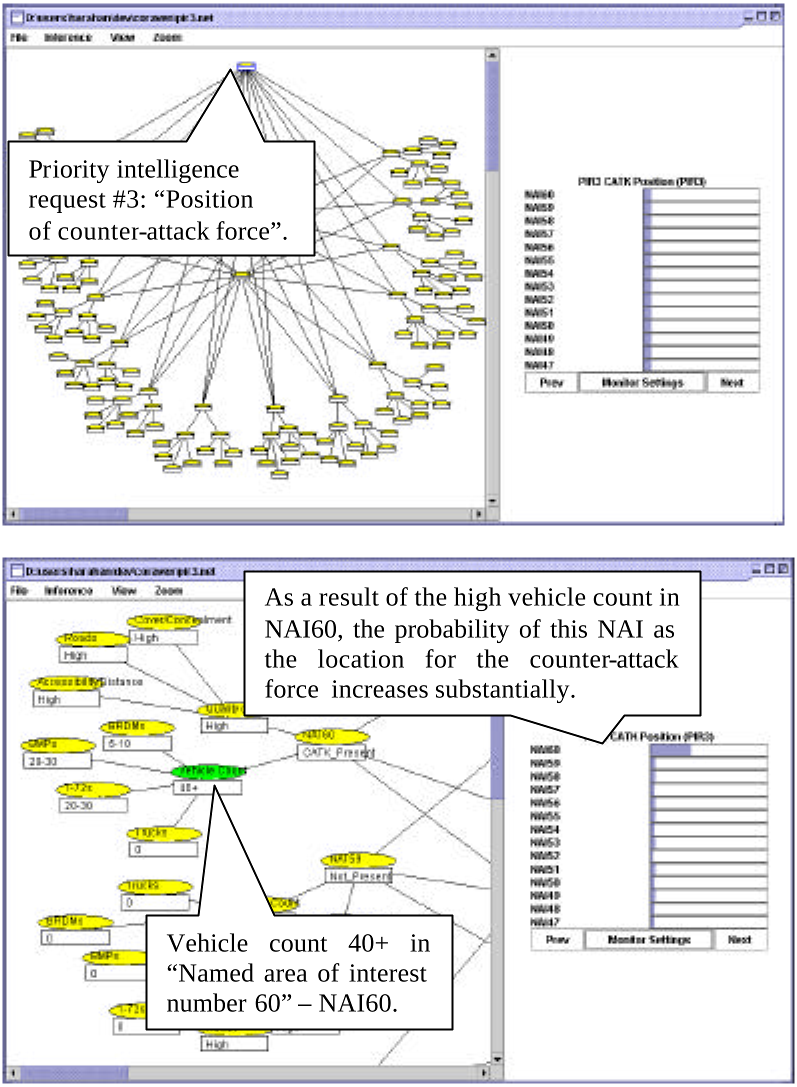

Figure 2. Bayesian network probabilities – before (top) and after (bottom) observations.

acquisition and understanding easier. Note that a BN can accept input at the same level as is given to the human, thus the problem is different than low-level sensor fusion. Fourth, BNs allow playing what-if games: the analyst can change the belief associated with various input reports and see how conclusions – in terms of distributions over PIRs – change.

4. Bayesian Network Computation

In this section, new heuristic algorithms for efficient inference are presented [Mengshoel, 1999a] [Mengshoel, 1999b]. These algorithms are implemented in the RAVEN software. This software includes a user interface front end, shown in Figure 2 and implemented in JAVA and Swing, and an algorithmic back end implemented in C++.

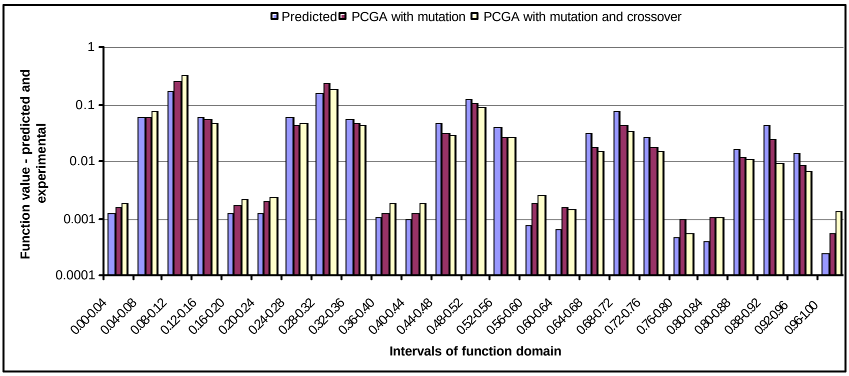

Figure 3. Predicted versus actual allocation of individuals to niches for a test function for the Probabilistic Crowding GA.

4.1 Probabilistic Crowding Genetic Algorithm

We have investigated the use of genetic algorithms (GAs) for approximate BN inference [Mengshoel, 1998c]. GAs are robust function optimizers that employ stochastic, instance-based (or population-based) search [Goldberg, 1989]. Since a BN represents a function, it is natural to consider using GAs to search in the space of BN explanations.

As part of this research, investigations have been made into niching genetic algorithms, which converge to multiple local optima. Local optimum is an important problem in BN inference as well as in other hard inference and optimization problems. We have introduced the Probabilistic Crowding niching genetic algorithm [Mengshoel, 1999a] [Mengshoel, 1999b], and presented theoretical and empirical results showing that Probabilistic Crowding gives predictable convergence, which at equilibrium is proportional to the utility function. Figure 3 presents empirical results for the Probabilistic Crowding GA (PCGA). The figure shows the predicted allocation of individuals – essentially the value of the utility function – as well as actual allocation of individuals for two PCGA variants: PCGA with mutation only, and PCGA with mutation and crossover. For both variants, there is a very good correspondence between empirical results and the predicted allocation.

4.2 Stochastic Greedy Search

Improvements have been made to stochastic local search algorithms for BN inference by introducing the Stochastic Greedy Search approach [Mengshoel, 1999a]. The Stochastic Greedy Search algorithm, which uses noisy and hill-climbing search steps, different measures of gain, and an operator-based approach, provides different ways to perform local search. Stochastic Greedy Search also introduced novel initialization algorithms. Comparisons to the state-of-the-art inference Hugin system, which implements the clustering algorithm [Lauritzen, 1988], show that Stochastic Greedy Search performs significantly better for satisfiability BNs as well as for certain application networks. In application networks, initialization algorithms prove to be very valuable. The reason for this is that they, by exploiting structure as well as conditional probability tables in these BNs, typically start hill-climbing steps from a better starting point than simple random initialization.

4.3 Abstraction in Bayesian Networks

We have also made improvements to the use and measurement of abstraction and aggregation to improve BN inference [Mengshoel, 1998b] [Mengshoel, 1999a]. A criterion is introduced that quantifies how different methods of abstraction and aggregation impact quality of inference. The criterion regards abstraction as noise and uses variance as a measure of quality. Results are presented that quantify how different methods and levels of abstraction effect accuracy.

5. Synthetic Bayesian Networks

Two major research results are presented that relate to creating hard synthetic BNs for empirical research on inference algorithms.

5.1 Hard Synthetic Bayesian Networks

The first method of creating hard synthetic Bayesian networks translates satisfiability problems into BNs [Mengshoel, 1999a]. Connectivity, value of conditional probability tables, and degree of regularity of the underlying graph are shown to affect the speed of inference for Hugin and Stochastic Greedy

Search. Figure 4 presents how regularity affects the speed of inference in the Hugin system. For both irregular and regular BNs, inference time increases with increasing *C/V*-ratio, as expected based on previous results for satisfiability (SAT). Interestingly, the regular BNs are on the average harder than the irregular BNs, and to our knowledge this is the first time this effect has been shown for a structure-based system such as Hugin.

5.2 Deceptive Bayesian Networks

The concept of deception was introduced in order to systematically investigate conditions under which GA schema processing might lead a GA away from utility function optima [Goldberg, 1989]. The second method of creating hard synthetic Bayesian networks is based on translating deceptive problems studied in genetic algorithms to a BN setting, showing that BNs can be deceptive [Mengshoel, 1999a] [Mengshoel, 1999b].

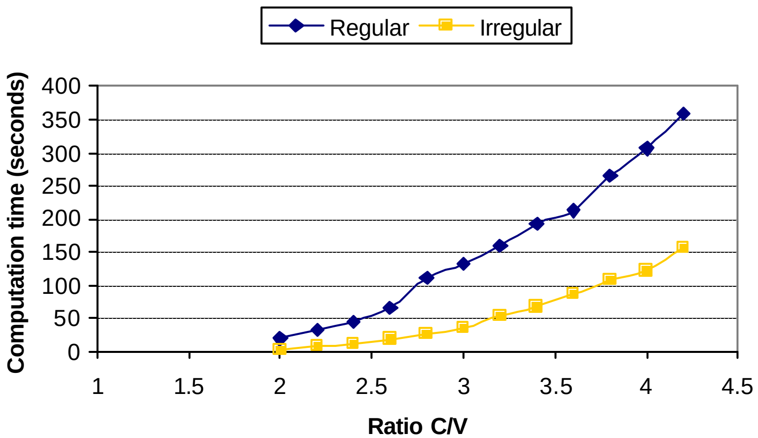

Figure 4. Computation time in Hugin, for regular and irregular BNs, as function of the ratio *C/V*, where *C* is the number of non-root nodes, *V* the number of root nodes. The mean for 100 BN instances is presented.

6. Summary and Conclusion

This chapter has illustrated three points, details of which are greatly expanded elsewhere, especially in [Mengshoel, 1999a]. First, we have presented an approach to filtering and integrating intelligence data reports using Bayesian networks. The approach addresses several of the challenges associated with intelligence analysis, including dealing with a large volume of reports as well as dealing with the uncertain and incomplete knowledge contained in the reports. Second, we have presented new BN inference approximation methods, in particular Probabilistic Crowding and Stochastic Greedy Search. Third, we have presented results related to construction of synthetic Bayesian networks for empirical investigations.

The Raven project has served as a basis of a number of other research efforts. The Co-Raven project [Jones, 1999; Wilkins, 1999] added sonification, collaboration, and a terrain map visualization interface to the Bayesian interface and reasoning core. Psychological studies have been performed relating to the accuracy of estimating Bayesian network probabilities, and the effect of specific estimates on global decision-making accuracy [Chernyshenko, 1999].

Acknowledgements

This research was performed while Ole J. Mengshoel was a Ph.D. student at the University of Illinois. We gratefully acknowledge the insights and assistance provided by Peter Asaro, David Fried, David Goldberg, Carline Hayes, Patricia Jones, Song Han, Tim Harahan, and Dan Roth. This research was funded in part by ARL FedLab grant DAAL01-96-2-0003 and ONR grant N00014-95-1-0749.

References

[Jones, 1998] P. M. Jones, C. C. Hayes, D. C. Wilkins, R. Bargar, J. Sniezek, P. Asaro, O. J. Mengshoel, D. Kessler, M. Lucenti. (1998). CoRaven: Modeling and Design of a Multimedia Intelligent Infrastructure for Collaborative Intelligence Analysis. *Proc. 1998 IEEE Intrnl Conf. on Systems, Man, and Cybernetics*, San Diego, Oct 1998.

[Mengshoel, 1997a] O. J. Mengshoel, and D. C. Wilkins, “Visualizing Uncertainty in Battlefield Reasoning Using Bayesian Networks.” In *Proc. ARL Advanced Displays and Interactive Displays Fed. Lab. First Annual Symposium*, Adelphi MD, January 1997.

[Mengshoel, 1998a] O. J. Mengshoel, D. C. Wilkins, and S. Uckun, “Filtering and Visualizing Uncertain Battlefield Data Using Bayesian Networks,” In *2nd Annual Fed. Lab. Symposium, Advanced Displays and Interactive Displays*, College Park MD, February 1998.

[Mengshoel, 1998b] O. J. Mengshoel and D. C. Wilkins, “Abstraction for Belief Revision: Using a Genetic Algorithm to Compute the Most Probable Explanation,” In *Proc. AAAI 1998 Spring Symposium on Satisficing Models*, Stanford University CA, pp. 46-53, March 1998.

[Mengshoel, 1998c] O. J. Mengshoel, and D. C. Wilkins, “Genetic Algorithms for Belief Network Inference: The Role of Scaling and Niching,” In *Proc. of the Seventh Annual Conference on Evolutionary Programming,* San Diego CA, pp. 547-556, March 1998.

[Mengshoel, 1998d] O. J. Mengshoel, D. E. Goldberg, and D. C. Wilkins, “Deceptive and Other Functions of Unitation as Bayesian Networks,” In *Genetic Programming 1998: Proc. of the Third Annual Conference*, University of Wisconsin, Madison, pp. 559-566, July 22-25, 1998.

[Mengshoel, 1999a] O. J. Mengshoel, “Efficient Bayesian Network Inference: Genetic Algorithms, Stochastic Local Search, and Abstraction,” *Ph.D. Dissertation,* Dept. of Computer Science, University of Illinois, Urbana-Champaign IL, May 1999.

[Mengshoel, 1999b] O. J. Mengshoel and D. E. Goldberg, “Probabilistic Crowding: Deterministic Crowding with Probabilistic Replacement.” In *Proc. of Genetic and Evolutionary Computation Conference (GECCO-99)*, Orlando FL, July 13-17, 1999.

[Chernyshenko, 1999] Chernyshenko, O. S., & Sniezek, J. A. Priming for expertise and confidence in choice: Evaluating the global improves calibration for the specific. Annual meeting of the Judgment and Decision Making Society, Dallas, November, 1998.

[Pearl, 1988] J. Pearl, “Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference,” Morgan Kaufmann, San Mateo CA, 1988.

[Wilkins, 1999] D. C. Wilkins, O. J. Mengshoel, O. Chernyshenko, P. M. Jones, C. C. Hayes, R. Bargar, "*Collaborative Decision Making and Intelligent Reasoning in Judge Advisor Systems*," In Proc. Hawaii International Conference on System Sciences (HICSS-32), January 1999.