A New Prediction Model Based on Belief Rule Base for System's Behavior Prediction

Xiao-Sheng Si, Chang-Hua Hu, Jian-Bo Yang, and Zhi-Jie Zhou

***Abstract*—In engineering practice, a system's behavior constantly changes over time. To predict the behavior of a complex engineering system, a model can be built and trained using historical data. This paper addresses the forecasting problems with a belief rule base (BRB) to trace and predict system performance in a more interpretable and transparent way. More precisely, it extends the BRB method to handle a system's behavior prediction, and a new prediction model based on BRB is presented, which can model and analyze prediction problems using not only numerical data but human judgmental information as well. The proposed forecasting model includes some unknown parameters that can be manually tuned and trained. To build an effective BRB forecasting model, a multiple-objective optimization model is provided to locally train the BRB prediction model by minimizing the mean square error (MSE). Finally, a practical case study is provided to illustrate the detailed implementation procedures and examine the feasibility of the proposed approach in engineering application. Furthermore, the comparative studies with other state-of-the-art prediction methods are carried out. It is shown that the proposed model is effective and can generate better prediction in terms of accuracy, as well as comprehensibility.**

***Index Terms*—Belief rule base (BRB), evidential-reasoning (ER) approach, expert system, nonlinear optimization, prediction.**

I. Introduction

A. Motivation

SAFE and reliable operation of technical systems is of great significance for the protection of human life and health, the environment, and the economy. The correct functioning of these systems also has a profound impact on production cost and product quality. Recently, many serious accidents have happened in the world where systems have been lager scale and complex, and they not only caused heavy damage and a social sense of instability but brought an unrecoverable bad influence on the living environment as well. Such disasters as the crash of a Chinese early warning aircraft, the sinking of a Russian submarine, and the explosion of "*Deepwater Horizon*" in the Gulf of Mexico could have been avoided if a sound system's behavior forecasting mechanism had been enforced. These are said to have occurred from various sources of equipment deterioration or characteristic parameters, such as the failure of the "*last line of defence*," i.e., the so-called blowout preventer in "*Deepwater Horizon*." As such, it is very important to predict the future behavior of the systems or characteristic parameters.

In the area of forecasting, a wide variety of prediction techniques have been introduced to date, going from simple linear regression to advanced nonlinear prediction methods. Although nonlinear techniques typically provide the most-accurate predictive models, they are often not suitable to be used in many practical application domains because of their lack of transparency and comprehensibility. In domains where validation of the underlying model is required, especially in safety-critical domains, a clear insight into the reasoning process of the prediction model is necessary and desired [11]. On the other side, prediction implies dealing with prognostic uncertainty partly or fully because the information is obtained as absent, incomplete, insufficient, imprecise, ambiguous, vague, etc. The acknowledgment of prediction uncertainty has become commonplace in the prediction studies community [47]. Thus far, several frameworks have been constructed to deal with uncertainty involved [64], such as Bayesian probability theory, Dempster–Shafer (D–S) theory [7]–[9], [40], and fuzzy set theory [2], [10], [67]. Each of these frameworks is aimed at a specific application environment and has its own features. For instance, Bayesian probability theory is suitable for handling with probabilistic uncertainty, and D–S theory is suitable for dealing with vagueness and incompleteness existing in knowledge, while fuzzy set theory is suitable for fuzzy uncertainty. In fact, a variety of uncertainties may coexist in real systems, e.g., fuzzy information may coexist with ignorance, thereby leading to the induction of knowledge without certainty, but only with degrees of belief or credibility [64]. Under such circumstance, it is desirable to develop a new prediction model to deal with different kinds of uncertainty in prediction. In the literature, a belief rule base (BRB) proposed in [64] is a generic rule-base inference methodology to handle hybrid information with uncertainty in general, and to deal with fuzziness and ignorance in data in particular, by combining fuzzy set theory with D–S theory. This methodology establishes a nonlinear relationship between antecedent attributes and an associated consequent and aims to handle hybrid information with uncertainty in knowledge inference [66].

Manuscript received November 13, 2009; revised March 17, 2010 and September 29, 2010; accepted February 8, 2011. Date of publication March 22, 2011; date of current version August 8, 2011. This work was supported by the National Natural Science Foundation of China under Grant 60736026 and Grant 61004069, the National Science Fund for Distinguished Young Scholars of China under Grant 61025014, and the Program for New Century Excellent Talents in University of China under Grant 07144.

X.-S. Si is with Xi'an Institute of Hi-tech, Xi'an 710025, China, and also with the Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: sxs09@mails.tsinghua.edu.cn).

C.-H. Hu and Z.-J. Zhou are with Xi'an Institute of Hi-tech, Xi'an 710025, China (e-mail: hch6603@263.net; zhouzj04@mails.tsinghua.edu.cn).

J.-B. Yang is with the Manchester Business School, University of Manchester, Manchester, M15 6PB, U.K. (e-mail: jian-bo.yang@mbs.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TFUZZ.2011.2130527

Together with these analyses, we can see that one thing shared in common by the prediction and BRB is to deal with uncertainty encountered in inference process. Furthermore, most of current applications of BRB are focused on decision making. Therefore, in this paper, we develop a new prediction model based on BRB to achieve behavior prediction, which can inherit the merits of BRB in dealing with uncertainty existing in data. This is the main motivation of this paper.

B. *Survey Over the Related Works*

Up to now, many prediction mechanisms and mathematical models have been proposed to achieve prediction by a large community of researchers [1], [5], [18], [21], [25], [29], [36], [44], [50], [60], [71]. These existing methods can be broadly classified into three categories [33], [38]: physical-model-based methods, qualitative-knowledge-based methods, and data-driven methods, in which the data-driven models usually include machine-learning methods and statistical data-driven methods. A state-of-the-art report on selecting prediction technologies has been given in [38]. Physical-model-based methods, such as Kalman predictor [61], strong tracking predictor [50], and particle predictor [3], use some of engineering laws to establish the prognostic models. As a premise, the physical model must be known *a priori* and accurate for these methods to be highly effective. However, the correct and accurate mathematical models of engineering systems are often difficult to obtain, since nonlinearity and uncertainty widely exist in system models. Qualitative-knowledge-based methods can be used to predict the future behaviors if some expert knowledge about the systems is available [1], [60]. As it is known, knowledge from domain-specific experts is usually inexact and experts' preferences are likely to be incomplete and uncertain. Therefore, reasoning on knowledge is often imprecise. In such a case, knowledge-based methods have limited capabilities for modeling both the observed data and uncertain knowledge and providing more robust problem solving. In contrast, data-driven prediction methods have gained much attention and have become a popular approach of prognostics and health management with success. These methods are aimed mainly at building predictive data models, adapting internal parameters of the data models to account for the known (training) data samples, and allowing for predictions to be made on the unknown (test) data samples [11]. Many of well-established methods in this category have been proposed to achieve prognostic applications, such as improved autoregressive moving-average-model (ARMA) methods [18], [36], gray methods [21], [25], neural networks [37], [68], [70], [71], support-vector machine (SVM) [5], [44], etc. However, most of the data-driven methods are modeled in a black-box style, and they achieve high accuracy using a large number of numerical parameters in a way that is incomprehensible to humans [11]. This leads to several dangers, such as overfitting and difficulty in explanation of their results. Therefore, the use of such black-box models in forecasting applications that require systematic reasoning and offer explanations of their results may not be possible. Especially, in safety-critical domains, such as nuclear power station or weapon control loop, black-box prediction is not satisfactory, and such risks may not be acceptable. In addition, the current version of data-driven methods only uses either numerical or subjective information of expert [42]. In such case, it is highly desirable to develop a nonblack-box prediction model that can model and analyze prediction problems using both numerical data and human judgmental information subject to uncertainty, and further provide some comprehensible prediction results as well as model parameters.

To the best of our knowledge, reasoning with logical rules is more acceptable to human users than the prediction made by black-box models, because such reasoning is comprehensible, provides explanations, and can be validated by human inspection. It can increase reliability of the system as well, and may help to discover important relationships and combination of features, if the expressive power of these rules is sufficient for that [11]. Yang *et al*. [64] proposed a generic rule-base inference methodology using the evidential-reasoning (ER) approach (RIMER) to handle hybrid information with uncertainty in human decision making [43], [49]. Since the main components of RIMER are belief rules, it is also termed as BRB system. The inference engine of BRB system is the ER approach, which is based on D–S theory [7]–[9], [40], decision theory [20], and fuzzy set [2], [67]. Compared with the traditional IF–THEN rule base [43] and the fuzzy IF–THEN rule base [4], [10], [17], [23], [29], the BRB approach provides a more informative and realistic scheme for knowledge representation and models numerical data and human judgmental information in a unified way. For example, in safety analysis and performance assessment, system's behavior can be described linguistically due to its easy interpretation. One may use such linguistic terms as "*low*," "*average*," and "*high*." Therefore, it is common to assess behavior level of system by degrees to which it belongs to such linguistic variables as "*low*," "*average*," and "*high*" that are referred to as behavior expression. For example, an expert may state that he is 20% sure that system's behavior is low and 80% sure that it is high. In the statement, "high" and "low" denote distinctive evaluation grades, and the percentage values of 20 and 80 are referred to as the degrees of belief, which indicate the extents that the corresponding grades are assessed to. The above assessment can be expressed as the following expression: $S(\text{behavior}) = \{(low, 0.2), (high, 0.8)\}$, where $S(\text{behavior})$ stands for the state of the system's behavior, and the real numbers 0.2 and 0.8 denote the degrees of belief of 20% and 80%, respectively. Among the current methods, BRB can provide such default form of the output structure so that the analysis results can provide a complete picture about the system's behavior state. In addition, all parameters in the BRB prediction model have a clear meaning in knowledge representation. For example, the attribute weight reflects the knowledge we have obtained for the belief rule. If the attribute weight is less than 1.0, then it represents that the corresponding rule has a low credibility, and some uncertainties exist in assessment of this rule. To strengthen the capability of BRB system, Yang *et al*. [66] proposed several parameter-estimation methods to train BRB system. Generally, there are several means of improving the interpretability of rule-based system, such as adjusting the structure of the rule-based system by rule reduction or rule simplification in several well-structured works in the literature [13], [14], [39] and improving the ability of knowledge representation of rule-based system [48], in which a probabilistic fuzzy system (PFS) was developed by van den Berg *et al*. to deal with the probabilistic and fuzzy uncertainties. The BRB method used in our paper belongs to the latter, i.e., improving the ability of knowledge representation. It is worth noting that the BRB model is similar to the PFS in [48]. However, BRB method differs from the PFS in two essential respects. First, the PFS is designed to combine the interpretability of the fuzzy system with the statistical properties of probabilistic systems. As such, it is suitable for dealing with the probabilistic and fuzzy uncertainties. However, as discussed above, BRB model is designed to handle hybrid information with uncertainty, such as incompleteness, fuzziness, imprecision, vagueness, etc. In addition, Yang *et al*. [64] have shown that the ER algorithm, which serves as inference engine of BRB, can also deal with the probabilistic and fuzzy uncertainties. The reasons for this are twofold. On the one hand, in the original interpretation of belief by Dempster [7], the terms lower and upper probability are adopted and the D–S theory can be considered as a generalization of probability theory, as discussed in [8]. This viewpoint was long ago recognized and was recently reviewed again by Dempster [9]. Some authors in [15] and [19] have given probabilistic interpretation of the D–S evidence theory either. For example, Halpern and Fagin [15] proposed that there were two useful and quite different ways of interpreting belief functions. The first is that a belief function is interpreted as a generalized probability function, and the second is that a belief function is used as a way for representation of evidence. In the used BRB model, the ER algorithm, which is based on the D–S theory, serves as the inference engine. Therefore, the output results of the BRB also have a belief structure as well as the combination results of the D–S theory. On the other hand, in the BRB model, the belief is assigned to the single element of the assessment framework using rule-based information-transformation technique [62]. With another viewpoint, the belief degree in such case also has a probabilistic interpretation according to the pignistic transformation developed by Smets and Kennes [41]. Therefore, we can conclude that BRB model can also deal with the probabilistic and fuzzy uncertainties. In this sense, BRB model is more general than the PFS [48] in dealing with knowledge uncertainty. Second, in the PFS, in order to incorporate the statistical properties of probabilistic systems, additional works are needed to approximate a probabilistic distribution by fuzzy-histogram technique. In contrast, rule-based transformation technique in BRB model is easier to implement than fuzzy-histogram technique in the PFS. Due to limited space, we do not summarize the rule-based transformation technique into this paper. For details, see [61] and [64]. Therefore, the BRB system provides a nonblack-box modeling approach since its parameters and output are comprehensible, and its structure can be directly adjusted according to expert intervention. Besides the above development, Liu *et al*. [26] presented a fuzzy version of BRB inference system that had the same structure in belief rule but considered fuzziness existing in decision-making problems. Thus far, the BRB system has already been applied to the safety analysis of offshore systems [27], the leak detection of oil pipelines [58], and multiple attribute decision analysis [64], etc. However, the current studies on the BRB systems mainly aimed at decision-making problems that are not appropriate to forecast the future behavior. At present, no attempt has been directly made to address the issue of how to deal with system's behavior prediction problems using BRB system.

C. Our Approach

In this paper, being motivated by the RIMER [64] and the leaning algorithms developed in [66], a new prediction model based on BRB system is developed in a more interpretable and transparent way in terms of knowledge representation and interpretation of the output results, which can model and analyze prediction problems using not only numerical data, but human judgmental information with uncertainty as well. In order to establish an effective BRB forecasting model, a multiple-objective optimization model is provided for locally training the BRB prediction model by minimizing the mean square error (MSE). In order to demonstrate the usefulness of the proposed model, we apply our method to analyze the behavior of dynamically tuned gyro (DTG), which is the main device of the inertial navigation systems (INSs) used in weapon systems and space equipments. The DTG is a sensor fixed in the INS to measure angular velocity. The DTG used here is a mechanical structure. When the INS is operating, the rotating wheel of DTG with very high speed can lead to rotation axis wear and, finally, result in gyro's drift. With accumulation of wear, the drift degrades and, finally, results in the failure of DTG. For a general description of an INS and gyros, see [55]. As such, the drift of DTGs is usually used as a performance indicator to evaluate the health condition of DTG and INS. In our case study, we used drift data of the DTG to predict its behavior. Overall, the main contributions of this paper are that the traditional BRB method is extended to handle system's behavior prediction, and a prediction model based on BRB is established with an optimization method tuning unknown parameters. Unlike other data-driven methods, the BRB prediction model is interpretable and transparent in terms of inference process and model parameters with the predicted results. Experimental studies are conducted to compare against several well-established methods in the field of prediction, such as the hybrid model of ARMA and generalized autoregressive conditional heteroscedasticity (ARMA/GARCH) [36], radial-basis-function neural networks (RBFNNs) [68], least-square SVMs (LSSVMs) [44], and fuzzy-c-means-clustering-based Takagi–Sugeno model (FCM-TS) [23]. The results show that the proposed model is effective and can generate better prediction in terms of accuracy as well as comprehensibility.

This paper is organized as follows. In Section II, the forecasting model of the system's behavior is constructed on the basis of the BRB. Section III presents a multiple-objective optimization model for training the forecasting model under belief distributed outputs. A practical case study is presented to verify the proposed approach in Section IV. Section V presents some conclusions and discussions.

TABLE I
PATTERN OF TRAINING DATASET

X				y
x_1	x_2	…	x_p	y_{p+1}
x_2	x_3	…	x_{p+1}	y_{p+2}
⋮	⋮	⋮	⋮	⋮
x_{M-p-1}	…	…	x_{M-1}	y_M

II. Behavior Prediction Model Based on Belief Rule Base

A. Model Structure and Representation

In this section, the prediction model is investigated in the BRB framework. It is assumed that a set of observed data is provided in the form of input–output pairs $(\mathbf{X}(t_m), y(t_m)), m = 1, \ldots, M$, with $\mathbf{X}(t_m)$ being an input vector of the actual system at time t_m and $y(t_m)$ being a scalar or vector representing the corresponding output value or subjective distribution value of the actual system at time t_m. In this paper, the input vector $\mathbf{X}(t_m)$ consists of the antecedent attributes associated with system output $y(t_m)$ at time t_m.

In prediction problem, the inputs used in prediction model are the past input vector, i.e., lagged observations of the current time, while the outputs are the future values. Each set of input patterns is composed of any moving fixed-length window within the time series of the input data. The general prediction model can be represented as

$$\hat{y}(t+k-1) = f(x_{t-1}, x_{t-2}, \ldots, x_{t-p}) \tag{1}$$

where $\hat{y}(t+k-1)$ represents the predicted value at time t, $(x_{t-1}, x_{t-2}, \ldots, x_{t-p})$ is a vector of lagged variables, and p represents the dimensions of the input vector (number of input nodes) or the number of past inputs related to the future value. Although multistep prediction may capture some system dynamics, the performance may be poor due to the accumulation of errors. In practice, one-step-ahead prediction results are more useful since they provide timely information for preventive and corrective maintenance plans [57]. In addition, in views of Wang [52] and Zhang *et al*. [69], the more the step is ahead, the less reliable the forecasting operation. Thus, this study only considers one-step-ahead prediction, that is, $k = 1$, in (1). For simplicity, let $\mathbf{X}(t) = (x_{t-1}, x_{t-2}, \ldots, x_{t-p})$ denotes the input vector of the prediction model.

B. Construction of Belief-Rule-Base System for Behavior Prediction

In order to construct BRB prediction model, suppose $x_{t-1}, x_{t-2}, \ldots, x_{t-p}$ are p antecedent attributes associated with system-prediction output $\hat{y}_t$, and the BRB system approach attempts to identity the appropriate internal representation between $\mathbf{X}(t)$ and $\hat{y}_t$.

Table I shows how training patterns can be designed in the BRB prediction model. In Table I, p denotes the number of lagged variables (i.e., so-called embedding dimension), while $t-p$ represents the total number of data samples. Moreover, $\boldsymbol{X}$ is the input nodes and y denotes the predicted output node. Following successful training, the BRB prediction model can predict future outcomes y_t.

In the BRB prediction model, the belief rules can be designed as follows:

$$\begin{aligned} &R_k\text{: IF } x_{t-1} \text{ is } A_1^k \wedge x_{t-2} \text{ is } A_2^k \cdots \wedge x_{t-p} \text{ is } A_{t-p}^k \\ &\text{THEN } \hat{y}_t \text{ is } \{(D_1, \beta_{1,k}), \ldots, (D_N, \beta_{N,k})\} \text{ with a} \\ &\text{rule weight } \theta_k \text{ and attribute weight } \delta_{1,k}, \delta_{2,k}, \ldots, \delta_{M_k,k} \end{aligned} \tag{2}$$

where $x_1, x_2, \ldots, x_{t-p}$ represents the antecedent attributes in the kth rule. A_i^k (with $i = 1, \ldots, p, k = 1, \ldots, L$) is the referential value of the ith antecedent attribute in the kth rule and $A_i^k \in A_i$. $A_i = \{A_{i,j}, j = 1, \ldots, J_i\}$ is a set of referential values for the ith antecedent attribute and J_i is the number of the referential values. $\theta_k (\in \mathrm{R}^+, k = 1, \ldots, L)$ is the relative weight of the kth rule, and $\delta_{1,k}, \delta_{2,k}, \ldots, \delta_{p,k}$ are the relative weights of the M_k antecedent attributes used in the kth rule. $\beta_{j,k} (j = 1, \ldots, N, k = 1, \ldots, L)$ is the belief degree assessed to D_j, which denotes the jth consequent corresponding to the output value at time t. If $\sum_{j=1}^{N} \beta_{j,k} = 1$, the kth rule is said to be complete, else it is incomplete. In addition, suppose that p is the total number of antecedent attributes used in the rule base.

A BRB given in (2) represents functional mappings between antecedents and consequents possibly with uncertainty. It provides a more informative and realistic scheme than a simple IF–THEN rule base for knowledge representation. Note that the degrees of belief $\beta_{i.k}$ and the weights could be assigned initially by experts or by common sense, and then trained or updated using dedicated learning algorithms [66]. For simplicity, in the following, let us assume that the relative weight of ith antecedent attribute in belief rule k is the same as the relative weight of this antecedent attribute in belief rule l, where $k = 1, \ldots, L, l = 1, \ldots, L, k \neq l$. As such, $\delta_{i,k}$ can be represented as $\delta_i (i = 1, \ldots, p)$.

C. Strategy for Converting Available Data to Belief Structure

When applying a BRB, the input of an antecedent is transformed into a belief structure over the referential values of an antecedent. The distribution is then used to calculate the activation weights of the rules in the rule base. The main advantage of doing so is that precise data, random numbers, and subjective judgments with uncertainty can be consistently modeled under the same framework [59], [63], [64], [66]. In addition, the ER algorithm is the inference engine of BRB system, but the ER algorithm is suitable for dealing data in the format of a belief structure [59], [62], [63]. Due to the fact that the input data $\mathbf{X}(t)$ may be a numerical value or a subjective distribution, there is a need to transform these data into the belief structure. In [62], a rule-based scheme to convert other inputs has been developed to deal with the input information involving quantitative data. In this paper, rule-based transformation technique can be used for data transformation. For more discussions on this issue, see [62] and [64]. As a result, each input can be represented as a distribution on the referential values using a belief structure.

By using the rule-based transformation technique [62], the input vector $X(t) = (x_{t-1}, x_{t-2}, \ldots, x_{t-p})$ can be described as

TABLE II
REFERENTIAL POINTS OF GYROSCOPIC DRIFT

Linguistic terms	*Small* (S)	*Medium*(M)	*Large*(L)
Numerical values(° / *hour*)	1.0	1.4	2.4

a distribution on referential values using a belief structure as follows:

$$S(x_{t-i}) = \{(A_{i,n}, \alpha_{n,i}(x_{t-i})), n = 1, \ldots, J_i\}, \quad i = 1, \ldots, p \tag{3}$$

where $A_{i,n}$ is the nth referential value of the attribute x_{t-i}, $\alpha_{n,i}(x_{t-i})$ is the matching degree, which measures the matching degree of the ith input to its nth referential value, and J_i is the number of the referential values used to describe the ith antecedent attribute. From (3), the mass can only be assigned to the single referential point of the assessment framework. Therefore, the focal elements under such case are always the single grade in the assessment framework.

In (3), $\alpha_{n,i}(x_{t-i})$ could be obtained using different ways in hand, depending on the nature of an antecedent attribute and data available. Generally, there is a scheme to deal with different types of input information, as summarized below [64], [66].

1) *Quantitative attributes assessed using referential terms:* In this case, if the antecedent attribute x_{t-i} can be assessed by defining independent crisp sets, then $\alpha_{n,i}(x_{t-i})$ can be obtain through rule-based transformation technique [62]. For example, in the Section IV-C, the assessment framework for the model inputs is $A_i = \{S, M, L\} = \{A_{i,1}, A_{i,2}, A_{i,3}\}$, for $i = 1$ and 2. For intuitive illustration, we used the 35th data point in our dataset to show how we can obtain the belief structure. As for the 35th data point, the specific inputs are $1.22^\circ\ \mathrm{h}^{-1}$ and $1.26^\circ\ \mathrm{h}^{-1}$. Taking the first input, i.e., $1.22^\circ\ \mathrm{h}^{-1}$, for instance, we can convert $1.22^\circ\ \mathrm{h}^{-1}$ as $S(1.22) = \{(A_{1,1}, 0.45), (A_{1,2}, 0.55), (A_{1,3}, 0)\}$ through rule-based transformation technique and the referential points in Table II. Specifically, we have $\alpha_{1,1}(1.22) = (1.4 - 1.22)/(1.4 - 1)$, and $\alpha_{2,1}(1.22) = (1.22 - 1)/(1.4 - 1)$ based on the rule-transformation technique and the referential points in Table II. Other data can be handled in a similar way, and thus, their description is omitted. As a result, each input can be represented as a distribution on the referential values using a belief structure. If the assessment of x_{t-i} involves fuzziness, then $A_{i,j}$ can be defined as fuzzy sets and $\alpha_{n,i}(x_{t-i})$ can be calculated via membership functions [65].
2) *Quantitative attributes assessed using interval:* In this case, there are two ways to model input information in BRB framework. First, the interval can be seen as a special form of the fuzzy linguistic value; therefore, $\alpha_{n,i}(x_{t-i})$ can be determined in a way similar to the case 1). The second method is that belief structure can be extended to an interval version [53], and then, $\alpha_{n,i}(x_{t-i})$ can be generated through rule-based transformation technique, but $\alpha_{n,i}(x_{t-i})$ is also in the format of interval in such a case. For details, see [53].
3) *Qualitative or symbolic attributes assessed using subjective judgments.* In this case, the belief degree $\alpha_{n,i}(x_{t-i})$ is assigned directly by the decision maker using his subjective judgments for each referential value or each symbolic term. For example, if $\varepsilon_{n,i}(x_{t-i})$ is the belief degree assigned to the symbolic term or the referential term $A_{i,j}$ by the decision maker, then $\alpha_{n,i}(x_{t-i}) = \varepsilon_{n,i}(x_{t-i})$.

D. Calculating the Output of Belief-Rule-Base Prediction Model

When the antecedent attribute, such as the inputs of the BRB, is available, the ER approach serves as the inference engine of BRB prediction model, which mainly consists of following two steps.

1) Step 1: Calculation of the Activation Weight of Belief Rule: The activation weight of the kth rule ω_k at time t is calculated by

$$\omega_k(t) = \frac{\theta_k \prod_{i=1}^{M_k} (\alpha_{j,i}^k(t-i))^{\bar{\delta}_i}}{\sum_{l=1}^{L} \theta_l \prod_{i=1}^{M_k} (\alpha_{j,i}^l(t-i))^{\bar{\delta}_i}} \quad \text{and}$$
$$\bar{\delta}_i = \frac{\delta_i}{\max_{i=1,\ldots,M_k}\{\delta_i\}} \tag{4}$$

where $\delta_{i,k} (\in \mathrm{R}^+, i = 1, \ldots, M_k)$ is the relative weight of the ith antecedent attribute that is used in the kth rule. Because ω_k will be eventually normalized so that $\omega_k \in [0, 1]$ using (4), θ_k and $\delta_{i,k}$ can be assigned to any value in R^+. Without loss of generality, however, we assume that $\theta_k \in [0, 1]$(for $k = 1, \ldots, L$) and $\delta_{i,k} \in [0, 1]$(for $i = 1, \ldots, M_k$). $\alpha_{i,j}^k$(for $i = 1, \ldots, M_k$), which is called the individual matching degree, is the degree of belief to its jth referential value $A_{i,j}^k$ in the kth rule. $\alpha_k = \prod_{i=1}^{M_k} (\alpha_{i,j}^k)^{\bar{\delta}_i}$ is called the normalized combined matching degree.

2) Step 2: Rule Inference Using the Evidential-Reasoning Approach: Using the analytical ER algorithms [51], the final conclusion $O(\hat{y}(t))$ that is generated by aggregating all rules that are activated by the actual input vector $\mathbf{X}(t) = (x_{t-1}, x_{t-2}, \ldots, x_{t-p})$ at time instant t can be represented as follows:

$$O(\hat{y}(t)) = F(\mathbf{X}(t)) = \{(D_j, \hat{\beta}_j(t)), j = 1, \ldots, N\} \tag{5}$$

where $O(\hat{y}(t))$ denotes the predicted output of prediction model in the format of belief structure, and $\hat{\beta}_j(t)$ denotes the predicted belief degree in D_j at time instant t. The aggregated result $O(\hat{y}(t))$ represents the overall assessment of system's behavior and provides a complete picture about the system state at time t, from which one can tell which assessment grades the system's behavior is assessed to, and what belief degrees are assigned to the defined assessment grades $D_n, n = 1, \ldots, N$. In (5), $\hat{\beta}_j(t)$ can be formulated as follows [51]:

$$\hat{\beta}_j(t) = \frac{\mu(t) \times \prod_{k=1}^{L} \left(\omega_k(t)\beta_{j,k}(t) + 1 - \omega_k(t)\sum_{i=1}^{N} \beta_{i,k}(t)\right)}{1 - \mu(t) \times \left[\prod_{k=1}^{L} (1 - \omega_k(t))\right]}$$
$$- \frac{\mu(t) \times \prod_{k=1}^{L} \left(1 - \omega_k(t)\sum_{i=1}^{N} \beta_{i,k}(t)\right)}{1 - \mu(t) \times \left[\prod_{k=1}^{L} (1 - \omega_k(t))\right]} \tag{6}$$

$$\mu(t) = \left[\sum_{j=1}^{N}\prod_{k=1}^{L}\left(\omega_k(t)\beta_{j,k}(t) + 1 - \omega_k(t)\sum_{i=1}^{N}\beta_{i,k}(t)\right) - (N-1)\prod_{k=1}^{L}\left(1 - \omega_k(t)\sum_{i=1}^{N}\beta_{i,k}(t)\right)\right]^{-1} \quad (7)$$

where $\omega_k(t)$ is calculated by (4). Note that $\hat{\beta}_j(t)$ is the function of $\beta_{i,k}$, θ_k, δ_i, and $\mathbf{X}(t)$.

As shown in (5), we can see that the default output format of BRB prediction model is belief distribution. However, in some engineering applications, a numerical output value $\hat{y}(t)$ is required and preferred, such as pipeline-leak detection in [58]. To achieve this aim, the concept of utility in decision theory and utility-based transformation technique [62] can be used to convert the distributed assessment $O(\hat{y}(t))$ to a numerical output $\hat{y}(t)$. It is demonstrated that the two are equivalent in the sense that they both represent the same states of the system [62]. For detailed implementation, see [58] and [62].

From (6) to (7), it can be seen that belief degrees $\beta_{i,k}$, attribute weights δ_m, and rule weights θ_k play a significant role in the final conclusion $O(\hat{y}(t))$. The degree to which the final output can be affected is determined by the magnitude of the belief degrees, attribute weights, and rule weights. On the other hand, if the parameters of the BRB prediction model, such as θ_k, δ_m, and $\beta_{j,k}$, are not given *a priori* or only known partially or imprecisely, they could be trained using observed input and output information. Therefore, the inference performance of BRB prediction systems and the proposed model can be improved if the foregoing unknown parameters with some constraints are adjusted by optimization algorithm. The construction of the constraints of the unknown parameters in BRB prediction model is given as follows [66].

1) A belief degree must not be less than 0 or more than 1, i.e.,

$$0 \le \beta_{j,k} \le 1, \quad j = 1, \ldots, N, \quad k = 1, \ldots, L. \quad (8)$$

2) If the kth belief rule is complete, its total belief degree in the consequent will be equal to 1, i.e.,

$$\sum\nolimits_{j=1}^{N} \beta_{j,k} = 1. \quad (9)$$

Otherwise, the total belief degree is less than 1.

3) A rule weight is normalized, so that it is between 0 and 1, i.e.,

$$0 \le \theta_k \le 1, \quad k = 1, \ldots, L. \quad (10)$$

4) A attribute weight is normalized so that it is between 0 and 1, i.e.,

$$0 \le \delta_i \le 1, \quad i = 1, \ldots, p \quad (11)$$

As such, there is a need to develop a method that can learn parameters of BRB prediction model using observed input and output information. This is exactly discussed in the following.

III. Parameters Optimization for Belief-Rule-Base Prediction Model

As discussed in [64] and [66], belief rule in a BRB may initially be provided by human experts based on individual experience and personal judgments and then optimally trained if the observed input–output data are available. In addition, a change in unknown parameters θ_k, δ_m, and $\beta_{j,k}$ may lead to changes in the performance of the BRB prediction model. Therefore, it is preferred to construct optimization model for training prediction model. Inherited from the training process for general BRB system developed in [66], one optimization model is developed to train the proposed prediction model.

First, we assume that a set of observation pairs $(\mathbf{X}(t_m), y(t_m)), m = 1, \ldots, M$ is available, where $\mathbf{X}(t_m)$is an input vector at time t_m, and $y(t_m)$ is the observed output accordingly. Depending on the types of input and output, the optimal learning models can be constructed in different ways [19], [28], [66]. In this paper, we only consider the case that the output is in the format of belief structure. Then, in the following, $y(t_m)$ can be written as

$$y(t_m) = \{(D_n, \beta_n(t_m)), n = 1, \ldots, N\} \quad (12)$$

where D_n is a system-state evaluation grade of system used in the BRB prediction model, and $\beta_n(t_m)$ is the degree of belief to which D_n is assessed for the mth pair of observed data at time t_m.

Using the same system-state evaluation grades as for the observed output $y(t_m)$, a belief distribution conclusion that is generated by aggregating all the activated rules in the BRB prediction system can also be represented as follows:

$$\hat{y}(t_m) = \{(D_n, \hat{\beta}_n(t_m)), n = 1, \ldots, N\} \quad (13)$$

where $\hat{\beta}_n(t_m)$ is generated by (6) and (7) for a given input vector $\mathbf{X}(t_m)$. It is desirable that, for a given input $\mathbf{X}(t_m)$, the BRB prediction model can generate an output $\hat{y}(t_m)$, as represented in (13), which can be as close to $y(t_m)$ as possible. In other words, for the mth pair of the observed data $(\mathbf{X}(t_m), y(t_m))$, it is required that the BRB prediction model is trained to minimize the MSE between the observed belief $\beta_n(t_m)$ and the belief $\hat{\beta}_n(t_m)$ that is generated by the BRB prediction model for each referential term. Consequently, the training problem under this case is a multiobjective optimization problem since the foregoing requirement is true for all evaluation grades. The *minimax* method can be used to solve such multiobjective optimization problem [28], [30], [31], where the objective function can be written as

$$\begin{aligned} &\min_{\mathbf{Q}} \quad \max_{j=1,\ldots,N} \{\phi_j \varphi_j(\mathbf{Q})\} \\ &\text{s.t.} \qquad (8)\text{–}(11) \end{aligned} \quad (14)$$

with

$$\xi_j(Q) = \frac{1}{M}\sum_{m=1}^{M}(\beta_j(t_m) - \hat{\beta}_j(t_m))^2, \quad j = 1, \ldots, N \quad (15)$$

$$\varphi_j(\mathbf{Q}) = \frac{\xi_j(\mathbf{Q}) - \xi_j^*}{\xi_j^+ - \xi_j^*} \quad (16)$$

$$\xi_j^+ - \xi_j^* > 0, \quad j = 1, \ldots, N \tag{17}$$

where $\mathbf{Q} = \{\beta_{j,k}, \theta_k, \delta_i\}$ is the training parameter vector with $j = 1, \ldots, N, k = 1, \ldots, L, i = 1, \ldots, p$, and $\xi_j(Q)$ is the MSE between the observed belief $\beta_j(t_m)$ and the predicted belief $\hat{\beta}_j(t_m)$ by the BRB prediction model for the jth referential term. In this case, the optimization problem involves N objective functions, $L + L \times N + p$ training parameters given as $\mathbf{Q}$ and $L \times N + 2L + p$ constraints.

In (14), ϕ_j is the weighting parameter representing the relative importance of the jth objective function constrained with $0 \le \phi_j \le 1$ and $\sum_{j=1}^{N} \phi_j = 1$. In this paper, for simplicity, ϕ_j is set to be equivalent for all objective functions, i.e., $\phi_j = (1/N), j = 1, \ldots, N$. ξ_j^+ and ξ_j^* are the feasible maximum and minimum values for the jth objective function, as defined in (15). It can be seen from (14) to (17) that the optimized objective is a function of parameter vector $\mathbf{Q}$, while the elements of $\mathbf{Q}$ are constrained in the continuous region, as defined in (8)–(11). In this sense, we can see that $\beta_{j,k}, \theta_k$, and δ_i are continuous variables, and there is no discrete variable optimized in (14). As shown in (6) and (7), $\hat{\beta}_j(t)$ is a function of $\beta_{j,k}, \theta_k$, and δ_i, and thus, the optimized objective function can be considered to be continuous. Therefore, such a *minimax* optimization model can be solved using MATLAB optimization toolbox with FMINIMAX function. In addition, we note that this function FMINIMAX has been used for a similar purpose in [28] and [66]. Once the input–output data become available, the related parameters to the BRB prediction model can be obtained rationally. Then, the system's behavior state can be predicted when the new information becomes available.

IV. Demonstration of Belief-Rule-Base Prediction Model: A Practical Case Study

In this section, a practical case study is examined to validate the BRB prediction model under the distributed outputs and to show the application potential of BRB prediction model in engineering practice.

As a key device of INSs in weapon systems and space equipments, an inertial navigation platform plays an irreplaceable important role and its operating state has a direct influence on navigation precision. In our tested inertial navigation platform, the sensors are fixed in it, which mainly include three DTGs and three accelerometers, measuring angular velocity and linear acceleration, respectively. Statistical analysis shows that almost 80% failures of inertial platforms result from gyroscopic drift. In our case, the gyro fixed on an inertial platform is a mechanical structure having 2 degrees of freedom from the driver and sense axis. For a general description of an inertial navigation platform and gyros, see [55]. When the inertial platform is operating, the rotating wheel of DTGs with very high speed can lead to rotation axis wear, and finally, result in gyro's drift. With accumulation of wear, the drift degrades, and finally, results in the failure of DTGs. Especially, several fails in launching, which are resulted by the abnormality of the DTGs, are strongly driving forces for building reliable and cost-effective prediction model that can provide the complete picture of the DTG by analyzing the changing trend of the DTG drift. As such, the drift of DTGs is usually used as a performance indicator to evaluate the health condition of an inertial platform. Since such DTG is widely used in aeronautic control system with safety-critical requirement, it is desired that the adopted model is not only comprehensible by human expert, but also transparent with inference process, and can provide some explanations of the predicted results. As analyzed above, the BRB prediction model can satisfy these desires in the sense that it can provide a complete picture on the state of DTG, such as the belief degree on each assessment grade. In our study, we only take the drift-degradation measurement along the sense axis for an illustration purpose since this variable plays a dominant role in the assessment of gyros degradation. The demonstration of the proposed model is conducted in the following and some comparative studies are provided as well.

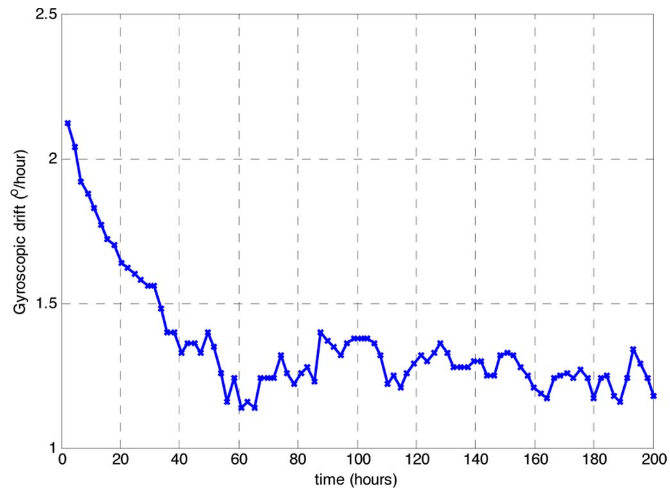


Fig. 1. Entire gyroscopic drift data collected in the trial.

A. Problem Description

In this study, the system's drift data are collected in a DTG performance reliability trial from leaving factory. For the DTG drift trial, some technical parameters include the sampling interval T, the acceleration of gravity g, and the geographic latitude R. In this experiment, $T = 2.2$ h, $g = 979.4121$ cm/s^2, and $R = 34.6006°$, where a PC-based data-acquisition system is used to acquire and store the drift data. After the experiment, we can collect all the drift data. The datasets include the time-to-drift data for 90 suits of gyroscope. The experiment results are illustrated in Fig. 1.

As shown in Fig. 1, it indicates that the gyroscopic drift is a time series. Therefore, it is reasonable to assume that the current gyroscopic drift value y_t is related to the most recent y_{t-1}, or even extended into the past values $y_{t-2}, \ldots, y_{t-p}$. This is because the next gyroscopic drift value is dependent on the current level of gyroscope state to a certain extent. As such, y_{t-i} can be treated as the antecedent attributes of the rule base, where $i = 1, \ldots, p$. In next section, the BRB prediction model will be established to predict the behavior of DTG at time t based on $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$.

B. Prediction Model Based on the Belief-Rule-Base Approach

The BRB system is used to capture the relationship among $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$ through system state after p steps, which captures the dynamics of DTG. This kind of BRB system can be considered as the forecasting model and used to predict the future behavior of the system. Thus, the prediction BRB system can be constructed as follows:

R_k: IF y_{t-1} is $A_1^k \wedge y_{t-2}$ is $A_2^k \cdots \wedge y_{t-p}$ is A_p^k, THEN system state at next step is $\{(D_1, \beta_{1,k}), \ldots, (D_N, \beta_{N,k})\}$ with a rule weight θ_k and attribute weight $\delta_1, \delta_2, \ldots, \delta_p$.

For simplicity, this study experimented with a relatively smaller number p, and then, two antecedent attributes are selected as an example. Consequently, we can transform 90 observed values to 88 sets of input–output patterns according Table I. Then, the belief rule can be represented as follows:

R_k: IF y_{t-1} is $A_1^k \wedge y_{t-2}$ is A_2^k, THEN system state at next step is $\{(D_1, \beta_{1,k}), \ldots, (D_N, \beta_{N,k})\}$ with a rule weight θ_k and attribute weight $\delta_1, \delta_2, \ldots, \delta_p$

where $R_k (k = 1, \ldots, L)$ is the belief rules of the BRB. In the BRB, A_1^k and A_2^k are the referential points of y_{t-1} and y_{t-2}, respectively. $D_l (l = 1, \ldots, N)$ are the assessment grades of system's behavior.

C. Referential Points of the Antecedents and Consequence

From Fig. 1, we can see that the original input data were all provided as numerical numbers; therefore, there is a need to equivalently transform the numerical value into the belief structures. Note that the referential values of an attribute and the types of input information are problem specific [26], [59]. In this case study, the technical requirement of DTG is that the drift value is not larger than 2.4° h^{-1}; therefore, the linguistic labels for y_{t-1}, y_{t-2} are defined as "*Small*" (S), "*Medium*" (M), and "*Large*" (L). i.e., $A_i^k \in \{S, M, L\}$ for $i = 1$ and 2. For the system's behavior state, three assessment grades of system's behavior for outputs of BRB prediction system are used, and they are "*Good*" (G), "*Average*" (A), and "*Poor*" (P); this is to say that $\mathbf{D} = (D_1, D_2, D_3) = (G, A, P)$. As such, in rule R_k $(k = 1, \ldots, L)$ in above section, $L = 9$ and $N = 3$.

The referential points of inputs defined above are in linguistic terms and need to be quantified. In [62], a scheme to convert other inputs to belief structure has been developed. In the proposed scheme, there is an important technique, i.e., rule-based information-transformation technique, which is used to deal with the input/output information involving quantitative data. After data transformation, quantitative data can be transformed to belief structures, and the two are equivalent in the sense that they both represent the same states of the system. For more details, see [62]. The quantified results in this case study are listed in Table II.

D. Simulation Results Under Belief Distribution Output

To validate the BRB prediction model, the available data are partitioned into a training dataset and a testing dataset. The training dataset is used to train the BRB-prediction-model parameters. In this case, the output is transformed into the following distributed output format:

$$S(\text{output}) = \{(D_1, \beta_1), (D_2, \beta_2), (D_3, \beta_3)\}$$

where $\beta_i, i = 1, 2, \text{ and } 3$ can be obtained by the rule-based-information transformation. Since the gyroscopic drift is a time series and the input–output patterns are constructed by Table I, we use the same referential points as given in Table II for the assessment grades of system's behavior $D_i, i = 1, 2,$ and 3, respectively. As such, according to Table II, one of transformation rules for system-assessment grade D_1 could be read as follows: If gyroscopic drift is 1.0° h^{-1}, then system's behavior state is ranked to be "*Good*" with the matched degree 1.0, i.e., $D_1 = 1.0$. The other two transformation rules can be interpreted in a similar way. For instance, using the defined referential points, the 35th data with output value 1.28° h^{-1} in our dataset can be transformed into S(output) = {(D_1,0.7),(D_2,0.3),(D_3,0)} based on rule-based transformation technique. Similar to the example in Section II-C, we have $\beta_1 = (1.4 - 1.28)/(1.4 - 1) = 0.3$, and $\beta_2 = (1.28 - 1)/(1.4 - 1) = 0.7$. In fact, this process is similar to fuzzification step used in fuzzy set theory [2], [10]. However, it is worth noting that the referential values used in the above rules are problem-dependent. In our case study, it is usually required that the drift of the DTG is not larger than 2.4° h^{-1} since the DTG is used in a safety-critical system.

In this case study, three BRB prediction systems are constructed for this validation analysis. The first BRB is directly constructed from expert knowledge about the relationship between system's behavior and drift, which simulated the uncertain information in system's behavior prediction. The second BRB is given by an engineering practitioner. Then, it is trained using the proposed optimization method and the data are generated from the first BRB, thus leading to the third optimally trained BRB. Finally, the trained BRB prediction model is used to predict outputs for the testing input data. In order to make our following listed steps clear, we select the 35th data used in our dataset to illustrate and track the changes in each step.

1) Step 1: Directly Construct a Benchmark Belief Rule Base According to the Prior Knowledge: In engineering practice, the DTG drift values change over time, as shown in Fig. 1. The inertial system presents great requirement to the precision of DTG. In practice, DTG drift as an indication of performance of DTG is the smaller the better, i.e., if the drift value is large, then the behavior of DTG can be assessed to "*Poor*" with great possibility. Based on this principle, a benchmark BRB is constructed through expert intervention, as shown in Table III.

In the benchmark BRB, the values of θ_k and δ_i are all set to 1, where $k = 1, \ldots, 9$, and $i = 1, \ldots, 2$, except $\theta_3, \theta_5,$ and θ_7. For example, the value of θ_3 is 0.9, which is less than 1.0, which represents that the rule 3 has a low credibility. To some extent, this reflects the knowledge that we have obtained is not complete and some uncertainties exist in assessment results. This BRB is used as a benchmark to check how closely a BRB, which is initialized either using historical or using expert knowledge, can be trained using the proposed training algorithm to simulate

TABLE III
BENCHMARK BRB PREDICTION SYSTEM

Rule number	Rule weight	y_{t-1} AND y_{t-2}	Output distribution{D_1,D_2,D_3}
1	1.0	S and S	{ (D_1, 0.8), (D_2, 0.1), (D_3,0.1) }
2	1.0	S and M	{ (D_1, 0.7), (D_2, 0.2), (D_3,0.1) }
3	0.9	S and L	{ (D_1, 0.64), (D_2, 0.12), (D_3,0.24) }
4	1.0	M and S	{ (D_1, 0.6), (D_2, 0.2096), (D_3,0.1904) }
5	0.95	M and M	{ (D_1, 0.55), (D_2, 0.25), (D_3,0.2) }
6	1.0	M and L	{ (D_1, 0.45), (D_2, 0.3), (D_3,0.25) }
7	0.85	L and S	{ (D_1, 0.27), (D_2, 0.18), (D_3,0.55) }
8	1.0	L and M	{ (D_1, 0.0), (D_2, 0.3), (D_3,0.7) }
9	1.0	L and L	{(D_1, 0.0087), (D_2, 0.1629), (D_3,0.8284) }

the true relationship. Therefore, the BRB shown in Table III is referred to as benchmark BRB for short.

To apply the benchmark BRB to construct the benchmark distributed outputs, the input values of y_{t-1} and y_{t-2} also need to be transformed and represented in terms of the referential values in the form of $A_i^k \in \{S, M, L\}$. Then, (6) and (7) are used to generate the distributed outputs of the benchmark BRB, as defined in Table III. The generated distributed outputs are then used to be the true behavior state for the benchmark distributed outputs. Since the inputs data of the 35th data used in our dataset has been converted into the required belief structure, the output of the benchmark BRB prediction system for this data point can be computed by (6) and (7) as $\{(D_1, 0.749), (D_2, 0.1425), (D_3, 0.1085)\}$. Other data can be obtained in a similar way. The distributed outputs of the benchmark BRB are shown graphically in Fig. 2(a)–(c).

2) Step 2: Set the Parameters of the Initial Belief Rule Base: The belief degrees in the initial BRB are given and listed in Table IV. The initial values of θ_k and δ_i are all set to 1, where $k = 1, \ldots, 9$, and $i = 1$ and 2. The initial belief degrees in Table IV are determined by the expert knowledge according to the running patterns and historical data of the DTG and the change of the drift values over time. For example, according to the historical information, if y_{t-1} is "*Small*" and y_{t-2} is "*Small*," the expert judges that the possibility of system in the "*Good*" state is larger than the other states. Therefore, the expert may assess that the belief degree to "*Good*" is 0.95 and the belief degree to "*Average*" is 0.05 and to "*Poor*" is 0, respectively. Thus, the initial belief rule can be obtained as the second row of Table IV. The other rules can be initialized in a similar way.

Once the initial BRB prediction system is constructed, it can be used to predict the system's behavior. For illustration purposes, all sets of sampling data are used for testing the performance of the initial BRB. As shown in Fig. 2(a)–(c), we can observe that the belief degrees to the consequents calculated by (6) and (7) using the initial BRB do not match well the distributed outputs generated by the benchmark BRB, as defined in Table III. Similar to step 1, the output of the initial BRB prediction system for the 35th data can be computed by (6) and (7) as $\{(D_1, 0.8579), (D_2, 0.1106), (D_3, 0.0314)\}$. Obviously, there are significant difference between the output of the benchmark BRB and the initial BRB. This means that the initial rule base

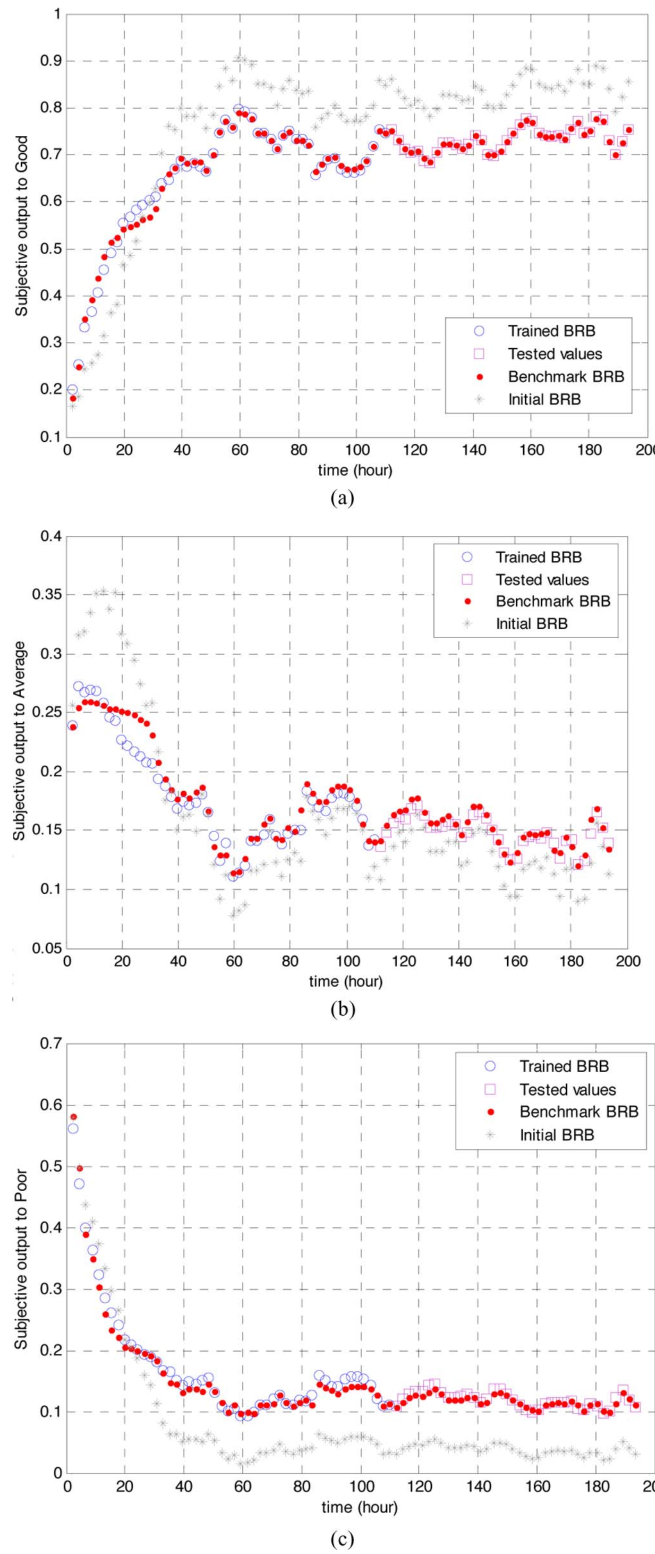


Fig. 2. Predicted results of BRB prediction model being (a) "good," (b) "average," and (c) "poor."

provided by the expert is not good enough. Therefore, it is necessary to use the available information to train the rule base by the optimization approach.

TABLE IV
INITIAL BRB PREDICTION SYSTEM

Rule number	Rule weight	y_{t-1} AND y_{t-2}	Output distribution{D_1, D_2,D_3}
1	1.0	S and S	{ (D_1, 0.95), (D_2, 0.05), (D_3,0) }
2	1.0	S and M	{ (D_1, 0.8), (D_2, 0.15), (D_3,0.05) }
3	1.0	S and L	{ (D_1, 0.75), (D_2, 0.2), (D_3,0.05) }
4	1.0	M and S	{ (D_1, 0.6), (D_2, 0.3), (D_3,0.1) }
5	1.0	M and M	{ (D_1, 0.5), (D_2, 0.3), (D_3,0.2) }
6	1.0	M and L	{ (D_1, 0.45), (D_2, 0.25), (D_3,0.3) }
7	1.0	L and S	{ (D_1, 0.35), (D_2, 0.4), (D_3,0.25) }
8	1.0	L and M	{ (D_1, 0.1), (D_2, 0.5), (D_3,0.4) }
9	1.0	L and L	{ (D_1, 0.05), (D_2, 0.05), (D_3,0.9) }

TABLE V
TRAINED BRB PREDICTION SYSTEM

Rule number	Rule weight	y_{t-1} AND y_{t-2}	Output distribution{D_1, D_2,D_3}
1	0.9755	S and S	{ (D_1, 0.8242), (D_2, 0.0833), (D_3,0.0925) }
2	0.9475	S and M	{ (D_1, 0.7321), (D_2, 0.1489), (D_3,0.1190) }
3	1.0	S and L	{ (D_1, 0.75), (D_2, 0.2), (D_3,0.05) }
4	0.9998	M and S	{ (D_1, 0.6560), (D_2, 0.2110), (D_3,0.1329) }
5	1.0	M and M	{ (D_1, 0.6477), (D_2, 0.1880), (D_3,0.1644) }
6	0.9867	M and L	{ (D_1, 0.4926), (D_2, 0.2419), (D_3,0.2655) }
7	1.0	L and S	{ (D_1, 0.35), (D_2, 0.40), (D_3,0.25) }
8	1.0	L and M	{ (D_1, 0.1194), (D_2, 0.5061), (D_3,0.3749) }
9	1.0	L and L	{ (D_1, 0.1019), (D_2, 0.1453), (D_3,0.7528) }

3) Step 3: Train the Belief Rule Base Constructed Initially in Step 2: After the input values y_{t-1} and y_{t-2} are transformed and represented in terms of referential values, based on the data generated for construction of the benchmark BRB, as shown in step 1, and the initial BRB, as given by the expert in step 2, the optimization models (14)–(17) are used to train the initial BRB. For illustration purposes, the first 50 sets of data are used as the training data for parameter estimation. In this simulation, the error tolerance is set to 0.05 and the maximum iteration is set to 100 to avoid dead loop in the optimal learning process. The trained results are listed in Table V.

After training, the remaining 38 sets of data are used for testing the trained BRB prediction model. The test results are illustrated in Fig. 2(a)–(c), where the comparisons are shown between the actual output and the predicted output that is generated using the trained BRB prediction model and the initial one. It is clear that the distributed outputs generated by the trained BRB system can match the benchmark BRB more closely than the initial BRB. Similar to the previous two steps, the output of the trained BRB prediction system for the 35th data can be obtained as $\{(D_1, 0.7748), (D_2, 0.1407), (D_3, 0.1145)\}$. This further demonstrates that the trained BRB can match the benchmark BRB accurately and the established optimization method can work well.

As discussed in Section I, we can see that probabilistic interpretation is possible in our case here since the belief degree is only assigned to the single grade of the assessment framework in the BRB model. This result arises from rule-based transformation technique and ER algorithm. Therefore, we can consider belief degree as a generalized probability from either Dempster's lower and upper probability [7]–[9] or Smets' pignistic probability [41] viewpoints. In the following, we can give a clear interpretation of the predicted outputs and assessed results. Taking the predicted curve being grade "*Good*" [see Fig. 2(a)], for example, we can read from probabilistic viewpoint that the probability of DTG system's behavior assigned to grade "*Good*" increases from 0.1998 to 0.724 monotonously and then experiences a fluctuating process around 0.72. The reason that the belief degree assigned to grade "*Good*" is below 0.5 in the first seven sampling period lies in the fact that this kind of DTG is not stable at the beginning of its operating time. As shown in Fig. 2(a), the belief degree assigned to grade "*Good*" is around 0.72 with the steady operation of DTG system, i.e., the function of DTG is good almost with the probability value 0.72. At the same time, the probability of system's behavior assigned to grade "*Poor*" is small. The other cases can be analyzed in a similar way. From the above analysis, we can conclude that the function and behavior of DTG is normal in this case and that the proposed method can give complete picture about the operating state of DTG. These results can provide valuable information for conducting cost-effective maintenance and avoiding a large calamity.

4) Step 4: Performance Evaluation: In this section, the results of predicted results using BRB prediction model will be discussed. The histograms of the absolute difference between the actual and the predicted by the initial BRB and the trained BRB are shown in Fig. 3.

It can be seen that the absolute errors of most of samples using trained BRB are less than 0.01 in terms of assessment "*Good*," as shown in Fig. 3(b), when compared with the range of the predicted results in Fig. 2(a). For the predicted values in terms of assessment "*Average*" and "*Poor*," the histograms of the absolute errors are shown in Fig. 3(b). All predicted results of trained BRB model are acceptable. In contrast, the absolute errors of most of samples using the initial BRB model are close to 0.1 in terms of assessment "*Good*," as shown in Fig. 3(a).

In order to further demonstrate the proposed method, five measures of forecasting performance are introduced [34], [45], [46], including mean absolute percentage error (MAPE), root MSE (RMSE), Teil's inequality coefficient (TIC), Teil's U-statistics (TUS), and modeling efficiency (ME). These indicators are all based on forecast residuals and are widely employed in the realm of forecasting practice. The detailed mathematical expressions are summarized in Table VI. The measures MAPE and RMSE have been widely used in prediction field to test the accuracy of the prediction model [6], [22], [24], [54], [56], where RMSE can, in part, measure the variance character in predicted results. Here, we report on additional measures TIC, TUS, and ME since they provide some simple performances on relative scale that can avoid the scaling problem of both MAPE and RMSE. If TIC $= 0$ or TUS $= 0$, then the model produces

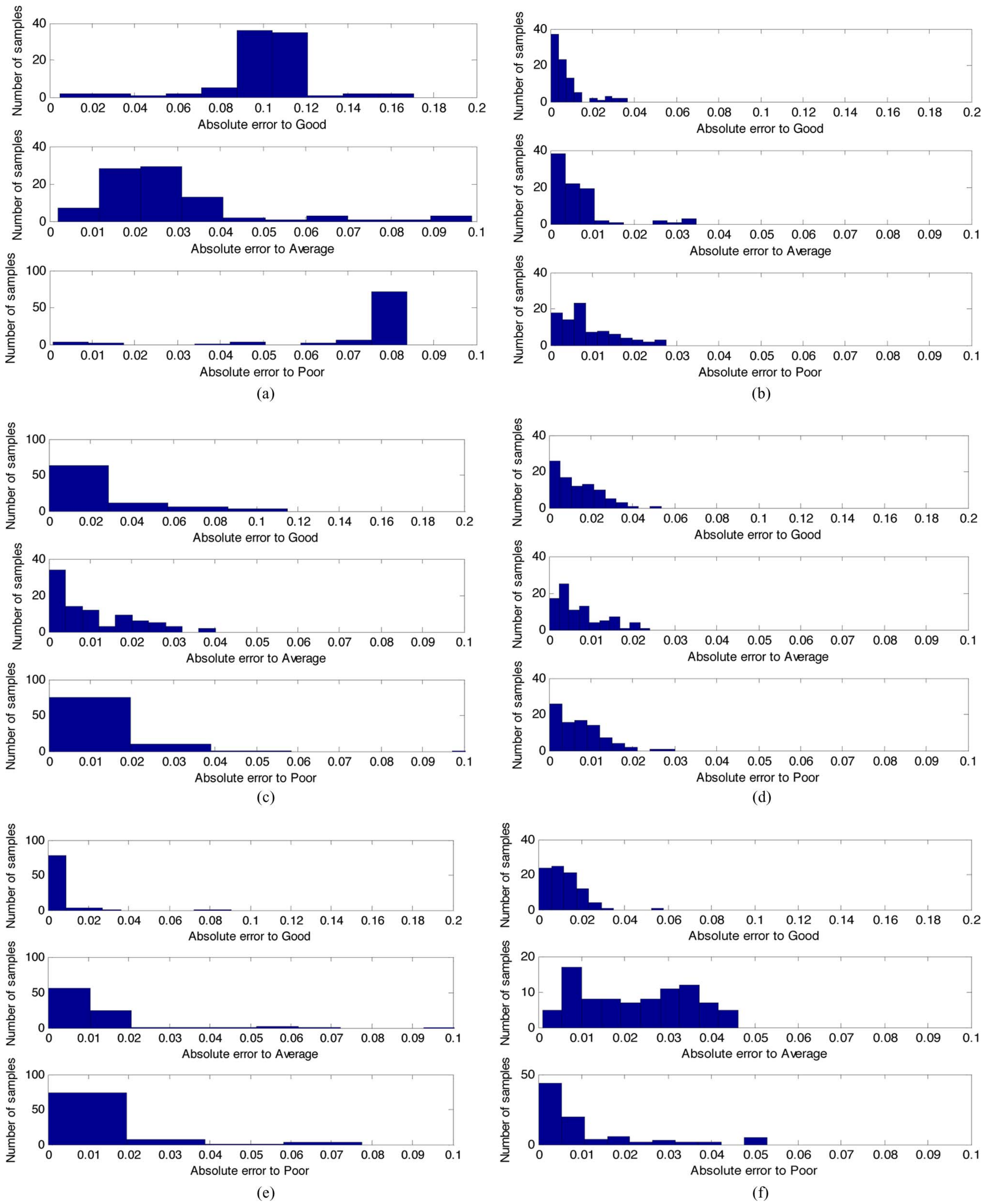


Fig. 3 Histogram of absolute difference by (a) initial BRB, (b) trained BRB, (c) BRF neural networks, (d) ARMA/GARCH, (e) FCM-TS model, and (f) LSSVM.

perfect predictions. In addition, ME $= 1$ indicates a perfect fit, ME $= 0$ reveals that the model is no better than a simple average, and ME < 0 indicates a poor forecasting performance.

In this case study, the above measures are used to evaluate the prediction performance. For example, the MAPE and ME between the benchmark BRB and the initial BRB in terms belief degrees to the linguistic term "*Good*" is 0.1923 and 0.0475, respectively. However, after training prediction model, the MAPE and ME in terms of the belief degrees to "*Good*" are 0.003 and 0.9935, respectively. This provides the evidence for the validity

TABLE VI
STATISTICAL MEASURES OF PREDICTION ACCURACY

Measures	Mathematic expression
MAPE	$\frac{1}{n}\sum_{i=1}^{n}\left\lvert\left(\text{Observed}_i - \text{Predicted}_i\right)/\text{Observed}_i\right\rvert$
RMSE	$\sqrt{\sum_{i=1}^{n}\left(\text{Observed}_i - \text{Predicted}_i\right)^2/n}$
TIC	$\sqrt{\sum_{i=1}^{n}\left(\text{Observed}_i - \text{Predicted}_i\right)^2}/\left(\sqrt{\sum_{i=1}^{n}\text{Observed}_i^2} + \sqrt{\sum_{i=1}^{n}\text{Predicted}_i^2}\right)$
TUS	$\sqrt{\sum_{i=1}^{n}\left(\text{Observed}_i - \text{Predicted}_i\right)^2/\sum_{i=1}^{n}\text{Observed}_i^2}$
ME	$1-\left(\sum_{i=1}^{n}\left(\text{Observed}_i - \text{Predicted}_i\right)^2/\sum_{i=1}^{n}\left(\text{Observed}_i - \overline{\text{Predicted}}\right)^2\right)$

TABLE VII
PERFORMANCE OF DIFFERENT MODELS USING THE SAME DTG-TESTING DATASET

	Initial BRB			Trained BRB			RBFNN [68]		
	Good	*Average*	*Poor*	*Good*	*Average*	*Poor*	*Good*	*Average*	*Poor*
MAPE	0.1923	0.2273	0.9238	0.003	0.0377	0.0457	0.0416	0.0849	0.0753
RMSE	0.1211	0.0298	0.0922	0.0022	0.0058	0.0105	0.0334	0.0137	0.0110
TIC	0.0662	0.0918	0.5072	0.0013	0.0165	0.0209	0.0199	0.0377	0.0394
TUS	0.1418	0.1692	0.6748	0.0026	0.0327	0.0423	0.0392	0.0774	0.0806
ME	0.0475	0.2079	0.0135	0.9935	0.8934	0.7369	0.2486	0.5324	0.2592
	ARMA/GARCH [36]			FCM-TS [23]			LSSVM [44]		
	Good	*Average*	*Poor*	*Good*	*Average*	*Poor*	*Good*	*Average*	*Poor*
MAPE	0.0240	0.0692	0.0758	0.004	0.0894	0.1404	0.0156	0.2232	0.0646
RMSE	0.0183	0.0109	0.0088	0.0044	0.0153	0.0196	0.0112	0.0322	0.0083
TIC	0.0107	0.0309	0.0319	0.0026	0.0427	0.0725	0.0066	0.0844	0.0301
TUS	0.0214	0.0620	0.0645	0.0051	0.0866	0.1431	0.0131	0.1825	0.0605
ME	0.5703	0.6011	0.3926	0.9754	0.2616	-1.6907	0.8388	0.0425	0.4436

of the developed prediction model and optimization method since the MAPE value is smaller and the ME value is very close to 1. The detailed results are listed in Table VII.

Based on the above experiment results, it is clear that the initial belief rules for system's behavior prediction given by an expert are not accurate. When the running information of system becomes available, the proposed optimization model can train the BRB effectively.

E. Comparative Studies

With the same dataset, we obtain the predicted results by ARMA/GARCH [36], RBFNN [68], LSSVM [44], and FCM-TS models [23] as a comparison of the measure of the learning ability of the proposed model. The histograms of the absolute difference between the actual and the predicted using these methods are shown in Fig. 3(c)–(f), and Table VII lists the measures of forecasting performance. Even though the performance of ARMA/GARCH, RBFNN, LSSVM, and FCM-TS predictor could be improved by optimizing the related parameters, it is clear that the proposed BRB predictor performs the best. The predicted results of the DTG behavior on testing dataset by these methods are shown in Fig. 4(a)–(c). It is worth noting, from Fig. 4(a), that both the proposed method and FCM-TS predictor generate better prediction on grade "*Good*" than other methods to some extent. However, under the same condition, our method can perform better on grades "*Average*" and "*Poor*" than the FCM-TS predictor, as shown in Fig. 4(b) and (c). In addition, the measures presented in Table VII further support this point.

In experiment, it should be noted that even though the BRB prediction model takes a little longer time (i.e., 6.21 s) in training than other predictors, it is still an effective tool in DTG-behavior forecasting application. The reasons are twofold. First, the sampling interval is about 2.2 h in DTG-monitoring process. Accordingly, the average training speed of 6.21 s still means an effective processing speed in this case study. Moreover, in most real-world applications, the time interval between two samples is usually in terms of hour(s) or day(s); hence, the model developed in this paper can perform well in real-world forecasting practice. Second, in contrast with real-time requirement in DTG-behavior prediction, safety and interpretability are more desired due to its safety-critical requirement in application. To further reveal the improvement of the proposed model, we have a look at the aspect of interpretability of the used models above. Here, the ARMA/GARCH, RBFNN, LSSVM, and FCM-TS predictor are used to show the learning ability of the BRB model. Therefore, the same training and testing datasets are used for ARMA/GARCH, RBFNN, LSSVM, FCM-TS, and BRN predictors. However, if we do not use the rule-based transformation technique under ER and BRB framework, the dataset cannot be converted into the belief structures, and thus, ARMA/GARCH,

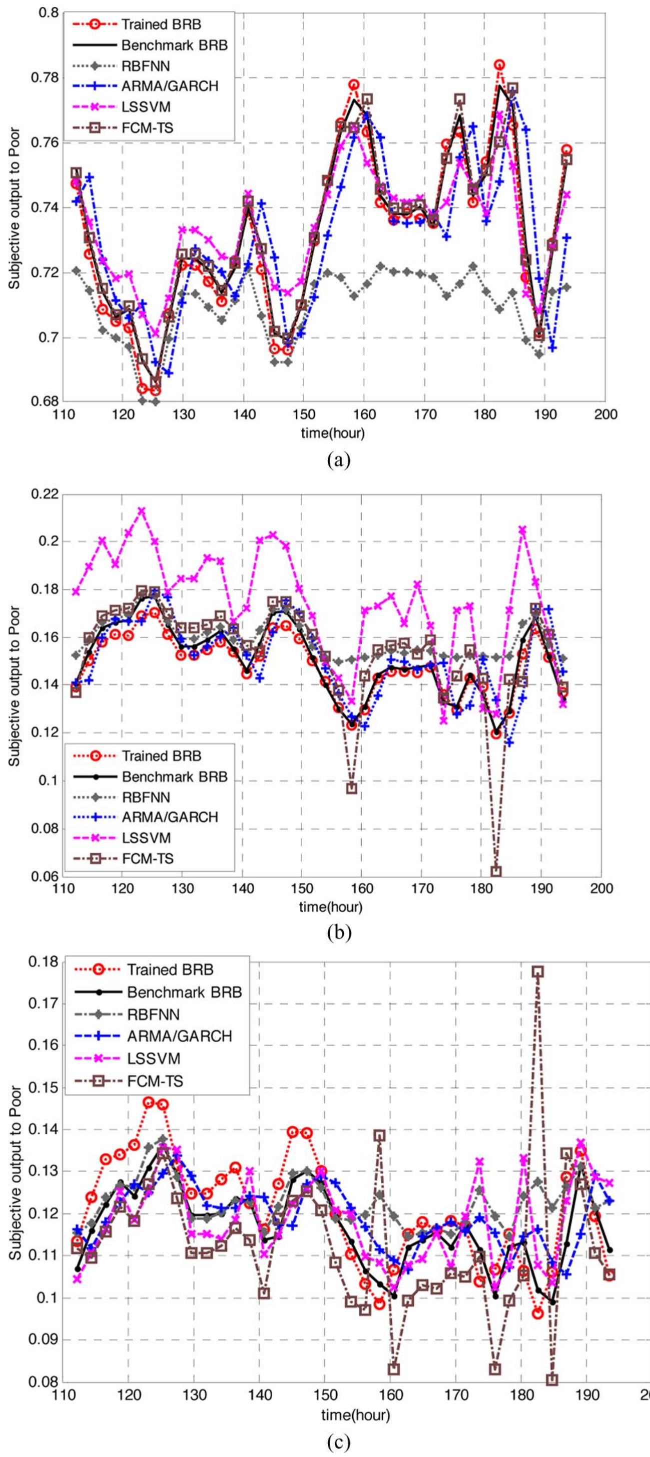


Fig. 4. Predicted results on testing dataset being (a) "*good*," (b) "*average*," and (c) "*poor*."

RBFNN, LSSVM, and FCM-TS predictors can neither directly deal with the inputs with belief structures nor generate the outputs with the belief structures as BRB model. However, the outputs with the belief structures are easy to interpret, as shown in Section IV-D. In addition, the parameters in ARMA/GARCH, RBFNN, LSSVM, and FCM-TS are difficult to interpret. For example, the connection weights of RBFNN are not easy to explain and the performance of LSSVM is heavily dependent on σ, ε, and C, but they are difficult to select [44]. Unlike these models, the parameters in BRB have clear meaning. For example, in the second rule of Table V, the trained rule weight is 0.9475, which reflects the credibility of this rule, and the estimated results in consequent part can also reflect a complete picture of the system's behavior. The interpretability of such structure has been illustrated in Sections I and IV-D. This contrasts sharply with ARMA/GARCH, RBFNN, LSSVM, and FCM-TS predictors if we do not model prediction problem under the ER and BRB framework.

F. Effect of the Number of Linguistic Labels in Antecedent on the Predicted Results

The number of linguistic labels determines granularity of the linguistic representation with respect to the numerical values. In foregoing simulation studies, a three-label set in antecedent and consequence is chosen as an example. Certainly, other choices can be used as well, but the final judgment will be based on the goodness-of-fit testing and the specific requirements in engineering applications. Therefore, the selection of the labels of the antecedent and consequence is problem-dependent. In the following, we give a simulation study to show the effect of the number of labels set in antecedent on the prediction results. The effect of the consequent partition can be analyzed in a similar way and, thus, is omitted due to limited space.

As a matter of fact, more label sets in prediction model means that there are more number of belief rules in BRB system since the relationship between the label set of antecedent and the number of belief rules can be written as $L = \prod_{i=1}^{p} J_i$. If all antecedent attributes share the same number of referential value J, as in this paper, then the numbers of belief rule can be written as $L = J^p$. Under such case, the processed problem can be represented more specifically. However, the increase of belief rule will cause that more unknown parameters need to be optimized since the number of unknown parameters is calculated by $L + L \times N + p$. In addition, the more labels we use, the lower the interpretability of the model. In order to show the effect of the number of label sets in antecedent on the predicted results, a simulation is given under different numbers of antecedents as in the following. Specifically, four-label, six-label, and ten-label sets are used to show the impact of the number of the labels on the predicted result. It corresponds to BRB prediction model having 16 rules, 36 rules, and 100 rules, respectively. In the simulation, the initial parameters of BRB are assigned randomly with constraints (8)–(11). In the obtained results, the MAPE fluctuates along with increasing number of labels—0.2% for four labels, 0.21% for six labels, and 0.18% for ten labels. However, the ME increases correspondingly—0.9955 for four labels, 0.9957 for six labels, and 0.9959 for ten labels. Table VIII shows the influence of the number of linguistic labels on the predicted results.

From Table VIII, we can see that the prediction accuracy can be improved for a greater label set in antecedent, but the improvement is a bit small in this simulation. A possible explanation for this is that when increasing the number of labels,

TABLE VIII
PERFORMANCE EVALUATION UNDER DIFFERENT NUMBER OF LABEL SET

	BRB(4)			BRB(6)			BRB(10)		
	Good	*Average*	*Poor*	*Good*	*Average*	*Poor*	*Good*	*Average*	*Poor*
MAPE	0.002	0.0344	0.0467	0.0019	0.0339	0.0462	0.0018	0.0320	0.0423
RMSE	0.0019	0.0052	0.0114	0.0018	0.0052	0.0110	0.0016	0.0048	0.0096
TIC	0.0011	0.0150	0.0213	0.0011	0.0147	0.0210	0.0010	0.0138	0.0192
TUS	0.0022	0.0297	0.0431	0.0022	0.0292	0.0425	0.0021	0.0247	0.0387
ME	0.9955	0.9102	0.7328	0.9957	0.9131	0.7381	0.9959	0.9230	0.7758
Training time (s)	9.064			17.327			39.128		

the number of belief rules increases correspondingly, e.g., 16, 36, and 100 for 4, 6, and 10 labels, respectively. However, the increased belief rules do not influence the predicted value as heavily, since many of belief rules cannot be activated by the input data. We also note that more time is needed to train the BRB prediction model for a greater label set and the average training time increases as well. The reason for such an unexpected case is that the number of belief rules increases as analyzed above. These results show that we cannot choose BRB arbitrarily since more time is needed to train the BRB prediction model for a greater label set and the interpretability of the model can be lowered for a greater label set. In this case study, a three-label set is a tradeoff between the prediction accuracy and training efficiency. The issues associated with how to select the number of the label set may be considered in further research.

G. Discussion

Based on the above experiments, it is clear that the initial belief rules for system's behavior prediction given by an expert are not accurate. With the accumulation of the new knowledge or input–output data, the proposed prediction model can be progressively trained to better simulate a real system. Once the BRB prediction is trained, its knowledge can be used to forecast future behavior. However, as known for rule-base system, there are some typical types of structural errors including conflict rules, missing rules, redundant rules, and circular depending rules [16], [33]. It is noted that the prediction model, in this paper, are initially built using expert knowledge or historical data, and then tuned by optimization. Hence, they are not yet equipped with the ability to automatically identify and rectify conflicting rules. If conflicting rules are identified by experts or certain relationships among rules are required, then additional constraints could be added to the optimization models to avoid conflicts and to meet the requirements. On the other side, for a complex real-world problem, prior knowledge may be limited, which may lead to the construction of an incomplete, or even inappropriate, initial BRB structure. If there are too many belief rules in an initial rule base, the training task may become too complicated to handle, or it is possible to result in overfitting. For example, in this paper, we consider all possible rules, but rules 3 and 7 in Table V were not trained at all as the belief degrees and rules weights of these rules remain unchanged after the training, as marked in gray in Section IV-D. However, in practice, when activated, such rules that are untouched during training may lead to an irrational conclusion if they were initially assigned randomly or without care. If there are too few rules in the initial rule base, this may lead to underfitting. To achieve an overall optimal BRB prediction model, however, the structure of a BRB system may need to be adjusted and optimized as well. This involves parameters identification and structure identification simultaneously. In fact, there are some good papers addressing optimal design of the rule-based system, such as rule-reduction or rule-simplification techniques in fuzzy-rule-base literature [13], [14], [39], as discussed previously. In this paper, we do not consider such issues; however, the above simulation results indeed suggest that considering the optimal design of the BRB system is necessary and valuable in future work. This can also improve the interpretability of the model since the unknown parameters may be reduced. One thing that is worth noting is that the measure *variance accounted for* (VAF) [12], [35] may be useful for quantifying the improvement of the optimized model in the case when the optimal design of the BRB is taken into account, since the VAF can measure the predictor variable importance. This is valuable for further research as well.

V. CONCLUSION

This paper addresses the forecasting problems with BRB to improve the interpretability of the predicted result. More precisely, it extends BRB to handle system's behavior prediction problem, and a new prediction model based on BRB is presented, which can model and analyze prediction problems using not only numerical data but human judgmental information as well. In order to build an effective BRB forecasting model, a multiple-objective optimization model is provided to locally train the BRB prediction model by minimizing the MSE criterion. Finally, a practical study is provided to illustrate the detailed implementation procedures of the proposed approach and examine the feasibility and validity in real-life behavior prediction of DTG used in the INS. Compared with ARMA/GARCH, RBFNN, LSSVM, and FCM-TS predictors, the proposed method is superior in terms of prediction accuracy and interpretability; hence, it can be regarded as an effective tool for prediction applications.

There are three features in the proposed model that are inherited from BRB system. First, the BRB prediction model provides a powerful simulator that can explicitly represent expert's domain-specific knowledge as well as common-sense judgments and, thus, can avoid generating obviously irrational conclusions. Second, the new model is capable of processing input and output information that can be numerical or distributed with an information-transformation technique, thereby providing a flexible way to represent and deal with hybrid information encountered in forecasting practice. Third, the BRB prediction model is designed to allow the direct intervention of experts in deciding the internal structure of a belief rule and is comprehensible to humans in aspects of model parameter as well as prediction results from a probabilistic point of view. In further research, we will consider the optimal design of the BRB model and the online parameter-estimation methods for the proposed model.

Acknowledgment

The authors would like to sincerely thank and acknowledge the tremendous work and the inputs performed by the Associate Editor and the three anonymous reviewers for their valuable suggestions and thorough review, which have greatly improved this paper. Particularly, they would like to thank one anonymous reviewer for his/her insightful comments, which have improved the representation significantly.

References

[1] C. Angeli and A. Chatzinikolaou, "Prediction and diagnosis of faults in hydraulic systems," *Proc. Inst. Mech. Eng.*, vol. 216, no. 2, pp. 293–297, 2002.

[2] R. Babuska, *Fuzzy Modeling for Control*. Boston, MA: Kluwer, 1998.

[3] M. Z. Chen, D. H. Zhou, and G. P. Liu, "A new particle predictor for fault prediction of nonlinear time-varying systems," *Developments Chemical Eng. Mineral Process.*, vol. 13, no. 3/4, pp. 379–388, 2005.

[4] W. T. Chen and M. Saif, "A novel fuzzy system with dynamic rule base," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 5, pp. 569–582, Oct. 2005.

[5] K. Y. Chen, "Forecasting system's reliability based on support vector regression with genetic algorithms," *Rel. Eng. Syst. Safety*, vol. 92, no. 4, pp. 423–432, 2007.

[6] L. H. Chen and C. C. Hsueh, "Fuzzy regression models using the least-squares method based on the concept of distance," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 6, pp. 1259–1272, Dec. 2009.

[7] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Ann. Math. Stat.*, vol. 38, no. 2, pp. 325–339, 1967.

[8] A. P. Dempster, "A generalization of Bayesian inference," *J. Royal Stat. Soc., Series B*, vol. 30, no. 2, pp. 205–247, 1968.

[9] A. P. Dempster, "The Dempster–Shafer calculus for statisticians," *Int. J. Approx. Reason.*, vol. 48, no. 2, pp. 365–377, 2008.

[10] J. M. de Costa Sousa and U. Kaymak, *Fuzzy Decision Making in Modeling and Control* (World Sci. Series Robotics Intell. Syst., vol. 27). River Edge, NJ: World Scientific, 2002.

[11] W. Duch, R. Setiono, and J. M. Zurada, "Computational intelligence methods for rule-based data understanding," *Proc. IEEE*, vol. 92, no. 5, pp. 771–805, May 2004.

[12] P. E. Green, J. Douglas, and W. S. Desarbo, "A new measure of predictor variable importance in multiple regression," *J. Market. Res.*, vol. 15, no. 3, pp. 356–360, 1978.

[13] S. Guillaume, "Designing fuzzy inference systems from data: An interpretability-oriented review," *IEEE Trans. Fuzzy Syst.*, vol. 9, no. 3, pp. 426–443, Jun. 2001.

[14] S. Guillaume and B. Charnomordic, "Generating an interpretable family of fuzzy partitions from data," *IEEE Trans. Fuzzy Syst.*, vol. 12, no. 3, pp. 324–335, Jun. 2004.

[15] Y. Halpern and R. Fagin, "Two views of belief: Belief as generalized probability and belief as evidence," *Artif. Intell.*, vol. 54, no. 3, pp. 275–317, 1992.

[16] X. D. He, W. L. C. Chu, and H. J. Yang, "A new approach to verify rule-based systems using Petri nets," *Inf. Software Technol.*, vol. 45, pp. 663–669, 2003.

[17] L. J. Herrera, H. Pomares, I. Rojas, O. Valenzuela, and A. Prieto, "Tase, a Taylor series-based fuzzy system model that combines interpretability and accuracy," *Fuzzy Sets Syst.*, vol. 153, pp. 403–427, 2005.

[18] S. L. Ho and M. Xie, "The use of ARIMA models for reliability forecasting and analysis," *Comput. Ind. Eng.*, vol. 35, no. 1/2, pp. 213–216, 1998.

[19] C. H. Hu, X. S. Si, and J. B. Yang, "System reliability prediction model based on evidential reasoning algorithm with nonlinear optimization," *Expert Syst. Appl.*, vol. 37, no. 3, pp. 2550–2562, 2010.

[20] C. L. Huang and K. Yong, *Multiple Attribute Decision Making Methods and Applications, a State-of Art Survey*. Berlin, Germany: Springer-Verlag, 1981.

[21] Y. Q. Jiang, "Applying grey forecasting to predicting the operating energy performance of air cooled water chillers," *Int. J. Refrigeration*, vol. 27, no. 4, pp. 385–392, 2004.

[22] C. F. Juang and C. D. Hsieh, "A locally recurrent fuzzy neural network with support vector regression for dynamic system modeling," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 2, pp. 261–273, Apr. 2010.

[23] E. Kim, M. Park, S. Ji, and M. Park, "A new approach to fuzzy modeling," *IEEE Trans. Fuzzy Syst.*, vol. 5, no. 3, pp. 328–337, Aug. 1997.

[24] L. W. Lee, L. H. Wang, S. M. Chen, and Y. H. Leu, "Handling forecasting problems based on two-factors high-order fuzzy time series," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 3, pp. 468–477, Jun. 2006.

[25] G. D. Li, S. Masuda, D. Yamaguchi, and M. Nagai, "A new reliability prediction model in manufacturing systems," *IEEE Trans. Rel.*, vol. 59, no. 1, pp. 170–177, Mar. 2010.

[26] J. Liu, J. B. Yang, 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.

[27] J. Liu, J. B. Yang, and H. S. Sii, "Engineering system safety analysis and synthesis using fuzzy rule based evidential reasoning approach," *Quality Rel. Eng. Int.*, vol. 21, no. 4, pp. 387–411, 2005.

[28] J. Liu, J. B. Yang, D. Ruan, L. Martinez, and J. Wang, "Self-tuning of fuzzy belief rule bases for engineering system safety analysis," *Ann. Operational Res.*, vol. 163, pp. 143–168, 2008.

[29] E. D. Lughofer, "FLEXFIS: A robust incremental learning approach for evolving Takagi–Sugeno fuzzy models," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 6, pp. 1393–1410, Dec. 2008.

[30] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Struct. Multidiscip. Optim.*, vol. 26, pp. 369–395, 2004.

[31] K. Miettine, *Nonlinear Multiobjective Optimization*. Boston, MA: Kluwer, 1999.

[32] D. L. Nazareth, "Issues in the verification of knowledge in rule-based systems," *Int. J. Man–Mach. Stud.*, vol. 30, pp. 255–271, 1989.

[33] G. Niu and B. S. Yang, "Dempster–Shafer regression for multi-step-ahead time-series prediction towards data-driven machinery prognosis," *Mech. Syst. Signal Process.*, vol. 23, pp. 740–751, 2009.

[34] N. Ogris and M. Jurc, "Sanitary felling of Norway spruce due to spruce bark beetles in Slovenia: A model and projection for various climate change scenarios," *Ecol. Model.*, vol. 221, no. 2, pp. 290–302, 2010.

[35] R. A. Peterson, "A meta-analysis of variance accounted for and factor loadings in exploratory factor analysis," *Marketing Lett.*, vol. 11, no. 3, pp. 261–275, 2000.

[36] H. T. Pham and B. S. Yang, "Estimation and forecasting of machine health condition using ARMA/GARCH model," *Mech. Syst. Signal Process.*, vol. 24, pp. 546–558, 2010.

[37] E. A. Rietman and M. Beachy, "A study on failure prediction in a plasma reactor," *IEEE Trans. Semicond. Manuf.*, vol. 11, no. 4, pp. 670–680, Nov. 1998.

[38] M. J. Roemer, C. S. Byington, G. J. Kacprzynski, and G. Vachtevanos, "An overview of selected prognostic technologies with reference to an integrated PHM architecture," in *Proc. IEEE Aerosp. Conf.*, Big Sky, MT, 2005.

[39] M. Setnes, R. Babuska, U. Kaymak, and H. R. van Nauta Lemke, "Similarity measures in fuzzy rule base simplification," *IEEE Trans. Syst., Man, Cybern., B, Cybern.*, vol. 28, no. 3, pp. 376–386, Jun. 1998.

[40] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton Univ. Press, 1976.

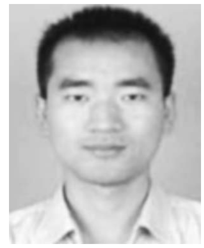
[41] P. Smets and R. Kennes, "The transferable belief model," *Artif. Intell.*, vol. 66, no. 2, pp. 191–243, 1994.

[42] W. Stach, L. Kurgan, and W. Pedrycz, "Numerical and linguistic prediction of time series with the use of fuzzy cognitive maps," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 1, pp. 61–72, Feb. 2008.

[43] R. Sun, "Robust reasoning: Integrating rule-based and similarity-based reasoning," *Artif. Intell.*, vol. 75, no. 2, pp. 241–295, 1995.

[44] J. A. K. Suykens, T. van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Support Vector Machines*. Singapore: World Scientific, 2002.

[45] H. Theil, *Economic Forecast and Policy*. Amsterdam, the Netherlands: North-Holland, 1961, p. 567.
[46] L. Trapani and G. Urga, "Optimal forecasting with heterogeneous panels: A Monte Carlo study," *Int. J. Forecasting*, vol. 25, pp. 567–586, 2007.
[47] M. B. A. van Asselt, J. Mesman, and S. A. van't Klooster, "Dealing with prognostic uncertainty," *Futures*, vol. 39, no. 6, pp. 669–684, 2007.
[48] J. van den Berg, U. Kaymak, and W.-M. van den Bergh, "Financial markets analysis by using a probabilistic fuzzy modelling approach," *Int. J. Approx. Reason.*, vol. 35, pp. 291–305, 2004.
[49] P. Walley, "Measures of uncertainty in expert system," *Artif. Intell.*, vol. 83, no. 1, pp. 1–58, 1996.
[50] D. Wang, D. H. Zhou, Y. H. Jin, and S. J. Qin, "A strong tracking predictor for nonlinear processes with input time delay," *Comput. Chem. Eng.*, vol. 28, no. 12, pp. 2523–2540, 2004.
[51] Y. M. Wang, J. B. Yang, and D. L. Xu, "Environmental impact assessment using the evidential reasoning approach," *Eur. J. Operational Res.*, vol. 174, no. 3, pp. 1885–1913, 2006.
[52] W. Wang, "An adaptive predictor for dynamic systems forecasting," *Mech. Syst. Signal Process.*, vol. 21, pp. 809–823, 2007.
[53] Y. M. Wang, J. B. Yang, D. L. Xu, and K. S. Chin, "The evidential reasoning approach for multiple attribute decision analysis using interval belief degrees," *Eur. J. Operational Res.*, vol. 175, no. 1, pp. 35–66, 2008.
[54] W. Wang and J. Vrbanek, "An evolving fuzzy predictor for industrial applications," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 6, pp. 1439–1449, Dec. 2008.
[55] O. J. Woodman. (2007). An introduction to inertial navigation, Comput. Lab., Univ. Cambridge, Cambridge, MA, Tech. Rep. [Online]. Available: http://www.cl.cam.ac.uk/techreport/
[56] D. Wang, X. J. Zeng, and J. A. Keane, "An evolving construction scheme for fuzzy systems," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 4, pp. 755–770, Aug. 2010.
[57] K. Xu, M. Xie, L. C. Tang, and S. L. Ho, "Application of neural networks in forecasting engine systems reliability," *Appl. Soft Comput.*, vol. 2, no. 4, pp. 255–268, 2003.
[58] D. L. Xu, J. Liu, J. B. Yang, G. P. Liu, J. Wang, I. Jenkinson, and J. Ren, "Inference and learning methodology of belief-rule-based expert system for pipeline leak detection," *Expert Syst. with Appl.*, vol. 32, no. 1, pp. 103–113, 2007.
[59] J. B. Yang and M. G. Singh, "An evidential reasoning approach for multiple attribute decision making with uncertainty," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 24, no. 1, pp. 1–18, Jan. 1994.
[60] S. K. Yang and T. S. Liu, "A Petri net approach to early failure detection and isolation for preventive maintenance," *Qual. Rel. Eng. Int.*, vol. 14, no. 5, pp. 319–330, 1998.
[61] S. K. Yang and T. S. Liu, "State estimation for predictive maintenance using Kalman filter," *Rel. Eng. Syst. Safety*, vol. 66, no. 1, pp. 29–39, 1999.
[62] J. B. Yang, "Rule and utility based evidential reasoning approach for multi-attribute decision analysis under uncertainties," *Eur. J. Operational Res.*, vol. 131, no. 1, pp. 31–61, 2001.
[63] 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.
[64] J. B. Yang, J. Liu, J. Wang, H. S. Sii, and H. W. Wang, "Belief rule-base inference methodology using the evidential reasoning approach—RIMER," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 36, no. 2, pp. 266–285, Mar. 2006.
[65] J. B. Yang, Y. M. Wang, D. L. Xu, and K. S. Chin, "The evidential reasoning approach for MADA under both probabilistic and fuzzy uncertainties," *Eur. J. Operational Res.*, vol. 171, no. 1, pp. 309–343, 2006.
[66] J. B. Yang, J. Liu, D. L. Xu, J. Wang, and H. W. Wang, "Optimal learning method for training belief rule based systems," *IEEE Trans. Syst., Man, Cybern. Part A, Syst. Humans*, vol. 37, no. 4, pp. 569–585, Jul. 2007.
[67] L. Z. Zadeh, "Fuzzy sets," *Inform. Control*, vol. 8, no. 3, pp. 338–353, 1965.
[68] A. S. Zhang and L. Zhang, "RBF neural networks for the prediction of building interference effects," *Comput. Struct.*, vol. 82, pp. 2333–2339, 2004.
[69] L. B. Zhang, Z. H. Wang, and S. X. Zhao, "Short-term fault prediction of mechanical rotating parts on the basis of fuzzy-grey optimizing method," *Mech. Syst. Signal Process.*, vol. 21, pp. 856–865, 2007.
[70] J. Zheng, "Predicting software reliability with neural network ensembles," *Expert Syst. Appl.*, vol. 36, pp. 2116–2122, 2009.
[71] P. Zito, H. B. Chen, and M. C. Bell, "Predicting real-time roadside CO and NO2 concentrations using neural networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 3, pp. 514–522, Sep. 2008.

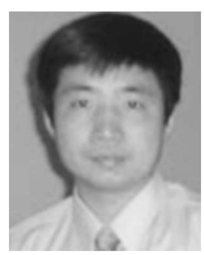


Xiao-Sheng Si received the B.Eng. and M.Eng. degrees, all in control engineering, from the High-Tech Institute of Xi'an, Xi'an, China, in 2006 and 2009, respectively. He is currently working toward the Ph.D. degree with Tsinghua University, Beijing, China.

He has authored or coauthored about 15 articles in several journals including the *European Journal of Operational Research*, the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS PART A, *Expert Systems with Applications*, *Science China Information Sciences*, etc. He has also been the Invited Reviewer for several international journals. His current research interests include evidence theory, expert systems, prognositics and health management, reliability estimation, predictive maintenance, and lifetime estimation.

Chang-Hua Hu received the B.Eng. and M.Eng. degrees from the High-Tech Institute of Xi'an, Xi'an, China, in 1987 and 1990, respectively, and the Ph.D. degree from the Northwestern Polytechnic University, Xi'an, in 1996.

He is currently a Professor with the High-Tech Institute of Xi'an. During September 2008–December 2008, he was a Visiting Scholar with the University of Duisburg, Duisburg, Germany. His current research was supported by the National Science Foundation of China. He has authored or coauthored two books and about 100 articles. His current research interests include fault diagnosis and prediction, life prognosis, and fault-tolerant control.



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

He is currently a Professor of decision and system sciences and the Director of the Decision Sciences Research Center, Manchester Business School, University of Manchester, Manchester, U.K. He is also a Specially Appointed Changjian Chair Professor with the School of Management, Hefei University of Technology, Hefei, China. He was a Faculty Member with the University of Manchester Institute of Science and Technology, in 1990 and again from 1998 and 2004, the University of Birmingham, Birmingham, U.K., from 1995 to 1997, the University of Newcastle upon Tyne, Newcastle upon Tyne, U.K., from 1991 to 1995, and Shanghai Jiao Tong University, Shanghai, China, from 1987 to 1989. Over 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 research interests include design decision making, risk modeling and 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. He has authored or coauthored three books, more than 140 papers in international journals/book chapters, and a similar number of papers in conferences. He has developed several software packages in these areas, including the Windows-based intelligent decision system via evidential reasoning.

Zhi-Jie Zhou received the B.Eng. and M.Eng. degrees from the High-Tech Institute of Xi'an, Xi'an, China, in 2001 and 2004, respectively. He is currently working toward the Ph.D. degree with Tsinghua University, Beijing, China.

He is currently a Lecturer with the High-Tech Institute of Xi'an. He was a Visiting Scholar with the University of Manchester, Manchester, U.K., from March to August 2009. He has authored or coauthored about 25 articles. His current research interests include belief rule bases, dynamic system modeling, hybrid quantitative and qualitative decision modeling and fault prognosis, and optimal maintenance of dynamic systems.