Journal of International Technology and Information Management

Volume 18
Issue 3 *Double Issue 3/4*

Article 9

2009

Inference-Guiding on Bayesian Knowledge-Based Systems

Jinchang Wang
Richard Stockton College of New Jersey

Inference-Guiding on Bayesian Knowledge-Based Systems

Jinchang Wang
Richard Stockton College of New Jersey
USA

ABSTRACT

Bayesian network is a robust structure for representing knowledge containing uncertainties in a knowledge-based system. In applications of expert systems and knowledge-based systems, it often happens that initial data are not sufficient to derive a conclusion of high enough certainty. Inference-guiding is in that case to identify the missing information, pursue its value, and lead inference to a conclusion. This paper presents and characterizes a criterion for effectively selecting key missing information, and thereby develops a "smart" inference approach with the inference-guiding function based on the newly developed criterion for uncertain inference in a Bayesian knowledge-based system.

INTRODUCTION

A knowledge-based system is a computer system that automates intelligent processes. Inference on a knowledge-based system is a field of artificial intelligence (AI). AI has found numerous applications in business management. An AI approach can assist international firms in screening markets (Fish, 2006), can help plan and control the performance of a just-in-time manufacturing system (Wray, Markham, & Mathieu, 2003), and can serve for monitoring and detecting financial frauds and abuses (Hall & McPeak, 2003). Knowledge-based systems are necessary components of decision support systems (DSS) that are now widely used in business to enhance managers' decision making process (McManus & Snyder, 2003; Hung, Tang, & Shu, 2008).

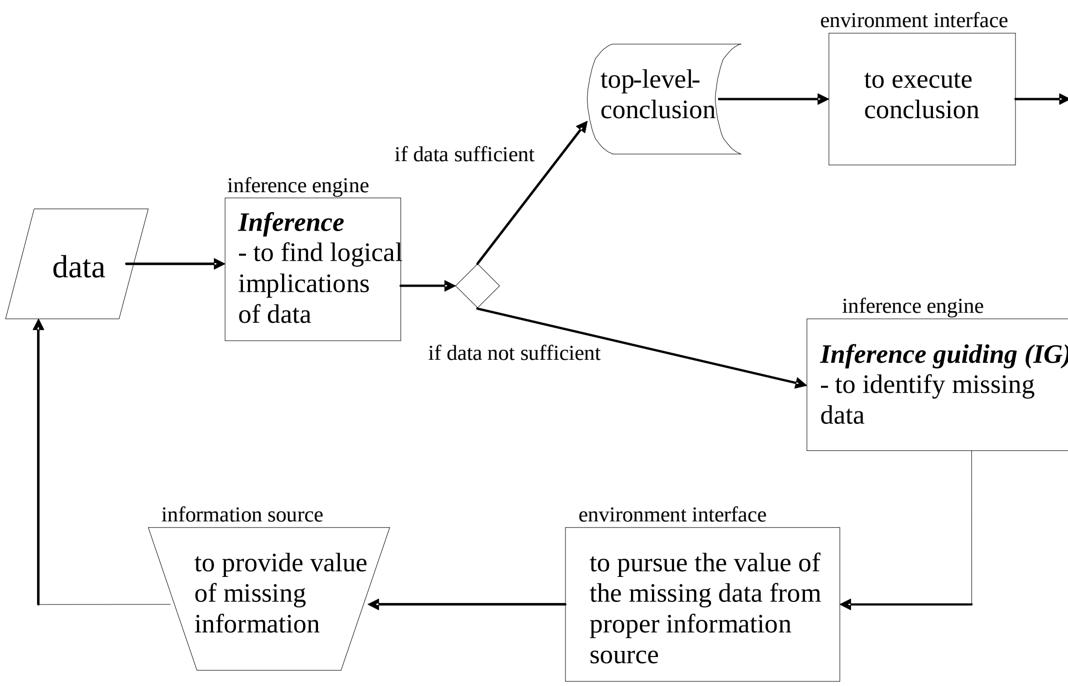
A knowledge-based system is composed of a knowledge base, an inference engine, and an environment interface. A *knowledge base* organizes and stores knowledge. An *inference engine*, which is composed of software for inference and reasoning, generates logical implications of given data based on the knowledge in the knowledge base. An *environment interface* consists of software and hardware, such as sensors, monitors, keyboards, speakers, and control mechanisms, for interacting with the environment. Knowledge-based systems have been successfully used in numerous applications such as locating fuel deposits, designing complex computer systems, analyzing electronic circuits, diagnosing diseases, assisting driving vehicles, and doing the jobs that are dangerous to humans.

An inference engine has two functions for practical applications: logical inference and inference-guiding. *Logical inference*, or simply *inference* or *reasoning*, generates facts that are logically implied by the knowledge and given facts. Inference is aimed at the top-level-conclusion (TLC) that is a conclusion in the desired domain. When data are not sufficient, logical inference cannot reach a TLC, and *inference-guiding (IG)*, the second function of an inference engine, is needed to identify missing data and lead inference to a top-level-conclusion. In a disease diagnosing system, for example, TLCs are possible ailments patients may have, and logical inference is the process of deducting the disease a patient has based on the patient's symptoms and doctor's knowledge. A patent's superficial symptoms and initial information are usually not sufficient for the doctor to determine the illness, and he/she has to ask the patient to provide more information and arrange some medical examinations. In a crime investigation example, TLCs are the suspects who may have committed a particular crime. Based on the information and evidences initially available, a detector deducts the logical implications. If the detector is not able to reach any conclusion due to the missed links, he/she has to identify the missing information and go ahead to find evidences so as to solve the criminal puzzle.

A good IG approach must be able to identify *relevant* missing information to pursue. Gynecological questions, for example, are not relevant for a male patient. A good IG approach must also be able to select *key* missing information so that confirmation of those pieces of missing information would quickly lead to a TLC. In other words, a good IG approach should lead to a TLC after pursuing minimum amount of missing information. A bad IG approach, on the other hand, would delay the inference process by asking irrelevant, silly, and off-the-point questions.

Inference and inference-guiding are conducted alternately in the process of reaching a TLC as shown in Figure 1. The process starts with the function of inference based on the currently known facts. If they are not sufficient to logically reach a TLC in the target domain, the inference-guiding function is kicked on to select a missing fact to confirm. The system's environment interface then pursues the value of the selected missing fact through a proper information source (a sensor, or a user, for example). The newly confirmed fact is added to the known data, and the process is repeated again. The process stops when a TLC is reached.

Figure 1: Inference and inference guiding in a knowledge-based system.



-

Knowledge can be imprecise and incomplete. Data can be inexact and fragmentary. Knowledge-based systems in many cases must draw conclusions and act under uncertainty that comes from *un-sureness of belief* and *un-sureness of truth*. Un-sureness of belief is often associated with a piece of knowledge. For instance, we do not believe with 100% confidence that "if someone breaks in, then the alarm will sound" because the alarm would not work if it is malfunctioned or the power is out. Un-sureness of belief can be associated with data. For example, we may reasonably doubt at the collected data that "57 people saw a UFO at Hutchinson, Kansas, at 1:35am 11/27/2001" since it could be simply an airplane instead of a UFO. Un-sureness of truth occurs where the definition of a truth value is ambiguous. For example, truth of "the patient has a headache" or "it is cloudy today" may not be definite since "headache" and "cloudy" are not defined clearly.

A couple of models have been developed to represent uncertainties in knowledge and data. Non-probabilistic approaches that were proposed in 1970s and 1980s include MYCIN's certainty factor method (Buchanan & Shortliffe, 1984; Shortliffe & Buchanan, 1975), the confidence factor union method (Hayes-Roth, Waterman, & Lenat, 1983), and the fuzzy set method (Zadeh, 1965; Kickert, 1978). They have been used in many expert system shells (Magill & Leech, 1991). Probability theory was considered in 1960s for dealing with uncertainty (Duda et al., 1976). Expert Edge (Human Edge, 1985), an expert system shell, for example, applies the Bayes' Theorem. The probabilistic models fell out of favor in the early 1970s, until in the recent decade when researchers realized the shortcomings of the non-probabilistic approaches and showed that probability systems had theoretical strength in many applications where the knowledge base is large and there exist complex interactions between pieces of knowledge (Russell & Norvig, 2003).

Probability has an intrinsic link with the uncertainty. Un-sureness of belief is subjective probability by its nature, whose calculations and propagations follow the probability theory. Assigning a probability of 0 to a given assertion corresponds to an unequivocal belief that the assertion is false, while assigning a probability of 1 corresponds to an unequivocal belief that the assertion is true. Probabilities between 0 and 1 correspond to intermediate degrees of belief in the truth of the assertion. Although un-sureness of truth is not probability by its nature, it can be viewed as probability and be taken into the uncertainty calculations and propagations by following the probability theory.

Bayesian network is a robust structure for representing knowledge with uncertainties, which has captured attentions of researchers (Ben-Gal, 2007; Pearl, 1988; Langseth & Portinale, 2007). Its structure makes it convenient to depict a complex knowledge base (Chavira & Darwiche, 2007). There are two exclusive advantages of Bayesian network. One is that Bayesian network is inherently capable of allowing uncertainties (Kanal & Lemmer, 1986). Another is that logical inference in Bayesian network becomes calculations of probabilities, which are supported by the probability theory (Wang, 2005b).

Uncertain inference has been a active research field in artificial intelligence for almost four decades. The approaches of uncertain inference available in literatures can be grouped into three categories. One category is for the rule-base, which was built on the success of logical rule-based systems by adding a sort of "fudge factor" to each rule to accommodate uncertainty. Certainty factor is an instance. The methods of this category were developed in the mid-1970s and formed the basis for a large number of expert systems in medicine and other areas. The second category of uncertainty reasoning methods is based on fuzzy logic that was developed in 1980s and specially good for calculating un-sureness of truth (Zadeh, 1965). The third category is based on the probability theory. By such an approach, the process of reasoning becomes a process of calculating probabilities. Researchers have developed a couple of algorithms to have more efficient inference in Bayesian networks more efficient, such as variational methods (Bishop, Spiegelhalter, & Winn, 2003), variable elimination method and likelihood weighting methods (Liu & Soetjipto, 2004), and those surveyed by Russell and Norvig in their book in their book (Russell & Norvig, 2003).

Compared to logical inference, inference-guiding has received less attention from researchers, though there have been some published literature on the topic. For exact inference that does not consider uncertainties, EXPERT uses pre-listed orderings of rules and questions for missing data selection (Hayes-Roth et al., 1983). KAS, a shell over PROSPECT, uses both forward and backward chaining, together with a scoring function, for selecting missing data (Duda, Hart, & Nilsson, 1979). Mellish gave a procedure, using a so-called "Alpha-beta pruning technique", to eliminate irrelevant questions for acyclic inference nets (Mellish, 1985). Wang and Vande Vate (1990) proved that the inference-guiding problem is NP-hard even in a Horn clause system. Based on a dynamic representation of Horn systems (Jerolow & Wang, 1989), they proposed an efficient heuristic strategy, called minimum usage set strategy (MUS). The experiments carried out by Wang and Triantaphyllou (1994) showed that MUS strategy performed well. The inference guiding strategy developed in (Wang, 2005a) was able to select the key pieces of missing information in such a way that the total cost of acquiring additional information for reaching a conclusion is minimized. For uncertain inference, the method presented in (Wang, 1994) is a heuristic for inference-guiding in a rule-base with certainty factors. The research on inference-guiding in the probabilistic Bayesian network started in (Wang, 2005b).

This paper presents the criterion for selecting key missing information for inference-guiding in a Bayesian network, and thereby develops an inference approach that aims at reaching a TLC after pursuing fewest data. Section 2 introduces the fundamental concepts of uncertain inference and the Bayesian network. Section 3 explores the criterion for effectively guiding inference in Bayesian network. An IG index is developed and characterized as an effective criterion. Section 4 incorporates the IG index into an inference approach that would lead inference quickly to a TLC. An example is given throughout the paper to illustrate Bayesian network, the IG index, and the new inference approach.

FUNDAMENTALS

An *assertion* is a statement can be either 'true' or 'false'. An assertion is *observable* if its value can be obtained directly from an information source in the environment (a user or a sensor, for example). An assertion is called *unconfirmed observable assertion (UOA)* if it is observable and its value is not yet known. UOAs represent the missing data. Selecting a missing data is to select a UOA. If assertion A_k's value is logically implied by assertions A_{j_1}, A_{j_2}, ..., then A_{j_1}, A_{j_2}, ..., are called the *premises* of A_k, and A_k is called the *inferred assertion*. As defined in Section 1, an assertion is a *top-level-conclusion (TLC)* if it is in the domain that inference is aimed at. For example, possible diseases are TLCs in a disease diagnosing system; possible faults in an engine are TLCs in a diagnosing system for aircraft maintenance; and required operating adjustments are TLCs of a real-time piloting control system.

If the currently known facts are not sufficient to reach a TLC, inference-guiding is needed to identify some UOAs, whose values are then pursued. The newly obtained data are added to the set of facts for purpose of further inference. If there are *n* UOAs, then there are *n* options in selecting the next UOA to pursue. Total number of UOAs that are selected to pursue before a TLC is reached can be large or small, depending on what UOAs are selected for pursuing and their sequence. Of cause, we want to reach a TLC after asking for as few UOAs as possible.

Uncertainties may occur in knowledge and facts. Inference on knowledge and facts with uncertainty is called *uncertain inference* or *uncertain reasoning*. A TLC is reached when its certainty to be true is high enough. For example, the doctor would not inform a patient about the disease he might have unless, in the doctor's mind, the probability that the patient has that disease is significantly high. Usually, people use a *threshold* value to represent the 'high enough' probability. Thus, a TLC is *reached* if its probability to be true is at or above the preset threshold.

The *Bayesian network* is composed of nodes and arcs. A node represents an assertion. There is a directed arc from node A_j to node A_k if A_j is a premise of A_k. In other words, there is an arc from A_j to A_k if A_j has a direct influence on A_k. Each node is labeled with the probabilities of possible values. If assertion A_k is observable, then the *prior probability* of 'A_k to be true', denoted as $P(A_k=T)$, is given as the label. If A_k is an inferred assertion with premises A_i and A_j, for example, then node A_k is labeled with full conditional probabilities, $\{P(A_k=T \mid A_i=T, A_j=T), P(A_k=T \mid A_i=F, A_j=T), P(A_k=T \mid A_i=T, A_j=F)$, and $P(A_k=T \mid A_i=F, A_j=F)\}$. The complementary conditional probabilities can be derived from the above probabilities, such as $P(A_k=F \mid A_i=T, A_j=T) = 1- P(A_k=T \mid A_i=T, A_j=T)$. The uncertainties are thus embedded in these probabilities. Given the full conditional probabilities, we can calculate any joint probability by applying the probability theory. The process of logical inference in a Bayesian network is therefore a process of probability calculations.

Example 1. A Bayesian network and calculations for inference.

Figure 2 shows a Bayesian network with five assertions. Suppose that A_1, A_2 and A_3 are UOAs, and A_4 and A_5 are inferred assertions and TLCs. The arcs show that A_1 and A_2 have direct influences on A_4, and A_2 and A_3 have direct influences on A_5. In terms of "if...then..." statements, that Bayesian network represent two rules: "If A_1, A_2 then A_4", and "If A_2 and A_3, then A_5". There are no arcs among A_1, A_2 and A_3, which means they are independent between each other. The prior probabilities are given to the observable assertions A_1, A_2 and A_3, while the full conditional probabilities are given to the inferred assertions A_4 and A_5 in form of tables.

With the given prior and conditional probabilities, we can calculate all joint probabilities. For instance, the joint probability:

$$P(A_4=T, A_1=T, A_2=F)$$
$$= P(A_4=T|A_1=T,A_2=F)*P(A_1=T,A_2=F)$$

$= P(A_4=T| A_1=T,A_2=F)*P(A_1=T)*(A_2=F)$ (since A_1 and A_2 independent)
$= 0.75*0.8*0.4 = 0.24.$

Similarly,

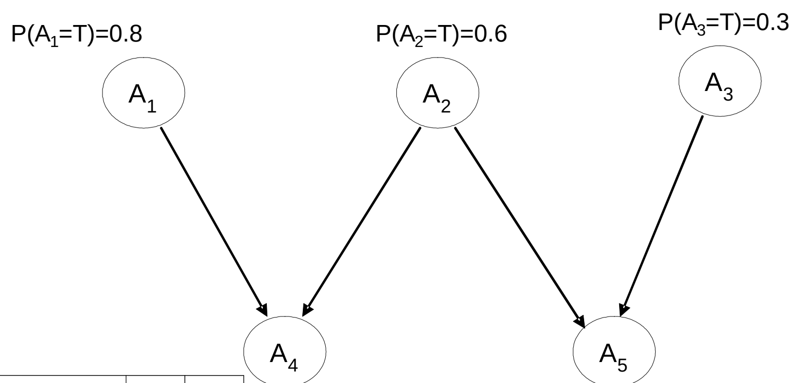
$P(A_4=T, A_1=T, A_2=T) = P(A_4=T|A_1=T,A_2=T)*P(A_1=T)*(A_2=T) = 0.9*0.8*0.6 = 0.432;$
$P(A_4=T, A_1=F, A_2=T) = P(A_4=T|A_1=F,A_2=T)*P(A_1=F)*(A_2=T) = 0.6*0.2*0.6 = 0.072;$
$P(A_4=T, A_1=F, A_2=F) = P(A_4=T|A_1=F,A_2=F)*P(A_1=F)*(A_2=F) = 0.2*0.2*0.4 = 0.016.$

With the joint probabilities, the prior probability of inferred assertion A_4 can be derived:

$P(A_4=T)$
$= P(A_4=T,A_1=T,A_2=T)+P(A_4=T,A_1=T,A_2=F)+P(A_4=T,A_1=F,A_2=T)+P(A_4=T,A_1=F,A_2=F)$
$= 0.432 + 0.24 + 0.072 + 0.016$
$= 0.76.$

Similarly, we can calculate the prior probability of A_5 and have $P(A_5=T) = 0.571$.

Figure 2: An example of a Bayesian network.



$P(A_4=T|A_1,A_2)$	A_1	A_2
0.9	T	T
0.75	T	F
0.6	F	T
0.2	F	F

$P(A_5=T|A_2,A_3)$	A_2	A_3
0.85	T	T
0.7	T	F
0.8	F	T
0.1	F	F

THE CRITERION FOR SELECTING KEY MISSING DATA

Let X represent the truth value A_x=true, ¬X represent the truth value A_x=false (i.e., $\neg A_x$=true). With this notation, we have P(X) for probability $P(A_x=\text{true})$, P(¬X) for $P(\neg A_x=\text{true})$ or $P(A_x=\text{false})$, and P(K|¬I, J) for $P(A_k=\text{true}|\neg A_i=\text{true}, A_j=\text{true}) = P(A_k=\text{true}|A_i=\text{false}, A_j=\text{true})$.

We define the *presumed contribution index of truth value U to proving truth value T, $CI_{U\rightarrow T}$*, as the difference between P(T|U) and P(T). That is, $CI_{U\rightarrow T} = P(T|U)-P(T) =$

$P(A_t{=}true|A_u{=}true){-}P(A_t{=}true)$. $CI_{\neg U\rightarrow T}$, $CI_{U\rightarrow\neg T}$, and $CI_{\neg U\rightarrow\neg T}$ are defined similarly. For example, $CI_{\neg U\rightarrow\neg T} = P(\neg T|\neg U){-}P(\neg T) = P(A_t{=}false|A_u{=}false){-}P(A_t{=}false)$.

$CI_{U\rightarrow T}$ measures how much it would help proving A_t=true if knowing A_u=true for any A_u and any A_t. If A_t is a TLC and A_u is a UOA, then $CI_{U\rightarrow T}$ shows that in case the UOA A_u's value is true, how much it would contribute to inferring that TLC A_t is true. The larger the $CI_{U\rightarrow T}$ is, the more significant 'A_u=true' is towards proving 'A_t=true'. If P_u does not have any influence on P_t, then $CI_{U\rightarrow T}= CI_{\neg U\rightarrow T}= CI_{U\rightarrow\neg T}= CI_{\neg U\rightarrow\neg T} = 0$. The presumed contribution index is an indicator of relevance between an UOA and a TLC.

However, the presumed contribution index has two flaws to be the criterion of selecting missing data. (1) Between A_u and A_t, there are four presumed contribution indices, $CI_{U\rightarrow T}$, $CI_{\neg U\rightarrow T}$, $CI_{U\rightarrow\neg T}$, and $CI_{\neg U\rightarrow\neg T}$. Which one should be used as the criterion? (2) The presumed contribution index may mislead inference sometime. If A_{u1} and A_{u2} are two UOAs, and $CI_{U1\rightarrow T} > CI_{U2\rightarrow T}$, then it looks that A_{u1} is more significant than A_{u2} in proving A_t=true if both A_{u1} and A_{u2} are true. But it is not always correct because A_{u1} and A_{u2}'s values can be false and we do not know their values when selecting one from them. If $P(A_{u1}{=}true)$ were much smaller than $P(A_{u2}{=}true)$, then selecting A_{u2} to pursue would be better than selecting A_{u1} even $CI_{U1\rightarrow T} > CI_{U2\rightarrow T}$.

To improve the above two flaws of the presumed contribution index, we define the *inference-guiding index (IG-index) of truth value A_u=true towards proving A_t=true, $IGI_{U\rightarrow T}$*, as the product of the presumed contribution index $CI_{U\rightarrow T}$ and the prior probability of U. That is,

$$IGI_{U\rightarrow T} = CI_{U\rightarrow T} * P(U).$$

$IGI_{U\rightarrow T}$ is a better indicator of contribution of A_u than $CI_{U\rightarrow T}$ since it takes both the contribution of A_u=true for deriving A_t=true and the prior probability of A_u=true into account. The IG-index $IGI_{U\rightarrow T}$ is high if both the presumed contribution of U and probability of U are high. With this improvement on the presumed contribution index, the IG-index can be called the "*true contribution index*" of A_u=true for deriving A_t=true. Thus, the second flaw of the presumed contribution index mentioned above is taken care of. But among the four IG-indices between A_u and A_t, $IGI_{U\rightarrow T}$, $IGI_{U\rightarrow\neg T}$, $IGI_{\neg U\rightarrow T}$, and $IGI_{\neg U\rightarrow\neg T}$, which one should be used as the criterion in inference-guiding? Theorem 1 below answers this question by showing that the absolute values of the four IG-indices are same. To prove Theorem 1, we need three lemmas.

Lemma 1.

$IGI_{\neg U\rightarrow\neg T} = IGI_{U\rightarrow T}$ for any assertion A_u and A_t.

Proof:

By definitions of $IGI_{\bullet\rightarrow\bullet}$ and $CI_{\bullet\rightarrow\bullet}$, and the probability theory, for any A_u and A_t we have:

$$\begin{aligned} IGI_{\neg U\rightarrow\neg T} &= CI_{\neg U\rightarrow\neg T} * P(\neg U) && \text{(by definition of } IGI_{\bullet\rightarrow\bullet}\text{)} \\ &= (P(\neg T|\neg U) - P(\neg T)) * P(\neg U) && \text{(by definition of } CI_{\bullet\rightarrow\bullet}\text{)} \end{aligned}$$

$= P(\neg T|\neg U) * P(\neg U) - P(\neg T) * P(\neg U)$
$= P(\neg T|\neg U) * P(\neg U) - (1- P(T)) * (1-P(U))$
$= P(\neg T|\neg U) * P(\neg U) - (1- P(T) - P(U) + P(T)*P(U))$
$= P(\neg T,\neg U) - 1+ P(T) + P(U) - P(T)*P(U))$
$= P(\neg T,\neg U) -1+ P(T,U) + P(T,\neg U) + P(U,T) + P(U,\neg T) -P(T)*P(U))$.

Since $P(\neg T,\neg U) + P(T, U) + P(T, \neg U) + P(U, \neg T) = 1$,

$IGI_{\neg U\to\neg T} = P(U, T) - P(T)*P(U)$
$= P(T, U) - P(T)*P(U)$
$= P(T|U)*P(U) - P(T)*P(U)$
$= (P(T|U) - P(T)) * P(U)$
$= CI_{U\to T} * P(U)$
$= IGI_{U\to T}$.

#

Lemma 2.

$IGI_{U\to\neg T} = IGI_{\neg U\to T}$, for any assertion A_u and A_t.

Proof:

Let A_v an assertions. By Lemma 1, $IGI_{\neg V\to\neg T} = IGI_{V\to T}$. Let $A_u=\neg A_v$. So, $\neg A_u=A_v$, i.e., $\neg U=V$. Make the substitution, we have $IGI_{U\to\neg T} = IGI_{\neg U\to T}$.

#

Lemma 3.

$IGI_{U\to T} = -IGI_{U\to\neg T}$ for any assertion A_u and A_t.

Proof:

$IGI_{U\to\neg T} = (P(\neg T|U) - P(\neg T)) * P(U)$ (by definition of $IGI_{\bullet\to\bullet}$)
$= (1 - P(T|U) - 1 + P(T)) * P(U)$
$= - (P(T|U) - P(T)) * P(U)$
$= - CI_{U\to T} * P(U)$
$= - IGI_{U\to T}$

#

Theorem 1.

$IGI_{U\to T} = IGI_{\neg U\to\neg T} = - IGI_{U\to\neg T} = - IGI_{\neg U\to T}$ for any A_u and A_t in a Bayesian network.

Proof:

Putting Lemma 1, 2, and 3 together, we have $IGI_{U\to T} = IGI_{\neg U\to\neg T} = - IGI_{U\to\neg T} = - IGI_{\neg U\to T}$ for any A_u and A_t in a Bayesian network.

#

Recall that $IGI_{U\rightarrow T}$ is interpreted as the contribution index of A_u=true for proving A_t=true. Then, $IGI_{\neg U\rightarrow\neg T}$ is interpreted as the contribution index of A_u=false for proving A_t=false. A negative IGI value is viewed as the contribution to 'disproving'. So, $-IGI_{U\rightarrow\neg T}$ represents the contribution of 'A_u=true' to disproving 'A_t=false'.

With the above interpretation, Theorem 1 tells that the true contribution of A_u=true to proving A_t=true is same as its true contribution to disproving A_t=false; and no matter whether A_u=true or A_u=false, it has contributions to confirming the value of A_t, either A_t=true or A_t=false, and the amounts of contributions are same. That is just the property we want the IG criterion to possess, - It would select such a UOA that once its value is known it would help confirm a TLC to a large extent.

By Theorem 1, absolute values of the four IG-indices between two assertions, A_u and A_t are same. Let $|IGI_{A_u\rightarrow A_t}|$ denote the absolute value of four IG-indices between A_u and A_t. For a TLC A_t, we would select a UOA A_u such that $|IGI_{A_u\rightarrow A_t}|$ is the largest comparing to the other UOAs, even we do not yet know the value of A_u at the time of selection.

Example 2. Illustration of Calculations and Property of IG-index

Continuing Example 1, in which A_1, A_2 and A_3 are UOAs, and A_4 and A_5 are TLCs.

For A_1, A_2, and A_4, the following probabilities are given in the Bayesian network: $P(A_1)=0.8$, $P(A_2)=0.6$, $P(A_4|A_1,A_2)=0.9$, $P(A_4|A_1,\neg A_2)=0.75$, $P(A_4|\neg A_1,A_2)=0.6$, $P(A_4|\neg A_1,\neg A_2)=0.2$. The calculation results of presumed contribution indices and IG-indices are shown in Table 1.

Table 1: Presumed contribution indices and IG-indices relative to A_4.

Conditional Probabilities		Presumed contribution index to A_4		IG-index to A_4	
$P(A_4\|A_1)$	0.84	$CI_{A_1\rightarrow A_4}$	0.08	$IGI_{A_1\rightarrow A_4}$	0.064
$P(A_4\|\neg A_1)$	0.44	$CI_{\neg A_1\rightarrow A_4}$	-0.32	$IGI_{\neg A_1\rightarrow A_4}$	-0.064
$P(\neg A_4\|A_1)$	0.16	$CI_{A_1\rightarrow\neg A_4}$	-0.08	$IGI_{A_1\rightarrow\neg A_4}$	-0.064
$P(\neg A_4\|\neg A_1)$	0.56	$CI_{\neg A_1\rightarrow\neg A_4}$	0.32	$IGI_{\neg A_1\rightarrow\neg A_4}$	0.064
$P(A_4\|A_2)$	0.84	$CI_{A_2\rightarrow A_4}$	0.08	$IGI_{A_2\rightarrow A_4}$	0.048
$P(A_4\|\neg A_2)$	0.64	$CI_{\neg A_2\rightarrow A_4}$	-0.12	$IGI_{\neg A_2\rightarrow A_4}$	-0.048
$P(\neg A_4\|A_2)$	0.16	$CI_{A_2\rightarrow\neg A_4}$	-0.08	$IGI_{A_2\rightarrow\neg A_4}$	-0.048
$P(\neg A_4\|\neg A_2)$	0.36	$CI_{\neg A_2\rightarrow\neg A_4}$	0.12	$IGI_{\neg A_2\rightarrow\neg A_4}$	0.048

For A_2, A_3, and A_5, the following probabilities are given in the Bayesian network: $P(A_2)=0.6$, $P(A_3)=0.2$, $P(A_5|A_2,A_3)=0.85$, $P(A_5|A_2,\neg A_3)=0.7$, $P(A_5|\neg A_2,A_3)=0.8$, $P(A_5|\neg A_2,\neg A_3)=0.1$. The calculation results of the presumed contribution indices and IG indices are shown in Table 2.

Table 2: Presumed contribution indices and IG-indices relative to A_5.

Conditional Probabilities		Presumed contribution index to A_5		IG-index to A_5	
$P(A_5\|A_2)$	0.745	$CI_{A_2\to A_5}$	0.174	$IGI_{A_2\to A_5}$	0.1044
$P(A_5\|\neg A_2)$	0.31	$CI_{\neg A_2\to A_5}$	-0.261	$IGI_{\neg A_2\to A_5}$	-0.1044
$P(\neg A_5\|A_2)$	0.255	$CI_{A_2\to\neg A_5}$	-0.174	$IGI_{A_2\to\neg A_5}$	-0.1044
$P(\neg A_5\|\neg A_2)$	0.69	$CI_{\neg A_2\to\neg A_5}$	0.261	$IGI_{\neg A_2\to\neg A_5}$	0.1044
$P(A_5\|A_3)$	0.83	$CI_{A_3\to A_5}$	0.259	$IGI_{A_3\to A_5}$	0.0777
$P(A_5\|\neg A_3)$	0.46	$CI_{\neg A_3\to A_5}$	-0.111	$IGI_{\neg A_3\to A_5}$	-0.0777
$P(\neg A_5\|A_3)$	0.17	$CI_{A_3\to\neg A_5}$	-0.259	$IGI_{A_3\to\neg A_5}$	-0.0777
$P(\neg A_5\|\neg A_3)$	0.54	$CI_{\neg A_3\to\neg A_5}$	0.111	$IGI_{\neg A_3\to\neg A_5}$	0.0777

In summary, $|IGI_{A_1\to A_4}|$ =0.064, $|IGI_{A_2\to A_4}|$ =0.048, $|IGI_{A_2\to A_5}|$ =0.1044, $|IGI_{A_3\to A_5}|$ =0.0777, and $|IGI_{A_1\to A_5}| = |IGI_{A_3\to A_4}| = 0$.

#

Further investigations have revealed more characteristics of the IG-index. Theorem 2 and Corollary 1 below establish the correspondence between IG-indices and covariances.

Theorem 2.

Let X_u and X_t be 0-1 random variables with 1 representing 'true' and 0 representing 'false'. Let U represent the truth value X_u=true, ¬U represent the truth value $\neg X_u$=true (i.e. X_u=false). Let T represent the truth value X_t=true, ¬T represent the truth value $\neg X_t$=true (i.e. X_t=false). Then, $IGI_{U\to T} = Cov(X_u, X_t)$, where $Cov(X_u, X_t)$ stands for covariance of X_u and X_t.

Proof:

By the definition of covariance and probability theory:

$$\begin{aligned} Cov(X_u, X_t) &= E[(X_u-E[X_u])(X_t-E[X_t])] \\ &= E[X_uX_t - X_uE[X_t] - X_tE[X_u] + E[X_u][X_t]] \\ &= E[X_uX_t] - E[X_u]E[X_t] - E[X_t]E[X_u] + E[X_u][X_t] \\ &= E[X_uX_t] - E[X_u]E[X_t] \end{aligned}$$

Let $Y=X_uX_t$. Y is such a 0-1 random variable that Y=1 if and only if both $X_u=1$ and $X_t=1$.
Then, $P(Y=1) = P(X_u=1, X_t=1)$, and

$$\begin{aligned} E[X_uX_t] &= E[Y] \\ &= P(Y=1) \\ &= P(X_u=1, X_t=1) \\ &= P(X_t=1, X_u=1) \end{aligned}$$

$= P(X_t=1|X_u=1)P(X_u=1)$
$= P(T|U)P(U).$

And $E[X_u] = P(X_u=1) = P(U)$, and $E[X_t] = P(X_t=1) = P(T)$.

Therefore,

$$\begin{aligned} Cov(X_u, X_t) &= E[X_u X_t] - E[X_u]E[X_t] \\ &= P(T|U)P(U) - P(U)P(T) \\ &= (P(T|U) - P(T))\, P(U) \\ &= IGI_{U\to T}. \end{aligned}$$

#

Theorem 2 shows that the inference-guiding index, $IGI_{U\to T}$, is the covariance of two random variables corresponding to U and T. Covariance has limited use in the probability theory (for example, Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)), and its properties are not explored sufficiently (Ross, 1985). Theorem 2 can be extended to the following corollary about other IG-indices and their corresponding covariances. Proofs are straightforward, so not provided.

Corollary 1.

Let us define random variables $\bar{X}_u = 1 - X_u$ and $\bar{X}_t = 1 - X_t$. $\bar{X}_u$ =1 if and only if X_u=0 (or X_u=false, or ¬U); and $\bar{X}_u$ =0 if and only if X_u=1 (or X_u is true, or U). $\bar{X}_t$ =1 if and only if X_t=0 (or X_t is false, or ¬U); and $\bar{X}_t$ =0 if and only if X_t=1 (or X_t is true, or U). Then:

$IGI_{\neg U\to T} = Cov(\bar{X}_u, X_t)$,
$IGI_{U\to \neg T} = Cov(X_u, \bar{X}_t)$,
$IGI_{\neg U\to \neg T} = Cov(\bar{X}_u, \bar{X}_t)$.

#

A REASONING APPROACH WITH INFERENCE-GUIDING

In this section a reasoned approach is presented that integrates logical inference and inference-guiding functions for a Bayesian knowledge-based system. The absolute value of IG-index, as we have investigated in Section 3, is used as the criterion in inference-guiding.

Let {TLC} be the index set of TLCs, {UOA} be the index set current UOAs. Let 1>h>0 be a preset threshold for the acceptable certainty level of a TLC. The reasoning approach with inference-guiding function for Bayesian knowledge-based systems is formalized as follows.

IG-index Inference Approach:

Step 1. Reasoning

Calculate $P(A_i)$ for each $i \in \{TLC\}$.
Pick up A_t such that $t \in \{TLC\}$ and $P(A_t) = Max_{i\in\{TLC\}} P(A_i)$.
If $P(A_t) \geq h$, Stop, - the inference is done with A_t being the inferred TLC; otherwise, go to Step 2.

Step 2. Inference-Guiding

If {UOA} is empty, then Stop, - no TLC with certainty level equal to or higher than h can be derived.

Calculate $|IGI_{A_i \to A_t}|$ for each $i \in \{UOA\}$.

Pick a UOA A_u such that $|IGI_{A_u \to A_t}| = \text{Max}_{i \in \{UOA\}} |IGI_{A_i \to A_t}|$.

Find the truth value of A_u from an information resource in environment.

Step 3. Updating

Update the set {UOA} such that $\{UOA\} = \{UOA\} \setminus \{A_u\}$.

Recalculate the conditional probabilities in Bayesian network with the newly obtained truth value of A_u.

Go back to Step 1.

#

Step 1 is doing logical inference. There are various inference algorithms available for the user to choose from, such as variable elimination, direct sampling, and Markov chain simulation (Russell & Norvig, 2003). The problem of logical inference in a Bayesian network is, same as in a knowledge base with other structures, computationally hard. The classic method based on the probability theory gives accurate results but are not efficient for large Bayesian networks. Approximate methods are efficient, but their results are more or less inaccurate. A user may choose an inference algorithm for this step at his/her discretion.

Step 2 is doing inference-guiding. It first calculates the IG-index for each UOA, A_i, associated with the TLC A_t selected in Step 1. To do it, recall that the absolute value of IG-index is calculated with the formula:

$$|IGI_{A_i \to A_t}| = |((P(A_t|A_i) - P(A_t)) * P(A_i)|.$$

The value of $P(A_i)$ is given in the network since A_i is a UOA. $P(A_t)$ was calculated in Step 1. To calculate $P(A_t|A_i)$, we set A_i to 'true' and then use the inference algorithm selected in Step 1 to figure out the probability of A_t=true under that circumstance.

The UOA with the largest IG-index absolute value is picked as the key missing data, and its value is pursued from an information source such as a sensor or the user.

After obtaining the truth value of the selected UOA, A_u, Step 3 removes A_u from the UOA set, since A_u is no longer unconfirmed, and recalculates the conditional probabilities in the Bayesian network. For each assertion influenced by A_u, its conditional probabilities must be recalculated. Suppose A_u=true from an information source, A_x is an inferred assertion that is influenced by A_u, and the conditional probabilities are in the format of $P(X|z, A_u)$ where z represents a set of assertions. For example, $z=\{A_1, A_2\}$, then $P(X|z, A_u) = P(X|A_1, A_2, A_u)$. By the probability theory, if A_u is 'true' (i.e., $P(A_u)=1$), then the new conditional probability is:

$$\begin{aligned} P(X|z) &= P(X,z) / P(z) = (P(X,z,A_u) + P(X,z,\neg A_u)) / P(z) \\ &= P(X,z,A_u) / P(z) \qquad \text{(since } P(\neg A_u)=0) \\ &= (P(X|z,A_u) * P(z,A_u)) / P(z) \end{aligned}$$

$= (P(X|z,A_u) * P(z|A_u) * P(A_u)) / P(z)$
$= P(X|z,A_u) * (P(z|A_u) / P(z))$ (since $P(A_u)=1$)

Similarly, if A_u is 'false' (i.e., $P(A_u)=0$), then the new conditional probability is:
$P(X|z) = (P(X|z,\neg A_u) * (P(z|\neg A_u) / P(z))$.

In the formulas for updating conditional probabilities, $P(X|z,A_u)$ and $P(X|z,\neg A_u)$ are currently given at the node A_x in the Bayesian network before A_u's truth value is obtained. So, the new conditional probability is calculated from the current probability by multiplying a ratio proportional to the correlation between A_u and z. If A_u is independent of z, then $P(z|A_u) = P(z|\neg A_u) = P(z)$, and the ratios $(P(z|A_u) / P(z))$ and $(P(z|\neg A_u) / P(z))$ will become one, so that $P(X|z)=P(X|z,A_u)$ if A_u=true, and $P(X|z)=P(X|z, \neg A_u)$ if A_u=false. If A_u is not independent of z, then the probabilities $P(z|A_u)$ and $P(z)$ are calculated from the joint probabilities with the formulas $P(z|A_u)=P(z,A_u)/P(A_u)$, and $P(z)=P(z, A_u)+P(z,\neg A_u)$.

The three steps are executed repeatedly until a TLC with certainty level higher than the preset threshold *h* so that we can claim the TLC is reached, or all UOAs are asked and no TLC reaches the threshold.

Example 3. A complete inference process with IG-index Inference Approach

Continuing Example 1 and Example 2. Recall that A_1, A_2, and A_3 are UOAs, and A_4 and A_5 are TLCs. Suppose we set threshold h=0.82. We have initially {UOA}={A_1, A_2, A_3} and {TLC}={A_4, A_5}.

Iteration 1.

Step 1. Calculate the probabilities of TLCs:
$P(A_4) = 0.76$, and $P(A_5) = 0.571$ (as calculated in Example 1 in Section 2).
Neither reaches the threshold h. We pick TLC A_4 since $P(A_4) > P(A_5)$.

Step 2. The two influential UOAs on A_4 are A_1 and A_2 (referring to Fig. 1), whose IG-indices with respective to A_4 are $|IGI_{A_1 \rightarrow A_4}|= 0.064$, and $|IGI_{A_2 \rightarrow A_4}|= 0.048$, as calculated in Example 2. Since $|IGI_{A_1 \rightarrow A_4}|$ is the larger, we pick UOA A_1 and pursue its truth value. Suppose an information source in environment gives A_1's value that is 'true'.

Step 3.
Update {UOA} so that the new {UOA} = {A_2, A_3}. The conditional probabilities at node A_4, after knowing A_1=T, are updated. Note that A_1 is independent of A_2. We have: $P(A_4=true|A_2=true)=0.9$, and $P(A_4=true|A_2=false)=0.75$.
The conditional probabilities at node A_5 do not change since A_1 has no influence on A_5.

Iteration 2.

Step 1. Calculate the probabilities of TLCs:
$P(A_4) = 0.84$, and

$P(A_5) = 0.571$.

Since $P(A_4) > h = 0.82$, we are done with this inference process and the TLC derived is A_4 whose certainty level is 84%.

#

CONCLUSION

IG-index and the inference approach presented in this paper are for the Bayesian network that is a robust structure of the knowledge base containing uncertainties. The IG-index provides an effective criterion in selecting the key missing information when given data is not sufficient to reach a top-level-conclusion. The reasoning approach, IG-index Inference Approach, integrates inference-guiding function with logical inference function, and "smartly" leads reasoning to a top-level-conclusion.

REFERENCES

Ben-Gal, I. (2007). Bayesian Networks, in F. Ruggeri, R. Kenett, & F. Faltin (editors), *Encyclopedia of Statistics in Quality and Reliability*, John Wiley & Sons.

Bishop, C. M., Spiegelhalter, D., & Winn, J. (2003). VIBES: A Variational Inference Engine for Bayesian Networks, on *Advances in Neural Information Processing Systems (NIPS)*, May, http://books.nips.cc/papers/files/nips15/AA37.pdf

Buchanan, B. G. & Shortliffe, E. H. (1984). *Rule-Based Expert Systems - The MYCIN Experiments of the Stanford Heuristic Programming Project*, Addison-Wesley, Reading, MA.

Chavira, M. & Darwiche, A. (2007). Compiling Bayesian Networks Using Variable Elimination. *Proceedings of IJCAI 2007*.

Duda, R., Gasching, J., & Hart, P. (1979). Model Design in the PROSPECT Consultant System for Mineral Exploration, Expert Systems in the Micro-Electronic Age, Edinburgh Univ. Press, UK.

Duda, R. O., Hart, P. E., & Nilsson, N. J. (1976). Subjective Bayesian Methods for Rulebased Inference Systems, *Proceedings of National Computer Conference,* 45, AFIPS, 1075-1082.

Fish, K. E. (2006). An Artificial Intelligence Approach to International Market Screening DSS. *Journal of International Technology and Information Management,* 15(2), 49-60.

Hall, O. P. Jr., & McPeak, C. J. (2003). Using Neural Net Technology to Analyze Corporate Restructuring Announcements. *Journal of International Technology and Information Management*, 12(2), 29-40.

Hayes-Roth, E., Waterman, D. A., & Lenat, D. B. (1983). Building Expert Systems, Addison-Wesley, Reading, MA

Human Edge Software Corporation, Expert Edge Manual for the IBM Personal Computer, California, 1985.

Hung, S. Y., Tang, K. Z., & Shu, T. C. (2008). Expanding Group Support System Capabilities from the Knowledge Management Perspective. *Journal of International Technology and Information Management*, 17(1) 21-42.

Ignizio, J. P. (1991). *An Introduction to Expert Systems*, McGraw-Hill, Inc.

Jensen, F. V. (2001). *Bayesian Networks and Decision Graphs*, Springer-Verlag, Berlin

Jeroslow, R. G. & Wang, J. (1989). Dynamic Programming, Integral Polyhedra and Horn Clause Knowledge Bases, *ORSA's Journal on Computing* 1, 7-19

Kanal, L. N. & Lemmer, J. F. (1986). *Uncertainty in Artificial Intelligence*, Elsevier/North-Holland, Amsterdam, London, New York.

Kickert, W. J. M. (1978). *Fuzzy Theories on Decision Making*, Leiden, Boston, Mass.

Langseth, H. & Portinale, L. (2007). Bayesian networks in reliability. *Reliability Engineering and System Safet,y* 92(1), 92–108.

Liu, R. F. & Soetjipto, R. (2004). *Analysis of Three Bayesian Network Inference Algorithms: Variable Elimination, Likelihood Weighting, and Gibbs Samplin.*, http://www.eecs.berkeley.edu/~rfl/publications/6825p2.pdf, rliu.rusmin@mit.edu, Nov. 4, 2004.

McManus, D. J. & Snyder, C. A. (2003). Knowledge Management: The Role of EPSS. *Journal of International technology and Information Management*, 12(2), 17-28.

Magill, W. G. W. & Leech, S. A. (1991). Uncertainty Techniques in Expert System Software. *Decision Support Systems*, 7, 55-65.

Mellish, C. S. (1985). Generalized Alpha-beta Pruning as a Guide to Expert System Question Selection, Expert System 85, *Proceedings 5th Technical Conference of the British Computer Society Specialist Group on Expert Systems*, Univ. of Warwick, (Cambridge Univ. Press, Cambridge CB2 1RP).

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann.

Ross, S. (1985). *Introduction to Probability Models, 3rd Edition,* Academic Press, Inc, London, 49-50.

Russell, S. & Norvig, P. (2003). *Artificial Intelligence, a Modern Approach,* Prentice Hall, Pearson Education Inc., Upper Saddle River, New Jersey.

Shortliffe, E. H. & Buchanan, B. G. (1975). A Model of Inexact Reasoning in Medicine. *Mathematical Biosciences*, 23, 351-379.

Triantaphyllou, T. & Wang, J. (1992). A Cost-Effective Question-Asking Strategy. *The Proceedings of the ACM/SIGAPP Symposium on Applied Computing*, Kansas City, 505-511.

Ullman, J. D. (1989). *Principles of Database and Knowledge-Based Systems*, Computer Science Press, Rockville, Maryland.

Wang, J. (2005a). Inference-Guiding for Intelligent Agents. *Journal of International Technology and Information Management*, 14(2) 93-108.

Wang, J. (2005-b) Inference-Guiding in a Knowledge Base with Bayesian Network Structure. *Proceedings of NEDSI Annual Conference*, Philadelphia, PA.

Wang, J. (1994). Identifying Key Missing Data for Inference under Uncertainty. *International Journal of Approximate Reasoning*, 10, 287-309.

Wang, J. & Triantaphyllou, E. (1994). The Problem of Asking the Minimum Number of Questions in Expert Systems with Horn Clause Knowledge Bases. *Mathematical and Computer Modeling*, 20(9), 75-87.

Wang, J. & Vande Vate, J. (1990). Question-Asking Strategies for Horn Clause Systems. *Annals of Mathematics and Artificial Intelligence*, 1, 359-370.

Wray, B. A., Markham, I. S., & Mathieu, R. G. (2003). An Artificial Neural Network Approach to Learning from Factory Performance in a Kanban-Based System. *Journal of International Technology and Information Management*, 12(2), 85-98.

Zadeh, L. A. (1965). *Fuzzy Sets, Information and Control*, 8, 338-353.