REPORT DOCUMENTATION PAGE

AFRL-SR-AR-TR-04-

0046

1. REPORT DATE *(DD-MM-YYYY)*
07-20-2000

2. REPORT TYPE
Final performance report

3. DATES COVERED *(From - To)*
12/1/1999 to 11/30/2002

4. TITLE AND SUBTITLE

Enhancements of Systems Based on Bayesian Networks and Structural Equation Models for Command and Control Support

5a. CONTRACT NUMBER

5b. GRANT NUMBER
F49620-00-1-0112

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)
Marek J. Druzdzel, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Pittsburgh
School of Information Sciences
135 North Bellefield Avenue
Pittsburgh, PA 15260

Phone: 412-624-9432
Fax: 412-624-2788
Email: marek@sis.pitt.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

20040203 050

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A
Approved for Public Release
~~Distribution Unlimited~~

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The performed project focused on a new paradigm of planning systems that are based on a combination of Bayesian networks and structural equation models. We focused on theoretical issues that surround combining the two in a practical planning system, developing the foundations for, and building a prototype of such system. The approach and the system built allow for efficient, yet normatively correct, treatment of various types of information, uncertainty, and utility. It is especially powerful in complex situations where the available information is heterogeneous and consists of a mixture of deterministic and uncertain relationships among discrete and continuous variables.
Our main contributions are: (1) two state of the art stochastic sampling algorithm for approximate inference in graphical models, both (2) analysis of problems related to combining probabilistic information, (3) an module for interactive construction of causal graphical models and search for opportunities, (4) algorithm for learning graphical models from small data sets, and (5) a prototype of the system, used by over 5,000 people world-wide.

15. SUBJECT TERMS
Bayesian networks, structural equation models, graphical models, uncertainty, decision making.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Marek J. Druzdzel
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)* 412-624-9432

Major Accomplishments

Major accomplishments of the project have been:

(1) two state of the art stochastic sampling algorithm for approximate inference in graphical models, both being two fastest algorithms available,

(2) theoretical analysis of problems related to combining information from various sources in building probabilistic models,

(3) an interactive module for construction of causal graphical models that deals with reversible causal mechanisms,

(4) a module that performs search for opportunities,

(5) algorithms for learning probabilities from small data sets, and

(6) a prototype of the system, used by over 5,000 people world-wide.

We briefly summarize each of these in the separate sections below.

Stochastic sampling algorithms

A system that is a combination of Bayesian networks and structural equation models needs to include algorithms that are flexible enough to work with both discrete (Bayesian networks) and continuous (structural equation models) variables. The algorithms have to accommodate arbitrary probability distributions and work with very large models. The only known classes of algorithms that will accommodate these requirements are stochastic sampling algorithms. In our work (starting with the previous AFOSR grant), we probed three directions: Latin hypercube sampling, quasi-Monte Carlo methods, and adaptive importance sampling. In some of the papers resulting from the previous grants we also acknowledged the current grant, as additional experiments or finishing touches on the papers were performed after the termination date of the previous grant.

The new directions of our work were analyzing the convergence of stochastic sampling and investigating the confidence intervals around the result of sampling algorithms. One of the useful results of this work is the ability of a stochastic sampling algorithm to self-reflect and predict how many more samples are needed to achieve a given precision. Two related publications in this area are an article in *Computational Statistics* (Cheng 2001) and another in the prestigious *Conference on Uncertainty in Artificial Intelligence* (Cheng and Druzdzel 2001).

Our later work on sampling algorithms has led to the design of the EPIS-BN (Estimated Posterior Importance Sampling), an algorithm that is even more efficient than the AIS-BN algorithm developed in our previous grant. EPIS-BN uses an algorithm known as Loopy Belief Propagation (LBP) to compute an estimate of the posterior probability distribution in a Bayesian network. The LBP algorithm is a modification of an exact belief propagation algorithm for singly-connected Bayesian networks proposed in mid-1980s by Judea Pearl. In case of singly-connected networks, its complexity is polynomial but unfortunately it does not extend to multiply-connected networks and suffers from possible infinite loops and local minima in terms of its precision. In the EPIS-BN algorithm, we rely on the fact that the LBP algorithm typically produces results that are close to the posterior probability distribution over the network. Once we have the results of the LBP algorithm, we can use these as the importance function in an importance sampling algorithm. This algorithm produces excellent results and does not require a costly learning stage of the AIS-BN algorithm that we developed previously.

Our results published in a paper on the AIS-BN algorithm were excellent – on real, hard cases, when the probability of evidence is very low, the algorithm has beaten previous algorithms by two orders of magnitude in terms of its precision. In terms of computing time required to reach the same precision, the results were even better. Here is a typical experimental result of the AIS-BN algorithm:

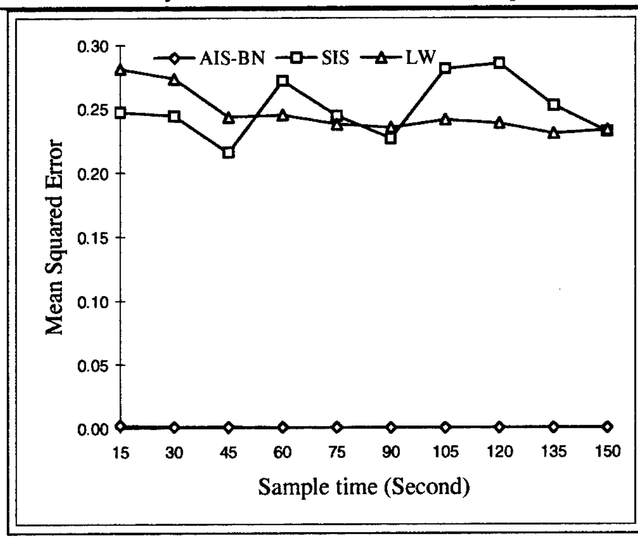

Figure 1: Observed example convergence rate improvement s in the proposed adaptive importance sampling algorithm for Bayesian networks (AIS-BN).

Figure 3 shows example performance comparison of the three algorithms. Figure 4 shows the performance of the AIS-BN algorithm at a finer scale.

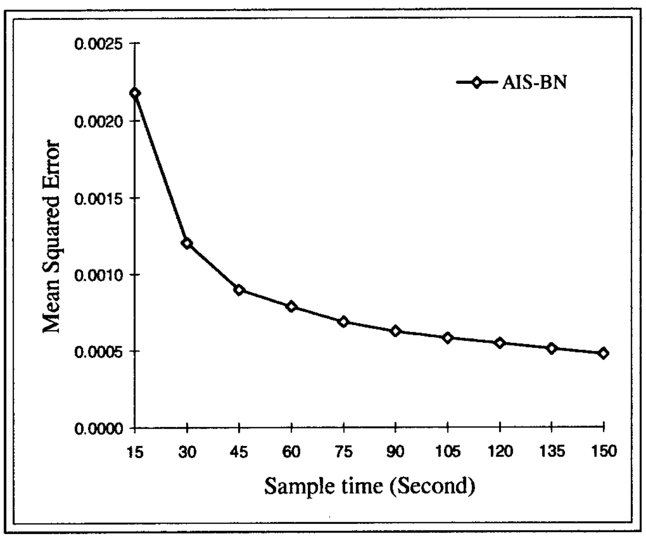

Figure 2: Observed example convergence rate improvements in the proposed adaptive importance sampling algorithm for Bayesian networks (AIS-BN): A close-up of the adaptive importance sampling algorithm in Figure 1).

The EPIS-BN algorithm improves these phenomenal results even further. We tested it on several large real Bayesian networks and compared the results with the AIS-BN algorithm. The empirical results showed that the EPIS-BN algorithm provides a considerable improvement over the AIS-BN algorithm, especially in those cases that are hard for the latter. Figures 3 and 4 show typical results obtained in our tests.

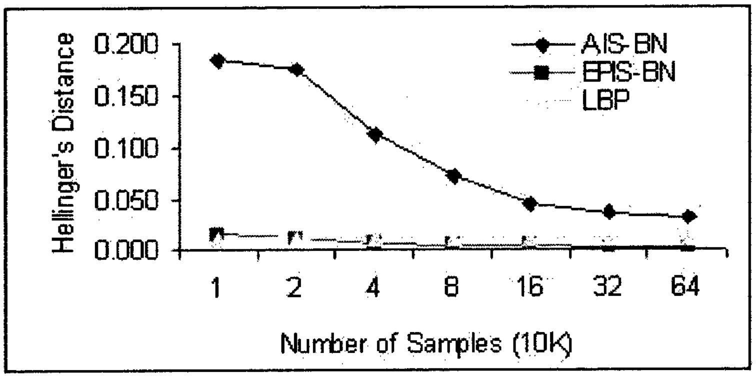

Figure 3: Convergence rate comparison for AIS-BN and EPIS-BN as a function of the number of samples on the ANDES network.

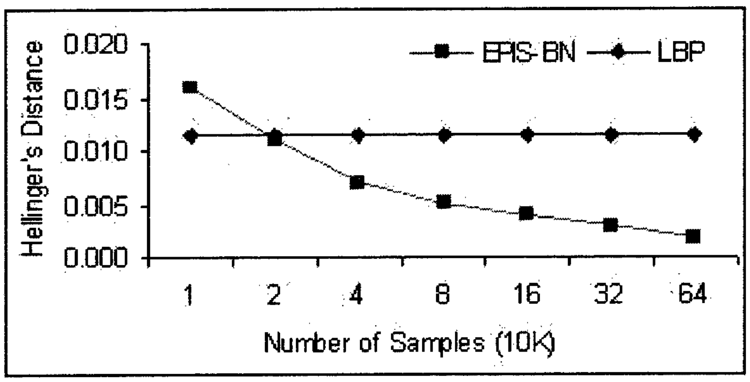

Figure 4: Convergence curve for EPIS-BN in a finer scale. The horizontal line shows the accuracy reached by loopy belief propagation.

We have tested to what degree these results can be improved further. Probabilistic logic sampling, the first stochastic sampling algorithm for Bayesian network, with no evidence is equivalent to importance sampling with a perfect importance function (the prior probability distribution!). When run on the networks that we used in our tests, the probabilistic logic sampling algorithm achieves precision on the order of 10^{-4}, which is only slightly better than the precision that the EPIS-BN algorithm reached. It seems, we conclude, that EPIS -BN is close to what sampling algorithms for Bayesian networks can achieve. We have presented the EPIS-BN algorithm in the prestigious *Conference on Uncertainty in Artificial Intelligence* this year and are working on a journal submission (*Mathematical and Computer Modelling*, special issue on *Optimization and Control for Military Applications*, edited by Dr. Juan Vasquez).

Combining information from various sources in building probabilistic models

One of the most serious hurdles in practical application of probabilistic methods is the effort that is required of model building and, in particular, of quantifying graphical models with numerical probabilities. Knowledge engineers quantifying probabilistic models usually combine various sources of information, such as existing textbooks, statistical reports, databases, and expert judgment. However, lack of attention to whether the sources are compatible and whether they can be combined may lead to erroneous behavior of the model. For instance, an unwary knowledge engineer might combine the prevalence of a certain disease, obtained from a general-population study, with the sensitivity and specificity of a certain test obtained at hospital. This combination of information may lead to a several orders of magnitude error in the computation of the posterior probabilities of interest. While most knowledge engineers realize the danger of misapplication of data that describe different population groups, they often fail to appreciate purely statistical effects that play a role in probabilistic information. Even though one might think that no experienced knowledge engineer would make such a mistake, the fact that sensitivity and specificity may be biased when obtained from a subpopulation has never been mentioned in Bayesian network literature. Even in medical literature, it is not uncommon to find values of sensitivity and specificity without an explanation of how they were obtained, because they are assumed to be invariant. After all, sensitivity and specificity do not depend on the prevalence. Builders of probabilistic models realize that different population characteristics, such as sex, race, diet, etc., can influence both sensitivity and specificity, but we forget about purely statistical phenomena such as conditioning.

Although variability of sensitivity and specificity has been reported in the medical literature for decades---see, for instance (Ransohoff 1978) and (Knottnerus 1987) – many of today's epidemiological studies on the assessment of diagnostic tests fail to mention it, and, to our knowledge, researchers in the area of artificial intelligence have never considered it when building probabilistic models. This entails a significant risk because, as we have shown, collecting these statistics in one setting and using them in another can lead to errors in posterior probabilities as large as several orders of magnitude. We used the framework of directed probabilistic graphs to systematize our observation, to explain the risks of naive knowledge combination, and to offer practical guidelines for combining knowledge correctly. The problems that we pointed out are due to purely statistical effects related to selection phenomena. They may occur when data or knowledge are collected from different subpopulations and subsequently combined into one model, or even when the parameters for a causal model are obtained from the same subpopulation in which the model is applied. On the contrary, these problems have nothing to do with small databases, missing data, or unreliable expert judgment.

On the other hand, an over-cautious position of never combining numerical data obtained from different sources would result in disregarding valuable information, which might be useful in model construction. In fact, we have shown that the criteria "do not combine knowledge from different sources" and "obtain all the data from the subpopulation in which the model will be applied" are neither necessary nor sufficient to guarantee the correctness of the model. For this reason, we have introduced a criterion for combining data from different sources, namely that the causal graph, built from expert knowledge, is *linearly ordered.* We have also offered an algorithm for making the graph linearly ordered by adding links that represent the probabilistic dependencies induced by selection mechanisms. Knowledge engineers must not ignore this property, because the absence of those links may lead to important errors in the computation of the probabilities, even when all the probabilities were obtained from the subpopulation in which the model is applied.

Our key results, published in *Journal of Machine Learning Research* (Druzdzel & Diez 2003) are captured in the following two theorems:

Theorem 1

Given a selection variable X_s in a Bayesian network and a node X_i (other than X_s), such that X_i is not an ancestor of X_s, the conditional probability distribution of X_i given Parents(X_i) is the same in the general population and in the subpopulation induced by value x_s, i.e.,

$$\Pr(x_i|\mathrm{Pa}(x_i), x_s) = \Pr(x_i|\mathrm{Pa}(x_i)) .$$

Definition 2

A graph is *linearly ordered* for X_s iff

$$\forall X_i, X_i \in \{X_s\} \cup Anc(X_s), \exists X_j, X_j \in Pa(X_i), \exists X_k, X_k \in Pa(X_i)$$
$$\Rightarrow (X_j = X_k) \vee (X_j \in Pa(X_k)) \vee (X_k \in Pa(X_j)) .$$

This property can be phrased as follows: if X_s or an ancestor of X_s (say X_i) has two parents (X_j and X_k), then one of the two must be a parent of the other. Obviously, if each ancestor of X_s has only one parent, then the graph is linearly ordered for X_s.

Definition 3

A *causal Bayesian network* is *linearly ordered* for X_s if its graph is linearly ordered for X_s.

Theorem 4

Given a Bayesian network that is linearly ordered for X_s, for each configuration $\mathbf{x_R}$ of the variables in $\mathbf{X_R} = \mathbf{X} \setminus \{\mathbf{X_s}\}$, it holds that

$$\Pr(\mathbf{x_R} \mid \mathbf{x_s}) = \Pi_{i \neq s} \Pr(x_i|Pa(x_i), x_s) .$$

The theorems, based on Markov condition will help the knowledge engineer determine whether some of those variables can be removed from the graph, provided that the conditional probabilities of their ancestors are coherently chosen. In contrast, when a node is not an ancestor of any of those selection variables, its conditional probability is invariant and can be obtained from any source.

The conclusions of our analysis are general, applicable in model building across domains. One example is medical or machine diagnosis, where models are built based on a combination of hospital/field experience, physiological model/device specification, and hospital/repair shop data. Yet another is fraud detection, where models are based on general population characteristics combined with customer transaction data. Yet another is detection and prevention of terrorist activities, where the information consists of intelligence reports, past cases, and surveillance data.

Our motivating examples were based on a medical data set, but the same argument can be made with respect to numbers obtained from human experts. Subjective probability judgments have been shown to rely on judgmental heuristics (Kahneman and Tversky 1982) and they are very sensitive to prior experiences (in fact prior experiences are often all that probability judgments are based on). Humans have been shown to be able to match the probability of observed events with an amazing precision in certain experiments (Estes 1976). Physicians working in a hospital will tend to match the sensitivity and specificity of medical symptoms and tests that they observe in their practice. These are often determined by the circumstances, such as what brought the patients to the hospital or clinic in the first place. Physician experts will tend to at least adjust the parameters to what they observe in their practice. While their experience is valuable for building decision models for the particular clinics where they have worked, in general they cannot be readily used in other settings. Similarly, one cannot assume that this knowledge can be combined with data originating from other settings.

Although the main focus or our work was knowledge engineering, it sheds light on other fields. From the point of view of machine learning, it emphasizes the importance of selection biases in the automatic construction of causal models from databases. It can also be useful when one or several agents look for information (for instance, by searching the Internet) and try to build a model by combining information extracted from several sources. In this scenario, the agent should use qualitative knowledge as a guide for combining numerical data. A particular case of this scheme would be the development of a tool for automated elicitation of knowledge through interaction with human experts, similar to those that exist for building rule-based expert systems. Finally, from the point of view of statistics, our work is useful for the application of causal models in epidemiology (Greenland 1999, Pearl 2000, Hernan 2002), in which the analysis of data (in general, selected data) is based on a causal graph built from expert knowledge. Our analysis might also be applied to meta-analysis, a technique that has become popular in the last years, especially in medicine, based on extracting data from different epidemiological studies published in the literature and combining them in order to draw more reliable or more precise conclusions. The data of each study and the collection of studies are prone to selection biases (see, for instance, Macaskill 2001).

Support for construction of causal graphical models

Causal models based on structural equations have become a major formalism for representation of causal relations and reasoning, such as predicting effects of actions, deriving causal relations from data, and generating causal explanations for observed events. Since the quality of causal reasoning depends directly on the quality of the underlying models, we focused our work on (1) providing a sound and effective methodology in constructing requisite causal models, (2) supporting derivation of the effects of actions with systems containing reversible mechanisms, and (3) assisting decision makers in achieving decision objectives by searching for novel interventions. This work led to the publications in the prestigious *Annual Conference on Uncertainty in Artificial Intelligence* (Lu and Druzdzel 2000), *European Conference on Symbolic and Qualitative Approaches to Reasoning with Uncertainty* (Lu and Druzdzel 2001), *European Workshop on Probabilistic Graphical Models* (Lu and Druzdzel 2002), and a doctoral dissertation (Lu 2003). We are planning journal submissions based on this work. In addition, we have developed a working system *ImaGeNIe* that supports causal model construction and utilization. Figure 1 shows the architecture of *ImaGeNIe.*

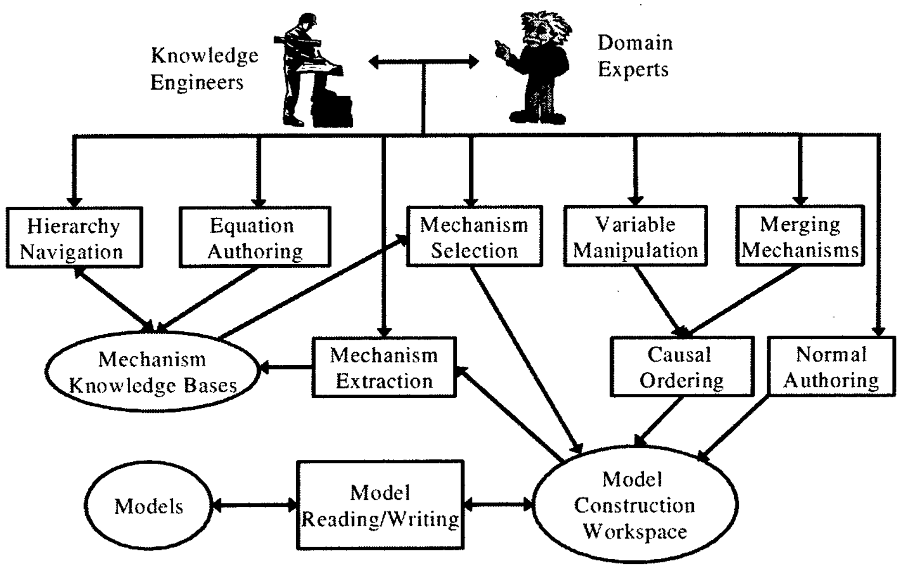

Figure 1: System architecture of *ImaGeNIe.*

ImaGeNIe includes three knowledge structures: *mechanism knowledge bases*, which holds domain knowledge expressed as causal mechanisms, *model building workspace*, which serves as a blackboard for model composition, and *models*. The domain knowledge can be maintained either by *equation authoring interface*, or by *mechanism extraction operation* that enables model builders to extract reusable mechanisms from existing models. Model builder can use *hierarchy navigation interface* to locate the mechanism of interest and select them into the model building workspace with assistance of the *mechanism selection operation*. In addition to mechanism selection and traditional model authoring operations, model builder can *manipulate variables* and *merge mechanisms* as the model building process evolves. The underlying *casual ordering* module restructures the models according to users' interactions with the system. Figure 2 shows a typical *ImaGeNIe* graphical user interface.

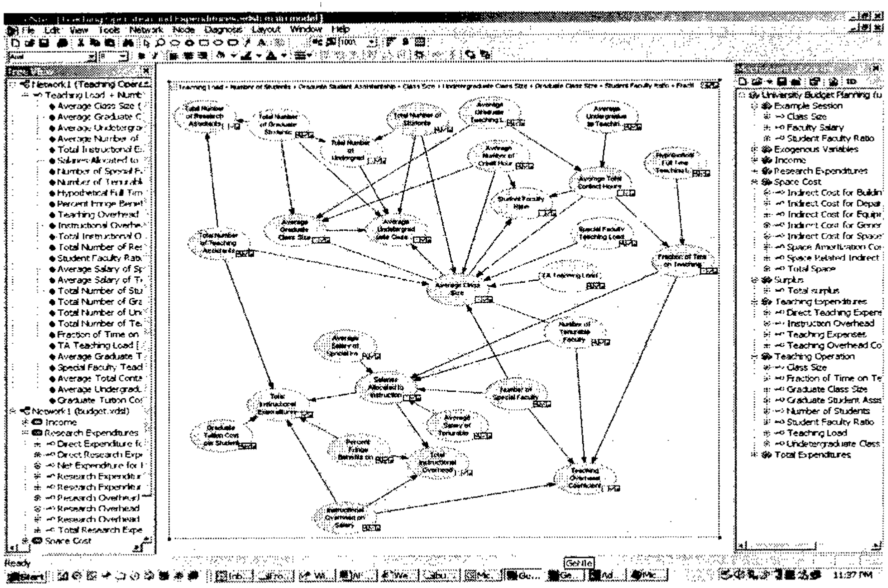

Figure 2: *ImaGeNIe* graphical user interface.

The mechanism-based view of causality, first proposed by Simon (1953) as the theory of causal ordering, is the theoretical foundation of the implementation of *ImaGeNIe.* The theory of causal ordering explicates the causal relations in a self-contained structure model into a causal graph. We extended the theory of causal ordering to explicate causal relations in an under-constrained structure model such that its graphical representation can represent decision makers' intermediate understanding of decision problems. The model construction process in *ImaGeNIe* can be viewed as the process of assembling mechanisms from under-constrained models into self-contained models. Figure 3 shows an under-constrained models and mechanisms that are ready to be merged into the under-constrained model. We have found in an empirical test that *ImaGeNIe* can effectively assist users in constructing causal models for causal reasoning.

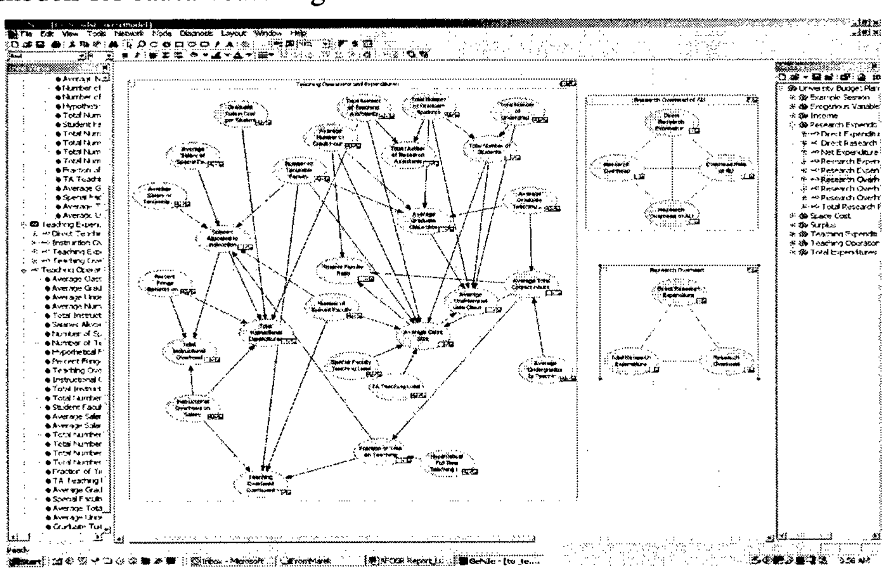

Figure 3: An example model session with *ImaGeNIe*.

In addition to providing decision makers with a sound methodology for building causal models, we assist decision makers in deriving the effects of manipulations on systems containing reversible mechanisms. The works of Pearl (1993, 2000) and Spirtes et al (1993, 2001) has focused on predicting the effects of actions for systems containing only irreversible mechanisms. Their approach of predicting the effect of action has been referred as the 'arc-cutting' approach. For example, the rain (R) can get us wet (W), R→W, however, wearing the rain coat can prevent us getting wet but it does not make the rain go away, i.e., the arc between R and W will never reverse. Druzdzel (1992) recognized that the causal reversibility in systems containing reversible mechanisms. For example, in the power train of a car, we normally have engine (E) that drives the wheel (W) through transmission (T), E→T→W, however, when we drive a car down a hill, it is common practice to slow down the care by switching to a lower gear. In other words, causal relations among the variables in question have reversed to E←T←W. This type of reasoning requires prior knowledge of what mechanisms will be brought into the system due to the manipulation and what mechanisms will be released from the system to maintain as self-contained. This reasoning is known as *changes in structure* in econometrics.

To support predicting the effects of actions for systems that consist of mixtures of mechanisms, we formalized the representations of causal reversibility and action operator. We defined the set of *effect variables* as a property of a mechanism. A mechanism can be categorized into three categories according to their *reversibility*: (1) *completely reversible*: every variable in the mechanism can be an effect variable, (2) *partially reversible*: some of the variables in the mechanism can be effect variable, and (3) *irreversible*: exactly one of the variables in the mechanism can be an effect variable. We draw the analogy between changes in structure and STRIPS-like action language (popular in AI) to define the action operator **Act(E, Epre, Eadd, Edel)** where **E** is the model that an action applies on, **Epre** is the set of preconditions that must be satisfied before an action can be applied, **Eadd** is the set of structural equations to be added into **E**, and **Edel** is the set of structural equations to be removed from **E.** In addition, we assist decision makers in deliberating an action, namely reasoning about which structural equations should included in **Eadd** or **Edel.** In particular, we developed algorithms to answer two types of queries: (1) When manipulating a causal model, which mechanisms are possibly invalidated and can be removed from the model? (2) Which variables may be manipulated in order to invalidate and, effectively, remove a mechanism from a model? Figure 4 shows the support of changes in structure in *ImaGeNIe* based on these two algorithms.

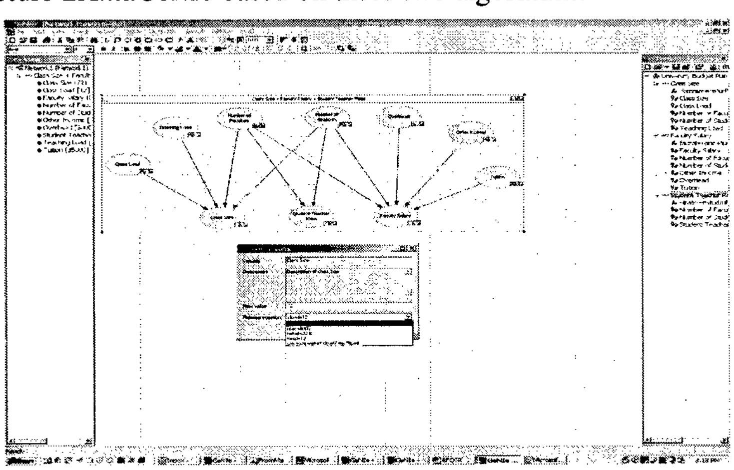

Figure 4: Changes in structure in *ImaGeNIe.*

Search for opportunities

Although changes in structure assist decision makers in predicting effects of actions, decision makers still need to provide partial parameters for an action operator, namely **Eadd** or **Edel**, for deliberating an action. We took a step further to address the decision scenarios in which none of **Eadd** or **Edel** is given but a causal model and a decision objective. This decision scenario happens when a decision maker who is confronted with a complex system does not know which variables to best manipulate or to observe to achieve a desired objective. We refer to this problem as *search for opportunities*, which amounts to both identifying the set of policy variables and computing their optimal setting for a given decision objective. To solve the problem of search for opportunities, we introduced the concept of *value of intervention* which arises from considering jointly the economic factors and effects of actions in causal models. We proposed *augmented causal models,* which allow molders to specify observability, manipulability, and focus as the property of variables, to describe a decision problem at hand. We developed myopic search algorithms to solve the problem of search for opportunities for systems. The algorithm looks one step ahead to compute the value of intervention for each manipulable variable in the model and yields the optimal sequence of actions. Figure 5 shows how the algorithms perform myopic search for opportunities.

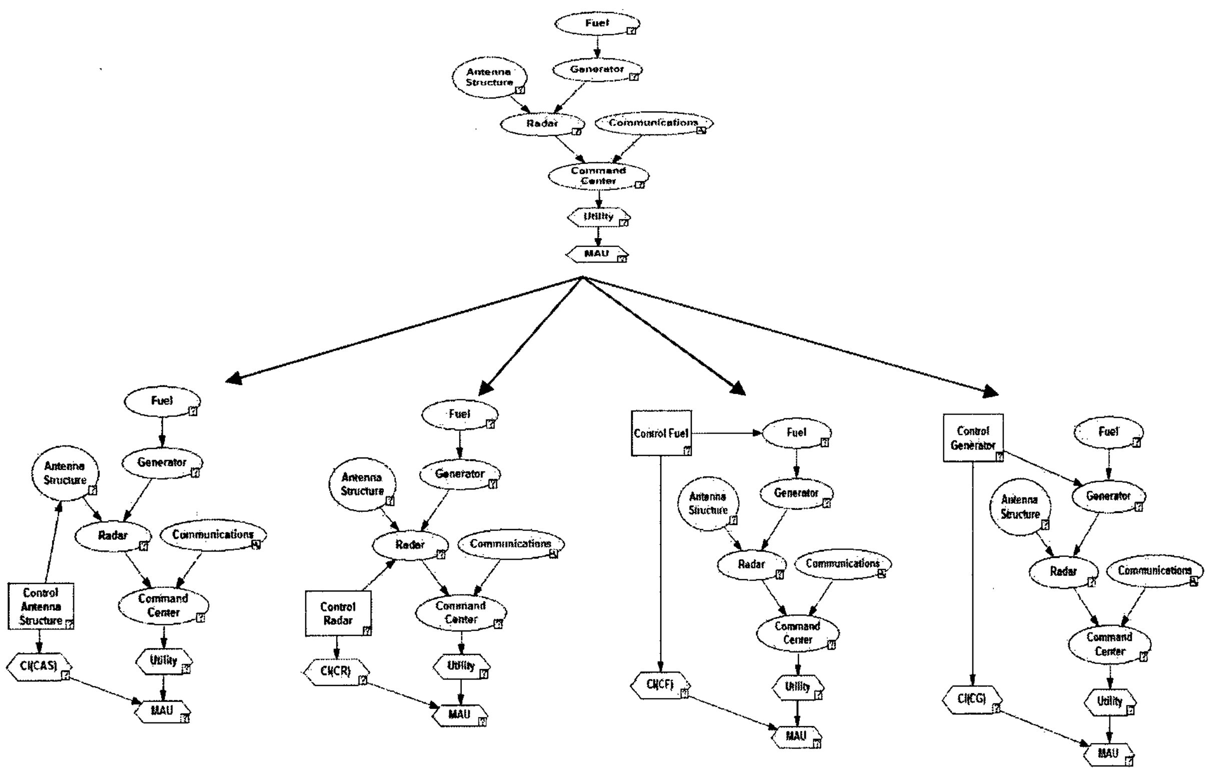

Figure 5: Myopic search for opportunities

The myopic search for opportunities can be applied by a robot to find out the next most effective action. It can also be used in an interactive modeling environment, where we present to users a list of actions ranked by their values of intervention computed by the myopic search for opportunities (shown in Figure 6). Users then have the option to override systems' suggestion to select the action that is not ranked highest in the list. This allows users to perform 'what if' analysis in generating decision sequences.

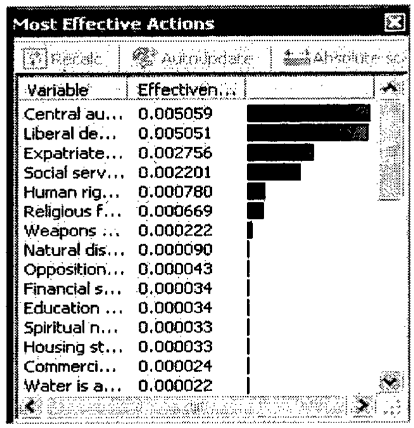

Figure 6: A ranked list of interventions computed by the myopic search for opportunities.

Learning probabilities from small data sets

The focus of our work in this area, published in the *International Journal of Approximate Reasoning*, was learning CPTs in Bayesian network models from small data sets given an existing network structure. Learning CPTs amounts essentially to counting data records for different conditions encoded in the network. Roughly speaking, prior probability distributions are obtained from relative counts of various outcomes for each of the nodes without predecessors. Conditional probability distributions are obtained from relative counts of various outcomes in those data records that fulfill the conditions described by a given combination of the outcomes of the predecessors (this combination of parents' outcomes is often referred to as *conditioning case*). While prior probabilities can be learned reasonably accurately from a database consisting of a few hundred records, learning CPTs is more daunting. In small data sets, many conditioning cases are represented by too few or no data records and they do not offer sufficient basis for learning conditional probability distributions. In cases where there are several variables directly preceding a variable in question, individual combinations of their values may be very unlikely to the point of being absent from the data file. In such cases, the usual assumption (direct or indirect, by means of Dirichlet priors) made in learning the parameters is that the distribution is uniform, i.e., the combination is completely uninformative.

A CPT offers a complete specification of a probabilistic interaction that is powerful in the sense of its ability to model any kind of probabilistic dependence between a discrete node Y and its parents $X_1, \ldots, X_n$. However, when learning the conditional probability distribution from data sets, this precision can be illusory. If the size of the data set is small, many of the CPT entries will have be learned from an insufficient number of records, undermining the very purpose of a full specification. We proposed enhancing the process of learning the CPTs from data by combining the data with structural and numerical information obtained from an expert. Given expert's indication that an interaction in the model can be approximated by a Noisy-OR gate (Henrion 1989, Pearl 1988), we first estimate the Noisy-OR parameters for this gate. Subsequently, in all cases of a small number of records for any given combination of parents of a node, we generate the probabilities for that case as if the interaction was a Noisy-OR gate. Effectively, we obtain a conditional probability distribution that has a higher number of parameters. At the same time, the learned distribution is smoothed out by the fact that in all those places where no data is available to learn it, it is reasonably approximated by a Noisy-OR gate. Noisy-OR distributions approximate CPTs using fewer parameters and learning distributions with fewer parameters is in general more reliable (Friedman et al. 1999).

We tested our approach on Hepar II, a Bayesian network model for diagnosis of liver disorders consisting of 73 nodes. The parameters of Hepar II are learned from a data set of 505 patient cases. We showed that the proposed method leads to an improvement in the quality of the model as measured by its diagnostic accuracy. While the observed improvement in accuracy were modest (only 6.7% and 14.3% in comparison to a multiple-disorder model and single-disorder model respectively), it was obtained at a negligible cost, which makes our method attractive in practice.

For each combination of a node and its parents (a family) in the multiple-disorder version of the Hepar II model, we verified with our expert whether the interaction could be approximately modeled by a Noisy-OR gate. The expert identified 25 nodes (from among the total of 62 nodes with parents) that could be reasonably approximated by Noisy-OR gates. Testing the Noisy-OR assumption for each of the gates with the expert was quite straightforward once the expert had understood the concept of independence of causal interaction. When deciding whether an interaction can be approximated by a Noisy-OR gate, we followed the criteria proposed by Diez (1997). An interaction can be approximated by a Noisy-OR gate if it meets the following three assumptions: (1) the child node and all its parents must be variables indicating the degree of presence of an anomaly, (2) each of the parent nodes must represent a cause that can produce the effect (the child variable) in the absence of the other causes, (3) there may be no significant synergy among the causes.

Each of the such identified Noisy-OR gates was subject to the following learning enhancement. Whenever there were sufficiently many records for a given conditioning case, we used these records to learn a corresponding element of the CPT. When there were no or very few data records, we generated the CPT entry from our Noisy-OR parameters. Effectively, the complete CPT, once learned, was a general CPT with a fraction of its elements generated using the Noisy-OR assumption. The assumption that we made was that a general conditional probability table will fit the actual distribution better than a Noisy-OR distribution. Noisy-OR will fit better than a uniform distribution in those cases when there was not enough data to learn a distribution.

We performed a series of empirical tests of diagnostic accuracy of various versions of the model. In order to make the comparison fair, we used the same data set for learning the parameters of each of the models. Our data set contained 505 patient records classified in 9 different disorder classes. In each case we used the same measure of accuracy: diagnostic performance using the leave-one-out method (Moore 1994). Essentially, given n=505 data records, we used n-1 of them for learning model parameters and the remaining one record to test the model. This procedure was repeated n times, each time with a different data record. In our tests, we used as observations only those findings that were actually reported in the data (i.e., we did not use the values that were missing, even though we used their assumed values in learning). The diagnosis for each patient case was calculated given the evidence, i.e., a subset of the 66 possible observations such as symptoms, signs and the laboratory tests results. These data did not include the results of a biopsy. By accuracy we mean the proportion of records that were classified correctly. Whenever we report accuracy within a class, we report the fraction of records within that class that were classified correctly.

Our second test aimed at comparing the diagnostic accuracy of the plain multiple-disorder model to the models whose probabilities were smoothed out using the Noisy-OR parameters. Here, we focused on three models: (1) the plain multiple-disorder model (i.e., general CPT) and two models enhanced with: (2) Noisy-OR parameters obtained from data, and (3) Noisy-OR parameters assessed by the expert.

Our enhancement process replaced those elements of the CPT that had not enough data records to learn a distribution reliably, i.e., when the number of records found in the data set was lower than a *replacement threshold* (we specified this threshold as a percentage of all records in the data set, i.e., a threshold of 10% corresponds roughly to 50 records). Figure 7 shows the relationship between the replacement threshold and the percentage of all CPT entries that were replaced by the Noisy-OR distributions. The percentage of replaced CPT entries seems to be directly proportional to the replacement threshold.

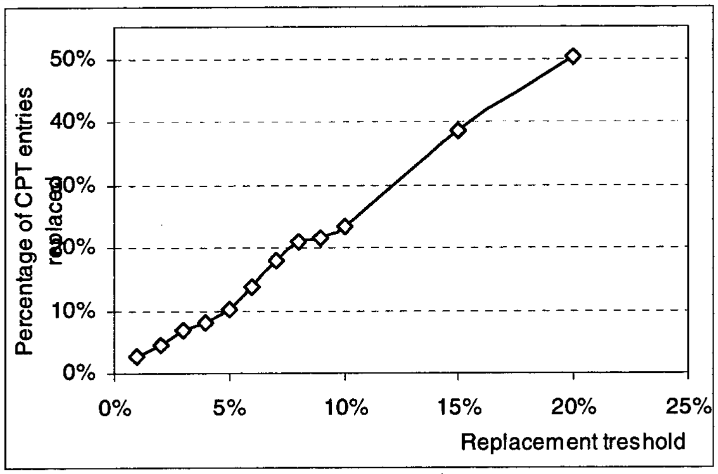

Figure 7: Percentage of conditional probability distribution entries replaced by Noisy-OR distributions as a function of the replacement threshold.

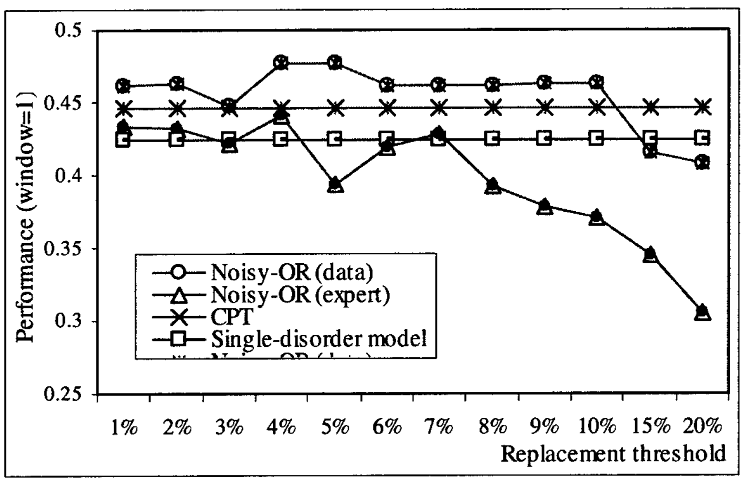

Figure 8: Diagnostic accuracy as a function of the replacement threshold, window=1.

Figure 8 shows the results for the three tested models for the window size of 1. It pictures the diagnostic accuracy of the models as a function of the replacement threshold. In addition we included the results for the single-disorder model. It appears that the highest accuracy was reached by the model whose CPTs were enhanced with the Noisy-OR parameters learned from data. The highest accuracy achieved by the models was 45%, 48%, and 46% for the CPT model, the data Noisy-OR model, and the expert Noisy-OR model respectively.

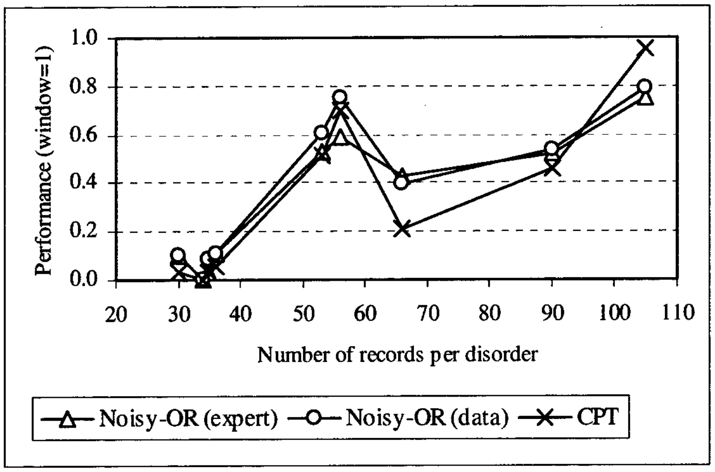

Figure 9: Diagnostic accuracy as a function of the number of disorder cases in the database (class size) for the CPT and two versions of the model with Noisy-OR parameters.

Figure 9 shows the performance within each class for the three models. Again we observed that for almost each of the disorders, the data Noisy-OR model performed better than the other models.

Diagnostic accuracy of the multiple-disorder model enhanced with the Noisy-OR parameters was 6.7% better than the accuracy of the plain multiple-disorder model and 14.3% better than the single-disorder diagnosis model. This increase in accuracy has been obtained with very modest means – in addition to structuring the model so that it is suitable for Noisy-OR nodes, the only knowledge elicited from the expert and entered in the learning process was which interactions can be viewed as approximately Noisy-OR. This knowledge was straightforward to elicit. We have found that whenever combining expert knowledge with data, and whenever working with experts in general, it pays off generously to build models that are causal and reflect reality as much as possible, even if there are no immediate gains in accuracy.

We have also observed that the diagnostic accuracy of the model based on numbers elicited from the expert (as opposed to learned from data) was quite good for diseases with well understood risk factors and symptoms. The accuracy tends to be lower in case of those diseases whose mechanisms are not exactly known, for example *Functional hyperbilirubinemia*, *Reactive hepatitis*, or *PBC*, even if the number of records in the data set was very small.

Other contributions

The Hepar II medical diagnostic system

In order to demonstrate the usefulness of our system in practical setting, we have continued our successful collaboration focusing on building a practical medical system for diagnosis of liver disorders. The resulting system, Hepar II uses our software at its core and consists of a Bayesian network model comprising over 60 variables, such as disorder variables, risk factors for various disorders, symptoms, and test results (Figures 5 and 7 show the model and the model as seen through **GeNIe 2.0** diagnostic interface). The system's parameters are obtained from a database of real patient cases collected at the Institute of Food and Feeding in Warsaw, Poland. The resulting system is applied both as a diagnostic tool in clinical setting and as a tool for training beginning diagnosticians. The results of this work have resulted in several joint publications (listed in the publication list).

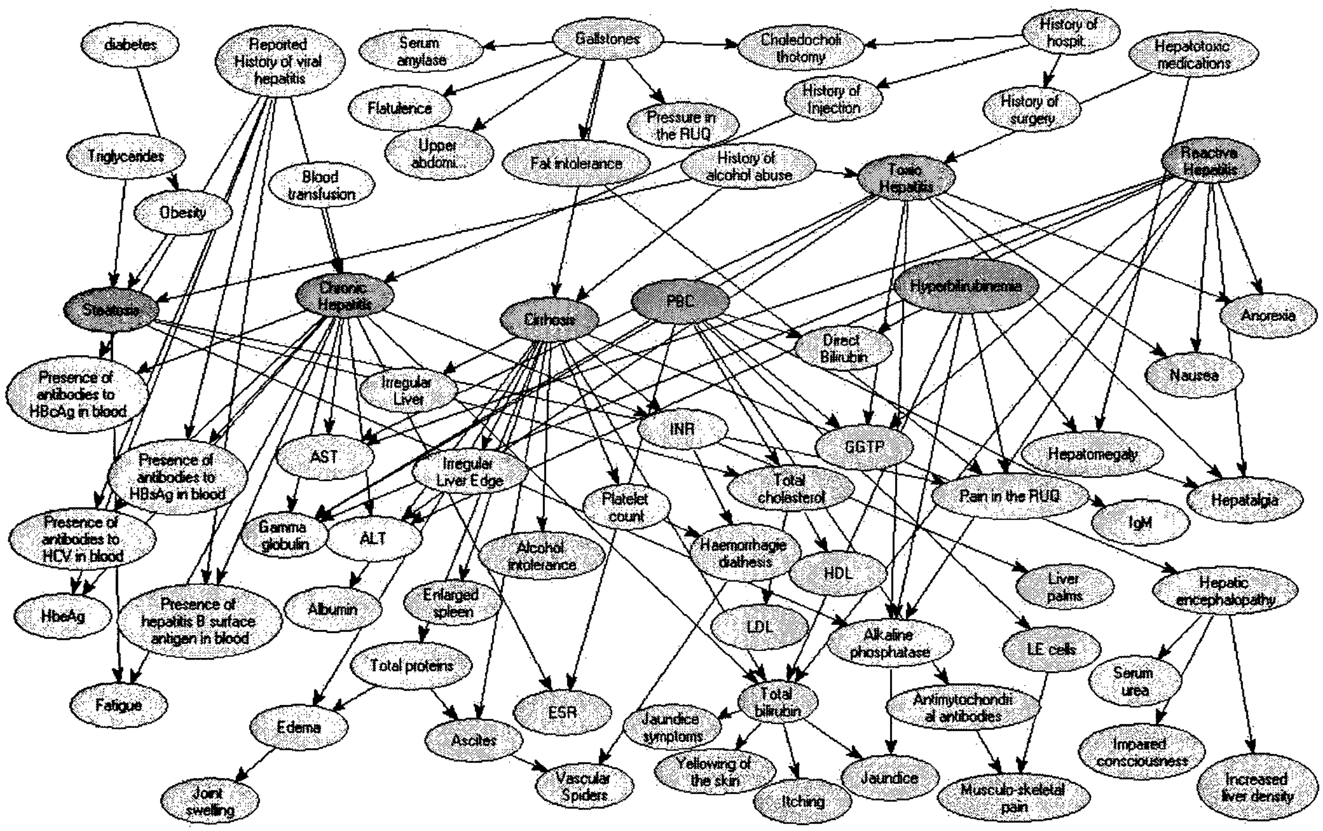

Figure 5: The Hepar II Bayesian network model.

A method for evaluating probability elicitation schemes

As more and more decision-analytic models are being developed to solve real problems in complex domains, extracting knowledge from experts is arising as a major obstacle in model building (Druzdzel & van der Gaag 2000). Quite a few methods have been proposed to elicit subjective probabilities from domain experts. These techniques balance quality of elicitation with the time required to elicit the enormous number of parameters associated with many practical models. Furthermore, the effectiveness of elicitation techniques is likely to task-dependent (Spetzler 1975) or even expert-dependent (Lopez1990), and there is no guidance as to how to select an appropriate method for various domains or experts. Structure elicitation is likewise a tedious problem and formal techniques for this task are even less mature. Systematic evaluation and comparison of different model elicitation methods are thus becoming of growing concern.

In Bayesian probabilistic models, encoded probabilities reflect the degree of personal beliefs of the experts. The sole purpose of probability elicitation is to extract an accurate description of the expert's personal beliefs. In order to judge whether the elicitation procedure has produced an accurate model, therefore, the elicitor must know intimate details about the expert's knowledge. Unfortunately, these details that the elicitor is seeking from the start are hidden from explicit expressions; so it has not been possible to evaluate elicitation schemes directly. Less direct methods are the only possibility.

In a paper published by *IEEE Transactions on Systems Man & Cybernetics* (Wang et al. 2002) we present an objective approach for evaluation of elicitation methods that avoids the assumptions and pitfalls of existing approaches. Our technique is much closer to the ideal "direct" comparison between the elicited network and the expert's beliefs. The main idea is to simulate the training/learning process of an expert by allowing the trainee to interact with a virtual domain. Underlying the domain is a Bayesian network that is used to stochastically update the state of the world in response to the subject's interaction. Then by recording every state of the world that is experienced by the trainee, we can effectively gain direct access to the trainee's knowledge. It is quite an established fact that people are able to learn observed frequencies with an amazing precision if exposed to them for a sufficient time (Estes 1976). Therefore, after training, the trainee obtains some level of knowledge of the virtual world and, consequently, becomes an expert at a certain proficiency level. This knowledge, in the form of a database of records, can be converted to an "expected" model of the expert by applying Bayesian learning algorithms to the database. Finally, this expected expert model can be directly compared to the model elicited from the expert to judge the accuracy of elicitation.

Our approach captures a subject's state of knowledge of the probabilistic events in the toy world. The subject's experience with the toy world, rather than the actual model underlying the world, forms the basis of his or her knowledge. For this reason, the learned model should be the standard used to evaluate the elicitation schemes, rather than the original toy model. This technique allows us to avoid the expensive process of training subjects to fully-proficient expertise. For example, our expert's experience may have led him to explore some states of the world very infrequently. In this case, even if our elicitation procedure is perfect, the elicited probabilities of these states may be significantly different from the underlying model. Using the expert's experience rather than the original model gets around this problem completely because we know precisely how many times our expert has visited any given state of the world.

We use these techniques along with a toy cat-mouse game to evaluate the accuracy of three methods for eliciting discrete probabilities from a fixed structure: (1) direct numerical elicitation, (2) the probability wheel (Spetzler 1975), and (3) the scaled probability bar (Wang & Druzdzel 2000). We use mean squared errors between the learned and the elicited probabilities to evaluate the accuracy of each of the three methods. We show that for our domain the scaled probability bar is the most effective and least time-consuming.

Comparison of rule-based expert systems and systems based on Bayesian networks

Two major classes of expert systems are those based on rules, known as *rule-based expert systems*, and those based on probabilistic graphical models, often referred to as *probabilistic expert systems* or *normative systems.* Rule-based expert systems, originating from the pioneering work of Buchanan and Shortliffe on the Mycin system (Mycin 1984), aim at capturing human expertise in terms of rules of the form **if** *condition* **then** *action*. There is overwhelming psychological evidence (e.g., Newell & Simon 1972) that such rules are capable of modeling the human thought process. A set of rules can capture a human expert's relevant knowledge of a domain and can be subsequently used to reproduce the expert's problem solving in that domain. Probabilistic expert systems originate from research at the intersection of statistics and artificial intelligence. Research on these systems focuses on the concepts of relevance and probabilistic independence and has led to the development of intuitive and efficient graphical tools for knowledge representation. A prominent tool for capturing expert knowledge in this approach are Bayesian networks. Bayesian networks, while also aim at capturing expert knowledge, are based on the mathematical foundations of probability theory. When used in reasoning, they apply mathematical formalism and make no claim about reproducing the expert's thought process.

Several authors have studied theoretical differences between rule-based expert systems and normative systems (e.g., Heckerman 1985, Lucas 2001, van der Gaag 1990), in particular with respect to handling uncertainty. Much less work, however, has been done on studying the implications that choosing one approach over the other has on the knowledge engineering effort and overall system performance. Today, theoretical developments and practical experiences with the probabilistic systems are matching those of rule-based expert systems. Both rule-based and probabilistic systems are in wide use and it is more than ever important to understand the advantages and drawbacks of each of the approaches.

Our work in this area focuses on comparing the two approaches in the context of a challenging practical problem that we worked on independently (Onisko & Druzdzel and our co-author, Peter Lucas), using both rule-based and probabilistic approaches: diagnosis of liver disorders. Expert systems that we have developed are of considerable size and have taken several years to build. Hepatology, the study of diseases of the liver and biliary tract, is an excellent domain for such comparison, as it is complex, contains both rare and frequently occurring disorders, disorders for which both much biomedical knowledge is available and which are described only in terms of symptoms and signs. The results of our comparison were published in the *European Conference on Artificial Intelligence in Medicine* (Onisko et al. 2001).

Quantitative experiments that we performed within the framework of this study have confirmed that a rule-based system can have difficulty with dealing with missing values: around 35% of the IFF patients (the data set used by

Dr. Lucas in building the rule-based version of Hepar) remained unclassified by the rule-based Hepar, while in Hepar-BN only 2% of IFF patients remained unclassified. This behavior might be due to the semantics of negation by absence, and in fact a deliberate design choice in rule-based systems. Refraining from classifying is better than classifying incorrectly, although it will be at the cost of leaving certain cases unclassified. In all cases, the true positive rate for Hepar-BN was higher than for the rule-based Hepar, although sometimes combined with a lower true negative rate.

Both systems were in general more accurate when dealing with their original datasets. The reason is that the systems were using then all available data, not only the common variables. We have noticed some indications of overfitting in case of Hepar-BN, visible especially in those results, where the system was trained and tested on different data sets.

Building the models in each of the two approaches has its advantages and disadvantages. One feature of the rule-based approach that we found particularly useful is that it allows testing models by following the trace of the system's reasoning. A valuable property of Bayesian network-based systems is that models can be trained on existing data sets. Exploiting available statistics and patient data in a Bayesian network is fairly straightforward. Fine-tuning a rule-based system to a given dataset is much more elaborate.

Rule-based systems capture heuristic knowledge from the experts and allow for a direct construction of a classification relation, while probabilistic systems capture causal dependencies, based on knowledge of pathophysiology, and enhance them with statistical relations. Hence, the modeling is more indirect, although in domains where capturing causal knowledge is easy, the resulting diagnostic performance may be good. Rule-based systems may be expected to perform well for problems that cannot be modeled using causality as a guiding principle, or when a problem is too complicated to be modeled as a causal graph.

GeNIe and SMILE©

A major accomplishment of the project is the implementation of the system. Since there is much interest now in Bayesian networks, influence diagrams, and decision-analytic systems, we have put much effort in making the implementation easy to use and robust and decided to share it with the community. We believe that this will bring a high payoff in the long run in terms of practical applications based on our system. We have written a comprehensive on-line help for **GeNIe** (the user interface running on Windows machines), useful for both beginning modelers and students in decision-analytic methods and a documentation for **SMILE**© (**S**tructural **M**odeling, **I**nference, and **L**earning **E**ngine), a portable library of C++ classes for decision-theoretic reasoning, **GeNIe**'s reasoning engine. We have also developed **SmileX**, an Active-X control version of **SMILE**© that allows the program to be used from most Windows applications, including Visual Basic, Excel, and HTML pages. We have made our programs available on the World Wide Web in July 1998 (the address to download the program is: http://www.sis.pitt.edu/~genie). There is a growing number of users of our software. Over 5,000 people from countries all over the world downloaded it since the release date. We have heard very positive feedback from these users.

During the period of the current grant, we have enhanced the module for assistance in model building based on causal mechanisms, the work on which started as a result of the previous AFOSR grant. We have also developed a specialized module for diagnosis, and a module for learning models from data. These modules have not been released on the World Wide Web yet because they are not sufficiently reliable (given the large number of users of our programs, we have adopted high quality standards for releasing our software).

We have a first implementation of the scheme for search for opportunities in causal models, i.e., such a mode of working of a system that allows for automatic and autonomous choice of policy variables.

We have advanced on the second generation of the program, **GeNIe 2.0**, which we plan to release in the last quarter of 2003. **GeNIe 2.0** has a much better user interface, it includes the diagnostic module. Its reasoning engine, **SMILE**©, available also separately, is much faster and it includes our recent additions to the stochastic sampling algorithms. We have replaced **SmileX** with **SMILE.NET**, which offers an even wider applicability, while being upward compatible with the Active-X standard.

Screenshots of **GeNIe 2.0** and its diagnostic interface are presented in Figures 5 and 6.

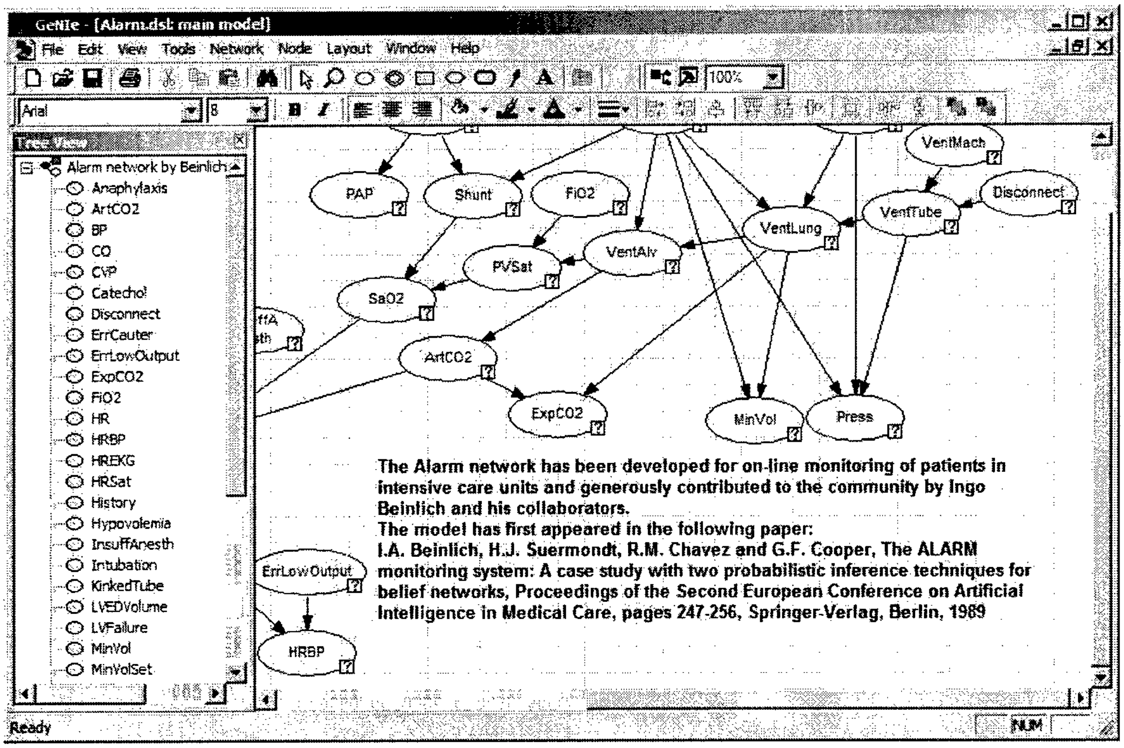

Figure 11: A screen shot of **GeNIe 2.0**.

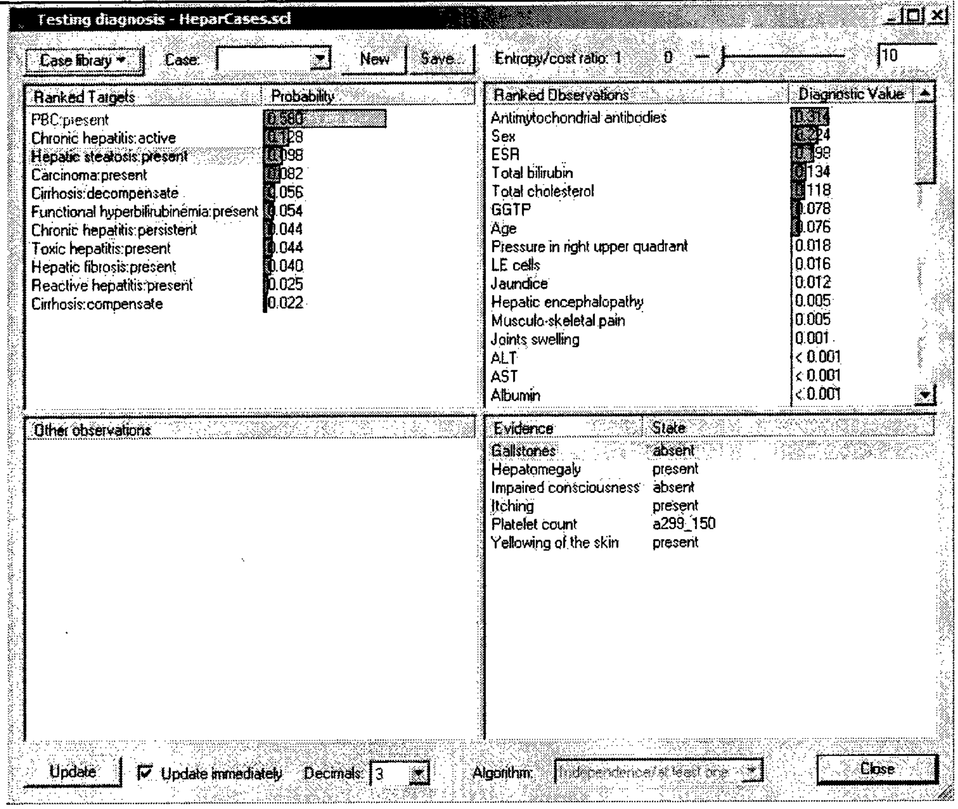

Figure 12: A screen shot of the diagnostic interface of **GeNIe 2.0** (the Hepar II model).

Publications acknowledging support from AFOSR

Journals:

Haiqin Wang, Denver H. Dash and Marek J. Druzdzel. A method for evaluating elicitation schemes for probabilistic models. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, 32(1):38-43, February 2002.

Agnieszka Onisko, Marek J. Druzdzel and Hanna Wasyluk.. Learning Bayesian network parameters from small data sets: Application of Noisy-OR gates. *International Journal of Approximate Reasoning*, 27(2):165-182, 2001.

Marek J. Druzdzel and Hans van Leijen. Causal reversibility in Bayesian networks. *Journal of Experimental and Theoretical Artificial Intelligence*, 13(1):45-62, 2001.

Jian Cheng. Sampling algorithms for estimating the mean of bounded random variables. *Computational Statistics*, 16(1): 1-23, 2001.

Hanna Wasyluk, Agnieszka Onisko and Marek J. Druzdzel. Support of diagnosis of liver disorders based on a causal Bayesian network model. *Medical Science Monitor*, 7(Suppl. 1):327-332, May 2001.

Jian Cheng and Marek J. Druzdzel. BN-AIS: An adaptive importance sampling algorithm for evidential reasoning in large Bayesian networks. *Journal of Artificial Intelligence Research*, 13:155-188, 2000.

Marek J. Druzdzel and Linda C. van der Gaag. Building probabilistic networks: Where do the numbers come from? Guest editors' introduction. *IEEE Transactions on Knowledge and Data Engineering*, 12(4):481-486, 2000.

Major peer reviewed conferences:

Denver H. Dash and Marek J. Druzdzel. Caveats for causal reasoning with equilibrium models. In *Proceedings of the Sixth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU-2001)*, S. Benferhat, P. Besnard (eds.), *Springer Lecture Notes in Computer Science; Lecture Notes in Artificial Intelligence, LNAI 2143*, Berlin Heidelberg: Springer-Verlag, pages 192-203, 2001.

Tsai-Ching Lu and Marek J. Druzdzel. Supporting changes in structure in causal model construction. In *Proceedings of the Sixth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU-2001)*, S. Benferhat, P. Besnard (eds.), *Springer Lecture Notes in Computer Science; Lecture Notes in Artificial Intelligence, LNAI 2143*, Berlin Heidelberg: Springer-Verlag, pages 204-215, 2001.

Agnieszka Onisko, Peter Lucas and Marek J. Druzdzel. Comparison of rule-based and Bayesian network approaches in medical diagnostic systems. *Proceedings of the Eighth Annual Conference on Artificial Intelligence in Medicine (AIME-2001)*, S. Quaglini, P. Barahona, S. Andreassen (eds.) *Artificial Intelligence in Medicine, Lecture Notes in Computer Science Subseries*, Springer Verlag, pages 281-292, 2001.

Haiqin Wang, Denver H. Dash and Marek J. Druzdzel. A method for evaluating elicitation schemes for probabilities. In *Proceedings of the Fourteenth International Florida Artificial Intelligence Research Society Conference (FLAIRS-2001)*, Ingrid Russell & John Kolen (eds), AAA Press, Key West, FL, pages 607-612, 2001.

Jian Cheng and Marek J. Druzdzel. Confidence inference in Bayesian networks. In *Proceedings of the Seventeenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-2001)*, pages 75-82, Morgan Kaufmann Publishers, Inc., San Francisco, CA, 2001.

Jian Cheng and Marek J. Druzdzel. Latin hypercube sampling in Bayesian networks. In *Proceedings of the Uncertain Reasoning in Artificial Intelligence track of the Thirteenth International Florida Artificial Intelligence Research Symposium Conference (FLAIRS-2000)*, pages 287-292, AAAI Press/The MIT Press, Menlo Park, CA, 2000.

Haiqin Wang and Marek J. Druzdzel. User interface tools for navigation in conditional probability tables and elicitation of probabilities in Bayesian networks. In *Proceedings of the Sixteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-2000)*, pages 617-625, Morgan Kaufmann Publishers, Inc., San Francisco, CA, 2000.

Tsai-Ching Lu, Marek J. Druzdzel and Tze-Yun Leong. Causal mechanism-based model construction. In *Proceedings of the Sixteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-2000)*, pages 353-362, Morgan Kaufmann Publishers, Inc., San Francisco, CA, 2000.

Jian Cheng and Marek J. Druzdzel. Computational investigation of low-discrepancy sequences in Bayesian networks. In *Proceedings of the Sixteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-2000)*, pages 72-81, Morgan Kaufmann Publishers, Inc., San Francisco, CA, 2000.

Other peer reviewed conferences, symposia, workshops, and book chapters:

Agnieszka Onisko, Marek J. Druzdzel and Hanna Wasyluk. HEPAR and HEPAR II - computer systems supporting a diagnosis of liver disorders. In *Proceedings of the Twelfth Conference on Biocybernetics and Biomedical Engineering*, Warsaw, Poland, November 28-30, 2001. (Best Young Investigator Paper award for Ms. Onisko).

Agnieszka Onisko, Marek J. Druzdzel and Hanna Wasyluk. An experimental comparison of methods for handling incomplete data in learning parameters of Bayesian networks. In *Intelligent Information Systems 2002: Proceedings of the IIS'2002 Symposium*, M. Klopotek, S.T. Wierzchon, M. Michalewicz (eds.), pages 351-360, *Advances in Soft Computing* Series, Physica-Verlag (A Springer-Verlag Company), Heidelberg, 2002.

F. Javier Diez and Marek J. Druzdzel. Fundamentals of canonical models. In *Proceedings of the IX Conferencia de la Asociacion Espanola para la Inteligencia Artificial (CAEPIA-TTIA 2001)*, pages 1125-1134, Gijon, Spain, 2001.

Agnieszka Onisko, Marek J. Druzdzel and Hanna Wasyluk. Extension of the Hepar II Model to Multiple-Disorder Diagnosis. In *Intelligent Information Systems*, M. Klopotek, M. Michalewicz, S.T. Wierzchon (eds.), pages 303-313, *Advances in Soft Computing* Series, Physica-Verlag (A Springer-Verlag Company), Heidelberg, 2000.

Marek J. Druzdzel and F. Javier Diez. Criteria for combining knowledge from different sources in probabilistic models. In *Working Notes of the workshop on "Fusion of Domain Knowledge with Data for Decision Support," Sixteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-2000)*, pages 23-29, Stanford, CA, 30 June 2000.

Agnieszka Onisko, Marek J. Druzdzel and Hanna Wasyluk. Learning Bayesian network parameters from small data sets: Application of Noisy-OR gates. In *Working Notes of the Workshop on Bayesian and Causal Networks: From Inference to Data Mining, 12th European Conference on Artificial Intelligence (ECAI-2000)*, Berlin, Germany, 22 August 2000.

Marek J. Druzdzel and Roger R. Flynn. Decision Support Systems. In *Encyclopedia of Library and Information Science*, Vol. 67, Suppl. 30, pages 120-133, Allen Kent (ed.), Marcel Dekker, Inc., New York, 2000.

Interactions / Transitions

a. Participation / presentations at meetings, conferences, seminars, etc.

The PI, Dr. Druzdzel, gave a lecture on augmenting human decision making through normative systems at the Air Force Rome Laboratories Decision Science Working Group (DSWG) meeting, George Mason University, October 2002.

The PI, Dr. Druzdzel, gave a lecture on the project at the National University for Distance Education, Madrid, Spain, May 2002.

The PI, Dr. Druzdzel, gave a lecture on the project at the University of Pittsburgh, May 2002.

The PI, Dr. Druzdzel, gave a lecture on the project at the New World Vistas progress meeting in Minnowbrook, NY, November 2001.

Doctoral student Mr. Denver Dash, gave two presentations of joint work with the PI at the Sixth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU-2001), September 2001.

The PI, Dr. Druzdzel, gave a lecture on augmenting human decision making through normative systems at the Biomedical Asia 2001 Conference, September 2001.

Doctoral students: Mr. Jian Cheng, Mr. Tsai-Ching Lu and Ms. Haiqin Wang, gave presentations of joint work with the PI at the 17th Annual Conference on Uncertainty in Artificial Intelligence (UAI-2001), July 2001.

Doctoral student Ms. Haiqin Wang, gave a presentation of joint work with the PI at the Fourteenth International Florida Artificial Intelligence Research Society Conference (FLAIRS-2001), May 2001.

The PI, Dr. Druzdzel, gave a lecture on the project in the Department of Statistics, University of Pitttsburgh, January 2001.

The PI, Dr. Druzdzel, gave a presentation during the annual New World Vista progress meeting at Lockheed Martin Electronics and Missiles Facility, Orlando, FL, September 2000.

Doctoral students: Mr. Jian Cheng, Mr. Tsai-Ching Lu and Ms. Haiqin Wang, gave presentations of joint work with the PI at the 16th Annual Conference on Uncertainty in Artificial Intelligence (UAI-2000), July 2000.

A doctoral student, Mr. Jian Cheng, gave a presentation of joint work with the PI at the 13th International Florida Artificial Intelligence Research Symposium Conference (FLAIRS-2000), May 2000.

The PI, Dr. Druzdzel, gave a lecture on the project at the Honors Day at the University of Pittsburgh, March 2000.

The PI, Dr. Druzdzel, gave a lecture on the qualitative aspects of graphical models at the Naval War College, The Center for naval Warfare Studies, March 2000.

b. Consultative and advisory functions to other laboratories

None so far

c. Applications of our software

Here are some of the applications of our results and our software:

Dr. John Lemmer (**John.Lemmer@rl.af.mil**) at the US Air Force Rome Laboratories will use the results of our research on stochastic sampling algorithms in his work on causal military planning.

Dr. Wojtek Przytula (**wojtek@hrl.com**) at the Hughes Raytheon Laboratories uses **GeNIe** and **SMILE**© in a diagnostic system for General Motors Diesel locomotives. Researchers at Boeing are also applying our software in the work on diagnosis. We have had initial contacts with researchers at Intel interested in applying GeNIe in their work.

GeNIe and **SMILE**© were applied in an intelligent tutoring system for teaching elementary physics, developed at University of Pittsburgh's Learning Research and Development Center (contact person is Prof. Kurt van Lehn, **vanlehn@cs.pitt.edu**). The system was aimed to be applied in teaching Navy cadets. Continuation of this work is at the University of British Columbia, Vancouver, Canada. The point of contact is Dr. Cristina Conati (**conati@cs.ubc.ca**).

Rockwell International Science Center, Palo Alto Laboratory, in collaboration with US Air Force Rome Laboratories applied **GeNIe**, **SMILE**© and **SmileX** to the problem of battle damage assessment. The contact persons there are Mark Peot (**peot@rpal.rockwell.com**) and John F. Lemmer.

The Decision Support Department of the United States Naval War College, Newport, RI, plans to use **GeNIe** and **SMILE**© in supporting a joint US NWC/NATO project on detection of sources of regional instabilities. The point of contact there is Bradd C. Hayes (**hayesb@nwc.navy.mil**).

In collaboration with a group of researchers in Poland, we have applied **GeNIe** and **SMILE**© to the problem of medical diagnosis of liver disorders. This problem is quite similar to the problem of battle damage assessment.

We have two current points of contact who are interested in using the results of our work when our system implements both Bayesian networks and structural equations: Dr. Patrick Love at the ALCOA Technical Center (**Patrick.Love@alcoa.com**), for strategic business planning at Aluminum Company of America, and Mr. Jeffrey Bolton (**jb5c+@andrew.cmu.edu**) and Mr. Kevin Lamb (**kl3g+@andrew.cmu.edu**) at the Carnegie Mellon University's Office of Planning and Budget, for strategic planning of university operations. These contacts will be followed up when **GeNIe** and **SMILE**© implement both equations and Bayesian networks.

Honors / Awards

2003 Robert R. Korfhage award (with Adam Zagorecki), awarded school-wide for the best paper co-authored between a student and a faculty member.

Best Young Investigator Paper award for Ms. Onisko for the paper *HEPAR and HEPAR II - computer systems supporting a diagnosis of liver disorders.* Twelfth Conference on Biocybernetics and Biomedical Engineering, Warsaw, Poland, November 28-30, 2001

2000 Robert R. Korfhage award (with Jian Cheng), awarded school-wide for the best paper co-authored between a student and a faculty member.