Research on reliability data mining technology of electronic components

Cui Yaru¹, Zhang Kui²*, Liu Huaguang², An Shengbiao³, Huang Jie²
¹Hebei College of Industry and Technology, Shijiazhuang, China
²National Semiconductor Device Quality Supervision and Inspection Center, Shijiazhuang, China
³Hebei Gigantech Electronic Technology Co., Ltd, Shijiazhuang, China
cuiyaru1909@126.com
test_tech@teah.net
zyjcdesign@126.com
²zhangk@cetc13.cn

Abstract. This paper deeply analyzed the types of information that can be mined in component reliability data. According to specific needs, the clustering measurement method of reliability data sample similarity is studied, the distance measurement strategy is obtained, the algorithm steps and analysis methods of Apriori association rule data mining are improved, the reliability data prediction and analysis methods based on statistical methods are compared. AR prediction model, MA model and other component reliability information time prediction methods have formed data mining software such as clustering, correlation analysis and prediction, discovered a large amount of hidden reliability information, and realized the comprehensive mining and utilization of electronic component reliability data.

1. Introduction
Since the 21st century, the level of electronic information technology has grown rapidly, society has entered a climax of technological development. Advances in technology promote the development of electronic components, which covered military, civil, aerospace, mobile, communications, computer networks and other fields, almost all projects are inseparable from the use of electronic components.

The basis of high-reliability equipment is high-reliability of electronic components, whose reliability directly determines the performance and weapon reliability of electronic equipment such as electronic countermeasure and radar. No reliable quality components, no reliable product. In fact, the reliability data of various electronic components are inability to provide timely and rapid reference due to the accumulation of excessive and unsystematic data.

With the rapid development of artificial intelligence, Big Data technology is widely used in various fields, such as voice, image and medicine[1]. McKinsey & Company had discovered that the massive personal information recorded on various online platforms has potential commercial value, so it invested a lot of manpower and material resources to conduct research. In June 2011, it released a report on "Big Data", which analyzes in detail key technologies, application areas, the impact of "Big Data", etc[2].
In the past five years, China has vigorously promoted the development of Big Data, including Big Data engineering, Big Data technology, Big Data application [3]. According to the ten Big Data predictions released by the Big Data Committee: Big Data has more "value", the form of Big Data is diversified, Big Data system security, Big Data prediction, Big Data visualization, etc. the rise of Big Data and the system environment of Big Data are more comprehensive [4,5].

Therefore, according to the current research status of reliability analysis and data mining, this article selects representative electric vacuum devices, hybrid integrated circuits, optoelectronic devices and other components and products to conduct corresponding reliability information data mining technology research.

2. Requirement analysis of component reliability data mining

In the design and selection of electronic components, the user units of components are concerned about the batch consistency, replaceable products and their reliability, and single / multi batch statistical information in addition to the performance index and quality level of the products. They hope to infer or predict the failure rate, service life and qualification rate of the components through the existing information, and finally complete the required components by comparison Preferred. Except for the product quality grade and reliability test results, the above reliability information can’t be obtained by conventional means. According to the survey and statistics, the user units have the demand of reliability data mining in the following aspects:

1) Same model product quality consistency
2) Recommended alternative products
3) Presumption of device reliability correlation
4) Information prediction such as product qualification rate
5) Input historical information for prediction
6) Statistical analysis of single/multi-batch products
7) Product reliability prediction
8) Product horizontal comparison and optimization

Therefore, in view of the gaps in this field, this article conducts in-depth research on data mining sub-fields such as clustering technology, association rule technology, and prediction technology corresponding to various needs, and completes each functional module. The specific system composition is shown in Figure 1.
3. Research on Reliability Data Mining Technology

3.1. Cluster analysis technology

According to different attributes of reliability information, the data types of cluster analysis can be divided into numerical type, discrete type, etc. The more typical ones are: continuous data, whose attribute values and characteristic values can take any value, generally only expressed by numerical values, which can be directly performed mathematical operations; discrete data, whose characteristic values need to be taken within a limited range, Attribute values and characteristic values are represented by numerical values, symbols or letters [6]. Because there are so many types of component reliability information, it is necessary to select different sample distance calculation methods to analyze separately according to the characteristics of various data. The distances used in this article include:

1) Euclidean distance
The normal form in multidimensional Euclidean distance space is defined as follows:

\[ d([x_1, x_2, \ldots, x_n], [y_1, y_2, \ldots, y_n]) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \]  

2) Minkoski distance

\[ d_{ij}(k) = \left[ \sum_{i=1}^{n} |x_i - y_i|^k \right]^{1/k} \quad (i, j = 1, 2, \ldots, n) \]

3) Jaeger's distance
The calculation formula is as follows:

\[ d_{ij}(J) = \left[ \sum_{k=1}^{n} \left( x_{ik} - \sqrt{\sum_{j=1}^{k} \sqrt{x_{jk}}} \right)^2 \right]^{1/2} \]

4) Oblique space distance

If there is a relationship between variables, the result of Euclidean distance is not accurate enough. For samples I and j, the oblique space distance can be used as the classification scale.

\[ d_{ij} = \left[ \frac{1}{q^2} \sum_{a=1}^{q} \sum_{b=1}^{q} (x_{ia} - x_{ja}) (x_{ia} - x_{ja}) r_{ab} \right]^{1/2} \]

5) Langmuir distance

\[ d_{ij}(L) = \frac{1}{p} \sum_{a=1}^{p} \frac{|x_{ia} - x_{ja}|}{x_{ia} + x_{ja}} \quad i, j = 1, \ldots, n \]

After calculating the distance, the distance matrix can be obtained:

\[ D = \begin{bmatrix}
  d_{11} & d_{12} & \ldots & d_{1n} \\
  d_{21} & d_{22} & \ldots & d_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  d_{n1} & d_{n2} & \ldots & d_{nn}
\end{bmatrix} \]

After the sample distance conversion of reliability information is completed, the information clustering needs to be realized by clustering algorithm. According to different sample distance calculation methods, this paper selects different clustering algorithms

1) K-means algorithm based on partition
The k-means algorithm is described in Figure 2.

The processing process of K-means algorithm is as follows:

① K objects are randomly selected to represent the center of the initial clustering.
② The remaining objects are divided into clusters with their nearest cluster centers.
③ The mean value of each cluster is recalculated to obtain a new cluster center value.
④ Repeat the process until each cluster is no longer changed.

2) Hierarchical clustering algorithm

The algorithm decomposes the data set in hierarchy until it reaches the termination condition. The algorithm is shown in Figure 3.

3) Density based DBSCAN algorithm

The algorithm has good noise resistance and low requirement for the shape of reliability data set [7]. The basic steps of DBSCAN algorithm are as follows:

① Select a data point O arbitrarily.

② If O is not a core object, it is marked as an isolated point. If O is the core object, all points in its neighborhood form a cluster C, and these points are stored as potential objects.

③ Check whether the point in the potential object is the core point, and if so, divide it into cluster C.

④ Repeat the above process until the potential object database is cleared.

3.2. Association rule technology

Apriori algorithm is the most widely used association rule algorithm. It uses iterative layer by layer search method, which is simple and easy to use. However, it also has defects such as frequent scanning of database, affecting operation efficiency, large number of candidates and so on. Therefore, this paper improves it and optimizes the flow chart of the algorithm.

The correlation coefficient refers to the index of the degree of correlation between variables. The correlation coefficient between the i-th device sample and the j-th device sample is:
### 3.3. Data prediction technology

When selecting and purchasing electronic components, users usually know the reliability information of the past or known test conditions of the device, but there is no corresponding reliability information for other test conditions or products not yet produced. Therefore, it is necessary to study the data prediction technology, extract the past product qualification rate and other reliability information from the massive reliability data, and predict the device reliability information through the prediction algorithm combined with the mathematical model in reliability physics, so as to provide more reliability information for users [8].

1) **Autoregressive (AR) prediction model**

The mathematical model is as follows:

$$X_t = \sum_{j=1}^p a_j X_{t-j} + \varepsilon_t$$  \hspace{1cm} (9)

Autoregressive model can describe the variables with certain relevance at different times, which is a dynamic model.

2) **Moving average (MA) model**

The model can be expressed as follows:

$$X_t = \varepsilon_t + \sum_{j=1}^q b_j \varepsilon_{t-j}$$  \hspace{1cm} (10)

3) **Autoregressive moving average (ARMA) model**

The formula is as follows:

$$X_t = \sum_{j=1}^p a_j X_{t-j} + \sum_{j=0}^q b_j \varepsilon_{t-j}$$  \hspace{1cm} (11)

### 4. Construction of reliability data mining system

Through the above data mining technology research, combined with the traditional statistical theory and the national standard of reliability prediction, the reliability information data mining system of electronic components is built according to the system functional structure shown in Figure 1.

Among them, the data clustering of the same model of components can cluster the reliability test parameters of the same type of devices, and display the model, batch, quantity, qualified rate, consistency coefficient, clustering parameters, one-dimensional data chart, clustering tree and two-dimensional clustering diagram in order to obtain the advantages and disadvantages information of device consistency, as shown in Fig.4. At the same time, the performance and reliability indexes of different types of devices can be clustered to obtain the substitutability of devices and the similarity between different types of devices, so as to provide more references and choices for the whole machine users.

In the association rule analysis, for example, the correlation analysis of the HSG28S5 DC/DC converter found that the two reliability parameters after the HSG28S5 DC/DC converter test have a strong positive correlation in Fig.5.

In the statistical analysis, 408 CX661d GaAs FET were analyzed, and the non-normal distribution was found for the first time, as shown in Fig.6.
Fig. 4 Cluster analysis

Fig. 5 The correlation diagram of HSG28S5 DC/DC reliability parameters

Fig. 6 Histogram of GaAs FET
5. Conclusion
In this paper, we study the measurement method of reliability data sample similarity, improve the Apriori association rule data mining algorithm, study the auto regression (AR prediction model), moving average (MA) model and other time prediction methods of component reliability information, and get a comprehensive mining method of electronic component reliability information. The related technologies are applied to the development of data mining system, and the reliability data mining system of electronic components is formed. Cluster analysis of more than 1000 models such as HSG28S5 is realized. The correlation among reliability parameters such as cx661d is found. At the same time, it has discovered the non-normal distribution of parameters that have never been found in many products, which provides a good data resource for the reliability of electronic components.

References
[1] Chen K C, Chiang M, Poor H V. From technological networks to social networks [J]. IEEE Journal on Selected Areas in Communications, 2013, 31(9): 548-572.
[2] Siddiqa A, Karim A, Gani A. Big data storage technologies: a survey [J]. Frontiers of Information Technology & Electronic Engineering, 2017, 18(8): 1040-1070.
[3] Han J, Pei J, Kamber M. Data mining: concepts and techniques [M]. Elsevier, 2011.
[4] Chen K C, Mankir B, Huang S L, et al. Communication theoretic inference on heterogeneous data [C]. Communications (ICC), 2016 IEEE International Conference on, 2016: 1-6.
[5] John Lu Z. The elements of statistical learning: data mining, inference, and prediction [J]. Journal of the Royal Statistical Society: Series A (Statistics in Society), 2010, 173(3): 693-694.
[6] Chuang T Y, Lu J P, Chen K C. Communication theoretic prediction on networked data[C]. Communications (ICC), 2015 IEEE International Conference on, 2015: 1201-1206.
[7] Giannakis G B, Cendrillon R, Cevher V, et al. Introduction to the issue on signal processing for big data [J]. IEEE Journal of Selected Topics in Signal Processing, 2015, 9(4):583-585.
[8] Chen K C, Huang S L, Zheng L, et al. Communication Theoretic Data Analytics [J]. IEEE Journal on Selected Areas in Communications, 2015, 33(4): 663-675.