Belief Rule-Base Inference Methodology Using the Evidential Reasoning Approach—RIMER

Jian-Bo Yang, Jun Liu, Jin Wang, *Member, IEEE*, How-Sing Sii, and Hong-Wei Wang, *Associate Member, IEEE*

***Abstract*—In this paper, a generic rule-base inference methodology using the evidential reasoning (RIMER) approach is proposed. Existing knowledge-base structures are first examined, and knowledge representation schemes under uncertainty are then briefly analyzed. Based on this analysis, a new knowledge representation scheme in a rule base is proposed using a belief structure. In this scheme, a rule base is designed with belief degrees embedded in all possible consequents of a rule. Such a rule base is capable of capturing vagueness, incompleteness, and nonlinear causal relationships, while traditional if–then rules can be represented as a special case. Other knowledge representation parameters such as the weights of both attributes and rules are also investigated in the scheme. In an established rule base, an input to an antecedent attribute is transformed into a belief distribution. Subsequently, inference in such a rule base is implemented using the evidential reasoning (ER) approach. The scheme is further extended to inference in hierarchical rule bases. A numerical study is provided to illustrate the potential applications of the proposed methodology.**

***Index Terms*—Decision-making, evidential reasoning approach, expert system, fuzzy sets, inference mechanisms, rule-based system, uncertainty.**

I. Introduction

AMONG many alternative means for knowledge representation, rules seem to be one of the most common forms for expressing various types of knowledge for a number of reasons [1]. It has been argued that other knowledge representation schemes can be transformed into logic (rule)-based schemes [2]–[4]. As such, knowledge-based systems (e.g., rule-based expert systems), usually constructed from human knowledge in forms of if–then rules, have become the most visible and fastest growing branch of artificial intelligence (AI) [1].

There are two essential components in a rule-based system: a knowledge base and an inference engine. They are combined to infer useful conclusions from rules established by experts and observation facts provided by users. In the design and implementation of rule-based systems for supporting human decision making, it is inevitable to deal with uncertainty caused by vagueness intrinsic to human knowledge and imprecision or incompleteness resulting from the limit of human knowledge [5], [6]. It is therefore necessary to use a scheme for representing and processing vague, imprecise, and incomplete information in conjunction with precise data.

The development of methods for dealing with uncertainty has received considerable attention in the last three decades. Several numerical and symbolic methods have been proposed for handling uncertain information. Three of the most common frameworks for representing and reasoning with uncertain knowledge are:

1) Bayesian probability theory;
2) Dempster–Shafer (D–S) theory of evidence;
3) fuzzy set theory.

Each of these frameworks is aimed at a special application environment and has its own features [7]–[12]. In fact, different kinds of uncertainty may coexist in real systems, e.g., fuzzy information may coexist with ignorance, leading to the induction of knowledge without certainty but only with degrees of belief or credibility regarding a hypothesis [13]. Therefore, it is highly desirable to develop a hybrid knowledge representation scheme and inference methodology to deal with different kinds of uncertainty. For instance, the benefit of combining fuzzy set theory with D–S theory of evidence may be substantial in situations where fuzziness and ignorance in data become prevalent [12], [14].

In addition, attributes involved in the premise of rules may not be of the same type. Indeed, they could be quantitative or qualitative in nature, and input data may be different both in type and in scale. Hence, there is also a need to establish a mathematical framework that can provide a basis for handling various types of input information.

This paper reports an investigation into the design and implementation of hybrid rule-based systems based on the D–S theory of evidence, decision theory, and fuzzy set theory. As a result of the investigation, a new methodology will be proposed for modeling a hybrid rule base using a belief structure and for inference in the rule-based system using the evidential

Manuscript received July 1, 2003; revised January 19, 2004 and May 30, 2004. This work forms part of the projects supported by the U.K. Engineering and Physical Sciences Research Council under Grants GR/R30624, GR/R32413, GR/S85498/01, and GR/S85504/01. This paper was recommended by Associate Editor Pan.

J.-B. Yang is with the Decision Sciences and Operations Management Group, Manchester Business School, The University of Manchester, Manchester M60 1QD, U.K., and also with the Institute of Systems Engineering, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China (e-mail: jian-bo.yang@manchester.ac.uk).

J. Liu is with School of Computing and Mathematics, University of Ulster at Jordanstown, Newtownabbey, BT37 0QB, Northern Ireland, U.K. (e-mail: j.liu@ulster.ac.uk).

J. Wang and H.-S. Sii are with the School of Engineering, Liverpool John Moores University, Liverpool L3 3AF, U.K. (e-mail: etmjwang@livjm.ac.uk; h.s.sii@livjm.ac.uk).

H.-W. Wang is with the Control Theory and Engineering Department, Systems Engineering Institute, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China (e-mail: hwwang@mail.hust.edu.cn).

Digital Object Identifier 10.1109/TSMCA.2005.851270

reasoning (ER) approach [15]–[20]. The methodology is referred to as a generic rule-base inference methodology using the evidential reasoning (RIMER) approach. In the RIMER approach, a detailed analysis is first conducted on the nature of antecedent attributes and various types of uncertainties in knowledge. A generic knowledge representation scheme is then proposed using a belief structure. A rule base designed on the basis of the belief structure, called belief rule base, is used to capture nonlinear causal relationships as well as uncertainty. Relevant knowledge representation parameters, including the weights of both attributes and rules, are also considered in the scheme.

In an established belief rule base, input for each antecedent is transformed into a distribution on the referential values of this antecedent. This distribution describes the degree of each antecedent being activated. Moreover, the antecedents of an if–then rule form an overall attribute, called a packet antecedent attribute. The activation weight of a rule can be generated by aggregating the degrees to which all antecedents in the rule are activated. In this context, an if–then rule can be considered as an evaluation problem of a packet antecedent attribute being assessed to an output term in the consequent of the rule with certain degrees of belief. Finally, the inference of a rule-based system is implemented using the ER approach. A further investigation into the inference problem in a hierarchical knowledge base is also provided. A numerical study is used to illustrate the application of the proposed methodology. The new methodology is also applied to the safety analysis of an offshore engineering system and is reported in another paper [21], [22].

The paper is organized as follows: In Section II, an investigation into rule-based system design is presented, including the structure of a knowledge base and knowledge representation schemes under uncertainty. Based on this investigation, a new generic knowledge-base inference methodology using the ER approach (RIMER) is proposed in Section III. A further investigation into knowledge representation and inference in a hierarchical knowledge base is given in Section IV. In Section V, a numerical study is provided to illustrate the methodology. The paper is concluded in Section VI.

II. Investigation Into Rule-Based System Design

A. Attribute Types in a Rule Base

1) Attribute as a Proposition: In rule-based systems, there are three main types of propositions: boolean, fuzzy, and random. Boolean propositions represent concepts that can only be evaluated to be either false or true. Fuzzy propositions are related to vague concepts. Take a rule in medical diagnosis for example [23].

"If presence of creatinine then renal failure is definite."

In the above rule, presence of creatinine is a fuzzy proposition, which can be quantified by the degree of membership of a numerical value to a fuzzy set. The presence of creatinine is modeled by defining a fuzzy set [24] that takes as argument the numerical value in millimolar per liter of creatinine.

On the other hand, the proposition "raining causes flooding" or "disease D is a common illness" is a probabilistic rather than deterministic or fuzzy proposition, which is determined by previous statistics.

2) Attribute as a Variable: This type of attribute is defined by a set of values that an attribute can take. Type distinction is one of the forms used to separate the different types of existing data.

Let $U = \{U_i;\ \ i = 1, \ldots, T\}$ be a set of attributes, and L_{U_i} be a set of feasible values for U_i $(i = 1, \ldots, T)$. Each attribute U_i $(\mathrm{i} = 1, \ldots, T)$ in U can be one of the following types [11]:

1) continuous, where L_{U_i} is a convex subset of real numbers (e.g., distances and measures on a continuous scale);
2) numeric, where L_{U_i} is a discrete finite set of real numbers (e.g., number of legs for mammals);
3) symbolic, where L_{U_i} is a discrete finite set of symbols (e.g., car models or telephone numbers);
4) ordered symbolic, where L_{U_i} is a discrete finite set of ordered symbols. For instance, colors can be ordered by their wavelength. Linguistic information is a kind of ordered symbolic information.

It is possible that some attributes can be measured numerically (e.g., age) and other attributes can only be described subjectively (e.g., good). In general, continuous and numerical attributes can be in quantitative format, while symbolic and ordered symbolic attributes can be in qualitative format.

B. Uncertainty in Data and Knowledge Bases and Their Representation

In conventional information processing techniques, it is often assumed that problems are well structured, complete information is always available, and information processing procedures can be clearly defined. However, in many real-world decision-making problems, this is not always the case, and decision making may be associated with uncertainty. Uncertainty can occur because information is not clearly described, or only by partial and imprecise evidence, which is a result of ill-defined concepts in observations, or due to the inaccuracy and poor reliability of instruments used to make the observations.

In the design of decision-making models based on multisource data, it poses a challenge to find an appropriate framework for an identified form of uncertainty and to combine different strategies for formulating a proposition or a variable that can express more than one type of uncertainty. In knowledge-based systems, there are situations where information cannot be acquired precisely in a quantitative form but may be extracted in a qualitative form. Human judgments and domain knowledge can be represented in forms of if–then rules, which are normally based on linguistic variables because they are more natural and expressive than numerical numbers. This makes it necessary to use a linguistic approach to process qualitative information. Fuzzy logic [24], [25] provides a mathematical framework to systematically process linguistic information.

Furthermore, propositional statements provided by domain experts may be crisp and uncertain, crisp and certain, fuzzy and certain, or fuzzy and uncertain, as illustrated by the following examples:

1) John is young with a certainty of 1 (fuzzy and certain).
2) John is young with a certainty of 0.8 (fuzzy and uncertain).
3) John is 25 years old with a certainty of 1 (crisp and certain).
4) John is 25 years old with a certainty of 0.9 (crisp and uncertain).

In (2), 0.8 means that "we are 80% sure" and that the remaining 20% is ignorance. In the following proposed rule-base inference procedure, fuzziness and ignorance in input information are both taken into account.

In a rule-based system, a rule is used to describe causal relationships between antecedent attributes and their associated consequent. In the design and implementation of rule-based systems, ignorance may be caused by weak implication that may occur when an expert is unable to establish a precise correlation between premise and conclusion [26] but only with degrees of belief or credibility [13]. The D–S theory of evidence [27], [28] is based on the concept of belief function and is well suited to modeling subjective credibility induced by partial evidence [29]. It describes and handles uncertainties using the concept of the degrees of belief, which can model ignorance explicitly. It also provides appropriate methods for computing belief functions for combination of multiple pieces of evidence.

Although the D–S theory was not specifically proposed in relation to AI, it has been increasingly used in AI and expert systems in the past two decades [30]–[36]. The D–S theory has also been used for multiple attribute decision analysis (MADA) under uncertainty, as shown in the development of the ER approach [14]–[20], [37]. Different from conventional MADA methods that describe a MADA problem using a decision matrix, the ER approach uses a belief decision matrix, in which the assessment of an alternative on each attribute is described by a distribution using a belief structure. The main advantage of doing this is that one can model precise data and also capture various types of uncertainties such as ignorance and vagueness in subjective judgments. In the following proposed new rule-based systems, the ER approach will be used as the basis for rule combination in the final steps of the inference procedure.

Fuzzy set theory is well suited to dealing with fuzziness, and D–S theory provides an ideal framework for handling ignorance. It seems reasonable to extend the fuzzy logic framework to cover credibility uncertainty. Several researchers have investigated the relationships between fuzzy sets and D–S theory and suggested different ways of integrating them. Ishizuka *et al.* [38], [39] extended the D–S theory to include fuzzy knowledge for structural damage assessment. Ogawa *et al.* [40] subsequently proposed an inexact inference procedure on the base of the D–S theory of evidence and fuzzy sets to make the structural damage assessment more general and practical. Chen [41] extended Ishizuka's rule-base inference for more general decision-making problems. Binaghi and Madella [13] used fuzzy D–S reasoning for rule-based classification. Yager [42], Yager and Filev [43], and Baldwin *et al.* [44] also considered fuzzy sets in the D–S belief structure to deal with probability uncertainty.

C. Basic Knowledge Base Structure

The starting point of constructing a rule-based system is to collect if–then rules from human experts or based on domain knowledge. A knowledge base and an inference engine are then designed to infer useful conclusions from the rules and observation facts provided by users.

Formally, a rule-based model is represented as

$$R = \langle U, A, D, F \rangle$$

where $U = \{U_i;\ i = 1, \ldots, T\}$ is the set of antecedent attributes, with each of them taking values (or propositions) from an array of finite sets $A = \{A_1, A_2, \ldots, A_T\}$. $A_i = \{A_{ij};\ j = 1, \ldots,\ J_i = |A_i|\}$ is a referential set of values (or propositions) for an attribute U_i $(i = 1, \ldots, T)$, and the values or propositions in A_i (e.g., A_{ij}) are referred to as referential values, which can be taken in different types of value, as mentioned in Section II-A. The array $\{U_1, U_2, \ldots, U_T\}$ defines a list of finite conditions, representing the elementary states of a problem domain, which may be linked by "$\wedge$" or "$\vee$" connectives. Note that "$\wedge$" is a logical connective to represent the "AND" relationship and "$\vee$" a logical connective to represent the "OR" relationship. $D = \{D_n;\ n = 1, \ldots, N\}$ is the set of all consequents, which can either be conclusions or actions. F is a logical function, reflecting the relationship between conditions and their associated conclusions.

Note that in a rule base, a referential set can be a set of meaningful and distinctive evaluation standards for describing an attribute by subjective linguistic terms. To estimate failure likelihood in safety analysis, for example, one may use such linguistic terms as "highly frequent," "frequent," "average," "low," and "very low." These linguistic terms are the referential values for an antecedent attribute "failure likelihood." In a general rule base, a referential set may be different in type.

To establish a rule base, one has to determine which referential set for each antecedent attribute needs to be used and how many referential values should be used. More specifically, the kth rule in a rule base in forms of a conjunctive "if–then" rule can be written as

$$R_k:\ \text{if } A_1^k \wedge A_2^k \wedge \cdots \wedge A_{T_k}^k,\ \text{then } D_k \tag{1}$$

where A_i^k $(i = 1, \ldots, T_k)$ is a referential value of the ith antecedent attribute in the kth rule, and T_k is the number of the antecedent attributes used in the kth rule. D_k $(\in D)$ is the consequent in the kth rule. For example, the following is an if–then rule for safety analysis [20].

If failure rate of a hazard is frequent and consequence severity is catastrophic and failure consequence probability is likely, then safety estimate is poor.

The linguistic terms frequent, catastrophic, and likely are the referential values of the attributes failure rate, consequence severity, and failure consequence probability, respectively. Poor is the consequent of the rule corresponding to the output attribute safety estimate.

A basic rule base is composed of a collection of such simple "if–then" rules. In the next section, more complicated rules are taken into consideration.

D. Additional Knowledge Representation Parameters

1) Relative Importance of an Antecedent Attribute (Attribute Weight): The relative importance of an attribute to its consequent (attribute weight) plays an important role in rule base inference. For example, in medical diagnostic systems, a particular symptom combined with other symptoms may lead to a possible disease. It is important to assign a weight to each symptom (attribute) in order to show the relative importance of each symptom to the consequent (disease).

2) Relative Importance of a Rule (Rule Weight): A relative weight can also be assigned to a rule, which is used to represent the relative importance of the rule to the associated conclusions. The weight of a rule can also take into account the degree to which the rule is activated, which is related to input information.

Weights can be assigned at the knowledge acquisition phrase when a rule base is established. Apart from the simple scaling methods, pairwise comparisons may also be used to estimate relative weights of attributes or rules in several approaches, including the eigenvector method [45], the geometric least squares method [46], and the geometric mean method [47].

III. New Generic Rule-Base Inference Methodology

A. Generic Rule-Base Structure and Representation Schema

To take into account belief degrees, attribute weights, and rule weights in a rule, the kth rule given in (1) is extended as

$$\begin{aligned} R_k:\ & \text{if } A_1^k \wedge A_2^k \wedge \cdots \wedge A_{T_k}^k \text{ then, } D_k \\ & \text{with a belief degree } \overline{\beta}_k,\ \text{a rule weight } \theta_k \\ & \text{and attribute weights } \delta_{k1}, \delta_{k2}, \ldots, \delta_{kT_k} \end{aligned} \tag{2}$$

where A_i^k $(i = 1, \ldots, T_k)$ is the referential value of the ith antecedent attribute in the kth rule, T_k the number of antecedent attributes used in the kth rule, and $\overline{\beta}_k$ the belief degree to which D_k $(\in D)$ is believed to be the consequent, given "$A_1^k \wedge A_2^k \wedge \cdots \wedge A_{T_k}^k$" in the kth rule. θ_k is the relative weight of the kth rule, and $\delta_{k1}, \delta_{k2}, \ldots, \delta_{kT_k}$ are the relative weights of the T_k antecedent attributes used in the kth rule.

Rule (2) can be further extended to a so-called packet rule using a belief structure, where all possible consequents are associated with belief degrees. A collection of packet rules constitute a rule base with a belief structure (called a belief rule base) as

$$\begin{aligned} R_k:\ & \text{if } A_1^k \wedge A_2^k \wedge \cdots \wedge A_{T_k}^k, \text{ then} \\ & \left\{(D_1, \overline{\beta}_{1k}), (D_2, \overline{\beta}_{2k}), \ldots, (D_N, \overline{\beta}_{Nk})\right\} \\ & \left(\sum_{i=1}^{N} \overline{\beta}_{ik} \le 1\right),\ \text{with a rule weight } \theta_k \\ & \text{and attribute weights } \delta_{k1}, \delta_{k1}, \ldots, \delta_{kT_k} \\ & k \in \{1, \ldots, L\} \end{aligned} \tag{3}$$

where $\overline{\beta}_{ik}$ $(i \in \{1, \ldots, N\}$ is the belief degree to which D_i is believed to be the consequent if, in the kth packet rule, the input satisfies the packet antecedents $A^k = \{A_1^k, A_2^k, \ldots, A_{T_k}^k\}$. L is the number of all packet rules in the rule base. If $\sum_{i=1}^{N} \overline{\beta}_{ik} = 1$, the kth packet rule is said to be complete; otherwise, it is incomplete. Note that $\sum_{i=1}^{N} \overline{\beta}_{ik} = 0$ denotes total ignorance about the output, given the input in the kth packet rule. Rule (3) is referred to as a belief rule in the paper.

Take, for example, the following belief rule in safety analysis:

$$\begin{aligned} R_k:\ & \text{if the failure rate is frequent and the} \\ & \text{consequence severity is critical and the failure} \\ & \text{consequence probability is unlikely,} \\ & \text{then the safety estimate is} \\ & \{(\text{good}, 0),\ (\text{average}, 0),\ (\text{fair}, 0.7),\ (\text{poor}, 0.3)\} \end{aligned}$$

where {(good, 0), (average, 0), (fair, 0.7), (poor, 0.3)} is a belief distribution representation for safety consequent, stating that it is 70% sure that safety level is fair and 30% sure that safety level is poor. In this belief rule, the total degree of belief is $0.3 + 0.7 = 1$, so that the assessment is complete. The referential value set for failure rate is given by $A_{\mathrm{FR}} =$ {very low, low, reasonably low, average, reasonably frequent, frequent, and highly frequent}.

Remark 1: Antecedent attributes or the number of attributes is not required to be the same from one rule to another, even though they share a common consequent set $D = \{D_n;\ n = 1, \ldots, N\}$.

Remark 2: Rules (1) and (2) are the special cases of rule (3), with $\{\overline{\beta}_{1k}, \overline{\beta}_{2k}, \ldots, \overline{\beta}_{Nk}\}$ being given special values. In fact, if $\overline{\beta}_{ik} = 1$, $\overline{\beta}_{jk} = 0$, $j \neq i$, $j = 1, \ldots, N$ are applied in rule (3), then rule (1) can be established; if $\overline{\beta}_{ik} = \overline{\beta}_k$, $\overline{\beta}_{jk} = 0$, $j \neq i$, $j = 1, \ldots, N$ are used, then rule (2) will be derived.

A belief rule base given in the form shown in (3) represents functional mappings between antecedents and consequents with uncertainty. It provides a more informative and realistic scheme for uncertain knowledge representation. Note that the degrees of belief $\overline{\beta}_{ik}$ could be assigned directly by experts, or more generally, they may be trained and updated using dedicated learning algorithms if *prior* or up-to-date information regarding the inputs and outputs of a rule-based system is available. Once such a belief rule base is established, the knowledge contained in the belief rule base can be used to perform inference for given inputs. The inference procedure is investigated in the following subsections.

B. Input Transformation

Before an inference process can start, the relationship between an input (fact) and each referential value in the antecedents of a rule needs to be determined so that an activation weight for each rule can be generated. The basic idea is to examine all the referential values of each attribute in order to determine a matching degree to which an input belongs to a referential value. This is equivalent to transforming an input into a distribution on referential values using belief degrees [17]. Once the matching degrees between an input and the referential values of all antecedents in a rule are determined, they are processed to generate an activation weight for the rule, which is used to measure the degree to which the packet antecedent of the rule is activated by the input.

In a belief rule base shown in (3), a general input form corresponding to all antecedent attributes is given as

$$(A_1^*, \varepsilon_1) \wedge (A_2^*, \varepsilon_2) \wedge \cdots \wedge (A_T^*, \varepsilon_T) \tag{4}$$

where ε_i expresses the degree of belief assigned to the input value A_i^* of the ith attribute $(i = 1, \ldots, T)$, which reflects the uncertainty of the input data, and T is the total number of different antecedent attributes involved in all the rules in a rule base. For example, (red, 0.9) means that there is a 90% certainty that the input can take the value "red" of the attribute "color."

To facilitate data collection, it is desirable to acquire assessment information in a manner appropriate to a particular attribute. By using the distribution assessment approach [17], a referential value of an attribute may in general be regarded as an evaluation grade, and the input (A_i^*, ε_i) for the ith attribute can be transformed to a distribution on the referential values of the attribute using belief degrees as

$$S(A_i^*, \varepsilon_i) = \{(A_{ij}, \alpha_{ij});\ j = 1, \ldots, J_i\}, \quad i = 1, 2, \ldots, T \tag{5}$$

where A_{ij} is the jth referential value of the ith attribute, α_{ij} the degree to which the input A_i^* belongs to the referential value A_{ij} with $\alpha_{ij} \geq 0$ and $\sum_{j=1}^{J_i} \alpha_{ij} \leq 1$ $(i = 1, 2, \ldots, T)$, and J_i the number of the referential values used for describing the ith antecedent attribute. The distributed assessment above reads that the input A_i^* is assessed to the referential value A_{ij} with the degree of belief of α_{ij} $(j = 1, 2, \ldots, J_i$ and $i = 1, 2, \ldots, T)$. An assessment $S(A_i^*, \varepsilon_i)$ is complete if $\sum_{j=1}^{J_i} \alpha_{ij} = 1$ and incomplete if $\sum_{j=1}^{J_i} \alpha_{ij} < 1$. A special case is $\sum_{j=1}^{J_i} \alpha_{ij} = 0$ (or $\alpha_{ij} = 0$ for all $j = 1, 2, \ldots, J_i$), which means that the input does not affect the ith attribute at all. Such incompleteness will be handled in the inference procedure of the rule base using the ER approach. α_{ij} could be generated using various ways, depending on the nature of an antecedent attribute, which will be investigated in Section IV.

For instance, in evaluation of qualitative antecedent attributes, subject judgments could be used. In assessment of the brakes of a motorcycle (if its referential set is {poor, indifferent, average, good, excellent}), for example, assessors may be

> "30% sure that its braking stability is at the average level and 70% sure that it is good."

Hence, $S(\text{brake}) = \{$(poor, 0), (indifferent, 0), (average, 0.3), (good, 0.7), (excellent, 0)$\}$.

Given the input for the packet antecedent A^k in the kth rule, denoted by A^{*k}, the corresponding activation weight for each rule, denoted by w_k, can be generated as discussed in the following subsection.

C. Activation Weight for the Packet Antecedent of a Rule

Consider an input given in a format shown in (5) corresponding to the kth rule defined as in (3)

$$\left(A_1^k, \alpha_1^k\right) \wedge \left(A_2^k, \alpha_2^k\right) \wedge \cdots \wedge \left(A_{T_k}^k, \alpha_{T_k}^k\right)$$

where $A_i^k \in \{A_{ij};\quad j = 1, \ldots, J_i\}$ and $\alpha_i^k \in \{\alpha_{ij};\quad j = 1, \ldots, J_i\}$. For example, an input in safety analysis transformed into the kth rule is given by [20].

> "Failure rate is (frequent, 0.45) and consequence severity is (critical, 0.8) and failure consequence probability is (unlikely, 0.2)."

The total degree α_k to which the input matches the packet antecedent A^k in the kth rule can be calculated using the following formula:

$$\alpha_k = \varphi\left(\left(\delta_{k1}, \alpha_1^k\right), \ldots, \left(\delta_{kT_k}, \alpha_{T_k}^k\right)\right). \tag{6}$$

Here, φ is an aggregation function that reflects the relationship among the T_k antecedents in the kth rule, and δ_{ki} $(i = 1, \ldots, T_k)$ is the relative weight of the ith antecedent attribute in the kth rule.

Care should be taken in selection of the aggregation function φ. Considering that α_k is a belief degree (or a subjective probability), the following pair of aggregation functions may be used for generating such subjective probabilities, which are referred to as the "probability-product" and "probability-sum" operators for the "$\wedge$" and "$\vee$" connectives in a rule-based system, i.e., $\varphi_{\text{product}}(a, b) = ab$ and $\varphi_{\text{sum}}(a, b) = a + b - ab$, respectively. This pair is also referred to as the triangular norm (t-norm) operator and the triangular conorm (t-conorm) operators [48], [49].

Suppose the "$\wedge$" connective is used for all antecedents in a rule, such as "if $A \wedge B \wedge C$." In other words, the consequent of a rule is not believed to be true unless all the antecedents of the rule are activated. In such cases, one may use the following simple weighted multiplicative aggregation function to calculate α_k:

$$\alpha_k = \prod_{i=1}^{T_k} \left(\alpha_i^k\right)^{\overline{\delta}_{ki}} \tag{6a}$$

where

$$\overline{\delta}_{ki} = \frac{\delta_{ki}}{\max\limits_{i=1,\ldots,T_k}\{\delta_{ki}\}} \text{ so that } 0 \leq \overline{\delta}_{ki} \leq 1.$$

The above aggregation function is used in the case study in Section V. Note that $0 \leq \alpha_k \leq 1$, $\alpha_k = 1$ if $\alpha_i^k = 1$ for all $i = 1, \ldots T_k$, and $\alpha_k = 0$ if $\alpha_i^k = 0$ for any $i = 1, \ldots T_k$. Furthermore, the contribution of an antecedent attribute towards α_k is positively related to the weight of the attribute. In other words, an important attribute plays a greater role in determining α_k, which is further explained as follows:

1) If $\overline{\delta}_{kl} = 0$ $(l \in \{1, \ldots T_k\})$, then $(\alpha_l^k)^{\overline{\delta}_{kl}} = 1$, which shows that an attribute with zero importance does not

have any impact on the aggregation process; if $\overline{\delta}_{kl} = 1$, then $(\alpha_l^k)^{\overline{\delta}_{kl}} = \alpha_l^k$, which shows that the most important antecedent has significant impact on the aggregation process proportional to the degree to which it is matched by the input.

2) Note that if $a > b$, then $(a)^\delta \geq (b)^\delta$ for $\delta > 0$, which means that the function $(\alpha)^\delta$ is monotonically nondecreasing in the argument α if $\delta > 0$. In other words, if the individual belief with regard to the antecedent attribute is increased, the overall belief should not decrease.

3) Furthermore, $(\alpha)^\delta$ is nonincreasing in δ for $0 \leq \alpha \leq 1$. In other words, if $\delta_1 < \delta_2$, then $(\alpha)^{\delta_1} \geq (\alpha)^{\delta_2}$ for $0 \leq \alpha \leq 1$. Therefore, as the ith antecedent attribute in the kth rule becomes less important, δ_{ki} decreases, which will decrease the possibility of α_k being dominated by α_i^k.

If the "$\vee$" connective is used for all antecedents in a rule, such as "if $A \vee B \vee C$", then one may use the following recursively defined weighted product–sum aggregation function to calculate α_k:

$$\begin{aligned} \alpha_{k(1)} &= h_1^k \\ &= \overline{\delta}_{k1} \times \alpha_1^k \\ \alpha_{k(i+1)} &= \alpha_{k(i)} + \left(1 - \alpha_{k(i)}\right) h_{i+1}^k \quad \text{for } i = 1, \ldots, T_k - 1 \\ \alpha_k &= \alpha_{k(T_k)} \end{aligned} \tag{6b}$$

where $h_j^k = \overline{\delta}_{kj} \times \alpha_j^k$, $j = 1, \ldots, T_k$. Note that $0 \leq \alpha_k \leq 1$, $\alpha_k = 1$ if $\alpha_i^k = 1$ and the ith antecedent attribute is the most important attribute in the kth rule, and $\alpha_k = 0$ if $\alpha_i^k = 0$ for all $i = 1, \ldots T_k$. In addition, the contribution of an antecedent attribute towards α_k is positively related to the weight of the attribute. In other words, an important attribute plays a greater role in determining α_k based on the definition of h_j^k, $j = 1, \ldots, T_k$, which can be explained in the similar way as for (6a).

However, the antecedents of a rule could be more complicated. The combination of conjunction and disjunction, like $(A \vee B \vee C) \wedge (A \vee B)$, is common in rule-based systems. On the other hand, the meanings of the relationships "AND" and "OR" are not fixed. In some cases, they correspond to disjunction and conjunction, and in other cases, their meanings may be reversed [50]. It is also possible that they do not correspond to either disjunction or conjunction, in particular, when various criteria support each other and trigger a rule collectively [51], or when the criteria display conjunctive and disjunctive behavior simultaneously [49], or when aggregation depends on the values of the criteria [52]. Consequently, it may not be feasible to predefine the universal form of a general aggregation function φ and use it throughout a rule base. Instead, each rule should be examined individually, and an aggregation function appropriate for each rule should be used [53].

A rule may use both the "$\vee$" connective and the "$\wedge$" connective. Such a rule can be decomposed into a collection of equivalent rules, using standard techniques from crisp logic. Suppose, for example, a rule is expressed as

$$R_k: \text{if } \left(A_1^k \wedge A_2^k \wedge \cdots \wedge A_m^k\right) \vee \left(A_{m+1}^k \wedge \cdots \wedge A_{T_k}^k\right), \text{ then } D.$$

This rule can be expressed as in the following two equivalent rules

$$R_k^{(1)}: \text{if } A_1^k \wedge A_2^k \wedge \cdots \wedge A_m^k, \text{ then } D$$

$$\text{and } R_k^{(2)}: \text{if } A_{m+1}^k \wedge \cdots \wedge A_{T_k}^k, \text{ then } D.$$

Then, aggregation functions similar to (6a) could be used for each of the two rules.

The activation weight w_k of the packet antecedent A^k in the kth rule is generated by weighting and normalizing the matching degree α_k given by (6) as

$$w_k = \frac{\theta_k \alpha_k}{\sum_{i=1}^{L} \theta_i \alpha_i} \tag{7}$$

where θ_k is the relative weight of the kth rule. Note that $0 \leq w_k \leq 1$ $(k = 1, \ldots, L)$, and $\sum_{i=1}^{L} w_i = 1$. Note also that $w_k = 0$ if the kth rule is not activated.

D. Degrees of Belief in the Consequent of a Rule

The degree of belief $\overline{\beta}_{ik}$ in the ith possible consequent term D_i in the kth rule is already given when a rule base is established with $0 \leq \sum_{i=1}^{N} \overline{\beta}_{ik} \leq 1$. The kth rule is complete if $\sum_{i=1}^{N} \overline{\beta}_{ik} = 1$, and it is incomplete if $\sum_{i=1}^{N} \overline{\beta}_{ik} < 1$. The incompleteness of the consequent of a rule can also result from its antecedents due to lack of data. For instance, the input for an antecedent attribute may not be available or may be only partially known. In the inference process, such incompleteness should be considered. An incomplete input for an attribute will lead to an incomplete output in each of the rules in which the attribute is used. The original belief degree $\overline{\beta}_{ik}$ in the ith consequent D_i of the kth rule is updated based on the actual input information as

$$\beta_{ik} = \overline{\beta}_{ik} \frac{\sum_{t=1}^{T_k} \left(\tau(t,k) \sum_{j=1}^{J_t} \alpha_{tj} \right)}{\sum_{t=1}^{T_k} \tau(t,k)} \tag{8}$$

where

$$\tau(t,k) = \begin{cases} 1, & \textit{if } U_t \text{ is used in defining } R_k \ (t = 1, \ldots, T_k) \\ 0, & \text{otherwise.} \end{cases}$$

α_{tj} is given in (5) with $\alpha_{tj} \geq 0$, and $\sum_{j=1}^{J_t} \alpha_{tj} \leq 1$. $\overline{\beta}_{ik}$ is given in (3). Note that $0 \leq \sum_{i=1}^{N} \beta_{ik} \leq 1$ for all k and $(1 - \sum_{i=1}^{N} \beta_{ik})$ represents both ignorance incurred in establishing R_k and the incompleteness that may exist in the input.

To demonstrate (8), suppose there are three antecedent attributes used to determine the output of a rule R_k. It is possible that an antecedent attribute may not be used in defining R_k. For example, suppose $T = 3$ in (4) and that the third attribute is not used to define the rule R_k, which is given by

$$R_k: \text{ if } A_1^k \wedge A_2^k, \text{ then } \left\{ (D_1, \overline{\beta}_{1k}), \ldots, (D_N, \overline{\beta}_{Nk}) \right\}.$$

Then, $\tau(1,k) = 1$, $\tau(2,k) = 1$, and $\tau(3,k) = 0$. Hence

$$\beta_{ik} = \overline{\beta}_{ik} \frac{\sum_{t=1}^{3} \left(\tau(t,k) \sum_{j=1}^{J_t} \alpha_{tj} \right)}{\sum_{t=1}^{3} \tau(t,k)} = \overline{\beta}_{ik} \frac{\sum_{j=1}^{J_1} \alpha_{1j} + \sum_{j=1}^{J_2} \alpha_{2j}}{2}.$$

If the input for the first two attributes is both complete, or $\sum_{j=1}^{J_1} \alpha_{1j} = 1$ and $\sum_{j=1}^{J_2} \alpha_{2j} = 1$, then

$$\beta_{ik} = \overline{\beta}_{ik} * \frac{2}{2} = \bar{\beta}_{ik}.$$

Suppose all the three attributes are used to define the rule R_k, but the input of an attribute is incomplete. For example, the input for the second antecedent is incomplete, and that for the first and the third is complete, e.g., $\sum_{j=1}^{J_2} \alpha_{2j} = 0.9 < 1$, and $\sum_{j=1}^{J_1} \alpha_{1j} = \sum_{j=1}^{J_3} \alpha_{3j} = 1$. Then

$$\beta_{ik} = \overline{\beta}_{ik} \frac{\sum_{t=1}^{3} \left(\tau(t,k) \sum_{j=1}^{J_t} \alpha_{tj} \right)}{\sum_{t=1}^{3} \tau(t,k)} = \overline{\beta}_{ik} * \frac{2.9}{3}.$$

Hence, the incompleteness in the input for the second attribute will be reflected in β_{ik}. Furthermore, if the input for the second attribute is unknown, or $\sum_{j=1}^{J_2} \alpha_{2j_2} = 0$, then $\beta_{ik} = \overline{\beta}_{ik} * (2/3)$.

E. Belief Rule Expression Matrix for a Belief Rule Base

Suppose a belief rule base is given by $R = \{R_1, R_2, \ldots, R_L\}$. The kth rule can be represented as

$$R_k: \text{ if } U \text{ is } A^k, \text{ then } D \text{ with belief degree } \beta^k \qquad (9)$$

where U represents the antecedent attribute vector $(U_1, U_2, \ldots, U_{T_k})$, A^k the packet antecedents $\{A_1^k, A_2^k, \ldots, A_{T_k}^k\}$, D the consequent vector $(D_1, D_2, \ldots, D_N)$, and β^k the vector of the belief degrees $(\beta_{1k}, \beta_{2k}, \ldots, \beta_{Nk})$ for $k \in \{1, \ldots, L\}$. This is the vector form of a belief rule, which can be explained as follows.

The packet antecedent A^k of a belief if–then rule can be considered as a global attribute, which is considered to be assessed to a consequent D_i with a belief degree of $\beta_{ik} (i \in \{1, \ldots, N\})$. This assessment can be represented by

$$S(A^k) = \{(D_i, \beta_{ik});\ i = 1, \ldots, N\} \qquad (10)$$

which is a distribution assessment and referred to as a belief structure, where β_{ik} measures the degree to which D_i is the consequent if the input activates the antecedent A^k in the kth rule for $i = 1, \ldots, N$, $k = 1, \ldots, L$. L is the number of rules in the rule base, and N is the number of possible consequents. If $\sum_{i=1}^{N} \beta_{ik} = 1$, then the kth rule is said to be complete; if $\sum_{i=1}^{N} \beta_{ik} = 1$ for all $k = 1, \ldots, L$, then the rule base is a complete rule base; otherwise, it is incomplete. Note that $\sum_{i=1}^{N} \beta_{ik} = 0$ denotes total ignorance about the output of the kth rule. Here, it is assumed that all the L rules are independent of each other, which means that the packet antecedents $A^1, \ldots, A^L$ are independent of each other.

TABLE I
BELIEF RULE EXPRESSION MATRIX FOR A RULE BASE

Output	Input					
	$A^1(w_1)$	$A^2(w_2)$	…	$A^k(w_k)$	…	$A^L(w_L)$
D_1	β_{11}	β_{12}	…	β_{1k}	…	β_{1L}
D_2	β_{21}	β_{22}		β_{2k}		β_{2L}
⋮	⋮	⋮	…	⋮	…	⋮
D_i	β_{i1}	β_{i2}	…	β_{ik}	…	β_{iL}
⋮	⋮	⋮	…	⋮	…	⋮
D_N	β_{N1}	β_{N2}	…	β_{Nk}	…	β_{NL}

A belief rule base established using belief rules given by (10) can be summarized using a belief rule expression matrix shown in Table I.

In the matrix, w_k is the activation weight of A^k, which measures the degree to which the kth rule is weighted and activated.

In the following section, the inference procedure is implemented in order to combine all rules for generating the final belief degrees for $D_1, \ldots, D_N$ based on the rule expression matrix.

F. Rule Inference Using the Evidential Reasoning Approach

Based on the above belief rule expression matrix, the ER approach can be used to combine rules and generate final conclusions.

Having represented each rule by using (10), the ER approach can be directly applied as follows. First, transform the degrees of belief β_{jk} for all $j = 1, \ldots, N$, $k = 1, \ldots, L$ into basic probability masses using the following ER algorithm [17], [18]:

$$m_{j,k} = w_k \beta_{j,k},\ j = 1, \cdots, N$$

$$m_{D,k} = 1 - \sum_{j=1}^{N} m_{j,k} = 1 - w_k \sum_{j=1}^{N} \beta_{j,k}$$

$$\overline{m}_{D,k} = 1 - w_k$$

$$\widetilde{m}_{D,k} = w_k \left(1 - \sum_{j=1}^{N} \beta_{j,k} \right)$$

for all $k = 1, \cdots, L$, where $m_{D,k} = \overline{m}_{D,k} + \widetilde{m}_{D,k}$ for all $k = 1, \cdots, L$ and $\sum_j^L w_j = 1$. The probability mass assigned to the consequent set D, which is unassigned to any individual consequent, is split into two parts: one caused by the relative importance of the kth packet antecedent A^k (or $\overline{m}_{D,k}$) and the other by the incompleteness of the kth packet antecedent A^k (or $\widetilde{m}_{D,k}$).

Then, aggregate all the packet antecedents of the L rules to generate the combined degree of belief in each possible consequent D_j in D. Suppose $m_{j,I(k)}$ is the combined degree of belief in D_j by aggregating the first k packet antecedents

$(A^1, \ldots, A^k)$, and $m_{D,I(k)}$ is the remaining degree of belief unassigned to any consequent. Let $m_{j,I(1)} = m_{j,1}$ and $m_{D,I(1)} = m_{D,1}$. Then, the overall combined degree of belief β_j in D_j is generated as

$$\{D_j\}: \; m_{j,I(k+1)} = K_{I(k+1)} \left[m_{j,I(k)} m_{j,k+1} + m_{j,I(k)} \times m_{D,k+1} + m_{D,I(k)} m_{j,k+1}\right]$$

$$m_{D,I(k)} = \overline{m}_{D,I(k)} + \widetilde{m}_{D,I(k)}, \quad k = 1, \cdots, L$$

$$\{D\}: \; \widetilde{m}_{D,I(k+1)} = K_{I(k+1)} \left[\widetilde{m}_{D,I(k)} \widetilde{m}_{D,k+1} + \widetilde{m}_{D,I(k)} \times \overline{m}_{D,k+1} + \overline{m}_{D,I(k)} \widetilde{m}_{D,k+1}\right]$$

$$\{D\}: \; \overline{m}_{D,I(k+1)} = K_{I(k+1)} \left[\overline{m}_{D,I(k)} \overline{m}_{D,k+1}\right]$$

$$K_{I(k+1)} = \left[1 - \sum_{j=1}^{N} \sum_{\substack{t=1 \\ t \neq j}}^{N} m_{j,I(k)} m_{t,k+1}\right]^{-1}$$

$$k = 1, \cdots, L-1$$

$$\{D_j\}: \; \beta_j = \frac{m_{j,I(L)}}{1 - \overline{m}_{D,I(L)}}, \quad j = 1, \cdots, N$$

$$\{D\}: \; \beta_D = \frac{\widetilde{m}_{D,I(L)}}{1 - \overline{m}_{D,I(L)}}.$$

β_D represents the remaining belief degrees unassigned to any D_j. It has been proven that $\sum_{j=1}^{N} \beta_j + \beta_D = 1$ [18]. The final conclusion generated by aggregating the L rules, which are activated by the actual input vector $A^* = \{A^{*k}, k = 1, \ldots, L\}$, can be represented as

$$S(A^*) = \{(D_j, \beta_j); \; j = 1, \ldots, N\}. \tag{11}$$

Suppose the utility (or score) of an individual consequent D_j is denoted by $u(D_j)$. The expected utility (or score) of $S(A^*)$ is given as [21]

$$u\left(S(A^*)\right) = \sum_{j=1}^{N} u(D_j)\beta_j. \tag{12}$$

Note that β_j denotes the lower bound of the likelihood that the output is assessed to D_j. The upper bound of the likelihood is given by $(\beta_j + \beta_D)$. Complementary to the distribution assessment as shown by (11), a utility interval can also be established [17] if the assessment is incomplete or imprecise, characterized by the maximum, minimum, and average utilities of $S(A^*)$ defined as follows, provided that $u(D_{n+1}) \geq u(D_n)$:

$$u_{\max}\left(S(A^*)\right) = \sum_{j=1}^{N-1} \beta_j u(D_j) + (\beta_N + \beta_D) u(D_N) \tag{12a}$$

$$u_{\min}\left(S(A^*)\right) = (\beta_1 + \beta_D) u(D_1) + \sum_{j=2}^{N} \beta_j u(D_j) \tag{12b}$$

$$u_{\text{avg}}\left(S(A^*)\right) = \frac{u_{\max}\left(S(A^*)\right) + u_{\min}\left(S(A^*)\right)}{2}. \tag{12c}$$

These utilities are used for characterizing an assessment but not used in the aggregation process. Note that if the assessment of $S(A^*)$ is complete, then $\beta_D = 0$, and $u(S(A^*)) = u_{\max}(S(A^*)) = u_{\min}(S(A^*)) = u_{\text{avg}}(S(A^*))$.

The logic behind the approach is that if the output in the kth rule includes D_i and the kth rule is activated, then the overall output must be D_i to a certain degree. The degree is measured by both the degree to which the kth rule is important to the overall output and the degree to which the antecedents of the kth rule are activated by the actual input.

The computational complexity of reasoning within the Dempster's rule of combination could be one of the major points of criticism if the combination rule is not used properly. In fact, Orponen [54] showed that the combination of mass functions or basic probability assignments (bpas) using Dempster's rule is #P-complete (the class #P is a functional analogue of the class NP of decision problems), but the computational complexity of reasoning using Dempster's rule based on the specific ER decision analysis framework above becomes linear rather than #P-complete [15]–[17]. It should also be noted that conflicting information can be explicitly modeled using the framework shown in Table I using the normalized activation weight w_k and can be logically processed using the ER algorithm described earlier in this subsection, thereby overcoming another drawback of the Dempster's original combination rule in dealing with conflicting evidence.

IV. Extension to Hierarchical Knowledge Base

As mentioned in Section II, attributes involved in the premise of a rule could be quantitative or qualitative, so that input data may be different both in type and in scale. Hence, there is a need to establish a framework that provides a basis for multidimensional information synthesis. In this section, a scheme for handling various types of input information will be investigated first, and the basic knowledge-base structure discussed in the previous section will be extended to a hierarchical structure.

A. Transformation of Various Types of Input Information

Transformation of input variables could be implemented using various ways, depending on the nature of an antecedent attribute. The following cases are discussed.

1) Quantitative Attribute Described Using Linguistic Terms:

a) Transformation based on fuzzy linguistic values using fuzzy membership functions: An antecedent attribute can be described by fuzzy linguistic values characterized using fuzzy membership functions. In this case, the referential values of an antecedent attribute are fuzzy linguistic values. An input may be uncertain and can be obtained from historical data and expert's experiences using the following numerical forms to suit conditions under study:

1) a single deterministic value with 100% certainty;
2) a closed interval defined by an equally likely range;
3) a triangular distribution defined by a most likely value, with lower and upper least likely values;
4) trapezoidal distribution defined by a most likely range, with lower and upper least likely values.

Suppose the ith antecedent attribute is a quantitative attribute described using linguistic terms. To get the belief degree α_{ij} as required in (5), the first step is to determine the degree to which an input matches each of the appropriate fuzzy sets via membership functions. In general, $S(A_i^*, \varepsilon_i)$ of (5) can be calculated using

$$S\left(A_i^*, \varepsilon_i\right) = s\left\{(A_{ij}, \alpha_{ij});\ j = 1, \ldots, J_i\right\},\quad i \in \{1, 2, \ldots, T\} \tag{13}$$

where

$$\alpha_{ij} = \frac{\tau\left(A_i^*, A_{ij}\right)\varepsilon_i}{\sum_{j=1}^{J_i}\left[\tau\left(A_i^*, A_{ij}\right)\right]}$$

where τ is a matching function, and $\tau_i = \tau(A_i^*, A_{ij})$ is called a matching degree. Note that $1 \geq \alpha_{ij} \geq 0$ and $\sum_{j=1}^{J_i} \alpha_{ij} \leq 1$. If A_i^* completely belongs to the jth linguistic expression, i.e., $\tau(A_i^*, A_{ij}) = 1$, then α_{ij} may not be equal to 1 due to ε_i.

In summary, $S(A_i^*, \varepsilon_i)$ is determined using the following three steps:

1) Determining the matching degree τ_i between A_i^* and the jth expression of the ith antecedent attribute by using a matching function, $j = 1, \ldots, J_i$.
2) Normalizing the matching degree τ_i.
3) Generating α_{ij} by combining the normalized matching degree with the belief degree ε_i.

One possible matching function τ for a quantitative attribute is given as follows. If an input A^* is one of the above four numerical forms (which can be considered as the special fuzzy sets) and A is a referential value characterized using fuzzy membership functions, then the matching degree between A^* and A is defined by

$$\tau(A^*, A) = \max_x\left[\min\left(A^*(x), A(x)\right)\right] \tag{14}$$

where x covers the domain of the input A^*. $A(x)$ [or $A^*(x)$] represents the membership function of A (or A^*). If A_i^* is a crisp value x_0, then $\tau(A^*, A) = A(x_0)$. Otherwise, $\tau(A^*, A)$ is the highest point of the intersection of the input fuzzy set A^* and the fuzzy set A.

The max–min operation is used for illustration purpose because it is a classical tool to set the matching degree between fuzzy sets, and it is shown as a similarity measure between the input fuzzy set A^* and the fuzzy antecedent A [49]. However, other functions could also be chosen, such as the t-conorm function and the t-norm function [49]. Alternatively, various similarity measures based on the geometric model [55], the set-theoretic approach and the matching function [56], [57], the best-fit method for discrete membership functions [58], and the grade matching method for continuous membership functions [14] may be used.

b) Rule or utility-based equivalence transformation techniques for quantitative data: For a quantitative attribute described by linguistic values, another way to generate α_{ij} in (5) is to use the rule or utility-based equivalence transformation techniques [17], which is outlined as follows.

Suppose the input of a quantitative antecedent attribute is given by numerical values. In this case, equivalence rules first need to be extracted from the decision maker to transform a value to an equivalent expectation, thereby relating a particular value to a set of referential values.

In general, suppose $U = \{U_i;\ i = 1, \ldots, T\}$ is the set of antecedent attributes, $A_i = \{A_{ij};\ j = 1, \ldots, J_i = |A_i|\}$ ($i = 1, \ldots, T$) is the referential set of the antecedent attribute U_i, and a value a_{ij} for an antecedent attribute U_i is judged to be equivalent to a referential value $A_{ij}(j = 1, \ldots, J_i)$, or

$$a_{ij} \text{ means } A_{ij}\ (j = 1, \ldots, J_i). \tag{15}$$

Then, it needs to represent the numerical value using an equivalent expectation.

Without loss of generality, suppose U_i is a "profit" attribute, that is, a large value $a_{i(j+1)}$ is preferred over a smaller value a_{ij}. Let a_{iJ_i} be the largest feasible value and a_{i1} the smallest. Then, an input value A_i^* for U_i may be represented using the following equivalent expectation:

$$S\left(A_i^*\right) = \{(a_{ij}, \gamma_{ij});\ j = 1, \ldots, J_i\} \tag{15a}$$

where

$$\gamma_{ij} = \frac{a_{i(j+1)} - A_i^*}{a_{i(j+1)} - a_{ij}}$$

$$\gamma_{i(j+1)} = 1 - \gamma_{ij} \quad \text{if} \quad a_{ij} \leq A_i^* \leq a_{i(j+1)} \tag{16a}$$

$$\gamma_{ik} = 0 \quad \text{for} \quad k = 1, \ldots, J_i,\ k \neq j,\ j+1. \tag{16b}$$

Note that representing A_i^* by (15a) means that A_i^* is calculated by the expected value of $S(A_i^*)$ denoted by $E(S(A_i^*))$, or $A_i^* = E(S(A_i^*))$, and the utility of A_i^* is calculated by $u(S(A_i^*))$. It has been proven that the above equivalence relations are justified if the underlying implicit utility function of the attribute U_i is assumed to be piecewise linear [17]. The use of a piecewise linear utility function is of general interest and wide applicability as any nonlinear function may be approximated by a piecewise linear function if a sufficient number of values of the nonlinear function are estimated [59], [60].

Given the equivalence rules described in (15), a value A_i^* can be represented by the following equivalence expectation:

$$S\left(A_i^*\right) = \{(A_{ij}, \alpha_{ij});\ j = 1, \ldots, J_i\} \tag{17}$$

where

$$\alpha_{ij} = \gamma_{ij}, \quad j = 1, \ldots, J_i. \tag{18}$$

Thus, given the equivalence rules between a_{ij} and A_{ij} and assuming a piecewise linear utility function for U_i, a numerical value is represented by an equivalent distribution using (16)–(18).

A quantitative antecedent attribute may also be a random variable and may not always take a single value but several values with different probabilities. To assess such an attribute U_i, the following distribution is used [17]:

$$S(U_i) = \{(A_l^*, p_l)\,;\ l = 1, \ldots, M\} \tag{19}$$

where A_l^* is a possible value that U_i may take, and p_l is the probability that U_i takes the value A_l^*, where $\sum_{l=1}^{M} p_l = 1$. The above distribution reads that an attribute U_i takes a value A_l^* with a probability p_l. Note that U_i taking a single value A_l^* for sure is a special case of (19) with $p_l = 1$ and $p_m = 0$ ($m = 1, \ldots, M, m \neq l$).

Assuming a piecewise linear utility function for U_i, the distribution $S(U_i)$ given by (19) can be equivalently represented using a_{ij} by

$$S(U_i) = \left\{(a_{ij}, \overline{\gamma}_{ij});\ j = 1, \ldots, J_i\right\} \tag{20}$$

where

$$\overline{\gamma}_{ij} = \begin{cases} \sum\limits_{l\in\pi_j} p_l\gamma_{lj}, & \text{for } j = 1 \\ \sum\limits_{l\in\pi_{j-1}} p_l\big(1-\gamma_{l(j-1)}\big) - \sum\limits_{l\in\pi_j} p_l\gamma_{lj}, & \text{for } 2 \leq j \leq J_i - 1 \\ \sum\limits_{l\in\pi_{j-1}} p_l\big(1-\gamma_{l(j-1)}\big), & \text{for } j = J_i \end{cases} \tag{21}$$

$$\pi_j = \left\{l \,\middle|\, a_{ij} \leq A_l^* \leq a_{i(j+1)},\ \ l = 1, \ldots, M\right\}, \quad j = 1, \ldots, J_i - 1 \tag{22}$$

and γ_{ij} is calculated by (16a) and (16b). Note that $\pi_n \cap \pi_m = \varnothing$ ($n, m = 1, \ldots, J_i - 1; n \neq m$) and

$$\bigcup_{j=1}^{J_i-1} \pi_j = \{1, 2, \ldots, M\}.$$

Given the equivalence rule represented by (15), $S(U_i)$ can be equivalently represented by the following expectation using the antecedent referential value set:

$$S(U_i) = \{(A_{ij}, \alpha_{ij});\ j = 1, \ldots, J_i\} \text{ with } \alpha_{ij} = \overline{\gamma}_{ij}. \tag{23}$$

2) Quantitative Attributes Described Using Interval Form: In this case, an antecedent attribute is described by an interval of values. In this case, the way to get α_{ij} is similar to that in the linguistic value case using (12), as discussed earlier in this section. The difference is the way to determine the matching function $\tau(A^*, A)$. The interval matching function, introduced by Bustince and Burillo [61] and Chen and Hsiao [62], can be used to determine $\tau(A^*, A)$.

3) Qualitative Attributes Assessed Using a Subjective Belief Distribution Vector: It is natural that qualitative attributes are assessed using human judgments, which are subjective in nature and are inevitably associated with uncertainties caused due to human being's inability to provide precise judgments, or the lack of information, or the vagueness of meanings about attributes and their assessments. Qualitative parameters in a model may be assessed using a subjective scale against which a range of linguistic values is mapped in domains defined by the model builder. A subjective scale is called a psychometric scale, since it comes from the model builder's minds [63]. The range of a psychometric scale is determined by the level of granularity and fine details in the model. Due to subjectivity, a qualitative attribute could be directly assessed to a distribution using linguistic terms with the degrees of belief based on subjective judgments. In other words, α_{ij} can be assigned directly by the decision maker using his subjective judgments for each A_{ij}. For example, if ε_{ij} is the degree of belief assigned to the association of A_{ij}, then $\alpha_{ij} = \varepsilon_{ij}$.

4) Symbolic Attributes Assessed Using a Subjective Belief Distribution Vector: Usually, a symbolic attribute is deterministic, but one may not be sure about which symbol may be taken for an attribute. Similar to a qualitative linguistic value, if ε_{ij} is the degree of belief assigned to the symbolic term A_{ij}, then $\alpha_{ij} = \varepsilon_{ij}$.

Based on the above transformation techniques, an input can be represented as a belief distribution, which can provide a panoramic view about the status of an attribute. Using such a distributed assessment framework, the features of a range of evidence can be catered for while the assessor is not forced to preaggregate various types of evidence into a single numerical value. The main advantage of doing so is that both precise data and subjective judgments with uncertainty, whether complete or incomplete, can be consistently modeled under the unified framework without loss of their original features.

In summary, an input can be expressed as

$$S\left(A_1^*, \varepsilon_1\right) \wedge S\left(A_2^*, \varepsilon_2\right) \wedge \cdots \wedge S\left(A_T^*, \varepsilon_T\right). \tag{24}$$

Comparing (5) with each rule given in (3), an input can be rewritten as

$$\left(A_1^k, \alpha_1^k\right) \wedge \left(A_2^k, \alpha_2^k\right) \wedge \cdots \wedge \left(A_{T_k}^k, \alpha_{T_k}^k\right)$$

where α_i^k $(i = 1, \ldots, T_k)$ is the degree of belief to which the input (A_i^*, ε_i) belongs to A_i^k of the ith individual antecedent in the kth rule, and $\alpha_i^k \in \{\alpha_{ij};\ \alpha_{ij} \geq 0 \text{ and } \sum_{j=1}^{J_i} \alpha_{ij} \leq 1,\ j = 1, \ldots, J_i\}$. α_{ij} is calculated using the above transformation techniques according to the type and nature of an attribute. The belief degrees were used in Section III-C for generating the activation weights of antecedent attributes. It should be noted that in the above discussion, the development of a belief rule-based system is divided into three parts: belief rule-base construction; input information transformation; and belief rule-base inference. The concepts and techniques in fuzzy set theory, utility theory, and belief theory are used for rule construction and input information transformation, while the inference is implemented using ER.

B. Hierarchical Knowledge Base

The rules of a knowledge base for a complex decision-making problem can be of a hierarchical structure. In general,

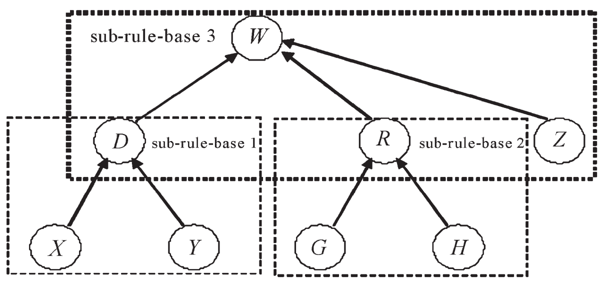


Fig. 1. Simple hierarchical structure.

a bottom-up approach can be used to solve such a problem. Pieces of evidence for the lowest level attributes are aggregated as evidence for the second lowest level attributes, which is, in turn, aggregated to produce evidence for higher level attributes.

Fig. 1 shows a hierarchical knowledge base, composed of three sub-rule bases: sub-rule base 1; sub-rule base 2; and sub-rule base 3. Information is propagated from the bottom level states (X, Y, G, H, and Z) up to the goal state W. Sub-rule base 1 and sub-rule base 2 are independent of each other. The output of sub-rule base 1 (i.e., D) and the output of sub-rule base 2 (i.e., R) together with the independent input state Z are taken as the input states to sub-rule base 3. Each of the three sub-rule bases constitutes a basic rule base, as investigated in Section III-A and can be dealt with using the methodology developed in Section III, where sub-rule base 1 and sub-rule base 2 are solved first, followed by the solution of sub-rule base 3.

C. Hierarchical Rule-Base Inference

Consider the simple case of two pieces of evidence shown in sub-rule base 1 in Fig. 1, where two evidential states X and Y determine the output state D.

1) X and Y are Independent Evidential States: Note that the evidential state X can consist of multiple interrelated attributes, which all affect the output state D and are considered simultaneously. In other words, X may be in forms of "$X_1 \wedge X_2 \wedge \cdots \wedge X_{\mathrm{T}}$." The evidential state Y may have a similar form.

Rules describing the relations between X and D and between Y and D can be formulated as follows.

1) Rule base describing the relations between X and D

$$\text{if } X \text{ is } A^1, \text{ then } D \text{ is } \{(D_1,\eta_{11}),(D_2,\eta_{21}), \ldots,(D_N,\eta_{N1})\}$$
$$\text{else if } X \text{ is } A^2, \text{ then } D \text{ is } \{(D_1,\eta_{12}),\ldots,(D_N,\eta_{N2})\}$$
$$\cdots$$
$$\text{else if } X \text{ is } A^i, \text{ then } D \text{ is } \{(D_1,\eta_{1i}),\ldots,(D_N,\eta_{Ni})\}$$
$$\cdots$$
$$\text{else } X \text{ is } A^L, \text{ then } D \text{ is } \{(D_1,\eta_{1L}),\ldots,(D_N,\eta_{NL})\}.$$

2) Rule base describing the relations between Y and D

$$\text{if } Y \text{ is } B^1, \text{ then } D \text{ is } \{(D_1,\gamma_{11}),\ldots,(D_N,\gamma_{N1})\}$$
$$\text{else if } Y \text{ is } B^2, \text{ then } D \text{ is } \{(D_1,\gamma_{12}),\ldots,(D_N,\gamma_{N2})\}$$
$$\cdots$$
$$\text{else if } Y \text{ is } B^j, \text{ then } D \text{ is } \{(D_1,\gamma_{1j}),\ldots,(D_N,\gamma_{Nj})\}$$
$$\cdots$$
$$\text{else } Y \text{ is } B^P, \text{ then } D \text{ is } \{(D_1,\gamma_{1P}),\ldots,(D_N,\gamma_{NP})\}$$

where A^i $(i=1,\ldots,L)$ denotes a packet antecedent attribute corresponding to $(X_1 \wedge \cdots \wedge X_T)$ in the ith rule, and η_{ki} is the degree of belief to which D is believed to take D_k due to X being A^i; B^j $(j=1,\ldots,P)$ denotes a packet antecedent attribute corresponding to $(Y_1 \wedge \cdots \wedge Y_T)$ in the jth rule, and γ_{kj} is the degree of belief to which D is believed to take D_k due to Y being B^j.

Once the inputs for X and Y are given, the output for D can be inferred using the sub-rule base and the proposed RIMER method based on X and Y, respectively. The final output for D is a combination of the outputs from X and Y, which can be done by directly using the ER algorithm.

Suppose there are several sources or experts, denoted by E_k $(k=1,\ldots,K)$, and A_k^* is the input derived from E_k for the evidential state X. For each input, the corresponding output $D_k^* = \{(D_1,\eta_{1k}^*),(D_2,\eta_{2k}^*),\ldots,(D_N,\eta_{Nk}^*)\}$ can be derived, which can be formulated as

$$\text{if } X \text{ is } A_1^*, \text{ then } D \text{ is } \{(D_1,\eta_{11}^*),\ldots,(D_N,\eta_{N1}^*)\}$$
$$\text{if } X \text{ is } A_2^*, \text{ then } D \text{ is } \{(D_1,\eta_{12}^*),\ldots,(D_N,\eta_{N2}^*)\}$$
$$\cdots$$
$$\text{if } X \text{ is } A_k^*, \text{ then } D \text{ is } \{(D_1,\eta_{1k}^*),\ldots,(D_N,\eta_{Nk}^*)\}$$
$$\cdots$$
$$\text{if } X \text{ is } A_K^*, \text{ then } D \text{ is } \{(D_1,\eta_{1L}^*),\ldots,(D_N,\eta_{NK}^*)\}.$$

Furthermore, it may be assumed that different experts/different sources may have different reliability weights w_{E_k} $(k=1,\ldots,K)$. Then the final output for D can be generated by taking into account both the experts' views and their weights.

2) X and Y are Dependent Evidential States: In this case, the rule base may be expressed as

$$\text{if } X \text{ is } A^1 \wedge Y \text{ is } B^1, \text{ then } D \text{ is } \{(D_1,\lambda_{11}),\ldots,(D_N,\lambda_{N1})\}$$
$$\text{else if } X \text{ is } A^2 \wedge Y \text{ is } B^2, \text{ then } D \text{ is } \{(D_1,\lambda_{12}), \ldots,(D_N,\lambda_{N2})\}$$
$$\cdots$$
$$\text{else if } X \text{ is } A^i \wedge Y \text{ is } B^i, \text{ then } D \text{ is } \{(D_1,\lambda_{1i}), \ldots,(D_N,\lambda_{Ni})\}$$
$$\cdots$$
$$\text{else if } X \text{ is } A^L \wedge Y \text{ is } B^L, \text{ then } D \text{ is } \{(D_1,\lambda_{1L}), \ldots,(D_N,\lambda_{NL})\}$$

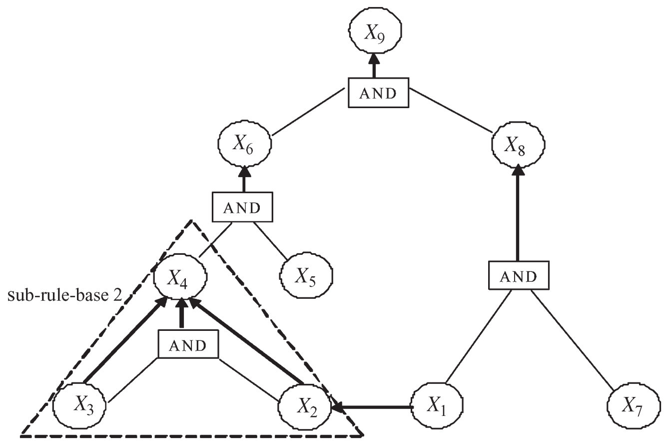


Fig. 2. Hierarchical structure of a small exploratory expert system.

where "∧" is a connective to reflect the corelationship between X and Y, which are combined to get the goal state D. The "∨" connective might also be used for this purpose. In this case, the RIMER approach can also be used to infer the goal state D.

Similarly, sub-rule base 2, and finally, sub-rule base 3, may be implemented to assess the final goal state W.

Note that the new RIMER method developed above requires that the input attributes of a rule are independent of each other. In a so-called "recursive" (not hierarchical) rule base, the consequent of a rule may be used as an attribute in another rule, and the consequent of the latter rule may then be used as an attribute in the former rule. Such a "recursive" rule base may not be allowed by RIMER if the independence requirement is violated.

V. Numerical Study

A. Problem Description

A numerical example, which is based on a fuzzy rule base for an exploratory expert system discussed by Hodges *et al.* [64], is studied in this section. The example aims to determine a confidence degree to which the system believes that a container may contain graphite.

The input variables defined in the exploratory expert system include the following:

1) accuracy of the weight measurement;
2) degree to which the calculated density is consistent with graphite;
3) observer's experience;
4) observer's confidence that the real-time radiography (RTR) shows graphites,

For illustration purposes, each of these input variables and the output variable (confidence to which a container contains graphite) are defined as having values of high (H), medium (M), or low (L). The system structure is shown in Fig. 2.

In Fig. 2, the parameters are defined as follows.

- X_1 Observer's experience.
- X_2 Accuracy of the fill height determination.
- X_3 Accuracy of the weight measurement.
- X_4 Accuracy of the calculated density.
- X_5 Consistency of the calculated density with graphite.
- X_6 Confidence that the density indicates graphite.
- X_7 Observer's confidence that the RTR shows graphite.
- X_8 Confidence that the RTR shows graphite.
- X_9 Confidence that the container contains graphite.

This example uses four input variables (X_1, X_3, X_5, X_7) and four intermediate variables (X_2, X_4, X_6, X_8) to predict X_9 in terms of qualitative linguistic values. The expert knowledge is coded as if–then rules, hierarchically organized in five sub-rule bases [64]. In sub-rule base 2, for example, the following if–then rules are included:

$$\text{if } X_2 \text{ is L, then } X_4 \text{ is L}$$
$$\text{if } X_3 \text{ is L, then } X_4 \text{ is L}$$
$$\text{if } X_2 \text{ is H} \wedge X_3 \text{ is H, then } X_4 \text{ is H}$$
$$\text{if } X_2 \text{ is H} \wedge X_3 \text{ is M, then } X_4 \text{ is H}$$
$$\text{if } X_2 \text{ is M} \wedge X_3 \text{ is H, then } X_4 \text{ is H}$$
$$\text{if } X_2 \text{ is M} \wedge X_3 \text{ is M, then } X_4 \text{ is M.}$$

Based on the rule base, fuzzy inference was used in [64]. The process of the fuzzy inference consists of the following steps:

1) Fuzzification of a crisp input by input membership functions.
2) Fuzzy aggregation of antecedents in each rule (∧ connective).
3) Implication relation for each individual rule (if–then connective).
4) Aggregation of the rules (also connective).
5) Deriving inference from the set of rules, using the crisp input to obtain the fuzzy output.
6) Defuzzification of the output.

There are a number of different ways to implement a fuzzy inference engine. Mamdani and Assilian [65] described an inference engine in terms of a fuzzy relation matrix and used the composition rule of inference (CRI) to arrive at an output fuzzy set for a given input fuzzy set. This CRI was applied in [64]. The results generated from 12 test runs conducted using CRI are summarized in the sixth column of Table III, as shown later in this section.

B. Extension of the Original Rules Using the Belief Structure

Each rule used in [64] has only one consequent with a belief degree being always exactly one, which is a special case of the belief structure discussed in Sections III and IV. In this subsection, the rules are extended using the belief structure to provide better flexibility and versatility for more precisely imitating human reasoning using rule-based systems. The definitions of the extended rules with the consequents having the dedicated degrees of belief are given in Table II. These degrees of belief in the consequents were assigned by the researchers as a result of the observation of the given expert judgments. In a more systematic scheme, the belief degrees could be trained using expert judgments as test data and may also be updated once new evidence becomes available. The rules are numbered and clustered in conjunction with the five sub-rule bases. The

TABLE II
NEW RULE BASE WITH THE BELIEF STRUCTURE

Number	W	Antecedent	Consequent
1	1	(X_1 is H)	X_2 is {(H, 1)}
2	1	(X_1 is M)	X_2 is {(M, 1)}
3	1	(X_1 is L)	X_2 is {(L, 1)}
4	1	(X_3 is L)	X_4 is {(L, 1)}
5	1	(X_2 is L)	X_4 is {(L, 1)}
6	1	(X_2 is H $\wedge X_3$ is H)	X_4 is {(H, 1)}
7	0.7	(X_2 is H $\wedge X_3$ is M)	X_4 is {(H, 0.3), (M, 0.7)}
8	0.7	(X_2 is M $\wedge X_3$ is H)	X_4 is {(H, 0.3), (M, 0.7)}
9	0.7	(X_2 is M $\wedge X_3$ is M)	X_4 is {(M, 1)}
10	1	(X_5 is H $\wedge X_4$ is H)	X_6 is {(H, 1)}
11	1	(X_5 is H $\wedge X_4$ is M)	X_6 is {(H, 0.4), (M, 0.6)}
12	1	(X_5 is H $\wedge X_4$ is L)	X_6 is {(M, 1)}
13	1	(X_5 is M $\wedge X_4$ is H)	X_6 is {(H, 0.2), (M, 0.8)}
14	0.4	(X_5 is M $\wedge X_4$ is M)	X_6 is {(M, 1)}
15	1	(X_5 is M $\wedge X_4$ is L)	X_6 is {(M, 0.2), (L, 0.8)}
16	0.2	(X_5 is L $\wedge X_4$ is H)	X_6 is {(H, 0.1), (M, 0.3), (L, 0.6)}
17	1	(X_5 is L $\wedge X_4$ is M)	X_6 is {(M, 0.2), (L, 0.8)}
18	1	(X_5 is L $\wedge X_4$ is L)	X_6 is {(L, 1)}
19	1	(X_7 is H $\wedge X_1$ is H)	X_8 is {(H, 1)}
20	0.2	(X_7 is H $\wedge X_1$ is M)	X_8 is {(H, 0.3), (M, 0.7))
21	0.8	(X_7 is H $\wedge X_1$ is L)	X_8 is {(M, 0.3), (L, 0.7)}
22	1	(X_7 is M $\wedge X_1$ is H)	X_8 is {(H, 0.4), (M, 0.6)}
23	0.4	(X_7 is M $\wedge X_1$ is M)	X_8 is {(M, 1)}
24	1	(X_7 is M $\wedge X_1$ is L)	X_8 is {(M, 0.1), (L, 0.9)}
25	1	(X_7 is L $\wedge X_1$ is H)	X_8 is {(H, 0.1), (M, 0.3), (L, 0.6)}
26	1	(X_7 is L $\wedge X_1$ is M)	X_8 is {(M, 0.3), (L, 0.7)}
27	1	(X_7 is L $\wedge X_1$ is L)	X_8 is {(L, 1)}
28	1	(X_6 is H $\wedge X_8$ is H)	X_9 is {(H, 1)}
29	0.6	(X_6 is H $\wedge X_8$ is M)	X_9 is {(H, 0.2) (M, 0.8)}
30	1	(X_6 is H $\wedge X_8$ is L)	X_9 is {(H, 0.1), (M, 0.2), (L, 0.7)}
31	0.6	(X_6 is M $\wedge X_8$ is H)	X_9 is {(H, 0.2) (M, 0.8)}
32	0.6	(X_6 is M $\wedge X_8$ is M)	X_9 is {(M, 1)}
33	1	(X_6 is M $\wedge X_8$ is L)	X_9 is {(M, 0.1) (L, 0.9)}
34	1	(X_6 is L $\wedge X_8$ is H)	X_9 is {(H, 0.1), (M, 0.2), (L, 0.7)}
35	1	(X_6 is L $\wedge X_8$ is M)	X_9 is {(M, 0.1) (L, 0.9)}
36	1	(X_6 is L $\wedge X_8$ is L)	X_9 is {(L, 1)}

extended rule base can imitate both discrete and continuous reasoning processes. The numbers shown in the second column represent the weights assigned to the rules, where H stands for high, M for medium, and L for low. For example, rule (7) means: if "the accuracy of the fill height determination" is high and "the accuracy of the weight measurement" is medium, then "the accuracy of the calculated density" is high with a belief degree of 0.3 and medium with a belief degree of 0.7 and the rule is assigned a weight of 0.7.

Note that rules (6)–(9) indicate that "the accuracy of the calculated density" depends on "the accuracy of the fill height determination" and "the accuracy of the weight measurement." However, rules (4) and (5) check each of these two factors (fill height accuracy and weight accuracy) separately rather than in a compound fashion as found in rules (6)–(9), i.e.,

$$\text{if } X_3 \text{ is L, then } X_4 \text{ is } \{(\text{L}, 1)\}$$
$$\text{if } X_2 \text{ is L, then } X_4 \text{ is } \{(\text{L}, 1)\}.$$

The definitions of the two rules result from the expert's experience that one negative piece of evidence, such as an inaccurate fill height determination, would make the confidence in the final result low, regardless of other evidence that may also be available. Furthermore, the differences between the rules with different antecedents can be shown using the belief degrees in the conclusion part, e.g.,

$$\text{if } X_2 \text{ is H} \wedge X_3 \text{ is H, then } X_4 \text{ is } \{(\text{H}, 1)\}$$
$$\text{if } X_2 \text{ is H} \wedge X_3 \text{ is M, then } X_4 \text{ is } \{(\text{H}, 0.3), (\text{M}, 0.7)\}$$
$$\text{if } X_2 \text{ is M} \wedge X_3 \text{ is H, then } X_4 \text{ is } \{(\text{H}, 0.3), (\text{M}, 0.7)\}$$

which, however, had the same consequent (H) in the original rule base in [64].

Test 3 is examined in detail in the following subsection to demonstrate the procedure involved in the inference engine. The other tests are conducted in the same way using the RIMER approach, and the generated results are summarized in Table III.

TABLE III
TEST RUN RESULTS COMPARED WITH THE EXPERT'S JUDGEMENT

Test Run	X_3	X_1	X_5	X_7	X_{91}	X_{92}	X_{93}	X_{9E}
1	0.98	1.00	1.00	1.00	0.82	1.00	*1.00*	1.00
2	0.98	0.80	0.80	0.80	0.82	0.91	*0.94*	0.90
3	0.98	0.80	0.20	0.80	0.50	0.44	*0.61*	0.60
4	0.98	0.40	0.40	0.80	0.50	0.50	*0.28*	0.30
5	0.98	0.40	0.60	0.80	0.65	0.53	*0.35*	0.40
6	0.98	1.00	0.00	0.80	0.50	0.50	*0.25*	0.20
7	0.98	0.00	0.00	0.00	0.18	0.00	*0.00*	0.00
8	0.98	1.00	1.00	0.20	0.50	0.50	*0.46*	0.30
9	0.98	0.40	0.60	0.20	0.30	0.11	*0.14*	0.10
10	0.98	0.60	0.40	0.20	0.18	0.06	*0.18*	0.20
11	0.98	0.80	0.20	0.20	0.18	0.11	*0.24*	0.20
12	0.98	0.80	0.80	0.20	0.50	0.50	*0.36*	0.40

C. Implementation of the RIMER Inference Process

1) Transformation of Input: Before the inference can be started, input values need to be transformed and represented in terms of the referential values: low, medium, and high using belief degrees to which the values belong to the referential values. Note that the transformation based on the max–min matching method, which was used by Hodges *et al.* in [64], seems to not be rational. For example, belief degrees would be calculated as follows using the max–min matching method in (13):

$$0.98 \text{ is transformed to } \{(\text{H}, 1)\}$$
$$1 \text{ is transformed to } \{(\text{H}, 1)\}$$
$$0.8 \text{ is transformed to } \{(\text{H}, 1)\}$$
$$0.4 \text{ is transformed to } \{(\text{L}, 0.31), (\text{M}, 0.69)\}$$
$$0.2 \text{ is transformed to } \{(\text{L}, 1)\}$$
$$0.1 \text{ is transformed to } \{(\text{L}, 1)\}$$
$$0.6 \text{ is transformed to } \{(\text{M}, 0.66), (\text{H}, 0.34)\}.$$

Note that in the above transformation 0.98, 1, and 0.8 are given the same matching degree of 1 to high, and 0.2 and 0.1 are given the same matching degree of 1 to low. This is not intuitively reasonable and would cause inaccuracy in subsequent inference, thereby sacrificing the credibility of results subsequently generated. Because of these concerns, the rule-based transformation technique [17] is used for the quantitative data transformation, shown by using (15)–(18). The equivalence rules are given as

$$1 \text{ is equivalently transformed to } \{(\mathrm{H},1)\} \text{ or } \mathrm{H}=1$$
$$0.5 \text{ is equivalently transformed to } \{(\mathrm{M},1)\} \text{ or } \mathrm{M}=0.5$$
$$0 \text{ is equivalently transformed to } \{(\mathrm{L},1)\} \text{ or } \mathrm{L}=0.$$

Hence

$$0.98 \text{ is equivalently transformed to } \{(\mathrm{H},0.96),(\mathrm{M},0.04)\}$$
$$0.8 \text{ is equivalently transformed to } \{(\mathrm{H},0.6),(\mathrm{M},0.4)\}$$
$$0.6 \text{ is equivalently transformed to } \{(\mathrm{H},0.2),(\mathrm{M},0.8)\}$$
$$0.4 \text{ is equivalently transformed to } \{(\mathrm{M},0.8),(\mathrm{L},0.2)\}$$
$$0.2 \text{ is equivalently transformed to } \{(\mathrm{M},0.4),(\mathrm{L},0.6)\}.$$

Furthermore, the "$\wedge$" connective in the antecedent is defined as the product operation, as shown in (6a). In the following, assume that all antecedent attributes in any rule have equal weight for illustration purpose, i.e., $\delta_{ki}=1$ for any $k=1,\ldots,36$; $i=1,2$.

2) Inference of the Sub-Rule Base 1: The input for X_1 is 0.8 or equivalently given by

$$X_1:\ \{(\mathrm{H},0.6),(\mathrm{M},0.4),(\mathrm{L},0)\}$$

which means that there is a 60% certainty that X_1 is H, 40% that X_1 is M, and 0% that X_1 is L.

In the sub-rule base for X_2 as shown in Table III, the activation weights w_k for the three rules R_k ($k=1,2,$ and 3) are generated by using (6a) and (7) by $w_1=0.60$, $w_2=0.4$, and $w_3=0$, respectively. Note that the rule weights θ_k are assumed to be equal to 1 for all $k=1$, 2, and 3.

Based on the sub-rule base for X_2 and using the window-based and graphically designed intelligent decision system (IDS) [66], the consequent estimates are generated. The final assessment result for X_2 is obtained as

$$X_2:\ \{(\mathrm{H},0.6923),(\mathrm{M},0.3077),(\mathrm{L},0)\}$$

which has a corresponding confidence score of 0.8462 generated by using (12).

3) Inference of the Sub-Rule Base 2: Note that the input for X_3 is given by

$$X_3:\ \{(\mathrm{H},0.96),(\mathrm{M},0.04),(\mathrm{L},0)\}.$$

Considering the sub-rule base for X_4 shown in Table III and since the "$\wedge$" connective is used in the antecedents, the activation weights w_k for the rules R_k (k $=4,\ldots,9$) are generated by using (6a) and (7) by $w_4=0$, $w_5=0$, $w_6=0.7390$, $w_7=0.0216$, $w_8=0.2299$, and $w_9=0.0096$, respectively. Note that the antecedent weight δ_i is 1 for all $i=1$, 2, and 3, and the rule weights are given by $\theta_4=\theta_5=\theta_6=1$, and $\theta_7=\theta_8=\theta_9=0.7$. The final assessment result for X_4 is obtained using IDS as

$$X_4 \text{ is } \{(\mathrm{H},0.9304),(\mathrm{M},0.0696),(\mathrm{L},0)\}$$

with a corresponding confidence score of 0.9652.

Note that the above result is determined by the activation weights, the rule weights, and the belief degrees for rules (4)–(9). Rule R_6 has the highest activation weight of 0.7390 and the highest rule weight of 1, and thus it plays the most important role in determining the output, followed by rule R_8. The other four rules have only made marginal or no contribution to the result. Moreover, the consequents for R_6 and R_8 are given by {(H, 1)} and {(H, 0.3), (M, 0.7)}, respectively. Hence, the final output generated should be close to high to a large degree. This is confirmed in the above result.

4) Inference of the Sub-Rule Base 3: In this rule base, X_6 is determined by X_4 and X_5. Note that the input for X_5 is given by

$$X_5:\ \{(\mathrm{H},0),(\mathrm{M},0.4),(\mathrm{L},0.6)\}.$$

From the sub-rule base for X_6 shown in Table III, the activation weight w_k for the rules R_k ($k=10,\ldots,18$) are generated by using (6a) and (7) by $w_{10}=0$, $w_{11}=0$, $w_{12}=0$, $w_{13}=0.6934$, $w_{14}=0.0207$, $w_{15}=0$, $w_{16}=0.2080$, $w_{17}=0.0778$, and $w_{18}=0$, respectively. The rule weights are $\theta_{10}=1$, $\theta_{11}=1$, $\theta_{12}=1$, $\theta_{13}=1$, $\theta_{14}=0.4$, $\theta_{15}=1$, $\theta_{16}=0.2$, $\theta_{17}=1$, and $\theta_{18}=1$, respectively. The final assessment result for X_6 is generated using IDS as

$$X_6:\ \{(\mathrm{H},0.1707),(\mathrm{M},0.7473),(\mathrm{L},0.0820)\}$$

with a confidence score of 0.5444.

This output for X_6 also seems logical as the activation weight w_{13} is as high as 0.6934, the rule weight θ_{13} is 1, and the consequent of the rule R_{13} is given by {(H, 0.3), (M, 0.7), (L, 0)}, which is assessed to medium to a large degree of 0.7. The second most important rule is R_{16} with $w_{16}=0.2080$, though its rule weight is only 0.2. The consequent of the rule R_{16} is given by {(H, 0.1), (M, 0.3), (L, 0.6)}), which is also assessed to medium to a degree of 0.3.

5) Inference of the Sub-Rule Base 4: Note that the input for X_7 is given by

$$X_7:\ \{(\mathrm{H},0.6),(\mathrm{M},0.4),(\mathrm{L},0)\}$$

and the input for X_1 is also given by

$$X_1:\ \{(\mathrm{H},0.6),(\mathrm{M},0.4),(\mathrm{L},0)\}.$$

From the sub-rule base for X_8 shown in Table III, the activation weights w_k for the rules R_k ($k=19,\ldots,27$) are generated by using (6a) and (7) by $w_{19}=0.5056$, $w_{20}=0.0674$, $w_{21}=0$, $w_{22}=0.3371$, $w_{23}=0.0899$, $w_{24}=0$, $w_{25}=0$, $w_{26}=0$, and $w_{27}=0$, respectively. The rule weights are $\theta_{19}=1$, $\theta_{20}=0.2$, $\theta_{21}=0.8$, $\theta_{22}=1$, $\theta_{23}=0.4$, $\theta_{24}=1$, $\theta_{25}=1$, $\theta_{26}=1$,

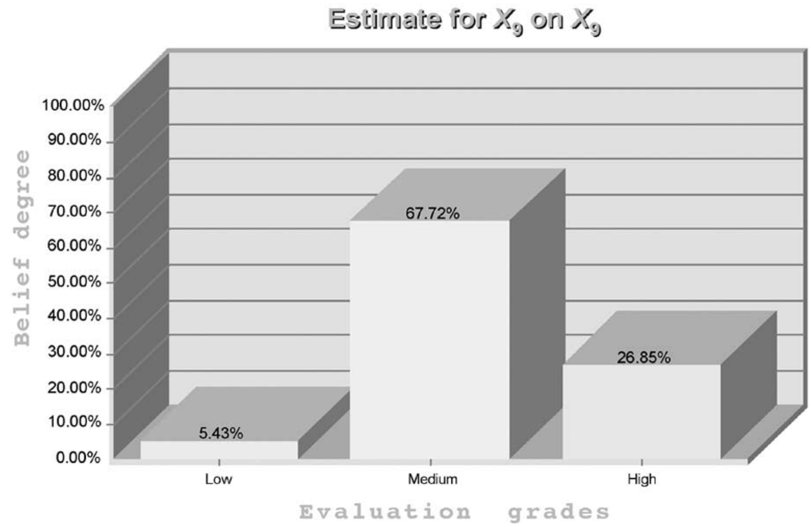


Fig. 3. Evaluation on the final output X_9.

and $\theta_{27} = 1$, respectively. The final assessment result for X_8 is generated using IDS as

$$X_8 : \{(\mathrm{H}, 0.7459), (\mathrm{M}, 0.2541), (\mathrm{L}, 0)\}$$

with a confidence score of 0.8730.

Note that for the rules R_{19} and R_{22}, the activation weights are $w_{19} = 0.5056$ and $w_{22} = 0.3371$, the rule weights are $\theta_{19} = 1$ and $\theta_{22} = 1$, and the corresponding consequents are given by {(H, 1)} and {(H, 0.4), (M, 0.6), (L, 0)}, respectively. The output for X_8 therefore should be to a large extent close to H, which is confirmed by the above result for X_8.

6) Inference of the Sub-Rule Base 5: Since X_6 has already been inferred from X_4 and X_5 by {(H, 0.1707), (M, 0.7473), (L, 0.0820)}, using the sub-rule base for X_9 shown in Table III, the activation weights w_k for the rules R_k ($k = 28, \ldots, 36$) are generated by using (6a) and (7) by $w_{28} = 0.1862$, $w_{29} = 0.0381$, $w_{30} = 0$, $w_{31} = 0.4892$, $w_{32} = 0.1666$, $w_{33} = 0$, $w_{34} = 0.0895$, $w_{35} = 0.0305$, and $w_{36} = 0$, respectively. The rule weights are $\theta_{28} = 1$, $\theta_{29} = 0.6$, $\theta_{30} = 1$, $\theta_{31} = 0.6$, $\theta_{32} = 0.6$, $\theta_{33} = 1$, $\theta_{34} = 1$, $\theta_{35} = 1$, and $\theta_{36} = 1$, respectively. The final assessment result for X_9 is obtained using IDS as

$$X_9 : \{(\mathrm{H}, 0.2685), (\mathrm{M}, 0.6772), (\mathrm{L}, 0.0543)\}$$

with a confidence score of 0.6071.

Note that for the rules R_{28}, R_{31}, and R_{32}, the activation weights are given by $w_{28} = 0.1862$, $w_{31} = 0.4892$, and $w_{32} = 0.1666$, the rule weights by $\theta_{28} = 1$, $\theta_{31} = 0.6$, and $\theta_{32} = 0.6$, and the consequents by {(H, 1)}, {(H, 0.2), (M, 0.8)}, and {(M, 1)}, respectively. Therefore, the output for X_9 should be close to medium to the largest degree, followed by high. Thus, the output for X_9 seems to be a logical result as well, which is shown in Fig. 3.

D. Test Runs Compared With the Expert's Opinions

The results are summarized in Table III, which are generated from 12 test runs conducted using the CRI approach [64] and the proposed RIMER approach. The final values generated using the RIMER approach, which can be represented by both the confidence scores and the distributed assessments with the belief degrees to which the system believes that the container contains graphite, are compared with the confidence scores provided by an expert. In the study, the same input values as given by Hodges *et al.* [64] are used.

Column 6 (X_{91}) summarizes the results generated using the CRI approach by Hodges *et al.* [64]. Column 7 (X_{92}) presents the results generated using the RIMER approach and the same rule base as the CRI approach used, i.e., without the belief structure. Moreover, the same matching function is used in these two approaches. Column 8 (X_{93}) summarizes the results generated using the RIMER approach and the new rule base given in Table II with the belief structure. The IDS software is used to generate the consequent estimates. Column 9 ($X_{9\mathrm{E}}$) summarizes the expert's confidence that contents are graphite.

The following remarks can be made from the results shown in Table III.

1) The results generated using the RIMER approach with the belief structure in the rule base are much closer to the expert's judgments compared with the CRI approach. RIMER also provides scope and flexibility for better knowledge acquisition, including the definitions of input and output variables, the construction of rule bases, and the assignment of rule weights.
2) The rule base with the belief structure is constructed on the basis of the original rule base provided by Hodges *et al.* [64]. The belief degrees are assigned to each rule by the researchers by examining the original rule base and perceiving the expert's way of making the judgments. For example, as shown in Table III, the expert's final confidence in the verification of the graphite code tends to be quite low in the presence of a piece of negative information such as the inexperience of the observer. Such a human reasoning process can be accurately imitated by, for example, assigning weights to antecedent attributes and/or by adjusting belief degrees in the consequents of rules in a systematic manner. This is one of the prominent

TABLE IV
SENSITIVE ANALYSIS FOR RULE WEIGHTS AND BELIEF DEGREES

Items	Original	Revised	Distribution Assessment	Degree
θ_{31}	0.6	1	(0.2373, 0.7315, 0.0312)	0.6031
θ_{29}	0.6	1	(0.2659, 0.6805, 0.0536)	0.6062
θ_{33}	1	0.6	(0.2685, 0.6772, 0.0543)	0.6071
Belief of R_{31}	(0.2, 0.8, 0)	(0.6, 0.4, 0)	(0.5372, 0.4129, 0.0544)	0.7391
		(0.3, 0.7, 0)	(0.3344, 0.6113, 0.0543)	0.6401
		(0.22, 0.78, 0)	(0.2817, 0.6640, 0.0543)	0.6137
Belief of R_{29}	(0.2, 0.8, 0)	(0.6, 0.4, 0)	(0.2831, 0.6623, 0.0546)	0.6142
		(0.3, 0.7, 0)	(0.2722, 0.6735, 0.0543)	0.6089
		(0.22, 0.78, 0)	(0.2693, 0.6765, 0.0543)	0.6075
Belief of R_{33}	(0, 0.1, 0.9)	(0, 1, 0)	(0.2685, 0.6772, 0.0543)	0.6071

features of the rule base with the belief structure, and the rules may even be trained in different ways.

3) If the same inputs and matching degrees are used, most of the results generated using the CRI approach and the RIMER approach are similar, though the results generated using the RIMER approach seem closer to the expert's judgments in a number of the tests.
4) Note that the activation weights for rules play an essential role in the inference procedure. The final output tends to be close to the consequents of rules with high activation weights. In addition, belief degrees in each rule and rule weights also affect output while activation weights are fixed. This feature can be demonstrated in the following sensitive studies.

Take test 3 for example. The global antecedent weights w_k for the rules R_k $(k = 28, \ldots, 36)$ are generated by $w_{28} = 0.1862$, $w_{29} = 0.0381$, $w_{30} = 0$, $w_{31} = 0.4892$, $w_{32} = 0.1666$, $w_{33} = 0$, $w_{34} = 0.0895$, $w_{35} = 0.0305$, and $w_{36} = 0$, respectively. The rule weights are $\theta_{28} = 1$, $\theta_{29} = 0.6$, $\theta_{30} = 1$, $\theta_{31} = 0.6$, $\theta_{32} = 0.6$, $\theta_{33} = 1$, $\theta_{34} = 1$, $\theta_{35} = 1$, and $\theta_{36} = 1$, respectively. Now consider the three rules with the highest global antecedent weight $w_{31} = 0.4892$, the lower weight $w_{29} = 0.0381$, and the lowest weight $w_{33} = 0$. The changes caused by adjusting the rule weights or the belief degrees are shown in Table IV. Note that the original output of X_9 is given as

$$X_9 : \{(\mathrm{H}, 0.2685), (\mathrm{M}, 0.6772), (\mathrm{L}, 0.0543)\}$$

with a confidence score of 0.6071.

For example, in Table IV, (0.2373, 0.7315, 0.0312) means the distribution assessment {(H, 0.2373), (M, 0.7315), (L, 0.0312)}. From Table IV, the following remarks may be made.

1) First, if the activation weight of a rule is equal to 0 (e.g., $w_{33} = 0$), then the weight and the belief degree of this rule will have no influence on the final output.
2) If the activation weight of a rule is not equal to 0, then the weight and the belief degrees of this rule will affect the final output. The degree to which the final output can be affected is determined by the magnitude of the activation weight and the belief degrees. Moreover, if a rule has the highest activation weight (e.g., R_{31}), a small change in its belief degrees [e.g., from (0.2, 0.8, 0) to (0.3, 0.7, 0)] will result in a significantly different final output (e.g., 0.6401) from the original output (0.6071). This seems logical and reasonable.
3) The distribution assessment provides a panoramic view about the output status, from which one can see the variation between the original output and the revised output on each linguistic term. A distribution is easy to understand and flexible to represent input information than a single average value.

TABLE V
RULE BASE FOR X_6 WITH THE UPDATED BELIEF DEGREE

w_k	Antecedent	Consequent
$w_{10} = 0$	(X_4 is H ∧ X_5 is H)	X_6 is {(H, 0.9)}
$w_{11} = 0$	(X_4 is H ∧ X_5 is M)	X_6 is {(H, 0.36), (M, 0.54)}
$w_{12} = 0$	(X_4 is H ∧ X_5 is L)	X_6 is {(M, 0.9)}
$w_{13} = 0$	(X_4 is M ∧ X_5 is H)	X_6 is {(H, 0.27), (M, 0.63)}
$w_{14} = 0$	(X_4 is M ∧ X_5 is M)	X_6 is {(M, 0.9)}
$w_{15} = 0$	(X_4 is M ∧ X_5 is L)	X_6 is {(M, 0.18), (L, 0.72)}
$w_{16} = 0.7278$	(X_4 is L ∧ X_5 is H)	X_6 is {(H, 0.09), (M, 0.27), (L, 0.54)}
$w_{17} = 0.2722$	(X_4 is L ∧ X_5 is M)	X_6 is {(M, 0.18), (L, 0.72)}
$w_{18} = 0$	(X_4 is L ∧ X_5 is L)	X_6 is {(L, 0.9)}

E. Inference Based on Incomplete Input Information

Both complete and incomplete inference can be accommodated in a unified manner within the proposed RIMER framework. To illustrate how incomplete input can be dealt with in the inference methodology, in the above case study, the same input for X_1, X_3, and X_7 is used, but the input for X_5 is modified to {(L, 0.8)}, so that the actual assessment for X_5 is assumed to be incomplete. This assumes that the experts are only 80% certain that the consistency of the calculated density with graphite is "low." In other words, the degree of ignorance is 0.2 in this assumption.

Note that the assessment for X_4 has already been generated by {(H, 0.9304), (M, 0.0696), (L, 0)}, which is complete. From the sub-rule base for X_6, the new activation weight w_k for the rules R_k $(k = 10, \ldots, 18)$ based on the revised input are generated by using (6a) and (7) by $w_{10} = 0$, $w_{11} = 0$, $w_{12} = 0$, $w_{13} = 0$, $w_{14} = 0$, $w_{15} = 0$, $w_{16} = 0.7278$, $w_{17} = 0.2722$, and $w_{18} = 0$, respectively. The rule weights are still $\theta_{10} = 1$, $\theta_{11} = 1$, $\theta_{12} = 1$, $\theta_{13} = 1$, $\theta_{14} = 0.4$, $\theta_{15} = 1$, $\theta_{16} = 0.2$, $\theta_{17} = 1$, and $\theta_{18} = 1$, respectively.

Due to the assumed incomplete input for X_5, the belief degree of the relevant rules needs to be updated to reflect the incompleteness. Using (8)

$$\beta_{ik} = \overline{\beta}_{ik} \frac{\sum_{t=1}^{2}\left(\tau(t,k)\sum_{j=1}^{3}\alpha_{tj}\right)}{\sum_{t=1}^{2}\tau(t,k)} = \overline{\beta}_{ik}\frac{1.8}{2}$$
$$= \overline{\beta}_{ik} * 0.9, \quad i = 1,2,3;\ k = 10,\ldots,18.$$

Therefore, $0 < \sum_{i=1}^{3}\beta_{ik} < 1$ for all rules that are associated with X_5. The sub-rule base for X_6 with the updated belief degrees is shown in Table V.

TABLE VI
RULE BASE FOR X_9 WITH THE UPDATED BELIEF DEGREE

w_k	Antecedent	Consequent
0.0624	(X_6 is H ∧ X_8 is H)	X_9 is {(H, 0.9572)}
0.0128	(X_6 is H ∧ X_8 is M)	X_9 is {(H, 0.1914) (M, 0.7658)}
0	(X_6 is H ∧ X_8 is L)	X_9 is {(H, 0.0957), (M, 0.1914), (L, 0.67)}
0.1325	(X_6 is M ∧ X_8 is H)	X_9 is {(H, 0.1914) (M, 0.7658)}
0.0451	(X_6 is M ∧ X_8 is M)	X_9 is {(M, 0.9572)}
0	(X_6 is M ∧ X_8 is L)	X_9 is {(M, 0.0957) (L, 0.8615)}
0.5573	(X_6 is L ∧ X_8 is H)	X_9 is {(H, 0.0957), (M, 0.1914), (L, 0.67)}
0.1899	(X_6 is L ∧ X_8 is M)	X_9 is {(M, 0.0957) (L, 0.8615)}
0	(X_6 is L ∧ X_8 is L)	X_9 is {(L, 0.0957)}

Using the above sub-rule base, the assessment result for X_6 is obtained using IDS as

$$X_6: \{(\mathrm{H}, 0.0679), (\mathrm{M}, 0.2402) \quad (\mathrm{L}, 0.6062), (\mathrm{Unknown}, 0.0856)\}$$

and the confidence score is given by 0.2309.

"Unknown" in the above result means that the output is also incomplete due to the incomplete input from X_5.

Since the assessment for X_8 has already been generated from X_1 and X_7, the output for X_9 can be inferred from the incomplete input from X_6 and the complete input from X_8. Due to the incomplete input from X_6, the belief degree of the relevant rules for X_9 needs to be updated to reflect the incompleteness, which are associated with X_6. Using (8)

$$\beta_{ik} = \overline{\beta}_{ik} \frac{\sum_{t=1}^{2} \left(\tau(t,k) \sum_{j=1}^{3} \alpha_{tj} \right)}{\sum_{t=1}^{2} \tau(t,k)} = \overline{\beta}_{ik} \frac{1.9144}{2}$$
$$= \overline{\beta}_{ik} * 0.9572, \quad i = 1,2,3;\ k = 28, \ldots, 36.$$

Therefore, $0 < \sum_{i=1}^{3} \beta_{ik} < 1$ for all rules associated with X_6. The sub-rule base for X_9 with the updated belief degrees is shown in Table VI.

Finally, X_9 can be inferred from the incomplete input from X_6 and the complete input from X_8 using IDS as

$$X_9: \{(\mathrm{H}, 0.1151), (\mathrm{M}, 0.248), (\mathrm{L}, 0.5998) \quad (\mathrm{Unknown}, 0.0371)\}$$

which is depicted as in Fig. 4, and the confidence score of X_9 is 0.2576 generated by using (12c).

VI. CONCLUSION

In this paper, existing knowledge representation and inference schemes were first investigated, and a new generic knowledge representation scheme was then proposed, based on the theory of evidence and fuzzy set theory, where a belief rule base was designed to capture uncertainty and nonlinear causal relationships. The inference process of such a rule-based system was characterized by a belief rule expression matrix and was implemented using the evidential reasoning (ER) approach. The methodology was further extended to the inference of hierarchical knowledge bases. A numerical example was used to illustrate the application of the proposed methodology.

In this methodology, various types of information from different sources can be transformed and used in the inference process. One of the unique features of the methodology is that if vague information coexists with ignorance or incompleteness caused by evidence not strong enough to make simple true or false judgments but with degrees of belief, the new methodology can provide a flexible and effective way to represent and a rigorous procedure to deal with such hybrid uncertain assessment information to arrive at rational conclusions. The evidence theory combined with fuzzy logic equips the new methodology with a flexible and versatile framework to represent not only precise data but also vagueness and ignorance in knowledge. The weights of antecedent attributes and rules were also taken into account in the framework to provide a more informative representation of knowledge and reflect the dynamic nature of decision-making problems.

Furthermore, the inference is implemented using the rigorous yet pragmatic ER approach, which establishes a nonlinear relationship between antecedent attributes and an associated consequent attribute. This is regarded as an advantage over traditional approaches, which suffer from lack of mathematical foundation. Moreover, this new methodology provides scopes and flexibility for rule training and self-learning/updating in a rule base.

Different from most conventional rule-base inference methods, the rule-base inference methodology using the evidential reasoning (RIMER) approach is characterized with certain unique features. First, each input can be represented as a distribution on referential values using a belief structure. The main advantage of doing so is that precise data, random numbers, and subjective judgments with uncertainty can be consistently modeled under the unified framework. Second, the ER approach provides a novel procedure for aggregating rules, which can preserve the original features of various types of information.

It is expected that the proposed RIMER approach can be used for rule-based evaluation analysis, classification, diagnosis, and decision making in a range of engineering, management, and medical fields. The application of the methodology to the rule-based safety analysis and synthesis of maritime systems is reported by Liu *et al.* in [20] and [21].

The problem of consistency of a rule base depends on how to generate rules. It would be usually thought to be trivial if rules are extracted from expert knowledge. However, if rules are generated from a set of data affected by noise, this problem can become serious. Consistency between generated rules with the intuition and common sense of human beings needs to be further investigated. For instance, rules are regarded to be inconsistent if they have very similar premise parts but possess significantly different consequents that conflict with expert knowledge or heuristics.

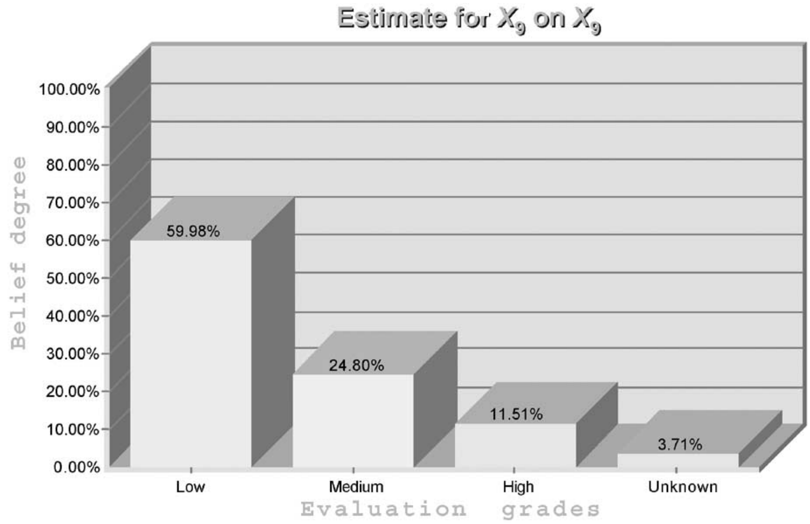


Fig. 4. Evaluation on X_9 based on the incomplete input.

ACKNOWLEDGMENT

The authors are grateful to the four anonymous referees for their constructive comments that have helped to improve the quality of the paper to its current standard.

REFERENCES

[1] R. Sun, "Robust reasoning: Integrating rule-based and similarity-based reasoning," *Artif. Intell.*, vol. 75, no. 2, pp. 241–295, 1995.
[2] P. L. Hayes, "In defence of logic," in *Proc. Int. Joint Conf. Artificial Intelligence (IJCAI)*, Cambridge, MA, 1977, pp. 559–565.
[3] J. Chomskey, "Rules and representation," *Behav. Brain Sci.*, vol. 3, no. 1, pp. 1–16, 1980.
[4] M. Nilsson, *Principle of Artificial Intelligence*. San Mateo, CA: Tiogo, 1980.
[5] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*. San Mateo, CA: Morgan Kaufmann, 1988.
[6] L. A. Zadeh, "Fuzzy logic," *Computer*, vol. 21, no. 4, pp. 83–93, 1988.
[7] S. J. Henkind and M. C. Harrison, "An analysis of four uncertainty calculi," *IEEE Trans. Syst. Man Cybern.*, vol. 18, no. 5, pp. 700–714, Sep.–Oct. 1988.
[8] D. A. Clark, "Numerical and symbolic approaches to uncertainty management in AI," *Artif. Intell. Rev.*, vol. 4, no. 2, pp. 109–146, 1996.
[9] S. Parson, "Current approaches to handling imperfect information in data and knowledge bases," *IEEE Trans. Knowl. Data Eng.*, vol. 8, no. 3, pp. 353–372, Jun. 1996.
[10] J. Hodges, S. Bridges, C. Sparrow, B. Wooley, B. Tang, and C. Jun, "The development of an expert system for the characterization of containers of contaminated waste," *Expert Syst. Appl.*, vol. 17, no. 3, pp. 167–181, 1999.
[11] H.-J. Zimmermann, "An application-oriented view of modeling uncertainty," *Eur. J. Oper. Res.*, vol. 122, no. 2, pp. 190–199, 2000.
[12] J. Liu, J. B. Yang, J. Wang, and H. S. Sii, "Review of uncertainty reasoning approaches as guidance for maritime and offshore safety-based assessment," *J. U.K. Safety Reliab. Soc.*, vol. 23, no. 1, pp. 63–80, 2003.
[13] E. Binaghi and P. Madella, "Fuzzy Dempster–Shafer reasoning for rule-based classifiers," *Int. J. Intell. Syst.*, vol. 14, no. 6, pp. 559–583, 1999.
[14] J. B. Yang, Y. M. Wang, and D. L. Xu, "The evidential reasoning approach for MCDA under both probabilistic and fuzzy uncertainties," *Eur. J. Oper. Res.*, to be published.
[15] J. B. Yang and M. G. Singh, "An evidential reasoning approach for multiple attribute decision making with uncertainty," *IEEE Trans. Syst. Man Cybern.*, vol. 24, no. 1, pp. 1–18, Jan. 1994.
[16] J. B. Yang and P. Sen, "A general multi-level evaluation process for hybrid MADM with uncertainty," *IEEE Trans. Syst. Man Cybern.*, vol. 24, no. 10, pp. 1458–1473, Oct. 1994.
[17] J. B. Yang, "Rule and utility based evidential reasoning approach for multi-attribute decision analysis under uncertainties," *Eur. J. Oper. Res.*, vol. 131, no. 1, pp. 31–61, 2001.
[18] J. B. Yang and D. L. Xu, "On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty," *IEEE Trans. Syst. Man Cybern., A, Syst. Humans*, vol. 32, no. 3, pp. 289–304, May 2002.
[19] ——, "Nonlinear information aggregation via evidential reasoning in multi-attribute decision analysis under uncertainty," *IEEE Trans. Syst. Man Cybern., A, Syst. Humans*, vol. 32, no. 3, pp. 376–393, May 2002.
[20] J. Liu, J. B. Yang, J. Wang, H. S. Sii, and Y. M. Wang, "Fuzzy rule-based evidential reasoning approach for safety analysis," *Int. J. Gen. Syst.*, vol. 33, no. 2–3, pp. 183–204, 2004.
[21] J. Liu, J. B. Yang, J. Wang, and H. S. Sii, "Engineering system safety analysis and synthesis using fuzzy rule-based evidential reasoning approach," *Qual. Reliab. Eng. Int.*, vol. 21, no. 4, pp. 387–411, 2005.
[22] L. Godo, R. López de Mántaras, J. Puyol-Gruart, and C. Sierra, "Renoir, Pneumon-IA and Terap-IA: Three medical applications based on fuzzy logic," *Artif. Intell. Med.*, vol. 21, no. 1–3, pp. 153–162, 2001.
[23] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 69, pp. 338–353, 1965.
[24] ——, "The concept of a linguistic variable and its application to approximate reasoning," *Inf. Sci.*, vol. 8, no. 3, pp. 199–249, 1975; pp. 301–357; vol. 9; pp. 43–80.
[25] P. P. Bonissone and R. M. Tong, "Reasoning with uncertainty in expert systems," *Int. J. Man-Mach. Stud.*, vol. 22, no. 3, pp. 241–250, 1985.
[26] A. P. Dempster, "A generalization of Bayesian inference," *J. R. Stat. Soc., Ser. B*, vol. 30, no. 2, pp. 205–247, 1968.
[27] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton Univ. Press, 1976.
[28] P. Smets, "Belief functions," in *Non-Standard Logics for Automated Reasoning*, P. Smets, E. H. Mamdani, D. Dubois, and H. Prade, Eds. New York: Academic, 1988, pp. 253–286.
[29] R. R. Yager, J. Kacprzyk, and M. Fedrizzi, *Advances in the Dempster–Shafer Theory of Evidence*. New York: Wiley, 1994.
[30] R. R. Yager, "Modeling uncertainty using partial information," *Inf. Sci.*, vol. 121, no. 3–4, pp. 271–294, 1999.
[31] P. Wallery, "Measures of uncertainty in expert systems," *Artif. Intell.*, vol. 83, no. 1, pp. 1–58, 1996.
[32] S. Anand, D. A. Bell, and J. G. Hughes, "EDM: A general framework for data mining based on evidence theory," *Data Knowl. Eng.*, vol. 18, no. 3, pp. 189–223, 1996.
[33] T. Denoeux, "Analysis of evidence-theoretic decision rules for pattern classification," *Pattern Recogn.*, vol. 30, no. 7, pp. 1095–1107, 1997.
[34] ——, "Reasoning with imprecise belief structures," *Int. J. Approx. Reason.*, vol. 20, no. 1, pp. 79–111, 1999.
[35] E. Hullermeier, "Similarity-based inference as evidential reasoning," *Int. J. Approx. Reason.*, vol. 26, no. 2, pp. 67–100, 2001.
[36] W. R. Liu and J. Hong, "Reinvestigating Dempster's idea on evidence combination," *Knowl. Inf. Syst.*, vol. 2, no. 2, pp. 223–241, 2000.

[37] J. B. Yang and P. Sen, "Multiple attribute design evaluation of large engineering products using the evidential reasoning approach," *J. Eng. Des.*, vol. 8, no. 3, pp. 211–230, 1997.

[38] M. Ishizuka, K. S. Fu, and J. T. P. Ya, "A rule-based inference with fuzzy sets for structural damage assessment," in *Approximate Reasoning in Decision Analysis*, M. M. Gupta and E. Sanchez, Eds. Amsterdam, The Netherlands: North Holland, 1982, pp. 261–268.

[39] ——, "Inference procedure under uncertainty for the problem-reduction method," *Inf. Sci.*, vol. 28, no. 3, pp. 179–206, 1982.

[40] H. Ogawa, K. S. Fu, and J. T. P. Yao, "An inexact inference for damage assessment of existing structures," *Int. J. Man-Mach. Stud.*, vol. 22, no. 3, pp. 295–306, 1985.

[41] L.-H. Chen, "An extended rule-based inference for general decision-making problems," *Inf. Sci.*, vol. 102, no. 1–4, pp. 247–261, 1997.

[42] R. R. Yager, "Generalized probabilities of fuzzy events from belief structures," *Inf. Sci.*, vol. 28, no. 1, pp. 45–62, 1982.

[43] R. R. Yager and D. P. Filev, "Including probabilistic uncertainty in fuzzy logic controller modeling using Dempster–Shafer theory," *IEEE Trans. Syst. Man Cybern.*, vol. 25, no. 8, pp. 1221–1230, Aug. 1995.

[44] J. F. Baldwin, T. P. Martin, and B. W. Pilsworth, *Fril—Fuzzy and Evidential Reasoning in AI*. New York: Research Studies/Wiley, 1995.

[45] T. L. Saaty, *The Analytic Hierarchy Process*. Pittsburgh, PA: Univ. Pittsburgh Press, 1988.

[46] G. Islei and A. G. Lockett, "Judgmental modeling based on geometric least square," *Eur. J. Oper. Res.*, vol. 36, no. 1, pp. 27–35, 1988.

[47] J. Barzilai, "Deriving weights from pairwise comparison matrices," *J. Oper. Res. Soc.*, vol. 48, no. 12, pp. 1226–1232, 1997.

[48] B. Schweizer and A. Sklar, *Probabilistic Metric Spaces*. Amsterdam, The Netherlands: North Holland, 1983.

[49] H.-J. Zimmermann, *Fuzzy Set Theory and Its Applications*. Norwell, MA: Kluwer, 1996.

[50] E. Hisdal, *Logical Structures for Representation of Knowledge and Uncertainty*. Hiedelberg, Germany: Physica-Verlag, 1998.

[51] N. Kasabov, *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*. Cambridge, MA: MIT Press, 1998.

[52] J.-L. Marichal, "On the associativity functional equation," *Fuzzy Sets Syst.*, vol. 114, no. 3, pp. 381–390, 2000.

[53] G. Beliakov and J. Warren, "Appropriate choice of aggregation operators in fuzzy decision support systems," *IEEE Trans. Fuzzy Syst.*, vol. 9, no. 6, pp. 773–784, Dec. 2001.

[54] P. Orponen, "Dempster's rule of combination is # P-complete," *Artif. Intell.*, vol. 44, no. 1–2, pp. 245–253, 1990.

[55] S. Gottwald and W. Perdycz, "On the methodology of solving fuzzy relational equations and its impact on fuzzy modeling," in *Fuzzy Logic in Knowledge-Based Systems, Decision and Control*, M. M. Gupta and T. Yamakawa, Eds. New York: Elsevier, 1988, pp. 197–210.

[56] S.-M. Chen, M.-S. Yeh, and P.-Y. Hsiao, "A comparison of similarity measures of fuzzy values," *Fuzzy Sets Syst.*, vol. 72, no. 1, pp. 79–89, 1995.

[57] W.-J. Wang, "New similarity measures on fuzzy sets and on elements," *Fuzzy Sets Syst.*, vol. 85, no. 3, pp. 305–309, 1997.

[58] J. Wang, J. B. Yang, and P. Sen, "Safety analysis and synthesis using fuzzy set modelling and evidential reasoning," *Reliab. Eng. Syst. Saf.*, vol. 47, no. 3, pp. 103–118, 1995.

[59] E. Jacquet-Lagreze and J. Siskos, "Assessing a set of additive utility function for multicriteria decision making the UTA method," *Eur. J. Oper. Res.*, vol. 10, no. 2, pp. 151–164, 1982.

[60] J. B. Yang and P. Sen, "Preference modeling by estimating local utility functions for multiobjective optimization," *Eur. J. Oper. Res.*, vol. 95, no. 1, pp. 115–138, 1996.

[61] H. Bustince and P. Burillo, "Mathematical analysis of interval-valued fuzzy relations: Application to approximate reasoning," *Fuzzy Sets Syst.*, vol. 113, no. 2, pp. 205–219, 2000.

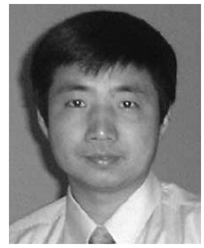
[62] S.-M. Chen and W.-H. Hsiao, "Bidirectional approximate reasoning for rule-based systems using interval-valued fuzzy sets," *Fuzzy Sets Syst.*, vol. 113, no. 2, pp. 185–203, 2000.

[63] E. D. Cox, "*The Fuzzy Systems Handbook: A Practitioner's Guide to Building, Using, and Maintaining Fuzzy Systems*," 2nd ed. Boston, MA: Academic, 1998.

[64] J. Hodges, S. Bridge, and S. Y. Yie, "Preliminary results in the use of fuzzy logic for radiological waste characterization expert system," Mississippi State Univ., MS, Tech. Rep. MSU-960626, Jun. 26, 1996.

[65] E. H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *Int. J. Man-Mach. Stud.*, vol. 7, no. 1, pp. 1–13, 1975.

[66] J. B. Yang and D. L. Xu, *Intelligent Decision System Via Evidential Reasoning, version 1.1*. Cheshire, U.K.: IDSL, 1999.



Jian-Bo Yang received the B.Eng. and M.Eng. degrees in control engineering from the North Western Polytechnic University, Xi'an, China, in 1981 and 1984, respectively, and the Ph.D. degree in systems engineering from Shanghai Jiao Tong University, Shanghai, China, in 1987.

He is a Professor of Decision and System Sciences and the Head (2003–2004) of the Decision Sciences and Operations Management Group, Manchester Business School, University of Manchester, U.K. He is also a Specially Appointed Professor of Huazhong University of Science and Technology, Wuhan, Hubei, China. Prior to his current appointment, he was a faculty member of the University of Birmingham, Birmingham, U.K., (1995–1997), the University of Newcastle-upon-Tyne, Newcastle-upon-Tyne, U.K., (1991–1995), University of Manchester Institute of Science and Technology (UMIST) (1990), and Shanghai Jiao Tong University (1987–1989). In the past two decades, he has been conducting research in multiple criteria decision analysis under uncertainty, multiple objective optimization, intelligent decision support systems, hybrid quantitative and qualitative decision modeling using techniques from operational research, artificial intelligence, and systems engineering, and dynamic system modeling, simulation, and control for engineering and management systems. His current applied research is in design decision-making, risk and safety analysis, production planning and scheduling, quality modeling and evaluation, supply chain modeling and supplier assessment, and the integrated evaluation of products, systems, projects, policies, etc. His current research has been supported by EPSRC, EC, SERC, and HKRGC. He has published widely and developed several software packages in these areas including the Windows-based intelligent decision system (IDS) via evidential reasoning.

Jun Liu received the B.Sc. and M.Sc. degrees in applied mathematics in 1993 and 1996, respectively, and the Ph.D. degree in information engineering and control in 1999 from the Southwest Jiaotong University, Chengdu, China.

He is currently a Lecturer at the School of Computing and Mathematics, University of Ulster at Jordanstown, Newtownabbey, Northern Ireland, U.K. Prior to his current appointment, he worked as a Research Associate at Manchester Business School of the University of Manchester, Manchester, U.K. He also worked as a postdoctoral staff member at the Belgian Nuclear Research Center (SCK • CEN), Mol, Belgium (March 2000 to February 2002). His research is focused on algebraic and nonclassical logic (many-valued logic, fuzzy logic), automated reasoning theory and methods, fuzziness and uncertainty theory and applications, intelligent decision support systems and evaluation analysis, information fusion, and data combinations. His current applied research covers intelligent decision analysis and support under uncertainties for design decision making and risk and safety analysis in engineering fields.

Jin Wang (S'03–A'04–M'04) received the B.Sc. degree in marine automation from Dalian Maritime University, Dalian, China, in 1983 and the M.Sc. degree in marine engineering and the Ph.D. degree (via staff registration) in marine safety engineering from the University of Newcastle-upon-Tyne, Newcastle-upon-Tyne, U.K., in 1989 and 1994, respectively.

He is a Professor of Marine Technology at the School of Engineering at Liverpool John Moores University, Liverpool, U.K. He has been involved in marine and offshore safety research for the past 14 years, with support from the EPSRC, EU, HSE, etc. His major research interests include safety and reliability-based design and operations of large marine and offshore systems, probabilistic and nonprobabilistic safety analysis and decision making, port safety assessment, and analysis of safety-critical systems.

How-Sing Sii received the bachelor's degree in mechanical engineering from the University of South Australia, Adelaide, Australia, the master's degree in engineering science from the University of Malaya, Kuala Lumpur, Malaysia, and the Ph.D. degree in marine and mechanical engineering from Staffordshire University, Staffordshire, U.K.

He has more than 15 years of industrial experience in various disciplines. He was a Postdoctoral Research Fellow at Liverpool John Moores University, Liverpool, U.K., working on two U.K. EPSRC-funded offshore safety research projects from 2000 to 2003. He is currently an academic member of the Institute BREW, Sarawak, Malaysia. He is also a safety consultant in the rail industry. His research and consultation interests are in the areas of quality, reliability, and safety engineering and management. He has published more than 50 technical papers in international conference proceedings and journals.

Hong-Wei Wang (M'04–A'05) received the Ph.D. degree in systems engineering from Huazhong University of Science and Technology, Wuhan, Hubei, China, in 1993.

He is a Professor at the Department of Control Science and Engineering, Huazhong University of Science and Technology. He is the Head of Department of Control Science and Engineering and the Director of the Institute of Systems Engineering. His research interests are in the fields of supply chain management, decision making and simulation of logistics management, and decision support systems.