To appear: AI Magazine Special Issue on Uncertainty in AI, Summer 1999.

Introduction to This Special Issue:

An Overview of Some Recent Developments in Bayesian Problem Solving Techniques

Peter Haddawy
Decision Systems and Artificial Intelligence Lab
Dept. of EE&CS
University of Wisconsin-Milwaukee
and
Intelligent Systems Lab
Faculty of Science & Technology
Assumption University, Thailand

Abstract

The last five years have seen a surge in interest in the use of techniques from Bayesian decision theory to address problems in AI. Decision theory provides a normative framework for representing and reasoning about decision problems under uncertainty. Within the context of this framework, researchers in uncertainty in the AI community have been developing computational techniques for building rational agents and representations suited to engineering their knowledge bases. This special issue reviews recent research in Bayesian problem-solving techniques. The articles cover the topics of inference in Bayesian networks, decision-theoretic planning, and qualitative decision theory. Here, I provide a brief introduction to Bayesian networks and then cover applications of Bayesian problem-solving techniques, knowledge-based model construction and structured representations, and the learning of graphical probability models.

The past five years or so have seen increased interest and tremendous progress in the development of Bayesian techniques for building problem solving systems. We have come a long way since the Uncertainty in AI Workshop was founded in 1985, an event precipitated in large part by the fact that the mainstream AI community at that time considered probabilistic approaches impractical for building intelligent systems. Since then the workshop has become the Conference on Uncertainty in AI, attracting high-quality contributions from researchers in a broad array of disciplines, including AI, statistics, operations research, and decision science. In the last several years, concepts from Bayesian Decision Theory, along with representational and computational techniques developed within the Uncertainty in AI community have found their way into main-stream AI and are appearing more and more routinely in papers not focused primarily on uncertainty. Areas include vision, natural language processing, robot navigation, planning, and machine learning. One sign of the importance now regarded Bayesian techniques in building and analyzing software systems is the fact that the *Journal of the ACM* recently introduced a track on Decisions, Uncertainty, and Computation.

Bayesian Decision Theory, like much of AI, is concerned with the characterization of rational behavior. According to Bayesian Decision Theory [Savage, 1954], a choice situation is characterized by a set of possible acts A, a probability distribution P over the set of possible states of the world *S*, the outcome of each act in each possible state *a(s)*, and a utility function *u* over the outcome space. The optimal act is the one that maximizes expected utility:

$$EU(a) = \sum_{s \in S} p(s) \cdot u(a(s))$$

Acts here can be physical actions, speech acts, deliberative acts, or complex plans composed of various kinds of actions. Decision Theory is interesting for AI because it provides a normative theory for designing agents capable of reasoning and acting under conditions of uncertainty. This normativity is expressed in the form of a representation theorem stating that if an agent's preferences obey a set of intuitively appealing constraints, then there exists a probability function *P* and a utility function *U*, such that the most preferred action is the one that maximizes expected utility. But what Decision Theory does not provide are the computational mechanisms for building rational agents and the representations suited to engineering their knowledge bases. These issues have been the primary focus and contribution of the work conducted by uncertainty researchers in AI. In this introduction I will discuss issues in the elicitation, representation and manipulation of probability models. I provide a brief discussion of Bayesian networks and then cover applications of Bayesian techniques, knowledge-based model construction and structured representations, and learning of graphical probability models. More recently, researchers have begun to explore these same issues with regard to utility models. The article by Jon Doyle and Richmond Thomason and the one by Jim Blythe that follow provide some discussion of this work. In the interests of brevity, in this introduction I have omitted discussion of exciting and valuable research contributions in many other subareas. I point the interested reader to the proceedings of the Conference on Uncertainty in AI [UAI-98], the Workshop on AI and Statistics [AIStat-99], and to the web page of the Association for Uncertainty in AI (www.auai.org) for further reading.

Bayesian Networks

The Bayesian network formalism is the single development most responsible for progress in building practical systems capable of handling uncertain information. The first book on Bayesian networks [Pearl 1988] was published just over ten years ago and since then several other text books have appeared [Neapolitan 1990, Jensen 1996, Castillo et al 1997]. A Bayesian network is a directed acyclic graph that represents a probability distribution. Nodes represent random variables and arcs represent probabilistic correlation between the variables. The types of paths (and lack thereof) between variables indicates probabilistic independence. Quantitative probability information is specified in the form of conditional probability tables. For each node the table specifies the probability of each possible state of the node given each possible combination of states of its parents. The tables for root nodes just contain unconditional probabilities.

The key feature of Bayesian networks is the fact that they provide a method for decomposing a probability distribution into a set of local distributions. The independence semantics associated with the network topology specifies how to combine these local distributions to obtain the complete joint probability distribution over all the random variables represented by the nodes in the network. This has three important consequences. First, naively specifying a joint probability distribution with a table requires a number of values exponential in the number of variables. In systems in which interactions among the random variables are sparse, Bayesian networks drastically reduce the number of values required. Second, efficient inference algorithms exist that work by transmitting information between the local distributions rather than working with the full joint distribution. Third, the separation of qualitative representation of the influences between variables from the numeric quantification of the strengths of the influences has a significant advantage for knowledge engineering. In building a Bayesian network model, one can first focus on specifying the qualitative structure of the domain and then focus on quantifying the influences. When finished, one is guaranteed to have a complete specification of the joint probability distribution.

The most common computation performed using Bayesian networks is determination of the posterior probability of some random variables, given the values of other variables in the network. Because of the symmetric nature of conditional probability, this computation can be used to perform both diagnosis and prediction. Other common computations are: computing the probability of the conjunction of a set of random variables, computing the most likely combination of values of the random variables in the network, and computing the piece of evidence that most influenced or will have the most influence on a given hypothesis. For a detailed discussion of Bayesian networks, focusing on inference techniques see the article by Bruce D'Ambrosio in this issue. Influence diagrams [Howard & Matheson, 1984] are a generalization of Bayesian networks for analyzing courses of action. In addition to chance nodes, they contain decision and value nodes. They share all the benefits of Bayesian networks.

Package	Company	Contact
Baron 2.0	KC Associates	kchang@gmu.edu
Analytica	Lumina Decision Systems	www.lumina.com
DX Solution Series	Knowledge Industries, Inc.	www.kic.com
Ergo	Noetic Systems, Inc.	www.noeticsystems.com
Graphical-Belief 2.0	MathSoft, Inc.	almond@acm.org or gmellman@statsci.com
HUGIN	Hugin Expert A/S	www.hugin.dk
Netica	Norsys Software Corp.	www.norsys.com

Table1: Commercial Bayesian network packages.

Package	Author(s)	Contact
BAYES		http://www.cs.cmu.edu/afs/cs/project/ai-repository/ai/areas/reasonng/probabl/bayes/0.html
BELIEF	Russell Almond	http://www.cs.cmu.edu/afs/cs/project/ai-repository/ai/areas/reasonng/probabl/belief/0.html
BN Toolbox	Kevin Patrick Murphy, U.C. Berkeley	http://www.cs.berkeley.edu/~murphyk/Bayes/bnt.html
BUGS	MRC Biostatistics Unit and Imperial College School of Medicine	http://www.mrc-bsu.cam.ac.uk/bugs/
IDEAL	Rockwell International	http://www.rpal.rockwell.com/ideal.html
JavaBayes	Fabio Cozman, U. São	http://www.cs.cmu.edu/~javabayes/Home/

	Paulo	
MacEvidence	Prakash Shenoy, U. Kansas	http://lark.cc.ukans.edu/~pshenoy/
MSBN	Microsoft Decision Theory and Adaptive Systems Group	http://www.research.microsoft.com/research/dtg/msbn/
Pulcinella	IRIDIA, Universite Libre de Bruxelles	http://iridia.ulb.ac.be/pulcinella/Welcome.html
Symbolic Probabilistic Inference (SPI)	Bruce D'Ambrosio	http://www.cs.orst.edu/~dambrosi/bayesian/frame.html
Genie/Smile	Decision Systems Laboratory, Univ of Pittsburgh	http://www2.sis.pitt.edu/~genie/
WebWeaver	Yang Xiang, U. of Regina	http://cs.uregina.ca/~yxiang/ww3/index.html

Table2: Free Bayesian network packages.

The practical value of Bayesian networks in building problem solving systems has spawned a small industry producing software for building and performing computations on Bayes nets. Table 1 lists some commercially available packages. Free demo versions can be downloaded for most of these packages and some packages are available free of charge or at greatly reduced prices to academic users for non-commercial purposes. The home pages for many of the companies listed also contain tutorials on Bayesian networks and archives of example networks. Table 2 lists several freely available packages.

For more information on available Bayesian network packages see the web sites
http://bayes.stat.washington.edu/almond/belief.html,
http://http.cs.berkeley.edu/~murphyk/Bayes/bnsoft.html, or
http://www.afit.af.mil/Schools/EN/ENG/LABS/AI/BayesianNetworks/tools3.html.

Applications of Bayesian Modeling and Inference Techniques

Perhaps the greatest testament to the usefulness of Bayesian problem solving techniques is the wealth of practical applications that have been developed in recent years. Here I sample a few of these in the areas of intelligent user interfaces, information filtering, autonomous vehicle navigation, weapons scheduling, and medical diagnosis. For a nice collection of papers on applications of Bayesian techniques, see the March 1995 special issue of *CACM* [CACM, 1995].

Lumière

Without a doubt, the single most widely distributed application of Bayesian inference techniques is Microsoft's Office Assistant, a Bayesian help system in the Office '97 suite of applications. The Office Assistant was based on prototypes developed within the Lumière project [Horvitz et al, 1998; Lumiere 1998] at the Decision Theory and Adaptive Systems Group of Microsoft Research. The goal of the Lumière project is the development and integration into computational systems of user models that continue to infer a user's goals and needs by considering the user’s background, actions, and queries. The approach taken is to develop Bayesian user models that capture the uncertain relationships among the goals and needs of a user and observations about program state, sequences of actions over time, and words in a user’s query. Observations are continuously input to a Bayesian model and a probability distribution over user needs is inferred. In addition, the system infers the likelihood that the user would like to receive assistance at the present moment. Ongoing and future research in the Lumière project includes learning Bayesian network models from user log data, using new sources of event information (such as data from automated vision and speech), and use of dialogue for obtaining information about user goals and needs. Other applications developed by the Decision Theory and Adaptive Systems Group include decision-theoretic troubleshooters that are available via the worldwide web.

Vista

In the Mission Control Center of the Johnson Space Center in Houston, teams of flight controllers work together around-the-clock monitoring and controlling each of the Space Shuttle orbiter's subsystems. Each team is responsible for making control decisions in high-stakes, time-critical situations. Project Vista [Horvitz & Barry, 1995; Project Vista] was initiated to develop techniques for providing on-line decision support to flight controllers, particularly by managing the complexity of the information displayed to them. Vista was developed by researchers at Rockwell Palo Alto Research Lab and Stanford University, working in close collaboration with an expert propulsion systems flight engineer at the Rockwell Space Operations Company. The system has been used at NASA Mission Control Center in Houston for several years. The system uses Bayesian networks to interpret live telemetry and provides advice on the likelihood of alternative failures of the space shuttle's propulsion systems. It provides a list of problems ordered by likelihood and separately by criticality. The system uses a model of time-criticality to control the level of detail displayed about particular subsystems, thereby directing flight controllers to the most important information. Software developers at Johnson Space Center are integrating the ideas from Vista into a variety of monitoring programs which are being installed in a new workstation-based Mission Control Center.

Lockheed Martin UUV

Lockheed Martin's Marine Systems in Sunnyvale, California and the Artificial Intelligence Center in Palo Alto are jointly developing an Autonomous Control Logic (ACL) system for demonstration in an Unmanned Underwater Vehicle (UUV) being developed by the United States Navy [Lockheed Martin, 1996]. The goal of the project is to develop software for a UUV that is capable of controlling planned and unanticipated events in a manner that minimizes risk of vehicle loss and maximizes the probability of successful completion of mission objectives. The ACL system will allow the UUV to monitor progress of its mission, analyze the health of its equipment, detect and analyze events that impact mission goals, make decisions and take actions to compensate for events, and modify its mission plan when the current one is no longer achievable. The ACL architecture is a hybrid of rule-based and Bayesian model-based techniques: the rule-based component provides real-time response, while the model-based component performs diagnosis, analysis, and decision-making about unanticipated events. The Model-Based Reasoner uses a

Bayesian network to model existing vehicle capabilities and the uncertainty regarding the state of those capabilities. It selects from the available alternatives the best response to the unanticipated event with the aim of maximizing the overall achievement of mission objectives.

Bayesian Ship Self Defense Tactics Engine

Scheduling ship self-defense systems is a complex problem due to the high-speeds and low trajectories of modern anti-ship missiles, which often makes them detectable only at close range. As a further complication, each shipboard self defense mechanism has constraints associated with it and the various systems can interact. The Bayesian Tactics Engine software [Musman & Lehner, 1999; Musman, 1999] is a real-time weapons scheduler designed to reside inside a Ship Self-Defense System. The Tactics Engine accounts for uncertainties caused by environmental conditions, sensor measurement errors, and threat identification errors. Bayesian networks are used to determine the optimal time to fire each self defense asset given the evidence from the ship's sensors. Because of limitations and constraints associated with each self defense asset, it is not always possible to implement a plan using the self defense assets in an optimal manner. Often conflicts occur such that a weapons system cannot fire at two or more different targets within the ideal plan. A transformational planner is used to resolve these conflicts and produce an optimal solution, working around the physical constraints on the self defense assets.

Microsoft Pregnancy and Child Care

In 1996 Microsoft's Health Product Unit released an online consumer health information service, Microsoft Pregnancy and Child Care, which has been previewed on the Microsoft Network. Bayesian-network models were constructed for different commonly occurring symptoms in children. At run time, an appropriate model is selected based on the chief complaint. The expert modules repeatedly determine the next best question to ask the parent, tailoring the multimedia presentations to the child's most likely health issues. Knowledge Industries, Palo Alto, CA working with the Decision Theory and Adaptive System Group at Microsoft Research developed and tested the Bayesian-network knowledge bases and associated inference procedures. Independent clinical testing was performed by a group of collaborating physicians affiliated with the University of Washington. The models were developed using the Microsoft Bayesian Network modeling and inference system, named MSBN. [AFIT Bayesian Networks, 1996] [Horvitz, 1999]

Pathfinder/Intellipath

Pathfinder is a Bayesian network based expert system for providing assistance with identification of disorders from lymph-node tissue sections [Heckerman et al, 1992]. The Pathfinder project at Stanford pioneered many technical and practical issues with the real-world use of large Bayesian networks. The success of Pathfinder led to its later commercialization as the Intellipath constellation of systems. The Intellipath set includes Bayesian models for lymph node pathology in addition to Bayesian models for 18 other tissue types, each representing a key area of expertise in the realm of surgical pathology. The initial lymph node models reason about 76 lymph node diseases and uses 105,000 subjectively-derived probabilities. Intellipath modules create a "differential diagnosis" of plausible diseases based on the histological features entered into the system. At any point in the diagnostic session, the user can ask the system to identify the features that would best help to distinguish among the competing diagnoses, considering the costs and benefits of each observation or test. Intellipath modules integrate Bayesian networks for pathology diagnosis with videodisc libraries of histological slides. An evaluation of the diagnostic accuracy of pathologists working with the assistance of the lymph node module concluded that pathologists working with the system produced significantly more correct diagnoses then those working without the system [Nathwani et al, 1997]. The assistance appears to be based in the information integration capabilities of the Bayesian model for lymph node diagnosis. Several hundred Intellipath systems are currently in use throughout the world. [Horvitz, 1999]

Knowledge-Based Model Construction and Structured Representations

The success of Bayesian networks lies largely in the fact that the formalism introduces structure into probabilistic modeling and cleanly separates the qualitative structure of a model from the quantitative aspect. Recent work has attempted to carry this theme further yet. Naïve use of Bayesian network technology would involve building one large network to represent a domain. For large systems, this is impracticable from an engineering standpoint. Although large domain models can often be decomposed into a collection of independent smaller networks, it is desirable, for systems that need to address a broad range of problems, to be able to assemble the needed model components dynamically. The lack of modularity in the representation also makes reuse of models difficult. A second limitation is that a Bayesian network is essentially a propositional representation of a domain: each node represents a multi-valued propositional variable. Thus it is not possible to express general relationships among concepts without enumerating all the potential instances in advance. This pre-enumeration is again impracticable when the system faces a broad range of dynamic decision situations. Researchers have endeavored to address these problems by augmenting the Bayesian network representation with concepts from programming languages and knowledge representation.

The first steps in this direction represented classes of Bayesian networks using sets of Horn-clauses with probabilities associated with them. Essentially, such a Horn-clause represented a node with its set of parents and the associated conditional probability table. Free variables in the Horn-clauses permitted expression of relationships among classes of individuals. Early work along this line produced algorithms for constructing Bayesian networks from such knowledge bases [Breese, 1992; Goldman & Charniak, 1990; Goldman & Charniak, 1993; Horsch & Poole, 1990; Wellman et al, 1992]. The algorithms were capable of producing small networks, tailored to the specific inference problem, resulting in computational savings in model evaluation. Later work provided a formal semantics for the knowledge base representation language and associated proofs of soundness and completeness for the process of constructing Bayesian networks and performing inference over them [Poole, 1991; Poole, 1993; Haddawy, 1994; Ngo & Haddawy, 1995; Ngo & Haddawy, 1997]. This knowledge-based model construction approach has been applied to problems such as military situation assessment [Mahoney & Laskey, 1998], student modeling for intelligent tutoring [Gertner et al, 1998], and synthesis of data analysis programs [Buntine et al, 1999]. Most recently, research has focused on yet more structured approaches, introducing concepts from object-oriented languages [Koller and Pfeffer, 1997] and frame-based languages [Koller and Pfeffer, 1998]. These languages provide support for structuring a model in terms of interacting components, as well as for building and reasoning about a domain model at different levels of abstraction.

Learning of Graphical Probability Models

Typically, the most difficult and time consuming part of the task in building a Bayesian network model is coming up with the probabilities to quantify it [Druzdzel et al, 1995]. Probabilities can be derived from various sources. They can be obtained by interviewing domain experts to elicit their subjective probabilities. They can be gathered from published statistical studies or can be derived analytically from the combinatorics of some problems, e.g. transmission of genes from parents to children. Finally, they can be learned directly from raw data. The learning of Bayesian networks and other graphical probability models has been one of the most active areas of research within the Uncertainty in AI community in recent years. Several excellent tutorials on learning of Bayesian networks from data are available [Buntine, 1996; Krause, 1998; Heckerman, 1998; Friedman & Goldszmidt, 1998], from which the following discussion is largely taken.

In addition to learning probabilities, we may wish to learn the structure of a Bayesian network. Learning of network structure can point out interesting relations in a domain, e.g. causal. There are also applications in which we simply have the need to learn autonomously, without a human providing the network structure. We can classify learning of Bayesian network models along two dimensions: data can be complete or incomplete and the structure of the network can be known or unknown. The following discussion touches on each of the four cases.

Known Structure

The most straightforward case is that in which the network structure is known and complete data is available for all variables in the network. A prior is assumed for each network parameter (probability table) and is updated using the available data. In the Bayesian learning literature, the Dirichlet distribution is commonly used as a prior for model parameters [Buntine, 1991]. (The special case in which the random variable has only two states is the well-known Binomial distribution.) The Dirichlet can express a large range of probability functions and its mathematical properties make the calculation of a posterior distribution from a prior relatively easy. The hyperparameters of the Dirichlet distribution have a natural interpretation in terms of the underlying sample size of the distribution. Thus in obtaining an estimate of a prior from a domain expert, we can ask how much past experience the estimate is based on, e.g., how many patient cases. Assuming that the model parameters are independent [Spiegelhalter & Lauritzen, 1990], the Dirichlet for each parameter can be updated independently.

The data from which we wish to learn a network may be incomplete for two reasons. First, some values may simply be missing. For example, in learning a medical diagnostic model, we may not have all symptoms for each patient. Here we can distinguish between values missing at random and values missing systematically. Values may be systematically missing because, for example certain tests are only run if certain readily observed symptoms are present. A classic approach to handling systematically missing values is to build a prior model about when the data will be missing and update the model using observed data [Rubin, 1974]. A second cause of incomplete data may be the lack of observability of some variables in the network. Such *hidden variables* can actually make the learning task easier in the sense that less data may be required than for the equivalent network in which all variables are observable [Russell et al, 1995].

Assuming that the data is missing at random, several techniques are available, of which the two most popular are Gibbs sampling and expectation-maximization. Both can handle continuous domain variables and dependent parameters. Gibbs sampling [Buntine, 1994] is a stochastic method that can be used to approximate any function of an initial joint distribution provided that certain conditions are met. First, for the distribution p(X) we must be able to sample any state of X given any possible initial state. This is satisfied if the joint distribution has no zeros. Second, each instantiation must be chosen infinitely often. This condition is met by iterating through the variables. Under these conditions, the average value of the sampled function approaches the expectation with respect to p(X) with probability 1 as the number of samples tends to infinity.

The expectation-maximization (EM) algorithm [Dempster et al, 1977] can be used to search for the maximum a posteriori (MAP) estimate of the model parameters [Lauritzen, 1995]. The EM algorithm iterates through two steps: the expectation step and the maximization step. In the first step, the expected sufficient statistics for the missing entries in the database D are computed. Any Bayesian network inference algorithm may be used to perform this step. In the second step, the expected sufficient statistics are taken as though they were the actual sufficient statistics for a database D' and the mean or mode of the parameters is calculated, such that the probability of D' given the network structure and parameters is maximized. The EM algorithm is fast but has the disadvantage of not providing a distribution over the model parameters. In addition, it can become stuck in local maxima, particularly when substantial amounts of data are missing.

Unknown Structure

The common approach to learning both structure and parameters from data is to introduce a scoring function that evaluates each network with respect to the training data, and then to search for the best network according to this metric. An obvious choice is the Bayesian score: the posterior probability of the network given the observed data. Unfortunately, this score is difficult to compute so that alternative criteria are typically used. The two most commonly used metrics are the belief scoring function [Cooper & Herskovits, 1992; Heckerman et al, 1995] and the minimal description length (MDL) based scoring function [Lam & Bacchus, 1994]. The MDL scoring function prefers networks that fit the data well and that are simple. It is an approximation of the Bayesian score: In the limit, as the number of cases in the database tends to infinity, MDL gives the same score as the Bayesian score, assuming Dirichlet distribution with uniform priors on structures. Both MDL and the belief scoring function use the likelihood function to measure how well the network fits the observed data. When the data is complete, the independencies in the network structure can be used to decompose the likelihood function into a product of terms. This allows for a modular evaluation of the candidate network and all local changes to it. Additionally, the evaluation of a particular change remains the same after changing a different part of the network.

When the data is incomplete, we can no longer decompose the likelihood function, and must perform inference to evaluate it, using either the EM algorithm or gradient descent [Binder et al, 1997]. The first step of the EM algorithm requires computing the probabilities of several events for each instance in the training data, and thus is inefficient. To make matters worse, a local change in one part of the network can affect the evaluation of a change in another part of the network, so that the neighbors of all networks visited must also be evaluated. This requires many calls to the EM procedure before making a single change to the current candidate network. Recently, Friedman introduced the innovation of performing the search for the best structure inside the EM procedure [Friedman, 1997]. He uses the current best estimate of the unknown distribution to complete the data and then use procedures that work efficiently for complete data. His approach maintains a current network candidate and at each iteration attempts to find a better network structure by computing the expected statistics needed to evaluate alternatives. Since this search is done in a complete data setting, it can exploit the decomposition properties of the scoring metrics. This algorithm applies only to scoring functions that approximate the Bayesian score, such as MDL. In more recent work, Friedman [1998] has extended the approach to work with the exact Bayesian score. There is evidence that the exact Bayesian score provides better assessment of the generalization properties of a model given the data. Furthermore, it provides a principled way of incorporating prior knowledge into the learning process.

Articles in This Special Issue

This introduction has provided only brief mention of the rich array of techniques available for inference in Bayesian networks. The article by Bruce D'Ambrosio provides a detailed discussion of Bayesian networks, briefly describing the representational aspects and then focusing on a variety of exact and approximate inference techniques and their mathematical foundations.

Since Decision Theory is concerned with rational choice among available actions, planning is a natural application for Bayesian techniques. The article by Jim Blythe shows how decision-theoretic planning extends the classical AI planning paradigm, outlines the central issues in decision-theoretic planning, and describes five alternative approaches that have been used to build decision-theoretic planners.

The article by Jon Doyle and Richmond Thomason is less a survey of previous accomplishments in the field and more a discussion of future directions in the development of decision-theoretic problem solving systems. They argue that the quantitative techniques of traditional Decision Theory have not proven fully adequate for supporting the attempts in Artificial Intelligence to automate decision making and that a more qualitative approach is called for. They provide an overview of the fundamental concepts of Decision Theory, a discussion of the need for the qualitative approach, and pointers to some recent work in this direction.

Acknowledgements
I thank Vu Ha, Eric Horvitz, Mark Kaplan, and Jude Shavlik for very helpful comments on drafts of this paper. This work was partially supported by NSF grant IRI-9509165.

References

[AFIT Bayesian Networks, 1996] Air Force Institute of Technology Artificial Intelligence Laboratory, WPAFB, Ohio. Bayesian Networks. http://www.afit.af.mil/Schools/EN/ENG/LABS/AI/BayesianNetworks/, 1996.

[AIStat-99] D. Heckerman and J. Whittaker (eds). *Proceedings of the Seventh International Workshop on Artificial Intelligence and Statistics*. Morgan Kaufmann Publishers, Inc., San Francisco, CA, 1999. Electronic version available at http://uncertainty99.microsoft.com/proceedings.htm.

[Binder et al, 1997] J. Binder, D. Koller, S. Russell, K. Kanazawa. Adaptive probabilistic networks with hidden variables. *Machine Learning* 29:213-244, 1997.

[Breese, 1992] J.S. Breese. Construction of belief and decision networks. *Computational Intelligence*, 8(4):624-647, 1992.

[Buntine et al, 1999] W. Buntine, K. Havelund, M. Lowry, T. Pressburger, S. Roach, P. Robinson, and J. Van Baalen. Transformation systems at NASA Ames. In*: Proc. of the Int'l Workshop on Software Transformation Systems*, Los Angeles, May 1999.

[Buntine, 1996] A guide to the literature on learning networks from data. *IEEE Transactions on Knowledge and Data Engineering,* 8(2):195-210, 1996.

[Buntine, 1994] W. Buntine. Operations for learning graphical models. *Journal of Artificial Intelligence Research*, 2:159-225, 1994.

[Buntine, 1991] W. Buntine. Theory refinement on Bayesian networks. In: *Proc. UAI-91*, 52-60, 1991.

[CACM, 1995] *Communications of the ACM,* Special issue on Real-World Applications of Bayesian Networks, D. Heckerman, A. Mamdani, M. Wellman (eds), 38(3), March 1995.

[Castillo et al, 1997] E. Castillo, J.M. Gutiérrez, A.S. Hadi. *Expert Systems and Probabilistic Network Models.* Springer Verlag, New York, 1997.

[Cooper & Herskovits, 1992] G. Cooper and E. Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning* 9:309-347, 1992.

[Dempster et al, 1977] A. Dempster, N. Laird, D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. B,* 39: 1-38, 1977.

[Druzdzel et al, 1995] M.J. Druzdzel, L.C. van der Gaag, M. Henrion, F.V. Jensen (eds). Working Notes of the IJCAI-95 Workshop on Building Probabilistic Networks: Where Do the Numbers Come From? Aug, 1995.

[Friedman, 1998] N. Friedman. The Bayesian structural EM algorithm. In *Proc. Fourteenth Conference on Uncertainty in Artificial Intelligence*, 129-138, 1998.

[Friedman, 1997] N. Friedman. Learning belief networks in the presence of missing values and hidden variables. In *Proc. Fourteenth Int'l Conf. on Machine Learning*, 125-133, 1997.

[Friedman & Goldszmidt, 1998] N. Friedman and M. Goldszmidt. Slides from the AAAI-98 tutorial on Learning Bayesian Networks from Data. http://www.cs.berkeley.edu/~nir/Tutorial/, 1998.

[Gertner et al, 1998] A.S. Gertner, C. Conati, K. VanLehn. Procedural help in Andes: Generating hints using a Bayesian network student model. In *Proc. AAAI-98*, 106-111, 1998.

[Goldman & Charniak, 1993] R.P. Goldman and E. Charniak. A Language for Construction of Belief Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(3):196-208, 1993.

[Goldman & Charniak, 1990] R.P. Goldman and E. Charniak. Dynamic construction of belief networks. In *Proc. of the Sixth Conference on Uncertainty in Artificial Intelligence,* 90-97, 1990.

[Haddawy, 1994] P. Haddawy. Generating Bayesian networks from probability logic knowledge bases. In *Proc. of the Tenth Conference on Uncertainty in Artificial Intelligence*, 262-269, 1994.

[Heckerman, 1998] D. Heckerman. A Tutorial on Learning with Bayesian Networks. Technical Report MSR-TR-95-06, Microsoft Research, Redmond, Washington. March 1995 (revised Nov 1996). Available at ftp://ftp.research.microsoft.com/pub/tr/TR-95-06.PS

[Heckerman et al, 1995] D. Heckerman, D. Geiger, and D.M. Chickering. Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning* 20:197-243.

[Heckerman et al, 1992] D. Heckerman, E. Horvitz, B. Nathwani. Towards normative expert systems: Part I. The Pathfinder project. *Methods of Information in Medicine*, 31(2):90-105, 1992.

[Horsch & Poole, 1990] M.C. Horsch and D. Poole. A Dynamic Approach to Probabilistic Inference using Bayesian Networks. In *Proc. of the Sixth Conference on Uncertainty in Artificial Intelligence,* 155-161, 1990.

[Horvitz, 1999] E. Horvitz. Personal communication. Jan 31, 1999.

[Horvitz et al, 1998] E. Horvitz, J. Breese, D. Heckerman, D. Hovel, and K. Rommelse. The Lumière project: Bayesian user modeling for inferring the goals and needs of software users. In: *Proc of the Fourteenth Conf on Uncertainty in AI,* 256-265, 1998.

[Horvitz & Barry, 1995] E. Horvizt and M. Barry. Display of Information for Time-Critical Decision Making. In: *Proc. Eleventh Conference on Uncertainty in Artificial Intelligence*, 296-305, 1995.

[Howard & Matheson, 1984] R.A. Howard and J.E. Matheson. Influence diagrams. In *Readings on the Principles and Applications of Decision Analysis*, R.A. Howard and J.E. Matheson (eds), vol. 2 719-762. Strategic Decisions Group, Menlo Park,CA, 1984.

[Jensen, 1996] F.V. Jensen. *An Introduction to Bayesian Networks.* Springer Verlag, New York, 1996.

[Koller, 1998] D. Koller. Structured Probabilistic Models: Bayesian Networks and Beyond. In *Proc. AAAI98*, 1210-1211, 1998.

[Koller & Pfeffer, 1997] D. Koller and A. Pfeffer. Object-oriented Bayesian networks. In: *Proc of the Thirteenth Conf on Uncertainty in AI*, 302-313.

[Krause, 1998] P.J. Krause. Learning Probabilistic Networks. Available at www.auai.org/bayesUS_krause.ps.gz, 1998.

[Lam & Bacchus, 1994] W. Lam and F. Bacchus. Learning Bayesian belief networks: An approach based on the MDL principle. *Computational Intelligence* 10(4):269-293, 1994.

[Lauritzen, 1995] S.L. Lauritzen. The EM algorithm for graphical association models with missing data. *Computational Statistics & Data Analysis*, 19:191-210, 1995.

[Lockheed Martin, 1996] Lockheed Martin Autonomous Control Logic to Guide Unmanned Underwater Vehicle. Press Release, Lockheed Martin Missiles and Space Communications Office, Palo Alto, CA, April 17, 1996.
http://lmms.external.lmco.com/newsbureau/pressreleases/1996/9604.html.

[Lumiere 1998] Lumière Project: Bayesian Reasoning for Automated Assistance.
http:\\research.microsoft.com\~horvitz\lum.htm.

[Mahoney & Laskey, 1998] S. M. Mahoney and K.B. Laskey. Constructing situation specific belief networks. In *Proc. Fourteenth Conf. on Uncertainty in AI*, 370-378, 1998.

[Musman & Lehner, 1999] S. Musman, P. Lehner. Real-Time Scheduling Under Uncertainty for Ship Self Defense, Submitted to *IEEE Expert*, Special Issue on Real-Time Intelligent System, 1999. Available at http://imsidc.com/~musman/personal/RT-Sched.ps

[Musman, 1999] S. Musman. Bayesian Ship Self Defense Tactics Engine.
http://imsidc.com/~musman/ssds/ssds.html, 1999.

[Nathwani et al, 1997] B.N. Nathwani, K. Clarke, M.C. Pike, S.P. Azen. Evaluation of an expert system on lymph node pathology. *Human Pathology* 28(9): 1097-1110, 1997.

[Neapolitan, 1990] R.E. Neapolitan. *Probabilistic Reasoning in Expert Systems.* John Wiley & Sons, Inc, New York, 1990.

[Ngo & Haddawy, 1997] L. Ngo and P. Haddawy. Answering Queries from Context-Sensitive Probabilistic Knowledge Bases. *Theoretical Computer Science*, 171(1-2):147-177, 1997.

[Ngo & Haddawy, 1995] L. Ngo and P. Haddawy. Probabilistic logic programming and Bayesian networks. In *Algorithms, Concurrency, and Knowledge (Proceedings ACSC95),* Lecture Notes in Computer Science, vol 1023, 286-300, Springer Verlag, 1995.

[Pearl, 1988] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo, CA, 1988. (Revised second printing 1991.)

[Poole, 1993] D. Poole. Probabilistic Horn abduction and Bayesian networks. *Artificial Intelligence*, 64(1):81-129, 1993.

[Poole, 1991] D. Poole. Representing Bayesian networks within probabilistic Horn abduction. In *Proc. of the Seventh Conference on Uncertainty in Artificial Intelligence*, 271-278, 1991.

[Project Vista, 1996] Project Vista: Managing the Display of Complex Information. http://www.rpal.rockwell.com/~barry/vista.html, 1996.

[Rubin, 1974] D.B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *J. of Edu. Psych.*, 66:688-701, 1974.

[Russell et al, 1995] S. Russell, J. Binder, D. Koller, K. Kanazawa. Local learning in probabilistic networks with hidden variables. In: *Proc of the Fourteenth Int'l Joint Conf on AI,* 1146-1152, 1995.

[Savage, 1954] L. J. Savage. *The Foundations of Statistics.* John Wiley & Sons, New York, 1954. (Second revised edition published 1972 by Dover Publications, Inc.)

[Spiegelhalter & Lauritzen, 1990] D.J. Spiegelhalter and S.L. Lauritzen. Sequential Updating of Conditional Probabilities on Directed Graphical Structures. *Networks*, 20:579-605, 1990.

[UAI-98] *Proc. of the Fourteenth Conf. on Uncertainty in Artificial Intelligence,* G.F. Cooper and S. Moral (eds), Morgan Kaufmann, San Francisco, 1998.

[Wellman et al, 1992] M. Wellman, J.S. Breese, R.P. Goldman. From knowledge bases to decision models. *The Knowledge Engineering Review*, 7(1):35-53, 1992.