Using data mining and root cause analysis method for failure analysis in electronic components

Hongjian Wang, Zhenbo Zhao, Zongbei Dai, Zeya Peng and Shaoping Li
Reliability Research and Analysis Center, China Electronic Product Reliability and Environmental Testing Research Institute, Guangzhou, 510610, P.R. China
E-mail: wanghongjian@ceprei.com

Abstract. The electronic components are the core of all electronic products, the reliability and maintainability directly affect manufacture products quality. And as products become more complex and more demanding. We need more reliable components. In order to solve this problem, we need to analysis how and why those components failed. Failure analysis cases are important data for analysis, data mining and improving reliability. China Electronic Product Reliability and Environment Testing Research Institute (CEPREI) is the earliest research organization engaged in reliability research and failure analysis in China, and has accumulated rich failure case data. By combining data mining methods such as Bayesian network, correlation analysis, we can perform statistical analysis and root cause analysis to assist failure analysis and improve product reliability. This paper will apply the failure analysis case of the electromagnetic relay, and explores and analyses related regular pattern.

1. Introduction
To improve product reliability and enhance product competitiveness, in addition to optimizing the design of product, it is also necessary to reversely analyse the cause of the product’s failure. Fault Tree Analysis (FTA) [1] and Failure Mode and Effect Analysis (FMEA) [2] are traditional and most popular reliability assessment methods. With these methods, product manufacturer can design more reliable products. FTA is a methodology to determine the potential causes of an undesired event, and FMEA analyses different failure modes and their effects on the system. Although these methods performed well, they are less effective in complex system [3].

In order to solve this problem, people introduced Bayesian networks [4], Markov chain [5], etc. BN perform the factorization of variables joint distribution based on the conditional dependencies [6]. BN can integrate important issues through design, produce, process, material, package, environment, stress and so on [7]. To sum up, BN can solve product design reliability issues.

Whether FTA or BN, they are qualitative analysis tools to aid design. FMEA is just analysis the effects about failure mode. None of them considered specific failure case data which is worth studying. Failure cases contain a lot of valuable data, failure analysis engineers will take a series of experiments during failure analysis to find the failure mode and failure mechanism. This process reveals the potential relationships between failure mechanism and environment stress [8].

Fault of machine is caused by various reasons, including environments factors, human factors, product design issues and other related factors. That makes the process of failure analysis extremely difficult, engineer need a lot of common sense to analysis and find failure reason [9]. It makes difficult
to cultivate failure analysis engineer and make the cost of failure analysis significant. So, it’s necessary to data mining of the cases of failure analysis to find some valuable rules.

Methods of data mining includes: classify, fitting and prediction, clustering, correlation analysis [10]. The goal of data mining is to find some invisible regular pattern and assist in analysis [11]. For example, we can use Bayesian dependence to find correlations between various factors; to predict most likely failure mechanism by Bayesian network. In short, through data mining, we can get useful knowledge and law [13,14,15].

CEPREI is the professional mechanism for failure analysis, and collect tens of thousands of cases of failure analysis, it’s a precious data to analysis, modeling and assessment. We extract 65 failure analysis cases of relay in this paper to investigate hidden information. All the test method meets the standard GJB548B and the physics of failure of relay. Fault tree of relay is established, we can find all the failure mode and failure mechanism in it.

Finally, the paper will introduce how to collect and preparation the dataset in section 2. At this part we take 100 failure analysis report to do the data mining, all the reports included is about failure relays. After that, we will use different data mining methods to modeling the data, mainly involve the extract, transform, load and data cleaning. The end of this paper will be summarized to conclusion and get recommendations for improving reliability of relay, we use FTA/Bayesian network to represent the fault of relay. All the fault reason and mechanism of failure of relay can be expressed on the Bayesian network and to get the quantify result of which fault reason is most likely to occur.

2. Preparation and data pre-processing
Our goal is investigating the potential rule in failure analysis. So, we collect all the related failure cases, failure analysis report, fault tree and physics of failure. We extract and refine features which is important to analysis. The features are: failure mode, failure mechanism, failure module, failure phase and the experimental results at each step.

2.1. Handling missing data
As we know, not every case has enough information because the difference of components, package form, root cause of failure and analysis process of engineers. There will be two different solutions to solve the problems.

Firstly, if the attribute of failure case is inherent but not written explicitly in the case, we should extract the attribute according to component model, component usage background information and so on. For example, we can get the package form information from datasheet, experiment results from experiment procedure.

Then, there are still much data that can not find in anywhere. It asks us to compute the dependencies between factors other than precise quantity. All we need is to calculate is the probability between the node in fault tree, which can be a massive aid in failure analysis.

2.2. Extract transformation and loading data
We sampling data from the electromagnetic relay failure analysis reports that are collected over the years. And only the typical of them about 64 cases will be extract to our dataset. These reports basically include all of the failure mode and failure mechanism of relay. These cases can reveal the effect among all the factors to failure, and the potential rule.

Due to all the reports are written by different failure engineers, they may use different term to express the same thing. It’s necessary to uniform the terminology, thereby we can get a concise dataset with a standardized database.

After we take ETL measures and cleaning data, we can get a standardized dataset which is easy to take data mining. The data we processed is focus on specific failure analysis related values and attributes that are likely to contain failure analysis knowledge.

Finally, we get a table refined by failure reports, has the attributes like: ID, sample name, type, manufacturer, package form, failure phase, failure mode, failure mechanism, background information,
conclusion, failure module, external inspection result, electrical test result, X-ray result, PIND test result, internal visual inspection result, seal test result, internal atmosphere test result.

3. Data mining
Here we take some common data mining methods to sought in-line with the dependability target such as failure root cause reasoning and training failure analysis engineer. We will first take advantage of conventional probability statistics to explore dependencies between different cases of failure analysis. Then we establish data model by Bayesian network and hidden Markov chain, these models can unearth some more important information.

3.1. Results of statistics
We can see the overview of our dataset in Table 1, the failure mechanism and failure mode are collected from 64 failure reports. In addition to this, there are attributes include failure module, experiments results, manufacturer, type, trademark and so on.

**Table 1.** overview of the failure reports

| Failure mechanism | Parameter drift | Contact sticking | short | Functional failure | Poor contact | open | leakage | normal | sum |
|-------------------|----------------|------------------|-------|--------------------|--------------|------|---------|--------|-----|
| arcing            | 4              |                  | 1     | 5                  |              |
| Foreign object    | 9              | 1                | 1     | 2                  | 2            | 9    | 1       | 25     |
| corrosion         |                | 1                | 1     | 2                  |              |
| Process defect    |                | 1                | 2     | 3                  |
| Electrical overstress |          | 2              | 2     | 2                  |
| Mechanical stress |                | 2                |       | 2                  |
| Metal electromigration |           | 1              | 2     | 3                  |
| Thermal stress    |                | 3                |       | 3                  |
| Water vapor infiltration |             | 1              |       | 1                  |
| normal            |                | 14               | 14    | 64                 |

At first glance of the table, we can get some intuitive observations like most failure mechanism is foreign objects enter the relay interior. And the most common failure mode is open. On this basis, we can proceed the slice, drill and rotate on the dataset to find more information.

Due to electromagnetic relay have two main parts which are load and control module. We can see the proportion of each module failure in figure 1. It’s clear that the load module has the highest failure rate. And we can find the foreign object is the most frequency mechanism when we continue to slice the load part. Obviously, the foreign object is the main reason of the failure in the module of load. It
means when we get a failure relay, we can first consider the whether it has some foreign object internal especially when the relay is abnormal of electrical parameters.

It’s complicated to analyze the failure data because it has much data missing though we take the ETL measures when we take failure analysis. We will take Apriori algorithm [12] to analysis the association rules of the attributes in failure analysis report.

Here we concentrate on failure mechanism. It’s easy to calculate the support and confidence of the 1-itemsets of failure mechanism. This result is obtained by table 1, the last column of the table can be calculate separately to sum and get the percentage. We can see the result in table 2.

| Table 2. the support of item failure mechanism |
|-----------------------------------------------|
| Foreign object | 39.06% |
| normal         | 21.88% |
| Electrical overstress | 9.38% |
| arcing         | 7.81%  |
| Metal electro-migration | 4.69% |
| Thermal stress | 4.69%  |
| Process defect | 4.69%  |
| corrosion      | 3.13%  |
| Mechanical stress | 3.13% |
| Water vapor infiltration | 1.56% |
| sum            | 100.00% |

Then we choose the top3 mechanism to calculate the support and confidence to other attributes. We just list a few interesting result below, which exclude the common sense like the dependencies between seal test and internal atmosphere test.

No matter what failure mechanism is, the most likely external inspection result is normal, it means we can not distinguish the mechanism by the outward appearance. Furthermore, we can neither find out the mechanism by X-ray test. No conclusion can be drawn from observation alone. Foreign material adhering on contact surface is the reason of foreign object, and it may cause the circuit open.

After we take the Apriori algorithm for the dataset, we can get a intuitive impression of how the relay fault and the reason of it failure. This points the way for our work and will make our job easier.

3.2. Modeling data by fault tree and Bayesian network

In essence, failure of component is caused by various factors. The faulty tree shows qualitative relationship of various factors. On this basis, Bayesian network expressed a more concise qualitative
calculation method. It’s easy to compute the potential cause of an undesired event, referred to as the top event. And using Boolean algebra to express the derived for the top event in terms of combinations of primary events. The probability of minimal cut-sets can express the occurrence of the primary event. But in the complex system, it difficulty to compute by fault tree cause the incomprehensible Boolean algebra, so a lot of researchers put forward the Bayesian network to replace the fault tree. BN is suitable to express and analyse the uncertain knowledge based on bidirectional inference mechanism and the state description. The bidirectional reasoning mechanism of BN can represent the gates and event of FT [16][17].

The faulty tree of relay is shown below in figure 2. We can compute the criticality of the cut-set, the possibility of primary event. We can map FT to BN to obtain more information. The dependencies between nodes of the BN and our failure case dataset will provide the quantitative result. First, we need to map the FT to BN according to methods, we transform the root of FT to top event in BN, and take dependencies relationship to represent the AND/OR gate in FT. When we get the BN model, we can take the qualitative and quantitative analysis to compute the unreliability of the TE and the unreliability of the given subsystem. The conditional probability table (CPT) are filled by the Boolean gates to compute the overall unreliability of the system. But it’s difficult if we just know the failure top event to calculate the most likely reason. In BN, if we want to take quantitative calculation, we must know the prior probability which is experienced and not precise. This reflects the importance of the failure cases and the data refined by them. According to the failure phenomenon and their reason, we can construct the posterior dependencies probability of the BN. Once we have the posterior dependencies, it can be used to find the most likely failure mechanism, pointing the way for failure analysis and improve reliability by probability of why it’s fault.

Let’s take a look at the BN we transformed in figure 3. The posterior probability is filled directly in the node of BN. This figure revealed all the cause of failure and how it may influence the reliability. We can also compute the criticality of every cut-sets by Bayesian dependency to take the similar operate of the BN. Each node and the arc of the BN we put in figure 3 represents the event and their dependencies, it’s easy to find each mechanism and their influence to root event.
Figure 3 shows the dependencies of node in fault tree, that makes failure analysis easier and can have a general direction when a electromagnetic relay fails. For example, if we have a failed relay, and want to find the reason of failing, we just need to analyze step by step follow the fault tree, find the most likely posterior probability in the FT’s node to seek the failure mechanism of the relay. If a relay is circuit open of contact, we will know the most likely cause is contact adhesion non-conductive excess, then we can concern this reason to saving time and resource, it can be an more effective method in failure analysis, and can be auxiliary to intern engineer.
4. Conclusions
This paper mainly proposed some data mining methods and models in failure analysis. We first introduced the data preparation methods. These methods reference the conventional data mining methods like extracting, transform and loading data, handling missing data, data standardization, data cleaning and so on. Second part of the paper proposed some statistical learning algorithm such as
Apriori, Bayesian network to the organized data. These methods deal with the data help us get some interesting and potential conclusion.

Furthermore, we can explore other electrical component’s failure reports by the same data mining methods until we get a useful tool in the area of electrical components failure analysis. The quantity of data should to continue to accumulate from failure report, and the experience obtained failure cases can help to take failure analysis easier.

Acknowledgements
This work was financially supported by the National Natural Science Foundation of China (NSFC, 21801048), thanks to the National New Material Testing and Evaluation Platform-Electronic Materials Industry Center (by MIIT).

References
[1] Haasl D F, Roberts N H, Vesely W E and et al 1981 Fault tree handbook
[2] Lee B H 2001 Using FMEA models and ontologies to build diagnostic models
[3] Weber P, Medina-Oliva G, Simon C and et al 2012 Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas J. Engineering Applications of Artificial Intelligence 25(4) 671-82
[4] Weber, P, Jouffe and et al 2003 Reliability modelling with dynamic bayesian networks J. IFAC Proceedings Volumes
[5] de Souza e Silva, Edmundo, Ochoa and Pedro Mejía. State space exploration in Markov models J. Acm Sigmetrics Performance Evaluation Review 20(1) 152-66
[6] Philippe Weber, Lionel Jouffe and Paul Munteanu 2004 Dynamic Bayesian Networks modelling the dependability of systems with degradations and exogenous constraints/ 11th IFAC Symposium on Information Control Problems in Manufacturing INCOM'04. Salvador-Bahia, Brazil
[7] Weber P and Jouffe L 2006 Complex system reliability modelling with Dynamic Object Oriented Bayesian Networks (DOOBN) J. Reliability Engineering & System Safety 91(2) 149-62
[8] Gergely Pintér, István Majzik and András Pataricza 2012 Application of Data Mining in Failure Data Analysis
[9] Donca Gheorghe 2010 Applying data mining for root-cause analysis of machine faults in complex continuous process J. Analele Universităţii din Oradea, Fascicula: Ecotoxicologie, Zootehnie şi Tehnologii de Industrie Alimentară pp 369-72 ref.11
[10] Wang X Z, Chen B H and Mcgreavy C 1997 Data Mining for Failure Diagnosis of Process Units by Learning Probabilistic Networks J. Process Safety & Environmental Protection 75(4) 210-6
[11] Amooee G, Minaeibidgoli B and Bagheridehnavi M 2011 A Comparison Between Data Mining Prediction Algorithms for Fault Detection(Case study: Ahanpishegan co.) J. 8(6)
[12] Srikant R and Agrawal R 1996 Mining Sequential Patterns: Generalizations and Performance Improvements J.
[13] Mi J, Li Y F, Peng W and et al 2018 Reliability analysis of complex multi-state system with common cause failure based on evidential networks J. Reliability Engineering & System Safety 174 71-81
[14] Deng X and Jiang W 2018 Dependence assessment in human reliability analysis using an evidential network approach extended by belief rules and uncertainty measures J. Annals of Nuclear Energy 117 183-93
[15] Liu R, Yang B, Zio E and et al 2018 Artificial intelligence for fault diagnosis of rotating machinery: A review J. Mechanical Systems and Signal Processing 108 33-47
[16] H. Wang, L. Liu, Y. Wang and Z. Peng 2019 An Overview of Failure Analysis Expert System Based on Machine Learning International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE) Zhangjiajie China 584-9

[17] H. Wang, L. Liu, Z. Peng and Y. Wang 2019 An overview of failure analysis expert system based on Bayesian networks 2019 IEEE 26th International Symposium on Physical and Failure Analysis of Integrated Circuits (IPFA) Hangzhou China 1-5