

## **A survey on failure prediction methods of information systems**

Yin Yang<sup>1,a</sup>, \*Wenyi Li<sup>1,b</sup>, Xiaoxiao Xu<sup>1,c</sup>, Ruoxi Feng<sup>1,d</sup> and Lei Zhu<sup>1,e</sup>

<sup>1</sup>*School of Accounting, Wuhan Textile University, Wuhan, China*

<sup>a</sup>*cs\_yangyin@hust.edu.cn,*

<sup>b</sup>*newsone260@126.com,* <sup>c</sup>*Xylona1228@outlook.com,* <sup>d</sup>*ruoxipaper@gmail.com,* <sup>e</sup>*jennyju@foxmail.com*

*\*Corresponding Author*

### **Abstract.**

Through analyzing the process of failure of computer information systems and confirming the relation of fault, error, symptom and failure, this paper proposes the corresponding prediction methods of diagnosing, detecting, monitoring and tracking, the research and development of failure prediction methods of fault diagnosing, detected error reporting, undetected error evaluating, symptom monitoring and failure tracking are summarized. Finally, the realize principle and fundamental framework of failure prediction methods are given, the failure prediction methods of computer relevant application are analyzed and discussed, the advantages and disadvantages of failure prediction methods are also compared and discussed.

*Keywords: failure; symptom; error; fault; perdition methods*

### **Introduction**

It is inescapable to go through the failure problems during the rapid developing process of computer application. The failure will not merely bring tremendous damages to computer applications and economy, but more importantly, it may cause the mental damages to the relevant systems of computer application, statistics and information. To prevent these from happening, it has become extremely urgent to predict the failure problems of computer relevant application. Therefore, the methods to predict failure are bound to develop rapidly and be improved.

The prediction methods apply in all fields of information systems. There are plentiful factors which cause the failure to the formation systems and the complete failure is a process generally containing fault, error, symptom and failure. Hence

this paper will firstly distinguish and establish the definitions and relationships between fault, error, symptom and failure and then put forward the predicting method of related diagnosis, inspection, monitoring and tracking. Final, it will summarize the failure predicting method of information systems mainly based on fault diagnosis, non-detected error evaluation, detected errors report, symptom monitoring and failure tracking.

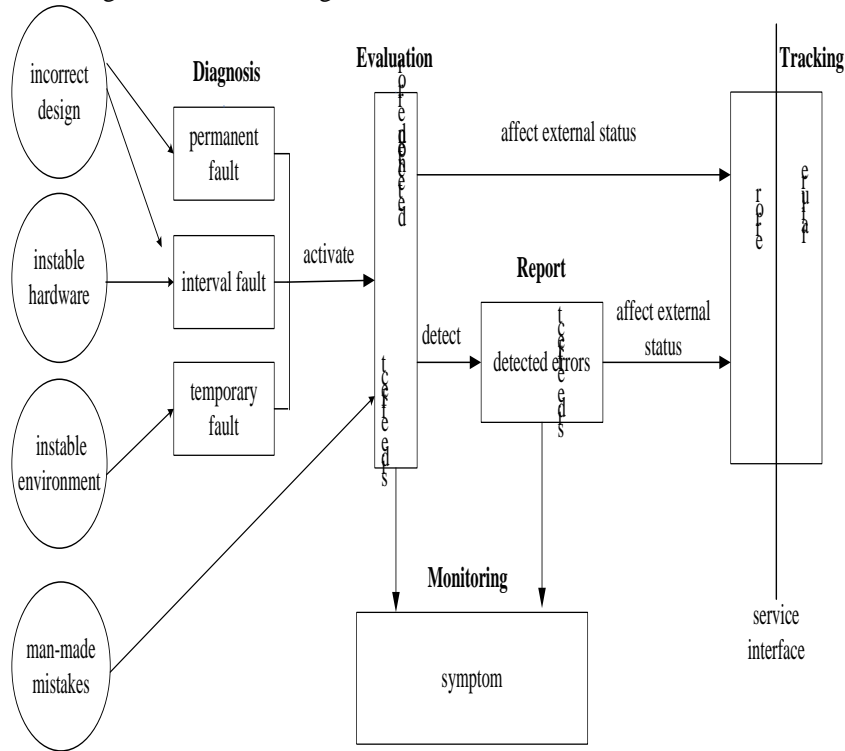


Fig. 1. The relationship among fault, error, symptom and failure.

Fault is the reason leading to errors but not all the faults that cause them. Fault is usually in static state. When the fault is activated by the system which is operating in some specific environment, the system will enter a wrong condition. In systematic mistakes, it is vital to know the degree of error was detected. It is impossible to detect errors completely and they can only be detected to some extent, thus errors can be divided into non-detected errors and detected errors [1]. If an error arises, it means that faults exist which is the deviation between validity and accuracy. Meanwhile, errors can arouse the action out of the systems and cause symptoms. There will come across a failure if certain function of systems are operating incorrectly due to an error when loses all or part of its functions, the situation tends to depart from the usual expectation and arrive at the service interface. The relationship among fault, error, symptom and failure is shown in Fig. 1. We can establish the classified method resembling Fig. 2 through the

predicting methods of researching fault diagnosis, non-detected error evaluation, detected errors report, symptom monitoring and failure tracking.

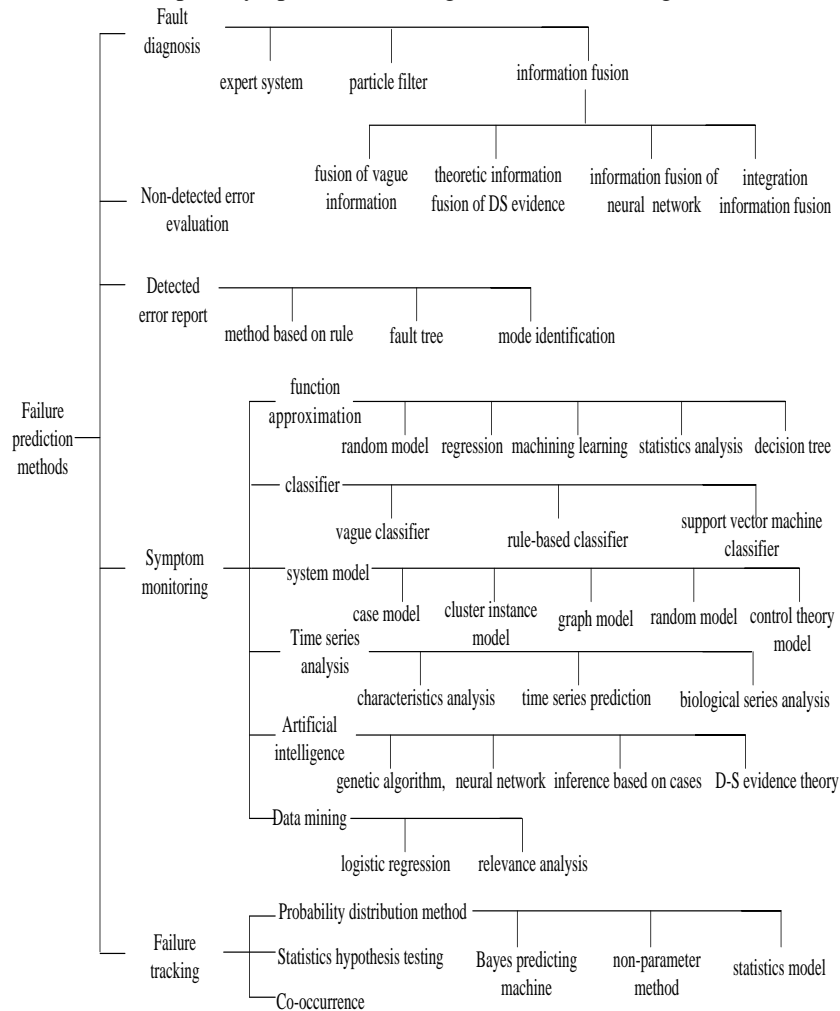


Fig. 2. The classified model of failure prediction methods.

## Fault Diagnosis

Fault which makes the system shows the unexpected characteristics can be understood as the nature of any abnormal phenomena of system. The diagnosis of fault utilize all types of the system's operating information and statuses and various of knowledge we have obtained now to work on the comprehensive processing of information and eventually gain the synthetic assessment of

system operating and fault's state . The major methods of faults' diagnosis include: expertise system, particle filter and information fusion, etc [2].

**Expertise System.** Expertise system is an intelligent system of computer programs which contains huge quantities of knowledge and experience that reach expert levels in a certain field. It can reasoning , judge and imitate the decision-making process of human specialists in order to solve complicated problems that need human specialists to tackle according to the knowledge and experience supplied by one or more experts in some fields.

**Particle Filter.** The particle filter estimates the system's situation and it will make pre-warning when discovering some abnormal phenomena (namely fault) happened to the system. It expresses its distribution through the random particle extracted from a posteriori probability, seeks for a group of random samples spread in state space to make it approximate the probability density function and replace the integral operation with sample average to get the distribution of the state minimum variance.

**Information Fusion.** The information fusion is the information process that data in multisensors are analysed automatically and then finish decision-making and evaluation in accordance with certain rules. It can be defined as a process of multiple layers and aspects to address detection, integration, estimation and combination of data with many resources in an effort to get accurate status, identification's estimation and complete and punctual threatening estimations. It can open up a new approach to settle complex systems' failure diagnosis by applying information fusion technique to fault diagnosis fields achieved by studying techniques of multiple information fusion and approaches to information with multiple dimensions.

## **Non-detected Error Evaluation**

Since errors cannot be detected entirely so the evaluation of non-detected errors is scarcely possible to achieve and at present we haven't found literature about the hardware's failure predicting method based on non-detected errors' evaluation. This part merely introduces several literatures about the failure predicting methods of software and the number of non-detected errors.

Software system may fail due to errors in demanding and design stages. To enhance the reliability of software, we need to study relative information in those two stages where some errors can be detected.

Ebrahimi comes up with a process that uses the frequency theorists' method to estimate the number of the non-detected errors [3]. The main strength of it is that it doesn't need to assume the independence between detectors. It sets up models of dependent detectors and meanwhile it uses nonparametric kernel level and Maximum-likelihood method to estimate the number of non-detected errors.

## **Detected Errors Report**

Detected events are always reported to users through the forms of information and signals when an error is detected. Using the incorrect report to be the way for inputting data must involve with the event-driven input data. Prior to now, the incorrect report has happened and it needs to analyze whether the confirmation of the wrong report will fail. The main method to detect incorrect reports includes: rule-based method, fault tree [4], identification mode, etc.

**Rule-based Method.** Rule-based method applies to a rule set; every rule set is made up by several error reporting. Failure prediction will happen once at least one situation of the rule set is accorded with. The aim of failure prediction arithmetic is to confirm rule set from the training data-set.

**Fault Tree.** Fault tree is a method that is to analyze from system to parts, then the components, following the declined trend. It analyzes the probability of faulted events' happening, through a branching diagram which is gradually expanded into a tree that drawn by the logic symbols. Fault tree carries out two stages: The first stage is the modeling and preparation for performing off-line. Collect the information about the influence that is sub-system's failure to system performance. Sub-system failure's reason is designed into the form of fault tree. The second stage is application. Installing a sensor on the system can provide the information of the present system performance. Therefore, the potential reasons can be deduced.

**Pattern Identification.** Pattern identification is the processes that describe, recognize, classify and explain objects or phenomena, through disposing and analyzing characteristic objects or phenomena's all kinds of information. Pattern identification puts forward that ranking value is distributed among continuous error reporting in the consecutive error events. If the distributed error reporting's ranking value is similar to pattern, the system will be failed.

## Symptom Monitoring

Symptom monitoring can ensure the coming failure through the recurrent analytic system's variable parameter, some errors influence system performance before they are monitored. Error information can be acquired by the system's side-effect (symptom), it can work out some failures that is non-failure to stop. These failures are difficult to obtain. The methods of symptom monitoring mainly include: function approximation, classifier, system model, analysis of time series, artificial intelligence, and data mining, etc [5].

**Function Approximation.** The basic question involved function approximation is function's approximation representation: find a function  $g$  among the elected functions; it is the approximation representation of the known function  $f$  in a certain sense. Through imitating target variable, the input data, as testing system's approximate function, can works out the given output data by approximate function. The method of function approximation mainly include: stochastic model, regression, machine learning, statistical analysis, decision tree, etc.

**Classifier.** The aim of classification is to structure a classification function or classification model based on data set's feature. This model can map the unknown category samples into someone of given category. The classifier evaluates system variable's existing value directly, judging if the system's current state is inclined to be failed or not through the classification. The judgment range of classification is from reference data-set, every data node can be used to judge that whether the failure will happen. The method of classifier is mainly including: vague classifier, rule-based classifier, support vector machine classifier, etc.

**System Model.** System model describes someone nature of system by a certain form. On the one hand, based on the different research purposes, the different system models can be established to the same system; On the other hand, the same system model can represent different systems. Comparing with classifier, it needs possessing the training data of failed tendency and non-failed tendency, relying on non-failed behavior model, and comparing the expected value calculated with model with the current test value of system. If the test value is different from the expected value apparently, the system may be at the abnormal state and it can predict the failure. The method of system model is mainly including: case model, the cluster instance model, stochastic model, graph model, control theory model, etc.

**Time Series Analysis.** Analysis of time series is a widely-used quantitative analysis method and statistic method of dynamic data processing. It is mainly used for describing and exploring quantitative regularity which is about the phenomenon is changing along with the time's development. Seeing the monitored continuous system variable as a time series, it means analyzing system variable's continuous samples. Analysis of time series determines if it is easy to be failed through calculating the current state's residual value. Analysis of time series' methods is including characteristics analysis, time series prediction, and biological series analysis.

**Artificial Intelligence.** Artificial intelligence is the science of technology studying, developing the methods, technology and application system which are imitating, extending and expanding human being's intelligence. Learn in the environment; make programming technology realized using computers. And then, the effective knowledge system can be formed. Thereby, carry out the knowledge learned from these environments to make the system appear effect of intelligence. The method of artificial intelligence is mainly including: genetic algorithm, neural network, inference based on case, DS evidence theory, etc.

**Data Mining.** Data mining helps people to promote data-application of the simple inquiry from low-level to the stage where mines knowledge from data, and then, provides decision-making. We know from the large, incomplete and blurry data. Above all, data mining is a process of extracting valid, previously unknown, comprehensible and actionable information or knowledge from the huge, incomplete, blurry and random practical data. The method of data mining mainly include: logistic regression, correlation analysis, etc.

## **Failure Tracking**

Failure tracking base on the situation which happened before can draw a conclusion of the possibility of forthcoming failure. This situation conclude the time and the type of failure happening. Method of failure tracking: probability distribution method, statistical hypothesis testing, co-occurrence [6].

**Probability Distribution Method.** Probability distribution method is to state the probability rule of choosing random variable. To make it easy, we will choose different pattern based on different random variable type to get the probability of the next failure. The method of probability distribution method mainly include: Bayes predicting machine, non-parametric method, and statistic model, etc. Bayesian predictor method puts forward two kinds of models, the first kind of Bayesian classifier, through the model prior probability set a weight for each naive Bayes model, through the expectation maximization algorithm iterative adjustment model and sub model probability prior probability. The proposed second methods on the same data set for training on standard naive Bias classifier.

**Statistical Hypothesis Testing.** The rank sum test is one of the basic concepts of statistical hypothesis test, based on the data according to the small to large array rank, rank sum method for the test of hypothesis. The process is first established to test the hypothesis, parameter does not involve assumptions; then RSR and calculate statistics; finally determine the probability value and conclusion. If the test the hypothesis, two groups of rank and should not be much difference.

**Co-occurrence.** Co-occurrence is the system in time or space nearest neighbor case possible failure, this situation can be used as a prediction system in the near future possible failure. If a failure occurs, so in a short period of time a failure is likely to occur; if some components have been reported more events, in this position may occur additional failure. Co-occurrence method uses the reliability, availability and serviceability of data log predictive IBM BlueGene/L failure. If a failure occurs, so in a short period of time a failure is likely to occur; if some components have been reported more events, in this position may occur additional failure.

## Summary

Our paper designed a classification method is easy to understand, better structure and classification of existing failure prediction technology, help researchers potential easy to find methods may attract their application environment. Failure prediction methods related to information systems according to process input data types are divided into five types: get status information (fault diagnosis); not be error detection failure (detection error report); through the evaluation study did not detect errors (non-detected error evaluation); monitoring data reflect the error causes the symptoms (symptom monitoring); analysis of previously failures (failure tracking). These five kinds of classification through the principle method is further subdivided into more sub class, give opinions each seed class a simple explanation, and gives a suitable reference.

## References

- [1] A. Avizienis, J. C. Laprie, B. Randell and C. Landwehr: IEEE. T. DEPEND. SECURE 1 (2004), p. 11
- [2] Y. N. Zhang, Q. W. Sun, H. Quan He, Y. G. Jin and T. F. Quan: J. SYST. ENG. ELECTRON 17 (2006), p. 495
- [3] N. Ebrahimi: IEEE. T. SOFTW.ARE ENG 23 (1997), p. 529
- [4] E. E. Hurdle, L. M. Bartlett and J. D. Andrews: RELIAB. ENG. SYST. SAFE 94 (2009), p. 1371
- [5] E. Kiciman and A. Fox: IEEE. T. NEUR. NETWOR 16 (2005), p. 1027
- [6] G. Hughes, J. Murray, K. Kreutz-Delgado and C. Elkan: IEEE. T. RELIAB 51 (2002), p. 350