A New Uncertainty Measure for Belief Networks with Applications to Optimal Evidential Inferencing

Jiming Liu, *Senior Member*, *IEEE*, David A. Maluf, and Michel C. Desmarais

Abstract—This paper is concerned with the problem of measuring the uncertainty in a broad class of belief networks, as encountered in evidential reasoning applications. In our discussion, we give an explicit account of the networks concerned, and coin them the Dempster-Shafer (D-S) belief networks. We examine the essence and the requirement of such an uncertainty measure based on well-defined discrete event dynamical systems concepts. Furthermore, we extend the notion of entropy for the D-S belief networks in order to obtain an improved optimal dynamical observer. The significance and generality of the proposed dynamical observer of measuring uncertainty for the D-S belief networks lie in that it can serve as a performance estimator as well as a feedback for improving both the efficiency and the quality of the D-S belief network-based evidential inferencing. We demonstrate, with Monte Carlo simulation, the implementation and the effectiveness of the proposed dynamical observer in solving the problem of evidential inferencing with optimal evidence node selection.

Index Terms—Belief networks, uncertainty modeling and management, discrete event dynamical systems, optimal evidential inferencing, controller, observer, entropy, user profile assessment.

1 Introduction

Belief networks are computational structures that are composed of clusters of nodes representing assertions or query/evidence variables interrelated by links signifying the independence relationships among the nodes [5], [17]. Some belief networks decompose the joint-probability distribution of real-world probabilistic knowledge with conditionals [16], while others focus on the belief-function measures of the nodes as supported by surrounding evidences. Falling into the latter category is the Dempster-Shafer (D-S) belief networks in which the probabilities of evidential support are explicitly represented. The D-S belief networks are frequently used as a knowledge representation scheme to handle situations where causal or diagnostic relationships are to be captured and reasoned about. Some examples of the D-S belief network application are diagnosis and multisensor integration [11], [12].

There exists a complete formalism of evidential reasoning for computing and propagating evidential support (whether confirming or disconfirming) throughout the network; the formalism is also known as the D-S theory of evidence. In the D-S theory of evidence, the deductions take place within logical constraints, and the belief information (i.e., the weight) is treated as metaconstraints that modify these logical constraints. The D-S evidential representation and inferencing scheme may be viewed as a simplified (but sound) theoretical deviation from the Bayesian theory [7], [18].

- *J. Lui is with the Department of Computer Science, Hong Kong Baptist University, Kowloon Tong, Hong Kong. E-Mail: jiming@comp.hkbu.edu.hk.*
- *D.A. Maluf is with RIACS/NASA, Ames Research Center, Mail Stop 269-2, Moffett Field, CA 94035. E-Mail: maluf@ptolemy.arc.nasa.gov.*
- *M.C. Desmarais is with Public Technology Multimedia, 1001 Sherbrooke E., Suite 700, Montreal, Quebec, Canada H2L 1L3. E-mail: desmarais@ptm.ca.*

Manuscript received 8 Aug. 1996; revised 1 Nov. 1999; accepted 19 Jan. 2000. For information on obtaining reprints of this article, please send e-mail to: tkde@computer.org, and reference IEEECS Log Number 104380.

1.1 Problem Statement

With the D-S approach to knowledge representation and reasoning, if some information on certain nodes or variables is obtained, its support over other unobserved nodes can readily be computed based on the structure of the network as well as the previous state of the network (i.e., probabilistic or belief measures of nodes). This paper is concerned with two important problems in the D-S belief network applications; namely,

1. Is it *necessary* to devise an optimal policy of selecting evidence nodes so that stronger belief functions can be achieved?
2. Furthermore, if necessary, then how can we computationally determine such an optimal policy?

1.2 Organization of the Paper

In this paper, we will, first of all, address the above two issues from a point of view of *discrete event dynamical systems* [4]. This will, in turn, enable us to better understand the nature of the problems at hand and to qualitatively study the requirements of our solutions. Following those requirements, we will then define a new entropy-oriented uncertainty measure for the D-S belief networks essential for solving the optimal evidence collection problem. In order to quantitatively demonstrate the necessity as well as the effectiveness of the proposed uncertainty measure application, we will conduct some Monte-Carlo simulation studies in which the performances of entropy-based evidential inferences are contrasted with those of random evidential inferences.

2 The D-S Belief Networks as Discrete Event Dynamical Systems

2.1 The D-S Belief Networks

In the D-S belief networks, the set of all possible outcomes of a node is called the *frame of discernment*, denoted by Θ. For instance, with respect to node x_i in the D-S network, its possible outcomes may be expressed as follows:

$$\Theta_{x_i} = \{a_1, a_2, \ldots, a_k, \ldots, a_n\}. \tag{1}$$

Here, the term "discern" entails that it is possible to differentiate the correct variable state from all the other possible states with respect to a specific node. One correct state requires that the set be exhaustive and that the subset be disjoint [7], [18].

The D-S theory of evidence accepts partial evidential specifications in the form of logical sentences and allows a *basic probability assignment* (bpa) to the subsets of a conclusion, as denoted by $m(*)$. Unlike the Bayesian approach, the D-S model does not allow a subset be proven by any rule set unless it appears in a consequent of at least one rule. Suppose that our frame of discernment for node x_i, Θ_{x_i}, is $\{a, \neg a\}$, where each element denotes a hypothesis induced from some observations. Thus, the confirmation $m(a)$, disconfirmation $m(\neg a)$, and the frame of discernment mass $m(\Theta_{x_i})$ will form the basic probability assignment for node x_i. Formally, a *bpa* of node x_i is a function:

$$m : 2^{\Theta}_{x_i} \rightarrow [0, 1],$$

where

$$m(\emptyset) = 0,$$

$$\sum_{c_j \subseteq \Theta_{x_i}} m(c_j) = 1.0. \tag{2}$$

The D-S theory distinguishes the state of *ignorance* about a variable from the relative weight afforded on the variable versus its negation. The ignorance is signified by the probability mass assigned to Θ_{x_i}, as denoted by $m(\Theta_{x_i})$.

Based on the notion of probability mass, a *belief function*, $Bel(c_j)$, over Θ_{x_i} can be defined as the total belief committed to all subsets of c_j, i.e.,

$$Bel(c_j) = \sum_{b \subseteq c_j} m(b). \tag{3}$$

The D-S theory of evidence offers a rigorous means for revising beliefs in the presence of new evidential support from distinct sources (i.e., accumulated evidence), known as *Dempster's rule of combination*. This rule states that two *bpa*s corresponding to two independent sources of evidence may be combined to yield a new *bpa*, as follows:

$$m(c) = \alpha \cdot \sum_{c' \cap c'' = c} m(c') \cdot m(c''), \tag{4}$$

where α is a normalization factor that ensures (4) be satisfied. Specifically,

$$\alpha = \frac{1}{1 - \sum_{c' \cap c'' = \emptyset} m(c') \cdot m(c'')}. \tag{5}$$

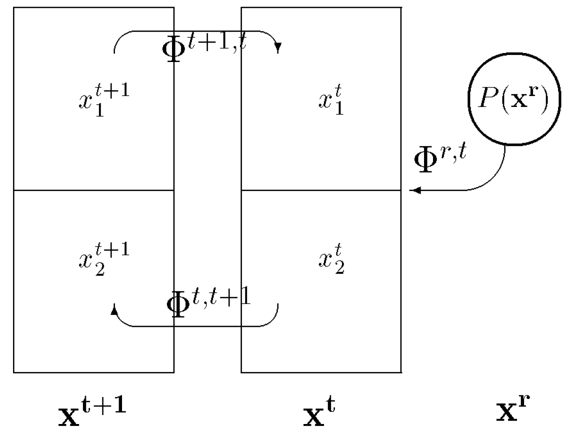

Fig. 1. A typical multivariate DEDS model of mixed belief functions. $\mathbf{x^t}$ and $\mathbf{x^{t+1}}$ are two state vectors, whereas $\mathbf{x^r}$ denotes a state vector whose basic probability assignment (*bpa*) function has been updated.

2.2 The Discrete Event Dynamical Systems (DEDS) Model

Having described the basic constructs of a D-S belief network, we can now take a close look at how such a network fits into the conventional discrete event dynamical systems (DEDS) model [4]. This treatment is essential for our later discussions on the uncertainty measure as used in optimal evidential reasoning.

In order to apply the DEDS model, we first represent the belief functions of the interconnected network nodes as a single vector, called *state vector* $\mathbf{x}$. The element of this vector is called a *vector node*, x_i. Subsequently, given the collection of derived belief functions, we can formulate a *discrete event dynamical systems* model for the D-S belief network in terms of the following quintuple:

$$\mathcal{S} = (X, U, Y, \Phi, \eta), \tag{6}$$

where X, U, and Y correspond to a finite set of state vectors, a finite set of evidence inputs, and a finite set of outputs of the network, respectively. Φ denotes a transition function of the state vectors, and η denotes an output function defined as $X \rightarrow Y$. The DEDS model is illustrated in Fig. 1.

Given the system of some finite number of vector nodes, we can make systems state transitions based on the input sequences. In other words, we can obtain new hypotheses (belief functions) based on the belief network by taking into account the input evidence that supports one or more vector nodes.

2.3 Transition Function Φ

Generally speaking, transition function Φ can be defined as follows:

$$\Phi : X \times U \rightarrow X. \tag{7}$$

In what follows, we formulate the exact transition function in our DEDS model (i.e., $\Phi^{t,t+1}$ in Fig. 1) based on our earlier description of the D-S belief networks.

Without loss of generality, we assume that there are only two possible outcomes for each vector node (i.e., a network node variable) x_i in both $\mathbf{x^t}$ and $\mathbf{x^{t+1}}$. Hence, our frame of discernment can be written as: $\Theta_{x_i} = \{a, \neg a\}$. Suppose that in state $\mathbf{x^t}$, node x_i^t receives certain updated evidential supports;

namely, k supports confirm a for the value of vector node x_i^t, as denoted by $\{m_{C1}, m_{C2}, \ldots, m_{Ck}\}^t$, and l supports disconfirm a, as denoted by $\{m_{D1}, m_{D2}, \ldots, m_{Dl}\}^t$.

First, we organize these supports by combining all *bpa*s for each of the possible outcomes into two composite evidential supports, one confirming a with a *bpa* equal to m_C^t and the other disconfirming a with m_D^t. By the definition of *bpa*, we know that m_C^t and m_D^t can both be derived by repeatedly applying (4) and (5). Hence, we have

$$m_C^t = 1 - \prod_{1 \le i \le k} (1 - m_{Ci}^t), \tag{8}$$

$$m_D^t = 1 - \prod_{1 \le j \le l} (1 - m_{Dj}^t). \tag{9}$$

As a result, we can derive a pair of new *bpa*s, m_C^{t+1}, and m_D^{t+1} for $\mathbf{x^{t+1}}$, representing the effect of propagating supports from the two composite evidential sources,

$$m_C^{t+1} = \alpha \cdot m_C^t \cdot (1 - m_D^t), \tag{10}$$

$$m_D^{t+1} = \alpha \cdot m_D^t \cdot (1 - m_C^t), \tag{11}$$

where

$$\alpha = \frac{1}{1 - m_C^t \cdot m_D^t}.$$

From the preceding discussions, we can readily work out a transition function of the discrete event dynamical system for automatically deriving the system state. The transition function, $\Phi^{t,t+1}$, that gives new belief functions is expressed as follows:

$$\mathbf{x^{t+1}} = \Phi^{t,t+1}(\mathbf{x^t}, \{m_C^t, m_D^t\}) = \{Bel_i^{t+1}(a), Bel_i^{t+1}(\neg a)\}, \tag{12}$$

where

$$Bel_i^{t+1}(a) = m_C^{t+1} \tag{13}$$

$$Bel_i^{t+1}(\neg a) = m_D^{t+1}. \tag{14}$$

It should be pointed out that due to the existence of node connectivity, the transition from state $\mathbf{x^t}$ to state $\mathbf{x^{t+1}}$ may involve a chain of vector node updating, in which one updated vector node further propagates evidential supports to other adjacent nodes. In the Appendix, we have provided a complete algorithm that governs the *repeated updating* of vector nodes during a state transition.

3 Optimal "Control" Policy for Choosing Informative Evidence Nodes

From the transition function of DEDS, we know that the next vector state in which a network will be is entirely dependent on the present vector state and the evidential input at one or more vector nodes. Our earlier empirical investigations of the D-S belief networks have also demonstrated that an arbitrary sequence of vector node observation $\{u^1, u^2, \ldots, u^t \ldots\}$ can dramatically change the belief functions, and as a result, reduce the overall uncertainty of the system to a varying degree [8].

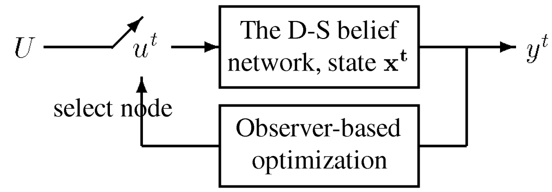

Fig. 2. A good measure of uncertainty associated with $\mathbf{x^t}$ can result in some feedback for determining input sequence u^{t+1}, which can, in turn, dictate the way in which a belief network is updated.

In the D-S belief network applications, various evidence node selection policies may be applied to determine which node is to be observed next. One approach is to *randomly* choose an evidence node from a complete node set, U. Another approach is to apply some optimization techniques and choose the most informative node. This approach requires a well-defined optimality function to evaluate the performance of evidential inferences based on the belief network.

In what follows, we are concerned with the construction of such an optimality function. More specifically, we are interested in the problem of how to unambiguously measure the degree of uncertainty reduction in the system so that sufficient feedback information can be obtained for choosing the input sequence $\{u^t\}$. If we have such an uncertainty measure, we can then address the problem of optimal evidential reasoning in which the purposefully selected node observation will rapidly bring the network to an equilibrium state with a minimum uncertainty.

As shown in Fig. 2, this is essentially an *optimal policy generation* problem from the point of view of the optimal DEDS control. An optimally selected input sequence will yield an optimal performance of the system (with respect to some specific optimality definition) [10]. The point of interest here is how to devise a robust *dynamical observer* for the DEDS that can unambiguously measure the performance of the system. Specifically, the observer should estimate the uncertainties associated with the systems vector nodes over a sequence of finite evidential observations, where normally each vector node consists of a probability space of more than one independent variable.

3.1 Observability of the D-S Belief Networks

In this section, we formulate the dynamical observer that incorporates an uncertainty measure for the system $\mathcal{S}$. The dynamical observer to be constructed must be capable of measuring the belief functions of each state vector in the system $\mathcal{S}$, hence satisfying the following observability axiom [1], [3]:

Definition of Observability. *A dynamical system that is described by (6) is said to be observable if given an arbitrary input u^t; there exists for every vector in X an output sequence described by output function η and, also,*

$$\forall\, u^t \subset U; \;\; X \times \{u^t, \Phi\}; \;\; \exists\, \eta : X \rightarrow y^t, \;\; y^t \neq y^{t-1}. \tag{15}$$

In order to satisfy the above observability condition, the construction of our dynamical observer for the DEDS, $\mathcal{S}$, has to take into account certain functional aspects that can readily identify which $u^t \in U$ has been taken such

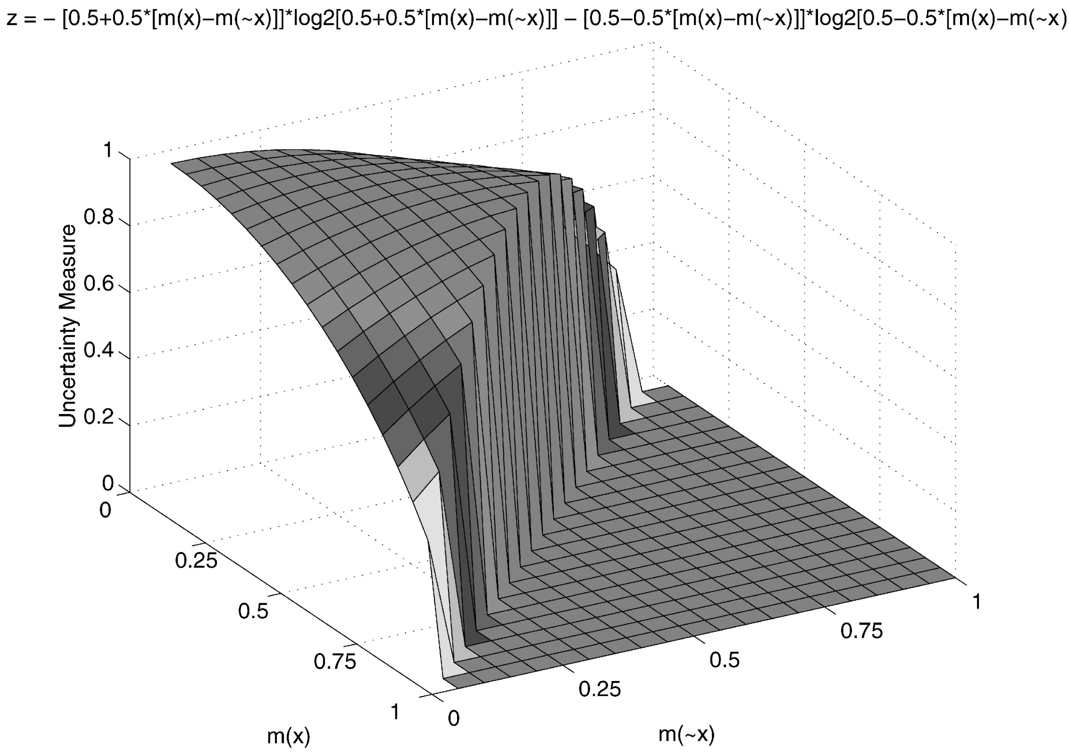

Fig. 3. Entropy computed for the D-S belief networks based on a Bayesian projection space. Notice the linearly projected entropy function for the D-S belief functions at values $\{0.5, 0.5\}$ and $\{0, 0\}$.

that $X \times \{u^t, \Phi\} \rightarrow X$. Here, we choose η as our output function that yields Y, an uncertainty measure, from state vectors in X.

In general, when the vector nodes face some alterations resulting from an arbitrary input u^t, the system $\mathcal{S}$ undergoes a state transition, as mentioned in the previous section. This state transition also produces a corresponding output Y using an output function η defined over all vector nodes. Formally, we define η as an N:1 function that maps the complete set of vector nodes into a measure of uncertainty, which is denoted as follows:

$$y^t = \eta(Bel_1^t, Bel_2^t, \ldots, Bel_i^t, \ldots, Bel_N^t), \tag{16}$$

where Bel_i^t represents the belief function of the $i\text{th}$ vector node in $\mathbf{x^t}$.

3.2 Entropy-Driven Optimal Evidence Node Selection

Whether for simple systems, or for systems that have a tendency to grow in complexity and size such as belief networks, a standard method of measuring the systems uncertainty is essential. In the classic information theory [19], formalisms were defined with an attempt to quantitatively measure an information process or mechanism. In the context of the D-S belief networks, we let the dynamical system $\mathcal{S}$ be measured with a similar means to acquire a significant and informative measurable index.

3.2.1 The Classic Entropy-Based Uncertainty Measure

Suppose that a system of some finite number of vector nodes is given, about which new hypotheses are to be analyzed resulting from the dynamics of some input to one or more vector nodes. And, the notion of entropy will be employed to evaluate the uncertainty change in the system as a function of the induced hypotheses. In what follows, we can observe that unlike the entropy computation for the Bayesian models, the entropy computation for the D-S belief networks cannot be carried out in a straightforward fashion using probability distributions.

For the sake of illustration, let us consider a D-S belief network which has a two-element frame of discernment for its state vector nodes, as denoted by $\Theta_{x_i} = \{a, \neg a\}$. First, we take a linear projection that transforms the D-S belief functions, Bel_i, into a probability value, for instance,

$$P_i(a) = \frac{1}{2} + \frac{1}{2}[Bel_i(a) - Bel_i(\neg a)]. \tag{17}$$

Further, we apply the conventional entropy measure and write:

$$H_{ds}(\mathbf{x}) = -\sum_i \sum_k P_i(a_k) \log_2 P_i(a_k). \tag{18}$$

Without lose of generality, Fig. 3 plots the corresponding entropy function of a *single* vector node system, $\mathcal{S}$, having linearly projected its D-S belief functions into their Bayesian counterparts. As shown in the figure, values $\{0.5, 0.5\}$ and $\{0, 0\}$ have the same entropy, although the two values carry different amounts of information concerning $\{a, \neg a\}$.

According to the discussion in Section 3.1, we know that the preceding linear projection of entropy has led to an unobservable state in the system $\mathcal{S}$. That is, there exists an input u that has no consequent change in the output measure y. This contradicts the observability.

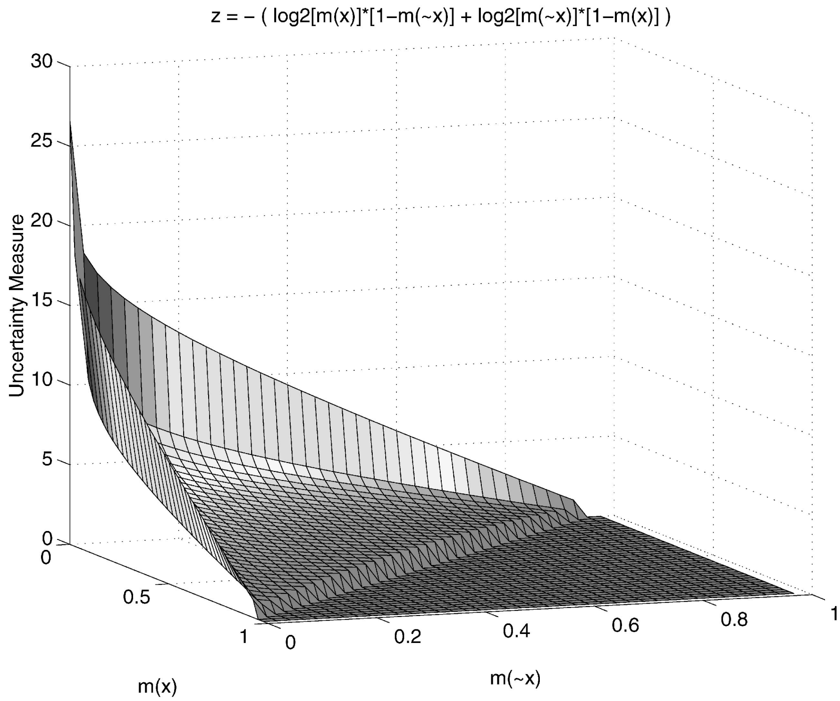

Fig. 4. Uncertainty measure computed for the D-S belief functions. Notice the function at values $\{0,0\}, \{1,0\}$, and $\{0,1\}$.

3.2.2 An Improved Entropy-Based Uncertainty Measure

Let us now modify the above classic entropy measure in order to cater to the D-S belief network semantics. According to the D-S network formalism, the belief functions associated with each variable signify the probabilities of evidential supports, rather than the probabilities of the variable itself. This is also to say, the uncertainty-oriented interpretation of entropy within the present context of the D-S networks has to be modified. The proper interpretation would be: *the extended entropy measure indicates the degree of the uncertainty associated with the evidential supports weighted by the degree to which the evidential supports do not disconfirm (although not necessarily always confirm) the variable.*

Since the degree to which the evidential supports do not disconfirm a node variable (which is sometimes referred to as a *plausibility function*) can be formally expressed as follows:

$$\begin{aligned} Pl_i(a) &= 1 - Bel_i(\neg a) \\ &= \sum_{b \subseteq \Theta_{x_i}} m(b) - \sum_{b \subseteq (\Theta_{x_i} - a)} m(b) = \sum_{a \cap b \neq \emptyset} m(b), \end{aligned} \tag{19}$$

we can write a generalization of the Shannon entropy definition as the uncertainty measure for state vector $\mathbf{x}$ as follows:

$$H_{ds}(\mathbf{x}) = -\sum_{x_i \in \mathbf{x}} \sum_{y \subseteq \Theta_{x_i}} Pl_i(y) \log_2[Bel_i(y)] \tag{20}$$

and based on (3) and (19), we can rewrite:

$$H_{ds}(\mathbf{x}) = -\sum_{x_i \in \mathbf{x}} \sum_{y \subseteq \Theta_{x_i}} \left[\sum_{y \cap c \neq \emptyset} m(c) \right] \log_2 \left[\sum_{b \subseteq y} m(b) \right], \tag{21}$$

which is a monotonically decreasing function.

The fact that (21) is monotonic and decreasing proves that this entropy measure is *observable*. For the sake of illustration, we have given in Fig. 4 a plot of the proposed entropy function for the D-S belief function in our above single-node, two-element frame-of-discernment example. Note that the entropy increases for values tending toward $\{0,0\}$, and reaches a minimum entropy at $\{1,0\}$ and $\{0,1\}$.

Klir [13] and Klir and Yuan [14] have provided an "entropy-like" measure called *dissonance*. If we follow thier original definition, we can further derive a detailed expression of dissonance for the D-S belief-network system as follows:

$$D_{ds}(\mathbf{x}) = -\sum_{x_i \in \mathbf{x}} \sum_{y \subseteq \Theta_{x_i}} \left[\sum_{b \subseteq y} m(b) \right] \log_2 \left[\sum_{y \cap c \neq - \emptyset} m(c) \right]. \tag{22}$$

Equation (22) appears to be a symmetrical function to (21) above. However, the two definitions have quite different semantic meanings: The former is concerned with the uncertainty associated with (or doubt about) the evidences for *all subsets* of Θ_{x_i} that have a nonnull intersection with x_i—the doubt about the plausibility, whereas the latter is concerned only with the uncertainty associated with (or doubt about) the total evidences committed *particularly* to x_i. In other words, (21) measures the doubt about our belief in x_i, which is exactly what we are interested in.

In addition, as shown in Fig. 5, a plot of dissonance for our single-node belief system, the measurements at $\{0,0\}$, $\{1,0\}$, and $\{0,1\}$ are equal under the dissonance computation, further reflecting the fact that Klir's dissonance does not express the uncertainty associated with our belief about the variable node. Thus, (22) cannot serve as our optimality measure.

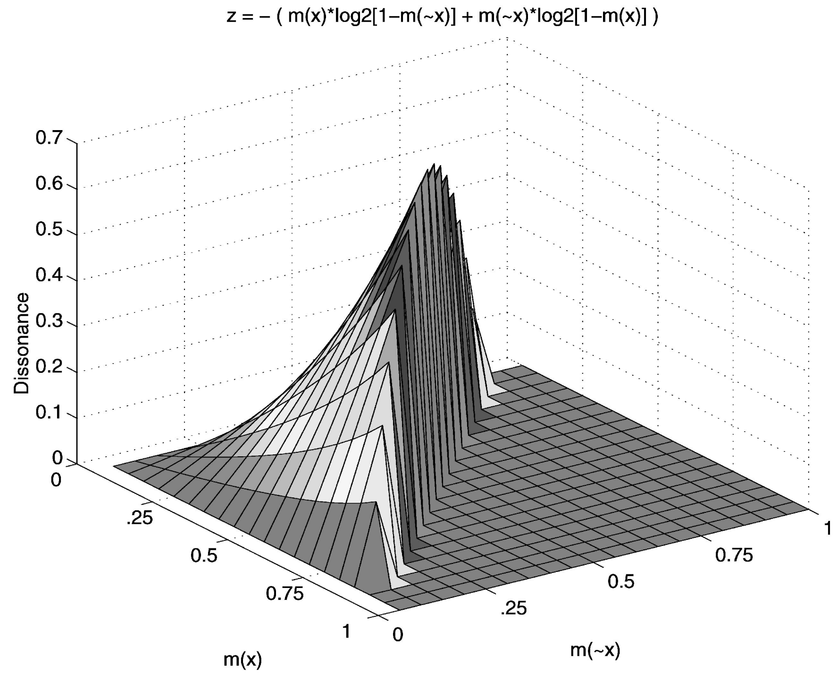

Fig. 5. Dissonance computed for the D-S belief functions. Notice the function at values $\{0,0\}$, $\{1,0\}$, and $\{0,1\}$.

3.2.3 Selecting Optimal Evidence Nodes Based on the Improved Uncertainty Measure

In what follows, we revisit the original problem of generating an optimal evidence node policy for the optimal DEDS "control," based on our proposed uncertainty measure of the D-S belief-network system. Specifically, here our problem is viewed as an optimization problem that is to minimize the doubt about the evidential supports (i.e., to maximize the belief yield in the DEDS) with the least number of evidence nodes. To do so, we incrementally choose a sequence of evidence nodes that have the highest chance of reducing the entropy in the system.

Our "controller" utilizes the uncertainty measure as defined above to predict the expected belief yield of each individual node over all the possible outcomes. The node that has the maximum expected belief yield is selected as the potentially most *informative* evidence node, which is to be observed next.

Based on the definition of our extended entropy computation for the D-S belief system, we can write the optimality criterion that x_i is most likely to reduce entropy in our two-element frame-of-discernment case, as follows:

$$H_{ds}(\mathbf{x} \mid x_i \text{ is observed}) = [Bel_i(a) \cdot H_{ds}(\mathbf{x} \mid x_i = a)] + [Bel_i(\neg a) \cdot H_{ds}(\mathbf{x} \mid x_i = \neg a)], \tag{23}$$

where $H_{ds}(\mathbf{x} \mid x_i = a)$ is the total entropy computed from the belief systems state if the evidence of a is observed and $H_{ds}(\mathbf{x} \mid x_i = \neg a)$ is the total entropy if the evidence of $\neg a$ is observed. Given the expected entropy value for every x_i, the problem of determining the most informative evidence node is hence reduced to that of finding the node with the lowest H_{ds} value.

4 An Optimal Evidential Inferencing Example

In this section, we examine the effectiveness of our proposed uncertainty measure in generating a sequence of optimal evidence nodes for reducing the uncertainty of a D-S belief network. Our examination will be based on a Monte-Carlo simulation study. More specifically, our study utilizes a small set of empirically obtained data samples to algorithmically induce a D-S belief-network system. Thereafter, based on such an induced network, we carry out optimal evidential inferencing, by selecting and observing a sequence of optimal evidence nodes from the network. The observations of the evidence nodes are simulated using the event variables from the empirical data samples, and correspondingly, the network-based inference results about the rest of unobserved nodes are validated using the values from the same data samples. From such an empirical validation, the amount of correct inferences (i.e., reduction in uncertainty), as resulted from the input evidence, can readily be calculated.

4.1 The D-S Belief-Network Induction Based on Empirical Data Samples

The empirical data used for building the D-S belief network consists of 26 complete samples, which were compiled based on the results of a questionnaire administered to a group of subjects. Each sample contains 191 variables (or nodes), covering the subjects' knowledge of using a commercial word processor. Each data sample can be viewed as a certain state vector from the point of view of the preceding DEDS model.

We then input those 26 data samples into an induction algorithm to construct a D-S belief-network system. This belief-network induction algorithm along with the study on the validity of induced networks has been reported elsewhere. Interested readers are referred to [15] for details.

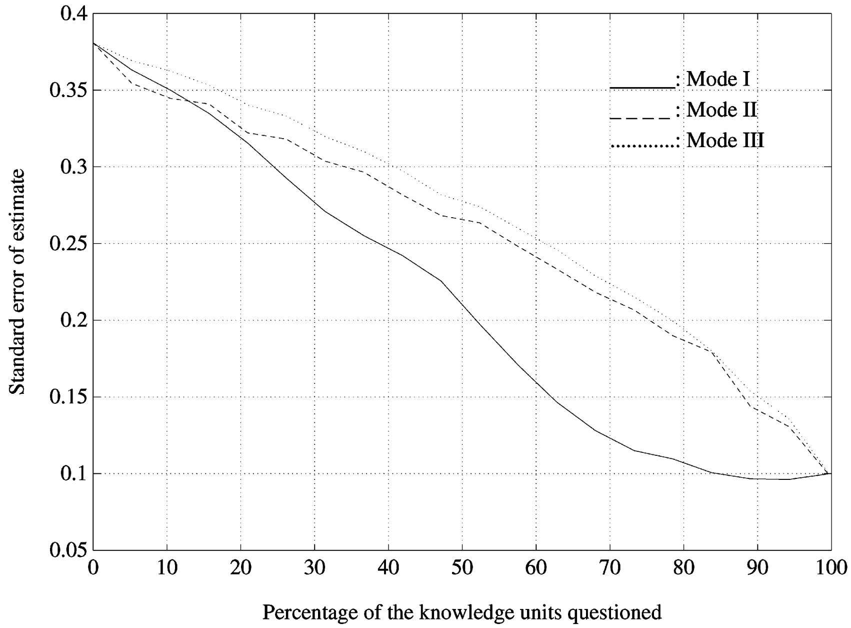

Fig. 6. *Individual-node-assessment* performance in three different modes of observation, measured in the standard error of estimate over 191 KUs and averaged for 10 subjects. Note: The solid line, the dashed line, and the dotted line correspond to Modes I, II, and III, respectively (refer to the text for details).

The set of data samples as used in this study induces $2,368$ statistically significant links among the 191 variable nodes. The specific meanings of the derived D-S belief-network system in this case can be stated as follows:

- A node represets a fine-grain *knowledge unit* (or KU), which may be a basic concept or elementary skill. We assume that each KU_i corresponds to a proposition, namely, "the individual knows KU_i."
- The weight for a KU_i indicates our belief that the KU_i is mastered.
- A link represents a gradation constraint, which indicates that if a certain concept or skill has been acquired then it can, to some extent, be inferred that another concept or skill is also acquired.

Therefore, the knowledge of an individual subject can be described using some subset (i.e., an *overlay model*) of all KUs.

One of the main applications of this knowledge assessment technique is to dynamically build fine-grain user profiles. Here, by *fine-grain* modeling we mean the characterization of an individual's knowledge, with respect to a set of knowledge units (KUs) consisting of either basic concepts or elementary skills.

4.2 Experimental Results on Optimal Evidential Inferencing

With the induced belief network, we conducted a series of user profile assessment simulations. We used a set of 10 testing data samples (*other than those for the network induction*) to simulate the observations of some network evidence nodes and, at the same time, let our reasoning program estimate the belief values for other unobserved nodes. Prior to the inferencing, all the nodes of the knowledge structure were assigned the same initial beliefs.

As a result of the evidential inferencing, a node with $Bel(a)$ above 0.85 is considered TRUE and $Bel(\neg a)$ above 0.85 is considered FALSE. This translates to the diagnoses of *known* and *unknown* knowledge units (KUs), respectively. The system does not produce any predictions if the weights associated with a node variable do not meet such thresholds. Such a bidirectional thresholding is typical of applications in which a partial diagnosis is acceptable [9].

After each observation-and-updating session, we examined the performance of the evidential inferencing by measuring the results with *the standard error of estimate*, defined as follows:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{10}\sum_{j=1}^{191}\left(x_{\mathrm{emp}_{ij}} - x_{\mathrm{est}_{ij}}\right)^2}{N_s \times n_{max}}}, \tag{24}$$

where n_{max} is the number of knowledge units (191). N_s is the number of subjects used for the test (10). $x_{\mathrm{emp}_{ij}}$ is equal to 1 if the *actual value in the subject* i's *empirical sample* corresponding to KU_j is *known*, and 0 otherwise. $x_{\mathrm{est}_{ij}}$ is the estimated belief.

The results of the systems performance in three different simulation modes are given in Fig. 6. They correspond, respectively, to the average standard error scores over 10 subjects. The three simulation modes are:

1. *Inferences based on the entropy-driven evidence node selection.* When a node, KU_i, is chosen based on entropy minimization, the belief for KU_i is assigned 0.9 for a successful occurrence of KU_i in the testing sample, and 0.1 otherwise. Inference propagation is performed around the observed evidence node, KU_i, according to the connectivity of the belief network.
2. *Inferences based on random sampling of the evidence nodes*. Same as (I) but nodes are chosen at random.
3. *No inference condition*. Same as (II) but no inference propagation is performed.

Note that we have assigned weights 0.9 and 0.1 for successful and unsuccessful occurrences, respectively, to

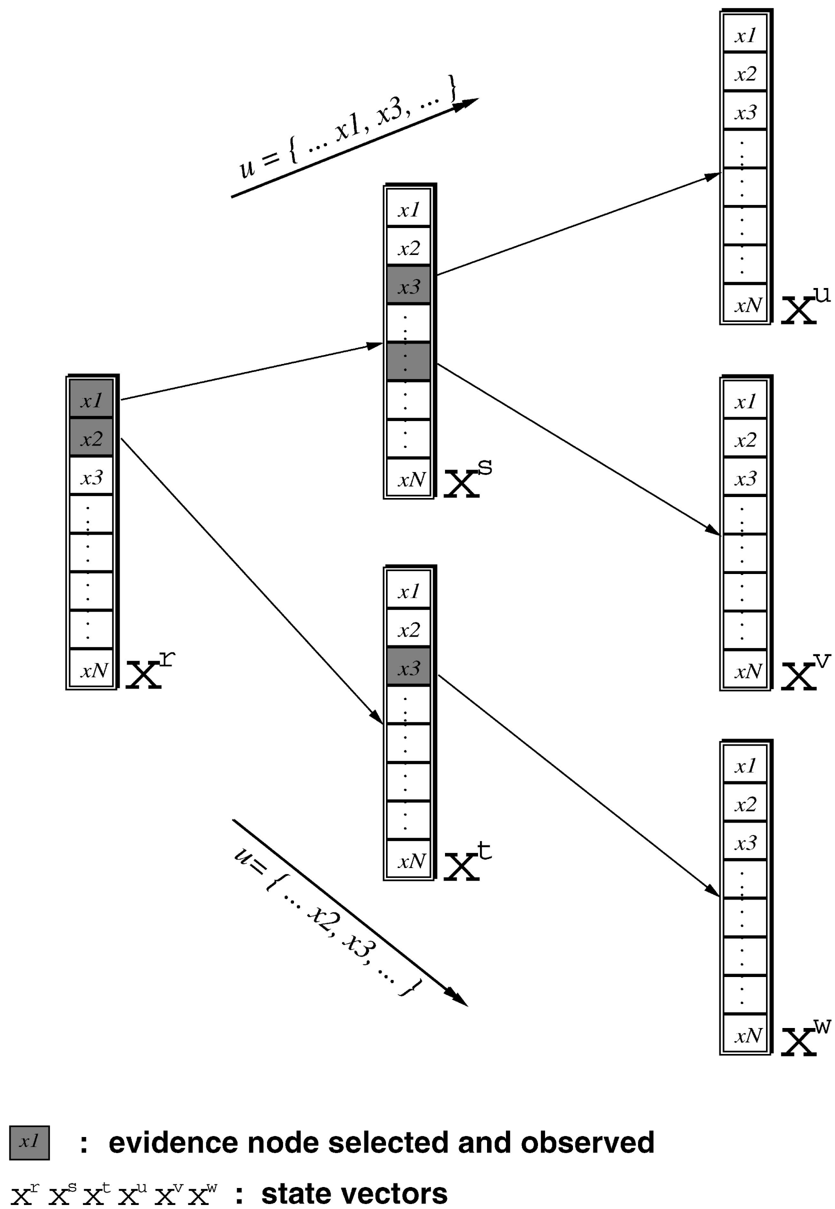

Fig. 7. Selection of one evidence node, e.g., x_1 or x_2, will affect future admissible selections of other nodes.

reflect the residual uncertainties associated with such a process (e.g., a person may produce good answers by chance or bad errors by mistake). As a result, the expected score at 100 percent observation is below the perfect score, since the nodes' weights are contrasted against 1.0 and 0, instead of 0.9 and 0.1.

4.3 Discussions

The results from Fig. 6 clearly indicate that the entropy-driven approach (Mode I) is more efficient in reducing the standard error of estimate. For instance, a close to perfect knowledge assessment was obtained after sampling 80 percent of a subject's knowledge units (i.e., 80 percent of evidence nodes). Furthermore, sampling 60 percent of the knowledge units would reduce the standard error score of estimate to about half of the error score in Mode III. Thus, the method was successful in reducing the number of questions (or evidence nodes) to be asked in order to assess a subject's profile.

The explicit algorithm used in the above entropy-driven evidence node selection process is essentially a technique of *hill climbing*. This technique benefits the incremental evaluation of the optimality criterion. As can be noted from the DEDS model derived in this study, an optimal evidence node selection will affect the future admissible selections. In other words, the selection of one node will change the topology of the state transition diagram; an example of such a case has been provided in Fig. 7. Therefore, we cannot directly apply the conventional *dynamic programming* algorithm to find the optimal policy of a DEDS that minimizes the "total cost" by proceeding from the terminal node *backwards* [2], [6].

One of the obvious limitations of the hill climbing technique is that the search may be trapped in local optima, hence affecting the final (global) search results (in our case, a sequence of evidence nodes). From Fig. 6, we notice that Mode I, although, in general, is consistently more efficient than Mode II in reducing the errors, gave poorer performance in evidential inferencing when the amount of observation was *less than 12 percent*. This is due to the fact that the entropy-driven evidential inferencing was trapped in a local minimum when fewer than 12 percent of evidence nodes were observed.

In order to both overcome the computational complexity and improve the overall search performance, we may further consider other search techniques well-known for handling NP-hard optimization problems, such as *simulated annealing*. With those techniques, the optimal sequence search starts with an initial sequence and makes randomized changes on the previous sequence in such a way that the sequence is biased towards a global optimal. The advantage of such approaches lies in that the search will not blindly search all local optima. An explicit treatment of those optimization techniques is beyond the scope of the present paper. Our future work will examine this issue in details.

5 Conclusion

In this paper, the problem of measuring the uncertainty associated with a Dempster-Shafer (D-S) belief network in order to determine a sequence of evidence nodes during reasoning has been addressed. This problem was interpreted with the existing concepts of optimal system control so that the nature of the problem as well as the requirement for such an uncertainty measure can better be examined. This was done by viewing the D-S belief network as a discrete event dynamical system (DEDS) and, subsequently, studying the possible formulation of the uncertainty measure for the DEDS. As it was shown, the classic entropy measure for the D-S belief system could lead to unobservable vector states. As an improved dynamical observer especially catering to the semantics of the D-S system, a new computation scheme was given. The necessity and effectiveness of the proposed uncertainty measure in the optimal evidential inferencing was shown in Monte Carlo simulation experiments that drew upon a hill climbing search technique.

Appendix

The Belief Revision Algorithm

Belief revision starts from each observed node, x_i, and propagates the belief to its neighboring nodes based on the inference rules of *modus ponens* and *modus tollens*. The algorithm can be stated as follows:

The Belief Revision Algorithm. {*Initially, all the observed nodes (i.e., the truth values of some nodes) are stored in a linked list,* $\texttt{link}_{observ}$. `insert` *and* `get_next_node` *are standard queuing functions.* `update_belief` *computes belief functions.* $\triangle Bel(\bullet)$ *denotes the net change in beliefs after updating*.}

Begin

for each observed node, x_i in $\texttt{link}_{observ}$, **do**
    `insert`(x_i `queue`);
    while `queue` is not empty, **do**
        starting `node` $\leftarrow$ `get_next_node`(`queue`);
        if starting node = `TRUE`, **then**
            for each rule: starting `node` $\Rightarrow x_j$;
                starting `node` $\Rightarrow \neg x_j$;
                $x_j \Rightarrow \neg$ starting `node`; $\neg x_j \Rightarrow \neg$ starting `node` **do**
                $Bel(x_j) \leftarrow$ `update_belief`(starting `node`, x_j);
                if $\triangle Bel(x_j)$ is greater than a threshold, θ, **then** `insert`(x_j, `queue`);

        else

            for each rule: $x_k \Rightarrow$ starting `node`;
                $\neg x_k \Rightarrow$ starting `node`;
                $\neg$ starting `node` $\Rightarrow x_k$;
                $\neg$ starting `node` $\Rightarrow \neg x_k$, **do**
                $Bel(x_k) \leftarrow$ `update_belief`(starting `node`, x_k);
                if $\triangle Bel(x_k)$ is greater than a threshold, θ, **then**
                `insert`(x_k, `queue`);

End

It should be pointed out that the D-S belief network may not always be a singly-connected graph. In order to handle the problem of multiple transitivity in the network, our present implementation of the belief updating algorithm allows the traversal from one node to another to be performed only once by randomly choosing one of the possible traversal paths. Thus, the path traversal in the multiple transitivity case may be regarded as being nondeterministic.

References

[1] K.J. Astrom and B. Wittenmark, *Computer-Controlled Systems: Theory and Design.* Englewood Cliffs, N.J.: Prentice Hall, 1990.

[2] R. Bellman, *Dynamic Programming.* Princeton Univ. Press, 1957.

[3] P.E. Caines, R. Greiner, and S. Wang, "Classical and Logic-Based Dynamic Observers for Finite Automata," *IMA J. Math. Control & Information,* 1991.

[4] C.G. Cassandras, *Discrete Event Systems: Modeling and Performance Analysis.* Homewood, Ill.: Aksen Assoc. Inc. Publishers and IRWIN, 1993.

[5] E. Charniak, "Bayesian Networks without Tears," *AI Magazine,* pp. 50-63, 1991.

[6] T.L. Dean and M.P. Wellman, *Planning and Control,* San Mateo, Calif.: Morgan Kaufmann, 1991.

[7] A.P. Dempster, "A Generalization of Bayesian Inference," *J. Royal Statistical Soc.,* vol. 30, pp. 205-247, 1968.

[8] M.C. Desmarais, L. Giroux, S. Larochelle, and S. Leclerc, "Assessing the Structure of Knowledge in a Procedural Domain," *Proc. Cognitive Science Soc.,* pp. 475-481, 1988.

[9] M.C. Desmarais, "Architecture et Fondements Empiriques d'un Système d'Aide Assistée par Ordinateur pour l'Édition de Texte," PhD thesis, Université de Montéal, Département de Psychologie, 1990.

[10] V.N. Fomin, *Discrete Linear Control Systems.* Dordrecht, The Netherlands: Kluwer Academic, 1991.

[11] T.D. Garvey, J.D. Lowrance, and M.A. Fischler, "An Inference Technique for Integrating Knowledge from Disparate Sources," *Proc. Int'l Joint Conf. Artifical Intelligence '81,* pp. 319-325, 1981.

[12] J. Gordon and E.H. Shortliffe, "The Dempster-Shafer Theory of Evidence," *Rule-Based Expert Systems,* B.G. Buchanan and E.H. Shortliffe, eds., Reading, Mass.: Addison-Wesley, 1984.

[13] G.J. Klir, "Generalized Information Theory," *Fuzzy Sets and Systems,* vol. 40, pp. 127-142, 1991.

[14] G.J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications.* Upper Saddle River, N.J.: Prentice Hall, 1995.

[15] J. Liu and M.C. Desmarais, "A Method of Learning Implication Networks from Empirical Data: Algorithm and Monte-Carlo Simulation-Based Validation," *IEEE Trans. Knowledge and Data Eng.,* vol. 9, no. 6, pp. 990-1004, Nov./Dec. 1997.

[16] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Mateo, Calif.: Morgan Kaufmann, 1988.
[17] S.J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach.* Englewood Cliffs, N.J.: Prentice Hall, 1995.
[18] G. Shafer, *A Mathematical Theory of Evidence.* Princeton, N.J.: Princeton Univ. Press, 1976.
[19] C.E. Shannon, "A Mathematical Theory of Communication," *Bell Systems Technical J.,* vol. 27, pp. 379-423, pp. 623-656, 1948.

Jiming Liu received the MS degree in educational technology from Concordia University and the ME and PhD degrees in electrical engineering from McGill University in Montreal Canada. Currently, he is an associate professor of computing studies at Hong Kong Baptist University. Prior to joining the university in 1993, he worked for several years as a software engineer, research associate, and senior research agent at R&D firms and government labs in Canada (e.g., Computer Research Institute of Montreal, Canadian Workplace Automation Research Centre/Government of Canada, KENTEK (Knowledge Engineering Technology Inc.), and Virtual Prototypes Inc.). In 1999, while on a six month sabbatical as a visiting scholar in the Computer Science Department at Stanford University, Dr. Liu was associated with the AI and Robotics Laboratory and taught advanced graduate classes on topics related to machine learning, neural robotics, and evolutionary robotics. He inititated and served as the program chair for the first Asian-Pacific Conference on Intelligent Agent Technology (ITAT '99). He is also the author of two forthcoming books, entitled *Multiagent Robotic Systems* (CRC Press) and *Autonomous Agents and Multiagent Systems: An Introduction* (World Scientific Publishing). He is also the editor of *Intelligent Agent Technology: Systems, Methodologies, and Tools* (World Scientific Publishing). His areas of expertise are artifical intelligence, autonomous agents and multiagent systems, learning self-adaptation and artificial life in software systems, robotics, intelligent agent-mediated electronic commerce (IamEC), agent-human interaction, virtual environments and animation, and applied dynamics of computation and complex systems. He is a senior member of the IEEE and the ACM.

David A. Maluf received the BE degree in electrical engineering from the American University of Beirut in 1987, the ME and PhD degrees in electrical engineering from McGill University in 1991 and 1995, respectively, as well as his premedical studies (concurrent to his PhD effort) from the faculty of science at McGill University. He worked for several years as an advisor, research associated, and senior researcher at both Institutes Center de Rescherche Informatique de Montreal and the government in Canada, prior to joinging Stanford Universtiy in 1996, where was involved in information technology and the database group. He was an adjunct professor in the Fauclty of Engeering at McGill University in Canada. He has been involved in research in academics and industry relating to the areas of: automation, intelligent information integration, databases and knowledge-bases, data mining and knowledge discovery, digital libraries, image understanding, information processing, systems & control, and foundations of artificial intelligence, and have yielded several millions of research dollars. Dr. Maluf has been a science reviewer for numerous organizations and for the US National Science Foundation. His focus on the data-intensive science problems in bioinformatics resulted in the first solution allowing high-throughput integration of genomic data for accelerated analysis. His technique has been integrated with Incyte Pharmaceutical (Palo Alto, California) proprietary data and has generated a new product and several millions of dollars of revenues. Currently, he is with the Computational Science Division at NASA Ames Research.

Michel C. Desmarais received the PhD degree in psychology from the University of Montreal. He is currently a team leader of the Artificial Intelligence Group at Public Technologies Multimedia, Montreal, Quebec, Canada. He has directed a number of R&D projects, from an expert system in meteorology to a computer coach for text-editing. His previous work experience include the Computer Research Institute of Montreal, Xerox PARC, Stanford University, and HP Labs. His areas of interest span from cognitive science and artificial intelligence to user interfaces and performance support systems.