Automatic Modeling Method Based on Bayesian Network

Lu Han, Xianjun Shi and Taoyu Wang
Naval Aviation University, Yantai, Shandong, China
Email: 953339581@qq.com

Abstract. In fact, the relationship between fault and test is uncertain, and the traditional testability model cannot deal with uncertain information. Bayesian network model with uncertain information reasoning ability and self-learning ability has become a research hotspot. Aiming at the difficulty and high cost of Bayesian network testability model, a automatic modeling method based on Bayesian network was proposed. This method uses the circuit functional structure relationship characterized by EDA to provide the required conditions for the K2 algorithm, and verifies the results of the structure and parameter learning. Through this method, the Bayesian network model can continuously update the model structure and parameters by combining expert knowledge, historical experience and later use data, making the model more and more accurate. After simulation and comparison, the algorithm can effectively build a model and reduce the workload of related personnel.

1. Introduction
Testability model is a model established for designing, analyzing, and evaluating the testability of a system. It mainly describes the relationship between failure and testing. According to GJB 2547A, the use of test models throughout the equipment’s entire life cycle plays an important role in equipment demonstration, research and development, and maintenance [1]. At present, the most commonly used test models are related models, which mainly include multi-signal flow graph models and information flow models. They are relatively simple to model and have low computational complexity. They are widely used in equipment test design. These two models are mainly based on the failure mode impact and hazard analysis (FMECA) [2], and there are various ways to obtain FMECA: you can conduct experiments, learn from the failure information of similar products, or use statistics during the use of equipment. However, the confidence of the information collected by different channels is uneven and may even be highly subjective [3]. In addition, the above two models have insufficient processing and learning update capabilities for FMECA, have low utilization of FMECA information, cannot use the uncertain information therein, and have a high degree of model distortion.

The Bayesian network testability model utilizes the powerful modeling capabilities and inference mechanism of the Bayesian network, which can effectively handle fault-related information, and can continuously learn and update the network based on the new fault test data information in the later stage [4]. It is suitable for each stage of equipment use. However, Bayesian network model modeling requires more information about fault testing to determine prior and conditional probability information, which is difficult to obtain in practice; Moreover, the existing Bayesian network testability models focus on the correlation between fault and test, but pay less attention to the relationship between fault and fault, which makes the built model complex. This method mainly aims at testability modeling of electronic devices, and collects information needed for Bayesian network parameters and structure learning through EDA simulation [5]. Meanwhile, through the physical
structure of the system reflected by EDA, a naive Bayesian network can be established to learn on this basis, reducing the difficulty of learning.

2. Sample Collection Based on EDA
Electronic Design Automation (EDA) technology is a tool that uses a computer as a tool to complete circuit design through a hardware description language. The main EDA software is Pspice, Multisim, Saber and so on. With the EDA tool, users can use a computer to design a simulated circuit model and view the circuit status. In the field of fault diagnosis, EDA tools can help developers to simulate circuit failure and verify testability design. Zhang simulated the fault with Pspice, analyzed the test results, and established the relationship between the fault and the test [6]. Wang proposed a comprehensive fault modeling and injection method to improve the efficiency of fault injection [7]. Dai established a Bayesian network model through a multi-signal flow graph model [8]. On the basis of obtaining fault test correlation matrix through EDA simulation, Wen Ye established a hierarchical testability model of missile launch and control circuit. However, Bayesian network has strong learning ability. fault injection simulation based on EDA can provide the sample data for Bayesian network model to learn. At the same time, EDA can be used to obtain prior knowledge and reduce the difficulty of network learning. Through EDA simulation results, the model can be evaluated and the network structure can be optimized.

2.1. Distribution of Fault Sample Size
As mentioned above, most Bayesian network models focus on the correlation between faults and tests, but pay little attention to the correlation between faults, which leads to the bloated structure of model network and low efficiency. In order to solve this problem, the expression of fault definition is adjusted to some extent. In this paper, it is stipulated that the output signal is fault signal, and the component fails. The change is mainly aimed at this situation: the former component has a fault, and the fault is transmitted through the output signal, which affects the subsequent component. If the later component lacks robustness, the output is still a fault signal. At this point, there is no fault inside the component, but according to the regulations, the component is also in a fault state.

Due to this change, the determination and allocation of the failure sample size can refer to the sample determination and allocation method in testability verification technology. The prior information that researchers can collect at the early stage is mainly the failure rate or failure mode frequency ratio of similar products and components. According to the types of prior information, the stratified sampling distribution method based on failure rate is adopted here.

The estimation of the minimum sample size can be expressed by the following formula:

\[ R_L = a^n \]

Among them, \( R_L \) is the lowest acceptable value of the test index; \( a = 1 - C, C \) is the credibility; \( n \) is the sample size and is a positive integer.

\[ n_i = nC_{pi} \]

\[ C_{pi} = \frac{Q_i T_i \lambda_i}{\sum Q_i T_i \lambda_i} \]

\( Q_i \) is the number of the i-th unit; \( \lambda_i \) is the failure rate of the i-th unit; \( C_p \) is the relative frequency of failure; \( T_i \) is the working time coefficient of the i-th unit.

2.2. Sample Labeling Based on Bayesian Method
Through EDA simulation fault injection, monte Carlo simulation was conducted for each fault for several times to obtain the untreated initial samples of fault. The set of initial samples of fault implied the fault detection rate, omission rate, false alarm rate, fault transmission intensity, fault influence
degree and other information of each measuring point. The hidden information in the set is the sample mark on each sample. For example, sample 1 is injected with fault mode $F_1$ into component $M_1$. When components $M_1$ and $M_2$ fail, $T_1$ and $T_2$ are tested to detect the fault. Among them, it is easy to judge whether a fault has been detected at the measuring point. It can be judged according to the detection threshold of sensor signals at the measuring point. The threshold setting method and principle are as follows:

Through the monte Carlo simulation of the normal state of the circuit by EDA, the signal sample set under the condition of no fault of the circuit is collected, and the normal signal is normally distributed. Since the sample set is small, the Bayesian method is adopted to provide the signal reference so as to set the signal. The principle is as follows:

Theorem 1: The joint conjugate distribution of the normal distribution population parameter $(\mu, \sigma^2)$ is the normal - inverse Gamma distribution, $(\mu, \sigma^2) \sim N - IGa\left(v_1, \mu_1, \sigma^2_1, k_1\right)$. Then the posterior distribution is $(\mu, \sigma^2 | X) \sim N - IGa\left(v_2, \mu_2, \sigma^2_2, k_2\right)$

Part of the sample set was selected as the prior information, and the mean $\mu$, variance $\sigma^2$, of signal distribution mean $\mu$, and the mean $\sigma^2$, and variance $s^2_\sigma$, of signal distribution variance $\sigma^2$ were calculated by Bootstrap method [9]. Taking mean $\mu$ and variance $s^2_\mu$ as the expectation and variance of $\mu$, and mean $\mu^2$ and variance $s^2_2$ as the expectation and variance of $\sigma^2$, the normal-inverse GAMMA distribution parameters are obtained as follows:

$$v_1 = \frac{2(\mu^2)}{s^2_\mu} + 4$$

$$\mu_1 = \bar{\mu}$$

$$\sigma^2_1 = \frac{v_1 - 2}{v_1} \mu^2$$

$$k_1 = \frac{\mu^2}{s^2_\mu}$$

The remaining samples are treated as new data, $(\mu, \sigma^2 | X) \sim N - IGa\left(v_2, \mu_2, \sigma^2_2, k_1\right)$

$$v_2 = v_1 + n$$

$$\mu_2 = \frac{k_1}{k_1 + n} \mu_1 + \frac{n}{k_1 + n} \bar{X}$$

$$\sigma^2_2 = \frac{1}{v_1} \left[ v_1 \sigma^2_1 + n S^2 + \frac{k_1 n}{k_1 + n} (\mu_1 - \bar{X})^2 \right]$$

$$k_2 = k_1 + n$$

Set the loss function as the square error function, and get:

$$\mu_E = E(\mu | X) = \mu_2$$

$$\sigma^2_E = E(\sigma^2 | X) = \frac{v_2 \sigma^2_2}{v_2 - 2}$$

It can be seen that Bayesian method has the ability of fusion of prior information and sample data, and can continuously learn from new data in the later stage, making the posterior information closer and closer to the reality.
The threshold setting of measuring points is determined according to the distribution of signals and test indicators. In fact, it is also limited by the selected sensor itself and can be changed according to the conditions of actual sensor configuration. The signal exceeds the threshold, the test system identifies a fault, and the element in the sample is marked as a fault.

For the fault transmission between different components, due to the adjustment of the fault definition, the subsequent component output signal is detected, and the component state is determined by comparing with the detection signal threshold of the first-order correlation test. This setting causes the state of the element to be the same as the state of the first-order correlation test for the element; If the component has no order correlation test, the component in the sample can be marked according to the normal signal distribution obtained above.

3. Testability Model Based on Bayesian Network

The testability model based on Bayesian network, with the help of the uncertain information processing ability and learning ability of Bayesian network, makes the testability model closer to reality, which can be continuously updated in the later period [10]. Moreover, through the network, the evidence can be quickly processed and the fault diagnosis conclusion can be reached. Bayesian network learning is mainly divided into structural learning and parametric learning, in which the parametric learning is relatively mature, and structural learning is a hotspot in recent years.

3.1. Parameter Learning

Parameter learning is based on the determination of the Bayesian network node connection relationship, and continues to mine the dependencies between variables in the data to obtain a Conditional Probabilities Table (CPT). Since the sample set obtained through EDA analysis is complete, the mature Bayesian method can be used for parameter learning.

3.2. Structure Learning of K2 Algorithm

Parameter learning is based on the determination of the Bayesian network node connection relationship, and continues to mine the dependencies between variables in the data to obtain a Conditional Probabilities Table (CPT). Since the sample set obtained through EDA analysis is complete, the mature Bayesian method can be used for parameter learning.

The K2 algorithm is a greedy search algorithm based on scores [11], and briefly describes how it works: the currently selected node has no parent nodes, adding no more than a specific number of parent nodes to the node, scoring the structure. Select the structure with the highest score, and determine the node at this time as its parent. K2 algorithm needs to specify the order between nodes, and only one-way connections can be made between the nodes in the order before and after, so there is no need to do loop check, and it is not necessary to select the parent node for each node separately. K2 algorithm has excellent learning performance and is a classical algorithm for Bayesian network structure learning, but it needs to meet two prerequisites: (1) providing topological ordering of all nodes; (2) specify the maximum number of parent nodes in the network. These two conditions greatly limit the application of K2 algorithm. In this paper, through the circuit diagram and EDA analysis, in addition to obtaining the tag sample set, we can also understand the correlation between components and component fault test, provide some prior information for K2 algorithm, specify the topological ordering and the maximum number of parent nodes. Therefore, there is no need to improve the K2 algorithm in this paper, and the prior information provided by EDA can be incorporated. Using the K2 scoring function simplified by Bayesian scoring:

$$
\log(P(G, D)) = \prod_{i=1}^{n} \left\{ \sum_{j=1}^{q_i} \left[ \log \left( \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \right) \right] + \sum_{k=1}^{r_i} \log \left( N_{ijk}! \right) \right\}
$$
where \( n \) is the number of network nodes, \( r_i \) is the number of values of node \( x_i \), and \( q_i \) is the sum of values of parent nodes of node \( x_i \).

\[
N_{ij} = \sum_{k=1}^{q_i} N_{ijk}.
\]

### 3.3. Evaluation and Verification of the Model

Through the structure learning and parameter learning of the sample data set, a test model based on Bayesian network is established. There is a comparison method for the verification of this model. The model is evaluated and verified by comparison with other test models. In addition, there is a prior information test method. Theorem 1 was used to determine the distribution of normal signals and test thresholds. In the same way, using the Bayesian method, you can obtain the distribution of fault signals on the basis of small samples, weight the distribution of all fault modes of components according to the frequency ratio of the fault modes, obtain the fault signal distribution of the components, and finally get the test fault detection rate, false alarm rate, and other information. Through these prior information, the testability model of Bayesian network is verified.

### 4. Modeling Examples

An instrument's amplifier circuit is shown in figure 1. According to the method in this paper, the test model of its Bayesian network is shown in figure 2.

![Figure 1. Instrument amplifier circuit diagram.](image1)

![Figure 2. Bayesian network testability model for instrumentation amplifier.](image2)

The prior information obtained by EDA is used to verify and evaluate the model. The evaluation results are shown in table 1.

| Prior Information | Bayesian network testability model |
|-------------------|-----------------------------------|
| \( P(T_1 \mid A_1) \) | 95.0% | 93.9% |
| \( P(T_2 \mid R_2) \) | 95.0% | 94.1% |
| \( P(T_3 \mid R_1) \) | 93.2% | 92.3% |
| \( P(T_4 \mid A_2) \) | 88.7% | 87.2% |
| \( P(T_4 \mid A_2) \) | 79.3% | 80.2% |
5. Conclusion
The Bayesian network testability modeling method based on EDA and K2 structure learning algorithm proposed in this paper has the following advantages after testing:

(1) Compared with the past, the Bayesian network test model established by this method takes care of the association between failures and faults, and the model structure is closer to the actual physical structure of the device. While characterizing the fault test correlation, it also reflects the fault propagation information;

(2) Solve the difficult problem of Bayesian network modeling. Through software programming, model generation can be realized automatically. Reduce workload;

(3) The model uses the Bayesian theory’s ability to fuse prior information with sample information. Through experiments and use, the parameters and structure are continuously updated, and the model can be applied to different stages of equipment use.

References
[1] Zhao J, Zhao W J and Zhang Q 2014 Design and realization of an avionics equipment testability model based on TADS Applied Mechanics and Materials 3468.
[2] Liu G, Yang S, Qiu J and Yang P 2013 A novel testability model for health management of heading attitude system Chinese Journal of Aeronautics 26 (1).
[3] Yang S M, Qiu J, Liu G J, Yang P and Zhang Y 2012 Testability requirement uncertainty analysis in the sensor selection and optimization model for PHM Journal of Physics: Conference Series 364 (1).
[4] Zhu C S, Zhang Q, Su F T and Ran H L 2012 Research on the multi-objective optimization model of system-level bit testability index determination Applied Mechanics and Materials 1498.
[5] Kout A, Toure F and Badri M 2011 An empirical analysis of a testability model for object-oriented programs ACM SIGSOFT Software Engineering Notes 36 (4).
[6] Zhang Q 2012 Multi Signal Modeling and Software Implementation of Analog Circuits Based on Multi Features (University of Electronic Science and technology).
[7] Wang Y 2013 Research on Comprehensive Failure Mode and Effect Analysis (FMEA) Based on Saber (Beijing University of Posts and Telecommunications).
[8] Dai J, Yu J, Zhang P and Li X 2009 Diagnostic Bayesian network modeling based on multi signal flow graph Journal of Beijing University of Aeronautics and Astronautics 35 (04) 472-475.
[9] Yang P 2008 Diagnosis Strategy Optimization Design Technology Based on Correlation Model (National University of Defense Science and Technology).
[10] Wang X, Sun B, Lv Y and Li L 2011 System testability modeling and analysis based on Bayesian networks China Testing 37 (05) 90-93
[11] Shen L, Yu J, Tang D and Liu H 2016 Bayesian network automatic modeling based on graph model and learning algorithm Journal of Beijing University of Aeronautics and Astronautics 42 (07) 1486-1493.