第 29 卷第 12 期
2012 年 12 月

控 制 理 论 与 应 用
Control Theory & Applications

Vol. 29 No. 12
Dec. 2012

文章编号: 1000−8152(2012)12−1579−08

基于进化信度规则库的故障预测

司小胜[1,2], 胡昌华[1], 张 琪[1], 李 娟[3]

(1. 第二炮兵工程大学 302教研室, 陕西 西安 710025;

2. 清华大学 自动化系, 北京 100084; 3. 青岛农业大学 机电工程学院, 山东 青岛 266109)

摘要: 在假设信度规则库(BRB)的输入为均匀分布的情况下, 已有文献提出了一种序贯自适应的学习算法以实现BRB的参数在线辨识和结构的自适应调整. 然而在实际问题中, 信度规则库的输入一般是未知的、难以得到的, 这在一定程度上限制了序贯自适应学习算法的实用性, 因此就需要研究一种改进的BRB学习算法以实现参数和结构的同时辨识. 本文在序贯自适应方法的基础上, 通过定义BRB的完整性准则, 提出了改进的BRB进化策略. 与现有方法相比, 该方法可以实现信度规则的自动增减, 且无需输入样本的概率密度函数. 此外, 该方法继承了BRB的特点, 仅需要部分的输入输出信息. 基于改进的进化策略, 提出了一种新的故障预测算法, 最后通过陀螺仪故障预测实验验证了本文方法的有效性.

关键词: 专家系统; 信度规则库; 证据推理; 故障预测

中图分类号: TP206+.3　　**文献标识码**: A

Fault prognosis based on evolving belief-rule-base system

SI Xiao-sheng[1,2], HU Chang-hua[1], ZHANG Qi[1], LI Juan[3]

(1. Department of Automation, Xi'an Institute of High-Tech, Xi'an Shaanxi 710025, China;
2. Department of Automation, Tsinghua University, Beijing 100084, China;
3. College of Mechanical and Electrical Engineering, Qingdao Agricultural University, Qingdao, Shandong 266109, China)

Abstract: Recently, a sequential adaptive learning algorithm has been developed for online constructing belief-rule-based (BRB) system. This algorithm is based on the assumption that the sample density function of the inputs to BRB system obeys the uniform distribution. However, in practice, the sample density function is not always available and is difficult to be determined; this really limits the applicability of the above method. As such, it is desired to develop an improved algorithm without requiring the sample density function. In this paper, on the basis of the sequential adaptive learning algorithm, we develop an improved evolving BRB learning algorithm based on the belief-incomplete criterion. Compared with the current algorithms, a belief rule can be automatically added into the BRB or pruned from the BRB without the need of the sample density function. In addition, our algorithm inherits the features of the BRB, in which only partial input and output information are required. Based on the improved algorithm, a fault prognosis method is presented. In order to verify the effectiveness of our algorithm, a practical case study for gyroscope fault prognosis is studied and examined to demonstrate how our algorithm can be implemented.

Key words: expert system; belief-rule-base; evidential reasoning; fault prognosis

1 Introduction

As a key element of condition-based maintenance (CBM), fault prognosis has become very important in prognosis and health management, and is currently an active research area in the world[1–3]. For example, in some systems like nuclear power station and missile control system, it is very significant that before the faults occur[4–6], they can be predicted so as to avoid large calamity. Moreover, it is critically important to conduct fault prognosis while in use since it has direct impacts on the planning of maintenance activities, spare parts provision, operational performance, and the profitability of the owner of the asset. Generally speaking, the current methods can be categorized into three kinds: physics of failure-based method, data-driven method, and fusion method which is the combination of previous two. However, for a specific system under consideration, many fault prediction problems involve quantitative data as well as qualitative knowledge, which may suffer various types of uncertainties such as incompleteness and fuzziness. However, afore-mentioned prediction models are limited in dealing with both numerical data and human judgment information under uncertainty[7–8]. In order to deal with uncertainty in-

Received 26 May 2012; Revised 26 July 2012.

This work was supported by the National Nature Science Foundation of China (No. 61025014, 61174030, 61104223) and the Shandong High School Science & Technology Fund Planning Project (J09LG26).

volved, several frameworks have been constructed[9], such as Bayesian probability theory, Dempster-Shafer (D–S) theory, and fuzzy set theory. Due to the power of the D-S theory in handling uncertainties, it has found wide applications in many areas such as expert systems to date[10–12].

Under the D-S inference framework, a generic rule-base inference methodology using the evidential reasoning approach was proposed to handle hybrid information with uncertainty, also named as belief-rule-base (BRB). This methodology establishes a nonlinear relationship between antecedent attributes and an associated consequence, and can reflect the dynamic nature of inference process. It is known that the IF-THEN rule-based method and the fuzzy IF-THEN rule-based method can only cope with fuzzy uncertainty and are not applicable in cases where there exists probabilistic uncertainty[13]. In contrast, the BRB approach provides a more informative and realistic scheme than the traditional IF–THEN rule-base with respect to knowledge representation. However, there is little literature on fault prognosis only using BRB. This paper fills this gap and only focuses on the BRB approach. In BRB, there are several types of unknown parameters including belief degrees, attribute weights and rule weights. These parameters need to be determined accurately for a specific application. However, it is often difficult to determine these parameters by human, particularly for a large scale rule base. As such, some optimization models have been proposed to train a BRB[10] in an off-line style. As such, the training process is time-consuming. Further, the recursive algorithms for online updating the BRB systems have also been developed and their calculation speed is high[11]. However, these optimal algorithms are all based on the determined structure of the BRB. To make joint adjustment of model structure and parameters to nonlinear systems, adopting the sequential learning algorithm is quite natural. There are some excellent works on sequential adaptive learning algorithms for nonlinear system identification, such as [14] and [15].

For example, a selective recursive kernel learning was proposed in [14] for online identification of nonlinear systems with NARX form and one of its appealing characteristics was that it can adaptively adjust the model structure to capture the process dynamics. It is noted that these algorithms were basically achieved by minimizing the mean-squared error (MSE). In order to enhance the capacity of BRB method, recently, a sequential adaptive learning algorithm has been developed for online constructing belief-rule-based (BRB) systems[12]. The above learning algorithm is based on the assumption that the sampling density function of the inputs of BRB system obeys the uniform distribution. However, the sample density function is not available and is difficult to obtain for a specific case, and this really limits the applicability of the above method. As such, it is desired to develop an improved algorithm without requirement for the sample density function.

In this paper, we first give a definition representing the activated degree of belief rule for the given inputs, and then used this definition as the belief completeness criterion for the BRB system. Furthermore, along the line of the sequential adaptive learning algorithm[12], we develop an improved evolving BRB learning algorithm based on the belief completeness criterion, in which the structure and parameters of BRB can be adjusted online. Compared with the original learning algorithms in [12], a belief rule can be automatically added into the BRB or pruned from the BRB without need of the sample density function and it can be implemented easily. Unlike the algorithms in [14] and [15], the proposed parameter estimation algorithm in this paper is done by maximizing the likelihood function. In addition, our algorithm inherits the feature of the BRB and the sequential adaptive algorithm, in which only partial input and output information are required. Based on the improved method, a fault prognosis is presented explicitly. In order to verify the effectiveness of our algorithm, a practical case study for gyroscope fault prognosis are provided and examined to demonstrate how our algorithm can be implemented. The results show that our algorithm may be widely used in fault prognosis practice.

2 Preliminaries

2.1 Belief rule base

A belief-rule-base (BRB), which represents the dynamics of a system, consists of a collection of belief rules defined as follows[10]:

$$\begin{aligned} R_k:\ & \text{If } x_1 \text{ is } A_1^k \ \wedge\ x_2 \text{ is } A_2^k\ \wedge \cdots \wedge\ x_{M_k} \text{ is } A_{M_k}^k, \\ & \text{Then } \{(D_1, \beta_{1,k}^1),\ \cdots, (D_N, \beta_{N,k}^1)\}, \\ & \text{with a ruleweight } \theta_k^1 \text{ and attribute weight} \\ & \delta_{1,k}^1, \delta_{2,k}^1, \cdots, \delta_{M_k,k}^1, \end{aligned} \tag{1}$$

where $x_1, x_2, \cdots, x_{M_k}$ represents the antecedent attributes in the kth rule. $A_i^k(i=1,\cdots,M_k,\ k=1,\cdots,L)$ is the referential value of the ith antecedent attribute in the kth rule and $A_i^k \in A_i$. $A_i=\{A_{i,j},\ j=1,\cdots,J_i\}$ is a set of referential values for the ith antecedent attribute and J_i is the number of the referential values. $\theta_k \in \mathbb{R}^+(k=1,\cdots,L)$ is the relative weight of the kth rule, and $\delta_{1,k}, \delta_{2,k}, \cdots, \delta_{M_k,k}$ are the relative weights of the M_k antecedent attributes used in the kth rule. $\beta_{j,k}(j=1,\cdots,N,\ k=1,\cdots,L)$ is the belief degree assessed to D_j which denotes the jth consequence. If $\sum\limits_{j=1}^{N}\beta_{j,k}=1$, the kth rule is said to be complete; otherwise, it is incomplete. Note that '$\wedge$' is a logical connective to represent the 'AND' relationship. In addition, we suppose that M is the total number of antecedent attributes used in the rule base. In the following, we use $\boldsymbol{x}(n)=[x_1(n)\ \ \cdots\ \ x_M(n)]^{\mathrm{T}}$ as the input vector of BRB at time instant n.

2.2 Rule-based information transformation technique for quantitative data

In this paper, we only consider the case that the input of BRB system is given by numerical values, that is, all elements of $\boldsymbol{x}(n)$ are numerical values. In this case, equivalence rules need to be extracted from the decision maker to achieve transformation from numerical value into belief structure[9–10, 16]. This can be used to transform a value to an equivalent expectation, thereby relating a particular value to a set of referential values, named as rule-based transformation technique [16]. In this technique, a value $\gamma_{i,j}(i=1,\cdots,M,\ j=1,\cdots,J_i)$ can be judged to be a referential value $A_{i,j}$ in a BRB, or

$$\gamma_{i,j}\text{means } A_{i,j},\ i=1,\cdots,M,\ j=1,\cdots,J_i. \quad (2)$$

Assume that a larger value $\gamma_{i,(j+1)}$ is preferred over a smaller value $\gamma_{i,j}$. Let γ_{i,J_i} and $\gamma_{i,1}$ be the largest and smallest feasible values, respectively. Then, each element $x_i(n)$ in the input vector $\boldsymbol{x}(n)$ can be represented using the following equivalent expectation:

$$S(x_i(n))=\{(\gamma_{i,j},\alpha_{i,j}(x_i(n))),\ i=1,\cdots,M,\ j=1,\cdots,J_i\}, \quad (3)$$

where $\alpha_{i,j}(x_i(n))$ can be calculated by

$$\begin{cases}\alpha_{i,j}(x_i(n))=\dfrac{\gamma_{i,j+1}-x_i(n)}{\gamma_{i,j+1}-\gamma_{i,j}},\\ \gamma_{i,j}\leqslant x_i(n)\leqslant\gamma_{i,j+1},\ j=1,\cdots,J_i-1,\end{cases} \quad (4)$$

$$\begin{cases}\alpha_{i,j+1}(x_i(n))=1-\alpha_{i,j}(x_i(n)),\\ \gamma_{i,j}\leqslant x_i(n)\leqslant\gamma_{i,j+1},\ j=1,\cdots,J_i-1,\end{cases} \quad (5)$$

$$\alpha_{i,s}(x_i(n))=0 \text{ for } s=1,\cdots,J_i,\ s\neq j,j+1. \quad (6)$$

The detailed rule-based transformation technique can be found in [14].

2.3 Belief-rule inference using the evidential reasoning approach

When the antecedent attributes, i.e., the inputs of the BRB are available, the evidential reasoning (ER) approach[17] is used as the inference tool. Using the ER analytical algorithms[10, 18], the final conclusion $\mathrm{O}(\hat{\boldsymbol{y}}(n))$ that is generated by aggregating all rules, activated by the actual input vector $\boldsymbol{x}(n)$, can be represented as follows:

$$\mathrm{O}(\hat{\boldsymbol{y}}(n))=h(\boldsymbol{x}(n))=\{(D_j,\beta_j(\boldsymbol{x}(n))),\ j=1,\cdots,N\}, \quad (7)$$

where $\beta_j(\boldsymbol{x}(n))$ denotes the belief degree in D_j at time instant n, and (8) and (9) hold.

$$\beta_j(\boldsymbol{x}(n))=\frac{\mu(\boldsymbol{x}(n))}{1-\mu(\boldsymbol{x}(n))\times[\prod\limits_{k=1}^{L}(1-\omega_k(\boldsymbol{x}(n)))]}\times [\prod_{k=1}^{L}(\omega_k(\boldsymbol{x}(n))\beta_{j,k}+1-\omega_k(\boldsymbol{x}(n))\sum_{i=1}^{N}\beta_{i,k})-\prod_{k=1}^{L}(1-\omega_k(\boldsymbol{x}(n))\sum_{i=1}^{N}\beta_{i,k})], \quad (8)$$

$$\mu(\boldsymbol{x}(n))=[\sum_{j=1}^{N}\prod_{k=1}^{L}(\omega_k(\boldsymbol{x}(n))\beta_{j,k}+1-\omega_k(\boldsymbol{x}(n))\sum_{i=1}^{N}\beta_{i,k})-(N-1)\prod_{k=1}^{L}(1-\omega_k(\boldsymbol{x}(n))\sum_{i=1}^{N}\beta_{i,k})]^{-1}, \quad (9)$$

where $\beta_j(\hat{\boldsymbol{x}}(n)t)$ is the function of the belief degrees $\beta_{i,k}\,(i=1,\cdots,N,k=1,\cdots,L)$, the rule weights $\theta_k\,(k=1,\cdots,L)$, the attribute weights $\bar{\delta}_i\,(i=1,\cdots,M)$, and the input vector $\boldsymbol{x}\,(n)$. $\omega_k\,(\boldsymbol{x}\,(n))$, the activation weight of the kth rule at time instant n, can be calculated by

$$\begin{cases}\omega_k(\hat{\boldsymbol{x}}(n))=\dfrac{\theta_k\prod\limits_{i=1}^{M}(\alpha_i^k(x_i(n)))^{\bar{\delta}_i}}{\sum\limits_{l=1}^{L}\theta_l\prod\limits_{i=1}^{M}(\alpha_i^l(x_i(n)))^{\bar{\delta}_i}},\\ \bar{\delta}_i=\dfrac{\delta_i}{\max\limits_{i=1,\cdots,M}\{\delta_i\}},\end{cases} \quad (10)$$

where $\delta_i\in\mathbb{R}^+(i=1,\cdots,M)$ is the relative weight of the ith antecedent attribute used in the kth rule. $\alpha_i^k(x_i(n))\in\{\alpha_{i,j}(x_i(n)),\ i=1,\cdots,M,\ j=1,\cdots,J_i\}$, the individual matching degree, is the degree of belief to its jth referential value $A_{i,j}^k$ in the kth rule at time instant n. $\alpha_k(\boldsymbol{x}(n))=\prod\limits_{i=1}^{M}(\alpha_i^k(x_i(n)))^{\bar{\delta}_i}$ is called the normalized combined matching degree.

The logic behind the approach is that, if the consequence in the kth rule includes D_i with $\beta_{i,k}>0$ and the kth rule is activated, then the overall output must be D_i to a certain degree. This degree is measured by both the degree to which the kth rule is important to the overall output and the degree to which the antecedents of the kth rule are activated by the actual input $\boldsymbol{x}(n)$[9–10]. Then for the numerical output case considered in this paper by using utility concept, the predicted output $\hat{\boldsymbol{y}}(n)$ is formulated as

$$\hat{\boldsymbol{y}}(n)=\sum_{j=1}^{N}\mu(D_j)\beta_j(\hat{\boldsymbol{x}}(n)), \quad (11)$$

where $\mu(D_j)$ represents the utility of an individual consequent D_j and $\beta_j(\boldsymbol{x}(n))$ is calculated by (8).

3 An evolving strategy for BRB

As analyzed in introduction, the current sequential learning algorithm[12] for BRB is based on the assumption that the sampling density function of the inputs of BRB system obeys the uniform distribution. However, the sample density function is not frequently available in practice. As such, developing an improved method without the sample density function required is desired from the application point of view. Similar to fuzzy rule system[19], we define a belief ε-completeness for rule adjustment in the following.

3.1 ε-completeness criterion for rule adjustment in BRB

Definition 1 ε-completeness of belief rules. For any inputs in the operating range, there exists at least one belief rule so that the matched degree is no less than ε. In application, the minimum value of ε is usually selected as $\varepsilon = 0.5$.

According to this definition, we can easily learn that $\omega_k(\boldsymbol{x}(n))$ in (10) can be used for calculating the matched degree between the given input $\boldsymbol{x}(n)$ and belief rule k at time instant n. As such, we can define the quantitative value of ε-completeness of belief rules as the minimum value of the activated degree of the belief rules which should be satisfied.

Definition 2 for a compact belief-rule-base, the following condition should be satisfied using ε-completeness of belief rules

$$\min_k\{\omega_k(\boldsymbol{x}(n)),\ k = 1,\cdots,L\} \geqslant \varepsilon. \tag{12}$$

Then from the viewpoint of belief rules, a belief rule is inherent to a local representation over a region defined in the input space. If a new input pattern satisfies the ε-completeness criterion of belief rules, the BRB system will not generate a new belief rule but accommodate the new input and output sample by updating the parameters of existing rules. In addition, if the activated degrees of some belief rules are very small, this means these rules are less important for current input pattern and so they are pruned to maintain the compact ability of BRB. In the following subsections, we give the detailed rule adjustment strategy according to Definitions 1 and 2.

3.2 Adding a belief rule

Firstly, suppose that in an initial BRB, there have been L belief rules, M antecedent attributes, $J_i(i = 1,\cdots,M)$ referential values of the ith antecedent attribute and N consequences. Furthermore, we assume that these L belief rules are all significant. Here adding a belief rule can be interpreted as follows: if the criteria as given later are satisfied on the basis of the available input and output information of the BRB system, some new referential values of the antecedent attribute are added.

According to the above ε-completeness of belief rules, when a new observed data pair $(\boldsymbol{x}(n), \boldsymbol{y}(n))$ arrives at time instant n, where $\boldsymbol{x}(n)$ is an input vector and $\boldsymbol{y}(n)$ is the corresponding output vector, the following two criteria may be used to determine whether a new belief rule is added[12,20].

$$\begin{cases} \|\boldsymbol{x}(n) - \bar{\boldsymbol{\gamma}}(n)\| > e_{\mathrm{g}}, \\ \min\limits_k\{\omega_k(\boldsymbol{x}(n)),\ k = 1,\cdots,L\} \geqslant \varepsilon, \end{cases} \tag{13}$$

where $\bar{\boldsymbol{\gamma}}(n) = [\bar{\gamma}_1(n)\ \ \cdots\ \ \bar{\gamma}_M(n)]^{\mathrm{T}},\ \bar{\gamma}_i(n) \in \{\gamma_{i,j},\ i = 1,\cdots,M,\ j = 1,\cdots,J_i\}$ and $\|\cdot\|$ is the Euclidean norm. $\bar{\boldsymbol{\gamma}}(n)$ denotes the referential vector of the antecedent attributes of a belief rule being nearest to $\boldsymbol{x}(n)$ under the Euclidean distance sense. ε is a threshold to be selected as 0.5 in this paper and e_{g} is the expected approximation accuracy. $L+1$ is the number of belief rule in the updated BRB.

Once the two criteria given in (13) are satisfied, a new belief rule, i.e., the $(L+1)$ th rule, may be added. The parameters of the new rule can be determined as follows[12]:

1) The referential value vector of the antecedent attributes is

$$\gamma_{L+1} = \boldsymbol{x}(n). \tag{14}$$

2) The belief degree $\beta_{j,L+1}(j = 1,\cdots,N)$ which D_j is assessed for $\boldsymbol{y}(n)$ can be determined using rule based information transformation technique[16]:

$$\begin{cases} \beta_{j,L+1} = \dfrac{u(D_{j+1}) - \boldsymbol{y}(n)}{u(D_{j+1}) - u(D_j)}, \\ u(D_j) \leqslant \boldsymbol{y}(n) \leqslant u(D_{j+1}),\ j = 1,\cdots,N-1, \end{cases} \tag{15}$$

$$\begin{cases} \beta_{j+1,L+1} = 1 - \beta_{j,L+1}, \\ u(D_j) \leqslant \boldsymbol{y}(n) \leqslant u(D_{j+1}),\ j = 1,\cdots,N-1, \end{cases} \tag{16}$$

$$\beta_{s,L+1} = 0 \text{ for } s = 1,\cdots,N,\ s \neq j, j+1. \tag{17}$$

3) The weight of antecedent attribute $\bar{\delta}_i(i = 1,\cdots,M)$ in the $(L+1)$th belief rule is the same as one in the other rules. The rule weight can be set as $\theta_{L+1} = 1$.

3.3 Parameter adjustment

Once the structure of a BRB is determined using the observed data pair $(\boldsymbol{x}(n), \boldsymbol{y}(n))$, some parameters, such as the rule weights, the attribute weights, and the belief degrees, should be updated using $(\boldsymbol{x}(n), \boldsymbol{y}(n))$. Along the line of the method developed in [16], first, we assume that if the inputs of the BRB $\boldsymbol{x}(1),\cdots,\boldsymbol{x}(n)$ are independent, the true outputs, $\boldsymbol{y}(1),\cdots,\boldsymbol{y}(n)$, can also be assumed to be independent. Therefore, there exists

$$f(\boldsymbol{y}(1),\cdots,\boldsymbol{y}(n)|\boldsymbol{x}(1),\cdots,\boldsymbol{x}(n),\boldsymbol{Q}) = \prod_{\tau=1}^{n} f(\boldsymbol{y}(\tau)|\boldsymbol{x}(\tau),\boldsymbol{Q}), \tag{18}$$

where $\boldsymbol{y}$ is numerical and is considered as a random variable. $f(\boldsymbol{y}(\tau)|\boldsymbol{x}(\tau),\boldsymbol{Q})$ is assumed to be the conditional probability density function (PDF) of $\boldsymbol{y}$ at time instant τ and $\boldsymbol{Q}$ is the unknown parameter vector.

The expectation of the log-likelihood of (18) at time instant n is defined as

$$\boldsymbol{L}_{n+1}(\boldsymbol{Q}) \triangleq \mathrm{E}\{\sum_{\tau=1}^{n} \log f(\boldsymbol{y}(\tau)|\boldsymbol{x}(\tau),\boldsymbol{Q})|\boldsymbol{x}(1), \boldsymbol{x}(2),\cdots,\boldsymbol{x}(n),\boldsymbol{Q}(n)\}, \tag{19}$$

where $\mathrm{E}\{\cdot|\cdot\}$ denotes the conditional expectation at $\boldsymbol{Q}=\boldsymbol{Q}(n)$.

The recursive formulation of (19) can be written as

$$\boldsymbol{L}_{n+1}(\boldsymbol{Q}) = \boldsymbol{L}_{n}(\boldsymbol{Q}) + \mathrm{E}\{\log f(\boldsymbol{y}(n)|\boldsymbol{x}(n),\boldsymbol{Q})|\boldsymbol{x}(n),\boldsymbol{Q}(n)\}. \tag{20}$$

Define

$$\varGamma(\boldsymbol{Q}(n)) \triangleq \nabla_{\boldsymbol{Q}} \log f(\boldsymbol{y}(n)|\boldsymbol{x}(n),\boldsymbol{Q}(n)), \tag{21}$$

$$\varXi(\boldsymbol{Q}(n)) \triangleq E\{-\nabla_{\boldsymbol{Q}}\nabla_{\boldsymbol{Q}}^{\mathrm{T}} \log f(\boldsymbol{y}(n)|\boldsymbol{x}(n),\boldsymbol{Q})|\boldsymbol{x}(n),\boldsymbol{Q}(n)\}. \tag{22}$$

Based on the recursive EM algorithm[21–23], the maximizing parameter $\boldsymbol{Q}(n+1)$ is given by [12] and [21],

$$\boldsymbol{Q}(n+1) = \boldsymbol{Q}(n) + \frac{1}{n}[\varXi(\boldsymbol{Q}(n))]^{-1}\varGamma(\boldsymbol{Q}(n)), \tag{23}$$

where $\boldsymbol{Q}$ consists of the rule weights, attribute weights, belief degrees satisfying the equality and inequality constraints[9]:

$$0 \leqslant \theta_k \leqslant 1,\ k=1,\cdots,L, \tag{24}$$

$$0 \leqslant \bar{\delta}_m \leqslant 1,\ m=1,\cdots,M, \tag{25}$$

$$0 \leqslant \beta_{j.k} \leqslant 1,\ j=1,\cdots,N,\ k=1,\cdots,L, \tag{26}$$

$$\sum_{j=1}^{N} \beta_{j,k} = 1,\ k=1,\cdots,L. \tag{27}$$

Hence, the recursive algorithm (23) can be revised as follows:

$$\boldsymbol{Q}(n+1) = \prod_{H}\{\boldsymbol{Q}(n) + \tfrac{1}{n}[\varXi(\boldsymbol{Q}(n))]^{-1}\varGamma(\boldsymbol{Q}(n))\}, \tag{28}$$

where H is a constraint set composed of the constraints (24)−(27), and $\prod_{H}\{\cdot\}$ is the projection onto the constraint set H_1, ensuring that the estimation of $\boldsymbol{Q}$ can satisfy the given constraints. The detailed algorithm of $\prod_{H}\{\cdot\}$ has been given in [11]. In order to obtain the analytic formulations of $\varXi(\boldsymbol{Q}(n))$ and $\varGamma(\boldsymbol{Q}(n))$, the following assumption is given.

Generally speaking, for prediction problem, we hope that for a given input $\boldsymbol{x}(n)$, the BRB system can generate an predicted output $\hat{\boldsymbol{y}}(n)$ as calculated by (11), as close to $\boldsymbol{y}(n)$ as possible. Here $\hat{\boldsymbol{y}}(n)$ is considered as a random variable and $\boldsymbol{y}(n)$ can be considered as its expectation. As such, we expect that $(\boldsymbol{y}(n)-\hat{\boldsymbol{y}}(n)) \sim N(0,\sigma^2)$. Hence we assume that the PDF of $\boldsymbol{y}(n)$ obeys the following normal distribution:

$$f(\boldsymbol{y}(n)|\boldsymbol{x}(n),\boldsymbol{Q}) = \frac{1}{\sqrt{2\pi\sigma}}\exp\{-\frac{(\boldsymbol{y}(n)-\hat{\boldsymbol{y}}(n))^2}{2\sigma}\}, \tag{29}$$

where $\boldsymbol{Q} = [\boldsymbol{V}^{\mathrm{T}}\ \ \sigma]^{\mathrm{T}}$ denotes the parameter vector and σ denotes variance. $\boldsymbol{V} = [\theta_k\ \ \bar{\delta}_m\ \ \beta_{j,k}]^{\mathrm{T}}$ is parameter vector of the BRB and $k=1,\cdots,L,\ m=1,\cdots,M,\ j=1,\cdots,N$. When (29) is put into (28), the analytic formulation of the recursive algorithm can be obtained. Due to the independence between the elements of $\boldsymbol{V}$ and the entries of σ, they can be updated independently using the following formulations:

$$\sigma(n) = \arg\max_{\sigma} \log f(\boldsymbol{y}(n)|\boldsymbol{x}(n),\boldsymbol{Q})|_{\boldsymbol{V}=\boldsymbol{V}(n)} = (\boldsymbol{y}(n)-\hat{\boldsymbol{y}}(n))^2|_{\boldsymbol{V}=\boldsymbol{V}}(n), \tag{30}$$

$$\boldsymbol{V}(n+1) = \prod_{H_1}\{\boldsymbol{V}(n) + \frac{1}{n}[\varXi'(\boldsymbol{Q}(n))]^{-1}\varGamma'(\boldsymbol{Q}(n))\}, \tag{31}$$

where $\varXi'(\boldsymbol{Q}(n))$ and $\varGamma'(\boldsymbol{Q}(n))$ can be obtained by (21) and (22) with respect to $\boldsymbol{V}$, since the elements of $\boldsymbol{V}$ and the entries of σ have been treated separately.

In addition, a group of belief rules in BRB model are activated at time instant n, which means that only some parameters can be updated, so the matrix $\varXi'(\boldsymbol{Q}(n))$ may be singular and needs to be revised. Thus, the final recursive algorithm can be written as follows:

$$\boldsymbol{V}(n+1) = \pi_2\{\boldsymbol{V}(n) + \frac{\alpha}{n}\pi_1\{\boldsymbol{V}(n)\} \times [\varXi'(\boldsymbol{Q}(n)) + \gamma\boldsymbol{I}_{\mathrm{p}}]^{-1}\varGamma'(\boldsymbol{Q}(n))\}, \tag{32}$$

where $\alpha > 0$ is the step factor and can change the convergence speed. The matrix $\varXi_1'(\boldsymbol{Q}(n))$ is amended using $\gamma\boldsymbol{I}_{\mathrm{p}}$ so that it becomes positive definite and $\gamma \geqslant 0$, which is termed as the revision factor. In (32), $\boldsymbol{\pi}_1\{\boldsymbol{V}(n)\}$ is used to deal with the equality constraints (27), defined as $(1-1/N)\boldsymbol{I}_N$ and $\boldsymbol{\pi}_2\{\cdot\}$ is used to deal with inequality constraints (24)−(26) so that the updated parameters can all be within the given bound. In this paper, we adopt a simple strategy that, if the updated parameter is within the bound, maintain as it is; if the updated parameter goes beyond the upper bound, let it equal to the upper bound, vice versa.

3.4 Pruning of a belief rule

After adding a new rule and parameters updating, if the statistical utility of the kth belief rule is less than the given threshold $e_{\rm p}$, i.e., this rule is insignificant, it should be removed. The criterion to prune a belief rule can be described by

$$\omega_k(\boldsymbol{x}(n)) < e_{\rm p},\ k = 1, \cdots, L. \tag{33}$$

In (13) and (33), parameters ε, $e_{\rm g}$ and $e_{\rm p}$ should be determined or chosen in advance. Obviously, if ε and $e_{\rm p}$ are smaller, the system performance is better, but the resulting BRB's structure is more complex, which is a disadvantage when there is high real-time requirement. Therefore, there should be a reasonable trade-off between the system performance and BRB's structure. In specific application, we should choose these parameters carefully according to the expected system performance.

3.5 Fault identification

As to fault prognosis, we consider the threshold-based fault identification method based on the principle that if a system is in normal state, the characteristic parameter expressing the system operating condition will be changed in a fixed bound; when the value the characteristic parameter goes beyond a constant level, fault is declared[6]. This critical level is defined as fault threshold $\bar{\boldsymbol{y}}$. In this paper, we use the concept of the distance between the predicted output of the BRB model and such threshold to define the occurrence of the fault from a conservative point of view, since our method will be demonstrated using drift data of gyro, which is a safety-critical device used widely in inertial navigation system and we must ensure enough time for scheduling maintenance or replacement if fault will occur. Considering that $\hat{\boldsymbol{y}}\,(n)$ is the predicted value of the characteristic parameter at time instant n, when the distance between $\hat{\boldsymbol{y}}\,(n)$ and $\bar{\boldsymbol{y}}$ go beyond a small index $\phi_{\rm th}$, we can declare that the operating system is in a failed state. Namely,

$$d_t = \|\hat{\boldsymbol{y}}(n) - \bar{\boldsymbol{y}}\| \geqslant \phi_{\rm th}, \tag{34}$$

where $\|\cdot\|$ represents the normal operator.

As a result of the above discussion, the procedure for fault prognosis using the improved BRB updating method may be summarized as follows.

Step 1 Initialization step–$\boldsymbol{Q}(0)$, α, γ, ε, $e_{\rm g}$, $e_{\rm p}$, and $\phi_{\rm th}$.

Step 2 Adding rule step–when $\boldsymbol{x}(n)$, $\boldsymbol{y}(n)$, $\boldsymbol{Q}(n)$ are available, then check (13), if satisfied, then a new rule is added using (14)−(17). Otherwise, go to next step.

Step 3 Parameters updating step–$\sigma(n)$ is obtained from (30). $\boldsymbol{V}(n+1)$ is calculated from (32).

Step 4 Pruning rule step–Check (33). If it is satisfied, then such rule is pruned.

Step 5 Prediction step–$\hat{\boldsymbol{y}}(n+1)$ is obtained by (11) using updated BRB.

Step 6 Fault identification step–Check (34). If it is satisfied, stop. Otherwise, let $n = n + 1$ and go to *Adding rule step*.

4 A practical case study

In this section, a practical case study is examined to validate the proposed model and to show the application potential in engineering practice. As a key device of the inertial navigation system in missiles and space equipments, gyro plays an irreplaceable important role and its operating state has a direct effect on navigation precision. According to the statistic analysis, it is shown that almost 80% faults of inertial navigation system result from gyros. As gyroscopic drift rises, the gyro performance begins to degrade, resulting in the excitation of gyroscopic faults. As such, building reliable and cost-effective prediction model is welcome and desired in recent years.

In this study, the drift data are collected in a gyroscope performance reliability test. For the gyroscope drift test, some technical parameters include the sampling interval T, the acceleration of gravity g, and the geographic latitude R. In this experiment, $T = 2.2\,\mathrm{h}$, $\mathrm{g} = 9.7941\,\mathrm{m/s^2}$ and $R = 34.6006°$. After the experiment, we can collect all the drift data. The data sets include the time-to-drift data for 90 suits of gyroscope. The experiment results are illustrated in Fig.1.

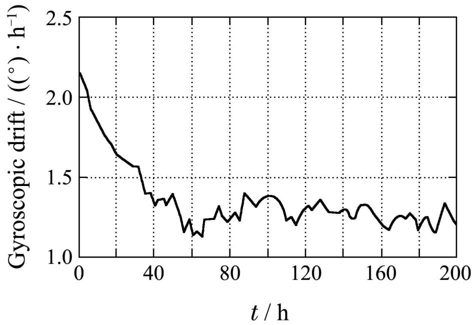

Fig. 1 All gyroscopic drift data collected in the test

As shown in Fig.1, it indicates that the gyroscopic drift is a time series. For illustration purpose, in this study six antecedent attributes are selected as an example. Consequently, we can transform 90 observed values to 84 sets of input-output patterns accordingly. Then the belief rule can be represented as follows:

$$R_k:\ \text{If } y_{t-1} \text{ is } A_1^k \wedge \cdots \wedge y_{t-6} \text{ is } A_2^k,$$

Then drift at next step is y_t with

$\{(D_1, \beta_{1,k}), \cdots, (D_N, \beta_{N,k})\}$ involved θ_k and $\delta_1, \delta_2, \cdots, \delta_6$,

where $R_k (k = 1, \cdots, L)$ is the belief rules of the BRB. In the BRB, A_i^k is the referential points of y_{t-i}, $i =$

$1, 2, \cdots, 6$, respectively. $D_l (l = 1, \cdots, N)$ are the assessment grades of system behavior. In this case study, the technical requirement of this kind of gyroscope is that the drift value is not larger than $2.4(°)/\mathrm{h}$, so the linguistic labels for $y_{t-1}, y_{t-2}, \cdots, y_{t-6}$ are defined as 'Small'(S), 'Medium'(M), and 'Large'(L) and $\bar{\boldsymbol{y}} = 2.4$. That is to say that $A_i^k \in \{\mathrm{S}, \mathrm{M}, \mathrm{L}\}$ for $i = 1, \cdots, 6$. And then three assessment grades of fault assessment of gyro are used: 'Good'(G), 'Average'(A), and 'Poor'(P), this is to say, $\boldsymbol{D} = (D_1, D_2, D_3) = (\mathrm{G}, \mathrm{A}, \mathrm{P})$. As such, if we enumerate all belief rules, there are 3^6 rules, which is difficult to apply in practice. Therefore, the evolving BRB method is a must. The referential points of inputs defined above are in linguistic terms and thus need to be quantified. In [16], a scheme to convert other inputs to belief structure has been developed. 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. In this case, we set the referential points as $\bar{\boldsymbol{\gamma}}(n) = [1.0 \quad 1.4 \quad 2.4]^{\mathrm{T}}$, corresponding to $A_i^k \in \{\mathrm{S}, \mathrm{M}, \mathrm{L}\}$. Since drift data is a time series, we set the utility used of $\boldsymbol{D}$ as used in (11) to be $\mu(D_1) = 1, \mu(D_2) = 1.4, \mu(D_3) = 2.4$, which are the same as $\bar{\boldsymbol{\gamma}}(n)$.

According to fault prognosis algorithm in Table 1, we first set the initial BRB involved the following two rules:

R_1 : If y_{t-1} is $S \wedge y_{t-2}$ is $S \wedge y_{t-3}$ is
$S \wedge y_{t-4}$ is $S \wedge y_{t-5}$ is $S \wedge y_{t-6}$ is S,
Then drift y_t at next step is
$\{(D_1, 0.95), (D_2, 0.05), (D_3, 0)\}$
with $\theta_1 = 1$ and $\delta_1 = \delta_2, \cdots, \delta_6 = 1$,

R_2 : If y_{t-1} is $L \wedge y_{t-2}$ is $L \wedge y_{t-3}$ is
$L \wedge y_{t-4}$ is $L \wedge y_{t-5}$ is $L \wedge y_{t-6}$ is L,
Then drift y_t at next step is
$\{(D_1, 0), (D_2, 0), (D_3, 1)\}$
with $\theta_1 = 1$ and $\delta_1 = \delta_2, \cdots, \delta_6 = 1$,

and $\alpha = 3, \gamma = 0.4, \varepsilon = 0.5, e_{\mathrm{g}} = 0.01, e_{\mathrm{p}} = 0.05$ while $\phi_{\mathrm{th}} = 0.3$. Then using algorithm summarized in Table 1 for 84 sets of drift data, we can obtain the predicted drift value to achieve fault prognosis through (11). The predicted results are illustrated in Fig.1. As shown in Fig.1, it is obvious that the predicted outcomes generated by the proposed method can trace the changes of the gyro drift well. In order to further demonstrate the accuracy of the proposed recursive algorithm, the mean absolute percentage error (MAPE) and root-mean-square error (RMSE) are used. Through calculation, the MAPE and RMSE between the actual values and the predicted values generated by the proposed method model is 4.11% and 0.0828, respectively.

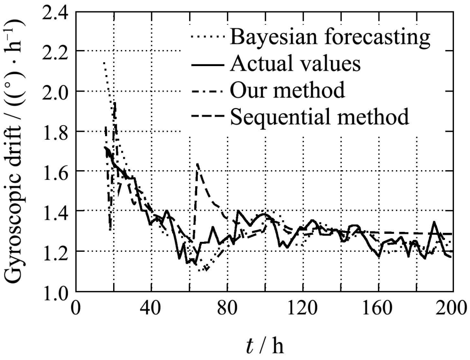

Fig. 2 Experimental results with different methods

It may be of interest to compare the results generated by the proposed method with the sequential learning method[12] and the classical Bayesian forecasting method[24–25], since the first one uses the same model structure and adjusting mechanism as adopted in this paper while the last one is based on the Gaussian distribution and independent assumption. In the simulation, the parameters of Bayesian forecasting method in [24] are updated using Kalman filter. The simulated results are illustrated in Fig.2 and the comparison results are summarized in Table1.

Table 1 Comparison of the predicted results

	The proposed method	Ref. [24]	Ref. [12]
MAPE	0.0411	0.0481	0.0553
RMSE	0.0828	0.0910	0.1022

Table1 shows that the proposed method has better prediction performance in this case study than other two methods in terms of prediction accuracy. One possible cause for lower prediction accuracy of sequential method is that it requires the input of BRB to obey the uniform distribution. However, it is clear that such requirement cannot be satisfied in this case study as shown in Fig.1. The performance between our method and sequential method is subtle. However, according to fault identification criterion (34), Bayesian forecasting method will generate false alarm at the first predicted point since there is no fault in our case study, as shown in Fig.2. This fact also provides evidence for the forecasting performance of the proposed fault prognosis method, and further demonstrates the effectiveness and feasibility of the developed evolving BRB method in practice.

5 Conclusions

This paper is concerned with constructing an evolving BRB system to achieve fault prognosis. First, we give a definition of belief completeness criterion to represent the activated degree of belief rule given the inputs. Then along the line of the sequential adaptive learning algorithm, we develop an improved evolving BRB learning algorithm based on the defined criterion, in which the structure and parameters of BRB can be adjusted online but without other additional requirement for the sample density function of inputs. Compared with the other learning algorithms, a belief rule can be automatically added into the BRB or pruned from the BRB without the need of the sample density function. In addition, our algorithm inherits the feature of the BRB, in which only partial input and output information is required. In order to verify the effectiveness of our algorithm, a practical case study of gyro fault prognosis are studied and examined to demonstrate how our algorithm can be implemented. Compared with Bayesian forecasting method and sequential learning method, the proposed method in this paper can generate more satisfactory results.

References:

[1] PECHT M. *Prognostics and Health Management of Electronics* [M]. New Jersey: John Wiley, 2008.

[2] WANG W, HUSSIN B. Plant residual time modeling based on observed variables in oil samples [J]. *Journal of the Operational Research Society*, 2009, 60(6): 789 – 796.

[3] SI X S, WANG W, HU C H, et al. Remaining useful life estimation-A review on the statistical data driven approaches [J]. *European Journal of Operational Research*, 2011, 213(1): 1 – 14.

[4] CHEN S M, CHUNG N Y. Forecasting enrollments using high-order fuzzy time series and genetic algorithm [J]. *International Journal of Intelligent Systems*, 2006, 21(5): 485 – 501.

[5] SI X S, HU C H, YANG J B, et al. A new prediction model based on belief-rule-base for system's behavior prediction [J]. *IEEE Transactions on Fuzzy Systems*, 2011, 19(4): 636 – 651.

[6] HU C H, SI X S, YANG J B. Systems reliability forecasting based on evidential reasoning algorithm with nonlinear optimization [J]. *Expert Systems with Applications*, 2009, 37(3): 2550 – 2562.

[7] STACH W, KURGAN L, PEDRYCZ W. Numerical and linguistic prediction of time series with the use of fuzzy cognitive maps [J]. *IEEE Transactions on Fuzzy Systems*, 2008, 16(1): 61 – 72.

[8] ZHOU M, LIU X B, YANG J B. Evidential reasoning-based nonlinear programming model for MCDA under fuzzy weights and utilities [J]. *International Journal of Intelligent Systems*, 2010, 25(1): 31 – 58.

[9] YANG J B, LIU J, WANG J, et al. Belief rule-base inference methodology using the evidential reasoning approach – RIMER [J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 2006, 36(2): 266 – 285.

[10] YANG J B, LIU J, XU D L, et al. Optimal learning method for training belief rule based systems [J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 2007, 37(4): 569 – 585.

[11] ZHOU Z J, HU C H, YANG J B, et al. Online updating belief-rule-based system for pipeline leak detection under expert intervention [J]. *Expert Systems with Applications*, 2009, 36(3): 7700 – 7709.

[12] ZHOU Z J, HU C H, YANG J B, et al. A sequential learning algorithm for online constructing belief-rule-based systems [J]. *Expert Systems with Applications*, 2010, 37(2): 1790 – 1799.

[13] YANG J B, WANG Y M, XU D L, et al. The evidential reasoning approach for MADA under both probabilistic and fuzzy uncertainties [J]. *European Journal of Operational Research*, 2006, 171(1): 309 – 343.

[14] LIU Y, WANG H Q, YU J, et al. Selective recursive kernel learning for online identification of nonlinear systems with NARX form [J]. *Journal of Process Control*, 2010, 20(2): 181 – 194.

[15] LIU Yi, YU Haiqing, GAO Zengliang, et al. Online adaptation of kernel learning adaptive predictive controller [J]. *Control Theory & Applications*, 2011, 28(9): 1099 – 1104.
(刘毅, 喻海清, 高增梁, 等. 核学习自适应预测控制器的在线更新方法 [J]. 控制理论与应用, 2011, 28(9): 1099 – 1104.)

[16] YANG J B. Rule and utility based evidential reasoning approach for multi-attribute decision analysis under uncertainties [J]. *European Journal of Operational Research*, 2001, 131(1): 31 – 61.

[17] YANG J B, XU D L. On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty [J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 2002, 32(3): 289 – 304.

[18] WANG Y M, YANG J B, XU D L. Environmental impact assessment using the evidential reasoning approach [J]. *European Journal of Operational Research*, 2006, 174(3): 1885 – 1913.

[19] CHEN W T, SAIF M. A novel fuzzy system with dynamic rule base [J]. *IEEE Transactions on Fuzzy Systems*, 2005, 13(5): 569 – 582.

[20] RONG H J, SUNDARARAJAN N, HUANG G B, et al. Sequential adaptive fuzzy inference system (SAFIS) for nonlinear system identification and prediction [J]. *Fuzzy Sets and Systems*, 2006, 157(9): 1260 – 1275.

[21] CHUNG P J, BOHME J F. Recursive EM and SAGE-inspired algorithms with application to DOA estimation [J]. *IEEE Transactions Signal Processing*, 2005, 53(8): 2664 – 2677.

[22] DEMPSTER A P, LAIRD N, RUBIN D B. Maximum likelihood from incomplete data via the EM algorithm [J]. *Journal of the Royal Statistical Society, Series B (Methodological)*, 1977, 39(1): 1 – 38.

[23] TITTERINGTON D M. Recursive parameter estimation using incomplete data [J]. *Journal of the Royal Statistical Society, Series B (Methodological)*, 1984, 46(2): 257 – 267.

[24] HARRISON P J, STEVENS C F. Bayesian forecasting [J]. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1976, 38(3): 205 – 247.

[25] WEST M. Robust sequential approximate Bayesian estimation [J]. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1981, 43(2): 157 – 166.

作者简介:

司小胜 (1984–), 男, 博士研究生, 目前研究方向为故障诊断与预测、可靠性、预测与健康管理、寿命估计, E-mail: sxs09@mails.tsinghua. edu.cn;

胡昌华 (1966–), 男, 教授, 博士生导师, 目前研究方向为故障诊断与容错控制、可靠工程, E-mail: hch6603@263.net;

张 琪 (1981–), 女, 博士, 目前研究方向为故障预测与健康管理, E-mail: zhangqi6530@163.com;

李 娟 (1969–), 女, 教授, 目前研究方向为故障诊断与容错控制、复杂系统的分析和设计, E-mail: lijuan291@sina.com.